Estimación por intervalos

Estimación por Intervalos

El problema

ENTRE EL 12% Y EL 16% PADECE OBESIDAD

Sanidade estima que entre un 25% y un 30% de la población infantil gallega tiene sobrepeso

Con un estimador, estimamos el parámetro con una cierta precisión, que depende:

- De la variabilidad del estimador
- Del tamaño de la muestra
- Del nivel de confianza de la estimación: cuan seguros queremos estar de que la estimación es correcta

El problema

Set de cada deu estudiants de la UIB practica el ciberplagi a l'hora de confeccionar els treballs acadèmics

Set de cada deu estudiants de la UIB (76,6 per cent) accepten haver copiat i aferrat fragments d'una web o un altre recurs obtingut a Internet i, sense esmentar-ne la procedència, haver-lo fet servir amb altres textos fets per ells per elaborar un

<u>Fixa tècnica de la mostra de la UIB</u>

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

Por lo tanto (y por el momento):

Con 95 % de confianza podemos afirmar que entre un 73.1 % y un 80.1 % de los estudiantes de la UIB aceptan ...

El problema

EL PAÍS

PORTADA | INTERNACIONAL | PO

ECONOMÍA

ECONOMÍA EMPRESAS MERCADOS BOLSA FINANZAS PERSONALES VIVIENDA TEC

ESTÁ PASANDO

MERCADO LABORAL

El paro baja en 72.800 personas por el empleo temporal del verano

- La tasa de desempleo baja ligeramente en el tercer trimestre hasta el 25,98%
- El empleo avanza en 39.500 personas, aunque se desploman los indefinidos
- Solo se crean puestos de trabajo en el sector servicios
- Radiografía del mercado laboral español en 10 titulares

MANUEL V. GÓMEZ | Madrid | 24 OCT 2013 - 21:29 CET 476

Definiciones básicas

En la Encuesta de Población Activa (EPA):

Errores de muestreo relativos, de la població[n]
de 16 y más años por comunidad autónoma
y relación con la actividad económica

Unidades:Porcentaje

	Ocupados	Parados
	2013TIII	2013TIII
Total Nacional	0,37	0,87

`http://www.ine.es/jaxi/tabla.do?per=03&type=db&divi=EPA&idtab=313`

estimación $\pm$ 1 vez el error de muestreo = intervalo de confianza del 67%.

estimación $\pm$ 2 veces el error de muestreo = intervalo de confianza del 95%.

estimación $\pm$ 3 veces el error de muestreo = intervalo de confianza del 99,7%.

`http://www.ine.es/docutrab/eval_epa/evaluacion_epa04.pdf`

El problema

EPA de octubre de 2013:

- El número estimado de parados a nivel nacional fue del 5 904 700
- El error de muestreo fue de un 0.87 %
- Por lo tanto, estamos bastante seguros (nivel de confianza del 95 %) de que el número de parados estaba entre

$$\begin{aligned}5\,904\,700-2\cdot 0{,}0087\cdot 5\,904\,700 &= 5\,904\,700-102\,742\\ &= 5\,801\,958 \quad \text{y}\\ 5\,904\,700+2\cdot 0{,}0087\cdot 5\,904\,700 &= 5\,904\,700+102\,742\\ &= 6\,007\,442\end{aligned}$$

- La EPA de junio del 2013 había estimado el número de parados en 5 977 500
- No hay evidencia de que el paro bajase

Definiciones básicas

Una estimación por intervalos de un parámetro poblacional es una regla para calcular, a partir de una muestra, un intervalo en el que, con una cierta probabilidad (nivel de confianza), se encuentra el valor verdadero del parámetro

Estas reglas definirán, a su vez, estimadores

Ejemplos

Ejemplo: Hemos escogido al azar 50 estudiantes de grado de la UIB, hemos calculada sus notas medias de las asignaturas del primer semestre, y la media de estas medias ha sido un 6.3, con una varianza muestral de 1.8.

Determinar un intervalo del que podamos afirmar con probabilidad 95 % que contiene la media real de las notas medias de los estudiantes de grado de la UIB este primer semestre.

Ejemplos

Ejemplo: En un experimento se ha medido el porcentaje de aumento de alcohol en sangre a 40 personas despuesde tomar 4 cañas de cerveza. La media y la desviación típica muestral de estos porcentajes de incremento han sido

$$\overline{x} = 41{,}2, \quad \widetilde{s} = 2{,}1$$

Determinar un intervalo que podamos afirmar con probabilidad 95 % que contiene el porcentaje de aumento medio de alcohol en sangre (verdadero) de una persona despuesde beber cuatro cañas de cerveza.

Definiciones básicas

Dado un parámetro θ, el intervalo $]A, B[$ es un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ para al parámetro θ cuando

$$P(A < \theta < B) = 1 - \alpha$$

El valor $(1-\alpha)\cdot 100\,\%$ (o contiene solo el $1-\alpha$) recibe el nombre de nivel de confianza

El valor α recibe el nombre de nivel de significación

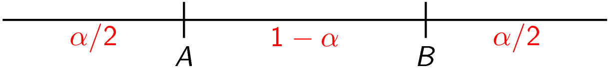
Ejemplo: $]A, B[$ es un intervalo de confianza del $95\,\%$ (o de nivel de significación de 0.05) si

$$P(A < \theta < B) = 0{,}95$$

Definiciones básicas

Por defecto, buscaremos intervalos tales que la cola de probabilidad sobrante α se reparta por igual a cada lado del intervalo:

$$P(\theta < A) = P(\theta > B) = \frac{\alpha}{2}$$



Ejemplo: Para buscar un intervalo de confianza $]A, B[$ del 95 %, buscaremos A, B de manera que

$$P(\theta < A) = 0{,}025 \quad \text{y} \quad P(\theta > B) = 0{,}025$$

Ejemplo: μ de población normal con σ conocida

Sea X una v.a. normal con media poblacional μ desconocida y desviación típica poblacional σ conocida (a la práctica, usualmente, estimada en un experimento anterior)

Sea $X_1, \ldots, X_n$ una m.a.s. de X, con media muestral $\overline{X}$

Queremos determinar un intervalo de confianza para a μ con un cierto nivel de confianza (digamos, 97.5 %): un intervalo $]A, B[$ tal que

$$P(A < \mu < B) = 0{,}975$$

Ejemplo: μ de población normal con σ conocida

Bajo estas condiciones, sabemos que

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

sigue una distribuciónn normal estándar

Comencemos calculando un intervalo centrado en 0 en el que Z tenga probabilidad 0,975:

$$0{,}975 = P(-\delta < Z < \delta) = F_Z(\delta) - F_Z(-\delta) = 2F_Z(\delta) - 1$$

$$F_Z(\delta) = \frac{1{,}975}{2} = 0{,}9875 \Rightarrow \delta = \texttt{qnorm(0.9875)} = 2{,}24$$

Ejemplo: μ de población normal con σ conocida

Por lo tanto

$$P(-2{,}24 < Z < 2{,}24) = 0{,}975$$

Substituyendo $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$

$$P\left(-2{,}24 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 2{,}24\right) = 0{,}975$$

$$P\left(\overline{X} - 2{,}24\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2{,}24\frac{\sigma}{\sqrt{n}}\right) = 0{,}975$$

Ejemplo: μ de població normal con σ conocida

Por lo tant, la probabilidad que la media μ de X se encuentre dentro del intevalo

$$\left]\overline{X} - 2{,}24\frac{\sigma}{\sqrt{n}}, \overline{X} + 2{,}24\frac{\sigma}{\sqrt{n}}\right[$$

es 0,975: es un intervalo de confianza del 97.5 %

Ejemplo: μ de població normal con σ conocida

Por lo tant, la probabilidad que la media μ de X se encuentre dentro del intevalo

$$\left]\overline{X} - 2{,}24\frac{\sigma}{\sqrt{n}}, \overline{X} + 2{,}24\frac{\sigma}{\sqrt{n}}\right[$$

es 0,975: es un intervalo de confianza del 97.5 %

Además tenemos que:

- Está centrado en $\overline{X}$
- El 0.025 de probabilidad restante está repartido por igual en los dos extremos del intervalo

Ejemplo: μ de població normal con σ conocida

- Como estimador: Un 97.5 % de las veces que tomemos una muestra de tamaño n de X, el verdadero valor de μ caerá dentro de este intervalo
- Para una mostra concreta: La probabilidad que, si una μ ha producido esta mostra, esté en este intervalo concreto, es del 97.5 %
- En ocasiones lo entenderemos como: "La probabilidad de que μ esté en este intervalo es del 97.5 %"
- Pero no la frease anterio es mentira (es un abusio de lenguaje): La μ concreta es un valor fijo, por lo tanto que pertenezca o no a aquest intervalo concreto tiene probabilidad 1 (si hi pertenece) y 0 (si no pertenece)

I.C. pera μ de población normal con σ conocida

Teorema

Sea $X \sim N(\mu, \sigma)$ con μ desconocida y σ conocida.

Tomamos una m.a.s. de X de medida n, con media $\overline{X}$.

Un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ pera μ es

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

donde $z_{1-\frac{\alpha}{2}}$ es el $(1-\frac{\alpha}{2})$-cuantil de la normal estándard Z (es decir, $z_{1-\frac{\alpha}{2}} = F_Z^{-1}(1-\frac{\alpha}{2})$, o $P(Z \leqslant z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$)

I.C. para μ de población normal con σ conocida

Si X es normal con σ conocida, un intervalo de confianza I.C. para μ de población normal con σ conocida μ del $(1-\alpha)\cdot 100\,\%$ es

$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} := \left]\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right[$$

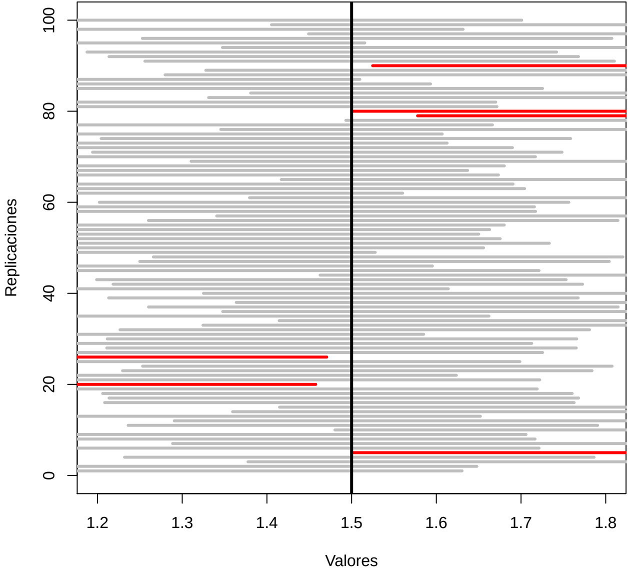
Observad que está centrado en $\overline{X}$

confianza $1-\alpha$	Significación α	$z_{1-\frac{\alpha}{2}}$
0.900	0.100	1.64
0.950	0.050	1.96
0.975	0.025	2.24
0.990	0.010	2.58

I.C. para μ de población normal con σ conocida

```
ICZ=function(x,sigma,alpha){
  c(mean(x)-qnorm(1-alpha/2)*sigma/sqrt(length(x)),
  mean(x)+qnorm(1-alpha/2)*sigma/sqrt(length(x)))}
set.seed(5)
mu=1.5; sigma=1; alpha=0.05
Poblacion=rnorm(10^6,mu,sigma)
M=replicate(100,ICZ(sample(Poblacion,50,replace=T),
 sigma,alpha))
plot(1:10,type="n",xlim=c(1.2,1.8),ylim=c(0,100),
xlab="Valores",ylab="Replicaciones")
seg.int=function(i){color="grey";
  if((mu<M[1,i]) | (mu>M[2,i])){color = "red"}
  segments(M[1,i],i,M[2,i],i,col=color,lwd=3)}
invisible(sapply(1:100,FUN=seg.int))
abline(v=mu,lwd=3)
```

I.C. para μ de población normal con σ conocida



I.C. para μ de población normal con σ conocida

¡Atención!

De media, un $\alpha 100\,\%$ de las veces, un intervalo de confianza del $(1-\alpha)100\,\%$ no contendrá el valor real del parámetro

Ejemplo: De media, un 5 % de las vegades un intervalo de confianza del 95 % no contendrá el valor real del parámetro

Ejemplo

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

Tomamos una m.a.s. de tamaño $n = 16$ de una v.a. normal con $\sigma = 4$ y μ desconocida. La media de la m.a.s. es $\overline{x} = 20$.

Calculad un intervalo de confianza del 97.5 % para μ de una población normal con σ conocidaμ

Ejemplo

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

Tomamos una m.a.s. de tamaño $n = 16$ de una v.a. normal con $\sigma = 4$ y μ desconocida. La media de la m.a.s. es $\overline{x} = 20$.

Calculad un intervalo de confianza del 97.5 % para μ de una población normal con σ conocidaμ

$$\left]20 - 2{,}24 \cdot \frac{4}{\sqrt{16}}, 20 + 2{,}24 \cdot \frac{4}{\sqrt{16}}\right[=]17{,}76, 22{,}24[$$

La probabilidad que un parámetro μ que haya producido la muestra esté en este intervalo es 0,975

Ejemplo

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

Tomamos una m.a.s. de tamaño $n = 16$ de una v.a. normal con $\sigma = 4$ y μ desconocida. La media de la m.a.s. es $\overline{x} = 20$.

Calculad un intervalo de confianza del 97.5 % para μ de una población normal con σ conocidaμ

$$\left]20 - 2{,}24 \cdot \frac{4}{\sqrt{16}}, 20 + 2{,}24 \cdot \frac{4}{\sqrt{16}}\right[=]17{,}76, 22{,}24[$$

La probabilidad que un parámetro μ que haya producido la muestra esté en este intervalo es $0{,}975$

"La probabilidad que el parámetro μ de la población que ha producido la muestra este en este intervalo es $0{,}975$*"*

Ejemplo

Queremos que analizar un sensor que mide la temperatura de un procesador en grados centígrados [1] que tiene como temperatura normal de 32° a 40°. Para saber si está bien calibrado, diseñamos un experimento en el que ponemos el procesador el procesador en las mismas condiciones y tomamos 40 muestras de su temperatura. Los resultados son los Seaentes:

```
temperatura=c(36,35,38,38,36,37,38,36,37,36,
              37,37,34,38,35,37,36,36,34,38,
              36,37,35,35,35,35,36,36,36,35,
              36,35,34,34,37,37,35,36,34,36)
mean(temperatura)

## [1] 35.975
```

[1] En concreto un Intel Core i7-2600K

Ejemplo

Supongamos que las medidas de nuestro sensor siguen una distribucíónn normal con varianza poblacional conocida $\sigma^2 = 1{,}44$. Calculad un intervalo de confianza del 90 % para al resultado medio de la temperatura del procesador.

Tenemos las Seaentes condiciones:

- Población normal con $\sigma = \sqrt{1{,}44} = 1{,}2$ conocida
- M.a.s. de tamaño $n = 40$
- media de la muestra $\overline{x} = 35{,}975$
- $1 - \alpha = 0{,}9 \Rightarrow \alpha = 0{,}1 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}95$
- $z_{0{,}95} \approx 1{,}64$
  - Con la tabla de Z, $P(Z \leqslant 1{,}64) = 0{,}9495 \approx 0{,}95$
  - Con R

```
qnorm(0.95)

## [1] 1.644854
```

Ejemplo

Aplicamos la fórmula

$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

con

$$\overline{x} = 35{,}975,\ z_{0,95} = 1{,}64,\ \sigma = \sqrt{1{,}44} = 1{,}2,\ n = 40$$

Obtenemos que el intervalo de confianza del 90 % pedido es

$$35{,}975 \pm 1{,}64 \cdot \frac{1{,}2}{\sqrt{40}} =]34{,}991, 36{,}959[$$

Amplitud

La amplitud A de un intervalo de confianza

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

es

$$\begin{aligned} A &= \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} - \left(\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \\ &= 2 \cdot z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \end{aligned}$$

El error máximo, al nivel de confianza $(1-\alpha)$, que cometemos al estimar μ por $\overline{X}$ es la mitad de la amplitud,

$$z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Amplitud

La Amplitud A del intervalo de confianza

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

es

$$A = 2 \cdot z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Observaciones

- I.C. para μ de una población normal con σ conocida n y α fijos, si σ crece, A crece
- I.C. para μ de población normal con σ conocida σ y α fijos, si n crece, A decrece
- I.C. para μ de población normal con σ conocida σ y n fijos, si $1-\alpha$ crece, A crece

Amplitud

Si queremos calcular el tamaño n de la muestra para asegurar que el interval de confianza per μ al nivel de confianza $(1-\alpha)$ tenga una amplitud prefijada máxima A_0 (o un error máximo $A_0/2$), podemos despejar el tamaño muestral n de:

$$A_0 \geqslant 2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \Rightarrow n \geqslant \left(2z_{1-\frac{\alpha}{2}}\frac{\sigma}{A_0}\right)^2$$

Dada A_0, tomaremos

$$n = \left\lceil \left(2z_{1-\frac{\alpha}{2}}\frac{\sigma}{A_0}\right)^2 \right\rceil$$

Ejemplo

Recordemos que las medidas de nuestro sensor de temperatura seguían una distribuciónn normal con varianza poblacional conocida $\sigma^2 = 1{,}44$, $\sigma = 1{,}2$

¿Cuántas medidas tendríamos que tomar para obtener la temperatura media con un error máxim0 de $0{,}05^\circ$ al nivel de confianza del $90\,\%$?

$$n = \left\lceil \left(2 z_{1-\frac{\alpha}{2}} \frac{\sigma}{A_0}\right)^2 \right\rceil$$

on

$$\frac{A_0}{2} = 0{,}05, \quad z_{1-\frac{\alpha}{2}} = 1{,}64, \quad \sigma = 0{,}1$$

Obtenemos $n = \lceil 10{,}76 \rceil = 11$

distribuciónn t de Student

Sea $X \sim N(\mu, \sigma)$

Sea $X_1, \ldots, X_n$ una m.a.s. de X, con media $\overline{X}$ y desviación típica muestral $\widetilde{S}_X$

Teorema

En estas condicions, la v.a.

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

sigue una distribuciónn t de Student con $n-1$ grados de libertad, t_{n-1}

$\widetilde{S}_X/\sqrt{n}$: el error muestral, estima el error estándar $\sigma/\sqrt{n}$

distribuciónn t de Student

La distribuciónn t de Student con ν grados de libertad, t_ν:

- Tiene densidad

$$f_{t_\nu}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

donde $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ si $x > 0$.

- La distribuciónn está tabulada (las tablas en el moodle de la asignatura), y con R est.

Distribución t de Student

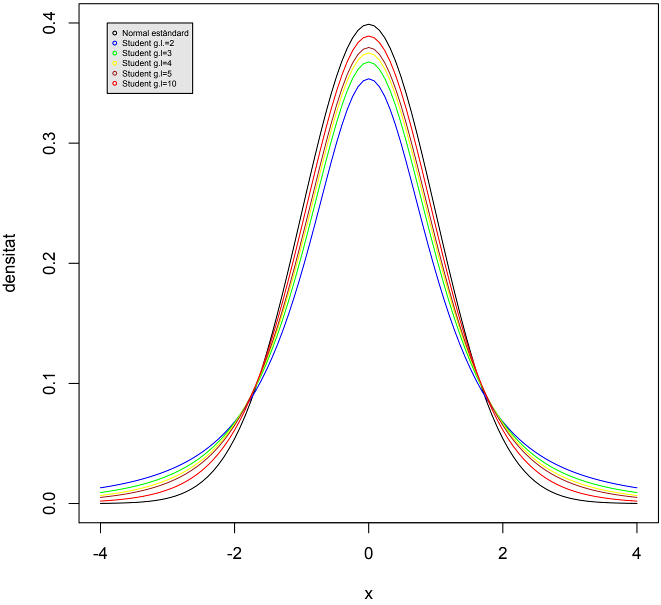
Sea t_ν una v.a. que sigue la distribución t de Student con ν grados de libertad

- $E(t_\nu) = 0$ si $\nu > 1$ y $Var(t_\nu) = \dfrac{\nu}{\nu - 2}$ si $\nu > 2$
- Su función de distribución es simétrica respecto de $E(t_\nu) = 0$ (como la de una $N(0,1)$):

$$P(t_\nu \leqslant -x) = P(t_\nu \geqslant x) = 1 - P(t_\nu \leqslant x)$$

- Si ν es grande, su distribución es aproximadamente la de $N(0,1)$ (pero con más variancia: un poco más aplastada)

Distribución t de Student

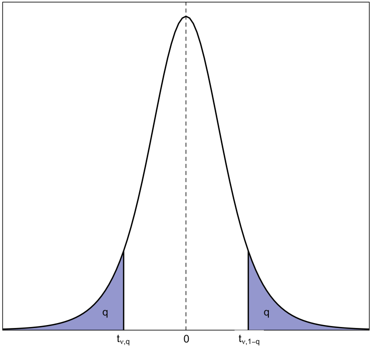


Distribución t de Student

Indicaremos con $t_{\nu,q}$ el q-cuantil de una v.a. X_{t_ν} que sigue una distribución t_ν:

$$P(X_{t_\nu} \leqslant t_{\nu,q}) = q$$

Por simetría, $t_{\nu,q} = -t_{\nu,1-q}$



μ de población normal con σ desconocida

Consideremos la situación siguiente :

- X una v.a. normal con μ y σ desconocidas
- $X_1, \ldots, X_n$ una m.a.s. de X de tamaño n, con media $\overline{X}$ y varianza muestral $\widetilde{S}_X^2$

Teorema

En estas condicions, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de una población normal con σ conocida μ es

$$\left]\overline{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

Ejemplo

La empresa *3D-print* ofrece una impresora industrial de papel en color de alta capacidad. En su publicidad afirma que sus cartuchos imprimen una media de 500 mil copias con la especificación:

`Datos tècnicos: Muestra de tamaño` $n = 100$, `población aproximadamente normal, nivel de confianza del 90 %`

La OCU (asociación de consumidores) desea comprobar estas afirmaciones y su laboratorio toma una muestra aleatoria de tamaño $n = 24$, obteniendo una media de $\overline{x} = 518$ mil impresiones y una desviación típica muestral $\widetilde{s} = 40$

¿Con esta muestra, la media poblacional anunciada per fabricante cae en el intervalo de confianza del 90 %?

Ejemplo

Cal calcular el intervalo de confianza I.C. para μ de población normal con σ conocidaμ con

$$n = 24, \overline{x} = 518, \widetilde{s} = 40, \alpha = 0{,}1$$

Será

$$\left]\overline{x} - t_{24,0,95}\frac{\widetilde{s}}{\sqrt{n}}, \overline{x} + t_{24,0,95}\frac{\widetilde{s}}{\sqrt{n}}\right[$$

Consultando las tablas de la distribución t de Student, Obtenemos $t_{24,0,95} = 1{,}71$

```
qt(0.95,24)

## [1] 1.710882

round(qt(0.95,24),2)

## [1] 1.71
```

Observaciones

- el intervalo de confianza obtenido está centrado en $\overline{X}$
- La fórmula

$$\left]\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

  I.C. para μ de población normal con σ conocida el intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ se puede utilizar cuando X es normal y n cualquiera

- Si n es grande $t_{n-1,1-\frac{\alpha}{2}} \approx z_{1-\frac{\alpha}{2}}$ y podemos aproximarlo mediante

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

μ I.C. para μ de población normal con σ conocida muestras grande

Consideremos la situación siguiente :

- X una v.a. cualquiera con media poblacional μ desconocida y desv. típ. σ conocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con media $\overline{X}$
- n es grande (pongamos que $n \geqslant 40$)

μ I.C. para μ de población normal con σ conocida muestras grande

Consideremos la situación siguiente :

- X una v.a. cualquiera con media poblacional μ desconocida y desv. típ. σ conocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con media $\overline{X}$
- n es grande (pongamos que $n \geqslant 40$)

En estas condicions (T.C.L.)

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0, 1)$$

μ I.C. para μ de población normal con σ conocida muestras grande

Consideremos la situación siguiente :

- X una v.a. cualquiera con media poblacional μ desconocida y desv. típ. σ conocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con media $\overline{X}$
- n es grande (pongamos que $n \geqslant 40$)

Teorema

En estas condiciones, podemos tomar como intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocida μ

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

μ I.C. para μ de población normal con σ conocida muestra grande

Consideremos la situación siguiente :

- X una v.a. cualquiera con media poblacional μ desconocida y desv. típ. σ desconocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con media $\overline{X}$ y desviación típica muestral $\widetilde{S}_X$
- n es grande (pogamos que $n \geqslant 40$)

"Teorema"

En estas condiciones, se recomienda tomar como intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ para μ de población normal con σ conocida μ

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

Ejemplo

La [green]https://twitter.com/guardiacivil/status/80515465397 civil informa... En un experimento se ha medido la tasa oficial de alcoholemia en sangre a 40 varones (sobrios) después de tomar 3 cañas de cerveza de 330 ml La media y la desviació típica muestral de porcentajes . El siguiente código simula este experimento supuesto que la tasa

```
tasa_alcoholemia=round(rnorm(40,mean=0.7,sd=0.1)
tasa_alcoholemia

##  [1] 0.66 0.71 0.54 0.55 0.62 0.75 0.83 0.82
## [15] 0.90 0.51 0.68 0.61 0.76 0.65 0.73 0.67
## [29] 0.92 0.66 0.69 0.64 0.62 0.67 0.75 0.70
```

$$\overline{x} = 41{,}2, \quad \widetilde{s} = 2{,}1$$

Calculad un intervalo del que podamos afirmar que con

Ejemplo

Nos piden un intervalo de confianza del 95 % para μ de población normal con σ conocida μ de la v.a. X "porcentaje de aumento de alcohol en sangre después de una persona después de beber cuatro cañas de cerveza" No conocemos la distribución de X, pero $n = 40$ es "grande"

Ejemplo

Podemos emplear

$$\left]\overline{x} - z_{1-\frac{\alpha}{2}}\frac{\widetilde{s}}{\sqrt{n}}, \overline{x} + z_{1-\frac{\alpha}{2}}\frac{\widetilde{s}}{\sqrt{n}}\right[$$

on

$$n = 40, \overline{x} = 41{,}2, \widetilde{s} = 2{,}1,$$

$$\alpha = 0{,}05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$$

Ejemplo

Podemos emplear

$$\left]\overline{x} - z_{1-\frac{\alpha}{2}}\frac{\widetilde{s}}{\sqrt{n}}, \overline{x} + z_{1-\frac{\alpha}{2}}\frac{\widetilde{s}}{\sqrt{n}}\right[$$

on

$$n = 40, \overline{x} = 41{,}2, \widetilde{s} = 2{,}1,$$

$$\alpha = 0{,}05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$$

$$]40{,}55, 41{,}85[$$

Podemos afirmar con un 95 % de confianza que el porcentaje medio de aumento de alcohol en sangre de una persona después de beber cuatro cañas de cerveza está entre el 40,55 % y el 41,85 %

Ejemplo

S'ha pres una mostra de sang a 1000 adults sans y s'hi ha mesurat la quantitat de calci (en mg per dl de sang). S'ha obtenido una media muestral de 9.5 mg/dl con una desviació típica muestral de 0.5 mg/dl.

Trobau un intervalo de confianza del 95 % I.C. para μ de población normal con σ conocidala quantitat media de calci en sang en un adult sa

Ejemplo

Com que $n = 1000$ es grande, podem emprar

$$\left]\overline{x} - z_{1-\frac{\alpha}{2}} \frac{\widetilde{s}}{\sqrt{n}}, \overline{x} + z_{1-\frac{\alpha}{2}} \frac{\widetilde{s}}{\sqrt{n}}\right[$$

on

$$\overline{x} = 9{,}5,\ \widetilde{s} = 0{,}5,\ \alpha = 0{,}05,\ z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$$

Dóna

$$]9{,}47, 9{,}53[$$

Podem afirmar con un 95 % de confianza que la quantitat media de calci en sang en un adult sa está entre 9,47 y 9,53 mg/dl

Amplitud

L'Amplitud de

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

es

$$A = 2z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}$$

Per determinar n (gran) que doni com a màxim una Amplitud A prefixada, ens cal $\widetilde{S}_X$, que depèn de la mostra.

Solucions:

- Si sabem la desv. típ. poblacional σ, l'empram en lloc de $\widetilde{S}_X$
- Si hem pres una mostra prèvia (pilot), n'empram la desviació típica muestral per estimar σ

Amplitud

D'una població X n'hem pres una m.a.s. pilot que ha tingut una desviació típica muestral $\widetilde{s}_{pilot}$.

Estimarem que la tamaño mínima n de una m.a.s. de X que doni un intervalo de confianza I.C. para μ de población normal con σ conocidaμ_X de nivell de confianza $1-\alpha$ y Amplitud màxima A_0 es

$$n = \left\lceil \left(2 z_{1-\frac{\alpha}{2}} \frac{\widetilde{s}_{pilot}}{A_0}\right)^2 \right\rceil$$

Ejemplo

Volem estimar l'alçada media dels estudiants de la UIB. Cercam un intervalo de confianza del 99 % con una precisió màxima de 1 cm. En una mostra pilot de 25 estudiants, obtinguérem

$$\overline{x} = 170 \text{ cm}, \widetilde{s} = 10 \text{ cm}$$

Basant-nos en estas dades, quina tamaño hauria de tenir la mostra per assolir el nostre objectiu?

Ejemplo

$$n = \left\lceil \left(2 z_{1-\frac{\alpha}{2}} \frac{\widetilde{S}_{pilot}}{A} \right)^2 \right\rceil = \left\lceil \left(z_{1-\frac{\alpha}{2}} \frac{\widetilde{S}_{pilot}}{A/2} \right)^2 \right\rceil$$

- Precisió = error màxim = $A/2 = 1$
- $\widetilde{S}_{pilot} = 10$
- $\alpha = 0{,}01 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0{,}995} = 2{,}58$

Dóna

$$n = \left\lceil \left(2{,}58 \cdot \frac{10}{1} \right)^2 \right\rceil = \lceil 665{,}64 \rceil = 666$$

p I.C. para μ de población normal con σ conocidamostres petites

Consideremos la situación siguiente :

- X una v.a. Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con nombre d'èxits x y per tant freqüència relativa d'èxits $\widehat{p}_X = x/n$

Recordau que x es$B(n, p)$

Mètode "exacte" de Clopper-Pearson

Un intervalo de confianza $]p_0, p_1[$ del $(1-\alpha)100\,\%$ I.C. para μ de población normal con σ conocida$p$ s'obté trobant el p_0 mes grande y el p_1 mespetit tals que

$$\sum_{k=x}^{n} p_0^k (1-p_0)^{n-k} \leqslant \frac{\alpha}{2}, \qquad \sum_{k=0}^{x} p_1^k (1-p_1)^{n-k} \leqslant \frac{\alpha}{2}$$

A mà (consultant taules) esuna feinada.

p I.C. para μ de población normal con σ conocidamostres petites

El paquet `epitools` porta

```
binom.exact(èxits, tamaño ,conf.)
```

per calcular-ho.

De 10 pacients tractats con un medicament, 2 s'han curat. Donau un intervalo de confianza del 95 % I.C. para μ de población normal con σ conocidala proporció p de pacients que aquest medicament cura.

```
> install.packages("epitools",dep=TRUE)
> library(epitools)
> round(binom.exact(2,10,0.95),3)
  x  n proportion lower upper conf.level
1 2 10        0.2 0.025 0.556       0.95
```

Dóna $]0{,}025, 0{,}556[$

p I.C. para μ de población normal con σ conocidamostres grans I

Consideremos ara la situación siguiente :

- X una v.a. Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con n gran (per Ejemplo, $n \geqslant 40$) y freqüència relativa d'èxits $\widehat{p}_X$

En estas condicions (pel T.C.L.),

$$Z = \frac{\widehat{p}_X - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

p I.C. para μ de población normal con σ conocidamostres grans I

Per tant

$$P\left(-z_{1-\frac{\alpha}{2}} \leqslant \frac{\widehat{p}_X - p}{\sqrt{\frac{p(1-p)}{n}}} \leqslant z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

i aïllant la p Obtenimos:

p I.C. para μ de población normal con σ conocidamostres grans I

Mètode de Wilson

En estas condicions, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocida$p$ es (posant $\widehat{q}_X = 1-\widehat{p}_X$)

$$\left]\frac{\widehat{p}_X+\frac{z_{1-\alpha/2}^2}{2n}-z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X\widehat{q}_X}{n}+\frac{z_{1-\alpha/2}^2}{4n^2}}}{1+\frac{z_{1-\alpha/2}^2}{n}},\ \frac{\widehat{p}_X+\frac{z_{1-\alpha/2}^2}{2n}+z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X\widehat{q}_X}{n}+\frac{z_{1-\alpha/2}^2}{4n^2}}}{1+\frac{z_{1-\alpha/2}^2}{n}}\right[$$

`binom.wilson` del paquet `epitools`

p I.C. para μ de población normal con σ conocidamostres grans II

Consideremos ara la situación siguiente :

- X una v.a. Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con n mes grande y $\widehat{p}_X$ enfora de 0 y 1. Per Ejemplo, tal que:

$$n \geqslant 100, n\widehat{p}_X \geqslant 10, n(1-\widehat{p}_X) \geqslant 10$$

Fórmula de Laplace (1812)

En estas condicions, es pot prendre com a intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocida$p$

$$\left]\widehat{p}_X - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}, \widehat{p}_X + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}\right[$$

Ejemplo

En una mostra aleatòria de 500 famílies con nins en edat escolar es va trobar que 340 introduïen fruita de forma diària en la dieta dels seus fills

Cercau un intervalo de confianza del 95 % I.C. para μ de población normal con σ conocidala proporció real de famílies d'aquesta ciutat con nins en edat escolar que incorporen fruita fresca de forma diària en la dieta dels seus fills

Ejemplo

X = "Aportar diàriament fruita a la dieta dels fills" es$Be(p)$, y cercam intervalo de confianza del 95 % I.C. para μ de población normal con σ conocida$p$

Com que $n = 500 \geqslant 100$, $n\widehat{p}_X = 340 \geqslant 10$ y $n(1-\widehat{p}_X) = 160 \geqslant 10$, podem emprar

$$\left]\widehat{p}_X - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}, \widehat{p}_X + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}\right[$$

con

$$n = 500, \widehat{p}_X = \frac{340}{500} = 0{,}68$$

Dóna (recordau $\alpha = 0{,}05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$)

$$]0{,}639, 0{,}721[$$

Ejemplo

Con els altres mètodes:

```
> round(binom.exact(340,500,0.95),3)
    x   n proportion lower upper conf.level
1 340 500       0.68 0.637 0.721       0.95
> round(binom.wilson(340,500,0.95),3)
    x   n proportion lower upper conf.level
1 340 500       0.68 0.638 0.719       0.95
```

Donen:

- Clopper-Pearson: $]0{,}637, 0{,}721[$
- Wilson: $]0{,}638, 0{,}719[$
- Laplace: $]0{,}639, 0{,}721[$

Ejemplo

En un assaig d'un nou tractament de quimioteràpia, en una mostra de n (gran) malalts tractats, cap desenvolupà càncer testicular com a efecte secundari. Trobau un intervalo de confianza al 95 % I.C. para μ de población normal con σ conocidala proporció de malalts tractats con aquesta quimio que desenvolupen càncer testicular.

Ejemplo

En un assaig d'un nou tractament de quimioteràpia, en una mostra de n (gran) malalts tractats, cap desenvolupà càncer testicular com a efecte secundari. Trobau un intervalo de confianza al 95 % I.C. para μ de población normal con σ conocidala proporció de malalts tractats con aquesta quimio que desenvolupen càncer testicular.

No podem emprar la fórmula de Laplace, perquè $\widehat{p}_X = 0$. Cal emprar el mètode de Wilson:

Ejemplo

$$\left] \frac{\widehat{p}_X + \frac{z_{1-\alpha/2}^2}{2n} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X \widehat{q}_X}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{z_{1-\alpha/2}^2}{n}}, \frac{\widehat{p}_X + \frac{z_{1-\alpha/2}^2}{2n} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X \widehat{q}_X}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{z_{1-\alpha/2}^2}{n}} \right[$$

Ejemplo

$$\left]\frac{\widehat{p}_X+\frac{z_{1-\alpha/2}^2}{2n}-z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X\widehat{q}_X}{n}+\frac{z_{1-\alpha/2}^2}{4n^2}}}{1+\frac{z_{1-\alpha/2}^2}{n}},\right.$$

$$\left.\frac{\widehat{p}_X+\frac{z_{1-\alpha/2}^2}{2n}+z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X\widehat{q}_X}{n}+\frac{z_{1-\alpha/2}^2}{4n^2}}}{1+\frac{z_{1-\alpha/2}^2}{n}}\right[$$

$$\left]\frac{\frac{1{,}96^2}{2n}-1{,}96\sqrt{\frac{1{,}96^2}{4n^2}}}{1+\frac{1{,}96^2}{n}},\frac{\frac{1{,}96^2}{2n}+1{,}96\sqrt{\frac{1{,}96^2}{4n^2}}}{1+\frac{1{,}96^2}{n}}\right[=\left]0,\frac{1{,}96^2}{n+1{,}96^2}\right[$$

Els metges empren $\left]0,\frac{3}{n}\right[$ (la regla del 3)

Observaciones

- El mètode de Wilson dóna un I.C. centrado en

$$\frac{\widehat{p}_X + \frac{z_{1-\alpha/2}^2}{2n}}{1 + \frac{z_{1-\alpha/2}^2}{n}} = \frac{2n\widehat{p}_X + z_{1-\frac{\alpha}{2}}^2}{2n + 2z_{1-\frac{\alpha}{2}}^2}$$

- No es coneix una fórmula per al centre de l'I.C. de Clopper-Pearson.
- La fórmula de Laplace dóna un I.C. centrado en $\widehat{p}_X$
- Quan n creix es redueix l'Amplitud de el intervalo de confianza

Amplitud

L'Amplitud de el intervalo de confianza de Laplace es

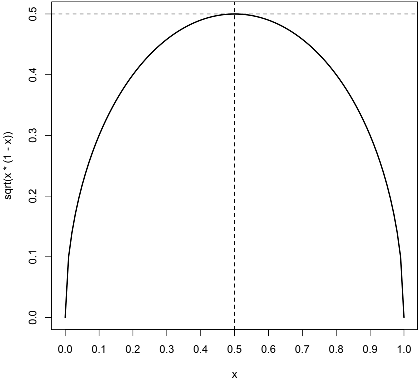
$$A = 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}$$

No podem determinar la tamaño de la mostra a fi que el intervalo de confianza tingui una certa Amplitud màxima sense conèixer $\widehat{p}_X$, que no coneixem sense una mostra

Amplitud

$$A = 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}$$

El màxim de $\sqrt{\widehat{p}_X(1-\widehat{p}_X)}$ s'assoleix a $\widehat{p}_X = 0{,}5$



Amplitud

$$A = 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}$$

El màxim de $\sqrt{\widehat{p}_X(1-\widehat{p}_X)}$ s'assoleix a $\widehat{p}_X = 0{,}5$

Per tant, calcularem n per obtenir una Amplitud com a màxim A_0 suposant el pitjor dels casos ($\widehat{p}_X = 0{,}5$):

$$A_0 \geqslant 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{0{,}5^2}{n}} = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \Rightarrow n = \left\lceil \frac{z_{1-\frac{\alpha}{2}}^2}{A_0^2} \right\rceil$$

Ejemplos

Set de cada deu estudiants de la UIB practica el ciberplagi a l'hora de confeccionar els treballs acadèmics

Fixa tècnica de la mostra de la UIB

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

$$\text{Error} =$$

Ejemplos

Set de cada deu estudiants de la UIB practica el ciberplagi a l'hora de confeccionar els treballs acadèmics

Fixa tècnica de la mostra de la UIB

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

$$\text{Error} = \frac{1{,}96 \cdot 0{,}5}{\sqrt{727}} \approx 0{,}0363$$

Ejemplo

Volem estudiar quina fracció de las morts per càncer corresponen a morts per càncer d'estómac. Per determinar aquesta fracció a un nivell de confianza del 95 % y garantir un error màxim de 0.05, de quina tamaño ha de ser la mostra en el pitjor dels casos?

$$n = \left\lceil \frac{z_{1-\frac{\alpha}{2}}^2}{A^2} \right\rceil$$

on

$$\frac{A}{2} = 0{,}05, \quad z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$$

Dóna $n = \lceil 384{,}16 \rceil = 385$.

varianza de una població normal

Consideremos ara la situación siguiente :

- X una v.a. normal con μ y σ desconocidas
- $X_1, \ldots, X_n$ una m.a.s. de X y varianza muestral $\widetilde{S}_X^2$

Teorema

En estas condicions

$$\frac{(n-1)\tilde{S}_X^2}{\sigma^2}$$

té distribución χ_{n-1}^2

varianza de una població normal

Consideremos ara la situación siguiente :

- X una v.a. normal con μ y σ desconocidas
- $X_1, \ldots, X_n$ una m.a.s. de X y varianza muestral $\widetilde{S}_X^2$

Teorema

En estas condicions, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ *I.C. para* μ *de población normal con* σ *conocida*σ^2 *es*

$$\left]\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right[,$$

on $\chi^2_{\nu,q}$ *esel q-quantil de la distribución* χ^2_ν

varianza de una població normal

En efecte

$$
\begin{aligned}
1-\alpha &= P\left(\chi^2_{n-1,\frac{\alpha}{2}} \leqslant \chi^2_{n-1} \leqslant \chi^2_{n-1,1-\frac{\alpha}{2}}\right) \\
&= P\left(\chi^2_{n-1,\frac{\alpha}{2}} \leqslant \frac{(n-1)\widetilde{S}_X^2}{\sigma^2} \leqslant \chi^2_{n-1,1-\frac{\alpha}{2}}\right) \\
&= P\left(\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \leqslant \sigma^2 \leqslant \frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)
\end{aligned}
$$

I ara χ^2_{n-1} no essimètrica, així que s'han de calcular $\chi^2_{n-1,\frac{\alpha}{2}}$ y $\chi^2_{n-1,1-\frac{\alpha}{2}}$

Observació: el intervalo de confianza per σ^2 no està centrado en $\widetilde{S}_X^2$

Ejemplo

Un índex de qualitat d'un reactiu químic esel temps que triga a actuar. L'estàndard esque aquest ha de ser $\leqslant 30$ segons. Se suposa que la distribución del temps d'actuació del reactiu es aproximadament normal.

Es realitzen 30 proves en las quals es mesura el temps d'actuació del reactiu:

12, 13, 13, 14, 14, 14, 15, 15, 16, 17, 17, 18, 18, 19, 19, 25, 25, 26, 27, 30, 33, 34, 35, 40, 40, 51, 51, 58, 59, 83

Es demana calcular un intervalo de confianza I.C. para μ de población normal con σ conocidala desviació típica al nivell 95 %

Ejemplo

$$\left]\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right[$$

```
> Temps=c(12,13,13,14,14,14,15,15,16,17,17,18,
 18,19,19,25,25,26,27,30,33,34,35,40,40,51,51,
 58,59,83)
> length(Temps) #n
[1]  30.0000
> var(Temps) # varianza mostral
[1]  301.5506
```

i $\alpha = 0{,}05$:

$$\chi^2_{29,0{,}975} = 45{,}72, \ \ \chi^2_{29,0{,}025} = 16{,}05$$

Ejemplo

el intervalo serà

$$\left]\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right[$$

Obtenimos

$$\left]\frac{29\cdot 301{,}5506}{45{,}72}, \frac{29\cdot 301{,}5506}{16{,}05}\right[=]191{,}27, 544{,}86[$$

Ejemplo

el intervalo serà

$$\left]\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right[$$

Obtenimos

$$\left]\frac{29\cdot 301{,}5506}{45{,}72}, \frac{29\cdot 301{,}5506}{16{,}05}\right[=]191{,}27, 544{,}86[$$

Aquest era I.C. para μ de población normal con σ conocidala variància! I.C. para μ de población normal con σ conocidala desviació típica

$$]\sqrt{191{,}27}, \sqrt{544{,}86}[=]13{,}83, 23{,}34[$$

"Poblacions finites"

Fins ara hem emprat mostres aleatòries simples

A la pràctica, es prenen mostres aleatòries sense reposició

Si la tamaño N de la població esmolt mes grande que la tamaño n de la mostra (posem $N \geqslant 40n$), las fórmules donades fins ara funcionen (aproximadament) bé

Però...

Fixa tècnica de la mostra de la UIB

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

"Poblacions finites"

Es dóna l'efecte de població finita quan N esrelativament petit

En aquest cas, a las fórmules que hem donat per als intervals de confianza I.C. para μ de población normal con σ conocidaμ o p cal multiplicar l'error estàndard o l'error muestral pel factor corrector

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

"Poblacions finites"

Consideremos la situación siguiente :

- X una població de tamaño N que sigue una distribución con media poblacional μ desconocida
- $X_1, \ldots, X_n$ una m.a. sense reposició de X, con media $\overline{X}$
- n es grande

"Teorema"

En estas condicions, es recomana prendre com a intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocidaμ

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}},\ \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right[$$

"Poblacions finites"

Consideremos la situación siguiente :

- X una població de tamaño N que sigue una distribución Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a. sense reposició de X, con n molt gran y con freqüència relativa d'èxits $\widehat{p}_X$ no extrema

"Teorema"

En estas condicions, es recomana prendre com a intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocida$p$

$$\left]\widehat{p}_X - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}\sqrt{\frac{N-n}{N-1}},\right.$$
$$\left.\widehat{p}_X + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}\sqrt{\frac{N-n}{N-1}}\right[$$

"Poblacions finites"

"Teorema"

En las condicions anteriors, per obtenir un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para μ de población normal con σ conocida$p$ en el pitjor dels casos caldrà prendre una mostra de tamaño

$$n = \left\lceil \frac{N z_{1-\frac{\alpha}{2}}^2}{A^2(N-1) + z_{1-\frac{\alpha}{2}}^2} \right\rceil$$

Ejemplo

Fixa tècnica de la mostra de la UIB

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

De la població total d'estudiants de grau de la UIB quants n'hem d'escollir de manera aleatòria sense reposició per estimar la proporció dels que han comès plagi, con un error del 3.52 % y un nivell de confianza del 95 %?

Ejemplo

Fixa tècnica de la mostra de la UIB

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

De la població total d'estudiants de grau de la UIB quants n'hem d'escollir de manera aleatòria sense reposició per estimar la proporció dels que han comès plagi, con un error del 3.52 % y un nivell de confianza del 95 %?

$$n = \left\lceil \frac{11797 \cdot 1{,}96^2}{0{,}0704^2 \cdot 11796 + 1{,}96^2} \right\rceil = \lceil 727{,}3854 \rceil = 728$$