Estadística Inferencial

Roberto S. Villamarín Guevara.[1]

UNIVERSIDAD NACIONAL DE CHIMBORAZO [1]
Facultad de Ciencias de la Educación
Pedagogía de las Ciencias Experimentales Matemáticas y Físicas

Contenido

Contenido

Lo que debe recordar!!

1. ¿Qué es la Estadística?

Lo que debe recordar!!

1. ¿Qué es la Estadística?
2. ¿Cuál es su clasificación?

Lo que debe recordar!!

1. ¿Qué es la Estadística?
2. ¿Cuál es su clasificación?
3. ¿Por qué es importante estudiar Estadística?

Lo que debe recordar!!

1. ¿Qué es la Estadística?
2. ¿Cuál es su clasificación?
3. ¿Por qué es importante estudiar Estadística?
4. Variables. Clasificación. Niveles de Medición

Lo que debe recordar!!

1. ¿Qué es la Estadística?
2. ¿Cuál es su clasificación?
3. ¿Por qué es importante estudiar Estadística?
4. Variables. Clasificación. Niveles de Medición
5. Distribución de datos. Distribución de Frecuencias. Representación Gráfica de Datos.

Lo que debe recordar!!

1. ¿Qué es la Estadística?
2. ¿Cuál es su clasificación?
3. ¿Por qué es importante estudiar Estadística?
4. Variables. Clasificación. Niveles de Medición
5. Distribución de datos. Distribución de Frecuencias. Representación Gráfica de Datos.
6. Medidas de Tendencia Central: (Datos Agrupados - no agrupados)
    1. Media
    2. Mediana
    3. Moda
    4. Media Geométrica

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
    1. Amplitud de Variación
    2. Desviación Media
    3. Varianza y Desviación Estándar

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
   1. Amplitud de Variación
   2. Desviación Media
   3. Varianza y Desviación Estándar
2. Teorema de Chevysev - Regla Empírica

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
    1. Amplitud de Variación
    2. Desviación Media
    3. Varianza y Desviación Estándar
2. Teorema de Chevysev - Regla Empírica
3. Dispersión Relativa

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
    1. Amplitud de Variación
    2. Desviación Media
    3. Varianza y Desviación Estándar
2. Teorema de Chevysev - Regla Empírica
3. Dispersión Relativa
4. Asimetría

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
    1. Amplitud de Variación
    2. Desviación Media
    3. Varianza y Desviación Estándar
2. Teorema de Chevysev - Regla Empírica
3. Dispersión Relativa
4. Asimetría
5. Cuartíles, Decíles, Percentíles

Lo que debe recordar!!

1. Medidas de Dispersión: (Datos Agrupados - no agrupados)
    1. Amplitud de Variación
    2. Desviación Media
    3. Varianza y Desviación Estándar
2. Teorema de Chevysev - Regla Empírica
3. Dispersión Relativa
4. Asimetría
5. Cuartíles, Decíles, Percentíles
6. Diagramas de Caja

Lo que debe recordar!!

1. ¿Qué es la probabilidad?

Lo que debe recordar!!

1. ¿Qué es la probabilidad?
2. Clasificación:
    - Clásica
    - Empírica
    - Subjetiva

Lo que debe recordar!!

1. ¿Qué es la probabilidad?
2. Clasificación:
   - Clásica
   - Empírica
   - Subjetiva
3. Reglas de la Probabilidad:
   - de la Adición
   - de la Multiplicación

Lo que debe recordar!!

1. ¿Qué es la probabilidad?
2. Clasificación:
   - Clásica
   - Empírica
   - Subjetiva
3. Reglas de la Probabilidad:
   - de la Adición
   - de la Multiplicación
4. Diagramas de Árbol

Lo que debe recordar!!

1. ¿Qué es la probabilidad?
2. Clasificación:
   - Clásica
   - Empírica
   - Subjetiva
3. Reglas de la Probabilidad:
   - de la Adición
   - de la Multiplicación
4. Diagramas de Árbol
5. Teorema de Bayes

Lo que debe recordar!!

1. Principios del Conteo
    1. Fórmula de la Multiplicación
    2. Fórmula de la Permutación
    3. Fórmula de la Combinación

Lo que debe recordar!!

1. Principios del Conteo
   1. Fórmula de la Multiplicación
   2. Fórmula de la Permutación
   3. Fórmula de la Combinación
2. Distribuciones de probabilidad Discretas
   1. Variables aleatorias: Discreta, continua
   2. Media, varianza, desviación estándar de una distribución de probabilidad
   3. Distribución probabilística binomial
   4. Distribuciones probabilísticas acumulativas
   5. Distribución probabilística Hypergeométrica
   6. Distribución probabilística de Poisson

Lo que debe recordar!!

1. Distribución Probabilística Normal
    1. Familias de distribuciones probabilísticas normales
    2. Distribución probabilística norma estándar
    3. Área bajo la curva normal
    4. Factor de Corrección por continuidad

Contenido

Introducción

- La estadística descriptiva: Describe el fenómeno tal y como ocurre, para hace uso de las medidas de tendencia central y de dispersión.
- Recuerde que, en la inferencia estadística, el objetivo es determinar (inferir) algo sobre una población a partir sólo de una muestra.
- La población es todo el grupo de individuos u objetos en estudio, y la muestra es una parte o subconjunto de dicha población.
- Las distribuciones de probabilidad abarcan todos los posibles resultados de un experimento, así como la probabilidad asociada con cada resultado.
- Mediante las distribuciones de probabilidad se evaluó la **probabilidad de que ocurra algo en el futuro**.

Objetivos

1. Estudiar los métodos para seleccionar una muestra de una población.

Objetivos

1. Estudiar los métodos para seleccionar una muestra de una población.
2. Construir una distribución de la media de la muestra para entenderla forma como las medias muestrales tienden a acumularse en torno a la media de la población.

Objetivos

1. Estudiar los métodos para seleccionar una muestra de una población.
2. Construir una distribución de la media de la muestra para entenderla forma como las medias muestrales tienden a acumularse en torno a la media de la población.
3. Demostrar que, para cualquier población, la forma de esta distribución de muestreo tiende a seguir la distribución de probabilidad normal

Contenido

El por qué del Muestreo

1. Establecer contacto con toda la población requeriría mucho tiempo
2. El costo de estudiar todos los elementos de una población resultaría prohibitivo.
3. Es imposible verificar de manera física todos los elementos de la población.
4. Algunas pruebas son de naturaleza destructiva.
5. Los resultados de la muestra son adecuados.

Muestreo aleatorio simple

Definición

Muestra seleccionada de manera que cada elemento o individuo de la población tenga las mismas posibilidades de que se le incluya.

Al emplear números aleatorios para seleccionar empleados, se elimina la influencia o sesgo del proceso de selección

- Tablas de números aleatorios
- Software. Ms. Excel. $(= aleatorio.entre(inferior; superior))$
- R, R-studio, R-comander

Rstudio

```
install.packages("dlpyr")
install.packages("sampling")

 # muestreo aleatoria simple.
library(dplyr)
library(sampling)
set.seed(109) # genera la semilla
base
muestra_aleatoria <- slice_sample(.data=base, n=8, replace = FALSE)
# slice_sample saca todos los elementos del registro
muestra_aleatoria


muestra_aleatoria <- slice_sample(.data=base, n=8, replace = FALSE)
# slice_sample saca todos los elementos del registro
muestra_aleatoria
```

Figura: Muestreo aleatorio simple en Rstudio

Origen de los datos: *base.xlsx*
$n = 10$

No	N. aleatorio	elemento
1	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F4;A1:D71;4;)
2	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F5;A2:D72;4;)
3	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F6;A3:D73;4;)
4	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F7;A4:D74;4;)
5	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F8;A5:D75;4;)
6	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F9;A6:D76;4;)
7	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F10;A7:D77;4;)
8	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F11;A8:D78;4;)
9	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F12;A9:D79;4;)
10	=ALEATORIO.ENTRE(1;70)	=BUSCARV(F13;A10:D80;4;)

Muestreo aleatorio sistemático

Definición

Se selecciona un punto aleatorio de inicio y posteriormente se elige cada **k-ésimo** miembro de la población.

Se aplica cuando el Muestreo aleatorio simple, tiene un coste asociado (tiempo, recurso, facilidad)

Procedimiento:

- Calcular el valor de $k = N/n$, se redondea si es necesario
- Para el primer elemento (semilla). Aplicar el muestreo aleatorio simple.
- Luego a partir de la semilla seleccionar el siguiente elemento sumando k al valor de la semilla.
- Evite aplicar este método cuando los datos están ordenados con algún criterio (ascendente, descendente), en este caso el método no garantiza una **muestra aleatoria**

M. A. Simple Vs. M. A. Sistemático

Origen de datos: base.xlsx $n = 10$ $N = 70$

No	N. aleatorio	elemento		No	Semilla	k	No aleatorio	Elemento
1	66	154		1	21	7	21	190
2	41	165		2			28	154
3	60	163		3			35	168
4	46	169		4			42	151
5	19	175		5			49	156
6	51	177		6			56	190
7	40	153		7			63	170
8	28	154		8			70	150
9	12	169		9			7	165
10	42	151		10			14	154

Código en Excel

Origen de datos: base.xlsx

$$n = 10$$

$$N = 70$$

No	Semilla	k	Elemento de la muestra	Elemento
1	=REDONDEAR(ALEATORIO()*70;0)	=70/10	=SI(J4>63;J4+K4-70;J4)	=BUSCARV(L4;$A$2:$D$71;4;1)
2			=SI(L4+$K$4¿70;L4+$K$4-70;L4+$K$4)	=BUSCARV(L5;$A$2:$D$71;4;1)
3			=SI(L5+$K$4>70;L5+$K$4-70;L5+$K$4)	=BUSCARV(L6;$A$2:$D$71;4;1)
4			=SI(L6+$K$4>70;L6+$K$4-70;L6+$K$4)	=BUSCARV(L7;$A$2:$D$71;4;1)
5			=SI(L7+$K$4>70;L7+$K$4-70;L7+$K$4)	=BUSCARV(L8;$A$2:$D$71;4;1)
6			=SI(L8+$K$4>70;L8+$K$4-70;L8+$K$4)	=BUSCARV(L9;$A$2:$D$71;4;1)
7			=SI(L9+$K$4>70;L9+$K$4-70;L9+$K$4)	=BUSCARV(L10;$A$2:$D$71;4;1)
8			=SI(L10+$K$4>70;L10+$K$4-70;L10+$K$4)	=BUSCARV(L11;$A$2:$D$71;4;1)
9			=SI(L11+$K$4>70;L11+$K$4-70;L11+$K$4)	=BUSCARV(L12;$A$2:$D$71;4;1)
10			=SI(L12+$K$4>70;L12+$K$4-70;L12+$K$4)	=BUSCARV(L13;$A$2:$D$71;4;1)

Muestreo aleatorio estratificado

Definición

Una población se divide en subgrupos, denominados estratos, y se selecciona al azar una muestra de cada estrato.

Se aplica cuando una población se divide en grupos a partir de ciertas características, con el fin de garantizar el hecho de que cada grupo se encuentre representado en la muestra. Ejem: Estudiantes por sexo (H,M), por facultad (E,S,P,I)

Procedimiento:

- Se definen los estratos
- Aplicar el muestreo aleatorio simple.

Muestreo Aleatorio Estratificado - Ejemplo

Los datos son los gastos en publicidad de las 352 empresas más grandes de Estados Unidos. Suponga que el objetivo del estudio consiste en determinar si las empresas con altos rendimientos sobre el capital gastan en publicidad la mayor parte del dinero ganado en ventas que las empresas con un registro de bajo rendimiento o déficit.

Estrato	Probabilidad	Nº Empresas	F. Relativa	Nº Muestras
1	30 % y más	8	0.02	1 [1]
2	20 % - 30 %	35	0.10	5
3	10 % - 20 %	189	0.54	27
4	0 % - 10 %	115	0.33	16
5	Déficit	5	0.01	1
Total	—	352	1.00	50

Cuadro: Número seleccionado para una muestra aleatoria estratificada proporcional

[1] 0.02 de 50 =1

Rstudio - MAE

```
# Muestreo aleatorio estratificiado no proporcional -
set.seed(10)
muestra_est_np<-strata(data = base,
                       stratanames = c("nivel_estudios"), # se pueden agregar mas variables
                       size = c(1,5,5), # la cantidad de elementos de c/estrato
                       method = "srswor") # srswor = m.a.s sin reemplazo
muestra_est_np
data.frame(table(base$nivel_estudios))
nrow(base)
###### M. a . estratificado proporcional
nrow(base)
data.frame(table(base$nivel_estudios))
n_superior <- sum(with(base, nivel_estudios == "Superior"))/nrow(base) # with identifica y coge  los elemetos que cumplen la condición
n_superior
n_basico <- sum(with(base, nivel_estudios == "Básico")) /nrow(base)
n_basico
n_medio_superior <- sum(with(base, nivel_estudios == "Medio Superior")) /nrow(base)
n_medio_superior
muestra_est_p<-strata(data = base,
                       stratanames = c("nivel_estudios"), # se pueden agregar mas variables
                       size = c(round(12*n_superior),round(12*n_basico),round(12*n_medio_superior)), # 12 la cantidad de elementos de c/estrato
                       method = "srswor") # srswor = m.a.s sin reemplazo
muestra_est_p
data.frame(table(muestra_est_p$nivel_estudios)) # para verficar que la muestra es de 12 elementos
```

Figura: Muestreo Aleatorio Estratificado

Muestreo por Conglomerado

Definición

Una población se divide en conglomerados a partir delos límites naturales geográficos o de otra clase. A continuación se seleccionan los conglomerados al azar y se toma una muestra de forma aleatoria con elementos de cada grupo.

Se emplea a menudo para reducir el costo de muestrear una población dispersa en cierta área geográfica.

Recuerde!!

Nota importante:

El estudio de los métodos de muestreo de las secciones anteriores no incluye todos los métodos de muestreo disponibles para el investigador.
Si usted emprendiera un proyecto de investigación importante de marketing, finanzas, contabilidad u otras áreas, necesitaría consultar libros dedicados exclusivamente a la teoría del muestreo y al diseño de muestras.

Rstudio-Muestreo por conglomerado

```
# Muestreo por conglomerados... (base_trabajadores)

library(readxl)
base_trabajadores <- read_excel("Downloads/base_trabajadores.xlsx")
View(base_trabajadores)

data.frame(table(base_trabajadores$sucursal)) # verificar los conglomerados

set.seed(10)

cluster(data = base_trabajadores,
        clustername = c("sucursal"),
        size = 3, # NUMERO DE CONGLOMERADOS PARA OBTENER 120 MAS O MENOS
        method = "srswor",
        description = TRUE)
```

Figura: Muestreo Aleatorio por Congolmerados

Tarea

Actividades de aprendizaje

Resolver los siguientes ejercicios:

Capítulo 8

Ejercicio 1 - 4

Estadística Aplicada a la Administración y Economía Lind & Marshall.

Contenido

Error de Muestreo

- En la sección anterior se estudiaron métodos de muestreo útiles para seleccionar una **muestra** que constituya una representación imparcial o **sin sesgos** de la población.
- Las muestras se emplean para determinar características de la población.
- Por ejemplo, con la media de una muestra se calcula la media de la población.
- No obstante, como la muestra forma parte o es una porción **representativa** de la población, es poco probable **que la media de la muestra sea exactamente igual a la media poblacional**.
- Por tanto, puede esperar una **diferencia entre un estadístico de la muestra y el parámetro de la población**

Error de muestreo

- Las muestras se emplean para determinar características de la población.
- No obstante,como la muestra forma parte o es una porción representativa de la población, es poco probable que la media de la muestra sea exactamente igual a la media poblacional.
- Asimismo, es poco probable que la **desviación estándar** de la muestra sea exactamente igual a la *desviación estándar de la población*.
- Por tanto, puede esperar una diferencia entre un estadístico de la muestra y el parámetro de la población correspondiente.
- Esta diferencia recibe el nombre de error **de muestreo**

Definición

Es la diferencia entre el estadístico de una muestra y el parámetro de la población correspondiente

Ejemplo

- La población es el número de habitaciones rentadas cada uno de los 30 días. Determine la media de la población.
  $NH : \{0, 2, 3, 2, 3, 4, 2, 3, 4, 7, 3, 4, 4, 4, 7, 0, 5, 3, 6, 2, 3, 2, 3, 6, 0, 4, 1, 1, 3, 3\}$
- Utilice Excel u otro software de estadística para seleccionar tres muestras aleatorias de cinco días.
- Calcule la media de cada muestra y compárela con la media poblacional.
- ¿Cuál es el error de muestreo en cada caso?
  $$\mu = \frac{\sum x}{N} = \frac{0+2+2+\cdots+3}{30} = \frac{94}{30}$$
- $$\overline{x}_1 = \frac{4+7+4+3+1}{5} = 3,80$$
  $$e_{m_1} = \overline{x}_1 - \mu = 0,67$$

Ejemplo cont..

- $\overline{x}_2 = \dfrac{3+3+2+3+6}{5} = 3,4$
  $e_{m_2} = \overline{x}_2 - \mu = 0,27$
- $\overline{x}_3 = 1,8$
  $e_{m_3} = \overline{x}_3 - \mu = -1,33$

¿Cuántas posibles muestras existen tomando cada una de ellas de 5 elementos?
Sol: 142506.
Justifique su respuesta

Recuerde!!

Cada muestra puede tener una media muestral diferente y, por consiguiente,un error de muestreo distinto.

Cont..

- El valor del error de muestreo se basa en el valor particular de las 142 506[2] posibles muestras seleccionadas.
- Por consiguiente, los errores de muestreo son aleatorios y se presentan al azar.
- Si determinara la suma de estos errores de muestreo en una gran cantidad de muestras, el resultado se aproximaría mucho a cero.
- Sucede así porque la media de la muestra constituye un **estimador sin sesgo de la media de la población**.

[2]total de combinaciones de 30 tomando 5 a la vez

Contenido

Distribución Muestral de la medias

- Las medias muestrales del ejemplo anterior varían de una muestra a la siguiente.
- La media de la primera muestra de 5 días fue de 3.80 habitaciones, y la media de la segunda muestra fue de 3.40 habitaciones.
- La media poblacional fue de 3.13 habitaciones.
- Si organiza las medias de todas las muestras posibles de 5 días en una distribución de probabilidad, el resultado recibe el nombre de

Distribución muestral de la media

Distribución de probabilidad de todas las posibles medias de las muestras de un determinado tamaño muestra de la población.

Ejemplo

Considere los siguientes datos:

Empleado	Joe	Sam	Bob	Sue	Jan	Art	Ted
Ingreso / hora (\$)	7	7	8	8	7	8	9

1.- ¿Cuál es la media de la población?

- $\mu = \frac{\sum}{\overline{X}} = \frac{7+7+8+8+8+8+9}{7} = \$7,71$

Ejemplo

2.- ¿Cuál es la distribución muestral de la media para muestras de tamaño 2?

Calcular todas las todas las muestras posibles de tamaño 2

$$_{N}C_{n} = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

Joe - Sam					
Joe - Bob	Sam - Bob				
Joe - Sue	Sam -Sue	Bob - Sue			
Joe - Jam	Sam -Jam	Bob - Jam	Sue - Jam		
Joe - Art	Sam - Art	Bob - Art	Sue - Art	Jam - Art	
Joe - Ted	Sam -Ted	Bob - Ted	Sue - Ted	Jam - Ted	Art - Ted

Distribución muestral para n=2

Media Muestral	Numero de Medias	probabilidad
7	3	,1419
7,5	9	,4285
8	6	,2857
8,5	3	,1429
Total	21	1,00

Ejemplo

3.- ¿Cuál es la media de la distribución muestral de la media?

- La media de la distribución muestral de la media se obtiene al sumar las medias muestrales y dividir la suma entre el número de muestras.
  La media de todas las medias muestrales se representa mediante $\mu_{\overline{X}}$.
- La μ recuerda que se trata de un valor poblacional, pues tomó en cuenta todas las muestras posibles.
- El subíndice $\overline{X}$ indica que se trata de la distribución muestral de la media.

$$\mu_{\overline{X}} = \frac{\sum \overline{X}}{N} = \frac{7 * 3 + 7{,}5 * 9 + 8 * 6 + 8,5 * 3}{21} = 7,71$$

Ejemplo

4.- ¿Qué observaciones es posible hacer sobre la población y la distribución muestral de la media?

1. La media de la distribución muestral es igual a la media de la distribución $\mu = \mu_{\overline{X}}$
2. La dispersión de la distribución muestral de las medias es menor que la dispersión de los valores de población.
   La media de las muestras varía de \$7,00 a\$8,50, mientras que los valores de población varían de \$7,00 a \$9,00.
3. Observe que, conforme se incrementa el tamaño de la muestra, se reduce la dispersión de la distribución muestral de las medias.
4. La forma de la distribución muestral de la media y la forma de la distribución de frecuencias de los valores de población son diferentes.
5. La distribución muestral de las medias tiende a adoptar más forma de campana y a aproximarse a la distribución de probabilidad normal.

Ejemplo ..

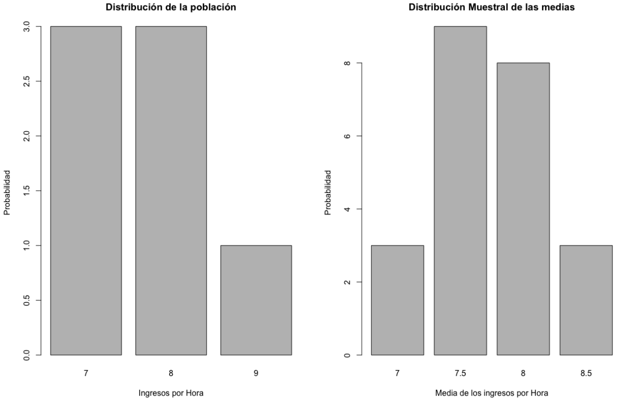


Figura: Distribución de los valores de población y distribución muestral de las medias

En resumen

En resumen, tome todas las posibles muestras aleatorias de una población y calcule un estadístico muestral (la media de los ingresos percibidos) para cada una. Este ejemplo ilustra las importantes relaciones entre la distribución poblacional y la distribución muestral de la media:

1. La media de las medias de las muestras es exactamente igual a la media de la población
2. La dispersión de la distribución muestral de la media es más estrecha que la distribución poblacional.
3. La distribución muestral de la media suele tener forma de campana y se aproxima a la distribución de probabilidad normal.

Actividades de aprendizaje

Ejercicios del 5 al 10
Cap. 8
Marshall, L., & Wathen, S. A. (2015). Estadística aplicada a los negocios y economía. Ciudad México.

Contenido

Teorema del Límite Central

- El teorema del límite central hace hincapié en que, en el caso de muestras aleatorias grandes, la forma de la distribución muestral de la media se aproxima a la distribución de probabilidad normal.
- La aproximación es más exacta en el caso de muestras grandes que en el de muestras pequeñas
  Ésta es una de las conclusiones más útiles de la estadística
- Permite razonar sobre la distribución de las medias muestrales sin ninguna información acerca de la forma de la distribución de población de la que se toma la muestra.
- En otras palabras, el teorema del límite central se cumple en el caso de todas las distribuciones.

Teorema

Si todas las muestras de un tamaño en particular se seleccionan de cualquier población, la distribución muestral de la media se aproxima a una distribución normal. **Esta aproximación mejora con muestras más grandes.**

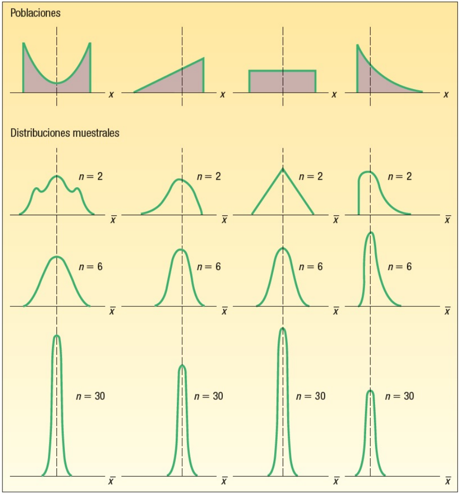
Teorema del Límite Central

- Si la población obedece a una distribución normal, entonces, en el caso de cualquier tamaño de muestra, la distribución muestral de las medias también será de naturaleza normal.
- Si la distribución poblacional es simétrica (pero no normal), se verá que la forma normal de la distribución muestral de las medias se presenta con muestras tan pequeñas como 10.
- Por otra parte, si se comienza con una distribución sesgada o con colas gruesas, quizá se requieran muestras de 30 o más para observar la característica de normalidad.

Recuerde!!

- Observe la convergencia hacia una distribución normal sin importar la forma de la distribución de población.
- Una muestra de 30 o mayor es lo bastante grande para aplicar el teorema del límite central.
- Si la población obedece a una distribución normal, entonces, en el caso de cualquier tamaño de muestra, la distribución muestral de las medias también será de naturaleza normal.
- Si la distribución poblacional es simétrica (pero no normal), se verá que la forma normal de la distribución muestral de las medias se presenta con muestras tan pequeñas como 10.
- Si se comienza con una distribución sesgada o con colas gruesas, quizá se requieran muestras de 30 o más para observar la característica de normalidad

Resultados del teorema del límite central para diversas poblaciones



Actividades de aprendizaje

Ed decide formar un comité de cinco empleados. Se pedirá al comité que estudie el tema del cuidado de la salud y haga alguna recomendación sobre el plan que mejor convenga a los empleados. Ed cree que el punto de vista de los empleados más recientes en relación con el cuidado de la salud difiere de los empleados con más experiencia.

Si Ed selecciona al azar este comité:

- ¿qué puede esperar en términos del promedio de años que llevan con Spence Sprockets los miembros del comité?
- ¿Cuál es la forma de la distribución de años de experiencia de todos los empleados (la población) en comparación con la forma de la distribución muestral de las medias?
- Los tiempos de servicio (redondeados al año inmediato) delos 40 empleados que actualmente están en nómina en Spence Sprockers, Inc., son los siguientes:

Ejercicio: Años de servicio

11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

1. Hallar la gráfica de la distribución de la población
2. Hallar la media de la población
3. Seleccione al azar 25 muestras de 5 empleados cada una, calcule la media de cada una de las muestras
4. Hallar la gráfica de la distribución de las medias
5. Compare las dos gráficas y establezca una conclusión
6. Halle el rango de los datos de la población y los rango de las medias muestrales ¿Qué puede concluir?
7. Repita el proceso anterior con 25 muestras de 20 empleados
8. Del análisis de los datos y las gráficas ¿Que puede concluir?

Para finalizar

- La forma de la distribución muestral de las medias es diferente a la de la población
- **A medida que incrementa el tamaño de la muestra, la distribución muestral de las medias se aproxima a la distribución de probabilidad normal**. Este hecho se ilustra con el teorema del límite central.
- Hay menos dispersión en la distribución muestral de las medias que en la distribución de la población.
- También puede comparar la media de las medias de la muestra con la media dela población. Éstas se encuentran muy próximas.
- El teorema del límite central indica que, sin importarla forma de la distribución de población, la distribución muestral de la media se aproximará a la distribución de probabilidad normal.

Error estándar de la media

- El teorema del límite central mismo no dice nada sobre la dispersión de la distribución muestral de medias, ni sobre la comparación entre la media de la distribución muestral de medias y la media de la población.
- La media de las medias de las muestras se encuentra cerca de la media de la población
- Se puede demostrar que la media de la distribución muestral es la media poblacional, es decir,que

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

  y si la desviación estándar de la población es σ, la desviación estándar de las medias muestrales es $\frac{\sigma}{\sqrt{n}}$, en la que n es el número de observaciones de cada muestra
- Entonces, $\frac{\sigma}{\sqrt{n}}$ es el error estándar de la media. En realidad, el nombre completo es **desviación estándar de la distribución muestral de medias.**

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

Conclusiones

1. La media de la distribución muestral de las medias será exactamente igual a la media poblacional si selecciona todas las muestras posibles del mismo tamaño deuna población dada. Es decir,

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

2. Aunque no seleccione todas las muestras, es de esperar que la media de la distribución muestral de medias se aproxime a la media poblacional.
3. Habrá menos dispersión en la distribución muestral de las medias que en la población.
4. Si la desviación estándar de la población es σ, la desviación estándar de la distribución muestral de medias es $\frac{\sigma}{\sqrt{n}}$
5. Se cumple que cuando se incrementa el tamaño de la muestra, disminuye el error estándar de la media.

Actividades de aprendizaje

Resolver los ejercicios 11-14
Pág 280
Marshall, L., & Wathen, S. A. (2015). Estadística aplicada a los negocios y economía. Ciudad México. 13º Edición

Contenido

Uso de la distribución muestral de las medias

- El análisis de las muestras o muestreo, es fundamental para la toma de decisiones (efectivas o no!)
- Con el análisis anterior es posible determinar si el error de muestreo, se debe al azar.
- Es posible determinar la probabilidad de que la media de una muestra se encuentre dentro de un determinado margen.
- La distribución de muestreo, seguirá la distribución de probabilidad normal con dos condiciones:
  1. Cuando se conoce que las muestras se toman de una población normal, en este caso el tamaño de la muestra no constituye un factor (será también normal)
  2. Cuando se desconoce la distribución de la población, o se conoce que no es normal y la muestra tiene al menos 30 observaciones. En este caso el **T. del límite Central** garantiza que la **Distribución muestral de las medias, tiene una distribución normal.**

Recuerde que:

- Valor z= distancia con signo entre el valor seleccionado, designado por X, y la media μ, dividida entre la desviación estándar σ

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

- De esta manera, el valor z es la distancia de la media, medida en unidades de desviación estándar.
- Cualquier distribución de probabilidad normal puede convertirse en una distribución de probabilidad normal estándar al restar la media de cada observación y dividir esta diferencia entre la desviación estándar.
- Los resultados reciben el nombre de valores z, o valores tipificados.

Cálculo de Z

1. Las decisiones (especialmente en negocios) se toman en función de la muestras, no en función de una sola observación.
2. Lo importante es la distribución de la media muestral, en lugar de una sola observación. Primer cambio importante en la fórmula anterior.
3. El segundo cambio es utilizar $\dfrac{\sigma}{\sqrt{n}}$

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Ejemplo

Los registros indican que la cantidad de bebida de cola tiene una distribución de probabilidad normal. La cantidad media por botella es de 31.2 onzas,y la desviación estándar de la población, de 0.4 onzas.
Se selecciona 16 botellas al azar de la linea de llenado. La media es de 31.38 onzas.

1. ¿Es un resultado poco probable?
2. ¿Es probable que el proceso permita colocar demasiada bebida en las botellas?
3. En otras palabras, ¿es poco común el error de muestreo de 0.18 onzas?

Solución

1. $z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{31{,}38 - 31{,}20}{\dfrac{0,4}{\sqrt{16}}} = 1{,}80$
2. El numerador de esta ecuación, $\overline{X} - \mu = 31{,}38 - 31{,}20 = 0{,}18$, es el error muestral
3. El denominador, $\dfrac{\sigma}{\sqrt{n}} = \dfrac{0,4}{\sqrt{16}} = 0{,}1$, es el error estándar de la distribución muestral de la media
4. Los valores z expresan el error muestral en unidades estándar; en otras palabras, el error estándar.
5. Calcule la probabilidad de un valor z mayor que 1,80. Este valor es de 0,4641
6. La probabilidad de un valor z mayor que 1.80 es de 0.0359, que se calcula con la resta 0.5000 – 0.4641

Solución - Conclusión

1. No es probable —menos de 4 % de probabilidad— que seleccione una muestra de 16 observaciones de una población normal con una media de 31,2 onzas y una desviación estándar poblacional de 0,4 onzas, y determine que la media de la muestra es igual o mayor que 31,38 onzas

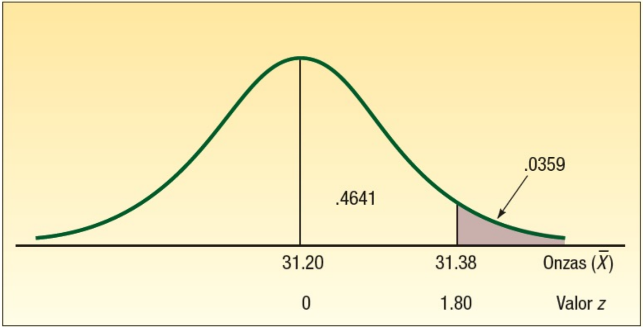


Figura: Distribución muestral de la cantidad media de bebida de cola en una botella gigante

Actividades de aprendizaje

1. Ejercicios 15- 18.
2. Ejercicios 21 - 45 (múltiplos de 3)

Contenido

Referencias

Marshall, L., & Wathen, S. A. (2015). Estadística aplicada a los negocios y economía. Ciudad México.