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

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.

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 poblacio[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 ± 1 vez el error de muestreo = intervalo de confianza del 67%.

estimación ± 2 veces el error de muestreo = intervalo de confianza del 95%.

estimación ± 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 en el que 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 desviación típica de esta tasa han sido

$$\overline{x} = 0.7, \quad \widetilde{s} = 0.1$$

Determinar un intervalo que podamos afirmar con probabilidad 95 % que contiene la tasa de alcoholemia media en sangre de una varón después de beber 3 cañas de cerveza de 330ml.

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

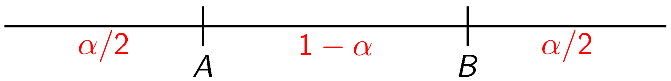
Ejemplo: $]A, B[$ es un intervalo de confianza del 95 % (o de nivel de significación de 0.05) si

$$P(A < \theta < B) = 0.95$$

Definiciones básicas

Por defecto, buscaremos intervalos bilaterales 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 distribución 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 $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$

$$P\left(-2.24 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 2.24\right) = 0.975$$

$$P\left(\overline{X} - 2.24\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.24\frac{\sigma}{\sqrt{n}}\right) = 0.975$$

Ejemplo: μ de población normal con σ conocida

Por lo tanto, la probabilidad que la media μ de X se encuentre dentro del intervalo

$$\left]\overline{X} - 2.24\frac{\sigma}{\sqrt{n}}, \overline{X} + 2.24\frac{\sigma}{\sqrt{n}}\right[$$

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 población 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 muestra concreta: La probabilidad de que la media μ de la población que ha producido esta muestra, 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 la frase anterior es mentira (es un abuso de lenguaje): La μ concreta es un valor fijo, por lo tanto que pertenezca o no a este intervalo concreto tiene probabilidad 1 (si pertenece) y 0 (si no

I.C. para 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 $(1-\alpha)\cdot 100\,\%$ para 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 estándar 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

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

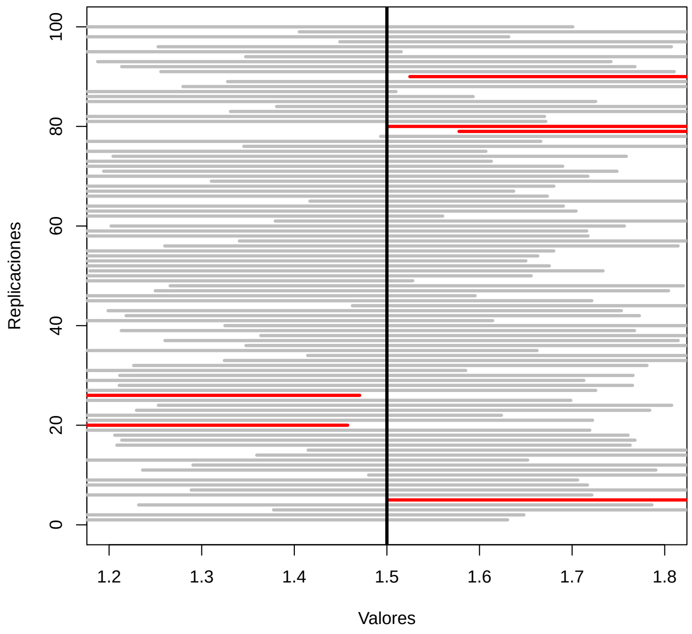
Observad que está centrado en $\overline{X}$

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
Población=rnorm(10^6,mu,sigma)
M=replicate(100,ICZ(sample(Población,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 $\alpha \cdot 100\,\%$ de las veces, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ no contendrá el valor real del parámetro

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

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 siguientes:

```
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

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

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 conocida

$$\left]20-2.24\cdot\frac{4}{\sqrt{16}},20+2.24\cdot\frac{4}{\sqrt{16}}\right[=]17.76,22.24[$$

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

Supongamos que las medidas de nuestro sensor siguen una distribución 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 siguientes condiciones:

- Población normal con $\sigma=\sqrt{1.44}=1.2$ conocida
- Muestra aleatoria simplede 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 % 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 intervalo de confianza para 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:

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

Dada A_0, tomaremos

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

Ejemplo

Recordemos que las medidas de nuestro sensor de temperatura seguían una distribución 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 máximo de 0.05° al nivel de confianza del 90 %?

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

donde

$$\frac{A_0}{2} = 0.05, \quad z_{1-\frac{\alpha}{2}} = 1.64, \quad \sigma = 0.1$$

Obtenemos $n = \lceil 10.76 \rceil = 11$

distribución t de Student

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

- Tiene densidad

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

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

- La distribución está tabulada (las tablas en el moodle de la asignatura), y con R es `t`.

Distribución 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 condiciones, la v.a.

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

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

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

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 $N(0,1)$):

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

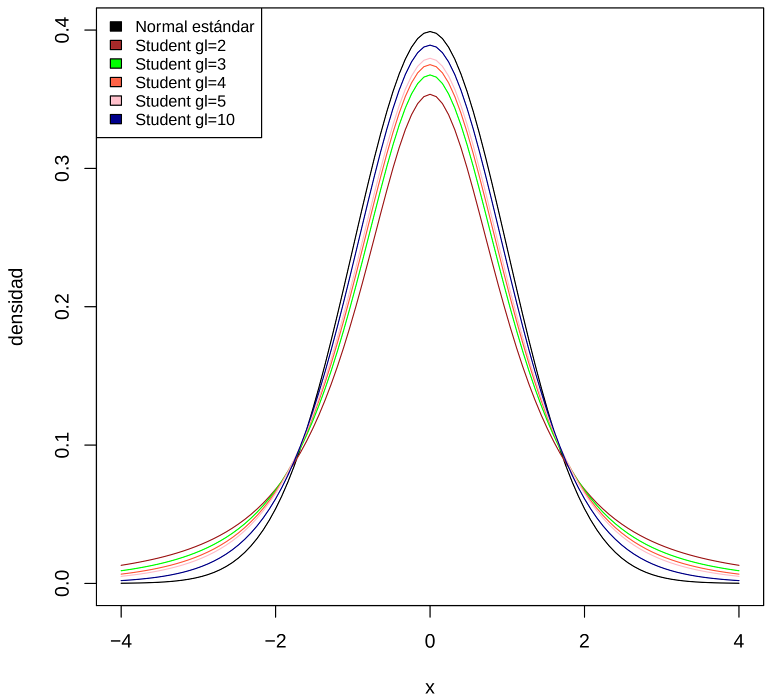
- Si ν es grande, su distribución es aproximadamente la de $N(0,1)$ (pero con más varianza: un poco más aplastada)

Distribución t de Student

EL siguiente código dibuja distintas densidades de la distribución t de Student

```
curve(dnorm(x, 0, 1), xlim = c(-4, 4),
      ylim = c(0, 0.4), col = "black",
      ylab = "densidad", xlab = "x")
legend("topleft",
        legend = c("Normal estándar", "Student gl=2",
                   "Student gl=3",  "Student gl=4",
                   "Student gl=5",  "Student gl=10"),
        fill = c("black", "brown", "green", "tomato",
                "pink", "darkblue"), cex = 0.8)
curve(dt(x, df = 2), col = "brown", add = TRUE)
curve(dt(x, df = 3), col = "green", add = TRUE)
curve(dt(x, df = 4), col = "tomato", add = TRUE)
curve(dt(x, df = 5), col = "pink", add = TRUE)
curve(dt(x, df = 10), col = "darkblue", add = TRUE)
```

Distribución t de Student

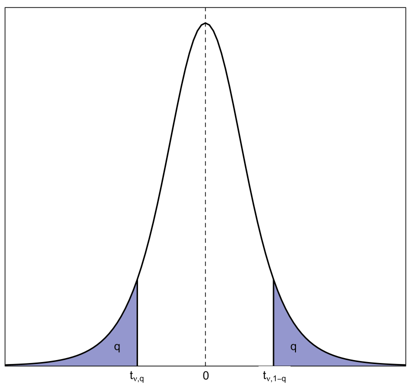


Distribución t de Student

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



I.C. para 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 condiciones, 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:

`Ficha técnica: 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 por fabricante cae en el intervalo de confianza del 90 %?

Ejemplo

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

```
qt(0.95,24)

## [1] 1.710882

round(qt(0.95,24),2)

## [1] 1.71
```

Operando: $]531.962, 504.038[$, y no contiene a 500 (¡pero se equivoca a favor del consumidor!)

Ejemplo

Con esta muestra ¿la media poblacional anunciada por fabricante cae en el intervalo de confianza del 90 %?
Hay que calcular el intervalo de confianza para la μ de una población normal con σ conocida μ 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 $t_{24,0.95} = 1.71$

Observaciones

- El intervalo de confianza obtenido está centrado en $\overline{X}$
- La fórmula
$$\left]\overline{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right[$$
nos da el I.C. para μ en una 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
$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right[$$

I.C. para μ una población normal con σ conocida y muestra 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 condiciones (T.C.L.)

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

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

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 (pongamos 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

Consultad el enlace Guardia civil informa...
Simulemos un experimento en el que se ha medido la tasa oficial de alcoholemia en sangre a 40 varones (sobrios) después de tomar 3 cañas de cerveza de 330 ml. El siguiente código simula este experimento supuesto que la tasa de alcoholemia siga una distribución normal de media 0.7 (supuestamente desconocida) y desviación típica 0.1 (supuestamente desconocida)

```
set.seed(1234)# por reproducibilidad
tasa_alcoholemia=round(rnorm(40,mean=0.7,sd=0.1),2)
head(tasa_alcoholemia,10)

##  [1] 0.58 0.73 0.81 0.47 0.74 0.75 0.64 0.65 0.64 0.61
```

Ejemplo

```
media=round(mean(tasa_alcoholemia),3)
media

## [1] 0.659

des.tip=round(sd(tasa_alcoholemia),3)
des.tip

## [1] 0.091
```

Así que tenemos los siguientes estadísticos

$$\overline{x} = 0.659, \quad \widetilde{s} = 0.091$$

Ejemplo

Calculad un intervalo del que podamos afirmar que con una probabilidad del 95 % contiene la media poblacional de la tasa de alcoholemia para hombres después de beber 3 cerveza de 330ml.

Nos piden un intervalo de confianza del 95 % para μ de una población normal con σ conocida de la v.a. X "tasa de alcoholemia para hombres después de beber 3 cervezas de 330ml"
No conocemos la distribución de X, pero $n = 40$ es "grande"

Ejemplo

Se ha tomado una muestra del tiempo de visualización de vídeo semanal en horas de 1000 usuarios de un canal de videos por internet. Se ha obtenido una media muestral de 9.5 horas/semana con una desviació típica muestral de 0.5 horas/semana.

Calculad un intervalo de confianza del 95 % para la media poblacional del número de horas visualizadas por semana supuesto que sigue aproximadamente una población normal con σ desconocida

Ejemplo

Podemos emplear

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

donde

$$n = 40, \overline{x} = 0.659, \widetilde{s} = 0.091,$$
$$\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

$$]0.631, 0.687[$$

Podemos afirmar con un 95 % de confianza que la tasa de alcoholemia media en sangre de la población de hombres después de beber 3 cervezas de 330ml está entre 0.631 y 0.687

Ejemplo

Como $n = 1000$ es grande, podemos utilizar

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

donde

$$\overline{x} = 9.5,\ \widetilde{s} = 0.5,\ \alpha = 0.05,\ z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

Con estos valores obtenemos que

```
media=9.5
sd=0.5
n=1000
alpha=0.5
```

Ejemplo

```
cuantil=qnorm(1-0.5/2)
cuantil

## [1] 0.6744898

a=round(media-cuantil*sd/sqrt(n),3)
a

## [1] 9.489

b=round(media+cuantil*sd/sqrt(n),3)
b

## [1] 9.511
```

Por lo tanto

Amplitud

De una población X hemos tomado una m.a.s. piloto que ha tenido una desviación típica muestral $\widetilde{s}_{pilot}$.

Estimaremos que el tamaño mínima n de una m.a.s. de X que dé un intervalo de confianza I.C. para μ de una población normal con σ desconocida de nivell de confianza $1-\alpha$ y amplitud máxima A_0 es

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

Amplitud

La 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

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

Para determinar n (grande) que dé cómo máximo una amplitud A prefijada, necesitamos $\widetilde{S}_X$, que depende de la muestra.

Soluciones:

- Si sabemos la desv. típ. poblacional σ, la utilizaremos en lugar de $\widetilde{S}_X$
- Si hemos tomado una muestra previa (piloto), emplearemos la desviación típica muestral para estimar σ

Ejemplo

Queremos estimar la estatura media de los estudiantes de la UIB. Queremos obtener un intervalo de confianza del 99 % con una precisión máxima de 1 cm. En una muestra piloto de 25 estudiantes, obtuvimos que

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

¿Basándonos en estos datos, cuál es el tamaño necesario de la muestra para poder alcanzar nuestro objetivo?

Ejemplo

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

- Precisión $=$ error máximo $= A/2 = 1$
- $\widetilde{s}_{pilot} = 10$
- $\alpha = 0.01 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$

Dóna

$$n = \left\lceil \left(2.58 \cdot \frac{10}{1}\right)^2 \right\rceil = \lceil 665.64 \rceil = 666$$

I.C para una proporción

Consideremos la situación siguiente :

- X una v.a. Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a.s. de X, con número de éxitos x y por lo tanto la frecuencia relativa de éxitos es $\widehat{p}_X = x/n$

Recordad que X es $B(n, p)$

Método "exacto" o de Clopper-Pearson

Un intervalo de confianza $]p_0, p_1[$ del $(1-\alpha)100\,\%$nivel de confianza para p de una población se obtiene encontrando el p_0 más grande y el p_1 más pequeño tales que

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

Manualmente (consultando las tablas) sería un trabajo muy pesado.

I.C. para una proporción

El paquete `epitools` incorpora la función

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

para calcularlo

De 10 pacientes tratados con un medicamento, 2 se han curado. Dar un intervalo de confianza del 95 % I.C. para la proporción poblacional p de pacientes que este medicamento sana

```
#descomentar para instalar
#install.packages("epitools",dep=TRUE)
library(epitools)
solución=round(binom.exact(2,10,0.95),3)
solución

##   x  n proportion lower upper conf.level
## 1 2 10        0.2 0.025 0.556       0.95
```

Da $]0.025, 0.556[$

I.C. para la proporción poblacional muestras grandes I

Consideremos 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 (por Ejemplo, $n \geqslant 40$) y frecuencia relativa de éxitos $\widehat{p}_X$

En estas condiciones (por el T.C.L.),

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

I.C. para la proporción p muestral muestras grandes I

Por lo tanto

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

El problema es que no conocemos p, la literatura plantea entre otras soluciones :

I) El método de Wilson
II) La solución de Laplace (1812)

Exponemos estas soluciones a continuación

I.C. para la proporción poblacional muestras grandes I

Método de Wilson

En estas condiciones, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ I.C. para p es (donde $\widehat{q}_X = 1-\widehat{p}_X$)

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

`binom.wilson` del paquete `epitools`

I.C. para la proporción poblacional muestras grandes II

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

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

Fórmula de Laplace (1812)

Bajo estas condiciones, se puede tomar como intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ para 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 muestra aleatoria de 500 familias con niños en edad escolar se encontró que 340 introducían fruta de forma diaria en la dieta de sus hijos

Calculad un intervalo de confianza del 95 % para conocida la proporción real de familias de esta ciudad con niños en edad escolar que incorporen fruta fresca de forma diaria en la dieta de sus hijos.

Ejemplo

$X =$ "Aportan diariamente fruta a la dieta de sus hijos" es $Be(p)$, y buscamos el intervalo de confianza del 95 % para p

Como que $n = 500 \geqslant 100$, $n\widehat{p}_X = 340 \geqslant 10$ y $n \cdot (1 - \widehat{p}_X) = 160 \geqslant 10$, podemos utilizar este intervalo

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

con

$$n = 500, \widehat{p}_X = \frac{340}{500} = 0.68$$

Que vale (recordad que $\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$)

$$]0.639, 0.721[$$

Ejemplo

Con los otros métodos:

```
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
```

Obtenemos:

- Clopper-Pearson: $]0.637, 0.721[$
- Wilson: $]0.638, 0.719[$
- Laplace: $]0.639, 0.721[$

Ejemplo

En un ensayo de un nuevo tractamiento de quimioterapia, en una muestra de n (gran) enfermos tratados, ninguno desarrolla cáncer testicular como efecto secundario. Calculad un intervalo de confianza al 95 % para la proporción poblacional de enfermos afectados con aquesta quimio que tienen cáncer testicular.

No podemos emplear la fórmula de Laplace, porqué $\widehat{p}_X = 0$. Tenemos que emplear el método de Wilson:

Ejemplo

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

$$\left]\frac{\frac{1.96^2}{2n} - 1.96\sqrt{\frac{1.96^2}{4n^2}}}{1 + \frac{1.96^2}{n}}, \frac{\frac{1.96^2}{2n} + 1.96\sqrt{\frac{1.96^2}{4n^2}}}{1 + \frac{1.96^2}{n}}\right[= \left]0, \frac{1.96^2}{n + 1.96^2}\right[$$

En el área de sanidad se emplea $\left]0, \frac{3}{n}\right[$ (la regla del 3)

Observaciones

- El método de Wilson da un I.C. centrado en

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

- No se conoce una fórmula para el centro del I.C. de Clopper-Pearson.
- La fórmula de Laplace dóna un I.C. centrado en $\widehat{p}_X$
- Cuando n crece se reduce la amplitud del intervalo de confianza

Amplitud

La amplitud del intervalo de confianza de Laplace es

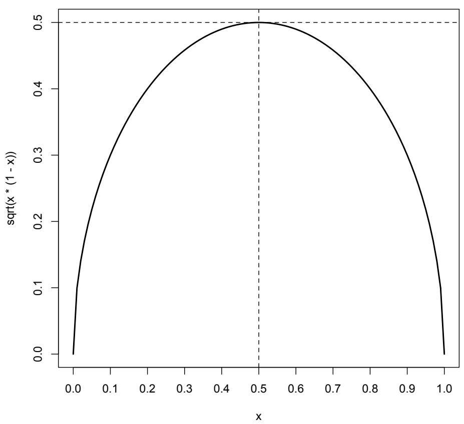
$$A = 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n}}$$

No podem determinar el tamaño de la mostra para que el intervalo de confianza tenga una cierta amplitud máxima sin conocer $\widehat{p}_X$, que obviamente no conocemos sin hacer una muestra

Amplitud

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

El máximo de $\sqrt{\widehat{p}_X(1-\widehat{p}_X)}$ se alcanza en $\widehat{p}_X = 0.5$



Amplitud

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

El máximo de $\sqrt{\widehat{p}_X(1-\widehat{p}_X)}$ se alcanza en $\widehat{p}_X = 0.5$

Por lo tanto, calcularemos n para obtener una amplitud máxima A_0 suponiendo el peor de los casos ($\widehat{p}_X = 0.5$):

$$A_0 \geqslant 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{0.5^2}{n}} = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \Rightarrow n = \left\lceil \frac{z^2_{1-\frac{\alpha}{2}}}{A_0^2} \right\rceil$$

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.

$$\text{Error} = \frac{1.96 \cdot 0.5}{\sqrt{727}} \approx 0.0363$$

Ejemplo

Quemos estudiar qué fración teléfonos móviles utilizan android. Para determinar esta proporción con un nivel de confianza del 95 % y garantizar un error máximo de 0.05, ¿de qué tamaño ha de ser la muestra en el peor de los casos?

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

on

$$\frac{A}{2} = 0.05, \quad z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

Obtenemos que $n = \lceil 384.16 \rceil = 385$.

I.C. de la varianza σ^2 de una población normal

Consideremos la siguiente situación:

- 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^2_{n-1}

I.C. de la varianza σ^2 de una població normal

Tenemos 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 condiciones, un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ *para varianza* σ^2 *de una población normal es*

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

donde $\chi^2_{\nu,q}$ *es el q-cuantil de la distribución* χ^2_ν

I.C. de la varianza σ^2 de una població normal

En efecto

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

Como χ^2_{n-1} no es simétrica, hemos de calcular $\chi^2_{n-1,\frac{\alpha}{2}}$ y $\chi^2_{n-1,1-\frac{\alpha}{2}}$

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

Ejemplo

Un algoritmo probabilístico depende de la semilla de aleatorización que se genera en cada paso. Para saber si la semilla influye mucho en el resultado se ejecuta el algoritmo varias veces hasta obtener un resultado similar y se estudia la varianza de su tiempo de ejecución.
Queremos que se esta varianza sea $\leqslant 30$
Se supone que la distribución del tiempo de ejecución del algoritmo es aproximadamente normal.

Se realizan 30 ejecuciones del algoritmo de las que se mide el tiempo de ejecución, los resultados son :

Nos piden calcular un intervalo de confianza para σ^2 de una población normal al nivel 95 %

Ejemplo

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

```
tiempo=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)
n=length(tiempo)
n

## [1] 30

var(tiempo)

## [1] 301.5506

sd(tiempo)

## [1] 17.36521
```

Ejemplo

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

```
#cuantiles
alpha=0.05
qchisq(alpha/2,n-1)

## [1] 16.04707

qchisq(1-alpha/2,n-1)

## [1] 45.72229
```

$\alpha = 0.05$:

$$\chi^2_{29,0.975} = 45.72,\ \chi^2_{29,0.025} = 16.05$$

Ejemplo

el intervalo será

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

Obtenemos

$$\left]\frac{29\cdot 301.5506}{45.72}, \frac{29\cdot 301.5506}{16.05}\right[=]191.27, 544.86[$$

Este es el I.C. para σ de una población normal con σ, para la desviación típica podemos hacer (hay otras fórmulas)

$$]\sqrt{191.27}, \sqrt{544.86}[=]13.83, 23.34[$$

"Poblaciones finitas"

Por el momento hemos utilizado muestras aleatorias simples

En la práctica, se toman muestras aleatorias sin sin reposición

Si la tamaño N de la población es mucho más grande que la tamaño n de la muestra (digamos que $N \geqslant 40 \geqslant n$), las fórmulas dadas hasta ahora funcionan (aproximadamente) bien

Pero. . .

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.

"Poblaciones finitas"

En este caso sí se da el efecto de población finita cuando N es relativamente pequeño

Así que en esta situación, en las fórmulas que hemos propuesto para los intervalos de confianza I.C. para μ o p hay que multiplicar el error estándar o el error muestral por el factor corrector de población finita

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

"Poblaciones finitas"

Consideremos la situación siguiente :

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

"Teorema"

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

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

"Poblaciones finitas"

Consideremos la situación siguiente :

- X una población de tamaño N que sigue una distribución Bernoulli con p desconocida
- $X_1, \ldots, X_n$ una m.a. sin reposición de X, con n muy grande y con frecuencia relativa de éxitos $\widehat{p}_X$ lejos de los extremos

"Teorema"

En estas condiciones, es recomienda tomar como intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ el parámetro p

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

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 población total de estudiantes de grado de la UIB ¿cuántos hemos de escoger de manera aleatoria y sin reposición para estimar la proporción de los que han cometido plagio, con un error del 3.52 % y un nivel de confianza del 95 %?

$$n = \left\lceil \frac{11797\cdot 1.96^2}{0.0704^2\cdot 11796 + 1.96^2}\right\rceil = \lceil 727.3854\rceil = 728$$

"Poblaciones finitas"

"Teorema"

En las condiciones anteriores, para obtener un intervalo de confianza del $(1-\alpha)\cdot 100\,\%$ para p en el peor de los casos necesitaremos tomar una muestra de tamaño

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