CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL IPN

Unidad Distrito Federal

DEPARTAMENTO DE MATEMÁTICA EDUCATIVA

Significados que estudiantes universitarios atribuyen a las distribuciones muestrales en un ambiente de simulación computacional y estadística dinámica

Tesis que presenta:

Santiago Inzunsa Cázares

para obtener el Grado de

Doctor en Ciencias

en la Especialidad de

Matemática Educativa

Director de tesis: Dr. Ernesto Alonso Sánchez Sánchez

México, Distrito Federal Febrero de 2006

Anexo 1

SIGNIFICADO INSTITUCIONAL DE LAS DISTRIBUCIONES MUESTRALES EN UN CURSO INTRODUCTORIO DE ESTADÍSTICA A NIVEL UNIVERSITARIO

1. Introducción

Como ya se señaló en el capítulo 1, en el marco teórico que se apoya nuestra investigación, el significado de los objetos matemáticos depende del contexto o institución donde son utilizados, y este, se concibe como el sistema de prácticas que las personas realizan para resolver problemas de los que emerge el objeto en cuestión. De tal manera, una investigación desde esta perspectiva teórica, debe tener como punto de partida la identificación y la descripción del significado institucional del objeto.

En nuestro caso, el objeto de interés son las distribuciones muestrales en un curso introductorio de estadística a nivel universitario, por lo que nos hemos planteado identificar el sistema de prácticas prototípicas que la institución de los educadores estadísticos (significado institucional) en este nivel, utiliza para resolver problemas de donde emerge este concepto. La pregunta anterior se ubica en la intersección de las categorías epistemológica y semiometría de la agenda para la investigación en educación matemática asociada al marco teórico, que se ha definido en Godino y Batanero (1998).

Determinar dicho significado, es importante en nuestra investigación, pues este será tomado como referencia para el diseño de actividades de enseñanza y para establecer una comparación con los significados que los estudiantes desarrollen en el proceso de instrucción a que serán expuestos durante el estudio. Este análisis a nivel macro, de la correspondencia entre significado personal de los estudiantes con el significado institucional de referencia, será un indicador de la comprensión y las dificultades de los estudiantes en el proceso de aprendizaje del concepto.

Para determinar dicho significado, hemos realizado un análisis epistémico de una muestra representativa de libros de texto de estadística para estudiantes universitarios, como son los de ingeniería, ciencias de la computación, económicas y administrativas. Dichos libros son una referencia confiable para determinar el significado, pues son utilizados por los profesores para impartir su clase de probabilidad y estadística, además, sus autores poseen reconocido prestigio en la materia, lo que ha dado lugar a que algunos de los textos hayan sido editados y reimpresos varias veces.

La relación de libros consultados es la siguiente:

1. *Probabilidad y Estadística para Ingenieros*. Sexta edición. (Walpole, Myers y Myers, 1999). Editorial Prentice Hall, México.
2. *Probabilidad y Estadística para Ingenieros de Miller y Freund.* Quinta edición. (Johnson, 1997). Editorial Prentice Hall, México
3. *Estadística Matemática con Aplicaciones.* Segunda edición. (Mendenhall, Wackerly y Scheaffer, 1994). Editorial Iberoamérica, México.
4. *Introducción a la Probabilidad y Estadística.* Primera edición. (Mendenhall, Beaver y Beaver, 2002). International Thompson Editores. México.
5. *Probabilidad y Estadística aplicada a la Ingeniería.* Segunda edición. (Montgomery y Runger, 2003). Editorial Limusa, México.
6. *Estadística Aplicada Básica.* (Moore, 1995). Antoni Bosch Editor, España.
7. *Estadística Aplicada a la Administración y a la Economía.* Tercera edición. (Kazmier, 1998). Editorial Mc Graw Hill, México.

8. *Probabilidad y Estadística para Ingeniería y Administración.* Tercera edición. (Hines y Montgomery, 1993). Editorial CECSA, México

Hemos restringido nuestro análisis a las distribuciones muestrales de la media y la proporción, por ser dos estadísticos muy importantes que cubren una amplia gama de aplicaciones en los textos de estadística, y a los que con frecuencia se reduce la enseñanza de las distribuciones muestrales en los primeros cursos universitarios.

2. Elementos de significado

2.1 Situaciones-problema (elementos fenomenológicos)

En todos los libros consultados, se observó la tendencia de considerar principalmente distribuciones muestrales normales o aproximadamente normales, ya sea porque se asume que la distribución de la población es normal, o porque el tamaño de la muestra es lo suficientemente grande, por lo general mayor a 30, -una consecuencia del teorema del limite central-.

Usualmente, las situaciones de normalidad en las distribuciones muestrales de los problemas abordados, las clasifican en diversos casos, -lo que muchas veces es fuente de confusión para los estudiantes-, de acuerdo a ciertos supuestos que hacen sobre la distribución, la desviación estándar o la varianza. La clasificación es la siguiente:

1) Población normal, σ conocida, cualquier tamaño de muestra.
2) Población cualquiera, tamaño de muestra grande ($n \geq 30$), σ conocida o desconocida. (aplican el teorema del limite central).
3) Población normal, tamaño de muestra pequeño ($n < 30$), σ conocida.
4) Población normal, tamaño de muestra pequeño ($n < 30$), σ desconocida.

En los primeros tres casos utilizan la distribución de probabilidad normal para calcular las probabilidades, mientras que en el último caso, utilizan la distribución t-student.

En cuanto a las formas de razonamiento implicadas para resolver los diversos problemas donde se aplican las distribuciones muestrales, identificamos dos categorías bien diferenciadas:

1. **Problemas que implican un razonamiento deductivo.**

En este tipo de problemas se suponen conocidos, la distribución de la población y los valores de sus parámetros, y a partir de ello, se plantea conocer aspectos como:

- Determinar la probabilidad de ciertos resultados muestrales.
- Construir intervalos de confianza en términos de desviaciones estándar, esto es, utilizando propiedades de la distribución normal.
- Determinar el tamaño de la muestra, para un error de muestreo y una confianza especificada.

Cabe señalar que es la primer situación la que predomina en todos los libros, y constituye el tipo de problemas más prototípicos de la categoría deductiva. Algunos ejemplos de este campo de problemas se muestran a continuación:

> Una empresa utiliza una máquina para llenar las botellas de un refresco. Se supone que las botellas contienen 300 ml. En realidad el contenido de las botellas varía según una distribución normal de media de 298 ml. y desviación estándar de 3 ml. ¿Cuál es la probabilidad que los contenidos de un paquete de 6 botellas sea menor que 295ml.? (Moore, 1995).

En este tipo de problemas, se sabe de antemano que la población posee distribución normal y la desviación estándar de la población es conocida, así aplicando el teorema del limite central sabemos que la distribución muestral tiene también una distribución normal con media $\mu_{\bar{x}} = 298$ y desviación estándar igual a $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{6}}$. Posteriormente se procede a estandarizar la distribución muestral ($z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$) y calcular el área que corresponde a z en las tablas de la distribución normal.

Otra situación frecuente es la que se presenta a continuación:

> Un ingeniero de control de calidad selecciona para su inspección una muestra aleatoria de 100 interruptores que forman parte de un gran envío. El ingeniero aceptará el envío si no encuentra más de 9 interruptores defectuosos en la muestra. El ingeniero no lo sabe pero el 10% de los interruptores del envío es defectuoso. ¿Cuál es la probabilidad de que no haya más de 9 interruptores defectuosos en la

> muestra de 100?, ¿Cuál es el número medio de interruptores que no pasarían la inspección en la situación planteada? (Moore, 1995)

En este caso se tiene una población con distribución binomial de parámetros $n = 100$ y $p = 0.10$. Dado que la muestra es grande, se supone que la distribución muestral de la proporción $\hat{p}$, será aproximadamente normal, como consecuencia del teorema del limite central. Los parámetros de la distribución muestral serían $\mu_{\hat{p}} = p = 0.10$ y $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.10 \times 0.90}{100}}$. Estandarizando la distribución muestral de $\hat{p}$ ($z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$) se calcula la probabilidad como el área correspondiente bajo la distribución normal.

Otro tipo de problemas es el siguiente:

> El ciclo de vida útil de la totalidad de los cinescopios de cierta marca para televisor es aproximadamente normal con media $\mu = 8900$ hrs. Se toma una muestra aleatoria de tamaño 15 de un gran lote y se determina que la desviación estándar de la muestra es $s = 500$ hrs. ¿Cuál es la probabilidad de que la vida útil promedio de la muestra sea menor a 8000 hrs.? (Kazmier, 1998).

En este ejemplo, la población es aproximadamente normal, pero la desviación estándar es desconocida, por lo es necesario estimarla con la desviación estándar muestral. De esta manera, en lugar de estandarizar los valores de la media mediante $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, se hace con $t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$, y la probabilidad se calcula mediante la distribución de probabilidad t-student .

2. **Problemas que implican un razonamiento inductivo**

En este tipo de problemas, a diferencia de los de carácter deductivo, los parámetros poblacionales son desconocidos y lo que se pretende es inferir su valor a partir de la información que proporcionan los datos de una muestra. En estos casos, el conocimiento de la distribución muestral del estadístico que se utiliza como estimador del parámetro, es fundamental para realizar las inferencias. Las dos principales técnicas de inferencia estadística que se utilizan en los libros seleccionados son la estimación por intervalos de confianza y las pruebas de hipótesis.

- *Intervalos de confianza*

Un intervalo de confianza consta de dos límites, uno inferior y otro superior, entre los cuales se espera se encuentre el parámetro estimado con una alta confianza o probabilidad que se fija de antemano. Para determinar el intervalo se debe conocer la distribución muestral del estadístico.

En este tipo de problemas, persiste la tendencia observada en todos libros, de plantear situaciones donde se supone que las distribuciones muestrales son normales o aproximadamente normales, con la misma clasificación que señalamos anteriormente. La forma que adoptan los intervalos de confianza para la media poblacional es la siguiente:

a) $\bar{x} \pm z_{\varepsilon/2} \frac{\sigma}{\sqrt{n}}$, población normal y σ conocida o población cualquiera y tamaño grande de muestra (aplicando teorema central del limite).

b) $\bar{x} \pm t_{\varepsilon/2} \frac{s}{\sqrt{n}}$, población normal y σ desconocida con tamaño de muestra pequeño.

c) $\bar{x} \pm z_{\varepsilon/2} \frac{s}{\sqrt{n}}$, población cualquiera, tamaño grande de muestra y σ desconocida.

Un ejemplo de intervalo de confianza se señala a continuación:

> ¿Cuánto tiempo gastan los usuarios en navegar por internet?. Un servidor de internet llevó a cabo una encuesta con 250 de sus clientes y encontró que la media de tiempo que los usuarios pasan en línea era de 10.5 horas por semana con una desviación estándar de 5.2 horas. Construya un intervalo de confianza de 95% para el tiempo medio que pasan en línea todos los usuarios de este servidor de internet en particular.

- *Prueba de hipótesis*

El propósito de una prueba de hipótesis es determinar si el valor supuesto de un parámetro poblacional debe aceptarse como válido, con base en la evidencia que proporcionan los datos de una muestra. De nueva cuenta, el caso prototípico en los textos analizados es considerar normal o aproximadamente normal a la distribución muestral del estadístico. Un ejemplo de prueba de hipótesis sería el siguiente:

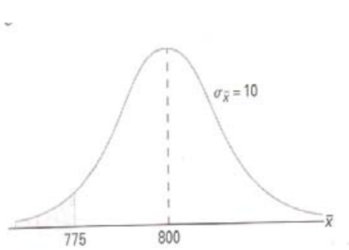
> Los ingresos promedio por semana para mujeres en puestos directivos son $5,300. ¿Los hombres en los mismos puestos tienen ingresos promedio por semana superiores a los de las mujeres?. Una muestra aleatoria de 40 hombres en puestos directivos mostró una media de $5500 y una desviación estándar de 150. Pruebe la hipótesis apropiada con un nivel de significación de 0.01.

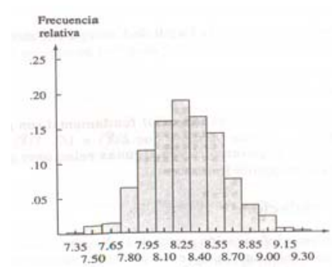
2.2 Lenguaje (elementos representacionales).

En esta sección abordaremos las distintas representaciones que se utilizan en la resolución de problemas donde intervienen las distribuciones muestrales. Hemos encontrado una gran uniformidad en este tipo de elementos en los diferentes textos, sin embargo, esto no ocurre con el enfoque y el tratamiento que se hace de ellos, ya que algunos textos son más formales que otros, y con frecuencia recurren a aspectos demostrativos.

- **Representaciones gráficas:**

El uso de este tipo de representaciones es generalizado en todos los textos, en algunos con mayor frecuencia que en otros. Las representaciones gráficas para distribuciones muestrales que se utilizan principalmente, son las distribuciones teóricas normales, sin embargo, algunos textos utilizan histogramas para representar distribuciones empíricas y ajustes de distribuciones empíricas a distribuciones teóricas, aunque en mucho menor medida y solo cuando introducen el concepto, como buscando que el lector establezca una relación entre ambas distribuciones. Otro tipo de representación gráfica utilizada con frecuencia por todos los textos, es la sombreada de áreas bajo la distribución, en la que se representa la probabilidad que se desea calcular. Los casos típicos son áreas de cola derecha, cola izquierda e intervalo central.





Ejemplos de representaciones gráficas de distribuciones muestrales

- **Representaciones numéricas**

Sin duda, la representación numérica más utilizada en todos los textos, es la tabla de probabilidad normal estandarizada ($\mu = 0$, $\sigma = 1$). En los libros que abordan distribuciones muestrales empíricamente –aunque de manera escasa-, (por ejemplo, Moore, 1995; Mendenhall, et. al., 2002), se utilizan tablas de datos y tablas de frecuencias con valores de estadísticos.

- **Representaciones simbólicas**

Hemos encontrando una gran uniformidad en las representaciones simbólicas de los conceptos más representativos, las cuales se muestran a continuación:

Concepto	**Simbología**
Población	N
Muestra	n
Media poblacional	μ
Media muestral	$\bar{x}$
Varianza poblacional	σ^2
Varianza muestral	s^2
Proporción poblacional	p
proporción muestral	$\hat{p}$
Distribución muestral de la media:	$\mu_{\bar{X}} = \mu$; $\sigma_{\bar{X}} = \sigma / \sqrt{n}$
Distribución muestral de la proporción:	$\mu_{\hat{p}} = p$; $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Expresiones de estandarización:	$z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$; $z = \dfrac{\hat{p} - p}{\sigma_{\hat{p}}}$
Expresiones para cálculo de probabilidades:	$P(\bar{X} < a)$, $P(\bar{X} > a)$, $P(a < \bar{X} < b)$ $P(\hat{p} < a)$, $P(\hat{p} > a)$, $P(a < \hat{p} < b)$

2.3 Acciones (elementos procedimentales)

En esta sección describiremos las acciones y los procedimientos que se utilizan en los libros de texto consultados para resolver las diversas situaciones-problema que tienen que ver con las distribuciones muestrales.

a) Problemas que implican un razonamiento deductivo

En todos los textos, predomina el enfoque basado en variables aleatorias y distribuciones de probabilidad para abordar los problemas de distribuciones muestrales. Hemos identificado las siguientes acciones y procedimientos para resolver este tipo de problemas:

1. Identificar si la distribución muestral del estadístico es normal o se puede considerar aproximadamente normal. Hemos señalados anteriormente tres posibles casos (para el caso de la media) para que esto suceda:

- Cuando la población es normal y σ es conocida, sin importar el tamaño de la muestra.
- Cuando $n \geq 30$, sin importar si la distribución sea normal o no, por efecto del teorema central del limite.
- Cuando $n < 30$, pero la población tiene distribución normal y σ es conocida.
- En el caso de población normal, con $n < 30$ y σ desconocida se usa la distribución t en lugar de la distribución normal.

Cuando el problema trata de proporciones, se debe comprobar previamente si la aproximación de la distribución binomial a la normal es apropiada. Esto se hace aplicando la regla ($np > 5$ y $n(1-p) > 5$).

Calcular los parámetros de la distribución muestral, a saber, $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$ para el caso de la media, $\mu_{\hat{p}}$ y $\sigma_{\hat{p}}$ para el caso de la proporción.

2. Estandarizar la distribución muestral del estadístico $z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ **;** $z = \dfrac{\hat{p} - p}{\sigma_{\hat{p}}}$
3. Construir una gráfica de ella indicando sus medidas descriptivas y sombreando el área cuya probabilidad se desea calcular.
4. Utilizar las tablas de probabilidad de la distribución normal estándar para calcular la probabilidad, o dada la probabilidad, encontrar los valores de z para determinar un intervalo de confianza.

Las acciones señaladas anteriormente son comunes en todos los libros consultados, sin embargo, algunos textos, especialmente el de Moore (1995), dan un espacio al proceso de obtener una distribución muestral en su forma empírica, generalmente apoyándose en el uso de tablas de números aleatorios. El enfoque planteado por Moore nos interesa pues aunque es el único texto que aborda de forma más sistemática el aspecto empírico de las distribuciones muestrales, de cualquier forma es una aportación al significado del concepto en estudio. Las acciones que hemos identificado para resolver un problema de distribuciones muestrales por esta vía son:

1. Tomar un gran número de muestras de una misma población.
2. Calcular el estadístico correspondiente (media o proporción) en cada muestra.
3. Construir una tabla de frecuencias con los valores obtenidos.
4. Dibujar una representación gráfica (usualmente histograma) con los valores del estadístico.
5. Calcular las medidas descriptivas de la distribución.

2.4 Conceptos y propiedades (elementos conceptuales)

a) Conceptos

Las distribuciones muestrales tienen la peculiaridad de ser un concepto en el que se involucran diversos a su vez diversos conceptos, tanto del ámbito de la probabilidad como de la estadística. Los conceptos que abordaremos en este apartado son: distribuciones muestrales, estadístico y muestra aleatoria, ya que están estrechamente relacionados entre sí con la temática que nos ocupa. Las diferentes definiciones que encontramos sobre estos conceptos están en función del enfoque (teórico o empírico) adoptado por el autor en el tratamiento del tema.

En algunos libros hemos observado que se plantean situaciones donde se construye una distribución muestral empíricamente a través de un muestreo repetido, pero sin aplicaciones posteriores, solo a nivel introductorio para explicar el concepto. Un ejemplo de ello se puede ver en los libros de Moore (1995) y Mendenhall, Beaver y Beaver (2002).

- Distribuciones muestrales:

Dos tipos de definiciones sobre distribuciones muestrales hemos encontrado en la revisión de los textos. Cada una de ellas tiene que ver con el enfoque (teórico o empírico) adoptado para el desarrollo y aplicación del concepto.

1. *La distribución muestral como la distribución teórica de probabilidad de un estadístico.*

Los textos que utilizan este tipo de definición tienen en común que enfatizan en el uso de variables aleatorias como base para desarrollar el concepto de distribución muestral. Este tipo de definición se da en la mayoría de los textos revisados.

2. *La distribución muestral como la distribución empírica de los valores del estadístico en todas las posibles muestras de igual tamaño de la misma población.*

Se utiliza el concepto de muestreo repetido como un medio para construir la distribución muestral de un estadístico. En el texto de Moore es donde encontramos mayor énfasis en este tipo de definición. Sin embargo, algunos textos que manejan rasgos de los dos tipos de definiciones anteriores. Por ejemplo en Mendenhall, et. al., (2002) se definen de la siguiente manera:

> "la distribución muestral de un estadístico es una distribución de probabilidad para los valores posibles del estadístico que resulta cuando se extraen repetidamente de la población las muestras aleatorias de tamaño n" (p. 251).

En Mendenhall, et. al., (1994) se señala:

> "Desde el punto de vista práctico, la distribución muestral representa un modelo teórico para el histograma de frecuencias relativas de los valores posibles del estadístico que se observarían en un muestro repetitivo" (p. 284).

En ambos casos, encontramos elementos en la definición como distribución empírica que hacen uso del concepto de muestreo, sin embargo, el enfoque que finalmente se utiliza en las diversas situaciones es a través de distribuciones teóricas de probabilidad. Mientras que otros textos que usan el primer tipo de definición, introducen el concepto de distribución muestral a través de un ejemplo en contexto de muestreo repetido, es decir, hacen uso del segundo tipo de definición (por ejemplo, Johnson, 1997, Mendenhall, et. al., 2002).

A manera de conclusión podemos decir, que la definición de distribución muestral que más aparece en los textos, es la definición como distribución de probabilidad del estadístico. En cuanto a la definición de distribución muestral como distribución empírica del estadístico en un muestreo repetido, solo aparece expresada en pocos textos y sin su correspondiente aplicación, salvo algún ejemplo aislado par mostrarla. Vemos en este tipo de definición un enfoque más estadístico que probabilístico, que lo hace apropiado para implementarlo por medio de un enfoque empírico basado en simulación.

- *Estadístico*

Otro concepto importante que es inherente al de distribuciones muestrales es el concepto de estadístico. Igualmente encontramos que en su definición se maneja diferente terminología según el enfoque adoptado por el libro de texto. Por ejemplo en Mendenhall, et. al (1994) que es el libro más formal de los que fueron analizados, un estadístico se define así:

> "un estadístico es una función de las variables aleatorias que se pueden observar en una muestra y de las constantes conocidas" (p. 284).

En Mendenhall, et, al., 2002 un estadístico se define como:

> "Las medidas descriptivas numéricas que se calculan de la muestra se llaman estadísticos" (p. 250).

Por su parte Moore (1995) lo define de la siguiente manera:

> "Un estadístico es un número que se puede calcular a partir de los datos de una muestra sin utilizar ningún parámetro desconocido" (p. 235).

En la mayoría de los textos, un estadístico es considerado como variable aleatoria, de ahí que este posea una distribución de probabilidad, llamada distribución muestral.

- *Muestra aleatoria*

Los estadísticos se utilizan como estimadores de parámetros poblacionales, sin embargo, para que la confiabilidad de dicha estimación pueda ser medible, es necesario que la muestra sea seleccionada en forma aleatoria. En los libros seleccionados, la definición de muestra aleatoria varía de un texto a otro. Por ejemplo, Hines y Montgomery (1993) la definen así:

> "Una muestra aleatoria es el conjunto de n observaciones $X_1, X_2, \ldots X_n$ tomadas en la variable aleatoria X, y con resultados numéricos $x_1, x_2, \ldots, x_n$" (p. 263).

Walpole et. al., (1999) definen a la muestra de la siguiente manera:

> "Sean $X_1, X_2, \ldots X_n$ variables aleatorias independientes, cada una con la misma distribución de probabilidad $f(x)$. Definimos entonces a $X_1, X_2, \ldots X_n$ como una muestra aleatoria de tamaño n de la población" (p. 201).

Mendenhall, et. al., (2002):

> Si se selecciona una muestra de n elementos de una población de N elementos, usando un plan de muestreo en el que cada una de las posibles muestras tiene la misma probabilidad de ser seleccionada, se dice que el muestreo es aleatorio y la muestra seleccionada resultante es una muestra aleatoria simple" (p. 246).

b) Propiedades (enunciados, proposiciones).

Un concepto de suma importancia que influye sobre la forma de las distribuciones muestrales es el teorema del límite central. Mendenhall, et. al., (1994) lo enuncia en forma de teorema:

> "Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes y distribuidas idénticamente con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Definimos $U_n = \sqrt{n}\left(\frac{\bar{Y} - \mu}{\sigma}\right)$. Entonces la función de distribución U_n converge a una función de distribución normal estándar cuando $n \to \infty$." (p, 299).

Mendenhall et. al., (2002):

> Si de una población no normal con media finita μ y desviación estándar σ se extraen muestras aleatorias de n observaciones, entonces, cuando n es grande, la distribución muestral de las medias muestrales $\bar{x}$ está aproximadamente distribuida de manera normal, con media $\mu_{\bar{X}} = \mu$ y desviación estándar $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. La aproximación se vuelve mas precisa a medida que aumenta *n*.

Walpole, et. al., (1999) :

Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n tomada de una población con media μ y varianza σ^2, entonces la forma límite de la distribución de $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ conforme $n \to \infty$, es la distribución normal estándar $n(z;0,1)$.

Otras propiedades son:

1. Si la población tiene distribución normal, la distribución muestral de $\bar{X}$ se distribuirá de manera exactamente normal, sin importar el tamaño de la muestra *n*.
2. El centro de la distribución muestral de la media es igual a la media de la población

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

3. La desviación estándar (error estándar) de la distribución muestral de la media es

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

4. La variabilidad de la distribución muestral de la media disminuye conforma aumenta el tamaño de la muestra.
5. La media de la distribución muestral de la proporción es igual a la proporción poblacional (p).

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

6. La desviación estándar (error estándar) de la distribución muestral de la proporción es

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \quad \text{donde } q = 1 - p$$

7. Cuando el tamaño de la muestra n es grande, la distribución muestral de $\hat{p}$ se puede aproximar mediante la distribución normal. La aproximación será adecuada si $np > 5$ y $n(1-p) > 5$.

2.5 Argumentaciones (elementos validativos)

Todas las acciones, las propiedades y los conceptos se relacionan entre sí mediante argumentos o razonamientos que se usan para optar por determinado procedimiento, fundamentar el uso de cierto procedimiento o estrategia, comprobar soluciones de los problemas o explicar a otros la solución obtenida. Por ejemplo, un tipo de argumentación utilizada ampliamente es considerar que la distribución muestral es aproximadamente normal cuando el tamaño de la muestra es mayor a 30, sea cual sea la distribución de la población, porque el teorema del limite central lo establece. Algunas de las argumentaciones que más se emplean en los libros consultados se muestran a continuación:

Validación a través de una representación gráfica:

Propiedades como la variabilidad de la distribución muestral, que depende del tamaño de la muestra y el centro de la distribución muestral que es igual al parámetro poblacional se pueden argumentar mediante el uso de representaciones gráficas. En muchos textos analizados observamos este tipo de validaciones. Otro tipo sería cuando se ajusta una distribución normal empírica a una distribución muestral teórica, con la finalidad de ver si los resultados obtenidas empíricamente mediante simulación se aproximan a los resultados teóricos. Encontramos que pocos textos explotan este tipo de argumentaciones. El caso más frecuente es el de usar gráficas con área sombreada como un medio para expresar la probabilidad que se desea calcular.

Validaciones deductivas

Este tipo de validaciones se presenta cuando se enuncia una propiedad por medio de una demostración. Encontramos algunas validaciones de este tipo en los libros consultados, sobre todo en los más formales como Mendenhall, et. al., (1994), Johnson (1997) y

Walpole et. al. (1999). Un ejemplo de ello se muestra en el Anexo 1, en donde se desarrollan las distribuciones muestral de la media y la proporción.

Validaciones informales

Frecuentemente este tipo de validación se da mediante el uso de representaciones gráficas y la comprobación de casos particulares conjuntamente para mostrar de manera intuitiva una propiedad. Por ejemplo, si aumentamos el tamaño de muestra varias veces y observamos que la media de la distribución muestral no es sensible a dicho aumento entonces decimos que el tamaño de la muestra no afecta su valor.

3. Resumen del significado institucional de referencia de las distribuciones muestrales.

Elementos de significado	**Elementos identificados**
Elementos representacionales	*a) Representaciones gráficas:* • Distribuciones teóricas (curvas suaves). • Distribuciones empíricas (histogramas). • Ajustes de distribuciones teóricas a empíricas. • Distribuciones teóricas con áreas sombreadas (cola derecha, cola izquierda e intervalo central). *b) Representaciones numéricas:* • Tabla de probabilidad normal estandarizada (enfoque teórico). • Tablas de datos y tablas de frecuencias de estadísticos (enfoque empírico). *c) Representaciones simbólicas:* • Población (N) • Muestra (n) • Media poblacional (μ) • Media muestral ($\bar{x}$) • Varianza poblacional (σ^2) • Varianza muestral (s^2) • Proporción poblacional (p) • Proporción muestral ($\hat{p}$) • Distribución muestral de la media: Media $\mu_{\bar{X}} = \mu$

	Desviación estándar (error estándar) $\sigma_{\bar{X}} = \sigma / \sqrt{n}$ • Distribución muestral de la proporción: Media $\mu_{\hat{p}} = p$ Desviación estándar $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$ • Expresiones de estandarización: $z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $z = \dfrac{\hat{p} - p}{\sigma_{\hat{p}}}$ • Expresiones para cálculo de probabilidades: $P(\bar{X} < a)$ $P(\bar{X} > a)$ $P(a < \bar{X} < b)$ $P(\hat{p} < a)$ $P(\hat{p} > a)$ $P(a < \hat{p} < b)$
Elementos fenomenológicos)	*a) Situaciones con distribuciones muestrales normales* • Porque la distribución poblacional es normal • Porque el tamaño de muestra es grande (mayor a 30) y se puede aplicar el teorema del limite central. *b) Por la forma de razonamiento para resolver las diversas situaciones* • Problemas que implican un razonamiento deductivo. • Problemas que implican un razonamiento inductivo. Intervalos de confianza. Prueba de hipótesis.
Elementos conceptuales	Conceptos: 1. La distribución muestral como la distribución teórica de probabilidad de un estadístico (enfoque teórico). 2. La distribución muestral como la distribución empírica de los valores del estadístico en todas las posibles muestras de igual tamaño de la misma población (enfoque empírico). 3. Un estadístico como función de variables aleatorias que se pueden observar en una muestra. .

	4. Un estadístico como medida descriptiva que se calculan de la muestra.
	Propiedades: *Sobre la forma y parámetros de la distribución muestral (teorema del limite central).* • Si de una población no normal con media finita μ y desviación estándar σ se extraen muestras aleatorias de *n* observaciones, entonces, cuando *n* es grande, la distribución muestral de las medias muestrales $\bar{x}$ está aproximadamente distribuida de manera normal, con media $\mu_{\bar{X}} = \mu$ y desviación estándar $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. La aproximación se vuelve mas precisa a medida que aumenta *n*. • Si la población tiene distribución normal, la distribución muestral de $\bar{X}$ se distribuirá de manera exactamente normal, sin importar el tamaño de la muestra *n*. • La variabilidad de la distribución muestral de la media disminuye conforma aumenta el tamaño de la muestra. • La media de la distribución muestral de la proporción es igual a la proporción poblacional (p). $$E(\hat{p}) = \mu_{\hat{p}} = p$$ • La desviación estándar (error estándar) de la distribución muestral de la proporción es $$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$ • Cuando el tamaño de la muestra *n* es grande, la distribución muestral de $\hat{p}$ se puede aproximar mediante la distribución normal. La aproximación será adecuada si $np > 5$ y $n(1-p) > 5$.
Elementos procedimentales	*Problemas que implican un razonamiento deductivo* *a) Siguiendo un enfoque teórico:* Siguiendo un enfoque basado en variables aleatorias y distribuciones de probabilidad como el que se emplea en la mayoría de los textos, identificamos las siguientes acciones para resolver este tipo de problemas:

<table>
<tr><td></td><td>

Identificar si la distribución muestral del estadístico es normal o se puede considerar aproximadamente normal.
Cuando el estadístico es la proporción se debe comprobar previamente si la aproximación normal a la distribución binomial es apropiada, es decir, aplicar la regla ($np > 5$ y $n(1-p) > 5$).
Calcular los parámetros de la distribución muestral, a saber, $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$ para el caso de la media, $\mu_{\hat{p}}$ y $\sigma_{\hat{p}}$ para el caso de la proporción.
Estandarizar la distribución muestral del estadístico
Construir una gráfica de ella indicando sus medidas descriptivas y sombreando el área cuya probabilidad se desea calcular.
Utilizar las tablas de probabilidad de la distribución normal estándar para calcular la probabilidad, o dada la probabilidad, encontrar los valores de z para determinar un intervalo de confianza.

b) Siguiendo un enfoque empírico:

Tomar un gran número de muestras de una misma población.
Calcular el estadístico correspondiente (media o proporción) en cada muestra.
Construir una tabla de frecuencias con los valores obtenidos.
Dibujar una representación gráfica (usualmente histograma) con los valores del estadístico.
Calcular las medidas descriptivas de la distribución.
Determinar la proporción de casos que caen en un intervalo dado.

</td></tr>
<tr><td>Elementos validativos</td><td>

Considerar que la distribución muestral de una proporción es una buena aproximación a la distribución normal a partir de aplicar el criterio $np > 5$ y $n(1-p) > 5$ (teorema central del límite).

Validaciones geométricas

Propiedades como la variabilidad de la distribución muestral, que depende del tamaño de la muestra, y el centro de la distribución muestral que es igual al parámetro poblacional se pueden argumentar mediante el uso de representaciones gráficas.
Ajustar una distribución normal empírica a una distribución muestral

</td></tr>
</table>

	teórica, con la finalidad de ver si los resultados obtenidas empíricamente mediante simulación se aproximan a los resultados teóricos. • Utilizar gráficas con área sombreada como un medio para expresar la probabilidad que se desea calcular. *Validaciones deductivas* • Este tipo de validaciones se presenta cuando se enuncia una propiedad por medio de una demostración. *Validaciones informales* • Este tipo de validación se da mediante el uso de representaciones gráficas y la comprobación de casos particulares conjuntamente para mostrar de manera intuitiva una propiedad. Por ejemplo, si aumentamos el tamaño de muestra varias veces y observamos que la media de la distribución muestral no es sensible a dicho aumento entonces decimos que el tamaño de la muestra no afecta su valor.

4. Significado local de las distribuciones muestrales

Dada la complejidad y extensión del concepto de distribuciones muestrales, identificada en el significado institucional de referencia, y de acuerdo con los principios del modelo teórico elegido, hemos delimitado el campo de problemas a un subconjunto representativo acorde al alcance y enfoque de nuestra la investigación. Este subconjunto del campo de problemas y los elementos de significado asociados a él, constituye el *significado institucional local* que hemos asumido para las distribuciones muestrales en el estudio definitivo. Este ha sido el referente más cercano que se tuvo en cuenta en el diseño de las actividades de enseñanza.

El enfoque de este estudio está orientado hacia el uso de simulación computacional, por lo que al momento de seleccionar los elementos de significado, nuestro interés no se centró en aspectos formales del tema, sino más bien en aquellos aspectos que pueden ser abordados desde un enfoque empírico apoyado en simulación. Además, por ser la simulación computacional una herramienta novedosa que aún no se incorpora en muchos libros de texto, proporciona elementos de significado propios que habrán de tomarse en cuenta en el significado institucional local.

4.1 Situaciones-problema (elementos fenomenológicos)

El subconjunto de situaciones-problema al que hemos acotado nuestra investigación, corresponde a la categoría de problemas que requieren de un enfoque deductivo para su solución, esto es, problemas en los que la distribución de la población y sus parámetros son conocidos, y lo que se plantea es calcular probabilidades de ciertos resultados muestrales. Además, nos restringiremos sólo a las distribuciones muestrales de la media y la proporción, por ser dos estadísticos que cubren una parte considerable del campo de problemas de donde emergen las distribuciones muestrales. Las distribuciones poblacionales que se abordaron en las actividades son normal, binomial, uniforme discreta e irregulares. Por limitaciones de tiempo, el campo de problemas que requiere de un razonamiento inductivo no será abordado en nuestro estudio.

4.2 Lenguaje (términos, expresiones, notaciones, gráficos).

- *Representaciones gráficas:* Distribuciones teóricas de una población (por ejemplo, normal, binomial, uniforme), distribuciones muestrales teóricas y empíricas (histogramas), representaciones de áreas bajo una distribución normal.

- *Representaciones numéricas:* Tablas de valores poblacionales y muestrales, tablas con frecuencias del estadístico, resúmenes estadísticos.

- *Representaciones simbólicas (notaciones):* Representaciones de parámetros (μ, p), estadísticos ($\bar{x}$, $\hat{p}$), tamaño poblacional (N), tamaño muestral (n), media de distribución muestral ($\mu_{\bar{X}} = \mu$, $\mu_{\hat{p}} = p$), desviación estándar de distribución muestral ($\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$), distribución muestral estandarizada ($z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, $z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$), intervalo de confianza en términos de desviaciones estándar ($\bar{x} \pm k\sigma_{\bar{x}}$).

- *Representaciones disponibles en Fathom*

Distribuciones teóricas de probabilidad (por ejemplo, *normalDensity(x, m, s)*), tabas de casos, muestras de casos, tablas resumen, histogramas, diagramas de puntos, formulas para cálculo de estadísticos (por ejemplo, *mean(x), stdDev(x),* cálculo de probabilidades, por ejemplo: *proportion(media>3).*

4.3 Acciones (operaciones, algoritmos, procedimientos).

El enfoque seguido por Moore (1995) se adapta en buena medida al enfoque de nuestro trabajo; dicho autor señala diversas acciones a realizar cuando se implementa una enfoque de simulación en la enseñanza de las distribuciones muestrales. Señalamos a continuación las acciones que hemos identificado como necesarias para el estudio de las distribuciones muestrales de acuerdo al enfoque de simulación:

1. Crear o generar una distribución poblacional.
2. Tomar muestras de la población haciendo uso de la opción *Sample cases*.
3. Definir el estadístico que se desea calcular en cada muestra mediante la opción *Measures.*
4. Seleccionar un número grande de muestras y calcular el estadístico en cada una de ellas mediante la opción *Collect measures*.
5. Construir una tabla con los valores de las medias y un histograma o diagrama de puntos.
6. Calcular medidas descriptivas de la distribución muestral como la media y la desviación estándar, haciendo uso de la opción *Sumary table.*
7. Superponer la distribución muestral teórica a la distribución empírica obtenida como un medio de validar los resultados.
8. Calcular la proporción de casos que caen dentro de cierto intervalo como una estimación de la probabilidad.

En la resolución de problemas a través de lápiz y papel se requiere estandarizar la distribución muestral, utilizar tablas para calculo de probabilidades y valores estandarizados.

4.4 Conceptos y propiedades (elementos conceptuales)

Las distribuciones muestrales son un concepto que para su comprensión y desarrollo involucra a su vez a diversos conceptos entre los cuales podemos mencionar población, muestra aleatoria, tamaño de muestra, variable aleatoria, estadístico, parámetro, distribución de probabilidad, distribución normal, centro, dispersión, distribución estandarizada, todos ellos aparecen cuando se trabaja en un ambiente computacional.

En algunos textos analizados estos conceptos se abordan, como lo señalamos con anterioridad, desde un enfoque que hace énfasis en el concepto de variable aleatoria, en otros sin embargo, el enfoque es menos formal y a una distribución muestral se le ve más como un conjunto de estadísticos producto de un muestreo repetitivo que contempla todas las muestras de un tamaño dado de una población, que como una distribución de probabilidad. Este último enfoque es el que tomamos para la selección de los elementos de significado.

En cuanto a las propiedades más importantes de las distribuciones muestrales, identificamos las siguientes:

- El centro de la distribución muestral de la media es igual a la media de la población.
- El centro de la distribución muestral de la proporción es igual a la proporción poblacional.
- La variabilidad de la distribución muestral (de la media y de la proporción) disminuye conforma aumenta el tamaño de la muestra.
- Si la distribución de la población no es normal, la distribución muestral (de la media y la proporción) será aproximadamente normal para muestras grandes (por lo general mayor a 30).

4.5 Argumentaciones (validaciones y explicaciones)

Las argumentaciones se usan para validar y explicar las propiedades de un concepto. En este caso, el elemento argumentativo más empleado será la misma simulación. Es decir, los estudiantes usarán los resultados de la simulación para validar experimentalmente propiedades y conceptos que tienen lugar en las distribuciones muestrales. Este es un aspecto que es poco abordado en los libros de texto, salvo algunas excepciones ya mencionadas donde se señala que la simulación es una herramienta para estudiar conceptos donde interviene el azar y validar resultados teóricos. En las actividades de enseñanza emplearíamos elementos argumentativos como el análisis y comprobación de propiedades de las distribuciones muestrales, bondad de ajuste de una distribución empírica a la distribución teórica, generalización de propiedades. En fin, no se pretende que las argumentaciones hagan énfasis en aspecto formales como demostraciones o teoremas.

Anexo 2

LAS ACTIVIDADES DE ENSEÑANZA

En este anexo se detallan las hojas de trabajo de cada una de las actividades de enseñanza de las que consistió el estudio definitivo.

ACTIVIDAD 1

Considera la población que se compone de todos los resultados que obtendríamos si lanzáramos una moneda al aire un número infinito de veces. Si la moneda es justa se espera que la proporción de águilas de esta población sea igual a 0.5

Primera parte

a) Toma 10 muestras de 10 lanzamientos cada una y obtén la proporción de águilas en cada muestra. Construye una grafica con las proporciones, calcula el centro y la dispersión de la distribución. Comenta tus observaciones.

b) Acumula tus resultados con los tus compañeros y construye una nueva grafica con los valores de las proporciones. Centra tu atención en el centro, la dispersión y la forma de la distribución. Comenta tus observaciones.

Segunda parte

c) Utiliza Fathom para simular el proceso de muestreo, considerando muestras de tamaño 10, 20 y 30. Registra tus resultados en el cuadro que se te proporciona.

d) Calcula el centro y la dispersión de las distribuciones muestrales teóricas y compara los resultados con los de las distribuciones muestrales obtenidas por simulación.

e) Coloca sobre la misma gráfica en Fathom, la distribución muestral teórica y la distribución muestral empírica. Comenta tus observaciones.

f) Guarda todo tu trabajo en un archivo cuyo nombre sea *Actividad1_tunombre.ftm*

Proporción de águilas en la población (p): ___________

Desviación estándar de la población (σ): ___________

Tamaño de muestra	Forma de la distribución muestral	Centro (Media)	Dispersión (Desviación estándar)
n=10			
n=20			
n=30			
n=1			

Observaciones:

ACTIVIDAD 2

En las bolsas de chocolates de la marca M&M´s en su presentación de chocolate con leche (bolsa café) aparecen chocolates de diversos colores. Según la compañía, el 30% de ellos son de color café.

Primera parte

a) Selecciona dos bolsas de chocolates y determina la proporción de chocolates color café en cada una. Acumula tus resultados con los de tus compañeros y construye una gráfica. Determina el centro y la dispersión de la distribución. Comenta tus observaciones centrado tu atención en la forma, centro y dispersión de la distribución muestral.

b) Determina la proporción de chocolates de color amarillo en cada bolsa y reúne tus resultados con los de tus compañeros. ¿Cuál sería tu estimación de la proporción de amarillos en la población de M&M´s?

Segunda parte

c) Simula el proceso anterior mediante Fathom, tomando muestras de tamaño 5, 15 y 30. Registra tus resultados en el cuadro que se te proporciona.

d) Compara los resultados teóricos con los que obtuviste mediante simulación.

Proporción de chocolates café en la población (p): __________

Tamaño de muestra	Centro	Dispersión
n=5		
n=15		
n=30		
n=1		

Compara las distribuciones muestrales en forma, centro y dispersión y anota tus conclusiones:

ACTIVIDAD 3

Considera como una población a los números que aparecen en las diferentes caras de un dado, es decir, la población consta de los números 1, 2, 3, 4, 5, 6.

Primera parte:

a) Construye una grafica con la distribución de la población, calcula su centro y su dispersión.

b) Selecciona 10 muestras de tamaño 5 y obtén la media de puntos en cada una de ellas. Construye una gráfica con dichos resultados.

c) Acumula tus resultados con los de tus compañeros y construye una nueva gráfica con los valores de las medias. Centra tu atención en el centro, la dispersión y la forma de la distribución muestral. Comenta tus conclusiones.

Segunda parte

d) Simula el proceso de muestreo anterior utilizando Fathom para muestras de tamaño10, 20 y 30. Calcula el centro y la dispersión para cada tamaño de muestra y regístrala en el cuadro que se te proporciona.

e) Si se toma una muestra de tamaño 30 de la población. ¿Cuál será la probabilidad de que la media sea menor o igual a 3 puntos?. Calcula el resultado de forma teórica y mediante simulación.

Media de la población (μ): ____________

Desviación estándar de la población (σ): ____________

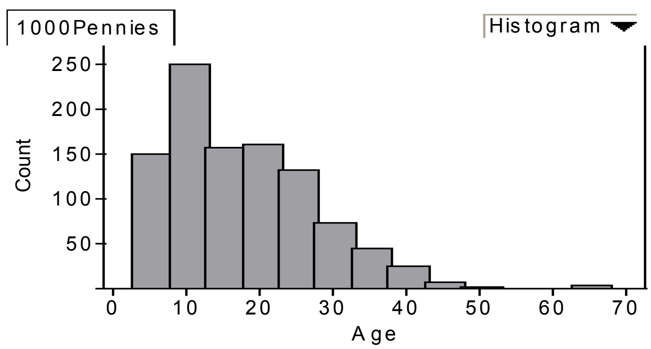
Tamaño de muestra	Forma de la distribución muestral	Media (Media)	Dispersión (Desviación estándar)
n=5			
n=20			
n=30			

n=1			

Observaciones:

ACTIVIDAD 4

Se han recolectado 1000 monedas y se observó su fecha de fabricación y su edad al presente año. Los datos han almacenado en un archivo de Fathom denominado *Monedas.ftm*.



Considera las 1000 monedas como una población.

a) Determina el centro y la dispersión de la población.
b) Utiliza Fathom para construir distribuciones muestrales de la población anterior para muestras de tamaño 1, 15 y 30. Registra los resultados en el cuadro que se te proporciona a continuación:

Media de la población (μ): ____________
Desviación estándar de la población (σ): ____________

Distribuciones muestrales (por simulación)

Tamaño de muestra	Centro	Dispersión
n=1		
n=15		
n=30		

c) Compara entre sí el centro y la dispersión de las cuatro distribuciones muestrales y saca tus conclusiones.

d) Compara el centro y la dispersión de las cuatro distribuciones muestrales con el centro y la dispersión de la población. Saca tus conclusiones.

e) Compara la forma de las distribuciones muestrales con la forma de la población. ¿Qué pasa cuando el tamaño de la muestra se incrementa?

f) Calcula el centro y la dispersión de las cuatro distribuciones muestrales utilizando las fórmulas que conoces y concentra los resultados en el cuadro siguiente:

Distribuciones muestrales (teóricas)

Tamaño de muestra	Centro	Dispersión
n=1		
n=15		
n=30		

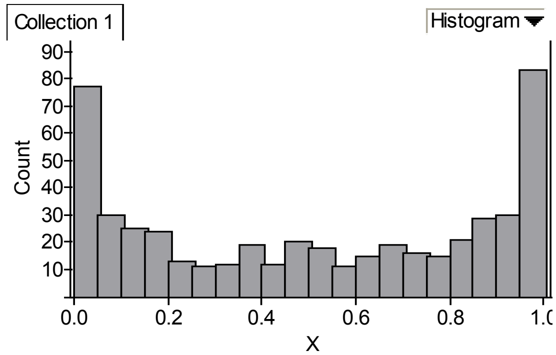
g) Compara los resultados que obtuviste mediante simulación con los que dan las fórmulas. Saca tus conclusiones.

h) Intenta enunciar en la forma más precisa posible, las propiedades (forma, centro y dispersión) de las distribuciones muestrales que has observado.

i) Ahora, ¿Cuál será la probabilidad que si se toma una muestra de tamaño 30, la edad promedio de las monedas sea mayor de 10 años? . Calcula dicha probabilidad tanto en forma teórica como con Fathom.

ACTIVIDAD 5

Consideremos los siguientes 500 datos como una población cuya forma se muestra en la siguiente gráfica:



Los datos se han almacenado en un archivo de Fathom cuyo nombre es *Beta.ftm*

a) Determina el centro y la dispersión de la población.

b) Utiliza Fathom para construir distribuciones muestrales de la población anterior para muestras de tamaño 1, 10, 20 y 30. Registra los resultados en el cuadro que se te proporciona a continuación:

Media de la población (μ): ____________
Desviación estándar de la población (σ): ____________

Distribuciones muestrales (por simulación)

Tamaño de muestra	Centro	Dispersión
n=1		
n=10		
n=20		
n=30		

c) Compara entre el centro y la dispersión de las cuatro distribuciones muestrales entre sí, y con el centro y la dispersión de la población. Saca tus conclusiones.

d) Compara la forma de las distribuciones muestrales con la forma de la población.

e) Calcula el centro y la dispersión de las cuatro distribuciones muestrales utilizando las fórmulas que conoces y concentra los resultados en el cuadro siguiente:

Distribuciones muestrales (teóricas)

Tamaño de muestra	Centro	Dispersión
n=1		
n=10		
n=20		
n=30		

f) Compara los resultados que obtuviste mediante simulación con los que dan las fórmulas. Saca tus conclusiones.

g) Intenta enunciar en la forma más precisa posible, las propiedades (forma, centro y dispersión) de las distribuciones muestrales que has observado.

h) Ahora, ¿Cuál será la probabilidad que si se selecciona una muestra aleatoria de tamaño 30, la media sea menor a 0.3? Calcula dicha probabilidad tanto en forma teórica como por simulación. Comenta tus conclusiones.

i) ¿Cuál será la probabilidad que si se selecciona una muestra aleatoria de tamaño 20, la media sea menor a 0.3? Calcula dicha probabilidad tanto en forma teórica como por simulación. Comenta tus conclusiones.

ACTIVIDAD 6

En una fábrica de cojinetes para automóvil se ha desajustado una máquina y el 30% de su producción está saliendo defectuosa.

a) ¿Cómo podrías simular físicamente el problema anterior utilizando canicas?

    1. Determina la cantidad y colores de las canicas que se requieren.

    2. Discute si es necesario el reemplazo.

b) Selecciona algunas 10 muestras de tamaño 10 y calcula la proporción de defectuosos en cada muestra. Determina entonces la media de las proporciones de defectuosos.

c) Construye una gráfica con las proporciones encontradas.

d) Si se toma una muestra de tamaño 80. ¿Cuántos cojinetes defectuosos esperarías en la muestra?
e) ¿Cuál es la probabilidad de que en una muestra de tamaño 80 existan 30 ó mas defectuosos?
f) ¿Cuál es la probabilidad de que aparezcan 20 o menos defectuosos en la muestra de tamaño 80?
g) Utiliza Fathom para construir la distribución muestral de la proporción de defectuosos.
h) Contesta los incisos del problema con los resultados de la simulación.
i) ¿Cómo podrías verificar que los resultados de la simulación están correctos?.
j) Realiza la verificación

ACTIVIDAD 7

Una máquina de refrescos se ajusta para que la cantidad de bebida que sirve promedie 240 ml con una desviación estándar de 5 ml. La máquina se verifica periódicamente tomando una muestra de 40 bebidas y se calcula el contenido promedio (media).

- ¿Cuál es la probabilidad de que en una muestra seleccionada al azar se tenga un contenido promedio menor que 239ml.?
- ¿Si la máquina se ajusta cuando el contenido promedio de la muestra sea menor a 239 ml y mayor a 241 ml. ¿Qué proporción será ajustada la máquina?

a) Construye con Fathom la distribución muestral de la media para muestras de tamaño 40 y contesta los incisos anteriores.
b) Resuelve teóricamente los incisos y compáralos con los resultados de la simulación.
c) Si en lugar de tomar una muestra de 40 bebidas, se toma una muestra de 20 bebidas. ¿Qué pasa con la proporción de veces que la máquina es ajustada?

ACTIVIDAD 8

Una tienda departamental ha decidido premiar la preferencia de sus clientes; para ello, dispone de una tómbola con 500 esferas, todas con premio y en igual proporción. Las esferas están marcadas con premios que van desde 1000 hasta 15000 pesos. El cliente selecciona de manera aleatoria 3 esferas y su premio consiste en la media de las cantidades seleccionadas. Después se regresan las esferas a la tómbola para que un nuevo cliente haga su selección.

Utiliza Fathom para contestar lo siguiente:

1. Construye la distribución muestral de la media.
2. ¿Cuál crees que sea la ganancia media esperada de los clientes?
3. ¿Cuál es la probabilidad de que el premio de una persona sea menor a 10000 pesos?
4. ¿Cuál es la probabilidad de que el premio de una persona se encuentre entre 3000 y 10000 pesos?
5. Interpreta los resultados. Explica que significan.
6. Comprueba teóricamente los resultados usando Fathom o papel y lápiz.

ACTIVIDAD 9

Considera como una población a los posibles resultados que pueden aparecer al lanzar un dado.

Primera parte:

a) Construye una grafica de la población, calcula su media y dispersión.
b) Utiliza Fathom para construir distribuciones muestrales de la media para muestras de tamaño 5 y 30. Registra los resultados en el cuadro que se te proporciona a continuación:

Media de la población (μ): ____________

Desviación estándar de la población (σ): ____________

Tabla 1: Valores de las distribuciones muestrales (por simulación)

Tamaño de muestra	Media	Desviación estándar
n=5		
n=30		

c) Analiza los resultados de la tabla 1 y comenta tus conclusiones (toma en cuenta los valores de la población y el tamaño de la muestra).

d) Ahora utiliza Fathom para simular las dos distribuciones muestrales sobre una misma gráfica y compara ambas distribuciones. Comenta tus conclusiones.

e) Calcula la media y la dispersión de las dos distribuciones muestrales utilizando las fórmulas que conoces y concentra los resultados en el cuadro siguiente:

Tabla 2: Distribuciones muestrales (teóricas)

Tamaño de muestra	Media	Desviación estándar
n=5		
n=30		

f) Compara los resultados que obtuviste mediante simulación con los que dan las fórmulas. Comenta.

Segunda parte:

g) Si se toma una muestra de tamaño 30 de la población, ¿cuál será la probabilidad de que la media de puntos sea mayor a 3?

h) Si se toma una muestra de tamaño 5 de la población. ¿Cuál será la probabilidad de que la media de puntos sea mayor a 3?

i) ¿A que atribuyes la diferencia en los resultados anteriores?. Comenta.

Anexo 3

Distribuciones muestrales de algunos estadísticos comunes

1. Distribución muestral del la media

La distribución de la media ($\bar{X}$) es la distribución de los valores de $\bar{X}$ de todas las muestras posibles del mismo tamaño extraídas de la misma población. Para conocer la distribución muestral exacta ó teórica se requiere hacer uso de teoría matemática.

Supongamos que $X_1, X_2, ..., X_n$ *es una muestra aleatoria de tamaño* n*, extraída de una población con distribución de probabilidad normal con media* μ *y varianza* σ^2*. Entonces las* X_i *son variables aleatorias independientes normalmente distribuidas con* $E(X_i) = \mu$ *y* $V(X_i) = \sigma^2$*, para* $i = 1,2,3,....,n$*. Además,* $\bar{X}$ *es una combinación lineal de* $X_1, X_2, ..., X_n$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1) + \frac{1}{n}(X_2) + ... + \frac{1}{n}(X_n),$$

Aplicando el teorema

$$E(X) = \sum_{i=1}^{n} a_i\mu_i = a_1\mu_1 + ... + a_n\mu_n \qquad y \qquad V(X) = \sum_{i=1}^{n} a_i^2\sigma_i^2 = a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2 .$$

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1) + ...+ \frac{1}{n}(X_n)\right] = \frac{1}{n}(\mu) + ...+ \frac{1}{n}(\mu) = \mu$$

$$V(\bar{X}) = V\left[\frac{1}{n}(X_1) + ...+ \frac{1}{n}(X_n)\right] = \frac{1}{n^2}(\sigma^2) + ...+ \frac{1}{n^2}(\sigma^2) = \frac{\sigma^2}{n}$$

Si la distribución de una población es normal con media μ y varianza σ^2, entonces la distribución muestral de la media ($\bar{X}$) para una muestra de tamaño *n*, también tiene distribución normal con media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$.

La relación entre la distribución muestral de $\bar{X}$ y la distribución poblacional está dada de la siguiente forma:

$$\mu_{\bar{X}} = \mu \quad \text{y} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Algunas propiedades de la distribución de $\bar{X}$ se puede deducir de su distribución muestral:

1. La media de la distribución es igual a la media de la población ($\mu_{\bar{X}} = \mu$), entonces $\bar{X}$ es un estimador insesgado de μ.
2. Para muestras grandes, los valores de $\bar{X}$ tienen menor dispersión, dado que $\sqrt{n}$ hace disminuir a la dispersión conforme aumenta *n*.

En el caso anterior, se partió del supuesto que la distribución de la población era normal, ¿pero que se puede decir acerca de la distribución muestral de $\bar{X}$, si la población de la que se selecciona la muestra no está no está normalmente distribuida?. Esto nos conduce al teorema central del limite, uno de los resultados estadísticos más importantes y que está estrechamente relacionado con el comportamiento de las distribuciones muestrales.

Supongamos que $X_1, X_2, ..., X_n$ *es una muestra aleatoria de tamaño* n*, extraída de una población con distribución de probabilidad cualquiera con media* μ *y varianza* σ^2*, las* X_i *son variables aleatorias independientes con* $E(X_i) = \mu$ *y* $V(X_i) = \sigma^2$*, para*

$i = 1,2,3,...,n$. *Entonces la distribución de* $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ *converge a la distribución normal estándar cuando* $n \to \infty$.

Sin perdida de generalidad, el teorema central del límite establece que la distribución muestral de $\bar{X}$ es aproximadamente normal, independientemente de cual sea la forma de la distribución de la población, siempre que el tamaño de muestra sea lo suficientemente grande. Usualmente un valor de *n* mayor que 30 asegura que la distribución de $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ pueda ser muy aproximada a la distribución normal.

Esto constituye un resultado de gran trascendencia en inferencia estadística, ya que permite usar la distribución normal como modelo de una distribución muestral para calcular la probabilidad de que ciertos valores del estadístico tengan lugar, sin tener que atender la forma de la distribución de la población de la que son extraídas las muestras.

2. Distribución muestral de una proporción

La distribución de una proporción ($\hat{p}$) es la distribución de los valores de $\hat{p}$ de todas las muestras posibles del mismo tamaño extraídas de una población. Podemos considerar como X, al número de elementos con cierta característica en una muestra, como una suma formada por ceros y unos, es decir ,

$$X = \sum_{i=1}^{n} Y_i$$

en donde

$$Y_i = \begin{cases} 1 \\ 0 \end{cases}, \qquad \text{1 si el i-ésimo elemento posee la característica y 0 en caso contrario.}$$

Entonces la proporción de elementos con la característica en la muestra es una media, porque

$$\frac{X}{n} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}$$

Por lo tanto es posible aplicar para el caso de la distribución muestral de una proporción, el mismo razonamiento que se aplicó en la distribución de la media. Además, como consecuencia del teorema central del límite, la distribución binomial con parámetros *n* y *p* se puede aproximar a la distribución normal, para lo cual $\mu = np$ y $\sigma^2 = npq$. Entonces, si $\hat{p} = \frac{X}{n}$ es la proporción muestral, donde X es la variable aleatoria que representa el número de elementos que poseen cierta característica en una muestra de tamaño n, se tiene

$$E(\hat{p}) = E\left(\frac{1}{n}X\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

$$V(\hat{p}) = V\left(\frac{1}{n}X\right) = \frac{1}{n}V(X) = \frac{1}{n^2}(npq) = \frac{pq}{n}$$

Si la distribución de una población es binomial con parámetros *n* y *p,* la cual puede ser aproximada a una distribución normal con media $\mu = np$ y varianza $\sigma^2 = npq$, entonces la distribución muestral de la proporción ($\hat{p}$) para una muestra de tamaño *n*, tiene distribución aproximadamente normal con media p y desviación estándar $\sqrt{\frac{np}{q}}$.

La relación entre la distribución muestral de $\hat{p}$ y la distribución poblacional está dada de la siguiente forma:

$$\mu_{\hat{p}} = p \qquad \text{y} \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Algunas propiedades de la distribución de $\hat{p}$ se puede deducir de su distribución muestral:

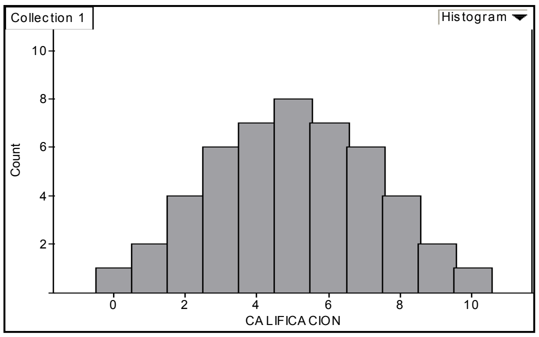
1. La media de la distribución es igual a la proporción poblacional ($\mu_{\hat{p}} = p$), entonces $\hat{p}$ es un estimador insesgado de p.
2. Para muestras grandes, los valores de $\hat{p}$ tienen menor dispersión, dado que $\sqrt{n}$ hace disminuir a la dispersión conforme aumenta *n*.

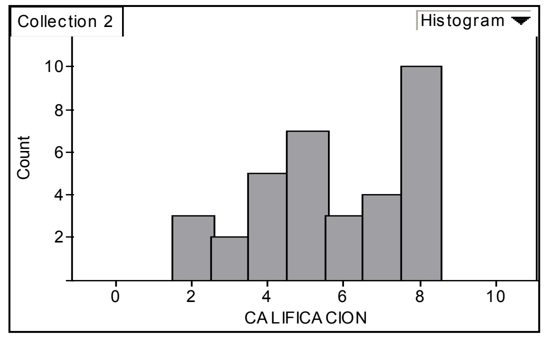
Cuestionario diagnóstico

Nombre:____ Edad__________

Lee cuidadosamente las instrucciones en cada pregunta y contesta lo que se te pide. Si te equivocas no borres, sólo encierra en un círculo lo que está incorrecto. Donde se te pide que expliques tus resultados, trata de hacerlo en la forma más detallada posible.

1. Marca con una X la distribución de datos que tiene más variabilidad.





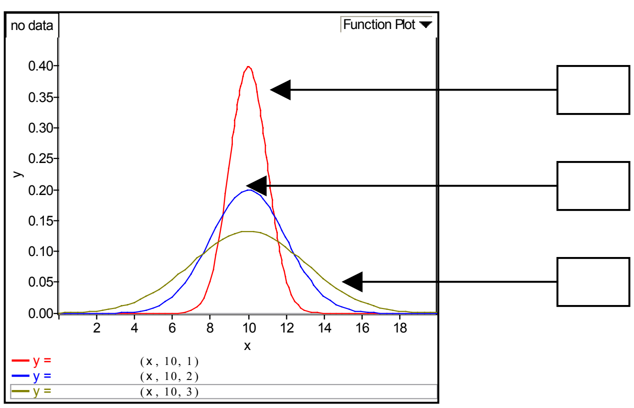
Explica con todo detalle las razones de tu elección.

____

____

____

2. En las siguiente grafica aparecen tres distribuciones poblacionales cuya media es $\mu = 10$ y sus desviaciones estándar son $\sigma = 1$, $\sigma = 2$ y $\sigma = 3$ respectivamente. Coloca sobre cada una de ellas la desviación estándar que le corresponde.



Explica las razones de tu asignación.

____

____

____

____

3. Si una moneda bien fabricada es lanzada una gran cantidad de veces, la proporción de águilas que aparecerá será muy cercana a 0.5. Supongamos que tomas 5 muestras de 10 lanzamientos cada una. Escribe cuántas águilas esperarías que aparecieran en cada una de las 5 muestras.

N° de águilas en la muestra 1: __________
N° de águilas en la muestra 2: __________
N° de águilas en la muestra 3: __________
N° de águilas en la muestra 4: __________
N° de águilas en la muestra 5: __________

Explica por qué.
____
____
____

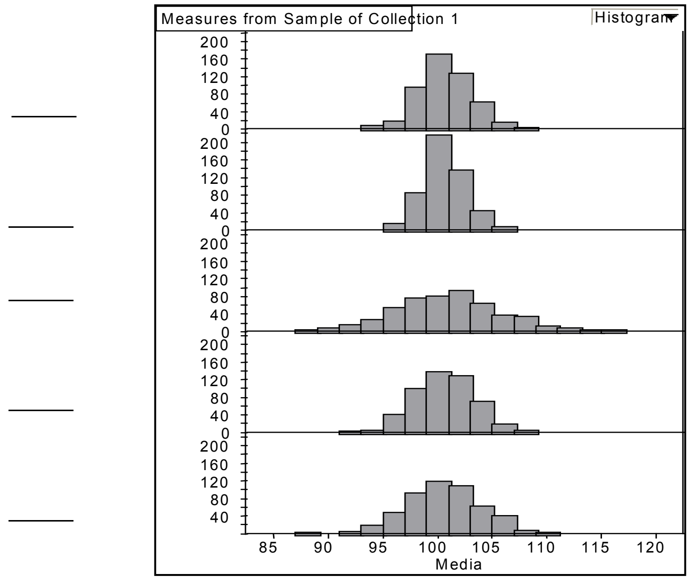
4. Se realizó una encuesta con una muestra aleatoria de 400 estudiantes universitarios con la finalidad de conocer su opinión sobre algunas reformas al reglamento escolar. Un estudiante duda de la validez del estudio argumentando que hay 4000 estudiantes en la universidad y sólo se encuestó a 400. Lee cuidadosamente cada enunciado que se muestra a continuación y selecciona la respuesta que consideres más razonable.

a. Estoy de acuerdo con él, 400 es una muestra muy pequeña (10%) de los 4000 estudiantes, para poder sacar conclusiones.
b. Estoy de acuerdo con él, debería tenerse una muestra del al menos 50% de la población para poder hacer inferencias.
c. Estoy de acuerdo con él, deberían encuestar a todos los estudiantes.
d. Estoy en desacuerdo con él, 400 es un número suficientemente grande para estos propósitos si la muestra fue seleccionada aleatoriamente.
e. Estoy en desacuerdo con él, si la muestra es aleatoria el tamaño de la muestra no importa.

5. Una empresa eléctrica fabrica focos que tienen una duración que se distribuye aproximadamente normal con media de 800 hrs. y una desviación estándar de 40 hrs. Si se toma una muestra de 16 focos, ¿Cuál será la proporción de focos con una vida promedio menor a 775 horas? Muestra todos tus cálculos.

6. De una población con distribución normal se extrajeron 500 muestras aleatorias de cada tamaño (5, 10, 15, 20 y 25). Se calculó la media de cada muestra y los resultados se dibujaron en los histogramas que se muestran en la siguiente figura.

a) Coloca a un lado de cada histograma, el tamaño de la muestra que corresponda.



Explica la razón de tu asignación

b) ¿Cuál es el valor de la media de la población de donde se extrajeron las muestras?. Explica por qué.

7. Si se selecciona una muestra aleatoria de una población. ¿Qué información puedes calcular de los datos de la muestra?. Subraya los símbolos que corresponden a dicha información y explica que significan.

σ $\qquad$ s $\qquad$ μ $\qquad$ $\bar{x}$

8. Subraya los símbolos que representan parámetros poblacionales y encierra en un círculo los que representan estadísticos muestrales.

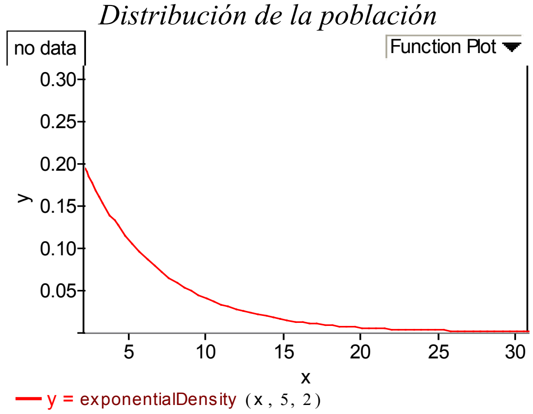
σ $\qquad$ s $\qquad$ μ $\qquad$ $\bar{x}$

9. Una muestra de 50 datos es seleccionada de una población de temperaturas obteniéndose una media muestral de 20 °C. ¿Cuál sería tu mejor estimación de la media poblacional μ?

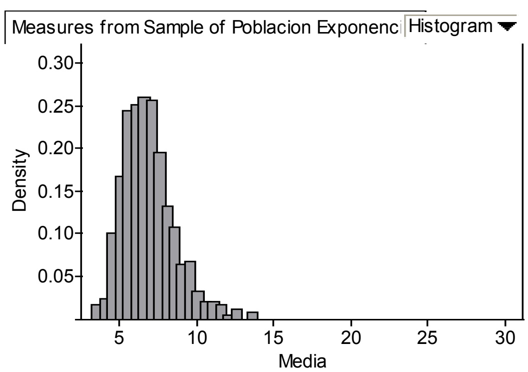
a) Sería exactamente 20 °
b) Sería cercana a 20 °
c) No sería posible hacer una estimación, pues μ es un parámetro desconocido y la información que se tiene se refiere sólo a una muestra.
d) Otra respuesta

10. En la figura se muestra la distribución de una población y dos distribuciones muestrales para muestras de tamaño 2 y 15. Coloca el tamaño de muestra a la distribución que corresponda. Nota: se tomaron 500 muestras en cada caso.

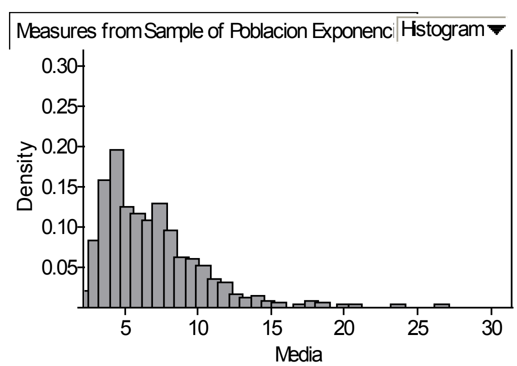
Distribución de la población



Distribuciones muestrales



n = __________



n =___________

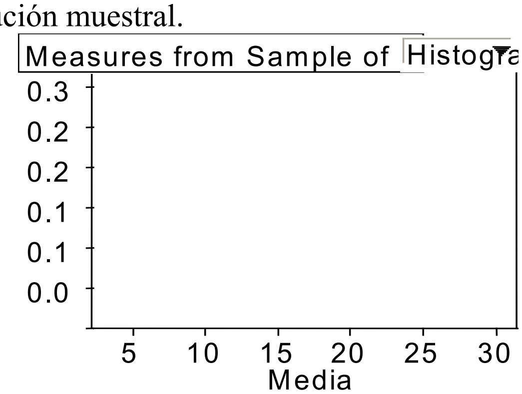
Explica las razones de tu asignación

____

____

____

11. En relación con el ítem anterior. Si la muestra fuera de tamaño 30, dibuja la forma que podría adoptar la distribución muestral.



Explica:

12. Consideremos como una población una gran urna que contiene el 70% de bolas rojas. Si se selecciona una muestra aleatoria de la población ¿Qué consideras que sea más probable de ocurrir?

a) 7 bolas rojas en una muestra de tamaño 10.
b) 70 bolas rojas en una muestra de tamaño 100.
c) Ambos casos se tiene la misma probabilidad de ocurrir

Explica.

13. La compañía de chocolates M&M´s señala que el 30% de los chocolates que vienen en la presentación de bolsa amarilla, son de color café. Se tomó una muestra de 10 bolsas y se resultó que la proporción de chocolates café era de 25%.

a) ¿Cómo explicas el resultado anterior?

b) ¿Cuál es el estadístico y cuál es el parámetro en el problema anterior?

14. Bajos condiciones similares, tres compañías encuestadoras han realizado un estudio para determinar la opinión de los ciudadanos sobre el desempeño del presidente de la república. ¿Esperarías que obtuvieran el mismo resultado? Explica tu respuesta.

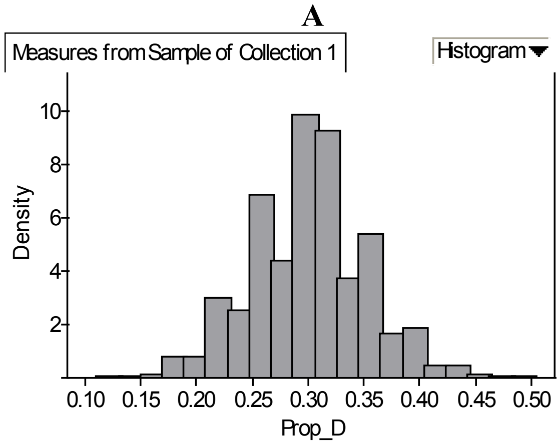
Cuestionario posterior

Nombre:____

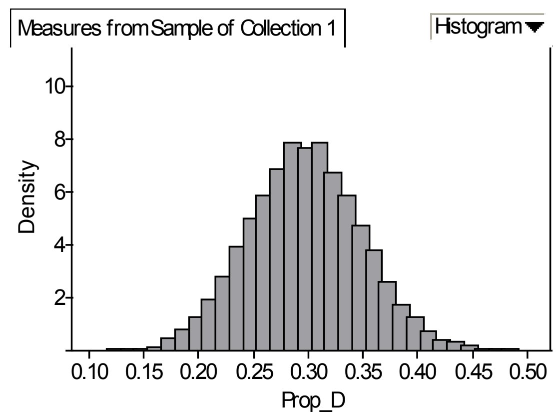
Lee cuidadosamente las instrucciones en cada pregunta y contesta lo que se te pide. Explica en la forma más detallada posible.

1. Observa las siguientes distribuciones muestrales y subraya la opción que consideres correcta.

A



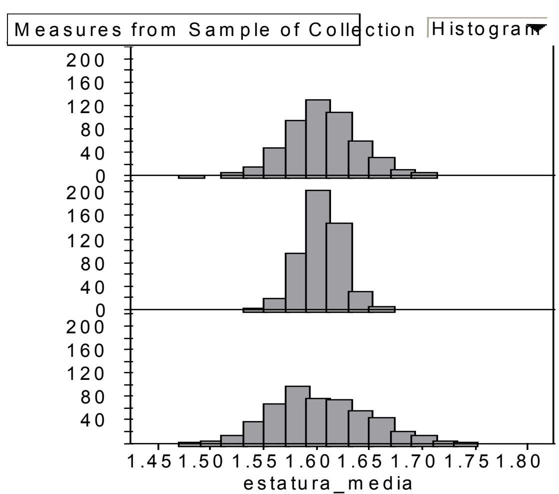
B



a) La distribución A tiene mayor variabilidad que la distribución B.
b) La distribución B tiene mayor variabilidad que la distribución A.
c) Ambas distribuciones tienen la misma variabilidad.

Explica con todo detalle las razones de tu elección.

2. A continuación se muestran tres distribuciones muestrales de la media (para muestras de tamaño n = 5, n = 10 y n = 30), las cuales se han construido tomando 500 muestras de una misma población.



a) Coloca sobre cada una de ellas el tamaño de muestra que le corresponde. Explica con todo detalle las razones de tu asignación.

b) Sobre la gráfica de las distribuciones, coloca 1 a la distribución de menor dispersión y 2 a la distribución de mayor dispersión. Explica con todo detalle las razones de tu asignación.

c) ¿Cuál será la media de la población de la que se extrajeron las muestras?. Explica.

d) ¿En cuál de las tres distribuciones será más probable obtener una media mayor a 1.65 cm.? (Obsérvese que los datos son estaturas). Explica en forma detallada.

3. En una urna se tienen 4 bolas etiquetadas con los números 2, 4, 6 y 8. Se seleccionan muestras de dos bolas al azar simultáneamente y se calcula la media de los números que aparecen en ellas. Describe mediante una tabla ó grafica la distribución muestral que resulta.

4. La compañía M&M´s dice que el 30% de sus chocolates en la presentación Milk son color café. Antes de ser empacados en bolsas los chocolates se encuentran en un gran depósito donde son mezclados de manera uniforme. Si se seleccionan de forma independiente 5 muestras de tamaño 10, una tras otra. ¿Cuántos chocolates café esperarías en cada muestra?

N° de chocolates café en la premier muestra: ____
N° de chocolates café en la segunda muestra: ____
N° de chocolates café en la tercer muestra: ____
N° de chocolates café en la cuarta muestra: ____
N° de chocolates café en la quinta muestra: ____

Explica:
5. Imagina que lanzas un dado 60 veces. Muestra en la tabla siguiente el número de veces que esperas que aparezca cada número.

Numero en el dado	Número de veces que se espera aparezca el número
1	
2	
3	
4	
5	
6	
Total	

Explica las razones de tu asignación:

6. Se midió la estatura a una muestra aleatoria de 100 estudiantes de la población del IPN y se encontró que la media era de 160 cm. ¿Cuál sería tu estimación de la media poblacional?

e) Sería exactamente 160 cm.
f) Sería cercana a 160 cm.
g) No sería posible hacer una estimación, pues la información obtenida se refiere sólo a una muestra.
h) Otra respuesta. ____________________

Explica:

7. Otra persona hizo lo mismo, tomó una muestra aleatoria de 100 estudiantes del IPN y encontró que la media era 157 cm. ¿Cómo podrías explicar la diferencia de resultados?

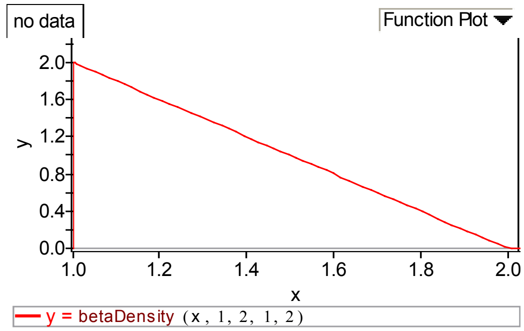
8. Si aumentamos a 200 el número de estudiantes en el problema anterior (tamaño de muestra). ¿Qué esperarías?:

a) La estimación estuviera más cercana a la media poblacional que cuando la muestra es de 100.
b) La estimación estuviera más lejana a la media poblacional que cuando la muestra es de 100.
c) La estimación estuviera a igual distancia de la media poblacional que cuando la muestra es de 100.

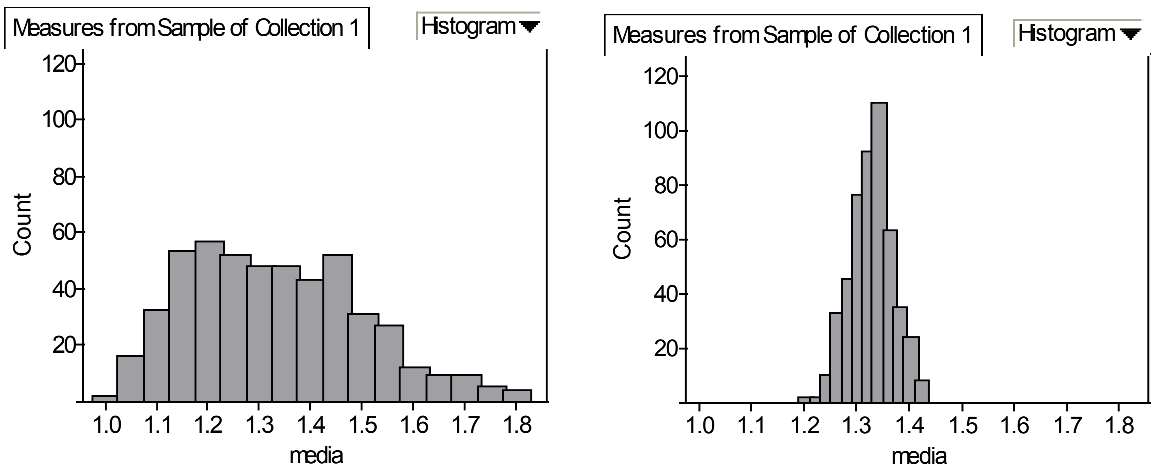
Explica.

9. En la figura se muestra la distribución de una población y dos distribuciones muestrales de medias para muestras de tamaño 2 y 30. Coloca el tamaño de muestra a la distribución que corresponda.

Distribución de la población



Distribuciones muestrales



n = _________ n =__________

Explica las razones de tu asignación.

10. En una población de estudiantes el peso promedio es de 60 kg. y una desviación estándar de 5 kg. Si se seleccionan dos muestras aleatorias de dicha población, una de tamaño 10 y otra de tamaño 100 en cual consideras que sea mas probable obtener una media muestral muestral menor a 55 kg.?

d) En la muestra de tamaño 10.
e) En la muestra de tamaño 100.
f) En ambos muestras se tiene la misma probabilidad.

Explica.

11.Una empresa eléctrica fabrica focos que tienen una duración que se distribuye aproximadamente normal con media de 700 hrs. y una desviación estándar de 32 hrs. Se ha tomado una muestra de 16 focos y se encontró que la duración media fue de 684 hrs.

a) Basado en los resultados obtenidos, ¿podrías concluir que la empresa está mintiendo sobre la duración de los focos?. Explica tu respuesta (puedes realizar cálculos si lo deseas).

b) ¿Crees que si se toma una muestra de tamaño 100, la duración media seguirá siendo de 684 hrs.?. Explica.

c) Tomando en cuenta la información de la población, ¿Cuál será la proporción de focos con una vida promedio menor a 708 horas?. Resuelve el problema con el método que te parezca más conveniente (simulación con Fathom o de manera teórica).