UNIVERSIDAD DE BUENOS AIRES

ESTADÍSTICA - CÁTEDRA I

CÓDIGO 60

PROF. ADJUNTA A CARGO: DRA. MARÍA SILVIA GALIBERT

TRABAJO PRÁCTICO Nro. 4

AUTORES

Profesores y Docentes de la Cátedra I

> "La vida sería intolerable si los fenómenos ocurrieran al azar de una forma completamente impredecible y carecería de interés si, en el otro extremo, todo fuera determinista y completamente predecible".
>
> Radhakrishna Rao

PRÁCTICA 4. Inferencia Estadística y Modelos de Probabilidad

1.- Para trabajar en clase. Realice una lectura comprensiva del resumen y el fragmento del artículo y responda las preguntas que se formulan a continuación del mismo.

RESUMEN

Antecedentes: durante la Educación Secundaria, etapa con riesgo de fracaso y abandono escolar, el apoyo social representa una variable contextual relevante para prevenir el desajuste escolar. El objetivo de este trabajo es examinar un modelo teórico sobre la capacidad explicativa del apoyo social sobre el ajuste escolar –implicación escolar y rendimiento académico percibido–. **Método:** participan 1.468 estudiantes (51% chicas; 49% chicos) con edades entre 12 y 17 años (M=14.03; DT=1.36) del País Vasco. Es un estudio con diseño ex post facto transversal. Las medidas empleadas son: TCMS –subescala apoyo profesorado–, AFA-R –subescalas apoyo familiar y apoyo amistades–, SEM –escala implicación escolar– y EBAE-10 –subescala rendimiento académico percibido–. Se comprueban varios modelos estructurales. **Resultados:** el modelo de primera elección es el de predicción del apoyo social sobre la implicación escolar con el rendimiento académico percibido como variable mediadora: predicen conjuntamente un 73% de la implicación y prevalece el efecto del apoyo del profesorado, seguido del apoyo familiar, frente a la ausencia de efecto directo de amistades sobre las variables de ajuste escolar. **Conclusiones:** el profesorado y la familia deben ofrecer apoyo social al alumnado para reforzar la percepción de autoeficacia académica y la implicación escolar.[1]

Participants

Participants were 1,457 students, 51% girls and 49% boys, from 5 public secondary schools (61%) and 4 semi-private ones (39%) in the Basque Country. The semi-private schools (i.e., private schools which receive some state-funding) are attended by students from families with a medium-high socioeconomic level, while those attending 3 of the public schools have a medium socioeconomic level and those attending the other 2 have a low level. Participants were selected by systematic random sampling from the official list published by the Basque Government Department of Education. Participants were all aged between 12 and 17 years (*M*=14.03; *SD*=1.36)

Fernández Lasarte, O., Ramos Díaz, Goñi Palacios, E. y Rodríguez Fernández, A. (2020). El papel del apoyo social en el ajuste escolar en Educación Secundaria. *Psicothema 32,1,100-107.*

---

[1] Los extractos que se leen en la consigna han sido transcriptos de manera textual del artículo científico allí citado. Mantuvimos el idioma original de la publicación ya que la intención es que ustedes tengan la oportunidad de leer cómo escriben los investigadores y puedan apreciar las decisiones que toman también a la hora de divulgar sus producciones científicas. Por ejemplo, decidir traducir los trabajos al inglés, en vez de publicarlos en el idioma nativo, puede relacionarse con la mayor difusión que tienen los escritos en dicho idioma. También habrán podido notar que existe una discrepancia entre los tamaños muestrales que se leen en el resumen en español y el apartado "participants" en inglés (1468 y 1457, respectivamente). Los autores mencionaron en otro lugar que hay 11 casos en los que se perdieron las observaciones de las variables sexo y edad. Sin embargo, ellos deberían haber utilizado el mismo criterio en el resumen y en el abstract.

a) ¿Quiénes constituyen la población objetivo y quiénes integran la muestra?
b) ¿Con qué método de muestreo se obtuvo la muestra y en qué consiste el mismo?
c) ¿Qué diseño se utilizó para el estudio? Para tener una idea de lo que significa, puede buscarlo en el siguiente enlace:https://www.studocu.com/es/document/universitat-jaume-i/metodos-y-tecnicas-de-investigacion/apuntes/disenos-ex-post-facto/2474601/view

2.- Un examen de elección múltiple consiste en 20 preguntas, cada una de las cuales tienen 4 alternativas con una sola verdadera (clave). La calificación del examen consiste en la cantidad de respuestas correctas. Suponga una población hipotética de alumnos que contestaran todo por puro azar (como sorteando entre las cuatro alternativas con igual probabilidad). El examen se aprueba con 4 puntos.
a) Identifique la variable Bernoulli asociada a cada pregunta y su parámetro.
b) Defina convenientemente una variable Binomial e identifique los parámetros de su distribución para responder a las siguientes preguntas:
i) ¿Cuál es la probabilidad de aprobar por azar?
ii) ¿Qué porcentaje de estudiantes en tal población hipotética aprobaría respondiendo por azar?
iii) ¿Cuál es la cantidad esperada de respuestas correctas por azar?
iv) ¿Cuál debería ser el punto de corte para que a lo sumo aprobara por azar el 10% de los estudiantes? (Considere en este punto la probabilidad redondeada a dos decimales.)
v) ¿Qué modificaciones en el diseño de la prueba propondría para disminuir la probabilidad de aprobación por azar?
vi) **Conceptualice:** ¿Cuáles son los supuestos (o condiciones) de una distribución binomial y qué es lo que en el presente ejercicio asegura que se satisfacen?

3.- Cierto psicólogo tiene la experiencia de que el 40% de los pacientes asistentes a un grupo de recuperación abandona los talleres antes de culminar el programa. ¿Qué probabilidad le asignaría a que en un nuevo grupo de 8 pacientes persevere la mayoría hasta el final?

4.- Cierta cátedra tiene 500 alumnos inscriptos y desea estimar la razón (o proporción) de estudiantes que se sienten insatisfechos con el dictado de la materia. Para ello va a encuestar a 50 estudiantes elegidos de manera aleatoria, quienes responderán de manera anónima.
i) ¿Qué cálculo haría para estimar la razón de interés?
ii) **Conceptualice:** ¿Qué vinculación hay entre ese cálculo y una variable Binomial?

5.- Para el artículo cuya referencia está a continuación se extrajeron un par de fragmentos. Intente resolver el problema planteado y luego haga una lectura comprensiva de la resolución que se encuentra inmediatamente a continuación del enunciado.

José Luis Marcos Malmierca y Alfonso Barca Lozano (2009). Efecto del nivel de consciencia de la presencia del estímulo sobre el aprendizaje de expectativa *Psicothema 21*,3,*397-402.*

El artículo completo se halla en www.psicothema.com/pdf/3644.pdf. A continuación se extractan algunos párrafos, se describe parte de la experiencia y se plantea un problema con su resolución.

Participantes (Textual)

La muestra inicial estaba compuesta por 92 estudiantes de las titulaciones de Logopedia y Terapia Ocupacional, que participaron de modo voluntario tras haber sido convenientemente informados sobre el experimento.

Procedimiento (Textual)

c) Fase de evaluación de la consciencia. Para evaluar el grado de consciencia del estímulo enmascarado E1 se efectuaron 80 ensayos de una tarea de identificación de elección forzada, bajo las mismas condiciones visuales y de duración que en las fases anteriores. E1 estímulo enmascarado E1 consistía en la letra «O» en la mitad de los ensayos y en la letra «X» en la otra mitad. Inmediatamente después de la presentación de cada secuencia el participante debía indicar qué letra había sido presentada («O» o «X»), pulsando las correspondiente teclas «O» y «X» del teclado del ordenador.

PROBLEMA

Los autores proceden luego a clasificar a los participantes teniendo en cuenta el nivel de conciencia del estímulo enmascarado de acuerdo con el desempeño que tuvieron en la tarea de elección forzosa. El criterio que adoptan para ello es clasificar como semiconscientes o conscientes a los sujetos cuya cantidad de identificaciones correctas superan cierto umbral que es poco probable (menos de 0,05), de alcanzar por azar; el resto es clasificado como no consciente.

¿Cuál es el umbral (o punto de corte) para dicha clasificación?

RESOLUCIÓN

X = Número de identificaciones correctas por azar entre las 80 presentaciones.
$X \sim B(80; 0,5)$
Debemos hallar r tal que $P(X > r) < 0,05$, lo cual es equivalente buscar el primer valor de r donde la probabilidad acumulada supera a 0,95. Ese valor es r = 47.

Veamos ahora lo que dicen los autores

Medida y análisis (Textual)

Los participantes que en la tarea de detección de elección forzada mostraban unos resultados de detección correcta inferior al 60% eran asignados al grupo no-consciente, ya que, según una distribución binomial, con N = 80 (ensayos) y p (identificación correcta)=0.5, la probabilidad de que ocurran 48 (60%) o más de 48 identificaciones correctas es de 0.046. Aplicando ese criterio, fueron asignados a este grupo no-consciente 32 sujetos. Los participantes con identificaciones correctas que oscilaban entre 48 (60%) y 64 (80%) fueron asignados al grupo semiconsciente, que quedó constituido por otros 32 sujetos. Finalmente, otros 28 participantes que habían logrado superar 64 detecciones correctas (+80%) conformaron el grupo consciente.

6.- El puntaje en un test de memoria se distribuye normalmente con una media de 20 puntos y una desviación estándar de 5 puntos en un grupo normativo. Calcule:
a) La probabilidad de que una persona elegida al azar de la población representada por el grupo normativo obtenga una puntuación
i) superior a 27.

ii) inferior a 10.
iii) entre 17 y 23.

b) La mínima puntuación que debería obtener una persona para considerarse dentro del 2% de mayores puntuaciones.

7.- El puntaje T en un test de habilidades espaciales tiene distribución normal en la población de interés. Se cosideran atípicos los sujetos que se encuentran a más de dos desviaciones estándar de la media. Halle:
a) el rango percentilar de 62.
b) el percentil 90.
c) el porcentaje de sujetos atípicos en la población.
d) los umbrales de los puntajes atípicos.

8.- Las calificaciones en un examen de ingreso puntuado de 0 a 100 se distribuyen normalmente en una población de aspirantes. El área bajo la curva normal entre los puntajes 70 y 80 es 0,15. ¿Qué representa esa área? Redacte una oración expresando ese resultado de manera coloquial como si lo estuviera informando.

9.- El tiempo de reacción ante un estímulo se distribuye normalmente con media 8 segundos y desviación estándar 1,2 segundos. Un psicólogo opina que hay que preocuparse por el grupo de las personas con reacción lenta definido como el de los que tardan tiempos mayores y conforman el 20% del total.
a) ¿Cuál es el tiempo de reacción a partir del cual hay que preocuparse?
b) ¿A cuántas desviaciones estándar, como mínimo, por encima de la media se hallan las personas con reacción lenta?
c) Entre 400 personas elegidas al azar de la población, ¿cuántas cabría esperar con reacción lenta?

10.- Suponga que el tiempo necesario para resolver un test se distribuye normalmente con una media de 50 minutos y una desviación estándar de 8 minutos en la población objetivo. Imagine ahora que dichos parámetros no son conocidos por los psicometristas que diseñan el test y que lo administrarán en una muestra piloto de 36 personas con el fin de estimar el tiempo promedio que llevaría realizarlo en la población general.
a) ¿Cuál es la probabilidad de que la media estimada sea
i) mayor a 52?
ii) mayor a 52 o menor a 48 (error muestral de más de 2 puntos)?

b) **Conceptualice.** Aun cuando no se tuvieran hipótesis sobre la media ni sobre la desviación estándar de la población, se puede conocer la probabilidad de cometer un error de estimación superior a, por ejemplo, 1,5 desviaciones estándar de la media. Plantee y calcule dicha probabilidad.

11.- Conceptualice: Deténgase a considerar y comprender que la media muestral, como variable aleatoria, tiene una distribución de probabilidades y que esa distribución es aproximadamente normal (por el Teorema Central del Límite) en torno al parámetro (μ) que desea estimar, cuando el tamaño de muestra es suficientemente grande. Asegúrese de comprender esto porque aquí va lo principal de la fundamentación de los métodos de inferencia estadística que se verán en adelante y es clave para su comprensión. Para ayudarse, entonces, complete la siguiente frase:

Sea X una variable aleatoria con esperanza (media) μ y desviación estándar σ. El promedio de los valores de X provenientes de muestras de tamaño n (media muestral: $\bar{X}$) es aproximadamente normal con esperanza: ……………… y desviación estándar ……………. para n suficientemente grande (n > 30). Realice un gráfico aproximado de la distribución de X y de la de $\bar{X}$ en torno a μ. Note que, como es deseable que ocurra, intervalos de valores de la media muestral próximos a la media poblacional tienen más probabilidad de aparecer en una muestra que intervalos de valores lejanos (consideramos para la comparación intervalos de igual longitud). Eso es bueno porque justamente se desea estimar el valor desconocido de μ a partir del valor conocido de una muestra. La diferencia entre ambos es el error de muestreo y es deseable que dicho error sea pequeño.

12.- Las siguientes afirmaciones resultan de teoremas que se necesitan para fundamentar los métodos de inferencia estadística. Las mismas están expresadas en símbolos.

a) Traduzca en lenguaje coloquial las expresiones simbólicas para asegurarse de que comprende el significado de los símbolos.

$$X \sim N(\mu, \sigma) \Rightarrow \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Si S es la desviación estándar muestral (correspondiente a dividir la suma de cuadrados de los desvíos por n-1) entonces:

$$X \sim N(\mu, \sigma) \Rightarrow \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Si n es grande ($n > 30$), por Teorema de Slutzky,

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$$

Si $\hat{p}$ es la frecuencia relativa o proporción muestral y p es la probabilidad de éxito o proporción poblacional que se estima a través de $\hat{p}$

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx N(0,1) \quad si \quad n\hat{p} \geq 5 \; y \; n(1-\hat{p}) \geq 5$$

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \quad si \quad np \geq 5 \; y \; n(1-p) \geq 5$$

Al considerar la tabla de contingencia de una distribución bivariada,

$$\sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(f_{ij}^{o} - f_{ij}^{e})^2}{f_{ij}^{e}} \sim \chi^2_{(f-1)\mathrm{x}(c-1)}$$

Donde las f_{ij} aluden a las frecuencias absolutas observadas y esperadas bajo la hipótesis de independencia en una tabla de contingencia.

b) Halle los percentiles 5 y 95 de las distribuciones Normal Estándar, t de Student con 9 grados de libertad y χ^2 con 9 grados de libertad. Elija una aplicación apropiada para hacerlo (PQRS, Excel, Probability Distribution, Statistix, Infostat, etc.)

c) **Conceptualice:** Deténgase a comprender los conceptos de Estadístico, Estimador y Parámetro. Reconozca en todas las expresiones anteriores quiénes son estadísticos, quiénes son estimadores y quiénes parámetros. Para ello complete la siguiente tabla tildando la celda que corresponda. La primera línea está hecha como ejemplo.

NOMBRE	EXPRESIÓN	ESTADÍSTICO	ESTIMADOR	PARÁMETRO
Media poblacional o esperanza	μ			✓
	σ			
	$\bar{X}$			
	$\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$			
	$\hat{p}$			
	p			
	$\dfrac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$			
	f_{ij}^{o}			
	$\sum_{i=1}^{f}\sum_{j=1}^{c}\dfrac{(f_{ij}^{o}-f_{ij}^{e})^2}{f_{ij}^{e}}$			

13.- Conceptualice. $\bar{X}$ y μ aluden a la media, σ y S a la desviación estándar, $\hat{p}$ y p a la proporción o frecuencia relativa.

a) ¿Por qué se utilizan dos símbolos diferentes para aludir a cada uno de estos conceptos? Indique cuál es la diferencia conceptual dentro de cada uno de estos pares de símbolos.

b) Asocie la palabra "poblacional" o bien "muestral" a cada uno de estos seis símbolos: $\bar{X}$, μ σ, S, $\hat{p}$ y p. Por ejemplo: $\bar{X} \rightarrow$ muestral.

14.- Dé un ejemplo de una población objetivo y de una variable cuantitativa por estudiar en esa población. Indique cuáles son las unidades de análisis. Para esa variable en esa población dé tres ejemplos de parámetros. Luego suponga que extrae una muestra aleatoria de dicha población. Dé tres ejemplos de estimadores de los correspondientes parámetros antes mencionados.

En los siguientes ítems de elección múltiple elija la única opción correcta.

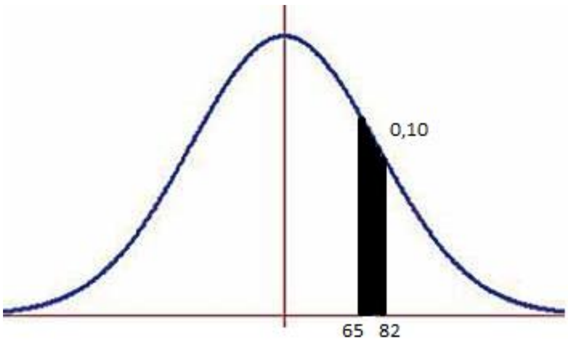
15.- Se denomina error estándar de un estimador a:
a) su desviación estándar.
b) su discrepancia con el parámetro que estima.
c) su probabilidad de distar más de una desviación estándar del parámetro.

16.- Un estadístico es
a) un valor numérico que se calcula a partir de una muestra aleatoria.
b) un valor característico de la población objetivo de la que se saca la muestra.
c) una variable cuyos valores dependen de una muestra aleatoria.

17.- Si z_p y t_p son los percentiles 70 de las distribuciones normal estándar y t de Student entonces
a) $z_p = t_p$
b) $z_p < t_p$
c) $z_p > t_p$

18.- Los puntajes en una escala de inhibición conductual se distribuyen normalmente en un grupo normativo de preescolares con una media de 35 y una desviación estándar de 8. Un establecimiento educativo decide brindar asistencia terapéutica a los niños correspondientes a la última interdecila. ¿Qué puntaje (redondeado) debe tener como mínimo un niño en esta escala para recibir asistencia terapéutica?
a) 90
b) 25
c) 45

19.- Los puntajes en un test de creatividad se distribuyen normalmente en la población de estudiantes de la escuela CREARTE. La región sombreada en la siguiente figura corresponde a un área de 0,10.



¿Cómo se interpreta esa información con respecto a dicha área sombreada?

a) Hay 10 alumnos o más en la escuela CREARTE que tienen puntaje superior a 65 en el test de creatividad.
b) El rango percentilar del puntaje 82 en el test de creatividad es 10 para los alumnos de la escuela CREARTE.
c) El 10% de los alumnos de la escuela CREARTE tiene un puntaje en el test de creatividad entre 65 y 82.

RESOLUCIÓN

1.- Para trabajar en clase. Realice una lectura comprensiva del resumen y el fragmento del artículo y responda las preguntas que se formulan a continuación del mismo.

RESUMEN

Antecedentes: durante la Educación Secundaria, etapa con riesgo de fracaso y abandono escolar, el apoyo social representa una variable contextual relevante para prevenir el desajuste escolar. El objetivo de este trabajo es examinar un modelo teórico sobre la capacidad explicativa del apoyo social sobre el ajuste escolar –implicación escolar y rendimiento académico percibido–. **Método:** participan 1.468 estudiantes (51% chicas; 49% chicos) con edades entre 12 y 17 años (M=14.03; DT=1.36) del País Vasco. Es un estudio con diseño ex post facto transversal. Las medidas empleadas son: TCMS –subescala apoyo profesorado–, AFA-R –subescalas apoyo familiar y apoyo amistades–, SEM –escala implicación escolar– y EBAE-10 –subescala rendimiento académico percibido–. Se comprueban varios modelos estructurales. **Resultados:** el modelo de primera elección es el de predicción del apoyo social sobre la implicación escolar con el rendimiento académico percibido como variable mediadora: predicen conjuntamente un 73% de la implicación y prevalece el efecto del apoyo del profesorado, seguido del apoyo familiar, frente a la ausencia de efecto directo de amistades sobre las variables de ajuste escolar. **Conclusiones:** el profesorado y la familia deben ofrecer apoyo social al alumnado para reforzar la percepción de autoeficacia académica y la implicación escolar.

Participants

Participants were 1,457 students, 51% girls and 49% boys, from 5 public secondary schools (61%) and 4 semi-private ones (39%) in the Basque Country. The semi-private schools (i.e., private schools which receive some state-funding) are attended by students from families with a medium-high socioeconomic level, while those attending 3 of the public schools have a medium socioeconomic level and those attending the other 2 have a low level. Participants were selected by systematic random sampling from the official list published by the Basque Government Department of Education. Participants were all aged between 12 and 17 years (*M*=14.03; *SD*=1.36)

Fernández Lasarte, O., Ramos Díaz, Goñi Palacios, E. y Rodríguez Fernández, A. (2020). El papel del apoyo social en el ajuste escolar en Educación Secundaria. *Psicothema 32,1,100-107.*

a) ¿Quiénes constituyen la población objetivo y quiénes integran la muestra?

La población está conformada por los estudiantes de secundaria de entre 12 y 17 años del País Vasco y la muestra está integrada por los 1468 participantes.

b) ¿Con qué método de muestreo se obtuvo la muestra y en qué consiste el mismo?

El método es de muestreo aleatorio sistemático. Consiste en elegir al azar el primer elemento de la lista; en este caso del listado oficial publicado por el Departamento de Educación del Gobierno Vasco, y seguir eligiendo los otros elementos a intervalos regulares.

c) ¿Qué diseño se menciona para el estudio?

Se menciona un diseño ex post facto trasversal. Ex post facto se refiere al hecho de que los valores de las variables se registran después de tomar los datos; es decir que, dentro del marco de muestreo elegido, las variables toman valores libremente, no están predeterminados. La transversalidad alude al hecho de que no se hace un seguimiento de las unidades experimentales a través del tiempo (lo que sería un diseño longitudinal).

2.- Un examen de elección múltiple consiste en 20 preguntas, cada una de las cuales tienen 4 alternativas con una sola verdadera (clave). La calificación del examen consiste en la cantidad de respuestas correctas. Suponga una población hipotética de alumnos que contestaran todo por puro azar (como sorteando entre las cuatro alternativas con igual probabilidad). El examen se aprueba con 4 puntos.
a) Identifique la variable Bernoulli asociada a cada pregunta y su parámetro.

La variable es "Condición de acertar o no acertar al responder al azar la pregunta presentada con 4 opciones de respuesta". El parámetro es la probabilidad de respuesta correcta por azar, que en al haber 4 opciones es $p = 1/4 = 0,25$.

b) Defina convenientemente una variable Binomial e identifique los parámetros de su distribución para responder a las siguientes preguntas:

La variable es "Cantidad de aciertos frente a 20 preguntas con 4 opciones de respuesta, respondiendo al azar".
$X \sim B(20; 0,25)$

i) ¿Cuál es la probabilidad de aprobar por azar?

Aprobar significa obtener como mínimo 4 puntos; en símbolos: $X \geq 4$.
Utilizando una aplicación calculamos $P(X \geq 4) = 0,7748$. La probabilidad de aprobar por azar es 0,7748.

ii) ¿Qué porcentaje de estudiantes en tal población hipotética aprobaría respondiendo por azar?

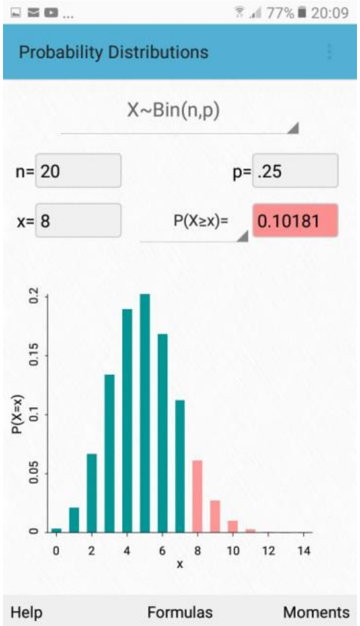
La probabilidad calculada en a): 0,7748, modeliza la proporción de estudiantes en la población que aprobaría por azar. Expresada porcentualmente, corresponde al 77,48%.

iii) ¿Cuál es la cantidad esperada de respuestas correctas por azar?

Es la esperanza de la variable: $E(X) = np = 20x0,25 = 5$. Habría en promedio 5 respuestas correctas en la población de alumnos que responden por azar.

iv) ¿Cuál debería ser el punto de corte para que a lo sumo aprobara por azar el 10% de los estudiantes?

Hay que buscar el valor x_0 de la variable tal que $P(X \geq x_0) \leq 0,10$. Si se usa la aplicación Probability, completando la probabilidad en la ventana de la derecha, devuelve dos valores 8 y 9. Muestra que la probabilidad a partir de 8 es 0,10181 y que la probabilidad a partir de 9 es 0,04093. Redondeando a dos decimales, puede tomarse 8 como nota mínima de aprobación. Si el problema se resuelve con otras aplicaciones, como EXCEL que acumulan la probabilidad a izquierda, hay que tomar el punto donde se acumula como máximo 0,90. En otras palabras, se pide el percentil 90 de la distribución.



v) ¿Qué modificaciones en el diseño de la prueba propondría para disminuir la probabilidad de aprobación por azar?

Se puede aumentar la cantidad de preguntas o la cantidad de opciones por pregunta o el puntaje mínimo de aprobación.

vi) **Conceptualice:** ¿Cuáles son los supuestos (o condiciones) de una distribución binomial y qué es lo que en el presente ejercicio asegura que se satisfacen?

Las condiciones son la ***independencia*** de las observaciones y la ***probabilidad de éxito constante*** en cada observación.

En este problema la ***independencia*** significa que: saber que se respondió bien (o mal) una pregunta no hace ni más ni menos probable que se responda bien (o mal) otra pregunta. Esa independencia está garantizada porque se está respondiendo todo el examen por azar; de modo el éxito o fracaso en la respuesta a cada pregunta no está asociado al éxito o fracaso en la respuesta a cualquier otra. La ***probabilidad de éxito es constante*** en cada observación. En este contexto, eso corresponde a que la probabilidad de contestar bien por azar es la misma para cada pregunta. Esa condición se satisface porque todas las preguntas tienen la misma estructura: 4 opciones con una sola verdadera. Si el examen hubiera tenido algunas preguntas de V/F, otras de cuatro opciones, otras de 3 opciones, etc., se habría cumplido la independencia pero no la probabilidad de éxito constante.

3.- Cierto psicólogo tiene la experiencia de que el 40% de los pacientes asistentes a un grupo de recuperación abandona los talleres antes de culminar el programa. ¿Qué probabilidad le asignaría a que en un nuevo grupo de 8 pacientes persevere la mayoría hasta el final?

X = Cantidad de pacientes, entre los 8 del grupo, que persevera en el programa hasta el final.
Como se informa que el 40% abandona entonces se sabe que el 60% persevera.
$X \sim B(8;0,6)$

La mayoría se define como la mitad más 1, como mínimo; por lo que la mayoría entre 8 es 5 o más.
Se pide calcular $P(X \geq 5) = 0,5941$

La probabilidad de que persevere la mayoría es 0,5941.

4.- Cierta cátedra tiene 500 alumnos inscriptos y desea estimar la razón (o proporción) de estudiantes que se sienten insatisfechos con el dictado de la materia. Para ello va a encuestar a 50 estudiantes elegidos de manera aleatoria, quienes responderán de manera anónima.

i) ¿Qué cálculo haría para estimar la razón de interés?

Calcularía la frecuencia relativa (razón muestral) de insatisfechos; es decir, la cantidad de insatisfechos en la muestra dividido 50.

ii) **Conceptualice:** ¿Qué vinculación hay entre ese cálculo y una variable Binomial?

El numerador de la frecuencia relativa, "Cantidad de insatisfechos en la muestra", es una variable Binomial de parámetros n=50 y el parámetro *p* es la proporción poblacional de insatisfechos, que es lo que se desea estimar.

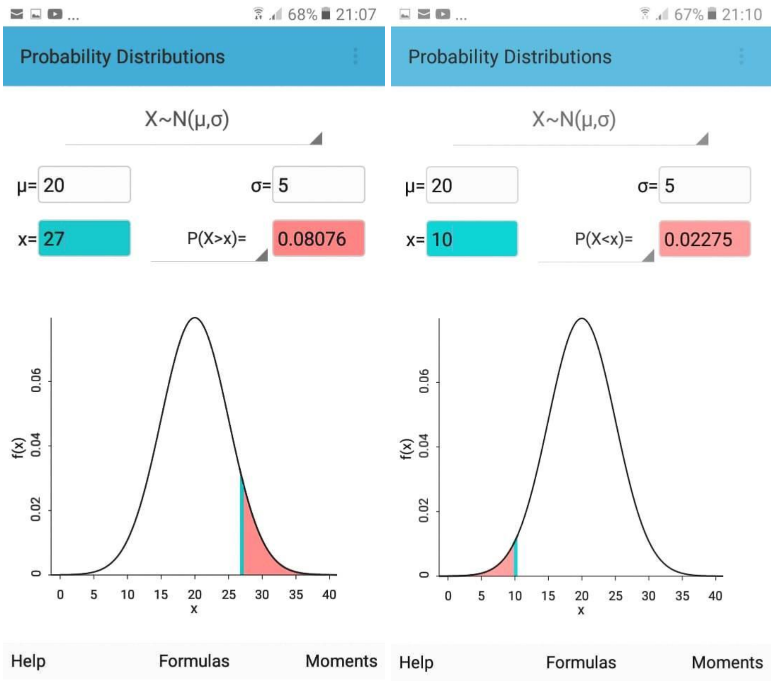
5.- Ya fue comentado.

6.- El puntaje en un test de memoria se distribuye normalmente con una media de 20 puntos y una desviación estándar de 5 puntos en un grupo normativo. Calcule, redondeando a 4 decimales:
a) La probabilidad de que una persona elegida al azar de la población representada por el grupo normativo obtenga una puntuación
i) superior a 27.
ii) inferior a 10.
iii) entre 17 y 23.
b) La mínima puntuación que debería obtener una persona para considerarse dentro del 2% de mayores puntuaciones.

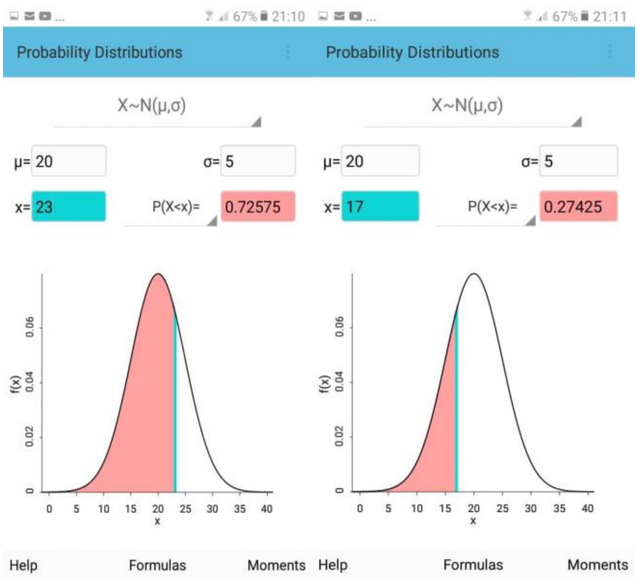
Repaso preliminar: Para realizar este ejercicio hay que tener en cuenta que en la curva normal, las áreas representan probabilidades y en el eje de abscisas, x, se representan los valores de la variable considerada.
Un área izquierda es la probabilidad de encontrar valores inferiores al x_i considerado.
Un área derecha es la probabilidad de encontrar valores superiores al x_i considerado.
El área total debajo de la campana es 1, y se corresponde con el 100% de todas las xi observadas.
Es necesario considerar que, por ejemplo, un 10% de los valores x_i se corresponde con un área de 0,1; y un 1% de los x_i, con el 0,01 del área total.
Si es necesario calcular un área comprendida entre dos x_i, puede ser necesario realizar restas de áreas obtenidas al área total 1.

a) X = Puntaje en un test de memoria.
X ~ N(20,5)
i) P(X>27) = 0,0808 ii) P(X<10) = 0,0228

Realizado con Probability Distribution:



iii) P(17<x<23) = P(X<23) – P(X<17) = 0,7258 – 0,2743 = 0,4515

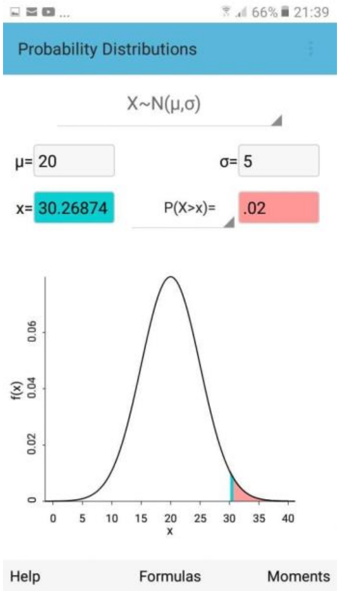


Respuestas: Si una persona es elegida al azar del grupo normativo, las probabilidades de que tenga un puntaje superior a 27, inferior a 10, entre 17 y 23 son respectivamente 0,0808, 0,0228 y 0,4515 respectivamente.

b) La mínima puntuación que debería obtener una persona para considerarse dentro del 2% de mayores puntuaciones.

Se pide el percentil 98 de la distribución; el valor x_0 / P(X> x_0)=0,02. Se completa la ventana de la derecha (roja) con 0,02 y el programa devuelve el valor de x en ventana izquierda (azul).

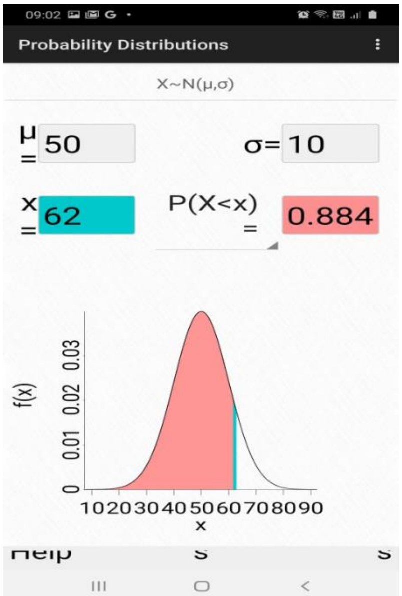
P(x> x_0) = 0,02 $\Rightarrow$ x_0 = 30,2687 puntos.



Redondeando a valores enteros, se necesitan como mínimo 30 puntos para pertenecer al grupo que constituye el 2% de mayores puntuaciones.

7.- El puntaje T en un test de habilidades espaciales tiene distribución normal en la población de interés. Se consideran atípicos los sujetos que se encuentran a más de dos desviaciones estándar de la media. Halle:
a) el rango percentilar de 62.

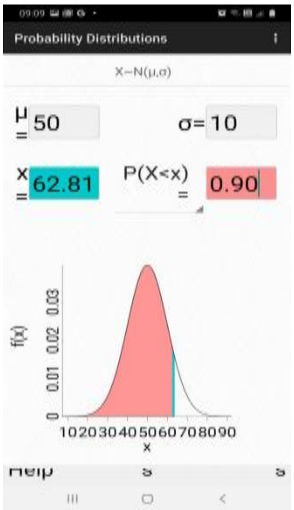
La escala T, utilizada en algunas pruebas psicométricas, tiene media 50 puntos y desviación estándar de 10 puntos. El rango percentilar de un puntaje es el porcentaje de valores que se encuentran por debajo del mismo; en este caso se busca el rango percentilar del valor T=62; se corresponderá al área izquierda de 62:



El rango percentilar de 62 es 88,4%

b) el percentil 90.

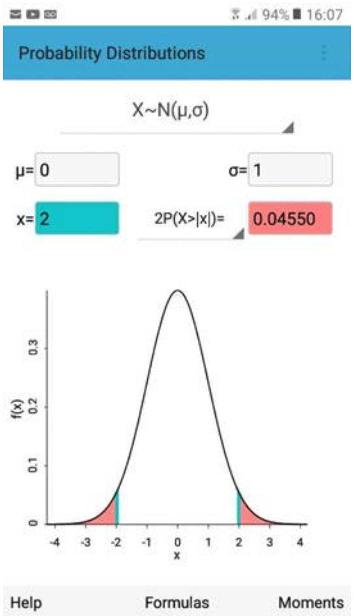
El percentil 90 es el puntaje T que deja por debajo suyo al 90% de las observaciones. Se completa la ventana de la derecha (roja) con 0.90 y devuelve a la izquierda en la ventana azul x = 62.81



El percentil 90 es el puntaje 62,81.

c) el porcentaje de sujetos atípicos en la población.

El puntaje Z expresa el número de desviaciones estándar por encima o por debajo de la media. Como el puntaje, según se informa en el enunciado, se distribuye normalmente y se consideran atípicos los que están a más de 2 desviaciones estándar de la media, basta buscar la probabilidad a la derecha de 2 y a la izquierda de -2 en la aplicación para la Normal Estándar. Ésta se abre por default en la aplicación Probability Distribution. La probabilidad de ambas colas puede hallarse en la tercera opción de la ventana de la derecha.



En la población hay un 4,6% de sujetos atípicos.

d) los umbrales de los puntajes atípicos.

Los umbrales de los puntajes atípicos son -2 y 2 en la escala Z y hay que expresar sus equivalentes en la escala T = 50 + 10Z.

Reemplazando Z por -2 y 2 se obtienen los umbrales: T = 50 - 10.2 = 30 y T = 50 + 10.2 = 70.

Los umbrales inferior y superior son respectivamente 30 y 70.

8.- Las calificaciones en un examen de ingreso puntuado de 0 a 100 se distribuyen normalmente en una población de aspirantes. El área bajo la curva normal entre los puntajes 70 y 80 es 0,15. ¿Qué representa esa área? Redacte una oración expresando ese resultado de manera coloquial como si lo estuviera informando.

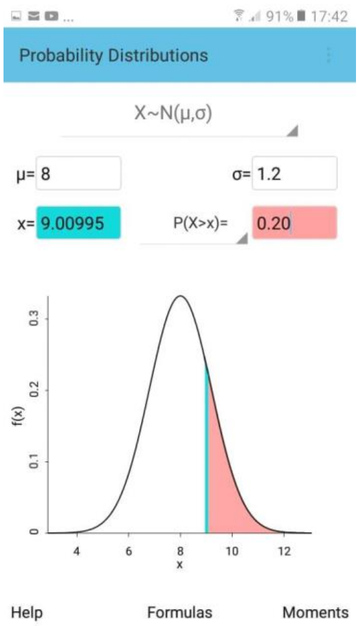
El 15% de los aspirantes que rindieron el examen de ingreso obtuvieron puntajes entre 70 y 80.

Explicación adicional. *La respuesta precedente se fundamenta en el hecho de que las áreas bajo una función de densidad de probabilidad, en este caso la normal, representan las probabilidades de los intervalos sobre los que se levantan las mismas. Aplicado a este ejemplo, eso significa que, si se elige al azar un aspirante de la población en cuestión, la probabilidad de que sea alguien con un puntaje entre 70 y 80 es 0,15. La probabilidad está asociada a la composición porcentual de la población, por lo que se sigue que el 15% de la población de aspirantes tiene un puntaje en dicho intervalo. Esto es congruente con el hecho de que la probabilidad modeliza la frecuencia relativa; así como al multiplicar la frecuencia relativa por 100 se obtiene la frecuencia en términos porcentuales, de la misma manera ocurre con la probabilidad en un modelo.*

9.- El tiempo de reacción ante un estímulo se distribuye normalmente con media 8 segundos y desviación estándar 1,2 segundos. Un psicólogo opina que hay que preocuparse por el grupo de las personas con reacción lenta definido como el de los que tardan tiempos mayores y conforman el 20% del total.
a) ¿Cuál es el tiempo de reacción a partir del cual hay que preocuparse?

Corresponde al punto que deja por encima de sí una probabilidad de 0,20. Es el percentil 80 de la distribución (P_{80}).

P_{80}= 9,009 segundos



Sería preocupante un tiempo de reacción superior a los 9 segundos

b) ¿A cuántas desviaciones estándar, como mínimo, por encima de la media se hallan las personas con reacción lenta?

Las personas con reacción lenta son las que tardan 9,009 segundos o más. Hay que averiguar a cuántos desvíos de la media se encuentra 9,009 seg. Para ello se debe calcular el z=(9,009-8)/1,2 = 0,84
Por tanto, se hallan a por lo menos 0,84 desvíos por encima de la media.

c) Entre 400 personas elegidas al azar de la población, ¿cuántas cabría esperar con reacción lenta?

Si en la muestra se mantiene la misma proporción que en la población, se esperaría que hubiera un 20% (quinta parte) de 400 con reacción lenta; es toes, 80 personas.

10.- Suponga que el tiempo necesario para resolver un test se distribuye normalmente con una media de 50 minutos y una desviación estándar de 8 minutos en la población objetivo. Imagine ahora que dichos parámetros no son conocidos por los psicometristas que diseñan el test y que lo administrarán en una muestra piloto de 36 personas con el fin de estimar el tiempo promedio que llevaría realizarlo en la población general.

a) ¿Cuál es la probabilidad de que la media estimada sea

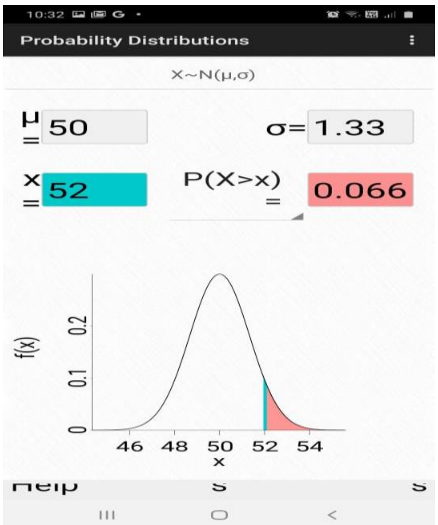
i) mayor a 52?

Para calcular la probabilidad de que la media muestral tome valores mayores a 52 debe considerarse su distribución de probabilidades que, por el Teorema Central del Límite, es aproximadamente normal. Los parámetros característicos de esta distribución, para la media muestral, son la misma media μ de la variable X y su desviación estándar dividida la raíz del tamaño de muestra: $\frac{\sigma}{\sqrt{n}}$. En símbolos:

$$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{X} \approx N\left(50, \frac{8}{\sqrt{36}}\right) = N(50; 1{,}33)$$

Entonces se puede calcular $P(\bar{X} > 52) = 0{,}066$



La probabilidad de que la media muestral tome valores mayores a 52 si la media poblacional fuera 50 y la desviación estándar 8, es 0,066. En otras palabras, la probabilidad de sobreestimar a la media en más de dos puntos es 0,066 (bajo los supuestos mencionados).

ii) mayor a 52 o menor a 48 (error muestral de más de 2 puntos)?

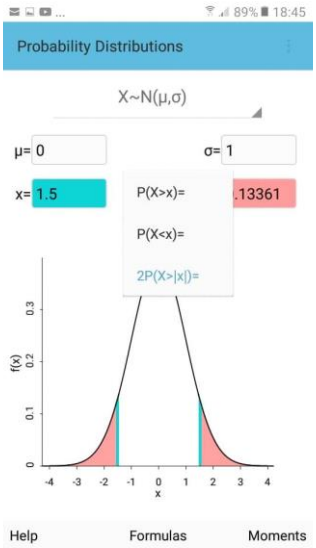
Bajo los mismos supuestos, la probabilidad de hallar medias menores a 48 (por ser la distribución simétrica), será también de 0,066. Por tanto, la probabilidad de que la media muestral sea mayor a 52 o menor a 48 será 0,066*2= 0,132. Es decir, que la probabilidad de que el error muestral de la estimación supere los dos puntos es 0,132.

b) **Conceptualice.** Aun cuando no se tuvieran hipótesis sobre la media ni sobre la desviación estándar de la población, se puede conocer la probabilidad de cometer un error de estimación superior a, por ejemplo, 1,5 desviaciones estándar de la media. Plantee y calcule dicha probabilidad.

Para calcular la probabilidad de cometer un error de estimación superior a 1,5 desvíos de la media, se puede considerar la media muestral tipificada y su distribución, aproximadamente normal estándar:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

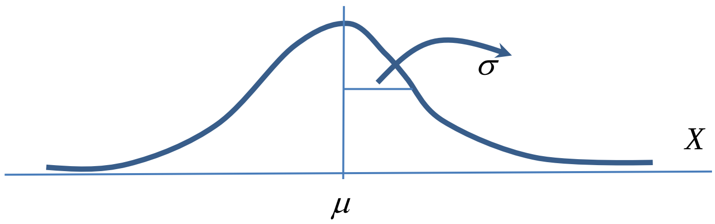
Como el error es la distancia entre el verdadero valor del parámetro y su estimación, un error mayor a 1,5 desvíos de la media corresponde a distancias superiores a 1,5 σ para ambos lados del parámetro (por exceso o por defecto). Por tanto lo que se desea calcular es P(|Z| > 1,5), lo que se halla en la tercera opción de la ventana derecha de Probability Distribution en la Normal Estándar.



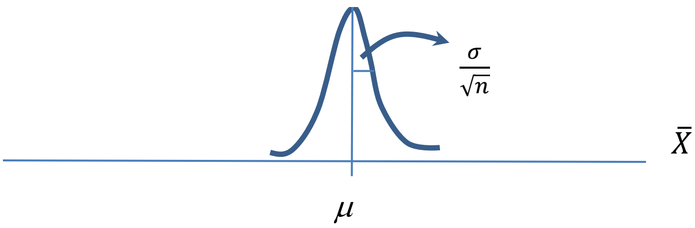
La probabilidad de cometer un error mayor a 1,5 unidades de desviación estándar es 0,1336.

11.- Conceptualice: Deténgase a considerar y comprender que la media muestral, como variable aleatoria, tiene una distribución de probabilidades y que esa distribución es aproximadamente normal (por el Teorema Central del Límite) en torno al parámetro (μ) que desea estimar, cuando el tamaño de muestra es suficientemente grande. Asegúrese de comprender esto porque aquí va lo principal de la fundamentación de los métodos de inferencia estadística que se verán en adelante y es clave para su comprensión. Para ayudarse, entonces, complete la siguiente frase:

Sea *X* una variable aleatoria con esperanza (media) μ y desviación estándar σ. El promedio de los valores de *X* provenientes de muestras de tamaño *n* (media muestral: $\bar{X}$) es aproximadamente normal con esperanza: ……… $\boldsymbol{\mu}$ ………. y desviación estándar ………$\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$…….. para *n* suficientemente grande (*n* > 30). Realice un gráfico aproximado de la distribución de X y de la de $\bar{X}$ en torno a μ. Note que, como es deseable que ocurra, intervalos de valores de la media muestral próximos a la media poblacional tienen más probabilidad de aparecer en una muestra que intervalos de valores lejanos (consideramos para la comparación intervalos de igual longitud). Eso es bueno porque justamente se desea estimar el valor desconocido de μ a partir del valor conocido de una muestra. La diferencia entre ambos es el error de muestreo y es deseable que dicho error sea pequeño.



Distribución de la variable X en la población



Distribución de la media muestral

12.- Las siguientes afirmaciones resultan de teoremas que se necesitan para fundamentar los métodos de inferencia estadística. Las mismas están expresadas en símbolos.

a) Traduzca en lenguaje coloquial las expresiones simbólicas para asegurarse de que comprende el significado de los símbolos.

$$X \sim N(\mu, \sigma) \Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Si una variable X sigue una distribución Normal con media μ y desvío σ, la distribución de la media muestral (de muestras tomadas al azar de la población cada una de ellas de tamaño n), también sigue una distribución normal con una media igual a μ pero con un desvío estándar menor al de la variable X: σ dividida la raíz de *n*. Al estandarizar o tipificar la media, ésta sigue la distribución normal estándar.

Explicaciones adicionales

Cada valor tipificado (z) de la media muestral indica a cuántos desvíos típicos de μ se encuentra la media muestral.

Sobre la base de los gráficos dados en el ejercicio 11, puede observarse que existen mayores semejanzas entre medias muestrales que entre valores de X. O sea, que es más probable que difieran *más entre sí dos individuos cualesquiera tomados al azar que los promedios de dos muestras de individuos tomadas también al azar. En otras palabras, el comportamiento de las muestras es más similar que el comportamiento de dos individuos. Por tanto, puede esperarse que las medias muestrales obtenidas sean más parecidas a μ que los valores individuales de la variable X. Es más factible que un conjunto de individuos promedien una media más cercana a μ que el valor que tuviera cada uno de ellos por separado.*

Como en la investigación de las ciencias empíricas se usan habitualmente las muestras, es de notable interés la distribución de las medias muestrales, ya que al obtener los datos de una de ellas, matemáticamente puede calcularse la probabilidad de que esos datos sean compatibles o no con los poblacionales que se tienen hasta el momento por correctos. Esta es la base matemática del método empleado para validar hipótesis científicas llamado "Prueba de Hipótesis".

Si S es la desviación estándar muestral (correspondiente a dividir la suma de cuadrados de los desvíos por n-1) entonces:

$$X \sim N(\mu, \sigma) => \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

La distribución de la media muestral tipificada, cuando la variable X es normal y la desviación estándar σ se estima con S, sigue la distribución t de Student con n-1 grados de libertad.

Explicación adicional

La t es se aproxima a la z a medida que el tamaño de muestra n es cada vez más grande. Las diferencias entre ambas distribuciones se acentúan para tamaños muestrales pequeños o inferiores a 30.

Tanto Z como t son distribuciones que estandarizan la media muestral indicando a cuántos desvíos se encuentra la media muestral de la μ poblacional.

Si n es grande ($n > 30$), por Teorema de Slutzky,

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$$

La media muestral estandarizada, estimando σ mediante *S*, sigue la distribución aproximadamente normal para valores de n grandes.

Si $\hat{p}$ es la frecuencia relativa o proporción muestral y p es la probabilidad de éxito o proporción poblacional que se estima a través de $\hat{p}$

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx N(0,1) \quad si \quad n\hat{p} \geq 5 \; y \; n(1-\hat{p}) \geq 5$$

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \quad si \quad np \geq 5 \; y \; n(1-p) \geq 5$$

La proporción muestral tipificada tiene distribución aproximadamente normal estándar bajo ciertas condiciones (las desigualdades mencionadas arriba). La diferencia entre ambas expresiones está en el denominador; donde figuran $\hat{p}$ (estimador muestral de p) o p (parámetro poblacional).

Explicaciones adicionales

Cuando la distribución considerada es sobre proporciones muestrales; por ejemplo, proporción de mujeres en una muestra de 35 personas, la "p" es la proporción poblacional, mientras que la p̂ es la proporción muestral. Es de esperar que si la proporción poblacional es "p", la p̂ muestral sea parecida a la poblacional. Cuando se verifican las desigualdades arriba señaladas, en ambos casos la distribución de las tipificaciones de las proporciones muestrales siguen una distribución aproximadamente normal estándar "Z". En la primera expresión se desconoce el error estándar del estimador y se lo estima (en el denominador) con p̂ mientras que en la segunda expresión se supone conocido; en el denominador está el valor del parámetro p y no su estimador. Esta última expresión se utiliza cuando hay algún valor hipotético de p (por ejemplo en una prueba de hipótesis).

Al considerar la tabla de contingencia de una distribución bivariada,

$$\sum_{i=1}^{f}\sum_{j=1}^{c}\frac{(f_{ij}^{o}-f_{ij}^{e})^2}{f_{ij}^{e}} \sim \chi^2_{(f-1)\mathrm{x}(c-1)}$$

Donde las f_{ij} aluden a las frecuencias absolutas observadas y esperadas bajo la hipótesis de independencia en una tabla de contingencia.

La suma de los cuadrados de las diferencias entre las frecuencias conjuntas observadas y las esperadas bajo la hipótesis de independencia tienen distribución Ji-Cuadrado cuyos grados de libertad es el producto entre la cantidad de filas menos 1 y la cantidad de columnas menos 1.

b) Halle los percentiles 5 y 95 de las distribuciones Normal Estándar, t de Student con 9 grados de libertad y χ^2 con 9 grados de libertad. Elija una aplicación apropiada para hacerlo (PQRS, Excel, Probability Distribution, Statistix, Infostat, etc.)

Distribución	Percenti 5	Percentil 95
Normal Estándar	-1,645	1,645
t de Student con 9 gl	-1,833	1,833
χ^2 con 9 gl	3,325	16,919

c) **Conceptualice:** Deténgase a comprender los conceptos de Estadístico, Estimador y Parámetro. Reconozca en todas las expresiones anteriores quiénes son estadísticos, quiénes son estimadores y quiénes parámetros. Para ello complete la siguiente tabla tildando la celda que corresponda. La primera línea está hecha como ejemplo.

NOMBRE	EXPRESIÓN	ESTADÍSTICO	ESTIMADOR	PARÁMETRO
Media poblacional o esperanza	μ			✓
	σ			✓
	$\bar{X}$	✓	✓	
	$\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$	✓		
	$\hat{p}$	✓	✓	
	p			✓
	$\dfrac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$	✓		
	f_{ij}^{o}	✓	✓	
	$\sum_{i=1}^{f}\sum_{j=1}^{c}\frac{(f_{ij}^{o}-f_{ij}^{e})^2}{f_{ij}^{e}}$	✓		

13.- Conceptualice. $\bar{X}$ y μ aluden a la media, σ y S a la desviación estándar, $\hat{p}$ y p a la proporción o frecuencia relativa.

a) ¿Por qué se utilizan dos símbolos diferentes para aludir a cada uno de estos conceptos? Indique cuál es la diferencia conceptual dentro de cada uno de estos pares de símbolos.

La diferencia está dada porque unos refieren a las muestras sobre las que se obtienen, y los otros refieren a las poblaciones de las que se obtienen; los primeros se denominas estadísticos (muestrales) y los segundos parámetros (poblacionales).

b) Asocie la palabra “poblacional” o bien “muestral” a cada uno de estos seis símbolos: $\bar{X}$, μ σ, S, $\hat{p}$ y p. Por ejemplo: $\bar{X} \rightarrow$ muestral.

$\mu \rightarrow$ poblacional σ $\rightarrow$ poblacional S $\rightarrow$ muestral $\hat{p} \rightarrow$ muestral $p \rightarrow$ poblacional

14.- Dé un ejemplo de una población objetivo y de una variable cuantitativa por estudiar en esa población. Indique cuáles son las unidades de análisis. Para esa variable en esa población dé tres ejemplos de parámetros. Luego suponga que extrae una muestra aleatoria de dicha población. Dé tres ejemplos de estimadores de los correspondientes parámetros antes mencionados.

Población objetivo: Niños escolarizados de CABA, de 6to y 7mo grado que recibieron clases virtuales.

Variable cuantitativa: Número de asistencias a las clases virtuales durante abril de 2020.

Unidades de análisis: Cada uno de los niños de 6to y 7mo grado de CABA que recibieron clases virtuales.

Parámetros: 1) Promedio de la cantidad de asistencias a las clases virtuales de los niños de 6to y 7mo grado de CABA durante abril de 2020.

2) Mediana de dicha cantidad de asistencias en la población ya mencionada.

3) Proporción de niños de 6to y 7mo grado de CABA que asistieron a menos de 15 clases en abril de 2020.

Se toma una muestra aleatoria de 600 niños de 6to y 7mo grado de las escuelas públicas y privadas de CABA que impartieron clases virtuales y se examina el registro de asistencia a clases virtuales durante el mes de abril de 2020. Se computa la cantidad total de asistencias durante ese mes.

Estimadores: 1) Promedio de la cantidad de asistencias a clases virtuales en abril de 2020 de los 600 niños seleccionados en la muestra.

2) Mediana de la cantidad de asistencias a clases virtuales en abril de 2020 de los 600 niños seleccionados en la muestra.

3) Cantidad de niños de la muestra que asistieron durante abril de 2020 a menos de 15 clases, dividida 600; es decir la proporción muestral o frecuencia relativa de quienes asistieron a menos de 15 clases.

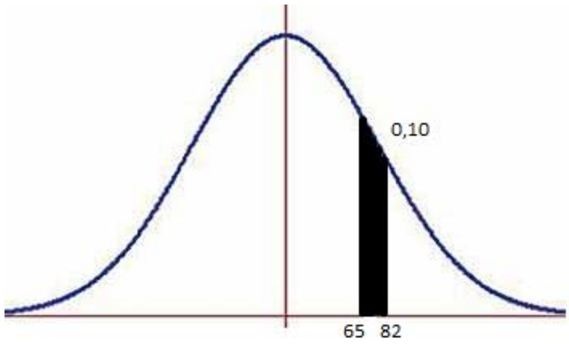
15.- Se denomina error estándar de un estimador a:
a) su desviación estándar.
b) su discrepancia con el parámetro que estima.
c) su probabilidad de distar más de una desviación estándar del parámetro.

16.- Un estadístico es
a) un valor numérico que se calcula a partir de una muestra aleatoria.
b) un valor característico de la población objetivo de la que se saca la muestra.
c) una variable cuyos valores dependen de una muestra aleatoria.

17.- Si z_p y t_p son los percentiles 70 de las distribuciones normal estándar y t de Student entonces
a) $z_p = t_p$
b) $z_p < t_p$
c) $z_p > t_p$

18.- Los puntajes en una escala de inhibición conductual se distribuyen normalmente en un grupo normativo de preescolares con una media de 35 y una desviación estándar de 8. Un establecimiento educativo decide brindar asistencia terapéutica a los niños correspondientes a la última interdecila. ¿Qué puntaje (redondeado) debe tener como mínimo un niño en esta escala para recibir asistencia terapéutica?
a) 90
b) 25
c) 45

19.- Los puntajes en un test de creatividad se distribuyen normalmente en la población de estudiantes de la escuela CREARTE. La región sombreada en la siguiente figura corresponde a un área de 0,10.



¿Cómo se interpreta esa información con respecto a dicha área sombreada?

a) Hay 10 alumnos o más en la escuela CREARTE que tienen puntaje superior a 65 en el test de creatividad.
b) El rango percentilar del puntaje 82 en el test de creatividad es 10 para los alumnos de la escuela CREARTE.
c) El 10% de los alumnos de la escuela CREARTE tiene un puntaje en el test de creatividad entre 65 y 82.