# Intervals de confiança

Àngel J. Gil Estallo

P08/05057/02307

Universitat Oberta de Catalunya
www.uoc.edu

# Índex

# Introducció als intervals de confiança.
# El cas de la mitjana aritmètica

La inferència estadística proporciona mètodes per obtenir conclusions a partir d'un conjunt de dades. La majoria de les vegades no tindrem una certesa absoluta de les conclusions a les quals arribem. La teoria de la probabilitat fonamentarà les conclusions obtingudes i permetrà d'establir la precisió dels mètodes utilitzats. A continuació, treballarem els anomenats *intervals de confiança*, començant pel cas més senzill, a partir del qual introduirem les definicions generals.

## 1. El concepte d'interval de confiança

A partir d'un cas concret anirem introduint gradualment les idees i les tècniques que sustenten la construcció dels anomenats *intervals de confiança*. En concret, ens plantegem estudiar la mitjana de les alçades dels estudiants de la UOC. Veurem, pas a pas, quin seria el procediment que seguiríem:

**1)** Establim algunes *hipòtesis prèvies,* que són les que determinen les distribucions que cal utilitzar en la construcció de l'interval. Aquestes hipòtesis prèvies permetran, en definitiva, d'utilitzar resultats de la teoria de la probabilitat.

En aquest exemple n'utilitzarem dues:

**a)** La distribució de les alçades segueix una llei normal, que tindrà una mitjana $\mu$ (que suposem desconeguda) i una desviació típica $\sigma$.

**b)** Coneixem, d'estudis anteriors, el valor de la desviació típica poblacional $\sigma$; suposem que tenim $\sigma = 10$ cm.

**2)** A continuació, efectuem la *recollida de les dades* adients al problema. Normalment, seleccionarem una mostra aleatòria simple de la població i obtindrem les dades requerides a partir dels individus de la mostra.

**La utilitat de les mostres**

En la immensa majoria dels casos no podem accedir a les dades de tota la població, bé perquè és inviable o perquè resulta massa car!

Suposem que obtenim una mostra aleatòria simple de 121 alumnes de la UOC, als quals en preguntem l'alçada.

**3)** A partir de les dades obtingudes, calculem resums numèrics adients.

Com que estem interessats en la mitjana de les alçades, calculem la mitjana d'aquestes 121 observacions. Suposem que obtenim que aquesta mitjana és $\bar{x} = 171$ cm.

En aquest cas, la mitjana poblacional $\mu$ és un **paràmetre** de la població que volem descriure i la mitjana mostral $\bar{x}$ és un **estadístic** que ens permet d'aproximar el valor del paràmetre.

**4)** Utilitzem la teoria de la probabilitat per a obtenir una relació entre el paràmetre i l'estadístic que permeti de construir un cert interval, anomenat *interval de confiança*. Aquest interval ha de tenir les propietats següents:

**a)** Cal que estigui centrat en el valor de l'estadístic (que és el valor que, efectivament, coneixem).

**b)** Ja que podem obtenir mostres diferents de la població, i cada mostra donarà lloc a una mitjana mostral diferent, hem de garantir, per a un percentatge alt de les possibles mostres, que el procediment produeixi intervals que continguin el veritable valor del paràmetre.

**La confiança en l'interval**

Hem de mesurar d'alguna manera la confiança que podem tenir en l'interval

Aquest percentatge de mostres que donen lloc a intervals que contenen el veritable valor del paràmetre és l'anomenat **nivell de confiança**.

Així, doncs, un **interval de confiança** per a un cert paràmetre amb un nivell de confiança de $C$% és un interval calculat a partir d'una mostra de manera que el procediment de càlcul garanteix que el $C$% de les mostres donen lloc a un interval que conté el valor real del paràmetre.

Veurem com es pot obtenir un interval amb aquestes característiques per a l'exemple de les alçades dels estudiants de la UOC.

Ara ens preguntem què podem dir de la mitjana aritmètica de les alçades de tots els estudiants de la UOC. Observem que aquesta mitjana és la mitjana poblacional $\mu$, ja que suposem que la distribució de les alçades és normal amb mitjana precisament $\mu$. També sabem que, llevat que preguntem a tots i cadascun dels estudiants de la UOC quina és la seva alçada, mai no aconseguirem de saber quin és el veritable valor de la mitjana poblacional $\mu$.

La pregunta és, doncs: quines conclusions podem extreure sobre el valor de la mitjana poblacional $\mu$, a partir de les dades de què disposem, i en concret a partir de la mitjana mostral $\bar{x} = 171$ cm? És a dir, ens preguntem quina relació podem establir entre la veritable mitjana poblacional $\mu$ i la mitjana mostral $\bar{x}$. Observeu que no podem dir que siguin iguals, ni que una sigui més gran que l'altra. A més, la mitjana mostral depèn de la mostra i això vol dir que mostres diferents proporcionen resultats diferents.

Haurem de recórrer als nostres coneixements de probabilitat; en concret, utilitzarem que per a la mitjana de les alçades de la UOC, la variable:

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X}-\mu}{\frac{10}{\sqrt{121}}}$$

segueix una distribució normal estàndard. Aquesta expressió permet d'establir una relació indirecta entre $\mu$ i $\bar{x}$ . Aprofitarem aquest fet per construir l'interval de confiança seguint la seva definició.

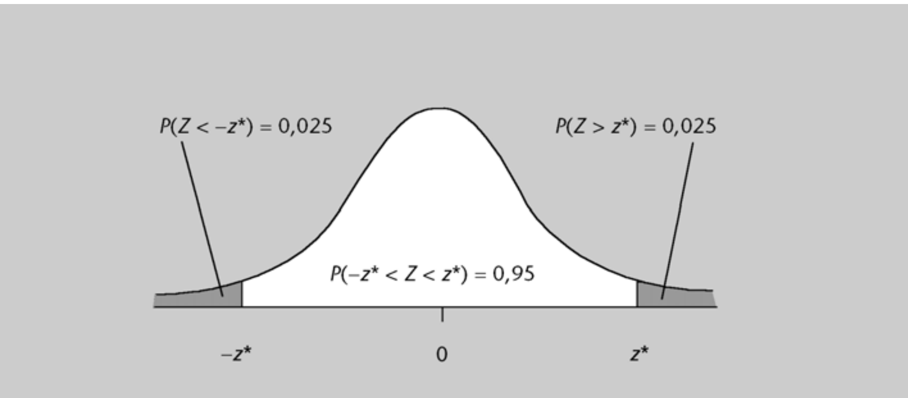
Suposem que volem un nivell de confiança del 95%. Començarem per construir un interval de la forma ($-z^*$, $z^*$) centrat en el valor 0, de manera que els valors que pren la variable aleatòria $Z$:

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

pertanyin a aquest interval amb una probabilitat del 0,95; és a dir, busquem uns valors $-z^*$ i $z^*$ per als quals:

$$P\left(-z^* \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z^*\right) = 0,95$$

En aquestes condicions, i tal com es veu en el gràfic següent:




el valor de $z^*$ és aquell valor tal que $P(Z \geq z^*) = (1 - 0,95)/2 = 0,025$ (ja que $Z$ segueix una llei normal estàndard) i, per tant, tenim que $z^* = 1,96$ (aquest valor es pot obtenir a partir de les taules de la distribució normal estàndard); és a dir:

$$P\left(-1,96 \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1,96\right) = 0,95$$

Ara, operant aquesta expressió amb l'objectiu d'aïllar el valor de μ:

$$P\left(-\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

i, finalment:

$$P\left(\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

**Feu memòria**

Recordeu que l'expressió $-a \leq b$ és equivalent a aquesta altra: $-b \leq a$.

Aquest resultat ens diu que la probabilitat que la mitjana poblacional μ pertanyi a un interval de la forma:

$$\left(\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}},\ \overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$

és de 0,95; o el que és el mateix: 95 de cada 100 vegades que escollim una mostra aleatòria simple i calculem el valor de la mitjana mostral, l'interval que obtindré substituint el valor de $\overline{X}$ per la mitjana corresponent a la mostra de què disposem contindrà el veritable valor de μ.

Finalment, l'únic que hem de fer és substituir $\overline{X}$, σ i $n$ pels valors corresponents, i recordar que treballem amb una confiança del 95%. Fent la substitució, doncs, obtenim que l'interval:

$$\left(171 - 1{,}96\frac{10}{\sqrt{121}},\ 171 + 1{,}96\ \frac{10}{\sqrt{121}}\right) = (169{,}22\ ,\ 172{,}78)$$

és un interval de confiança amb un nivell de confiança del 95% per a la mitjana de les alçades dels alumnes de la UOC.

Abans de continuar, és important remarcar alguns fets:

**a)** L'interval està centrat en el valor de la mitjana de la mostra.

**b)** No sabem si l'interval conté o no la mitjana poblacional μ i no hi ha manera de saber-ho, llevat que coneguem el valor de μ. (I si coneixem el valor de μ no cal que ens esforcem a buscar-hi el seu interval de confiança.)

**c)** Per a cada mostra obtinguda de la població tenim un valor de la mitjana mostral i, per tant, un interval de confiança que pot ser diferent del que hem obtingut.

**d)** L'expressió "confiança del 95%" indica "confiança en el mètode" emprat, de manera que el 95% de les vegades que apliquem el mètode a la mateixa població obtindrem intervals que sí contenen la mitjana poblacional μ.

**No s'ha de confondre**

Dir que μ pertany a l'interval amb una probabilitat del 0,95 és incorrecte, ja que la μ pertany a aquest interval (amb la qual cosa la probabilitat és u) o no hi pertany (amb la qual cosa la probabilitat que hi pertanyi és zero).

Resumim a continuació les idees i procediments que ens han portat a la construcció del nostre interval de confiança. Evidentment, la construcció depèn de les hipòtesis de partida i, com veurem al final de la sessió, diferents hipòtesis donen lloc a diferents tipus d'intervals.

## 2. Interval de confiança per a la mitjana aritmètica quan la població és normal i coneixem la desviació típica

Suposem que la variable que volem estudiar segueix una llei normal de mitjana μ (desconeguda) i desviació típica σ coneguda i que disposem d'una mostra aleatòria simple de mida $n$ i del valor de la mitjana de la mostra $\bar{x}$. Aleshores:

**Notació**

El nivell de confiança també es denota per (1 – α)100%; normalment considerarem (1 – α) igual a 90%, 95% o 99%.

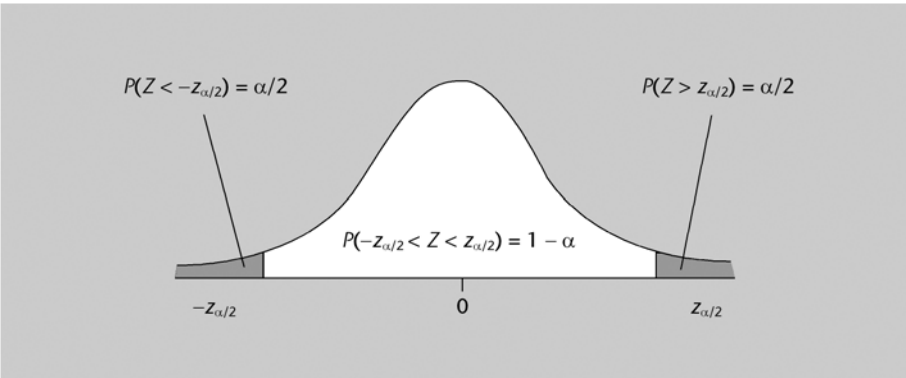
**a)** Fixem el nivell de confiança (en forma de percentatge), que habitualment escriurem com a (1 – α).

**b)** Calculem l'error estàndard de la mitjana $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

**c)** Calculem l'anomenat *valor crític*, que és aquell punt $z_{\alpha/2}$ tal que $P(Z \geq z_{\alpha/2}) = \alpha/2$, on $Z$ és una variable $N(0, 1)$. Gràficament:

**Notació del valor crític**

En l'exemple de les alçades dels estudiants de la UOC s'ha utilitzat la notació $z^*$, que ara és substituïda per $z_{\alpha/2}$ per a poder precisar el valor crític segons el nivell de confiança.




Per als nivells de confiança usuals, els valors crítics corresponents són els següents:

- $(1 - \alpha) = 90\% = 0{,}9$, $\alpha = 0{,}1$ i $z_{\alpha/2} = z_{0,05} = 1{,}645$
- $(1 - \alpha) = 95\% = 0{,}95$, $\alpha = 0{,}05$ i $z_{\alpha/2} = z_{0,025} = 1{,}96$
- $(1 - \alpha) = 99\% = 0{,}99$, $\alpha = 0{,}01$ i $z_{\alpha/2} = z_{0,005} = 2{,}575$

**d)** Calculem l'anomenat **marge d'error** (també anomenat **precisió de l'estimació**) com a $z_{\alpha/2}$ per a l'error estàndard, és a dir:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

**e)** L'interval de confiança obtingut amb la mostra de partida és el següent:

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

**Observació**

Per tant, el marge d'error és la meitat de la longitud d'interval de confiança.

Mentre que $(1 - \alpha)$ és el nivell de confiança, $\alpha$ és l'anomenat **nivell de significació** i es correspon a la proporció de mostres a partir de les quals l'interval construït segons el procediment explicat no conté el veritable valor del paràmetre que es vol aproximar.

### 2.1. L'efecte de la mida de la mostra

Moltes vegades, fixat el nivell de confiança, ens marcarem com a objectiu donar el valor del paràmetre $\mu$ amb una certa precisió. L'única manera d'obtenir la precisió desitjada consisteix a modificar adequadament la mida de la mostra. Suposem que desitgem una precisió o marge de error *ME*; com que sabem que

$$ME = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

aïllant *n* obtenim:

$$\text{Grandària de la mostra} = n \geq (z_{\alpha/2})^2 \frac{\sigma^2}{ME^2}$$

**Mida de la mostra**

És fàcil veure que si volem reduir l'amplada de l'interval de confiança a la meitat, haurem de prendre una mostra quatre vegades més gran.

**Temps de connexió al Campus Virtual**

En un estudi fet a la UOC es pren una mostra aleatòria de 150 estudiants i es demana quant temps van estar connectats al Campus Virtual durant el mes d'abril de l'any 2000. S'obté una mitjana mostral de 120 minuts. Suposem, a més, que el temps de connexió al Campus Virtual durant el mes d'abril de 2000 segueix una distribució normal amb desviació típica de deu minuts. Podem calcular un interval de confiança del 95% per al temps de connexió durant aquest mes, considerant la mitjana mostral $(\bar{x} = 120)$ i l'error estàndard de la mitjana, que és:

$$\sigma_{\bar{x}} = \frac{10}{\sqrt{150}} = 0{,}816$$

Per tant, l'interval de confiança amb un nivell de confiança del 95% és $(120 \pm 1{,}96 \cdot 0{,}816) = (120 \pm 1{,}59936) = (118{,}40\ ,\ 121{,}60)$. Si volem que la precisió del nostre interval sigui de cinc punts percentuals, haurem d'aconseguir un marge d'error inferior a 5% = 0,05. Per tant:

$$1{,}96\sigma_{\bar{x}} = 1{,}96\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{10}{\sqrt{n}} < 0{,}05$$

i aïllant $n$ obtenim $n > 153.664$ i, per tant, necessitaríem una mostra inabastable, ja que a hores d'ara la UOC no té tants estudiants.

## 2.2. Consideracions sobre la normalitat

El procediment que s'ha presentat és vàlid per a variables que segueixen lleis normals de mitjana μ, ja que en aquest cas la variable següent segueix una llei normal estàndard:

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

D'altra banda, el teorema del límit central afirma que donada qualsevol variable aleatòria $X$ amb mitjana μ, si la mida de les mostres considerades és $n > 30$, aleshores la variable següent també es comporta com una distribució normal estàndard:

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

**Conseqüència del teorema del límit central**

L'únic cas en què no sabrem calcular intervals de confiança per a la mitjana (suposant coneguda la desviació típica) serà quan les poblacions siguin no normals de mesura menor que trenta.

Aquest fet fa que el procediment de càlcul d'un interval de confiança per a la mitjana sigui vàlid, encara que la variable que cal estudiar no sigui normal, sempre a condició que la mida de la mostra sigui superior a trenta.

# 3. Interval de confiança per a la mitjana quan la població és normal i desconeixem la desviació típica

En aquest cas, procedirem tal com es fa quan s'estudia la distribució de la mitjana mostral quan la desviació típica és desconeguda: estimarem la desviació típica usant els valors mostrals i treballarem amb la distribució de la mitjana mostral $\overline{X}$, ja que, per un procediment semblant a la estandardització, la podem relacionar amb una altra variable que segueix una distribució de Student. A continuació, repetirem el cas de les alçades dels alumnes de la UOC perquè pugueu comparar tots dos casos:

**Mitjana de les alçades dels estudiants de la UOC si la desviació típica és desconeguda**

**1)** La hipòtesi prèvia serà la següent: la distribució de les alçades segueix una llei normal, que tindrà una certa mitjana μ i una certa desviació típica σ, totes dues desconegudes.

**2)** La recollida de dades consisteix a seleccionar una mostra aleatòria simple en la població. Suposem que obtenim una mostra aleatòria simple de 121 individus, als quals preguntem la seva alçada.

**3)** En l'apartat de càlculs ara necessitem:

**a)** La mitjana d'aquestes 121 observacions. Suposem que obtenim $\bar{x} = 171$ cm.
**b)** La desviació típica d'aquestes observacions:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{121}(x_i - \bar{x})^2}$$

Suposem que en el nostre exemple obtenim $s$ = 11 cm.

**4)** Fixem un nivell de confiança; suposem que interessa un nivell de $(1 - \alpha) = 95\%$. Per a calcular l'interval de confiança, en aquest cas hem de treballar amb la variable aleatòria següent:

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

Com sabem, aquesta variable segueix una distribució $t_{n-1}$ de Student amb $n - 1$ graus de llibertat. Ara necessitem un interval de la forma $(-t^*, t^*)$, centrat també en 0, ja que la distribució és simètrica al voltant de 0, i de manera que:

$$P\left(-t^* \le \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \le t^*\right) = 0{,}95$$

En aquestes condicions, i per un raonament anàleg al seguit quan la desviació típica era coneguda, el valor de $t^*$ és aquell valor que verifica:

$$P(t_{n-1} > t^*) = P(t_{120} > t^*) = (1 - 0{,}95)/2 = 0{,}025$$

Per tant, consultant a les taules, per exemple, tenim que $t^* = 1{,}98$; és a dir:

$$P\left(-1{,}98 \le \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \le 1{,}98\right) = 0{,}95$$

Ara, operant aquesta expressió amb l'objectiu d'aïllar el valor de $\mu$ obtenim:

$$P\left(\overline{X} - 1{,}98\frac{s}{\sqrt{n}} \le \mu \le \overline{X} + 1{,}98\frac{s}{\sqrt{n}}\right) = 0{,}95$$

Aquest resultat ens diu que el 95% dels intervals de la forma següent:

$$\left(\overline{X} - 1{,}98\frac{s}{\sqrt{n}}, \overline{X} - 1{,}98\frac{s}{\sqrt{n}}\right)$$

contenen el veritable valor de la mitjana poblacional. I ara ja arribem al final, l'únic que hem de fer és substituir els valors de la mitjana mostral, $s$ i $n$, pels valors obtinguts en el nostre estudi, amb la confiança del 95% que l'interval que obtenim contindrà el veritable valor de $\mu$. Fent la substitució, doncs, obtenim que l'interval:

$$\left(171 - 1{,}98\frac{11}{\sqrt{121}}, 171 + 1{,}98\frac{11}{\sqrt{121}}\right) = (169{,}02\ ,172{,}98)$$

és un interval de confiança amb un nivell del 95% per a la mitjana de les alçades dels estudiants de la UOC. És a dir, en el nostre exemple, la mitjana de les alçades dels estudiants de la UOC es troba entre 169,02 cm. i 172,98 cm, amb una confiança del 95%.

Resumim a continuació el procediment per a construir un interval de confiança per a la mitjana aritmètica quan la població és normal i desconeixem la desviació típica. Suposem que la variable que volem estudiar segueix una llei normal de mitjana $\mu$ (desconeguda) i desviació típica $\sigma$ també desconeguda i que disposem d'una mostra aleatòria simple de mida $n$ i de valor de la mitjana de la mostra $\bar{x}$ . Aleshores:

**1)** Fixem el nivell de confiança, que habitualment l'escriurem com $(1 - \alpha)\%$.

**2)** Calculem la desviació típica mostral: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

**3)** Calculem l'error estàndard de la mitjana $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

**4)** Calculem l'anomenat *valor crític*, que és aquell punt $t_{\alpha/2,n-1}$ tal que:

$$P(t_{n-1} \geq t_{\alpha/2,\, n-1}) = \alpha/2$$

on $t_{n-1}$ és una variable Student amb $n$–1 graus de llibertat.

**5)** Calculem l'anomenat *marge d'error* (també anomenat *precisió de l'estimació*) com $t_{\alpha/2,n-1}$ per a l'error estàndard, és a dir, com:

$$t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}}$$

**6)** L'interval de confiança obtingut amb la mostra de partida és el següent:

$$(\bar{x} \pm t_{\alpha/2,\, n-1}s_{\bar{x}}) = \left(\bar{x} - t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}}\right)$$

### 3.1. La mida de la mostra

En aquest cas no tenim un procediment directe per a trobar la mida d'una mostra que produeix un determinat marge d'error, ja que el marge d'error depèn de la desviació típica mostral ($s$), que és un valor que calculem precisament un cop la mostra ja ha estat seleccionada. En cas que necessitem aproximar la mida de la mostra, prendrem una mostra de prova que ens doni un valor aproximat de $s$. Després calcularem el marge d'error per a aquest valor $s$ i prendrem una mostra d'amplada suficient per a garantir aquest marge d'error.

D'altra banda, i tal com passava en el cas de l'interval de confiança de la mitjana amb la desviació típica coneguda, el procediment de càlcul de l'interval de confiança quan la desviació típica és desconeguda també és vàlid, sempre que la mostra sigui major que trenta, encara que la variable que cal estudiar no segueixi una llei normal.

**Temps de connexió al Campus Virtual amb desconeixement de la desviació típica poblacional**

En un estudi fet a la UOC es pren una mostra aleatòria de 150 estudiants i es demana quant temps van estar connectats al Campus Virtual durant el mes d'abril de l'any 2000. S'obté una mitjana mostral de 120 minuts i una desviació típica mostral de 10 minuts. Com que la mostra és gran ($n > 30$), podem calcular un interval de confiança del 95% per al temps de connexió durant aquest mes, considerant la mitjana mostral ($\bar{x} = 120$) i l'error estàndard de la mitjana, que és:

$$s_{\bar{x}} = \frac{10}{\sqrt{150}} = 0{,}816$$

Ens cal calcular $t_{\alpha/2,\, n-1} = t_{0{,}025,\, 149}$, que és aquell valor per al qual $P(t_{149} > t_{0{,}025,149}) = 0{,}025$ ; utilitzant algun programa d'ordinador, obtenim $t_{0{,}025,\, 149} = 1{,}976$. Per tant, l'interval de confiança amb un nivell de confiança del 95% és $(120 \pm 1{,}976 \cdot 0{,}816) = (120 \pm 1{,}612416) = (118{,}39,\ 121{,}61)$.

En cas que volguéssim obtenir un interval de confiança del 99%, hauríem de calcular $t_{0{,}005,\, 149} = 2{,}6092$ i l'interval de confiança seria $(120 \pm 2{,}6092 \cdot 0{,}816) = (117{,}87,\ 122{,}13)$.

Si volem que la precisió del nostre interval (al nivell de confiança del 95%) sigui de cinc punts percentuals, haurem d'aconseguir un marge d'error inferior al 5% = 0,05. Podem utilitzar la mostra que tenim per a donar una aproximació al valor de *s* i, per tant, podem aplicar la fórmula:

$$1{,}976\sigma_{\bar{x}} = 1{,}976\frac{s}{\sqrt{n}} = 1{,}976\frac{10}{\sqrt{n}} < 0{,}05$$

aïllant $n$ obtenim $n > 156.183{,}04$ la qual cosa vol dir que tornem a necessitar una mostra inabastable.

## 4. Comparació entre els casos estudiats

Hem vist com hem de construir, a partir dels resultats obtinguts de les observacions d'una mostra aleatòria simple, els anomenats *intervals de confiança*. Es parteix d'un paràmetre poblacional desconegut i d'algunes hipòtesis sobre la distribució de la variable d'interès. Fixat un cert nivell de confiança C%, el mètode de construcció dels intervals garanteix que el C% de les mostres produeixen un interval que conté el veritable valor del paràmetre desconegut.

**Notació**

En molts llibres el nivell de confiança es denota per $(1 - \alpha)100\%$.

En cas de voler trobar intervals de confiança per a la mitjana aritmètica i suposant que la variable que cal considerar segueix una distribució normal, trobem que els intervals de confiança estan centrats en la mitjana mostral i, per tant, tenen la forma següent:

$$(\bar{x} \pm \text{Marge d'error})$$

El marge d'error es calcula multiplicant un factor associat al nivell de confiança per l'error estàndard de la mitjana. Per a calcular correctament el marge d'error, cal distingir entre dos casos:

**a)** Si coneixem la desviació típica poblacional $\sigma$, el marge d'error es calcula com:

$$z_{\alpha/2}\sigma_{\bar{x}}$$

on $z_{\alpha/2}$ és aquell valor tal que en el qual $P(Z \geq z_{\alpha/2}) = \alpha/2$, essent $Z$ una variable $N(0, 1)$ i $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, l'error estàndard de la mitjana.

**b)** Si desconeixem la desviació típica poblacional, el marge d'error es calcula com:

$$t_{\alpha/2,\, n-1} s_{\bar{x}}$$

on $t_{\alpha/2,\, n-1}$ és el valor tal que $P(t_{n-1} \geq t_{\alpha/2,\, n-1}) = \alpha/2$, on $t_{n-1}$ és una variable Student amb $n - 1$ graus de llibertat i $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ és l'error estàndard de la mitjana, calculat a partir de la desviació típica mostral $s$.

Finalment, s'ha considerat l'efecte de la mida de la mostra, tant per a obtenir intervals de confiança amb els marges d'error desitjats com per a observar que en mostres de mesura major que trenta, i en virtut del teorema del límit central, es poden calcular els intervals de confiança, encara que la població no sigui normal.

## 5. Resum

En aquesta sessió es descriuen procediments per a obtenir intervals de confiança per a la mitjana en aquests casos:

- Quan la variable segueix una distribució normal de desviació típica coneguda.
- Quan la variable segueix una distribució normal de desviació típica desconeguda.

També es mostra com aquests procediments es poden estendre a variables que no segueixen una distribució normal, a condició que la mesura de la mostra considerada sigui suficientment gran (en concret, $n > 30$).

En tots els casos es considera l'efecte que la mida de la mostra té sobre l'interval de confiança i es mostra com s'ha de calcular la mesura de la mostra que produeix un marge d'error estipulat *a priori*.

Per a poder construir els intervals de confiança, s'han introduït els conceptes de paràmetre, estadístic i els nivells de confiança i significació.

## Exercicis

**1.** El temps (en segons) que triga a arrancar l'última versió del programa Macrohard Phrase segueix una distribució normal de desviació típica de quaranta segons. En vuitanta-un ordinadors s'ha mesurat el temps que triga a arrencar i s'ha trobat que la mitjana dels temps d'arrancada mesurats és de 158,3 segons.

**a)** Doneu un interval de confiança del 90% per a la mitjana de temps d'arrancada del programa.

**b)** Interpreteu l'interval de confiança.

**c)** El fabricant afirma que la mitjana del temps d'arrancada del programa és de 140 segons. Això és possible, segons el que hem obtingut amb l'interval de confiança?

**d)** Quina hauria de ser la mesura de la mostra per a reduir la longitud de l'interval de confiança a la meitat?

**2.** El temps (en segons) que triga a arrencar l'última versió del programa Macrohard Phrase segueix una distribució normal. En vuitanta-un ordinadors s'ha mesurat el temps que triga a arrencar i s'ha trobat que la mitjana dels temps d'arrancada mesurats és de 158,3 segons i la desviació típica de la mostra és de dotze segons.

**a)** Doneu intervals de confiança del 90% i del 95% per a la mitjana dels temps d'arrancada del programa.

**b)** Compareu els intervals de confiança obtinguts en l'apartat anterior.

**c)** El fabricant afirma que la mitjana del temps d'arrancada del programa és de 140 segons. Això és possible, segons el que hem obtingut amb l'interval de confiança?

**3.** El fabricant d'una determinada marca de iogurts afirma que els seus envasos contenen en mitjana 150 g de iogurt. Hem anat al supermercat, hem comprat deu iogurts, hem pesat el seu contingut i hem obtingut les dades següents (en grams): 148, 149, 147, 146, 149, 146, 149, 148, 149, 149.

**a)** Construïu un interval de confiança del 95% per al pes dels iogurts, suposant que el pes segueix una distribució normal de desviació típica de 3 g.

**b)** D'acord amb el resultat anterior, i ja que els pesos dels iogurts que hem comprat són tots menors que 150, podem afirmar que el fabricant no és prou sincer en el pes dels seus productes?

**4.** Repetiu l'exercici anterior, però suposant ara que el pes segueix una distribució normal de desviació típica desconeguda.

**Solucionari**

**1.**

**a)** Com que la població és normal i coneixem la desviació típica poblacional, procedirem de la forma següent:

1) Fixem el nivell de confiança $(1 - \alpha) = 0{,}9$.

2) Calculem l'error estàndard de la mitjana $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{40}{\sqrt{81}}$.

3) Calculem el valor crític; com $(1 - \alpha) = 0{,}9$, $\alpha = 0{,}1$ i $z_{\alpha/2} = z_{0{,}05} = 1{,}645$.

4) Calculem el marge d'error (també anomenat *precisió de l'estimació*) com a $z_{\alpha/2}$ per a l'error estàndard, és a dir, com:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}645 \cdot \frac{40}{\sqrt{81}} = 7{,}31$$

5) L'interval de confiança és:

$$(\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}) = (158{,}3 - 7{,}31,\ 158{,}3 + 7{,}31) = (150{,}99,\ 165{,}61)$$

**b)** La interpretació és la següent: en el 90% de les mostres de vuitanta-un ordinadors, el valor de la mitjana mostral obtinguda fa que l'interval contingui el veritable valor de la mitjana del temps que triga a arrancar el programa.

**c)** L'interval de confiança obtingut no conté el valor 140; a més, l'extrem esquerre de l'interval està molt allunyat del valor 140. Això ens diu que és poc probable que la veritable mitjana sigui de 140 segons.

**d)** Si la longitud de l'interval de confiança ha de ser la meitat, aleshores l'error estàndard també ha de ser la meitat; per tant, tenim que:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}645 \cdot \frac{40}{\sqrt{n}} = \frac{7{,}31}{2} = 3{,}655$$

resolent l'equació obtenim:

$$n = 1{,}645^2\frac{40^2}{3{,}655^2} \approx 324$$

Observem que equival a multiplicar per quatre la mesura de la mostra inicial.

**2.**

**a)** En aquest cas, no coneixem la desviació típica poblacional, per tant, aplicarem el procediment següent (calcularem simultàniament els dos intervals demanats):

**1)** Fixem els nivells de confiança; en el primer cas, $1 - \alpha = 0{,}9$ i en el segon, $1 - \alpha = 0{,}95$.

**2)** Calculem la desviació típica mostral, que és $s = 12$, tal com ens diu l'enunciat.

**3)** Calculem l'error estàndard de la mitjana $s_{\bar{x}} = \dfrac{s}{\sqrt{n}} = \dfrac{12}{\sqrt{81}} = 1{,}33$.

**4)** Calculem els valors crítics:

- Cas $1 - \alpha = 0{,}9 \Rightarrow t_{\alpha/2,\, n-1} = t_{0{,}05;80} = 1{,}6641$
- Cas $1 - \alpha = 0{,}95 \Rightarrow t_{\alpha/2,\, n-1} = t_{0{,}025;80} = 1{,}9901$

**5)** Calculem el marge d'error com a $t_{\alpha/2,\, n-1}$ per a l'error estàndard, és a dir:

- Cas $1 - \alpha = 0{,}9 \Rightarrow s_{\bar{x}} = t_{\alpha/2,\, n-1}\dfrac{s}{\sqrt{n}} = 1{,}664 \cdot 1{,}33 = 2{,}2132$
- Cas $1 - \alpha = 0{,}95 \Rightarrow s_{\bar{x}} = t_{\alpha/2,\, n-1}\dfrac{s}{\sqrt{n}} = 1{,}9901 \cdot 1{,}33 = 2{,}6468$

**6)** Finalment, els intervals de confiança demanats són els següents:

- Cas $1 - \alpha = 0{,}9 \Rightarrow (158{,}3 - 2{,}2132,\ 158{,}3 + 2{,}2132) = (156{,}09,\ 160{,}51)$
- Cas $1 - \alpha = 0{,}95 \Rightarrow (158{,}3 - 2{,}6468,\ 158{,}3 + 2{,}6468) = (155{,}65,\ 160{,}95)$

**b)** En augmentar el nivell de confiança, l'interval es fa més llarg. Això és degut al fet que com que hem de garantir que més intervals continguin la mitjana poblacional, els intervals, que sempre estan centrats en la mitjana mostral, han de ser més llargs.

**c)** No sembla possible, ja que els valors de l'interval de confiança estan molt allunyats del valor 140.

**3.**

**a)** Es tracta d'un interval de confiança per a una variable normal de la qual coneixem la desviació típica poblacional; per tant, té la forma següent:

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \left(148 \pm 1{,}96 \cdot \frac{3}{\sqrt{10}}\right) = (146{,}14,\ 149{,}8)$$

**b)** L'interval obtingut no conté el valor 150; per tant, la conclusió a què arribem és que, tot i que és possible que 150 sigui la mitjana poblacional (ja que està molt a prop dels límits de l'interval), podem dir que no ho és amb una confiança del 95%, ja que sabem que el 95% dels intervals contenen la mitjana, i aquest no la conté.

**4.**

**a)** Es tracta d'un interval de confiança per a una variable normal de la qual desconeixem la desviació típica poblacional; per tant, té la forma següent:

$$\left(\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = \left(148 \pm 2{,}2621\ \frac{1{,}247}{\sqrt{10}}\right) = (147{,}11,\ 148{,}89)$$

on $\bar{x}$ és la mitjana dels valors de la mostra i $s$ és la desviació típica dels valors de la mostra.

**b)** Tampoc ara l'interval obtingut conté el valor 150; per tant, arribem a la mateixa conclusió que en l'exercici anterior amb la diferència que ara el valor 150 es troba més lluny de l'interval.

Observem que, en no assumir la desviació típica poblacional igual a 3, el marge d'error es fa més petit, ja que la desviació típica mostral és molt menor que 3.

## Interval de confiança per a la proporció

En aquesta sessió ens dedicarem a l'estudi de l'interval de confiança per a una proporció. Començarem amb un exemple. Suposem que volem estudiar la proporció d'estudiants de la UOC que han visitat la vall de Núria algun cop (per veure si paga la pena fer-ne publicitat, per exemple). Seguirem els passos següents:

**1)** Suposem que la proporció real dels d'estudiants de la UOC que han visitat algun cop la vall de Núria és $p$.

**2)** Escollim una mostra aleatòria de, suposem, $n = 136$ estudiants.

**3)** En aquesta mostra, el 75% dels estudiants declaren haver estat a Núria.

Denotarem per $\hat{p}$ la proporció obtinguda a partir d'una mostra.

Aleshores, en el nostre cas $\hat{p} = 0{,}75$.

Quines conclusions podem treure ara sobre la relació entre la proporció real, respecte de tota la població $p$, i la proporció $\hat{p}$ obtinguda a partir d'una mostra?

Com en el cas de la mitjana, haurem de recórrer a la teoria de la probabilitat per a conèixer la distribució de la proporció, segons la qual, si la mesura de la mostra és prou gran, la variable següent segueix una distribució normal estàndard:

**Notació**

$p$ és el paràmetre poblacional i $\hat{p}$ és l'estadístic que utilitzem per a estimar-lo. I denotem per $\hat{P}$ tots els possibles valors de les proporcions en totes les mostres de la mateixa mida.

$$\frac{\hat{P}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

Ara, si fixem un nivell de confiança del 95%, per exemple, i com que estem usant una distribució normal estàndard, tenim que:

$$P\left(-1{,}96 \le \frac{\hat{P}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \le 1{,}96\right) = 0{,}95$$

Operant aquesta expressió amb l'objectiu d'aïllar el valor de $p$ tenim que:

$$P\left(\hat{P}-1{,}96\sqrt{\frac{p(1-p)}{n}} \le p \le \hat{P}+1{,}96\sqrt{\frac{p(1-p)}{n}}\right)$$

A diferència del que passava en els exemples anteriors, ara ens trobem amb un problema afegit, i és que l'error estàndard:

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

també depèn del valor de $p$, que és precisament el valor que volem estimar, i que, per tant, és desconegut.

En aquest cas, el que farem serà aproximar el valor de l'error estàndard pel valor que obtindríem a partir del valor de la proporció en la mostra que tenim, és a dir, aproximarem el valor de l'error estàndard per:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Per tant, en el cas de l'interval de confiança de la proporció, si el nivell de confiança és del 95%, treballarem amb intervals de la forma:

$$\left(\hat{p}-1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \ \hat{p}+1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

**Nivell de confiança**

La confiança aproximada serà del 95%, que ens diu que noranta-cinc de cada cent vegades que escollim una mostra aleatòria simple i fem la substitució, l'interval que obtindrem contindrà el veritable valor de *p*.

Ara, l'únic que hem de fer és substituir els valors de $\hat{p}$ i $n$ pels valors obtinguts en el nostre estudi, amb la qual cosa construïm l'interval següent:

$$\left(\hat{p}-1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \ \hat{p}+1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) =$$
$$= (0{,}75-1{,}96\cdot 0{,}03713,\ 0{,}75+1{,}96\cdot 0{,}03713) = (0{,}6772,\ 0{,}8227)$$

que és un interval de confiança amb un nivell del 95% per a la proporció dels estudiants de la UOC que han visitat alguna vegada la vall de Núria.

Per tant, la proporció d'estudiants de la UOC que, en el nostre exemple, han visitat la vall de Núria està entre el 67,72% i el 82,27%, amb una confiança del 95%.

Com en el cas de la mitjana aritmètica, descriurem detalladament el procediment de construcció dels intervals de confiança.

## 1. Procediment per a construir un interval de confiança per a la proporció

Suposem que disposem d'una mostra aleatòria simple de mida $n$ i del valor de la proporció calculat a partir de la mostra $\hat{p}$ i que la mida de la mostra és prou gran, en concret nosaltres exigirem sempre que la mida sigui superior a cent. A continuació:

**Mostra sempre superior a 100**

La mida de la mostra ha de ser superior a cent per a garantir que es pugui aplicar el teorema del límit central i per a poder obtenir simultàniament marges d'error acceptables.

**1)** Fixem el nivell de confiança, que habitualment escriurem com a $(1-\alpha)$%.

**2)** Calculem l'aproximació a l'error estàndard següent:

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

**Error estàndard de la mostra**

Escrivim $s_{\hat{p}}$ en comptes de $\sigma_{\hat{p}}$, ja que es calcula a partir del valor de la proporció en una mostra.

**3)** Calculem l'anomenat **valor crític**, que és aquell punt $z_{\alpha/2}$ pel qual $P(Z \geq z_{\alpha/2}) = \alpha/2$, on $Z$ és una variable $N(0, 1)$.

**4)** Calculem l'anomenat *marge d'error* (també anomenat *precisió de l'estimació*) com a $z_{\alpha/2}$ per a la nostra aproximació a l'error estàndard, és a dir, com a:

$$z_{\alpha/2}s_{\hat{p}} = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

**5)** L'interval de confiança obtingut amb la mostra de partida és:

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

## 2. L'efecte de la mida de la mostra

Moltes vegades, fixat el nivell de confiança, ens marcarem com a objectiu donar el valor del paràmetre $p$ amb una certa precisió. L'única manera d'obtenir la precisió desitjada serà modificant adequadament la mida de la mostra. Suposem que desitgem una precisió o marge d'error $ME$; si igualem el marge d'error amb l'error estàndard, tenim que:

$$ME = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

aïllant $n$ obtenim:

$$n \geq (z_{\alpha/2})^2\frac{p(1-p)}{ME^2}$$

Ara ens tornem a trobar amb el problema d'abans, ja que, abans de fer l'estudi, no sabem quin és el valor de $p$.

En aquest cas, s'acostuma a escollir un valor de $\hat{p}$ que ens sembli apropiat (com més a prop del veritable i desconegut valor de $p$, millor) d'acord amb anteriors estudis o a algun fet destacat.

En resum, determinarem la mida de la mostra així:

$$\text{Mida de la mostra} = n = (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{ME^2}$$

En cas que no disposem de cap informació sobre el possible valor de $\hat{p}$, ens haurem d'assegurar que la mesura de la mostra és prou gran per a assolir el nostre objectiu. Com que $z_{\alpha/2}$ i el marge d'error estan fixats, utilitzarem el fet que el producte $\hat{p}(1-\hat{p})$ és sempre menor o igual que 1/4 i que precisament $\hat{p}(1-\hat{p}) = 1/4$, en el cas que $\hat{p} = 0{,}5$. Així, doncs, suposant que $\hat{p} = 0{,}5$, obtenim la mida mostral màxima per a cada marge d'error. Aquest és el procediment que segueixen molts sondejos i estudis electorals, que en definitiva es basa en el fet que:

**Això és degut al fet que $x(1-x)$ assoleix el seu valor màxim quan $x = 0{,}5$.**

$$\text{Mida de la mostra} = n = (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{ME^2} \leq \frac{1}{4}\frac{(z_{\alpha/2})^2}{ME^2}$$

Per tant, una mostra de mida $n = \frac{1}{4}\left(\frac{z_{\alpha/2}}{ME}\right)^2$ ens garanteix sempre el marge d'error desitjat.

**Taula de grandàries**

Fent servir la fórmula anterior, obtenim la taula de grandàries mostrals següent segons el marge d'error (per a un nivell de confiança del 95%):

| Marge d'error | Mida màxima |
|---|---|
| 5% | 384 |
| 2% | 2.401 |
| 1% | 9.640 |
| 0,5% | 38.416 |
| 0,02% | ≈24.010.000 |

**Proporció de catalans que aniran a votar**

De cent catalans escollits a l'atzar, seixanta-cinc ens diuen que aniran a votar a les properes eleccions. Amb aquestes dades podem determinar un interval de confiança amb un nivell de confiança del 95% per la proporció de catalans que aniran a votar d'aquesta manera:

$$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = \left(\frac{65}{100} \pm 1{,}96\sqrt{\frac{\frac{65}{100}\left(1-\frac{65}{100}\right)}{100}}\right) = (0{,}5565,\ 0{,}7434)$$

Per tant, podem esperar que el percentatge de votants estarà entre el 55% i el 74%, amb un nivell de confiança del 95%.

Realment obtenim un interval de confiança molt ample, però és que només hem preguntat a cent individus.

**Mida de la mostra**

Si sabem que a les anteriors eleccions van anar a votar un 70% dels catalans, podem utilitzar aquesta informació per a determinar quina ha de ser la mida d'una mostra que permeti de construir un interval de confiança com el de l'exemple anterior, però amb marge d'error menor que el 5%. En concret, haurem d'aplicar la fórmula següent per a obtenir la mida de la mostra:

$$n = (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{ME^2} = 1{,}96^2 \frac{0{,}7 \cdot 0{,}3}{0{,}05^2} = \frac{0{,}806736}{0{,}05^2} = 322{,}69$$

És a dir, necessitaríem dades sobre la intenció de participar de 323 catalans.

## 3. Resum

En aquesta sessió hem construït l'interval de confiança per a la proporció i hem estudiat la relació entre el marge d'error i la mida de la mostra.

# Exercicis

**1.**
Hem agafat una mostra aleatòria de 400 estudiants de la UOC i hem observat que 180 d'ells són fumadors.

**a)** Doneu un interval de confiança del 95% per a la proporció d'estudiants de la UOC que fumen. Quina és la longitud de l'interval de confiança?

**b)** Quants estudiants hauríem de seleccionar per a reduir la longitud de l'interval de confiança del 95% a la meitat? I si volem que el marge d'error sigui inferior a l'1%?

**c)** El Departament de Sanitat de la Generalitat de Catalunya afirma que el 40% de la població catalana és fumadora. Quines conseqüències podem treure dels estudiants de la UOC, comparant aquesta dada amb l'interval de confiança?

**2.**
En l'edició de Catalunya del dia 28-4-2001, el diari *El Periódico* publicava la fitxa tècnica següent d'un sondeig efectuat amb motiu de les eleccions autonòmiques al País Basc del dia 13-5-2001:

> "Error de la mostra: l'error en la mostra associat a un nivell de confiança del 95,5% (sigma = = 2) i p = q = 50% és del +/–2,57". Mida de la mostra = 1.514.

Comproveu que l'error (marge d'error) està calculat correctament.

## Solucionari

**1.**
**a)** L'interval de confiança té la forma següent:

$$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = \left(\frac{180}{400} \pm 1{,}96\sqrt{\frac{\frac{180}{400}\left(1-\frac{180}{400}\right)}{400}}\right) = (0{,}4012,\ 0{,}4988)$$

i la seva longitud és (0,4012, 0,4988).

**b)** Si volem reduir la longitud de l'interval a la meitat, haurem de multiplicar la mesura de la mostra per quatre. Això es veu fàcilment substituint a la fórmula de la mida mostral $ME$ per $ME/2$, amb la qual cosa tenim que:

Mida de la mostra = $n = (z_{\alpha/2})^2 \dfrac{\hat{p}(1-\hat{p})}{(ME/2)^2} = 4(z_{\alpha/2})^2 \dfrac{\hat{p}(1-\hat{p})}{ME^2}$

I, per tant, la nova mida ha de ser quatre vegades l'antiga.

Si volem que el marge d'error sigui inferior a 0,01 s'haurà de complir:

$$n \geq (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{(0{,}01)^2} = 1{,}96^2 \frac{\frac{180}{400}\left(1-\frac{180}{400}\right)}{1{,}01^2} = 9.507{,}96$$

**c)** Com que el 95% dels intervals de confiança contenen la veritable proporció poblacional i com que aquest interval no conté el 40%, hem de pensar que la proporció de fumadors a la UOC supera la proporció de fumadors dins la població catalana.

**2.**

En aquest cas tenim $p = q = 0{,}5$, és a dir, es calcula la mida màxima de la mostra per al nivell de confiança donat. El nivell de confiança és 95,5%, habitualment usat en els sondejos, ja que en aquest cas el valor $z_{\alpha/2} = 2$ (o "2 sigma", com diu la fitxa tècnica, ja que representa dues desviacions típiques de la mitjana). Aplicant la fórmula del marge d'error obtenim que:

$$ME = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 2\sqrt{\frac{1/4}{1.514}} = 0{,}025700$$

per tant, el marge d'error és 2,57% (en els sondejos s'utilitza l'expressió +/–2,57, que indica una possible variació cap amunt o avall del 2,57%).