BLOC 2= Tema 2 + Tema 3. Estimació Puntual + Estimació per intervals.

TEMA 3. Intervals de confiança

LLIÇONS

3.1 Concepte d'interval de confiança.

3.2 Interval de confiança per a la mitjana: cas normal i cas general.

3.3 Interval de confiança per a una proporció.

3.4 Intervals de confiança per a la variància. Cas normal

3.5 Intervals calculats amb R.

3.1 Concepte d'interval de confiança

Un interval de confiança (IC) per a un paràmetre θ (no observable) és un tipus d'estimació de θ, complementari a l'estimació puntual donada per $\hat{\theta}$.

Un interval de confiança es calcula a partir de les observacions, i per tant difereix de mostra a mostra.

Un interval de confiança freqüentment inclou el valor del paràmetre d'interès θ.

La freqüència en que l'interval observat conté el paràmetre es determina pel nivell de confiança.

Si es construissin molts intervals de confiança a partir de molts anàlisis de les dades per separat, la proporció de tals intervals que contenen el veritable valor del paràmetre θ coincidiria amb el nivell de confiança donat.

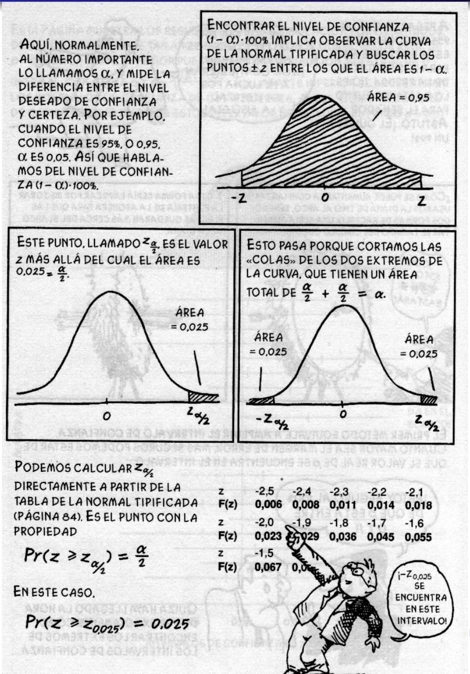
Interval de confiança per μ a partir d'una m.a.s de $X \sim N(\mu, 1)$, $\sigma^2 = 1$ amb nivell de confiança del 99%

Ja sabem que $\overline{X}$ és un bon estimador de μ. Ara ens plantegem trobar un interval $[\overline{X} - a, \overline{X} + a]$ amb nivell de confiança $1 - \alpha = 0.99$, és a dir, $\mathrm{Prob}\{\overline{X} - a \leq \mu \leq \overline{X} + a\} = 0.99$ per un cert valor a.

Com trobem a? En aquest cas ho resolem a partir d'una mostra X_1, X_2, X_3, X_4.

1. Calculem $\overline{X} = \frac{1}{n}\sum_{i=1}^{4} X_i$. En realitat hauriem d'escriure $\overline{X}_4$, pero s'omet ara per fer-ho mes senzill de llegir.
2. Com que $\overline{X} \sim N(\mu, 1/4)$, estandaritzant $\frac{\overline{X}-\mu}{1/2} = Z \sim N(0,1)$
3. $0.99 = \mathrm{Prob}\{\overline{X} - a \leq \mu \leq \overline{X} + a\} = \mathrm{Prob}\{-2a \leq \frac{\overline{X}-\mu}{1/2} \leq 2a\}$ $0.99 = \mathrm{Prob}\{-2a \leq Z \leq 2a\} = 1 - 2\mathrm{Prob}\{Z \geq 2a\} \implies$ $\mathrm{Prob}\{Z \geq 2a\} = 0.005 = \alpha/2$
4. $z_{\alpha/2} = z_{0.005} = 2.575 \implies 2a = 2.575 \Rightarrow a = 1.2875$
5. $\mathrm{Prob}\{-2.575 \leq \frac{\overline{X}-\mu}{1/2} \leq 2.575\} = 0.99$
6. $[\overline{X} - 1.2875, \overline{X} + 1.2875]$ interval de confiança per μ amb nivell de confiança del 99%

Trobant els valors de les cues a les taules



Interval de confiança amb nivell de confiança del 99%

Suposem que hem observat:
$x_1 = 3.4, x_2 = 1.9, x_3 = 2.8, x_4 = 4.1 \Rightarrow \overline{x} = 3.05$. Estimem μ per $\overline{X}$ i l'estimació serà $\overline{x} = 3.05$ amb $S^2 = \sum_{i=1}^{4}(x_i - \overline{x})^2/3 = 0.87$, $S = 0.93$.

Amb les dades de l'exemple l'interval
$[3.05 - 1.2875, 3.05 + 1.2875] = [1.7625, 4.3375]$ conté μ amb una confiança del 99%.

Fixem-nos que els extrems de l'interval de confiança depenen de:

1. Les dades $x_1 = 3.4, x_2 = 1.9, x_3 = 2.8, x_4 = 4.1 \Rightarrow \overline{x} = 3.05$
2. La desviació estàndard. En aquest cas $\sigma = 1$ coneguda
3. El nivell de confiança $1 - \alpha$. En aquest cas $\alpha = 0.01$
4. La llei de $\overline{X}$. En aquest cas $\overline{X} \sim N(\mu, 1/4)$
5. $\alpha = 0.01$ i $\overline{X} \sim N(\mu, 1/4) \Rightarrow z_{\alpha/2} = 2.575$

El preu que hem pagat per aconseguir una CONFIANÇA del 99%, en comptes d'un 95%, ha estat un interval MÉS AMPLE i per tant menys precisió per estimar μ:
passem de $[2.05, 4.05]$ (confiança 95%) a

Definició interval de confiança: FORMALMENT

Definició

Sigui $X_1, X_2, \ldots, X_n \sim X \sim$ model probabilístic que depèn de θ.
Un **interval de confiança** $[\hat{\theta}_1, \hat{\theta}_2]$ **amb nivell de confiança** $100(1-\alpha)\%$ per a θ és un interval aleatori tal que $\mathrm{Prob}\{[\hat{\theta}_1, \hat{\theta}_2] \quad \mathrm{conte} \quad \theta\}$ amb probabilitat $1-\alpha$.

Els extrems de l'interval $\hat{\theta}_1$ i $\hat{\theta}_2$ es calculen a partir de

- les dades mostrals $(x_1, x_2, \ldots, x_n)$
- el nivell de confiança $100(1-\alpha)\%$

En l'exemple anterior, el paràmetre $\theta = \mu$

- per un nivell de confiança del $1-\alpha = 99\%$
  $\hat{\mu}_1 = \hat{\mu}_1(X_1, X_2, \ldots, X_n, 0.99) = \overline{X} - 1.2875$ i
  $\hat{\mu}_2 = \hat{\mu}_2(X_1, X_2, \ldots, X_n, 0.99) = \overline{X} + 1.2875 \Rightarrow$
  $[\overline{X} - 1.2875, \overline{X} + 1.2875]$
- per un nivell de confiança del $1-\alpha = 95\%$
  $\hat{\mu}_1 = \hat{\mu}_1(X_1, X_2, \ldots, X_n, 0.95) = \overline{X} - 1$ i
  $\hat{\mu}_2 = \hat{\mu}_2(X_1, X_2, \ldots, X_n, 0.95) = \overline{X} + 1 \Rightarrow [\overline{X} - 1, \overline{X} + 1]$

LLEGIR PEÑA 8.1 PÀGINES 319-321

Per que funcionen els intervals de confiança?

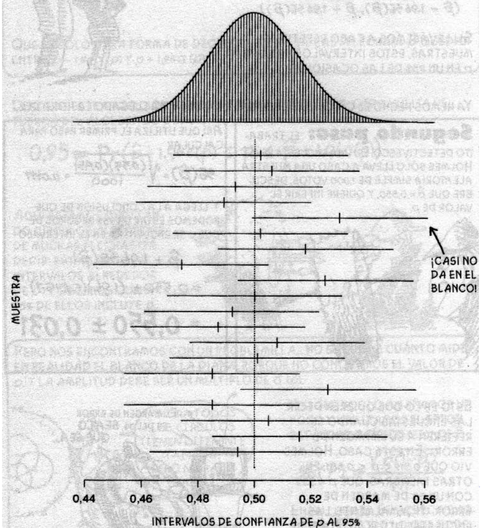
$X_1, X_2, \ldots, X_n \sim X \sim$ model probabilístic que depèn de θ
Busquem un interval $[\hat{\theta}_1, \hat{\theta}_2]$ tal que $\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$
$\hat{\theta}_1$ i $\hat{\theta}_2$ es calculen a partir de la mostra $X_1, X_2, \ldots, X_n$ de forma tal que si:

- agafessim MOLTES mostres,
- totes de la mateixa grandària n
- per cada una fessim un interval $\hat{\theta}_1^{(i)} \leq \theta \leq \hat{\theta}_2^{(i)}$
- podriem afirmar que el $100(1-\alpha)\%$ (per exemple 95%) dels intervals construits contindrien θ, el veritable valor del paràmetre.

Veiem-ho gràficament

Interpretant el nivell de confiança

ESTA PÁGINA MUESTRA LOS RESULTADOS DE UNA SIMULACIÓN POR ORDENADOR DE VEINTE MUESTRAS DE TAMAÑO $n = 1.000$. SUPONEMOS QUE EL VALOR REAL DE $p = 0,5$. EN LA PARTE SUPERIOR PUEDES VER LA DISTRIBUCIÓN MUESTRAL DE $\hat{p}$ (NORMAL, CON MEDIA p Y $\sigma = \sqrt{\frac{p(1-p)}{n}}$). EN LA PARTE INFERIOR SE ENCUENTRAN LOS INTERVALOS DE CONFIANZA DE CADA MUESTRA, AL 95%. COMO MEDIA, UNO DE CADA VEINTE (O UN 5%) DE ESTOS INTERVALOS NO INCLUIRÁ EL PUNTO $p = 0,5$.



3.2 Intervals de confiança per a μ: cas normal i cas general

Sigui $X_1, \ldots, X_n$ una m.a.s. de X. Distingim els següents 3 casos:

1. Interval de confiança per a la mitjana d'una $N(\mu, \sigma^2)$ amb σ^2 coneguda $\Rightarrow$ Prova Z. Fet a l'apartat 3.1.
2. Interval de confiança per a la mitjana d'una $N(\mu, \sigma^2)$ amb σ^2 desconeguda $\Rightarrow$ Prova t de Student
3. Interval de confiança per a la mitjana μ amb σ^2 desconeguda i $n > 30$ $\Rightarrow$ Prova Z

Interval de confiança per a μ quan les dades són $N(\mu,\sigma^2)$ amb σ^2 coneguda. Com a 3.1 amb α i σ^2 qualsevols

Sigui $X_1,\ldots,X_n$ una m.a.s. de $X \sim N(\mu,\sigma^2)$, amb σ^2 coneguda.
Interval de confiança de nivell de confiança $1-\alpha$ pel paràmetre μ.

- Escollim $1-\alpha \Rightarrow$ Calculem $\alpha/2$
- Determinem $z_{\alpha/2}$ fent servir que $Z \sim N(0,1)$

$$\mathrm{Prob}\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha$$

- $X \sim N(\mu,\sigma^2) \Rightarrow \overline{X} \sim N(\mu,\frac{\sigma^2}{n}) \Rightarrow Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow$

$$\begin{aligned}
\mathrm{Prob}\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) &= 1-\alpha \\
\mathrm{Prob}\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &= 1-\alpha \\
\mathrm{Prob}\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &= 1-\alpha
\end{aligned}$$

LLEGIR PEÑA 8.3.1 PÀGINES 323-324

Interval de confiança per a μ quan les dades són $N(\mu, \sigma^2)$ i σ^2 coneguda

- Interval de confiança de nivell de confiança $1-\alpha$:

$$\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- Donades les dades: $x_1, x_2, \cdots, x_n$, calculem $\overline{x}$, l'interval

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

  conté a la mitjana μ amb probabilitat $1-\alpha$.
- Usualment expressem:

$$IC_{1-\alpha}(\mu) = \overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Exemple: Si $\sigma = 1.5$, $n = 12$ i $\overline{x} = 3.6$,
$1-\alpha = 0.95 \Rightarrow z_{\alpha/2} = z_{0.05/2} = 1.96$
$\left(3.6 - 1.96\frac{1.5}{\sqrt{12}}, 3.6 + 1.96\frac{1.5}{\sqrt{12}}\right) = (2.751, 4.449)$ és un interval de confiança al 95% per la mitjana poblacional μ.

Interval de confiança per a μ quan les dades són $N(\mu, \sigma^2)$ i σ^2 desconeguda

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$, amb σ^2 desconeguda. Interval de confiança de nivell de confiança $1-\alpha$ pel paràmetre μ.

- Escollim $1-\alpha \Rightarrow$ Calculem $\alpha/2$
- Calculem $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
- L'estadístic

$$T_{n-1} = \sqrt{n}\frac{\overline{X}-\mu}{S} = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

  es distribueix com una t de Student amb $n-1$ g.ll.
- Es tracta de trobar a i b tal que

$$\mathrm{Prob}\left(a < \frac{\overline{X}-\mu}{S/\sqrt{n}} < b\right) \quad = \quad 1-\alpha$$

- En comptes de Z usarem t_{n-1}

LLEGIR PEÑA 8.3.2 PÀGINES 325-326

LECTURA INDEPENDENT: t de Student amb notació matemàtica

- Y, Z v.a.'s independents
- $Z \sim N(0,1)$
- $Y \sim \chi^2_\nu$

la distribució de la v.a. $t_\nu = \frac{Z}{\sqrt{Y/\nu}}$ s'anomena **distribució t de Student amb ν graus de llibertat**

- Quan n és gran, t_{n-1} s'aproxima MOLT a una $N(0,1)$
- Amb efectes pràctics es pot fer servir la Normal en comptes de la t.
- Per $\alpha/2 = 0.025$ ($\Rightarrow 1 - \alpha/2 = 0.975$), $t_{30;0.975} = 2.042$, $t_{35;0.975} = 2.030$, $t_{60;0.975} = 2.00$, $t_{90;0.975} = 1.987$, i quan $n \to \infty$ $t_{n;0.975} = 2.045 \approx z_{0.975} = 1.96$

Interval de confiança amb nivell de confiança $1-\alpha$ per a μ basat en una t de Student

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$, amb σ^2 desconeguda.

- Escollim $1-\alpha \Rightarrow$ calculem $\alpha/2$
- Trobem a i b tal que

$$\mathrm{Prob}\left(a < \frac{\overline{X}-\mu}{S/\sqrt{n}} < b\right) \quad = \quad 1-\alpha$$

  Com?
- Degut a la simetria de t de Student repartim α de forma equànim a cada cua assignant $\alpha/2$.
- Calculem el percentil $t_{n-1;1-\frac{\alpha}{2}}$. Definim $a = -t_{n-1;1-\frac{\alpha}{2}}$ i $b = t_{n-1;1-\frac{\alpha}{2}}$
-

$$\mathrm{Prob}\left(-t_{n-1;1-\frac{\alpha}{2}} < t_{n-1} = \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{n-1;1-\frac{\alpha}{2}}\right) \quad = \quad 1-\alpha$$

Interval de confiança basat en una t de Student

Fent servir les mateixes manipulacions que abans obtenim

$$
\begin{aligned}
\mathrm{Prob}\left(-t_{n-1;1-\alpha/2} < t_{n-1} = \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{n-1;1-\alpha/2}\right) &= 1-\alpha \\
\mathrm{Prob}\left(-t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}} < \overline{X}-\mu < t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha \\
\mathrm{Prob}\left(\overline{X}-t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X}+t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha
\end{aligned}
$$

L'interval de confiança amb nivell de confiança $1-\alpha$ per a la mitjana μ és

$$\left(\overline{x}-t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{x}+t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

EXERCICI VOLUNTARI 2: Alçada dels estudiants del Grau d'Estadística

Les alçades dels dels estudiants del Grau d'Estadística UB-UPC es distribueixen com una Normal (μ_A, σ_A^2).

1. A quants estudiants haurem de mesurar si, amb una probabilitat del 95% (o superior), volem estimar la mitjana μ_A de l'alçada amb una precisió de 3cm?. Suposeu que la desviació estàndar de les alçades és $\sigma_A = 8.5$cm.
2. A quants estudiants haurem de mesurar si, amb una probabilitat del 95% (o superior), volem l'estimació de μ_A amb una precisió $0.64\sigma_A$ cm i l'estimació de σ_A amb una precisió de $0.8\sigma_A$ cm.
3. Suposat que l'alçada mitjana de la població d'estudiants és de $\mu_A = 174$ cm amb desviació estàndar $\sigma_A = 8.5$cm, simuleu (genereu), amb R, aleatòriament $n = 50$ alçades d'aquesta població. Podeu fer servir la funció `rnorm(n,mu,sigma)=rnorm(50,174,8.5)`.

EXERCICI VOLUNTARI 2: Alçada dels estudiants del Grau d'Estadística

Calculeu $\sum_{i=1}^{50} x_i$ i $\sum_{i=1}^{50} x_i^2$ amb les 50 dades simulades i feu servir aquests valors per construir els següents intervals de confiança per la mitjana poblacional de les alçades dels estudiants μ_A :

4. Calculeu intervals de confiança per μ_A al 90% i al 95%, suposant que $\sigma_A = 8.5$cm.
5. Calculeu intervals de confiança per μ_A al 90% i al 95%, estimant la desviació estàndar de la població a partir de les dades simulades
6. Entre quins valors es troba la mitjana poblacional de l'alçada dels estudiants del Grau d'Estadística i perquè?

EXERCICI VOLUNTARI 2: Alçada dels estudiants del Grau d'Estadística

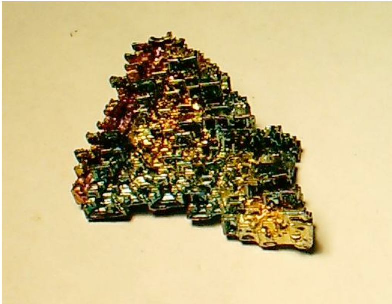
Simuleu $k = 100$ mostres d'alçades de grandària $n = 50$ a partir d'una variable aleatòria Normal de mitjana $\mu_A = 174$ cm i desviació estàndar $\sigma_A = 8.5$cm.

7. Calculeu els $k = 100$ intervals de confiança per μ_A al 90% deduits de les 100 mostres simulades havent estimat la desviació estàndar de la població a partir de les dades simulades
8. Quina proporció d'intervals conté el valor 174? Noteu que heu generat dades d'una normal amb mitjana $\mu_A = 174$.
9. Escriviu amb les vostres pròpies paraules el que enteneu per interval de confiança a partir de les conclusions d'aques exercici.

RESOLEU-LO i LLIUREU-LO EL DIVENDRES 1 D'ABRIL

Exemple Bismut: element químic, Bi, nombre atòmic 83

El contingut de bismut en un aliatge en 10 mostres és (en ppm): 19.7, 21.4, 12.5, 13.8, 13.5, 20.4, 18.2, 16.6, 15.2, 16.3



$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{167.6}{10} = 16.76$$

$$S^2 = \frac{1}{9}\left(\sum_{i=1}^{10} X_i^2 - 10(16.76)^2\right) = \frac{85.704}{9} = 9.52 \Rightarrow S = 3.08$$

La mitjana del contingut de Bismut en aquest aliatge és de 16.76 ppm amb una desviació estàndar de 3.086 ppm.

Interval de confiança al 95% pel contingut mig de Bismut

Un interval de confiança al 95% $\Longrightarrow 1-\alpha=0.95$
L'interval de confiança general és:

$$\left(\overline{x}-t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}},\overline{x}+t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

que traduït al nostre exemple serà:

$$\left(16.76-t_{9;1-0.05/2}\frac{3.086}{\sqrt{10}},16.76+t_{9;1-0.05/2}\frac{3.086}{\sqrt{10}}\right)$$

$$\left(16.76-2.208,16.76+2.208\right)=\left(14.552,18.968\right)$$

Com que $n=10$, hem mirat a la taula de la distribució de la t d'Student amb $n-1=10$ g.ll.: $t_{9;0.975}=2.2622$ i com $1-\alpha=0.95$ hem agafat $\alpha/2=0.025$.)

Interval de confiança al 99% pel contingut mig de Bismut

Si ara en volguéssim un interval de confiança al 99%, en aquest cas $\alpha = 1 - 0.99 = 0.01$ i només cal que substituïm $t_{9;0.975} = 2.208$ per $t_{9;0.995} = 3.2498$ i l'interval quedaria de la forma:

$$\left(16.76 - t_{9;1-0.01/2}\frac{3.086}{\sqrt{10}}, 16.76 + t_{9;1-0.01/2}\frac{3.086}{\sqrt{10}}\right) =$$

$$\left(16.76 - t_{9;0.995}0.976, 16.76 + t_{9;0.995}0.976\right)$$

per tant s'obté l'interval

$$\left(13.588, 19.932\right)$$

Comparem-lo amb l'interval de confiança al 95%: $\left(14.552, 18.968\right)$. Notem que continua centrat en la mitjana 16.76 però l'interval s'ha fet més gran ja que estem exigint més precisió al resultat.

Interval de confiança per a la mitjana: cas general

Sigui $X_1, X_2, \ldots, X_n \sim X$ i no coneixem la distribució de X, podria ser exponencial, lognormal, normal, etc.
Sigui μ la mitjana de la població ($E(X) = \mu$) i σ^2 la variància de la població ($\sigma^2 = \mathrm{Var}(X)$.)
Volem un interval de confiança amb nivell de confiança $1 - \alpha$ per μ.
Ens basarem en el Teorema Central del Límit (TCL) i en la llei dels Grans Nombres (LGN) que ens asseguren, per mostres de grandària n gran, que

- La mitjana mostral $\overline{X} \sim N(\mu, \sigma^2/n)$ (TCL)
- La variància mostral S^2 es comporta com σ^2 (LGN)
- $$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$
- Podem determinar el valor $z_{\alpha/2}$ de manera que

$$\mathrm{Prob}\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

Interval de confiança per a la mitjana: cas general

Per a mostres grans de qualsevol població, l'interval

$$\left(\overline{x} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

obtingut a partir d'un valor qualsevol de $\overline{x}$, conté a la mitjana μ amb una probabilitat $1-\alpha$ i per tant

$$IC_{1-\alpha}(\mu) = \overline{x} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$$

és un interval de confiança per μ de nivell $1-\alpha$.
LLEGIR PEÑA 8.4 PÀGINA 326

3.3 Interval de confiança per a una proporció

Sigui p la proporció desconeguda d'una certa característica C en una població. Com fer per estimar-la?

- Prenem una mostra $X_1, X_2, \ldots, X_n$ mostra aleatòria de X on $X = 1$ si l'individu presenta C, $X = 0$ si l'individu no presenta C
- De fet $X \sim \mathrm{Bern}(p)$
- Estimem p mitjançant la proporció d'individus que presenten C $\Rightarrow \hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}$
- Recordem que $\sum_{i=1}^n X_i \sim \mathrm{Bin}(n, p) =$ nombre d'individus que verifiquen la característica d'interès i $\mathrm{Var}(\sum_{i=1}^n X_i) = np(1-p)$
- $\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\mathrm{Var}(\sum_{i=1}^n X_i)}{n^2} = \frac{p(1-p)}{n}.$

Interval de confiança per a una proporció per n gran

Suposem que ens basem en un nombre d'individus prou gran, posem $n > 30$

- Pel Teorema Central del Límit, $\sum_{i=1}^{n} X_i \sim N(np, np(1-p))$
- $\hat{p} \sim N(p, \frac{p(1-p)}{n})$
- $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
- $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$
- Un interval de confiança per p amb un nivell de confiança $100(1-\alpha)$ ve donat per:

$$\left(\hat{p} - z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

Recordem que es vol estimar el percentatge de nens obesos en la comunitat Americana-Latina (no mexicana) i Americana-Xinesa. A l'article parlant de la Prevalença de sobrepes.

- *Overweight prevalence*=Prevalença de sobrepes: quantifica la proporció de nens obesos en un període donat
- Com es defineix el sobrepes infantil?
- Hi ha el mateix llindar per qualsevol edat?
- Hi ha el mateix llindar per nens i nenes?
- La resposta és negativa: El llindar de sobrepes infantil és específic per grups d'edat i per gènere

HEALTHY FOR LIFE: Interval de confiança per la prevalença de sobrepes

If a simple random sample design had been used, overweight prevalence could have been estimated by $\hat{p} = x/n$, where x is the number of overweight children in the sample or relevant subsample (e.g., children aged 2–5) and n is the total sample or subsample size. However, the sample design employed needs to be taken into account for the analysis.

In a simple random sample, the variance of $\hat{p}$ could have been estimated by $\hat{p}(1-\hat{p})/n$, and 95% confidence intervals given by $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$. The use of the sample weights to

L'interval de confiança per la prevalença és:

$$\left(LI, LS\right) = \left(\hat{p} - z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

on LI i LS indiquen el límit inferior i superior de l'interval.

Healthy for Life: Intervals de confiança per la prevalença

- A partir de les dades de l'article ($n = 2474$)
- 990 nens: 490 nens i 500 nenes entre 2 i 5 anys
- 1484 nens: 740 nens i 744 nenes entre 6 i 11 anys
- NS: nombre de nens/nenes a cada categoria amb sobrepes
- LI i LS el límit inferior i superior de l'interval de confiança per la proporció de sobrepes al 95%.

Els intervals de confiança per la prevalença de sobrepes (proporció de nens/nenes amb sobrepes) en cada categoria d'edat.

	EDAT	n	NS	$\hat{p}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	LI	LS
Both	2-5	990	216	0.218	0.0131	0.1925	0.2439
Both	6-11	1484	353	0.238	0.0111	0.2162	0.2595
Nens	2-5	490	112	0.229	0.0190	0.1914	0.2658
Nens	6-11	740	172	0.232	0.0155	0.2020	0.2629
Nenes	2-5	500	104	0.208	0.0182	0.1724	0.2436
Nenes	6-11	744	181	0.243	0.0157	0.2124	0.2741

A la WHO: World Health Organization (Organización Mundial de la Salud (OMS))

http://www.who.int/mediacentre/factsheets/fs311/es/index.ht

¿Qué son la obesidad y el sobrepeso? *La obesidad y el sobrepeso se definen como una acumulación anormal o excesiva de grasa que puede ser perjudicial para la salud. El índice de masa corporal (IMC) -el peso en kilogramos dividido por el cuadrado de la talla en metros (kg/m^2)- es una indicación simple de la relación entre el peso y la talla que se utiliza frecuentemente para identificar el sobrepeso y la obesidad en los adultos, tanto a nivel individual como poblacional. La OMS define sobrepeso cuando IMC ≥ 25, y obesidad cuando IMC ≥ 30. Los nuevos Patrones de crecimiento infantil presentados por la OMS en abril de 2006 incluyen tablas del IMC para lactantes y niños de hasta 5 años. La medición del sobrepeso y la obesidad en niños de 5 a 14 años es difícil porque no hay una definición normalizada de la obesidad infantil que se aplique en todo el mundo.*

Consulteu el CDC: Center for Disease Control.

`http://www.cdc.gov/healthyweight/assessing/bmi`

Calculating and interpreting BMI for children and teens involves the following steps:

1. Obtain accurate height and weight measurements.

2. Calculate the BMI and percentile using the Child and Teen BMI Calculator.

3. Review the calculated BMI-for-age percentile and results.

BMI is both age-and sex-specific for children and teens because:.

- *The amount of body fat changes with age.*
- *The amount of body fat differs between girls and boys.*

4. Find the weight status category for the calculated BMI-for-age percentile as shown in the following table:

- Underweight ⇔ Less than the 5th percentile
- Healthy weight ⇔ 5th percentile to less than the 85th percentile
- Overweight ⇔ 85th to less than the 95th percentile
- Obese ⇔ Equal to or greater than the 95th percentile

3.4 Interval de confiança per la variança d'una població normal

Sigui $X \sim N(\mu, \sigma^2)$ i volem construir l'interval de confiança per σ^2 a partir d'una mostra $X_1, \ldots, X_n \sim X$.
Calcularem la variància mostral S^2 i farem servir la distribució χ^2 amb $n-1$ graus de llibertat, tenint en compte que

$$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

Es tracta de trobar a i b tal que

$$\mathbf{P}\left(a < \frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1} < b\right) \;=\; 1-\alpha$$

En aquest cas la llei χ^2 amb $n-1$ graus de llibertat no és simètrica, i per tant encara que repartim $1-\alpha$ entre les dues cues, no podrem fer com abans $a = -b$, sino que haurem de trobar dos valors diferents $a = \mathcal{X}^2_{n-1;\alpha/2}$ i $b = \mathcal{X}^2_{n-1;1-\alpha/2}$.

LLEGIR PEÑA 8.5 PÀGINES 327-330

Interval de confiança per la variància d'una Normal

Els dos valors diferents $a = \mathcal{X}^2_{n-1;\alpha/2}$ i $b = \mathcal{X}^2_{n-1;1-\alpha/2}$ són tals que

$$\mathbf{P}\Big(\frac{(n-1)S^2}{\sigma^2} < \mathcal{X}^2_{n-1;\alpha/2} = a\Big) = \frac{\alpha}{2}$$

$$\mathbf{P}\Big(\frac{(n-1)S^2}{\sigma^2} < \mathcal{X}^2_{n-1;1-\alpha/2} = b\Big) = 1 - \frac{\alpha}{2}$$

i per tant

$$\mathrm{Prob}\left(a = \mathcal{X}^2_{n-1;\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \mathcal{X}^2_{n-1;1-\alpha/2} = b\right) = 1 - \alpha$$

o equivalentment

$$\mathrm{Prob}\left(\frac{1}{b} = \frac{1}{\mathcal{X}^2_{n-1;1-\alpha/2}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\mathcal{X}^2_{n-1;\alpha/2}} = \frac{1}{a}\right) = 1 - \alpha$$

Interval de confiança per la variància d'una Normal

Manipulem aquesta expressió per tal "d'aïllar" σ^2. Teniem

$$\mathrm{Prob}\Big(\frac{1}{\mathcal{X}^2_{n-1;1-\alpha/2}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\mathcal{X}^2_{n-1;\alpha/2}}\Big) = 1-\alpha$$

o, el que és el mateix

$$\mathrm{Prob}\Big(\frac{(n-1)S^2}{\mathcal{X}^2_{n-1;1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\mathcal{X}^2_{n-1;\alpha/2}}\Big) = 1-\alpha$$

Obtenim, doncs, l'interval de confiança per σ^2:

$$IC_{1-\alpha}(\sigma^2) = \left(\frac{(n-1)S^2}{\mathcal{X}^2_{n-1;1-\alpha/2}}, \frac{(n-1)S^2}{\mathcal{X}^2_{n-1;\alpha/2}}\right)$$

Notem que l'interval no és simètric respecte del valor de S^2.

Exemple Bismut

Recordem que $n = 10$, $\overline{x} = 16.76$ i $S^2 = 9.52$ i un interval de confiança al 95% per μ és $(14.552, 18.968)$.
Trobem a la taula de la χ^2_9:

$$a = \mathcal{X}^2_{n-1;\alpha/2} = \mathcal{X}^2_{9;0.025} = 2.70$$

$$b = \mathcal{X}^2_{n-1;1-\alpha/2} = \mathcal{X}^2_{9;0.975} = 19.02$$

i per tant l'interval de confiança per σ^2 al 95% serà:

$$\left(\frac{9 \cdot 9.52}{19.02}, \frac{9 \cdot 9.52}{2.7}\right) = \left(\frac{85.68}{19.02}, \frac{85.68}{2.7}\right) = (4.505, 31.73)$$

En resum, les dades recollides del bismut indiquen que la millor estimació de la mitjana μ de l'aliatge de Bismut és 16.76 ppm i l'interval $(14.5, 19)$ conté μ amb alta probabilitat. La millor estimació de la desviació típica és 3.08 ($\sqrt{9.52}$) i l'interval $(2.12, 5.63) = (\sqrt{4.5}, \sqrt{31.7})$ conté σ amb alta probabilitat.