EXPLICACIÓN SOBRE EL MANEJO DE ESTA GUÍA

Esta guía comprende todos los temas del programa, de ahí que su estudio deba de ser de la primera a la última página, si se desea un máximo de posibilidades de éxito en el examen extraordinario. Para resolverla no basta con leerla como se hace con textos de otro tipo, es necesario participar resolviendo **todos** los ejercicios y problemas.

SUGERENCIAS DE ACTIVIDADES DE APRENDIZAJE TEÓRICO PRÁCTICAS

¿Cómo estudiar esta guía?

- Hazlo **relajado** siempre, no bajo presión o urgencia, ni con distractores.
- Mantente en **alerta** para seguir al pie de la letra las instrucciones, comprender los conceptos y no equivocarte al contestar.
- Cuando inicies el estudio de un tema, **conclúyelo**.
- Al estudiar aprovecha **al máximo** el tiempo, con cero interrupciones.
- Usa lápiz, sacapuntas, goma y calculadora. Ten disponibles escuadras, transportador, compás y una computadora con impresora.

El **repaso** es un hábito clave para el aprendizaje y esta guía no es la excepción. Al realizar un estudio sostenido, con repasos **diarios**, estarás en el camino del aprendizaje perdurable y significativo.

¿Cómo repasar?

1º Comienza con una lectura de todo lo visto antes.

2º Cuando aparezcan problemas y ejercicios ya estudiados, en un cuaderno aparte resuélvelos nuevamente y sólo al terminar consulta la solución.

De hacer lo anterior, al final adquirirás las competencias matemáticas previstas, obtener buena calificación y estar satisfecho del aprendizaje logrado.

En ocasiones se sugieren actividades de campo: por ejemplo investigar en qué consiste el Teorema de Chebishev (*Distribución normal, ejercicios*), obtener una muestra aleatoria (*El teorema central del límite*, *tabla V.3*), etc. Esas actividades propician el reforzamiento del aprendizaje, por lo cual es necesario realizarlas.

FORMAS DE VERIFICAR EL APRENDIZAJE LOGRADO

En cada tema, los ejercicios (algunos desarrollados y con sus respectivas respuestas) tienen el propósito de reforzar tu aprendizaje. Su resolución es **indispensable** para comprobar tu nivel de comprensión. No dejes de resolverlos y presenta esta Guía resuelta cuando presentes el examen.

BIBLIOGRAFÍA

- Daniel W. W. Bioestadística. Ed. Limusa Wiley. México 2007.
- Mendenhall W., Beaver R. J. y Beaver B. M. Introducción a la probabilidad y Estadística. Ed. Thomson. México 2002.
- Pagano R. R. Estadística para las ciencias del comportamiento. Ed. Thomson. México 2006.
- Weimer R. C. Estadística. Ed. CECSA. México 1999.

ÍNDICE

UNIDAD I
VARIABLE ALEATORIA Y FUNCIÓN DE PROBABILIDAD

COMPETENCIAS POR ADQUIRIR

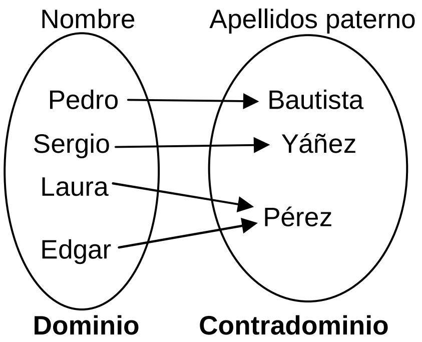
En un problema dado:

- **Identificar la variable aleatoria y calcular sus valores**
- **Calcular la varianza y la desviación estándar de la variable aleatoria**
- **Calcular los valores solicitados de la función de probabilidad**
- **Calcular el valor esperado de la variable y su esperanza matemática**

Al estudiar un fenómeno probabilístico se pueden asociar los eventos del espacio muestral con un tipo especial de función, conocida como ***variable aleatoria***, la cual puede ser discreta o continua. En términos de esa variable queda definida otra función muy importante: la ***función de probabilidad*** o ***función de densidad***.

Ambos conceptos, ***variable aleatoria*** y ***función de probabilidad***, además de la **esperanza matemática** y la **desviación estándar**, son el objeto de esta Unidad.

¿TE ACUERDAS DEL CONCEPTO DE FUNCIÓN? . . .



Para los científicos una función es una **relación** que se da entre los elementos de un conjunto (llamado **Dominio**) y los de otro (llamado **Contradominio**), con la condición: de que a un elemento determinado del Dominio le corresponda **sólo uno** del Contradominio (al cual se le llama imagen).

La vida cotidiana está plagada de funciones. Por ejemplo, a cada trabajador en una fábrica le corresponde sólo un salario; legalmente a cada nombre de persona le corresponde sólo un apellido paterno; al contar los árboles de un huerto, a cada árbol le hacemos corresponder sólo un número natural. . . y así puedes hallar muchos casos más.

En el diagrama, ¿cuál es la imagen de Sergio? _______.

APRÉNDETE ESTAS DOS FORMULAS

Media poblacional o esperanza matemática: $\mu = E(x) = \sum_{i=i}^{i=n} X_i \; P(X = X_i)$ (I.1)

Varianza poblacional: $\sigma^2 = \sum_{i=1}^{i=n} (x_i - \mu)^2 P(x_i)$ (I.2)

Problema. 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.

¿Cuántos ocupantes en promedio iban en los automóviles? ¿Cuánta dispersión tuvo la variable aleatoria X: ocupantes en un automóvil?

Ocupantes en un automóvil (X)	f_i	$x_i f_i$	$\frac{f_i}{n}$
1	208	208	0.208
2	318	636	0.318
3	169	507	0.169
4	188	752	0.188
5	92	460	0.092
6	25	150	0.025
Total	1000		1.00

Tabla I.1

Utilizando la fórmula I.1 se tiene que:

$$\mu = E(X) = \sum_{i=i}^{i=n} X_i \, P(X = X_i)$$

$$\mu = 1(0.208) + 2(0.318) + 3(0.169) + 4(0.188) + 5(0.092) + 6(0.025)$$
$$= 2.713 \text{ ocupantes/automóvil}$$

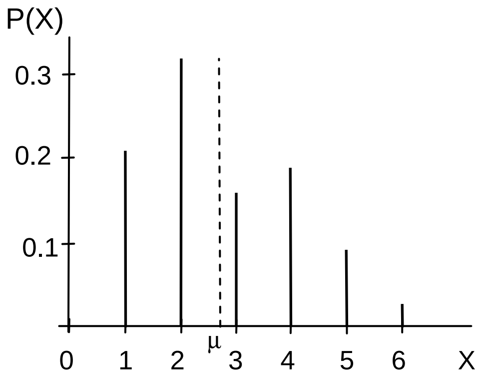
La medida cuantitativa de la dispersión puede ser la desviación estándar o la varianza.

Empleando la fórmula I.2:

$$\sigma^2 = \sum_{i=1}^{i=n} (x_i - \mu)^2 P(x_i)$$

$$= (-1.713)^2(0.208) + (-0.713)^2(0.318) + (0.287)^2(0.169) +$$
$$+ (1.287)^2(0.188) + (2.287)^2(0.092) + (3.287)^2(0.025)$$
$$= 1.848 \text{ (ocupantes/automóvil)}^2$$

Por lo tanto: $\sigma = = 1.359$ ocupantes/automóvil.



Al elaborar la gráfica de la distribución de probabilidad de la variable aleatoria X, se puede apreciar la ubicación del valor esperado E(X) o media μ (izquierda).

¿Cuántos valores de X quedan comprendidos en el intervalo que va de $\mu + \sigma$ a $\mu - \sigma$? _____. ¿Qué porcentaje son del total? _____%.

Problema. Calcula el valor esperado del juego. Un jugador lanza un dado legal (que no está cargado). Si sale un número primo, gana el mismo número de pesos pero si no sale un número primo, entonces pierde ese número de pesos. Los resultados posibles x_i del juego con sus respectivas probabilidades $P(x_i)$ son (completa la tabla con las fracciones que faltan, los números negativos ocurren cuando no sale un número primo):

x_i	2	3	5	-1	-4	-6
$P(x_i)$	$\frac{1}{6}$	___	___	___	___	___

Problema. Rodolfo Moreno es veterinario y tiene cuatro gatos birmanos (b_1, b_2, b_3, b_4), dos siameses (s_1, s_2) y dos de angora (a_1, a_2). Para llevar a cabo un experimento debe seleccionar dos gatos al azar, ¿cómo puede determinar la variable aleatoria X: *número de gatos siameses*? **(Nota: *determinar la variable aleatoria* significa definir su dominio y su contradominio).**

Antes que nada, debe calcular cuántos eventos elementales forman el dominio Ω de la variable aleatoria X. Puesto que de **ocho** gatos serán seleccionados **dos** (completa):

$$C_2^8 = \frac{\quad}{\quad}$$

$$=$$

Los valores posibles de X son 0, 1 y 2. Esto es así porque de los dos gatos seleccionados pudiera ser que ninguno fuera siamés, o sólo uno, o ambos.

X = 0 cuando: los 2 gatos son birmanos (de 4): $C_2^4 = 6$

ó 1 gato es birmano (de 4) y 1 de angora (de 2): $C_1^4\, C_1^2$ = 8

ó dos gatos son de angora (de 2): $C_2^2 = 1$

Al sumar, resulta que 15 de los resultados posibles no incluyen un gato siamés.

X = 1, o un gato es siamés, cuando (escribe lo que falta):

Uno de los gatos es birmano (de 4) y otro es siamés (de 2):

$$C\ \ C\ =$$

ó uno de los gatos es de angora (de 2) y otro es siamés (de 2):

$$C\ \ C\ =$$

Entonces, de los resultados posibles, ¿cuántos incluyen un gato siamés? ___.

Finalmente, **X = 2**, o dos gatos son siameses, cuando:

Los 2 gatos son siameses (de 2): $C_2^2 =$ __

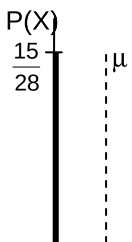
Los resultados anteriores se organizaron en esta tabla:

EVENTOS ELEMENTALES DE Ω	VALOR DE X (NÚM. DE GATOS SIAMESES)	MANERA DE CALCULARLOS
b_1b_2, b_1b_3, b_1b_4, b_2b_3, b_2b_4, b_3b_4 b_1a_1, b_1a_2, b_2a_1, b_2a_2, b_3a_1, b_3a_2, b_4a_1, b_4a_2 a_1a_2	0	$C_2^4 + C_1^4\, C_1^2 + C_2^2$
b_1s_1, b_1s_2, b_2s_1, b_2s_2 b_3s_1, b_3s_2, b_4s_1, b_4s_2 a_1s_1, a_1s_2, a_2s_1, a_2s_2	1	$C_1^4\, C_1^2 + C_1^2\, C_1^2$
s_1s_2	2	C_2^2

Si tomas como dominio a los eventos elementales de Ω (primera columna de la tabla) y como imágenes a los valores de X (segunda columna de la tabla), la relación entre ambos es la función: **variable aleatoria X**.

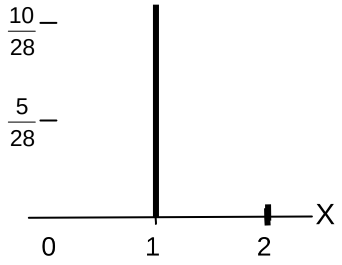
¿Cómo es X, discreta o continua? ___________.

A su vez, si las imágenes de X, que son 0, 1, y 2, se toman como dominio de la **función de densidad** P, ésta se define así:



$$P: \{0, 1, 2\} \xrightarrow{P(X)} [0,1]$$

Donde (escribe lo que falta):



Gráfica I.2

$$P(0) = \frac{15}{28},\ P(1) = \frac{\quad}{\quad},\ P(2) = \frac{\quad}{\quad}$$

Las funciones X y P están relacionadas de esta manera:

$$\Omega \xrightarrow{X(x)} \{0, 1, 2\} \xrightarrow{P(X)} [0,1]$$

La esperanza matemática del número de gatos siameses se calcula empleando la igualdad I.1.

$$E(X) = 0\left(\frac{15}{28}\right) + 1\left(\frac{\quad}{\quad}\right) + 2\left(\frac{\quad}{\quad}\right)$$

$$= \frac{\quad}{\quad}$$

$$= 0.5$$

En la gráfica I.2, ¿por qué el valor de E(X) no se encuentra en medio de la distribución, donde X = 1? ____.

O sea, Rodolfo debe esperar que cuando seleccione al azar dos gatos, uno o ninguno sea siamés (puesto que el resultado no puede ser 0.5).

Imagina: se selecciona una pareja de gatos un millón de veces (por supuesto reemplazando los gatos cada vez), ¿cuántos gatos siameses tendría la mayoría de las parejas seleccionadas? (marca con ✓):

Dos ( ) Uno ( ) Ninguno ( ) Uno o ninguno ( )

Calcula el valor de la varianza y de la desviación estándar de X:

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

$$=$$

$$=$$

$$=$$

$$=$$

$$\sigma =$$

Sobre el eje X de la gráfica I.2 toma como punto de referencia a μ y marca σ unidades tanto a la derecha como a la izquierda.

Así se formó un intervalo (o una franja) de 2σ unidades de ancho, centrada en μ.

¿Cuántos valores de X están en el intervalo? ____ . Es decir, el ____%.

EJERCICIOS COMPLEMENTARIOS I

1. Completa la tabla:

Fenómeno o Experimento	Variable Aleatoria X	Tipo de Variable Aleatoria	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 cms.
Encuestar a 100 personas sobre su preferencia por un candidato a la presidencia.	Proporción de la preferencia por un candidato.	Continua	
Deportación a ilegales mexicanos 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.		

(1) Los puntos suspensivos indican que la numeración continúa hasta un número indeterminedo.

2. Un consorcio habitacional decide emplear a tres de los seis arquitectos (A_1, A_2, A_3, A_4, A_5, A_6) que tiene contratados. Cada uno tiene diseñadas la cantidad de casas modelo que se indican en la siguiente tabla:

ARQUITECTOS	CASAS MODELO DISEÑADAS
A_2, A_3, A_5	2
A_1, A_6	3
A_4	4

Se seleccionan los tres arquitectos al azar y se define la variable aleatoria X: *suma de casas modelo diseñadas* por los tres arquitectos.

a. Elabora una tabla para indicar los valores posibles de X (1ª columna) y los respectivos valores elementales del espacio muestral (2ª columna).

b. Determina el valor de la probabilidad para cada valor de X.

c. Calcula el valor de la esperanza de X.

Cantidad de enfermos de hepatitis C	Probabilidad
0	0.10
1	0.35
2	0.17
3	0.30
4	0.08

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 acudiera una cantidad de enfermos determinada al mes (ver la tabla).
a. ¿Cuál es la cantidad esperada de enfermos de hepatitis C?
b. ¿Cuál es la desviación estándar?

4. 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?

5. Sea X: *número de “águilas” obtenidas al tirar tres monedas ideales*. Obtén el valor esperado de X.

6. Una tienda de artículos electrónicos vende cierto modelo de computadora portátil, del cual se tienen cuatro en existencia. El gerente se pregunta cuál será la demanda hoy para dicho modelo. El Departamento de Ventas le informa que la distribución de probabilidad para X: *demanda diaria para la computadora portátil*, es la siguiente:

Determina la media, la varianza y la desviación estándar de X.

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

¿Cuál es la respuesta a la pregunta que se hace el gerente? ____________________

7. Para un sorteo en beneficio del cuerpo de bomberos se venderán 8 000 boletos, a $50.00 cada uno. Si el premio es un automóvil de $200 000.00 y una persona compra dos boletos,

a. ¿cuál es su ganancia esperada?

b. ¿cuál es su ganancia esperada si compra tres boletos?

8. En la ciudad de Los Angeles, en Estados Unidos de Norteamérica, la probabilidad de que una casa de cierto tipo quede destruida por un incendio en un año es 0.005. Una compañía de seguros le ofrece al propietario una póliza de seguro contra incendio por 20 000 USD y a un año por una prima de 150 000 USD.

a. ¿Qué es una póliza de seguro? ____

____

b. ¿Qué es una prima de seguro? ______________________________________

Calcula la ganancia esperada de la compañía

9. Una variable aleatoria X puede tomar cinco valores: 0, 1, 2, 3, 4. Enseguida se muestra una parte de la distribución de probabilidad, a) completa la tabla.

X	0	1	2	3	4
P(X)	0.10	0.30	0.30		0.10

Calcula:

b) La media, varianza y desviación estándar.

c) La probabilidad de que X sea mayor que 2.

d) La probabilidad de que X sea 3 o menor

UNIDAD II
DISTRIBUCIÓN BINOMIAL DE PROBABILIDAD

COMPETENCIAS POR ADQUIRIR
- Identificar un experimento de Bernoulli.
- Para un problema dado,
  a. Construir la distribución de probabilidad de la variable aleatoria discreta involucrada.
  b. Calcular la probabilidad de una variable aleatoria.
  c. Calcular la esperanza matemática de una variable aleatoria binomial.

Existe una diversidad de distribuciones de *variable discreta*, entre las más importantes está la distribución binomial, que es una distribución de probabilidad del número de éxitos en una secuencia de n experimentos independientes.

La función de probabilidad P de una **variable aleatoria binomial** es tal que:

$$\sum_{i=1}^{i=n} P(X = X_i) = 1$$

$$\text{y} \quad P(x) = C_x^n\, p^x q^{n-x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ 3.1$$

Recuerda siempre esta fórmula, con ella se calcula la probabilidad de obtener x éxitos en un número n, fijo, de intentos, si se conoce la probabilidad de un éxito aislado.

Lo anterior se cumple siempre y cuando se trate de un **experimento de Bernoulli**, como en el caso estudiado por Martín López.

Son cuatro las características de un experimento de Bernoulli:
- **Sólo tiene dos resultados posibles: éxito y fracaso.**
- **Se puede repetir n veces.**
- **Las repeticiones son independientes entre sí.**
- **La probabilidad de éxito es la misma en cada repetición.**

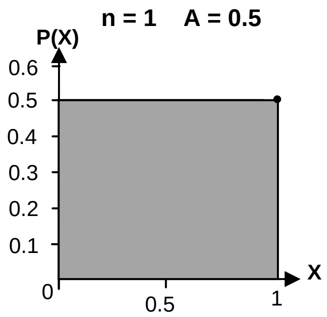
La ventaja de la fórmula 3.1 es que permite calcular las demás probabilidades (para n = 4, n = 5, etc.), sin necesidad de obtener los valores de Ω. Completa la tabla 3.6 empleando dicha fórmula.

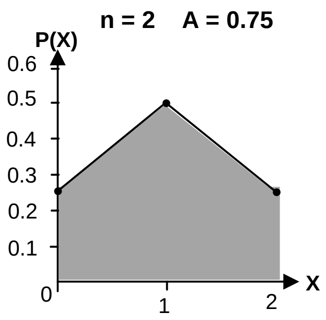
n \ X	0	1	2	3	4	5	6
1	0.5	0.5					
2	0.25	0.5	0.25				
3	0.125	0.375	0.375	0.125			
4							
5							
6							

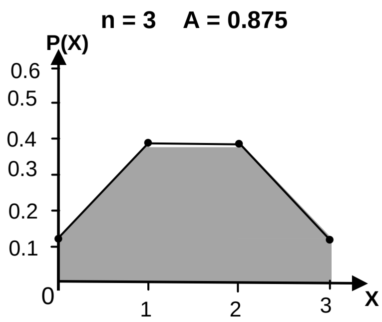
Tabla 3.6

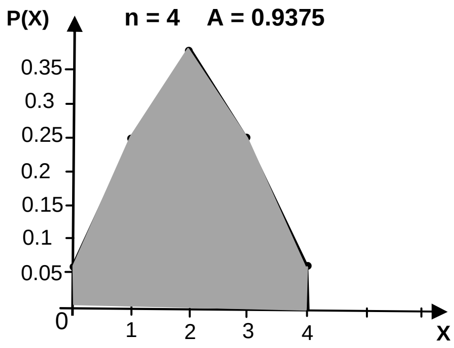
¿Hay simetría en cada renglón de la tabla 3.6? ___ .
Una pareja de recién casados planea tener 4 hijos (n = 4). ¿Qué probabilidad tiene de que 3 sean varones (X = 3)?_______.
Si planea tener 5 hijos, ¿cuál es la probabilidad de que ninguno sea varón?_______. ¿Cuál es la probabilidad de que 2 o 3 sean varones?_______.

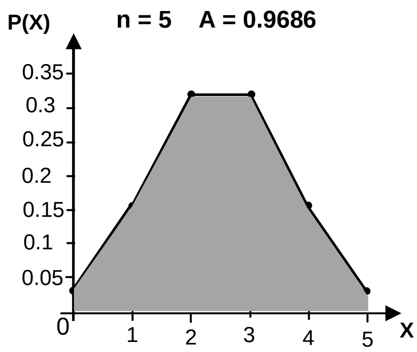
La representación gráfica de P(X) para cada valor de *n* queda así (en cada una se indica el área bajo la curva:











La escala en el eje P de las probabilidades se formó con números decimales porque son valores que no tienen más cifras en su parte decimal. Cuando el número de cifras en la parte decimal es infinito, o es elevado, conviene trabajar con fracciones para no perder exactitud.

A partir de las gráficas anteriores, si aumenta *n*, entonces el área bajo la "curva", ¿aumenta o disminuye?__________. Si *n* se hace cada vez más grande, ¿a qué valor tiende el área bajo la curva? ___ .

Si se traza una línea vertical por el centro de cada una de las gráficas, se hace más evidente su propiedad de ___________; lo cual sucede porque $p = q$.
Cuando $p \neq q$ en el experimento de Bernoulli, se puede construir la distribución de probabilidad como lo hizo Martín López. Por ejemplo, supón que $q = 0.3$ y $p = 0.7$ son las probabilidades de fracaso y éxito respectivamente. Escribe los valores de la probabilidad en la tabla que sigue, para calcularlos aplica la fórmula 3.1 de la **distribución binomial**, $P(x) = C^{n}_{x}\, p^{x}q^{n-x}$, dándole un valor a *n* y a sus respectivos valores de *x*.

X / ***n***	**0**	**1**	**2**	**3**	**4**	**5**
1	0.3	0.7				
2	0.09	0.42	0.49			

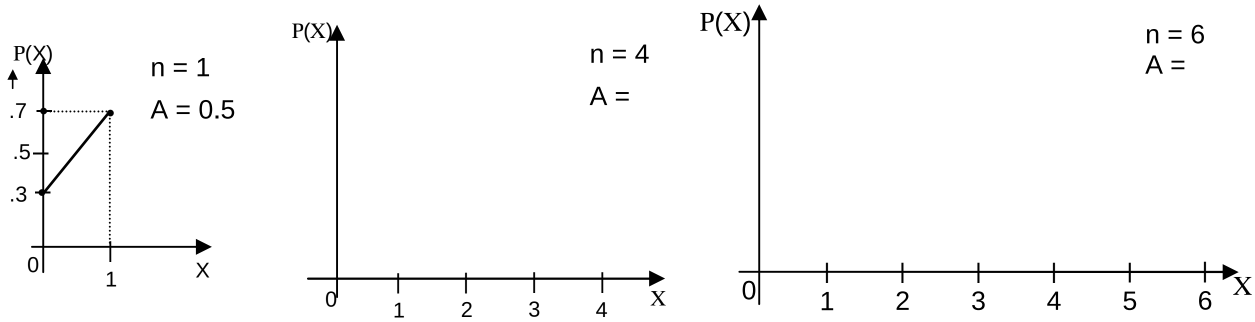
3	0.027	0.189	0.441	0.343		
4						
5						
6						

Tabla 3.7

¿Hay simetría en cada renglón de la tabla 3.7? ___ .Cuando *p* y *q* son **iguales**, sí hay simetría, ahora en cada valor de *n* (renglones en la tabla) se advierte que las **probabilidades** de la variable aleatoria **no forman un arreglo simétrico**.

A cada valor de *n* le corresponde una distribución de probabilidad y sus gráficas nos permiten visualizar de una manera rápida el comportamiento de la función de probabilidad.

Termina las gráficas de abajo y calcula el área bajo la curva; para hacerlo es recomendable hacer divisiones en forma de triángulos y rectángulos.



¿A qué valor tiende el área A cuando *n* se hace más grande? ___.

¿Son simétricas las gráficas? ___. Esto sucede porque $p \neq q$.

TABLAS BINOMIALES DE PROBABILIDAD

Observa la tabla 3.6.

Esa tabla se construyó para n = 0, 1, 2, 3, 4, 5, 6 suponiendo p = q. En caso de que $p \neq q$, dependiendo de los valores de de p y de q, se pueden hacer los cálculos y construir la tabla correspondiente (es el caso de la tabla 3.7).

Existen tablas más amplias que contienen las probabilidades binomiales para diferentes valores de "n" y de "x" (ver Apéndice II-A y II-B), las cuales ahorran algunos cálculos.

Las tablas binomiales de probabilidad se presentan en diferentes formas.

Por ejemplo, localiza la tabla del Apéndice II-A. Ésta indica que si se realizan 7 intentos, cada uno con una probabilidad de éxito de 0.5, entonces la probabilidad de obtener 3 éxitos es P(x=3) = 0.2734.

¿Cuál es la probabilidad de obtener 2 éxitos en 10 intentos si la probabilidad "p" de éxito en cada intento es de 0.7? (utiliza el último renglón de la tabla) _______ .

Para calcular, por ejemplo, la probabilidad de obtener 0, 1 ó 2 éxitos en 10 intentos, con probabilidad de éxito 0.5, normalmente se suman las probabilidades respectivas (completa consultando el Apéndice II-A)

P(x =0, 1 ó 2) = P(x = 0) + P(x = 1) + P(x = 2)

= _______ + _______ + _______

= _______

Sin embargo, existen tablas, como la del Apéndice II-B, que de manera directa proporcionan el resultado acumulado de P(x = 0) + P(x = 1) + P(x = 2). Localiza el valor de la probabilidad en dicha tabla, tomando n = 10, p = 0.5, k = 2; ¿cuánto es? _______. ¿Coincide este valor con el resultado que obtuviste antes? ___.

Una probabilidad binomial se puede calcular por tres métodos:

b) Por medio de la fórmula $P(x) = C_x^n p^x q^{n-x}$.

c) Utilizando las tablas binomiales.

a) Empleando un programa de cómputo, por ejemplo, Minitab.

Problema. Un medicamento causa efectos secundarios en cinco de cada cien pacientes. Si se eligen al azar ocho pacientes que tomaron el medicamento, encuentra:

a. La probabilidad de que ningún paciente muestre efectos secundarios (encuentra la respuesta utilizando la tabla del Apéndice II-A o la del Apéndice II-B).

P(x=0) = _______

b. La probabilidad de que muestren efectos secundarios a lo más tres pacientes (encuentra la respuesta de manera directa utilizando la tabla del Apéndice II-B).

P(x = 0, 1, 2 ó 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) = _________.

Problema. El médico Juan Manuel Lozano es acupunturista y ha descubierto que un tratamiento para curar la obesidad ha dado resultado favorable en dos de cada tres de sus pacientes, quienes disminuyeron 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 quiere saber el número más probable de pacientes que se curará.

El Dr. Lozano advierte lo siguiente:

n = 4 (*se someten al tratamiento 4 pacientes).*

X = 0, 1, 2, 3, 4 (*número de pacientes que se curan).*

$p = \frac{2}{3}$, $q = \frac{1}{3}$

Se trata de un experimento de Bernoulli.

Por lo tanto, decide aplicar la fórmula de la **distribución binomial**: $P(x) = C_x^n\, p^{x} q^{n-x}$

a. 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:

$$P(3) + P(4) = C_3^4 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^1 + C_4^4 \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^0$$

$$= \frac{32}{81} + \frac{16}{81}$$

$$= \frac{48}{81}$$

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

$$P(2) = C_2^4 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2$$

$$= \frac{24}{81}$$

c. La probabilidad de que se cure un paciente significa que sucede x = ___:

P(1) = _________

= ——

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

Recordando que la suma de probabilidades de los eventos elementales es ___:

P(0) + P(1) + P(2) + P(3) + P(4) = ___

Al despejar:

P(0) = ___ - [P(1) + P(2) + P(3) + P(4)]

Los valores de las probabilidades que están dentro de los corchetes ya se calcularon antes, por lo tanto:

P(0) = ___ - [____ + ____ + ____ + ____]

P(0) = __ - ____

= ____

e. Cuando la variable aleatoria tiene una distribución binomial de probabilidad, las ecuaciones generales para calcular el valor esperado y la varianza se pueden simplificar de tal manera que:

E(x) = μ = np **(donde μ es la media aritmética de la distribución binomial)** 2.2

Var (x) = σ^2 = npq 2.3

En el problema del Dr. Lozano, n = 4, $p = \frac{2}{3}$, $q = \frac{1}{3}$, en consecuencia:

$$\mu = \frac{8}{3} = 2.66$$

Con la fórmula general, ¿cuál es el resultado?

$$\sum_{j=0}^{j=4} X_i P(X_i) = \frac{8}{81} + 2\frac{24}{81} + 3\frac{32}{81} + 4\frac{16}{81}$$

$$= \frac{216}{81}$$

= ______

$$\sigma^2 = \frac{8}{9} = 0.88$$

$$\sigma = 0.938$$

Esto significa que, en promedio, el Dr. Lozano debiera esperar que con mayor probabilidad se curen ___ pacientes.

Problema. 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. ¿Cuál es la probabilidad de que los clientes seleccionen menos de 2 autos rojos?

Se trata de un experimento aleatorio **con reemplazo**, puesto que todos y cada uno de los clientes tendrán a su disposición ocho autos para seleccionar uno. Siendo así, tiene las características de un experimento de Bernoulli, o sea:

1. Hay una situación idéntica de una acción a otra.
2. En cada experimento son posibles dos resultados: el cliente selecciona un auto rojo (éxito) y el cliente no selecciona un auto rojo (__________).
3. La probabilidad de éxito y de fracaso no varían de un cliente a otro.
4. La decisión de elegir de un cliente es independiente de la decisión del anterior.

El éxito en este caso se daría cuando se seleccionara uno de los 5 autos rojos. Si X es el número de autos rojos que pueden seleccionar los 4 clientes:

$$X = 0, 1, 2, 3, \text{ y } 4.$$

Además, $n = 4$, $p = \frac{5}{8}$

La probabilidad que buscamos es $P(X < 2)$:

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) \\ &= C_0^4 \left(\frac{5}{8}\right)^0 \left(\frac{3}{8}\right)^4 + C_1^4 \left(\frac{5}{8}\right)^1 \left(\frac{3}{8}\right)^3 \\ &= \frac{81}{4096} + \frac{540}{4096} \\ &= 0.1516 \end{aligned}$$

Si los cuatro clientes llegaran al mismo tiempo, el primero podría escoger el auto de entre 8 disponibles, el segundo lo escogería de entre 7, el tercero de entre 6 y el cuarto de entre 5. En ese caso, ¿dejarían de cumplirse las características de un experimento de Bernoulli? ___.

¿Cuáles? ____

____

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

$$E(X) =$$

$$=$$

Calcula la desviación estándar.

$$\sigma =$$

$$=$$

$$=$$

Problema. 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?

El número de experimentos es $n = 7$ y la probabilidad de éxito es la probabilidad de que al tomar un reloj éste salga defectuoso: $p = 0.1$

Aunque el muestreo se hace **sin reemplazo**, puesto que la cantidad de artículos que hay en el lote es mucho mayor comparada con los que se escogen, se puede suponer, sin perder mucha exactitud, que son válidos los supuestos del experimento de Bernoulli.

Si X es el *número de artículos defectuosos*, entonces: X = ________________________

La probabilidad solicitada es:

$P(X > 1)$ = _______ + _______ + _______ + _______ + _______ + $P(X = 7)$

El cálculo de las 6 probabilidades de la igualdad resulta laborioso, pero notemos que:

$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)+ P(X = 6)+ P(X = 7)$ = ___

Lo anterior es lo mismo que: $P(X \leq 1) + P(X > 1)$ = ___

Al despejar: $P(X > 1)$ = _________

Al consultar el Apéndice II-B para n = 7 y p = 0.10, se obtiene:

$P(X \leq 1) = P(X = 0) + P(X = 1)$

= _________

La media o valor esperado μ y la varianza serían: $\mu = np$ = ______

= _____

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

= __________

EJERCICIOS COMPLEMENTARIOS II

1. ¿La erupción del volcán de Colima es un fenómeno equivalente a un experimento de Bernoulli? ¿Por qué?

____

2. De acuerdo a una encuesta del Instituto Mexicano de la Juventud, alrededor del 40% 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. ¿Por qué 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) y construye su gráfica.

c. Calcula la probabilidad de que sean católicos practicantes:

a. Todos __________________
b. Ninguno __________________
c. Al menos 5 __________________
d. Más de 2 pero menos de 6 __________________

d. Construye las gráficas y calcula las áreas bajo las curvas que corresponden a $n = 2$ y $n = 7$. ¿Cuál de las áreas es mayor?

e. ¿Cuál es el valor de la esperanza matemática?

3. Si en una población de 100 000 habitantes, 55 de cada 100 personas son mujeres y 45 son varones. En una familia de tres hijos:

a. ¿Cuál es la probabilidad de que a lo más 2 sean mujeres?

b. ¿Cuántas mujeres debiera esperar un matrimonio que planea tener tres hijos?

4. 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?

5. El 10% de cierta población es daltónica. Si al azar se selecciona una muestra de 15 personas, calcula la probabilidad de que sean daltónicos:

a. Cuatro o menos.

b. Cinco o más.

c. Entre tres y seis, inclusive.

6. De los botones producidos por una máquina, 95% no tiene defectos. Al tomar una muestra al azar de 12 piezas, ¿cuál es la probabilidad de que nueve no tengan defectos?

7. Durante un largo tiempo se ha observado que un soldado da en el blanco, con un solo disparo, con probabilidad igual a 0.80. Supón que dispara cuatro veces al blanco, calcula la probabilidad de que dé en el blanco:

a. Dos veces

b. Al menos una vez

8. Se sabe que siete de cada 10 pacientes que toman cierta medicina se curan. De 30 que han tomado la medicina, ¿cuál es la probabilidad de que se curen 20?

9. En cierta ciudad la necesidad de dinero para comprar alimentos se establece como el motivo del 75% de los robos. Encuentra la probabilidad de que entre los siguientes cinco casos de robo:

a) Dos resulten de la necesidad de dinero para comprar alimentos.

b) Al menos tres resulten de la necesidad de dinero para comprar alimentos.

c) Representa esta distribución binomial por medio de una gráfica.

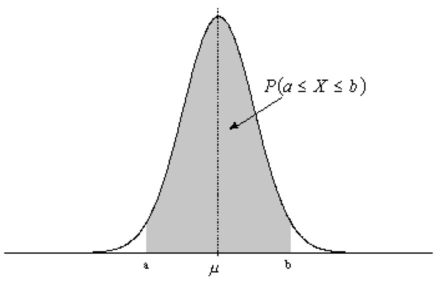
d) Calcula la media y la desviación estándar de esta distribución binomial.

UNIDAD III
FUNCIÓN DE DISTRIBUCIÓN NORMAL

COMPETENCIAS POR ADQUIRIR

- **Expresar las propiedades de la curva normal.**
- **Resolver problemas aplicando la teoría de la distribución normal de probabilidad.**

Una variable numérica puede ser discreta o **continua**. Cuando es **continua**, el modelo probabilístico más utilizado es la función de distribución **normal**, la cual provee una descripción adecuada para la distribución de una gama muy amplia de ese tipo de variables.



Campana de Gauss

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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in R$$

Esto significa que, con valores dados de μ y σ, al darle valores a *x*, se obtiene que *f(x)* es una campana de Gauss **centrada en μ**.

Algunas propiedades de la distribución normal

- Es **simétrica** con respecto a un eje que pasa por μ, su media aritmética.
- Se extiende desde –∞ hasta +∞.
- Es **asintótica** al eje *x*: nunca lo toca por mucho que se extienda tanto a la derecha como a la izquierda.
- La esperanza matemática de la variable aleatoria X coincide con el valor del parámetro μ, la media aritmética poblacional, y la varianza de X es σ^2:

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

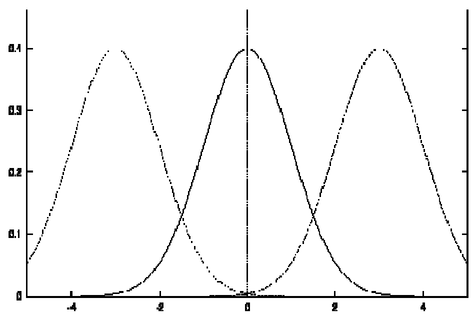


Figura 3.3 Distribuciones gaussianas con distintas medias pero igual dispersión.

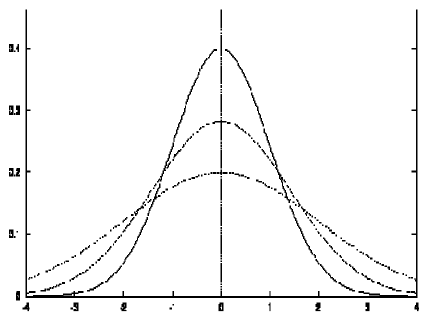


Figura 3.4 Distribuciones gaussianas con distintas dispersiones pero igual media.

- La **posición** de la curva depende del valor de la media aritmética μ. Si se modifica μ, se desplaza pero mantiene su forma (figura 3.3).
- Su **forma** depende de la desviación σ (parámetro de dispersión). Cuanto menor sea σ, mayor probabilidad se concentrará alrededor de la media (gráfica de *f* muy "apuntada" cerca de μ), y cuanto mayor sea σ más "aplastada" estará (figura 3.4).
- El valor de la media μ coincide con los valores de la mediana y la moda.
- En un intervalo (*a, b*) sobre el eje *x*, **el área bajo la curva representa un valor de la probabilidad** P(*a* < X < *b*). El área total equivale a una probabilidad de 1.

En una distribución normal, el **valor *z*** se define así*:*

$$z = \frac{x-\mu}{\sigma}$$ **(Recuerda siempre esta definición)**

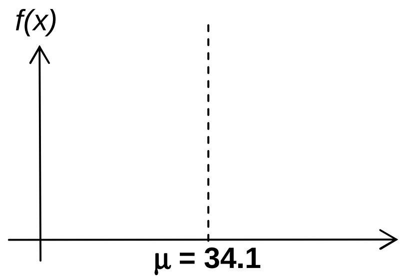
o bien, $z\sigma = x - \mu$ **(*z* veces σ es igual a la separación entre "*x*" y μ, lo cual quiere decir que *z* es el número de desviaciones estándar que hay de diferencia entre "*x*" y μ).**

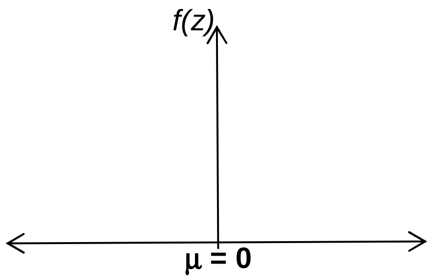
Si en *f*(*x*) se substituye *z*, se produce una transformación. Observa:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \longrightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{donde } \mu = \mathbf{0},\ \sigma = \mathbf{1}$$

Al hacer el cambio de variable (en vez de *x* utilizar *z*) se dice que se **estandariza** la variable aleatoria X. La función *f(z)* se conoce como **función de densidad estandarizada** y su gráfica es la **curva normal estandarizada**, para la cual la media aritmética es 0 y la desviación estándar es uno.

Abajo, dibuja la curva normal para *f(x)* y la normal estandarizada para *f(z)*, ambas tienen forma de campana y están centradas alrededor de la media.





Volvamos al asunto de los lobos, Juan se plantea la pregunta

> **¿Cuál es la probabilidad de que al seleccionar un lobo al azar, de entre los 90, pese menos de 35 kilos?**

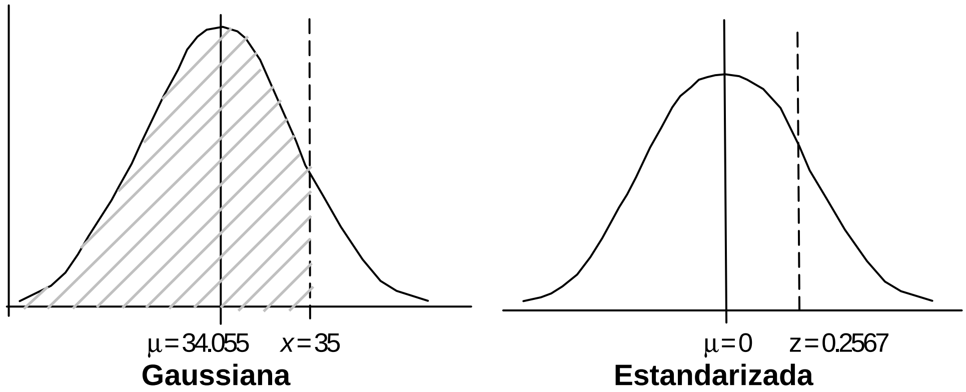
Hallar esa **probabilidad** equivale a calcular un **área bajo la curva normal estandarizada**.
Para ello conviene seguir estos tres pasos **(recuérdalos siempre)**:

Paso I. Estandarizar los valores *x* de interés, lo cual se hace aplicando $z = \frac{x-\mu}{\sigma}$.

Juan sabe que *x* = 35 es el valor de interés, σ = 3.6811 y μ = 34.055, por lo tanto:

$$z = \frac{35 - 34.055}{3.8522} = 0.2453$$

Paso II. Dibujar un esbozo de las curvas normales, tanto de la gaussiana *f*(*x*) como de la estandarizada *f*(*z*), y ubicar a *"x"* a la derecha o a la izquierda de μ, lo mismo que a "*z*", (sombrea el área de interés bajo la normal estandarizada):



Paso III. Encontrar el valor del área de interés a partir de los valores *z* obtenidos. Esto se hace utilizando tablas como la del Apéndice III.

En esa tabla, Juan determina que el área entre -∞ y 0.2453 es 0.59871

Puesto que un área bajo la curva *f*(*z*) es una probabilidad, el resultado es:

P(X < 35) = 0.59871 **(significa que al seleccionar un lobo mexicano al azar, la probabilidad de que pese menos de 35 kg es de 0.59871).**

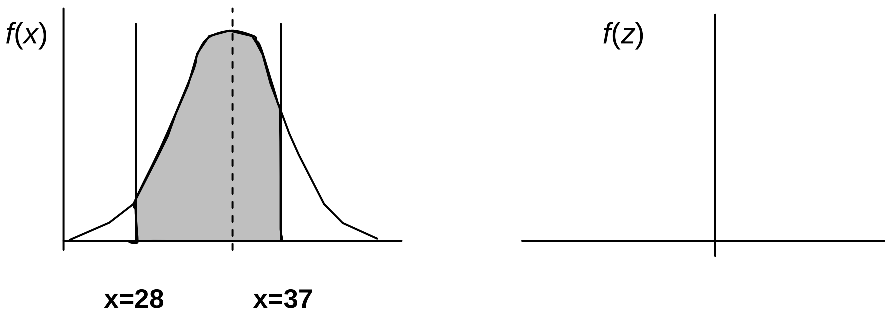
Dicho de otra manera, 59.871 % de los lobos que sean seleccionados al azar pesará menos de 35 Kg.
La pregunta: **¿Cuál es la probabilidad de que un lobo seleccionado al azar tenga un peso comprendido entre 28 y 37 Kg?**

Juan la responde siguiendo los pasos vistos arriba (llena lo que hace falta en cada paso).

Paso I. **Estandariza** los valores *x* de interés por medio de $z = \frac{x - \mu}{\sigma}$.

Los valores de la desviación y de la media son σ = 3.8522 y μ = 34.055
Si x_1 = 28, entonces z_1 = -1.5718; si x_2 = 37, entonces z_2 = _______

Paso II. **Dibuja** dos esbozos de las curvas normales: uno de la gaussiana *f*(*x*) donde marca los valores de *x*, y otro de la estandarizada *f*(*z*), donde marca los valores *z*, además sombrea las áreas de interés (dibuja el otro esbozo):



Paso III. Mediante la tabla del Apéndice III se **encuentra el valor del área** de interés:

Para z_1 = -1.5718, el área en la cola izquierda nos da P(X < 28) = 0.05821

Para z_2 = 0.7645, el área en la cola izquierda nos da P(X < 37) = ______

Puesto que interesa el área entre -1.5718 y 0.7645, se debe hacer una resta de sus respectivas áreas (esto es más claro si observas el esbozo que hiciste):

$$P(X < 37) - P(X < 28) = \_\_\_\_\_\_ - \_\_\_\_\_\_\_$$

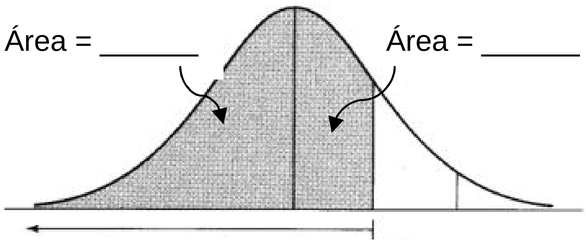
$$=$$

Por lo tanto, si se selecciona un lobo al azar, la probabilidad de que pese entre 28 y 37 kilos es de _______. ¿De qué otra manera se puede expresar el resultado? _________
____.

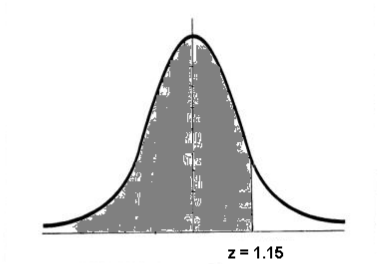
EJERCICIOS III

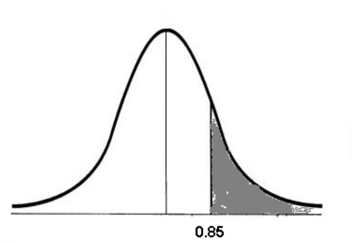
1. Escribe las cantidades correctas en la figura.

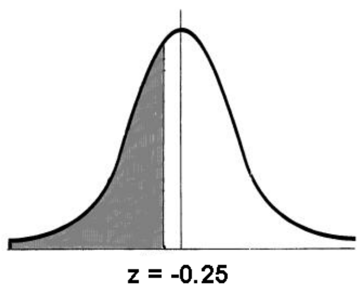
Área = ______ Área = ______

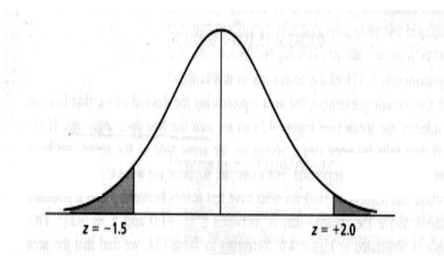


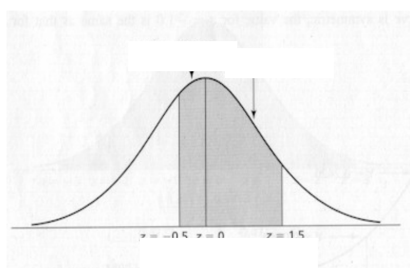
Área total sombreada = _______ z = 0.85

<table>
<tr>
<td>2. a. Encuentra el porcentaje de puntuaciones donde z es menor que 1.15


____________</td>
<td>b. Encuentra el porcentaje de puntuaciones donde z es mayor que 0.85


____________</td>
<td>c. Encuentra el porcentaje de puntuaciones donde z es mayor que 1.15
________________________</td>
</tr>
<tr>
<td>d. ¿Cuál es el porcentaje de puntuaciones donde z es menor que -0.25?


____________</td>
<td>e. ¿Cuál es el porcentaje de puntuaciones z menores que -1.5 pero mayores que 2.0?


____________</td>
<td>e. ¿Cuál es el porcentaje de puntuaciones donde z está entre -0.5 y 1.5?

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

3. Un médico registró el tiempo de vida de 2 000 enfermos a partir del momento en que se les detectó cierta enfermedad. Para poder predecir el tiempo de vida de un enfermo determinado, hace un estudio en el que:

a.- Toma muestras. ¿Qué debe hacer para que sean representativas? ¿De qué tamaño le conviene tomar cada muestra?

____.

b.- Obtiene la media de cada muestra. ¿Cuál de las medias será el mejor estimador de la media poblacional? ____.

c.- Construye la gráfica de la distribución de las medias muestrales. ¿Qué forma tendrá la distribución? _____________.

d.- Calcula la media de las medias muestrales $\bar{\bar{X}}$. ¿El resultado será un buen estimador de la media poblacional? ___________________________________.

e.- Calcula la desviación estándar $s_{\bar{x}}$ de las medias muestrales. ¿El resultado será un buen estimador de la desviación estándar poblacional? ____. A partir del valor de $s_{\bar{x}}$, ¿cómo se puede obtener el valor de la desviación estándar poblacional? ______ _______________________.

4. El salario mensual de los directores de 500 empresas se distribuye normalmente y su media aritmética es μ = \$51 800.00, con desviación estándar σ = \$730.30.

Si se obtiene una muestra aleatoria de tamaño 30, ¿cuál es la probabilidad de que la media aritmética de la muestra difiera cuando mucho por \$500.00 de la media de la población? ___________________

5. La tabla de abajo contiene datos obtenidos por un criadero de perros.
Ahí se muestra la distribución de las temperaturas, en grados centígrados, de 194 hembras pastor alemán al momento del parto.

a. Dibuja el histograma y traza su polígono de frecuencias "suavizado" (si lo deseas, utiliza computadora).

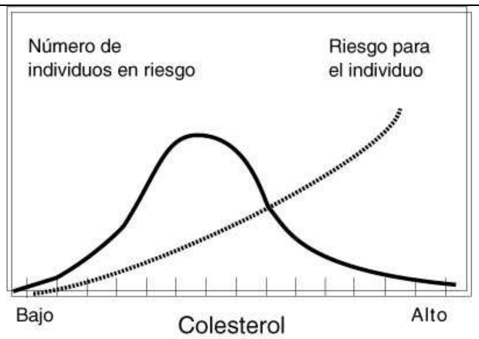
Temperatura	Frecuencia
37.15 – 37.35	6
37.35 – 37.55	13
37.55 – 37.75	19
37.75 – 37.95	26
37.95 – 38.15	37
38.15 – 38.35	35
38.35 – 38.55	25
38.55 – 38.75	17
38.75 – 38.95	11
38.95 – 39.15	5

b. Calcula la probabilidad de que al seleccionar una "pastor" al azar su temperatura sea:

i. Menor de 38.5℃

ii. Entre 37.5 ℃ y 38.7℃

iii. Mayor que 38.5℃ (aprovecha el resultado del p rimer inciso y el hecho de que el área total bajo la curva es 1).

	**6.** En la figura de la izquierda se muestra la curva de riesgo individual progresivo, por el nivel de colesterol, y la curva del nivel de colesterol en una población. La distribución de los individuos en riesgo, ¿es gaussiana?____________ _________. ¿Cómo es la proporción de individuos en alto riesgo con respecto a los de bajo riesgo? ______.

7. Una empresa construye 247 puertas de madera que se instalarán en un edificio Al medirlas, pasado un mes, cuando la madera ya sufre poca deformación, y antes de colocarlas, resulta que la media aritmética de su ancho es de 85.74 cm., con una desviación típica de 0.42 cm. El hueco dispuesto para las puertas permite que midan 85 cm. como mínimo y 86 cm. como máximo.

Además, la altura media de las puertas es de 1.91 m., con desviación típica de 0.5 cm., y el espacio donde se colocarán admite alturas de 1.92 ± 0.03 m.

Los anchos y las alturas de las puertas se distribuyen de manera normal.

a) Determina el porcentaje de puertas que habrán de repararse de lo ancho para poderse colocar. ____________.

b) Determina el porcentaje de puertas que habrán de repararse de lo alto para poderse colocar. ____________.

c) ¿Cuál sería el gasto extra de la empresa, si reparar cada puerta de lo ancho significa 60 pesos de gasto y 40 pesos repararla de lo alto? ____________.

8. Se considera que las calificaciones de cierta prueba de aprendizaje se distribuyen normalmente, con una media de 500 puntos y desviación típica de 100 puntos. ¿Cuál es la probabilidad de que una calificación seleccionada al azar sea mayor de 700 puntos?

9. Los estudios muestran que el uso de gasolina para automóviles compactos vendidos en Estados Unidos tiene una distribución normal, con una media de 25.50 millas por galón (mpg) y una desviación estándar de 4.50 mpg.

a. ¿Qué porcentaje de automóviles compactos rinde 30 mpg o más?

b. Un fabricante desea construir un automóvil compacto que supere en economía de combustible al 95% de los automóviles compactos actuales, ¿cuál debe ser el consumo de gasolina para el nuevo automóvil?

10. El departamento de carnes en un supermercado prepara específicamente sus paquetes de “un kilo” de carne molida. Se sabe que los pesos de los paquetes varían, algunos con un poco más y otros con un poco menos de un kilogramo. Los pesos de estos paquetes tienen una distribución normal con una media de 1.00 kg. y desviación estándar de 0.15 kg.

a. ¿Qué proporción de los paquetes pesará más de un kilogramo?

b. ¿Qué proporción de los paquetes pesará entre 0.95 y 1.05 kilogramos?

c. ¿Cuál es la probabilidad de que un paquete de carne molida seleccionado al azar pese menos de 0.80 kg.?

11. En el siguiente resumen faltan algunas expresiones, procede a escribirlas.

RESUMEN

En una **muestra,** a la media aritmética y a la desviación estándar se les denomina ______________ ________________. En una **población**, la media se representa por μ y la desviación estándar por σ y se denominan ______________**.**

Una variable aleatoria continua *X* sigue una **distribución normal,** de parámetros μ y σ, si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in R$$

Al cambiar la variable *x* por $z = \dfrac{x-\mu}{\sigma}$ la función anterior se transforma en:

f(z) = donde μ = ___, σ = __.

La distribución normal N(μ, σ) tiene las siguientes propiedades:

- Es **simétrica** con respecto a su ________ _____________.
- Se extiende desde –∞ hasta _____.
- Es **asintótica** al eje ___, nunca lo toca por mucho que se extienda.
- La media aritmética de la población coincide con la esperanza matemática de la variable aleatoria X y σ^2 es la ___________.
- Su **posición** depende del valor de la ________ _____________..
- Su **forma** depende de la _____________ ___________..
- El valor de la media coincide con el valor de la _______ y con el valor de la ___________.
- El **área bajo la campana representa un valor de la** ________________ P(*a*< X <*b).*

Hallar una **probabilidad** implica calcular un **área bajo la curva normal** ________________.

Para ello se utilizan tablas y conviene seguir estos 3 pasos:

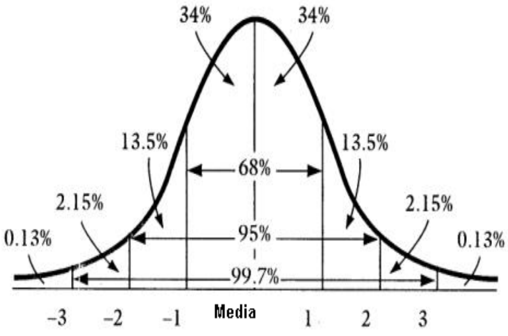
Paso I. **Estandarizar** los valores *x* de interés, mediante la fórmula z = ―――.

Paso II. **Hacer** un esbozo de las curvas normales, tanto de la gaussiana como de la _______________, marcar ahí los valores de *x* y de *z*; además, sombrear el área de interés en la normal estandarizada.

Paso III. **Encontrar el área de interés** mediante una tabla a partir de los valores *z* obtenidos.

12. Un automóvil viaja en un área residencial a 30 millas por hora (mph), la distancia necesaria para detenerse al frenar sigue una distribución normal, con media de 50 pies y desviación estándar de 8 pies. Si otro automóvil se para de pronto en su carril a 60 pies de distancia, al aplicar el conductor los frenos, calcula la probabilidad de: **a.** Detenerse en 40 pies o menos; **b.** Detenerte en 50 pies o menos; **c.** Evitar la colisión si sólo puede frenar hasta detenerse.

INVESTIGA: La denominada “regla empírica” tiene relación con diagramas como el de la derecha. Investiga el contenido del *Teorema de Chebishev* y da una explicación del diagrama.



Teorema de Chebishev: ____________________

____

____

____

____

____

Explicación del diagrama: ____

____

____

UNIDAD IV
EL TEOREMA CENTRAL DEL LÍMITE

Competencias por adquirir

- **Explicarás bajo qué condiciones se cumple y en qué consiste el Teorema Central del Límite.**
- **Verificarás el cumplimiento del Teorema Central del Límite.**

En general, para cualquier conjunto de datos está demostrado lo siguiente:

Si el tamaño *n* de la muestra se hace cada vez más grande, la distribución de la variable $\overline{X}$, media muestral, se aproxima más a una normal…………I

Una muestra se considera grande si tiene 30 o más elementos.

Está demostrado que la media de todas las medias de las muestras posibles, $\mu_{\bar{x}}$, tiene un valor igual a la media poblacional μ de los datos:

$$\mu_{\bar{x}} = \mu \text{…………………} \mathbf{II}$$

En la medida en que el tamaño ***n*** de la muestra sea mayor, se sabe que la desviación estándar de las medias muestrales, $s_{\bar{x}}$, se acerca cada vez más al valor de $\frac{\sigma}{\sqrt{n}}$.

Si la población es muy grande, la desviación $\sigma_{\bar{x}}$ de las medias de **todas las muestras posibles** y la desviación de la población, σ, se relacionan así:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{…………………} \mathbf{III}$$

De acuerdo a la igualdad (III), cuando el tamaño de la muestra es n = 2 o más, ¿cómo es σ comparada con $\sigma_{\bar{x}}$, mayor o menor? _________.

El valor de $\sigma_{\bar{x}}$ es una medida de que tan alejados están los valores $\bar{x}_i$ con respecto a la media poblacional μ. A partir de la igualdad III, al aumentar el tamaño ***n*** de la muestra, el valor de $\sigma_{\bar{x}}$, ¿aumenta o disminuye? ____________.

Es decir, a mayor tamaño de la muestra, menos dispersión de las medias $\bar{x}_i$ con respecto a μ (más pequeños los errores muestrales, que se definen como la distancia entre la media de una muestra y la media poblacional, $|\bar{x}_i - \mu|$).

El cociente $\frac{\sigma}{\sqrt{n}}$ se denomina **error estándar de la media**.

Al aumentar el tamaño de la muestra, el error estándar de la media ¿aumenta o disminuye? _____________.

Unidas, las igualdades I, II y III constituyen el

TEOREMA CENTRAL DEL LÍMITE

Cuando **X** es una variable **continua** con **parámetros** μ y σ (media y desviación estándar poblacionales) y $\overline{\mathrm{X}}$ es la variable cuyos valores son las medias aritméticas de muestras aleatorias de tamaño ***n*** (tomadas de la población de valores de **X**), con **parámetros** $\mu_{\overline{X}}$ y $\sigma_{\overline{X}}$, entonces:

I. Mientras mayor sea el tamaño *n* de la muestra, más cerca de ser normal estará la distribución de la variable $\overline{x}$

II. $\mu_{\overline{X}} = \mu$

III. $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Reconoce las notaciones empleadas.

VARIABLE	VALORES DE LA VARIABLE	MEDIA MUESTRAL	MEDIA POBLACIONAL	DESVIACIÓN ESTÁNDAR MUESTRAL	DESVIACIÓN ESTÁNDAR POBLACIONAL
X	x_i	$\overline{x}$	μ	s	σ
$\overline{\mathrm{X}}$	$\overline{x}_i$	$\overline{\overline{X}}$	$\mu_{\overline{X}}$	$s_{\overline{X}}$	$\sigma_{\overline{X}}$

EJERCICIOS COMPLEMENTARIOS IV

1. Una urna contiene cinco bolas numeradas del 1 al 5. Cada número indica su valor en pesos.

a. ¿Cuál es la media poblacional μ de los valores en pesos?

b. ¿Cuál es el valor de la desviación estándar poblacional σ?

c. Una persona extrae dos bolas. Escribe todas las muestras posibles, con reemplazo, junto a cada muestra escribe el valor $\overline{x}_i$ de su correspondiente media.

Muestra									
$\overline{x}_i$									

Muestra									
$\overline{x}_i$									

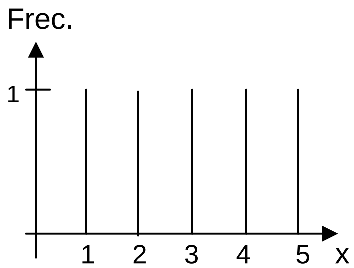
Muestra							
$\overline{x}_i$							

¿Cuántas muestras son? ____.

d. Construye la tabla de frecuencias de la media muestral $\overline{x}$.

e. La distribución de la población tiene forma "rectangular" (derecha).
¿Qué forma tendrá la distribución de las medias muestrales? ____________.

f. Compruébalo construyendo su distribución en forma semejante a la de la derecha (hazlo debajo de ésta).



g. La tabla de frecuencias para la media muestral (inciso d) constituye la distribución muestral de la media, calcula su media aritmética y desviación estándar. ________ ________

h. ¿Coincide la desviación estándar σ de la población de datos (calculada en el inciso b) con la desviación estándar de la distribución muestral (calculada en el inciso g)? ____

i. Calcula el error estándar de la media. _____

j. ¿Coincide la desviación estándar de la distribución muestral (calculada en el inciso g) con el error estándar de la media (calculada en el inciso i)? ____

2. El médico Juan Barradas registra el tiempo que duraron con vida 2 000 enfermos a partir del momento en que se les detectó cierta enfermedad. Para poder predecir el tiempo que durará con vida un enfermo determinado hace un estudio en el que:

a. Toma muestras. ¿Qué debe hacer para que sean representativas? ¿De qué tamaño le conviene tomar cada muestra?
b. Obtiene la media de cada muestra. ¿Cuál de las medias será el mejor estimador de la media poblacional?
c. Grafica la distribución de las medias muestrales. ¿Qué forma tendrá la distribución?
d. Calcula la media de las medias muestrales. ¿El resultado será un buen estimador de la media poblacional?
e. Calcula la desviación estándar de las medias muestrales. ¿El resultado será un buen estimador de la desviación estándar poblacional?

3. En el cuadro siguiente redacta un resumen del tema *Teorema Central del Límite.*

UNIDAD V-A
ESTIMACIÓN POR INTERVALO DE LA MEDIA ARITMÉTICA DE UNA POBLACIÓN

COMPETENCIAS POR ADQUIRIR
- **Para resolver un problema dado, determinarás un intervalo de confianza para la media aritmética de la población.**
- **Interpretarás el significado del tamaño del intervalo de confianza, relacionándolo con la precisión.**
- **Identificarás las propiedades de la distribución t de Student.**

Intervalo de confianza para la media poblacional μ cuando σ es conocida

Cuando se pretende calcular la media aritmética poblacional de una variable **continua,** pero no se pueden abarcar todos los datos de la población, lo más práctico es tomar una muestra aleatoria y calcular su media muestral $\bar{x}$, cuyo valor es una aproximación, o **estimación**, del valor del parámetro poblacional μ.

A partir de una muestra, no se puede saber con total seguridad si el valor del estimador $\bar{x}$ es igual al valor del parámetro μ, pero se puede averiguar en términos probabilísticos qué tanto se aproxima $\bar{x}$ a μ:

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 - α)100% de confianza** para μ es:

$$\left(\bar{x} - z\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z\frac{\sigma}{\sqrt{n}}\right)$$

(z es el valor de la distribución normal estándar que corresponde a "1 - α en área central" o a "α en dos colas").

El intervalo

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

permite hacer una **inferencia –o predicción- sobre μ** a partir del valor $\bar{x}$, siempre y cuando se conozca la desviación estándar σ de la población.

Las inferencias implican siempre un riesgo de cometer error, o sea, α nunca podrá ser cero. Por lo mismo la confiabilidad 1 - α nunca podrá ser ______.

Lo anterior tiene un inconveniente: en casos prácticos es difícil saber de antemano la desviación estándar poblacional σ.

Intervalo de confianza para la media poblacional μ cuando σ no se conoce

Si la variable continua de interés X se distribuye normalmente, entonces las medias muestrales $\bar{x}_i$ se distribuyen ______________. Esto se puede asegurar por el Teorema __________ del _________.

Como la distribución de $\bar{X}$ es normal, se puede utilizar la tabla de la distribución normal estándar para el valor *z* de la variable Z, calculado así:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Como *s* es un estimador de σ, ¿en vez de la expresión anterior podría utilizarse

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} ?$$

La diferencia entre $\frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$ y $\frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$ se debe a la diferencia que hay entre el parámetro σ y su estimador *s.*

Pero en la medida en que el tamaño *n* de la muestra sea más grande, *s* estima mejor a σ, es decir, tienen un valor más cercano entre sí.

En la fórmula para calcular la desviación típica *s* de la muestra:

$$s = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n-1}},$$

al aumentar *n*, por estar en el denominador de la fórmula, provoca una disminución de *s* y al mismo tiempo que tenga un valor cada vez más cercano a σ. O sea, si en la fórmula para *s* el denominador *n* - 1 aumenta, entonces *s* estimará mejor a σ.

¿Qué representa “n”? ___________________________________.

Al valor de *n*-1 se le conoce como los **grados de libertad** (g. l.) de *s*.

Por lo tanto, entre más *g.* l. haya, más se aproximará *s* a σ; a la vez, más próximo estará $\frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$ de $\frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$.

La distribución de $\frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$ no es normal estándar porque contiene *s* en vez de σ, pero debe parecerse más a la normal estándar mientras más próxima esté *s* de σ, o sea, mientras más sean los *g*. l.

Cuando la variable X se distribuye normalmente, los valores de

$$\frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

siguen la distribución **“*t* de Student”**, la cual permite hacer **estimaciones por intervalo** de la media aritmética poblacional.

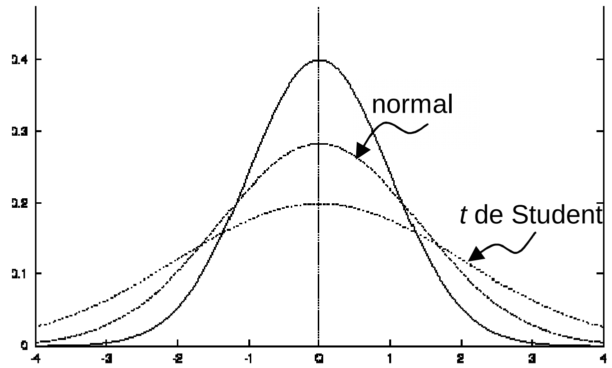
La distribución *t* de Student (o simplemente *t*) se caracteriza por lo siguiente:

1. Existe una distribución *t* para cada valor *n* – 1 de los grados de libertad.

2. Es **simétrica** con respecto a un eje vertical en $t = 0$, lo mismo que la distribución normal estándar con respecto al eje $z = 0$.
3. Su media aritmética es $t = 0$, para la distribución normal estándar es $z = 0$.
4. Su desviación típica es mayor que la unidad (**$\sigma > 1$**), lo cual significa una mayor dispersión comparada con la distribución normal estándar, para la cual $\sigma = 1$.
5. A medida que es **mayor** el tamaño n de la muestra, **los *g. l.* aumentan** y la distribución t de Student se parece más a la normal:

Teóricamente, en el límite ***n* = ∞** las distribuciones normal estándar y t de Student coinciden.

6. Mientras la distribución normal se puede transformar en una distribución normal estándar, esto **no** es posible en el caso de una distribución t de Student.



Asocia las columnas mediante flechas

- Permite estimar μ cuando se conoce σ
- Su desviación típica es mayor que 1
- Grados de libertad
- Se decide arbitrariamente
- Permite estimar μ cuando no se conoce σ
- Longitud del intervalo de confianza
- Media aritmética cero
- Forma familias de acuerdo a los valores de n-1

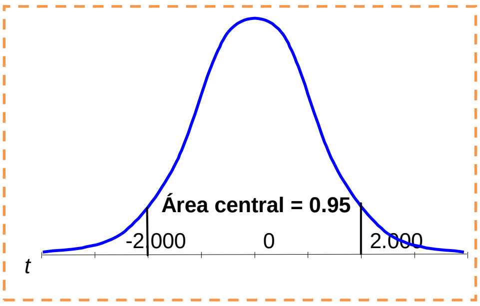
Distribución “t” de Student
Distribución normal
Distribución normal estandarizada
Intervalo de confianza
Media muestral
Precisión
Tamaño de la muestra

ESTIMACIÓN POR INTERVALO DE LA MEDIA ARITMÉTICA DE UNA POBLACIÓN

Con los diferentes valores de n - 1 se puede construir una tabla de valores t, (Apéndice V). En esa tabla, ¿cómo son los valores del último renglón comparados con los del último renglón de la tabla de valores z? __________. ¿Por cuál de las propiedades de la distribución “t” se explica esta coincidencia? ______________________________.

Puesto que cada renglón de la tabla del Apéndice V corresponde a una distribución distinta, nos referiremos a los valores de t indicando los grados de libertad con un subíndice. Por ejemplo, $t_{(16)}$ se refiere a un valor de t con 16 grados de libertad y por ende a una muestra de 17 datos.

En la tabla del Apéndice V, en el renglón “1- α en área central”, hay distintos valores de 1- α. Si por ejemplo 1- α = 0.95 y *g. l.* = 60, entonces $t_{(60)} = 2.000$; esto significa que la probabilidad de que un valor t esté comprendido entre -2.000 y 2.000 es 0.95, es decir:



$$P(-2.000 < t < 2.000) = 1- \alpha$$
$$= 0.95$$

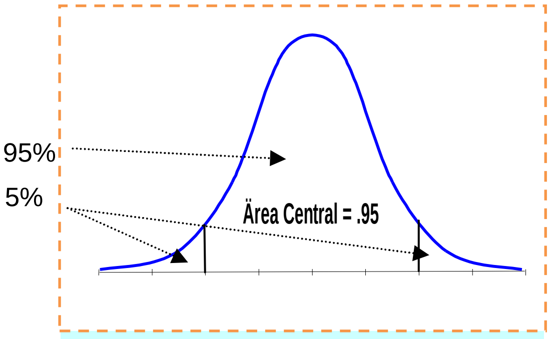
¿Qué se representa en el eje de las ordenadas de la distribución t?

______________________.

El área que corresponde a 1- α es la central y es el 95% del área total.

El valor 1- α indica la **confiabilidad** con la cual, al tomar una muestra de tamaño 61 y calcular $\dfrac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$, este valor se localizará entre -2.000 y 2.000. Dicho de otra manera: si se pudieran tomar todas las muestras posibles de tamaño 61 (o un número muy grande de ellas) y para cada una calcular $\dfrac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$, además de distribuirse normalmente, el 95% de los valores calculados estaría entre –2.000 y 2.000.

En consecuencia, 5% serían "menores que -2.000 o mayores que 2.000", o sea:



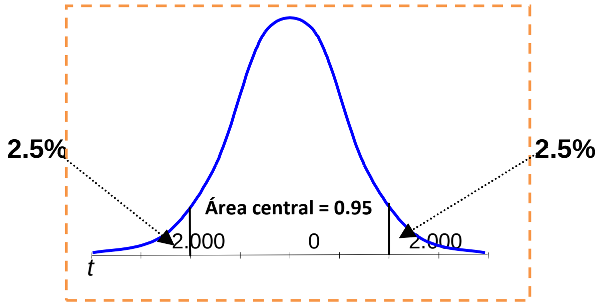
P (-2.000 > t > 2.000) = α
= 0.05

El área de ambas colas es el 5% del área total y se reparte equitativamente, entonces el área de cada una es 2.5%.

Es decir, α = 0.05; $\frac{\alpha}{2}$ = 0.025, por lo cual $t_{(60)}$ = 2.000 también corresponde a α = 0.05 en el renglón "α en dos colas" y a α = 0.025 en el renglón "α en una cola". Entonces

P (t > 2.000) = $\frac{\alpha}{2}$
= 0.025

y por simetría: P (t < -2.000) = $\frac{\alpha}{2}$
= 0.025



Escribe los valores de:

$t_{(20)}$ con 1- α = .80 en área central _____

$t_{(40)}$ con α = .02 en dos colas _____

$t_{(30)}$ con α = 0.005 en una cola _____

Los límites del intervalo de confianza para μ, cuando se conoce σ, se calculan por medio de:

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} \textbf{ (límite inferior)}, \quad \bar{x} + z\frac{\sigma}{\sqrt{n}} \textbf{ (límite superior)}$$

En casos más apegados a la realidad, cuando la variable bajo estudio X se distribuye en forma parecida a una normal, con σ desconocida, los límites del intervalo (abierto) de confianza para μ se pueden calcular por medio de:

$$\bar{x} - t_{(n-1)}\frac{s}{\sqrt{n}} \textbf{ (límite inferior)}, \quad \bar{x} + t_{(n-1)}\frac{s}{\sqrt{n}} \textbf{ (límite superior)}$$

Problema. En La Paz, Baja California Sur, la Dra. Bertha Leyva está interesada por saber la estatura promedio de las estudiantes que asisten a las universidades del Estado. Con ese propósito toma una muestra aleatoria de 121 jóvenes y resulta que su estatura promedio es $\bar{x}$ = 1.62 m. y que s = 0.16 m.

De inicio supone que las estaturas de la población bajo estudio (variable continua) **se distribuyen normalmente** y procede a construir un intervalo de confianza para μ. Con ese fin elige una confiabilidad de 60%.

Para encontrar el intervalo de confianza

$$\left(\bar{x} - t_{(n-1)} \frac{s}{\sqrt{n}}\ ,\ \bar{x} + t_{(n-1)} \frac{s}{\sqrt{n}}\right),$$

los datos conocidos son:

$$\bar{x} = 1.62$$
$$s = 0.16$$
$$n = 121$$

El valor de t lo encuentra en la Tabla de la distribución "t de Student" con 1 - α en área central = 0.60 y 121 – 1 = 120 g. l.; ahí observa que $t_{(120)} = 0.845$

Al sustituir los datos conocidos obtiene:

$$\bar{x} - t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 - 0.845 \frac{0.16}{\sqrt{121}}$$
$$= 1.607$$
$$\bar{x} + t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 + 0.845 \frac{0.16}{\sqrt{121}}$$
$$= 1.632$$

Por lo anterior, la doctora Leyva interpreta que: con una confiabilidad del 60%, la estatura promedio de las jóvenes que asisten a las universidades del Estado de Baja California tiene un valor comprendido entre 1.607 m. y 1.632 m.

Al reflexionar en el resultado anterior, la doctora prefiere tener una estimación de μ con mayor confiabilidad y decide que ésta sea de 90% (las confiabilidades preferidas en las investigaciones son de 90 y de 95 por ciento).

La tabla de la distribución "t de Student", para 1 - α en área central, indica $t_{(120)} = 1.658$. En consecuencia:

$$\bar{x} - t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 - 1.658 \frac{0.16}{\sqrt{121}}$$
$$= 1.596$$
$$\bar{x} + t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 + 1.658 \frac{0.16}{\sqrt{121}}$$
$$= 1.644$$

Es decir, con una confiabilidad del 90%, la estatura promedio de las jóvenes que asisten a las universidades del Estado de Baja California tiene un valor comprendido entre 1.596 m. y 1.644 m.

Con la finalidad de obtener todavía mayor confiabilidad, la doctora Leyva decide fijarla en 99%. Siendo así, ¿qué valor de $t_{(120)}$ se obtiene de la tabla para la distribución t de Student? _____ .

Por lo tanto:

$$\bar{x} - t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 - \_\_\_\_\_\_\_ \frac{0.16}{\sqrt{121}}$$
$$= \_\_\_\_\_\_\_\_$$

$$\bar{x} + t_{(n-1)} \frac{s}{\sqrt{n}} = 1.62 + \underline{\qquad\qquad} \frac{0.16}{\sqrt{121}}$$

$$= \underline{\qquad\qquad}$$

Entonces, con una confiabilidad del 99%, la estatura promedio de las jóvenes que asisten a las universidades del Estado de Baja California tiene un valor comprendido entre 1.582 m. y 1.658 m. ¿De acuerdo?

En esta tabla se reúnen los resultados:

CONFIABILIDAD	INTERVALO	TAMAÑO DEL INTERVALO
60%	(1.607 , 1.632)	0.025
90%	(1.596 , 1.644)	0.048
99%	(1.582 , 1.658)	0.076

Observa los intervalos obtenidos. Cuando la confiabilidad se incrementa, ¿qué pasa con el tamaño del intervalo, aumenta o disminuye? __________ .

Por lo tanto, si se quiere mayor confiabilidad se debe estar dispuesto a sacrificar la exactitud. En el límite, una confiabilidad del 100% implica un intervalo de tamaño infinitamente grande. Por lo tanto, en un problema real, ¿es posible tener una confiabilidad del 100%? ____.

Problema. La bióloga Silvia Lozano realiza un estudio del tiempo de gestación de los gorilas hembra en Nigeria, El Congo y Camerún, África. De los 30 ejemplares que observó, obtuvo una media aritmética de 270 días y una desviación típica de 47 días. Después calculó, con una confiabilidad de 95%, el tiempo de gestación de la población de gorilas de la región, de lo cual obtuvo el intervalo, en días, de (252.45, 287.54) ¿Es correcto su resultado?, compruébalo enseguida.

Los datos son:

$\bar{x}$ = ____ días

s = ___ días

n = ___

El valor de t en la tabla de la distribución "t de Student", con $1 - \alpha$ en área central = 0.95 y con 30 – 1 = 29 g. l., es

$$t_{(29)} = \underline{\qquad\qquad}$$

Al sustituir los datos conocidos se obtienen los límites del intervalo:

$$\bar{x} - \frac{\qquad}{\underline{\qquad\qquad}} =$$

$$=$$

$$=$$

$$\bar{x} + \frac{\qquad}{\underline{\qquad\qquad}} =$$

$$=$$

$$=$$

¿Son correctos los resultados obtenidos por la Dra. Silvia? ____.

EJERCICIOS COMPLEMENTARIOS V.1

1. El físico Roberto Ortega investiga la penetración en la roca de las partículas β (emitidas por el uranio radiactivo 238). Después de hacer una gran cantidad de mediciones, Roberto quiere calcular un intervalo de confianza para la media poblacional; para ello toma una muestra y calcula su media aritmética.
Marca con ✓ si la afirmación es correcta, con × si es incorrecta:

- Si Roberto conoce la desviación estándar poblacional, debe aplicar $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ ( )
- Si desconoce la desviación estándar poblacional puede aplicar $\bar{x} \pm t_{(n-1)}\frac{s}{\sqrt{n}}$ ( )
- Para aplicar cualquiera de los criterios anteriores la distribución de las mediciones debe ser normal y la variable debe ser continua ( )
- Si el número de mediciones es tan grande que puede considerarse infinita, en vez de $\frac{\sigma}{\sqrt{n}}$ deberá utilizar $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ ( )
- La confiabilidad la debe decidir según sea el tamaño de la población ( )

2. 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 49, 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 ponérsela de nuevo es de 3.6 minutos. Si se sabe que σ = 0.7, con 95% de confianza, estima por intervalo un valor para la media poblacional.

3. En el Municipio de Tatalpa, al suroeste del Estado de Jalisco, se miden las alturas de una muestra aleatoria de 121 saucos, *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 aritmética de los saucos de la región?

4. En la empresa *RAPIDITO* de camiones foráneos se interesan por saber la velocidad promedio con que viajan sus unidades al ir del Distrito Federal a Acapulco. Con ese propósito seleccionaron al azar 31 viajes y encontraron que la media aritmética de sus velocidades promedio fue de 74.8 Km./h. y su desviación típica de 21.3 Km./h.

Con una confiabilidad de 90%, ¿qué velocidad promedio debe esperar Esther Oseguera de un camión que aborda en el Distrito Federal para ir a Acapulco?

5. Determina un intervalo de confianza de 95% para los casos siguientes:

a. $\bar{x} = 15.0$	b. $\bar{x} = 15.0$
$\sigma_x = 2.0$	$\sigma_x = 2.0$
n = 100	n = 16
N = 1000	N = 200
____________________	____________________

Idea: primero determina si la muestra, de tamaño n, es mayor del 5% de la población.

Dependiendo del resultado decide si se utiliza $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$, o bien $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

UNIDAD V-B
ESTIMACIÓN POR INTERVALO DE PROPORCIONES POBLACIONALES

COMPETENCIAS POR ADQUIRIR

- **Resolver problemas que requieran determinar un intervalo de confianza para la proporción poblacional, el error de estimación o el tamaño de la muestra.**
- **Interpretar el significado del tamaño y conceptos inherentes al intervalo de confianza**
- **Relacionar el intervalo de confianza con la precisión, confiabilidad, tamaño de la muestra y error de estimación.**

Cuando la variable es categórica y el parámetro es la proporción *p* con la que ocurre cierta categoría de la variable, el procedimiento a seguir es semejante al estudiado cuando se obtuvo un intervalo de confianza para la media aritmética de una población.

Es decir, a partir del valor $\hat{p}$ con que ocurre la proporción en una muestra, se estima del valor de *p*. La estimación podrá ser puntual o por intervalo.

ESTIMACIÓN PUNTUAL DE PROPORCIONES POBLACIONALES

La estimación puntual de la proporción *p* poblacional se obtiene así:

·· Al azar, se toma una muestra de la población.

·· De la muestra tomada se calcula el porcentaje (proporción) de datos que tienen cierta característica deseada.

·· Se infiere: La proporción obtenida para la muestra se acepta como válida para toda la población.

Por ejemplo, de los miles de migrantes centroamericanos que llegan por la frontera sur de México cada año, interesa saber cuántos son mujeres. Una estimación puntual sería tomar una muestra aleatoria de 100 migrantes, calcular el porcentaje de mujeres que la componen y suponer que el resultado es válido para toda la población.

Si dos personas, de manera independiente, hacen una estimación puntual de una proporción, ¿obtendrán el mismo resultado? ______________________________. ¿Cómo se podría decidir cuál de las dos hizo la mejor estimación? ________________

____.

ESTIMACIÓN POR INTERVALO DE PROPORCIONES POBLACIONALES

Una proporción nos indica la parte del todo que posee determinada característica, en particular cuando se trata de una **variable categórica**.

Se habla de una proporción cuando se dice que en un cine 56% de los espectadores son menores de edad, o cuando se afirma que 2 de cada 10 habitantes de una comunidad viven del comercio...

Problema. En la imprenta *Nuevo Mundo* se hace un tiraje de 30 000 ejemplares de un libro y al editor Fernando del Río le interesa saber la proporción *p* (de ese tiraje) que se imprimió con algún defecto.

Primero efectúa una **estimación puntual** calculando la proporción *p* para una muestra de 3 000 libros que selecciona al azar. Resulta que de los 3 000 ejemplares, 45 están defectuosos, es decir, $\hat{p} = \frac{45}{3000} = 0.015$, pero sabe que su resultado puede variar si toma otra muestra; además, al calcular dos o más proporciones para distintas muestras, tendría dificultad para decidir cuál de ellas sería el mejor estimador, pues no sabría cuál estaría más cerca de la media poblacional. ¿Qué le conviene hacer a Fernando?

En adelante, la proporción con la que ocurre una categoría de la variable **en una población,** se denotará por *p*; la proporción con la que ocurre la categoría en diferentes **muestras** de tamaño *n* por $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$, o en una muestra cualquiera por $\hat{p}$.

Damos por válido este

Teorema. Si de una población se extraen muestras de tamaño *n*:

A. Los valores $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$ de las proporciones con que se presenta en cada muestra la categoría de interés, se distribuyen en forma normal, si *n* es grande.

B. La media aritmética poblacional del conjunto $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$, la cual denotaremos por $\mu_{\hat{p}}$, es igual a la proporción poblacional *p*:

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

C. La desviación estándar de los valores $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$, cuya notación será $\sigma_{\hat{p}}$, se obtiene mediante:

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

(La raíz se denomina “error estándar de la proporción estimada $\hat{p}$”)

CONTINÚA SÓLO SI PUEDES EXPRESAR, SIN LEERLO, EL TEOREMA ANTERIOR.

Antes, *n* se consideró grande cuando era mayor o igual que 30; ahora *n* se considerará grande cuando se cumplan estas dos condiciones:

$$np > 5, \; n(1 - p) > 5 \qquad \text{(I)}$$

La proporción *p* con la que ocurre en la población una de las categorías de una variable categórica, se puede estimar construyendo un intervalo de confianza para *p* con una confiabilidad predeterminada de 1 - α. FALTA

En una población en la que ocurre una categoría de cierta variable categórica, si se extrae una muestra aleatoria de tamaño *n* para la cual la cual la categoría se presenta en una proporción $\hat{p}$ y si además n$\hat{p}$ > 5 y $n(1 - \hat{p}) > 5$, entonces un intervalo de (1 - α) 100% de confianza para la proporción poblacional *p* es:

$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\;,\; \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

***z* en la distribución normal estándar corresponde a “(1 - α) en área central” o a “α en dos colas”.**

Este resultado es muy útil para estimar la proporción en que una de las categorías de una variable categórica se presenta en la población, basta con extraer una muestra grande, calcular la proporción en la que ahí se manifiesta la categoría y calcular los límites del intervalo de confianza utilizando la distribución normal estándar.

Calcular $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ equivale a construir un intervalo de confianza predeterminada, centrado en $\hat{p}$ y de longitud igual a dos veces el **error de estimación** $z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Regresemos al problema del editor Fernando del Río. Él quiere conocer la proporción poblacional, *p*, de libros que se imprimieron con algún defecto, en un tiraje de 30 000 ejemplares. En este caso la variable categórica pudiera ser "calidad de las impresiones" y sus categorías: "con defectos" y "sin defectos".

Al revisar una muestra aleatoria de 3 000 ejemplares obtiene que 45 tienen defectos de impresión $\frac{45}{3000}$ = 0.015; por lo cual sus datos conocidos son:

$$n = 3\,000,\ \hat{p} = 0.015$$

Ya identificados sus datos, verifica que se cumplan las dos condiciones para aceptar que el tamaño de la muestra es grande (escribe los signos de desigualdad):

$n\hat{p}$ __ 5: $\quad 3\,000(\frac{15}{1000})$ __ 5

45 __ 5 ¿Se cumple la desigualdad? ___.

$n(1-\hat{p})$ __ 5 $\quad 3\,000(1 - \frac{15}{1000})$ __ 5

$3\,000(\frac{985}{1000})$ __ 5

2 955 __ 5 ¿Se cumple la desigualdad? ___.

Después, Fernando fija una confiabilidad de 95%. Para ese porcentaje la tabla indica el valor crítico z = 1.96; con lo cual procede a calcular los límites del intervalo de confianza:

$$\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.015 - 1.96\sqrt{\frac{0.015(0.985)}{3000}} = 0.015 - .004349$$

$$= 0.0106$$

$$\hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.015 + 1.96\sqrt{\frac{0.015(0.985)}{3000}} = 0.015 + .004349$$

$$= 0.019349$$

Esto le indica que, con una confiabilidad de 95%, en los 30 000 ejemplares impresos, la proporción de libros defectuosos está entre 0.0106 y 0.019349; o sea: con una confianza de 95%, por cada 100 libros impresos, entre 1 y casi 2 salieron defectuosos.

En el supuesto de que Fernando repitiera el procedimiento 1 000 veces, tomando muestra del mismo tamaño, y calculara en cada ocasión un intervalo de confianza, ¿obtendría el mismo intervalo siempre? ___. ¿Cuál sería la razón? ________________

__________________. De los 1 000 intervalos, ¿cuántos podría asegurar que contienen el valor real de la proporción poblacional *p*? _____.

Problema. En Altamira, Tamaulipas, se realizó la vacunación contra la hepatitis B. Se consideraron personas mayores de 17 años que se vacunaron; a 130 de ellas se les pregunta al respecto, resultando que sólo 44 se vacunaron. Con 90% de confiabilidad, ¿entre cuáles valores se encuentra la proporción de personas vacunadas en Altamira?

Solución: $n = 130$, $\hat{p} = \frac{44}{130} = 0.34$ y $(1 - \alpha)\ 100\% = 90\%$

Puesto que: $n\hat{p} = 130(0.34) = 44.2 > 5$

$n(1 - \hat{p}) = 130(0.66) = 85.8 > 5$

$\therefore$ la muestra es grande

Entonces: $\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.34 - 1.645\sqrt{\frac{0.34(1-0.34)}{130}} = 0.272$

$\hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.34 + 1.645\sqrt{\frac{0.34(1-0.34)}{130}} = 0.408$

Por lo tanto, la proporción de personas mayores de 17 años que se vacunaron en Altamira está entre _____% y _____% con una confiabilidad de 90%.

Problema. Mediante una muestra aleatoria de tamaño 500, se calcula la proporción de varones adultos, residentes en una población, con obesidad severa (Índice de Masa Corporal, IMC, menor que 40 pero mayor que 30). Se obtiene una estimación de varones con obesidad severa del 18%. Utilizando un nivel de confianza del 98%, los médicos afirman que el error máximo que se comete al estimar esa proporción, por medio de un intervalo de confianza, es de 0.04, ¿tienen razón?

Datos: Confiabilidad requerida _____, n = _____, $\hat{p}$ = _____

En tabla se encuentra que el "valor crítico" z es ______. La expresión que indica el **error de estimación** es: __________

Al sustituir los valores se encuentra la solución:

Calcula tu IMC dividiendo tu peso en Kg. Entre el cuadrado de tu estatura en m.

Mi IMC = _____.

Si resultó mayor que 25, es mala señal, si resultó mayor que 30, recomendable acudir de inmediato al médico.

EJERCICIOS COMPLEMENTARIOS V-2

1. Un vendedor de libros se interesa en saber a qué proporción *p* de las personas de la Delegación Coyoacán, donde trabaja, le gusta leer novelas. Para ello, elige al azar a 50 personas y encuentra que a 37 les gusta leer novelas. Calcula un intervalo de 95% de confianza para *p*. ¿Conviene vender novelas en Coyoacán?.

2. En un anuncio publicitario se afirma que 8 de cada 10 médicos utilizan o recomiendan cierto producto. Un estudiante desconfiado elige al azar a 100 médicos y encuentra que 30 de ellos utilizan o recomiendan el citado producto. Con una confianza de 99%, realmente, a. ¿cuántos médicos de cada cien utilizan o recomiendan el producto?, b. ¿crees que lo que se afirma en el anuncio publicitario sea correcto?

3. En una muestra aleatoria de 40 padres de familia, se encontró que 27% no le dedicaba por lo menos 15 minutos diarios a conversar con sus hijos. Estima con una confianza de 95% el porcentaje de padres que no conversan con sus hijos ese tiempo.

4. En una encuesta aplicada a 16 niños seleccionados aleatoriamente en una ciudad, se encontró que 25% de los niños no tenían una alimentación adecuada. **a.** Calcula un intervalo de 95% de confianza para la proporción de niños de toda la ciudad que no reciben alimentación adecuada. **b.** Haz el cálculo con el 90% de confianza.

5. En Miahuatlán, Oax., de 100 votantes seleccionados al azar y entrevistados acerca de su preferencia sobre los candidatos a presidente municipal, 59 se manifestaron a favor de Ceferino Gil. **a.** Con una confianza del 99.9%, ¿le recomendarías a Ceferino comenzar a prepararse para la toma de posesión del cargo? **b.** ¿Se lo recomendarías con una confianza del 50%?

UNIDAD VI

PRUEBA DE HIPÓTESIS PARA UNA MEDIA POBLACIONAL

COMPETENCIAS POR ADQUIRIR

- **Identificar los conceptos básicos de la prueba de hipótesis: tipos de hipótesis, estimador y regla de decisión.**
- **Resolver problemas aplicando el procedimiento de prueba de hipótesis para conjeturas sobre una media poblacional, tanto si la desviación estándar poblacional es conocida como si es desconocida.**

¿QUÉ ES UNA HIPÓTESIS?. . .

Es una conjetura o proposición tentativa acerca de la relación entre dos o más variables o fenómenos.

La **hipótesis estadística** es una proposición, o supuesto, acerca del valor de un parámetro poblacional.

¿EN QUÉ FORMA SE ESCRIBE?. . .

La hipótesis puede aparecer en forma de oración afirmativa, negativa o interrogativa, y relaciona general, o específicamente, variables con variables.

¿QUÉ EXPRESAN?. . .

Expresan una solución o explicación **tentativa**, **racional** y **verificable** de un problema científico.

Escribe una hipótesis acerca de un tema mundial que consideres de interés.

____

I. **PRUEBA DE HIPÓTESIS A PARTIR DE UNA POBLACIÓN NORMAL CON DESVIACIÓN ESTÁNDAR DESCONOCIDA**

La **prueba de hipótesis** es un procedimiento que termina en una decisión acerca de una hipótesis planteada inicialmente.

Problema 1. La QFB Consuelo Ortiz trabaja en una empresa dedicada al mantenimiento de albercas, en Acapulco, Guerrero. Ella realiza un estudio en la zona “dorada” con el fin de comparar el pH de dos poblaciones: por un lado las albercas particulares y por otro las destinadas al uso turístico. Un mes antes su empresa había dejado a todas las albercas con el mismo nivel de pH; ahora, debido a las diferentes condiciones de uso y conservación, ese nivel sería diferente de una alberca a otra.

Ella supone que la variable

X: *nivel de pH*

se distribuye normalmente. **(El termino pH expresa el grado de acidez o alcalinidad del agua. Un bajo pH causa corrosión a las piscinas; un alto pH puede traer consigo generación de bacterias y algas en las tuberías).**

Consuelo tiene esta **hipótesis**:

El nivel promedio de pH de las albercas particulares, después de un mes, es diferente del nivel promedio de pH de las albercas turísticas.

Después de examinar **todas** las albercas de uso turístico que recibieron servicio de su empresa, encontró que su media aritmética del nivel de pH era:

$\mu = 7.5$ *(población de albercas turísticas)*

Sin embargo, por diferentes motivos no podía efectuar el estudio en todas las albercas particulares de la población, por lo que decidió hacerlo **en una muestra** aleatoria de 30 albercas.

Para la muestra de 30 albercas particulares, la media aritmética muestral del nivel de pH fue:

$\bar{x} = 7.1$ *(muestra de albercas particulares)*

con desviación estándar muestral:

$s = 0.64$

¿Difiere la media **poblacional** de las albercas turísticas de la media **muestral** obtenida para la **muestra** de albercas particulares? ____.

¿Los resultados permiten concluir, como lo supuso Consuelo al principio, que el nivel de pH de las albercas turísticas difiere del nivel de pH de las albercas particulares? _____ .

Si decidiéramos aceptar una **estimación puntual**, tomaríamos el valor de la media muestral $\bar{x}$, de las albercas particulares, como su media poblacional. Entonces podríamos comparar dos medias poblacionales (7.5 y 7.1), y la respuesta a la pregunta anterior sería sí, que los niveles de pH difieren, pero esto sería muy incierto, porque al tomar otra muestra su media podría ser igual o mayor que 7.5, llevándonos a una conclusión diferente.

La hipótesis de Consuelo se puede plantear, en lenguaje más preciso, así:

La media aritmética del nivel de pH de las albercas turísticas ***difiere*** *de la media aritmética del nivel de pH de las albercas particulares, al pasar un mes desde que tenían los mismos niveles.*

La hipótesis de Consuelo será llamada en adelante **hipótesis de investigación** (H_{inv}).

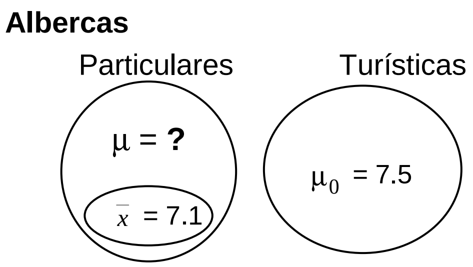


Diagrama VI.1

Si μ denota la media aritmética del nivel de pH de las albercas particulares, y $\mu_0 = 7.5$ es la media para las albercas turísticas, entonces la hipótesis de investigación queda expresada así:

H_{inv}: $\mu \neq \mu_0$, es decir: $\mu \neq 7.5$

En Estadística, la hipótesis de investigación se denomina **hipótesis alternativa;** aquí la representaremos por H_1:

H_1: $\mu \neq \mu_0$, es decir: $\mu \neq 7.5$

La hipótesis que **contradice** a la de investigación se denota por H_0 y se denomina **hipótesis nula** (porque considera que la diferencia entre μ y μ_0 es **nula**):

H_0: $\mu = \mu_0$, es decir: μ _____

Cuando se resuelve un problema, las referencias se hacen en términos de la **hipótesis nula.**

Verifica si entendiste.

Supón que Consuelo hubiera planteado como hipótesis **de investigación**:

H_{inv}: $\mu > \mu_0$, es decir: $\mu > 7.5$

En ese caso la hipótesis **alternativa** hubiera sido (escribe los símbolos de relación):

H_1: μ __ μ_0 , es decir: μ __ 7.5

Y como lo contrario de ">" es "≤", la hipótesis **nula** hubiera sido:

H_0: μ __ μ_0 , es decir: μ __ 7.5

Otra posibilidad de hipótesis **de investigación** hubiera sido:

H_{inv}: $\mu < \mu_0$, es decir: $\mu < 7.5$

Entonces, la hipótesis **alternativa** se hubiera escrito así:

H_1: μ __ μ_0 , es decir: μ __ 7.5

Y como lo contrario de "<" es "≥", la hipótesis **nula** hubiera sido:

H_0: μ __ μ_0 , es decir: μ __ 7.5

Completa la tabla:

H_{inv}:	H_1:	H_0:
$\mu \neq \mu_0$		
$\mu < \mu_0$		
$\mu > \mu_0$		

Riesgos de cometer error

A Consuelo le fue posible obtener la media del nivel de pH para la población de albercas turísticas de la zona dorada de Acapulco, (obtuvo μ_0 = 7.5), pero le fue imposible medir el pH de todas las albercas particulares, sólo pudo calcular una media muestral ($\bar{x}$ = 7.1).

Para poder comparar las medias de sus dos poblaciones necesita el valor de μ (la media aritmética del nivel de pH de las albercas particulares) ¿Cómo lo puede calcular?

Lo más sencillo es hacer una estimación puntual: aceptar que a partir del valor de la media muestral, $\bar{x}$ = 7.1, se **infiere** que la media poblacional de las albercas particulares es μ = 7.1, pero ya se mencionaron los inconvenientes de hacerlo.

De manera intuitiva, si el estimador $\bar{x}$ = 7.1 difiere "mucho" de μ_0 = 7.5, la hipótesis nula (H_0: μ = 7.5) se contradice. Si por el contrario, $\bar{x}$ = 7.1 difiere "poco" de μ_0 = 7.5, pareciera que la hipótesis nula es válida.

Esta es la diferencia entre las medias muestral y poblacional conocidas por Consuelo:

$$\bar{x} - \mu_0 = 7.1 - 7.5$$
$$= -0.4$$

¿Difieren mucho?

Es difícil contestar, para unas personas puede ser mucho y para otras puede ser poco, por ello se requiere aplicar un criterio lo más objetivo posible.

Ahora bien, la hipótesis nula puede ser verdadera o falsa, y la conclusión a la que se llegue, conocida como **conclusión estadística**, tiene dos posibilidades:

- **Se rechaza** la hipótesis nula.
- **No se rechaza** la hipótesis nula.

Nótese que se dice "rechazar" y "no rechazar", no se dice "rechazar" y "aceptar".

Por lo anterior, en la conclusión estadística se pueden cometer dos tipos de error:

Error de tipo I: Se comete al **rechazar** la hipótesis nula cuando en realidad es **verdadera**.

Error de tipo II: Se comete al **no rechazar** la hipótesis nula cuando en realidad es **falsa**.

En cambio, **no hay error** en la conclusión estadística si:

Se rechaza la hipótesis nula cuando en realidad es **falsa**.

No se rechaza la hipótesis nula cuando en realidad es **verdadera**.

Consuelo cometería el error de tipo I si rechazara la hipótesis nula:

$$H_0: \quad \mu = 7.5$$

cuando en realidad fuera verdadera.

Cometería el error de tipo II si no rechazara la hipótesis nula:

$$H_0: \quad \mu = 7.5$$

cuando en realidad fuera falsa.

La inferencia del parámetro μ a partir del valor de la media muestral $\bar{x}$ estará sujeta, como pasa con todas las estimaciones, a cierto margen de error. Por supuesto, siempre se pretenderá que la probabilidad de cometer un error, ya sea de tipo I o de tipo II, sea mínima.

REGLA DE DECISIÓN

Enseguida se darán los pasos hacia lo que se denomina la "regla de decisión"; para ello es necesario que la variable bajo estudio (en este caso X: *nivel de pH*) se distribuya normalmente. De cumplirse este requisito,

▸ Se **d**ecide la **probabilidad** α con la que se está dispuesto a cometer **error** en la estimación por intervalo de μ.

Si se decide que la probabilidad de cometer error sea α = 0.05, entonces el intervalo de confianza para μ será de: (1 - α)100% = 95% .

Si la decisión de que la probabilidad de cometer error sea α = 0.10, entonces el intervalo de confianza para μ será de: (1 - α)_____ = ___% .

▸ Se construye **el intervalo de confianza** para el denominado "**estadístico de prueba"**

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Si Consuelo se decide por α = 0.05, y sabiendo que: $n - 1 = 29$ *g. l.*, puede hallar en la tabla de valores t (Apéndice V):

$$t_{(n-1)} = t_{(29)} = 2.045$$

Por lo tanto: $- 2.045 < \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} < 2.045$ **)**

Para no cargar con toda la expresión del **estadístico de prueba**, conviene denotarlo con t_p:

$$t_p = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

El valor del estadístico de prueba t_p se puede calcular cuando se conocen los valores de $\bar{x}$, s, n y μ (éste se sustituye por μ_0 si se acepta como verdadera la hipótesis nula).

▸ Se supone que la hipótesis nula (H_0: $\mu = \mu_0$) es **verdadera** y se aplica la regla de decisión.

Al **suponer H_0 verdadera,** el valor del estadístico de prueba t_p, puede, o no, caer en el intervalo (-2.045, 2.045); con este motivo se aplica esta **“regla de decisión”**:

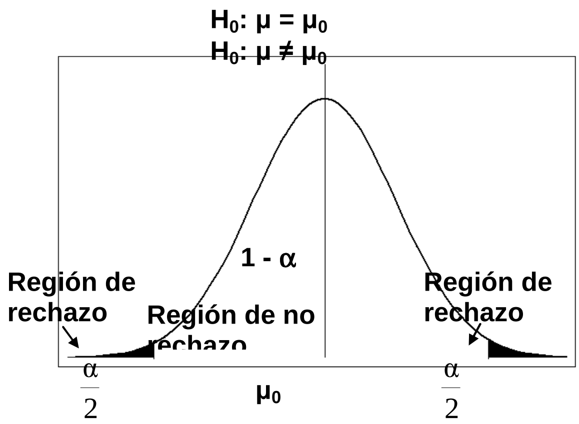
- Si t_p cae **dentro del intervalo** -lo cual es posible con una probabilidad muy alta ($1 - \alpha = 0.95$)-, entonces el supuesto de que H_0 es verdadera **no se contradice**. Decisión: **no se rechaza H_0**.
- Si t_p cae **fuera del intervalo** -lo cual es posible con una probabilidad muy baja ($\alpha = 0.05$)-, entonces **se contradice** el supuesto de que H_0 es verdadera. Decisión: **se rechaza H_0**.

La **regla de decisión** también es el criterio para determinar con mayor objetividad si el estimador $\bar{x}$ difiere “mucho” de μ_0, o difiere “poco”. Para entenderlo, analicemos la expresión para el estadístico de prueba $t_p = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

O bien $t_p \dfrac{s}{\sqrt{n}} = \bar{x} - \mu_0$ Los valores de s y n están fijos (lado izquierdo de la igualdad), entonces, si t_p es pequeño, provoca que la diferencia $\bar{x} - \mu_0$ sea pequeña (lado derecho), lo cual quiere decir que $\bar{x}$ casi es igual a μ_0.

- Si el valor de t_p cae dentro del intervalo de confianza, significa que **difiere poco** de cero, o sea, $\bar{x}$ difiere poco de μ_0. En este caso **no se debe rechazar la hipótesis nula**.
- Si el valor de t_p cae fuera del intervalo de confianza, significa que **difiere mucho** de cero, o sea, $\bar{x}$ difiere mucho de μ_0. En este caso **se debe rechazar la hipótesis nula**.

Abajo aparece la curva normal con sus regiones de rechazo y no rechazo:



Si el estadístico de prueba no se localiza dentro del intervalo de confianza, α se puede interpretar como la probabilidad de que se rechace H_0, siendo verdadera En otras palabras, es la probabilidad de cometer el error de tipo I al probar la hipótesis; y si β denota la probabilidad de cometer el error de tipo II

$$\alpha = P(E_I),\ \beta = P(E_{II})$$

En síntesis:

CONCLUSIÓN ESTADÍSTICA	**VALOR DE VERDAD DE LA HIPÓTESIS NULA**	
	H_0 VERDADERA	**H_0 FALSA**
SE RECHAZA H_0	Se comete el error de tipo I con una probabilidad α.	Con una probabilidad de 1 - β no se comete error.
NO SE RECHAZA H_0	Con una probabilidad de 1 - α no se comete error.	Se comete el error de tipo II con una probabilidad β.

Si la QFB Consuelo Ortiz aplica la regla de decisión:

Los datos son $\bar{x} = 7.1$, $\mu_0 = 7.5$, $s = 0.64$, $n = 30$.

$$t_p = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$t_p = \frac{7.1 - 7.5}{\frac{0.64}{\sqrt{30}}} = 3.423$$

¿Cae $t_p = 3.423$ dentro del intervalo de no rechazo (-2.045, 2.045)? ____. ¿Debe Consuelo **rechazar** la hipótesis nula? ___. ¿Difiere mucho o poco $\bar{x}$ de μ? ________.

Al rechazar la hipótesis nula, implícitamente se acepta como verdadera la hipótesis de investigación, con una confianza de 95%. Entonces, la conclusión de Consuelo es:

Con una confianza de 95%, después de transcurrido un mes, la media poblacional del nivel de pH de las albercas particulares es diferente a la media poblacional de pH de las albercas públicas.

PRUEBA DE HIPÓTESIS A PARTIR DE UNA POBLACIÓN NORMAL CON DESVIACIÓN ESTÁNDAR DESCONOCIDA

Resolución por pasos

Problema 2. Una fábrica de llantas produce las marcas Faires y Gudrich empleando procesos diferentes. Por estudios previos, se sabe que la duración de ambas marcas se distribuye de manera normal y que la duración media de la marca Faires es de 25 000 millas, desconociéndose la de la marca Gudrich. De ésta se selecciona una muestra aleatoria de 121 llantas producidas y se calculan su media y desviación estándar, obteniéndose 25 477 y 3 300 respectivamente, entonces, ¿la duración promedio de las llantas marca Gudrich es igual a la duración promedio de las llantas Faires?

Llena los huecos.

Paso 1. Se indican los datos del problema: $\mu =$, $n =$, $\bar{x} =$, $s =$

Paso 2. Se plantea la hipótesis nula. H_0: $\mu_0 =$ (es lo que se debe probar)

La hipótesis alternativa se puede enunciar así:

La media aritmética de la duración de las llantas marca Gudrich es diferente de la media aritmética de la duración de las llantas Faires.

Hipótesis alternativa. H_1: μ (es la hipótesis de investigación)

Paso 3. Se identifica la fórmula del estadístico de prueba t_p. En este caso se está probando una hipótesis acerca de la media de una población, cuyos valores siguen una distribución normal y se desconoce su desviación estándar:

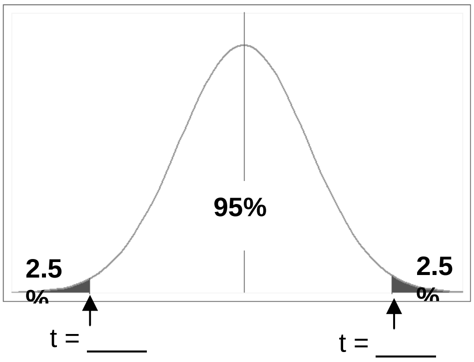
$$t_p = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

(el estadístico de prueba sigue una distribución "t de Student" con n – 1 = grados de libertad, si H_0 es verdadera)

Paso 4. Se determina el "nivel de significación", es decir, la probabilidad α de cometer el error de tipo (rechazar la hipótesis nula siendo).

Suponemos α = 0.5

Paso 5. Partiendo de que α = 0.5, se contesta la pregunta ¿Cuál es el valor de t a la derecha del cual está el 2.5% del área bajo la curva t de Student?



La respuesta está en el cuerpo de la tabla de valores t (Apéndice V). Para ello se considera que el valor de t a la derecha del cual está el 2.5% del área bajo la curva t de Student, es igual al valor t que tiene un área de 0.975 entre este valor y -∞:

t =

Asimismo, ¿Cuál es el valor de t a la izquierda del cual está el 2.5% del área bajo la curva t de Student?, la respuesta es:

t =

La parte sombreada de la figura es la "región de rechazo" o "región ", la parte no sombreada es la "región de ". A los valores que separan las regiones de rechazo y no rechazo (y) se les llama "valores ".

Regla de decisión:

Rechazar H_0 si el valor del estadístico de prueba está en la región de rechazo determinada por el intervalo:

≥ t ≥ (Se lee a partir del centro: "t es menor o igual que , y mayor o igual que ")

Paso 6. Se calcula el **estadístico de prueba** t_p y se determina la decisión estadística**.**

$$t_p = \qquad = \frac{25477 - 25000}{\frac{3300}{10}}$$

$$= 1.445$$

El valor del estadístico de prueba está en la región de , por lo tanto, con un nivel de significación de 5%, no se rechaza la hipótesis nula:

$$H_0\colon \mu = 25\,000$$

Viéndolo de otra manera, el intervalo de confianza para μ es

$$\bar{x} \pm t\,\frac{s}{\sqrt{n}} = 25\,477 \pm 1.980\,\frac{3300}{10}$$

$$= 25\,477 \pm 653.4$$

Es decir: (24 823.6, 26 130.4)

Dado que el intervalo de confianza incluye a $\bar{x}$ = 25 477, no se rechaza H_0, es decir, se llega a la misma conclusión.

Problema 3. Al estudiar el efecto del estrés sobre la presión arterial, en un libro, el doctor Othón Cortés lee: "*La presión sistólica media en varones jóvenes (entre 15 y 20 años) estresados, es de 19 cm de Hg".* De sus pacientes, tomó una muestra aleatoria de 36 jóvenes en ese rango de edad y obtuvo $\bar{x}$ = 19.7, lo cual lo llevó a plantear que en realidad *la presión sistólica media es mayor que 19 cm de Hg.* Suponiendo que las mediciones se distribuyen normalmente, ¿está en lo cierto el Dr. Othón?

Procede por pasos para averiguarlo.

1. Datos del problema:

____

2. Hipótesis nula:

_________ (es la hipótesis que se debe probar)

La hipótesis alternativa se puede enunciar así:

____

____.

H_1: ______ (es la hipótesis de investigación)

3. La fórmula del estadístico de prueba. En este problema, ¿los valores siguen una distribución normal? ___ ¿se conoce su desviación estándar? ____.

El estadístico de prueba tiene esta forma matemática: t_p = _____

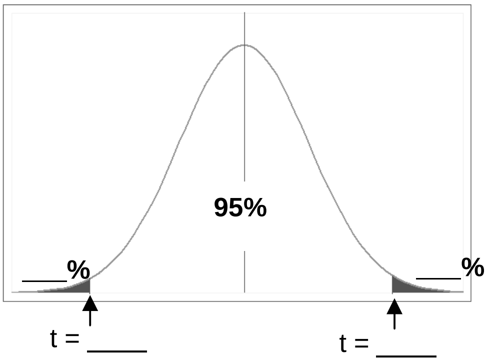
¿Qué tipo de distribución tiene el estadístico de prueba? ______________.

¿Cuántos son los grados de libertad, si H_0 es verdadera? ____.

4. Se determina el “nivel de significación”, es decir, la probabilidad α de cometer el error de tipo __, que consiste en rechazar la hipótesis ______ siendo ____________.

Supón $\alpha = 0.5$

5. Partiendo del valor de α seleccionado, se contesta la pregunta ¿Cuál es el valor de t a la derecha del cual está el ____% del área bajo la curva? ¿Cuál curva? _____________.



El valor de t (a la derecha del cual está el 2.5% del área bajo la curva t de Student), entre -∞ y él, tiene un área de _______ Su valor es:

t = _______

El valor de t, a la izquierda del cual está el 2.5% del área bajo la curva t de Student es:

t = ______

La parte sombreada de la figura es la ________ de __________ o región _________, la parte no sombreada es la ________ de ___ __________. A los valores que separan las regiones de rechazo y no rechazo, es decir, a ______ y ______ se les llama _________ ___________.

La regla de decisión consiste en:

____

____

6. Cálculo del **estadístico de prueba** t_p**:**

$t_p =$

¿El valor del estadístico de prueba está en la región de rechazo? ____. Por lo tanto, con un nivel de significación de 5% (marca con ✓),

Se rechaza la hipótesis nula ( ) No se rechaza la hipótesis nula ( )

Viéndolo de otra manera, el intervalo de confianza para μ es

____

____

____

Es decir: (________, ________)

¿El intervalo de confianza incluye al valor de $\bar{x}$ ____? En consecuencia, ¿Se rechaza H_0? _____. ¿Se llegó a la misma conclusión? ____.

Problema 4. En el año 2005, el Consejo Nacional de Población (Conapo) informó que el promedio de vida de los mexicanos era de 73 años en los hombres, pero mayor en el caso de las mujeres. Al leer esta noticia, la licenciada Norma Alicia Reyes, encargada de registrar los decesos en el Instituto Mexicano del Seguro Social, quiso saber la vida media de las mujeres. Con ese propósito, de sus registros tomó 31 casos al azar y calculo la media de su edad al fallecer, obteniendo $\bar{x}$ = 74 años con s = 12.7. Suponiendo que la población de donde se tomó la muestra es normal, ¿cuál es el promedio de vida de las mujeres?

Llena donde haga falta.

1. Se indican los datos del problema:

2. Se plantea la hipótesis nula. (es la hipótesis que se debe)

Enuncia la hipótesis alternativa:

Hipótesis alternativa. H_1: (es la hipótesis de investigación)

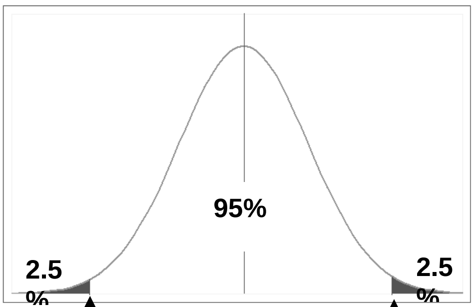
3. Se identifica la fórmula del estadístico de prueba. En este caso se está probando una hipótesis acerca de la media de una población, cuyos valores siguen una distribución y se desconoce su :

t_p = (el estadístico de prueba sigue una distribución con $n - 1$ = grados de libertad, si H_0 es verdadera)

4. Se determina el "nivel de significación", es decir, la probabilidad α de cometer el error de tipo (rechazar la hipótesis nula siendo).

Suponemos α = 0.5

5. Partiendo de que α = 0.5, se contesta la pregunta:



t = ____ t = ____

La respuesta está en el cuerpo de la tabla de valores t (Apéndice V). Para ello se considera que el valor de t a la derecha del cual está el % del área bajo la curva t de Student, es igual al valor t que tiene un área de entre este valor y -∞:

t =

Asimismo, ¿Cuál es el valor de t a la izquierda del cual está el 2.5% del área bajo la curva t de Student?, la respuesta es:

t =

La parte sombreada de la figura es la “región de ” o “región ”, la parte no sombreada es la “región de ”. A los valores que separan las regiones de rechazo y no rechazo (y) se les llama .

Enuncia la regla de decisión:

6. Se calcula el **estadístico de prueba** y se determina la decisión estadística**.**

El valor del estadístico de prueba está en la región de no rechazo, por lo tanto, con un nivel de significación de 5%, no se rechaza la hipótesis nula:

H_0:

Viéndolo de otra manera, el intervalo de confianza para μ es

Es decir: (,)

El intervalo de confianza incluye al valor de $\bar{x}$ = 74. En consecuencia, .

PRUEBA DE HIPÓTESIS A PARTIR DE UNA POBLACIÓN NORMAL CON DESVIACIÓN ESTÁNDAR CONOCIDA

Resolución por pasos

Problema 5. En una empresa de cobertura nacional se adquirieron 15 000 computadoras. Los fabricantes garantizaron que, en promedio, gastan 150 w. de energía eléctrica por hora, con desviación σ = 12.4 w. El gerente de compras de la empresa se resiste a creerlo y procede a realizar una prueba estadística para salir de dudas; para ello toma al azar una muestra de 25 computadoras y obtiene que su media del gasto de energía, por hora, es $\bar{x}$ = 156 w. ¿Mintieron los fabricantes? Averígualo mediante una prueba de hipótesis.

La solución siguiente supone que la muestra tomada por el gerente proviene de una población de datos con distribución normal, completa los pasos.

1. Datos del problema: $\mu = 150$, $n = 25$, $\bar{x} = 156$, $\sigma = 12.4$

2. Hipótesis nula. H_0: $\mu_0 = 150$ (es la hipótesis que se debe probar)

Hipótesis alternativa. H_1: _____ $\mu \neq 150$ (es la hipótesis del gerente)

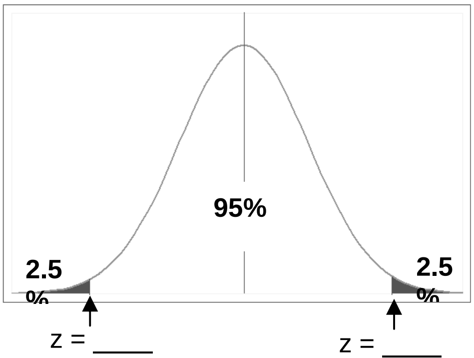
3. Fórmula del estadístico de prueba, puesto que se está probando una hipótesis acerca de la media de una población, cuyos valores siguen una distribución normal y se conoce su desviación estándar:

$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$ (el estadístico de prueba sigue una distribución normal con desviación igual a 1 y media cero)

4. Se determina el "nivel de significación", es decir, la probabilidad α de cometer el error de tipo I (rechazar la hipótesis nula siendo verdadera).

Suponemos $\alpha = 0.5$

5. Partiendo de que $\alpha = 0.5$, se contesta la pregunta ¿Cuál es el valor de z a la derecha del cual está el 2.5% del área bajo la curva normal?



La respuesta está en el cuerpo de la tabla de valores z (Apéndice IV). Para ello se considera que el valor de z a la derecha del cual está el 2.5% del área bajo la curva normal, es igual al valor z que tiene un área de 0.975 entre este valor y $-\infty$:

$$z = 1.96$$

Asimismo, ¿Cuál es el valor de z a la izquierda del cual está el 2.5% del área bajo la curva normal?, la respuesta consignada en la tabla es:

$$z = -1.96$$

La parte sombreada de la figura se conoce como "región de rechazo" o "región crítica", la parte no sombreada como "región de no rechazo". A los valores que separan las regiones de rechazo y no rechazo (-1.96 y 1.96) se les llama "valores críticos".

Regla de decisión:

Rechazar H_0 si el valor del estadístico de prueba está en la región de rechazo que es:

$-1.96 \geq z \geq 1.96$ (Se lee a partir del centro: "z es menor o igual que -1.96, y mayor o igual que 1.96")

6. Cálculo del **estadístico de prueba** y decisión estadística**.**

$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{156 - 150}{\frac{12.4}{5}}$$

$$= 2.42$$

El valor del estadístico de prueba está en la región de rechazo, por lo tanto, con un nivel de significación de 5%, se rechaza la hipótesis nula y no se rechaza la del gerente de compras:

$$H_1: \mu \neq 150$$

El intervalo de confianza para μ es

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = 156 \pm 1.96 \frac{12.4}{5}$$

$$= 156 \pm 4.86$$

$$(151.14, 160.86)$$

Dado que el intervalo no incluye a $\mu = 150$ (la cifra dada por el fabricante), se rechaza H_0, es decir, se llega a la misma conclusión.

Problema 6. Escitalopram es un agente antidepresivo que se llega a prescribir para pacientes con ansiedad y ataques de angustia (pánico). Los fabricantes afirman que la vida media de eliminación, tras dosis múltiples, es de 30 h por vía hepática (metabólica) y renal, con $\sigma = 4$ h. En un laboratorio de análisis clínicos, al estudiar a 20 pacientes tomados al azar, encontraron que la vida media de eliminación del escitalopram fue de 34 h. ¿Este resultado confirma lo que sostiene el fabricante?

La solución siguiente supone que la muestra tomada por el laboratorio proviene de una población de datos con distribución normal. Completa los pasos.

1. Datos del problema: $\mu = 30$ h, $n = 20$, $\bar{x} = 34$, $\sigma = 12$ h.

2. Hipótesis nula. H_0: $\mu = 30$ h (es la hipótesis que se debe probar)

Hipótesis alternativa. H_1: $\mu \neq 30$ (es la hipótesis del laboratorio)

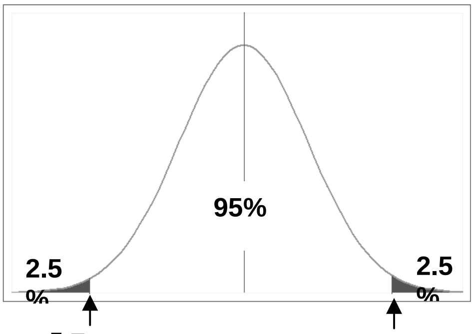
3. Fórmula del estadístico de prueba, puesto que se está probando una hipótesis acerca de la media de una población, cuyos valores siguen una distribución normal y se conoce su desviación estándar:

$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$ (el estadístico de prueba sigue una distribución normal con desviación igual a 1 y media cero)

4. Se determina el "nivel de significación", es decir, la probabilidad α de cometer el error de tipo I (rechazar la hipótesis nula siendo verdadera).

Suponemos $\alpha = 0.5$

5. Partiendo de que $\alpha = 0.5$, se contesta la pregunta ¿Cuál es el valor de z a la derecha del cual está el 2.5% del área bajo la curva? ¿Cuál curva? __________.



z = ____ z = ____

El valor de z (a la derecha del cual está el 2.5% del área bajo la curva normal), entre -∞ y él, tiene un área de _______ Su valor es:

z = _______

El valor de z, a la izquierda del cual está el 2.5% del área bajo la curva normal es:

z = ______

La parte sombreada de la figura es la ________ de __________ o región _________, la parte no sombreada es la ________ de ___ __________. A los valores que separan las regiones de rechazo y no rechazo, es decir, a ______ y ______ se les llama _________ ___________.

La regla de decisión consiste en: ____

____

6. Cálculo del **estadístico de prueba** t_p**:** $t_p =$

¿El valor del estadístico de prueba está en la región de rechazo? ____. Por lo tanto, con un nivel de significación de 5% (marca con ✓),

Se rechaza la hipótesis nula ( ) No se rechaza la hipótesis nula ( )

Viéndolo de otra manera, el intervalo de confianza para μ es

____

____

____

Es decir: (________, ________)

¿El intervalo de confianza incluye al valor de μ ____? En consecuencia, ¿Se rechaza H_0? _____. ¿Se llegó a la misma conclusión? ____.

Problema 7. El peso de las maletas de los pasajeros de un avión tiene una distribución normal con desviación estándar de 30 lb. Al tomar una muestra de las maletas de 36 pasajeros, se obtuvo que $\bar{x}$ = 160 lbs. Con un nivel de significancia de 0.05, ¿se puede concluir que el peso promedio de todos los pasajeros es menor que 170 lbs?

1. Datos del problema: _______________________.

2. Hipótesis nula. __________ (es la hipótesis que se debe _________)

Hipótesis alternativa. __________ (es la hipótesis dada)

3. Se identifica la fórmula del estadístico de prueba.

¿Los valores siguen una distribución normal? ___ ¿se conoce su desviación estándar? ____.

Forma matemática del estadístico de prueba: ——

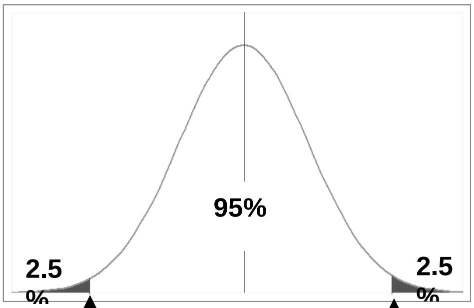
¿Qué tipo de distribución tiene el estadístico de prueba? ______________.

4. Se determina el "nivel de significación", es decir, la probabilidad ______________________ (o rechazar la hipótesis nula siendo ____________).

Supón α = 0.5

5. Partiendo de que α = 0.5, se contesta la pregunta ______________________________

____



z = ____   z = ____

Encuentra los valores críticos:

¿Cómo se denominan las partes debajo de la curva? ______________________

____

Regla de decisión: ____

____

6. Calcula el **estadístico de prueba:**

Enuncia la decisión estadística ____

____

¿El valor del estadístico de prueba está en la región de rechazo? ____, por lo tanto, con un nivel de significación de 5%, ¿se rechaza la hipótesis nula? ____.
Calcula el intervalo de confianza para μ:

¿El intervalo encontrado incluye incluye a μ_0? ____. ¿Se rechaza H_0? ____. ¿Se llega a la misma conclusión? ____.

Tabla resumen

Nombre	H_0	Prueba	H_1:≠	H_1:>	H_1:<
Media con varianza desconocida	$\mu - \mu_0$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z \leq z_{\alpha/2}$ $z \geq z_{1-\alpha/2}$	$z \geq z_{1-\alpha}$	$z \leq z_{\alpha}$
Media para varianza desconocida	$\mu - \mu_0$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t \leq t_{\alpha/2,n-1}$ $t \geq t_{1-\alpha/2,n-1}$	$t \geq t_{1-\alpha,n-1}$	$t \leq t_{\alpha,n-1}$

EJERCICIOS COMPLEMENTARIOS VI

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

1. ¿Cuál es la hipótesis de investigación de Ho: μ < 3.45, donde Ho representa la hipótesis nula y μ es la media aritmética de la población?

A. $\mu \geq 3.45$ **B.** $\mu > 3.45$ **C.** $\mu < 3.45$ **D.** $\mu \leq 3.45$ **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 H_{inv} cuando en realidad es falsa.
E. Rechazar H_{inv} 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 inocente
D. Castigar a un inocente **E.** Castigar a un culpable ( )

4. La expresión para el estadístico de prueba es:

A. $\dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ B. $\dfrac{s}{\sqrt{n}}$ C. $\dfrac{\bar{x} - \mu_0}{\sqrt{\frac{s}{n}}}$ D. $t_{n-1} \dfrac{s}{\sqrt{n}}$ E. $\bar{x} \pm t_{n-1} \dfrac{s}{\sqrt{n}}$ ( )

5. Una buena decisión es considerar:

A. H_{inv} verdadera sólo si la información contenida en la muestra da suficiente evidencia para aceptar H_0.

B. H_1 verdadera sólo si la información contenida en la muestra da suficiente evidencia para no rechazar Ho.

C. H_{inv} verdadera sólo si la información contenida en la muestra da suficiente evidencia para rechazar Ho.

D. Ho verdadera sólo si la información contenida en la muestra da suficiente evidencia para rechazar Hinv.

E. Ho verdadera sólo si la información contenida en la muestra da suficiente evidencia para no rechazar H_1.. ( )

6. ¿En qué consiste la regla de decisión? (t_p = estadístico de prueba; i_c = intervalo de confianza, $\in$ = queda contenido en ; $\notin$ = no queda contenido en; $\rightarrow$ = entonces).

A. $t_p \notin i_c \rightarrow$ no rechazamos Ho
$t_p \in i_c \rightarrow$ rechazamos Ho

B. $t_p \notin i_c \rightarrow$ rechazamos Ho
$t_p \in i_c \rightarrow$ no rechazamos Ho

C. $t_p \notin i_c \rightarrow$ rechazamos H_1
$t_p \in i_c \rightarrow$ no rechazamos H_1

D. $t_p \notin i_c \rightarrow$ no rechazamos H_1
$t_p \in i_c \rightarrow$ rechazamos H_1

E. Ninguna de las anteriores. ( )

7. ¿Qué falta en la 2ª, 3ª y 4ª columna respectivamente? ($s = 0.71$, $\mu_0 = 7.5$, $\bar{x} = 7.86$, $\alpha = .05$).

n	t_c	*Región de no rechazo de Ho* $<t_{(n-1)}, t_{(n-1)}>$	*Conclusión estadística: ¿se rechaza Ho y se acepta H_1?*
10	1.603	<-2.262, 2.262>	no
15		<-2.145, 2.145>	no
20	2.268	<-2.093, 2.093>	
25	2.535		sí
30	2.778	<-2.045, 2.045>	sí

A. 2.268, <-2.064, 2.064>, no **B.** 1.964, <-2.064, 2.064>, sí **C.** 1.964, <-2.262, 2.262>, sí

D. 2.268, <-2.064, 2.064>, sí **E.** 1.964, <-2.064, 2.064>, no ( )

8. ¿Qué conclusión verdadera se obtiene de la tabla anterior?

A. Mientras mayor es el tamaño de la muestra, menor es t_c.
B. Si la muestra es grande, se acepta la hipótesis nula.
C. Si la muestra es pequeña se rechaza la hipótesis nula.
D. A mayor tamaño de la muestra, menor tamaño de la *región de no rechazo de Ho*.
E. Al aumentar el estadístico de prueba, se acepta H_0. ( )

9. ¿Qué falta en la 2ª, 3ª y 4ª columna respectivamente? ($\mu = 7.5$, $\bar{x} = 7.86$, $n = 30$ y $\alpha = .05$).

s	t_c	*Región de no rechazo de Ho* $<t_{(n-1)}, t_{(n-1)}>$	*Conclusión estadística: ¿se rechaza Ho y no se rechaza H_1?*
.71	2.778	<-2.045, 2.045>	sí
.90	2.191		sí
1.10		<-2.045, 2.045>	
1.30	1.517	<-2.045, 2.045>	no

A. 1.517, <-2.045, 2.045>, sí **B.** 1.517, <0, 2.045>, no **C.** 1.517, <-2.045, 2.045>, no

D. 1.793, <-2.045, 2.045>, sí **E.** 1.793, <-2.045, 2.045>, no ( )

10. En una escuela de idiomas, los estudiantes habían obtenido, en un curso, una calificación promedio de 7.4. Para validar el *proyecto alfa* (un nuevo método de aprendizaje), fueron elegidos al azar 25 estudiantes, quienes obtuvieron al final del curso un promedio mensual de 7.6, con desviación estándar de 0.6. Con una confianza del 90%, ¿es válido afirmar que el promedio de calificaciones es mayor cuando se usa el *proyecto alfa*?

11. Por una encuesta aplicada a 64 empleados de una fábrica, se concluyó que el tiempo medio de la duración de un empleo en la misma es de 6.5 años con desviación típica de 4 años. Con un nivel de significación de 5%, ¿se puede aceptar que el tiempo medio de empleo en esa fábrica es menor o igual que 6 años?

SOLUCIÓN A LOS EJERCICIOS

PÁGINA 7 EJERCICIOS COMPLEMENTARIOS I

Fenómeno o Experimento		Variable Aleatoria X	Tipo de Variable Aleatoria		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		X = 0, 1, 2, 3, 4, 5
Comprar tres computadoras.		Computadora defectuosa.	Discreta		X = 0, 1, 2, 3
Las mariposas “monarca” llegarán a Michoacán el próximo año.		Número de mariposas “monarca“ que mueren en Michoacán.	Discreta		X = 0, 1, 2, 3, . . .(1)
Se mide la estatura de un adulto.		Altura en centímetros.	Continua.		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		Cualquier valor entre 0 y 100%.
Deportación a ilegales mexicanos en Estados Unidos de Norteamérica.		Número de indocumentados mexicanos que deportará Estados Unidos el próximo mes.	Discreta		X = 0, 1, 2, 3, . . .
Vacunar a 67 personas contra la Hepatitis B.		Número de casos de hepatitis B.	Discreta		X = 0, 1, 2, 3, . . . 67
Sorteo con 12 premios mayores de la Lotería Nacional si se vendieron todos los boletos.		Premios que otorgará la Lotería Nacional.	Discreta.		X = 0, 1, 2, 3, . . . 12

(1) Los puntos sucesivos al final indican una indeterminación.

2. a.

X	Ω
6	A_2, A_3, A_5
7	(A_1, A_2, A_3), (A_1, A_2, A_5), (A_1, A_3, A_5), (A_2, A_3, A_6), (A_2, A_5, A_6), (A_3, A_5, A_6)
8	(A_1, A_2, A_6), (A_1, A_3, A_6), (A_1, A_5, A_6), (A_2, A_3, A_4), (A_2, A_4, A_5), (A_3, A_4, A_5)
9	(A_1, A_2, A_4), (A_1, A_3, A_4), (A_1, A_4, A_5), (A_2, A_4, A_6), (A_3, A_4, A_6), (A_4, A_5, A_6)
10	(A_1, A_4, A_6)

b. $P(X = 6) = \left(\frac{1}{20}\right)$ $P(X = 7) = \left(\frac{6}{20}\right)$ $P(X = 8) = \left(\frac{6}{20}\right)$ $P(X = 9) = \left(\frac{6}{20}\right)$ $P(X = 10) = \left(\frac{1}{20}\right)$; **c.** $E(X) = 8$

3. $\mu = 1.91$, $\sigma = 1.167$

4.

6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	**2**	**3**	**4**	**5**	**6**

Dado que

P(X:múltiplo de 3)= $\frac{1}{3}$;

P(X:suma es 7)= $\frac{1}{6}$ y

P(X:suma no es 7 ni múltiplo de 3)= $\frac{1}{2}$,

entonces:

$\mu = E(X) = 200\frac{1}{3} + 100\frac{1}{6} - 50\frac{1}{2}$

$= 58.333$ Por lo tanto, debo esperar ganancia.

5. Ω: (aaa), (aas), (ass) y (sss).

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

$\mu = E(X) = 0(\frac{1}{4}) + 1(\frac{1}{4}) + 2(\frac{1}{4}) + 3(\frac{1}{4}) = 1.5$

6. $\mu = E(X) = 0(0.10) + 1(0.40) + 2(0.20) + 3(0.15) + 4(0.10) + 5(0.05) = 1.75$

$\sigma^2 = (-1.75)^2(\frac{1}{4}) + (-0.75)^2(\frac{1}{4}) + (0.25)^2(\frac{1}{4}) + (1.25)^2(\frac{1}{4}) + (2.25)^2(\frac{1}{4}) + (3.25)^2(\frac{1}{4}) = 5.531$

$\sigma = 2.352$

7. La ganancia puede ser pérdida de $100.00, con una probabilidad de 7998/8000, o bien ganar $199 990.00 con una probabilidad de 2/8 000:

X	P(X)	$\mu = 0.00025$
-100	7998/8000	
$199 990.00	2/8 000	

8. $\mu = 0.005(-150\,000) + 0.995(20\,000) = 19\,150$ USD

9. a)

X	0	1	2	3	4
P(X)	0.10	0.30	0.30	0.20	0.10

b) $\mu = E(X) = 0(0.10) + 1(0.30) + 2(0.30) + 3(0.20) + 4(0.10) = 1.9$

$\sigma^2 = (-1.90)^2(0.10) + (-0.90)^2(0.30) + (0.10)^2(0.30) + (1.10)^2(0.20) + (2.10)^2(0.10)$

$= 0.361 + 0.243 + 0.003 + 0.242 + 0.441 = 1.29$

$\sigma = 1.136$

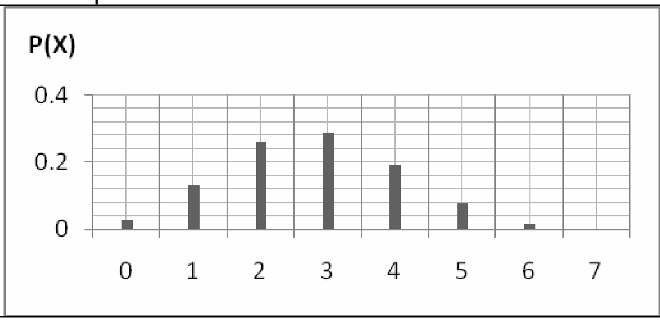
c) $P(X>2) = 0.30$ **d)** $P(X\leq3) = 0.90$

PÁGINA 17 EJERCICIOS COMPLEMENTARIOS II

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.

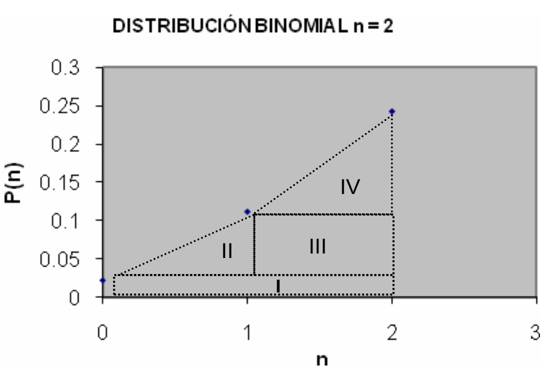
b.
$P(0) = 0.0280$
$P(1) = 0.1306$
$P(2) = 0.2613$
$P(3) = 0.2903$
$P(4) = 0.1935$
$P(5) = 0.0774$
$P(6) = 0.0172$
$P(7) = 0.0016$

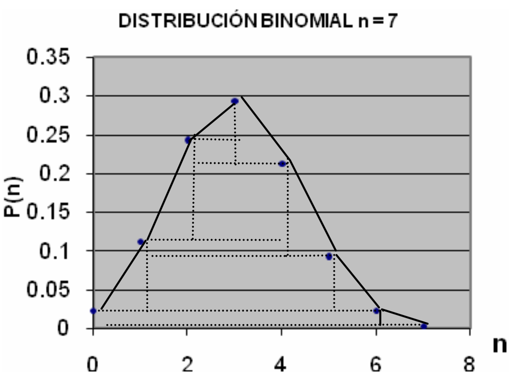


c. i. $P(x = 7) = 0.0016$ **ii.** $P(x = 0) = 0.0280$

iii. P (al menos 5 sean católicos practicantes) = 0.09812 **iv.** 0.5612

d.





Área (n = 2) = área I + área ΔII + área □III + área ΔIV

$= 2(0.02207) + \frac{1}{2}(1)(0.08985) + (1)(0.08985) + \frac{1}{2}(0.13122)$

$= 0.244525$

Área (n = 7) = 0.999965

e. $\mu = 2.8$

3. $n = 3$ y $p = 0.55$, $q = 0.45$.

a. $P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$
$= 0.1664 + 0.4084 + 0.3341$
$= 0.9089$

b. Dos

4. 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 $p = \frac{5}{8} = 0.625$. La probabilidad que necesitamos es $P(X < 2)$:

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

5. $n = 15$, $p = 0.10$, $q = 1 – p = 0.90$

a. $P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4) = 0.2059 + 0.3431 + 0.2669 + 0.1285 + 0.0429 = 0.9873$

b. $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.9873 = 0.0127$

c. $P(3 \leq X \leq 6) = P(3) + P(4) + P(5) + P(6) = 0.1285 + 0.0429 + 0.0105 + 0.0019 = 0.1838$

6. $n = 12$, $p = 0.95$, $q = 1 – p = 0.05$, $X = 9$
$P(X = 9) = 0.0196$

7. a. $n = 4$, $p = 0.80$, $q = 1 – p = 0.20$, $P(X = 2) = 0.1536$

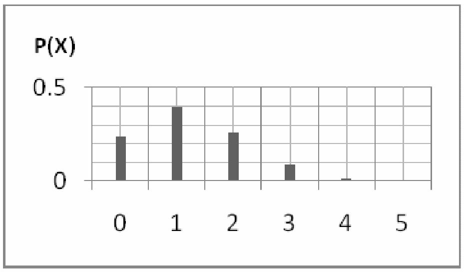
b. $P(X \geq 1) = 1 – 0.4096 = 0.5904$

8. $n = 30$, $p = 0.70$, $q = 1 – p = 0.30$, $X = 20$

$P(X = 2) = C^{30}_{20}(0.70)^{20}(0.30)^{10} = 0.14156$

9. a. $p = 0.75$, $q = 0.25$ $n = 5$, $k = 2$, $P(X = 2) = 0.2637$

b. $P(X \geq 3) = 0.0879 + 0.0146 + 0.0010 = 0.1035$

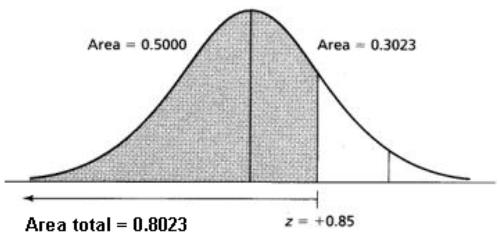


d) $\mu = 5(0.75) = 3.75$
$\sigma^2 = 5(0.75)(0.25) = 0.9375$
$\sigma = 0.968$

10. $\mu = 10\,000(0.002) = 20$
$\sigma^2 = 10\,000(0.002)(0.998) = 199$
$\sigma = 4.468$

PÁGINA 23 EJERCICIOS COMPLEMENTARIOS III

1.



2.a. 87.49%; **b.** 19.77%; **c.** 12.51%; **d.** 40.13%; **e.** 8.96%; **f.** 62.47%

3. a.- Para que las muestras sean representativas, debe seleccionarlas al azar. La muestra debe ser lo más grande posible o al menos mayor que 30.

b. - No se puede saber cuál de las muestras es el mejor estimador.

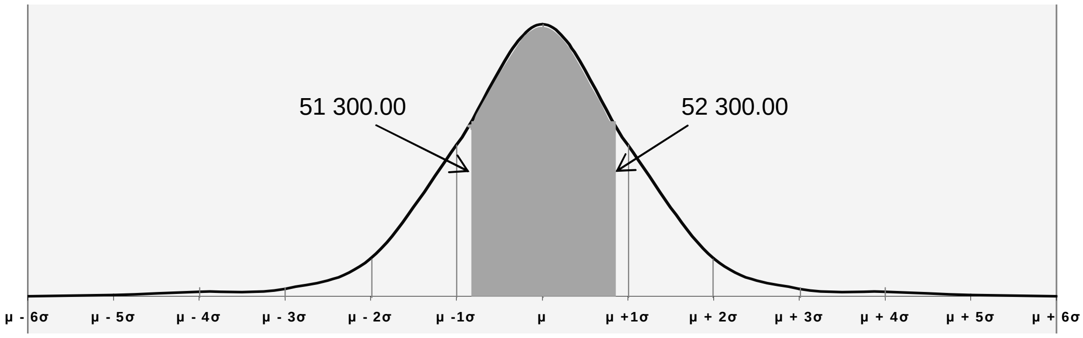
c. - La distribución de las medias muestrales tendrá una forma normal.

d. - La media de las medias muestrales sí es un buen estimador de la media poblacional.

e. - La desviación estándar muestral no es buen estimador de la desviación estándar poblacional.

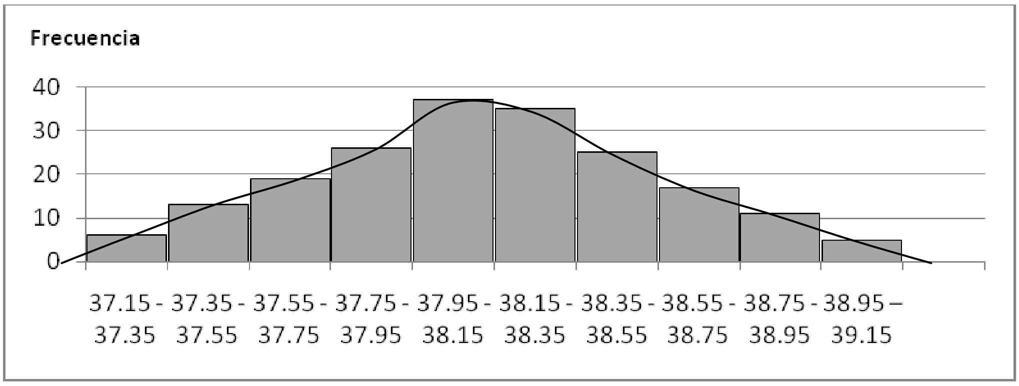
4. Se pregunta sobre la probabilidad de que la media de la muestra esté entre

$51 300.00 y $52 300.00, es decir, que esté dentro del área sombreada de la figura:



Como μ = $51 800.00 y σ = $730.30, entonces $z = \frac{51300 - 51800}{730.30}$ = -0.68 y puesto que la distribución muestral es normal, al consultar la tabla de la distribución normal estándar, se obtiene que el área entre z = 0 y z = -0.68 es 0.2518. Por lo tanto la probabilidad de que el valor de la media caiga entre los valores $51 300.00 y $52 300 es 0.2518 + 0.2518 = 0.5036.

5.



b. μ = 38.1304, σ = 0.4252 **i.** 0.80785; **ii.** 0.87235; **iii.** 0.19215

6. Es aproximadamente simétrica o gaussiana, los individuos en la zona de riesgo alto son la mayoría.

7. a. 30.68%; **b.** 69.59%; **c.** $4546.78 + $6875.49 = $11422.27

8. Denótese con x a la calificación lograda en la prueba de cualquier estudiante; x tiene una **media de 500 y una desviación típica de 100. Aplicando la ecuación:**

$$z = \frac{x - \mu}{\sigma} = \frac{700 - 500}{100} = 2.00$$

Utilizando la gráfica de la curva normal cola izquierda (Apéndice III) se obtiene que para una z = 2.00 el área bajo la curva es 0.97725, entonces, la probabilidad buscada es 0.02275.

9. a. La proporción de automóviles compactos que consiguen 30 mpg o más, está dada por el área sombreada en la figura de la tabla normal estándar (cola derecha). Para resolver el problema debes comenzar por encontrar el valor de z que corresponde a x = 30. Al sustituir en la fórmula para z, se obtiene:

$$z = \frac{x - \mu}{\sigma} = \frac{30 - 25.50}{4.50} = 1.00$$

El área determinada en la tabla que corresponde con un valor de z = 1.00 es 1 – 0.84134 = 0.15866, por lo tanto, el porcentaje que excede 30 mpg o más es 15.87%.

b. La proporción en el uso de gasolina x tiene una distribución normal con una media de 25.50 mpg y una desviación estándar de 4.50 mpg. Se necesita encontrar un valor particular, x_o, tal que $P(x \leq x_o) = 0.95$.

Este es el 95º percentil de la distribución de la proporción en el uso de gasolina x. Puesto que la única información disponible acerca de las probabilidades normales está en términos de la variable aleatoria normal estándar z, se debe estandarizar el valor de x_o:

$$z_0 = \frac{x_0 - \mu}{\sigma} = \frac{x_0 - 25.50}{4.50}$$

como el valor de z_o corresponde a x_o, también debe tener un área 0.95 a su izquierda, lo cual significa que el área entre $z = -\alpha$ y $z = z_o$ debe ser 0.95. Por tanto, el área remanente es
1.00 – 0.95 = 0.05. Al localizar dicho valor en la tabla normal estándar se tiene z_o = 1.6

Al despejar x_o se obtiene: $x_o = \mu + \sigma z_o$; x_o = 25.50 + (4.50)(1.645) = 32.90

El nuevo automóvil compacto del fabricante debe lograr entonces 32.90 mpg para superar el desempeño del 95% de los automóviles compactos que se venden en la actualidad en Estados Unidos.

10. a. 50%; **b.** P(z_1 = -0.333) = 0.3707, P(z_2 = 0.333) = 0.62552, P(0.95 < x < 1.05) = 0 0.25482;

c. P(x < 0.80) 0.09176

11.

RESUMEN

En una **muestra,** a la media aritmética y a la desviación estándar se les denomina **estimadores estadísticos**. En una **población**, la media se representa por μ y la desviación estándar por σ y se denominan **parámetros.**

Una variable aleatoria continua *X* sigue una **distribución normal,** de parámetros μ y σ, si su función de densidad es:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x\in R$$

Al cambiar la variable *x* por $z=\frac{x-\mu}{\sigma}$ la función se transforma en:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \quad donde\, \mu=0,\ \sigma=1$$

*La distribución normal N(*μ, σ) y tiene las siguientes propiedades:

- Es **simétrica** con respecto a su media aritmética.
- Se extiende desde –∞ hasta +∞.
- Es **asintótica**: al eje x, nunca lo toca por mucho que se extienda.
- La media aritmética de la población coincide con la esperanza matemática de la variable aleatoria X y σ^2 es la varianza.
- Su **posición** depende del valor de la media aritmética.
- Su **forma** depende de la desviación estándar.
- El valor de la media coincide con el valor de la moda y con el valor de la mediana.
- El **área bajo la campana representa un valor de la probabilidad** P(*a*< X <*b).*

Hallar una **probabilidad** implica calcular un **área bajo la curva normal** estandarizada. Para ello se utilizan tablas y conviene seguir estos 3 pasos:

Paso I. **Estandarizar** los valores *x* de interés, mediante $z=\frac{x-\mu}{\sigma}$.

Paso II. **Hacer** un esbozo de las curvas normales, tanto de la gaussiana como de la estandarizada, marcar ahí los valores de *x* y de *z*; además, sombrear el área de interés en la normal estandarizada.

Paso III. **Encontrar el área de interés** mediante una tabla a partir de los valores *z* obtenidos.

12. a. P(x < 40) = 0.1056; **b.** P(x < 50) = 0.5; **c.** 0.8944

PÁGINA 29 EJERCICIOS COMPLEMENTARIOS IV

1. **a.** μ = $3.00; **b.** σ = $\sqrt{2}$ = 1.4142: **c.**

Muestra	1,1	1,2	1,3	1,4	1,5	2,1	2,2	2,3	2,4
$\bar{x}_i$	($\bar{x}$ = 1)	($\bar{x}$ = 1.5)	($\bar{x}$ = 2)	($\bar{x}$ = 2.5)	($\bar{x}$ = 3)	($\bar{x}$ = 1.5)	($\bar{x}$ = 2)	($\bar{x}$ = 2.5)	($\bar{x}$ = 3)

Muestra	2,5	3,1	3,2	3,3	3,4	3,5	4,1	4,2	4,3
$\bar{x}_i$	($\bar{x}$ = 3.5)	($\bar{x}$ = 2)	($\bar{x}$ = 2.5)	($\bar{x}$ = 3)	($\bar{x}$ = 3.5)	($\bar{x}$ = 4)	($\bar{x}$ = 2.5)	($\bar{x}$ = 3)	($\bar{x}$ = 3.5)

Muestra	4,4	4,5	5,1	5,2	5,3	5,4	5,5
$\bar{x}_i$	($\bar{x}$ = 4)	($\bar{x}$ = 3.5)	($\bar{x}$ = 3)	($\bar{x}$ = 3.5)	($\bar{x}$ = 4)	($\bar{x}$ = 4.5)	($\bar{x}$ = 5)

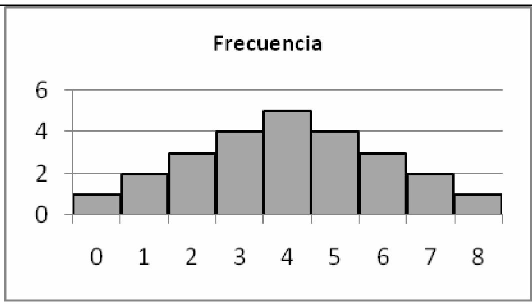
25 muestras.

d.

$\bar{x}$	Frecuencia
0	**1**
1	**2**
2	**3**
3	**4**
4	**5**
5	**4**
6	**3**
7	**2**
8	**1**

e. Acampanada o gausiana.

f. (Derecha)



g. $\mu_{\bar{x}} = 3 \quad \sigma_{\bar{x}} = 1$; **h.** No coinciden ; **i.** $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{2}} = 1$; **j.** Sí coinciden.

2. a. Para que las muestras sean representativas, debe seleccionarlas al azar. La muestra debe ser lo más grande posible o al menos mayor que 30.
b. No se puede saber cuál de las muestras es el mejor estimador.
c. La distribución de las medias muestrales tendrá una forma normal.
d. La media de las medias muestrales sí es un buen estimador de la media poblacional.
e. La desviación estándar de las medias muestrales no es un buen estimador de la desviación estándar de la población.

SOLUCIÓN A LOS EJERCICIOS V.1

1. ✓, ✓, ✓, ×, ×. **2.** P(3.404 min. < μ < 3.796 min.) **3.** (6.7575 m, 7.842 m) **4.** 68.31 Km/h < v < 81.29 Km/h

5. a. $\frac{n}{N} = \frac{100}{1000} = 10\%$. Por tanto, se debe utilizar el factor finito de corrección. Los límites del intervalo de confianza se obtienen de $\bar{x} \pm z \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$, al sustituir los datos del problema.

$$15.0 \pm 1.96 \frac{2.0}{\sqrt{100}} \sqrt{\frac{1000-100}{1000-1}} = 15.0 \pm 1.96(0.2)\sqrt{0.901} = 15.0 \pm 0.372$$

b. $\frac{n}{N} = \frac{16}{200} = 8\%$. Cuando se desconoce σ_x y $n \le 30$, la distribución ***t*** es apropiada (suponiendo que la población que se muestra sea normal), y la fórmula para el intervalo de confianza es $\bar{x} \pm t \frac{S_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$. Con los datos del problema se obtienen (15 grados de libertad):

$$15.0 \pm 2.131 \frac{2.0}{\sqrt{16}} \sqrt{\frac{200-16}{200-1}} = 15.0 \pm 2.131(0.5)\sqrt{0.925} = 15.0 \pm 1.025$$

PÁGINA 42 EJERCICIOS V-2

1. P(0.618 < X < 0.861) Con una confianza del 95%, el porcentaje de personas que les gusta leer novelas en Coyoacán está entre 61.8% y 86.1%. Es muy conveniente.
2. a. P(0.1819 < μ < 0.418) = 0.99. Con una confianza de 99%, en realidad el porcentaje de médicos que utilizan o recomiendan el producto es, cuando mucho, 41.8%.
b. Con esa misma confianza se puede afirmar que el anuncio publicitario miente.
3. P(0.1324 < p < 0.4076) = 0.95. Con una confiabilidad de 95%, el porcentaje de los que no dedican la menos 15 min. diarios para conversar con sus hijos está entre 13.24% y 40.76%.
4. No se cumple que np > 5, en consecuencia no se puede aplicar el método.
5. a. Con una confianza de 99.9%, la fracción votante que apoyará a Ceferino estará entre 0.432 y 0.748, por lo tanto no sería aconsejable que se comenzara a preparar para tomar posesión. **b.** Con una confianza de 50%, la fracción votante que apoyará a Ceferino estará entre 0.558 y 0.622, por lo tanto sería aconsejable que se comenzara a preparar para la toma de posesión.

PÁGINA 58 EJERCICIOS COMPLEMENTARIOS VI

1. A; **2.** B; **3.** B; **4.** A; **5.** C; **6.** B; **7.** B; **8.** D; **9.** E
10. $t_c = 1.667$. Región de no rechazo: (-∞, 1.711). No se rechaza H_0. Como no se rechaza H_0: μ ≤ 7.4, con una confianza de 95% no hay evidencia para considerar que, con el *proyecto alfa*, el promedio de calificaciones sea mayor que el que se obtiene sin aplicarlo.

11. μ = 6, s = 4, n = 64; H_0: μ ≤ 6; H_1 : μ > 6; región de no rechazo: (-∞, 6.8225); región de rechazo: (6.8225, ∞); resulta que 6.5 está en la región de no rechazo, se acepta la hipótesis nula con 95% de confianza.