Unidad 6

Distribuciones de probabilidad continua, muestreo y distribución de muestras

Introducción

La unidad 5 se enfocó en el estudio de las distribuciones de probabilidad discreta, entre las cuales se abordaron las siguientes: la binomial, la hipergeométrica, la de Poisson y la geométrica. Estas distribuciones tienen la peculiaridad de basarse en las variables aleatorias discretas, a las que se definió como *aquellas que toman valores expresados con números enteros*.

Una vez que analizamos las distribuciones de probabilidad discretas, deberemos estudiar las distribuciones de probabilidad continuas, las cuales tienen importantes aplicaciones en distintos campos de estudio como las áreas administrativas, financieras y de producción, ya que, a diferencia de las distribuciones discretas, en las continuas se consideran todos los valores existentes en un conjunto de datos. Por ejemplo, al estudiar el tiempo que una persona permanece en un banco o el que tarda para producir un artículo, los valores que toma la variable tiempo, no son expresados únicamente en números enteros, ya que el proceso puede tardar 2.5 minutos o más.

Otro aspecto importante es el tipo de muestra con la que se trabaja, ya que las distribuciones de probabilidad continua se caracterizan por emplear muestras grandes ($n \geq 30$), aunque también se da el caso en el que se trabaja con muestras pequeñas, como la distribución *t-student* (que veremos posteriormente). A pesar de que la distribución *t* emplea muestras pequeñas, se considera dentro de las distribuciones continuas porque los valores que toma la variable aleatoria se encuentran dentro de un intervalo.

Por otro lado, es importante mencionar que también se estudian algunas distribuciones muestrales para lo cual es necesario obtener muestras que sean representativas, lo que permitirá conducir algunos aspectos acerca de la población en estudio.

Existen tres métodos para lograr que una muestra sea representatíva: muestreo aleatorio, muestreo estratificado y muestreo sistemático, los cuales se abordarán en esta unidad.

6.1. Modelos de probabilidad para variables aleatorias continuas

Una distribución de probabilidad continua se asocia con las variables aleatorias continuas, las cuales comprenden medidas como peso, estatura, distancia y una de las más importantes, el tiempo; su definición es:

Una variable aleatoria continua es aquella que toma cualquier valor dentro de un conjunto de datos, es decir, no sólo toma valores enteros sino también valores en fracciones o de cualquier otro tipo distinto a los enteros y puede asumir un número infinito de valores.

Algunos ejemplos de variables aleatorias continuas son:

- El tiempo que tarda el servicio en un restaurante de comida rápida.
- La cantidad de mililitros de refresco que contiene cada botella.
- El peso de las mercancías que va a transportar una empresa de autotransportes.
- El tiempo que tarda un empleado en producir 100 artículos.
- La estatura promedio de los habitantes de una población.
- El riesgo diario en que incurren algunos inversionistas en la bolsa.
- El precio de cambio de una divisa.

Si quisiéramos ver las aplicaciones de algunos de los ejemplos anteriores, tenemos:

- Nos interesa medir la cantidad de refresco que contienen las botellas producidas por una empresa con el fin de conocer la probabilidad de que alguna de ellas contenga una menor cantidad que la especificada y así el departamento de producción no reciba reclamaciones por inconformidad de los clientes.
- Es importante conocer el peso de mercancías transportadas para así estimar la probabilidad de que los paquetes transportados no excedan el peso especificado y, de esa manera, reducir el riesgo de accidentes e infracciones.
- Es importante el tiempo que tarda un empleado en producir cierta cantidad de artículos para conocer su productividad y así determinar la probabilidad de que el empleado incremente su productividad mediante los estímulos necesarios.
- Es de interés conocer la estatura promedio de una población, si una compañía de prendas de vestir está interesada en instalarse en esa población y de esa manera conocer la probabilidad de que las prendas que fabrica cumplan con las necesidades de la población.
- Queremos conocer el riesgo que implica invertir en la bolsa para determinar la probabilidad de que haya un riesgo muy grande al elegir algunas acciones.
- Queremos conocer el historial del precio de cambio de una divisa para observar la probabilidad de ganancia si se invierte en esa divisa.

Al emplear las distribuciones de probabilidad continua, regularmente se utilizan modelos matemáticos con fórmulas definidas para cada distribución, de tal forma que se puedan obtener resultados únicos en cada una.

A continuación se tratará con detalle la distribución normal, parte elemental para la construcción de modelos de variables aleatorias continuas.

Ejercicio 1

1. Una variable aleatoria continua adopta valores:

   a) No solamente específicos, sino cualquiera que esté dentro de un intervalo.
   b) Únicamente específicos, es decir, números enteros.
   c) Definidos dentro de un intervalo, excepto aquellos que sean fracciones.
   d) Únicamente los valores que se encuentren expresados en fracciones.

2. Algunos ejemplos de variables aleatorias continuas son:

   a) Tiempo y pantalones producidos por una empresa.
   b) Coches y estatura.
   c) Cuatrimestres estudiados y programa de estudios.
   d) Distancia y volumen.

6.2. Distribución de probabilidad normal

En gran parte de los casos prácticos se trabaja con muestras grandes donde existe un elevado número de datos, por lo que se hace necesario el uso de técnicas adecuadas que permitan tratar datos cuyos valores no sean únicamente números enteros. La distribución de probabilidad normal resulta ser un instrumento adecuado para efectuar mediciones de interés porque no sólo trabaja con muestras, sino principalmente con poblaciones.

Por esta razón la distribución continua de probabilidad más relevante en todo el campo de la estadística es la distribución normal.

La distribución normal es importante por dos aspectos: en primer lugar, tiene algunas propiedades que la hacen aplicable a un gran número de situaciones en donde se toman muestras grandes, en segundo lugar, se ajusta a distribuciones de frecuencias observadas en muchos fenómenos, incluyendo áreas como administración y finanzas.

La distribución normal se puede representar a través de una gráfica que tiene forma acampanada y recibe el nombre de curva normal (véase la figura 6.1). La curva normal depende de dos parámetros, de la media μ y de la desviación estándar σ. La media señala la parte central de la distribución y es ahí donde se espera esté la mayor parte de los datos con el fin de que no exista una gran dispersión entre ellos. La varianza y la desviación estándar son importantes debido a que indican si existe alguna dispersión entre los datos y de qué magnitud es tal dispersión en caso de que se presente.

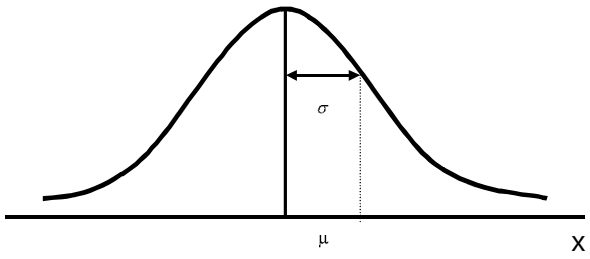


Figura 6.1. La curva normal.

Características de la distribución normal

La curva normal posee algunas características especiales:

1. La curva tiene forma de campana y presenta un punto máximo que se encuentra en el centro de la distribución, en ese punto la media, la mediana y la moda son iguales.

2. Es simétrica con respecto a la media de la distribución, es decir, el índice de asimetría de la distribución normal es cero.

3. La distribución normal no es ni tan puntiaguda ni tan plana, es decir, es una distribución mesocúrtica, por lo que su índice de kurtosis es igual a 3.

4. La curva normal se extiende horizontalmente de menos infinito a más infinito ($-\infty$ a $+\infty$). Esto quiere decir que pueden existir valores positivos extremadamente altos (a la derecha de la parte central) o valores negativos (a la izquierda de la parte central).

5. El área total bajo la curva normal se considera que es de 100%, ya que la suma de las probabilidades a lo largo de la distribución es uno.

6. Cada distribución normal está completamente especificada por su media y su desviación estándar, existiendo una distribución normal diferente para cada combinación de media y de desviación estándar, dependiendo del grado de dispersión que exista.

Las características mencionadas anteriormente se pueden presentar a través de las figuras 6.2. y 6.3.

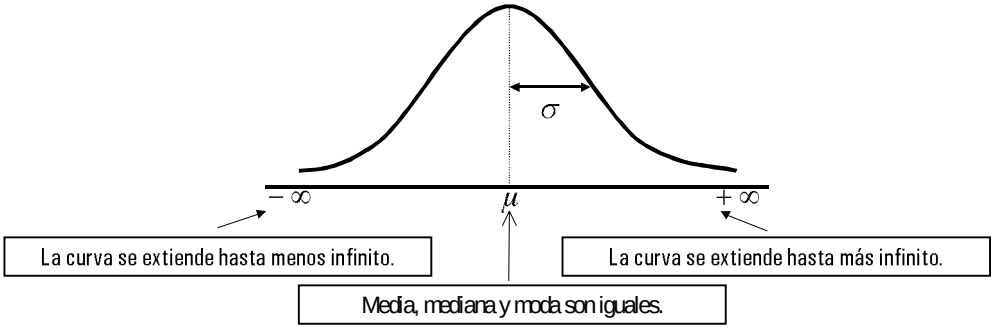


Figura 6.2. Características de la curva normal.

En la figura 6.3. se puede apreciar que la forma de las curvas normales es diferente cuando las medias y las desviaciones estándar no coinciden, en este caso $\mu_2 > \mu_1$ y $\sigma_1 < \sigma_2$.

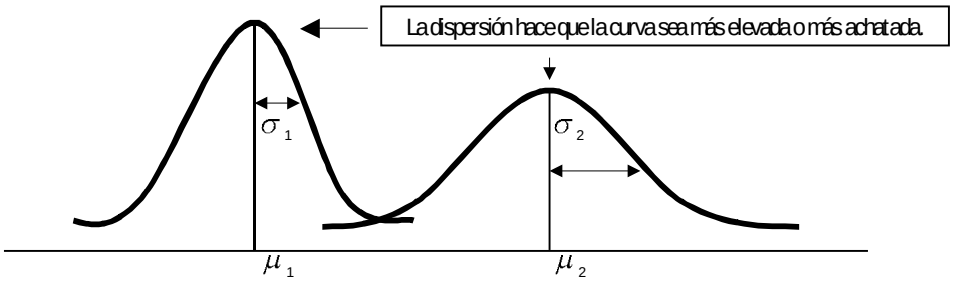


Figura 6.3. Curvas normales cuando las medias y las desviaciones estándar son diferentes.

Áreas bajo la curva normal

Sin importar cuáles sean los valores de la media y de la desviación estándar para una distribución de probabilidad normal, el área bajo la curva tiene un valor de 1, de manera que se puede pensar en áreas bajo la curva como si fueran probabilidades, donde en cada mitad de la distribución la suma de probabilidades es 0.5 y cualquier valor que se encuentre en la distribución tiene una probabilidad de ocurrencia. La forma más sencilla de plantear esto es trazar áreas limitadas por desviaciones estándar con respecto de la media.

Con base en la regla empírica de la teoría estadística se puede plantear tres aseveraciones bajo la curva normal.

1. Se sabe que aproximadamente 68% del área bajo la curva está comprendida en un intervalo de $\mu \pm \sigma$, lo anterior indica que los valores de la distribución normal se encuentran en un rango que va desde $\mu - \sigma$ hasta $\mu + \sigma$ ($-\sigma < \mu < +\sigma$), es decir, 68% de los datos se encuentra a una distancia de una desviación estándar a la derecha y una desviación estándar a la izquierda del valor de la media.
2. Aproximadamente 95% de todos los valores de una población normalmente distribuida se encuentra en un rango que comprende $\mu \pm 2\sigma$, lo anterior es indicativo de que los valores de la distribución normal están en un intervalo que va de $\mu - 2\sigma$ a $\mu + 2\sigma$ ($-2\sigma < \mu < +2\sigma$), es decir, 95% de los datos se encuentra a una distancia de dos desviaciones estándar a la derecha y dos desviaciones estándar a la izquierda del valor de la media.
3. Aproximadamente 99% de todos los valores de una población normalmente distribuida se encuentra en un rango que va de $\mu \pm 3\sigma$, lo anterior quiere decir que la distribución normal determina un intervalo comprendido de $\mu - 3\sigma$ a $\mu + 3\sigma$ ($-3\sigma < \mu < +3\sigma$), es decir, 99% de los datos se encuentra a una distancia de tres desviaciones estándar a la derecha y tres desviaciones estándar a la izquierda del valor de la media.

Las áreas bajo la curva normal basadas en las tres aseveraciones mencionadas pueden representarse a través de la figura 6.4.

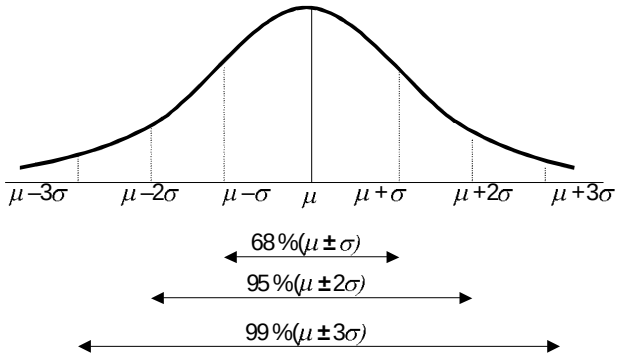


Figura 6.4. Áreas bajo la curva normal.

Ejemplo 1

El sueldo mensual que reciben los empleados de una empresa dedicada a la producción de plástico, sigue una distribución normal con una media de $8 000 y una desviación estándar de $700. La empresa desea conocer:

a) El rango de valores entre los que se encuentra aproximadamente 68% de los sueldos de los empleados.

b) El rango de valores entre los que se encuentra aproximadamente 95% de los sueldos los empleados.

c) El rango de valores entre los que se encuentra aproximadamente 99% de los sueldos de los empleados.

Solución

a) Como se mencionó anteriormente, 68% de los datos se encuentra a una desviación estándar a la derecha y a una a la izquierda con respecto al valor de la media.

Donde:
μ = \$8 000
σ = \$700

La distancia entre la media y la desviación estándar es $\mu \pm \sigma$
Al sustituir $\mu \pm \sigma$ en se obtiene:

$\mu - \sigma$ = 8 000 – 700 = 7 300
$\mu + \sigma$ = 8 000 + 700 = 8 700

Por lo tanto, 68% de los sueldos se encuentra en un intervalo entre \$7 300 y \$8 700. Ello quiere decir que en promedio, de 68% de los sueldos, el sueldo mínimo que podrán recibir los empleados es de \$7 300 y el salario máximo que pueden recibir es de \$8 700.

b) De la distribución 95% de probabilidad normal se encuentra 2 desviaciones estándar a la derecha y dos desviaciones estándar a la izquierda con respecto al valor de la media, con base en esto se puede proceder de la siguiente manera:

$\mu \pm 2\sigma$
2 (σ) = 2 (700)
2 (σ) = 1 400

Sustituimos:

$\mu - 2\sigma$ = 8 000 – 1 400 = 6 600
$\mu + 2\sigma$ = 8 000 + 1 400 = 9 400

De esta manera, 95% de los sueldos de los empleados se encuentra en el intervalo entre \$6 600 y \$9 400, por lo que al considerar 95% de los sueldos, el sueldo mínimo que podrían percibir los empleados es de \$6 600 y el sueldo máximo que podrían percibir es de \$9 400.

c) El 99% de la distribución de probabilidad normal se encuentra 3 desviaciones estándar a la derecha y 3 desviaciones estándar a la izquierda con respecto al valor de la media, con base en lo anterior se puede proceder de la siguiente manera:

$\mu \pm 3\sigma$
3 (σ) = 3 (700)
3 (σ) = 2 100

Se sustituye:

$\mu - 3\sigma$ = 8 000 – 2 100 = 5 900
$\mu + 3\sigma$ = 8 000 + 2 100 = 10 100

Por lo tanto, 99% de los sueldos de los empleados se encuentra en el intervalo entre $5 900 y $10 100. Al considerar 99% de los sueldos, los empleados recibirán como mínimo un sueldo de $5 900 y como máximo uno de $10 100.

Los resultados también se pueden presentar mediante una gráfica (véase la figura 6.5.)

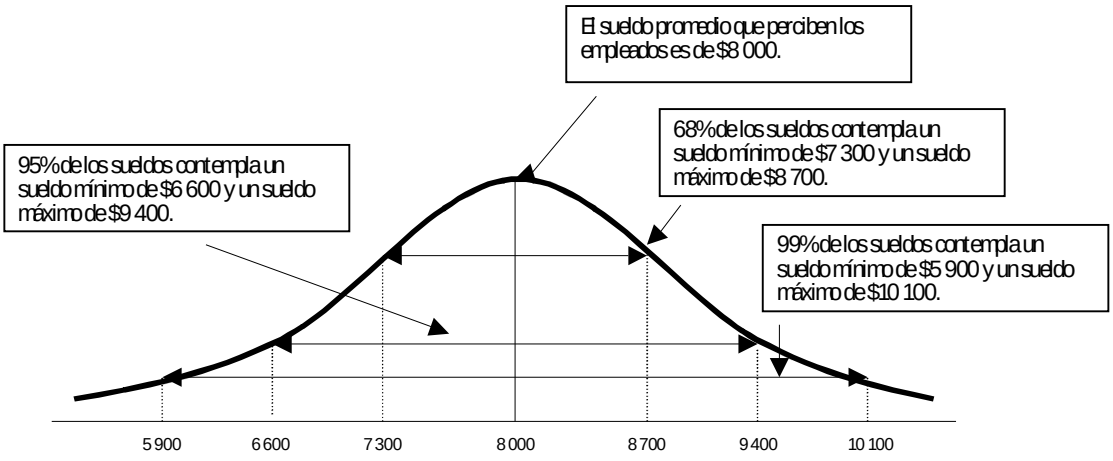


Figura. 6.5. Áreas bajo la curva normal.

Distribución de probabilidad normal estandarizada

Como se explicó en los apartados anteriores, al analizar fenómenos reales en los distintos campos de estudio, existe un número infinito de distribuciones normales, donde cada una de estas distribuciones tiene una media y una desviación estándar distinta. En términos operativos resultaría demasiado complicado proporcionar resultados para cada combinación de media y desviación estándar. El problema de expresar un número infinito de distribuciones normales se aminora si se hace uso de una estandarización de los datos que tenga la variable aleatoria X, donde se hace uso de una regla de transformación que cambia sus valores X por otros transformados que se llaman Z (la variable estandarizada) y sin que importen las distintas combinaciones de los parámetros μ, σ que posean los datos. Éstos se pueden convertir a una escala estandarizada aplicando la siguiente fórmula:

$$Z = \frac{X - \mu}{\sigma}$$

Donde:

X = Es el valor de la variable aleatoria en estudio.

μ = Es el valor de la media de la distribución de la variable aleatoria.

σ = Es la desviación estándar de la distribución.

La fórmula expresada en términos de Z *mide la distancia entre el valor específico de* **X** *y el valor de la media, utilizando la desviación estándar como unidad de medida.*

De esta manera, se puede convertir cualquier variable aleatoria normal X en una variable aleatoria normal estandarizada Z. Regularmente los datos de la variable aleatoria tienen una media y una desviación estándar definida, sin embargo, la variable aleatoria estandarizada Z siempre tiene una media igual a cero ($\mu = 0$) y una desviación estándar igual a uno ($\sigma = 1$). Por lo tanto, ya que partimos del hecho de que los datos estandarizados siempre tendrán valores de $\mu = 0$ y $\sigma = 1$, sólo es necesario generar y después tabular una distribución.

Como trabajamos con una forma estandarizada, debemos buscar la probabilidad de ocurrencia de un valor Z en tablas, y para ello es necesario considerar que la tabla únicamente muestra dos decimales, es decir, se observa un valor de Z que contiene hasta centésimos. La columna de la izquierda indica los valores enteros de Z con un decimal y el primer renglón muestra las centésimas. Por ejemplo, si se quiere encontrar la probabilidad de ocurrencia de algún valor que se encuentre entre la media de la distribución normal y el valor estandarizado Z = 0.25, en el extremo izquierdo de la tabla se busca 0.2 y en la parte superior 5; el punto donde se intersectan esos valores será la probabilidad de ocurrencia de Z, que en este caso es 0.0987 y lo podemos apreciar en la tabla siguiente (ver anexo 1):

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879

Tabla 6.1. Segmento de la tabla de valores de la distribución normal estándar.

Ejemplo 2

Una compañía productora de llantas realiza un estudio sobre el tiempo de vida útil de las llantas, del estudio resulta que las llantas tienen una duración promedio de 35 000 kilómetros y una desviación estándar de 4 000 kilómetros. El gerente de la empresa está interesado en saber:

a) ¿Qué probabilidad existe de que las llantas tengan un tiempo de vida superior a 38 000 kilómetros?
b) ¿Qué proporción de estas llantas tiene un tiempo de vida inferior a 32 000 kilómetros?
c) ¿Qué proporción de estas llantas tiene un tiempo de vida entre 32 000 y 38 000 kilómetros?

Solución:

a) El objetivo en este punto es conocer cuál es la probabilidad de que las llantas tengan una duración superior a 38 000 km.
Los datos son:

μ = 35 000
σ = 4 000
X = 38 000
P(X > 38 000)

El objetivo es encontrar la probabilidad de que X sea superior a 38 000 km.
Sustituyendo en la fórmula se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{38\,000 - 35\,000}{4\,000} = \frac{3\,000}{4\,000} = 0.75$$

El valor en las tablas para Z = 0.75 es de 0.2734

Se tiene que el valor de tablas para la zona que va de 0 a 0.75 es de 0.2734 (véase la figura 6.6). Como ya habíamos argumentado, el valor para cada una de las mitades de la curva normal es de 0.5. Como el problema radica en encontrar la probabilidad de que las llantas tengan una vida útil superior a 38 000 kilómetros, ello equivale a decir que interesa conocer la probabilidad de que en términos estandarizados Z sea superior a 0.75, por lo que con la mitad izquierda de la distribución se tiene 0.5 y hay que restarle al 0.5 de la mitad derecha el valor de tablas de 0.2734, con lo cual se encuentra la probabilidad que nos interesa, es decir, el área localizada en el extremo derecho, por lo tanto:

$P(X > 38\,000) = P(Z > 0.75) = 0.5 - 0.2734 = 0.2266$

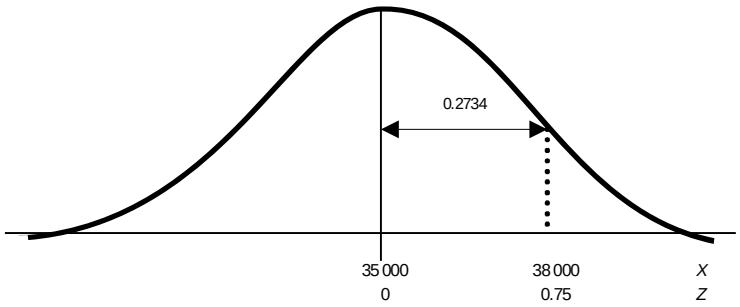


Figura 6.6. Distribución normal de la vida útil de llantas.

En conclusión, la probabilidad de que las llantas tengan una duración superior a los 38 000 km es 0.2266 o de 22.66%.

b) En este punto se busca que las llantas tengan un tiempo de duración inferior a 32 000 km.

Los datos son:

μ = 35 000
σ = 4 000
X = 32 000
$P(X < 32\,000)$

Al sustituir en la fórmula se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{32\,000 - 35\,000}{4\,000} = -\frac{3\,000}{4\,000} = -0.75$$

El valor en las tablas para Z = 0.75 es de 0.2734

Se tiene que el valor de tablas para la zona que va de 0 a –0.75 es de 0.2734, por lo que interesa el área ubicada a la izquierda de –0.75. Para encontrar la probabilidad que queremos conocer, es

necesario restar a 0.5 el valor 0.2734, con el fin de garantizar que el valor que se va a encontrar es el que asegura las condiciones del problema (véase la figura 6.7.), por tanto:

$P(X < 32\,000) = P(Z < -0.75) = 0.5 - 0.2734 = 0.2266$

Figura 6.7. Distribución normal de la vida útil de llantas.

En conclusión, la probabilidad de que las llantas tengan una duración inferior a 32 000 km es 0.2266 o de 22.66%.

c) En este punto se busca la probabilidad de que las llantas tengan una vida útil entre 32 000 y 38 000 km.
Los datos son:

$\mu = 35\,000$
$\sigma = 4\,000$
$X_1 = 32\,000$
$X_2 = 38\,000$
$P(32\,000 < X < 38\,000)$

$$Z_1 = \frac{X - \mu}{\sigma} = \frac{32\,000 - 35\,000}{4\,000} = -\frac{3\,000}{4\,000} = -0.75$$

$$Z_2 = \frac{X - \mu}{\sigma} = \frac{38\,000 - 35\,000}{4\,000} = \frac{3\,000}{4\,000} = 0.75$$

Lo que se busca es la probabilidad de que la vida útil de las llantas esté entre 32 000 y 38 000 kilómetros, por lo que es necesario conocer la probabilidad de que Z esté entre – 0.75 y 0.75, para lo cual es necesario sumar al valor de tablas del extremo izquierdo, el valor de tablas (0.2734) del extremo derecho. En este caso no es necesario restar el valor de tablas al 0.5 que le corresponde a cada extremo de la distribución, debido a que, como ya hemos explicado, el valor de tablas es un valor acumulativo que comprende la distancia desde que Z vale 0, hasta que Z toma un valor de 0.75, teniéndose:

$P(32\,000 < X < 38\,000) = P(-0.75 < Z < 0.75) = (0.2734) + (0.2734) = 0.5468$

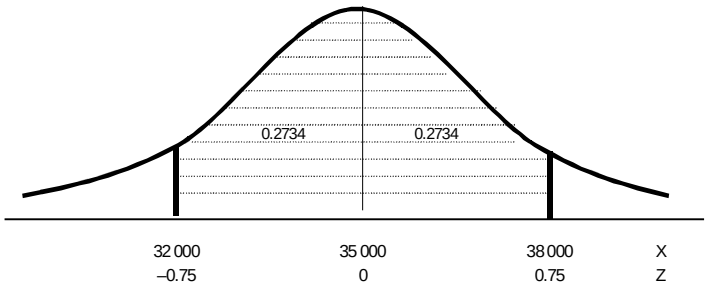


Figura 6.8. Distribución normal de la vida útil de llantas.

Por lo tanto, la probabilidad de que las llantas tengan una duración entre 32 000 y 38 000 km es de (0.2734) (2) = 0.5468 o el 54.68% (véase la figura 6.8.).

Ejemplo 3

Una empresa de mercadotecnia logró firmar un contrato con un importante grupo financiero; sin embargo, esto ha implicado una serie de nuevos costos. Se estimó que el costo promedio de la empresa es de $50 000 y una desviación estándar de $5 000. La empresa requiere saber cuál es la probabilidad de que el costo de ejecutar el contrato se encuentre entre $46 000 y $54 000.

Los datos son:

μ = $50 000
σ = $5 000
X_1 = $46 000
X_2 = $54 000
$P(46\,000 < X < 54\,000)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{46\,000 - 50\,000}{5\,000} = -\frac{4\,000}{5\,000} = -0.8$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{54\,000 - 50\,000}{5\,000} = \frac{4\,000}{5\,000} = 0.8$$

El valor en las tablas tanto para Z_1 como para Z_2 es 0.2881. Para determinar la probabilidad requerida es necesario sumar las probabilidades de que Z sea mayor que –0.8 y que Z sea menor que 0.8, no siendo necesario realizar resta alguna por la condición de que los valores estandarizados de tablas son acumulados (véase la figura. 6.9). Con lo anterior se tiene:

$P(46\,000 < X < 54\,000) = P(-0.8 < Z < 0.8)$
$= 0.2881 + 0.2881 = 0.5762$

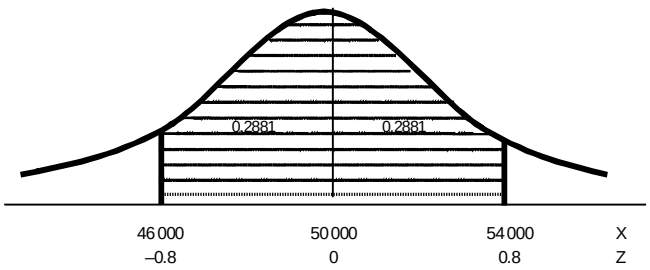


Figura 6.9. Distribución normal del costo de ejecución de un contrato de una empresa.

En conclusión, la probabilidad de que el costo de ejecución del contrato se encuentre entre $46 000 y $54 000 es de 0.5762 o de 57.62%.

Ejercicio 2

1. Una empresa de automóviles realizó un estudio de tiempos y movimientos, en dicho estudio se detectó que el ensamblado de un automóvil sigue una distribución normal con una media de 27.8 minutos y una desviación estándar de 4.0 minutos.

   a) ¿Cuál es la probabilidad de que este tipo de automóvil se pueda ensamblar en menos de 25 minutos?
   b) ¿Cuál es la probabilidad que se encuentre entre 26 y 30 minutos?

2. Una empresa paga a sus empleados un salario promedio de $20 por hora con una desviación estándar de $2. Si los salarios están aproximadamente distribuidos en forma normal, ¿qué porcentaje de los trabajadores recibe salarios entre $18 y $23 por hora?

3. Se sabe que el ciclo de vida de un componente eléctrico sigue una distribución normal con una media de 2 000 horas y una desviación estándar de 200 horas. Calcula la probabilidad de que un componente aleatoriamente seleccionado dure entre 2 000 y 2 400 horas.

4. La demanda anticipada de un producto en el próximo mes para cierta compañía puede representarse como una variable aleatoria normal, con una media de 1 200 unidades y desviación típica de 100 unidades. ¿Cuál es la probabilidad de que la demanda sea superior a 1 000 unidades?

5. Una compañía de reparación de fotocopiadoras encuentra, revisando sus expedientes, que el tiempo invertido en realizar un servicio se representa como una variable normal con media de 65 minutos y desviación estándar de 20 minutos. Calcula:

   a) La proporción de servicios que se hacen en menos de 60 minutos.
   b) La proporción de servicios que se hacen en menos de 90 minutos.

6.3. Relación entre la distribución normal y la binomial

Cuando se analizaron las distribuciones de probabilidad discretas se pudo apreciar que una de las distribuciones más útiles es la distribución binomial, la cual estaba asociada al experimento, que consiste en considerar dos posibles resultados "éxito" o "fracaso".

Debido a que se presentan dos posibles resultados, es importante conocer cuántos éxitos (y/o fracasos) se espera obtener dentro de un espacio muestral que contiene n ensayos. Como se incluye un número de éxitos (k) y el número de fracasos ($n-k$) en n ensayos, para conocer el número de formas totales se obtiene el número de combinaciones de n elementos de los cuales k pertenece a una clase y $n-k$ a otra clase distinta, entonces la probabilidad de que ocurra un evento es:

$$P(k) = P(k, n, p) = \frac{n!}{k!(n-k)!} \cdot p^k \cdot q^{n-k}$$

Donde:

$P(k)$ = Probabilidad de ocurrencia de la variable X tome un valor cualquiera.
k = Número de éxitos.
$n-k$ = Número de fracasos.
p = Probabilidad de éxito.
q = Probabilidad de fracaso.

En el ejemplo, la fórmula $\frac{n!}{k!(n-k)!}$ se refiere a las posibles combinaciones que pueden darse con el conjunto de datos con el que se cuenta, es decir, se muestra cuántos posibles resultados se pueden obtener si se desean sólo éxitos, dentro de un conjunto de datos.

Una de las limitaciones de la distribución de probabilidad binomial es que únicamente tiene aplicaciones donde la muestra es relativamente pequeña, los cálculos rara vez se extienden más allá de $n = 30$, debido a que al calcular el factorial de números mayores a 30, se tendrían problemas para efectuar tales operaciones. Además, es posible trabajar la distribución binomial mediante tablas, pero también tenemos el problema de que la tabla sólo abarca hasta la observación 30.

Cuando el número de observaciones es relativamente grande, el empleo de la distribución de probabilidad normal resulta ser muy útil para dar una aproximación a la distribución binomial. Como se observó en la sección anterior, no es difícil el empleo de la distribución normal. La distribución normal es más efectiva cuando la probabilidad de éxito está próxima a 0.5 y dicha aproximación se incrementa a medida que aumenta el número de observaciones.

La relación que existe entre ambas distribuciones se da cuando dentro de la distribución binomial se quiere conocer la probabilidad de ocurrencia de que la variable X tome un valor en particular y se debe obtener la media y la desviación estándar de un número grande de datos, ya que si se cuenta con un tamaño de muestra con muchos datos no es posible calcular en la mayoría de las ocasiones el factorial de un número, por lo que se tiene que suponer que los datos se comportan de manera normal y, al emplear el tamaño de muestra y la probabilidad asociada, es posible efectuar el cálculo de los parámetros de la población. Es entonces cuando al conocer tales parámetros se puede inferir acerca del comportamiento de una variable a través del valor Z.

Una de las dificultades que se presenta cuando se quiere emplear la distribución de probabilidad normal como una aproximación de la distribución binomial es que la distribución normal es continua, en tanto que la distribución binomial es discreta. Es importante recordar que las variables discretas emplean únicamente valores enteros, mientras que las variables continuas emplean todos los valores que se encuentran dentro de un intervalo, incluyendo enteros.

Como los datos de la distribución continua no son enteros, el problema se resuelve construyendo intervalos teóricos para poder representar valores enteros que sean parecidos a los que toman las

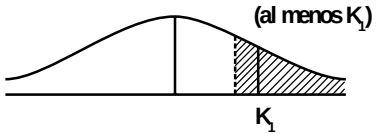
variables discretas. Esto se facilita mediante el empleo de una herramienta estadística conocida como factor de corrección por continuidad.

Debemos hacer tal ajuste porque se utiliza una distribución continua (distribución normal) para aproximar a una distribución discreta (binomial), pues de lo contrario, si únicamente se trabajara con cifras fraccionales, al plantear un problema no podría decirse que se quiere conocer la probabilidad de éxito de que se contraten a 3.4 personas o que se vendan 1.7 artículos.

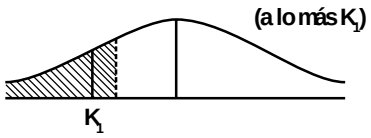
El factor de corrección por continuidad se representa a través del valor 0.5 para garantizar la simetría de la distribución normal y se suma o se resta dependiendo de cómo se haya diseñado el problema.

A continuación se presentan las distintas modalidades para el uso del factor de corrección:

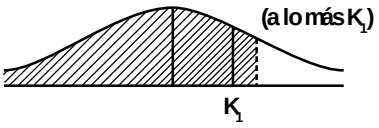
a) Si se trabaja en el extremo derecho de la distribución, restar 0.5 cuando se desea conocer $P(X \geq K_1)$.



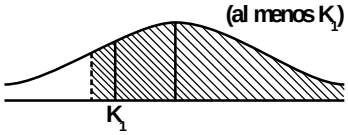
b) Si se trabaja en el extremo izquierdo de la distribución, sumar 0.5 cuando se desea conocer $P(X \leq K_1)$.



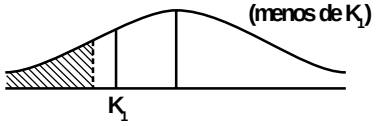
c) Si se trabaja en el extremo derecho de la distribución, sumar 0.5 cuando se desea conocer $P(X \leq K_1)$.



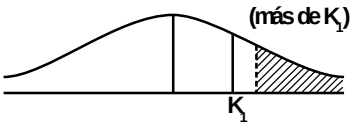
d) Si se trabaja en el extremo izquierdo de la distribución, restar 0.5 cuando se desea conocer $P(X \geq K_1)$.



e) Si se trabaja en el extremo izquierdo de la distribución, restar 0.5 cuando se desea conocer $P(X < K_1)$.



f) Si se trabaja en el extremo derecho de la distribución, sumar 0.5 cuando se desea conocer $P(X > K_1)$.



Para determinar el valor de la normal Z si se está haciendo una aproximación y sólo se conocen los datos que caracterizan a una distribución binomial, es necesario conocer la media μ y la desviación estándar σ. Ambas medidas se obtienen con:

$$\mu = np \qquad \text{y} \qquad \sigma = \sqrt{npq}$$

Ejemplo 4

Una empresa de construcción está contratando personal por expansión; 40% de las solicitudes que llegan son aceptadas, ¿cuál es la probabilidad de que en un grupo seleccionado al azar de 65 solicitudes se acepten más de 30?

Datos:
$p = 0.40$; $q = 0.60$; $n = 65$; $X = 30$; $P(X > 30)$

Por las características del problema y como habíamos mencionado (*f*), hay que sumar 0.5 a X = 30, para establecer un número similar a un límite inferior de clase, por lo que realmente X = 30 + 0.5 = 30.5

$$\mu = np = (65)\,(0.40) = 26 \qquad \sigma = \sqrt{npq} = \sqrt{(65)\,(0.40)\,(0.60)} = \sqrt{15.6} = 3.95$$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{30.5 - 26}{3.95} = \frac{4.5}{3.95} = 1.14$$

El valor en las tablas para Z = 1.14 es 0.3729
Al restar 0.3729 de 0.5 queda 0.1271 (véase la figura 6.10). Por lo tanto, la probabilidad de que se acepten más de 30 solicitudes es 0.1271 o de 12.71%.

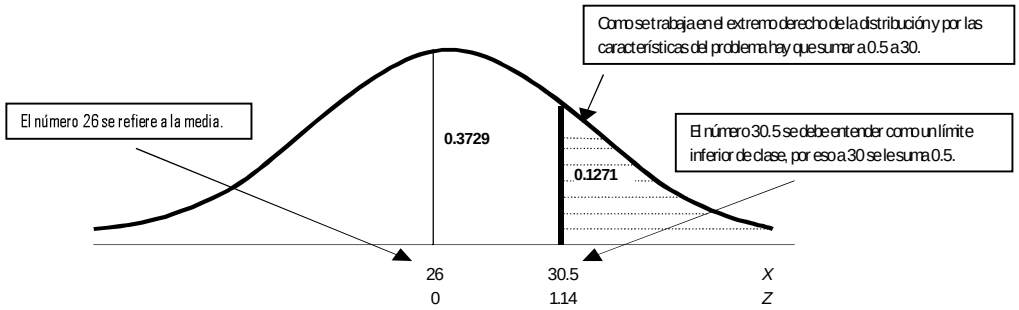


Figura. 6.10. Distribución de la contratación de personal.

Ejemplo 5

La gerencia de finanzas de una empresa detectó que el departamento de crédito y cobranza tiene 25% del total de las facturas con un atraso en el cobro de un mes. Si se toma una muestra aleatoria de 45 facturas, ¿cuál es la probabilidad de que sean menos de 10 las facturas atrasadas?

Datos:
$p = 0.25$
$q = 0.75$
$n = 45$
$X = 10$
$P(X < 10)$

Como ahora se trabaja en el extremo izquierdo de la distribución (e) $X = 10 - 0.5 = 9.5$

$$\mu = np = (45)\,(0.25) = 11.25 \qquad \sigma = \sqrt{npq} = \sqrt{(45)(0.25)(0.75)} = \sqrt{8.44} = 2.90$$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{9.5 - 11.25}{2.90} = \frac{1.75}{2.90} = -0.60$$

El valor en las tablas para $Z = -0.60$ es 0.2257
Al restar 0.2257 de 0.5 nos queda 0.2743 (véase la figura 6.11). La probabilidad de que existan menos de 10 facturas atrasadas es 0.2743 o de 27.43%.

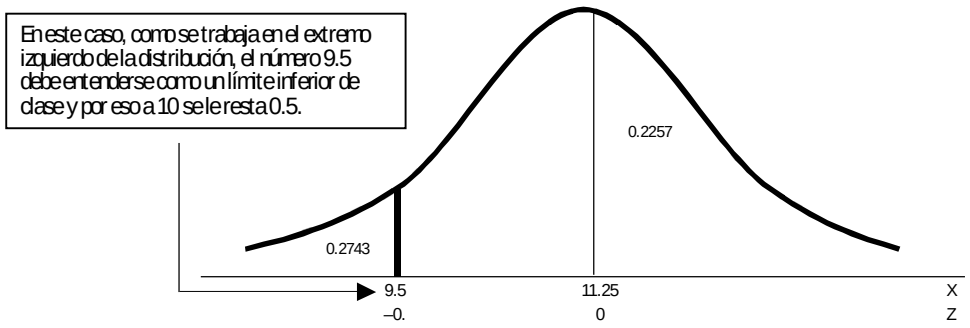


Figura 6.11. Distribución de las facturas sin cobrar.

Ejemplo 6

Se sabe que 10% de las unidades producidas por un proceso de fabricación resultan defectuosas. De la producción total de un día, se seleccionan 100 unidades aleatoriamente, ¿cuál es la probabilidad de que al menos 8 unidades resulten defectuosas?

Datos:
$q = 0.90$
$p = 0.10$
$n = 100$

$X = 8$

$P(X \geq 8)$

$\mu = nq = (100)\ (0.10) = 10$ $\qquad$ $\sigma = \sqrt{npq} = \sqrt{(100)\ (0.10)\ (0.90)} = \sqrt{9} = 3$

Sustituyendo en la fórmula de la normal se obtiene:
Al centrar la atención en el extremo izquierdo (*a*) $X = 8 - 0.5 = 7.5$

$$Z = \frac{X - \mu}{\sigma} = \frac{7.5 - 10}{3} = -\frac{2.5}{3} = -0.83$$

El valor de tablas para $Z = -0.83$ es 0.2967. Como el problema está pidiendo que al menos 8 unidades sean defectuosas, se utiliza el símbolo $\geq$, definiendo de esta manera que para determinar la probabilidad requerida, es necesario que a 0.2967 se le sume 0.5000 quedando el área bajo la curva igual a 0.7967 (véase la figura 6.12).

La probabilidad de que al menos 8 unidades resulten defectuosas es 0.7967 o de 79.67%.

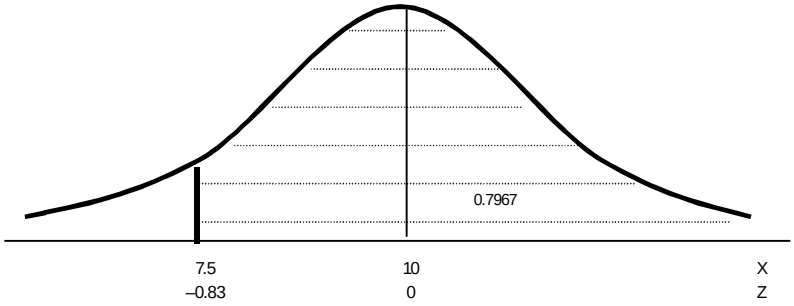


Figura 6.12. Distribución de producción defectuosa.

Ejercicio 3

1. Una empresa realiza un estudio de mercado para saber si es viable la introducción de un nuevo detergente en el mercado. El estudio reporta que aproximadamente 75% de las mujeres opina que el detergente es bueno. De las siguientes 80 personas entrevistadas,

   a) ¿Cuál es la probabilidad de que al menos 50 sean de la misma opinión?
   b) ¿Cuál es la probabilidad que más de 56 personas sean de la misma opinión?

2. La administración de una empresa de reconocido prestigio ha decidido ofrecer una agresiva política de servicio a clientes, dicha política consiste en aceptar devoluciones sin discusión alguna. El número promedio de clientes que regresan la mercancía es de 10% por día; si se elige una muestra al azar de 70 clientes, ¿cuál es la probabilidad de que más de 5 clientes regresen la mercancía?

3. En relación con un grupo extenso de prospectos de venta se ha observado que 30% de los contactados personalmente por un representante de ventas realizará una compra. Si un representante de ventas contacta a 30 prospectos, determina la probabilidad de que 10 o más realicen una compra.

4. Una tienda departamental efectúa un estudio y determina que 70% de los clientes que acude realizan al menos una compra. En una muestra de 50 individuos, ¿cuál es la probabilidad de que al menos 40 personas realicen una compra o más cada uno?

6.4. Distribución *t*, uniforme y exponencial

6.4.1 Distribución *t*

En los apartados anteriores se utilizó la distribución normal, ya que resulta ser un buen instrumento para realizar inferencias cuando se trabaja con muestras grandes ($n \geq 30$) y se conoce la desviación estándar σ, su fórmula aplicada a la media muestral está dada por:

$$Z = \frac{\overline{X} - \mu}{\sigma}$$

Existen situaciones donde únicamente se dispone de muestras pequeñas ($n < 30$) y la desviación estándar (σ) no se conoce. El desconocimiento de la desviación estándar se debe a que en un determinado experimento el número de observaciones con que se cuenta no es lo suficientemente grande para representar las características de una población.

Para emplear una teoría que sea correspondiente con el problema a tratar, y que sea útil para realizar estudios con muestras pequeñas, se debe suponer que la muestra obtenida de la población sigue una distribución normal y, por lo tanto, se puede basar el estudio en la distribución *t*.

La distribución *t* es un conjunto de distribuciones que tienen un comportamiento muy similar a la distribución normal, con la salvedad de que sus datos tienen mayor dispersión. Se aplica para realizar inferencias cuando la muestra con la que se está trabajando es pequeña y además se desconoce la desviación estándar poblacional.

Para la aplicación de la distribución *t* se utiliza una fórmula estandarizada especialmente construida para trabajar con muestras pequeñas. Como en este caso no se conoce la desviación estándar de la población (σ), se debe emplear la desviación estándar de una muestra representada por "*S*". La distribución se puede expresar en los siguientes términos:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Donde:

$\overline{X}$ = Media muestral.
μ = Media poblacional.
S = Desviación estándar muestral como aproximación a la desviación estándar de la población σ.
n = Número de observaciones.

La fórmula de la distribución *t* muestra la relación que existe entre la diferencia de la media muestral $\overline{X}$ y la poblacional μ con respecto a la aproximación de la desviación estándar *S*, cabe mencionar que el valor de *S* es influido por los grados de libertad.

Los grados de libertad se obtienen restando uno al tamaño de la muestra ($n - 1$), cuando se está analizando una sola variable, por ejemplo X; los grados de libertad están relacionados con la varianza muestral "S^2". La noción de grado de libertad se emplea para denotar que se pierde un dato por cada parámetro que se calcula.

La gráfica de la distribución *t* es muy similar a la de la distribución normal y es simétrica con respecto al valor de la media μ. La forma exacta de la distribución *t* depende de los grados de libertad (*gl*).

Una diferencia de la distribución *t* con respecto de la distribución normal es que la primera presenta dispersiones mayores que la segunda y esa mayor variabilidad de la distribución *t* se debe a que los cálculos dependen tanto de la media muestral $\overline{X}$ como de la aproximación a la desviación estándar "*S*", mientras que los cálculos de la distribución normal dependen únicamente de la media $\overline{X}$ ya que la desviación estándar σ se conoce. Por esta razón, la distribución *t* es platicúrtica, es decir, más plana que la distribución normal.

Para poder determinar los valores en tablas de la distribución *t*, es necesario conocer tanto el nivel de confianza como el nivel de significancia con que se trabaja.

El nivel de confianza ($1 - \alpha$) es el grado o porcentaje de confiabilidad de que las decisiones que se toman son correctas, por ejemplo, el caso más usual es trabajar a 95% de nivel de confianza, el cual indica que existe una confiabilidad de 95% de que las decisiones tomadas sean las correctas. El nivel de significancia (α) se refiere al grado o porcentaje de error con que se espera estar trabajando o tomando decisiones erróneas.

El comportamiento de la distribución *t* es como el de la normal Z, tiene forma acampanada y es perfectamente simétrica con respecto al punto medio de la curva, sin embargo, es más plana que la distribución normal (véase la figura 6.13).

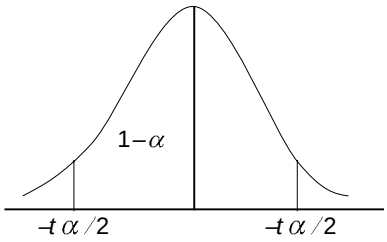


Figura 6.13. Área bajo la curva de la distribución *t*.

En la gráfica puedes observar que el *nivel de confianza* se representa por ($1 - \alpha$) y el nivel de significancia por (α). En algunas ocasiones, cuando se pretende estimar una medida de tendencia central de la población con base en la información proporcionada por una muestra, al trabajar con una distribución como *t* el *nivel de significancia* debe partirse a la mitad ya que existe una perfecta simetría en la distribución, mostrando que existen dos colas (extremos), por esta razón se debe buscar en las tablas el nivel de significancia partido a la mitad ($\alpha/2$).

Regularmente se trabaja con 90, 95 y 99% de confianza y, por lo tanto, con un 0.10, 0.05 y 0.01 de significancia.

Por ejemplo, cuando $1 - \alpha = 95\%$ la significancia es entonces, $\alpha = 1 - .95$ o $\alpha = 0.05$

Por último, es importante destacar que aunque la distribución *t* se emplea para realizar inferencias con muestras pequeñas, los valores críticos que aparecen en la tabla se basan en el supuesto de que la muestra extraída de la población tiene una distribución de probabilidad normal.

Para obtener el valor de tablas es necesario conocer tanto el nivel de significancia con el que se va a trabajar como los grados de libertad. Una vez que se eligió el nivel de significancia y se calcularon los grados de libertad, en la tabla se buscarán en la columna de la izquierda los grados de libertad y en la parte superior el nivel de significancia. Por ejemplo, al buscar un valor en la tabla se procede de la siguiente manera: si $n = 10$ y se trabaja a 95% de confianza, entonces:

$1 - \alpha = 95\%$
$\alpha = 5\%$ o $\alpha = 0.05$
Los grados de libertad son $n - 1 = 10 - 1 = 9$
Con $\alpha = 0.05$ y $n - 1 = 9$ el valor de $t_{\alpha} = 1.833$, y si el valor $t_{\alpha/2} = 2.262$,
Esto puede apreciarse en la tabla siguiente (gráfica de una sola cola):

Grados de libertad	0.15	0.1	0.05	0.025	0.01	0.005
1	1.963	3.078	6.314	12.706	31.821	63.657
2	1.386	1.886	2.920	6.314	6.965	9.925
3	1.250	1.638	2.353	4.303	4.541	5.841
4	1.190	1.533	2.132	2.776	3.747	4.604
5	1.156	1.476	2.015	2.571	3.365	4.032
6	1.134	1.440	1.943	2.447	3.143	3.707
7	1.119	1.415	1.895	2.365	2.998	3.499
8	1.108	1.397	1.860	2.306	2.896	3.355
9	1.100	1.383	1.833	2.262	2.821	3.250
10	1.093	1.372	1.812	2.228	2.764	3.169

Tabla 6.2. Segmento de la tabla *t-student.*

Ejercicio 4

1. La distribución *t student* se utiliza:

   a) Para estimar el desempeño de los estudiantes en una universidad.
   b) Para estimar la varianza poblacional con base en una muestra.
   c) Para hacer inferencia cuando se dispone de muestras pequeñas.
   d) Para hacer inferencia cuando se dispone de muestras grandes.

2. La distribución *t student* se diferencia de la distribución normal porque:

   a) La distribución *t* es más platicúrtica y la normal es mesocúrtica.
   b) La distribución *t* no es simétrica mientras que la normal sí lo es.
   c) La distribución *t* es acampanada, mientras que la normal no lo es.
   d) La distribución *t* es leptocúrtica mientras que la normal es mesocúrtica.

3. La distribución *t student* tiene:

   a) Menor asimetría que la distribución normal.
   b) Mayor asimetría que la distribución normal.
   c) Menor dispersión en sus datos que la distribución normal.
   d) Mayor dispersión en sus datos que la distribución normal.

4. Cuando los grados de libertad son obtenidos mediante la expresión $(n - 1)$:

   a) Se está analizando una sola variable.
   b) Se están analizando dos variables.
   c) Se está analizando un conjunto de variables de manera simultánea.
   d) No existe ninguna variable que se esté examinando.

5. El nivel de significancia representa:

   a) Qué tan significativo es realizar una estimación.
   b) El grado o porcentaje de error con que se espera estar trabajando o tomando decisiones erróneas.
   c) El grado de confiabilidad que representa la estimación que se desea llevar a cabo.
   d) El nivel óptimo de llevar a cabo un método de estimación con muestras pequeñas.

6. Si $n = 20$ y se trabaja con un nivel de confianza de 95% para estimar una variable, los valores t_{α} y $t_{\alpha/2}$ son:

   a) $t_{\alpha} = 1.724$ y $t_{\alpha/2} = 2.086$
   b) $t_{\alpha} = 1.729$ y $t_{\alpha/2} = 2.093$
   c) $t_{\alpha} = 1.325$ y $t_{\alpha/2} = 1.724$
   d) $t_{\alpha} = 0.05$ y $t_{\alpha/2} = 0.025$

6.4.2. Distribución uniforme

Existen situaciones donde todos los eventos posibles de un experimento tienen la misma probabilidad de ocurrir, por ejemplo, si 100 personas compran un número de lotería las 100 tienen la misma probabilidad de ganar o si 20 personas con el mismo grado de estudios y capacidades similares se presentan al departamento de recursos humanos de una empresa a solicitar un empleo, las 20 tienen la misma probabilidad de ser aceptadas.

Cuando una variable aleatoria asume una serie de valores en una escala continua entre dos puntos, de tal manera que ninguno de los valores tenga más probabilidad que los demás de ocurrir, entonces la probabilidad relacionada con la variable aleatoria continua se puede presentar mediante una distribución uniforme.

Una distribución de probabilidad uniforme contiene todos los valores posibles que puede tomar una variable aleatoria continua y todos estos valores tienen la misma probabilidad de ser tomados por la variable aleatoria.

A diferencia de la distribución normal estandarizada y de la distribución *t*, la distribución de probabilidad uniforme se puede obtener sin necesidad de recurrir al uso de fórmulas estandarizadas ni al empleo de tablas.

Para una mejor comprensión de lo que es la distribución uniforme se hace uso de un gráfico, mediante un rectángulo (véase la figura 6.14).

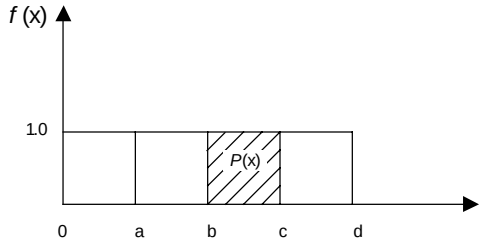


Figura 6.14. Distribución uniforme.

La altura del rectángulo de la figura 6.14 es igual a 1.0 y el área a 100%, como puedes observar, el rectángulo está dividido en cuatro partes con una misma probabilidad, es decir, cada una con una probabilidad de $\frac{1}{4}$. Por lo tanto, el área bajo el rectángulo entre dos puntos cualesquiera, por ejemplo *c* y *d*, es igual al porcentaje o área del intervalo total incluido entre *b* y *c*. Una fórmula para representar esto sería:

$$P(b \leq X \leq c) = \frac{c-b}{d-a}$$

Donde:
a = Valor mínimo de la distribución.
d = Valor máximo de la distribución.
b = Valor mínimo del rango con que se trabaja.
c = Valor máximo del rango con que se trabaja.

La fórmula muestra que si se quiere conocer la probabilidad de que el valor que toma una variable está entre dos puntos, *b* y *c*, hay que tomar la diferencia existente entre los valores que toma la variable en esos puntos y esa diferencia dividirla entre la resta de los valores máximo y mínimo de la distribución.

Si el objetivo fuera encontrar la probabilidad entre *a* y *c*, entonces la fórmula estaría dada por:

$$P(a \leq X \leq c) = \frac{c-a}{d-a}$$

Como no sólo nos interesa la probabilidad de que el valor que toma una variable esté en cierto intervalo, existen algunas aplicaciones en el mundo real donde es necesario el uso de fórmulas especiales sobre la media y la desviación estándar para una distribución de probabilidad uniforme. Las fórmulas para obtener la media y la desviación estándar están dadas por:

Media: $\mu = \frac{a+d}{2}$

$Var(X) = \sigma^2 = \frac{(d-a)^2}{12}$

Desviación estándar $\sigma = \sqrt{Var\,(X)}$

Ejemplo 7

Se espera que las ventas de computadoras de una importante empresa sigan una distribución de probabilidad uniforme. Debido a las limitantes del mercado, las ventas mensuales no pueden ser menores de 5 000 computadoras o superiores a 25 000.

a) ¿Cuál será la probabilidad de que al menos se vendan 20 000 computadoras?
b) ¿Cuál será la probabilidad de que se vendan entre 10 000 y 15 000 computadoras?

Antes de resolver este punto es importante trazar un gráfico que represente cada uno de los puntos que se van a considerar (véase la figura 6.15)

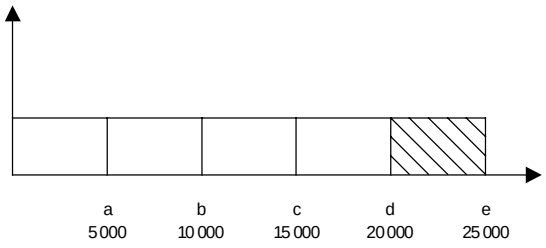


Figura 6.15. Distribución uniforme de las ventas mensuales de computadoras.

El valor medio viene dado por $\mu = \frac{a+e}{2} = \frac{5\,000+25\,000}{2} = \frac{30\,000}{2} = 15\,000$

Como el objetivo es encontrar la probabilidad de que al menos se vendan 20 000, entonces hay que centrar la atención en el intervalo que corresponde a los puntos *d* y *e*. Por lo tanto:

$$P(d \leq X \leq e) = \frac{e-d}{e-a} = \frac{25\,000-20\,000}{25\,000-5\,000} = \frac{5\,000}{20\,000} = 0.25$$

La probabilidad de que las ventas alcancen por lo menos 20 000 computadoras es de 25%.

b)

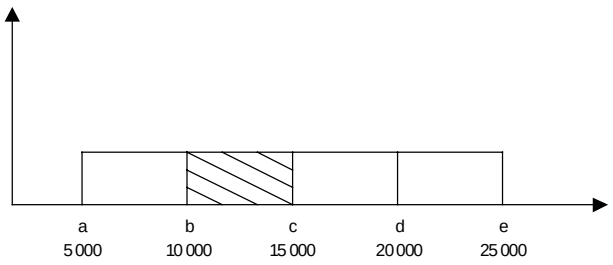


Figura 6.16. Distribución uniforme de las ventas mensuales de computadoras.

Como el objetivo es encontrar la probabilidad de que las ventas sean de 10 000 a 15 000 computadoras, nos interesa el intervalo que corresponde a los puntos *c* y *b*, por lo tanto la fórmula puede ser:

$$P(b \leq X \leq c) = \frac{c-b}{e-a} = \frac{15\,000-10\,000}{25\,000-5\,000} = \frac{5\,000}{20\,000} = 0.25$$

La probabilidad de que las ventas se encuentren entre 10 000 y 15 000 computadoras es 25%.

Como puedes apreciar, la distribución uniforme es útil cuando queremos conocer la probabilidad de que un determinado valor que ha de tomar alguna variable a estudiar se encuentre en un intervalo o rango de valores perfectamente definido.

Ejercicio 5

1. Las ventas de una gasolinera alcanzan en promedio los 40 000 litros diarios y un mínimo de 30 000, si las ventas del combustible siguen una distribución uniforme, ¿cuál es la probabilidad de que las ventas de gasolina excedan los 35 000 litros?

2. Una compañía productora de acero corta y vende tubos con medidas que van de 1 a 5 metros, estas medidas son las más demandadas en el mercado

   a) ¿Cuál es la medida promedio de un tubo?
   b) ¿Cuál es la probabilidad de que un tubo sea mayor de 3 metros?

3. Los ingresos familiares en una colonia determinada se encuentran entre 4 800 y 7 200 pesos mensuales. Si a un especialista en tendencias de consumo le interesa determinar el ingreso promedio con el fin de establecer una estrategia publicitaria sobre algunos artículos, calcula la probabilidad de que los ingresos familiares estén entre 6 000 y 7 200 pesos.

4. Un consultor comienza a trabajar en un proyecto. El beneficio esperado oscila entre 30 000 y 70 000 pesos. ¿Cuál es la probabilidad de que el beneficio del consultor esté entre 50 000 y 60 000 pesos?

5. Un vendedor recibe un salario anual de entre 150 000 y 250 000 pesos, según su productividad. Calcula la probabilidad de que:

   a) Tenga ingresos superiores a 175 000 pesos.
   b) Sus ingresos sean menores a 200 000 pesos.

6.4.3. Distribución exponencial

Otra de las distribuciones discretas relevantes que se analizaron en la unidad anterior fue la distribución de Poisson. Cuando el número de datos con que se quiere trabajar es muy grande y ocurren en un proceso de Poisson, es decir, se caracterizan por un entorno de tiempo o de espacio, se dice entonces que sigue una distribución exponencial de probabilidad. Por lo tanto, se puede aseverar que existe una estrecha relación entre la distribución de Poisson y la distribución exponencial.

La distribución exponencial aborda fenómenos cuya probabilidad se refiere a tiempo y distancias entre la ocurrencia de experimentos con respecto a un intervalo continuo.

Existe una gran cantidad de fenómenos asociados a la distribución exponencial. Algunos ejemplos son:

- Número de automóviles vendidos por día.
- Número de defectos por lote en un proceso de producción.
- Número de clientes atendidos en un banco en un día.
- El tiempo entre llegadas de clientes a un supermercado.

Una de las diferencias entre la distribución de probabilidad de Poisson y la exponencial es que, mientras el proceso de Poisson es estacionario, es decir, la probabilidad es la misma para todos los eventos a lo largo de todo el intervalo de tiempo, la distribución exponencial se puede aplicar cuando el interés consiste únicamente en conocer el tiempo o la distancia hasta el primer evento, el tiempo o la distancia entre dos eventos sucesivos.

Las probabilidades exponenciales se pueden expresar en términos de tiempo o distancia. Al igual que las otras distribuciones de probabilidad es posible utilizar una fórmula para calcular la distribución exponencial de probabilidad, la fórmula varía según el caso que se esté tratando. A continuación se presentan dos casos de interés:

Cuando λ es el número medio de ocurrencias y el objetivo es encontrar la probabilidad de que el evento no suceda en el intervalo especificado, entonces la fórmula es:

$$P(T > t) = e^{-\lambda t}$$

Donde:
e = 2.71828, la base de los logaritmos naturales.
λ = Constante positiva igual a la media de la distribución.
t = Tiempo.
T = Evento que se quiere delimitar.

De igual manera, cuando λ es el número medio de ocurrencias y lo que se busca es la probabilidad de que un evento ocurra en el curso del intervalo establecido, entonces la fórmula es:

$$P(T \leq t) = 1 - e^{-\lambda t}$$

Ejemplo 8

El departamento de servicio a clientes de una empresa de teléfonos celulares recibe tres llamadas en un promedio de 15 minutos, las llamadas provienen de clientes a quienes los celulares les han salido defectuosos.

a) ¿Cuál es la probabilidad de que las tres llamadas ocurran en un tiempo mayor de 15 minutos?
b) ¿Qué probabilidad existe de que el tiempo sea de 15 minutos o menos?

a) Lo primero que debemos obtener es el valor de λ, es decir, el número de llamadas por minuto.

Se tiene que $\lambda = \frac{3}{15} = 0.2$ llamadas por minuto.

Como se considera que el tiempo sea mayor a 15 minutos, se estaría trabajando fuera del intervalo considerado y el siguiente paso es sustituir en la primer fórmula y resolver:

$$P(T > t) = e^{-\lambda t}$$

$$P(T > 15) = e^{-0.2(15)} = e^{-3} = 0.049$$

La probabilidad de que ocurran tres llamadas en un tiempo mayor de 15 minutos es 0.049 o de 4.9%.

b) Para resolver este inciso, debido a que se trabaja dentro del intervalo considerado, se utiliza la fórmula $P(T \leq t) = 1 - e^{-\lambda t}$, por lo que, sustituyendo se obtiene:

$$P(T \leq 15) = 1 - e^{-3} = 1 - 0.049 = 0.95$$

La probabilidad de que las llamadas ocurran en 15 minutos o menos es 0.95 o de 95%.

Ejemplo 9

Una empresa de telas ha detectado que en un rollo de 100 metros de tela hay un metro que está dañado en promedio, ¿cuál es la probabilidad de que el metro de tela dañado se encuentre en los primeros 40 metros de tela?

Como $\lambda = \frac{1}{100} = 0.01 \qquad \lambda t = 0.01\,(40) = 0.4$

$$P(T \leq 40) = 1 - e^{-\lambda t} = 1 - e^{-0.01(40)} = 1 - e^{-0.4} = 1 - 0.67 = 0.33$$

La probabilidad de que el metro de tela dañado se encuentre en los primeros 40 metros es 0.33 o de 33%.

EJERCICIOS

Ejercicio 6

1. Las llamadas de emergencia que recibe un hospital durante las primeras horas del día lunes siguen un modelo exponencial, con un tiempo medio de 20 minutos por cada llamada.

   a) Calcula la probabilidad de que el tiempo en que se espera una llamada sea mayor a 20 minutos.
   b) Obtén la probabilidad de que el tiempo en que se tarda en recibir una llamada sea igual o menor a 20 minutos.
   c) Encuentra la probabilidad de que el tiempo de espera de una llamada sea de 10 minutos o menos.

2. En promedio 5 personas tardan 10 minutos en retirar dinero de un cajero automático, ¿cuál es la probabilidad de que tarden más de 5 minutos?

3. El tiempo de atención al cliente en un servicio de información de una biblioteca sigue una distribución exponencial, con un tiempo de servicio medio de 3 consultas cada 5 minutos, ¿cuál es la probabilidad de que las 3 consultas se realicen en más de 5 minutos?

4. En unos grandes almacenes, la oficina de atención al cliente recibe en promedio 6 reclamaciones cada 30 minutos sobre la calidad del servicio, ¿cuál es la probabilidad de que se reciban esas 6 reclamaciones en más de 30 minutos?

5. Un analista hace predicciones sobre las ganancias de una corporación. Si las ganancias promedio son de 60 000 mensuales, realizando 4 servicios cada 5 días, cuál es la probabilidad de que:

   a) Se hagan 4 servicios en más de 5 días.
   b) Se efectúen 4 servicios en 5 días o menos.

6.5. Tipos de muestreo

Usualmente se recurre a la realización de muestras para describir las características de una población. Para que los resultados obtenidos en el proceso de inferencia estadística sean confiables, se debe tener una muestra que represente en gran medida al conjunto total de datos, es decir, que contenga las características esenciales de la población.

Mediante una buena metodología, una muestra elegida adecuadamente suministra la información más relevante de una población, tendiendo diversas ventajas sobre un censo; por ejemplo, una muestra genera un menor costo, mayor rapidez y precisión en la obtención de los resultados, además de recolectar la información deseada en aquellos casos en los que es imposible realizar un censo.

Por ejemplo, si una compañía se dedica a realizar estudios de mercado, le interesará efectuar una encuesta sobre los gustos y preferencias de los consumidores con el fin de observar si la introducción de un nuevo producto al mercado es viable o no. En este caso, sólo se encuesta a una parte de la población a quienes irá dirigido el producto. Elegir una muestra representativa de los consumidores implica que el costo de la encuesta (entre otros factores) es menor que si se considerara al total de los consumidores.

Al proceso de diseñar y analizar una muestra se le conoce con el nombre de *muestreo*, el cual se puede definir como:

El muestreo es un proceso donde se elige una muestra que sea capaz de representar a la población de manera que no se pierdan los rasgos y características más relevantes de ésta.

El muestreo suele aplicarse en las distintas áreas económico administrativas, tales como:

- En los estudios de mercado. Por ejemplo, si una empresa de la industria cigarrera está interesada en introducir un nuevo tipo de cigarro en el mercado y quiere saber si el producto tendrá aceptación, acude a la realización de un muestreo. Si se conoce que aproximadamente hay una población de 15 millones de personas que fuman, es posible encuestar a 1 500 fumadores que conformen una muestra representativa.

- En el control de calidad. Por ejemplo, cuando el departamento de ventas de una empresa de pinturas envía una muestra de 100 litros de pintura de un nuevo color a cada uno de sus mejores clientes con el fin de conocer si la pintura cumple con los requerimientos de calidad de los clientes. Los resultados permitirán determinar si el producto será o no aceptado, además de hacer posible la estimación del monto de ventas que se ha de alcanzar.

- En el diseño de política de desarrollo. Por ejemplo, si el gobierno de una ciudad está interesado en introducir una nueva ruta de transporte colectivo, primero debe analizar si existe la demanda para este tipo de servicio. Para ello realiza un muestreo con el fin de conocer a cuánta gente beneficiaría la nueva ruta, observando durante algunos días el número de posibles pasajeros, realizando encuestas a algunos de ellos y, de esta manera, realizar el análisis de costo beneficio para saber qué tan rentable le resulta impulsar este proyecto de desarrollo.

- Para analizar fenómenos macroeconómicos. Por ejemplo, el Banco de México estima el Índice Nacional de Precios al Consumidor (INPC), recolectando y procesando la información estadística mediante la toma de aproximadamente 450 bienes y servicios que forman una muestra representativa de los precios y cantidades de la economía de todo del país.

- Para medir las preferencias electorales. Por ejemplo, durante las campañas presidenciales los partidos políticos y los medios de comunicación llevan a cabo distintos tipos de muestreo para conocer las preferencias de los electores y así diseñar estrategias de campaña o informar al público sobre el desarrollo del proceso electoral.

- Para medir la opinión pública. Por ejemplo, algunos noticieros o gobiernos de las entidades federativas elaboran métodos de muestreo con el propósito de conocer la opinión pública sobre la aceptación o no de alguna política que se desea llevar a cabo.

- Para diseñar estrategias de publicidad. Por ejemplo, cuando una agencia desea realizar un comercial con el objetivo de incrementar las ventas de cerveza de una compañía. Si la agencia estima que el mercado está constituido por 10 millones de consumidores, puede evaluar a una muestra de 2 000, mostrándoles distintos comerciales para saber cuál sería el comercial más efectivo en cuanto a la impresión que les causó.

Existe una serie de argumentos de por qué es preferible llevar a cabo muestreo y no censos, dentro de tales argumentos se pueden señalar los siguientes:

a) Si se trabaja con muestras se minimizan los costos que representa la recolección de la información, ya que sólo se necesita conocer una parte de la población, es decir, el número de observaciones "*n*" es menor.

b) En un muestreo existe una mayor exactitud en cuanto a los datos obtenidos en comparación con los censos, esto se debe a que la información recabada es seleccionada y minuciosamente analizada.

c) El grado de error suele ser menor en una muestra que en un censo, por lo complejo que resulta hacer un conteo de la totalidad de los elementos.

d) La información proporcionada por un muestreo se obtiene con mucha mayor rapidez que si se deseara llevar a cabo un censo.

e) En ocasiones resulta imposible elaborar un censo debido a la naturaleza del problema o al tipo de población que se desea examinar.

Un censo es un proceso en el que se recolecta la información contenida en el total de una población. A pesar de sus desventajas ante la muestra, en ocasiones se lleva a cabo este tipo de mediciones. Por ejemplo, a nivel nacional se realizan diversos censos para medir algunos fenómenos poblacionales o económicos. Por ejemplo, el Censo General de Población y Vivienda que se realiza cada 10 años o los Censos Económicos que se elaboran cada cinco años en nuestro país.

Antes de examinar los distintos tipos de muestreo, es importante señalar que en la estadística inferencial existen dos tipos de poblaciones: las poblaciones finitas y las poblaciones infinitas.

Población finita es aquella que está compuesta de un número determinado de elementos, objetos u observaciones.

Algunos ejemplos de población finita son:

- El número de vuelos diarios en el aeropuerto de la Ciudad de México.

- El número de empleados en un banco.
- La producción diaria de computadoras en una fábrica.
- La cartera de clientes de un ejecutivo de cuenta de una casa de bolsa.

Población infinita es aquella en la que no es posible conocer el número determinado de elementos u objetos que la componen.

Algunos ejemplos de poblaciones infinitas pueden ser los siguientes:

- El número de robos al año no denunciados ante el ministerio público.
- El número de niños de la calle que no cenaron en la Nochebuena.
- Las operaciones diarias de compra-venta de dólares alrededor del mundo.
- El número de artículos de contrabando introducidos al país cada año.

Cabe destacar que en aplicaciones de casos reales usualmente se emplea la población finita, aunque no se descarta el uso de poblaciones infinitas, ya que cuando la población es muy grande se considera como una población infinita.

El muestreo debe garantizar que la muestra estadística sea lo suficientemente representativa y describa plenamente a la población. De esta manera, la problemática principal que enfrenta el muestreo consiste en extraer la información de las variables de la población que se piensa analizar.

Como existen distintas maneras en que las poblaciones se encuentran estructuradas, dependiendo de sus características, es necesario utilizar el tipo de muestreo más adecuado para realizar las respectivas inferencias.

Este apartado tiene como finalidad hacer una revisión de los distintos tipos de muestreo, dentro de los que se pueden destacar: el *muestreo aleatorio*, el *muestreo estratificado* y el *muestreo sistemático*.

6.5.1. Muestreo aleatorio simple

Uno de los modelos de muestreo más conocido y que más se emplea es el denominado *muestreo aleatorio simple*, el cual se relaciona con el tema de las probabilidades.

El muestreo aleatorio simple para poblaciones finitas es un proceso mediante el cual se obtiene una muestra de tal manera, que tuvo la misma probabilidad de ser seleccionada que el resto de las muestras posibles que pudieron recolectarse.

Por ejemplo, si se desea conocer la opinión de los suscriptores de un programa de cliente frecuente de una aerolínea, y para ello se selecciona un vuelo de la Ciudad de México a Tijuana, esta muestra del tamaño de n clientes que se tome del total de la población de N suscriptores a dicho programa, debió tener la misma probabilidad de ser seleccionada que si se hubiera seleccionado cualquier otro vuelo de esta compañía. Nota que el tamaño de esta población es finito.

El muestreo aleatorio simple para poblaciones infinitas es un proceso en el cual se obtiene una muestra donde cada uno de sus elementos fue recolectado de la misma población y de manera independiente a los demás.

Por ejemplo, si se desea conocer el tipo de cambio (peso-dólar) a una determinada hora del día, la población bajo estudio es infinita, pues resulta imposible conocer el total de las transacciones de compra venta entre el peso y el dólar realizadas en todo el mundo a esa hora determinada. En este sentido, las transacciones que únicamente deben ser consideradas son aquellas que se realizan entre pesos y dólares, pues es la población bajo estudio; es decir, no debe tomarse en cuenta las operaciones realizadas entre el dólar y otras monedas distintas al peso. Por otra parte, cada una de las transacciones consideradas en la muestra deben ser independientes entre sí, es decir, que las personas seleccionadas tengan distintos motivos a las demás para realizar operaciones de compra venta de la divisa, por ejemplo, para especular, para salir de viaje, para importar productos extranjeros, etc. y no tomar en cuenta únicamente las transacciones realizadas por un solo cliente, un solo banco o por una sola institución financiera.

Aunque el término *aleatorio* implica que los elementos que forman parte de la muestra se seleccionan de manera fortuita o al azar, lo cierto es que el muestreo aleatorio no necesariamente debe cumplir este requisito; más bien requiere un proceso cuidadoso en cuanto al diseño con el fin de asegurar la independencia de los elementos, es decir, se busca que el resultado de un experimento no dependa de algún resultado que se obtuvo anteriormente.

Por lo tanto, para llevar a cabo un muestreo aleatorio se deben tomar en cuenta los siguientes aspectos:

1. Definir la población objetivo; es decir, identificar cuáles son las personas, productos o servicios que se desea estudiar.

2. Diseñar un método para llevar a cabo el muestreo aleatorio; es decir, definir la manera en que serán seleccionados los elementos de la muestra.

3. Llevar a cabo el muestreo aleatorio; es decir, seleccionar y recabar la información en cada uno de los elementos de la muestra.

Cuando se manejan poblaciones finitas y cuando sea posible, es recomendable tener acceso al marco muestral para diseñar el método de muestreo. El marco muestral es una lista del total de los elementos de la población. Por ejemplo, si se desea realizar un muestreo para conocer la preferencia política en un proceso electoral, el marco muestral podría ser el padrón electoral. Si se desea conocer el poder adquisitivo de los clientes de un banco, el marco muestral sería el listado de todos los clientes de ese banco.

Para seleccionar los elementos que compone la muestra se utilizan distintas técnicas, por ejemplo, tablas de números aleatorios o paquetes computacionales que proporcionan distintas series de números aleatorios.

Algunos ejemplos de los errores que se cometen al realizar un muestreo aleatorio son:

- Elaborar un muestreo para conocer las preferencias electorales de un país, encuestando únicamente a los participantes de un mitin político.

- Realizar una encuesta a la salida de una estación del metro para conocer la preferencia del tipo de aditivo o lubricante para el motor de los automóviles.

- Realizar preguntas tendenciosas que confundan a los encuestados o los incite a mentir.

El muestreo aleatorio permite una adecuada extracción de un grupo de elementos para que formen parte de una muestra representativa y descriptiva de una población. Los elementos que conforman la muestra dependen de las probabilidades que tienen de ser escogidas; además, todos los elementos tienen las mismas probabilidades de ser seleccionados.

Si en el muestreo aleatorio se realiza un experimento *con reemplazo* en una población finita de tamaño *N*, entonces cada elemento de la población tendrá la misma probabilidad de $\frac{1}{N}$ de ser extraído; por otro lado, si el experimento se realiza *sin reemplazo*, entonces la probabilidad de ser extraído el primer elemento es $\frac{1}{N}$, el segundo elemento $\frac{1}{N-1}$, el tercer elemento $\frac{1}{N-2}$ y así sucesivamente.

En la práctica, regularmente se utilizan dentro del muestreo aleatorio experimentos *sin reemplazo*, debido a que no interesa que se repita un resultado. Una fórmula para conocer la probabilidad de obtener una muestra entre las distintas muestras posibles que se puede extraer de una población, se representa por:

$$P = \frac{1}{{}_N C_n} \qquad \text{Fórmula 6.1}$$

Donde:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

${}_N C_n$ = Número de combinaciones que existe para obtener "*n*" muestra de *n* elementos de una población de "N" elementos.
N = Número de elementos que contiene la población.
n = Número de elementos contenidos en la muestra.

Esto indica la probabilidad de obtener cada una de las distintas muestras que se pueden obtener al combinar los elementos de la población. La fórmula toma en cuenta las *combinaciones* en el denominador. Las distintas combinaciones de selección de los elementos que conforman la muestra es importante, ya que si en un banco hay cinco personas y sólo tres cajas disponibles, la primera persona puede acceder a la caja 1, pero también existe la posibilidad de que se le atienda en las cajas 2 y 3, al igual que a los otros cuatro clientes, por lo que se pueden dar diferentes combinaciones.

Ejemplo 10

El departamento de producción de una empresa de químicos se interesa en saber cuántas muestras distintas de 3 químicos se pueden obtener de una población de 10 químicos.

En este caso la población viene representada por $N = 10$ y la muestra por $n = 3$
Al aplicar la fórmula de combinaciones obtenemos:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

$${}_{10} C_3 = \frac{10!}{3!(10-3)!} = \frac{10 \times 9 \times 8 \times 7!}{(3 \times 2 \times 1)\,(7!)} = \frac{720}{6} = 120$$

La empresa puede obtener 120 muestras distintas y cada muestra está compuesta de 3 químicos. La probabilidad de que cada una de las 120 muestras contenga 3 químicos se define por:

$$\frac{1}{{}_N C_n} = \frac{1}{{}_{10} C_3} = \frac{1}{120} = 0.0083$$

La probabilidad de obtener cada muestra de 3 químicos es 0.0083.

Ejemplo 11

El jefe de personal de una empresa de ropa para niño necesita contratar dos costureras. A la entrevista acuden 8 personas. El jefe de personal quiere conocer cuántas muestras diferentes se pueden obtener y cuál es la probabilidad de obtener cada muestra.

$n = 2 \qquad N = 8$

$${}_N C_n = \frac{N!}{n!(N-n)!} = \frac{8!}{2!(8-2)!} = \frac{8 \times 7 \times 6!}{(2 \times 1)(6!)} = \frac{56}{2} = 28$$

Se pueden obtener 28 muestras distintas donde cada una está integrada por dos costureras. La probabilidad de que cada muestra contenga dos costureras es:

$$\frac{1}{{}_N C_n} = \frac{1}{{}_8 C_2} = \frac{1}{28} = 0.036$$

La probabilidad de obtener cada muestra donde haya dos costureras es 0.036.

6.5.2. Muestreo estratificado

Otra manera que se emplea en forma regular para seleccionar muestras es el muestreo estratificado, este tipo de muestreo retoma la información que se conoce acerca de la población.

El muestreo estratificado puede emplearse principalmente cuando las varianzas de los distintos estratos son diferenciadas.

El muestreo estratificado es un proceso mediante el cual la población es dividida en distintos grupos o estratos, a cada uno de ellos se le extrae una muestra aleatoria que sea proporcional al tamaño de la población de ese estrato y en forma independiente al resto de los estratos. Los elementos contenidos en un mismo estrato deben poseer características similares o comunes entre sí.

El muestreo estratificado puede aplicarse principalmente cuando en una población existen grupos diferenciados.

Algunos ejemplos donde puede aplicarse el muestreo estratificado son los siguientes:

- Las empresas que producen diferentes diseños de automóviles y desean estimar la demanda para las distintas regiones del país.

- Cuando se realiza un estudio para conocer los ingresos que obtienen las personas que se encuentran en distintos puestos y en distintos sectores de la economía.

- Las preferencias en la programación de un canal de televisión entre las distintas generaciones en que se encuentra compuesta la población.

- La opinión pública que la población tiene hacia una política en la Ciudad de México, la cual es desagregada en distintas clases sociales.

Este método es eficiente cuando puede aplicarse una estratificación efectiva, es decir, cuando en la población existen grupos diferenciados, pues de lo contrario el muestreo estratificado no aporta ninguna ventaja sustancial con respecto al muestreo aleatorio y al sistemático.

En este tipo de muestreo, el número de elementos seleccionados que conforman la muestra es proporcional al número de objetos con que cuenta cada estrato.

En términos generales, si se divide una población finita de N elementos en K grupos o estratos de tamaño $N_1, N_2,...,N_K$ y se desea tomar una muestra de tamaño n, se toman K muestras aleatorias de tamaño $n_1, n_2,...,n_K$ en cada estrato, las cuales deben ser proporcionales a la población de cada estrato y si se suman entre sí deben ser igual a n. Para ello se utiliza la siguiente fórmula:

$$n_K = nP_K \quad P_K = \frac{N_K}{N}$$ Fórmula 6.2

Donde:

n_K = Es el número de elementos que componen una muestra del estrato K.
n = Número total de elementos de la muestra elegida.
P_K = Es la proporción de los elementos de la población incluidos en el estrato K.
N_K = Número de elementos que componen el estrato K de la población.
N = Número total de elementos de la población.

La fórmula $n_K = nP_K$ nos indica el número de elementos dentro de una muestra que pertenecen a un estrato, es decir, en qué proporción participa cada uno de los estratos en la composición de la muestra.

Ejemplo 12

Una agencia de publicidad realiza una encuesta a una muestra de 500 profesionistas de una población de 6 000, la cual está compuesta por 3 000 administradores, 1 200 abogados, 600 mercadólogos y 1 200 médicos. Si la asignación es proporcional, ¿cuántos elementos de cada estrato forman parte de la muestra?

La fórmula que se emplea es $n_K = nP_K$ donde $P_K = \frac{N_K}{N}$

Los datos son:

N = 6 000 población
N_1 = 3 000 administradores
N_2 = 1 200 abogados
N_3 = 600 mercadólogos
N_4 = 1 200 médicos
n = 500

$$P_1 = \frac{N_1}{N} = \frac{3\,000}{6\,000} = 0.50 \qquad n_1 = nP_1 = (500)\,(0.50) = 250$$

$$P_2 = \frac{N_2}{N} = \frac{1\,200}{6\,000} = 0.20 \qquad n_2 = nP_2 = (500)\,(0.20) = 100$$

$$P_3 = \frac{N_3}{N} = \frac{600}{6\,000} = 0.10 \qquad n_3 = nP_3 = (500)\,(.10) = 50$$

$$P_4 = \frac{N_4}{N} = \frac{1\,200}{6\,000} = 0.20 \qquad n_4 = nP_4 = (500)\,(.20) = 100$$

Por lo tanto, la muestra elegida estará compuesta por 250 administradores, 100 abogados, 100 médicos y 50 mercadólogos.

Ejemplo 13

Con propósitos de seguridad nacional y de prevenir que México albergue a grupos terroristas, la oficina de migración desea inspeccionar la situación migratoria de los extranjeros de cinco países en la Ciudad de México. Ante la prioridad del tiempo y ante el número limitado de agentes, se decidió seleccionar a 1 000 extranjeros de los que ingresaron de enero a diciembre de 2001:

País	Número
Colombia	5 120
España	2 594
Perú	1 499
Japón	1 100
Federación Rusa	790

Tabla 6.3.

Fuente: Programa Bianual de Mejora Regulatoria 2001-2003 de la Secretaría de Gobernación, www.cofemer.gob.mx

Si la asignación es proporcional, ¿cuántos elementos de cada estrato forman parte de la muestra?

N = 11 103 población
N_1 = 5 120 Colombia
N_2 = 2 594 España
N_3 = 1 499 Perú
N_4 = 1 100 Japón
N_5 = 790 Federación Rusa
n = 1 000

$$P_1 = \frac{N_1}{N} = \frac{5\,120}{11\,103} = 0.461 \qquad n_1 = nP_1 = (1\,000)\,(0.461) = 461$$

$$P_2 = \frac{N_2}{N} = \frac{2\,594}{11\,103} = 0.233 \qquad n_2 = nP_2 = (1\,000)\,(0.233) = 233$$

$$P_3 = \frac{N_3}{N} = \frac{1\,499}{11\,103} = 0.135 \qquad n_3 = nP_3 = (1\,000)\,(0.135) = 135$$

$$P_4 = \frac{N_4}{N} = \frac{1\,100}{11\,103} = 0.099 \qquad n_4 = nP_4 = (1\,000)\,(0.099) = 99$$

$$P_5 = \frac{N_5}{N} = \frac{790}{11\,103} = 0.072 \qquad n_4 = nP_4 = (1\,000)\,(0.072) = 72$$

Por lo tanto, la inspección migratoria se realizaría a 461 colombianos, 233 españoles, 135 peruanos, 99 japoneses y 72 rusos.

6.5.3. Muestreo sistemático

Si la población es excesivamente numerosa, como por ejemplo, la producción de tornillos en una empresa en sus distintas presentaciones, el marco muestral es difícil de obtener, pues ante la inmensidad de datos resultaría demasiado laborioso presentar la lista de los elementos de toda la población. Para solucionar este problema se acude al muestreo sistemático.

El muestreo sistemático es un proceso mediante el cual se elige sistemáticamente un elemento de cada *P* observaciones realizadas en la población.

Mediante este tipo de muestreo se pueden conformar distintas muestras que toman en cuenta a los distintos grupos representados en la población. Una de las ventajas más importantes de este tipo de muestreo es que aparecen elementos de todos los estratos de la población.

Existe una gran variedad de casos prácticos donde resulta más fácil seleccionar a los elementos, por ejemplo, 1 de cada 10 focos para saber si están defectuosos, 1 de cada 20 personas para conocer sus gustos y preferencias o 1 de cada 5 personas que viajan en avión.

Para determinar los datos que conformarán la muestra sistemática, se define k, el cual será el primer dato que es seleccionado de manera aleatoria, y P el cual indica cada cuántos números se realizará un corte; así, $k + P$ será el segundo dato seleccionado, $k + 2P$ el tercer dato seleccionado y así sucesivamente. Éste es un muestreo sistemático por el hecho de que se emplea el dato $k + (i + 1)P$ de los P grupos en que se segmenta la población.

Ejemplo 14

El departamento de servicio a clientes de una empresa registró el número de llamadas que recibe en 25 días hábiles. Las llamadas recibidas son:

Semana 1	
Lu	23
Ma	25
Mi	21
Ju	30
Vi	31

Semana 2	
Lu	31
Ma	42
Mi	43
Ju	28
Vi	27

Semana 3	
Lu	18
Ma	26
Mi	33
Ju	41
Vi	39

Semana 4	
Lu	13
Ma	48
Mi	34
Ju	29
Vi	48

Semana 5	
Lu	12
Ma	15
Mi	19
Ju	46
Vi	25

Tabla 6.4.

El departamento podría seleccionar cierto número de muestras sistemáticas. Por ejemplo, si se desea determinar cinco muestras con el fin de que éstas representen el comportamiento de las llamadas que se realizaron durante estas cinco semanas y que los datos seleccionados no se repitan nuevamente, se procede de la siguiente manera:

En la primera muestra se define $P = 5$ y K inicia con el primer día de los 25 seleccionados y se tomarán elementos de 5 en 5 hasta que se agoten todos los elementos (véase la tabla 6.4).

La primera muestra se eligió de la siguiente manera: se sabe que se inició adoptando el día lunes de la primera semana; el segundo término a ser seleccionado es $K + P$, es decir el dato que corresponde al día $1 + 5 = 6$, el cual es el día lunes de la segunda semana cuyo valor es 31; el tercer término a ser seleccionado es $K + 2P$, es decir, $1 + 2(5) = 11$, o sea el lunes de la tercera semana cuyo valor es 18 y así sucesivamente.

En la muestra dos el valor de K corresponde al segundo día de la primera semana y P continúa con un valor de 5, por lo tanto, el primer elemento es el día martes de la primera semana con un valor de 25, el segundo día a ser seleccionado que se considera en la muestra $(2 + 5)$ se refiere al día martes de la segunda semana con un valor de 42 llamadas y así sucesivamente.

	Muestra 1	**Muestra 2**	**Muestra 3**	**Muestra 4**	**Muestra 5**
K	23	25	21	30	31
$K + P$	31	42	43	28	27
$K + 2P$	18	26	33	41	39
$K + 3P$	13	48	34	29	48
$K + 4P$	12	15	19	46	25

Tabla 6.5. Llamadas recibidas por el departamento de servicio a clientes.

Este tipo de muestreo es útil cuando a partir del número de elementos que contiene una población se desea determinar no una muestra sino varias que tengan características similares, de tal forma que se elija la que muestre un comportamiento más parecido al de la población.

Ejercicio 7

1. Uno de los aspectos que se debe cuidar al elegir una muestra es:

   a) Que estén contenidos el total de los datos de la población.
   b) Que los datos se encuentren organizados mediante un muestreo estratificado.
   c) Que los datos sean de tipo cuantitativo.
   d) Que sea representativa de la población de donde se recolecta.

2. Son algunas ventajas que ofrecen las muestras respecto a un censo:

   a) Son atractivas para hacer inferencias al seleccionar a toda la población.
   b) Son menos costosas, estéticas y representativas de la población.
   c) Los parámetros que emanan de la muestra son utilizados para inferir.
   d) Son menos costosas y sus resultados se obtienen de manera relativa y con rapidez.

3. Es el proceso donde se elige una muestra que sea capaz de representar a la población de manera que no se pierdan los rasgos y las características más relevantes de la población:

   a) Muestreo.
   b) Población infinita.
   c) Población finita.
   d) Censo.

4. Es el proceso en el que se recolecta la información contenida en el total de una población:

   a) Muestreo.
   b) Población infinita.
   c) Población finita.
   d) Censo.

5. Es aquella compuesta de un número determinado de elementos, objetos u observaciones:

   a) Muestreo.
   b) Población infinita.
   c) Población finita.
   d) Censo.

6. Es aquella en que no es posible conocer el número determinado de elementos u objetos que la componen:

   a) Muestreo.
   b) Población infinita.
   c) Población finita.
   d) Censo.

7. Es un ejemplo de una población finita:

   a) El número de las estrellas en el universo.
   b) Los barriles de petróleo que PEMEX exporta cada año.

c) Las compras navideñas en el comercio ambulante de la Ciudad de México.
d) Las sonrisas expresadas durante el año por los niños de un colegio.

8. Es un ejemplo de una población infinita:

a) El número de comerciales transmitidos por una televisora durante un día.
b) El número de libros de texto impresos por la Secretaría de Educación.
c) Las llamadas locales realizadas por tu familia durante un mes.
d) Las "mordidas" realizadas por los agentes de tránsito al año en tu ciudad.

9. Es un proceso mediante el cual se obtiene una muestra de tal manera, que tuvo la misma probabilidad de ser seleccionada que el resto de las muestras posibles que pudieron recolectarse:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

10. Es un proceso en el cual se obtiene una muestra donde cada uno de sus elementos fue recolectado de la misma población y de manera independiente a los demás:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

11. Es un proceso mediante el cual se elige sistemáticamente un elemento de cada *P* observaciones realizadas en la población:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

12. Es un proceso en el que se divide la población en distintos grupos, extrayéndole a cada uno una muestra aleatoria proporcional al tamaño de la población de ese grupo, y en forma independiente, al resto de los grupos:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

13. Un error al realizar un muestreo podría ser:

a) Conocer la opinión sobre el lanzamiento de una revista para el público gay al finalizar un partido de soccer en el estadio Azteca.

b) Realizar una encuesta en una universidad sobre la demanda de libros didácticos a nivel superior.

c) Investigar el precio de introducción de un shampoo contra la calvicie mediante una muestra de hombres mayores a los 25 años.

d) Conocer la calidad del servicio telefónico seleccionando una muestra de los usuarios de este servicio.

14. Son algunos pasos que deben tomarse en cuenta en el proceso de muestreo:

a) Diseñar el muestreo, encontrar el espacio muestral y su distribución.
b) Definir y diseñar la distribución de probabilidades de la población.
c) Definir la población objetivo, diseñar el muestreo y llevarlo a cabo.
d) Llevar a cabo el muestreo mediante una encuesta telefónica.

15. El marco muestral se define como:

a) Es un listado del total de los elementos de la población
b) Es el espacio muestral probabilístico.
c) Son los resultados obtenidos en la muestra.
d) Es el número total de elementos que integran la muestra.

16. El jefe de producción de una empresa selecciona 3 lámparas de un total de 9 para saber si hay alguna que esté defectuosa, ¿cuántas muestras diferentes de tres elementos se pueden obtener? y ¿cuál es la probabilidad de obtener cada muestra?

17. La Secretaría de Fomento Industrial decide tomar una muestra estratificada de mil empresas de una población de 50 000, los estratos están compuestos por 20 000 de tamaño micro, 15 000 pequeñas, 10 000 medianas y 5 000 grandes. ¿Cuál es el tamaño de cada estrato que conforma la muestra?

18. Una empresa produce 100 000 tornillos por día, éstos se dividen en tres estratos, un primer estrato está compuesto por 40 000 tornillos de 1 pulgada, 35 000 tornillos de 1 ½ pulgadas y 25 000 tornillos de 2 pulgadas. Si la asignación es proporcional y selecciona una muestra de 1 400 elementos, ¿de qué tamaño se debe tomar una muestra de cada estrato?

6.6. Distribuciones muestrales y teorema del límite central

En esta sección se presentará el concepto de la distribución muestral, la cual es un caso particular de las distribuciones de probabilidad y uno de los conceptos más importantes en la estadística inferencial. Particularmente se centrará la atención en un modelo conocido como la distribución muestral de la media.

Al inicio de esta unidad se expuso que en una población es posible extraer distintas muestras aleatorias; de hecho, utilizando la fórmula de las combinaciones se obtenía el número exacto de las posibles muestras aleatorias que podían extraerse de una misma población. De ahí que cada muestra tenía una cierta probabilidad de ser recolectada en un proceso de muestreo.

Por otra parte, de las distintas muestras posibles se pueden obtener estadísticos; es decir, medidas descriptivas que utilizan la información contenida en cada una de las muestras posibles de la población. Por ejemplo, en el caso de la empresa que tenía una población de 10 químicos, recuerda que se podían obtener 120 muestras distintas conformadas por 3 elementos. Si a cada una de estas muestras se le calculara un estadístico, por ejemplo su media muestral, se obtendrían 120 medias muestrales para cada una de las muestras. Estas medias muestrales pueden tener distintos valores entre sí y también una cierta probabilidad de ser recolectadas en un proceso de muestreo aleatorio.

Por esta razón, los estadísticos son considerados variables aleatorias, pues toman diferentes valores según la muestra con la que se estimen, es decir, al ser los datos muestrales observaciones de variables aleatorias, los estadísticos también lo son y, por lo tanto, su comportamiento puede ser descrito mediante su distribución de probabilidad, similar a las expuestas en las unidades anteriores.

Una distribución muestral es un caso particular de las distribuciones de probabilidad donde la variable aleatoria es un estadístico muestral. Es decir, al considerar que el valor de los estadísticos varía de muestra en muestra, los estadísticos muestrales tienen una distribución de probabilidad conocida como distribución muestral.

Existe un tipo de distribución muestral particular denominada distribución muestral de la media, la cual es muy importante para llevar a cabo técnicas de estadística inferencial.

Una distribución muestral de la media es una distribución de probabilidad donde la media muestral es la variable aleatoria. Es decir, es una tabla, gráfica o regla que señala la relación de todos los posibles valores que adquiere la variable aleatoria "media muestral", con sus respectivas probabilidades.

La distribución muestral de la media se basa en el hecho de que al tener una población, es posible obtener varias muestras y cada una de ellas tiene una media determinada para encontrar posteriormente la probabilidad de ocurrencia de cada media. Al calcular un promedio de las medias de todas las muestras que fueron obtenidas de una población, se puede inferir el comportamiento que debe tener la media de la población.

Ejemplo 15

Un supervisor tiene seis empleados cuyas experiencias (medidas en años de trabajo) son 2, 3, 4, 6, 7 y 8. El supervisor elige al azar dos de estos empleados y les asigna una nueva tarea.

a) Calcula la media de los años de experiencia de esta población.
b) Tabula las distribuciones muestrales con sus respectivas probabilidades.
c) Calcula la media de la distribución de medias muestrales.

Solución

a) Se calcula la media de la población, la cual se obtiene a partir de:

$$\mu = \frac{\sum x}{N} = \frac{2+3+4+6+7+8}{6} = \frac{30}{6} = 5$$

El resultado muestra que los empleados que constituyen la población tienen en promedio 5 años de experiencia.

b) Para obtener las distribuciones muestrales de la media es necesario construir una tabla en la cual se indique el número de muestras distintas que se pueden construir. Como se tiene una población de $N = 6$ elementos, y una muestra de $n = 2$ elementos, el número de formas existentes para formar muestras de tamaño dos es:

$${}_N C_n = \frac{N!}{n!(N-n)!} = \frac{6!}{2!(6-2)!} = \frac{6 \times 5 \times 4!}{(2 \times 1)(4!)} = \frac{30}{2} = 15$$

El resultado indica que existen quince maneras diferentes para construir una muestra de tamaño dos. Con ello, la siguiente tabla presenta las 15 posibles medias muestrales que se pueden obtener de esta población:

Muestras	Medidas muestrales ($\bar{X}$)
2 y 3	2.5
2 y 4	3
2 y 6	4
2 y 7	4.5
2 y 8	5
3 y 4	3.5
3 y 6	4.5
3 y 7	5
3 y 8	5.5
4 y 6	5
4 y 7	5.5
4 y 8	6
6 y 7	6.5
6 y 8	7
7 y 8	7.5
	$\sum \bar{X} = 75$

La primera muestra indica que seleccionan dos personas con 2 y 3 años de experiencia.

$\frac{2+3}{2} = 2.5$

Tabla 6.6. Medias muestrales de los años de experiencia que tienen los empleados en 15 medias posibles.

Con base en la tabla 6.6. se puede construir una segunda tabla donde podrá apreciarse cómo quedan distribuidas las distintas muestras. En la tabla 6.7. se presenta la distribución muestral para la media de los años de experiencia.

Debido a que se tiene que establecer una distribución muestral es necesario conocer la probabilidad de ocurrencia de los eventos, es decir, la probabilidad de que se genere una media en específico. La primera columna se representa por el valor de la media muestral, cabe mencionar que hay algunos valores de las medias muestrales que se repiten varias veces; por ejemplo: el 4.5 se repite

dos veces, el 5 se repite tres veces y así sucesivamente; en la segunda columna se representan las probabilidades de que sea seleccionada cada media muestral en un proceso de muestreo.

Cada probabilidad se obtiene según el número de veces que se repita un valor de la media muestral y se divide entre el número total de elementos, por ejemplo, para obtener la primera probabilidad se observa que el valor medio 2.5 sólo se da una vez, por lo que este valor se divide entre el total de elementos que es 15 y así sucesivamente.

$\overline{X}$	$P(\overline{X})$
2.5	1/15
3	1/15
3.5	1/15
4	1/15
4.5	2/15
5	3/15
5.5	2/15
6	1/15
6.5	1/15
7	1/15
7.5	1/15
	1

Tabla 6.7. Distribución muestral de los años de experiencia.

c) Ahora bien, es importante determinar los valores para la media de esa distribución muestral de las medias $\mu_{\overline{x}}$, de esa manera se puede tener un comparativo entre los valores de la población y los de la muestra.

La media de la distribución muestral de las medias se obtiene a partir de:

$$\mu_{\overline{X}} = \frac{\sum \overline{X}}{n_{\overline{X}}} = \frac{75}{15} = 5$$

El resultado indica que la media de la distribución muestral de la media es de 5 años en promedio de experiencia. De esta manera se obtiene como conclusión que la media poblacional es igual a la media de la distribución muestral de la media, esto es $\mu = \mu_{\overline{x}}$. A esta propiedad se le denomina *imparcialidad*, la cual será de gran utilidad para realizar inferencia estadística de la media poblacional.

El concepto de imparcialidad se puede formalizar matemáticamente obteniendo el valor esperado de la media muestral. Si se tiene una colección X_1, X_2, ..X_N de variables aleatorias independientes provenientes de una población de cualquier tipo de distribución y con media o valor esperado $E(X) = \mu$, entonces esta media poblacional μ es la misma que el valor esperado de la media muestral $E(\overline{X})$:

$$E(\overline{X}) = E\left[\frac{\Sigma X}{n}\right] = E\left[\frac{X_1 + X_2 + ...X_n}{n}\right] = \frac{1}{n}\{E(X_1) + E(X_2) + ... + E(X_n)\} = \frac{1}{n}(n\mu) = \mu$$

Sin embargo, en el caso de la varianza de la media muestral, su resultado no es el mismo que el valor de la varianza de la población; es decir, no coinciden. Lo anterior se debe a la presencia de un tipo de error conocido como el *error estándar de la media*.

$$V(\bar{X}) = V\left[\frac{\Sigma X}{n}\right] = V\left[\frac{X_1 + X_2 + \ldots X_n}{n}\right] = \frac{1}{n^2}\{V(X_1) + V(X_2) + \ldots + V(X_n)\} = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Es decir, la media muestral $\bar{X}$ tiene un valor esperado igual al parámetro de la población $E(\bar{X}) = \mu$, pero una menor varianza $\sigma^2_{\bar{X}} = V(\bar{X}) = \sigma^2 / n$ que la poblacional σ^2. Por lo tanto, la desviación estándar de una distribución de $\bar{X}$ está dada por $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ conocida como el error estándar de la media.

Un rasgo común de las poblaciones es que poseen características y distribuciones distintas, por lo que los valores de sus parámetros (media, varianza, moda, etc.) también son distintos. Sin embargo, mediante la distribución muestral de la media se facilita el proceso de inferencia estadística de los parámetros, pues no importa qué tipo de distribución tenga la población de la variable $X_1, X_2, \ldots X_N$, la media de la distribución muestral de la media coincide con el valor de la media poblacional μ, pero con una desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ menor a la poblacional σ, conocida como el error estándar de la media.

6.6.1. Teorema central del límite

Como ya se estableció con anterioridad, la media muestral $\bar{X}$ de cada una de las muestras seleccionadas es una variable aleatoria y, como tal, tiene su propia distribución de probabilidad, a lo que suele designarse como "distribución muestral de la media". Se observará que conforme se incremente el tamaño n de cada muestra posible que se extrae de una población de tamaño N, la distribución muestral de la media irá adquiriendo la forma de una distribución normal. En forma más precisa, a este concepto se le conoce como el teorema del límite central, el cual es el teorema más importante de la estadística inferencial.

Como ya se ha mencionado, una población compuesta por $X_1, X_2, \ldots X_N$ con cualquier tipo de distribución, con media μ y varianza σ^2, la distribución muestral de la media tiene una media $E(\bar{X}) = \mu$ que coincide con el valor de la media poblacional μ, y una desviación estándar menor a la poblacional σ, conocida como el error estándar de la media.

De una población compuesta por variables aleatorias independientes $X_1, X_2, \ldots X_N$ con media μ y varianza σ^2, el teorema central del límite señala que la distribución muestral de la media $\bar{X}$ se aproxima a una distribución normal con media μ y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, a medida de que se incremente el número de elementos n que conforman el tamaño de las muestras posibles que se obtienen de la población, no importando el tipo de distribución que tengan $X_1, X_2, \ldots X_N$

Es decir, conforme se incremente el tamaño n de cada muestra posible que se extrae de una población de tamaño N, la distribución muestral de la media irá adquiriendo la forma de una distribución normal, sin importar que la población de la que se extrae no tenga una distribución normal.

La aproximación a la normal de la distribución muestral de la media se cumple si $n \geq 30$ sin importar cuál sea la forma de la población. Si $n < 30$, la aproximación es válida sólo si la población no difiere mucho de una distribución normal, y, cuando la distribución de la población es normal, la distribución muestral de $\bar{X}$ seguirá exactamente una distribución normal, sin importar qué tan pequeño sea el tamaño de las muestras.

Por ejemplo, si se tiene una población de tamaño *N* cuya distribución no es normal, como la de habitantes en México, la cual es sesgada positiva y leptocúrtica, su distribución de frecuencias se representa por la siguiente gráfica:

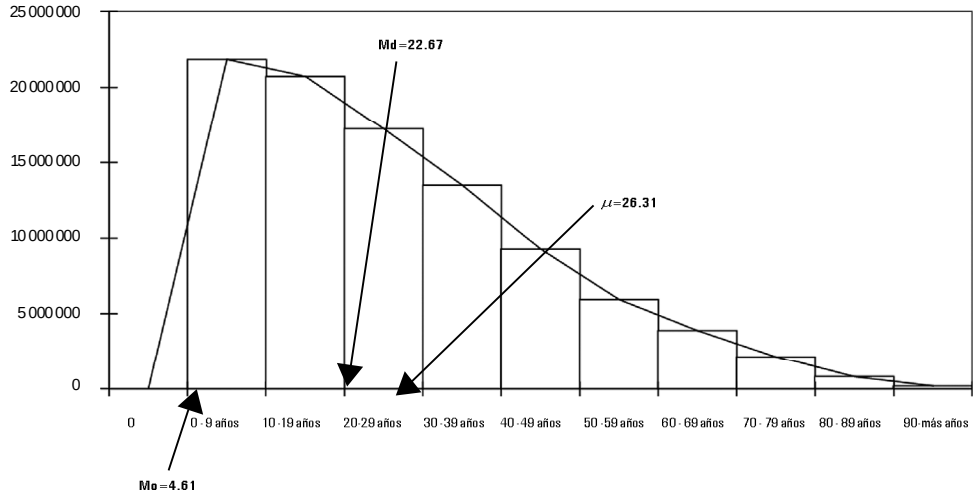


Figura 6.17. Distribución de la población de habitantes en México.

Si se tomaran todas las muestras posibles de tamaño $n \geq 30$, su distribución muestral de la media tendría un distribución normal, es decir, una distribución simétrica cuyos índices de asimetría y de kurtosis serían 0 y 3 respectivamente, con una media $\mu = 26.31$ años.

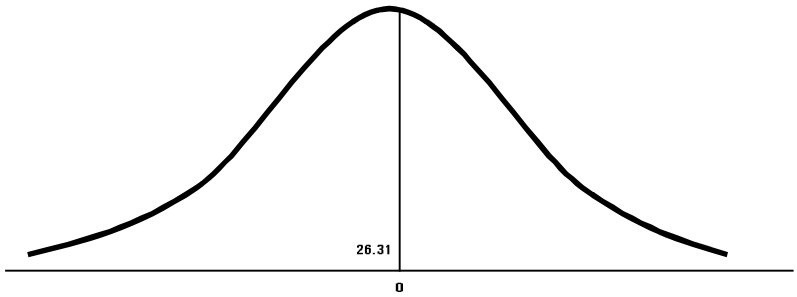


Figura 6.18. Distribución muestral de la media de edades de la población en México.

En general, si $\overline{X}$ es la media de una muestra de tamaño $n \geq 30$, tomada de una población con media μ y desviación estándar σ, su distribución estandarizada es:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \qquad \text{Fórmula 6.3}$$

Es decir, es aproximadamente una normal estandarizada con media uno y varianza igual a cero. Este resultado es de fundamental importancia en estadística, haciendo mayor aún la utilidad de la distribución normal y acrecentando la popularidad de la media aritmética como medida de tendencia central.

El teorema del límite central es de vital relevancia en problemas prácticos, ya que permite dar respuesta a una gran variedad de fenómenos mediante el uso de la curva normal; se aplica a poblaciones infinitas e finitas.

Ejemplo 16

El tiempo de vida útil de cierto tipo de neumático para automóvil tiene una media de 60 000 km y una desviación estándar de 7 500 km, ¿cuál es la probabilidad de que el promedio de una muestra conformada por 100 neumáticos, seleccionada de manera aleatoria, tenga una vida útil mayor de 58 000 km?

Definimos:

μ = 60 000
σ = 7 500
n = 100
$\overline{X}$ = 58 000

Como podemos apreciar, se conoce la desviación estándar poblacional σ, sin embargo, se desconoce el valor de la desviación estándar de la distribución muestral de la media $\sigma_{\overline{X}}$. Por lo tanto, se procede a calcular la $\sigma_{\overline{X}}$, mediante la fórmula del error estándar de la media:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{7\,500}{\sqrt{100}} = 750$$

Ahora se sustituye el valor de $\sigma_{\overline{X}}$ en la fórmula de la normal estandarizada para la distribución muestral de la media, de lo que resulta:

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{58\,000 - 60\,000}{750} = -2.66$$

El valor en tablas para Z = –2.66 es de 0.4961

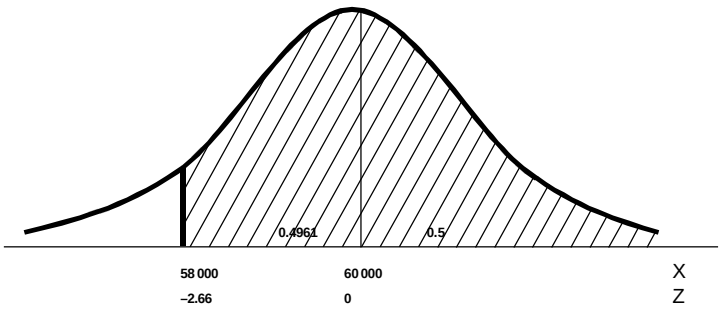


Figura. 6.19. Distribución muestral de la media de la vida útil de neumáticos.

Se tiene que el valor de tablas para la región que va de 0 a –2.66 es de 0.4961, por lo que se procede a sumar este valor a 0.5, de lo cual resulta un valor de 0.9961 que es la probabilidad que nos interesa (véase la figura 6.19), por tanto:

0.5 + 0.4961 = 0.9961

En conclusión, la probabilidad de que el promedio de vida útil de una muestra de 100 neumáticos seleccionada aleatoriamente sea mayor a 58 000 kilómetros es de 0.9961.

Ejemplo 17

Un auditor de un despacho contable toma una muestra de $n = 40$ de una población de 1 100 cuentas por cobrar. El promedio de las cuentas por cobrar de la población viene dado por $\mu = 260$, con una desviación estándar poblacional de $\sigma = 60$, ¿cuál es la probabilidad de que la media muestral $\overline{X}$ sea menor a 240?

Definimos:

$\mu = 260$
$\sigma = 60$
$n = 40$
$\overline{X} = 240$

En primer término, se procede a calcular la $\sigma_{\overline{x}}$ mediante la siguiente fórmula:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{40}} = 9.49$$

Ahora se sustituye el valor de $\sigma_{\overline{X}}$ en la fórmula de la normal estandarizada:

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{240 - 260}{9.49} = -2.1$$

El valor en tablas para $Z = -2.1$ es de 0.4821.
Se tiene que el valor de tablas para la zona que va de 0 a –2.1 es de 0.4821, por lo que se procede a restar este valor a 0.5 de lo cual resulta 0.0179 que es la probabilidad que nos interesa (véase la figura 8.4), por tanto:

0.5 – 0.4821 = 0.0179

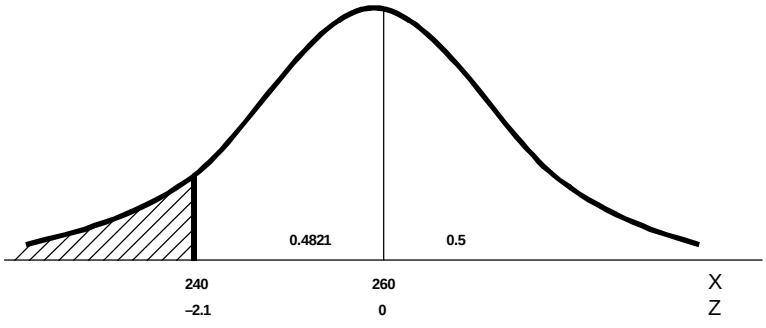


Figura 6.20. Distribución de cuentas por cobrar.

En conclusión la probabilidad de que la media muestral sea inferior a 240 cuentas es del 0.0179.

Ejercicio 8

1. Señala que la media poblacional es igual a la media de la distribución muestral de la media, esto es $\mu - \mu_{\bar{x}}$.

   a) Teorema del límite central.
   b) Distribución muestral.
   c) Imparcialidad.
   d) Distribución muestral de la media.

2. Es un caso particular de las distribuciones de probabilidad donde la variable aleatoria es un estadístico muestral:

   a) Teorema del límite central.
   b) Distribución muestral.
   c) Imparcialidad.
   d) Distribución muestral de la media.

3. Es una distribución de probabilidad donde la media muestral es la variable aleatoria:

   a) Teorema del límite central.
   b) Distribución muestral.
   c) Imparcialidad.
   d) Distribución muestral de la media.

4. Si se compara la desviación estándar de la media muestral con la desviación estándar de la población, la desviación estándar de la media muestral es:

   a) Menor a la desviación estándar de la población.
   b) Mayor a la desviación estándar de la población.
   c) Igual a la desviación estándar de la población.
   d) Imparcial a la desviación estándar de la población.

5. El teorema de límite central señala que al incrementar el tamaño n de cada muestra posible que se extrae de una población de tamaño N, la distribución muestral de la media irá adquiriendo la forma de una distribución normal:

   a) Siempre que la población de donde se extrae la muestra tenga una distribución normal.
   b) Siempre que las desviaciones estándar de las muestras sean menores a las desviaciones estándar de la población.
   c) Siempre que las desviaciones estándar de las muestras sean iguales a las desviaciones estándar de la población.
   d) Sin importar que la población de donde se extrae la muestra no tenga una distribución normal.

6. Una distribuidora de automóviles tomó una muestra de 2 vendedores de una población de 5, cuyas ventas diarias de automóviles son 4, 5, 7, 6 y 3, respectivamente:

   a) ¿Cuál es la media y la desviación estándar de la población?
   b) ¿Cuántas muestras se pueden obtener?
   c) ¿Cuál es la distribución de probabilidades de las medias muestrales?

d) ¿Cuál es la media?
e) ¿Qué valor toma el error estándar de la media?

7. La duración de cierto tipo de focos tiene una distribución aproximadamente normal, con una media igual a 1 200 hrs. y una desviación estándar de 50hrs., ¿cuál es la probabilidad de que una muestra aleatoria de 16 focos tenga una vida promedio de más de 1 220 hrs.?

8. Un vendedor contacta telefónicamente con clientes potenciales para estudiar si merece la pena una visita a domicilio. Su experiencia le indica que en promedio 40 de sus contactos por teléfono vienen seguidos de una visita a domicilio y se tiene una desviación estándar de 50. Si selecciona una muestra de 100 personas contactadas por teléfono, ¿cuál es la probabilidad de que se realicen entre 45 y 50 visitas como resultado?

9. Una oficina de defensa del consumidor recibe en promedio 25 llamadas por día con una desviación estándar de 40. Si selecciona una muestra de 64, calcula la probabilidad de que el promedio de llamadas recibidas en un día esté entre 20 y 30.

10. Un hospital encuentra que en promedio 20 facturas tienen un retraso de un mes con una desviación estándar de 4 facturas. Si se selecciona aleatoriamente una muestra conformada por 36 facturas, ¿cuál es la probabilidad de que menos de 20 facturas tengan retraso de un mes?

6.7. Aplicaciones para el caso de estudios de mercado

El uso de distribuciones de probabilidad dentro de la actividad económica y de negocios es muy común y coadyuva en gran medida a poder tomar decisiones con mayor efectividad, en este apartado podremos ver algunos ejemplos de ello.

1. Un asesor financiero ha sugerido a una institución bancaria se otorgue un incentivo equivalente a 10% sobre su saldo diario promedio como premio a la constancia de sus ahorradores, con el propósito de incrementar el volumen de captación de efectivo.

   Para aprobar esta propuesta el director del área de ahorros solicitó información sobre el saldo promedio diario de los ahorradores. Las cifras presentadas al director estiman que éste es de $15 000 y que la desviación estándar es de $4 000.

   a) ¿Qué porcentaje de cuentas se esperaría que tuvieran un saldo promedio diario mayor a $18 000?

   b) Para asignar el incentivo al ahorro los clientes deberán incrementar y sostener su nivel de ahorro por un tiempo preestablecido, ¿cuál debe ser el saldo diario promedio para que a lo más 20% de las cuentas que tengan saldo menor al estipulado se beneficien con el incentivo?

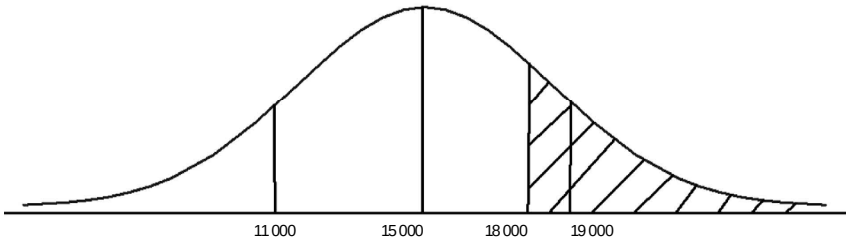


Figura 6.21.

Solución

a) En la gráfica podemos identificar los valores para μ = 15 000 donde el área sombreada representa a todos los clientes que pueden mantener un saldo diario promedio mayor a $18 000.

Datos:

X = 18 000
μ = 15 000
σ = 4 000
P(X > 18 000) = ?

Sustituyendo en la fórmula se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{18\,000 - 15\,000}{4\,000} = 0.75$$

El valor de tablas para Z = 0.75 es 0.2734. Se requiere conocer la probabilidad de que Z sea superior a 0.75, por lo tanto a 0.5 hay que restarle 0.2734.

0.5 – 0.2734 = 0.2266

Por lo tanto, podemos afirmar que 22.66% de los ahorradores mantienen un saldo diario promedio mayor a $18 000.

b) Mediante las tablas de distribución normal estándar determinamos el valor que debe tener Z para que los valores menores a él representen un área de 0.2000 (el valor que más se aproxime al área buscada) para nuestro caso Z = –0.84.

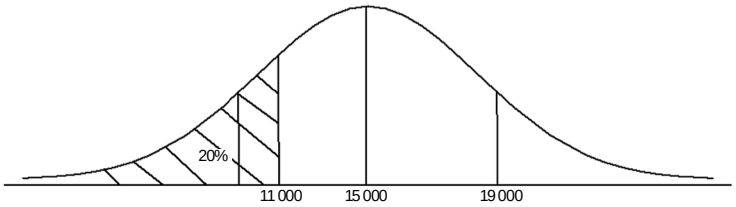


Figura 6.22.

$$Z = \frac{X - \mu}{\sigma} = -0.84$$

$X - \mu = -0.84\,\sigma$
$X = \mu - 0.84\,\sigma$
$X = 15\,000 - 0.84\,(4\,000) = 11\,640$

Con base en lo anterior determinamos que el saldo diario promedio mínimo que se debe tener para obtener este incentivo y sólo otorgarlo a 20% de los ahorradores es de $11 640.

2. Los datos arrojados por un estudio de mercado solicitado por una pequeña empresa embotelladora de agua potable muestran que el consumo promedio diario por persona de agua purificada embotellada es de 1.5 litros diarios. Suponiendo que el consumo de agua potable tiene una distribución normal y que la desviación estándar es de 0.5 litros determina:

a) ¿Cuál es la probabilidad de que una persona consuma más de 2.5 litros de agua embotellada diariamente?
b) ¿Cuál es la probabilidad de que el consumo de una persona esté entre 1 y 3 litros diarios?

Solución

a) Debemos determinar la probabilidad de que el consumo diario por persona de agua purificada embotellada sea mayor a 2.5 litros, la información con la que contamos es la siguiente:

$\mu = 1.5$
$\sigma = 0.5$
$X = 2.5$
$P(X > 2.5) = ?$

Sabemos que $Z = \frac{X - \mu}{\sigma} = \frac{2.5 - 1.5}{0.5} = \frac{1}{0.5} = 2$

Buscando el valor para Z = 2 en la tabla de distribución normal estándar obtenemos un valor de 0.4772.

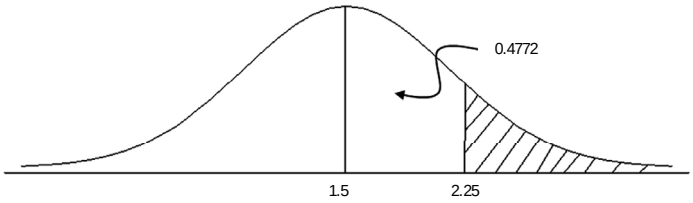


Figura 6.23.

De acuerdo con la gráfica el área sombreada es de nuestro interés y su valor es:

$P(Z > 2) = 0.5 - 0.4772 = 0.0228$

Por lo tanto, la probabilidad de que el consumo promedio diario por persona de agua purificada embotellada sea mayor a 2.5 litros es de 0.0228 o 2.28%.

b) El objetivo es determinar la probabilidad de que las personas consuman entre 1 y 3 litros de agua embotellada diariamente.

La información con la que contamos es la siguiente:

$\mu = 1.5$
$\sigma = 0.5$
$X_1 = 1$
$X_2 = 3$
$P(1 < X < 3) = ?$

Gráficamente podemos apreciar que el área en cuestión es la que se delimita entre 1 y 3.

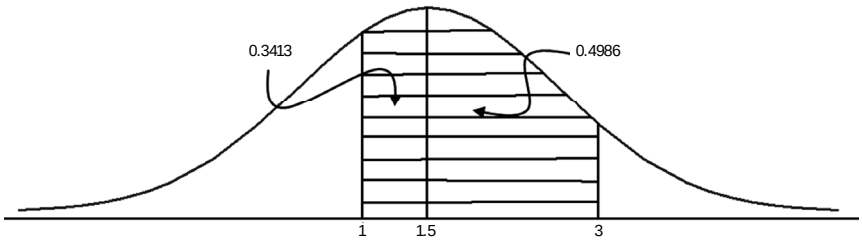


Figura 6.24.

Procedemos a calcular los valores para Z

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{1 - 1.5}{0.5} = \frac{-0.5}{0.5} = -1$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{3 - 1.5}{0.5} = \frac{1.5}{0.5} = 3$$

Los valores para Z_1 y Z_2 obtenidos en la tabla de distribución estándar son:

$Z_1 = 0.3413$
$Z_2 = 0.4986$

Luego entonces: $P(1 < X < 3) = P(-1 < Z < 3) = 0.3413 + 0.4986 = 0.8399$

Por lo tanto, la probabilidad de que el consumo diario de agua embotellada por persona sea de 1 a 3 litros es de 83.99%.

3. Una microempresa que comercializa productos de limpieza necesita reajustar su volumen de ventas de jabón en polvo para lo cual se realizó un muestreo aleatorio de 25 clientes obteniendo que el promedio de compra fue de 5 kg con una desviación estándar de 0.5 kg.

   El empresario necesita estimar la cantidad promedio real que consumen sus clientes teniendo 95% de confianza de que el intervalo obtenido incluye el promedio real de la población, ¿cuál es el valor de $t_{\alpha/2}$ y los grados de libertad correspondientes?

Solución

Como el nivel de confianza es 98%, entonces $1 - \alpha = 95\% = 0.95$ por lo tanto

$$\alpha = 1 - 0.95 = 0.05 \Rightarrow \alpha/2 = 0.025$$

Los grados de libertad son n – 1 sustituyendo: 25 – 1 = 24

Buscando en la tabla de distribución *t* con nivel de significancia de 0.025 y 24 como gl (grado de libertad), se tiene: $t_{\alpha/2} = 2.0639$

4. En un estudio de mercado que se realizó para incorporarlo en un plan de negocio para una empresa de soporte técnico de computadoras en la zona norte de la ciudad se encontró que 35% de los hogares encuestados cuentan con equipo de cómputo propio. Si consideramos una muestra de 80 hogares:

   a) ¿Cuál es la probabilidad de que más de 40 hogares cuenten con un equipo de cómputo propio?

   b) ¿Cuál es la probabilidad de que menos de 35 hogares cuenten con un equipo de cómputo propio?

Solución

En apariencia este problema se puede resolver por medio de la distribución binomial, sin embargo, el tamaño de la muestra es muy grande y esto complica fuertemente los cálculos. Para dar solución aplicaremos la distribución normal como una aproximación a la binomial, aplicando el factor de corrección apropiado.

   a) ¿Cuál es la probabilidad de que más de 40 hogares cuenten con un equipo de cómputo propio?

Datos:

$p = 0.35$
$q = 0.65$
$n = 80$
$X = 40$
$P(X > 40) = ?$

$\mu = np = 80\ (0.35) = 28$ $\qquad \sigma = \sqrt{npq} = \sqrt{80(0.35)(0.65)} = \sqrt{18.2} = 4.27$

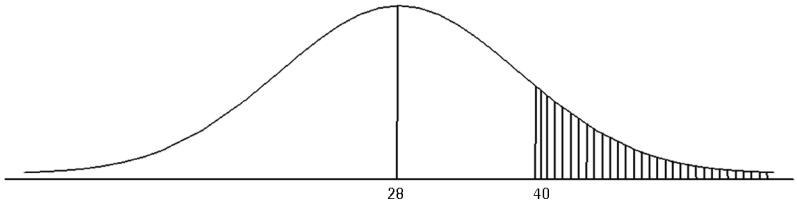


Figura 6.25.

Como se pide que más de 40 hogares cuenten con equipo de cómputo, para localizar el valor de Z se le suma 0.5 a 40, por lo tanto X = 40 + 0.5 =40.5, entonces:

$$Z = \frac{X - \mu}{\sigma} = \frac{40.5 - 28}{4.27} = 2.93$$

Buscando el valor en tablas para el área mayor a Z = 2.93 es 0.0017

Por lo tanto, la probabilidad de que más de 40 hogares tengan equipo de cómputo propio es de 0.17%.

b) ¿Cuál es la probabilidad de que menos de 35 hogares cuenten con un equipo de cómputo propio?

En este caso se solicita la probabilidad de que de la misma muestra, menos de 35 hogares, cuenten con equipo de cómputo.

$p = 0.35$
$q = 0.65$
$n = 80$
$X = 30$
$P(X < 35) = ?$

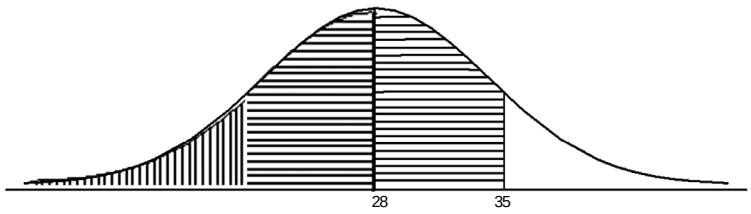


Figura 6.26.

Sabemos que:

$\mu = 28$ y $\sigma = 4.27$

Como se requiere X < 35, se resta 0.5 a 35, por tanto X = 35 – 0.5 = 34.5

Ahora procedemos a estimar el valor de Z para X = 34.5

$$Z = \frac{X - \mu}{\sigma} = \frac{34.5 - 28}{4.27} = 1.52$$

Buscando el valor en tablas para Z = 1.52 es 0.4357

P(X > 35) = 0.5 + 0.4357 = 0.9357

Por lo tanto, la probabilidad de que menos de 35 hogares tengan equipo de cómputo propio es de 93.57%.

Ejercicios

1. El director del área de inversiones de un banco solicitó información sobre el saldo promedio diario de los inversionistas. Las cifras presentadas al director estiman que el saldo promedio diario de sus clientes es de $15 000 y que la desviación estándar es de $1 500 (redondea el valor de Z a 2 decimales).

   a) ¿Cuál es la probabilidad de que un nuevo inversionista invierta con un saldo superior a $17 000?

   b) ¿Cuál es la probabilidad de que un nuevo inversionista inicie con un saldo menor a $14 000?

Soluciones

a) 0.0918; 9.18%

b) 0.2514; 25.14%

2. Un negocio de pizzas vende los fines de semana pizzas tamaño familiar a un precio de $80, si por cada pedido que levanta de sus clientes se adicionan otros productos como refrescos de lata, postres y sopas, en promedio sus ventas son de $150 por familia con una desviación estándar de $25.

   El dueño del negocio planea ofrecer un combo que incluye 1 pizza familiar, 4 sopas y 4 refrescos por $170.

   a) ¿Qué probabilidad existe de que los clientes compren este nuevo paquete?

   b) ¿Cuál es precio al que se debe ofrecer el paquete para que 60% de los clientes prefieran el nuevo combo?

Soluciones

a) 0.2119; 21.19%
b) $143.75

Autoevaluación

1. Una variable aleatoria continua se define como:

   a) Aquella que toma solamente valores enteros.
   b) Aquella que toma valores dentro de un intervalo.
   c) Aquella que toma números irracionales.
   d) Aquella en la cual los valores no se pueden fraccionar.

2. Un ejemplo de variable continua es:

   a) El número de artículos defectuosos en un lote de producción.
   b) El número de alumnos en una universidad.
   c) El tiempo que tarda la fabricación de un avión.
   d) El número de nacimientos en el Distrito Federal.

3. La distribución de probabilidad normal se aplica a muestras cuyo número de observaciones es:

   a) $n \leq 20$
   b) $n > 10$
   c) $n \geq 50$
   d) $n \geq 30$

4. Son características de la distribución normal:

   a) Es asimétrica, su forma es acampanada y se extiende de $-\infty$ a $+\infty$.
   b) El área bajo la curva normal se considera al 100%, su forma es acampanada y se extiende de 0 a $+\infty$.
   c) Se extiende de $-\infty$ a 0, es acampanada y asimétrica.
   d) Es simétrica, se extiende de $-\infty$ a $+\infty$ y es acampanada.

5. La fórmula de la normal estandarizada mide:

   a) La diferencia entre el valor de X y la σ
   b) La diferencia entre la μ y la σ
   c) La diferencia entre X y la σ
   d) La diferencia entre el valor específico de X y la μ en términos de la σ

6. Si X = 20 $\mu = 16$ y $\sigma = 2$, el valor de "Z" es:

   a) 1
   b) 5
   c) 3
   d) 2

7. Cuando Z = 0.4 el valor de tablas es:

   a) 0.1591
   b) 0.1179
   c) 0.1554
   d) 0.2531

8. Cuando Z = 2.35 el valor de tablas es:

   a) 0.4904
   b) 0.4906
   c) 0.4984
   d) 0.4750

9. La fórmula para calcular la desviación estándar en un experimento cuando la binomial se aproxima a la normal es:

   a) np

   b) $\sqrt{npq}$

   c) $\sqrt{n\mu q}$

   d) $\sqrt{n\mu p}$

10. Una diferencia importante entre la distribución normal y la distribución t es que:

    a) La primera mide variables discretas mientras que la segunda mide variables continuas.
    b) Para la normal la curva tiene forma simétrica y para la distribución t asimétrica.
    c) Una se utiliza en muestras grandes y la otra en muestras pequeñas.
    d) Una se calcula con una fórmula estandarizada y la otra no.

11. Los grados de libertad cuando se analiza una sola variable se obtienen a través de:

    a) $n - k$
    b) $n - 2$
    c) $n - 1$
    d) $n - \infty$

12. La distribución exponencial se utiliza cuando se trabaja con variables como:

    a) Tiempo y distancia.
    b) Tiempo y natalidad.
    c) Productividad y número de empleados.
    d) Ingreso y población.

13. Si en un intervalo todos los eventos tienen la misma probabilidad de ocurrir se está haciendo referencia a la distribución:

    a) Discreta.
    b) Exponencial.
    c) Uniforme.
    d) Normal.

14. Si $\alpha = 0.05$ y $n - 1 = 15$ el valor de t_α es:

    a) 1.7531
    b) 1.7613

c) 2.1315
d) 1.3406

15. La distribución t se aplica a muestras donde:

a) $n \geq 20$
b) $n < 30$
c) $n \leq 10$
d) $n > 30$

16. El tipo de muestreo que subdivide en grupos a una población es:

a) Muestreo estratificado.
b) Muestreo sistemático.
c) Muestreo aleatorio.
d) Muestreo por conglomerados.

17. En este tipo de muestreo todas las muestras tienen la misma probabilidad de ser elegidas:

a) Muestreo sistemático.
b) Muestreo estratificado.
c) Muestreo aleatorio.
d) Muestreo por conglomerados.

18. El teorema del límite central establece que, no importando el tipo de distribución que tenga la población, la distribución muestral de la media será normal, siempre y cuando el tamaño de la población sea:

a) $n < 30$
b) $n > 50$
c) $n \geq 30$
d) $n \leq 30$

19. Si la varianza de una población es $\sigma^2 = 45$ y el tamaño de la muestra $n = 5$, la desviación estándar de la media muestral es aproximada a:

a) 9
b) 8.2
c) 3
d) 20.1

20. El teorema del límite central señala que, no importando la distribución de la población, si el tamaño de la muestra se incrementa:

a) La distribución de la media muestral se aproxima a una normal.
b) La distribución de la población es normal y la de la muestra no.
c) Que la distribución muestral nunca será muestral.
d) Que mientras más pequeña se "n" más se aproxima la media muestral a tener un comportamiento como la normal.

21. Si $\overline{X} = 10$ $\mu_{\overline{X}} = 20$ $\sigma_{\overline{X}} = 5$, el valor de Z es:

a) – 4
b) 2.5
c) 5
d) – 2

22. La fórmula que se emplea para obtener la probabilidad de la distribución muestral con aproximación normal es:

a) $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

b) $Z = \frac{\overline{X} - \mu}{\sigma^2}$

c) $Z = \frac{\overline{X} - \mu}{\sigma^2}$

d) $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$

23. Si Z = 1.39, el valor de tablas es:

a) 0.4162
b) 0.4319
c) 0.4175
d) 0.4177

24. La inferencia estadística tiene como objetivo:

a) Analizar a una muestra.
b) Buscar valores representativos de una población.
c) Deducir una serie de conclusiones a partir de una muestra.
d) Analizar la distribución de una población.

25. Si se tiene una población de tamaño $N = 10$ y se desea seleccionar una muestra de $n = 2$, el número de posibles muestras que se pueden obtener de esta población es:

a) 40
b) 45
c) 30
d) 25

26. Un parámetro es:

a) Un valor numérico que sirve para resumir la totalidad de los elementos que forman parte de una población.
b) El valor que se utiliza para describir una muestra.

c) Un valor cualitativo.
d) Un valor tabulado.

27. Señala que la media poblacional es igual a la media de la distribución muestral de la media, esto es $\mu = \mu_{\bar{x}}$.

a) Teorema del límite central.
b) Distribución muestral.
c) Imparcialidad.
d) Distribución muestral de la media.

28. Es un caso particular de las distribuciones de probabilidad donde la variable aleatoria es un estadístico muestral:

a) Teorema del límite central.
b) Distribución muestral.
c) Imparcialidad.
d) Distribución muestral de la media.

29. Es una distribución de probabilidad donde la media muestral es la variable aleatoria:

a) Teorema del límite central.
b) Distribución muestral.
c) Imparcialidad.
d) Distribución muestral de la media.

30. El marco muestral se define como:

a) Es un listado del total de los elementos de la población.
b) Es el espacio muestral probabilística.
c) Son los resultados obtenidos en la muestra.
d) Es el número total de elementos que integran a la muestra.

31. Es un proceso mediante el cual se elige sistemáticamente un elemento de cada *F* observaciones realizadas en la población:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

32. Es un proceso mediante el cual se obtiene una muestra de tal manera que tuvo la misma probabilidad de ser seleccionada que el resto de las muestras posibles que pudieron recolectarse:

a) Muestreo aleatorio simple para poblaciones finitas.
b) Muestreo estratificado.
c) Muestreo aleatorio simple para poblaciones infinitas.
d) Muestreo sistemático.

EJERCICIOS

33. Es un proceso en el cual se obtiene una muestra donde cada uno de sus elementos fue recolectado aleatoriamente de la misma población y de manera independiente a los demás:

    a) Muestreo aleatorio simple para poblaciones finitas.
    b) Muestreo estratificado.
    c) Muestreo aleatorio simple para poblaciones infinitas.
    d) Muestreo sistemático.

34. Es el proceso en el que se recolecta la información contenida en el total de una población:

    a) Muestreo.
    b) Población infinita.
    c) Población finita.
    d) Censo.

35. Es aquella compuesta de un número determinado de elementos, objetos u observaciones:

    a) Muestreo.
    b) Población infinita.
    c) Población finita.
    d) Censo.

36. Es aquella en la que no es posible conocer el número determinado de elementos u objetos que la componen:

    a) Muestreo.
    b) Población infinita.
    c) Población finita.
    d) Censo.

37. Uno de los aspectos que se debe cuidar al elegir una muestra es:

    a) Que estén contenidos el total de los datos de la población.
    b) Que se encuentren organizados mediante un muestreo estratificado.
    c) Que los datos sean de tipo cuantitativo.
    d) Que sea representativa de la población de donde se recolecta.

38. Algunas ventajas que ofrecen las muestras respecto a un censo:

    a) Son atractivas para hacer inferencias al seleccionar a toda la población.
    b) Son menos costosas, estéticas y representativas de la población.
    c) Los parámetros que emanan de la muestra son utilizados para inferir.
    d) Son menos costosas y sus resultados se obtienen con relativa rapidez.

39. Es el proceso donde se elige una muestra que sea capaz de representar a la población de manera que no se pierdan los rasgos y las características más relevantes de la población:

    a) Muestreo.
    b) Población infinita.

c) Población finita.
d) Censo.

40. Si se compara la desviación estándar de la media muestral con la desviación estándar de la población, la desviación estándar de la media muestral es:

a) Menor a la desviación estándar de la población.
b) Mayor a la desviación estándar de la población.
c) Igual a la desviación estándar de la población.
d) Imparcial a la desviación estándar de la población.

Respuestas a los ejercicios

Ejercicio 1

1. a)
2. d)

Ejercicio 2

1. a) Los datos son:

$\mu = 27.8$
$\sigma = 4$
$X = 25$
$P(X < 25)$

Sustituyendo en la fórmula se obtiene,

$$Z = \frac{X - \mu}{\sigma} = \frac{25 - 27.8}{4} = -\frac{2.8}{4} = -0.7$$

El valor de tablas para Z = 0.7 es de 0.2580

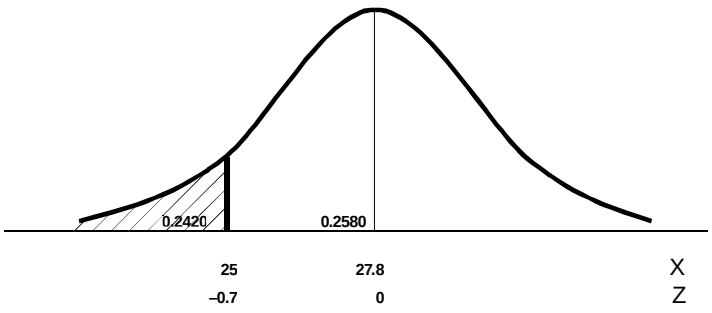


Figura 6.27. Distribución del tiempo de ensamblado de automóviles.

El valor de tablas para la zona que va de 0 a –0.7 es de 0.2580; como ya se había argumentado anteriormente, el valor para cada una de las mitades de la curva normal es de 0.5. Si a 0.5 se le resta 0.2580 nos queda 0.2420.

De lo anterior concluimos que la probabilidad de que el automóvil se pueda ensamblar en menos de 25 minutos es 0.2420 o 24.20%.

b) Los datos son:

$\mu = 27.8$ min.
$\sigma = 4$ min.
$X_1 = 26$ min.
$X_2 = 30$ min.
$P(26 < X < 30)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{26 - 27.8}{4} = \frac{1.8}{4} = -0.45$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{30 - 27.8}{4} = \frac{2.2}{4} = 0.55$$

El valor de tablas para $Z_1 = -0.45$ es 0.1736
El valor de tablas para $Z_2 = 0.55$ es 0.2088

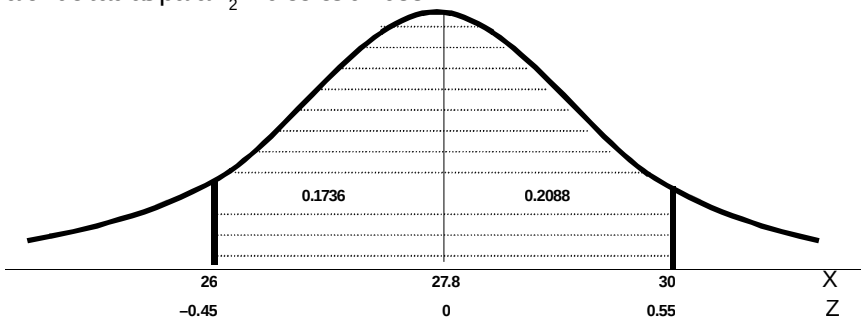


Figura 6.28. Distribución del tiempo de ensamblado de automóviles.

En conclusión, la probabilidad de que el automóvil se pueda ensamblar entre 26 y 30 minutos es de 0.1736 + 0.2088 = 0.3824 o el 38.24%.

2. Los datos son:

μ = \$20
σ = \$2
X_1 = \$18
X_2 = \$23
$P(18 < X < 23)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{18 - 20}{2} = -\frac{2}{2} = -1 \quad Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{23 - 20}{2} = \frac{3}{2} = 1.5$$

El valor de tablas para $Z_1 = 1$ es 0.3413
El valor de tablas para $Z_2 = 1.5$ es 0.4332

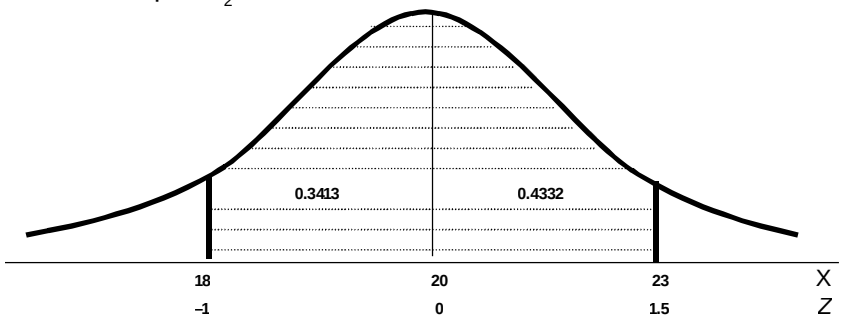


Figura 6.29. Distribución de los salarios pagados por una empresa.

En conclusión, el porcentaje de trabajadores que reciben salarios entre \$18 y \$23 es de 0.3413 + 0.4332 = 0.7745 o 77.45%.

3. Los datos son:

$\mu = 2\,000$
$\sigma = 200$
$X_1 = 2\,000$
$X_2 = 2\,400$
$P(2\,000 < X < 2\,400)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{2\,000 - 2\,000}{200} = -\frac{0}{200} = 0 \quad Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{2\,400 - 2\,000}{200} = \frac{400}{200} = 2$$

El valor de tablas para $Z_1 = 0$ es 0
El valor de tablas para $Z_2 = 2$ es 0.4772

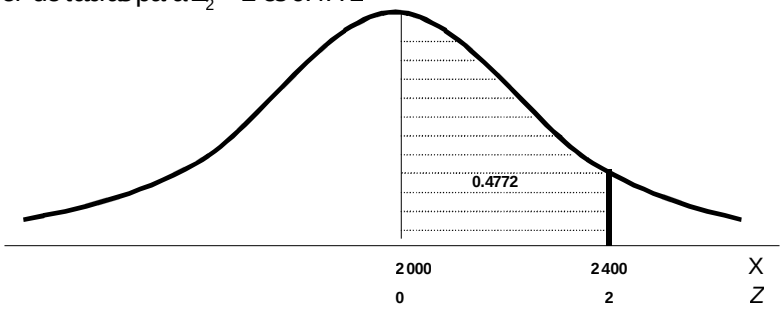


Figura 6.30. Distribución de la duración en horas de un componente eléctrico.

En conclusión, el porcentaje de componentes eléctricos que dura entre 2 000 y 2 400 horas es de 0 + 0.4772 = 0.4772 o 47.72%.

4. Los datos son:

$\mu = 1\,200$
$\sigma = 100$
$X = 1\,000$
$P(X > 1\,000)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{1\,000 - 1\,200}{100} = -\frac{200}{100} = -2$$

El valor de tablas para $Z = -2$ es 0.4772

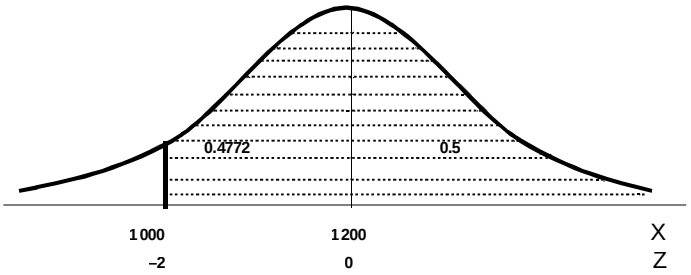


Figura 6.31. Distribución de las unidades demandadas de un producto.

En conclusión, el porcentaje de unidades de producto demandadas superiores a 1 000 es de 0.4772 + 0.5 = 0.9772 o 97.72%.

5. a) Los datos son:

$\mu = 75$
$\sigma = 20$
$X = 60$
$P(X < 60)$

Sustituyendo en la fórmula se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{60 - 75}{20} = -\frac{15}{20} = -0.75$$

El valor de tablas para $Z = 0.75$ es de 0.2734

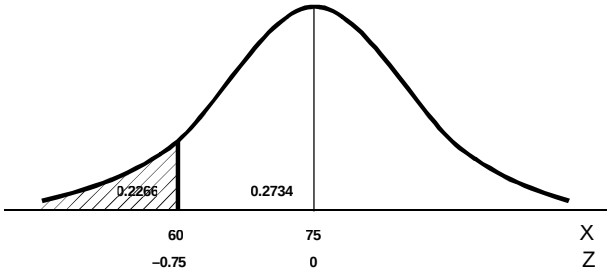


Figura 6.32. Distribución del tiempo de reparación de una fotocopiadora.

De lo anterior se concluye que la proporción de fotocopiadoras que son reparadas en menos de 60 minutos es 0.2266 o de 22.66%.

b) Los datos son:
$\mu = 75$
$\sigma = 20$
$X = 90$
$P(X < 90)$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{90 - 75}{20} = \frac{15}{20} = 0.75$$

El valor de tablas para $Z_1 = 0.75$ es 0.2734

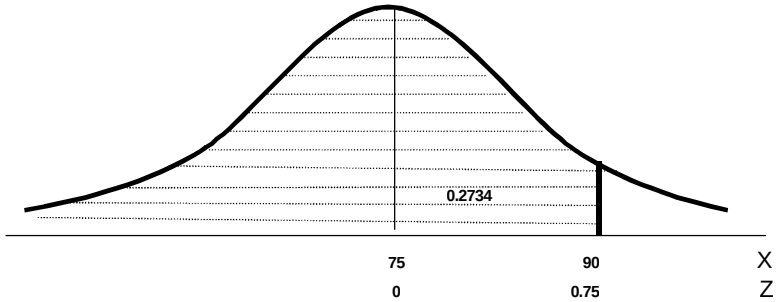


Figura 6.33. Distribución del tiempo de reparación de una fotocopiadora.

De lo anterior se concluye que la proporción de fotocopiadoras que son reparadas en menos de 90 minutos es 0.7734 o de 77.34%.

Ejercicio 3

1. a) Datos:

$p = 0.75$
$q = 0.25$
$n = 80$
$X = 50$
$P(X \geq 50)$

$\mu = np = (80)\ (0.75) = 60 \qquad \sigma = \sqrt{npq} = \sqrt{(80)\ (0.75)\ (0.25)} = \sqrt{15} = 3.82$

Sustituyendo en la fórmula de la normal se obtiene:

$X = 50 - 0.5 = 49.5$

$$Z = \frac{X - \mu}{\sigma} = \frac{49.5 - 60}{3.872} = -\frac{10.5}{3.872} = -2.71$$

El valor de tablas para $Z = -2.71$ es 0.4966. Como el problema está pidiendo *al menos* se utiliza el símbolo $\geq$, de esta manera a 0.4966 se le suma 0.5000 quedando el área bajo la curva igual a 0.9660 (véase la figura 6.34).

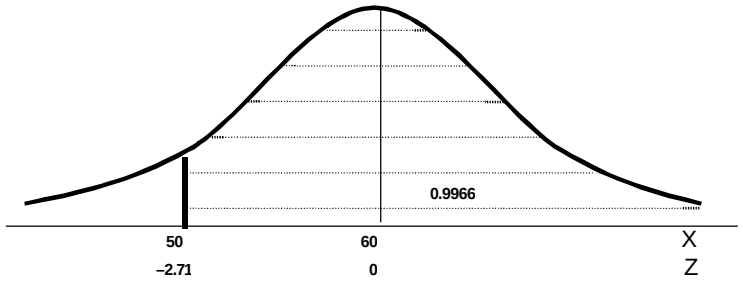


Figura 6.34. Distribución de la introducción de un nuevo detergente.

La probabilidad de que al menos 50 mujeres sean de la misma opinión es de 96.60%.

b) Datos:

$p = 0.75$
$q = 0.25$
$n = 80$
$X = 56$
$P(X > 56)$

$\mu = np = (80)\ (0.75) = 60 \qquad \sigma = \sqrt{npq} = \sqrt{(80)\ (0.75)\ (0.25)} = \sqrt{15} = 3.82$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{56.5 - 60}{3.872} = -\frac{10.5}{3.872} = -2.90$$

El valor de tablas para Z = –2.90 es 0.3159. Si a 0.5000 se le suma 0.3159 el área bajo la curva igual a 0.8159 (véase la figura 6.13).

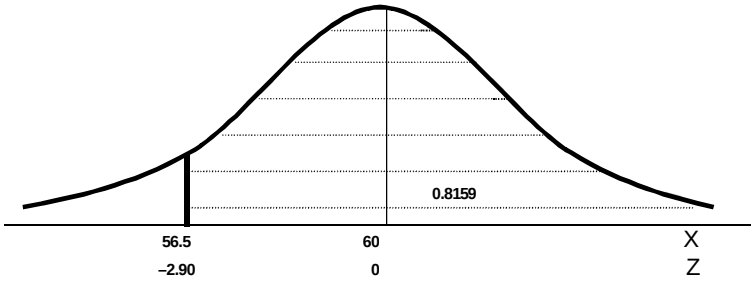


Figura 6.35. Distribución de la introducción de un nuevo detergente.

La probabilidad de que más de 56 mujeres tengan la misma opinión es de 81.59%.

2. Datos:

$p = 0.10$
$q = 0.90$
$n = 70$
$X = 5$
$P(X > 5)$

$$\mu = np = (70)\ (0.10) = 7 \qquad \sigma = \sqrt{npq} = \sqrt{(70)\ (0.10)\ (0.90)} = \sqrt{6.3} = 2.51$$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{5.5 - 7}{2.51} = -\frac{1.5}{2.51} = -0.6$$

El valor de tablas para Z = 0.6 es 0.2743

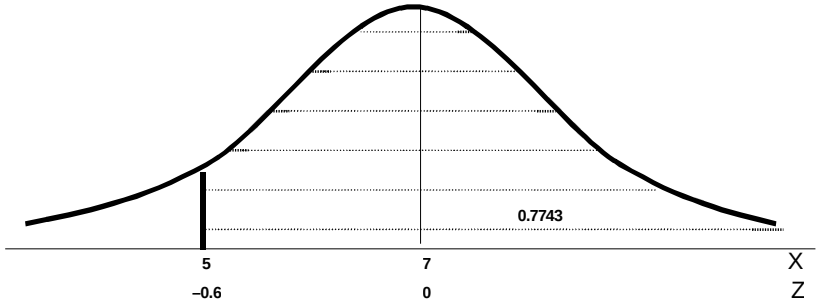


Figura 6.36. Distribución de la devolución de mercancías.

En este caso para obtener la probabilidad hay que sumar 0.5 + 0.2743 = 0.7743, por lo que la probabilidad de que más de 5 clientes regresen la mercancía es del 77.43%

3. Datos:

$p = 0.30$
$q = 0.70$
$n = 30$
$X = 10$
$P(X \geq 10)$

$$\mu = np = (30)\ (0.30) = 9 \qquad \sigma = \sqrt{npq} = \sqrt{(30)\ (0.30)\ (0.70)} = \sqrt{6.3} = 2.51$$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{9.5 - 9}{2.51} = \frac{0.5}{2.51} = 0.2$$

El valor de tablas para Z = 0.2 es 0.0793

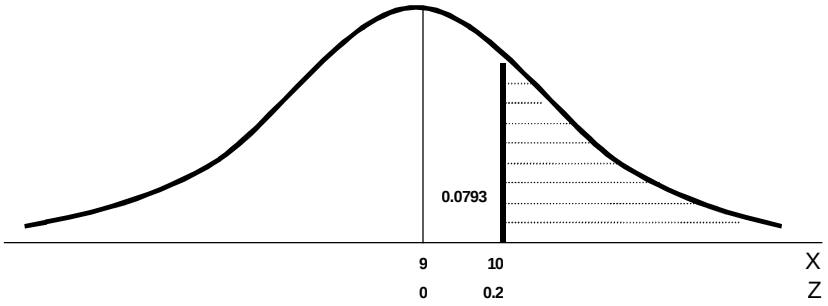


Figura 6.37. Distribución de las compras realizadas.

En este caso para obtener la probabilidad hay que restar 0.5 – 0.0793 = 0.4203, por lo que la probabilidad de que más de 10 clientes realicen una compra es de 42.07%

4. Datos:

$p = 0.70$
$q = 0.30$
$n = 50$
$X = 40$
$P(X \geq 40)$

$$\mu = np = (50)\ (0.70) = 35 \qquad \sigma = \sqrt{npq} = \sqrt{(50)\ (0.70)\ (0.30)} = \sqrt{10.5} = 3.24$$

Sustituyendo en la fórmula de la normal se obtiene:

$$Z = \frac{X - \mu}{\sigma} = \frac{39.5 - 35}{3.24} = \frac{4.5}{3.24} = 1.39$$

El valor de tablas para Z = 1.39 es 0.4177

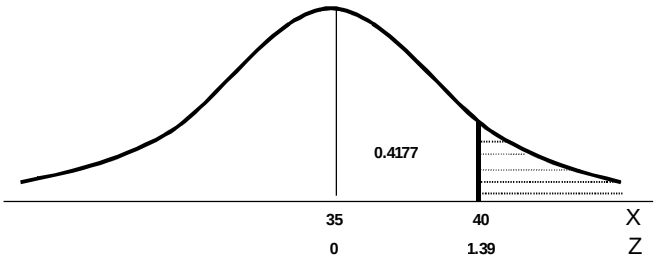


Figura 6.38. Distribución de las compras realizadas.

En este caso para obtener la probabilidad hay que restar 0.5 –0.4177 = 0.0823, por lo que la probabilidad de que más de 40 clientes compren uno o más productos es de 8.23%.

Ejercicio 4

1. c)
2. a)
3. d)
4. a)
5. b)
6. b)

Ejercicio 5

1. Antes de resolver este punto sería importante trazar un gráfico que representará cada uno de los puntos que se van a considerar (véase la figura 6.33).

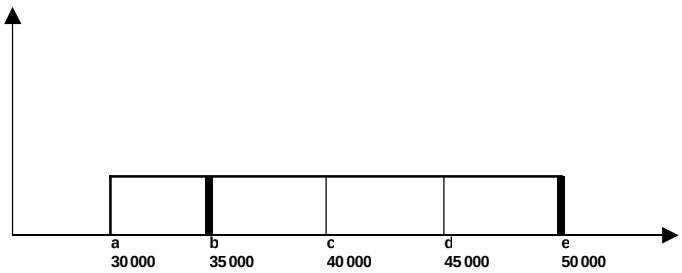


Figura. 6.39. Distribución uniforme de las ventas mensuales de computadoras.

Como el objetivo es encontrar la probabilidad de que las ventas sean mayores a los 35 000 litros, entonces nos centramos en el intervalo que corresponde a los puntos b y e, por lo tanto:

$$P(b \leq X \leq e) = \frac{e-b}{e-a} = \frac{50\,000 - 35\,000}{50\,000 - 30\,000} = \frac{15\,000}{20\,000} = 0.75$$

La probabilidad de que las ventas rebasen 35 000 litros de gasolina es 0.75 o de 75%.

2. a) Para calcular la media se emplea la fórmula: $\mu = \frac{a+e}{2} = \frac{1+5}{2} = 3$

Por lo tanto, la medida promedio de los tubos es de 3 metros.

b)

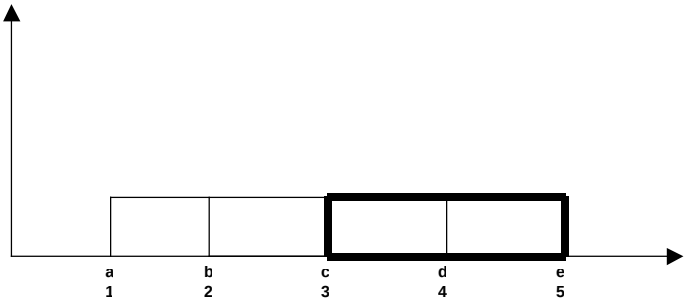


Figura 6.40. Distribución uniforme de las ventas mensuales de computadoras.

Como el objetivo es encontrar la probabilidad de que los tubos sean mayores de 3 metros, nos centramos en el intervalo comprendido entre c y e, la fórmula que se emplea es:

$$P(c \le X \le e) = \frac{e-c}{e-a} = \frac{5-3}{5-1} = \frac{2}{4} = 0.50$$

La probabilidad de que los tubos sean mayores de 3 metros es de 50%.

3. El ejercicio indica que al tener un ingreso mínimo de 4 800 y uno máximo de 7 200, se debe determinar la probabilidad de que el ingreso promedio sea entre 6 000 y 7 200:

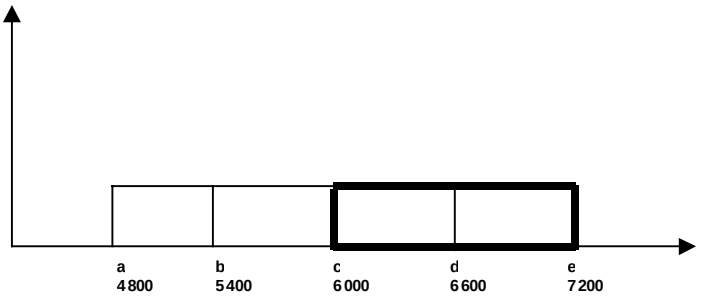


Figura 6.41. Distribución uniforme de los ingresos familiares mensuales.

Como el objetivo es encontrar la probabilidad de que los ingresos familiares sean entre 6 000 y 7 200, entonces nos centramos en el intervalo que corresponde a los puntos c y e, por lo tanto:

$$P(c \le X \le e) = \frac{e-c}{e-a} = \frac{7\,200-6\,000}{7\,200-4\,800} = \frac{1\,200}{2\,400} = 0.5$$

La probabilidad de que los ingresos familiares sean entre 6 000 y 7 000 mensuales es de 0.5 o de 50%

4. El ejercicio indica que al tenerse un beneficio mínimo de 30 000 y uno máximo de 70 000, se debe determinar la probabilidad de que el beneficio promedio esté entre 50 000 y 60 000.

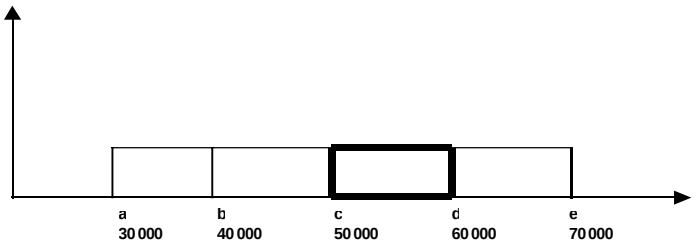


Figura 6.42. Distribución uniforme del beneficio de un consultor.

Como el objetivo es encontrar la probabilidad de que el beneficio del consultor esté entre 50 000 y 60 000, nos centramos en el intervalo que corresponde a los puntos c y d, por lo tanto:

$$P(c \leq X \leq d) = \frac{d-c}{e-a} = \frac{60\,000 - 50\,000}{70\,000 - 30\,000} = \frac{10\,000}{40\,000} = 0.25$$

La probabilidad de que el beneficio del consultor esté entre 50 000 y 60 000 es 0.25 o de 25%.

5. a) El ejercicio indica que al tener un salario anual mínimo de 150 000 y uno máximo de 250 000, se debe determinar la probabilidad de que el salario promedio anual esté entre 200 000 y 250 000.

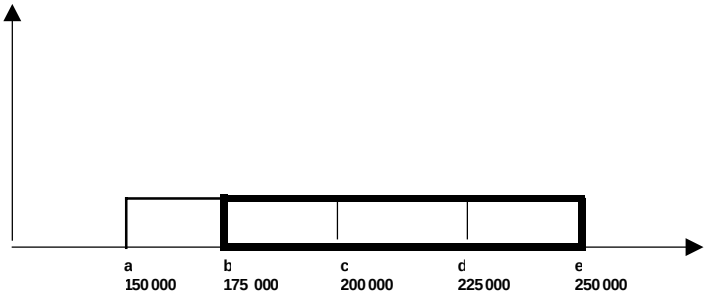


Figura 6.43. Distribución uniforme de salario de un vendedor.

Como el objetivo es encontrar la probabilidad de que el salario del vendedor oscile entre 175 000 y 250 000, nos centramos en el intervalo que corresponde a los puntos b y e, por lo tanto:

$$P(b \leq X \leq e) = \frac{e-b}{e-a} = \frac{250\,000 - 175\,000}{250\,000 - 150\,000} = \frac{75\,000}{100\,000} = 0.75$$

La probabilidad de que el salario de un vendedor oscile entre 175 000 y 250 000 es de 75%.

b) El ejercicio indica que al tenerse un salario anual mínimo de 150 000 y uno máximo de 250 000, se debe determinar la probabilidad de que el salario promedio anual sea menor de 200 000.

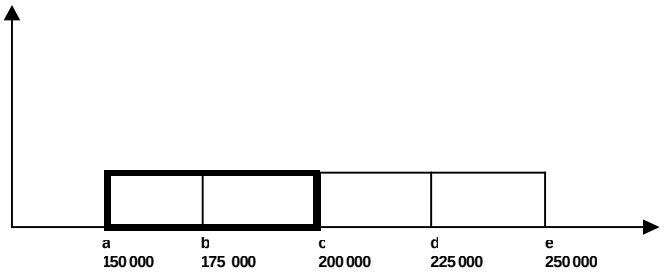


Figura 6.44. Distribución uniforme del salario de un vendedor.

Como el objetivo es encontrar la probabilidad de que el salario del vendedor oscile entre 150 000 y 200 000, nos centramos en el intervalo que corresponde a los puntos a y c, por tanto:

$$P(c \leq X \leq a) = \frac{c-a}{e-a} = \frac{200\,000-150\,000}{250\,000-150\,000} = \frac{50\,000}{100\,000} = 0.5$$

La probabilidad de que el salario de un vendedor esté entre 150 000 y 200 000 es de 50%.

Ejercicio 6

1. a) Lo primero que debemos obtener es el valor de λ, es decir, el número de llamadas por minuto, se tiene que $\lambda = \frac{1}{20} = 0.05$ por minuto.

El siguiente paso es sustituir en la primera fórmula y resolver:

$$P(T > t) = e^{-\lambda t}$$

$$P(T > 20) = e^{-0.05(20)} = e^{-1} = 0.3679$$

La probabilidad de que ocurran llamadas después de 20 minutos es de 36.49%.

b) Para resolver este inciso se utiliza la fórmula $P(T \leq t) = 1 - e^{-\lambda t}$, sustituyendo se obtiene

$$P(T \leq 20) = 1 - e^{-1} = 1 - 0.3679 = 0.6321$$

La probabilidad de que las llamadas tarden 20 minutos o menos es de 63.21%.

c) $\lambda = \frac{1}{10} = 0.1$

$$P(T \leq 10) = 1 - e^{-0.1} = 1 - 0.9048 = 0.095$$

La probabilidad de que las llamadas tarden 10 minutos o menos es 9.5%.

2. $\lambda = 5/10 = 0.5$

$P(T > t) = e^{-\lambda t}$

$P(T > 10) = e^{-0.5(5)} = e^{-2.5} = 0.082$

La probabilidad de que cinco personas tarden más de cinco minutos es de 8.2%.

3. $\lambda = 3/5 = 0.6$

$P(T > t) = e^{-\lambda t}$

$P(T > 5) = e^{-0.6(5)} = e^{-3} = 0.049$

La probabilidad de que las 3 consultas ocurran en más de 5 minutos es de 4.9%.

4. $\lambda = 6/30 = 0.2$

$P(T > t) = e^{-\lambda t}$

$P(T > 30) = e^{-0.2(30)} = e^{-6} = 0.0025$

La probabilidad de que se reciban 6 reclamaciones en más de 30 minutos es de 0.25%.

5. a) $\lambda = 4/5 = 0.8$

$P(T > t) = e^{-\lambda t}$

$P(T > 5) = e^{-0.8(5)} = e^{-4} = 0.0183$

La probabilidad de que se realicen los 4 servicios en más de 5 días es de 1.83%.

b) $\lambda = 4/5 = 0.8$

$P(T > t) = e^{-\lambda t}$

$P(T \leq 5) = 1 - e^{-0.8(5)} = 1 - e^{-4} = 1 - 0.0183 = 0.9817$

La probabilidad de que se realicen los 4 servicios en 5 días o menos es de 98.17%.

Ejercicio 7

1. d)
2. d)
3. a)
4. d)
5. c)
6. b)
7. b)
8. d)
9. a)
10. c)
11. d)
12. b)

13. a)
14. c)
15. a)
16. Datos:

$N = 9$
$n = 3$

Con la fórmula, se obtiene:

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{9!}{3!(9-3)!} = \frac{504}{6} = 84$$

El jefe de producción puede seleccionar 84 muestras distintas de focos.
Mientras que la probabilidad viene dada por:

$$\frac{1}{_NC_n} = \frac{1}{84} = 0.012$$

La probabilidad de obtener cada muestra es de 1.2%

17. Fórmula $N_K = nP_K$

Datos:

$N = 50\,000$
$N_1 = 20\,000$
$N_2 = 15\,000$
$N_3 = 10\,000$
$N_4 = 5\,000$
$n = 1\,000$

$$P_1 = \frac{N_1}{N} = \frac{20\,000}{50\,000} = 0.4 \qquad n_1 = (1\,000)\,(0.4) = 400$$

$$P_2 = \frac{N_2}{N} = \frac{15\,000}{50\,000} = 0.3 \qquad n_2 = (1\,000)\,(0.3) = 300$$

$$P_3 = \frac{N_3}{N} = \frac{10\,000}{50\,000} = 0.2 \qquad n_3 = (1\,000)\,(0.2) = 200$$

$$P_4 = \frac{N_4}{N} = \frac{5\,000}{50\,000} = 0.10 \qquad n_4 = (1\,000)\,(0.10) = 100$$

Conclusión. La muestra estará compuesta por 400 empresas micro, 300 pequeñas, 200 medianas y 100 empresas grandes.

18. Fórmula $N_K = nP_K$

Datos:

$N = 100\,000$
$N_1 = 40\,000$
$N_2 = 35\,000$

$N_3 = 25\,000$
$n = 1\,400$

$$P_1 = \frac{N_1}{N} = \frac{40\,000}{100\,000} = 0.40 \qquad n_1 = (1\,400)\,(0.40) = 560$$

$$P_2 = \frac{N_2}{N} = \frac{35\,000}{100\,000} = 0.35 \qquad n_2 = (1\,400)\,(0.35) = 490$$

$$P_3 = \frac{N_3}{N} = \frac{25\,000}{100\,000} = 0.25 \qquad n_3 = (1\,400)\,(0.25) = 350$$

La muestra estará compuesta por 560 tornillos de una pulgada, 490 tornillos de 1.5 pulgadas y 350 tornillos de 2 pulgadas.

Ejercicio 8

1. c)
2. b)
3. d)
4. a)
5. d)
6. a) En primera instancia lo que se establece es la media de la población, la cual se obtiene a partir de:

$$\mu = \frac{\sum X}{N} = \frac{4+5+7+6+3}{5} = \frac{25}{5} = 5$$

La varianza viene dada por:

$$\sigma^2 = \frac{1}{N}\sum (X-\mu)^2 = \frac{(4-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(7-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(3-5)^2}{5} = \frac{10}{5} = 2$$

La desviación estándar de la población es:

$$\sigma = \sqrt{\sigma^2} = \sqrt{2} = 1.41$$

b)

Muestras	Medias muestrales
4-5	4.5
4-7	5.5
4-6	5
4-3	3.5
5-7	6
5-6	5.5
5-3	4
7-6	6.5
7-3	5
6-3	4.5
10	$\sum \overline{X} = 50$

Tabla 6.8. Posibles muestras de ventas diarias de automóviles.

En la primera columna se aprecia que son 10 las muestras que se pueden obtener:

c)

$\overline{X}$	Probabilidad $\overline{X}$
4.5	2/10
5.5	2/10
5	1/10
3.5	1/10
6	1/10
4	1/10
6.5	2/10
	1

Tabla 6.9. Distribución de frecuencias muestrales de la media de ventas de automóviles.

d) La media muestral se obtiene a partir de:

$$\mu_{\overline{X}} = \frac{\sum \overline{X}}{n_{\overline{X}}} = \frac{50}{10} = 5$$

e) A partir del valor de la varianza se obtiene el error estándar de la media:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.41}{\sqrt{5}} = 0.63$$

7. Datos

μ = 1 200 hrs.
σ = 50 hrs.
n = 16 hrs.
$\overline{X}$ = 1 220 hrs.

En primer lugar se procede a calcular la $\sigma_{\overline{X}}$, mediante la siguiente fórmula:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.41}{\sqrt{5}} = 0.63$$

Ahora se sustituye el valor de $\sigma_{\overline{X}}$ en la fórmula de la normal estandarizada, de lo que resulta:

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{1\ 220 - 1\ 200}{12.5} = 1.6$$

El valor en tablas para Z = 1.6 es de 0.4452. Se tiene que el valor de tablas para la zona que va de 0 a 1.6 es de 0.4452, por lo que se procede a restar este valor a 0.5 de lo cual resulta 0.0548, que es la probabilidad que nos interesa, por lo tanto:

0.5 – 0.4452 = 0.0548

En conclusión, la probabilidad de que una muestra de 16 focos tenga una vida promedio de 1 220 hrs. es de 0.0548.

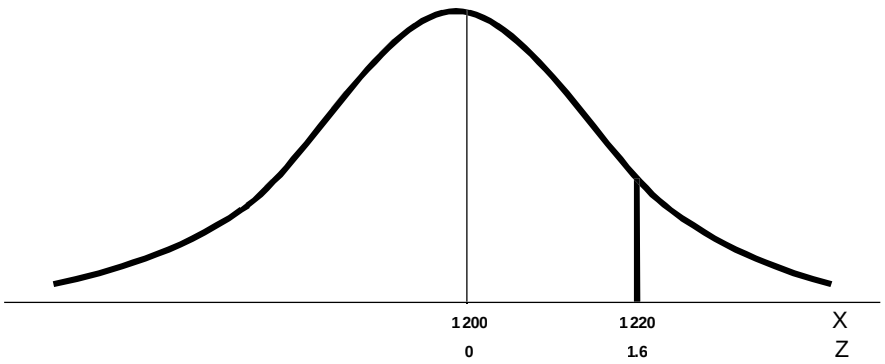


Figura 6.45. Distribución muestral de la media de la vida de focos.

8. Datos:

$45 < \overline{X} < 50$
$\sigma = 50$
$n = 100$
$\mu = 40$

Se procede a calcular la $\sigma_{\overline{X}}$ mediante la siguiente fórmula:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$$

Ahora se sustituye el valor de $\sigma_{\overline{X}}$ en la fórmula de la normal estandarizada, de lo que resulta:

$$Z_1 = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{45 - 40}{5} = 1 \quad \text{y} \quad Z_2 = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{50 - 40}{5} = 2$$

El valor en tablas para $Z_1 = 1$ es de 0.3413 y para $Z_2 = 2$ es de 0.4772, con ello la probabilidad de que se realicen entre 45 y 50 llamadas es:

$0.4772 - 0.3413 = 0.1359$

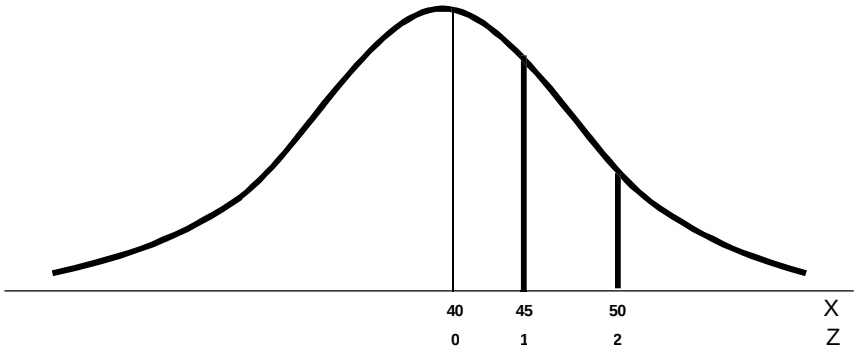


Figura 6.46. Distribución muestral de la media de las llamadas de un vendedor.

En conclusión, la probabilidad de que de una muestra de 100 llamadas se realicen entre 45 y 50 visitas es de 0.1359.

9. Datos:

$20 < \bar{X} < 30$
$\sigma = 40$
$n = 64$
$\mu = 25$

Se procede a calcular la $\sigma_{\bar{X}}$ mediante la siguiente fórmula:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

Ahora se sustituye el valor de $\sigma_{\bar{X}}$ en la fórmula de la normal estandarizada, de lo que resulta:

$$Z_1 = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{20 - 25}{5} = -1 \quad \text{y} \quad Z_2 = \frac{\bar{X}_2 - \mu}{\sigma_{\bar{X}}} = \frac{30 - 25}{5} = 1$$

El valor en tablas para $Z_1 = -1$ es de 0.3413 y para $Z_2 = 1$ es de 0.3413, con ello la probabilidad de que se realicen entre 20 y 30 llamadas es:

0.3413 + 0.3413 = 0.6826

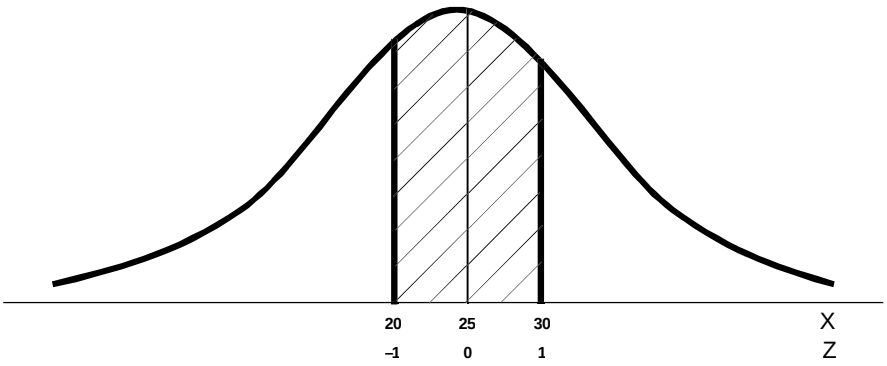


Figura 6.47. Distribución muestral de la media de las llamadas a una oficina de defensa del consumidor.

En conclusión, la probabilidad de que en una muestra de 64 llamadas recibidas por una oficina de protección al consumidor se realicen entre 20 y 30 llamadas es de 0.6826.

10. Datos:

$\bar{X} < 20$
$\sigma = 4$
$n = 36$
$\mu = 20$

Se procede a calcular la $\sigma_{\bar{X}}$, mediante la siguiente fórmula:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{36}} = 0.67$$

Ahora se sustituye el valor de $\sigma_{\bar{X}}$ en la fórmula de la normal estandarizada, de lo que resulta:

$$Z_1 = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{20 - 20}{0.67} = 0$$

El valor en tablas para Z = 0 es de 0.0000, con ello la probabilidad de que menos de 20 facturas estén atrasadas es 0.5 – 0.0000 = 0.5

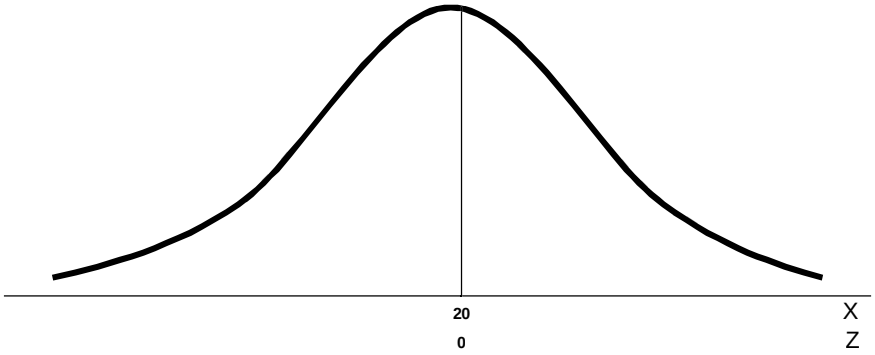


Figura 6.49. Distribución muestral de la media de las facturas atrasadas.

En conclusión, la probabilidad de que en una muestra de 36 facturas menos de 20 facturas estén atrasadas es de 0.50.

Respuestas a la autoevaluación

1. b)
2. c)
3. d)
4. d)
5. d)
6. d)
7. c)
8. b)
9. b)
10. c)
11. c)
12. a)
13. c)
14. a)
15. b)
16. a)
17. c)
18. c)
19. c)
20. a)
21. d)
22. d)
23. d)
24. c)
25. b)
26. a)
27. c)
28. b)
29. d)
30. a)
31. d)
32. a)
33. c)
34. d)
35. c)
36. b)
37. d)
38. d)
39. a)
40. a)