Estimación de Intervalos de Confianza

Roberto S. Villamarín G.

PCEMYF

3 de mayo de 2025

Objetivos

1. Definir un estimador puntual
2. Definir el nivel de confianza
3. Construir un intervalo de confianza para la media poblacional cuando se conoce la desviación estándar de la población
4. Construir un intervalo de confianza para la media poblacional cuando no se conoce la desviación estándar de la población
5. Construir un intervalo de confianza para una proporción de la población.
6. Determinar el tamaño de la muestra para un muestreo de atributos y variables

Introducción

1. Recordemos algunas razones para el muestreo
   1. Estudiar toda la población es costoso y poco práctico
   2. Por lo general, los resultados de la muestra son adecuados
   3. Algunas pruebas son de naturaleza destructiva
   4. Es imposible revisar todos los elementos
2. La media muestral ofrece una buena aproximación de la media poblacional

Considere los siguientes cuestionamientos:

1. ¿Con cuántos electores debe ponerse en contacto una compañía dedicada a realizar encuestas con el fin de predecir los resultados de las elecciones?
2. ¿Cuántos productos se necesitan analizar para garantizar el nivel de calidad?
3. ¿Cuál es el tamaño de una muestra adecuada?

Estimadores puntuales e intervalos de confianza de la media

El análisis de los estimadores puntuales y los intervalos de confianza comienza con el estudio del cálculo de la media poblacional. Se deben considerar dos casos:

1. Se conoce la desviación estándar de la población (σ)
2. Se desconoce la desviación estándar de la población (σ). En este caso se sustituye la desviación estándar de la muestra (s) por la desviación estándar de la población (σ)

Existen importantes distinciones en los supuestos entre estos dos casos. **Primero se considera el caso en el que σ se conoce.**

1. En la mayoría de los casos no se conoce la población, o es demasiado grande.
2. Esto hace que se deba confiar en la información de la muestra
3. Es decir, no se conoce el parámetro poblacional

Estimador puntual

Es un estadístico calculado a partir de la información de la muestra para estimar el parámetro poblacional

Ejemplo selecciona una muestra aleatoria de 50 compradores recientes, determina la edad de cada comprador y calcula la edad media de los compradores de la muestra. Entonces **La media de esta muestra es un estimador puntual** de la media de la población.

Ejemplo

- La media $(\overline{X})$ es un **estimador puntual** de la media poblacional (μ)
- p un proporción muestral, es un **estimador puntual** de π la **proporción poblacional**
- s la **desviación estándar muestral**, es un estimador puntual de (σ), la **desviación estándar poblacional**

Intervalo de confianza

Los estimadores puntuales son aproximaciones de los valores poblacionales
Pregunta:¿Es posible determinar cuan cercano esta un estimador del valor real?

Intervalo de confianza

Conjunto de valores formado a partir de una muestra de datos de forma que exista la posibilidad de que el parámetro poblacional ocurra dentro de dicho conjunto con una probabilidad específica. **La probabilidad específica recibe el nombre de nivel de confianza.**

Los resultados del teorema del límite central (para muestras grandes) permiten afirmarlo siguiente:

1. Noventa y cinco por ciento de las medias muestrales seleccionadas de una población se encontrará a $\pm 1{,}96$ desviaciones estándares de la media poblacional, μ
2. Noventa y nueve por ciento de las medias muestrales se encontrará a $\pm 2{,}58$ desviaciones estándares de la media poblacional.

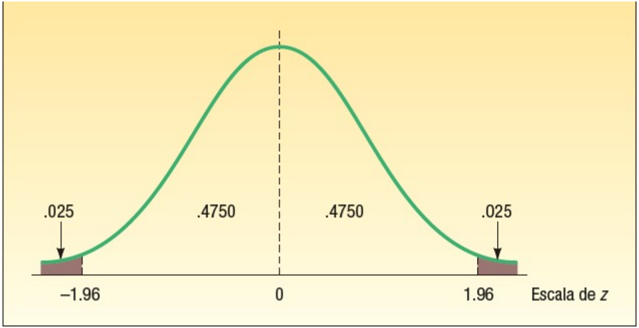


Figura: Intervalos de confianza

1. Fórmula para un 95 % de confianza $\overline{X} \pm \mathbf{1.96}\frac{\sigma}{\sqrt{n}}$
2. Fórmula para un 99 % de confianza $\overline{X} \pm \mathbf{2.58}\frac{\sigma}{\sqrt{n}}$

Nota

NO existe restricciones sobre los niveles de confianza. Es posible seleccionar cualquier nivel de confianza entre 0 y 100 % y hallar el valor de z correspondiente.

$$\overline{X} \pm \mathbf{z}\frac{\sigma}{\sqrt{n}}$$

Nivel de Confianza	Probabilidad más cercana	Valor z
80 %	0,3997	1,28
90 %	0,4505	**1,65**
94 %	0,4699	1,88
95 %	0,475	**1,96**
96 %	0,4798	2,05
99 %	0,4951	**2.58**

Ejemplo

Una muestra aleatoria de 256 gerentes revela una media muestral de $45420

La desviación estándar de esta muestra es de $2050

1. ¿Cuál es la media de la población?
2. ¿Cuál es un conjunto de valores razonable para la media poblacional?
3. ¿Cómo se deben interpretar estos resultados?

¿Cuál es la media de la población?

- Se desconoce la media de la población.
- La mejor estimación del valor de población sea el estadístico de la muestra correspondiente
- La media de la muestra de $45420 constituye un **estimador puntual** de la media poblacional desconocida

¿Cuál es el conjunto de valores razonable para la media poblacional?

Con un nivel de confianza del 95 %

$$\begin{aligned} \overline{X} \pm z\frac{\sigma}{\sqrt{n}} &= 45420 \pm 1,96\frac{2050}{\sqrt{256}} \\ &= 45420 \pm 251 \\ &= 45169 \leftrightarrow 45671 \end{aligned}$$

- Los valores extremos $45169 \leftrightarrow 45671$ reciben el nombre de los **límites de confianza**
- El valor del 95 % recibe el nombre del **nivel de confianza**
- El **intervalo de confianza** abarca desde 45169 a 45671
- El valor ± 251 se le conoce como el **margen de error**

¿Cómo se deben interpretar estos resultados?

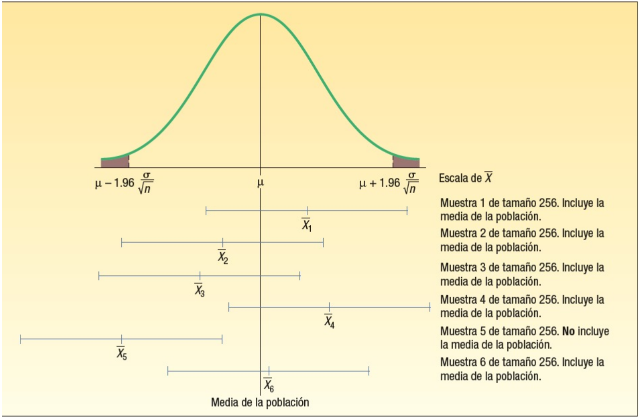


Figura: Probabilidad de el 95 % de los intervalos de confianza contengan la media de la población

Actividades de aprendizaje

Ejercicios:
1 al 8
Pág 301 -302
Estadística Aplicada a los negocios y economía Lind and Marsal 15ª Edición

1. En la mayoría de los casos de muestreo, no se conoce la desviación estándar de la población σ
2. Se utiliza la **desviación estándar de la muestra** *para estimar la desviación estándar poblacional.*
3. Es decir, se utiliza s, la desviación estándar de la muestra, para estimar σ, la desviación estándar de la población
4. No se puede utilizar la fórmula $\overline{X} \pm z\dfrac{\sigma}{\sqrt{n}}$
5. Como no conoce σ, no puede utilizar la distribución z. Sin embargo,hay una solución: utilizar la desviación estándar de la media y **sustituir la distribución z con la distribución t.**

La distribución t

- Es una distribución de probabilidad continua, con muchas características similares a las de la distribución z.
- Proviene del estudio del comportamiento de

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

- Siendo s un estimador de σ

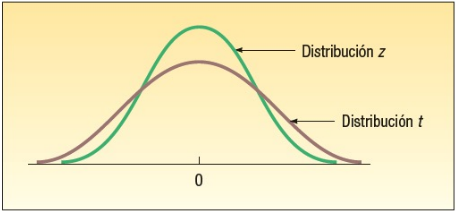


Figura: Distribución normal estándar y distribución t de Student

Características de t

1. La distribución t es más plana y que se extiende más que la distribución normal estándar. Puesto que desviación estándar de la distribución t es mayor que la distribución normal estándar
2. Es una distribución continua
3. Tiene forma de campana y es simétrica
4. No existe una distribución t, sino una familia de distribuciones t.
5. Todas las distribuciones t tienen una media de 0, y sus desviaciones estándares difieren de acuerdo con el tamaño de la muestra, n
6. La distribución t se extiende más y es mas plana por el centro que la distribución normal estándar.
7. Conforme se incrementa el tamaño de la muestra, la distribución t se aproxima a la distribución normal estándar, pues los errores que se cometen al utilizar s para estimar σ disminuyen con muestras más grandes.

Características de t

- t de Student posee mayor dispersión que la distribución z
- Los valores de t para un nivel de confianza dado tiene una magnitud mayor que el valor z correspondiente

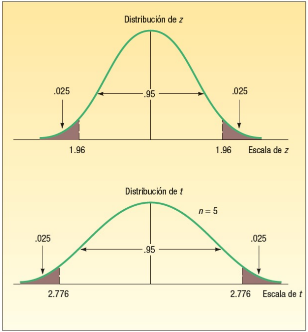


Figura: Valores de z y t para el nivel de confianza de 95 %

INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL CON σ DESCONOCIDA

Para crear un intervalo de confianza para la media poblacional con una desviación estándar desconocida:

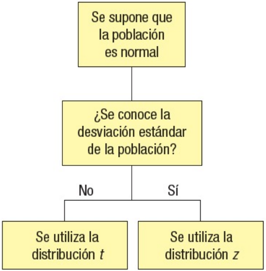


Figura: Determinar cuándo usar la distribución z o la distribución t

Ejemplo

Una muestra (n) de 10 llantas, para recorrer 50 mil millas, tiene una $\overline{X} = 0{,}32$, y $s = 0,09$ pulgadas.

- Construya un intervalo de confianza de 95 % para la media poblacional
- ¿Sería razonable que el fabricante concluyera que después de 50 mil millas la cantidad media poblacional de cuerda restante es de 0.30 pulgadas?

Solución

1. Suponer que la distribución de la población es normal.
2. Se conoce la desviación estándar de la muestra $s = 0{,}09$ (no de la población σ)
3. Aplica la fórmula $\overline{X} \pm t \cdot \frac{s}{\sqrt{n}}$

$$\begin{aligned}\overline{X} \pm t \cdot \frac{s}{\sqrt{n}} &= 0,32 \pm \mathbf{2{,}262}\frac{0,09}{\sqrt{10}} \\ &= 0,32 \pm 0,064 \\ &= 0,256 \leftrightarrow 0,384\end{aligned}$$

4. ¿Cómo interpretar este resultado?
   - Resulta razonable concluir que la media poblacional se encuentra en este intervalo.
   - El fabricante puede estar seguro (95 % seguro) de que la profundidad media de las cuerdas oscila entre 0,256 y 0,384 pulgadas
   - Posiblemente la media de la población sea de 0,30, se encuentre en el intervalo de confianza

Ejemplo 2

Un articulo de prensa sostiene que el tiempo medio de venta de una residencia es de 60 días. Se selecciona una muestra aleatoria de 20 residencias y se encuentra que el tiempo de venta es de 65 días. Se crea un intervalo de confianza del 95 % para la media de la población, se obtiene que los puntos extremos del intervalo de confianza son: 62 y 68 días.

¿Cómo interpreta este resultado?

- Note que el valor propuesto es de 60 días, que no se encuentra en el intervalo de confianza, por lo que no es probable (o es muy poco probable) que la media sea 60 días.
- La evidencia indica que lo afirmado por la prensa, puede no ser correcto!
- $\therefore$Parece poco razonable obtener la muestra que usted tomó de una población que tenía un tiempo de venta medio de 60 días

Aplicación en R

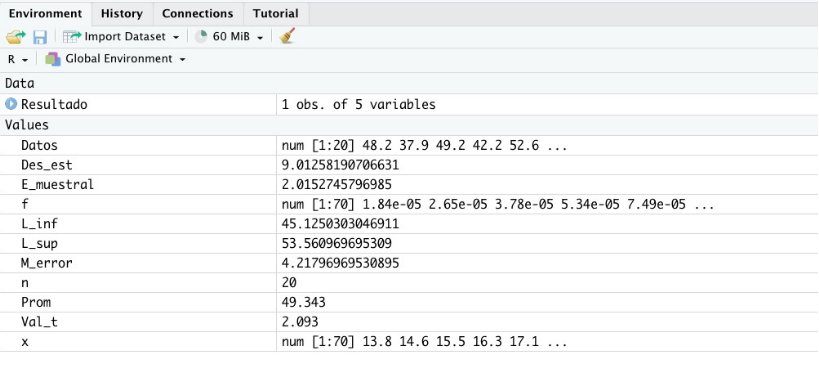
Una muestra de 20 clientes revela los siguientes datos, sobre la cantidad de dinero que gastan en el centro comercial

48,16	42,22	46,82	51,45	23,78	41,86	54,86
37,92	52,64	48,59	50,82	46,94	61,83	61,69
49,17	61,46	51,35	52,68	58,84	43,88	

1. ¿Cuál es la mejor estimación de la media poblacional?
2. Determine un intervalo de confianza de 95 %.
3. Interprete el resultado
4. ¿Concluiría de forma razonable que la media poblacional es de 50? ¿Y de 60?

```r
rm(list =ls())

Datos<-c(48.16, 37.92, 49.17, 42.22, 52.64, 61.46, 46.82, 48.49, 51.35,
         51.45, 50.82, 52.68, 23.78, 46.94, 58.84, 41.86, 61.83, 43.88, 54.86, 61.69)

Prom<-mean(Datos)
Des_est<-sd(Datos)

#hallar el valo de t, para 19 grados de libertad y 95% de confianza = 2.093
n<-20
Val_t<-2.093
#error muestral s/sqrt(n)

E_muestral<-Des_est/sqrt(n)
#margen de error
M_error<-Val_t*E_muestral

#calculomos el intrevalo de confianza
L_inf<-Prom - M_error
L_sup<-Prom + M_error
Resultado<-data.frame(Prom, Des_est, M_error, L_inf, L_sup)

hist(Datos, prob = TRUE,
     main = "Histograma con curva normal", ylab = "Densidad")
x <- seq(min(Datos), max(Datos), length = 70)
f <- dnorm(x, mean = mean(Datos), sd = sd(Datos))
lines(x, f, col = "red", lwd = 2)
```

Figura: Código en R



Environment | History | Connections | Tutorial

Import Dataset | 60 MiB

R | Global Environment

Data	
Resultado	1 obs. of 5 variables
Values	
Datos	num [1:20] 48.2 37.9 49.2 42.2 52.6 ...
Des_est	9.01258190706631
E_muestral	2.0152745796985
f	num [1:70] 1.84e-05 2.65e-05 3.78e-05 5.34e-05 7.49e-05 ...
L_inf	45.1250303046911
L_sup	53.560969695309
M_error	4.21796969530895
n	20
Prom	49.343
Val_t	2.093
x	num [1:70] 13.8 14.6 15.5 16.3 17.1 ...

Figura: Resultados en R

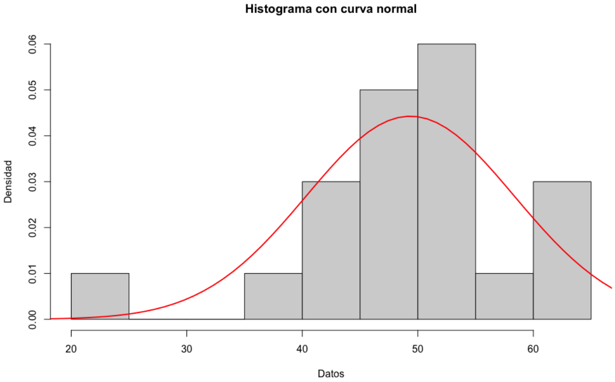


Figura: Distribución de los datos

Respuestas a las preguntas

1. ¿Cuál es la mejor estimación de la media poblacional?
   La media de la muestra
2. Determine un intervalo de confianza de 95 %.
   $45{,}12 \leftrightarrow 53{,}56$
3. Interprete el resultado
   Es razonable concluir que la media de la población se encuentra en el intervalo calculado
4. ¿Concluiría de forma razonable que la media poblacional es de 50? ¿Y de 60?
   La media de 50 es razonable
   La media de 60 no es razonable

Actividades de aprendizaje

Ejercicios:
9 al 14
Pág 308 -309
Estadística Aplicada a los negocios y economía Lind and Marshal 15ª Edición

Introducción

1. Los datos trabajados hasta ahora son de nivel de **razón** o **intervalo**
2. ¿Que pasaría si se desea trabajar con datos de tipo *nominal* u *ordinal*?

Proporción

Fracción, razón o porcentaje que indica la parte de la muestra de la población que posee un rasgo de interés particular

Si p representa la **proporción de la muestra**, X el *número de éxitos* y n el *número de elementos de la muestra*, se determina una proporción muestral de la siguiente manera:

Proporción Muestral

$$p = \frac{X}{n}$$

La **proporción de la población** se define por medio de π
Para crear un intervalo de confianza para una proporción, es necesario cumplir con los siguientes supuestos

1. Condiciones binomiales
   1. Los datos de la muestra son resultados de conteo
   2. Solo hay dos posibles resultados (éxito o fracaso)
   3. La probabilidad de un éxito permanece igual de una prueba a la siguiente
   4. Las pruebas son independientes. Esto significa que el resultado de la prueba no influye en el resultado de otra
2. Los valores $n\pi$ y $n \cdot (1{-}\pi)$ deben ser mayores o iguales que 5. Esta condición permite recurrir al teorema del límite central y emplear la distribución normal estándar, es decir, z, para completar un intervalo de confianza.

Intervalo de confianza para de la proporción de una población

$$p \pm z \cdot \sqrt{\frac{p \cdot (1 - p)}{n}}$$

Ejemplo

Dos empresas desean fusionarse, esta operación debe ser aprobada por lo menos por las 3/4 partes del sindicato. Se tiene una muestras de 2000 miembros actuales de los cuales 1600 planean votar afirmativamente por la propuesta.

Preguntas

1. ¿Qué es el estimador de la proporción poblacional?
2. Determine un intervalo de confianza de 95 % para la proporción poblacional.
3. Fundamente su decisión en esta información de la muestra: ¿puede concluir que la proporción necesaria de miembros del sindicato favorece la fusión? ¿Por qué?

Ejemplo - solución

1. Calcular la proporción del muestra

$$p = \frac{X}{n} = \frac{1600}{2000} = 0{,}80$$

Se puede entender que el 80 % esta de acuerdo con la fusión.

2. Determinamos el intervalo de confianza (95 %) (Valor z para 95 % de confianza de 1,96)

$$\begin{aligned} p \pm z \cdot \sqrt{\frac{p(1-p)}{n}} &= 0,8 \pm 1,96 \cdot \sqrt{\frac{0,8(1-0,8)}{2000}} \\ &= 0,8 \pm 0,018 \end{aligned}$$

3. Intervalo $0,782 \leftrightarrow 0,818$

Ejemplo - Respuestas

1. ¿Qué es el estimador de la proporción poblacional?
   Proporción poblacional p
2. Determine un intervalo de confianza de 95 % para la proporción poblacional. Intervalo de confianza $0,782 \leftrightarrow 0,818$
3. Fundamente su decisión en esta información de la muestra: ¿puede concluir que la proporción necesaria de miembros del sindicato favorece la fusión? ¿Por qué? Si
   Dado que el intervalo de confianza es superior al mínimo requerido $(0,75)$

Actividad de aprendizaje

Implemente la solución en R

Código de la implementación de los cálculo en R

```r
rm(list = ls()) #borra todo lo que existe

Icprop = function(x,n,z=1.96)
  {
  p=x/n #proporcion
  se=sqrt((p*(1-p))/n) #error estandar
  icprop = c(p - z*se, p + z*se) #calcula los limites inferior y superior
  return(icprop)
}
```

Figura: Códido de los cálculos, realizados en una función en R

```
Console   Jobs
R 4.1.3 · ~/Documents/CURSO R - MATERIAL/
> source("~/Documents/CURSO R - MATERIAL/int_conf_proporcion.R")
> x=1600
> n=2000
> z=1.96
> Icprop(x,n,z)
[1] 0.7824692 0.8175308
>
```

Figura: Resultado de la aplicación de la función

Ejemplo 2

Cliff es candidato. Se entrevista a los electores que acaban de votar y 275 indicaron que votaron por Clic. Considere que 500 electores es una muestra aleatoria de quienes votan en ese distrito. Esto significa que 55 % de los electores votó por Cliff.

$$p = \frac{X}{n} = \frac{275}{500} = 0,55$$

Este es un **estimador puntual**. No se conoce el porcentaje de la población que finalmente votará por él.
¿Es posible tomar una muestra de 500 electores de una población en la que 50 % o menos de los electores apoye a Cliff para encontrar que 55 % de la muestra lo apoya? En otras palabras, ¿el error de muestreo, que es $p – \pi = 0{,}55 – 0{,}50 = 0{,}05$, se debe al azar, o la población de electores que apoya a Cliff es superior a 0,50?

Ejemplo 2 - cont

Hallar el intervalo para 95 % de confianza

$$p \pm z \cdot \sqrt{\frac{p(1-p)}{n}} = 0,55 \pm 1,96 \cdot \sqrt{\frac{0,55(1-0,55)}{500}}$$
$$= 0,55 \pm 0,044$$

Los puntos extremos del intervalo son $0,506 \leftrightarrow 0,594$.
Note: El valor de $0,50$ no pertenece al intervalo.
Por tanto, se concluye que probablemente más de 50 % de los electores apoya a Cliff, lo cual es suficiente para que salga electo.

¿Siempre se utiliza este procedimiento?

Sí. Es exactamente el procedimiento de las cadenas de televisión, revistas de noticias y sondeos en la noche de las elecciones

Solución Optimizada

Implemente la solución en R

Código de la implementación de los cálculo en R

```r
rm(list = ls()) #borra todo lo que existe

Icprop = function(x,n,nc=0.95)
{
  p=x/n #proporcion
  p<-round(p,3)
  se=sqrt((p*(1-p))/n) #error estandar
  se<-round(se,3)
  z<- round(qnorm((1+nc)/2),2) #calcula el valor de z, a partir del nivel de confianza
  z<-round(z,3)
  me<-z*se
  me<-round(me,3)
  li<-p - me
  ls<-p + me
  Resultado<-data.frame(p,z, se, me, li, ls)
  return(Resultado)
}
```

Figura: Códido de los cálculos, realizados en una función en R

```
Console   Jobs
R 4.1.3 · ~/Documents/CURSO R - MATERIAL/
> source("~/Documents/CURSO R - MATERIAL/int_conf_proporcion.R")
> x=275
> n=500
> nc=0.95
> Icprop(x,n,nc)
     p    z    se    me    li    ls
1 0.55 1.96 0.022 0.043 0.507 0.593
> 
```

Figura: Resultado de la aplicación de la función

Actividades de aprendizaje

Ejercicios:
15 al 18
Pág 312
Estadística Aplicada a los negocios y economía Lind and Marshal 15ª Edición

Factor de Corrección de una Población Finita

Las poblaciones de las que se han tomado muestras hasta ahora han sido muy grandes o infinitas.

¿Qué sucedería si la población de la que se toma la muestra no fuera muy grande?

Es necesario realizar algunos ajustes en la forma de calcular el error estándar de las medias muestrales y del error estándar de las proporciones muestrales

Una población con un límite superior es una población finita

Una población finita puede **ser muy pequeña**; puede constar de todos los estudiantes registrados para este curso.

También puede ser **muy grande,** como todas las personas de la tercera edad que viven en un estado.

Factor de Corrección de una Población Finita

En el caso de una población finita, en la que el número total de objetos o individuos es **N** y el número de objetos o individuos en la muestra es **n**, es necesario ajustar los errores muestrales en las fórmulas de los intervalos de confianza.

En otras palabras, para determinar el intervalo de confianza para la media, se ajusta el error estándar de la media en las fórmulas

$$\overline{X} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

y

$$\overline{X} \pm t \cdot \frac{s}{\sqrt{n}}$$

Si está determinando el intervalo de confianza para una proporción, necesita ajustar el error estándar de la proporción en la fórmula

$$p = \frac{X}{n}$$

Factor de Corrección de una Población Finita

Este ajuste recibe el nombre de **factor de corrección de una población finita**. Con frecuencia se le abrevia **FCP,** el cual es:

$$FCP = \sqrt{\frac{N-n}{N-1}}$$

¿Por que es necesario?

- Si la muestra es un porcentaje significativo de la población, el estimador es mas preciso
- La mayoría de las fórmulas utilizadas para calcular errores estándar se basan en la idea de que:
  1. las muestras se seleccionan con reemplazo o que
  2. las muestras se seleccionan de una población infinita.

  En una investigación real, ninguna de estas ideas es cierta. Afortunadamente, esto no suele ser un problema si el tamaño de la muestra es inferior al 5 % del tamaño total de la población.

Ejemplo

Supongamos que:

- N=1000
- n=100
- $\sqrt{\dfrac{N-n}{N-1}} = \sqrt{\dfrac{1000-100}{1000-1}} = \sqrt{\dfrac{900}{999}} = 0{,}9492$ *Factor de Corrección*
- $\sqrt{\dfrac{N-n}{N-1}} \cdot \sigma_{\overline{X}} = 5\,\%(1 - 0{,}9492 = 0,0508)$
- Esta reducción en la magnitud del error estándar da como resultado un intervalo menor de valores al calcular la media poblacional o la proporción poblacional.
- Para el caso de $n = 200$?

Regla

Si la razón n/N es menor que 0,05, se ignora el factor de corrección.

Tamaño (muestra)	Fracción de la población	Factor de Corrección
10	0,010	0,9955
25	0,025	0,9879
50	0,050	0,9752
100	0,100	0,9492
200	0,200	0,8949
500	0,500	0,7075

Para construir un intervalo de confianza para la media a partir de una población finita sin conocer la desviación estándar de la población, tenemos:

$$\overline{X} \pm t \cdot \frac{s}{\sqrt{n}} \left(\sqrt{\frac{N-n}{N-1}} \right)$$

y

$$p = \frac{X}{n} \cdot \sqrt{\frac{N-n}{N-1}}$$

Ejemplo de FCP

Hay 250 familias, una muestra aleatoria de 40 de estas familias revela que la contribución anual media a la iglesia fue de \$450, y la desviación estándar, de \$75. ¿La media poblacional puede ser de \$445 o \$425?

Preguntas:

1. ¿Cuál es la media de la población? ¿Cuál es el mejor estimador de la media poblacional?
2. ¿Se debe aplicar el FCP? ¿Por qué?
3. Construya un intervalo de confianza de 90 % para la media de la población.¿Cuáles son los puntos extremos del intervalo de confianza?
4. Interprete el intervalo de confianza

Ejemplo de FCP - Solución

Observe que la población es finita $N = 250$

1. No se conoce la media poblacional. Media de la muestra, mejor estimador. (\$450)
2. La muestra es el 16 % de la población. $n/N = 40/250 = 0,16 > 0{,}05$ se debe aplicar el FCP.
3. El intervalo de confianza se calcula mediante:

$$\overline{X} \pm t \cdot \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$\overline{X} = 450$; $s = 75, N = 250, n = 40$. Se utiliza t con $gl = 39$. $t_{39} = 1,685$

$$\overline{X} \pm t \cdot \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 450 \pm 18{,}35 \therefore [431,65 \leftrightarrow 468,35]$$

4.

Es probable que la media poblacional sea de más de \$431,65 e inferior a \$468,35.
En otras palabras, ¿la media de la población puede ser de \$445?
Sí, pero **no es probable** que sea de \$425.

¿Por qué?

Porque el valor de \$445 se encuentra dentro del intervalo de confianza y \$425 no pertenece al intervalo de confianza.

Actividades de aprendizaje

Ejercicios 19-22
Pag 314-315
Estadística Aplicada a los negocios y economía Lind and Marshal 15ª Edición

¿Cuántos elementos debe haber en una muestra?.

El tamaño adecuado de una muestra depende de tres factores:

Primero El nivel de confianza deseado.

1. Los niveles de confianza de 95 y 99 % son los más comunes, aunque es posible cualquier valor entre 0 y 100 %
2. Mientras más alto sea el nivel de confianza elegido, mayor será el tamaño de la muestra correspondiente

Segundo El margen de error que tolerará el investigador.

1. Designado E,es la magnitud que se suma y resta de la media muestral
2. También es la mitad de la amplitud del correspondiente intervalo de confianza
3. Un error admisible más pequeño requerirá una muestra mayor. Un error admisible grande permitirá una muestra menor

Tres La variabilidad de la población que se estudia (desviación estándar de la población.)

1. Si la población se encuentra **muy dispersa, se requiere una muestra grande**
2. si la población se encuentra concentrada (homogénea),el tamaño de muestra que se requiere será menor.
3. No obstante, puede ser necesario utilizar un estimador para la desviación estándar de la población.

Criterios para la selección de un estimador

1. **Utilice un estudio comparativo**. Aplique este enfoque cuando se encuentre disponible un estimador de la dispersión de otro estudio. Si se considera confiable una desviación estándar de un estudio anterior, se puede utilizar en el estudio actual como ayuda para obtener el tamaño aproximado de una muestra.
2. **Emplee un enfoque basado en el intervalo**. Para aplicar este enfoque necesita conocer o contar con un cálculo de los valores máximo y mínimo de la población. Puede calcular la desviación estándar como un sexto del rango (Casí todos los datos se encuentran a $-3s$, y $+3s$ de la media)
3. **Realice un estudio piloto**. Método más común. Para probar la validez del cuestionario, se aplica a una pequeña muestra de estudiantes.A partir de esta pequeña muestra se calcula la desviación estándar, y se utiliza este valor para determinar el tamaño adecuado de la muestra.

Tamaño de la muestra para estimar la media de la población

La interacción entre estos tres factores y el tamaño de la muestra se expresa con la siguiente fórmula:

$$E = z \cdot \frac{\sigma}{\sqrt{n}}$$

Al despejar n, se obtiene

Tamaño de la muestra para estimar la media de la población

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

donde:

- **n** es el tamaño de la muestra
- **z** es el valor normal estándar correspondiente al nivel de confianza deseado
- σ es la desviación estándar de la población
- **E** es el error máximo admisible

El resultado debe ser redondeado pues **n** es un dato discreto.
Nota: *Redondeo siempre hacia arriba.*

Ejemplo Tamaño de la muestra

Un estudiante de administración pública desea determinar la cantidad media que ganan al mes los miembros de los consejos ciudadanos de las grandes ciudades.
El error al calcular la media debe ser inferior a \$100, con un nivel de confianza de 95 %.
El estudiante encontró un informe del Departamento del Trabajo en el que la desviación estándar es de \$1000.
¿Cuál es el tamaño de la muestra que se requiere?

- Error admisible. $E = \$100$
- $z_{95\,\%} = 1,96$
- Estimador de la desviación estándar: $\sigma = 1000$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{1,96 \cdot 1000}{100}\right)^2 = (19,6)^2 = 384,16 \approx 385$$

- $z_{99\,\%} = 2,58$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{2,58 \cdot 1000}{100}\right)^2 = (25,8)^2 = 665,64 \approx 666$$

Note que:

Pasar de $z_{95\,\%}$ al $z_{99\,\%}$ incrementa en 281 observaciones, lo cual implica costo (tiempo y dinero), razón por la cual se debe elegir con cuidado el nivel de confianza.

Tamaño de la muestra en el caso de una proporción

Es necesario especificar tres elementos:

1. Nivel de confianza deseado
2. Margen de error en la proporción de la población
3. Una aproximación de la proporción de la población
4. Fórmula:
$$n = p(1-p)\left(\frac{z}{E}\right)^2$$
5. Si se conoce p a partir de otra fuente (estudio piloto) se puede utilizar.
6. $p \cdot (1-p)$ jamas será mayor a $0,5$

Ejemplo Proporciones

En el estudio del ejemplo anterior también se calcula la proporción de ciudades que cuentan con recolectores de basura privados.

El estudiante desea que el **margen de error** se encuentre a 0,10 de la proporción de la población; el nivel de confianza deseado es de 90 %, y no se encuentra disponible ningún estimador para la proporción de la población.
¿Cuál es el tamaño de la muestra que se requiere?

Ejemplo Proporciones

Solución

- $E = 0,10$
- $nc = 90\,\% \rightarrow z_{90\,\%} = 1,65$
- Estimado de la proporción **desconocido**. $p = 0,50$

$$n = (0,5)(1 - 0,5)\left(\frac{1,65}{0,10}\right)^2 = 68,062 \approx 69$$

Aplicaciones en R.

Desarrolle en R, una aplicación para calcular el tamaño de la muestra para estimar la media de la población, y también para estimar el tamaño de la muestra en el caso de una proporción

Calculadoras online

Otras formas de calcular el tamaño de la muestra

1. Busque información en la web, sobre como calcular el tamaño de la muestra utilizando las calculadoras online.
2. Contraste los resultados con los que usted obtendría calculando con R, o con la calculadora
3. ¿Existe diferencias entre estas distintas formas de calcular el tamaño de la muestra?. Argumente su respuesta.

Actividades de aprendizaje

Ejercicios 23-30
Pág 317-318
Estadística Aplicada a los negocios y economía Lind and Marshal 15ª Edición

Actividades de aprendizaje - Fin de capítulo

Ejercicios 31-70
Pág 319-323
Estadística Aplicada a los negocios y economía Lind and Marshal 15ª Edición