

**LA ESTADÍSTICA Y SUS APLICACIONES EN LA INGENIERÍA DE
PETRÓLEOS**

JORGE ENRIQUE PARRA CAMARON

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO – QUÍMICAS
ESCUELA DE INGENIERÍA DE PETRÓLEOS
BUCARAMANGA**

2012

**LA ESTADÍSTICA Y SUS APLICACIONES EN LA INGENIERÍA DE
PETRÓLEOS**

JORGE ENRIQUE PARRA CAMARON

Trabajo de grado para optar el título de INGENIERO DE PETRÓLEOS

Director

OSCAR VANEGAS ANGARITA

Ingeniero de petróleos

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO – QUÍMICAS
ESCUELA DE INGENIERÍA DE PETRÓLEOS
BUCARAMANGA**

2012

AGRADECIMIENTOS

El autor expresa su agradecimiento a:

Al director de la tesis, el ingeniero **Oscar Vanegas Angarita**, por ser un amigo incondicional, por su constante apoyo, motivación y por la orientación que me brindo en cada uno de los aspectos de este trabajo.

A mis amigos y a todas las demás personas que de alguna forma ayudaron a la culminación EXITOSA del proyecto.

DEDICATORIA

A mis maravillosos padres **Rosa Camarón Suarez** y **Alirio Parra Aguirre** por darme la oportunidad de estudiar y son ellos los que han hecho esto posible, ya que han soportado con mucha tranquilidad todos mis caprichos y serán los que disfruten todos mis logros durante mi vida profesional. A mis amigos de carrera y a todas aquellas personas que me acompañaron durante todo el camino recorrido.

En especial a mi tía **Miriam** y su esposo **paco**.

Muchas, muchas gracias a **Suly Díaz**, que durante los últimos meses me ha acompañado y apoyado en los momentos más difíciles de mi vida y por brindarme su amor incondicional.

JORGE ENRIQUE PARRA CAMARON

CONTENIDO

	Pág.
INTRODUCCIÓN	20
1. ESTADISTICA DESCRIPTIVA	21
1.1 PARÁMETROS DE TENDENCIA CENTRAL	21
1.1.1 Parámetros de tendencia central a partir de datos sueltos o no agrupados (muestras de tamaño menor a 30 datos)	22
1.2 PARAMETROS DE DISPERSION	28
1.2.1 Parámetros de dispersión a partir de datos sueltos o no agrupados	28
1.2.2 Parámetros de dispersión a partir de datos agrupados	31
1.3 PARAMETROS DE POSICION	33
1.3.1 Parámetros de posición a partir de datos sueltos o no agrupados	33
1.3.2 Parámetros de posición a partir de datos agrupados	35
1.4 PARAMETROS DE FORMA Y CONCENTRACION	37
1.4.1 Parámetros de forma	37
1.4.2 Parámetros de concentración	40
1.5 EJERCICIOS RESUELTOS	45
1.6 EJERCICIOS PROPUESTOS	55
2. PROBABILIDAD	62
2.1 TECNICAS DE CONTEO	62
2.2 ENFOQUES DE LA PROBABILIDAD A PRIORI	64
2.2.1 Enfoque clásico	64
2.2.2 Enfoque frecuencial	64
2.2.3 Enfoque subjetivo	64
2.3 RELACIONES ENTRE SUCESOS Y EVENTOS	65
2.4 PROBABILIDAD CONDICIONAL O A POSTERIORI	65
2.5 REGLAS DE LA PROBABILIDAD	66

2.5.1 Regla de la adición.	66
2.5.2 Regla de la multiplicación.	66
2.6 TEOREMA DE BAYES	67
2.7 ARBOLES DE DECISION	68
2.8 EJERCICIOS RESUELTOS	69
2.9 EJERCICIOS PROPUESTOS	74
3. DISTRIBUCIONES DE PROBABILIDAD	80
3.1 DISTRIBUCIONES DE VARIABLE DISCRETA	80
3.1.1 Distribución binomial	80
3.1.2 Distribución hipergeometrica	81
3.1.3 Distribución de Poisson	82
3.2 DISTRIBUCIONES DE VARIABLE CONTINUA	83
3.2.1 Distribución normal:	83
3.2.2 Distribución t student	86
3.2.3 Distribución χ^2 – cuadrado (χ^2).	87
3.2.4 Distribución F.	89
3.3 EJERCICIOS RESUELTOS	90
3.4 EJERCICIOS PROPUESTOS	95
4. ESTADISTICA INFERENCIAL	97
4.1 DISTRIBUCION MUESTRAL	97
4.1.1 Distribución muestral de medias.	97
4.1.2 Distribución muestral de proporciones	101
4.1.3 Métodos de muestreo.	102
4.1.4 Tamaño de la muestra para medias usando muestreo aleatorio simple.	103
4.1.5 Tamaño de la muestra para proporciones usando muestreo aleatorio simple.	105
4.2 ESTIMACION POR INTERVALOS DE CONFIANZA	106
4.2.1 El fundamento de un intervalo de confianza.	107
4.2.2 Intervalo de confianza para la media poblacional - muestras grandes	107
4.2.3 Intervalo de confianza para la media poblacional - muestras pequeñas	108

4.2.4 Intervalo de confianza para la proporción poblacional.	109
4.2.5 Intervalo de confianza para la diferencia entre dos medias – muestras grandes.	110
4.2.6 Intervalo de confianza para la diferencia entre dos medias – muestras pequeñas.	111
4.2.7 Intervalo de confianza para la diferencia entre proporciones.	115
4.3 PRUEBAS DE HIPÓTESIS	115
4.3.1 El concepto de prueba de hipótesis.	116
4.3.2 Tipos de errores.	117
4.3.3 Pasos para pruebas de hipótesis.	118
4.3.4 Prueba de hipótesis para la media poblacional	119
4.3.5 Prueba de hipótesis para la proporción poblacional.	121
4.3.6 Prueba de hipótesis para la diferencia entre dos medias.	123
4.3.7 Prueba de hipótesis para la diferencia entre dos proporciones.	126
4.3.8 Prueba de Bondad de Ajuste.	129
4.3.9 Prueba de independencia.	131
4.3.10 Comparación de la varianza de dos poblaciones normales.	131
4.3.11 Análisis de varianza – ANOVA.	133
4.4 EJERCICIOS RESUELTOS	139
4.5 EJERCICIOS PROPUESTOS	171
5. DISEÑO DE EXPERIMENTOS	179
5.1 METODOLOGIA DEL DISEÑO DE EXPERIMENTOS	179
5.2.1 La suma de cuadrados.	183
5.2.2 Cuadrados medios:	185
5.3 PRUEBAS PARA LA DIFERENCIA ENTRE PARES DE MEDIAS	187
5.3.1 Pruebas para diseños balanceados.	187
5.3.2 Pruebas para diseños no balanceados.	188
5.4 ANÁLISIS DE VARIANZA A DOS VÍAS	189
5.5 DISEÑOS FACTORIALES	193
5.5.1 Interacción entre dos factores.	193

5.5.2 Diseños factoriales para dos factores.	194
5.5.3 Análisis estadístico para modelo de efectos fijos.	195
5.5.4 Modelo de efectos aleatorios.	199
5.5.5 Modelos mixtos.	201
5.5.6 El experimento factorial generalizado	203
5.6 DISEÑOS FACTORIALES 2^K	206
5.6.1 Diseño 2^2 .	206
5.6.2 Diseño factorial 2^{3k} .	210
5.6.3 Diseño General 2^K .	214
5.6.4 Diseño factorial 2^K con una sola replicación.	216
5.7 EJERCICIOS RESUELTOS	217
5.8 EJERCICIOS PROPUESTOS	246
6. SERIES CRONOLOGICAS	252
6.1 CONCEPTOS GENERALES	252
6.2 COMPONENTES DE UNA SERIE CRONOLOGICA	253
6.2.1 Tendencia secular.	253
6.2.2 Variaciones estacionales.	254
6.2.3 Variaciones cíclicas.	254
6.2.4 Variaciones irregulares.	254
6.3 MODELOS DE SERIES CRONOLOGICAS	255
6.4 TECNICAS DE SUAVIZAMIENTO	256
6.4.1 Promedios móviles.	256
6.4.2 Suavizamiento exponencial.	257
6.5 ANALISIS DE TENDENCIA	258
6.5.1 Fundamentos de regresión.	259
6.5.2 Coeficiente de correlación.	259
6.5.3 Tendencia lineal.	261
6.5.4 Tendencia Cuadrática.	262
6.5.5 Tendencia exponencial.	263
6.6 DESCOMPOSICION DE LAS SERIES CRONOLOGICAS	264

6.6.1 Aislamiento de la componente estacional.	264
6.6.2 Aislamiento de la variación cíclica.	266
6.6.3 Variación irregular.	267
6.7 EJERCICIOS RESUELTOS	267
6.8 EJERCICIOS PROPUESTOS	284
7. CONCLUSIONES	286
8. RECOMENDACIONES	287
BIBLIOGRAFIA	288
ANEXOS	289

LISTA DE FIGURAS

	Pág.
Figura 1. Distribuciones de parámetros de forma	38
Figura 2. Curva de Lorenz (Ejercicio 3)	55
Figura 3. Árbol de decisión	68
Figura 4. Árbol de decisión (Ejercicio 5)	73
Figura 5. Distribución normal	84
Figura 6. Distribución t	87
Figura 7. Distribución ji – cuadrado	88
Figura 8. Distribución F	90
Figura 9. Distribución normal (Ejercicio 4 Numeral a)	94
Figura 10. Distribución normal (Ejercicio 4 Numeral c)	95
Figura 11. Prueba de hipótesis de dos colas para la media poblacional	119
Figura 12. Prueba de hipótesis para la proporción poblacional	122
Figura 13. Prueba de hipótesis con muestras grandes	124
Figura 14. Prueba de hipótesis con muestras pequeñas	126
Figura 15. Prueba de hipótesis para la diferencia entre dos proporciones	128
Figura 16. Distribución F	133
Figura 17. Prueba dos colas para la media poblacional (Ejercicio 11)	152
Figura 18. Prueba hipótesis para la media poblacional (Ejercicio 12)	154
Figura 19. Prueba de hipótesis para la proporción poblacional (Ejercicio13)	155
Figura 20. Prueba para la diferencia entre dos medias (Ejercicio 14)	157
Figura 21. Prueba para la diferencia entre dos medias (Ejercicio15)	159
Figura 22. Prueba para la diferencia entre dos proporciones (Ejercicio 16)	162
Figura 23. Distribución ji – cuadrado (Ejercicio 17)	164
Figura 24. Distribución ji – cuadrado (Ejercicio 18)	166
Figura 25. Distribución F (Ejercicio 19)	168

Figura 26. Distribución F (Ejercicio 1)	220
Figura 27. Diseño 2^2 (Ejercicio 7)	241
Figura 28. Componentes De Una Serie Cronológica	253
Figura 29. Ciclos de producción futura (Ejercicio 1)	272
Figura 30. Tendencia de Producción (Ejercicio 2)	281
Figura 31. Índice cíclico – irregular (Ejercicio 2)	283

LISTA DE TABLAS

	Pág.
Tabla 1. Frecuencias(Parámetros de Concentración)	43
Tabla 2. Reservas probadas para 11 campos (Ejercicio 1)	46
Tabla 3.Tabla de frecuencia (Ejercicio 2)	50
Tabla 4. Permeabilidades y frecuencias del núcleo (Ejercicio3)	53
Tabla 5.Tabla de frecuencia (Ejercicio3)	54
Tabla 6. Datos de porosidad y permeabilidad (Ejercicio 2)	56
Tabla 7. Datos de empleados (Ejercicio 3)	57
Tabla 8. Distribución capacidad de almacenamiento (Ejercicio 4)	57
Tabla 9. Ingresos de los 20 campos (Ejercicio 5)	58
Tabla 10. Tabla de frecuencia (Ejercicio 7)	59
Tabla 11. Años de experiencia de ingenieros (Ejercicio 9)	60
Tabla 12. Datos PVT (Ejercicio 11)	61
Tabla 13. Datos de porosidad (Ejercicio 12)	61
Tabla 14. Características de las bombas (Ejercicio 1)	69
Tabla 15. Probabilidades de Ecopetrol (ejercicio 5)	73
Tabla 16. Reservas de crudo (Ejercicio 1)	74
Tabla 17. Datos de volumen y probabilidad (Ejercicio 2)	75
Tabla 18. Flujos de caja (Ejercicio 5)	77
Tabla 19. Ventas de la compañía (Ejercicio 1)	140
Tabla 20. Frecuencias de las empresas (Ejercicio 17)	163
Tabla 21. Resultados de la prueba (Ejercicio18)	164
Tabla 22.Resultados calculados (Ejercicio18)	165
Tabla 23. Promedio de rentabilidad de los campos (Ejercicio 20)	169
Tabla 24. Numero de licitaciones de cada empresa (ejercicio 17)	176
Tabla 25. Valor del bono de cada área (Ejercicio 18)	177

Tabla 26. Cantidad de petróleo separado (Ejercicio 20)	178
Tabla 27. Niveles de interacción entre dos factores	193
Tabla 28. Signos algebraicos para calcular los efectos en un diseño 2^3	213
Tabla 29. Reservas de las compañías en cada bloque (ejercicio 1)	218
Tabla 30. Costos de los tipos de contrato (Ejercicio 2)	221
Tabla 31. Resumen de los estadísticos (Ejercicio 2)	222
Tabla 32. Rendimiento de los sistemas (Ejercicio 3)	224
Tabla 33. Resumen de los estadísticos (Ejercicio 3)	226
Tabla 34. Perdidas de lodo en el pozo (Ejercicio 4)	228
Tabla 35. Resumen de los estadísticos (Ejercicio 4)	230
Tabla 36. Resumen de las sumas de cuadrados (Ejercicio 5)	232
Tabla 37. Producción de los pozos (Ejercicio 6)	234
Tabla 38. Resumen de los estadísticos para ab (Ejercicio 6)	235
Tabla 39. Resumen de los estadísticos para ac (Ejercicio 6)	236
Tabla 40. Resumen de los estadísticos para bc (Ejercicio 6)	237
Tabla 41. Resumen de los estadísticos para abc (Ejercicio 6)	237
Tabla 42. Resumen de los estadísticos (Ejercicio 6)	239
Tabla 43. Desviación estándar del volumen específico (ejercicio 7)	240
Tabla 44. Resumen de los estadísticos (Ejercicio 7)	242
Tabla 45. Producción del campo (Ejercicio 8)	243
Tabla 46. Resumen de los estadísticos (Ejercicio 8)	245
Tabla 47. Datos de concentración (Ejercicio 1)	246
Tabla 48. Producción de los pozos (Ejercicio 2)	246
Tabla 49. Resistencias a la tensión (Ejercicio 3)	247
Tabla 50. Resistencia al esfuerzo cortante (Ejercicio 4)	248
Tabla 51. Grado de absorción CO_2 (Ejercicio 6)	249
Tabla 52. Datos obtenidos del experimento (Ejercicio 7)	250
Tabla 53. Producción en porcentaje del pozo (Ejercicio 8)	251
Tabla 54. Datos de producción (Ejercicio 1)	268
Tabla 55. Desarrollo estadístico para datos de producción anual	269

Tabla 56. Obtención de las relativas cíclicas – irregulares	271
Tabla 57. Descomposición de la serie (Ejercicio 2)	274
Tabla 58. Cálculo del índice estacional S_i (Ejercicio 2)	280
Tabla 59. Datos de producción (Ejercicio 1)	285
Tabla 60. Datos de producción (Ejercicio 2)	285

LISTA DE ANEXOS

	Pág.
ANEXO A. Glosario estadístico	289
ANEXO B. Glosario tecnico	294
ANEXO C. Respuestas a algunos ejercicios	297
ANEXO D. Distribución Binomial	299
ANEXO E. Distribución Normal	310

RESUMEN

TITULO: LA ESTADÍSTICA Y SUS APLICACIONES EN LA INGENIERÍA DE PETRÓLEOS*

AUTOR: JORGE ENRIQUE PARRA CAMARON**

PALABRAS CLAVES: Intervalos de confianza, probabilidades, producción, presiones, pruebas de hipótesis, series, permeabilidad, muestras.

DESCRIPCIÓN

La estadística cuenta con herramientas diversas y versátiles que permiten medir la confiabilidad en los resultados de los estudios y procesos, permitiendo la estimación del comportamiento de las variables petrofísicas y de producción que juegan un papel importante en la cuantificación de los factores como caudales futuros, reservas, vida útil y rentabilidad. Estas aplicaciones son las que permiten visualizar la importancia del uso de la estadística como herramienta para la correcta recolección de información, el correcto manejo de estos datos, y su uso para la estimación de las variables aleatorias, permitiendo con ellas hacer inferencias sobre los comportamientos futuros, midiendo sus errores y peso probabilístico que permita respaldar las decisiones que se tomen.

El proyecto se fundamenta en la elaboración de un manual guía, donde cada capítulo se inicia con definiciones, conceptos y principios, pertinentes y claros, con material gráfico, y de varios ejercicios resueltos y propuestos

El Capítulo 1 da la introducción a la estadística descriptiva, utilizando datos agrupados y no agrupados. Más adelante, el Capítulo 2 habla de la teoría de probabilidad. Después, el Capítulo 3 habla de las distribuciones discretas, y de las distribuciones continuas. El Capítulo 4 describe todo lo relacionado con la estadística inferencial. El capítulo 5 trata los diseños fundamentales de un experimento. Finalmente, el capítulo 6 trata las series cronológicas o de tiempo, tomando como base el método exponencial de producción.

* Trabajo de Grado

** Facultad de ingenierías Físicoquímicas. Escuela de ingeniería de petróleos. Director: Oscar Vanegas Angarita

ABSTRACT

TITLE: STATISTICS AND ITS APPLICATIONS TO THE PETROLEUM ENGINEERING*

AUTHOR: JORGE ENRIQUE PARRA CAMARÓN**

KEY WORDS: Confidence intervals, probabilities, production, pressures, hypothesis tests, series, permeability, samples.

DESCRIPTION

Statistics counts with diverse and versatile tools, those allow to measure the reliability in the results of the processes and studies, allowing to estimate the behavior of the petro physical and production variables that play an important role in the quantification of factors like future flows, reserves, lifetime and rentability. These applications let us to visualize the importance of the use of statistics as a tool for the right gathering of information, the correct treatment of these data and its use for the estimation of the random variables, allowing to make inferences over the future performance, measuring its errors and probabilistic weight that help to support the decisions to be made.

The project is based on the elaboration of a guide manual, each chapter stars with definitions concepts and principles, pertinent and clear, with graphic material and various exercises proposed and solved.

The chapter 1 gives the introduction to the descriptive statistic, using grouped and separated data. Later, the chapter 2 is about the theory of the probability. Afterwards, the chapter 3 deals with the discrete distributions and the continuous distributions. The chapter 4 describes all the related with inferential statistics. Chapter 5 treats the fundamental designs of an experiment. Finally, the chapter 6 treats whit the chronological or time series, taking as a base the exponential production method.

*Undergraduate Project

**Faculty of Physical – Chemical Engineering. Petroleum Engineering. Director: Oscar VanegasAngarita

INTRODUCCIÓN

La estadística y los métodos probabilísticos o estocásticos juegan un papel muy importante en todas las áreas de la industria de los hidrocarburos. El uso de la probabilidad y de la estadística se ha extendido, no tan solo a las áreas tradicionales universitarias o escolásticas, sino también a todos los campos de la ingeniería, en este caso particular, la ingeniería de petróleo. En donde la estadística, como rama de las matemáticas, es una ciencia ingenieril de gran aplicación, no tan solo en la toma de decisiones, si no en los procesos investigativos, para el correcto manejo y tratamiento de la información, sobre todo con los datos cuantitativos, dándole confiabilidad a las investigaciones al enfrentar correctamente las hipótesis que se plantean.

Este texto guía de probabilidad y de estadística, está diseñado para cursos de pregrado de la Ingeniería de petróleo y para la aplicación en la industria de los hidrocarburos como tal. En donde, para un buen entendimiento del texto, se requiere que el lector tenga conocimientos previos de estadística, El texto es una compilación teórica de más de 6 libros de referencias bibliográficas de probabilidad y de estadística, donde sus ejemplos están orientados tan solo a la ingeniería de petróleo y la industria de los hidrocarburos, por otra parte el texto incluye un glosario con las palabras más relevantes de cada sección.

El texto consta de más de 250 páginas que incluyen conceptos teóricos, varios ejercicios prácticos, resueltos y propuestos. El propósito del texto es el de ayudar al lector a entender los conceptos, ideas y funciones de la probabilidad y de la estadística aplicados a problemas de la ingeniería de petróleo. El texto deberá ser también útil para aquellos estudiosos quienes deseen hacer aplicaciones de la probabilidad y de estadística a problemas de la ingeniería en términos generales, así como también a la investigación.

1. ESTADISTICA DESCRIPTIVA

La estadística descriptiva se encarga de los procedimientos matemáticos que tienen por objeto calcular los parámetros estadísticos que describen el comportamiento del conjunto de datos a analizar. Estos parámetros pueden ser de tendencia central, de dispersión, de posición, de forma y de concentración.

Se le aclara al lector que la nomenclatura utilizada en las fórmulas de esta sección, únicamente hace referencia a datos procedentes de muestras. Para el cálculo de los parámetros estadísticos poblacionales se utilizan las mismas fórmulas, teniendo en cuenta el uso de la correspondiente nomenclatura; por ejemplo, para calcular la media aritmética de una muestra se utiliza x como nomenclatura, pero para la media aritmética poblacional se debe utilizar la letra μ , teniendo el cuidado de reemplazar dentro de la fórmula el tamaño de la muestra n , por el tamaño de la población N . De acuerdo a lo anterior, la estadística descriptiva es la segunda etapa a desarrollar en un análisis de información, después del muestreo.

1.1 PARÁMETROS DE TENDENCIA CENTRAL

Los parámetros de tendencia central son valores numéricos que tienden a localizar la parte central de un conjunto de datos. Es un valor que se puede tomar como representativo de todos los datos. Entre los parámetros de tendencia central se encuentran la media aritmética, la media ponderada, la media geométrica, la mediana y la moda.

1.1.1 Parámetros de tendencia central a partir de datos sueltos o no agrupados (muestras de tamaño menor a 30 datos)

1.1.1.1 Media aritmética: También denominada promedio¹, es la que se utiliza principalmente y se define como la suma de los valores de todas las observaciones x_i divididas por el número total de datos n .

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

Propiedades de la media aritmética

- La sumatoria de las desviaciones de cada término respecto de la media es igual a cero.

$$\sum (x_i - \bar{x}) = 0$$

- La suma de los cuadrados de las desviaciones de los valores de la variable con respecto a un número cualquiera se hace mínima cuando dicho número coincide con la media aritmética.

$$\sum (x_i - \bar{x})^2 = \text{mínimo}$$

- Si todos los valores de la variable se les suma un mismo número, la media aritmética queda aumentada en dicho número.

¹Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

- Si todos los valores de la variable se multiplican por un mismo número, la media aritmética queda multiplicada por dicho número.

1.1.1.2 Media ponderada: En ocasiones no todos los valores de la variable tienen el mismo peso. Esta importancia que asignamos a cada variable, es independiente de la frecuencia absoluta que tenga. Será como un aumento del valor de esa variable, en tantas veces como consideremos su peso.

Por lo tanto la media aritmética ponderada se utiliza cuando a cada valor de la variable x_i se le otorga una ponderación o peso distinto de la frecuencia o repetición. Para poder calcularla se tendrá que tener en cuenta los pesos w_i de cada uno de los valores que tenga la variable.

$$x_w = \frac{x_1w_1 + x_2w_2 + x_3w_3 + \dots + x_nw_n}{w_1 + w_2 + w_3 + \dots + w_n} = \frac{x_iw_i}{w_i} \quad (1 - 2)$$

En donde x_w Es la media ponderada

x_i Es la observación individual

w_i Es el peso asignado a cada observación

1.1.1.3 Media geométrica: La media geométrica proporciona una medida promedio precisa de una variable que mide los cambios porcentuales en una serie de valores positivos. Se halla tomando la raíz enésima del producto de n valores porcentuales.

$$MG = \sqrt[n]{x_1x_2x_3 \dots x_n} \quad (1 - 3)$$

1.1.1.4 Media armónica: Es la inversa de la media aritmética de los inversos de los n valores de la variable, se representa por la letra H y responde a la siguiente expresión:

$$H = \frac{n}{\frac{\sum x_i}{n}} \quad (1 - 4)$$

Esta media no es aconsejable en distribuciones de variables con valores pequeños y unidades simples. Se utiliza para promediar variables con unidades complejas o compuestas del área de la física y la química, tales como velocidades, aceleraciones, trabajo, energía, etc.

1.1.1.5 Mediana: Se define como el valor que deja igual número de observaciones a su izquierda que a su derecha, es decir divide el conjunto de datos en dos partes iguales y se denota por Me . Si el conjunto de datos tiene un número impar de observaciones, una vez que los datos han sido **ordenados** en orden creciente o decreciente², la posición de la mediana es:

$$\text{posición de la mediana} = \frac{n + 1}{2} \quad (1 - 5)$$

La Me será el valor del dato observado X_i que esté ubicado en la posición calculada mediante la ecuación (1 - 5). Si el conjunto de datos observados es par, la Me será la media aritmética de los datos observados en las posiciones centrales:

²John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \quad (1 - 6)$$

1.1.1.6 Moda: Es el valor de la variable que más veces se repite, es decir, aquella cuya frecuencia absoluta es mayor. Puede haber más de una moda en una distribución. Se denota por *Mo*.

1.1.2 Parámetros de tendencia central a partir de datos agrupados (muestras mayores o iguales a 30 datos). Para calcular los parámetros estadísticos (tanto los de tendencia central, como los de dispersión, posicionamiento, forma y concentración) cuando el tamaño de la muestra es mayor o igual a 30 datos, éstos previamente se deben ordenar de menor a mayor o de mayor a menor, y luego proceder a agruparlos por clases o intervalos en una tabla de frecuencias, siguiendo los siguientes pasos o indicaciones:

Agrupamiento:

- Calcular el número de clases o intervalos = (\bar{n}) , se recomienda que el número de clases este entre 5 y 15.
- Rango = $X_n - X_1$
- Amplitud de clase (A)= rango / N_o de clase

Tabla de frecuencias:

- Punto medio o marca de clase: El centro de la clase, es el valor de los datos que se ubican en la posición central de la clase y representan todos los demás valores de esa clase. Este valor se utiliza para el cálculo de la media aritmética.

- Frecuencia absoluta (f_i): es el número total de valores (datos), que se encuentran presentes en una clase determinada.
- Frecuencia relativa (h_i): es aquella que resulta de dividir cada una de las frecuencias absolutas entre el número total de datos (n). Se representa como ($h_i = f_i / n$).
- Frecuencia absoluta acumulada (F): es aquella que se obtiene de la suma sucesiva de las f_i que integran cada una de las clases de una distribución de frecuencia de clase, esto se logra cuando la acumulación de las frecuencias se realiza tomando en cuenta la primera clase hasta alcanzar la última.
- Frecuencia relativa acumulada (H)³: es aquella que se obtiene de la suma sucesiva de las h_i que integran cada una de las clases de una distribución de frecuencia de clase, esto se logra cuando la acumulación de las frecuencias se realiza tomando en cuenta la primera clase hasta alcanzar la última.

1.1.2.1 Media aritmética: Para calcular la media de datos agrupados se puede usar la ecuación 1-7, donde x_c es la marca de clase en cada intervalo, f_i es la frecuencia absoluta de cada clase y n es la suma de frecuencias absolutas.

$$x = \frac{f_i x_c}{n} \quad (1 - 7)$$

O puede usarse la ecuación 1-8, donde x_c es la marca de clase en cada intervalo y h_i es la frecuencia relativa de cada clase.

³Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

$$x = \sum h_i x_c \quad (1 - 8)$$

1.1.2.2 Media geométrica: Para datos agrupados la media geométrica es la raíz n -ésima del producto de los valores de las marcas de clase, elevado cada uno a su respectiva frecuencia absoluta, donde k es el número de clases y n es la suma de frecuencias absolutas. Se calcula mediante la siguiente ecuación.

$$MG = \sqrt[n]{x_{c1}^{f_1} x_{c2}^{f_2} x_{c3}^{f_3} \dots x_{ck}^{f_n}} \quad (1 - 9)$$

1.1.2.3 Mediana: En el caso de datos agrupados, se debe encontrar primero la clase mediana. La clase mediana es la clase cuya frecuencia acumulada es mayor o igual a $n/2$. Se ubica el intervalo en el que está la mediana, y su valor se obtiene a través de interpolación mediante la siguiente expresión:

$$\text{Mediana}(Me) = L_{i\ md} + A \frac{\frac{n}{2} - F}{f_{md}} \quad (1 - 10)$$

En donde $L_{i\ md}$ Es el límite inferior de la clase de la mediana

f_{md} Es la frecuencia absoluta de la clase de la mediana

A Es la amplitud del intervalo de la clase mediana

n Es el tamaño de la muestra

F Es la frecuencia absoluta acumulada de la clase que antecede a la clase mediana.

1.1.2.4 Moda: Es la observación que ocurre con mayor frecuencia. Primero se hallará la clase que tenga la frecuencia más alta, llamada clase modal. Para estimar la moda en este caso se utiliza la siguiente ecuación.

$$\text{Moda } (Mo) = L_{i\ mo} + A \frac{\Delta_a}{\Delta_a + \Delta_b} \quad (1 - 11)$$

En donde $L_{i\ mo}$ Es el límite inferior de la clase modal

Δ_a Es la diferencia entre la frecuencia absoluta de la clase modal y la clase que la antecede.

Δ_b Es la diferencia entre la frecuencia absoluta de la clase modal y la clase que le sigue.

A Es la amplitud del intervalo de clase de la clase modal

1.2 PARAMETROS DE DISPERSION

Los parámetros de dispersión⁴ indican la mayor o menor concentración de los datos con respecto a las medidas de centralización. Miden que tanto se desvían las observaciones alrededor de su media.

1.2.1 Parámetros de dispersión a partir de datos sueltos o no agrupados

1.2.1.1 Rango: La medida de dispersión más simple (y menos útil) es el rango o recorrido. El rango es simplemente la diferencia entre la observación más alta y la más baja. Su ventaja es que es fácil de calcular y su desventaja es que se considera dos de los cientos de observaciones que se encuentran en un conjunto de datos. Se calcula de la misma forma para datos agrupados o no agrupados.

⁴ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

1.2.1.2 Varianza: La varianza es el promedio de las desviaciones cuadráticas de cada dato (x_i) respecto a su media aritmética (x).

$$s^2 = \frac{(x_1 - x)^2 + (x_2 - x)^2 + (x_3 - x)^2 + \dots + (x_n - x)^2}{n - 1}$$

$$s^2 = \frac{(x_i - x)^2}{n - 1} = \frac{x_i^2 - \frac{(x_i)^2}{n}}{n - 1} \quad (1 - 12)$$

En donde $x_1, x_2, x_3, \dots, x_n$ Son las observaciones individuales

x Es la media de la muestra

n Es el tamaño de la muestra

Propiedades de la varianza

- La varianza será siempre un valor positivo o cero, en el caso en que las puntuaciones sean iguales.
- Si a todos los valores de la variable se les suma un número, varianza no varía.
- Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.
- Si se tienen k distribuciones con la misma media y se conocen sus respectivas varianzas, se puede calcular la varianza total.

Si todas las k muestras tiene el mismo tamaño

$$\sigma^2 = \frac{s_1^2 + s_2^2 + \dots + s_n^2}{k} \quad (1 - 13)$$

Si las k muestras tienen distinto tamaño

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2 + \dots + n_k s_n^2}{n_1 + n_2 + \dots + n_k} \quad (1 - 14)$$

1.2.1.3 Desviación estándar: También llamada desviación típica, es una medida de dispersión usada en estadística que nos dice cuánto tienden a alejarse los valores puntuales del promedio en una distribución. Específicamente, la desviación estándar es "el promedio de la distancia de cada punto respecto del promedio". Es la raíz cuadrada de la varianza. Para calcular la desviación estándar para una muestra se utiliza la siguiente expresión.

$$s = \sqrt{s^2} \quad (1 - 15)$$

1.2.1.4 Desviación media: Es la media aritmética de todas las diferencias absolutas entre cada observación individual y la media aritmética del conjunto de datos. La desviación media de la muestra se calcula utilizando la siguiente ecuación.

$$Dm = \frac{|x_i - \bar{x}|}{n} \quad (1 - 16)$$

En donde \bar{x} Es la media de la muestra

n Es el número total de observaciones o tamaño de la muestra

1.2.1.5 Coeficiente de variación: Un uso importante de la desviación estándar es servir como medida de dispersión⁵. Sin embargo, se aplican ciertas limitaciones. Cuando se consideran dos o más distribuciones que tienen medias significativamente diferentes, o que están medidas en unidades distintas, es peligroso sacar conclusiones respecto a la dispersión solo con base o teniendo en cuenta la desviación estándar.

Por tanto, con frecuencia debemos considerar el coeficiente de variación (CV), el cual sirve como medida relativa de dispersión. El coeficiente de variación determina el grado de dispersión de un conjunto de datos relativo a su media. Se calcula dividiendo la desviación estándar de una distribución por su media y multiplicando por 100.

Coeficiente de variación para la muestra

$$CV = \frac{S}{x} * 100 \text{ (1 - 17)}$$

1.2.2 Parámetros de dispersión a partir de datos agrupados

1.2.2.1 Varianza: Si los datos están agrupados en una tabla de frecuencia, la varianza de la muestra puede calcularse mediante la siguiente ecuación.

$$s^2 = \frac{f_i (x_c - x)^2}{n - 1} \text{ (1 - 18)}$$

⁵John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

En donde \bar{x} Es la media muestral

n Total de observaciones

x_c Punto medio de la clase

f_i Frecuencia absoluta de la clase

1.2.2.2 Desviación estándar: La desviación estándar de una muestra es la raíz cuadrada de la varianza y se puede calcular con la siguiente expresión.

$$s = \sqrt{s^2} \quad (1 - 19)$$

1.2.2.3 Desviación media: Si los datos vienen agrupados en una tabla de frecuencias, la ecuación para la desviación media de una muestra es la siguiente.

$$Dm = \frac{f_i |x_c - \bar{x}|}{n} \quad (1 - 20)$$

En donde \bar{x} Es la media muestral

n Total de observaciones

x_c Punto medio de la clase

f_i Frecuencia absoluta de la clase

1.2.2.4 Coeficiente de variación: Es una medida relativa de variabilidad de los datos. Permite comparar la variabilidad de dos o más conjuntos de datos expresados en unidades diferentes. El Coeficiente de variación para la muestra se calcula mediante la siguiente expresión:

$$CV = \frac{S}{x} * 100(1 - 21)$$

1.3 PARAMETROS DE POSICION

Los parámetros de posición dividen un conjunto de datos en grupos con el mismo número de individuos, para calcular los parámetros de posición⁶ es necesario que los datos estén ordenados de mayor a menor.

1.3.1 Parámetros de posición a partir de datos sueltos o no agrupados

1.3.1.1 Percentiles: Son 99 valores que dividen en cien porciones iguales el conjunto de datos ordenados. Los **percentiles** dan los valores correspondientes al 1%, al 2%.y al 99% de los datos. La posición de la variable de los Percentiles se puede calcular con la siguiente ecuación.

$$\text{Posición de la variable del } P_k = \frac{Kn}{100} (1 - 22)$$

En donde $P_k = P_1, P_2, P_3, \dots, P_{99}$

$$K = 1, 2, 3, \dots, 99$$

n Total de observaciones

La anterior ecuación se utiliza cuando el total de observaciones (n) es par, cuando el total de observaciones es impar, la ecuación es:

$$\text{Posición de la variable del } P_k = \frac{K(n + 1)}{100} (1 - 23)$$

⁶ John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

1.3.1.2 Deciles: Son los nueve valores que dividen al conjunto de datos ordenados en diez porciones iguales, son también un caso particular de los percentiles, ya que corresponden a los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90. La posición⁷ de la variable de los Deciles se puede calcular con la siguiente ecuación.

$$\text{Posición de la variable del } D_k = \frac{Kn}{10} \quad (1 - 24)$$

En donde $D_k = D_1, D_2, D_3, \dots, D_9$

$$K = 1, 2, 3, \dots, 9$$

n Total de observaciones

La anterior ecuación se utiliza cuando el total de observaciones (n) es par, cuando el total de observaciones es impar, la ecuación es:

$$\text{Posición de la variable del } D_k = \frac{K(n + 1)}{10} \quad (1 - 25)$$

1.3.1.3 Cuartiles: Son los tres valores que dividen al conjunto de datos ordenados en cuatro porciones iguales, son un caso particular de los percentiles, correspondiendo a los percentiles 25, 50 y 75.

- El primer cuartil Q1 es el valor de la variable que deja a la izquierda el 25% de la distribución.
- El segundo cuartil Q2 (la mediana), es el valor de la variable que deja a la izquierda el 50% de la distribución.

⁷ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

- El tercer cuartil Q3 es el valor de la variable que deja a la izquierda el 75% de la distribución.

La posición de la variable que representa cada uno de los cuartiles, se calcula con la siguiente ecuación.

$$\text{Posición de la variable del } Q_k = \frac{Kn}{4} \quad (1 - 26)$$

En donde $Q_k = Q_1, Q_2, Q_3$

$$K = 1, 2, 3$$

n Total de observaciones

La anterior ecuación se utiliza cuando el total de observaciones (n) es par, cuando el total de observaciones es impar⁸, la ecuación es:

$$\text{Posición de la variable del } Q_k = \frac{K(n + 1)}{4} \quad (1 - 27)$$

1.3.2 Parámetros de posición a partir de datos agrupados

1.3.2.1 Percentiles: Para datos agrupados en una tabla de frecuencias, los percentiles se calculan con la siguiente ecuación:

$$P_k = L_i + A \frac{\frac{Kn}{100} - F_{i-1}}{f_i} \quad (1 - 28)$$

En donde L_i Es el límite inferior de la clase del percentil k

⁸Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

n Es la suma de las frecuencias absolutas

A Es la amplitud de la clase del percentil k

f_i Frecuencia de la clase del percentil k

F_{i-1} Es la frecuencia acumulada de la clase que antecede a la clase del percentil k . ($K = 1,2,3, \dots, 9$)

1.3.2.2 Deciles: Para datos agrupados en una tabla de frecuencias, los deciles se calculan con la siguiente ecuación.

$$D_k = L_i + A \frac{\frac{Kn}{10} - F_{i-1}}{f_i} \quad (1 - 29)$$

En donde L_i Es el límite inferior de la clase del decil k

n Es la suma de las frecuencias absolutas

A Es la amplitud de la clase del decil k

f_i Frecuencia de la clase del decil k

F_{i-1} Es la frecuencia acumulada de la clase que antecede a la clase del decil k . ($K = 1,2,3, \dots, 9$)

1.3.2.3 Cuartiles: Cuando los datos están agrupados en una tabla de frecuencias. Los cuartiles⁹ se calculan con la siguiente ecuación.

⁹ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

$$Q_k = L_i + A \frac{\frac{Kn}{4} - F_{i-1}}{f_i} \quad (1 - 30)$$

En donde L_i Es el límite inferior de la clase del cuartil k

n Es la suma de las frecuencias absolutas

A Es la amplitud de la clase del cuartil k

f_i Frecuencia de la clase del cuartil k

F_{i-1} Es la frecuencia acumulada de la clase que antecede a la clase del cuartil k. ($K = 1, 2, 3$)

1.4 PARAMETROS DE FORMA Y CONCENTRACION

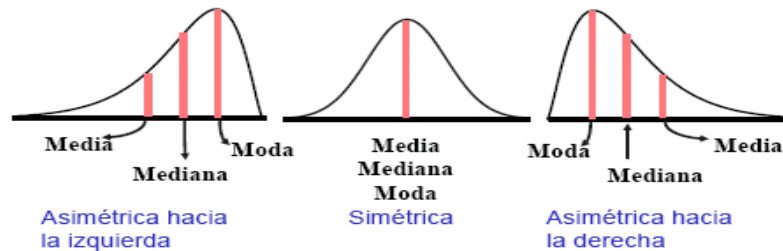
1.4.1 Parámetros de forma: Las medidas de forma de una distribución se pueden clasificar en dos grandes grupos o bloques: *medidas de asimetría* y *medidas de curtosis*. Una distribución de frecuencias puede ser simétrica o asimétrica. Para saber si es simétrica se tiene que tomar una referencia, es decir, ver respecto a qué es simétrica.

Cuando al trazar una vertical, en el diagrama de barras o histograma, de una variable, según sea esta discreta o continua, por el valor de la media, esta vertical, se transforma en eje de simetría, decimos que la distribución es simétrica.

Diremos pues, que es simétrica, cuando a ambos lados de la media aritmética haya el mismo número de valores de la variable, equidistantes de dicha media dos a dos, y tales que cada par de valores equidistantes tiene la misma frecuencia absoluta.

En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría. El coeficiente de asimetría de Pearson y el coeficiente de asimetría de Fisher¹⁰, miden la desviación de la simetría.

Figura 1. Distribuciones de parámetros de forma



a.

b.

c.

Fuente: www.vitutor.com/estadistica/descriptiva/a_7.html - España

- **Coeficiente de asimetría de Fisher**

Para calcular la asimetría una posibilidad es mediante el coeficiente de Fisher que se representara como (g_1) , y se representara mediante la siguiente expresión:

$$g_1 = \frac{(x_i - \bar{x})^3 f_i}{n s^3} \quad (1-31)$$

En donde \bar{x} Es la media aritmética de la muestra

s Es la desviación estándar de una muestra

n Es la frecuencia de cada valor

x_i Cada uno de los valores

¹⁰John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

- Si $g_1 = 0$... la distribución es simétrica, en ese caso las desviaciones a la derecha y a la izquierda de la media se compensan. Figura 1.b
- Si $g_1 < 0$... la distribución es asimétrica negativa o desplazada o sesgada hacia la izquierda. Figura 1.a
- Si $g_1 > 0$... la distribución es asimétrica positiva o desplazada o sesgada hacia la derecha. Figura 1.c

- **Coefficiente de asimetría de Pearson**

Otra forma de calcular la asimetría, es por medio del coeficiente de Pearson (A_p), el cual responde a la siguiente expresión:

$$A_p = \frac{(x - M_o)}{s} \quad (1 - 32)$$

En donde \bar{x} Es la media aritmética

s Es la desviación estándar de una muestra

M_o Es la Moda

En la práctica este coeficiente es más fácil de calcular que el anterior, pero casi no se utiliza ya que sólo es cierto cuando la distribución tiene las siguientes condiciones:

1. Datos ordenados de menor a mayor
2. unimodal
3. Campaniforme
4. Moderada o ligeramente asimétrica

- Si $A_p = 0$... la distribución es simétrica, en ese caso las desviaciones a la derecha y a la izquierda de la media se compensan. Figura 1.b

- Si $A_p < 0$... la distribución es asimétrica negativa ya que la mayoría de las observaciones están a la derecha de la proyección de la media. Figura 1.a
- Si $A_p > 0$... la distribución es asimétrica positiva ya que la mayoría de las observaciones están a la izquierda de la proyección de la media. Figura 1.c

- **Curtosis: Coeficiente de Fisher**

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del *Coeficiente de Curtosis*, podemos identificar si existe una gran concentración de valores (*Leptocúrtica*), una concentración normal (*Mesocúrtica*) o una baja concentración (*Platicúrtica*)¹¹. Para calcularlo se utilizara la siguiente expresión:

$$g_2 = \frac{(x_i - x)^4 f_i}{n s^4} - 3 \quad (1 - 33)$$

- Si $g_2 = 0$ la distribución será mesocurtica o normal.
- Si $g_2 < 0$ la distribución será platicurtica o menos apuntada que lo normal (aproximada a la distribución t de Student).
- Si $g_2 > 0$ la distribución será leptocurtica o apuntada (aproximada a la distribución de Fisher).

1.4.2 Parámetros de concentración: Los parámetros de concentración tratan de poner de relieve el mayor o menor grado de igualdad en el reparto del total de los valores de la variable, son por tanto indicadores del grado de distribución de la variable. Para este fin están concebidos los estudios sobre concentración.

¹¹John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

En donde se denomina concentración a la mayor o menor equidad en el reparto de la suma total de los valores de la variable considerada, en este caso puede ser (permeabilidad, porosidad, saturación, etc.). Las infinitas posibilidades que pueden adoptar los valores, se encuentran entre los dos extremos:

1. Concentración máxima: cuando uno solo percibe el total y los demás nada, en este caso, nos encontramos ante un reparto no equitativo:

$$X_1 = X_2 = X_3 = \dots = X_{n-1} = 0 \quad \text{y} \quad X_n = \text{todo}$$

2. Concentración mínima: cuando el conjunto total de valores de la variable está repartido por igual, en este caso se estará ante un reparto equitativo:

$$X_1 = X_2 = X_3 = \dots = X_{n-1} = X_n$$

De las diferentes medidas de concentración que existen solo se hará referencia a:

Índice de Gini, coeficiente, por tanto será un valor numérico.

Curva de Lorenz, gráfico, por tanto será una representación en ejes coordenados.

Sea, por ejemplo, una distribución de permeabilidades (X_i vs f_i) de la que formaremos una tabla con las siguientes columnas:

1. Los productos $X_i f_i$. Donde X_i es la permeabilidad marca de clase en md y f_i es el espesor en pies.
2. Las frecuencias absolutas acumuladas F_i .
3. Los totales acumulados u_i que se calculan de la siguiente forma:

$$\begin{aligned}
u_1 &= X_1 f_1 \\
u_2 &= X_1 f_1 + X_2 f_2 \\
u_3 &= X_1 f_1 + X_2 f_2 + X_3 f_3 \\
u_4 &= X_1 f_1 + X_2 f_2 + X_3 f_3 + X_4 f_4 \\
u_n &= X_1 f_1 + X_2 f_2 + X_3 f_3 + X_4 f_4 + \dots + X_n f_n
\end{aligned}$$

Por tanto podemos decir que el total acumulado de todas las clases es:

$$u_n = \sum X_i f_i \quad (1 - 34)$$

4. La columna total de frecuencias relativas acumuladas, que expresaremos en tanto por ciento y se representara como p_i .

$$p_i = \frac{F_i}{n} \quad (1 - 35)$$

5. A cada total acumulado U_i de cada clase lo dividimos en U_n , que dada en tanto por ciento se representara como q_i y que responderá a la siguiente notación:

$$q_i = \frac{u_i}{u_n} \quad (1 - 36)$$

Por tanto ya podemos confeccionar la tabla que será la siguiente:

Tabla 1. Frecuencias (Parámetros de Concentración)

X_i	f_i	F_i	$X_i f_i$	U_i	$q_i = \frac{u_i}{u_n}$	$p_i = \frac{F_i}{n}$	$p_i - q_i$
X_1	f_1	$F_1 = f_1$	$X_1 f_1$	$U_1 = X_1 f_1$	q_1	$p_1 = (F_1/n)$	$p_1 - q_1$
X_2	f_2	$F_2 = f_1 + f_2$	$X_2 f_2$	$U_2 = U_1 + X_2 f_2$	q_2	$p_2 = (F_2/n)$	$p_2 - q_2$
...
X_n	f_n	$\sum f_i$	$X_n f_n$	$U_n = U_{n-1} + X_n f_n$	Q_n	P_n	$p_n - q_n$

Fuente: www.ucm.es/info/eiop/licenciaturas/pdfs

Como se puede observar la última columna es la diferencia entre las dos penúltimas, esta diferencia sería 0 para la concentración máxima ya que $p_i = q_i$ y por tanto su diferencia sería cero.

Si esto se representa gráficamente se obtiene la **curva de concentración o curva de Lorenz**. La manera de representarlo será, en el eje de las X los valores p_i en % y en el de las Y los valores de q_i en %. Al ser un %, el gráfico siempre será un cuadrado, y la gráfica será una curva que se unirá al cuadrado por los valores (0,0), y (100,100), y quedará por debajo de la diagonal si los datos se ordenan de menor a mayor o por encima de la diagonal si los datos se ordenan de mayor a menor.

La manera de interpretarla será: cuanto más cerca se sitúe esta curva a la diagonal, mayor concentración habrá, o más homogeneidad en el yacimiento. Cuanto más se acerque a los ejes, por la parte inferior o superior del cuadrado, menor concentración de la variable a un único valor, o más heterogeneidad del yacimiento.

El **índice de Gini se calculará** analíticamente, el cual responde a la siguiente ecuación:

$$I_G = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i} \quad (1 - 37)$$

Este índice tomará los valores de **$I_G = 0$ cuando $p_i = q_i$ (concentración máxima)** y de **$I_G = 1$ cuando $q_i = 0$.**

Otra manera de medir la concentración es a través del Coeficiente de Variación de Dykstra – Parsons. En donde **Dykstra y Parsons** introdujeron el concepto de coeficiente de variación de permeabilidad, V , el cual es una medida estática de la no uniformidad de un conjunto de datos. Esto es generalmente aplicado a la propiedad de permeabilidad pero puede ser extendido para tratar otras propiedades de la roca.

La información de permeabilidad presenta una distribución log-normal. Esto significa que los procesos geológicos que crearon la permeabilidad en una roca de yacimiento aparecieron para dejar permeabilidades distribuidas alrededor de un

promedio geométrico. Dykstra y Parsons reconocieron este comportamiento e introdujeron la variación en la permeabilidad que caracteriza una distribución en particular. A continuación se presenta un resumen de los pasos a seguir para determinar el coeficiente de variación de permeabilidad (V):

- Colocar las muestras de núcleo en orden decreciente de permeabilidad.
- Para cada muestra, calcular el porcentaje del espesor con la permeabilidad más grande de la muestra.
- Usando un papel log – probabilístico, graficar los valores de permeabilidad en la escala logarítmica y el porcentaje del espesor en la escala probabilística.
- Dibujar la mejor recta que se ajuste a los puntos graficados.
- Leer el valor de permeabilidad correspondiente a 50% y 84.1% de espesor, estos valores son designados como K_{50} y $K_{84.1}$.
- La variación de Dykstra – Parsons está definida por la siguiente expresión

$$V = \frac{K_{50} - K_{84.1}}{K_{50}} \quad (1 - 38)$$

1.5 EJERCICIOS RESUELTOS

1. Ejercicio datos no agrupados

La Agencia Nacional de Hidrocarburos (ANH) ha reportado el tamaño de las reservas probadas de petróleo descubiertas en 11 campos en el último año en Colombia, las cuales se muestran en la siguiente tabla de forma descendente en millones de barriles (MBIs), por campo.

Tabla 2. Reservas probadas para 11 campos (Ejercicio 1)

Nombre del Campo	Reservas Probadas (MBIs)	(X_i)
Vananín	7.4	
El Camarón	5.0	
El Caimán	3.5	
El Óscar	2.1	
UIS 32		1.5
Roberta		1.2
Los Cucarrones		1.0
Emiliana		1.0
Guadalupe	0.4	
Tolomeo		0.3
La Castellana	0.2	

Se quieren calcular los parámetros estadísticos suficientes para analizar el riesgo geológico de las cuencas sedimentarias colombianas y el nivel de atracción exploratorio del País para los inversionistas. ¿Calcule los parámetros de tendencia central, de dispersión, y para los parámetros de posición halle el segundo cuartil, el sexto decil y el percentil número 35 para el tamaño de las reservas probadas?

Solución

X_i = observación individual de cada campo

n = número total de datos

• Media Aritmética

$$\bar{x} = \frac{\sum x_i}{n} = \frac{23.6}{11} = 2.1455 \text{ MBIs}$$

Los yacimientos de petróleo descubiertos en Colombia en el último año tienen en promedio 2'145.500 barriles de reservas probadas.

- **Media Geométrica**

$$MG = \sqrt[n]{x_1 x_2 x_3 \dots x_n} = \sqrt[11]{7.4 * 5 * 3.5 * \dots * 0.2} = 1.251 \text{ MBls}$$

- **Media Armónica**

$$H = \frac{n}{\frac{n_i}{x_i}} = \frac{11}{\frac{1}{7.4} + \frac{1}{5} + \dots + \frac{1}{0.2}} = 0.713 \text{ MBls}$$

- **Mediana**

$$\text{posición de la mediana} = \frac{n + 1}{2} = \frac{11 + 1}{2} = 6$$

Me(Mediana) = 1.2MBls.El 50% de los yacimientos de petróleo descubiertos en Colombia en el último año tienen menos de 1'200.000 barriles de reservas probadas.

- **Moda**

Moda = 1.0 MBls (dato que se repite dos veces).

El tamaño de reservas probadas más frecuente de los yacimientos descubiertos en Colombia en el último año es de un millón de barriles.

- **Varianza**

$$s^2 = \frac{(x_i - \bar{x})^2}{n-1} = \frac{x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{102.4 - \frac{(23.6)^2}{11}}{11-1} = 5.18 \text{MBls}^2$$

- **Desviación estándar**

$$s = \sqrt{s^2} = \sqrt{5.18} = 2.276 \text{MBls}$$

- **Coefficiente de variación**

$$CV = \frac{s}{\bar{x}} * 100 = \frac{2.276}{2.1455} * 100 = 106\%$$

Como es mayor al 30%, indica que la información es muy heterogénea o dispersa y que la media aritmética no es un parámetro de centramiento confiable. Por lo cual, el parámetro recomendado como promedio es la mediana; es decir que 1'200.000 barriles son las reservas probadas promedio de los yacimientos descubiertos en el último año en Colombia.

- **Cuartiles**

$$\text{Posición del } Q_2 = \frac{K(n+1)}{4} = \frac{2(11+1)}{4} = 6$$

$Q_2 = 1.2$ MBIs. El 50% de los yacimientos descubiertos en el último año en Colombia tienen menos de 1.2 millones de barriles.

- **Deciles**

$$\text{Posición del } D_6 = \frac{K(n+1)}{10} = \frac{6(11+1)}{10} = 7.2 \approx 7$$

$D_6 = 1.0$ MBls. El 40% de los yacimientos tienen reservas probadas inferiores a un millón de barriles, y el 60% más de un millón.

• **Percentiles**

$$\text{Posición del } P_{35} = \frac{K(n + 1)}{100} = \frac{35(11 + 1)}{100} = 4.2 \approx 4$$

$P_{35} = 2.1$ MBls. El 35% de los yacimientos tienen reservas probadas superiores a 2.1 millón de barriles, y el 65% menos.

2. Ejercicio datos agrupados

La empresa Wentinton Oil Company (WOC), es la encargada de la parte de yacimientos y pruebas de presión del campo Kalapas. En los últimos 9 años WOC le ha realizado al pozo Kalapas 32 una serie de pruebas de presión en fondo de pozo (well bore). Durante este tiempo la empresa le realizó al pozo Kalapas 32 un total de 36 pruebas de presión, cuyos datos en psi son los siguientes:

Presiones ordenadas de menor a mayor (psi)

900, 910, 920, 990, 1100, 1500, 1600, 1650, 1680, 1700, 1750, 1900, 2000, 2050, 2300, 2350, 2400, 2500, 2600, 2650, 2700, 2760, 2800, 2920, 3000, 3005, 3005, 3020, 3110, 3200, 3300, 3350, 3390, 3400, 3500, 3600.

Se quieren calcular los parámetros estadísticos suficientes para analizar el comportamiento de la presión en el fondo del pozo y con ello tener un estimativo de la caída de presión en el yacimiento.

Calcule los parámetros de tendencia central, de dispersión, y para los parámetros de posición halle el tercer cuartil, el segundo decil y el percentil número 40.

Solución

X_i = observación individual de las presiones del pozo

n = número total de datos

• **Número de clases de intervalos:** $\bar{n} = \overline{36} = 6$

• **Rango:** $x_n - x_1 = 3600 - 900 = 2700 \text{ psi}$

• **Amplitud de clase**

$$A = \frac{\text{Rango}}{\text{Numero de clases}} = \frac{2700}{6} = 450 \text{ psi}$$

• **Tabla de frecuencia**

Tabla 3. Tabla de frecuencia (Ejercicio 2)

Clases	Marca de clase (X_c)	Frecuencia absoluta (f_i)	Frecuencia abs. Acumulada (F_i)	Frecuencia relativa (h) (f_i/n)	Frec. relativa Acumulada (H)
900 –1350	1125	5	5	0.139	0.139
1351–1800	1575	6	11	0.167	0.306
1801–2250	2025	3	14	0.083	0.389
2251–2700	2475	7	21	0.194	0.583
2701–3150	2925	8	29	0.223	0.806
3151–3600	3375	7	36	0.194	1.00
		n = 36		1.00	

- **Media Aritmética**

$$x = \frac{fx_c}{n} = \frac{5 * 1125 + 6 * 1575 + \dots + 7 * 3375}{36} =$$

$$= 2387.5 \text{ psi}$$

- **Moda**

$$(Mo) = L_{i\ mo} + A \frac{\Delta_a}{\Delta_a + \Delta_b} = 2701 + 450 \frac{1}{1+1} = 2926 \text{ psi}$$

- **Mediana**

$$(Me) = L_{i\ md} + A \frac{\frac{n}{2} - F}{f_{md}} = 2251 + 450 \frac{18 - 14}{7} = 2508.1 \text{ psi}$$

- **Varianza**

$$s^2 = \frac{f_i (x_i - x)^2}{n - 1}$$

$$= \frac{5(1125 - 2387.5)^2 + 6(1575 - 2387.5)^2 + \dots + 7(3375 - 2387.5)^2}{36 - 1}$$

$$s^2 = 614732.1 \text{ psi}^2$$

- **Desviación estándar**

$$s = \sqrt{s^2} = \sqrt{614732.1} = 784 \text{ psi}$$

- **Desviación media**

$$Dm = \frac{f_i |x_i - x|}{n} =$$

$$= \frac{5|1125 - 2387.5| + 6|1575 - 2387.5| + \dots + 7|3375 - 2387.5|}{36}$$

$$Dm = \frac{24100}{36} = 669.4psi$$

- **Coeficiente de variación**

$$CV = \frac{S}{x} * 100 = \frac{784}{2387.5} * 100 = 32.84 \%$$

- **Cuartiles**

Primero se determina la clase donde está el tercer cuartil, así:

$KN/4 = (3*36)/4 = 27$, o sea que el Q_3 es el 27° término de la serie y este queda en la clase 2701–3150

$$Q_3 = L_i + A \frac{\frac{KN}{4} - F_{i-1}}{f_i} = 2701 + 450 \frac{\frac{3*36}{4} - 21}{8} = 3038.5psi$$

- **Deciles**

Primero se determina la clase donde está el segundo decil, así:

$KN/10 = (2*36)/10 = 7.2$, o sea que el D_2 es el 7.2° término de la serie y este queda en la clase 1351–1800

$$D_2 = L_i + A \frac{\frac{KN}{10} - F_{i-1}}{f_i} = 1351 + 450 \frac{\frac{2*36}{10} - 5}{6} = 1516 \text{ psi}$$

- **Percentiles**

Primero se determina la clase donde está el percentil 40, así:

$KN/100 = (40*36)/100 = 14.4$, o sea que el P_{40} es el 14.4° término de la serie y este queda en la clase 2251 -2700

$$P_{40} = L_i + A \frac{\frac{KN}{100} - F_{i-1}}{f_i} = 2250 + 450 \frac{\frac{40*36}{100} - 14}{7} = 2275.7 \text{ psi}$$

3. Ejercicio parámetros de concentración

La compañía Cotracal Geology extrajo de una cuenca sedimentaria ubicada en el sur – occidente de Colombia varios núcleos de la formación cachemira, la firma cuenta con laboratorios calificados para realizarle a los núcleos las diferentes pruebas petrofísicas. Una de las pruebas exigidas al laboratorio fue la de permeabilidad, cuyos resultados fueron reportados en la siguiente tabla de frecuencias en milidarcy (md):

Tabla 4. Permeabilidades y frecuencias del núcleo (Ejercicio3)

Clases ($L_i - L_s$)md	Marca de Clase X_i	Frecuencia absoluta f_i
0 – 50	25	23
50 – 100	75	72
100 – 150	125	48
150 – 200	175	19
250 – 300	225	8
300 – 350	275	14
350 – 400	325	7
400 – 450	425	5
450 – 500	475	2

La compañía quiere conocer si este yacimiento tiene una permeabilidad homogénea o heterogénea. Usted como ingeniero de yacimientos lo debe responder utilizando el índice de GINI y la curva de Lorenz para dar una respuesta a la compañía.

Solución

Los datos obtenidos se tabulan en una tabla de frecuencias.

Tabla 5. Tabla de frecuencia (Ejercicio3)

CLASES Li - Ls	Marca xi	Frecuencia Absolutas		xi*fi	ui	qi = (ui/Un)	pi = (Fi/n)	pi - qi
		fi	Fi					
0 - 50	25	23	23	575	575	0,0148	0,0885	0,0737
50 - 100	75	72	95	5400	5975	0,1538	0,3654	0,2116
100 - 150	125	62	157	7750	13725	0,3533	0,6038	0,2506
150 - 200	175	48	205	8400	22125	0,5695	0,7885	0,2190
200 - 250	225	19	224	4275	26400	0,6795	0,8615	0,1820
250 - 300	275	8	232	2200	28600	0,7362	0,8923	0,1561
300 - 350	325	14	246	4550	33150	0,8533	0,9462	0,0929
350 - 400	375	7	253	2625	35775	0,9208	0,9731	0,0522
400 - 450	425	5	258	2125	37900	0,9755	0,9923	0,0168
450 - 500	475	2	260	950	38850	1,0000	1,0000	0,0000
		260		38850			6,5115	1,2548

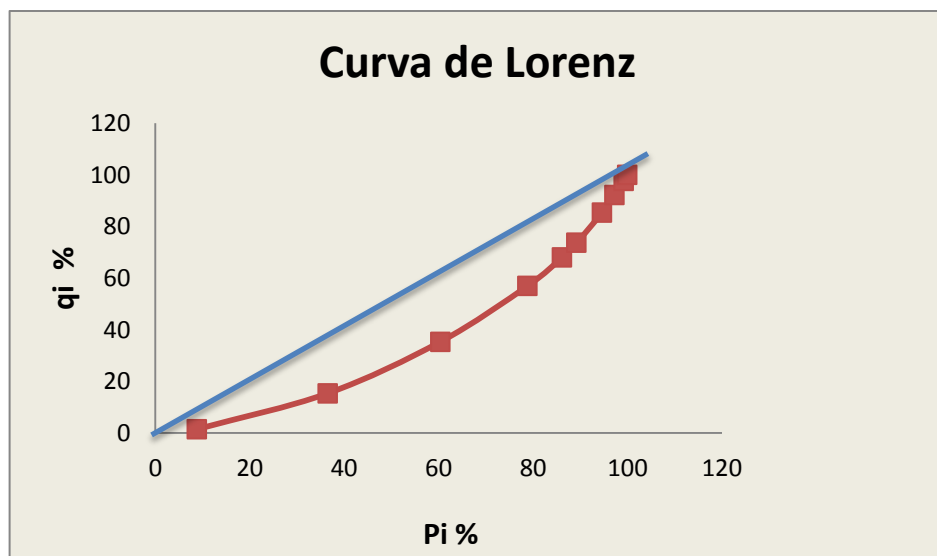
a. Índice de concentración de GINI

$$I_G = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = \frac{1.2548}{6.5115} = 0.193$$

Se observa que no hay buena concentración ya que el índice de concentración de GINI hallado se encuentra lejos de cero. Por lo general se dice que un yacimiento es homogéneo si el índice de concentración de GINI es menor a 0.083, valor que corresponde a una distribución completamente normal.

b. Curva de Lorenz

Figura 2. Curva de Lorenz (Ejercicio 3)



La curva se obtuvo alejada de la diagonal, lo que indica que hay poca concentración, o sea que el yacimiento es heterogéneo.

1.6 EJERCICIOS PROPUESTOS

1. Oscar Vanegas, experto en economía petrolera, ha decidido comprobar una serie de hipótesis sobre el comportamiento del precio WTI y hacer unos estimativos sobre el futuro precio del petróleo. Por esta razón tomó los precios WTI promedio

de los últimos 28 años, que se dan a continuación, con el fin de calcular el precio promedio general y su aleatoriedad: 25, 20, 15, 15, 17, 14, 18, 22, 19, 16, 19, 19, 18, 19, 17, 14, 19, 21, 23, 25, 27, 38, 76, 103, 77, 80, 88, 87.

Igualmente se quiere calcular el precio de quiebre donde el 50% de ellos han sido superiores e inferiores, y cuál es el precio que se ha presentado con más frecuencia en dicho periodo. Ayúdele a calcular dichos parámetros.

2. Se desea estimar la porosidad y permeabilidad promedio de un pozo, junto con sus desviaciones estándares. Igualmente, se desea conocer el P10 y el P90. Los datos petrofísicos fueron obtenidos en laboratorio de muestras cilíndricas de un corazón de 10 ft (pie). Use el promedio aritmético ponderado.

Tabla 6. Datos de porosidad y permeabilidad (Ejercicio 2)

Profundidad (ft)	Porosidad (%)	Permeabilidad milidarcy (md)
3690-3691	20	10
3691-3693	23	100
3693-3694	21	50
3694-3697	26	200
3697-3698	18	70
3698-3700	22	12

3. Los sindicalistas de la Unión Sindical Obrera (USO) argumentan que, en contravención de la convención colectiva de trabajo, los ingenieros de campo del área de producción, hacen un promedio salarial mensual menor y con una mayor variabilidad, que los empleados de oficina. Una muestra de 10 ingenieros de campo y 10 empleados de oficina arroja los siguientes datos de salarios mensuales en millones de pesos:

Tabla 7. Datos de empleados (Ejercicio 3)

Empleados	de producción	de oficina
1	2.15	5.12
2	8.17	8.42
3	9.42	7.12
4	5.17	6.92
5	8.63	8.15
6	6.42	5.81
7	5.49	9.12
8	8.73	9.15
9	9.12	8.73
10	8.36	9.66

¿El sindicato de la USO tiene razón en su reclamo? Explique su respuesta.

4. El sistema de bombeo gas lift en algunos pozos petroleros del campo Orito, de propiedad de Ecopetrol, utiliza Timer para controlar la inyección de gas, que funcionan con energía solar, los cuales tienen una capacidad de almacenamiento de calor en días, diversa. A continuación se presenta una distribución de la capacidad de almacenamiento de calor (en días) de 30 Timer del campo Orito.

Tabla 8. Distribución capacidad de almacenamiento (Ejercicio 4)

Días	# de Timer	Días	# de Timer
0 – 0.99	3	4 – 4.99	5
1 – 1.99	4	5 – 5.99	3
2 – 2.99	6	6 – 6.99	2
3 – 3.99	7		

Ecopetrol quiere saber qué tan asimétrica y dispersa o concentrada está la capacidad de almacenamiento de calor en días de los Timer que tiene instalados en el campo de Orito, para evaluar su confiabilidad. Utilice los coeficientes de Fisher y Pearson, al igual que el Índice de Gini y la curva de Lorenz para ayudarle a Ecopetrol a sacar sus conclusiones.

5. Los costos de exploración y desarrollo de un campo petrolero en la cuenca “La Pelucera” es estimado mediante la correlación:

$$Y = 100X + 100(X)^{1/2}$$

Donde “X” es el tamaño del yacimiento de acuerdo a las reservas probadas, que se han catalogado discretamente de 1 a 6, donde 1 es el tamaño mínimo y 6 el máximo. “Y” está en millones de dólares actuales. La siguiente tabla muestra los ingresos obtenidos históricamente, en millones de dólares actuales, de los 20 campos maduros explotados en la cuenca, según su tamaño:

Tabla 9. Ingresos de los 20 campos (Ejercicio 5)

<u>“X”</u>	<u>INGRESOS</u>
1	500
2	500, 600, 700, 800
3	600, 700, 800, 900
4	700, 900, 1000, 1000, 1200
5	900, 1100, 1200, 1400
6	1300, 1500

Si usted fuera a explorar un área en dicha cuenca, ¿cuál será la utilidad esperada para el campo que llegare a descubrir?

6. Los ingresos por venta de gas, las compañías gasíferas lo reportan al MINMINAS cada trimestre. Los siguientes datos representan los ingresos en millones de US\$ durante el primer trimestre de 2011 para 9 compañías: 16, 18, 11, 17, 13, 10, 22, 15, 16.

- (a) Calcule la media, la mediana, la moda y la desviación estándar.
- (b) Si hay un total de 35 compañías gasíferas en el país, estime el ingreso total por ventas de gas al trimestre y durante todo el año.

- (c) Estime, para el trimestre y para el año, la desviación estándar de la suma total de ingresos por ventas de gas en el país.

7. Complete la siguiente tabla:

Tabla 10. Tabla de frecuencia (Ejercicio 7)

Temperatura °C	Xc		Hi	fi	Fi
					25
		0.15			
-18.5				50	
			0.45		
		0.2			
3.5			0.75		
		0.05			
				75	
TOTALES		1		500	

8. Una empresa petrolera ha pagado a sus empleados un salario promedio mensual de \$4.5 millones, con una desviación estándar de \$800 mil, sin subsidios ni bonificaciones. Este año han creado un sindicato y están solicitando un aumento del 10%, una bonificación mensual de \$300 mil, un descuento para el sindicato del 2%, un subsidio de transporte de \$100 mil y un subsidio de alimentación de 200 mil. El descuento para salud es del 4.5% y para pensión del 6%. Calcule:

- El nuevo salario promedio y la nueva varianza.
- Si la nómina mensual costaba 360 millones, ¿cuánto constará la nueva?

9. A continuación se muestra el número de ingenieros de petróleos por rango de años de experiencia en una empresa petrolera multinacional.

Tabla 11. Años de experiencia de ingenieros (Ejercicio 9)

AÑOS DE EXPERIENCIA	NÚMERO DE INGENIEROS
Menos de un año	35
Entre 1 y 5 años	45
Entre 5 y 10 años	78
Entre 10 y 15 años	129
Entre 15 y 20 años	56
Más de 20 años	34

- Represente gráficamente la distribución.
- Calcule la moda, la media y la mediana.
- Calcule la medida de dispersión más adecuada.
- Comente los resultados.

10. Dos fabricantes de bombas de subsuelo para bombeo mecánico anuncian que la vida promedio de sus productos es de 7 años. Sin embargo, Ecopetrol analiza los archivos de cada fabricante y extrae dos muestras aleatorias sobre la vida en años de sus bombas, obteniendo:

Fabricante A: 5,5,5,6,6,6,6,7,7,7,7,7,7,8,8,8,8,9,9,9

Fabricante B: 2,3,4,5,5,5,5,6,6,6,7,7,7,7,7,8,8,20,20,20

Determine:

- ¿Cuál medida de tendencia central coincide con el anuncio de cada fabricante?
- ¿Con cuál fabricante se sentiría usted más seguro de que la vida promedio es realmente 7 años? Justifique su respuesta.

11. Se desea estimar el factor volumétrico, la solubilidad del gas en solución y la Presión de burbuja promedios de un yacimiento. Los datos del B_o , R_{so} y P_b fueron obtenidas de pruebas PVT para tres muestras recombinadas en superficie de tres pozos. Use el promedio aritmético y el aritmético ponderado.

Tabla 12. Datos PVT (Ejercicio 11)

No Pozo	Pb Psia	B _o RB/STB	R _s SCF/STB	Área Acres	Espesor Ft
1	1875	1.41	310	120	35
2	1685	1.38	280	80	25
3	1950	1.52	450	100	40

12. Se toman varias muestras de corazones de un yacimiento, a los cuales mediante pruebas de laboratorio se les determina la porosidad. En la siguiente tabla se encuentran los rangos de porosidad y su respectiva frecuencia. ¿Usted como encargado del laboratorio petrofísico del ICP, se le pide que determine si este yacimiento tiene una porosidad homogénea o heterogénea? Utilice el índice de GINI y la curva de Lorenz para determinarlo.

Tabla 13. Datos de porosidad (Ejercicio 12)

Clases $L_i - L_{s\emptyset}$ (%)	Frecuencia f_i
6 – 8	5
8 – 10	37
10 -12	75
12 – 14	90
14 – 16	96
16 - 18	101
18 – 20	256
20 – 22	300

2. PROBABILIDAD

La probabilidad es la medición de la posibilidad de que algo ocurra, y ese algo puede ser un suceso o un evento. Esta medición comprende valores entre 0 y 1. Entre mayor sea la posibilidad de que ocurra un evento, su probabilidad asignada estará más próxima a 1. La probabilidad de certeza plena es 1, mientras que la probabilidad de algo imposible de ocurrir es 0.

El proceso que produce un evento se llama acción y puede ocurrir experimentalmente o al azar. Un experimento es toda acción bien definida que conlleva a un resultado único bien definido, lanzar un dado es un experimento. El resultado obtenido después del evento se llaman suceso, y el conjunto de todos los sucesos es el espacio muestral.

Para el cálculo de las probabilidades se deben tener presentes las técnicas de conteo, con lo cual se estima el espacio muestral. Aquí se repasan algunos métodos para su cálculo.

2.1 TECNICAS DE CONTEO

Muchas decisiones comerciales requieren que se cuente el número de subconjuntos que se pueden obtener de un conjunto. Se analizarán cuatro técnicas de conteo: combinaciones, permutaciones, escogencia múltiple y multiplicación.

Al seleccionar los elementos en los subconjuntos, la distinción entre permutaciones y combinaciones depende de si el orden de las selecciones hace la diferencia. Si un orden es suficiente para construir otro subconjunto, entonces se trata de permutaciones. Si dos subconjuntos se consideran iguales debido a que

simplemente se han reordenado los mismos elementos, entonces se involucran las combinaciones.

Si se considera que dos subconjuntos son diferentes porque el orden es diferente, se considera como permutaciones. Si se considera que dos subconjuntos son idénticos y constituyen el mismo subconjunto porque ambos tienen los mismos elementos sin considerar el orden, son denominados combinaciones. Es decir con las permutaciones el orden es el que hace la diferencia.

Dado un conjunto de n elementos, el número de permutaciones, cada uno de tamaño r , se determina como:

$${}_n P_r = \frac{n!}{n-r!} \quad (2-1)$$

En donde $n!$ se lee como “ n factorial” y significa el producto de todos los números de 1 a n . El número de combinaciones de n elementos tomados r a la vez es:

$${}_n C_r = \frac{n!}{n-r!} \quad (2-2)$$

Ni las permutaciones ni las combinaciones permiten que se seleccione un elemento más de una vez. Si se permite la duplicación, se utilizaría el método de escogencia múltiple de conteo. El número de arreglos de escogencia múltiple de n elementos tomados r a la vez es:

$${}_n M_r = n^r \quad (2-3)$$

Si se requiere escoger un elemento de dos o más conjuntos, es adecuado el proceso de multiplicación. Este principio requiere que simplemente se multiplique el número de elementos en cada conjunto.

$$n_1 * n_2 * n_3 \dots * n_n \quad (2-4)$$

2.2 ENFOQUES DE LA PROBABILIDAD A PRIORI

A pesar de la difundida aplicación de los principios de la probabilidad, existen solo tres formas generalmente aceptadas para calcular la probabilidad a priori: (1) el enfoque clásico (2) el enfoque frecuencial, y (3) el enfoque subjetivo.

2.2.1 Enfoque clásico: El enfoque clásico dice que si hay X posibles resultados favorables a la ocurrencia de un evento, y **que todos los resultados son igualmente posibles**¹² y mutuamente excluyentes (no pueden ocurrir los dos al mismo tiempo), entonces la probabilidad de que ocurra el evento se determina mediante:

$$P E = \frac{N^{\circ} \text{ de formas en las que puede ocurrir un evento}}{N^{\circ} \text{ total de posibles resultados}} \quad (2 - 5)$$

2.2.2 Enfoque frecuencial: El enfoque frecuencial utiliza el cálculo de la frecuencia relativa a partir de datos que se han observado empíricamente mediante una experimentación o información histórica registrada y tabulada; donde se contabiliza el número de veces que se presenta algún evento o suceso en particular, y se estima la probabilidad de que éste ocurra nuevamente con base en frecuencia relativa de estos datos. La probabilidad de un evento o suceso con base en el enfoque frecuencial se determina mediante:

$$P E = \frac{N^{\circ} \text{ de veces que ha ocurrido el evento en el pasado}}{N^{\circ} \text{ total de observaciones}} \quad (2 - 6)$$

2.2.3 Enfoque subjetivo: El enfoque subjetivo es un enfoque personalista y se utiliza cuando no se tiene información registrada, pero se cuenta con la experiencia de la ocurrencia de los hechos. Si el evento o suceso nunca ha

¹²Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia.

ocurrido con anterioridad se acude a la intuición. Debido a que no hay datos sobre los cuales confiar, se deben analizar las opiniones y creencias personales para obtener una estimación subjetiva, por lo cual no existen fórmulas para calcularla.

2.3 RELACIONES ENTRE SUCESOS Y EVENTOS

Sucesos mutuamente excluyentes. Dos o más sucesos son mutuamente excluyentes, si no pueden ocurrir simultáneamente. Es decir, la ocurrencia de un suceso impide automáticamente la ocurrencia del otro suceso (o sucesos).

Sucesos mutuamente incluyentes. Dos o más sucesos son mutuamente incluyentes, cuando es posible que ocurran ambos. Esto no indica que necesariamente deban ocurrir estos sucesos en forma simultánea.

Eventos independientes. Dos o más eventos son independientes cuando la ocurrencia o no-ocurrencia de un evento no tiene efecto sobre la probabilidad de ocurrencia del otro evento (o eventos).

Eventos dependientes. Dos o más eventos serán dependientes cuando la ocurrencia o no-ocurrencia de uno de ellos afecta la probabilidad de ocurrencia del otro (u otros). Esto conduce a la probabilidad condicional.

2.4 PROBABILIDAD CONDICIONAL O A POSTERIORI

Con frecuencia se desea determinar la probabilidad de algún evento A, dado que antes otro evento B ya haya ocurrido, siempre y cuando los eventos sean dependientes entre sí. Lógicamente, esta es llamada probabilidad condicional. Se denota como $P(A|B)$ y se lee la “probabilidad de A dado B”. Esta es la ecuación general para la probabilidad condicional del evento A, dado que se conoce que el evento B ya ocurrió.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) P(A|B)}{P(B)} \quad (2 - 7)$$

2.5 REGLAS DE LA PROBABILIDAD

Existen dos reglas básicas para calcular la probabilidad de eventos más complejos: la regla de la multiplicación y la regla de la adición, cada una se utiliza para propósitos específicos.

2.5.1 Regla de la adición. La regla de la adición establece que si dos sucesos “a” y “b” son mutuamente excluyentes la probabilidad de que uno u otro suceso ocurra es igual a la suma de sus probabilidades.

$$P(a \cup b) = P(a) + P(b) \quad (2 - 8)$$

Cuando los sucesos “a” y “b” son mutuamente incluyentes, ambos sucesos pueden ocurrir al mismo tiempo. En este caso, la fórmula requiere que se reste la probabilidad del suceso conjunto “a” y “b”.

$$P(a \cup b) = P(a) + P(b) - P(a \cap b) \quad (2 - 9)$$

La razón por la cual se debe restar la probabilidad conjunta cuando los sucesos son mutuamente incluyentes es para evitar el doble conteo.

2.5.2 Regla de la multiplicación. La regla de la multiplicación tiene como propósito determinar la probabilidad del evento conjunto $P(A \cap B)$. Es decir, que para encontrar la probabilidad de A y B, simplemente se multiplican sus respectivas probabilidades. El procedimiento exacto depende de si A y B son dependientes o independientes. Los eventos A y B son independientes si $P(A) = P(A|B)$. Es decir, la probabilidad de A es la misma bien se considere o no el evento B. de igual forma, si A y B son independientes si $P(B) = P(B|A)$. Para eventos independientes la probabilidad de dos eventos se vuelve.

$$P A \cap B = P A * P B \quad (2 - 10)$$

Si los eventos son dependientes, entonces, por definición, se debe considerar el primer evento al determinar la probabilidad del segundo. Es decir, la probabilidad del evento B depende de la condición que A ya haya ocurrido.

$$P A \cap B = P A * P B A \quad (2 - 11)$$

2.6 TEOREMA DE BAYES

El Teorema de Bayes¹³, proporciona la distribución de probabilidad condicional de un evento "B" dado otro evento "A" (probabilidad a posteriori), en función de la distribución de probabilidad condicional del evento B dado A y de la distribución de probabilidad marginal del evento B (probabilidad simple o a priori conjunta). O sea, el Teorema de Bayes proporciona una fórmula que permite calcular una de las probabilidades condicionales si se conoce la otra.

Sean B_k sucesos mutuamente excluyentes tales que, cualquier suceso B_k en el espacio muestral pertenece a uno y sólo a uno de estos sucesos. Entonces la probabilidad de que ocurra cualquier suceso B_k dado que ha ocurrido el evento " A_j " se calculará mediante:

$$P B_k \cap A_j = P A * P B_k A_j \quad (2 - 12)$$

$$P B_k A_j = \frac{P B_k \cap A_j}{P A} \quad (2 - 13)$$

¹³John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

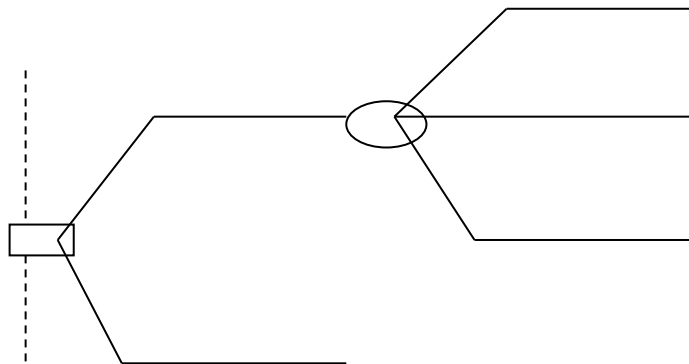
Por lo tanto sustituyendo la ecuación de probabilidad condicional, se obtiene la ecuación general para el Teorema de Bayes.

$$P(B_k | A_j) = \frac{P(B_k \cap A_j)}{P(A_j)} = \frac{P(B_k) * P(A_j | B_k)}{P(B_k) * P(A_j | B_k) + \dots} \quad (2 - 14)$$

2.7 ARBOLES DE DECISION

Un árbol de decisión es un método gráfico que expresa en orden cronológico las alternativas viables para la toma de decisiones como se puede observar en la figura 3. Representa una serie de cuestiones o pautas basadas en un conjunto de variables que explican la variabilidad de la variable dependiente o de resultado, que se plasman como un camino que recorre el árbol.

Figura 3. Árbol de decisión



- Punto de decisiones
- Nodo o Punto de Evento
- Decisiones o Sucesos

Fuente: www.mitecnologico.com/Main/ArbolesDeDecision

2.8 EJERCICIOS RESUELTOS

1. Ejercicio probabilidad condicional

El taller de bombas de subsuelo de la Superintendencia del Rio de la Gerencia Magdalena Medio de Ecopetrol S.A, desea optimizar y programar el mantenimiento y reparación de las bombas utilizadas para el levantamiento artificial por bombeo mecánico del petróleo que se produce en los campos que administra y opera. Allí se realiza el mantenimiento y la reparación de dos tipos de bombas de subsuelo: THC, RHAC, algunas de las cuales han sido acondicionadas con cámaras laterales para el venteo de gas (CLVG). La siguiente tabla resume la proporción de ellas que tienen o no el sistema de venteo, de acuerdo a su tipo:

Tabla 14. Características de las bombas (Ejercicio 1)

Tipo de Bomba	Con CLVG	Sin CLVG	
THC	0.003	0.017	0.020
RHAC	0.003	0.977	0.980
	0.006	0.994	1.000

Del total de bombas de los campos de la Superintendencia del Rio, la probabilidad de que una bomba tenga cámara lateral de venteo es $P(A) = 0.006$, y la probabilidad de que sea del tipo THC y que tenga cámara lateral de venteo es $P(A \cap B) = 0.003$. Dado que una bomba seleccionada al azar sea del tipo THC, ¿Cuál es la probabilidad de que tenga cámara lateral de venteo?

Solución

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.003}{0.020} = 0.15$$

Es decir que dentro de las bombas tipo THC existe un 15% de ellas que tienen cámara lateral de venteo de gas (CLVG).

Ahora se calcula la probabilidad de que una bomba tenga CLVG dado que es del tipo RHAC.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.003}{0.980} = 0.003$$

Es decir que solo el 0,3% de las bombas RHAC, tienen cámara lateral de venteo de gas.

2. Ejercicio regla de la adición

Revisando los registros de un campo petrolero se encuentra que de un total de 1000 pozos, 800 tienen registros Gamma Ray (GR), 600 tienen registros SP y 500 tienen ambos. ¿Cuál es la probabilidad de que un pozo seleccionado al azar tenga cualquiera de los dos registros, ya sea GR o SP?

Solución

$$P(a \cup b) = P(a) + P(b) - P(a \cap b)$$

$$P(a \cup b) = 0.8 + 0.6 - 0.5 = 0.9$$

O sea que el 10% de los pozos (100) no tienen ninguno de los dos registros.

3. Ejercicio regla de la multiplicación

Una compañía de servicio de lodos para la perforación de pozos petroleros está considerando introducir dos nuevos tipos de lodos base aceite al mercado local. El gerente cree que la posibilidad de éxito es alrededor del 50% para el primero y

75% para el segundo. ¿Cuál es la probabilidad de que ambos productos tengan éxito? Ambas probabilidades son subjetivas porque están basadas en la opinión del gerente al no haber aún estadísticas de ventas. También se hace la suposición de que el éxito de los dos tipos de lodos es independiente el uno del otro.

Solución

$$P(A \cap B) = P(A) * P(B)$$

$$P(A \cap B) = 0.50 * 0.75 = 0.375$$

Existe alrededor de un 37.5% de probabilidad de que ambos lodos tengan éxito simultáneamente.

4. Ejercicio Teorema de Bayes

Una empresa petrolera tiene tres campos (A, B, C) productores de crudo pesado donde algunos pozos tienen problemas de parafinas. Los campos A, B y C tienen respectivamente el 50%, 30% y 20% de los pozos activos totales, cuyos registros de producción se archivan conjuntamente. El porcentaje de pozos con el problema de parafinas, en cada campo, es 3%, 4% y 5% respectivamente. Si se selecciona una carpeta de un pozo al azar y éste muestra problemas de parafinas, ¿cuál será la probabilidad de que pertenezca al campo C?

Solución

$$P(A) = 0.50 \quad P(P/A) = 0.03$$

$$P(B) = 0.30 \quad P(P/B) = 0.04$$

$$P(C) = 0.20 \quad P(P/C) = 0.05$$

$$P(C|P) = \frac{P(C \cap P)}{P(P)}$$

$$P(C|P) = \frac{P(C) * P(P|C)}{P(A) * P(P|A) + P(B) * P(P|B) + P(C) * P(P|C)}$$

$$P(C|P) = \frac{0.2 * 0.05}{0.5 * 0.03 + 0.3 * 0.04 + 0.2 * 0.05} = 0.27$$

Existe un 27% de probabilidad de que el pozo seleccionado con problema de parafina pertenezca al campo C.

5. Ejercicio árboles de decisión

Es el 28 de enero y Tomás Róbaló consultor jurídico de ECOPETROL afronta un difícil problema. Parece que la empresa se enfrenta a dos litigios correlacionados con el ministerio del medio ambiente por daños generados por derrames de crudo. Para ambos litigios tiene la opción de un arreglo directo, de lo contrario se irían a juicio. La fecha del juicio número 1 está programada para el 15 de julio y el juicio número 2 para el 8 de octubre. Los costos de preparación de un juicio se estiman en \$50 millones; sin embargo, si la empresa prepara ambos el precio del segundo será de \$30 millones. Esto puede evitarse mediante un arreglo directo. Para el primer litigio el arreglo directo consiste en pagar una multa de \$400 millones, y para el litigio 2 de \$250 millones. Estas multas ya le fueron notificadas por el ministerio de medio ambiente.

Si la empresa va al juicio 1 y lo gana, no pagará la multa; pero si lo pierde, la penalización puede ser de \$1000, \$1500 o \$2000 millones, según estimaciones del abogado; con probabilidades del 50%, 30% y 20% respectivamente. Tomás considera que la probabilidad de ganar el juicio número 1 es del 50%.

El juicio número 2 puede tener uno de los siguientes resultados: a) que lo gane la empresa, en cuyo caso no pagará la penalización, b) que la empresa sea declarada culpable sin intención (CS), y deberá pagar la multa, c) que la empresa sea declarada culpable con alta responsabilidad (CC), por lo cual puede ser condenada a pagar \$1000, \$800 o \$500 millones, con probabilidades del 60%, 30% y 10%, respectivamente.

Los resultados del juicio 2 están condicionados por el resultado del juicio 1, pues el juez lo considera como un precedente importante. Tomás ha valorado las

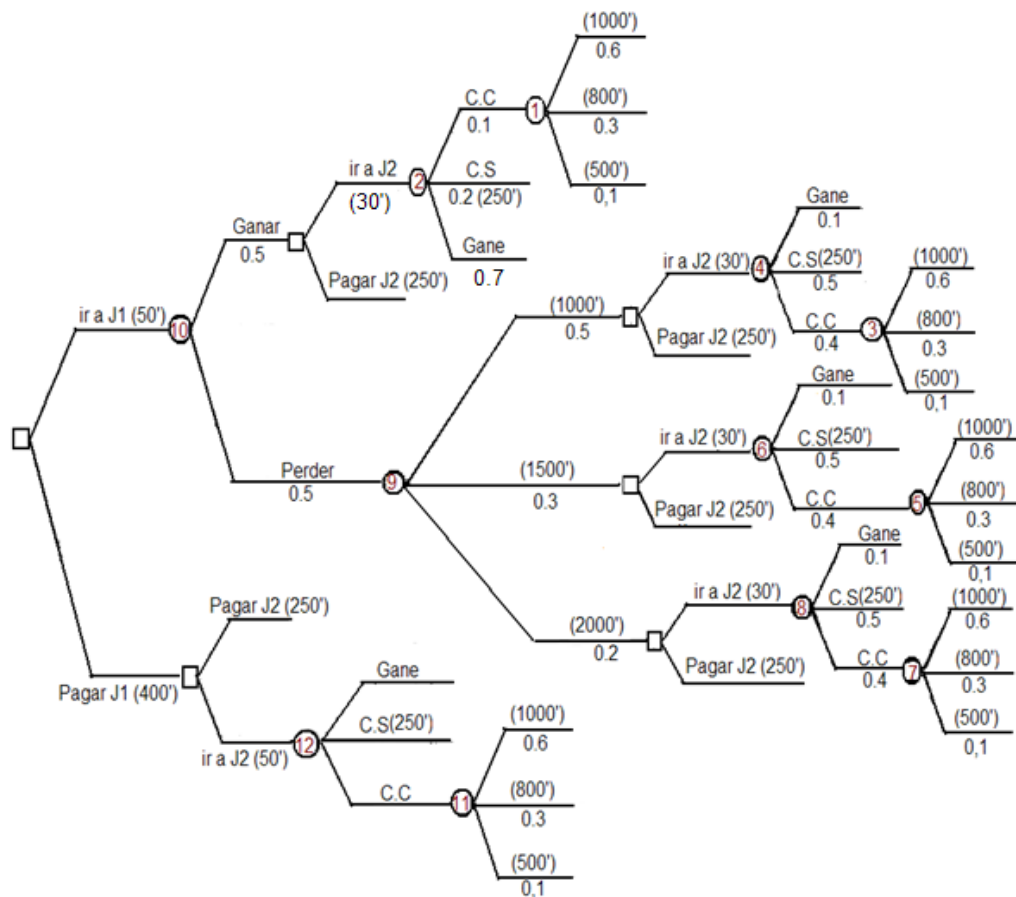
probabilidades de los resultados del juicio 2 bajo las condiciones del juicio 1 mediante el teorema de Bayes. Las probabilidades se encuentran resumidas en la siguiente tabla. ¿Encuentre la estrategia óptima para la empresa?

Tabla 15. Probabilidades de Ecopetrol (ejercicio 5)

Resultado J2	No Hubo J1	Gana J1	Pierde J2
Gane	0.3	0.7	0.1
Culpable sin Responsabilidad	0.3	0.2	0.5
Culpable con Responsabilidad	0.4	0.1	0.4

Solución

Figura 4. Árbol de decisión (Ejercicio 5)



VE: Valor esperado

$$VE_1 = 1080(0.6) + 880(0.3) + 580(0.1) = 970$$

$$VE_2 = 80(0.7) + 330(0.2) + 970(0.1) = 219$$

$$VE_3 = 2080(0.6) + 1880(0.3) + 1580(0.1) = 1970$$

$$VE_4 = 1080(0.1) + 1330(0.5) + 1970(0.4) = 2101$$

$$VE_5 = 2580(0.6) + 2380(0.3) + 2080(0.1) = 2470$$

$$VE_6 = 1580(0.1) + 1830(0.5) + 2470(0.4) = 2061$$

$$VE_7 = 3080(0.6) + 2880(0.3) + 2580(0.1) = 2970$$

$$VE_8 = 2080(0.1) + 2330(0.5) + 2970(0.4) = 2561$$

$$VE_9 = 1300(0.5) + 1800(0.3) + 2300(0.2) = 1650$$

$$VE_{10} = 219(0.5) + 1650(0.5) = 934.5$$

$$VE_{11} = 1450(0.6) + 1250(0.3) + 950(0.1) = 1340$$

$$VE_{12} = 450(0.3) + 700(0.3) + 1350(0.4) = 881$$

La mejor opción para la empresa es pagar las multas de los litigios, haciendo un arreglo directo, sin ir a los juicios.

2.9 EJERCICIOS PROPUESTOS

1. Entre los 304 campos de petróleo crudo que se han descubierto en Colombia, 50 tienen más de 100 millones de barriles (MBbls) de reservas originales (OOIP), y entre los 20 campos de gas natural, 4 tienen reservas originales superiores a 100 millones de barriles equivalentes (MBE):

Tabla 16. Reservas de crudo (Ejercicio 1)

	< 100 MB	> 100 MB	Total
Petróleo (P)	254	50	304
Gas (G)	16	4	20
Total	270	54	324

Para realizar pronósticos de reservas a priori, encuentre la probabilidad (Probabilidad condicional) de que un campo que se encuentra en exploración:

- a. Tenga > 100 MB, dado que se acaba de probar que es petróleo crudo.
- b. Tenga < de 100 MBE, dado que se acaba de probar que es de gas.
- c. Sea de petróleo crudo si se sabe que la estructura contiene > 100 MB.

2. La compañía Dow Química tiene una máquina automática que llena canecas de 55 galones con el químico D-100 para romper emulsiones petróleo-agua. La mayoría de las canecas se llenan en forma adecuada; sin embargo, algunas se llenan más de lo necesario y a otras les falta líquido. Una muestra aleatoria de 1000 canecas observadas dio como resultado:

Tabla 17. Datos de volumen y probabilidad (Ejercicio 2)

Suceso	Galones	Número de Canecas	Probabilidad
A	< 55	20	0.02
B	55	940	0.94
C	> 55	40	0.04

¿Cuál es la probabilidad de que una caneca cualquiera, seleccionada al azar, esté más llena o menos llena de lo que debe? (Regla de la adición)

3. Los datos históricos muestran que el 15% de las Bombas PCP (bombas de cavidades progresivas) de un campo petrolero necesitan motor nuevo durante los primeros dos años de operación. De aquellas que necesitan motores nuevos, el 80% también necesita una correa de transmisión nueva durante el mismo periodo. ¿Cuál es la probabilidad de que una bomba PCP necesite tanto un nuevo motor como una correa de transmisión nueva durante los primeros dos años?

4. Oscar Pinzón Rico, el gerente de producción del campo La Pelucera, está tratando de mejorar la eficiencia de su cuadrilla de workover durante la semana. Cada lunes por la mañana se reúne con ellos y les da una charla de motivación. Oscar Pinzón Rico dice que la charlas ale como lo espera solo el 70% de las veces. También estudia los reportes de mantenimiento de pozos y descubre que cuando la charla sale bien, su cuadrilla realiza excelentemente el 95% de las tareas de la semana (calificado como desempeño bueno), pero si no sale como lo esperaba, su cuadrilla solo logra realizar de forma excelente el 40% de las tareas de mantenimiento (desempeño regular).

Oscar decide revisar el reporte de una semana cualquiera y encuentra que el desempeño de la cuadrilla fue bueno. ¿Cuál es la probabilidad de que esa semana la charla de motivación haya salido como lo esperaba?

5. VANANAS Oil Company, empresa entre cuyos accionistas el 80% son ingenieros de petróleos, ha recibido una invitación de ECOPETROL para entrar en asociación bajo riesgo en un porcentaje del 50%, para perforar un pozo exploratorio a 15 mil pies (ft) de profundidad, en un terreno que según estudios sísmicos es muy prometedor, habiéndose evaluado una probabilidad de éxito del 30%, en contraste con la probabilidad general del 5% que existe para el país. Pero, allí existen dos formaciones que han manifestado problemas durante la perforación de pozos en campos aledaños, una intermedia de arcillas poco consolidadas que ha generado pegas y otra profunda de presión subnormal que ha presentado pérdidas de circulación, por lo cual los costos de perforación tienen tres escenarios: 50, 60 o 80 millones de dólares, con probabilidades de 0.2, 0.3 y 0.5 respectivamente. Pero, si la empresa acepta la invitación de ECOPETROL, deberá aplazar el proyecto que se encuentre evaluando en este momento, aunque cuenta con un capital de riesgo de 100 millones de dólares. Este último proyecto consiste en perforar un pozo exploratorio que tiene como prospecto en un bloque asignado por la ANH, del cual la empresa VANANAS Oil Company tiene contrato

de concesión moderna. Si la empresa perfora su pozo o se asocia con ECOPETROL y no tiene éxito, fracasará y liquidará. Pero si tiene éxito se desarrollará el campo en cualquiera de los casos, obteniéndose los valores presentes netos (VPN en millones de dólares) de los flujos de caja (después de impuestos) proyectados para los siguientes 24 años, tal como se muestra en la tabla, dependiendo del estimativo de reservas y de proyección de precios del barril en el mercado internacional.

Tabla 18. Flujos de caja (Ejercicio 5)

VPN EN MM US\$							
RA PA	RA PB	RM PA	RM PB	RR PA	RR PB	RB PA	RB PB
10000	5000	6000	3000	1500	700	800	350

Las probabilidades, si perfora en su bloque, el tamaño de reservas a descubrir, deben ser calculadas tomando como base campos productores existentes en Colombia, entre los cuales 19 tiene reservas OOIP superiores a 500 MMBIs (Reservas Altas – RA), 35 entre 100 y 500 MMBIs (Reserva Moderadas – RM), 25 entre 50 y 100 MMBIs (Reservas Regulares – RR), y 225 con menos de 50 MMBIs (Reservas Bajas – RB). Si se asocia con ECOPETROL, debe tenerse en cuenta que esta empresa en su historia a descubierto, solo o asociado, 5 campos con más de 500 MMBIs, 15 entre 100 y 500 MMBIs, 13 entre 50 100 MMBIs y 155 con menos de 50 MMBIs. Además, se estima que durante los próximos 24 años los precios del crudo permanecerán por encima de los 80 US\$/Bbl (Precios Altos – PA) con una probabilidad del 60%, y por debajo de los 80 US\$/Bbl (Precios Bajos – PB) con la probabilidad del 40%.

Los costos de exploración y perforación del pozo en su bloque, se estiman en 25, 35 o 45 millones de dólares, con probabilidades del 0.40, 0.35 y 0.25

respectivamente. ¿Cuál es la mejor decisión para la empresa VANANAS Oil Company? , ¿Si se asocia con ECOPETROL, cuál sería su participación óptima?

6. FEPCO recibió una invitación de la SLACOP a participar en la licitación para la fabricación de Boquillas de Chorro y niebla para los próximos 3 años, para lo cual deberá diseñar y elaborar muestras para ser probadas por PARKER DRILL, incurriendo en un costo de US\$ 10.000 que será perdido si no se gana la licitación, para lo cual existe una probabilidad del 60% de ganar. Si gana, puede elaborarlas en una de las dos líneas de producción. La línea 1 involucra costos fijos de US\$ 50.000 y posibles costos variables de US\$8 ó US\$12 por unidad, con probabilidades del 50% cada una. La línea 2 tendrá un costo fijo de US\$ 80.000 y costos variables de US\$5 u US\$8 por unidad, con probabilidades de 0.6 y 0.4 respectivamente.

También puede subcontratar a Metalúrgicas Andina quien le garantiza una ganancia neta US\$ 200.000 en valor presente; pero esta empresa ha incumplido en el 30% de los trabajos, y si esto ocurre deberán pagar una multa por incumplimiento de US\$ 30.000.

FEPCO puede pensar en asociarse con talleres mayo en un 80%, contando con un presupuesto al riesgo de US\$ 500.000, existiendo las probabilidades de 50%, 30% y 20% de producirlas por la L1, la L2 ó subcontratar a Metalúrgicas Andina. Qué decisión debe tomar FEPCO. Si la Licitación es de 10.000 boquillas que serán pagas a 80 US\$ la unidad?

7. Para el próximo año la política de exploración petrolera puede ser buena, regular o mala con probabilidades de 0.6, 0.3 y 0.1 respectivamente. GEOLOG, empresa de servicios ha entrado a licitar en ECOPETROL, para realizar los estudios sísmicos que necesite la estatal durante el próximo año, para lo cual se han

estimado contratos por 60 40 ó 15 millones al año, según el precio por Km licitado y la política de exploración que se presente.

Estos estudios se han venido realizando a través de la OTIS, a un precio por Km del 20% más que el licitado por GEOLOG. Pero, no podemos olvidar ni desconocer, que OTIS ha venido corriendo los registros de pozos a precios muy por debajo de los cobrados por otras compañías, y ha informado a ECOPETROL que si no le renuevan la exclusividad para realizar los estudios sísmicos, aumentará el próximo año, en un 25% el precio por pie registrado. Por lo cual se estiman contratos por 100, 70 ó 35 millones anuales, de acuerdo a la política petrolera. Además, si ECOPETROL decide quitar la exclusividad de los estudios sísmicos a la OTIS, esta puede decidir en la convención de junio del próximo año:

- a) Pedir exclusividad para realizar los estudios sísmicos a partir de ese mes, o si no cancelará los contratos de registros.
- b) Que se retire y cancele contratos.
- c) No se toma ninguna decisión al respecto.

Se estiman las posibilidades del 10%, 50% y 40%, respectivamente, para cada una de las posibles reacciones de la OTIS en la convención.

Si la ORIS decide en junio cancelar los contratos, ECOPETROL llamará a la DOWELL para que corra los registros de los últimos 6 meses del año, estimándose contratos por 70, 40 ó 20 millones para ese período, según la política de exploración petrolera. Evalúe el problema por medio de un árbol de decisión.

3. DISTRIBUCIONES DE PROBABILIDAD

Una distribución de probabilidad es una distribución de frecuencia relativa de una variable aleatoria, por intervalos o puntos, donde se describen todos los posibles resultados de un experimento junto con la probabilidad de ocurrencia de cada resultado. Las distribuciones de probabilidad corresponden a funciones o proposiciones matemáticas que describen el comportamiento de la frecuencia relativa o la magnitud de la barra de frecuencia a partir de las marcas de clase, para cada uno de los valores de la variable, ya sean puntuales o por intervalos.

Las distribuciones de probabilidad existen tanto para variables discretas (medibles en forma puntual y entera, generalmente por conteo, dándose separaciones entre valores observables sucesivos), como para variables continuas (medibles con valores a lo largo de un intervalo continuo, con valores potenciales entre dos valores cualesquiera observables).

3.1 DISTRIBUCIONES DE VARIABLE DISCRETA

Se denomina distribución de variable discreta a aquella que puede asumir solo ciertos valores, con frecuencia números enteros, y resulta principalmente del conteo¹⁴.

3.1.1 Distribución binomial. La distribución binomial es una distribución de probabilidad discreta que mide la posibilidad de ocurrencia de un determinado número de X éxitos, en una secuencia de “ n ” ensayos independientes de Bernoulli con una probabilidad fija “ p ” de ocurrencia del éxito en cualquier prueba.

¹⁴John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, sólo son posibles dos resultados que son mutuamente excluyentes. A uno de estos se le denomina éxito y tiene una probabilidad de ocurrencia “**p**” y al otro se le denomina fracaso, con una probabilidad “**q**”(1- p). En la distribución binomial el anterior experimento se repite “**n**” veces, de forma independiente, y se trata de determinar la probabilidad de un determinado número de éxitos¹⁵.

La función de probabilidad de una distribución binomial es:

$$p(x) = \frac{n!}{x! (n-x)!} p^x (1-p)^{n-x}$$

En donde **p** ^x probabilidad de *x* éxitos en *n* pruebas

x número de éxitos

n número de pruebas

p probabilidad de éxito en cualquier prueba

Los valores de *n*, *x* y *p*, se encuentran tabulados en la tabla A, del anexo D al final del manual.

3.1.2 Distribución hipergeométrica: La distribución binomial supone que la población de la que se toma la muestra es infinita. Por esta razón, la probabilidad de éxito no cambia con las pruebas. Cuando no se puede suponer que la poblaciones infinita, o ni siquiera muy grande, la distribución binomial no es apropiada. Esto se debe a que cada vez se elige un elemento de la población, su tamaño se reduce y la probabilidad de éxito en la siguiente prueba cambia. La distribución hipergeométrica se usa para modelar estas situaciones.

¹⁵John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

La distribución hipergeométrica¹⁶ se usa para modelar situaciones de población finita en las que se toman muestras sin reemplazo y donde se quiere calcular la probabilidad de cierto número de éxitos. La fórmula para esta distribución es:

$$P_x = \frac{C_{n-x}^{N-r} \cdot C_x^r}{C_n^N} \quad (3 - 2)$$

En donde N tamaño de la población

n tamaño de la muestra

r número de éxitos en una población

x número de éxitos en una muestra para los cuales se desea la probabilidad

C combinación

3.1.3 Distribución de Poisson: La distribución de Poisson¹⁷ se usa para modelar situaciones en las que hay ocurrencias aleatorias de sucesos por unidad de espacio o tiempo y en donde se desea conocer la probabilidad de un número específico de éxitos. Son necesarios dos supuestos para la aplicación de la distribución de Poisson:

- La probabilidad de ocurrencia del evento es constante para dos intervalos cualesquiera de tiempo o espacio.
- La ocurrencia del evento en un intervalo es independiente de la ocurrencia de otro intervalo cualquiera.

¹⁶Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

¹⁷Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

Dados estos supuestos, la función de probabilidad de Poisson puede expresarse como:

$$p_x = e^{-\mu} \frac{\mu^x}{x!} \quad (3 - 3)$$

En donde x número de ocurrencias cuya probabilidad se desea conocer
 μ número promedio de ocurrencias por unidad de tiempo o espacio.
 e base de los logaritmos naturales (2,71828...)

En la tabla B del anexo D, se encuentran el valor de probabilidad p , para los diferentes valores de x y μ .

3.2 DISTRIBUCIONES DE VARIABLE CONTINUA

Se denomina distribución de variable continua a aquella que resulta principalmente de la medición y puede tomar cualquier valor dentro de un intervalo dado.

3.2.1 Distribución normal: La distribución normal o también conocida como distribución de gauss¹⁸. Es la distribución de probabilidad más importante. Se utiliza para reflejar la distribución de variables tales como estaturas, pesos, distancias y otras medidas de la naturaleza que tienen observaciones numéricas que tienen a agruparse alrededor de la media y son divisibles infinitamente. En otras palabras, es más probable que una observación esté cerca de la medida del conjunto de datos que lejos. Tales variables continuas son el resultado de la medida. Existen varias formas de definir una distribución de probabilidad normal, la más visual es mediante su función de densidad, la cual está dada por:

¹⁸Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad (3-4)$$

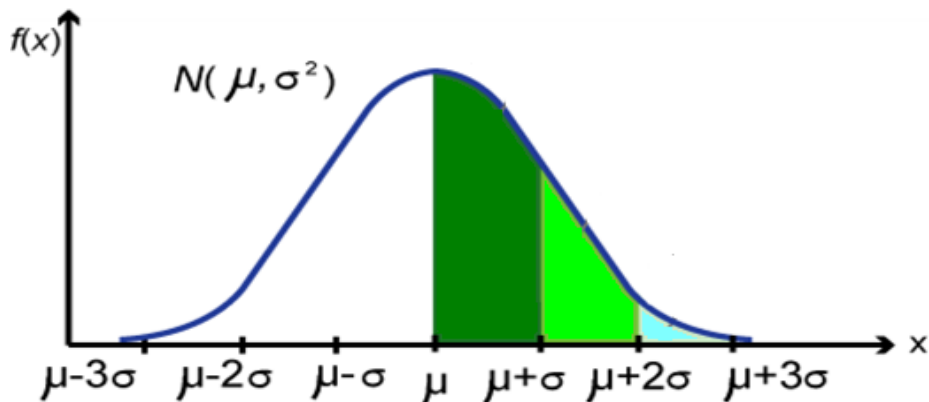
En donde x algún valor específico de la variable aleatoria

σ desviación estándar

μ media

La gráfica de la función de densidad tiene una forma de campana y es simétrica respecto de un determinado parámetro. Esta curva como se muestra en la figura 5 se conoce como campana de Gauss.

Figura 5. Distribución normal



La forma de la campana de Gauss depende de los parámetros μ y σ . La media indica la posición de la campana, de modo que para diferentes valores la gráfica es desplazada a lo largo del eje horizontal.

Por otra parte, la desviación estándar determina el grado de apuntamiento de la curva. Cuanto mayor sea el valor de σ , más se dispersarán los datos en torno a la media y la curva será más plana. Un valor pequeño de este parámetro indica, por tanto, una gran probabilidad de obtener datos cercanos

al valor medio de la distribución. La distribución normal y su curva tienen las siguientes características.

- La curva tiene forma de campana. La media, la moda y la mediana de la distribución son iguales y se localizan en el centro de la distribución.
- La escala horizontal de la curva se mide en desviaciones estándar.
- La distribución normal es simétrica alrededor de su media. Por lo tanto la mitad del área bajo la curva está antes del punto central y la otra mitad después, es decir la mitad de la curva tiene un área de 0,5.

Puede existir un número infinito de distribuciones normales posibles, cada una con su propia media y su desviación estándar. Ya que obviamente no se puede analizar un número tan grande de posibilidades, es necesario convertir todas estas distribuciones normales a una forma estándar. Esta conversión a la distribución normal estándar se efectúa con la fórmula de conversión (o fórmula z) que se representa de la siguiente forma:

$$Z = \frac{x - \mu}{\sigma} \quad (3 - 5)$$

En donde Z es la desviación normal y x es algún valor específico de la variable aleatoria. Después de este proceso de conversión, la media de la distribución es 0 y la desviación estándar es 1. Es decir, sin considerar lo que valen la media y la desviación estándar, se miden en las unidades originales en la distribución, después de que se ha aplicado la fórmula de conversión la media es 0 y la desviación estándar es 1.

En otras palabras el valor Z, es el número de desviaciones estándar¹⁹ a las que una observación está por encima o por debajo de la media. La tabla C del anexo E, proporcionará el área bajo la curva desde la media hasta algún valor por encima o por debajo de ésta.

3.2.2 Distribución t student. La distribución t student²⁰ es una distribución de probabilidad que estima la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño ($n < 30$). La distribución t de Student se utiliza cuando se cumple tres condiciones:(1) la muestra es pequeña, (2) σ es desconocida, y (3) la población es normal o casi normal.

La distribución t tiene una media de cero, es simétrica con respecto a la media y oscila entre menos infinito e infinito, la varianza de la distribución t es mayor que 1 y está dada por:

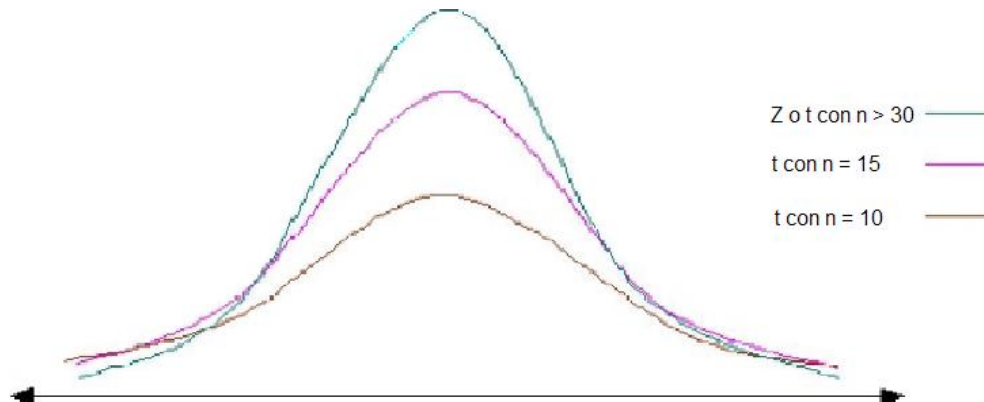
$$\sigma^2 = \frac{n - 1}{n - 3} (3 - 6)$$

La distribución t es una familia de distribuciones cada una con su propia varianza como se puede observar en la figura 6.

¹⁹ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

²⁰John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

Figura 6. Distribución t



La varianza depende de los $n-1$ grados de libertad, definidos como el número de observaciones que se pueden escoger libremente. Es el número de observaciones menos el número de restricciones impuestas sobre tales observaciones, en donde una restricción es algún valor que tales observaciones deben poseer. El estadístico t está dado por:

$$t = \frac{x - \mu}{S_x} \quad (3 - 7)$$

Los valores de la distribución t Student se encuentran tabulados en la tabla D, del anexo E al final del manual.

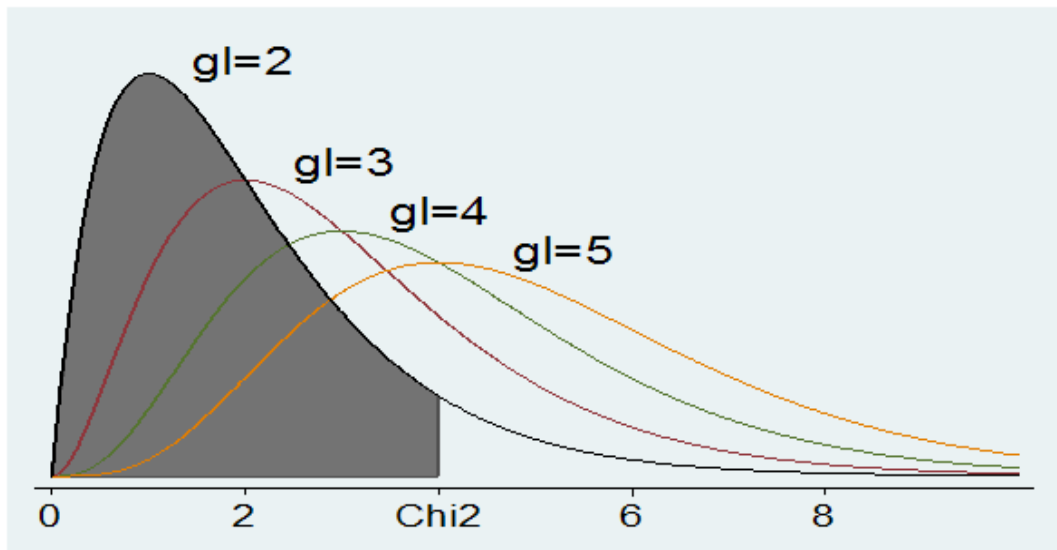
3.2.3 Distribución ji – cuadrado (x2). Al igual que la distribución t , la distribución ji-cuadrado es toda una familia de distribuciones²¹²¹. Existe una distribución ji-cuadrado para cada grado de libertad, como se puede observar en la figura 7. La figura muestra que a medida que se incrementa el número de grados de libertad ($k-1$), la distribución ji-cuadrado se vuelve menos sesgada. Las dos aplicaciones más comunes de ji-cuadrado son: (1) pruebas de bondad de ajuste y (2) pruebas

²¹ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

de independencia. Las cuales se verán en detalle en el capítulo de estadística inferencial.

La distribución ji-cuadrado se usa para comparar los resultados observados de los resultados esperados por una hipótesis y si la desviación obtenida no es significativa y puede atribuirse al azar o es significativa y otras variables diferentes al azar están influyendo en nuestros resultados.

Figura 7. Distribución ji – cuadrado



Fuente: www.monografias.com > Matemáticas > Estadística

La distribución ji – cuadrado se representa como:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (3 - 8)$$

En donde O_i Es la frecuencia de los eventos observados en los datos muestrales

k Es el número de categorías o clase

E_i Es la frecuencia de los eventos esperados

Los valores de la distribución ji – cuadrado se encuentran tabulados en la tabla F, del anexo E al final del manual.

3.2.4 Distribución F. La distribución F, es una distribución de probabilidad de gran aplicación en la estadística inferencial, especialmente en la comparación de la varianza de dos poblaciones normales, y, fundamentalmente en el análisis de varianza. La distribución F aparece frecuentemente como la distribución nula de una prueba estadística.

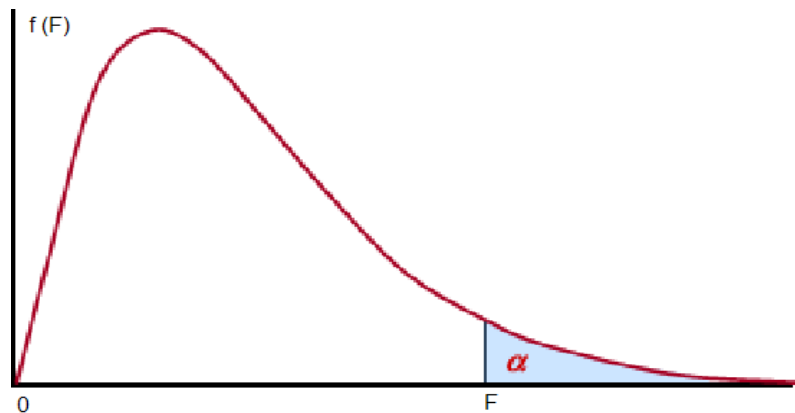
La variable aleatoria F se define como el cociente de dos variables aleatorias ji-cuadrado independiente²², cada una dividida entre sus respectivos grados de libertad. Esto es:

$$F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}} (3 - 9)$$

Donde U y V son variables aleatorias independientes que tienen distribución ji – cuadrado con $n_1 - 1$ y $n_2 - 1$ grados de libertad del numerador y del denominador. La variable aleatoria F es no negativa, y la distribución tiene un sesgo hacia la derecha, como se observa en la figura 8. Los valores de la distribución F se encuentran tabulados en la tabla E del anexo E al final del manual.

²²John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

Figura 8. Distribución F



Fuente: www.monografias.com › Matemáticas › Estadística

Si s_1^2 y s_2^2 son las varianzas muestrales independientes de tamaño n_1 y n_2 tomadas de poblaciones normales con varianzas σ_1^2 y σ_2^2 , respectivamente entonces:

$$F = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} = \frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2} \quad (3 - 9)$$

3.3 EJERCICIOS RESUELTOS

1. Ejercicio distribución binomial

Se dice que el 75% de los accidentes que ocurren en la planta de gas que opera la empresa Cumberger en Yopal se atribuyen a errores operacionales. Si en un periodo de tiempo dado, se suscitan 5 accidentes, determine la probabilidad de que: a) dos de los accidentes se atribuyan a errores operacionales, b) como máximo uno de los accidentes se atribuya a errores de tipo operacional, c) tres de los accidentes no se atribuyen a errores operacionales.

Solución

a) $n = 5$

$x = 2$

$p = 0.75$

$$p_2 = \frac{5!}{2! 5 - 2 !} 0.75^2 * (1 - 0.75)^{5-2}$$

$$p(2) = \frac{5 * 4 * 3 * 2 * 1}{2 * 1 3 * 2 * 1} (0.75^2) * (0.25)^3 = 0.08789$$

b) $n = 5$

$x = 0,1$

$p = 0.75$

$$p_x = p_0 + p_1$$

$$p_0 = \frac{5!}{0! 5 - 0 !} 0.75^0 * (1 - 0.75)^{5-0} = 0.000976$$

$$p_1 = \frac{5!}{1! 5 - 1 !} 0.75^1 * (1 - 0.75)^{5-1} = 0.01464$$

$$p_x = 0.000976 + 0.01464 = 0.01562$$

c) en este caso se cambiara el valor de p.

$n = 5$

$x = 3$

$p = 0.25$

$$p_3 = \frac{5!}{3! 5 - 3 !} 0.25^3 * (1 - 0.25)^{5-3} = 0.08789$$

2. Ejercicio distribución hipergeometrica

Una compañía de perforación y completamiento ha decidido revisar y verificar la calidad de la tubería que está implementando en sus operaciones (tubing y casing). Debido a que en los últimos 3 envíos la tubería ha presentado fallas, y por

consiguiente problemas en las operaciones, la compañía ha decidido seleccionar 5 tubos (Casing) al azar de un paquete de 15 para su revisión. Después de la revisión por parte de la empresa, se descubrió que el proveedor de la tubería había probado los 15 tubos (Casing) del paquete y que solo 10 estaban en buenas condiciones. El ingeniero de la compañía encargado de esta operación espera que no haya tubería defectuosa en este envío. ¿Cuál es la probabilidad de tener un envío con toda la tubería en buen estado?

Solución

$$N = 15$$

$$n = 5$$

$$r = 10$$

$$x = 5$$

$$P_x = \frac{C_{n-x}^{N-r} - C_x^r}{C_n^N} = \frac{C_{5-5}^{15-10} - C_5^{10}}{C_5^{15}} = \frac{\frac{5!}{0!5!} * \frac{10!}{5!5!}}{\frac{15!}{5!10!}} = \frac{1 * (252)}{3003} = 0.0839$$

Es evidente que existe una probabilidad muy pequeña (0.0839) de haber mandado un envío sin tubería defectuosa. La gerencia de la compañía debe localizar al proveedor para disponer un control del envío y reemplazar la tubería defectuosa.

3. Ejercicio distribución Poisson

La refinería kualampa de Arabia Saudita, recibe un promedio de 3.5 MBIs de petróleo por año de varios campos del oriente medio. Estas ocurrencias se producen al azar, es decir, no existe un patrón durante el año, o de un año a otro. ¿Cuál es la probabilidad de que se reciban exactamente 4 MBIs en un año?

Solución

$$\mu = 3.5$$

$$x = 4$$

$$p_x = e^{-\mu} \frac{\mu^x}{x!} = (2.71828)^{-3.5} \frac{3.5^4}{4!} = 0.189$$

4. Ejercicio distribución normal

TELHQE asociados es una compañía que presta servicios de instalación y monitoreo de bombas PCP (bombas de cavidades progresivas) a diferentes empresas petroleras en Colombia. Los funcionarios de TELHQE en su base de datos tienen un promedio de 150 bombas PCP por cada cliente, con una desviación estándar de 15 bombas. Las bombas parecen estar distribuidas normalmente.

Para estimar de manera apropiada la demanda del cliente por los servicios de la compañía y establecer una estructura de tarifas que maximice las utilidades corporativas. TELHQE debe determinar qué tan probable es que algunos servicios por parte de la compañía se presenten. El director de servicios desea que usted proporcione estimados de la probabilidad de que una empresa petrolera instale o monitoree:

- a) Entre 125 y 150 bombas
- b) Menos de 125 bombas
- c) Entre 145 y 155 bombas

Solución

$$\sigma = 15$$

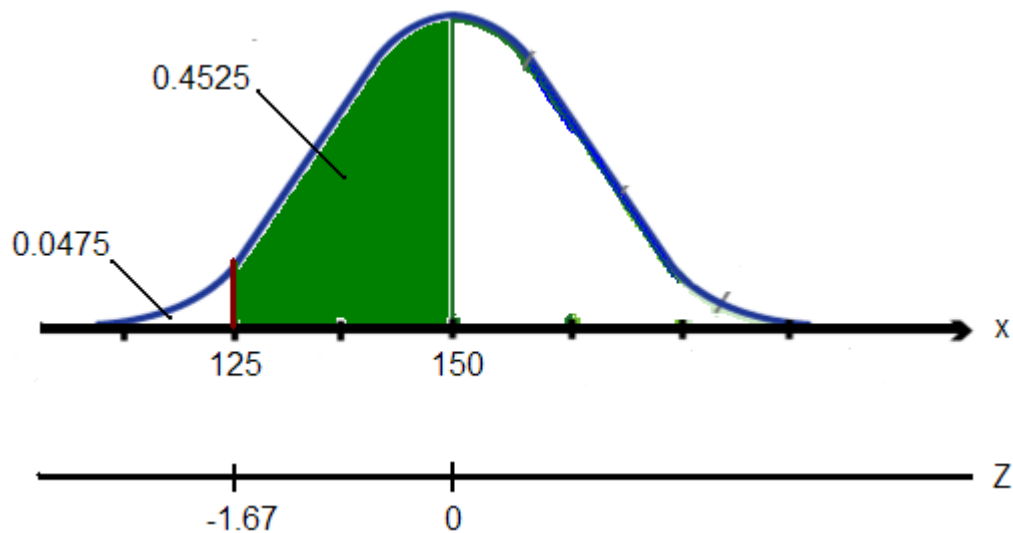
$$\mu = 150$$

a) Entre 125 y 150 bombas

$$Z = \frac{x - \mu}{\sigma} = \frac{125 - 150}{15} = -1.67$$

De la tabla C del anexo E con un valor Z de 1.67 da un área de 0.4525.

Figura 9. Distribución normal (Ejercicio 4 Numeral a)



Por tanto la probabilidad de que una empresa petrolera instale o monitoree entre 125 y 150 bombas es del 45.25%.

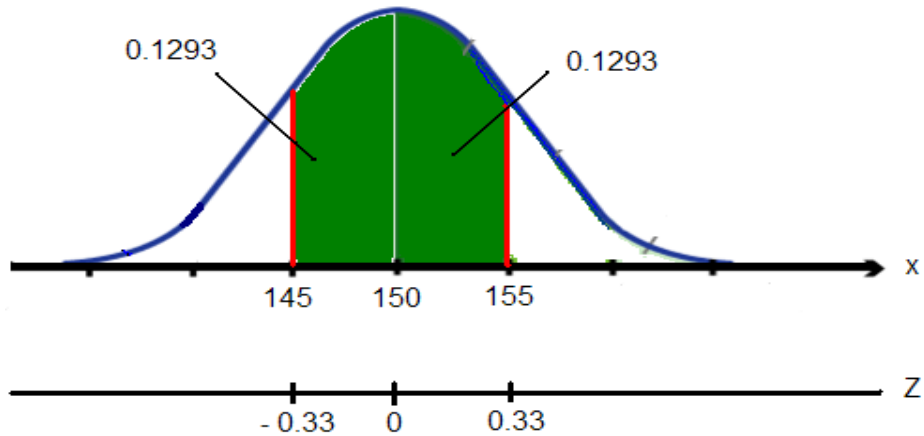
b) Menos de 125 bombas

Si el 45.25% del área está entre 125 y 150, entonces $0.5000 - 0.4525$, o el 4.75% de todas las empresas petroleras requieren menos de 125 bombas. La probabilidad de que cualquier empresa seleccionada aleatoriamente requiera 125 bombas o menos es del 4.75%.

c) Entre 145 y 155 bombas

$$Z = \frac{x - \mu}{\sigma} = \frac{145 - 150}{15} = -0.33$$

Figura 10. Distribución normal (Ejercicio 4 Numeral c)



Dado $Z = -0.33$, el área que está entre 145 y 150 es 0.1293. Debido a que 155 está a una distancia por encima de la media de 150 igual a la que 145 está por debajo de la media, el área entre 150 y 155 también es de 0.12293. Por tanto $p(145 \leq x \leq 155) = 0.1293 + 0.12293 = 0.2586$. Por tanto la probabilidad de que una empresa petrolera instale o monitoree entre 145 y 155 bombas es del 25.86%.

3.4 EJERCICIOS PROPUESTOS

1. Si la probabilidad de que el vapor de agua de un gas se condense en un tramo de la línea de transporte antes de llegar a la planta de deshidratación a 800 psi de presión es de 0.40, si se decide probar 12 tubos del tramo bajo esas condiciones, determine la probabilidad de que: a) el vapor de agua se condense en 4 de los tubos, b) en más de 2 tubos se condense el vapor, c) el vapor se condense en exactamente 5 tubos.

2. Una compañía manufacturera de productos para lodos de perforación utiliza un esquema para la aceptación de sus productos en las diferentes compañías de perforación. Se preparan 25 clases de lodos con productos diferentes y se selecciona una muestra de 3 lodos para verificar si los productos químicos utilizados por la compañía tienen un buen rendimiento en la formación y bajos costos. Si se encuentra que uno de los lodos no cumple con las necesidades requeridas, los productos ofrecidos por la compañía no se tendrán en cuenta y se regresaran para verificar el 100% de los productos. Si se encuentra un buen rendimiento y bajos costos, los 25 productos ofrecidos por la compañía serán comprados. a) ¿Cuál es la probabilidad de que los productos utilizados en 3 lodos tengan un bajo rendimiento?, b) ¿Cuál es la probabilidad de utilizar un producto con bajo rendimiento en el lodo y se devuelva la totalidad para su verificación?

3. El ingeniero Oscar Machuca, jefe de producción del campo Charlottesville de la compañía Winnsted, está preocupado por la producción de unos cuantos pozos del campo en los últimos días. Estos pozos han sido sometidos a periodos cortos de cierre en un promedio de 4 horas y un periodo de producción de 3 horas, los cuales han arrojado una disminución de la producción significativamente. Oscar ha decidido que si con un cierre de 5 horas o más, los pozos no tiene una producción mayor que 0.3, entonces estos pozos serán sometidos a un proceso de inyección de agua para aumentar su producción. ¿Deberá hacer esto?

4. Como ingeniero de lodos, usted es el encargado de la compra y utilización de bolsas de barita de un promedio de 50 libras, con una desviación estándar de 5.2 libras. Desde la última perforación usted se ha dado cuenta que el peso de cada bolsa ha aumentado significativamente, ya que lo ha visto reflejado en que produce una mayor cantidad de lodo. ¿Cuál es la probabilidad que las bolsas de barita tengan 60 libras?

4. ESTADISTICA INFERENCIAL

La estadística inferencial es la parte de la estadística que incluye los métodos utilizados para tomar decisiones o para obtener conclusiones sobre una característica desconocida de la población a partir de la información contenida en una o más muestras representativas de esa población. La herramienta teórica que utiliza es la distribución de la probabilidad a través del teorema de límite central.

4.1 DISTRIBUCION MUESTRAL

La distribución muestral resulta de considerar todas las muestras posibles que pueden ser tomadas de una población. Su estudio permite calcular la probabilidad que se tiene, dada una sola muestra, de acercarse al parámetro de la población. Mediante la distribución muestral se puede estimar el error para un tamaño de muestra dado.

4.1.1 Distribución muestral de medias. Cada muestra de tamaño n que se puede extraer de una población proporciona una media. Si se considera cada una de estas medias como valores de una variable aleatoria, se puede estudiar su distribución, que se llamará distribución muestral de medias.

4.1.1.1 La media de las medias muestrales: La distribución muestral de las medias muestrales es simplemente una lista de todas las medias muestrales posibles. Estas medias muestrales, al igual que cualquier lista de números, tienen una media denominada “la media de las medias muestrales” o la gran media. Esta medida de las medias se calcula de la forma usual: las observaciones individuales (medias muestrales) se suman y el resto se divide por el número de observaciones (muestras).

$$\bar{x} = \frac{\sum x}{k} (4 - 1)$$

En donde \bar{x} Es la media muestral

k Es el número de muestras en la distribución muestral

\bar{x} Es la media de las medias muestrales

La media de la distribución muestral siempre será igual a la media poblacional original.

$$\bar{x} = \mu (4 - 2)$$

4.1.1.2 La varianza y el error estándar de las medias muestrales: La distribución de las medias muestrales también tiene una varianza. La varianza en las medias muestrales es como cualquier otra varianza. Mide la dispersión²³ de las observaciones individuales (medias muestrales) alrededor de su media (la gran media). Además, esta varianza se calcula al igual que cualquier otra varianza. Es la Medel cuadrado de las desviaciones con respecto a su media. Se halla así:

- Determinando la cantidad por la cual cada una de las observaciones (medias muestrales) difiere de su media (la gran media)
- Elevando al cuadrado tales desviaciones.

²³Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

- Promediando las desviaciones al cuadrado y dividiendo por el número de medias muestrales K, se tiene:

$$\sigma_x^2 = \frac{(x - \bar{x})^2}{k} = \frac{(x - \mu)^2}{k} \quad (4 - 3)$$

Si se tuviera que sacar la raíz cuadrada de la varianza en la distribución de estas medias muestrales, se tendría el error estándar de la distribución muestral, por lo tanto:

$$\sigma_x = \sqrt{\sigma_x^2} \quad (4 - 4)$$

El error estándar de la distribución muestral²⁴ (o error estándar) es una medida de la dispersión de las medias muestrales alrededor de μ . Como se habrá notado la ecuación anterior requiere de mucha aritmética de tercer grado para calcular la varianza de la distribución muestral. Una aproximación cercana puede obtenerse mediante:

$$\sigma_x^2 = \frac{\sigma^2}{n} \quad (4 - 5)$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \quad (4 - 6)$$

Se requiere la suposición que la varianza poblacional σ^2 sea conocida. En donde la ecuación anterior es apropiada solo si el muestreo se realiza con reemplazo, o si la muestra se toma de una población muy grande. Si el muestreo se realiza sin

²⁴Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

reemplazo y si el tamaño de la muestra es más del 5% de la población²⁵, $n > 0.05N$, debe aplicarse el factor de corrección para poblaciones finitas (fpc). La ecuación apropiada para el error estándar entonces es:

$$\sigma_x = \frac{\sigma}{n} \sqrt{\frac{N-n}{N-1}} \quad (4-7)$$

En donde $\sqrt{\frac{N-n}{N-1}}$ es el factor de corrección para poblaciones finitas.

Si n es pequeña respecto a N (menos del 5%), el fpc se aproxima a 1 y por tanto es innecesario.

4.1.1.3 Teorema del límite central: El teorema del límite central hace referencia qué, para una población cualquiera, a medida que n aumenta, la distribución de las medias muestrales se aproxima a una distribución normal con una media:

$$\bar{x} = \mu \quad (4-8)$$

Y un error estándar

$$\sigma_x = \frac{\sigma}{n} \quad (4-9)$$

En donde σ_x Es el error Estándar

σ Es la desviación estándar

²⁵Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

N Es el tamaño de la población

n Es el tamaño de la muestra

Por tanto, incluso si la población no está distribuida normalmente, la distribución de muestreo de las medias muestrales será normal si n es lo suficientemente grande. La regla general es que si n es como mínimo 30, el teorema del límite central asegurará una distribución normal en las medias muestrales, incluso si la población no es normal. Si n es menor o igual a 30, la distribución que más se ajusta es la t student.

4.1.2 Distribución muestral de proporciones. El proceso de las proporciones muestrales es muy similar al de las medias. De cualquier población es posible obtener muchas muestras diferentes de un tamaño dado²⁶. En donde cada muestra tendrá su propia proporción de “éxitos” p ; sin embargo, al igual que con las medias, el valor esperado de la distribución muestral de las proporciones muestrales será igual a la proporción de éxitos en la población: $E(p) = P$. En donde la proporción²⁷ de “éxitos” p es igual a:

$$p = \frac{x}{n} = \frac{\text{número de éxitos}}{\text{número de ensayos}} \quad (4 - 10)$$

El valor esperado (media) de la distribución muestral de la proporción muestral es:

$$E p = \frac{p}{K} = P \quad (4 - 11)$$

²⁶ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

²⁷ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

Y el error estándar es

$$\sigma_p = \frac{\sqrt{P(1-P)}}{n} \quad (4 - 12)$$

De la misma manera que con las medias, si $n > 0.05N$, se requiere el fpc y el error estándar se vuelve

$$\sigma_p = \frac{\sqrt{P(1-P)}}{n} \sqrt{\frac{N-n}{N-1}} \quad (4 - 13)$$

Las herramientas recientemente desarrolladas para las proporciones muestrales permiten determinar las probabilidades que pueden ser muy útiles en la toma de decisiones importantes. Esto se logra aplicando la desviación normal a la distribución de proporciones muestrales:

$$Z = \frac{p - P}{\sigma_p} \quad (4 - 14)$$

4.1.3 Métodos de muestreo. Existen dos métodos para seleccionar muestras de poblaciones. El primero es el muestreo no aleatorio o de juicio, en el cual se emplea el conocimiento y la opinión personal para identificar aquellos elementos de la población que debe incluirse dentro de la muestra²⁸. El segundo es el muestreo aleatorio o de probabilidad, en el cual todos los elementos de la población tienen la probabilidad de ser escogidos para la muestra. Dentro de este

²⁸ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

tipo de muestreo se encuentran, el muestreo aleatorio simple, el muestreo sistemático, el muestreo estratificado y el muestreo por conglomerados.

Muestreo aleatorio simple. Es un método de selección de muestras que permite que cada muestra posible pueda ser elegida con la misma probabilidad. Por su parte cada elemento de la población tiene la misma oportunidad de ser incluido en la muestra.

Muestreo sistemático. Método en el cual los elementos que se muestrearán se seleccionan de la población de un intervalo uniforme que se mide con respecto al tiempo al orden o al espacio.

Muestreo estratificado. Método en el que la población se divide en grupos homogéneos, o estratos, y después se toma una aleatoria simple de cada estrato. Aquí la variabilidad dentro de cada grupo es pequeña y entre los grupos es grande.

Muestreo por conglomerados. Método en el que la población se divide en grupos o conglomerados de elementos, y luego se selecciona una muestra aleatoria de estos conglomerados. La variabilidad dentro de cada grupo es grande y entre los grupos es pequeña, es como si cada conglomerado fuese una pequeña representación de la población.

4.1.4 Tamaño de la muestra para medias usando muestreo aleatorio simple.

Para conocer el tamaño de la muestra es preciso comprender perfectamente el concepto de distribuciones muestrales²⁹. En donde tres factores afectan a la determinación del tamaño de la muestra para estimar la media de una población: 1) el nivel de confianza, 2) el error tolerable máximo y 3) la variación de la población.

²⁹ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

En el tamaño de la muestra para medias, si el estadístico de la muestra tiene distribución normal de acuerdo con el teorema del límite central, el tamaño de la muestra se puede estimar usando los siguientes cuatro pasos.

1. Determinar el nivel de confianza. Éste debe ser superior al 80% para que la estimación hecha con ese tamaño de muestra sea aceptable, pues un nivel de confianza menor llevaría a una muestra no representativa de la población. Por lo general se usan niveles de confianza entre el 90 y el 99%. Un nivel de confianza más alto requiere un tamaño de muestra mayor, que se acercaría al tamaño de la población, convirtiéndose en un censo; lo cual sería antieconómico y dispendioso, perdiéndose la filosofía del muestreo. Con el nivel de confianza utilizado se encuentra el valor Z en la distribución normal.
2. Determinar cuál es el error E o diferencia máxima permitida entre la media muestral y la media poblacional. La magnitud del error es subjetiva y dependerá de lo crítica que sea la exactitud de la estimación.
3. Si no se conoce la desviación estándar de la población³⁰, debe estimarse a partir de la desviación típica de muestras pasadas (*s*) o a partir de muestreos pilotos o preliminares.

3. Si la población es infinita debe utilizarse la siguiente ecuación:

4.

$$n = \frac{Z^2 \sigma^2}{E^2} = \frac{Z^2 \sigma^2}{(x - \mu)^2} \quad (4 - 15)$$

En donde Z valor de la tabla normal estándar que refleja el nivel de confianza

³⁰ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

σ es la desviación estándar poblacional

n es el tamaño de la muestra

E Error tolerable máximo (generalmente se utiliza un % de X , no mayor al 10% para no perder confiabilidad en el muestreo)

5. Si la población es finita debe utilizarse la siguiente ecuación:

$$n = \frac{NZ^2\sigma^2}{E^2 N - 1 + Z^2\sigma^2} \quad (4 - 16)$$

En donde N es el tamaño de la población.

4.1.5 Tamaño de la muestra para proporciones usando muestreo aleatorio simple. El procedimiento para determinar el tamaño de la muestra apropiado para estimar una proporción es similar al que se usa para las medias muestrales³¹. Los pasos 1 y 2 son en esencia los mismos. Los pasos 3 y 4 si cambian y son los siguientes:

1. Si no se conoce la proporción de la población (P), debe estimarse a partir de la proporción de muestras pasadas (p) o a partir de muestreos pilotos o preliminares.
2. Si la población es infinita debe utilizarse la siguiente ecuación:

$$n = \frac{Z^2 P(1 - P)}{E^2} \quad (4 - 17)$$

³¹Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

En donde Z valor de la tabla normal estándar que refleja el nivel de confianza

P proporción de la población

n es el tamaño de la muestra

E error porcentual tolerable máximo (en decimal)

El error tolerable máximo se calcula mediante

$$E = Z * \frac{\sqrt{P(1-P)}}{n} \quad (4-18)$$

O en su defecto utilizar un % subjetivo no mayor al 10%.

3. Si la población es finita debe utilizarse la siguiente ecuación:

$$n = \frac{NZ^2P(1-P)}{E^2(N-1) + Z^2P(1-P)} \quad (4-19)$$

En donde N es el tamaño de la población.

4.2 ESTIMACION POR INTERVALOS DE CONFIANZA

Una estimación por intervalos de confianza especifica el rango dentro del cual estará un parámetro³² desconocido. Tal intervalo con frecuencia va acompañado de una afirmación sobre el nivel de confianza que se da en su exactitud. En realidad hay tres niveles de confianza relacionados comúnmente con los intervalos de confianza: 99, 95 y 90%. Se podría calcular cualquier otro intervalo de

³²Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

confianza si se deseara, pero siempre superior al 80%. Pero estos tres niveles de confianza, denominados coeficientes de confianza, son simplemente convencionales.

Los intervalos de confianza tienen un valor alfa (α), que es la probabilidad de error tipo I, o la probabilidad de que un intervalo dado no contenga la media poblacional desconocida (1- coeficiente de confianza).

Las estimaciones por intervalo gozan de ciertas ventajas sobre las estimaciones puntuales (media o proporción muestral, por ejemplo). Debido al error de muestreo, probablemente la media muestral no será igual la media poblacional³³. Sin embargo no hay manera de saber que tan grande es el error de muestreo. Por tanto, los intervalos se utilizan para explicar esta discrepancia desconocida, el cual está dado por la diferencia entre uno de los límites del intervalo y la media muestral.

4.2.1 El fundamento de un intervalo de confianza. Un intervalo de confianza tiene un límite inferior de confianza (LI) y un límite superior de confianza (LS). Estos límites se hallan calculando primero el parámetro estadístico muestral (por ejemplo, la media o la proporción). Luego a este parámetro muestral se suma el error E o diferencia máxima permitida entre el parámetro muestral y el poblacional, para obtener el LS, y la misma cantidad se resta para obtener el LI.

4.2.2 Intervalo de confianza para la media poblacional - muestras grandes. Uno de los usos más comunes de los intervalos de confianza es estimar la media poblacional³³. Se debe recordar que el intervalo de confianza se forma utilizando la media muestral como una estimación puntual para el cual se adiciona y se resta

³³John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

un cierto valor para obtener los límites superior e inferior del intervalo de confianza.

Por lo tanto el intervalo de confianza para estimar μ cuando σ es conocido es:

$$I.C \text{ para estimar } \mu = x \pm Z\sigma_x \text{ (4-20)}$$

La ecuación anterior requiere la suposición que la desviación estándar σ es conocida. En el evento probable que σ sea desconocida, la desviación estándar debe substituirse:

$$I.C \text{ para estimar } \mu = x \pm ZS_x \text{ (4-21)}$$

En donde

$$S_x = \frac{s}{n} \text{ (4-22)}$$

El valor apropiado de Z puede hallarse en la tabla C del anexo E.

4.2.3 Intervalo de confianza para la media poblacional - muestras pequeñas.

En ocasiones no es posible obtener por lo menos 30 observaciones. Entonces es cuando debe tomarse una muestra pequeña. En este caso la distribución normal puede no aplicarse. El teorema del límite central asegura normalidad en el proceso de muestreo solo si la muestra es grande. Cuando se utiliza una muestra pequeña, puede ser necesaria una distribución alternativa, la distribución t Student. Específicamente, la distribución t se utiliza cuando se cumplen las tres condiciones: (1) la muestra es pequeña, (2) σ es desconocido y (3) la población es normal o casi normal.

El estadístico t se calcula como:

$$t = \frac{x - \mu}{S_x} \quad (4 - 23)$$

Reescribiendo algebraicamente para expresarlo como un intervalo de confianza para estimar μ , se tiene que:

$$I.C \text{ para estimar } \mu = x \pm t S_x = x \pm t \frac{S}{n} \quad (4 - 24)$$

El valor apropiado de t puede hallarse en la tabla D del anexo E, para un nivel de significancia α y un grado de libertad ($\gamma = n - 1$).

4.2.4 Intervalo de confianza para la proporción poblacional. Las decisiones dependen con frecuencia de parámetros que son binarios, parámetros con solo dos posibles categorías dentro de las cuales pueden clasificarse las respuestas. En este evento, el parámetro de interés es la proporción poblacional. La distribución de las proporciones muestrales será normal y la distribución muestral de la proporción muestral tendrá una media igual a la proporción poblacional P y un error estándar de P :

$$\sigma_p = \frac{\sqrt{P(1 - P)}}{n} \quad (4 - 25)$$

Sin embargo, como la ecuación anterior contiene P, el parámetro que se desea estimar, por tanto la proporción muestral p se utiliza como estimador de P. Entonces la ecuación anterior puede replantearse como:

$$S_p = \frac{\overline{p(1-p)}}{n} \quad (4 - 26)$$

El intervalo de confianza es entonces

$$I.C \text{ para estimar } P = p \pm ZS_p \quad (4 - 27)$$

El valor apropiado de Z puede hallarse en la tabla C del apéndice I.

4.2.5 Intervalo de confianza para la diferencia entre dos medias – muestras grandes. Aquí el interés está en estimar la diferencia de dos medias poblacionales ($\mu_1 - \mu_2$). La estimación puntual de la diferencia entre ($\mu_1 - \mu_2$) está dada por la diferencia entre las dos medias muestrales. Ya que muchas muestras diferentes pueden tomarse de cada población, resulta toda una distribución de diferencias de estas medias muestrales. Si tanto n_1 y n_2 son grandes, la distribución de las diferencias entre las medias muestrales es una distribución normal centrada en ($\mu_1 - \mu_2$).

Dada esta distribución normal de las diferencias entre las medias muestrales, la desviación normal Z puede utilizarse para construir el intervalo³⁴. El procedimiento es muy parecido al desarrollado para una sola población.

³⁴Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm Z\sigma_{x_1-x_2} \quad (4-28)$$

En donde el error estándar de las diferencias entre medias muestrales se obtiene de:

$$\sigma_{x_1-x_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (4-29)$$

De la misma manera que con cualquier error estándar, se mide la tendencia que tienen a variar las diferencias entre las medias muestrales; en donde σ_1^2 y σ_2^2 son las dos varianzas poblacionales. En el evento probable que σ_1^2 y σ_2^2 sean desconocidas, deben utilizarse las varianzas muestrales s_1^2 y s_2^2 . El estimado del error estándar se vuelve entonces:

$$S_{x_1-x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (4-30)$$

El intervalo de confianza cuando las varianzas poblacionales son desconocidas es:

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm ZS_{x_1-x_2} \quad (4-31)$$

4.2.6 Intervalo de confianza para la diferencia entre dos medias – muestras pequeñas. Si cualquier muestra es pequeña (menor que 30), no se puede asumir que la distribución de las diferencias en las medias muestrales se ajusta a una

distribución normal³⁵. Por tanto, se debe utilizar la desviación *t* Student, verificando qué:

- Las poblaciones están distribuidas normalmente o distribuidas casi normalmente.
- Las varianzas poblacionales son desconocidas.

Para aplicar adecuadamente la distribución *t*, también se debe determinar si las varianzas de las poblaciones son iguales. Se analizarán los intervalos de confianza cuando: (a) se tienen muestras independientes, asumiendo que las varianzas de las dos poblaciones son iguales, (b) se tienen muestras independientes, asumiendo que las varianzas sean diferentes, y (c) se tienen muestras dependientes.

a. Suposición de muestras independientes cuyas varianzas son iguales ($\sigma_1^2 = \sigma_2^2$) pero desconocidas: si las varianzas de las dos poblaciones son iguales, existe alguna varianza σ^2 común a ambas poblaciones. Es decir $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Sin embargo, debido al error de muestreo, se toma una muestra de cada población. Como las dos varianzas de las muestras probablemente difieran una de la otra, así como de la varianza común, los datos de ambas muestras pueden mancomunarse para obtener un solo estimado de S_D^2 . Esto se hace calculando el promedio ponderado de las dos varianzas de la muestra, en donde los pesos son los grados de libertad ($\gamma = n_1 + n_2 - 2$) para ambas muestras. Esta estimación mancomunada de la varianza poblacional común es:

³⁵ Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

$$S_D^2 = \frac{s_1^2 n_1 - 1 + s_2^2 n_2 - 1}{n_1 + n_2 - 2} \quad (4 - 32)$$

El intervalo de confianza para la diferencia entre las dos medias poblacionales³⁶ se halla entonces con una distribución t con un nivel de significancia (α) y $n_1 + n_2 - 2$ grados de libertad (γ).

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm t \sqrt{\frac{S_D^2}{n_1} + \frac{S_D^2}{n_2}} \quad (4 - 33)$$

b. Muestras independientes con varianzas desiguales: si las varianzas de las poblaciones son desiguales o no existe una evidencia para asumir una igualdad, los procedimientos descritos anteriormente no se aplican directamente porque la distribución de las diferencias entre las medias muestrales no se ajustan a una distribución t con $n_1 + n_2 - 2$ grados de libertad. De hecho, no se ha encontrado que una distribución exacta describa este proceso de muestreo, y solo se han desarrollado aproximaciones. Se ha propuesto una aproximación tal que utilice el estadístico t con grados de libertad levemente alterados. En el evento que σ_1^2 y σ_2^2 sean diferentes, se encuentra que los grados de libertad son:

$$g.l = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} \quad (4 - 34)$$

³⁶Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

Debido a que los grados de libertad se calculan de esta manera alterada, el estadístico t se simboliza con t' . El intervalo de confianza entonces será calculado como:

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm t' \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (4 - 35)$$

c. Muestras dependientes: Dos muestras se dicen que son dependientes, si el resultado o conocimiento de la una, afecta el resultado de la otra. Con mucha frecuencia ocurre que las observaciones de una variable se toman por parejas para las muestras. Cada observación de una muestra se asocia con alguna observación de la otra muestra; es decir, los dos elementos de una pareja de datos son dependientes, por lo que el conocer el uno, reduce el error de encontrar el otro. Cuando se toman muestra pequeñas dependientes, y como los elementos están por parejas, se tomarán entonces las muestras de un igual tamaño. El valor de la varianza poblacional estimada de las diferencias se calculará mediante la siguiente relación:

$$S_D^2 = \frac{D_i^2 - (D_i)^2/n}{n - 1} \quad (4 - 36)$$

El intervalo de confianza es entonces

$$I.C \text{ para } \mu_1 - \mu_2 = D \pm t \sqrt{\frac{S_D^2}{n}} \quad (4 - 37)$$

Donde $D: x_1 - x_2$

4.2.7 Intervalo de confianza para la diferencia entre proporciones. En ocasiones surgen situaciones en las cuales es necesario comparar las proporciones de dos poblaciones diferentes. En general, muchos asuntos de la industria del petróleo están determinados mediante la estimación de proporciones relativas de dos poblaciones, el procedimiento es similar al del intervalo de confianza para una proporción poblacional. En donde el error estándar de la diferencia entre dos proporciones muestrales ($p_1 - p_2$) se estima mediante:

$$S_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (4 - 38)$$

El termino $s_{p_1-p_2}$ reconoce que si se tomaran varios pares de muestras de cada población ($p_1 - p_2$) variaría. La ecuación anterior tiene en cuenta dicha variación, el intervalo de confianza es:

$$I.C \text{ para } P_1 - P_2 = p_1 - p_2 \pm ZS_{p_1-p_2} \quad (4 - 39)$$

El valor apropiado de Z puede hallarse en la tabla C del anexo E.

4.3 PRUEBAS DE HIPÓTESIS

La estadística inferencial se basa en información muestral, a partir de la cual se quieren estimar parámetros poblacionales; sin embargo, muchas veces se requiere comprobar el comportamiento supuesto o teórico de uno o varios parámetros poblacionales. Esta comprobación se hace mediante inferencia estadística a partir de datos muestrales. El procedimiento matemático se llama Prueba de Hipótesis, donde la suposición o afirmación es la hipótesis y la prueba se hace infiriendo el parámetro muestral a través de la teoría del límite central. La

aprobación o no de la hipótesis depende del respaldo probabilístico dado por el intervalo de confianza.

4.3.1 El concepto de prueba de hipótesis. Para realizar una prueba de hipótesis, se hace alguna inferencia, afirmación o supuesto con sentido acerca de una variable de la población. En donde una hipótesis nula (H_0) se contrasta con una hipótesis alternativa (H_A) que establece lo contrario. El término nula no necesariamente implica algo nulo; pero sí, esta hipótesis tradicionalmente está referenciada a uno de los siguientes signos: "=", " \leq " ó " \geq "; y la hipótesis alternativa al signo contrario, o sea: " \neq ", ">", ó "<", respectivamente. En el planteamiento de las hipótesis la afirmación o el supuesto que se quiere probar se ubica en la hipótesis que se ajuste al signo (H_0 o H_A).

La prueba consiste en establecer los límites de confianza de la zona de ocurrencia de la variable poblacional; es decir, el intervalo de valores en el cual la variable puede ocurrir o presentarse en un evento cualquiera al azar. Esta zona se llama "zona de aceptación de H_0 ". Si el signo de la hipótesis alternativa es " \neq ", se deben calcular los dos límites de confianza y se dice que la prueba es de dos colas. Si el signo de la hipótesis alternativa es "<", se debe calcular el límite inferior y se dice que la prueba es de cola inferior; y si es de signo ">", la prueba es de cola superior y se debe calcular el límite superior.

Con base en los datos muestrales se calcula el parámetro para la muestra, el cual se contrasta con la zona de ocurrencia o de aceptación de la hipótesis nula. Si el parámetro muestral cae dentro de la zona de aceptación, se dice que la hipótesis nula no es rechazada; en caso contrario, o sea si cae fuera de esta zona, es rechazada. Aclaro que nunca se puede "aceptar" la hipótesis nula como verdadera. El no rechazo de la hipótesis nula solamente significa que la evidencia muestral no es lo suficientemente fuerte como para llevar a su rechazo. Cuando se

realiza una prueba de hipótesis, la hipótesis nula se supone que es “verdadera” hasta que una preponderancia de la evidencia muestral indica que es falsa.

Cuando la diferencia entre el valor de la media poblacional bajo la hipótesis y el valor de la media muestral es lo suficientemente pequeña como para atribuirla a un error de muestreo, se dice que esta diferencia estadísticamente es insignificante.

Si la diferencia entre el valor de la media bajo la hipótesis y el hallado en la muestra es insuficiente para rechazar la hipótesis nula, entonces el asunto se convierte simplemente en un inferencial que me permite estimar qué tan grande debe ser la diferencia para que sea estadísticamente significativa y conduzca a un rechazo de la hipótesis nula. De la misma manera, la aceptación o el rechazo de la hipótesis nula puede medirse probabilísticamente. Esta probabilidad se conoce como el “respaldo probabilístico” de la decisión tomada. Si ese respaldo es menor del 80%, la decisión tomada no tiene el suficiente respaldo, lo que indica que con una muestra más grande, la decisión puede cambiar.

4.3.2 Tipos de errores. Al probar una hipótesis se pueden cometer dos tipos de errores. Un error tipo I es rechazar una hipótesis nula siendo verdadera. Es la probabilidad o posibilidad de rechazar, con el estadístico muestral, el parámetro supuesto poblacional, siendo verdadero. Esta probabilidad es igual al nivel de significancia o valor de α . Lo opuesto es el nivel de confianza o $1 - \alpha$, que indica la probabilidad de no rechazarla siendo verdadera.

El error tipo II es no rechazar una hipótesis nula que es falsa. Mientras que la probabilidad de un error tipo I es igual al valor α seleccionado, la probabilidad de un error tipo II, representado por la letra β , no se determina fácilmente. No se puede asumir que $\alpha + \beta = 1$.

Los niveles de significancia, o valores α , comúnmente seleccionados para pruebas de hipótesis son del 10%, 5% y 1%. Sin embargo, al igual que con los intervalos de confianza no hay nada especial o mágico sobre estos valores de α . La selección de un valor α depende del tipo de error, tipo I o tipo II. Vale la pena recordar que el valor de α de la prueba es la probabilidad de error tipo I. si se rechaza una hipótesis verdadera (error tipo I) es más serio que si se rechaza una hipótesis falsa (error tipo II). Se desearía seleccionar un valor de α bajo, como 1% o 5%, para minimizar la probabilidad de cometer un error tipo I. por otra parte, si no rechazar una hipótesis falsa (error tipo II) es más serio, en este caso es preferible un valor de α mas alto como 10%. Resumiendo, aunque los dos errores no son complementarios, el asumir un error tipo I alto, implícitamente implica considerar un error tipo II bajo, y viceversa.

4.3.3 Pasos para pruebas de hipótesis. Para realizar una prueba de hipótesis hay cuatro pasos involucrados:

Paso 1: plantear la hipótesis.

Paso 2: establecer los límites de confianza para el parámetro poblacional (zona de aceptación de H_0) utilizando el nivel de significancia α para calcular el valor del coeficiente de distribución Z o t . Se utilizan las fórmulas vistas en la sección anterior sobre intervalos de confianza, cambiando la variable muestral por la poblacional; o sea x por μ , o p por P .

Paso 3: se realiza la prueba de contraste comparando el estadístico muestral con el intervalo de confianza.

Paso 4: interpretación y conclusiones: Si el estadístico muestral (x o p) queda dentro de la zona de aceptación o intervalo de confianza, se acepta la H_0 , o no hay evidencia significativa para rechazarla. En caso contrario se rechaza la H_0 y se acepta la H_A .

4.3.4 Prueba de hipótesis para la media poblacional

4.3.4.1 Prueba de dos colas para la media poblacional – muestras grandes

a. Planteamiento de la hipótesis:

$H_0: \mu =$ al valor de la media poblacional supuesto que se quiere probar.

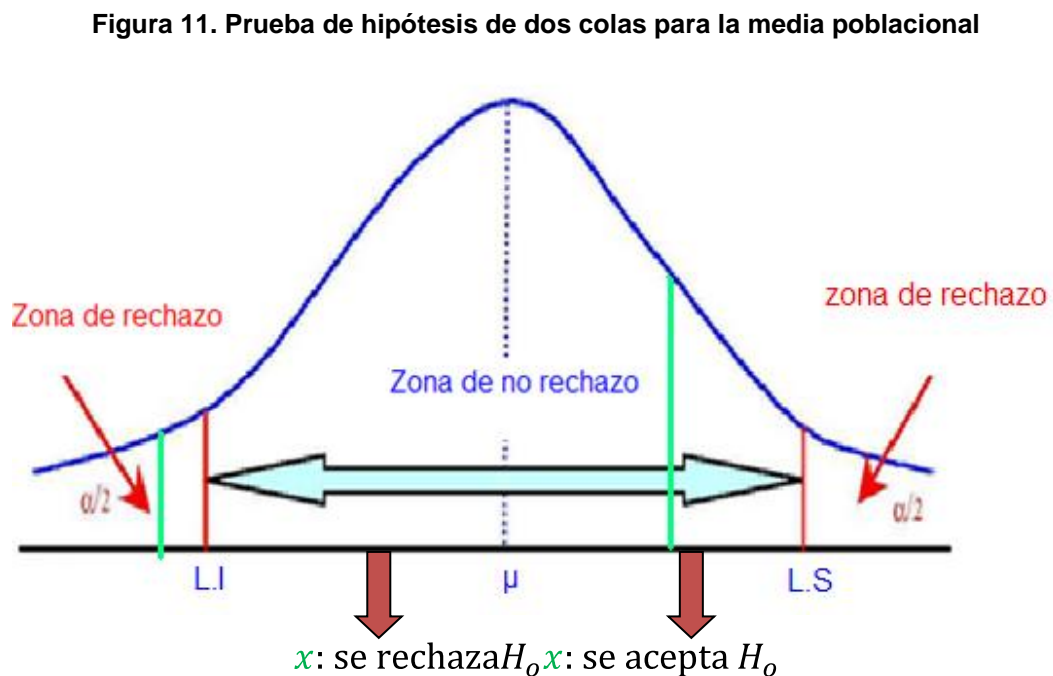
$H_A: \mu \neq$ al valor de la media poblacional supuesto que se quiere probar.

α : nivel de significancia con el cual se calcula el estadístico Z.

b. Cálculo de la zona de aceptación de H_0 :

$$L.C = \mu \pm Z \frac{S}{n} (4 - 40)$$

c. Prueba de contraste: se grafica la media muestral x en la campana de Gauss.



d. Interpretación: si la x queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si x queda fuera de la zona, se rechaza H_0 y se acepta H_A .

4.3.4.2 Prueba de una cola para la media poblacional: Las pruebas realizadas anteriormente eran de dos colas debido a que había zonas de rechazo en ambas, en donde la hipótesis alternativa no especifica dirección. Sin embargo hay ocasiones en las que se está interesado solo en un extremo u otro. En pruebas de una sola cola la hipótesis alternativa indica una dirección (ya sea mayor que o menor que). Vale la pena destacar que tanto la prueba de cola a la izquierda como en la prueba de cola a la derecha, el signo igual se coloca en la hipótesis nula. Esto porque la hipótesis nula se está probando a un valor α específico (como por ejemplo 5%) y el signo da a la hipótesis nula un valor específico para probarla.

4.3.4.3 Prueba de hipótesis para la media poblacional – muestras pequeñas: Cuando el tamaño de la muestra es pequeño (menor a 30) debe suponerse que la distribución de la población de la que se elige la muestra es aproximadamente normal. Si puede hacerse esta suposición y se conoce la desviación estándar poblacional, puede suponerse que la distribución muestral de las medias de la muestra es normal. Si no se conoce la desviación estándar de la población, la distribución muestral de las medias de la muestra se describe por una distribución t student.

a. Planteamiento de la hipótesis:

$H_0: \mu =$ al valor de la media poblacional supuesto que se quiere probar.

$H_A: \mu \neq$ al valor de la media poblacional supuesto que se quiere probar.

α : nivel de significancia con el cual se calcula el estadístico t.

b. Cálculo de la zona de aceptación de H_0 :

$$L.C = \mu \pm t S_x = \mu \pm t \frac{s}{n} \quad (4-41)$$

- c. Prueba de contraste: se grafica la media muestral \bar{x} en la campana de Gauss.
- d. Interpretación: si la \bar{x} queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si \bar{x} queda fuera de la zona, se rechaza H_0 y se acepta H_A .

Para muestras pequeñas en una prueba de dos colas, t se halla de manera normal de las tablas. Pero para una prueba de una cola se realiza una leve alteración. Cuando se utiliza la tabla t , se deben utilizar valores para pruebas de una cola.

4.3.5 Prueba de hipótesis para la proporción poblacional. El proceso de prueba de hipótesis para la proporción³⁷ poblacional P es muy similar al de μ tanto para pruebas de hipótesis de una o dos colas. La metodología de la prueba depende de si se usa una muestra pequeña o una muestra grande. Para muestras pequeñas ($n \leq 30$) se usa la distribución binomial. Para muestras grandes ($n > 30$) es aceptable la distribución normal si $nP > 5$ y $n(1 - P) > 5$. Como la mayor parte del análisis de proporciones maneja muestras grandes, solo se estudiará en este manual la prueba de hipótesis para este tipo de muestras.

a. Planteamiento de la hipótesis:

$H_0: P =$ al valor de la proporción poblacional supuesto que se quiere probar.

$H_A: P \neq$ al valor de la proporción poblacional supuesto que se quiere probar.

³⁷John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

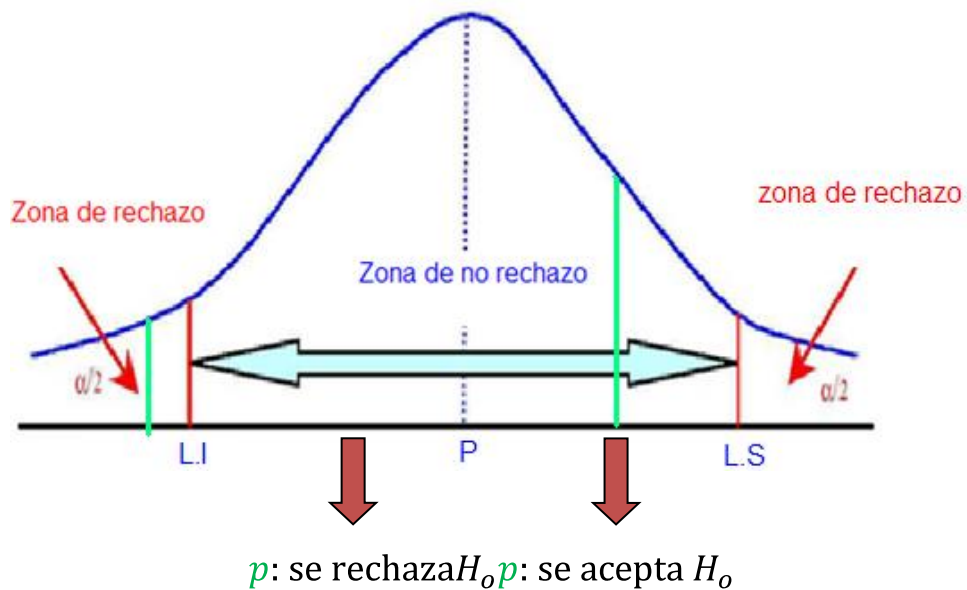
α : nivel de significancia con el cual se calcula el estadístico Z.

b. Cálculo de la zona de aceptación de H_0 :

$$L.C = P \pm Z \sqrt{\frac{P(1 - P)}{n}} \quad (4 - 42)$$

c. Prueba de contraste: se grafica la proporción muestral x en la campana de Gauss.

Figura 12. Prueba de hipótesis para la proporción poblacional



d. Interpretación: si la p queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si p queda fuera de la zona, se rechaza H_0 y se acepta H_A .

Al igual que con las pruebas para las medias, las pruebas de hipótesis de una sola cola para las proporciones pueden tener cola a la izquierda o la derecha. El valor p se puede calcular de forma similar que para las medias poblacionales.

4.3.6 Prueba de hipótesis para la diferencia entre dos medias. Cuando se prueba la igualdad de las medias de dos poblaciones, la hipótesis nula afirma que las medias de las dos poblaciones son iguales. Esta hipótesis no se descartará a menos que se reúna suficiente evidencia para rechazarla con una probabilidad de error pequeña. La hipótesis alternativa³⁸ puede usar el signo desigual, mayor que o menor que. Esta elección depende de la afirmación que se haga sobre las poblaciones de interés. Para una prueba de dos colas, las hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

O el equivalente

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

Para una prueba de una cola, las hipótesis nula y alternativa se establecen como uno de los dos grupos siguientes:

$$H_0: \mu_1 - \mu_2 \leq 0 \quad \text{o} \quad H_0: \mu_1 - \mu_2 \geq 0$$

$$H_A: \mu_1 - \mu_2 > 0 \quad \text{o} \quad H_A: \mu_1 - \mu_2 < 0$$

³⁸John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

4.3.6.1 Pruebas con muestras grandes: La prueba de hipótesis para la diferencia entre medias de dos poblaciones es muy parecida a la de una sola población. Se incluyen los mismos 4 pasos:

a. Planteamiento de la hipótesis:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

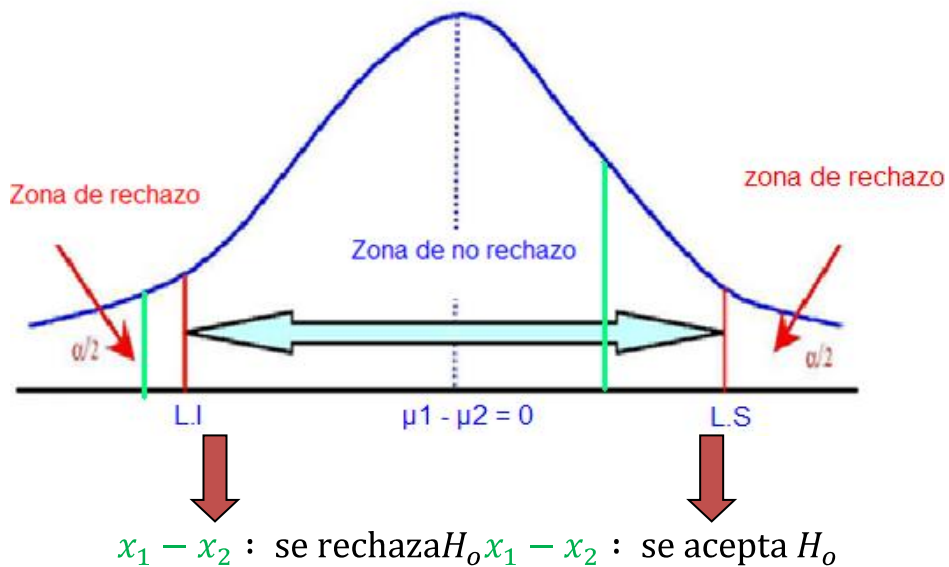
α : nivel de significancia con el cual se calcula el estadístico Z.

b. Cálculo de la zona de aceptación de H_0 :

$$L.C = \mu_1 - \mu_2 \pm Z \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (4 - 43)$$

c. Prueba de contraste: se grafica la diferencia entre dos medias muestrales $x_1 - x_2$ en la campana de Gauss.

Figura 13. Prueba de hipótesis con muestras grandes



d. Interpretación: si la diferencia $x_1 - x_2$ queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si $x_1 - x_2$ queda fuera de la zona, se rechaza H_0 y se acepta H_A .

4.3.6.2 Pruebas con muestras pequeñas: La prueba es igual que la de las medias de dos poblaciones para muestras grandes, aunque se usa la distribución t en lugar de la curva normal. La prueba t se considera apropiada cuando cualquier tamaño de muestra es menor que 30.

a. Planteamiento de la hipótesis:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

α : nivel de significancia con el cual se calcula el estadístico t.

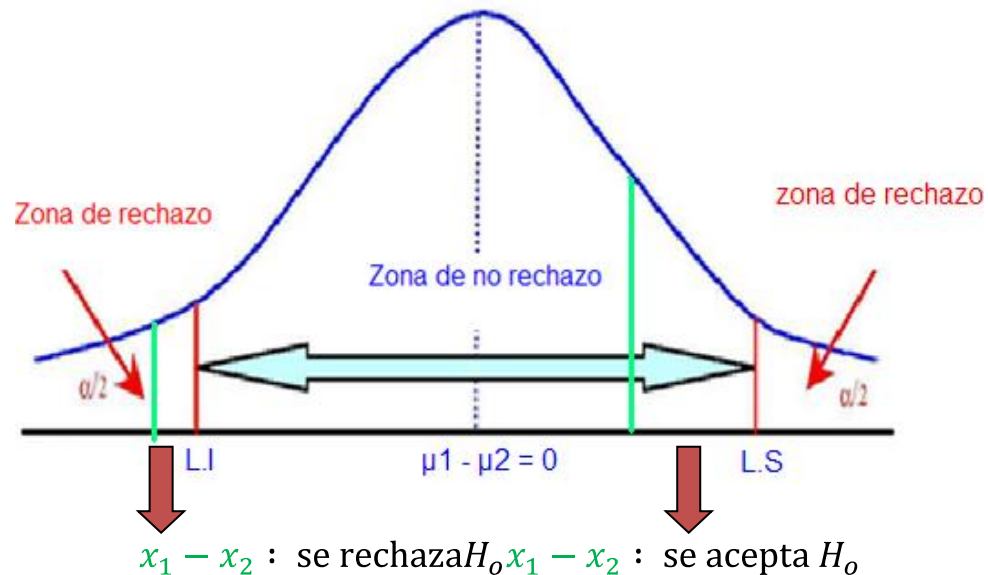
b. Cálculo de la zona de aceptación de H_0 :

$$L.C = \mu_1 - \mu_2 \pm t \sqrt{\frac{S_D^2}{n_1} + \frac{S_D^2}{n_2}} \quad (4 - 44)$$

$$\text{Donde } S_D^2 = \frac{s_1^2 n_1 - 1 + s_2^2 n_2 - 1}{n_1 + n_2 - 2}$$

c. Prueba de contraste: se grafica la diferencia entre dos medias muestrales $x_1 - x_2$ en la campana de Gauss.

Figura 14. Prueba de hipótesis con muestras pequeñas



d. Interpretación: si la diferencia $x_1 - x_2$ queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si $x_1 - x_2$ queda fuera de la zona, se rechaza H_0 y se acepta H_A .

4.3.7 Prueba de hipótesis para la diferencia entre dos proporciones. Para probar la diferencia entre dos proporciones poblacionales se toma una muestra aleatoria de cada población y se calcula la proporción muestral de cada una. La diferencia entre las proporciones muestrales es el estadístico de prueba, su distribución muestral es la distribución normal. La hipótesis nula que se quiere probar es que las dos poblaciones tienen la misma proporción. Esta hipótesis no se rechaza a menos que se reúna suficiente evidencia para rechazarla con una probabilidad pequeña de error. Para la prueba de dos colas la hipótesis nula y alternativa son:

$$H_0: P_1 = P_2$$

$$H_A: P_1 \neq P_2$$

O el equivalente

$$H_0: P_1 - P_2 = 0$$

$$H_A: P_1 - P_2 \neq 0$$

Para una prueba de una cola, las hipótesis nula y alternativa se establecen como uno de los grupos siguientes:

$$H_0: P_1 - P_2 \leq 0 \quad \text{o} \quad H_0: P_1 - P_2 \geq 0$$

$$H_A: P_1 - P_2 > 0 \quad \text{o} \quad H_A: P_1 - P_2 < 0$$

a. Planteamiento de la hipótesis:

$$H_0: P_1 - P_2 = 0$$

$$H_A: P_1 - P_2 \neq 0$$

α : nivel de significancia con el cual se calcula el estadístico Z.

b. Cálculo de la zona de aceptación de H_0 :

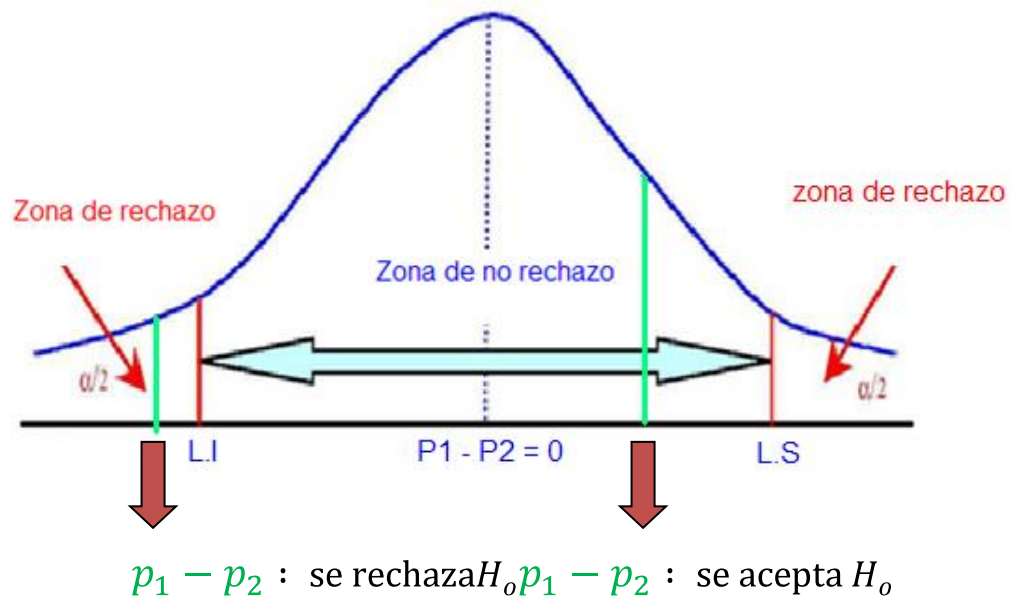
$$L.C = P_1 - P_2 \pm Z \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}} \quad (4 - 45)$$

En la prueba para proporciones con dos muestras, la estimación de la proporción poblacional se calcula con el promedio de las proporciones de las dos muestras³⁹.

$$P = \frac{x_1 + x_2}{n_1 + n_2} = P_1 = P_2(4 - 46)$$

c. Prueba de contraste: se grafica la diferencia entre dos medias muestrales $p_1 - p_2$ en la campana de Gauss.

Figura 15. Prueba de hipótesis para la diferencia entre dos proporciones



d. Interpretación: si la diferencia $p_1 - p_2$ queda dentro de los límites de confianza, se acepta la H_0 ; o sea que no existen evidencias significativas para rechazarla. En caso contrario, si $p_1 - p_2$ queda fuera de la zona, se rechaza H_0 y se acepta H_A .

³⁹John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

4.3.8 Prueba de Bondad de Ajuste. Con frecuencia, las decisiones requieren que se pruebe alguna hipótesis sobre la distribución poblacional desconocida. La prueba de bondad de ajustese utiliza para determinar si la distribución de los valores en la población se ajusta a una forma en particular planteada como hipótesis, en este caso una distribución uniforme. De la misma manera que con todas las pruebas estadísticas de esta naturaleza, los datos muestrales se toman de la población y estos constituyen la base de los hallazgos. Las hipótesis que se probarían son:

H_0 = La distribución poblacional es uniforme

H_A = La distribución poblacional no es uniforme

Si existe gran diferencia entre lo que realmente se observa en la muestra y lo que se esperaría observar si la hipótesis nula fuera verdadera, en tal caso es menos probable que la hipótesis nula sea verdadera. Es decir, la hipótesis nula debe rechazarse cuando las observaciones obtenidas en la muestra difieren mucho del patrón que se espera que ocurra si la distribución planteada como hipótesis si se presenta.

Para contrastar la hipótesis relativa a una distribución poblacional⁴⁰, se debe analizar la diferencia entre las expectativas con base en la distribución planteada como hipótesis y los datos reales que aparecen en la muestra. Esto es precisamente lo que hace la prueba chi-cuadrado de bondad de ajuste. Determina si las observaciones muestrales “se ajustan” a las expectativas. La prueba toma la siguiente forma:

⁴⁰ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (4 - 47)$$

En donde O_i Es la frecuencia de los eventos observados en los datos muestrales

k Es el número de categorías o clase

E_i Es la frecuencia de los eventos esperados si la hipótesis nula es correcta

La prueba tiene $K - m - 1$ grados de libertad, en donde “m” es el número de parámetros a estimar. El impacto exacto de “m” será más evidente a medida que la discusión progrese. Vale la pena anotar que el numerador de la ecuación anterior mide la diferencia entre las frecuencias de los eventos observados y la frecuencia de los eventos esperados al cuadrado. Cuando esas diferencias son grandes, haciendo que chi-cuadrado se incremente, debería rechazarse la hipótesis nula. El valor χ^2 debe compararse con un valor crítico de χ^2 tomado de la tabla.

Sin embargo surgen muchos casos en los cuales las frecuencias se prueban contra un parámetro determinado, en el cual las frecuencias esperadas no son todas iguales. En su lugar debe determinarse:

$$E_i = np_i \quad (4 - 48)$$

En donde n es el tamaño de la muestra.

p_i es la probabilidad de cada categoría como se especificó en la hipótesis nula.

4.3.9 Prueba de independencia. La prueba de independencia, permite la comparación de dos atributos para determinar si existe una relación entre ellos. En donde las hipótesis que se probarían son:

H_0 = Los atributos son independientes

H_A = Los atributos no son independientes

Los grados de libertad de la prueba están dados por el número de filas menos uno ($f - 1$) multiplicado por el número de columnas menos uno ($c - 1$). Probar las hipótesis requiere de una comparación de E_i y O_i , sobre las fc (celdas), utilizando la ecuación:

$$\chi^2 = \sum_{i=1}^{rc} \frac{(O_i - E_i)^2}{E_i} \quad (4 - 49)$$

En donde O_i es la frecuencia de los eventos observados en los datos muestrales
 rc es el número de categorías o clase

E_i es la frecuencia de los eventos esperados si la hipótesis nula es correcta

4.3.10 Comparación de la varianza de dos poblaciones normales. Cuando se comparan las varianzas de dos poblaciones, se toma una muestra de cada población. Las varianzas de la muestra sirven como estimados de sus varianzas poblacionales respectivas. Una distribución F se forma por la razón de estas dos varianzas muestrales. La razón F es

$$F = \frac{S_l^2}{S_s^2} \quad (4 - 50)$$

En donde el numerador es la más grande de las dos varianzas muestrales y el denominador es la más pequeña de las dos varianzas muestrales. Se asume que la varianza de la segunda muestra excede la de la primera. En ese caso, la razón F es:

$$F = \frac{s_2^2}{s_1^2} (4 - 51)$$

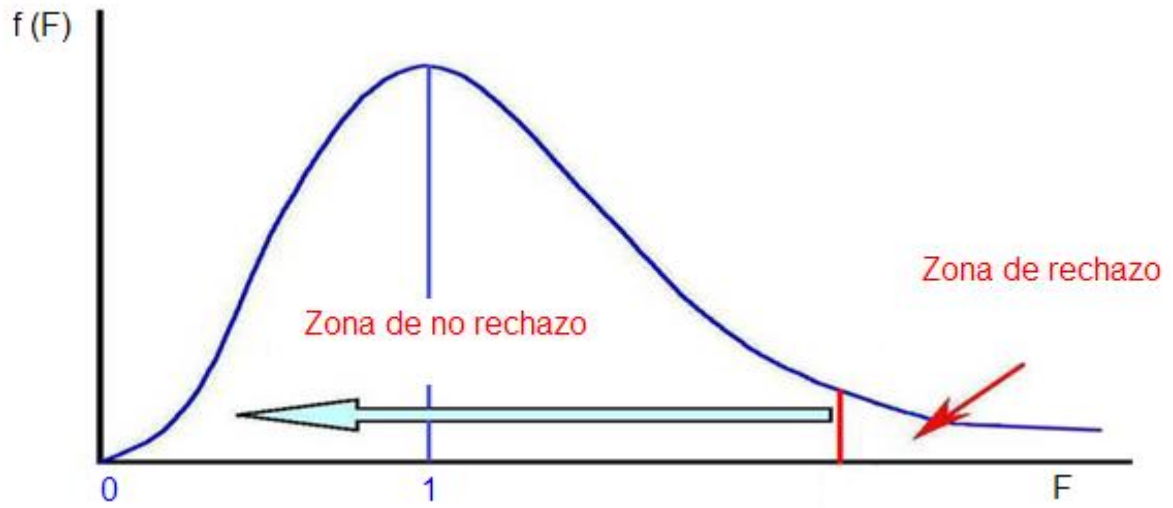
Entre más exceda s_2^2 a s_1^2 menos probable es que $\sigma_1^2 = \sigma_2^2$ y más grande será la razón F. Por tanto, un valor de F grande conllevará a un rechazo de la hipótesis nula

$$H_o: \sigma_1^2 = \sigma_2^2$$

$$H_A: \sigma_1^2 \neq \sigma_2^2$$

Manipular la razón F para garantizar que la varianza de la muestra más alta se coloque en el numerador asegurará que el valor F siempre sea mayor que 1. Como se muestra en la figura 9, la distribución F no es simétrica y está limitada por cero en el extremo inferior. Debido a que la razón F está restringida para exceder siempre de 1, resulta solo una zona de rechazo en la cola derecha. Está prohibida toda zona de rechazo que hubiera aparecido de otro modo en la cola izquierda. Por tanto, solo la mitad del área bajo la distribución es accesible como zona de rechazo y es necesario dividir por 2 el valor de α seleccionado e identificar una zona de rechazo única en la zona derecha, con un área igual a $\alpha/2$.

Figura 16. Distribución F



La distribución F tiene dos grados de libertad⁴¹: uno asociado al numerador, el cual es igual a $n_1 - 1$ y otro grado para el denominador, el cual es igual a $n_2 - 1$, en donde n_1 y n_2 son los tamaños de las muestras en el numerador y en el denominador, respectivamente. Donde los valores apropiados de F se hallan en la tabla E del apéndice I.

4.3.11 Análisis de varianza – ANOVA. Muchas decisiones en la industria petrolera requieren de la comparación de tres o más de dos poblaciones; es aquí donde el análisis de varianza (ANOVA)⁴² es de gran utilidad. ANOVA está diseñado específicamente para probar si tres o más poblaciones tienen la misma media. Aun cuando el propósito de ANOVA es hacer pruebas para hallar diferencias en las medias poblacionales, implica un examen de las varianzas muestrales; de allí el término análisis de varianza. Más específicamente, el procedimiento se puede utilizar para determinar si cuando se aplica un

⁴¹ Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.

⁴² John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

“tratamiento” en particular a una población, este tendrá un impacto significativo en su media.

La forma como se seleccionan los tratamientos determina si se está utilizando un modelo de efectos fijos o un modelo de efectos aleatorios. El modelo de efectos fijos es donde se seleccionan tratamientos específicos o se fijan antes del estudio, y el modelo de efectos aleatorios es donde los niveles (tratamientos) utilizados en el estudio se seleccionan aleatoriamente de una población de niveles posibles. Para la aplicación de ANOVA son esenciales tres suposiciones.

- Todas las poblaciones involucradas son normales.
- Todas las poblaciones tiene la misma varianza.
- Las muestras se seleccionan independientemente.

Si un número de tratamientos se designa como c , el conjunto de hipótesis de prueba es

$$H_0: \mu_1 = \mu_2 = \mu_3 \dots = \mu_n$$

$$H_A: \text{No todas las medias son iguales}$$

La letra c se utiliza para el número de tratamientos debido a que en una tabla de ANOVA, cada tratamiento se especifica en su propia columna.

4.3.11.1 Fundamentos del ANOVA: Para determinar si tratamientos diferentes tienen efectos diferentes en sus respectivas poblaciones, se realiza una comparación de la variación *dentro* de las muestras y la variación *entre* muestras⁴³. El tratamiento en sí mismo no producirá ninguna variación en las

⁴³John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

observaciones *dentro* de una muestra, debido a que todas las observaciones en dicha muestra reciben el mismo tratamiento.

Es un asunto diferente con la variación *entre* muestras. La variación *entre* muestras (de una muestra a la siguiente) puede producirse por el mismo efecto que produjo la variación *dentro* de una muestra, mas toda influencia adicional que puede tener los tratamientos diferentes. Puede existir un efecto del tratamiento *entre* muestras debido a que cada muestra tiene un tratamiento diferente.

Si un efecto del tratamiento *entre* muestras existe, puede detectarse comparando la variación *entre* las muestras y la variación *dentro* de las muestras. Si la variación *entre* las muestras es mayor que la variación *dentro* de las muestras, un fuerte efecto de tratamiento está presente *entre* las muestras que las hace diferentes. Esta diferencia de la variación *entre* muestras y la variación *dentro* de las muestras es precisamente lo que mide el análisis de varianza mediante un cociente o relación entre ambas variaciones, siendo la varianza *entre* muestras el numerador. Si los tratamientos en las diferentes muestras tienen efectos diferentes, la variación *entre* muestras crecerá, haciendo que la razón aumente. Esta razón se basa en la distribución F.

Hablando de muestras de diferentes poblaciones, cuando las medias poblacionales son diferentes, el efecto del tratamiento está presente y la desviación *entre* las muestras será grande comparada con la desviación *dentro* de una muestra; por tanto, el valor F aumentará.

4.3.11.2 Cálculo de la varianza *ENTRE* las medias de las muestras: El paso 1 en el análisis de varianza indica que se debe obtener una estimación de la varianza de la población a partir de la varianza *entre* las medias de las muestras. La estimación de la varianza común de la población produce una estimación

válidas sólo si la hipótesis nula es cierta⁴⁴. Para entenderlo mejor se debe recordar el teorema del límite central, el cual establece que la distribución de las medias muestrales tiende a una distribución normal conforme crece el tamaño de la muestra, con una media μ y una desviación estándar σ/\sqrt{n} . Si el error estándar de la media es σ/\sqrt{n} , entonces la varianza de la distribución es igual al error estándar al cuadrado (σ^2/n).

Esta varianza es una medida de las diferencias entre todas las medias muestrales que puedan obtenerse de la distribución y la media poblacional. En ANOVA, para estimar la varianza de la distribución muestral de medias, se debe estimar primero la media poblacional. La media de todos los valores muestrales proporciona esa estimación. Después se determina la diferencia entre la media de cada grupo y esta media poblacional estimada, y estas diferencias se elevan al cuadrado y se suman. Esta suma se divide entre el número adecuado de grados de libertad para obtener la estimación de la varianza poblacional de la distribución *entre* medias de las muestras.

$$S_x^2 = \frac{\sum_j (x_j - \bar{x})^2}{k - 1} \quad (4 - 52)$$

En donde S_x^2 estimación de la varianza de la distribución *entre* las medias

x_j media del grupo j

\bar{x} gran media (media de todos los valores), usada como estimación de μ

k número de grupos (número de muestras)

⁴⁴John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

Para la distribución muestral *entre* medias partimos de la ecuación del error estándar de las medias al cuadrado $\sigma_x^2 = \sigma^2/n$, en donde “n” es el tamaño de la muestra, o el número de elementos de cada grupo. Al despejar de ésta la varianza (σ^2) se obtiene:

$$\sigma^2 = n\sigma_x^2$$

Análogamente, por tanto, se puede calcular una estimación de S^2 si se multiplica “n” por la estimación de la varianza de la distribución muestral de medias S_x^2 (ecuación 4 – 52), es decir:

$$S^2 = nS_x^2$$

Así, la estimación de la varianza de la distribución muestral *entre* medias S_b^2 queda:

$$S_b^2 = \frac{n \sum_j (x_j - x)^2}{k - 1} \quad (4 - 53)$$

En donde S_b^2 estimación de la varianza poblacional a partir de la varianza *entre* las medias de las muestras.

x_j media del grupo j

x gran media (media de todos los valores), usada como estimación de μ

k número de grupos (número de muestras)

n número de elementos de la muestra en cada grupo si el elemento de observaciones en cada uno es el mismo.

El valor adecuado de los grados de libertad para el cálculo de la Varianza entre las Medias de las Muestras es $(K - 1)$.

4.3.11.3 Cálculo de la varianza *DENTRO* de las muestras: El paso 2 en ANOVA requiere una segunda estimación de la varianza de la población, basada en la varianza *dentro* de las muestra. La estimación de la varianza produce una estimación válida sin importar si la hipótesis nula de medias poblacionales es cierta. Esto se debe a que la variabilidad de los valores de la muestra se determina comparando cada elemento en los datos con la media muestral. Cada valor de la muestra obtenido de la población A se compara con la media muestral A; cada elemento de la población B se compara con la media muestral B, y así sucesivamente. La siguiente ecuación se usa para calcular la varianza *dentro* de las muestras.

$$S_w^2 = \frac{\sum_j \sum_i (x_{ij} - x_j)^2}{k(n - 1)} \quad (4 - 54)$$

En donde S_w^2 estimación de la varianza poblacional a partir de la varianza *dentro* de las muestras.

x_{ij} i – esimo elemento de los datos en el grupo j

x_j media del grupo j

k número de grupos (número de muestras)

n número de elementos de la muestra en cada grupo

Primero se encuentran las diferencias entre cada valor x y la media de su grupo, se elevan al cuadrado y se suman. Después, se agregan estas sumas para cada grupo. El resultado es la suma del cuadrado de las desviaciones entre cada medida de cada muestra y la media de su grupo. Esta suma se divide después entre el número adecuado de grados de libertad para producir una estimación de la varianza desconocida de la población.

4.3.11.4 Cálculo e interpretación de la estadística F: En el paso 3 de la ANOVA se comparan las dos estimaciones de la varianza de la población mediante el cálculo de su cociente en la siguiente forma:

$$F = \frac{S_b^2 \text{ (estimación de } \sigma^2 \text{ por el método entre)}}{S_w^2 \text{ (estimación de } \sigma^2 \text{ por el método dentro)}} \quad (4 - 55)$$

Si la hipótesis nula es cierta, tanto el numerador como el denominador de la ecuación (4 – 55) son estimaciones válidas de la varianza común de las poblaciones que se estudian. Este cociente, entonces, se ajusta a la distribución F^{45} . Sin embargo, si la hipótesis nula es falsa, el numerador de la ecuación (4 – 55) en realidad es una estimación inflada de la σ^2 , aunque el denominador siga siendo una estimación válida. Bajo estas condiciones, el valor F será muy grande, y se puede concluir que la hipótesis nula es falsa.

4.4 EJERCICIOS RESUELTOS

1. Ejercicio distribución muestral de medias

Las ventas de crudo en miles de dólares para East Coast Petroleum Company durante los últimos 5 meses fueron de 68, 73, 65, 80 y 72. Asumiendo que estos 5 meses constituyen la población, la media es $\mu = 71.6$. Como director del departamento comercial se desea estimar este μ “desconocido” tomando una

muestra de tamaño $n = 3$. Se espera que el error de muestreo que es probable que ocurra sea relativamente pequeño. Realice la distribución muestral y haga comentarios sobre el posible error de muestreo.

Solución

Hay ${}_5C_3 = 10$ muestras en la distribución muestral

Tabla 19. Ventas de la compañía (Ejercicio 1)

Númerode la muestra	Elementos de la muestra	Media muestral \bar{x}
1	68, 73, 65	68.67
2	68, 73, 80	73.67
3	68, 73, 72	71.00
4	68, 65, 80	71.00
5	68, 65, 72	68.33
6	68, 80, 72	73.33
7	73, 65, 80	72.67
8	73, 65, 72	70.00
9	73, 80, 72	75.00
10	65, 80, 72	72.33

La media de la distribución muestral es:

$$\bar{x} = \frac{\sum x}{k} = \frac{68.67 + 73.67 + 71.0 + \dots + 72.33}{10} = 71.6 = \mu$$

La varianza y el error estándar se la distribución muestral son:

$$\sigma_x^2 = \frac{(x - \mu)^2}{k}$$

$$= \frac{(68.67 - 71.6)^2 + (73.67 - 71.6)^2 + \dots + (72.33 - 71.6)^2}{10}$$

$$\sigma_x^2 = 4.31 \text{ miles de doslares al cuadrado}$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{4.31} = 2.08 \text{ miles de dolares}$$

La media de la distribución muestral es igual a la media de la población original $\mu = 71.6$. El error estándar, el cual mide el grado de dispersión de las 10 medias muestrales alrededor de μ , indica en cuanto puede variar la media muestral de la media poblacional. ¿Esto es lo que se tiene en mente cuando se esperaba un error de muestreo relativamente pequeño?

2. Ejercicio distribución muestral de proporciones

La ANH está considerando en hacer pública la cantidad de pozos perforados en el año 2011 por ECOTRAL. Según la ANH esta empresa tuvo un porcentaje en comparación con otras empresas de 25. Silvia laser, directora de relaciones públicas de la ANH, sospecha que el porcentaje real es menor que el 25%. Contrata a una agencia independiente para tomar una muestra aleatoria de 750 compañías petroleras las cuales hayan realizado perforaciones en el año 2011. La agencia encuentra que 175 compañías habían realizados trabajos de perforación. ¿Presentan estos datos suficiente evidencia para contradecir la versión de la ANH?

Solución

$$p = \frac{x}{n} = \frac{\text{número de éxitos}}{\text{número de ensayos}} = \frac{175}{750} = 0.233$$
$$\sigma_p = \frac{P(1-P)}{n} = \frac{0.25(1-0.25)}{750} = 0.0158$$
$$Z = \frac{p-P}{\sigma_p} = \frac{0.233-0.25}{0.0158} = -1.08$$

Con Z en la tabla de distribución normal se halla un área bajo la curva de 0.3599 (área entre 0.25 y 0.233). Entonces $0.5000 - 0.3599 = 0.1401$ (área a la izquierda de 0.233). La probabilidad de encontrar 175 compañías o menos que realizaron trabajos de perforación en una muestra de 750 es aproximadamente el 14% si el porcentaje verdadero de la población es 25%.

3. Ejercicio tamaño de la muestra para medias

Carolina rueda desea estimar el peso medio de las bolsas de cemento utilizadas en la lechada en la operación de completamiento. La media muestral no debe diferir en más de dos libras de la media poblacional con un 95% de confianza. Se estima que la desviación estándar de la población es 10 libras. ¿Qué tamaño deberá tener la muestra?

Solución

Con el valor de 0.95 se busca el valor Z en la tabla de distribución normal. $Z = 1.96$

$$n = \frac{1.96^2(10)^2}{2^2} = 96.04$$

Redondeando a 97, por lo tanto un tamaño de muestra de 97 bolsas de cemento dará la seguridad del 95% de que la media de la muestra no se aleje más de 2 libras de la media de la población.

4. Ejercicio tamaño de la muestra para proporciones

Richard es gerente de Alcopol, una compañía que se especializa en la instalación de sistemas de levantamiento artificial en pozos petroleros, quiere tener una idea de la proporción de clientes que continúan utilizando bombeo mecánico como sistema de levantamiento artificial. Acaba de leer un artículo sobre los efectos y problemas que este sistema atrae al yacimiento. Debido a la creciente importancia que ha tomado otros sistemas de levantamiento artificial en el mundo. Richard cree que debería plantear la idea a sus clientes de las ventajas de estos sistemas e implementarlas en sus pozos. ¿Cuál debe ser el tamaño de la muestra de clientes para probar la efectividad de los sistemas? Richard quiere estar bastante seguro (99% de confianza) de que la estimación muestral no se aleja más de 0.03 de la proporción poblacional verdadera. Alcopol no tiene idea de que proporción de clientes siguen utilizando bombeo mecánico, se usa 0.5 para obtener el tamaño de la muestra como una estimación de la proporción de la población.

Solución

Con el valor de 0.99 se busca el valor Z en la tabla de distribución normal. $Z = 2.575$

$$n = \frac{2.575^2(0.5)(1 - 0.5)}{0.03^2} = 1841.8$$

Redondeando a 1842. Un tamaño de muestra de 1842 proporcionara un nivel de confianza del 99% de que la muestra de los clientes que siguen utilizando bombeo mecánico estará a 3 puntos porcentuales del valor verdadero de la población.

5. Ejercicio intervalo de confianza para la media poblacional - muestras grandes

Gattos, operadora de la planta deshidratadora de gas chupanagua está considerando en aumentar la calidad de su gas, para incrementar así el tamaño de los clientes que quieran comprar su gas, con ello aumentar la rentabilidad de la compañía. Gattos estima el tamaño promedio de los clientes en la actualidad, los cuales adquieren su gas a diferentes precios. Selecciona una muestra de 36 clientes los cuales adquieren el ft^3 de gas cuya media es de \$14.37 con una desviación estándar de \$4.02.

- Construya una estimación por intervalo con el 95% de confianza
- Si se conoce la desviación estándar de la población como \$3.50, construya una estimación por intervalo con el 90% de confianza.

Solución

a. Con 0.95 se calcula Z de la tabla de distribución normal. $Z = 1.96$

$x = 14.37$, $S = 4.02$ y $n = 36$

$$I.C \text{ para estimar } \mu = x \pm Z \frac{S}{n} = 14.37 \pm 1.96 \frac{4.02}{36}$$

$$I.C \text{ para estimar } \mu = 14.37 \pm 1.31$$

$$I.C \text{ para estimar } \mu = \$13.06 \text{ a } 15.68$$

Gattos puede decir con una confianza del 95% que el costo promedio de un ft^3 de gas para un cliente está entre \$13.06 y \$15.68.

b. Con 0.90 se calcula Z de la tabla de distribución normal. $Z = 1.645$ y $S = 4.02$

$$I.C \text{ para estimar } \mu = x \pm Z \frac{S}{n} = 14.37 \pm 1.645 \frac{3.50}{36}$$

$$I.C \text{ para estimar } \mu = 14.37 \pm 0.96$$

$$I.C \text{ para estimar } \mu = \$13.41 \text{ a } 15.33$$

Gattos puede decir con una confianza del 90% que el costo promedio de un ft³ de gas para un cliente está entre \$13.41 y \$15.33

6. Ejercicio intervalo de confianza para la media poblacional - muestras pequeñas

Gilberto, es el asistente del ingeniero de lodos y debe estimar el número promedio de galones de lodo – pie perforado. Desde hace 22 días Gilberto viene haciendo este cálculo y hallo un promedio de galones –pies perforados de 3.2, con una desviación estándar de 0.6. Construya una estimación por intervalo con el 99 % de confianza para los galones – pies perforados promedio en el pozo e interprete el significado de este intervalo. Se sabe que la distribución se acerca mucho a la normal.

Solución

El valor apropiado de t se halla de la tabla de distribución t , para un nivel de significancia $\alpha = 0.01$ y un grado de libertad ($\gamma = 22 - 1 = 21$). $t = 2.831$

$$I.C \text{ para estimar } \mu = x \pm t S_x = 3.2 \pm 2.831 \frac{0.6}{22}$$

$$I.C \text{ para estimar } \mu = 3.2 \pm 0.36$$

$$I.C \text{ para estimar } \mu = 2.84 \text{ a } 3.56$$

Gilberto puede ahora informar al ingeniero con un 99% de confianza, que el número promedio de galones – pie perforado durante los 22 días estuvo entre 2.84 y 3.56.

7. Ejercicio intervalo de confianza para una proporción poblacional

La corporación Kenloss ha estado recibiendo reclamos por parte de diferentes compañías petroleras que utilizan el chiller que comercializa desde hace cinco años. Para resolver este problema, la compañía quiere conocer primero los porcentajes de chiller vendidos que tienen problemas. Durante los últimos seis meses ha estado vigente un programa de descuentos, de manera que la empresa decide usar los nombres de la lista de descuentos como la población de la que se obtendrá la muestra. Se escoge una muestra aleatoria de 1800 nombres a partir de esta lista, y se les envía por correo un breve cuestionario a las compañías junto con un vale de descuento por valor de \$500. Contestaron un total de 1675 encuestados, lo que representa un buen porcentaje, 93%. La compañía piensa que obtuvo una muestra representativa. El 12% de los encuestados dijo haber experimentado problemas. La estimación por intervalo se calcula con un nivel de confianza del 99%.

Solución

Para encontrar el valor de Z correcto para un 99%, se divide este valor entre 2, lo que produce un área en la tabla de distribución normal de 0.4950. Así el valor de Z = 2.575

$$I.C \text{ para estimar } P = 0.12 \pm 2.575 \frac{0.12(1 - 0.12)}{1675}$$

$$I.C \text{ para estimar } P = 0.12 \pm 0.02$$

$$I.C \text{ para estimar } P = 0.10 \text{ a } 0.14$$

Kenloss puede establecer con el 99% de confianza que el porcentaje de compradores que han tenido problemas con el chiller esta entre el 10 y 14 %. La dirección de la compañía no considera satisfactorio este nivel de calidad e inicia un programa de mejora de la calidad.

8. Ejercicio intervalo de confianza para la diferencia entre dos medias – muestras grandes

Carlos, ingeniero de yacimientos de la corporación Natrac, la cual se encarga de prestar asesorías en simulación de yacimientos a diferentes compañías petroleras, recientemente instituyo dos programas de capacitación para los nuevos ingenieros de la corporación. Para probar la efectividad relativa de cada programa, a 45 ingenieros($n_1= 45$) entrenados en el primer programa se les hizo la prueba de competencia. El puntaje promedio fue $\bar{x}_1= 76.00$ con $S_1 = 13.50$ puntos. Las 40 personas ($n_2= 40$) entrenadas bajo el segundo programa reportaron un puntaje promedio de $\bar{x}_2= 77.97$ con $S_2 = 9.05$. La gerencia desea saber si un programa de entrenamiento es más efectivo que el otro. Siendo usted la persona seleccionada para tomar esta determinación, usted decide construir un intervalo de confianza del 99% para hallar la diferencia entre los puntajes de competencia promedio de los empleados entrenados que se encuentran en cada programa. Usted también es responsable de recomendar cual programa de entrenamiento deberá utilizar la compañía exclusivamente para un mejor desempeño.

Solución

El error estándar de la diferencia entre la media muestral es

$$S_{x_1-x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(13.50)^2}{45} + \frac{(9.05)^2}{40}} = 2.47$$

El intervalo del 99%

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm ZS_{x_1-x_2}$$

$$I.C \text{ para } \mu_1 - \mu_2 = 76.00 - 77.97 \pm 2.58 \text{ (2.47)}$$

$$I.C \text{ para } \mu_1 - \mu_2 = -1.97 \pm 6.3726$$

$$I.C \text{ para } \mu_1 - \mu_2 = -8.34 \text{ a } 4.40$$

Usted puede tener un 99% de seguridad en que $(\mu_1 - \mu_2)$ está entre -8.34 y 4.40 puntos. Usted puede reportar con un 99% de confianza que los puntajes promedios de quienes recibieron entrenamiento con el programa 2 están entre 4.40 puntos menos y 8.34 puntos más que los que entrenaron con el programa 1. Es importante saber que debido a que el intervalo contiene cero, usted puede estar 99% seguro de que no hay diferencia en los puntajes promedio. Se puede concluir que cualquiera de los programas puede utilizarse para entrenar a los ingenieros recientemente contratados.

9. Ejercicio intervalo de confianza para la diferencia entre dos medias – muestras pequeñas

Agméld, utiliza dos tipos de brocas para sus taladros de perforación. Las pruebas de desgaste para medir la durabilidad revelaron que 13 brocas del tipo duraron un promedio de 11.3 días, con una desviación estándar de 3.5 días, mientras que 10 del tipo 2 duraron un promedio de 7.5 días, con una desviación estándar de 2.7 días, el tipo 1 es más costoso para fabricar y el director ejecutivo de Agmé no desea utilizarlo a menos que tenga un promedio de duración de por lo menos de 8 días más que el tipo 2. El director ejecutivo tolerara una probabilidad de error de solo el 20%. No existe evidencia que sugiera que las varianzas de la duración de los dos productos sean iguales.

Solución

$$x_1 = 11.3 \quad x_2 = 7.5$$

$$n_1 = 13 \quad n_2 = 10$$

$$s_1 = 3.5 \quad s_2 = 2.7$$

$$g.l = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

$$g.l = \frac{[(3.5)^2/13 + (2.7)^2/10]^2}{[(3.5)^2/13]^2/(13 - 1) + [(2.7)^2/10]^2/(10 - 1)} \approx 20$$

Un intervalo de confianza del 98% ($\alpha = 0.02$) con 20 g.l. requiere un valor t de 2.528.

$$I.C \text{ para } \mu_1 - \mu_2 = x_1 - x_2 \pm t' \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$I.C \text{ para } \mu_1 - \mu_2 = 11.3 - 7.5 \pm 2.528 \sqrt{\frac{(3.5)^2}{13} + \frac{(2.7)^2}{10}}$$

$$I.C \text{ para } \mu_1 - \mu_2 = 3.8 \pm 3.3$$

$$I.C \text{ para } \mu_1 - \mu_2 = 0.5 \text{ a } 7.1 \text{ semanas}$$

Agmé puede estar 98% seguro de que el tipo1 dura entre 0.5 y 7.1 semanas más que el tipo 2. Debido a que la diferencia requerida de ocho semanas no está en el intervalo. El director ejecutivo puede estar 98% seguro de que no desea utilizar el tipo 1.

10. Ejercicio intervalo de confianza para la diferencia entre dos proporciones

Una empresa de salud ocupacional y riesgos profesionales es la encargada de realizar un estudio en el campo petrolero de chambe que para determinar si los accidentes de los trabajadores en el turno del día es diferente al de los trabajadores del turno de la noche. Se realiza una comparación de 150 trabajadores de cada turno. Los resultados muestran que 37 trabajadores diurnos han sufrido por lo menos cinco accidentes de trabajo durante el año anterior, mientras que 52 trabajadores nocturnos han sufrido por los menos cinco accidentes. ¿Qué revelan estos datos sobre la tendencia a la accidentabilidad entre los trabajadores? Calcule un intervalo de confianza del 90% para la diferencia entre las proporciones de trabajadores de los dos turnos que se accidentaron cinco veces o más.

Solución

$$p_1 = \frac{37}{150} = 0.25, \quad p_2 = \frac{52}{150} = 0.35$$

$$S_{p_1-p_2} = \frac{0.25(1-0.25)}{150} + \frac{0.35(1-0.35)}{150} = 0.0526$$

$$I.C \text{ para } P_1 - P_2 = p_1 - p_2 \pm ZS_{p_1-p_2}$$

$$I.C \text{ para } P_1 - P_2 = 0.25 - 0.35 \pm 1.65 (0.0526)$$

$$I.C \text{ para } P_1 - P_2 = -0.10 \pm 0.087$$

$$I.C \text{ para } P_1 - P_2 = -0.187 a - 0.013$$

Debido a que la proporción de trabajadores nocturnos que se accidentaron cinco veces o más, se restó de la proporción de trabajadores diurnos que se accidentaron, la empresa puede estar 90% segura de que la proporción de

trabajadores nocturnos accidentados en cinco o más oportunidades esta entre 1.3% y 18.7% más alta que los del turno diurno.

11. Ejercicio prueba de dos colas para la media poblacional – muestras grandes

¿El tiempo promedio de centrifugado de un crudo, para retirarle el agua libre en el laboratorio de fluidos ha variado de 3.8 minutos (la media de hace dos años)? Esta es una pregunta que hace uno de los estudiantes al laboratorista. El profesor de la asignatura está familiarizado con artículo de una revista internacional que habla sobre el tiempo mínimo de centrifugado para retirarle el agua libre al crudo, debido a esta inquietud el docente decide hacer un seguimiento cuidadoso al tiempo mínimo de centrifugado.

Una muestra aleatoria de 100 clases de fluidos de diferentes campos revela una media muestral de 4 minutos con una desviación estándar de 0.5 minutos. Se escoge un nivel de significancia de 0.02 para esta prueba de hipótesis.

Solución

Como la pregunta se refiere a un cambio en cualquier dirección en relación con el promedio anterior, la hipótesis alternativa se establecerá como una prueba de dos colas, la hipótesis nula y alternativa son:

$$H_0: \mu = 3.8$$

$$H_A: \mu \neq 3.8$$

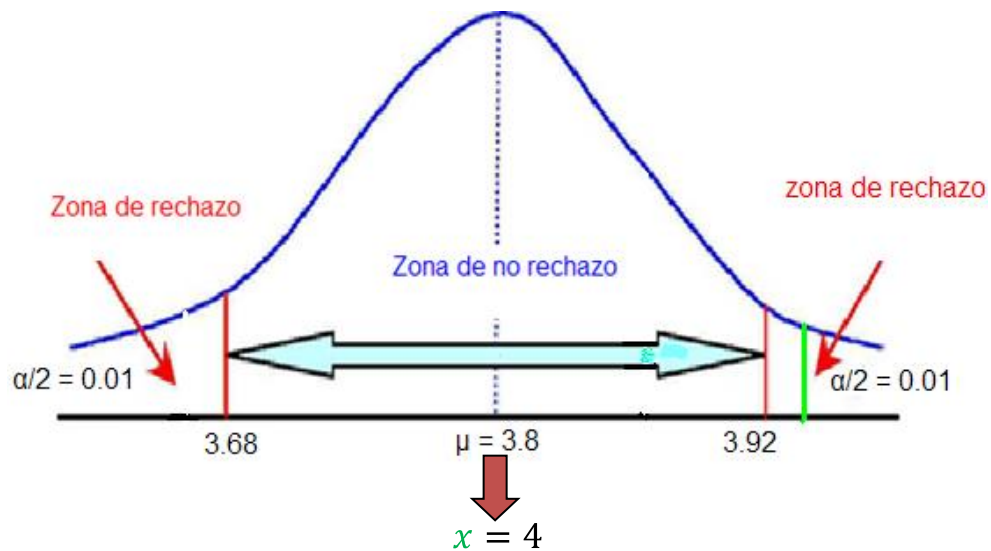
$\alpha = 0.02$, nivel de significancia con el cual se calcula el estadístico Z. $Z = 2.33$. Y

Se grafica la media muestral \bar{X} en la campana de Gauss.

$$L.C = \mu \pm Z \frac{s}{n} = 3.8 \pm (2.33) \frac{0.5}{100} = 3.8 \pm 0.1165$$

$$L.C = 3.68 \text{ a } 3.92$$

Figura 17. Prueba dos colas para la media poblacional (Ejercicio 11)



Como la x queda por fuera de los límites de confianza, se rechaza la H_0 ; y se acepta H_A . Ósea, el tiempo mínimo requerido para retirarle el agua libre al crudo por medio de la centrifuga es diferente a 3.8 minutos.

12. Ejercicio prueba de hipótesis para la media poblacional – muestras pequeñas

La corporación Tapico desea saber cuál es la máxima tensión de ruptura que soportan los cables de acero que fabrica para el sostenimiento de la sarta de perforación. Un cliente importante está interesado en la compra de un numero grande de cables y ha establecido que el punto promedio de ruptura no debe ser menor que una tonelada. Tapico piensa que una tonelada es aproximadamente el punto de ruptura de los cables. Pero decide probar la hipótesis de que la tensión de ruptura promedio (media) es una tonelada. Tapico decide que puede costear la

ruptura de 10 cables y diseña un proceso para hacerlo, elige un nivel de confianza de 0.05 para la prueba, con una media muestral de 0.96 y una desviación estándar de 0.15 toneladas.

Solución

Cada uno de los cables se romperá y se registra el peso en el punto de ruptura. El promedio de estos puntos se usara para probar la hipótesis nula. Una muestra pequeña es apropiada porque cada elemento muestreado se destruye en el proceso. Una muestra grande sería demasiado costosa. La hipótesis nula y alternativa son:

$$H_0: \mu \geq 1$$

$$H_A: \mu < 1$$

$\alpha = 0.05$ nivel de significancia y con 9 grados de libertad con el cual se calcula el estadístico t. $t = 1.833$

$$\mu = 1$$

$$S = 0.15$$

$$n = 10$$

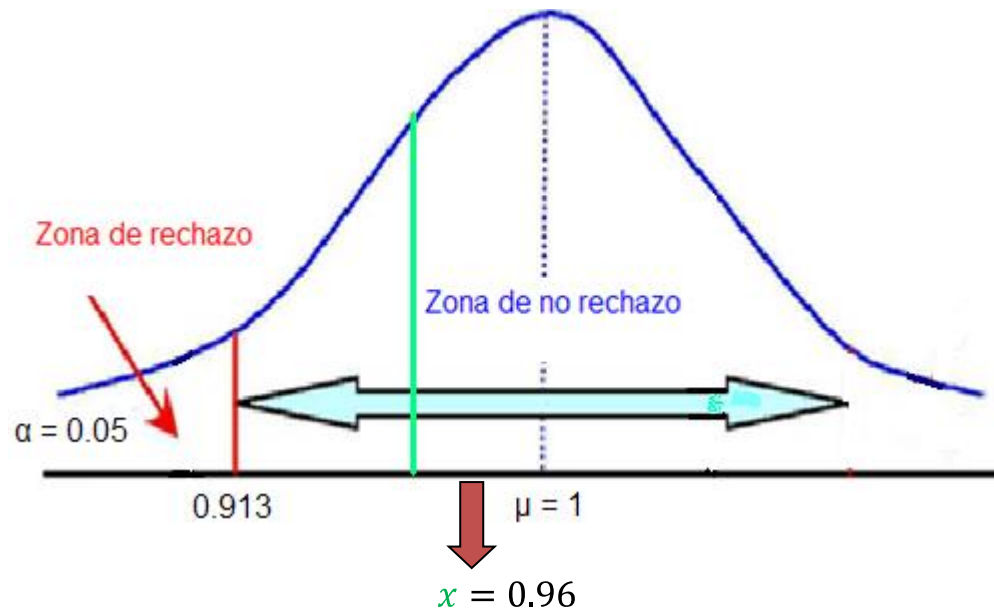
$$L.C = \mu - t \frac{S}{n} = 1 - (1.833) \frac{0.15}{10}$$

$$L.C = 1 - (0.0869)$$

$$L.C = 0.913$$

Se grafica la media muestral \bar{x} en la campana de Gauss.

Figura 18. Prueba hipótesis para la media poblacional (Ejercicio 12)



Como la x no es menor al límite de confianza inferior, la hipótesis nula (H_0) no se rechaza, según la prueba de hipótesis, se justifica la afirmación de la compañía sobre el punto de ruptura promedio de una tonelada.

13. Ejercicio prueba de hipótesis para la proporción poblacional

Como director de operaciones de bombas electrosumergibles de diferentes pozos en el centro de Barrancabermeja, usted considera que el 60% de las bombas presentan un desgaste considerable y necesitan un rápido mantenimiento. Usted intenta establecer una importante política respecto a la estructura de precios sobre esta proporción. Una muestra de 800 bombas instaladas revela que 492 bombas necesitan un mantenimiento, produciendo una proporción muestral de $p = 492/800 = 0.615$. A un nivel de significancia del 5%, ¿Qué puede concluir sobre la proporción de todas las bombas que requieren mantenimiento?

Solución

Las hipótesis son:

$$H_0: P = 0.60$$

$$H_A: P \neq 0.60$$

$\alpha = 0.05$, nivel de significancia con el cual se calcula el estadístico Z. $Z = 1.96$

$n = 800$

$p = 0.6$

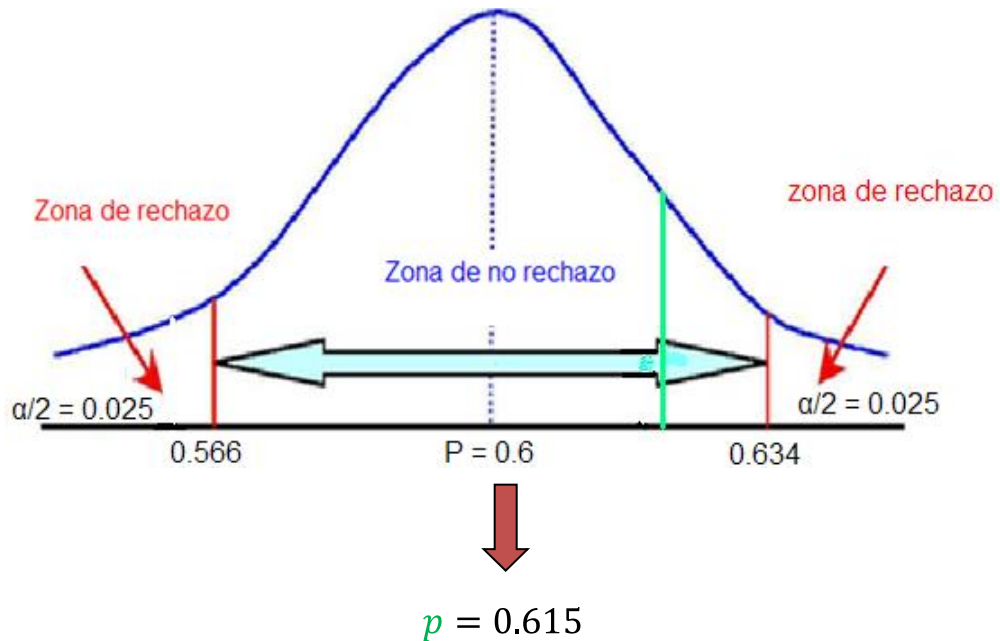
$$L.C = P \pm Z \frac{\sqrt{P(1 - P)}}{\sqrt{n}} = 0.6 \pm (1.96) \frac{\sqrt{0.6(1 - 0.6)}}{\sqrt{800}}$$

$$L.C = 0.6 \pm 0.034$$

$$L.C = 0.566 \text{ a } 0.634$$

Se grafica la media muestral p en la campana de Gauss.

Figura 19. Prueba de hipótesis para la proporción poblacional (Ejercicio13)



El $p = 0.615$ está en la zona de no rechazo. La evidencia de la muestra confirma que $P = 0.6$. Ahora es posible desarrollar su política de precios con base en esta conclusión.

14. Ejercicio prueba de hipótesis para la diferencia entre dos medias – muestras grandes

“Se prevé una mayor dependencia del petróleo extranjero con el desplome de la producción en Estados Unidos”, según publicó el New York Times en primera página, el 20 de enero de 1990. La compañía Purget Power se pregunta si ello provocara un incremento de los precios del petróleo, lo que a su vez podría originar un incremento en el uso de electricidad para la calefacción de las casas. Para medir los cambios experimentados desde el año anterior, se elige una muestra al azar de 40 casas en el mes de enero de 1989, y se compara con una muestra de 50 casas en el mes de enero de 1990. Los resultados muestrales son:

$$\begin{array}{lll} x_1 = 1645 \text{ Kwh} & S_1 = 298 \text{ Kwh} & n_1 = 40(1989) \\ x_2 = 1803 \text{ Kwh} & S_2 = 305 \text{ Kwh} & n_2 = 50(1989) \end{array}$$

Realice una prueba con un nivel de significancia de 0.10 para verificar si el promedio de consumo de electricidad por casa ha cambiado durante el mes de enero. ¿A qué conclusión deberá llegar Purget Power?

Solución

Se trata de una prueba de dos colas al nivel de significancia de 0.10. El nivel de significancia se divide a la mitad $0.10/2 = 0.05$, con el valor de 0.05 se busca el valor de Z de la tabla de distribución normal. En este caso el valor apropiado de Z es 1.645

La hipótesis nula y alternativa son:

$$H_0: \mu_1 - \mu_2 = 0$$

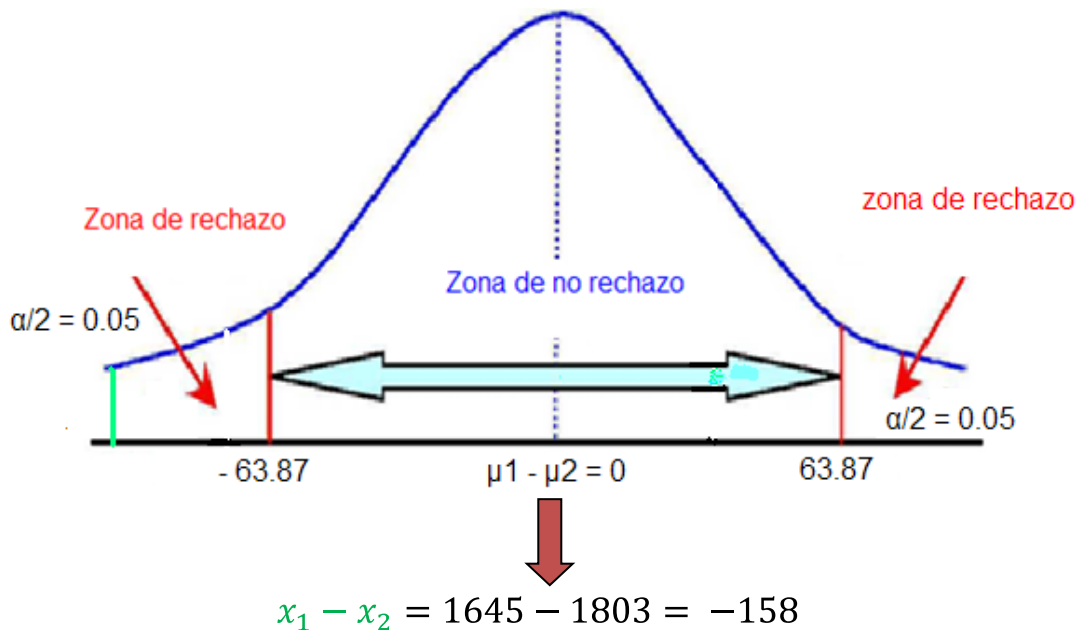
$$H_A: \mu_1 - \mu_2 \neq 0$$

$$L.C = \mu_1 - \mu_2 \pm Z \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0 \pm (1.645) \sqrt{\frac{298^2}{40} + \frac{305^2}{50}}$$

$$L.C = -63.87 \text{ a } 63.87$$

Se grafica la diferencia entre las dos medias muestrales $x_1 - x_2$ en la campana de Gauss.

Figura 20. Prueba para la diferencia entre dos medias (Ejercicio 14)



Como la diferencia $x_1 - x_2$ queda por fuera de los límites de confianza, se rechaza la H_0 ; y se acepta H_A . Por lo tanto se concluye que el consumo promedio de electricidad por casa ha cambiado durante el mes de enero.

15. Ejercicio prueba de hipótesis para la diferencia entre dos medias – muestras pequeñas

La compañía Battes ha desarrollado un nuevo proceso de manufactura para fabricar los tubos que usan los oleoductos en Colombia de alta calidad. Como la resistencia es un elemento tan importante en el diseño del oleoducto, deberán someterse algunos de estos tubos a una prueba de tensión, y los puntos defectuosos se comparan con los de los tubos convencionales fabricados por la compañía. Los tubos son costosos tanto en términos del material como de la mano de obra, por lo que se decide someter a la prueba de tensión a 8 tubos nuevos y siete antiguos, se aplicara sobre cada tubo una fuerza de carga creciente hasta que falle. La compañía va a usar una máquina que ofrece una lectura de tensión combinada al aumentar la fuerza de carga, Se escoge un nivel de significancia de 0.05 para la prueba.

Se seleccionaron las muestras y se obtienen los siguientes resultados.

Tubos nuevos: $x_1 = 27.8 \text{ Unid}$ $S_1 = 3.4 \text{ Unid}$ $n_1 = 8 \text{ unid}$ Tubos
antiguos: $x_2 = 23.7 \text{ unid}$ $S_2 = 2.9 \text{ unid}$ $n_2 = 7 \text{ Unid}$

Solución

La hipótesis nula y alternativa son:

$$H_0: \mu_1 - \mu_2 \leq 0.$$

$$H_A: \mu_1 - \mu_2 > 0$$

Con el nivel de significancia de 0.05 y con los grados de libertad ($gl = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$) se calcula el estadístico t de la tabla de distribución t Student, para una prueba de una sola cola. En donde $t = 1.771$.

Se calcula S_D^2 :

$$S_D^2 = \frac{s_1^2 n_1 - 1 + s_2^2 n_2 - 1}{n_1 + n_2 - 2}$$

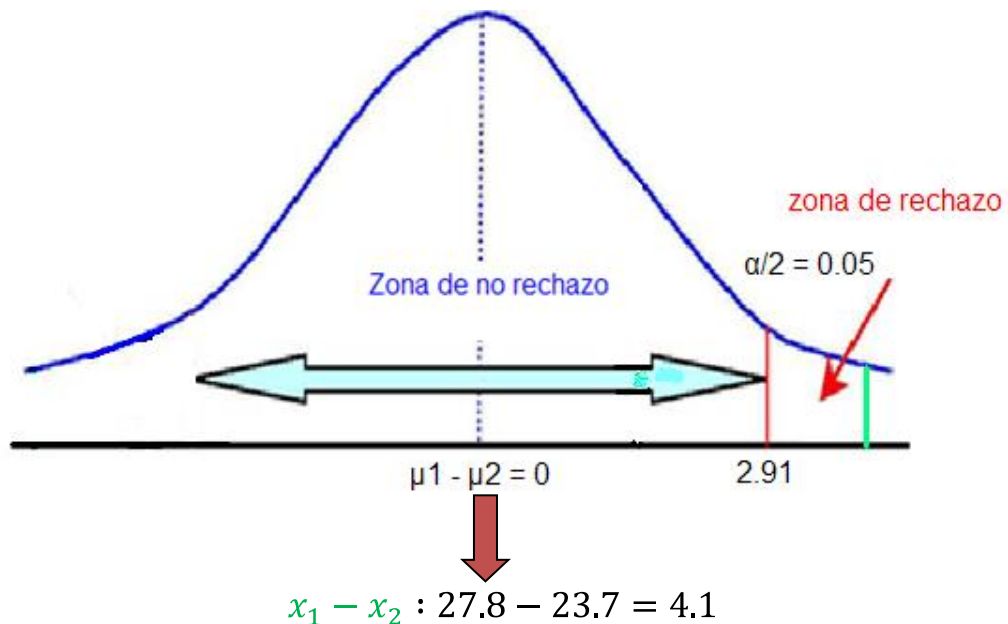
$$S_D^2 = \frac{3.4^2 8 - 1 + 2.9^2 7 - 1}{8 + 7 - 2} = 10.106$$

$$L.C = \mu_1 - \mu_2 + t \frac{S_D^2}{n_1} + \frac{S_D^2}{n_2} = 0 + (1.771) \frac{10.106}{8} + \frac{10.106}{7}$$

$$L.C = 2.91$$

Se grafica la diferencia entre dos medias muestrales $x_1 - x_2$ en la campana de Gauss.

Figura 21. Prueba para la diferencia entre dos medias (Ejercicio15)



Como la diferencia $x_1 - x_2$ es mayor que el límite de confianza superior, se rechaza la H_0 ; y se acepta H_A . Batters concluye que, basándose en una muestra pequeña, que los nuevos tubos si sobrepasan a los anteriores en términos de resistencia.

16. Ejercicio prueba de hipótesis para la diferencia entre dos proporciones

Joseph Conrad, ingeniero de producción de ECOPECROL, quiere determinar si el índice de productividad (IPR) es el mismo o no en dos pozos que pertenecen a campos diferentes. Joseph decide usar un nivel de significancia de 0.05 para la prueba, también elige muestras aleatorias de pozos de cada campo y calcula estos errores

$$\begin{array}{ll} x_1 = 55 \text{ pozos con el mismo IPR} & n_1 = 550 \text{ pozos} \\ x_2 = 90 \text{ pozos con el mismo IPR} & n_2 = 750 \text{ pozos} \end{array}$$

IPR: índice de productividad

Solución

La hipótesis nula y alternativa es:

$$H_0: P_1 - P_2 = 0$$

$$H_A: P_1 - P_2 \neq 0$$

Se trata de una prueba de dos colas al nivel de significancia de 0.05. El nivel de significancia se divide a la mitad ($\alpha = 0.05/2 = 0.025$), con el valor de 0.025 se busca el valor de Z de la tabla de distribución normal. En este caso el valor apropiado de Z es 1.96.

En la prueba para proporciones con dos muestras, la estimación de la proporción poblacional se calcula con el promedio de las proporciones de las dos muestras.

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{55 + 90}{550 + 750} = 0.1115 = P_1 = P_2$$

Cálculo de la zona de aceptación de H_0 :

$$L.C = P_1 - P_2 \pm Z \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}$$

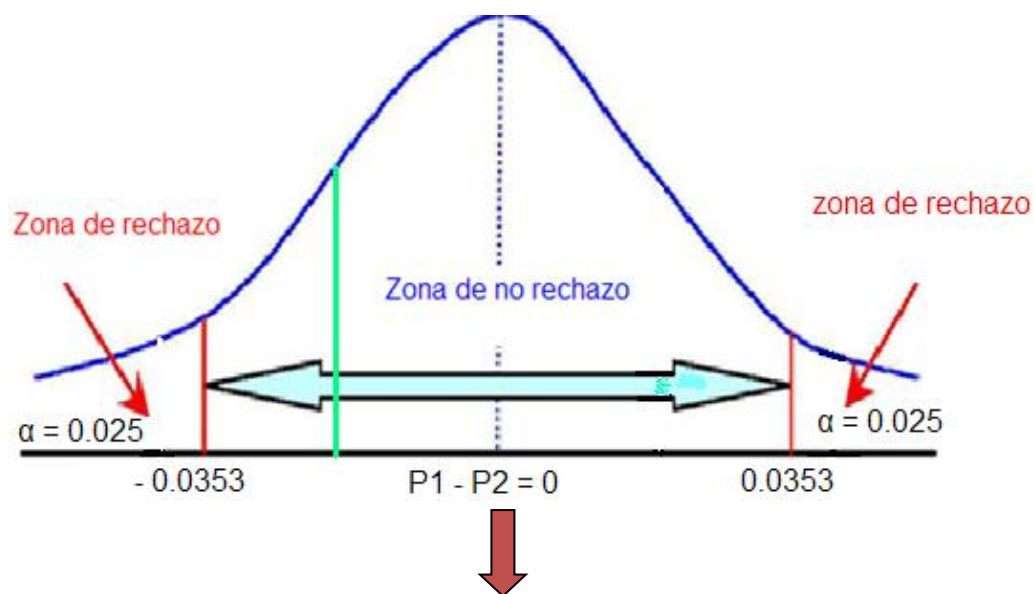
$$L.C = 0 \pm 1.96 \sqrt{\frac{0.1115(1 - 0.1115)}{550} + \frac{0.1115(1 - 0.1115)}{750}}$$

$$L.C = -0.0353 \text{ a } 0.0353$$

$$p_1 = \frac{55}{550} = 0.10 \text{ y } p_2 = \frac{90}{750} = 0.12$$

Se grafica la diferencia entre dos medias muestrales $p_1 - p_2$ en la campana de Gauss.

Figura 22. Prueba para la diferencia entre dos proporciones (Ejercicio 16)



$$p_1 - p_2 = 0.10 - 0.12 = -0.02$$

Como la diferencia $p_1 - p_2$ quedo dentro de los límites de confianza, se acepta la H_0 ; La evidencia muestral no es suficientemente fuerte para rechazar la hipótesis nula de que los dos pozos tienen el mismo IPR.

17. Ejercicio prueba de bondad de ajuste

John Dillinger gerente de la ANH, trata de seguir una política de extender un 60% de los bloques de exploración a empresas multinacionales, un 10% a empresas privadas colombianas y un 30% a la empresa estatal ECOPETROL. Para determinar si la política se está siguiendo, Hoover lozano, vicepresidente del departamento geológico, selecciona aleatoriamente 85 bloques que se aprobaron recientemente. Encuentra que 62 de tales bloques se otorgaron a multinacionales, 10 a empresas privadas colombianas y 13 a ECOPETROL. Al nivel del 10%, ¿parece que el patrón de bloques deseado se preserva?

Solución

H_0 : se mantuvo el patrón deseado: 60% multinacionales, 10% empresas privadas y 30% son de ECOPETROL

H_A : el patrón deseado no se mantuvo

Si la hipótesis nula es correcta, el señor Hoover esperaría que el 60% de los 85 bloques de la muestra sean bloques de las multinacionales. De manera que para la primera categoría $E_i = np_i = (85 \cdot 0.60) = 51$ bloques de multinacionales, además espera que $(85 \cdot 0.10) = 8.5$ de los bloques fueran otorgados a empresas privadas colombianas y $(85 \cdot 0.30) = 25.5$ bloques a ECOPETROL. Los datos se reducen en la siguiente tabla.

Tabla 20. Frecuencias de las empresas (Ejercicio 17)

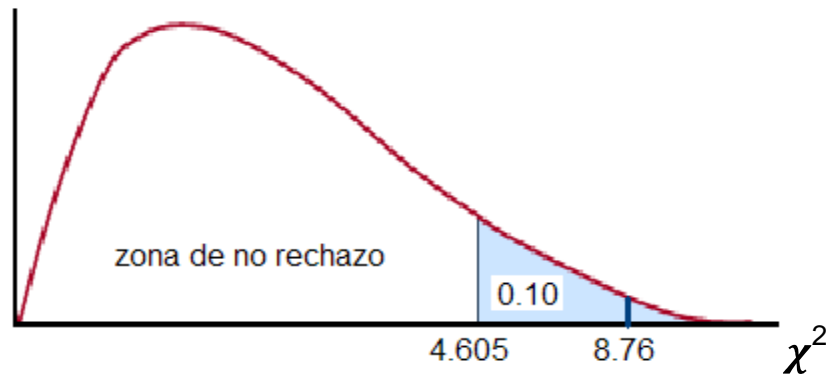
BLOQUES A:	Frecuencias Observadas (O_i)	Frecuencias Esperadas (E_i)
multinacional	62	51
Emp. Privadas	10	8.5
ECOPETROL	13	25.5
	85	85.0

El valor χ^2 es:

$$\chi^2 = \frac{(62 - 51)^2}{51} + \frac{(10 - 8.5)^2}{8.5} + \frac{(13 - 25.5)^2}{25.5} = 8.76$$

No se estimaron parámetros y $m = 0$. Con un nivel de significancia del 10% y $K = 3$ categorías de empresas (multinacionales, privadas y ECOPETROL), existen $k - m - 1 = 3 - 0 - 1 = 2$ grados de libertad. Con estos valores el señor Hoover encuentra que el valor crítico de $\chi^2_{0.10,2} = 4.605$.

Figura 23. Distribución ji – cuadrado (Ejercicio 17)



La hipótesis nula se rechazaría debido a que $8.76 > 4.605$. Las diferencias entre lo que el señor Hoover observó y lo que esperaba observar si el patrón de bloques deseado se alcanzaba era demasiado grande como para ocurrir por simple azar. Existe solo un 10% de probabilidad de que una muestra de 85 bloques seleccionados aleatoriamente pudiera producir las frecuencias observadas aquí demostradas.

18. Ejercicio prueba de independencia

Wentinton Oil Company, desea determinar si existe alguna relación entre el ingreso de su personal de operaciones nuevo y la importancia que dan al sueldo que la empresa cancela por sus servicios. Los ingenieros están agrupados en tres niveles de ingreso y se les pide asignar un nivel de significancia de 1% para poner precio a su salario. Los resultados se muestran en la siguiente tabla.

Tabla 21. Resultados de la prueba (Ejercicio18)

	Ingreso			
Nivel de importancia	Bajo	Medio	Alto	Total
Grande	$O_i = 83$	$O_i = 62$	$O_i = 37$	182
Moderado	$O_i = 52$	$O_i = 71$	$O_i = 49$	172
Poco	$O_i = 63$	$O_i = 58$	$O_i = 63$	184
	198	191	149	538

Solución

La hipótesis nula y alternativa son:

H_0 : Ingreso e importancia del sueldo son independientes

H_A : Ingreso e importancia del sueldo no son independientes

Debido a que $182/538 = 33.83\%$ de todos los que respondieron a la encuesta agregan un nivel de importancia “grande” al sueldo, entonces si el ingreso y el sueldo no están relacionados, se esperaría que el 33.83% de ellos, en cada clasificación de ingresos respondan que el sueldo era de “gran” importancia. Por tanto, los E_i para un nivel de importancia “bajo” son $(198 \cdot 0.3383) = 66.98$, $(191 \cdot 0.3383) = 64.62$, y $(149 \cdot 0.3383) = 50.41$. De la misma forma se calcula E_i para el nivel de importancia “moderado” y “poco”. Estos valores se resumen en la siguiente tabla.

Tabla 22. Resultados calculados (Ejercicio18)

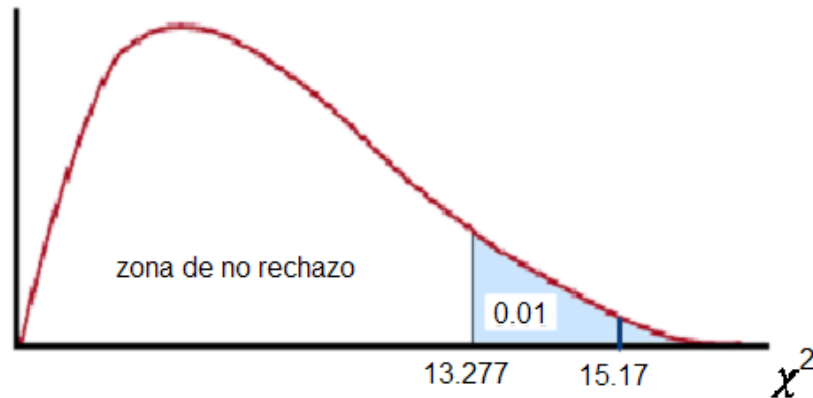
Nivel de importancia	Ingreso			
	Bajo	Medio	Alto	Total
Grande	$O_i = 83$ $E_i = 66.98$	$O_i = 62$ $E_i = 64.62$	$O_i = 37$ $E_i = 50.41$	182
Moderado	$O_i = 52$ $E_i = 63.32$	$O_i = 71$ $E_i = 61.06$	$O_i = 49$ $E_i = 47.64$	172
Poco	$O_i = 63$ $E_i = 67.72$	$O_i = 58$ $E_i = 65.32$	$O_i = 63$ $E_i = 50.96$	184
	198	191	149	538

El valor χ^2 es:

$$\chi^2 = \frac{(83 - 66.98)^2}{66.98} + \frac{(62 - 64.62)^2}{64.62} + \dots + \frac{(63 - 50.96)^2}{50.96} = 15.17$$

Con un nivel de significancia del 1% y con $(f - 1)(c - 1) = (3 - 1)(3 - 1) = 4$ grados de libertad. Se encuentra que el valor crítico de $\chi^2_{0.01,4} = 13.277$

Figura 24. Distribución ji – cuadrado (Ejercicio 18)



La hipótesis nula se rechaza. Existe solo un 1% de probabilidad de que si no existe relación entre ingreso y significancia del sueldo, las diferencias entre O_i y E_i serían lo suficientemente grandes como para producir un ji – cuadrado más grande que 13.277. Existe evidencia de una relación entre el ingreso de los ingenieros y la importancia dada al sueldo de cada uno.

19. Ejercicio comparación de la varianza de dos poblaciones normales

Oscar Mitchell, analista de los laboratorios Ebbott, un fabricante nacional de desemulsificantes, está preocupada por la calidad de uno de sus productos. Ebbott compra el material para fabricar este producto a dos proveedores. El nivel de defectos en la materia prima es aproximadamente el mismo entre los dos proveedores. Pero Oscar está preocupado por la variabilidad que existe de un embarque a otro. Si el nivel de defectos tiende a variar en forma excesiva para un proveedor, puede afectar la calidad del desemulsificante. Para comparar la variación relativa de los dos proveedores, Oscar selecciona 11 embarques de cada uno y mide los porcentajes de defectos en la materia prima, junto con la desviación estándar. Los resultados son:

$$S_1 = 0.61 \quad n_1 = 11 \text{ (Proveedor 1)}$$

$$S_2 = 0.29 \quad n_1 = 11 \text{ (Proveedor 2)}$$

¿Pruebe si la variabilidad del nivel de defectos por embarque para el proveedor 1 es mayor que para el proveedor 2, para un nivel de significancia de 0.05?

Solución

La hipótesis nula y alternativa son:

$$H_o: \sigma_1^2 - \sigma_2^2 \leq 0$$

$$H_A: \sigma_1^2 - \sigma_2^2 > 0$$

El valor F crítico se halla, primero calculando los grados de libertad para el numerador y el denominador.

$$\text{gl (numerador)} = (11 - 1) = 10; \quad \text{gl (denominador)} = (11 - 1) = 10$$

Luego con los grados de libertad del numerador y del denominador, y con el nivel de significancia de 0.05, se utiliza la tabla de distribución F para encontrar el valor crítico. En este caso $F = 2.97$.

El valor F es:

$$F = \frac{s_1^2}{s_2^2} = \frac{(0.61)^2}{(0.29)^2} = 4.42$$

Figura 25. Distribución F (Ejercicio 19)



Una de las varianzas muestrales es 4.42 veces la otra. La hipótesis nula se rechaza porque el estadístico de prueba (4.42) excede el valor crítico (2.97). Oscar debe concluir que la variabilidad en los niveles de defectos en los embarques para el proveedor 1 es mayor que para los del proveedor 2.

20. Ejercicio análisis de varianza – ANOVA

Erika Herr, una analista de la multinacional petrolera Best Foods, quiere saber si los tres campos que tienen en Colombia tienen el mismo promedio en dólares de rentabilidad. Se elige una muestra aleatoria de la rentabilidad durante seis meses en cada campo. La tabla presenta los datos recolectados de esta muestra junto con las medias muestrales para cada campo y la gran media de todos los datos. Erika hará una prueba con un nivel de significancia de 0.01. ¿Utilice el método *entre y dentro*, para estimar la varianza de las tres poblaciones?

Tabla 23. Promedio de rentabilidad de los campos (Ejercicio 20)

	Campo 1(MUS\$)	Campo 2(MUS\$)	Campo 3(MUS\$)
	12.05	15.17	9.48
	23.94	18.52	6.92
	14.63	19.57	10.47
	25.78	21.40	7.63
	17.52	13.59	11.90
	18.45	20.57	5.92
Media	18.73	18.14	8.72
Gran media: $\bar{x} = 15.20$, $k = 3$, $n = 6$			

(MUS\$): Millones de dólares

Solución

La hipótesis nula que se quiere probar es que todas las poblaciones de las que se obtuvieron los datos muestrales tienen la misma media. La hipótesis alternativa es que las poblaciones no tienen la misma media.

Cálculo de la Varianza *ENTRE* las Medias

$$S_b^2 = \frac{6[(18.73 - 15.20)^2 + (18.14 - 15.20)^2 + (8.72 - 15.20)^2]}{3 - 1}$$

$$S_b^2 = 189.27$$

Cálculo de la Varianza *DENTRO* de las Muestras:

Tienda 1:

$$(12.05 - 18.73)^2 + (23.94 - 18.73)^2 + \dots + (12.05 - 18.45)^2 = 139.82$$

Tienda 2

$$(15.17 - 18.14)^2 + (18.52 - 18.14)^2 + \dots + (20.57 - 18.14)^2 = 48.25$$

Tienda 3

$$(9.48 - 8.72)^2 + (6.92 - 8.72)^2 + \dots + (5.92 - 8.72)^2 = 26.02$$

$$S_w^2 = \frac{\sum_j \sum_i (x_{ij} - x_j)^2}{k(n-1)} = \frac{139.82 + 48.25 + 26.02}{3 \cdot 6 - 1} = 14.27$$

El valor F crítico se halla, primero calculando los grados de libertad para el numerador y el denominador.

$$\text{gl (numerador)} = (k - 1) = (3 - 1) = 2$$

$$\text{gl (denominador)} = k(n - 1) = 3(6 - 1) = 15$$

Luego con los grados de libertad del numerador y del denominador, y con el nivel de significancia de 0.01, se utiliza la tabla de distribución F para encontrar el valor crítico. En este caso $F = 6.36$.

El valor F es:

$$F = \frac{s_b^2}{s_w^2} = \frac{189.27}{14.27} = 13.26$$

Como el valor F calculado de 13.26 es mayor que el valor crítico, lo que significa que se tiene suficiente evidencia muestral para rechazar la hipótesis nula de medias poblacionales iguales. Como una de las medias de grupos es menor que las otras, Erika concluye que aunque los campos 1 y 2 pueden tener promedios de rentabilidad iguales, parece que el campo 3 está por debajo del promedio de los otros dos.

4.5 EJERCICIOS PROPUESTOS

1. La producción en barriles por día del pozo Singer 2, en los últimos días es 200, 250, 150, 200, y 350. Realice una distribución muestral y calcule la media de las medias y el error estándar para las muestras de tamaño $n = 2$.
2. Una compañía prestadora de servicios de bombeo PCP, sabe que el 8% de sus bombas cumplen con los requerimientos del pozo durante los primeros 5 años. ¿Cuál es la probabilidad de que 350 bombas, menos de 50 bombas cumplan durante los primeros 5 años?
3. Ramón, un distribuidor de brocas de perforación, es consciente que la perforación de un pozo esta influenciada por las boquillas de la broca. Ramón quiere conocer con una certeza del 98% la importancia del diámetro medio verdadero de la boquilla en la broca, con una diferencia de no mas de 0.08 pulgadas. La experiencia indica que puede suponerse que la desviación estándar es de 0.20 pulgadas.
 - a. ¿Qué tamaño de muestra se requiere para reducir el error tolerable máximo a 0.05 pulgadas?
 - b. Calcule el error tolerable máximo para una muestra de 100 observaciones.

4. Una compañía nacional que se encarga de pruebas de presiones en diferentes campos petroleros en el país, quiere estimar la proporción de pozos que tendrán la misma presión la próxima semana. El ingeniero de análisis de presiones decide seleccionar una muestra aleatoria y preguntar a los ingenieros de campo como se ha mantenido la presión. Los ingenieros de campo le suministran los datos de las presiones del 3 de mayo de 2011, los cuales mostraron que los pozos tenían en un 23% la misma presión.

¿Cuántos pozos debe elegir al azar para tener una confianza del 98% para la proporción verdadera de presiones si desea un error tolerable máximo de 3 puntos porcentuales?

5. La compañía Mutual Ltda. Es la encargada de hacer cumplir las normas HSEQ en el campo Kalapas. En los últimos meses ha tenido problemas con los empleados en lo que se refiere a accidentes de trabajo, para lo cual necesita estudiar las reclamaciones sobre los seguros de riesgos profesionales anuales de hombres entre 40 y 50 años para determinar la necesidad de aumentar la tasa de riesgo. Una muestra aleatoria de 450 reclamaciones produce una media muestral de \$742 y una desviación estándar muestral de \$53. Averigüe la estimación por intervalo con el 98% de confianza para la reclamación media poblacional.

6. Un ingeniero de calidad de la compañía, llevo a cabo una prueba de resistencia a la tracción en 15 cables utilizados en una perforación y determino una resistencia a la tracción media de 95 con una desviación estándar de 7.6. Construya una estimación por intervalo con el 98% para la resistencia a la tracción promedio de los cables de.

7. El departamento contable de GOOSERVIS OIL COMPANY quiere estimar el número total campos que tuvieron una auditoria en sus propias declaraciones el año pasado. Si una muestra aleatoria de 100 declaraciones indico que habían sido

auditadas, elabore una estimación por intervalo con un 90% de confianza para la proporción verdadera de campos cuyas declaraciones fueron auditadas el año pasado.

8. Dos procesos de producción se utilizan para producir tubos de acero (tubing). Una muestra de 100 tubos tomada del primer proceso de producción tiene una longitud promedio de 27.3 pies y $S = 10.3$ pie. Las cifras correspondientes para los 100 tubos producidos por el segundo método son 30.1 pies y 5.2 pies. ¿Qué revela un intervalo del 99% sobre la diferencia en las longitudes promedio de los tubos producidos por estos dos métodos?

9. Growservis vende un emulsificante comercial producido en dos plantas de Atlanta y Dallas. Las quejas recientes de una compañía de lodos sugieren que a los envíos de Atlanta les falta peso, comparados con los envíos de Dallas. ¿Si 10 bolsas de la planta de Atlanta tienen un promedio de 96.3 libras con $S = 12.5$ y 15 bolsas de la planta de Dallas tienen un promedio de 101.1 con $S = 10.3$, un intervalo de confianza del 99% confirma esta queja? Se asumen varianzas iguales.

10. Su campo utiliza dos sistemas de recobro diferentes, se han presentado problemas en cuanto al ajuste apropiado, debido al funcionamiento de los equipos. Como ingeniero de yacimientos su trabajo es estimar la diferencia en la proporción de defectos producidos por cada sistema, Se tomaran muestras de tamaño $n_1 = 120$ y $n_2 = 95$, el primer sistema produjo 38% de defectos y el segundo sistema 43% de defectos. Si la evidencia sugiere que la diferencia en la proporción de defectos excede del 5%, se implementara en el yacimiento el sistema que menores defectos tenga. ¿Qué decisión tomaría usted?

11. Una de las refinerías que existe en Colombia gasto remodelando sus instalaciones miles de dólares para el mejoramiento de sus productos y servicios, el gerente comercial de la compañía espera que con esta inversión se mejore la

calidad del proceso y pueda tener una mayor competitividad en el mercado petrolero, y que nuevos clientes quieren adquirir sus servicios. Antes de realizar la remodelación la compañía promediaba U\$ 3.5 millones por semana, ahora que ha terminado la remodelación, el gerente toma una muestra de 36 semanas para ver si la construcción afecto de alguna manera a la refinería. Se reportó una media de U\$ 3.71 millones por semana y una desviación media de U\$ 0.6 millones. ¿Qué puede decir el gerente a un nivel de significancia del 1%?

12. Como supervisor de los productos que se utilizan para preparar el lodo de perforación, es su responsabilidad garantizar que las bolsas que llegan del viscosificante que vende la firma BAKER posean en promedio 25 libras. Urgido por la preocupación de que esta especificación del peso no se cumpla, usted selecciona 25 bolsas y encuentra una media de 23.8 libras con una desviación estándar de 6.6 libras. ¿Debería ordenar a la BAKER que se hagan los ajustes en el proceso de llenado de las bolsas? Para minimizar el error tipo I, escoja un valor de significancia de 1.

13. Un artículo de una revista petrolera habla sobre las nuevas tecnologías que están implementado las compañías para aumentar el recobro de crudo en sus campos. Toroto internacional, una de estas compañías, esta intentando aumentar el recobro a través de la implementación de la tecnología llamada DOWS (DOWNHOLE WATER - OIL SEPARATION). Durante varios años, el porcentaje de recobro de uno de sus campos ha sido de alrededor del 6%. El departamento de yacimientos considera que tiene evidencia de que las condiciones actuales del yacimiento han sido la causa de este porcentaje. Durante el año anterior Toroto realizo chequeos a 478 pozos. De estos, 42 tenían una baja producción, la evidencia muestral es $n = 478$, $p = 42/478 = 0.0879$. El departamento de yacimientos señala que el recobro del campo aumentara a 9% si se implementa la tecnología DOWS, como una de las medidas que debe tomarse para asegurar la implementación de esta tecnología en el campo. Toroto toma como nivel de

significancia un valor del 5%. ¿El departamento de yacimientos tendrá la razón en que el recobro se aumentara hasta el 9%?

14. MINMINAS lleva a cabo estudios diseñados para estimar las partes por millón (ppm) de CO₂ emitidas en los campos petroleros. Se ha pedido a MINMINAS que compare las ppm que emite un campo que utiliza un sistema de inyección de CO₂ con un campo que no utiliza este sistema de recobro de crudo. MINMINAS elige 80 pozos y prueba la cantidad de CO₂ emitida por los 40 pozos que utilizan inyección de CO₂ y de los 40 que utilizan otro sistema, los resultados son:

$$\begin{array}{lll} x_1 = 29.3 \text{ ppm} & S_1 = 1.9 \text{ ppm} & n_1 = 40(\text{inyeccion } CO_2) \\ x_2 = 28.9 \text{ ppm} & S_2 = 1.7 \text{ ppm} & n_2 = 40(\text{otro sistema}) \end{array}$$

¿Cuál será la conclusión de MINMINAS si la prueba se hace con un nivel de significancia de 0.01?

15. El campo de crudo liviano Tendy quiere determinar si aumentara la productividad del campo utilizando un sistema de levantamiento de bombeo mecánico. El ingeniero de petróleos decide utilizar 13 bombas en pozos seleccionados al azar y comparar los resultados con 15 bombas que están instaladas en un campo vecino, el número de barriles producidos (Bpd) por día es:

$$\begin{array}{lll} x_1 = 450 \text{ Bpd} & S_1 = 30 \text{ Bpd} & n_1 = 13(\text{bombas campo teddy}) \\ x_2 = 430 \text{ Bpd} & S_2 = 20 \text{ Bpd} & n_2 = 15(\text{bombas campo vecino}) \end{array}$$

¿Cuál será la conclusión del ingeniero si la prueba se lleva a cabo con un nivel de significancia de 0.01?

16. La Pacifica Rubiales, una importante compañía que suministra gas y petróleo, tiene intensiones en comprar una empresa más pequeña: Idaho Power. El asesor financiero de Pacifica informa que hay una proporción mayor de hombres accionistas que de mujeres accionistas que apoyan la compra. Kim Gamble, presidenta de Pacifica, ha pedido a Oscar Vanegas Angarita, el estadístico de la empresa, que lleve a cabo una encuesta de una muestra aleatoria de accionistas para confirmar este informe. Oscar muestrea 1000 accionistas (400 hombres y 600 mujeres) y determina que 220 hombres y 320 mujeres apoyan la compra. Oscar utiliza un nivel de significancia de 0.05. ¿Cuál deberá ser la conclusión de Oscar?

17. La compañía de petróleos Shaga de china está a punto de introducir en el mercado latinoamericano sus servicios de perforación y quiere saber si las compañías Latinoamérica prefieren una empresa determinada en un contrato por prestación de servicios. Se ofrecen sus servicios en 200 licitaciones, donde solo se permite participar 4 compañías, los resultados de las pruebas (licitaciones) son:

Tabla 24. Numero de licitaciones de cada empresa (ejercicio 17)

Cantidad de contratos que gana cada compañía			
A	B	C	D
60	43	46	51

¿Con un nivel de significancia de 0.01, pruebe si la gente prefiere una compañía a otra?

18. Las grandes multinacionales de petróleos dan a sus empleados bonos de campo por cada día que laboren en este. Leo Guzmán, quiere saber si el valor del bono tiene alguna relación con el área de desempeño, los datos recogidos de una muestra de varios empleados son:

Tabla 25. Valor del bono de cada área (Ejercicio 18)

Área	Valor del bono			Total
	\$10	\$50	\$100 o mas	
Perforación	22	26	54	102
Producción	33	31	22	86
Yacimientos	41	43	19	103

¿Pruebe si el valor del bono se relaciona con el área de desempeño del empleado, con un nivel de significancia de 0.01?

19. La compañía Ventaga Ltda. utiliza dos máquinas para producir empaques para ensambles de BOP (Preventoras). Es esencial que el diámetro de estos empaques no varíe en forma significativa de las especificaciones en más de 0.3 pulgadas. El gerente de producción, Pedro Hernández, está preocupado porque una de las maquinas está produciendo empaques que varían en una cantidad excesiva, y decide comparar las dos máquinas, elige una muestra aleatoria de 20 empaques del lote de producción de cada máquina y calcula la desviación estándar de los diámetros. ¿Pruebe si la variabilidad de los diámetros de los empaques difiere en las dos máquinas, con un nivel de significancia de 0.05?

$$S_1 = 0.002 \quad n_1 = 20 \text{ (maquinas)}$$

$$S_2 = 0.0015 \quad n_2 = 20 \text{ (maquinas)}$$

20. La dueña de la corporación NACFRAT, Marlene Mojica, decide reemplazar varios separadores bifásicos en el campo que opera su empresa. Después de investigar la situación, concluye que cuatro marcas parecen comparables en términos de costos y vida útil proyectada. Marlene determina que el factor decisivo entre las cuatro marcas es la cantidad de fluido que puede separar en una operación normal en un día a una determinada presión. Mide entonces el volumen

de petróleo separado, en barriles. Con los siguientes resultados que se muestran en la tabla. ¿A qué conclusión deberá llegar Marlene, si prueba la hipótesis con un nivel de significancia de 0.01?

Tabla 26. Cantidad de petróleo separado (Ejercicio 20)

Marca A	Marca B	Marca C	Marca D
5.4	6.1	8.2	7.2
5.9	5.9	8.5	6.5
6.2	6.3	6.9	6.8
7.0	6.5	9.4	7.1
5.1	7.2	7.9	7.4
5.5	6.9	8.6	6.7

21. una compañía de perforación prueba dos brocas perforando a un máximo de 112 pies y registrando el número de horas que tomo el procedimiento. La primera broca se utilizó en 12 casos, resultando un tiempo promedio de $x_1 = 27.3$ horas y $s_1 = 8.7$ horas. Se excavaron 10 pozos con una segunda broca, produciendo un $x_2 = 31.7$ horas y $s_2 = 8.3$ horas. **a)** ¿Parece que una broca es más efectiva que la otra? Sea $\alpha = 0.1$. Ninguna evidencia sugiere que las varianzas son iguales. Responda construyendo el intervalo de confianza apropiado. **b)** Si por otro lado todos los pozos fueron perforados con el mismo equipo de perforación y en el mismo suelo, y por otras razones, la compañía de perforación tendría varianzas iguales, ¿Cómo se diferenciaría en parte la prueba?

5. DISEÑO DE EXPERIMENTOS

Los modelos de diseño de experimentos son modelos estadísticos clásicos cuyo objetivo es averiguar si unos determinados factores influyen en una variable de interés y, si existe influencia de algún factor, cuantificar dicha influencia.

La metodología del diseño de experimentos se basa en la experimentación. Es sabido que si se repite un experimento, en condiciones indistinguibles, los resultados presentan una cierta variabilidad. Si la experimentación se realiza en un laboratorio donde la mayoría de las causas de variabilidad están muy controladas, el error experimental será pequeño y habrá poca variación en los resultados del experimento. Pero si se experimenta en procesos industriales o Administrativos la variabilidad será mayor en la mayoría de los casos.

5.1 METODOLOGIA DEL DISEÑO DE EXPERIMENTOS

Se propone una metodología para desarrollar con éxito un experimento, la cual se puede extender a cualquier prueba de hipótesis estadística. Esta metodología consta de 10 pasos básicos.

Paso 1: Definir claramente la hipótesis a comprobar

Es de importancia fundamental identificar de forma muy específica el objetivo del Experimento, es decir, la pregunta exacta que se quiere contestar o la hipótesis que se necesita contrastar. Esta pregunta básica se formulará por escrito.

Paso 2: Identificar la variable independiente

La variable independiente representa la característica que, suponemos, influye sobre los valores de la variable dependiente. Puesto que, para la realización del

Experimento, se le asignarán diferentes valores, hay que asegurarse que esté en nuestro poder manipularla.

Paso 3: Establecer los tratamientos

En base a la naturaleza de la variable, las condiciones reales del proceso o situación y la pregunta específica que se quiere contestar, se identificarán los valores o el recorrido de valores de la variable independiente, relevantes para el Experimento y se establecerán los tratamientos a efectuar.

Paso 4: Decidir el número de repeticiones para cada tratamiento

Es absolutamente aconsejable realizar varias observaciones para cada nivel de tratamiento (condición experimental), para que los errores de medida e influencias no controladas de variables extrañas puedan contrarrestarse entre sí.

Paso 5: Definir los sujetos sobre los que se va a realizar la medida

Paso 6: Determinar la variable dependiente

Sólo puede existir una única variable dependiente. Esta deberá tener, necesariamente, un nivel de medida continuo, o lo más próximo a ese extremo que sea posible. Cuantas más posibilidades de apreciar diferencias entre distintas observaciones ofrezca la variable dependiente, más se favorecerá la sensibilidad de la misma a los distintos tratamientos.

Paso 7: Explicitar el procedimiento de aleatorización

Esta es una parte muy importante del Diseño, ya que asegurará que las diferencias que se encuentren entre los tratamientos son debidas a ellos mismos y no a efectos laterales no deseados. Un procedimiento muy fácil y común para asignar los sujetos a los diferentes tratamientos al azar es el siguiente:

Numerar los sujetos y hacer para cada sujeto un papel con el número correspondiente. Poner todos los papeles en una bolsa y mezclarlos. Extraer, para cada tratamiento, tantos papeles "a ciegas", cuantas son las observaciones planificadas para el mismo.

Paso 8: Identificar posibles factores de "ruido" y/o variables de bloqueo

Analizar la futura situación experimental e identificar los factores que puedan, además de la variable independiente, influir sobre los valores de la variable dependiente. Según su relevancia y las posibilidades técnicas, la decisión a adoptar respecto a los factores de ruido y/o variables de bloqueo, será la siguiente:

- O bien tenerlos bajo control (constantes), a lo largo de todas las observaciones.
- O bien integrarlos en el Diseño, como variable de bloqueo.
- O bien transformarlos en una variable independiente. Esto será necesario, aunque se complique notablemente el Diseño de Experimento desde el punto de vista estadístico, cuando su influencia sobre la variable dependiente resulte ser relevante.
- O bien se espera que la aleatorización sea suficiente para que sus efectos se contrarresten en las repeticiones de cada tratamiento. Esta posibilidad será aceptable sólo si la variable en cuestión está fuera de nuestro control y se considera que su influencia es bastante limitada. Su efecto se englobará dentro del "error experimental", o "ruido".

Paso 9: Asegurarse de la aptitud del Diseño del Experimento para contestar la pregunta inicial

Comprobar que el tipo de resultados que se obtienen del Experimento tal y como se ha planificado, proporcione efectivamente la información que se necesita. El resultado de este paso será el Diseño, propiamente dicho, del Experimento.

Paso 10: Realización del Experimento

Se crearán las condiciones experimentales (tratamientos) y se efectuarán las observaciones según el plan establecido, teniendo un cuidado particular en evitar posibles influencias extrañas sobre los valores de la variable dependiente. Se registrarán los resultados del Experimento, anotando además toda la información posiblemente relevante sobre las circunstancias prácticas de cada observación. (Persona que realiza la medición, hora, temperatura, etc.).

5.2 ANÁLISIS DE VARIANZA A UNA VÍA

Hay varias formas en las cuales se puede diseñar un experimento ANOVA. Quizá el más común es el diseño completamente aleatorio o ANOVA a una vía. El término proviene del hecho que varios sujetos o unidades experimentales se asignan aleatoriamente a diferentes niveles de un solo factor. En donde la gran media de todas las observaciones “n” del experimento se calculan mediante:

$$\bar{x} = \frac{\sum x_{ij}}{n} \quad (5- 1)$$

El análisis de varianza se basa en una comparación de la cantidad de variación en cada uno de los tratamientos. Si de un tratamiento al otro la variación es significativamente alta, puede concluirse que los tratamientos tienen efectos diferentes en las poblaciones. Se pueden identificar tres tipos o fuentes de variación: la primera es la suma de las otras dos, y se llama variación total (existe variación entre el número total de observaciones); la segunda se denomina variación entre muestras (existe variación entre tratamientos); y la tercera variación se denomina variación dentro de la muestra (existe variación dentro de un tratamiento dado). Al comparar estas fuentes de variación se puede utilizar el análisis de varianza para probar la igualdad de las medias de poblaciones diversas.

5.2.1 La suma de cuadrados. El reconocimiento de estas tres fuentes de variación permite la división de la suma de los cuadrados, un procedimiento que es necesario para el análisis de varianza⁴⁵. Cada uno de los tres tipos de variación produce una suma de cuadrados. Existe: 1) la suma de cuadrados total ($S.C_{Tot}$), 2) la suma de cuadrados de los tratamientos ($S.C_{Trat}$), y 3) la suma de cuadrados del error ($S.C_E$). Como era de esperarse.

$$S.C_{Tot} = S.C_{Trat} + S.C_E \quad (5 - 2)$$

Esto ilustra que ($S.C_{Tot}$) puede dividirse en sus dos componentes: ($S.C_{Trat}$) y ($S.C_E$). Se pueden utilizar estas sumas de cuadrados para probar la igualdad de las medias poblacionales. Vale la pena recordar que la varianza muestral se calcula como:

$$s^2 = \frac{(x_i - \bar{x})^2}{n - 1} \quad (5 - 3)$$

El numerador es la suma de los cuadrados de las desviaciones de la media. De esta forma, la suma de los cuadrados se utiliza para medir la variación. El denominador es el número de grados de libertad. Esta ecuación sirve como patrón que puede aplicarse a la suma de cuadrados en el análisis de varianza.

Sea X_{ij} la observación i ésima en la muestra j ésima. Entonces la siguiente ecuación representa la suma de cuadrados total.

⁴⁵John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

$$S.C_{Tot} = \sum_{i=1}^r \sum_{j=1}^c (x_{ij} - \bar{x})^2 \quad (5 - 4)$$

La gran media se le resta a cada una de las observaciones. Las diferencias se elevan al cuadrado y se suman. Como muestra el signo de doble sumatoria en la ecuación, esto se hace a través de las filas y a través de todas las columnas. Debería notarse que $S.C_{Tot}$ es simplemente la variación de las observaciones alrededor de la gran media. Para la suma de cuadrados de los tratamientos se tiene que:

$$S.C_{Trat} = \sum_j r_j (\bar{x}_j - \bar{x})^2 \quad (5 - 5)$$

El número de observaciones o filas en cada tratamiento, r_j , se multiplica por las diferencias cuadradas entre la media de cada tratamiento y la gran media. Los resultados se suman para todos los tratamientos. La ecuación pide que se multiplique el número de filas en la j ésima columna por la desviación de la media elevada al cuadrado de dicha columna de la gran media. La $S.C_{Trat}$ refleja la variación en las medias de la columna alrededor de la gran media. La suma de cuadrados del error se expresa como:

$$S.C_E = \sum_{i,j} (x_{ij} - \bar{x}_j)^2 \quad (5 - 6)$$

$S.C_E$ mide la variación aleatoria de los valores dentro de un tratamiento alrededor de su propia media.

5.2.2 Cuadrados medios: La ecuación para la varianza muestral dice que después de obtener la suma de los cuadrados, cada una se divide por los grados de libertad. Una suma de cuadrados dividida por sus grados de libertad produce un cuadrado medio. Es decir, si se divide una suma de cuadrados por sus grados de libertad, se obtiene un cuadrado medio. En donde los grados de libertad son como el número total de observaciones del conjunto de datos menos toda “restricción” que pueda aplicarse. Una restricción fue todo valor que se calcula a partir del conjunto de datos.

En este aspecto, se nota que al calcular $S.C_{Tot}$, se utilizó todo el conjunto de datos de “n” observaciones para calcular un valor. Ese valor único era la gran media, la cual representa una restricción. Por tanto $S.C_{Tot}$ tiene $n - 1$ grados de libertad. El cálculo de $S.C_{Trat}$ involucra el uso de c medias muestrales de las cuales se puede calcular la gran media. Las medias muestrales por tanto se ven como puntos de datos individuales y la gran media se toma como restricción. $S.C_{Trat}$ tiene entonces $c - 1$ grados de libertad. Finalmente se calcula $S.C_E$ sumando la desviación de “n” observaciones de “c” medias muestrales. Por tanto, $S.C_E$ tiene $n - c$ grados de libertad. Se tiene que:

$$g.l \text{ para } S.C_{Tot} = g.l \text{ para } S.C_{Trat} + g.l \text{ para } S.C_E \quad (5 - 7)$$

$$n - 1 = c - 1 + n - c$$

Se halla entonces la media total de los cuadrados, o cuadrado medio total ($C.M_{Tot}$) como:

$$C.M_{Tot} = \frac{S.C_{Tot}}{n - 1} \quad (5 - 8)$$

El cuadrado medio del tratamiento ($C.M_{trat}$) es

$$C.M_{Trat} = \frac{S.C_{Trat}}{C - 1} \quad (5 - 9)$$

Y el cuadrado medio del error ($C.M_E$) es:

$$C.M_E = \frac{S.C_E}{n - c} \quad (5 - 10)$$

Estos tres cuadrados medios están moldeados a partir de la ecuación para la varianza muestral. Son sumas de los cuadrados divididos por sus grados de libertad, y como tales son varianzas. Es la razón de las dos últimas $C.M_{trat}$ y $C.M_E$, la que se utiliza como base del análisis de varianza para probar la hipótesis respecto a la igualdad de las medias. Como se pudo observar anteriormente, esta razón se ajusta a la distribución F, y se expresa como:

$$F = \frac{C.M_{Trat}}{C.M_E} \quad (5 - 11)$$

$C.M_{Trat}$ mide la variación entre tratamientos. Si los tratamientos tienen efectos diferentes, $C.M_{Trat}$ lo reflejara a través de su incremento. Entonces, la razón F en si misma se incrementa. Por tanto, si la razón F se vuelve significativamente grande es porque $C.M_{Trat}$ excede a $C.M_E$ en una cantidad grande, y se reconoce que los efectos del tratamiento probablemente existen. Es probable que tratamientos diferentes tengan efectos diferentes en las medias de sus poblaciones respectivas, y podrá rechazarse la hipótesis nula $\mu_1 = \mu_2 = \mu_3$. El valor crítico de F que es considerado significativamente grande puede encontrarse en tablas.

5.3 PRUEBAS PARA LA DIFERENCIA ENTRE PARES DE MEDIAS

Como se explicó anteriormente, el análisis de varianza dice si todas las medias son iguales. Sin embargo, cuando se rechaza la hipótesis nula, el análisis de varianza no revela cual(es) media(s) es (son) diferente(s) del resto. Se deben utilizar otras pruebas estadísticas para tomar esta determinación. Estas pruebas consisten en una comparación por pares, de todos los pares de medias posibles. Si el valor absoluto (ignorando los signos) de la diferencia entre dos medias muestrales cualquiera es mayor que algún estándar, se observa como una diferencia significativa, y se concluye que las medias poblacionales respectivas son diferentes. Se puede determinar este estándar debido a una diversidad de procedimientos estadísticos incluyendo el método de Tukey (Too' Key) y la diferencia mínima significativa (DMS).

5.3.1 Pruebas para diseños balanceados. Tanto el método de Tukey como el primero de los dos métodos DMS que aparecen aquí, se utilizan si existe igual número de observaciones en cada muestra. Se dice que tales diseños del análisis de varianza son balanceados. Si diseño no está balanceado porque las muestras son de diferentes tamaños, debe utilizarse un método DMS alternativo. El método Tukey requiere el cálculo del criterio de Tukey, T, como aparece en la ecuación:

$$T = q_{\alpha, c, n-c} \frac{\overline{C \cdot M_E}}{r} \quad (5 - 12)$$

En donde “q” tiene una distribución de rangos con c y n – c grados de libertad y α es el valor seleccionado. Vale la pena recordar que “c” es el número de muestras o tratamientos (columnas), y “n” es el número total de observaciones en todas las muestras combinadas.

El criterio estándar de Tukey⁴⁶ (T), se compara con la diferencia absoluta entre cada par de medias muestrales. Si cualquier par de medias muestrales tiene una diferencia absoluta mayor que el valor de T, se puede concluir que sus medias poblacionales respectivas no son iguales. La diferencia entre las medias muestrales es demasiado grande como para concluir que viene de poblaciones similares. Estos resultados pueden resumirse mediante el subrayado común en el cual las líneas que conectan las medias muestrales que éstas no difieren significativamente. Las medias muestrales primero deben ponerse en una serie ordenada, generalmente del más bajo al más alto.

El método de diferencia mínima significativa es muy similar al método de Tukey. Compara el criterio de la diferencia menos significativa con la diferencia absoluta en las medias muestrales. Si el diseño está balanceado, el criterio DMS es:

$$DMS = \frac{2(C \cdot M_E) F_{\alpha, 1, n-c}}{r} \quad (5 - 13)$$

Vale la pena destacar que al utilizar el método DMS, F tiene 1 y n – c grados de libertad. De igual forma que con el método Tukey, se compara el DMS con las diferencias absolutas de las medias muestrales. El método DMS es más conservador en que, dado un conjunto de condiciones cualquiera, el criterio DMS será menor que el valor de Tukey. Con cualquiera de los métodos, pueden surgir inconsistencias.

5.3.2 Pruebas para diseños no balanceados. Si el diseño es no balanceado, el método de Tukey y el método DMS, discutidos anteriormente, simplemente no se

⁴⁶ Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia

aplican, en su lugar, se puede utilizar un método DMS alternativo. Para comparar las muestras j – esima y K – esima, la ecuación para DMS se vuelve:

$$DMS_{j,k} = \frac{1}{r_j} + \frac{1}{r_k} * C.M_E F_{\alpha,c-1,n-c} \quad (5 - 14)$$

En donde el r_j es el número de observaciones en la muestra j – esima y r_k es el número de observaciones en la muestra k – esima. El valor de DMS será diferente para cada par de comparaciones por pares, debido a que el número de observaciones no es el mismo en cada muestra.

5.4 ANÁLISIS DE VARIANZA A DOS VÍAS

Con el análisis de varianza a una vía, se pensó que solo un factor influenciaba las unidades experimentales. Sin embargo, con frecuencia se encuentra que una segunda influencia exterior puede impactar las unidades experimentales. Para obtener una medida decisiva, se debe “bloquear” el factor externo, colocando las observaciones en grupos homogéneos. Así, las observaciones se clasifican tanto por bloques como por tratamientos. El propósito del bloque es reducir la variación dentro de un tipo de tratamiento. Este diseño experimental se llama diseño experimental aleatorio en bloques⁴⁷.

Si los bloques se realizan de manera efectiva y se basan en un factor que verdaderamente afecte la productividad, se obtiene una media más pura del efecto del tratamiento. Sin embargo, si el factor seleccionado para el bloque no afecta la productividad, los resultados pueden ser engañosos. Es importante determinar si

⁴⁷John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

el bloque se hace o no correctamente, y si el factor en el que se basa el bloque si tiene cierto impacto.

Con el análisis de varianza a dos vías, la suma de los cuadrados total se divide en tres partes: la suma de cuadrados de tratamiento ($S.C_{Trat}$), suma de cuadrados del error, y la suma de cuadrados de bloques ($S.C_{Bloq}$), por tanto

$$S.C_{Tot} = S.C_{Trat} + S.C_E + S.C_{Bloq} \quad (5 - 15)$$

$S.C_{Tot}$ y $S.C_{Trat}$ se calculan de la misma forma que en el análisis de varianza de una vía, sin embargo $S.C_E$ se subdivide en una media para $S.C_E$ y $S.C_{Bloq}$, en donde la suma de cuadrados de bloques es:

$$S.C_{Bloq} = \sum C_i(x_i - \bar{x})^2 \quad (5 - 16)$$

El número de tratamientos en cada bloque, c_i , se multiplica por la diferencia al cuadrado entre la media para cada bloque (fila), y la gran media. Los resultados se suman para todos los bloques. El símbolo c_i se utiliza para indicar el número de tratamientos en un bloque (fila). La suma de los cuadrados del bloque mide el grado de variación de las medias del bloque (fila) alrededor de la gran media. La suma de cuadrados del error $S.C_E$, se calcula como

$$S.C_E = S.C_{Tot} - S.C_{Trat} - S.C_{Bloq} \quad (5 - 17)$$

En donde $S.C_{Tot}$ y $S.C_{Trat}$, son:

$$S.C_{Tot} = \sum_{i=1}^r \sum_{j=1}^c (x_{ij} - \bar{x})^2 \quad (5-18)$$

$$S.C_{Trat} = \sum_j r_j (x_j - \bar{x})^2 \quad (5-19)$$

En donde hay r bloques y c tratamientos, y hay $n = rc$ observaciones. Los grados de libertad para cada una de las sumas de cuadrados de los valores de la ecuación son:

$$S.C_E = S.C_{Tot} - S.C_{Trat} - S.C_{Bloq} \quad 5-20$$

$$r - 1 \quad c - 1 \quad = \quad n - 1 \quad - \quad c - 1 \quad - \quad (r - 1)$$

El cuadrado medio total y el cuadrado medio del tratamiento son, al igual que antes, la suma de sus cuadrados dividida por sus grados de libertad, por tanto el cuadrado medio total es:

$$C.M_{Tot} = \frac{S.C_{Tot}}{n - 1} \quad (5-21)$$

El cuadrado medio del tratamiento (CMTR) es:

$$C.M_{Trat} = \frac{S.C_{Trat}}{c - 1} \quad (5-22)$$

Y en el análisis de varianza a dos vías, el cuadrado medio del error (CME) es:

$$C.M_E = \frac{S.C_E}{r - 1 (c - 1)} \quad (5 - 23)$$

Y el cuadrado medio del bloque se halla mediante:

$$C.M_{bloq} = \frac{S.C_{Bloq}}{r - 1} \quad (5 - 24)$$

Los valores F se calculan de la misma manera que en el análisis de varianza a una vía:

$$F = \frac{C.M_{Trat}}{C.M_E} \quad (5- 25)$$

Vale la pena notar que se calculan dos valores F – uno que utiliza C.M_{Trat} y uno que utiliza C.M_{Bloq}. El valor F para C.M_{Bloq} se calcula para determinar si los bloques se realizaron de manera efectiva. Si el bloqueo se basa en un factor externo que no afecta, los resultados pueden ser engañosos. Por lo tanto, se deben hacer pruebas para ver si existe una diferencia significativa entre las medias del bloque (fila). Si no existe diferencia significativa entre las medias de los bloques, entonces la experiencia no es un factor crítico. En este caso, se debería abandonar el análisis de varianza a dos vías, y se necesitaría regresar al análisis de varianza a una sola vía.

5.5 DISEÑOS FACTORIALES

Muchos experimentos requieren el estudio de dos o más factores. Se puede mostrar que, en general, los experimentos factoriales son diseños más eficientes para este tipo de análisis. Se entiende por experimento factorial aquel en el cual cada ensayo completo o replicación del experimento todas las posibles combinaciones de los niveles de los factores son observadas. Por ejemplo si hay a niveles del factor A y b niveles del factor B, cada replicación contiene todas las ab combinaciones del tratamiento. Cuando los factores son arreglados en un experimento factorial se dice también que están cruzados.

El efecto de un factor se define como el cambio en la respuesta producida por un cambio en el nivel del factor. Esto se llama frecuentemente un efecto principal puesto que se refiere a los factores primarios de interés en el experimento.

5.5.1 Interacción entre dos factores. En algunos experimentos la diferencia en la respuesta entre niveles de un factor no es la misma en todos los niveles de los otros factores. Considere los datos siguientes:

Tabla 27. Niveles de interacción entre dos factores

		B	
A		B_1	B_2
A_1		50	20
A_2		30	42

En el nivel 1 de B el efecto de A es: $30 - 50 = -20$. Pero en el nivel 2 de B el efecto de A es: $42 - 20 = 22$. Puesto que el efecto de A depende del nivel escogido para el factor B, se ve que hay una interacción entre A y B.

Cuando la interacción es grande, el cálculo del efecto principal de un factor no se puede efectuar de la forma que se hizo anteriormente, no tendría significado el resultado. El resultado tiende a hacer creer que no hay efecto de A. Sin embargo, si se calcula este efecto para cada nivel de B, el resultado es bien diferente. A tiene efecto, pero depende del nivel del factor de B.

5.5.2 Diseños factoriales para dos factores. Uno de los diseños de experimentos factoriales más sencillo involucra solo dos factores o conjuntos de tratamientos. El factor A tiene a niveles y el factor B tiene b niveles y estos son arreglados en un diseño factorial, es decir, cada replicación del experimento contiene todas las ab combinaciones. Asume que hay n repeticiones del experimento y sea Y_{ijk} la notación para la observación tomada bajo el nivel i del factor A, el nivel j para el factor B, en la k – esima replicación. El orden de experimentación de las abn mediciones es seleccionado al azar, de modo que el diseño es completamente aleatorio. Cada observación se puede describir mediante el modelo lineal estadístico:

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + E_{ijK} \quad (5 - 26)$$

i = 1,2,..., a

j = 1,2,..., b

K = 1,2,..., n

Donde μ es la media general, A_i es el efecto del nivel i del factor A, B_j es el efecto del nivel j del factor B, $(AB)_{ij}$ es el efecto de la interacción entre A_i y B_j , E_{ij} es el componente del error aleatorio. Inicialmente ambos factores se asumen fijos y los efectos de los tratamientos se definen como desviaciones de la media general, es decir:

$$A_i = \mu_{i..} - \mu_{..} \quad y \quad B_j = \mu_{.j.} - \mu_{..} \quad (5 - 27)$$

De modo que $A_i = 0$, y, $B_j = 0$. De modo similar los efectos de la interacción se definen como la desviación.

$$(\mu_{ij.} - \mu) - (\mu_{i..} - \mu) - (\mu_{.j.} - \mu) = \mu_{ij.} - \mu_{i.} - \mu_{.j.} + \mu \quad (5 - 28)$$

Entonces para el modelo de efecto fijo:

$$\sum_{i=1}^a \sum_{j=1}^b (AB)_{ij} = 0 \quad (5 - 29)$$

El procedimiento apropiado para probar las hipótesis es el análisis de varianza.

5.5.3 Análisis estadístico para modelo de efectos fijos. Sea Y_{ij} para denotar el total de las observaciones en la combinación ij o celda ij . Y_i denota el total de las observaciones bajo el nivel i del factor A. Y_j denota el total de las observaciones bajo el nivel j del factor B y $Y_{...}$ denota el total de todas las observaciones. Se denotan por $Y_{ij}, Y_{i.}, Y_{.j}, Y_{...}$ los promedios de las celdas, el nivel i del factor A, el nivel j del factor B y el total de todas las observaciones respectivamente. Las anteriores definiciones se expresan matemáticamente como:

$$Y_{ij} = \sum_{K=1}^n Y_{ijk} \quad Y_{ij} = \frac{Y_{ij.}}{n}$$

$$Y_{i..} = \sum_j Y_{ij} = \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} , \quad Y_{i..} = \frac{Y_{i..}}{bn}$$

$$\begin{aligned}
Y_{.j.} &= \sum_{i=1}^a Y_{ij.} = \sum_{i=1}^a \sum_{k=1}^n Y_{ijk}, & Y_{.j.} &= \frac{Y_{.j.}}{an} \\
Y_{...} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}, & Y_{...} &= \frac{Y_{...}}{abn} \quad (5-30)
\end{aligned}$$

La suma de cuadrados se puede expresar como:

$$S.C_{Tot} = S.C_A + S.C_B + S.C_{AB} + S.C_E \quad (5-31)$$

Se observa que para poder tener información sobre la $S.C_E$ debe haber más de una replicación ya que si, $n = 1$, $Y_{ijk} = Y_{ij}$. El número de grados de libertad de cada tratamiento son, para el efecto A los grados de libertad son $(a - 1)$, para el efecto B los grados de libertad son $(b - 1)$, para el efecto AB los grados de libertad son $(a - 1)(b - 1)$ y los grados de libertad causados por el error son $ab(a - 1)$. Los grados de libertad de la interacción se determinan al restar a los grados de libertad de las celdas los grados de libertad del efecto de A y del efecto B:

$$(ab - 1) - (a - 1) - (b - 1) = (a - 1)(b - 1)$$

Las esperanzas de los cuadrados medios están dadas por:

$$E C.M_A = \sigma^2 + \frac{bn}{(a - 1)} A_i^2$$

$$E C.M_B = \sigma^2 + \frac{an}{(b - 1)} B_j^2$$

$$E C.M_{AB} = \sigma^2 + \frac{a}{(a-1)(b-1)} \sum_{ij} (AB)_{ij}^2$$

$$E C.M_E = \sigma^2 \quad (5 - 32)$$

Las hipótesis a probar son:

1. $H_0: A_i = 0 \quad i = 1, 2, \dots, a$
 $H_A: \text{Al menos uno es diferente de cero}$
2. $H_0: B_j = 0 \quad j = 1, 2, \dots, b$
 $H_A: \text{Al menos uno es diferente de cero}$
3. $H_0: (AB)_{ij} = 0 \quad i = 1, 2, \dots, a \quad \text{y} \quad j = 1, 2, \dots, b$
 $H_A: \text{Al menos uno es diferente de cero}$

Para validar o rechazar estas hipótesis se usan los estadísticos, respectivamente.

$$F = \frac{C.M_A}{C.M_E}$$

$$F = \frac{C.M_B}{C.M_E}$$

$$F = \frac{C.M_{AB}}{C.M_E} \quad (5 - 33)$$

Si los errores son normales y las hipótesis son ciertas los estadísticos tienen distribuciones F con $(a - 1)$ y $ab(n - 1)$, $(b - 1)$ y $ab(n - 1)$, $(a - 1)(b - 1)$ y $ab(n - 1)$ grados de libertad respectivamente. Por lo tanto, las reglas de decisión para cada hipótesis serán, respectivamente:

Rechazar H: a un nivel significativo α , si F es mayor que $F_{\alpha, (a-1), ab(n-1)}$, $F_{\alpha, (b-1), ab(n-1)}$, $F_{\alpha, (a-1)(b-1), ab(n-1)}$. Las formulas computacionales para las sumas de cuadrados están dadas por:

$$S.C_{Tot} = \sum Y_{ijk}^2 - F.C \quad (5-34)$$

$$S.C_A = \frac{1}{bn} \sum_{i=1}^a Y_{i..}^2 - F.C \quad (5-35)$$

$$S.C_B = \frac{1}{an} \sum_{j=1}^b Y_{.j.}^2 - F.C \quad (5-36)$$

Para calcular la suma de cuadrados de la interacción, se obtiene primero la suma de cuadrados de las celdas y a esta se le resta las sumas de cuadrados de A y B.

$$S.C_{celdas} = \sum_i^a \sum_j^b \frac{Y_{ij.}^2}{n} - F.C \quad (5-37)$$

$$S.C_{AB} = S.C_{celdas} - S.C_A - S.C_B \quad (5-38)$$

$$F.C = \frac{Y^2 \dots}{abn} \quad 5-39$$

La suma de cuadrados del error se calcula por diferencia.

$$S.C_E = S.C_{Tot} - S.C_{celdas} \quad (5-40)$$

El primer estadístico F de la ecuación (5 – 33) se usa para verificar las hipótesis:

$$H_0: A_i = 0 \quad i = 1, 2, \dots, a$$

H_A : Al menos uno es diferente de cero

El segundo estadístico F de la misma ecuación se utiliza para probar el efecto principal del factor B, y el tercer estadístico F se usa para verificar la significancia de la interacción.

5.5.4 Modelo de efectos aleatorios. Considérese la situación en la cual los niveles del factor A y B son escogidos al azar de una población de niveles. Este es el modelo de efectos aleatorios o componentes de varianza. Puesto que los a niveles de A y los b niveles de B fueron escogidos aleatoriamente, las inferencias serán válidas sobre todos los niveles en la población bajo estudio. Las observaciones se pueden representar por el modelo lineal estadístico.

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + E_{ijk} \quad (5 - 41)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$K = 1, 2, \dots, n$$

Donde A_i , B_j , $(AB)_{ij}$ y E_{ijk} son variables aleatorias. La varianza de una observación es:

$$V(Y_{ijk}) = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2 \quad (5 - 42)$$

Y σ^2_A , σ^2_B , σ^2_{AB} y σ^2 se llaman los componentes de varianza, en donde las hipótesis que se deben probar son:

$$H_0: \sigma^2_A = 0, \quad H_0: \sigma^2_B = 0, \quad \text{y} \quad H_0: \sigma^2_{AB} = 0,$$

$$S.C_{Tot} = S.C_A + S.C_B + S.C_{AB} + S.C_E \quad (5 - 43)$$

Los grados de libertad son los mismos que en el modelo de efecto fijo. Sin embargo, las esperanzas de los C.M son diferentes:

$$E(C.M_A) = \sigma^2 + n\sigma_{AB}^2 + bn\sigma_A^2$$

$$E(C.M_B) = \sigma^2 + n\sigma_{AB}^2 + an\sigma_B^2$$

$$E(C.M_{AB}) = \sigma^2 + n\sigma_{AB}^2$$

$$E C.M_E = \sigma^2 \quad (5 - 44)$$

De las esperanzas de los C.M se observa que el test estadístico apropiado para probar H: $\sigma^2_{AB} = 0$ contra A: $\sigma^2_{AB} > 0$ es:

$$F = \frac{C.M_{AB}}{C.M_E} \quad (5 - 45)$$

De manera semejante para probar H: $\sigma^2_A = 0$ contra A: $\sigma^2_A > 0$ es:

$$F = \frac{C.M_A}{C.M_{AB}} \quad (5 - 46)$$

Para probar la $H: \sigma_B^2 = 0$ contra $A: \sigma_B^2 > 0$ el estadístico es:

$$F = \frac{C \cdot M_B}{C \cdot M_{AB}} \quad (5 - 47)$$

Las esperanzas de los C.M siempre son usadas como guías para la construcción de los test estadísticos. Los componentes de varianza pueden estimarse igualando los C.M observados a los valores esperados y resolviendo. Esto da:

$$\begin{aligned} \sigma^2 &= C \cdot M_E \\ \sigma_{AB}^2 &= \frac{C \cdot M_{AB} - C \cdot M_E}{n} \\ \sigma_B^2 &= \frac{C \cdot M_B - C \cdot M_{AB}}{an} \\ \sigma_A^2 &= \frac{C \cdot M_A - C \cdot M_{AB}}{bn} \quad (5 - 48) \end{aligned}$$

5.5.5 Modelos mixtos. La situación es la misma que en los dos modelos anteriores, es decir, hay dos factores A y B. Aquí se considera sin embargo que los niveles de A son fijos (a) y los niveles de B son aleatorios (b). El modelo lineal estadístico es el mismo que para el modelo de efectos aleatorios. Las esperanzas en este modelo se expresan como:

$$\begin{aligned} E C \cdot M_A &= \sigma^2 + n\sigma_{AB}^2 + \frac{bn}{(a-1)} A_i^2 \\ E(C \cdot M_B) &= \sigma^2 + an\sigma_B^2 \\ E(C \cdot M_{AB}) &= \sigma^2 + n\sigma_{AB}^2 \\ E C \cdot M_E &= \sigma^2 \quad (5 - 49) \end{aligned}$$

Por lo tanto, el test apropiado para validar $H: A_i = 0$ contra $A: A_i$ diferente de cero es:

$$F = \frac{C \cdot M_A}{C \cdot M_{AB}} \quad (5 - 50)$$

El cual se distribuye como una $F(a - 1, (a - 1)(b - 1))$ si $H: A_i = 0$, es cierta. Para probar la $H: \sigma^2_B = 0$ contra $A: \sigma^2_B > 0$ el estadístico es:

$$F = \frac{C \cdot M_B}{C \cdot M_E} \quad (5 - 51)$$

El cual se distribuye como una $F(b - 1, ab(n - 1))$ si $H: \sigma^2_B = 0$, es cierta. Finalmente Para probar la $H_0: \sigma^2_{AB} = 0$ contra $A: \sigma^2_{AB} > 0$ el estadístico es:

$$F = \frac{C \cdot M_{AB}}{C \cdot M_E} \quad (5 - 52)$$

El cual se distribuye como una $F(a - 1)(b - 1, ab(n - 1))$. En el modelo mixto es posible estimar los efectos del factor fijo como:

$$\mu = Y \dots \quad i = 1, 2, \dots a$$

$$A_i = Y_{i..} - Y \dots \quad (5 - 53)$$

Los componentes de varianza, se pueden estimar como en el modelo anterior.

$$\sigma^2 = C \cdot M_E$$

$$\sigma_{AB}^2 = \frac{C \cdot M_{AB} - C \cdot M_E}{n}$$

$$\sigma_B^2 = \frac{C \cdot M_B - C \cdot M_E}{an} \quad (5 - 54)$$

En los modelos mixtos es posible probar la hipótesis acerca de las medias individuales para el factor fijo. Al usar el test del rango múltiple de Duncan, el error estándar de la media del tratamiento del efecto fijo está dado por:

$$S_{YA} = \frac{\overline{C \cdot M_A}}{bn} \quad (5 - 55)$$

5.5.6 El experimento factorial generalizado. Los resultados para el análisis de varianza para dos métodos de clasificación puede extenderse al caso general donde hay (a) niveles del factor A, b niveles del factor B, c niveles del factor C y así sucesivamente, arreglados en un experimento factorial. En general, habrá abc... n observaciones en total, si hay n replicaciones del experimento completo. Nuevamente, obsérvese que se debe tener al menos dos replicaciones ($n \geq 2$) para poder determinar la S.C debido al error aleatorio si todas las interacciones son incluidas en el modelo.

Si todos los factores en el experimento son fijos se puede formular y probar fácilmente hipótesis acerca de los efectos principales interacciones. Para modelos de efectos fijos, los test para probar las hipótesis sobre cada efecto principal y las interacciones se pueden construir dividiendo el C.M correspondiente del efecto o la interacción por el C.M_E. Los grados de libertad para cada efecto principal son los niveles del factor menos uno, por ejemplo (a - 1), y el número de grados de libertad para cada interacción es el producto del número de grados de libertad

asociado con los componentes individuales de esta. Por ejemplo para AB, serán $(a - 1)(b - 1)$, para ABC, serán $(a - 1)(b - 1)(c - 1)$, y así con las demás interacciones. Por ejemplo considérese el análisis de varianza el modelo de tres factores.

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + C_k + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + E_{ijkl} \quad (5 - 56)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, c$$

$$l = 1, 2, \dots, n$$

Las ecuaciones de cálculo para las S.C están dadas por:

$$S.C_{Tot} = \sum_{ijkl} Y_{ijkl}^2 - F.C \quad (5 - 57)$$

Para los efectos principales:

$$S.C_A = \frac{1}{bcn} \sum_{i=1}^a Y_{i..}^2 - F.C \quad (5 - 58)$$

$$S.C_B = \frac{1}{acn} \sum_{j=1}^b Y_{.j.}^2 - F.C \quad (5 - 59)$$

$$S.C_C = \frac{1}{abn} \sum_{j=1}^b Y_{..K}^2 - F.C \quad (5-60)$$

Para calcular las S.C de las interacciones de dos factores se calculan las S.C de las celdas correspondientes y se les resta las S.C de los dos efectos principales. Es de ayuda si los datos se reducen a tablas de dos factores para calcular las S.C de las celdas correspondientes. Las S.C se obtienen así:

$$S.C_{AB} = \sum_i^a \sum_j^b \frac{Y_{ij.}^2}{cn} - F.C - S.C_A - S.C_B$$

$$S.C_{AB} = S.C_{celdas} - S.C_A - S.C_B \quad (5-61)$$

$$S.C_{AC} = \sum_i^a \sum_K^c \frac{Y_{i.K.}^2}{bn} - F.C - S.C_A - S.C_C$$

$$S.C_{AC} = S.C_{celdas} - S.C_A - S.C_C \quad (5-62)$$

$$S.C_{BC} = \sum_j^b \sum_K^c \frac{Y_{.Kj.}^2}{cn} - F.C - S.C_B - S.C_C$$

$$S.C_{AB} = S.C_{celdas} - S.C_B - S.C_C \quad (5-63)$$

$$F.C = \frac{Y^2 \dots}{abcn} \quad (5-64)$$

Las S.C de la interacción ABC se calcula de los totales de las celdas (Y_{ijk}) como:

$$S.C_{ABC} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \frac{Y_{ijk}^2}{n} - F.C - S.C_A - S.C_B - S.C_C - S.C_{AB} - S.C_{AC} - S.C_{BC} \quad (5 - 65)$$

Las S.C_E se calcula por diferencia

$$S.C_E = S.C_{Tot} - S.C_{celdas\ abc} \quad (5 - 66)$$

5.6 DISEÑOS FACTORIALES 2^k

El diseño 2^k es muy útil en las primeras etapas de una investigación experimental cuando hay muchos posibles factores que pueden afectar la variable respuesta. Este diseño provee el menor número de combinaciones de tratamientos en el cual se pueden estudiar k factores en un arreglo completamente factorial. Se asume que la variación de la variable respuesta es lineal entre los dos niveles de cada factor. Los niveles se asumen fijos y el orden de experimentación es completamente aleatorio.

5.6.1 Diseño 2². Este diseño es el más elemental de la serie 2^k, es decir, un diseño donde solo hay dos factores, por ejemplo A y B, cada uno a dos niveles, este se denomina diseño factorial 2². Los niveles de los factores se denominan arbitrariamente “bajo” y “alto” y se suelen denotar en número como 0 y 1, respectivamente.

Por convención se denota el efecto principal de un factor con letras latinas mayúsculas. Así, A se refiere al efecto principal del factor A y, B se refiere al

efecto principal del factor B y, AB se refiere al efecto de la interacción de los factores A y B.

Los totales por combinación se pueden representar mediante los vértices de un cuadrado. En donde las coordenadas de los vértices del cuadrado representan las cuatro combinaciones de tratamientos, así: 00 que representan a los dos factores en su nivel bajo, 10 representan al nivel alto de A y el nivel bajo de B, 01 representa el nivel bajo de A y el nivel alto de B y 11 representan los niveles altos de ambos factores. Estas combinaciones usualmente se representan con letras minúsculas.

Se define el efecto promedio de un factor como el cambio en la respuesta producido por un cambio en el nivel del factor, promediado sobre los niveles del otro factor. También las combinaciones (1), a, b, ab, representan el total de todas las n replicaciones tomadas en estas combinaciones de tratamientos. El efecto de A en el nivel alto de B es $[ab - a]/ n$. promediando estas dos cantidades se obtiene:

$$A = \frac{1}{2n} \{ ab - b + a - 1 \}$$

$$A = \frac{1}{2n} - 1 + a - b + ab \quad (5 - 67)$$

El efecto medio de B se halla promediando el efecto de B en el nivel bajo de A, $[b - (1)]/ n$, con el efecto de B en el nivel alto de A, $[ab - a]/ n$.

$$B = \frac{1}{2n} \{ ab - a + b - 1 \}$$

$$B = \frac{1}{2n} \{ 1 - a + b + ab \} \quad (5 - 68)$$

Se define el efecto de la interacción AB como la diferencia media entre el efecto de A en el nivel alto de B y el efecto de A en el nivel bajo de B, así:

$$AB = \frac{1}{2n} \{ ab - b - a - 1 \}$$

$$AB = \frac{1}{2n} \{ 1 - a - b + ab \} \quad (5 - 69)$$

De manera similar se puede definir AB como la diferencia media entre el efecto de B en el nivel alto de A y el efecto de B en el nivel bajo de A:

$$AB = \frac{1}{2n} \{ ab - a - b - 1 \}$$

$$AB = \frac{1}{2n} \{ 1 - a - b + ab \} \quad (5 - 70)$$

Para determinar la suma de cuadrados de los efectos, se debe observar que las ecuaciones (5 - 67), (5 - 68) y (5 - 69) son contrastes puesto que $C_i = 0$. Más aun, estos contrastes son ortogonales entre si, como puede verificarse. La suma de cuadrados de un contraste, como se recordara, está dado por:

$$S.C_{C_m} = \frac{C_m^2}{n C_i^2} \quad (5 - 71)$$

Por lo tanto, las sumas de cuadrados de los efectos están dados por:

$$S.C_A = \frac{[-1 + a - b + ab]^2}{4n} \quad (5 - 72)$$

$$S.C_B = \frac{[-1 - a + b + ab]^2}{4n} \quad (5 - 73)$$

$$S.C_{AB} = \frac{[1 - a - b + ab]^2}{4n} \quad (5 - 74)$$

Las combinaciones de tratamiento se escriben en el orden denominado estándar: (1), a, b, ab. Al usar este orden, se ve que los coeficientes de los contrastes usados para estimar los efectos son:

	(1)	a	b	ab
A	-1	+1	-1	+1
B	-1	-1	+1	+1
AB	+1	-1	-1	+1

Los coeficientes del contraste para estimar el efecto de la interacción se obtiene multiplicando los coeficientes correspondientes para los dos efectos principales. Entonces para obtener los efectos basta multiplicar el total por combinación por el signo respectivo y dividir el resultado por 2n. Por ejemplo para calcular el efecto de A:

$$\frac{[-1 + a - b + ab]}{2n}$$

No sobra advertir que esta es una metodología especial para este diseño, pero los resultados de las sumas de cuadrados son idénticos si se calculan usando las ecuaciones dadas en la sección anterior.

5.6.2 Diseño factorial ²³. En este diseño se están estudiando tres factores A, B, y C cada uno a dos niveles. El diseño se denomina diseño factorial 2^3 y las ocho combinaciones de tratamiento se pueden representar por los vértices de un cubo. El eje x corresponde al factor A, y sobre este se marcan los dos niveles 0 y 1. El eje y corresponde al factor B y sobre este se marcan los niveles 0 y 1, el eje z corresponde al factor C y también se marcan los dos niveles 0 y 1. Para calcular los efectos principales y las interacciones se procede según se indica a continuación.

Para el efecto de A:

El efecto de A cuando B y C están en el nivel cero es $[a - (1)] / n$. El efecto de A cuando B está en el nivel uno y C en el nivel cero es $(ab - a) / n$. El efecto de A cuando B está en el nivel cero y C en el nivel uno es $(ac - c) / n$. Finalmente el efecto de A en el nivel uno de B y C es $(abc - bc) / n$. Entonces el efecto de A es el promedio de estos cuatro efectos parciales, es decir:

$$A = \frac{1}{4n} [a - 1 + ab - b + ac - c + abc - bc] \quad (5 - 75)$$

Esta ecuación se puede desarrollar como un contraste entre las cuatro combinaciones de tratamiento de la cara derecha del cubo (donde a está en el nivel alto) y las cuatro combinaciones de tratamiento de la cara izquierda del cubo (donde a está en el nivel bajo), combinando estas se obtiene:

$$A = \frac{1}{4n} [a + ab + ac + abc - 1 - b - c - bc]$$

Al escribir el resultado anterior en el orden estándar se obtiene:

$$A = \frac{1}{4n} - 1 + a - b + ab - c + ac - bc + abc \quad (5 - 76)$$

De manera similar, el efecto de B es un contraste entre las cuatro combinaciones⁴⁸ de la cara del fondo del cubo (donde B está en el nivel alto) y las cuatro combinaciones de la cara frontal (donde B está en el nivel bajo), esto resulta en:

$$B = \frac{1}{4n} b + ab + bc + abc - 1 - a - c - ac \quad (5 - 77)$$

Que está escrita en el orden estándar de:

$$B = \frac{1}{4n} - 1 - a + b + ab - c - ac + bc + abc \quad (5 - 78)$$

El efecto de C es un contraste entre las cuatro combinaciones de tratamiento en la cara superior del cubo y las cuatro de la cara baja, es decir:

$$C = \frac{1}{4n} c + ac + bc + abc - 1 - a - b - ab \quad (5 - 79)$$

O en el orden estándar:

$$C = \frac{1}{4n} - 1 - a - b - ab + c + ac + bc + abc \quad (5 - 80)$$

⁴⁸Humberto Gutiérrez pulido / Román de la Vara Salazar. Análisis y Diseño de Experimentos. Segunda Edición. Ed. McGraw – Hill. Interamericana S.A. México

Cuando c está en el nivel bajo el efecto de la interacción AB es la diferencia media⁴⁹ en el efecto de A en los niveles de B, es decir:

$$\frac{[ab - b - a - (1)]}{2n}$$

Cuando c está en el nivel alto, la interacción es:

$$\frac{[(abc - bc) - (ac - c)]}{2n}$$

AB es el promedio de estos dos:

$$AB = \frac{1}{4n} ab - b - a + 1 + abc - bc - ac + c \quad (5 - 81)$$

O en el orden estándar:

$$AB = \frac{1}{4n} 1 - a - b + ab + c - ac - bc + abc \quad (5 - 82)$$

Similarmente, los efectos medios de AC y BC son:

$$AC = \frac{1}{4n} 1 - a + b - ab + ac - bc + abc \quad (5 - 83)$$

$$BC = \frac{1}{4n} 1 + a - b - ab - c - ac + bc + abc \quad (5 - 84)$$

⁴⁹Humberto Gutiérrez pulido / Román de la Vara Salazar. Análisis y Diseño de Experimentos. Segunda Edición. Ed. McGraw – Hill. Interamericana S.A. México

El efecto de la interacción ABC se define como la diferencia promedio entre la interacción AB para los niveles de c, es decir:

$$ABC = \frac{1}{4n} abc - bc - ac - c - ab - b + (a - 1) \quad (5 - 85)$$

$$ABC = \frac{1}{4n} - 1 + a + b - ab + c - ac - bc + abc \quad (5 - 86)$$

Los signos para los efectos principales se determinan asociando un signo más con el nivel alto del factor y un signo menos con el nivel bajo. Una vez que se han establecido los signos para los efectos principales, los signos para las columnas restantes se pueden obtener multiplicando las columnas precedentes apropiadas, fila por fila. Por ejemplo, los signos en la columna AB son el producto de los signos de las columnas A y B en cada fila. El contraste para cualquier efecto se puede obtener fácilmente de la siguiente manera.

TABLA 28. Signos algebraicos para calcular los efectos en un diseño 2^3

	I	A	B	AB	C	AC	BC	ABC
(1).	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	+	+	+	+	+	+	+

Fuente: Wilfrido Moreno T. (1993). Aplicaciones al Diseño y Análisis de Experimentos. Bucaramanga: UIS. Departamento de Ingeniería Industrial.

La tabla 11 tiene varias propiedades. (1) Excepto para la columna I, cada columna tiene igual número de signos menos y más. (2) La suma de los productos de los signos de columnas es cero. (3) Al multiplicar la columna I cualquier número de veces por cualquier columna deja la columna sin cambio. Es decir, I es un elemento de identidad. (4) El producto de cualquier dos columnas da una columna en la tabla.

Las sumas de cuadrados para los efectos se determinan fácilmente puesto que cada efecto tiene un contraste correspondiente con un grado de libertad. En el diseño 2^3 con n replicaciones, la suma de cuadrados para cualquier efecto es.

$$S.C = \frac{(\text{contraste})^2}{8n} \quad (5 - 87)$$

5.6.3 Diseño General 2^K . Los métodos de análisis que fueron presentados pueden generalizarse al caso de un diseño factorial 2^k , es decir, un diseño con k factores cada uno a dos niveles. El modelo estadístico para un diseño 2^k incluirá k efectos principales, $\binom{k}{2}$ interacciones de dos factores, $\binom{k}{3}$ interacciones de tres factores, etc. Y una interacción de los K factores. La notación introducida antes para las combinaciones de tratamiento también se usa aquí, por ejemplo para un diseño de 2^5 , abd denota la combinación de tratamiento con los tratamientos A, B y D en el nivel alto y los factores C y E en el nivel bajo. Las combinaciones de tratamiento se pueden escribir en el orden estándar introduciendo los factores uno a la vez, cada nuevo factor se combina sucesivamente con los anteriores a este.

Para estimar un efecto o para calcular la suma de cuadrados para el efecto, primero se determina el contraste asociado con este efecto. Esto se puede hacer siempre usando una tabla de signos más y signos menos. Sin embargo, para valores de K mayores de 3 esto es engorroso y se prefiere un método más

eficiente. En general, se determina el contraste para el efecto $AB \dots_k$ expandiendo el lado derecho de:

$$\text{Contraste } AB \dots_k = a \pm 1 \quad b \pm 1 \quad \dots \quad K \pm 1 \quad (5 - 88)$$

Al explicar la anterior ecuación, se utiliza el álgebra ordinaria pero 1 se reemplaza por (1) en la expresión final. El signo en cada conjunto de paréntesis es negativo si el factor está incluido en el efecto y positivo si el factor no está incluido⁵⁰. Para ilustrar el uso de la ecuación (5 - 88), considere el diseño factorial 2^4 , el contraste para AB será.

$$AB = a - 1 \quad b - 1 \quad c + 1 \quad d + 1$$

$$AB = ab - a - b + 1 \quad (cd + c + d + 1)$$

$$\begin{aligned} AB &= abcd - abc - abd + ab - acd - ac - ad - a - bcd - bc \\ &\quad - bd - b + cd + c + d + 1 \\ &= 1 - 1 - b + ab + c - ac - bc + abc + d - ad - bd + abd \\ &\quad - cd - acd - abcd \end{aligned}$$

Una vez que se han calculado los contrastes para los efectos, se pueden estimar los efectos y calcular las sumas de cuadrados de acuerdo a:

$$AB \dots_k = \frac{2}{n2^k} \text{ contraste } AB \dots_k \quad (5 - 89)$$

$$S.C_{AB \dots_k} = \frac{(\text{contraste } AB \dots_k)^2}{n2^k} \quad (5 - 90)$$

⁵⁰Humberto Gutiérrez pulido / Román de la Vara Salazar. Análisis y Diseño de Experimentos. Segunda Edición. Ed. McGraw - Hill. Interamericana S.A. México

Respectivamente, donde n es el número de replicaciones.

5.6.4 Diseño factorial 2^k con una sola replicación. Aun para un número moderado de factores, el número total de combinaciones de tratamiento en un diseño factorial 2^k es grande. Por ejemplo, un diseño 2^5 tiene 32 combinaciones de tratamiento, uno de 2^6 tiene 64 combinaciones de tratamiento, y así sucesivamente. Debido a las limitaciones naturales de dinero, tiempo equipos, etc., el número de replicaciones que se pueden efectuar puede ser restringido, es frecuente que solo se pueda efectuar una replicación, a menos que se desea omitir algunos de los factores originales.

Con una sola replicación en un diseño 2^k no se puede calcular una estimación del error experimental, es decir, el cuadrado medio del error. Así que parece como si las hipótesis sobre los efectos principales y las interacciones no se pudieran probar. Sin embargo, el método usual para el análisis en un diseño factorial 2^k con una sola replicación es asumir que ciertas interacciones de alto orden no existen y entonces las esperanzas de sus cuadrados medios son iguales a σ^2 y se pueden combinar para estimar el error experimental.

Estas esperanzas de los cuadrados medios se deben seleccionar antes de efectuar el análisis de los datos. Usualmente, el diseño más pequeño para este procedimiento debe ser de uno de 2^4 factorial. El peligro que se corre al efectuar estas combinaciones de cuadrados medios es que algunas interacciones del cuadrado medio del error se aumenta. Como resultado de esto, otros efectos significantes no se pueden detectar.

Como regla general, es probable que sea poco razonable asumir que las interacciones de dos factores sean cero sin tener una información previa. Si la mayoría de las interacciones dobles son pequeñas es porque es muy probable que las interacciones de alto orden no sean significantes.

Sin embargo, en cada caso se debe usar el conocimiento del fenómeno bajo estudio y el sentido común para el análisis del diseño. Por ejemplo, si en el análisis de un diseño 2^5 los efectos principales de A, B y C como también AB y AC son grandes, entonces es probable que ABC sea grande. Entonces ABC no se debe incluir en las interacciones usadas como error. Esta decisión es mejor tomarla antes de que la interacción ABC sea analizada.

5.7 EJERCICIOS RESUELTOS

1. Ejercicio análisis de varianza a una vía

Robert Share es vicepresidente de la ANH. Es el encargo de implementar nuevas alternativas para atraer compañías extranjeras al país, para que realicen trabajos de exploración en 50 bloques que están aún sin explorar, para esto ha incluido algunos beneficiosas las últimas cuatro compañías petroleras que iniciaron trabajos de exploración en Colombia, los cuales consisten en menores regalías para el estado o un contrato de concesión de 40 años o más si llegan a encontrar hidrocarburos. Share está convencido de que estos beneficios atraerán a diferentes multinacionales. Algunas compañías prefieren menores regalías que tengan que pagar al estado y otras pueden sentirse más atraídas por un contrato de concesión de 40 años o más. Share decide utilizar el número de reservas probadas en millones de barriles (MBIs) que tiene cada compañía en 7 bloques seleccionados al azar. El desea determinar si existe una diferencia en el nivel promedio de reservas probadas entre las cuatro compañías y desea probar la hipótesis con un nivel de significancia del 5%. Si se llega alguna diferencia de reservas encontradas alta entre las compañías, share ofrecerá más beneficios a las compañías que quieran ingresar a Colombia a explorar, ya que algunas recuperaran su inversión rápidamente y obtendrán mayores ganancias que otras, debido a que tendrán un mayor de reservas. Los datos seleccionados aleatoriamente de cada compañía se muestran en la siguiente tabla.

TABLA 29. Reservas de las compañías en cada bloque (ejercicio 1)

Bloque	Compañía 1	Compañía 2	compañía 3	compañía 4
1	5.1	1.9	3.6	1.3
2	4.9	1.9	4.2	1.5
3	5.6	2.1	4.5	0.9
4	4.8	2.4	4.8	1.0
5	3.8	2.1	3.9	1.9
6	5.1	3.1	4.1	1.5
7	4.8	2.5	5.1	2.1
x_j	4.87	2.29	4.31	1.46

Solución

Hay $c = 4$ tratamientos (muestras) y $r_j = 7$ observaciones en cada tratamiento. El número total de observaciones es $n = rc = 28$.

Share desea probar la hipótesis al nivel del 5% que:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_A: \text{No todas las medias son iguales}$$

Se calcula la gran media de todas las observaciones del experimento

$$\bar{x} = \frac{\sum X_{ij}}{n} = \frac{5.1 + 4.9 + 5.6 + \dots + 2.1}{28} = 3.23$$

Se halla la suma de cuadrados:

$$S.C_{Tot} = \sum_{i=1}^r \sum_{j=1}^c (X_{ij} - X)^2$$

$$S.C_{Tot} = (5.1 - 3.23)^2 + (4.9 - 3.23)^2 + \dots + (2.1 - 3.23)^2 = 61$$

$$S.C_{Trat} = \sum r_j (X_j - X)^2$$

$$S.C_{Trat} = 7(4.87 - 3.23)^2 + 7(2.29 - 3.23)^2 + \dots + 7(1.46 - 3.23)^2 \\ = 55.33$$

$$S.C_E = \sum (X_{ij} - X_j)^2$$

$$S.C_E = (5.1 - 4.87)^2 + \dots + (4.8 - 4.87)^2 + (1.9 - 2.29)^2 + \dots \\ + (2.5 - 2.29)^2 + (3.6 - 4.31)^2 + \dots + (5.1 - 4.31)^2 \\ + (1.3 - 1.46)^2 + \dots + (2.1 - 1.46)^2 = 5.67$$

Se halla los cuadrados medios:

$$C.M_{Trat} = \frac{S.C_{Trat}}{C - 1} = \frac{55.33}{3 - 1} = 18.44$$

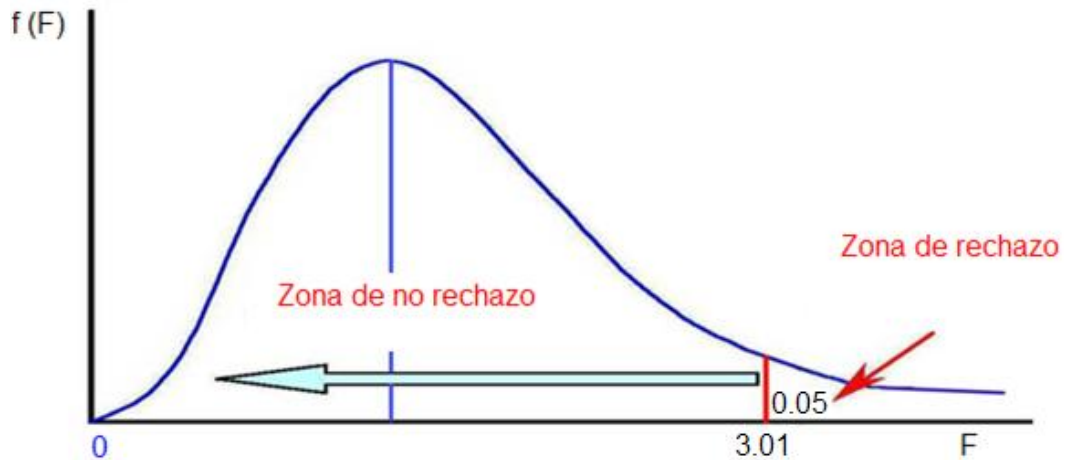
$$C.M_E = \frac{S.C_E}{n - c} = \frac{5.67}{28 - 4} = 0.236$$

Entonces la razón F es:

$$F = \frac{C.M_{Trat}}{C.M_E} = \frac{18.44}{0.236} = 78.14$$

Share debe utilizar 3 y 24 grados de libertad, ya que los gl para S.C_{Trat} = c – 1 = 3 y gl para S.C_E = n – c = 24. Y u nivel de significancia de 0.05 para hallar F en la tabla de la distribución F. en este caso $F_{0.05, 3, 24} = 3.01$

Figura 26. Distribución F (Ejercicio 1)



Debido a que $F = 78.14$, Share debe rechazar la hipótesis nula. Puede estar 95% seguro de que las reservas probadas encontradas por cada compañía no son iguales. Si considera que las diferentes compañías se sienten atraídas por diferentes tipos de beneficios. Deberá diseñar esquemas alternativos para atraer nuevas compañías debido a que el riesgo de encontrar hidrocarburos y recuperar la inversión es muy alto.

2. Ejercicio pruebas para la diferencia entre pares de medias

Cada vez más empresas petroleras buscan en reducir costos, por tal motivo acuden a empresas prestadoras de servicios para que realicen diferentes tipos de trabajos en las distintas áreas de la industria o a bolsas de empleo para su

personal operativo. Outdoor Well, una de estas empresas, contrato a un consultor para estudiar esta situación que se está presentado en la industria.

Parte del estudio realizado por el consultor requería una comparación de los costos que realizan las empresas en los tres tipos de contrato que están manejando, contratación directa, contratación con corporaciones prestadoras de servicios o con bolsas de empleo. En la tabla aparecen los datos seleccionados aleatoriamente de 6 petroleras en Colombia que cuentan con los tres tipos de contratación. Se determina si existe alguna diferencia promedio en los costos del tipo de contratación que manejan las empresas. ¿Si las compañías decidieran no contar con los tres contratos, sino solo con dos, cual es la mejor combinación para reducir costos? El consultor debe utilizar un nivel de significancia del 0.05.

TABLA 30. Costos de los tipos de contrato (Ejercicio 2)

Compañía	Cont. Directa	Bolsa Empleo	Prest. Servicios
1	(US\$/dia) 38.00	(US\$/dia) 30.00	(US\$/dia) 19.00
2	32.00	25.00	35.00
3	35.00	31.00	20.00
4	36.00	35.00	22.00
5	38.00	**	25.00
6	32.00	**	**
x_j	(US\$) 35.17	(US\$) 30.25	(US\$) 24.20

Solución

El consultor desea probar la hipótesis al nivel del 0.05 que:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_A: \text{No todas las medias son iguales}$

El valor crítico de F se halla de la tabla de distribución F. en este caso $F_{\alpha, c-1, n-c} = F_{0.05, 2, 12} = 3.89$.

La siguiente tabla es un resumen de la suma de cuadrados, grados de libertad, cuadrado medio y el valor F, para ser comparado con el valor crítico de F.

TABLA 31. Resumen de los estadísticos (Ejercicio 2)

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Tratamiento	328.0	2	164.0	7.74
Error	254.4	12	21.2	
Total	582.4	14	20.00	

La hipótesis nula se rechaza debido a que el valor de F (7.74) es mayor al valor crítico (3.89). Por lo tanto el consultor deberá utilizar las comparaciones por pares para determinar cuáles difieren del resto. Si α es 0.05, $F_{\alpha, c-1, n-c} = F_{0.05, 2, 12} = 3.89$. La comparación para la actividad de contratación directa y bolsa de empleo, utilizando la ecuación es:

$$DMS_{j,k} = \frac{1}{r_j} + \frac{1}{r_k} * C.M_E F_{\alpha, c-1, n-c}$$

$$DMS_{C,D,B,E} = \frac{1}{6} + \frac{1}{4} * 21.2 (3.89) = 5.85$$

Una comparación entre contratación directa y prestadora de servicios revela que:

$$DMS_{C.D,P.S} = \frac{1}{6} + \frac{1}{5} * 21.2 (3.89) = 5.48$$

Y la última comparación entre bolsa de empleo y prestadora de servicios produce:

$$DMS_{B.E,P.S} = \frac{1}{4} + \frac{1}{5} * 21.2 (3.89) = 6.08$$

Las diferencias entre las medias y si exceden o no su valor DMS respectivo son:

$$x_{C:D} - x_{B:E} = 35.17 - 30.25 = 4.92 < 5.85$$

$$x_{C:D} - x_{P.S} = 35.17 - 24.20 = 10.97 < 5.48$$

$$x_{B.E} - x_{P.S} = 30.25 - 24.20 = 6.05 < 6.08$$

Se puede concluir a un nivel de significancia del 5% que solo la contratación de empresas que prestan servicios y la contratación directa difieren significativamente. El consultor puede utilizar esta información para exponer ante la junta de Outdoor Well que la mayor diferencia en costos se encuentra entre tener contratación directa y empresas prestadoras de servicios y que es mas rentable para la compañía contar solo con esta combinación ya que se pueden reducir costos. Las otras dos combinaciones no difieren demasiado, por lo tanto es igual en términos de costos una que la otra.

3. Ejercicio análisis de varianza a dos vías

Una publicación reciente de una revista petrolera describió los esfuerzos realizados por una importante empresa para desarrollar varios sistemas que se pueden utilizar para prevenir la precipitación de parafinas en el pozo. Se seleccionaron cinco de estos sistemas aleatoriamente y se evaluó su rendimiento de 1 a 50 en cuatro pozos. Los resultados, junto con las medias de las filas y las columnas, aparecen en la siguiente tabla.

TABLA 32. Rendimiento de los sistemas (Ejercicio 3)

	Pozo (Tratamiento)				
Sistema	1	2	3	4	x_i
1	31	35	46	38	37.50
2	29	32	45	36	35.50
3	13	17	35	20	21.25
4	28	38	52	39	39.25
5	14	20	40	20	23.50
x_j	23	28.4	43.6	30.6	$\bar{x} = 31.4$

¿El gerente de un campo con problemas de parafinas desea saber si existe diferencia en el rendimiento promedio de los cuatro pozos con un nivel de significancia de 0.01?

Solución

El Gerente desea probar la hipótesis al nivel del 0.01, del rendimiento promedio dadas por cada uno de los cinco sistemas (filas),

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A : No todas las medias de las filas son iguales

Y también desea probar el rendimiento promedio de los cuatro pozos (columnas)

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_A : No todas las medias de las columnas son iguales

Se halla la suma de cuadrados:

$$S.C_{Tot} = \sum_{i=1}^r \sum_{j=1}^c (X_{ij} - X)^2$$

$$S.C_{Tot} = (31 - 31.4)^2 + (29 - 31.4)^2 + \dots + (20 - 31.4)^2 = 2344.8$$

$$S.C_{Trat} = \sum r_j (X_j - X)^2$$

$$S.C_{Trat} = 5(23 - 31.4)^2 + 5(28.4 - 31.4)^2 + \dots + 5(30.6 - 31.4)^2 \\ = 1145.2$$

$$S.C_{Bloq} = \sum C_i (X_i - X)^2$$

$$S.C_{Bloq} = 4(37.5 - 31.4)^2 + 4(35.5 - 31.4)^2 + \dots + 4(23.5 - 31.4)^2 \\ = 1124.3$$

$$S.C_E = S.C_{Tot} - S.C_{Trat} - S.C_{Bloq}$$

$$S.C_E = 2344.8 - 1145.2 - 1124.3 = 73.3$$

La tabla ANOVA a dos vías se convierte en:

TABLA 33. Resumen de los estadísticos (Ejercicio 3)

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre Muestras (Tratamiento)	1145.2	3	381.73	60.79
Entre Bloques	1124.3	4	281.08	44.76
Dentro de Muestras (Error)	75.3	12	6.28	
Variación Total	2344.8	19		

Si $\alpha = 0.01$, El valor crítico de F apropiado para las (filas) se halla de la tabla de distribución F. en este caso $F_{0.01, 4, 12} = 5.41$.

Ahora el gerente puede determinar si hay una diferencia significativa en el rendimiento promedio dados por cada uno de los cinco sistemas (filas), las cuales requerirán hacer bloques sobre los sistemas. Como el valor F de los cinco sistemas (filas) 44.76 es mayor que el valor crítico de F (5.41). La hipótesis nula se rechaza, y se determina, que los rendimientos promedio hechos por los cinco sistemas son diferentes y se necesita el bloqueo.

Si $\alpha = 0.01$, El valor crítico de F apropiado para las (columnas) se halla de la tabla de distribución F. en este caso $F_{0.01, 3, 12} = 5.95$.

El gerente puede probar ahora su hipótesis inicial respecto a los rendimientos promedio de los cuatro pozos (columnas). Como el valor F de los 4 pozos es mayor (60.79) que el valor crítico de F (5.95). La hipótesis nula se rechaza.

Al incluir un factor de bloqueo, el gerente pudo detectar una diferencia significativa en el rendimiento promedio de los pozos realizado por parte de los cinco sistemas. Sin el factor de bloqueo, la variación en los promedios, debido a los bloques (diferencias en el desempeño del sistema), se hubiera incluido en el factor de error de SCE.

4. Ejercicio análisis estadístico para modelo de efectos fijos

Se desea verificar si las pérdidas de lodo en el pozo, en barriles, ocasionadas en la operación de perforación dependen del peso del lodo que se está utilizando (lb/gal) y la cantidad de producto empleada para controlar el filtrado en el tratamiento. Para el experimento se utilizó pesos del lodo de 24, 22, 18 y 14 lb/gal. Por lo tanto los niveles de este factor es cualitativo, además se escogieron tres niveles de cantidad de producto utilizado para el control de filtrado: 50, 100 y 150 lb. Otros factores como galones por minuto bombeado de lodo, equipo y perforador se mantuvieron constantes durante el experimento. Para el factor peso del lodo use un nivel de significancia del 1%, y para la cantidad de producto utilizado para controlar el filtrado y la interacción entre ambas utilice un nivel de significancia del 2.5%. ¿Cuál de los factores influye en la pérdida de lodo hacia el pozo? La siguiente tabla contiene los datos.

TABLA 34. Perdidas de lodo en el pozo (Ejercicio 4)

Peso Lodo (A)	Cantidad de producto para el filtrado utilizada (B)						Y _i
	50	100	150	50	100	150	
	0.8	1.2	1.8	1.3	0.5	0.9	
24	1.0	0.8	1.0	1.1	0.7	1.1	
3.8	4.2		3.2		11.2		
	1.8	1.6	1.7	1.3	1.1	0.9	
22	2.0	1.6	1.4	1.3	1.1	1.0	
	7.0		5.7		4.1		16.8
	2.2	2.0	2.2	1.8	1.4	1.2	
18	2.2	1.4	2.0	1.9	1.5	1.3	
	7.8		7.9		5.4		21.1
	15.2	4.8	2.0	3.6	2.2	4.4	
14	10.0	4.0	4.2	2.6	5.8	3.2	
	34.0		12.4		15.6		62.0
Y _{.j.}	52.6		30.2		28.3		111.1

La cifra que aparece debajo de las cuatro medidas en cada celda corresponde al total de esta. El orden del experimento fue completamente aleatorio para las 48 mediciones, por lo tanto el experimento factorial es 4*3.

Solución

Se efectúan los cálculos de los componentes de las sumas de cuadrados:

$$S.C_{Tot} = Y_{ijk}^2 - \frac{Y^2 \dots}{abn} = 554.03 - \frac{(111.1)^2}{48} = 296.87$$

$$S.C_A = \frac{1}{bn} \sum_{i=1}^4 Y_{i..}^2 - \frac{Y^2 \dots}{abn}$$

$$S.C_A = \frac{(11.2)^2 + (16.8)^2 + (21.1)^2 + (62)^2}{3 * 4} - \frac{(111.1)^2}{48} = 134.25$$

$$S.C_B = \frac{1}{an} \sum_{j=1}^3 Y_{.j.}^2 - \frac{Y^2 \dots}{abn}$$

$$S.C_B = \frac{(52.6)^2 + (30.2)^2 + (28.3)^2}{4 * 4} - \frac{(111.1)^2}{48} = 22.83$$

$$S.C_{Celdas} = \sum_{i,j} \frac{Y_{ij.}^2}{n} - \frac{Y^2 \dots}{abn}$$

$$S.C_{Celdas} = \frac{(3.8)^2 + (4.2)^2 + (3.2)^2 + \dots + (15.6)^2}{4} - \frac{(111.1)^2}{48}$$

$$S.C_{Celdas} = 461.5375 - 257.15021 = 204.38$$

$$S.C_{AB} = S.C_{celd} - S.C_A - S.C_B = 204.38 - 134.25 - 22.83 = 47.29$$

La suma de cuadrados del error se calcula por diferencia.

$$S.C_E = S.C_{Tot} - S.C_{celdas} = 296.87 - 204.38 = 92.49$$

El análisis de varianza se muestra en la siguiente tabla.

TABLA 35. Resumen de los estadísticos (Ejercicio 4)

Fuente De variación	Sumas De Cuadrados	Grados De Libertad	C. Medio	F _o
Peso lodo (A)	134.25	3	44.75	17.42
Cantidad (B) Producto	22.83	2	11.41	4.44
Interacción (AxB)	47.29	6	7.88	3.07
Error	92.49	36	2.56	
Total	296.87	47		

Al observar los resultados de los estadísticos se concluye que:

- el factor peso del lodo es altamente significativo, es decir que el porcentaje de perdidas depende de este, puesto que 17.42 es mayor que el valor crítico F, $F_{0.01, 3, 36} = 4.40$.
- la cantidad de producto utilizada para controlar el filtrado es significativa al 2.5 %, puesto que 4.44 es mayor que el valor crítico F, $F_{0.025, 2, 36} = 4.1$.
- Existe interacción significativa al 2.5%, puesto que 3.07 es mayor que el valor crítico F, $F_{0.025, 6, 36} = 2.80$.

Se puede concluir que las pérdidas de lodo se deben tanto al peso del lodo, como a la cantidad de producto utilizado para controlar el filtrado. Por lo tanto el ingeniero de lodos debe utilizar un lodo con otras características.

5. Ejercicio modelo de efectos aleatorios

Supóngase que en el ejercicio anterior (4) los cuatro pesos de lodo y las tres cantidades de producto utilizado para controlar el filtrado han sido seleccionados al azar de un número mayor de niveles. El modelo es ahora de efectos aleatorios tal como se definió mediante la ecuación (5 – 41).

Solución

Las cuatro primeras columnas son idénticas a las del ejercicio (5.7.4). Los test estadísticos F son calculados con las siguientes ecuaciones.

Para verificar el efecto de A

$$F = \frac{C \cdot M_A}{C \cdot M_{AB}} = \frac{44.75}{7.88} = 5.68$$

Para verificar el efecto de B

$$F = \frac{C \cdot M_B}{C \cdot M_{AB}} = \frac{11.41}{7.88} = 1.45$$

Par verificar el efecto de la interacción

$$F = \frac{C \cdot M_{AB}}{C \cdot M_E} = \frac{7.88}{2.56} = 3.07$$

El análisis de varianza se muestra en la siguiente tabla.

TABLA 36. Resumen de las sumas de cuadrados (Ejercicio 5)

Fuente De variación	Sumas De Cuadrados	Grados De Libertad	C. Medio	F _o
Peso lodo (A)	134,25	3	44,75	5,68
Cantidad (B) Producto	22,83	2	11,41	1,45
Interacción (AxB)	47,29	6	7,88	3,07
Error	92,49	36	2,56	
Total	296,87	47		

De los test estadísticos obtenidos se concluye que la cantidad de producto no tiene efecto significativo al 20%, un resultado diferente si se compara con el ejercicio (4), donde se encontró significancia al 2.5%. El peso del lodo tiene efecto significativo, aunque es menor el valor observado de F. El cálculo del test para verificar la significancia de la interacción no se modifica. De los resultados anteriores se puede concluir que cuando hay efecto de la interacción el cambiar de un modelo de efectos fijos a un modelo de efectos aleatorios, esto tiende a ocultar los efectos principales de los factores. Una razón de esto es que es más difícil detectar si hay efecto en una población grande de niveles (infinitos) que una población finita y fija de niveles (a).

Las estimaciones de los componentes de varianza son:

$$\begin{aligned}\sigma^2 &= C.M_E = 2.569 \\ \sigma_{AB}^2 &= \frac{C.M_{AB} - C.M_E}{n} = \frac{7.88 - 2.56}{4} = 1.32 \\ \sigma_B^2 &= \frac{C.M_B - C.M_{AB}}{an} = \frac{11.41 - 7.88}{4 * 4} = 0.22 \\ \sigma_A^2 &= \frac{C.M_A - C.M_{AB}}{bn} = \frac{44.75 - 7.88}{4 * 4} = 3.07\end{aligned}$$

6. Ejercicio el experimento factorial generalizado

El gerente de la compañía petrolera NARCISA desea desarrollar los cálculos y el análisis, en relación a la producción promedio de petróleo en cuatro campos, los campos implementa cada uno, cuatro sistemas de levantamiento en diferentes pozos, cada sistema se prueba en zonas del yacimiento que no tienen gas libre y en zonas que poseen gas libre, los datos de producción en millones de barriles de cada pozo se encuentran consignados en la siguiente tabla. En donde este experimento corresponde a un modelo de un diseño factorial de 4x2x3 combinaciones de tratamiento. ¿Con un nivel de significancia del 1%, diga si los sistemas de levantamiento y sus interacciones tienen algún efecto sobre la producción de los campos?

TABLA 37. Producción de los pozos (Ejercicio 6)

Sistema De Levantamiento (A)									
No de Pozos (C)	B. Mecánico		B. Electrosumergible		B. Hidráulico		B. PCP		
(B)	(B)		(B)		(B)		(B)		
	Sin Gas	gas libre	Sin Gas	gas libre	Sin Gas	gas libre	Sin Gas	gas libre	
1	0.85	0.59	1.52	1.01	1.25	0.80	2.41	1.45	
	0.75	0.54	1.43	0.98	1.18	0.87	2.35	1.39	
	0.73	0.48	1.20	0.85	1.23	0.84	2.30	1.38	
3.94	6.99		6.07		11.28				
2	0.92	0.75	1.74	1.20	1.37	0.90	2.50	1.50	
	0.85	0.60	1.50	1.12	1.25	0.85	2.42	1.45	
	0.85	0.58	1.43	0.97	1.20	0.84	2.35	1.42	
	4.55		7.96		6.41		11.64		
3	1.20	0.86	1.78	1.30	1.42	0.98	2.51	1.53	
	1.25	0.95	1.50	1.04	1.31	0.91	2.45	1.48	
	1.34	0.98	1.45	1.02	1.28	0.84	2.38	1.39	
	6.58		8.09		6.74		11.74		

Solución

Se efectúan los cálculos de los componentes de las sumas de cuadrados:

$$S.C_{Tot} = Y_{ijkl}^2 - \frac{Y^2 \dots}{abcn} = 136.8 - \frac{(91.99)^2}{72} = 19.27$$

$$S.C_A = \frac{1}{bcn} \sum_{i=1}^4 Y_{i..}^2 - \frac{Y^2 \dots}{abcn}$$

$$S.C_A = \frac{(15.07)^2 + (23.04)^2 + (19.2)^2 + (34.7)^2}{2 * 3 * 3} - \frac{(91.99)^2}{72} = 11.84$$

$$S.C_B = \frac{1}{acn} \sum_{j=1}^2 Y_{.j.}^2 - \frac{Y^2 \dots}{abcn}$$

$$S.C_B = \frac{(55.45)^2 + (36.54)^2 + \dots}{4 * 3 * 3} - \frac{(91.99)^2}{72} = 4.96$$

$$S.C_C = \frac{1}{abn} \sum_{k=1}^3 Y_{..k.}^2 - \frac{Y^2 \dots}{abcn}$$

$$S.C_C = \frac{(28.28)^2 + (30.56)^2 + (33.15)^2}{4 * 2 * 3} - \frac{(91.99)^2}{72} = 0.49$$

Para determinar la $S.C_{AB}$ se calcula primero la $S.C_{celdas\ ab}$. Los totales por celdas ab son:

TABLA 38. Resumen de los estadísticos para ab (Ejercicio 6)

	Sistema De Levantamiento (A)							
	1		2		3		4	
B	1	2	1	2	1	2	1	2
	8.74	6.33	13.55	9.49	11.49	7.43	21.67	12.99

$$S.C_{celdas\ ab} = \sum_{i,j} \frac{Y_{ij.}^2}{cn} - \frac{Y^2 \dots}{abcn}$$

$$S. C_{celd ab} = \frac{(8.74)^2 + (6.33)^2 + \dots + (12.99)^2}{3 * 3} - \frac{(91.99)^2}{72} = 18.05$$

Por tanto la suma de cuadrados de la interacción será:

$$S. C_{AB} = S. C_{celdas} - S. C_A - S. C_B = 18.05 - 11.84 - 4.96 = 1.24$$

Los totales por celdas ac son:

TABLA 39. Resumen de los estadísticos para ac (Ejercicio 6)

Sistema De Levantamiento(A)				
C	1	2	3	4
1	3.94	6.99	6.07	11.28
2	4.55	7.96	6.41	11.64
3	6.58	8.09	6.74	11.74

$$S. C_{celdas ac} = \sum_i^4 \sum_K^3 \frac{Y_{i.K.}^2}{bn} - \frac{Y^2 \dots}{abcn}$$

$$S. C_{celd ac} = \frac{(3.94)^2 + (4.55)^2 + \dots + (11.74)^2}{2 * 3} - \frac{(91.99)^2}{72} = 12.65$$

Por tanto la suma de cuadrados de la interacción será:

$$S. C_{AC} = S. C_{celdas} - S. C_A - S. C_C = 12.65 - 11.84 - 0.49 = 0.319$$

Los totales por celdas bc son:

TABLA 40. Resumen de los estadísticos para bc (Ejercicio 6)

C	B	
	1	2
1	17.20	11.08
2	18.38	12.18
3	19.87	13.28

$$S. C_{celdas\ bc} = \sum_j \sum_K \frac{Y_{jk.}^2}{an} - \frac{Y^2 \dots}{abcn}$$

$$S. C_{celd\ bc} = \frac{(17.2)^2 + (18.38)^2 + \dots + (13.28)^2}{4 * 3} - \frac{(91.99)^2}{72} = 5.46$$

Por tanto la suma de cuadrados de la interacción será:

$$S. C_{BC} = S. C_{celdas} - S. C_B - S. C_C = 5.46 - 4.96 - 0.49 = 0.0052$$

Los totales por celdas abc son:

TABLA 41. Resumen de los estadísticos para abc (Ejercicio 6)

C	Sistema De Levantamiento (A)							
	1		2		3		4	
	1	2	1	2	1	2	1	2
1	2.33	1.61	4.15	2.84	3.66	2.41	7.06	4.22
1	2.62	1.93	4.67	3.29	3.82	2.59	7.27	4.37
1	3.79	2.79	4.73	3.36	4.01	2.73	7.34	4.40

$$S.C_{celdas\ abc} = \sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^3 \frac{Y_{ijk}^2}{n} - \frac{Y^2 \dots}{abcn}$$

$$S.C_{celd\ abc} = \frac{(2.33)^2 + (2.62)^2 + \dots + (4.40)^2}{4 * 3} - \frac{(91.99)^2}{72} = 18.87$$

Por tanto la suma de cuadrados de la interacción será:

$$S.C_{ABC} = S.C_{celd\ abc} - S.C_A - S.C_B - S.C_C - S.C_{AB} - S.C_{AC} - S.C_{BC}$$

$$S.C_{ABC} = 18.87 - 11.84 - 4.96 - 0.49 - 1.24 - 0.319 - 0.0052 = 0.0059$$

Las $S.C_E$ se calcula por diferencia

$$S.C_E = S.C_{Tot} - S.C_{celdas\ abc} = 19.27 - 18.87 = 0.398$$

La siguiente tabla muestra el análisis de varianza correspondiente.

TABLA 42. Resumen de los estadísticos (Ejercicio 6)

Fuente De variación	Sumas De Cuadrados	Grados De Libertad	C. Medio	F ₀
Sistema levantamiento (A)	11.84	3	3.94	474.89
Tipo yacimiento. (B)	4.96	1	4.96	5.97.57
No. De Pozos (C)	0.49	2	0.24	29.77
Interacción (AxB)	1.24	3	0.41	49.85
Interacción (AxC)	0.319	6	0.053	6.41
Interacción (BxC)	0.0052	2	0.0026	0.32
Interacción (AxBxC)	0.0059	6	0.00099	0.12
Error	0.398	48	0.0083	
Total	19.27	71		

Se hallan el valor crítico de F, para los factores y para las interacciones con un nivel de significancia del 1%. $F_{0.01, 3, 48} = 4.22$, $F_{0.01, 1, 48} = 7.19$, $F_{0.01, 2, 48} = 5.08$, $F_{0.01, 6, 48} = 3.20$.

Los resultados de los test estadísticos muestran que hay efecto significativo al 1% de los tres factores A, B, C, así como también de las interacciones (AB) y (AC). Las demás interacciones no son significantes y el nivel significativo crítico es mayor que 0.25.

7. Ejercicio diseño 2²

La oficina de control de pesos y medidas de las facilidades de superficie del campo cachemira, desean saber si dos empresas distribuidoras de tanques de almacenamiento de crudo, están vendiendo los tanques con las capacidades de almacenamiento con las que se les está especificando en la compra. Se escogieron dos tamaños de tanques con capacidades de 20 y 40 barriles, y la variable respuesta se midió como una desviación porcentual del volumen específico. La siguiente tabla muestra la información obtenida. El nivel de 20 barriles se asocia con "0" y el de 40 barriles con "1". La empresa x se asocia con el nivel "0" y la empresa y con el nivel "1". ¿Con que nivel de significancia la capacidad de almacenamiento real de los tanques se ve afectada por el tipo de empresa o por el volumen que dice tener cada tanque y su interacción?

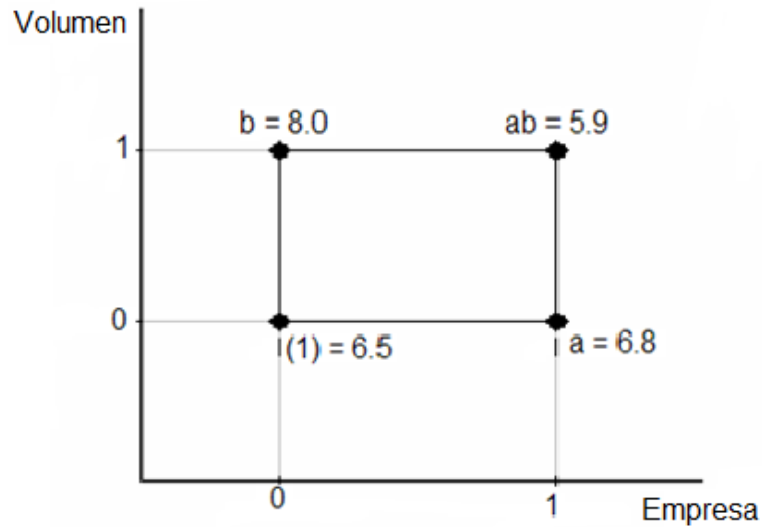
TABLA 43. Desviación estándar del volumen específico (ejercicio 7)

Volumen	Empresa			
	X (0)		Y (1)	
20 (0)	2.5		3.0	
	1.9		1.5	
	2.1	6.5	2.3	6.8
40 (1)	2.7		1.7	
	3.0		2.2	
	2.3	8.0	2.0	5.9

Solución

Los totales por combinación (los cuales aparecen al lado derecho en cada combinación) se pueden representar mediante los vértices de un cuadrado como indica la siguiente figura.

Figura 27. Diseño 2² (Ejercicio 7)



Las sumas de cuadrados están dados por:

$$S.C_A = \frac{[-1 + a - b + ab]^2}{4n} = \frac{[-6.5 + 6.8 - 8 + 5.9]^2}{4 * 3} = 0.27$$

$$S.C_B = \frac{[-1 - a + b + ab]^2}{4n} = \frac{[-6.5 - 6.8 + 8 + 5.9]^2}{4 * 3} = 0.03$$

$$S.C_{AB} = \frac{[1 - a - b + ab]^2}{4n} = \frac{[6.5 - 6.8 - 8 + 5.9]^2}{4 * 3} = 0.48$$

La suma de cuadrados del total y del error se calcula de la forma usual.

$$S.C_{Tot} = 64.12 - \frac{(27.2)^2}{4 * 3} = 2.466$$

$$S.C_E = S.C_{Tot} - S.C_A - S.C_B - S.C_{AB}$$

$$S.C_E = 2.466 - 0.27 - 0.03 - 0.48 = 1.6866$$

El análisis de varianza se muestra en la siguiente tabla.

TABLA 44. Resumen de los estadísticos (Ejercicio 7)

Fuente De variación	Sumas De Cuadrados	Grados De Libertad	C. Medio	F _o
(A)	0.27	1	0.27	1.28
(B)	0.03	1	0.03	0.14
(AB)	0.48	1	0.48	2.277
Error	1.6866	8	0.2108	
Total	2.466	11		

Los resultados observados para F indican que no hay efecto significativo de los dos factores principales, ni tampoco para la interacción. En conclusión, los tanques de almacenamiento si vienen con el volumen con el cual se les pide a las empresas distribuidoras.

8. Ejercicio diseño factorial 2³

Se llevó a cabo un experimento en una multinacional petrolera para verificar si dos condiciones de trabajo (sin bonos de campo y con bonos de campo), dos turnos (7 a.m. – 3 p.m. y 3 p.m. – 11 p.m.), y dos días típicos de trabajo (lunes y miércoles), afectan la producción diaria del campo. Los datos de producción en miles de barriles por día se muestran en la tabla. ¿Con que nivel de significancia afectan los factores o sus interacciones la producción del campo?

TABLA 45. Producción del campo (Ejercicio 8)

		TURNO (B)			
Condición De trabajo	0 DIA (C)		1 DIA (C)		
	0	1	0	1	
0	75	78	73	81	
	79	98	77	77	
	77	80	75	80	
	$(1) = 231$		$c = 238$	$b = 225$	$c = 238$
1	85	90	82	93	
	83	92	84	91	
	87	93	80	95	
	$a = 255$	$ac = 275$	$ab = 246$	$abc = 279$	

Solución

Usando los totales bajo las combinaciones de tratamientos mostrados en la tabla, se pueden obtener los efectos medios

$$A = \frac{1}{4n} - 1 + a - b + ab - c + ac - bc + abc$$

$$A = \frac{1}{4 * 3} - 231 + 255 - 225 + 246 - 238 + 275 - 238 + 279 = \frac{123}{12} = 10.25$$

$$B = \frac{1}{4n} - 1 - a + b + ab - c - ac + bc + abc$$

$$B = \frac{1}{4 * 3} - 231 - 255 + 225 + 246 - 238 - 275 + 238 + 279 = \frac{-11}{12} = -0.91$$

$$C = \frac{1}{4n} - 1 - a - b - ab + c + ac + bc + abc$$

$$C = \frac{1}{4n} - 231 - 255 - 225 - 246 + 238 + 275 + 238 + 279 = \frac{73}{12} = 6.08$$

$$AB = \frac{1}{4n} - 1 - a - b + ab + c - ac - bc + abc$$

$$AB = \frac{1}{4n} 231 - 255 - 225 + 246 + 238 - 275 - 238 + 279 = \frac{1}{12} = 0.083$$

$$AC = \frac{1}{4n} 1 - a + b - ab + ac - bc + abc$$

$$AC = \frac{1}{4n} 231 - 255 + 225 - 246 - 238 + 275 - 238 + 279 = \frac{33}{12} = 2.75$$

$$BC = \frac{1}{4n} 1 + a - b - ab - c - ac + bc + abc$$

$$BC = \frac{1}{4n} 231 + 255 - 225 - 246 - 238 - 275 + 238 + 279 = \frac{19}{12} = 1.58$$

$$ABC = \frac{1}{4n} - 1 + a + b - ab + c - ac - bc + abc$$

$$ABC = \frac{1}{4n} -231 + 255 + 225 - 246 + 238 - 275 - 238 + 279 = \frac{7}{12} 0.583$$

De acuerdo a la ecuación (5 – 86), las sumas de cuadrados son:

$$S.C_A = \frac{(123)^2}{24} = 630.375$$

$$S.C_B = \frac{(-11)^2}{24} = 5.043$$

$$S.C_C = \frac{(73)^2}{24} = 222.042$$

$$S.C_{AB} = \frac{(1)^2}{24} = 0.042$$

$$S.C_{AC} = \frac{(33)^2}{24} = 45.374$$

$$S.C_{BC} = \frac{(19)^2}{24} = 15.042$$

$$S.C_{ABC} = \frac{(7)^2}{24} = 2.042$$

La suma de cuadrados del total es $S.C_{Tot} = 975.96$, y por sustracción, la suma de cuadrados del error es $S.C_E = 56.0$

TABLA 46. Resumen de los estadísticos (Ejercicio 8)

Fuente De variación	Sumas De Cuadrados	Grados De Libertad	C. Medio	F_o
Cond. de trabajo (A)	630.375	1	630.375	180.11 (1)
Turno (B)	5.043	1	5.043	1.44
Dia (C)	222.042	1	222.042	63.44 (1)
AB	0.042	1	0.042	0.012
AC	45.374	1	45.374	12.96 (1)
BC	15.042	1	15.042	4.30 (2)
ABC	2.042	1	2.042	0.58
Error	56.0	16	3.5	
Total	975.96	23		

La tabla anterior muestra el análisis de varianza correspondiente. Donde, los resultados observados para F indican que ahí significancia de los factores de la siguiente manera.

(1)significancia al 1%

(2) significancia al 2%

5.8 EJERCICIOS PROPUESTOS

1. Los laboratorios Loden, le realizan pruebas a cuatro rompedores de emulsión(Agua – crudo) que pueden afectar la concentración de un componente en la mezcla líquida, el laboratorio desea determinar si existe una diferencia significativa en el nivel promedio de las pruebas entre los cuatro rompedores de emulsión con un nivel de significancia del 1%. En la siguiente tabla aparecen 5 pruebas seleccionadas aleatoriamente de cada rompedor.

TABLA 47. Datos de concentración (Ejercicio 1)

Prueba	Rompedor			
	A	B	C	D
1	58.2	56.3	50.1	52.9
2	57.2	54.5	54.2	49.9
3	58.4	57	55.4	50
4	55.8	55.3	53.7	51.7
5	54.9	54.92	54.6	56.5

2. Un ingeniero de producción desea comparar el factor caudal del pozo en barriles por día (BPD), utilizando cuatro sistemas de levantamiento diferentes. Si se aumenta la producción, el sistema seleccionado se implementará en todo el campo. Se escogen 4 pozos y a cada uno se le implementa los cuatro sistemas de levantamiento, el caudal de cada pozo con los diferentes sistemas se muestran en la siguiente tabla.

TABLA 48. Producción de los pozos (Ejercicio 2)

Pozo	Sistema de levantamiento			
	B. Hidráulico	Gas lift	B. PCP	B. Mecánico
1	790	690	830	850
2	820	520	790	780
3	570	620	850	780
4	790	610	780	730

¿A un nivel de significancia del 1%, use el método de Tukey para probar las diferencias y determinar si el ingeniero debería utilizar algún sistema de levantamiento o evitar su uso ya que el pozo produce naturalmente 780 BPD?

3. Un ingeniero químico quiere probar el efecto de 4 agentes químicos sobre la resistencia de una lechada en un completamiento de pozos. Debido a que podría haber variabilidad de una lechada a otra, el ingeniero decide usar un diseño de bloques aleatorizados, con los barriles de lechada considerados como bloques. Selecciona 5 lechadas y aplica los 4 agentes químicos de manera aleatoria a cada lechada. A continuación se presentan las resistencias a la tensión resultantes. Analizar los datos de este experimento (utilizar $\alpha=0.05$) y sacar las conclusiones apropiadas.

TABLA 49. Resistencias a la tensión (Ejercicio 3)

Agente Químico	Lechada				
	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

4. Se desea verificar mediante una prueba, si la resistencia de un aislante especial que se utiliza en los oleoductos puede ser afectada por la presión y la temperatura aplicada por el hidrocarburo que transporta. Para el experimento se utilizó una presión de 120, 130, 140 y 150 (Psi), además se escogieron tres niveles de temperatura: 150, 160 y 170 (°F), donde los niveles de los factores son fijos y equiespaciados. Se utilizaron 36 tubos iguales a los cuales se les aplico el aislante en las mismas condiciones. El equipo usado para medir la resistencia al esfuerzo cortante fue el mismo. Otros factores como tiempo de la prueba y operador se

mantuvieron constantes durante el experimento. Para el factor de presión use un nivel de significancia del 2.5%, y para la temperatura y la interacción entre ambas utilice un nivel de significancia del 1%. La siguiente tabla contiene los datos.

TABLA 50. Resistencia al esfuerzo cortante (Ejercicio 4)

Presión (A)	Temperatura De Operación (B)			Y _{i..}
	50	100	150	
120	9.58	11.51	9.00	90.1
	9.75	11.28	9.12	
	9.60	11.18	9.08	
	(28.93)	(33.97)	(27.20)	
120	9.82	10.69	9.57	88.73
	9.69	10.28	9.40	
	9.79	10.10	9.39	
	(29.30)	(31.07)	(28.36)	
120	8.96	10.85	9.20	86.71
	8.43	11.01	9.03	
	9.10	10.89	9.15	
	(26.49)	(32.84)	(27.38)	
120	8.82	10.28	9.75	88.47
	9.98	10.44	9.80	
	9.35	10.12	9.93	
	(28.15)	(30.84)	(29.48)	
Y _{.j.}	112.87	128.72	112.42	354.01

La cifra que aparece debajo de las tres medidas en cada celda corresponde al total de esta. El orden del experimento fue completamente aleatorio.

5. Para el ejercicio anterior (4), asuma que las cuatro presiones y las tres temperaturas usadas, han sido seleccionadas al azar de un número mayor de

niveles. El modelo es ahora de efectos aleatorios tal como se definió mediante la ecuación (5 – 41).

6. Se efectuó un experimento para determinar si algunos factores influyen en el grado de absorción de CO₂ en el yacimiento del campo Tumbaleke en un proceso de inyección. Se escogieron como factores, la densidad del crudo (A), medida en libra por pie cubico, la presión a la que se incorpora el CO₂ (B), medida en libras por pulgada cuadrada, y la temperatura del CO₂ a la cual se incorpora en el yacimiento(C), medida en °F, la variable respuesta se mide en volúmenes absorbidos de CO₂ por parte del crudo, lo cual significa que si un volumen de CO₂ se disuelve en el mismo volumen de líquido (crudo), se dice que dicho liquido contiene un volumen de CO₂, si se disuelven dos volúmenes, el líquido contiene dos volúmenes, etc. La siguiente tabla contiene la información para un experimento factorial 3 * 2 * 2, con dos replicaciones por combinación.

TABLA 51. Grado de absorción CO₂ (Ejercicio 6)

Presión (B)				
Densidad (A)	1000		1450	
	Temperatura (C) (100)	Temperatura (C) (150)	Temperatura (C) (100)	Temperatura (C) (150)
10	1.5	1.9	2.1	1.9
	1.7	1.8	2.0	2.0
15	1.4	1.8	2.3	2.1
	1.5	2.0	1.9	2.1
20	1.6	2.1	2.0	2.2
	1.4	1.7	1.9	1.8

¿Cuáles son los factores o combinaciones que influyen en el grado de absorción del CO₂ en el crudo, utilizando un nivel de significancia del 5%?

7. Considere una investigación llevada a cabo para estudiar el efecto que tiene la concentración de un viscosificante y la cantidad de barita utilizada en un lodo de perforación base agua. Sea la concentración del viscosificante el factor A con dos niveles de interés, 15 y 20%, la cantidad de barita constituye el factor B; el nivel alto o superior denota el uso de dos sacos de barita y el nivel bajo denota el uso de un saco. El experimento se realiza (“replica o repite”) tres veces, y los datos son como sigue:

TABLA 52. Datos obtenidos del experimento (Ejercicio 7)

Cantidad Barita (B)	Concentración Viscosificante (A)	
	15% (0)	20% (1)
28		36
1 Saco (0)	25	
27 80		32 100
18		31
2 Sacos (1)	19	
23 60		29 90

¿Verificar a qué nivel son significativos los efectos principales e interacciones de los factores, si los hay?

8. Se midió la producción de crudo en porcentaje de un campo petrolero, como función de la presión natural del yacimiento, la presencia o ausencia de un sistema de levantamiento, y dos tipos de choques utilizados, la temperatura y el tiempo se mantuvieron constantes y se utilizó el mismo pozo. Los datos obtenidos fueron.

TABLA 53. Producción en porcentaje del pozo (Ejercicio 8)

Condición De operación		Presión Psi (B)					
		890 Choque (C)			900 Choque (C)		
1	2	1	2	1	2	1	2
Sin sistema Levantamiento	75.4	78.7	73.4			80.1	80.3
		76.2		79.8		80.5	
Con sistema Levantamiento		76.8	77.5			82.1	80.0
		79.7	80.6			87.8	90.6
	80.2	83.4		89.1		89.3	
	84.3	81.9	90.2	88.7			

¿Verificar a qué nivel son significativos los efectos principales e interacciones de los factores, si los hay?

6. SERIES CRONOLOGICAS

Los pronósticos o predicciones, son una herramienta esencial en cualquier proceso de toma de decisiones, sus usos varían desde la determinación de requerimientos de inventario de una empresa hasta la estimación de la producción de un pozo de petróleo. La calidad de las predicciones que se pueden efectuar está estrechamente relacionada con la información que se puede extraer y utilizar de los datos que se tengan. El análisis de series cronológicas es un método cuantitativo que se utiliza para determinar patrones en los datos recolectados a través del tiempo.

6.1 CONCEPTOS GENERALES

Una serie cronológica, está formada por un conjunto de observaciones de una variable, ordenadas en función del tiempo. Su ámbito de aplicación, no está limitado a la esfera estrictamente económica. Su metodología puede utilizarse en la medicina (electrocardiograma, electroencefalograma, etc.), agricultura (evolución de las lluvias en las diferentes estaciones), psicología (evolución del coeficiente intelectual de una persona), ingeniería de petróleos (producción) y en muchas otras disciplinas. El propósito perseguido con el análisis de series, consiste en predecir los valores futuros de la variable estudiada.

El proceso de desarrollar un pronóstico comienza con la recolección de datos anteriores durante varios periodos. El conjunto de datos resultantes se denomina una serie de tiempo o serie cronológica porque contiene observaciones para alguna variable durante el tiempo. Los periodos de tiempo varían en duración. Pueden ser anuales, trimestrales, mensuales o incluso diarios. Los periodos de solo una hora pueden utilizarse para variables altamente volátiles como el precio del petróleo o para las acciones transadas en una de las bolsas reconocidas.

6.2 COMPONENTES DE UNA SERIE CRONOLOGICA

Se utiliza el término serie cronológica para referirse a cualquier grupo de información estadística que se acumula a intervalos regulares⁵¹. Hay cuatro tipos o componentes de cambio o variación implicados en el análisis de series cronológicas, estos son:

6.2.1 Tendencia secular. La tendencia secular, o simplemente tendencia, Figura 28, es la conducta a largo plazo de la variable durante un periodo de longitud prolongado. Refleja la dirección general de la serie de tiempo como ascendente o descendente. Para poder detectarla es necesario que la serie conste de un número de observaciones elevada, a lo largo de muchos años, para que se pueda determinar si la serie muestra un movimiento a largo plazo que responda a una determinada ley de crecimiento, decrecimiento o estabilidad. Ese comportamiento tendencial puede responder a distintos perfiles: lineal, exponencial, parabólico, etc.

Figura 28. Componentes De Una Serie Cronológica



Fuente: www.uca.es/uca/.../f.../CC%20Trabajo%20Complemento%201.pdf

⁵¹John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

6.2.2 Variaciones estacionales. Este tipo de variación implica patrones de cambio en el lapso de un año que tiende a repetirse anualmente. Son movimientos de la serie que se repiten de forma periódica, como se puede apreciar en el gráfico de color verde de la figura 28. La razón de estas variaciones se basa en causas de tipo climático (por ejemplo producción) o de ordenamiento del tiempo (los días de la semana condicionan el comportamiento de ciertas series). La periodicidad es generalmente es el año, aunque puede ser el mes, la semana o incluso el día.

6.2.3 Variaciones cíclicas. Muchas variables presentan una tendencia a la variación por encima y por debajo de la tendencia a largo plazo, durante un periodo prolongado de tiempo. Como se puede apreciar en el gráfico de color morado de la figura 28. Tales fluctuaciones se denominan variaciones cíclicas o ciclos comerciales. Abarcan periodos mucho más prolongados que las variaciones estacionales, y con frecuencia incluyen tres o más años de duración.

Un ciclo contiene cuatro fases: 1) Alza o expansión, durante la cual el nivel de actividad, por ejemplo de la producción se acelera, 2) el pico, en el cual la tasa de actividad de la producción tiene “el punto más alto”, 3) el descenso o contracción, la actividad merma, y 4) la depresión, cuando la actividad está en su punto mínimo. Un ciclo se va desde una fase hasta la siguiente.

6.2.4 Variaciones irregulares. Las series de tiempo también contienen variaciones irregulares o aleatorias, producidas por sucesos inusuales, que producen movimientos sin un patrón discernible. Tales movimientos son, como las huellas, únicas y probablemente no ocurran de nuevo de la misma manera. Pueden ser producidos por eventos tales como guerras, inundaciones, terremotos o embargos de petróleo.

6.3 MODELOS DE SERIES CRONOLOGICAS

La interacción de los cuatro componentes mencionados en la sección 6.2 genera la serie cronológica⁵². La forma en que se combinen puede ser muy variada, pero tradicionalmente se ha optado por dos modelos distintos. El aditivo y el multiplicativo, aunque en algunas ocasiones se mezclan ambos, el modelo aditivo se expresa como:

$$Y_t = T_t + S_t + C_t + I_t \quad (6- 1)$$

En donde Y_t es el valor de la serie cronológica para el periodo t , y la tendencia son los valores al lado derecho, la variación estacional, la variación cíclica y la variación aleatoria o irregular, respectivamente para el mismo periodo. En el modelo aditivo, todos los valores están expresados en unidades originales, y S , C e I son desviaciones alrededor de T . El modelo aditivo sufre del supuesto, algo irreal, de que los componentes son independientes uno del otro. Esto rara vez es el caso en la vida real. En la mayoría de los casos los movimientos de un componente tendrán un impacto en los otros componentes, por tanto se niega el supuesto de independencia. O quizás de manera más común, con frecuencia se encuentran que ciertas fuerzas simultáneamente afectan dos o más componentes. De nuevo, el supuesto de independencia no se cumple.

Como resultado, el modelo multiplicativo es el que se prefiere con frecuencia, supone que los componentes interactúan entre si y no se mueven independientemente. El modelo multiplicativo se expresa como:

$$Y_t = T_t * S_t * C_t * I_t \quad (6- 2)$$

⁵²John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

En el modelo multiplicativo solo T se expresa en las unidades originales, y S, C e I se expresan en términos de porcentajes. Debido a que las fluctuaciones ocurren dentro de periodos de menos de un año, podrían no estar reflejados en los datos anuales. Una serie de tiempo para datos anuales se expresaría como:

$$Y_t = T_t * C_t * I_t \text{ (6- 3)}$$

6.4 TECNICAS DE SUAVIZAMIENTO

El comportamiento general de una variable, con frecuencia puede analizarse mejor observando su tendencia a largo plazo. Sin embargo si la serie de tiempo contiene demasiadas variaciones o cambios estacionales a corto plazo, la tendencia puede ser algo confusa y difícil de observar. Es posible eliminar muchos de estos factores que se pueden confundir promediando los datos de varios periodos. Esto se logra utilizando ciertas técnicas de suavizamiento⁵³ o suavización que eliminan las fluctuaciones aleatorias de la serie, proporcionando por tanto un panorama menos distorsionado de la conducta real de la serie. Se analizarán dos métodos comunes de suavizamiento de los datos de series de tiempo: promedio móvil y suavizamiento exponencial.

6.4.1 Promedios móviles. Un promedio móvil (PM) tendrá el efecto de “suavizamiento” de los datos, produciendo un movimiento con menos picos y valles. Se calcula promediando los valores en la serie cronológica sobre un número fijo de periodos. El mismo número de periodos se mantiene para cada promedio, eliminando la observación más antigua y recogiendo la más reciente.

⁵³John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

Cuando un número impar de periodos se utiliza en el promedio móvil, los resultados pueden centrarse automáticamente en el periodo de la mitad. Sin embargo, si existe un número par de periodos en el promedio, debe hacerse un ajuste ya que no existe una observación en la mitad en la cual se centre automáticamente el valor.

Los promedios móviles pueden utilizarse para eliminar variaciones irregulares y estacionales. Cada entrada en el PM se deriva de cuatro trimestres, es decir, el valor de todo un año. Por tanto, el promedio móvil “promedia” toda variación estacional que pueda ocurrir dentro del año, eliminándolas de manera efectiva y dejando solo la tendencia y las variaciones cíclicas. En general, si el número de periodos en un promedio móvil es suficiente como para abarcar todo un año (si se utilizan datos de 12 meses, 52 si se utilizan datos semanales), las variaciones estacionales se promedian y se eliminan de la serie. Entonces se afirma que los datos están desestacionalizados.

6.4.2 Suavizamiento exponencial. Como su nombre lo indica, tiene el efecto de suavizar una serie. También proporciona un medio efectivo de predicción. Suavizamiento exponencial de primer orden se utiliza cuando los datos no presentan ningún patrón de tendencia

El modelo contiene un mecanismo de autocorrección que ajusta los pronósticos en dirección opuesta a los errores pasados. La ecuación es:

$$F_{t+1} = \alpha A_t + (1 - \alpha) F_t \quad (6-4)$$

En donde F_{t+1} Es el pronóstico para el siguiente periodo

A_t Es el valor real observado para el periodo corriente

w_i Es la proyección hecha previamente para el periodo corriente

A es una constante de suavizamiento a la cual se le da un valor entre 0 y 1. Teniendo en cuenta que los datos no tienen tendencia ascendente o descendente, pero varían alrededor de algún promedio a largo plazo, se toma el valor F_{t+1} como pronóstico para cualquier periodo futuro. El valor seleccionado para α es crucial. Debido a que se desea producir un pronóstico con el error más pequeño posible, el valor α que minimiza el cuadrado medio del error (CME) debe ser óptimo. El ensayo y error sirve con frecuencia como el mejor método para determinar el valor α apropiado. El CME se calcula como:

$$CME = \frac{(F_t - A_t)^2}{n - 1} \quad (6- 5)$$

6.5 ANALISIS DE TENDENCIA

Si una serie cronológica tiene una tendencia ascendente o descendente a largo plazo, el análisis de tendencia puede ser útil para desarrollar pronósticos. Es decir, si una tendencia está presente debido a que los datos no varían alrededor de algún promedio a largo plazo, los métodos de suavizamiento como el promedio móvil y el suavizamiento exponencial no son apropiados. En su lugar, se puede estimar una recta de tendencia o cualquier otro comportamiento no lineal, como una cuadrática, una exponencial o un comportamiento logarítmico, cuyos coeficientes se estiman utilizando las técnicas para la regresión simple. La elección de la tendencia se hará a partir de la gráfica de la serie, estudiando qué tipo de función se ajusta mejor al desarrollo descrito en ella, o a partir de la verificación de determinados supuestos de crecimiento y evolución de los datos.

6.5.1 Fundamentos de regresión. El análisis de regresión⁵⁴ se utiliza en la predicción, para el caso de dos variables; se desarrolla un modelo que utiliza la variable independiente (t) para obtener una mejor predicción de la otra variable, la variable dependiente Y. el análisis de correlación por contraste con el de regresión, se utiliza para medir la fuerza de asociación entre las variables. En este caso el objetivo no es utilizar una variable para predecir la otra sino solo medir la fuerza de asociación o “covarianza” entre las dos variables. En un análisis de regresión, los valores individuales se trazan en una gráfica bidimensional llamada “Diagrama de Dispersión”, cada valor se traza en sus coordenadas Y y t particulares.

El análisis de regresión se clasifica generalmente en dos clases, simple y múltiple. En este manual se trabajara solo la regresión simple, ya que es aquella en la cual entran solamente dos variables, tales como la regresión de Y respecto a t. este tipo de relación puede ser lineal, curvilínea o exponencial.

6.5.2 Coeficiente de correlación. En la mayoría de los casos el principal interés del investigador no solamente está en poder medir la relación que pueda existir, entre las dos variables, sino que además se concentra en determinar si están o no correlacionadas⁵⁵, y en caso de estarlo, en hallar qué tan fuerte es este grado de relación. Esta técnica analítica que se utiliza en hallar este grado de relación entre las dos variables, recibe el nombre de “Análisis de Correlación”. El parámetro utilizado para medir este grado de relación entre dos variables recibe el nombre de “Coeficiente de Correlación”.

Este valor, simbolizado por r, puede asumir valores de -1 y $+1$. Si la relación entre dos variables es perfectamente lineal e inversa, o sea pendiente negativa, el

⁵⁴ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

⁵⁵ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

coeficiente de correlación $r = -1$. Si la relación es perfectamente lineal y directa, pendiente positiva, entonces $r = 1$.

Cuando las dos variables no están correlacionadas, entonces $r = 0$, y sus puntos en el diagrama estarán dispersos. El coeficiente de correlación entre dos variables Y y t tendrá siempre el mismo signo que la pendiente de la de la ecuación de regresión de Y sobre t . con base en estos hechos, se puede deducir que mientras más cerca este el valor numérico del coeficiente de correlación -1 o $+1$, entonces más estrecho será el grado de relación entre las variables Y y t . Por tanto, la magnitud d r (coeficiente de correlación) nos da una indicación de la fuerza o grado de relación entre las dos variables.

El coeficiente de determinación (r^2),servirá como medida de la bondad de ajuste de la tendencia estimada a la evolución temporal de los datos de la serie. En este caso se podrá calcular de la siguiente forma:

$$r^2 = \frac{S_{tY}^2}{S_t^2 * S_Y^2} \quad (6- 6)$$

El coeficiente de correlación (r) se puede obtener de la ecuación anterior así:

$$r = \frac{\frac{\sum tY}{n} - tY}{\sqrt{\left(\frac{\sum t^2}{n} - t^2\right) * \left(\frac{\sum Y^2}{n} - Y^2\right)}} \quad (6- 7)$$

En la regresión lineal simple r toma el signo de b_1 . Si b_1 es positivo, r es positivo, si b_1 es negativo, r es negativo, Si b_1 es cero, r es cero. La ecuación (6 – 6) presenta el coeficiente de relación de la muestra en términos de la variación

explicada. Esta fórmula es apropiada cuando se hace el análisis de correlación junto con la regresión. Pero, si se hace solo la correlación, el coeficiente r de correlación se calcula así:

$$r = \frac{(t_i - t)(Y_i - Y)}{(t_i - t)^2 * (Y_i - Y)^2} \quad (6- 8)$$

6.5.3 Tendencia lineal. En el caso de tendencia lineal se supone que el crecimiento o decrecimiento de los datos de la serie se produce con una variación constante. Esto se puede comprobar obteniendo la serie transformada $\Delta Y_t = Y_t - Y_{t-1}$, de primeras diferencias, y verificar que el resultado es una serie estacionaria con valores muy parecidos unos a otros. Si es así, el esquema adecuado sería el aditivo, donde la variable dependiente es la serie cronológica que se desea pronosticar y el tiempo se utiliza como variable independiente. El modelo a estimar es simplemente⁵⁶.

$$Y_t = b_0 + b_1 t \quad (6- 9)$$

Los valores para t se obtienen codificando el periodo, iniciando con 1 para el primer periodo, 2 para el segundo, y así sucesivamente. La ecuación (6 – 6) supone un modelo de regresión lineal simple donde la variable exógena es el tiempo y los parámetros se estimarán por mínimos cuadrados.

Pendiente de la recta de tendencia:

$$b_1 = \frac{Y t - nYt}{t^2 - nt^2} \quad (6- 10)$$

⁵⁶ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

Intercepto de la recta de tendencia:

$$b_0 = Y - b_1 t \quad (6- 11)$$

En donde Y valores de la variable dependiente

X valores de la variable independiente

\bar{Y} media de los valores de la variable independiente

n número de puntos de dato en la serie cronológica

6.5.4 Tendencia Cuadrática. En la sección (6.5.3) se ha descrito el método de ajustar una línea recta a una serie cronológica. Pero muchas de estas series se describen mejor mediante curvas, no líneas rectas. En estos casos, el modelo lineal no describe de manera adecuada el cambio en la variable conforme avanza el tiempo⁵⁷. Para salvar este problema, a menudo se utiliza una curva parabólica, que es descrita matemáticamente por una ecuación de segundo grado. La forma general de una ecuación de segundo grado estimada es:

$$Y_t = b_0 + b_1 t + b_2 t^2 \quad (6- 12)$$

De nuevo se utiliza el método de mínimos cuadrados para determinar las constantes numéricas de la ecuación de segundo grado que describe el mejor ajuste, se puede determinar el valor de las constantes numéricas (b_0 , b_1 , b_2) a partir de las siguientes tres ecuaciones:

$$Y = b_0 n + b_2 t^2 \quad (6- 13)$$

⁵⁷ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

$$t^2 Y = b_0 t^2 + b_2 t^4 \quad (6-14)$$

$$b_1 = \frac{Y t}{t^2} \quad (6-15)$$

6.5.5 Tendencia exponencial. El supuesto, en este caso, es que los datos evolucionan a una tasa de variación constante. Esto se podrá confirmar obteniendo la serie transformada $\Delta \text{Log} Y_t$, que supone una aproximación lineal de la tasa de variación⁵⁸, y verificar, como en el caso de la tendencia lineal, que el resultado es una serie estacionaria con valores casi constantes. Si comprobamos esto, el esquema más apropiado sería el multiplicativo, quedando:

$$Y_t = c b^t \quad (6-16)$$

Esto se puede expresar como:

$$\log Y = \log c + t \log b \quad (6-17)$$

Las ecuaciones normales para el cálculo de los parámetros c y b, utilizando el método de los mínimos cuadrados será:

$$\log Y = n \log c + t \log b \quad (6-18)$$

$$t \log Y = t \log c + t^2 \log b \quad (6-19)$$

⁵⁸ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

Las ecuaciones anteriores se utilizan cuando la sumatoria de t es diferente de cero, si la sumatoria de t es igual a cero, entonces los parámetros se obtienen directamente de la siguiente forma:

$$\log c = \frac{\log Y}{n} \quad (6- 20)$$

$$\log c = \frac{t \log Y}{t^2} \quad (6- 21)$$

6.6 DESCOMPOSICION DE LAS SERIES CRONOLOGICAS

Con frecuencia es útil descomponer una serie cronológica “desglosando” cada uno de sus cuatro componentes. Así, cada componente se puede examinar individualmente. La tendencia histórica puede reflejar patrones anteriores de comportamiento, permitiendo ganar discernimiento en cuanto a los movimientos a largo plazo de las variables que se desea examinar. Esto permite el desarrollo de modelos de tendencia, útiles en la proyección y pronóstico. Tal como se estudió en el análisis de tendencia. Al analizar el factor estacional se puede determinar si la actividad comercial presenta alguna variación estacional, que pueda considerarse planes futuros.

6.6.1 Aislamiento de la componente estacional. El primer paso en el desglosamiento⁵⁹ es obtener un índice estacional. En donde, primero se debe calcular un promedio móvil centrado. Este promedio móvil elimina los movimientos estacionales recurrentes, así como todo efecto aleatorio. Así, dando un modelo multiplicativo $Y = T * C * S * I$, el promedio móvil elimina S e I y contiene solo T y C. es decir, $PM = T * C$.

⁵⁹ Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS

Ahora es posible calcular la razón por promedio móvil. Para hacerlo, se divide el valor de la serie original Y por el promedio móvil. Este resultado produce los componentes S e I de la serie cronológica.

$$\frac{Y}{PM} = \frac{T * C * S * I}{T * C} = S * I \quad (6- 22)$$

Al dividir los valores de la serie cronológica por el promedio móvil, se llega a la razón por promedio móvil, la cual contiene solo dos componentes S e I . el componente I se eliminara en breve. En resumen, se busca aislar y analizar el componente estacional, extrañamente se comienza eliminando S (e I) calculando el promedio móvil. Luego se restaura el componente estacional calculando la razón por promedio móvil.

Ahora se debe calcular una razón media por promedio móvil para cada uno de los tiempos (por ejemplo para cada uno de los doce meses de un año). Esto se logra promediando la razón por promedio móvil para cada (mes). Por tanto, se deben normalizar estas razones para obtener un índice estacional. Esto se logra dividiendo 12 (debido a que hay doce periodos) por la suma de las razones promedio por promedio móvil, este valor es la razón normalizada.

Esta razón normalizada se multiplica por cada una de las razones promedio, para obtener el índice estacional para cada periodo. Este proceso de eliminación ha eliminado toda actividad irregular, dejando solo el factor estacional.

Los índices se pueden utilizar para desestacionalizar los datos, esto tiene el efecto de eliminar la variación estacional de una serie para determinar lo que serán los valores ante la ausencia de variaciones estacionales. Este valor desestacionalizado se halla dividiendo el valor real durante el (mes) por el índice

estacional en dicho (mes). Los valores desestacionalizados también denominados corregidos estacionalmente, porque dicen lo que los valores serían si se corrigieran las influencias estacionales.

6.6.2 Aislamiento de la variación cíclica. La obtención de la variación cíclica⁶⁰ es, sin lugar a dudas, la que entraña más problemas. Esto es así por tres razones como mínimo. En primer lugar no siempre existe. En segundo lugar hace falta mucha información, series muy largas, para que pueda detectarse la presencia de esta componente. En tercer lugar, es la menos sistemática, pues los ciclos, cuando existen, no siempre tienen la misma longitud y, además, se puede dar el caso de que se superpongan más de un ciclo de distintas longitudes de onda. Todo ello hace que sea muy frecuente el tratamiento de las series en el que se prescindir del estudio separado de los ciclos y, en su lugar, se trabaja con la componente mixta ciclo-tendencia, a la que también se le conoce como extra estacional. Pero para no dejar olvidada esta componente vamos a seguir con el enfoque multiplicativo usado hasta el momento e indicar los pasos a seguir para su obtención.

El componente cíclico puede identificarse primero obteniendo la tendencia y el componente estacional tal y como se describió anteriormente. Se multiplica la norma estadística multiplicando la proyección de tendencia por el índice estacional. A esta se le denomina la norma porque representa los valores que ocurrirán si solo estuvieran presentes la tendencia y las variaciones estacionales.

Los componentes cíclicos e irregulares se obtienen dividiendo los datos originales por la norma estadística, la cual contiene T y S, es decir, debido a que $Y = T * S * C * I$

⁶⁰John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

$$\frac{Y}{T * S} = \frac{T * C * S * I}{T * S} = C * I \quad (6- 23)$$

Luego los resultados se multiplican por 100 para expresar la respuesta en forma de porcentaje, estos resultados contienen tanto componentes cíclicos como irregulares, este último se elimina tomando un promedio móvil de cuatro periodos, dejando solo el factor cíclico. Los valores de la serie de tiempo constarían de $Y = T * C * I$. los componentes C e I se encuentran directamente al dividir solo por los valores de tendencia.

$$\frac{Y}{T} = \frac{T * C * I}{T} = C * I \quad (6- 24)$$

6.6.3 Variación irregular. La última componente de una serie cronológica es la variación irregular⁶¹. Después que se han eliminado las variaciones de tendencia, cíclica y estacional de una serie cronológica, todavía queda un factor impredecible. Típicamente, la variación irregular se presenta en intervalos cortos y sigue un patrón aleatorio. Debido a lo impredecible de la variación irregular, no se trata de describirla matemáticamente. Sin embargo a menudo se pueden aislar sus causas.

6.7 EJERCICIOS RESUELTOS

1. Ejercicio datos anuales

En la tabla 54 se listan los datos de producción en barriles por año de petróleo (BOPA), del pozo llanito 10 desde el año 1981 hasta el año 1987, utilizando las series cronológicas cual será la producción real del pozo en 1995.

⁶¹John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.

TABLA 54. Datos de producción (Ejercicio 1)

Año	Datos de producción anual (y = q_i)
1981	415263
1982	333381
1983	241525
1984	195634
1985	146399
1986	111309
1987	86121

Solución

Cuando los datos históricos (q_i) están dados anualmente, la ecuación del método multiplicativo de series cronológicas no contiene la componente estacional (S_i), quedando así:

$$q_i = T_i * C_i * I_i$$

La componente de tendencia (T_i) se calcula por mínimos cuadrados, mediante la técnica de regresión. Utilizando los datos de la tabla 55. La variable (C_i) representa los cambios cíclicos que ocurren en períodos de uno, dos o más años, que como será demostrado más adelante, para datos anuales, estas variaciones son despreciables o insignificantes en la producción de un pozo, si el índice esta entre 0.95 y 1.05 por lo tanto esta componente será aproximada a 1. La componente irregular está representada por la letra (I_i), la cual no se puede cuantificar por ser una variable aleatoria que se presenta al azar y no tiene ningún sentido de pronóstico.

TABLA 55. Desarrollo estadístico para datos de producción anual

Año	(y = q_i)	Log y	t (años)	t log y	t²	n
1981	415263	5.618323	0	0	0	1
1982	333381	5.522940	1	5.522940	1	2
1983	241525	5.382962	2	10.765924	4	3
1984	195634	5.284232	3	15.854198	9	4
1985	146399	5.165538	4	20.662152	16	5
1986	111309	5.046531	5	25.232651	25	6
1987	86121	4.935109	6	29.610654	36	7
Σ	1526632	36.956136	21	107.64852	91	

Utilizando mínimos cuadrados, ecuación (6 – 18) y (6 – 19), se hallan las constantes c y b.

$$36.956136 = 7 \log c + 21 \log b$$

$$107.64852 = 21 \log c + 91 \log b$$

Se tienen dos ecuaciones con dos incógnitas, resolviendo simultáneamente se halla que c y b son:

$$b = 0.7673685$$

$$c = 421144.2387$$

Tomando la analogía de la ecuación estadística y la ecuación de producción de declinación exponencial, se hallan los parámetros de interés q_i y D_i.

$$Y_t = cb^t = q$$

$$q = q_i e^{-D_i t}$$

Remplazando los valores hallados y haciendo el procedimiento matemático, se obtienen los parámetros de interés:

$$D_i = 0.26478802$$

$$q_i = c = 421144.2387$$

De esta manera se halla la ecuación de producción para el ejercicio, $q = T$.

$$q = 421194,2387 * e^{-0.26478802t} = T \dots$$

el método de mínimos cuadrados prevé valores de T_i de tendencia “ajustados” para cada año en la serie, se puede eliminar con facilidad la componente (T) del modelo multiplicativo, por tanto la componente de la tendencia se puede eliminar mediante una división así:

$$\frac{q_t}{T_i} = C_i * I_i$$

Las proporciones entre los datos observados y los valores de las tendencias se llaman relativas cíclicas irregulares. Estos valores que fluctúan alrededor de una base de 1, muestran tanto la actividad cíclica como irregular en la serie. En la tabla 56 se encuentran los pesos insignificativos que tienen estos dos componentes sobre la producción, pues ninguna de estas relativas es menor de 0.95, ni mayor a 1.05. Además en la figura 29 se observa un ciclo regular que pueda ser estandarizado para hacer los pronósticos de la producción futura.

Si se llegara al caso en el cual se observara un ciclo de amplitud determinada, por ejemplo, tres años, se calculan los promedios aritméticos o medianas de las

relativas para los múltiplos anuales de tres, y con este parámetro se hará el pronóstico.

Si se considera que los ciclos del ejercicio son significativos, en la figura 29 se observa una amplitud aproximada de dos años, lo que llevaría a los siguientes promedios.

TABLA 56. Obtención de las relativas cíclicas – irregulares

Año	(y = q_i)	t (años)	T_i	q_i / T_i	C_i * I_i Promedio
1981	415263	0	421149.24	0.986	0.9915
1982	333381	1	323176.70	1.031	1.012
1983	241525	2	247995.65	0.974	0.9915
1984	195634	3	190304.08	1.012	1.012
1985	146399	4	146033.37	1.002	0.9915
1986	111309	5	112061.42	0.993	1.012
1987	86121	6	85992.42	1.001	0.9915

$$Años\ impares = \frac{0.986 + 0.974 + 1.002 + 1.001}{4} = 0.9915$$

$$Años\ pares = \frac{1.031 + 1.012 + 0.993}{3} = 1.012$$

De esta forma, se puede hacer un pronóstico más real de la producción futura del pozo o del yacimiento, para un año n, se tendría que multiplicar la tendencia de dicho año por su relativa cíclica irregular que le corresponda, según sea n par o impar.

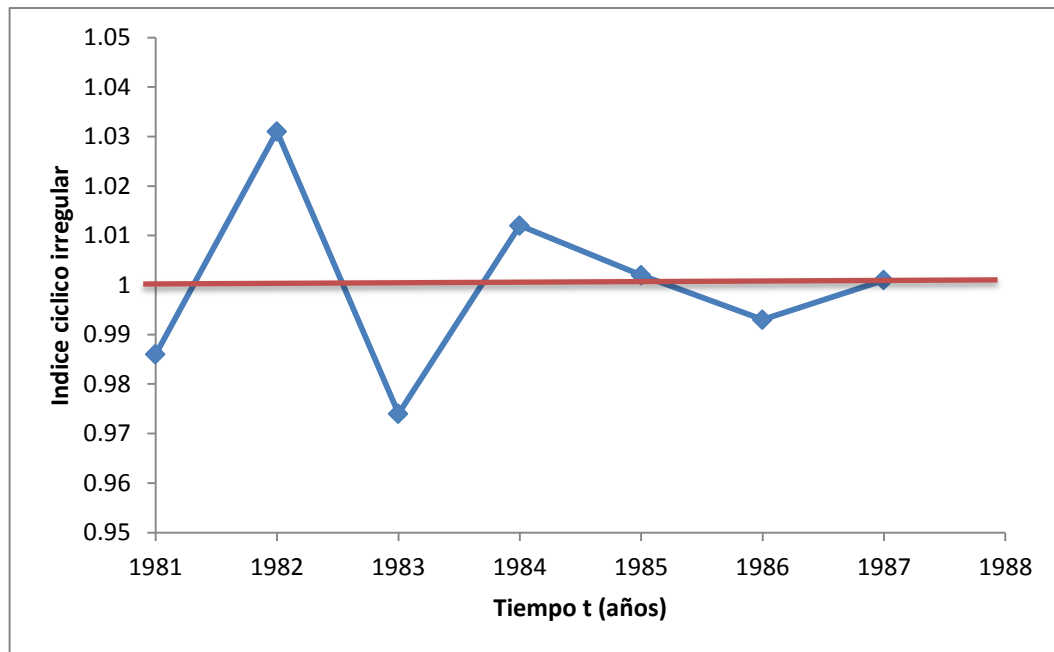
Para el caso del ejercicio, se desea determinar la producción futura para el año 1995, ($t = 14$) la producción pronosticada real será:

$$q_i = T_i * C_i * I_i = T_{(95)} * C_{(95)} = T_{(t=14)} * C_{(impar)}$$

$$q_{(95)} = 421194,2387 * e^{-0.26478802 * 14} * (0.9915)$$

$$q_{(95)} = 10251.44 \text{ Bbls}$$

Figura 29. Ciclos de producción futura (Ejercicio 1)



Con datos anuales, ya no se continua la “descomposición” de las series cronológicas. Aunque para datos anuales la cíclica irregular es despreciable, se observa en este ejercicio, un cambio leve en el pronóstico de la producción por regresión de mínimos cuadrados, evaluados por la función de declinación exponencial, comparado con el ajuste de series cronológicas, o sea, la declinación exponencial pronostica una producción anual en 1995 de 10339.32 Bbls, mientras

que el ajuste por series cronológicas pronostica 10251.44 Bbls, es decir 87.88 Bbls menos.

2. Ejercicio datos mensuales

En la tabla 57 se listan los datos de producción en barriles por mes de petróleo (BOPM), del pozo Orito 15 desde enero de 1981 hasta diciembre de 1987. Se quiere estimar la producción del pozo para el mes de enero de 1990 y para el mes de febrero del 2000.

Solución

Para series cronológicas mensuales como estas, el modelo multiplicativo incluye la componente estacional (S_i), columna (9), además de la tendencia (T_i), columna (5), que ha sido calculada por la técnica de regresión de mínimos cuadrados. Las componentes cíclica e irregular, se observan en la columna (11).

El modelo se expresa como:

$$q_i = T_i * S_i * C_i * I_i$$

Para lograr el objetivo básico de pronosticar los futuros movimientos mensuales, se debe descomponer la serie cronológica, aislando el componente estacional, para eliminarlo de la serie y calcular la tendencia por mínimos cuadrados, promedios móviles o suavización exponencial, para aislar la componente cíclica – irregular. Con estos componentes aislados, se puede proceder a realizar extrapolaciones o pronosticar la futura producción mensual del pozo, con resultados cercanos a la realidad. La componente irregular no es aislada, por ser un parámetro incierto, producido al azar, y por tanto inadecuado para pronósticos. En este caso se deja aliada al componente cíclico, el cual, en series cronológicas con características como las de producción de petróleo, es despreciable porque no difieren de la unidad (1) en más de 5 %.

TABLA 57.Descomposición de la serie (Ejercicio 2)

Año (1)	q_i (BOPM) (2)	t_i (3)	T_i (4)	Total Móvil 12 meses (5)	Promedio Móvil Centrado (T.C) (6)	$(q_i / T.C) = S_i$ (7) = (2)/(6)	S_i (8)	$(q_i / S_i) = T.C_i$ (9) = (2)/(8)	$(T.C_i / T_i) = C.I$ (10) = (9)/(4)	Cíclica C (11)
1981 ENE	38688	0	39501				0.9891	391114.3464	0.9902	
FEB	37833	1	38639				1.0380	36447.9769	0.9433	0.946
MAR	37251	2	37796				1.0861	34297.9468	0.9074	0.915
ABR	36283	3	36971				1.0853	33431.3093	0.9042	0.967
MAY	35052	4	36164				1.1316	31432.4850	1.1505	1.026
JUN	35052	5	35375	415263			1.1034	31767.2648	0.8980	1.014
JUL	33790	6	34603	406320	34233	0.9871	0.8817	38323.6929	1.1075	1.067
AGO	33948	7	33848	398412	33531	1.0124	0.8672	39146.6789	1.1565	1.132
SEP	32730	8	33109	391665	32920	0.9942	0.8927	36664.0528	1.1074	1.105
OCT	33348	9	32387	384426	32337	1.0313	0.9815	339765665	1.0491	1.061
NOV	31482	10	31680	377401	31743	0.9918	0.9579	32865.6436	1.0374	1.021
DIC	29282	11	30989	371568	31207	0.9385	0.9852	29728.9890	0.9593	0.987
1982 ENE	29745	12	30312	365560	30714	0.9684	0.9891	30072.7934	0.9921	0.979
FEB	29925	13	29651	358528	30170	0.9919	1.0380	28829.4797	0.9723	0.976
MAR	30504	14	29004	351837	29599	1.0306	1.0861	28085.8116	0.9683	0.963
ABR	29044	15	28371	343590	28976	1.0023	1.0853	26761.2642	0.9433	0.941
MAY	28544	16	27752	337898	28395	1.0052	1.1316	25224.4609	0.9089	0.934
JUN	29219	17	27146	333381	27970	1.0447	1.1034	26480.8773	0.9755	1.012
JUL	27782	18	26554	326757	27506	1.0100	0.8817	31509.5638	1.1866	1.136
AGO	26916	19	25974	319095	26911	1.0002	0.8672	31037.8229	1.1949	1.181
SEP	26039	20	25407	310300	26225	0.9929	0.8927	29168.8137	1.1481	1.130
OCT	25101	21	24853	302628	25639	0.9828	0.9815	25574.1212	1.0290	1.078
NOV	25790	22	24311	293688	24847	1.0380	0.9579	26923.4784	1.1075	1.075

Año (1)	q_i (BOPM) (2)	t_i (3)	T_i (4)	Total Móvil 12 meses (5)	Promedio Móvil Centrado (T.C) (6)	$(q_i / T.C) = S_i$ (7) = (2)/(6)	S_i (8)	$(q_i / S_i) = T.C.I$ (9) = (2)/(8)	$(T.C.I / T_i) = C.I$ (10) = (9)/(4)	Cíclica C (11)
DIC	24772	23	23780	284180	24078	1.0288	0.9852	25144.1332	1.0574	1.057
1983 ENE	23121	24	23261	275986	23340	0.9906	0.9891	23375.7962	1.0049	1.002
FEB	22263	25	22754	268630	22692	0.9811	1.0380	21447.9769	0.9426	0.947
MAR	21709	26	22257	261571	22092	0.9827	1.0861	19988.0306	0.8980	0.911
ABR	21372	27	21771	255212	21533	0.9925	1.0853	19692.2510	0.9045	0.880
MAY	19604	28	21296	247547	20948	0.9358	1.1316	17324.1428	0.8135	0.847
JUN	19711	29	20831	241525	20378	0.9672	1.1034	17863.8753	0.8576	0.905
JUL	19588	30	20377	236444	19915	0.9836	0.8817	2216.1733	1.0902	1.042
AGO	19560	31	19932	231819	19511	1.0025	0.8672	22555.3506	1.1316	1.111
SEP	18980	32	19497	192876	17696	1.0725	0.8927	21261.3420	1.0904	1.078
OCT	18742	33	19072	187956	15868	1.1811	0.9815	19095.2624	1.0012	1.027
NOV	18125	34	18655	185852	15575	1.1637	0.9579	18921.5993	1.0143	1.018
DIC	18750	35	18248	182121	15332	1.2230	0.9852	19031.6687	1.0429	1.030
1984 ENE	18040	36	17850	178222	15014	1.2015	0.9891	18238.8030	1.0218	1.015
FEB	17638	37	17461	143615	13410	1.3152	1.0380	16992.2929	0.9732	1.0974
MAR	17234	38	17079	139324	11789	1.4619	1.0861	15867.7838	0.9291	0.935
ABR	16452	39	16707	136207	11480	1.4331	1.0853	15158.9422	0.9673	0.923
MAY	17500	40	16342	132663	11203	1.5621	1.1316	15464.8286	0.9463	0.926
JUN	15780	41	15986	128072	10864	1.4710	1.1034	14482.5086	0.9059	0.973
JUL	15689	42	15637	129002	10711	1.4647	0.8817	17794.0342	1.1379	1.077
AGO	15047	43	15295	125063	10586	1.4214	0.8672	17351.2454	1.1344	1.125
SEP	14689	44	14962	120942	10250	1.4330	0.8927	16454.5760	1.0997	1.105
OCT	15625	45	14635	117333	9928	1.5738	0.9815	15919.5109	1.0878	1.085

Año (1)	q_t (BOPM) (2)	t_t (3)	T_t (4)	Total Móvil 12 meses (5)	Promedio Móvil Centrado (T.C) (6)	$(q_t / T.C) = S.I$ (7) = (2)/(6)	S_t (8)	$(q_t / S_t) = T.C.I$ (9) = (2)/(8)	$(T.C.I / T_t) = C.I$ (10) = (9)/(4)	Cíclica C (11)
NOV	14581	46	14316	112181	9563	1.5247	0.9579	15221.8394	1.0633	1.060
DIC	14159	47	14003	108190	9182	1.5420	0.9852	14371.7012	1.0263	1.037
1985 ENE	13970	48	13698	104748	8872	1.5746	0.9891	14123.9510	1.0311	1.018
FEB	13699	49	13399	100951	8571	1.5983	1.0380	13197.4952	0.9849	0.981
MAR	13120	50	13106	97548	8271	1.5863	1.0861	12079.9190	0.9215	0.938
ABR	12836	51	12820	93471	7959	1.6128	1.0853	11827.1446	0.9225	0.909
MAY	12348	52	12541	90144	7651	1.6139	1.1316	10914.8767	0.8703	0.887
JUN	11989	53	12267	86837	7374	1.6258	1.1034	10865.5066	0.8857	0.950
JUL	12247	54	11999	83492	7097	1.7256	0.8817	13890.2121	1.1576	1.077
AGO	11250	55	11737	79921	6809	1.6522	0.8672	12972.7860	1.1053	1.117
SEP	11286	56	11481	77170	6545	1.7243	0.8927	12542.5451	1.1012	1.089
OCT	11548	57	11230	74441	6317	1.8281	0.9815	11765.6648	1.0477	1.067
NOV	11254	58	10986	71661	6088	1.8486	0.9579	11748.6168	1.0694	1.053
DIC	10852	59	10746	68152	5826	1.8627	0.9852	11015.0223	1.0250	1.035
1986 ENE	10652	60	10511	65068	5551	1.9140	0.98891	10742.0889	1.0220	1.004
FEB	10128	61	10282	62767	5326	1.9016	1.0380	9757.2254	0.9489	0.967
MAR	10369	62	10058	60631	5137	2.0184	1.0861	9747.0030	0.9492	0.948
ABR	10107	63	9838	57857	4933	2.0488	1.0853	9312.6324	0.9466	0.930
MAY	9568	64	9623	54563	4684	2.0427	1.1316	8455.2845	0.8786	0.880
JUN	8480	65	9412	51747	4430	1.9142	1.1034	7685.3362	0.8165	0.910
JUL	9163	66	9208	49116	4203	2.1801	0.8817	10392.4237	1.1286	1.055
AGO	8949	67	9007	47057	4007	2.2333	0.8672	10319.4188	1.1457	1.143
SEP	9050	68	8810	44486	3814	2.3728	0.8927	10137.7842	1.1507	1.124

Año	q_i (BOPM)	t_i	T_i	Total Móvil 12 meses	Promedio Móvil Centrado (T.C)	$(q_i/T.C) = S_i$ (7) = (2)/(6)	S_i (8)	$(q_i/S_i) = T.C.I$ (9) = (2)/(8)	$(T.C.I/T_i) = C.I$ (10) = (9)/(4)	Cíclica C
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
OCT	8874	69	8618	41879	3599	2.4657	0.9815	9041.2633	1.0491	1.059
NOV	7960	70	8430	39661	3398	2.3425	0.9579	8309.8444	0.9857	1.002
DIC	8036	71	8246	38369	3251	2.719	0.9852	8156.7194	0.9892	0.992
1987 ENE	7994	72	8066	36271	3110	2.5704	0.9891	8082.0948	1.0020	0.995
FEB	8069	73	7890	34116	2933	2.7511	1.0380	7773.6030	0.9852	0.976
MAR	7795	74	7718	31966	2753	2.8315	1.0861	7177.0555	0.9299	0.940
ABR	7500	75	7546	29824	2575	2.9126	1.0853	6910.5316	0.9150	0.910
MAY	7350	76	7385	28353	2424	3.0322	1.1316	6495.2280	0.8795	0.894
JUN	7188	77	7223	26562	2288	3.1416	1.1034	6514.4100	0.9019	0.954
JUL	7065	78	7066				0.8817	8012.9296	1.1340	1.076
AGO	6794	79	6911				0.8672	7834.4095	1.1336	1.136
SEP	6900	80	6761				0.8927	7729.3603	1.11432	1.140
OCT	6732	81	6613				0.9815	6858.8694	1.0372	1.066
NOV	6489	82	6469				0.9579	6774.1935	1.0472	1.033
DIC	6245	83	6328				0.9852	6338.8415	1.0017	

El cálculo de la tendencia se hace sometiendo los datos históricos originales o reales al proceso matemático de regresión, usando mínimos cuadrados. Este procedimiento no difiere en absoluto del método tradicional que se ha venido utilizando en la industria petrolera, que en este caso conlleva el modelo exponencial.

El procedimiento del cálculo de la tendencia es similar al que se realizó en el ejercicio (1). Obteniéndose la siguiente expresión:

$$q = 39501.34187 * e^{-0.0220646t} = T \dots$$

La expresión anterior fue utilizada para calcular la columna (4), luego de remplazar el valor t_i de la columna (3), como se observa en la tabla 57 del ejercicio 2. En la figura 30 se observa la tendencia superpuesta sobre el tren de producción real.

En el cálculo del índice estacional (S_i), es importante aislar los movimientos estacionales en la serie cronológica de la producción, porque se puede, con facilidad, ajustar y mejorar las extrapolaciones de la tendencia para fines de pronóstico. Si como se supone a menudo, los movimientos estacionales son razonablemente constantes con el tiempo, la construcción del índice estacional se puede ilustrar con las tablas 57 y 58 para empezar, se obtiene una serie de totales móviles de 12 meses (o trimestres, si la producción esta reportada en acumulativos de tres meses). Pero, como lo indica la columna (5) de la tabla 57, al registrar estos totales móviles, los resultados se centran entre los dos meses centrales que constituyen cada total móvil respectivo. Por ejemplo, el primer total móvil que consiste en la sumatoria de la producción de los 12 meses de 1981, se registra entre junio y julio de 1981; el segundo total móvil que consiste en los meses de febrero de 1981 hasta enero de 1982, se registra entre julio y agosto de 1981, y así sucesivamente.

Para encontrar o centrar estos resultados dentro de un mes en particular, se obtienen totales móviles de 2 meses de los totales móviles de 12 meses y se dividen en 24 como se indica en la columna (6) de la tabla 57, obteniéndose el producto de las variable o componente cíclica y tendencia de la serie. Después, los datos originales, columna (2), se dividen entre los respectivos promedios móviles centrados, columna (6), y arrojan las componentes estacional e irregular, columna (7).

$$\frac{q_t}{T_i * C_i} = S_i * I_i$$

Para formar el índice estacional, se reordenan los datos d la columna (7) de acuerdo con los valores indicados en la tabla58. En la tabla 58 se ve que para cada mes, se pueden eliminar las variaciones irregulares, si se utiliza la mediana de las diversas proporciones obtenidas con los promedios móviles. Como se indica en la tabla 58, después se ajustan estos valores de la mediana, de modo que el valor total de los índices estacionales en el año sea 12.0 y el valor promedio de cada índice estacional (mensual) sea de 1.0. Por tanto, un índice estacional de 0.98 para el mes de enero, indica que la producción de enero es el 98% del promedio mensual. El índice estacional se puede representar como:

$$Indice Estacional = S_i = \frac{12 (Mediana)}{16.8391}$$

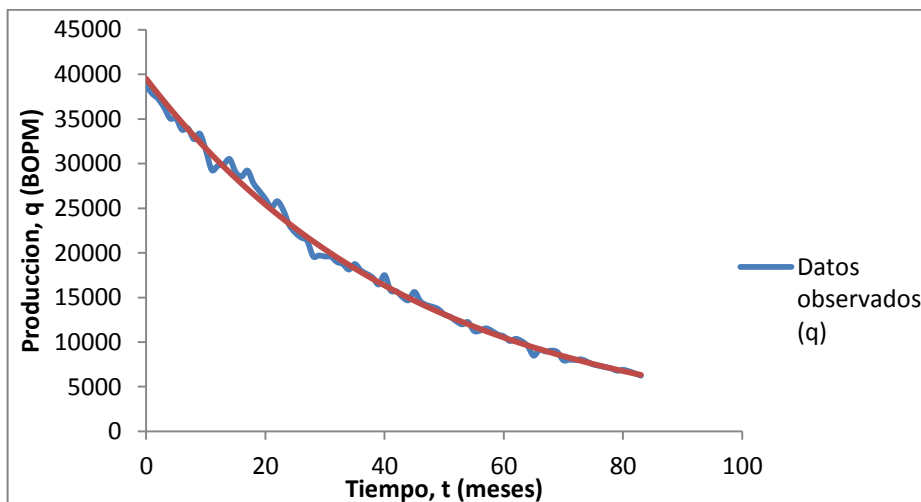
TABLA 58. Cálculo del índice estacional S_i (Ejercicio 2)

AÑOS									
Mes	1981	1982	1983	1984	1985	1986	1987	Mediana	Índice estac.
ENE		0.96	0.99	1.20	1.57	1.91	2.57	1.3880	0.9891
FEB		0.99	0.98	1.31	1.59	1.90	2.75	1.4567	1.0380
MAR		1.03	0.98	1.46	1.58	2.01	2.83	1.5241	1.0861
ABR		1.00	0.99	1.43	1.61	2.04	2.91	1.5229	1.0853
MAY		1.00	0.93	1.56	1.61	2.04	3.03	1.5880	1.1316
JUN		1.04	0.96	1.47	1.62	1.91	3.14	1.5484	1.1034
JUL	0.98	1.01	0.98	1.46	1.72	2.18		1.2373	0.8817
AGO	1.01	1.00	1.00	1.42	1.65	2.23		1.2169	0.8672
SEP	0.99	0.99	1.07	1.43	1.72	2.37		1.2527	0.8927
OCT	1.03	0.98	1.18	1.57	1.82	2.46		1.3774	0.9815
NOV	0.99	1.03	1.16	1.52	1.84	2.34		1.3442	0.9579
DIC	0.93	1.02	1.22	1.54	1.862	2.471		1.3825	0.9852
Total								16.8391	12.0163

El índice estacional se utiliza, junto con la tendencia, para aislar la componente cíclica, que para este caso de aplicación, se mantendrá integrada con la componente irregular, si gráficamente (figura 31) son identificables los ciclo, en caso contrario, se deberá aislar la variable o componente cíclica de la irregular aplicando promedios móviles para 3 meses, con doble peso para el segundo mes, eliminando las distorsiones que la componente irregular le puede ocasionar a los ciclos.

Para desestacionar los datos solo se divide cada valor observado en la serie cronológica mensual, columna (2), entre el índice estacional para cada mes, columna (8). Los resultados se muestran en la columna (9) de la tabla 57.

Figura 30. Tendencia de Producción (Ejercicio 2)



En términos del modelo multiplicativo de series cronológicas, la desestacionalizada se expresa como:

$$\frac{q_t}{S_i} = \frac{T_i * S_i * C_i * I_i}{S_i} = T_i * C_i * I_i$$

El siguiente paso consiste en eliminar el componente de tendencia, la que llevaría a una serie de relativas cíclicas – irregulares, $C_i * I_i$. Entonces con los datos de la tendencia, columna (4), se dividen los datos desestacionalizados, columna (9), y se obtienen las relativas cíclicas irregulares de la columna (10), es decir:

$$\frac{T_i * C_i * I_i}{T_i} = C_i * I_i$$

Estas relativas cíclicas – irregulares se trazan en la figura 31, donde se observa que no hay una identificación clara de los ciclos, por tal razón hay necesidad de aislar los ciclos de la relativa cíclica – irregular $C_i * I_i$, lo cual se logra aplicándole a la columna (10) un promedio móvil de 3 meses, con doble peso para el segundo,

eliminando de esta manera la influencia de la componente irregular que distorsiona los ciclos, columna (11).

En la figura 31 se grafica la componente cíclica aislada, observándose ciclos claros de una amplitud de 13 meses, que para efectos de simplicidad en los pronósticos se aproximará a un año (12 meses). De esta manera, la componente cíclica para pronósticos se hará ANUAL con un peso apropiado para cada mes, el cual será calculado como el promedio de los índices cíclicos de cada mes respectivo, es decir, se suman todos los índices cíclicos de enero y se dividen en el número de índice sumados, así:

$$C_{enero} = \frac{C_{enero}}{n}$$

Donde “n” es el número de eneros sumados. Para el ejercicio se calculara el índice cíclico para los meses de enero y febrero, los otros meses se calculan de forma similar.

$$Enero = \frac{0.979 + 1.002 + 1.015 + 1.018 + 1.004 + 0.995}{6} = 1.00217$$

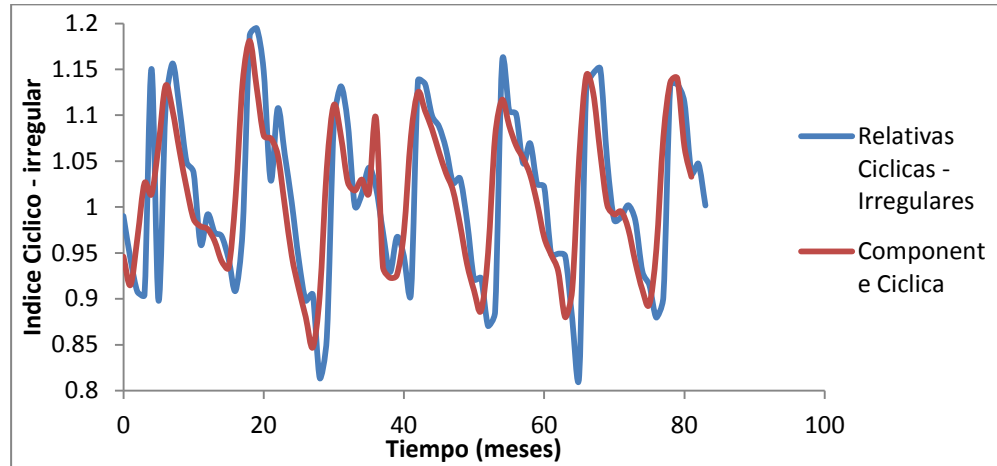
$$Febrero = \frac{0.946 + 0.976 + 0.947 + 0.974 + 0.981 + 0.976 + 0.976}{7}$$

$$Febrero = 0.96671$$

Así, para efectos de pronósticos futuros, cuando se vaya a estimar la producción, por ejemplo, para el mes de enero de un año cualquiera, se deberá multiplicar la componente de tendencia (T_i) y la componente estacional de enero (S_i) por el índice cíclico del mes de enero (1.00217), que corresponde a la componente cíclica de enero. El hecho de que la componente cíclica se haya presentado anual

en el ejercicio, no significa que ésta amplitud se presentará anual en todos los casos, pues los ciclos tienen amplitudes de uno, dos o más años.

Figura 31. Índice cíclico – irregular (Ejercicio 2)



Pronósticos

- Como se quiere estimar la producción del pozo, para el mes de enero de 1990, se debe calcular la componente de tendencia, utilizando la expresión encontrada con mínimos cuadrados, para un $t = 108$, sabiendo que $t = 0$ corresponde a Enero de 1981, mediante la expresión:

$$T_{(108)} = 39501.34187 * e^{-0.0220646(108)}$$

$$T_{(108)} = 3645 \text{ BOPM}$$

Luego se utiliza el modelo multiplicativo de series cronológicas:

$$q_i = T_i * C_i * S_i$$

Remplazando $T_{(108)}$, $C_{(enero)}$ y $S_{(enero)}$, por los valores correspondientes así:

$$q_{(enero 1990)} = 3645 * 1.00217 * (0.9891)$$

$$q_{(enero 1990)} = 3613 \text{ BOPM}$$

O sea, que en el mes de enero de 1990, se espera una producción de 3613 barriles de petróleo por mes, y no 3645 que sería el pronóstico obtenido utilizando la correlación empírica tradicional que dio el método exponencial.

- También se quiere estimar la producción del pozo, para el mes de febrero del año 2000, para un $t = 229$, la producción real del pozo será:

$$T_{(229)} = 39501.34187 * e^{-0.0220646(229)}$$

$$T_{(229)} = 252.47 \text{ BOPM}$$

Utilizando el modelo multiplicativo de series cronológicas:

$$q_i = T_i * C_i * S_i$$

Remplazando $T_{(229)}$, $C_{(febrero)}$ y $S_{(febrero)}$, por los valores correspondientes así:

$$q_{(febrero 2000)} = 252.47 * 0.96671 * (1.038)$$

$$q_{(febrero 2000)} = 253 \text{ BOPM}$$

La producción para febrero del 2000 solo difiere en un barril del pronóstico tradicional.

6.8 EJERCICIOS PROPUESTOS

1. En la tabla 59 se listan los datos de producción en barriles por año de petróleo (BOPA), del pozo Emirato 20 desde el año 1998 hasta el año 2009, cuál será la producción real del pozo en 2013 y 2020.

TABLA 59. Datos de producción (Ejercicio 1)

Año	Datos de producción anual (y = q_i)	Año	Datos de producción anual (y = q_i)
1998	995263	2004	806121
1999	933381	2005	798345
2000	921525	2006	795948
2001	895634	2007	788990
2002	846399	2008	777837
2003	811309	2009	750873

2. En la tabla60se listan los datos de producción en barriles por trimestre de petróleo (BOPT), del pozo Maxiwell 12 desde el primer trimestre del año 1991 hasta el cuarto trimestre del año 1994. Se quiere estimar la producción del pozo para el segundo trimestre del año 1997 y para el tercer trimestre del año 2000.

TABLA 60. Datos de producción (Ejercicio 2)

Año - Trimestre	q_i (BOPT)	Año - Trimestre	q_i (BOPT)
1991 – 1	998688	1993 – 1	992730
1991 – 2	997833	1993 – 2	993348
1991 – 3	997251	1993 – 3	991482
1991 – 4	996283	1993 – 4	899282
1992 – 1	995052	1994 – 1	898745
1992 – 2	995052	1994 – 2	897925
1992 – 3	993790	1994 – 3	887504
1992 – 4	993948	1994 – 4	885044

7. CONCLUSIONES

- La industria de los hidrocarburos, debido a su alto riesgo y complejidad tecnológica, requiere de decisiones eficientes y bien documentadas. La estadística cuenta con las bases necesarias para tales decisiones. Las soluciones a estos problemas son esenciales para la optimización de los procesos y operaciones, así como para la maximización de las ganancias.
- La estadística es una herramienta valiosa en el área de yacimientos, en el momento de cuantificar que tan probable es que un nuevo método de recobro sea exitoso. Tales aplicaciones y muchas otras, hacen del continuo mejoramiento en la toma de decisiones en la industria petrolera.
- La estadística es un conjunto de técnicas que partiendo de la observación de fenómenos, permiten al investigador la estimación de los factores que afectan o alteran el comportamiento de las variables en un proceso, como por ejemplo en el área de fluidos o de lodos de perforación.
- La estadística evoluciona a la par con el desarrollo de las matemáticas, en donde se hace importante la estimación de procesos futuros, aplicando la estadística para la predicción de caudales futuros a través del tiempo, ya que se tienen en cuenta otros factores que la ecuación de declinación exponencial de producción no contempla.
- Esta obra fue escrita con la idea de proporcionar al lector no solamente la teoría, que es fundamental, sino también el enfoque práctico, que es indispensable para desarrollar los modelos planteados y poder llegar a resultados útiles.

8. RECOMENDACIONES

- Emplear una herramienta multimedia para la implementación de este manual en la pedagogía de la asignatura de estadística aplicada de la escuela de ingeniería de petróleos - UIS.
- La manipulación estadística de datos requiere el empleo constante de operaciones aritméticas. Es por ello que se recomienda al lector el uso de paquetes como el MINITAB, el SAs y el CBS, para el análisis de estos datos en los ejercicios resueltos y propuestos para obtener un mejor análisis del resultado.
- Utilizar un sistema computacional interactivo para la ayuda de toma de decisiones (Arboles de decisiones), se recomienda al lector implementarla herramienta SDSS (Crossland et al. 1995) que se basa en la tecnología SIG. O también el SPSS Decisión Trees 17.0.
- Desarrollar los ejercicios propuestos de todos los capítulos y proponer más ejercicios aplicados a petróleos, para construir un solucionario de estadística aplicada a la industria de los hidrocarburos. Con el cual los estudiantes de ingeniería de petróleos y los profesionales, puedan contar para terminar de obtener bases sólidas y profundizar en esta ciencia ingenieril.

BIBLIOGRAFIA

- Allen L. Webster, (2000). Estadística Aplicada a los negocios y a la Economía. Tercera Edición. Ed. McGraw – Hill. Interamericana S.A. Colombia
- Humberto Gutierrez Pulido / Román de la Vara Salazar, (2004). Análisis y Diseño de Experimentos. Editorial Mc Graw – Hill. Segunda Edición. México.
- Ignacio Vélez Pareja. (2003). Decisiones Empresariales bajo Riesgo e Incertidumbre. Primera Edición. Editorial Norma. Colombia.
- Jiménez Barajas Gilberto (1995). Ajuste de Curvas de Declinación por Series Cronológicas. Tesis (Ingeniero de Petróleos). Bucaramanga – UIS.
- John E. Hanke / Arthur G. Reitsch, (1995). Estadística para Negocios. Segunda Edición. Editorial Erwin. España.
- Richard I. Levin / David S. Rubin (1996). Estadística para Administradores. Editorial Prentice Hall. Hispanoamericana S.A. Sexta Edición. México.
- Wilfrido Moreno T. (1993). Aplicaciones al Diseño y Análisis de Experimentos. Bucaramanga: UIS. Departamento de Ingeniería Industrial.

ANEXOS

ANEXO A. Glosario estadístico

Amplitud: es la distancia entre el valor máximo observado y el valor mínimo observado en un conjunto o distribución de datos.

Análisis de varianza (ANOVA): un procedimiento estadístico para determinar si las medias de tres o más poblaciones son iguales.

Al azar, estocástico: este término representa una idea que debe ser expresada en términos del concepto de probabilidad. Se tiene la noción de que un fenómeno ocurre en forma aleatoria cuando no sigue un patrón particular que se pueda describir directamente por ecuaciones. Así, no se puede hacer una predicción perfecta del resultado que se obtendrá del fenómeno. Al decir que un proceso es aleatorio estamos diciendo que sigue alguna distribución de probabilidad.

Coefficiente de correlación: medida de la intensidad de la relación entre dos variables.

Coefficiente de determinación: medida en la cual las variaciones de una variable se pueden atribuir a las variaciones en la otra variable.

Confiabilidad: indica cuán seguros podemos estar que el proceso seguido resulte en valores que representen verdaderamente la población. Se usa más comúnmente con intervalos de confianza. En sentido probabilístico, si tuviéramos una confiabilidad del 95%, decimos que si repitiéramos el proceso muchas veces, en cerca del 95% de las veces obtendríamos resultados que reflejan verdaderamente la realidad. Cerca del 95% de los intervalos así contruidos contendrían el valor desconocido del parámetro.

Datos: valores que se obtienen al observar directamente los resultados de una variable en la muestra o población. Pueden ser numéricos o cualitativos.

Distribución bimodal: distribución de datos que tiene dos modas.

Distribución multimodal: distribución de datos que tiene más de una moda.

Distribución unimodal: distribución de datos que tiene una sola moda.

Encuesta: método de obtener datos de una población o muestra, sin ejercer control alguno sobre los factores que pueden afectar las características de interés o resultados de la encuesta.

Error tipo I: ocurre cuando rechazamos la hipótesis nula siendo ésta cierta.

Error tipo II: ocurre cuando no rechazamos la hipótesis nula siendo ésta falsa.

Error muestral: es la diferencia entre un estadístico muestral y su parámetro poblacional correspondiente.

Error no muestral: son errores que ocurren en la selección, recolección, anotación y tabulación de los datos. Son usualmente resultado de error humano.

Espacio muestral: es el conjunto de todos los posibles resultados de un experimento.

Estadística Descriptiva: método que se utiliza para describir los datos que se han obtenido de la muestra o población. Sirve para presentar una idea de la realidad y para hacer inferencia informal.

Estadístico: elemento que describe una muestra y sirve como una estimación del parámetro de la población correspondiente.

Estadística Inferencial: métodos probabilísticos que se utilizan para tomar decisiones, estimar, predecir o hacer generalizaciones sobre una población basados en una muestra.

Estadística prueba: cantidad calculada de los datos muestrales, que se usa para decidir si se rechaza o no la hipótesis nula. Generalmente, un valor grande de esta estadística es un indicador que nos apunta hacia el rechazo de la hipótesis nula.

Estimación por intervalo: un intervalo dentro del cual es muy probable que se encuentre el parámetro poblacional.

Estimación puntual: un solo valor que se mide a partir de una muestra y se usa como una estimación del parámetro poblacional correspondiente.

Evento: conjunto o colección de uno o más posibles resultados de un experimento. Un evento ocurre cuando cualquier resultado contenido en el evento es observado.

Eventos dependientes. Dos o más eventos serán dependientes cuando la ocurrencia o no-ocurrencia de uno de ellos afecta la probabilidad de ocurrencia del otro (u otros).

Eventos independientes: dos eventos son independientes si el que uno ocurra no afecta la probabilidad de que el otro ocurra.

Experimento: es un proceso que cuando se lleva a cabo resulta en uno y solo uno de los posibles resultados que se podría obtener (probabilidad). Generalmente, son datos observados de los miembros de la población o muestra, ejerciendo control sobre uno o más de los factores que podrían alterar la característica de interés o los resultados del experimento (ciencia).

Grados de libertad: se refiere al número de elementos en los datos que pueden variar libremente.

Hipótesis nula: es una aseveración sobre el valor de un parámetro desconocido de una población. Se presume cierta hasta tanto se demuestre lo contrario. Esta hipótesis se rechaza o no dependiendo del valor de la estadística prueba o del valor p al nivel de significancia deseado.

Hipótesis alternativa: la afirmación sobre la población que debe ser cierta si la hipótesis nula es falsa.

Individuos: se llama unidad estadística o individuo a cada uno de los elementos que componen la población estadística. El individuo es un ente observable que no tiene por qué ser una persona, puede ser un objeto, un ser vivo, o incluso algo abstracto.

Método de mínimos cuadrados: método para ajustar una recta, que hace mínimo el promedio de los errores de las estimaciones de Y , a partir de X .

Método *entre*: un método para estimar la varianza de las poblaciones que produce una estimación válida sólo si la hipótesis nula es cierta.

Método *dentro*: un método para estimar la varianza de las poblaciones que produce una estimación válida sea o no cierta la hipótesis nula.

Muestra: es una parte representativa de la población que se selecciona para ser estudiada, ya que la población es demasiado grande como para analizarla en su totalidad.

Nivel de confianza: indica la probabilidad de que una estimación por intervalo contenga el parámetro poblacional. El cual se expresa como un porcentaje.

Nivel de significancia: probabilidad de rechazar la hipótesis nula cuando es cierta. Probabilidad de cometer un error tipo I. Este nivel es seleccionado por el investigador antes de realizar el experimento. Los valores más comúnmente seleccionados son niveles de .01, .05 y .10.

Parámetro: cualquier característica medible de una población.

Prueba de hipótesis: es un procedimiento por el cual establecemos hipótesis nula y alterna con el fin de resolver un problema. El procedimiento incluye el diseño y selección de la muestra. Luego de tomados los datos de la muestra, se calcula el valor de una estadística prueba. A un nivel de significancia previamente seleccionado, la estadística prueba se compara con el valor obtenido de la tabla de la distribución estadística apropiada. Esa comparación nos lleva a tomar la decisión de rechazar o no la hipótesis nula.

Población: es la recolección completa de todas las observaciones de interés para el investigador.

Regla de decisión: una regla de la prueba de hipótesis que especifica la acción a tomar para cada resultado muestral posible.

Regresión: relación que se da entre una variable independiente y otra dependiente.

Sesgo muestral: es la tendencia a favorecer la selección de ciertos elementos de muestra en lugar de otros.

Suceso: es el resultado posible de un experimento.

Sucesos mutuamente excluyentes. Dos o más sucesos son mutuamente excluyentes, si no pueden ocurrir simultáneamente. Es decir, la ocurrencia de un suceso impide automáticamente la ocurrencia del otro suceso (o sucesos).

Sucesos mutuamente incluyentes. Dos o más sucesos son mutuamente incluyentes, cuando es posible que ocurran ambos. Esto no indica que necesariamente deban ocurrir estos sucesos en forma simultánea.

Significancia: corresponde a la probabilidad de error tipo I que se esta dispuestos a permitir cuando se hace una prueba de hipótesis. Usualmente se expresa como un porcentaje. Los valores más comunes son 1%, 5%, 10%. Una significancia del 5% quiere decir que de cada cien pruebas donde rechazamos la hipótesis nula, nos permitimos la posibilidad de haberla rechazado en 5 ocasiones a pesar de ser cierta. El nivel de significancia se selecciona de acuerdo con una amplia gama de criterios que incluyen el costo de cometer error tipo I y la tradición en el área de contenido sobre el cual se esta haciendo la prueba.

Tabla de contingencia: es una tabla que sirve para clasificar a los miembros de un grupo de acuerdo con algunas características cualitativas o cuantitativas.

Variable: es una característica de la población que se esta analizando en un estudio estadístico.

Variable aleatoria: es una función que asigna un valor numérico a cada suceso elemental del espacio muestral. Es decir, una variable aleatoria es una variable cuyo valor numérico esta determinado por el resultado del experimento aleatorio. La variable aleatoria la notaremos con letras en mayúscula X, Y, \dots y con las letras en minúscula x, y, \dots sus valores.

Variable aleatoria discreta: se dice que una variable aleatoria X es discreta si puede tomar un número finito o infinito, pero numerable, de posibles valores. Una variable aleatoria discreta se obtiene después de sumar o contar; trabaja con números enteros.

Variable aleatoria continua: se dice que una variable aleatoria X es continua si puede tomar un número infinito (no numerable) de valores, o bien, si puede tomar un número infinito de valores correspondientes a los puntos de uno o más intervalos de la recta real. Una variable aleatoria continua se obtiene después de hacer una medición; trabaja con números reales.

ANEXO B. GLOSARIO TECNICO

Almacenamiento: Instalación que cuenta con uno o varios depósitos con la finalidad de acopiar los combustibles líquidos y gaseosos.

Barril: Unidad de medida volumétrica empleada en varios países, entre ellos E.E.U.U. Un barril de petróleo equivale a 159, litros, o sea que un metro cúbico de petróleo equivale a 6,29 barriles.

Cementación: Proceso por el cual se bombea al pozo una mezcla de cemento que al fraguarse o endurecerse proporciona sustentación a la tubería de revestimiento dando hermeticidad contra la filtración de fluidos de formación.

Concesión de explotación: Decisión gubernamental que da derecho de explotar o utilizar una cosa o bien público. En la minería se trata de un título que otorga el derecho de explotar a continuación del descubrimiento de un yacimiento comercializable.

Condensado de gas: Hidrocarburo que se mantiene en estado gaseoso en las condiciones de su depósito natural pero por las altas presiones se licua en las condiciones superficiales normales. En otros países se lo conoce como líquido del gas natural.

Cuenca sedimentaria: Área de corteza terrestre que puede abarcar extensas regiones que han sufrido hundimientos donde se acumulan importantes depósitos de rocas sedimentarias en capas superpuestas que llegan a tener hasta más de 10.000 metros de espesor. Bajo determinadas condiciones y por descomposición de la materia orgánica se pueden generar hidrocarburos.

Emulsión: El petróleo crudo que se produce en un campo se encuentra en la mayoría de los casos mezclado con agua, en cantidades o porcentajes que varían en un rango muy amplio de acuerdo con varios factores, como la vida productiva del pozo, la tasa de producción y la procedencia del agua, entre otros.

Exploración: Es la búsqueda de yacimientos de petróleo y gas y comprende todos aquellos métodos destinados a detectar yacimientos comercialmente explotables.

Explotación (producción): Operación que consiste en la extracción de petróleo y/o gas de un yacimiento.

Factor de recuperación: Porcentaje del petróleo extraído de un yacimiento con relación al volumen total contenido en el mismo.

Gas asociado al petróleo: Gas que se presenta en el yacimiento junto al petróleo. Puede estar en el yacimiento como una capa libre, también mezclado con el petróleo y presentarse como condensado formando una sola faz líquida con él en determinadas condiciones de temperatura y presión.

Gas natural: Gas que se presenta natural en el subsuelo y está constituido principalmente por metano.

Gasoducto: Tubería para el transporte de gas natural a alta presión y grandes distancias. Los gasoductos pueden ser nacionales e internacionales, y suministran a una sola o varias regiones.

Gas seco: Gas natural cuyo contenido de agua ha quedado reducido por un proceso de deshidratación.

Inyección (lodo de perforación): Mezcla de arcilla, agua y ciertos productos químicos inyectado en forma continua durante las operaciones de perforación. El lodo sirve para evacuar los recortes, lubricar y enfriar el trépano, sostener las paredes de los pozos y equilibrar la presión de los fluidos contenidos en las formaciones.

Oleoducto: Tubería generalmente subterránea para transportar petróleo a cortas y largas distancias. En estas últimas se utilizan estaciones de bombeo.

Perforación: Operación que consiste en perforar el subsuelo con la ayuda de herramientas apropiadas para buscar y extraer hidrocarburos.

Permeabilidad: Es la conductividad de un cuerpo poroso a los fluidos o capacidad de los fluidos para desplazarse entre los espacios que conectan los poros de una masa porosa.

Porosidad: Porcentaje del volumen total de una roca constituido por espacios vacíos. La porosidad efectiva es el volumen total de los espacios porosos interconectados de manera que permitan el paso de fluidos a través de ellos.

Pozo: Denominación dada a la abertura producida por una perforación.

Pozo de inyección: Pozo a través del cual se inyecta agua para mantener la presión de un yacimiento en la operación de recuperación secundaria.

Refinación: Conjunto de procesos industriales empleados para transformar los petróleos crudos en productos derivados; nafta, gas-oil, querosene, solventes, lubricantes, asfalto, etc.

Regalías: Proporción del valor del petróleo crudo o gas medido en boca de pozo que se reconoce al Estado propietario, provincial o nacional, según corresponda, de los yacimientos.

Relación Gas/Petróleo: Volumen de gas producido simultáneamente por un pozo con relación a cada metro cúbico de petróleo.

Reservas de hidrocarburos: Son los volúmenes de hidrocarburos líquidos o gaseosos existentes en el subsuelo y que han sido objetos de evaluación.

Reservas posibles: Son las situadas en áreas en las que no se ha encontrado hidrocarburos luego de llevar a cabo pozos de exploración. No obstante, las características geológicas permiten presumir la presencia de los hidrocarburos.

Reservas probables: Reservas cuya presencia en una zona determinada están claramente demostradas pero que las condiciones técnicas y económicas actuales impiden extraerlas, ya sea por el alto costo de extracción o por la poca fluidez de los petróleos.

Reservas probadas: Volúmenes de hidrocarburos recuperables de un yacimiento, de una cuenca, de una provincia o de un país con tecnología disponible y las condiciones económicas actuales.

Saturación: Volumen de petróleo o de gas en una roca con respecto al volumen de agua.

Tuberías de revestimiento (casing): Serie de tubos que se colocan en el pozo mientras progresa la perforación para prevenir derrumbes de las paredes y para la extracción de los hidrocarburos en la fase de la producción.

Yacimiento de petróleo o gas: Formación geológica continua de roca porosa y permeable por la que pueden circular los hidrocarburos, agua y otros gases.

ANEXO C. RESPUESTAS A ALGUNOS EJERCICIOS

CAPITULO 1

Ejercicio 2. Porosidad: $x_w = 22.7 \%$, $S = 2.95 \%$, $P_{10} = 20 \%$, $P_{90} = 18 \%$.

Permeabilidad: $x_w = 94.5 \text{ md}$, $S = 74.72 \text{ md}$, $P_{10} = 10 \text{ md}$, $P_{90} = 70 \text{ md}$.

Ejercicio 3. Empleados producción: $x = 7.166$ millones de pesos, $S = 2.27$ millones de pesos.

Empleados oficina: $x = 7.82$ millones de pesos, $S = 1.45$ millones de pesos.

El sindicato tiene la razón, ya que los salarios de los empleados de producción son inferiores a los de oficina, y su variabilidad es mayor. La variabilidad de estos salarios se midió con la desviación estándar.

CAPITULO 2

Ejercicio 3. La probabilidad de que necesite un nuevo motor o una nueva correa es 0.12.

Ejercicio 4. La probabilidad de que la charla hubiera salido como lo esperaba es de 0.85.

CAPITULO 3

Ejercicio 1. a) $p = 1.33$, b) $p = 0.0638$, c) $p = 3.54$

Ejercicio 3. $p = 0.156$, el pozo debe ser sometido a un proceso de inyección.

CAPITULO 4.

Ejercicio 3. a) $n = 87$ y b) $E = 0.047$

Ejercicio 4. $n = 1068$ pozos

Ejercicio 8. $-5.77 \leq (\mu_1 - \mu_2) \leq 0.177$

Ejercicio 9. $-17.65 \leq (\mu_1 - \mu_2) \leq 8.05$, el intervalo contiene cero, sugiriendo que no hay diferencia.

Ejercicio 10. – $0.1821 \leq (\mu_1 - \mu_2) \leq 0.0821$, debido a que el intervalo contiene cero, no hay evidencia de que exista alguna diferencia.

Ejercicio 19. Se rechaza H_0 si $F > 2.21$. Como $F = 1.78$. No se rechaza la hipótesis nula, la variabilidad de los diámetros de los empaques es la misma en las dos máquinas.

Ejercicio 20. La hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_A : no todas las poblaciones tienen la misma media

La hipótesis nula se rechaza ya que el estadístico de prueba (16.36) es mayor que el valor crítico (4.94). Por lo tanto Marlene debe concluir que los volúmenes de petróleo separado difieren entre estas cuatro marcas de separadores bifásicos.

ANEXO D. Distribución Binomial

Tabla A Distribución Binomial

<i>n</i>	<i>X</i>	π									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
	4	0.0000	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1563
	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
	2	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156
	7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.2573	0.3720	0.3960	0.3670	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547
	2	0.0406	0.1240	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.2140	0.1641
	3	0.0036	0.0230	0.0617	0.1147	0.1730	0.2269	0.2679	0.2903	0.2918	0.2734
	4	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734
	5	0.0000	0.0002	0.0012	0.0043	0.0115	0.0250	0.0466	0.0774	0.1172	0.1641
	6	0.0000	0.0000	0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.0320	0.0547
	7	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0016	0.0037	0.0078
	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.2793	0.3826	0.3847	0.3355	0.2670	0.1977	0.1373	0.0896	0.0548	0.0313
	2	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.2090	0.1569	0.1094
	3	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2188
	4	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734
	5	0.0000	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2188
	6	0.0000	0.0000	0.0002	0.0011	0.0038	0.0100	0.0217	0.0413	0.0703	0.1094
	7	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0313
	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0017	0.0039
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176

Tabla A Distribución Binomial (Continuación)

<i>n</i>	<i>X</i>	π									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	2	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703
	3	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
	4	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461
	5	0.0000	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
	6	0.0000	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641
	7	0.0000	0.0000	0.0000	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
	8	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0020
	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
2	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439	
3	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172	
4	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051	
5	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461	
6	0.0000	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051	
7	0.0000	0.0000	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172	
8	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439	
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098	
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.3293	0.3835	0.3248	0.2362	0.1549	0.0932	0.0518	0.0266	0.0125	0.0054
	2	0.0867	0.2131	0.2866	0.2953	0.2581	0.1998	0.1395	0.0887	0.0513	0.0269
	3	0.0137	0.0710	0.1517	0.2215	0.2581	0.2568	0.2254	0.1774	0.1259	0.0806
	4	0.0014	0.0158	0.0536	0.1107	0.1721	0.2201	0.2428	0.2365	0.2060	0.1611
	5	0.0001	0.0025	0.0132	0.0388	0.0803	0.1321	0.1830	0.2207	0.2360	0.2256
	6	0.0000	0.0003	0.0023	0.0097	0.0268	0.0566	0.0985	0.1471	0.1931	0.2256
	7	0.0000	0.0000	0.0003	0.0017	0.0064	0.0173	0.0379	0.0701	0.1128	0.1611
	8	0.0000	0.0000	0.0000	0.0002	0.0011	0.0037	0.0102	0.0234	0.0462	0.0806
	9	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018	0.0052	0.0126	0.0269
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0021	0.0054
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.3413	0.3766	0.3012	0.2062	0.1267	0.0712	0.0368	0.0174	0.0075	0.0029
	2	0.0988	0.2301	0.2924	0.2835	0.2323	0.1678	0.1088	0.0639	0.0339	0.0161
	3	0.0173	0.0852	0.1720	0.2362	0.2581	0.2397	0.1954	0.1419	0.0923	0.0537
	4	0.0021	0.0213	0.0683	0.1329	0.1936	0.2311	0.2367	0.2128	0.1700	0.1208
	5	0.0002	0.0038	0.0193	0.0532	0.1032	0.1585	0.2039	0.2270	0.2225	0.1934
	6	0.0000	0.0005	0.0040	0.0155	0.0401	0.0792	0.1281	0.1766	0.2124	0.2256
	7	0.0000	0.0000	0.0006	0.0033	0.0115	0.0291	0.0591	0.1009	0.1489	0.1934
	8	0.0000	0.0000	0.0001	0.0005	0.0024	0.0078	0.0199	0.0420	0.0762	0.1208
	9	0.0000	0.0000	0.0000	0.0001	0.0004	0.0015	0.0048	0.0125	0.0277	0.0537
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0025	0.0068	0.0161
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0029
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.3512	0.3672	0.2774	0.1787	0.1029	0.0540	0.0259	0.0113	0.0045	0.0016

Tabla A Distribución Binomial (Continuación)

<i>n</i>	<i>X</i>	π									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
14	2	0.1109	0.2448	0.2937	0.2680	0.2059	0.1388	0.0836	0.0453	0.0220	0.0095
	3	0.0214	0.0997	0.1900	0.2457	0.2517	0.2181	0.1651	0.1107	0.0660	0.0349
	4	0.0028	0.0277	0.0838	0.1535	0.2097	0.2337	0.2222	0.1845	0.1350	0.0873
	5	0.0003	0.0055	0.0266	0.0691	0.1258	0.1803	0.2154	0.2214	0.1989	0.1571
	6	0.0000	0.0008	0.0063	0.0230	0.0559	0.1030	0.1546	0.1968	0.2169	0.2095
	7	0.0000	0.0001	0.0011	0.0058	0.0186	0.0442	0.0833	0.1312	0.1775	0.2095
	8	0.0000	0.0000	0.0001	0.0011	0.0047	0.0142	0.0336	0.0656	0.1089	0.1571
	9	0.0000	0.0000	0.0000	0.0001	0.0009	0.0034	0.0101	0.0243	0.0495	0.0873
	10	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0022	0.0065	0.0162	0.0349
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0012	0.0036	0.0095
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016
	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
	14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002
1		0.3593	0.3559	0.2539	0.1539	0.0832	0.0407	0.0181	0.0073	0.0027	0.0009
2		0.1229	0.2570	0.2912	0.2501	0.1802	0.1134	0.0634	0.0317	0.0141	0.0056
3		0.0259	0.1142	0.2056	0.2501	0.2402	0.1943	0.1366	0.0845	0.0462	0.0222
4		0.0037	0.0349	0.0998	0.1720	0.2202	0.2290	0.2022	0.1549	0.1040	0.0611
5		0.0004	0.0078	0.0352	0.0860	0.1468	0.1963	0.2178	0.2066	0.1701	0.1222
6		0.0000	0.0013	0.0093	0.0322	0.0734	0.1262	0.1759	0.2066	0.2088	0.1833
7		0.0000	0.0002	0.0019	0.0092	0.0280	0.0618	0.1082	0.1574	0.1952	0.2095
8		0.0000	0.0000	0.0003	0.0020	0.0082	0.0232	0.0510	0.0918	0.1398	0.1833
9		0.0000	0.0000	0.0000	0.0003	0.0018	0.0066	0.0183	0.0408	0.0762	0.1222
10		0.0000	0.0000	0.0000	0.0000	0.0003	0.0014	0.0049	0.0136	0.0312	0.0611
11		0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0010	0.0033	0.0093	0.0222
12		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0019	0.0056
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0009	
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.3658	0.3432	0.2312	0.1319	0.0668	0.0305	0.0126	0.0047	0.0016	0.0005
	2	0.1348	0.2669	0.2856	0.2309	0.1559	0.0916	0.0476	0.0219	0.0090	0.0032
	3	0.0307	0.1285	0.2184	0.2501	0.2252	0.1700	0.1110	0.0634	0.0318	0.0139
	4	0.0049	0.0428	0.1156	0.1876	0.2252	0.2186	0.1792	0.1268	0.0780	0.0417
	5	0.0006	0.0105	0.0449	0.1032	0.1651	0.2061	0.2123	0.1859	0.1404	0.0916
	6	0.0000	0.0019	0.0132	0.0430	0.0917	0.1472	0.1906	0.2066	0.1914	0.1527
	7	0.0000	0.0003	0.0030	0.0138	0.0393	0.0811	0.1319	0.1771	0.2013	0.1964
	8	0.0000	0.0000	0.0005	0.0035	0.0131	0.0348	0.0710	0.1181	0.1647	0.1964
	9	0.0000	0.0000	0.0001	0.0007	0.0034	0.0116	0.0298	0.0612	0.1048	0.1527
	10	0.0000	0.0000	0.0000	0.0001	0.0007	0.0030	0.0096	0.0245	0.0515	0.0916
	11	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0074	0.0191	0.0417
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0016	0.0052	0.0139
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0032	
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.3706	0.3294	0.2097	0.1126	0.0535	0.0228	0.0087	0.0030	0.0009	0.0002
	2	0.1463	0.2745	0.2775	0.2111	0.1336	0.0732	0.0353	0.0150	0.0056	0.0018

Tabla A Distribución Binomial (Continuación)

<i>n</i>	<i>X</i>	π										
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	
10	3	0.0359	0.1423	0.2285	0.2463	0.2079	0.1465	0.0888	0.0468	0.0215	0.0085	
	4	0.0061	0.0514	0.1311	0.2001	0.2252	0.2040	0.1553	0.1014	0.0572	0.0278	
	5	0.0008	0.0137	0.0555	0.1201	0.1802	0.2099	0.2008	0.1623	0.1123	0.0667	
	6	0.0001	0.0028	0.0180	0.0550	0.1101	0.1649	0.1982	0.1983	0.1684	0.1222	
	7	0.0000	0.0004	0.0045	0.0197	0.0524	0.1010	0.1524	0.1889	0.1969	0.1746	
	8	0.0000	0.0001	0.0009	0.0055	0.0197	0.0487	0.0923	0.1417	0.1812	0.1964	
	9	0.0000	0.0000	0.0001	0.0012	0.0058	0.0185	0.0442	0.0840	0.1318	0.1746	
	10	0.0000	0.0000	0.0000	0.0002	0.0014	0.0056	0.0167	0.0392	0.0755	0.1222	
	11	0.0000	0.0000	0.0000	0.0000	0.0002	0.0013	0.0049	0.0142	0.0337	0.0667	
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011	0.0040	0.0115	0.0278	
	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0029	0.0085	
	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018	
	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	
	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
		1	0.3741	0.3150	0.1893	0.0957	0.0426	0.0169	0.0060	0.0019	0.0005	0.0001
2		0.1575	0.2800	0.2673	0.1914	0.1136	0.0581	0.0260	0.0102	0.0035	0.0010	
3		0.0415	0.1556	0.2359	0.2393	0.1893	0.1245	0.0701	0.0341	0.0144	0.0052	
4		0.0076	0.0605	0.1457	0.2093	0.2209	0.1868	0.1320	0.0796	0.0411	0.0182	
5		0.0010	0.0175	0.0668	0.1361	0.1914	0.2081	0.1849	0.1379	0.0875	0.0472	
6		0.0001	0.0039	0.0236	0.0680	0.1276	0.1784	0.1991	0.1839	0.1432	0.0944	
7		0.0000	0.0007	0.0065	0.0267	0.0668	0.1201	0.1685	0.1927	0.1841	0.1484	
8		0.0000	0.0001	0.0014	0.0084	0.0279	0.0644	0.1134	0.1606	0.1883	0.1855	
9		0.0000	0.0000	0.0003	0.0021	0.0093	0.0276	0.0611	0.1070	0.1540	0.1855	
10		0.0000	0.0000	0.0000	0.0004	0.0025	0.0095	0.0263	0.0571	0.1008	0.1484	
11		0.0000	0.0000	0.0000	0.0001	0.0005	0.0026	0.0090	0.0242	0.0525	0.0944	
12		0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0081	0.0215	0.0472	
13		0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0021	0.0068	0.0182	
14		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0016	0.0052	
15		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	
16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001		
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000	
	1	0.3763	0.3002	0.1704	0.0811	0.0338	0.0126	0.0042	0.0012	0.0003	0.0001	
	2	0.1683	0.2835	0.2556	0.1723	0.0958	0.0458	0.0190	0.0069	0.0022	0.0006	
	3	0.0473	0.1680	0.2406	0.2297	0.1704	0.1046	0.0547	0.0246	0.0095	0.0031	
	4	0.0093	0.0700	0.1592	0.2153	0.2130	0.1681	0.1104	0.0614	0.0291	0.0117	
	5	0.0014	0.0218	0.0787	0.1507	0.1988	0.2017	0.1664	0.1146	0.0666	0.0327	
	6	0.0002	0.0052	0.0301	0.0816	0.1436	0.1873	0.1941	0.1655	0.1181	0.0708	
	7	0.0000	0.0010	0.0091	0.0350	0.0820	0.1376	0.1792	0.1892	0.1657	0.1214	
	8	0.0000	0.0002	0.0022	0.0120	0.0376	0.0811	0.1327	0.1734	0.1864	0.1669	
	9	0.0000	0.0000	0.0004	0.0033	0.0139	0.0386	0.0794	0.1284	0.1694	0.1855	
	10	0.0000	0.0000	0.0001	0.0008	0.0042	0.0149	0.0385	0.0771	0.1248	0.1669	
	11	0.0000	0.0000	0.0000	0.0001	0.0010	0.0046	0.0151	0.0374	0.0742	0.1214	
	12	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0047	0.0145	0.0354	0.0708	
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0045	0.0134	0.0327		

Tabla A Distribución Binomial (Continuación)

<i>n</i>	<i>X</i>	π									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
19	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011	0.0039	0.0117
	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0031
	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006
	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.3774	0.2852	0.1529	0.0685	0.0268	0.0093	0.0029	0.0008	0.0002	0.0000
	2	0.1787	0.2852	0.2428	0.1540	0.0803	0.0358	0.0138	0.0046	0.0013	0.0003
	3	0.0533	0.1796	0.2428	0.2182	0.1517	0.0869	0.0422	0.0175	0.0062	0.0018
	4	0.0112	0.0798	0.1714	0.2182	0.2023	0.1491	0.0909	0.0467	0.0203	0.0074
	5	0.0018	0.0266	0.0907	0.1636	0.2023	0.1916	0.1468	0.0933	0.0497	0.0222
	6	0.0002	0.0069	0.0374	0.0955	0.1574	0.1916	0.1844	0.1451	0.0949	0.0518
	7	0.0000	0.0014	0.0122	0.0443	0.0974	0.1525	0.1844	0.1797	0.1443	0.0961
	8	0.0000	0.0002	0.0032	0.0166	0.0487	0.0981	0.1489	0.1797	0.1771	0.1442
	9	0.0000	0.0000	0.0007	0.0051	0.0198	0.0514	0.0980	0.1464	0.1771	0.1762
20	10	0.0000	0.0000	0.0001	0.0013	0.0066	0.0220	0.0528	0.0976	0.1449	0.1762
	11	0.0000	0.0000	0.0000	0.0003	0.0018	0.0077	0.0233	0.0532	0.0970	0.1442
	12	0.0000	0.0000	0.0000	0.0000	0.0004	0.0022	0.0083	0.0237	0.0529	0.0961
	13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0024	0.0085	0.0233	0.0518
	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0082	0.0222
	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0022	0.0074
	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018
	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003
	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.3774	0.2702	0.1368	0.0576	0.0211	0.0068	0.0020	0.0005	0.0001	0.0000
	2	0.1887	0.2852	0.2293	0.1369	0.0669	0.0278	0.0100	0.0031	0.0008	0.0002
	3	0.0596	0.1901	0.2428	0.2054	0.1339	0.0716	0.0323	0.0123	0.0040	0.0011
	4	0.0133	0.0898	0.1821	0.2182	0.1897	0.1304	0.0738	0.0350	0.0139	0.0046
5	0.0022	0.0319	0.1028	0.1746	0.2023	0.1789	0.1272	0.0746	0.0365	0.0148	
6	0.0003	0.0089	0.0454	0.1091	0.1686	0.1916	0.1712	0.1244	0.0746	0.0370	
7	0.0000	0.0020	0.0160	0.0545	0.1124	0.1643	0.1844	0.1659	0.1221	0.0739	
8	0.0000	0.0004	0.0046	0.0222	0.0609	0.1144	0.1614	0.1797	0.1623	0.1201	
9	0.0000	0.0001	0.0011	0.0074	0.0271	0.0654	0.1158	0.1597	0.1771	0.1602	
10	0.0000	0.0000	0.0002	0.0020	0.0099	0.0308	0.0686	0.1171	0.1593	0.1762	
11	0.0000	0.0000	0.0000	0.0005	0.0030	0.0120	0.0336	0.0710	0.1185	0.1602	
12	0.0000	0.0000	0.0000	0.0001	0.0008	0.0039	0.0136	0.0355	0.0727	0.1201	
13	0.0000	0.0000	0.0000	0.0000	0.0002	0.0010	0.0045	0.0146	0.0366	0.0739	
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0049	0.0150	0.0370	
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0013	0.0049	0.0148	
16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0013	0.0046	
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011	
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Tabla B Distribución De Poisson

x	μ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	0.3679
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	0.1839
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	0.0613
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	0.0153
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	0.0031
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

x	μ									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.3662	0.3614	0.3543	0.3452	0.3347	0.3230	0.3106	0.2975	0.2842	0.2707
2	0.2014	0.2169	0.2303	0.2417	0.2510	0.2584	0.2640	0.2678	0.2700	0.2707
3	0.0738	0.0867	0.0998	0.1128	0.1255	0.1378	0.1496	0.1607	0.1710	0.1804
4	0.0203	0.0260	0.0324	0.0395	0.0471	0.0551	0.0636	0.0723	0.0812	0.0902
5	0.0045	0.0062	0.0084	0.0111	0.0141	0.0176	0.0216	0.0260	0.0309	0.0361
6	0.0008	0.0012	0.0018	0.0026	0.0035	0.0047	0.0061	0.0078	0.0098	0.0120
7	0.0001	0.0002	0.0003	0.0005	0.0008	0.0011	0.0015	0.0020	0.0027	0.0034
8	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0005	0.0006	0.0009

x	μ									
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.2572	0.2438	0.2306	0.2177	0.2052	0.1931	0.1815	0.1703	0.1596	0.1494
2	0.2700	0.2681	0.2652	0.2613	0.2565	0.2510	0.2450	0.2384	0.2314	0.2240
3	0.1890	0.1966	0.2033	0.2090	0.2138	0.2176	0.2205	0.2225	0.2237	0.2240
4	0.0992	0.1082	0.1169	0.1254	0.1336	0.1414	0.1488	0.1557	0.1622	0.1680
5	0.0417	0.0476	0.0538	0.0602	0.0668	0.0735	0.0804	0.0872	0.0940	0.1008
6	0.0146	0.0174	0.0206	0.0241	0.0278	0.0319	0.0362	0.0407	0.0455	0.0504
7	0.0044	0.0055	0.0068	0.0083	0.0099	0.0118	0.0139	0.0163	0.0188	0.0216
8	0.0011	0.0015	0.0019	0.0025	0.0031	0.0038	0.0047	0.0057	0.0068	0.0081
9	0.0003	0.0004	0.0005	0.0007	0.0009	0.0011	0.0014	0.0018	0.0022	0.0027
10	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0004	0.0005	0.0006	0.0008
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002
12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

x	μ									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1397	0.1304	0.1217	0.1135	0.1057	0.0984	0.0915	0.0850	0.0789	0.0733
2	0.2165	0.2087	0.2008	0.1929	0.1850	0.1771	0.1692	0.1615	0.1539	0.1465
3	0.2237	0.2226	0.2209	0.2186	0.2158	0.2125	0.2087	0.2046	0.2001	0.1954
4	0.1733	0.1781	0.1823	0.1858	0.1888	0.1912	0.1931	0.1944	0.1951	0.1954
5	0.1075	0.1140	0.1203	0.1264	0.1322	0.1377	0.1429	0.1477	0.1522	0.1563

Tabla B Distribución De Poisson (Continuación)

x	μ									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
6	0.0555	0.0608	0.0662	0.0716	0.0771	0.0826	0.0881	0.0936	0.0989	0.1042
7	0.0246	0.0278	0.0312	0.0348	0.0385	0.0425	0.0466	0.0508	0.0551	0.0595
8	0.0095	0.0111	0.0129	0.0148	0.0169	0.0191	0.0215	0.0241	0.0269	0.0298
9	0.0033	0.0040	0.0047	0.0056	0.0066	0.0076	0.0089	0.0102	0.0116	0.0132
10	0.0010	0.0013	0.0016	0.0019	0.0023	0.0028	0.0033	0.0039	0.0045	0.0053
11	0.0003	0.0004	0.0005	0.0006	0.0007	0.0009	0.0011	0.0013	0.0016	0.0019
12	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006
13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

x	μ									
	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0.0166	0.0150	0.0136	0.0123	0.0111	0.0101	0.0091	0.0082	0.0074	0.0067
1	0.0679	0.0630	0.0583	0.0540	0.0500	0.0462	0.0427	0.0395	0.0365	0.0337
2	0.1393	0.1323	0.1254	0.1188	0.1125	0.1063	0.1005	0.0948	0.0894	0.0842
3	0.1904	0.1852	0.1798	0.1743	0.1687	0.1631	0.1574	0.1517	0.1460	0.1404
4	0.1951	0.1944	0.1933	0.1917	0.1898	0.1875	0.1849	0.1820	0.1789	0.1755
5	0.1600	0.1633	0.1662	0.1687	0.1708	0.1725	0.1738	0.1747	0.1753	0.1755
6	0.1093	0.1143	0.1191	0.1237	0.1281	0.1323	0.1362	0.1398	0.1432	0.1462
7	0.0640	0.0686	0.0732	0.0778	0.0824	0.0869	0.0914	0.0959	0.1002	0.1044
8	0.0328	0.0360	0.0393	0.0428	0.0463	0.0500	0.0537	0.0575	0.0614	0.0653
9	0.0150	0.0168	0.0188	0.0209	0.0232	0.0255	0.0281	0.0307	0.0334	0.0363
10	0.0061	0.0071	0.0081	0.0092	0.0104	0.0118	0.0132	0.0147	0.0164	0.0181
11	0.0023	0.0027	0.0032	0.0037	0.0043	0.0049	0.0056	0.0064	0.0073	0.0082
12	0.0008	0.0009	0.0011	0.0013	0.0016	0.0019	0.0022	0.0026	0.0030	0.0034
13	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009	0.0011	0.0013
14	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005
15	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002

x	μ									
	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	0.0061	0.0055	0.0050	0.0045	0.0041	0.0037	0.0033	0.0030	0.0027	0.0025
1	0.0311	0.0287	0.0265	0.0244	0.0225	0.0207	0.0191	0.0176	0.0162	0.0149
2	0.0793	0.0746	0.0701	0.0659	0.0618	0.0580	0.0544	0.0509	0.0477	0.0446
3	0.1348	0.1293	0.1239	0.1185	0.1133	0.1082	0.1033	0.0985	0.0938	0.0892
4	0.1719	0.1681	0.1641	0.1600	0.1558	0.1515	0.1472	0.1428	0.1383	0.1339
5	0.1753	0.1748	0.1740	0.1728	0.1714	0.1697	0.1678	0.1656	0.1632	0.1606
6	0.1490	0.1515	0.1537	0.1555	0.1571	0.1584	0.1594	0.1601	0.1605	0.1606
7	0.1086	0.1125	0.1163	0.1200	0.1234	0.1267	0.1298	0.1326	0.1353	0.1377
8	0.0692	0.0731	0.0771	0.0810	0.0849	0.0887	0.0925	0.0962	0.0998	0.1033
9	0.0392	0.0423	0.0454	0.0486	0.0519	0.0552	0.0586	0.0620	0.0654	0.0688
10	0.0200	0.0220	0.0241	0.0262	0.0285	0.0309	0.0334	0.0359	0.0386	0.0413
11	0.0093	0.0104	0.0116	0.0129	0.0143	0.0157	0.0173	0.0190	0.0207	0.0225
12	0.0039	0.0045	0.0051	0.0058	0.0065	0.0073	0.0082	0.0092	0.0102	0.0113
13	0.0015	0.0018	0.0021	0.0024	0.0028	0.0032	0.0036	0.0041	0.0046	0.0052

Tabla B Distribución De Poisson (Continuación)

<i>x</i>	μ									
	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
14	0.0006	0.0007	0.0008	0.0009	0.0011	0.0013	0.0015	0.0017	0.0019	0.0022
15	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
16	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001

<i>x</i>	μ									
	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0137	0.0126	0.0116	0.0106	0.0098	0.0090	0.0082	0.0076	0.0070	0.0064
2	0.0417	0.0390	0.0364	0.0340	0.0318	0.0296	0.0276	0.0258	0.0240	0.0223
3	0.0848	0.0806	0.0765	0.0726	0.0688	0.0652	0.0617	0.0584	0.0552	0.0521
4	0.1294	0.1249	0.1205	0.1162	0.1118	0.1076	0.1034	0.0992	0.0952	0.0912
5	0.1579	0.1549	0.1519	0.1487	0.1454	0.1420	0.1385	0.1349	0.1314	0.1277
6	0.1605	0.1601	0.1595	0.1586	0.1575	0.1562	0.1546	0.1529	0.1511	0.1490
7	0.1399	0.1418	0.1435	0.1450	0.1462	0.1472	0.1480	0.1486	0.1489	0.1490
8	0.1066	0.1099	0.1130	0.1160	0.1188	0.1215	0.1240	0.1263	0.1284	0.1304
9	0.0723	0.0757	0.0791	0.0825	0.0858	0.0891	0.0923	0.0954	0.0985	0.1014
10	0.0441	0.0469	0.0498	0.0528	0.0558	0.0588	0.0618	0.0649	0.0679	0.0710
11	0.0244	0.0265	0.0285	0.0307	0.0330	0.0353	0.0377	0.0401	0.0426	0.0452
12	0.0124	0.0137	0.0150	0.0164	0.0179	0.0194	0.0210	0.0227	0.0245	0.0263
13	0.0058	0.0065	0.0073	0.0081	0.0089	0.0099	0.0108	0.0119	0.0130	0.0142
14	0.0025	0.0029	0.0033	0.0037	0.0041	0.0046	0.0052	0.0058	0.0064	0.0071
15	0.0010	0.0012	0.0014	0.0016	0.0018	0.0020	0.0023	0.0026	0.0029	0.0033
16	0.0004	0.0005	0.0005	0.0006	0.0007	0.0008	0.0010	0.0011	0.0013	0.0014
17	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006
18	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001

<i>x</i>	μ									
	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003
1	0.0059	0.0054	0.0049	0.0045	0.0041	0.0038	0.0035	0.0032	0.0029	0.0027
2	0.0208	0.0194	0.0180	0.0167	0.0156	0.0145	0.0134	0.0125	0.0116	0.0107
3	0.0492	0.0464	0.0438	0.0413	0.0389	0.0366	0.0345	0.0324	0.0305	0.0286
4	0.0874	0.0836	0.0799	0.0764	0.0729	0.0696	0.0663	0.0632	0.0602	0.0573
5	0.1241	0.1204	0.1167	0.1130	0.1094	0.1057	0.1021	0.0986	0.0951	0.0916
6	0.1468	0.1445	0.1420	0.1394	0.1367	0.1339	0.1311	0.1282	0.1252	0.1221
7	0.1489	0.1486	0.1481	0.1474	0.1465	0.1454	0.1442	0.1428	0.1413	0.1396
8	0.1321	0.1337	0.1351	0.1363	0.1373	0.1381	0.1388	0.1392	0.1395	0.1396
9	0.1042	0.1070	0.1096	0.1121	0.1144	0.1167	0.1187	0.1207	0.1224	0.1241
10	0.0740	0.0770	0.0800	0.0829	0.0858	0.0887	0.0914	0.0941	0.0967	0.0993
11	0.0478	0.0504	0.0531	0.0558	0.0585	0.0613	0.0640	0.0667	0.0695	0.0722
12	0.0283	0.0303	0.0323	0.0344	0.0366	0.0388	0.0411	0.0434	0.0457	0.0481
13	0.0154	0.0168	0.0181	0.0196	0.0211	0.0227	0.0243	0.0260	0.0278	0.0296
14	0.0078	0.0086	0.0095	0.0104	0.0113	0.0123	0.0134	0.0145	0.0157	0.0169

Tabla B Distribución De Poisson (Continuación)

x	μ									
	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
15	0.0037	0.0041	0.0046	0.0051	0.0057	0.0062	0.0069	0.0075	0.0083	0.0090
16	0.0016	0.0019	0.0021	0.0024	0.0026	0.0030	0.0033	0.0037	0.0041	0.0045
17	0.0007	0.0008	0.0009	0.0010	0.0012	0.0013	0.0015	0.0017	0.0019	0.0021
18	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006	0.0006	0.0007	0.0008	0.0009
19	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003	0.0003	0.0004
20	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001

x	μ									
	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1	0.0025	0.0023	0.0021	0.0019	0.0017	0.0016	0.0014	0.0013	0.0012	0.0011
2	0.0100	0.0092	0.0086	0.0079	0.0074	0.0068	0.0063	0.0058	0.0054	0.0050
3	0.0269	0.0252	0.0237	0.0222	0.0208	0.0195	0.0183	0.0171	0.0160	0.0150
4	0.0544	0.0517	0.0491	0.0466	0.0443	0.0420	0.0398	0.0377	0.0357	0.0337
5	0.0882	0.0849	0.0816	0.0784	0.0752	0.0722	0.0692	0.0663	0.0635	0.0607
6	0.1191	0.1160	0.1128	0.1097	0.1066	0.1034	0.1003	0.0972	0.0941	0.0911
7	0.1378	0.1358	0.1338	0.1317	0.1294	0.1271	0.1247	0.1222	0.1197	0.1171
8	0.1395	0.1392	0.1388	0.1382	0.1375	0.1366	0.1356	0.1344	0.1332	0.1318
9	0.1256	0.1269	0.1280	0.1290	0.1299	0.1306	0.1311	0.1315	0.1317	0.1318
10	0.1017	0.1040	0.1063	0.1084	0.1104	0.1123	0.1140	0.1157	0.1172	0.1186
11	0.0749	0.0776	0.0802	0.0828	0.0853	0.0878	0.0902	0.0925	0.0948	0.0970
12	0.0505	0.0530	0.0555	0.0579	0.0604	0.0629	0.0654	0.0679	0.0703	0.0728
13	0.0315	0.0334	0.0354	0.0374	0.0395	0.0416	0.0438	0.0459	0.0481	0.0504
14	0.0182	0.0196	0.0210	0.0225	0.0240	0.0256	0.0272	0.0289	0.0306	0.0324
15	0.0098	0.0107	0.0116	0.0126	0.0136	0.0147	0.0158	0.0169	0.0182	0.0194
16	0.0050	0.0055	0.0060	0.0066	0.0072	0.0079	0.0086	0.0093	0.0101	0.0109
17	0.0024	0.0026	0.0029	0.0033	0.0036	0.0040	0.0044	0.0048	0.0053	0.0058
18	0.0011	0.0012	0.0014	0.0015	0.0017	0.0019	0.0021	0.0024	0.0026	0.0029
19	0.0005	0.0005	0.0006	0.0007	0.0008	0.0009	0.0010	0.0011	0.0012	0.0014
20	0.0002	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0005	0.0005	0.0006
21	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0003
22	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

x	μ									
	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
1	0.0010	0.0009	0.0009	0.0008	0.0007	0.0007	0.0006	0.0005	0.0005	0.0005
2	0.0046	0.0043	0.0040	0.0037	0.0034	0.0031	0.0029	0.0027	0.0025	0.0023
3	0.0140	0.0131	0.0123	0.0115	0.0107	0.0100	0.0093	0.0087	0.0081	0.0076
4	0.0319	0.0302	0.0285	0.0269	0.0254	0.0240	0.0226	0.0213	0.0201	0.0189
5	0.0581	0.0555	0.0530	0.0506	0.0483	0.0460	0.0439	0.0418	0.0398	0.0378
6	0.0881	0.0851	0.0822	0.0793	0.0764	0.0736	0.0709	0.0682	0.0656	0.0631
7	0.1145	0.1118	0.1091	0.1064	0.1037	0.1010	0.0982	0.0955	0.0928	0.0901

Tabla B Distribución De Poisson (Continuación)

x	μ									
	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
8	0.1302	0.1286	0.1269	0.1251	0.1232	0.1212	0.1191	0.1170	0.1148	0.1126
9	0.1317	0.1315	0.1311	0.1306	0.1300	0.1293	0.1284	0.1274	0.1263	0.1251
10	0.1198	0.1210	0.1219	0.1228	0.1235	0.1241	0.1245	0.1249	0.1250	0.1251
11	0.0991	0.1012	0.1031	0.1049	0.1067	0.1083	0.1098	0.1112	0.1125	0.1137
12	0.0752	0.0776	0.0799	0.0822	0.0844	0.0866	0.0888	0.0908	0.0928	0.0948
13	0.0526	0.0549	0.0572	0.0594	0.0617	0.0640	0.0662	0.0685	0.0707	0.0729
14	0.0342	0.0361	0.0380	0.0399	0.0419	0.0439	0.0459	0.0479	0.0500	0.0521
15	0.0208	0.0221	0.0235	0.0250	0.0265	0.0281	0.0297	0.0313	0.0330	0.0347
16	0.0118	0.0127	0.0137	0.0147	0.0157	0.0168	0.0180	0.0192	0.0204	0.0217
17	0.0063	0.0069	0.0075	0.0081	0.0088	0.0095	0.0103	0.0111	0.0119	0.0128
18	0.0032	0.0035	0.0039	0.0042	0.0046	0.0051	0.0055	0.0060	0.0065	0.0071
19	0.0015	0.0017	0.0019	0.0021	0.0023	0.0026	0.0028	0.0031	0.0034	0.0037
20	0.0007	0.0008	0.0009	0.0010	0.0011	0.0012	0.0014	0.0015	0.0017	0.0019
21	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006	0.0006	0.0007	0.0008	0.0009
22	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004
23	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001

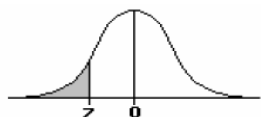
x	μ									
	11	12	13	14	15	16	17	18	19	20
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0010	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0037	0.0018	0.0008	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0102	0.0053	0.0027	0.0013	0.0006	0.0003	0.0001	0.0001	0.0000	0.0000
5	0.0224	0.0127	0.0070	0.0037	0.0019	0.0010	0.0005	0.0002	0.0001	0.0001
6	0.0411	0.0255	0.0152	0.0087	0.0048	0.0026	0.0014	0.0007	0.0004	0.0002
7	0.0646	0.0437	0.0281	0.0174	0.0104	0.0060	0.0034	0.0019	0.0010	0.0005
8	0.0888	0.0655	0.0457	0.0304	0.0194	0.0120	0.0072	0.0042	0.0024	0.0013
9	0.1085	0.0874	0.0661	0.0473	0.0324	0.0213	0.0135	0.0083	0.0050	0.0029
10	0.1194	0.1048	0.0859	0.0663	0.0486	0.0341	0.0230	0.0150	0.0095	0.0058
11	0.1194	0.1144	0.1015	0.0844	0.0663	0.0496	0.0355	0.0245	0.0164	0.0106
12	0.1094	0.1144	0.1099	0.0984	0.0829	0.0661	0.0504	0.0368	0.0259	0.0176
13	0.0926	0.1056	0.1099	0.1060	0.0956	0.0814	0.0658	0.0509	0.0378	0.0271
14	0.0728	0.0905	0.1021	0.1060	0.1024	0.0930	0.0800	0.0655	0.0514	0.0387
15	0.0534	0.0724	0.0885	0.0989	0.1024	0.0992	0.0906	0.0786	0.0650	0.0516
16	0.0367	0.0543	0.0719	0.0866	0.0960	0.0992	0.0963	0.0884	0.0772	0.0646
17	0.0237	0.0383	0.0550	0.0713	0.0847	0.0934	0.0963	0.0936	0.0863	0.0760
18	0.0145	0.0255	0.0397	0.0554	0.0706	0.0830	0.0909	0.0936	0.0911	0.0844
19	0.0084	0.0161	0.0272	0.0409	0.0557	0.0699	0.0814	0.0887	0.0911	0.0888
20	0.0046	0.0097	0.0177	0.0286	0.0418	0.0559	0.0692	0.0798	0.0866	0.0888
21	0.0024	0.0055	0.0109	0.0191	0.0299	0.0426	0.0560	0.0684	0.0783	0.0846
22	0.0012	0.0030	0.0065	0.0121	0.0204	0.0310	0.0433	0.0560	0.0676	0.0769
23	0.0006	0.0016	0.0037	0.0074	0.0133	0.0216	0.0320	0.0438	0.0559	0.0669

Tabla B Distribución De Poisson (Continuación)

x	μ									
	11	12	13	14	15	16	17	18	19	20
24	0.0003	0.0008	0.0020	0.0043	0.0083	0.0144	0.0226	0.0328	0.0442	0.0557
25	0.0001	0.0004	0.0010	0.0024	0.0050	0.0092	0.0154	0.0237	0.0336	0.0446
26	0.0000	0.0002	0.0005	0.0013	0.0029	0.0057	0.0101	0.0164	0.0246	0.0343
27	0.0000	0.0001	0.0002	0.0007	0.0016	0.0034	0.0063	0.0109	0.0173	0.0254
28	0.0000	0.0000	0.0001	0.0003	0.0009	0.0019	0.0038	0.0070	0.0117	0.0181
29	0.0000	0.0000	0.0001	0.0002	0.0004	0.0011	0.0023	0.0044	0.0077	0.0125
30	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0013	0.0026	0.0049	0.0083
31	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0007	0.0015	0.0030	0.0054
32	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0004	0.0009	0.0018	0.0034
33	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0010	0.0020
34	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0012
35	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0007
36	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004
37	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
38	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
39	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

ANEXO E. Distribución Normal

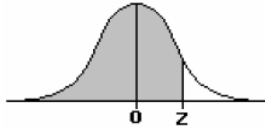
Tabla C Distribución Normal



La tabla proporciona, para cada valor de z , el área que queda a su izquierda.

z	0'00	0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09
-4'4	0'00001	0'00001	0'00001	0'00000	0'00000	0'00000	0'00000	0'00000	0'00000	0'00000
-4'3	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001
-4'2	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001	0'00001
-4'1	0'00002	0'00002	0'00002	0'00002	0'00002	0'00002	0'00002	0'00002	0'00002	0'00001
-4'0	0'00003	0'00003	0'00003	0'00003	0'00003	0'00003	0'00002	0'00002	0'00002	0'00002
-3'9	0'00005	0'00005	0'00004	0'00004	0'00004	0'00004	0'00004	0'00004	0'00003	0'00003
-3'8	0'00007	0'00007	0'00007	0'00006	0'00006	0'00006	0'00006	0'00005	0'00005	0'00005
-3'7	0'00011	0'00010	0'00010	0'00010	0'00009	0'00009	0'00009	0'00008	0'00008	0'00008
-3'6	0'00016	0'00015	0'00015	0'00014	0'00014	0'00013	0'00013	0'00012	0'00012	0'00011
-3'5	0'00023	0'00023	0'00022	0'00021	0'00020	0'00019	0'00019	0'00018	0'00017	0'00017
-3'4	0'00034	0'00033	0'00032	0'00030	0'00029	0'00028	0'00027	0'00026	0'00025	0'00024
-3'3	0'00049	0'00047	0'00045	0'00044	0'00042	0'00041	0'00039	0'00038	0'00036	0'00035
-3'2	0'00069	0'00067	0'00064	0'00062	0'00060	0'00058	0'00056	0'00054	0'00052	0'00050
-3'1	0'00097	0'00094	0'00091	0'00088	0'00085	0'00082	0'00079	0'00077	0'00074	0'00071
-3'0	0'00135	0'00131	0'00127	0'00123	0'00119	0'00115	0'00111	0'00107	0'00104	0'00101
-2'9	0'00187	0'00181	0'00175	0'00169	0'00164	0'00159	0'00154	0'00149	0'00144	0'00139
-2'8	0'00256	0'00248	0'00240	0'00233	0'00226	0'00219	0'00212	0'00205	0'00199	0'00193
-2'7	0'00347	0'00336	0'00326	0'00317	0'00307	0'00298	0'00289	0'00280	0'00272	0'00264
-2'6	0'00466	0'00453	0'00440	0'00427	0'00415	0'00402	0'00391	0'00379	0'00368	0'00357
-2'5	0'00621	0'00604	0'00587	0'00570	0'00554	0'00539	0'00523	0'00508	0'00494	0'00480
-2'4	0'00820	0'00798	0'00776	0'00755	0'00734	0'00714	0'00695	0'00676	0'00657	0'00639
-2'3	0'01072	0'01044	0'01017	0'00990	0'00964	0'00939	0'00914	0'00889	0'00866	0'00842
-2'2	0'01390	0'01355	0'01321	0'01287	0'01255	0'01222	0'01191	0'01160	0'01130	0'01101
-2'1	0'01786	0'01743	0'01700	0'01659	0'01618	0'01578	0'01539	0'01500	0'01463	0'01426
-2'0	0'02275	0'02222	0'02169	0'02118	0'02068	0'02018	0'01970	0'01923	0'01876	0'01831
-1'9	0'02872	0'02807	0'02743	0'02680	0'02619	0'02559	0'02500	0'02442	0'02385	0'02330
-1'8	0'03593	0'03515	0'03438	0'03362	0'03288	0'03216	0'03144	0'03074	0'03005	0'02938
-1'7	0'04457	0'04363	0'04272	0'04182	0'04093	0'04006	0'03920	0'03836	0'03754	0'03673
-1'6	0'05480	0'05370	0'05262	0'05155	0'05050	0'04947	0'04846	0'04746	0'04648	0'04551
-1'5	0'06681	0'06552	0'06426	0'06301	0'06178	0'06057	0'05938	0'05821	0'05705	0'05592
-1'4	0'08076	0'07927	0'07780	0'07636	0'07493	0'07353	0'07214	0'07078	0'06944	0'06811
-1'3	0'09680	0'09510	0'09342	0'09176	0'09012	0'08851	0'08692	0'08534	0'08379	0'08226
-1'2	0'11507	0'11314	0'11123	0'10935	0'10749	0'10565	0'10383	0'10204	0'10027	0'09853
-1'1	0'13567	0'13350	0'13136	0'12924	0'12714	0'12507	0'12302	0'12100	0'11900	0'11702
-1'0	0'15866	0'15625	0'15386	0'15150	0'14917	0'14687	0'14457	0'14231	0'14007	0'13786
-0'9	0'18406	0'18141	0'17879	0'17619	0'17361	0'17106	0'16853	0'16602	0'16354	0'16109
-0'8	0'21186	0'20897	0'20611	0'20327	0'20045	0'19766	0'19489	0'19215	0'18925	0'18673
-0'7	0'24196	0'23885	0'23576	0'23270	0'22965	0'22663	0'22363	0'22065	0'21770	0'21476
-0'6	0'27425	0'27093	0'26763	0'26435	0'26109	0'25785	0'25463	0'25143	0'24825	0'24510
-0'5	0'30854	0'30503	0'30153	0'29806	0'29550	0'29116	0'28774	0'28434	0'28096	0'27760
-0'4	0'34446	0'34090	0'33724	0'33360	0'32997	0'32636	0'32276	0'31918	0'31561	0'31207
-0'3	0'38209	0'37828	0'37448	0'37070	0'36693	0'36317	0'35942	0'35569	0'35197	0'34827
-0'2	0'42074	0'41683	0'41294	0'40905	0'40517	0'40129	0'39743	0'39358	0'38974	0'38591
-0'1	0'46017	0'45620	0'45234	0'44828	0'44433	0'44038	0'43644	0'43251	0'42858	0'42465
-0'0	0'50000	0'49601	0'49202	0'48803	0'48405	0'48006	0'47608	0'47210	0'46812	0'46414

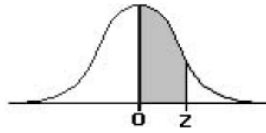
Tabla C Distribución Normal (Continuación)



La tabla proporciona, para cada valor de z , el área que queda a su izquierda.

z	0'00	0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09
0'0	0'50000	0'50399	0'50798	0'51197	0'51595	0'51994	0'52392	0'52790	0'53188	0'53586
0'1	0'53983	0'54380	0'54766	0'55172	0'55567	0'55962	0'56356	0'56749	0'57142	0'57535
0'2	0'57926	0'58317	0'58706	0'59095	0'59483	0'59871	0'60257	0'60642	0'61026	0'61409
0'3	0'61791	0'62172	0'62552	0'62930	0'63307	0'63683	0'64058	0'64431	0'64803	0'65173
0'4	0'65554	0'65910	0'66276	0'66640	0'67003	0'67364	0'67724	0'68082	0'68439	0'68793
0'5	0'69146	0'69497	0'69847	0'70194	0'70450	0'70884	0'71226	0'71566	0'71904	0'72240
0'6	0'72575	0'72907	0'73237	0'73565	0'73891	0'74215	0'74537	0'74857	0'75175	0'75490
0'7	0'75804	0'76115	0'76424	0'76730	0'77035	0'77337	0'77637	0'77935	0'78230	0'78524
0'8	0'78814	0'79103	0'79389	0'79673	0'79955	0'80234	0'80511	0'80785	0'81075	0'81327
0'9	0'81594	0'81859	0'82121	0'82381	0'82639	0'82894	0'83147	0'83398	0'83646	0'83891
1'0	0'84134	0'84375	0'84614	0'84850	0'85083	0'85313	0'85543	0'85769	0'85993	0'86214
1'1	0'86433	0'86650	0'86864	0'87076	0'87286	0'87493	0'87698	0'87900	0'88100	0'88298
1'2	0'88493	0'88686	0'88877	0'89065	0'89251	0'89435	0'89617	0'89796	0'89973	0'90147
1'3	0'90320	0'90490	0'90658	0'90824	0'90988	0'91149	0'91308	0'91466	0'91621	0'91774
1'4	0'91924	0'92073	0'92220	0'92364	0'92507	0'92647	0'92786	0'92922	0'93056	0'93189
1'5	0'93319	0'93448	0'93574	0'93699	0'93822	0'93943	0'94062	0'94179	0'94295	0'94408
1'6	0'94520	0'94630	0'94738	0'94845	0'94950	0'95053	0'95154	0'95254	0'95352	0'95449
1'7	0'95543	0'95637	0'95728	0'95818	0'95907	0'95994	0'96080	0'96164	0'96246	0'96327
1'8	0'96407	0'96485	0'96562	0'96638	0'96712	0'96784	0'96856	0'96926	0'96995	0'97062
1'9	0'97128	0'97193	0'97257	0'97320	0'97381	0'97441	0'97500	0'97558	0'97615	0'97670
2'0	0'97725	0'97778	0'97831	0'97882	0'97932	0'97982	0'98030	0'98077	0'98124	0'98169
2'1	0'98214	0'98257	0'98300	0'98341	0'98382	0'98422	0'98461	0'98500	0'98537	0'98574
2'2	0'98610	0'98645	0'98679	0'98713	0'98745	0'98778	0'98809	0'98840	0'98870	0'98899
2'3	0'98928	0'98956	0'98983	0'99010	0'99036	0'99061	0'99086	0'99111	0'99134	0'99158
2'4	0'99180	0'99202	0'99224	0'99245	0'99266	0'99286	0'99305	0'99324	0'99343	0'99361
2'5	0'99379	0'99396	0'99413	0'99430	0'99446	0'99461	0'99477	0'99492	0'99506	0'99520
2'6	0'99534	0'99547	0'99560	0'99573	0'99585	0'99598	0'99609	0'99621	0'99632	0'99643
2'7	0'99653	0'99664	0'99674	0'99683	0'99693	0'99702	0'99711	0'99720	0'99728	0'99736
2'8	0'99744	0'99752	0'99760	0'99767	0'99774	0'99781	0'99788	0'99795	0'99801	0'99807
2'9	0'99813	0'99819	0'99825	0'99831	0'99836	0'99841	0'99846	0'99851	0'99856	0'99861
3'0	0'99865	0'99869	0'99873	0'99877	0'99881	0'99885	0'99889	0'99893	0'99896	0'99899
3'1	0'99903	0'99906	0'99909	0'99912	0'99915	0'99918	0'99921	0'99923	0'99926	0'99929
3'2	0'99931	0'99933	0'99936	0'99938	0'99940	0'99942	0'99944	0'99946	0'99948	0'99950
3'3	0'99951	0'99953	0'99955	0'99956	0'99958	0'99959	0'99961	0'99962	0'99964	0'99965
3'4	0'99966	0'99967	0'99968	0'99970	0'99971	0'99972	0'99973	0'99974	0'99975	0'99976
3'5	0'99977	0'99977	0'99978	0'99979	0'99980	0'99981	0'99981	0'99982	0'99983	0'99983
3'6	0'99984	0'99985	0'99985	0'99986	0'99986	0'99987	0'99987	0'99988	0'99988	0'99989
3'7	0'99989	0'99990	0'99990	0'99990	0'99991	0'99991	0'99991	0'99992	0'99992	0'99992
3'8	0'99993	0'99993	0'99993	0'99994	0'99994	0'99994	0'99994	0'99995	0'99995	0'99995
3'9	0'99995	0'99995	0'99996	0'99996	0'99996	0'99996	0'99996	0'99996	0'99997	0'99997
4'0	0'99997	0'99997	0'99997	0'99997	0'99997	0'99997	0'99998	0'99998	0'99998	0'99998
4'1	0'99998	0'99998	0'99998	0'99998	0'99998	0'99998	0'99998	0'99998	0'99999	0'99999
4'2	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999
4'3	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999	0'99999
4'4	0'99999	0'99999	0'99999	1'00000	1'00000	1'00000	1'00000	1'00000	1'00000	1'00000

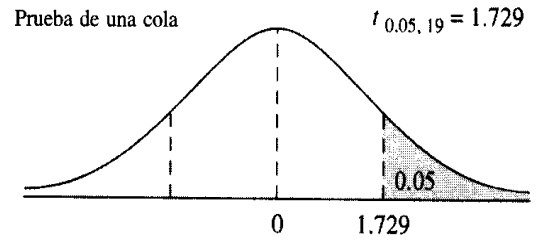
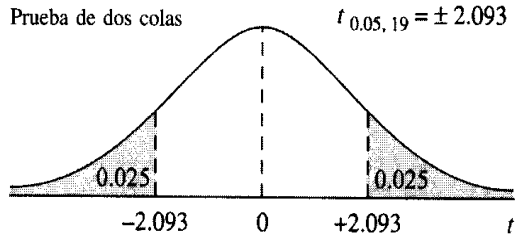
Tabla C Distribución Normal (Continuación)



La tabla proporciona el área que queda comprendida entre 0 y z.

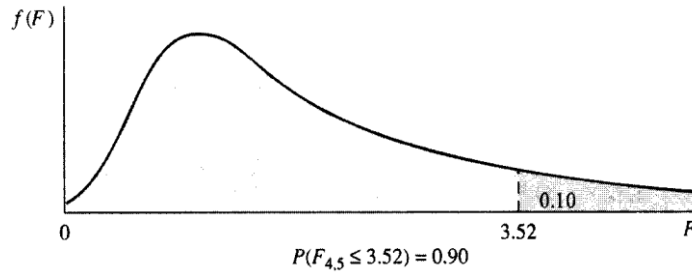
z	0'00	0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09
0'0	0'00000	0'00399	0'00798	0'01197	0'01595	0'01994	0'02392	0'02790	0'03188	0'03586
0'1	0'03983	0'04380	0'04766	0'05172	0'05567	0'05962	0'06356	0'06749	0'07142	0'07535
0'2	0'07926	0'08317	0'08706	0'09095	0'09483	0'09871	0'10257	0'10642	0'11026	0'11409
0'3	0'11791	0'12172	0'12552	0'12930	0'13307	0'13683	0'14058	0'14431	0'14803	0'15173
0'4	0'15554	0'15910	0'16276	0'16640	0'17003	0'17364	0'17724	0'18082	0'18439	0'18793
0'5	0'19146	0'19497	0'19847	0'20194	0'20450	0'20884	0'21226	0'21566	0'21904	0'22240
0'6	0'22575	0'22907	0'23237	0'23565	0'23891	0'24215	0'24537	0'24857	0'25175	0'25490
0'7	0'25804	0'26115	0'26424	0'26730	0'27035	0'27337	0'27637	0'27935	0'28230	0'28524
0'8	0'28814	0'29103	0'29389	0'29673	0'29955	0'30234	0'30511	0'30785	0'31075	0'31327
0'9	0'31594	0'31859	0'32121	0'32381	0'32639	0'32894	0'33147	0'33398	0'33646	0'33891
1'0	0'34134	0'34375	0'34614	0'34850	0'35083	0'35313	0'35543	0'35769	0'35993	0'36214
1'1	0'36433	0'36650	0'36864	0'37076	0'37286	0'37493	0'37698	0'37900	0'38100	0'38298
1'2	0'38493	0'38686	0'38877	0'39065	0'39251	0'39435	0'39617	0'39796	0'39973	0'40147
1'3	0'40320	0'40490	0'40658	0'40824	0'40988	0'41149	0'41308	0'41466	0'41621	0'41774
1'4	0'41924	0'42073	0'42220	0'42364	0'42507	0'42647	0'42786	0'42922	0'43056	0'43189
1'5	0'43319	0'43448	0'43574	0'43699	0'43822	0'43943	0'44062	0'44179	0'44295	0'44408
1'6	0'44520	0'44630	0'44738	0'44845	0'44950	0'45053	0'45154	0'45254	0'45352	0'45449
1'7	0'45543	0'45637	0'45728	0'45818	0'45907	0'45994	0'46080	0'46164	0'46246	0'46327
1'8	0'46407	0'46485	0'46562	0'46638	0'46712	0'46784	0'46856	0'46926	0'46995	0'47062
1'9	0'47128	0'47193	0'47257	0'47320	0'47381	0'47441	0'47500	0'47558	0'47615	0'47670
2'0	0'47725	0'47778	0'47831	0'47882	0'47932	0'47982	0'48030	0'48077	0'48124	0'48169
2'1	0'48214	0'48257	0'48300	0'48341	0'48382	0'48422	0'48461	0'48500	0'48537	0'48574
2'2	0'48610	0'48645	0'48679	0'48713	0'48745	0'48778	0'48809	0'48840	0'48870	0'48899
2'3	0'48928	0'48956	0'48983	0'49010	0'49036	0'49061	0'49086	0'49111	0'49134	0'49158
2'4	0'49180	0'49202	0'49224	0'49245	0'49266	0'49286	0'49305	0'49324	0'49343	0'49361
2'5	0'49379	0'49396	0'49413	0'49430	0'49446	0'49461	0'49477	0'49492	0'49506	0'49520
2'6	0'49534	0'49547	0'49560	0'49573	0'49585	0'49598	0'49609	0'49621	0'49632	0'49643
2'7	0'49653	0'49664	0'49674	0'49683	0'49693	0'49702	0'49711	0'49720	0'49728	0'49736
2'8	0'49744	0'49752	0'49760	0'49767	0'49774	0'49781	0'49788	0'49795	0'49801	0'49807
2'9	0'49813	0'49819	0'49825	0'49831	0'49836	0'49841	0'49846	0'49851	0'49856	0'49861
3'0	0'49865	0'49869	0'49873	0'49877	0'49881	0'49885	0'49889	0'49893	0'49896	0'49899
3'1	0'49903	0'49906	0'49909	0'49912	0'49915	0'49918	0'49921	0'49923	0'49926	0'49929
3'2	0'49931	0'49933	0'49936	0'49938	0'49940	0'49942	0'49944	0'49946	0'49948	0'49950
3'3	0'49951	0'49953	0'49955	0'49956	0'49958	0'49959	0'49961	0'49962	0'49964	0'49965
3'4	0'49966	0'49967	0'49968	0'49970	0'49971	0'49972	0'49973	0'49974	0'49975	0'49976
3'5	0'49977	0'49977	0'49978	0'49979	0'49980	0'49981	0'49981	0'49982	0'49983	0'49983
3'6	0'49984	0'49985	0'49985	0'49986	0'49986	0'49987	0'49987	0'49988	0'49988	0'49989
3'7	0'49989	0'49990	0'49990	0'49990	0'49991	0'49991	0'49991	0'49992	0'49992	0'49992
3'8	0'49993	0'49993	0'49993	0'49994	0'49994	0'49994	0'49994	0'49995	0'49995	0'49995
3'9	0'49995	0'49995	0'49996	0'49996	0'49996	0'49996	0'49996	0'49996	0'49997	0'49997
4'0	0'49997	0'49997	0'49997	0'49997	0'49997	0'49997	0'49998	0'49998	0'49998	0'49998
4'1	0'49998	0'49998	0'49998	0'49998	0'49998	0'49998	0'49998	0'49998	0'49999	0'49999
4'2	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999
4'3	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999	0'49999
4'4	0'49999	0'49999	0'49999	0'50000	0'50000	0'50000	0'50000	0'50000	0'50000	0'50000

Tabla D Distribución t



	0.900	0.700	0.500	0.300	0.200	0.100	0.050	0.020	0.010	Valor α	} Prueba de dos colas
	0.100	0.300	0.500	0.700	0.800	0.900	0.950	0.980	0.990	IC	
	0.450	0.350	0.250	0.150	0.100	0.050	0.025	0.010	0.005	Valor α	} Prueba de una cola
	0.550	0.650	0.750	0.850	0.900	0.950	0.975	0.990	0.995	IC	
g.l.	Valores de t										
1	0.158	0.510	1.000	1.963	3.078	6.314	12.706	31.821	63.657		
2	0.142	0.445	0.816	1.386	1.886	2.920	4.303	6.965	9.925		
3	0.137	0.424	0.765	1.250	1.638	2.353	3.182	4.541	5.841		
4	0.134	0.414	0.741	1.190	1.533	2.132	2.776	3.747	4.604		
5	0.132	0.408	0.727	1.156	1.476	2.015	2.571	3.365	4.032		
6	0.131	0.404	0.718	1.134	1.440	1.943	2.447	3.143	3.707		
7	0.130	0.402	0.711	1.119	1.415	1.895	2.365	2.998	3.499		
8	0.130	0.399	0.706	1.108	1.397	1.860	2.306	2.896	3.355		
9	0.129	0.398	0.703	1.100	1.383	1.833	2.262	2.821	3.250		
10	0.129	0.397	0.700	1.093	1.372	1.812	2.228	2.764	3.169		
11	0.129	0.396	0.697	1.088	1.363	1.796	2.201	2.718	3.106		
12	0.128	0.395	0.695	1.083	1.356	1.782	2.179	2.681	3.055		
13	0.128	0.394	0.694	1.079	1.350	1.771	2.160	2.650	3.012		
14	0.128	0.393	0.692	1.076	1.345	1.761	2.145	2.624	2.977		
15	0.128	0.393	0.691	1.074	1.341	1.753	2.131	2.602	2.947		
16	0.128	0.392	0.690	1.071	1.337	1.746	2.120	2.583	2.921		
17	0.128	0.392	0.689	1.069	1.333	1.740	2.110	2.567	2.898		
18	0.127	0.392	0.688	1.067	1.330	1.734	2.101	2.552	2.878		
19	0.127	0.391	0.688	1.066	1.328	1.729	2.093	2.539	2.861		
20	0.127	0.391	0.687	1.064	1.325	1.725	2.086	2.528	2.845		
21	0.127	0.391	0.686	1.063	1.323	1.721	2.080	2.518	2.831		
22	0.127	0.390	0.686	1.061	1.321	1.717	2.074	2.508	2.819		
23	0.127	0.390	0.685	1.060	1.319	1.714	2.069	2.500	2.807		
24	0.127	0.390	0.685	1.059	1.318	1.711	2.064	2.492	2.797		
25	0.127	0.390	0.684	1.058	1.316	1.708	2.060	2.485	2.787		
26	0.127	0.390	0.684	1.058	1.315	1.706	2.056	2.479	2.779		
27	0.127	0.389	0.684	1.057	1.314	1.703	2.052	2.473	2.771		
28	0.127	0.389	0.683	1.056	1.313	1.701	2.048	2.467	2.763		
29	0.127	0.389	0.683	1.055	1.311	1.699	2.045	2.462	2.756		
30	0.127	0.389	0.683	1.055	1.310	1.697	2.042	2.457	2.750		
40	0.126	0.388	0.681	1.050	1.303	1.684	2.021	2.423	2.704		
60	0.126	0.387	0.679	1.045	1.296	1.671	2.000	2.390	2.660		
120	0.126	0.386	0.677	1.041	1.289	1.658	1.980	2.358	2.617		
∞	0.126	0.385	0.674	1.036	1.282	1.645	1.960	2.326	2.576		

Tabla E Distribución F



Grados de libertad del denominador	$F_{0.90}; \alpha = 0.10$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.13	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.95	1.85	1.77	1.72	1.67	1.63

Tabla E Distribución F (Continuación)

Grados de libertad del denominador	$F_{0.90}; \alpha = 0.10$									
	Grados de libertad del numerador									
	10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Grados de libertad del denominador	$F_{0.95}; \alpha = 0.05$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00

Tabla E Distribución F (Continuación)

Grados de libertad del denominador	$F_{0.95}; \alpha = 0.05$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88

Grados de libertad del denominador	$F_{0.95}; \alpha = 0.05$									
	Grados de libertad del numerador									
	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71

Tabla E Distribución F (Continuación)

Grados de libertad del denominador	$F_{0.05}; \alpha = 0.05$									
	Grados de libertad del numerador									
	10	12	15	20	24	30	40	60	120	∞
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Grados de libertad del denominador	$F_{0.975}; \alpha = 0.025$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31

Tabla E Distribución F (Continuación)

Grados de libertad del denominador	$F_{0.975}; \alpha = 0.025$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Grados de libertad del denominador	$F_{0.975}; \alpha = 0.025$									
	Grados de libertad del numerador									
	10	12	15	20	24	30	40	60	120	∞
1	968.6	976.7	984.9	993.1	997.2	1001.4	1005.6	1009.8	1014.0	1018.2
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.41	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.73
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32

Tabla E Distribución F (Continuación)

Grados de libertad del denominador	$F_{0.975}; \alpha = 0.025$									
	Grados de libertad del numerador									
	10	12	15	20	24	30	40	60	120	∞
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.95	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Grados de libertad del denominador	$F_{0.99}; \alpha = 0.01$								
	Grados de libertad del numerador								
	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52

Tabla E Distribución F (Continuación)

		$F_{0.99}; \alpha = 0.01$								
Grados de libertad del denominador	Grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	

		$F_{0.99}; \alpha = 0.01$									
Grados de libertad del denominador	Grados de libertad del numerador										
	10	12	15	20	24	30	40	60	120	∞	
1	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.8	
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	

Tabla E Distribución F (Continuación)

		$F_{0.99}; \alpha = 0.01$									
Grados de libertad del denominador	Grados de libertad del numerador										
	10	12	15	20	24	30	40	60	120	∞	
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81	
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	
∞	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00	

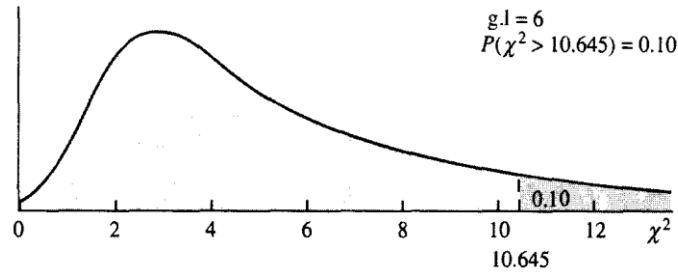
		$F_{0.995}; \alpha = 0.005$								
Grados de libertad del denominador	Grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	
2	198.50	199.00	199.17	199.25	199.30	199.33	199.36	199.37	199.39	
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	

Tabla E Distribución F (Continuación)

		$F_{0.995}; \alpha = 0.005$								
Grados de libertad del denominador	Grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.75	2.62	

		$F_{0.995}; \alpha = 0.005$									
Grados de libertad del denominador	Grados de libertad del numerador										
	10	12	15	20	24	30	40	60	120	∞	
1	24224	24426	24630	24836	24940	25044	25148	25253	25359	25464	
2	199.40	199.42	199.43	199.45	199.46	199.47	199.47	199.48	199.49	199.50	
3	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83	
4	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.33	
5	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14	
6	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88	
7	8.38	8.18	7.97	7.75	7.64	7.53	7.42	7.31	7.19	7.08	
8	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95	
9	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19	
10	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64	
11	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.45	4.34	4.23	
12	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90	
13	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65	
14	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44	
15	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26	
16	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11	
17	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98	
18	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87	
19	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78	
20	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69	
21	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61	
22	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55	
23	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48	
24	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43	
25	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38	
26	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33	
27	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29	
28	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25	
29	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21	
30	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18	
40	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93	
60	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69	
120	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43	
∞	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00	

Tabla F Distribución Ji - Cuadrado



g.l.	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.700}$	$\chi^2_{0.500}$	$\chi^2_{0.300}$	$\chi^2_{0.200}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.020}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$
1	0.000	0.000	0.001	0.004	0.016	0.148	0.455	1.074	1.642	2.706	3.841	5.024	5.412	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.713	1.386	2.408	3.219	4.605	5.991	7.378	7.824	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.424	2.366	3.665	4.642	6.251	7.815	9.348	9.837	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	2.195	3.357	4.878	5.989	7.779	9.488	11.143	11.668	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	3.000	4.351	6.064	7.289	9.236	11.070	12.833	13.388	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.828	5.348	7.231	8.558	10.645	12.592	14.449	15.033	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.671	6.346	8.383	9.803	12.017	14.067	16.013	16.622	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.527	7.344	9.524	11.030	13.362	15.507	17.535	18.168	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	6.393	8.343	10.656	12.242	14.684	16.919	19.023	19.679	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	7.267	9.342	11.781	13.442	15.987	18.307	20.483	21.161	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	8.148	10.341	12.899	14.631	17.275	19.675	21.920	22.618	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	9.034	11.340	14.011	15.812	18.549	21.026	23.337	24.054	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	9.926	12.340	15.119	16.985	19.812	22.362	24.736	25.472	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.821	13.339	16.222	18.151	21.064	23.685	26.119	26.873	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.721	14.339	17.322	19.311	22.307	24.996	27.488	28.259	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	12.624	15.338	18.418	20.465	23.542	26.296	28.845	29.633	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	13.531	16.338	19.511	21.615	24.769	27.587	30.191	30.995	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	14.440	17.338	20.601	22.760	25.989	28.869	31.526	32.346	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	15.352	18.338	21.689	23.900	27.204	30.144	32.852	33.687	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	16.266	19.337	22.775	25.038	28.412	31.410	34.170	35.020	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	17.182	20.337	23.858	26.171	29.615	32.671	35.479	36.343	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	18.101	21.337	24.939	27.301	30.813	33.924	36.781	37.659	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	19.021	22.337	26.018	28.429	32.007	35.172	38.076	38.968	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.943	23.337	27.096	29.553	33.196	36.415	39.364	40.270	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	20.867	24.337	28.172	30.675	34.382	37.652	40.646	41.566	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	21.792	25.336	29.246	31.795	35.563	38.885	41.923	42.856	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	22.719	26.336	30.319	32.912	36.741	40.113	43.194	44.140	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	23.647	27.336	31.391	34.027	37.916	41.337	44.461	45.419	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	24.577	28.336	32.461	35.139	39.087	42.557	45.722	46.693	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	25.508	29.336	33.530	36.250	40.256	43.773	46.979	47.962	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	34.872	39.335	44.165	47.269	51.805	55.758	59.342	60.436	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	44.313	49.335	54.723	58.164	63.167	67.505	71.420	72.613	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	53.809	59.335	65.227	68.972	74.397	79.082	83.298	84.580	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	63.346	69.334	75.689	79.715	85.527	90.531	95.023	96.388	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	72.915	79.334	86.120	90.405	96.578	101.879	106.629	108.069	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	82.511	89.334	96.524	101.054	107.565	113.145	118.136	119.648	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	92.129	99.334	106.906	111.667	118.498	124.342	129.561	131.142	135.807	140.169