ESTADÍSTICA Y PROBABILIDAD II

Programa de apoyo al egreso

Autores

Álvarez García Carlos Alberto
Padilla Alemán David Bryan
Rivera Vargas Héctor Gabriel
Zecua Fernández Aquilino
Zepeda Aguillón Víctor Hugo

Coordinador

Rivera Vargas Héctor Gabriel

2022

COLEGIO DE CIENCIAS Y HUMANIDADES PLANTEL NAUCALPAN

Contenido

Presentación

El presente material tiene como finalidad ser usado en los grupos de Estadística y Probabilidad II del Programa de Apoyo y Egreso, tratando de apoyar al profesor y a los estudiantes por medio de ejemplos, ejercicios y algunos videos, sin que con esto, se pretenda sustituir al profesor ni a la vasta gama de materiales didácticos y bibliográficos que profundizan en los temas.

El material está basado en el Programa de Estudio de Estadística y Probabilidad I y II, por lo tanto, se organiza en 3 unidades: Unidad 1 Modelos de probabilidad y sus aplicaciones, Unidad 2 Estimadores e introducción a la inferencia estadística y Unidad 3 Inferencia estadística

Los comentarios, criticas, aportaciones o sugerencias respecto al material, ayudarán en gran medida a mejorar.

Esperamos que este sencillo, pero esmerado trabajo, sea un producto que sirva tanto a nuestros compañeros como a los estudiantes.

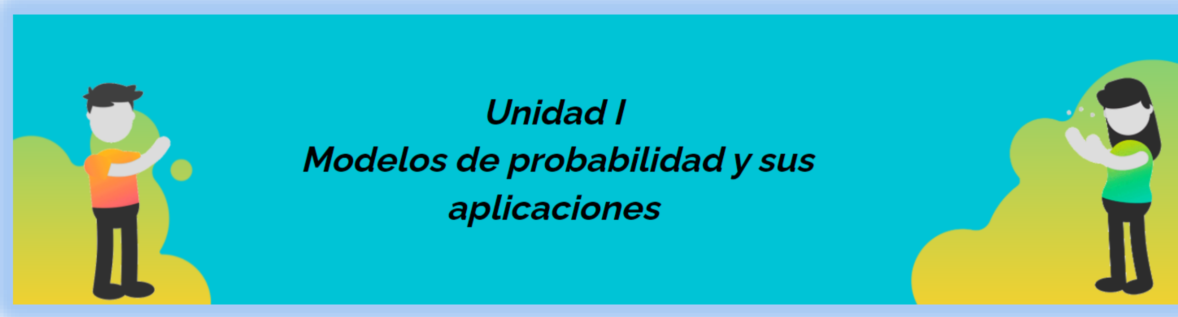
Los autores

Contenido Temático por unidades

UNIDAD	NOMBRE	HORAS
I	Modelos de probabilidad y sus aplicaciones	24
II	Estimadores e introducción a la inferencia estadística	14
III	Inferencia estadística	26

Bibliografía sugerida

- Castillo Juana, Estadística Inferencial Básica, Iberoamérica, 1998
- Johnson R, Estadística Elemental, Iberoamérica, 1980
- Stevenson, W., Estadística y Probabilidad para Administración y Economía, Harla, 1994
- Chao, L., Introducción a la Estadística, CECSA, 1987
- Mendenhall, W., Estadística Matemática con Aplicaciones, Iberoamérica, 1986
- Christensen H. B., Estadística Paso a paso, Trilla



Unidad I Modelos de probabilidad y sus aplicaciones

Aprendizajes

- Identifica el concepto de variable aleatoria.
- Diferencia entre variables aleatorias discretas y continuas.
- Examina los conceptos de distribución de probabilidad, esperanza matemática y desviación estándar.
- Construye la distribución de probabilidad para una variable aleatoria. discreta y su modelo de simulación, físico o por medio de la computadora.
- Identifica las características de un proceso binomial.
- Construye el modelo para la distribución binomial, apoyándose en la simulación física o con la computadora.
- Aplica el modelo binomial, su valor esperado y su desviación estándar a fenómenos contextualizados que se ajusten a este modelo, interpretando los resultados, obtenidos desde la distribución o de tablas.
- Deduce que, en el caso de variables aleatorias continuas, la probabilidad debe calcularse para valores dentro de un intervalo
- Esboza la curva de densidad para una variable aleatoria aproximadamente normal, a partir de la suavización del polígono de frecuencias de un ejemplo sencillo.
- Deduce el proceso de estandarización de la distribución normal, aplicando la Regla Empírica, dentro de un problema contextualizado.

- Utiliza la tabla para valores bajo la curva de la distribución normal estandarizada como el recurso para el cálculo de probabilidades o de valores z para dicha distribución.
- Calcula probabilidades por medio de la distribución normal estandarizada dentro de problemas contextualizados, interpretando los resultados.
- Contrasta la gráfica para una situación de comportamiento aproximadamente normal con su correspondiente gráfica en el modelo estandarizado.
- Concluye que una variable aleatoria, discreta o continua, puede describirse y analizarse por su tendencia, dispersión y distribución.

Variable aleatoria (v. a.)

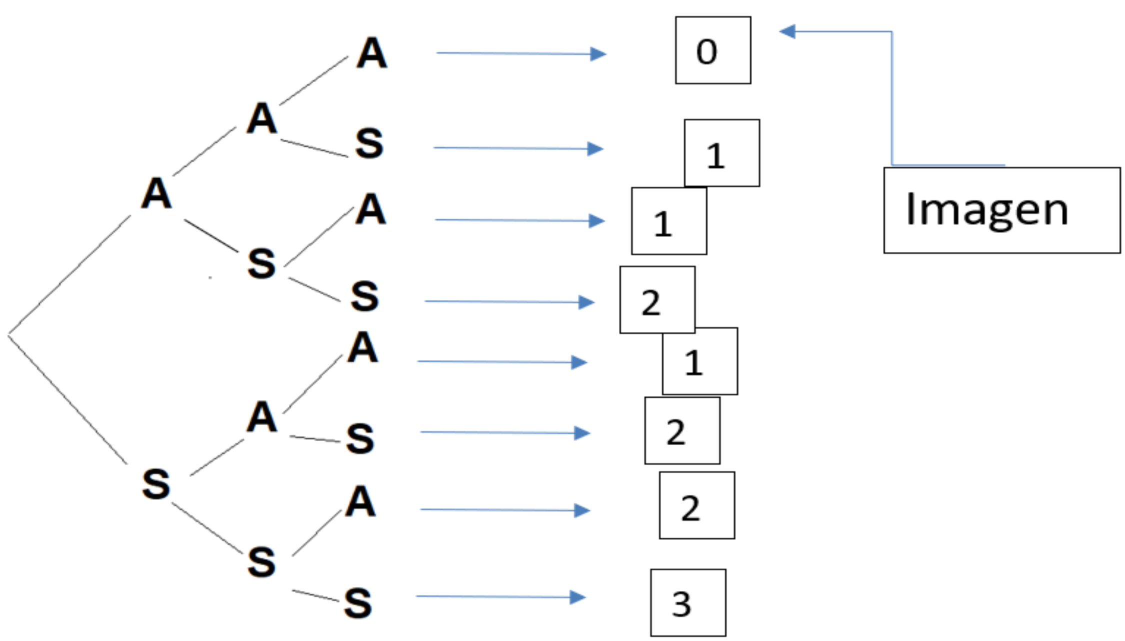
Una v. a. es una función cuyos valores son números reales, definida en un espacio muestral.

Dicho de otra manera, una v. a. es una variable que toma un valor único para cada uno de los resultados en el espacio muestral de un experimento probabilístico. Recordemos que es el espacio muestral es el conjunto de posibles resultados al efectuarse un experimento probabilístico.

Ejemplo

Se lanzan 3 monedas y la v. a. X se define como el número de soles que se observan al terminar el experimento. ¿Qué valores puede tomar X y cuál es el más probable?

Representemos por medio de un diagrama de árbol:



$S = \{AAA, AAS, ASA, ASS, SAA, SAS, SSA\ Y\ SSS\}$

¿Qué valores puede tomar X? $0, 1, 2\ y\ 3$

¿Qué valor es más probable? $1\ y\ 2$

Variable aleatoria discreta (v. a. d.)

Una v. a. X se dice discreta si solamente puede tomar un conjunto numerable de valores. La distribución de probabilidad (d. de p.) para una v. a. X se puede representar por una fórmula, una tabla o una gráfica que indique las probabilidades $p(x)$ correspondientes a cada valor x de X.

Ejemplo

Se escogen tres artículos de un proceso de manufactura y se desea saber el número de artículos defectuosos que se escogieron, ¿qué valores puede tomar la v. a. X? Solución: $0, 1, 2$ o 3 artículos defectuosos.

Variable aleatoria continua (v. a. c.)

Ejemplo

En una fábrica de chocolates se seleccionaron 100 de sus productos para sumar sus respectivos pesos en gramos, obtener una media y con esto llevar un cierto control de calidad. ¿Cuáles son los posibles valores de la v. a. de interés?

Solución. Los posibles valores de la v. a. X son todas las $x > 0$.

Esta es una v. a. c. Si las observaciones para una v. a. se obtienen como resultado de medir, tal y como sucede con las estaturas de los estudiantes de una escuela, se trata de una v. a. c. Esto a diferencia de los valores de las observaciones para una v. a. d., los cuales se obtiene como resultado de contar.

Distribución de probabilidad (d. de p.)

Es una distribución de probabilidades cada una de las cuales está asociada con uno de los posibles valores diferentes de la v. a. Toda d. de p. de v. a. d., independientemente de la forma en la que se le represente, cumple con las siguientes propiedades: $\boldsymbol{0 \leq p(x) \leq 1}$ **y que para toda** $\boldsymbol{x}$ $\sum \boldsymbol{p(x)} = \boldsymbol{1}$, donde la suma se toma sobre todos valores de y con probabilidad diferente de cero.

Ejemplo

Verifica si $M(x)=\frac{6-|x-7|}{36}$ para $x = 2,3,4,5,6,7,8,9,10,11,12$ representa una función de probabilidad (f. de p.)

Obtengamos

$$M(2)=\frac{6-|2-7|}{36}=\frac{6-|-5|}{36}=\frac{6-5}{36}=\frac{1}{36}$$

$$M(3)=\frac{6-|3-7|}{36}=\frac{6-|-4|}{36}=\frac{6-4}{36}=\frac{2}{36}$$

$$M(4)=\frac{6-|4-7|}{36}=\frac{6-|-3|}{36}=\frac{6-3}{36}=\frac{3}{36}$$

$$M(5)=\frac{6-|5-7|}{36}=\frac{6-|-2|}{36}=\frac{6-2}{36}=\frac{4}{36}$$

$$M(6)=\frac{6-|6-7|}{36}=\frac{6-|-1|}{36}=\frac{6-1}{36}=\frac{5}{36}$$

$$M(7)=\frac{6-|7-7|}{36}=\frac{6-|0|}{36}=\frac{6-0}{36}=\frac{6}{36}$$

$$M(8)=\frac{6-|8-7|}{36}=\frac{6-|1|}{36}=\frac{6-1}{36}=\frac{5}{36}$$

$$M(9)=\frac{6-|9-7|}{36}=\frac{6-|2|}{36}=\frac{6-2}{36}=\frac{4}{36}$$

$$M(10)=\frac{6-|10-7|}{36}=\frac{6-|3|}{36}=\frac{6-3}{36}=\frac{3}{36}$$

$$M(11)=\frac{6-|11-7|}{36}=\frac{6-|4|}{36}=\frac{6-4}{36}=\frac{2}{36}$$

$$M(12)=\frac{6-|12-7|}{36}=\frac{6-|5|}{36}=\frac{6-5}{36}=\frac{1}{36}$$

Sumando la probabilidad de cada posible valor tenemos $\frac{36}{36} = 1$, por lo tanto, se verifica que $M(x) = \frac{6-|x-7|}{36}$ para $x = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ si representa una función de probabilidad (f. de p.).

Ejemplo. Listado

Encuentra la d. de p. para la v. a. definida como el número de canciones de jazz, cuando se eligen al azar 2 de una colección que consta de 4 canciones de jazz, 2 canciones de rock y 1 de reggaetón. Expresa la d. de p. través de una tabla usando:

a) una lista de todas las posibilidades
b) un diagrama de árbol
c) fórmula

Solución a)

La v. a. se define como el # canciones jazz en la muestra aleatoria y la llamaremos X. Se seleccionan a 2 aleatoriamente. Hay 4 de jazz (J) y 3 que no son de jazz (N), en total son 7. Una canción la definimos con el número 1, otra como 2 y así cada una de las 7. Además, consideremos que la canción 1, 2, 3 y 4 son de jazz.

Por lo tanto, tenemos la siguiente tabla conformada por los elementos del espacio muestral y su imagen de acuerdo con la definición de la v. a.:

S	Imagen
1,2	2
1,3	2
1,4	2
1,5	1
1,6	1
1,7	1
2,3	2
2,4	2
2,5	1
2,6	1
2,7	1
3,4	2
3,5	1
3,6	1
3,7	1
4,5	1
4,6	1

4,7	1
5,6	0

→

5,7	0
6,7	0

La distribución de probabilidad de la v. a. es

X	0	1	2
$P(X=x)$	$3/21$	$12/21$	$6/21$

Ejemplo. Fórmula

Solución c)

Consideremos que hay 4 de jazz (J) y 3 que no son de jazz (N), en total son 7

Se seleccionan a 2 aleatoriamente y la v. a. X se define como el # canciones jazz seleccionados, entonces

$$P(X=0)=\frac{C(4,0)C(3,2)}{C(7,2)}=\frac{(1)(3)}{21}=\frac{3}{21}$$

$$P(X=1)=\frac{C(4,1)C(3,1)}{C(7,2)}=\frac{(4)(3)}{21}=\frac{12}{21}$$

$$P(X=2)=\frac{C(4,2)C(3,0)}{C(7,2)}=\frac{(6)(1)}{21}=\frac{6}{21}$$

La distribución de probabilidad de la v. a. es

X	0	1	2
$P(X=x)$	$3/21$	$12/21$	$6/21$

Videos

https://www.youtube.com/watch?v=U5QyGHq-MRU pertenece a un video donde se explica la obtención de la distribución de probabilidad por medio del listado.

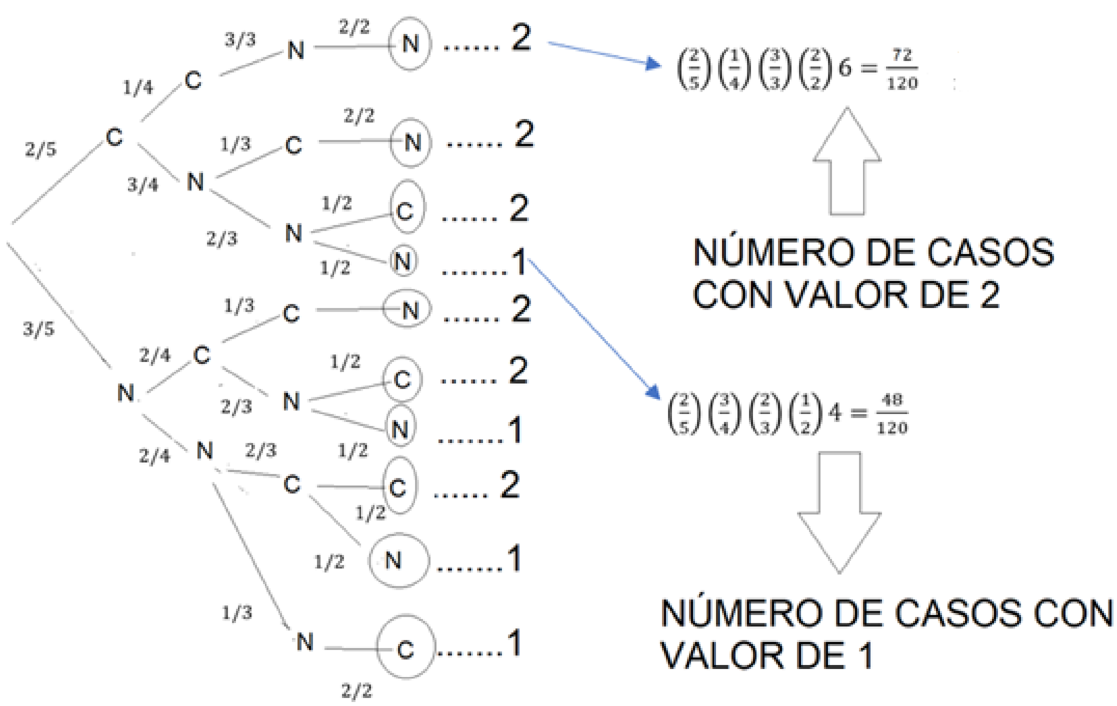
https://www.youtube.com/watch?v=Mq8xUSpxOIE pertenece a un video donde se explica la obtención de la distribución de probabilidad por medio del diagrama de árbol.

https://www.youtube.com/watch?v=xX_JkhKKwNc pertenece a un video donde se explica la obtención de la distribución de probabilidad por medio de la fórmula.

Ejemplo. Diagrama de árbol

En una farmacia se tienen 5 pepto en venta de los cuales 2 ya caducaron. El farmacéutico selecciona a 4 de manera aleatoria para vendérselos a una persona que los requiere. Si se define a la v. a. X como el número de Pepto caducados adquiridos por la persona. Obtenga la d. de p. de v. a. X usando un diagrama de árbol.

Solución b) Muestra aleatoria de 4, 2 caducos, 3 NO caducos, v. a. X se define como el # de pepto caducos



Simplificando cada fracción, la distribución de probabilidad de la v. a. X es

X	1	2
$P(X = x)$	$2/5$	$3/5$

Video

https://www.youtube.com/watch?v=3QPWHErrwmw pertenece a un video donde se explica la obtención de la distribución de probabilidad por medio del diagrama de árbol.

Ejercicios de distribuciones de probabilidad de v. a. d.

1. Verifica si las siguientes expresiones son funciones de probabilidad (f. de p.)

   a) $P(x) = \dfrac{5-x}{10}$ para $x = 1, 2, 3, 4.$

   b) $Q(x) = \dfrac{x^2-1}{50}$ para $x = 2, 3, 4, 5.$

Video

Puedes apoyarte con https://www.youtube.com/watch?v=jWyfH3b91eI

2. Encuentre el valor de k para que la siguiente tabla represente una d. de p. de la v. a. d. X

X	1	3	5
$P(X)$	k	$2k$	$0.5k$

3. Se tiene una urna con 11 papelitos y se selecciona 1 papelito de manera aleatoria: Se define la v. a. X de la siguiente manera:

$$X = \begin{cases} \text{multiplicar por 2 al número obtenido en el papelito si es impar y primo} \\ \text{restar 4 al número obtenido en el papelito si es solamente impar} \\ \text{multiplicar por 1.5 al número obtenido en el papelito en cualquier otro caso} \end{cases}$$

Obtén la d. de p. de X .

4. Se selecciona de manera aleatoria una pelota de una urna que contiene 1 pelota verde con el número 2, 1 pelota azul con el número 2 y una pelota roja con el número 3. Se define a la v. a. X como: sumar 2 al número de la pelota seleccionada si ésta es verde, en caso contrario, multiplicar por 2 al número de la pelota seleccionada. Obtenga la distribución de probabilidad de la v. a. X construyendo el modelo de simulación físico o por medio de la computadora.

5. De 6 estudiantes, 3 tienen habilidades artísticas sobresalientes. Si se escogen al azar 2 estudiantes de los 6 y se define la v. a. X como el número de estudiantes seleccionados con habilidades artísticas sobresalientes. Obtén la distribución de probabilidad de X.

6. En un café-internet tienen en servicio 2 computadoras ACER, 2 HP y 1 Lenovo. Si se seleccionan de manera aleatoria a 4 de esas 5 máquinas y se define la v. a. X como el número de computadoras ACER seleccionadas obtenga la d. de p. de X. Resuélvelo por:
   a) listado de posibilidades
   b) fórmula
   c) el modelo de simulación físico o computacional

7. Profeco realizará un muestreo aleatorio de 6 productos de un lote de 50 (45 se encuentran en buen estado y 5 no). Si se define la v. a. X como el número de artículos seleccionados en buen estado obtenga la distribución de probabilidad de X.

8. Sea X una v. a. d. Determinar el valor de k para que la función $p(x) = \frac{k}{x}, \quad x = 1, 2, 3, 4$, sea la f. de p. de X. Representar por medio de una tabla la d. de p.

9. Una compañía fabrica agujas para la inyección de insulina y las empaca en cajas de 100 unidades. Durante algunos años se han hecho muestreos de estas cajas, por lo que se sabe que el 90% contiene agujas no defectuosas, 7% contiene exactamente una aguja defectuosa, y 3% exactamente dos con defecto. Tomando como base esta información, ¿cuál es la distribución de probabilidades para X, donde X indica el número de agujas defectuosas por caja?

10. De una caja que contiene $3\ monedas\ de\ \$5, 2\ de\ \$1\ y\ 2\ de\ \$0.50$ se seleccionan al azar y **sin reemplazo** a tres de ellas. Encuentra la d. de p. para el total T de dinero de las tres monedas. Utiliza un listado de todas las posibilidades. Además, construye el modelo de simulación físico o computacional para simular la situación y comparar con la d. de p. obtenida con el listado.

Esperanza matemática y varianza de una v. a. d.

El valor esperado de una v. a. d. se obtiene sumando los diferentes valores multiplicados por sus probabilidades correspondientes. En términos simbólicos:

$$\mu = \mathrm{E(X)} = \sum \mathrm{X_i P(X_i)}$$

La desviación estándar es: $\sigma = \sqrt{\mathrm{E(X^2)} - \mu^2}$

Ejemplo

Se tiene la siguiente distribución de probabilidad

X	5	7.5	10
$P(X)$	0.40	$\square$	0.50

Obtenga $\mathrm{E(X)}$ y σ.

La probabilidad de 7.5 es 0.10, ya que la tabla representa a una distribución de probabilidad, entonces:

$E(X) = (5)(0.40) + (7.5)(0.10) + (10)(0.5) = 7.75$

$\sigma^2 = (5)^2(0.40) + (7.5)^2(0.10) + (10)^2(0.50) - (7.75)^2 = 5.5625$

entonces $\sigma = 2.3585$

Ejercicios de esperanza matemática, varianza y desviación estándar

1. Se tiene la siguiente distribución de probabilidad:

X	1	3	5
$P(X)$	$\square$	0.25	0.50

Obtenga la esperanza matemática de X y su desviación estándar.

Video Puedes apoyarte con

https://www.youtube.com/watch?v=rREKYlZqxu0

2. Un señor apuesta \$5 pesos al número 3 al lanzar un dado corriente, y pierde \$2 en caso de no observar el número 3 en el dado, ¿cuál es el valor esperado del juego? ¿Y cuál es la varianza de la v. a. X?

Video Puedes apoyarte con

https://www.youtube.com/watch?v=clNzNXMPnl0

3. Al invertir en unas acciones financieras, una persona puede lograr una ganancia de \$4000 en un año con probabilidad de 0.3 o bien tener una pérdida de \$1000 con una probabilidad de 0.7. ¿Cuál sería la ganancia esperada de esta persona? Solución: $\$500$.

4. A un trabajador de un establecimiento de lavado de automóviles se le paga según el número de autos que entran al servicio. Suponga que las probabilidades son $1/12, 1/12, 1/4, 1/4, 1/6$ y $1/6$, respectivamente, de que el trabajador reciba $\$70, \$90, \$110, \$130, \$150$ ó $\$170$ entre las 16:00 y las 17:00 horas, en cualquier viernes de la semana. Determina las ganancias del trabajador para este periodo en particular.

5. Una urna contiene 5 papeletas que no pueden distinguirse. Tres de ellas están marcadas con $2 y el resto con $4 cada una. Un jugador saca al azar 2 papeletas de la urna sin reponerlas y gana una cantidad igual a la suma de las marcas en las dos papeletas que ha sacado. Si el costo del juego es de $5.60, ¿se trata de un juego justo?

Distribución Bernoulli (extraído de Chao, Introducción a la estadística)

Los experimentos que contienen dos posibles resultados se denominan **ensayos de Bernoulli**, generalmente se denominan a uno de los dos posibles resultados de un ensayo de Bernoulli **éxito** y al otro **fracaso**. Estos dos términos se utilizan sólo para identificar los resultados y no implican que un resultado sea más deseable que el otro. A cualquiera de los dos posibles resultados puede denominársele éxito y al otro fracaso.

Se acostumbra a asignar al valor de 1 al evento denominado éxito y 0 al evento denominado fracaso. Sea ***X*** una v. a. de Bernoulli que tiene los posibles valores de *0* y *1*. Además, sea $\boldsymbol{p}$ la probabilidad de éxito y $(\mathbf{1}-\boldsymbol{p}) = \boldsymbol{q}$ la probabilidad de fracaso. La distribución de probabilidad de este ensayo es la siguiente:

X	***éxito***	***fracaso***
p(X)	***p***	***q***

Ejemplo

En una compañía dedicada a la investigación, 75% de los empleados tienen grado universitario y el 25% no. Sea W la v. a. que toma el valor de 1 cuando un empleado seleccionado aleatoriamente tiene grado universitario y 0 cuando el empleado no lo tiene. Representa la distribución por medio de una tabla:

W	***1***	***0***
p(W)	***0.75***	***0.25***

Distribución binomial.

El modelo probabilístico de Bernoulli constituye la base de la distribución binomial. Si en un experimento binomial $\boldsymbol{p}$ es la probabilidad de éxito y $\boldsymbol{q}$ la de fracaso en un ensayo simple, entonces la probabilidad $\boldsymbol{P(x)}$ de que haya exactamente $\boldsymbol{x}$ éxitos en $\boldsymbol{n}$ ensayos es:

$$\boldsymbol{P(x) = C(n,x)p^x q^{n-x}} \textbf{ para } \boldsymbol{x = 0, 1, 2, 3, \dots, n}$$

Propiedades de la distribución binomial

1. Cada ensayo tiene dos resultados posibles (éxito, fracaso).
2. Hay n ensayos Bernoulli independientes repetidos.
3. $\boldsymbol{P(éxito) = p}$ $\boldsymbol{P(fracaso) = q}$**;** las cuales se mantienen constantes en cada ensayo Bernoulli.
4. La v. a. binomial $\boldsymbol{X}$ es el número de éxitos, por lo tanto $\boldsymbol{x}$ puede tomar cualquier valor entero entre $\boldsymbol{0}$ y $\boldsymbol{n}$**.**

La media y la desviación estándar de una distribución de probabilidad binomial teórica puede obtenerse utilizando las dos fórmulas siguientes:

$$\boldsymbol{\mu = np}$$

$$\boldsymbol{\sigma = \sqrt{npq}}$$

Video Para resolver los siguientes ejercicios puedes apoyarte con

https://www.youtube.com/watch?v=tX90Qw60eN0

https://www.youtube.com/watch?v=685bM_4gvgA

https://www.youtube.com/watch?v=pjnypN5LYmY

Ejercicios distribución binomial

1. Una prueba contiene 20 preguntas de verdadero-falso. Si un estudiante contesta las preguntas adivinando, ¿cuál es la probabilidad de que conteste correctamente a 7 preguntas?

2. Cierta medicamento contra COVID19 tiene una efectividad del 60%; es decir, de cada 100 pacientes que la toman, 60 se curan. Sea ***X*** el número de pacientes curados en una muestra de 30 pacientes que han tomado el medicamento. Obténgase la probabilidad de que
   a) por lo menos 1 se cure
   b) 2 o menos pacientes se curen
   c) 29 o menos pacientes NO se curen
   d) ninguno se cure

3. Supóngase que diez aparatos de radar están operando independientemente uno del otro, y que la probabilidad que cualquier radar detecte un cohete enemigo es de 0.80. ¿Cuál es la probabilidad de que 9 aparatos de radar detecten el cohete?

4. Si se sabe que el 90% de los estudiantes que toman el curso de E. y P. II aprueban, ¿Cuál es la probabilidad de que por lo menos 1 estudiantes en una clase de 15 <u>reprueben</u> el curso?

5. Supóngase que el 10% de los reproductores Blue Ray de cierta marca fallarán antes de que expire su garantía. Obténgase la probabilidad de que 30 reproductores Blue Ray vendidos, 15 fallen antes de que termine su garantía.

6. n
SOL hacia arriba es de 0.7. Sea X el número de SOLES obtenidos en 20 tiradas de esta moneda. Obténgase las siguientes probabilidades:

$$a. P(X = 10)$$

$$b. P(X \leq 19)$$

7. Se afirma que un procedimiento terapéutico nuevo es exitoso en el 80% de las veces. Si la terapia se realiza 5 veces y si suponemos que los resultados son independientes entre sí,
   a) ¿Cuál es la probabilidad de que las 5 terapias son exitosas?
   b) ¿Cuál es la probabilidad de que por lo menos 1 **no** sea exitosa?

Distribución Normal

"Es una distribución probabilística para una v. a. c., la cual tiene simetría perfecta, forma de campana unimodal. La media, la moda y la mediana de la distribución son todas iguales y están localizadas al centro de la distribución." (Christensen, 1983, p. 242).

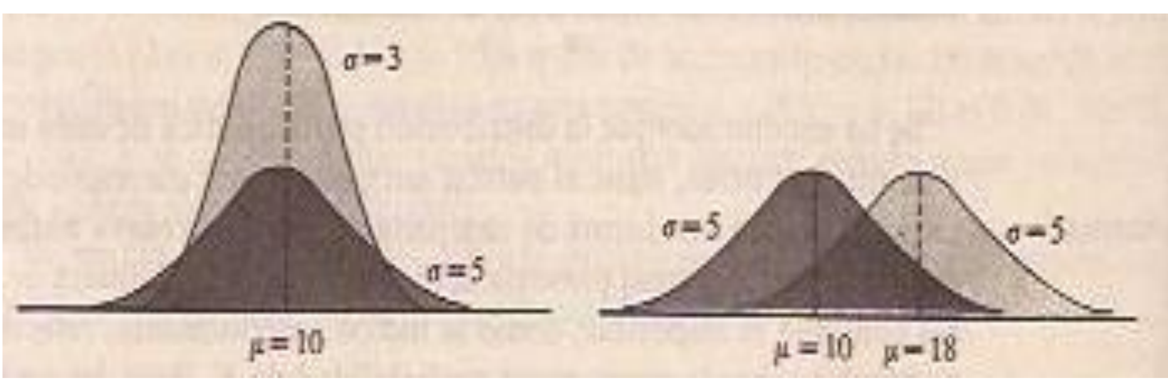
Se ha encontrado que la distribución probabilística de muchas variables sigue el patrón normal, lo cual corresponde gráficamente a una curva simétrica con forma de campana denominada curva normal.

Debido a que el hecho de que la curva normal presenta la distribución probabilística de una v. a. c., es imposible referirse a algún punto en particular sobre la curva como probabilidad de X. Para determinar probabilidades, es necesario hacer referencia a intervalos, tales como el intervalo de a y b.

El área sombreada bajo la curva proporciona la probabilidad de que la v. a. tome cualquier valor entre a y b, y el área bajo la curva es igual a 1.

La curva de la distribución normal es simétrica con respecto a su media (μ). Teóricamente, las colas del lado izquierdo y derecho de la curva se extienden hasta $-\infty$ y $+\infty$, respectivamente.

La distribución normal tiene dos parámetros: la media (μ) y la desviación estándar (σ). Estos dos parámetros determinan la forma específica y ubicación de la curva normal.



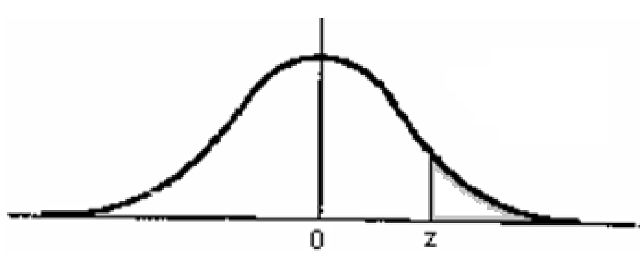
Importancia de la distribución normal

“La distribución normal es sumamente importante debido a que puede utilizarse como una aproximación para cierto número de otras distribuciones; por ejemplo, la distribución binomial se acerca al patrón de la distribución normal conforme n se vuelve grande. Otra razón para la importancia de la distribución normal estriba en el hecho de que numerosas variables aleatorias parecen seguir un patrón de distribución que es semejante a la curva normal. Entre estas variables aleatorias se encuentran las mediciones de pesos, alturas y duración en servicio de varios artículos de consumo.

La distribución de la media muestral es aproximadamente normal sin importar la forma de la distribución de la población, siempre y cuando el tamaño de la muestra sea grande. Esta interesante idea se conoce como el teorema central de límite. Este es de primordial importancia en la estadística inferencial debido a que un tamaño de muestra que se considere suficientemente grande para aplicar la normalidad a la distribución muestral de la media es bastante fácil de lograrse en la mayoría de las situaciones prácticas.” (Chao, 1989, p. 214-215).

Distribución normal estándar.

“Si estandarizamos todas las mediciones en una distribución normal que tiene una media μ y una desviación estándar σ, llamamos a la distribución resultante distribución normal estándar. Esta tiene una media igual a cero y una varianza y desviación estándar iguales a 1. En otras palabras, si Y es una v. a. distribuida normalmente, entonces $z = \frac{y-\mu}{\sigma}$ que está también normalmente distribuida, con una media 0 y una varianza igual a 1. (Christensen, 1983, p. 243). La gráfica de la distribución normal estándar es la siguiente:

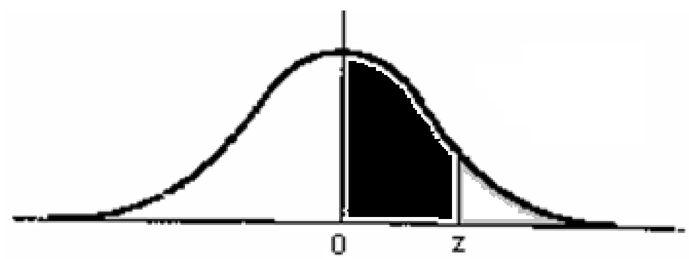


Obtención de áreas bajo la curva normal estándar

Para obtener áreas bajo la curva normal estándar, debes considerar que:

- el área total bajo la curva es igual a 1
- existe simetría con respecto a la media, por lo tanto 0.5 del lado izquierdo y 0.5 del lado derecho

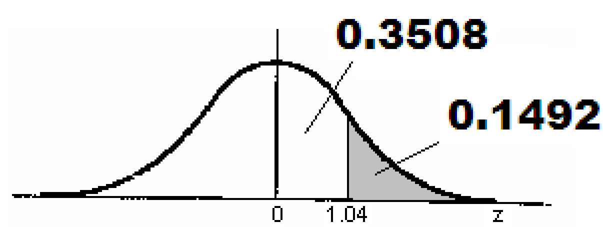
La tabla que se utilizará proporciona áreas de 0 a un valor específico z, esto es:



Ejemplos

1. Obtener el área bajo la curva normal estándar a la derecha de $z = 1.04$

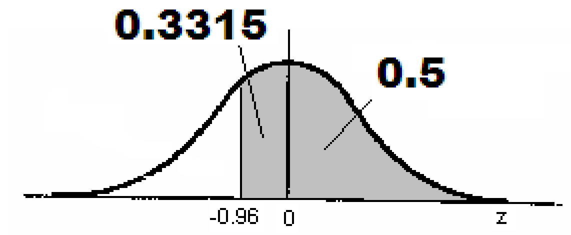
   Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 1.04$, teniendo un valor de 0.3508



   Entonces, el área a la derecha de $z = 1.04$ es $0.5 - 0.3508 = 0.1492$, la cual puede representarse como $P\ (Z > 1.04) = 0.1492$

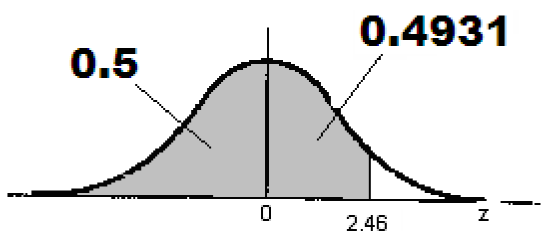
2. Obtener el área bajo la curva normal estándar a la derecha de $z = -0.96$

   Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 0.96$, teniendo un valor de 0.3315



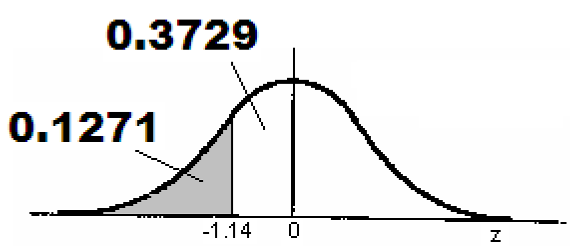
Entonces, el área a la derecha de $z = -0.96$ es $0.5 + 0.3315 = 0.6315$, la cual puede representarse como $P\ (Z\ >\ -0.96)\ =\ 0.8315$

3. Obtener el área bajo la curva normal estándar a la izquierda de $z\ =\ 2.46$
Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 2.46$, teniendo un valor de 0.4931



Entonces, el área a la izquierda de $z = 2.46$ es $0.5 + 0.4931 = 0.9931$, la cual puede representarse como $P\ (Z <\ 2.46)\ =\ 0.9931$

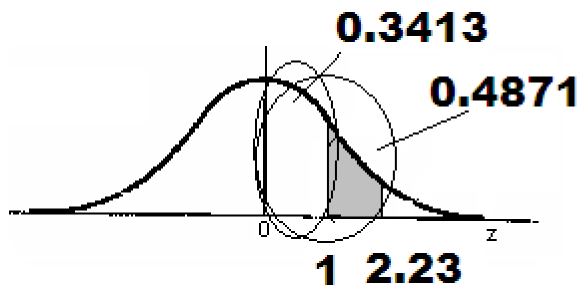
4. Obtener el área bajo la curva normal estándar a la izquierda de $z\ =\ -1.14$
Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 1.14$, teniendo un valor de 0.3729



Entonces, el área a la izquierda de $z = -1.14$ es $0.5 - 0.3729 = 0.1271$, la cual puede representarse como $P\ (Z < \ -1.14)\ =\ 0.1271$

5. Obtener el área bajo la curva normal estándar entre $z = 1$ y $z = 2.23$

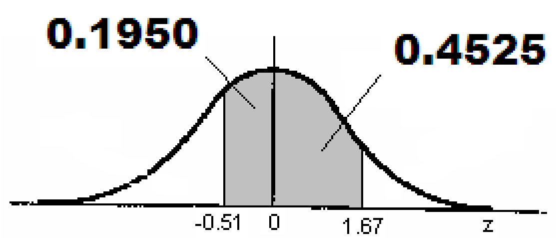
Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 1$, teniendo un valor de 0.3413. También debemos obtener el área bajo la curva entre 0 y $z = 2.23$, cuyo valor es 0.4871. Observa la gráfica:



Entonces, el área entre $z = 1$ y $z = 2.23$ es $0.4871 - 0.3413 = 0.1458$, la cual puede representarse como $P\ (1 < Z < \ 2.23) \ = \ 0.1458$

6. Obtener el área bajo la curva normal estándar entre $z = \ -0.51$ y $z \ = \ 1.67$

Solución. La tabla nos proporciona el área bajo la curva entre 0 y $z = 0.51$, teniendo un valor de 0.1950.



Entonces, el área entre $z = -0.51$ y $z = 1.67$ es $0.1950 + .4525 = 0.6475$, la cual puede representarse como $P\ (-0.51 < Z < \ 1.67) \ = \ 0.6475$

Video

Puedes apoyarte con

https://www.youtube.com/watch?v=4WKzfOorrKs

https://www.youtube.com/watch?v=kkLtvA6UBkQ

Ejercicios. Áreas bajo la curva normal estándar

Dibuja y obtén el área bajo la curva normal estándar...

a) a la derecha de $z = 2.11$

b) entre $z = 0$ y $z = 1.57$

c) a la izquierda de $z = 1.2$

d) a la derecha de $z = -1.07$

e) a la izquierda de $z = -0.23$

f) entre $z = -1$ y $z = 1$

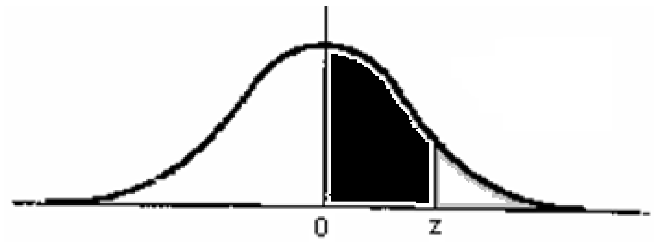
g) entre $z = -2.29$ y $z = -1.19$

h) entre $z = 0$ y $z = 0.49$

i) entre $z = 1$ y $z = 2.13$

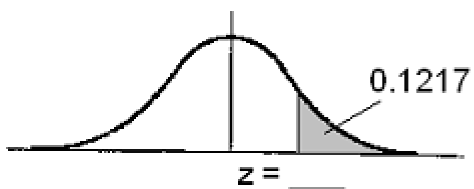
Ejercicios. Correspondiente valor de z

Para obtener el correspondiente valor de z debe considerar que el área total bajo la curva es igual a 1, la existencia de simetría con respecto a la media y que la tabla proporciona áreas de 0 a un valor específico z, esto es:

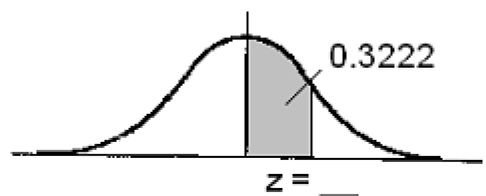


Obtén el correspondiente valor de z en cada una de las siguientes gráficas:

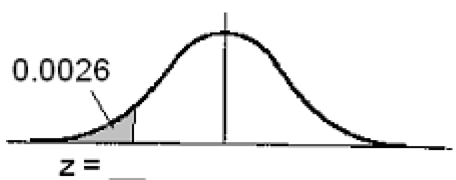
a)



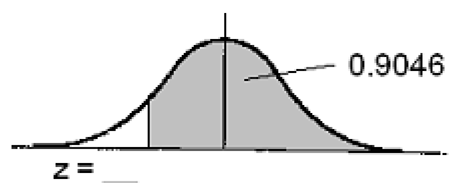
b)



c)

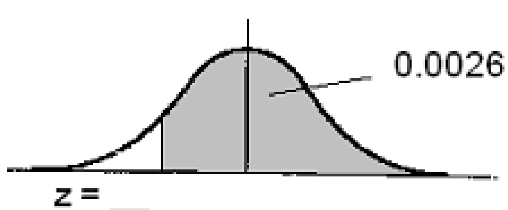


d)

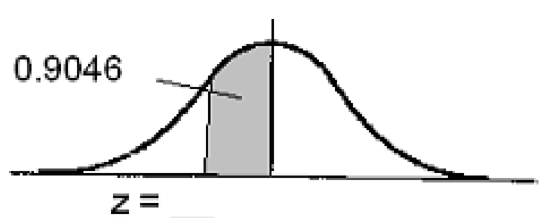


En el inciso e y f hay error, ¿por qué?

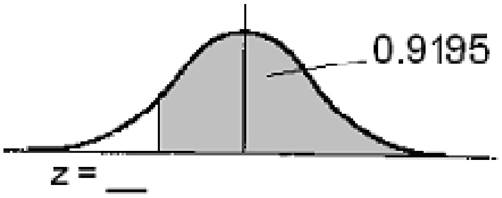
e)



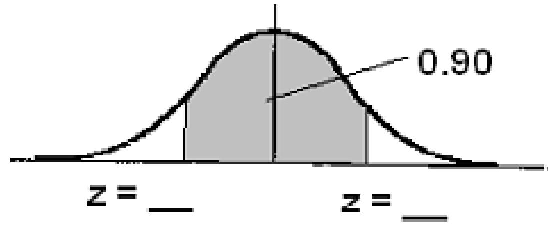
f)



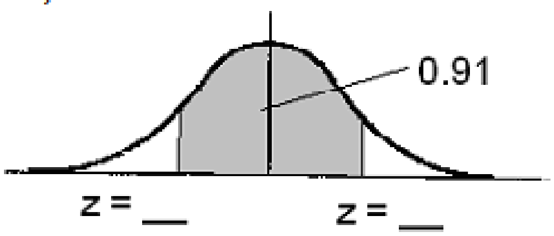
g)



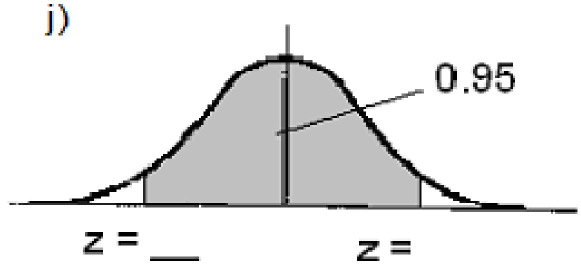
h)



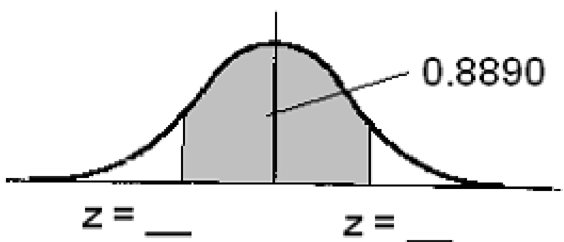
i)



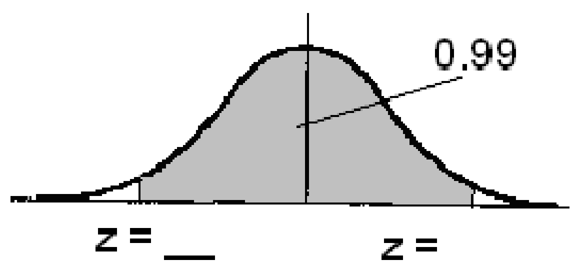
j)



k)



l)



Ejercicios. Correspondiente valor z

a) El área a la derecha de un valor de z es igual a 0.8923, ¿Cuál es el valor de z?

b) El área a la izquierda de un valor de z es igual a 0.6909, ¿Cuál es el valor de z?

c) El área a la derecha de un valor de z es igual a 0.1245, ¿Cuál es el valor de z?

d) El área a la derecha de un valor de z es igual a 0.0998, ¿Cuál es el valor de z?

e) El área a la derecha de un valor de z es igual a 0.4027, ¿Cuál es el valor de z?

Aplicación de la distribución normal en la resolución de problemas

Resolvamos algunos ejercicios de aplicación:

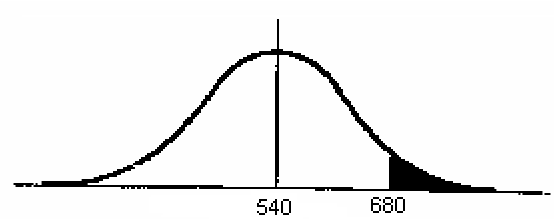
1. Las puntuaciones en una prueba nacional de aprovechamiento tuvieron una distribución normal con media de 540 y desviación estándar de 110.

a) Si la puntuación que obtuviste fue de 680. ¿Qué porcentaje de aquellos que tomaron la prueba obtuvieron mayor calificación que tú?

Solución:

$$\mu = 540 \qquad \sigma = 110$$

El área que representa el porcentaje es:



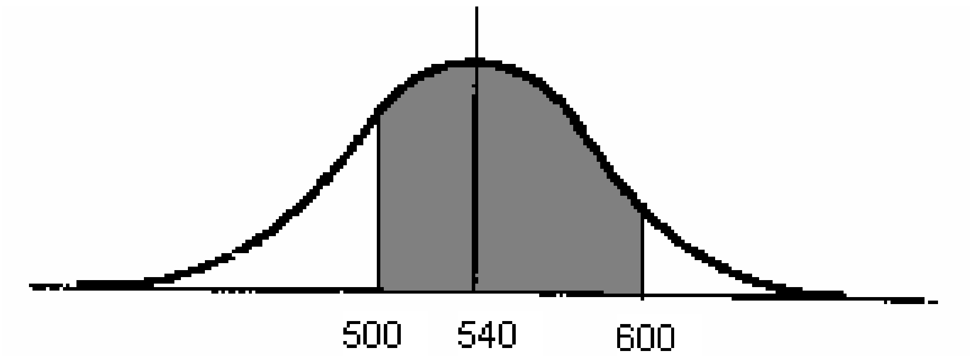
Estandarizando:

$$z = \frac{x-\mu}{\sigma} = \frac{680-540}{110} = 1.27$$

Buscando en tablas el área correspondiente a la curva normal estándar y graficando obtén el porcentaje:

b) ¿Qué porcentaje obtuvo una puntuación entre 500 y 600?

Graficando:

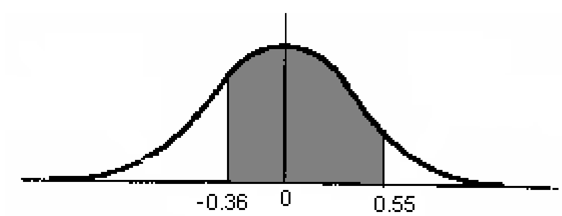


Estandarizando:

$$z = \frac{x-\mu}{\sigma} = \frac{500-540}{110} = -0.36$$

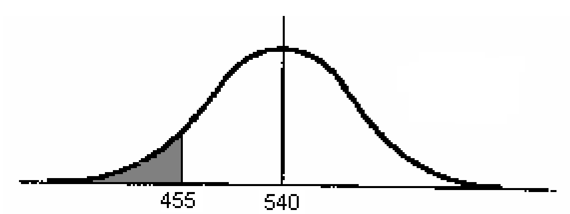
$$z = \frac{x-\mu}{\sigma} = \frac{600-540}{110} = 0.55$$

Una vez estandarizados los valores y utilizando las tablas, obtenemos:



Entonces, el valor del área que representa el porcentaje es ________________

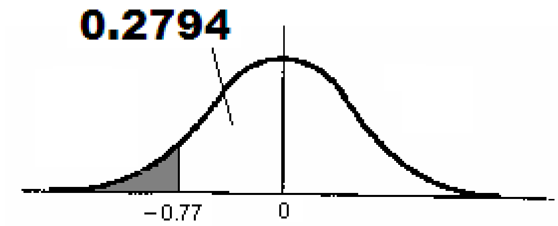
c) ¿Qué porcentaje obtuvo una puntuación menor a los 455?



Estandarizando

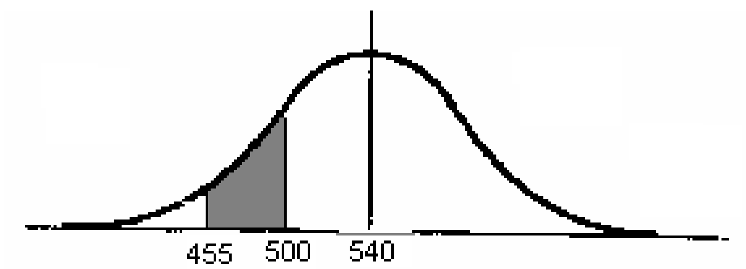
$$z = \frac{x-\mu}{\sigma} = \frac{455-540}{110} = -0.77$$

Tenemos la siguiente gráfica:



Y por lo tanto el área sombreada tiene un valor de 0.2206

d) ¿Qué porcentaje obtuvo una puntuación entre 455 y 500?



Estandarizando:

$$z = \frac{x-\mu}{\sigma} = \frac{\quad -540}{110} = -0.36$$

$$z = \frac{x-\mu}{\sigma} = \frac{455-\quad}{110} = \underline{\qquad}$$

Entonces, ¿Qué porcentaje obtuvo una puntuación entre 455 y 500? ___________

Video Puedes apoyarte con
https://www.youtube.com/watch?v=zxH1joFNNVU

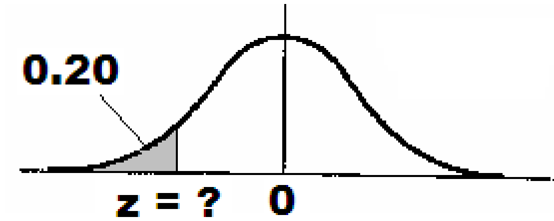
Ejercicios de aplicación

1. Las edades de los trabajadores de una gran industria están distribuidas normalmente con una media de 50 años y una desviación estándar de 5 años. Si se escoge aleatoriamente a un trabajador de esta industria:
   a) ¿Cuál es la probabilidad de que su edad esté entre 50 y 52.5 años? Solución. 0.1915

   b) ¿Cuál es la probabilidad de que su edad no sea mayor de 45 años? Solución. 0.1587

   c) ¿Cuál es la probabilidad de que su edad esté entre 41 y 58 años? Solución. 0.9093

d) Si el 20% de los trabajadores son los más jóvenes, ¿Cuál es la edad máxima para que a un trabajador aún se le considere joven?



De acuerdo a la tabla, la cual proporciona áreas bajo la curva normal estándar entre $z = 0$ y un valor específico de z, busca un área que se aproxime a 0.30 (no busques $z = 0.30$), te darás cuenta de que el valor que más se aproxima es el área con valor de 0.2995 que corresponde al área a la derecha de $z = 0.84$, pero como el área que representa a los más jóvenes está del lado izquierdo de la media (como se trata de la normal estándar el valor de la media es 0) entonces este valor es negativo. Continúa con el procedimiento para llegar a la solución de 45.8 años, que equivale a $45\ años\ 9\ meses\ y\ 22\ días$.

e) Si usted es trabajador de esa industria y su edad es de 47 años con 9 meses, ¿qué porcentaje de los trabajadores es menor que usted? Solución. 32.64%

2. Se sabe que los C. I. de cierta población rural de México se distribuye normalmente con media de 106 y varianza de 144.
   a) Si se quiere seleccionar aleatoriamente a una persona de esta población, ¿Cuál es la probabilidad de que la persona elegida tenga un C. I. entre 100 y 112?

   Solución. 0.383

   b) Se sabe que los llamados “genios” obtienen las calificaciones más altas y que son el 0.5% de la población, ¿Cuál es el C. I. mínimo para que una persona sea considerada como “genio”? Solución. 136.96.

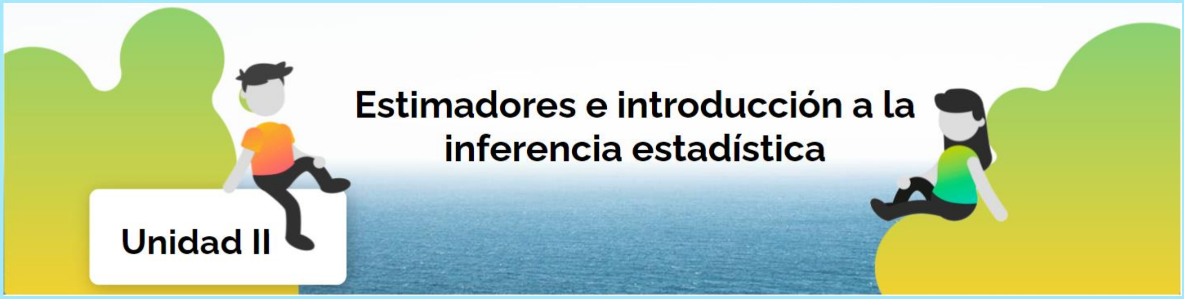
3. El contenido neto de la población de las bolsas chicas de papas fritas Zavrithas se distribuye de manera normal con $\mu = 40$ gramos y $\sigma = 2$ gramos, si se selecciona aleatoriamente una bolsa de esta población,
   a) ¿cuál es la probabilidad de que su contenido neto sea mayor a 42.5 gramos?

   b) ¿cuál es la probabilidad de que su contenido neto esté entre los 38 y 40 gramos?

   c) La empresa sabe que sólo el 1% de la producción de esta presentación se desecha debido a que contienen “demasiado producto”, ¿cuál es el contenido neto mínimo para que una bolsa sea desechada por esta razón?

Propuesta de examen para la Unidad I Modelos de probabilidad y sus aplicaciones

1. Sólo el 10% de los clientes de un restaurante realizan un consumo que genera un pago de más de $600.
   a) ¿Cuál es la probabilidad de que de los siguientes 8 clientes por lo menos 2 realicen un consumo que genere un pago de más de $600?
   b) ¿Cuál es la probabilidad de que de los siguientes 8 clientes más de 6 NO realicen un consumo que genere un pago de más de $600?
   c) De 20 clientes, ¿Cuántos se espera realicen un consumo que genere un pago de más de $600?
2. El periódico Universo sabe que la media de edad de sus suscriptores es de 32 años con una desviación estándar de 14.75 años. Si se selecciona aleatoriamente a un suscriptor,
   a) ¿Cuál es la probabilidad de que su edad sea mayor a los 40 años?
   b) ¿Cuál es la probabilidad de que su edad esté entre los 32 y 40 años?
   c) El periódico menciona que solamente el 1% de sus suscriptores corresponde a adultos con edad muy avanzada, ¿Desde qué edad el periódico está considerando a un suscriptor con edad muy avanzada?
3. Se tiene la siguiente información respecto a un juego de azar: 35% de las veces el jugador gana $120, 30% de las veces gana $80 y en las veces restantes pierde $480. ¿Cuál es el valor esperado del juego de azar?



Unidad II Estimadores e introducción a la inferencia estadística

Aprendizajes

- Establece hipótesis o conjeturas acerca del comportamiento de una variable en una población, a partir de los datos de una muestra, de manera informal, en el contexto de una investigación o un problema.
- Valora la importancia del azar en los procesos de muestreo.
- Valora a los estimadores como variables aleatorias y como indicadores del posible valor puntual de sus correspondientes parámetros.
- Inspecciona el comportamiento de la media muestral y de la proporción muestral como variables aleatorias, obtenidas por medio de la simulación física y/o computacional dentro del contexto de un problema o una investigación y en términos de tendencia, dispersión y distribución.
- Infiere que los estimadores media y proporción se distribuyen de manera aproximadamente *normal,* al trabajar con muestras grandes.
- Construye las distribuciones muestrales para la media y la proporción, bajo las condiciones del Teorema del Límite Central y a partir de la expresión para estandarizar la distribución normal.
- Formula juicios acerca de la representatividad de una muestra, a partir de la probabilidad para algún valor de la media o proporción muestrales, obtenidas por medio de la computadora o la calculadora, dentro del contexto de un problema o una investigación.

Introducción

Lee las siguientes situaciones y contesta lo que se te pide en las líneas correspondientes:

Situación 1. Papas Sabritas

Considera la población de bolsas únicas de papas Sabritas de 40 gramos de la producción de una fábrica en un día en particular. El productor desea que exista la menor variabilidad posible entre el contenido de llenado de cada una de las bolsas, es decir, debe buscar que la media de contenido neto de todas las bolsas, es decir (μ), sea de ___ gramos y la desviación estándar de la población (σ) tenga el menor valor posible.

Sabiendo que la media de contenido neto de los elementos de esa población es $\mu = 40$ gramos, y, si seleccionamos aleatoriamente a 200 elementos de esta población, con la finalidad de obtener la media de contenido neto de esas 200 bolsas, ¿Qué valor obtendríamos? ____ El valor que obtendríamos no estaría determinado y tendríamos una infinidad de posibles valores, es decir, el valor de la media muestral ($\bar{x}$) no estaría determinado y estaría en función de la aleatoriedad.

> **A las características numéricas de una población les llamamos parámetros y a las características numéricas de una muestra les llamamos estadísticos.**

Situación 2. Electores

Considera la población de electores en una comunidad. Varias empresas han realizado un muestreo aleatorio de 500 electores y obtienen la media de edad, ¿Todas las empresas obtendrían la misma media? _________. ¿Por qué? ____

____

El valor que cada empresa obtendría es una media muestral debido a que la media es obtenido mediante los elementos de una ___________. La media muestral la denotamos por ______. La media poblacional la denotamos por ______ y este valor no cambia.

Ahora, si esas empresas también obtienen, de ese muestreo de 500 electores, la proporción de electores que votaría por MO-REÑA, estarían obteniendo una proporción muestral ($\hat{p}$) y no necesariamente cada empresa obtendría el mismo valor. Sin embargo, si considerarán a todos los electores para obtener el porcentaje de electores que votaría por el MO-REÑA, que, por cierto, por lo regular no es posible o no es viable considerar a todos los elementos de una población, se estaría obteniendo el valor de p, la cual denota a la ________________________________

¿Cuál es el valor mínimo que puede tomar una proporción? ____________________

Y, ¿Cuál es el valor máximo que puede tomar una proporción? _________________

Situación 3. CCHacheros

Considera la población de estudiantes del CCH Naucalpan, cuyo media de edad es de 17.8 años, es decir, es el valor de la media poblacional,

¿cómo se denota? _________________

Si realizamos un muestreo aleatorio de 200 estudiantes de esta población, con la finalidad de obtener la media de edad de esos 200 estudiantes, ¿Qué valor obtendríamos? ¿lo podemos determinar? ______. El valor no lo sabemos, este valor está en función de la aleatoriedad, ¿cómo denotamos a la media muestral? _____

Situación 4. NFL

El artículo publicado por Sergio Sarmiento en http://www.sergiosarmiento.com/index.php/columnas/reforma/635-futbolista-gay dice lo siguiente:

> “Nadie sabe qué porcentaje de la población de un país es homosexual. La represión y la discriminación hacen que muchos oculten su inclinación. En Estados Unidos, un país en el que cuando menos los homosexuales no son linchados, el 1 por ciento de los hombres que reportan sus preferencias sexuales en Facebook afirman ser homosexuales en Mississippi y el 3 por ciento en California. El hecho de que California sea un estado más tolerante tiene al parecer mucho que ver con esta manifestación pública. Pero aun en California muchos homosexuales prefieren ocultar su inclinación.
>
> Seth Stephens-Davidowitz calculaba en un artículo en el New York Times el 7 de diciembre de 2013, "How Many American Men Are Gay?", que un 5 por ciento de los hombres estadounidenses son homosexuales. Entre otros argumentos señalaba que el 5 por ciento de las búsquedas de pornografía en Internet es de escenas de hombres homosexuales en toda la Unión Americana.
>
> Si el 5 por ciento de la población masculina de Estados Unidos es gay, esto significaría que de los 1,696 jugadores en activo de la NFL (32 equipos que en temporada regular pueden tener 53 jugadores en activo) un 5 por ciento, unos 85, serían homosexuales. En los cientos de equipos colegiales la cifra sería, por supuesto, mucho mayor.
>
> El que en toda la historia del futbol americano ningún jugador haya declarado abiertamente su homosexualidad es un testimonio de lo cerrado del futbol americano. La cultura del juego es muy especial. Los quarterbacks ponen las manos entre las piernas de los centros para iniciar muchas jugadas, todos los jugadores utilizan pantalones ajustados que resaltan sus nalgas y genitales, y en las celebraciones es muy común que los jugadores se peguen nalgadas unos a otros. Pero casi nadie ha querido ver los aspectos homosexuales de estos rituales.
>
> Michael Sam, un ala defensiva reconocido como All-American o el mejor en su

posición en todo el futbol universitario estadounidense, ha hecho sin duda una gran apuesta al dar a conocer públicamente su homosexualidad. Para empezar, se ha alejado de su familia y en particular de su padre, que mantiene una cultura de rechazo a la homosexualidad. Quizá haya limitado también sus esperanzas de ser escogido en el próximo draft de la NFL, porque muchos jugadores han expresado que no les gustaría jugar y compartir baños y habitación de hotel con un compañero homosexual. En cambio, Sam ha logrado el apoyo público del presidente.

Barack Obama y de su esposa Michelle... en un país en el que la fama -buena o mala- vale más que cualquier otro factor para lograr el éxito económico.

La aceptación de un jugador homosexual sería un paso gigantesco en el proceso, lento pero inevitable, de aceptación de los homosexuales en la sociedad estadounidense. Es un caso como el de Jackie Robinson de los Brooklyn Dodgers, que en 1947 se convirtió en el primer jugador negro de las ligas mayores de béisbol. Michael Sam está abriendo terreno. Pronto llegará un momento en que los jugadores tendrán que ser juzgados sólo por su calidad y ya no por el color de su piel o por sus preferencias sexuales."

En el tercer párrafo se menciona una proporción poblacional del 0.05 (5%), ¿Cómo denotamos a la proporción poblacional? ________ En este mismo párrafo, se está concluyendo erróneamente, ¿Cuál es la conclusión errónea? _________________

____

____

____

____

Critica cada una de las aseveraciones que Sergio Sarmiento hace en este artículo, pueden ser varias, algunas relacionadas con la estadística y otras no. Escribe tus críticas:

____

____

Situación 5. Gasto por transporte

Considera la siguiente situación hipotética: “En el salón de clases se conformarán equipos de 4 personas. Cada equipo realizará un muestreo aleatorio de 50 estudiantes para preguntarles el gasto por transporte a la semana. Usando esos datos muestrales, los alumnos obtendrán la media ($\bar{x}$ porqué proviene de una ______ y no μ porque no es un valor obtenido de todos los elementos que conforman la __________) ¿Necesariamente cada equipo obtendría la misma $\bar{x}$? ______ ¿Por qué? ____________________

También al realizar el muestreo, obtendrán la proporción de mujeres seleccionadas, es decir, obtendrán la __________ muestral, ¿Cómo denotamos a la proporción muestral? ________ ¿Necesariamente cada equipo obtendrá el mismo valor para la proporción muestral? __________________________ ¿Por qué? ____________________

Como te percataste, el valor de $\bar{x}$ y $\hat{p}$ son características numéricas de la muestra, es decir son ____________. El valor de cada uno de ellos no está determinado y puede variar de muestra a muestra y su valor está en función de la aleatoriedad. Esto conlleva a pensar que existe una cierta probabilidad para los valores de los estadísticos.

Teorema de Límite Central

1. Si la población que se muestreo está distribuida de manera normal, la distribución muestral de medias estará normalmente distribuidos respecto a todos los tamaños de muestra.
2. Si la población no es normal, la distribución muestral de medias será aproximadamente normal respecto a un tamaño de muestra grande.

Distribución muestral de la media de la muestra

"Supongamos que seleccionamos una muestra de tamaño n de una población que tiene una media μ y desviación estándar σ. Entonces calculamos la media de esta muestra $\bar{x}$. Si repetimos este procedimiento básico de muestreo hasta que todos los sucesos posibles estén considerados, el conjunto resultante de $\bar{x}$ produce una distribución probabilística. Esta distribución es llamada distribución muestral de la media de la muestra $\bar{x}$." (Christensen, 1983, p.284-285).

La media de estas medias muestrales, simbolizada $\mu_{\bar{x}}$, es la misma que la media de la población μ, esto es

$$\mu_{\bar{x}} = \mu.$$

Si seleccionamos las muestras con reemplazamiento, o de una población infinitamente grande, entonces la desviación estándar de la distribución de todas las posibles medias muestrales, que se simboliza por $\sigma_{\bar{x}}$, es igual a la desviación estándar de la población dividida entre la media n de la muestra. Esto es:

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

Para descubrir estas leyes generales, repetimos un procedimiento de muestreo dado, un gran número de veces, y calculando cada vez la media de la muestra $\bar{x}$.

La media de la muestra resultante crea una distribución muestral de medias. Esta distribución muestral de $\bar{x}$, tiende al centro, alrededor de la media μ de la población. También la distribución muestral de medias tiende a mostrar menor variación y dispersión que la población original. Esto es porque el cálculo de una media regula las influencias de los valores extremos que pueden crear variación en la población.

Ejemplo

Considera el conjunto de los números enteros impares {1, 3, 5, 7, 9} y la lista de todas las muestras de tamaño 2 que puedan ser seleccionadas con reemplazo de este conjunto. Construye la distribución de muestreo de las medias muestrales para muestras de tamaño 2 seleccionadas de este conjunto.

$$(1,1), (3,1), (5,1), (7,1), (9,1)$$

$$(1,3), (3,3), (5,3), (7,3), (9,3)$$

$$(1,5), (3,5), (5,5), (7,5), (9,5)$$

$$(1,7), (3,7), (5,7), (7,7), (9,7)$$

$$(1,9), (3,9), (5,9), (7,9), (9,9)$$

La población de medias maestrales es:

1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8
5	6	7	8	9

Por lo tanto:

$\bar{x}$	1	2	3	4	5	6	7	8	9
$P(\bar{x})$	1/25	2/25	3/25	4/25	5/25	4/25	3/25	2/25	1/25

Video

Puedes apoyarte con https://www.youtube.com/watch?v=as83ZYbl-8E

https://www.youtube.com/watch?v=PmOhKT4wOJU

Distribución muestral de la proporción de la muestra

Supongamos que usted selecciona una muestra de tamaño n de una población en la que la proporción p de cierta característica. Suponga, además, que calcula la proporción de la muestra y le llamamos $\hat{p}$. Si usted repite este experimento básico de muestreo hasta que sean considerados todos los posibles sucesos muestrales el conjunto resultante de proporciones muestrales p de la muestra. La media de estas proporciones muestrales $\mu_{\hat{p}}$ es igual a la proporción de la población, esto es,

$$\mu_{\hat{p}} = p.$$

Si usted selecciona las muestras con reemplazamiento o de una población infinitamente grande, la varianza de la distribución de todas las posibles proporciones muestrales, a la cual simbolizamos $\sigma_{\hat{p}}$, está relacionada con p, en donde n es el tamaño de la muestra, tal como sigue:

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

Ejemplo en video

Con la ayuda de tu profesor(a) y/o el video al cual puedes accesar por medio del enlace https://www.youtube.com/watch?v=q3ysJacFT3Y realiza y analiza el siguiente ejercicio:

Considera la población {Ana, Beatriz, Gustavo} y todas las muestras posibles de tamaño 2 con sustitución para obtener la proporción de mujeres en cada muestra. Comprueba que $\mu_{\hat{p}} = p$ y que $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

Ejercicios. Relación de expresiones matemáticas

1. Relaciona con una línea cada óvalo de la columna de la izquierda con uno de la columna derecha, de tal forma que ninguno quedé con un ovalo que lo relacione conceptualmente de manera correcta:

$\sigma_{\bar{x}}$	$\sqrt{\frac{p(1-p)}{n}}$
DESVIACIÓN ESTANDAR	$\hat{p}$
MEDIA MUESTRAL	n
MEDIA POBLACIONAL DE LAS MEDIAS MUESTRALES	σ
Tamaño de la muestra	$\bar{x}$
proporción muestral	μ
MEDIA POBLACIONAL DE LAS PROPORCIONES MUESTRALES	$\sigma_{\hat{p}}$
MEDIA POBLACIONAL	$\frac{\sigma}{\sqrt{n}}$
$\mu_{\hat{p}}$	$\mu_{\bar{x}}$
Proporción poblacional	p

2. Llena los espacios vacíos con la opción correcta. Las opciones son

p	$\sigma_{\hat{p}}$
$\mu_{\hat{p}}$	$\hat{p}$
n	$\sqrt{\frac{p(1-p)}{n}}$
1	

Considera una muestra de tamaño _________de una población y la proporción ________de cierta característica. Supón que calculas la proporción de la muestra y le llamamos _________. Si repites este experimento básico de muestreo hasta que sean considerados todos los posibles sucesos muestrales, el conjunto resultante de proporciones muestrales $\hat{p}$ tiene una media denotada por ___________, la cual es igual a la proporción de la población, esto es, ___________ $= p$

Si seleccionas las muestras con reemplazamiento o de una población infinitamente grande, la desviación estándar de la distribución de todas las posibles proporciones muestrales, a la cual simbolizamos ____________, está relacionada con n y p, en donde n es el tamaño de la muestra, tal como sigue: $\sigma_{\hat{p}}$ =_______________

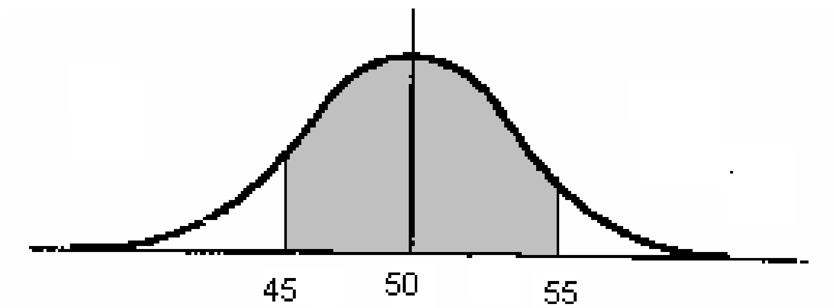
¿Qué rango de valores tomará p? De 0 a ____

Ejemplo

Se va a seleccionar una muestra aleatoria de tamaño 36 de una población que tiene media $\mu = 50$ y desviación estándar $\sigma = 10$.

a) ¿Cuál es la probabilidad de que esta media muestral esté entre 45 y 55?

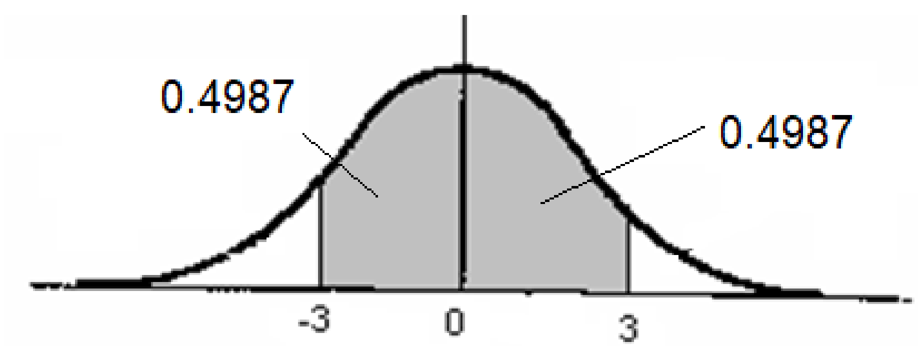
Solución:



$$P(45 \leq \bar{x} \leq 55) = ?$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{45 - 50}{10/\sqrt{36}} = -3$$

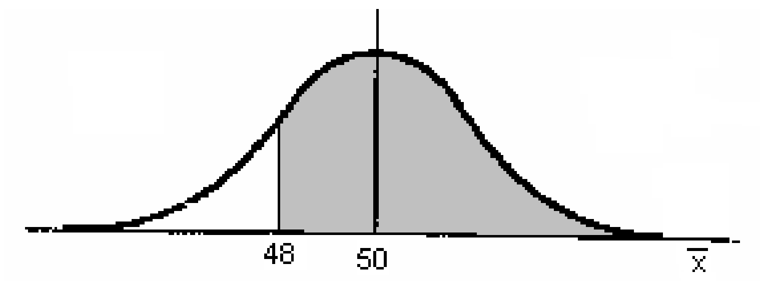
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{55 - 50}{10/\sqrt{36}} = 3$$



$$P(45 \leq \bar{x} \leq 55) = 0.4987 + 0.4987 = 0.9974$$

b) ¿Cuál es la probabilidad de que la media muestral tenga un valor mayor que 48?

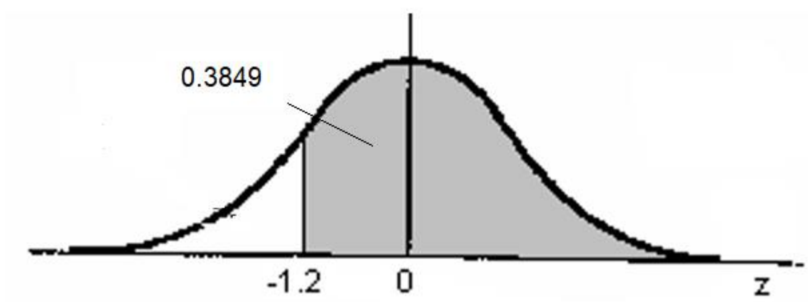
Solución:



$P(\bar{x} > 48) = ?$

Estandarizando:

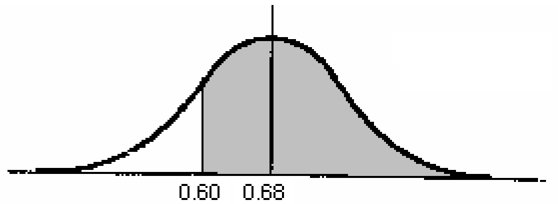
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{48 - 50}{10 / \sqrt{36}} = -1.2$$



$$P(\bar{x} > 48) = 0.5 + 0.3849 = 0.8849$$

Se sabe que la proporción de estudiantes del sexo femenino del plantel Naucalpan es de 68%. La directora del plantel seleccionará a 49 estudiantes en forma aleatoria. ¿Cuál es la probabilidad de que la selección tenga una proporción de mujeres mayor al 60 %? Solución.

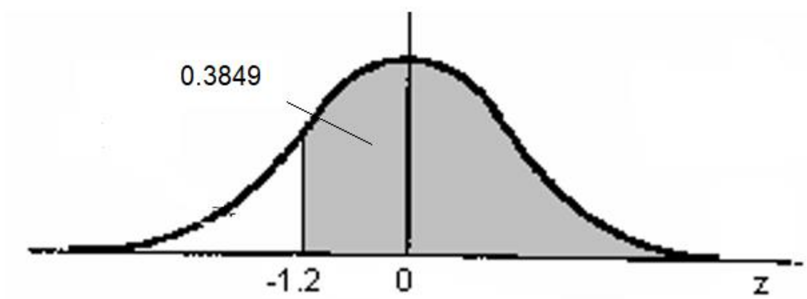
$p = 0.68, \quad n = 49$



estandarizando:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\frac{0.60-0.68}{\sqrt{\frac{(0.68)(0.32)}{49}}} = -1.2$$

entonces:



Por lo tanto, $P(\hat{p} > 0.60) = 0.5 + 0.3849 = 0.8849$

Ejercicios. Serie 1

Ejercicio 1. Considera la población {Ana, Beatriz, Gustavo} y todas las muestras posibles de tamaño 2 con sustitución para obtener la proporción de mujeres en cada muestra. Comprueba que $\mu_{\hat{p}} = p$ y que $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Video

Puedes apoyarte con https://www.youtube.com/watch?v=q3ysJacFT3Y

Ejercicio 2. Considera la siguiente población de tres bicicletas familiares que se rentan en el Bosque de Aragón $\{b1, b2, b3\}$ en donde sólo $b1$ no está en buenas condiciones. También considera todas las muestras posibles de tamaño 2 con reemplazo. Con esto:

a) Obtén la distribución muestral de las proporciones muestrales del número de bicicletas familiares de cada muestra que no están en buenas condiciones

b) Obtén $\mu_{\hat{p}}$
c) Obtén p
d) Verifica que $\mu_{\hat{p}} = p$
e) Obtén $\sigma_{\hat{p}}$ mediante la distribución muestral
f) Obtén $\sigma_{\hat{p}}$, mediante $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
g) Verifica que lo obtenido en e) y f) son iguales

Ejercicio 3. Considera el conjunto de siguientes números enteros pares {0, 2, 4, 6}. Haz una lista de todas las muestras de tamaño 2 que puedan ser seleccionadas con reemplazo de este conjunto. Construye la distribución de muestreo de las medias muestrales para muestras de tamaño 2 seleccionadas de este conjunto.

Ejercicio 4. La delincuencia en el "Municipio de Fonseka" tiene una media de 9 delitos al día con una desviación estándar de 3 delitos. ¿Cuál es la probabilidad de que en los próximos 28 días la media de delitos…

a) sea mayor a los 8 delitos?
b) sea menor a los 8.5 delitos?

Ejercicio 5. Se sabe que la proporción de estudiantes del sexo femenino del plantel Naucalpan es de 68%. La directora del plantel seleccionará a 49 estudiantes en forma aleatoria. ¿Cuál es la probabilidad de que la selección tenga una proporción de mujeres mayor al 60 %?

Ejercicio 6. PROFECO realizará un muestreo de 90 teléfonos celulares a la empresa PANASOUND, esto para saber el porcentaje de teléfonos celulares defectuosos que produce. La empresa sabe que la proporción de teléfonos celulares defectuosos que produce es de únicamente el 2%, con esto información, ¿qué probabilidad hay de que la muestra contenga...

a) más de 5% teléfonos celulares defectuosos?
b) por mucho 3 teléfonos celulares defectuosos?

Ejercicios. Serie 2

Ejercicio 1. El 5% de los clavos que produce una fábrica son defectuosos. Si se realiza un muestreo de 50 clavos como parte de un estudio de control de calidad, ¿Cuál es la probabilidad de que la muestra contenga

a) entre el 4% y el 5% de clavos defectuosos?
b) más del 8% de clavos defectuosos?
c) menos del 2% de clavos defectuosos?
d) menos de 3 clavos defectuosos?
e) más de 5 clavos defectuosos?
f) más de 7 clavos defectuosos?

Video Puedes apoyarte con https://www.youtube.com/watch?v=VULy5bvIAn4

Ejercicio 2. De los videos de sus prácticas de bateo ha podido calcular los distintos ángulos de sus batazos, teniendo una media de 42.5^0 con una desviación estándar de 6^0. ¿Qué probabilidad hay de que en los siguientes 50 batazos la media de su ángulo de bateo esté entre los 43^0 y 44^0?

Ejercicio 3. "Proféko" realizará un muestreo aleatorio de 200 envases de refresco de "cuca-cola" en su presentación de $600\ mls$, debido al gran número de demandas que la empresa tiene por contener menos refresco de lo que se especifica. Si la media de contenido de refresco de estos refrescos es de $600\ mls$ con una desviación estándar de $15\ mls$,

a) ¿cuál es la probabilidad de que la media de contenido de refresco de la muestra sea menor a los $598\ mls$?
b) La empresa será demandada si se obtiene una media inferior a los $597\ mls$, ¿cuál es la probabilidad de ser demandada?
c) La empresa parará el proceso de producción, si la media de llenado es superior a los $603\ mls$, ¿cuál es la probabilidad de que la empresa pare por esta razón?

Ejercicio 4. La media de vida útil de los focos "Duraluz" es de 200 horas con una desviación estándar de 40 horas. Si se seleccionan aleatoriamente 30 de estos focos, ¿cuál es la probabilidad de que la media de vida útil de la muestra sea mayor a 195 horas?

Ejercicio 5. Se sabe que el porcentaje de mujeres en cierta comunidad es del 65%. Si se seleccionan aleatoriamente a 80 personas, ¿cuál es la probabilidad de que la muestra contenga

a) más de 70% de mujeres?
b) más de 60 mujeres?
c) menos del 60% de mujeres?
d) menos de 54 mujeres

Propuesta de examen para la Unidad II Estimadores e introducción a la inferencia estadística

1. Se sabe que una máquina genera 3% de artículos defectuosos. Si se seleccionan de manera aleatoria 150 artículos generados por esta máquina
   a) ¿Cuál es la probabilidad de que la muestra contenga por lo menos 3 artículos defectuosos?
   b) ¿Cuál es la probabilidad de que la muestra contenga más de 2% artículos defectuosos?
2. La media de vida útil de la lámpara de los proyectores de la marca EPSON es de 4000 horas con una desviación estándar de 520 horas. Si una empresa compra 200 proyectores de esa marca,
   a) ¿Cuál es la probabilidad de que la media de la vida útil de la lámpara de esos 200 proyectores sea más de 4100 horas?
   b) ¿Cuál es la probabilidad de que la media de la vida útil de la lampara de esos 200 proyectores sea por lo menos de 3880 horas?

Unidad III Inferencia estadística

Aprendizajes

- Construye el concepto de estimación por intervalo a partir de un problema.
- Deduce las expresiones para el cálculo de intervalos de confianza para media o para proporción, bajo las condiciones del Teorema del Límite Central.
- Estima la media y proporción poblacionales por medio del intervalo de confianza correspondiente, que haya generado en el contexto de una investigación o un problema, comunicando su interpretación.
- Analiza el concepto de prueba de hipótesis, para una media y para una proporción, preferentemente con el uso de la computadora o la calculadora, en el contexto de una investigación o un problema.
- Mide la validez de una hipótesis estadística para la media o para la proporción, en el contexto de una investigación o un problema.
- Evalúa las características de interés en una población, de manera formal, a partir de los datos de una muestra, en el contexto de una investigación o un problema.

Introducción

“La mayoría de nosotros encontramos pronósticos, proyecciones, estimaciones y aproximaciones en alguna forma u otra casi a diario. El pronóstico del Servicio Meteorológico acerca de la probabilidad de que llueva; la proyección del economista del último giro de la economía, el informe mensual del índice de precios al mayoreo el índice de precios al consumidor son todos ejemplos comunes de estimaciones.” (Estadística paso a paso, Christensen, 1983, p. 309)

¿Qué estimaciones realizas a diario?

____
____
____

¿Qué información utilizas para realizar tus estimaciones?

____
____
____

¿Qué estimaciones has oído que se realizan en la radio, periódico, tv o revistas?

____
____
____

¿Cuál es tu definición de estimación?

____
____
____

Estimación

"Es el procedimiento utilizado cuando se responde a una pregunta que pide el valor de un parámetro poblacional. Por ejemplo, ¿Cuál es la distancia media en un solo sentido que deben viajar los estudiantes que asisten a la universidad local?

Si es necesario contestar a esta pregunta debe tomarse una muestra de la población y calcular la media muestral $\bar{x}$. Supóngase que se selecciona una muestra aleatoria de 100 distancias en un solo sentido y que resulta una media de 10.22 millas. ¿Cuál es la estimación del valor medio de la población? Si se toma la media muestral como dicha estimación, se estará efectuando una estimación puntual. " (Estadística Elemental, Johnson, p. 297).

Supongamos ahora que se selecciona una muestra aleatoria de 120 bolsas chicas de papas Zavrithas para estimar la media de contenido neto de papas fritas. Y después de efectuado el muestreo, se obtiene una media de 39.5 gramos. Es decir, la media muestral es $\bar{x} = 39.5$, con ello se ha estimado que la media de contenido de papas fritas de bolsas chicas Zavrithas en 39.5 gramos. Esto no quiere decir que la media poblacional μ sea exactamente igual a 39.5, sino que se interpreta esta estimación como "μ está próxima a 39.5". El término de proximidad es relativo, pero quizás en este caso "próximo" debe ser definido arbitrariamente como estar "dentro de 1 gramo de μ". Si 1 gramo satisface la idea intuitiva de proximidad, entonces decir "μ está próxima a 39.5 "es comparable a decir "μ está entre $38.5\ (39.5-1)$y $40.5\ (39.5+1)$". Este tipo de estimación utiliza el concepto de intervalo y asigna una medida a la confiabilidad de éste en la estimación del parámetro en cuestión.

Nivel de confianza $(1-\propto)$

Probabilidad de que la muestra por seleccionarse produzca valores límite que se localicen en lados opuestos del parámetro que se estima. Algunas veces el nivel de confianza se llama coeficiente de confianza.

"Sabemos que parte de la idea de la Estadística es utilizar los resultados obtenidos de la muestra como valores aproximados a los que se supone operan para la población; es decir, usamos los estadísticos para “tantear” o estimar los parámetros, y es eso precisamente a lo que se refiere el término estimación. Existen dos formas de realizar estimación: puntual y por intervalos. La estimación puntual consiste en calcular un estadístico y asumir que su valor se aproxima al valor del parámetro. La estimación por intervalos consiste en generar precisamente un intervalo a partir de valores obtenidos de una muestra, y dentro del que se espera esté contenido algún parámetro en particular. Ahora, se sabe que una muestra es tomada de manera aleatoria, por lo que deberá de incluirse en la estimación un elemento de probabilidad, y este elemento nos permitirá dar cierto grado de certidumbre a nuestras estimaciones; esto es, si sabemos cómo se comporta la población de interés, o al menos el estimador con el que se trabaje, en términos de distribución de probabilidad, entonces podemos decidir el porcentaje de seguridad para nuestra estimación; esto es, podemos establecer de antemano la probabilidad de que al construir el intervalo de confianza, efectivamente consigamos que dentro de éste se encuentre el parámetro a estimar.” (Estadística y probabilidad II, Materia de Apoyo para el PAE, Álvarez y otros, p. 30).

Estimación Puntual

Sea una variable aleatoria X que sigue una distribución con Función de Distribución $F(x;\theta)$ y Función de Densidad $f(x;\theta)$. De esa población extraemos una muestra $(x_1, x_2, ..., x_n)$ y a partir de esos datos intentamos acotar o conocer el valor del parámetro θ que nos es desconocido. Al proceso de obtener un valor aproximado del parámetro a partir de los datos de la muestra se le llama estimación.

Estimador

Llamaremos estimador del parámetro θ y lo designaremos por $\bar{\theta}$ a cualquier función de los valores de la muestra

$$\bar{\theta} = \bar{\theta}(x_1, x_2, ..., x_n)$$

Cuyos valores se asigna al parámetro.

Propiedades de los estimadores

La definición que se ha dado de estimador es tan general que para estimar un parámetro existen infinitos estimadores potenciales en consecuencia necesitamos criterios que nos permitan seleccionar los estimadores adecuados a una situación dada. Como quiera que los estimadores son variables aleatorias en el muestreo estos criterios harán referencia a propiedades de su distribución.

Estimador Insesgado

Se dice que un estimador es insesgado o centrado si la media o esperanza matemática del estimador coincide con el verdadero valor del parámetro. Es decir, si: $E[\bar{\theta}] = \theta$

Se llama sesgo de un estimador a la diferencia entre su media y el verdadero valor:

$$sesgo(\bar{\theta}) = E[\bar{\theta}] - \theta$$

Estimador Eficiente

Aunque la propiedad de ser insesgado es deseable, no es un criterio suficiente para seleccionar un estimador, ya que pueden existir estimadores centrados con distribuciones muy diferentes, para completar la anterior propiedad se introduce el concepto de eficiencia.

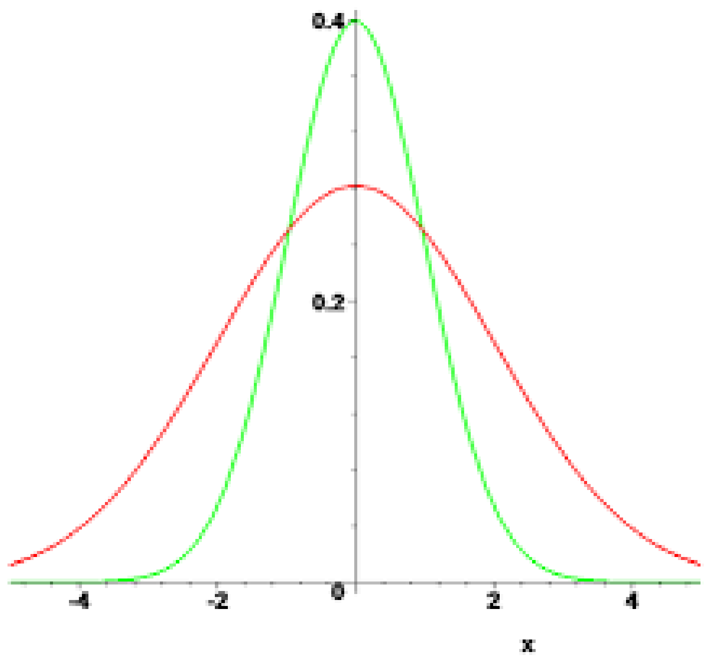
Dados dos estimadores de un mismo parámetro $\bar{\theta}_1$ y $\bar{\theta}_2$ diremos que $\bar{\theta}_1$ es más eficiente que $\bar{\theta}_2$, si teniendo ambos la misma media $\bar{\theta}_1$ tienen una varianza menor que $\bar{\theta}_2$. Esta propiedad nos permite decir entre dos estimadores cual es más conveniente. En términos absolutos establecemos que:

Se dice que un estimador es eficiente si es insesgado y de mínima varianza.

Obviamente si existe un estimador eficiente éste será el más adecuado, de acuerdo con los criterios tradicionales. Sin embargo, no siempre es posible determinar la existencia de un estimador eficiente.

Ejemplo

En dos distribuciones muestrales, la varianza σ_1^2 de (color verde) es mínima en comparación con la varianza del otro estimador σ_2^2 (color rojo). Se puede observar que cuando un estimador tiene la varianza mínima y es **insesgado** se le denomina estimador **Eficiente.**



$$\sigma_1^2 < \sigma_2^2$$

Estimador Consistente

Cuando el estimador no es insesgado en primera medida, lo que sería lo idóneo, se requiere al menos que su valor oscile cerca del valor del parámetro para tamaños de muestra grandes, es decir, un estimador es consistente cuando:

$$Lim_{n+\infty} E(\bar{\theta}) = \theta$$

$$Lim_{n+\infty} V(\bar{\theta}) = 0$$

De esta manera, estimador es consistente cuando en un estudio probabilístico se está aumentando el tamaño de la muestra $[n \to \infty]$, la media de la muestra $\bar{x}\left[\mu_{\bar{x}}\right]$, se acerca a la media de la población μ; es decir que $\bar{x}$ es un estimador consistente de μ.

$$\bar{x} \approx \mu \text{, si } n \to \infty$$

Estimador Suficiente

Intuitivamente un estimador es suficiente para un parámetro si toda la información acerca del parámetro está contenida en la muestra.

Formalmente seria: un estadístico $\bar{\theta}$ se dice suficiente para θ basado en una muestra aleatoria $x_1, x_2, ..., x_n$ de una población con función de densidad de población $f(x;\theta)$. Si la distribución condicional de las variables aleatorias $x_1, x_2, ..., x_n$ dado $\bar{\theta}$ no dependen del parámetro θ, es decir, $\bar{\theta}$ es un estimador suficiente de θ si:

$$f_x(x_1, x_2, ..., x_n \mid \bar{\theta} = \theta) = g(\bar{x}):$$

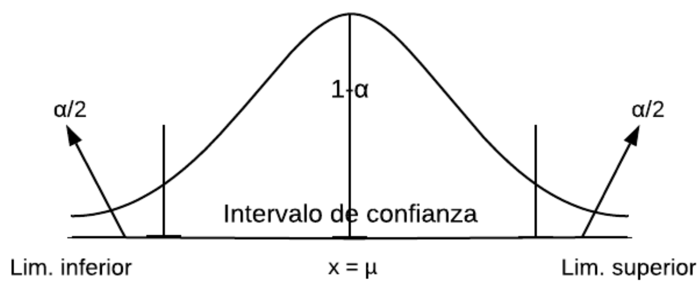
Donde $g(\bar{x}) = (x_1, x_2, ..., x_n)$.

Por lo tanto, un estimador es suficiente con relación a un parámetro si se maneja toda la información estadística de la muestra para realizar los cálculos como puede ser la media aritmética, la mediana, la moda, la proporción, la varianza y la deviación estándar; es decir utilizar información suficiente para realizar estimaciones.

Estimación por Intervalos

Una estimación por intervalos de un parametro $\theta\left[\mu, P, \sigma^2\right]$de la población, es un intervalo de la forma que tendrá un límite inferior y un límite superior de tal manera que se cumpla el intervalo de confianza; donde $\mathrm{L}_{\inf erior.}$ = límite inferior y

$L_{superior.}$ = límite superior del intervalo. Estos intervalos se construyen de tal manera que van a depender del resultado final del procedimiento matemático al aplicar los conceptos estadísticos de los estimadores de la media $\overline{x}$, y de la proporción $\overline{p}$, para una muestra en particular de $\overline{\theta}\left[\overline{x}, \overline{p}, S_{\overline{x}} = \sigma_{\overline{x}}\right]$.



$1-\alpha$ = Intervalo de Confianza

α = Significación

$P(L_{inf\,erior} \leq \mu \leq L_{superior}) = 1-\alpha$

Estimación por Intervalos

Estimación por Intervalos para Media de la Muestra de la Población

Para la estimación por intervalos de la media muestral de la población se considera el valor Z, para una muestra que se aproximación a una distribución normal:

$$Z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}}$$

Utilizando el Teorema Central de Límite para la media muestral el valor de Z es: $Z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ Despejando el estimador de la muestra " $\overline{x}$ " y sustituyendo en la ecuación de estimación por intervalos nos queda la siguiente expresión matemática:

$$1-\alpha = \bar{x} \pm Z_{\frac{\alpha}{2}}(\sigma_{\bar{x}}) \qquad \text{donde} \qquad \bar{x} = \mu$$

Con esta expresión nos permite calcular los límites del intervalo de confianza:

$$1-\alpha = \mathrm{P}(\mathrm{L}_{\inf erior} \le \mu \le \mathrm{L}_{superior})$$

Sustituyendo los Limites en la expresión del intervalo de confianza, queda la siguiente expresión:

$$1-\alpha = \mathrm{P}(\bar{x} - Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right))$$

Donde él "Límite inferior" y "Límite superior" son:

$$\mathrm{L}_{\inf erior} = \bar{x} - Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \qquad \text{y} \qquad \mathrm{L}_{superior} = \bar{x} + Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

También, es importante mencionar que al producto del valor de **"Z"** (valor critico) multiplicado por el cociente que resulta de dividir la desviación estándar entre la raíz cuadrada del tamaño de la muestra se le llama **"error de estimación"** el cual es:

$$e = Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Para calcular un intervalo de confianza ($1-\alpha$) del 100% para la media de la muestra "$\mu_{\bar{X}}$"se supone que se conoce la desviación estándar de la muestra "S", en caso contrario de no conocerse se reemplazará por desviación estándar de la población "σ", siempre y cuando el tamaño de la muestra sea mayor que 30 $(n \ge 30)$.

Ejemplo

En un estudio realizado a 81 personas que acuden a un centro comercial a comprar su despensa en periodos de quincena, se encontró una media de gasto de $1400 pesos con una desviación estándar $600 pesos. El gerente de la tienda requiere estimar lo intervalos de confianza al 99% para cuando acuden a ese centro comercial los clientes.

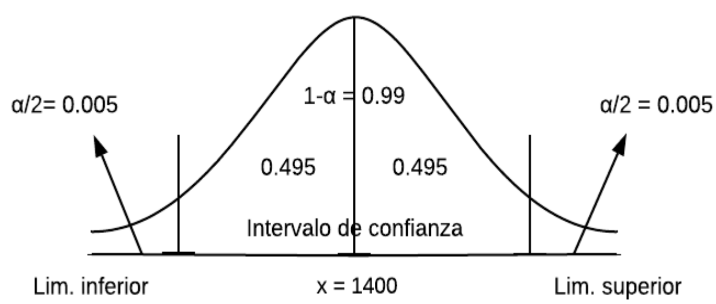
Solución

Datos

$\mathrm{n} = 81 \qquad \bar{x} = \$1400$

$\sigma = \$600$

$1 - \alpha = 99\% = 0.99 \qquad \alpha = 1\% = 0.1 \qquad \dfrac{\alpha}{2} = 0.5\% = 0.005$



$$1 - \alpha = \mathrm{P}(\mathrm{L}_{\inf erior} \leq \mu \leq \mathrm{L}_{superior})$$

Para encontrar el valor de Z se busca **el porcentaje de 0.495** dentro de la tabla de la Distribución Normal Estándar. El valor corresponde a una **Z=2.58**

Calculando el valor del error de estimación: $e = Z_{\frac{\alpha}{2}}\left(\dfrac{\sigma}{\sqrt{n}}\right) = (2.58)\left(\dfrac{600}{\sqrt{81}}\right) = 172$

Sustituyendo en el Intervalo de Confianza:

$$1 - \alpha = \bar{x} \pm Z_{\frac{\alpha}{2}}\left(\dfrac{\sigma}{\sqrt{n}}\right)$$

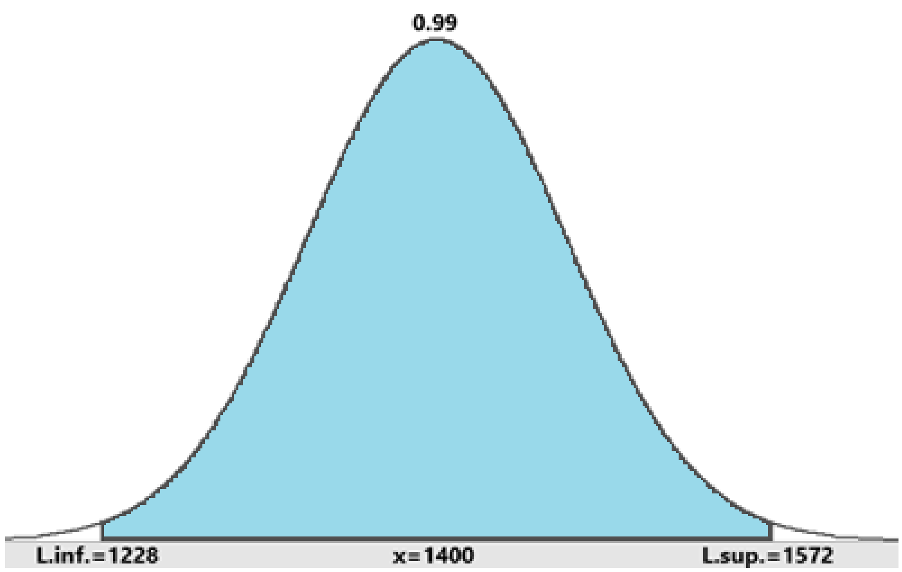
$$1-\alpha = 1400 \pm (2.58)\left(\frac{600}{\sqrt{81}}\right)$$

$$1-\alpha = 1400 \pm 172$$

$$1-\alpha = \begin{Bmatrix} L_{inferior} = 1400 - 172 = 1228 \\ L_{superior} = 1400 + 172 = 1572 \end{Bmatrix}$$

$$1-\alpha = P(L_{inferior} \leq \mu \leq L_{superior})$$

$$1-\alpha = P(1228 \leq \mu \leq 1572)$$



Interpretación del resultado

Se estima que la media de gasto de los clientes al comprar su despensa en periodos de quincena en el centro comercial está entre $1228 a $1572 pesos con una confianza del 99%.

Estimación por Intervalos para Proporción de la Muestra de la Población

Para la estimación por intervalos de la proporción muestral de la población se considera el valor de Z, para una muestra que se aproxima a una distribución normal:

$$Z = \frac{\bar{p} - P}{\sigma_{\bar{P}}}$$

Utilizando el Teorema Central de Límite para la proporción muestral de Z es:

$$Z = \frac{\bar{p} - P}{\sqrt{\frac{p*(1-p)}{n}}}.$$

Despejando el estimador de la muestra “ $\bar{p}$ ” y sustituyendo en la ecuación de estimación por intervalos nos queda la siguiente expresión matemática:

$$1-\alpha = \bar{p} \pm Z_{\frac{\alpha}{2}}(\sigma_{\bar{P}}) \qquad \text{donde} \qquad \bar{p} = P$$

$$1-\alpha = \bar{p} \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right)$$

Con esta expresión nos sirve para calcular los límites del intervalo de confianza

$$1-\alpha = \mathrm{P}(\mathrm{L}_{inferior} \leq P \leq \mathrm{L}_{superior})$$

Sustituyendo lo Limites en la expresión del intervalo de confianza, queda la siguiente expresión:

$$1-\alpha = \mathrm{P}(\bar{p} - Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right) \leq P \leq \bar{p} + Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right))$$

Donde él “Límite inferior” y “Límite superior” son:

$$\mathrm{L}_{inferior} = \bar{p} - Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right) \quad \text{y} \quad \mathrm{L}_{superior} = \bar{p} + Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right)$$

También, es importante mencionar que al producto del valor de **"Z"** multiplicado por el cociente que resulta de dividir la desviación estándar entre la raíz cuadrada del tamaño de la muestra se le llama **"error de estimación"** el cual es: $e = Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right)$

Ejemplo

En un estudio realizado en un colegio, una empresa de cigarros realizó una encuesta a 400 estudiantes del colegio sobre el hábito de fumar y encontró que el 80% de los jóvenes fuman cigarros sin importar la marca. La empresa quiere conocer la estimación de los intervalos de confianza al 90%, para la proporción de lo jovenes que fuman actualmente.

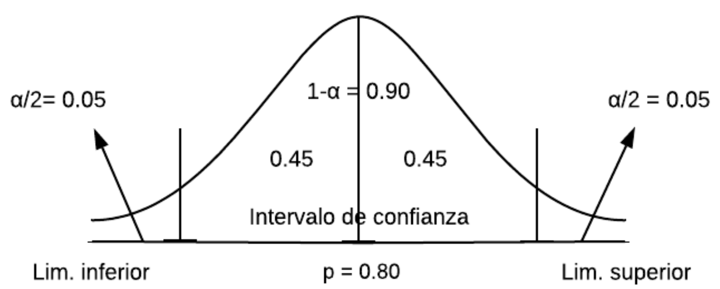
Solución

$n = 400 \qquad \bar{p} = 0.80$

$P(fumar) = P = 80\% = 0.80$

$q(no \quad fumar) = 1 - p = 1 - 0.8 = 0.20$

$1 - \alpha = 90\% = 0.90 \qquad \alpha = 10\% = 0.10 \qquad \frac{\alpha}{2} = 5\% = 0.05$



$$1 - \alpha = \mathrm{P}(\mathrm{L}_{inf\,erior} \le P \le \mathrm{L}_{superior})$$

Para encontrar el valor de Z se busca **el porcentaje de 0.4505** dentro de la tabla de la Distribución Normal Estándar. El valor que corresponde es de **Z=1.65**

Calculando el valor del error de estimación:

$$e = Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right) = (1.65)\left(\sqrt{\frac{(0.8)(0.2)}{400}}\right) = 0.033$$

Sustituyendo en el Intervalo de Confianza:

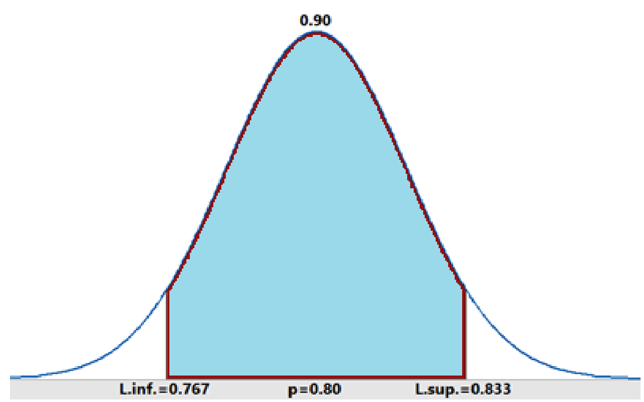
$$1-\alpha = \bar{p} \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p*(1-p)}{n}}\right)$$

$$1-\alpha = 0.8 \pm (1.65)\left(\sqrt{\frac{(0.8)*(0.2)}{400}}\right)$$

$$1-\alpha = 0.8 \pm 0.033$$

$$1-\alpha = \begin{Bmatrix} \mathrm{L}_{\inf erior} = 0.8 - 0.033 = 0.767 \\ \mathrm{L}_{superior} = 0.8 + 0.033 = 0.833 \end{Bmatrix}$$

$$1-\alpha = \mathrm{P}(0.767 \leq P \leq 0.833)$$



Interpretación de Resultados

Se estima que la proporción de jóvenes fumadores de cigarros de cualquier marca fluctúa entre 76.7% y 83.3%, con una confianza del 90%.

Ejercicios de estimación

1. Supón que un psicólogo desea realizar una estimación de intervalo de la media verdadera de los C.I. de alumnos de cierto grupo étnico. Se sabe que los C.I. se distribuyen de manera normal con desviación estándar de 15. Construye un intervalo de confianza del 95% para la media verdadera (μ) con base a una muestra de 25 alumnos con una media muestral de 105.

Interpretación:

2. Se realizó una entrevista telefónica para determinar si quien responde es o no es propietario de una casa. Se selecciona del directorio telefónico una muestra aleatoria de 110 personas y 52 de ellas resultaron ser propietarias. Obtén un intervalo de confianza del 98% para la proporción de propietarios.

Interpretación:

____

____

____

3. En una muestra aleatoria de 64 muchachas universitarias de primer año, el 8% resultaron ser casadas. Determina el intervalo de confianza del 98.8% para la proporción ρ de mujeres universitarias casadas.

Interpretación:

____

____

____

4. Se realizó una investigación de teleaudiencia. En una muestra de 900 espectadores, el número de ellos que veían un programa en particular fue de 180. Determina un intervalo de confianza del 99% para estimar la proporción de espectadores que ven este programa en particular.

Interpretación:

____

____

____

5. El dueño de una tienda de abarrotes consideró los registros de ventas de 40 días seleccionados aleatoriamente. Sabiendo que la desviación estándar es de $\$1200$ y la media muestral obtenida del muestreo es $\$9870$, establezca un intervalo de confianza del 98% para estimar la media de ventas diarias de la tienda de abarrotes.

Interpretación:

____

____

____

6. Se van a realizar, durante un mes, pruebas de mercado de un nuevo cereal en los supermercados. Los resultados para una muestra de 144 tiendas señalaron ventas con una media de $\$120000$. Se conoce que la desviación estándar poblacional es de $\$1800$. Estima un intervalo de confianza del 99% para la media de las ventas de este nuevo cereal.
   Solución. $[\$119\ 613, \$120\ 387]$.

Interpretación:

---

7. Supóngase que está interesado en estimar la media del número de horas que un joven de bachillerato dedica al estudio diariamente.
   Se sabe que $\sigma = 1.2$ horas. De una muestra de 50 alumnos de bachillerato se obtuvo una media de 3.25 horas. Utilice esta información para establecer un intervalo de confianza de 95.5% para estimar la media de horas por día que los estudiantes de bachillerato dedican al estudio. Solución $[2.9089\ hrs., 3.5911\ hrs.]$. Interpreta el resultado.

Interpretación:

8. El banco “Ilusiones Pasajeras” desea estimar el número de minutos que tarda un cajero en atender a una persona una vez que esta llega a cajas. Se conoce que =1.1 minutos. Se escoge una muestra aleatoria de 55 transacciones hechas con el cajero de 5.2 minutos. Obtén un intervalo de confianza del 95% para estimar el tiempo medio que tardan los cajeros en atender a los clientes una vez que llegan a ellos. Solución $[4.9093 min., 5.4907 min.]$. Interpreta el resultado.

Interpretación:

9. Un investigador desea estimar la proporción de hombres que fuman en el municipio de Atizapán de Zaragoza. En una muestra de 100 de ellos, 27 le mencionaron que tienen el hábito de fumar. Utilice está información para establecer un intervalo de confianza del 96.5 % para la proporción de hombres que radican en el municipio de Atizapán de Zaragoza fumadores. Solución [0.1763, 0.3637]. Interprete el resultado obtenido.

Interpretación:

Prueba de hipótesis.

“Una hipótesis estadística es una declaración o afirmación tentativa acerca del valor de un parámetro o parámetros de una población. Tal declaración se considera tentativa debido a que los verdaderos valores de los parámetros en cuestión se desconocen. Los ejecutivos de negocios, científicos, estrategas militares, políticos y muchas otras personas toman decisiones relativas a parámetros de la población. Algunos simplemente adivinan los valores de los parámetros, mientras que otros realizan inferencias acerca de ellos basándose en datos muestrales incompletos. Las pruebas de hipótesis pueden mostrar si una declaración tentativa se ve apoyada o rechazada por la evidencia de la muestra.” (Chao, Lincoln, L., Introducción a la Estadística, 1985, p. 254).

Planteamiento de las hipótesis.

Hay dos tipos de hipótesis. Uno, denominado hipótesis nula, se forma principalmente para determinar si puede rechazarse. Tal hipótesis se denota por convención mediante el símbolo H_0, con la letra H sugiriendo hipótesis y el subíndice 0 indicando “nula” o “inexistente”.

La hipótesis nula, H_0, es una declaración tentativa de que un parámetro de la población es igual a un valor específico. A menudo en tal declaración está implícita la idea de que “no hay diferencia” y de ahí el nombre de hipótesis “nula”.

Por ejemplo, supón que se ha el Municipio de Naucalpan ha publicado que la mediade delitos al día en ese municipio no rebasa los 7 delitos por día. Entonces, la hipótesis nula es H_0: $\mu = 7$, debido a que debe suponerse que es 7 o menor.

La hipótesis alternativa, denotada mediante el símbolo H_a o H_1, ésta se acepta cuando se rechaza la hipótesis nula. La hipótesis alternativa H_1, es una declaración tentativa de que el mismo parámetro de la población tiene un valor diferente del especificado en la hipótesis nula.

Respecto a la media de delitos al día en el Municipio de Naucalpan de que no rebasa los 7 delitos por día, la hipótesis alternativa es $H_1: \mu > 7$ que significa que la media si rebasa los 7 delitos al día.

Se obtendrán datos muestrales para determinar si existe evidencia suficiente para apoyar la hipótesis alternativa o no. Se apoyaría H_1 si los datos muestrales proporcionan una media significativamente alta con lo que H_0 sería rechazada.

Una analogía.

“El razonamiento empleado al determinar si una hipótesis nula puede rechazarse tiene una parecido notable con el empleado en una corte criminal. La corte considera que el acusado es inocente o no culpable, hasta que su culpabilidad se determine. El fiscal obtiene y presenta evidencias en un intento de eliminar la consideración de no culpable, es decir la hipótesis nula. Si no se rechaza ésta, no habrá condena y el acusado quedará libre. Esto no necesariamente significa que en realidad el acusado no sea culpable. Puede ser que en verdad sea culpable, pero la evidencia disponible no es suficiente para condenarlo.” (Chao, Lincoln, L., Introducción a la Estadística, 1985, p. 255).

De acuerdo a la analogía y al supuesto referido al Municipio de Naucalpan:

¿Qué es el “acusado? _______________

¿Qué es el fiscal? (acusador) ________________

¿Con qué se obtendrían las evidencias? ________________________________

Al final de una prueba de hipótesis, el investigador o tomador de decisiones podrá

- ✓ Rechazar una hipótesis verdadera, con lo que estaría cometiendo un error, que le llamaremos Error Tipo I o nivel de significación.
- ✓ Aceptar una hipótesis verdadera, con lo que estaría decidiendo acertadamente.

¿Cuáles son las otras dos situaciones en las que el investigador o tomador de decisiones podrá estar al final de una prueba de hipótesis?

- ✓ ____
- ✓ ____

“Las diferencias entre muestras se deben al azar: a los factores accidentales que determinan la selección de observaciones de cualquier muestra.” (Chao, Lincoln, L., Introducción a la Estadística, 1985, p. 256).

Considerando las hipótesis antes planteadas: $H_0: \mu = 7$ y $H_1: \mu > 7$, para realizar la prueba de hipótesis se realiza un muestreo y se calcula la media. Debido a la naturaleza aleatoria del muestreo, la $\bar{x}$ puede ser significativamente mayor a 7, que hará que se rechace la hipótesis nula cuando en realidad sea verdadera. “¿Cuándo puede considerarse una media muestral como ‘significativamente mayor’? La respuesta a esta pregunta depende del nivel de error que se desea tolerar; es decir, de la probabilidad de que la muestra haya proporcionado una media lo suficientemente mayor que el valor hipotético debido a factores aleatorios.”

El nivel de significación es la probabilidad de rechazar una hipótesis nula verdadera o de cometer lo que se denomina error tipo I; a esta probabilidad comúnmente se denota mediante $\propto$. El valor de $\propto$ afecta a la decisión de considerar significativa cualquier diferencia entre el valor muestral observado y el valor hipotético de la población, o si se considera demasiado extrema para atribuirse al azar. La selección de $\propto$ es arbitraria; depende de que tanto riesgo puede tomarse para rechazar incorrectamente una hipótesis nula verdadera.

Estadístico de prueba

Es una v. a., cuyo valor se utiliza para llegar a la decisión de rechazar o no la hipótesis nula.

Región Crítica

Los valores del estadístico de prueba se dividen en dos categorías: la región de rechazo y la región de aceptación. También se conoce a la región de rechazo como <u>región crítica.</u> La región crítica es el conjunto de valores para el estadístico de prueba que provocará la aceptación de la hipótesis nula.

El valor que separa a las dos regiones se le llama **valor crítico**.

Ejemplo

Se ha informado a los alumnos que la media del promedio de sus calificaciones es de 6.5 con una desviación estándar de 2. Los alumnos indignados por tal aseveración y pensando que la media es mucho mayor a lo mencionado, realizaron un muestreo de 500 estudiantes, obteniendo una media muestral de 6.9. ¿Hay suficiente evidencia para rechazar la información proporcionada a los alumnos con respecto a la media y afirmar que este es mayor a lo mencionado? Utilice un nivel de significación $\propto = 0.05$.¿Cuáles son las dos hipótesis que se deben plantear?

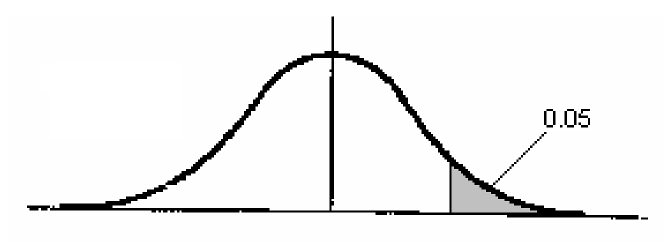
$H_o :$ ______

$H_a :$ ______

¿Cuál es el tamaño de la muestra? ________

¿Cuál es el valor del estadístico de prueba?

$$
\begin{aligned}
Z_{prueba} &= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \\
&= \frac{6.9 - \_\_\_\_}{\_\_\_\_ / \sqrt{\_\_\_}} \\
&= 4.47
\end{aligned}
$$



¿Qué valor tiene el nivel de significación? _________

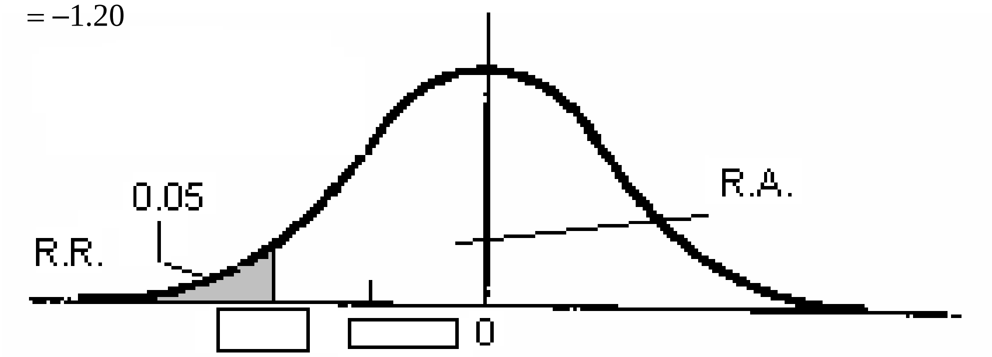
¿Cuál es el valor crítico? _______

Como $Z_{prueba} = 4.47$ se rechaza la hipótesis nula, por lo tanto, hay suficiente evidencia para decir que la media de calificaciones de los alumnos es superior a 6.5 con un nivel de significación del 5%.

Ejemplo

Supóngase que en el CCH-N se ha asegurado que la media del peso de las alumnas es de 54.4 kilogramos con una desviación estándar de 5.4 kilogramos. Una de las profesoras no cree que tal aseveración sea correcta, mencionando que dicha media es aún menor. Con el fin de contrastar la afirmación reúne una muestra aleatoria de 100 alumnas para registrar su peso, obteniéndose una media de 53.75 kilogramos. ¿Es esta evidencia suficiente para considerar que efectivamente la media es menor a lo asegurado?

$$H_o : \mu = 54.4$$
$$H_a : \mu < 54.4$$
$$Z_{prueba} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{53.75 - 54.4}{5.4 / \sqrt{100}} = -1.20$$



Como $Z_{prueba} = -1.20$ cae en la región de aceptación, la hipótesis nula es aceptada, por lo tanto, no hay suficiente evidencia para decir que la media de peso de las alumnas del CCH-N sea menor al 54.4 kilogramos con un nivel de significación del 5%.

Ejercicios de prueba de hipótesis

1. Carlos Avitia quien se dedica a la venta de accesorios de teléfonos celulares, está a punto de realizar una compra de un gran número de fundas para un modelo en particular, él advierte al vendedor que realizará un muestreo aleatorio de 120 fundas y de obtener un porcentaje mayor al 10% de fundas defectuosas, no realizará la compra. Al realizar dicho muestreo observa que 14 de las 120 fundas inspeccionadas son defectuosas. Con el muestreo realizado, ¿Carlos tiene suficiente evidencia para no realizar la compra a un nivel de significación del 5%?

2. Un comerciante de focos afirma que su producto tiene una vida útil de por lo menos 100 horas. Al seleccionar a 30 de sus focos se obtuvo una media de 95 horas, ¿hay suficiente evidencia para refutar lo afirmado por el comerciante, sabiendo que la desviación estándar es de 10 horas y usando un nivel de significación del 5%?

3. Se sabe que aproximadamente 1 de cada 10 fumadores prefieren cigarrillos de la marca A. Después de una campaña de promoción en una región de ventas dada, se entrevistó a una muestra de 200 fumadores para determinar la efectividad de la campaña. En la encuesta, un total de 26 personas expresaron su preferencia por la marca A. ¿Presentan estos datos suficiente evidencia que indique un aumento en la preferencia de la marca A en la región? Utiliza un nivel de significación del 5%.

4. Se lanza un dado 500 veces, de las cuales en 90 ocasiones se obtuvo el número 6, Utilizando un nivel de significación del 1%, ¿hay evidencia suficiente para suponer que el dado está cargado?

5. Un comunicador asegura que la publicidad tiene una influencia en el público no mayor al 78.4%. Un estudio realizado por una compañía de publicidad reveló que el 80% de un total de 500 consumidores realizaron sus compras influenciados por la publicidad. ¿Contradice el estudio el supuesto del consumidor, a un nivel de significación del 5%?

6. Una empresa afirma que la media de contenido neto de las cajas de cereal que vende es de 560 gramos, con una desviación estándar de 15 gramos. Si en un muestreo aleatorio de 50 cajas de este cereal, se obtuvo una media de 555 gramos, ¿existe suficiente evidencia para decir que la media es inferior a lo afirmado por la empresa? Utilizar un nivel de significación del 5%.

Propuesta de examen para la Unidad III Inferencia estadística

1. De 200 denuncias de robo de automóvil, 114 fueron de automóvil color blanco. Usa esta información para establecer un intervalo de confianza del 90% para estimar el porcentaje de denuncias de robo que se refieren a automóvil color blanco. Interpreta el resultado.

2. Se tomó, en estado de reposo, la frecuencia cardiaca a 50 personas mayores de 70 años. Se obtuvo una media de 90 latidos por minuto con una desviación estándar de 12. Estima por medio de un intervalo de confianza del 95%, la media de frecuencia cardiaca en estado de reposo para esta población de personas. Interpreta el resultado.

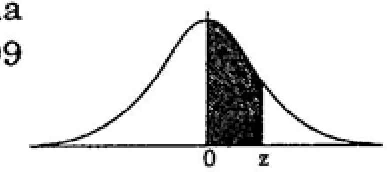
3. La media de delitos al día en el municipio de Naucalpan es de 5, con una desviación estándar de 2.77. Las autoridades del municipio han implementado en todo el municipio un programa titulado "patrullando tu colonia" y en un mes de 31 días se obtuvo una media de 4 delitos al día, ¿Podría pensarse que el programa está funcionando? Utiliza un nivel de significación del 5%

4. Cierto partido político afirma que el porcentaje de electores que votará a su favor en las próximas elecciones es de por lo menos 56%. De una muestra aleatoria de 200 electores, 110 mencionaron votarían por ese partido, ¿existe suficiente evidencia para refutar lo afirmado por el partido político? Utiliza un nivel de significación del 5%.

5. En una encuesta realizada en México se entrevistó a 800 adultos para preguntarles respecto si están de acuerdo con la política de seguridad "Abrazos no balazos" del presidente López Obrador. El 52% de los entrevistados estuvo a favor de la política, ¿Concluirías que la mayoría de los adultos en México está a favor de la política "Abrazos no balazos"? Utiliza un nivel de significación del 5%.

Áreas de la curva de probabilidad normal

Esta tabla proporciona las áreas bajo la distribución normal entre z = 0 a z = 3,99 en pasos de 0,01.



.	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4961	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000