GUÍA PARA PREPARAR EL EXAMEN EXTRAORDINARIO ESTADÍSTICA Y PROBABILIDAD II

Becerril Montes Helios
Candanosa Aranda Carlos
Lara Álvarez Alicia
León Cano María Eugenia

Mayo de 2019

INSTRUCCIONES PARA EL MANEJO DE ESTA GUÍA

Esta guía comprende todos los temas del programa, te recomendamos que el estudio deba de ser de la primera a la última página si se deseas un máximo de posibilidades de éxito en el examen extraordinario.

Está diseñada de tal manera que un estudiante puede resolverla sólo o en compañía de sus compañeros.

SUGERENCIAS DE ACTIVIDADES DE APRENDIZAJE TEÓRICO PRÁCTICAS

Al iniciar el estudio de un tema es conveniente concluirlo para lograr una visión completa de los conceptos y procedimientos que impliquen.

Está demostrado que el repaso es un hábito clave para el aprendizaje y esta guía no es la excepción. Al realizar un estudio sostenido, con repasos diarios de cada tema o unidad, estarás en el camino del aprendizaje perdurable y significativo.

UNIDAD 1. MODELOS DE PROBABILIDAD Y SUS APLICACIONES

Propósito

En esta unidad continuarás desarrollando tu pensamiento estadístico, apropiándote del concepto de variable aleatoria, y construyendo modelos de probabilidad en términos de su tendencia, variabilidad y distribución.

APRENDIZAJES
➢ Identifica el concepto de variable aleatoria. ➢ Diferenciarás 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.

1.1 Antecedentes

Continuando con los conceptos vistos en Estadística I, recordemos algunos de ellos.

Todos los hechos que ocurren se denominan **fenómenos**.

Un **fenómeno aleatorio** es aquel que tiene varios resultados posibles y estos no se pueden predecir con certeza, pues obedecen las leyes del azar. Por ejemplo, el CENAPRED[1] informó, en su reporte de las 16:00h, que en las últimas 24 horas del 23 de abril de 2019, se identificaron 55 exhalaciones, por medio de los sistemas de monitoreo del volcán Popocatépetl, ¿cuántas exhalaciones se reportarán durante las próximas 24 horas?

Para estudiar los fenómenos de nuestro entorno, lo que hacemos son experimentos. Un experimento es algún proceso u operación que lleva a resultados bien definidos.

Un **Experimento Aleatorio** es aquel con la característica de que se conocen los resultados que se pueden obtener, pero no es posible determinar cuál de esos resultados se obtendrá antes de realizar el experimento. Y esto sucede cada vez que el experimento se va a llevar a cabo. Por ejemplo, lanzar una moneda común, es una *operación o experimento*, si cae “sol (o águila)” ese es un *resultado*.

Un **resultado** es lo que se obtiene en un solo ensayo de un experimento.

Un **espacio muestral**, o espacio muestra, de un experimento es el conjunto de todos los posibles resultados distintos del experimento. Generalmente se denota por **Ω**, o por **S**.

Ejemplo.

Experimento: se lanza una moneda legal, al caer la moneda, se observa la cara superior.

El espacio muestral es: Ω = {“sol”, “águila”}

Un resultado es: “sol”

Ejemplo

[1] Centro Nacional de Prevención de Desastres.
http://www.cenapred.gob.mx/reportesVolcanGobMX/Procesos?tipoProceso=detallesUltimoReporteVolcan

Experimento: se lanza una moneda legal dos veces consecutivas, al caer, se observa la cara superior.

El espacio muestral es:

$$\Omega = \{\text{“sol-sol”, “águila-sol”, “sol-águila”, “águila-águila”}\}$$

Un resultado es “águila-sol”

Todos los resultados de cada uno de estos experimentos son fenómenos aleatorios, por lo que un resultado de cualquier experimento con frecuencia se denomina **punto muestral**. Estos eventos pueden clasificarse como **simples** o **compuestos**.

Un **evento aleatorio simple** es el resultado de un solo punto muestral en cualquier experimento.

Un **evento aleatorio compuesto** es un subconjunto del espacio muestral, que contiene dos o más puntos muestrales.

Ejemplo

Se lanza una moneda legal dos veces consecutivas, al caer, se observa la cara superior. Si cae sol escribimos ***s***, si cae águila escribimos **a**

El espacio muestral es **Ω = {ss, as, sa, aa}**

- Si interesa que caiga un sol en el primer lanzamiento y águila en el segundo lanzamiento, el suceso **{sa}** es un evento simple ya que es un único suceso del espacio muestral que cumple esa condición.
- Si interesa que caiga exactamente un sol, el suceso que cumple es **{as, sa}** que corresponde a un evento compuesto.

1.2 Variable aleatoria

Variable*.* Es la característica de interés en una población o muestra. Como las características no se mantienen constantes de un individuo a otro se les considera aleatorias.

Ejemplo

Se eligen al azar seis estudiantes de la clase de Computación del Colegio y se les pregunta cuánto tiempo tardan de su casa la escuela; la carrera que elegirán; cuantos hermanos tienen; su estatura; su peso.

Para el enunciado anterior, identifique las variables aleatorias son

- Tiempo de traslado de su casa a la escuela.
- La carrera que elegirán al concluir el bachillerato.
- El número de hermanos que tiene cada alumno
- La estatura de cada estudiante, en centímetros.
- El peso de cada estudiante, en kilogramos.

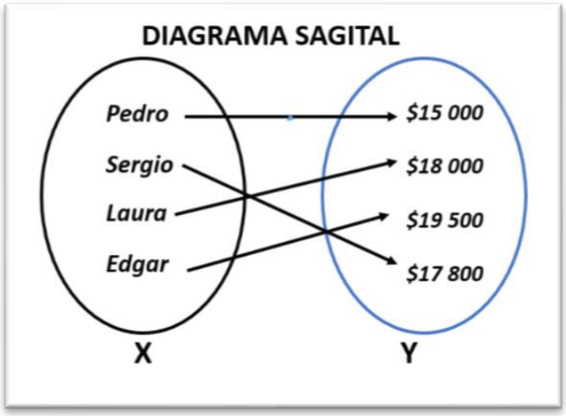
Una función es una asociación especial entre un elemento X de un conjunto, (llamado dominio), y un elemento Y de otro conjunto, (llamado contradominio), en la cual cada elemento de X se asocia con uno y solo uno de los elementos de Y.

Por ejemplo, legalmente a cada persona le corresponde un apellido; a cada alumno del colegio le corresponde un número de cuenta; al contar los árboles de un huerto, a cada árbol le hacemos corresponder un número natural; al medir la estatura de las personas, a cada persona le corresponde una estatura; etc.

Ejemplo

En una oficina del colegio se eligen a cuatro trabajadores especializados. A cada trabajador le hacemos corresponder solo un salario.

Esto se observa en el siguiente diagrama sagital



Definición de variable aleatoria.

Sea Ω un espacio muestral asociado con un experimento aleatorio. Una variable aleatoria continua X es una función que asocia a cada evento aleatorio simple del espacio muestral un número real $X(E_i) = x$, donde E_i son los eventos elementales simples de Ω

$$X : \Omega \rightarrow R$$

Indicaremos con R_X el *rango*, o *recorrido*, de la variable aleatoria, quiere decir el conjunto de valores numéricos de la variable aleatoria X

Variable aleatoria discreta

Una ***variable aleatoria es discreta*** si toma u número finito o infinito numerable de valores en un intervalo. Su recorrido R_X es un conjunto discreto. Se obtiene a través de un proceso de conteo.

Los siguientes enunciados corresponden a variables aleatorias discretas:

- Número de soles obtenidos al lanzar tres monedas distinguibles.
- Número de libros en la biblioteca.
- Número de lobos en una manada.
- Número de hermanos de los estudiantes del grupo de Estadística.

Ejercicio

Se lanza una moneda legal dos veces consecutivas al aire, al caer, se observan los soles de la cara superior.

a) Escribe el espacio muestral Ω.
b) Identifica la variable aleatoria X.
c) Relaciona los eventos aleatorios simples y sus valores correspondientes de la variable aleatoria X, a través de un diagrama sagital.
d) Escribe su recorrido R_X de la variable X
e) ¿Qué tipo de variable aleatoria es X?

Variable aleatoria continua

Una **variable aleatoria es continua** si puede tomar un número infinito de valores entre un intervalo dado.

Del ejemplo 3, “se eligen al azar seis estudiantes de la clase de Computación del Colegio y se les pregunta: <u>cuánto tiempo tardan de su casa la escuela</u>; la carrera que elegirán; cuantos hermanos tienen; <u>su estatura</u>; <u>su peso</u>.”

A partir de los enunciados subrayados, identificar las variables aleatorias y escribe su recorrido.

Variable aleatoria X	Recorrido R_X
○ Tiempo de traslado de su casa a la escuela, en minutos:	(20, 150)
○ Estatura de cada estudiante, en centímetros:	(145, 190)
○ Peso de cada estudiante, en kilogramos:	(50.4, 92.5)

Se observa que los resultados se obtuvieron al medir el tiempo de traslado, medir su estatura y medir el peso, de los estudiantes del colegio,

Estas características se obtuvieron a través de un proceso de medición. Su recorrido R_X es un conjunto no numerable, esto es, abarca un intervalo de números reales.

Ejercicio

Se lanza un dado cúbico legal dos veces consecutivas y se suman los puntos de la cara superior.

a) Escribe el espacio muestral del experimento aleatorio.
b) Identifica la variable aleatoria X.
c) Escribe su recorrido R_X de la variable X.

1.3 Distribución de probabilidad

La **distribución de probabilidad** de una variable aleatoria discreta X es una función que asocia a cada elemento de X su respectiva probabilidad. Se denota por P(x).

$$P(x): R \rightarrow [0,1]$$

Propiedades

Para la distribución de probabilidad de X se deben cumplir las siguientes propiedades

- $0 \leq P(x) \leq 1$
- $\sum_{x \in R_X} P(x) = 1$

Ejercicio

Se lanza una moneda corriente tres veces al aire, al caer, se observan los soles de la cara superior.

a) Define la variable aleatoria X de este experimento aleatorio.
b) ¿Cuáles son los eventos aleatorios simples y los valores correspondientes de la variable aleatoria X?
c) ¿Cuál es la distribución de probabilidad de la variable aleatoria X?

1.4 Esperanza matemática de la variable aleatoria discreta X

En Estadística, la esperanza matemática de una variable aleatoria X es su valor promedio, es decir, la media aritmética de los valores de la variable aleatoria (μ **)**. También se conoce como valor esperado.

La **Esperanza Matemática,** o **Valor Esperado,** de una variable aleatoria discreta X es igual a la suma de los diferentes valores de X, multiplicados por sus probabilidades correspondientes. Se denota por E[X], μ

$$\mu = E(x) = \sum_{i=1}^{n} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + \ldots + x_n P(x_n)$$

donde *n* es el número de diferentes valores que toma la variable aleatoria X.

Ejemplo

Sea X es la variable aleatoria definida como el número de soles al lanzar una moneda legal tres veces. ¿Cuál es el valor esperado de X?

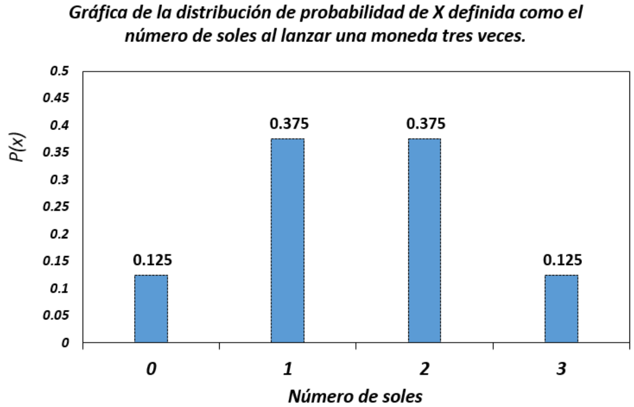
Al utilizar la fórmula anterior y sustituir los valores de la variable aleatoria y sus respectivas probabilidades, se obtiene:

$$\mu = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = 1.5$$

El valor esperado de 1.5 soles es el promedio, a largo plazo, de soles obtenidos al lanzar una moneda tres veces.

"Se espera, en promedio, obtener 1.5 soles al lanzar una moneda tres veces."

x	P(x)	x*P(x)
0	0.125	0
1	0.375	0.375
2	0.375	0.75
3	0.125	0.375
	1	1.5



1.5 Desviación estándar de la variable aleatoria discreta X.

La desviación estándar es una medida de la variación de todos los valores de una variable aleatoria X.

Se requiere obtener la varianza.

La varianza de X se define como la suma de los cuadrados de las desviaciones de los distintos valores de la variable aleatoria X respecto del valor esperado µ, multiplicado por sus probabilidades correspondientes. Se denota por $\boldsymbol{\sigma^2}$

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i) \qquad \text{donde } \mu = E(x)$$

que es equivalente a

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 P(x_i) - \left(\sum_{i=1}^{n} x_i P(x_i) \right)^2 \qquad \text{o} \qquad \sigma^2 = E(x^2) - E^2(x)$$

La desviación estándar de la variable aleatoria X se define como la raíz cuadrada positiva de la varianza. Se denota por σ

Ejemplo

Se lanza una moneda legal tres veces consecutivas al aire, al caer, se observan los soles de la cara superior. ¿En promedio, que tan lejos están los valores de la media?

Solución

Se requiere calcular primero la varianza. Se utiliza la expresión anterior

$$\sigma^2 = (0-1.5)^2 \times \frac{1}{8} + (1-1.5)^2 \times \frac{3}{8} + (2-1.5)^2 \times \frac{3}{8} + (3-1.5)^2 \times \frac{1}{8} = 0.75\, soles^2$$

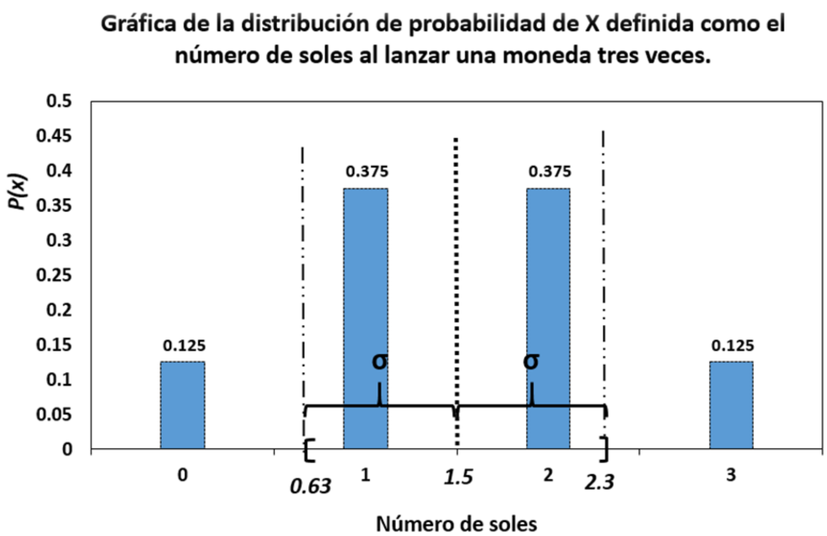
O

$$\sigma^2 = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} - 1.5^2 = 0.75\, soles^2$$

Y la desviación estándar de X, es

$$\sigma = \sqrt{0.75\,soles^2} \approx 0.866\;soles$$

Por lo tanto, en promedio, se desvían aproximadamente 0.866 soles alrededor de la media



Ejercicio

Toño Lee participa en un torneo de tae kwon do. En la primera etapa, “todos contra todos”, participara en *tres* encuentros y por cada partido en que resulte vencedor recibirá un premio en efectivo; de esa manera, al descontar sus gastos, saldría *ganando* $4000 por partido, pero en caso de ser derrotado en un partido, nada recibiría, lo cual le provocaría una *pérdida* de $1000. Si el torneo está tan parejo que su probabilidad de ganar es igual a la de perder.

Realiza lo siguiente

a) Identifica la Variable aleatoria
b) Escribe el espacio muestral Ω.
c) Relaciona los eventos aleatorios simples y sus valores correspondientes de la variable aleatoria X, a través de un diagrama sagital.
d) Escribe su recorrido R_X de la variable aleatoria X
e) Construye su distribución de probabilidad de la variable X
f) ¿Debe Toño esperar ganancia de dinero después de sus tres partidos?
g) ¿En promedio, que tan lejos están los valores de la media de ganancia de dinero?

Ejercicio

Al realizar una investigación acerca del uso del automóvil en carretera, se contó el número de ocupantes en cada uno de 1000 automóviles y se organizaron los datos en la tabla de la derecha.

a) Identifica la variable aleatoria X
b) Construye una tabla de distribución de probabilidad para la variable aleatoria X.
c) ¿Cuántos ocupantes en promedio iban en los automóviles?
d) ¿Cuánta dispersión tuvo la variable aleatoria X: ocupantes en un automóvil?

Ocupantes en un automóvil	Número de automóviles
1	208
2	318
3	169
4	188
5	92
6	25

Ejercicios adicionales.

1. Un jugador lanza un dado legal. Si sale un número primo gana dicho número de pesos, pero si no sale un número primo entonces pierde ese número de pesos. Los resultados posibles ***xi*** del juego con sus respectivas probabilidades ***P(xi)*** son como sigue:

xi	2	3	5	-1	-4	-6
P(xi)	1/6	1/6	1/6	1/6	1/6	1/6

Los números negativos –1, -4, y –6 corresponden al hecho de que el jugador pierde si sale un número primo.

a) Identifica la variable aleatoria X.
b) Construye una tabla de distribución de probabilidad para la variable aleatoria X.
c) ¿Cuál es el valor esperado del juego?
d) ¿Cuánta dispersión tuvo la variable aleatoria X?

2. Completa la tabla:

Fenómeno o Experimento	**Variable Aleatoria X**	**Tipo de Variable Aleatori a**	**Valores que puede tomar la Variable Aleatoria X**
Lanzar la pelota 5 veces a la canasta en un juego de básquetbol.	La pelota entra a la canasta.	Discreta	
Comprar tres computadoras.	Computadora defectuosa.		
Las mariposas “monarca” llegarán a Michoacán el próximo año.	Número de mariposas “monarca“ que morirán en Michoacán el próximo año.		X = 0, 1, 2, 3, . . .(1)
Se mide la estatura de un adulto.			Cualquier valor entre 50 y 300 cm

Encuestar a 100 personas sobre su preferencia por un candidato a la presidencia.	Proporción de la preferencia por un candidato.	Continua	
Deportación a ilegales en Estados Unidos de Norteamérica.	Número de indocumentados mexicanos que deportará Estados Unidos el próximo mes.		X = 0, 1, 2, 3, . . .
Vacunar a 67 personas contra la Hepatitis B.	Número de casos de hepatitis B.		
Sorteo con 12 premios mayores de la Lotería Nacional si se vendieron todos los boletos.	Premios que otorgará la Lotería Nacional.		

3. A un hospital de Venezuela acuden de 0 a 4 enfermos de hepatitis C en un mes cualquiera. Empleando un registro histórico se obtuvieron las probabilidades de que acuda una cantidad de enfermos determinada al mes (véase la tabla).

a) ¿Cuál es la cantidad esperada de enfermos de hepatitis C?

b) ¿Cuál es su dispersión?

Cantidad deEnfermos de hepatitis C	**Probabilidad P(x)**
0	0.10
1	0.35
2	0.17
3	0.30
4	0.08

4. Una empresa de bienes raíces tiene un registro de los terrenos vendidos en los 365 días del año anterior (véase la tabla).

Número de días del año anterior	**Terrenos vendidos en el día**	**Probabilidad de los terrenos vendidos en un día**
210	0	
70	1	
30	2	
15	3	
30	4	
0	5	
10	6	

Supón que el experimento consiste en seleccionar un día del año y que se define X como el número de terrenos vendidos en un día.

a) Completa la tabla. Al hacerlo habrás obtenido la distribución de probabilidad para X.

b) ¿Cuánto suman las probabilidades de que sucedan los valores de X?

c) Si las condiciones de venta fueran similares este año, la empresa ¿qué promedio de terrenos debería esperar vender?

d) ¿Qué tanta dispersión tuvieron las ventas por día del año pasado?

e) Por medio de una gráfica muestra cómo es la distribución de probabilidad de X.

f) ¿Qué porcentaje de valores de X están comprendidos en el intervalo cerrado $[-\sigma, \sigma]$ centrado en el valor esperado μ?

5. Un juego consiste en lanzar dos dados de distinto color, de tal manera que:

a) Si la suma de los números es múltiplo de 3, gano $200.00
b) Si la suma es 7, gano $100.00
c) Si la suma no es 7 ni múltiplo de 3, pierdo $50.00

¿Debo esperar ganancia o pérdida?, ¿cuánto?

6. Sea X el número de caras obtenidas al tirar tres monedas ideales. Obténgase el valor esperado de X.

7. Una tienda de artículos electrónicos vende un modelo particular de computadora portátil. Hay solo cuatro computadoras en el almacén, y el gerente se pregunta cuál será la demanda el día de hoy para este modelo particular. El departamento de mercadotecnia le informa que la distribución de la probabilidad para x, la demanda diaria para la computadora portátil, es la siguiente:

X	0	1	2	3	4	5
f(x)	0.10	0.40	0.20	0.15	0.10	0.05

Determine la media, la varianza y desviación estándar de x.

8. Una lotería organizada a beneficio del cuerpo de bomberos venderá 8000 boletos de 5 dólares cada uno. El premio es un automóvil de 12,000 USD. Si usted compra dos boletos, ¿cuál es su ganancia esperada?

9. Una variable aleatoria x puede asumir cinco valores: 0, 1, 2, 3, 4. Se muestra enseguida una parte de la distribución de probabilidad:

X	0	1	2	3	4
P(x)	0.10	0.30	0.30	¿?	0.10

a) Escribe las propiedades de una distribución de probabilidad.
b) Encontrar P(3)
c) Calcular la media, varianza y desviación estándar
d) ¿cuál es la probabilidad de que x sea mayor que 2?
e) ¿cuál es la probabilidad de que x sea 3 o menor?

1.6 Distribución Binomial

Aprendizajes
➢ 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 propia distribución o de tablas. ➢ Deduce que en el caso de variables aleatorias continuas, la probabilidad debe calcularse para valores dentro de un intervalo.

Iniciemos con el Experimento de Bernoulli que consiste en un experimento aleatorio en el que sólo se pueden obtener dos resultados: éxito y fracaso.

Jacob Bernoulli (1654 – 1705) Matemático y científico suizo. Sus contribuciones a la geometría analítica, a la teoría de probabilidades, al, Cálculo diferencial, a la teoría de números, etc

Por ejemplo, se le pregunta a un alumno si el pasado fin de semana fue al cine, su respuesta es “si” o “no”; si lanzas una moneda, al caer, su resultado puede ser “sol” o “águila”; si el semáforo de alerta volcánica del Popocatépetl continua en *Amarillo fase 3* o no; si en un accidente automovilístico hubo heridos o no; si un nuevo fármaco para el tratamiento de la diabetes mellitus tipo 1 ha tenido el éxito esperado o no; etc.

A un resultado de este experimento se le denomina *éxito* y tiene una probabilidad de ocurrencia p y al otro resultado se le denomina *fracaso* y tiene una probabilidad de $q = 1-p$.

La distribución de probabilidad de Bernoulli es $f(x) = p^x(1-p)^{1-x}$

Experimento Binomial

El modelo de Bernoulli forma la base para la distribución binomial.

Ejemplo

Se lanza una moneda legal tres veces consecutivas al aire, al caer, se observan los soles de la cara superior. ¿Cuál es la probabilidad de que caigan exactamente dos soles?

Solución

Del espacio muestral, los eventos aleatorios simples, tienen la peculiaridad que de los tres lanzamientos, dos son soles.

sas
ssa → **2 soles**
ass

Si cae sol se considera un éxito, si cae águila se considera un fracaso.

La probabilidad de que caiga sol en una sola ejecución, es decir, en un solo lanzamiento es: $p = \frac{1}{2}$

Por lo tanto, la probabilidad de obtener dos soles al lanzar la moneda tres veces, queda:

$$p^2q + p^2q + p^2q = 3\,p^2q$$

Al sustituir la probabilidad de obtener un sol es: $p = \frac{1}{2}$

Y la probabilidad de no obtener sol es $q = 1 - p = \frac{1}{2}$

$$P(obtener\ dos\ soles) = 3\,p^2q = 3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) = \frac{3}{8}$$

Para este ejemplo se consideró lo siguiente:

1. Las veces que se lanza la moneda son tres.
2. Resultado: éxito: que caiga en la cara superior “sol”
   Fracaso: que no caiga en la cara superior “sol”
3. Los resultados de cada lanzamiento son independientes.
4. Probabilidad de éxito, es decir, la probabilidad de que caiga “sol” en una sola ejecución es $p=\frac{1}{2}$
   Probabilidad de fracaso, es decir, la probabilidad de que no caiga “sol” en una sola ejecución es $q = 1 - p = \frac{1}{2}$
5. La variable aleatoria binomial se define como el número de veces que cae “sol” al lanzar la moneda tres veces.
6. Los valores de la variable aleatoria binomial son: X = 0, 1, **2**, 3. Interesa X= 2

Características de un proceso binomial

- ✓ El experimento aleatorio se puede repetir n veces.
- ✓ Cada prueba que se realiza tiene solo dos posibles dos resultados: Éxito y Fracaso.
- ✓ El resultado de cada prueba es independiente entre sí.
- ✓ La probabilidad de éxito es la misma en cada repetición. Se denota por *p*. y la probabilidad de fracaso es también la misma en cada repetición. Se denota por *q*; *p+q=1*
- ✓ La variable aleatoria binomial se define como el número *X* de éxitos obtenidos en *n* pruebas realizadas
- ✓ Los valores de la variable aleatoria binomial son: X = 0, 1, 2, 3, 4, . . . , n

El modelo matemático para la distribución binomial con *n* ensayos y probabilidad de éxito *p* para una variable aleatoria X es:

$$P(n, p; x) = C_x^n p^x q^{(n-x)}$$

Donde $C_x^n = \frac{n!}{(n-x)!x!}$

Este modelo matemático se utiliza para calcular la probabilidad de obtener *x* éxitos en un número fijo de intentos, *n*, siempre y cuando se conozca la probabilidad de un éxito aislado *p*.

Propiedades

7. $P(n,p;\, x) \geq 0$
8. $0 \leq P(n,p;\, x) \leq 1$
9. $\sum_{x=0}^{n} B(n, p; x) = \sum_{x=0}^{n} C_x^n p^x q^{n-x} = 1$

Valor esperado de la variable aleatoria binomial La Esperanza Matemática, o Valor Esperado**,** de una variable aleatoria discreta X de tipo binomial, es

$$E(x) = \mu = np$$

Desviación estándar de la variable aleatoria binomial. La **Varianza** de la variable aleatoria binomial se obtiene al aplicar la expresión matemática

$$\sigma^2 = npq$$

Por lo tanto, la desviación estándar para una variable aleatoria binomial es

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

Ejemplo

Juan Manuel Lozano es acupunturista y ha descubierto un tratamiento para curar la obesidad, de tal manera que dos de cada tres pacientes que lo siguen logran curarse al disminuir su peso hasta un valor adecuado a su estatura. Próximamente cuatro pacientes se someterán a su tratamiento y se interesa por conocer la probabilidad de que:

a) Se cure uno.
b) Se curen dos.
c) Se curen al menos tres.
d) No se cure ninguno.
e) También se interesa por saber cuál es el número más probable de pacientes que se curará.
f) ¿Qué dispersión tiene la variable x?

¿Se puede considerar como un experimento de tipo binomial?

Solución

Para resolverse, debe cumplir con las características de un experimento binomial.

Características de un proceso binomial	
✓ El experimento aleatorio se puede repetir n veces. ✓ Cada prueba que se realiza solo son posibles dos resultados: Éxito y Fracaso. ✓ El resultado de cada prueba es independiente entre sí. ✓ La probabilidad de éxito es la misma en cada repetición. Se denota por *p*. ✓ La probabilidad de fracaso es la misma en cada repetición. Se denota por *q*; Se cumple *p+q=1* ✓ La variable aleatoria binomial se define como el número *X* de éxitos obtenidos en *n* pruebas realizadas ✓ Los valores de la variable aleatoria binomial son: X = 0, 1, 2, 3, 4, ..., n	-*Éxito*: que el paciente al someterse al tratamiento se cure. -*Fracaso*: que el paciente al someterse al tratamiento no se cure. -El resultado de cada prueba es independiente. -La probabilidad de que el paciente al someterse al tratamiento se cure: $p=\frac{2}{3}$ -La probabilidad de que el paciente al someterse al tratamiento no se cure: $q=\frac{1}{3}$ Se debe cumplir $p+q=\frac{2}{3}+\frac{1}{3}=1$ La variable aleatoria binomial se define como el número de pacientes que se someten al tratamiento y se curan -Su recorrido X = 0, 1, 2, 3, 4 Por lo tanto, se considera que es un experimento de tipo binomial

a) La probabilidad de que se cure un paciente significa que *x* = 1, con $p=\dfrac{2}{3}$ y $q=\dfrac{1}{3}$

$$P(n=4, p=\frac{2}{3}; x=1)=P(x=1)=C_1^4\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^3=\frac{8}{81}\approx 0.0987$$

Por lo tanto, la probabilidad de que se cure un paciente al someterse al tratamiento 4 personas es de aproximadamente 0.0987.

b) La probabilidad de que se curen dos pacientes significa que sucede *x* = 2:

$$P(n=4, p=\frac{2}{3}; x=2)=P(x=2)=C_2^4\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2=\frac{24}{81}\approx 0.2962$$

Por lo tanto, la probabilidad de que se curen dos pacientes al someterse al tratamiento 4 personas es de aproximadamente 0.2962

c) La probabilidad de que se curen al menos tres pacientes significa que sucede *x* = 3 ó *x* = 4; por lo tanto, se suman las probabilidades:

$$P(n=4, p=\frac{2}{3}; x=3, x=4)=P(x=3)+P(x=4)=C_3^4\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1+C_4^4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^0=\frac{32}{81}+\frac{16}{81}=\frac{48}{81}\approx 0.5925$$

Por lo tanto, la probabilidad de que se curen al menos tres pacientes al someterse al tratamiento de un total de 4 personas es de aproximadamente 0.5925

d) Para calcular la probabilidad de que no se cure paciente alguno: *x* = 0.

$$P(n=4, p=\frac{2}{3}; x=2) = P(x=0) = C_0^4\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^4 = \frac{1}{81} \approx 0.01234$$

Por lo tanto, la probabilidad de que se no se cure ningún paciente al someterse al tratamiento 4 personas es de aproximadamente 0.01234.

e) El número más probable de pacientes que se curarán se obtiene al calcular el valor esperado

$$\mu = np = 4\times\frac{2}{3} = \frac{8}{3} = 2.\overline{6}\ \textit{pacientes}$$

Se espera que se curen en promedio, aproximadamente 2.66 pacientes.

(Bajo el contexto, el número de pacientes que se espera que se curen son 3)

f) Su dispersión se calcula a partir de la desviación estándar:

$$\sigma = \sqrt{npq} = \sqrt{4\times\frac{2}{3}\times\frac{1}{3}} = \sqrt{\frac{8}{9}} \approx 0.9428\ \textit{pacientes}$$

En promedio se desvía aproximadamente una unidad alrededor de la media

Ejercicio

Una distribuidora automotriz dispone de 8 autos para demostración, 5 son rojos y 3 azules. A diferentes horas de cierto día se presentan 4 clientes y cada uno selecciona al azar un auto de los 8 para manejarlo a prueba.

a) ¿Cuál es la probabilidad de que los clientes seleccionen menos de 2 autos rojos?
b) ¿Qué valor tiene la media aritmética de la distribución?
c) ¿Cuál es su dispersión de la variable binomial?

Se trata de un experimento aleatorio **con reemplazo**, puesto que cada cliente tiene la posibilidad de seleccionar un auto de los 8 disponibles.

¿Se puede considerar como un experimento de tipo binomial?

Ejercicio

En un lote con 1,000 relojes de pulsera hay 10% defectuosos. Si se escogen al azar 7, ¿cuál es la probabilidad de que haya más de 1 defectuoso?

¿Se puede considerar como un experimento de tipo binomial?

Ejercicios adicionales.

1**.** ¿La erupción del volcán de Colima es un fenómeno equivalente a un experimento de Bernoulli? ¿Por qué? [Hint: consulta las características de un experimento binomial]

2**.** De acuerdo con una encuesta del Instituto Mexicano de la Juventud, el 42% de los jóvenes mexicanos con edades entre 12 y 29 años se declara católico practicante.

Se realiza una selección al azar de 7 jóvenes con edades en ese rango:

a) ¿Es válido afirmar que la selección de un joven es un experimento de Bernoulli?

b) Calcula la distribución de probabilidad (los valores de la probabilidad para cada valor de la variable aleatoria con n = 7).

c) Calcula la probabilidad de que sean católicos practicantes:

i. Todos
ii. Ninguno
iii. Al menos 5
iv. Más de 2 pero menos de 6

a) ¿Cuál es el valor de la esperanza matemática?

3**.** Se extraen 4 canicas **con reemplazo** de una urna que tiene 5 blancas y 3 negras. ¿Cuál es la probabilidad de que salgan menos de 2 blancas?

4. Se conoce que 10% de cierta población es daltónica. Si se selecciona una muestra de 15 personas al azar de esta población, calcule las probabilidades siguientes:

a) 4 o menos sean daltónicos

b) 5 o más sean daltónicos

c) entre 3 y 6 inclusive sean daltónicos

5**.** Se conoce que el 95% de las piezas producidas por una máquina son perfectas. Se toma una muestra de 12 piezas al azar y se desea saber la probabilidad de que exactamente 9 sean perfectas.

6**.** Durante un largo tiempo se ha observado que un soldado puede dar en el blanco con un solo disparo con probabilidad igual a 0.80. Suponga que dispara cuatro tiros al blanco.

a) ¿cuál es la probabilidad de que dé en el blanco exactamente dos veces?

b) ¿cuál es la probabilidad de que dé en el blanco al menos una vez?

7**.** Obtener la media, varianza y distribución estándar de X, definidas como el número de caras obtenidas en la tirada de cuatro monedas balanceadas.

8**.** Suponga que el 70% de todos los pacientes que han tomado cierta medicina se curarán, ¿cuál es la probabilidad de que de una muestra de 30 pacientes que han tomado la medicina se curen?

1.7 DISTRIBUCIÓN NORMAL

APRENDIZAJES
✓ Utiliza la tabla para valores bajo la curva de la distribución normal estandarizada como el recurso para el cálculo de probabilidades o de valores de z para dicha distribución ✓ Contrasta la gráfica para una situación de comportamiento aproximadamente normal con su correspondiente gráfica en el modelo estandarizado ✓ Calcula probabilidades por medio de la distribución normal estandarizada dentro de problemas contextualizados, interpretando los resultados

Introducción

Como se vio en secciones anteriores, una variable numérica se puede clasificar como discreta o continua, dependiendo de si toma un número contable o no contable de valores.

Una variable continua se asocia con la idea de "medir" utilizando fracciones y decimales, además de enteros, a diferencia de una discreta que se asocia con la idea de "contar".

Cuando la variable es continua, el modelo probabilístico más utilizado es la Distribución Normal, su importancia se debe principalmente a que hay muchos fenómenos que siguen el comportamiento de esta distribución:

- Caracteres morfológicos de individuos de una especie (personas, animales, plantas,...): talla, peso, longitud, diámetro, perímetro,...
- Caracteres fisiológicos: efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.
- Caracteres sociológicos: consumo de cierto producto por un mismo grupo de individuos, puntuaciones de examen.
- Caracteres psicológicos: cociente intelectual, grado de adaptación a un medio,...
- Errores cometidos al medir ciertas magnitudes.
- Valores estadísticos muestrales: la media, la proporción.
- Otras distribuciones como la Binomial o la de Poisson se pueden aproximar a una Normales
- En general cualquier característica que se obtenga como suma de muchos factores.

Ejemplo

Un biólogo realiza un estudio acerca del peso del lobo mexicano *canis lupus baileyi*, una especie en extinción. Después de varios meses de trabajo logra medir los pesos de 90 machos, datos que quiere analizar para conocer cómo se comportan y determinar la probabilidad de que al seleccionar uno al azar resulte de cierto peso en un intervalo determinado.

Peso en kg de 90 lobos mexicanos machos

29	33	31	31	33	35	40	29	32	31	36	43	38	35	27
43	34	25	35	34	36	37	33	28	35	35	37	33	35	36
36	33	31	29	33	33	32	32	34	34	38	33	33	36	38
32	30	40	34	39	36	26	36	36	32	37	34	40	27	33
37	35	42	37	29	35	33	36	35	37	39	37	35	34	32
27	33	28	31	44	37	34	30	29	39	31	38	33	31	31

Solución
La variable peso es continua, aunque aquí se discretizó.

Los datos se organizan en una tabla de frecuencias por intervalos y se construyen su histograma y su polígono de frecuencias. (Figura 1.3.1)

Peso	Frecuencia
24.5 – 26.5	2
26.5 – 28.5	5
28.5 – 30.5	7
30.5 – 32.5	14
32.5 – 34.5	21
34.5 – 36.5	19
36.5 –38.5	12
38.5 – 40.5	6
40.5 – 42.5	3
42.5 – 44.5	1

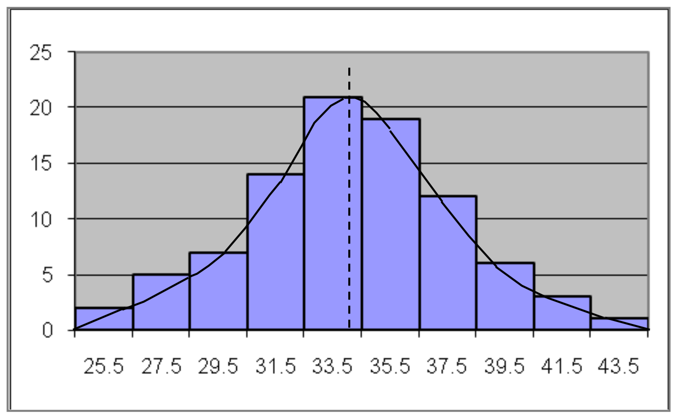


Figura 1.3.1

Obsérvese que el polígono de frecuencias es casi simétrico con respecto a un eje vertical central. Cuando la simetría es perfecta, la figura es conocida como *curva normal* o *campana de Gauss*.

Propiedades de la Distribución Normal

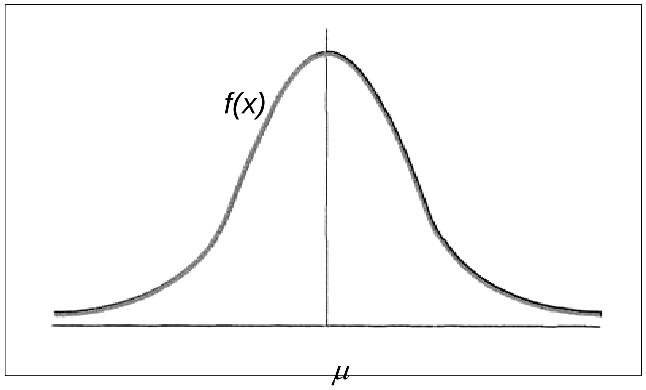
Se dice que una variable aleatoria continua X tiene distribución normal (o gaussiana), de parámetros μ y σ, si su función de densidad es del tipo:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

donde μ es la media de la población, σ es la desviación estándar de la población y x es un valor de la variable continua

Como e y π son constantes, la forma de la curva normal depende solamente de los dos parámetros de la distribución normal, μ y σ.

Si se grafica la función anterior, para valores fijos de μ y σ se obtiene una campana de Gauss centrada en μ.



Cuando una variable aleatoria X tiene distribución normal, se representa simbólicamente como $X \sim N(\mu, \sigma^2)$ y tiene las propiedades que se indican a continuación.

- Toma valores desde $-\infty$ hasta $+\infty$.
- Su gráfica es simétrica con respecto a la recta vertical que pasa por μ
- Es asintótica al eje x: nunca lo toca por mucho que se extienda tanto a la derecha como a la izquierda.
- La esperanza matemática o valor esperado coincide con μ, y la varianza es σ^2:

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

- Cuando se modifica μ la ubicación de la campana cambia pero mantiene su forma (figura 1.3.2).
- Cuanto menor sea σ, mayor probabilidad se concentrará alrededor de la media, cuanto mayor sea sigma, más dispersa es la distribución (figura 1.3.3)
- El valor de la media μ coincide con los valores de la mediana y la moda.
- La probabilidad de que la variable tome valores en un intervalo (a, b), equivale al área bajo la curva de gauss en dicho intervalo.
- El área total bajo la curva equivale a una probabilidad de 1.

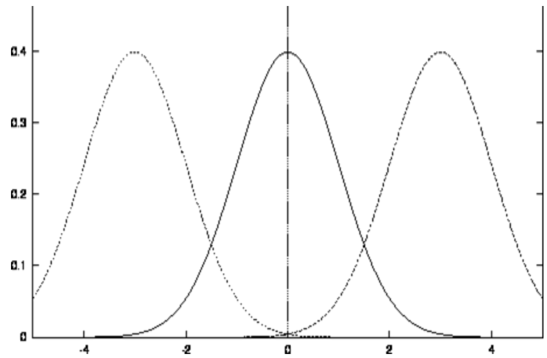


Figura 1.3.2 Distribuciones normales con distintas medias pero igual dispersión.

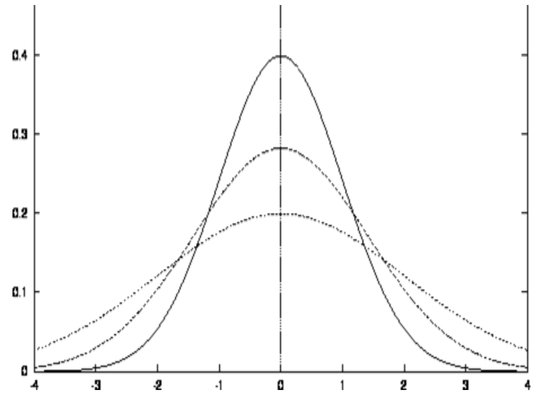


Figura 1.3.3 Distribuciones normales con distintas dispersiones pero igual media.

1.8 Distribución Normal Estándar

Puesto que hay un número infinito de combinaciones para los dos parámetros, hay un número infinito de curvas normales diferentes, pero existe una distribución Normal muy particular e importante llamada Estándar y se reserva el uso de Z para representar a la variable aleatoria correspondiente (en lugar de X). Algunas de sus propiedades son

* La media de la distribución es cero $\mu = 0$
* La desviación estándar o típica es igual a uno. $\sigma = 1$
* La curva es simétrica con respecto al eje X
* El área bajo la curva corresponde a la probabilidad en el intervalo dado
* Las siguientes tres características son conocidas como *regla empírica* (figura 1.3.4)

a) el 68% de la probabilidad se encuentra a una desviación estándar de la media, en ambas direcciones, es decir, de – 1 a 1

b) el 95% se halla a dos desviaciones estándar, o sea entre – 2 y 2

c) y el 99.7% a tres desviaciones estándar, es decir, de – 3 a 3

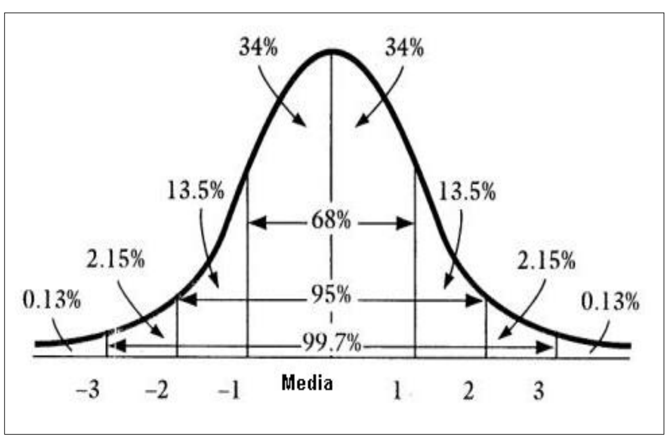


Figura 1.3.4

Para obtener la probabilidad no es necesario calcular el área directamente, para eso existen tablas con valores ya calculados.

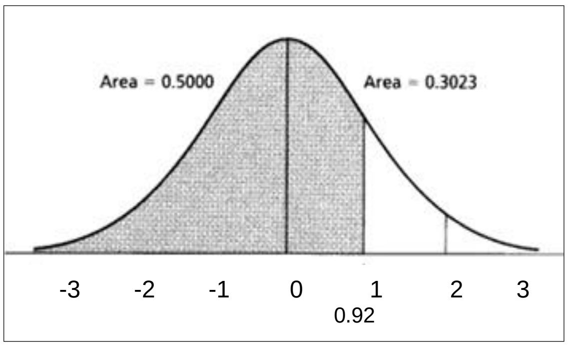
Existen diferentes tablas de distribución normal estándar, pero en el fondo son los mismos valores, sólo que con distintas presentaciones. En esta guía se te explica cómo utilizar una de tales versiones, pero ten presente que en el examen extraordinario te pueden proporcionar otra versión, así que te recomendamos que lleves tu propia tabla.

Ejemplos

1) Si Z es una variable aleatoria que tiene distribución Normal Estándar, para determinar la probabilidad de que Z sea menor de 0.92 y su gráfica, se procede como sigue

Aunque los valores de Z van desde – ∞ hasta ∞, basta con graficar más o menos de – 3.5 a 3.5, pues, fuera de dicho intervalo, la probabilidad es extremadamente pequeña.

Una vez trazada la curva, se sombrea el área que corresponde a la probabilidad buscada.



Observa que el área en este ejemplo es justo como muestra la imagen en la tabla, debido a esto, se busca de la siguiente manera:

* en la primera columna de la tabla se busca el valor de Z hasta décimos

* en ese renglón, se lee hacia la derecha hasta encontrar la columna con el valor de los centésimos,

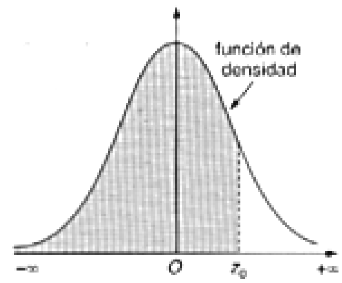
* la probabilidad se encuentra en la intersección de ambas direcciones.

Por lo tanto, la probabilidad buscada es:

P (Z < 0.92) = 0.8212   (Ver la tabla en la siguiente figura)

TABLA DE LA DISTRIBUCIÓN NORMAL $N(0, 1)$

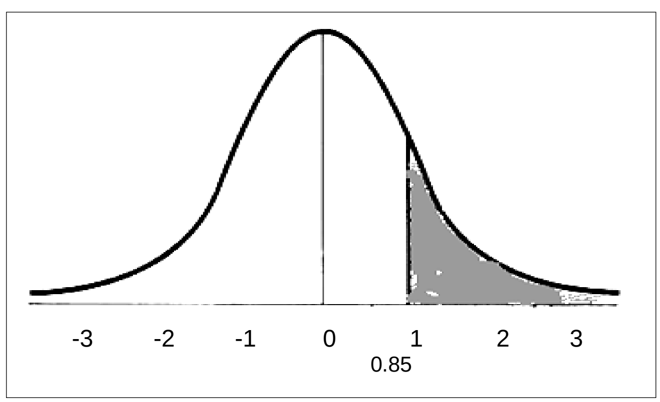
$P(Z \leq z_0) = \left\{ \begin{array}{c} \text{área del recinto} \\ \text{coloreado} \end{array} \right\}$



z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	$\boxed{.8461}$	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

2) Calcular la probabilidad de que Z sea mayor de 0.85

Se traza la gráfica y se sombrea el área correspondiente



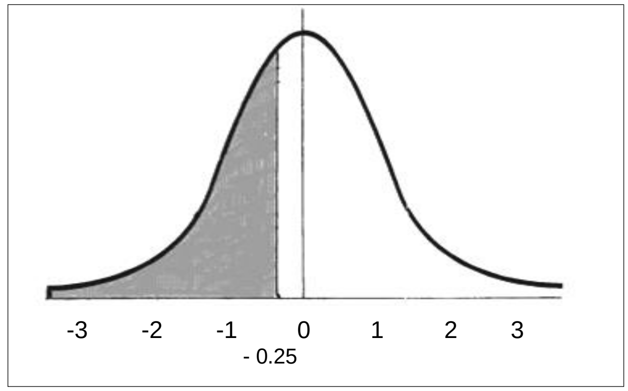
Obsérvese que esta vez, es el área no sombreada la que se asemeja al área de la tabla, por lo que al buscar en ella se obtiene la probabilidad “complementaria” no la indicada.

Como el área total bajo la curva es igual a 1, entonces la probabilidad buscada se obtiene restando apropiadamente.

$$P(Z > 0.85) = 1 – P(Z < 0.85)$$
$$= 1 – 0.8023 = 0.1976$$

3) Calcular la probabilidad de que Z sea menor que – 0.25

Al trazar la gráfica y sombrear el área correspondiente se obtiene la imagen de la derecha.



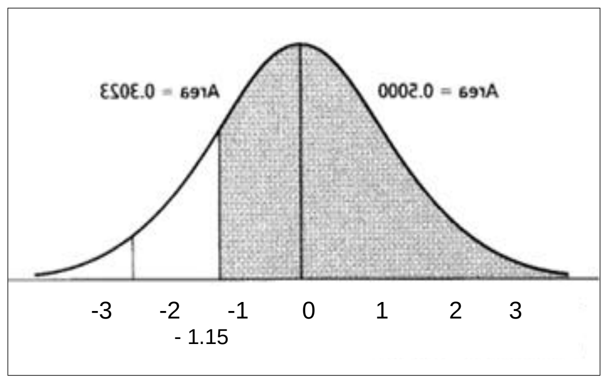
Si bien la tabla no tiene valores de Z negativos, es posible calcular el área indicada debido a la simetría de la curva. Reflejando verticalmente la situación e interpretando el valor de z como si fuese positivo, se obtiene el área no sombreada.

Como el área total bajo la curva es igual a 1, restando los valores apropiados se obtiene la probabilidad buscada.

P(Z < – 0. 25) = 1 – P(Z < 0.25) = 1 – 0. 5887 = 0. 4013

4) Calcular la probabilidad de que Z sea mayor – 1.15.

Al trazar la gráfica y sombrear el área correspondiente, se obtiene la imagen de la derecha.

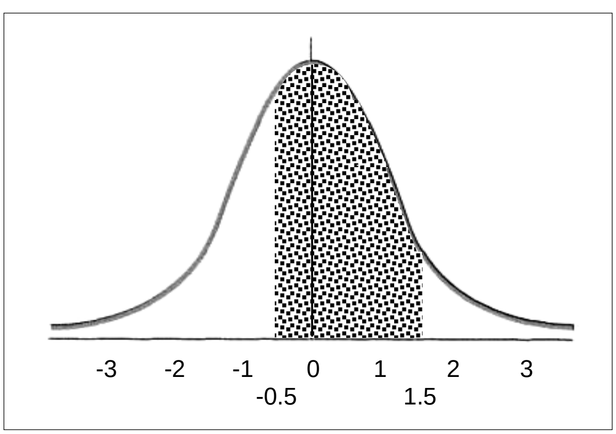


Aunque el valor de z es negativo, por la simetría de la curva la probabilidad buscada es equivalente a la probabilidad reflejada, considerando z positiva.

Por lo tanto, esta vez no se resta con el 1.

P(Z > – 1.15) = P(Z < 1.15) = 0.8749

5) Calcular la probabilidad de que Z tome valores mayores a – 0.5 pero menores a 1.5.



Como se desea el área intermedia, la probabilidad buscada se obtiene restando ambas probabilidades “parciales”.

P(-0.5 < Z < 1.5) =

= P(Z < 1.5) –(1 – P(Z ≤ 0.5)) =

= 0.9332 – 0.6915 = 0.2417

1.9 Estandarización

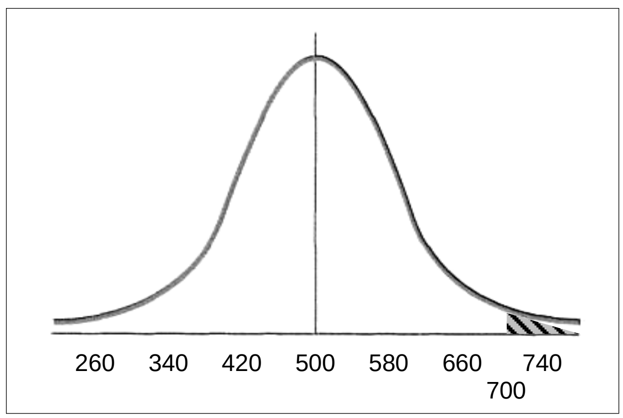
Cuanto una variable aleatoria tiene distribución Normal pero con parámetros distintos de 0 y 1, y se desean calcular sus probabilidades, se recurre a un procedimiento llamado *estandarización*, que significa determinar una nueva variable aleatoria con base en la original, pero que tenga distribución normal estándar.

Teorema
Si una variable aleatoria X tiene distribución Normal con parámetros μ y σ^2, entonces la variable aleatoria $Z = \frac{X-\mu}{\sigma}$ tiene distribución Normal estándar.

Ejemplo
Considérese que las calificaciones en cierta prueba de aprendizaje, aplicada a todos los graduados de educación media superior en una ciudad, tienen distribución normal, con una media de 500 puntos y una desviación estándar de 80 puntos.
Si se selecciona al azar una prueba, ¿cuál es la probabilidad de que la calificación sea mayor de 700 puntos?

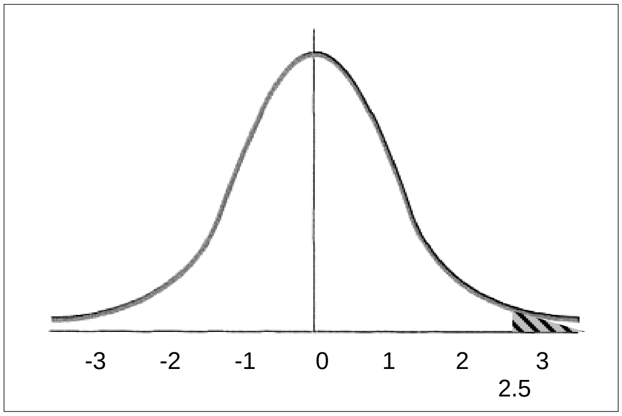
Solución.
Variable aleatoria: calificación en la prueba de aprendizaje, X.
Información: $X \sim N(500, 80^2)$.

Si se grafica la campana de Gauss para X y se sombrea el área que corresponde, se observaría la siguiente imagen



Se estandariza la variable aleatoria X realizando la transformación $Z = \frac{X-\mu}{\sigma}$

P(X > 700) = P($\frac{x-\mu}{\sigma} > \frac{700-500}{80}$) = P(Z > 2.5)



Si ahora se traza la gráfica correspondiente a Z, se observa la misma imagen, pero con escala diferente, debido a la estandarización.

Por lo tanto, utilizando la tabla presentada se obtiene
P(X > 700) = P(Z > 2.5) = 1 – 0.9938 =
= 0.0062
Es decir, la probabilidad de que la prueba elegida a la azar tenga una calificación mayor a 700 puntos es de 0.0062

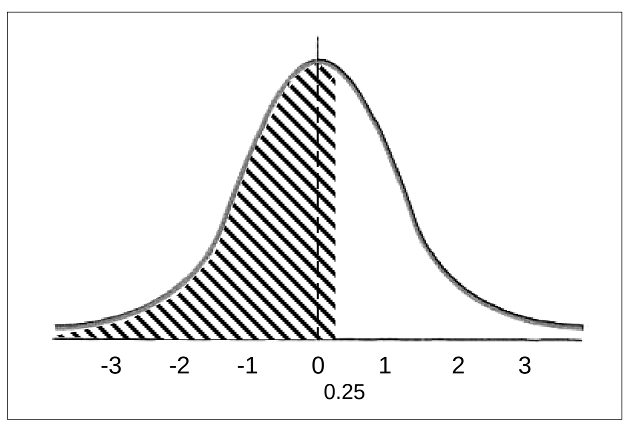
Ejemplo
Regresando al tema del peso de lobos mexicanos que se te presento en la introducción de esta sección, considera que se distribuye de manera Normal con una media de 34.05 kg y una desviación estándar de 3.68 kg. Si se elige al azar a uno de tales lobos, a) ¿cuál es la probabilidad de que pese menos de 35 kg? b) ¿con que probabilidad tendrá un peso de 28 a 37 kg?

Solución
X ~ N (34.05 , 3.68^2).

a) P(X < 35).
Se grafica la campana de Gauss para X y se sombrea el área que corresponde o se estandariza primero y sólo se traza la gráfica para Z, pues ya se sabe que la imagen y el área son las mismas, y se estandariza la variable aleatoria X realizando la transformación $Z = \frac{X-\mu}{\sigma}$

P(X < 35) = P($\frac{x-\mu}{\sigma} < \frac{35-34.05}{3.68}$)

= P(Z < 0.25) = 0.5987



Por lo tanto, la probabilidad de que al elegir al azar un lobo mexicano su peso sea menor a 35 kg es del 0.5987.

Dicho de otra manera, el 59.87 % de los lobos mexicanos que sean elegidos al azar tendrán un peso menor a 35 kg.

b) P(28 < X < 37).

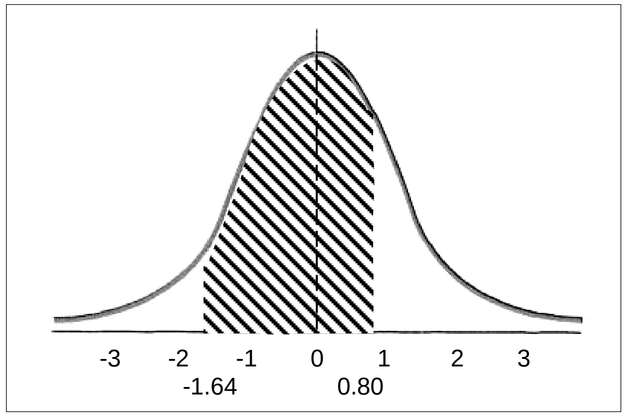
Se escribe el equivalente de esta probabilidad, aprovechando la simetría de la distribución como: P(28 < X < 37) = P(X < 37) – P(X < 28), y se estandariza la variable aleatoria X realizando la transformación $Z = \frac{X-\mu}{\sigma}$ para ambos valores

P(X < 37) = P($\frac{x-\mu}{\sigma} < \frac{37-34.05}{3.68}$) = P(Z < 2.51)

P(X < 28) = P($\frac{x-\mu}{\sigma} < \frac{28-34.05}{3.68}$) = P(Z < 0.80)

De este modo se obtiene

(28 < X < 37) = P(X < 37) – P(X < 28)
= P(Z < 0.80) – P(Z < –1.64)
= 0.7881 – 0.0505
= 0.7376



Por lo tanto, si se elige al azar un lobo mexicano, la probabilidad de que pese entre 28 y 37 kg es de 0.7376

Dicho de otra manera, el 73.76 % de los lobos mexicanos que sean elegidos al azar tendrán un peso entre 28 kg y 37 kg.

Ejercicios

1. La tabla a continuación muestra la distribución de las temperaturas, en grados centígrados, de 194 hembras pastor alemán al momento del parto.

Dibuja el histograma, traza su polígono de frecuencias "suavizado" y describe qué observas.

Temperatura	Frecuencia
37.15 – 37.35	6
37.35 – 37.55	13
37.55 – 37.75	19
37.75 – 37.95	26
37.95 – 38.15	37
38.15 – 38.35	35
38.35 – 38.55	25
38.55 – 38.75	17
38.75 – 38.95	11
38.95 – 39.15	5

2. Los estudios muestran que el uso de gasolina para automóviles compactos vendidos en Estados Unidos tiene una distribución normal, con una media de 25.5 millas por galón (mpg) y una desviación estándar de 4.5 mpg. ¿Qué porcentaje de automóviles compactos consigue 30 mpg o más?

3. Si las concentraciones de colesterol total para cierta población están distribuidas en forma aproximadamente normal, con una media de 200 unidades (medidas en mg/100 ml), y una desviación estándar de 20 unidades, determina la probabilidad de que un individuo de dicha población seleccionado al azar tenga una concentración de colesterol:

a) Entre 170 y 230 unidades.
b) Mayor de 155 unidades.

4. Los paquetes de carne, en cierto supermercado, indican en su etiqueta que contienen 1 kg. Cómo se sabe, los pesos de los paquetes varían debido al proceso de envasado, por lo que algunos contienen un poco más de un kilo y otros un poco menos. Si el peso de los paquetes de carne tiene una distribución normal con media de 1 kg y desviación estándar de 15 g, calcula:

a) el porcentaje de paquetes que pesará más de un kilogramo
b) la probabilidad de que un paquete elegido al azar pese entre 0.950 y 1.030 kg

5. Para cierta población de trabajadores, el número de días de incapacidad al año tiene distribución normal con media de 5.4 y desviación estándar de 2.8 días. Determina la probabilidad de que un trabajador elegido al azar en esa población tenga, al año, incapacidad de:

a) Más de 6 días
b) Entre 4 y 6 días
c) Menos de 5 días

UNIDAD 2. ESTIMADORES E INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

Propósito:
El alumno analizará el comportamiento de los estimadores media y proporción, a través del modelo Normal, para construir un vínculo entre la Probabilidad y 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 y de la proporción muestrales 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 de la proporción muestrales, obtenidas por medio de la computadora o la calculadora, dentro del contexto de un problema o una investigación.

2.1 Población y muestra

Muestreo aleatorio simple, con y sin reemplazo.

De acuerdo con lo estudiado con anterioridad, el muestreo es un concepto importante para la inferencia estadística.

El muestreo puede ser probabilístico y no probabilístico, siendo el aleatorio simple el que nos interesa en este curso, el cual puede ser sin reemplazo, donde cada individuo de la población tiene la misma posibilidad de ser elegido para la muestra sólo una vez o con reemplazo donde cada individuo de la población tiene la misma posibilidad de ser elegido para la muestra más de una vez.

Así, el número de muestras de tamaño n que se pueden obtener de una población de N individuos elegidos sin reemplazo es C_n^N y el número n de muestras de tamaño n posibles que se pueden obtener con reemplazo es N^n.

Como tema central de esta unidad tenemos a las distribuciones muestrales, concepto que es fundamental en la comprensión de la inferencia estadística, las cuales podemos definir como la distribución de todos los valores posibles que pueden ser adquiridos por algún estadístico, calculado a partir de muestras de igual tamaño extraídas aleatoriamente de la misma población.

Una distribución muestral puede construirse cuando de una población finita, discreta, de tamaño N, se extraen todas las muestras posibles de tamaño n, enseguida se calcula el estadístico de interés de cada muestra y finalmente se construye una tabla de frecuencias con los valores observados del estadístico y su frecuencia correspondiente.

2.2 Distribución de la media muestral

Si se tiene la población $S = \{2, 4, 6, 8\}$

Para calcular la media aritmética de la población:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

al sustituir

$$\mu = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

La varianza de la población, se define como:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

al sustituir

$$\sigma^2 = \frac{(2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2}{4} = \frac{9+1+1+9}{4} = \frac{20}{4} = 5$$

La desviación estándar de la población es una medida de la variación de todos los valores de una variable aleatoria, con respecto a la media aritmética de la población.
La desviación estándar se define como la raíz cuadrada positiva de la varianza.

La desviación estándar de la población es:

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{5} \approx 2.2360679$$

Muestreo sin reemplazo

De la población, S, se eligen muestras de tamaño dos, sin reemplazo, y se calcula la media de cada muestra.
El número de muestras posibles de tamaño 2, (n=2) es $C_2^4 = 6$, las cuales son:

$m_1 = \{2, 4\} \quad \bar{x}_1 = 3$

$m_2 = \{2, 6\} \quad \bar{x}_2 = 4$

$m_3 = \{2, 8\} \quad \bar{x}_3 = 5$

$m_4 = \{4, 6\} \quad \bar{x}_4 = 5$

$m_5 = \{4, 8\} \quad \bar{x}_5 = 6$

$m_6 = \{6, 8\} \quad x_6 = 7$

La media aritmética de las medias muestrales $\bar{x}_i$ es:

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{m} \bar{x}_i}{m}$$

Al calcular se tiene:

$$\mu_{\bar{x}} = \frac{3+4+5+5+6+7}{6} = \frac{30}{6} = 5$$

La varianza de las medias muestrales $\bar{x}_i$, se calcula a partir de:

$$\sigma_{\bar{x}}{}^2 = \frac{\sum_{i=1}^{m}(\bar{x}_i - \mu_{\bar{x}})^2}{m}$$

$$\sigma_{\bar{x}}^2 = \frac{(3-5)^2+(4-5)^2+(5-5)^2+(5-5)^2+(6-5)^2+(7-5)^2}{6} = \frac{4+1+0+0+1+4}{6} = \frac{10}{6} = \frac{5}{3} = 1.\overline{6}$$

Por lo tanto, la desviación estándar de las medias muestrales $\bar{x}_i$ es:

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2}$$

$$\sigma_{\bar{x}} = \sqrt{\tfrac{5}{3}} = 1.29$$

Muestreo con reemplazo

El número de muestras posibles de tamaño 2, (*n*=2) extraídas con reemplazo es $4^2 = 16$, las cuales son

$m_1 = \{2,4\}$ $\bar{x}_1 = 3$ $m_9 = \{4,2\}$ $\bar{x}_9 = 3$

$m_2 = \{2,6\}$ $\bar{x}_2 = 4$ $m_1 = \{6,2\}$ $\bar{x}_{10} = 4$

$m_3 = \{2,8\}$ $\bar{x}_3 = 5$ $m_1 = \{8,2\}$ $\bar{x}_{11} = 5$

$m_4 = \{4,6\}$ $\bar{x}_4 = 5$ $m_1 = \{6,4\}$ $\bar{x}_{12} = 5$

$m_5 = \{4,8\}$ $\bar{x}_5 = 6$ $m_1 = \{8,4\}$ $\bar{x}_{13} = 6$

$m_6 = \{6,8\}$ $\bar{x}_6 = 7$ $m_1 = \{8,6\}$ $\bar{x}_{14} = 7$

$m_7 = \{2,2\}$ $\bar{x}_7 = 2$ $m_1 = \{6,6\}$ $\bar{x}_{15} = 6$

$m_8 = \{4,4\}$ $\bar{x}_8 = 4$ $m_1 = \{8,8\}$ $\bar{x}_{16} = 8$

Al calcular la media aritmética de las medias muestrales $\bar{x}_i$, se obtiene:

$$\mu_{\bar{x}} = \frac{3+4+5+5+6+7+2+4+3+4+5+5+6+7+6+8}{16} = \frac{80}{16} = 5$$

La varianza de las medias muestrales $\bar{x}_i$queda:

$$\sigma_{\bar{x}}{}^2 = \frac{2(3-5)^2+3(4-5)^2+4(5-5)^2+1(8-5)^2+3(6-5)^2+2(7-5)^2+1(2-5)^2}{16}$$

$$= \frac{8+3+0+9+3+8+9}{16} = \frac{40}{16} = \frac{5}{2}$$

Como puedes observar la media aritmética poblacional es igual a la media de las muestras de tamaño igual extraídas con reemplazo y de la media de las muestras extraídas sin reemplazo.
Sin embargo, la varianza de las medias extraídas con reemplazo es igual a la varianza poblacional σ entre el tamaño de muestra n y la varianza de las medias muestrales extraídas sin reemplazo es igual a la varianza poblacional entre el tamaño de muestra multiplicado por $\frac{N-n}{N-1}$ llamado corrección (o factor de corrección por población finita y puede ignorarse cuando el tamaño de la muestra es pequeño, comparado con el tamaño de la población.
A la raíz cuadrada de la varianza de las medias muestrales $\sqrt{\sigma_{\bar{x}}^{2}}=\sqrt{\frac{\sigma^{2}}{n}}=\frac{\sigma}{\sqrt{n}}$ se le llama ***error estándar de la media***, o simplemente, error estándar.
Cuando el muestreo es a partir de poblaciones normalmente distribuidas, las distribuciones de las medias muestrales, cumplirá con:
a) La distribución de la $\bar{x}$ será normal.
b) La media de las medias muestrales de tamaño igual $\mu_{\bar{x}}$ será igual a la media μ de la población de la cual se extranjeros las muestras
c) La varianza $\sigma_{\bar{x}}^{2}$ de la distribución de medias muestrales, será igual a la varianza de la población, de la cual se extrajeron las muestras, dividida entre el tamaño de muestra.
Cuando la población no está distribuida normalmente, haremos uso del Teorema del Límite Central o Teorema Central del Límite, que establece:
Para una población de cualquier comportamiento funcional, con una media μ y una varianza finita σ^{2}, la distribución de las medias muestrales $\bar{x}$ de igual tamaño n extraídas de esa población, estará distribuida aproximadamente en forma normal con $\mu_{\bar{x}}$ igual a la media de la población y varianza $\sigma_{\bar{x}}^{2}$ igual a la varianza de la población dividida entre el tamaño de muestra, siempre y cuando el tamaño de las muestras sea grande.
Para calcular probabilidades, utilizaremos el mismo procedimiento visto en el cálculo de probabilidades en la Distribución Normal, sólo que el cálculo se z, se realizará como sigue:
Para distribución de medias muestrales $z_{\bar{x}}=\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ que será equivalente a $z_{\bar{x}}=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Lo anterior, nos lleva a buscar un comportamiento estadístico de las medias muestrales, lo que encontraremos en el Teorema del Límite Central.

Ejemplos

a) Para una población que se distribuye normalmente.
Si se tiene que el gasto de los alumnos en pasajes diarios para asistir a la escuela se distribuye normalmente con media y desviación típica de $25.00 y $8.00 respectivamente. Determina la probabilidad de que una muestra de alumnos de tamaño 16 tenga un gasto promedio por día en pasajes de al menos de $21.50

Solución:

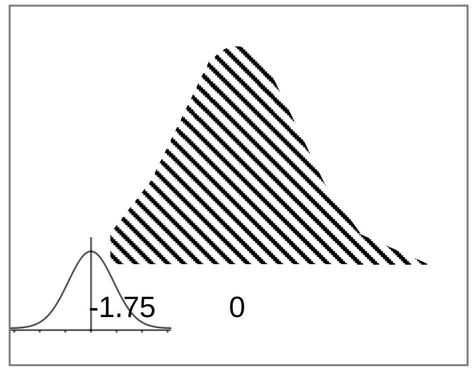
$P(\bar{x}\geq 21.50)$

Como la población se distribuye normalmente, el tamaño de muestra no afecta a la distribución de las medias muestrales y el valor de z es

$$z_{\bar{x}} = \frac{21.50 - 25}{\frac{8}{\sqrt{16}}} = -1.75$$

Por lo tanto $P(\bar{x} \geq 21.50) = P(z \geq -1.75)$

Al trazar la gráfica, se observa que la probabilidad buscada queda a la derecha del valor de Z



Consultando la tabla de áreas bajo la curva normal estándar, se obtiene

$$P(\bar{X} \geq 21.50) = P(Z \geq -1.75) = 0.9599$$

b) Para una población de la cual no se sabe cuál es su distribución.

Se sabe que tiempo promedio dedicado por los empleados de una empresa para realizar ejercicio es de 40 minutos con una desviación típica de 9 minutos, si se elige al azar una muestra de tamaño 36 de esos empleados, determina la probabilidad de que en promedio:

dediquen menos de 37 minutos para realizar ejercicio:

Solución:

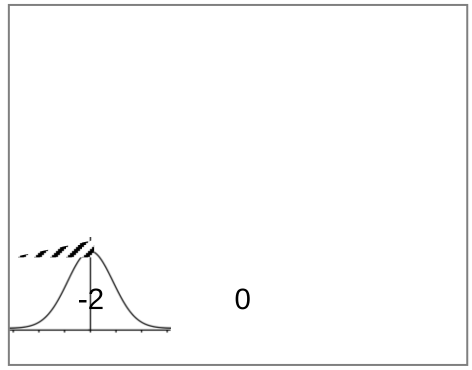
$P(\bar{x} < 37)$

Como el tamaño de muestra es grande, apegados a Teorema del Límite Central, la distribución de las medias muestrales será aproximadamente normal y el valor de z es

$$z_{\bar{x}} = \frac{37 - 40}{\frac{9}{\sqrt{36}}} = -2$$

Por lo tanto $P(\bar{x} < 37) = P(z < -2)$

Al trazar la gráfica, se observa que la probabilidad buscada queda a la izquierda del valor de Z



Consultando la tabla de áreas bajo la curva normal estándar, se obtiene

$$P(\overline{x} < 37) = P(z < -2) = 0.0228$$

Ejercicios

1. Dada una población normalmente distribuida con una $\mu = 100$ y una $\sigma^2 = 400$, determina para una muestra de tamaño 40:
   a) $P(\overline{x} \geq 100) =$
   b) $P(96 \leq \overline{x} \leq 108) =$
   c) $P(\overline{x} > 110) =$
2. Dado una distribución normal con $\mu = 50$ y $\sigma = 16$ y $n = 64$, determina:
   a) $P(46 \leq \overline{x} \leq 55) =$
   b) $P(\overline{x} > 53) =$
   c) $P(\overline{x} < 48) =$
   d) $P(45 \leq \overline{x} \leq 49) =$
3. Por estudios realizados, se sabe que en cierta población humana grande, la longitud craneal está distribuida aproximadamente en forma normal con una media de 185.6 mm. y una desviación estándar de 12.7 mm. ¿Cuál es la probabilidad de obtener una muestra aleatoria de tamaño 10 de esa población que tenga una media:
   a) mayor de 190 mm.?
   b) de al menos 180 mm.?
   c) mayor de 182 mm. y menor de 189 mm.?
4. Suponiendo que se sabe que los salarios por hora de cierto tipo de empleados bancarios de limpieza están distribuidos aproximadamente en forma normal con una media y una desviación típica de $45.00 y $ 5.00 respectivamente. Si se selecciona al azar una muestra de tamaño 16 empleados de esa población, determina la probabilidad de que la media del salario por hora para la muestra sea:
   a) mayor que $42.50.
   b) entre $43.00 y $47.50.
   c) de por lo menos $ 48.00.
   d) a lo más de $ 42.00.
5. Se ha encontrado que siguiendo un periodo de entrenamiento, el tiempo medio requerido por ciertas personas impedidas para realizar una tarea particular es de 25 segundos con una desviación estándar de 5 segundos. Determina la probabilidad de que una muestra de 32 individuos proporcionen una media:
   a) de 26 segundos o más.

b) entre 24 y 27 segundos.
c) a lo más de 23 segundos.
d) mayor de 22 segundos.

6. Si las concentraciones de ácido úrico en los adultos de sexo masculino están distribuidos aproximadamente de manera normal con una $\mu = 5.7$ mgs. por ml (mg/ml) y una $\sigma = 1$ mg/ml, determinar la probabilidad de que una muestra de tamaño 9 tenga una media:
   a) mayor que 6 mg/ml.
   b) entre 5 y 6 mg/ml.
   c) menor que 5.2 mg/ml.
   d) entre 5.8 y 6.1 mg/ml.
   e) menor de 5.2 o mayor de 5.9 mg/ml.

2.3 Distribución de la proporción muestral.

Así también, en la distribución de las proporciones de la muestra, si el tamaño de muestra es grande, es aplicable el Teorema del Límite Central, por lo que la media de las medias de las proporciones muestrales, será igual a la proporción verdadera de la población, es decir, $\mu_{\overline{p}} = p$ y la varianza de la distribución de proporciones muestrales será igual a $\frac{p(1-p)}{n}$, que para considerar el tamaño de muestra grande se debe cumplir con que los productos np y $n(1-p)$ sean ambos ser mayores de 5.

Así, para estandarizar, tenemos

$$z_{\overline{p}} = \frac{\overline{p} - \mu_{\overline{p}}}{\sigma_{\overline{p}}}$$

que es equivalente a

$$z_{\overline{p}} = \frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Ejemplos

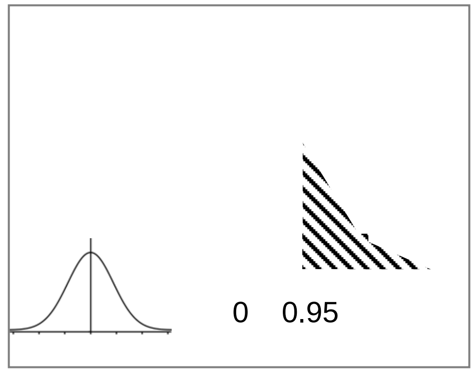
a) Para una población que se distribuye normalmente.
Si de los alumnos de sexto semestre se tiene la información que solo el 31% de ellos egresan sin haber reprobado materias durante sus estudios en el CCH ¿Si la proporción de materias reprobadas, se distribuye normalmente, cuál es la probabilidad de que en una muestra aleatoria de 10 alumnos que egresan se obtenga un porcentaje de ellos que no hayan reprobado materias de al menos mayor 45%?

Solución:
$P(\overline{p} \geq 0.45)$
Como la población se distribuye normalmente, el tamaño de muestra no afecta a la distribución de las proporciones muestrales y el valor de z es

$$z_{\overline{p}} = \frac{0.45 - 0.31}{\sqrt{\frac{0.31(1-0.31)}{10}}} = 0.95$$

Por lo tanto $\quad P(\overline{p} \geq 0.45) = P(z \geq 0.95)$

Al trazar la gráfica, se observa que la probabilidad buscada queda a la derecha del valor de Z



Por lo tanto

$$P(\overline{p} \geq 0.45) = P(z \geq 0.95) = 0.1711$$

Para una población de la cual no se sabe cuál es su distribución.

Se sabe que en una colonia del sur de la CDMX la proporción de lectores de revistas de espectáculos es de 0.43. ¿Cuál es la probabilidad de que una muestra aleatoria de 25 habitantes de esa colonia, se encuentre entre el 30% y el 62% de lectores de revistas de espectáculos?

Solución:
$P(0.30 < \overline{p} < 0.62)$

Como el tamaño de muestra no es grande, apegados al Teorema del Límite Central, la distribución de las medias muestrales será aproximadamente normal porque $25 \times 0.43 = 10.75$ y $25 \times 0.57 = 14.25$, ambos productos mayores de 5 y los valores de z son

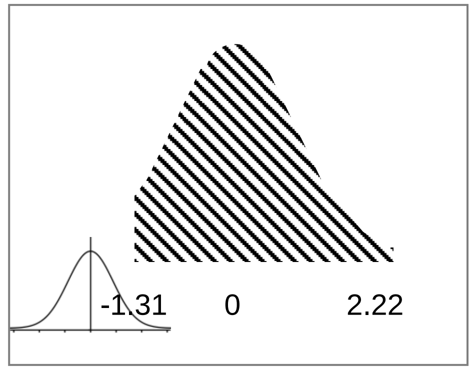
$$z_{\overline{p}} = \frac{0.30 - 0.43}{\sqrt{\frac{0.43(1-0.43)}{25}}} = -1.31 \qquad \text{y} \qquad z_{\overline{p}} = \frac{0.62 - 0.43}{\sqrt{\frac{0.43(1-0.43)}{25}}} = 2.22$$

Por lo tanto $P(0.30 < \overline{p} < 0.62) = P(-1.31 < z < 2.22)$

Al trazar la gráfica, se observa que la probabilidad buscada queda comprendida entre los dos valores de Z

Por lo que se obtiene

$$P(0.30 < \overline{p} < 0.62) = P(-1.31 < z < 2.22) = 0.8917$$



Ejercicios

1. Dada una población en la que $p = 0.6$ y una muestra aleatoria de tamaño 100 extraída de
   esa población, encontrar:
   a) $P(\overline{p} \geq 0.65) =$
   b) $P(\overline{p} \leq 0.58) =$
   c) $P(0.56 < \overline{p} \leq 0.66) =$

2. Si en una población de 500 mujeres, se encontró que 75 de ellas están sometidas a cierta dieta, ¿cuál es la
   probabilidad de que una muestra de tamaño 100 dé una proporción de aquellas mujeres que se encuentran a dieta:
   a) mayor o igual que 0.2?
   b) entre 0.1 y 0.22?
   c) no mayor que 0.12?

3. En cierta ciudad, se observa que el 20% de las familias tienen al menos un miembro que está sufriendo de algún malestar debido a la contaminación del aire. Una muestra aleatoria de 150 familias dio $\tilde{p} = 0.27$. Si el valor del 20% es correcto, ¿Cuál es la probabilidad de obtener una proporción de la muestra así o mayor?

4. Supóngase que se sabe que en cierta población humana, el 0.08 son daltónicos. Si se eligen aleatoriamente 150 individuos de esta población, ¿cuál es la probabilidad de que la
   proporción de los que son daltónicos sean tan grande como 0.15?

5. Si se sabe que en cierta población de clientes de un banco el 90% de ellos ha tenido cuentas en otro banco y se extrae una muestra aleatoria de tamaño 200 de esa población ¿cuál es la probabilidad de que la proporción de la muestra que hayan tenido cuentas en otro banco sea:
   a) menor que 0.85?
   b) entre 0.87 y 0.89?
   c) al menos de 0.92?
   d) mayor de 0.86 pero menor de 0.93?
   e) menor de 0.84 o mayor de 0.88?

3. INFERENCIA ESTADÍSTICA

Propósito

El estudiante realizará inferencias formales sobre los valores de los parámetros, a partir del análisis de los estimadores, para fundamentar la toma de decisiones en una investigación estadística, consolidando la formación de su pensamiento estadístico.

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 proporción, bajo las condiciones del Teorema del Límite Central. ➢ Estimar la media y proporción poblacionales por medio del intervalo de confianza correspondiente. ➢ Analizar el concepto de pruebas de hipótesis, para una media y para una proporción, bajo las condiciones del Teorema del Límite Central. ➢ Construye pruebas de hipótesis para la media y para la proporción ➢ Mide la validez de una hipótesis estadística, en el contexto de una investigación

3.1 Introducción

La Inferencia estadística es la rama de la Estadística relacionada con el uso de los conceptos probabilísticos para manejar la toma de decisiones bajo condiciones de incertidumbre.
La inferencia estadística comprende básicamente dos áreas principales: estimación y prueba de hipótesis. La estimación consiste en "aproximar" los valores de los parámetros de la población bajo estudio, mientras que las pruebas de hipótesis constituyen el proceso de aceptar o rechazar declaraciones relacionadas con dichos parámetros poblacionales; ambos procedimientos con base en la información que arroja una muestra aleatoria tomada en dicha población.
En la unidad 2 se estudiaron los conceptos básicos de distribuciones muestrales que sirven como apoyo a la esta parte la estadística.

3.2 Estimación

La estimación forma parte de la vida cotidiana en un sin número de lugares y en los distintos campos del quehacer humano. Cuando tratamos de atravesar una calle, estimamos la velocidad del auto y nuestra propia rapidez en cruzar la calle. Luego de hacer rápidamente estos estimados, se decide si debemos atravesar o esperar. La estimación es un procedimiento de la inferencia estadística mediante el cual se realizan cálculos con los datos de una muestra para obtener valores o resultados que describan las características de la población.

Cuando se usa un estadístico muestral para estimar un parámetro de la población es llamado *estimador*. Por ejemplo, la media muestral se utiliza como un estimador de la media de la población.
Cuando se observa un valor numérico específico de estimador, se le llama un *estimado*. En otras palabras, un estimado es un valor específico observado de un estadístico muestral.
Ejemplo 1

Poblaciones, parámetros de la población, estimadores y estimados

Población en la que se está interesado	Parámetros de la población que se desean estimar	Estadística muestral que se usará como estimador	Estimado que se hace
Empleados en una gran organización	Rotación media por año.	Rotación media por un período de 1 mes	8.90% de rotación por un año
Candidatos para una posición gerencial	Educación formal media (años)	Educación media formal para cada 5 candidatos	1.20% años de educación formal

Se pueden hacer dos tipos de estimados acerca de la población: un estimado puntual y un estimado por intervalo. Un estimado puntual es un valor único que pretende estimar el valor del parámetro poblacional de interés; un estimado por intervalo es un conjunto infinito de valores limitado por dos extremos, en donde se pretende se encuentra el valor del parámetro bajo estudio.

Un estimado puntual es a menudo insuficiente porque está correcto o está errado, pero no se tiene una medida de su precisión al aproximar el verdadero valor del parámetro.

Un estimado por intervalo indica su error en dos formas: por la extensión de su rango y por la probabilidad de que el verdadero parámetro de la población esté dentro de ese rango.

Ejemplos de estimados puntuales

a) La longitud promedio de una muestra de 50 varillas producidas por la máquina "A" es 40.70 metros.

b) En una muestra de 200 votantes, se obtuvo que 75% están a favor del candidato demócrata.

Ejemplos de estimados por intervalos

a) La longitud promedio de todas las varillas producidas por la máquina "A", se encuentra entre 40.20 y 41.10 metros

b) La proporción de votantes que están a favor del candidato demócrata, se encuentra entre 70% y 78%.

3.3 Estimación puntual y por intervalos para la media de la población.

Estimación puntual

La media muestral $\bar{x} = \frac{\sum x}{n}$ es el mejor estimador puntual de la media μ de la población. Es no sesgada, consistente y el estimador más eficiente; y a medida que la muestra sea suficientemente grande, su distribución muestral puede ser aproximada por la distribución normal (Teorema del Límite Central).

Estimación por intervalo

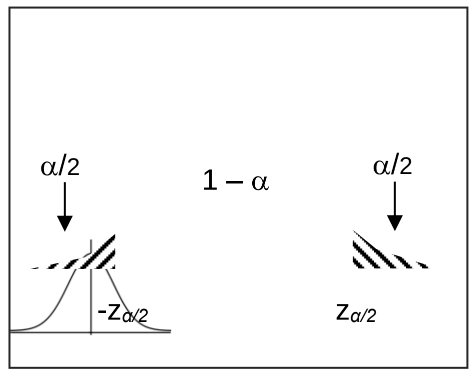
Con el fin de obtener una medida de la precisión de la estimación, se utiliza un intervalo, el cual incluirá el valor del parámetro con una probabilidad prefijada *1 – α*.

Comúnmente, a este intervalo de estimación se denomina *intervalo de confianza* del parámetro. El símbolo α representa la probabilidad de que el intervalo no contenga el verdadero valor del parámetro.

De una población normal, con media μ desconocida y varianza σ² conocida, se extrae una muestra aleatoria de n observaciones independientes. Como se sabe la variable aleatoria $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ tiene distribución normal estándar, por lo que $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, para un valor específico $z_{\alpha/2}$.

Para encontrar dicho valor específico, observa que es necesario tomar α/2 pues la probabilidad α queda repartida en las dos "colas" de la gráfica.

La probabilidad 1 – α aparece en blanco en la gráfica de la derecha



Nota:
Si se utiliza una tabla de "cola izquierda", como la presentada en la unidad 1, se tiene que sumar 1 – α más α/2 y buscar en la tabla de forma inversa.

Sustituyendo Z por su equivalente, se obtiene: $P(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$

Multiplicando los términos de la desigualdad por $\sigma/\sqrt{n}$, posteriormente por (–1) y sumando $\bar{x}$ a cada término, se obtiene $P(\bar{x} - z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{x} + z_{\alpha/2}(\sigma/\sqrt{n})) = 1 - \alpha$

Lo que se interpreta como: si la media de una población normal se estima usando la media de una muestra de tamaño *n*, entonces la probabilidad de que el intervalo

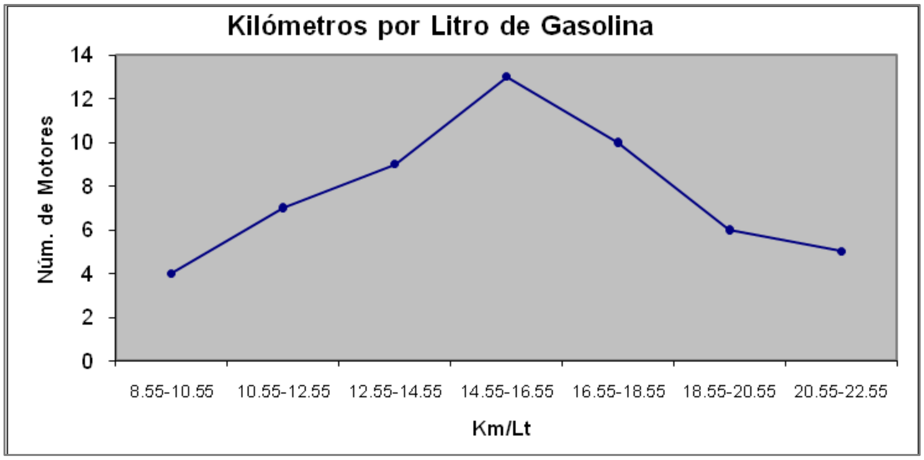
$$\left(\bar{x} - z_{\alpha/2}(\sigma/\sqrt{n}) \quad , \quad \bar{x} + z_{\alpha/2}(\sigma/\sqrt{n})\right)$$

contenga a la media de la población es 1 – α. Otra forma de entenderlo es: el (1 – α)% de los intervalos construidos de esta manera contendrán el valor del parámetro.
A la cantidad $z_{\alpha/2}(\sigma/\sqrt{n})$ se le conoce como error máximo de estimación de la media.

Ejemplo 2
El gerente de ventas, Carlos Soto, de una planta donde se arman motores de automóvil, sabe por resultados previos que el kilometraje por litro de gasolina se distribuye normalmente con una desviación típica o estándar de 1.6 km. por litro, pero con media desconocida, de la cual quiere hacer una estimación.

Para estimar μ, Carlos hace pruebas en una muestra de 54 motores recién armados y escogidos al azar. Después de hacer cálculos, obtiene un rendimiento promedio de $\bar{x}$= 14.7 Km. por litro.

Con base en el Teorema Central del Límite, él sabe que la variable $\bar{X}$: *promedio de kilómetros por litro de gasolina,* con *n* = 54 tiene distribución normal con desviación estándar $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.6}{\sqrt{54}} = 0.2177$ y media $\mu_{\bar{x}} = \mu$ desconocida.

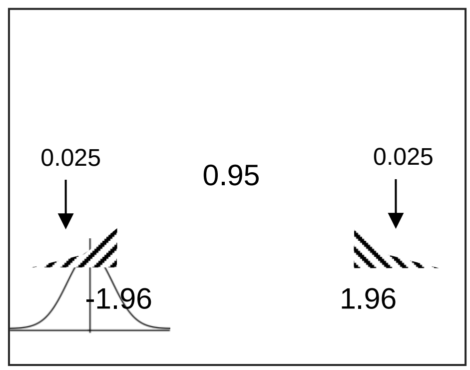


Si se fija una probabilidad (nivel de confianza) de 95% y se busca en la tabla de la distribución normal estándar el valor de z que le corresponde, se obtiene *z* = 1.96.

En la gráfica, la probabilidad de 0.95 está en blanco.

Por lo tanto P (–1.96 < Z < 1.96) = 0.95

Nota: 0.025 + 0.95 = 0.975 que en la tabla "de cola izquierda" presentada en la unidad 1 corresponde a z = 1.96.



Para cada valor de la variable continua $\bar{X}$, es posible calcular un correspondiente valor de la variable Z, utilizando la transformación $z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$. Carlos sustituye sus valores en esta igualdad y en el intervalo anterior y obtiene: $-1.96 < \frac{14.7 - \mu}{0.2177} < 1.96$

Al multiplicar la desigualdad por 0.2177, se obtiene: $-0.42669 < 14.7 - \mu < 0.42669$. Si resta 14.7, queda: $-15.1266 < -\mu < -14.2733$. Finalmente, al multiplicar por -1 e invertir el sentido de las desigualdades: $14.2733 < \mu < 15.1266$. Y esta es una estimación por intervalo para μ, al 95% de confianza.

Entonces, Carlos puede afirmar con 95% de confianza que el kilometraje medio poblacional está entre 14.2733 y 15.1266. Es decir, existe un 5% de probabilidad de que lo afirmado por Carlos no sea verdad.

Carlos observa que el valor $\bar{x}$= 14.7 *Km./l.* calculado a partir de su muestra de 54 motores de automóvil es justo el punto medio del intervalo de confianza que obtuvo.

Si Carlos calculara la media de otra muestra de igual tamaño y aplicara los pasos anteriores, obtendría otro intervalo. Si Carlos repitiera el cálculo para 100 muestras, en 95 de los 100 intervalos construidos de esa manera debería esperar que sí estuviera contenido el valor de la media poblacional.
Para resolver el problema de Carlos se decidió que la confiabilidad fuera del 95%. ¿Por qué se tomó ese valor y no otro? Cada investigador determina a su arbitrio la confiabilidad, para ello debe tener en cuenta que si se aumenta la confiabilidad, aumenta el tamaño del intervalo, es decir tiene menos precisión; y, al disminuir la confiabilidad, el tamaño del intervalo también, o sea, aumenta la precisión.
Los eventos “cometer error”, cuya probabilidad es α, y “no cometer error”, cuya probabilidad es 1 – α, se complementan, por lo tanto la suma de sus probabilidades es 1.

Los pasos que dio Carlos permitieron determinar un intervalo en el cual, con una confiabilidad predeterminada, o con un riesgo predeterminado de cometer error, se puede afirmar que está el valor de la media poblacional.

La siguiente afirmación generaliza lo anterior:

Si X es una variable poblacional continua, que se distribuye normalmente con desviación estándar σ conocida, al tomar una muestra de tamaño *n* cuya media es $\bar{x}$, un intervalo de $(1-\alpha)100\%$ de confianza para μ es $(\bar{x} - z_0\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_0\frac{\sigma}{\sqrt{n}})$, donde z_0 es el valor de la distribución normal estándar que corresponde a 1 – α en área central o a α “en dos colas”.

No es necesario deducir el intervalo en cada aplicación, basta con verificar si se cumplen las condiciones del Teorema Central del Límite y, si es así, utilizar el resultado del recuadro anterior.

Para el ***ejemplo*** del gerente de ventas, Carlos Soto, se tiene que el kilometraje por litro de gasolina se distribuye normalmente, por lo que es válido aplicar el Teorema Central del Límite.
Además sabe que la desviación estándar es de 1.6, y de una muestra aleatoria de n = 54 motores obtiene un rendimiento promedio de $\bar{x}$= 14.7 Km por litro. Y para un nivel de confianza del 95% se obtiene un valor z_0 = 1.96.

Sustituyendo estos valores en el intervalo, se obtiene:

$$(\bar{x} - z_0\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_0\frac{\sigma}{\sqrt{n}})$$

$$(14.7 - 1.96\ (0.2177)\ ,\ 14.7 + 1.96\ (0.2177))$$

$$(14.7 - 0.42669\ ,\ 14.7 + 0.42669)$$

$$(14.2733\ ,\ 15.1266)$$

Y se concluye que el kilometraje medio poblacional está entre 14.2733 y 15.1266 Km/L, con 95% de confianza.

Ejercicios

1. En las condiciones de densidad del aire y presión atmosférica que privan a una altura de 7 600 m., se investiga el tiempo de resistencia de 235 candidatos a piloto. Para lo cual se toma una muestra de 50 candidatos, quienes portando mascarilla de oxígeno suben a un avión que se eleva a la altura mencionada. En ese punto se quitan las mascarillas de oxígeno y se la vuelven a poner sólo cuando la situación les resulta insoportable. El tiempo promedio que tardan en volverse a poner la mascarilla es de 3.6 minutos. Estima por intervalo, al 98%, un valor para la media poblacional si se sabe que $\sigma = 0.8$

2. En el Municipio de Tatalpa, al suroeste del Estado de Jalisco, se miden las alturas de una muestra aleatoria de 83 saúcos, *sambucus mexicana*, una de las especies de árbol que crecen en la región. De las mediciones resulta una media aritmética de 7.3 m. con desviación típica igual a 3.02 m.
Con una confiabilidad de 95%, ¿entre qué valores está la media de los saúcos de la región?

3.4 Estimación para la proporción poblacional

Cuando el parámetro que se desea estimar, de manera puntual, es la proporción poblacional (es decir la proporción de individuos en la población que poseen la característica de interés), se puede utilizar la proporción muestral $\hat{p}$, que se define como $\hat{p} = \frac{Y}{n}$, donde *n* es el tamaño de muestra y *Y* es el número de individuos en ella con la característica de interés.

Se sabe, por el Teorema Central del Límite, que los parámetros de la variable aleatoria $\hat{p}$ (proporción muestral) son $\mu_{\hat{p}} = p$ y $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, y además que $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ tiene distribución aproximadamente normal estándar.

Así, $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$ y sustituyendo Z se obtiene $P(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2}) = 1 - \alpha$

Y realizando los ajustes pertinentes: $P(-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} - p < z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}) = 1 - \alpha$

$$P(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}) = 1 - \alpha$$

Aunque los límites de intervalo contienen al parámetro desconocido *p*, es posible, sustituirlo por su estimador puntual $\hat{p}$, y se obtiene

$$P(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 1 - \alpha$$

En resumen, la proporción poblacional se encuentra entre los valores $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ y $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, con $(1 - \alpha)100\%$ de confianza.

Ejemplo

En el municipio de Altamira, Tamaulipas, se hizo una campaña de vacunación contra la hepatitis B y se desea estimar la proporción de personas mayores de 17 años que asistieron a vacunarse. Con este propósito, el consultor Fernando Juárez tomó una muestra aleatoria de 130 personas en ese rango de edad, y observó que sólo 44 se vacunaron. Con 90% de confiabilidad, ¿entre cuáles valores se encuentra la proporción de personas vacunadas en Altamira?

Solución
Fernando tiene la siguiente información: n = 130 personas mayores de 17 años elegidas al azar, Y = 44 personas vacunadas, y $(1 - \alpha)100\%$ = 90% de confianza.

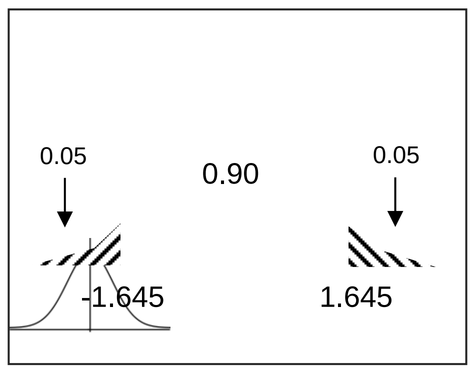
Por un lado, $\hat{p} = \frac{Y}{n} = \frac{44}{130}$ = 0.34, lo que se puede interpretar como: "el 34% de las personas mayores de 17 años acudieron a vacunarse".

Por otro lado, puesto que: $n\hat{p}$ = 130(0.34) = 44.2 > 5 y $n(1 - \hat{p})$ = 130(0.66) = 85.8 > 5, Fernando puede concluir que la muestra es suficientemente grande y es válido aplicar el Teorema Central del Límite.

Primero observa que:

$$P(-1.645 < Z < 1.645) = 0.90$$

Por lo que construye el intervalo de confianza.



$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad , \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

$$\left(0.34 - 1.645\sqrt{\frac{0.34(1-0.34)}{130}} \, , \, 0.34 + 1.645\sqrt{\frac{0.34(1-0.34)}{130}} \right)$$

$$(0.272 \quad , \quad 0.408)$$

Así que, Fernando puede afirmar, con 90% de confianza, que la proporción de personas mayores de 17 años que se vacunaron en Altamira está entre 27.2% y 40.8%.

Ejercicios

1. Un vendedor de libros se interesa en saber a qué proporción p de las personas de la ciudad donde vive les gusta leer novelas. Para ello, elige al azar a 50 personas y encuentra que a 37 de ellos les gusta leer novelas. Calcula una estimación para p, a) de manera puntual, b) con un intervalo de 96% de confianza.
2. En una escuela se planea la construcción de espacios para practicar deporte y se detecta que, de 48 alumnos elegidos al azar, 33 prefieren como deporte el básquetbol. Con 97% de confiabilidad calcula un intervalo de confianza para estimar la proporción de alumnos de toda la escuela que prefieren el básquetbol. Interpreta tu resultado.

3.5 Cálculo del tamaño de la muestra dado un error en la estimación de la media.

En las secciones anteriores se desarrolló la teoría necesaria para encontrar el intervalo de confianza para el parámetro μ, la media de una población normal. Notemos que α = probabilidad de que el intervalo no contenga la media, es generalmente un valor pequeño, que fijamos a nuestro criterio. En algunas ocasiones, es también factible que esté en nuestras manos el poder fijar el tamaño de la muestra que nos servirá para realizar la estimación. Para ello, debemos observar que el error máximo en la estimación de μ -que denotaremos por "e"- viene dado por: $e = Z_{\alpha 2}(\sigma / \sqrt{n})$

De esta igualdad, se puede ver que este error está en función:
a. el nivel de confianza deseado
b. de la desviación estándar de la población, y
c. del tamaño de la muestra

Para el tamaño muestral, observamos que a mayor valor de "n", menor error en la estimación (y viceversa). Si nos preguntamos que tan grande debe ser la muestra que tomemos con el fin de no exceder cierto error especificado, podemos despejar "n" de la ecuación anterior, y se obtiene

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

Con esta ecuación se concluye que fijando el error máximo en la estimación de μ (*e*), el nivel de confianza (1- α), y conociendo la desviación estándar de la población estudiada, se puede obtener el tamaño mínimo de la muestra que se adecua a estas condiciones.

3.6 Prueba de hipótesis

Muchos problemas en estadística se relacionan con decisiones acerca de declaraciones hechas referentes a parámetros poblacionales. A estas declaraciones se les llama hipótesis, y al procedimiento que se sigue para decidir acerca de la *verdad* o *falsedad* de una hipótesis se le conoce como prueba de hipótesis.
Las pruebas de hipótesis son una de las partes fundamentales del área de la estadística conocida como estadística inferencial o inferencia estadística.

Definición: Una hipótesis es una suposición acerca de la distribución probabilística o los parámetros de una variable aleatoria.
Ejemplos:
H: $\mu = 10$ (la media poblacional es igual a 10)
H: $\sigma^2 < 6$ (la varianza poblacional es menor a 6)
H: La población tiene distribución normal

Definición: Prueba de hipótesis es el procedimiento mediante el cual se acepta o rechaza una hipótesis, con base en la información contenida en una muestra de la población.

Los elementos de la prueba de hipótesis son:
1) Hipótesis nula (Ho): La hipótesis que se desea probar
2) Hipótesis alternativa (Ha): Una hipótesis complementaria
3) Estadístico de prueba: El que recoge la información de la muestra

4) Región de rechazo (o región critica): Forma parte de una regla de decisión

Al establecer una hipótesis y probarla, tenemos la posibilidad de cometer errores de aceptación o de rechazo. A estos errores se les conoce como Error tipo I y Error tipo II.

Definición: Error tipo I, rechazar la hipótesis nula cuando es verdadera. Con la letra griega α se designa a la probabilidad de cometer este error.

Definición: Error tipo II, aceptar la hipótesis nula cuando es falsa. Con la letra griega β se designa a la probabilidad de cometer error tipo II. Lo anterior se puede resumir en la siguiente tabla:

	La hipótesis nula es:	
DECISIÓN	Verdadera	Falsa
Aceptar	No hay error	Error tipo II
Rechazar	Error tipo I	No hay error

Definición: La probabilidad máxima con la que al probar una hipótesis se puede cometer error tipo I (α), se llama nivel de significancia de la prueba.

Dado el diseño de una prueba de hipótesis (Ho, Ha, estadístico de prueba, región de rechazo, decisión), en muchas ocasiones es posible calcular la probabilidad de cometer Error tipo I, es decir, conocer el nivel de significancia de la prueba. Sin embargo, el error tipo II generalmente no lo se puede determinar de antemano.

Por esto, se dice en estadística que rechazar una hipótesis es llegar a una conclusión fuerte (se tiene una medida de la probabilidad de estar equivocados), mientras que aceptarla es una conclusión débil (no se puede, en la mayoría de los casos, medir la probabilidad de estar equivocados).

Una forma de hacer la prueba más poderosa, es aumentar el tamaño de la muestra. Es importante para un experimentador considerar estas relaciones (α, β, n) para optimizar el diseño y los riesgos asociados al momento de tomar una decisión.

Ejemplo

Se desea contrastar con un nivel de significancia del 5% la hipótesis de que la estatura media de los hombres de 18 o más años de un país es igual a 180 cm, suponiendo que la desviación típica de las estaturas de la población es igual a 4 cm, contra la hipótesis alternativa de que es distinta.

Solución

1. Las hipótesis son:

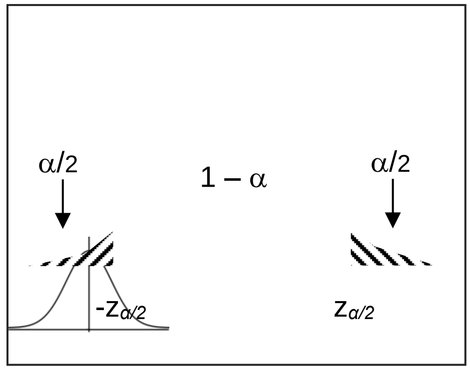
Ho: $\mu = 180$ (hipótesis nula)

Ha: $\mu \neq 180$ (hipótesis alterna)

2. Con base en el nivel de significancia, α, se obtienen los valores de $Z_{\alpha/2}$ de la distribución normal estándar.

La probabilidad α aparece “partida” en dos y sombreada en la gráfica de la derecha.

Para una significancia de α = 0.05 se tiene un valor de Z = 1.96

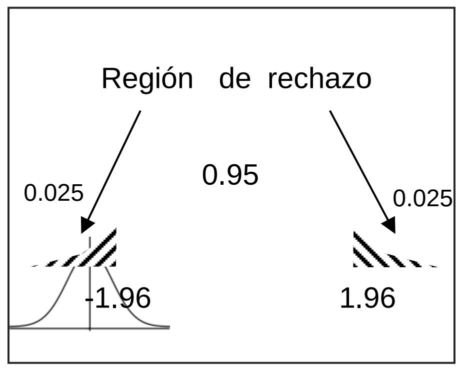


3. Se toma una muestra de n = 15 hombres seleccionados al azar, cuyas alturas son: 167, 167, 168, 168, 168, 169, 171, 172, 173, 175, 180, 180, 182, 182 y 194 cm. Con esto se calcula la media $\bar{x}$ = 174.4 y sustituyendo en la expresión del estadístico de prueba, se obtiene: $Z = \frac{(174.4-180)}{(\frac{4}{\sqrt{15}})} = -5.42$

El valor del estadístico de prueba está en la zona de rechazo.

4. Por lo que se rechaza la hipótesis nula que establece una talla igual a 180 cm

5. Conclusión
Se tiene suficiente evidencia para afirmar, con 5% de significancia, que la estatura media de los hombres de 18 años o más en ese país no es de 180 cm.



Ejercicios

En el paréntesis de la derecha escribe la letra del inciso que corresponda:

1. ¿Cuál es la hipótesis de investigación, o hipótesis alterna, de Ho: μ < 3?45, donde Ho representa la hipótesis nula y μ es la media aritmética de la población? ( )

A) μ ≥ 3.45 B) μ > 3.45 C) μ < 3.45 D) μ ≤ 3.45 E) Ninguna de las anteriores

2. El error de tipo II consiste en: ( )

A) Rechazar Ho cuando en realidad es falsa.
B) No rechazar Ho cuando en realidad es falsa.
C) No rechazar Ho cuando en realidad es verdadera.
D) No rechazar Ha cuando en realidad es falsa.
E) Rechazar Ha cuando en realidad es verdadera.

3. Un error de tipo I consiste en castigar a un acusado que es inocente, por lo tanto, el error de tipo II consiste en: ( )

A) Perdonar a un culpable
B) No castigar a un culpable
C) No castigar a un inocenteμ
D) Castigar a un inocente
E) Castigar a un culpable

4. La etiqueta de los refrescos Pascual sabor uva indica un contenido medio de 2.5 gr de azúcar por 355 ml. En un estudio con 36 refrescos se encontró una concentración promedio de azúcar de 2.6 gr por 355 ml. Con un nivel de significancia del 2% y una desviación estándar poblacional de 0.3 gr, ¿se tiene suficiente evidencia para rechazar lo que indica la etiqueta de los refrescos?

APÉNDICES

APÉNDICE I

Números aleatorios

51772	74640	42331	29044	46621
24033	23491	83587	06568	21960
45939	60173	52078	25424	11645
30586	02133	75797	45406	31041
03585	79353	81938	82322	96799
64937	03355	95863	20790	65304
15630	64759	51135	98527	62586
09448	56301	57683	30277	94623
21631	91157	77331	60710	52290
91097	17480	29414	06829	87843
62898	93582	04186	19640	87056
21387	76105	10863	97453	90581
55870	56974	37428	93507	94271
86707	12973	17169	88116	42187
85659	36081	50884	14070	74950
55189	00745	65253	11822	15804
41889	25439	88036	24034	67283
85418	68829	06652	41982	49159
16835	48653	71590	16159	14676
28195	27279	47152	35683	47280

APÉNDICE II

TABLA 5 Probabilidades binomiales

Los elementos de la tabla son la probabilidad de x éxitos en n intentos de un experimento binomial, siendo p la probabilidad de que se tenga un éxito en un intento. Por ejemplo, con 6 intentos y $p = .05$, la probabilidad de dos éxitos es .0305.

		p								
n	***x***	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
2	0	.9801	.9604	.9409	.9216	.9025	.8836	.8649	.8464	.8281
	1	.0198	.0392	.0582	.0768	.0950	.1128	.1302	.1472	.1638
	2	.0001	.0004	.0009	.0016	.0025	.0036	.0049	.0064	.0081
3	0	.9703	.9412	.9127	.8847	.8574	.8306	.8044	.7787	.7536
	1	.0294	.0576	.0847	.1106	.1354	.1590	.1816	.2031	.2236
	2	.0003	.0012	.0026	.0046	.0071	.0102	.0137	.0177	.0221
	3	.0000	.0000	.0000	.0001	.0001	.0002	.0003	.0005	.0007
4	0	.9606	.9224	.8853	.8493	.8145	.7807	.7481	.7164	.6857
	1	.0388	.0753	.1095	.1416	.1715	.1993	.2252	.2492	.2713
	2	.0006	.0023	.0051	.0088	.0135	.0191	.0254	.0325	.0402
	3	.0000	.0000	.0001	.0002	.0005	.0008	.0013	.0019	.0027
	4	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
5	0	.9510	.9039	.8587	.8154	.7738	.7339	.6957	.6591	.6240
	1	.0480	.0922	.1328	.1699	.2036	.2342	.2618	.2866	.3086
	2	.0010	.0038	.0082	.0142	.0214	.0299	.0394	.0498	.0610
	3	.0000	.0001	.0003	.0006	.0011	.0019	.0030	.0043	.0060
	4	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0002	.0003
	5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
6	0	.9415	.8858	.8330	.7828	.7351	.6899	.6470	.6064	.5679
	1	.0571	.1085	.1546	.1957	.2321	.2642	.2922	.3164	.3370
	2	.0014	.0055	.0120	.0204	.0305	.0422	.0550	.0688	.0833
	3	.0000	.0002	.0005	.0011	.0021	.0036	.0055	.0080	.0110
	4	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005	.0008
	5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
7	0	.9321	.8681	.8080	.7514	.6983	.6485	.6017	.5578	.5168
	1	.0659	.1240	.1749	.2192	.2573	.2897	.3170	.3396	.3578
	2	.0020	.0076	.0162	.0274	.0406	.0555	.0716	.0886	.1061
	3	.0000	.0003	.0008	.0019	.0036	.0059	.0090	.0128	.0175
	4	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0011	.0017
	5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001
	6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
8	0	.9227	.8508	.7837	.7214	.6634	.6096	.5596	.5132	.4703
	1	.0746	.1389	.1939	.2405	.2793	.3113	.3370	.3570	.3721
	2	.0026	.0099	.0210	.0351	.0515	.0695	.0888	.1087	.1288
	3	.0001	.0004	.0013	.0029	.0054	.0089	.0134	.0189	.0255
	4	.0000	.0000	.0001	.0002	.0004	.0007	.0013	.0021	.0031
	5	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0002
	6	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

APÉNDICE III

PROPORCIÓN ACUMULADA (COLA IZQUIERDA) P DE UNA PUNTUACIÓN Z NORMAL

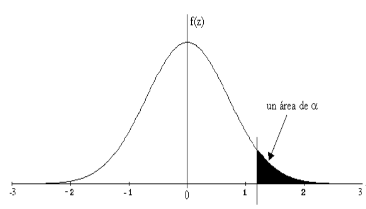
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.9999	0.9999	0.9999	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-4	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.0001	0.0001	0.0001	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.0002	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.0003	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.0004	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.0006	0.00058	0.00056	0.00054	0.00052	0.0005
-3.1	0.00097	0.00094	0.0009	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.001
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.0024	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.0028	0.00272	0.00264
-2.6	0.00466	0.00453	0.0044	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.0057	0.00554	0.00539	0.00523	0.00508	0.00494	0.0048
-2.4	0.0082	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.0099	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.0139	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.0116	0.0113	0.01101
-2.1	0.01786	0.01743	0.017	0.01659	0.01618	0.01578	0.01539	0.015	0.01463	0.01426
-2	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.0197	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.0268	0.02619	0.02559	0.025	0.02442	0.02385	0.0233
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.0392	0.03836	0.03754	0.03673
-1.6	0.0548	0.0537	0.05262	0.05155	0.0505	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.0778	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.3	0.0968	0.0951	0.09342	0.09176	0.09012	0.08851	0.08692	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
-1.1	0.13567	0.1335	0.13136	0.12924	0.12714	0.12507	0.12302	0.121	0.119	0.11702
-1	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.2327	0.22965	0.22663	0.22363	0.22065	0.2177	0.21476
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.2451
-0.5	0.30854	0.30503	0.30153	0.29806	0.2946	0.29116	0.28774	0.28434	0.28096	0.2776
-0.4	0.34458	0.3409	0.33724	0.3336	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
-0.3	0.38209	0.37828	0.37448	0.3707	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
-0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
-0.1	0.46017	0.4562	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0	0.5	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.4721	0.46812	0.46414

APÉNDICE IV

Distribución normal estándar
Valores de z para algunas probabilidades

Área central para $(-z, z)$ $1-\alpha$ en área central	.10	.30	.50	.60	.70	.80	.90	.95	.98	.99	.995	.998	.999
Área de dos colas $(-\infty, -z] \cup [z, \infty)$	.90	.70	.50	.40	.30	.20	.10	.05	.02	.01	.005	.002	.001
Área de la cola derecha $[z, \infty)$	.45	.35	.25	.20	.15	.10	.05	.025	.01	.005	.0025	.001	.0005
Valores de z > 0	**.126**	**.385**	**.674**	**.842**	**1.036**	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**	**2.807**	**3.090**	**3.291**



RESPUESTAS DE LOS EJERCICIOS

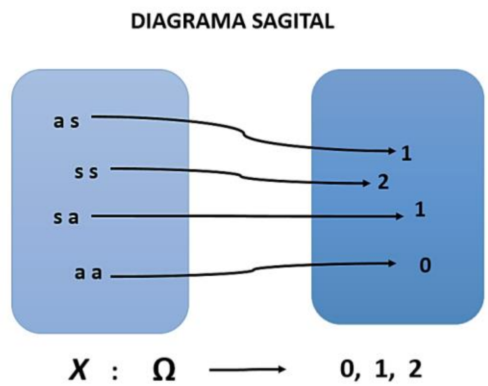
Ejercicio. Pagina 6

a) Espacio muestral

$$\Omega = \{as, ss, sa, aa\}$$

b) La variable aleatoria X se define como el número de soles que aparecen al lanzar una moneda, dos veces consecutivas.

c) Diagrama sagital para la variable aleatoria X



d) Recorrido: $R_X = 0, 1, 2$

e) El tipo de variable aleatoria es discreta, ya que el resultado se obtuvo a través de un proceso de conteo.

Ejercicio. Página 7

a) Espacio muestral:

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	(1 1)	(1 2)	(1 3)	(1 4)	(1 5)	(1 6)
⚁	(2 1)	(2 2)	(2 3)	(2 4)	(2 5)	(2 6)
⚂	(3 1)	(3 2)	(3 3)	(3 4)	(3 5)	(3 6)
⚃	(4 1)	(4 2)	(4 3)	(4 4)	(4 5)	(4 6)
⚄	(5 1)	(5 2)	(5 3)	(5 4)	(5 5)	(5 6)
⚅	(6 1)	(6 2)	(6 3)	(6 4)	(6 5)	(6 6)

Ω = {(1 1), (1 2), (1 3), (1 4), (1 5), (1 6), (2 1), (2 2), (2 3), (2 4), (2 5), (2 6),

(3 1), (3 2), (3 3), (3 4), (3 5), (3 6),

(4 1), (4 2), (4 3), (4 4), (4 5), (4 6),

(5 1), (5 2), (5 3), (5 4), (5 5), (5 6),

(6 1), (6 2), (6 3), (6 4), (6 5), (6 6) }

b) Resultado de la suma de los puntos, al lanzar un dado dos veces.

X es una función que asigna un valor numérico a cada uno de los eventos aleatorios simples del espacio muestral.

$$\begin{array}{ccc} X : \Omega & \longrightarrow & \mathfrak{R} \\ (1\ 1) & \longrightarrow & 2 \\ (1\ 2) & \longrightarrow & 3 \\ \cdots & & \cdots \\ (6\ 6) & \longrightarrow & 12 \end{array}$$

c) Recorrido $R_X = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$

Ejercicio. Página 9

a) La variable aleatoria X se define como el número de soles obtenidos al lanzar una moneda corriente tres veces.

b) Cada uno de los eventos aleatorios simples del espacio muestral corresponden a un valor de la variable aleatoria X

La variable aleatoria X definida como el número de soles obtenidos al lanzar una moneda común tres veces consecutivas

Espacio muestral	Número de soles
aaa	0
aas	1
asa	1
saa	1
ssa	2
sas	2
ass	2
sss	3

c) Cada uno de los valores de la variable aleatoria corresponden a una probabilidad de X

Probabilidades para la variable aleatoria X definida como el número de soles obtenidos al lanzar una moneda común tres veces consecutivas

Número de soles X	Probabilidad P(X)
0	1/8
1	1/8
1	1/8
1	1/8
2	1/8
2	1/8
2	1/8
3	1/8
	1

La distribución de probabilidad de la variable aleatoria X definida como el número de soles al lanzar una moneda común tres veces.

Tabla 3. Distribución de probabilidad de la variable aleatoria X definida como el número de soles obtenidos al lanzar una moneda común tres veces

Número de soles X	Probabilidad P(X)
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
	1

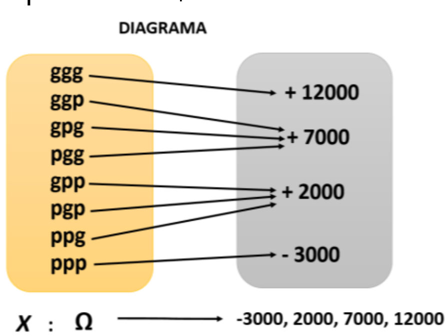
Ejercicio. Página 10

a) La variable aleatoria X se define como la ganancia, o pérdida, de dinero (en pesos), al participar Toño Lee en el torneo de tae kwon do.

b) De los tres encuentros, si gana lo denotamos por **g**, si pierde lo denotamos por **p**.

$$\Omega = \{ggg, ggp, gpg, pgg, gpp, pgp, ppg, ppp\}$$

c) Por cada encuentro ganado recibe \$4000, por cada encuentro perdido le provoca una pérdida de \$1000.



En el diagrama se observan las ganancias y pérdidas, en pesos Existe una relación entre el conjunto Ω y el conjunto de números reales.
Entre dos valores de sus imágenes no es posible otro valor se considera variable aleatoria discreta.

d) El recorrido de la variable X

$$R_X = \mathbf{-3000, 2000, 7000, 12000}$$

e) Según el enunciado *“Si el torneo está tan parejo que su probabilidad de ganar es igual a la de perder,”*

la probabilidad de que se presente cualquier resultado del espacio muestral es de $\frac{1}{8}$

La distribución de probabilidad de la variable aleatoria X definida como la ganancia, o pérdida, de dinero (en pesos), al participar Toño Lee en el torneo de tae kwon do.

Distribución de probabilidad de la variable aleatoria X definida como la ganancia, o pérdida, de dinero (en pesos), al participar Toño Lee en el torneo de tae kwon do.

Ganancia o pérdida, en pesos, X	Probabilidad P(x)	Interpretación de P(x)
-3000	$\frac{1}{8}$	Es la probabilidad de no ganar un solo encuentro
2000	$\frac{3}{8}$	Es la probabilidad de ganar un encuentro y perder dos
7000	$\frac{3}{8}$	Es la probabilidad de ganar dos encuentros y perder uno.
12000	$\frac{1}{8}$	Es la probabilidad de ganar los tres encuentros.
	1	

Además, cumple con las propiedades de distribución de probabilidad de la variable X.

f) ¿Debe Toño ***esperar*** ganancia de dinero después de sus tres encuentros?

Para obtener la respuesta se debe calcular el promedio de sus posibles ganancias a través del valor esperado de X.

El **valor esperado** o promedio μ se calcula *ponderando* cada ganancia o pérdida con su probabilidad de ocurrencia, es decir, dándoles peso con su respectiva probabilidad:

Valores P_i

$$\mu = 12\,000\frac{1}{8} + 7\,000\frac{3}{8} + 2\,000\frac{3}{8} - 3\,000\frac{1}{8}$$

Valores x_i

$$\mu = \frac{12+21+6-3}{8}(1000) = \frac{36}{8}(1000) = 4\,500$$

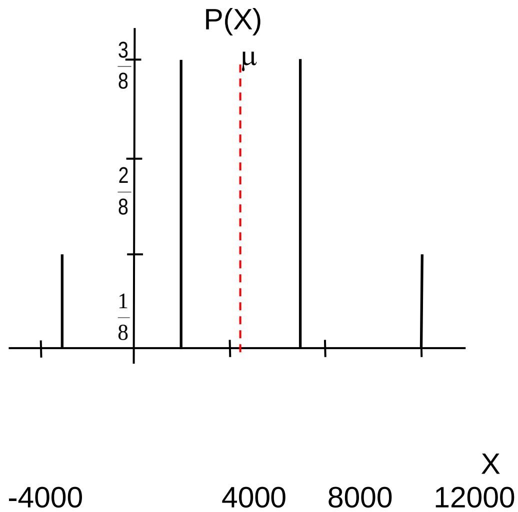
Otra forma es construir la tabla de distribución de probabilidad, para calcular el valor esperado, μ.

Distribución de probabilidad de la variable aleatoria X definida como la ganancia, o pérdida, de dinero (en pesos), al participar Toño Lee en el torneo de tae kwon do.

Ganancia, o pérdida, en pesos. X	Probabilidad P(X)	***X*P(X)***
-3000	$\frac{1}{8}$	$-3000 * \frac{1}{8} = \frac{-3000}{8}$
2000	$\frac{3}{8}$	$2000 * \frac{3}{8} = \frac{6000}{8}$
7000	$\frac{3}{8}$	$7000 * \frac{3}{8} = \frac{21000}{8}$
12000	$\frac{1}{8}$	$12000 * \frac{1}{8} = \frac{12000}{8}$
	1	$\frac{36000}{8} = 4500$

Por lo que tenemos que $\mu = \$4,500.00$, es decir, Toño Lee espera, en promedio, ganar $4500.00 a largo plazo, como ganancia al participar en sus tres encuentros por cada partido.

Lo anterior quiere decir que Toño debiera esperar una ganancia de \$4 500.00, lo cual no significa que suceda con seguridad, puede tener un mal día y resultar una pérdida final de \$3 000.00, pero será más probable una ganancia cercana a \$4500.00



Gráfica I.1

En forma abreviada:

$$\mu = E(X) = \sum_{i=i}^{i=n} \mathbf{X}_i \; P(X = X_i) \; . \; . \; . \; . \; . \; . \; . \; \text{I.1}$$

La gráfica I.1 representa la distribución de probabilidad para X. En ella podemos apreciar simetría con respecto a su promedio μ, lo cual sucede porque para Toño la probabilidad de ganar un encuentro es la misma que de perderlo.

g) Para medir la dispersión de la variable aleatoria, recuerda que se debe calcular la desviación estándar de la variable X, con la fórmula adecuada

La **variabilidad** o **dispersión** de los valores de la variable aleatoria**,** en el problema de Toño Lee, resulta:

$$\sigma^2 = (12000 - 4500)^2 \frac{1}{8} + (7\,000 - 4500)^2 \frac{3}{8} + (2\,000 - 4500)^2 \frac{3}{8} + (-3\,000 - 4500)^2 \frac{1}{8}$$

$$= (7500)^2 \frac{1}{8} + (2500)^2 \frac{3}{8} + (-2500)^2 \frac{3}{8} + (-7500)^2 \frac{1}{8}$$

$$= 7\,031\,250 + 2\,343\,750 + 2\,343\,750 + 7\,031\,250$$

$$= 18\,750\,000 \text{ (pesos)}^2$$

Por consiguiente: σ = \$4 330.127

es lo que se dispersa, (se aleja), del promedio μ.

Sobre el eje X de la gráfica I.1, al hacer una marca σ unidades a la derecha de μ (o sea, a 4 330.127 a la derecha) y también a la izquierda de μ.

Así se formó un intervalo de 2σ unidades de ancho (8 660.254) sobre el eje X; una franja del mismo ancho en el plano, centrada en μ.

Ejercicio. Página 17

Debe cumplir con las características de un experimento binomial.

Características de un proceso binomial	
✓ El experimento aleatorio se puede repetir n veces. ✓ Cada prueba que se realiza solo son posibles dos resultados: Éxito y Fracaso. ✓ El resultado de cada prueba es independiente entre sí. ✓ La probabilidad de éxito es la misma en cada repetición. Se denota por *p*. ✓ La probabilidad de fracaso es la misma en cada repetición. Se denota por *q*; Se cumple *p+q=1* ✓ La variable aleatoria binomial se define como el número *X* de éxitos obtenidos en *n* pruebas realizadas ✓ Los valores de la variable aleatoria binomial son: X = 0, 1, 2, 3, 4, ..., n	-n = 4 clientes -*Éxito*: el cliente seleccione un auto rojo. -*Fracaso*: el cliente no seleccione un auto rojo. -La decisión de un cliente es independiente de la decisión de otro cliente. -La probabilidad de que el cliente seleccione un auto rojo: $p = \frac{5}{8}$ -La probabilidad de que el cliente no seleccione un auto rojo: $q = \frac{3}{8}$ Se debe cumplir $p + q = \frac{5}{8} + \frac{3}{8} = 1$ La variable aleatoria binomial se define como el número de clientes que seleccionan un auto rojo. -Su recorrido X = 0, 1, 2, 3, 4 POR LO TANTO, SE CONSIDERA QUE ES UN EXPERIMENTO DE TIPO BINOMIAL

b) Para calcular la probabilidad de que los clientes seleccionen menos de 2 autos rojos significa que sucede X=0 ó x=1, por lo tanto se suman las probabilidades:

$$P(x<2) = P(x=0) + P(x=1) = C_0^4 \times \left(\frac{5}{8}\right)^0 \left(\frac{3}{8}\right)^4 + C_1^4 \times \left(\frac{5}{8}\right)^1 \left(\frac{3}{8}\right)^3 = \frac{81}{4096} + \frac{540}{4096} = 0.1516$$

Por lo que se puede afirmar que la probabilidad de que los clientes seleccionen menos de dos autos rojos de un total de cuatro es de aproximadamente 0.1516

Calcula la media aritmética (o esperanza matemática) de la distribución

$$\mu = np; \quad \mu = 4 \times \frac{5}{8} = \frac{20}{8} = 2.5 \text{ clientes eligen acuto rojo}$$

En promedio, se espera que 2.5 clientes elijan un auto rojo. (Bajo el contexto, se espera que elijan un auto rojo entre 2 y 3 clientes.)

c) Su dispersión se calcula a partir de la desviación estándar:

$$\sigma = \sqrt{npq}; \quad \sigma = \sqrt{4 \times \frac{5}{8} \times \frac{3}{8}} = 0.9682 \text{ clientes}$$

En promedio se desvía un cliente alrededor de la media.

Ejercicio. Página 17

Debe cumplir con las características de un experimento binomial.

Características de un proceso binomial	
✓ El experimento aleatorio se puede repetir n veces. ✓ Cada prueba que se realiza solo son posibles dos resultados: Éxito y Fracaso. ✓ El resultado de cada prueba es independiente entre sí. ✓ La probabilidad de éxito es la misma en cada repetición. Se denota por *p*. ✓ La probabilidad de fracaso es la misma en cada repetición. Se denota por *q*; Se cumple *p+q=1* ✓ La variable aleatoria binomial se define como el número *X* de éxitos obtenidos en *n* pruebas realizadas ✓ Los valores de la variable aleatoria binomial son: X = 0, 1, 2, 3, 4, ..., n	-n = 7 relojes de pulsera. -*Éxito*: el reloj de pulsera esté defectuoso. -*Fracaso*: el reloj de pulsera no esté defectuoso. -La decisión de un cliente es independiente de la decisión de otro cliente. -La probabilidad de que el reloj de pulsera esté defectuoso: $p = \frac{10}{100}$ -La probabilidad de que el reloj de pulsera no esté defectuoso: $q = \frac{90}{100}$ Se debe cumplir $p + q = \frac{10}{100} + \frac{90}{100} = 1$ La variable aleatoria binomial se define como el número de relojes de pulsera estén defectuosos. -Su recorrido X = 0, 1, 2, 3, 4, 5, 6, 7 POR LO TANTO, SE CONSIDERA QUE ES UN EXPERIMENTO DE TIPO BINOMIAL

a) Para calcular la probabilidad de que escojan más de un reloj defectuoso, significa que sucede X = 2, 3, 4, 5, 6, 7.

   Se puede calcular X = 0, 1 que es el complemento, por lo que quedaría:

$$P\left(n = 7, p = \frac{10}{100};\ x = 2{,}3{,}4{,}5{,}6{,}7\right) = 1 - [C_0^7(\frac{10}{100})^0(\frac{90}{100})^{(7-0)} + C_1^7(\frac{10}{100})^1(\frac{90}{100})^{(7-1)}] =$$

$$= 1 - [0.4782969 + 0.3720087] = 0.1496944$$

Por lo tanto, la probabilidad de que haya más de un reloj defectuoso de un total de 7 es del 0.1496 aproximadamente.

La media o valor esperado $\mu = np; \quad \mu = 7 \times 0.1 = 0.7$

Se espera, en promedio entre 0 y 1 relojes defectuosos de un total de 7.

Su dispersión, se calcula mediante la desviación estándar:

$\sigma = \sqrt{npq}; \quad \sigma = \sqrt{7 \times 0.1 \times 0.9} = 0.7937$ relojes

Ejercicios adicionales. Página 18

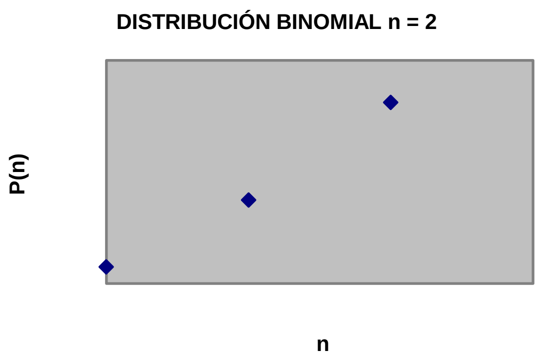
1. Sí, con reserva, porque sólo tiene dos resultados posibles: éxito y fracaso, se pudo repetir n veces, pero las repeticiones no necesariamente son independientes entre sí.

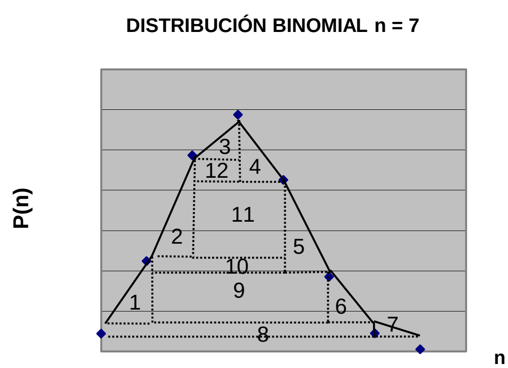
2. a) Porque sólo tiene dos resultados posibles: éxito y fracaso (es católico practicante o no lo es); se puede repetir n veces (la selección) y las repeticiones son independientes entre sí.

b) P(0) = 0.02207 P(1) = 0.11192 P(2) = 0.24314
P(3) = 0.29344 P(4) = 0.21249 P(5) = 0.09232
P(6) = 0.02228 P(7) = 0.002305

c) i. P(n = 7) = 0.002305 ii. P(n = 0) = 0.02207 iii. P (al menos 5 sean católicos practicantes) = 0.116915 iv. 0.59825

d)





Área (n = 2) = $\frac{1}{2}$ 2(0.24314 - 0.02207) + 2(0.02207) = 0.81932 + 0.04414 = 0.86346
Área (n = 7) = 0.999965
E(X) = *np*=2.94

3. Como la elección se hace con reemplazo, se cumplen los requisitos de un experimento binomial. X es el número de canicas blancas, n = 4 y p = 5/8 = 0.625. La probabilidad que necesitamos es P(X < 2) y ésta es igual a:

$$P(X < 2) = P(X = 0) + P(X = 1) = 0.375^4 + (4)(0.625)(0.375)^3 = 0.1516$$

4. n = 15, p = 0.10, q = 1 - p = 0.90

a) $$P(X \le 4) = \sum_{x=0}^{4} f(x) = P(0) + P(1) + P(2) + P(3) + P(4)$$

$$= 0.2059 + 0.3431 + 0.2669 + 0.1285 + 0.0429 = 0.9873$$

b) $P(X \geq 5) = 1 - \sum_{x=0}^{4} f(x) = 1 - 0.9873 = 0.0127$

c) $P(3 \leq X \leq 6) = \sum_{x=3}^{6} f(x) = P(3) + P(4) + P(5) + P(6)$

$= 0.1285 + 0.0429 + 0.0105 + 0.0019 = 0.1838$

5. n = 12, p = 0.95, q = 1 - p = 0.05; x = 9

a) $P(X = 9) = {}_{12}C_9 (0.95)^9 (0.05)^3 = 0.0173$

6. n=4, p=0.80, q=0.20

a) $P(X = 2) = {}_4C_2 (0.80)^2 (0.20)^2 = 0.1536$

b) $P(X \geq 1) = \sum_{x=1}^{4} f(x) = 1 - P(0) = 1 - 0.0016 = 0.9984$

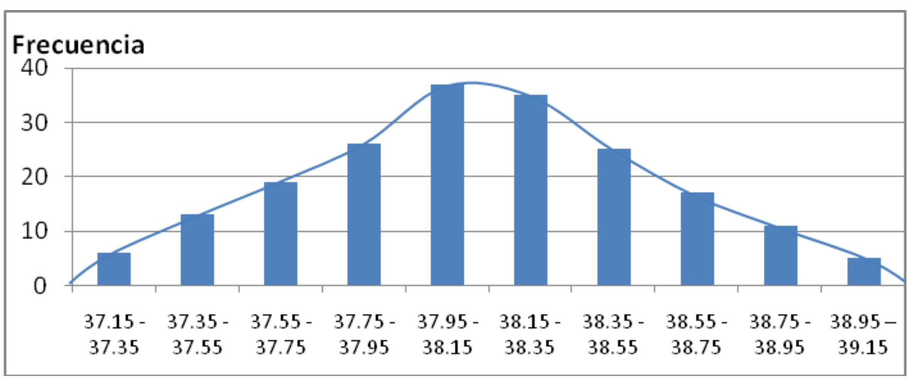
7. E(X) = np = 4(0.50) = 2, $\sigma_x^2 = npq = 4(0.50(0.50) = 1$, $\sigma_x = \sqrt{npq} = 1$

8. n = 30; p = 0.70; q = 1- p = 1 – 0.70 = 0.30; X = 20

$P(X = 20) = {}_{30}C_{20} (0.70)^{20} (0.30)^{10} = 0.14156$

Ejercicios. Página 28

SOLUCIONES



1.

La distribución es aproximadamente normal o gaussiana, las temperaturas se agrupan alrededor de 38^0.

2. P(X > 30) = P($\frac{x-\mu}{\sigma} > \frac{30-25.5}{4.5}$) = P(Z > 1) = 0.1587

El 15.87% de los automóviles compactos consigue 30 mpg o más.

3. a) P(170 < X < 230) = P(x < 230) – P(X < 170) = P(Z < 1.5) – P(X < –1.5) = 0.9332 – 0.0668 = 0.8664.

La probabilidad de que un individuo seleccionado al azar en esa comunidad tenga un nivel de colesterol entre 180 y 220 unidades es del 86.64%.
b) P(X > 155) = P(Z > – 2.25) = 0.9878. La probabilidad de que un individuo seleccionado al azar en esa comunidad tenga un nivel de colesterol mayor de 155 unidades es del 98.78%.

4. a) P(X > 1) = P(Z > 0) = 0.5000. El 50% de los paquetes de carne pesará más de 1 kg.
b) P(950 < X < 1030) = P(X < 1030) – P(X < 950) = P(Z < 2) – P(Z < – 3.33) = 0.9772 – 0.005 = 0.9767. La probabilidad de que un paquete elegido al azar pese entre 0.950 y 1.030 kg es del 97.67%.

5. a) P(X > 6) = P(Z > 0.28) = 0.3897. La probabilidad de que un trabajador elegido al azar en esa población tenga, al año, incapacidad de más de 6 días es de 38.97%
b) a) P(X < 5) = P(Z < –0.14) = 0.4443. La probabilidad de que un trabajador elegido al azar en esa población tenga, al año, incapacidad de más de 6 días es de 44.43%.

Ejercicios. Página 35

SOLUCIÓN

1.a) $P(\bar{x} \geq 100) = P\left(z \geq \frac{100-100}{\frac{20}{\sqrt{40}}}\right) = P(z \geq 0) = 0.5000$

b) $P(96 \leq \bar{x} \leq 105) = P\left(\frac{96-100}{\frac{20}{\sqrt{40}}} \leq z \leq \frac{105-100}{\frac{20}{\sqrt{40}}}\right) = P(-1.26 \leq z \leq 2.53) = 0.8905.$

c) $P(\bar{x} \geq 110) = P\left(z \geq \frac{110-100}{\frac{20}{\sqrt{40}}}\right) = P(z \geq 3.16) = 0.0008$

2.a) $P(46 \leq \bar{x} \leq 55) = P\left(\frac{46-50}{\frac{16}{\sqrt{64}}} \leq z \leq \frac{55-50}{\frac{16}{\sqrt{64}}}\right) = P(-2 \leq z \leq 2.5) = 0.9710$

b) $P(\bar{x} > 53) = P\left(z > \frac{53-50}{\frac{16}{\sqrt{64}}}\right) = P(z > 1-5) = 0.4332$

c) $P(\bar{x} < 48) = P\left(z < \frac{48-50}{\frac{16}{\sqrt{64}}}\right) = P(z < -1) = 0.1587$

d) $P(45 \leq \bar{x} \leq 49) = P\left(\frac{45-50}{\frac{16}{\sqrt{64}}} \leq z \leq \frac{49-50}{\frac{16}{\sqrt{64}}}\right) = P(-2.5 \leq z \leq -0.50) = 0.3023$

3. a) $P(\bar{x} > 190) = P\left(z > \frac{190-185.6}{\frac{12.7}{\sqrt{10}}}\right) = P(z > 1.1) = 0.1357$

b) $P(\bar{x} \geq 180) = P\left(z \geq \frac{180-185.6}{\frac{12.7}{\sqrt{10}}}\right) = P(z \geq -1.39) = 0.9177$

c) $P(182 < \bar{x} < 189) = P\left(\frac{182-185.6}{\frac{12.7}{\sqrt{10}}} < z < \frac{189-185.6}{\frac{12.7}{\sqrt{10}}}\right) = P(-1.79 < z < 0.85) = 0.7656$

4. a) $P(\bar{x} > 42.5) = P\left(z > \frac{42.5-45}{\frac{5}{\sqrt{16}}}\right) = P(z > -2) = 0.9772$

b) $P(43 \leq \bar{x} \leq 47.5) = P\left(\frac{43-45}{\frac{5}{\sqrt{16}}} \leq z \leq \frac{47.5-45}{\frac{5}{\sqrt{16}}}\right) = P(-1.60 \leq z \leq 2) = 0.9224$

c) $P(\bar{x} \geq 48) = P\left(z \geq \frac{48-45}{\frac{5}{\sqrt{16}}}\right) = P(z \geq 2.4) = 0.0082$

d) $P(\bar{x} \leq 42) = P\left(z \leq \frac{42-45}{\frac{5}{\sqrt{16}}}) = P(z \leq -2.4\right) = 0.0082$

5. a) $P(\bar{x} \geq 26) = P\left(z \geq \frac{26-25}{\frac{5}{\sqrt{32}}}\right) = P(z \geq 1.13) = 0.1292$

b) $P(24 \leq \bar{x} \leq 27) = P\left(\frac{24-25}{\frac{5}{\sqrt{32}}} \leq z \leq \frac{27-25}{\frac{5}{\sqrt{32}}}\right) = P(-1.13 \leq z \leq 1.44) = 0.7953$

c) $P(\bar{x} \leq 23) = P\left(z \leq \frac{23-25}{\frac{5}{\sqrt{32}}}\right) = P(z \leq -1.44) = 0.0749$

d) $P(\bar{x} > 22) = P\left(z > \frac{22-25}{\frac{5}{\sqrt{32}}}\right) = P(z > -3.39) = 0.9997$

6. a) $P(\bar{x} > 6) = P\left(z > \frac{6-5.7}{\frac{1}{\sqrt{9}}}\right) = P(z > 0.9) = 0.1841$

b) $P(5 \leq \bar{x} \leq 6) = P\left(\frac{5-5.7}{\frac{1}{\sqrt{9}}} \leq z \leq \frac{6-5.7}{\frac{1}{\sqrt{9}}}\right) = P(-2.1 \leq z \leq 0.9) = 0.7980$

c) $P(\bar{x} < 5.2) = P\left(z < \frac{5.2-5.7}{\frac{1}{\sqrt{9}}}\right) = P(z < -1.5) = 0.0668$

d) $P(5.8 \leq \bar{x} \leq 6.1) = P\left(\frac{5.8-5.7}{\frac{1}{\sqrt{9}}} \leq z \leq \frac{6.1-5.7}{\frac{1}{\sqrt{9}}}\right) = P(0.3 \leq z \leq 1.2) = 0.2670$

Ejercicios. Página 38

1. a) P($\hat{p}$ ≥ 0.65) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{0.65-0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}}$) = P(Z ≥ 1.02) = 0.1539

b) P($\hat{p}$ ≤ 0.58) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.58-0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}}$) = P(Z ≤ -0.4) = 0.3446

c) P(0.56 < $\hat{p}$ < 0.66) = P($\frac{0.56-0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.66-0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}}$) = P(-0.81< Z < 1.22) = 0.6798

2. a) P($\hat{p}$ ≥ 0.2) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{0.2-0.15}{\sqrt{\frac{(0.15)(0.85)}{100}}}$) = P(Z ≥ 1.4) = 0.0808

La probabilidad de que la muestra dé una proporción de mujeres sometidas a dieta mayor o igual que 0.2 es del 0.0808

b) P(0.1 < $\hat{p}$ < 0.22) = P($\frac{0.1-0.15}{\sqrt{\frac{(0.15)(0.85)}{100}}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.22-0.15}{\sqrt{\frac{(0.15)(0.85)}{100}}}$) = P(-1.4 < Z < 1.96) = 0.8942

La probabilidad de que la muestra arroje una proporción de mujeres sometidas a dieta entre 0.1 y 0.22 es de 0.8942

c) P($\hat{p}$ ≤ 0.12) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.12-0.15}{\sqrt{\frac{(0.15)(0.85)}{100}}}$) = P(Z ≤ - 0.84) = 0.2911

La probabilidad de que la proporción de mujeres en la muestra sometidas a dieta sea no mayor que 0.12 es de 0.2911

3. P($\hat{p}$ ≥ 0.27) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{0.27-0.2}{\sqrt{\frac{(0.2)(0.8)}{150}}}$) = P(Z ≥ 2.14) = 0.0162.

La probabilidad de que la proporción de familias en la muestra, con al menos un miembro que sufre algún malestar debido a la contaminación del aire, sea 0.27 o mayor es de 0.0162

4. P($\hat{p}$ ≥ 0.15) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{0.15-0.08}{\sqrt{\frac{(0.08)(0.92)}{150}}}$) = P(Z ≥ 3.16) = 0.0008.

La probabilidad de que la proporción de individuos daltónicos en la muestra tan grande como 0.15 es apenas de 0.0008

5. a) P($\hat{p}$ < 0.85) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.85-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$) = P(Z < 2.35) = 0.9906.

La probabilidad de que la proporción de clientes en la muestra con cuentas en otro banco sea menor a 0.85 es de 0.9906

b) P(0.87 < $\hat{p}$ < 0.89) = P($\frac{0.85-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.89-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$) = P(-1.41 < Z < -0.47) = 0.2399

La probabilidad de que la proporción de clientes en la muestra con cuentas en otro banco este entre 0.87 y 0.89 es de 0.2399

c) P($\hat{p}$ > 0.92) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.92-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$) = P(Z > 0.94) = 0.1736.

La probabilidad de que la proporción de clientes en la muestra con cuentas en otro banco sea al menos de 0.92 es de 0.1736

d) P(0.86 < $\hat{p}$ < 0.93) = P($\frac{0.86-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.93-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$) = P(-1.88 < Z < 1.41) = 0.8903

La probabilidad de que la proporción de clientes en la muestra con cuentas en otro banco sea mayor de 0.86 pero menor de 0.93 es de 0.8903

e) P($\hat{p}$ < 0.84) + P($\hat{p}$ > 0.88) = P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.84-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$) + P($\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.88-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}}$)

= P(Z < -2.82) + P(Z > -0.94) = 0.0024 + 0.8264 = 0.8288

La probabilidad de que la proporción de clientes en la muestra con cuentas en otro banco sea menor a 0.84 o mayor a 0.88 es de 0.8288

Ejercicios. Página 44

1. El tiempo medio que tardan los candidatos a piloto en volverse a poner la mascarilla está entre 3.33 y 3.86 minutos, con 98% de confianza.

2. La media de los saúcos de esa región está entre 6.65 y 7.94 m, con una confiabilidad del 65%.

Ejercicios. Página 45

1. a) La proporción de personas en la ciudad que les gusta leer novelas es de aprox. 74%
b) La proporción de personas en la ciudad que les gusta leer novelas se encuentra entre 61.22 y 86.77%, con una confianza del 98%.

2. La proporción de alumnos de toda la escuela que prefieren el básquetbol está entre 54,23 y 83.26% con una confianza del 97%.

Ejercicio. Página 48

1. A; 2. D; 3. B; 4. El valor del estadístico de prueba está en la zona de rechazo, por lo tanto se rechaza la hipótesis nula Ho: $\mu = 2.5$ y se concluye que existe suficiente evidencia para rechazar lo que indica la etiqueta.

BIBLIOGRAFÍA

Bibliografía recomendada

→ *Castillo, J*. Estadística Inferencial Básica. México. GEI. 2000

→ *Chao, L*. Introducción a la Estadística. México. CECSA.1989

→ Christensen, H. Estadística paso a paso. Trillas, 1990 (reimp. 2006)

→ *Daniel, W*. Estadística aplicada a las Ciencias Sociales y a la Educación

→ Johnson, R. y Kuby, P. Estadística Elemental. 11a. Edición. Iberoamérica, 1990

→ Lind, D. Marchal, W. y Mason, R. Estadística para Administración y Economía. Alfaomega, 2004

→ *Triola, M*. Estadística. 11a Ed. México. Pearson Eduacación. 2013

→ Willowghby, S. Probabilidad y Estadística. PCSA, 1993

→ Wonnacott, T. Fundamentos de Estadística para Administración y Economía. Limusa, 1989

→ Spiegel, M. Probabilidad y Estadística. Mc Graw Hill, 1975