

**DETERMINACIÓN DE POTENCIALES (IPR) EN POZOS HORIZONTALES Y  
DESVIADOS PARA FLUJO EN DOS FASES USANDO UNA HERRAMIENTA  
SOFTWARE BASADA EN SIMULACIÓN MONTECARLO**

**ANDRÉS FERNANDO OROZCO IBARRA  
NELSON ARMANDO ROCHA SALDARRIAGA**

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

**2008**

**DETERMINACIÓN DE POTENCIALES (IPR) EN POZOS HORIZONTALES Y  
DESVIADOS PARA FLUJO EN DOS FASES USANDO UNA HERRAMIENTA  
SOFTWARE BASADA EN SIMULACIÓN MONTECARLO**

**ANDRÉS FERNANDO OROZCO IBARRA  
NELSON ARMANDO ROCHA SALDARRIAGA**

Trabajo de grado para optar por el título de Ing. de Petróleos

**DIRECTORES**

**Fernando Calvete, M. Sc.**  
Docente UIS

**Aristóbulo Bejarano W., M.Sc.**  
Docente UIS

**CO-DIRECTOR**  
**Jaime Flórez, M.Sc**  
Gerente PETROTECH

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-QUÍMICAS  
ESCUELA DE INGENIERÍA DE PETRÓLEOS  
BUCARAMANGA  
2008**

## DEDICATORIA

A Dios, por darme la fuerza para levantarme después de cada caída, y ser mi guía y mi protector.

A mi Papá, por su voz de aliento cada vez que lo necesité, por su comprensión, apoyo y cariño. Tus palabras ayudaron a calmar mi soberbia y mi orgullo. Sin ti, nada de esto fuera realidad. Te quiero papá.

A mi Madre, el ser que más amo en el mundo entero, el motor que mueve mis sueños y la fuerza de mi espíritu, tú eres quien me motivo a ser alguien en la vida. Cada uno de mis éxitos tiene tu firma y es para ti. Gracias mamá por tu amor, tu comprensión y por estar allí siempre que te necesité. TE AMO.

A mis hermanos, Karol Enrique, Jairo Luís y Guille por haberme acompañado durante mi niñez y haber ayudado a convertirme en el hombre que soy hoy.

A mis tías y tíos, Clara, Yolanda, Mirza, Cielo, Nelson, Carmelina, Rafa, Nene, Zulma por estar pendientes de mi y regalarme una sonrisa y un abrazo cuando lo necesite.

A Ilba y Jairo, mi segunda familia, gracias por acogerme como uno más, los quiero.

A Pablo, y Jair, mis otros hermanitos, las personas que me enseñaron lo bueno y lo malo de la vida, con ustedes aprendí a caminar, a reír y a llorar. Los quiero.

A Andrés, Fabi y Ricardo, Amigos como ustedes solo se consiguen una vez en la vida, espero siempre estar allí para ustedes.

A Yeniffer, gracias por llegar a mi vida y ser tan especial

A Adriana, gracias por ser tan especial conmigo, Cada minuto a tu lado, me hizo sentir al hombre más afortunado del mundo, Te quiero Adri.

A Lucho, Víctor, Poncho, Alfonso F, Mamfre, Dan, Kissy, Víctor F, Joselo, Juanfe, Clau, Juanis, Edwin, Nata, Ley, Pas, Luís Eduardo (Pitu) y Carlos gracias por su cariño, compañía y consejos, los quiero amigos.

A todas las personas que hicieron posible este logro.....

**NELSON ARMANDO**

## DEDICATORIA

A DIOS, mi eterno guía, quien no me abandona en ninguno de los eventos más importantes de mi vida.

A mi madre y mi padre, quienes son el más claro ejemplo del amor que se puede sentir por un hijo, y en últimas, es gracias a su infinito esfuerzo por sacarme adelante por el cual yo estoy donde estoy.

A mis hermanos Carlos y Daniel, gracias por su apoyo y cariño para conmigo.

A mi novia July, por darme su sincero amor, por apoyarme y complacerme todos los días de mi vida con su bonita compañía.

A Jáder, Eduardo, Jesús David, Nelson, Jose, Rodri, Leydi y Pilín, verdaderos amigos y hermanos, por estar a mi lado cuando estaba feliz y cuando ciertas circunstancias me bajaron el ánimo.

A Don Gilberto y Doña Amparo, quienes sin conocerme me acogieron como un miembro más de su hermosa familia.

A Carmen, mi segunda mamá, quien merece de mi todo el cariño y admiración por ser tan buena con toda mi familia.

A Mi tía Fidiam, de quien no olvido todo el apoyo, que en los momentos más cruciales, dio a mí y a mi familia.

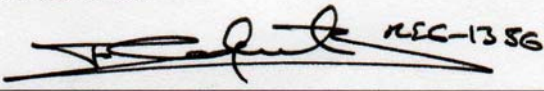
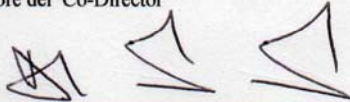


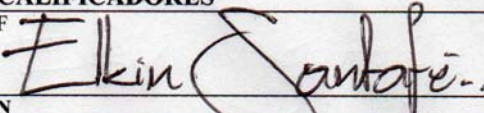
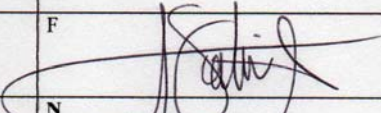
A mi prima Alma delia, a quien doy las gracias por su bondad con todos los suyos, y a quien Dios ha de darle muchísimos años más de vida.

A todos los que hicieron posibles esta realidad.

**ANDRÉS**

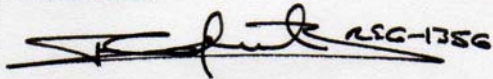



NOMBRE DEL ESTUDIANTE		CODIGO	
NELSON ARMANDO ROCHA SALDARRIAGA		2030915	
TITULO DEL PROYECTO:			
<i>“DETERMINACIÓN DE POTENCIALES (IPR) EN POZOS HORIZONTALES Y DESVIADOS PARA FLUJO EN DOS FASES USANDO UNA HERRAMIENTA SOFTWARE BASADA EN SIMULACIÓN MONTECARLO”</i>			
Registros No.	FACULTAD CIENCIAS FISICO-QUIMICAS	CARRERA ESCUELA INGENIERIA DE PETROLEOS	
Calificación (letra y número)			Créditos
APROBADO			


Nombre del Director	Firma del Director
M.Sc. FERNANDO ENRIQUE CALVETE	 REC-1356
Nombre del Co-Director	Nombre del Co-Director
MGT, ARISTOBULO BEJARANO WALLENS	
Nombre del Co-Director	Nombre del Co-Director
Msc. JAIME H. FLOREZ	
Nombre del Co-Director	Nombre del Co-Director
Ph.D. LUIGI SAPUTELLI	
<b>CALIFICADORES</b>	
F 	F 
N ING. ELKIN RODOLFO SANTAFE RANGEL	N ING. JORGE ANDRES SÁCHICA AVILA

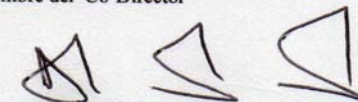




NOMBRE DEL ESTUDIANTE		CODIGO
NELSON ARMANDO ROCHA SALDARRIAGA		2030915
TITULO DEL PROYECTO:		
<p align="center"><i>“DETERMINACIÓN DE POTENCIALES (IPR) EN POZOS HORIZONTALES Y DESVIADOS PARA FLUJO EN DOS FASES USANDO UNA HERRAMIENTA SOFTWARE BASADA EN SIMULACIÓN MONTECARLO”</i></p>		
Registros No.	FACULTAD	CARRERA
	CIENCIAS FISICO-QUIMICAS	ESCUELA INGENIERIA DE PETROLEOS
Calificación (letra y número)		Créditos
CINCO, CERO (5,0)		

Nombre del Director	Firma del Director
M.Sc. FERNANDO ENRIQUE CALVETE	

Nombre del Co-Director	Nombre del Co-Director
MGT, ARISTOBULO BEJARANO WALLENS	

Nombre del Co-Director	Nombre del Co-Director
Msc. JAIME H. FLOREZ	

Nombre del Co-Director	Nombre del Co-Director
Ph.D. LUIGI SAPUTELLI	

<b>CALIFICADORES</b>	
F 	F 
N ING. ELKIN RODOLFO SANTAFE RANGEL	N ING. JORGE ANDRES SÁCHICA AVILA

## **AGRADECIMIENTOS**

Al profesor Aristóbulo Bejarano por su inmensa colaboración en el desarrollo y culminación de este proyecto.

Al Ingeniero Jaime Flórez por sacar un poco de su valioso tiempo para atender, con sus sabios consejos, todas nuestras inquietudes.

A nuestro Director Fernando Calvete, por entregarnos todos sus conocimientos para afrontar este trabajo.

Al GITAH (Grupo de Investigación en Tecnologías Alternativas Aplicadas a Hidrocarburos) y a todos sus integrantes, por ayudarnos con cada una de sus sugerencias en pro del óptimo desenvolvimiento de esta tesis.

## CONTENIDO

	<b>Pág.</b>
INTRODUCCIÓN	29
1. ASPECTOS GENERALES DEL POTENCIAL DE PRODUCCIÓN DE LOS POZOS	31
1.1 EL ÍNDICE DE PRODUCTIVIDAD J	31
1.1.1 Ecuaciones Involucradas en el cálculo del IP en este Trabajo	34
1.2 RELACIÓN DE RENDIMIENTO DE INFLUJO	37
1.2.1 Ecuaciones analíticas del flujo radial en medio poroso	39
1.2.2 Relación de rendimiento de influjo (IPR) para flujo bifásico en pozos verticales.	44
1.2.3 Relación de Rendimiento de Influjo de flujo bifásico en pozos desviados y horizontales	45
1.2.3.1 Modelo de Bendakhlia & Aziz	45
1.2.3.2 Modelo de Cheng	48
1.2.3.3 Modelo de Retnanto & Economides	49
1.2.4 Ecuación Modificada de Vogel para pozos horizontales	50
2. CONCEPTOS ESTADÍSTICOS GENERALES	55
2.1 ESTADÍSTICA DESCRIPTIVA	55
2.1.1 Distribuciones de frecuencia	55
2.1.1.1 Definición	55
2.1.1.2 Elaboración	56
2.1.1.3 Propiedades de las distribuciones de frecuencia.	57
2.1.1.4 Representaciones gráficas de una distribución de frecuencias.	58
2.1.1.5. Distribuciones de frecuencias acumuladas.	60
2.2 MEDIDAS DE TENDENCIA CENTRAL	61
2.2.1 Media Aritmética	61

2.2.1.1 Media poblacional	62
2.2.1.2 Media muestral	62
2.2.2 Media ponderada.	63
2.2.3 Mediana	63
2.2.4 Moda.	64
2.3 MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS.	64
2.3.1 Media	64
2.3.2 Mediana	65
2.3.3 Moda.	66
2.4 MEDIDAS DE DISPERSIÓN	66
2.4.1 Amplitud o intervalo de variación	66
2.4.2 Desviación media	67
2.4.3 Varianza	67
2.4.4 Desviación estándar	68
2.5 MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS	68
2.5.1 Amplitud de variación	68
2.5.2 Desviación estándar	69
2.6 DECILES, CUARTILES Y CENTILES	69
2.7 INTERPRETACIÓN Y USOS DE LA DESVIACIÓN ESTÁNDAR	73
2.8 PROBABILIDADES	77
2.8.1 Definiciones Previas	77
2.8.1.1 Experimento	78
2.8.1.2 Evento	78
2.8.1.3 Espacio muestral	80
2.8.2 Definición de Probabilidad de un evento	81
2.8.3 Puntos de vista de la probabilidad	81
2.8.3.1 Probabilidad objetiva	82
2.8.3.2 Probabilidad subjetiva	83
2.8.4 Reglas de Probabilidad	83
2.8.4.1 Regla de la adición	83

2.8.4.2 Regla de la multiplicación	84
2.8.4.3 Regla del complemento	85
2.8.5 Regla de la Probabilidad Total	85
2.8.6. Distribuciones de Probabilidad	86
2.8.6.1 Definición	87
2.9 DISTRIBUCIÓN DE PROBABILIDAD DISCRETA	92
2.9.1 Función de distribución acumulativa	92
2.9.2 Media, varianza y desviación estándar de una distribución de probabilidad discreta	93
2.9.3 Tipos de Distribuciones de Probabilidad Discreta	94
2.9.3.1 Distribución Binomial	94
2.9.3.2 Distribución De Poisson	96
2.9.3.3 Distribución Geométrica	97
2.10 DISTRIBUCIÓN DE PROBABILIDAD CONTINUA	99
2.10.1 Función de distribución acumulativa	100
2.10.2 Media, varianza y desviación estándar de una distribución de probabilidades continua.	100
2.10.3 Tipos de Distribuciones de Probabilidad Continua	101
2.10.3.1 La distribución Normal	102
2.10.3.2 Distribución Log Normal	105
2.10.3.3 Distribución Triangular	106
2.10.3.4 Distribución Gamma	109
2.10.3.5 Distribución Chi Cuadrado	110
2.10.3.6 Distribución t Student	111
3. SIMULACIÓN MONTECARLO	113
3.1 HISTORIA	113
3.2 ¿QUÉ ES LA SIMULACIÓN MONTECARLO?	114
3.2.1 Características básicas.	115
3.3 ALGORITMO PARA IMPLEMENTAR UNA SIMULACIÓN MONTECARLO	116
3.4 LAS MATEMÁTICAS DETRÁS DE LA SIMULACIÓN MONTECARLO	117

3.5 ALGUNAS DISTRIBUCIONES DE PROBABILIDAD COMÚNMENTE UTILIZADAS EN LA SMC	118
3.6 DISTRIBUCIONES DE PROBABILIDAD ASOCIADAS A CADA UNA DE LAS INCERTIDUMBRES EN UN YACIMIENTO PETROLÍFERO	120
3.7 HERRAMIENTAS SOFTWARE MÁS UTILIZADAS PARA EL ANÁLISIS DE UN PROBLEMA MEDIANTE SIMULACIÓN MONTECARLO	122
3.7.1 Simulación Montecarlo En Excel	123
3.7.1.1 La función ALEATORIO () de Excel	123
3.7.2 Simulación Montecarlo en Lumenaut	125
4. METODOLOGÍA PARA LA APLICACIÓN DE LA SIMULACIÓN MONTECARLO EN LA DETERMINACIÓN DE LOS POTENCIALES DE POZOS DESVIADOS Y HORIZONTALES.	126
4.1 METODOLOGIA PARA POZOS DESVIADOS. CAMPO COLORADO POZO COLORADO 38. IMPLEMENTACIÓN EN HOJA DE CÁLCULO	127
4.1.1. Pasos para la aplicación de la metodología.	130
5. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	194
5.1 CÁLCULO DE POTENCIALES DE PRODUCCIÓN PARA EL POZO COLORADO 38. CORRIDA EN EXCEL	194
5.2 CALCULO DE IP PARA EL POZO COLORADO 38 ZONA B REPRESENTANDO LA VISCOSIDAD Y EL FACTOR VOLUMÉTRICO CON UNA FUNCIÓN DE DENSIDAD TRIANGULAR Y LA PERMEABILIDAD Y EL ESPESOR NETO CON UNA FUNCIÓN DE DISTRIBUCION DE PROBABILIDAD LOG-NORMAL (2) Y COMPARACIÓN CON LOS RESULTADOS OBTENIDOS CUANDO SE USÓ LA DISTRIBUCION TRIANGULAR (1) PARA LA DESCRIPCION DE TODAS LAS VARIABLES.	196
5.3. CÁLCULO DE POTENCIALES DE PRODUCCIÓN PARA EL POZO HORIZONTAL DE PAPER SPE 107669-MS-P:	198
5.4 COMPARACION CORRIDA POZO HORIZONTAL EXCEL Y CON SOFTWARE PARA SIMULACION MONTECARLO, LUMENAUT (DISTRIBUCIÓN TRIANGULAR EN TODOS LOS CASOS):	199

6. CONCLUSIONES	200
RECOMENDACIONES	203
BILIOGRAFÍA	205
ANEXO	208

## LISTA DE FIGURAS

	<b>Pág.</b>
Figura 1. Grafica Del Abatimiento	33
Figura 2. Esquema Del Comportamiento De La Presion De Fondo Fluyendo	34
Figura 3. División Del Problema 3d De Pozos Horizontales, En Dos Problemas 2d	36
Figura 4. Curva de Relación de Rendimiento de Influjos	38
Figura 5. Plano de un flujo geométrico radial	40
Figura 6. Modelo de Yacimiento de Vogel	44
Figura 7. Modelo de Bendakhlia y Aziz y modelo de Cheng	46
Figura 8. Parámetros V y n para la ecuación de Bendakhlia y Aziz.	47
Figura 9. Modelo de yacimiento de la ecuación de Helmy & Wattenbarger	54
Figura 10. Histograma	58
Figura 11. Polígono de frecuencias	59
Figura 12. Distribución de Frecuencia acumulada	60
Figura 13. Esquematación de las posiciones de A) Cuartiles, B) Deciles, C) Centiles	70
Figura 14. Polígono de frecuencias con 7 clases	73
Figura 15. Polígono de frecuencias con 14 clases	74
Figura 16. Polígono de frecuencias con infinitas clases	74
Figura 17. Distribución de frecuencias simétrica o de campana.	75
Figura 18. Distribución de frecuencias con simetría positiva y ubicación de las medidas de tendencia central	76
Figura 19. Distribución de frecuencias con simetría negativa y ubicación de las medidas de tendencia central	76
Figura 20. Partición del espacio muestral S, compuesta por los eventos E1, E2, E3,, En	85
Figura 21. Partición del espacio muestral S, compuesta por los eventos E1, E2, E3,,..., En y el evento cualquiera B	86

Figura 22. Diagrama de árbol para las probabilidades de sacar 2 esferas en el ejemplo 1	88
Figura 23. Diagrama de Venn para la variable aleatoria que cuenta el número de esferas negras en el ejemplo 1.	90
Figura 24. Gráfica de la distribución de probabilidades o función de densidad de probabilidad de la variable aleatoria que cuenta el número de esferas negras en el ejemplo 1.	91
Figura 25. Curva de una distribución normal, con la media y una desviación estándar $\sigma$	103
Figura 26. Porcentaje de área comprendida entre $\mu - \sigma$ y $\mu + \sigma$ .	104
Figura 27. Porcentaje de área comprendida entre $\mu - 2\sigma$ y $\mu + 2\sigma$ .	104
Figura 28. Curva De Una Distribución Log Normal	106
Figura 29. Representación Gráfica de la Función de Densidad Triangular dependiendo de que	107
Figura 30. Función de Distribución Gamma.	109
Figura 31. Distribución Chi Cuadrado.	110
Figura 32. Distribución t Student	112
Figura 33. Función ALEATORIO de Excel	124
Figura 34. Calculo De Ip Mediante La Generación De Muestras Aleatorias, Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.	132
Figura 35. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.	133
Figura 36. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1812-1829, Pozo Colorado 38, Zona B.	134
Figura37. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B	136
Figura 38. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B.	136
Figura 39. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B.	138

Figura 40. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B.	138
Figura 41. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B.	140
Figura 42. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B.	140
Figura 43. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B	142
Figura 44. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B.	142
Figura 45. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel. Restricciones Del Tubo. Pozo Colorado 38, Zona B	144
Figura 46. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel Y Tomando Como Referencia El Ip Del P10. Restricciones Del Tubo, Pozo Colorado 38, Zona B.	149
Figura 47. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C	151
Figura 48. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C.	151
Figura 49. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C	153
Figura 50. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C.	153
Figura 51. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C	155
Figura 52. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C.	155
Figura 53. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C	157

Figura 54. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C.	157
Figura 55. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C	159
Figura 56. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C.	159
Figura 57. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel. Restricciones Del Tubo. Pozo Colorado 38, Zona C	160
Figura 58. Esquema Que Ilustra La Manera De Generar La Función De Distribución De Densidad Para La Viscosidad.	165
Figura 59. Esquema Que Ilustra La Forma De Realizar Un Análisis De Datos Luego De La Aplicación Del Método Montecarlo En Excel.	166
Figura 60. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 600 Metros	167
Figura 61. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 600 Metros.	168
Figura 62. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 800 Metros	169
Figura 63. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 800 Metros.	170
Figura 64. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 1000 Metros	171
Figura 65. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 1000 Metros.	172
Figura 66. Cuadro De Frecuencia Acumulada (Percentiles) Versus Índices De Producción. Comparación Entre Las Longitudes De Pozo 600, 800 Y 1000 Metros	173
Figura 67. Función De Densidad De Probabilidad Para La Permeabilidad Horizontal.	176
Figura 68. Función De Densidad De Probabilidad Para El Espesor Neto	177

Figura 69. Función De Densidad De Probabilidad Para La Viscosidad.	178
Figura 70. Función De Densidad De Probabilidad Para Beta.	179
Figura 71. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia.Longitud De Pozo 600 Metros	181
Figura 72. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia.Longitud De Pozo 800 Metros	182
Figura 73. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia.Longitud De Pozo 1000 Metros	183
Figura 74. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1812-1829 Ft Usando Un Función De Densidad Triangular Para La Viscosidad Y El Factor Volumetrico, Y Una Función De Densidad Log-Normal Para La Permeabilidad Y El Espesor Neto, Pozo Colorado 38, Zona B	186
Figura 75. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1812-1829, Pozo Colorado 38, Zona B. Corrida 2	187
Figura 76. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1862-1878, Pozo Colorado 38, Zona B. Corrida 2	188
Figura 77. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1905-1909, Pozo Colorado 38, Zona B. Corrida 2	189
Figura 78. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1938-1943, Pozo Colorado 38, Zona B. Corrida 2	190
Figura 79. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 2110-2118, Pozo Colorado 38, Zona B. Corrida 2	191
Figura 80. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel. Restricciones Del Tubo. Pozo Colorado 38, Zona B. Corrida 2	192

## LISTA DE TABLAS

Pág.

Tabla 1. Constantes usadas en la correlación de Cheng.	49
Tabla 2. Porcentajes que separan los cuartiles de un conjunto de datos.	71
Tabla 3. Porcentajes que separan los deciles de un conjunto de datos.	71
Tabla 4. Porcentajes que separan los centiles de un conjunto de datos.	72
Tabla 5. Distribución de probabilidades para el ejemplo planteado.	89
Tabla 6. Distribución de Probabilidades y variables aleatoria que cuenta las esferas negras en el ejemplo 1.	90
Tabla 7. Función de distribución de la variable aleatoria X que cuenta el número de esferas negras del ejemplo 1.	93
Tabla 8. Distribuciones de Probabilidad asociadas a cada una de las incertidumbres en un yacimiento petrolífero.	121
Tabla 9. Intervalos Productores Para El Pozo Colorado 38 Zona B	130
Tabla 10. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1812-1829 Ft) Colorado 38, Zona B.	131
Tabla 11. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.	135
Tabla 12. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1862-1878 Ft) Pozo Colorado 38, Zona B.	135
Tabla 13. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B.	137
Tabla 14. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1905-1909 Ft) Pozo Colorado 38, Zona B.	137
Tabla 15. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B	139
Tabla 16. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1938-1943 Ft) Pozo Colorado 38, Zona B.	139

Tabla 17. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B	141
Tabla 18. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (2110-2118 Ft) Pozo Colorado 38, Zona B.	141
Tabla 19. Marcas De Clase (Ip) Y Frecuencia, Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B	143
Tabla 20. Índice De Productividad, Presión De Fondo Fluyendo Y Presión De Yacimiento De La Zona B, Pozo Colorado 38	144
Tabla 21. Presiones De Fondo Fluyendo Como Datos De Entrada Para La Ecuación De Vogel Y Obtención De Caudales (Ipr) Pozo Colorado 38, Zona B.	145
Tabla 22. Caudal Esperado De La Zona B A La Presión A La Cual Se Encuentra Produciendo	146
Tabla 23. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.	146
Tabla 24. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos, Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B	147
Tabla 25. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B	148
Tabla 26. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B	148
Tabla 27. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B	149
Tabla 28. Caudal Esperado De La Zona B A La Presión A La Cual Se Encuentra Produciendo. Referencia Ip Del P10	150
Tabla 29. Intervalos Productores Para El Pozo Colorado 38 Zona C	150
Tabla 30. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3383-3386 Ft) Colorado 38, Zona C.	150
Tabla 31. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C.	152

Tabla 32. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3453-3459 Ft) Colorado 38, Zona C.	152
Tabla 33. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C.	154
Tabla 34. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3488-3506 Ft) Colorado 38, Zona C.	154
Tabla 35. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C.	156
Tabla 36. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3705-3713 Ft) Colorado 38, Zona C.	156
Tabla 37. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C.	158
Tabla 38. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (4006-4018 Ft) Colorado 38, Zona C.	158
Tabla 39. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C.	160
Tabla 40. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Involucrada En El Cálculo Del Ip Según Ecuación De Joshi.	164
Tabla 41. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 600 Metros	168
Tabla 42. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 800 Metros	170
Tabla 43. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 1000 Metros	172
Tabla 44. Valores Mínimo, Más Probable Y Máximo Para La Permeabilidad Horizontal	176
Tabla 45. Valores Mínimo, Más Probable Y Máximo Para El Espesor Neto	177
Tabla 46. Valores Mínimo, Más Probable Y Máximo Para La Viscosidad	178
Tabla 47. Valores Mínimo, Más Probable Y Máximo Para Beta	179

Tabla 48. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitu De Pozo 600 Metros.	181
Tabla 49. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitu De Pozo 800 Metros.	183
Tabla 50. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitu De Pozo 1000 Metros.	184
Tabla 51. Funciones De Distribución A Usar Para Una Nueva Corrida, De Acuerdo A tabla 8.	185
Tabla 52. Media Geométrica Y Desviación Estandar Del Logaritmo Natural De Valores De Permeabilidad	185
Tabla 53. Presiones De Fondo Fluyendo Como Datos De Entrada Para La Ecuación De Vogel Y Obtención De Caudales (Ipr) Pozo Colorado 38, Zona B. Corrida 2	193
Tabla 54. Índices De Producción Obtenidos Para La Zona B Del Pozo Colorado 38. Implementación En Hoja De Cálculo Usando Una Función De Densidad De Probabilidad Triangular Para Describir Todas Las Variables.	195
Tabla 55. Índices De Producción Obtenidos Para La Zona C Del Pozo Colorado 38. Implementación En Hoja De Cálculo Usando Una Función De Densidad De Probabilidad Triangular Para Describir Todas Las Variables.	195
Tabla 56. Comparación De Ip Al Describir Las Variables Según Tabla 8, Y Al Usar La Función Triangular Para Describir Todas Las Variables.	196
Tabla 57. Comparación De Ipr'S Corrida 1 Y Corrida 2	197
Tabla 58. Comparación Entre Los Ip Hallados Por Los Autores Y Los Que Muestra El Paper Spe 107669	198
Tabla 59. Comparación Entre Los Ip Hallados Excel Y Los Hallados En Lumenaut. Confrontación Con Artículo 107669	199

## LISTA DE ANEXOS

	<b>Pág.</b>
ANEXO A. INTRODUCCIÓN A LA SIMULACIÓN CON EXCEL	208
A.1 NÚMEROS ALEATORIOS	208
A.1.1 PROCEDIMIENTOS RELACIONADOS	208
A.1.2 HERRAMIENTA EXCEL PARA LA GENERACIÓN DE NÚMEROS ALEATORIOS	208
A.1.3 GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE FÓRMULAS	215
A.2 DISTRIBUCIÓN DE FRECUENCIAS	218
A.2.1 PROCEDIMIENTOS RELACIONADOS	218
A.2.2 ANÁLISIS DE DATOS: HISTOGRAMA	218
A.2.3 LA FUNCIÓN FRECUENCIA	219
A.3 MEDIDAS DE TENDENCIA CENTRAL, VARIACIÓN Y FORMA	220
A.3.1 PROCEDIMIENTOS RELACIONADOS	221
A.3.2 ANÁLISIS DE DATOS ESTADÍSTICA DESCRIPTIVA	221
A.3.3 FUNCIONES PARA EL CÁLCULO DE LA TENDENCIA CENTRAL.	222
A.3.4 FUNCIONES PARA EL CÁLCULO DE LA VARIACIÓN	223
A.3.5 FUNCIONES PARA EL CÁLCULO DE LA FORMA	225

**TITULO: DETERMINACIÓN DE POTENCIALES DE PRODUCCIÓN EN POZOS HORIZONTALES Y DESVIADOS PARA FLUJO EN DOS FASES USANDO UNA HERRAMIENTA SOFTWARE BASADA EN SIMULACIÓN MONTECARLO\***

**AUTORES: ANDRÉS FERNANDO OROZCO IBARRA, NELSON ARMANDO ROCHA SALDARRIAGA\*\***

**PALABRAS CLAVES:** Simulación Montecarlo, Estadística, Funciones de distribución, potencial de producción (IP), relación de rendimiento de injerto (IPR), Histograma, Números aleatorios.

**RESUMEN**

Existen muchas incertidumbres asociadas en la predicción del índice de productividad. Estas están presentes en las propiedades de la roca (Permeabilidad, espesor neto) y en las propiedades de los fluidos (Bo, Viscosidad).

Este proyecto pretendía predecir los potenciales de producción en pozos horizontales y desviados mediante simulación Montecarlo. En el desarrollo de la tesis fueron construidos dos modelos de estudio, uno para cada caso, dado que las ecuaciones para calcular el índice de productividad para pozos horizontales y desviados son diferentes. Para tal fin fueron usadas tanto hojas de cálculo como una herramienta software. Finalmente se hicieron comparaciones.

Para el desarrollo de esta metodología, se deben tener en cuenta algunos aspectos estadísticos importantes como son los números aleatorios y las funciones de distribución. La metodología principalmente se basa en los siguientes pasos: i) tener claro el modelo estudio, ii) escoger la o las funciones de distribución de densidad que mejor nos modelan las variables asociadas al modelo de estudio, iii) generar los números aleatorios de acuerdo a las funciones de distribución de densidad, ya sea con Excel, Visual Basic para aplicaciones o algún software comercial o de libre distribución, iv) simular el modelo estudio mediante la combinación aleatoria de las variables asociadas al modelo ya sea mediante Excel o algún software, v) obtener el histograma de la simulación, vi) determinar los escenarios mas probables o P10, P50 y P90 y los escenarios determinísticos..

---

\* Trabajo de Investigación

\*\* Facultad de Ingenierías Físico-Química. Escuela de Ingeniería de Petróleos. Fernando Calvete.

**TITLE: PRODUCTIVITY INDEX DETERMINATION IN HORIZONTAL AND SLANT WELL FOR TWO PHASE FLOW USING A SOFTWARE TOOL BASED ON MONTECARLO SIMULATION\***

**AUTHORS: ANDRÉS FERNANDO OROZCO IBARRA, NELSON ARMANDO ROCHA SALDARRIAGA\*\***

**KEYWORDS:** Montecarlo's Simulation, Statistics, Distribution functions, Production index (PI), Inflow performance relationship (IPR), Histogram, Random numbers.

**ABSTRACT**

There are many uncertainties associated in the determining the Productivity index. Those uncertainties include the rock properties (and permeability, net pay) and also fluid properties (Bo, viscosity).

This project objective was try to predict the productivity index in horizontal wells and slant wells using Montecarlo simulation. In the develop of the thesis two models were made, one for each case, it is because the equations to can calculate production index in horizontal and slant wells are different. Spreadsheets and a software tool were used. Finally the results were compared.

For the methodology, statistical aspects like random numbers and distribution functions were considered. Basically the methodology is based on the followings: i) Set clear objectives of the study model, ii) choose density distribution or distributions that best model the variables associated to the model of study, iii) generate random numbers agreeing with the density distribution functions; it can be done with Microsoft Excel, Visual Basic, other commercial software or free wave software, iv) simulate using a random combination of the variables associated to the model, it can be with Microsoft Excel or any software, v) obtain the histogram of the simulation, vi) determine the most probably scenarios or P10, P50 and P90 and the deterministic scenarios.

---

\* Research Work

\*\* Faculty of Physic-Chemical Engineering, Petroleum Engineering School. Fernando Calvete.

## INTRODUCCIÓN

Siendo consecuentes con el gran momento que está atravesando la industria de los hidrocarburos, resulta totalmente indispensable estar al tanto de todos aquellos métodos que de una u otra forma ayuden, fácil y efectivamente, a determinar todos esos parámetros de los pozos o de los yacimientos petrolíferos que indiquen las condiciones de rendimiento actual; uno de estos parámetros es el índice de producción. El índice de producción y, de la mano con este, la relación de rendimiento de flujo son el indicativo más dicente de la capacidad que tienen los pozos para entregar fluidos.

Existen muchas incertidumbres asociadas en la predicción del índice de productividad. Estas incertidumbres están presentes en las propiedades de la roca como el espesor neto y la permeabilidad absoluta y en las propiedades de los fluidos como la viscosidad y el factor volumétrico de formación del aceite. En este orden de ideas, el método de Montecarlo surge como una técnica que combina conceptos estadísticos (muestreo aleatorio) con la capacidad que tienen los ordenadores para generar números pseudo-aleatorios y automatizar cálculos. El método es aplicable a cualquier tipo de problema, ya sea estocástico o determinístico.

Este proyecto busca determinar el índice de productividad y con él la relación de rendimiento de flujo de los pozos mediante las combinaciones aleatorias de las variables involucradas en el cálculo del mismo. El presente trabajo está desarrollado en 5 capítulos: en el primer capítulo se presentan los conceptos generales relacionados con el índice de producción y con las relaciones de rendimiento de flujo, tanto para pozos verticales y desviados como para pozos horizontales. En el segundo capítulo se hace un barrido general de los conceptos

estadísticos involucrados a priori de estudiar el método de Montecarlo, haciendo especial énfasis en las funciones de densidad de probabilidad continua, que en últimas son las que describen las variables involucradas en un cálculo con simulación montecarlo. En el capítulo 3 se aborda todo lo relacionado con el método de Montecarlo, desarrollando desde su definición y su forma de aplicación hasta las recomendaciones de algunos autores al momento de asignar una función de densidad de probabilidad cuando se quiera representar una variable involucrada en el cálculo del parámetro deseado. Seguidamente, en el capítulo 4 se muestra detalladamente todas las corridas realizadas usando el método de Montecarlo en la predicción del índice de productividad tanto para el pozo colorado 38 como para el pozo horizontal, haciendo uso tanto de la hoja de cálculo como de la herramienta software. En el capítulo 5 se analizan los resultados obtenidos y de forma comparativa se estudian las posibles desviaciones. Por último, el capítulo 6 muestra las conclusiones de este trabajo de una manera corta y precisa.

## 1. ASPECTOS GENERALES DEL POTENCIAL DE PRODUCCIÓN DE LOS POZOS

En muchos casos es necesario tener un estimativo cuantitativo de cuál es la habilidad o capacidad que tienen los pozos de producir fluidos y, a las condiciones actuales del pozo, cual sería el caudal esperado. Desde este punto de vista se requiere definir el índice de productividad de los pozos y la relación de rendimiento de influjo.

### 1.1 EL ÍNDICE DE PRODUCTIVIDAD J

Se conoce como el abatimiento al cambio de presión entre la presión estática del yacimiento y la presión de fondo fluyendo, es decir,  $\text{abatimiento} = P_s - P_{wf}$ . Por otra parte, la relación que existe entre el gasto de producción de un pozo y el abatimiento, se denomina índice de productividad, el cual se simboliza con la letra J; si la producción,  $q$ , está en bbls/día de líquido a condiciones de almacenamiento y el abatimiento esta expresado en libras sobre pulgadas cuadradas, el IP (J) se define de la siguiente manera:

$$J = \frac{q}{(P_s - P_{wf})} \quad (1)$$

A menos que se especifique otra cosa el IP se basa en la producción neta de líquidos (Aceite y agua).

El IP específico  $J_s$  es el número total de barriles de líquido, a condiciones de almacenamiento, producidos por día por cada  $\frac{lb}{in^2}$  de abatimiento y por cada pie de espesor de la formación productora, es decir,

$$J_s = \frac{J}{h} = \frac{q}{h(P_s - P_{wf})} \quad (2)$$

La ecuación anterior se puede definir para el flujo radial de la producción de un líquido homogéneo de poca compresibilidad que está contenido en un yacimiento horizontal y uniforme, su ecuación es:

$$J = \frac{q}{(P_s - P_{wf})} = \frac{0,007082Kh}{B_o \mu \ln\left(\frac{r_e}{r_w}\right)} \quad (3)$$

Donde:

$q$  = Caudal de líquido en *bbbls/día*

$P_s$  = Es la presión estática del yacimiento en *PSI*

$P_{wf}$  = Es la presión de fondo fluyendo en *PSI*

$K$  = Permeabilidad en *md*

$h$  = Espesor *ft*

$B_o$  = Factor Volumétrico *Rb/STB*

$\mu$  = Viscosidad en *cp*

$r_e$  = radio de drenaje *ft*

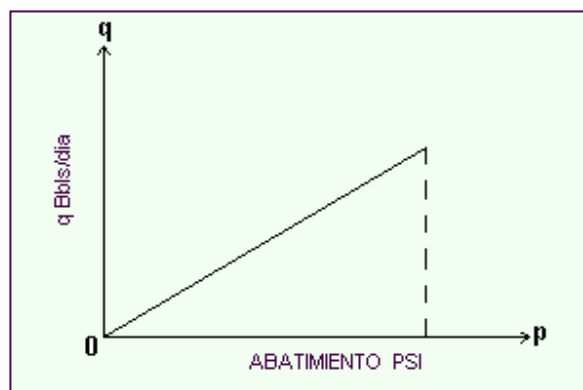
$r_w$  = radio del pozo *ft*

De igual manera se puede expresar el índice de productividad específico para un flujo radial.

$$J_s = \frac{J}{h} = \frac{0,007082k}{B_o \mu \ln\left(\frac{r_e}{r_w}\right)} \quad (4)$$

Si el valor del IP del pozo se toma como constante, independientemente de la producción actual del pozo, se puede decir que  $q = J * \Delta P$ , en la que  $\Delta P$  es el abatimiento. Así es evidente que la relación entre  $q$  y  $\Delta P$  es una línea recta que pasa por el origen y tiene una pendiente  $J$ . Figura 1.

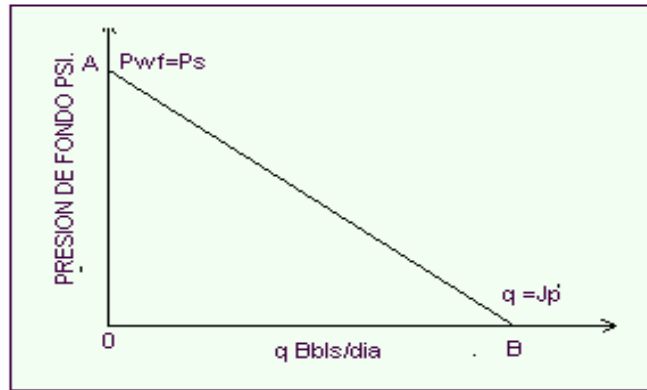
**FIGURA 1. GRAFICA DEL ABATIMIENTO**



**FUENTE:** T.E.W Nind, Fundamentos de Producción y Mantenimiento de Pozos Petroleros, Editorial Limuza México 1987.

En un momento particular de la vida del yacimiento,  $P_s$  tiene un valor específico, por lo que si  $J$  es constante, al graficar  $P_{wf}$  contra  $q$  se obtiene una línea recta, como se observa en la figura 2.

**FIGURA 2. ESQUEMA DEL COMPORTAMIENTO DE LA PRESION DE FONDO FLUYENDO**



**FUENTE:** T.E.W Nind, Fundamentos de Producción y Mantenimiento de Pozos Petroleros, Editorial Limuza México 1987.

El valor de  $q$  en el punto B, es decir,  $J_{ps}$ , se conoce como el potencial del pozo y se representa con el símbolo  $q'$ . Se debe hacer énfasis que lo anterior se refiere el gasto máximo al cual la formación puede entregar líquido hacia el pozo, lo que se presenta cuando la  $P_{wf}$  es cero.

### 1.1.1 Ecuaciones Involucradas en el cálculo del IP en este Trabajo

➤ **Para Pozos Verticales.** La ecuación implicada en el cálculo del índice de productividad para este trabajo, es la misma ecuación 3 con la modificación referente a la distancia recorrida por el trasiente de presión, y al mismo tiempo, a la forma que posea el yacimiento. De esta manera:

$$IP = \frac{kh}{141.2 * \mu_o B_o * \left[ \ln \left( \frac{R_e}{r_w} \right) - \frac{3}{4} \right]} \quad (5)$$

Donde:

- $h$  es el espesor neto promedio de petróleo del yacimiento;
- $\mu$  es la viscosidad del petróleo;
- $B_o$  es el factor volumétrico de formación del petróleo;
- $r_w$  es el radio del pozo;
- $R_e$  es el radio de drenaje del pozo;
- $k$  es la permeabilidad efectiva del yacimiento.

➤ **Para Pozos Horizontales.** En un pozo horizontal, se deben considerar muchos aspectos para calcular la productividad. JOSHI (1988) presentó una metodología, considerando muchos aspectos involucrados en el problema, como la excentricidad del pozo y la influencia de la anisotropía en la permeabilidad. En esta tesis, se desprecian tanto la excentricidad del pozo como el factor de daño.

El índice de productividad para pozos horizontales se calcula por la siguiente ecuación:

$$IP = \frac{2\pi k_h h / (\mu B_o)}{\ln \left[ \frac{a + \sqrt{a^2 - (L/2)^2}}{(L/2)} \right] + \frac{\beta h}{L} \ln \left( \frac{\beta h}{2r_w} \right)} \quad (6)$$

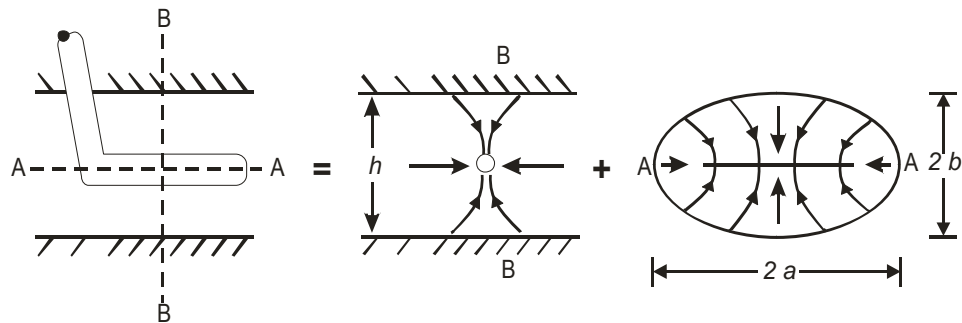
Donde:

- $h$  es el espesor neto promedio de petróleo del yacimiento;
- $\mu$  es la viscosidad del petróleo;
- $B_o$  es el factor volumétrico de formación del petróleo;
- $L$  es la longitud horizontal del pozo;
- $r_w$  es el radio del pozo;

- $\beta = \left( \frac{k_h}{k_v} \right)^{0,5}$
- $k_h$  y  $k_v$  son las permeabilidades efectivas horizontales y verticales del yacimiento;
- $a$  es la mitad del eje mayor de la elipse de drenaje en un plano horizontal donde el pozo esta localizado (Figura 3) y se define como :

$$a = \frac{L}{2} \left[ \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{(L/2r_{eH})^4}} \right]^{\frac{1}{2}} \quad (7)$$

**FIGURA 3. División del problema 3D de pozos horizontales, en dos problemas 2D**



**Fuente:** Determinación probabilística del índice de productividad para un pozo horizontal en un yacimiento de aceite pesado, SPE 107669

donde  $r_{eH}$  es el radio de drenaje del pozo horizontal. Se implemento una corrección para considerar los efectos de la anisotropía en el radio del pozo (JOSHI, 1988). Con el objetivo de determinar el radio de drenaje del pozo horizontal, para esto se usó un método de solución de ecuaciones simultáneas para satisfacer la ecuación  $ar_{eH} = r_e^2$ , donde  $r_e$  es el radio de drenaje del pozo, y es dado por el usuario.

## 1.2 RELACIÓN DE RENDIMIENTO DE INFLUJO

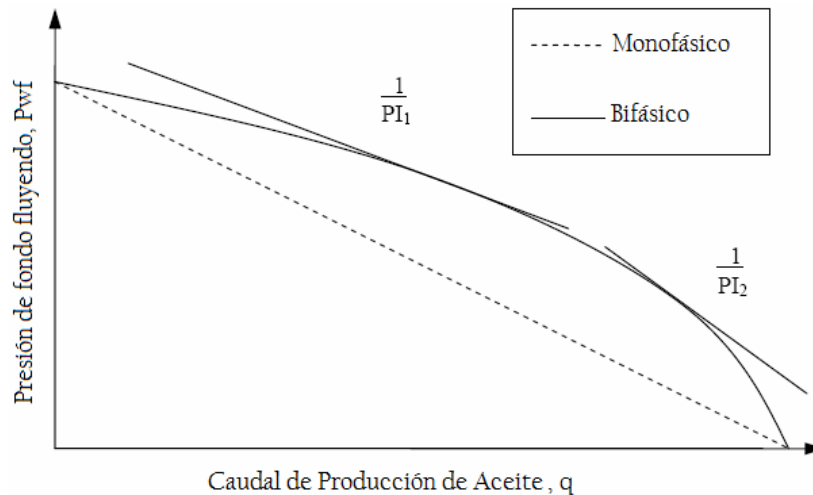
Una relación de rendimiento de influjo es la relación entre la presión de fondo fluyendo y el caudal de producción. Generalmente, esta relación se presenta en forma gráfica dibujando la presión de fondo fluyendo en el eje X y el caudal de producción en el eje Y. Esta curva es llamada curva de relación de rendimiento de influjo o curva IPR.

Las curvas IPR son usadas para analizar la Entregabilidad de los pozos estimando la rata de producción para una presión de fondo fluyendo dada y viceversa. Las curvas IPR son aquellas que reflejan la habilidad del yacimiento de entregar fluidos a la cara del pozo. La combinación de una curva IPR y una curva que refleje el rendimiento de la tubería de producción identifica un punto de operación del pozo. Por lo tanto, las curvas IPR son una información importante para analizar la entregabilidad de los pozos.

Para un flujo monofásico, los caudales de producción en un pozo son directamente proporcionales a la diferencia entre la presión de fondo fluyendo y la presión del yacimiento, también llamado descenso.

Las curvas IPR para flujo monofásico dan como resultado líneas rectas con pendientes de uno sobre el *Índice de productividad (1/IP)*. Sin embargo, Muskat sugirió que para flujo bifásico de líquido o gas, esta relación lineal no existe. Estas curvas IPR para flujo bifásico pueden tener índices de productividad con más de un valor, como se muestra en la Figura 4.

**Figura 4. Curva de Relación de Rendimiento de Influjo**



**FUENTE:** Análisis de rendimiento de influjo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

El flujo bifásico de aceite y gas comúnmente ocurre en producción, especialmente en yacimientos de empuje de gas en solución. La producción causa la reducción de la presión del yacimiento debajo de la presión de punto burbuja. De esta forma, el gas se libera del aceite saturado. Como consecuencia de lo anterior el flujo en el yacimiento debajo de la presión de punto de burbuja, consiste de un flujo bifásico de aceite y gas.

Desde 1980, los pozos horizontales han sido perforados incrementalmente. Desde comienzo de los 90, los pozos horizontales han sido populares en el desarrollo de campos de aceite y gas.

Sus aplicaciones incluyen yacimientos delgados o gruesos con alta permeabilidad vertical. Los pozos horizontales han sido usados para controlar problemas severos de conificación agua o gas, y también pueden acelerar el proceso de recobro. Además, los pozos horizontales son parte esencial de pozos multilaterales o pozos complejos. Recientemente, los pozos complejos han sido perforados

extensivamente para mejorar la productividad de los pozos, especialmente para producción de aceite pesado.

Una ecuación empírica para calcular el IPR de pozos verticales en flujo bifásico, fue presentada por *Vogel* en 1968. Del éxito de la ecuación de *Vogel*, muchos estudios han sido publicados bajo el tema de relación de rendimiento de influjo (IPR) de pozos horizontales en yacimientos con empuje de gas en solución, y alguno de esos estudios presentaron ecuaciones analíticas siguiendo el trabajo dejado por *Vogel*. *Bendakhlia* Y *Aziz* presentaron una ecuación analítica de relación de rendimiento de influjo (IPR) en 1989 para pozos cuyos yacimientos poseen empuje de gas en solución. *Cheng* introdujo una ecuación analítica en 1990 para flujo bifásico en pozos desviados y horizontales. *Retnanto* Y *Economides* desarrollaron otra ecuación de relación de rendimiento de influjo (IPR) de pozos multiramados y horizontales en yacimientos de empuje de gas en solución. Aunque estas ecuaciones son diferentes, todas ellas están basadas en la ecuación de *Vogel*. Las correlaciones en estas ecuaciones cambian para diferentes condiciones tales como factores de Recobro, presión de punto de burbuja, y presión de yacimiento.

### 1.2.1 Ecuaciones analíticas del flujo radial en medio poroso

La geometría de flujo de un fluido desde un medio poroso homogéneo dentro de un pozo vertical, es de tipo radial. El comportamiento de flujo en un medio poroso puede ser explicado por la ley de Darcy. La ecuación de la ley de Darcy en coordenadas radiales es:

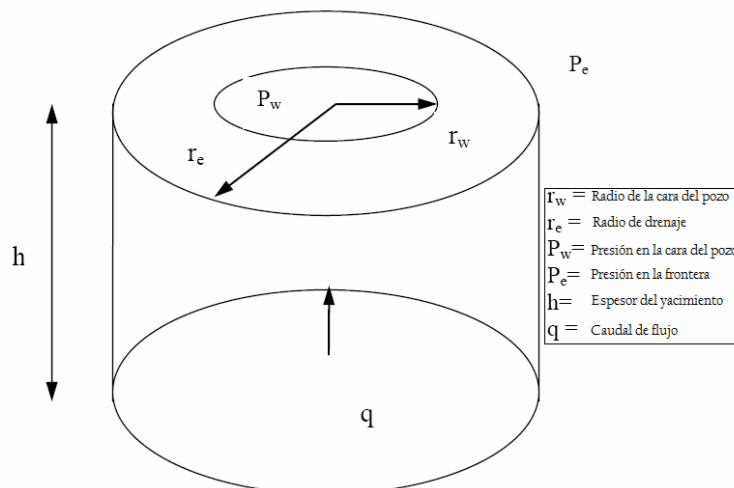
$$q = \frac{kA}{\mu} \frac{dP}{dr} \quad (8)$$

Donde,

$q$  = Caudal de flujo  
 $K$  = Permeabilidad  
 $A$  = Área de drenaje  
 $P$  = Presión  
 $r$  = Distancia en la dirección de Flujo  
 $\mu$  = Viscosidad del Fluido

Para resolver los perfiles de presión, son requeridas algunas asunciones. Es asumido que la formación es homogénea e isotrópica. El fluido tiene una viscosidad constante, y una compresibilidad pequeña y constante. El flujo es isotérmico y laminar. Los efectos gravitacionales son insignificantes. El espesor del yacimiento es constante con capas impermeables encima y debajo de la zona permeable. El flujo en el yacimiento es monofásico. La figura 2 muestra el plano de un flujo geométrico radial.

**Figura 5. Plano de un flujo geométrico radial**



**FUENTE:** Análisis de rendimiento de influjo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

Aplicando las asunciones de encima y usando la ley de Darcy para un sistema en coordenadas radiales, la ecuación de difusividad para flujo radial puede ser resuelta para la distribución de presión. La ecuación de difusividad es:

$$\frac{\partial^2 P}{\partial r^2} + \frac{1}{r} \frac{\partial P}{\partial r} = \frac{\phi \mu c_t}{k} \frac{\partial P}{\partial t} \quad (9)$$

Para resolver la ecuación de difusividad, son necesitadas unas condiciones iniciales y de frontera. La condición inicial especificada es que la presión del yacimiento es uniforme y constante en su totalidad. La solución que domina esta condición es:

$$\frac{\partial p}{\partial r} = \frac{q\mu}{2\pi kh} \frac{1}{r_w} \quad \text{en } r=r_w \quad (10)$$

Usualmente, la condición de límite interna para un sistema ideal es:

$$\lim_{r \rightarrow 0} \left( r \frac{\partial p}{\partial r} \right) = \frac{q\mu}{2\pi kh} \quad (11)$$

Estas ecuaciones se aplican a pozo con radio finito en la cara del pozo.

Tres tipos de condiciones externas de fronteras son: yacimiento actuando como infinito, frontera de no flujo, y presión constante en la frontera. La primera condición de frontera (yacimiento actuando como infinito) sucede al comienzo de la producción del yacimiento. El perfil de presión no ha alcanzado aún la frontera, por lo tanto este es una función del tiempo. La condición de yacimiento infinito es:

$$\lim_{r \rightarrow \infty} P(r,t) = P_i \quad (12)$$

La segunda condición es la frontera de no flujo. En esta condición, no hay flujo a través del límite, y esto es expresado como:

$$\frac{\partial p}{\partial r} = 0 \quad \text{en} \quad r = r_e \quad (13)$$

Finalmente, la condición de frontera de yacimiento constante asume que la presión en el límite externo es constante. Esta condición está dada por:

$$P(r, t) = P_e \quad \text{en} \quad r = r_e \quad (14)$$

Después de aplicar las condiciones inicial y límite, la solución analítica para cada condición es obtenida de resolver la ecuación de difusividad. La solución analítica para la condición de yacimiento actuando como infinito o flujo trasiente, condición de yacimiento limitado o flujo en estado pseudoestable, y condición de yacimiento constante o flujo en estado estable, son presentadas por las ecuaciones 15 a 17 en unidades de campo respectivamente,

Flujo trasiente,

$$P_{wf} = P_i - \frac{162,6q\mu B}{kh} \left[ \log(t) + \log\left(\frac{k}{\phi\mu C_i r_w^2}\right) - 3,23 \right] \quad (15)$$

Flujo en estado pseudoestable

$$\bar{P} = P_{wf} + \frac{q\mu B}{0,00708kh} \left( \ln\left(\frac{r_e}{r_w}\right) - \frac{3}{4} \right) \quad (16)$$

, Y flujo en estado estable

$$P_e = P_{w_f} + \frac{q\mu B}{0,00708kh} \left( \ln \left( \frac{r_e}{r_w} \right) \right) \quad (17)$$

Donde,

$P_i$  = Presión inicial del yacimiento (psia)

$\bar{P}$  = Presión promedio del yacimiento (psia)

$P_{w_f}$  = Presión de fondo fluyendo

$q$  = Caudal de flujo (STB/día)

$k$  = Permeabilidad (mildarcy)

$h$  = Espesor de la formación (ft)

$B$  = Factor volumétrico de formación (fracción)

$\mu$  = Viscosidad del fluido (cp)

$r_e$  = Radio de drenaje (ft)

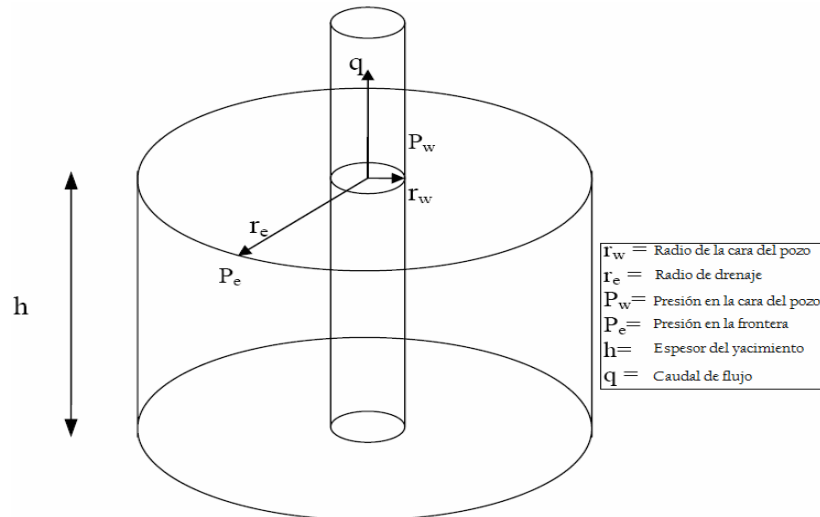
$r_w$  = Radio de la cara del pozo

La ecuación analítica para flujo monofásico en pozos verticales es resuelta aplicando la ley de Darcy junto con algunas asunciones para obtener la solución de la ecuación de difusividad. Así las condiciones inicial y de frontera son requeridas para resolver la ecuación de difusividad. De acuerdo a este procedimiento, la ecuación analítica para flujo bifásico es difícil de obtener cuando las permeabilidades relativas y las transmisibilidades de fase son tenidas en cuenta. Por lo tanto, no hay solución analítica para flujo bifásico. Sin embargo, Vogel presentó una ecuación empírica en 1968. Su ecuación estuvo basada en resultados de simulación de yacimientos.

### 1.2.2 Relación de rendimiento de influjo (IPR) para flujo bifásico en pozos verticales.

En 1968, Vogel utilizó un programa de computador generado por Weller en 1966 para yacimientos con empuje de gas en solución para predecir las curvas IPR. Su modelo usó un yacimiento circular, con un pozo vertical completado en el centro de la formación. El yacimiento limitado es penetrado a través del espesor de la formación. La formación es uniforme e isotrópica con saturación de agua constante. Los efectos de gravedad y compresibilidad de la roca son despreciables. La composición y el equilibrio son constantes para aceite y agua. Las presiones son diferentes en la fase aceite y en la fase gas. El estado semiestable asume que la rata de desaturación del tanque de aceite es la misma en todos los puntos a un instante dado. El modelo de yacimiento de Vogel es mostrado en la figura 6.

Figura 6. Modelo de Yacimiento de Vogel



FUENTE: Análisis de rendimiento de influjo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

Vogel usó la aproximación de Weller para pozos con empuje de gas en solución para cubrir un amplio rango de condiciones. Estas condiciones incluyeron diferentes características del aceite crudo, los efectos de espaciado de pozo, fracturamiento, y restricciones de daño. Él obtuvo una ecuación empírica de IPR para flujo bifásico en pozos verticales por la generación de las curvas adimensionales de IPR. La ecuación empírica es dada por:

$$\frac{q_0}{q_{0\max}} = 1 - 0,20 \left( \frac{P_{wf}}{P_R} \right) - 0,80 \left( \frac{P_{wf}}{P_R} \right)^2 \quad (18)$$

La ecuación de Vogel ha sido usada extensiva y exitosamente para analizar la relación de rendimiento de influjo de pozos de aceite fluyendo en yacimientos de empuje de gas en solución.

### **1.2.3 Relación de Rendimiento de Inlujo de flujo bifásico en pozos desviados y horizontales.**

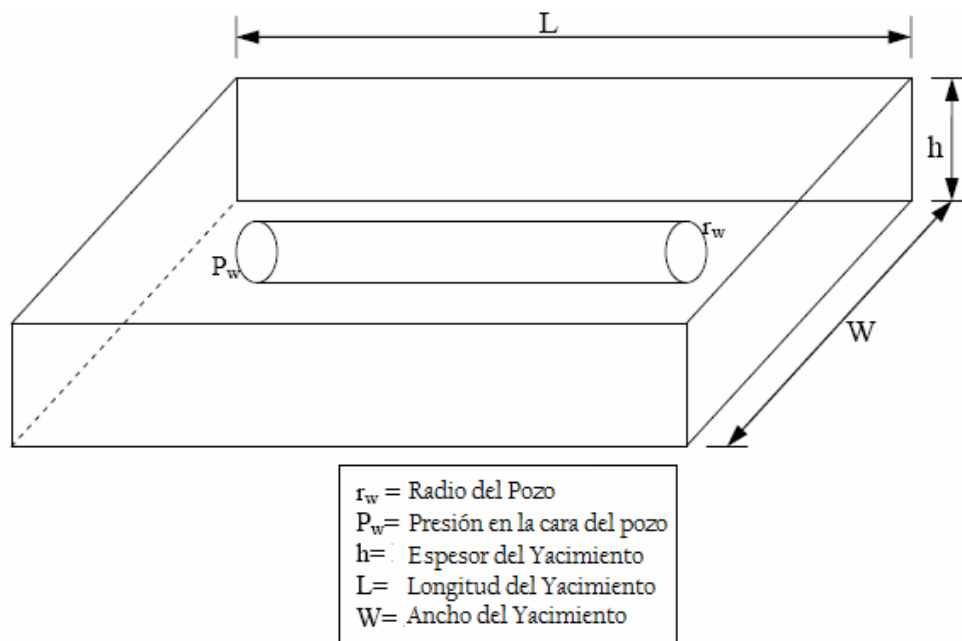
Debido a que la ecuación empírica de Vogel ha sido usada exitosamente en la industria, muchos estudios presentaron ecuaciones similares para pozos horizontales basados en el trabajo de Vogel. Estas incluyen el modelo de Bendakhlia & Aziz, Modelo de Cheng, y modelo de Retnanto & Economides.

#### **1.2.3.1 Modelo de Bendakhlia & Aziz**

En 1989, Bendakhlia & Aziz desarrollaron una ecuación de IPR para pozos horizontales. Sus datos son generados de una simulación de yacimientos. Su modelo físico es un yacimiento en forma de caja con un área de drenaje cuadrada. La permeabilidad de la formación es homogénea e isotrópica, y la compresibilidad

de la formación es constante. Un pozo horizontal es localizado en la mitad del yacimiento, está centrado a media altura y completado a lo largo de la longitud total del yacimiento, como se muestra en la figura 7

**Figura 7. Modelo de Bendakhlia y Aziz y modelo de Cheng**



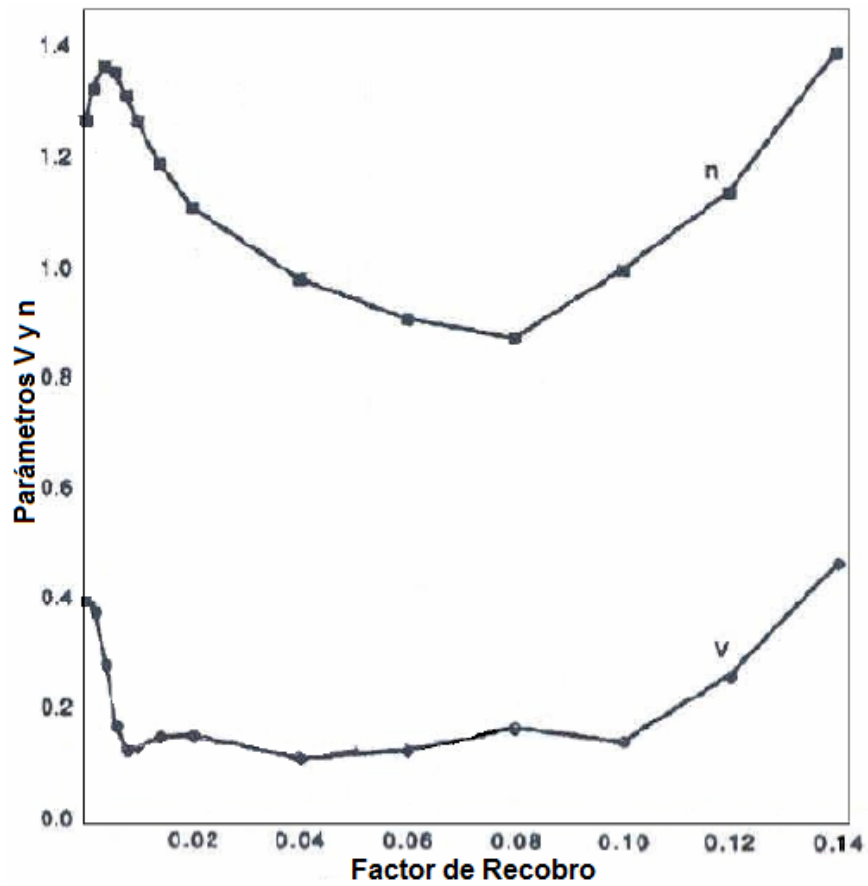
**FUENTE:** Análisis de rendimiento de influjo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

Las curvas IPR fueron generadas sobre un amplio rango de propiedades de fluidos, condiciones de yacimiento y características de pozo. Las curvas IPR resultantes fueron hechas adimensionalmente para comparar sus curvaturas y la tasa de cambio de la producción de aceite con la presión de fondo fluyendo. De los resultados, una combinación de la ecuación de Vogel y la ecuación de Fetkovich fue sugerida por la variación de la ecuación de correlación con el porcentaje de recobro. La ecuación de Bendakhlia y Aziz para calcular la curva IPR para pozos horizontales es:

$$\frac{q_0}{q_{0\max}} = \left[ 1 - V \left( \frac{P_{wf}}{\bar{P}_R} \right) - (1 - V) \left( \frac{P_{wf}}{\bar{P}_R} \right)^2 \right]^n \quad (19)$$

Donde, V y n ecuación de encima son parámetros que varían con los factores de recobro, como se muestra en la figura 8

**Figura 8. Parámetros V y n para la ecuación de Bendakhlia y Aziz.**



**FUENTE:** Análisis de rendimiento de flujo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

### 1.2.3.2 Modelo de Cheng

En 1990 Cheng generó una inecuación para calcular la curva de IPR para pozos desviados y horizontales. Él usó el simulador de NIPER de pozos horizontales y desviados para generar curvas IPR y entonces desarrolló una ecuación basada en los resultados de simulación. Su modelo es un yacimiento rectangular limitado con un pozo en la mitad del yacimiento. El yacimiento es homogéneo e isotrópico con una saturación de agua constante. La saturación de agua es inmóvil durante la producción. Por lo tanto, en el yacimiento solo hay gas y petróleo. Las fuerzas de presión capilar de los fluidos del yacimiento son despreciables. El pozo horizontal esta produciendo bajo condiciones de estado semiestable. El modelo de yacimiento es el mismo que se muestra en la figura 7.

La ecuación de Cheng genera curvas de IPR para pozos horizontales variando el ángulo de un pozo inclinado desde un pozo vertical hasta un pozo horizontal. La correlación que presentó es:

$$\frac{q_o}{q_{o\max}} = a_0 - a_1 \left( \frac{P_{wf}}{P_R} \right) - a_2 \left( \frac{P_{wf}}{P_R} \right)^2 \quad (20)$$

La tabla 1 muestra las constantes que son usadas en la ecuación de Cheng. Nótese que no hay diferencia en la ecuación de Cheng tanto para pozos horizontales como para pozos verticales, excepto el ángulo de desviación del pozo. Las constantes en la ecuación 20 solo varían con el ángulo de desviación del pozo.

**Tabla 1. Constantes usadas en la correlación de Cheng.**

Angulo de desviación (Grados)	Parámetros		
	a <sub>0</sub>	a <sub>1</sub>	a <sub>2</sub>
0(Vertical)	1.0000	0.2000	0.8000
15	0.9998	0.2210	0.7783
30	0.9969	0.1254	0.8582
45	0.9946	0.0221	0.9663
60	0.9926	-0.0549	1.0395
75	0.9915	-0.1002	1.0829
85	0.9915	-0.1120	1.0942
90(Horizontal)	0.9885	-0.2055	1.1818

**FUENTE:** Análisis de rendimiento de influjo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

### 1.2.3.3 Modelo de Retnanto & Economides

Retnanto & Economides presentaron una ecuación para calcular el IPR para flujo en 2 fases en pozos horizontales en 1998. El simulador numérico, VIP, fue usado para generar curvas de IPR de pozos horizontales bajo diferentes condiciones, y la correlación para el modelo fue generada basándose en los resultados de la simulación numérica.

La ecuación analítica del modelo de Retnanto & Economides es:

$$\frac{q_o}{q_{o\max}} = 1 - 0.25 \left( \frac{P_{wf}}{P_R} \right) - 0.75 \left( \frac{P_{wf}}{P_R} \right)^n \quad (21)$$

Donde:

$$n = \left( -0.27 + 1.46 \left( \frac{P_{wf}}{P_b} \right) - 0.96 \left( \frac{P_{wf}}{P_b} \right)^2 \right) \left( 4 + 1.66 \times 10^{-3} P_b \right) \quad (22)$$

La ecuación 22 muestra que n es un parámetro que depende de la presión de punto de burbuja del fluido del yacimiento.

#### 1.2.4 Ecuación Modificada de Vogel para pozos horizontales

En 1992, Kabir mostró que una vez el potencial absoluto de entrega de flujo (AOFP) es evaluado tanto por las correlaciones de Vogel y la de Fetkovich para pozos verticales, puede ser usado para describir el IPR. El AOFP puede ser computado usando cualquiera de las expresiones para el índice de productividad desarrolladas para pozos horizontales para varias condiciones límites. La expresión para el potencial absoluto de entrega de flujo (AOFP) en la ecuación de Vogel es:

$$q_{o,\max} = \frac{JP_b}{1.8} \quad (23)$$

, y para la ecuación de Fetkovich es:

$$q_{o,\max} = \frac{JP_b}{2} \quad (24)$$

donde,

J= Índice de Productividad

$P_b$  = Presión de punto burbuja

Para un pozo vertical en un área de drenaje circular,  $J$  esta dada en unidades de campo por:

$$J = \frac{k_o h}{141.2 \mu_o B_o \left( \ln \left( 0.472 \frac{r_e}{r_w} \right) + s \right)} \quad (25)$$

donde,

- $k_o$  = Permeabilidad para la fase aceite, md
- $h$  = Espesor de la formacion, pies
- $\mu_o$  = Viscosidad del aceite, cp
- $B_o$  = Factor volumétrico del aceite, fraccion
- $r_e$  = Radio externo del yacimiento, pies
- $r_w$  = Radio del pozo, pies
- $S$  = Facor de daño

Para límites de drenaje no circulares,  $r_e/r_w$  debería ser remplazado por un factor de forma que depende a su vez de la configuración del pozo y del yacimiento.

En 1998, Helmy & Wattenbarger presentó una ecuación analítica para calcular el índice de productividad para pozos horizontales produciendo a presión constante o a caudal constante. El mostró que la expresión para el índice de productividad (PI) para pozos horizontales con presión constante es:

$$J_{CP} = \frac{q}{P_r - P_{wf}} = \frac{k_o b_{eq}}{141.2 \mu_o B_o \left( \frac{1}{2} \ln \left( \frac{4A_{eq}}{r^2_{weq}} - \frac{1}{2} \ln C_{ACP} + s_{PCP} \right) \right)} \quad (26)$$

donde,

$$\begin{aligned} \ln C_{ACP} = & 2.607 - \left( 4.74 - 10.353 \left( \frac{x_{weq}}{a_{eq}} \right)^{1.115} + 9.165 \left( \frac{x_{weq}}{a_{eq}} \right)^{2.838} \right) \left( \frac{a_{eq}}{h_{eq}} \right)^{1.01} \\ & + 1.810 \ln \left( \operatorname{sen} \left( \frac{\pi z_{eq}}{h_{eq}} \right) \right) + 2.056 \left( \frac{a_{eq}}{h_{eq}} \right) \end{aligned} \quad (27)$$

$$s_{PCP} = \left( \left( \frac{b_{eq}}{L_{eq}} \right)^{1.233} - 1 \right) \times \left[ 2.897 + 0.003 \ln C_{ACP} - 0.453 \ln \left( \frac{h_{eq}}{a_{eq}} \right) + AA \right] \quad (28)$$

$$AA = \frac{\left[ 0.388 - 1.278 \left( \frac{y_{weq}}{b_{eq}} \right) + 0.715 \left( \frac{y_{weq}}{b_{eq}} \right)^2 + 1.278 \left( \frac{L_{eq}}{b_{eq}} \right) - 1.215 \left( \frac{L_{eq}}{b_{eq}} \right)^2 \right]}{\left( \frac{h_{eq}}{a_{eq}} \right) \left( \frac{a_{eq}}{b_{eq}} \right)^{1.711}} \quad (29)$$

$$k_{eq} = \sqrt[3]{k_x k_y k_z} \quad (30)$$

$$a_{eq} = \sqrt{\frac{k_{eq}}{k_y}} a \quad (31)$$

$$b_{eq} = \sqrt{\frac{k_{eq}}{k_x}} b \quad (32)$$

$$h_{eq} = \sqrt{\frac{k_{eq}}{k_z}} h \quad (33)$$

$$L_{eq} = \sqrt{\frac{k_{eq}}{k_y}} L \quad (34)$$

$$A_{eq} = a_{eq} h_{eq} \quad (35)$$

$$r_{weq} = \frac{1}{2} r_w \left( \sqrt[4]{\frac{k_x}{k_z}} + \sqrt[4]{\frac{k_z}{k_x}} \right) \quad (36)$$

$$x_{weq} = \sqrt{\frac{k_{eq}}{k_x}} x_w \quad (37)$$

$$y_{weq} = \sqrt{\frac{k_{eq}}{k_y}} y_w \quad (38)$$

$$z_{weq} = \sqrt{\frac{k_{eq}}{k_z}} z_w \quad (39)$$

, y para una condición de tasa constante es:

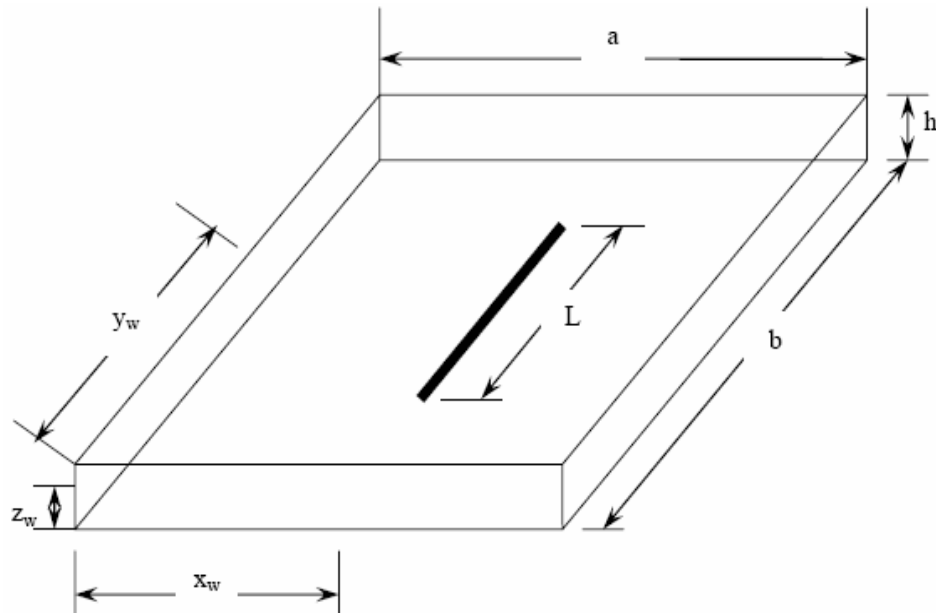
$$J = \frac{q}{P_r - P_{wf}} = \frac{k_{eq} b_{eq}}{141.2 \mu_o B_o \left( \frac{1}{2} \ln \left( \frac{4A_{eq}}{r_{weq}^2} - \frac{1}{2} \ln C_A + s_p \right) \right)} \quad (40)$$

$$\ln C_A = 4.485 - \left( 4.187 - 12.56 \left( \frac{x_{weq}}{a_{eq}} \right) + 12.56 \left( \frac{x_{weq}}{a_{eq}} \right)^2 \right) \left( \frac{a_{eq}}{h_{eq}} \right) + 2 \ln \left( \operatorname{sen} \left( \frac{\pi z_{eq}}{h_{eq}} \right) \right) + 2.056 \left( \frac{a_{weq}}{h_{eq}} \right) \quad (41)$$

$$s_p = \left( \left( \frac{b_{eq}}{L_{eq}} \right)^{0.858} - 1 \right) \times \left[ 0.025 + 0.022 \ln C_A - 3.781 \ln \left( \frac{h_{eq}}{a_{eq}} \right) + BB \right] \quad (42)$$

$$BB = \frac{\left[ 1.289 - 4.751 \left( \frac{y_{weq}}{b_{eq}} \right) + 4.652 \left( \frac{y_{weq}}{b_{eq}} \right)^2 + 1.654 \left( \frac{L_{eq}}{b_{eq}} \right) - 1.718 \left( \frac{L_{eq}}{b_{eq}} \right)^2 \right]}{\left( \frac{h_{eq}}{a_{eq}} \right) \left( \frac{a_{eq}}{b_{eq}} \right)^{1.472}} \quad (43)$$

**Figura 9. Modelo de yacimiento de la ecuación de Helmy & Wattenbarger**



**FUENTE:** Análisis de rendimiento de flujo bifásico en pozos horizontales, Rungtip Kamkom, M.S, Universidad de Texas 2004

## **2. CONCEPTOS ESTADÍSTICOS GENERALES**

### **2.1 ESTADÍSTICA DESCRIPTIVA**

Muchas veces se dispone de una cantidad muy grande de datos que para poder analizarlos se hace necesario organizarlos. Tal análisis puede consistir donde se agrupan la mayor cantidad de los datos o como varían para poder detectar cualquier tendencia. La organización de los datos puede hacerse mediante gráficas, diagramas y tablas que sean de fácil entendimiento y manejo. A continuación veremos las técnicas para construir estas formas de presentación.

#### **2.1.1 Distribuciones de frecuencia**

Las distribuciones de frecuencia son una herramienta importante al momento de realizar un análisis detallado de los datos sabiendo que mediante estas, podemos definir con más claridad el comportamiento que están siguiendo los datos y operarlos posteriormente.

##### **2.1.1.1 Definición**

Es un método de clasificación y agrupamiento de datos estadísticos en clases o intervalos, de tal manera que se pueda establecer el número o porcentaje de cada clase y así manipularlos en grandes cantidades. El número o porcentaje de cada clase se le conoce como *frecuencia de clase*.

### 2.1.1.2 Elaboración

Los pasos para la construcción de una distribución de frecuencias son:

✓ PASO 1: *Determinar el número de clase*: Consiste en identificar los intervalos en los cuales vamos a dividir el conjunto de datos. Demasiadas clases o muy pocas pueden no revelar la forma básica del conjunto de datos. Siempre es recomendable que el límite inferior de la primera clase sea un poco menor que el valor más pequeño del conjunto de datos. Además, que el límite superior de la última clase sea un poco mayor que el valor más grande del conjunto de datos. En general, para grandes cantidades de datos se requiere más clases que para pequeñas cantidades.

La fórmula que a continuación se nombra, permite calcular el número de clases.

$$\text{N}^\circ \text{ de Intervalos} = k = \begin{cases} \sqrt{n} & \text{si } n \text{ no es muy grande} \\ 1 + 3.22 \text{Log} & \text{en otro caso} \end{cases} \quad (44)$$

✓ PASO 2: *Determinar la amplitud de clase o intervalo de clase*: Se debe tomar el intervalo igual para todas las clases, sin embargo hay ocasiones en que se pueden tomar amplitudes de clase desiguales para evitar clases vacías o casi vacías. Del número de datos se busca el valor más alto (A) y el valor más bajo (B) y el intervalo se halla por medio de la fórmula

$$I = \frac{A - B}{K} \quad (45)$$

Donde K es el número de clases.

✓ PASO 3: *Creación de las clases*: El límite superior de cada clase corresponde al intervalo siguiente. Para ser más estrictos, en términos matemáticos, cada clase sería un intervalo semiabierto a la derecha [ ).

✓ PASO 4: *Contar el número de elementos de cada clase*: Hay que recordar que el número de observaciones de cada clase se llama frecuencia de clase. Las frecuencias asociadas a cada clase se simbolizan  $f_i$  y las frecuencias relativas  $h_i$ .

- *Punto medio de clase*: También conocida como marca de clase, se ubica a la mitad entre los límites inferiores de dos clases consecutivas. Se calcula sumando el límite inferior al límite superior de una clase y dividiendo esta suma por 2. La marca de clase se simboliza  $x_i$  donde  $i$  es el número de la clase correspondiente.

### 2.1.1.3 Propiedades de las distribuciones de frecuencia.

Las distribuciones de frecuencia cumplen las siguientes propiedades:

- Las frecuencias de clase son enteros no negativos, es decir,  $f_i \geq 0$ .
- Las frecuencias relativas son números no negativos menores que 1, o sea

$$0 \leq h_i \leq 1.$$

- La suma de todas las frecuencias relativas es igual a 1, esto es,  $\sum_{i=1}^n h_i = 1$  donde  $n$  es el número de datos.

- La suma de todas las frecuencias es igual a la totalidad de los datos, es decir,

$$\sum_{i=1}^n f_i = n$$

### 2.1.1.4 Representaciones gráficas de una distribución de frecuencias.

Generalmente resulta muy conveniente resaltar información contenida en una tabla de una distribución de frecuencias. Para tal fin se utilizan diagramas o gráficas.

Existen diagramas tales como el histograma, el polígono de frecuencias, el polígono de frecuencias acumuladas, entre otros, que representan de una manera adecuada la distribución de frecuencias.

Al construir una gráfica en el plano cartesiano se acostumbra a ubicar a las clases en el eje de las abcisas (eje x) y las frecuencias en el eje de las ordenadas (eje y).

### Histogramas

Es uno de los diagramas que se utiliza con mayor frecuencia. Es una gráfica hecha con rectángulos donde la base y la altura de cada uno son la clase y la frecuencia correspondiente, respectivamente. Cada rectángulo se coloca adyacente al inmediatamente anterior al él como se indica en la figura 10.

**Figura 10. Histograma**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

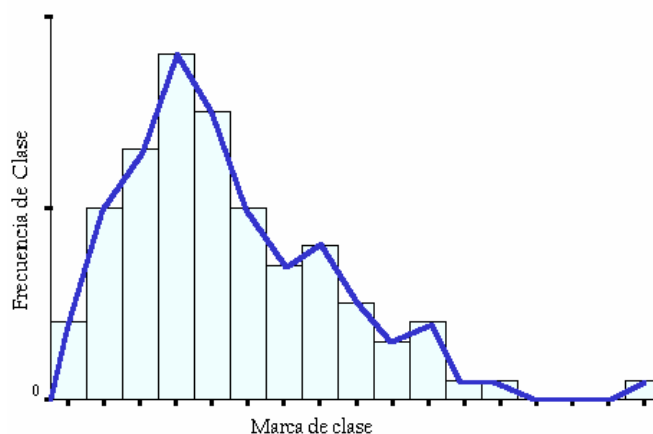
## Polígono de frecuencias

Es similar al histograma, está formado por segmentos de rectas que unen puntos de coordenadas  $(x_i, f_i)$  donde  $x_i$  es el punto medio de cada clase y  $f_i$  es la frecuencia de clase  $i$ . Se añaden dos clases, una antes de la primera clase y la otra después de la última clase. En estas dos clases hay cero observaciones, por tanto sus frecuencias son cero.

Se encuentran los puntos medios de estas clases y se ancla el polígono a cero, de tal manera que este se empieza a construir desde el eje X.

Un polígono de frecuencias que utiliza frecuencias relativas de puntos de dato en cada una de las clases, en lugar del número real de puntos, se conoce como *polígono de frecuencias relativas*. Este polígono tiene la misma forma que el polígono de frecuencias construido a partir del mismo conjunto de datos, pero con una escala diferente en los valores del eje vertical. La figura 11 muestra un polígono de frecuencias.

**Figura 11. Polígono de frecuencias**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

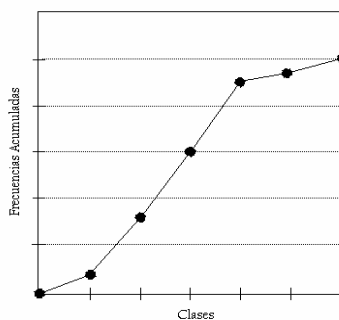
De las ventajas de los polígonos de frecuencia se puede decir:

- Es más sencillo que su correspondiente histograma.
- Traza con más claridad el perfil del patrón de los datos.
- Se vuelve cada vez más liso y parecido a una curva conforme se aumenta el número de clases y el número de observaciones.
- Se pueden superponer sobre unos mismos ejes dos o más polígonos de frecuencia para la comparación de sus correspondientes distribuciones de frecuencia.

#### 2.1.1.5. Distribuciones de frecuencias acumuladas.

La frecuencia acumulada para la primera clase es su misma frecuencia por ser la primera, es decir, no acumula porque antes no hay frecuencias. Las demás frecuencias acumuladas se calculan sumando la frecuencia acumulada inmediatamente anterior. Las frecuencias acumuladas reemplazan a las frecuencias en el polígono de frecuencias, o sea, van ubicadas en el eje de las ordenadas (eje Y). En la figura 12 se ve un polígono de frecuencias acumuladas.

**Figura 12. Distribución de Frecuencia acumulada**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Las características principales del polígono de frecuencias acumuladas son:

- El polígono empieza en cero, una clase antes de la primera clase y termina en la última clase.
- El polígono no cierra.
- Cada punto del polígono tiene coordenadas  $(x_i, f_i)$ , donde  $x_i$  es la marca de clase.

## **2.2 MEDIDAS DE TENDENCIA CENTRAL**

Al tener un conjunto de datos es posible describirlos de dos maneras: la primera como se hizo en la explicación anterior de estadística descriptiva, por medio de una distribución de frecuencias, un histograma, un polígono de frecuencias, entre otros.

La segunda forma, como se explicará a continuación, es describirlos mediante un único valor que sea representativo de todos los datos. A este valor se le conoce como medida de tendencia central. Se hace necesario saber que tan dispersos están todos los datos respecto a esta medida de tendencia central, y es por eso que además se estudian las medidas de dispersión.

### **2.2.1 Media Aritmética**

También se llama promedio aritmético. Si se suman cada uno de los valores de una variable  $x$  y se divide el resultado por el número de valores sumados, se obtiene la media aritmética o promedio. Cuando se trabaja con poblaciones se

calcula la media poblacional y cuando se trabaja con muestras se calcula la media muestral.

### 2.2.1.1 Media poblacional

Se simboliza con la letra griega  $\mu$ . Es la suma de todos los valores de la población, dividida entre el número total de dichos datos. Se calcula mediante la fórmula:

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \quad (46)$$

$\mu$ : Es la media poblacional

$n$ : Es el número total de datos de la población.

$X_i$ : Valor  $i$  de la población.

### 2.2.1.2 Media muestral

Se simboliza por  $\bar{X}$  y es la suma de los valores de una muestra, dividida entre el número total de los mismos. Se calcula mediante la fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (47)$$

La media muestral, o cualquier otra medida basada en una muestra se le conoce como dato estadístico.

### 2.2.2 Media ponderada.

Es un caso especial de la media aritmética. Se aplica cuando hay observaciones con un mismo valor, lo cual puede ocurrir si los datos se han agrupado en una distribución de frecuencias. Se calcula mediante la fórmula:

$$\bar{X}_p = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \quad (48)$$

Donde  $f_i$  es la frecuencia con que se repite el dato, también llamado peso, y  $n$  es el número de datos.

### 2.2.3 Mediana

Si en un conjunto de valores se observa que la tendencia de los datos está sesgada (inclinada) hacia los valores altos o hacia los bajos, es conveniente utilizar una medida más representativa llamada mediana. La mediana es el valor que corresponde al punto medio de los datos luego de ordenarlos de menor a mayor o viceversa de tal manera que el 50% de dichos datos son mayores que ella y el otro 50% son menores.

Para cuando el número de datos es pequeño la mediana se puede calcular de dos formas:

- **Si el número de datos es impar.** Una vez que los datos estén organizados en orden creciente o decreciente, entonces la mediana estará a la mitad de los datos. Por ejemplo si se tienen los siguientes datos 1,7,25,4,7,2,8,7,1,6,6. Al

organizarlos de mayor a menor, se tiene: 25,8,7,7,7,6,6,4,2,1,1. El dato central es 6, que es la mediana del conjunto de datos.

- **Si el número de datos es par.** Después de organizar los datos, la mediana será la media de los dos datos centrales. Por ejemplo, dado los datos 2,3,6,5,4,9,10,11,13,16 y 18, al organizarlos se tiene: 2,3,4,5,5,6,9,10,11,13,16,18. Los datos centrales son 6 y 9 y su media es 7.5 que es la mediana del conjunto de datos.

#### **2.2.4 Moda.**

Es el valor de la observación que aparece con más frecuencia. La moda es muy útil para describir datos de nivel nominal y ordinal, aunque se puede utilizar en cualquier nivel de datos.

### **2.3 MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS.**

Cuando se tiene una gran cantidad de datos es muy tedioso hallarle sus medidas de tendencia central a menos que ellos se encuentren agrupados en una distribución de frecuencias.

#### **2.3.1 Media**

Se considera que las observaciones en cada clase están representadas por la marca de clase (punto medio de cada clase), para su cálculo se emplea la fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \quad (49)$$

Donde  $\bar{X}$  es la media aritmética,  $x_i$  es la marca de clase,  $f_i$  es la frecuencia de clase y  $n$  es el número total de clases.

### 2.3.2 Mediana

Para datos agrupados no es posible encontrar la mediana exacta pues los datos no son identificables. Sólo se puede hacer una estimación de ella. Para estimar la mediana se utiliza la fórmula:

$$Med = L + \frac{\frac{n}{2} - F_i}{f_{clase}} I \quad (50)$$

Donde:

$L$  Es el límite inferior de la clase que contiene la mediana.

$n$  Es el número total de frecuencias.

$f_{clase}$  Es la frecuencia de la clase que contiene la mediana.

$F_i$  Es la frecuencia acumulada de las clases que están antes de la clase que contiene la mediana.

$I$  Es la amplitud del intervalo o clase que contiene la mediana.

### **2.3.3 Moda.**

La moda se aproxima al punto medio de la clase que contenga el mayor número de frecuencias de clase.

## **2.4 MEDIDAS DE DISPERSIÓN**

La media, la mediana y la moda sólo localizan en centro de los datos, pero no dicen nada acerca de cómo están dispuestos respecto a la medida de tendencia central. No es significativo saber cual es la medida de tendencia central de un conjunto de datos si no se sabe que tan apartados están los datos de esa medida. Es decir, tomar una decisión con solo saber una medida de tendencia central es arriesgado. Por ejemplo si usted va a cruzar un río y ve un aviso que dice que la profundidad promedio es de 1.5 metros, ¿Usted lo cruzaría con toda confianza? Muy seguramente si decide o no sería muy apresurado de su parte. Pero si en el aviso se anuncia que las profundidades del río a lo ancho no están muy alejadas del promedio, digamos que como máximo 0,2 metros, la decisión ahora de cruzar o no el río tiene mayor respaldo. En conclusión se necesitaría saber como varían las profundidades respecto al promedio, o sea, que tan dispersas están las medidas de las profundidades respecto al promedio.

### **2.4.1 Amplitud o intervalo de variación**

Es la diferencia entre el valor más grande y el valor más pequeño de un conjunto de datos. Por ejemplo, si los datos son 2,7,8,9,6,3,5 y 2, entonces la amplitud de variación es  $Av=9-2=7$ .

### 2.4.2 Desviación media

Es el promedio de los valores absolutos de las desviaciones con respecto a la media aritmética. La desviación de un dato respecto a la media es la diferencia entre estos valores. Por ejemplo, si la media de un conjunto de datos es 5,3 y si de ese conjunto de datos se toma a 5,8, entonces la desviación media es  $|5,3 - 5,8| = 0,5$ . La fórmula para calcular la desviación media es:

$$Dm = \frac{\sum_{i=1}^n |\bar{X} - x_i|}{n} \quad (51)$$

Donde  $\bar{X}$  es la media,  $x_i$  es el dato  $i$  y  $n$  es el número de datos. El valor absoluto asegura que las diferencias entre la media y cada dato sea positiva.

### 2.4.3 Varianza

Otra forma de evitar que las desviaciones respecto a la media aritmética sean negativas es elevarlas al cuadrado. La varianza es la media aritmética de las desviaciones cuadradas. La varianza puede ser poblacional, que se simboliza por la letra sigma al cuadrado ( $\sigma^2$ ); o puede ser muestral, simbolizada por la letra  $s$  al cuadrado ( $s^2$ ). Las fórmulas para cada caso son:

$$\sigma^2 = \frac{\sum_{i=1}^n (\mu - x_i)^2}{n} \quad (52)$$

$$s^2 = \frac{\sum_{i=1}^n (\bar{X} - x_i)^2}{n-1} \quad (53)$$

Donde  $\mu$  es la media poblacional,  $\bar{X}$  es la media muestral,  $x_i$  es el dato  $i$  y  $n$  es el número total de datos de la muestra o población, según el caso.

#### 2.4.4 Desviación estándar

Es la raíz positiva de la varianza. También puede ser poblacional ( $\sigma$ ) o muestral ( $s$ ). Las fórmulas respectivas son:

$$\sigma^2 = \sqrt{\frac{\sum_{i=1}^n (\mu - x_i)^2}{n}} \quad (54)$$

$$s^2 = \sqrt{\frac{\sum_{i=1}^n (\bar{X} - x_i)^2}{n-1}} \quad (55)$$

### 2.5 MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS

Es posible encontrar las medidas de dispersión para datos agrupados en una distribución de frecuencias.

#### 2.5.1 Amplitud de variación

Es la diferencia entre el límite inferior de la clase más baja y el límite superior de la clase más alta.

### 2.5.2 Desviación estándar

La fórmula de la desviación estándar para datos agrupados en una distribución de frecuencias es:

$$s = \sqrt{\frac{\sum_{i=1}^n f_i x_i^2 - \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}}{\sum_{i=1}^n f_i - 1}} \quad (56)$$

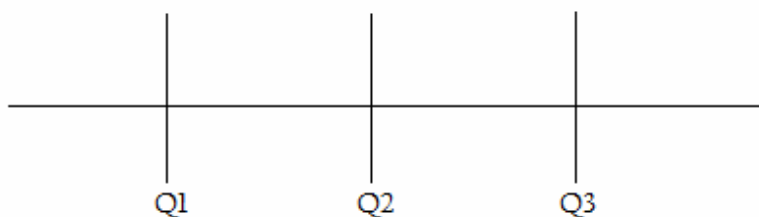
Donde  $s$  es la desviación estándar muestral,  $x_i$  es la marca de clase,  $f_i$  es la frecuencia de clase y  $n$  es el número de clases.

### 2.6 DECILES, CUARTILES Y CENTILES

Son medidas de dispersión que dividen los datos en partes iguales, así los deciles los dividen en 10 partes iguales, los cuartiles en cuatro partes iguales y los centiles en 100 partes iguales. La figura 13 esquematiza la posición de estas medidas.

**FIGURA 13. Esquematación de las posiciones de A) Cuartiles, B) Deciles, C) Centiles**

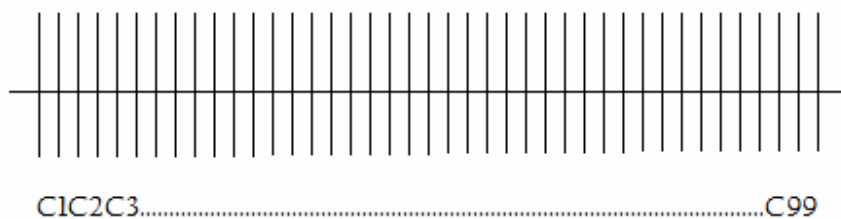
A)



B)



C)



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Se podría pensar que los deciles, cuartiles y centiles no son medidas de dispersión porque sólo dividen el conjunto de datos en partes iguales, como lo hace la mediana; pero la dispersión se nota cuando se habla que un porcentaje de los datos está por encima de una de estas medidas. Según el esquema de la figura 13 se puede ver por ejemplo que el cuartil Q1 separa el 25% del 75% restante de la derecha, el decil D7 separa el 70% del 30% restante de la derecha y

el centil C33 separa el 33% de los datos del 67% restante de la derecha. Es decir, cada una de estas medidas separa un cierto porcentaje de datos a la derecha o a la izquierda. Las tablas 2, 3 y 4 muestran los porcentajes que separan cada una de estas medidas de dispersión.

**Tabla 2. Porcentajes que separan los cuartiles de un conjunto de datos.**

Cuartel	Porcentaje Menor	Porcentaje Mayor
Q1	25%	75%
Q2	50%	50%
Q3	75%	25%

**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Es de notar que el cuartil Q2 es equivalente a la mediana porque separa el 50% de los datos. También el decil D5 y en centil C50 son equivalentes a la mediana.

**Tabla 3. Porcentajes que separan los deciles de un conjunto de datos.**

Decil	Porcentaje Menor	Porcentaje Mayor
D1	10%	90%
D2	20%	80%
D3	30%	70%
D4	40%	60%
D5	50%	50%
D6	60%	40%
D7	70%	30%
D8	80%	20%
D9	90%	10%

**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

**Tabla 4. Porcentajes que separan los centiles de un conjunto de datos.**

Centil	Porcentaje Menor	Porcentaje Mayor
C1	1%	99%
C2	2%	98%
C3	3%	97%
C4	4%	96%
C5	5%	95%
C6	6%	94%
C7	7%	93%
.		.
C98	98%	2%
C99	99%	1%

**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Para ubicar un cuartil, un decil, o un centil se utiliza la fórmula:

$$L = (n + 1) \frac{P}{100} \quad (57)$$

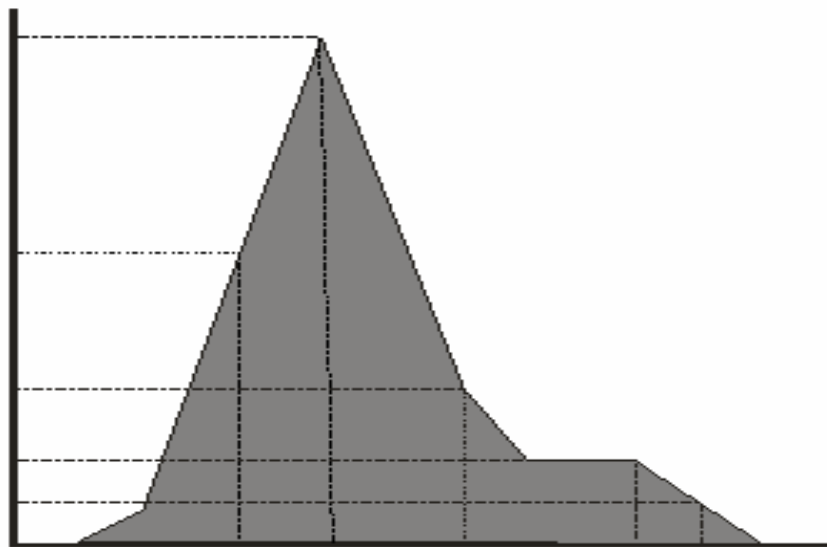
donde  $n$  es el total de observaciones y  $P$  puede ser el cuartil, decil o centil que se quiere ubicar y toma el valor de porcentaje menor que la medida separa. Por ejemplo si se quiere situar el tercer cuartil, entonces  $P=75$  ya que este cuartil separa el 75% de los datos del 25% que queda mayor (ver tabla 2).

De una manera más práctica si utilizamos la notación “el percentil  $P_K$ ” entonces “ $K$ ” es el porcentaje de datos acumulados.

## 2.7 INTERPRETACIÓN Y USOS DE LA DESVIACIÓN ESTÁNDAR

Si se tiene un polígono de frecuencias con 7 clases como se muestra en la figura 14:

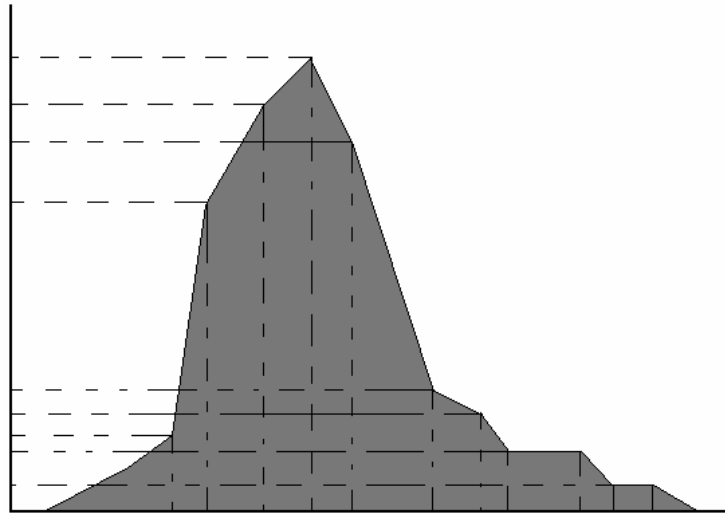
**Figura 14. Polígono de frecuencias con 7 clases**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Si se aumenta el número de clases a 14, se obtendrá un polígono como se muestra en la figura 15:

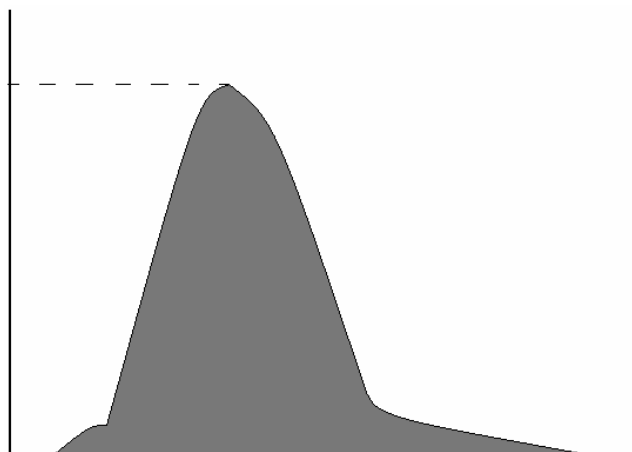
**Figura 15. Polígono de frecuencias con 14 clases**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Se nota que al aumentar el número de frecuencias, el polígono tiende a ser una curva suave. Si el número de clases crece hasta el infinito, entonces el polígono de frecuencias se convierte en una curva continua, como se muestra en la figura 16

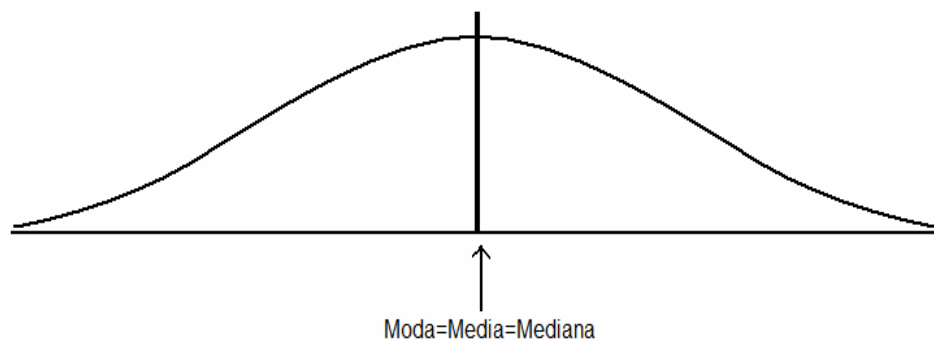
**Figura 16. Polígono de frecuencias con infinitas clases**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Dentro de los polígonos de frecuencia de curva suave se destaca el de la distribución de frecuencia tipo campana o también llamado distribución simétrica, lo cual significa que la distribución tiene la misma forma en ambos lados del eje central. Si el polígono se doblara por la mitad, ambas partes coincidirían. En una distribución de este tipo la moda, la mediana y la media están en el eje de simetría. La figura 17 muestra una distribución de frecuencias simétricas.

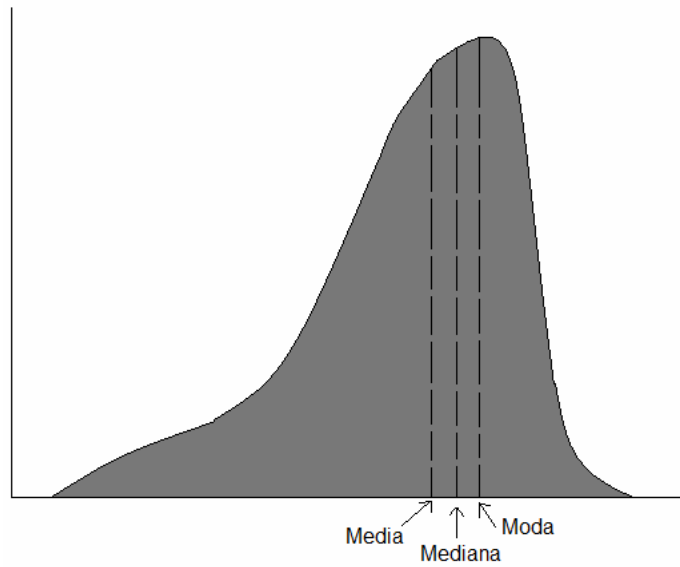
**Figura 17. Distribución de frecuencias simétrica o de campana.**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

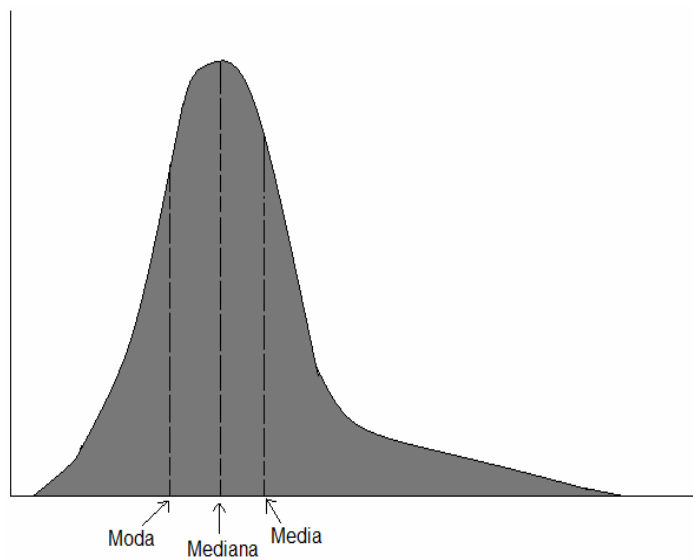
Cuando una distribución de frecuencias de curva suave no es distribuida simétricamente, se dice que el conjunto de datos es *asimétrico o sesgado*. La asimetría puede ser positiva o negativa. Una asimetría positiva significa que la mayoría de los datos están a la derecha de la línea central. Una asimetría negativa significa que la mayoría de los datos está hacia el lado izquierdo. Desde el punto de vista gráfico una distribución con simetría positiva está inclinada hacia la derecha y una distribución con simetría negativa está inclinada a la izquierda. Las figuras 18 y 19 muestran respectivamente, distribuciones con simetría positiva y negativa así como también como las posiciones de las medidas de tendencia central moda, media y mediana.

**Figura 18. Distribución de frecuencias con simetría positiva y ubicación de las medidas de tendencia central**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

**Figura19. Distribución de frecuencias con simetría negativa y ubicación de las medidas de tendencia central**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

## **2.8 PROBABILIDADES**

En temas anteriores se trató sobre la estadística descriptiva, es decir, como organizar los datos de una muestra o población y que tanto estaban tan dispersos respecto a una medida de tendencia central.

La atención se fija ahora en la estadística inferencial. Esta se ocupa del estudio de la posibilidad de que algo ocurra en el futuro. Esta rama de la estadística se encarga de obtener conclusiones de una población a partir de una muestra. Es claro que la toma de decisiones implica una cierta incertidumbre, es decir no se sabe si la decisión es la correcta; pero si se puede saber la probabilidad que resulte confiable tal decisión. Es por ello que la teoría de las probabilidades constituye la base fundamental para la estadística inferencial.

La teoría de la probabilidad es una teoría muy intrincada y desarrollada para describir los sucesos aleatorios. La palabra aleatorio proviene del vocablo *alea*, el cual significa suerte o azar. Un fenómeno aleatorio es aquel cuyo resultado está fuera de control y que depende del azar.

Qué tanta incertidumbre se tiene respecto a un fenómeno constituye la probabilidad de ocurrencia de un evento. La palabra evento, suceso o acontecimiento significan lo mismo y más adelante se definirán.

### **2.8.1 Definiciones Previas**

Para poder adelantarse al concepto de probabilidad se hace necesario definir algunos conceptos importantes:

### **2.8.1.1 Experimento**

Es un proceso que conduce a que ocurra una y solo una de varias observaciones posibles. El experimento relativo a la probabilidad no es algo que se hace en el laboratorio, es casi siempre una actividad imaginaria. Los experimentos aleatorios cumplen las siguientes características:

- \* Se puede repetir las veces que el experimentador requiera, es decir, está sujeto a repetición.
- \* Su resultado es incierto, o sea, está fuera de control, depende del azar.
- \* Se puede anticipar la totalidad de los resultados posibles.
- \* Se observa en él un patrón de regularidad que se hace más evidente a medida que se aumentan las repeticiones.

Lanzar una moneda al aire es un experimento y las observaciones posibles son “que salga cara” y “que salga sello”. Otro experimento puede ser: el jugador de fútbol X ejecuta un tiro penal, las observaciones posibles son “que anote gol” y “que erre el gol”.

### **2.8.1.2 Evento**

Un evento está relacionado con el azar, o sea, su ocurrencia depende exclusivamente de este, es por esto que algunas veces se le conoce como evento aleatorio. Se puede definir como el conjunto de uno o más resultados de un experimento. Para el jugador que ejecuta el penal, un evento puede ser:

E1: marca el gol; otro puede ser:

E2: La pelota impacta en el travesaño.

La unión de estos eventos es un evento:

$E3 = E1 \cup E2$ : la pelota impacta en el travesaño o marca gol.

La intersección es otro evento:

$E4 = E1 \cap E2$ : La pelota impacta en el travesaño y marca gol.

En el experimento "se lanza un dado" se pueden tener los siguientes eventos:

E1: Que se obtenga un número par

E2: Que se obtenga un 2

E3: Que se obtenga un número par mayor que 2

E4: Que se obtenga un número impar

Y se pueden describir una infinidad de eventos. Es posible que dos eventos sucedan al tiempo; en el evento del lanzamiento del dado, E1 y E2 pueden suceder al tiempo, es decir se puede obtener un 2 y este es un número par. Los eventos E1 y E4 no pueden ocurrir simultáneamente ya que con un dado no se puede obtener simultáneamente un número par y un número impar.

### ***Eventos mutuamente excluyentes***

Son eventos en que la ocurrencia de unos implica la no ocurrencia de los otros, es decir son eventos que no pueden ocurrir simultáneamente. En el ejemplo del párrafo anterior E1 y E4 son mutuamente excluyentes; pero E1 y E2 no lo son.

### 2.8.1.3 Espacio muestral

Es un conjunto universal que enmarca todas las posibles ocurrencias de un experimento aleatorio. Generalmente el espacio muestral se denota por la letra  $S$ . Los elementos del espacio muestral son eventos, así por ejemplo, si el experimento consiste en el nacimiento de un bebé, cuatro eventos podrían ser:

E1: Que el bebé sea de sexo masculino.

E2: Que el bebé sea de sexo femenino.

E3: Que el bebé sea rubio.

E4: Que el bebé sea moreno.

El espacio muestral para este experimento es  $S_1 = \{E1, E2, E3, E4\}$ . En este ejemplo el espacio muestral contiene, aparte de estos cuatro eventos uniones o intersecciones de ellos, por ejemplo:

E5 =  $E1 \cup E4$ : que el bebé sea de sexo masculino o que sea moreno.

E6 =  $E2 \cap E3$ : que el bebé sea de sexo femenino y que sea rubio.

Estos eventos derivados de los cuatro primitivos hacen parte, en forma tácita, del espacio muestral  $S$  y no se describen explícitamente sólo por comodidad.

La forma en que se defina un espacio muestral de un experimento dependerá del problema que se enfrente. Por ejemplo, si en el experimento del nacimiento del bebé, solo interesa saber si es de sexo masculino o femenino, entonces se utilizará el espacio muestral  $S_2 = \{E1, E2\}$ . Es decir que se pueden utilizar diferentes espacios muestrales para un experimento.

### 2.8.2 Definición de Probabilidad de un evento

La probabilidad es un concepto abstracto que se usa para describir el grado de incertidumbre de un evento. La probabilidad es un número entre cero y uno, inclusive que describe la probabilidad relativa de que ocurra un evento. Si se denota la probabilidad de un evento con la letra  $P$ , entonces  $0 \leq P \leq 1$ . Si la probabilidad de un evento es 0, entonces el evento es imposible, si es 1, entonces el evento es seguro. Por ejemplo ¿cuál es la probabilidad de que una persona muera? Como es un evento seguro, la probabilidad de que suceda es 1. ¿Cuál es la probabilidad de que un hombre salga embarazado? Hasta lo que va del desarrollo de la ciencia parece que esto es imposible, por tanto la probabilidad de que suceda es 0.

Una probabilidad se puede expresar como una fracción decimal tal como 0.23, 0.40 ó 0.134, como un número racional tal como  $\frac{2}{3}$ ,  $\frac{1}{5}$ , ó  $\frac{3}{10}$  ó en porcentaje así como 15%, 82%, ó 23.7%.

### 2.8.3 Puntos de vista de la probabilidad

El análisis de las probabilidades se puede hacer desde distintos puntos de vista: desde el punto de vista objetivo, que a su vez puede ser clásico o empírico; y desde el punto subjetivo.

### 2.8.3.1 Probabilidad objetiva

De manera intuitiva, se acostumbra a considerar la probabilidad de un evento como un número. Este número puede ser calculado de acuerdo con la proporción de casos favorables o de veces que ocurrió un evento en un tiempo determinado.

#### ***Probabilidad clásica***

Se basa en la consideración de que todos los resultados de un experimento son igualmente posibles. La fórmula para el cálculo de esta probabilidad es:

$$P(E) = \frac{\text{Número de resultados a favor del evento}}{\text{Número de resultados posibles}} \quad (58)$$

donde  $P(E)$  es la probabilidad de que ocurra el evento E.

#### ***Probabilidad empírica***

Se basa en las frecuencias relativas de eventos que han sucedido en un tiempo pasado determinado. Se calcula observando el número de veces en que el evento ocurrió en un tiempo determinado. La fórmula para calcular este tipo de probabilidad es:

$$P(E) = \frac{\text{Número de veces que ocurrió el evento en el pasado}}{\text{Número total de observaciones}} \quad (59)$$

### 2.8.3.2 Probabilidad subjetiva

Estas probabilidades se pueden interpretar como evaluaciones personales o subjetivas. Se asignan de acuerdo a cualquier información que se disponga con base en suposiciones razonadas, intuición u otras subjetividades. Por todo lo anterior no existe una fórmula para calcular la probabilidad subjetiva, pues es un valor que se asigna de acuerdo a lo que se piensa, podría ser.

### 2.8.4 Reglas de Probabilidad

Las reglas de probabilidad son fórmulas que se aplican a eventos que son y no son mutuamente excluyentes.

#### 2.8.4.1 Regla de la adición

Si  $E_1$  y  $E_2$  son eventos (no importa si son o no son mutuamente excluyentes) entonces la probabilidad de que ocurra  $E_1$  o  $E_2$  está dado por la fórmula  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ . Se utilizan los símbolos de unión ( $\cup$ ) e intersección ( $\cap$ ) porque los eventos son conjuntos. La unión se refiere a la disyunción (o) y la intersección se refiere a la conjunción (y). Si los eventos fueran mutuamente excluyentes, entonces no pueden suceder al mismo tiempo, es decir no podría ocurrir  $E_1$  y  $E_2$ , por tanto la probabilidad  $P(E_1 \cap E_2) = 0$  y la fórmula de la adición quedaría  $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ .

La regla de la adición se puede escribir para tres eventos, así:

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3) \quad (60)$$

#### 2.8.4.2 Regla de la multiplicación

A veces la probabilidad de un evento determinado depende de que otro evento se halla producido o no con anterioridad. Es decir, en algunas ocasiones la aparición de un evento puede influir o no en la aparición de otro, o bien que el otro se haga mas o menos probable. Cuando la aparición de un evento influye en la probabilidad del otro, se dice que los dos eventos son dependientes, en caso contrario se dice que los eventos son independientes.

La probabilidad de eventos dependientes se les conoce como probabilidades condicionadas. Se simbolizan  $P(A/B)$  y se leen “La probabilidad de que se de el evento A dado que B ocurrió”.

La regla de la multiplicación para dos eventos independientes se enuncia:  $P(E1 \cap E2) = P(E1) \cdot P(E2)$ , y en general para n eventos independientes

$$P(E1 \cap E2 \cap E3 \cap \dots \cap En) = P(E1) \cdot P(E2) \cdot P(E3) \dots P(En) \quad (61)$$

Para dos eventos dependientes se enuncia:  $P(E1 \cap E2) = P(E1/E2) \cdot P(E2)$ . Al generalizar para n eventos dependientes:

$$P(E1 \cap E2 \cap E3 \cap \dots \cap En) = P(En/E_{n-1}, \dots, E3, E2, E1) \dots P(E3/E2, E1) P(E2/E1) P(E1). \quad (62)$$

El primer miembro de la derecha se lee: “La probabilidad de que ocurra el evento  $E_n$ , dado que ocurrieron  $E_{n-1}, \dots, E3, E2, E1$ ”

### 2.8.4.3 Regla del complemento

La regla del complemento se utiliza para determinar la probabilidad de que ocurra un evento restando del número 1 la probabilidad de que un evento no ocurra.

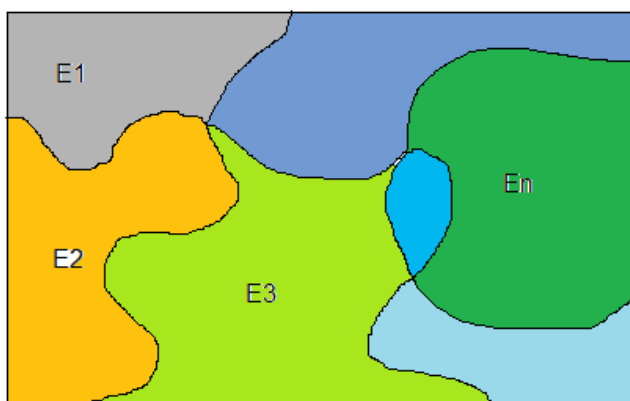
Si  $P(E1)$  es la probabilidad del evento  $E1$  y  $P(\sim E1)$  es el complemento de  $E1$ , es decir que no ocurra  $E1$ , entonces,

$P(E1)+P(\sim E1)=1$  y despejando se tiene:  $P(E1)=1- P(\sim E1)$  o también  $P(\sim E1)=1- P(E1)$ .

### 2.8.5 Regla de la Probabilidad Total

Se llama partición a un conjunto de sucesos  $E_i$  mutuamente excluyentes de tal manera que la unión de todos ellos sea el espacio muestral  $S$ . En términos matemáticos si  $E1 \cup E2 \cup E3 \cup \dots \cup E_n = S$  y además  $E_i \cap E_j \neq \phi$ . La figura 20 esquematiza una partición.

**Figura 20. Partición del espacio muestral  $S$ , compuesta por los eventos  $E1$ ,  $E2$ ,  $E3$ ,  $E_n$**



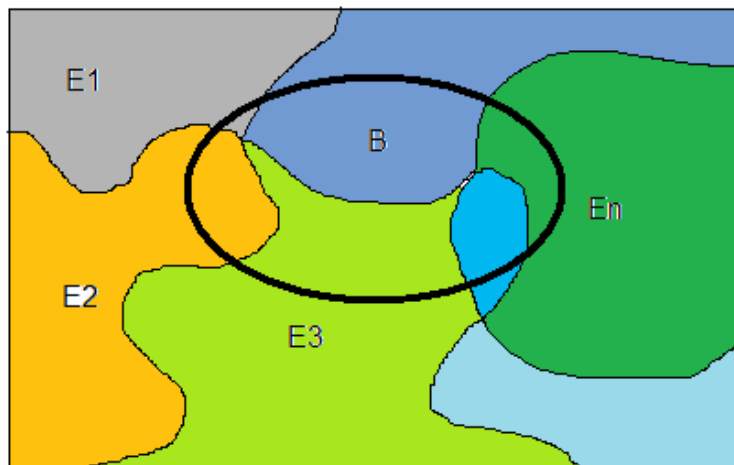
**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

La regla de la probabilidad total establece que si un conjunto de eventos  $E_i$  forma una partición del espacio muestral  $S$  y si la probabilidad de cada uno de estos eventos es diferente de ceros, es decir  $P(E_i) \neq 0$ , para todo  $E_i$ , entonces para cualquier otro suceso  $B$  se cumple que:

$$P(B) = P(B/E_1)P(E_1) + P(B/E_2)P(E_2) + P(B/E_3)P(E_3) + \dots + P(B/E_n)P(E_n) \quad (63)$$

La figura 21 ilustra esta regla.

**FIGURA 21. Partición del espacio muestral  $S$ , compuesta por los eventos  $E_1$ ,  $E_2$ ,  $E_3, \dots$ ,  $E_n$  y el evento cualquiera  $B$**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

### 2.8.6. Distribuciones de Probabilidad

Una distribución de probabilidad hace referencia a toda una gama de probabilidad de los eventos de un experimento y muestra que tan probable es un evento futuro, relacionado con dicho experimento. Se puede concebir como una distribución teórica de frecuencias, es decir, una distribución que describe cómo se espera que

varíen las probabilidades para el experimento dado. Las distribuciones de frecuencias son modelos de gran utilidad, pues ayudan mucho en las inferencias y en la toma de decisiones en condiciones de incertidumbre.

### 2.8.6.1 Definición

Una distribución de probabilidad es un modelo matemático que organiza los eventos posibles de un experimento y que les asocia su probabilidad correspondiente. Con el siguiente ejemplo podremos tener una mejor comprensión.

Ejemplo 1:

*Supongamos que se seleccionan al azar dos esferas donde hay 5 blancas y 3 negras. ¿Cuál es la distribución de probabilidades de este experimento?.*

SOLUCIÓN:

Dado que se puede sacar una esfera blanca seguida de una negra o una blanca, o primero una esfera negra seguida de una negra o una blanca, entonces el espacio muestral es:  $S = \{BB, BN, NB, NN\}$ . Los eventos son:

E1: Sacar una esfera blanca y luego una blanca (BB).

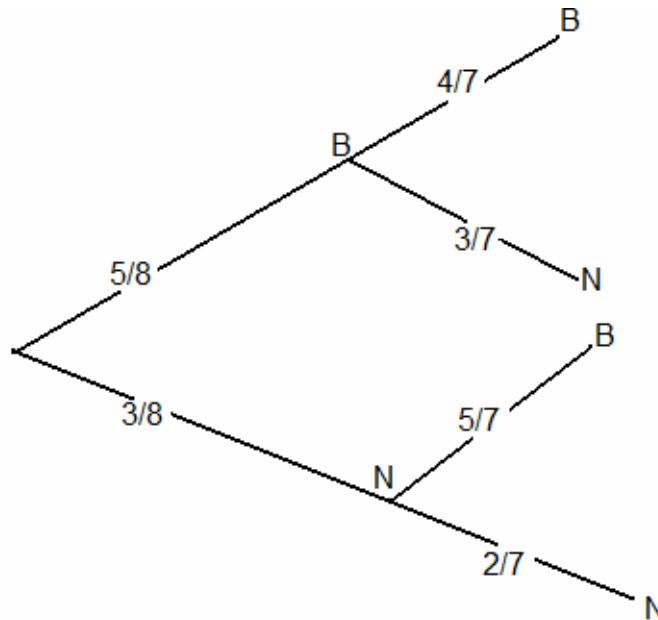
E2: Sacar una esfera blanca y luego una negra (BN)

E3: Sacar una esfera negra y luego una blanca (NB)

E4: Sacar una esfera negra y luego una negra (NN)

Las probabilidades de cada evento se calculan con la ayuda de un diagrama de árbol figura 22

**Figura 22. Diagrama de árbol para las probabilidades de sacar 2 esferas en el ejemplo 1**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Las probabilidades son:

$$P(E1) = \frac{5}{8} * \frac{4}{7} = \frac{20}{56} = \frac{5}{14}, \quad P(E2) = \frac{5}{8} * \frac{3}{7} = \frac{15}{56}, \quad P(E3) = \frac{3}{8} * \frac{5}{7} = \frac{15}{56},$$

$$P(E4) = \frac{3}{8} * \frac{2}{7} = \frac{6}{56} = \frac{3}{28}$$

Al resumir estas probabilidades con sus eventos se obtiene una distribución de probabilidades como se muestra en la tabla 5

**Tabla 5. Distribución de probabilidades para el ejemplo planteado.**

Evento	Probabilidad	Número de Blancas	Número de Negras
E1	5/14	2	0
E2	15/56	1	1
E3	15/56	1	1
E4	3/28	0	2

**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Es de notar que la suma de las probabilidades de una distribución de frecuencias es igual a 1.

### ***Variable Aleatoria***

Los eventos de un experimento aleatorio se pueden contar. Se puede definir una función que cuente los resultados de los eventos posibles de un experimento. Si en el ejemplo 1 se quisiera contar el número de esferas negras, es decir, definir el evento “que salga esfera negra”, entonces se utiliza la variable aleatoria  $X$  así:

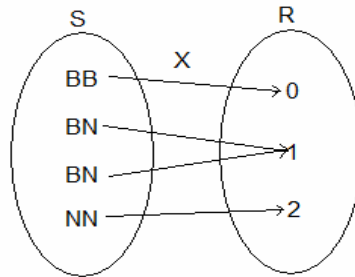
$X=0$ , hay cero esferas negras.

$X=1$ , hay una esfera negra.

$X=2$ , has dos esferas negras.

El diagrama de Venn para esta variable aleatoria se puede observar en la figura 23.

**Figura 23. Diagrama de Venn para la variable aleatoria que cuenta el número de esferas negras en el ejemplo 1.**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Note que la variable aleatoria es *una función* que asocia un número real, perfectamente definido, a cada resultado posible de un experimento. La tabla 5 se puede reescribir para la variable aleatoria en mención para así obtener la tabla 6, que es en si una distribución de probabilidades.

**Tabla 6. Distribución de Probabilidades y variables aleatoria que cuenta las esferas negras en el ejemplo 1.**

Evento	Probabilidad	X
BB	5/14	0
BN	15/56	1
NB	15/56	1
NN	3/28	2

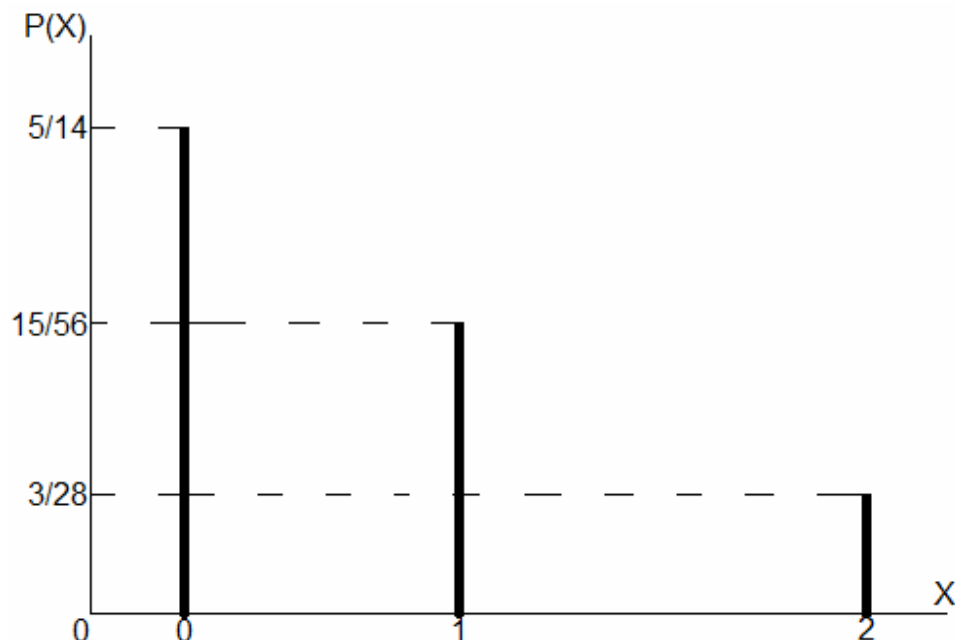
**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

De la tabla 6 se puede calcular, por ejemplo, la probabilidad de que la variable aleatoria  $X$  tenga un valor de 2, así  $P(X = 2) = \frac{3}{28}$ .

De lo anterior se puede deducir que las probabilidades son funciones de las variables aleatorias. Por esta razón es que a las distribuciones de probabilidad se les conoce como *funciones de densidad de probabilidad (fdp)*. Una fdp asocia a cada valor de una variable aleatoria una probabilidad.

Para el ejemplo 1 se puede obtener la gráfica de la distribución de frecuencias en función de la variable aleatoria que cuenta las esferas negras, es decir, la gráfica de la función de densidad de probabilidad. Esta gráfica se muestra en la figura 24.

**Figura 24. Gráfica de la distribución de probabilidades o función de densidad de probabilidad de la variable aleatoria que cuenta el número de esferas negras en el ejemplo 1.**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

## 2.9 DISTRIBUCIÓN DE PROBABILIDAD DISCRETA

Como se mencionó anteriormente, la distribución de probabilidad discreta nace de una variable aleatoria discreta. Existen varios tipos especiales de esta clase de distribución, y entre las más conocidas están: la *distribución binomial*, la *distribución geométrica*, la *distribución de Poisson*,

### 2.9.1 Función de distribución acumulativa

Hay muchos problemas donde es importante conocer la probabilidad de que el valor de una variable aleatoria sea menor o igual a un número real  $a$ . Se puede escribir, por consiguiente, la probabilidad de que la variable aleatoria  $X$  tome valores menores o iguales que  $a$  como la función  $F$  tal que  $F(X) = P(X \leq a)$ , a esta función se le conoce como función de distribución acumulativa o simplemente función de distribución de la variable aleatoria  $X$ . Matemáticamente esta función se define por:

Si  $X$  es una variable aleatoria discreta,  $F(X) = P(X \leq a) = \sum_{i \leq a}^n P(X)$ , es decir la

función de distribución acumulativa es la suma de todas las probabilidades que están por debajo de la probabilidad de que la variable aleatoria tome el valor de  $a$ . Para el ejemplo 1 la tabla 7 muestra los valores de la función de distribución acumulativa de la variable aleatoria  $X$  que cuenta el número de esferas negras.

**Tabla 7. Función de distribución de la variable aleatoria X que cuenta el número de esferas negras del ejemplo 1.**

X	P(X)	F(X)
0	5/14	5/14
1	15/56	35/56
1	15/56	50/56
2	3/28	(56/56)=1

**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Si se quisiera calcular la probabilidad de que el número de esferas negras sea menor que dos, entonces se suman las probabilidades correspondientes a los valores de la variable aleatoria menor que dos., es decir,  $(50/56) \approx 0.9$ .

### **2.9.2 Media, varianza y desviación estándar de una distribución de probabilidad discreta**

Al igual que en las distribuciones de frecuencias, en las distribuciones de probabilidad se pueden calcular la media, como una medida de tendencia central; la desviación estándar y la varianza como una medida de dispersión.

➤ **La media:** Es un valor típico que representa a la distribución de probabilidad. Se simboliza por  $\mu$  y viene dada por:

$$\mu = \sum_x XP(X) \quad (64)$$

donde  $P(X)$  es la probabilidad de que la variable aleatoria  $X$  tome un valor entre  $x$  real. Algunas veces se habla del valor esperado de una variable aleatoria  $E(X)$  en vez de la media de la distribución. Los significados son idénticos.

➤ **La varianza:** Como se ha visto anteriormente la varianza es una medida que muestra la dispersión de los datos. En el caso de las distribuciones de probabilidad la varianza  $\sigma^2$  se calcula mediante la fórmula

$$\sigma^2 = \sum_x [(X - \mu)^2 P(X)] \quad (65)$$

➤ **La desviación estándar:** Es la raíz cuadrada de la varianza, se simboliza por  $\sigma$  y viene dada por la fórmula:

$$\sigma = \sqrt{\sum_x [(X - \mu)^2 P(X)]} \quad (66)$$

### 2.9.3 Tipos de Distribuciones de Probabilidad Discreta

A continuación se dará una definición breve de las distribuciones de probabilidad discreta más usadas.

#### 2.9.3.1 Distribución Binomial

Es una distribución de probabilidad discreta del número de éxitos en una secuencia de  $n$  experimentos independientes, cada uno de los cuales tiene probabilidad  $\theta$  de ocurrir. (La distribución de Bernoulli es una distribución binomial con  $n = 1$ ). Su función de masa de probabilidad está dada por:

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad (67)$$

Para  $x = 0, 1, 2, \dots, n$  siendo  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$  las combinaciones de  $n$  en  $x$  ( $n$  elementos tomados de  $x$  en  $x$ ).

Por ejemplo, la distribución binomial se usa para encontrar la probabilidad de sacar 5 caras y 7 cruces en 12 lanzamientos de una moneda. En este caso se tiene  $x=5$ ,  $n=12$ ,  $\theta=0.5$  y resulta:

$$b(5;12,0.5) = \binom{12}{5} 0.5^5 (1-0.5)^{12-5} = 0.19 \quad (68)$$

Su media y su varianza son:

$$\begin{aligned} \mu &= n\theta \\ \sigma^2 &= n\theta(1-\theta) \end{aligned} \quad (69)$$

La variable aleatoria binomial y su distribución están basadas en un experimento que satisface las siguientes condiciones:

- El experimento consiste en una secuencia de  $n$  intentos, donde  $n$  se fija antes del experimento.
- Los intentos son idénticos, y cada uno de ellos puede resultar en dos posibles resultados, que se denotan por éxito ( $S$ ) o fracaso ( $F$ ) ( $p(S)+p(F)=1$ ).
- Los intentos son independientes, por lo que el resultado de cualquier intento en particular no influye sobre el resultado de cualquier otro intento.
- La probabilidad de éxito es constante de un intento a otro.

Siguiendo estas premisas, la variable aleatoria binomial  $X$  está definida como

$X$  = el número de  $S$  entre los  $N$  intentos.

### 2.9.3.2 Distribución De Poisson

Expresa la probabilidad de un número de eventos ocurriendo en un tiempo fijo si estos eventos ocurren con una tasa media conocida, y son independientes del tiempo desde el último evento.

La distribución fue descubierta por Siméon-Denis Poisson (1781–1840) que publicó, junto con su teoría de probabilidad, en 1838 en su trabajo *Recherches sur la probabilité des jugements en matières criminelles et matière civile* ("Investigación sobre la probabilidad de los juicios en materias criminales y civiles"). El trabajo estaba enfocado en ciertas variables aleatorias  $N$  que cuentan, entre otras cosas, un número de ocurrencias discretas (muchas veces llamadas "arribos") que tienen lugar durante un intervalo de tiempo de duración determinada. Si el número esperado de ocurrencias en este intervalo es  $\lambda$ , entonces la probabilidad de que haya exactamente  $k$  ocurrencias (siendo  $k$  un entero no negativo,  $k = 0, 1, 2, \dots$ ) es igual a:

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (70)$$

donde

$e$  es el base del logaritmo natural ( $e = 2.71828\dots$ ),

$k!$  es el factorial de  $k$ ,

$k$  es el número de ocurrencias de un evento,

$\lambda$  es un número real positivo, equivalente al número esperado de ocurrencias durante un intervalo dado. Por ejemplo, si los eventos ocurren de media cada 4 minutos, y se está interesado en el número de eventos ocurriendo

en un intervalo de 10 minutos, se usaría como modelo una distribución de Poisson con  $\lambda = 2.5$ .

Por ejemplo, si 2% de los libros encuadernados en cierto taller tiene encuadernación defectuosa, obtener la probabilidad de que 5 de 400 libros encuadernados en este taller tengan encuadernaciones defectuosas.  
 $k = 5, \lambda = 400(0,02) = 8$

$$P(5;8) = \frac{8^5 e^{-8}}{5!} = 0,092 \quad (71)$$

Su media y su varianza son:

$$\begin{aligned} \mu &= \lambda \\ \sigma^2 &= \lambda \end{aligned} \quad (72)$$

Como una función de  $k$ , ésta es la función probabilidad de masa. La distribución de Poisson puede ser vista como un caso limitante de la distribución binomial, es decir, que una distribución binomial en la que  $n \rightarrow \infty$  y  $\theta \rightarrow 0$  se puede aproximar por una distribución de Poisson de valor  $\lambda = n\theta$

La distribución Poisson es también llamada Poissoniana, análogamente al término Gaussiana para una distribución de Gauss o distribución normal.

### 2.9.3.3 Distribución Geométrica

En teoría de probabilidad y estadística, la distribución geométrica es cualquiera de las dos distribuciones de probabilidad discretas siguientes:

- La distribución de probabilidad del número  $X$  del ensayo de Bernoulli necesaria para obtener un éxito, contenido en el conjunto  $\{ 1, 2, 3, \dots \}$  ó
- La distribución de probabilidad del número  $Y = X - 1$  de fallos antes del primer éxito, contenido en el conjunto  $\{ 0, 1, 2, 3, \dots \}$ .

Cuál de éstas es la que uno llama "la" distribución geométrica, es una cuestión de convención y conveniencia.

Si la probabilidad de éxito en cada ensayo es  $p$ , entonces la probabilidad de que  $n$  ensayos sean necesarios para obtener un éxito es

$$P(X = n) = (1 - p)^{n-1} p \quad (73)$$

para  $n = 1, 2, 3, \dots$  equivalentemente, la probabilidad de que haya  $n$  fallos antes del primer éxito es

$$P(Y = n) = (1 - p)^n p \quad (74)$$

para  $n = 0, 1, 2, 3, \dots$

En ambos casos, la secuencia de probabilidades es una secuencia geométrica.

Por ejemplo, supongamos que un dado ordinario es lanzado repetidamente hasta que aparece "1" por primera vez. La distribución de probabilidad del número de veces que el dado es lanzado se encuentra en el conjunto infinito  $\{1, 2, 3, \dots\}$  y es una distribución geométrica con  $p=1/6$ .

El valor esperado de una variable aleatoria  $X$  distribuida geoméricamente es  $1/p$  y su varianza es  $(1 - p)/p^2$ ;

$$E(X) = \frac{1}{p}, \quad \text{var}(X) = \frac{1 - p}{p^2} \quad (75)$$

Equivalentemente, el valor esperado de una variable aleatoria distribuida geoméricamente  $Y$  es  $(1 - p)/p$ , y su varianza es  $(1 - p)/p^2$ .

$$E(Y) = \frac{1-p}{P}, \quad \text{var}(Y) = \frac{1-P}{P^2} \quad (76)$$

## 2.10 DISTRIBUCIÓN DE PROBABILIDAD CONTINUA

Una distribución de probabilidad continua está sujeta a una variable aleatoria continua. Las distribuciones de probabilidad continua más conocidas están: *la distribución de probabilidad normal, la distribución log-normal, la distribución triangular (usada en este trabajo), la distribución gama, la distribución Ji-cuadrado, la distribución t de student*, entre otras.

Una función  $f$  definida para un conjunto de valores se le conoce como función de densidad de probabilidad para una variable continua  $X$  si  $P(a \leq X \leq b) = \int_a^b f(x)dx$

para cualesquiera constantes  $a$  y  $b$ . Es decir que para calcular la probabilidad de que el valor de la variable continua esté entre  $x=a$  y  $x=b$  se debe encontrar el área debajo de la curva definida mediante la función de densidad de probabilidad entre esos valores.

La probabilidad de que una variable aleatoria continua tome un valor de  $x=a$  es

$$P(a) = \int_a^a f(x)dx = 0, \text{ es decir, cuando se trabaja con variables aleatorias continuas,}$$

no tiene sentido preguntar por la probabilidad de un valor particular si no por la probabilidad en un intervalo de valores. Por lo tanto, para el caso continuo tiene mayor importancia hablar de la distribución de probabilidad acumulada y no de la distribución de probabilidades en si.

Para que una función  $f$  sea considerada una función de densidad de probabilidad, se deben cumplir las dos condiciones siguientes:

- $f$  debe ser positiva o cero para cualquier valor de la variable aleatoria continua, es decir,  $f(x) \geq 0$  para todo  $x \in (-\infty, \infty)$ . Esto significa que la gráfica de la función de densidad de probabilidad debe estar por encima del eje horizontal (eje de las abscisas).

- El área debajo de la gráfica de la función  $f$  entre menos infinito y el infinito, debe ser igual a 1. Esto es  $\int_{-\infty}^{\infty} f(x)dx = 1$ .

### 2.10.1 Función de distribución acumulativa

Para hablar de las distribuciones de probabilidad acumulada para el caso continuo se hace necesario definir la función de distribución acumulativa o función de distribución.

La función de distribución de una variable continua  $X$  viene dada por

$$F(X) = P(X \leq a) = \int_{-\infty}^a f(t)dt$$
, es decir la función de distribución acumulativa es el

área bajo la curva de la función de densidad de probabilidad  $f$  entre menos infinito y el valor  $a$ .

### 2.10.2 Media, varianza y desviación estándar de una distribución de probabilidades continua.

En las distribuciones de probabilidad continuas también se puede calcular la media, la varianza y la desviación estándar.

➤ **La media o valor esperado:** La media o el valor esperado se calcula por la fórmula

$$\mu = \int_{-\infty}^{\infty} xf(x)dx. \quad (77)$$

➤ **La varianza:**  $\sigma^2$  se calcula mediante la fórmula

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx. \quad (78)$$

➤ **La desviación estándar:** Es la raíz cuadrada de la varianza, se simboliza por  $\sigma$  y viene dada por lo fórmula

$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx} \quad (79)$$

### 2.10.3 Tipos de Distribuciones de Probabilidad Continua

A continuación se explicará brevemente las distribuciones de probabilidad continua más conocidas y usadas. Vale la pena aclarar que al momento de la aplicación del método de Montecarlo, se debe trabajar con una distribución de probabilidad de este tipo.

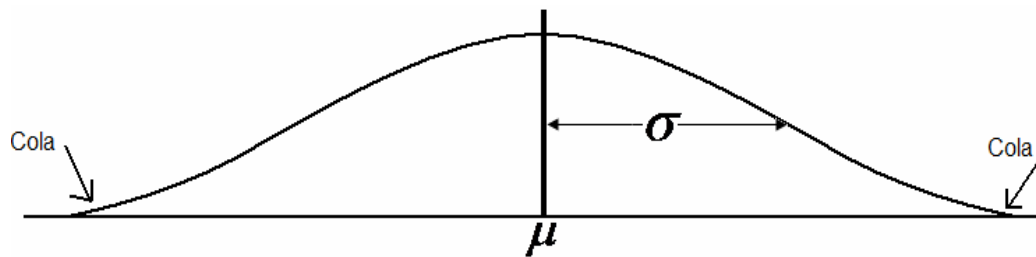
### 2.10.3.1 La distribución Normal

Debería llamarse distribución acumulativa normal, pues como se ha dicho, en las distribuciones continuas tiene mayor sentido hablar de las distribuciones acumulativas. Para mayor simplicidad se le llamarán distribución normal y a la función de densidad de probabilidad se le llamará curva normal. La variable aleatoria continua con una distribución normal es una de las distribuciones más importantes en el campo de la estadística. Las características de la distribución normal son:

- La curva normal es en forma de campana y presenta un pico en centro de la distribución. La media, la mediana y la moda de la distribución son iguales y están localizadas en el centro.
- La distribución normal es simétrica respecto a su media, por lo tanto el área de la curva que está a la derecha de la media es igual a la que está a la izquierda.
- La curva normal decrece uniformemente en ambas direcciones a partir del valor central. Es asintótica respecto al eje horizontal. Las terminaciones de la curva a ambos lados de la media se le llaman *colas*.
- La gráfica típica de una distribución normal se muestra en la figura 25. En esta figura se muestra que la distancia entre el eje de simetría (media) y el punto de inflexión de la curva normal es una desviación estándar. La función de densidad de probabilidad (curva normal) viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (80)$$

**Figura 25. Curva de una distribución normal, con la media y una desviación estándar  $\sigma$**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Para este tipo de distribución la *media*, o *esperanza* o *valor esperado* viene dada por:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_x(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[ -e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = 0 \quad (81)$$

La *varianza* se define como (teniendo en cuenta que el valor esperado es 0):

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \quad (82)$$

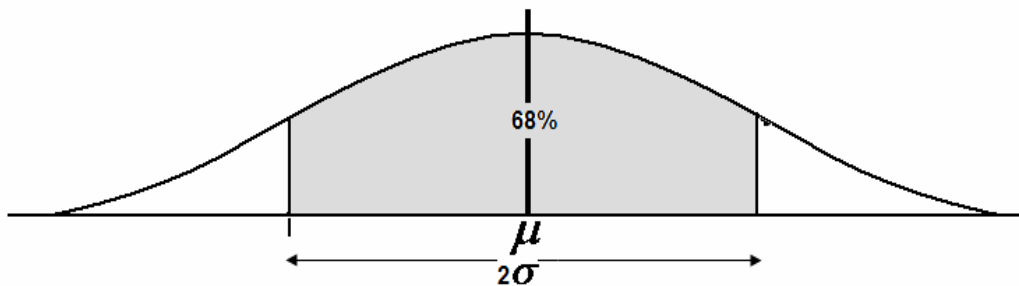
$$Var(X) = E(X^2) - [E(X)]^2 = 1 \quad (83)$$

Dependiendo de los valores que tome la media  $\mu$  y la desviación estándar  $\sigma$  la gráfica será más alargada o achatada, pero en cualquier caso tendrá las mismas condiciones de simetría.

El área total bajo la curva es igual a 1. El área debajo de la curva comprendida entre  $\mu - \sigma$  y  $\mu + \sigma$  es aproximadamente igual a 0,68 del área total; entre  $\mu - 2\sigma$  y

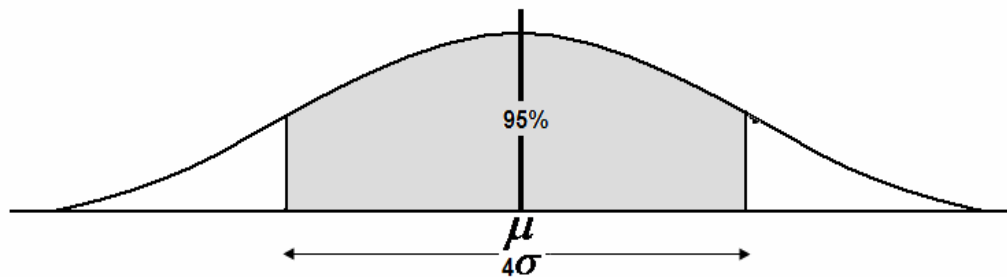
$\mu + 2\sigma$  es aproximadamente igual a 0,95 del área total. Las figuras 26 y 27 ilustran las áreas descritas.

**Figura 26. Porcentaje de área comprendida entre  $\mu - \sigma$  y  $\mu + \sigma$ .**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

**Figura 27. Porcentaje de área comprendida entre  $\mu - 2\sigma$  y  $\mu + 2\sigma$ .**



**FUENTE:** Estadística y Probabilidades; Juan W. Arteaga 2006

Los únicos parámetros necesarios para dibujar un gráfico de la distribución normal son  $\mu$  y  $\sigma$ . Con estos dos parámetros se puede situar la campana (en el punto correspondiente a la media) y cual es su ancho (determinado por la desviación estándar). Cuando se encuentra con una población de observaciones que siga una distribución normal, solo hace falta estimar la media y la desviación estándar para obtener toda la información necesaria acerca de dicha población.

### 2.10.3.2 Distribución Log Normal

La distribución log-normal (utilizada en este trabajo en la aplicación del método de montecarlo) es una distribución de probabilidad de cualquier variable aleatoria con su logaritmo normalmente distribuido (la base de una función logarítmica no es importante ya que,  $\log_a X$  está distribuida normalmente si y solo si  $\log_b X$  está distribuida normalmente). Si  $X$  es una variable aleatoria con una distribución normal, entonces  $\exp(X)$  tiene una distribución log-normal.

Una variable puede ser modelada como log-normal si puede ser considerada como un producto multiplicativo de muchos pequeños factores independientes. Un ejemplo típico es un retorno a largo plazo de una inversión: puede considerarse como un producto de dos retornos diarios.

La distribución log-normal tiende a la función densidad de probabilidades:

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2} \quad (84)$$

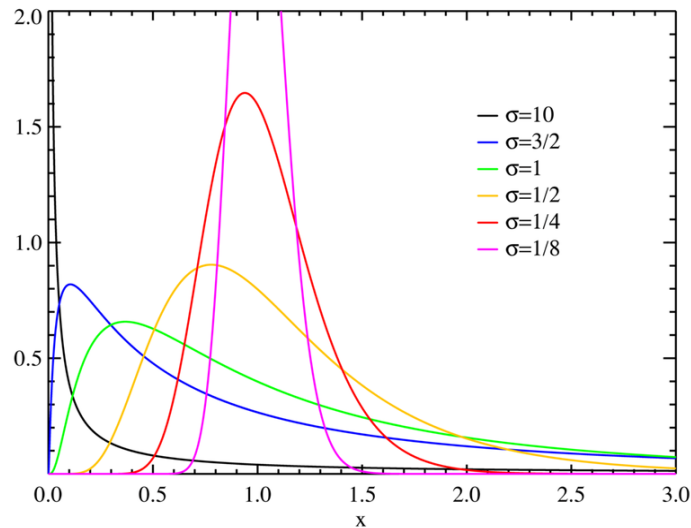
Para  $x > 0$ , donde  $\mu$  y  $\sigma$  son la media y la desviación estándar del logaritmo de la variable. El valor esperado es:

$$E(X) = e^{\mu + \sigma^2 / 2} \quad (85)$$

Y la varianza es:

$$\text{var}(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \quad (86)$$

**Figura 28. Curva De Una Distribución Log Normal**



**FUENTE:** [http://gl.wikipedia.org/wiki/Imaxe:Lognormal\\_distribution\\_PDF.png](http://gl.wikipedia.org/wiki/Imaxe:Lognormal_distribution_PDF.png)

### 2.10.3.3 Distribución Triangular

Este tipo de distribución (utilizada en este trabajo en la aplicación del Método de Montecarlo) se emplea básicamente en aquellos problemas en los cuales se conocen muy pocos datos o ningún dato. La distribución triangular es útil como una aproximación inicial en situaciones para las que no se dispone de datos confiables.

Esta distribución tiene 3 parámetros,  $a$  (límite inferior de la variable);  $m$  (la moda) y  $b$  (el límite superior de la variable).

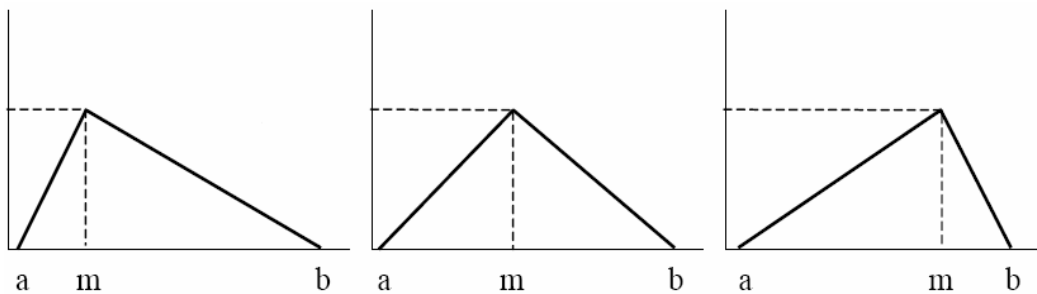
Sea  $X$  una variable aleatoria valuada en  $(a,b) \subset \mathbb{V}$ , se dice que se distribuye según

una distribución triangular si su función de densidad responde a la siguiente expresión:

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(m-a)} & \text{Si } a < x \leq m \\ \frac{2(b-x)}{(b-a)(b-m)} & \text{Si } m \leq x < b \\ 0 & \text{en otro caso} \end{cases} \quad (87)$$

La representación Gráfica de tal función de densidad es:

**Figura 29. Representación Gráfica de la Función de Densidad Triangular dependiendo de que**



$$\frac{a+b}{2} > m, \frac{a+b}{2} = m, \text{ ó } \frac{a+b}{2} < m$$

**FUENTE:** Distribuciones de tipo continuo, Estadística empresarial II, 2005

Por tanto, tal distribución puede ser simétrica o tener una asimetría a la derecha o a la izquierda.

Para este tipo de distribución *la media* se puede hallar así:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^m \frac{2}{(b-a)(m-a)} x(x-a)dx + \int_a^b \frac{2}{(b-a)(b-m)} x(b-x)dx =$$

$$\frac{2}{(b-a)} \left[ \left[ \frac{2m^2 - ma - a^2}{6(m-a)} \right] + \left[ \frac{b^2 + mb - 2m^2}{6(b-m)} \right] \right] = \frac{b^2 + mb - ma - a^2}{3(b-a)} = \frac{a + m + b}{3}$$

(88)

Y la *varianza* es igual a:

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2 - (b-m)(m-a)}{18}$$

(89)

Su función de distribución viene dada por:

$$f(x) = \begin{cases} 0 & \text{Si } x \leq a \\ \frac{(x-a)^2}{(b-a)(m-a)} & \text{Si } a < x \leq m \\ 1 - \frac{(b-x)^2}{(b-a)(b-m)} & \text{Si } m \leq x < b \\ 1 & \text{Si } x \geq b \end{cases}$$

(90)

### 2.10.3.4 Distribución Gamma

Es una distribución de probabilidad continua con dos parámetros  $k$  y  $\lambda$  cuya función de densidad para valores  $x > 0$  es:

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)} \quad (91)$$

Aquí  $e$  es el número  $e$  y  $\Gamma$  es la función gamma dada por:

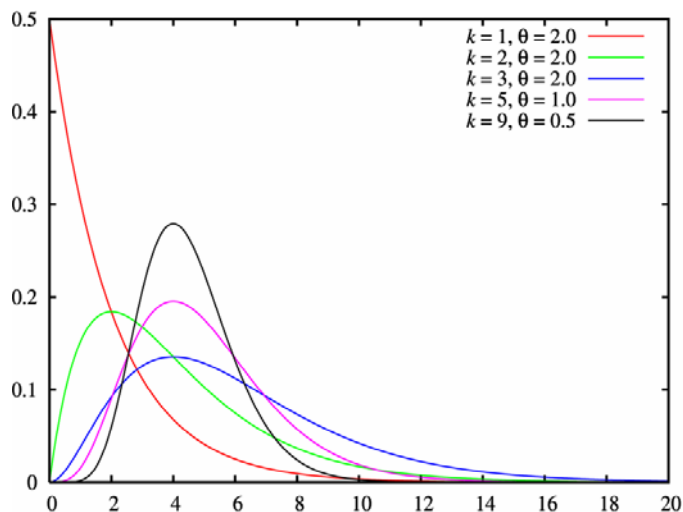
$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (92)$$

El *valor esperado* y la *varianza* de una variable aleatoria  $X$  de distribución gamma son:

$$E[X] = k / \lambda \quad (93)$$

$$V[X] = k / \lambda^2$$

**Figura 30. Función de Distribución Gamma.**



**FUENTE:** [http://es.wikipedia.org/wiki/Distribuci%C3%B3n\\_gamma](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_gamma)

### 2.10.3.5 Distribución Chi Cuadrado

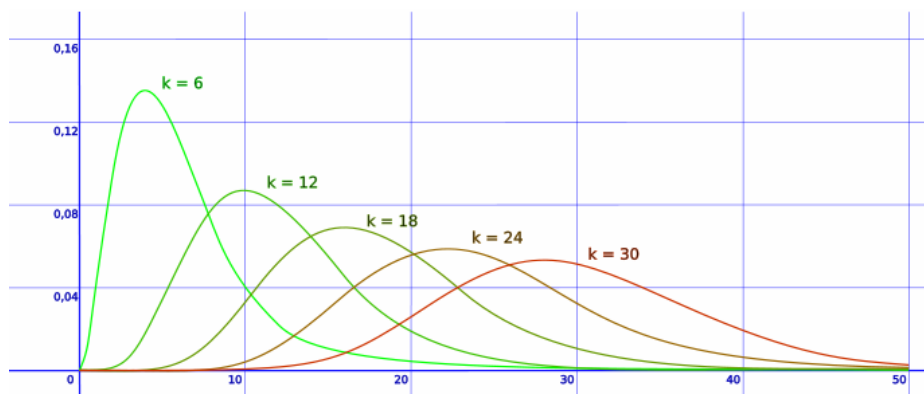
La distribución ji-cuadrado, también denominada ji-cuadrado de Pearson, es una distribución de probabilidad continua con un parámetro  $k$  que representa los grados de libertad de la variable aleatoria:

$$X = Z_1^2 + \dots + Z_k^2 \quad (94)$$

donde  $Z_i$  son variables de distribución normal, de media cero y varianza uno.

Esta distribución se expresa habitualmente  $cX \sim X_k^2$  donde el subíndice  $k$  de  $X_k^2$ , es el número de sumandos y se denomina grados de libertad de la distribución.

**Figura 31. Distribución Chi Cuadrado.**



**FUENTE:** [http://es.wikipedia.org/wiki/Distribuci%C3%B3n\\_chi-cuadrado](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_chi-cuadrado)

Se suele usar la denominada prueba ji-cuadrado como test de independencia y como test de bondad de ajuste.

La función de densidad ji-cuadrado es:

$$f_k(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2} \quad (95)$$

Donde  $x \geq 0$  y  $f_k(x) = 0$  para  $x \leq 0$ .  $\Gamma$  es la función Gamma.

La función de distribución es:

$$F_k(x) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)} \quad (96)$$

Donde  $\gamma(k, z)$  es la función gamma incompleta.

El valor esperado y la varianza de una variable aleatoria X con distribución Chi-cuadrada son:

$$E[X] = k \quad (97)$$

$$V[X] = 2k$$

### 2.10.3.6 Distribución t Student

Es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño. Ésta es la base del popular test de la t de Student para la determinación

de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones.

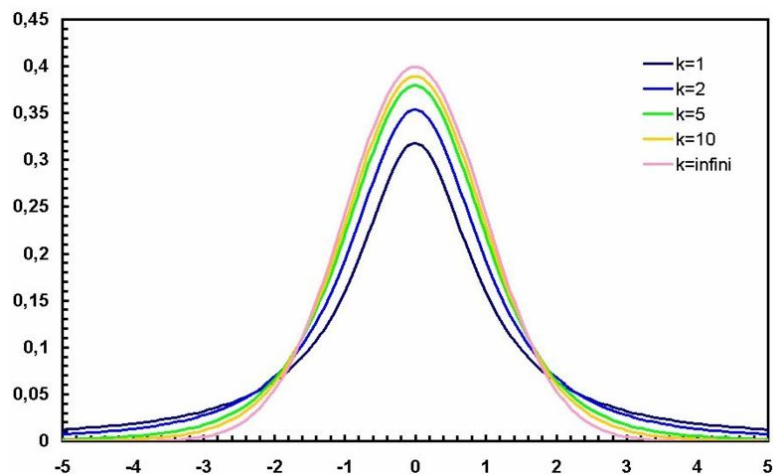
La distribución  $t$  surge, en la mayoría de los estudios estadísticos prácticos, cuando la desviación típica de una población se desconoce y debe ser estimada a partir de los datos de una muestra.

Su función de densidad viene dada por:

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1+t^2/\nu\right)^{-(\nu+1)/2} \quad (98)$$

Con  $\nu$  igual a  $n-1$ . Este parámetro se llama convencionalmente el número de grados de libertad..

**Figura 32. Distribución t Student**



**FUENTE:** [http://es.wikipedia.org/wiki/Distribuci%C3%B3n\\_t\\_de\\_Student](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_t_de_Student)

### **3. SIMULACIÓN MONTECARLO**

Después de haber abordado conceptos estadísticos básicos, se está en capacidad de enfrentar todo lo concerniente al método de Montecarlo como tema central de este trabajo.

#### **3.1 HISTORIA**

El método fue llamado así por el principado de Mónaco por ser "la capital del juego de azar", al tomar una ruleta como un generador simple de números aleatorios. El nombre y el desarrollo sistemático de los métodos de Monte Carlo datan aproximadamente de 1944 con el desarrollo de la computadora. Sin embargo hay varias instancias (aisladas y no desarrolladas) en muchas ocasiones anteriores a 1944.

El uso real de los métodos de Monte Carlo como una herramienta de investigación, proviene del trabajo de la bomba atómica durante la Segunda Guerra Mundial. Este trabajo involucraba la simulación directa de problemas probabilísticos de hidrodinámica concernientes a la difusión aleatoria de neutrones en material de fusión.

Aún en la primera etapa de estas investigaciones, John von Neumann y Stanislaw Ulam refinaron esta simulación directa con ciertas técnicas, en particular "La Ruleta rusa", dando uso a esta técnica ya que los problemas eran complicados para tratarse analíticamente y muy costosos y peligrosos para resolverse por medio de la experimentación física. El nombre que escogieron parece bastante apropiado, puesto que el principio básico es el mismo que se encuentra en operación en el casino de Mónaco; se emplean mecanismos o dispositivos para producir muestras al azar de poblaciones bien definidas, como por ejemplo la ruleta, los dados, o la baraja.

La posibilidad de aplicar el Método de Montecarlo a los problemas determinísticos fue notada por Fermi, Von Neumann y Ulam y fue popularizada por ellos en los años de post-guerra inmediatos. Aproximadamente 1948 Fermi, Metrópoli y Ulam obtuvieron estimaciones con el método de Montecarlo para los autovalores de la ecuación de Schrodinger.

El Dr. Stephen Brush (del Laboratorio de la Radiación a Livermore), quien tiene un interés particular en la historia de la matemática, ha desenterrado un artículo de Kelvin, increíblemente moderno de las técnicas de Montecarlo, en el que hace 60 años aparece una discusión de la ecuación de Boltzmann. Parece completamente correcto y apropiado que Ulam, Von Neumann, y Fermi deben tomar el crédito por redescubrir independientemente no sólo los Métodos Montecarlo sino también de asegurar que sus colegas científicos se dieran cuenta de las posibilidades, potencialidades, y las aplicaciones.

### **3.2 ¿QUÉ ES LA SIMULACIÓN MONTECARLO?**

La Simulación Montecarlo es una técnica que convierte la incertidumbre en variables de entrada a un modelo dentro de distribuciones de probabilidad. Mediante la combinación de las distribuciones y la selección de valores aleatorios recalcula el modelo simulado muchas veces y proporciona la probabilidad de los valores de salida.

La clave de la simulación MC consiste en crear un modelo matemático del sistema, proceso o actividad que se quiere analizar, identificando aquellas variables (*inputs* del modelo) cuyo comportamiento aleatorio determina el comportamiento global del sistema. Una vez identificados dichos *inputs* o variables aleatorias, se lleva a cabo un experimento consistente en (1) *generar* – con ayuda del ordenador- *muestras aleatorias* (valores concretos) para dichos *inputs*, y (2)

analizar el comportamiento del sistema ante los valores generados. Tras repetir  $n$  veces este experimento, se dispondrá de  $n$  observaciones sobre el comportamiento del sistema, lo cual será de utilidad para entender el funcionamiento del mismo –obviamente, el análisis será tanto más preciso cuanto mayor sea el número  $n$  de experimentos que llevemos a cabo.

### 3.2.1 Características básicas.

- La simulación Montecarlo permite que varios inputs sean utilizados al mismo tiempo para crear la distribución de probabilidades de uno o varios outputs (o valores de salida).
- Diferentes tipos de distribuciones de probabilidad pueden ser asignadas a los inputs del modelo. Cuando la distribución no se conoce, el número 1, que representa el mejor ajuste, podría ser elegido.
- La utilización de los números aleatorios caracterizan a la simulación Montecarlo como un método estocástico. Los números aleatorios tienen que ser independientes, no debería existir correlación entre ellos (*características de una variable aleatoria, capítulo 2*).
- La simulación Montecarlo genera los resultados como un rango en lugar de un valor fijo y muestra cuál es la probabilidad de que el resultado final se dé dentro de ese rango.

### 3.3 ALGORITMO PARA IMPLEMENTAR UNA SIMULACIÓN MONTECARLO

El algoritmo de Simulación Monte Carlo Crudo o Puro está fundamentado en la generación de números aleatorios basado en las distribuciones acumuladas de frecuencias:

- Determinar la/s variables aleatorias y sus distribuciones acumuladas(F)
- Generar un número aleatorio uniforme  $\in (0,1)$ .
- Determinar el valor de la Variable Aleatoria para el número aleatorio generado de acuerdo a las clases que tengamos.
- Calcular media, desviación estándar error y realizar el histograma.
- Analizar resultados para distintos tamaños de muestra.

En los pasos dos y tres se debe iterar tantas veces como sea necesario.

Otra opción para trabajar con Monte Carlo, cuando la variable aleatoria no es directamente el resultado de la simulación o existen relaciones entre variables es la siguiente:

- Diseñar el modelo lógico de decisión
- Especificar distribuciones de probabilidad para las variables aleatorias relevantes.
- Incluir posibles dependencias entre variables

- Muestrear valores de las variables aleatorias
- Calcular el resultado del modelo según los valores del muestreo (iteración) y registrar el resultado.
- Repetir el proceso hasta tener una muestra estadísticamente representativa
- Obtener la distribución de frecuencias del resultado de las iteraciones
- Calcular media, desvío.
- Analizar los resultados

### 3.4 LAS MATEMÁTICAS DETRÁS DE LA SIMULACIÓN MONTECARLO

Si se considera una función real valuada  $g(x)$  con una función de frecuencia de probabilidad  $P(x)$  (si  $x$  es discreta) o función de densidad  $f(x)$  (si  $x$  es continua). Se puede definir entonces el valor esperado de  $g(x)$  en términos discretos y continuos respectivamente:

$$E(g(x)) = \sum_{-\infty}^{+\infty} g(x)P(x), \text{ donde } P(x) > 0 \text{ y } \sum_{-\infty}^{+\infty} P(x) = 1$$

$$E(g(x)) = \int_{-\infty}^{+\infty} g(x)f(x)dx, \text{ donde } f(x) > 0 \text{ y } \int_{-\infty}^{+\infty} f(x)dx = 1$$
(99)

Luego, al hacer  $n$  pruebas aleatorias de  $X$  ( $X_1, \dots, X_n$ ), también llamado correr una prueba o hacer correr una simulación, calcular  $g(x_1) \dots g(X_n)$  y encontrar la media de  $g(x)$  de la muestra:

$$\mu g(x) = \frac{1}{n} \sum_{i=1}^n g(x_i), \quad (100)$$

la cual representa el valor final simulado de  $E(g(X))$

Por tanto,  $\mu g_n(x) = \frac{1}{n} \sum_{i=1}^n g(X)$  será el estimador de Monte Carlo de  $E(g(x))$ . Como  $n \rightarrow \infty$ ,  $\mu g_n(x) \rightarrow E(g(X))$ , se está entonces en condiciones de computar la dispersión en torno a la media estimada con la varianza imparcial de  $\mu g_n(x)$ :

$$Var(\mu g_n(X)) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^n (g(x_i) - \mu g_n(x))^2 \quad (101)$$

### 3.5 ALGUNAS DISTRIBUCIONES DE PROBABILIDAD COMÚNMENTE UTILIZADAS EN LA SMC

Previamente, en el capítulo 2, se explicaron las diferentes distribuciones de probabilidad más comunes. A continuación se verá cuando es apropiado usar una u otra distribución de probabilidad en el momento de la aplicación del método de Montecarlo.

- *Distribución Normal o de Gauss*: Distribución continua aplicada en situaciones donde la media y la desviación estándar están dados y la media representa además el valor más probable de la variable. Esta simetría en los entornos de la media no tiene límites.
- *Distribución Log-Normal*: Distribución continua especificada por media y desviación estándar. Es apropiada para variables que se mueven entre cero e infinito, con simetría positiva y con la particularidad de que el logaritmo natural de la variable está distribuido normalmente.
- *Distribución Triangular*: Distribución continua con valores mínimo y máximo fijos. Está limitada por los valores mínimos y máximos y puede ser tanto simétrica (moda=media=mediana) o asimétrica.
- *Distribución Uniforme*: Distribución continua limitada por valores mínimo y máximo conocidos. En contraste con la distribución triangular, la probabilidad de ocurrencia de los valores entre el mínimo y el máximo es la misma.
- *Distribución Exponencial*: Distribución continua utilizada para ilustrar el tiempo entre ocurrencias independientes, con la condición de que la tasa de ocurrencia es conocida.

En la realización de este trabajo se utilizó inicialmente una función de distribución triangular porque como se explicó en el capítulo 2 esta es bastante útil cuando se conocen muy pocos datos (o ningún dato) y como una aproximación preliminar cuando dichos datos no son muy confiables. Por todo lo anterior, y por simplicidad en su aplicación, se optó por trabajar con la distribución triangular. En un trabajo

posterior se realizó una corrida de acuerdo a lo sugerido en la tabla 8, con el único objetivo de realizar un trabajo comparativo.

### **3.6 DISTRIBUCIONES DE PROBABILIDAD ASOCIADAS A CADA UNA DE LAS INCERTIDUMBRES EN UN YACIMIENTO PETROLÍFERO**

Atendiendo a muchas recomendaciones arrojadas de la experiencia, a continuación se hará un breve resumen de cuales podrían ser las distribuciones de probabilidad más apropiadas para tratar cada una de las incertidumbres asociadas a un yacimiento petrolífero.

Las incertidumbres están asociadas a los datos geológicos y/o ingenieriles en áreas perforadas y/o bloques de falla y a la precisión en la medida de tales datos. Estas incluyen:

- Volumen bruto de roca en área perforada.
- Cambio en el espesor neto rentable y relacionado a variaciones espaciales.
- Propiedades petrofísicas y PVT del sistema Roca-Fluido.
- Localización de los contactos entre los fluidos.
- Distribución espacial de permeabilidad.
- Grado de compartimentación del yacimiento.
- Área de drenaje de pozos individuales.
- Eficiencia de recobro de aceite, gas y condensado.

De las variables de encima, algunas son aptas para un análisis estadístico, y estas son identificadas como incertidumbres ingenieriles. Estas incluyen: Porosidad, saturación de agua, propiedades PVT, eficiencias de recobro, cambios en el espesor neto rentable y áreas de drenaje de pozos individuales.

Inicialmente es difícil definir las funciones de densidad de probabilidad de los parámetros de encima, por esto, lo siguiente puede ser dicho acerca del tipo de función de densidad de probabilidad que puede ser utilizada.

**Tabla 8. Distribuciones de Probabilidad asociadas a cada una de las incertidumbres en un yacimiento petrolífero.**

Parámetro	Tipo de distribución de frecuencia
Área-Acres	Una distribución Log Normal. Los parámetros requeridos son la media y la desviación estándar.
Espesor Neto	La distribución Log Normal es la mejor aproximación, típicamente exhibe inclinación positiva; puede existir covarianza con porosidad límite y saturación de agua intersticial. Ingresar la media y la desviación estándar.
Porosidad	Típicamente puede tratarse con una distribución Normal o Gaussiana. Exhibe ocasionalmente una tendencia al sesgo positivo o log normal y una covarianza con la saturación de agua inicial. Los parámetros aquí son también la media y la desviación estándar.
Saturación de agua intersticial	Típicamente exhibe un ligero sesgo negativo aproximándose a una distribución simétrica.
Permeabilidad	Una distribución Log Normal es la mejor aproximación, pero la naturaleza de la distribución es controlada por el tipo de ambiente deposicional, y/o diagénesis post-deposicional; típicamente presenta una covarianza con la porosidad.

Parámetro	Tipo de distribución de frecuencia
Viscosidad	Recomendado usar una distribución triangular con los parámetros: Mínimo, más probable y máximo.
Factor volumétrico de formación del aceite	Recomendado usar una distribución triangular con los parámetros: Mínimo, más probable y máximo.
Eficiencia de recobro	La mejor aproximación es una distribución Log Normal, típicamente con sesgo positivo; puede exhibir covarianza con porosidad, saturación de agua irreducible, permeabilidad y espesor neto dependiendo del mecanismo de empuje.

**FUENTE:** Estimation And Classification Of Reserves Of Crude Oil, Natural Gas, And Condensate”: Chapman Cronquist, Society Of Petroleum Engineers, Richardson, Texas 2001

La covarianza mencionada en los comentarios de arriba sobre los tipos de distribuciones de frecuencia, es una medida estadística del grado de correlación entre los conjunto de datos. Las variables estocásticas que exhiben un alto grado de covarianza no pueden ser independientes, por tanto deben tomarse cuidados en el tratamiento de ellos como variables independientes en los cálculos probabilísticos.

### **3.7 HERRAMIENTAS SOFTWARE MÁS UTILIZADAS PARA EL ANÁLISIS DE UN PROBLEMA MEDIANTE SIMULACIÓN MONTECARLO**

Hoy día existen una gran cantidad de herramientas software para llevar a cabo un análisis utilizando la Simulación Montecarlo. Estas van desde las más

convencionales como las hojas de cálculo hasta la gran variedad de programas que llevan a cabo estudios de riesgos y pronósticos.

En la realización de este trabajo la aplicación del método Montecarlo se llevó a cabo utilizando una herramienta software al alcance de muchos usuarios como lo es la Hoja de cálculo de Excel adjunto al paquete de Microsoft office; por otro lado, y con el fin de tener un soporte más amplio se utilizó un programa llamado Lumenaut (que funciona como un complemento de Excel) el cual está destinado a quitar la incertidumbre mediante el examen de predicciones y superar limitaciones que tienen las hojas de cálculo.

### **3.7.1 Simulación Montecarlo En Excel**

Son muchos los autores que han apostado por utilizar hojas de cálculo para realizar simulación Montecarlo. La potencia de las hojas de cálculo reside en su universalidad, en su facilidad de uso, en su capacidad para recalcular valores y, sobre todo, en las posibilidades que ofrece con respecto al análisis de escenarios (“*what-if analysis*”). Las últimas versiones de Excel incorporan, además, un lenguaje de programación propio, el *Visual Basic for Applications*, con el cual es posible crear auténticas aplicaciones de simulación destinadas al usuario final.

#### **3.7.1.1 La función ALEATORIO () de Excel**

Las hojas de cálculo como Excel (y cualquier lenguaje de programación estándar) son capaces de generar números pseudo-aleatorios provenientes de una distribución uniforme entre el 0 y el 1. Este tipo de números pseudo-aleatorios son los elementos básicos a partir de los cuales se desarrolla cualquier simulación por ordenador.

En Excel, es posible obtener un número pseudo-aleatorio -proveniente de una distribución uniforme entre el 0 y el 1 usando la función ALEATORIO:

**Figura 33. Función ALEATORIO de Excel**



The image shows a screenshot of an Excel spreadsheet. The active cell is A1, and the formula bar displays the function `=ALEATORIO()`. The spreadsheet grid shows columns A, B, and C, and rows 1 and 2. Cell A1 contains the value 0,6645232.

	A	B	C
1	0,6645232		
2			

**FUENTE:** SIMULACIÓN DE MONTE CARLO CON EXCEL, Javier Faulín, Ángel A. Juan, UOC 2006

1. Cada vez que se usa la función ALEATORIO, cualquier número real entre 0 y 1 tiene la misma probabilidad de ser generado (de ahí el nombre de distribución uniforme).
2. Los diferentes números generados son estadísticamente independientes unos de otros (es decir, el valor del número generado en un momento dado no depende de los generados con anterioridad).

La función ALEATORIO es una función volátil de Excel. Esto significa que cada vez que se pulsa la tecla F9 o se cambie alguno de los *inputs* del modelo, todas las celdas donde aparezca la función ALEATORIO serán recalculadas de forma automática. Se pueden encontrar ejemplos del uso de ALEATORIO en el propio menú de ayuda de Excel. En el Anexo A se podrá ver la manera de obtener números aleatorios distribuidos según una forma determinada ya sea recurriendo al complemento de *Análisis de Datos* o Utilizando la Función *ALEATORIO*. En general en este Anexo se podrá encontrar además todo lo relacionado con las funciones de densidad de probabilidad al momento de la generación de números

aleatorios y el análisis de datos llevado a cabo posteriormente a la aplicación del método.

### **3.7.2 Simulación Montecarlo en Lumenaut**

El software Lumenaut como complemento de Excel acentúa en un análisis de riesgo usando simulación Montecarlo, un paquete de análisis mediante un árbol de decisión y un paquete estadístico paramétrico y no paramétrico, proporcionando una poderosa configuración de herramientas para la ingeniería o cualquier otro tipo de negocio.

Lumenaut es un software de fácil uso en el ámbito profesional, comercial y académico a través de todo el mundo.

Para un análisis Montecarlo, Lumenaut proporciona un rango de herramientas que le permiten al usuario interactuar fácil y rápidamente al momento de realizar una simulación Montecarlo en Excel.

Lumenaut es fácil de aprender y fácil de usar, y ha sido desarrollado teniendo en mente una amplia gama de usuarios y de usos de hojas de cálculo. No es necesario tener un conocimiento altamente avanzado de estadísticas o de computación para usar Lumenaut en su máximo potencial. Solamente necesita conocimientos básicos del funcionamiento de su computadora y la capacidad de crear modelos de hoja de cálculo. (Todas las características y ejemplos de uso de este software se pueden encontrar en <http://www.lumenaut.com/montecarlo.htm>)

#### **4. METODOLOGÍA PARA LA APLICACIÓN DE LA SIMULACIÓN MONTECARLO EN LA DETERMINACIÓN DE LOS POTENCIALES DE POZOS DESVIADOS Y HORIZONTALES.**

Después de haber abarcado toda la base teórica necesaria para la utilización del método de Montecarlo, se está en capacidad de aplicar este para determinar los índices de productividad y posteriormente los IPR en caso de que poseamos una data de presiones. En el desarrollo de este capítulo se realizarán las siguientes acciones:

- En primer lugar se calculará los índices de producción para los pozos desviados (en este caso aplicando la metodología al campo colorado, pozos 38 el cual posee las Zonas B y C) usando la hoja de cálculo Excel. En esta primera aproximación se trabajará con una función de densidad de probabilidad triangular por las razones explicadas anteriormente (2.10.3.3). Luego de calcular el índice de producción se calculará el IPR
- La metodología explicada para hallar el índice de producción mediante simulación Montecarlo en pozos desviados y usando una hoja de cálculo, se hará extensiva para pozos horizontales, donde teniendo en cuenta unos datos de un pozo horizontal (artículo SPE 107669-MS-P) se hallará el índice de producción. No se calculará IPR pues no se contaba con una data de presiones. Para este caso también se utilizará una función de densidad de probabilidad triangular.
- Se realizará una corrida en el Software Lumenaut para el mismo pozo horizontal referenciado en el ítem anterior, donde también se anotará una metodología. Veremos que los resultados obtenidos con la hoja de cálculo y con el software no difieren significativamente.

- Como una forma de explorar varias posibilidades, y basándose en lo propuesto por la experiencia de algunos autores (“Estimación y clasificación de reservas de aceite crudo, gas natural, y condensado” Chapman Cronquist, Sociedad de ingenieros de petróleos, Richardson, Texas 2001.) se realizó una nueva corrida usando una hoja de cálculo con la finalidad de comparar los resultados. En esta parte, la corrida se realizó para la Zona B del pozo colorado 38 usando una función de densidad de probabilidad log normal en la generación de valores para la permeabilidad y para el espesor neto; para la viscosidad y el factor volumétrico de formación se usó la función de distribución triangular (ver tabla 8)

#### **4.1 METODOLOGIA PARA POZOS DESVIADOS. CAMPO COLORADO POZO COLORADO 38. IMPLEMENTACIÓN EN HOJA DE CÁLCULO**

Generalmente en la industria petrolífera, los pozos perforados no son cien por ciento verticales, y en la mayoría de los casos poseen una desviación con respecto a la línea vertical.

El pozo Colorado 38 está compuesto por varias arenas productoras y por tanto la presente metodología debe ser aplicada a cada intervalo productor para el cálculo del IP. Estos intervalos productores se obtuvieron de la interpretación de registros hecha a los pozos del campo, con base en la información disponible.

El índice de productividad para pozos verticales se puede calcular mediante la ecuación:

$$IP = \frac{kh}{141.2 * \mu_o B_o * \left[ \ln\left(\frac{R_e}{r_w}\right) - \frac{3}{4} \right]} \quad (5)$$

Donde:

- $h$  es el espesor neto promedio de petróleo del yacimiento;
- $\mu$  es la viscosidad del petróleo;
- $B_o$  es el factor volumétrico de formación del petróleo;
- $r_w$  es el radio del pozo;
- $R_e$  es el radio de drenaje del pozo;
- $k$  es la permeabilidad efectiva del yacimiento;

El objetivo es hallar el IP en pozos verticales usando Simulación Monte Carlo. En este orden es importante destacar la naturaleza variable de las propiedades de un yacimiento petrolífero, es decir, la idea de que las propiedades en los yacimientos no son constantes, y por tanto la estimación del índice de productividad generará incertidumbre y en ocasiones conllevará a predecir un valor no acertado o que está muy desviado del valor esperado. La simulación Monte Carlo viene en este caso a ser una opción bastante viable ya que, aunque no provee un resultado exacto, es una buena aproximación con un nivel de incertidumbre aceptable.

Las variables de yacimiento como son porosidad, permeabilidad, espesores netos productores, viscosidades y factores volumétricos varían su valor en la gran mayoría de los yacimientos.

La simulación Monte Carlo puede ser aplicada en dos casos: cuando se tiene muy poca información de un pozo en un yacimiento, y/o cuando las variables en el pozo varían de tal forma que al calcular un valor promedio de estas, el resultado obtenido llevará consigo un error bastante elevado.

Teniendo en cuenta lo anterior, y para calcular los potenciales de producción, la simulación Montecarlo sería válida de la siguiente manera:

Caso i: En este caso que es donde se cuenta con la data necesaria del pozo, pero las propiedades  $(k, \Phi, h, \mu)$  varían de tal manera que al sacar un promedio de éstas para el cálculo de los índices de productividad, se enfrenta un error considerable al calcular un valor sobre estimado o subestimado del índice de productividad IP.

Caso ii: Cuando la información del pozo es escasa, o no se cuenta con el valor de alguna variable  $(k, \Phi, h, \mu)$ , la mejor aproximación es buscar otro pozo que se encuentre en la misma formación productora y mediante simulación Monte Carlo, sacar el valor esperado de la variable o variables no conocidas y luego aplicar nuevamente simulación Monte Carlo para predecir el valor de IP del pozo del cual los datos son escasos o muy pocos.

En el caso del pozo Colorado 38, se contaba con información (obtenida de registros) de las permeabilidades, porosidades, espesores, viscosidades, y Factores volumétricos. Se puede afirmar que se cuenta con información suficiente del pozo, pero el problema radica en que cada variable involucrada en el cálculo del potencial de producción del pozo varía de tal forma que si se sacara un promedio para calcular el IP se podría conllevar a un error. En este punto cabe destacar que según la información que se obtuvo del pozo, este posee 5 intervalos productores principales en la zona B y 5 intervalos productores principales en la Zona C. Como una restricción a la aplicación de la metodología es importante decir que la simulación Monte Carlo debe ser llevada a cabo para cada arena por separado, o en su defecto si existiera una sola arena que produjera todos los HCs y en la cual las propiedades variarían.

#### 4.1.1. Pasos para la aplicación de la metodología.

Inicialmente se realizará la aplicación para la zona B. La zona C se tratará posterior a esto. Los cinco intervalos productores que posee el pozo Colorado 38 en la zona B son los siguientes:

**Tabla 9. Intervalos Productores Para El Pozo Colorado 38 Zona B**

### ZONA B

INTERVALOS DE PRODUCCION POZO COL-38 ZONA B
INTERVALO 1812-1829 ft
INTERVALO 1862-1878 ft
INTERVALO 1905-1909 ft
INTERVALO 1938-1943 ft
INTERVALO 2110-2118 ft

**FUENTE:** El autor.

- El primer paso a seguir en la aplicación de la metodología es definir que tipo de función de distribución se va a usar para aplicar la simulación Monte Carlo en el cálculo del IP. Para el caso estudio (Pozo Colorado 38) se definió que la función de distribución triangular sería una buena primera aproximación para el modelo. Esta función de distribución triangular tiene en cuenta 3 valores principales (como se explicó en el capítulo 2, el valor mínimo de las variables, el máximo y el valor esperado o moda). Por tanto, del set de datos se procedió a sacar estos valores para cada variable en cada uno de los intervalos productores, de esta forma:

**Tabla 10. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1812-1829 Ft) Colorado 38, Zona B.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	10,85	<b>Espesor del intervalo productor(ft)</b>	Mínimo	4
	Probable	19,88		Probable	12,6
	Máximo	23,9		Máximo	16
<b>Viscosidad (Cp)</b>	Mínimo	0,9137	<b>Bo (Bls/STB)</b>	Mínimo	1,1875
	Probable	0,9140		Probable	1,1890
	Máximo	0,9144		Máximo	1,195

**FUENTE:** El autor.

La simulación Monte Carlo nos permite, mediante combinación aleatoria de variables, obtener el valor esperado del índice de productividad.

- El siguiente paso es generar una muestra aleatoria de cada variable involucrada en el cálculo del IP. Normalmente Microsoft Excel permite generar una muestra aleatoria de algunas funciones de distribución que vienen ya en el paquete. En este caso la función de distribución triangular no venía en el paquete de Excel pero la forma de su aplicación fue encontrada en la bibliografía (Ver Anexo A). La fórmula es:

$$F(x) = c + (a + U * (b - a) - c) * \sqrt{U} \quad (102)$$

Donde:

- a: Valor mínimo de la variable evaluada.
- b: Valor más probable de la variable evaluada.
- c: Valor máximo de la variable evaluada.
- U: Función aleatorio () de Excel.

Se generan las muestras aleatorias de cada variable involucradas en el cálculo del IP (permeabilidad, espesor, viscosidad y factor volumétrico) con la ecuación (102). Cada valor aleatorio que toman estas variables se introducen en la ecuación para el cálculo del IP (ecuación 5) tal como se muestra en la figura (34).

Se comenzará con el *INTERVALO 1812-1829 ft*, para explicar la metodología y luego se extenderá a los otros intervalos.

**Figura 34. Calculo De Ip Mediante La Generación De Muestras Aleatorias, Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.**

Datos de entrada						
K(mD)	Minimo		10,86	Espesor del intervalo productor(ft)	Minimo	4
	Probable		19,88		Probable	12,6
	Maximo		23,9		Maximo	16
Viscosidad(Cp)	Minimo		0,913722715	Bo(Bls/STB)	Minimo	1,187589305
	Probable		0,914083917		Probable	1,187590332
	Maximo		0,914445054		Maximo	1,187591358
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
			INTERVALO 1812-1829			
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
17,69991775	15,42900358	0,913961085	1,187590251	0,347185733		

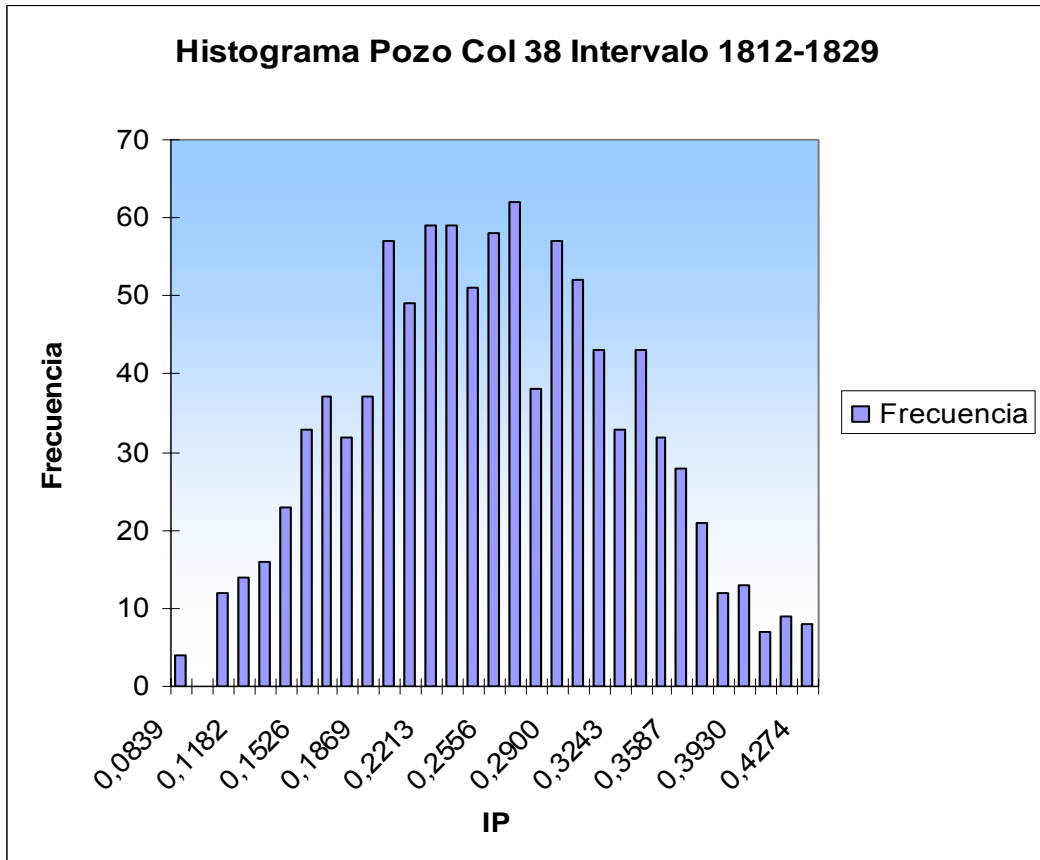
**FUENTE:** El autor.

- El paso siguiente es escoger las celdas a las cuales se le ha asignado el valor aleatorio y desplazarlas hacia abajo tantas veces como iteraciones se quiera. Cada valor hacia abajo nos generará un valor aleatorio del IP que resultará de la combinación al azar de las muestras de permeabilidad, el espesor, el factor volumétrico y la viscosidad.

En este caso se hizo una corrida de 1000 iteraciones para cada intervalo productor, los resultados para cada intervalo se anexarán a continuación.



**Figura 36. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1812-1829, Pozo Colorado 38, Zona B.**



**FUENTE:** El autor.

En este histograma se presentan las marcas o punto medio de clase (Capítulo 2) en donde la muestra que más se repite (frecuencia más alta) será el valor esperado del IP para esta zona. La tabla de datos se anexa para poder leer el valor que más se repite.

**Tabla 11. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.**

IP	Frecuencia	IP	Frecuencia
0,0839	4	0,2671	62
0,0953	0	0,2785	38
0,1068	12	0,2900	57
0,1182	14	0,3014	52
0,1297	16	0,3129	43
0,1411	23	0,3243	33
0,1526	33	0,3358	43
0,1640	37	0,3472	32
0,1755	32	0,3587	28
0,1869	37	0,3701	21
0,1984	57	0,3816	12
0,2098	49	0,3930	13
0,2213	59	0,4045	7
0,2327	59	0,4159	9
0,2442	51	0,4274	8
0,2556	58	y mayor...	2

**FUENTE:** El autor.

Del análisis del histograma el *IP esperado para este intervalo es 0.2671bbl/psi* ya que es el valor que mas se repite al hacer la simulación.

El mismo procedimiento se aplicó a los demás intervalos productores. A continuación se presentan los datos y los resultados:

- ✓ INTERVALO 1862-1878 ft.
- ✓

**Tabla 12. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1862-1878 Ft) Pozo Colorado 38, Zona B.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	22	<b>Espesor del intervalo productor(ft)</b>	Mínimo	4
	Probable	27		Probable	16
	Máximo	30		Máximo	16
<b>Viscosidad(Cp)</b>	Mínimo	0,912064083	<b>Bo(BIs/STB)</b>	Mínimo	1,187595405
	Probable	0,912181104		Probable	1,187595737
	Máximo	0,91229812		Máximo	1,187596069

**FUENTE:** El autor.



**Tabla 13. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B.**

<i>IP</i>	<i>Frecuencia</i>	<i>IP</i>	<i>Frecuencia</i>
0,1540	2	0,3801	44
0,1682	2	0,3943	42
0,1823	8	0,4084	54
0,1964	5	0,4225	61
0,2105	13	0,4366	51
0,2247	10	<b>0,4508</b>	<b>62</b>
0,2388	18	0,4649	50
0,2529	18	0,4790	54
0,2671	31	0,4932	50
0,2812	18	0,5073	52
0,2953	29	0,5214	45
0,3095	28	0,5356	42
0,3236	32	0,5497	25
0,3377	31	0,5638	22
0,3519	49	0,5780	10
0,3660	36	y mayor...	7

**FUENTE:** El autor

Del análisis del histograma el *IP esperado para este intervalo es 0.4508 bbl/psi* ya que es el valor que mas se repite al hacer la simulación.

✓ INTERVALO 1905-1909 ft

**Tabla 14. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1905-1909 Ft) Pozo Colorado 38, Zona B.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	14	<b>Espesor del intervalo productor(ft)</b>	Mínimo	3
	Probable	16		Probable	4
	Máximo	17		Máximo	5
<b>Viscosidad(Cp)</b>	Mínimo	0,910318126	<b>Bo(Bls/STB)</b>	Mínimo	1,187600538
	Probable	0,910403348		Probable	1,18760078
	Máximo	0,910488566		Máximo	1,187601021

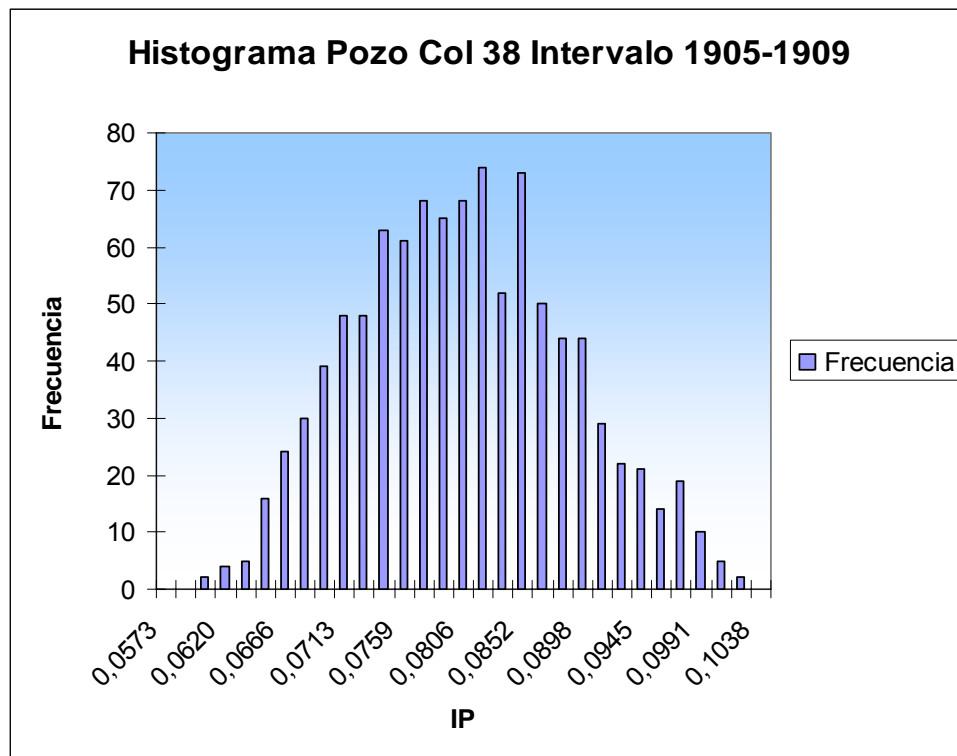
**FUENTE:** El autor

**Figura 39. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B.**

Datos de entrada						
K(mD)	Minimo		14	Espesor del intervalo productor(ft)	Minimo	3
	Probable		16		Probable	4
	Maximo		17		Maximo	5
Viscosidad(Cp)	Minimo		0,910318126	Bo(Bls/STB)	Minimo	1,187600538
	Probable		0,910403348		Probable	1,18760078
	Maximo		0,910488566		Maximo	1,187601021
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 1905-1909						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
16,39014414	4,1264865	0,910455318	1,187600857	0,086313978		
14,75151523	4,473495422	0,91040831	1,187600771	0,084221702		
15,22460508	3,990654159	0,910445148	1,187600852	0,0775377		
15,35774694	4,000613792	0,910339199	1,187600743	0,07842012		
15,16615455	3,647812464	0,910415798	1,187600618	0,070606527		
14,77636792	4,021608274	0,910382719	1,187600714	0,075843799		
15,67288336	4,408408425	0,91042461	1,18760072	0,088178632		
15,23266188	3,611256395	0,910455029	1,187600773	0,070202443		

FUENTE: El autor

**Figura 40. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B.**



FUENTE: El autor

**Tabla 15. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B**

IP	Frecuencia	IP	Frecuencia
0,0573	0	0,0821	74
0,0589	0	0,0837	52
0,0604	2	0,0852	73
0,0620	4	0,0868	50
0,0635	5	0,0883	44
0,0651	16	0,0898	44
0,0666	24	0,0914	29
0,0682	30	0,0929	22
0,0697	39	0,0945	21
0,0713	48	0,0960	14
0,0728	48	0,0976	19
0,0744	63	0,0991	10
0,0759	61	0,1007	5
0,0775	68	0,1022	2
0,0790	65	0,1038	0
0,0806	68	y mayor...	1

FUENTE: El autor

Del análisis del histograma el *IP esperado para este intervalo es 0.0821 bbl/psi* ya que es el valor que mas se repite al hacer la simulación.

✓ INTERVALO 1938-1943 ft

**Tabla 16. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (1938-1943 Ft) Pozo Colorado 38, Zona B.**

Datos de entrada			Datos de entrada		
K(mD)	Mínimo	14	Espesor del intervalo productor(ft)	Mínimo	4
	Probable	17		Probable	5
	Máximo	20		Máximo	6
Viscosidad(Cp)	Mínimo	0,908889821	Bo(Bls/STB)	Mínimo	1,187604524
	Probable	0,908985799		Probable	1,187604796
	Máximo	0,909081774		Máximo	1,187605068

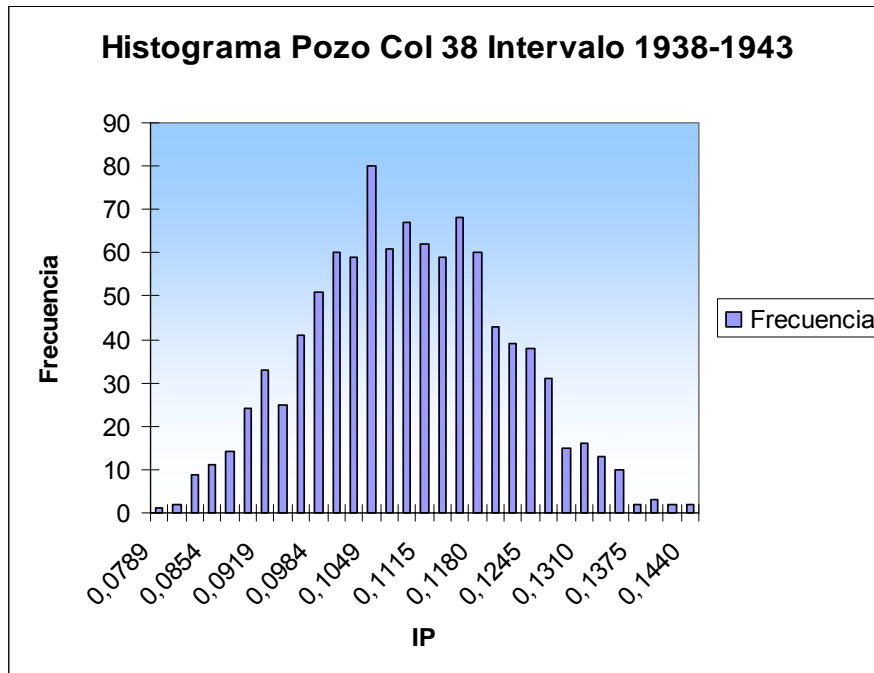
FUENTE: El autor

**Figura 41. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B.**

Datos de entrada						
K(mD)	Minimo		14	Espesor del intervalo productor(ft)	Minimo	4
	Probable		17		Probable	5
	Maximo		20		Maximo	6
Viscosidad(Cp)	Minimo		0,908888621	Bo(Bls/STB)	Minimo	1,187604524
	Probable		0,908985799		Probable	1,187604796
	Maximo		0,909081774		Maximo	1,187605068
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 1938-1943						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
18,2665137	5,158095628	0,909021891	1,187604671	0,120564968		
17,36274561	5,064835171	0,909027063	1,187604627	0,112404085		
15,94297739	4,559990834	0,909032852	1,187604778	0,092924227		
18,03424585	4,962027578	0,908983318	1,187604756	0,114366926		
16,63447952	4,92667506	0,909023422	1,187604777	0,104752211		
19,00057871	4,951486246	0,908908895	1,187604737	0,12026997		
17,2319577	4,864379682	0,908972116	1,187604797	0,107148637		
17,3704022	5,743702006	0,909058751	1,187604918	0,127521938		

FUENTE: El autor

**Figura 42. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B.**



FUENTE: El autor

**Tabla 17. Marcas De Clase (Ip) Y Frecuencia, Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B**

<i>IP</i>	<i>Frecuencia</i>	<i>IP</i>	<i>Frecuencia</i>
0,0789	1	0,1136	59
0,0811	2	0,1158	68
0,0832	9	0,1180	60
0,0854	11	0,1201	43
0,0876	14	0,1223	39
0,0897	24	0,1245	38
0,0919	33	0,1267	31
0,0941	25	0,1288	15
0,0963	41	0,1310	16
0,0984	51	0,1332	13
0,1006	60	0,1353	10
0,1028	59	0,1375	2
<b>0,1049</b>	<b>80</b>	0,1397	3
0,1071	61	0,1419	2
0,1093	67	0,1440	2
0,1115	62	y mayor...	0

**FUENTE:** El autor

Del análisis del histograma el *IP esperado para este intervalo es 0.1049 bbl/psi* ya que es el valor que mas se repite al hacer la simulación.

✓ INTERVALO 2110-2118 ft

**Tabla 18. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (2110-2118 Ft) Pozo Colorado 38, Zona B.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	13	<b>Espesor del intervalo productor(ft)</b>	Mínimo	7
	Probable	20		Probable	8
	Máximo	25		Máximo	9
<b>Viscosidad(Cp)</b>	Mínimo	0,90138152	<b>Bo(Bls/STB)</b>	Mínimo	1,1876253
	Probable	0,901553129		Probable	1,187625783
	Máximo	0,901724724		Máximo	1,187626266

**FUENTE:** El autor



**Tabla 19. Marcas De Clase (Ip) Y Frecuencia, Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B**

IP (bbl/psi)	Frecuencia	IP (bbl/psi)	Frecuencia
0,1276	0	0,1973	57
0,1320	1	0,2016	57
0,1363	3	0,2060	55
0,1407	7	0,2103	63
0,1450	10	0,2147	67
0,1494	12	0,2190	55
0,1537	18	0,2234	57
0,1581	16	0,2277	41
0,1624	26	0,2321	39
0,1668	31	0,2365	30
0,1711	37	0,2408	24
0,1755	39	0,2452	20
0,1799	39	0,2495	19
0,1842	39	0,2539	12
0,1886	51	0,2582	4
0,1929	59	y mayor...	13

**FUENTE:** El autor

Del análisis del histograma el *IP esperado para este intervalo es 0.2147 bbl/psi* ya que es el valor que mas se repite al hacer la simulación.

Finalmente se calculó el IPR de la zona B del pozo colorado 38. Debemos tener en cuenta que además de este cálculo se debe hacer otro cálculo de IP para la zona C, para así saber cual seria el aporte total del pozo. Para el cálculo del IPR la ecuación usada fue la de Vogel.

$$IPR = \frac{IP * P_{zona B}}{1.8} \left( 1 - 0.2 * \left[ \frac{P_{wf}}{P_{zona B}} \right] - 0.8 * \left[ \frac{P_{wf}}{P_{zona B}} \right]^2 \right) \quad (103)$$

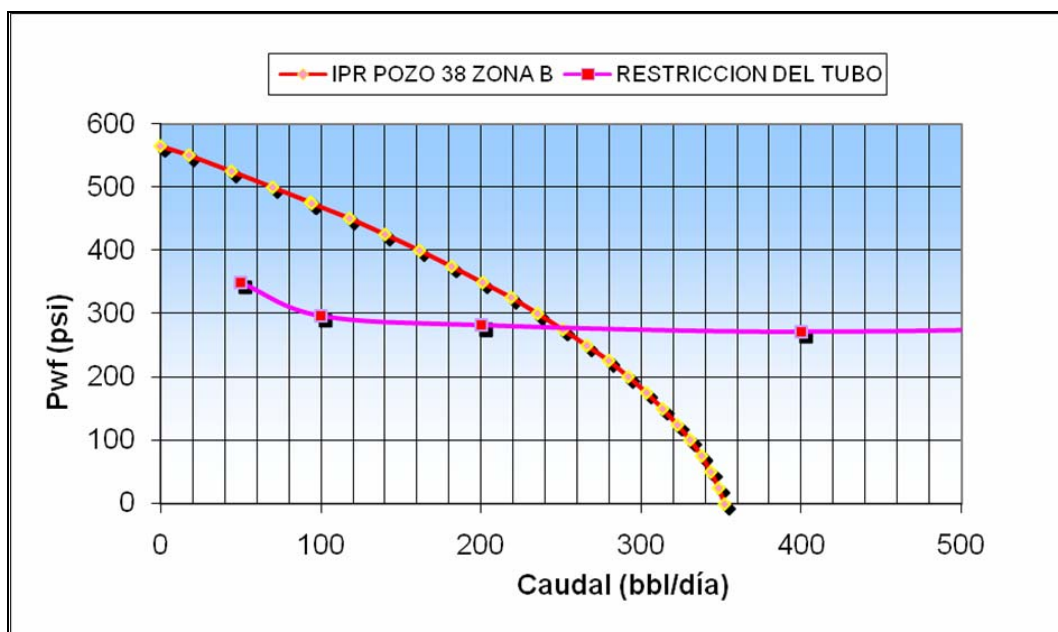
**Tabla 20. Índice De Productividad, Presión De Fondo Fluyendo Y Presión De Yacimiento De La Zona B, Pozo Colorado 38**

<b>INDICE DE PRODUCTIVIDAD ZONA B (IP)</b>	
1,1196	Bbl/PSI
<b>PRESIONES</b>	
<b>Pwf</b>	
475,9	PSI
<b>P ZONA B</b>	
566	PSI
<b>P-Pwf</b>	
90,1	PSI

**FUENTE:** El autor

En este caso, se hace variar la Pwf desde la presión de yacimiento hasta la mínima presión que en este caso se considera 0. Se tiene en cuenta, además, las restricciones de la tubería para el cálculo del caudal esperado del pozo. Estas restricciones muy seguramente disminuirán el potencial de flujo de las arenas productoras.

**Figura 45. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel. Restricciones Del Tubo. Pozo Colorado 38, Zona B**



**FUENTE:** El autor

**Tabla 21. Presiones De Fondo Fluyendo Como Datos De Entrada Para La Ecuación De Vogel Y Obtención De Caudales (Ipr) Pozo Colorado 38, Zona B.**

Pwf (psi)	Qo (bbl/día)
566	0
550	17,6887
525	44,4261
500	70,0646
475,9	93,7397
475	94,6041
450	118,0447
425	140,3864
400	161,6291
375	181,7728
350	200,8176
325	218,7634
300	235,6103
275	251,3583
250	266,0072
225	279,5572
200	292,0083
175	303,3605
150	313,6137
125	322,7679
100	330,8233
75	337,7796
50	343,6370
25	348,3955
0	352,0549

**FUENTE:** El autor

El caudal de la zona B lo obtenemos de la misma ecuación de Vogel:

$$IPR = \frac{IP * P_{zona B}}{1.8} \left( 1 - 0.2 * \left[ \frac{P_{wf}}{P_{zona B}} \right] - 0.8 * \left[ \frac{P_{wf}}{P_{zona B}} \right]^2 \right) \quad (103)$$

Solo se debe reemplazar los valores de la P en la zona B, y la Pwf a la cual esta produciendo el pozo actualmente (Tabla 21)

**Tabla 22. Caudal Esperado De La Zona B A La Presión A La Cual Se Encuentra Produciendo**

CAUDAL ZONA B
94 Bls/día

**FUENTE:** El autor

El anterior análisis fue hecho tomando como base para el cálculo el valor esperado del IP (también conocido como el percentil 50), pero para un resultado con una incertidumbre mucho menor y que dará un numero mas representativo será el valor del P10 del conjunto de datos analizados, que en este caso representan el índice de productividad (IP), por lo tanto, se procederá a sacar el valor del P10 intervalo a intervalo para hacer el análisis un poco mas profundo.

Estos valores fueron obtenidos con la herramienta de Excel llamada análisis de datos, en la opción Histograma.

✓ INTERVALO 1812-1829 ft

**Tabla 23. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1812-1829 Ft, Pozo Colorado 38, Zona B.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0691	0	0,00%	0,2475	70	43,8%
0,0810	0	0,00%	0,2594	62	50,0%
0,0929	3	0,30%	0,2713	59	55,8%
0,1047	6	0,90%	0,2832	55	61,3%
0,1166	7	1,60%	0,2951	51	66,4%
0,1285	12	2,80%	0,2951	51	66,4%
0,1404	20	4,80%	0,3070	58	72,2%
0,1523	23	7,09%	0,3189	60	78,2%
0,1642	27	10,00%	0,3308	51	83,3%
0,1761	16	11,39%	0,3427	48	90,0%
0,1880	48	16,2%	0,3546	41	92,2%
0,1999	44	20,6%	0,3665	20	94,2%
0,2118	49	25,5%	0,3784	24	96,6%
0,2237	58	31,3%	0,3903	10	97,6%
0,2356	55	36,8%	0,4022	11	98,7%
			0,4141	8	99,5%

**FUENTE:** El autor

Por tanto, en esta siguiente parte del análisis se usará el valor del P10, para que el resultado obtenido del IP tenga una mayor confiabilidad.

*IP P10 intervalo 1812-1829 ft: 0,16 bbl/psi*

✓ INTERVALO 1862-1878 ft

**Tabla 24. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos, Intervalo 1862-1878 Ft, Pozo Colorado 38, Zona B**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,1400	0	0,00%	0,3619	50	34,9%
0,1548	4	0,40%	0,3767	44	39,3%
0,1696	7	1,10%	0,3915	33	42,6%
0,1844	8	1,90%	0,4063	46	47,2%
0,1992	12	3,10%	0,4211	59	50,0%
0,2140	11	4,20%	0,4359	49	57,9%
0,2288	25	6,69%	0,4507	58	63,7%
0,2436	19	10,00%	0,4655	51	68,8%
0,2584	29	11,49%	0,4803	58	74,6%
0,2731	29	14,39%	0,4950	50	79,6%
0,2879	23	16,7%	0,5098	50	84,6%
0,3027	39	20,6%	0,5246	62	90,0%
0,3175	30	23,6%	0,5394	40	94,8%
0,3323	33	26,9%	0,5542	25	97,3%
0,3471	30	29,9%	0,5690	17	99,0%

**FUENTE:** El autor

*IP P10 intervalo 1862-1878 ft: 0,24 bbl/psi*

✓ INTERVALO 1905-1909 ft

**Tabla 25. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1905-1909 Ft, Pozo Colorado 38, Zona B**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0573	0	0,00%	0,0802	75	50,0%
0,0588	0	0,00%	0,0817	69	58,6%
0,0603	0	0,00%	0,0832	63	64,9%
0,0619	5	0,50%	0,0848	64	71,3%
0,0634	12	1,70%	0,0863	54	76,7%
0,0649	21	3,80%	0,0878	55	82,2%
0,0665	26	6,39%	0,0893	36	85,8%
0,0680	30	10,00%	0,0909	38	90,0%
0,0695	34	12,79%	0,0924	40	93,6%
0,0710	47	17,48%	0,0939	22	95,8%
0,0726	46	22,1%	0,0954	14	97,2%
0,0741	42	26,3%	0,0970	10	98,2%
0,0756	63	32,6%	0,0985	6	98,8%
0,0771	52	37,8%	0,1000	4	99,2%
0,0787	65	44,3%	0,1015	6	99,8%

FUENTE: El autor

*IP P10 intervalo 1905-1909 ft: 0,068 bbl/psi*

✓ INTERVALO 1938-1943 ft

**Tabla 26. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 1938-1943 Ft, Pozo Colorado 38, Zona B**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0753	0	0,00%	0,1089	82	50,0%
0,0776	0	0,00%	0,1111	66	59,5%
0,0798	1	0,10%	0,1133	69	66,4%
0,0820	2	0,30%	0,1156	69	73,3%
0,0843	8	1,10%	0,1178	49	78,2%
0,0865	4	1,50%	0,1201	42	82,4%
0,0888	19	3,40%	0,1223	37	86,1%
0,0910	27	6,09%	0,1245	33	90,0%
0,0932	29	10,00%	0,1268	30	92,4%
0,0955	34	12,39%	0,1290	30	95,4%
0,0977	54	17,8%	0,1312	18	97,2%
0,0999	56	23,4%	0,1335	6	97,8%
0,1022	53	28,7%	0,1357	11	98,9%
0,1044	72	35,9%	0,1379	6	99,5%
0,1066	89	44,8%	0,1402	3	99,8%

FUENTE: El autor

IP P10 intervalo 1938-1943 ft: 0,0932 bbl/psi

✓ INTERVALO 2110-2118 ft:

**Tabla 27. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 2110-2118 Ft, Pozo Colorado 38, Zona B**

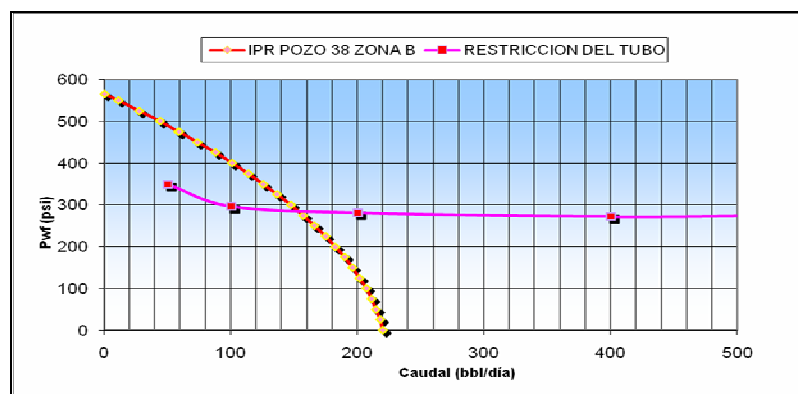
IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,1296	3	0,30%	0,1979	73	46,0%
0,1342	4	0,70%	<b>0,2025</b>	<b>80</b>	<b>50,0%</b>
0,1387	5	1,20%	0,2070	48	58,7%
0,1433	7	1,90%	0,2116	68	65,5%
0,1478	10	2,90%	0,2161	48	70,3%
0,1524	19	4,80%	0,2207	54	75,7%
0,1569	20	6,79%	0,2252	53	81,0%
<b>0,1615</b>	<b>31</b>	<b>10,00%</b>	0,2298	37	84,7%
0,1660	31	12,99%	0,2343	41	88,8%
0,1706	25	15,48%	<b>0,2389</b>	<b>24</b>	<b>90,0%</b>
0,1751	35	19,0%	0,2434	15	92,7%
0,1797	45	23,5%	0,2480	20	94,7%
0,1842	43	27,8%	0,2525	19	96,6%
0,1888	56	33,4%	0,2571	12	97,8%
0,1934	53	38,7%	0,2617	6	98,4%

FUENTE: El autor

IP P10 intervalo 2110-2118 ft: 0,1615 bbl/psi

Finalmente el IPR para el IP P10 es el presentado en la figura anexa:

**Figura 46. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel Y Tomando Como Referencia El Ip Del P10. Restricciones Del Tubo, Pozo Colorado 38, Zona B.**



FUENTE: El autor

**Tabla 28. Caudal Esperado De La Zona B A La Presión A La Cual Se Encuentra Produciendo. Referencia Ip Del P10**

CAUDAL ZONA B CON IP P10
58,6 Bbls/día

FUENTE: El autor

## ZONA C

De la misma forma para la zona C se predice el potencial de producción y el IPR, esta zona al igual que la zona B tiene 5 intervalos productores, los cuales se muestran a continuación.

**Tabla 29. Intervalos Productores Para El Pozo Colorado 38 Zona C**

INTERVALOS DE PRODUCCION POZO COL-38 ZONA C
INTERVALO 3383-3386 ft
INTERVALO 3453-3459 ft
INTERVALO 3488-3506 ft
INTERVALO 3705-3713 ft
INTERVALO 4006-4018 ft

FUENTE: El autor.

✓ INTERVALO 3383-3386 ft

**Tabla 30. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3383-3386 Ft) Colorado 38, Zona C.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	8	<b>Espesor del intervalo productor(ft)</b>	Mínimo	2
	Probable	10		Probable	3
	Máximo	11		Máximo	4
<b>Viscosidad(Cp)</b>	Mínimo	0,947231927	<b>Bo(Bls/STB)</b>	Mínimo	1,187779184
	Probable	0,947402894		Probable	1,187779577
	Máximo	0,94757385		Máximo	1,187779969

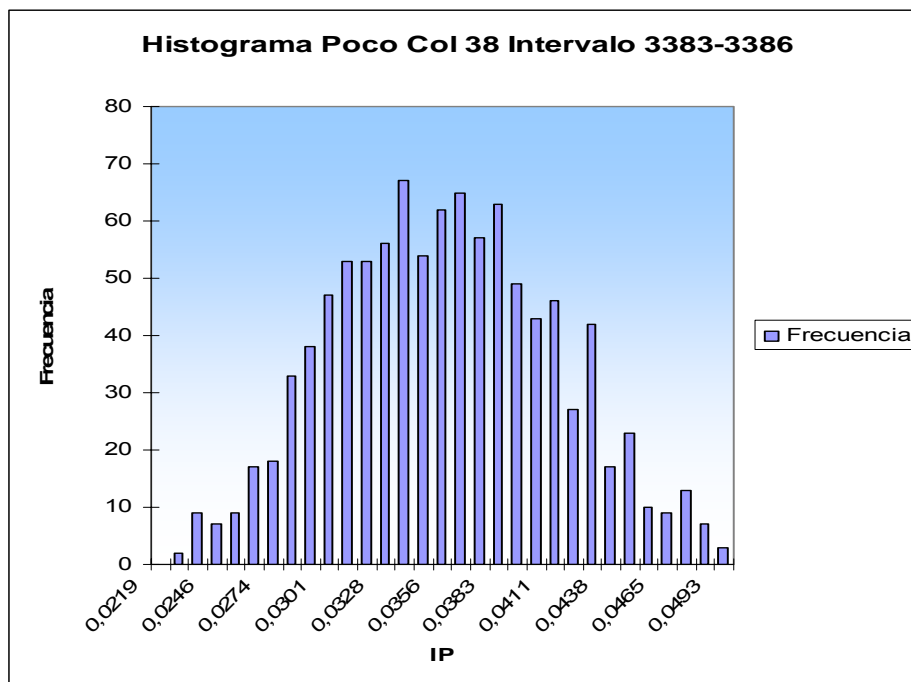
FUENTE: El autor.

**Figura 47. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C**

Datos de entrada						
K(mD)	Minimo		8	Espesor del intervalo productor(ft)	Minimo	2
	Probable		10		Probable	3
	Maximo		11		Maximo	4
Viscosidad(Cp)	Minimo		0,947231927	Bo(Bls/STB)	Minimo	1,18779184
	Probable		0,947402894		Probable	1,18779577
	Maximo		0,94757385		Maximo	1,18779969
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 3383-3386						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
9,763457618	3,145623604	0,947454156	1,1877952	0,037657339		
10,53755705	3,196063867	0,947328238	1,18779325	0,041301537		
10,26149267	2,58026325	0,947373164	1,1877968	0,032468682		
8,942746894	3,802579308	0,947260984	1,18779303	0,041705265		
9,859754415	2,757427549	0,947441831	1,18779644	0,033337179		
9,781248783	3,368354859	0,947436919	1,18779556	0,040411222		
10,03584024	2,597411165	0,947242248	1,1877953	0,031970152		
8,602199429	3,099330338	0,947367475	1,18779266	0,032694164		

FUENTE: El autor

**Figura 48. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C.**



FUENTE: El autor

**Tabla 31. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3383-3386 Ft, Pozo Colorado 38, Zona C.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0219	0	0,00%	0,0356	62	50,0%
0,0228	2	0,20%	0,0365	65	58,9%
0,0237	9	1,10%	0,0374	57	64,6%
0,0246	7	1,80%	0,0383	63	70,9%
0,0255	9	2,70%	0,0392	49	75,8%
0,0265	17	4,40%	0,0402	43	80,1%
0,0274	18	6,19%	0,0411	46	84,7%
0,0283	33	10,00%	0,0420	27	87,4%
0,0292	38	13,3%	0,0429	42	90,0%
0,0301	47	18,0%	0,0438	17	93,3%
0,0310	53	23,3%	0,0447	23	95,6%
0,0319	53	28,6%	0,0456	10	96,6%
0,0328	56	34,2%	0,0465	9	97,5%
0,0338	67	40,9%	0,0475	13	98,8%
0,0347	54	46,3%	0,0484	7	99,5%

FUENTE: El autor

*IP esperado: 0.0356 bbl/psi*

*IP P10: 0.0283 bbl/psi*

✓ INTERVALO 3453-3459 ft

**Tabla 32. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3453-3459 Ft) Colorado 38, Zona C.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	6	<b>Espesor del intervalo productor(ft)</b>	Mínimo	5
	Probable	8		Probable	6
	Máximo	11		Máximo	7
<b>Viscosidad(Cp)</b>	Mínimo	0,943623931	<b>Bo(BIs/STB)</b>	Mínimo	1,187787519
	Probable	0,943782121		Probable	1,187787881
	Máximo	0,943940301		Máximo	1,187788243

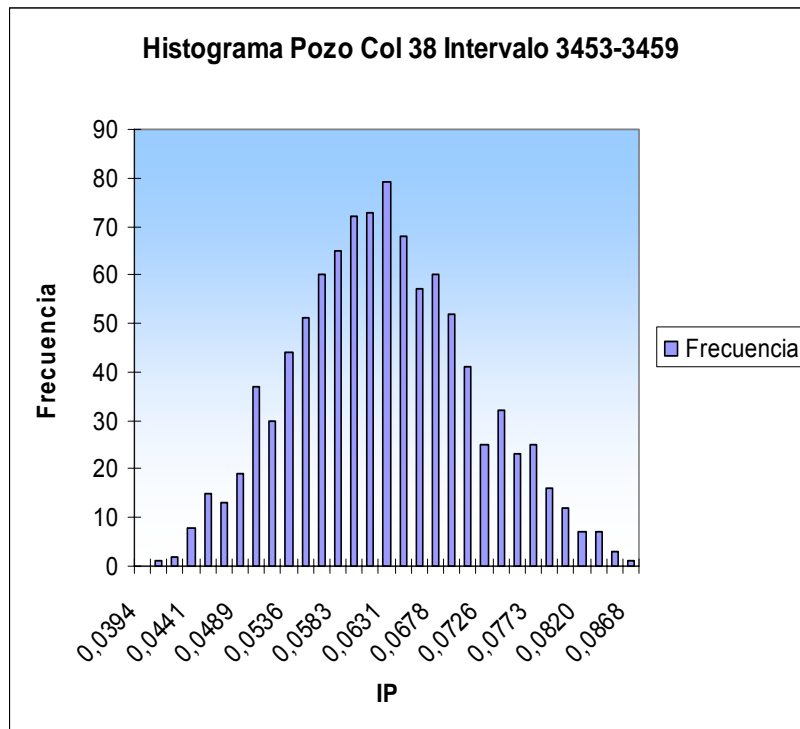
FUENTE: El autor.

**Figura 49. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C**

Datos de entrada						
K(mD)	Minimo		6	Espesor del intervalo productor(ft)	Minimo	5
	Probable		8		Probable	6
	Maximo		11		Maximo	7
Viscosidad(Cp)	Minimo		0,943623931	Bo(Bls/STB)	Minimo	1,187787519
	Probable		0,943782121		Probable	1,187787881
	Maximo		0,943940301		Maximo	1,187788243
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 3453-3459						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
7,746766825	5,61435473	0,943833877	1,187787589	0,053534415		
7,83936765	5,407233464	0,943712787	1,187788178	0,05218244		
8,20598142	6,084330372	0,943720801	1,187787618	0,061462195		
7,954843347	5,908474515	0,943810376	1,187787832	0,057853608		
7,866285505	5,834419605	0,943643373	1,187787831	0,056502501		
9,089433298	6,168795108	0,943930901	1,187787877	0,069008904		
7,606078325	6,177546177	0,943707414	1,187787891	0,057842573		
7,2479303	6,228905942	0,943826152	1,18778766	0,05557021		

FUENTE: El autor

**Figura 50. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C.**



FUENTE: El autor

**Tabla 33. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3453-3459 Ft, Pozo Colorado 38, Zona C.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0394	0	0,00%	0,0631	79	56,8%
0,0410	1	0,10%	0,0647	68	63,6%
0,0426	2	0,30%	0,0662	57	69,3%
0,0441	8	1,10%	0,0678	60	75,3%
0,0457	15	2,60%	0,0694	52	80,5%
0,0473	13	3,90%	0,0710	41	84,6%
0,0489	19	5,79%	0,0726	25	87,1%
0,0505	37	10,00%	0,0741	32	90,0%
0,0520	30	12,49%	0,0757	23	92,6%
0,0536	44	16,88%	0,0773	25	95,1%
0,0552	51	22,0%	0,0789	16	96,7%
0,0568	60	28,0%	0,0805	12	97,9%
0,0583	65	34,5%	0,0820	7	98,6%
0,0599	72	41,7%	0,0836	7	99,3%
0,0615	73	50,0%	0,0852	3	99,6%

FUENTE: El autor

*IP esperado: 0.0615 bbl/psi*

*IP P10: 0.0505 bbl/psi*

✓ INTERVALO 3488-3506 ft

**Tabla 34. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3488-3506 Ft) Colorado 38, Zona C.**

Datos de entrada			Datos de entrada		
K(mD)	Mínimo	7	Espesor del intervalo productor(ft)	Mínimo	8
	Probable	15		Probable	9
	Máximo	23		Máximo	10
Viscosidad(Cp)	Mínimo	0,941143419	Bo(BIs/STB)	Mínimo	1,187791746
	Probable	0,941618698		Probable	1,187792833
	Máximo	0,94209388		Máximo	1,187793921

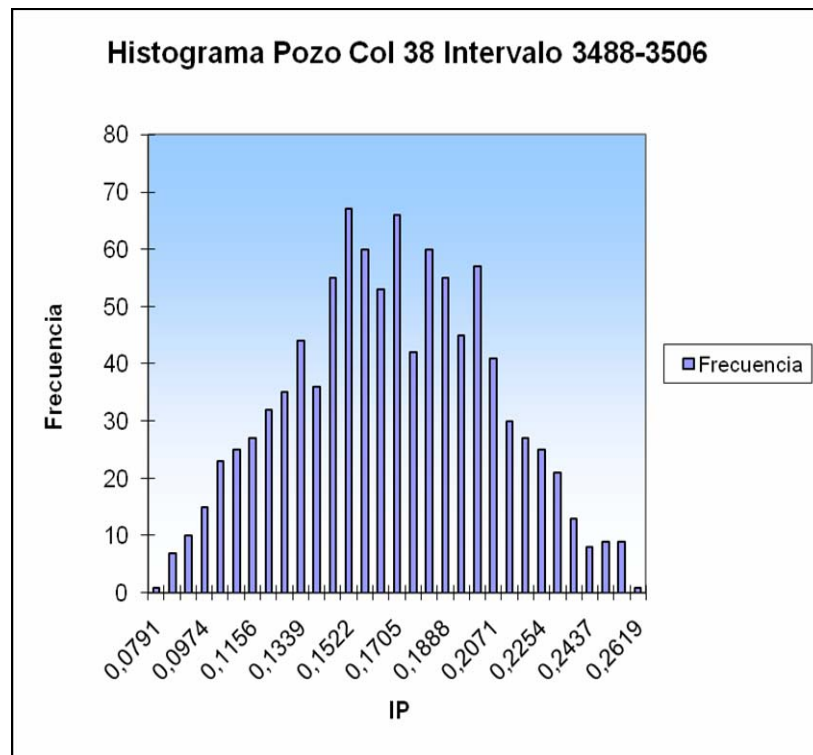
FUENTE: El autor.

**Figura 51. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C**

Datos de entrada						
K(mD)	Minimo		7	Espesor del intervalo productor(ft)	Minimo	8
	Probable		15		Probable	9
	Maximo		23		Maximo	10
Viscosidad(Cp)	Minimo		0,941143419	Bo(Bls/STB)	Minimo	1,187791746
	Probable		0,941618698		Probable	1,187792833
	Maximo		0,94209388		Maximo	1,187793921
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 3488-3506						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad		Valor aleatorio de Bo		PI COL 38 ZONA B
15,36746665	8,787802575	0,941435687		1,187792284		0,166647367
12,30463599	8,992348607	0,941889531		1,18779298		0,136473476
14,96898816	8,942888342	0,941487149		1,187793402		0,165159663
10,15879225	9,00414729	0,941377257		1,187793056		0,112882666
13,95838576	8,490999433	0,941639651		1,187792702		0,146223
10,29118339	8,849257361	0,941389834		1,187793045		0,112385151
18,56060746	8,617379892	0,941603832		1,187792454		0,197335726
7,703044201	9,298131684	0,941832954		1,18779328		0,088346715

FUENTE: El autor

**Figura 52. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C.**



FUENTE: El autor

**Tabla 35. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3488-3506 Ft, Pozo Colorado 38, Zona C.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,0791	1	0,10%	0,1705	66	55,5%
0,0852	7	0,80%	0,1766	42	59,7%
0,0913	10	1,80%	0,1827	60	65,7%
0,0974	15	3,30%	0,1888	55	71,2%
0,1034	23	5,59%	0,1949	45	75,7%
0,1095	25	8,09%	0,2010	57	81,4%
0,1156	27	10,00%	0,2071	41	85,5%
0,1217	32	13,99%	0,2132	30	90,0%
0,1278	35	17,48%	0,2193	27	91,2%
0,1339	44	21,88%	0,2254	25	93,7%
0,1400	36	25,5%	0,2315	21	95,8%
0,1461	55	31,0%	0,2376	13	97,1%
0,1522	67	37,7%	0,2437	8	97,9%
0,1583	60	43,7%	0,2498	9	98,8%
0,1644	53	50,0%	0,2559	9	99,7%

FUENTE: El autor

IP esperado: 0,1644 bbl/psi

IP P10: 0,1156 bbl/psi

✓ INTERVALO 3705-3713 ft

**Tabla 36. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (3705-3713 Ft) Colorado 38, Zona C.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	13	<b>Espesor del intervalo productor(ft)</b>	Mínimo	13
	Probable	17		Probable	14
	Máximo	20		Máximo	15
<b>Viscosidad(Cp)</b>	Mínimo	0,930169987	<b>Bo(BIs/STB)</b>	Mínimo	1,187817957
	Probable	0,930382787		Probable	1,187818441
	Máximo	0,930595568		Máximo	1,187818924

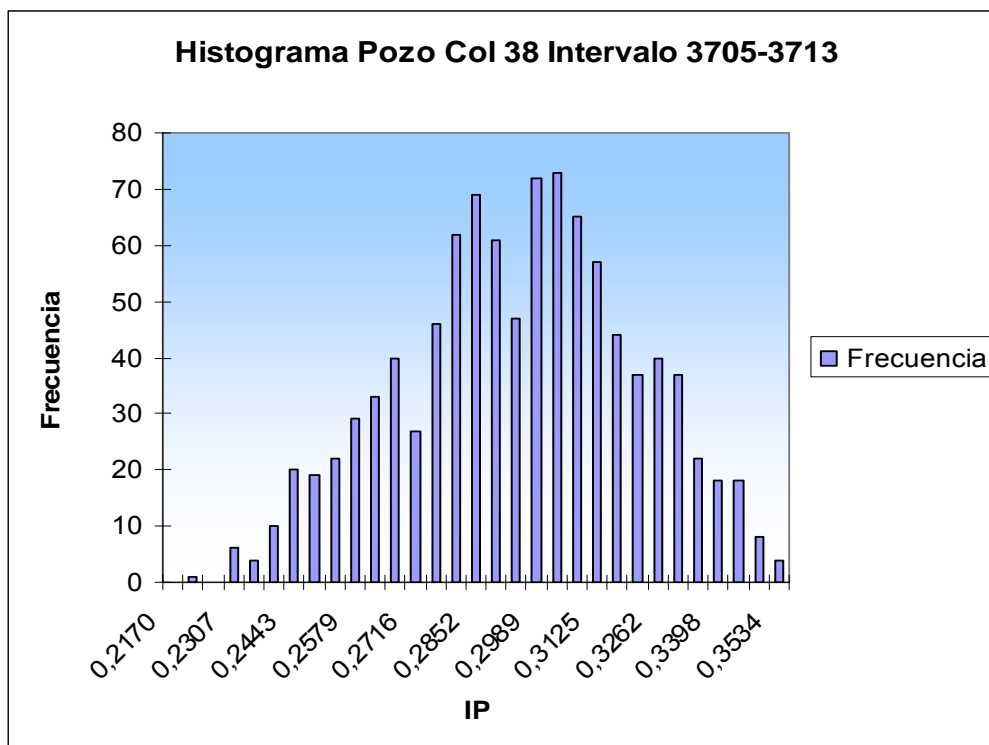
FUENTE: El autor.

**Figura 53. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C**

Datos de entrada						
K(mD)	Minimo		13	Espesor del intervalo productor(ft)	Minimo	13
	Probable		17		Probable	14
	Maximo		20		Maximo	15
Viscosidad(Cp)	Minimo	0,930169987		Bo(Bls/STB)	Minimo	1,187817957
	Probable	0,930382787			Probable	1,187818441
	Maximo	0,930595668			Maximo	1,187818924
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 3705-3713						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
14,76684816	13,47675234	0,930282003	1,187818426	0,248516627		
15,68801277	14,0028664	0,930382255	1,18781885	0,274296533		
16,35265443	14,4268773	0,930483718	1,187818626	0,294543036		
17,60218982	13,70615025	0,930280262	1,187818567	0,301276559		
17,62682385	13,96501336	0,930275546	1,187818492	0,307397834		
15,44081084	14,43782592	0,930195852	1,187818088	0,278416289		
14,29222689	14,0973074	0,930371587	1,187818711	0,251580245		
18,4395837	13,47174602	0,930404701	1,187818539	0,310170203		

FUENTE: El autor

**Figura 54. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C.**



FUENTE: El autor

**Tabla 37. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 3705-3713 Ft, Pozo Colorado 38, Zona C.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,2170	0	0,00%	0,2852	69	38,8%
0,2216	1	0,10%	0,2898	61	44,9%
0,2261	0	0,10%	<b>0,2943</b>	<b>47</b>	<b>50,0%</b>
0,2307	6	0,70%	0,2989	72	56,7%
0,2352	4	1,10%	0,3034	73	64,0%
0,2397	10	2,10%	0,3080	65	70,5%
0,2443	20	4,10%	0,3125	57	76,2%
0,2488	19	5,99%	0,3171	44	80,6%
0,2534	22	8,19%	0,3216	37	84,3%
<b>0,2579</b>	<b>29</b>	<b>10,00%</b>	<b>0,3262</b>	<b>40</b>	<b>90,0%</b>
0,2625	33	14,4%	0,3307	37	92,0%
0,2670	40	18,4%	0,3353	22	94,2%
0,2716	27	21,1%	0,3398	18	96,0%
0,2761	46	25,7%	0,3444	18	97,8%
0,2807	62	31,9%	0,3489	8	98,6%

FUENTE: El autor

*IP esperado: 0.2943 bbl/psi*

*IP P10: 0,2579 bbl/psi*

✓ INTERVALO 4006-4018 ft

**Tabla 38. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Para El Intervalo Productor (4006-4018 Ft) Colorado 38, Zona C.**

Datos de entrada			Datos de entrada		
<b>K(mD)</b>	Mínimo	12	<b>Espesor del intervalo productor(ft)</b>	Mínimo	7
	Probable	17		Probable	8
	Máximo	23		Máximo	9
<b>Viscosidad(Cp)</b>	Mínimo	0,913853506	<b>Bo(BIs/STB)</b>	Mínimo	1,187854315
	Probable	0,91417621		Probable	1,18785504
	Máximo	0,914498868		Máximo	1,187855764

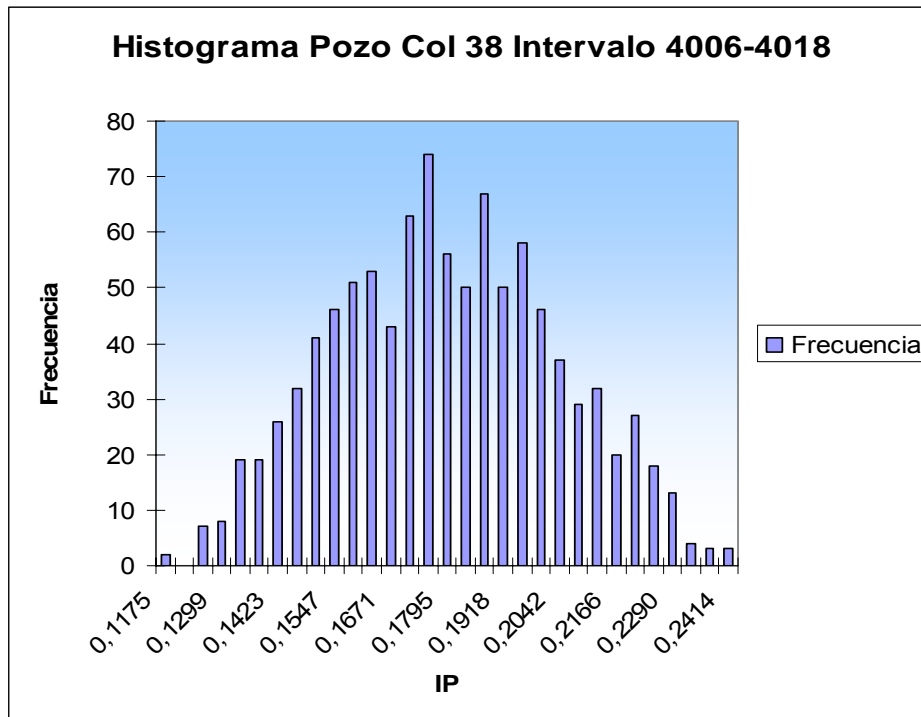
FUENTE: El autor.

**Figura 55. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C**

Datos de entrada						
K(mD)	Minimo		12	Espesor del intervalo productor(ft)	Minimo	7
	Probable		17		Probable	8
	Maximo		23		Maximo	9
Viscosidad(Cp)	Minimo		0,913853506	Bo(Bls/STB)	Minimo	1,187854315
	Probable		0,91417621		Probable	1,18785504
	Maximo		0,914498868		Maximo	1,18785764
Radio del pozo (ft)		0,276				
Re (ft)		98,99				
INTERVALO 4006-4018						
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Bo	PI COL 38 ZONA B		
14,24451738	7,278888911	0,914127956	1,187855047	0,131761783		
18,05070582	8,108686136	0,914080509	1,187854801	0,186013277		
21,85911685	8,086269293	0,91423638	1,187854535	0,224598124		
14,41504264	8,775073698	0,914198538	1,187854923	0,160734776		
16,95452531	7,851967452	0,914153155	1,187854997	0,169172125		
14,8458086	7,396667845	0,914143341	1,187855056	0,139541518		
17,12497754	8,210529699	0,913993809	1,187855145	0,17870698		
20,78685976	8,424898337	0,914202319	1,187855013	0,22253323		

FUENTE: El autor

**Figura 56. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C.**



FUENTE: El autor

**Tabla 39. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Intervalo 4006-4018 Ft, Pozo Colorado 38, Zona C.**

IP	Frecuencia	Percentil	IP	Frecuencia	Percentil
0,1175	2	0,20%	0,1795	56	53,9%
0,1216	0	0,20%	0,1836	50	58,9%
0,1258	7	0,90%	0,1877	67	65,6%
0,1299	8	1,70%	0,1918	50	70,6%
0,1340	19	3,60%	0,1960	58	76,4%
0,1382	19	5,49%	0,2001	46	81,0%
0,1423	26	8,09%	0,2042	37	84,7%
0,1464	32	10,00%	0,2084	29	87,6%
0,1505	41	15,38%	0,2125	32	90,0%
0,1547	46	19,98%	0,2166	20	92,8%
0,1588	51	25,1%	0,2207	27	95,5%
0,1629	53	30,4%	0,2249	18	97,3%
0,1671	43	34,7%	0,2290	13	98,6%
0,1712	63	41,0%	0,2331	4	99,0%
0,1753	74	50,0%	0,2373	3	99,3%

FUENTE: El autor

IP esperado: 0,1753 bbl/psi

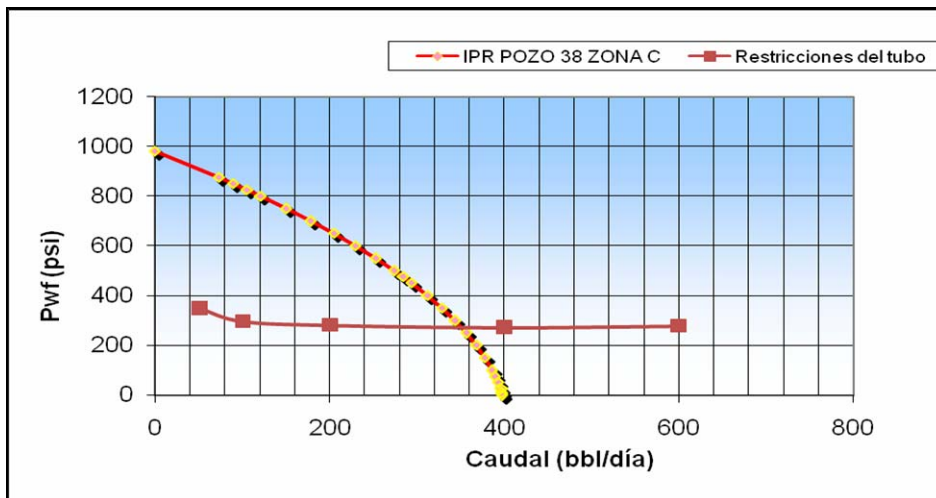
IP P10: 0,1464 bbl/psi

IP esperado total zona C: 0.73 bbl/psi

Finalmente se calculan los IPR:

IPR esperado:

**Figura 57. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel. Restricciones Del Tubo. Pozo Colorado 38, Zona C**



FUENTE: El autor

Vale la pena aclarar que al aplicar la metodología para el pozo colorado 38, se tomaron los datos históricos más no los actuales. Lo anterior, obviamente, da como resultado un potencial de producción mucho mayor al actual, donde parámetros significativos, como la presión, han caído notablemente desde sus inicios hasta la fecha. Lo anterior es válido, porque no se puede confundir esto con una posible falla de la metodología desarrollada.

#### 4.2. METODOLOGÍA PARA POZOS HORIZONTALES. IMPLEMENTACIÓN EN HOJA DE CÁLCULO.

A continuación, se extenderá la metodología explicada anteriormente para analizar los pozos horizontales. Las ecuaciones implicadas para el cálculo del índice de productividad en pozos horizontales son:

$$IP = \frac{2\pi k_h h / (\mu B_o)}{\ln \left[ \frac{a + \sqrt{a^2 - (L/2)^2}}{(L/2)} \right] + \frac{\beta h}{L} \ln \left( \frac{\beta h}{2r_w} \right)} \quad (6)$$

$$a = \frac{L}{2} \left[ \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{(L/2r_{eH})^4}} \right]^{\frac{1}{2}} \quad (7)$$

Donde el significado de cada variable ya fue explicado en el capítulo 1. Para la aplicación en pozos horizontales se tomaron los datos del artículo SPE-107669-MS-P. En cuanto al yacimiento en estudio en este artículo cabe anotar lo siguiente:

Existen muchas incertidumbres involucradas en la predicción de los potenciales de producción de los pozos horizontales. Estas incertidumbres, como ya hemos tratado anteriormente, se presentan en las propiedades de las rocas, como el espesor neto y las permeabilidades absolutas, en las propiedades del fluido, como la viscosidad, y las propiedades que dependen de ambos, como las permeabilidades relativas. En un pozo horizontal, la permeabilidad vertical es también un factor importante para predecir la productividad.

Por lo tanto, para modelar el comportamiento del yacimiento, un buen estimativo del IP del pozo es necesario. Dependiendo de este valor, la tasa inicial de aceite puede ser considerablemente diferente. Este valor es mucho mas importante cuando los valores de IP son bajos, y este es el caso de los yacimientos de crudo pesado. Para este caso, debido al reto que representaba el campo en desarrollo- el bajo API y el petróleo tan viscoso, y el espesor del yacimiento que varía de 15 a 35 metros- se requiere el uso de tecnologías recientes como largos pozos horizontales y líneas térmicas de flujo aisladas.

El campo está localizado en aguas profundas que varían de 800 metros a 2000 metros, con una pequeña cubierta de sedimentos de solo 500 metros en el área de estudio. El yacimiento es una arena del Mioceno con alta porosidad y permeabilidad. La distribución de fluidos en el yacimiento es un poco compleja.

El desarrollo del campo comprende 17 pozos de producción y 15 pozos de inyección, todos horizontales y la producción del campo esta alrededor de 170 000 BPD, con 14 pozos de producción en operación.

Para este caso las incertidumbres son:

- Espesor neto promedio del yacimiento ( $h$ ) – un valor mas probable de 25 metros fue considerado (en un rango de 15 metros a 35 metros), estos basados en las incertidumbres de los datos sísmicos, debido a que ningún pozo estaba perforado en la región de estudio.
- Viscosidad del petróleo ( $\mu$ ) – de acuerdo a los análisis PVT de los pozos cercanos, la viscosidad varia de 24 a 150 Cp, y el valor mas probable fue tomado como 100 Cp
- Las permeabilidades horizontales ( $k_h$ ) – de acuerdo a las pruebas a la formación, los resultados de producción e inyección en las regiones análogas, el valor mas probable fue tomado como 2230 mD, variando de 600 mD a 9000 mD.

- La relación de permeabilidades verticales-horizontales ( $k_h/k_v$ ) - de acuerdo a las pruebas a la formación, los resultados de producción e inyección en las regiones análogas, el valor mas probable fue considerado como 4% y variaba de 0.6 % a 18%.
- El radio de drenaje del pozo ( $r_e$ ) – basados en los datos sísmicos, este parámetro variaba de 600 metros a 1000 metros, y el valor mas probable fue considerado como 600 metros.

### **IMPLEMENTACION DE LA METODOLOGIA:**

De la misma forma que para pozos desviados, se trabajó en una hoja de cálculo de Microsoft Excel. En orden para calcular en índice de producción, se usó la ecuación de Joshi mencionada previamente. Los parámetros de entrada de la función son  $k_h, k_v, h, \mu, Bo, L, r_w$ , los cuales fueron definidos antes. Otros parámetros que se necesitan son el factor de daño, S, el radio de drenaje del pozo,  $r_e$ , y la posición vertical del pozo horizontal,  $Z_w$ . Para este estudio el factor de daño y la excentricidad del pozo no fueron considerados, esto indica que,  $S=0$  y  $Z_w = h/2$ . Los valores de salida son los correspondientes a los resultados de la función de índices de producción para pozos horizontales.

Una vez más, la distribución triangular fue usada. Sus parámetros de entrada son números aleatorios entre 0 y 1, valores mínimos, máximos y más probables. Por lo tanto, debemos construir una función de distribución triangular para cada una de las variables involucradas en el cálculo del índice de producción según la ecuación de Joshi.

**Tabla 40. Valores Mínimo, Máximo Y Más Probable De Cada Variable Aleatoria Involucrada En El Cálculo Del Ip Según Ecuación De Joshi.**

Datos de entrada			Datos de entrada		
Kh(mD)	Mínimo	600	Espesor neto de la formación(m)	Mínimo	15
	Probable	2230		Probable	25
	Máximo	9000		Máximo	35
Kv/Kh	Mínimo	0,60%	Viscosidad(Cp)	Mínimo	24
	Probable	4,00%		Probable	100
	Máximo	18,00%		Máximo	150
Longitud horizontal del pozo (m)	Inicial	600	Radio de drenaje	Mínimo	800
	Final	1000		Probable	800
	Incremento	200		Máximo	800
Radio del pozo(m)	rwh	0,078105	Bo(r <sup>3</sup> /m <sup>3</sup> )		1,097
N° de puntos para Montecarlo	10000		Factor de daño		0

**FUENTE:** El autor.

De nuevo se generan los números aleatorios de cada una de las variables involucradas en el cálculo del IP, y de la misma forma, se hará uso de la función aleatorio de Excel en combinación con la función de densidad de probabilidad triangular. De esta forma, se utiliza la misma ecuación (102)

$$F(x) = c + (a + U * (b - a) - c) * \sqrt{U} \quad (102)$$

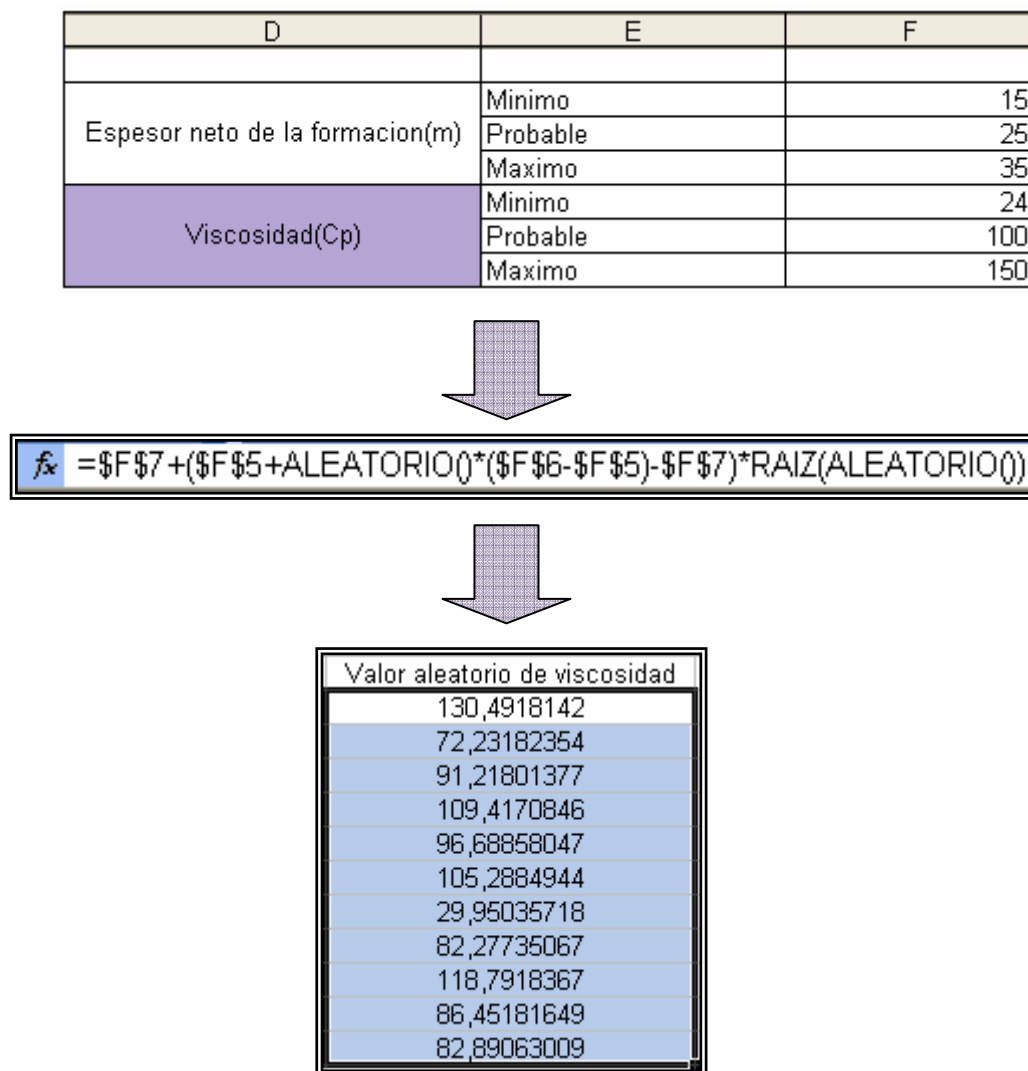
Donde:

- a: Valor mínimo de la variable evaluada.
- b: Valor más probable de la variable evaluada.
- c: Valor máximo de la variable evaluada.
- U: Función aleatorio () de Excel.

En el artículo ejemplo se desean predecir los IP de pozos horizontales usando simulación Monte Carlo, para esto se comparan los IP de 3 diferentes longitudes de pozos, 600 metros, 800 metros y 1000 metros. Ya que se tienen los valores mínimos, máximos y más probables de todas las variables involucradas en el cálculo de los IP, se procede a generar la función de distribución de densidad para cada variable usando una hoja de cálculo en

Excel. Por ejemplo, para la viscosidad ( $\mu$ ), y usando la ecuación (102), podemos crear números aleatorios colocándonos en la celda y arrastrando hacia abajo el cursor, tantas veces como se quiera; cada celda hacia abajo es considerada como una iteración. En la figura 58 se ilustra esto.

**Figura 58. Esquema Que Ilustra La Manera De Generar La Función De Distribución De Densidad Para La Viscosidad.**

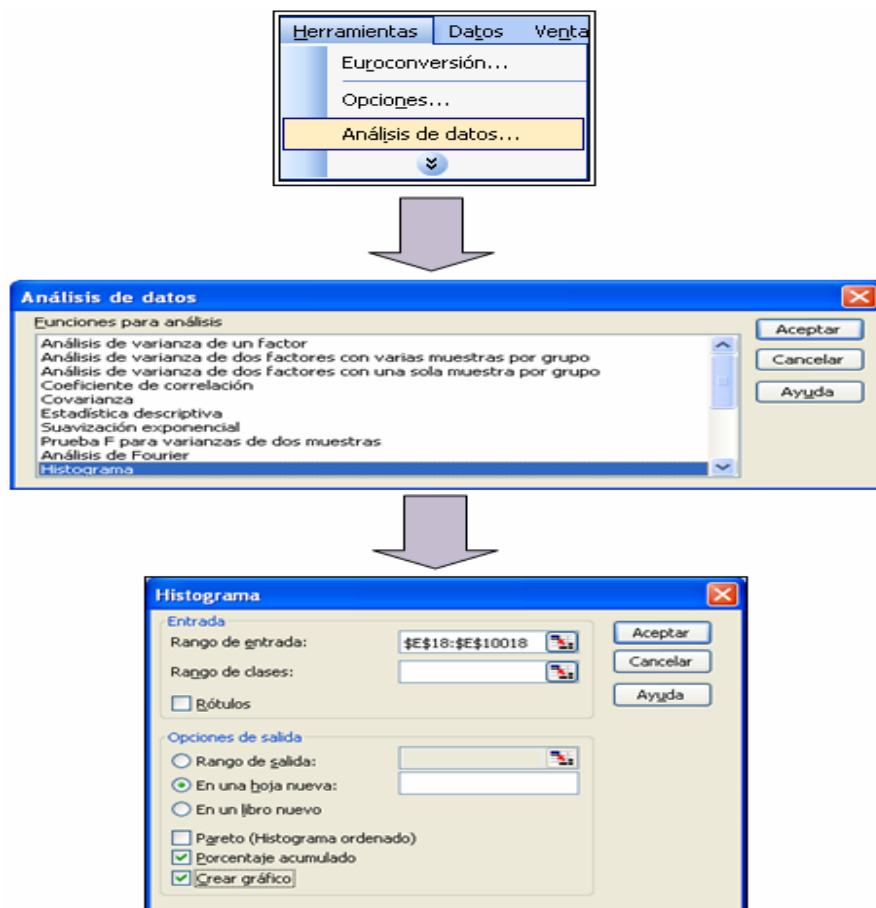


**FUENTE:** EL AUTOR

Este proceso se repite con las demás variables, tales como son  $K_h$ ,  $L$ ,  $\beta$ ,  $(k_h/k_v)$ ,  $h$ . Estas variables se colocan en la ecuación (6), y se obtiene el IP de cada combinación aleatoria de variables para cada uno de las longitudes de

pozo anteriormente dichas. Posteriormente se realiza un análisis de datos igual que se hizo para los pozos desviados, esto es, Microsoft Excel posee una opción en herramientas llamada análisis de datos. Internamente de esta elección, existe un completo set de herramientas para hacer análisis estadístico de datos, que es el paso siguiente luego de hacer la SMC. En este caso se usará la herramienta Histograma, aunque cabe resaltar que se pueden usar las diversas herramientas, tales como estadística descriptiva, etc. Dentro de esta opción lo que se encontrará será algo así como lo que se muestra en el final del esquema mostrado en la figura 59, en esta se seleccionará el rango o los datos a ser analizados para construir un histograma y ver cuales es el dato de la combinación aleatoria de variables (IP) que mas se repita, esto se hace para las diferentes longitudes.

**Figura 59. Esquema Que Ilustra La Forma De Realizar Un Análisis De Datos Luego De La Aplicación Del Método Montecarlo En Excel.**

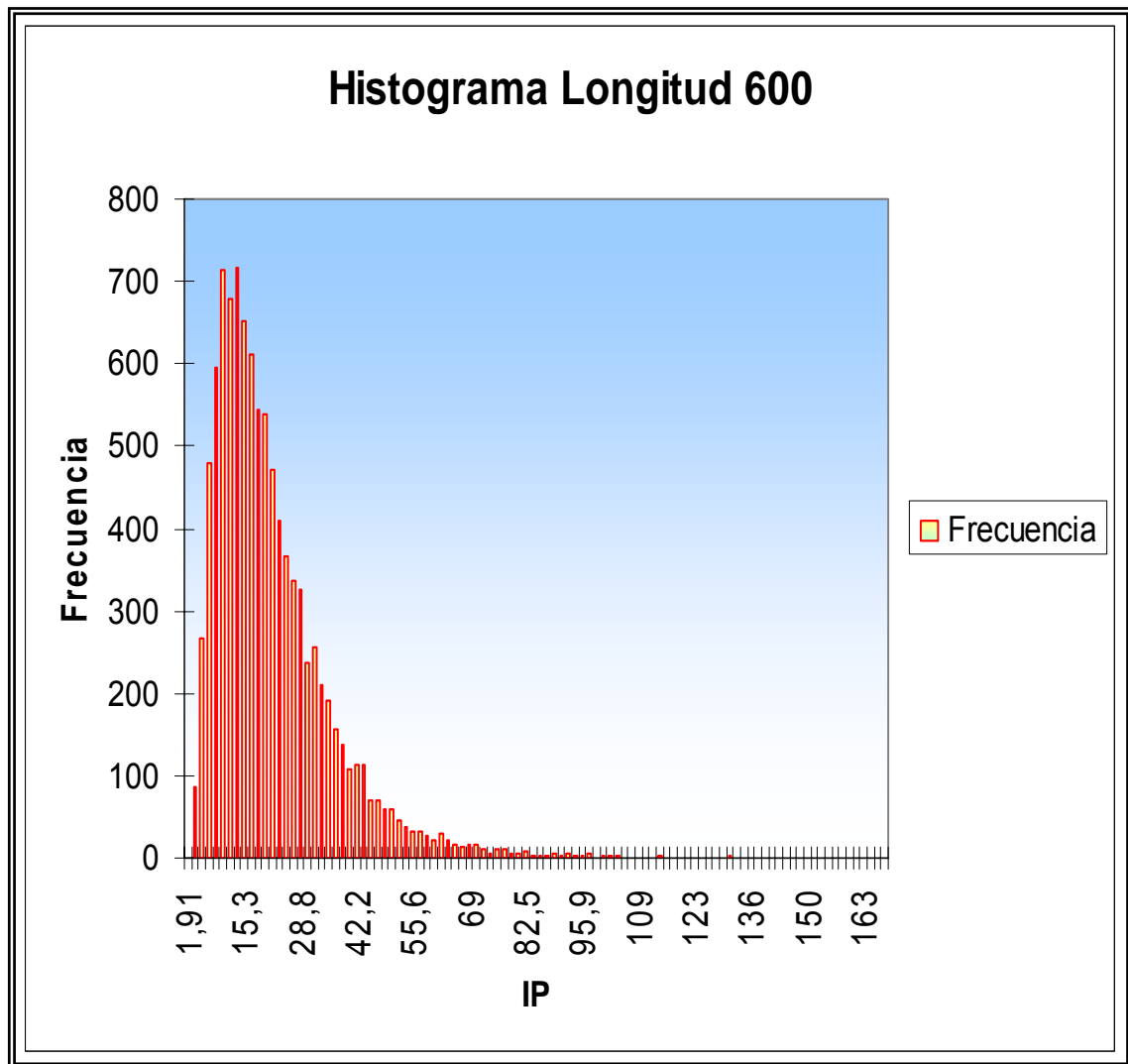


FUENTE: EL AUTOR

Los resultados se muestran continuación:

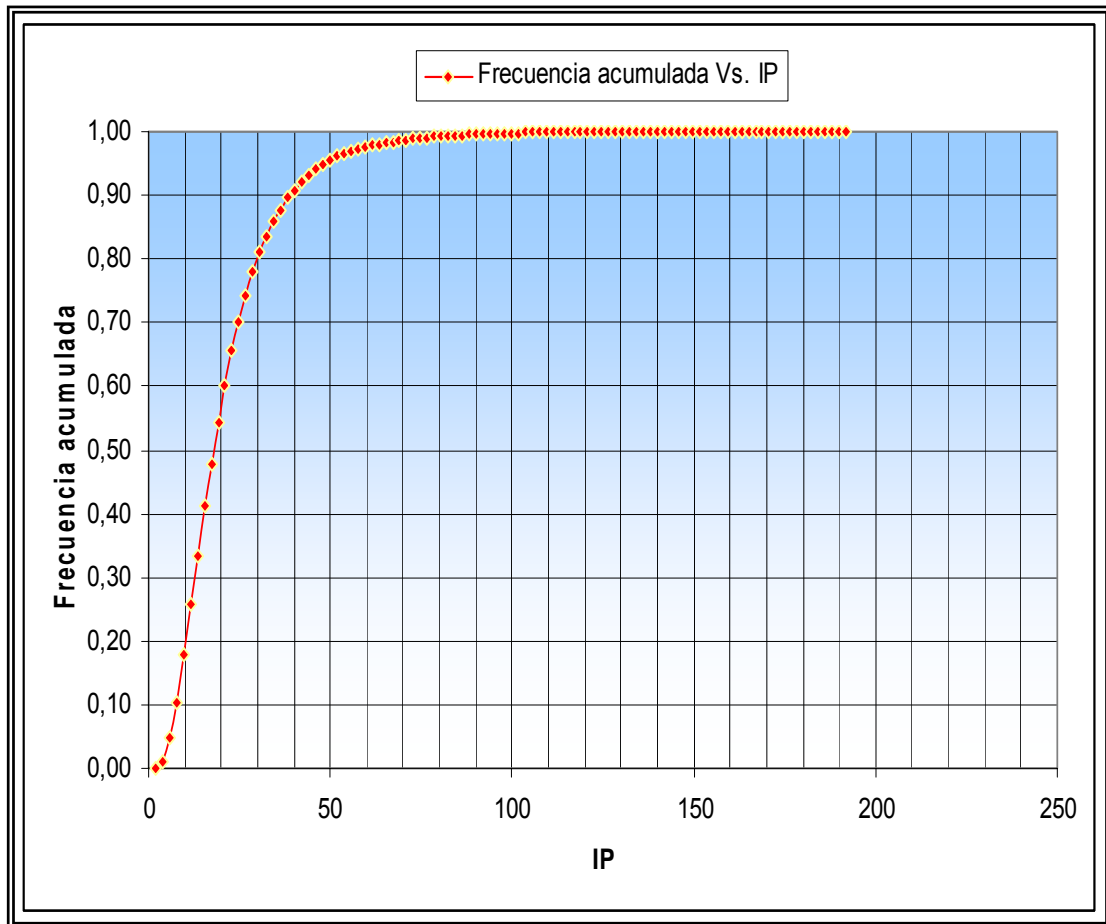
✓ Longitud 600:

**Figura 60. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 600 Metros**



FUENTE: El autor

**Figura 61. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 600 Metros.**



FUENTE: EL AUTOR

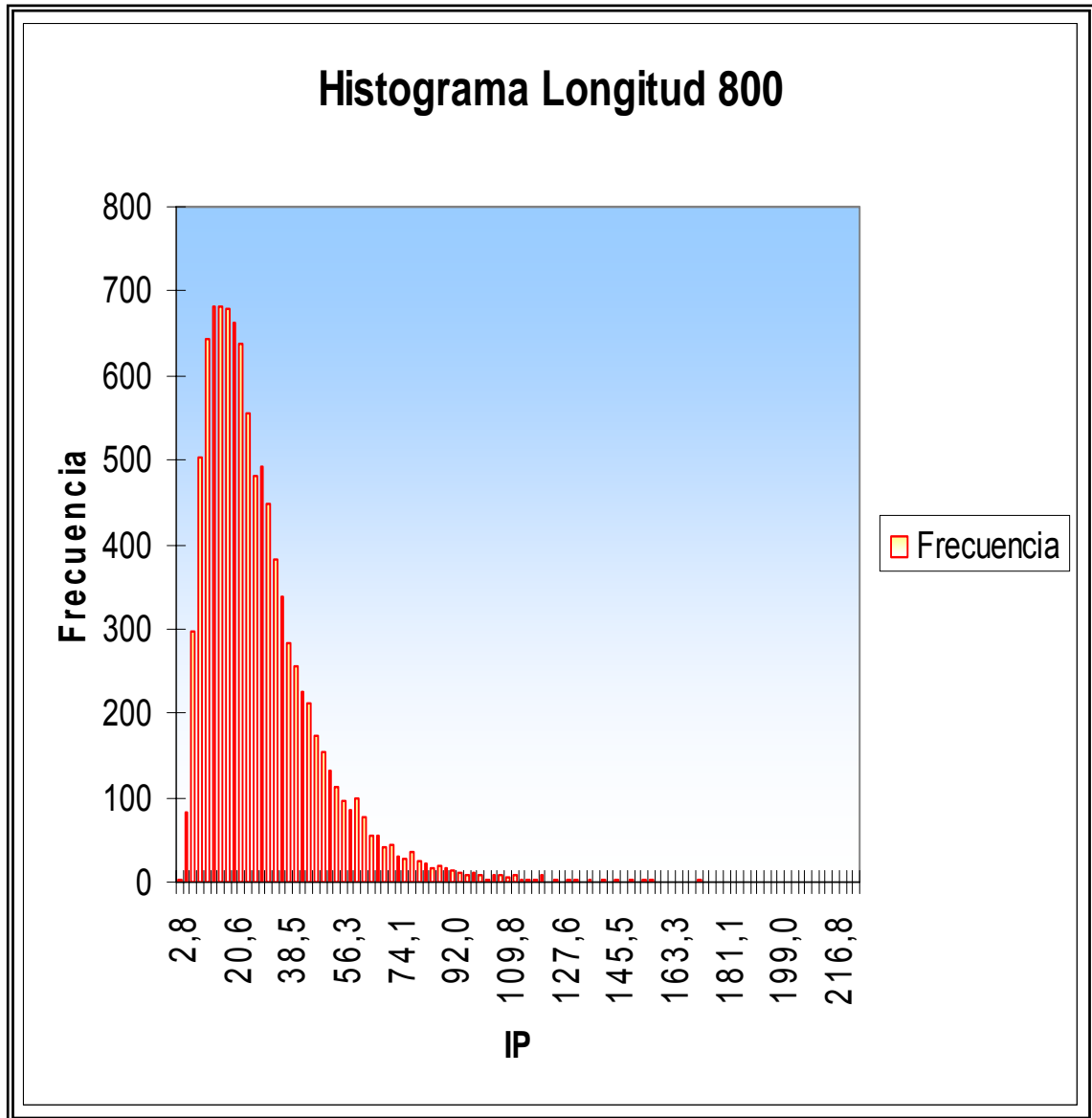
**Tabla 41. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 600 Metros**

Percentil	IP (bbl/día)
P10	7,62
P50	17,22
P90	38,33

FUENTE: EL AUTOR

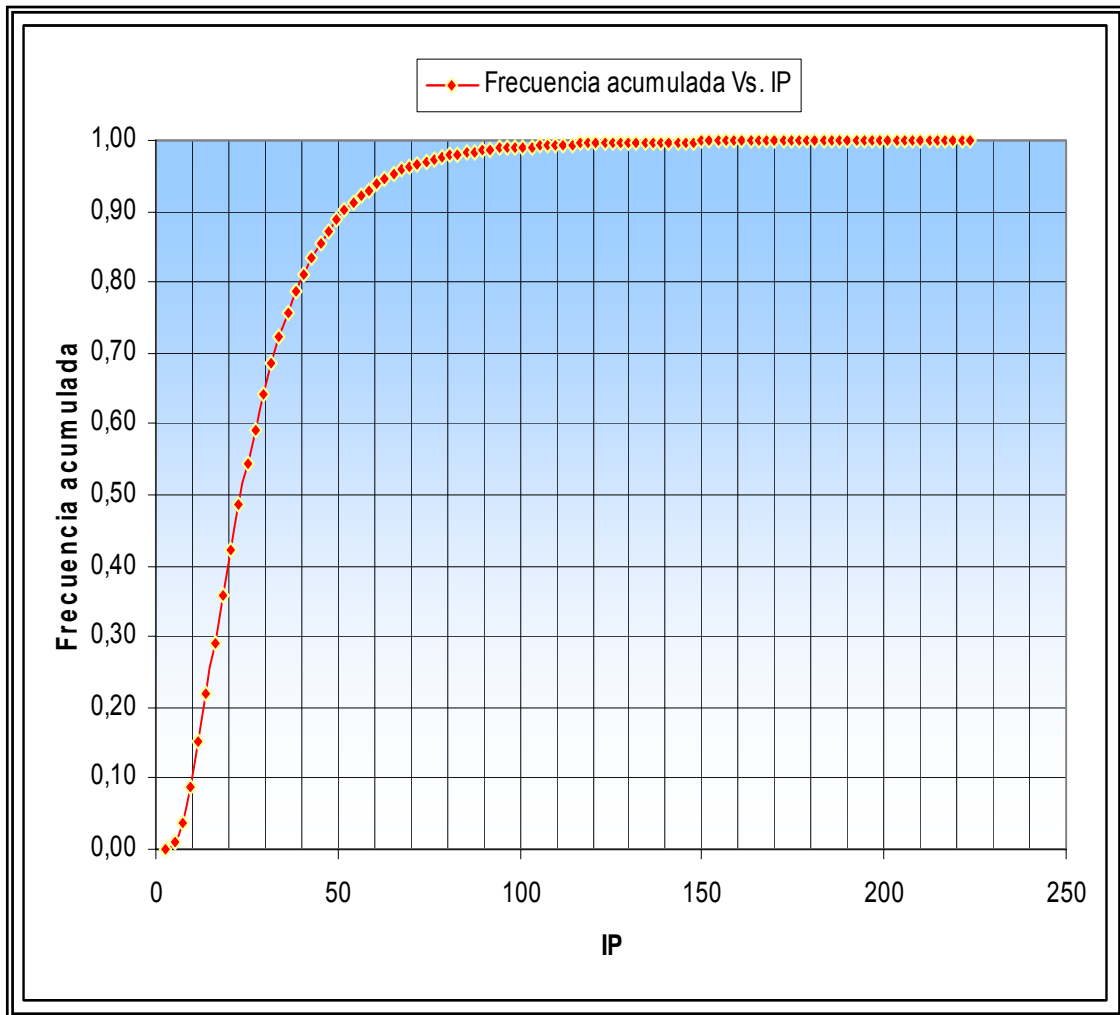
✓ Longitud 800:

**Figura 62. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia.Longitud De Pozo 800 Metros**



**FUENTE:** El autor

**Figura 63. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 800 Metros.**



FUENTE: EL AUTOR

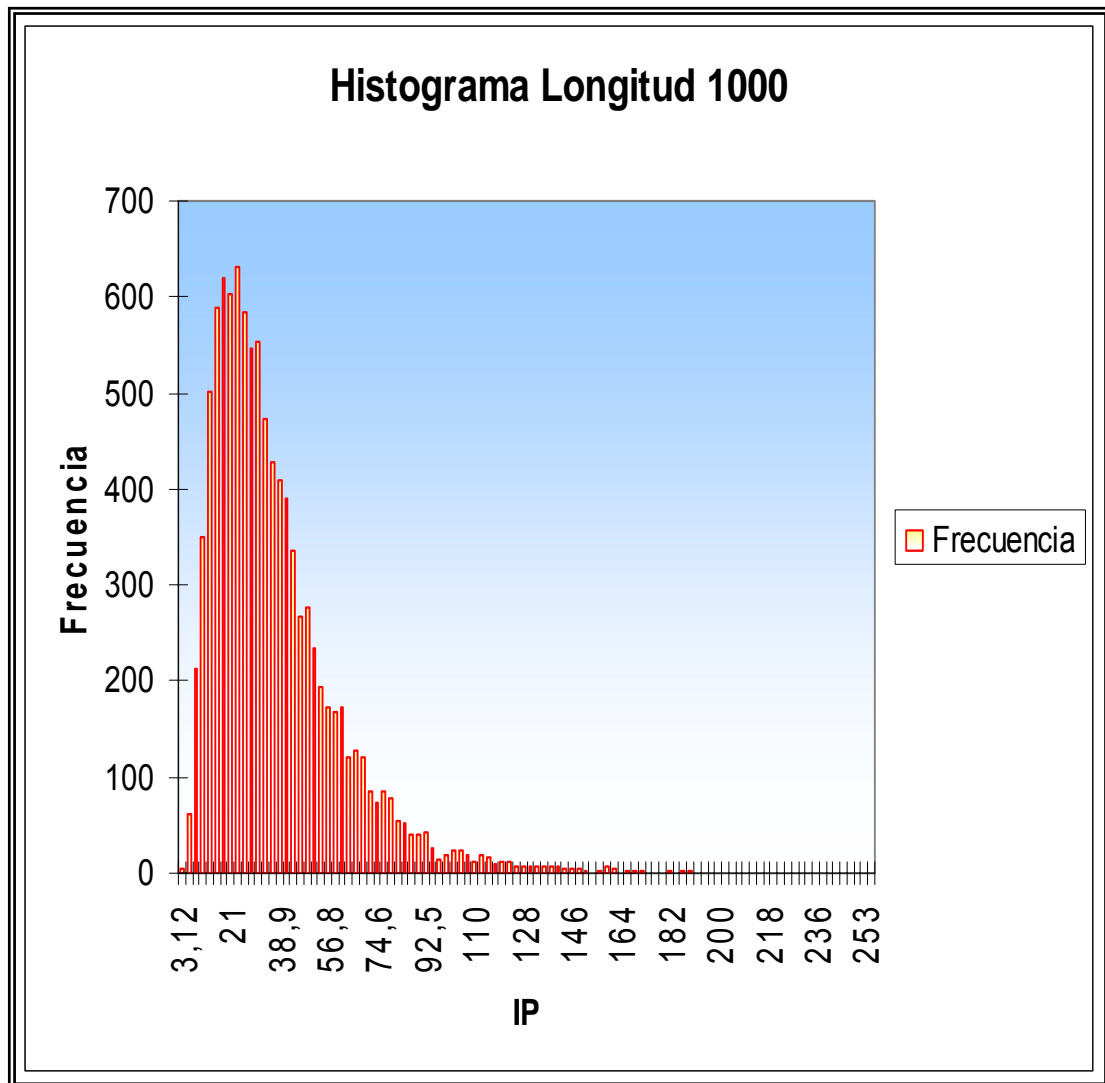
**Tabla 42. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 800 Metros**

Percentil	IP (bbl/día)
P10	9,48
P50	22,86
P90	51,84

FUENTE: EL AUTOR

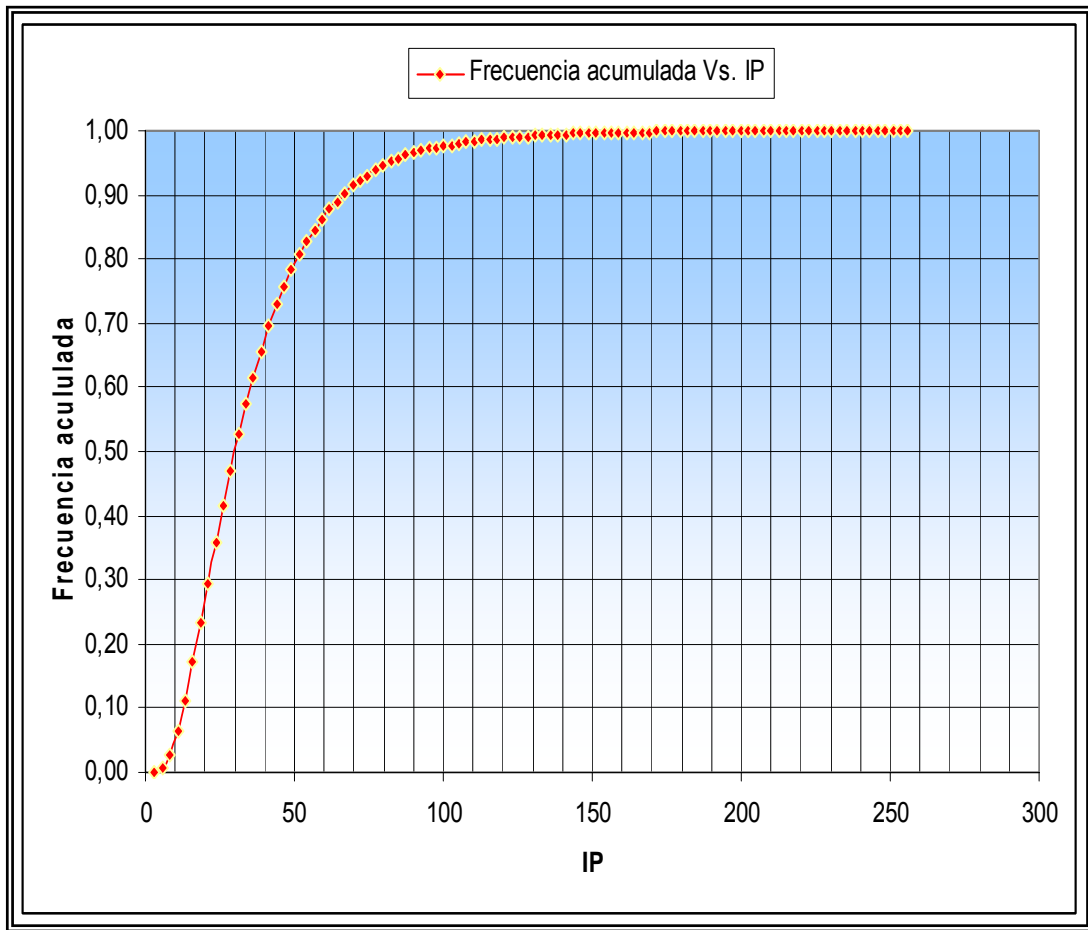
✓ Longitud 1000:

Figura 64. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia.Longitud De Pozo 1000 Metros



FUENTE: El autor

**Figura 65. Cuadro De Frecuencia Acumulada Versus Ip Para Una Longitud De Pozo De 1000 Metros.**



FUENTE: EL AUTOR

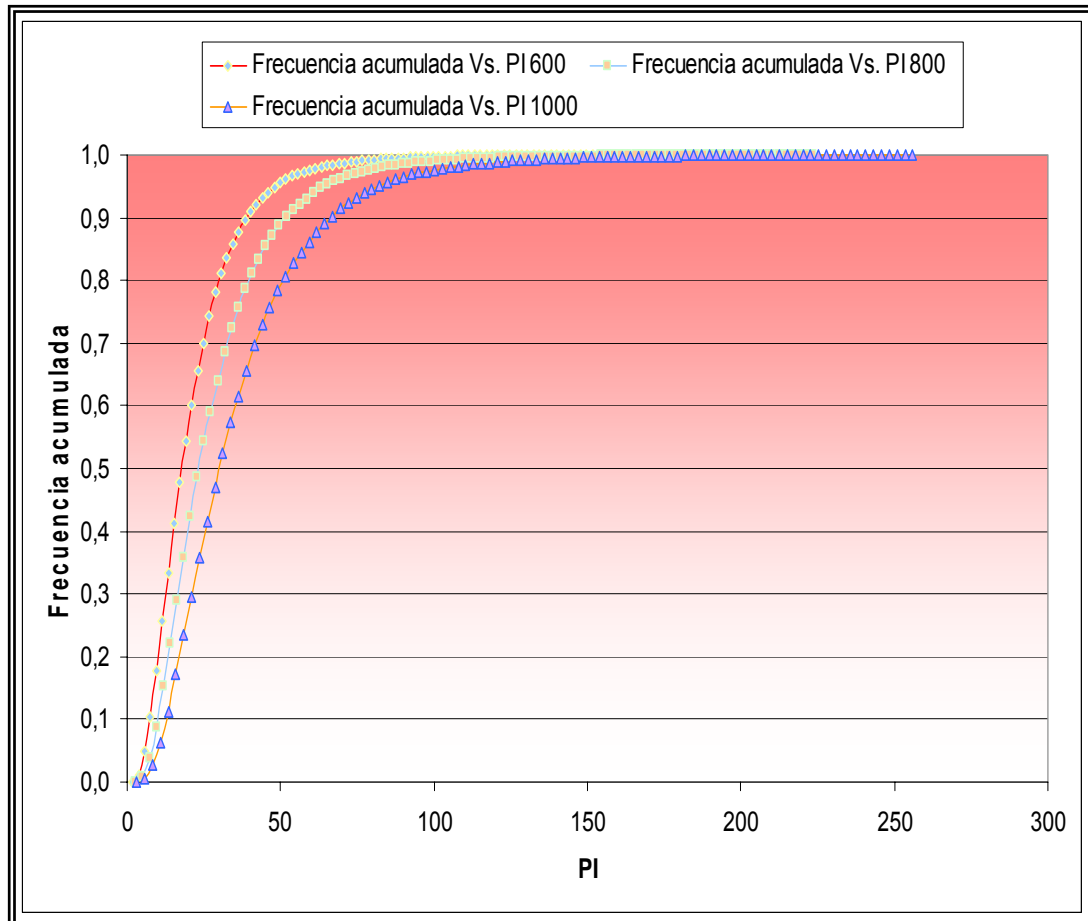
**Tabla 43. Percentiles 10, 50 Y 90 De Los Ip Hallados Para El Pozo Horizontal De Longitud 1000 Metros**

<i>Percentil</i>	<i>IP (bbl/psi)</i>
P10	13,34
P50	28,67
P90	66,98

FUENTE: EL AUTOR

Comparación de las 3 longitudes de pozo:

**Figura 66. Cuadro De Frecuencia Acumulada (Percentiles) Versus Índices De Producción. Comparación Entre Las Longitudes De Pozo 600, 800 Y 1000 Metros.**



FUENTE: EL AUTOR

#### **4.3 METODOLOGÍA PARA EL CÁLCULO DEL ÍNDICE DE PRODUCCIÓN EN POZOS HORIZONTALES USANDO SIMULACIÓN MONTECARLO. IMPLEMENTACIÓN EN LUMENAUT (SOFTWARE).**

En las anteriores secciones de la tesis se vio como se haría una corrida de simulación MC sin la ayuda de ningún software, solo usando Microsoft Excel. Este proceso puede ser un poco tedioso, y en ocasiones no se dispone del tiempo para hacer todos los pasos que anteriormente se enumeraron. Bajo esta óptica es preciso decir que existen numerosos software comerciales que

permiten correr la Simulación MC en un menor tiempo y que arrojaría todos los gráficos que se necesitan (CrystallBall, Lumenaut..etc), estos software contienen la mayoría de las funciones de distribución de probabilidad y permiten generar las muestras aleatorias solo escogiendo la función de distribución de probabilidad sin necesidad de recurrir a ecuaciones ni nada por el estilo.

En este caso el software a usar (Lumenut) tiene una licencia libre para estudiantes, es de fácil aplicación y automáticamente genera el grafico de distribución de probabilidad junto con las funciones de distribución triangular para cada variable.

La aplicación con el software no se correrá al campo colorado, ya que la confiabilidad de los datos es muy baja, razón por la cual lo que se hará será correr los datos en Lumenaut para el pozo horizontal y confrontar los datos a ver qué tan precisa es la metodología. Esto llevará a pensar que si los resultados obtenidos con el software y con la metodología son parecidos, la metodología aplicada puede tener una gran confiabilidad y ser aplicada a diferentes ámbitos. Se debe tener en cuenta que la función de distribución de densidad usada es la triangular, pero solo como aproximación inicial.

El software Lumenaut funciona como un add-in de Microsoft Excel, y su uso es fácil. A continuación se explicara cómo usar el software para hacer la corrida y calcular en IP de un yacimiento de crudo pesado en el cual se piensa perforar un pozo horizontal. Las corridas serán hechas para diferentes longitudes de pozo para observar el comportamiento del IP.

Inicialmente hay que saber cuáles son las entradas al modelo de simulación MC, que para este caso son las variables que se intervienen en el cálculo del IP, tales son beta, permeabilidad, espesor, viscosidad...etc. A cada una de estas variables se le generara su función de distribución triangular (aproximación inicial).

El índice de productividad para pozos horizontales se calcula por la ecuación (6):

$$IP = \frac{2\pi k_h h / (\mu B_o)}{\ln \left[ \frac{a + \sqrt{a^2 - (L/2)^2}}{(L/2)} \right] + \frac{\beta h}{L} \ln \left( \frac{\beta h}{2r_w} \right)} \quad (6)$$

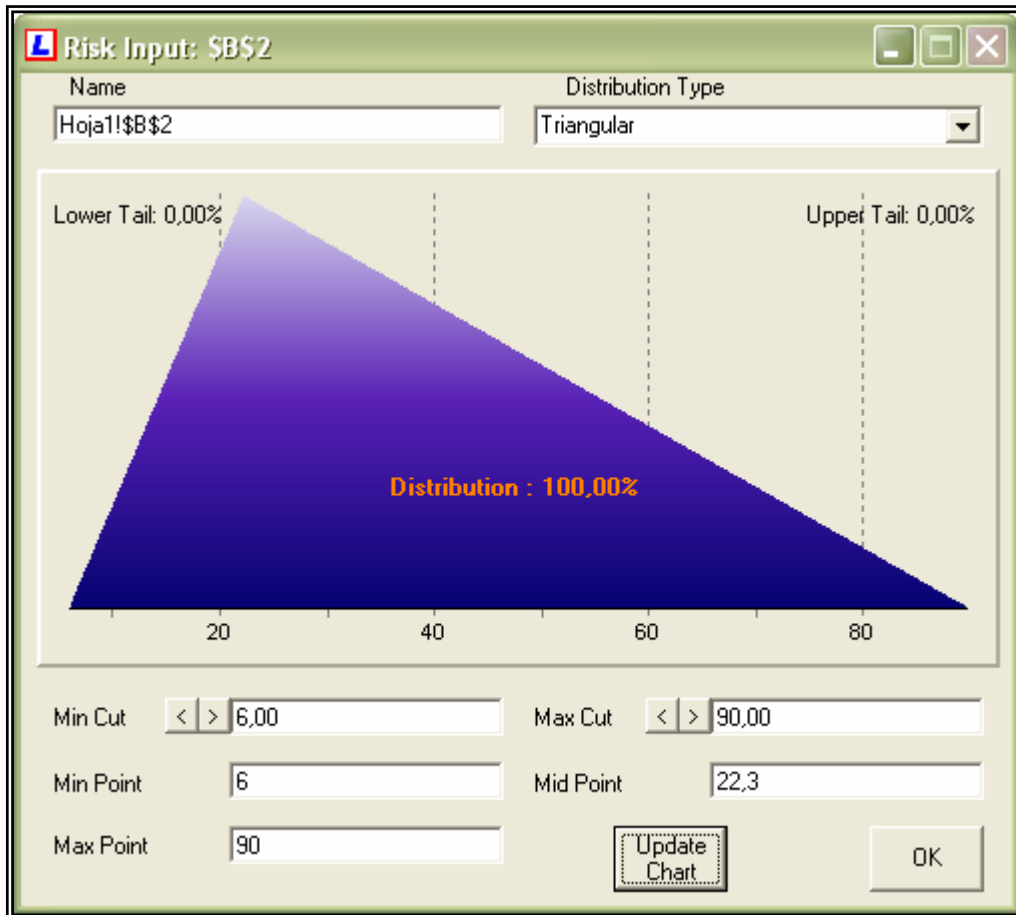
$$a = \frac{L}{2} \left[ \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{(L/2r_{eH})^4}} \right]^{\frac{1}{2}} \quad (7)$$

Donde el significado de cada variable ya fue explicado en el capítulo 1. Procederemos a generar los gráficos de la función de densidad triangular para cada una de las variables antes mencionadas, las cuales serán las mismas para todos los casos, lo único que se variará será la longitud de pozo (SPE-107669-MS-P)

Para generar las funciones de densidad triangular procedemos así: señalamos una celda, se selecciona herramientas de la ventana principal de Excel, dentro de herramientas se busca la opción que dice Lumenaut, simulación Montecarlo, Set as a model input, esto se hace para la permeabilidad horizontal, factor volumétrico de formación, viscosidad y espesor.

Por tanto, para *permeabilidad horizontal (Kh)*:

**Figura 67. Función De Densidad De Probabilidad Para La Permeabilidad Horizontal.**



FUENTE: SOFTWARE LUMENAUT

**Tabla 44. Valores Mínimo, Más Probable Y Máximo Para La Permeabilidad Horizontal**

<b>Kh(mD)</b>	<b>Mínimo</b>	6
	<b>Probable</b>	22,3
	<b>Máximo</b>	90

FUENTE: EL AUTOR

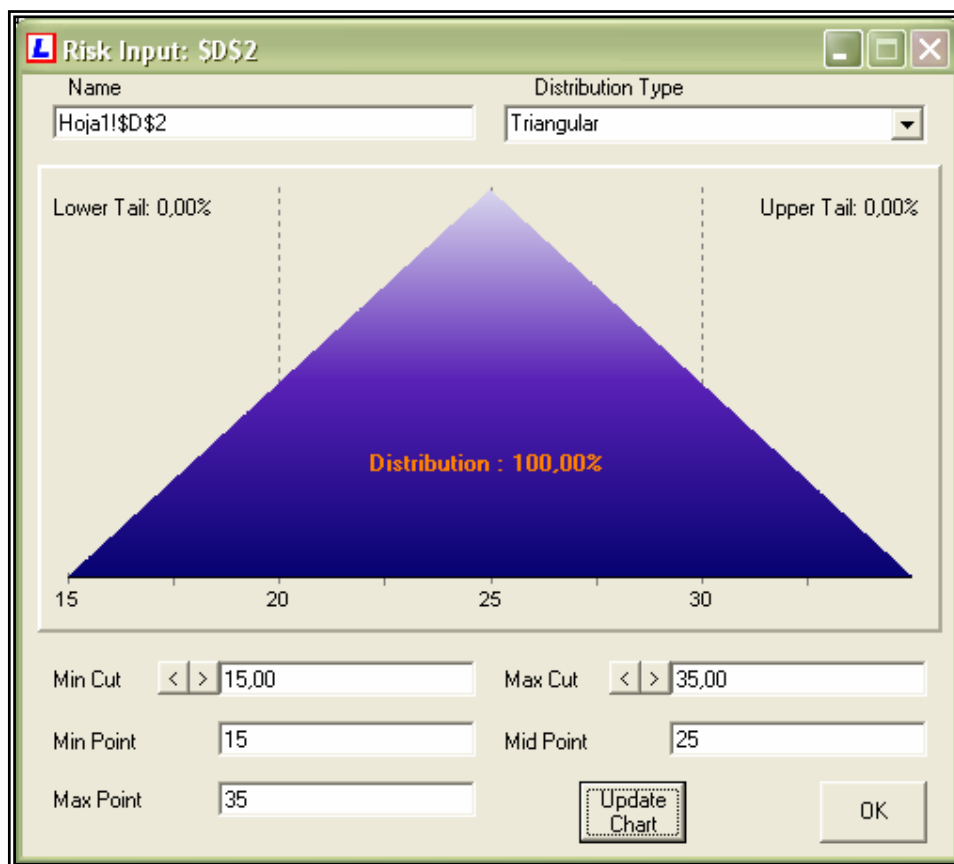
Como la función que se usó de densidad fue la triangular, se deben establecer los valores máximo, mínimo y más probable. El software construye

automáticamente el gráfico. Estos valores fueron los mismos usados en la corrida con Microsoft Excel.

Se procede ahora a hacer las gráficas para las variables restantes:

*Espesor:*

**Figura 68. Función De Densidad De Probabilidad Para El Espesor Neto.**



FUENTE: SOFTWARE LUMENAUT

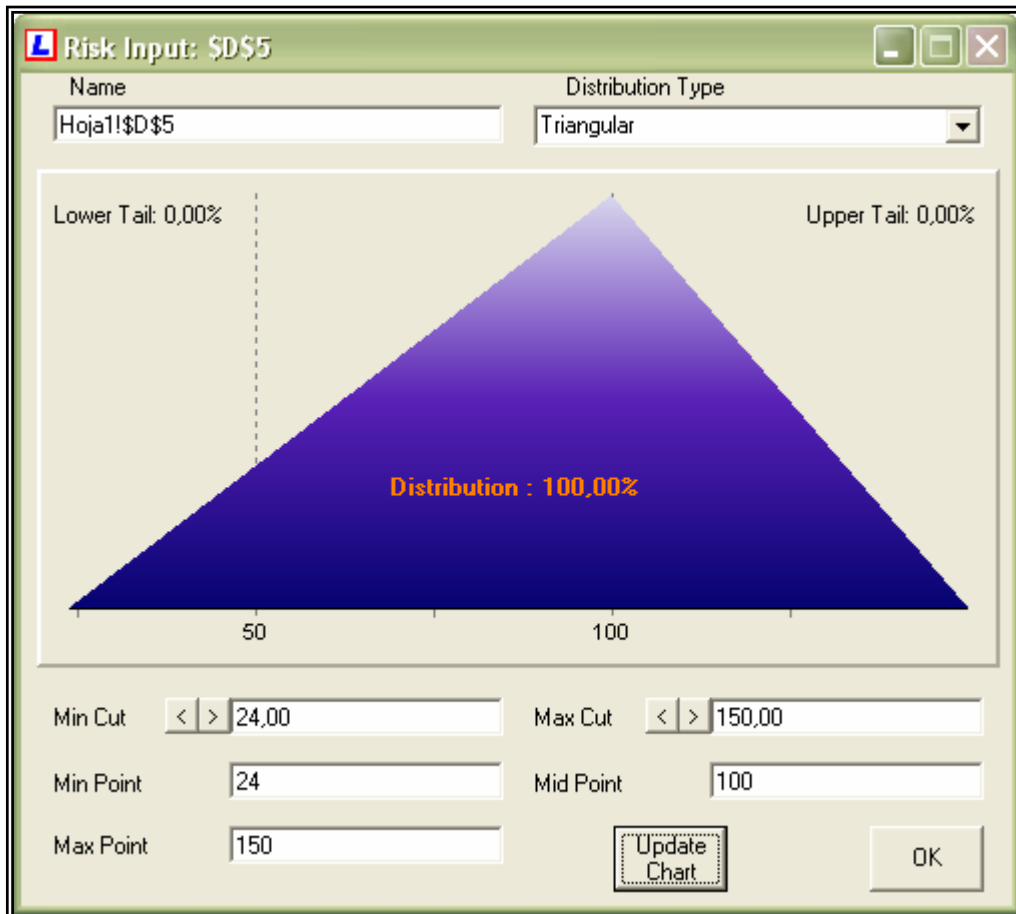
**Tabla 45. Valores Mínimo, Más Probable Y Máximo Para El Espesor Neto**

<b>Espesor neto de la formación(m)</b>	<b>Mínimo</b>	15
	<b>Probable</b>	25
	<b>Máximo</b>	35

FUENTE: EL AUTOR

Viscosidad:

Figura 69. Función De Densidad De Probabilidad Para La Viscosidad.



FUENTE: SOFTWARE LUMENAUT

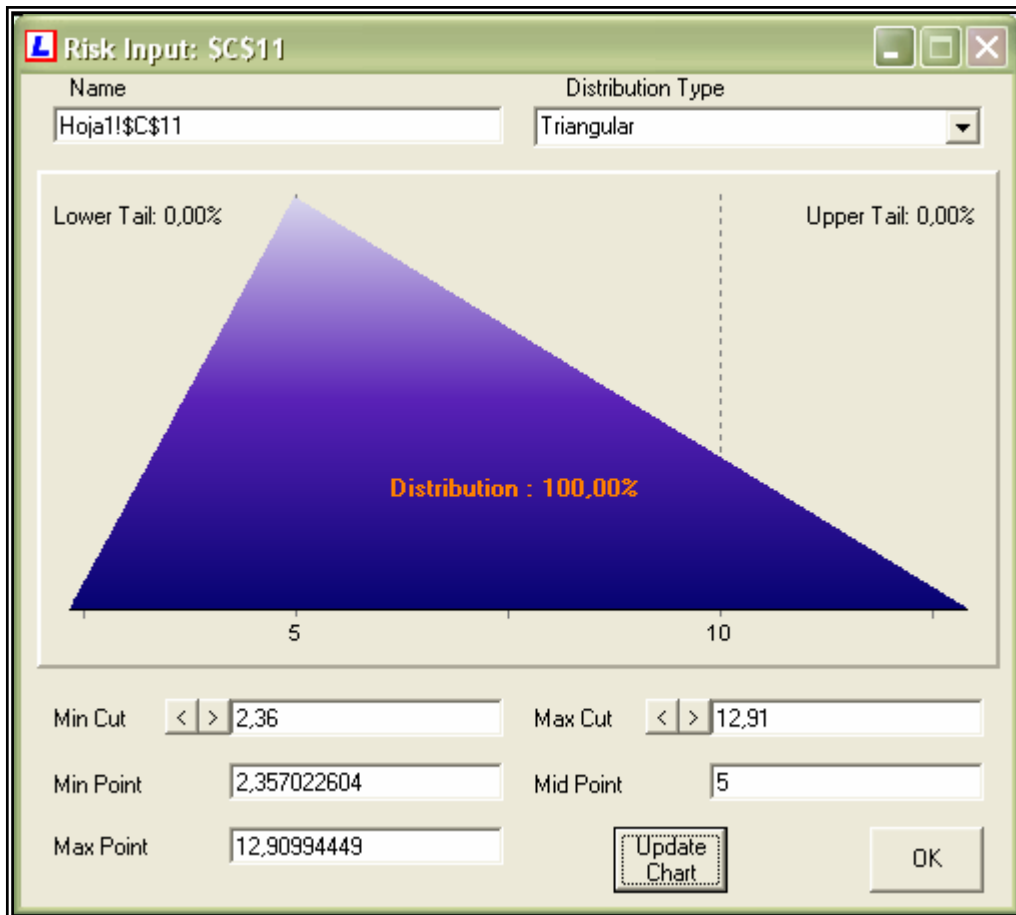
Tabla 46. Valores Mínimo, Más Probable Y Máximo Para La Viscosidad

Viscosidad(Cp)	Mínimo	24
	Probable	100
	Máximo	150

FUENTE: EL AUTOR

$$\beta = \left( \frac{k_h}{k_v} \right)^{0,5}, \text{ Beta:}$$

**Figura 70. Función De Densidad De Probabilidad Para Beta.**



FUENTE: SOFTWARE LUMENAUT

**Tabla 47. Valores Mínimo, Más Probable Y Máximo Para Beta**

<b>B</b>	<b>Mínimo</b>	2,4
	<b>Probable</b>	5,0
	<b>Máximo</b>	12,9

FUENTE: EL AUTOR

Una vez hecho esto, se selecciona una celda que será el output o celda de salida. Para tal objetivo, luego de elegir una celda cualquiera, se combinan las variables de entrada (inputs del modelo) de acuerdo a la ecuación que se desee modelar que en este caso es la ecuación para un pozo horizontal (Ecuación 6), es decir, los input del modelo se multiplican o dividen de acuerdo al lugar que ocupen en la ecuación del IP. Luego se selecciona en herramientas, la opción que dice Lumenaut, simulación Montecarlo, Set as a model output y se colocan los demás valores tales como  $a$ , que es una constante obtenida anteriormente para cada longitud,  $L$  y  $r_w$ , que también son constantes.

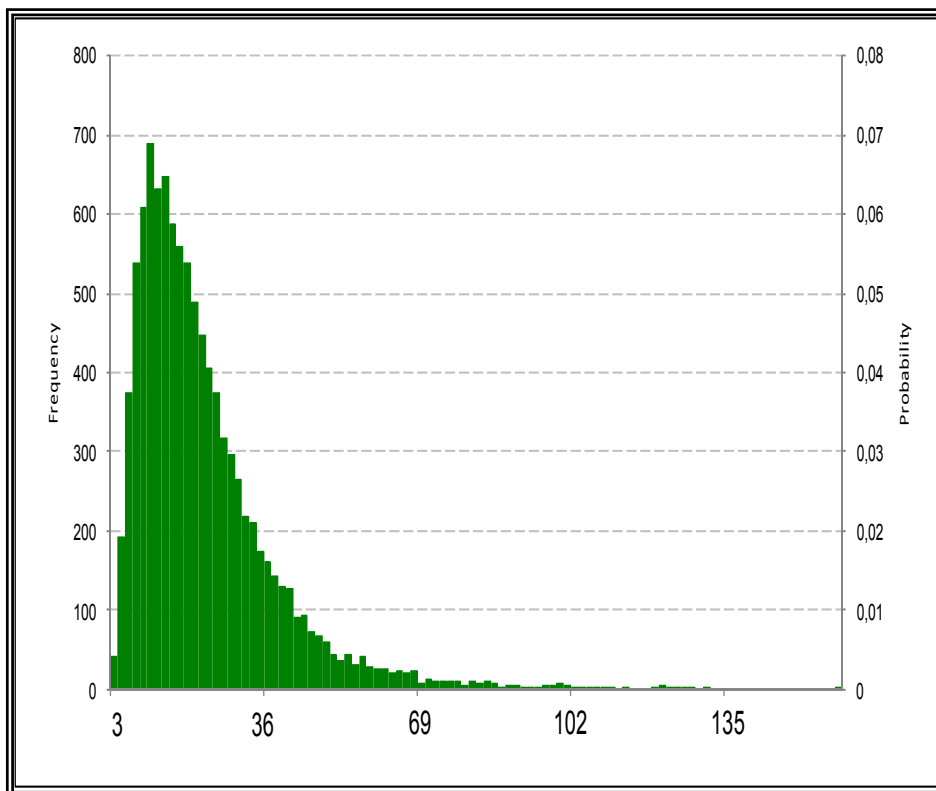
Una vez hecho esto se procede a hacer la corrida con el software. La ecuación para  $L=600$  reducida a su mínima expresión es:

$$IP_{600} = \frac{2\pi k_h h / (\mu B_o)}{1,3552077 + \frac{\beta h}{600} \ln(6,4016 * \beta h)} \quad (104)$$

Una vez indicados los inputs, y los outputs de la ecuación, se procede a correr la simulación. El procedimiento para lograr dicho fin es seleccionar en herramientas, la opción que dice Lumenaut, Run Monte, escoger el número de iteraciones y elegir la opción Run. En la corrida que se muestra a continuación fueron 10000 iteraciones para todos los casos ( $L=600$ ,  $L=600$  y  $L=1000$ ). Seguidamente se muestran los resultados:

Longitud L=600 metros.

**Figura 71. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 600 Metros**



FUENTE: El autor

**Tabla 48. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitu De Pozo 600 Metros.**

Percentil	IP (bbl/psi)
0%	1,5900
10%	7,4100
20%	10,0500
30%	12,4400
40%	14,9000
50%	17,7300
60%	20,9100
70%	24,7300
80%	30,0200
90%	39,2000
100%	158,9400

FUENTE: El autor

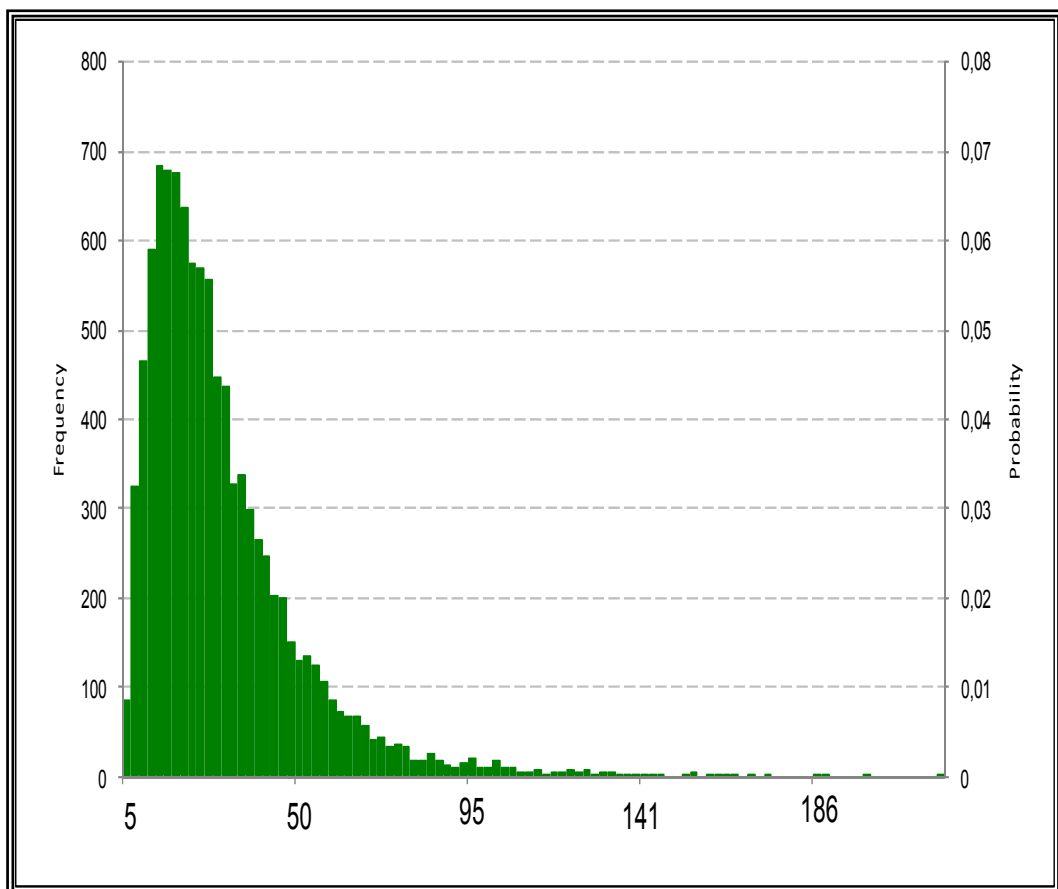
De igual manera se sacó una ecuación simplificada para L=800 y L=1000, que se presentarán a continuación. A cada una de las ecuaciones se les corrió su respectiva simulación calculando los P10, P50 y P90.

$$IP\ 800 = \frac{2\pi k_h h / (\mu B_o)}{1,04406 + \frac{\beta h}{800} \ln(6,4016 * \beta h)} \quad (105)$$

$$IP\ 1000 = \frac{2\pi k_h h / (\mu B_o)}{10,793 + \frac{\beta h}{1000} \ln(6,4016 * \beta h)} \quad (106)$$

*Longitud L=800 metros.*

**Figura 72. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 800 Metros**



**FUENTE:** El autor

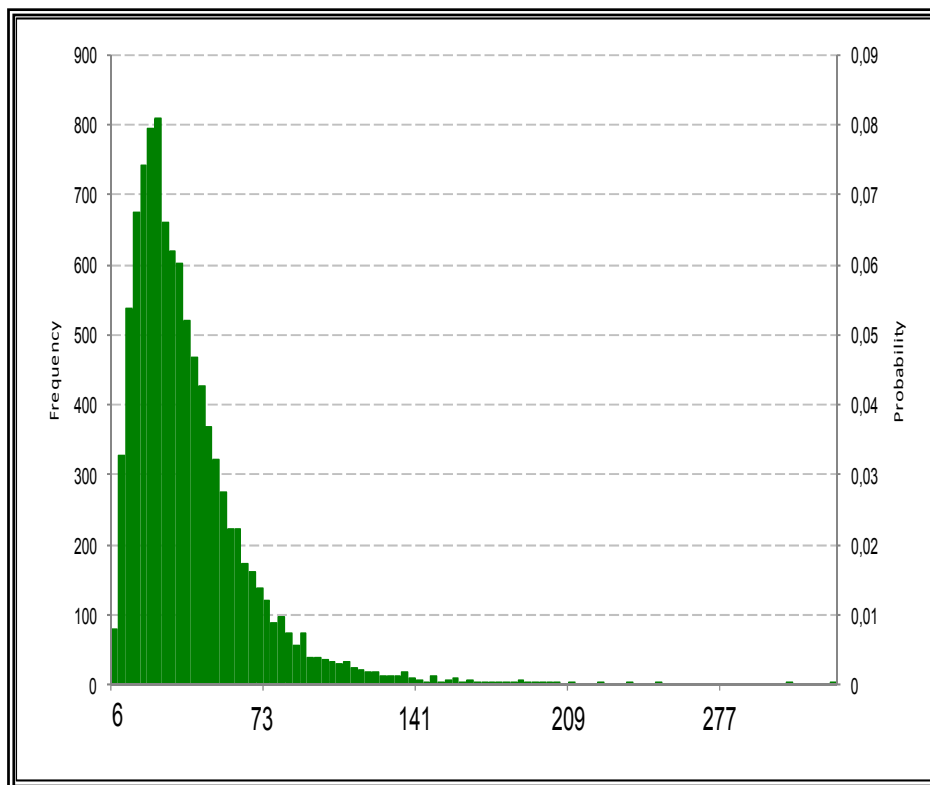
**Tabla 49. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitud De Pozo 800 Metros.**

Percentil	IP (bbl/psi)
0%	2,8000
10%	9,8200
20%	13,2100
30%	16,3300
40%	19,6600
50%	23,4200
60%	27,3500
70%	32,8300
80%	40,1200
90%	52,6600
100%	219,2600

FUENTE: El autor

Longitud  $L=1000$ .

**Figura 73. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia. Longitud De Pozo 1000 Metros**



FUENTE: El autor

**Tabla 50. Percentil 10, 50 Y 90 De Los Índices De Producción Obtenidos Para Longitu De Pozo 1000 Metros.**

Percentil	IP (bbl/psi)
0%	2,68
10%	12,63
20%	17,28
30%	21,35
40%	25,41
50%	30,41
60%	35,75
70%	42,59
80%	51,75
90%	67,59
100%	324,98

**FUENTE:** El autor

**4.4. CÁLCULO DEL ÍNDICE DE PRODUCCIÓN EN POZO COLORADO 38, ZONA B. IMPLEMENTACIÓN EN HOJA DE CÁLCULO. UTILIZACIÓN DE FUNCIONES DE DENSIDAD DE PROBABILIDAD DE ACUERDO A TABLA No. 8.**

Se decidió hacer una corrida más para los intervalos productores del pozo Colorado 38 en la zona B. En esta corrida se cambiarán las funciones de distribución para algunas variables, tales como son la permeabilidad y los espesores para las diferentes zonas. Teniendo en cuenta lo anterior, las funciones de distribución que se usaron en la nueva corrida son las siguientes:

**Tabla 51. Funciones De Distribución A Usar Para Una Nueva Corrida, De Acuerdo A Tabla 8.**

<b>Permeabilidad</b>	Log Normal
<b>Espesor</b>	Log Normal
<b>Viscosidad</b>	Triangular
<b>Factor Volumétrico</b>	Triangular

**FUENTE:** TABLA 8

El proceso seguido fue el siguiente para todos los intervalos:  
Se sacó el logaritmo natural a los valores de permeabilidad y espesor para cada intervalo productor. Como se muestra en la tabla 52.

**Tabla 52. Media Geométrica Y Desviación Estandar Del Logaritmo Natural De Valores De Permeabilidad**

Logaritmo Natural (K)	Media Geométrica	Desviacion Estándar
2,3843	2,9679	0,1884
2,9556		
2,9492		
2,9861		
3,0040		
2,9473		
3,0257		
3,1184		
3,1749		
3,1153		
3,1162		
3,1566		

**FUENTE:** EL AUTOR

Esto se hizo para cada intervalo de interés tanto para la permeabilidad como para los espesores.

Luego, a estos valores de logaritmo natural de la permeabilidad y espesor, se le saca la media geométrica y desviación estándar, valores

necesarios para poder usar la función de distribución de densidad Log-Normal.

Esta función de distribución de densidad Log-Normal es de la forma:

$$= \text{DISTR.LOG.INV}(\text{ALEATORIO}(); \text{Media Geométrica}; \text{desviación Estándar})$$

Donde la función Aleatorio () da el valor de un Número aleatorio, y la función de distribución Log normal devuelve el valor de la variable evaluada (Ya sea permeabilidad o espesor), teniendo en cuenta los valores de desviación estándar y media geométrica del logaritmo natural la permeabilidad y el espesor para cada intervalo en la zona B.

Tomando como ejemplo el intervalo 1812 – 1829, los datos montados al modelo quedan de la siguiente forma:

**Figura 74. Resultado De Arrastrar Las Celdas Para Hallar Un Nuevo Valor De Ip Intervalo 1812-1829 Ft Usando Un Función De Densidad Triangular Para La Viscosidad Y El Factor Volumetrico, Y Una Función De Densidad Log-Normal Para La Permeabilidad Y El Espesor Neto, Pozo Colorado 38, Zona B**

Datos de entrada					
K(mD)	Media	2,967	Espesor del intervalo productor(ft)	Media	2,136
	Desviación Estándar	0,1884		Desviación Estándar	0,74
Viscosidad(Cp)	Mínimo	0,913722715	Eo(Els/STB)	Mínimo	1,187589305
	Probable	0,914083917		Probable	1,187590332
	Máximo	0,914445054		Máximo	1,187591368
Radio del pozo (ft)	0,276				
Re (ft)	98,99				
INTERVALO 1812-1829					
Valor aleatorio de K	Valor aleatorio de h	Valor aleatorio de viscosidad	Valor aleatorio de Eo	PI COLO 38 ZONA B	
17,80385205	15,1747622	0,914181354	1,187589888	0,343387644	
24,94300842	7,661545061	0,914128279	1,187590568	0,242906077	
19,6779632	12,1126255	0,913853681	1,187590717	0,303055413	
18,85648698	8,399750663	0,914153804	1,187590799	0,201320572	
21,76937973	2,597424441	0,914038074	1,187590808	0,071879481	

**FUENTE:** El autor

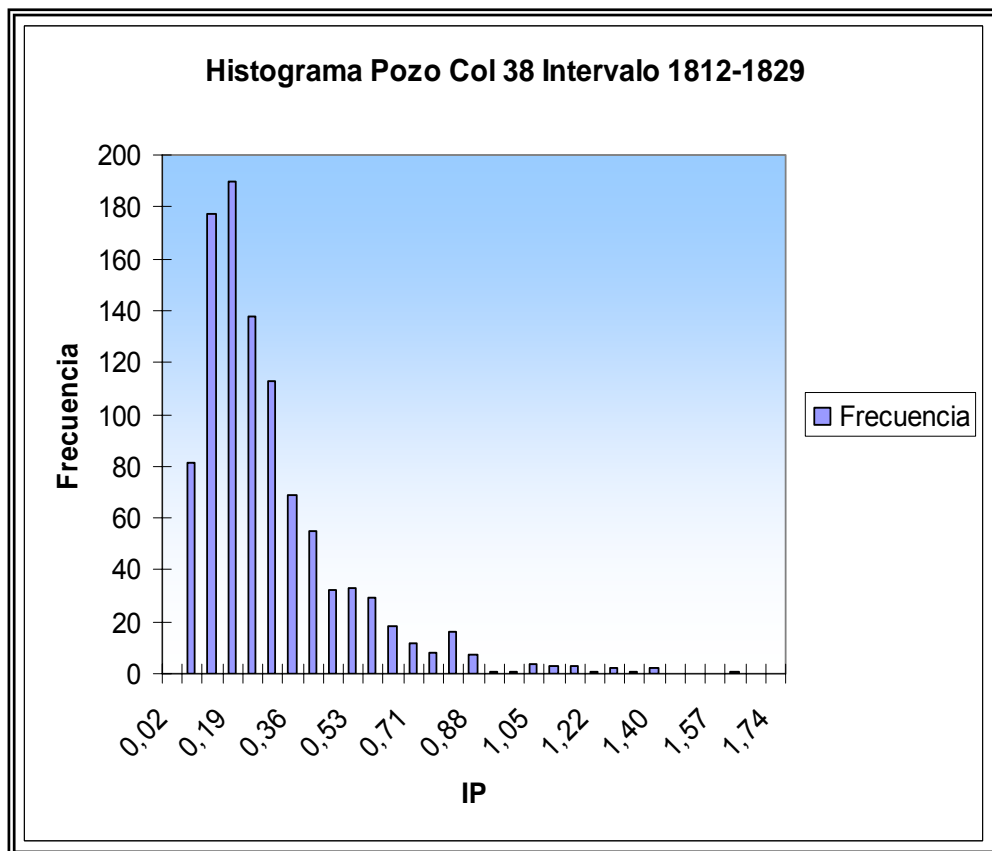
Luego se procedió a combinar las variables de acuerdo a la ecuación 5:

$$IP = \frac{kh}{141.2 * \mu_o B_o * \left[ \text{Ln} \left( \frac{R_e}{r_w} \right) - \frac{3}{4} \right]} \quad (5)$$

Donde el significado de cada variable ya fue definido en el capítulo 1. Seguidamente se escogen todas las celdas involucradas en el cálculo y se despliegan hacia abajo tantas veces como iteraciones se requiera, idénticamente como se anotó en las metodologías anteriores implementadas en las hojas de cálculo. Igualmente se trabajarán 1000 iteraciones. Los resultados obtenidos se muestran a continuación:

- ✓ Intervalo 1812-1829 ft

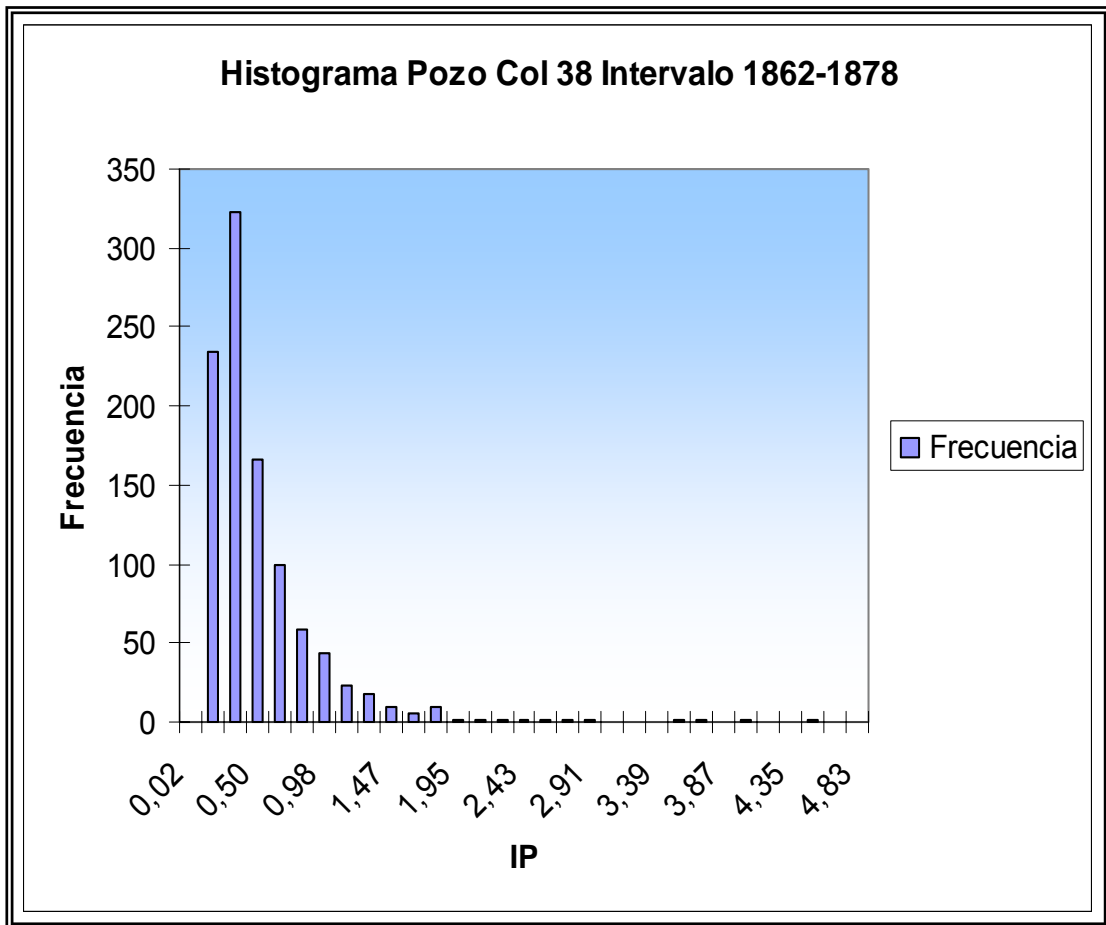
**Figura 75. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1812-1829, Pozo Colorado 38, Zona B. Corrida 2**



*IP esperado: 0.19 bbl/psi*

✓ Intervalo 1862-1878 ft

**Figura 76. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1862-1878, Pozo Colorado 38, Zona B. Corrida 2**

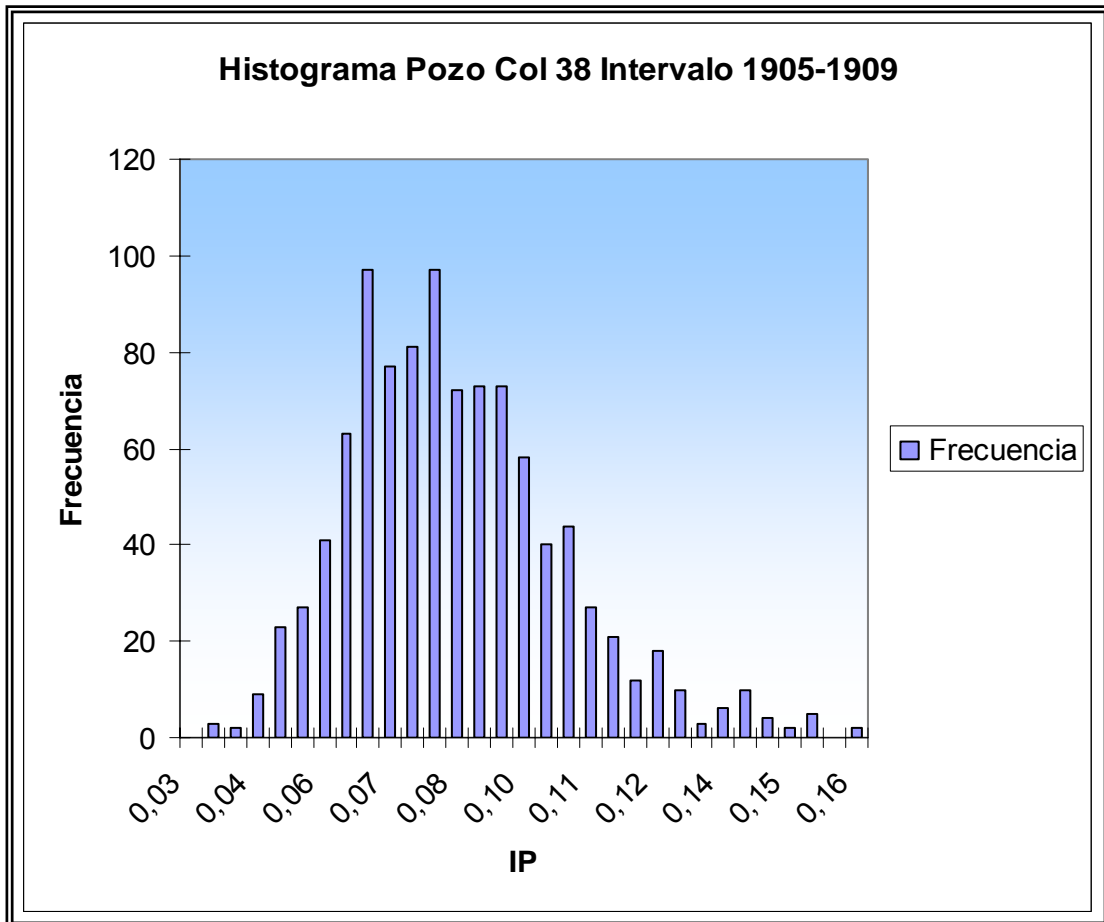


**FUENTE:** El autor.

*IP esperado: 0.3442 bbl/psi*

✓ Intervalo 1905-1909 ft

**Figura 77. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1905-1909, Pozo Colorado 38, Zona B. Corrida 2**

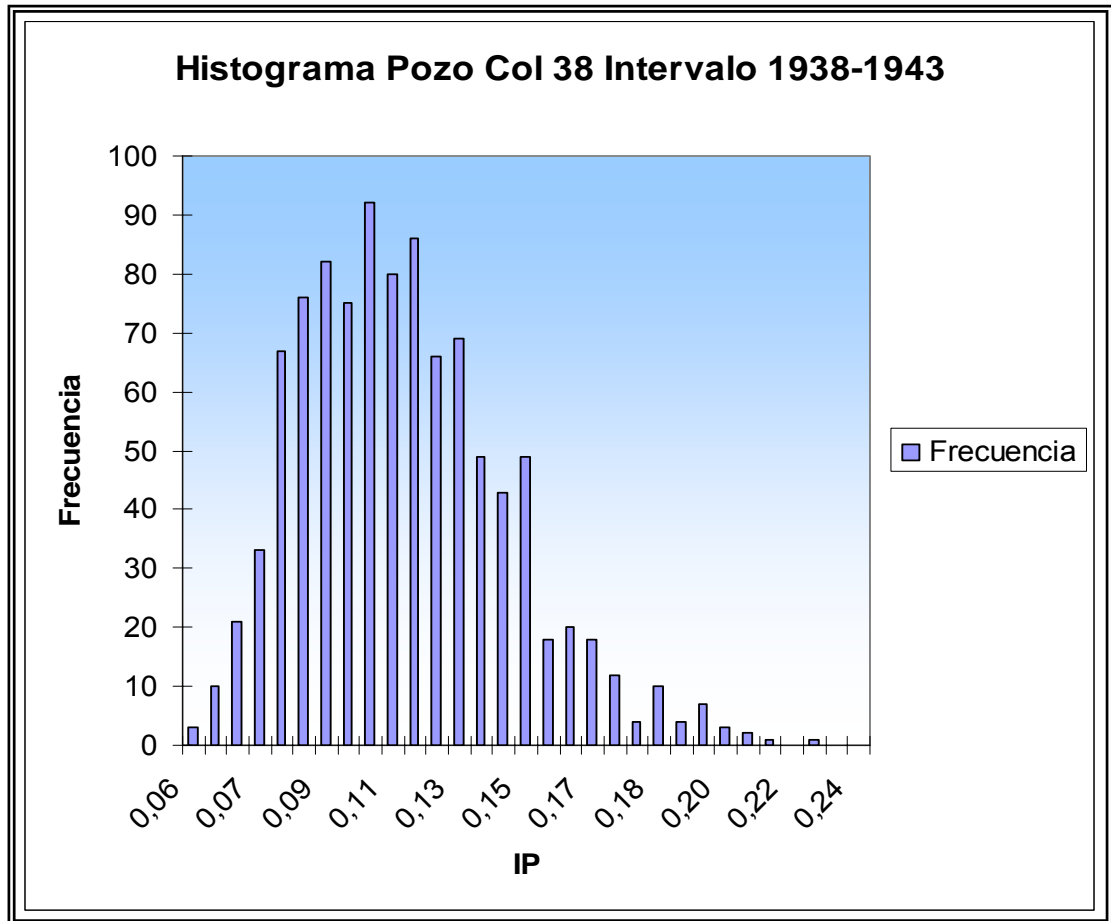


FUENTE: El autor.

*IP esperado: 0.080 bbl/psi*

✓ Intervalo 1938-1943 ft

**Figura 78. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 1938-1943, Pozo Colorado 38, Zona B. Corrida 2**

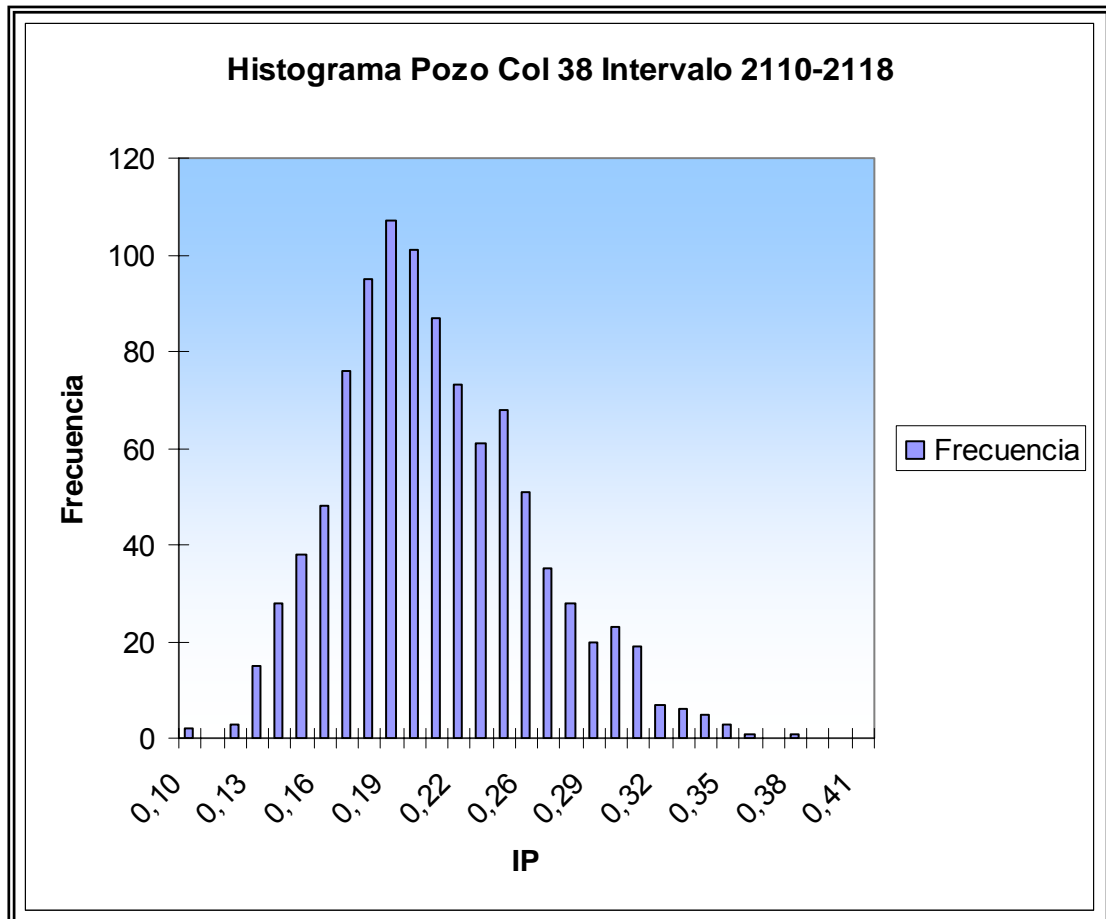


FUENTE: El autor.

*IP esperado: 0.1053 bbl/psi*

✓ Intervalo 2110-2118 ft

**Figura 79. Histograma Que Relaciona Los Puntos Medios De Clase De Ip Con Su Frecuencia, Intervalo 2110-2118, Pozo Colorado 38, Zona B. Corrida 2**



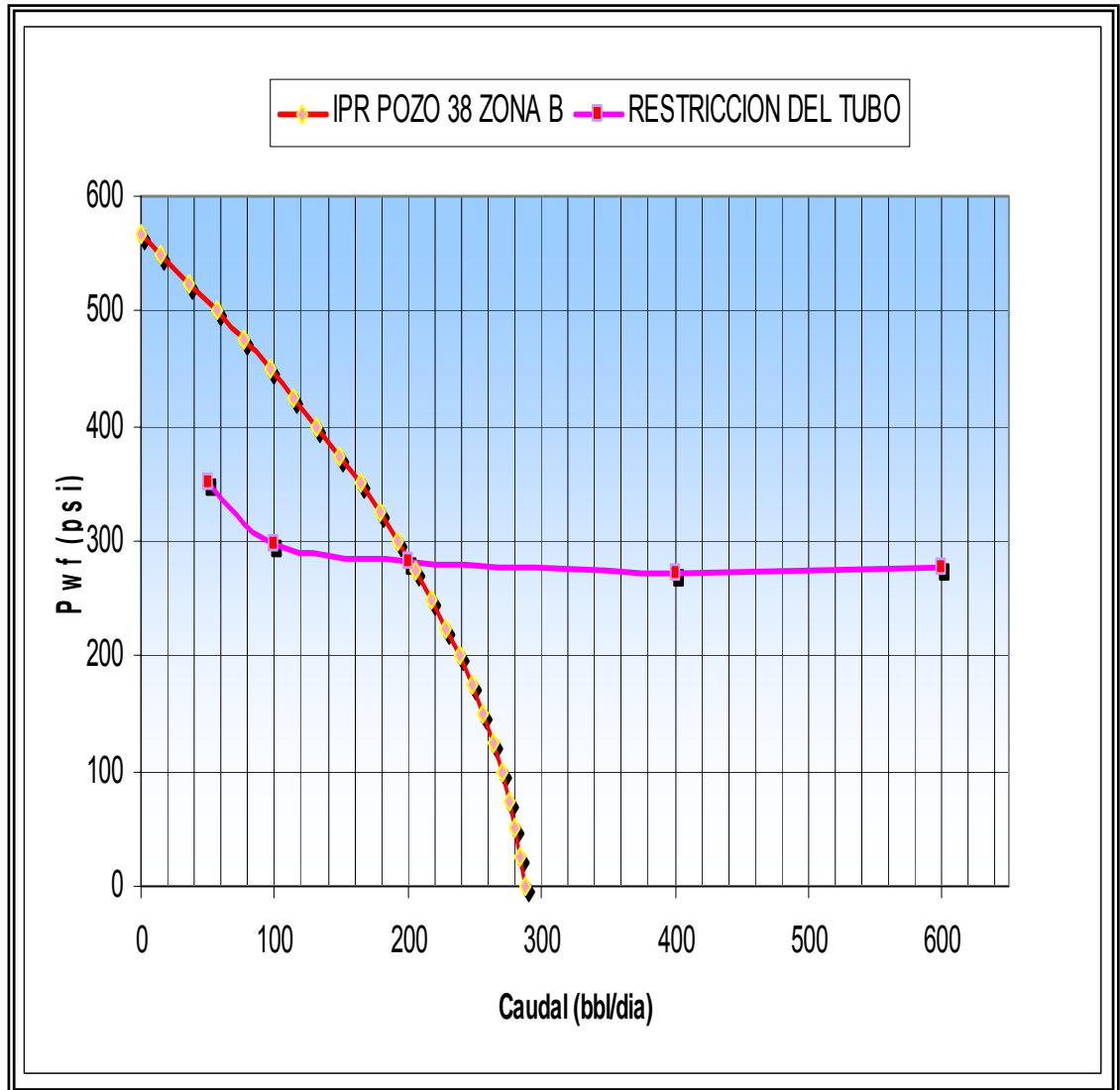
**FUENTE:** El autor.

***IP esperado: 0.19 bbl/psi***

Finalmente, se sumaron los IP de los intervalos, y se obtuvo el IP de la zona B.

***IP zona B= 0,91 bbl/psi*** con este valor de IP se construyó la curva IPR para el intervalo, la cual se anexa a continuación.

**Figura 80. Gráfico De Pwf Vs Caudal Mediante La Ecuación De Vogel.  
Restricciones Del Tubo. Pozo Colorado 38, Zona B. Corrida 2**



**FUENTE:** El autor

**Tabla 53. Presiones De Fondo Fluyendo Como Datos De Entrada Para La Ecuación De Vogel Y Obtención De Caudales (lpr) Pozo Colorado 38, Zona B. Corrida 2**

Pwf (psi)	qo (bbl/día)
566	0
550	14,4367
525	36,2586
500	57,1836
475,9	76,5062
475	77,2116
450	96,3428
425	114,5771
400	131,9144
375	148,3548
350	163,8983
325	178,5449
300	192,2946
275	205,1473
250	217,1032
225	228,1621
200	238,3242
175	247,5893
150	255,9575
125	263,4288
100	270,0032
75	275,6806
50	280,4611
25	284,3448
0	287,3315

**FUENTE:** El autor

## 5. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

A continuación se analizarán los resultados obtenidos en todas las corridas realizadas con el objetivo de darle una correcta explicación a los índices de producción de cada zona del pozo colorado 38 y del pozo horizontal. Se examinarán los efectos de usar una hoja de cálculo ó un software. Se compararán los índices de producción de la zona B al usar la función de densidad de probabilidad triangular para todas las variables implicadas en el cálculo del mismo y al utilizar una función de densidad de probabilidad log-normal para describir únicamente la permeabilidad y el espesor neto.

### 5.1 CÁLCULO DE POTENCIALES DE PRODUCCIÓN PARA EL POZO COLORADO 38. CORRIDA EN EXCEL

Inicialmente se calculó el índice de productividad del pozo colorado 38, usando la función de distribución Triangular en la generación de los números aleatorios para todas las variables involucradas en el cálculo del mismo ( $B_o, k, \mu, h$ ). La razón de haber escogido esta función de distribución de probabilidad, para este caso, fue la facilidad en su aplicación, sumado al hecho que se puede usar como una aproximación inicial cuando no se tiene mayor certeza respecto a cuál es la mejor función de distribución para describir la variable. De acuerdo a lo anterior, los resultados obtenidos fueron:

- ✓ Zona B

**Tabla 54. Índices De Producción Obtenidos Para La Zona B Del Pozo Colorado 38. Implementación En Hoja De Cálculo Usando Una Función De Densidad De Probabilidad Triangular Para Describir Todas Las Variables.**

INTERVALO (ft)	INDICE DE PRODUCTIVIDAD (bbl/psi)
1812-1829	0.2671
1862-1878	0.4508
1905-1909	0.0821
1938-1943	0.1049
2110-2118	0.2146
<b>IP total aportado</b>	<b>1,1195</b>

**FUENTE:** EL AUTOR

✓ Zona C

**Tabla 55. Índices De Producción Obtenidos Para La Zona C Del Pozo Colorado 38. Implementación En Hoja De Cálculo Usando Una Función De Densidad De Probabilidad Triangular Para Describir Todas Las Variables.**

INTERVALO (ft)	INDICE DE PRODUCTIVIDAD (bbl/psi)
3383-3386	0.0356
3453-3459	0.0615
3488-3506	0,1644
3705-3713	0.2943
4006-4018	0,1753
<b>IP total aportado</b>	<b>0,7311</b>

**FUENTE:** EL AUTOR

Para este pozo se tiene una presión de fondo fluyendo ( $P_{wf}$ ) de **476 psi**, lo cual permite tener un caudal para la zona B de **94 bbl/día** y para la zona C de **284 bbl/día**. A estas condiciones, el **máximo potencial de entrega de este pozo es 352 bbl/día para la zona B** y de **397,7 bbl/día**

**para la zona C.** Se puede ver que a pesar que el potencial en la zona C es menor que en la zona B, el caudal que entregaría el pozo a la misma  $P_{wf}$  es mayor en C que en B, y esto es debido a que la presión de yacimiento en la zona C es de 980 psi, mientras que en la zona B, la presión de yacimiento es de 566 psi, lo cual da una idea del efecto que tiene la presión del yacimiento en la Entregabilidad del pozo, y es la razón final por la cual siempre se busca mantener la presión del yacimiento lo más alta posible.

**5.2 CALCULO DE IP PARA EL POZO COLORADO 38 ZONA B REPRESENTANDO LA VISCOSIDAD Y EL FACTOR VOLUMÉTRICO CON UNA FUNCIÓN DE DENSIDAD TRIANGULAR Y LA PERMEABILIDAD Y EL ESPESOR NETO CON UNA FUNCIÓN DE DISTRIBUCION DE PROBABILIDAD LOG-NORMAL (2) Y COMPARACIÓN CON LOS RESULTADOS OBTENIDOS CUANDO SE USÓ LA DISTRIBUCION TRIANGULAR (1) PARA LA DESCRIPCION DE TODAS LAS VARIABLES.**

**Tabla 56. Comparación De Ip Al Describir Las Variables Según Tabla 8, Y Al Usar La Función Triangular Para Describir Todas Las Variables.**

<b>INTERVALO (ft)</b>	<b>INDICE DE PRODUCTIVIDAD (bbl/psi) (2)</b>	<b>INDICE DE PRODUCTIVIDAD (bbl/psi) (1)</b>
1812-1829	0.1900	0.2671
1862-1878	0.3442	0.4508
1905-1909	0.0800	0.0821
1938-1943	0.1053	0.1049
2110-2118	0.1900	0.2146
<b>IP total aportado</b>	0.91	1,12

**FUENTE: EL AUTOR**

De acuerdo a una comparación de los IPR tenemos lo siguiente:

**Tabla 57. Comparación De Ipr'S Corrida 1 Y Corrida 2**

Pozo colorado 38	Caudal esperado a $P_{wf}=476\text{psi}$ (bbl/día)	Máximo caudal que podría entregar el pozo (bbl/día)
Zona B (2)	76,5	287
Zona B (1)	94	352

**FUENTE:** EL AUTOR

De acuerdo a la tabla 57 se puede observar que se presenta una variación en los IPR teniendo en cuenta su dependencia de los índices de productividad. A continuación, se trata de dar una explicación a dicho fenómeno.

Un análisis comparativo al calcular el índice de producción usando la función de distribución triangular para describir todas las variables involucradas, con un cálculo posterior utilizando la función de distribución de probabilidad Log-Normal, para describir la permeabilidad y el espesor neto, conduce a lo siguiente:

Claramente, la mayor variación se produjo en el intervalo 1862-1878 ft. De acuerdo a se podría decir tentativamente que la corrida 2 puede ser un poco más confiable, pues esta es el fruto de la experiencia de algunos autores quienes después de muchos trabajos han optado por afirmar cual sería la mejor función de densidad de probabilidad para describir algunas de la variables involucradas en el cálculo de algunos parámetros del pozo o del yacimiento. De acuerdo a la referencia de la tabla 8, aunque son funciones de distribución recomendadas para describir las variables implicadas en el cálculo volumétrico de reservas, estas mismas funciones se pueden utilizar con buena aproximación para representar esas mismas variables pero involucradas en el cálculo de otros parámetros

como el índice de producción de los pozos. Cabe aclarar que no por esto se le debe quitar credibilidad a la corrida 1 puesto que cuando no se tenga certeza de cuál es la mejor función de distribución de probabilidad para describir una u otra variable, el mejor ejercicio debe consistir en representar todas las variables por la función de distribución triangular. Como podemos observar en la tabla 56 no existe mucha variación entre la corrida 1 y la 2, situación que sustenta lo antes dicho.

### 5.3. CÁLCULO DE POTENCIALES DE PRODUCCIÓN PARA EL POZO HORIZONTAL DE PAPER SPE 107669-MS-P:

**Tabla 58. Comparación Entre Los Ip Hallados Por Los Autores Y Los Que Muestra El Paper Spe 107669**

LONGITUD DE POZO (ft)	IP Excel	IP Paper SPE
600	17,2200	17,1100
800	22,8600	21,8800
1000	28,6700	27,0700

**FUENTE:** LOS AUTORES

Los resultados obtenidos mediante la aplicación de la metodología propuesta, fueron comparados con los resultados arrojados por el artículo SPE 107669-MS-P. Se puede observar que la variación es muy poca, pues ambas corridas se llevaron a cabo en hojas de cálculo. Los datos usados para ambas corridas fueron los mismos, pero cabe aclarar que en el paper solo hacían las corridas y sin explicar como funciona la simulación montecarlo, caso abordado en esta tesis.

#### 5.4 COMPARACION CORRIDA POZO HORIZONTAL EXCEL Y CON SOFTWARE PARA SIMULACION MONTECARLO, LUMENAUT (DISTRIBUCIÓN TRIANGULAR EN TODOS LOS CASOS):

**Tabla 59. Comparación Entre Los Ip Hallados Excel Y Los Hallados En Lumenaut. Confrontación Con Artículo 107669**

LONGITUD DE POZO (m)	IP Excel	IP Paper SPE	IP Lumenaut
600	17,2200	17,1100	17,7200
800	22,8600	21,8800	23,4200
1000	28,6700	27,0700	30,400

**FUENTE:** LOS AUTORES

Como se puede observar, con el Software Lumenaut se obtuvo muy poca variación con los resultados obtenidos anteriormente en las corridas hechas en el artículo y en el documento en Excel. Lo anterior le confiere validez a las metodologías explicadas. Las pequeñas diferencias mostradas son consecuencias de método estadístico mismo.

## 6. CONCLUSIONES

- Es importante que al momento de calcular el índice de productividad las variables involucradas tengan la mínima incertidumbre posible, ya que de esta forma los cálculos del IP se verán menos afectados y serán más precisos. Es de destacar que las incertidumbres se encuentran presente en las propiedades de la roca tales como espesor neto y permeabilidad absoluta, y en las propiedades de los fluidos como la viscosidad y el factor volumétrico de formación del aceite. Existirán yacimientos en los cuales estas propiedades no tienen mucha variación y por tanto pueden ser consideradas constantes (su incertidumbre puede considerarse baja).
- La simulación Montecarlo es una herramienta muy poderosa de cálculo, ya que esta combina conceptos estadísticos (muestreo aleatorio) con la capacidad que tienen los ordenadores para generar números pseudo aleatorios y automatizar cálculos. Si en un yacimiento se tiene gran variación en las propiedades, como en los casos presentados en este trabajo, la simulación Montecarlo se convierte en una gran ayuda para disminuir la incertidumbre en el cálculo del IP y dando una medida de tendencia del mismo que se ve reflejada en los Histogramas.
- Para flujo monofásico la grafica  $P_{wf}$  vs Caudal da una línea recta cuya pendiente es equivalente a  $(1/P)$ , esto tiene que ver con que no hay desprendimiento de gas en el yacimiento. Para flujo bifásico como fueron los casos estudiados en este trabajo lo que se obtiene son curvas que reflejan la entregabilidad del pozo y cuyo caudal esperado es menos a que si fuera flujo monofásico, dado el desprendimiento de gas en el yacimiento; por tanto es importante definir que tipo de fluido es el analizado, si es monofásico o bifásico, ya que un error en la

determinación del tipo de fluido nos puede sobre estimar o sub estimar los cálculos que se hagan.

- El índice de producción del pozo colorado 38, ha cambiado mucho desde sus comienzos hasta el día como se ve reflejado en los resultados obtenidos en este proyecto (que fueron calculados de acuerdo a datos históricos) al ser comparados con la producción actual del pozo. Esto puede ser consecuencia de la disminución notable que ha tenido la presión en todo este tiempo.
- Para tomar decisiones acertadas en el momento que se está analizando un conjunto de datos, no solo es importante saber la media, la mediana y la moda de los mismos, sino también es importante saber como están dispuestos los datos respecto a la medida de tendencia central, esto con el fin de saber que tan dispersos están los datos para tratar de evitar incurrir en errores en el análisis de los mismos.
- La toma de decisiones implica cierta incertidumbre, es decir no se sabe si la decisión es correcta, pero si se puede saber la probabilidad que resulte confiable tal decisión. Esto está ligado a la simulación Montecarlo, ya que esta no da un resultado concreto al combinar aleatoriamente las variables involucradas en el modelo sino nos da un margen en el cual se puede mover la variable que se pretende predecir con el modelo. Esto nos disminuye la incertidumbre asociada a cada variable.
- En el caso de las variables asociadas a la roca y a los fluidos es importante tener en cuenta el tipo de función de distribución de probabilidad que se va a escoger para cada variable, entre las más comunes están la normal, la log-normal (usada en este trabajo), la triangular (usada en este trabajo), la gamma, la distribución ji-cuadrado, la distribución t-student entre otras, en este trabajo se usaron principalmente la triangular como primera aproximación y luego una combinación entre

triangular y log-normal, que según dado que para algunas variables se determinó según la literatura que la función de distribución log-normal modelaba mejor su comportamiento.

- Existen muchas herramientas para hacer un análisis usando simulación Montecarlo. Estas van desde las convencionales como las hojas de cálculo hasta gran variedad de programas que llevan a cabo estudios de riesgo y pronósticos. En este trabajo se mostró como sería la metodología para montar un modelo de simulación Montecarlo para predecir IP. Esta metodología puede hacerse extensiva a otros cálculos que tengan o no que ver con la industria petrolera y es la metodología con la que trabajan los diversos software comerciales, además teniendo claros todos los conceptos estadísticos anteriormente se llegó al desarrollo de una aplicación partiendo de la metodología antes expuesta.

- Para los pozos horizontales, además de las incertidumbres asociadas a las variables antes mencionadas (permeabilidad horizontal, espesor, factor volumétrico y viscosidad) hay que tener en cuenta el efecto que tiene la permeabilidad vertical, ya que esta es una variable aleatoria de entrada al modelo que tiene gran incidencia en los cálculos de IP.

## RECOMENDACIONES

- Para un completo análisis teniendo en cuenta la Simulación Montecarlo, es necesario tener unos conceptos estadísticos muy claros que involucren la estadística descriptiva, las medidas de tendencia central, las medidas de dispersión y probabilidad.
- Si se quieren identificar puntos de operación del pozo, las curvas de IPR deberían ser combinadas con una curva que refleje el rendimiento de la tubería de producción.
- Si bien no se cuenta con una ecuación analítica para calcular la relación de rendimiento de influjo en flujo bifásico, se recomienda usar la ecuación de Vogel, pues arroja muy buenos resultados a través de los últimos 40 años.
- Es importante para no incurrir en errores identificar qué función de densidad de probabilidad está más asociada al comportamiento de cada variable.
- La hoja de cálculo es una opción recomendable cuando no se posee un software con licencia que permita correr la simulación, está al alcance de todos y permite entender realmente las bases bajo las cuales funciona el método.
- Los software que se encuentran disponibles presentan una gran versatilidad y fácil manejo, sumado al hecho que tienen una mayor

precisión por trabajar en mejores condiciones con un número de iteraciones más alto.

- Cuando se tenga poca información de la caracterización de un yacimiento, es decir se cuente con un alto grado de incertidumbre en las variables involucradas, la Simulación Montecarlo es una de las mejores opciones para obtener resultados con un alto grado de credibilidad.
- En el caso de la simulación MonteCarlo entre mas representativo se quiere que sean los resultados obtenidos se necesita hacer mas combinaciones aleatorias. En un trabajo posterior se podrían realizar más de diez mil combinaciones y así tener resultados más precisos pero sin descuidar el hecho de que simulaciones que requieren muchas combinaciones pueden ser muy costosas de correr.
- La gran cantidad de detalles disponible puede en algunos casos puede no ser soportada por la maquina, por lo tanto es importante regular las iteraciones.
- Esta metodología puede extenderse para el cálculo de algún parámetro del yacimiento o del pozo que requiera la combinación aleatoria de otras variables. En este sentido solo es necesario identificar cuales son las variables que exhiben cierto grado de aleatoriedad y aplicar el método.
- Dado el hecho de que la salida de la simulación es una distribución de frecuencia, el reporte del analista debería incluir todo tipo de gráficas.

## BILIOGRAFÍA

- CRONQUIST CH. Estimation and Classification of Reserves of Crude Oil, Natural Gas, and Condensate. Society of Petroleum Engineers. Ed Richardson, Texas 2001.
- ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.
- FACULTAD DE CIENCIAS EXACTAS. Simulación, Método de Montecarlo. Universidad Nacional del Centro de la Pcia de Buenos Aires. Argentina, 2005.
- FAULÍN, J. ÁNGEL, J. Simulación de Montecarlo Con Excel. Universidad Abierta de Cataluña. 2003
- JAY. L. DEVORE. Probabilidad y Estadística para ingeniería y ciencias. Madrid. Thomson – Paraninfo. S.A. 2002.
- J.K. PUCKNELL, J.N.E MASON, and E.G. VERVEST, An Evaluation of Recent “Mechanistic” Models of Multiphase Flow for Predicting Pressure Drops in Oil and Gas Wells, BP Exploration, SPE 26682, 1993.
- KAMKOM, R. Analysis of Two-Phase Inflow Performance In Horizontal Wells. University of Texas at Austin. Faculty of the graduate School. May, 2004.

- LISBOA, E.A., and CASTRO M.R. Probabilistic Determination of Productivity Index for a Horizontal Well in a Heavy-Oil Reservoir. SPE 107669. 2007
- M.A. AGGOUR, SPE, and H.Y. AL-YOUSEF, SPE, King Fahd U. of Petroleum & Minerals and A.J. AL-MURAIKHI, Saudi Aramco, Vertical Multiphase Flow Correlations for High Production Rates and Large Tubulars, SPE Production and Facilities, February 1996.
- MARIA A. BARRUFET and AHMED RASOOL, Petroleum Engineering Department, Texas A&M University 77843-3116, and MOHAMED AGGOUR, KING FAHD U. of Petroleum & Minerals, Daharan 31261, Saudi Arabia. Prediction of Bottomhole Flowing Pressures in multiphase Systems Using a Thermodynamic Equation of State. SPE 29479. 1995
- MASO, Robert D; LIND, Douglas A; MARCHAL, Willian G. Estadística para administración y Economía. Ed Alfaomega, 2004.
- MENDENHALL, W. Introducción a la probabilidad y estadística. México: Thomson, 2002.
- SADA D, JOSHI. Horizontal Well Technology. Ed PennWell Books Publishing Company, 1991.
- T.E.W Nind, Fundamentos de Producción y Mantenimiento de Pozos Petroleros, Editorial Limuza México 1987.

## WEBGRAFIA

- <http://es.wikipedia.org/wiki/Portada>
- <http://www.geocities.com/CollegePark/Quad/2435/index.html>
- <http://www.uc3m.es/portal/page/portal/inicio>
- <http://www.uoc.edu/portal/castellano/index3.html>
- <http://www.lumenaut.com/>
- [www.bcr.com.ar](http://www.bcr.com.ar)

## **ANEXO A. INTRODUCCIÓN A LA SIMULACIÓN CON EXCEL**

### **A.1 NÚMEROS ALEATORIOS**

Se explicará a continuación la forma de realizar un análisis de datos cuando se está trabajando en tareas que involucren números aleatorios.

#### **A.1.1 PROCEDIMIENTOS RELACIONADOS**

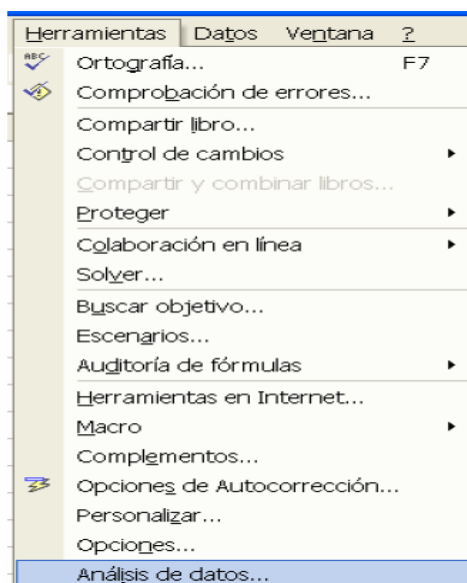
Excel cuenta con dos procedimientos para obtener números aleatorios distribuidos según una forma determinada:

- a) Recurriendo al complemento de *Análisis de Datos*.
- b) Utilizando la función ALEATORIO ().

#### **A.1.2 HERRAMIENTA EXCEL PARA LA GENERACIÓN DE NÚMEROS ALEATORIOS**

Excel proporciona un conjunto de herramientas para el análisis de los datos denominado **Análisis de Datos** que podrá utilizar para ahorrar pasos en el desarrollo de análisis estadísticos. Cuando utilice una de estas herramientas, deberá proporcionar los datos y parámetros para cada análisis; la herramienta utilizará las funciones de macros estadísticas o técnicas correspondientes y, a continuación, mostrará los resultados en una tabla de resultados. Algunas herramientas generan gráficos además de tablas de resultados.

Para ver una lista de las herramientas de análisis, elija Análisis de datos en el menú Herramientas.



FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

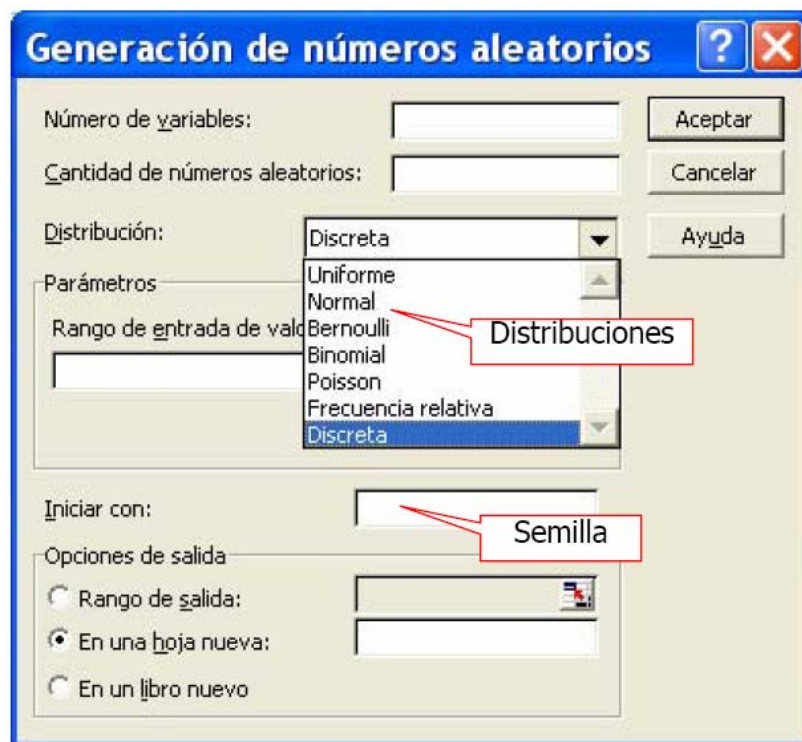
Si este comando no está en el menú, ejecute el programa de instalación para instalar las Herramientas para análisis de la forma siguiente:

Activar las Herramientas para análisis a) En el menú **Herramientas**, elija Macros automáticas. Si en la lista del cuadro de diálogo Macros automáticas no aparece Herramientas para análisis, haga clic en el botón "Examinar" y busque la unidad, directorio y archivo de la macro automática Herramientas para análisis, o bien ejecute el programa de instalación. b) Seleccione la casilla de verificación "**Herramientas para análisis**". Las macros automáticas que instale en Microsoft Excel permanecerán activas hasta que las quite.



**FIGUR 2A. MANERA DE INSTALAR LAS HERRAMIENTAS DE ANÁLISIS CUANDO ESTAS NO ESTÁN INICIALMENTE EN LA VENTANA DE EXCEL**

**FUENTE:** ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.



**FIGURA 3A. PLANTILLA GENERAL PARA LA GENERACIÓN DE NÚMEROS ALEATORIOS**

**FUENTE:** ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

- Número de variables: Introduzca el número de columnas de valores que desee incluir en la tabla de resultados. Si no introduce ningún número, Microsoft Excel rellenará todas las columnas del rango de salida que se haya especificado.

- Cantidad de números aleatorios: Introduzca el número de puntos de datos que desee ver. Cada punto de datos aparecerá en una fila de la tabla de resultados. Si no introduce ningún número, Microsoft Excel rellenará todas las columnas del rango de salida que se haya especificado.

- Distribución: Haga clic en el método de distribución que desee utilizar para crear los valores aleatorios.

- **Uniforme:** Caracterizado por los límites inferior y superior. Se extraen las variables con probabilidades iguales de todos los valores del rango.

- **Normal:** Caracterizado por una media y una desviación estándar.

- **Bernoulli:** Caracterizado por la probabilidad de éxito (valor  $p$ ) en un ensayo dado. Las variables aleatorias de Bernoulli tienen el valor 0 o 1.

- **Binomial:** Caracterizado por una probabilidad de éxito (valor  $p$ ) durante un número de pruebas.

- **Poisson:** Caracterizado por un valor  $\lambda$ , igual a  $1/\text{media}$ .

- **Frecuencia relativa:** Caracterizado por un límite inferior y superior, un incremento, un porcentaje de repetición para valores y un ritmo de repetición de la secuencia.

- **Discreta** Caracterizado por un valor y el rango de probabilidades asociado. El rango debe contener dos columnas. La columna izquierda deberá contener valores y la derecha probabilidades asociadas con el valor de esa fila. La suma de las probabilidades deberá ser 1.

## UNIFORME

Genera muestras de una distribución  $U[A;B]$

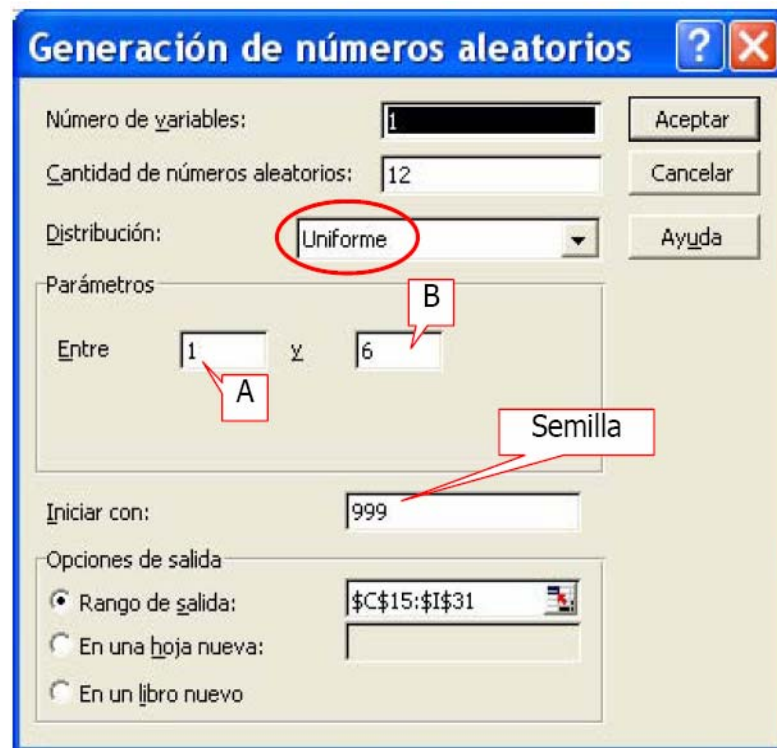
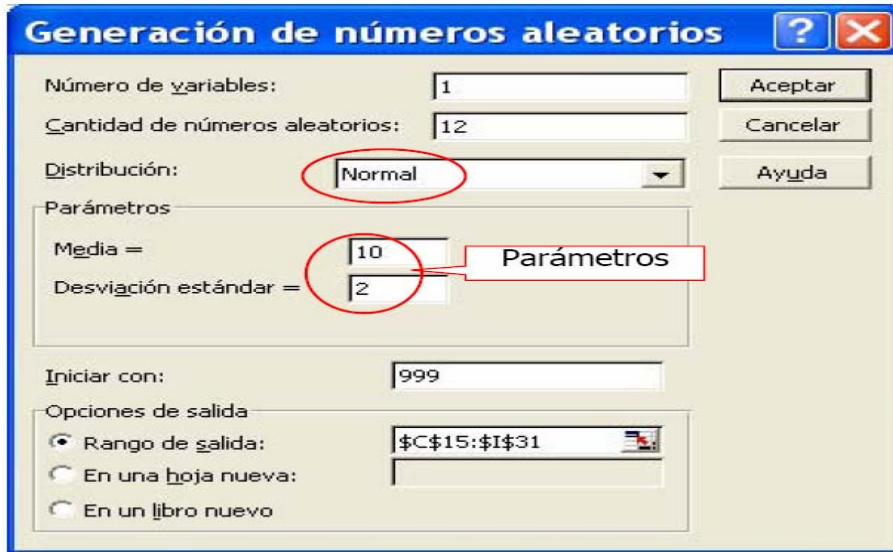


FIGURA 4A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE UNA FUNCIÓN UNIFORME.

FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

## ✓ NORMAL

Genera muestras de una distribución  $N[\mu;\sigma]$



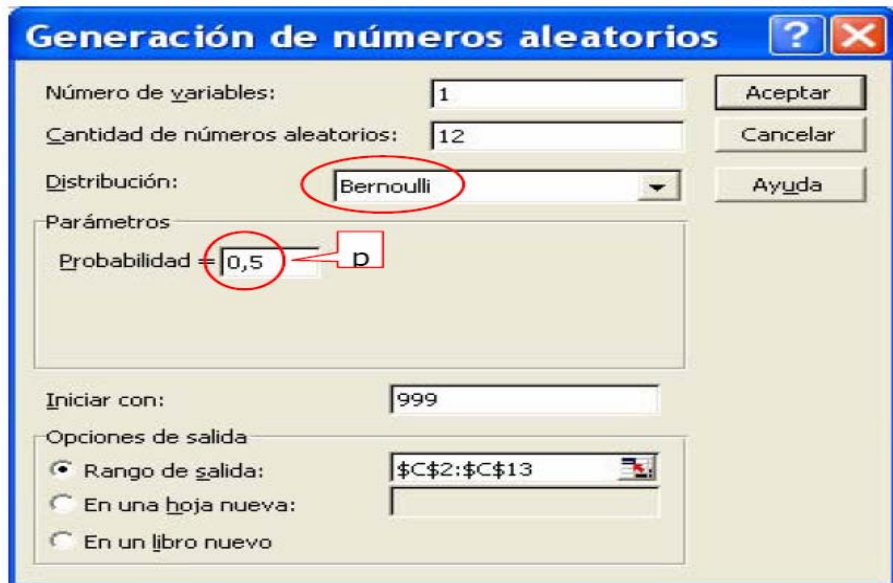
7,445
12,850
8,800
11,015
11,234
7,113
7,882
11,295
11,100
9,988
10,877
8,320

FIGURA 5A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE UNA FUNCIÓN DE DISTRIBUCIÓN NORMAL.

FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

✓ **BERNOULLI**

Genera muestras de una distribución de Bernoulli(p)



1
0
1
0
0
1
1
0
0
1
0
1

FIGURA 6A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE UNA FUNCIÓN DE DISTRIBUCIÓN DE BERNOULLI.

FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

✓ **BINOMIAL**

Genera muestras de una distribución de B(n,p)

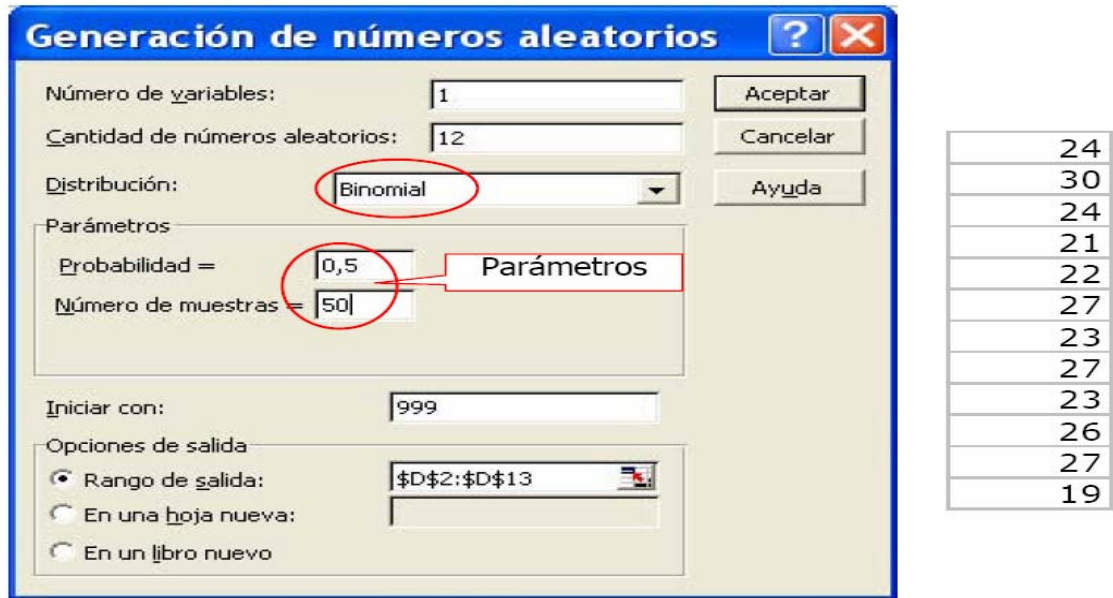


FIGURA 7A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE UNA FUNCIÓN DE DISTRIBUCIÓN BINOMIAL.

FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

✓ **POISSON**

Genera muestras de una distribución de **Poisson**( $\lambda$ )

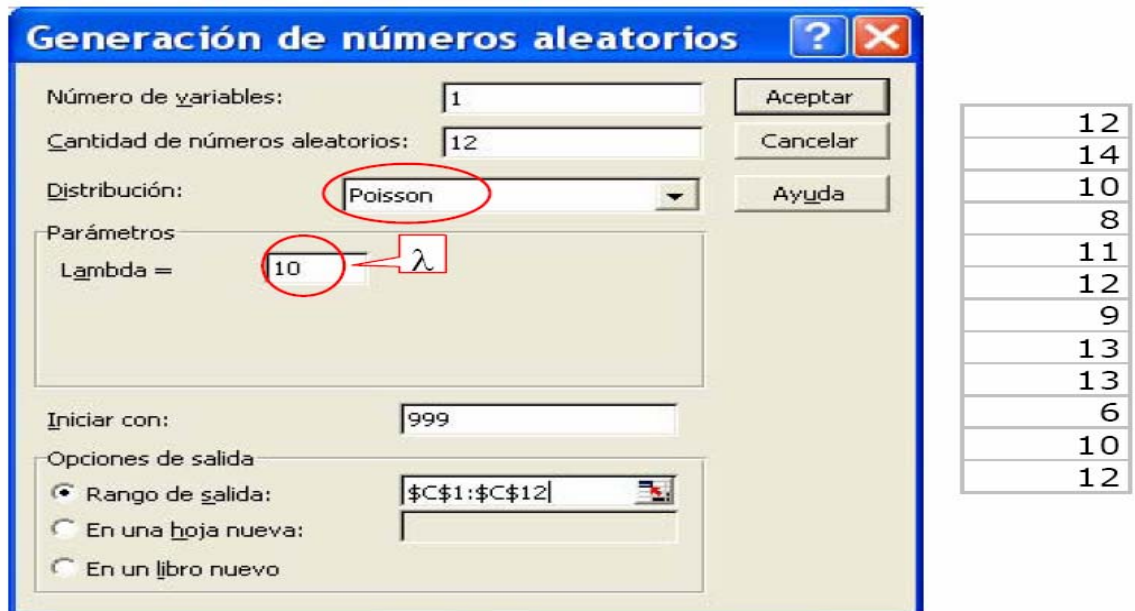


FIGURA 8A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE UNA FUNCIÓN DE DISTRIBUCIÓN DE POISSON

FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

✓ **FRECUENCIA RELATIVA**

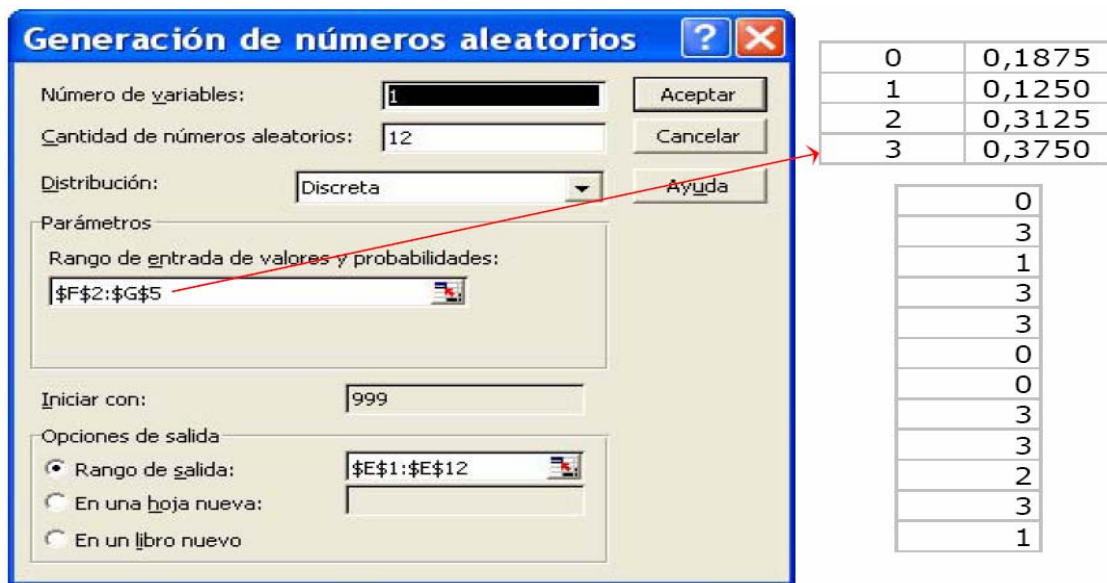
En realidad genera series o secuencias de números



**FIGURA 9A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE FRECUENCIA RELATIVA**  
 FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

✓ **DISCRETA**

Genera números dada una distribución de frecuencias relativas



**FIGURA 10A. GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE FUNCIÓN DE DISTRIBUCIÓN DISCRETA**  
 FUENTE: ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

**A.1.3 GENERACIÓN DE NÚMEROS ALEATORIOS MEDIANTE FÓRMULAS**

Excel cuenta con dos funciones para la generación de números de una distribución Uniforme: *ALEATORIO()* según una  $U[0;1]$ ;

*ALEATORIO.ENTRE(a;b)*: según una  $U[a;b]$  discreta. Estas funciones son volátiles de manera que se recalcularán cada vez (si la opción de cálculo está puesta en automático).

- **ALEATORIO**: Devuelve un número aleatorio mayor o igual que 0 y menor que 1, distribuido uniformemente. Cada vez que se calcula la hoja de cálculo, se devuelve un número aleatorio nuevo. Su sintaxis es:

### **ALEATORIO( )**

- **ALEATORIO.ENTRE**: Devuelve un número aleatorio entre los números que especifique. Devuelve un nuevo número aleatorio cada vez que se calcula la hoja de cálculo. Si esta función no está disponible, ejecute el programa de instalación e instale las Herramientas para análisis. Para instalar este complemento, elija **Complementos** en el menú **Herramientas** y active la casilla correspondiente.

Su sintaxis es

### **ALEATORIO.ENTRE(inferior; superior)**

- **Inferior**: es el menor número entero que la función *ALEATORIO.ENTRE* puede devolver.
- **Superior** es el mayor número entero que la función *ALEATORIO.ENTRE* puede devolver.

A pesar de contar únicamente con funciones para generar números distribuidos de forma uniforme podemos gracias a éstas, generar prácticamente cualquier distribución utilizando bien algoritmos descritos en la literatura, bien las funciones inversas cuando éstas están implementadas en Excel.

A continuación se presenta este procedimiento para algunas de las funciones más comunes, siendo  $U = \text{Aleatorio}()$ :

- **DISCRETAS**

**Binomial (n, p)** =BINOM.CRIT(n; p; U)

**Poisson (λ)** =BINOM.CRIT(λ/0,001;0,001;U)

**Uniforme (a,b)** =ALEATORIO.ENTRE(a; b)

**Geométrica (p)** =REDONDEAR.MENOS(LN(U)/LN(1-p);0)

**BinomNegativa(n,p)** =BINOM.CRIT(DISTR.GAMMA.INV(U; n;(1-p)/p)/ε;ε;U)

- **CONTINUAS**

**Beta (α,β)** =DISTR.BETA.INV(U; α; β; a , b )

**Exponencial(λ)** =(1/λ) \* -LOG(U)

**Gamma(α,β)** =DISTR.GAMMA.INV(U; α; β)

**LogNormal(π,σ)** =DISTR.LOG.INV(U; μ; σ)

**Normal(π,σ)** =μ + σ\*(RAIZ(-2\*LOG(U))\*SENO(2\*PI()\*U))

**Pareto(α,β)** =β\*(U<sup>-1/α</sup>)

**Uniforme (a,b)** =α + (β-α)\*U

**Triangular (a,b,c)** = c + (a+U\*(b-a)-c)\*MAX(U;U)  
c + (a+U\*(b-a)-c)\*RAIZ(U)

- **OTRAS**

**X<sup>2</sup><sub>GL</sub>** = PRUEBA.CHI.INV(U;GL)

**F<sub>(GL1;GL2)</sub>** =DISTR.F.INV(U;GL1;GL2)

**TGL** =DISTR.T.INV(U;GL)\*SIGNO(U-0,5)

## A.2 DISTRIBUCIÓN DE FRECUENCIAS

En la siguiente sección se presentarán todos los puntos necesarios para realizar un análisis de datos estudiando una distribución de frecuencias.

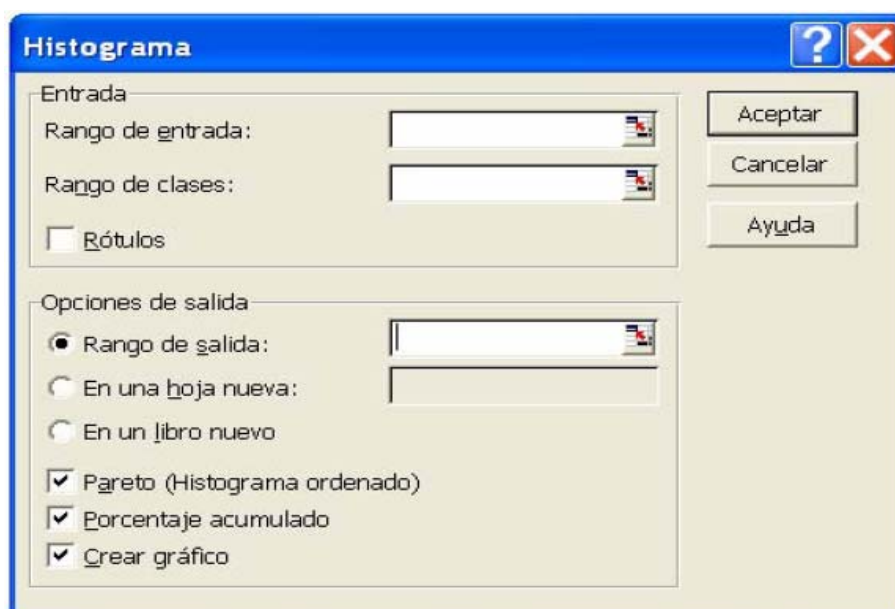
### A.2.1 PROCEDIMIENTOS RELACIONADOS

Excel cuenta con dos procedimientos para obtener la distribución de frecuencias de una variable:

- a) Recurrir al complemento de **Análisis de Datos (HISTOGRAMA)**.
- b) Utilizar la función **FRECUENCIA**.

### A.2.2 ANÁLISIS DE DATOS: HISTOGRAMA

Obtiene la distribución de frecuencias de un conjunto de datos. Dibuja un histograma y el diagrama de Pareto.



**FIGURA 11A. FORMA DE REALIZAR UN HISTOGRAMA TENIENDO EN CUENTA UN CONJUNTO DE DATOS**

**FUENTE:** ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

\* **Rango clases** (opcional): Introduzca la referencia de celda para un rango que contenga un conjunto opcional de valores límite que definan rangos de clase. Estos valores deberán estar en orden ascendente. Microsoft Excel contará el número de puntos de datos que hay entre el número de clases actual y el número de clases mayor, si lo hay.

- Se contará un número de una clase determinada si es igual o menor que el número de clase situado por debajo de la última clase. Todos los valores por debajo del primer valor de clase se contarán juntos, como los valores por encima del último valor de clase.
- Si omite el rango de clase, Excel creará un conjunto de clases distribuidas uniformemente entre los valores mínimo y máximo de los datos.

\* **Pareto** (Histograma ordenado): Active esta casilla para presentar los datos en orden de frecuencia descendente en la tabla de resultados. Si esta casilla está desactivada, Excel presentará los datos en orden ascendente y omitirá las tres columnas situadas más a la derecha que contienen los datos ordenados.

\* **Porcentaje acumulado**: Active esta casilla para generar una columna de tabla de resultados con porcentajes acumulados y para incluir una línea de porcentaje acumulado en el gráfico de histograma. Desactívela para omitir los porcentajes acumulados.

\* **Crear gráfico**: Active esta casilla para generar un gráfico de histograma incrustado con la tabla de resultados.

### **A.2.3 LA FUNCIÓN FRECUENCIA**

La función de Excel para la obtención de las distribuciones de frecuencias es:

- **FRECUENCIA:** Devuelve una distribución de frecuencia como una matriz

Vertical Su sintaxis es:

**FRECUENCIA(datos; grupos)**

- **Datos:** es una matriz de un conjunto e valores o una referencia a un conjunto de valores cuyas frecuencias desea contar. Si datos no contiene ningún valor, FRECUENCIA devuelve una matriz de ceros.
- **Grupos:** es una matriz de intervalos o una referencia a intervalos dentro de los cuales desea agrupar los valores del argumento datos. Si grupos no contiene ningún valor, FRECUENCIA devuelve el número de elementos contenido en datos.

Observaciones

- FRECUENCIA se introduce como una fórmula matricial después de seleccionar un rango de celdas adyacentes en las que se desea que aparezca el resultado de la distribución.
- El número de elementos de la matriz devuelta supera en una unidad el número de elementos de grupos. El elemento adicional de la matriz devuelta devuelve la suma de todos los valores superiores al mayor intervalo.

### **A.3 MEDIDAS DE TENDENCIA CENTRAL, VARIACIÓN Y FORMA**

Seguidamente de que tenemos un conjunto de datos como solución a un problema, se hace necesario definir y analizar las medidas de tendencia central y las medidas de desviación.

### A.3.1 PROCEDIMIENTOS RELACIONADOS

Excel cuenta con dos procedimientos para obtener la descripción mediante estadísticos de una muestra o una población:

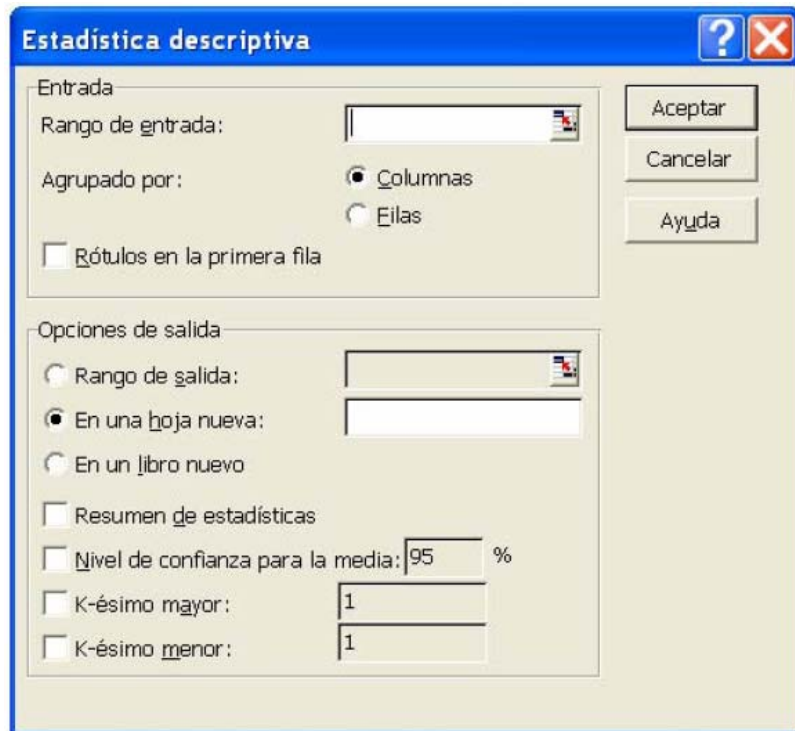
- a) Recurrir al complemento de Análisis de Datos (Estadística Descriptiva).
- b) Utilizar las funciones relacionadas expuestas en los apartados 3.3, 3.4 y 3.5

### A.3.2 ANÁLISIS DE DATOS ESTADÍSTICA DESCRIPTIVA

Calcula los estadísticos básicos de un conjunto de datos, para una o varias variables.

- **Nivel de confianza para la media:** Active esta casilla si desea incluir una fila correspondiente al nivel de confianza de la media en la tabla de resultados. En el cuadro, escriba el nivel de confianza que desee utilizar. Por ejemplo, un valor de 95 % calculará el nivel de confianza de la media con un nivel de importancia del 5 %.
- **K-ésimo mayor:** Active esta casilla si desea incluir una fila correspondiente al valor k-ésimo mayor de cada rango de datos en la tabla de resultados. En el cuadro, escriba el número que va a utilizarse para k. Si escribe 1, esta fila contendrá el máximo del conjunto de datos.
- **K-ésimo menor:** Active esta casilla si desea incluir una fila correspondiente al valor k-ésimo menor de cada rango de datos en la tabla de resultados. En el cuadro, escriba el número que va a utilizarse para k. Si escribe 1, esta fila contendrá el mínimo del conjunto de datos.
- **Resumen de estadísticas:** Seleccione esta opción si desea que Excel genere un campo en la tabla de resultados por cada una de las siguientes variables estadísticas: *Media*, *Error típico (de la media)*, *Mediana*, *Moda*,

*Desviación estándar, Varianza, Curtosis, Coeficiente de asimetría, Rango, Mínimo, Máximo, Suma, Cuenta, Mayor (#), Menor (#) y Nivel de confianza.*



**FIGURA 12A. FORMA DE REALIZAR UN ANÁLISIS DE DATOS DE ACUERDO A LO QUE IMPLICA LA ESTADÍSTICA DESCRIPTIVA**

**FUENTE:** ESTADÍSTICA ACTUARIAL I. Introducción a la Simulación con Excel. Universidad Carlos III de Madrid.

### **A.3.3 FUNCIONES PARA EL CÁLCULO DE LA TENDENCIA CENTRAL.**

#### **\* Media**

- **PROMEDIO:** Devuelve la media aritmética de los argumentos.
- **PROMEDIOA:** Devuelve la media incluidos texto y valores lógicos.
- **MEDIA.ACOTADA:** Devuelve la media recortada de un conjunto de datos MEDIA.ACOTADA(matriz ; porcentaje).

**Matriz** es la matriz o el rango de valores que desea acotar y de los cuales se calculará la media. **Porcentaje** es el número fraccionario de puntos de datos que se excluyen del cálculo. Por ejemplo, si porcentaje = 0,2, se eliminarán cuatro puntos de un conjunto de datos de 20 puntos (20 x 0,2), dos de la parte superior y dos de la parte inferior.

- **MEDIA.ARMO**: Devuelve la media armónica.
- **MEDIA.GEOM**: Devuelve la media geométrica.

**\*Mediana.**

- **MEDIANA**: Devuelve la mediana de los números dados.

**\*Moda.**

- **MODA**: Devuelve el valor más frecuente en un conjunto de datos.

#### **A.3.4 FUNCIONES PARA EL CÁLCULO DE LA VARIACIÓN**

**\*Rango medio**

- *Usar MAX y MIN:*

$$RM = \text{PROMEDIO}(\text{MAX}(\text{Datos}) + \text{MIN}(\text{Datos}))$$

**\*Cuartiles**

- **CUARTIL:**

CUARTIL(matriz ; cuartil)

**Matriz**, es la matriz o rango de celdas de valores numéricos cuyo cuartil desea obtener. **Cuartil**, indica el valor que se devolverá, el código es (0 = mínimo; 1 = primer cuartil; 2 = mediana; 3 = tercer cuartil; 4 = máximo).

- *PERCENTIL*: Devuelve el percentil k-ésimo de los valores de un rango *PERCENTIL(matriz ; k)*.

**Matriz** es la matriz o rango de datos que define la posición relativa. **K** es el valor de percentil en el intervalo de 0 a 1, inclusive.

- *RANGO.PERCENTIL*: Devuelve el % del los valores que son menores que cifra dentro de matriz .

*RANGO.PERCENTIL(matriz;x;cifra\_significativa)*.

**Matriz** es la matriz o rango de datos con valores numéricos que define la posición relativa. **X**, es el valor cuyo rango percentil desea conocer. **Cifra significativa** es un valor opcional que identifica el número de cifras significativas para el valor de porcentaje devuelto. Si se omite este argumento, *RANGO.PERCENTIL* utiliza tres dígitos.

#### \*Rango intercuartílico

-*PERCENTIL* (RI = *PERCENTIL(datos;0,75)*-*PERCENTIL(datos;0,25)*)

-*CUARTIL* (RI = *CUARTIL(datos;3)*- *CUARTIL(datos;1)*).

#### \*Varianza y desviación típica.

- **VAR(A)**: Calcula la (cuasi) varianza de una muestra.
- **VARP(A)**: Calcula la varianza de la población.
- **DESVEST(A)**: Calcula la (cuasi) desviación estándar de una muestra.
- **DESVESTP(A)**: Calcula la desviación estándar de la población total.

#### \*Coeficiente de variación

- Usar *PROMEDIO* y *DESVEST*

### A.3.5 FUNCIONES PARA EL CÁLCULO DE LA FORMA

#### \*Simetría

- **COEFICIENTE.ASIMETRIA**: Devuelve el sesgo de una distribución

#### \*Curtosis

- **CURTOSIS**: Devuelve el coeficiente de curtosis de un conjunto de datos