# CONCEPCIONES PERSONALES DE LA DISTRIBUCIÓN BINOMIAL EN UN AMBIENTE COMPUTACIONAL: UN ESTUDIO CON PROFESORES EN FORMACIÓN

**MÓNICA ANDREA VERGARA CHÁVEZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**FACULTAD DE CIENCIAS**

**ESCUELA DE MATEMÁTICAS**

**BUCARAMANGA**

**2008**

# CONCEPCIONES PERSONALES DE LA DISTRIBUCIÓN BINOMIAL EN UN AMBIENTE COMPUTACIONAL: UN ESTUDIO CON PROFESORES EN FORMACIÓN

**MÓNICA ANDREA VERGARA CHÁVEZ**

**Trabajo de grado presentado para optar al título de Licenciada en Matemáticas**

**Director**
**GABRIEL YAÑEZ CANAL**
**Doctor en Matemática Educativa**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**
**FACULTAD DE CIENCIAS**
**ESCUELA DE MATEMÁTICAS**
**BUCARAMANGA**
**2008**

## AGRADECIMIENTOS

*En esta página quiero agradecer muy sinceramente...*

*A Dios por iluminarme y permitirme seguir adelante.*

*A mi familia por estar siempre a mi lado, por confiar en mí y por su apoyo incondicional.*

*Al Dr. Gabriel Yáñez Canal por sus grandes aportes a esta investigación y por haber puesto su confianza en mí.*

*A los estudiantes partícipes de esta investigación, gracias por la colaboración; sus participaciones fueron muy fructíferas en la investigación.*

*A EDUMAT-UIS, especialmente a Matemática Recreativa, por esos años de formación y crecimiento como persona y futura profesora.*

*A todos los profesores de la Escuela de Matemáticas por dar en mí parte de su conocimiento y la vocación docente.*

*A Abelito por brindarme su amor y comprensión durante este tiempo.*

*A María Isabel por las lecturas constructivas que hacías a la presente investigación.*

*A mis compañeros de la Licenciatura, los llevo en el corazón.*

*A mí mamá, mi papá y mi hermano por ser las personas más importantes en mi vida, por estar siempre a mi lado y porque siempre han confiado en mí, enseñándome que todo se logra con esfuerzo y dedicación.*

*¡LOS QUIERO MUCHO!*

## RESUMEN

**TÍTULO:** CONCEPCIONES PERSONALES DE LA DISTRIBUCIÓN BINOMIAL EN UN AMBIENTE COMPUTACIONAL: UN ESTUDIO CON PROFESORES EN FORMACIÓN.♦

**AUTOR:** MÓNICA ANDREA VERGARA CHÁVEZ♦♦

**PALABRAS CLAVES**: Distribución Binomial,
Regla del Producto,
Componente Combinatorio
Aproximación y Cálculo de probabilidades
Simulación computacional.

**DESCRIPCIÓN O CONTENIDO**

Esta investigación se orientó a caracterizar las concepciones personales construidos por los profesores en formación en lo referente a la distribución binomial en un ambiente computacional. La metodología utilizada durante el proceso de enseñanza fue la de Resolución de Problemas, la cual permitió que, mediante la búsqueda de la solución construyeran el nuevo concepto.

Las actividades se diseñaron en tres etapas, siguiendo el marco teórico de la Teoría de las Situaciones Didácticas de Brousseau (1986). Dichas actividades se implementaron a través de simulación computacional y experimentación física; para las simulaciones computacionales se utilizó Fathom (2000) como software didáctico.

El análisis de los datos se hizo teniendo en cuenta las siguientes categorías:

- Identificar situaciones binomiales.
- Determinar los parámetros *n* y *k* en situaciones binomiales.
- Tener presente la regla del producto para *n* eventos independientes donde *k* son los éxitos y *n-k* los fracasos.
- Tener en cuenta el componente combinatorio para contar el número de *k* éxitos y *n-k* fracasos al realizar *n* repeticiones del experimento.

Los resultados de esta investigación mostraron los avances alcanzados por los profesores en formación respecto a las distribuciones binomiales junto con algunas dificultades presentadas en el mismo. Además, se mostró la importancia de la simulación computacional en situaciones binomiales porque reflejó la comprensión, interpretación y argumentación de tipo probabilístico que se presentaron al momento de utilizar software computacional en clase.


---

♦ Trabajo de Grado.


♦♦ Facultad de Ciencias. Escuela de Matemáticas. Dr. Gabriel Yánez Canal.

## ABSTRACT

**TITTLE:** PERSONAL MEANINGS OF BINOMIAL DISTRIBUTION IN A COMPUTATIONAL ATMOSPHERE: A STUDY WITH PROFESSORS IN FORMATION. ♦

**AUTHOR:** MÓNICA ANDREA VERGARA CHÁVEZ.♦♦

**KEY WORDS:** Binomial Distribution
Normal Distribution
Central- limit Theorem
Aproximation and calculation of probabilities
Computacional Simulation

**DESCRIPTION OR CONTENT**

This investigation was oriented to characterize the personal meanings that acquired professors in formation around the binomial distribution in a computational atmosphere. The methodology used during the teaching process was the Problems Resolutions, which allows that, by means of the search of the solution, to construct the new concept.

The activities were designed in three times, continuing the theoric marc Didactic Situations of Brousseau (1986). This activities were implemented through computational simulation and physical experimentation; for computational simulation was used Fathom (2000) like didactic software.

The analyses of the dates were made continuing the following characteristics:

- To identify binomial situations.
- To determinate the parameters *n* and *k* in binomial situations.
- To have in mind the Product Rule for *n* independent events where *k* are the successes and *n-k* the failures.
- To have in mind the combinatory component to count the number of *k* successes and *n-k* failures to realize *n* repetitions of the experiment.

The results of this investigation shown the advance by the professor in formation respect to the binomial distributions and some difficult presented in the same. Also, it showed the importance of computational simulation in binomial situations because it shown the understanding, interpretation and argument of probabilistic type presented in the moment to use computational software in the classroom.


♦ Grade Work
♦♦ Faculty of Science. Licentiate in Mathematics. Dr. Gabriel Yánez Canal.

## TABLA DE CONTENIDO

## LISTADO DE FIGURAS

## LISTA DE ANEXOS

# PRESENTACIÓN

La variable aleatoria binomial es una variable discreta que sirve de modelo para una gran cantidad de situaciones prácticas, lo que hace que su cabal entendimiento se constituya en un objetivo básico de la enseñanza de la probabilidad. Iniciamos este apartado con una explicación breve sobre las situaciones aleatorias que se pueden modelar con la distribución binomial.

Suponga que se tiene un experimento aleatorio que consta de *n* repeticiones y cumple las siguientes características:

- Tiene sólo dos posibles resultados: éxito y fracaso (Ensayos Bernoulli).
- Cada repetición es independiente de las anteriores, es decir, el resultado de cada repetición no afecta a las demás.
- La probabilidad de obtener un éxito es constante, es decir, tiene el mismo valor entre una repetición y otra.

Un experimento con las anteriores características tiene asociada una variable aleatoria *X* que expresa el número de éxitos obtenido en *n* repeticiones, la cual sigue una distribución Binomial con parámetros *(n,p)* donde *n* indica el número de repeticiones y *p* la probabilidad de obtener éxito. La probabilidad de obtener *k* éxitos está dada por la función de probabilidad $P(X=k)=\binom{n}{k}p^k q^{n-k}$, donde $q=1-p$, obtenida como una consecuencia de la regla de producto generalizada para *n* eventos independientes y de la consideración de todos los casos en que se obtienen *k* éxitos favorables o éxitos.

Pero… *¿Qué sucede con el combinatorio* $\binom{n}{k}$ *al aumentar el valor de n y k respectivamente?*

Para responder a esta pregunta, es necesario tener presente que $\binom{n}{k} = \dfrac{n!}{(n-k)!k!}$ donde $n! = 1*2*3*...*n$ y $0! = 1! = 1$. Debido que para valores grandes de *n* y *k,* el combinatorio $\binom{n}{k}$ se hace difícil de calcular, siendo necesario aproximar el valor de probabilidad de algún modo. Así fue como, De Moivre-Laplace afirmaban que "*la curva de la variable normal debe ajustarse estrechamente al histograma de probabilidades binomiales mediante el área bajo la curva; para ello la media y la varianza de la variable normal y la variable binomial deben ser iguales".* Más formalmente, el Teorema De Moivre-Laplace afirma lo siguiente:

*Sea* $X_i$ *una variable aleatoria de Bernoulli y* $S_n = \sum_{i=1}^{n} X_i$ *una variable aleatoria con distribución binomial* $B(n, p)$ *cuya esperanza matemática es* $E(S_n) = np$ *y cuya varianza es* $Var(S_n) = npq$*. Cuando* $n$ *tiende a infinito, la función de distribución* $S_n$ *tiende a la distribución normal con la misma media y varianza, esto es* $N(np, npq)$.

Para conocer las concepciones personales de los estudiantes respecto a la distribución binomial, se realizó la investigación con estudiantes de Licenciatura en Matemáticas de la Universidad Industrial de Santander, quienes en el año 2007, cursaban la asignatura Estadística dirigida por el profesor Gabriel Yáñez Canal, director de esta investigación.

La investigación tenía por objeto conocer las concepciones personales de los estudiantes respecto a la distribución binomial en situaciones reales utilizando simulación computacional; por esa razón tuvimos en cuenta los siguientes elementos, los cuales se convirtieron en nuestras categorías de estudio, para detectar las posibles dificultades presentadas en situaciones binomiales, ellas fueron:

- Identificar el experimento Bernoulli asociado a la situación binomial
- Establecer los parámetros *n* y *p* de la situación binomial presentada.
- Utilizar la regla generalizada del producto para *n* eventos independientes donde *k* son los casos exitosos y *n-k* los fracasos.
- Tener presente el componente combinatorio $\binom{n}{k}$ que cuenta el número de veces que se obtiene *k* éxitos y *n-k* fracasos al realizar *n* repeticiones del experimento.

Más concretamente, y partiendo del análisis de datos realizado que tuviera en cuenta las anteriores categorías, la pregunta de investigación surgida fue la siguiente: *¿Qué dificultades presentan los profesores en formación al enfrentarse con situaciones problemas de tipo binomial utilizando simulación computacional?*

Para responder a esta pregunta, se realizaron actividades que llevaran a los estudiantes a enfrentarse con situaciones problemas de tipo binomial para que, en primera medida, las establecieran como tales, establecieran los respectivos parámetros *n* y *p y* tuvieran presente el componente combinatorio y probabilístico al utilizar software computacional, con el fin de mostrar las argumentaciones de los estudiantes y de allí extraer las ideas, dificultades y los errores presentados durante el proceso de enseñanza. Por esta razón, se adoptó como metodología de estudio la Resolución de Problemas y el marco teórico de las Situaciones Didácticas de Brosseau (1986).

Las actividades desarrolladas integraron la teoría con la práctica buscando que los estudiantes construyeran las concepciones personales válidas de la distribución binomial y su aplicabilidad a situaciones cotidianas; por ello se utilizó simulación física (modelos de urna) y simulación computacional. Durante los periodos de trabajo práctico se utilizó Fathom como software didáctico.

Esta investigación está organizada en cinco capítulos. El primer capítulo titulado "*Antecedentes"* describe algunas investigaciones previas relacionadas con el enfoque frecuencial de probabilidad, la distribución binomial, la aproximación normal de la binomial e intervalos de confianza.

El segundo capítulo titulado *"Diseño de la Investigación"* presenta la estructura de la investigación, comentando la metodología, el marco teórico, el método de recolección y análisis de datos, los estudiantes partícipes, las actividades y evaluaciones llevadas a cabo durante la investigación. En este apartado se comentan los objetivos de las situaciones problemas presentadas en las actividades y de la evaluación final.

El tercer capítulo titulado *"La distribución binomial",* presenta las argumentaciones dadas por los estudiantes al enfrentarse a situaciones binomiales, lo que lleva a comentar las dificultades presentadas, los logros alcanzados y los errores que persisten finalizando el proceso de enseñanza.

El cuarto capítulo titulado *"Manejo de Fathom y su uso en la resolución de situaciones problemas binomiales"* presenta algunas programaciones realizadas por los estudiantes durante el proceso de enseñanza. El objetivo de este capítulo es analizar el efecto de la simulación computacional en la construcción de concepciones personales respecto a la distribución binomial, puesto que la utilización de Fathom refleja la interpretación del estudiante de la situación, lo que muestra la comprensión que los estudiantes tienen de la distribución binomial.

El quinto capítulo titulado *"Conclusiones"* presenta los resultados alcanzados en esta investigación. Después de presentar la bibliografía utilizada en esta investigación, se encuentran los anexos. En ella se muestran las actividades y las dos evaluaciones realizadas por los estudiantes.

# 1. ANTECEDENTES

A continuación se describen algunas investigaciones previas relacionadas con el enfoque frecuencial de la probabilidad, la distribución binomial, la aproximación normal de la binomial y los intervalos de confianza.

**Mantilla y Martínez (2007)**

La investigación de Mantilla y Martínez (2007) en su trabajo de grado presentado para merecer el título de Licenciado en Matemáticas *“Construcción de significados del concepto de probabilidad en un ambiente computacional. Una experiencia con profesores en formación”* presenta una experiencia de aula cuyo propósito era caracterizar los significados que los profesores en formación le concedían a la probabilidad frecuencial en un ambiente computacional.

En la investigación participaron 19 estudiantes de la carrera de Licenciatura en Matemáticas de la Universidad Industrial de Santander, quienes cursaban la asignatura Estadística en el primer semestre de 2007. La investigación fue de tipo cualitativo y la metodología utilizada fue la de resolución de problemas mediado con simulación computacional.

En primera medida, se presentó un cuestionario diagnóstico para identificar las concepciones e intuiciones que presentaban los profesores en formación antes de iniciar el proceso de instrucción. Posteriormente, se desarrollaron dos actividades: *El Casino y un Cuestionario de ocho preguntas* en donde se abordaron temas como: la regla del producto para eventos independientes, la regla de la suma, el principio fundamental del conteo, el conteo, la distribución binomial e ideas fundamentales del enfoque frecuencial de la probabilidad.

Las actividades planteadas se basaban fundamentalmente en el marco de *la Teoría de las Situaciones Didácticas de Brousseau (1986)*, y queriendo hacer una integración entre la teoría y la práctica, las actividades se desarrollaron en sesiones teórico-prácticas utilizando software didáctico Fathom (Finzer 2000) como simulador computacional.

Algunas conclusiones de esta investigación son las siguientes:

- Aunque los estudiantes concibieron el valor de probabilidad desde el punto de vista frecuencial, ellos se identificaron más con el enfoque clásico de probabilidad; en algunos casos se presentaban contradicciones cuando los procedimientos algebraicos no coincidían con lo teórico y computacional.

- Se resaltó la importancia de la simulación computacional para conjeturar o refutar algunas conclusiones de los problemas, ayudando a visualizar la situación, analizando la variabilidad del experimento y en las argumentaciones generadas.

- En algunas ocasiones se presentó dificultad a la hora de representar en Fathom el experimento real, lo que dificultó la realización de las sesiones. Además, los profesores en formación, para transformar el experimento en el programa, planteaban la situación en lápiz y papel y posteriormente traducían lo escrito en Fathom.

**Alvarado (2007)**

La investigación de Alvarado (2007) en su tesis doctoral *"Significados Institucionales y Personales del Teorema Central del Límite en la Enseñanza de Estadística en Ingeniería"* presenta los resultados de un estudio sistemático del significado institucional pretendido e implementado sobre el Teorema Central del

Límite dirigido a alumnos de carreras de ingenierías de la Universidad Católica de la Santísima Concepción de Chile.

“El objetivo de esa tesis era diseñar y evaluar un proceso de estudio dirigido sobre el Teorema Central del Límite a estudiantes de ingeniería que tuvieran en cuenta la evolución histórica del teorema, el contexto educativo, tipos de estudiantes y el marco teórico de referencia” (Alvarado, 2007; pág. 18); para ello realizó una descripción epistemológica de la evolución histórica del Teorema Central del Límite, después hizo un estudio exhaustivo sobre el significado del Teorema Central del Límite en libros de estadística para ingenierías y construyó una propuesta didáctica para el Teorema Central del Límite.

La propuesta didáctica se llevó a cabo en el segundo semestre del año 2005 y contó con la participación de 132 estudiantes de segundo año de ingenierías (acuicultura, marítimo y portuario, informática, industrial y civil) de la Universidad Católica de la Santísima Concepción de Chile, quienes cursaban la asignatura de Estadística. Los estudiantes participantes poseían conocimientos previos de probabilidad, puesto que la asignatura cursante tiene como prerrequisito el curso de probabilidades.

La investigación era de tipo cualitativo con una metodología intuitiva, basada en la simulación manual y computacional. Las tres actividades llevadas a cabo se fundamentaron en el marco teórico del enfoque ontosemiótico de la cognición e instrucción matemática (Godino, 2003) teniendo en cuenta las configuraciones epistémicas (manipulativa, algebraica, argumentativa). Además, elaboró un cuestionario con preguntas de selección múltiple y problemas abiertos con el fin de evaluar el aprendizaje de los estudiantes, comparando el significado institucional previsto con el observado en los estudiantes.

Algunas conclusiones generadas de esta investigación son las siguientes:

- El estudio de los libros de texto y la descripción de la evolución histórica del Teorema Central del Límite reflejan, por un lado, la complejidad del teorema al requerir otros conceptos estadísticos y sus relaciones; y por otro, muestra las numerosas aplicaciones del Teorema Central del Límite (necesidad de aproximar la distribución binomial para valores grandes del parámetro *n*, hasta llegar a la generalización del teorema para la suma de variables aleatorias independientes) en la teoría de la probabilidad y la estadística aplicada.
- Los alumnos presentaron una mejor comprensión de la distribución aproximada de la media muestral de variables aleatorias independientes provenientes de poblaciones binomiales.
- No se presentaron dificultades en el lenguaje simbólico. El procedimiento algebraico y la tipificación fueron los más utilizados.
- La simulación computacional apoyada en el uso de Excel ayudó a la argumentación de los estudiantes mediante la representación gráfica de las situaciones planteadas.
- En situaciones binomiales, los alumnos no aplicaban corrección de continuidad; además, no fueron capaces de dar una decisión con base al análisis del Teorema Central del Límite.
- Los alumnos confundieron el parámetro con el estadístico y persistió la heurística de la representatividad[1]. Algunos estudiantes definieron inadecuadamente la variable aleatoria que modela el experimento y reconocieron parcialmente la distribución de probabilidad.
- Los estudiantes no diferenciaron entre una variable aleatoria discreta de una continua. En algunos casos, consideraron en el cálculo de la probabilidad la varianza en la tipificación en lugar de la desviación estándar.

[1] La heurística de la representatividad descrita por Kahneman *et al* (1982) consiste en evaluar la probabilidad de un suceso en base a la representatividad del mismo respecto a la población que proviene. En este tipo de razonamiento se prescinde del tamaño de la muestra y, con ello, del estudio de la variabilidad del muestreo, produciéndose una confianza indebida en pequeñas muestras.

**Tauber (2001)**

La investigación de Tauber (2001) en su tesis doctoral *La Construcción del Significado de la distribución Normal a partir de Actividades de Análisis de Datos* presenta los resultados de una experiencia de enseñanza de la distribución normal con estudiantes de primer año universitario, dentro de un curso de análisis de datos mediado con simulación computacional y fundamentado en el marco teórico referente al significado y comprensión de los objetos matemáticos (Godino y Batanero, 1994; 1998).

La investigación se llevó a cabo con alumnos de primer año universitario que cursaban una asignatura de libre configuración de 9 créditos desarrollado en la Facultad de Educación de la Universidad de Granada durante 1998-1999 y 1999-2000.

El objetivo de la investigación era *la enseñanza y aprendizaje de la distribución normal*, más específicamente, se interesó en *la problemática que presentaba este tema a los alumnos que realizaban un curso introductorio a la estadística con un enfoque basado en el uso de programas estadísticos.* Por tal motivo, las sesiones tenían carácter teórico –práctico.

Durante la investigación se utilizó el programa Statgraphics para simular las situaciones planteadas a los alumnos a partir de ficheros de datos y realizar un análisis de los datos. La metodología llevada a cabo fue la resolución de problemas partiendo de un enfoque cuasi-experimental al inicio, siendo descriptivo y experimental, pasando a un enfoque interpretativo y explicativo en la última fase.

Una vez recogidos los datos (diario de observación de la investigadora, informes escritos de las actividades teóricas y prácticas hechas por los estudiantes, un cuestionario, una prueba de ensayo para ser resuelta por medio de la simulación computacional) se dio paso al análisis de datos con el fin de describir las principales características en el significado personal que los alumnos habían construido acerca de la distribución normal utilizando el paquete estadístico Statgraphics.

El análisis de la información obtenida se centró en:

- Fijar el significado institucional de referencia que dieron a la distribución normal, realizando para ello, un análisis epistémico del tema en textos universitarios.
- Elaborar una secuencia didáctica sobre la distribución normal de acuerdo al marco teórico, incorporando la simulación computacional como herramienta didáctica, el significado de referencia, las características del curso, las experiencias previas de los profesores en el mismo y las características de los alumnos.
- Describir los elementos de significado efectivamente observados en la secuencia de enseñanza.

Para evaluar el conocimiento adquirido se utilizaron dos instrumentos:

- Un cuestionario escrito.
- Tres tareas relacionadas a un fichero de datos para ser resueltas con la ayuda del programa estadístico.

Algunas conclusiones generadas de esta investigación son las siguientes:

- Los estudiantes no dominaron conceptos previos como: simetría, el significado de área en los histogramas, la determinación del intervalo donde cae la mediana en una tabla de frecuencia, algunas propiedades básicas de la mediana o la identificación e interpretación de los distintos tipos de frecuencia; dificultando así la comprensión de la distribución normal.
- El uso del paquete estadístico Statgraphics para el segundo año ayudó a clarificar algunos conceptos, especialmente el estudio descriptivo y la

comparación visual al poder realizar simulaciones y establecer una relación más estrecha entre la teoría y la práctica.

- En algunas ocasiones se presentó dificultad en el manejo del paquete estadístico al no encontrar o conocer los comandos del programa, mostrando la complejidad que implica organizar una clase práctica.
- Se observó dificultad en la interpretación de los coeficientes de asimetría y de curtosis, de los cuartiles y percentiles. Además, resulta complicada la interpretación de probabilidades como áreas bajo la curva especialmente cuando los intervalos a calcular resultan de una operación entre intervalos.
- Los estudiantes se ayudaron de la representación gráfica para validar, comprobar y aplicar algunas propiedades de manera descriptiva, olvidando utilizar las debidas argumentaciones que subyacen de ellas.
- Se mostró la importancia del análisis de datos en diferentes campos de problemas con fines inferenciales: ajuste de modelos a datos, aproximación de modelos a variables discretas, la obtención de distribuciones muestrales exactas y aproximadas, estimación y ajuste; para que los alumnos apreciaran la estadística y las múltiples posibilidades de aplicación en diferentes campos.
- Se muestra que los alumnos aprenden a utilizar el software, pero presentan cierta dificultad en la discriminación entre datos empíricos y los modelos matemáticos, en interpretar ciertos gráficos y resúmenes estadísticos y escasa capacidad de análisis y síntesis.

**Olivo y Batanero (2007)**

La investigación de Olivo y Batanero (2007) titulada *Un estudio exploratorio de las dificultades de comprensión del intervalo de confianza*, presentó los resultados de un cuestionario desarrollado por 48 estudiantes de ingeniería del Instituto Tecnológico y Estudios Superiores de Monterrey. Su objetivo era evaluar las dificultades de comprensión, cálculo e interpretación de intervalos de confianza, incluyendo algunas propiedades o el cálculo de intervalos de confianza para algunos parámetros.

Como la investigación se centró en estudiantes de ingenierías, se realizó un análisis del contenido relacionado con intervalos de confianza en una muestra de 18 libros de estadística dirigida a la enseñanza en ingenierías para deducir propiedades y relaciones de los intervalos de confianza con otros objetos matemáticos, y los procedimientos de construcción de intervalos de confianza.

El cuestionario estuvo compuesto por 11 ítems de opción múltiple, cuyos objetivos eran:

- Detectar algunos sesgos en la definición de intervalos de confianza (ítem 1).
- Evaluar la comprensión del efecto del tamaño de muestra sobre la precisión (ancho del intervalo) cuando se mantiene constante el nivel de confianza (ítem2).
- Evaluar la comprensión del efecto del nivel de confianza sobre el ancho del intervalo (ítem 3 y 4). Y el efecto de la varianza sobre el ancho de los intervalos (ítem 5).
- Evaluar la comprensión de la variabilidad del intervalo en diferentes muestras y la interpretación de su resultado (ítem 6).
- Evaluar el conocimiento procedimental en la construcción de intervalos de confianza para la media (ítem 9).
- Evaluar el conocimiento en la determinación de valores críticos en la distribución del estadístico (ítem10).
- Evaluar la interpretación de intervalos de confianza para toma de decisiones sobre la diferencia de medias (ítem 11).

Además, el cuestionario contenía cuatro ítems abiertos cuyo objetivo era evaluar las estrategias de resolución y argumentación de los estudiantes. Los resultados del cuestionario mostraron:

- Los estudiantes visualizaron los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial.
- Se notaba debilidad conceptual en la comprensión acerca de las relaciones de los distintos elementos asociados con los intervalos de confianza.
- Los estudiantes no concibieron el intervalo como un conjunto de valores plausibles del parámetro, llegando a justificaciones erradas para tomas de decisiones válidas sobre las hipótesis.

Olivo y Batanero (2007) recomiendan diseñar actividades basadas en la simulación para que los estudiantes exploren y experimenten el significado de los intervalos de confianza y el efecto del tamaño de muestra, varianza y nivel de confianza. Además, relacionar el tema con las distribuciones de probabilidad, mostrando la importancia de la distribución muestral en el cálculo de los intervalos de confianza.

**Observaciones acerca de los antecedentes a tener en cuenta en la investigación.**

De acuerdo con los antecedentes anteriormente descritos, nos permitimos identificar diversas dificultades e ideas importantes que soportan la presente investigación:

En cuanto a las dificultades que se presentan a la hora de enseñar el concepto de distribución (binomial, normal) se tienen las siguientes:

- Los conceptos previos a la distribución normal no están suficientemente claros en los estudiantes, lo que dificulta la apropiación del nuevo concepto. (Tauber, 2001).
- El enfoque clásico de probabilidad es aún más evidente en los estudiantes contrastado con el enfoque frecuencial del valor de la probabilidad.

Fundamentarse en la aritmetización de la probabilidad sin ninguna o poca intuición de su significado dificulta el paso al concepto de distribución (Mantilla y Martínez, 2007).

- La inferencia estadística se dificulta porque los estudiantes realizan un estudio de una parte de la población (muestra), conjeturan y dan conclusiones erradas de la misma al no tener en cuenta el tamaño de la muestra, el comportamiento y el valor de los parámetros del experimento. (Tauber, 2001; Alvarado, 2007).
- Aunque los alumnos resuelven diversos problemas de probabilidad, no distinguen entre una distribución discreta de una distribución continua, lo que dificulta concebir la probabilidad de una distribución continua como una aproximación de una distribución discreta al aumentar el tamaño muestral (Alvarado, 2007).

Algunas recomendaciones recogidas de los antecedentes son las siguientes:

- La simulación física (material concreto) y la computacional (paquete o software estadístico), ayudan al estudiante a comprender mejor el problema y a visualizar algunas propiedades que de forma teórica sería imposible percibir. De allí la importancia de las sesiones prácticas en las clases de probabilidad. (Tauber, 2001; Mantilla y Martínez, 2007; Alvarado, 2007).
- Plantear a los estudiantes situaciones problemas reales muestran la aplicabilidad de la probabilidad en diferentes campos (Tauber, 2001; Mantilla y Martínez, 2007; Alvarado, 2007; Olivo y Batanero, 2007).

# 2. DISEÑO DE LA INVESTIGACIÓN

A continuación presentamos el marco teórico y la metodología que sustentan esta investigación. De igual forma, se presenta el diseño y la descripción de las sesiones y actividades realizadas durante el proceso de enseñanza.

## 2.1. MARCO TEÓRICO

### 2.1.1. Significado y comprensión de la distribución binomial

Como el desarrollo teórico de esta investigación se centró en la distribución binomial mediado por simulación computacional, nuestra atención principal era analizar las concepciones personales construidas por los estudiantes de la Licenciatura en Matemáticas respecto a la distribución binomial.

Para ello es necesario tener en cuenta que la distribución binomial tiene como parámetros: la probabilidad constante de éxito *p* y el tamaño muestral *n* y que la función de probabilidad, dada por la expresión $P(X=k)=\binom{n}{k}p^k q^{n-k}$ donde $q=1-p$, tiene inmersa los siguientes componentes:

- **El experimento Bernoulli básico:** da cuenta de la situación como variable aleatoria discreta, necesaria para la distribución binomial al ser ésta un conjunto de experimentos Bernoulli. Identificar situaciones binomiales implica identificar el experimento Bernoulli asociado a la situación.

- **Los parámetros *n* y *p* en situaciones binomiales:** determina el comportamiento de la distribución binomial. En algunos casos suelen presentarse inconvenientes al establecerlos al no identificar el experimento básico Bernoulli de la situación o tomar como parámetros valores de la situación que no son necesarios.

- **El componente combinatorio:** El combinatorio $\binom{n}{k}$ cuenta el número de veces que se obtiene k éxitos y *n*-k fracasos al realizar *n* repeticiones del experimento. Se trata de la expresión en la binomial de la regla de la suma para eventos disyuntos donde el orden de los resultados juega un papel importante para modelar adecuadamente el valor de probabilidad.

- **El componente probabilístico:** hace referencia a la regla generalizada del producto para *n* eventos independientes de los cuales *k* son exitosos y *n-k* son fracasos.

### 2.1.2. Significado y comprensión del Teorema De Moivre-Laplace

Para el cálculo de probabilidades binomiales en situaciones con tamaños muestrales pequeños se utiliza la función de probabilidad $P(X = k) = \binom{n}{k} p^k q^{n-k}$. Sin embargo, cuando el tamaño muestral aumenta ése valor se hace difícil calcular mediante la expresión antes descrita, lo que lleva a que se trate de aproximar ese valor mediante otro procedimiento.

El inconveniente de la función de probabilidad $P(X = k) = \binom{n}{k} p^k q^{n-k}$ radica en el componente combinatorio $\binom{n}{k}$ puesto que al aumentar el valor de *n*, éste aumenta progresivamente llegando al caso de ser imposible de calcular. Por esa razón, no se utiliza esa expresión y se aproxima dicho valor con otro procedimiento; de hecho, De Moivre-Laplace afirmaban que "*la curva de la variable normal debe ajustarse estrechamente al histograma de probabilidades binomiales mediante el área bajo la curva; para ello la media y la varianza de la variable normal y la* *variable binomial deben ser iguales".* Más formalmente, el Teorema De Moivre-Laplace afirma lo siguiente:

*Sea $X_i$ una variable aleatoria de Bernoulli y $S_n = \sum_{i=1}^{n} X_i$ una variable aleatoria con distribución binomial $B(n,p)$ cuya esperanza matemática es $E(S_n) = np$ y cuya varianza es $Var(S_n) = npq$. Cuando $n$ tiende a infinito, la función de distribución $S_n$ tiende a la distribución normal con la misma media y varianza, esto es $N(np, npq)$.*

Debido a la similitud entre los histogramas binomiales y la curva normal al aumentar el valor de *n* se procede a utilizar la distribución normal para aproximar el valor de probabilidad binomial.

### 2.1.3. Significado y comprensión de la distribución normal

La distribución normal se puede obtener al considerar el modelo básico de una variable aleatoria binomial cuando el número de repeticiones se vuelve cada vez más grande.

Sea $X \sim N(\mu, \sigma^2)$ donde $E(X) = \mu$ es el valor con mayor probabilidad y centro de la curva normal, y $V(X) = \sigma^2$ es una medida de dispersión de los datos, éstos son los parámetros de la variable normal; además, se puede establecer $Z = \frac{X-\mu}{\sigma}$ como una variable normal estándar con $E(Z) = 0$ y $V(X) = 1$ que representa las distancia de *X* a partir de su media en términos de desviaciones estándar. Para establecer el valor de probabilidad en una variable normal se calcula el área bajo la curva mediante la expresión $P(X \le x) = P(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}) = P(Z \le z)$.

Ahora bien, como el valor de probabilidad binomial se aproxima mediante la normal cuando el tamaño muestral es grande, se hace necesario aplicar procedimientos normales conservando la media y varianza de la situación binomial. Así pues, si $X \sim B(n, p)$ se procede de la siguiente manera para establecer el valor de probabilidad:

- Establezco la variable normal estándar de la variable binomial *X* mediante la expresión $Z = \dfrac{X - np}{np(1-p)}$ donde $E(X) = np$ y $V(X) = np(1-p)$.
- Calculo los valores de probabilidad pedidos mediante la expresión $P(X \leq x) = P(\dfrac{X - np}{\sqrt{np(1-p)}} \leq \dfrac{x - np}{\sqrt{np(1-p)}}) = P(Z \leq z)$.

Por el Teorema Central del Límite se tiene que si el número de experimentos *n* es grande, entonces $X \sim N(np, np(v1-p))$. Además, como la distribución binomial es discreta y se busca aproximar su probabilidad mediante la distribución normal que es continua, se hace necesario un ajuste de esa probabilidad para mejorara la exactitud de la aproximación, el cual se conoce como corrección de continuidad. La corrección de continuidad se calcula tomando 0.5 de más del valor de probabilidad para tener la totalidad del histograma bajo la curva normal, porque de otra manera faltaría una parte del área bajo la curva que no sería tenida en cuenta y generaría error en la probabilidad.

**Algunas notas aclaratorias.**

- El establecimiento de los parámetros *n* y *p* en situaciones binomiales son fundamentales puesto que de ellas depende la obtención correcta de las media y varianza de la situación. La media y la desviación (la raíz cuadrada de la varianza) son los valores claves al momento de establecer el valor aproximado de probabilidad binomial.

- Concebir el área bajo la curva como el valor de probabilidad servirá para que los estudiantes tengan una idea de los valores obtenidos como porcentajes.
- En toda estimación, va a existir un margen de error mostrando que los valores obtenidos serán un valor aproximado del valor teórico de probabilidad.

## 2.2. METODOLOGÍA DE LA INVESTIGACIÓN

Al ser esta investigación de tipo cualitativo, se quiso analizar las concepciones personales construidas por los estudiantes respecto a la distribución binomial. Nuestro interés se centra en los razonamientos de los estudiantes alrededor del anterior concepto mediado por simulación computacional.

### 2.2.1. Fases de la Investigación

Durante el proceso de enseñanza – aprendizaje, se distinguieron dos etapas fundamentales:

#### 2.2.1.1. Diseño y aplicación de las actividades durante el proceso de enseñanza.

Queriendo que los estudiantes comprendieran el concepto de distribución binomial, se realizaron actividades que permitieran a los estudiantes enfrentarse a tales situaciones mediante la manipulación de material concreto y la utilización de simulación computacional. La metodología llevada a cabo fue la resolución de problemas mediada con un enfoque práctico y computacional.

El diseño y la implementación de las actividades fueron realizadas por el profesor Gabriel Yáñez Canal, profesor del curso de estadística y director de esta investigación. La autora de este trabajo asistió a las clases y tomó nota de las acciones y argumentos de los estudiantes.

. Los  recursos utilizados durante el proceso de enseñanza fueron:

- Los modelos de urnas formados con caja de cartón y bolas de ping pong
- El software didáctico Fathom (Finzer, 2000) utilizado para simular los experimentos.

Cada una de las actividades fueron socializadas en el aula de clase; con anterioridad, los estudiantes entregaban los informes escritos y las programaciones en Fathom con el fin de observar los errores, dificultades y avances que presentaron los estudiantes antes de la *institucionalización* de las actividades.

Las actividades se propusieron a través de guías de trabajo. En ellas se abordaron temas de la distribución binomial, la importancia del tamaño muestral y el valor de *p* al calcular el valor aproximado de la probabilidad, la distribución normal, la estimación puntual y los  intervalos de confianza.

#### 2.2.1.2. Evaluación y análisis de los resultados

Finalizado el proceso de enseñanza, se realizó un cuestionario de evaluación.  La estructura del cuestionario de evaluación fue la siguiente:

- *Situaciones problemas que debían resolverse bien en forma algebraica o con ayuda de Fathom*. Se pretendía conocer la capacidad de modelación de los estudiantes a situaciones problema que respondían a un modelo binomial. Igualmente se quería saber si utilizaban sólo procedimientos algebraicos o ambos simultáneamente.

- *Proposiciones de Verdadero o Falso.* El objetivo primordial era observar la comprensión que los estudiantes tenían de las propiedades de la distribución binomial, de la normal y de su papel como aproximación de la distribución binomial, evaluando los argumentos dados en las respuestas.

### 2.2.2. Estructura de las actividades realizadas

Las actividades realizadas se diseñaron teniendo en cuenta las tres etapas planteadas por Brousseau (1986) en su Teoría de las Situaciones Didácticas, citado por Mantilla y Martínez (2007):

- *Situación de Acción (Fase de trabajo individual):* los estudiantes intentan resolver la situación problema planteada en forma individual.

- *Situación de Formulación (Fase de trabajo en pequeños grupos):* los estudiantes, por parejas o pequeños grupos, formulan algunas estrategias o posibles soluciones al problema para ser discutido en el grupo. El objetivo de esta fase es llegar a un acuerdo de la posible solución del problema formada o enriquecida entre los integrantes del grupo. La posible solución será socializada en la siguiente fase.

- *Situación de Validación (Fase de socialización):* se expone a los demás las posibles soluciones halladas por cada grupo en la fase anterior con el fin de aceptarlas, rechazarlas o modificarlas para llegar a construir el concepto.

Por último se realizaban sesiones de *institucionalización*, a cargo del profesor de la asignatura, con el fin de concretar los resultados teóricos que surgieron a través del proceso, es decir, se llegaba al concepto verdadero partiendo de los respuestas dadas por los estudiantes durante el desarrollo de las actividades.

Al respecto Brousseau (1986), citado en Mantilla M. y Martínez M. (2007), comenta: *"(…) la institucionalización define las relaciones que pueden tener los comportamientos o las producciones "libres" del alumno con el saber cultural o científico y con el proyecto didáctico; da una lectura y les da un status (…)"*

### Población y muestra

La población de interés en este estudio son los profesores de matemáticas en formación, más específicamente, los estudiantes de la carrera de Licenciatura en Matemáticas de la Universidad Industrial de Santander. La muestra estuvo formada por 16 estudiantes, quienes en el primer semestre del 2007 cursaban la asignatura Estadística. Los nombres de los estudiantes participantes de esta investigación han sido cambiados para reservar su identidad.

### Instrumentos de recolección de datos

Durante la investigación se utilizaron los siguientes instrumentos de recolección de datos:

- Diario de campo.
- Se filmaron o grabaron una buena cantidad de las sesiones de trabajo, en especial, las discusiones que se dieron en la situación de institucionalización.
- Informes escritos de actividades teóricas y prácticas.
- La programación de las actividades realizada en Fathom.
- El cuestionario de evaluación.

### Técnicas de análisis de datos

Como esta investigación es de tipo cualitativo, se analizaron las respuestas y argumentaciones de los estudiantes a las preguntas abiertas, problemas planteados, guías de trabajo y cuestionario de evaluación. De igual forma, se analizaron los procedimientos algebraicos y las programaciones en Fathom realizadas por los estudiantes para resolver las situaciones problemas planteadas.

Para el análisis de los resultados se establecieron dos núcleos de interés:

- *La Distribución Binomial.* A través de las respuestas de los estudiantes se analiza las concepciones construidas y su comprensión en situaciones binomiales y, particularmente, de la capacidad de reconocer sus parámetros a la hora de modelar una situación real. Además, se comenta las dificultades y errores presentados al trabajar con situaciones binomiales.

- *Manejo de Fathom y su uso en la resolución de situaciones problemas binomiales.* Este apartado, se describen las dificultades que tuvieron en la interpretación de la situación real al momento de programar en Fathom y además, analizar las distintas programaciones realizadas que dan cuenta del uso de Fathom en situaciones binomiales.

## 2.3. DISEÑO Y DESARROLLO DE LAS SESIONES Y ACTIVIDADES REALIZADAS.

### 2.3.1. Primera Sesión. La Distribución Binomial

Esta sesión tuvo una duración de cuatro horas. En ella se implementó un cuestionario de seis actividades (Ver Anexo 1) con el objetivo de analizar el comportamiento de la distribución binomial y descubrir la manera como los parámetros *n* y *p* afectan a dicha distribución. El análisis fue de tipo gráfico. En esta sesión los estudiantes realizaron las cuatro primeras actividades del cuestionario; las dos últimas se dejaron como actividad extraclase.

Con el objetivo de generar intuición en los estudiantes acerca del comportamiento de *n* y *p,* se propusieron cuatro actividades que los estudiantes debían resolver con ayuda de Fathom. En esta sesión se pedía a los estudiantes construir la gráfica de la distribución binomial, que analizaran el comportamiento de la gráfica cuando varía el valor de *p* por medio del *slider*[2] y obtuvieran algunas conclusiones.

[2] El *slider* es un comando de Fathom, sirve para que un parámetro o valor de interés pueda tomar un sinnúmero de valores dentro de un determinado rango; en nuestro caso, el parámetro *p* varía entre 0 y 1.

Para el desarrollo de las actividades, se introdujeron algunas funciones básicas de Fathom como *caseIndex, Combinations, graph* con el fin de construir la gráfica de la distribución binomial; además, se introdujo el comando *slider* para hacer variar el valor de *p* entre 0 y 1.

A continuación describimos las fases desarrolladas en esta actividad:

- ***Situación de Acción (Fase de Trabajo Individual)***

La primera y segunda actividad buscaba que los estudiantes, de manera individual, construyeran la gráfica de la distribución binomial teniendo *n* fijo y variando el valor de *p* entre 0 y 1. El objetivo era analizar el comportamiento de la distribución cuando variaba el valor de probabilidad *p* y el tamaño muestral *n* era fijo; nos interesaba que los estudiantes dedujeran algunas propiedades del comportamiento de la distribución binomial.

***Primera Actividad***

*Utilizando Fathom construya una colección de datos que le permita realizar un gráfico de la distribución binomial. Como valor n tome el producto de 10 por el número que le adjudique el profesor. Como la idea es que el parámetro p varíe libremente entre 0 y 1, utilice un slider para definirlo. Para ello basta arrastrar el icono de slider, definirlo como p y restringir su intervalo de variación entre 0 y 1 en la misma forma como se cambian las escalas en los ejes de los gráficos.*

***Segunda Actividad***

*Construya el gráfico que relaciona el valor de probabilidad con los posibles valores de k de la distribución binomial y active el slider para observar la forma como cambia la distribución binomial. Saque algunas conclusiones respecto a la relación que pueda existir entre la forma del gráfico de la distribución binomial y los diferentes valores de p.*

- ***Situación de Formulación (Fase de Trabajo en Pequeños Grupos)***

La tercera y cuarta actividad invitaba a los estudiantes a debatir las conclusiones construidas individualmente con otro compañero. El objetivo era compartir, construir y mejorar las conclusiones para ser discutidas en la siguiente fase y además analizaran el comportamiento de las medias muestrales a medida que se aumentaba el tamaño muestral.

***Tercera Actividad***

*Reúnase con dos compañeros que hayan trabajado con valores distintos de n, analicen sus respectivos gráficos y escriban sus conclusiones respecto a la forma como los parámetros p y n afectan el gráfico.*

***Cuarta Actividad***

*Repita las actividades anteriores cambiando los k-valores por las proporciones o probabilidades posibles de éxitos. Escriba sus conclusiones individuales y participe en la escritura de las conclusiones de su grupo de trabajo.*

La situación antes descrita hace referencia al comportamiento gráfico de las medias muestrales binomiales.

- ***Situación de Validación (Fase de Socialización)***

Finalizada las cuatro actividades, se llevó a cabo la socialización de las argumentaciones construidas por los estudiantes, en forma individual y por parejas, acerca del comportamiento de la gráfica binomial cuando variaban los parámetros $n$ y $p$. Finalizada la fase de socialización se realizó un resumen de las principales conclusiones construidas culminado el proceso de enseñanza (Institucionalización).

La quinta y sexta actividad se dejaron como trabajo extraclase para ser entregado en la siguiente sesión. Estas actividades abordaban ideas intuitivas de las pruebas de hipótesis e intervalos de confianza y en cierta forma que recordaran la actividad del Casino (Mantilla y Martínez, 2007).

***Quinta Actividad***

*En una urna existe una cierta composición de bolas amarillas que se pretende averiguar. Para ello se realizan 100 extracciones con reposición y se obtienen 35 bolas amarillas.*

*a). Con la información obtenida ¿considera plausible el valor 0.29 para la proporción de amarillas? Explique su respuesta.*

*b) ¿Y el valor 0.42? Explique su respuesta.*

*c) Construya un intervalo de valores posibles (Intervalos de Confianza) y explique los motivos para su construcción.*

*Dé un algoritmo que le permita construir "intervalos de confianza" para estimar el valor de la probabilidad asociado a un cierto éxito.*

***Sexta Actividad***

*El rector de una universidad dice "el 99% de los estudiantes apoya el nuevo calendario académico". Para corroborar la información del rector se realizó una encuesta y se entrevistó a 200 estudiantes.*

*a). Si 180 de esos estudiantes manifestaron estar de acuerdo con el nuevo calendario, ¿se puede decir que el rector estaba en lo cierto?*

### 2.3.2. Segunda Sesión. Cálculo De Probabilidades

Esta sesión tuvo una duración de seis horas, en ella se implementó un cuestionario compuesto por 5 situaciones problemas (Ver Anexo 2) que buscaba analizar la importancia de las medidas de dispersión al comparar la variabilidad en un conjunto de datos en la distribución binomial, y establecieran valores de probabilidades binomiales en pequeñas muestras.

La primera situación fue realizada en el aula de clase, las otras cuatro actividades fueron trabajadas en parejas como informes para entregar, de todos modos las actividades fueron debidamente socializadas en clase de estadística.
En la segunda y tercera actividad se pedía a los estudiantes calcular la probabilidad de intervalos de valores en distribuciones binomiales $B(10{,}0.5)$ y $B(100{,}0.3)$. La idea era introducirlos en los intervalos de confianza y, de cierta manera, preparar el terreno para la aproximación de la binomial a través de la distribución normal. (Ver Anexo 2).

***Actividad 2***

*Dada la variable aleatoria X con distribución binomial B(10,0.5) calcule la probabilidad de los siguientes eventos:*

a) *Que la variable tome valores mayores que 2 y menores que 7, es decir, halle la probabilidad* $P(X\in(2,7))$.
b) *Que el variable tome valores mayores o iguales que 3, es decir, halle la probabilidad* $P(X\geq 3)$.
c) *Que el variable tome valores menores que 8, es decir, halle la probabilidad* $P(X<8)$.

Cuando el tamaño muestral puede ser demasiado grande en una variable aleatoria *X* que sigue una distribución binomial, el coeficiente binomial se hace imposible de calcular, razón por la cual tal valor no se puede establecer mediante la expresión $P(X=k)=\binom{n}{k}p^k q^{n-k}$; así pues, se hace necesario aproximar ese valor de probabilidad de algún modo. Así pues, la actividad 3 buscaba que los estudiantes intuyeran el procedimiento a realizar mediante la visualización gráfica de la binomial.

***Actividad 3***

*Dada la variable aleatoria X con distribución binomial B(100,0.3) calcule la probabilidad de los siguientes eventos:*

*a) Que la variable tome valores mayores que 12 y menores que 85, es decir, halle la probabilidad* $P(X \in (12,85))$.

*b) Que la variable tome valores mayores o iguales que 23, es decir, halle la probabilidad* $P(X \geq 23)$.

*c) Que la variable tome valores menores que 79, es decir, halle la probabilidad* $P(X<79)$.

*d) Comente las dificultades que tuvo calculando las probabilidades anteriores y sugiera algún mecanismo más cómodo que pudiera facilitarle los cálculos.*

### 2.3.3. Tercera Sesión. Teorema De Moivre-Laplace e Intervalos de Confianza

Durante la situación de institucionalización de las actividades anteriores, se presentó la sugerencia, por parte de algunos estudiantes, de recurrir al área bajo la curva normal obtenida mediante la unión de los rectángulos del histograma de la distribución binomial para estimar los valores de probabilidades binomiales. Se puede decir, entonces, que los estudiantes intuyeron el teorema de De Moivre Laplace que permite estimar probabilidades binomiales a través de la distribución normal. Para confirmar lo acertado de su intuición se les pidió que estudiaran el teorema en cualquier texto de estadística para que lo expusieran en la siguiente clase.

Esta sesión tuvo una duración de dos horas. En esta oportunidad, un grupo de estudiantes presentó una exposición de la distribución normal, más específicamente el Teorema De Moivre-Laplace y posteriormente se realizó la guía que constaba de dos situaciones problemas, los cuales fueron discutidos en clase. El análisis realizado fue de tipo gráfico utilizando Fathom.

El primer problema indagó a los estudiantes por valores de probabilidades en situaciones binomiales con tamaño muestral grande, la cual debía aproximarse mediante la distribución normal teniendo en cuenta el Teorema De Moivre-Laplace.

Para calcular el valor de probabilidad pedido, los estudiantes notaron la similitud del histograma de probabilidades binomiales con una curva normal, razón por la cual, para estimar el valor de probabilidad, calcularon el área bajo la curva normal comprendido hasta un número asignado. Para el desarrollo y análisis de este problema se utilizó Fathom, por lo cual el análisis fue de tipo gráfico.

***Ejercicio 1.***

*Se lanza una moneda honesta 100 veces.*

*(a) ¿Cuál es la probabilidad de que el número de caras no sea superior a 77?*

*(b) ¿Cuál es la probabilidad de que el número de caras no sea inferior a 10?*

*(c) ¿Cuál es la probabilidad de que el número de caras obtenidas sea al menos 17 y no mayor que 39?*

El segundo problema tenía como finalidad construir un intervalo de confianza de menor tamaño que tuviera una mayor probabilidad de éxito es decir que fuese de tamaño pequeño pero que tuviera un nivel de confianza grande, para tener mayor probabilidad de ganar el juego y tener que pagar poco dinero. Se pretendía que los estudiantes vivenciaran el dilema de la escogencia de la confianza y el tamaño muestral cuando el aumento de casos tenía un costo.

***Ejercicio 2.***

*Si ahora se tratara de realizar un juego de adivinación del número de caras que van a resultar y usted puede participar escogiendo tantos números como quiera siempre y cuando pague por ellos una cierta cantidad de* *dinero, ¿Cuáles números escogería para estar seguro que en el largo plazo ganaría y además que le cuesten la menor cantidad de dinero?*

### 2.3.4. Cuarta Sesión. La Distribución Normal

Esta sesión tuvo una duración de dos horas. Los estudiantes realizaron sus exposiciones y ejemplos respecto a la distribución normal. Esta sesión ayudó a aclarar algunas ideas y propiedades de la distribución normal, así como a mostrar la aplicabilidad de la misma en diversas situaciones.

Antes de finalizar la sesión, se realizó una pequeña evaluación con el fin de observar los conocimientos adquiridos hasta el momento referente a la distribución normal como aproximación de la distribución binomial (Ver Anexo 7). La evaluación buscaba observar si los estudiantes eran capaces de enfrentarse con situaciones problemas binomiales y que requerían del Teorema De Moivre-Laplace para calcular la probabilidad pedida. Así pues, se les presentó la siguiente pregunta:

***Pregunta de la evaluación***

*Una encuesta de opinión pregunta a una muestra de 500 adultos si están a favor de dar subsidios a los padres que tengan niños en edad escolar. Suponga que el 45% de la población está a favor de la idea. ¿Cuál es la probabilidad que más de la mitad de la muestra está a favor de la idea?*

### 2.3.5. Quinta Sesión. Estimación de Parámetros

Esta sesión tuvo una duración de cuatro horas. En ella, se implementó una guía compuesta por cinco actividades de un experimento (Ver Anexo 5) enfocado a determinar la proporción de bolas amarillas *(p)* que contenía una urna, realizando extracciones con sustitución.

En primera instancia, los estudiantes simulaban el experimento en Fathom por parejas, luego compartieron los datos con los compañeros del curso para finalmente analizar el comportamiento de la distribución de las medias muestrales al aumentar el tamaño de la misma. Luego, con base en la distribución muestral de las medias debían calcular intervalos de confianza para el valor del parámetro *p* de la distribución binomial en juego. En esta guía, como en las otras ya descritas, es necesario enfatizar en el enfoque constructivista de la propuesta metodológica implementada que no da soluciones ni presenta formas de hacer las cosas, sino que, por el contrario, plantea problemas que al intentar resolverlos crean las condiciones para abordar el conocimiento institucionalizado.

El objetivo de la guía era estimar la proporción de bolas amarillas *(p)* que contiene una urna, realizando extracciones con sustitución, y así poder analizar el comportamiento de las proporciones muestrales y calcular el margen de error generado en toda estimación. El objetivo primordial era analizar el comportamiento de las medias muestrales y notar que, aunque se disminuya la variabilidad al aumentar el tamaño muestral, cuando se estimar el valor de probabilidad de las medias existe un margen de error al estar trabajando con variables aleatorias.

1. *Forme pareja con un compañero y tomen una de las urnas disponibles. El problema consiste en estimar la proporción de bolas amarillas (p) que contiene la urna, realizando extracciones con sustitución. Realicen 10 veces el experimento de realizar 20 extracciones de la urna, y calculen la proporción de urnas amarillas obtenidas. Con estos 10 valores realicen una estimación de la proporción de amarillas que tiene la urna argumentando claramente su respuesta. No se niega el uso de Fathom.*

2. *Junten sus 10 resultados con los de las demás parejas, grábenlos en Fathom y realicen un nuevo análisis que les permita estimar el valor de la proporción de amarillas en la urna.*

3. *Para saber si el comportamiento de las proporciones muestrales (en este caso con muestras de tamaño 20) es semejante al observado en los numerales anteriores, y se puede hacer generalizaciones, realicen simulaciones en Fathom. Para poder hacer simulaciones en Fathom se requiere conocer el valor de p (asumen el que el profesor les asigne). La idea es suponer conocido el valor de p con el ánimo de saber qué propiedades tiene la distribución de las medias muestrales extraídas de esa población. Tomen 1000 muestras de tamaño 20, calculen la media de cada muestra y analicen la distribución de esos valores. Con base en los resultados, refinen, si lo consideran necesario, la estimación de la proporción de amarillas en la urna.*

### 2.3.6. Sexta Sesión. Evaluación Final

La evaluación final de estadística (Ver Anexo 6) se llevó a cabo en la última sesión, con una duración de cuatro horas. El cuestionario de evaluación consta de ocho ítems, de los cuales dos ítems son de selección múltiple (Ver Anexo 6, ítems 1 y 4); cuatro son ítems de problemas abiertos para justificar (Ver Anexo 6, ítems 2, 3, 5 y 6) y dos ítems de preguntas Verdadero/Falso (Ver Anexo 6, ítems 7 y 8). Aunque la evaluación final consta de 8 ejercicios, sólo cuatro situaciones problemas, el 1°, 3°, 4° y 5°, serán analizados en esta investigación.

La evaluación estaba enfocada a identificar y analizar las concepciones personales que los estudiantes atribuyen a la distribución binomial, al cálculo de probabilidades binomiales, a la aproximación normal de la binomial y a la estimación de parámetros e intervalos de confianza. De igual forma, se observaba los efectos de la utilización y manejo de Fathom para la interpretación, manera de abordar y comprensión de las actividades.

Como la utilización de Fathom era opcional, posiblemente en el análisis de las programaciones realizadas a las actividades de la evaluación, no se tendrá en cuenta a los 15 estudiantes sino solamente aquellos estudiantes que simularon las actividades. Por esta razón, podría suceder que hubiésemos realizado un análisis de las programaciones en Fathom con un pequeño grupo de estudiantes. En el capítulo 4 de esta investigación se detalla la muestra tenida en cuenta para el análisis de las programaciones en Fathom.

La primera situación propuesta fue un problema clásico que buscaba observar si los estudiantes intuían el efecto del tamaño muestral en el valor de probabilidades. Se podía enfrentar la situación bien recurriendo al enfoque frecuencial de la probabilidad, haciendo uso del hecho de que las frecuencias relativas son menos variables al aumentar el tamaño muestral, o bien, modelando la situación mediante la distribución binomial.
Como esta situación ya había sido propuesto en la Prueba diagnóstica al comienzo del semestre (Ver Mantilla y Martínez, 2007; Prueba Diagnóstica), se esperaba que los estudiantes, después del intenso trabajo computacional y teórico realizado alrededor de la probabilidad frecuencial y de la distribución binomial, demostraran mayor madurez al enfrentar nuevamente esta situación. En la prueba diagnóstica la gran mayoría de los estudiantes había recurrido a la proporción que existe entre las frecuencias relativas para seleccionar la opción c), poniendo en evidencia un típico argumento aritmético muy utilizado para resolver cuestiones de probabilidad.

1. *El notario de cierta población está muy interesado en prever el sexo de los recién nacidos para un estudio demográfico. Si se sabe por experiencia que las posibilidades son iguales para cada uno de los sexos, ¿Cuál de estos casos le parece más probable?*
    a) *Que en los próximos 10 nacimientos 8 o más sean varones.*
    b) *Que entre los próximos 100 nacimientos, 80 o más sean varones.*

c) *Son igualmente probables.*

El tercer punto presentaba una situación problema modelada por la distribución binomial con media 10 y probabilidad 1/5. Se pedía calcular valores de probabilidades en términos de expresiones *"al menos",* cuya comprensión es de difícil interpretación para los estudiantes.

El interés principal de esta pregunta estaba asociada con la capacidad de modelarla mediante una variable aleatoria binomial identificando acertadamente sus parámetros: la probabilidad de ocurrencia y el número de repeticiones; junto con el componente combinatorio y probabilístico inmerso en la expresión binomial.

*3. Un estudiante responde un examen de escogencia múltiple en el cual hay 10 preguntas cada una con 5 alternativas, de las cuales solo una es la respuesta correcta. El estudiante no sabe nada del cuestionario y resuelve contestar al azar.*

a) *¿Cuál es la probabilidad de que pase el examen (al menos 6 respuestas buenas)?*
b) *¿Cuál es la probabilidad de que conteste bien al menos una pregunta?*

El cuarto ejercicio buscaba que los estudiantes, además de reconocer la situación como binomial, analizaran el comportamiento de la variabilidad al aumentarse el tamaño muestral. El ingrediente "al menos" fue nuevamente incluido con el ánimo de evitar las rápidas comparaciones aritméticas que tanto les gustan a los estudiantes.

4. *Samuel Pepys, cuyos diarios relatan la vida del siglo XVIII en Inglaterra, era amigo del Sir Isaac Newton. Su interés en el juego lo llevó a preguntarle a Newton cuál de los siguientes resultados es más probable de obtenerse:*

a) *Al menos un 6 cuando se lanzan seis dados.*

b) *Al menos dos 6 cuando se lanzan 12 dados.*

c) *Al menos tres 6 cuando se lanzan 18 dados.*

*La pareja intercambió varias cartas antes de que Newton fuera capaz de convencer a Pepys de que ( ) era más probable. Llene el espacio y de sus razones para ello.*

El objetivo principal del quinto ejercicio era analizar las estrategias, argumentaciones y soluciones que los estudiantes dan a situaciones con ingredientes reales en las cuales se les requiere construir un criterio de decisión en un marco aleatorio.

5. *Suponga que la experiencia ha demostrado que solo 1/3 de todos los pacientes que tienen una cierta enfermedad se recobrarán si se les administra el tratamiento estándar. Una nueva droga va a ser probada con 12 voluntarios. Si las regulaciones de la salud requieren que al menos 7 de estos pacientes deben sanarse antes de que la nueva droga se permita, ¿cuál es la probabilidad de que la droga no sea acreditada incluso si aumenta la tasa de recuperación a ½?*

En el tercer y cuarto capítulo de esta investigación se presenta el análisis de las respuestas a las actividades y a la evaluación final a la luz de los dos núcleos de interés descritos previamente.

# 3. LA DISTRIBUCIÓN BINOMIAL

La variable aleatoria binomial es una variable discreta que sirve de modelo para una gran cantidad de situaciones prácticas, lo que hace que su cabal entendimiento se constituya en un objetivo básico de la enseñanza de la probabilidad.

Iniciamos este apartado con una explicación breve sobre las situaciones aleatorias que se pueden modelar con la distribución binomial para luego analizar las estrategias, dificultades y errores presentados por los estudiantes al momento de resolver situaciones problemas binomiales.

**La Distribución Binomial**

Suponga que se repite $n$ veces un experimento aleatorio que cumple las siguientes características:

- Tiene sólo dos posibles resultados: éxito y fracaso (Ensayos Bernoulli).
- Las repeticiones son independientes, esto es, el resultado de cada repetición no afecta el resultado de las demás.
- La probabilidad de obtener un éxito es constante, es decir, tiene el mismo valor cada vez que se repite el experimento.

Un experimento con las anteriores características tiene asociada una variable aleatoria $X$ que expresa el número de éxitos obtenido en $n$ repeticiones, la cual sigue una distribución Binomial con parámetros $(n,p)$ donde $n$ indica el número de repeticiones y $p$ la probabilidad de obtener éxito. La probabilidad de obtener $k$

éxitos está dada por la función de probabilidad $P(X=k)=\binom{n}{k}p^k q^{n-k}$ donde $q=1-p$ y se obtiene como una consecuencia de la regla del producto para *n* eventos independientes y de las consideración de todos los casos en que se obtienen *k* resultados exitosos.

**Elementos para el análisis de la comprensión de la distribución binomial**

Comprender la distribución binomial significa comprender los diversos elementos contenidos en la expresión $P(X=k)=\binom{n}{k}p^k q^{n-k}$ y su cabal adaptación a la situación particular a la que se refiere el problema que se desea modelar. Estos elementos, que van a servir de guía para el análisis del trabajo de los estudiantes, se constituyen en las categorías que asumimos como referentes para el análisis de los niveles de comprensión de los estudiantes, y son las siguientes:

1. La identificación de la distribución binomial como el modelo adecuado para resolver el problema propuesto. Es decir, antes de proceder a identificar los parámetros de la distribución binomial se requiere que el estudiante identifique que está en la presencia de esta distribución.
2. La identificación del experimento Bernoulli básico asociado a la situación y la determinación los parámetros *n* y *p* que definen completamente la función de probabilidad y que, en ciertos contextos, pueden confundirse o dar lugar a interpretaciones confusas.
3. La utilización de la regla generalizada del producto para *n* eventos independientes de los cuales *k* son exitosos y *n-k* no lo son.
4. El componente combinatorio $\binom{n}{k}$, que da cuenta del número de veces en que realizando *n* repeticiones se pueden obtener *k* éxitos y *n-k* fracasos. Se trata de la expresión en la binomial de la regla de la suma para eventos disyuntos donde el orden de los resultados juega un papel importante para modelar adecuadamente el valor de probabilidad.

## 3.1. CATEGORÍAS DE ESTUDIO

Teniendo en cuenta los elementos antes descritos, se establecieron las siguientes categorías para el análisis de nuestra investigación.

### 3.1.1. Categoría 1. IDENTIFICACIÓN DEL MODELO BINOMIAL

En esta categoría se encuentran los razonamientos que no concibieron las situaciones como binomiales. Observando las justificaciones, notamos que las concepciones personales de los estudiantes estuvieron más relacionadas con otros conceptos probabilísticos como: enfoque clásico de probabilidad, igualdad de proporciones, igual valor para $p$ y $q$ y probabilidad clásica. En algunos casos se presentaron explicaciones descontextualizadas.

A continuación presentamos los diversos razonamientos que hicieron los estudiantes al enfrentar situaciones binomiales pero que no las identificaron como tales.

#### 3.1.1.1 Enfoque Frecuencial de Probabilidad

Algunos estudiantes argumentaron sus respuestas a situaciones binomiales con base en el enfoque frecuencial de la probabilidad evitando el uso de la función de probabilidad binomial. Es el caso de Mauricio y Andrés al resolver la situación del estudio demográfico (Ver Anexo 6, ejercicio 1).

> *Mauricio: "Como vimos en el curso con muestras muy pequeñas, no se puede tener un valor aproximado (bueno), también sabemos y vimos que entre menos intentos o extracciones o datos, los cambios van a ser muy grandes, como vimos en las gráficas al principio varían mucho, a medida que aumentamos la cantidad de datos, esta variación va a disminuir y se va a parecer, va a tender al infinito; o con muestras muy grandes al valor de probabilidad".*

> *Andrés: "Por todo lo visto en clase se pudo notar que a mayor casos las posibilidades se acercaban más a la verdadera posibilidad; así que no son igualmente probables y para cada uno de los sexos las posibilidades son iguales. Por lo tanto decir de que 100 nacimientos 80 o más sean varones, no es tan probable ya que tenemos 100 casos, así la probabilidad debe acercarnos un poco más a la verdadera. Por lo tanto si de 10 nacimientos, 8 o más son varones es algo más probable puesto que tenemos menos casos, y la probabilidad aún varía".*

Mauricio y Andrés, dando argumentos frecuenciales de probabilidad, afirman que la variabilidad de la frecuencia relativa es bastante grande en muestras pequeñas, lo que hace que haya más valores de las frecuencias relativas, es decir, para muestras pequeñas es más probable tener cualquier valor posible de las frecuencias relativas que cuando las muestras son grandes. En otras palabras, a mayor variabilidad, mayor probabilidad de obtener valores lejanos del valor del parámetro $p$, por lo tanto es más probable obtener 8 varones o más en 10 nacimientos que 80 o más en 100 nacimientos. Aunque estas ideas no son incorrectas desde el punto de vista probabilístico, si hay que destacar su carácter intuitivo formado, seguramente, como resultado de las prácticas simuladas en Fathom alrededor del enfoque frecuencial de la probabilidad; el aspecto cuantitativo acerca de estos eventos que se obtiene cuando se reconoce en la situación el modelo binomial, no fue considerado por estos estudiantes.

#### 3.1.1.2. Igualdad de proporciones

Algunos estudiantes asumieron la igualdad de probabilidades basados en la igualdad de proporciones, lo cual es erróneo ya que se descarta el efecto que tiene el tamaño muestral sobre el valor de probabilidad. Un ejemplo de esto, se presentó en los argumentos de  Angélica en la situación del estudio demográfico (Ver Anexo 6, ejercicio 1).

*Angélica: "Son igualmente probables ya que la probabilidad de que 10 nacimientos 8 sean varones es de 0.8 y de 100 nacimientos que 80 sean varones es 0.8. Luego ambos casos son igual de probables".*

Cuando Angélica afirmaba que ambas opciones (10 nacimientos, 8 o más eran varones; 100 nacimientos, 80 o más eran varones) eran igualmente probables al existir igualdad de proporciones está utilizando un argumento esencialmente aritmético que asume una relación lineal entre el número de repeticiones y el valor de probabilidad. Esta universalidad de los modelos lineales es típica de muchos estudiantes que adoptan "la regla de tres" como argumento infalible para resolver problemas de diferente tipo.

En el caso de Fabián este argumento aritmético es combinado con algunos elementos de la distribución binomial para generar un argumento aún más incomprensible, ya que si se hubiera tomado el trabajo de realizar los cálculos hubiera percibido la no igualdad que propone:

*Fabián: "En el primero me está diciendo* $\frac{8}{10}$ *y en el segundo* $\frac{80}{100}$ *son fracciones equivalentes y como es un resultado existe la misma probabilidad.* $\left(\frac{8}{10}\right)\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{2}=\left(\frac{80}{100}\right)\left(\frac{1}{2}\right)^{80}\left(\frac{1}{2}\right)^{20}$ *".*

### 3.1.1.3. Igual valor de *p y q.*

Otra argumentación utilizada por algunos estudiantes en el problema de los nacimientos (Ver Anexo 6, ejercicio 1) y que les impidió reconocer la situación como binomial, y actuar en consecuencia para calcular las respectivas probabilidades, fue el hecho de que los valores de los parámetros $p$ y $q$ eran iguales. Veamos las argumentaciones de Camila y Gustavo que respaldan nuestra afirmación:

*Camila: "Son igualmente probables: porque si sabemos por experiencia que las posibilidades para ambos sexos son iguales y sabemos que entre más datos tengamos, mayor se acerca a la probabilidad y si decimos por experiencia quiere decir que con muchísimos datos entonces las probabilidades van a tender a ser igual en ambos sexos por los datos que obtenemos (lo que nos dan)".*

*Gustavo: "Sexo masculino igual probabilidad sexo femenino. Que entre los próximos 100 nacimientos, 80 o más sean varones ya que hay más opciones de escoger, pero de todas maneras las probabilidad no es muy buena ya que si se sabe teóricamente que la probabilidad para ambos sexos es igual porque de 100, 80 son varones".*

Como se observa ambos estudiantes son muy claros tanto en su respuesta como en su argumentación: Los eventos son igualmente probables porque la probabilidad para ambos es la misma, y las frecuencias relativas tienden a ese valor común, sin importar que en un caso se asuman 10 repeticiones y en el otro 100. Las argumentaciones contienen elementos del enfoque frecuencial de la probabilidad pero no distinguen el comportamiento de las frecuencias relativas para distintos tamaños muestrales; es asumir la convergencia de las frecuencias relativas al valor de la probabilidad de forma uniforme, esto es, sin ninguna referencia al proceso de acercamiento en la medida que se aumenta el tamaño muestral. Desde otro punto de vista se podría pensar que los estudiantes no asumen el experimento compuesto que resulta de considerar un determinado número de repeticiones del experimento.

#### 3.1.1.4. Probabilidad clásica.

En esta categoría incluimos los razonamientos que asimilaban el número de repeticiones del experimento como el espacio muestral y el número de éxitos como los resultados favorables para luego, utilizando probabilidad clásica, calcular la probabilidad requerida. Para ejemplificar esta peculiar forma de afrontar situaciones binomiales presentamos el argumento utilizado por Andrés en el problema que hacia referencia a las preguntas de un cierto examen (Ver Anexo 6, ejercicio 3).

*Andrés : Probabilidad de que la respuesta esté bien es de 1/5.*

*a) Debe ser* $\frac{6}{10} = \frac{3}{5}$ *b) 1/5*

Como se trataba de 10 preguntas y se debía contestar al menos 6 para aprobar, Andrés toma el cociente 6/10 como el valor de la probabilidad requerida. Esta argumentación que además olvida que también se pasa el examen respondiendo 7, 8, 9 y 10 preguntas, asimila un experimento compuesto (obtener 6 éxitos en 10 repeticiones) con un experimento simple que consiste en "seleccionar una respuesta buena cuando existen 5 de ellas". Para responder a la pregunta que indaga sobre la probabilidad de responder bien al menos una pregunta, Andrés responde 1/5 que es la probabilidad de responder bien una pregunta concreta y no una entre las 10 propuestas que es la forma como debía interpretarse la pregunta.
Este *sesgo clásico* se observa también en las respuestas que Cristina y Gustavo dieron al problema de los dados propuesto a Newton (Ver Anexo 6, ejercicio 4):

*Cristina: "todas eran igualmente probables".*

- *Al lanzar un dado la probabilidad de obtener 6 es* $\frac{1}{6}$

  *Con 6 dados tengo 36 posibles resultados y de ellos 6 posibles seis. Entonces* $P(6) con 6 dados : \frac{6}{36} = 0.166$

- *Con 12 dados tengo 72 y de ellos 12 son a favor del seis.*

*Entonces* $P(6) con 12 dados: \frac{12}{72} = 0.166$

- *Con 18 dados, n=108 y de ellos 18 a favor:*

  $P(6) con 18 dados: \frac{18}{108} = 0.166$

Cristina asumió que las tres opciones eran igualmente probables desde la probabilidad clásica. Aunque asumió correctamente el valor de *p*=1/6, no lo aplicó en la situación; más bien, basó sus argumentos en el número de posibilidades de los dados, como si se tratara de eventos simples y no compuestos que implica la realización conjunta de varios eventos y por lo tanto el uso de la regla del producto.

Aunque Gustavo también dio una justificación a su procedimiento mediante probabilidad clásica no tuvo presente que en cualquier dado puede resultar el número seis; más bien estableció un cociente entre el número de dados y obtener un seis, dos seis y tres seis en cualquier dado.

*Gustavo: "las probabilidades son igualmente probables".*

- *Al menos un seis en 6 dados. 1 seis de 6 dados* $P = \frac{1}{6} = 0.16$
- *Al menos dos seis en 12 dados.* $P = \frac{2}{12} = 0.16$
- *Al menos tres seis en 18 dados.* $P = \frac{3}{18} = 0.16$

Tanto Cristina como Gustavo asumieron la expresión *"(…) al menos (…)"* como *"(…) igual a (…)"* lo que los llevó a establecer erradamente las probabilidades $P(X=1)$, $P(X=2)$ y $P(X=3)$ en lugar $P(X \geq 1)$, $P(X \geq 2)$ y $P(X \geq 3)$. De aquí surge la inquietud de saber si Cristina y Gustavo hubieran continuado utilizando su estrategia clásica para resolver las preguntas tal y como se plantearon, es decir, $P(X \geq 1)$, $P(X \geq 2)$ y $P(X \geq 3)$.

Otra evidencia del *sesgo clásico* se presentó con Susana en el problema del examen (Ver Anexo 6, ejercicio 3).

> *Susana: "La probabilidad de que pase el examen es de 0.6 hacia arriba, es decir, 0.6, 0.7, 0.8, 0.9, 1".*
>
> *a) La probabilidad es 0.1, 0.2, 0.3, 0.4, 0.4, 0.6, 0.7, 0.8, 0.9, 1*

Susana, al contrario de Andrés, va más allá y considera las otras posibilidades que permiten también pasar el examen, esto es, responder bien 7, 8, 9 o 10 preguntas.

### 3.1.2. Categoría 2. DETERMINACIÓN DE EXPERIMENTOS BERNOULLI Y DE LOS PARÁMETROS *n* y *p.*

En esta categoría consideramos los argumentos que, aún identificando la situación como binomial, determinaron incorrectamente los parámetros *n* y *p.* En esta categoría se presentan los siguientes casos:

#### 3.1.2.1. Identificación parcial del experimento Bernoulli con *n* incorrecto.

Una idea clave al tratar de modelar situaciones problemas binomiales es el establecimiento del experimento básico, experimento Bernoulli, para luego establecer el parámetro *p* y el número de repeticiones, *n*, del experimento. Determinar el experimento Bernoulli representó algunas dificultades tal y como se evidencia en la argumentación de Camila para responder el problema del examen (Ver Anexo 6, ejercicio 3).

*n=50* $P(exacta) = \frac{10}{50} = 0.2$ $\mu = np = 10$ $q = 0.8$

$N(10{,}2.82)$ $\sigma = \sqrt{npq}$

$\sigma = \sqrt{50*0.2*0.8} = 2.82$

*A)* $P(X \geq 6) = P\left(\frac{X-10}{2.82} \geq \frac{6-10}{2.82}\right) = P\left(\frac{X-10}{2.82} \geq -1.41\right)$

$\Rightarrow P(Z \geq -1.41) = 1 - P(Z \leq 1.41) = 1 - 0.0793 = 0.9207 = 92\%$

*b)* $P(X \geq 1) = P\left(\frac{X-10}{2.82} \geq \frac{1-10}{2.82}\right) = P\left(\frac{X-10}{2.82} \geq -3.19\right)$

$\Rightarrow P(Z \geq -3.19) = 1 - P(Z \leq 3.19) = 1 - 0.007 = 0.9993 = 99\%$

Camila estableció incorrectamente el tamaño muestral, asumiendo el total de opciones de las 10 preguntas, es decir las 50 opciones, como el tamaño muestral en lugar de las 10 preguntas de la evaluación. Es decir, se podría interpretar su argumentación pensando que para ella el modelo Bernoulli lo establecía cada ítem y que se trataba de seleccionarlo o no. El problema de esta interpretación es que derivaría en un valor de *p* igual a ½ y no el 1/5 que ella indirectamente encuentra.

Asumir a *n =50 opciones* como tamaño muestral en lugar de *n =10 preguntas,* hizo que Camila aproximara el valor de probabilidad pedido mediante la distribución normal. Aclaramos que el procedimiento sería correcto siempre y cuando el valor de *n* fuese 50, pero como son únicamente 10 preguntas debían calcularlo mediante la función de probabilidad binomial $P(X = k) = \binom{n}{k} p^k q^{n-k}$ ya que en este caso la aproximación normal no es muy acertada.

Hubo una situación muy particular, en donde se evidencia la comprensión de situaciones binomiales justificadas con procedimientos algebraicos. Como las excepciones pueden existir, presentamos a continuación el argumento de un estudiante quien asumió la situación como binomial y estableció en forma algebraica los valores de probabilidad teniendo en cuenta los elementos de la función de probabilidad binomial: componente probabilístico (regla del producto) y componente combinatorio.

Aunque Fabián tuvo en cuenta el componente combinatorio al establecer los valores de probabilidad, no concibe que la evaluación (Ver Anexo 6, ejercicio 3) sea respondida correctamente porque no tuvo en cuenta la posibilidad de contestar las 10 preguntas como correctas. Además, asumió erradamente el valor de *p* como ½, es decir, asumió las preguntas como correctas o incorrectas y olvidó que cada preguntas tenía 5 opciones de respuestas y así obtener el verdadero valor de *p=1/5.*

*Fabián:* $$B=\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{4}*210+\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{3}*120+\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{2}*45+\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right)^{1}*10$$

$$=\left(\frac{1}{2}\right)^{10}*(210+120+45+10)=\left(\frac{1}{2}\right)^{10}*385=0.3759765$$

Aunque Fabián encontró el experimento Bernoulli de la evaluación, éste es incorrecto puesto que la convirtió en una situación simple (contestar bien o mal la pregunta) en lugar de una situación compuesta (contestar bien la pregunta de cinco opciones). En cierta forma, asumió equiprobable la posibilidad de contestar bien o mal la pregunta contrario a lo que indicaba la situación; pero sin embargo, resaltamos el procedimiento algebraico realizado por Fabián a la situación binomial.

#### 3.1.2.2. No identificación del experimento Bernoulli con *p* incorrecto.

La no identificación del experimento Bernoulli en la situación de la evaluación (Ver Anexo 6, ejercicio 2) llevó a que Mauricio estableciera incorrectamente el valor de *p* y además, que la situación no pudiese ser modelada mediante la distribución binomial al quedarse sin repeticiones pues reunió los datos para establecer los parámetros *n* y *p* respectivamente.

> *Mauricio: Se tendría un total de 50 opciones y 5 buenas respuestas, esto sería igual a tener 5/50 ó 1/10 sería tener una opción entre diez. Con ayuda de Fathom miré cuál era la probabilidad de obtener la respuesta esperada entonces:*
>
> a) *Obtener al menos 6 la probabilidad es 0.651322.*
>
> b) *Obtener al menos 1 la probabilidad es 0.3874.*

Mauricio estableció incorrectamente el valor de *p* al tomar las 50 opciones de las 10 preguntas como el tamaño muestral y las 5 opciones de cada pregunta las concibió como el número de preguntas de la evaluación. En cierta forma, no hizo distinción entre el número de opciones y el número de preguntas en la evaluación, lo que llevó a que se quedara sin repeticiones. Además, confundir el valor de tamaño muestral con las opciones posibles, hizo que asignara un valor equivocado a la probabilidad de éxito, puesto que calculó *p=5/50=1/10 (5 preguntas, 50 opciones)* en lugar de *p=1/5 (1 opción correcta de 5 posibles).*

Aunque es contradictorio, Mauricio concibe más probable obtener al menos seis respuestas correctas que al menos una respuesta correcta, reflejando incorrectamente que $P(X \geq 6) > P(X \geq 1)$ puesto que intuitivamente habrá mayor valor de probabilidad en $P(X \geq 1)$ al haber más valores de probabilidad.

Se desconoce la programación en Fathom realizada por Mauricio y nos queda la duda acerca de la utilización de Fathom para justificar sus argumentos puesto que Mauricio se quedó sin repeticiones. Sumado a ello, establece erróneamente el tamaño muestral de la situación a las 5 opciones de cada pregunta en lugar de las 10 preguntas del examen.

### 3.1.3. Categoría 3. REGLA DEL PRODUCTO Y EL COMPONENTE COMBINATORIO

En esta categoría consideramos los razonamientos que incluyeron algunos elementos del modelo binomial pero sin completar la expresión para la función de probabilidad. Nos referimos concretamente a la utilización de la regla del producto que surgió al considerar la situación propuesta como un experimento compuesto y a la mención del componente combinatorio. En esta categoría pudimos encontrar las siguientes características:

#### 3.1.3.1 Regla del producto para *k* éxitos.

En esta categoría encontramos aquellas argumentaciones que si bien no asumieron la situación como binomial, poseen ciertos elementos que muestra avance en la construcción y concepción de situaciones binomiales. A continuación presentamos la argumentación dada por Sebastián:

> *Sebastián: a) La probabilidad de que en cada punto de respuesta correcta es* $\frac{1}{5}$ *entonces la probabilidad de que le queden 6 bien es* $\frac{1}{5}*\frac{1}{5}*\frac{1}{5}*\frac{1}{5}*\frac{1}{5}*\frac{1}{5}=6.4*10^{-5}$
>
> *a) La probabilidad de que conteste bien al menos 1 pregunta es* $\frac{1}{5}$*.*

Sebastián, por ejemplo, asume el examen como una situación compuesta centrando su atención en las 5 posibles respuestas de cada pregunta para obtener un valor de p=1/5. Además, asume la probabilidad de responder bien una pregunta independientemente, es decir, contestar bien una pregunta no lleva a que otra pregunta sea resuelta correctamente, puede que no sea correcta.

Sebastián asumió contestar bien seis preguntas como un producto entre el valor de *p* de contestar bien una pregunta y el número de preguntas correctas para pasar la evaluación, es decir, asumió parcialmente la regla del producto para *k* éxitos (seis respuestas correctas) en *n* eventos independientes (10 preguntas de la evaluación) olvidando el efecto que tiene los *n-k* fracasos (4 preguntas) y el respectivo valor de *q=4/5* como fracaso y el componente combinatorio de la situación. Además, Sebastián no comprendió la expresión *"(…) al menos (…)"* pues la concibió como *"(…) igual a (…)".*

### 3.1.3.2 Regla del producto para *k* éxitos y *n-k* fracasos.

En este apartado presentamos las argumentaciones que, aunque no propusieron la expresión de la distribución de probabilidad binomial, sí contiene parte de ella como es el producto de las probabilidades de los éxitos requeridos y los consecuentes fracasos.

En nuestra investigación se evidenció que Susana, al trabajar con la situación de los dados (Ver Anexo 6, ejercicio 4), aplicó la regla del producto para *n* eventos independientes donde *k* eran los éxitos y *n-k* los fracasos, resaltando que las tres opciones eran tres situaciones independientes que poseían igual valor para *p* y *q* y con tamaños muestrales (*n= números de dados)* diferentes.

*Susana:* $P(a)=\frac{1}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}=0.06$

$$P(b)=\frac{1}{6}*\frac{1}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}=0.0044$$

$$P(c)=\frac{1}{6}*\frac{1}{6}*\frac{1}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}=\frac{3.051757812^{\ 10}}{1.015599567^{\ 14}}$$

*El (a) es más probable puesto que me da la probabilidad más alta.*

Aunque resaltamos la interpretación de Susana de la situación en términos de la regla del producto, vemos que Susana no comprendió la expresión *“ (…) al menos (…)”* puesto que la concibe como *“ (…) igual a (…)”* porque calculó incorrectamente la probabilidad de obtener exactamente un seis, dos seis y tres seis en el lanzamiento de 6, 12 y 18 dados respectivamente.

### 3.1.3.3 El principio fundamental de conteo, probabilidad clásica y regla del producto.

Otra estrategia utilizada para resolver el problema de los tres dados se caracteriza por la utilización de estrategias combinatorias (el principio fundamental de conteo), la probabilidad clásica y la regla del producto. Es el caso de Sebastián:

- *Como los números de posibilidades* $(X_1, X_2, X_3, X_4, X_5, X_6)$ *son* $6*6*6*6*6*6 = 46656$ *entonces* $\frac{6}{46656} \approx 0.0001$ *es la probabilidad de obtener un seis si se lanzan seis dados.*

- *Al menos dos seis cuando se lanzan 12 dados.*
  $6*6*6*6*6*6*6*6*6*6*6*6 = 2176782336$

$$\left(\frac{12}{2176782336}\right)*\left(\frac{12}{2176782336}\right) = \frac{144}{(2176782336\ )^2} \approx 3.03*10^{-17}$$

- *Al menos tres seis cuando se lanzan 18 dados.*
  $6*6*6*6*6*6*6*6*6*6*6*6*6*6*6*6*6*6 = 2821109907456$

$\left(\frac{18}{2821109907456}\right)^3$ *Probabilidad 3 seis cuando se lanzan 18 dados.*

*SOLUCIÓN GENERAL:* ***un seis cuando se lanzan seis dados, es la de mayor probabilidad.***

El argumento de Sebastián termina siendo, en últimas, la mezcla de la estrategia combinatoria con la utilización de la regla del producto cuando en verdad cualquiera de las dos eran válidas para responder las preguntas asociadas. Inicia calculando el número de 6-uplas, 12-uplas y 18-uplas posibles al lanzar 6, 12 y 18 dados respectivamente que asumió como el número de resultados posibles, es decir, el denominador en la expresión para la probabilidad clásica. Para el numerador asumió en el primer caso el valor 6 dando a entender que asumió la expresión *“ (…) al menos un seis (…)* como la probabilidad de obtener *un seis en cada dado* y como eran seis dados, se podía obtener seis veces el número seis. Para los casos en que se lanzaban 12 y 18 dados, aplicando el mismo argumento, obtuvo 12 y 18 posibilidades de obtener un 6, y como le requerían dos y tres 6 resolvió multiplicar las probabilidades de obtener un 6 dos y tres veces. Es decir, el exponente de la expresión se puede justificar en términos de las posibilidades de obtener el determinado número de seis pedidos en cada caso.

La argumentación de Sebastián refleja una falta de comprensión de cada uno de los métodos que combinó: El método combinatorio responde a una concepción global que da cuenta de una sola vez del carácter compuesto o múltiple del experimento, no tiene ninguna necesidad de aplicar la regla del producto que precisamente responde a otra estrategia de solución del problema: se calcula la probabilidad de éxito de cada experimento simple y luego se calcula la del experimento compuesto utilizando la regla del producto.

# 4. MANEJO DE FATHOM Y SU USO EN LA RESOLUCIÓN DE SITUACIONES PROBLEMAS BINOMIALES

La simulación computacional ayuda en la resolución de un problema pues sirve para corroborar o no una idea simulando la experiencia muchas veces y observando su comportamiento para realizar un análisis más detallado y argumentativo del mismo. Con respecto a esto Biehler (1991, p.193), citado por Mantilla y Martínez (2007), afirma: *"Usar simulación para permitir a los estudiantes experimentar un ambiente de azar en lugar de analizar un conjunto finito dado de datos reales puede tener varias ventajas. Los estudiantes participan en el proceso de generación de datos haciendo predicciones e hipótesis y experimentando incertidumbre (…)"*

Iniciamos este capítulo con una breve descripción de los comandos, utilidades y manejo del software computacional Fathom para luego analizar las programaciones realizadas por los estudiantes al resolver situaciones problemas. Para ello, tendremos en cuenta algunos elementos de análisis que se convertirán en nuestras categorías de estudio.

Nuestro interés se centrará en estudiar las programaciones realizadas por seis estudiantes con el fin de observar y analizar los razonamientos puestos en juego al enfrentarse con situaciones binomiales al momento de utilizar Fathom.

## Software Computacional Fathom

Para realizar una buena programación en Fathom es necesario tener en cuenta que las unidades fundamentales de la simulación computacional son los comandos, operadores y funciones que posee Fathom para definir los atributos. Las unidades que se requirieron para representar las situaciones problemas tratadas en esta investigación, más específicamente, en situaciones binomiales fueron:

- *Operadores:* Aritméticos: + , -, *, /, $\wedge$; lógicos: *and, or, not:* de comparación: >,<, =, ≤, ≥..
- *Ordenadores: CaseIndex:* es igual al número de la fila, es decir, indica el número de casos realizados hasta el caso en que se está considerando; *prev:* es igual al caso previo. Si se está en el primer valor *prev* retorna 0, *combinatios* establece el valor combinatorio en la variable de interés.
- *Funciones aleatorias: randomCumulative( ):* da el valor de probabilidad acumulado de una variable aleatoria normal con media 0 y desviación estándar 1, *randomQuantile( ):* da el valor de Z que indica la distancia en desviaciones estándar de la media, *randompick( ):* función simuladora de procesos aleatorios.
- *Condicionales:* crea un sentencia la cual tiene validez si se cumplen ciertas condiciones; por ejemplo, si se introduce la expresión $if(X<5)=\begin{cases}1\\0\end{cases}$ y se cumple la condición expresada dentro del paréntesis después del *if* entonces se cumple la dicho en la opción superior después de la llave, en este caso tomaría el valor de 1; en cambio si la expresión del paréntesis es falsa sucede la expresión de abajo, es decir se le asignaría el valor de 0.
- *Definición Modelo Binomial:* Para la simulación de situaciones binomiales se requiere de las siguientes tres colecciones: la colección *"Collection"* define el espacio muestral del experimento, la colección *"Sample"* define las muestras y la colección *"Measure"* define las medidas de las muestras, es decir, los resultados de cada muestra del experimento.
- La función *proportion( ) y Summary Table* para calcular las frecuencias relativas y las medias de las situaciones. Además, el comando *combination* establece el componente combinatorio de la expresión binomial y, mediante el comando *max( )* toma el máximo valor de un conjunto de datos.

- La función *Graph* es utilizada para visualizar gráficamente el comportamiento de los experimentos binomiales. Además, el comando *Slider* está concebido para que el parámetro pueda determinar un rango de posibles valores. Las dos anteriores funciones fueron de gran importancia en las actividades realizadas puesto que los estudiantes observaron los distintos cambios gráficos sufridos al variar el parámetro determinado.

Una de las características de Fathom es la posibilidad que brinda para generar variadas representaciones y percibirlas simultáneamente. Además, la facilitad y rapidez al realizar los cálculos permitiría que el estudiante haga un mayor análisis de la situación, construyera nuevas significados y adoptara nuevas estrategias en la resolución de situaciones problemas. En Fathom, los objetos que se construyen son representaciones formales de objetos reales puesto que se realiza una abstracción de la situación al definirlo como modelo de urna.

**Elementos para el análisis del manejo de Fathom**

Para utilizar adecuadamente a Fathom como simulador de experimentos aleatorios es necesario conocer sus comandos y establecer correctamente el espacio muestral, las muestras, las medidas a calcular y la generación de ellas.

Sin embargo, puede suceder que los estudiantes hayan concebido a Fathom como una calculadora en donde simplemente registran fórmulas usando sus comandos, lo que en la práctica significa el no uso de las posibilidades del paquete y el descarte de la simulación computacional como apoyo para la resolución de problemas. Por estas razones, los elementos a tener en cuenta y que se convertirán en nuestras categorías de estudio son:

1. Construcción del modelo básico de simulación de una experiencia aleatoria: Definición del espacio muestral del experimento, definición de las muestras y de las medidas asociadas.

2. Uso del Fathom para resolver problemas que responden a un modelo binomial.

## 4.1. CONSTRUCCIÓN DEL MODELO BÁSICO DE SIMULACIÓN: IDENTIFICACIÓN DEL EXPERIMENTO BERNOULLI, LAS MUESTRAS Y LAS MEDIDAS ASOCIADAS.

En este apartado tenemos en cuenta aquellas programaciones que presenten y muestren las situaciones binomiales planteadas como modelos de urna mediante los comandos de Fathom. Es decir, analizaremos el muestreo de los estudiantes en situaciones binomiales mediados con Fathom.

Una buena comprensión de situaciones binomiales está relacionada con la identificación del experimento Bernoulli básico, puesto que la binomial es un conjunto de tales elementos. Por esa razón, intentamos observar y analizar si los estudiantes concibieron el experimento básico Bernoulli en toda situación binomial a través de las programaciones realizadas en Fathom.

Hemos observado que aquellos estudiantes que simulan la situación binomial como modelo de urna reflejan en su programación el experimento Bernoulli básico del experimento. En nuestra investigación, el anterior caso se presentó con Abel y por esa razón nuestro interés se centrará en las programaciones realizadas por este estudiante para intentar explicar y ampliar esta categoría.

- **Identificación del Experimento Bernoulli básico.**

Con respecto a la situación del examen (Ver Anexo 6, ejercicio 1), Abel convirtió la situación en un modelo de urna compuesta por cinco bolas siendo una de ellas distinta a las demás. Contestar aleatoriamente el examen de 10 preguntas lo asimiló con la realización de 10 extracciones con sustitución de una bola de la urna. La pregunta se consideraba bien respondida si se extraía la bola distinta.

*Programación de Abel*

- Definió el espacio muestral como un modelo de urna mediante un atributo con cinco valores, cuatro “0” y un “1” dando buena cuenta del experimento Bernoulli, base del problema, pues contestar bien una pregunta entre cinco opciones es equivalente a obtener el valor “1” de la urna construida.

examen

| | examen |
|---|---|
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 1 |

Fig 4.1. Espacio Muestral. Evaluación Abel.

- Para simular el examen de 10 preguntas definió una muestra *(sample)* con 10 extracciones; las 10 extracciones hacen referencia a las 10 preguntas de la evaluación. La muestra, caracterizada por su tamaño, hace referencia a las veces que se repite el experimento, en este caso, a las 10 preguntas del examen.
- Luego definió la medida número de unos en cada muestra mediante el comando *Measure*, que hace el conteo de los resultados obtenidos en las muestras, para dar cuenta del número de respuestas acertadas en cada repetición del examen. Contar el valor de 1 cuando se contesta bien una pregunta estaría mostrando el número de respuestas correctas que obtuvo en una evaluación de 10 preguntas.
- Generó 1000 casos, es decir, simuló 1000 veces la acción de hacer 10 extracciones. Simular el experimento muchas veces busca observar el comportamiento de la situación generando la distribución de los valores posibles con mayor exactitud, en particular, tratando de estabilizar el valor de interés, es decir, de encontrar el valor de probabilidad requerido.
- Con ayuda del *Summary* y con el uso de la función *proportion( )* estableció los valores de probabilidad para $P(X \geq 6)$ y $P(X \geq 1)$ respectivamente, lo que le permitió obtener con bastante precisión las probabilidades pedidas.

Measures from Sample of examen

| | |
|---|---|
| comolefue | 2,0050975<br>0,0061969015<br>0,89045477 |

S1 = mean ( )
S2 = proportion (comolefue ≥ 6)
S3 = proportion (comolefue ≥ 1)

Fig. 4.2 Media y proporción Evaluación Abel

| | examen |
|---|---|
| 1 | 1 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |
| 8 | 0 |
| 9 | 0 |
| 10 | 1 |

Fig. 4.3. Tabla: Número de preguntas de la evaluación

Measures from Sample of e

| | comolefue |
|---|---|
| 1 | 3 |
| 2 | 4 |
| 3 | 1 |
| 4 | 1 |
| 5 | 4 |
| 6 | 1 |
| 7 | 5 |
| 8 | 5 |
| 9 | 1 |
| 10 | 2 |
| 11 | 3 |

Fig. 4.4. Tabla: Número de preguntas correctas en 10000 repeticiones.

Aunque el valor de la media no incidía en nuestra pregunta, sí muestra que si un estudiante contesta aleatoriamente el examen  es muy probable que conteste bien sólo dos de las diez preguntas del examen, valor éste que coincide con el valor teórico *np* para la media de una distribución binomial. Así pues, también se observa la poca probabilidad (6%) de pasar el examen contestando al azar y la probabilidad grande para contestar al menos una pregunta bien (89%).

De hecho, las programaciones de Abel en Fathom son las únicas, en todas las realizadas,  que identifican el experimento Bernoulli de las situaciones binomiales planteadas, lo que indica un primer paso para una buena comprensión e identificación de situaciones binomiales. Sin embargo, establecer el experimento Bernoulli básico no es garantía absoluta de buena concepción en situaciones binomiales puesto que puede suceder que no identifiquen adecuadamente el tamaño muestral, quedándose sin repeticiones, o den una argumentación incorrecta con base en una medida errada, como veremos más adelante.

- **No identificación del Experimento Bernoulli Básico.**

La no identificación del experimento Bernoulli básico es un punto de referencia para afirmar que los estudiantes no concibieron las situaciones planteadas como binomiales. Y si lo estudiantes no concibieron las situaciones como binomiales entonces nos surge las siguientes inquietudes: ¿Cómo intentaron dar alguna justificación al respecto? ¿Qué procedimientos llevaron a cabo los estudiantes?

Aunque las situaciones presentadas fueron binomiales, algunos estudiantes no las concibieron como tales posiblemente a que no identificaron, en primera medida, que podía haber situaciones compuestas de dos binomiales. La situación del estudio demográfico (Ver Anexo 6, Ejercicio 1) es un claro ejemplo de lo descrito anteriormente puesto que no la concibieron como una situación compuesta por dos binomiales; o no establecieron el experimento Bernoulli ni los parámetros *n* y *p* de la situación binomial, lo que llevó a que algunos estudiantes no hicieran distinción con los datos dados e intentaran dar justificaciones erradas con base en la distribución normal y la probabilidad clásica. Las respectivas justificaciones a lo mencionado anteriormente se encuentran en el capítulo 3 de esta investigación.

### 4.1.1. Uso del Teorema De Moivre-Laplace

En nuestra investigación, las programaciones de Marcela y Cristina presentaron la anterior situación y pudimos notar que las estudiantes intentaron dar justificaciones a las situaciones mediante probabilidad clásica y la distribución normal.

Marcela por ejemplo, intenta resolver el problema con base en la distribución normal al no identificar las dos situaciones binomiales ni los respectivos parámetros *n* y *p* de la situación planteada.

***Programación de Marcela***

- Definió la situación mediante una *Collection* llamada PUNTO1.
- Estableció en la columna **niños** de un *Table,* los valores de interés: 80 y 8 de las situaciones, haciendo referencia a los

PUNTO1

| | niños | area |
|---|---|---|
| 1 | 8 | 0,97111 |
| 2 | 80 | 1 |

Fig. 4.5. Estudio Demográfico Marcela

nacimientos ocurridos en los dos hospitales. Se puede pensar que ella concibió la situación compuesta por dos situaciones binomiales al establecer correctamente los valores de interés en cada una de ellas.

- Luego, En la segunda columna llamada **area** de *Table*, calcula el valor de probabilidad mediante la expresión $area = normalCumulative(niños;5;\sqrt{2.5})$. Para ello, establece la media y la desviación en términos del tamaño muestral *n* y el valor de *p.*

Mediante la expresión $area = normalCumulative(niños;5;\sqrt{2.5})$ estableció el área bajo la curva normal hasta el valor de interés, en nuestro caso serían 80 y 8 respectivamente, que en términos del Teorema De Moivre-Laplace equivale al valor aproximado de la probabilidad binomial obtenido mediante procedimientos normales.

Aunque Marcela realizó esta programación con el fin de aproximar el valor de probabilidad binomial mediante la normal por el tamaño muestral grande de la situación (*n=80)* procedió en forma incorrectamente al no identificar la situación compuesta por dos situaciones binomiales con tamaños muestrales 10 y 100 respectivamente y confundir los valores de interés (80 y 8 nacimientos) con los tamaños muestrales de las situaciones (100 y 10).

Además, si la situación hubiese sido una normal con *n=80 varones*, Marcela presenta dificultad para establecer la varianza al tomar como tamaño muestral a *n=8 varones* en lugar de *n=80 varones.* Al final, a Marcela se le dificulta distinguir diversas situaciones binomiales en una situación e identificar el tamaño muestral de cada una de ellas.

### 4.1.2. Probabilidad Clásica

De otro lado, y como evidencia de la influencia de la simulación computacional sobre las intuiciones previas y los argumentos que las respaldan, presentamos la programación de Cristina. Centramos nuestra atención en la programación de Cristina a la situación de los juegos con los dados (Ver Anexo 6, ejercicio 4) quien programando la situación como binomial, organizó sus ideas en base a la probabilidad clásica.

Analizaremos la programación realizada por Cristina para tratar de buscar las razones por las cuales llegó a la errada conclusión de que las tres opciones eran igualmente probables.

***Programación Cristina***

- Primero, definió los espacios muestrales de las situaciones mediante tres "*Colection"* llamadas: CON 6 DADOS, CON 12 DADOS y CON 18 DADOS y un slider fijo[3] en p=1/6. Resaltamos el hecho que Cristina concibió las tres situaciones binomiales con distintos tamaños muestrales, asignó correctamente el valor de *p*, pero no concibió el experimento Bernoulli básico del mismo.

- Después, estableció en un *Table* los posibles *k* valores de interés 36, 72 y 108 para las colecciones CON 6 DADOS, CON 12 DADOS CON 18 DADOS mediante la expresión $k\_valor = caseIndex - 1$. Los valores 36, 72 y 108 los concibió como el número de posibilidades que tenía para obtener

[3] Usar un slider fijo a un parámetro no tiene sentido ¿por qué lo harían? El slider está concebido precisamente para percibir los efectos de la variación de un parámetro sobre alguna expresión. En nuestra investigación se evidenció la importancia de tal comando cuando, por ejemplo, quisimos analizar el comportamiento de la distribución binomial cuando el tamaño muestral era fijo y el valor de *p* variaba entre 0 y 1 (Ver Anexo 1, Actividades 1-4).
Quizás querían observar el comportamiento del experimento cuando se variaba el valor de *p*, pero en la situación del juego con los dados, se necesitaba el valor de *p=0.2 fijo*.

un seis, dos seis y tres seis al lanzar 6 dados, 12 dados y 18 dados respectivamente.

- Luego, estableció el componente combinatorio de la expresión binomial mediante las expresiones: $combinar = combinations(36; k\_valor)$ ,
  $combinar = combinations(72; k\_valor)$ ,
  $combinar = combinations(108; k\_valor)$ .

Aunque parecía que Cristina había identificado los parámetros $n$ y $p$ de la situación, al momento de establecer el componente combinatorio de la expresión binomial notamos que confunde el número de caras o posibles valores de los dados con el tamaño muestral de la situación al acomodar los posibles valores como tamaño muestral sin hacer distinción alguna entre ellos.

- Posteriormente y mediante las siguientes expresiones programó en Fathom la expresión completa de la binomial:
  $$probabilid = combinar(p^{k\_valor})(1-p)^{36-k\_valor} ,$$
  $$probabilid = combinar(p^{k\_valor})(1-p)^{72-k\_valor}$$
  $$probabilid = combinar(p^{k\_valor})(1-p)^{108-k\_valor}$$

- Por último estableció las siguientes expresiones $k = \frac{k\_valor}{36}$, $k = \frac{k\_valor}{72}$ y $k = \frac{k\_valor}{108}$ , y centró su atención en el último valor dado (en los tres cocientes da como resultado 1) para dar sus argumentaciones.

CON 6 DADOS

| | K_VALOR | combinar | probabi | k |
|---|---|---|---|---|
| 28 | 27 | 94143280 | 1,6114e-14 | 0,75 |
| 29 | 28 | 30260340 | 1,03093e-15 | 0,777778 |
| 30 | 29 | 8347680 | 5,6606e-17 | 0,805556 |
| 31 | 30 | 1947792 | 2,62895e-18 | 0,833333 |
| 32 | 31 | 376992 | 1,01278e-19 | 0,861111 |
| 33 | 32 | 58905 | 3,14974e-21 | 0,888889 |
| 34 | 33 | 7140 | 7,59912e-23 | 0,916667 |
| 35 | 34 | 630 | 1,33459e-24 | 0,944444 |
| 36 | 35 | 36 | 1,51793e-26 | 0,972222 |
| 37 | 36 | 1 | 8,3925e-29 | $\boxed{1}$ |

Fig. 4.6. Juego con 6 Dados Diana.

Ahora que ya estableció la expresión binomial en comandos de Fathom, observamos que Cristina no identificó el tamaño muestral de la situación y sí en cambio dio doble uso a los posibles valores de interés: como tamaño muestral y valor de interés, es decir, tomó como tamaño muestral el número de casos posibles de la situación al definir la expresión binomial.

Por último, centró su atención en los cocientes de las tres programaciones; de hecho, los argumentos dados por la estudiante tuvieron en cuenta los resultados obtenidos en los tres cocientes. Aunque intentó y gastó tiempo en realizar una programación de la situación como binomial, sus argumentos se basaron únicamente en los tres cocientes de las programaciones de tipo clásico.

CON 12 DADOS

| | k_valor | combinar | probabilid | k |
|---|---|---|---|---|
| 61 | 60 | 1,53633e... | 2,79874e-35 | 0,833333 |
| 62 | 61 | 3022285... | 1,09586e-36 | 0,847222 |
| 63 | 62 | 5362119... | 3,86989e-38 | 0,861111 |
| 64 | 63 | 8511300... | 1,22264e-39 | 0,875 |
| 65 | 64 | 1196901... | 3,42219e-41 | 0,888889 |
| 66 | 65 | 1473109... | 8,38346e-43 | 0,902778 |
| 67 | 66 | 156238908 | 1,76978e-44 | 0,916667 |
| 68 | 67 | 13991544 | 3,15455e-46 | 0,930556 |
| 69 | 68 | 1028790 | 4,6168e-48 | 0,944444 |
| 70 | 69 | 59640 | 5,32714e-50 | 0,958333 |
| 71 | 70 | 2556 | 4,54422e-52 | 0,972222 |
| 72 | 71 | 72 | 2,54784e-54 | 0,986111 |
| 73 | 72 | 1 | 7,0434e-57 | 1 |

Fig. 4.7. Juego con 12 Dados Diana.

CON 18 DADOS

| | k_valor | combinar | probabili | k |
|---|---|---|---|---|
| 99 | 98 | 3,87228e... | 2,34542e-64 | 0,907407 |
| 100 | 99 | 3911395... | 4,71549e-66 | 0,916667 |
| 101 | 100 | 3520256... | 8,44718e-68 | 0,925926 |
| 102 | 101 | 2788321... | 1,33175e-69 | 0,935185 |
| 103 | 102 | 1913554... | 1,81912e-71 | 0,944444 |
| 104 | 103 | 111469176 | 2,1092e-73 | 0,953704 |
| 105 | 104 | 5359095 | 2,01835e-75 | 0,962963 |
| 106 | 105 | 204156 | 1,53042e-77 | 0,972222 |
| 107 | 106 | 5778 | 8,62119e-80 | 0,981481 |
| 108 | 107 | 108 | 3,20742e-82 | 0,990741 |
| 109 | 108 | 1 | 5,91117e-85 | 1 |

Fig. 4.8. Juego con 18 Dados Diana.

Marcela afirmó que las tres opciones eran igualmente probables *(obtener más de un seis con 6 dados, obtener más de un seis en 12 dados, obtener tres seis en 18 dados)* al encontrar un valor en ellas igual a 1. Los tres cocientes son arbitrarios puesto que establecer un cociente entre "supuestamente" dos variables, que para Marcela fue una variable con dos utilidades (*k_valores y tamaño muestral)* es a fin de cuenta un mismo valor (número de caras de los dados) y si realizo una suma acumulada mediante la expresión *prev( )* va a dar al final 1. Cómo Marcela centró su atención únicamente en el último valor dado en las expresiones, dedujo erróneamente que las tres opciones eran igualmente probables.

Aunque al principio pensamos que Cristina identificó la situación como binomial, al concebir las tres situaciones como binomiales independientes e identificar el valor de *p* y posiblemente el tamaño muestral *n,* su programación y argumentación dan cuenta de otra situación. La no identificación del experimento Bernoulli y del tamaño muestral llevó a que se quedara sin repeticiones y diera doble uso a los valores de interés: como tamaño muestral y la función de ella misma. Además, expresar la función de probabilidad en comandos de Fathom no fue de gran importancia puesto que sus argumentos se apoyaron en los cocientes de tipo clásico presentado al final de la programación.

## 4.2 USO DE FATHOM PARA RESOLVER SITUACIONES BINOMIALES.

Aunque existieron programaciones en Fathom que se limitaron a construir un atributo con la misma fórmula de la binomial mediante la trascripción de la expresión binomial $P(X=k)=\binom{n}{k}p^k q^{n-k}$ en comandos de Fathom, se presentaron ciertos errores en la programación, posiblemente, porque no identificaron las situaciones como binomiales o las consideraron parcialmente como tales.

Aunque las programaciones pueden que no presenten ningún error trascriptivo de la expresión en comandos de Fathom, en ellas no se refleja si los estudiantes identificaron el experimento Bernoulli básico asociado a la situación, sino que simplemente identificaban la situación como binomial y establecían correctamente los parámetros *n* y *p.* Algunas programaciones encontradas son:

***Programación de Gerardo al Juego de los dados (Anexo 6, ejercicio 4)***

- Define tres "*collection"* llamadas 6 DADOS, 12 DADOS Y 18 DADOS y un *slider* fijo en *p=0.166* *p*ara establecer el tamaño muestral *n* y el valor de *p* de la situación; nuevamente se concibe el *slider* como un función determinista (la variable toma un solo valor o) en lugar de una función variacional (pueda tomar varios valores en un tiempo corto).

Además, identifica la situación compuesta por tres binomiales con igual valor de *p* y distinto tamaño muestral. Establecer los valores de interés mediante el conteo del número de dados de cada situación ayudó al momento de la programación. Los valores de interés de cada situación están registrados en la columna **valor.**

- Establece el valor de probabilidad binomial en la comuna **binomial** mediante la expresión $(combinations(\max(valor));valor))(p^{valor})(1-p)^{\max()-valor})$ con el fin de agilizar los cálculos operativos puesto que ésta expresión es la trascripción de la expresión binomial en comandos de Fathom.

- Mediante las expresiones $if(valor \geq 1)\begin{cases} binomial \\ 0 \end{cases}$, $if(valor \geq 2)\begin{cases} binomial \\ 0 \end{cases}$ y $if(valor \geq 3)\begin{cases} binomial \\ 0 \end{cases}$ en las tres situaciones binomiales (6 DADOS, 12 DADOS Y 18 DADOS) toma las probabilidades binomiales que cumplan la condición dada en cada Collection.

- Por último, mediante la expresión *prev(sumx )+x* realiza una suma acumulada de las probabilidades seleccionadas anteriormente para obtener la probabilidad pedida.

Resaltamos de la programación de Mauricio el hecho que concibió la situación compuesta por tres situaciones binomiales con igual valor de *p=1/6* y diferentes tamaños muestrales, pero a la vez no se evidencia que Mauricio haya identificado el experimento Bernoulli básico de la situación.

12 DADOS

| | valor | binomial | x | sumx |
|---|---|---|---|---|
| 2 | 1 | 0,269111 | 0 | 0 |
| 3 | 2 | 0,296094 | 0,296094 | 0,296094 |
| 4 | 3 | 0,197443 | 0,197443 | 0,493537 |
| 5 | 4 | 0,0888707 | 0,0888707 | 0,582407 |
| 6 | 5 | 0,0284455 | 0,0284455 | 0,610853 |
| 7 | 6 | 0,006638... | 0,006638... | 0,617492 |
| 8 | 7 | 0,001138... | 0,001138... | 0,61863 |
| 9 | 8 | 0,000142... | 0,000142... | 0,618772 |
| 10 | 9 | 1,26546e... | 1,26546e... | 0,618785 |
| 11 | 10 | 7,59456e... | 7,59456e... | 0,618786 |
| 12 | 11 | 2,76232e... | 2,76232e... | 0,618786 |
| 13 | 12 | 4,60497e... | 4,60497e... | $\boxed{0,618786}$ |

Fig 4.10. Juego con 12 dados. Gerardo

SEIS DADOS

| | valor | binomial | x | sumx |
|---|---|---|---|---|
| 1 | 0 | 0,334818 | 0 | 0 |
| 2 | 1 | 0,401878 | 0,401878 | 0,401878 |
| 3 | 2 | 0,200987 | 0,200987 | 0,602865 |
| 4 | 3 | 0,0536094 | 0,0536094 | 0,656474 |
| 5 | 4 | 0,008043... | 0,008043... | 0,664517 |
| 6 | 5 | 0,000643... | 0,000643... | 0,665161 |
| 7 | 6 | 2,14592e... | 2,14592e... | 0,665182 |

Fig 4.9. Juego con 6 dados. Gerardo

Aunque establece correctamente los valores de interés de cada situación y la expresión binomial en cada una de ellas. El uso del *slider* fue innecesario en esta situación puesto que el parámetro de esta situación era fijo en 1/6.

Las expresiones utilizadas para calcular el valor de probabilidad dan cuenta de la comprensión de aquellas situaciones donde se plantea calcular la probabilidad con expresiones como *" al menos (…)"* puesto que Mauricio la concibió como *"(…) igual y mayor a (…)".* Lo anterior lo notamos en las siguientes expresiones:

$$if\,(valor \geq 1)\begin{cases} binomial \\ 0 \end{cases}$$

$$if\,(valor \geq 2)\begin{cases} binomial \\ 0 \end{cases} \qquad if\,(valor \geq 3)\begin{cases} binomial \\ 0 \end{cases}$$

18 DADOS

| | valor | binomial | x | sumx |
|---|---|---|---|---|
| 6 | 5 | 0,103033 | 0,103033 | 0,532174 |
| 7 | 6 | 0,0446583 | 0,0446583 | 0,576832 |
| 8 | 7 | 0,0153151 | 0,0153151 | 0,592148 |
| 9 | 8 | 0,004212... | 0,004212... | 0,59636 |
| 10 | 9 | 0,000936... | 0,000936... | 0,597297 |
| 11 | 10 | 0,000168... | 0,000168... | 0,597465 |
| 12 | 11 | 2,45277e... | 2,45277e... | 0,59749 |
| 13 | 12 | 2,86225e... | 2,86225e... | 0,597493 |
| 14 | 13 | 2,64271e... | 2,64271e... | 0,597493 |
| 15 | 14 | 1,8881e-... | 1,8881e-... | 0,597493 |
| 16 | 15 | 1,00723e... | 1,00723e... | 0,597493 |
| 17 | 16 | 3,77802e... | 3,77802e... | 0,597493 |
| 18 | 17 | 8,89158e... | 8,89158e... | 0,597493 |
| 19 | 18 | 9,88191e... | 9,88191e... | 0,597493 |

Fig 4.11. Juego con 18 dados. Gerardo

Por último, sumó las probabilidades de interés en cada situación para obtener la probabilidad pretendida y poderlas comparar. Los resultados que tomó para comparar se encuentran resaltados al final de cada tabla, y en base a ellos argumentó que la primera opción era la más probable (*obtener un seis con 6 dados)* al tener un mayor valor de probabilidad.

Aunque nuestro interés principal radicaba en el análisis de las programaciones y argumentaciones dadas por los estudiantes a situaciones binomiales, también quisimos observar además, si identificaban el experimento Bernoulli básico del mismo, si concebían la situación como binomial y si establecían correctamente los parámetros *n* y *p* para calcular valores de probabilidad en tamaños muestrales pequeños.

Por ejemplo, la situación de la evaluación fue concebida como binomial por la mayoría de los estudiantes que realizaron las programaciones en Fathom, identificando correctamente el tamaño muestral *n=10 (número de preguntas del examen)* y el valor de *p=1/5 (la posibilidad de escoger la respuesta correcta de las cinco opciones de cada pregunta).* Sin embargo, se presentaron ciertos errores de tipo procedimental que llevaron a erradas conclusiones.

Por esa razón, analizaremos la programación de Marcela a la situación de la Evaluación (Ver Anexo 6, ejercicio 1), quien ya anteriormente había dado muestras de dificultades para programar en Fathom. Para calcular la probabilidad, Marcela realizó la siguiente programación:

***Programación de Marcela***

- *Definió una colección y un slider fijo en 0.2 (p=1/5).* El valor de 0.2, aunque la programación de la estudiante no refleja el origen de este valor como en las simulaciones computaciones mencionadas en la sección anterior, hace referencia al experimento Bernoulli básico del experimento: contestar bien una pregunta es escoger la respuesta correcta de 5 opciones dadas.

- Definió los 10 valores de interés con *PREGUNTAS=caseIndex-1*, para hacer referencia a las 10 preguntas de la evaluación, que vendrían a convertirse en las repeticiones del experimento.
- Estableció el componente combinatorio de PREGUNTAS mediante la expresión $COMBINACIONES = (combinations(10; PREGUNTAS))$. Además, agregó a la expresión anterior el componente probabilístico: regla del producto mediante la expresión

$RES = (combinations(10; PREGUNTAS))(P^{PREGUNTAS})(1-P)^{10-PREGUNTAS}$ para tener la expresión binomial completa.

- Por último, estableció ciertas condiciones a las expresiones para obtener los valores de probabilidad pedidos mediante las expresiones $SUMAS = if(PREGUNTAS \geq 6)\begin{cases} RES \\ 0 \end{cases}$ **)** **y** $SUMAS1 = if(PREGUNTAS < 1)\begin{cases} RES \\ 0 \end{cases}$ **.** Estas expresiones tomaban los valores de probabilidad que cumplían con la expresión descrita en el paréntesis después del *if ( ).*

Analizando la programación se observa que Marcela estableció los valores de interés (en este caso **PREGUNTAS:** los números del 0 al 10 hacían referencia al número de preguntas de la evaluación), definió un *slider* fijo en 0.2 ( $p = \frac{1}{5}$ que indicaba la posibilidad de contestar correctamente a una pregunta que contenía 5 posible respuestas).

Calculó el valor de probabilidad binomial a cada PREGUNTA y tomó las probabilidades que le interesaba (para el ítem a) mediante la expresión **SUMAS=**$if(PREGUNTAS \geq 6)\begin{cases} RES \\ 0 \end{cases}$ **),** es decir la probabilidad de contestar correctamente 6, 7, 8, 9 y 10 preguntas respectivamente; afirmando el valor de probabilidad para $P(X \geq 6)$ era el valor $P(X = 10) = 1.024 * 10^{-7}$, olvidando sumar las probabilidades antes calculadas; en realidad el verdadero valor de probabilidad para $P(X \geq 6) = 0.00636938$. Nuevamente, Marcela olvidó sumar el conjunto de los valores de probabilidad de la variable aleatoria de interés.

Para que la programación de Marcela pueda obtener el valor de $P(X \geq 6)$ descrito anteriormente, sólo falta sumar los valores de probabilidad para 6, 7, 8, 8 y 10 así:

PUNTO3

| | PREGUNTAS | COMBIN... | RES | SUMAS | prob6a10 |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 0,107374 | 0 | 0 |
| 2 | 1 | 10 | 0,268435 | 0 | 0 |
| 3 | 2 | 45 | 0,30199 | 0 | 0 |
| 4 | 3 | 120 | 0,201327 | 0 | 0 |
| 5 | 4 | 210 | 0,0880804 | 0 | 0 |
| 6 | 5 | 252 | 0,0264241 | 0 | 0 |
| 7 | 6 | 210 | 0,005505... | 0,005505... | 0,00550502 |
| 8 | 7 | 120 | 0,000786... | 0,000786... | 0,00629146 |
| 9 | 8 | 45 | 7,3728e-... | 7,3728e-... | 0,00636518 |
| 10 | 9 | 10 | 4,096e-06 | 4,096e-06 | 0,00636928 |
| 11 | 10 | 1 | 1,024e-07 | 1,024e-07 | 0,00636938 |

$P(X \geq 6) = 0.00636938$

Fig 4.12. Situación Evaluación de Marcela *"pasar el examen"*

**prob6a10=** $prev(prob6a10) + SUMAS$

Ahora pues, al momento de establecer la probabilidad de contestar *al menos una pregunta bien,* Marcela presentó dificultad al establecer $P(X < 1)$ en lugar de $P(X \geq 1)$, es decir, no comprendió la expresión: *"(…) de que conteste bien al menos una pregunta".* El error en la programación se presenta al establecer SUMAS1= $if(PREGUNTAS < 1)\begin{cases} RES \\ 0 \end{cases}$ en lugar de $if(PREGUNTAS \geq 1)\begin{cases} RES \\ 0 \end{cases}$ para obtener los valores de probabilidad para $P(X \geq 1)$ pedidos. La columna **probm1** muestra los valores de probabilidad para los valores comprendidos en $P(X \geq 1)$ y en la columna **mas1** la suma de estas probabilidades para establecer $P(X \geq 1)$

PUNTO3

| | PREGUNTAS | COMBIN... | RES | SUMAS | SUMAS1 | probm1 | mas1 |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0,107374 | 0 | 0,107374 | 0 | 0 |
| 2 | 1 | 10 | 0,268435 | 0 | 0 | 0,268435 | 0,268435 |
| 3 | 2 | 45 | 0,30199 | 0 | 0 | 0,30199 | 0,570425 |
| 4 | 3 | 120 | 0,201327 | 0 | 0 | 0,201327 | 0,771752 |
| 5 | 4 | 210 | 0,0880804 | 0 | 0 | 0,0880804 | 0,859832 |
| 6 | 5 | 252 | 0,0264241 | 0 | 0 | 0,0264241 | 0,886256 |
| 7 | 6 | 210 | 0,005505... | 0,005505... | 0 | 0,005505... | 0,891761 |
| 8 | 7 | 120 | 0,000786... | 0,000786... | 0 | 0,000786... | 0,892548 |
| 9 | 8 | 45 | 7,3728e-... | 7,3728e-... | 0 | 7,3728e-... | 0,892622 |
| 10 | 9 | 10 | 4,096e-06 | 4,096e-06 | 0 | 4,096e-06 | 0,892626 |
| 11 | 10 | 1 | 1,024e-07 | 1,024e-07 | 0 | 1,024e-07 | 0,892626 |

Fig 4.13. Evaluación Marcela. "Contestar bien una pregunta".

Como la utilización del software computacional se basó en la trascripción de la expresión binomial en comandos en Fathom, no se presentaron inconvenientes de tipo probabilístico y combinatorio, lo que da a pensar que Marcela concibió la situación desde un principio como binomial. Más bien, los inconvenientes presentados fueron de tipo interpretativo, argumentativo y procedimental al momento de establecer los valores de probabilidad.

Debido a la familiaridad de los estudiantes a situaciones binomiales, notamos que algunos de ellos redujeron el espacio muestral a los valores de interés llegando a conclusiones válidas mediante las programaciones realizadas en Fathom. Ejemplo de ello es la programación de Andrea quien, en la situación de la no acreditación del nuevo medicamento (Ver Anexo 6, ejercicio 5), aunque transcribió la expresión de la binomial en comandos de Fathom redujo los valores posibles de la situación a los valores de interés del mismo.

Los estudiantes que programaron esta situación mediante la trascripción en comandos de Fathom, identificaron el tamaño muestral como *n=12 pacientes* (tamaño muestral) y el valor de *p=1/2* (posibilidad de sanarse o no con el nuevo medicamento)

***Programación de Andrea***

- Define en la columna **v** los valores 7, 8, 9, 10, 11 y 12 como los valores de interés para hacer referencia a los 7, 8, 9, 10, 11 u 12 pacientes que se recuperan suministrándoles el nuevo medicamento.
- Establece el componente combinatorio $combi = combinations(12;v)$. En este caso, el valor de 12 hace referencia a las repeticiones de la situación binomial, los cuales se convirtieron en los 12 pacientes que tomaron el nuevo medicamento para la situación tratada.
- Establece la probabilidad binomial a cada valor de interés mediante la expresión $proba = combi(0.5)^{v}(0.5)^{12-v}$. Aunque posiblemente la estudiante no haya percibido con detalle el experimento Bernoulli de la situación, al

momento de establecer el valor de *p=0.5* está mostrando que el experimento básico tratado es el sanarse o no con el nuevo medicamento.

- Establece la expresión $suma = prev(suma) + proba$ para establecer el valor de la probabilidad pedida como una suma de las probabilidades obtenidas individualmente.

Luego, en la columna **proba** estableció los valores de probabilidad binomial a cada valor mediante la expresión $(combi)(0.5)^{v}(0.5)^{12-v}$ donde $n = 12$ es el número de pacientes de la muestra, $p = \frac{1}{2}$ la probabilidad de sanarse o no con el nuevo medicamento y **combi=** $combinations(12; v)$ son las combinaciones que se originaron de los **v** valores y 12.

Como su interés recaía en calcular el valor de probabilidad de $P(X \geq 7)$, y en la columna **proba** se encuentran los valores de probabilidad para $P(X = 7)$, $P(X = 8)$, $P(X = 9)$, $P(X = 10)$, $P(X = 11)$ y $P(X = 12)$, se hizo necesario hacer la suma de esos valores de probabilidad mediante la expresión $prev(suma) + proba$ para obtener finalmente $P(X \geq 7) = 0.387207$. Por último, Andrea afirma que 0.387207, es decir el 38.72%, es la probabilidad de que el medicamento nuevo sea acreditado.

Nuevamente se evidencia la poca comprensión del lenguaje oral en las situaciones, puesto que la pregunta planteada de la situación hacía referencia a la *no acreditación del medicamento* y la mayoría de los estudiantes dieron respuesta a la acreditación del mismo; únicamente Mauricio da argumentos de la no acreditación del medicamento.

Collection 1

| | v | combi | proba | suma |
|---|---|---|---|---|
| 1 | 7 | 792 | 0,193359 | 0,193359 |
| 2 | 8 | 495 | 0,12085 | 0,314209 |
| 3 | 9 | 220 | 0,0537109 | 0,36792 |
| 4 | 10 | 66 | 0,0161133 | 0,384033 |
| 5 | 11 | 12 | 0,002929... | 0,386963 |
| 6 | 12 | 1 | 0,000244... | 0,387207 |

Fig. 4.14. Situación Medicamento Andrea.

**A manera de conclusiones:**

Las posibles incomprensiones que podrían generarse en torno a la distribución binomial teniendo presente la utilización de Fathom serían:

- Olvidar el componente combinatorio y probabilístico de la expresión binomial. Sin embargo, mediante la utilización de Fathom no se presenta tal situación; puesto que una vez identificada la situación como binomial, la programación de los estudiantes en Fathom es la trascripción de la expresión binomial.
- La poca comprensión de enunciados dificulta establecer las desigualdades pedidas. Se nota dificultad, principalmente, para establecer las desigualdades de probabilidades en preguntas que tengan inmersa la expresión *"(…) al menos…".*
- En algunos casos, olvidan sumar las probabilidades binomiales para obtener el valor de probabilidad pedido.
- La trascripción de la expresión binomial en comandos de Fathom no refleja si los estudiantes identifican el experimento Bernoulli básico asociado a toda situación binomial. Mientras que al realizar una simulación de la situación como modelo de urna sí se evidencia tal experimento.
- La no identificación de la situación como binomial, o la identificación o confusión de los parámetros $n$ y $p$ llevó a que los estudiantes realizaran diversos procedimientos en base a probabilidad clásica y normal.

# 5. CONCLUSIONES

En este capítulo presentamos las conclusiones generales de la investigación en donde analizamos las dificultades presentadas, las concepciones construidas y los errores que persistieron en los estudiantes al enfrentarse con situaciones binomiales a través de la simulación computacional.

- En algunos estudiantes persistió la heurística de la representatividad al no tener presente el efecto del tamaño muestral en la variabilidad de las frecuencias relativas, argumentando igualdad de probabilidad al existir igualdad de proporciones en los tamaños muestrales. Esto se evidenció en la evaluación final (Ver Capítulo 3).

- Aún persiste el sesgo de equiprobabilidad al afirmar que las situaciones son igualmente probables si el valor de probabilidad entre ellas son iguales independientemente del tamaño muestral de la situación binomial. Estos estudiantes suelen afirmar que las opciones son igualmente probables si *p=0.5 y q=0.5* sin tener en cuenta el tamaño muestral de ambas situaciones, lo cual lleva a errados argumentos al no concebir el efecto del tamaño muestral en los valores de probabilidades. Esto se evidenció en la situación: Estudio Demográfico (Ver Capítulo 3).

- En la evaluación final, los estudiantes no identificaron situaciones binomiales al no tener presente, en la mayoría de los casos, el componente combinatorio o probabilístico de la función de probabilidad $P(X=k)=\binom{n}{k}p^k q^{n-k}$. Esta dificultad se evidencia en los estudiantes que resolvieron las situaciones problemas en forma algebraica, porque aquellos estudiantes que utilizaron Fathom para simular el experimento llegaron a válidos resultados mostrando comprensión de situaciones binomiales realizando programaciones en Fathom.

La no identificación de situaciones binomiales puede haberse debido a que no identificaron el experimento Bernoulli o no establecieron correctamente los parámetros $n$ y $p$ de la situación planteada.

- El establecimiento incorrecto del tamaño muestral en situaciones binomiales llevó a que los estudiantes realizaran procedimientos de tipo normal, teniendo en cuenta el Teorema De Moivre.-Laplace, o realizaran procedimientos de tipo clásico (Ver Capítulo 3, Situaciones: Estudio demográfico y Juego Dados).

  Los errores al establecer el tamaño muestral $n$ se presentaron, en la mayoría de los casos, porque no tenían presente que la situación problema estaba compuesta por dos situaciones binomiales y tomaban el mayor valor, fuese el tamaño muestral o no (Ver Capítulo 3, Situación: Estudio Demográfico). También cuando confundían los valores de interés con el tamaño muestral ó establecían una relación entre ellos quedándose sin repeticiones, y por ende sin binomial (Ver Capítulo 3, Situación: Evaluación).

- Cuando los estudiantes no percibían las situaciones como binomiales, solían realizar procedimientos de tipo clásico o mediante la regla del producto. Aunque los procedimientos de la regla del producto no fueron del todo válidos a las situaciones planteadas, consideramos que fue un buen acercamiento a situaciones binomiales que no tuvo en cuenta el componente probabilístico de la binomial.

  Aplicar la regla del producto en tales situaciones mostró que los estudiantes son concientes del efecto del tamaño muestral y el valor de $p$ y $q$ en los valores de probabilidad pedidos.

- Al principio se presentó dificultad en el cálculo de probabilidades binomiales para muestras grandes, la cual fue superada mediante las simulaciones computacionales realizadas en Fathom mostrando gráficamente la similitud del histograma binomial con una curva normal.

- Se presentaron dificultades al calcular valores de probabilidad para eventos definidos mediante la expresión *"(…) al menos (…)"* puesto que los estudiantes concebían tal expresión como *"(…) igual a (…)"* en lugar de *"(…)mayor o igual a (…)".* La situación anterior se presentó en nuestra investigación (Ver Capítulo 3) llevando a respuestas erradas debido a la mala interpretación de enunciados.

- En algunas ocasiones, los estudiantes presentaron dificultad al momento de simular computacionalmente el experimento, posiblemente porque no identificaban las características propias del experimento real o porque desconocían las funciones en Fathom que hacían posible la representación computacional del experimento.

- Durante las simulaciones de situaciones binomiales en Fathom se presentaron dos tipos de programaciones: la primera y más utilizada fue la trascripción de la expresión binomial en comandos de Fathom que ayudaba a agilizar y conservar los cálculos obtenidos en forma algebraica para dedicar más tiempo a la argumentación de la situación; y la segunda y menos utilizada, pero realmente más significativa desde el punto de vista de simulación computacional: la generación de muestras que llevaba una mejor comprensión de situaciones binomiales puesto que, al establecer el espacio muestral como modelo de urna, las muestras y las medidas se hacía referencia al experimento Bernoulli y a las repeticiones asociadas a situaciones binomiales.

- A través del proceso de instrucción llevado a cabo, la mayoría de los estudiantes lograron identificar las diversas situaciones planteadas como binomiales, al poder establecer los experimentos Bernoulli asociados a cada situación, los parámetros *n* y *p,* y el cálculo de probabilidades en pequeñas y grandes muestras.

- Comprender el concepto de distribución binomial está relacionado en la utilización de este concepto en otros temas como: cálculo de probabilidades, estimación puntual de parámetros, intervalos de confianza y pruebas de hipótesis. Aunque no se realizó un análisis de esta temática, resaltamos la importancia de tratar esos temas porque muestran la aplicabilidad de la binomial en estadística y probabilidad.

- La simulación computacional realizada en Fathom ayudó a que los estudiantes hicieran un análisis más detallado de la situación planteada, al identificarla como binomial al momento de programar.

# BIBLIOGRAFÍA


Alvarado, H. (2007). Significados Institucionales y Personales del Teorema Central del Límite en la enseñanza de estadística en ingeniería. Tesis Doctoral. Universidad de Granada.

Batanero, C. (2001). Investigaciones sobre Razonamiento Estadístico y Dificultades. En: Didáctica de la Estadística, (pp. 55-113) Departamento de Didáctica de las Matemáticas. Universidad de Granada.

Biehler, R. (1991). Computer in Probability Education. En Kapadia, R., Borovnik, M : (editores). *Chance Encounters :Probability in Educations,* Cap 6, 169-211. Kluwer Academic Publishers.

Brousseau, G. (1986). Fondements et Méthodes de la Didactique des Mathématiques*. Recherches en Didactique des Mathématiques*. 7 (2), 33-105.

Cumming, G. y Fidler, F. (2005) Interval estimates for statistical communication: problems and possible solutions. IASE/ISI.

Cumming , G., Williams, J, y Fidler, F. (2004). Replication, and researches´ understanding of confidence intervals and standard error bars. Understanding Statistic, 3, 299-311.

Finzer, W., Erickson, T., Binker, J., (2000), Fathom:Dymanic Statistic. Key Currículum Press.

Godino, J. (1996) Significado y comprensión de los objetos matemáticos. En: L.Puig, y A. Gutierrez (Eds.), *Proceedings of the 20th Conference of the Internacional Group for the Psychology of Mathematics Education* (v.2, pp. 417-424). Universidad de Valencia.

Godino, J. y Batanero, C. (1998) Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. En: A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics Education as a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Mantilla, M. y Martínez, M. (2007) Construcción de significados del concepto de probabilidad frecuencial en un ambiente computacional. Una experiencia con profesores en formación. Trabajo de Grado Licenciatura en Matemáticas. Escuela de Matemáticas. Universidad Industrial de Santander.

Olivo, E y Batanero C. (2007) Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. En: Revista Iberoamericana de Educación Matemática UNIÓN, N° 12, pp. 37-51.

Tauber, L. (2001) La construcción del significado de la distribución normal a partir de actividades de análisis de datos. Tesis Doctoral. Departamento Didáctica de la Matemática. Universidad de Sevilla.

Yáñez, G. (2003) Estudios sobre el papel de la simulación computacional en la comprensión de las secuencias aleatorias, la probabilidad y la probabilidad condicional. Tesis doctoral no publicada. CINVESTAV-IPN, México.

## Anexo 1. Guía: LA DISTRIBUCIÓN BINOMIAL

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**10 de agosto de 2007**

### Objetivo

- Se pretende analizar el comportamiento de la distribución binomial y descubrir la manera como los parámetros *n* y *p* la afectan. El análisis es completamente gráfico.

### Primera Actividad

Utilizando Fathom construya una colección de datos que le permitan realizar un gráfico de la distribución binomial. Como valor *n* tome el producto de 10 por el número que le adjudique el profesor. Como la idea es que el parámetro *p* varíe libremente entre 0 y 1, utilice un *slider* para definirlo. Para ello basta arrastrar el icono de slider, definirlo como *p* y restringir su intervalo de variación entre 0 y 1 en la misma en como se cambian las escalas en los ejes de los gráficos.

### Segunda Actividad

Construya el gráfico que relaciona el valor de probabilidad con los posibles valores de *k* de la distribución binomial y active el slider para observar la forma como cambia la distribución binomial. Saque algunas conclusiones respecto a la relación que pueda existir entre la forma del gráfico de la distribución binomial y los diferentes valores de *p*.

### Tercera Actividad

Reúnase con dos compañeros que hayan trabajado con valores distintos de *n,* analicen sus respectivos gráficos y escriban sus conclusiones respecto a la forma como los parámetros *p* y *n* afectan el gráfico.

**Cuarta Actividad**

Repita las actividades anteriores caminando los k-valores por las proporciones o probabilidades posibles de éxitos. Escriba sus conclusiones individuales y participe en la escritura de las conclusiones de su grupo de trabajo.

**Quinta Actividad**

En una urna existe una cierta composición de bolas amarillas que se pretende averiguar. Para ello se realizan 100 extracciones con reposición y se obtienen 35 bolas amarillas.

a). Con la información obtenida ¿considera plausible el valor 0.29 para la proporción de amarillas? Explique su respuesta.

b) ¿Y el valor 0.42? Explique su respuesta.

c) Construya un intervalo de valores posibles *(Intervalos de Confianza)* y explique los motivos para su construcción.

Dé un algoritmo que le permita construir "intervalos de confianza" para estimar el valor de la probabilidad asociado a un cierto éxito.

**Sexta Actividad**

El rector de una universidad dice "el 99% de los estudiantes apoya el nuevo calendario académico". Para corroborar la información del rector se realizó una encuesta y se entrevistó a 200 estudiantes.

a). Si 180 de esos estudiantes manifestaron estar de acuerdo con el nuevo calendario, ¿se puede decir que el rector estaba en lo cierto?

b) A tu juicio, ¿cuál debe ser el número mínimo de estudiantes que debe responder positivamente para que no se diga que el rector se equivocó?

**Anexo 2. Segunda Guía: CÁLCULO DE PROBABILIDADES**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**22 de agosto de 2007**

Nombre:____________________________________________________

**Actividad 1.**

La siguiente tabla reporta las temperaturas promedio mensuales para dos ciudades A y B en un país con estaciones.

| | Ene | Feb | Mar | Abr | May | Jun | Jul | Ago | Sep | Oct | Nov | Dic |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ciudad A | 4 | 6 | 10 | 15 | 19 | 23 | 26 | 25 | 22 | 16 | 11 | 6 |
| Ciudad B | 9 | 11 | 12 | 13 | 14 | 17 | 17 | 18 | 18 | 16 | 13 | 9 |

1. ¿Cuál de las dos ciudades tiene mejor clima? Explique su respuesta.
2. ¿Cómo sugiere que podría medir la variabilidad de la temperatura en ambas ciudades?
3. Reúnase con un compañero y discuta su medida de variabilidad y traten de llegar a un acuerdo sobre cuál de las dos medidas es mejor.

**Actividad 2**

Dada la variable aleatoria X con distribución binomial $B(10,0.5)$ calcule la probabilidad de los siguientes eventos:

d) Que el valor tome valores mayores que 2 y menores que 7, es decir, halle la probabilidad $P(X \in (2,7))$.

e) Que el valor tome valores mayores o iguales que 3, es decir, halle la probabilidad $P(X \geq 3)$.

f) Que el valor tome valores menores que 8, es decir, halle la probabilidad $P(X < 8)$.

**Actividad 3**

Dada la variable aleatoria $X$ con distribución binomial $B(100,0.3)$ calcule la probabilidad de los siguientes eventos:

a) Que el valor tome valores mayores que 12 y menores que 85, es decir, halle la probabilidad $P(X \in (12,85))$.

b) Que el valor tome valores mayores o iguales que 23, es decir, halle la probabilidad $P(X \geq 23)$.

c) Que el valor tome valores menores que 79, es decir, halle la probabilidad $P(X < 79)$.

d) Comente las dificultades que tuvo calculando las probabilidades anteriores y sugiera algún mecanismo más cómodo que pudiera facilitarle los cálculos.

**Actividad 4.**

b) Dada la variable aleatoria $X$ con distribución binomial $B(10,0.5)$ halle dos intervalos de valores cuya suma de probabilidades de 0.4. Halle el intervalo más pequeño que satisfaga la misma condición, es decir, halle el intervalo $(a,b)$ de menor longitud tal que $P(X \in (a,b)) = 0.4$.

c) Dada la variable aleatoria $X$ con distribución binomial $B(10,0.8)$ halle dos intervalos de valores cuya suma de probabilidades de 0.95. Halle el intervalo más pequeño que satisfaga la misma condición, es decir, halle el intervalo $(a,b)$ de menor longitud tal que $P(X \in (a,b)) = 0.95$.

d) Describa el método que utilizó para responder las dos preguntas anteriores y describa el algoritmo general que utilizaría para calcular el intervalo mínimo para cualquier distribución binomial.

e) Comente las dificultades que tuvo hallando el menor intervalo y sugiera algún mecanismo más cómodo que pudiera facilitarle los cálculos.

**Anexo 3. Guía: TEOREMA DE MOIVRE-LAPLACE**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**29 de agosto de 2007**

Nombre:_______________________________________________________

**Ejercicio 1.** Se lanza una moneda honesta 100 veces.

a) ¿Cuál es la probabilidad de que el número de caras no sea superior a 77?
b) ¿Cuál es la probabilidad de que el número de caras no sea inferior a 10?
c) ¿Cuál es la probabilidad de que el número de caras obtenidas sea al menos 17 y no mayor que 39?

**Ejercicio 2.**

Si ahora se tratara de realizar un juego de adivinación del número de caras que van a resultar y usted puede participar escogiendo tantos números como quiera siempre y cuando pague por ellos una cierta cantidad de dinero, ¿Cuáles números escogería para estar seguro que en el largo plazo ganaría y además que le cuesten la menor cantidad de dinero?

## Anexo 4. Guía: ESTIMACIÓN PUNTUAL DE PARÁMETROS

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**7 de septiembre de 2007**

Nombre:_______________________________________________________

6. Forme pareja con un compañero y tomen una de las urnas disponibles. El problema consiste en estimar la proporción de bolas amarillas *(p)* que contiene la urna, realizando extracciones con sustitución. Realicen 10 veces el experimento de realizar 20 extracciones de la urna, y calculen la proporción de urnas amarillas obtenidas. Con estos 10 valores realicen una estimación de la proporción de amarillas que tiene la urna argumentando claramente su respuesta. No se niega el uso de Fathom.

7. Junten sus 10 resultados con los de las demás parejas, grábenlos en Fathom y realicen un nuevo análisis que les permita estimar el valor de la proporción de amarillas en la urna.

8. Para saber si el comportamiento de las proporciones muestrales (en este caso con muestras de tamaño 20) es semejante al observado en los numerales anteriores, y se puede hacer generalizaciones, realicen simulaciones en Fathom. Para poder hacer simulaciones en Fathom se requiere conocer el valor de *p* (asumen el que el profesor les asigne). La idea es suponer conocido el valor de *p* con el ánimo de saber qué propiedades tiene la distribución de las medias muestrales extraídas de esa población. Tomen 1000 muestras de tamaño 20, calculen la media de cada muestra y analicen la distribución de esos valores. Con base en los resultados, refinen, si lo consideran necesario, la estimación de la proporción de amarillas en la urna.

9. Si conocido el valor de *p* se conoce también la distribución de las medias muestrales (y por tanto su media y su desviación estándar) la pregunta es: ¿Cómo utilizamos esta información cuando no conocemos *p* y debemos precisamente estimarlo cuando solo se tiene una media muestral?

10. Por la variabilidad intrínseca que tienen las medias muestrales tenemos que aceptar que el estimador $\overline{X}_n$ (media muestral) obtenido con base en una sola muestra no tiene el mínimo chance de coincidir con el parámetro *p* que estamos buscando. Por esta razón, lo mejor es aceptar un error $\varepsilon$ , es decir, asumir el intervalo $\overline{X}_n \pm \varepsilon$ como una mejor estimación, esto es, en lugar de dar un solo valor posible de *p*, damos todo un intervalo de valores $(\overline{X}_n - \varepsilon, \overline{X}_n + \varepsilon)$. La pregunta ahora es: ¿Cómo calculamos ese intervalo, mejor dicho, cómo calculamos el *margen de error* $\varepsilon$*?* Para responder la pregunta, en el gráfico de la distribución muestral de las proporciones trace el valor de una de las proporciones muestrales obtenidas en el primer numeral. Con base en este valor, construya un intervalo que contenga el valor de *p.* Repita el mismo procedimiento para las demás medias muestrales construyendo intervalos que contengan el valor de *p.* Con base en los intervalos construidos diseñe un algoritmo que permita construir un intervalo que contenga la media poblacional *p* cuando solo se tenga una media o proporción muestral. Como es claro que no hay certeza absoluta de que todos los intervalos contengan el valor de *p,* ¿qué probabilidad de contener el valor *p* tienen los intervalos que construyeron?

**Anexo 5. Guía: INTERVALOS DE CONFIANZA**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**14 de septiembre de 2007**

Nombre:________________________________________________

**Ejercicio 1.**

Se lanza una moneda honesta 100 veces. Con confianza 90% halle el número de caras que se espera obtener.

**Ejercicio 2.**

¿Son comunes las conductas sexuales que implican un riesgo de contraer el Sida? La Encuesta Nacional sobre el Sida en USA encuestó una muestra aleatoria de 2673 adultos heterosexuales. De éstos 170 tuvieron más de una pareja en el último año. Basándonos en esto, ¿qué se puede decir sobre el porcentaje de todos los adultos heterosexuales que tienen múltiples parejas?

**Ejercicio 3.**

(a) Si al realizar 10 extracciones de una urna se obtiene una proporción de bolas negras igual a 0.37, halle los intervalos de confianza que garantice que la probabilidad (nivel de confianza) de que la proporción real de negras se encuentre en ese intervalo sea

(1) 80% (2) 90% (3) 95% (4) 99%

(b) Responda la misma pregunta pero ahora con 50 extracciones.

(c) Responda la misma pregunta pero ahora con 100 extracciones.

(d) Con base a las respuestas dadas a las preguntas anteriores, saque conclusiones acerca de la relación entre el tamaño del intervalo de confianza, el nivel de confianza y el tamaño muestral.

**Ejercicio 4.**

(a) ¿Cómo afecta el nivel de confianza el tamaño del intervalo de confianza? Es decir, se quiere saber qué sucede con el tamaño del intervalo si se aumenta (disminuye) el nivel de confianza. Demuestre su conjetura.

(b) ¿Cómo afecta el tamaño muestral el tamaño del intervalo de confianza? Es decir, se quiere saber qué sucede con el tamaño del intervalo si se aumenta (disminuye) el nivel de confianza. Demuestre su conjetura.

**Ejercicio 5.**

Se toman muestra de tamaño 10 de una distribución normal y se calcula la media muestral. Se quiere saber cuál es la distribución de probabilidad de estas medias muestrales. (FATHOM).

## Anexo 6. EVALUACIÓN FINAL

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**21 de septiembre de 2007**

Nombre:_______________________________________________________

Nota: para responder las preguntas que aquí se propone, puede utilizar todos los medios a su alcance: los desarrollos teóricos o la simulación utilizando Fathom.

2. El notario de cierta población está muy interesado en prever el sexo de los recién nacidos para un estudio demográfico. Si se sabe por experiencia que las posibilidades son iguales para cada uno de los sexos, ¿Cuál de estos casos le parece más probable?
   a) Que en los próximos 10 nacimientos 8 o más sean varones.
   b) Que entre los próximos 100 nacimientos, 80 o más sean varones.
   c) Son igualmente probables.

3. Complete la siguiente tabla: En los espacios, dé los intervalos donde considere se pueden encontrar los valores de las frecuencias relativas al repetir el experimento las veces indicadas en las columnas.

| Probabilidad/Repeticiones | 50 | 100 | 500 |
|---|---|---|---|
| 0.2 | | | |
| 05 | | | |
| 0.8 | | | |

4. Un estudiante responde un examen de escogencia múltiple en el cual hay 10 preguntas cada una con 5 alternativas, de las cuales solo una es la

respuesta correcta. El estudiante no sabe nada del cuestionario y resuelve contestar al azar.

c) ¿Cuál es la probabilidad de que pase el examen (al menos 6 respuestas buenas)?
d) ¿Cuál es la probabilidad de que conteste bien al menos una pregunta?

5. Samuel Pepys, cuyos diarios relatan la vida del siglo XVIII en Inglaterra, era amigo del Sir Isaac Newton. Su interés en el juego lo llevó a preguntarle a Newton cuál de los siguientes resultados es más probable de obtenerse:
   a) Al menos un 6 cuando se lanzan seis dados.
   b) Al menos dos 6 cuando se lanzan 12 dados.
   c) Al menos tres 6 cuando se lanzan 18 dados.

   La pareja intercambió varias cartas antes de que Newton fuera capaz de convencer a Pepys de que ( ) era más probable. Llene el espacio y de sus razones para ello.

6. Suponga que la experiencia ha demostrado que solo 1/3 de todos los pacientes que tienen una cierta enfermedad se recobrarán si se les administra el tratamiento estándar. Una nueva droga va a ser probada con 12 voluntarios. Si las regulaciones de la salud requieren que al menos 7 de estos pacientes deben sanarse antes de que la nueva droga se permita, ¿cuál es la probabilidad de que la droga no sea acreditada incluso si aumenta la tasa de recuperación a ½?

7. De una urna que contiene bolas blancas y negras se realizan 50 extracciones con sustitución y se obtienen 30 bolas blancas.
   a) Construya un intervalo de confianza del 95% de confianza para el valor verdadero de la proporción de bolas blancas en la urna.

b) ¿Qué tan factible es que en la urna hayan 15 bolas blancas? Explique su respuesta.
c) ¿Y que hayan 40 bolas blancas? Explique su respuesta.

8. Encierre en un círculo la expresión adecuada para obtener una afirmación verdadera:
   a) Una forma de aumentar la precisión de la estimación de la probabilidad de éxito en una distribución binomial es aumentando/disminuyendo el nivel de confianza.
   b) Otra forma de aumentar la precisión de la estimación de la probabilidad de éxito en una distribución binomial es aumentando/disminuyendo el tamaño de la muestra.

9. Responsa Verdadero o Falso las siguientes afirmaciones explicando sus respuestas.

i. Para mejorar la precisión de una estimación por intervalos lo mejor es aumentar tanto el nivel de confianza como el tamaño muestral. (V) (F).
ii. El tamaño de un intervalo de confianza es directamente proporcional a su nivel de confianza. (V) (F).
iii. Al disminuir el tamaño muestral se aumenta el tamaño del intervalo de confianza. (V) (F)
iv. Un intervalo del 90% de confianza para estimar la media de una binomial contiene el 90% de los éxitos posibles. (V) (F).
v. El nivel de confianza del 90% significa en promedio de cada 100 intervalos de confianza que se construyan con muestras del mismo tamaño, 10 intervalos no contienen el valor del parámetro buscado.

## Anexo 7. PEQUEÑA EVALUACIÓN

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**PEQUEÑA EVALUACIÓN**

- Una encuesta de opinión pregunta a una muestra de 500 adultos si están a favor de dar subsidios a padres que tengan niños en edad escolar. Suponga que el 45% de la población está a favor de la idea. ¿Cuál es la probabilidad que más de la mitad de la muestra está a favor de la idea?