

**Desarrollo del Razonamiento Inferencial Informal sobre las Pruebas de
Hipótesis en un ambiente de simulación computacional**

Cristian Eduardo Carreño Granados

Trabajo de grado para obtener el título de Magíster en Educación Matemática

Director

Gabriel Yáñez Canal
Doctor en Ciencias de la Especialidad Educación Matemática

Universidad Industrial de Santander
Facultad de Ciencias
Escuela de Matemáticas
Maestría en Educación Matemática
Bucaramanga
2018

Agradecimientos

A Dios, sin el nada sería posible

A mis padres Miriam y Vicente, que nunca dejaron de brindarme su apoyo incondicional, y su sabio consejo. A mi Tía Adela, por confiar en mí, brindarme su apoyo y confiar en mis habilidades.

Al profesor Gabriel Yáñez, director de esta tesis, y una de las personas que más admiro. Sus consejos y enseñanzas hicieron de realidad esta tesis, y me hicieron una mejor persona. Gracias profe.

A Laura, por brindarme su compañía, amor y comprensión.

Gracias a mis compañeros de maestría, Jonathan, Sully, Edwin, Ana y Giovanni, por recorrer este camino conmigo.

A mis profesores de la maestría Jorge Fiallo, y Sandra Parada, por su compromiso con mi formación y la de mis compañeros.

Dedicatoria

Dedico este trabajo a mis padres y mi familia, porque nunca dejaron de apoyarme. A mi director, Gabriel Yáñez, por aguantarme tanto durante este proceso de formación. Y especialmente a Laura Tatiana Jaimes, que siempre estuvo a mi lado en los momentos difíciles, y nunca dejó de animarme para cumplir esta meta.

Tabla de contenido

Introducción	22
1. Antecedentes.....	27
1.1. Dificultades y errores alrededor de las pruebas de hipótesis	27
1.2. El Razonamiento Inferencial Informal: Pruebas de Hipótesis.....	33
1.3. Sobre los elementos de la investigación	37
1.3.1. Sobre la Taxonomía SOLO.....	37
1.3.2. La simulación de datos como herramienta didáctica	40
2.- Referentes teóricos y conceptuales.....	43
2.1. Eje Conceptual.....	44
2.1.1. Inferencia Estadística: Prueba de Hipótesis.....	44
2.1.2. La Inferencia Estadística Informal y el Razonamiento Inferencial Informal	49
2.2. Eje Didáctico.....	53
2.2.1. Taxonomía SOLO.....	54
3.- Metodología.....	58
3.1. Generalidades.....	58

3.1.1. Participantes.....	58
3.1.2. <i>Fathom</i> como herramienta didáctica.....	59
3.1.3. Recolección de datos.....	60
3.2. Fases de la investigación.....	60
3.2.1. Fase I: Diseño de Instrumentos.....	60
3.2.1.1. Diseño de las Pruebas	60
3.2.1.2. Diseño de Talleres.....	61
3.2.1.3. Componentes conceptuales.....	63
3.2.2. Fase II: Aplicación de Instrumentos	65
3.2.3. Fase III: Análisis de Resultados.....	66
4.- Análisis de Resultados.....	67
4.1. Prueba Inicial	68
4.1.1. Análisis de Componentes.....	68
4.1.2. Análisis por estudiante	72
4.1.3. Conclusiones	88
4.2. Taller 1	88
4.2.1. Análisis de Componentes.....	89
4.2.2. Experimentación <i>Fathom</i>	91
4.2.3. Análisis por estudiante	107
4.2.4. Conclusiones	121

4.3. Taller 2	122
4.3.1. Análisis de Componentes.....	122
4.3.2. Experimentación <i>Fathom</i>	123
4.3.3. Análisis por estudiante	127
4.3.4. Conclusiones	143
4.4. Taller 3	143
4.4.1. Análisis de Componentes.....	143
4.4.2. Experimentación <i>Fathom</i>	146
4.4.3. Análisis por estudiante	150
4.4.4. Conclusiones	172
4.5. Taller 4	173
4.5.1. Análisis de Componentes.....	173
4.5.2. Experimentación <i>Fathom</i>	175
4.5.3. Análisis por estudiante	179
4.5.4. Conclusiones	197
4.6. Prueba Final	197
4.6.1. Análisis de Componentes.....	197
4.6.2. Análisis por pregunta	211
4.6.3. Conclusiones	229
4.7. Contraste Prueba Inicial versus Prueba Final	230
5.- Conclusiones.....	234

5.1. Los Estudiantes	234
5.1.1. Variabilidad de la proporción muestral y sus implicaciones (Dist. Discreta).....	234
5.1.2. Variabilidad de la muestral y del estimador muestral (Dist. Continua).....	235
5.1.3. Influencia del tamaño muestral	235
5.1.4. Región de rechazo/Nivel de significancia.....	236
5.1.5. Contraste de hipótesis/p-valor/Razonamiento Inferencial	236
5.1.6. Contraste de hipótesis (análisis gráfico)	236
5.2. La experimentación.....	237
5.3. Recomendaciones	239
Referencias bibliográficas.....	240

Lista de figuras

Figura 1. Esquema de los referentes de la investigación	43
Figura 2. Mapa conceptual de los elementos de las pruebas de hipótesis	48
Figura 3. Simulación de resultados para el juguete "Helper".	51
Figura 4. Distribución Prueba Inicial-1	68
Figura 5. Distribución Prueba Inicial-2	69
Figura 6. Distribución Prueba Inicial-3	69
Figura 7. Taller 1-Simulación esperada 1	91
Figura 8. Taller 1-Simulación esperada 2	91
Figura 9. Taller 1-Simulación esperada 3	92
Figura 10. Taller 1-Simulación esperada 4	92
Figura 11. Taller 1-Simulación esperada 5	93
Figura 12. Taller 1-Simulación esperada 6	93
Figura 13. Taller 1-Simulación esperada 7	94

Figura 14. Taller 1-Simulación esperada 8	95
Figura 15. Taller 1-Simulación esperada 9	95
Figura 16. Taller 1-Simulación esperada 10	96
Figura 17. Taller 1-Simulación esperada 11	96
Figura 18. Taller 1-Simulación esperada 12	97
Figura 19. Taller 1-Simulación esperada 13	97
Figura 20. Taller 1-Simulación esperada 14	98
Figura 21. Taller 1-Simulación esperada 15	98
Figura 22. Taller 1-Simulación esperada 16	99
Figura 23. Taller 1-Simulación esperada 17	100
Figura 24. Taller 1-Simulación esperada 18	101
Figura 25. Taller 1-Simulación esperada 19	101
Figura 26. Taller 1-Simulación esperada 20	102
Figura 27. Taller 1-Simulación esperada 21	102
Figura 28. Taller 1-Simulación esperada 22	103
Figura 29. Taller 1-Simulación esperada 23	103

RAZONAMIENTO INFERENCIAL INFORMAL EN P. DE HIPÓTESIS	12
Figura 30. Taller 1-Simulación esperada 24	104
Figura 31. Taller 1-Simulación esperada 25	105
Figura 32. Taller 1-Simulación esperada 26	105
Figura 33. Taller 1-Simulación esperada 27	106
Figura 34. Taller 1-Simulación esperada 28	106
Figura 35. Taller 1-Simulación esperada 29	107
Figura 36. Taller 1-Simulación esperada 30	107
Figura 37. Evidencia Gabriela Taller 1-1	115
Figura 38. Taller 2-Simulación esperada 1	124
Figura 39. Taller 2-Simulación esperada 2	124
Figura 40. Taller 2-Simulación esperada 3	125
Figura 41. Taller 2-Simulación esperada 4	126
Figura 42. Taller 2-Simulación esperada 5	126
Figura 43. Taller 2-Simulación esperada 6	127
Figura 44. Taller 2-Simulación esperada 7	127
Figura 45. Evidencia Andrés Taller 2-1	129

Figura 46. Evidencia Luis Taller 2-1	140
Figura 47. Evidencia Luis Taller 2-2	140
Figura 48. Evidencia Luis Taller 2-3	141
Figura 49. Taller 3-Simulación esperada 1	146
Figura 50. Taller 3-Simulación esperada 2	147
Figura 51. Taller 3-Simulación esperada 3	147
Figura 52. Taller 3-Simulación esperada 4	147
Figura 53. Taller 3-Simulación esperada 5	148
Figura 54. Taller 3-Simulación esperada 6	148
Figura 55. Taller 3-Simulación esperada 7	149
Figura 56. Taller 3-Simulación esperada 8	149
Figura 57. Taller 3-Simulación esperada 9	149
Figura 58. Evidencia Andrés Taller 3-1	151
Figura 59. Evidencia Andrés Taller 3-2	152
Figura 60. Evidencia Andrés Taller 3-3	154
Figura 61. Evidencia Andrés Taller 3-4	156

Figura 62. Evidencia Sofía Taller 3-1	157
Figura 63. Evidencia Sofía Taller 3-2.....	158
Figura 64. Evidencia Sofía Taller 3-3.....	161
Figura 65. Taller 4-Simulación esperada 1	176
Figura 66. Taller 4-Simulación esperada 2	176
Figura 67. Taller 4-Simulación esperada 3	176
Figura 68. Taller 4-Simulación esperada 4	177
Figura 69. Taller 4-Simulación esperada 5	177
Figura 70. Taller 4-Simulación esperada 6	178
Figura 71. Taller 4-Simulación esperada 7	178
Figura 72. Taller 4-Simulación esperada 8	179
Figura 73. Evidencia Andrés Taller 4-1.....	180
Figura 74. Evidencia Andrés Taller 4-2.....	181
Figura 75. Evidencia Sofía Taller 4-1	184
Figura 76. Evidencia Sofía Taller 4-2.....	186
Figura 77. Evidencia Sofía Taller 4-3.....	187

RAZONAMIENTO INFERENCIAL INFORMAL EN P. DE HIPÓTESIS	15
Figura 78. Evidencia Gabriela Taller 4-1	189
Figura 79. Evidencia Gabriela Taller 4-2	190
Figura 80. Prueba Final gráfico 1	204
Figura 81. Prueba Final gráfico 2	209
Figura 82. Análisis Prueba Final gráfico 1	221
Figura 83. Análisis Prueba Final gráfico 2	226

Lista de tablas

Tabla 1. Concepciones erróneas sobre las pruebas de hipótesis	31
Tabla 2. Niveles SOLO para Andrés en la prueba Inicial.....	75
Tabla 3. Niveles SOLO para Sofía en la prueba Inicial.....	80
Tabla 4. Niveles SOLO para Gabriela en la prueba Inicial	84
Tabla 5. Niveles SOLO para Luis en la prueba Inicial	87
Tabla 6. Taller 2-Composiciones de la Urna	122
Tabla 7. Taller 2-Composiciones de la Urna Andrés 1	128
Tabla 8. Taller 2-Composiciones de la Urna Andrés 2.....	128
Tabla 9. Taller 2-Composiciones de la Urna Sofía 1	131
Tabla 10. Taller 2-Composiciones de la Urna Sofía 2.....	132
Tabla 11. Taller 2-Composiciones de la Urna Gabriela 1	135
Tabla 12. Taller 2-Composiciones de la Urna Gabriela 2	135
Tabla 13. Taller 2-Composiciones de la Urna Luis 1	138
Tabla 14. Taller 2-Composiciones de la Urna Luis 2	138

Tabla 15. Prueba Final Pregunta 1	198
Tabla 16. Prueba Final Pregunta 2	199
Tabla 17. Prueba Final Pregunta 3	200
Tabla 18. Prueba Final Pregunta 4	202
Tabla 19. Prueba Final Pregunta 5	203
Tabla 20. Prueba Final Pregunta 6	205
Tabla 21. Prueba Final Pregunta 7	206
Tabla 22. Prueba Final Pregunta 8	207
Tabla 23. Prueba Final Pregunta 9	209
Tabla 24. Prueba Final Pregunta 10	210
Tabla 25. Análisis Prueba Final Pregunta 1	212
Tabla 26. Nivel SOLO prueba Final pregunta 1	213
Tabla 27. Análisis Prueba Final Pregunta 2	214
Tabla 28. Nivel SOLO prueba Final pregunta 2	215
Tabla 29. Análisis Prueba Final Pregunta 3	216
Tabla 30. Nivel SOLO prueba Final pregunta 4	217

Tabla 31. Análisis Prueba Final Pregunta 4.....	218
Tabla 32. Nivel SOLO prueba Final pregunta 4.....	219
Tabla 33. Análisis Prueba Final Pregunta 5.....	219
Tabla 34. Nivel SOLO prueba Final pregunta 5.....	220
Tabla 35. Análisis Prueba Final Pregunta 6.....	221
Tabla 36. Nivel SOLO prueba Final pregunta 6.....	222
Tabla 37. Análisis Prueba Final Pregunta 7.....	223
Tabla 38. Nivel SOLO prueba Final pregunta 7.....	224
Tabla 39. Análisis Prueba Final Pregunta 8.....	224
Tabla 40. Nivel SOLO prueba Final pregunta 8.....	225
Tabla 41. Análisis Prueba Final Pregunta 9.....	226
Tabla 42. Nivel SOLO prueba Final pregunta 9.....	227
Tabla 43. Análisis Prueba Final Pregunta 10.....	227
Tabla 44. Nivel SOLO prueba Final pregunta 10.....	228
Tabla 45. Respuestas correctas prueba Final.....	229
Tabla 46. Contraste de resultados para Andrés.....	230

Tabla 47. Contraste de resultados para Sofía..... 231

Tabla 48. Contraste de resultados para Gabriela 232

Tabla 49. Contraste de resultados para Luis 232

Resumen

Título: Desarrollo del Razonamiento Inferencial Informal sobre las Pruebas de Hipótesis en un ambiente de simulación computacional¹

Autor: Cristian Eduardo Carreño Granados²

Palabras claves: Educación estadística, Razonamiento Inferencial Informal, Pruebas de Hipótesis, Taxonomía SOLO

Descripción:

En el siguiente trabajo abordamos la problemática presente alrededor de la enseñanza y el aprendizaje de la Inferencia Estadística, específicamente la relacionada con las pruebas de hipótesis.

Teniendo en cuenta las concepciones erróneas y las dificultades que presentan los estudiantes al realizar una prueba de hipótesis; y el Razonamiento Inferencial Informal como alternativa para la enseñanza de inferencia estadística, nos planteamos el objetivo de caracterizar cómo cambian los razonamientos inferenciales informales de los estudiantes de licenciatura en matemáticas respecto a las pruebas de hipótesis después de trabajar en situaciones de inferencia con simulación computacional.

Con el fin realizar la caracterización de los razonamientos informales de los estudiantes, diseñamos una prueba que nos permitió identificar el razonamiento inferencial informal de los estudiantes, para posteriormente desarrollarlo a través de una serie de talleres mediados por simulación computacional con Fathom. Para el diseño de los talleres y el análisis de los resultados nos apoyamos en la taxonomía SOLO, la cual nos permitió clasificar las respuestas de los estudiantes de acuerdo al número de componentes conceptuales que utilizaron y relacionaron. Los resultados que obtuvimos muestran que la mediación computacional potencia el razonamiento informal, llevando a estudiantes sin conocimiento alguno sobre las pruebas de hipótesis, a realizar inferencias informales e interiorizar nociones de nivel de significancia, región de rechazo y p-valor.

¹ Trabajo de grado

² Facultad de Ciencias. Escuela de Matemáticas. Director: Gabriel Yáñez Canal, Doctor en Ciencias de Especialidad en Matemáticas.

Abstract

Title: Development of Informative Inferential Reasoning on Hypothesis Tests in a computational simulation environment³

Author: Cristian Eduardo Carreño Granados⁴

Keywords: Statistical Education, Informal Inferential Reasoning, Hypothesis Tests, Taxonomy SOLO

Description:

In the following work we address the current problems around the teaching and learning of Statistical Inference, specifically that related to hypothesis testing.

Taking into account the misconceptions and difficulties that students present when performing a hypothesis test; and Informal Inferential Reasoning as an alternative for the teaching of statistical inference, we set ourselves the objective of characterizing how informal inferential reasoning of undergraduate students in mathematics changes with respect to hypothesis testing after working in inference situations with computational simulation.

In order to carry out the characterization of students' informal reasoning, we designed a test that allowed us to identify students' informal inferential reasoning, to later develop it through a series of workshops mediated by computer simulation with Fathom. For the design of the workshops and the analysis of the results, we relied on the SOLO taxonomy, which allowed us to classify the students' answers according to the number of conceptual components they used and related. The results we obtained show that computational mediation enhances informal reasoning, leading students without any knowledge about hypothesis testing, to make informal inferences and internalize notions of significance level, rejection region and p-value.

³ Thesis

⁴ Facultad de Ciencias. Escuela de Matemáticas. Director: Gabriel Yáñez Canal, Doctor en Ciencias de Especialidad en Matemáticas.

Introducción

La estadística juega un papel importante en nuestro diario vivir, cada día se hace más indispensable su uso en infinidad de situaciones formales e informales. Cada vez es más frecuente que las personas, en sus actividades profesionales y en su vida diaria se vean en la necesidad de saber interpretar y comprender información sobre diversidad de temas (economía, política, negocios y finanzas, salud, demografía, deportes, etc.); además de tomar decisiones involucrando conceptos matemáticos de carácter cuantitativo y probabilístico.

Entre los temas que trata la estadística, la inferencia es uno de los más importantes. La podemos definir como los métodos para hacer juicios o sacar conclusiones acerca de una población, a partir de los datos de una muestra aleatoria. Dada la gran relevancia de la inferencia estadística, su estudio forma parte de los currículos de la educación básica y de los primeros cursos universitarios (NCTM, 2000; MEN 2006). Por ejemplo, en los principios y estándares del NTCM (2000) en el estándar de Análisis de datos y Probabilidad se dice que los programas de enseñanza de todas las etapas deberían capacitar a todos los estudiantes para desarrollar y evaluar inferencias y predicciones basadas en datos.

El NCTM (2000) considera que los estudiantes deberían llegar a comprender los elementos básicos del análisis estadístico: seleccionar una muestra adecuada, recoger datos de esta muestra, describir la muestra y hacer inferencias razonables que relacionen la muestra y la población.

Además, agrega que los estudiantes al finalizar la educación secundaria deberían usar simulaciones para aprender sobre distribuciones muestrales y hacer inferencias informales.

De igual forma, El Ministerio de Educación Nacional de Colombia (2006), en sus Estándares básicos de competencias en matemáticas hace alusión a la inferencia desde los primeros grados de educación; en el pensamiento aleatorio y sistemas de datos dice que los estudiantes deberían estar en la capacidad de juzgar si la probabilidad de ocurrencia de un evento es mayor que la de otro. Y en grados avanzados argumenta que los estudiantes deberían justificar o refutar inferencias basadas en razonamientos estadísticos a partir de resultados de estudios publicados en los medios o diseñados en el ámbito escolar; además de proponer inferencias a partir del estudio de muestras probabilísticas.

Pero a pesar de su evidente inclusión en los programas de estudio, la literatura ha mostrado las muchas dificultades presentes que encierra la inferencia estadística en todos los niveles académicos (Castro Sotos, Vanhoof, Van den Noortgate y Onghena, 2007).

Entre los métodos estadísticos propios de la inferencia estadística, uno de los que presenta mayor complejidad para ser aprendido y enseñado son las pruebas de hipótesis, las cuales, debido a la gran cantidad de conceptos y términos que intervienen en ella resultan sientos el talón de Aquiles de la inferencia estadística (Peskun, 1987; Falk, 1986; Vallecillos, 1991). Como consecuencia de esta dificultad la comunidad de educación estadística creó como mecanismo remedial la línea de investigación de la inferencia estadística informal, que busca la manera de acercarse a la inferencia formal a partir de inferencia informal.

Para trabajar la inferencia informal es necesario analizar, en primera medida, el razonamiento inferencial informal, que no es otra cosa que la manera en la cual los estudiantes usan su conocimiento estadístico informal para formular argumentos que apoyen sus conclusiones sobre una población.

Dentro de las estrategias que se pueden proponer para la enseñanza de la inferencia estadística sobresale la simulación computacional.

Precisamente los resultados de la investigación que presentamos es la respuesta a la pregunta: *¿Cómo cambian los razonamientos inferenciales informales de los estudiantes de licenciatura en matemáticas respecto a las pruebas de hipótesis después de trabajar en situaciones de inferencia con simulación computacional?*

Con el fin de saber cómo cambian los razonamientos inferenciales informales de los estudiantes, hicimos una experimentación con cuatro estudiantes de licenciatura en matemáticas que, en ese momento, no habían realizado cursos de estadística a nivel universitario. La experimentación contó con tres etapas: en la primera los estudiantes presentaron una prueba inicial relacionada con situaciones que implican pruebas de hipótesis, con el fin de hacer una primera caracterización de sus razonamientos informales. Posterior a la prueba inicial, el grupo de estudiantes participó en cuatro sesiones de clase en las cuales se trabajaron situaciones relacionadas con inferencia informal y razonamiento inferencial informal sobre pruebas de hipótesis, mediadas por simulación computacional. En cada una de las sesiones de clase se realizó un taller, en el cual fueron analizadas y caracterizadas cada una de las respuestas de los estudiantes. En la última etapa de la experimentación los estudiantes presentaron una prueba final, similar a la prueba inicial, con el fin

de hacer una segunda caracterización de sus razonamientos informales y hacer el contraste entre ambas pruebas. Utilizamos el paquete estadístico *Fathom* y la taxonomía SOLO para describir los niveles de desarrollo que sufrieron los estudiantes en su razonamiento inferencial informal alrededor de las pruebas o contrastes de hipótesis.

Entre los hallazgos encontrados se destacan el desarrollo conceptual adquirido por los estudiantes respecto a los conceptos de variabilidad y distribución muestral. Partiendo de una concepción determinista y de un sesgo de equiprobabilidad respecto a los valores posibles de los estimadores, evolucionaron hasta alcanzar una concepción de omnipresencia de la variabilidad asociada a los valores del estimador muestral cuyo comportamiento, no obstante su carácter aleatorio, se manifiesta en unas regularidades marcadas por su distribución de probabilidad. La formación de concepciones intuitivas sobre los valores de significancia así como la distinción de las regiones de aceptación y rechazo también fueron evidentes en el proceso de instrucción llevado a cabo.

Este informe está constituido por 5 capítulos, los cuales se describen a continuación:

Antecedentes, en este capítulo presentamos una revisión de la literatura, donde analizamos las diferentes publicaciones (investigaciones, artículos, libros, etc.) relacionadas con las pruebas de hipótesis, sus dificultades, malas concepciones, su tratamiento informal y la simulación como estrategia docente. El capítulo está dividido en tres apartados: primero se exponen los trabajos de investigación relacionados con las dificultades y errores presentes en los estudiantes respecto a las pruebas de hipótesis; posteriormente, abordamos el razonamiento inferencial informal como

respuesta al problema anterior y, finalmente, el tercer apartado hace referencia a investigaciones relacionadas con la taxonomía SOLO y simulación estadística.

Referentes Teóricos y Conceptuales, En este capítulo se exponen los fundamentos teóricos y conceptuales de cada una de las tres directrices de nuestra investigación, a saber: las pruebas de hipótesis, el razonamiento inferencial informal y la taxonomía SOLO.

Metodología, En este capítulo exponemos el tipo de investigación que realizamos, la población de estudio, describimos la herramienta de simulación que utilizamos: *FATHOM*; exponemos los métodos de recolección de datos y presentamos las fases de la investigación: Diseño de instrumentos, Aplicación de instrumentos y Análisis de resultados.

Análisis de resultados, en este capítulo analizamos y caracterizamos las respuestas de los estudiantes en cada uno de los 6 instrumentos aplicados, (2 pruebas y 4 talleres) bajo el marco de la taxonomía SOLO.

Por último el capítulo de las *Conclusiones* donde se destacan los logros más importantes producto de la investigación realizada.

Incluimos, además, una sección adicional de *Anexos*, donde se incluyen los textos de los talleres realizados y de las pruebas aplicadas.

1. Antecedentes

Este capítulo comprende una revisión bibliográfica de investigaciones sobre el razonamiento inferencial informal y las pruebas de hipótesis, que se convirtieron en el punto de partida de nuestra investigación

Con el fin de ordenar de una manera lógica la revisión bibliográfica hemos decidido dividir el capítulo en tres apartados:

En el primer apartado se muestran las dificultades y errores que presentan los estudiantes, e inclusive algunos profesores, a la hora de afrontar la enseñanza y el aprendizaje de la inferencia estadística; haciendo especial hincapié en las relacionadas con las pruebas de hipótesis.

En el segundo apartado se presenta la inferencia estadística informal, específicamente el razonamiento inferencial informal como una respuesta de la comunidad de los investigadores en estadística educativa a la problemática tratada en el apartado anterior. Además se presentan algunas investigaciones de este tipo, referentes a las pruebas de hipótesis.

En el tercer apartado se presentan algunas investigaciones que involucran ciertos elementos importantes para nuestra investigación como son la taxonomía SOLO y el software de simulación estadística *Fathom*.

1.1. Dificultades y errores alrededor de las pruebas de hipótesis

Son muchas las investigaciones que se han realizado desde hace ya un tiempo en torno a las pruebas de hipótesis, esta parte de la inferencia es probablemente la peor comprendida, más confundida y de la que más se ha abusado en toda la estadística (Batanero, 1994). Una de las razones en la que se fundamenta el argumento anterior es la gran diversidad de conceptos implicados que deben ser comprendidos: Población y muestra, estadístico y parámetro, hipótesis (nula y alternativa), nivel de significación, valor-p, potencia, región crítica y de aceptación, errores de tipo I y de tipo II.

Peskun (1987) menciona dificultades de los estudiantes en los siguientes aspectos: la determinación de la hipótesis nula H_0 y la hipótesis alternativa H_1 ; la distinción entre los errores Tipo I y Tipo II; la comprensión del propósito y uso de las curvas características operativas o curvas de potencia; y la comprensión de la terminología empleada al establecer la decisión. En relación con este último aspecto, investigadores como Falk (1986) y Vallecillos (1991) evidenciaron dificultades y errores en los estudiantes con respecto a la comprensión del nivel de significancia (probabilidad de rechazar la hipótesis nula, cuando en realidad es cierta), que se expresa de la siguiente manera:

$$\alpha = P((\text{Rechazar } H_0 | H_0 \text{ es cierta}))$$

Ambos autores señalan el error que cometen los estudiantes al intercambiar los sucesos condición y condicionado en la definición anterior. Por ejemplo, Vallecillos (1991) evidenció este hecho cuando propuso a sus estudiantes valorar entre verdaderas y falsas las siguientes dos interpretaciones del nivel de significancia, y percatarse que muchos estudiantes catalogaban como verdadera la segunda interpretación.

- i) Un nivel de significación del 5% supone que, en promedio, 5 de cada 100 veces que la hipótesis nula es cierta, la rechazaremos.
- ii) Un nivel de significación del 5% implica que, en promedio, 5 de cada 100 veces que rechazemos la hipótesis nula, estaremos equivocados.

De igual forma, Vallecillos distinguió cuatro aspectos importantes para las pruebas de hipótesis y los errores alrededor de ellos:

- a) *El contraste de hipótesis como problema de decisión:* El contraste de hipótesis es un problema de decisión entre dos hipótesis complementarias y excluyentes, con la posible consecuencia de cometer dos tipos de error, incompatibles pero no complementarios. Aquí, Vallecillos encontró estudiantes que interpretan los errores tipo I y II como sucesos complementarios.
- b) *Las probabilidades de error y relación entre las mismas:* los dos tipos de error tienen probabilidades α (Tipo I) y β (Tipo II). Es necesaria la comprensión de las probabilidades condicionales que intervienen en la definición de α y β , de la dependencia de β como función del parámetro desconocido y de las relaciones entre α y β .
- c) *Nivel de significación como riesgo del decisor:* los valores de α y β determinan los riesgos que el decisor está dispuesto a asumir y servirán, junto con las hipótesis, para adoptar un criterio de decisión. En este aspecto encontró estudiantes que creen que el cambio del nivel de significación no afecta al riesgo de error en la decisión.
- d) *Interpretación de un resultado significativo.* La obtención de un resultado estadísticamente significativo lleva al rechazo de la hipótesis nula, aunque no implica necesariamente

ninguna relevancia desde el punto de vista práctico. Por ejemplo, una pequeña diferencia entre la media en dos poblaciones puede dar un resultado significativo si se toma una muestra de gran tamaño. Algunos estudiantes confunden la significación estadística y práctica o bien asocian un resultado significativo como uno que corrobora la hipótesis nula.

Otras de las dificultades que está bien referenciada por la literatura, consiste en que los estudiantes relacionan la prueba de hipótesis con una demostración matemática. Respecto al determinismo matemático aplicado en la inferencia, algunos estudiantes ven la prueba de hipótesis como una prueba matemática y, por lo tanto, al finalizar se ha probado la verdad o falsedad de una hipótesis (Vallecillos, 1999). Esta creencia es siempre errónea; la tienen los alumnos que tienen pocas bases matemáticas y una fe ciega en las matemáticas. De igual forma, Batanero (2015) refiriéndose al hecho de ver la prueba de hipótesis como un conjunto de reglas (una receta) dice: “la aplicación correcta de una prueba de hipótesis no implica que el estudiante comprenda y discrimine todos estos conceptos ni que haya adquirido suficiente razonamiento estadístico, sino que recuerda y sabe aplicar una serie de fórmulas, que quizás no comprenda” (p. 2).

Castro Sotos, Vanhoof, Van den Noortgate y Onghena (2007) interesados en conocer las principales dificultades y concepciones erróneas que tienen los estudiantes universitarios sobre los conceptos de inferencia estadística, realizaron una minuciosa revisión de publicaciones en las principales fuentes bibliográficas del campo. En dicha investigación encontraron sólo diecisiete estudios que proporcionaban evidencia empírica en un total de más de 500 artículos que fueron publicados sobre el tema en el período de 1990 a 2006. Los autores concluyen que muchas de las concepciones erróneas sobre las pruebas de hipótesis son derivadas de los libros de texto y que los profesores, incluso algunos estadísticos, comparten las mismas concepciones erróneas de

estudiantes. En la siguiente tabla se muestran las concepciones erróneas identificadas acerca de los conceptos que intervienen en el proceso de una prueba de hipótesis, algunas de las cuales ya fueron descritas en párrafos anteriores.

Tabla 1. *Concepciones erróneas sobre las pruebas de hipótesis*

Concepciones erróneas sobre las pruebas de hipótesis	
<i>Tipo de concepción errónea</i>	<i>Descripción</i>
Sobre los diferentes enfoques de las pruebas de hipótesis.	<ul style="list-style-type: none"> - Falta de distinción del paralelismo entre pruebas de hipótesis y procesos de decisión derivado de los enfoques de Fisher y Neyman-Pearson.
Sobre la definición de hipótesis.	<ul style="list-style-type: none"> - Confusión en la definición de hipótesis nula e hipótesis alternativa. - Confusión de la hipótesis nula con la región de aceptación. - Creencia de que una hipótesis puede referirse tanto a la población como a la muestra.
Sobre la naturaleza condicional del nivel de significancia.	<ul style="list-style-type: none"> - Invertir el condicional del nivel de significancia. - Interpretar el nivel de significancia como la probabilidad de que una de las hipótesis sea cierta. - Interpretar el nivel de significancia como la probabilidad de cometer un error. - Interpretar el valor de p como la probabilidad de que el evento sucedió por azar.
Sobre la interpretación de valores numéricos del valor de p.	<ul style="list-style-type: none"> - Interpretar el valor numérico del valor de p como un indicador de la intensidad del efecto del tratamiento o la variable bajo prueba.
Sobre la naturaleza de las pruebas de hipótesis.	<ul style="list-style-type: none"> - Considerar una prueba de hipótesis como una demostración matemática. - Considerar una prueba de hipótesis como una demostración probabilística de una de las hipótesis. Esta concepción también es llamada <i>ilusión de demostración probabilística por contradicción</i> o <i>ilusión de lograr la improbabilidad</i>.
Sobre la interpretación de la significancia estadística.	<ul style="list-style-type: none"> - Confundir significancia práctica y significancia estadística.

Nota. Cudaro tomado de Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W.; Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review* 2(2), 98-113.

Posteriormente, Castro Sotos, Vanhoof, Van den Noortgate y Onghena (2009) realizaron una investigación fundamentada en que a pesar de los considerables esfuerzos que se habían hecho durante las últimas décadas para reconceptualizar los cursos de estadística, los estudiantes todavía

sostienen ideas erróneas sobre la inferencia estadística. El estudio presentado por los investigadores aborda la necesidad de investigar más a fondo los conceptos erróneos sobre las pruebas de hipótesis desde dos puntos de vista: (1) la documentación de los conceptos erróneos más comunes entre los estudiantes universitarios de cursos introductorios de estadística; (2) Concentrar la investigación en las ideas erróneas que aún no han recibido mucha atención hasta ahora. Para tal fin, los investigadores aplicaron un cuestionario, en el cual, se pidió a 144 estudiantes universitarios seleccionar un nivel de confianza de 0 a 10 en cada una de sus respuestas. Los investigadores identificaron algunos de los errores que habían mencionado en la revisión bibliográfica de su investigación anterior: la definición de prueba de hipótesis, el valor p y el nivel de significancia. Una de las sugerencias que brinda la investigación para hacer frente a estos errores es enfrentar a los estudiantes con sus ideas falsas y con el hecho de que algunas ideas correctas podrían dar lugar a ideas erróneas en ciertas situaciones.

En el ámbito nacional, encontramos dos investigaciones: la primera sobre la práctica pedagógica de docentes de estadística inferencial y la segunda sobre el diseño de situaciones problemáticas referentes a la prueba de hipótesis.

Ortiz, Gaviria y Gil (2010) realizaron una investigación que buscaba caracterizar la práctica pedagógica de los docentes de Estadística Inferencial en las carreras de Administración de Empresas, Psicología e Ingeniería Industrial de la Pontificia Universidad Javeriana Cali. Los investigadores recolectaron datos a partir de entrevistas semiestructuradas, cuestionarios para los estudiantes y rejillas para la observación de la práctica pedagógica; y después de analizar los datos, concluyeron que los cursos de inferencia estadística abarcan muchos temas, y por consecuencia no

se puede profundizar en cada uno de ellos. Además, evidencian que los contenidos están enfocados únicamente al conocimiento teórico dejando de lado su comprensión.

En el libro “Diseño, desarrollo y evaluación de situaciones problemáticas de estadística. El caso de los tests de hipótesis” Fernández, Monroy y Rodríguez (1998) presentan una investigación alrededor de la prueba de hipótesis teniendo en cuenta tres facetas fundamentales: el contenido, el aprendizaje y la enseñanza. Respecto a la faceta del aprendizaje los investigadores consideran conceptos tales como error, dificultad y obstáculo presentes en las pruebas de hipótesis; y además, presentan una caracterización informal del aprendizaje y la comprensión de los estudiantes considerados en el estudio. Para finalmente culminar con tres situaciones problemáticas relacionadas con inferencia acerca de proporciones, de la media de dos poblaciones y del coeficiente de correlación de Pearson de una población.

1.2. El Razonamiento Inferencial Informal: Pruebas de Hipótesis

A raíz de la situación descrita en el apartado anterior, ha surgido una línea de investigación que se conoce como inferencia informal que está cobrando un gran auge dentro de la educación estadística. La Inferencia Estadística Informal nace de la necesidad de explorar otros caminos para abordar la inferencia estadística, pues investigadores como Batanero (2011) y Rossman (2011) han revelado las dificultades que presentan los estudiantes al enfrentarse a la inferencia estadística formal. La inferencia estadística informal se basa en un tipo de razonamiento que está a medio camino entre el análisis exploratorio de datos y la inferencia estadística formal. El uso de la palabra informal aquí sólo pretende destacar la amplia aplicación de razonamiento deductivo y abre la

posibilidad de considerar la inferencia estadística fuera de los procedimientos formales (García, 2013).

Uno de los primeros investigadores en trabajar esta nueva línea de investigación desde una perspectiva amplia fue Rossman (2008) quien expone una serie de ejemplos (en los cuales es necesario hacer inferencia de una muestra de datos) con los cuales se puede trabajar la inferencia estadística informal y el razonamiento inferencial informal. En cada uno de los ejemplos planteados Rossman hace la inferencia correspondiente después de realizar un análisis mediado por un software de simulación estadística. Finalmente, menciona algunas pautas que se deben tener en cuenta a la hora de trabajar problemas que requieran inferencia. Afirma, en primer lugar, que es importante que los estudiantes desarrollen la capacidad de no tomar decisiones apresuradas a la luz de los datos sino que tengan en cuenta el contexto, la naturaleza del experimento y las posibles variables que pueden haber influido en los resultados; también asegura que es más importante que los estudiantes comprendan la lógica de la inferencia estadística y la aleatoriedad por medio de las simulaciones que con el uso convencional de cálculos basados en la distribución normal.

De igual forma, uno de los grupos que se interesó en trabajar esta línea de investigación fue The International Collaboration for Research on Statistical Reasoning Thinking and Literacy (SRTL) quienes desde 1999 ofrecen encuentros cada dos años para fomentar los estudios de investigación actuales e innovadores que examinan la naturaleza y el desarrollo de la alfabetización, el razonamiento y el pensamiento estadístico. En el SRTL-5 realizado en el 2007 Zieffler, Garfield, Delmas, y Reading (2007) hacen una revisión bibliográfica en la que buscan unificar las diferentes apreciaciones sobre la Inferencia Estadística Informal (IEI) y el Razonamiento Inferencial

Informal (RII); definiendo el Razonamiento Inferencial Informal como *“la forma en que los estudiantes usan sus conocimientos informales de estadística para crear argumentos basados en muestras observadas que sustenten las inferencias hechas sobre la población desconocida”* (p. 44). Dentro de este trabajo los autores identifican tres tipos de problemas que se pueden abordar desde el RII: Hacer gráficos a partir de muestras, Comparación de dos muestras y Elegir entre modelos. Este último tipo hace referencia a las pruebas de hipótesis que trabajamos en esta investigación.

Además, en este marco de apoyo para la investigación del RII se sugieren dos objetivos: 1) descubrir y describir formas en que sea posible que los estudiantes desarrollen ideas centrales de la inferencia estadística sin utilizar el aparato matemático que las fundamenta, y 2) crear un repertorio de problemas y actividades relacionadas con la inferencia estadística informal que jueguen un papel de antecedente sobre el cual se puedan construir los conocimientos formales de la inferencia estadística. A partir de estos dos objetivos se han realizado algunas investigaciones que buscan en la inferencia informal y el razonamiento inferencial informal un salvavidas para la enseñanza de la inferencia estadística. Como es evidente que el objetivo de esta publicación es explicar y brindar bases para las investigaciones enmarcadas dentro de la IEI y el RII; por lo cual, la retomaremos en la siguiente sección para explicar el razonamiento inferencial informal referente a las pruebas de hipótesis.

Una de las investigaciones que se hicieron a partir del marco de apoyo de Zieffler y sus colegas fue realizada por García y Sánchez (2014) quienes trabajaron las pruebas de hipótesis con Razonamiento Inferencial Informal en estudiantes de bachillerato. En esta investigación se

centraron en los elementos que intervienen en el razonamiento del estudiante de bachillerato al hacer inferencias estadísticas sin los métodos y técnicas formales. Es importante resaltar la definición que dan dentro de su marco conceptual: La inferencia estadística informal es una generalización probabilística (no determinista) de los patrones que son revelados por los datos disponibles, y esta generalización es el producto final de un Razonamiento Inferencial Informal. Una inferencia estadística informal se representa mediante un enunciado, mientras que el razonamiento inferencial informal es el proceso mediante el cual se descubren y establecen dichos enunciados.

La experiencia realizada por García y Sánchez se llevó a cabo con 16 estudiantes entre los 16-17 años a los cuales se les aplicó una prueba escrita con dos problemas relacionados con pruebas de hipótesis. El método de análisis no utilizó entrevista, se limitó únicamente al análisis y clasificación de las respuestas de cada estudiante en tres aspectos fundamentales: La hipótesis nula, la argumentación y la inferencia estadística informal. Después del análisis los investigadores detectaron tres tipos de razonamiento en las respuestas dadas: 1) considerar un modelo o distribución inicial de la población (hipótesis nula) y comparar los datos de la muestra con este modelo, 2) apoyar el razonamiento en el uso de su conocimiento informal y no analizar los datos de la muestra y 3) combinar los dos razonamientos anteriores. De igual manera, concluyen que el contexto juega un papel importante; puesto que los estudiantes utilizaron conocimiento informal sobre el contexto del problema al establecer la hipótesis nula y al explicar las causas de los datos; por lo tanto, si el conocimiento informal sobre el contexto del problema es erróneo las inferencias no serán las adecuadas.

Siguiendo esta línea de trabajo Batanero y Díaz (2015) proponen aproximaciones informales a las pruebas de hipótesis, en donde resaltan la importancia que tiene hoy en día la inferencia estadística y la gran cantidad de conceptos y fórmulas que se deben tener presente a la hora de hacer inferencia con pruebas de hipótesis. La propuesta de las investigadoras es abordar la inferencia estadística con ayuda de la tecnología, utilizando un *applet* que permite hacer inferencias informales, prescindiendo de los mecanismos formales de la estadística.

1.3. Sobre los elementos de la investigación

Este apartado incluye algunas investigaciones que trabajan con dos de los elementos que usamos en nuestra investigación: La taxonomía SOLO y la simulación estadística. En la primera parte se presentan investigaciones relacionadas con la taxonomía SOLO, la cual usamos en nuestra investigación para clasificar los razonamientos que presenten los estudiantes. En la segunda parte se presentan investigaciones y observaciones de importantes investigadores en torno a la utilización del *software* de simulación de datos como herramienta didáctica. La inclusión de este tipo de investigaciones se justifica, en la utilización del *software Fathom* como herramienta de simulación.

1.3.1. Sobre la Taxonomía SOLO

Inzunsa y Jiménez (2013) hacen una caracterización utilizando la taxonomía SOLO del razonamiento estadístico de estudiantes universitarios sobre las pruebas de hipótesis. Los investigadores trabajaron con estudiantes de Licenciatura en Matemáticas que tomaron un curso de formación estadística que contenía el tema de prueba de hipótesis. El objetivo de la

investigación era caracterizar los razonamientos de los estudiantes al enfrentarse a un cuestionario referente al tema.

En el análisis de resultados los investigadores notaron que los razonamientos de los estudiantes se encuentran su gran mayoría en los dos primeros niveles del modelo SOLO: Preestructural y Uniestructural. De igual manera, los resultados de la investigación mostraron que las pruebas de hipótesis son un concepto complejo para los estudiantes universitarios, aun cuando han tomado cursos de estadística matemática y fundamentos de teoría de la probabilidad. Finalmente, concluyen que el razonamiento estadístico de los estudiantes que participaron en la investigación se caracteriza por ser aislado en relación con los diversos conceptos que se involucran en las pruebas de hipótesis, esto derivado de la falta de comprensión y por creencias erróneas sobre diversos conceptos involucrados.

Jaimes (2011) realizó una investigación para identificar y evaluar los niveles de razonamiento probabilístico involucrando la noción de distribución de una variable con estudiantes de tercer grado de secundaria. Para tal fin, diseñó y aplicó un instrumento de evaluación apoyado en el modelo de taxonomía SOLO (Bigg y Collis, 1982) para valorar las respuestas de los estudiantes. El instrumento consistió en doce preguntas asociadas a doce tareas específicas como la generación de secuencias aleatorias, predicciones tabulares y gráficas de frecuencias absolutas y relativas a corto, mediano y largo plazo, así como con tareas de argumentación de sus respuestas y a la escritura de nuevas conjeturas asociadas a la noción de probabilidad frecuencial. Este instrumento se aplicó antes de la instrucción como test de diagnóstico y después de la instrucción como test de

post-experimentación física y computacional, para comparar las respuestas y categorizar las respuestas de acuerdo al modelo SOLO.

En esta investigación de Jaimes (2011) es importante destacar la manera como caracteriza cada uno de los niveles de SOLO para cada una de las preguntas definiendo componentes de respuestas que deben presentar los estudiantes. Por ejemplo, para la siguiente situación:

La familia Pérez está compuesta por el señor Carlos, su esposa Ana y su hijo Beto. Todas las noches, por lo general, después de cenar se reúne la familia a ver la televisión, pero nunca están de acuerdo para ver un mismo programa. Al señor Carlos le gusta ver programas deportivos, a la señora Ana las novelas y a Beto las películas de acción.

Como sólo hay un televisor en la casa, lo más sencillo sería que se turnaran el control de la televisión diariamente, pero Beto les propone a sus padres algo más divertido: “¡A la suerte!”. Propone rifar el control jugando a los volados con dos monedas de la siguiente manera: Si no sale ningún águila en los dos volados gana la Sra. Ana; si sale exactamente un águila gana el niño Beto y si salen dos águilas gana el Sr. Carlos. A los padres les parece justo y aceptan su propuesta.

Propone la tarea: *Escribe la secuencia de resultados que crees que piensas pueden ocurrir al jugar durante los primeros 20 días.* En esta tarea de generar distribuciones aleatorias se tuvieron en cuenta los siguientes componentes para evaluar la respuesta de cada estudiante:

- a. La escogencia de una persona por sorteo. Verificar que en el total, la suma corresponda al número de sorteos.

- b. La consideración de las rachas. Una secuencia adecuada tiene al menos dos rachas de longitud 3, 4 o más.

- c. Asignar mayor frecuencia al evento con mayor probabilidad (gana Beto).
- d. La no presencia de patrones determinístico en la secuencia (sentido aleatorio).

La manera en que se clasifican las respuestas de los estudiantes en los niveles de SOLO está dada por el número de componentes presentes en la respuesta; es decir, si la respuesta no presentaba ninguna componente, se ubicaba en un nivel Preestructural; si la respuesta presentaba una componente, se ubicaba en el nivel Uniestructural; pero, si por el contrario, la respuesta presentaba dos componentes, se ubicaba en el nivel Multiestructural; si presenta tres de las componentes, se ubica en el nivel relacional; y si la respuesta presenta las cuatro componentes, se ubica dentro del nivel de Abstracción Extendida.

1.3.2. La simulación de datos como herramienta didáctica

Según Cross Bu (2005) en su libro “*Simulación. Un enfoque Práctico*” dice: “Los *software* de simulación de datos y las técnicas de simulación pueden ser utilizadas como un instrumento pedagógico para enseñar a los estudiantes habilidades básicas en análisis estadístico, análisis teórico, etc.” (p. 17). En este mismo libro, Cross Bu (2005) enuncia algunas ventajas y desventajas de la simulación aplicables a las simulaciones estadísticas. Entre las ventajas están el uso de la simulación puede ser utilizada para experimentar nuevas situaciones, de las que no se tiene información suficiente por lo que ayudaría al mejor entendimiento de dicho sistema y a conocer el comportamiento de este; también ayuda a conocer cómo funciona un sistema al introducirse nuevos elementos y, además, en la simulación se tiene control sobre las variables para así poder generar las condiciones necesarias para cumplir con los objetivos. Entre las desventajas: La

simulación es imprecisa, porque no se puede simular toda la realidad; se pueden obtener resultados falsos si no se tienen las variables correctas o no se realizan suficientes repeticiones.

En el caso concreto de la simulación estadística, Batanero (2009) ve la simulación como una herramienta creada como ayuda para entender los problemas de probabilidad. En palabras de Batanero “*La ventaja de la simulación es obvia, ... permite condensar el experimento en un tiempo y espacio concreto*” (p. 45). De igual manera, otro aspecto destacable de estos *software* de simulación es la capacidad repetir un experimento cuantas veces sea necesario.

Un trabajo específico acerca de la simulación con *Fathom* (*software* de simulación de datos que utilizaremos en nuestra investigación) es el realizado por Ramírez (2008) quien investiga acerca de las formas de razonamiento que muestran estudiantes de maestría de Matemática Educativa sobre la distribución normal mediante problemas de simulación en *Fathom*. En sus conclusiones hace referencia al *software Fathom* de la siguiente manera: “los estudiantes usaron el *software* como un instrumento que permitió erradicar o modificar las conceptualizaciones erróneas que tenían los estudiantes sobre la distribución normal. El *software* llega a dar evidencias creíbles para el estudiante, esto se manifiesta a través del orden, la consistencia y la retroalimentación inmediata que ofrece el programa. Además, permite que el estudiante fije su atención en el proceso de creación y caracterización de la distribución más que en el simple cálculo de una probabilidad donde se apliquen de manera mecánica una serie de fórmulas.” (p. 9).

Finalmente, a pesar de las evidentes ventajas de los *software* de simulación de datos, existe el riesgo de utilizar mal esta tecnología. Batanero (2008) resalta este hecho en algunas propuestas didácticas sobre inferencia informal, donde sólo se limitan a enseñar la inferencia informal como

un conjunto de recetas: a) simular la extracción de muestras; b) formar una distribución muestral empírica; c) calcular la probabilidad de obtener el valor observado del estadístico usando esta distribución muestral empírica; d) tomar una decisión.

2.- Referentes teóricos y conceptuales

El capítulo de referentes teóricos y conceptuales se divide en dos partes en las cuales se tratan las tres directrices fundamentales de esta investigación: *la inferencia estadística a través de la prueba de hipótesis, el razonamiento inferencial informal y la taxonomía SOLO*. El siguiente esquema esboza cómo vinculamos los elementos presentes en esta investigación.



Figura 1. Esquema de los referentes de la investigación

De afuera hacia dentro del esquema: utilizando la taxonomía SOLO, la cual permite clasificar las respuestas de los estudiantes a partir de niveles, observaremos el razonamiento informal de los estudiantes cuando se enfrentan a situaciones estadísticas en las que deban hacer inferencias, más específicamente, pruebas de hipótesis.

2.1. Eje Conceptual

2.1.1. Inferencia Estadística: Prueba de Hipótesis

Moore (2005) define de forma simple, pero concreta la inferencia estadística como la parte de la estadística que proporciona métodos que permiten sacar conclusiones de una población a partir de los datos de una muestra. Por su parte, Batanero (2001) dice que “la inferencia estadística estudia los resúmenes de datos con referencia a un modelo de distribución probabilístico o una familia de modelos, determinando márgenes de incertidumbre en la estimación de los parámetros desconocidos del mismo. Se supone que el conjunto de datos analizados es una muestra de una población y el interés principal es predecir el comportamiento de la población a partir de los resultados en la muestra” (p. 10).

Según Moore (2005) los métodos más utilizados en inferencia estadística son los intervalos de confianza y las pruebas de hipótesis. Ambos métodos son un producto del siglo XX. A partir de un complejo y a veces confuso origen, las pruebas estadísticas tomaron su forma actual de los escritos de R. A. Fisher y los intervalos de confianza aparecieron en 1934 gracias al ingenio de Jerzy Neyman (1894-1981). Neyman dio un nuevo enfoque a las pruebas de significación al incluir la hipótesis alternativa.

Continúa diciendo Moore (2005) que las pruebas de hipótesis, tienen como objetivo valorar la evidencia proporcionada por los datos a favor de alguna hipótesis sobre la población. Aunque estas pruebas utilizan una terminología muy elaborada, la idea básica es sencilla: si suponemos que una determinada afirmación es cierta y bajo esta suposición observamos que un determinado suceso ocurre muy raramente, esto indica que la afirmación no es cierta. El razonamiento utilizado con

las pruebas de hipótesis, se basa en preguntar lo que ocurriría si repitiéramos el muestreo o el experimento muchas veces.

Conceptos

A continuación se definen algunos de los conceptos involucrados con las pruebas de hipótesis:

HIPÓTESIS NULA H_0 : La afirmación que se contrasta en una prueba estadística se llama hipótesis nula. Las pruebas de hipótesis se diseñan para valorar la fuerza de la evidencia en contra de la hipótesis nula. En general, la hipótesis nula es una afirmación de “ausencia de efecto” o de “no diferencia”.

HIPÓTESIS ALTERNATIVA H_a : Es la afirmación en relación con la población sobre la cual queremos hallar evidencia a favor, designada como H_a .

P-VALOR: La probabilidad, calculada suponiendo que H_0 es cierta, de que el resultado tome un valor al menos tan extremo como el observado se llama *valor p* de la prueba de significación. Cuanto menor sea el *valor p*, más fuerte es la evidencia que proporcionan los datos en contra de H_0 .

NIVEL DE SIGNIFICANCIA: Algunas veces damos un último paso para valorar la evidencia en contra de H_0 . Comparamos el valor p con un valor previamente determinado que consideramos decisivo. Esto equivale a decidir de antemano cuál consideramos que tiene que ser la evidencia en contra de H_0 . El valor p decisivo se llama nivel de significación.

ESTADÍSTICO DE CONTRASTE: Una prueba de significación utiliza los datos en forma de estadístico de contraste. Éste se basa, normalmente, en un estadístico que estima el parámetro que aparece en las hipótesis.

ERRORES: De igual forma, es importante destacar que las pruebas de hipótesis tienen unos errores ligados a la elección de la hipótesis. Si rechazamos H_0 cuando en realidad H_0 es cierta, cometemos un Error Tipo I. Si no rechazamos H_0 cuando en realidad H_a es cierta, cometemos un Error Tipo II.

La Prueba de Hipótesis

En esta investigación nos centramos en dos formas de la prueba de hipótesis:

La prueba de hipótesis para la media poblacional

La prueba consiste en contrastar la hipótesis nula $H_0: \mu = \mu_0$ a partir de una muestra aleatoria simple de tamaño n de una población con media desconocida (μ) y desviación típica (σ) conocida.

Se calcula el *estadístico de contraste* z

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

En términos de una variable Z que tiene una distribución normal estandarizada.

Para contrastar H_0 en contra de las siguientes hipótesis alternativas, los p-valores son los siguientes:

$$H_a: \mu > \mu_0 \text{ es } P(Z \geq z)$$

$$H_a: \mu < \mu_0 \text{ es } P(Z \leq z)$$

$$H_a: \mu \neq \mu_0 \text{ es } 2P(Z \leq |z|)$$

Estos p-valores son exactos si la distribución poblacional es normal y aproximadamente correctos para valores de n grandes, en los restantes casos.

Prueba de hipótesis para una proporción poblacional

La prueba consiste en contrastar la hipótesis nula $H_0: p = p_0$ a partir del cálculo del estadístico:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

En términos de una variable Z que tiene una distribución normal estandarizada.

Para contrastar H_0 en contra de las siguientes hipótesis alternativas, los p-valores son los siguientes:

$$H_a: p > p_0 \text{ es } P(Z \geq z)$$

$$H_a: p < p_0 \text{ es } P(Z \leq z)$$

$$H_a: p \neq p_0 \text{ es } 2P(Z \leq |z|)$$

Ahora bien, como se puede apreciar, las pruebas de hipótesis involucran una gran diversidad de conceptos que son necesarios para comprender su correcta aplicación e interpretación. En el siguiente mapa conceptual adaptado de Lipson (2000) se muestran los conceptos que intervienen en una prueba de hipótesis y la forma como estos se relacionan:

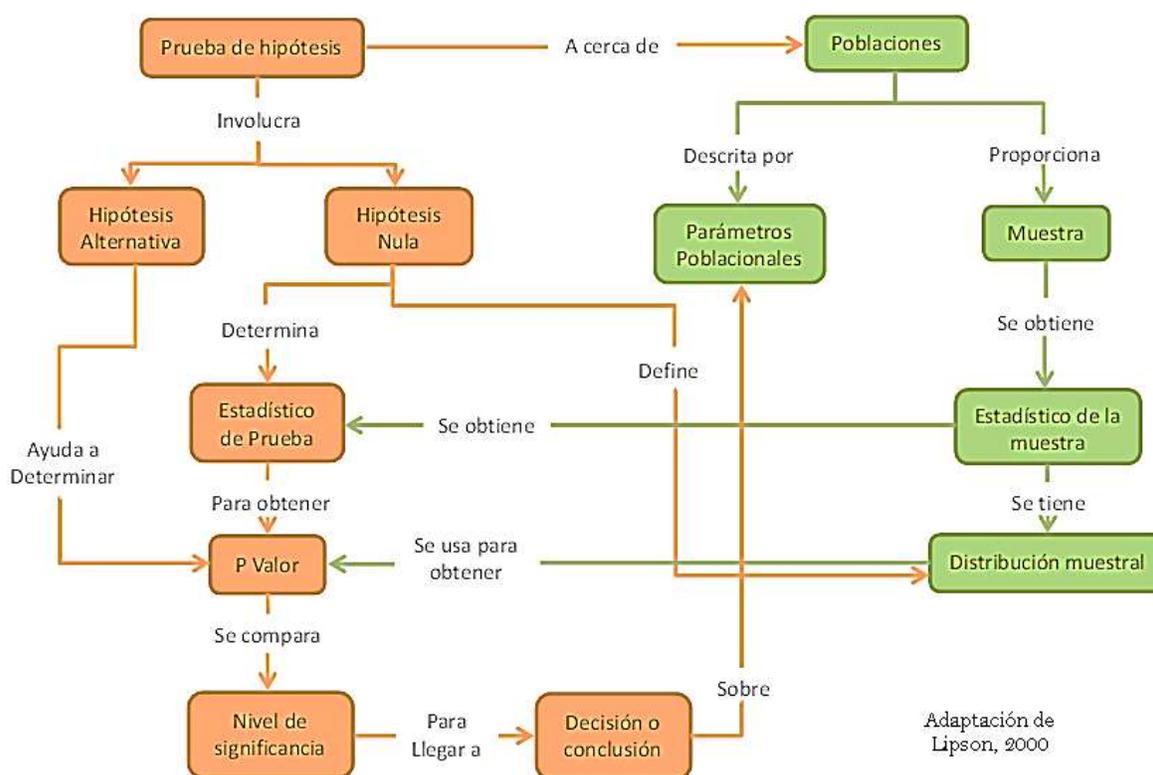


Figura 2. Mapa conceptual de los elementos de las pruebas de hipótesis, Adaptado de Lipson 2000

Descripción del mapa conceptual: las pruebas de hipótesis se hacen a muestras aleatorias para generar conclusiones sobre la población a la cual pertenecen dichas muestras, la prueba implica dos tipos de hipótesis, a saber, hipótesis nula e hipótesis alternativa; a partir de la hipótesis nula se determina el estadístico de la prueba, el cual se utiliza para obtener el p-valor. El p-valor obtenido se compara con un nivel de significancia previamente asignado para así generar una conclusión sobre los parámetros poblacionales que describen la población de la cual se sacó la muestra.

Ahora bien, a partir de la muestra proporcionada por la población se obtiene el estadístico de la muestra, que además de ser necesario para hallar el estadístico de la prueba previamente descrito,

también proporciona la distribución muestral. Finalmente, esta distribución muestral está definida por la hipótesis nula y se usa para obtener el p-valor.

2.1.2. La Inferencia Estadística Informal y el Razonamiento Inferencial Informal

2.1.2.1. Razonamiento Informal

La inferencia estadística informal es una línea de investigación de la educación estadística; que al ser relativamente nueva no tenía una definición concreta y consensada por la comunidad estadística de los conceptos y términos presentes en ella. Por ejemplo, Pfannkuch (2006) afirma que “el razonamiento inferencial informal está interconectado con el razonamiento de las distribuciones, el razonamiento de medidas de tendencia central, y el razonamiento de muestreo dentro de un ciclo de la investigación empírica” (p. 1). Mientras que Ben-Zvi (2006) comenta que la inferencia informal está relacionada con actividades de argumentación, derivando conclusiones lógicas a partir de los datos acompañados de argumentos persuasivos basados en el análisis de los datos.

Como se mencionó en los antecedentes, Zieffler y sus colaboradores (2008) con el fin de unificar todas las concepciones que se tenían hasta ese entonces sobre el razonamiento inferencial informal realizaron un marco de apoyo para la investigación sobre este tema. La definición que presentan Zieffler y sus colaboradores para el razonamiento informal es la que adoptaremos en esta investigación, *el razonamiento inferencial informal es la manera en la cual los estudiantes usan sus conocimientos estadísticos para formular argumentos que apoyen sus decisiones.*

Ahora bien, *el conocimiento informal*, según Sánchez y Ríos (2013) puede ser visto de dos formas: un conocimiento cotidiano del mundo que los estudiantes poseen con base en las experiencias fuera de la escuela; o un conocimiento menos formal de los temas que resultan de la enseñanza formal previa. En esta investigación adoptaremos el conocimiento informal como la articulación de las dos formas mencionadas.

Finalmente, según Zieffler y colaboradores (2008) la literatura de investigación sobre el razonamiento informal y el conocimiento informal sugiere que las tareas deben ser diseñadas para provocar múltiples argumentos de estudiantes, y a su vez deben desafiar al estudiante a:

1) Hacer juicios, o predicciones sobre una población a partir de muestras, pero sin utilizar procedimientos y métodos (por ejemplo, p-valor, pruebas t) estadísticos formales.

2) Aprovechar, utilizar e integrar el conocimiento previo (formal e informal) en la medida en que este conocimiento esté disponible.

3) Articular argumentos basados en la evidencia para emitir juicios, reclamos, y predicciones sobre poblaciones basados en muestras.

En el siguiente ejemplo propuesto Rossman (2008) se describe como se trabajan los problemas a través del razonamiento inferencial informal

Se presenta a los estudiantes un estudio aplicado a un grupo de 16 bebés de 10 meses de edad en el cual se pretendía estudiar si los bebés generan algunas preferencias por las otras personas basadas en sus acciones. Para tal estudio se presentó a los bebés dos escenas diferentes sobre un escalador.

En la primera el escalador es ayudado a escalar por otro personaje al cual llamaremos “helper”, y en la segunda el escalador es retrasado por un personaje diferente al que llamaremos “hinderer”. Después de mostrar la escenas en varia ocasiones a los bebés se les permite elegir entre uno de los dos personajes para jugar. De los 16 bebes, 14 prefirieron jugar con “helper”.

Se le preguntó a los estudiantes si este resultado proporciona evidencia de una preferencia genuina, ya sea entre una mayor población de niños, o entre estos mismos 16 bebés si fueran a ser probado repetidamente.

Simulando esta situación para 1000 muestras de 16 bebés se puede observar que es poco probable ($2/1000$) que se obtengan 14 éxitos solo por el azar, lo que hace sospechar que el resultado no ha sido solo producto de la aleatoriedad.

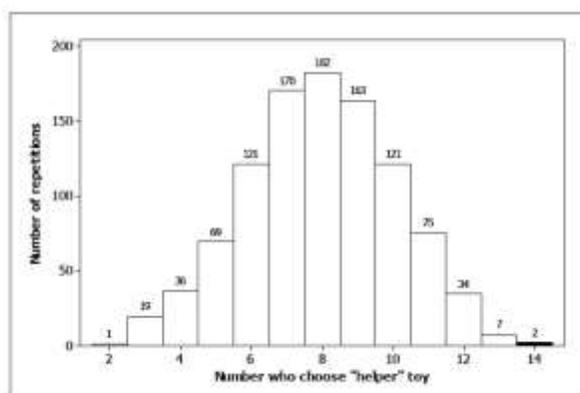


Figura 3. Simulación de resultados para el juguete "Helper".

Este tipo de razonamiento se considera informal, pues en ningún momento se habla de calcular una probabilidad exacta usando una distribución binomial, ni del cálculo de un p-valor basado en una distribución normal.

2.1.2.2. Razonamiento informal para la prueba de hipótesis

Dentro de los tres tipos de problemas que se pueden abordar desde el razonamiento inferencial informal, esta investigación se centra en el tercer tipo: La elección entre modelos, puesto que estos problemas están relacionados con las pruebas de hipótesis. Para los problemas de elección de modelos Zieffler y sus colaboradores (2008) identifican tres estilos de problemas para abordar el tema.

El primer estilo es utilizar los datos generados por un dispositivo de probabilidad, y usa proporciones o porcentajes que se resumen los datos de la muestra. Zieffler (2008) referencia este estilo de problema citando a Stohl y Tarr (2002) quienes utilizaron el problema *Schoolopoly* que pedía a estudiantes de sexto grado juzgar si un dado era justo o no, a partir de lanzamientos observados a través de un simulación por ordenador.

El segundo estilo consiste en determinar si un cambio particular en algún proceso, haya ocurrido o no ocurrido durante un intervalo de tiempo determinado. Un ejemplo de este estilo fue utilizado por Rubin, Hammerman, y Konold (2006) a través de la siguiente situación:

La Mus-Brush Company produce cepillos de hongos, utilizando una máquina de gran tamaño cuya salida es en promedio 215 cepillos cada dos minutos si está funcionando normalmente. Si se interrumpe la electricidad

a la máquina, aunque sea por un breve tiempo, será más lento de tal manera que la salida de la máquina será un 10% inferior en promedio. La Mus-Brush Company fue robada la noche anterior; forzaron la puerta para abrirla, el ladrón interrumpió la electricidad y la máquina se hizo menos productiva a partir de entonces. Hay un sospechoso que tiene una coartada entre la medianoche y las 3 a.m. (fue visto en un bar), por lo que la policía tiene un interés especial en la determinación de si el robo se produjo antes de la medianoche o después de 3, ya que el sospechoso no tiene coartada para ese intervalo de tiempo. Tenemos datos sobre la producción de Mus-Brush cada dos minutos desde las 8 p.m. hasta las 5 am. Nuestro trabajo es decidir si hay suficiente evidencia para argumentar que el robo ocurrió entre 12 y 3, consiguiendo así el sospechoso estar fuera del caso.

El tercer estilo consiste en decidir qué tan raro es un resultado en una muestra de datos cuantitativos. El problema de los bebés planteado por Rossman (2008) hace parte de este estilo de problemas.

Las situaciones que planteamos en la investigación están enfocadas en este último estilo de problemas.

2.2. Eje Didáctico

2.2.1. Taxonomía SOLO

La taxonomía SOLO (The Structure of Observed Learning Outcomes: Estructura de los resultados observados del aprendizaje) de Biggs y Collis (1991). Esta teoría de la inteligencia permite clasificar las respuestas de los estudiantes de acuerdo al número de elementos, o componentes conceptuales, o procedimientos que utilicen y relacionen.

Los aspectos centrales de la taxonomía SOLO son tres: la *forma particular del conocimiento* (se refiere a las expresiones de competencia de los estudiantes en cualquier contexto), los *modos de funcionamiento o representación* (relacionado con la etapas piagetianas del desarrollo cognitivo) y los *ciclos del aprendizaje o niveles de logros* (para describir la estructura de cualquier respuesta).

2.2.1.1. Formas de Conocimiento

El resultado del comportamiento competente es el logro de una forma particular de conocimiento como los siguientes:

- Conocimiento tácito: es aquel que se manifiesta haciendo y normalmente no es accesible verbalmente. Se evidencia cuando la persona desarrolla una actividad, aunque no tenga un fundamento teórico-argumentativo que lo sustente.
- Conocimiento intuitivo: se percibe directamente o se siente, se manifiesta cuando se da solución a un problema antes de que se pueda expresar simbólicamente.
- Conocimiento declaratorio: es expresado a través de un sistema simbólico, que es altamente entendible.

- Conocimiento teórico: está en un nivel superior de abstracción, está fundamentado en teorías racionales.

Destacamos que las dos primeras formas de conocimiento: tácito e intuitivo son propios del razonamiento inferencial informal que trabajamos en esta investigación.

2.2.1.2. Modos de funcionamiento

El origen de la taxonomía SOLO se puede encontrar en formulaciones piagetianas y reformulaciones neopiagetianas posteriores (Biggs y Collis, 1991) de las etapas del desarrollo cognitivo de Piaget. Este modelo teórico surgió de ciertas críticas a la tendencia piagetiana, la cual considera que el desarrollo cognoscitivo se da por etapas discretas evolutivas, en las cuales cada vez se alcanzan dominios más estructurados y complejos.

Para los neopiagetianos era engañoso clasificar a los estudiantes basándose en los estadios de Piaget, porque era muy común en el contexto escolar encontrar que no hay uniformidad en los desempeños: Un estudiante podía ser “Preformal” en matemática mientras que en historia podría ser “Preconcreto” o, incluso, “Preformal” en matemática un día y “Concreto” el siguiente (Biggs y Collis, 1982). Estas observaciones, dicen, no pueden indicar cambios en el desarrollo cognitivo, sino más bien, cambios en constructos más próximos como el aprendizaje, la actuación o la motivación de los estudiantes. Sin embargo los neopiagetianos como Biggs y Collis (1991) están de acuerdo en que los modos de funcionamiento, son niveles de abstracción que van progresando desde las acciones concretas a los principios y conceptos abstractos.

Huerta (1999) describe los modos de funcionar y las edades en las que normalmente emergen:

a) *Sensoriomotor* (desde el nacimiento): El niño sólo puede interaccionar con el mundo de la manera más concreta: dando una respuesta motriz a un estímulo sensorial.

b) *Icónico* (desde los 18 meses): Si una acción se vuelve más abstracta debe ser representada de alguna forma. La forma más simple de interiorizar una acción es imaginarla, formando una imagen interna o icono.

c) *Simbólico concreto* (desde los 6 años): Se inicia un cambio significativo de la abstracción desde la simbolización directa del mundo a través del lenguaje oral al escrito, sistemas de símbolos que se aplican al mundo de las experiencias.

d) *Formal* (desde los 16 años): El razonamiento formal se produce en un sistema abstracto supraordinado en el que pueden generarse formas alternativas de ordenar el mundo.

e) *Postformal* (desde los 18 años): El pensamiento postformal supone preguntarse por las fronteras convencionales de la teoría y de la práctica. Es ajeno a cualquier nivel educativo y puede aparecer en el campo de la investigación.

2.2.1.3. Ciclos del aprendizaje

El logro o calidad del desempeño se refleja “(...) en el nivel de complejidad con el cual se estructura el resultado del aprendizaje, independientemente de que el ítem aprendido sea una habilidad, un concepto o un problema” (Biggs, 1988, p. 197).

Biggs y Collis (1991) describen cinco niveles básicos en el ciclo de aprendizaje que no clasifican el rendimiento del estudiante, pero sí clasifican las respuestas en una escala de cinco

categorías en un tiempo determinado y respecto a una tarea específica. Los cinco niveles básicos en el ciclo de aprendizaje son:

1. **Prestructural** Respuestas centradas en los aspectos irrelevantes de la propuesta de trabajo, con contestaciones evasivas o tautológicas del enunciado.

2. **Unistructural** Respuestas que contienen datos informativos obvios, los cuales han sido extraídos directamente del enunciado.

3. **Multiestructural** respuestas que requieren la utilización de dos o más informaciones del enunciado, las cuales siendo obtenidas directamente de éste, son analizadas separadamente, no de forma interrelacionada.

4. **Relacional** Respuestas extraídas tras el análisis de los datos del problema, integrando la información en un todo comprensivo. Los resultados se organizan formando una estructura.

5. **Abstracción Extendida** Respuestas que manifiestan la utilización de un principio general y abstracto que puede ser inferido a partir del análisis sustantivo de los datos del problema y que es generalizable a otros contextos.

Los niveles más elevados de la taxonomía corresponden a un aprendizaje más profundo, a una interpretación personal del contenido que relaciona la tarea con situaciones alejadas del contexto inmediato, que establece relaciones con otros conocimientos relevantes y con materiales procedentes de diferentes fuentes de información. Contrariamente, los niveles inferiores corresponden al tratamiento de la información de manera aislada y reproductiva.

3.- Metodología

En este capítulo se mostrará cómo se llevó a cabo la investigación, la población participante en la experimentación, el *software* de simulación que se utilizó, la recolección de datos y las fases de la investigación.

Esta investigación es un estudio de casos, en el que se caracterizaron los razonamientos informales de los estudiantes cuando resuelven situaciones de inferencia relacionadas con pruebas de hipótesis. Esta caracterización se hizo a partir de los razonamientos que presentaron en sus respuestas los estudiantes, en dos pruebas referentes a las pruebas de hipótesis aplicadas en distintos periodos de tiempo.

3.1. Generalidades

A continuación se describirán los aspectos generales de la metodología de la investigación.

3.1.1. Participantes

La población que hizo parte de la experimentación fueron ocho estudiantes de primeros niveles de la Universidad Industrial de Santander (UIS), pertenecientes al programa de Licenciatura en Matemáticas. En esta investigación presentamos los resultados de cuatro de los estudiantes, escogidos a partir de los siguientes criterios: Casos representativos, los estudiantes escogidos presentaron respuestas similares a los demás participantes de la experimentación; Realización completa de la experimentación, los estudiantes participaron en cada una de las sesiones de la experimentación.

La razón de la elección de estudiantes de Licenciatura en Matemáticas de primeros niveles se debió a las siguientes razones:

Los estudiantes fueron seleccionados del programa de Licenciatura en Matemáticas, puesto que junto con la Maestría en Educación Matemática pertenecen a la Escuela de Matemáticas de la Universidad Industrial de Santander; lo cual garantizaría, de cierto modo, un mayor compromiso por parte de los estudiantes con la experimentación, y así evitar en abandono de la misma.

Los estudiantes no habían tomado hasta el momento ninguno de los cursos de formación estadística, con los que dispone el programa de Licenciatura en Matemáticas de la Universidad Industrial de Santander. Por lo cual carecen de los formalismos propios de la inferencia, incluso de una formación matemática sólida. Esta condición los hace los sujetos de estudio indicados, pues al momento de enfrentarse a las situaciones planteadas se verán obligadas a recurrir a sus conocimientos informales, los cuales proporcionan los razonamientos informales que estamos interesados en estudiar.

Los cuatro estudiantes que participaron de la experimentación los llamaremos, a partir de ahora: Andrés, Sofía, Gabriela y Luis. Al momento de la experimentación, Andrés y Luis se encontraban en segundo semestre de la licenciatura; Gabriela se encontraba en primer semestre; mientras que Sofía se encontraba en tercer semestre.

3.1.2. *Fathom* como herramienta didáctica

En esta investigación se pretendía potenciar el desarrollo del razonamiento informal de los estudiantes a partir de la simulación computacional, por esta razón se optó por el software *Fathom*.

La elección de *Fathom* está dada porque el *software*, además de haber sido diseñado para niveles desde secundaria hasta universidad; proporciona a los alumnos la posibilidad y la motivación para reunir, explorar y analizar datos a profundidad gráficamente, y comprender conceptos de matemáticas y estadística de una manera más sencilla, integrando instantáneamente las representaciones utilizadas.

3.1.3. Recolección de datos

Los métodos de recolección de datos que utilizamos en la investigación fueron dos: las *evidencias escritas* de los estudiantes (a partir de las pruebas y la experimentación) y *entrevistas semiestructuradas* de preguntas abiertas. Las entrevistas fueron realizadas con el fin de revalidar o comprender los razonamientos informales que presenten los estudiantes.

3.2. Fases de la investigación

La investigación contempla las siguientes tres fases claramente definidas: Diseño de Instrumentos, Aplicación de los Instrumentos y Análisis de los Resultados.

3.2.1. Fase I: Diseño de Instrumentos

3.2.1.1. Diseño de las Pruebas

Para dar respuesta a la pregunta planteada en la investigación, fue necesaria la creación de dos pruebas. Llamamos a estas pruebas la prueba inicial y la prueba final, sobre las cuales se realizó en contraste de los razonamientos inferenciales informales de los estudiantes. Las situaciones presentadas en las pruebas son creaciones propuestas por el autor de la tesis, teniendo en cuenta

las dificultades y errores que presentan los estudiantes cuando se enfrentan a las pruebas de hipótesis.

La prueba inicial se realizó sin ayuda computacional y antes de comenzar la experimentación con *Fathom*, con el fin de caracterizar ese primer razonamiento informal con el que llegaron los estudiantes.

La prueba final se realizó tres meses después de terminada la experimentación con *Fathom*, y al igual que la prueba inicial, los estudiantes no contaron con ayuda computacional. El objetivo de realizar la prueba final, pasado un tiempo de la experimentación con *Fathom*, era tener seguridad de que las respuestas de los estudiantes provenían del conocimiento y las intuiciones interiorizadas sobre las pruebas de hipótesis, y no de un conocimiento a corto plazo. El impedimento de la ayuda computacional, se hizo con el fin de tener las mismas condiciones de la prueba inicial.

Las pruebas Inicial y Final, se encuentran en la sección de Anexos, y sus análisis en el capítulo de Análisis de Resultados.

3.2.1.2. Diseño de Talleres

Para la experimentación con *Fathom* se realizaron cuatro talleres, que permitieron a los estudiantes trabajar la inferencia estadística relacionada con las pruebas de hipótesis, sin la

necesidad de recurrir a sus formalismos. Tres de los talleres fueron sobre distribuciones discretas y uno sobre distribuciones continuas.

El primer taller fue diseñado con una doble intención, la primera es que fuera un taller guía para los estudiantes sobre el manejo de *Fathom*, sus herramientas y la simulación; y la segunda intención fue enfrentar a los estudiantes con una situación que los obligara a realizar una inferencia a partir de una proporción muestral, y así tuvieran el primer contacto con los elementos que intervienen en una prueba de hipótesis: Hipótesis nula y alternativa, muestra, estimador muestral, distribuciones de probabilidad, gráficos de distribución, regiones de rechazo y p-valor. La situación que se escogió para el primer taller, comprende una distribución uniforme con dos eventos posibles A o B, y un contexto que no condiciona la inferencia de los estudiantes (García y Sánchez, 2014).

En el segundo taller, al igual que el primero, se creó una situación sobre distribución de variables discreta y con un contexto que no condicionará las respuestas (y por ende los razonamientos informales) de los estudiantes. Este taller también tiene la intención de enfrentar a los estudiantes con la inferencia a partir de un contraste de hipótesis; pero, a diferencia del primer taller, la situación planteada centra el trabajo en la identificación y manejo, a través de la simulación, de distribuciones de probabilidad no uniforme.

El tercer taller, se diseñó en la misma línea de los dos anteriores, con el fin de reforzar lo trabajado en el momento, este taller reúne tanto el trabajo con los elementos que intervienen en una prueba de hipótesis, como el manejo de distribuciones de probabilidad no uniforme. Además

de lo anterior, las preguntas del taller están enfocadas en trabajar la influencia del tamaño muestral en la región de rechazo/nivel de significancia.

Finalmente, para el cuarto taller se diseñó una situación para variable continua, la cual contó a diferencia de los tres talleres anteriores con una simulación preestablecida, debido a que la situación requería de una base de datos. El objetivo de este taller es que los estudiantes, extrapolen lo aprendido en pruebas de hipótesis y simulación de distribuciones de variable discreta a distribuciones de variable continua.

3.2.1.3. Componentes conceptuales

Según la taxonomía SOLO (Biggs y Collis, 1991) la clasificación de las respuestas de los estudiantes se debe al número de componentes conceptuales, propias de la pregunta, que presenten los estudiantes y a las relaciones que formen entre ellas. Como nuestro objetivo es caracterizar los razonamientos inferenciales informales de los estudiantes, es decir, los argumentos que crean a partir de su conocimiento informal para dar sus respuestas; es necesario tener claridad sobre las componentes conceptuales que se van a trabajar en cada una de las preguntas.

Por esta razón, además del diseño de los instrumentos, y con el fin de facilitar la caracterización de los razonamientos inferenciales informales de los estudiantes participantes de la experimentación, hemos decidido, realizar un análisis de componentes conceptuales a cada una de las preguntas de los instrumentos utilizados en la experimentación, de la siguiente manera:

Cada pregunta que se planteó en la investigación tiene vinculadas ciertas componentes conceptuales propias de las pruebas de hipótesis. Hemos definido para cada pregunta dos tipos de

componentes: componentes principales y componentes secundarias. Las primeras son las que están más directamente vinculadas a la pregunta y sobre las cuales centraremos el análisis; las segundas son componentes complementarias que también pueden formar parte de los argumentos de los estudiantes.

El siguiente ejemplo muestra cómo se realizó el análisis de componentes:

Pregunta: Si se realizaran 500 lanzamientos en vez de 200, ¿cuántas CARAS máximo aceptaría que sacara Ana para no afirmar que la moneda está arreglada? Nota. Responda con un número y justifíquelo.

❖ Componentes conceptuales

1. Principal: Nivel de significancia/Región de rechazo.
2. Principal: Influencia del tamaño muestral.
3. Secundario: Contraste de hipótesis.
4. Secundario: Realización del experimento vs Distribución de probabilidad del espacio muestral. (Experimento → Teoría)
5. Secundario: Variación del Estimador Proporcional.

Esta pregunta cuenta con dos componentes principales, relacionadas con el nivel de significancia y la influencia del tamaño muestral; y tres componentes secundarias relacionadas con el contraste de la hipótesis, la variación del estimador y la relación entre experimento teoría.

3.2.2. Fase II: Aplicación de Instrumentos

Para esta investigación, como se mencionó en el apartado de Participantes, ocho fueron los estudiantes de Licenciatura en Matemáticas que participaron de la experimentación; de los cuales cuatro fueron seleccionados para el análisis: Andrés, Sofía, Gabriela y Luis.

En primera instancia se aplicó a los ocho estudiantes la prueba inicial, los estudiantes no contaron con ayuda computacional y se les dio un tiempo máximo de 2 horas para que respondieran todas las preguntas, con sus respectiva justificación.

Para la experimentación se contó con aulas de cómputo especializadas, cada estudiante contó con su computador, con el *software Fathom* instalado, y los talleres. Las sesiones de trabajo tuvieron la siguiente metodología: en la primera parte de la sesión los estudiantes resolvían el taller sin ayuda computacional, y sin la intervención del profesor, esto se hizo con el fin de recopilar los razonamientos informales iniciales. En la segunda parte, el estudiante resolvía de nuevo el taller, pero esta vez con la ayuda de la simulación estadística de *Fathom*, y con la intervención del profesor para solucionar inquietudes sobre el manejo de software. En la tercera parte se hacía una corta socialización, en donde a través de la mediación del profesor, se generalizaban los resultados y los estudiantes sacaban sus propias conclusiones de lo trabajado en el taller.

Finalmente, pasados tres meses de la experimentación con *Fathom*, se realizó la prueba final, en la que solo participaron los cuatro estudiantes seleccionados para el análisis. La prueba Final, al igual que la inicial no contó con ayuda computacional, ni intervención del profesor; y contó con un tiempo de 2 horas para su solución.

En total se realizaron 6 sesiones de trabajo: 2 para la aplicación de las pruebas, y 4 sesiones para la aplicación de los talleres. En total, la experimentación tuvo una duración aproximada de 14 horas.

3.2.3. Fase III: Análisis de Resultados

Esta última Fase comprende el análisis de los resultados obtenidos a partir de la recolección de datos de la fase anterior. El análisis se realizó sobre las respuestas escritas por los estudiantes, y en algunos casos se tuvo en cuenta, algunas entrevistas hechas a los estudiantes con el objetivo de dar claridad a lo que querían expresar con sus respuesta, y a los argumentos de los razonamientos informales utilizados.

Las respuesta de los estudiantes fueron analizadas, clasificadas y caracterizadas, en los niveles de la taxonomía SOLO; teniendo en cuenta el análisis de componentes conceptuales evaluadas en cada pregunta. El análisis se realizó a cada uno de los instrumentos aplicados en la experimentación, haciendo al final de cada uno de ellos, una conclusión de los avances que iban teniendo los estudiantes, los cambios en sus razonamientos informales, y las dificultades que van presentando.

Para finalizar y dar respuesta a la pregunta de investigación, se realizó un contraste, con la taxonomía SOLO, entre la prueba Inicial y la prueba Final de los razonamientos informales que utilizaron los estudiantes en cada una de las componentes conceptuales de la pruebas de hipótesis.

4.- Análisis de Resultados

Este capítulo comprende el análisis de los resultados encontrados en cada uno de los 6 instrumentos aplicados en la experimentación.

Para las pruebas Inicial y Final se realizaron tres apartados, en el primero se analizan las pruebas y se establecen las componentes conceptuales de las pruebas de hipótesis que se tendrán en cuenta en cada pregunta; en el segundo apartado se analizan las respuestas de los estudiantes y sus argumentos informales, con el fin de clasificarlos y caracterizarlos, en los niveles de la taxonomía SOLO; y en el tercer apartado se describen las conclusiones generales del análisis.

Para los talleres, se realizaron cuatro apartados; en el primero se establecen las componentes conceptuales de las pruebas de hipótesis que se tendrán en cuenta en cada pregunta; en el segundo se expone brevemente la simulación que se espera realicen los estudiantes; y en el tercer apartado se realiza la caracterización y clasificación de los razonamientos informales de cada uno de los estudiantes en los niveles de la taxonomía SOLO; y finalmente en el cuarto apartado se describen las conclusiones generales del análisis.

Para finalizar el capítulo se realiza el contraste entre la prueba Inicial y la Prueba Final, en donde se podrán observar el desarrollo que tuvieron los estudiantes en su razonamiento inferencial informal.

4.1. Prueba Inicial

4.1.1. Análisis de Componentes

SITUACIÓN 1 (Distribución Continua)

Se sabe que las notas obtenidas por un grupo de 1000 estudiantes de cálculo I se distribuyen tal como se muestra en la siguiente gráfica. El promedio de las notas es 3,7.

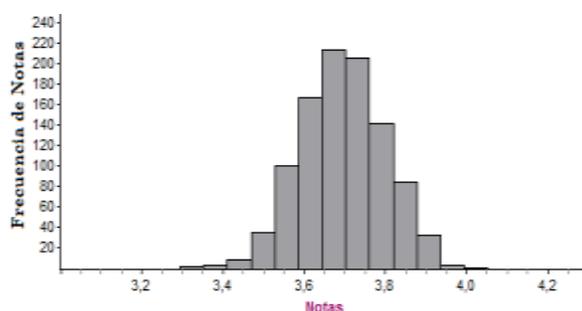


Figura 4. Distribución Prueba Inicial-1

Pregunta 1

Se extrae una muestra de 20 estudiantes de ese grupo y se calcula el promedio de sus notas. ¿Cuál cree que será el valor de ese promedio? Explique claramente su respuesta.

❖ Componentes conceptuales

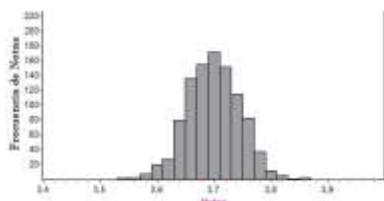
1. Principal: La Muestra como representante de la población. (Muestra → Población).
2. Principal: Carácter aleatorio de una muestra y del estimador muestral.
3. Principal: Interpretación gráfica de una distribución poblacional.

- Principal: Relación entre la distribución poblacional dada en forma gráfica y la distribución muestral de la media.

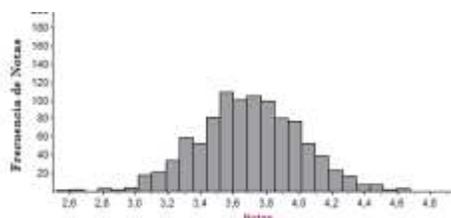
Pregunta 2

Responda de nuevo la pregunta 1 para cada uno de los siguientes gráficos de distribución de notas, teniendo en cuenta que el promedio de las mismas sigue siendo 3,7.

a.



b.



❖ Con *Figura 5. Distribución Prueba Inicial-2*

Figura 6. Distribución Prueba Inicial-3

- Principal: Relevancia de la desviación sobre la variabilidad del valor muestral de la media.
- Secundario: Muestra como representante de la población. (Muestra → Población).
- Secundario: Carácter aleatorio de una muestra y del estimador muestral.
- Secundario: Interpretación gráfica de una distribución poblacional.

Pregunta 3

Suponga ahora que no se conocen las notas de los 1000 estudiantes y se quiere conocer su promedio sin tener que revisar las notas de todos. Para ello, se decide seleccionar aleatoriamente a 20 estudiantes y mirar sus notas. El promedio de las notas de estos 20 estudiantes resulta ser 3,9.

Con base en la muestra, ¿se podría pensar que el promedio de todos los estudiantes es superior 3,7? Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Razonamiento inferencial (Muestra → Población).
2. Principal: Contraste de hipótesis.
3. Principal: Muestra como representante de la población. (Muestra → Población).
4. Secundario: Carácter aleatorio de una muestra.

Pregunta 4

Si en lugar de 20 estudiantes, la muestra se compone de 40 estudiantes y el promedio de las notas de los estudiantes de la muestra es 3,9.

¿Se podría pensar que el promedio de todos los 1000 estudiantes es superior a 3,7?

Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Razonamiento inferencial (Muestra → Población).
2. Principal: Contraste de hipótesis.
3. Principal: Influencia del tamaño muestral en la inferencia.
4. Principal: Muestra como representante de la población. (Muestra → Población)
5. Secundario: Carácter aleatorio de una muestra.

SITUACIÓN 2 (Distribución Discreta)

Ana y Pablo están jugando a lanzar la moneda, Ana escogió CARA y Pablo se quedó con el SELLO. Jugaron a 200 lanzamientos y ganó Ana obteniendo 170 CARAS.

Pablo no está de acuerdo con el juego, él dice: ¡Esa moneda esta arreglada, es imposible que salgan tantas CARAS y tan pocos SELLOS!

Pregunta 1

¿Qué piensa de la afirmación de Pablo, cree usted que tiene razón? Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Razonamiento inferencial.
2. Principal: Contraste de hipótesis.
3. Principal: Relación entre la distribución poblacional y los resultados experimentales.
4. Principal: Variación del Estimador Proporcional.

Pregunta 2

Pablo le propone a Ana realizar un nuevo juego con otra moneda. Esta vez Pablo gana con 110 sellos. Ahora es Ana quien protesta: ¡Esa moneda está arreglada!, ¿Tiene Ana razón? Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Razonamiento inferencial.
2. Principal: Contraste de hipótesis.
3. Principal: Relación entre la distribución poblacional y los resultados experimentales.
4. Principal: Variación del Estimador Proporcional.

Pregunta 3

Si se realizaran 500 lanzamientos en vez de 200, ¿cuántas CARAS máximo aceptaría que sacara Ana para no afirmar que la moneda está arreglada? Nota. Responda con un número y justifíquelo.

❖ Componentes conceptuales

1. Principal: Nivel de significancia/Región de rechazo.
2. Principal: Influencia del tamaño muestral.
3. Secundario: Contraste de hipótesis.
4. Secundario: Realización del experimento vs Distribución de probabilidad del espacio muestral. (Experimento → Teoría)
5. Secundario: Variación del Estimador Proporcional.

4.1.2. Análisis por estudiante

ANDRÉS

SITUACIÓN 1

Pregunta 1

3,7; tenemos que 3,7 es la nota promedio del grupo por ende en cualquier grupo de personas que se derive del primero la nota promedio estará cercana a esa cantidad.

Andrés infiere que la nota promedio es 3,7 o valores cercanos a esa cantidad, evidenciando tener intuición de la variabilidad muestral alrededor del promedio muestral. Además, Andrés muestra explícitamente en la parte final de su respuesta, la relación entre el promedio poblacional y el promedio muestral, no obstante da muestra de poseer el sesgo de los pequeños números (Kahneman y Tversky, 1983) porque en su argumento no da lugar a la existencia de muestras con

valores alejados a 3,7. La respuesta de Andrés es de nivel Uniestructural, sus argumentos no son lo suficientemente claros.

Pregunta 2

- a. *3,7 mismas razones que antes.*
- b. *3,6 la extensión de las notas es mayor (hago referencia a que va de 2,6 a 4,1) y tiene un crecimiento constante lo que me hace “deducir” que sería el promedio de las notas más altas y las notas más bajas.*

Andrés no consideró que la desviación en la gráfica (a) es menor que la desviación de la gráfica inicial y responde que es lo mismo que antes. Andrés, en resumen, no considera que la disminución de la variabilidad poblacional pueda afectar los resultados muestrales. Sin embargo, en la gráfica (b), Andrés sí consideró la mayor desviación en la distribución de las notas: *“la extensión de las notas es mayor”*, sugiriendo que el promedio puede ser 3,6. Andrés evidencia una forma de calcular los posibles valores muestrales: *sería el promedio de las notas más altas y las notas más bajas*, lo que conduce a un mismo valor para el promedio de cualquier muestra. Andrés da muestra de una incoherencia en su razonamiento aleatorio: existe la variación pero no existe la variación, la respuesta depende de lo que se le pregunte. Este tipo de argumentación contradictoria refleja la falta de ideas claras acerca de los experimentos aleatorios. Las respuestas de Andrés son de nivel Preestructural.

Pregunta 3

No, el subgrupo puede tener un promedio tanto mayor como menor y será cada vez más exacto dependiendo de qué tan grande sea esta (subgrupo).

Andrés no entendió ni la información ni la pregunta. Sin embargo, transluce la idea de que entre mayor sea la muestra mayor será la exactitud de la estimación. Respuesta de nivel Preestructural.

Pregunta 4

No, la misma razón del anterior aunque en una probabilidad mayor.

Nuevamente la respuesta de Andrés desconcierta: no puede ser mayor que 3,7 dice inicialmente, luego añade: “*aunque en una probabilidad mayor*”. Esta respuesta es Preestructural, pues va en contravía del efecto del tamaño muestral; pues según su argumento ahora es menos probable que el promedio sea superior a 3,9.

Será que lo que quiso decir es que ahora es menos probable que antes, esto es, ahora menos se puede creer que el promedio sea superior a 3,9, conclusión que va en contravía del efecto del tamaño muestral. Completamente Preestructural.

SITUACIÓN 2

Pregunta 1

La probabilidad de cada lanzamiento es de $\frac{1}{2}$ lo que no implica que por cada lanzamiento caiga el contrario al anterior.

Andrés destaca el hecho de que la distribución muestral no se corresponde exactamente con los resultados experimentales, pero hasta ahí, no responde la pregunta formulada. No obstante,

aunque relaciona lo teórico con lo empírico no se compromete con las diferencias que podrían darse entre lo teórico y lo experimental. Su argumento es de nivel Uniestructural.

Pregunta 2

No la misma respuesta que el anterior.

Andrés responde la pregunta correctamente, pero se basa en el argumento de la respuesta anterior; el cual es limitado para validar la repuesta. Por lo tanto la respuesta de Andrés es de nivel Uniestructural.

Pregunta 3

500 veces, misma respuesta que la anterior.

Andrés acepta la variabilidad del estimador, sin embargo, posee el sesgo de equiprobabilidad: todos los sucesos son igualmente probables. Esta respuesta de Andrés no cumple con ninguna de las componentes de la pregunta por lo cual está en nivel Preestructural.

Categorización de las respuestas de Andrés según SOLO

Tabla 2. Niveles SOLO para Andrés en la prueba Inicial

Andrés	Situación 1				Situación 2		
	1	2	3	4	1	2	3
		a	b				
No Respuesta							
Preestructural							

Uniestructural		
Multiestructural		
Relacional		

SOFÍA

SITUACIÓN 1

Pregunta 1

El valor de este grupo de 20 estudiantes estaría entre 3,7 y 3,8 pues casi la mitad del grupo de los 1000 están en 3,7, por lo tanto al tomar 20 estudiantes al azar varios estudiantes corresponderán a este promedio.

Sofía tiene en cuenta la distribución poblacional para asumir un intervalo donde posiblemente se encuentre el valor de la media muestral. Al decir “*varios estudiantes corresponderán a este promedio*” muestra que su argumentación está relacionada con el valor más probable de obtenerse que no es otro que el valor modal. Sin embargo en alguna forma reconoce que la muestra seleccionada puede tener otros valores que pueden afectar el promedio: *estaría entre 3,7 y 3,8*. La respuesta de Sofía es de nivel Uniestructural, pues aunque muestra intuición de los componentes de la pregunta, no evidencia una buena comprensión de ellos.

Pregunta 2

a. Para mi sacamos 20 estudiantes, su promedio seria también entre 3,7 y 3,8 pues los datos tomados entre la primera gráfica y la segunda son muy similares, por lo tanto sus resultados no varían tanto.

b. Si sacamos 20 estudiantes su promedio estaría entre 3,6 y 3,8 pues la gran mayoría de estudiantes está entre estos promedios, por lo tanto si sacamos esos 20, algunos de estos fijo serán 3,7.

Al igual que los demás estudiantes, Sofía tampoco percibió la disminución de la desviación de la notas en la gráfica (a), este fenómeno ya recurrente puede deberse a que la gráfica inicial y la gráfica (a) tienen la misma forma de campana que no alerta ningún cambio si no se examinan sus ejes; mientras que en la gráfica (b) su forma si cambia considerablemente haciendo más evidente el aumento de la desviación de la notas. Al no percibir el cambio Sofía maneja la misma respuesta y el mismo argumento de la pregunta anterior para la gráfica (a). Respuesta Uniestructural. No, capta el efecto de la variación de la distribución poblacional en la estimación aumentando el intervalo de estimación: *estaría entre 3,6 y 3,8.*

Pregunta 3

Si, pues el promedio de 20 entre los 1000 está en 3,9 por lo tanto muchos tienen notas altas y promedios superiores al 3,7.

Sofía infiere que la nota de los 1000 estudiantes sí es superior a 3,7, porque si el promedio de la muestra es 3.9 se puede deducir que existen muchos valores superiores a 3,7. Esta respuesta, evidencia la creencia en la representatividad que a priori tiene una muestra respecto a la población de la cual fue seleccionada. Es la misma idea que mostró en las primeras preguntas donde asume que la muestra representa bastante bien a su población, sin apreciar el efecto que pueda tener la selección específica de la muestra.

Al preguntarle a Sofía sobre el significado de promedio, respondió: *“es como el dato que más se repite”*; ¿entonces el promedio y la moda son lo mismo? Se le preguntó, para lo que ella respondió: *“yo creo que sí”*. Así pues, muestra una incomprensión del significado de promedio y lo confunde con la moda, tal vez dada la influencia del primer gráfico que muestra una distribución simétrica en la cual el promedio es prácticamente el mismo valor modal.

Esta respuesta está en un nivel Preestructural: desarrolla todas las componentes de la pregunta pero sobre una base equivocada del promedio. Sin embargo, da muestras de poseer algunos elementos de aleatoriedad, de representatividad y variabilidad muestral, aunque muy en ciernes.

Pregunta 4

Si, pues si ahora tenemos una muestra de 40 estudiantes y su promedio es de 3,9, significa que hay muchos estudiantes con notas entre 3,7 y 3,9.

Esta respuesta está en un nivel Preestructural: Sofía no considera para nada el aumento del tamaño muestral, su respuesta es idéntica a la anterior. Su idea de representatividad muestral es independiente al tamaño muestral.

SITUACIÓN 2

Pregunta 1

Pues mirando desde Pablo se puede pensar eso, pero pues es una probabilidad, no es nada seguro de que deban salir equitativamente lanzamientos, así que se vale que tenga Ana esos resultados.

Sofía tiene conciencia de la variabilidad de los resultados experimentales cuando afirma: *“no es nada seguro que deban salir equitativamente lanzamientos”*. El problema es que asume que el suceso presentado también se puede dar, evidencia el sesgo de equiprobabilidad mencionado previamente. Sin duda, el resultado es posible, el asunto está en que su probabilidad es muy pequeña. Esta respuesta es de nivel Preestructural, los argumentos no son suficientes para validar la respuesta.

Pregunta 2

Falso, pues las posibilidades de caer cara o sello son iguales para los dos y pues puede caer tantas veces la misma parte de la moneda como quiera, eso no depende de si la moneda esta arreglada o no.

Esta respuesta tiene el mismo argumento de la anterior, por lo cual, es de nivel Preestructural.

Pregunta 3

Unas 300 sería bien, pero si saca más se vería otra vez sospechoso.

Sofía, extrañamente se contradice con sus respuestas a las dos primeras preguntas, ahora dice que obtener *unas 300 sería bien*. Esto quiere decir que Sofía no está tan segura de sus argumentos anteriores. *“pero si saca más se vería otra vez sospechoso”*. Tal vez en este caso, el gran aumento del tamaño muestral le hizo dudar de sus ideas previas, generándole alguna intuición de la región de rechazo. La respuesta de Sofía es de nivel Preestructural, porque aunque da un valor, no da ningún argumento de su elección.

Categorización de las respuestas de Sofía según SOLO

Tabla 3. Niveles SOLO para Sofía en la prueba Inicial

Sofía	Situación 1				Situación 2			
	1	2 a	2 b	3	4	1	2	3
No Respuesta								
Preestructural								
Uniestructural								
Multiestructural								
Relacional								

GABRIELA

SITUACIÓN 1

Pregunta 1

3,7, ya que el promedio de todas las notas es esto, se podría decir que no interesa la cantidad de estudiantes este valor o número se mantendrá, y que la nota que más cantidad de estudiantes sacaron fue esta.

Gabriela evidencia el uso determinista de la heurística de representatividad: la muestra debe tener las mismas propiedades que la población, incluso afirma que sin importar “la cantidad de estudiantes”, Gabriela añade que 3,7 es el valor que más se repite, tal vez para reforzar su respuesta. Al parecer Gabriela establece una relación entre promedio y moda. La respuesta de Gabriela es de nivel Preestructural, pues no hay evidencia de que Gabriela perciba la variabilidad de la muestra.

Pregunta 2

- a. *Hay mayor probabilidad de que la nota promedio esté en el intervalo 3,66 y 3,72, ya que la mayoría de estudiantes están en este rango.*
- b. *Está en un intervalo entre 3,5 y 4,0, ya que en este caso la cantidad de estudiantes en cada nota están juntos, sus valores en cantidad son cercanos.*

Gabriela es la primera estudiante que presenta evidencia del cambio de desviación de las notas en la gráfica (a), argumentando que el promedio está en el intervalo (3,66-3,72), evidenciando de paso una intuición de la variabilidad del estimador muestral que no presentó en la primera pregunta y una intuición clara del efecto de la variabilidad poblacional sobre la distribución del estimador que la conduce a realizar una estimación por intervalo, incluso con elementos intuitivos sobre el nivel de confianza cuando afirma: *Hay mayor probabilidad*. Al preguntar a Gabriela por su contradicción con su respuesta a la primera pregunta, ella respondió “*no entendí la pregunta cómo era, me equivoque*” su corrección a la primera pregunta fue “*Estaría en un intervalo alrededor de 3,7*”. La respuesta de Gabriela está en un nivel Multiestructural, pues aunque maneja argumentos ingenuos, identifica el componente principal de la pregunta.

Pregunta 3

Si, ya que esto significa que no interesa el espacio muestral, porque la mayoría de los estudiantes tuvieron una nota superior a 3,7.

Aunque Gabriela tiene la misma respuesta que Sofía, su argumento es netamente aritmético determinístico; es decir, como ya una muestra me dio mayor a 3,7 entonces ya puedo asegurar que el promedio de los 1000 estudiantes también lo es. Esta respuesta está en un nivel Preestructural.

Pregunta 4

Si, ya que esto significa que por lo menos entre 40 estudiantes su promedio es mayor a 3,9.

La respuesta de Gabriela es Uniestructural, pues su inferencia sigue siendo la misma y su argumento resalta el hecho que la muestra es mayor.

SITUACIÓN 2

Pregunta 1

No tiene razón, ya que la probabilidad de que Ana saque cara y Pablo sello es la misma: $\frac{1}{2}$, por lo tanto puede ser posible, se le podría llamar suerte.

Gabriela explicita la función de probabilidad del experimento de lanzar la moneda, no justifica su inferencia: “*por lo tanto puede ser posible*”. Finalmente asume que el asunto de obtener uno u otro resultado, es cuestión de suerte.

Esta respuesta es de nivel Preestructural.

Pregunta 2

No tiene razón, todo es probabilidad en cada lanzamiento opino de igual manera que en el punto anterior.

Aunque en este caso la inferencia como tal está bien “*No tiene razón*”, los argumentos para sustentarla no son los correctos. Por lo cual esta respuesta está en un nivel Preestructural.

Pregunta 3

$$170 \rightarrow 200$$

$$90 \rightarrow 200$$

$$x \rightarrow 500$$

$$x \rightarrow 500$$

$$x = 425 \text{ de } 500$$

$$x = 225 \text{ de } 500$$

344, más o menos deben estar cerca de 250 caras y 250 sellos, una diferencia de 50 es muy pequeña, debe ser casi 100 pero no 100 más o menos 94.

Gabriela se arriesga a dar un valor de 344, considerando que la diferencia con 250 debe ser mayor que 50 y cercana a 100. Al entrevistar a Gabriela sobre el porqué de ese argumento responde repitiendo lo mismo que había escrito en la hoja. Este P- valor de Gabriela es muy superior al que se podría calcular formalmente (la probabilidad con 278 caras ya es menor de 0.01). Aunque en las respuestas anteriores, Gabriela apoyó su argumento en la suerte, para esta pregunta evidencia, involuntariamente, que en realidad no acepta todos los posibles resultados de la moneda, es decir, no solo es suerte.

La respuesta de Gabriela es, por consiguiente, de nivel Preestructural.

Categorización de las respuestas de Gabriela según SOLO

Tabla 4. Niveles SOLO para Gabriela en la prueba Inicial

Gabriela	Situación 1				Situación 2			
	1	2 a	2 b	3	4	1	2	3
No Respuesta								
Preestructural	■		■		■			■
Uniestructural					■			
Multiestructural		■	■					
Relacional								

LUIS

SITUACIÓN 1

Pregunta 1

Nuevamente el promedio es 3,7 ya que se obtendría de sumar las notas y dividirlos por la cantidad de estudiantes.

Luis presentó un argumento basado en la representatividad determinística, escribiendo en forma textual cómo se calcula un promedio, sin dar ninguna explicación del por qué, para qué o cómo usarlo en la situación. La respuesta de Luis es de nivel Preestructural.

Pregunta 2

a. Exactamente como en la primera situación el promedio sería 3,7, ya que no se ve afectado por cómo se adquiere las notas, simplemente es una cantidad igual.

Ejemplo: tengo 4 notas, (1, 2, 3, 4) y su promedio es 2,5, al escoger 2 grupos, diferentes de (1, 4) y (3, 2), porque sería claro que el promedio es 2,5 en ambos casos, tales como (1, 3) y (2, 4) sus promedios son 2 y 3 respectivamente, pero al sumarlos y obtener un promedio nuevo, este sería 2,5 igual al promedio de todas las notas.

b. Misma que la anterior.

Luis utiliza un ejemplo particular, el cual resulta ser contradictorio a sus argumentos. Luis afirma que no importa cómo se tome la muestra que el promedio siempre seguirá siendo el mismo, pero en su ejemplo utilizó las muestras (1, 3) y (2, 4) las cuales tienen promedios diferentes. Además, Luis no muestra ninguna evidencia ni implícita, ni explícita de haber percibido el cambio en la desviación de las notas en ninguna de las dos gráficas; por consiguiente sus respuestas son de nivel Preestructural.

Pregunta 3

No, ya que al seleccionar los promedios aleatoriamente puede pasar lo planteado en mi ejemplo anterior. Ejemplo: 4 notas (1, 2, 3, 4) Promedio 2,5.

Si tomo 2 grupos (2, 3) y (4, 1) el promedio es 2,5 en ambos casos, pero para los grupos (2, 4) y (3, 1) el promedio del primer grupo sería 3, (más alto que el promedio del grupo) y el segundo es 2. Pero el promedio de la suma de esos dos promedios es 2.5 que sería igual al promedio del grupo inicial; simulé que en este caso el 3.9 está representado por el promedio del grupo (2, 4), que es 3, pero el promedio de todas las notas tomando en cuenta todos los grupos vuelve a ser el inicial. "2,5" que sería el 3,7.

Luis afirma que no cree que el promedio sea superior a 3,7 y utiliza el ejemplo de la pregunta anterior para probar su punto. El uso del ejemplo presenta un problema al poner en correspondencia dos distribuciones distintas. La respuesta de Luis es de nivel Preestructural.

Pregunta 4

El promedio es el mismo.

Para Luis el tamaño de la muestra es indiferente debido a la utilización inadecuada del ejemplo (notas 1, 2, 3 y 4), y lo revalida al no evidenciar cambio cuando se duplica la muestra. Las respuestas de Luis son de nivel Preestructural.

SITUACIÓN 2

Pregunta 1

La probabilidad 50/50 pero la cantidad de intentos serían infinitos por lo tanto es difícil concluir a qué cantidad de intentos caerían las mismas caras que sellos.

Luis tiene la intuición correcta de que la probabilidad se cumple en el infinito, el problema es que esta noción le impide inferir qué pasa en un experimento con lanzamientos finitos. Por lo cual, Luis no pudo responder la pregunta y su argumento “es difícil concluir a qué cantidad de intentos caerían las mismas caras que sellos” pareciera dar a pensar que él aceptaría cualquier evento posible. La respuesta de Luis es Uniestructural, pues no aborda el tratamiento de la hipótesis, pero sí tiene intuición de la interpretación frecuencial de la probabilidad.

Pregunta 2

Uniestructural	
Multiestructural	
Relacional	

4.1.3. Conclusiones

En general las respuestas de los estudiantes se movieron entre los niveles preestructural y uniestructural de la taxonomía SOLO; muy pocos mostraron argumentos que les permitieran clasificar sus respuestas en un nivel multiestructural; y ninguno logró llegar al nivel relacional.

Como era de esperarse, los estudiantes presentaron en sus razonamientos muchos sesgos y dificultades descritos en la literatura. El sesgo de los pequeños números, entendido como la tendencia de la gente a creer que la información obtenida en una pequeña muestra, será representativa de la población total; o el sesgo de equiprobabilidad, donde se consideran todos los sucesos son igualmente probables. Entra las dificultades más notorias, están las contradicciones entre los argumentos de los estudiantes, evidencia de unas malas bases estadísticas.

Además, los estudiantes no considera la variabilidad muestral, ni poblacional en la distribuciones de variable continua, no obstante, parecen si aceptan el carácter variacional de una proporción muestral. Por ejemplo, Sofía tiene conciencia de la variabilidad de los resultados experimentales cuando afirma: *“no es nada seguro que deban salir equitativamente lanzamientos”*

Finalmente, se puede destacar que algunos estudiantes evidencian ideas intuitivas de región de rechazo/aceptación.

4.2. Taller 1

4.2.1. Análisis de Componentes

SITUACIÓN

Un estudio pretende averiguar si los habitantes de una ciudad con aproximadamente 20000 habitantes, tienen alguna preferencia entre los equipos de fútbol A y B. Para tal fin, se les preguntó a 300 personas seleccionadas aleatoriamente, cuál de los dos equipos era su favorito. Después de hacer el respectivo conteo, los resultados obtenidos arrojaron que de las 300 personas encuestadas 170 preferían al “Equipo A”.

Pregunta 1

¿Se puede afirmar que los habitantes de esa ciudad prefieren el equipo A? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis. (Si hay evidencia de preferencia)
2. Principal: Variabilidad Muestral. (Muestra → Población)
3. Secundario: Inferencia.

Pregunta 2

Si no fueron 170 sino 158 habitantes los que prefirieron el “Equipo A”. ¿Este resultado proporciona evidencia de una preferencia hacia el Equipo A? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis. (No hay evidencia de preferencia)

2. Principal: Variabilidad Muestral. (Muestra \rightarrow Población)
3. Secundario: Inferencia.

Pregunta 3

Y si en realidad fueron 180 habitantes los que prefirieron el equipo A ¿Este resultado proporciona evidencia de una preferencia hacia el Equipo A? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis. (Si hay evidencia de preferencia, superior a la pregunta 1)
2. Principal: Variabilidad Muestral. (Muestra \rightarrow Población).
3. Secundario: Inferencia.

Pregunta 4

¿Cuántos habitantes como mínimo deben preferir el “Equipo A” para que proporcionen evidencia de una preferencia? Marque y Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: P-valor informal
2. Principal: Región de rechazo

4.2.2. Experimentación *Fathom*

1. Inicie *Fathom* y cree una colección al hacer clic sostenido en el ícono *Collection*, y arrástrelo (haciendo clic sostenido) hacia la zona de trabajo.



Figura 7. Taller 1-Simulación esperada 1

2. Para darle nombre a la colección haga doble clic en donde aparece el nombre de la colección, en donde indica la flecha de la siguiente imagen.



Figura 8. Taller 1-Simulación esperada 2

Aparecerá un recuadro llamado *Rename Collection* en donde debe digitar el nombre que desea, para este caso será **Equipo Preferido**, luego haga clic en ok. El icono de la colección aparecerá con el nombre definido.

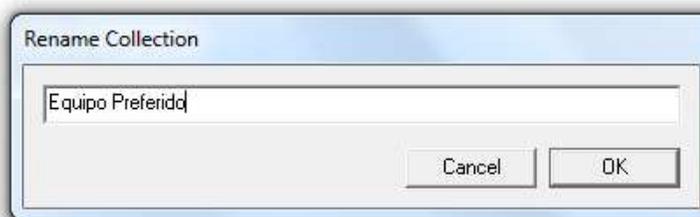


Figura 9. Taller 1-Simulación esperada 3

3. Teniendo seleccionada la colección (la colección queda seleccionada cuando le damos clic sobre ella) haga clic sostenido en el ícono **Table**, y arrástrela sobre el área de trabajo.

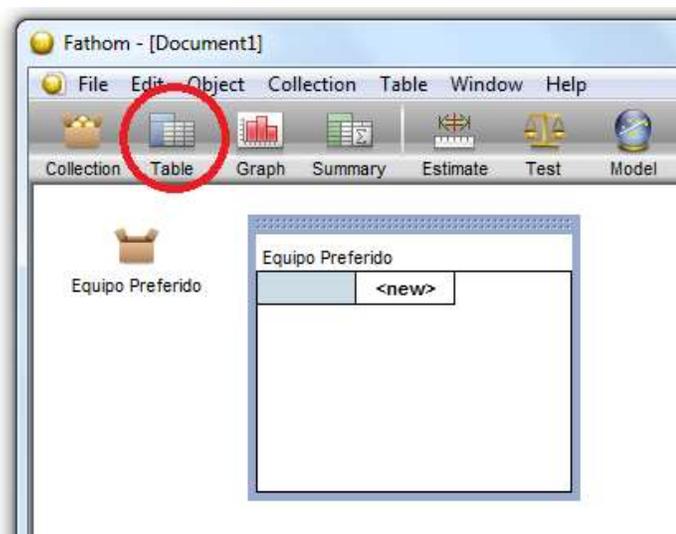


Figura 10. Taller 1-Simulación esperada 4

Las columnas de la tabla son llamadas **Attributes** (variables)

4. Para dar un nombre a la variable damos clic derecho sobre cualquier lugar de la tabla, se despliega una pestaña en cual daremos clic sobre **New Attribute...**

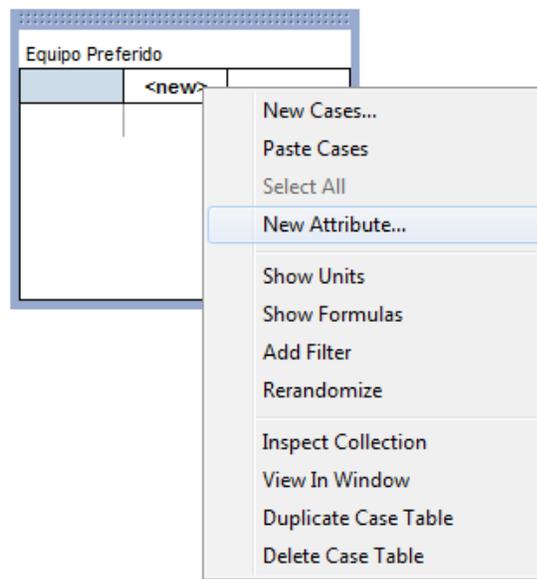


Figura 11. Taller 1-Simulación esperada 5

Aparecerá una ventana emergente donde nombraremos nuestra variable con el nombre de **Equipo** y damos OK.

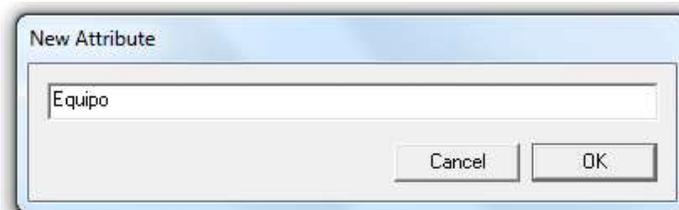


Figura 12. Taller 1-Simulación esperada 6

Ahora en la tabla aparecerá una columna con el nombre que indicamos.

Equipo Preferido

Equipo	<new>

Figura 13. Taller 1-Simulación esperada 7

5. En este punto es necesario definir un modelo de distribución para seleccionar entre el Equipo A y el Equipo B. La situación nos pregunta si existe una preferencia en la ciudad por el Equipo A, a partir de una muestra en la que 170 habitantes prefirieron dicho equipo. El razonamiento a seguir es: ¿Qué tan frecuente resulta ser una preferencia de 170 o más en una muestra de 300 habitantes?; es decir, si repetimos la misma experiencia con muestras de 300 habitantes de la ciudad, 10, 100, o 1000 veces ¿una preferencia de 170 o más es un resultado normal? o, al contrario, es poco inusual.

Ahora bien, para poder simular la experiencia repetidas veces y saber si existe alguna preferencia hacia el Equipo A debemos partir del supuesto de que no haya preferencia alguna por alguno de los equipos. Así pues, nuestro modelo de distribución para los datos será de 50-50, es decir, existen tantos habitantes seguidores de un equipo como del otro.

Para definir un modelo de distribución 50-50 sobre la variable **Equipo** damos clic derecho sobre **Equipo**, se desplegará una pestaña donde daremos clic sobre **Edit Formula**

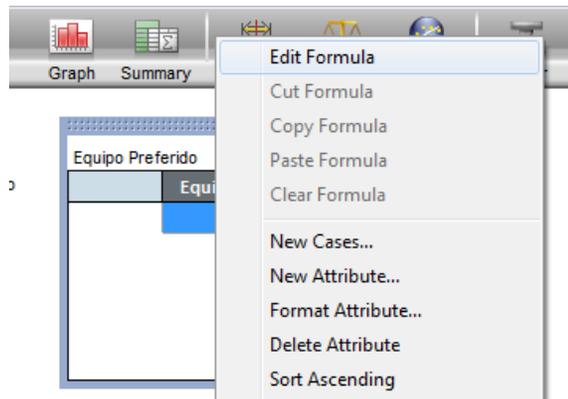


Figura 14. Taller 1-Simulación esperada 8

Aparecerá una ventana como la siguiente

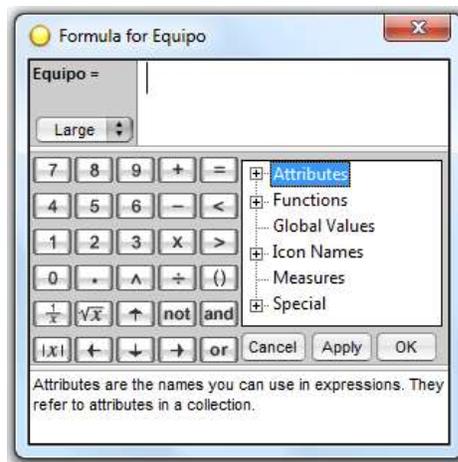


Figura 15. Taller 1-Simulación esperada 9

En la parte superior digitamos la manera como estará definido o como se obtendrán los valores para nuestra variable, a través de funciones o comandos específicos del Software. Para el modelo 50-50 desplegamos la opción de **Functions**, en el menú, dando *click* en el “+”, luego desplegamos **Random numbers** y entre ellos aparecerá el comando **randomPick**, damos doble clic en él. Si el comando se usa de manera correcta aparecerá

de color azul en la barra de entrada. Entre el paréntesis y comillas separadas por punto y coma se digitan las opciones entre las que se desea escoger, para nuestro caso “Equipo A” y “Equipo B”. Finalmente damos clic en OK.

RandomPick se traduce como “seleccione aleatoriamente” entre los valores que aparecen entre paréntesis. Como cada opción aparece una sola vez la selección se realiza con probabilidades iguales.

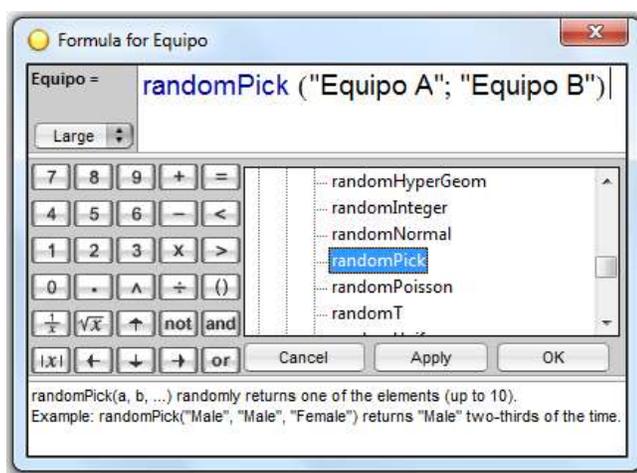


Figura 16. Taller 1-Simulación esperada 10

6. De nuevo en la ventana principal no ha aparecido nada en el tabla, estos se debe a que no tenemos casos a los cuales aplicar el paso anterior. Para crear los casos, en nuestra situación 300, damos clic derecho sobre la variable **Equipo**, y hacemos clic en *New Cases...*

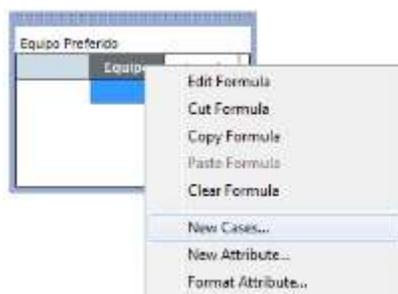


Figura 17. Taller 1-Simulación esperada 11

Aparecerá una ventana donde escribimos el número de casos que queremos que tenga nuestra tabla, 300 para nuestra situación. Clic en OK.

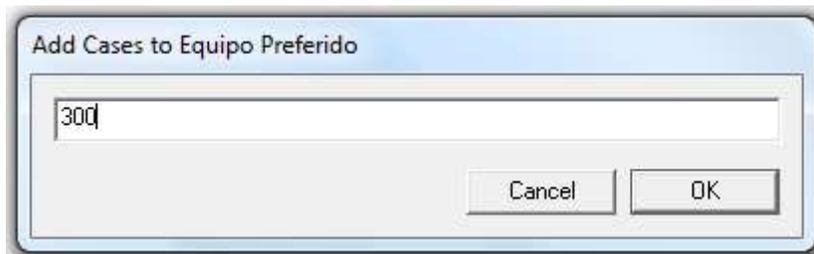
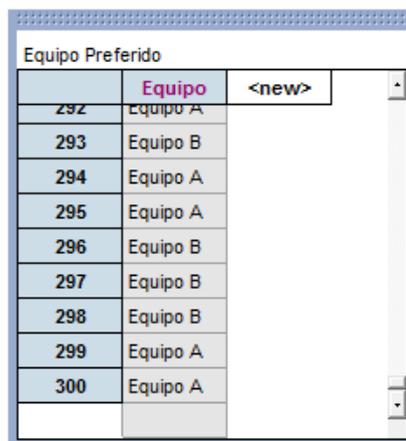


Figura 18. Taller 1-Simulación esperada 12

Ahora en la tabla ya aparecerán 300 casos en los que se eligió aleatoriamente entre los Equipos A y B, cada uno con igual probabilidad de ser elegido.



	Equipo	<new>
292	Equipo A	
293	Equipo B	
294	Equipo A	
295	Equipo A	
296	Equipo B	
297	Equipo B	
298	Equipo B	
299	Equipo A	
300	Equipo A	

Figura 19. Taller 1-Simulación esperada 13

Ya tenemos una muestra aleatoria de preferencia 50-50 (para ambos equipos) de tamaño 300. El siguiente paso es contar cuántas personas de esta muestra eligieron el Equipo A.

7. Para contar cuántas personas de la muestra de 300 prefirieron el Equipo A damos clic derecho sobre la colección, y hacemos clic en el *Inspect Collection* (Inspector de la colección).

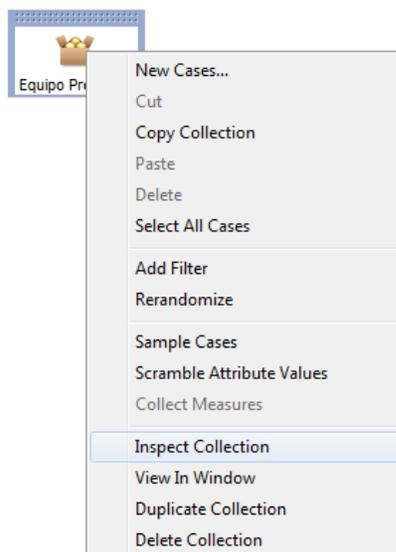


Figura 20. Taller 1-Simulación esperada 14

Aparecerá una nueva ventana y daremos clic en la pestaña *Measures* (Medidas), allí podremos definir una medida que sea contar la preferencia del Equipo A.

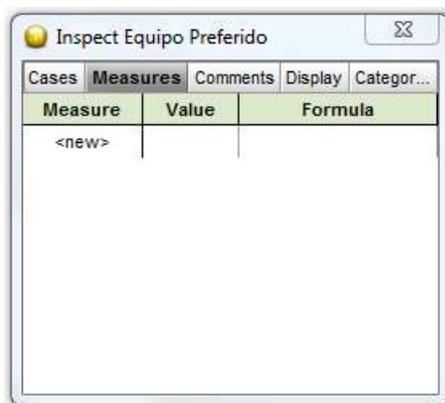


Figura 21. Taller 1-Simulación esperada 15

En la columna *Measure* se nombra la medida, en este caso pondremos, por ejemplo, **PEquipoA**, que hace referencia a preferencia por el Equipo A; en la segunda columna *Value* nos arrojará el valor de la medida, pero como no hemos definido cómo medirla aún no aparece ningún valor; finalmente en la tercera columna *Formula* definiremos la medida, damos doble clic sobre el espacio en blanco bajo *Formula*; nos desplegará una ventana para editar fórmulas como la que utilizamos para definir el modelo 50-50; en la barra de entrada escribiremos el siguiente comando *count* () (si está bien escrito tomará color azul), este comando nos permite contar los elementos que pongamos dentro del paréntesis. Como queremos contar cuántos hombres de la columna Equipo Prefieren el Equipo A entonces dentro del paréntesis escribiremos *count (Equipo = "Equipo A")*, finalmente damos clic en OK. La siguiente imagen muestra el comando.

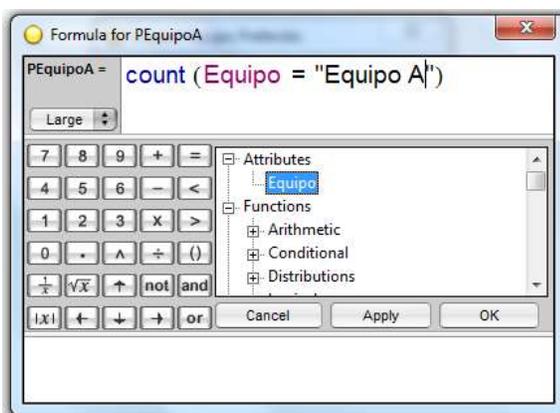
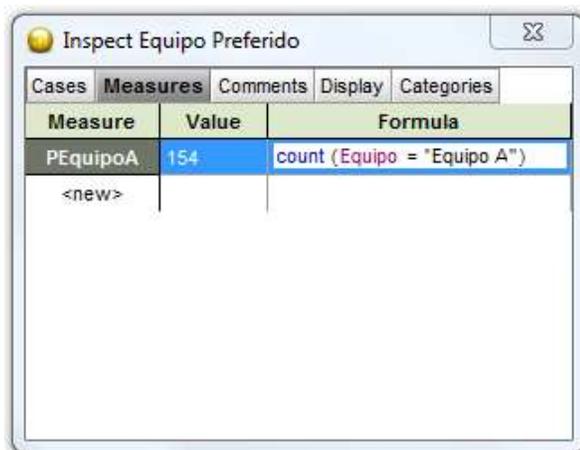


Figura 22. Taller 1-Simulación esperada 16

Regresamos a la ventana del inspector, la cual ya nos arroja cuántos de los 300 habiyantes de la muestra prefieren el Equipo A, en la siguiente imagen vemos que para esta muestra 154 personas prefirieron el Equipo A; es importante tener en cuenta que este valor depende

de la muestra y como las muestras son aleatorias lo más probable es que al replicar la situación se obtenga otro valor.



Measure	Value	Formula
PEquipoA	154	count (Equipo = "Equipo A")
<new>		

Figura 23. Taller 1-Simulación esperada 17

El *software Fathom* tiene la opción de hacer remuestreo, basta con dejar oprimida la tecla “ctrl” y pulsar la tecla “y”, cada vez que se pulse “y” se genera una nueva muestra de tamaño 300 y por consiguiente un nuevo **PEquipoA**. En este punto podemos utilizar el remuestreo, mirar cómo son las preferencias por el Equipo A en una distribución 50-50. Pero tendríamos que estar anotando en una libreta cada **PEquipoA** para cada muestra distinta. Pero existe una opción de *Fathom* que nos permite registrar las medidas **PEquipoA** de la cantidad de muestras que queramos.

8. Como no basta con tener una muestra y contar cuántos prefieren el Equipo A, vamos a hallar las preferencias del Equipo A para 1000 muestras de tamaño 300. Damos clic derecho en la colección y en la pestaña que se despliega damos clic en **Collect Measures** (Coleccionar Medidas).

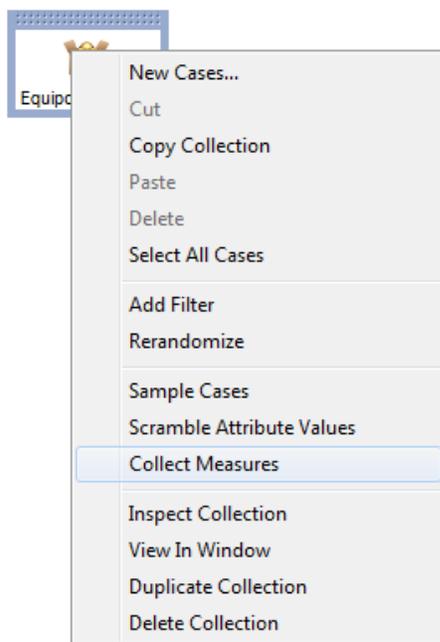


Figura 24. Taller 1-Simulación esperada 18

Automáticamente el software creará una nueva colección donde recogerá la medida de **PEquipoA** (preferencia por el Equipo A) para 5 muestras de tamaño 300.

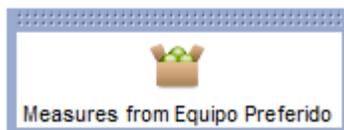


Figura 25. Taller 1-Simulación esperada 19

Para ver estos 5 valores de **PEquipoA**, seleccionamos la nueva colección **Measures from Equipo Preferido** y al igual como hicimos con la primera colección arrastramos un ícono de tabla. Aparecerá una tabla con los 5 valores de **PEquipoA** para 5 muestras distintas de tamaño 300.

Measures from Equipo Preferido	
	PEquipoA <new>
1	146
2	157
3	153
4	156
5	153

Figura 26. Taller 1-Simulación esperada 20

Como queremos 1000 valores de **PEquipoA** en vez de 5, vamos al **Inspect Collection** de la nueva colección. En esta ventana vamos a desmarcar la animación quitando la x de **Animation on** y cambiaremos donde dice measures 5 por 995 y así obtendremos las 1000 medidas de **PEquipoA** que queremos. Finalmente clic en **Collect More Measures**.

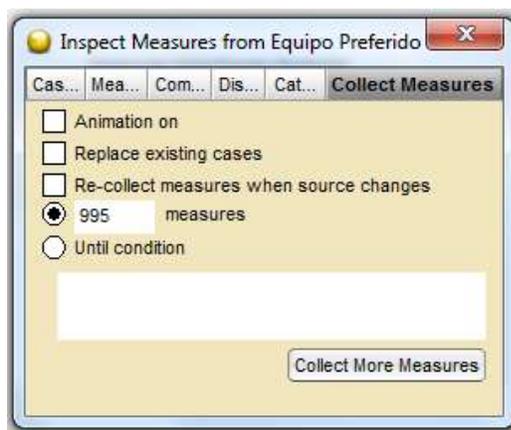


Figura 27. Taller 1-Simulación esperada 2

La tabla de la nueva colección tendrá ahora 1000 valores de **PEquipoA** de 1000 muestras aleatorias de tamaño 300.

	PEquipoA	<new>
994	164	
995	151	
996	138	
997	168	
998	142	
999	139	
1000	137	

Figura 28. Taller 1-Simulación esperada 22

9. Para ver de una mejor manera los valores de PEquipoA podemos utilizar una gráfica, la cual se puede hacer seleccionando la nueva colección y arrastrando de la barra de herramientas un ícono de **Graph** a la hoja de trabajo.



Measures from Equipo Preferido

	PEquipoA	<new>
994	164	
995	151	
996	138	
997	168	
998	142	
999	139	
1000	137	

no data Empty Plot

Drop an attribute here.

Figura 29. Taller 1-Simulación esperada 23

Arrastramos el nombre **PEquipoA** sobre el eje x de la tabla donde dice *Drop an attribute here*. Inmediatamente aparecerá una gráfica de puntos que nos permite ver la distribución que siguen los valores de **PEquipoA**.

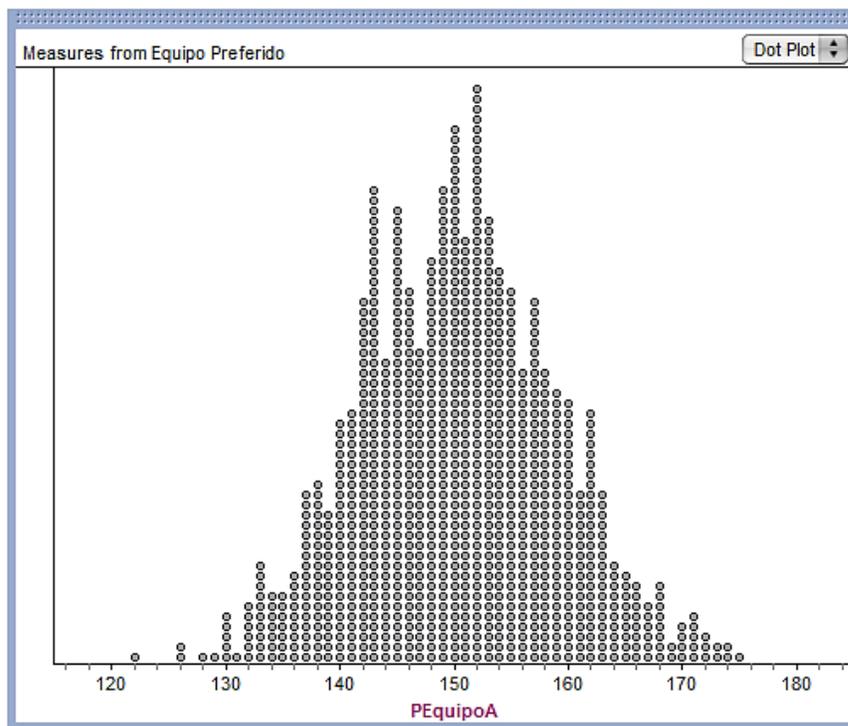


Figura 30. Taller 1-Simulación esperada 24

A partir de la gráfica ya podemos tener ideas de qué tan frecuente es que 170 personas o más prefieran el Equipo A en un muestra de 300. Pero para tener un valor exacto podemos calcular esta proporción.

10. Para calcular la proporción de valores mayores iguales que 170 en la simulación, hacemos clic derecho sobre la gráfica y abrimos el *Inspect Graph*.

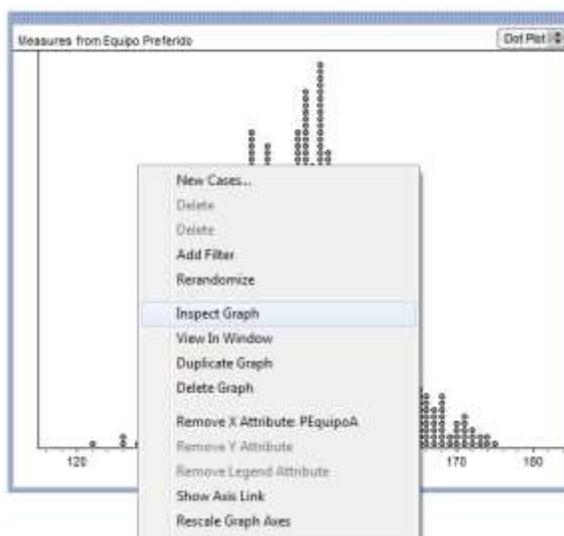


Figura 31. Taller 1-Simulación esperada 25

Por defecto aparecerá uno de los 1000 valor para **PEquipoA**, en este caso como se indica en la parte inferior el 158 es el valor 1000.

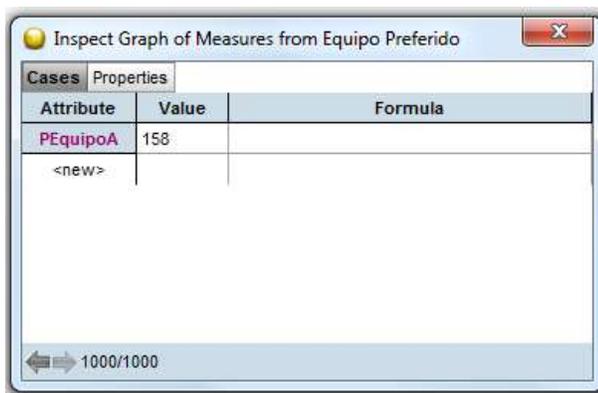


Figura 32. Taller 1-Simulación esperada 26

El *Inspect Graph* tiene el mismo funcionamiento que el *Inspect Collection*. Primero contemos cuantos valares son iguales o superiores a 170, para ello creamos un nuevo

atributo llamado **P170** y utilizamos en comando *count()* en la fórmula de la siguiente manera:

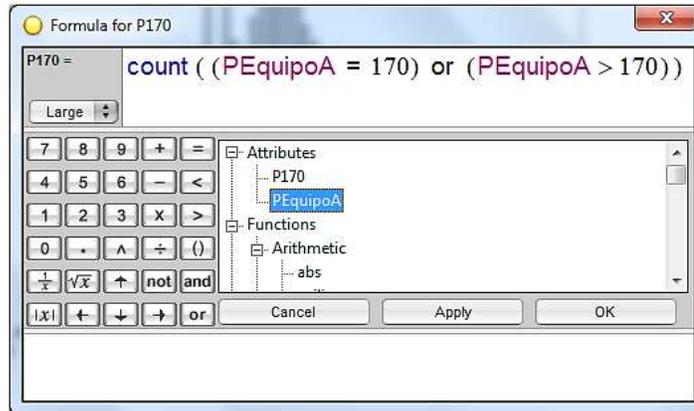


Figura 33. Taller 1-Simulación esperada 27

Ok para confirmar. Ahora en el inspector de la gráfica aparecerá cuantos valores son mayores o iguales a 170. En este caso 17 de los 1000 son mayores iguales a 170.

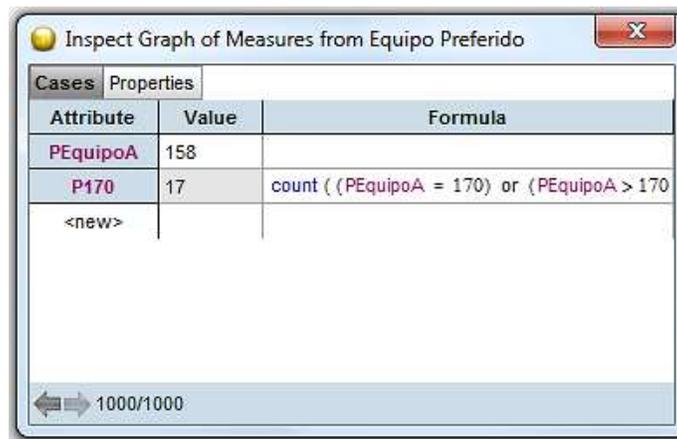


Figura 34. Taller 1-Simulación esperada 28

Para saber estos 17 valores qué proporción de los 1000 valores son, creamos un nuevo atributo llamado **Proporción** en el *Inspect Graph* y utilizamos en comando *Porportion* () de la siguiente manera:

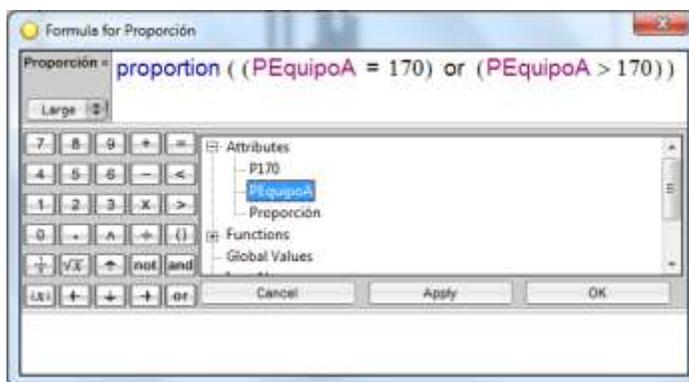


Figura 35. Taller 1-Simulación esperada 29

OK para confirmar. En el inspector de la gráfica aparecerá que proporción de los 1000 valores son iguales o superiores a 170. En este caso fue de 0,017

Attribute	Value	Formula
PEquipoA	158	
P170	17	count ((PEquipoA = 170) or (PEquipoA > 170))
Proporción	0,017	proportion ((PEquipoA = 170) or (PEquipoA > 170))
<new>		

Figura 36. Taller 1-Simulación esperada 30

4.2.3. Análisis por estudiante

ANDRÉS

Pregunta 1***Antes de Experimentación***

No, la muestra tomada es demasiado baja por lo tanto solo podemos afirmar que un sector de la ciudad prefiere el equipo A.

Después de Experimentación

Sí, hay una tendencia de que si son superiores a 170 decimos que si hay preferencia.

Hicimos la gráfica para una muestra donde no hubiese preferencia si se salía del centro de la gráfica significaba que había una preferencia ya sea por una o por la otra entonces con 170 está en el borde de la gráfica significa que si hay preferencia por el equipo A.

Preestructural → Multiestructural

Andrés pasa de una respuesta donde considera el tamaño de la muestra demasiado pequeño para hacer una inferencia, a una donde con ayuda del gráfico de *Fathom* pudo hacer una inferencia sobre la población. Andrés entiende la simulación cuando afirma “*la gráfica para una muestra donde no hubiese preferencia*” haciendo referencia al modelo de preferencia 50-50 que se planteó para contrastar la hipótesis de preferencia. Además, hace una asociación adecuada para rechazar la hipótesis nula (no hay preferencia) a partir del análisis gráfico de su distribución.

En la actividad con *Fathom* se realizó el cálculo del p-valor a partir de la simulación, el cual arrojó menos del 0.02 para valores superiores iguales a 170; Andrés no hizo uso de este valor para sus argumentos, Andrés se contenta con el análisis gráfico, no lo asocia con una probabilidad.

Pregunta 2*Antes de Experimentación*

No, igualmente el porcentaje de la población que se toma es demasiado bajo para determinar.

Después de Experimentación

No, ahí los valores son demasiado cercanos a la mitad teniendo en cuenta la gráfica para una muestra donde no hay preferencia, lo que nos dice que no hay preferencia.

Preestructural → Multiestructural

Al igual que la pregunta anterior, inicialmente se mantiene el argumento de la pequeñez del tamaño de la muestra, impidiendo que Andrés realice el contraste de la hipótesis. Ya después de la experimentación, Andrés vuelve a utilizar el gráfico de la distribución de no preferencia (modelo 50-50) para hacer el contraste de la hipótesis. Esta vez, como el valor de 158 se encuentra muy cercano a la mitad del gráfico, se garantiza para hacer su inferencia y no negar la hipótesis.

A pesar que Andrés ya conocía el proceso para calcular la probabilidad para cualquier valor, no lo realizó para 158, reafirmando que para él el gráfico es suficiente garantía para realizar la inferencia.

Pregunta 3*Antes de Experimentación*

No, mismas razones anteriores.

Después de Experimentación

Si decimos que hay preferencia, ya que la probabilidad de que 180 pertenezca a una muestra donde no hay preferencia es demasiado baja, por lo tanto decimos que si hay preferencia.

Preestructural → Multiestructural

Al igual que en las preguntas anteriores, Andrés siguió condicionando su respuesta a la pequeñez de la muestra, pero ya después de la experimentación evidencia que hay preferencia por el equipo A argumentando con ayuda de *Fathom* que la probabilidad es muy baja, aunque parece referirse al valor puntual de 180 dejando de lado valores superiores. Si bien en sus respuestas anteriores Andrés no dio muestras de considerar la probabilidad como argumento para contrastar las hipótesis en juego, ahora sí lo utiliza, no obstante que no dice explícitamente cuál fue su valor calculado.

Pregunta 4

Antes de Experimentación

270, tenemos que el porcentaje con respecto a la población es demasiado bajo para determinar, sin embargo si se tiene una preferencia en esa muestra del 90% considero que es posible que haya preferencia.

Después de Experimentación

En 160 se considera que pertenece a una muestra sin preferencia y 170 a una muestra que si tiene preferencia en los 165 hay un valor intermedio por lo tanto superior a eso se puede considerar que hay preferencia.

Uniestructural → Multiestructural

En esta última pregunta, Andrés presenta un conflicto con su argumento sobre la pequeñez del tamaño muestral pues considera que hay preferencia si el 90% de la muestra prefiere el equipo A. Después de la experimentación, Andrés no utilizó el cálculo de la probabilidad para seleccionar un, se basó únicamente en el análisis del gráfico para dar uno; sin embargo, su análisis es válido tomando informalmente un nivel de significancia del 5% puesto que el p-valor a partir de 165 es menor a 0.05.

SOFÍA

Pregunta 1

Antes de Experimentación

Pues se podría decir que probablemente no, pues 300 es una muestra pequeña de los 20000, lo que significa que no se puede afirmar que el equipo A sea preferido.

Después de Experimentación

Si, pues la preferencia es muy extraña de que 170 hombres prefieran el equipo A en este modelo.

Preestructural → Uniestructural

Sofía sigue el común denominador de sus compañeros anteriores afirmando antes de la experimentación que el tamaño de la muestra no es lo suficientemente grande para hacer una inferencia sobre la población. Después de la experimentación, su respuesta está basada en el análisis de la simulación con *Fathom*, pero sus argumentos son limitados y poco explicativos. Sofía no explica que el modelo sobre el cual hizo la inferencia es un modelo de distribución 50-50, ni argumenta por qué 170 es un valor extraño dentro de esta distribución, tampoco menciona un valor de probabilidad.

Preguntas 2

Antes de Experimentación

No, pues 300 sigue siendo una muestra muy pequeña y 158 de estos 300 sacados de los 2000 sigue siendo muy poco para asegurar que el equipo A sea preferido.

Después de Experimentación

No pues este grupo de 158 no muestran preferencia pues este grupo está entre el modelo y es más repetitivo que salga, por lo tanto no es extraño entre este grupo de datos.

Preestructural → Uniestructural

Esta vez después de la experimentación, Sofía mejora su argumento con respecto al de la pregunta anterior, aunque de nuevo continúa sin describir o especificar que el modelo al que se refiere es al de distribución 50-50. Al argumentar que “*está entre el modelo*”, Sofía hace referencia

a que 158 es un valor que hace parte de los valores donde se encuentra la mayor acumulación de la distribución 50-50. Para destacar en la argumentación de Sofía, la mención a la frecuencia de veces que puede salir una muestra de 158 personas con preferencia por el equipo A: *y es más repetitivo que salga*, dando muestras de poseer una intuición del enfoque frecuencial de la probabilidad al estilo de Aristóteles: “lo más probable es lo que más se repite”.

Pregunta 3

Antes de Experimentación

No porque y son pocos hombres entre los 300, si son 180 con el equipo A, 120 del B, por lo que se ve, no hay mucha diferencia entre la preferencia de los equipos.

Después de Experimentación

Si pues en este modelo el valor de 180 hombres que eligieron A es extraño por lo tanto me puede mostrar una preferencia.

Preestructural → Uniestructural

Para esta pregunta, Sofía deja de lado su argumento de la pequeñez del tamaño, tomando ahora el argumento que la diferencia entre preferencias no es considerable (180-120). Después de trabajar con *Fathom*, Sofía muestra en su respuesta que entendió el contraste de hipótesis que se planteó con la simulación, pero sus argumentos de nuevo son limitados y no explicativos.

Pregunta 4

Antes de Experimentación

No estoy de acuerdo con una muestra de 300 personas, pues es una muestra muy pequeña para afirmar la preferencia de los hombres.

Después de Experimentación

Pues 165 ya es un valor que se ve extraño entre este modelo. Por lo tanto los valores superiores a 165 me generan una preferencia.

Preestructural → Uniestructural

A pesar de no mencionarlo en la pregunta anterior, Sofía continúa con su argumento de la pequeñez de la muestra; así las 300 personas eligieran el equipo A, ella no contempla una evidencia a favor de una preferencia. Para después de la experimentación, Sofía sigue el mismo modelo de respuesta de las preguntas anteriores, donde se puede ver que entiende el objetivo de la simulación, la identificación de los valores extraños, tal como ella lo manifiesta. Finalmente el nivel de significancia adoptado por Sofía es de aproximadamente el 5%.

GABRIELA**Pregunta 1*****Antes de Experimentación***

No, ya que 20000 →100%, 300 solo equivale al 1,5% de la población, y esto no sería suficiente para afirmar esto.

Después de Experimentación

Si, ya que en la gráfica se evidencia que los valores mayores o iguales a 170, son extraños, por lo tanto, se puede decir que hay una preferencia. De 5000 muestras que hicimos 54 de ellas son mayores o iguales a 170 que eligen el equipo A, con la probabilidad de un 1%.

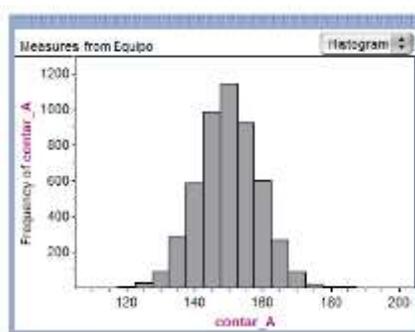


Figura 37. Evidencia Gabriela Taller 1-1

Uniestructural → Multiestructural

Al igual que sus compañeros, Gabriela sigue la constante de respuesta negativa sustentada en la pequeñez de la muestra; con la diferencia que su argumento está más elaborado afirmando que un 1.5% de la población es muy poco para poder inferir sobre ella. Gabriela utilizó el trabajo de la simulación con *Fathom* para dar una respuesta soportada con argumentos proporcionados por el software. Ella basa sus argumentos tanto en el gráfico, como en la probabilidad de la proporción de valores “*mayores iguales a 170*”. Gabriela es el único estudiante que inclusive presenta el gráfico como evidencia de sus argumentos.

Preguntas 2

Antes de Experimentación

Sucedo lo mismo que en el caso anterior la muestra es muy pequeña, esto solo equivale al 0,79% de toda la población.

Después de Experimentación

No, ya que 158 es un valor más común, se podría decir que no existe preferencia por ningún equipo ya que sería un valor más cercano a cincuenta-cincuenta, sería un valor más común en una ciudad en donde no existe preferencia.

Preestructural → Multiestructural

Gabriela extrapola el valor de la preferencia a la población, aunque en los argumentos que venía dando decía que la muestra es muy pequeña, con los argumentos para esta pregunta parece dejar de lado el tamaño de la población y comparar el valor de la preferencia directamente en la población; es decir, para Gabriela si sólo el 0,79% de la población prefiere el equipo A, no se puede decir que haya preferencia. Después de la experimentación, aunque Gabriela no explicita la utilización del gráfico, si evidencia su comprensión y aplicación en su argumento, “ningún equipo ya que sería un valor más cercano a cincuenta-cincuenta” haciendo referencia al modelo de distribución utilizado como hipótesis inicial de no preferencia.

Pregunta 3***Antes de Experimentación***

Esta cantidad de igual manera sigue representando una mínima de la población 0,9%

Después de Experimentación

Si, ya que si yo afirmo 170 es un valor para decir que hay preferencia, los valores mayores a este número, también lo son.

Preestructural → Relacional

Al igual que en la pregunta anterior, Gabriela dejó de lado el tamaño de la muestra, discriminándolo y evaluando la preferencia directamente en la población. Con la utilización de nuevo de este argumento, Gabriela muestra no contradicción con sus argumentos de la prueba diagnóstica donde evidenciaba el sesgo de representatividad determinista: la muestra debe tener las mismas propiedades que la población. Después de la experimentación, ella utiliza los argumentos de la pregunta 1 para relacionarlos con esta pregunta. Gabriela dice “*los valores mayores a este número, también lo son*” mostrando de nuevo una excelente interpretación de la situación y de las bases de una prueba de hipótesis.

Pregunta 4***Antes de Experimentación***

300, si escojo 300 de la población significa que el 1,5% de esta tiene preferencia por el equipo A, pero esto sigue siendo poco, ya que está un 98,5% de la población.

Después de Experimentación

Entre las 5000 muestras que hicimos podemos aceptar que, si entre una comunidad se le preguntan a 300 hombre y 164 de ellos eligen el equipo A, se puede aceptar que hay una

preferencia, ya que la probabilidad es de 6,1%, este es un valor aceptable, para decir que existe preferencia.

Preestructural → Relacional

Antes de trabajar con *Fathom*, Gabriela continúa ignorando el tamaño de la muestra y analizando la proporción que respecto a la población representan las personas que prefieren al equipo A. Queda la inquietud de saber qué porcentaje de personas de la muestra favorables al equipo A haría que Gabriela optara por aceptar la preferencia de la población por el equipo A. Después de la experimentación, Gabriela utiliza el software y el cálculo de la probabilidad de la proporción que aprendió, para dar un valor de 164 con el cual, como lo indica en su argumento, opta por un nivel de significancia inferior al 7%. De igual forma, la descripción de parte de la simulación evidencia de nuevo la comprensión del contraste de hipótesis que se planteó.

LUIS

Pregunta 1

Antes de Experimentación

No, si los 300 hombres son seleccionados aleatoriamente y la cantidad total son 20000, sería erróneo concluir que por 170 de 300 los otros 19600 arrojen el mismo resultado.

Después de Experimentación

Si porque 170 es un dato lejano a lo que la no preferencia es. El 170 demuestra preferencia hacia el equipo A.

Preestructural → Uniestructural

En su respuesta inicial, Luis nuevamente destaca la pequeñez de la muestra para realizar inferencia a toda la población. . Ya después de la experimentación el pensamiento de Luis es otro: ahora tiene en consideración una zona de valores donde no existe preferencia y otra, fuera de ella, donde sí existe. Esto es, Luis determina una zona de aceptación y otra de rechazo. El asunto ahora, es saber cómo determinó la frontera de las dos zonas.

Preguntas 2

Antes de Experimentación

También es incorrecto afirmar que en general hay una preferencia por el “Equipo A” debido a que el grupo del cual se está tomando la muestra es muy pequeño en comparación al número total de hombres.

Después de Experimentación

No, porque 158 se encuentra dentro del rango de valores que muestran una preferencia por el equipo A o B.

Preestructural → Multiestructural

Inicialmente antes de la experimentación se sigue la constante que presentaron todos los estudiantes: el tamaño de la muestra no es lo suficientemente grande para poder hacer una inferencia sobre preferencia. Ya después de la experimentación, Luis de nuevo, aunque no lo muestra en su respuesta utilizó el gráfico para realizar la inferencia

Pregunta 3***Antes de Experimentación***

Entre más se acerque el número a 300 por la preferencia del “Equipo A” mayor probabilidad que la conclusión sea asertiva.

Después de Experimentación

Si, con mayor razón demuestra preferencia por el equipo A ya que se aleja aún más del rango que nos haría saber que no hay preferencia.

Preestructural → Multiestructural

Aunque Luis no responde directamente la pregunta que se le hizo, sí afirma algo que es cierto: *Entre más se acerque el número a 300 por la preferencia del “Equipo A” mayor probabilidad que la conclusión sea asertiva.* Interesante, además, que utilice la palabra probabilidad, cosa que no había hecho en sus argumentaciones anteriores.

Ya después de la experimentación, Luis nuevamente recurre a su zona de rechazo sin explicitar ningún valor de probabilidad asociado.

Pregunta 4***Antes de Experimentación***

270, mi respuesta intuitiva serian 270 hombres al ser un número bastante por encima del 50% de los hombres de un grupo de 300.

Después de Experimentación

Al escoger 165 vemos que es un número suficientemente fuera del rango que consideraríamos como no preferencial respecto al equipo A o B.

Preestructural → Uniestructural

Como se mencionó para la pregunta anterior, aunque Luis argumentaba que la muestra de 300 es muy pequeña, ahora cuando se le pide escoger un valor, él adopta un valor lo suficientemente grande para generarle la confianza de poder inferir sobre la existencia de preferencia; es decir, pareciera que la evidencia es tan fuerte que el tamaño de la muestra pasa a un segundo plano. Después de la experimentación, Luis nuevamente recurre a su zona de rechazo que ahora define claramente a partir del valor 165, un nivel de significancia de aproximadamente 5%.

4.2.4. Conclusiones

En este taller se evidenció que los estudiantes antes de la experimentación con *Fathom* presentaron dificultades para aceptar el tamaño de la muestra como suficientemente representativo para realizar una inferencia. Ya después de la experimentación con *Fathom*, algunos lograron superar esa dificultad y lograron ver la proporción muestral como representante de la proporción poblacional.

A partir de la simulación algunos de los estudiantes crearon un criterio de proximidad para realizar inferencia basada en el análisis gráfico; es decir, los estudiantes contrastan la hipótesis nula a partir de advertir que tan lejos está la proporción muestral del centro del gráfico de distribución.

Gracias a la simulación guiada, algunos estudiantes empezaron a crearse una idea del nivel de significancia, a partir de la elección de la región de rechazo. Los niveles de significancia intuitivos de los estudiantes esta alrededor de 10% y 5%, nada mal para ser su primer acercamiento.

4.3. Taller 2

4.3.1. Análisis de Componentes

SITUACIÓN

En una urna hay 10 bolas entre rojas y verdes. Al hacer 100 extracciones con remplazo se obtuvieron 73 bolas rojas y 27 bolas verdes.

Pregunta 1

A partir de la información de las extracciones ¿Cuál o cuáles de las siguientes composiciones podría tener la urna? ¿Justifique para cada una?

Tabla 6. Taller 2-Composiciones de la Urna

Composición de la Urna	JUSTIFICACIÓN
10 Rojas	
9 Rojas y 1 Verde	
8 Rojas y 2 Verdes	
7 Rojas y 3 Verdes	
6 Rojas y 4 Verdes	
5 Rojas y 5 Verdes	
4 Rojas y 6 Verdes	
3 Rojas y 7 Verdes	
2 Rojas y 8 Verdes	

1 Roja y 9 Verdes
10 Verdes

❖ Componentes conceptuales

5. Principal: Inferencia sobre composición de la urna.
6. Secundario: Variabilidad muestral.

Pregunta 2

Ana afirma que la composición de la urna es 7 Rojas y 3 Verdes, Edinson que es 8 Rojas y 2 Verdes, mientras que Jonatan asegura que la distribución puede ser 1 Roja y 9 Verdes. ¿Cuál o cuáles afirmaciones considera válidas? ¿Cuál o cuáles no? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis.
2. Principal: Inferencia.
3. Secundario: Variabilidad Muestral.

4.3.2. Experimentación *Fathom*

Para esta experimentación se les proporcionará a los estudiantes otra forma de simular distribuciones diferentes al *randomPick*. Los pasos para la simulación esperada son los siguientes:

1. La simulación se realizará para la el supuesto de una urna con 7 bolas Rojas y 3 Verdes.

El primer paso es crear una colección llamada Urna, y una variable llamada **B_Urna**.

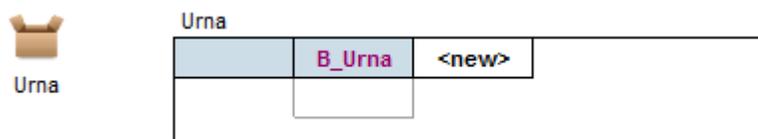


Figura 38. Taller 2-Simulación esperada 1

2. En el editor de formula digite el siguiente comando

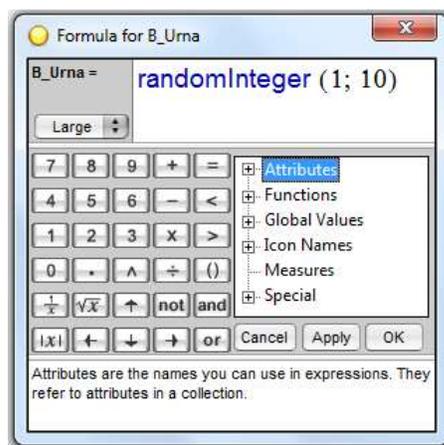
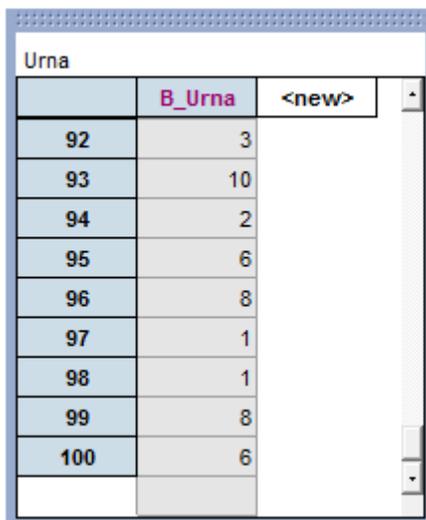


Figura 39. Taller 2-Simulación esperada 2

El comando randomInteger permite elegir aleatoriamente entre números enteros definidos en un intervalos cerrado; en este caso el comando escoge un número aleatorio entre 1 y 10.

3. Se crea una muestra de 100 casos como lo indica la situación.



	B_Urna	<new>
92	3	
93	10	
94	2	
95	6	
96	8	
97	1	
98	1	
99	8	
100	6	

Figura 40. Taller 2-Simulación esperada 3

4. El siguiente paso es asignarle un color a cada número, en este caso como se quieren 7 bolas rojas, los número 1, 2, 3, 4, 5, 6 y 7 representaran las bolas rojas y los números 8, 9 y 10 las bolas verdes.

Se crea la variable BR (bola roja) y se define el siguiente comando:

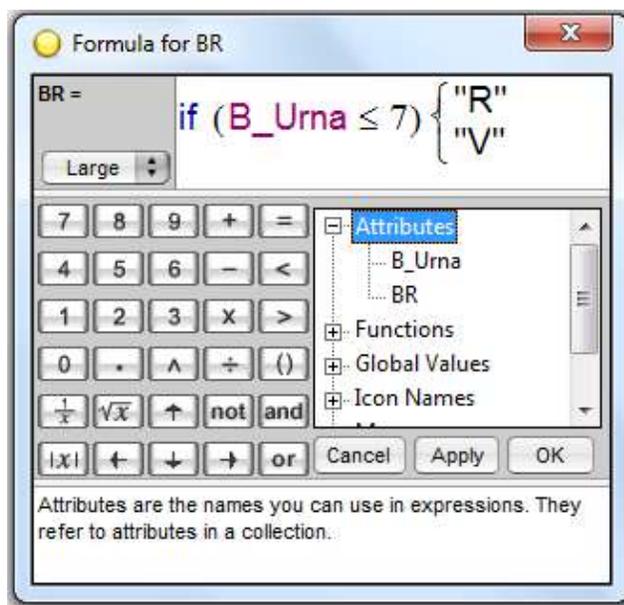


Figura 41. Taller 2-Simulación esperada 4

El comando **If**, convierte los números menores iguales a 7 en bolas rojas y los demás en bolas verdes.

5. Ya se cuenta con una muestra de 100 extracciones de una urna que contiene 7 bolas rojas y 3 bolas verdes.

Urna		
	B_Urna	BR
92	2	R
93	2	R
94	5	R
95	1	R
96	1	R
97	3	R
98	4	R
99	4	R
100	5	R

Figura 42. Taller 2-Simulación esperada 5

6. El proceso que se sigue a continuación es igual a realizado en el taller anterior. Entonces tenemos:

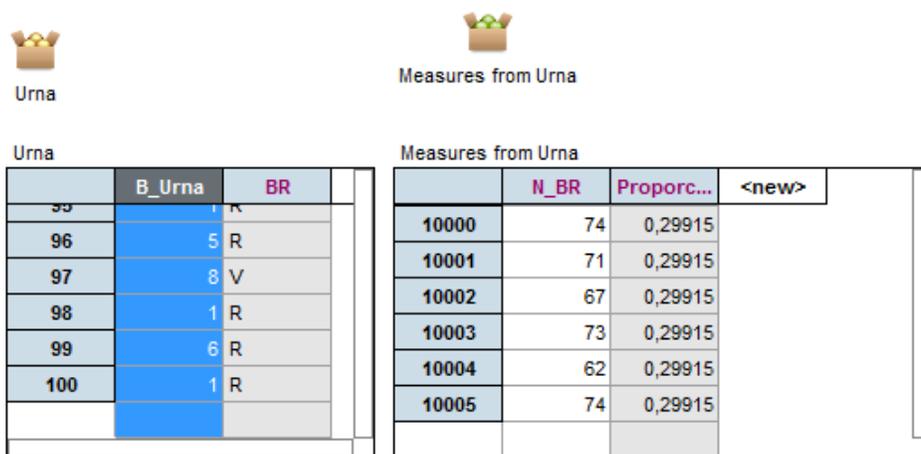


Figura 43. Taller 2-Simulación esperada 6

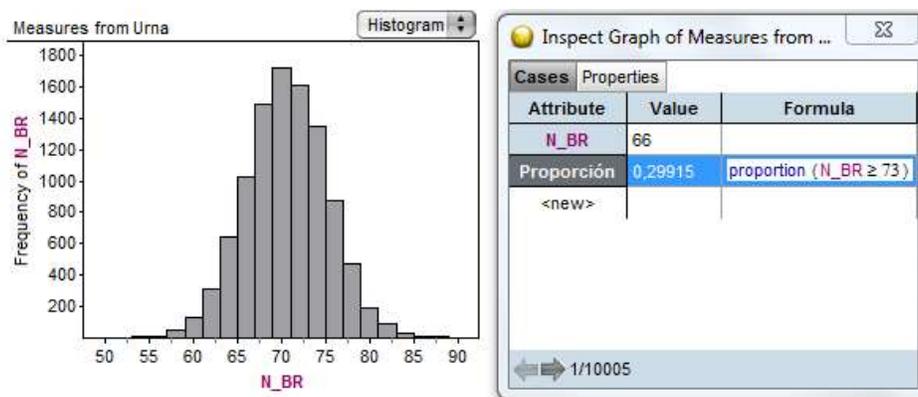


Figura 44. Taller 2-Simulación esperada 7

No se puede rechazar la hipótesis nula de que la urna tenga una composición de 7 bolas rojas y 3 negras pues el p-valor es de aproximadamente el 0.29.

7. Este mismo proceso se repite con las demás hipótesis de supuestas composiciones de la urna.

4.3.3. Análisis por estudiante

ANDRÉS

Pregunta 1

Antes de Experimentación

Tabla 7. Taller 2-Composiciones de la Urna Andrés 1

Composición de la Urna	JUSTIFICACIÓN
10 R	No, porque siempre saldrán rojas, o sea las 100 bolas hubieran sido rojas.
9 R y 1 V	No, porque la probabilidad de que salieran bolas rojas sería próximo a la 90/100.
8 R y 2 V	Sí, tenemos 73/100 en bolas rojas y 27/100 bolas verdes, si dividimos en 10/10 podemos apreciar valores próximos.
7 R y 3 V	Si, misma explicación anterior
6 R y 4 V	No, porque se aleja demasiado de 7.3 y 2.7 lo que me hace pesar que es poco probable.
5 R y 5 V	No, deberían estar los valores rodeando las 50 bolas.
4 R y 6 V	No, habría más probabilidad de que salieran bolas verdes.
3 R y 7 V	No, más probabilidad de verdes.
2 R y 8 V	No, más probabilidad de verdes.
1 R y 9 V	No, más probabilidad de verdes.
10 V	No, más probabilidad de verdes.

Después de Experimentación

Tabla 8. Taller 2-Composiciones de la Urna Andrés 2

Composición de la Urna	JUSTIFICACIÓN
10 R	No, porque entonces en las 100 extracciones hubiesen salido todas rojas
9 R y 1 V	No, porque los valores estarían cercanos entre 90 verdes y 10 rojas
8 R y 2 V	Observando el anexo “Measures from Collection 2” podemos observar que los valores dados están muy lejanos a esta probabilidad

7 R y 3 V	Si, observando el anexo “Measures from Collection 1” podemos observar que es muy probable que esta sea la distribución de las bolas ya que 73 está casi en el centro de la grafica
6 R y 4 V	No, porque los valores estarían cercanos entre 60 verdes y 40 rojas
5 R y 5 V	No, porque los valores estarían rodeando los 50 en ambos casos
4 R y 6 V	No, porque es más probable que haya un mayor número de verdes
3 R y 7 V	No, porque es más probable que haya un mayor número de verdes
2 R y 8 V	No, porque es más probable que haya un mayor número de verdes
1 R y 9 V	No, porque es más probable que haya un mayor número de verdes
10 V	No, porque entonces en las 100 extracciones hubiesen salido todas verdes

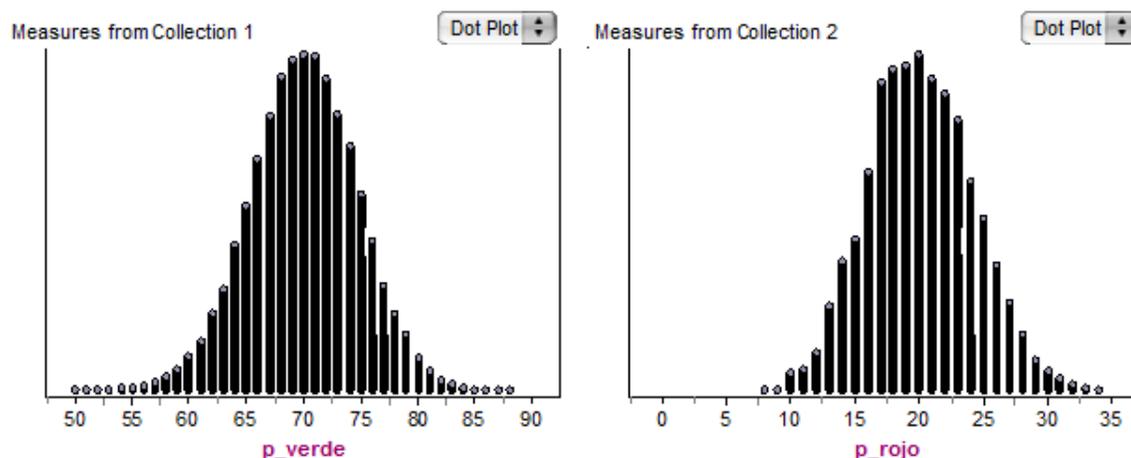


Figura 45.Evidencia Andrés Taller 2-1

Uniestructural → Uniestructural

Inicialmente, Andrés hace uso de la frecuencia relativa para hacer su inferencia, tomando como válidas las composiciones próximas a 7.3 bolas rojas; las cuales son (7R, 3V) y (8R, 2V). Además, evidencia comprensión de que la muestra manifiesta la composición de la urna, así pues, como muy bien menciona no pueden haber menos bolas rojas porque *“habría más probabilidad de que salieran bolas verdes”*.

Durante la experimentación, Andrés cambió bolas rojas por verdes, por lo cual sus respuestas y su simulación están bajo la premisa de 73 verdes y 27 rojas. Andrés hace la simulación de las composiciones (7R, 3V) y (8R, 2V), y sus respectivas gráficas.

Ya después de la experimentación Andrés, al igual que en el Taller 1, se basa sólo en el gráfico para hacer la inferencia; en consecuencia sigue utilizando los términos “*cercanos*” y “*muy lejanos*” que no le permiten cuantificar la inferencia.

Pregunta 2

Antes de Experimentación

Ana y Edinson podrían tener razón ya que la probabilidad que ellos plantean está cercana a los resultados arrojados. La de Jonathan no porque habría más probabilidad de que salieran bolas verdes.

Después de Experimentación

Considero que Ana tiene razón ya si analizamos el anexo “Measures from Collection 1” podemos observar que 73 está casi en el centro de la gráfica lo que me da una gran probabilidad de que sea de esta manera.

Los demás no porque están muy lejanos a los valores como por ejemplo Edinson. Podemos observar en “Measures from Collection 2” que el valor 27 está muy lejano del centro de la gráfica; y Jonathan está demasiado lejos de ambos valores haciendo referencia a que es más probable que las bolas con mayor numero sean las verdes.

Multiestructural → Multiestructural

Como se mencionó anteriormente, Andrés inconscientemente cambió el color de las bolas para su análisis. Respecto a la respuesta, en primera instancia podemos evidenciar que Andrés percibe la relación entre la variabilidad de la proporción muestral y la composición de la urna, al no aceptar las composiciones alejadas de dicha proporción; por ejemplo (1R, 9V).

Ya después de la experimentación vemos que Andrés, al igual que para el inciso anterior, basa su análisis en el gráfico y no utiliza el cálculo de la proporción para cuantificar su inferencia. El argumento “*el valor 27 está muy lejano del centro de la gráfica*” y los términos “*cercanos*” y “*muy lejanos*” utilizados en este taller para hacer la inferencia son similares al utilizado en el primer taller “*si se salía del centro de la gráfica significaba que había una preferencia*”. La recurrencia de este método, muestra que Andrés no encuentra la necesidad de cuantificar su respuesta con un valor de probabilidad.

SOFÍA

Pregunta 1

Antes de Experimentación

Tabla 9. Taller 2-Composiciones de la Urna Sofía 1

Composición de la Urna	JUSTIFICACIÓN
10 R	No pertenece a la composición pues en la urna hay tanto bolas rojas y verdes
9 R y 1 V	No pertenece pues es una proporción que no se acerca a 73 es a 27

8 R y 2 V	Esta proporción podría ser una composición de la urna porque hay más rojas que verdes en una diferencia grande
7 R y 3 V	Esta es la composición de la urna porque es la que más se acerca a la proporción de 73 rojas a 27 verdes
6 R y 4 V	Esta proporción se acerca un poco a las extracciones pero no es muy exacta
5 R y 5 V	No pertenece pues no hay igualdad de extracciones de la urna
4 R y 6 V	No pertenece por que en la composición deben haber más bolas rojas que verdes
3 R y 7 V	No pertenece por que en la composición deben haber más bolas rojas que verdes
2 R y 8 V	No pertenece por que en la composición deben haber más bolas rojas que verdes
1 R y 9 V	No pertenece por que en la composición deben haber más bolas rojas que verdes
10 V	No pertenece a la composición pues en la urna hay tanto bolas rojas y verdes

Después de Experimentación

Tabla 10. Taller 2-Composiciones de la Urna Sofía 2

Composición de la Urna	JUSTIFICACIÓN
10 R	No pertenece a la composición pues en la urna hay tanto bolas rojas y verdes
9 R y 1 V	No pertenece pues es una proporción que no se acerca a 73 es a 27 , pues está muy lejos de que sea esta composición
8 R y 2 V	Esta proporción no es la composición de la urna pues la proporción de que salgan 73 bolas rojas es del 0,0613466
7 R y 3 V	Esta es la composición de la urna porque es la que más se acerca a la proporción de 73 rojas a 27 verdes pues la proporción de que sean rojas en el modelo es de 0,297257
6 R y 4 V	Esta proporción no corresponde pues la proporción es de 0.00997
5 R y 5 V	Esta composición no pertenece pues en la urna no hay igualdad de bolas
4 R y 6 V	Esta composición no pertenece pues en la urna hay más bolas rojas que verdes por lo tanto la composición debe ser más bolas rojas que verdes
3 R y 7 V	Esta composición no pertenece pues en la urna hay más bolas rojas que verdes por lo tanto la composición debe ser más bolas rojas que verdes
2 R y 8 V	Esta composición no pertenece pues en la urna hay más bolas rojas que verdes por lo tanto la composición debe ser más bolas rojas que verdes

1 R y 9 V	Esta composición no pertenece pues en la urna hay más bolas rojas que verdes por lo tanto la composición debe ser más bolas rojas que verdes
10 V	No pertenece a la composición pues en la urna hay tanto bolas rojas y verdes

Multiestructural → Multiestructural

Sofía antes de la experimentación aceptó tres posibles composiciones de la urna con 6, 7 o 8 bolas rojas. Sofía argumentó que la composición (7R, 3V) es la más cercana a la proporción de la muestra, pero tuvo en cuenta la variabilidad muestral para admitir que la muestra podía provenir de una urna con composición (8R, 2V) o una composición (6R, 4V) en menor medida que la anterior. Para las composiciones menores de 5 bolas rojas el argumento fue *“No pertenece porque en la composición deben haber más bolas rojas que verdes”*, ratificando que la noción de proporción muestral como un representante de la proporción poblacional evidenciada en la prueba diagnóstica no fue casualidad.

Después de la experimentación, Sofía instintivamente utilizó la estimación de máxima verosimilitud para decidir la composición de la urna, escogiendo solo como composición de la urna la que se acerca más a la proporción muestral. En la simulación que realizó, Sofía calculó la probabilidad de que la muestra perteneciera a cada una de las composiciones que creyó correctas antes de la experimentación; finalmente eligió la composición de la urna que generó la mayor probabilidad (7R, 3V).

Pregunta 2

Antes de Experimentación

Yo considero que la afirmación de Ana es la correcta pues se aproxima más a las extracciones de 73 bolas rojas a 27 bolas verdes, pues la proporción $\frac{73}{27}$ como $\frac{7}{3}$, por lo tanto la afirmación de Edinson no es tan proporcional a 73 es a 27 y la de Jonathan es totalmente equivocada pues existen mas bolas rojas que verdes por lo tanto su afirmación es errónea.

Después de Experimentación

La afirmación de Ana es totalmente correcta pues esta composición corresponde con la proporción de 73 a 27, y las afirmaciones de Edinson y Jonathan no corresponden pues la proporción de Edinson es de 6,1% y la Jonathan es mucho más menor.

Multiestructural → Multiestructural

Antes de la experimentación, Sofía, a diferencia de lo expuesto en la tabla de la primera pregunta, decide elegir únicamente la composición de urna (7R, 3V). Después de la experimentación, Sofía reafirma su respuesta anterior, incluyendo porcentajes como argumentos para cuantificar sus respuestas (p-valor inform). Además, teniendo en cuenta el 6.1% sobre el cual Sofía basa su argumento, se puede intuir que Sofía maneja informalmente un nivel de significancia superior al 5%.

GABRIELA

Pregunta 1

Antes de Experimentación

Tabla 11. Taller 2-Composiciones de la Urna Gabriela 1

Composición de la Urna	JUSTIFICACIÓN
10 R	No, debido a que también salieron bolas verdes, entonces debe haber por lo menos una bola verde.
9 R y 1 V	Podría tener 9 rojas y una verde, ya que la cantidad de bolas rojas que extrajeron fue mayor, por lo tanto debe haber una mayor cantidad de bolas rojas.
8 R y 2 V	Podría tener 8 rojas y 2 verde, ya que la cantidad de bolas rojas que extrajeron fue mayor, por lo tanto debe haber una mayor cantidad de bolas rojas.
7 R y 3 V	Podría tener 7 rojas y 3 verde, ya que la cantidad de bolas rojas que extrajeron fue mayor, por lo tanto debe haber una mayor cantidad de bolas rojas.
6 R y 4 V	Podría tener 6 rojas y 4 verde, ya que la cantidad de bolas rojas que extrajeron fue mayor, por lo tanto debe haber una mayor cantidad de bolas rojas.
5 R y 5 V	No, ya que si hubiera la misma cantidad de bolas la cantidad de bolas extraídas serian valores cercanos a 50.
4 R y 6 V	No, porque si hubieran más bolas verdes que rojas, hubiera sacado más bolas verdes y eso no sucedió.
3 R y 7 V	No, porque si hubieran más bolas verdes que rojas, hubiera sacado más bolas verdes y eso no sucedió.
2 R y 8 V	No, porque si hubieran más bolas verdes que rojas, hubiera sacado más bolas verdes y eso no sucedió.
1 R y 9 V	No, porque si hubieran más bolas verdes que rojas, hubiera sacado más bolas verdes y eso no sucedió.
10 V	No, debido a que también salieron bolas rojas, por lo tanto debe haber una bola roja.

Después de Experimentación

Tabla 12. Taller 2-Composiciones de la Urna Gabriela 2

Composición de la Urna	JUSTIFICACIÓN
10 R	No, ya que en la urna hay bolas verdes como rojas.
9 R y 1 V	No, ya que si la siguiente urna no es suficiente para afirmar que esa muestra pertenece, para una urna con mayores bolas rojas tampoco.

8 R y 2 V	No, ya que la probabilidad que sean 73 bolas rojas en una urna de este estilo es del 6,1%, es un porcentaje muy pequeño, para poder decir que ese valor pertenece a esta urna.
7 R y 3 V	Si, ya que al realizar las muestras hay una probabilidad de 29,97% de al realizar la situación 100 veces se obtengan 73 veces bolas rojas, por lo tanto, se puede afirmar que este valor pertenece a esta urna.
6 R y 4 V	No, al hacer las muestras la probabilidad es muy mínima de que se obtengan 73 veces bolas rojas, no es suficiente para poder afirmar que este valor pertenece a esta urna.
5 R y 5 V	No, porque la urna no puede ser 50-50, si fuera de este estilo las extracciones de bolas rojas y verdes, serian valores cercanos.
4 R y 6 V	No, ya que, al realizar las extracciones, hay más bolas rojas que verdes, por lo tanto, en la urna debe haber mayor cantidad de bolas rojas que verdes.
3 R y 7 V	No, ya que, al realizar las extracciones, hay más bolas rojas que verdes, por lo tanto, en la urna debe haber mayor cantidad de bolas rojas que verdes.
2 R y 8 V	No, ya que, al realizar las extracciones, hay más bolas rojas que verdes, por lo tanto, en la urna debe haber mayor cantidad de bolas rojas que verdes.
1 R y 9 V	No, ya que, al realizar las extracciones, hay más bolas rojas que verdes, por lo tanto, en la urna debe haber mayor cantidad de bolas rojas que verdes.
10 V	No, ya que en la urna hay bolas verdes como rojas.

Uniestructural → Multiestructural

Para Gabriela la urna puede estar compuesta por 6, 7, 8 o 9 bolas rojas, evidenciando al igual que en la prueba diagnóstica concepción de la variabilidad de la proporción muestral. Gabriela fue la estudiante que más composiciones posibles aceptó, sólo descartó las composiciones de urna más propensas a ser descartadas: los casos imposibles con 0 y 10 bolas rojas; los casos donde hay menos bolas rojas 1, 2, 3 y 4; y el caso donde hay igual cantidad de bolas rojas y verdes.

Gabriela después de la experimentación utiliza, al igual que Sofía, recurre al estimador de máxima verosimilitud, eligiendo como composición correcta (7R, 3V), por el hecho de tener más probabilidad de ocurrencia que las otras composiciones.

Pregunta 2

Antes de Experimentación

Las afirmaciones de Ana y Edinson son válidas ya que hay una mayor probabilidad de que salga roja, debido a que hay mayor cantidad, hay el doble o más respecto a las bolas verdes.

Después de Experimentación

Como se justifica en la tabla anterior, la composición de Ana es la afirmativa, debido ese valor pertenece a esa urna, entonces podemos negar la de Edinson, porque si hay cuatro veces más bolas rojas que verdes al momento de hacer la extracción debería haber más bolas rojas, pero de igual manera podemos negar la de Jonatán, ya que no puede haber más bolas verdes que rojas, porque no se cumpliría con la muestra obtenida.

Uniestructural → Multiestructural

Guardando relación con la respuesta anterior Gabriela antes de la experimentación, toma como válidas las afirmaciones de Ana y Edinson. Ya después de la experimentación Gabriela optó por la opción de Ana, convirtiendo la afirmación de Edinson en incorrecta. Se descarta en la respuesta

de Gabriela el hecho de notar y explicitar las implicaciones de la proporción muestral “*ya que no puede haber más bolas verdes que rojas, porque no se cumpliría con la muestra obtenida*”.

LUIS

Pregunta 1

Antes de Experimentación

Tabla 13. Taller 2-Composiciones de la Urna Luis 1

Composición de la Urna	JUSTIFICACIÓN
10 R	No porque es evidente que hay bolas verdes
9 R y 1 V	No porque el valor del número de veces que extraemos bolas verdes sería más bajo
8 R y 2 V	Sí, es un porcentaje de bolas que puede dar con ese rango
7 R y 3 V	Sí, aún se podría obtener una cantidad similar al número de bolas rojas y verdes en la extracción
6 R y 4 V	No porque es un porcentaje de bolas muy cercano al 50% de una a la otra.
5 R y 5 V	No porque es un porcentaje de bolas muy cercano al 50% de una a la otra.
4 R y 6 V	No porque es un porcentaje de bolas muy cercano al 50% de una a la otra.
3 R y 7 V	No, puedo intuir que hay más bolas rojas que verdes debido al resultado de la extracción.
2 R y 8 V	No, puedo intuir que hay más bolas rojas que verdes debido al resultado de la extracción.
1 R y 9 V	No, puedo intuir que hay más bolas rojas que verdes debido al resultado de la extracción.
10 V	No porque evidentemente tenemos bolas rojas

Después de Experimentación

Tabla 14. Taller 2-Composiciones de la Urna Luis 2

Composición de la Urna	JUSTIFICACIÓN
10 R	No pueden ser 10 rojas puesto que es evidente que obtenemos bolas verdes.
9 R y 1 V	No es posible que esta sea la composición ya que es aún más complicado obtener el resultado de 27 bolas verdes que en el caso de 8 bolas rojas y 2 verdes.
8 R y 2 V	Es poco probable que esta sea la composición ya que siendo verde el número menor de bolas extraídas solo hay un 5% de probabilidad de obtener un número igual o mayor a 27 bolas.
7 R y 3 V	En este caso la probabilidad de obtener un número igual o menor a 27 bolas verdes es del 30% lo cual comparado con el 5% de los dos casos más cercanos a esta extracción es bastante alto.
6 R y 4 V	La probabilidad de que esta sea la composición disminuye nuevamente ya que solo un aproximado del 5% va a ser menor o igual a 27 bolas.
5 R y 5 V	No es posible porque el número de bolas verdes extraídas y la cantidad de bolas rojas debería ser más cercano.
4 R y 6 V	No puede ser la composición, ya que obtendríamos más bolas verdes que rojas. Lo cual no está sucediendo
3 R y 7 V	No puede ser la composición, ya que obtendríamos más bolas verdes que rojas. Lo cual no está sucediendo
2 R y 8 V	No puede ser la composición, ya que obtendríamos más bolas verdes que rojas. Lo cual no está sucediendo
1 R y 9 V	No puede ser la composición, ya que obtendríamos más bolas verdes que rojas. Lo cual no está sucediendo
10 V	No es posible que sean 10 bolas verdes ya que es evidente que tenemos bolas rojas al ser posible extraerlas.

(8 rojas 2 verdes)

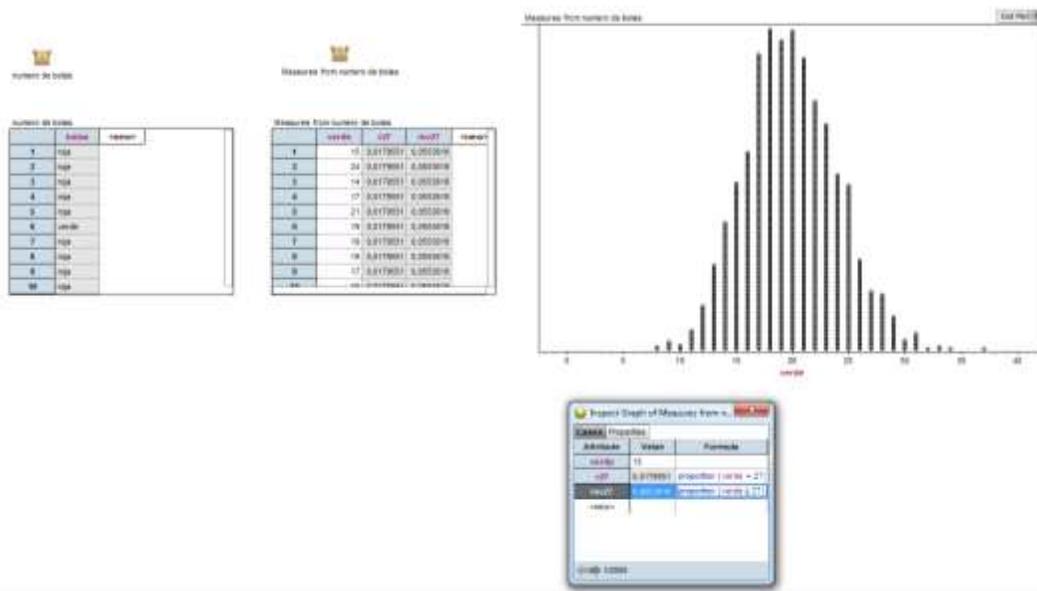


Figura 46. Evidencia Luis Taller 2-1

(7 rojas 3 verdes)

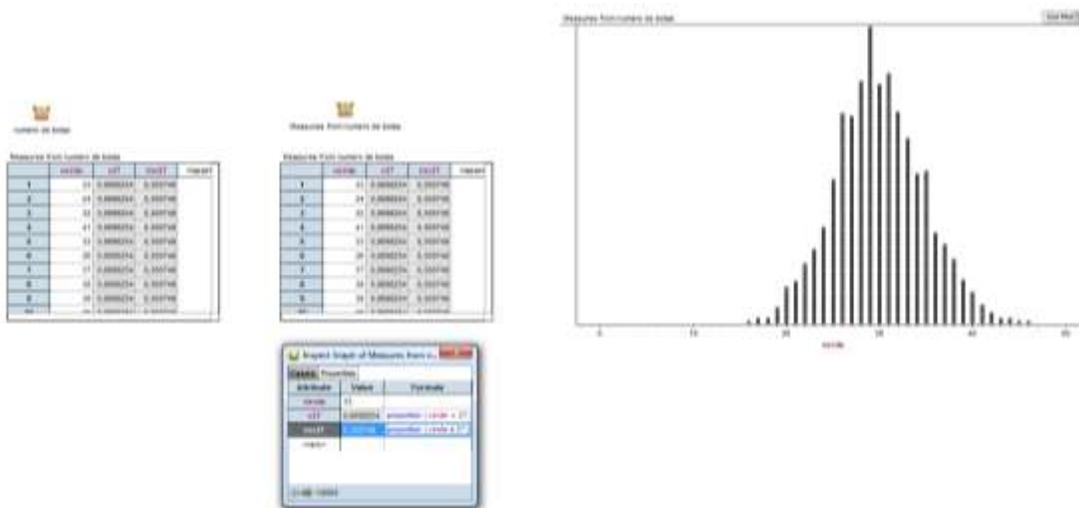


Figura 47. Evidencia Luis Taller 2-2

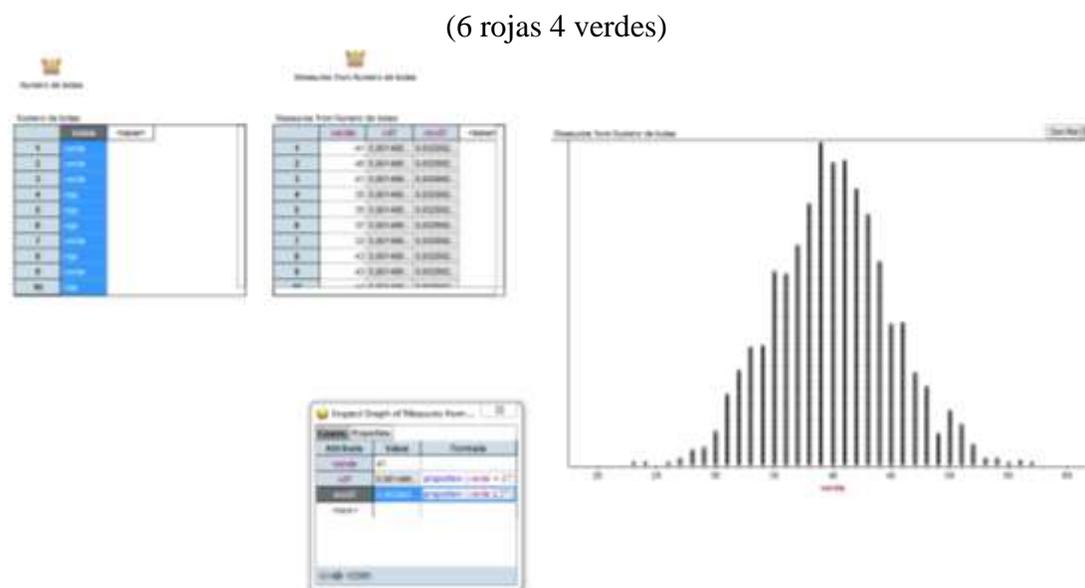


Figura 48. Evidencia Luis Taller 2-3

Uniestructural → Multiestructural

Al igual que Andrés, Luis aceptó las composiciones con 7 y 8 bolas rojas utilizando el argumento del porcentaje de bolas rojas respecto al total de bolas. Según Luis “*es un porcentaje de bolas que puede dar con ese rango*”, evidencia de nociones de variabilidad (que no había presentado en las actividades anteriores). Es importante destacar este avance, porque de los cuatro estudiantes que participaron en la experimentación, Luis era el que tenía más arraigadas las concepciones deterministas.

A pesar de que inicialmente solo consideró las composiciones (8R, 2V) y (7R, 3V), para la experimentación con *Fathom*, Luis también trabajó la composición (6R, 4V), evidenciando que la variabilidad de la muestra no generó la suficiente confianza en la primera pregunta. Después de la experimentación, Luis incluye cada uno de los modelos que trabajó como argumentos de soporte

de sus respuestas. Al igual que sus compañeras, escoge la urna que tiene mayor probabilidad (porcentaje en este caso), aunque vale la pena destacar que Luis no descarta rotundamente las composiciones (8R, 2V) y (6R, 4V) utilizando expresiones como “*es poco probable*” y “*La probabilidad de que esta sea la composición disminuye*”.

Pregunta 2

Antes de Experimentación

Las composiciones de 8 y 2, 7 y 3 me parecen razonables ya que el porcentaje de bolas verdes y rojas puede coincidir con el número de extracciones, pero 9 y 1 no porque el número de extracciones de bolas verdes sería bastante menor.

Después de Experimentación

Estoy de acuerdo con Ana ya que podemos ver que en el caso de 7 bolas rojas y 3 bolas verdes la probabilidad de obtener un resultado así es bastante alta.

Multiestructural → Multiestructural

Luis inicialmente argumenta su respuesta basado en la tabla que completó para la primera pregunta, con la idea de que la proporción muestral es una representante de la composición real de la urna. Después de la experimentación con *Fathom*, Luis eligió como correcta solo la afirmación de Ana mostrando de nuevo nociones intuitivas de máxima verosimilitud.

4.3.4. Conclusiones

Durante el desarrollo de taller se evidenció como los estudiantes aplicaron la proporción muestral y su variabilidad para inferir cuales podrían ser las composiciones de las urnas.

El análisis de la proximidad en los gráficos como criterio para realizar inferencia, se convirtió en una constante en los estudiantes. No obstante, algunos estudiantes ven la necesidad de cuantificar esa proximidad dada por los gráficos.

Respecto al contraste de hipótesis, los estudiantes ya están interiorizando un nivel de significancia informal, y un p-valor informal; inclusive realizan el contraste de hipótesis a partir de la relación de estos dos conceptos.

4.4. Taller 3

4.4.1. Análisis de Componentes

SITUACIÓN

En una urna hay 13 bolas entre rojas y verdes. Ana le dice a Edinson que ella puede decir cuántas bolas hay de cada color sin sacarlas todas a la vez de la urna, simplemente haciendo 150 extracciones con remplazo (es decir, sacar una bola, ver su color y regresarla a la urna de nuevo).

En las 150 extracciones, Ana obtuvo 47 bolas rojas y 103 bolas verdes, y a partir de esto ella afirma que en la urna hay 4 bolas rojas y 9 verdes.

Pregunta 1

¿Está usted de acuerdo con la afirmación de Ana? ¿Por qué?

❖ Componentes conceptuales

1. Principal: Contraste de hipótesis.
2. Principal: Inferencia.
3. Secundaria: Variabilidad Muestral.

Pregunta 2

Si en vez de 47 bolas rojas, Ana hubiera obtenido 65 bolas rojas y 85 verdes ¿Estaría de acuerdo con la afirmación de que en la urna hay 4 bolas rojas y 9 verdes? ¿Por qué?

❖ Componentes conceptuales

1. Principal: Variabilidad Muestral.
2. Principal: Contraste de hipótesis.
3. Principal: Inferencia.

Pregunta 3

¿Cuántas bolas rojas como máximo podrían obtenerse para que usted estuviera de acuerdo con la afirmación de Ana? ¿Por qué?

❖ Componentes conceptuales

3. Principal: p-valor informal.
4. Principal: Región de rechazo.
5. Secundaria: Prueba de hipótesis

6. Secundaria: Inferencia

Pregunta 4

Si la muestra de Ana se duplica; es decir, realiza 300 extracciones ¿Cuántas bolas rojas como máximo podrían obtenerse para que usted estuviera de acuerdo con la afirmación de Ana? Justifique.

❖ Componentes conceptuales

1. Principal: No proporcionalidad de la región de rechazo (Muestra duplicada).
2. Principal: p-valor informal.
3. Principal: Región de rechazo.
4. Secundaria: Contraste de hipótesis
5. Secundaria: Inferencia

Pregunta 5

Y si por el contrario, la muestra se reduce a la mitad, 75 extracciones ¿Cuántas bolas rojas como máximo podrían obtenerse para que usted estuviera de acuerdo con la afirmación de Ana? Justifique.

❖ Componentes conceptuales

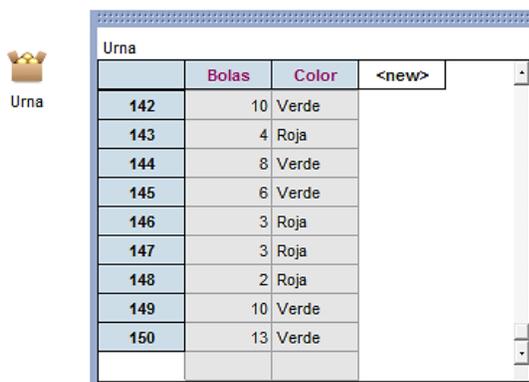
1. Principal: No proporcionalidad de la región de rechazo (Mitad de la muestra inicial).
2. Principal: p-valor informal.
3. Principal: Región de rechazo.
4. Secundaria: Contraste de hipótesis

5. Secundaria: Inferencia

4.4.2. Experimentación *Fathom*

La simulación con *Fathom* que se espera que realicen los estudiantes para la primera pregunta es la siguiente:

1. Crear una colección (Urna) bajo el supuesto de Ana (la distribución de la urna es 4 bolas rojas y 9 verdes). Crear una muestra con 150 extracciones como indica la situación.



	Bolas	Color	<new>
142	10	Verde	
143	4	Roja	
144	8	Verde	
145	6	Verde	
146	3	Roja	
147	3	Roja	
148	2	Roja	
149	10	Verde	
150	13	Verde	

Figura 49. Taller 3-Simulación esperada 1

Hasta el momento los estudiantes conocen dos formas de definir la distribución: con *RandomPick* (definiendo el color de cada una de las 13 bolas) o con *RandomInteger* e *If* (dando valores enteros aleatorios entre 1 y 13, para luego condicionar el color). En esta simulación utilizamos la segunda opción.

```
randomPick ("R"; "R"; "R"; "R"; "V"; "V"; "V"; "V"; "V"; "V"; "V"; "V"; "V")
```

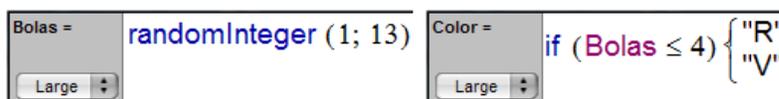


Figura 50. Taller 3-Simulación esperada 2

2. Luego se toma la medida que se desea simular, en este caso la cantidad de bolas rojas en las 150 extracciones (también puede ser la cantidad de bolas verdes).

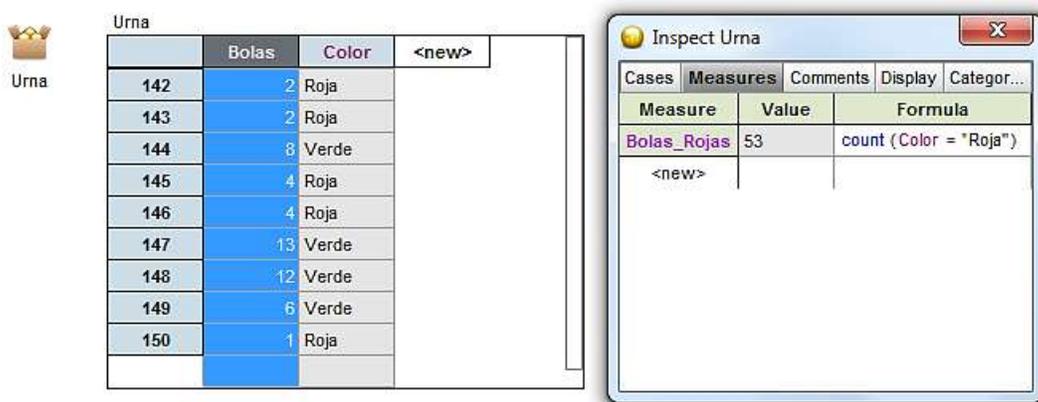


Figura 51. Taller 3-Simulación esperada 3

3. Se realiza la simulación de muchas muestras de 150 extracciones (10005 en este caso), registrando la cantidad de bolas rojas obtenidas en cada muestra.

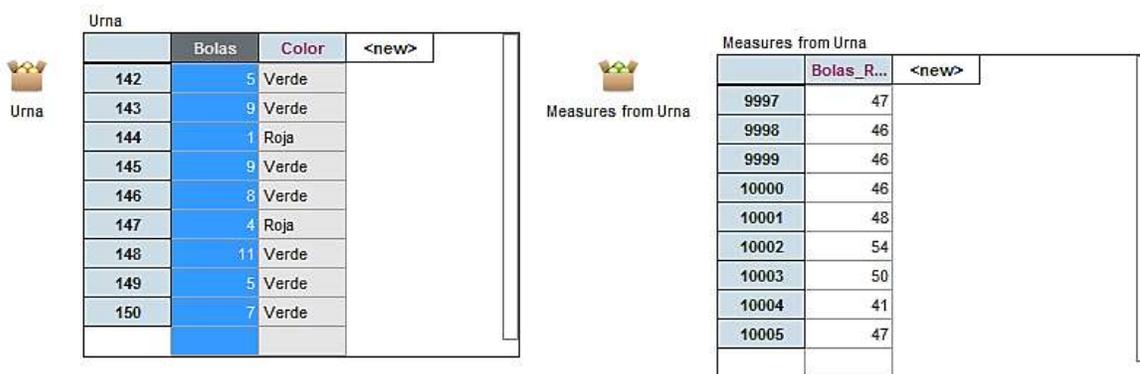


Figura 52. Taller 3-Simulación esperada 4

4. Se realiza el gráfico de la distribución de las bolas rojas.

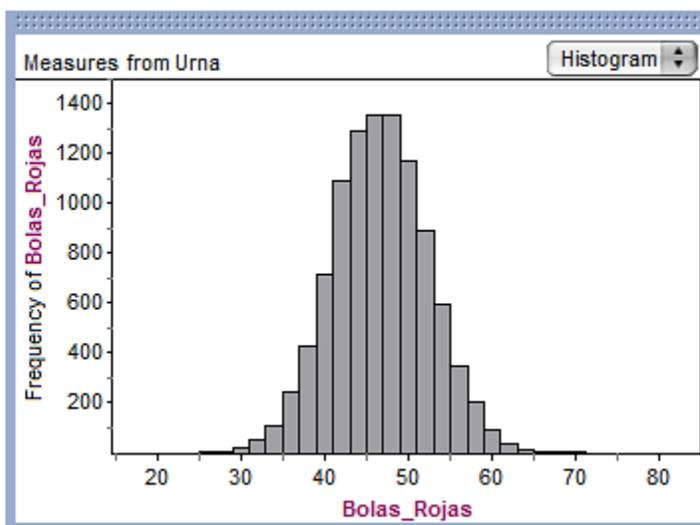


Figura 53. Taller 3-Simulación esperada 5

5. Se calcula la proporción de las muestras en las que se obtuvieron 47 o más bolas rojas.

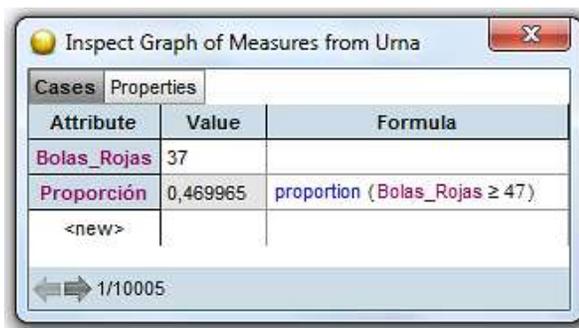


Figura 54. Taller 3-Simulación esperada 6

6. Se asimila esa proporción como probabilidad para afirmar que:
 “la probabilidad de que se obtengan 47 o más bolas rojas en una urna con 9 bolas verdes y 4 rojas es del 0.47, es decir, es muy probable lo que le ocurrió Ana, luego su estimación del número de bolas rojas y verdes no se puede contradecir con los resultados obtenidos”.

7. Para las demás preguntas el proceso de simulación es similar al de la pregunta 1; simplemente se cambia el tamaño de la muestra y se trabaja con la proporción para estimar el punto de rechazo.

Para una muestra de 75 extracciones.

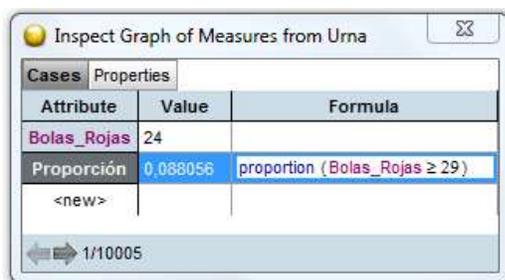


Figura 55. Taller 3-Simulación esperada 7

Para una muestra de 150 extracciones.

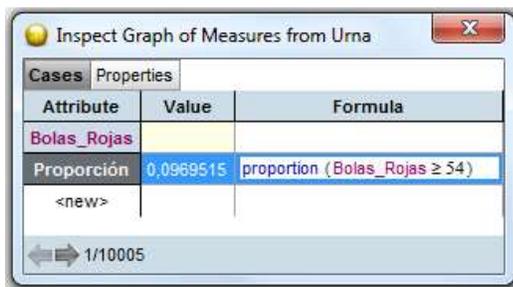


Figura 56. Taller 3-Simulación esperada 8

Para una muestra de 300 extracciones.

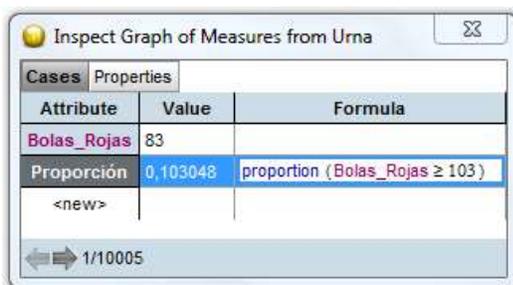


Figura 57. Taller 3-Simulación esperada 9

4.4.3. Análisis por estudiante

ANDRÉS

Pregunta 1

Antes de Experimentación

*Sí. En relación con el 150, 100 son dos tercias partes de las bolas totales; por lo tanto, las bolas verdes son dos tercias partes de las totales, si tenemos $(13/3)*2$ es 8.6 aproximamos a 9 y la resta entre 13 y 9 es 4.*

Después de Experimentación

Sí, podemos observar que 103 está en el centro de la gráfica lo que me indica que la relación podría llegar a ser 9 a 4. El hecho de que esté en el centro de la gráfica me dice que cumple con la relación propuesta.

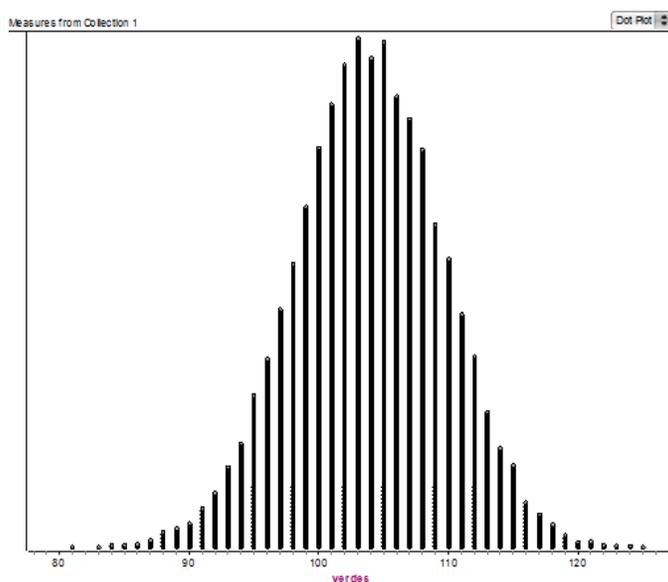


Figura 58. Evidencia Andrés Taller 3-1

Uniestructural → Multiestructural

Andrés utilizó el argumento de la proporcionalidad para hacer el contraste de la hipótesis (como lo trabajó en el taller anterior). Su argumento, que podría reflejar un proceso de estimación con componentes deterministas, es válido para la situación; lo asumió como único criterio para el contraste de la hipótesis, y no tuvo consideración alguna sobre la variabilidad muestral.

En la experimentación, Andrés hizo la simulación hasta realizar el gráfico de la distribución de las bolas (cuarto punto de la simulación esperada); para hacer el contraste de la hipótesis. Al igual que en los talleres anteriores, Andrés recurrió sólo al gráfico de la distribución de las bolas rojas para hacer su inferencia, continúa sin cuantificar con un valor de probabilidad. Vale la pena resaltar que Andrés sigue haciendo buenas inferencias a partir de la gráfica, argumentadas en la proximidad: “103 está en el centro de la gráfica”.

Pregunta 2

Antes de Experimentación

No, $150/86 = 1.8$ por lo tanto hacemos relación $13/1.8 = 7.2$ próximo a 7, por lo tanto diría que hay 7 bolas verdes y 6 rojas.

Después de Experimentación

Podemos observar que 85 está en el borde de la gráfica lo que me indica que es poco probable esa relación el hecho de que esté en el centro de la gráfica me dice que cumple con la relación propuesta; al estar por fuera me dice que no cumple con la relación.

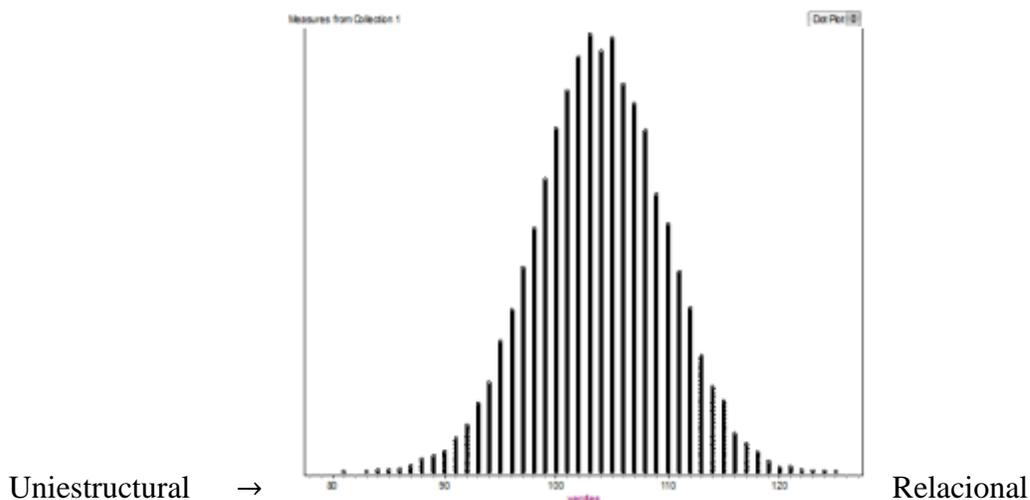


Figura 59. Evidencia Andrés Taller 3-2

Al igual que para la pregunta anterior, Andrés utilizó la proporción como único criterio para hacer el contraste de la hipótesis, dejando de lado la variabilidad muestral. De hecho, con base en la nueva información, conjetura una composición de la urna distinta a la propuesta por Ana.

En la experimentación Andrés utilizó el gráfico realizado para la primera pregunta y argumentó: *“Podemos observar que 85 está en el borde de la gráfica lo que me indica que es poco probable esa relación”*. El argumento utilizado por Andrés es más elaborado que en talleres anteriores, la adopción de la expresión *“poco probable”* evidencia desapego del determinismo y conciencia de la variabilidad (no mostrada antes de la experimentación). Como se pudo observar, Andrés divide el rango de los valores en los centrales y en los bordes, de tal forma que si los resultados caen en la zona central no se pueden rechazar, y si caen en los bordes se rechazan. No obstante que Andrés continúa sin calcular la proporción (probabilidad) se puede pensar que posee una idea intuitiva de la región de rechazo para realizar el contraste de la hipótesis. Queda pendiente la medida probabilística que le permita superar la indefinición propia que se genera cuando se habla simplemente de zona central y de zona de los bordes.

Pregunta 3

Antes de Experimentación

105, $150/105=1.43$ y $13/1.43=9.06$ como es próximo a 9 diría que es posible que 9 sea la cantidad de bolas verdes que hay

Después de Experimentación

104 es el centro de la gráfica lo que me indica que es el valor más probable para la relación

Preestructural → Preestructural

Como la pregunta iba dirigida a dar un valor, Andrés asumió el valor central como referente, contrayendo al máximo la zona central observada en la pregunta anterior. Vale resaltar que Andrés trabajó con las verdes y no con las rojas como estaba planteada la pregunta.

En el primer taller se formuló una pregunta con la misma intención y en esa ocasión Andrés sí comprendió lo que se le preguntaba, por esta razón parece que el problema de Andrés en esta pregunta pasa por una mala interpretación lectora.

Pregunta 4

Antes de Experimentación

El doble si decía 105 entonces sería 210 ya que la muestra se duplico, mi número se comporta de la misma manera

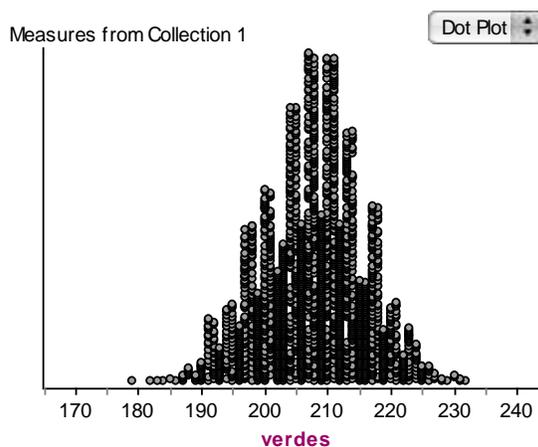


Figura 60. Evidencia Andrés Taller 3-3

Después de Experimentación

El doble si decía 104 entonces sería 208 ya que la muestra se duplico mi número se comporta de la misma manera

Preestructural → Preestructural

Andrés asume, lo que podríamos denominar, el sesgo de proporcionalidad del tamaño muestral, resultado que era de esperar ya que Andrés no traduce la información suministrada por el gráfico en términos probabilísticos.

Pregunta 5

Antes de Experimentación

De igual manera mi número se comporta igual que la muestra por lo tanto si disminuye a la mitad mi número también $105/2$ que será igual a 52.5.

Después de Experimentación

Con base a lo anterior puedo generalizar y decir que la proporción se comporta de igual manera que la muestra.

De igual manera mi número se comporta igual que la muestra por lo tanto si disminuye a la mitad mi número también $104/2$ que será igual a 52

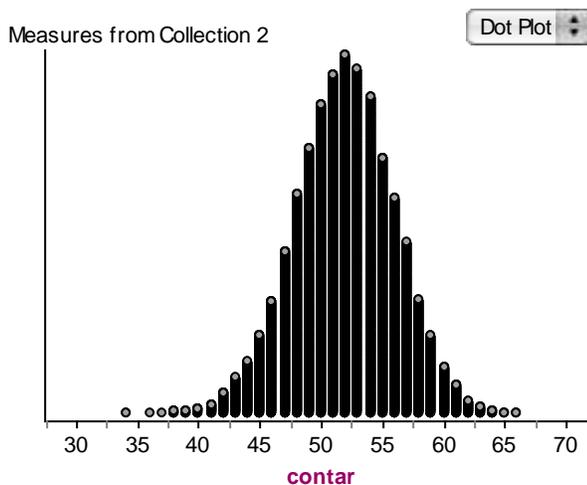


Figura 61. Evidencia Andrés Taller 3-4

Preestructural → Preestructural

Nuevamente, Andrés asumió el sesgo de proporcionalidad del tamaño muestral.

SOFÍA

Pregunta 1

Antes de Experimentación

Si, pues se acerca a la composición de la urna pues 4 es a 9 sí es proporcional a 47 es a 103.

Después de Experimentación

Si, porque la proporción 4 rojas es a 9 verdes tiene probabilidad de 45 %.

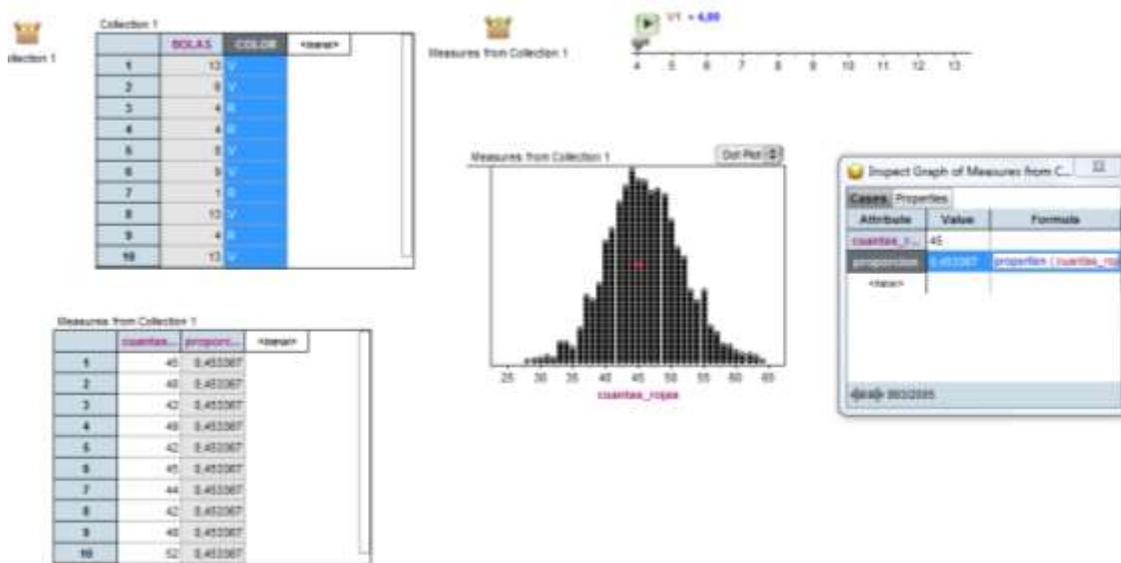


Figura 62. Evidencia Sofía Taller 3-1

Uniestructural → Relacional

Sofía utilizó antes de la experimentación, el criterio de la proporcionalidad, muy válido, ya que es el estimador generado por la muestra de extracciones. Pero al igual que Andrés, en su argumento no hay consideración alguna sobre la variabilidad muestral.

Ya después de la experimentación, Sofía, realizó toda la simulación esperada (los 6 puntos expuestos en la experimentación con *Fathom*), calculando la proporción de las bolas rojas, como se ve en la imagen continua a su respuesta. Sofía basó su argumento en encontrar 47 o más bolas rojas en muestras de 150 extracciones. La proporción encontrada fue de 0.45, la cual interpreta como evidencia suficiente para hacer la inferencia.

Pregunta 2

Antes de Experimentación

No pues la composición de la urna cambia por lo tanto su proporción cambia y ahora sería más aproximada a 6 es 8.

Después de Experimentación

No porque en un modelo de 4 rojas y 9 verdes en una muestra de 2005 experimentos no salió ninguna con esta referencia por no tanto no es cierta esta afirmación.

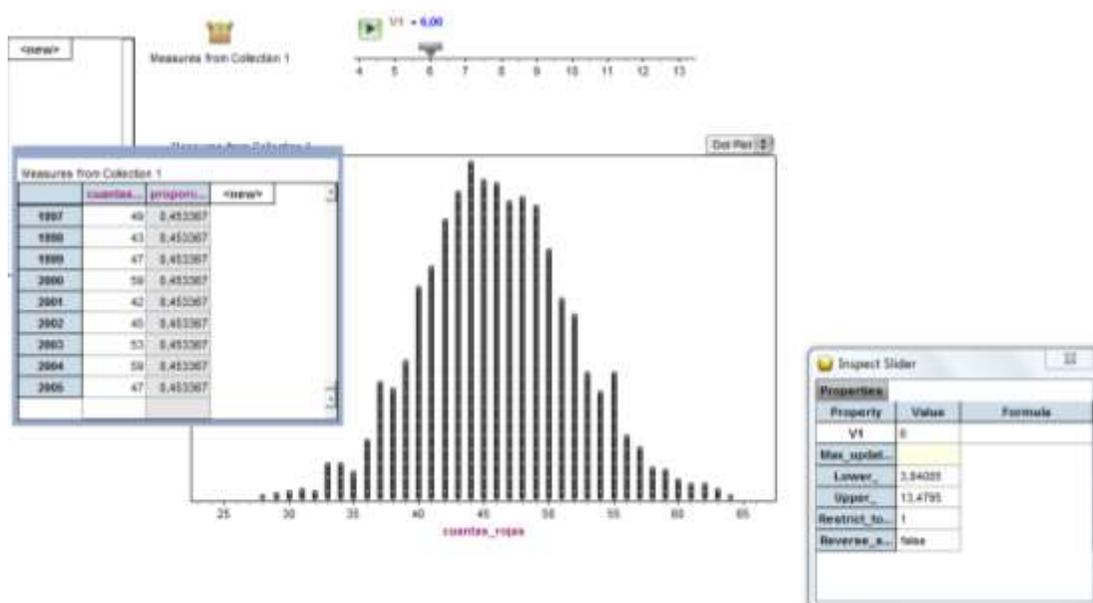


Figura 63. Evidencia Sofía Taller 3-2

Uniestructural → Multiestructural

Al igual que para la primera pregunta, el argumento de Sofía se centró únicamente en el criterio de la proporcionalidad de la muestra, no permitiendo evidenciar en su respuesta los elementos de la variabilidad.

Ya después de la experimentación, Sofía realizó la inferencia a partir de la gráfica de distribución construida en la simulación de la primera pregunta. Su argumento central es que de las 2005 muestras de tamaño 150, en ninguna aparecieron 65 o más bolas rojas. Este argumento muestra del proceso de simulación para hacer el contraste (utilizar el gráfico como criterio de inferencia). Los argumentos de Sofía, después de la experimentación, siguen sin mostrar conciencia de variabilidad, como sí lo evidenció en el taller anterior.

Pregunta 3

Antes de Experimentación

49, porque para que dé esta proporción de 4 bolas rojas y 9 verdes deben haber proporcionalmente extracciones es decir 49 extracciones rojas y 99 aproximadamente.

Después de Experimentación

Porque con 53 bolas como máximo obtengo un porcentaje de 12% lo que me evidencia que esa cantidad si pertenece a la proporción de 4 bolas rojas y 9 verdes, y porque porcentajes menores al 10 % no lo considero evidencia a favor.

Uniestructural → Relacional

Antes de la experimentación, de nuevo, Sofía utiliza el criterio de proporción como herramienta para la inferencia, sin consideraciones sobre la variabilidad.

Ya después de la experimentación, Sofía muestra nociones informales de p-valor, región de rechazo (“con 53 bolas como máximo obtengo un porcentaje de 12%”); y nivel de significancia (“porque porcentajes menores al 10 % no los considero evidencia”).

Con esta respuesta Sofía evidencia conciencia de elementos fundamentales del contraste de hipótesis, e inclusive logra una dar visos de la conexión entre ellos.

Pregunta 4

Antes de Experimentación

Deben obtenerse como máximo 90 extracciones de bolas rojas que corresponderían a 4 bolas rojas.

Después de Experimentación

Deben obtenerse como máximo 102 extracciones de bolas rojas que corresponderían al 12% que para mí ya es evidencia, y todas las extracciones que me den un porcentaje menor a 12% no es evidencia para mí.

Multiestructural → Multiestructural

En la primera parte, antes de la experimentación, Sofía, no presenta un argumento para su respuesta de 90 extracciones máximas; pero se puede notar, que no consideró las 90 extracciones como proporcionales a la muestra inicial (90 no es el doble de 49). En esta respuesta Sofía no evidencia el sesgo de proporcionalidad del tamaño muestral.

Después de la experimentación, Sofía estableció el 12% como nivel de significancia, el cuál considera como evidencia en contra para contrastar la hipótesis. La incorporación del nivel de significancia le permitió a Sofía dejar de lado criterio de proporcionalidad; pero no hace mención de no contemplar la no proporcionalidad.

Pregunta 5

Antes de Experimentación

Pues sería una proporción de 25 extracciones de bolas rojas y 45 extracciones de bolas verdes.

Después de Experimentación

Deberían obtenerse como máximo 28 extracciones de bolas rojas las cuales corresponden al 12% que ya es para mí evidencia de que los valores de ahí en adelante ya no pertenecen.

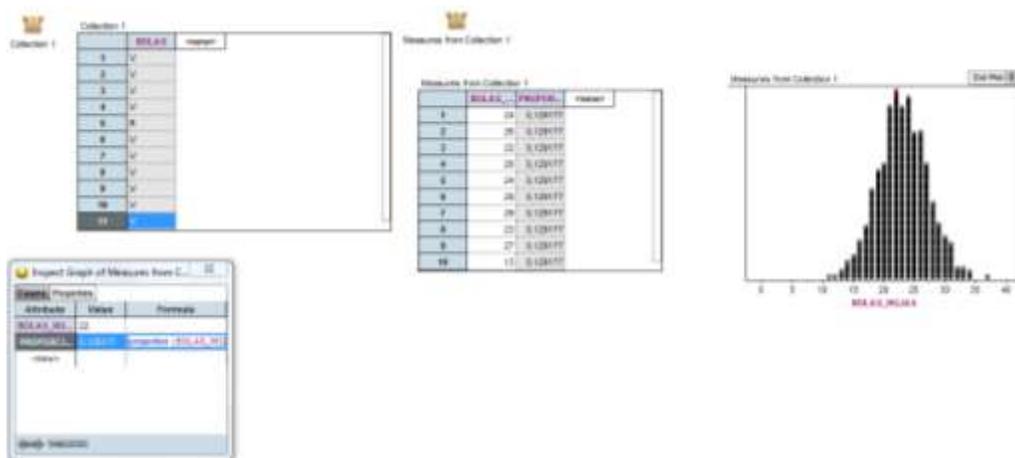


Figura 64. Evidencia Sofía Taller 3-3

Uniestructural → Relacional

Aunque en la pregunta anterior, Se presumía que Sofía no presentaba el sesgo de proporcionalidad del tamaño muestral; para esta pregunta Sofía tomó exactamente el valor proporcional al de la muestra inicial. Así, Sofía nos da a entender que no tiene conciencia aun del sesgo.

Ya después de la experimentación, Sofía evidencia dos aspectos importantes: el primero fue que ella adopta el p-valor para 28 extracciones de bolas rojas, dejando de lado el supuesto de proporcionalidad que exponía antes de la experimentación. El segundo es que confirma su nivel de significancia del 12%.

GABRIELA

Pregunta 1

Antes de Experimentación

Si, ya que la cantidad de bolas verdes debe de ser mayor a las rojas, además entre las 150 extracciones, la cantidad de bolas verdes fue el doble de rojas, entonces por lo tanto se podría decir que esto puede servir al momento de dar la cantidad de cada color en la urna.

Después de Experimentación

Si, debido a las 2000 muestras que hicimos en donde dentro de la urna había 4 balotas rojas y 9 verdes, la probabilidad de que salgan 47 balotas rojas, de 150 extracciones es de un 46,7%, por lo tanto, se puede afirmar lo que dijo Ana.

Multiestructural → Multiestructural

Gabriela presentó una respuesta más elaborada que la de sus compañeros antes de la experimentación. Aunque en de su argumento se encuentra el criterio de proporcionalidad, en la primera parte menciona “*la cantidad de bolas verdes debe de ser mayor a las rojas*” evidenciando apropiación de lo que se trabajó en talleres anteriores: las extracciones como representantes de la composición de la urna.

Ya después de la experimentación, aunque no presentó imágenes de evidencia como sus compañeros (a pesar de advertirle que las imágenes también hacían parte del argumento), Gabriela evidencia en su respuesta, comprensión del proceso de simulación para hacer el contraste de hipótesis. El argumento de su respuesta se centró en describir qué se hizo en la simulación y cómo la interpretó. Gabriela obtuvo una probabilidad de 46.7% en su simulación, la cual ella cataloga como evidencia suficiente para estar de acuerdo con Ana.

Pregunta 2

Antes de Experimentación

No, porque la cantidad de bolas verdes y rojas son cercanas, entonces la cantidad de bolas de cada color que tiene la urna deben ser valores cercanos.

Después de Experimentación

No, ya que, si la urna tuviera 4 bolas rojas y 9 verdes, la probabilidad de obtener 65 bolas rojas, de 150 extracciones es muy mínima, como se puede observar en la gráfica ese valor está alejado.

Multiestructural → Multiestructural.

De nuevo, las extracciones como representantes de la composición de la urna es el argumento fundamental de Gabriela para hacer su inferencia antes de la experimentación. Para ella, como las extracciones de bolas rojas y verdes son muy cercanas, entonces en la urna, la cantidad de bolas rojas y verdes debe ser muy parecida. Esta es la primera respuesta hasta el momento que antes de la experimentación no utiliza el criterio de proporcionalidad como argumento de la respuesta.

Después de la experimentación, la respuesta de Gabriela se centra en el análisis del gráfico de la primera simulación, pero no presenta el gráfico que analizó.

Ahora bien, conociendo la simulación y el gráfico que generaba la simulación, el argumento de Gabriela es válido pues como mostraron las gráficas de sus compañeros 65 extracciones de bolas rojas en un resultado con muy poca probabilidad.

Pregunta 3***Antes de Experimentación***

Yo aceptaría como máximo 56 bolas rojas, porque Ana dice hay cuatro bolas rojas, entonces la cantidad de bolas que aceptaría como máximo sería esa, porque si acepto una cantidad más grande, quiere decir que hay mayor cantidad de bolas rojas.

Después de Experimentación

Aceptaría como máximo 56 bolas rojas, ya que para mí aceptaría un 5,1% de probabilidad, esto quiere decir que sería como un 45% probable de que si me salen 56 bolas rojas o menos de 150 extracciones puedo afirmar lo que dice Ana.

Uniestructural → Relacional

Gabriela da un valor para el extremo de la región de rechazo considerablemente bueno antes de la experimentación, pero su argumento es limitado, no presenta explicación para la elección de este valor.

Ya después de la experimentación, Gabriela permite evidenciar que estableció un nivel de significancia alrededor del 5%. Con la siguiente oración: “*sería como un 45% probable de que si me salen 56 bolas rojas o menos de 150 extracciones puedo afirmar lo que dice Ana*”, Gabriela evidencia intuición de región de aceptación y región de rechazo.

Pregunta 4

Antes de Experimentación

Aceptaría el doble de la cantidad de bolas que acepte en donde las extracciones son el doble que la anterior por lo tanto aceptaría 112.

Después de Experimentación

Aceptaría como máximo 106 bolas rojas, ya que como he dicho aceptaría un 5% como en este caso, por lo tanto, si la muestra es de 300 extracciones, aceptaría ese valor, también porque como las muestras no son proporcionales, yo no podría decir que como en el caso de 150 extracciones acepté 56, en este aceptaría el doble, pero vemos que 106 es un valor cercano a 112.

Uniestructural → Relacional

Gabriela antes de la experimentación evidencia el sesgo de proporcionalidad del tamaño muestral. Inclusive lo hace evidente al decir que tomaría “*el doble*” de las extracciones anteriores.

Después de la experimentación, Gabriela no sólo reafirma el nivel de significancia interiorizado en la pregunta anterior, dando un valor de 106 bolas rojas (5%). Sino que también, se percata de la imposibilidad de utilizar el criterio de proporcionalidad e intenta explicar la no proporcionalidad presente en la situación.

Pregunta 5***Antes de Experimentación***

Aceptaría la mitad de la cantidad de bolas que acepte en la pregunta 3.

Después de Experimentación

Aceptaría 30 bolas rojas como máximo, ya que de acuerdo a mi criterio acepto un 5,6%; ya que en una muestra de 75 extracciones este valor es confiable, para poder afirmar la cantidad de color de cada urna.

Uniestructural → Relacional

Al igual que para la pregunta anterior, Gabriela incurrió al sesgo de proporcionalidad del tamaño muestral.

Ya después de la experimentación, Gabriela adopta el nivel de significancia como un criterio para definir la región de rechazo y lo aplica para definir la cantidad de bolas rojas que aceptaría.

Finalmente se puede percibir que Gabriela superó el sesgo de proporcionalidad del tamaño muestral, que era el objetivo principal del taller.

LUIS

Pregunta 1

Antes de Experimentación

Sí, porque en total tenemos 13 balotas y para una muestra de 150 extracciones con 47 balotas rojas y 103 balotas verdes somos capaces de intuir que hay más balotas verdes que rojas, y 4 rojas y 9 verdes luce como una proporción razonable.

Después de Experimentación

Sí, porque desde el momento en el que extraemos la medida de la tabla en la que tenemos una muestra con 150 extracciones con 4 bolas rojas obtengo 46 bolas rojas, el cual ya es un número bastante cercano al de la extracción. Después colectamos 2000 mil muestras más y obtenemos la gráfica a partir de ellas la cual en la inspección nos deja ver que la proporción para un número menor o igual a 47 es de aproximadamente el 48% lo me hace concluir que la afirmación de Ana es certera.

Uniestructural → Multiestructural

Luis al igual que Gabriela utilizó el argumento de las extracciones como representante de la composición de la urna.

Después de la simulación, Luis presenta la misma dificultad de Gabriela, no presentó las gráficas como parte de su argumento. Respecto a su respuesta, Luis evidencia comprensión del proceso de simulación para hacer el contraste de hipótesis, pero en su inferencia, da una respuesta que no da cabida a la variabilidad y al error “*me hace concluir que la afirmación de Ana es certera*”.

Pregunta 2

Antes de Experimentación

No, en este caso sigue habiendo una preferencia por las balotas verdes, pero no es tan alta como para concluir que hay 5 balotas verdes más que la cantidad de balotas rojas.

Después de Experimentación

No. Porque de las 2000 muestras obtenidas la probabilidad de que el valor de bolas rojas sea 65 es prácticamente nula, respetando su mínimo margen de error. En la gráfica el 65 no está representado por una cantidad de bolas en la distribución de 4 rojas y 9 verdes. Así mismo podemos ver que el número 65 es bastante lejano al 47 lo cual nos hacen concluir que no deben tener la misma distribución.

Uniestructural → Relacional

Al igual que para la pregunta anterior, Luis utilizó las extracciones como represente la composición de la urna como argumento para su respuesta.

El argumento de Luis después de la experimentación, es muy completo, pues tiene en cuenta todos los factores que intervienen: dice que la probabilidad que salgan 65 bolas rojas es prácticamente nula (presumiendo el análisis del gráfico de la simulación), pero entiende que aunque este valor es determinante en su simulación, es un valor variable y lo fundamenta con el argumento “*respetando su mínimo margen de error*”. Además, utiliza la validación visual al argumentar que “*65 es bastante lejano al 47*”.

Pregunta 3

Antes de Experimentación

Máximo 52 balotas rojas. Luce como un número en el que hay suficiente superioridad de las balotas verdes para que haya una cantidad de 4 balotas rojas y 9 verdes

Después de Experimentación

Para estar de acuerdo con la distribución de Ana no aceptaría un número mayor a 54. Ese sería el valor máximo de bolas rojas que deberíamos obtener porque la proporción para valores mayores es el 10% y por ejemplo con 53 es del 14%.

Uniestructural → Relacional

La respuesta de Luis está ligada a lo trabajado en los anteriores talleres; Luis ya es consciente que el número de bolas rojas no puede ser muy alejado del ideal del modelo sobre el que está trabajando. Por esta razón, Luis elige 52 bolas rojas antes de la experimentación y lo argumenta con “*Luce como un número en el que hay suficiente superioridad de las balotas verdes*”.

Posterior a la de la experimentación, Luis muestra que experimenta con diferentes opciones para las bolas rojas y evalúa su probabilidad; adoptando un valor del 10% para su nivel de significancia personal. Razón por la cual acepta hasta 54 bolas rojas.

Pregunta 4

Antes de Experimentación

Exactamente el doble que en la anterior tabla. Es decir 104 balotas rojas ya que si la muestra aumenta en su número de extracciones también aumenta en el número extracciones que obtenemos de cada color. Pero guardando sus proporciones, ya que el número de bolas siempre es el mismo.

Después de Experimentación

El valor máximo que de bolas rojas sería 102 bolas rojas y 198 verdes. Con una proporción de que el valor mayor a las 102 bolas rojas del 10% tendría evidencia a favor de que la distribución sea de 4 bolas rojas y 9 bolas verdes.

Uniestructural → Relacional

Luis evidencia el sesgo de proporcionalidad del tamaño muestral, de la forma evidente a mencionar “*Exactamente el doble*”.

Después de la experimentación, Luis toma conciencia del impedimento del uso del criterio de proporcionalidad; y continúa manteniendo un nivel de significancia del 10%.

Pregunta 5

Antes de Experimentación

La mitad de las extracciones del primer caso, es decir, 26 balotas rojas, ya que esta vez la cantidad de extracciones totales se reduce y por ende el número de extracciones de bolas de cada color también lo hace, pero otra vez guardando las mismas proporciones. Debido a que el número de bolas no cambian.

Después de Experimentación

Aunque la muestra y su cantidad de extracciones no disminuye o aumenta proporcionalmente en este caso el valor máximo de bolas rojas que aceptaría serían 28 bolas rojas para una distribución de 4 bolas rojas y 9 bolas verdes. Tengo en la proporción

de la gráfica un 13% a favor para esta cantidad de bolas rojas extraídas y para 29 ya disminuye por debajo del 10%.

Uniestructural → Relacional

De nuevo, Luis evidencia el sesgo de proporcionalidad del tamaño muestral, que se quería trabajar con este taller.

Después de la experimentación, Luis al igual que Gabriela, hace mención la imposibilidad del criterio de proporcionalidad para la situación “*la muestra y su cantidad de extracciones no disminuye o aumenta proporcionalmente*”; con el fin de aclarar por qué su elección coincidentalmente corresponde a la mitad de la muestra inicial.

Finalmente se puede ver que Luis ya interiorizó para sí un nivel de significancia alrededor 10%.

4.4.4. Conclusiones

Antes de la experimentación, para la primera y segunda pregunta, el criterio de proporcionalidad y las extracciones como representantes de la composición de la urna, fueron los dos argumentos utilizados por los estudiantes para realizar el contraste de hipótesis. Después de la experimentación, la ayuda de la simulación les proporcionó a los estudiantes las herramientas necesarias (gráficas y probabilidades) para fundamentar mejor sus inferencias.

El taller permitió a los estudiantes constituir un nivel de significancia personal. Además, se percibieron nociones intuitivas de p-valor y regiones de rechazo/aceptación.

Respecto al sesgo de proporcionalidad del tamaño muestral, fue evidente que los cuatro estudiantes presentaron esta dificultad antes de la experimentación, pero al ver que la simulación contradecía la proporcionalidad, notaron su error y encontraron en su nivel de significancia personal y el p-valor informal, nuevos argumentos para justificar sus respuestas.

Finalmente, se advierte la necesidad de un trabajo previo con los estudiantes sobre la interpretación de las distribuciones muestrales, en particular, la conversión de lo gráfico a lo cuantitativo y viceversa; porque aunque se han percibido avances, algunos estudiantes evidencian dificultades en dicho aspecto.

4.5. Taller 4

4.5.1. Análisis de Componentes

SITUACIÓN (Distribución Continua)

Según el DANE el promedio de vida en el año 2015 para una mujer en el departamento de Santander fue de 78,5 años, mientras que para un hombre es de 72,1 años. Con el fin de saber si el promedio de vida ha cambiado para este año para las mujeres, el DANE ha tomado una muestra de 100 mujeres fallecidas, y ha calculado su promedio de vida: 80,8 años.

Pregunta 1

Con base en la muestra, ¿Cree usted que el promedio de vida hoy en día es mayor a 78,5 años para las mujeres santandereanas? Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis.
2. Principal: Inferencia.
3. Secundario: Variabilidad Muestral.

Pregunta 2

Una muestra aleatoria de 150 muertes de hombres registradas este año en el departamento muestra una vida promedio de 60,2 años. ¿Cree usted que el promedio de vida hoy en día es menor a 72,1 años para los hombres santandereanos? Explique claramente su respuesta.

❖ Componentes conceptuales

1. Principal: Contraste de la hipótesis.
2. Principal: Inferencia.
3. Secundario: Variabilidad Muestral.

Pregunta 3

¿Cuál sería el promedio de vida mínimo que usted aceptaría, para poder afirmar que el promedio de vida no ha cambiado este año para los hombres? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: p-valor informal. Principal: Región de rechazo.

Pregunta 4

Se tienen las bases de datos de las muertes de hombres y mujeres en el año 2015 para el departamento de Santander. Si se extrae una muestra de 250 hombres de esta base de datos ¿Cuál cree que sería la media de vida de los hombres de esta muestra? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Muestra con representante de la población.
2. Principal: Variabilidad Muestral.

Pregunta 5

Si se extrae una muestra de 500 hombres ¿Cuál cree que será la media de vida de los hombres de esta muestra? Justifique su respuesta.

❖ Componentes conceptuales

1. Principal: Muestra con representante de la población.
2. Principal: Influencia del tamaño muestral.
3. Principal: Variabilidad Muestral.

4.5.2. Experimentación *Fathom*

La simulación con *Fathom* que se espera que realicen los estudiantes para la situación es la siguiente:

1. Esta simulación, a diferencia de las otras, trae una base de datos, que contiene las edades de los hombres y las mujeres que murieron durante el 2015.



Figura 65. Taller 4-Simulación esperada 1

2. Se crea la tabla para la base de datos (por defecto trae una muestra de 10 datos)

	Hombres	Mujeres	<new>
2	69,2413	72,944	
3	63,8845	81,6006	
4	71,3508	76,5415	
5	75,8124	77,5393	
6	70,4524	60,2409	
7	67,7267	91,9522	
8	84,3386	79,5411	
9	71,0248	83,3916	
10	73,8894	83,6056	

Figura 66. Taller 4-Simulación esperada 2

3. En el inspector de la muestra (base de datos) se cambian los siguiente valores

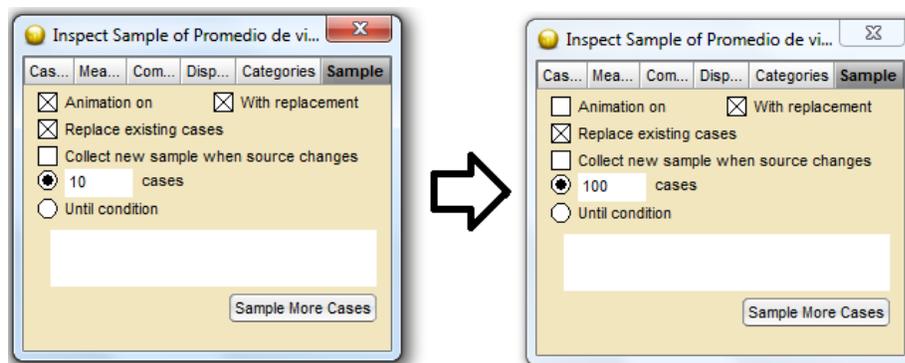


Figura 67. Taller 4-Simulación esperada 3

Se da clic en *Sample More Cases* y en la tabla aparecerán una muestra de 100 edades de defunción (tanto para hombres como mujeres) durante el año 2015.

	Hombres	Mujeres	<new>
92	72,4976	59,1371	
93	74,1199	67,3159	
94	65,2108	69,4397	
95	81,178	79,3841	
96	70,2489	82,7643	
97	62,8665	91,2436	
98	74,1892	77,2872	
99	57,9905	71,6154	
100	76,6087	78,1499	

Figura 68. Taller 4-Simulación esperada 4

- De nuevo en el inspector de la muestra, se toma la medida que se desea simular, en este caso la edad promedio de muerte de las 100 mujeres de la muestra.

Measure	Value	Formula
Promedio	80,0809	mean (Mujeres)
<new>		

Figura 69. Taller 4-Simulación esperada 5

- Se realiza la simulación de muchas muestras de tamaño 100 (6005 en este caso), registrando el promedio de edad de muerte de las mujeres.

Measures from Sample of Promedio de vida	
	Promedio <new>
5997	79,6759
5998	79,8223
5999	78,5448
6000	78,7507
6001	78,3926
6002	78,2017
6003	78,9125
6004	79,5046
6005	77,6142

Figura 70. Taller 4-Simulación esperada 6

6. Se realiza el gráfico de la distribución de los promedios.

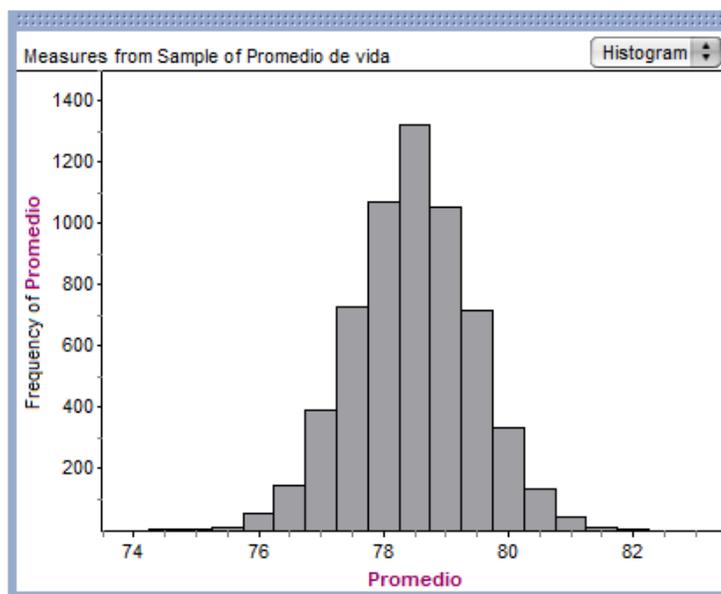


Figura 71. Taller 4-Simulación esperada 7

7. Se calcula la probabilidad de las muestras en las que se obtuvieron un promedio de 80.8 años o más

Attribute	Value	Formula
Promedio	78,486	
Probabilidad	0,007826...	proportion (Promedio ≥ 80,8)
<new>		

Figura 72. Taller 4-Simulación esperada 8

“la probabilidad de que el promedio de muerte sea de 80.8 años o superior es de 0.007, es decir, es muy probable que el promedio de vida hoy en día es mayor a 78.5 años para las mujeres santandereanas”

- Para la pregunta 2 y 3 se realiza la misma simulación cambiando el tamaño de la muestra a 150 y trabajando con la variable Hombres.

4.5.3. Análisis por estudiante

ANDRÉS

Pregunta 1

Antes de Experimentación

Sí, aunque vemos una variación relativamente pequeña podemos observar un incremento esto me da a pensar que hay probabilidad de que el promedio de muertes para mujeres en 2016 sea mayor que para el 2015.

Después de Experimentación

Si, tomando en cuenta la gráfica el valor de 80.5 tiene una probabilidad del 2% de que pertenezca a el promedio por lo tanto podría decir es posible que el promedio de muerte haya aumentado

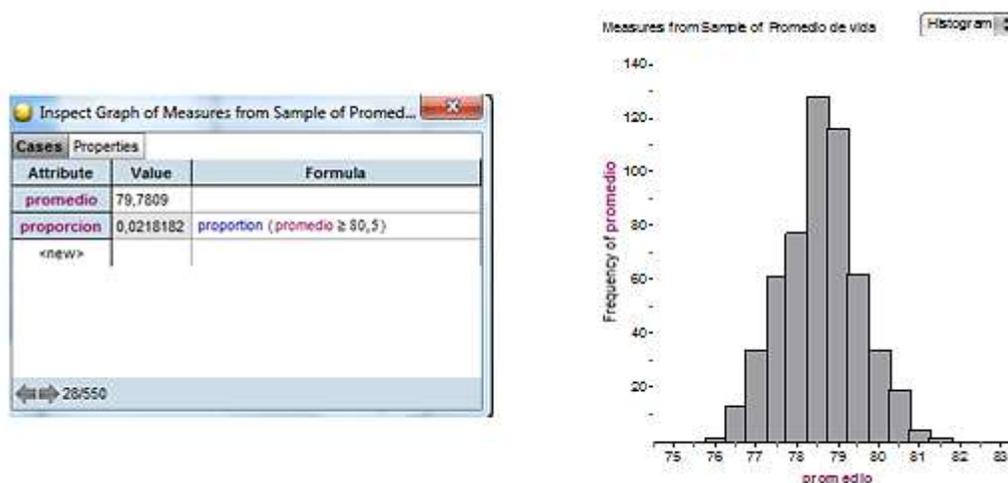


Figura 73. Evidencia Andrés Taller 4-1

Multiestructural → Relacional

A través de los talleres, los argumentos de Andrés han ido mejorando, utilizando un lenguaje más claro. Andrés antes de la experimentación ya es consciente de la variación y además, el argumento *“hay probabilidad de que el promedio de muertes para mujeres en 2016 sea mayor”*, evidencia que Andrés dejó de lado el determinismo matemático como argumento de respuesta.

Ya después de la experimentación, Andrés, qué generalmente basaba su inferencia en el gráfico, esta vez, recurrió al cálculo de la probabilidad para realizar su inferencia. Esta es la primera vez que Andrés, ve la necesidad de cuantificar para poder realizar la inferencia. Al parecer, Andrés adoptó el nivel de significancia más bajo de todos los compañeros.

Preguntas 2

Antes de Experimentación

Sí. Podemos observar una variación decreciente en el promedio de vida de los hombres en Santander lo que me lleva a pensar que cabe la posibilidad de que este promedio haya disminuido en comparación 2015, 2016

Después de Experimentación

Sí, como podemos observar en la gráfica de promedios el 60.2 no registra, lo que me permite sacar la conclusión de que la probabilidad de que haya disminuido el promedio de vida sea muy alta.

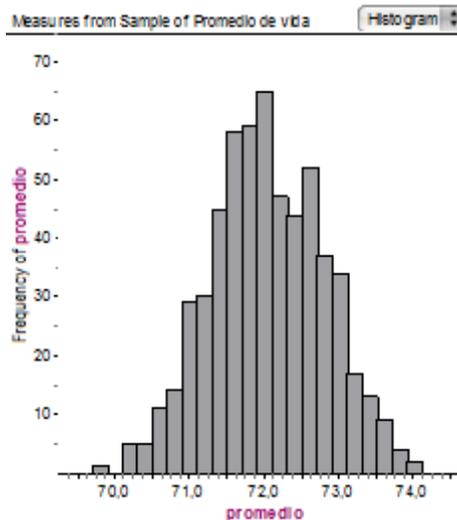


Figura 74. Evidencia Andrés Taller 4-2

Uniestructural → Multiestructural

Al igual que para la pregunta anterior, Andrés dejó de lado el determinismo matemático y habla de posibilidades.

Después de la experimentación, Andrés hizo la gráfica de promedios de vida de los hombres a partir de la extracción de muestras de 150 hombres de la base de datos que se le suministró. Como 60.2 es un valor que no está presente en el gráfico, Andrés infiere que la probabilidad de que haya disminuido es muy alta.

Pregunta 3

Antes de Experimentación

Para una muestra de 150 personas diría que podría oscilar entre 69.1 y 75.1 esto ya que son valores relativamente próximos a el promedio en 2015 que era 72.1

Después de Experimentación

De 71 hasta 73, ya que son los valores en los cuales la gráfica empieza a cambiar drásticamente pero aun así hacen parte del centro de la misma

Multiestructural → Relacional

Aunque la situación se planteó para la cola inferior, Andrés, antes de la experimentación trabajó el problema a dos colas, lo que le permitió construir un intervalo de confianza informal centrado en la media.

De nuevo, después de la experimentación, Andrés utiliza el gráfico como único argumento para su inferencia. Aunque, esta vez, Andrés menciona cuál es su criterio de rechazo para el gráfico:

“la gráfica empieza a cambiar drásticamente pero aun así hacen parte del centro de la misma”; en otras palabras, Andrés se fija en la dispersión que presentan los datos para realizar su inferencia.

Pregunta 4

72 años, ya que lo más probable es que cualquier muestra que tomemos si sacamos el promedio de vida será próximo a el promedio de vida general (72.1)

Relacional

Andrés evidencia progreso en sus concepciones y sus argumentos. Mientras que en la prueba inicial, Andrés, daba muestras del sesgo de los pequeños números (Kahneman y Tversky, 1983), ahora parece haberlo superado. En su argumento dejó de lado el determinismo; con la expresión *“ya que lo más probable”* da a entender la existencia de muestras alejadas pero poco probables.

Pregunta 5

Antes de Experimentación

72.1, entre más grande sea la muestra más exactitud podremos tener a la hora de analizar los datos por ende el valor será más próximo a el promedio de vida de los hombres.

Relacional

Andrés cumple con las componentes propuestas para esta pregunta, identificó la muestra como representante de la población, la influencia del tamaño de la muestra y la variabilidad muestral.

SOFÍA

Pregunta 1

Antes de Experimentación

Si, por que la muestra de 100 muertes nos arroja un promedio de vida de 80.8 años por lo tanto podemos deducir que el promedio de vida este año si es mayor a 78,5.

Después de Experimentación

Si ha aumentado pues la probabilidad de que un promedio de 80,8 salga en las muestras es de un 6% por lo tanto esto me asegura que el promedio si ha aumentado con respecto al promedio inicial de 78,5

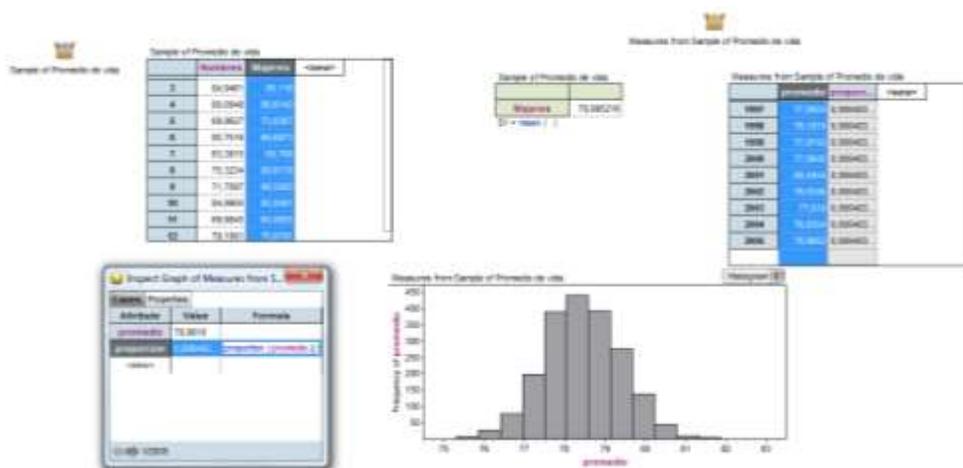


Figura 75. Evidencia Sofía Taller 4-1

Uniestructural → Multiestructural

Antes de la experimentación, aunque la respuesta es correcta, el argumento de Sofía es limitado, no presenta ningún argumento que sustente su respuesta.

Después de la experimentación, Sofía realizó el remuestreo necesario para realizar la simulación, y encontró que la probabilidad de encontrar valores tan extraños como 80.8 es muy baja. No obstante, la probabilidad de su simulación, como se ve en la imagen es 0.006 correspondiente a un 0.6% no a un 6% como ella indicó. Además, Sofía utiliza la expresión “*esto me asegura que*”, ignorando el error presente en la prueba y optando por un determino.

Preguntas 2

Antes de Experimentación

Si, por que el promedio de la muestra es menor a 72,1 por lo tanto el promedio total de los santandereanos si es menor porque con que sepamos en promedio de una muestra de esta población ya podemos acérmanos al promedio general, y en este caso sabemos que el promedio de la muestra es 60,2 por lo tanto si es menor que el promedio tomado en el año 2015

Después de Experimentación

Si disminuyo pues en este año la muestra no corresponde a un promedio de 72,1 pues en el año 2015 si estaban cerca ese promedio pero este año se sale y disminuye mucho

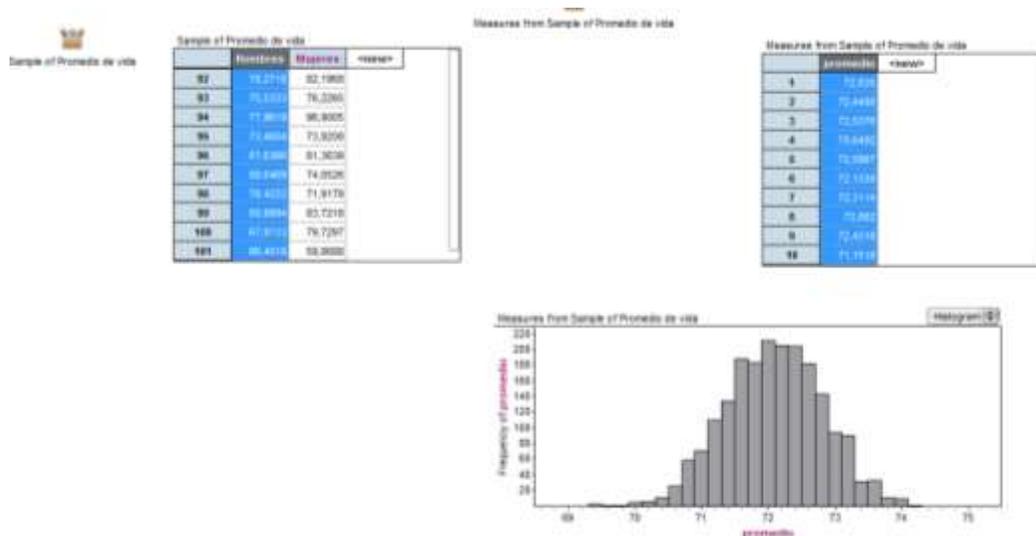


Figura 76. Evidencia Sofía Taller 4-2

Multiestructural → Multiestructural

Antes de la experimentación, Sofía utiliza el criterio de la muestra como representante de la población para realizar la inferencia. En su argumentación se percibe que Sofía podría estar pensando que sí el promedio de muestra es menor significa que pertenece a una distribución distinta a la que se está analizando. Este argumento es una de las bases de la prueba de hipótesis, pero no se puede asegurar que Sofía maneja esta noción, porque no hace referencia a proximidad a la media, simplemente resalta el hecho de ser menor.

Después de la experimentación, Sofía utiliza el gráfico de promedios de muestras como único argumento para su inferencia. Al igual que Andrés, identificó que el valor de 60.2 es un valor demasiado extraño “*este año se sale y disminuye mucho*” que ni siquiera hace parte del gráfico.

Pregunta 3

Antes de Experimentación

Aceptaría un promedio de 70,0 años esto para mí es como si el promedio no ha cambiado casi pues no se aleja mucho del promedio del año anterior que fue 72,1

Después de Experimentación

Yo acepto el promedio mínimo de 71 años de vida para los hombres

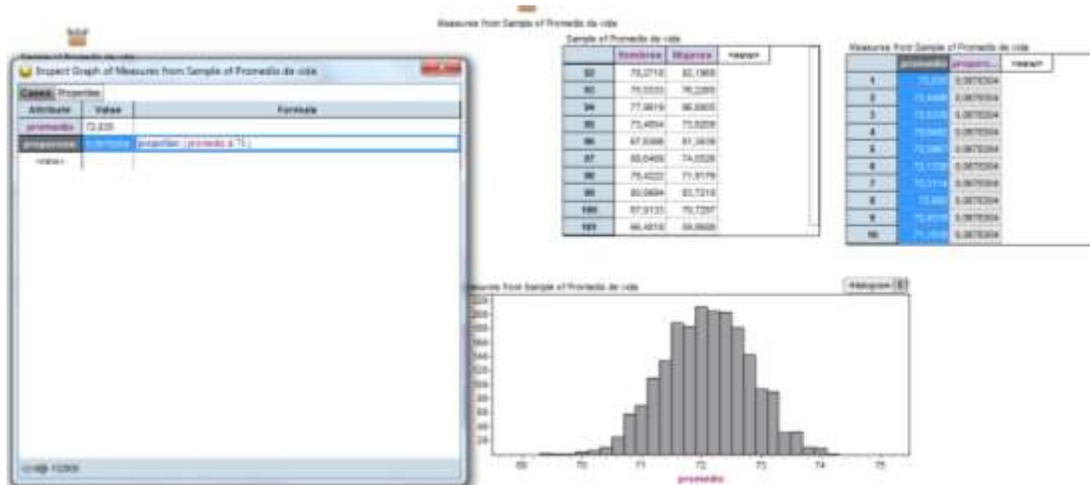


Figura 77. Evidencia Sofía Taller 4-3

Uniestructural → Uniestructural

Antes de la experimentación, Sofía utiliza realiza una buena inferencia a partir del argumento de la proximidad. Aunque en el enunciado no se mencionó la desviación, Sofía intuyó, a través de la proximidad que el p-valor informal considerablemente bueno.

Después de la experimentación, aunque Sofía no presentó ningún argumento para su respuesta, no obstante, revisando su simulación se puede ver que en el inspector de la gráfica, Sofía calculó que probabilidad de encontrar valores menores o iguales a 71 es de 0.06. Y recordando que el nivel de significancia adoptado en su taller anterior fue de 10%, de ahí pudo darse su respuesta.

Pregunta 4

Pues no muy diferente al promedio de la población pues aunque el tamaño de la muestra cambie no significa que el promedio tenga que cambiar mucho.

Uniestructural

La respuesta de Sofía es muestra evidencia de identificar la muestra como representante de la población, un aspecto que no evidenciaba en la prueba inicial.

Pregunta 5

Pues sería aproximadamente al promedio primero pues aunque la muestra cambie de valor y sea un valor muy mayor o menor su promedio no cambiara ni variaría mucho pues es la misma población y con una muestra mínima de 100 es suficiente para calcular el promedio de la población por lo tanto la media de vida sería aproximadamente 70 años

Uniestructural

Esta respuesta revalida el análisis de la pregunta anterior, Sofía identifica la muestra como representante de la población. No obstante, no evidencia en su argumento comprender la influencia de aumentar o disminuir el tamaño muestral.

GABRIELA**Pregunta 1**

Antes de Experimentación

Aunque 100 mujeres es un valor pequeño respecto a toda la población de mujeres, esto puede ser un valor representativo por lo tanto creo que el promedio de vida hoy en día es mayor al del año anterior.

Después de Experimentación

Como podemos observar este es un valor extraño, porque está alejado del promedio de la población, por lo tanto podemos decir que el promedio de vida hoy en día es mayor a 78,5.

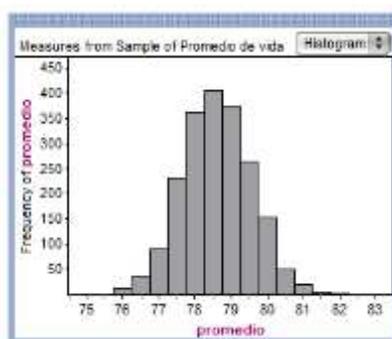


Figura 78. Evidencia Gabriela Taller 4-1

Multiestructural → Multiestructural

Gabriela antes de la experimentación, evidencia haber superado la dificultad del tamaño muestral que presentó en la prueba inicial; consideraba la muestra muy pequeña para hacer una inferencia. Esta vez, Gabriela, menciona que aunque la muestra es pequeña, es representativa para realizar una inferencia. Respecto a la inferencia podemos ver que Gabriela, al igual que Sofía, utiliza la muestra como representante de la población como argumento para su inferencia.

Ya después de la experimentación, Gabriela utiliza el gráfico como argumento para realizar la inferencia, recurriendo al criterio de la proximidad y dejando de lado alguna cuantificación para la inferencia.

Preguntas 2

Antes de Experimentación

Esta es una muestra más grande, pero podría aceptar un valor inferior al del año anterior 72,1, pero no decir que el promedio de vida está debajo de 60,2

Después de Experimentación

Si, ya que este valor está muy lejos del promedio de vida, en la gráfica se puede apreciar que este valor, ni siquiera aparece, por lo tanto, es un valor extraño y eso es suficiente para decir que el promedio de vida hoy en día es menor que 72,1.

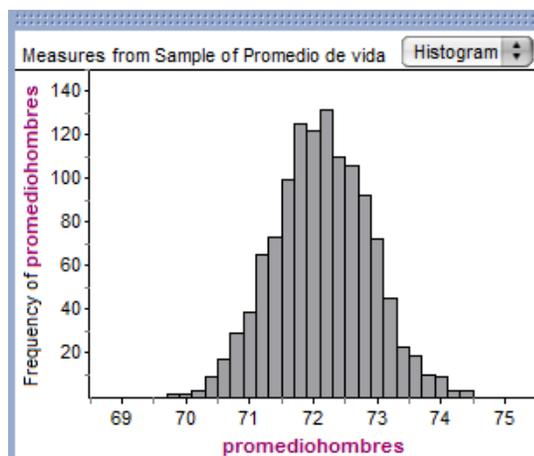


Figura 79. Evidencia Gabriela Taller 4-2

Uniestructural → Multiestructural

Antes de la experimentación, parece que Gabriela aun continua se cierta forma resistiva a que el tamaño de la muestra es muy pequeño para hacer una inferencia. Pero aun así, realiza una buena inferencia.

Ya después de la experimentación, Gabriela, al igual que en la primera pregunta, utiliza el gráfico como argumento para sustentar su respuesta. El valor “*ni siquiera aparece, por lo tanto, es un valor extraño*”, evidencia que Gabriela comprensión de la inferencia a partir del gráfico.

Pregunta 3

Antes de Experimentación

Aceptaría el promedio de vida mínimo seria de un 70 ya que no es un valor tan lejano al promedio del año pasado.

Después de Experimentación

Aceptaría como mínimo 71 años, ya que de acuerdo a mi criterio acepto en este caso un 6%, teniendo en cuenta que el promedio de vida es de 72,1 este es un valor confiable.

Multiestructural → Multiestructural

Antes de la experimentación, Gabriela utiliza el criterio de la proximidad como argumento para asignar el p-valor informal.

Ya después de la experimentación, Gabriela aunque no muestra evidencia de la simulación, calculó en mismo 6% que encontró Sofía. En el taller anterior Gabriela había adoptado un nivel de confianza del alrededor del 5%, por esta razón es que 70 resulta ser un valor confiable para ella.

Pregunta 4

Sería la misma, ya que sea una muestra más grande o más pequeño, es representativa, por lo tanto, podría ser la misma o variar, pero muy poco, ya que el tamaño de la muestra no interfiere.

Multiestructural

Gabriela identifica la muestra como representante de la población y nociones de variabilidad muestral. Si retomamos la prueba inicial, podemos percibir el cambio de pensamiento de Gabriela, puesto que inicialmente admitía que la media de cualquier muestra debía ser igual a la de la población.

Pregunta 5

Yo creo que sería inferior a la del año anterior, es decir a 70,2 porque si para una muestra de 150, fue de 60,2 se podría decir que el promedio de vida del hombre va disminuyendo, por lo tanto se puede decir inferior a 70,2

Preestructural

La frase “Yo creo que sería inferior a la del año anterior”, muestra que Gabriela no comprendió la pregunta, y la relacionó con la pregunta 2.

LUIS

Pregunta 1

Antes de Experimentación

No. Es claro que la cantidad total de mujeres que murieron en el 2015 es un número bastante alto. Con las 100 muestras no estamos haciendo al total de ellas y por esta razón no es válido afirmar inmediatamente que el promedio de la muestra es proporcional al promedio de todas las mujeres fallecidas.

Después de Experimentación

Si pienso que el promedio de vida es mayor porque al simular 2000 muestras de 100 muertes de con el promedio del año 2015 tenemos que el promedio de 80.8 años no pertenece a esa distribución ya que la probabilidad de que esa edad o una mayor es incluso menor al 1 por ciento.

Preestructural → Relacional

Antes de la experimentación, Luis no considera el tamaño de la muestra como representativo de la población, Su argumento “*Con las 100 muestras no estamos haciendo al total de ellas y por esta razón no es válido afirmar*” es válido, puesto que una prueba de hipótesis no permite afirmar, sino que presenta evidencia a favor o en contra de una hipótesis. Con esta respuesta pareciera que Luis no comprende aun la base de una prueba de hipótesis

Después de la experimentación, Luis es el único de los estudiantes que no presenta la simulación como parte de su argumento. El argumento de Luis se centra en la descripción de la simulación que hizo; y aunque no hay evidencia de la simulación, Luis, a diferencia de sus compañeros, no basó su inferencia en el gráfico, sino que cuantificó un valor de probabilidad. La respuesta de Luis es muy bien estructurada y relaciona correctamente los conceptos que trabajó.

Es extraño el contraste entre las respuesta de Luis antes y después de la experimentación. Puesto que inicialmente no parece tener comprensión de lo que significa una prueba de hipótesis, pero después de la simulación realiza un inferencia correcta.

Preguntas 2

Antes de Experimentación

Sí. Porque en este caso los años de vida disminuyen notablemente esto sucede ya que la muestra es bastante pequeña en comparación al número de hombre que en realidad fallecieron en ese año.

Después de Experimentación

En esta ocasión gracias a la muestra de 150 muertes de hombres tenemos suficiente evidencia a favor de que el promedio de vida si cambio, ya que en la gráfica es bastante notorio que el promedio de 60.2 no pertenece a información de las muertes registradas el año pasado.

Uniestructural → Multiestructural

Antes de la experimentación, Luis tampoco considera la muestra de 150 como representativa de la población, e indirectamente utiliza la ley de los pequeños números para justificar su respuesta, como la muestra es muy pequeña la media muestral es muy variable.

Después de la experimentación, al igual que los otros tres estudiantes, Luis hace su inferencia a partir del gráfico. Su argumento es claro “*ya que en la gráfica es bastante notorio que el promedio de 60.2 no pertenece a información de las muertes registradas el año pasado*”.

Pregunta 3

Antes de Experimentación

No menos de 69 años. Es un numero lo suficientemente cercano al 72.1 como para afirmar que el promedio no ha cambiado. También el porcentaje de valores a obtener debajo de los 69 sería muy pequeño lo que no concordaría con el promedio obtenido en primera instancia en el año 2015.

Después de Experimentación

Para poder afirmar que el promedio de vida no ha cambiado para una muestra con un promedio de vida dentro de los 71.3 y 73 años ya que en estas ocasiones la probabilidad de obtener edades en ese intervalo es mayor al 11 por ciento por la izquierda y al 12 por ciento por la derecha.

Multiestructural → Relacional

Antes de la experimentación, Luis, lo solo da un valor para la región de rechazo informal, sino que presume que el p-valor va a ser muy pequeño.

Ya después de la experimentación y con ayuda de la simulación; Luis, al igual que Andrés, toma la pregunta a dos colas (a pesar de ser a una cola). Esta situación llevó a Luis a crear un intervalo de confianza del 80% sin saberlo. Luis fue más allá de lo pedido, por lo cual su respuesta está en un nivel relacional

Pregunta 4

Para esta muestra la vida media se debe asemejar bastante al promedio de la vida de todos los hombres en el 2015. Es decir, un número muy cercano a los 72.1 años. Serían 70 años.

Multiestructural

Al igual que sus compañeros, Luis ya identifica la muestra como representante de una población. Su cambio de pensamiento es evidente, porque en la prueba inicial, al igual que Gabriela, consideraba la media de la muestra igual a la de la población.

Pregunta 5

A pesar de que en las muestras obtengamos valores diferentes a los del promedio del año 2015, la vida media realmente no cambia, la cantidad de años que aceptaría en este caso sería también muy parecida a la del año 2015 en general.

Uniestructural

La respuesta de Luis, habla de la variabilidad muestral, y de la muestra como representante de la población. Pero la influencia del tamaño muestral, componente principal de la pregunta, es ignorada por Luis.

4.5.4. Conclusiones

La inferencia a partir del gráfico de la distribución muestral sigue siendo el argumento predilecto por lo estudiantes, muy pocos ven la necesidad de cuantificar una probabilidad. Mientras el gráfico sea lo suficientemente evidente para realizar la inferencia, los estudiantes no ven la necesidad de calcular el p-valor.

Al no contar con la desviación de la población, en esta distribución normal (continua), los estudiantes antes de la experimentación recurrieron en su gran medida al criterio de la proximidad como argumento base para sus inferencias.

Los estudiantes no lograron establecer la influencia del tamaño de muestra en la media muestral; solo un estudiante lo logró. Pero, sí fue evidente que los estudiantes superaron en su gran mayoría sus dificultades relacionadas con la muestra como representante de la población.

4.6. Prueba Final

4.6.1. Análisis de Componentes

Situación 1: Una urna tiene 10 bolas entre blancas (B) y negras (N). Después de hacer 200 extracciones (con remplazo) se obtuvieron 112 bolas blancas y 88 bolas negras.

Pregunta 1

¿Cuál de las siguientes afirmaciones sobre la composición de la urna es correcta?

Componente

Reconocer la variabilidad de la muestra de extracciones y el estimador proporcional.

Tabla 15. Prueba Final Pregunta 1

Opciones	Razonamientos Informales	Nivel SOLO
a. No se puede decir nada porque las extracciones son muy pocas.	Se presentó mucho en la experimentación, no se reconoce la muestra como representante de la población, y por lo tanto no se desarrolla el objetivo de la pregunta.	<i>Preestructural:</i> Reconoce la variabilidad de la muestra y el estimador muestral.
b. Es muy probable que la urna contenga 5B y 5N.	112 bolas blancas equivalen a una proporción de $\frac{56}{100}$ (0.56). La composición dada en la opción b. está inmediatamente por debajo de esta composición.	<i>Uniestructural:</i> Reconoce el estimador puntual, pero no de la variabilidad de la muestra. <i>Multiestructural:</i> Reconoce el estimador puntual y la variabilidad de la muestra, pero no la relación entre ellos.
c. Es muy probable que la urna contenga 6B y 4N.	112 bolas blancas equivalen a una proporción de $\frac{56}{100}$ (0.56). La composición de la respuesta c. está inmediatamente por encima de esta composición.	<i>Uniestructural:</i> Reconoce el estimador puntual, pero no la variabilidad de la muestra. <i>Multiestructural:</i> Reconoce el estimador puntual y la variabilidad de la muestra, pero no la relación entre ellos.

d. La composición de la urna puede ser cualquiera.	Refleja el sesgo de Equiprobabilidad.	<i>Uniestructural:</i> Reconoce la variabilidad de la muestra, pero no del estimador puntual.
e. Es muy probable que la composición de la urna no tenga más de 7 bolas blancas.	Respuesta correcta, tiene en cuenta la muestra obtenida, la variabilidad y el estimador muestral.	<p><i>Multiestructural:</i> Reconoce el estimador puntual y de la variabilidad de la muestra, pero no las relaciona adecuadamente.</p> <p><i>Relacional:</i> Reconoce el estimador muestral y la variabilidad; y las relaciona entre ellos.</p>

Pregunta 2

Si en 10 extracciones se obtuvieron n 5B y 5N. Se puede afirmar sobre la composición de la urna que:

Componente

Contrastar la ley de los pequeños números

Tabla 16. Prueba Final Pregunta 2

Opciones	Razonamientos Informales	Nivel SOLO
a. No se puede decir nada porque las extracciones son muy pocas.	Respuesta que se presentó mucho en la experimentación. Aunque en esta ocasión la muestra si es pequeña, si es posible hacer una afirmación sobre la composición de la urna.	<i>Uniestructural:</i> Evidencia conciencia de las implicaciones del tamaño muestral sobre la variabilidad de la proporción, pero las extrapola erróneamente a la composición de la urna.

b. Es muy probable que la urna contenga 5B y 5N.	Respuesta determinista, ligada a la creencia de la ley de los pequeños números.	<i>Preestructural:</i> Evidencia del sesgo de la ley de los pequeños números.
c. La composición de la urna puede ser cualquiera.	Respuesta que ignora la muestra pues la composición de la urna no puede tener 10B o 10N.	<i>Uniestructural:</i> Evidencia consciencia de las implicaciones del tamaño muestral sobre la variabilidad de la proporción, pero las extrapola erróneamente a la composición de la urna.
d. Hay entre 4 y 6 blancas, las demás son negras.	Respuesta no determinista, pero ligada a la creencia de la ley de los pequeños números.	<i>Preestructural:</i> Evidencia consciencia de variabilidad, pero no identifica las implicaciones del tamaño de la muestra. <i>Uniestructural:</i> Evidencia consciencia de variabilidad, e identifica las implicaciones del tamaño de la muestra, pero las extrapola erróneamente a la composición de la urna.
e. Hay entre 2 y 8 blancas, las demás son negras.	Respuesta correcta.	<i>Multiestructural:</i> Identifica las proporciones de la composición de la urna inhabilitas en la situación <i>Relacional:</i> Identifica la implicación del tamaño muestral sobre la variabilidad de las proporciones muestrales; pero no lo extrapola a la composición de la urna.

Pregunta 3

Supongamos que se hacen 60 extracciones. ¿Cuántas bolas blancas como máximo aceptaría que salieran para afirmar que la urna está compuesta por 6B y 4N?

Componente

Nivel de significancia informal (región de rechazo).

Tabla 17. Prueba Final Pregunta 3

Opciones	Razonamientos Informales	Nivel SOLO
a. Menos de 36 bolas blancas.	Incomprensión del nivel de significancia.	<i>Preestructural:</i> Evidencia incomprensión del nivel de significancia (región de rechazo).
b. 38 bolas blancas.	Relaciona el nivel de significancia con los valores cercanos a la proporción de la composición de la urna.	<i>Uniestructural:</i> Comprende la situación y utiliza el argumento determinista de “mayor que”.
c. 45 Bolas blancas.	Respuesta correcta, el nivel de significancia oscila entre 0.147 y 0.009.	<i>Uniestructural:</i> En su respuesta utiliza argumentos poco claros o inadecuados. <i>Multiestructural:</i> Comprende la situación y utiliza argumentos de proximidad, como sustento para la región de rechazo. <i>Relacional:</i> Relaciona las posibles composiciones de la urna con los argumentos de proximidad para definir la región de rechazo.
d. 60 bolas blancas.	Respuesta que se presentó en la experimentación, relacionada con que todos los eventos son equiprobables.	<i>Uniestructural:</i> Comprende la situación, pero no acepta región de rechazo.
e. Cualquier cantidad.	Incomprensión de la situación.	<i>Preestructural:</i> No comprende ni la situación, ni el nivel de significancia informal.

Pregunta 4

Si se quintuplican las extracciones a 300, ¿cuál es el número máximo de bolas blancas que usted aceptaría que salieran para afirmar que la urna está compuesta por 6B y 4N?

Componente

El tamaño de la muestra no es proporcional con la región de rechazo.

Tabla 18. Prueba Final Pregunta 4

Opciones	Razonamientos Informales	Nivel SOLO
a. Menos de 180 bolas blancas.	Incomprensión del nivel de significancia. Proporcional a los valores anteriores	<i>Preestructural:</i> Evidencia de incomprensión del nivel de significancia.
b. 190 bolas blancas.	Relaciona el nivel de significancia con los valores cercanos a la proporción de la composición de la urna. Proporcional a los valores anteriores.	<i>Uniestructural:</i> Asume el criterio de proporcionalidad como argumento de respuesta.
c. 200 bolas blancas.	Respuesta correcta, el nivel de significancia oscila entre 0.12 y 0.009.	<i>Uniestructural:</i> En su respuesta utiliza argumentos poco claros o inadecuados. <i>Multiestructural:</i> Utiliza el criterio de proximidad como argumento de respuesta. <i>Relacional:</i> Además de utilizar el criterio de proximidad, tiene en cuenta la no proporcionalidad.
d. 225 bolas blancas.	Respuesta proporcional a la respuesta correcta de la pregunta anterior.	<i>Uniestructural:</i> Asume el criterio de proporcionalidad como argumento de respuesta.
e. Otro valor ¿Cuál?_____	Incomprensión de la situación.	

Pregunta 5

Supongamos que un estudiante realizó 300 extracciones, y obtuvo 165B y 135N. Con base en estos resultados el estudiante afirma que la composición de la urna probablemente es 5B y 5N. Si se repite 10.000 veces el experimento de extraer con repetición 300 bolas de una urna que contiene 5 bolas blancas y 5 bolas negras, y se calcula la proporción de casos en los que se obtiene 165 o más bolas blancas. Obteniéndose el valor 4.5% ¿Qué se puede decir de la afirmación del estudiante?

Componente

Contraste de hipótesis (análisis del p-valor)

Tabla 19. Prueba Final Pregunta 5

Opciones	Razonamientos Informales	Nivel SOLO
a. El estudiante tiene razón.	Respuesta incorrecta, interpretación errónea del p-valor.	<i>Uniestructural:</i> Realiza el contraste de hipótesis haciendo una interpretación incorrecta del p-valor.
b. Hay evidencia en contra de la afirmación del estudiante.	Respuesta correcta, interpretación adecuada del p-valor.	<i>Uniestructural:</i> Realiza el contraste de hipótesis, pero sus argumentos no son claros. <i>Multiestructural:</i>

		Realiza el contraste de hipótesis enfocado solo en el p-valor. <i>Relacional:</i> Realiza el contraste de hipótesis teniendo en cuenta la relación entre el p-valor y el nivel de significancia informal.
c. El estudiante está equivocado porque la única forma de conocer la composición de la urna es sacando las 10 bolas.	Respuesta incorrecta, ignora el p-valor.	<i>Preestructural:</i> Ignora por completo en significado de una prueba de hipótesis.
d. La información no es suficiente para dar una apreciación.	Respuesta incorrecta, interpretación errónea del p-valor	<i>Preestructural:</i> No comprende el proceso de simulación de una prueba de hipótesis.

Pregunta 6

El siguiente gráfico es de una simulación de una urna con composición 7B y 3N, donde se tomaron 10000 muestras de tamaño 250 en las que se contó el número de bolas blancas.

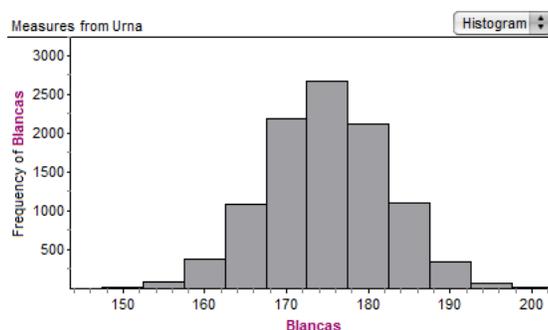


Figura 80. Prueba Final gráfico 1

Según el gráfico, ¿Se puede decir que una muestra de 195B y 105N es fácilmente extraíble de esa urna?

Componente

Contraste de hipótesis (análisis gráfico)

Tabla 20. Prueba Final Pregunta 6

Opciones	Razonamientos Informales	Nivel SOLO
a. No se sabe, el gráfico no brinda la información necesaria.	La respuesta muestra no comprensión al proceso de simulación.	<i>Preestructural:</i> No comprende el proceso de simulación, y la inferencia a partir del gráfico.
b. Sí, el gráfico muestra que es posible, que se dé esa muestra.	La respuesta olvida el aspecto probable de la muestra y se queda solo con que es posible.	<i>Uniestructural:</i> Utiliza un argumento basado en la posibilidad y no en la probabilidad.
c. El gráfico muestra que es poco probable que la muestra pertenezca a esta composición de urna.	Respuesta correcta, inferencia adecuada a partir del gráfico y de la comprensión de la simulación.	<i>Uniestructural:</i> En su respuesta utiliza argumentos poco claros o inadecuados. <i>Multiestructural:</i> Utiliza el criterio de la proximidad como único argumento de respuesta. <i>Relacional:</i> Utiliza el criterio de la proximidad como argumento de respuesta; y relaciona su inferencia con la posible composición de la urna.

Situación 2: El profesor de educación física de cierto colegio quiere saber si el índice de masa corporal, IMC (es el resultado de dividir el peso de la persona entre su longitud en metros al cuadrado) de sus 500 estudiantes disminuyó después de la implementación de un plan de ejercicios propuesto para el año escolar. En el mes de enero el promedio de IMC de los 500 estudiantes era

25.3; pero para el mes de noviembre el profesor sólo contó con 100 de sus estudiantes. El promedio de IMC de esos 100 estudiantes fue de 24.7.

Pregunta 7

Los estudiantes del grado noveno constituyen una muestra de 95 estudiantes de los 500 estudiantes de educación física. ¿Cuál cree que es el promedio del IMC de los estudiantes de noveno en el mes de enero (primera toma del IMC)?

Componente

Muestra como representante de la población, estimador muestral

Tabla 21. Prueba Final Pregunta 7

Opciones	Razonamientos Informales	Nivel SOLO
a. 25.3	Respuesta netamente determinística.	<i>Preestructural:</i> Concibe que la media de la muestra deba ser igual a la media de la población.
b. 24.7	Incomprensión de la pregunta.	<i>Preestructural:</i> No comprende de la situación planteada.
c. Entre 25.1 y 25.5	Evidencia de la muestra como representante de la población.	<i>Multiestructural:</i> Identifica a la muestra como representante de la población, pero el argumento para la

		respuesta no es claro o adecuado. <i>Relacional:</i> Evidencia comprensión de las implicaciones de la muestra como representante de la población.
d. Entre 25.3 y 26.3	Respuesta que se presentó en la prueba inicial, donde solo se toman valores iguales o superiores a la media.	<i>Uniestructural:</i> Identifica la variabilidad de la muestra, pero no las características del estimador muestral.
e. No, se puede saber porque es un valor variable.	Noción de variabilidad, pero incomprensión de la muestra como representante de la población.	<i>Preestructural:</i> No identifica la muestra como representante de la población.

Pregunta 8

¿Considera usted que el profesor puede afirmar que el IMC de sus 500 estudiantes disminuyó después del año escolar?

Componente

Relevancia de la desviación.

Tabla 22. Prueba Final Pregunta 8

Opciones	Razonamientos Informales	Nivel SOLO
a. Sí, porque 24.7 es menor que 25.3.	Respuesta netamente determinística/ Experiencia de los talleres.	<i>Preestructural:</i> Utiliza el argumento “menor que” como base de su respuesta. <i>Multiestructural:</i> Identifica la muestra como representante de

		la población para realizar la inferencia.
b. No, porque no se conoce cómo se comporta la distribución de promedios	Respuesta correcta, la justificación tiene en cuenta que la media no es suficiente para comprender el comportamiento de la distribución.	<i>Uniestructural:</i> No amplia los argumentos dados en la misma respuesta. <i>Multiestructural:</i> Advierte la necesidad de conocer la desviación. <i>Relacional:</i> Advierte la necesidad de conocer la desviación, pero se arriesga a tomar una decisión a la luz de los datos.
c. Si, después de hacer ejercicio durante un año el IMC debe haber disminuido, porque el ejercicio quema calorías y grasa.	Respuesta basada en el contexto de la situación.	<i>Uniestructural:</i> Realiza la inferencia influenciado por el contexto de la situación.
d. No, necesita tener el IMC de los 500 estudiantes para poder hacer esa afirmación.	Respuesta ignora la muestra como representante de la población.	<i>Preestructural:</i> Desestima el tamaño muestral para realizar la inferencia.
e. No, porque la muestra no es lo suficientemente grande para poder hacer esa afirmación.	Respuesta que se presentó en la experimentación. Donde no se considera el tamaño muestral como indicado para hacer la inferencia.	<i>Preestructural:</i> Desestima el tamaño muestral para realizar la inferencia.

Pregunta 9

El ayudante del profesor de educación física cuenta con el gráfico de la distribución del IMC de los 500 estudiantes que se hizo al iniciar el año escolar.

El ayudante afirma que no existe evidencia de que el promedio del IMC de los 500 haya disminuido después del año escolar. ¿Qué piensa de la afirmación del ayudante?

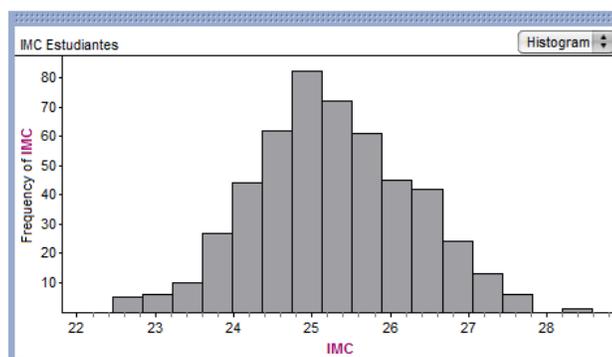


Figura 81. Prueba Final gráfico 2

Componente

Análisis de la distribución para hacer el contraste de hipótesis. (Observar si el estudiante, comprendió el proceso de la situación)

Tabla 23. Prueba Final Pregunta 9

Opciones	Razonamientos Informales	Nivel SOLO
a. Es correcta, porque en el gráfico se puede ver que IMC de 24.7 fueron de los que más se presentaron, alrededor de 60 de 500 estudiantes.	Confusión entre el gráfico de IMC y el gráfico de promedios de IMC necesario para el contraste de hipótesis. Inferencia a partir del análisis puntual de un valor teniendo en cuenta su frecuencia Situación vista en la prueba diagnóstica inicial.	<i>Preestructural:</i> Confunde el gráfico de IMC y el gráfico de promedios de IMC necesario para el contraste de hipótesis.
b. Es errónea, porque si alrededor de 60 estudiantes tuvieron un IMC de 24.7, eso significa que sólo un 12% de los estudiantes tiene este IMC. Como esta probabilidad es muy baja, hay evidencia en contra de	Confusión entre el gráfico de IMC y el gráfico de promedios de IMC. Análisis puntual sobre IMC de 24.7 y no sobre el promedio de 24.7	<i>Preestructural:</i> Confunde el gráfico de IMC y el gráfico de promedios de IMC necesario para el contraste de hipótesis.

la afirmación del ayudante		
c. Es correcta, como 24.7 está muy cercano al centro de la gráfica entonces o hay evidencia.	Confusión entre el gráfico de IMC y el gráfico de promedios de IMC. Utilizando el criterio de proximidad que se vio en la experimentación.	<i>Preestructural:</i> Confunde el gráfico de IMC y el gráfico de promedios de IMC necesario para el contraste de hipótesis.
d. Es errónea, el ayudante no puede hacer esa afirmación porque el gráfico no es el adecuado.	Respuesta correcta, el gráfico suministrado sólo permite ver la distribución de los IMC de los 500 estudiantes, no es una distribución de promedios de muestras de tamaño 100.	<i>Multiestructural:</i> Identifica como errónea la afirmación de estudiante. <i>Relacional:</i> Identifica como errónea la afirmación de estudiante, explicando la imposibilidad de inferencia sobre el gráfico suministrado.

Pregunta 10

La probabilidad de obtener 24.7 o menos en el promedio del IMC de los 100 estudiantes es menor del 0.01. A partir de la información se puede decir que hay evidencia de que el promedio del IMC disminuyó.

Componente

Contraste de hipótesis. (Análisis del p-valor)

Tabla 24. Prueba Final Pregunta 10

Opciones	Razonamientos Informales	Nivel SOLO
a. Si disminuyó.	Respuesta correcta, interpretación adecuada del p-valor.	<i>Uniestructural:</i> Identifica la respuesta correcta, pero no la justifica. <i>Multiestructural:</i>

		Identifica la inferencia correcta, y utiliza el p-valor para realizar la inferencia. <i>Relacional:</i> Identifica la inferencia correcta, y utiliza el p-valor y el nivel de significancia informal para realizar la inferencia.
b. No disminuyó.	Respuesta incorrecta, no hay interpretación adecuada del p-valor.	<i>Preestructural:</i> Presenta la respuesta sin justificación. <i>Uniestructural:</i> La respuesta incluye argumentos válidos, pero utilizados de manera errónea.
c. No se puede decir nada.	Respuesta que ignora los datos suministrados en la pregunta.	<i>Preestructural:</i> Ignora por completo el significado del proceso de simulación.
d. Aunque la probabilidad es muy baja, es de una muestra de 100 estudiantes, se necesitan los 500 estudiantes para poder decir que el promedio del IMC si disminuyó.	Inferencia errónea no se identifica la muestra como representante de la población.	<i>Preestructural:</i> Desestima el tamaño muestral para realizar la inferencia.

4.6.2. Análisis por pregunta

La prueba final se realizó sin ayuda de medios digitales y 3 meses después de la implementación de los trabajos, con el fin de evidenciar que concepciones interiorizaron los estudiantes.

Situación 1: Una urna tiene 10 bolas entre blancas (B) y negras (N). Después de hacer 200 extracciones (con remplazo) se obtuvieron 112 bolas blancas y 88 bolas negras.

1. ¿Cuál de las siguientes afirmaciones sobre la composición de la urna es correcta?

- a. No se puede decir nada porque las extracciones son muy pocas.
- b. Es muy probable que la urna contenga 5B y 5N.
- c. Es muy probable que la urna contenga 6B y 4N.
- d. La composición de la urna puede ser cualquiera.
- e. Es muy probable que la composición de la urna no tenga más de 7 bolas blancas.

Tabla 25. *Análisis Prueba Final Pregunta 1*

Estudiante	Respuesta	Justificación
Andrés	c	La proporción sería 6/4 por lo tanto los resultados de 200 extracciones rodearía 120 bolas blancas.
Sofía	e	Si porque la probabilidad de que salgan más de 7 bolas blancas es baja ya que no sería proporcional a 112 bolas B y 88 bolas N
Gabriela	b	112 es un valor que puede estar muy cercano al centro de una urna con 5B y 5N
Luis	c	Con una extracción de 112 B y 88 N. Pienso que es muy probable que contengan 6B y 4N porque después de hacer 200 extracciones podríamos analizar el comportamiento de la distribución y dar con que la probabilidad de extraer un número cercano a 112 bolas blancas o 88 bolas negras.

Andrés utiliza el argumento de la proporción que usó en segundo taller (sin simulación), aunque esta vez, aceptó como válida la composición que más se ajustó a la proporción calculada. En el segundo taller sólo una composición de urna tenía suficiente evidencia para ser escogida; tal vez, por esa razón decidió optar por una única composición. Andrés no tuvo en cuenta que valor de la proporción (0.56) no permitía despreciar la composición con 5 bolas blancas. De igual forma, Andrés dejó de lado la opción a., que fue su argumento en el primer taller. También, Andrés tiene

en cuenta la variabilidad al mencionar que el resultado de las extracciones “rodearía” las 120 bolas blancas.

Gabriela fue el caso contrario a Andrés, se inclinó por la composición con 5B, que también tiene alta probabilidad de ser la composición de la urna. En su argumento evidencia las nociones de variabilidad que ganó a lo largo de los talleres, al hablar de 112 como un valor “*cercano al centro de una urna con 5B y 5N*”. Al parecer, al igual que Andrés, Gabriela se ve influenciada por la elección de una única composición (segundo taller).

La opción e., es la más acertada para esta pregunta. Sofía la eligió utilizando el argumento de la variabilidad y en la proporción del número de extracciones (desarrollada en el segundo taller) al no inclinarse por una composición en particular, sino en descartar las composiciones de urnas que tenían menos probabilidad de ser correctas. Sofía fue la única estudiante que identificó que debido a la proporción del número de extracciones no se podía escoger una sola composición para la urna.

Luis también optó por la urna con 6B, la cual tiene mayor probabilidad de ser la composición de la urna. Según Luis, las extracciones son suficientes para poder hacer la inferencia, “*después de hacer 200 extracciones podríamos analizar*”, siendo el único estudiante que expresó la relevancia del número de extracciones para poder hacer la inferencia.

Tabla 26. Nivel SOLO prueba Final pregunta 1

Estudiante	Nivel
Andrés	Multiestructural
Sofía	Relacional
Gabriela	Multiestructural
Luis	Multiestructural

2. Si en 10 extracciones se obtuvieron 5B y 5N. Se puede afirmar sobre la composición de la urna que:
- No se puede decir nada porque las extracciones son muy pocas.
 - Es muy probable que la urna contenga 5B y 5N.
 - La composición de la urna puede ser cualquiera.
 - Hay entre 4 y 6 blancas, las demás son negras.
 - Hay entre 2 y 8 blancas, las demás son negras.

Tabla 27. *Análisis Prueba Final Pregunta 2*

Estudiante	Respuesta	Justificación
Andrés	b	Si tenemos que de un número x de extracciones la mitad de las bolas es de un color determinado podemos decir que esa relación es 50/50. Ejemplificando esto tenemos el caso anterior.
Sofía	a	Me parece muy pocas extracciones pues hay 10 bolas como total y 10 extracciones no me generan una buena muestra.
Gabriela	a	10 extracciones son muy pocas con eso no se puede afirmar que la urna tiene 5B y 5N, o que tiene muchas de un color como 8N y 2B.
Luis	a	No se puede decir aun con 10 extracciones de que la composición sea esa porque el número de extracciones es muy pequeño. Al hacer más extracciones nos acercaríamos a la verdadera composición.

Para Sofía, Gabriela y Luis “*No se puede decir nada porque las extracciones son muy pocas*”, los tres encaminaron sus argumentos a que 10 extracciones no eran suficientes para hacer una inferencia.

Luis, complementó su respuesta con el argumento: “*Al hacer más extracciones nos acercaríamos a la verdadera composición*”. Este argumento, evidencia que Luis comprende el comportamiento y la relación entre el número de extracciones y la composición de la urna.

Por su parte, Andrés, paso de no considerar, en la prueba inicial, las extracciones como representantes de la composición de la urna a aceptarlas de forma determinística, haciendo caso omiso de la influencia del número de extracciones sobre la variabilidad de las proporciones muestrales.

El error que cometieron los estudiantes, fue olvidar que no obstante el número reducido de extracciones, la proporción muestral tiene una distribución y que, por lo tanto, existen valores para las proporciones más posibles que otros. Dicho en otras palabras, olvidaron que la implicación del tamaño muestral es básicamente sobre la variabilidad de las proporciones muestrales lo que hace que la incertidumbre sobre la verdadera composición sea mayor.

Tabla 28. Nivel SOLO prueba Final pregunta 2

Estudiante	Nivel
Andrés	Preestructural
Sofía	Uniestructural
Gabriela	Uniestructural
Luis	Uniestructural

3. Supongamos que se hacen 60 extracciones. ¿Cuántas bolas blancas como máximo aceptaría que salieran para afirmar (no rechazar la afirmación) que la urna está compuesta por 6B y 4N?
- a. Menos de 36 bolas blancas.
 - b. 38 bolas blancas.
 - c. 45 bolas blancas.
 - d. 60 bolas blancas.
 - e. Cualquier cantidad.

Tabla 29. Análisis Prueba Final Pregunta 3

Estudiante	Respuesta	Justificación
Andrés	c	Tenemos que la relación es 60% a 40%, el 60% de 60 es 36 por lo tanto los resultados (en tanto a bolas blancas) deben ser superiores a esa cantidad.
Sofía	c	Como máximo 45 pues lo ideal sería 32 y 40 para que sea una composición de 6 y 4.
Gabriela	c	Es un intervalo que no esta tan lejos, ni tan cerca del centro.
Luis	c	Aceptaría como máximo entre las 40 y 45 bolas, dado que 60 extracciones sigue siendo un número pequeño con el que se puede presentar un error mayor respecto a la cantidad máxima de bolas de un color que se puede considerar aceptable.

Los cuatros estudiantes eligieron la opción correcta, evidenciando comprensión de la situación, consciencia de variabilidad y nociones informales de región de rechazo y nivel de significancia.

Andrés utilizó el porcentaje, calculó la proporción ideal de bolas blancas (36), y como él mismo lo indica, el máximo para la aceptación “*deben ser superiores a esa cantidad*”. Aunque no lo hace explicito, Andrés superó la dificultad evidenciada en la prueba inicial, donde no consideraba la variabilidad y solo las condiciones ideales permitían hacer afirmaciones. Sin embargo, Andrés deja

dudas sobre el sentido de su respuesta cuando afirma: “*los resultados (en tanto a bolas blancas) deben ser superiores a esa cantidad*”, no es claro a que hace referencia cuando habla de los resultados.

Sofía asumió una región de aceptación con un intervalo entre 32 y 40; y una región de rechazo fuera de él. Gabriela realizó un análisis basado en el posible gráfico de la distribución, utilizando como argumento el criterio de proximidad, muy usado durante el desarrollo de los talleres.

Por su parte, Luis, además de escoger la opción c, deja claro el efecto que sobre la variabilidad y, por ende, la exactitud de la estimación, tiene el tamaño muestral.

Tabla 30. Nivel SOLO prueba Final pregunta 4

Estudiante	Nivel
Andrés	Uniestructural
Sofía	Multiestructural
Gabriela	Multiestructural
Luis	Multiestructural

4. Si se quintuplican las extracciones a 300, ¿cuál es el número máximo de bolas blancas que aceptaría que salieran para afirmar que la urna está compuesta por 6B y 4N?
- Menos de 180 bolas blancas.
 - 190 bolas blancas.
 - 200 bolas blancas.
 - 225 bolas blancas.
 - Otro valor ¿Cuál? _____

Tabla 31. *Análisis Prueba Final Pregunta 4*

Estudiante	Respuesta	Justificación
Andrés	c	El 60% de 300 es 180 por lo tanto los valores deben ser superiores a esta cantidad pero cercanos a esa cantidad.
Sofía	c	Porque si 180 es el valor ideal para que sea 6 a 4, el máximo que puede ser sería 200, las extracciones por encima de este ya no pertenecen a esta composición.
Gabriela	c	Como se quintuplicó la muestra y además no es proporcional la región de rechazo. Entonces debe ser un valor menor que 200 y por encima de 180.
Luis	c	No aceptaría más de esas bolas blancas ya que con 300 extracciones se puede inferir que el número de extracciones de ese color estará por arriba pero cerca al centro de las extracciones.

Los 4 estudiantes respondieron correctamente la pregunta, evidenciando consciencia de la no proporcionalidad entre el número de extracciones y la región de rechaza/nivel de significancia informal. Es decir, si el tamaño de la muestra se duplica el valor asignado para la región de rechazo no cambia en las mismas proporciones.

Andrés, Sofía y Gabriela, hacen explícita la proporción ideal de la urna con 180 bolas blancas. Luis, aunque no lo dice explícitamente, da a entenderla cuando hace referencia al “*centro de las extracciones*”.

Andrés y Luis utilizan el criterio de proximidad como argumento base de su inferencia. Por su parte, Sofía, aunque presentó la respuesta correcta, no logró concretar ningún argumento para sustentarla.

Finalmente, aunque se evidencia comprensión de la no proporcionalidad por parte de los estudiantes, solo Gabriela utilizó este criterio con argumento en su respuesta “*además no es proporcional la región de rechazo*”; utilizando inclusive el término región de rechazo.

Tabla 32. Nivel SOLO prueba Final pregunta 4

Estudiante	Nivel
Andrés	Multiestructural
Sofía	Uniestructural
Gabriela	Relacional
Luis	Multiestructural

5. Supongamos que un estudiante realizó 300 extracciones, y obtuvo 165B y 135N. Con base en estos resultados el estudiante afirma que la composición de la urna probablemente es 5B y 5N. Se repite 10.000 veces el experimento de extraer con repetición 300 bolas de una urna que contiene 5 bolas blancas y 5 bolas negras, y se calcula la proporción de casos en los que se obtiene 165 o más bolas blancas obteniéndose el valor 4.5% ¿Qué se puede decir de la afirmación del estudiante?
- El estudiante tiene razón.
 - El estudiante está equivocado.
 - El estudiante está equivocado porque la única forma de conocer la composición de la urna es sacando las 10 bolas.
 - La información no es suficiente para dar una apreciación.

Tabla 33. Análisis Prueba Final Pregunta 5

Estudiante	Respuesta	Justificación
-------------------	------------------	----------------------

Andrés	b	El porcentaje de veces que se cumple esta situación es demasiado bajo por lo tanto diría que está equivocado.
Sofía	b	Pues la afirmación del estudiante de que la composición es de 5 B y 5N es incorrecta pues si da un valor de 4,5 % significa que hay menos bolas negras que blancas por lo tanto no sería una proporción de 5B y 5N.
Gabriela	b	Como la probabilidad es tan baja significa que está lejos del centro de la gráfica entonces la urna no puede tener por 5B y 5N.
Luis	b	La probabilidad es muy baja para aceptar que la composición de la urna sean 5B y 5N, lo aceptaría si la probabilidad supera por lo menos el diez por ciento. Con esa cantidad de extracciones y repeticiones es muy probable que la composición sea 6B y 4N

Los cuatro estudiantes respondieron correctamente la pregunta, evidenciando manejo del proceso de simulación y en algunos casos nociones de nivel de significancia. Aspectos que se trabajaron en cada uno de los talleres trabajados.

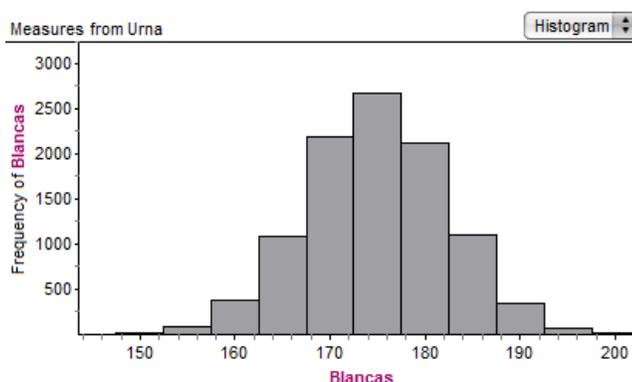
Andrés argumentó “*El porcentaje de veces que se cumple esta situación es demasiado bajo*”, haciendo una correcta interpretación del 4.5%. De igual forma, Sofía también realizó una correcta interpretación del 4.5%, haciendo un tratamiento adecuado a la hipótesis del estudiante, y realizando una correcta inferencia a la luz del porcentaje obtenido. Gabriela utilizó como argumento el p-valor informal y la región de rechazo, a partir de la interpretación del 4.5% en el posible gráfico de distribución.

Luis, por su parte, fue un poco más allá que los demás estudiantes, exteriorizando su nivel de significación: 10% tal como lo había trabajado en el segundo y tercer taller.

Estudiante	Nivel
Andrés	Multiestructural
Sofía	Multiestructural
Gabriela	Multiestructural
Luis	Relacional

6. El siguiente gráfico es de una simulación de una urna con composición 7B y 3N, donde se tomaron 10000 muestras de tamaño 300, en las que se contó el número de bolas blancas.

Figura 82. Análisis Prueba Final gráfico 1



Según el gráfico, ¿Se puede decir que una muestra de 195B y 105N es muy probable que haya sido extraída de la urna que contiene 7B y 3N?

- a. No se sabe, el gráfico no brinda la información necesaria.
- b. Si, el gráfico muestra que es posible, que se dé esa muestra.
- c. El gráfico muestra que es poco probable que la muestra pertenezca a esta composición de urna.

Tabla 35. Análisis Prueba Final Pregunta 6

Estudiante	Respuesta	Justificación
Andrés	c	Los valores dados están por fuera del centro de la gráfica.
Sofía	c	Es muy poco probable pues 195 no se encuentra en esta gráfica.
Gabriela	c	195 está muy lejos de centro de la gráfica.

Luis	c	El gráfico muestra que es poco probable que la muestra pertenezca a esta composición de urna.
-------------	---	---

Todos los estudiantes eligieron la opción correcta, era de esperarse pues durante la experimentación, la inferencia a partir del gráfico fue la herramienta más utilizada.

Andrés y Gabriela, utilizaron el criterio centro-afuera del centro del gráfico (Región de aceptación-Región de rechazo). Por su parte, Sofía eligió la opción correcta, pero su argumento no es del todo válido; ella dice “195 no se encuentra en esta gráfica”, lo cual es falso porque en el gráfico se alcanzan a observar valores hasta de 200. Luis también eligió la respuesta correcta, pero su argumento es limitado, simplemente reescribió el enunciado de la respuesta.

Tabla 36. Nivel SOLO prueba Final pregunta 6

Estudiante	Nivel
Andrés	Multiestructural
Sofía	Uniestructural
Gabriela	Multiestructural
Luis	Uniestructural

Situación 2: El profesor de educación física de cierto colegio quiere saber si el índice de masa corporal, IMC (es el resultado de dividir el peso de la persona entre su longitud en metros al cuadrado) de sus 500 estudiantes disminuyó después de la implementación de un plan de ejercicios propuesto para el año escolar. En el mes de enero el promedio de IMC de los 500 estudiantes era 25.3; pero para el mes de noviembre el profesor sólo contó con 100 de sus estudiantes. El promedio de IMC de esos 100 estudiantes fue de 24.7.

7. Los estudiantes del grado noveno constituyen una muestra de 95 estudiantes de los 500 estudiantes de educación física. ¿Cuál cree que es el promedio del IMC de los estudiantes de noveno en el mes de enero (primera toma del IMC)?
- 25.3
 - 24.7
 - Entre 25.1 y 25.5
 - Entre 25.3 y 26.3
 - No, se pude saber porque es un valor variable.

Tabla 37. *Análisis Prueba Final Pregunta 7*

Estudiante	Respuesta	Justificación
Andrés	c	Los estudiantes pertenecen a la muestra inicial por lo tanto el promedio de su IMC debe ser próximo a el promedio de IMC general.
Sofía	c	El promedio estaría entre esto pues 95 pertenece a los 500 por lo tanto su IMC será cercano al de los 500.
Gabriela	c	Los 95 estudiantes son una muestra del total entonces su promedio debería estar alrededor del promedio total.
Luis	c	Ya que no estamos hablando de los 500 estudiantes sino de un valor bastante pequeño respecto al total podemos inferir que el IMC de los estudiantes de noveno que tan solo son 95 varia dentro de ese intervalo aparentemente grande.

Los 4 estudiantes respondieron correctamente la pregunta, superando la dificultad evidenciada en la prueba inicial, donde no tenían conciencia ni de la variabilidad muestral, ni de la muestra como representante de la población.

El argumento que utilizaron los estudiantes fue general, los 95 estudiantes son una muestra, por lo tanto su promedio debe estar “*próximo*” o “*cercano*” o “*alrededor*” al promedio de IMC de los 500 estudiantes.

Cabe resaltar que Luis, continúa con su resistencia al tamaño de la muestra, esta vez mencionó “*un valor bastante pequeño respecto al total*”. No obstante, la desconfianza de Luis sobre el tamaño muestral, le permitió comprender la relación entre tamaño muestral y variabilidad muestral.

Tabla 38. Nivel SOLO prueba Final pregunta 7

Estudiante	Nivel
Andrés	Relacional
Sofía	Relacional
Gabriela	Relacional
Luis	Relacional

8. ¿Considera usted que el profesor puede afirmar que el IMC de sus 500 estudiantes disminuyó después del año escolar?
- Sí, porque 24.7 es menor que 25.3.
 - No, porque no se conoce como se comporta la distribución de promedios.
 - Si, después de hacer ejercicio durante un año el IMC debe haber disminuido, porque el ejercicio quema calorías y grasa.
 - No, necesita tener el IMC de los 500 estudiantes para poder hacer esa afirmación.
 - No, porque la muestra no es lo suficientemente grande para poder hacer esa afirmación.

Tabla 39. Análisis Prueba Final Pregunta 8

Estudiante	Respuesta	Justificación
-------------------	------------------	----------------------

Andrés	a	Los 100 estudiantes hacen parte del promedio final de IMC por lo tanto el promedio de estos será similar a el promedio general.
Sofía	c	La respuesta.
Gabriela	b	La respuesta.
Luis	b	La respuesta.

Andrés asumió que sí hay evidencia a favor de la disminución. Su justificación se centró en el criterio de la muestra como representante de la población.

Sofía eligió la opción ligada al contexto de la situación; esto evidencia como mencionan García y Sánchez (2014) que el contexto juega un papel importante en las pruebas de hipótesis, puesto que los estudiantes utilizan su conocimiento informal sobre el contexto del problema para establecer las hipótesis, o como en este caso para realizar la inferencia.

Gabriela y Luis, eligieron la opción correcta, b., presumiendo que ambos identificaron la necesidad de conocer el comportamiento de la distribución.

Tabla 40. Nivel SOLO prueba Final pregunta 8

Estudiante	Nivel
Andrés	Multiestructural
Sofía	Uniestructural
Gabriela	Uniestructural
Luis	Uniestructural

9. El ayudante del profesor de educación física cuenta con el gráfico de la distribución del IMC de los 500 estudiantes que se hizo al iniciar el año escolar.

El ayudante afirma que no existe evidencia de que el promedio del IMC de los 500 haya disminuido después del año escolar. ¿Qué piensa de la afirmación del ayudante?

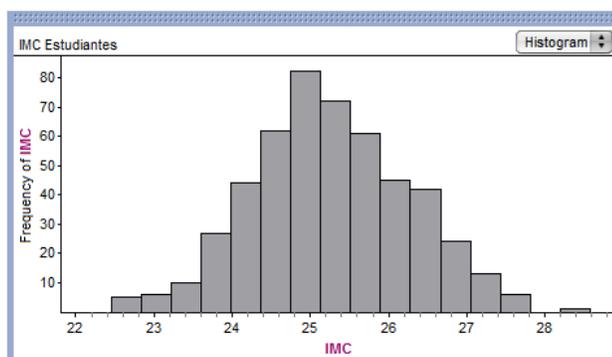


Figura 83. Análisis Prueba Final gráfico 2

- a. Es correcta, porque en el gráfico se puede ver que IMC de 24.7 fueron de los que más se presentaron, alrededor de 60 de 500 estudiantes.
- b. Es errónea, porque si alrededor de 60 estudiantes tuvieron un IMC de 24.7, eso significa que sólo un 12% de los estudiantes tiene este IMC. Como esta probabilidad es muy baja, hay evidencia en contra de la afirmación del ayudante.
- c. Es correcta, como 24.7 está muy cercano al centro de la gráfica entonces hay evidencia.
- d. Es errónea, el ayudante no puede hacer esa afirmación porque el gráfico no es el adecuado.

Tabla 41. Análisis Prueba Final Pregunta 9

Estudiante	Respuesta	Justificación
Andrés	d	Solo está teniendo en cuenta el dato de inicio de año. Si no tiene con que comparar si no tiene en cuenta los datos de final de año, no puede hacer una comparación.
Sofía	d	No se puede hacer una inferencia con una gráfica de enero, pues necesitaríamos una gráfica actual.
Gabriela	d	La gráfica no es la adecuada la gráfica correcta sería la de final de año.
Luis	c	En la respuesta.

Andrés, Sofía y Gabriela, se percataron de la imposibilidad de hacer la inferencia sobre el gráfico de enero. Evidenciando comprensión del proceso de simulación y de la prueba de hipótesis.

Por su parte, Luis, realizó la inferencia sobre el gráfico de enero, sin percatarse que no era el gráfico sobre el cual debía realizar la inferencia. Esto evidencia que Luis no comprende el proceso de simulación de la prueba, o simplemente tuvo una distracción. Al no haber justificación no se puede certeza de lo sucedido.

Tabla 42. Nivel SOLO prueba Final pregunta 9

Estudiante	Nivel
Andrés	Relacional
Sofía	Relacional
Gabriela	Relacional
Luis	Uniestructural

10. La probabilidad de obtener 24.7 o menos en el promedio del IMC de los 100 estudiantes es menor del 0.01. A partir de la información se puede decir que hay evidencia de que el promedio del IMC disminuyó.

- a. Si disminuyó.
- b. No disminuyó.
- c. No se puede decir nada.
- d. Aunque la probabilidad es muy baja, es de una muestra de 100 estudiantes, se necesitan los 500 estudiantes para poder decir que el promedio del IMC si disminuyo.

Tabla 43. Análisis Prueba Final Pregunta 10

Estudiante	Respuesta	Justificación
-------------------	------------------	----------------------

Andrés	a	El promedio disminuye ya que la probabilidad de que fuese el mismo sería 0.01.
Sofía	a	Pues la evidencia se genera más o menos desde el 12% hacia abajo, por lo tanto el 1% si es evidencia.
Gabriela	a	Mi criterio de rechazo es del 6%. Hay evidencia que si disminuyó.
Luis	d	En esta situación la primer muestra tiene 500 estudiantes y la segunda tan solo 100 por lo tanto debemos manejar un margen de error respecto a si disminuye el promedio.

De nuevo, Andrés, Sofía y Gabriela eligieron la respuesta correcta; pero esta vez con justificaciones distintas.

Andrés, asumió el p-valor como probabilidad de un evento; es decir, la probabilidad de que el promedio sea el mismo es 0.01, y la probabilidad de que haya disminuido es de 0.99. Esto evidencia una mala comprensión del significado del p-valor o, tal vez, del significado del valor de probabilidad. .

Sofía y Gabriela utilizaron el su nivel de significancia informal con criterio para realizar la inferencia. Sofía con un nivel de significancia del 12% y Gabriela con uno del 6%. Ambas evidenciaron conocimiento de la relación entre el p-valor y el nivel de significancia, pues los utilizaron correctamente para realizar la inferencia.

Luis, debido a su dificultad de aceptar e tamaño muestral como suficiente para hacer una inferencia, eligió la opción d., no aceptar la muestra como representante de la población. No obstante, analizando su justificación, Luis, parece aceptar la muestra como representante, pero debido a la variabilidad que presenta, es consciente de que existe la posibilidad del error.

Tabla 44. Nivel SOLO prueba Final pregunta 10

Estudiante	Nivel
-------------------	--------------

Andrés	Uniestructural
Sofía	Relacional
Gabriela	Relacional
Luis	Preestructural

Tabla general

Tabla 45. *Respuestas correctas prueba Final*

Respuestas Generales											
	1	2	3	4	5	6	7	8	9	10	Total
Andrés	c	b	c	c	b	c	c	a	d	a	7
Sofía	e	a	c	c	b	c	c	c	d	a	8
Gabriela	b	a	c	c	b	c	c	b	d	a	8
Luis	c	a	c	c	b	c	c	b	c	d	6
R. Correcta	e	e	c	c	b	c	c	b	d	a	10

4.6.3. Conclusiones

Al realizar la prueba sin computador y después de 3 meses de terminada la implementación de los talleres, podemos ver con buenos ojos que los avances evidenciados a lo largo de los talleres no fueron olvidados y parecen estar interiorizados por parte de los estudiantes.

La prueba permitió evidenciar que los estudiantes superaron, en su gran mayoría, las dificultades y los sesgos que tenían cuando trabajaron la prueba inicial. El reconocimiento de la muestra como representante de la población y la consciencia de la variabilidad muestral fueron dos logros de los estudiantes.

Los estudiantes, en su gran mayoría comprendieron las nociones básicas inmersas en un contraste de hipótesis, y de algunos de los conceptos más relevantes sobre el tema. La región de rechazo/Nivel de significancia informal; junto con el criterio de proximidad en los gráficos, fueron los elementos predilectos por los estudiantes para justificar sus respuesta.

4.7. Contraste Prueba Inicial versus Prueba Final

Teniendo en cuenta las componentes conceptuales propias de las pruebas de hipótesis en los siguientes cuadros se presentan los niveles alcanzados por cada uno de los estudiantes de acuerdo a la teoría SOLO, en esas componentes tanto en la prueba diagnóstica como en la prueba final.

Es importante resaltar que al aplicar la prueba final 3 meses después de la finalización de la experimentación con los talleres, evaluamos realmente llegaron a comprender los estudiantes sobre las pruebas de hipótesis y el nivel de su razonamiento informal

Tabla 46. *Contraste de resultados para Andrés*

		Andrés		
	Componente Conceptual	Prueba Inicial Nivel SOLO	Prueba Final Nivel SOLO	PROGRESO
Distribución Discreta	Variabilidad del estimador proporcional y sus implicaciones	Uniestructural	Multiestructural	↑
	Influencia del tamaño muestral	Preestructural	Multiestructural	↑
	Región de rechazo/Nivel de significancia	Preestructural	Uniestructural	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Uniestructural	Multiestructural	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Multiestructural	-
Distribución Continua	Variabilidad Muestral y del estimador muestral	Uniestructural	Relacional	↑
	Muestra representativa de la población	Uniestructural	Relacional	↑
	Influencia del tamaño muestral	Preestructural	Relacional	↑

	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Preestructural	Uniestructural	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Relacional	-

Tabla 47. *Contraste de resultados para Sofía*

Sofía				
	Componente Conceptual	Prueba Inicial Nivel SOLO	Prueba Final Nivel SOLO	PROGRESO
Distribución Discreta	Variabilidad del estimador proporcional y sus implicaciones	Preestructural	Relacional	↑
	Influencia del tamaño muestral	Preestructural	Uniestructural	↑
	Región de rechazo/Nivel de significancia	Preestructural	Multiestructural	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Preestructural	Multiestructural	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Uniestructural	-
Distribución Continua	Variabilidad Muestral y del estimador muestral	Uniestructural	Relacional	↑
	Muestra como representante de la población	Uniestructural	Relacional	↑
	Influencia del tamaño muestral	Uniestructural	Relacional	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Uniestructural	Relacional	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Relacional	-

Tabla 48. *Contraste de resultados para Gabriela*

Gabriela				
	Componente Conceptual	Prueba Inicial Nivel SOLO	Prueba Final Nivel SOLO	PROGRESO
Distribución Discreta	Variabilidad del estimador proporcional y sus implicaciones	Preestructural	Multiestructural	↑
	Influencia del tamaño muestral	Preestructural	Relacional	↑
	Región de rechazo/Nivel de significancia	Preestructural	Multiestructural	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Preestructural	Multiestructural	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Multiestructural	-
Distribución Continua	Variabilidad Muestral y del estimador muestral	Preestructural	Relacional	↑
	Muestra como representante de la población	Preestructural	Relacional	↑
	Influencia del tamaño muestral	Uniestructural	Relacional	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Preestructural	Relacional	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Relacional	-

Tabla 49. *Contraste de resultados para Luis*

Luis				
-------------	--	--	--	--

	Componente Conceptual	Prueba Inicial Nivel SOLO	Prueba Final Nivel SOLO	PROGRESO
Distribución Discreta	Variabilidad del estimador muestral (proporción) y sus implicaciones	Uniestructural	Multiestructural	↑
	Influencia del tamaño muestral	Preestructural	Multiestructural	↑
	Región de rechazo/Nivel de significancia	Preestructural	Multiestructural	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Uniestructural	Relacional	↑
	Contraste de hipótesis (análisis gráfico)	No se dio	Uniestructural	-
Distribución Continua	Variabilidad Muestral y del estimador muestral	Preestructural	Relacional	↑
	Muestra como representante de la población	Preestructural	Relacional	↑
	Influencia del tamaño muestral	Preestructural	Relacional	↑
	p-valor/Contraste de hipótesis/Razonamiento Inferencial	Preestructural	Preestructural	=
	Contraste de hipótesis (análisis gráfico)	No se dio	Uniestructural	-

El progreso que tuvieron los 4 estudiantes en evidenciable en los cuadros de contraste de las pruebas, los estudiantes mejoraron considerablemente sus argumentos y por consiguiente sus respuestas. Pasaron en su gran mayoría de tener respuesta preestructurales y uniestructurales a respuesta multiestructurales e inclusive relacionales.

5.- Conclusiones

En este capítulo se expondrán las conclusiones que surgieron de esta investigación, teniendo en cuenta dos ejes centrales: la caracterización de los razonamientos informales de los estudiantes participantes asociada a las pruebas de hipótesis, y las implicaciones que sobre el desarrollo de estos razonamientos tuvo la experimentación computacional.

Igualmente se propondrán algunas recomendaciones que den continuidad a esta investigación y que permitan superar sus deficiencias. De esta forma, el capítulo está estructurado en tres apartados; el de los estudiantes, el de la experimentación y las recomendaciones.

5.1. Los Estudiantes

A continuación se presenta la evolución conceptual de los razonamientos inferenciales informales, lograda por los estudiantes para cada uno de los componentes conceptuales que intervienen en los procesos de pruebas o contrastes de hipótesis.

5.1.1. Variabilidad de la proporción muestral y sus implicaciones (Dist. Discreta)

En un principio, las respuestas de los estudiantes se encontraban en el nivel preestructural de SOLO, ya que solo tenían conciencia de la variabilidad de la proporción muestral, pero no establecían ninguna relación con la proporción poblacional, ni percibían su distribución. Mejor dicho, asumían la distribución uniforme asociado al sesgo de equiprobabilidad, esto es, creían que todos los eventos posibles eran igualmente probables. Al finalizar la experimentación las respuesta alcanzaron el nivel multiestructural en su gran mayoría, los estudiantes lograron superar el sesgo

de equiprobabilidad y, además, entendieron el significado de la proporción muestral como estimador de la proporción poblacional.

5.1.2. Variabilidad de la muestral y del estimador muestral (Dist. Continua)

En esta componente, los estudiantes pasaron de niveles preestructurales o uniestructurales nivel relacional. En la prueba inicial los estudiantes no tenían en cuenta la variabilidad muestral y, por consiguiente, tampoco la variabilidad del estimador muestral, algunos asumían un total determinismo afirmando que la media muestral era igual a la media poblacional. Con el transcurrir de los talleres y el uso de la simulación, los estudiantes asumieron la variabilidad muestral como un hecho, y fueron capaces de analizar la distribución del estimador muestral, destacando los valores más probables y los menos probables, conceptos que les permitieron realizar inferencias mejor fundamentadas.

5.1.3. Influencia del tamaño muestral

En un principio, los estudiantes no aceptaban la representatividad de la muestra obtenida si su tamaño no era suficientemente grande. Después de realizar variadas simulaciones con muestras de diferente tamaño empezaron a percibir las implicaciones del tamaño muestral en la dispersión del estimador muestral y, por ende, en el tamaño de la región de aceptación y de rechazo. La percepción de esta relación entre tamaño muestral y región de rechazo que, inicialmente la asumieron directamente proporcional evolucionó hacia relaciones no lineales que, por supuesto, no lograron plasmar algebraicamente.

5.1.4. Región de rechazo/Nivel de significancia

Respecto a la región de rechazo y su relación con el nivel de significancia, se evidenció un desarrollo positivo. Inicialmente los estudiantes asumían prácticamente valores iguales entre el valor hipotético y el estimador para no rechazar la hipótesis nula; la visualización de la distribución muestral del estimador los condujo a desarrollar criterios de significación y de región de rechazo que oscilaban alrededor del 6% previa aceptación de la hipótesis nula.

5.1.5. Contraste de hipótesis/p-valor/Razonamiento Inferencial

Esta componente relaciona el p-valor con nivel de significancia para contrastar la hipótesis. Los estudiantes lograron asociar el nivel de significancia informal con el límite de la región de rechazo/aceptación. Además, evidenciaron el vínculo entre el p-valor y el nivel de significancia. De igual forma, como en otras componentes, el lenguaje estadístico y el sentido de la prueba de hipótesis fueron dando forma a las respuestas de los estudiantes. Pasaron de respuestas preestructurales y uniestructurales, caracterizadas por el error de asumir la prueba de hipótesis como una demostración matemática, como un resultado como verdad absoluta (Vallecillos, 1999); a respuestas de nivel multiestructural, en donde hicieron uso de un lenguaje propio de las pruebas de hipótesis, “hay evidencia a favor de”, “hay evidencia en contra de”; y realizaron inferencia correctas a partir de la relación entre el p-valor y el nivel de significancia.

5.1.6. Contraste de hipótesis (análisis gráfico)

Esta componente no se evaluó en la prueba inicial, porque surgió a partir de la interacción de los estudiantes con el *software Fathom*. Al realizar la experimentación guiada de la simulación en el primer taller de *Fathom*, los estudiantes empezaron a crear un criterio de proximidad para

realizar inferencia basada en el análisis gráfico. Empezaron utilizando expresiones como, “cercano”, “alejado”, “próximo”, “alrededor” y “al borde”, para indicar que el estimador muestral se encontraba en una región de rechazo o aceptación. A partir del uso recurrente de estas expresiones y la mejora consecuente del lenguaje estadístico, que incluía la distribución de probabilidad del estimador y del cálculo de probabilidades, el criterio de proximidad de los gráficos se convirtió en uno de los argumentos más utilizados por los estudiantes para realizar inferencias correctas, y dar respuestas de nivel multiestructural en la mayoría de los casos.

5.2. La experimentación

La premisa que asumimos durante esta investigación es que el razonamiento inferencial informal es el punto de partida para el razonamiento inferencial formal. Entendiendo este razonamiento informal como el razonamiento a partir de intuiciones primarias, si no se ha realizado estudio previo sobre el tema; o de intuiciones secundarias, las cuales son producto de un conocimiento previamente estudiado.

Empezar la experimentación a partir de estas intuiciones de los estudiantes, implicó que el primer acercamiento a las distribuciones muestrales se diera durante el desarrollo de esta investigación. Por esta razón, inicialmente las dificultades sobre este tema fueron notorias en los 4 estudiantes participantes. El manejo continuo del tema y las mediación de *Fathom*, permitieron a los estudiantes superar esa dificultad entre la primera y segunda experimentación con *Fathom*.

Respecto al *software Fathom*, inicialmente, como era de esperarse, los estudiantes presentaron dificultades para comprender el proceso de simulación, dificultades derivadas del inexistente

manejo de distribuciones que tenían hasta el momento. Con el transcurrir de los talleres los estudiantes se familiarizaron más con el software y la simulación; superaron sus dificultades, y pasaron de usar la simulación como una validadora de respuestas y utilizarla como un soporte de sus respuestas.

La cronología de la experimentación permitió sacar el mayor provecho de las actividades, trabajar primero con variables discretas facilitó el camino para luego abordar simulaciones con variables continuas. Los estudiantes utilizaron las herramientas interiorizadas para las simulaciones de distribuciones discretas y las utilizaron cuando se les presentó la simulación pre-establecida de una distribución continua.

Una de las principales dificultades que encontramos en la literatura respecto a las pruebas de hipótesis, es ver la prueba como un conjunto de reglas, una receta. Batanero (2015), dice que la aplicación correcta de fórmulas no implica comprensión de la situación y adquisición de razonamiento estadístico; por esta razón, recurrimos al razonamiento inferencial informal basado en las simulaciones y sus representaciones gráficas para alejarnos del formalismo de las pruebas. Pero trabajar con simulación no garantiza la comprensión del razonamiento alrededor de una prueba de hipótesis, es necesario tener cierta precaución y no convertir la simulación en un procedimiento ya que es fácil verla como un paso a paso, que puede conducir fácilmente a asumirla como un procedimiento algorítmico al igual que la aplicación insensata de fórmulas.

5.3. Recomendaciones

Se recomienda un trabajo de simulación intenso previo al trabajo computacional que familiarice a los estudiantes con los conceptos de variabilidad y distribución muestral. Este trabajo se puede iniciar en los primeros años de la educación secundaria, siempre en la perspectiva de informalidad que asumimos en esta investigación.

Las pruebas de hipótesis abarcan demasiados conceptos y son aplicables a diferentes distribuciones y situaciones, en esta investigación centramos los talleres en las pruebas de hipótesis para medias, tanto proporciones en el caso discreto, como media en el caso continuo. Las simulaciones no tienen alto grado de complejidad. Se realizaron tres talleres para variable discreta, y solo se trabajó un taller para variable continua, a partir de una simulación pre-establecida. Es conveniente ampliar el repertorio de talleres, agregando más material sobre variables continuas, e inclusive otras situaciones de fácil simulación, como por ejemplo, pruebas de hipótesis para comparación de medias.

El *software* estadístico *Fathom*, es la base de las simulaciones y las actividades realizadas, pero presenta el inconveniente de no tener licencia libre y aunque no es muy costoso puede estar fuera del alcance de muchas instituciones educativas. Por esta razón, es recomendable la consecución de un *software* de simulación estadístico de licencia libre, al cual puedan ajustarse las actividades. Puede pensarse en trabajar con Excel, no obstante que no tenga el carácter dinámico y de interacción que caracteriza a *Fathom*.

Referencias bibliográficas

- Batanero, C., Godino, J. D., Green, D. R., Holmes, P., & Vallecillos, A. (1994). Errores y dificultades en la comprensión de los conceptos estadísticos elementales. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.
- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Universidad de Granada.
- Batanero, C. (2009). La simulación como instrumento de modelización en probabilidad. *Revista Educación y Pedagogía*, 15(35).
- Batanero, C., Díaz, C. (2015). Aproximación informal al contraste de hipótesis.
- Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil.
- Biggs, J. B. y Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO taxonomy*. Nueva York: Academic Press.
- Biggs, J. B. (1988). The role of the metacognition in enhancing learning. *Australian Journal of Education*, 32,127-138.
- Biggs, J.B. y Collis (1991). Multimodal Learning and the Quality of Intelligent Behavior, en Rowe, H. (ed.) *Intelligence: Reconceptualization and Measurement*. LEA, Australian Council for Educational Research, 57-76.
- Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W.; Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review* 2(2), 98-113.
- Castro Sotos, A., Vanhoof, S., den Noortgate, W., & Onghena, P. (2009). How confident are students in their misconceptions about hypothesis tests. *Journal of Statistics Education*, 17(2), n2.

- Cross Bu, R. (2005). Simulación, Un enfoque práctico. Limusa (Noriega Editores).
- Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, 8 3-96.
- Fernández, F.; Monroy, O.; Rodríguez, L. (1998). Diseño, desarrollo y evaluación de situaciones problemáticas de estadística. El caso de los tests de hipótesis. Bogotá: una empresa docente.
- García y Ríos, N. (2013). Inferencias estadísticas informales en estudiantes mexicanos. En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 343-357). Granada, Departamento de Didáctica de la Matemática de la Universidad de Granada.
- García, V., Sánchez, E. (2014). Razonamiento inferencial informal: el caso de la prueba de significación con estudiantes de bachillerato. En González, María Teresa; Codes, Myriam; Arnau, David; Ortega, Tomás (Eds.), *Investigación en educación matemática* (pp. 345-354). Salamanca: Sociedad Española de Investigación en Educación Matemática.
- Huerta, M. (1999). Los niveles de Van Hiele y la taxonomía solo. In *Enseñanza de las Ciencias* (Vol. 17, pp. 291-309).
- Inzunsa, S., Jiménez, J. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista Latinoamericana de Investigación en Matemática Educativa, RELIME*, 16(2), 179-211.
- Jaimes, E. (2011). Niveles de razonamiento probabilístico con énfasis en la noción de distribución de estudiantes de secundaria en tareas de experimentación y simulación computacional. Tesis de Maestría en Ciencias con Especialidad en Matemática Educativa, no publicada, Centro de Investigaciones y Estudios Avanzados del Instituto Politécnico Nacional, Cinvestav-IPN, México, DF
- Lipson, K. (2000). The role of the sampling distribution in developing understanding of statistical inference. Doctoral Thesis unpublished, Swinburne University of Technology, Swinburne, Australia.
- Ministerio de Educación Nacional (2006) Estándares básicos de competencias en Matemáticas.

- Moore, D. S. (2005). Estadística aplicada básica. Antoni Bosch editor.
- Ortiz, V., Gaviria, M., & Gil, A. (2010). La enseñanza de la estadística inferencial. Un estudio de caso en la Pontificia Universidad Javeriana Cali. *Pensamiento Psicológico*, 1(6).
- Peskun, P. H. (1987). Constructing symmetric test of hypotheses. *Teaching Statistics*, 9 (1), 19-23.
- Pfannkuch, M. (2006). Informal inferential reasoning. En A. Rossman y B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil.
- Principios, N. C. T. M. (2000). Estándares para la educación matemática. Sociedad Andaluza de Educación.
- Ramírez Arce, G. (2008). Formas de razonamiento que muestran estudiantes de maestría de matemática educativa sobre la distribución normal mediante problemas de simulación en Fathom. *Revista electrónica de investigación en educación en ciencias*, 3(1), 10-23.
- Rossman, A. J. (2008). Reasoning about Informal Statistical Inference: One Statistician's View. *Statistics Education Research Journal*, 7(2), 5-19.
- Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In *Proceedings of the Seventh International Conference on Teaching Statistics*.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315.
- Vallecillos, A. (1992). Nivel de significación en un contraste estadístico de hipótesis. Un estudio teórico-experimental de errores en estudiantes universitarios. *Memoria de Tercer Ciclo*. Departamento de Didáctica de la Matemática. Universidad de Granada.
- Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Proceedings of the 52 session of the International Statistical Institute (Vol.2, pp. 201–204)*. Helsinki: International Statistical Institute.

Zieffler, A., Garfield, J. B., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 5-19. Online: www.stat.auckland.ac.nz/serj/.