UNIVERSITAT DE BARCELONA

MOSTREIG ESTADÍSTIC: PROBLEMES I PRÀCTIQUES

Manuela Alcañiz Zanón (coord.)

Mònica Bécue

Catalina Bolancé Losilla

Montserrat Guillén Estany

Dídac Planas Paz

Departament d'Econometria, Estadística i Economia Espanyola

Mostreig Estadístic: problemes i pràctiques

Agraïm l'ajut rebut de la Generalitat de Catalunya dins la convocatòria d'*Ajuts per al finançament de projectes per a la millora de la qualitat docent a les universitats catalanes* per a l'any 2009, MQD-00166.

Dipòsit legal:

ISBN:

Impressió: Gráficas Rey, S.L.

ÍNDEX

Presentació **5**

1. Mostreig probabilístic i estimadors **7**

1.1. Exercici resolt 8
1.2. Exercicis proposats 11
1.3. Solucions 15

2. Mostreig aleatori simple **19**

2.1. Exercici resolt 20
2.2. Exercicis proposats 23
2.3. Solucions 30

3. Mostreig aleatori estratificat **35**

3.1. Exercici resolt 36
3.2. Exercicis proposats 38
3.3. Solucions 48

4. Mostreig sistemàtic **53**

4.1. Exercici resolt 54
4.2. Exercicis proposats 57
4.3. Solucions 62

5. Mostreig per conglomerats **65**

5.1. Exercici resolt 66
5.2. Exercicis proposats 68
5.3. Solucions 72

Annex A. Exercicis en SPSS i R **75**

A.1. Mostreig en SPSS i R 76
A.2. Exercicis proposats 78

Annex B. Formularis **85**

B.1. Formulari de mostreig probabilístic i estimadors 86
B.2. Formulari de mostreig aleatori simple 87
B.3. Formulari de mostreig aleatori estratificat 89
B.4. Formulari de mostreig sistemàtic 91
B.5. Formulari de mostreig per conglomerats 92

Annex C. Taules estadístiques 95

C.1. Distribució Normal estàndard 96

C.2. Distribució T-Student 97

Bibliografia 99

PRESENTACIÓ

La present publicació desenvolupa el contingut aplicat de l'assignatura Mostreig Estadístic, que s'ofereix al Grau d'Estadística UB-UPC. Malgrat no se segueix estrictament el programa de la mateixa, les diferents seccions del llibre són fàcilment reconeixibles. S'inicia amb un bloc introductori referit al mostreig probabilístic i als estimadors i segueixen les seccions corresponents al mostreig aleatori simple, mostreig aleatori estratificat, mostreig sistemàtic i, en darrer lloc, mostreig per conglomerats.

La voluntat de la publicació és la d'esdevenir un recull exhaustiu d'exercicis relacionats amb els diferents mètodes de mostreig estadístic que es presenten a l'assignatura. És per aquest motiu que la part teòrica de l'assignatura queda exclosa de la publicació, i només s'inclouen uns formularis a l'*Annex B*, per a que l'alumne disposi de les fórmules necessàries per a la realització dels exercicis, sense haver de recórrer a d'altres materials.

Cadascun dels blocs corresponents als diferents mètodes de mostreig s'organitza de la mateixa manera: es presenta un exercici resolt amb tot el desenvolupament necessari; i després, s'inclou un llistat d'exercicis proposats i les solucions als mateixos.

El bloc d'annexos està dividit en tres seccions. A l'*Annex A*, trobareu una explicació detallada dels procediments necessaris per a dur a terme diferents tipus mostreig amb els paquets estadístics SPSS[1] i R. A més, es proposen quatre exercicis[2] a resoldre usant els programes esmentats. A continuació, s'inclouen els formularis per a poder solucionar els exercicis proposats. I en darrer terme, l'*Annex C* inclou les taules estadístiques de la distribució Normal estàndard i de la distribució t d'Student, les dues distribucions rellevants per a la realització dels exercicis.

Els objectius que es persegueixen a través de l'assignatura Mostreig Estadístic i, per extensió, amb la present publicació són, en referència als coneixements:

- Assimilar les diferències entre els conceptes de població, mostra, paràmetre d'interès, estimació i error de mostreig.
- Conèixer els diferents mètodes de mostreig estadístic i les seves particularitats.
- Distingir quin mètode és el més adequat en cada cas.
- Calcular els errors mostrals, en funció del mètode triat, la grandària de la mostra i la grandària de la població.
- Calcular la mida mostral necessària per aconseguir una precisió donada, en funció del mètode de mostreig triat.

[1] Les darreres versions del programa (a partir de la versió 17.0.3, de l'any 2009) es comercialitzen amb el nom de PASW Statistics. Degut a l'ampli coneixement de l'anterior denominació, serà el nom d'SPSS el que s'usarà en la publicació.

[2] Les bases de dades corresponents a aquests exercicis es troben disponibles al Campus Virtual de l'assignatura o demanant-les per correu electrònic a la professora Manuela Alcañiz (malcaniz@ub.edu).

Pel que fa a competències i habilitats que l'estudiant adquireix, la present publicació busca:

- Proposar mètodes de selecció aleatòria d'unitats poblacionals, per a formar una mostra estadística.
- Identificar dissenys de selecció de les mostres en casos reals concrets.
- Conèixer els diferents tipus de mostreig estadístic, especialment el mostreig aleatori simple, el mostreig sistemàtic, el mostreig estratificat i el mostreig per conglomerats.
- Distingir en quina situació és més adequat aplicar un mètode de mostreig o altre, i les implicacions de cadascun d'ells.
- Calcular els errors mostrals comesos, en funció del mètode triat, la grandària de la mostra i la grandària de la població.
- Determinar la grandària mostral necessària per garantir un cert nivell de precisió, en funció del mètode de mostreig triat, per a estimacions de proporció mitjana, total i total de classe.
- Classificar els diferents errors de mostreig i saber calcular-los.
- Simular el comportament de mostres aleatòries i comparar-lo enfront del comportament d'una població coneguda.
- Desenvolupar la capacitat d'anàlisi i de síntesi, així com el sentit crític per avaluar resultats i mètodes.
- Ser capaç d'enfrontar-se amb èxit a problemes de diversa dificultat.
- Desenvolupar recursos propis per a la resolució de problemes.
- Treballar la capacitat per utilitzar el raonament lògic i els instruments matemàtics.

Barcelona, abril de 2011

Tema 1. MOSTREIG PROBABILÍSTIC I ESTIMADORS

MOSTREIG PROBABILÍSTIC I ESTIMADORS

1.1 EXERCICI RESOLT

El propietari d'un 4x4 fa viatges privats amb turistes que visiten el Pirineu, ensenyant-los diversos paratges. En funció de la durada i dificultat del trajecte, el preu de l'excursió és més o menys elevat.
En una setmana, el conductor ha fet un total de 5 viatges, i ha percebut les següents quantitats, en euros:

$$120, 200, 140, 100, 180$$

Dels 5 grups que van fer l'excursió, 4 van quedar satisfets.

Tot i que siguin pocs valors, són les dades poblacionals, ja que el conductor no va fer més viatges.

a) Quina és la despesa mitjana poblacional? I la despesa total poblacional? Què significa?

$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{740}{5} = 148$ € de despesa mitjana per viatge.

T= $\sum_{i=1}^{n} x_i = 740$ € de despesa total poblacional.

b) Quina és la proporció poblacional de grups satisfets?

$p = \frac{1}{n}\sum_{i=1}^{n} A_i = \frac{4}{5} = 80\%$ de turistes satisfets, on $A_i = \begin{cases} 1, \text{ pels turistes satisfets} \\ 0, \text{ pels insatisfets} \end{cases}$

Suposem ara que traiem mostres de mida $n = 3$ de les quantitats que ha cobrat el conductor. Noteu que en un cas real, només trauríem una mostra.

c) Quantes mostres diferents de mida 3 es poden obtenir (sense reposició i sense que importi l'ordre)?

$C_{N,n} = C_{5,3} = \binom{5}{3} = 10$ mostres diferents.

d) Obteniu l'espai mostral format per aquestes mostres de mida 3.

$S = \{(120,200,140), (120,200,100), (120,200,180), (120,140,100), (120,140,180), (120,100,180), (200,140,100), (200,140,180), (200,100,180), (140,100,180)\}$

e) Si tots els valors tenen la mateixa probabilitat de formar part de la mostra, quina és la probabilitat de cadascuna de les mostres possibles?

$$P(S_i) = {}^{1}\!/_{10}, \forall i$$

Considerem ara la mitjana mostral del preu pagat, com a estimador de la mitjana poblacional.

f) Escriviu la distribució de probabilitat de la mitjana mostral ($\bar{x}$). Recordeu que, per a fer-ho, heu de trobar tots els possibles valors que pren la mitjana mostral (per a cada mostra, heu de calcular quant val la mitjana mostral), i veure amb quina probabilitat els pren.

S_i	$\bar{x}_i$	$P(\bar{x} = \bar{x}_i)$
120, 200, 140	153,333	1/10
120, 200, 100	140	1/10
120, 200, 180	166,667	1/10
120, 140, 100	120	1/10
120, 140, 180	146,667	1/10
120, 100, 180	133,333	1/10
200, 140, 100	146,667	1/10
200, 140, 180	173,333	1/10
200, 100, 180	160	1/10
140, 100, 180	140	1/10

Distribució de la mitjana mostral:

$$\bar{x} = \begin{cases} 120 \\ 133{,}333 \\ 140 \\ 146{,}667 \\ 153{,}333 \\ 160 \\ 166{,}667 \\ 173{,}333 \end{cases}, \quad \begin{matrix} P(\bar{x} = \bar{x}_i) = 1/10 \\ P(\bar{x} = \bar{x}_i) = 1/10 \\ P(\bar{x} = \bar{x}_i) = 1/5 \\ P(\bar{x} = \bar{x}_i) = 1/5 \\ P(\bar{x} = \bar{x}_i) = 1/10 \\ P(\bar{x} = \bar{x}_i) = 1/10 \\ P(\bar{x} = \bar{x}_i) = 1/10 \\ P(\bar{x} = \bar{x}_i) = 1/10 \end{matrix}$$

g) Calculeu l'esperança o valor esperat de la mitjana mostral i comenteu si és un estimador no esbiaixat de la mitjana poblacional que heu trobat abans.

$$E(\bar{x}) = \sum_i \bar{x}_i \, P(\bar{x} = \bar{x}_i) = \frac{120}{10} + \frac{133{,}333}{10} + \frac{140}{5} + \frac{146{,}667}{5} + \frac{153{,}333}{10} + \\ + \frac{160}{10} + \frac{166{,}667}{10} + \frac{173{,}333}{10} = 148$$

$B(\bar{x}) = E(\bar{x}) - \mu = 0 \;\rightarrow$ Estimador no esbiaixat

Considereu ara el valor mínim mostral. Per exemple, per la mostra $\{120, 200, 140\}$ seria 120.

h) Escriviu la distribució de probabilitat del mínim mostral.

S_i	$Min(x_i)$	$P[Min(x) = Min(x_i)]$
120, 200, 140	120	1/10
120, 200, 100	120	1/10
120, 200, 180	120	1/10
120, 140, 100	100	1/10
120, 140, 180	120	1/10
120, 100, 180	100	1/10
200, 140, 100	100	1/10
200, 140, 180	140	1/10
200, 100, 180	100	1/10
140, 100, 180	100	1/10

Distribució del mínim mostral:

$$Min(x) = \begin{cases} 100 & P[Min(x) = Min(x_i)] = 3/5 \\ 120\,, & P[Min(x) = Min(x_i)] = 3/10 \\ 140 & P[Min(x) = Min(x_i)] = 1/10 \end{cases}$$

i) Calculeu el valor esperat del mínim mostral.

$$E[Min(x)] = 100 * \frac{3}{5} + 120 * \frac{3}{10} + 140 * \frac{1}{10} = 110$$

j) És el mínim mostral un bon estimador de la mitjana poblacional? Per què? Quina propietat li manca?

El mínim mostral no és un bon estimador de la mitjana poblacional perquè te biaix (negatiu); és a dir, condueix a infravalorar de forma sistemàtica la mitjana:

$$B[Min(x)] = E[Min(x)] - \mu = 110 - 148 = -38$$

A més, s'observa com l'Error Quadràtic Mitjà per la mitjana mostral és inferior al del mínim mostral, quan volem estimar la mitjana poblacional:

$$Var(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2 = \frac{120^2 + 133{,}333^2 + (140^2 * 2) + \cdots}{10} - 148^2 =$$

$$= 229{,}333$$

$$EQM(\bar{x}) = Var(\bar{x}) + [B(\bar{x})]^2 = 229{,}333$$

$$Var[Min(x)] = E\{[Min(x)]^2\} - \{E[Min(x)]\}^2 =$$
$$= \frac{(100^2 * 6) + (120^2 * 3) + 140^2}{10} - 110^2 = 180$$
$$EQM[Min(x)] = Var[Min(x)] + \{B[Min(x)]\}^2 = 180 + (-38)^2 = 1624$$

1.2 EXERCICIS PROPOSATS

1. Sigui una població de mida $N = 5$. Es mesura la variable X, els valors de la qual es mostren a la següent taula:

i	1	2	3	4	5
X_i	25	20	28	23	19

a) Obteniu l'espai mostral format per mostres de mida $n = 3$, escollides sense reposició i sense importar l'ordre.

b) Calculeu la probabilitat d'obtenir cadascuna de les mostres de l'apartat anterior.

c) Obteniu la distribució de la mitjana mostral (arrodoniu a zero decimals).

d) Obteniu la distribució del valor mínim mostral.

e) Calculeu l'esperança, la variància i l'error quadràtic mitjà de la mitjana mostral i del mínim.

2. Sigui una població de mida $N = 8$ de la que s'extreu una mostra de mida $n = 3$. Els valors de la variable mesurada en la població són:

i	1	2	3	4	5	6	7	8
X_i	92	132	84	103	79	163	99	128

Per tal d'estimar la mitjana poblacional, μ, es proposen dos plans de mostreig:

PLA 1:

Nº de mostra	*Mostra* S_j	$P(S_j)$
1	1,2,5	1/8
2	1,3,8	1/8
3	1,5,6	1/8
4	2,5,7	1/8
5	2,6,8	1/8
6	3,5,8	1/8
7	3,6,7	1/8
8	4,6,7	1/8

PLA 2:

Nº de mostra	*Mostra* S_j	$P(S_j)$
1	2,4,6	1/4
2	3,6,7	1/2
3	5,7,8	1/4

a) Calculeu el valor de la mitjana poblacional μ.

b) Per a cadascun dels plans de mostreig, obteniu la distribució de la mitjana mostral, i calculeu la seva esperança, variància, biaix i error quadràtic mitjà.

c) Argumenteu quin pla de mostreig es consideraria millor.

3. En una població de 15 elements $(x_i, i = 1,\ldots,15)$, es volen extreure mostres de mida $n = 3$. Per a fer-ho, s'escull un nombre aleatori k entre 1 i 5, de manera que la mostra obtinguda esdevé $\{x_k, x_{k+5}, x_{k+10}\}$, essent $x_i = i$.
Es pretén estimar el màxim poblacional $[\theta = Max(x_i)]$ i es consideren dos estimadors diferents: $\hat{\theta}_1 = Max(x_i), i = 1,\ldots,n$, o $\hat{\theta}_2 = 2\bar{x} - 1$. Avalueu-ne la precisió i escolliu el més adequat per a l'estimació.

4. En una distribució binomial $B(2,p)$ es prenen mostres aleatòries simples de mida $n = 2$ i es consideren els dos estimadors de p següents:

$$\hat{p}_1 = \frac{\bar{x}}{2} \quad \text{i} \quad \hat{p}_2 = \frac{\bar{x}}{3}$$

a) Trobeu les distribucions en el mostreig dels dos estimadors, i feu-les servir per a veure si es tracta d'estimadors no esbiaixats.

Indicació: $X \sim B(2,p) \rightarrow P(X = k) = \binom{2}{k} * p^k * q^{2-k}$

k	$P(X = k)$
0	
1	
2	

b) Se us acut alguna manera més directa d'estudiar si els estimadors tenen biaix?

5. Sigui X una variable aleatòria distribuïda uniformement a l'interval $[0, b]$. A partir d'una mostra $X_1, X_2, \ldots, X_n$ d'observacions independents de la variable X, es proposen dos estimadors pel paràmetre b:

$$\hat{\theta}_1 = 2\bar{x} \quad \text{i} \quad \hat{\theta}_2 = x_i + x_j,\ i \neq j$$

a) Estudieu si tenen biaix.

b) Trobeu l'error quadràtic mitjà de cada estimador.

c) Quin és preferible?

d) Considereu ara l'estimador $\hat{\theta}_3 = Max(X_i),\ i = 1,\ldots,n$. Argumenteu si té biaix. En té assimptòticament?

6. La durada d'unes bombetes és una variable aleatòria amb distribució normal amb variància $\sigma^2 = 400$. Es pren una mostra de 30 bombetes i s'obté una durada mitjana de 1.510 hores. Calculeu:

a) L'interval de confiança al 95% per a la mitjana poblacional.

b) El mateix, però al 99%. Observeu la diferència amb el resultat anterior.

7. En un estudi, l'objectiu del qual és determinar el temps mitjà de fabricació d'un producte, s'agafen 20 operaris i s'obté un temps mitjà de fabricació de $42{,}5$ minuts i una desviació estàndard de $3{,}7$ minuts. Suposant que el temps de fabricació segueix una llei normal, calculeu:

a) Un interval de confiança al 99% per a la mitjana poblacional μ.

b) El mateix que a l'apartat anterior, però considerant una mostra de mida 100, per la qual s'obtenen els mateixos resultats.

8. En una població normal on $\sigma^2 = 25$, es desitja extraure una mostra que permeti assegurar, per a una probabilitat del 95%, que la mitjana mostral diferirà com a màxim 2 unitats de la mitjana veritable. Quina ha de ser la mida de la mostra?

9. D'una població amb distribució desconeguda i amb variància $\sigma^2 = 12$, s'extreu una mostra aleatòria simple de mida $n = 36$, essent la suma dels seus elements:

$$\sum_{i=1}^{n} x_i = 180$$

a) Calculeu l'interval de confiança per a la mitjana poblacional, al 90%, suposant que no es pot assumir normalitat.

b) Repetiu el càlcul suposant que sí que es pot suposar normalitat i compareu el resultat.

10. Un procés està caracteritzat per una variable aleatòria discreta, infinita i numerable. Es selecciona una mostra de 5 unitats i s'obté:

$$\sum_{i=1}^{n} x_i = 60 \text{ i } \sum_{i=1}^{n} x_i^2 = 768$$

a) Estimeu l'esperança i la variància (corregida) de la població.

b) Obteniu un interval de confiança al 95% per a μ:
   i. Suposant normalitat.
   ii. Sense suposar normalitat.

c) Igual que l'apartat *b)*, però suposant $n = 100$ (i que s'obtenen els mateixos resultats que per $n = 5$).

11. Una empresa de màrqueting vol conèixer el rendiment que s'obté en utilitzar un fitxer de dades de clients, que conté més d'un milió de registres (a efectes de càlcul, es pot considerar que la població és infinita).

Sigui p la proporció d'individus que accepten una oferta que es va realitzar a tots els clients del fitxer, es pren una mostra de 10.000 direccions i entre elles accepten l'oferta 230 clients. Calculeu un interval de confiança al 95% per al paràmetre p.

12. Per tal de conèixer l'impacte de la publicitat destinada a la promoció d'un nou producte, l'endemà del llançament es realitza una enquesta a tot el país, escollint una mostra aleatòria simple de 1.000 individus majors de 18 anys. D'aquests, 132 han vist la publicitat.

a) Realitzeu una estimació puntual i per interval (al 95%) de la proporció d'individus majors de 18 anys que han vist la publicitat.

b) El mateix per al total d'individus majors de 18 anys que han vist la publicitat, sabent que $N = 12.500.000$.

13. S'extreu una mostra aleatòria de 350 famílies amb fills amb edats entre 2 i 3 anys i s'observa la següent distribució de freqüències associada a l'edat en que els nens comencen a parlar:

Edat (en mesos)	9	10	11	12	13	14	15	16	17	18	19	20
Nº de nens	27	51	64	78	55	41	19	7	4	2	1	1

a) Obteniu l'histograma associat a la variable edat (en mesos) i considereu si és adequat el supòsit de distribució normal.

b) Obteniu una estimació puntual i per interval (amb nivell de confiança del 95%) per a la mitjana poblacional.

c) Obteniu un interval de confiança al 99% per a la mitjana poblacional i compareu-lo amb l'obtingut a l'apartat anterior.

14. Una universitat té 436 professors l'any 2009. La llista del personal acadèmic i el nombre de congressos internacionals a què han assistit els darrers dos anys es troba disponible. S'extreu una mostra aleatòria de 70 professors, que es reprodueix a la taula següent:

Nº de congressos internacionals	0	1	2	3	4	5	6	7	8	9	10
Nº de professors	41	6	3	4	6	3	2	0	2	2	1

a) Obteniu l'histograma associat a la variable nombre de congressos a què han assistit els professors i observeu si és adequat considerar que la distribució de la variable és normal.

b) Per a la mitjana de la variable anterior, obteniu una estimació puntual i un interval de confiança al 95%.

c) Estimeu la proporció de professors que no han assistit a cap congrés internacional els darrers dos anys i un interval de confiança al 95% per la mateixa.

15. En dues consultes mèdiques de tractament de la hipertensió, es va realitzar una enquesta de salut a una mostra aleatòria dels pacients citats al llarg d'un mes. En base a les dades d'aquesta enquesta, s'estima la proporció de pacients que han complert correctament el tractament prescrit durant el període d'observació.

A la primera consulta (A), el nombre de pacients citats per al proper mes és de 200, i se selecciona una mostra de 50 pacients (fracció de mostreig del 25%). A la segona consulta (B), se cita a 2000 pacients amb tractament prescrit i es constitueix una mostra de 200 pacients (fracció de mostreig del 10%).

Si no hi ha raons per a pensar que la proporció que es vol estimar és diferent en ambdues consultes, a quina consulta s'estimarà la proporció amb major precisió?

16. Es desitja estimar la mitjana d'una variable quantitativa en dues poblacions A i B de mida $N_A = 3N_B$. Per a realitzar l'estimació, se selecciona una mostra d'igual mida n en ambdues i s'observa que l'error de mostreig és igual per a les dues poblacions. Quina és la relació entre les variàncies mostrals de la variable?

1.3 SOLUCIONS

1. a) $S = \{(25,20,28), (25,20,23), (25,20,19), (25,28,23), (25,28,19), (25,23,19),$
$(20,28,23), (20,28,19), (20,23,19), (28,23,19)\}$

b) $P(S_i) = \frac{1}{10}, \forall i$

c) $\bar{x} = \begin{cases} 21, \ p = 1/5 \\ 22, \ p = 1/5 \\ 23, \ p = 1/5 \\ 24, \ p = 3/10 \\ 25, \ p = 1/10 \end{cases}$

d) $Min(x_i) = \begin{cases} 19, \ p = {}^3/_5 \\ 20, \ p = {}^3/_{10} \\ 23, \ p = {}^1/_{10} \end{cases}$

e)

	Mitjana	*Mínim*
Esperança	22,9	19,7
Variància	1,69	1,41
EQM	1,7	1,9

2. a) $\mu = 110$

b) PLA 1:

$$\bar{x} = \begin{cases} 97; \ \ p = {}^1/_8 \\ 101; \ \ p = {}^1/_8 \\ 101{,}333; \ \ p = {}^1/_8 \\ 103{,}333; \ \ p = {}^1/_8 \\ 111{,}333; \ \ p = {}^1/_8 \\ 115{,}333; \ \ p = {}^1/_8 \\ 121{,}667; \ \ p = {}^1/_8 \\ 141; \ \ p = {}^1/_8 \end{cases}$$

	Mitjana
Esperança	111,5
Biaix	1,5
Variància	184,863
EQM	187,113

PLA 2:

$$\bar{x} = \begin{cases} 132{,}667; \ p = {}^1/_4 \\ 115{,}333; \ p = {}^1/_2 \\ 102; \ p = {}^1/_4 \end{cases}$$

	Mitjana
Esperança	116,333
Biaix	6,333
Variància	118,559
EQM	158,666

3.

	θ_1	θ_2
Esperança	13	15
Biaix	-2	0
Variància	2	8
EQM	6	8

4. a)

$$p_1 = \begin{cases} 0, & P(p_1 = 0) = q^4 \\ {}^1/_4, & P(p_1 = {}^1/_4) = 4pq^3 \\ {}^1/_2, & P(p_1 = {}^1/_2) = 6p^2q^2 \\ {}^3/_4, & P(p_1 = {}^3/_4) = 4p^3q \\ 1, & P(p_1 = 1) = p^4 \end{cases} \qquad p_2 = \begin{cases} 0, & P(p_1 = 0) = q^4 \\ {}^1/_6, & P(p_1 = {}^1/_6) = 4pq^3 \\ {}^1/_3, & P(p_1 = {}^1/_3) = 6p^2q^2 \\ {}^1/_2, & P(p_1 = {}^1/_2) = 4p^3q \\ {}^2/_3, & P(p_1 = {}^2/_3) = p^4 \end{cases}$$

$$E(p_1) = p \to B(p_1) = 0$$

$$E(p_2) = \frac{2}{3}p \to B(p_2) = \frac{1}{3}p$$

b) $X \sim B(n,p) \ \to E(X) = np \ \to \begin{cases} E(p_1) = p \\ E(p_2) = \frac{2}{3}p \end{cases}$

5. a) i b)

	θ_1	θ_2
Esperança	b	b
Biaix	0	0
Variància	${b^2}/_{3n}$	${b^2}/_6$
EQM	${b^2}/_{3n}$	${b^2}/_6$

d) $E(\theta_3) = E(x_n) < b;\ B(\theta_3) < 0$
$n \to N:\ E(\theta_3) \to b;\ B(\theta_3) \to 0$

6. a) $\mu \in (1.502{,}843;\ 1.517{,}157)$

b) $\mu \in (1.500{,}579;\ 1.519{,}421)$

7. a) $\mu \in (40{,}133;\ 44{,}867)$

b) $\mu \in (41{,}528;\ 43{,}472)$

8. $n \geq 25$

9. a) $\mu \in (3{,}174;\ 6{,}826)$

b) $\mu \in (4{,}053;\ 5{,}947)$

10. a) $E(X) = 12;\ Var(X) = 12$

b) $\mu \in (7{,}699;\ 16{,}301);\ \mu \in (5{,}072;\ 18{,}928)$

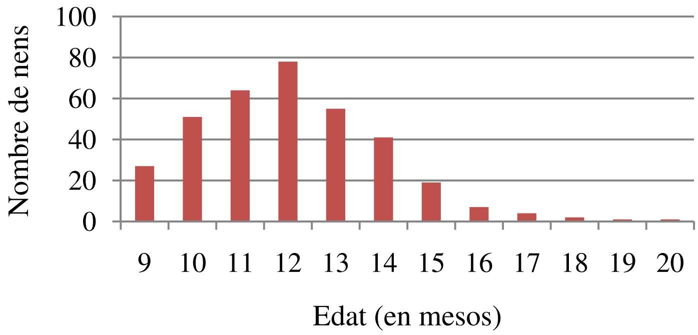
c) $\mu \in (11{,}313;\ 12{,}687)$; $\mu \in (10{,}451;\ 13{,}549)$

11. $p \in (0{,}020;\ 0{,}026)$

12. a) $\hat{p} = 0{,}132$; $p \in (0{,}111;\ 0{,}153)$

b) $\hat{A} = 1.650.000$; $A \in (1.387.752;\ 1.912.248)$

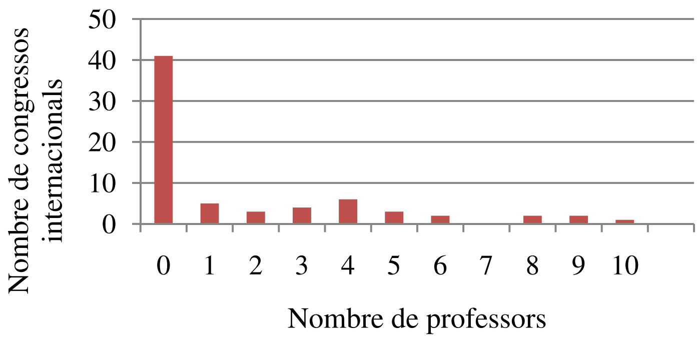
13. a)



b) $\hat{\mu} = 12{,}063$; $\mu \in (11{,}8609;\ 12{,}265)$

c) $\mu \in (11{,}797;\ 12{,}329)$

14. a)



b) $\hat{\mu} = 1{,}7$; $\mu \in (0{,}283;\ 3{,}117)$

c) $\hat{p} = 58{,}57\%$; $p \in (0{,}322;\ 0{,}849)$

15. Consulta B.

16. $S_A^2 = \frac{3N_B - 3n}{3N_B - n} S_B^2$.

Tema 2. MOSTREIG ALEATORI SIMPLE

MOSTREIG ALEATORI SIMPLE

2.1 EXERCICI RESOLT

El Ministeri d'Hisenda desitja estimar la proporció de persones que estan a favor d'una reforma en l'impost sobre la renda. La població investigada està formada per 10.000 persones i s'extreu una mostra aleatòria simple (sense reposició) de 200 persones, resultant que 50 estan a favor de la reforma.

a) Quina és la fracció de mostreig?

$$f = \frac{n}{N} = \frac{200}{10000} = 2\%$$

b) A quantes persones de la població representa cada persona de la mostra? Aquest valor s'anomena factor d'elevació.

$factor\ d'elevació = \frac{N}{n} = \frac{10000}{200} = 50$ → Cada persona de la mostra representa a 50 individus de la població.

c) Com es relaciona el factor d'elevació amb la fracció de mostreig?

$$factor\ d'elevació = \frac{1}{f} = \frac{N}{n}$$

d) Estimeu la proporció i el total poblacional de persones que estan a favor de la reforma.

$$\hat{p} = \frac{\sum_{i=1}^{n} A_i}{n} = \frac{50}{200} = 0{,}25 = 25\%$$

$$\text{on } A_i = \begin{cases} 1, \text{ si la persona està a favor de la reforma} \\ 0, \text{ en cas contrari} \end{cases}$$

$\hat{A} = N * \hat{p} = 10000 * 0{,}25 = 2500$ individus a favor de la reforma.

e) Per a la proporció, quin serà l'error absolut de mostreig? I el relatiu?

Error absolut de mostreig: $e(\hat{p}) = \hat{\sigma}(\hat{p}) = \sqrt{\widehat{Var}(\hat{p})} = \sqrt{\frac{1}{n-1}(1-f)\hat{p}(1-\hat{p})} = 0{,}0304$

Error relatiu de mostreig: $e_r(\hat{p}) = \frac{e(\hat{p})}{\hat{p}} = \frac{\hat{\sigma}(\hat{p})}{\hat{p}} = \frac{0{,}0304}{0{,}25} = 0{,}1216$

f) Calculeu l'interval de confiança al 95% per a la proporció poblacional i interpreteu el resultat.

$p \in \left[\hat{p} \pm z_{\alpha/2}\, \hat{\sigma}(\hat{p})\right] = (0{,}1904;\ \ 0{,}3096) \rightarrow$ La proporció de població favorable a la reforma de l'impost es troba entre el 19,04% i el 30,96%, amb una confiança del 95%.

g) Suposant que l'interval de l'apartat anterior es considera poc precís, calculeu la mida mostral necessària per tal que l'estimació de la proporció tingui un error màxim de $\pm 0{,}03$ amb un probabilitat del 95%.

n tal que $e_{\propto} = \pm 0{,}03$, per a $\hat{p}$ amb $\propto = 5\%$:

$$n = \frac{z_{\propto/2}^2\, N\, p(1-p)}{(N-1)\, e_{\propto}^2 + z_{\propto/2}^2\, p(1-p)} = 741{,}095 \rightarrow n = 742 \text{ individus}$$

h) Quin és el marge d'error de l'interval anterior?

Marge d'error: $e_{\propto} = \pm 0{,}03$.

Comprovem-ho:

$$\hat{\sigma}(\hat{p}) = \sqrt{\frac{1}{n-1}\,(1-f)\,\hat{p}(1-\hat{p})} = \sqrt{\frac{1}{741}\left(1 - \frac{742}{10000}\right) 0{,}25 * 0{,}75}$$
$$= 0{,}0153$$
$$e_{\propto} = z_{\propto/2}\, \hat{\sigma}(\hat{p}) = 0{,}03$$

i) Si volguéssiu que el marge d'error fos de 0,015, quina hauria de ser la mida mostral aproximada? Responeu sense fer cap càlcul.

Per a reduir el marge d'error a la meitat ($\pm 0{,}015$), caldria multiplicar la mida mostral, aproximadament, per 4 (amb un resultat de gairebé 3000 persones a la mostra).

j) Quin valor de p utilitzarieu a les fórmules del càlcul de n si no s'hagués pres prèviament la mostra de 200 persones? Justifiqueu la resposta.

Si no es pogués disposar d'una estimació prèvia de p, caldria utilitzar un valor de $p = 0{,}5$ per a fer el càlcul de la mida mostral, atès que, per aquest valor, s'assoleix la màxima variància de p i, per tant, es podrà assegurar que l'error de mostreig de l'estimador derivat de la mostra en qüestió és inferior o igual al prèviament fixat.

A les 200 persones de la mostra se'ls demana també quin són els seus ingressos bruts anuals. Resulta una mitjana de 25 u.m. (milers d'euros), amb una quasidesviació estàndard de 5 u.m.

k) Quin serà l'error absolut de mostreig en l'estimació dels ingressos mitjans?

$$\text{Error absolut de mostreig: } e(\bar{x}) = \hat{\sigma}(\bar{x}) = \sqrt{\widehat{Var}(\bar{x})} = \sqrt{(1-f)\,\frac{\hat{S}_x^2}{n}} = 0{,}35$$

l) Calculeu l'interval de confiança al 95% per a la mitjana dels ingressos.

$$\mu \in \left[\bar{x} \pm z_{\propto/2}\,\hat{\sigma}(\bar{x})\right] = (24{,}314;\ 25{,}686)$$

Per Teorema Central del Limit, es pot utilitzar la distribució normal en comptes de la t d'Student, en ésser $n > 100$.

m) Quin marge d'error en l'estimació resulta de l'interval anterior?

$$\frac{25{,}686 - 24{,}314}{2} = 0{,}686 \ \rightarrow\ e_\alpha = \pm 0{,}686$$

n) Com es relaciona la quantitat anterior amb l'error absolut que heu calculat a l'apartat *k)*?

$$e_\propto = z_{\propto/2}\,\hat{\sigma}(\bar{x}) = z_{\propto/2}\,e(\bar{x}) = 1{,}96 * 0{,}35 = 0{,}686$$

o) Estimeu l'ingrés total poblacional.

$$\hat{T} = N\,\bar{x} = 10000 * 25 = 250.000 \text{ milers d'€}$$

p) Construïu un interval de confiança al 95% per al paràmetre anterior.

$$\hat{\sigma}(\hat{T}) = N\,\hat{\sigma}(\bar{x}) = 10000 * 0{,}35 = 3500$$

$$T \in \left[\hat{T} \pm z_{\propto/2}\,\hat{\sigma}(\hat{T})\right] = (243.140;\ 256.860)$$

q) La població estudiada paga a Hisenda un impost fix, equivalent al 15% dels seus ingressos. Construïu un interval de confiança al 95% per a la quantitat total que Hisenda espera recaptar en impostos.

$$\hat{R} = 15\%\,\hat{T} = 37.500 \text{ u.m.}$$

$$\hat{\sigma}(\hat{R}) = 0{,}15\,\hat{\sigma}(\hat{T}) = 0{,}15 * 3500 = 525$$

$$R \in \left[\hat{R} \pm z_{\propto/2}\,\hat{\sigma}(\hat{R})\right] = (36.471;\ 38.529)$$

r) En realitat, Hisenda rep cada any una quantitat de 36.200 u.m. en impostos. És sorprenent aquest resultat? A què es podria atribuir?

Una recaptació de 36.200 u.m. és, efectivament, un resultat sorprenent; ja que, amb un 95% de seguretat, podem dir que la recaptació no serà d'aquesta quantitat. Aquest resultat, doncs, pot ésser degut a l'extracció d'una mostra atípica entre la població d'estudi.

Una altra causa plausible per a l'obtenció d'aquests resultats no coincidents amb la realitat és la possibilitat que s'estigui incomplint alguna de les hipòtesis; entre d'altres, el supòsit de normalitat de les dades potser no és correcte, malgrat treballar amb una mida mostral de 200 individus.

s) Seria la conclusió la mateixa si l'interval de l'apartat *q)* s'hagués calculat al 99% de confiança?

La conclusió difereix en aquest cas, ja que, amb un nivell de confiança del 99%, sí que podem afirmar que una recaptació de 36.200 u.m. no és una recaptació atípica:

$$R \in \left[\hat{R} \pm z_{\propto/2}\, \hat{\sigma}\left(\hat{R}\right)\right] = (36.145{,}5;\ 38.854{,}5), \text{ on } z_{\propto/2} = z_{0{,}5\%} = 2{,}58$$

t) A què es deu la diferència entre la conclusió de l'apartat *q)* i la de *s)*? Creieu que és una solució pujar la confiança si es vol variar la conclusió?

La diferent conclusió que s'assoleix en ambdós apartats rau en la variació del nivell de confiança $1-\propto$: en disminuir $\propto$, l'estimació del paràmetre perd en precisió, l'interval de confiança resultant és més ample i engloba el valor de la recaptació obtinguda.

No es tracta, però, d'una solució recomanable, ja que, per qüestions de consistència i rigorositat estadística, cal mantenir un nivell de confiança invariable al llarg d'un estudi i no variar-lo a voluntat d'obtenir uns resultats més o menys convenients en cada cas.

2.2 EXERCICIS PROPOSATS

1. Es vol estimar la proporció i el total de peces correctes produïdes en un procés industrial, en el què es fabriquen un total de 8.000 unitats. Per fer-ho, es desitja fer servir un mostreig aleatori simple sense reposició. Una mostra pilot conté 1/4 d'unitats correctes.

a) Trobeu la mida de la mostra necessària perquè l'error de mostreig sigui de 0,1 en estimar la proporció de peces correctes produïdes en el procés industrial.

b) Trobeu la mida de la mostra necessària perquè l'error relatiu de mostreig sigui del 20% en la mateixa estimació.

c) Quina condició és més exigent quant a precisió, la de l'apartat *a)* o la del *b)*?

d) Trobeu la mida de la mostra necessària perquè el marge d'error sigui de 600 unitats en estimar el total de peces correctes amb un coeficient de confiança del 99%.

e) Si volguéssim trobar un marge d'error relatiu equivalent a un marge d'error absolut de 600 unitats, quin seria el seu valor?

f) Trobeu la mida de la mostra en les condicions de l'apartat *d)*, però per un marge d'error relatiu del 8%. Abans de fer els càlculs, comenteu si s'espera que la mida de la mostra sigui més gran o més petita que la de l'apartat *d)*.

2. En una regió amb $N = 1.000$ habitatges, determineu la mida de la mostra necessària perquè, amb un nivell de confiança del 99%, l'estimació de la proporció de habitatges sense gas natural no difereixi en més de 0,2 del valor vertader. Suposeu mostreig aleatori simple sense reposició.

3. D'una població de $N = 150$ unitats s'ha extret una mostra de mida $n = 10$ mitjançant mostreig aleatori simple sense reposició, i s'han obtingut les següents dades d'una variable X, mesurada sobre elles: $\{27, 39, 26, 31, 35, 33, 38, 28, 30, 36\}$.

a) A partir d'aquesta mostra, estimeu la mitjana i el total poblacional de X, així com els seus errors absoluts i relatius de mostreig.

b) Quina mida de la mostra seria necessària perquè l'error de mostreig fos de 2 unitats en estimar la mitjana i 75 en estimar el total?

c) I perquè l'error relatiu fos del 6% en tots dos casos?

d) Contesteu també a les dues preguntes anteriors amb un coeficient de confiança del 95% (entenent els errors com marges d'error).

e) A partir de la mostra, estimeu la proporció i el total de nombres parells a la població. Trobeu els errors absoluts i relatius de mostreig.

f) Quina mida de la mostra seria necessària perquè l'error de mostreig fos de $0{,}08$ en estimar la proporció de nombres parells, i de 15 en estimar el total?

g) I perquè l'error relatiu fos del 8% en tots dos casos?

h) Contesteu també a les dues preguntes anteriors amb un coeficient de confiança del 95% (entenent els errors com marges d'error).

4. En una mostra extreta d'una població de 45 milions d'habitants, hi ha 8.000 persones que es consideren població activa; i d'aquestes, 150 estan a l'atur. Si la mostra és de 15.000 individus:

a) Calculeu el percentatge de població activa.

b) Calculeu el nombre total de persones actives que es troben en situació d'atur.

c) Calculeu els errors absoluts i relatius en les dues estimacions anteriors.

d) Trobeu els intervals de confiança amb un risc del 3 per mil per les estimacions d'*a)* i *b)*.

e) Segons resultats censals provisionals, s'estima que en aquesta població hi ha un 52% d'actius. Quantes persones seria necessari incloure en la mostra per estimar la taxa d'activitat amb un error de mostreig absolut de 0,02 i una probabilitat del 95%?

f) Contesteu a la pregunta de *e)* però amb un marge d'error relatiu del 5%.

g) Quina condició és més exigent quant a precisió, la de l'apartat *e)* o la de *f)*?

5. Es desitja estimar la proporció d'empreses que han tingut problemes de solvència durant els darrers dos anys, en una regió on hi ha censades un total de 1.000 empreses. En particular, es vol construir un interval de confiança al 95% per al paràmetre esmentat, que asseguri una precisió de $\pm 0{,}06$.

a) Suposant que es vol utilitzar un mostreig aleatori simple sense reemplaçament, calculeu la mida mostral necessària per a aconseguir-ho.

La mida mostral que s'obté a *a)* es considera fora de les possibilitats del pressupost de què es disposa, de manera que, a efectes orientatius, es decideix començar per prendre una mostra de 10 empreses, que donarà una estimació prèvia de p. El resultat mostra que 2 empreses han tingut dificultats econòmiques.

b) A partir del resultat de la mostra pilot, torneu a calcular la mida mostral que es demanava a *a)*.

c) Comenteu a què són degudes les diferències que s'observen entre les mides mostrals que heu obtingut a *a)* i a *b)*.

d) Suposeu que ja s'ha fet el mostreig, amb la mida mostral que s'ha obtingut a *b)*, i que s'observa que el 25% de les empreses de la mostra tenen problemes de solvència. Trobeu l'interval de confiança al 95% per la proporció poblacional.

e) Comenteu si l'interval que heu trobat a l'apartat anterior verifica les expectatives de precisió que teníeu sobre el mateix. Què ha passat?

f) En general, en poblacions grans, si per una mostra de mida n es té un error de mostreig e, quina serà la mida mostral si es vol que l'error sigui $0{,}5e$?

6. Suposem una població de 18.500 treballadors en una determinada regió. Estem interessats a estudiar la variable X = nombre d'hores diàries treballades. Arrel una enquesta pilot se sap que la quasivariància poblacional estimada és $S^2 = 7{,}341$.

a) Suposant població finita (univers petit):

i. Calculeu *n* per estimar la mitjana d'hores treballades diàriament per cada individu, de manera que hi hagi un error màxim de ± 10 minuts. Treballeu amb $z_{\alpha/2} = 1$, tot comentant a quin nivell de confiança correspon aquest valor.

ii. Trobeu l'error absolut de mostreig de l'estimació anterior.

iii. Feu un comentari crític sobre els resultats anteriors.

iv. Repetiu els mateixos apartats, però ara amb $z_{\alpha/2} = 2$, i un error màxim de ± 5 minuts.

v. Calculeu *n* per estimar el total d'hores treballades diàriament pel col·lectiu, de manera que hi hagi un error màxim de 1.500 hores. Treballeu amb $z_{\alpha/2} = 1$.

vi. Repetiu l'apartat anterior, però amb $z_{\alpha/2} = 2$, i un error màxim de ± 1.000 hores.

vii. Calculeu *n* per estimar la proporció d'individus que treballen cada dia 9 hores o més, de manera que hi hagi un error màxim de $\pm 0{,}025$ (quan es parla de proporcions, se sol dir del 2,5%). Treballeu amb $z_{\alpha/2} = 1$, tot suposant que tenim una estimació inicial, obtinguda a partir d'una enquesta pilot, que mostra $p = 0{,}42$.

viii. Trobeu l'error absolut de mostreig de l'estimació anterior.

ix. Repetiu els dos apartats anteriors, però ara amb un error màxim de $\pm 0{,}012$, i $z_{\alpha/2} = 2$.

b) Suposant població infinita (univers gran), calculeu *n* per estimar la proporció de treballadors que superen les 9 hores diàries, en els supòsits següents. Trobeu, també, en cada cas, l'error absolut de mostreig i la variància de la variable dicotòmica: $p\,(1-p)$.

i. $z_{\alpha/2} = 2,\ P = 0{,}5,\ e_{\propto} = 0{,}02$

ii. $z_{\alpha/2} = 2,\ P = 0{,}5,\ e_{\propto} = 0{,}04$

iii. $z_{\alpha/2} = 2,\ P = 0{,}35,\ e_{\propto} = 0{,}015$

iv. $z_{\alpha/2} = 2,\ P = 0{,}35,\ e_{\propto} = 0{,}03$

7. Un cens proporciona les següents dades per a certes ciutats:

Ciutat	Població
A	5.283
B	47.825
C	2.184
D	1.243.186
E	129.639

Suposem que es vol estimar el percentatge de persones que han estat objecte de robatoris al llarg del darrer any en cadascuna de les ciutats. Es desitja extreure una mostra aleatòria simple sense reemplaçament entre els habitants de cadascuna de les ciutats.

a) Per un error relatiu del $\mathbf{3}\%$, determineu quina ha de ser la mida mostral en cadascuna de les ciutats.

b) Per a quines ciutats podeu afirmar que la correcció per mostra finita és significativa?

8. En una enquesta d'opinió s'entrevista a una mostra de 2.500 adults; se sap que el marc de mostreig conté 25 milions d'adults, entre els quals s'extreu la mostra aleatòria simple sense reemplaçament.

a) Calculeu la probabilitat que l'individu i sigui seleccionat per a la mostra.

b) Si s'extraguessin 1.000 mostres independents, calculeu la probabilitat que l'individu i no estigui en cap de les mostres.

c) Quantes mostres haurien de seleccionar-se per a què l'individu i posseeixi una probabilitat de 0,5 d'estar en una d'elles?

Indicació: $(1-\pi_i)^x = 0{,}5$

9. Per tal d'estimar el percentatge de vivendes sense aigua corrent que hi ha a una regió amb 10.000 habitatges, s'obté una mostra aleatòria simple sense reemplaçament de mida 26.

a) Quin error absolut de mostreig es garanteix en l'estimació?

b) A partir de la mostra extreta s'obté una estimació del paràmetre: $\hat{p} = 0{,}2$. Quin és l'error real que s'està garantint en l'estimació?

10. L'estimació de la proporció de vivendes amb connexió a Internet es duu a terme en una zona amb 3.200 vivendes.

a) Sabent que estudis previs fixen l'esmentada proporció en un 40%, quina mida mostral cal seleccionar per a garantir un error absolut de mostreig inferior a 0,15?

b) Si es publiquen nous estudis que estableixen que la proporció de vivendes amb connexió a la xarxa està entorn del 60%, com fa variar això la mida mostral calculada?

c) I si volem garantir un error relatiu de mostreig del 15%? Repetiu *a)* i *b)*.

11. En una població de mida $N = 9$ formada per les xifres:

$$2, 37, 12, 13, 35, 17, 29, 23, 6$$

a) Determineu la mida mostral perquè en una m.a.s. sense reposició, l'error de mostreig per a l'estimació de la proporció de nombres senars sigui 0,25.

b) Determineu la mida mostral perquè en una m.a.s. sense reposició, l'error relatiu de mostreig per a l'estimació de la proporció de nombres senars sigui del 15%.

c) Determineu la mida mostral perquè en una m.a.s. sense reposició, el marge d'error relatiu per a l'estimació de la proporció de nombres senars sigui del 12%, amb un nivell de confiança del 90%.

12. A partir d'una població de mida $N = 8.000$, se selecciona una mostra aleatòria simple amb $n = 350$:

a) Obteniu el marge d'error absolut de mostreig, amb un nivell de confiança $1-\propto=$ 95%, per a l'estimació de la mitjana de la variable quantitativa *X*. Estudis anteriors demostren que la quasivariància de la variable és 18,5.

b) Un cop mesurada la variable, s'obté:

$$\sum_{i=1}^{n} X_i = 2457 \text{ i } \sum_{i=1}^{n} X_i^2 = 24563$$

Obteniu l'interval de confiança per μ al 95% i compareu-lo amb el que es deriva de l'apartat *a)*.

Indicació: $\hat{S}_{n-1}^2 = \frac{1}{n-1} \left[\sum_i {x_i}^2 - n\bar{x}^2\right]$

13. A una regió de França, 145 famílies de turistes han gastat en mitjana diàriament 830€. La desviació estándar de la despesa familiar diària puja a 210€. Sabent que la regió on es duu a terme l'enquesta va rebre 50.000 famílies de turistes, què pot dir-se de la despesa global diària del conjunt d'aquestes famílies? Calculeu el seu interval de confiança al 90%, sota el supòsit que el mètode de mostreig és l'aleatori simple sense reemplaçament.

14. Una mostra aleatòria simple amb reemplaçament està formada per 400 propietaris de vehicles d'un país donat, dels quals 40 són propietaris d'un vehicle de marca A. Es demana:

a) Construïu un interval de confiança, a un nivell de confiança del 95% per la veritable proporció de propietaris de vehicles de marca A en aquell país.

b) Suposeu que l'extracció és sense reemplaçament i $N = 5.000$; construïu l'interval de confiança al 95% per la veritable proporció de propietaris de vehicles de marca A.

c) Suposeu que l'extracció és sense reemplaçament i $N = 1.000.000$; construïu l'interval de confiança al 95% per la veritable proporció de propietaris de vehicles de marca A.

d) Si es desitgés un error absolut de l'1%, quines haurien de ser les mides mostrals d'extraccions sense reemplaçament per poblacions de mida 5.000 i $1.000.000$, respectivament?

15. Una empresa de màrqueting té accés a un fitxer amb $N = 200.000$ individus. Sigui p el rendiment desconegut del fitxer a una oferta d'inscripció a preu reduït, és a dir, $\hat{p}$ és la proporció d'individus que s'inscriurien si se'ls proposés l'oferta a tots els individus del fitxer.

Per a estimar el rendiment p, s'acostuma a procedir a partir d'un test sobre una mostra d'n individus, triats amb probabilitats iguals i sense reemplaçament. L'experiència de l'empresa és que el rendiment en aquests tipus d'oferta no sol superar el 3$\%$.

a) Quina és la mida mostral necessària per estimar p amb una precisió absoluta del $0,5\%$ a un nivell de confiança del 95$\%$?

b) Quina és la mida mostral necessària per estimar p amb una precisió absoluta del $0,3\%$ a un nivell de confiança del 95$\%$?

c) Quina és la mida mostral necessària per estimar p amb una precisió absoluta del $0,1\%$ a un nivell de confiança del 95$\%$?

d) Si finalment la mostra tingué una mida $n = 10.000$ i es comptabilitzen 230 inscripcions, quin és l'interval de confiança al 95% pel rendiment p i pel número total d'abonats en proposar l'oferta a tots els individus del fitxer?

16. Es duu a terme un mostreig d'opinió preelectoral per a recollir informació sobre l'opinió general d'una personalitat política i s'obté un percentatge d'opinions favorables del 20%.

Si l'extracció és sense reemplaçament, quantes persones han estat enquestades per a poder dir, amb un nivell de confiança del 95%, que el veritable percentatge d'opinions favorables en la població no es desvia més de 2 punts ($\pm 0{,}02$) de $\hat{p}$?

17. Un institut d'estudis sociològics s'encarrega d'estudiar la població de lectors d'un setmanari. S'empra un procediment de selecció de la mostra assimilable a un mostreig aleatori simple sense reposició per a triar $n = 2.000$ individus més grans de 15 anys. Els resultats del mostreig es troben a la taula següent:

	Àmbit Rural	*Àmbit Urbà*	*Total*
Lectors	64	476	540
No Lectors	576	884	1.460
Total	**640**	**1.360**	**2.000**

a) Estimeu de forma puntual i per interval la proporció de lectors del setmanari en l'àmbit urbà, al rural i al conjunt de la població.

b) Existeix un *efecte rural versus urbà* en las proporcions de lectors (és a dir, existeixen diferències significatives entre ambdues estimacions)?

2.3 SOLUCIONS

1. a) $n = 18{,}71 \rightarrow 19$ peces

b) $n = 74{,}31 \rightarrow 75$ peces

c) A l'apartat *a)*, $e_r = 40\%$, que és menys exigent que a l'apartat *b)*.

d) $n = 215{,}92 \rightarrow 216$ peces

e) $e_{r\propto} = 30\%$

f) $n = 2.244{,}9 \rightarrow 2.245$ peces

2. $n = 39{,}98 \rightarrow 40$ habitatges

3. a) $\hat{\mu} = 32{,}3 \rightarrow e(\hat{\mu}) = 1{,}411$

$e_r(\hat{\mu}) = 4{,}37\%$

$\hat{T} = 4.845 \rightarrow e(\hat{T}) = 211{,}713$

$e_r(\hat{T}) = 4{,}37\%$

b) $\left.\begin{array}{r} e(\hat{\mu}) = 2 \leftrightarrow n = 5{,}15 \rightarrow n = 6 \text{ unitats} \\ e(\hat{T}) = 75 \leftrightarrow n = 54{,}41 \rightarrow n = 55 \text{ unitats} \end{array}\right\} n = 55 \text{ unitats}$

c) $n = 5{,}48 \rightarrow n = 6$ unitats

d) $\left.\begin{array}{r} e_\alpha(\hat{\mu}) = 2 \leftrightarrow n = 18{,}03 \rightarrow n = 19 \text{ unitats} \\ e_\alpha(\hat{T}) = 75 \leftrightarrow n = 102{,}93 \rightarrow n = 103 \text{ unitats} \end{array}\right\} n = 103 \text{ unitats}$

$e_{r\alpha} = 6\% \leftrightarrow n = 19{,}06 \rightarrow n = 20$ unitats

e) $\hat{p} = 50\% \rightarrow e(\hat{p}) = 0{,}161$

$e_r(\hat{p}) = 32{,}203\%$

$\hat{A} = 75 \rightarrow e(\hat{A}) = 24{,}152$

$e_r(\hat{A}) = 32{,}203\%$

f) $\left.\begin{array}{r} e(\hat{p}) = 0{,}08 \leftrightarrow n = 31{,}16 \rightarrow n = 32 \text{ unitats} \\ e(\hat{A}) = 15 \leftrightarrow n = 21{,}54 \rightarrow n = 22 \text{ unitats} \end{array}\right\} n = 32 \text{ unitats}$

g) $n = 76{,}78 \rightarrow n = 77$ unitats

h) $\left.\begin{array}{r} e_\alpha(\hat{p}) = 0{,}08 \leftrightarrow n = 75{,}27 \rightarrow n = 76 \text{ unitats} \\ e_\alpha(\hat{a}) = 15 \leftrightarrow n = 58{,}79 \rightarrow n = 59 \text{ unitats} \end{array}\right\} n = 76 \text{ unitats}$

$e_{r\alpha} = 8\% \leftrightarrow n = 120{,}17 \rightarrow n = 121$ unitats

4. a) $\hat{p}_{Act} = 53{,}333\%$

b) $\hat{A}_{At} = 450.000$

c) $\hat{p}_{Act} \rightarrow e(\hat{p}_{Act}) = 0{,}004073$

$e_r(\hat{p}_{Act}) = 0{,}764\%$

$\hat{A}_{At} \rightarrow e(\hat{A}_{At}) = 36392{,}465$

$e_r(\hat{A}_{At}) = 8{,}087\%$

d) $p_{Act} \in (52{,}124\%;\ \ 54{,}543\%)$

$A_{At} \in (341.914;\ \ 558.086)$

e) $n = 623{,}99 \rightarrow n = 624$ persones

f) $n = 1.418{,}39 \rightarrow n = 1.419$ persones

g) A l'apartat *e)*, $e_{r\propto} = 7{,}35\%$, que és menys exigent que a l'apartat *f)*.

5. a) $n = 210{,}76 \rightarrow n = 211$ empreses

b) $n = 145{,}96 \rightarrow n = 146$ empreses

d) $p \in (18{,}487\%;\ \ 31{,}513\%)$

f) $n' = 4 * n$

6. a) i. $n = 260{,}55 \rightarrow n = 261$ treballadors

ii. $e(\bar{x}) = {}^{1}/_{6}$

iv. $n = 3.441{,}76 \rightarrow n = 3.442$ treballadors

v. $n = 1.053{,}08 \rightarrow n = 1.054$ treballadors

vi. $n = 6.512{,}19 \rightarrow n = 6.513$ treballadors

vii. $n = 318{,}74 \rightarrow n = 319$ treballadors

viii. $e(\hat{p}) = 0{,}025$

ix. $n = 4.954{,}68 \rightarrow n = 4.955$ treballadors
$e(\hat{p}) = 0{,}012$

b) i. $n = 2.500$ treballadors
$e(\hat{p}) = 0{,}01$
$p * (1 - p) = 0{,}25$

ii. $n = 625$ treballadors
$e(\hat{p}) = 0{,}02$
$p * (1 - p) = 0{,}25$

iii. $n = 4.044{,}44 \rightarrow n = 4.045$ treballadors
$e(\hat{p}) = 0{,}0075$
$p * (1 - p) = 0{,}2275$

iv. $n = 1.011{,}11 \rightarrow n = 1.012$ treballadors
$e(\hat{p}) = 0{,}015$
$p * (1 - p) = 0{,}2275$

7.

Ciutat	n_{CPF}	f	n_{∞}	*CPF significativa*
A	919	17,395%	1.111	Sí
B	1.086	2,271%	1.111	No
C	737	33,746%	1.111	Sí
D	1.110	0,089%	1.111	No
E	1.102	0,850%	1.111	No

8. a) $\pi_i = 0{,}01\%$

b) $Prob(i \notin 1000 \text{ mostres}) = 90{,}483\%$

c) 6.932 mostres

9. a) $e(\hat{p}) = 0{,}099$

b) $e(\hat{p}) = 0{,}079$

10. a) $n = 10{,}63 \rightarrow n = 11$ llars

b) $n = 10{,}63 \rightarrow n = 11$ llars

c) $p = 0{,}4 \Longrightarrow n = 65{,}33 \rightarrow n = 66$ llars

$p = 0{,}6 \Longrightarrow n = 29{,}37 \rightarrow n = 30$ llars

11. a) $n = 2{,}77 \rightarrow n = 3$ xifres

b) $n = 6{,}62 \rightarrow n = 7$ xifres

c) $n = 8{,}29 \rightarrow n = 9$ xifres

12. a) $e_{\propto} = 0{,}441$

b) $\mu \in (6{,}551;\ 7{,}489)$

$S_{n-1}^2 = 18{,}5 \rightarrow\ \mu \in (6{,}579;\ 7{,}641)$

13. $\hat{T} = 41.500.000$ €; $T \in (40.070.000;\ 42.930.000)$

14. a) $\hat{p} = 10\%$; $p \in (7{,}1\%;\ \ 12{,}9\%)$

b) $p \in (7{,}2\%;\ \ 12{,}8\%)$

c) $p \in (7{,}1\%;\ \ 12{,}9\%)$

d) Per $N = 5.000,\ n = 2.045$; per $N = 1.000.000,\ n = 3.447$.

15. a) $n = 4.375$ individus

b) $n = 11.696$ individus

c) $n = 71.710$ individus

d) $p \in (2{,}014\%;\ 2{,}586\%)$; $T \in (4.017;\ \ 5.173)$

16. a) $n = 1538$ persones

b) $n = 1538$ persones

17. a) $\hat{p}_{Urbà} = 35\%$; $p_{Urbà} \in (32{,}464\%;\ \ 37{,}536\%)$

$\hat{p}_{Rural} = 10\%$; $p_{Rural} \in (7{,}674\%;\ \ 12{,}326\%)$

$\hat{p}_{Conjunt} = 27\%$; $p_{Conjunt} \in (25{,}054\%;\ \ 28{,}946\%)$

b) Existeix *efecte rural vs urbà*, ja que ambdós intervals de confiança tenen intersecció nul·la.

Tema 3. MOSTREIG ALEATORI ESTRATIFICAT

MOSTREIG ALEATORI ESTRATIFICAT

3.1 EXERCICI RESOLT

Una entitat financera ofereix un nou tipus de llibreta amb una taxa d'interès atractiva per a les persones que dipositen diners a termini fix. Per tal de saber quina serà l'actitud dels seus clients, decideix dur a terme una enquesta a 500 clients. A les persones seleccionades, se'ls preguntarà quina quantitat de diners pensen dipositar al llarg del següent any.

Es decideix dur a terme un mostreig estratificat amb afixació proporcional, fent servir l'edat com a criteri d'estratificació. Es disposa de la següent informació:

Classe d'edat	*Núm. de clients*
De 18 a 24 anys	500
De 25 a 50 anys	1.000
Majors de 50 anys	1.500

a) Per què s'ha optat pel mostreig aleatori estratificat, enlloc d'escollir un mostreig aleatori simple? Què fa pensar que el m.a.e. pot millorar els resultats del m.a.s.?

La principal raó per la qual es tria el mostreig de tipus estratificat és que la variable edat és un factor d'estratificació que pot introduir diferències significatives en la quantitat dels dipòsits bancaris, transformant una població força heterogènia en subpoblacions més homogènies. És per aquest motiu que aquest tipus de mostreig pot ser més adequat que l'aleatori simple per a dur a terme l'estudi en qüestió.

b) Amb el mètode proposat a l'enunciat, calculeu la mida de la submostra de cada estrat.

Classe d'edat	W_h	n_h
De 18 a 24 anys	$1/6$	83
De 25 a 50 anys	$1/3$	167
Majors de 50 anys	$1/2$	250
		500

Suposem que ja heu pres la mostra per afixació proporcional. Els resultats, en euros, sobre la quantitat que pensen dipositar en aquest compte són:

Classe d'edat	$\bar{x}_h$	$\hat{S}_h$
De 18 a 24 anys	15	10
De 25 a 50 anys	64	22
Majors de 50 anys	150	40

c) Estimeu de forma puntual i per interval (al 95%) la quantitat mitjana que pensen invertir els clients del banc.

$$\bar{x} = \sum\nolimits_h (W_h \, \bar{x}_h) = 98{,}83€$$

$$\widehat{Var}(\bar{x}) = \sum\nolimits_h \left[W_h^2 \, (1 - f_h) \, \frac{\hat{S}_h^2}{n_h} \right] = 1{,}6295$$

$$\mu \in \left[\bar{x} \pm z_{\alpha/2} \, \hat{\sigma}(\bar{x}) \right] = (96{,}33; \; 101{,}34)$$

d) Estimeu de forma puntual i per interval (al 95%) la quantitat total que pensen invertir els clients.

$$\hat{T} = N \sum\nolimits_h (W_h \, \bar{x}_h) = N * \bar{x} = 296.499{,}9€$$

$$\hat{\sigma}(\hat{T}) = \sqrt{\widehat{Var}(\hat{T})} = \sqrt{N^2 \, \widehat{Var}(\bar{x})} = 3000\sqrt{1{,}6295} = 3829{,}5561$$

$$T \in \left[\hat{T} \pm z_{\alpha/2} \, \hat{\sigma}(\hat{T}) \right] = (288.993{,}97; \; 304.005{,}83)$$

e) Després d'observar aquesta primera mostra, si haguéssiu de repetir el mostreig, us semblaria interessant escollir les mides de les submostres de forma diferent? Quin tipus d'afixació proposaríeu? Raoneu la resposta.

Un cop obtinguts els resultats de la primera mostra, es veu que les variàncies dels estrats són molt diferents, cosa que fa pensar que una correcció segons la variància de l'estrat milloraria l'eficiència dels estimadors. Es proposa, doncs, el mètode d'afixació de variància mínima.

f) Calculeu la mida de les submostres segons el mètode que acabeu de proposar.

$$n_h = n \, \frac{N_h \, S_h}{\sum_h (N_h \, S_h)} = 500 * \frac{N_h \, S_h}{500 * 10 + 1000 * 22 + 1500 * 40}$$

$$n_h = \begin{cases} 28{,}74 \rightarrow n_1 = 29 \\ 126{,}44 \rightarrow n_2 = 127 \\ 344{,}83 \rightarrow n_3 = 345 \end{cases}$$

g) Suposant que els resultats de l'enquesta fossin els mateixos, calculeu ara la variància dels estimadors de la mitjana i el total poblacionals. En quin sentit s'han modificat?

$$\widehat{Var}(\bar{x}) = \sum\nolimits_h \left[W_h^2 \, (1 - f_h) \, \frac{\hat{S}_h^2}{n_h} \right] = 1{,}3560$$

$$\hat{\sigma}(\hat{T}) = \sqrt{\widehat{Var}(\hat{T})} = \sqrt{N^2 \, \widehat{Var}(\bar{x})} = 3000\sqrt{1{,}356} = 3493{,}4398$$

Tal com era d'esperar, l'ús del mètode d'afixació de variància mínima assoleix la mínima variància dels estimadors possible. En aquest cas, la variància es redueix en un 16,78%.

h) Calculeu de nou l'estimació per interval al 95% de la quantitat mitjana, tot comentant les diferències amb el resultat obtingut a *c)*.

$$\mu \in \left[\bar{x} \pm z\alpha_{/2}\, \hat{\sigma}(\bar{x})\right] = (96{,}55;\ 101{,}12)$$

En disminuir la variància dels estimadors, la longitud de l'interval de confiança disminueix (en 0,54€, en aquest cas), assolint una major precisió de l'estimació de la que es tenia amb l'afixació proporcional.

i) Redacteu breument una conclusió raonant quin dels dos mètodes d'afixació ha donat millors resultats.

A partir dels resultats que s'ha obtingut en els apartats anteriors, és obvi que el mètode d'afixació de variància mínima proporciona millors estimacions en el cas d'estudi. La raó que es donava a *e)*, la diferent variància dels subgrups, es confirma com a motiu perquè el mètode d'afixació de variància mínima proporcioni millors resultats.

j) Suposem ara que l'interval obtingut a *h)* no ens sembla prou precís, i que volem estimar la mitjana amb un precisió de $\pm 1{,}5$, és a dir, que volem un interval de 3€ de longitud. Quina seria la mida de la mostra necessària, si es pensa utilitzar la millor de les dues afixacions?

$$n = \frac{\left[\sum_h (W_h * S_h)\right]^2}{\frac{e_\propto{}^2}{z\propto_{/2}{}^2} + \frac{1}{N} * \sum_h (W_h * S_h^2)} = 922{,}46 \rightarrow 923 \text{ individus}$$

3.2 EXERCICIS PROPOSATS

1. Una població de mida $N = 1000$ està dividida en tres estrats dels què es coneixen les desviacions típiques poblacionals $\sigma_1 = 4$, $\sigma_2 = 12$ i $\sigma_3 = 80$ i els seus pesos dins la població: $W_1 = 0{,}6$, $W_2 = 0{,}3$ i $W_3 = 0{,}1$. Es vol estimar la mitjana poblacional amb un error de mostreig igual a $\sqrt{5}$.

a) Determineu la mida mostral necessària per a dur a terme una afixació proporcional.

b) Realitzeu l'afixació proporcional, és a dir, determineu la mida de la mostra de cada estrat.

c) Repetiu els dos apartats anteriors, però amb afixació de mínima variància. Compareu els resultats.

d) Determineu la mida de la mostra per afixació òptima, tot sabent que el cost d'obtenció d'informació d'una unitat és el següent, per cadascun dels estrats: $C_1 = 1000$, $C_2 = 1200$ i $C_3 = 2000$.

e) Seguint amb l'apartat anterior, realitzeu l'afixació òptima. Comproveu que els resultats coincideixen amb els de variància mínima quan els costos són unitaris als tres estrats.

2. Suposem conegudes les següents dades d'una població dividida en tres estrats: $S_1^2 = 9$, $S_2^2 = 225$, $S_1^3 = 1600$, $N_1 = 1000$, $N_2 = 600$, $N_3 = 200$, $C_1 = 1000$, $C_2 = 1200$ i $C_3 = 2000$.

a) Determineu el cost d'una mostra estratificada que proporciona un error relatiu de mostreig del 5% per estimar la mitjana, tot considerant afixació proporcional. Se sap que $\bar{x} = 22$ i que la funció de cost és lineal.

b) Repetiu el mateix, però per afixació de mínima variància.

c) Repetiu de nou el mateix; en aquest cas, per afixació òptima.

d) Amb quina de les 3 mostres s'obtindrà un interval de confiança més curt?

e) Donada una mostra de mida *n*, amb quina de les tres afixacions s'obtindria un interval més precís?

f) Suposem que volem escollir una mostra de mida $n = 100$. Feu les tres afixacions anteriors i, per a cadascuna, trobeu l'error mostral, l'IC al 95%, i el cost total d'extracció de la mostra. Comenteu els resultats.

3. En mesurar la variable *X*, que representa el nombre de vegades que els habitants d'una població han anat en avió, s'obté la següent distribució de freqüències:

X	1	2	4	8	16	32	64	128
N_h	135	141	124	157	76	42	31	9

Amb l'objectiu d'establir grups homogenis, s'estratifica la població fent servir el criteri donat per: $1 \leq X < 8, 8 \leq X < 64$, $64 \leq X \leq 128$. Per a l'extracció d'una mostra de mida $n = 100$, determineu l'afixació de mínima variància per a l'estimació del nombre mitjà de vols en avió. Comenteu els resultats.

Indicació:

	N_h	S_h	$\bar{x}_h$
$1 \leq X < 8$			
$8 \leq X < 64$			
$64 \leq X \leq 128$			

4. Una productora de televisió local vol estimar la proporció de llars d'un cert municipi del Pirineu català on es veu un determinat tipus de programa televisiu. El poble té un total de 400 llars i es pot dividir en tres estrats, en funció de la llengua que parla la família. Es selecciona una mostra estratificada de $n = 48$ llars amb afixació proporcional. Les dades són les següents:

Estrats	*Mides mostrals*	*Núm. de llars de la mostra on es veu el programa*
1	24	17
2	18	11
3	6	5

a) Estimeu la proporció *p* de llars del municipi on es veu el programa televisiu.

b) Trobeu els errors absolut i relatiu de mostreig de l'anterior estimació.

c) Trobeu una estimació per interval al 90% per *p*.

d) Calculeu la mida mostral necessària perquè, amb un nivell de confiança del 90%, l'estimació de *p* no difereixi en més de 0,05 del vertader valor del paràmetre.

5. Desitgem conèixer el nombre mitjà d'empleats per empresa en una regió molt industrial. En aquesta regió existeixen un total de 10.000 empreses que es reparteixen en: 8.000 petites (1-50 empleats), 1.800 mitjanes (51-1.000 empleats) i 200 grans (més de 1.000 empleats). S'extreu una mostra de 100 empreses, utilitzant mostreig estratificat amb afixació proporcional.

Les mitjanes i quasivariàncies mostrals del nombre d'empleats per cada tipus d'empresa es recullen a la següent taula:

Estrat	*Mitjana*	*Quasivariància*
Petites	12	100
Mitjanes	70	3600
Grans	1120	22500

a) Quina ha estat la mida mostral a cada estrat?

b) Estimeu el nombre mitjà d'empleats per empresa a la regió estudiada.

c) Estimeu el paràmetre anterior per interval, al 90% de confiança.

d) En general, si es desitja estimar una mitjana poblacional, quina condició han de complir els estrats per tal que el mostreig estratificat superi en precisió al mostreig aleatori simple?

6. Es vol estudiar l'hàbit de fumar en els espanyols, en funció del nivell d'educació. En particular, l'estudi es vol fixar en el percentatge que afirmen no haver fumat mai. S'ha fet un mostreig aleatori estratificat amb afixació proporcional, amb una mostra total de mida 2.498 i s'ha obtingut els següents resultats:

ESTUDIS	*Mida mostral (m.e.p.)*	*% no han fumat mai*	*Error de mostreig*	*Marge d'error (95,45%)*
< Primaris	762	61		
Primaris	899	48		
Batxillerat	578	33		
Superiors	242	27		
NS/NC	17	53		
TOTAL	**2498**			

a) Completeu la taula anterior.

Indicació: El mètode de mostreig aleatori estratificat rep aquest nom, degut a que els elements de la mostra pertanyents a cada estrat s'escullen segons un mostreig aleatori simple. Les variàncies dels paràmetres que en deriven es calculen en conseqüència. Així, és només pel total de la mostra que s'aplica les fórmules pròpies del mostreig estratificat.

b) Compareu la precisió dels diferents estrats amb la del total.

c) Construïu l'interval de confiança al 95,45% per a la proporció de gent amb estudis superiors que no ha fumat mai. Es pot considerar que aquesta proporció és diferent a la de persones amb Batxillerat que no han fumat mai?

d) Es pot dir que la precisió del grup d'estudis superiors és suficient per garantir una bona estimació?

e) Suposem que en el grup d'estudis superiors es vol aconseguir una precisió en l'estimació de $\pm 0{,}036$ (amb un 95,45% de confiança). Quantes entrevistes farien falta?

f) Si es volgués conservar l'afixació proporcional, calculeu quantes entrevistes caldrien a cadascun dels grups, i quantes en total.

g) Quina precisió s'obtindria aleshores en el grup d'estudis primaris?

h) Suposem que els grups que més interès tenim a estudiar per separat són els de batxillerat i estudis superiors. En tots dos volem una precisió de $\pm 0{,}036$ (amb confiança $1-\propto = 95{,}45\%$). Proposeu una afixació alternativa a la proporcional, que no impliqui treballar amb una mostra tan gran com l'obtinguda a *f)*. Comenteu quines limitacions tindria, en contrapartida.

i) A partir de la mostra obtinguda a l'apartat anterior, estimeu de nou el percentatge de població que no ha fumat mai, i comenteu el resultat.

7. Una població d'1.000.000 de llars està formada per un 30% de cases, un 10% de dúplex i un 60% de pisos. Estudis preliminars indiquen que el consum mitjà d'aigua a les cases és el doble que als dúplex, i el triple que als pisos; i també s'intueix que la desviació típica del consum és directament proporcional a la mitjana. Si se selecciona una mostra de 1.000 llars per a fer l'esmentat anàlisi, quina és la forma òptima de distribuir aquesta mostra entre cases, dúplex i pisos? Realitzeu l'afixació seleccionada.

8. Un investigador divideix una població en tres estrats d'igual mida, selecciona una mostra aleatòria en els mateixos i hi calcula la mitjana i la variància d'una variable d'estudi. S'observa que:

a) Les mitjanes estimades són gairebé iguals, mentre que les variàncies són molt diferents. Vistos els resultats, avalueu la conveniència de dur a terme un mostreig estratificat o un aleatori simple.

b) Les mitjanes estimades són molt diferents, mentre que els tres estrats tenen variàncies estimades molt semblants. Avalueu, també, la conveniència de dur a terme un mostreig estratificat o un aleatori simple.

9. Una empresa publicitària està interessada a saber què ha d'emfatitzar en la publicitat televisiva en un municipi determinat. Així, decideix realitzar una enquesta per a estimar el nombre d'hores setmanals que la població veu la televisió a les llars del municipi. Aquest està format per dos pobles (A, amb 155 habitants i B, amb 62 habitants) i una zona rural (amb 93 habitants).

Es duu a terme un mostreig estratificat, del que s'obté la següent informació:

Estrat	n_h	S_h^2	$\bar{x}_h$
A	20	35,358	33,9
B	8	232,411	25,125
Rural	12	87,636	19

a) Amb un nivell de significació $1-\propto= 0{,}95$, obteniu l'error d'estimació i l'interval de confiança de la mitjana d'hores setmanals que es veu la televisió en el conjunt del municipi.

b) A partir de la informació de la taula anterior, es podrien reduir els errors d'estimació modificant el mètode d'afixació (mantenint, però, la mida mostral $n = 40$)? Obteniu les noves mides mostrals per cada estrat i torneu a calcular l'error de mostreig.

c) En base a la informació anterior, l'empresa fa la següent aproximació de les quasidesviacions típiques de cada estrat: $\sigma_1 = 6$, $\sigma_2 = 15$ i $\sigma_3 = 9$. Calculeu la mida mostral total (n) i per cadascun dels estrats (n_h) que permetran estimar, amb una afixació de mínima variància, la mitjana d'hores setmanals, per a obtenir una estimació amb un marge d'error de ± 2 hores i $\propto = 0{,}05$.

d) L'empresa estima uns costs d'obtenció de les dades de 9€ a cadascun dels pobles, quantitat que incrementa en un 50% per la zona rural. A partir de les mateixes aproximacions de les quasidesviacions dels estrats que a *c)*, obteniu la mida mostral (n) i dels estrats (n_h) que, amb l'afixació de cost mínim, permetran a l'empresa estimar la mitjana d'hores setmanals, amb un marge d'error de ± 2 hores i $\propto = 0{,}05$.

e) Compareu el cost per l'empresa a *c)* i *d)*.

10. Una coneguda marca de cereals vol llençar al mercat un nou producte i desitja saber quina rebuda tindrà. Es decideix fer l'estudi a una gran superfície comercial, on gran part dels clients compren mostrant una targeta que recull les seves dades personals. Així, se sap que, durant el darrer any, van comprar en algun moment cereals de qualsevol marca 6.000 clients d'edats entre 18 i 25 anys; 45.000 clients d'edats entre 26 i 40 anys; i 32.000 clients de més de 40 anys. Totes aquestes persones es consideren potencials compradors del nou producte i constituiran la població objectiu.

Suposant que l'estructura poblacional es manté constant, es pren una mostra estratificada per edats de 1.000 compradors de cereals fent servir afixació proporcional, i se'ls demana que diguin quina quantitat del nou producte estimen que compraran al llarg d'un any. Traduint els resultats a la despesa que suposa, s'obté la següent taula:

Classe d'edat	*Mitjana*	*Quasidesviació típica*
De 18 a 25 anys	2.000	4000
De 26 a 40 anys	6.000	1000
Més de 40 anys	3.000	2000

a) Raoneu per què s'ha optat pel mostreig estratificat, enlloc d'escollir un mostreig aleatori simple.

b) Calculeu la mida de la submostra que s'ha pres a cada estrat.

c) Estimeu de forma puntual i per interval (al 90%) la quantitat mitjana que pensen gastar en un any els potencials compradors del nou producte.

d) L'empresa aspira a vendre 300 milions d'u.m. del nou producte en el pròxim any, dins la gran superfície considerada. Necessita dur a terme alguna mena de publicitat? Justifiqueu quantitativament la resposta.

e) Després d'observar aquesta primera mostra, si haguéssiu de repetir el mostreig, us semblaria interessant treballar amb un altre tipus d'afixació? Raoneu la resposta.

f) Calculeu la mida de les submostres segons el mètode que acabeu de proposar.

g) Suposant que els resultats de l'enquesta fossin els mateixos, calculeu novament la variància de l'estimador de la mitjana. En quin sentit s'ha modificat?

h) Calculeu de nou l'estimació per interval, al 90%, de la quantitat mitjana, tot comentant les diferències amb el resultat obtingut a *c)*.

i) Suposem ara que l'interval obtingut a *h)* no ens sembla prou precís, i que, mantenint el mateix nivell de confiança, volem estimar la mitjana amb un precisió de ± 50. Quina seria la mida de la mostra necessària, si es pensa utilitzar la millor de les dues afixacions?

11. L'exercici *14*, del tema *1*, conté dades sobre l'assistència a congressos internacionals per part una mostra de 70 professors d'una universitat en els darrers dos anys. A partir d'aquelles dades, es vol dur a terme un mostreig estratificat, on els estrats representen els diferents departaments de la universitat en qüestió. Si s'usa l'afixació proporcional, els resultats que s'obtenen són els següents:

Departament	N_h	n_h
Biologia	52	8
Economia	160	26
Estadística	106	17
Humanitats	118	19
TOTAL	436	70

Les freqüències del nombre de congressos a què han assistit els professors a cada departament es recullen a continuació:

Nombre de congressos internacionals	*Biologia*	*Economia*	*Estadística*	*Humanitats*
0	5	15	10	11
1	1	3	0	2
2	1	1	1	0
3	0	0	2	2
4	0	2	2	2
5	0	1	1	1
6	1	1	0	0
7	0	0	0	0
8	0	2	0	0
9	0	0	1	1
10	0	1	0	0

a) Estimeu el nombre mitjà de congressos internacionals a què han assistit els professors de la universitat, per mitjà de predicció puntual i amb un interval de confiança, amb $\propto = 5\%$.

Indicació:

	W_h	$\bar{x}_h$	$\hat{S}_h^2$
Biologia			
Economia			
Estadística			
Humanitats			

b) Estimeu la proporció de professors que no han assistit a cap congrés durant els darrers dos anys, mitjançant una predicció puntual i amb un interval de confiança, amb $\propto = 5\%$.

12. Considerem la variable despesa anual en transport (en centenars d'unitats monetàries). Com a resultat de mesurar-la en una població de 940 persones, s'obté la següent distribució de freqüències:

X	1	2	5	8	12	15	20	24	30	37	42	60	85	120
Freqüència	25	40	55	120	180	250	110	65	40	25	15	10	3	2

Per tal d'establir pautes per a futures enquestes on es pregunti sobre aquest tipus de despesa, s'estratifica la població de dues formes lleugerament diferents.

Mètode A:

Estrat	*Interval*
1	$1 \leq X \leq 8$
2	$9 \leq X \leq 24$
3	$25 \leq X \leq 120$

Mètode B:

Estrat	*Interval*
1	$1 \leq X \leq 12$
2	$13 \leq X \leq 30$
3	$31 \leq X \leq 120$

a) Si es vol extreure una mostra de 150 persones, obteniu l'afixació proporcional i de variància mínima per cadascun dels mètodes d'estratificació. Compareu els resultats, en base a l'error de mostreig en l'estimació de la mitjana poblacional.

b) Feu el mateix pel mètode d'afixació òptima, si:

$$C_{1A} = 1 \quad C_{2A} = 25 \quad C_{3A} = 16$$
$$C_{1B} = 9 \quad C_{2B} = 9 \quad C_{3B} = 36$$

On C_{ij} és el cost d'accedir i entrevistar una persona de l'estrat *i* segons el mètode *j*. Quin és ara el millor mètode d'estratificació?

c) En base al mètode triat a l'apartat *a)*, quina mida mostral seria necessària per a fer una enquesta en la població, si es vol aconseguir un error de mostreig de $0{,}75$ per a l'estimació de la mitjana de la despesa en transport i usant l'afixació de mínima variància?

d) I si es vol un error relatiu del 10% en estimar el total de despesa en transport i usant l'afixació proporcional?

13. Un estadístic disposa d'un pressupost de 20.000€ per a dur a terme un estudi a 1.000 persones sobre el nombre d'hores diàries de connexió a internet a les llars d'una regió formada per dos municipis. Els costos fixos de realitzar l'estudi són de 8.000€.

S'estima que el cost d'entrevista per telèfon és de 10€ i el cost d'entrevista personal és de 15 i 18€, respectivament (el 90% de les llars del poble A tenen telèfon, mentre que al poble B, només el 85%).

Les quasivariàncies pels estrats (definits pel poble de residència) són iguals a 25 i les mides de les poblacions són les següents: $N_A = 5000$, $N_B = 7000$.

Per quina de les següents alternatives haurà d'optar, si el límit pressupostari no és alterable?

1. Totes les entrevistes es fan personalment.
2. A les famílies amb telèfon se les entrevista telefònicament.

14. Una empresa realitza una enquesta sobre el conjunt del seu personal (format per 10.000 persones). Estudis preliminars han demostrat que les variables d'interès estan fortament correlacionades amb l'edat dels individus i que es poden establir tres categories d'edat, cadascuna de les quals constitueix un estrat.

Es proposa un pla de mostreig com si es volgués estudiar l'edat dels individus i es coneix l'edat de tot el personal, informació que es pot sintetitzar a la següent taula:

Estrat	W_h	S_h
1	0,2	18
2	0,3	12
3	0,5	3,6
Conjunt	1	16

a) Sigui μ l'edat mitjana de la població i $\bar{y}$ l'estimador mitjana mostral procedent d'una extracció aleatòria simple sense reposició de $n = 100$. Quin és l'error estàndard d'$\bar{y}$?

b) Es decideix efectuar l'extracció dels $n = 100$ individus de manera estratificada segons les 3 categories del personal. Quina és l'afixació proporcional? Quina és la desviació estàndard de l'estimador de μ resultant? Compareu els resultats amb l'apartat anterior.

c) Quina seria una millor afixació per la mostra? Quina seria la desviació estàndard de l'estimador de μ resultant? Compareu els resultats amb els apartats anteriors.

15. Una determinada població d'estudi es composa de 2 estrats dels quals es coneix la mida $(N_h,\ h = 1,2)$ i la seva variància poblacional corregida $(\sigma_h,\ h = 1,2)$. Es disposa d'un pressupost C per a obtenir una estimació de μ. La funció de cost s'escriu $C(n) = C + C_1 n_1 + C_2 n_2$ i $n_1 + n_2 = n$, i es proposa una extracció aleatòria sense reposició a cada estrat.

a) Si $n = 1000$, $N_1 = 10.000$, $N_2 = 20.000$, $\sigma_1 = 2$, $\sigma_2 = 1, C_1 = 4$, $C_2 = 9$ i $C = 1000$ calculeu l'afixació òptima en costos de la mostra.

b) Apliqueu també les dades a l'afixació proporcional i avalueu la pèrdua de precisió entre l'afixació òptima en costos i l'afixació proporcional per a l'estimació de la mitjana de la variable.

16. Una empresa que està formada per 400 persones de suport i 100 directius vol avaluar l'índex de satisfacció (Y) del seu personal elaborat a partir d'un conjunt de preguntes a una mostra de $n = 100$ individus.

a) Se suposa que la dispersió de la variable satisfacció és la mateixa en els dos estrats del personal. En aquest cas, quin és el mètode de mostreig indicat si es vol obtenir la millor precisió possible sobre el valor mig de l'índex de satisfacció del personal?

b) Finalment es realitza l'enquesta segons l'afixació de l'apartat anterior, tot obtenint els següents resultats: $\bar{y}_1 = 13$, $\bar{y}_2 = 15$, $\hat{\sigma}_1^2 = 9$ i $\hat{\sigma}_2^2 = 36$. Calculeu l'interval de confiança per la mitjana de l'índex de satisfacció del personal al nivell de confiança del 95%.

17. Una empresa especialitzada és l'encarregada de realitzar un mostreig en dues regions sobre una mostra total de $n = 500$ explotacions ramaderes, amb la finalitat d'avaluar el nombre mig d'animals per explotació (μ). El nombre d'explotacions total és de 50.000, distribuïdes de la següent manera: 40.000 a la regió 1 i 10.000 a la regió 2. Es disposa d'una base de mostreig amb l'adreça de cada explotació.

Dades històriques sobre les regions mostren que la quasidesviació típica del nombre d'animals per explotació a la regió 1 era de 20 i a la regió 2 de 40.

a) Descriviu com es construiria una mostra estratificada amb afixació proporcional prenent com criteri d'estratificació la regió. Quina és la diferència respecte haver considerat una mostra aleatòria simple sense estratificació pel conjunt del territori?

b) Quin és el marge d'error en l'estimació de la mitjana d'animals per explotació a cada regió i en el conjunt d'ambdues segons el mostreig aleatori estratificat amb afixació proporcional, a un nivell de confiança del 95%?

c) Quina seria la mida mostral a considerar si es volgués obtenir la mateixa precisió (un marge d'error absolut de ± 5 animals) en l'estimació del nombre mitjà d'animals per explotació a cada regió? Quina seria la precisió de l'estimador del nombre mig d'animals per explotació al conjunt del territori a un nivell de confiança del 95%?

d) Avaluant el cost de l'enquesta, l'empresa detecta que el cost unitari per cada unitat mostrejada no és el mateix a les dues regions: $C_1 = 200$ i $C_2 = 300$. Quina seria la mida mostral en les regions que garanteix un cost global mínim per una variància de l'estimador de μ fixada a $1{,}276$?

e) Calculeu el cost global de les enquestes descrites a *c)* i a *d)* i compareu-los.

3.3 SOLUCIONS

1. a) $n = 121{,}70 \rightarrow 122$ persones

b) $n_h = \begin{cases} 73{,}2 \rightarrow n_1 = 73 \text{ persones} \\ 36{,}6 \rightarrow n_2 = 37 \text{ persones} \\ 12{,}2 \rightarrow n_3 = 12 \text{ persones} \end{cases}$

c) $n = 34{,}43 \rightarrow 35$ persones

$n_h = \begin{cases} 6 \rightarrow n_1 = 6 \text{ persones} \\ 9 \rightarrow n_2 = 9 \text{ persones} \\ 20 \rightarrow n_3 = 20 \text{ persones} \end{cases}$

d) $n = 35{,}18 \rightarrow 36$ persones

e) $n_h = \begin{cases} 7{,}62 \rightarrow n_1 = 8 \text{ persones} \\ 10{,}43 \rightarrow n_2 = 11 \text{ persones} \\ 17{,}96 \rightarrow n_3 = 18 \text{ persones} \end{cases}$

2. a) $n_h = \begin{cases} 106{,}11 \rightarrow n_1 = 106 \text{ persones} \\ 63{,}67 \rightarrow n_2 = 64 \text{ persones} \\ 21{,}22 \rightarrow n_3 = 21 \text{ persones} \end{cases} \longrightarrow C = 224.800$ u.m.

b) $n_h = \begin{cases} 13{,}80 \rightarrow n_1 = 14 \text{ persones} \\ 41{,}40 \rightarrow n_2 = 41 \text{ persones} \\ 36{,}80 \rightarrow n_3 = 37 \text{ persones} \end{cases} \longrightarrow C = 137.200$ u.m.

c) $n_h = \begin{cases} 16{,}71 \rightarrow n_1 = 17 \text{ persones} \\ 45{,}78 \rightarrow n_2 = 46 \text{ persones} \\ 31{,}52 \rightarrow n_3 = 32 \text{ persones} \end{cases} \longrightarrow C = 136.200$ u.m.

f) Afix. Proporcional: $n_h = \begin{cases} 55{,}56 \rightarrow n_1 = 56 \text{ persones} \\ 33{,}33 \rightarrow n_2 = 33 \text{ persones} \\ 11{,}11 \rightarrow n_3 = 11 \text{ persones} \end{cases}$

$\hat{\sigma}(\bar{x}) = 1{,}568$

$\mu \in (18{,}927; 25{,}073)$

$C = 117.600$ u.m.

Afix. de mínima variància: $n_h = \begin{cases} n_1 = 15 \text{ persones} \\ n_2 = 45 \text{ persones} \\ n_3 = 40 \text{ persones} \end{cases}$

$\hat{\sigma}(\bar{x}) = 1{,}045$

$\mu \in (19{,}952; 24{,}048)$

$C = 149.000$ u.m.

Afix. òptima: $n_h = \begin{cases} 17{,}78 \rightarrow n_1 = 18 \text{ persones} \\ 48{,}69 \rightarrow n_2 = 49 \text{ persones} \\ 33{,}52 \rightarrow n_3 = 34 \text{ persones} \end{cases}$

$\hat{\sigma}(\bar{x}) = 1{,}0499$

$\mu \in (19{,}942; 24{,}058)$

$C = 144.800$ u.m.

3. $n_h = \begin{cases} 12{,}58 \rightarrow n_1 = 13 \text{ persones} \\ 59{,}63 \rightarrow n_2 = 60 \text{ persones} \\ 27{,}79 \rightarrow n_3 = 28 \text{ persones} \end{cases}$

4. a) $\hat{p} = 68{,}75\%$

b) $e(\hat{p}) = \hat{\sigma}(\hat{p}) = 0{,}064; e_r(\hat{p}) = 9{,}3\%$

c) $p \in (58{,}27\%;\ \ 79{,}223\%)$

d) $n = 251{,}30 \rightarrow n = 252$ llars

5. a) $n_h = \begin{cases} n_1 = 80 \text{ empreses} \\ n_2 = 18 \text{ empreses} \\ n_3 = 2 \text{ empreses} \end{cases}$

b) $\bar{x} = 44{,}6$ treballadors

c) $\mu \in (38{,}999;\ \ 50{,}201)$

6. a)

ESTUDIS	*Mida mostral (m.e.p.)*	*% no han fumat mai*	*Error de mostreig*	*% de marge d'error (95,45%)*
< Primaris	762	61	0,0177	3,54
Primaris	899	48	0,0167	3,33
Batxillerat	578	33	0,0196	3,92
Superiors	242	27	0,0286	5,72
NS/NC	17	53	0,1248	24,95
TOTAL	**2498**	**46'4941**	**0,0097**	**1,94**

c) $p_{Est.Sup.} \in (21{,}28\%;\ 32{,}72\%)$

$p_{Batx.} \in (29{,}08\%;\ 36{,}92\%)$

e) $n = 608{,}33 \rightarrow n = 609$ individus f)

ESTUDIS	*Mida mostral (m.e.p.)*	$n_{f)}$
< Primaris	762	1916
Primaris	899	2262
Batxillerat	578	1455
Superiors	242	609
NS/NC	17	43
TOTAL	**2498**	**6285**

g) $e_{\propto;\, Est.Prim.} = 2{,}10\%$

h) $n = 682{,}41 \rightarrow n = 683$ individus

ESTUDIS	*Mida mostral (m.a.e., a.p.)*	*Afixació mixta* (n_h)
< Primaris	762	762
Primaris	899	899
Batxillerat	578	683
Superiors	242	609
NS/NC	17	17
TOTAL	**2498**	**2970**

Ponderacions: $\lambda_{<Prim.} = \lambda_{Prim.} = \lambda_{NS/NC} = 1$

$\lambda_{Batx.} = {}^{578}/_{683}$

$\lambda_{Sup.} = {}^{242}/_{609}$

i) $\hat{p} = \sum_h (\hat{p}_h\, \lambda_h w_h) = 46{,}494\%$

7. Afixació de Variància Mínima: $n_h = \begin{cases} 545{,}45 \rightarrow n_C = 545 \text{ cases} \\ 90{,}91 \rightarrow n_D = 91 \text{ dúplexs} \\ 363{,}64 \rightarrow n_P = 364 \text{ pisos} \end{cases}$

9. a) $e(\bar{x}) = 1{,}403$

$\mu \in (24{,}924; 30{,}426)$

b) Afixació de Variància Mínima: $n_h = \begin{cases} 13{,}47 \rightarrow n_A = 13 \text{ vivendes} \\ 13{,}81 \rightarrow n_B = 14 \text{ vivendes} \\ 12{,}72 \rightarrow n_R = 13 \text{ vivendes} \end{cases}$

$e(\bar{x}) = 1{,}288$

c) Afixació de Variància Mínima: $n = 57{,}22 \rightarrow n = 58$ llars

$$n_h = \begin{cases} n_A = 20 \text{ llars} \\ n_B = 20 \text{ llars} \\ n_R = 18 \text{ llars} \end{cases}$$

d) Afixació de Cost Mínim: $n = 57{,}72 \rightarrow n = 58$ llars

$$n_h = \begin{cases} n_A = 21 \text{ vivendes} \\ n_B = 21 \text{ vivendes} \\ n_R = 16 \text{ vivendes} \end{cases}$$

e) $C_{c)} = 603€$

$C_{d)} = 594€$

10. b) $n_h = \begin{cases} 72{,}29 \rightarrow n_1 = 72 \text{ persones} \\ 542{,}17 \rightarrow n_2 = 542 \text{ persones} \\ 385{,}54 \rightarrow n_3 = 386 \text{ persones} \end{cases}$

c) $\bar{x} = 4.554{,}217$ u.m., $\mu \in (4.461{,}374;\ 4.647{,}06)$

d) No cal fer publicitat: $T \in (370.293.985{,}6;\ 385.706.036{,}4)$

f) $n_h = \begin{cases} 180{,}56 \rightarrow n_1 = 181 \text{ persones} \\ 338{,}37 \rightarrow n_2 = 338 \text{ persones} \\ 481{,}23 \rightarrow n_3 = 481 \text{ persones} \end{cases}$

g) $e(\bar{x}) = 50{,}311$

h) $\mu \in (4.471{,}707;\ 4.636{,}727)$

i) Afixació de Mínima Variància: $n = 2.652$ individus

11. a) $\bar{x} = 1{,}696$ congressos, $\mu \in (1{,}1203;\ 2{,}2717)$

b) $\hat{p} = 58{,}596\%$, $p \in (47{,}698\%;\ 69{,}494\%)$

12. a) La millor estratificació és la del mètode B per afixació proporcional i l'A per l'afixació de mínima variància.

Afixació proporcional:

$n_{Ah} = \begin{cases} 38{,}30 \rightarrow n_{A1} = 38 \text{ pers.} \\ 96{,}54 \rightarrow n_{A2} = 97 \text{ pers.} \\ 15{,}16 \rightarrow n_{A3} = 15 \text{ pers.} \end{cases}; \ e_A(\bar{x}) = 0{,}479$

$n_{Bh} = \begin{cases} 67{,}02 \rightarrow n_{B1} = 67 \text{ pers.} \\ 74{,}20 \rightarrow n_{B2} = 74 \text{ pers.} \\ 8{,}78 \rightarrow n_{B3} = 9 \text{ pers.} \end{cases}; \ e_B(\bar{x}) = 0{,}462$

Afixació de Variància Mínima:

$n_{Ah} = \begin{cases} 21{,}18 \rightarrow n_{A1} = 21 \text{ pers.} \\ 76{,}26 \rightarrow n_{A2} = 76 \text{ pers.} \\ 52{,}56 \rightarrow n_{A3} = 53 \text{ pers.} \end{cases}; \ e_A(\bar{x}) = 0{,}342$

$n_{Bh} = \begin{cases} 49{,}14 \rightarrow n_{B1} = 49 \text{ pers.} \\ 68{,}56 \rightarrow n_{B2} = 69 \text{ pers.} \\ 32{,}30 \rightarrow n_{B3} = 32 \text{ pers.} \end{cases}; \ e_B(\bar{x}) = 0{,}369$

b) La millor estratificació és la del mètode B per l'afixació òptima:

$n_{Ah} = \begin{cases} 64{,}12 \rightarrow n_{A1} = 64 \text{ pers.} \\ 46{,}13 \rightarrow n_{A2} = 46 \text{ pers.} \\ 39{,}75 \rightarrow n_{A3} = 40 \text{ pers.} \end{cases}; \ e_A(\bar{x}) = 0{,}417$

$$n_{Bh} = \begin{cases} 55{,}07 \;\rightarrow\; n_{B1} = 55 \text{ pers.} \\ 76{,}84 \;\rightarrow\; n_{B2} = 77 \text{ pers.} \\ 18{,}11 \;\rightarrow\; n_{B3} = 18 \text{ pers.} \end{cases};\; e_B(\bar{x}) = 0{,}389$$

c) $n_A = 39{,}65 \rightarrow\; n_A = 4$

d) $n_B = 15{,}25 \rightarrow\; n_B = 1$

13. Entrevistes telefòniques per aquelles persones que disposen de telèfon: $C_2 = 18.902{,}5€$.

14. a) $e(\hat{\mu}) = 1{,}59$

b) $e(\hat{\mu}) = 1{,}06$

c) Afixació de mínima variància: $e(\hat{\mu}) = 0{,}89$

15. L'afixació més precisa és la òptima:

$$n_{h;Opt.} = \begin{cases} 600 \\ 400 \end{cases};\;\; C_{Opt.} = 7.000;\;\; \hat{\sigma}(\bar{x}) = 0{,}0423$$

$$n_{h;Prop.} = \begin{cases} 333 \\ 667 \end{cases};\;\; C_{Opt.} = 8.335;\;\; \hat{\sigma}(\bar{x}) = 0{,}044$$

Pèrdua relativa de precisió: 4,0189%

16. a) Mostreig aleatori estratificat, amb afixació proporcional.

b) $\bar{x} = 13{,}4$ punts, $\mu \in (12{,}7348;\; 14{,}0652)$

17. a) $n_{h;Prop.} = \begin{cases} 400 \\ 100 \end{cases}$

b) $\widehat{Var}(\hat{\mu}_1) = 0{,}99 \;\rightarrow e_{\propto}(\hat{\mu}_1) = 1{,}9502$

$\widehat{Var}(\hat{\mu}_2) = 15{,}84 \;\rightarrow e_{\propto}(\hat{\mu}_2) = 7{,}8007$

$\widehat{Var}(\hat{\mu}) = 1{,}2672 \;\rightarrow e_{\propto}(\hat{\mu}) = 2{,}2064$

c) $\left.\begin{array}{r} n_1 = 61{,}37 \rightarrow n_1 = 62 \\ n_2 = 239{,}96 \rightarrow n_2 = 240 \end{array}\right\} \rightarrow\; n = 302;\;\; \hat{\sigma}(\hat{\mu}) = 2{,}0935$

d) Afixació de Cost Mínim: $n = 354{,}23 \rightarrow\; n = 355$ explotacions

$$n_h = \begin{cases} 252{,}09 \rightarrow n_1 = 252 \text{ explotacions} \\ 102{,}91 \rightarrow n_2 = 103 \text{ explotacions} \end{cases}$$

e) $C_{c)} = 84.400€;\;\; C_{d)} = 81.300€$

Tema 4. MOSTREIG SISTEMÀTIC

MOSTREIG SISTEMÀTIC

4.1 EXERCICI RESOLT

El Museu d'Història d'una determinada ciutat vol estimar quina proporció de visitants arriben en transport públic. Durant un matí, un total de 20 persones han visitat el Museu, i han estat preguntades sobre aquesta qüestió en el moment de pagar l'entrada. Les persones que han fet servir el transport públic són, per ordre d'arribada, les que ocupen les posicions: $\{1, 3, 4, 5, 7, 9, 10, 12, 14, 20\}$.

Suposem ara que volguéssim fer servir una mostra sistemàtica d'1 en 5 per estimar el paràmetre d'interès.

a) Abans de fer cap càlcul, creieu que pot ser adequat utilitzar un mostreig sistemàtic enlloc d'un altre mètode de mostreig? Contesteu tenint present que la variable que estem estudiant és binària (pren els valors 0 o 1) i, per tant, caldrà treballar amb proporcions enlloc de mitjanes.

   El mostreig sistemàtic pot ésser més adient que altres tipus de mostreig en diferents casos. D'una banda, és millor que un mostreig aleatori simple quan la variabilitat intramostral és més elevada que la variabilitat total de la variable (és a dir, les mostres són molt heterogènies en el seu si, però molt homogènies entre elles). D'altra banda, el mostreig sistemàtic es demostra millor que un d'estratificat quan els estrats són molt homogenis, però molt diferents entre si.

   En el cas que ens ocupa, el fet d'escollir una mostra sistemàtica implica estudiar el transport utilitzat tant per persones que visiten el museu a primera hora del matí com per la tarda. Per tant, podem pensar que la mostra recollirà prou variabilitat i que pot superar en precisió a una m.a.s.

b) Escriviu les dades poblacionals en una taula $4x5$, que tingui les possibles mostres en columnes. Calculeu la proporció mostral de persones que arriben en transport públic per a cada mostra possible, i la proporció poblacional.

$n\ /\ k$	1	2	3	4	5	
1	1	0	1	1	1	
2	0	1	0	1	1	
3	0	1	0	1	0	
4	0	0	0	0	1	
$\hat{\boldsymbol{p}}_{\boldsymbol{j}}$	$\mathbf{1/4}$	$\mathbf{1/2}$	$\mathbf{1/4}$	$\mathbf{3/4}$	$\mathbf{3/4}$	$\hat{\boldsymbol{p}} = \mathbf{1/2}$

c) Calculeu la quasivariància poblacional, la intramostral i la intermostral.

$$S_{BS}^2 = \frac{n\sum_j(\hat{p}_j - \hat{p})^2}{k-1} = 1/4$$

$$S_{WS}^2 = \frac{\sum_{i,j}(p_{ij} - \hat{p}_j)^2}{k(n-1)} = 4/15$$

$$S^2 = \frac{N}{N-1}\hat{p}\hat{q} = 5/19$$

d) Construïu la taula ANOVA per al problema.

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	1	$k-1=4$	$1/4$
Intramostral (WS)	4	$k(n-1)=15$	$4/15$
Total	**5**	$\boldsymbol{nK-1=19}$	$\boldsymbol{5/19}$

e) Compareu la precisió de la mostra sistemàtica utilitzada, front a una mostra aleatòria simple de la mateixa mida, a partir del valor de la quasivariància intramostral.

$S_{WS}^2 > S^2 \rightarrow$ La mostra sistemàtica és més precisa que una mostra aleatòria simple de la mateixa mida.

f) Repetiu l'apartat anterior, però ara raonant a partir del valor del coeficient de correlació intramostral.

$$Var(\hat{p}) = \frac{n}{N}\sum_j(\hat{p}_j - \hat{p})^2 = 1/20$$

$$Var(\hat{p}) = \frac{N-1}{N}\frac{S^2}{n}[1+(n-1)\rho_W]$$

$$\frac{1}{20} = \frac{19}{20}\frac{5/19}{4}\ (1+3\rho_W)\ \rightarrow\ \rho_W = -1/15$$

L'obtenció d'un coeficient intramostral negatiu, reafirma que la mostra sistemàtica és més precisa que la mostra aleatòria simple.

g) Trobeu la taula ANOVA per estratificació (considereu com a fonts de variació les variacions entre estrats i dins els estrats).

n / k	1	2	3	4	5	$\hat{\boldsymbol{p}}_i$
1	1	0	1	1	1	$\boldsymbol{4/5}$
2	0	1	0	1	1	$\boldsymbol{3/5}$
3	0	1	0	1	0	$\boldsymbol{2/5}$
4	0	0	0	0	1	$\boldsymbol{1/5}$
						$\hat{\boldsymbol{p}} = \boldsymbol{1/2}$

$$S_{BSt}^2 = \frac{k\sum_i(\hat{p}_i - \hat{p})^2}{n-1} = 1/3$$

$$S_{WSt}^2 = \frac{\sum_{i,j}(p_{ij} - \hat{p}_i)^2}{n(k-1)} = 1/4$$

$$S^2 = \frac{N}{N-1}\hat{p}\hat{q} = 5/19$$

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre estrats (BSt)	1	$n-1=3$	$1/3$
Intraestrats (WSt)	4	$n(k-1)=16$	$1/4$
Total	**5**	$\boldsymbol{nK-1=19}$	$\mathbf{5/19}$

h) Trobeu el coeficient de correlació estratal. Interpreteu el resultat.

$$Var(\hat{p}) = (1-\frac{n}{N})\ \frac{S_{WSt}^2}{n}[1+(n-1)\rho_{St}]$$

$$\frac{1}{20} = (1-\frac{4}{20})\ \ \frac{1/4}{4}\ \ (1+3\rho_{St})\ \rightarrow\ \ \rho_{St} = 0$$

Un coeficient de correlació estratal igual a 0 indica que el mostreig sistemàtic és tant precís com el mostreig estratificat, amb elecció d'un element per estrat.

i) Utilitzant la següent fórmula del mostreig sistemàtic:

$$Var(\hat{p}) = (1-f)\frac{S_{WSt}^2}{n}[1+(n-1)\rho_{St}]$$

calculeu la mida mostral necessària per tenir un error absolut de mostreig de 0.16 en l'estimació de la proporció de visitants que arriben en transport públic.

$$e(\hat{p}) = 0{,}16 \rightarrow Var(\hat{p}) = 0{,}16^2;\ \ \hat{\rho}_{St} = 0$$

$$0{,}0256 = \left(1-\frac{n}{N}\right)\frac{S_{WSt}^2}{n}$$

$$0{,}0256 = \left(1-\frac{n}{20}\right)\frac{1/4}{n}\ \rightarrow\ \ n = 6{,}56 \rightarrow n = 7$$

j) Quina seria aquesta mida mostral si es treballés amb mostreig aleatori simple?

Mitjançant les fórmules de càlcul d'n pel mostreig aleatori simple:

$$n = \frac{N\,p(1-p)}{p(1-p)+e^2(N-1)} = \frac{20 * 0{,}5^2}{0{,}5^2 + 0{,}16^2(20-1)} = 6{,}79 \rightarrow n = 7$$

k) Compareu els resultats obtinguts als dos apartats anteriors, i feu un comentari per justificar-los. Trobeu alguna contradicció aparent?

Que la mateixa mida mostral garanteixi una mateixa precisió tant per un mostreig aleatori simple com per un d'estratificat sí que sembla contradictori; atès que, sota les condicions establertes, el mostreig estratificat és més precís que l'aleatori simple (mostreig sistemàtic més precís que l'aleatori simple i igual de precís que l'estratificat).

Si considerem els decimals (no arrodonim n), la conclusió és que la contradicció era només aparent, ja que amb menor mida mostral ($n_{me} = 6{,}56$), el mostreig estratificat permet assolir la mateixa precisió en l'estimador que l'aleatori simple ($n_{mas} = 6{,}79$).

4.2 EXERCICIS PROPOSATS

1. Les 45 vivendes d'un carrer es numeren de l'1 al 45 segons la seva data de construcció. Les vivendes amb una superfície de més de 100 m^2 són les que tenen els números: $\{1, 4, 5, 6, 8, 9, 10, 12, 16, 17, 19, 20, 23{,}25, 28, 29, 30, 32, 33, 34, 36, 38, 39, 42\}$. Considereu una mostra sistemàtica d'1 en 5, per estimar la proporció d'immobles més grans de 100 m^2.

a) Trobeu la taula ANOVA per aquest problema.

b) Compareu la precisió d'una mostra sistemàtica d'1 en 5 amb una mostra aleatòria simple de la mateixa mida, a partir del valor de la quasivariància intramostral.

c) Repetiu l'apartat anterior, però ara raonant a partir del valor del coeficient de correlació intramostral.

d) Trobeu la taula ANOVA per estratificació (considereu com a fonts de variació les variacions entre estrats i dins els estrats).

e) Calculeu el coeficient de correlació estratal, i comenteu si és millor el mostreig sistemàtic o l'estratificat per estimar el paràmetre.

2. Un investigador desitja determinar la quantitat d'una determinada substància que es troba a les fulles dels gira-sols d'una finca, repartits de forma natural en 10 files. El nombre total plantes és desconegut. Se sap que el terreny i les condicions ambientals són homogenis a tota la finca (penseu quina implicació té això sobre el valor de ρ_w). L'investigador decideix fer servir una mostra sistemàtica d'1 en 10.

A la taula adjunta es troben les dades sobre el contingut de la substància en qüestió a les fulles dels gira-sols mostrejats, que són, en total, $n = 400$:

Arbre	*Contingut en sucre*
1	80
2	86
3	72
…	…
398	64
399	90
400	84

Fent els càlculs, es troba que: $\sum_{i=1}^{400} x_i = 32.568$ i $\sum_{i=1}^{400} x_i^2 = 2.795.475$.

a) Estimeu el contingut mitjà de substància a les fulles dels gira-sols de la finca.

b) Trobeu els errors absolut i relatiu de l'estimació (penseu que es pot considerar que les plantes estan repartides de forma aleatòria arreu de la finca).

c) Realitzeu l'estimació mitjançant un interval de confiança al 95%.

3. Considerem una població de $N = 100$ unitats, inicialment dividida en 6 estrats. Estudiant la disposició dels elements de la població, s'arriba a la conclusió que ρ_{st} és pròxim a zero. Es desitja estimar la mitjana de la població, i la variància d'aquesta mitjana. Per tal de poder-ho fer, s'agrupen els 6 estrats en 3, i es pren una mostra de mida 2 de cadascun dels 3 estrats. Les dades són: 2 i 4, 1 i 3, 0 i 5.

a) Estimeu la mitjana de la població.

b) Estimeu la variància de l'estimador de la mitjana.

c) Construïu un interval de confiança al 90% per a la mitjana.

4. Els 20 alumnes d'una classe de secundària es numeren de l'1 al 20 segons la seva data de naixement, començant pel més jove. Els alumnes que estan a favor d'anar de viatge de final de curs a Itàlia són els que ocupen els llocs: $\{5, 7, 9, 12, 14, 15, 16, 18, 19, 20\}$. Considereu una mostra sistemàtica d'1 en 5, per a estimar la proporció d'alumnes que volen anar a Itàlia de viatge.

a) Calculeu la quasivariància i les quasivariàncies entre mostres i dins de les mostres.

b) Construïu la taula ANOVA del problema.

c) Compareu la precisió de la mostra sistemàtica utilitzada, front a una mostra aleatòria simple de la mateixa mida, a partir del valor de la quasivariància intramostral.

d) Repetiu l'apartat anterior, però ara raonant a partir del valor del coeficient de correlació intramostral.

5. Una empresa de fabricació de components per a automòbils encarrega una auditoria de qualitat per tal d'analitzar la producció manual d'un cert tipus de peces per part de 10 operaris. Per a fer-ho, s'analitza el nombre de peces produïdes per la maquinària durant els primers 3 minuts d'ésser posades en funcionament. Els resultats es mostren a la taula següent:

Nº d'operari	1	2	3	4	5	6	7	8	9	10
Nº de peces produïdes	7	8	7	5	6	9	6	7	7	8

L'estudi del nombre de peces defectuoses es duu a terme mitjançant un mostreig sistemàtic d'1 en 7 (l'ordenació de les peces és la següent: s'inicia per les peces del primer treballador i quan s'acaben, es passa al segon i, així, successivament).

Si sabem que només la primera peça que produeix cada operari és defectuosa:

a) Calculeu el coeficient de correlació intramostral i l'estratal, i comenteu-ne les conclusions que se'n deriven.

b) Estimar la variància de la proporció d'unitats defectuoses, suposant:

   i. El mostreig dut a terme és equiparable al m.a.s.

   ii. El mostreig realitzat és similar al m.a.e.

6. Una taula ANOVA per a comparar un mostreig sistemàtic i un aleatori simple dóna els valors següents:

Font de variació	*Sumes de quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	123		2,795
Intramostres (WS)			
Total	**3691**	**224**	

a) Completeu la taula.

b) Quina mida té la població?

c) Quina mida tindria cada mostra (en el mostreig sistemàtic)?

d) Quin mètode suposa menor error de mostreig: el mostreig aleatori simple o el sistemàtic? Per què?

e) Calculeu el coeficient de correlació intramostral.

7. Amb un mostreig sistemàtic d'1 en 10, s'ha extret una mostra de 100 unitats en una cadena de producció i s'ha observat si la unitat és correcta (0) o defectuosa (1). Es vol estimar l'error de mostreig, sabent que les 100 unitats s'han agrupat en parelles respecte l'ordre d'extracció amb els resultats següents:

Nombre de parelles amb resultat $(0,0) \rightarrow 46$

Nombre de parelles amb resultat $(1,0)$ o $(0,1) \rightarrow 3$

Nombre de parelles amb resultat $(1,1) \rightarrow 1$

a) Estimeu la proporció de peces defectuoses.

b) Calculeu l'error absolut de mostreig suposant que el mostreig realitzat és equiparable a un aleatori simple.

c) Calculeu l'error absolut de mostreig suposant mostreig estratificat amb selecció de dues unitats per estrat, i digueu si és més gran o més petit que a l'apartat *b)*.

8. A una conferència, hi assisteixen 200 persones que seuen en files de 10. En acabar, es vol fer un estudi sobre la seva situació laboral; en concret, el nombre d'anys que porten treballant. Per tal d'estimar-ne la mitjana, es duu a terme un mostreig sistemàtic d'1 en 10.

S'observa la posició dels assistents i es veu que els joves se situen en les darreres files de la sala, mentre que els més grans seuen a les primeres. L'ordenació dels individus és, doncs, decreixent en edat des de la primera fila fins a la vintena.

a) Un cop es disposi dels resultats, com s'haurà de calcular la variància de l'estimador de la mitjana d'anys d'experiència laboral?

Se selecciona una mostra de les possibles i s'obtenen els següents resultats:

Files	x_{h1}	x_{h2}
1 – 2	38	40
3 – 4	32	35
5 – 6	27	28
7 – 8	26	24
9 – 10	24	25
11 – 12	20	18
13 – 14	18	15
15 – 16	16	12
17 – 18	11	9
19 – 20	7	5

b) Estimeu la mitjana poblacional de forma puntual i amb un interval al 95% de confiança.

9. Es vol estimar la quantitat de taronges que maduren setmanalment en un camp de tarongers (x_i). El terreny, segons diu el pagès propietari, és uniforme; és a dir, les propietats del sòl són molt similars arreu del mateix i els arbres són de la mateixa varietat i tenen igual nombre d'anys. Els tarongers se situen en files de 10 i es duu a terme un mostreig sistemàtic que resulta en una mostra de 150 arbres.

Els resultats obtinguts es mostren a continuació:

$$\sum_{i=1}^{150} x_i = 3.428 \text{ i } \sum_{i=1}^{150} x_i^2 = 82.865$$

a) Obteniu un interval de confiança per la quantitat mitjana de taronges que maduren a la setmana a cada arbre, amb un nivell de confiança del 90%.

b) El pagès ens demana un estudi similar per un terreny de presseguers, on els arbres no tenen la mateixa edat i estan ordenats dels més vells als més joves. El mostreig dut a terme i els resultats són els mateixos que anteriorment. Creus que l'interval calculat és adequat? Justifica la resposta.

10. Es vol realitzar un mostreig sistemàtic d'1 en 3 per a estimar la mitjana de la variable X en la següent població de 9 individus:

i	1	2	3	4	5	6	7	8	9
x_i	2	6	4	3	2	4	5	7	3

a) Determineu l'espai mostral.

b) Seleccioneu la millor mostra i compareu-ne la precisió respecte a un mostreig aleatori simple.

Indicació: La millor mostra és aquella amb menor EQM.

11. Els 36 metges d'un hospital es numeren de l'1 al 36 segons l'edat, començant pel més jove. Els metges que estan descontents amb el seu sou són els que ocupen els llocs: $\{2, 4, 5, 8, 14, 15, 17, 19, 21, 27, 30, 31, 32, 36\}$. Considerem una mostra sistemàtica d'1 en 6 per estimar la proporció de metges insatisfets.

a) Construïu la taula ANOVA per al problema.

b) Compareu la precisió de la mostra sistemàtica utilitzada, front a una mostra aleatòria simple de la mateixa mida, a partir del valor de la quasivariància intramostral.

c) Repetiu l'apartat anterior, però ara raonant a partir del valor del coeficient de correlació intramostral.

d) Construïu la taula ANOVA per estratificació.

e) Trobeu el coeficient de correlació estratal. Interpreteu el resultat.

f) Utilitzant la següent fórmula del mostreig sistemàtic:

$$Var(\hat{p}) = (1-f)\frac{S_{WSt}^2}{n}[1+(n-1)\rho_{St}]$$

calculeu la mida mostral necessària per tenir un error absolut de mostreig de 0,15 en l'estimació de la proporció de metges descontents.

g) Quina seria aquesta mida mostral si es treballés amb mostreig aleatori simple?

h) Compareu els resultats obtinguts als dos apartats anteriors, i feu un comentari per justificar-los.

i) Suposem ara que els resultats poblacionals són desconeguts, i que només coneixem la informació que ens dóna la mostra sistemàtica, que recull 4 metges satisfets i 2 insatisfets. A partir del resultat que mostra que ρ_w és pròxim a zero, estimeu la variància de l'estimador de la proporció de metges insatisfets.

j) Existeix algun problema per a trobar l'interval de confiança de l'estimador de la proporció?

4.3 SOLUCIONS

1. a)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	$14/45$	4	$7/90$
Intramostral (WS)	$98/9$	40	$49/180$
Total	$\mathbf{56/5}$	$\mathbf{44}$	$\mathbf{14/55}$

b) $S_{Ws}^2 = \frac{49}{180} > S^2 = \frac{14}{55}$ → És més precís el mostreig sistemàtic.

c) $\rho_W = \frac{-3}{32}$ → És més precís el mostreig sistemàtic.

d)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre estrats (BSt)	2	8	$1/4$
Intraestratal (WSt)	$46/5$	36	$23/90$
Total	$\mathbf{56/5}$	$\mathbf{44}$	$\mathbf{14/55}$

e) $\rho_{St} = \frac{-2}{23}$ → És més precís el mostreig sistemàtic.

2. a) $\hat{\mu} = 81{,}42$

b) $e(\bar{x}) = 0{,}9005;\ e_r(\bar{x}) = 1{,}106\%$

c) $\mu \in (79{,}655;\ \ 83{,}185)$

3. a) $\hat{\mu} = 2{,}5$

b) $Var(\bar{x}) = 0{,}862$

c) Supòsit: $x \sim N \rightarrow \mu \in (0{,}629;\ 4{,}37)$

Supòsit: $x \nsim N \rightarrow \mu \in (-0{,}436;\ \ 5{,}436)$

4. a) $S^2_{Bs} = \frac{1}{4};\ S^2_{Ws} = \frac{4}{15};\ S^2 = \frac{5}{19}$

b)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	1	4	$1/4$
Intramostral (WS)	4	15	$4/15$
Total	**5**	**19**	$\mathbf{5/19}$

c) $S^2_{Ws} = \frac{4}{15} > S^2 = \frac{5}{19}$ → És més precís el mostreig sistemàtic.

d) $\rho_W = \frac{-1}{15}$ → És més precís el mostreig sistemàtic.

5. a) $\rho_W = 0{,}013$ → És més precís el mostreig aleatori simple.

$\rho_{St} = 0{,}022$ → És més precís el mostreig aleatori estratificat.

b) i. $Var(\hat{p}) = 0{,}0117$

ii. $Var(\hat{p}) = 0{,}0203$

6. a)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	123	44	2,795
Intramostral (WS)	3568	180	19,822
Total	**3691**	**224**	**16,478**

b) $N = 225$

c) $n = 5$

d) $S^2_{Ws} = 19{,}822 > S^2 = 16{,}478$ → És més precís el mostreig sistemàtic.

e) $\rho_w = -0{,}208$ → És més precís el mostreig sistemàtic.

7. a) $\hat{p} = 5\%$

b) $e(\hat{p}) = 0{,}021$

c) $e(\hat{p}) = 0{,}009$

8. b) Supòsit: $x \sim N \rightarrow \mu \in (20{,}757;\ 22{,}243)$

Supòsit: $x \nsim N \rightarrow \mu \in (19{,}912;\ 23{,}088)$

9. a) $\hat{\mu} = 22{,}853;\ \mu \in (22{,}153;\ 23{,}553)$

b) No, l'equiparació amb el mostreig aleatori estratificat és més adient en aquest cas.

10. a) $S = \{(2,3,5), (6,2,7), (4,4,3)\}$

b)

	θ_1	θ_2	θ_3
$E(\theta_i)$	$10/3$	5	$11/3$
$B(\theta_i)$	$-2/3$	1	$-1/3$
$Var(\theta_i)$	$14/9$	$14/3$	$2/3$
$\boldsymbol{EQM(\theta_i)}$	$\mathbf{2}$	$\mathbf{17/3}$	$\mathbf{7/9}$

$S_3 = \{4,4,3\} \rightarrow S_{Ws}^2 = \frac{29}{9} > S^2 = 3 \rightarrow$ Mostreig sistemàtic supera en precisió l'aleatori simple.

11. a)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre mostres (BS)	$8/9$	5	$8/45$
Intramostral (WS)	$23/3$	30	$23/90$
Total	$\mathbf{77/9}$	$\mathbf{35}$	$\mathbf{11/45}$

b) $S_{Ws}^2 = \frac{23}{90} > S^2 = \frac{11}{45} \rightarrow$ És més precís el mostreig sistemàtic.

c) $\rho_W = -0{,}063 \rightarrow$ És més precís el mostreig sistemàtic.

d)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre estrats (BSt)	$5/9$	5	$1/9$
Intraestratal (WSt)	8	30	$4/15$
Total	$\mathbf{77/9}$	$\mathbf{35}$	$\mathbf{11/45}$

e) $\rho_{St} = \frac{-1}{15} \rightarrow$ És més precís el mostreig sistemàtic.

f) $\frac{4}{8100}n^2 - \frac{1561}{32400}n + \frac{64}{225} = 0 \Longrightarrow n = 6{,}31 \rightarrow n = 7$

g) $n = 8{,}67 \rightarrow n = 9$

i) $\tilde{p}_j = \frac{1}{3}$ i $\rho_W = 0 \rightarrow Var(\hat{p}) = \frac{1}{27}$

j) $p \in (-4{,}387\%;\ 71{,}054\%) \rightarrow$ L'ús de la distribució Normal és dubtós per una mostra de 6 elements.

Tema 5. MOSTREIG PER CONGLOMERATS

MOSTREIG PER CONGLOMERATS

5.1 EXERCICI RESOLT

Una empresa productora de tabac desitja estudiar el contingut mitjà en nicotina dels 1.000 paquets de cigarretes que ha produït en la última hora. Per a fer-ho, pren una mostra de 6 paquets i, mesurant cadascuna de les 20 cigarretes del paquet, obté els següents continguts mitjans per cigarreta en cada paquet (en mg): 1,25, 1,15, 1,30, 1,18, 1,27 i 1,23. Calculant, també obté que $\sum_{ij} x_{ij}^2 = 195{,}82$.

a) Estimeu el contingut mitjà de nicotina per cigarreta.

$$\hat{\mu} = \frac{1}{n}\sum_i \bar{x}_i = 1{,}23 \text{ mg/cigarret}$$

b) Calculeu un interval de confiança al 90% per al paràmetre anterior. Comenteu si es pot suposar normalitat.

$$\hat{S}_{BC}^2 = \frac{M}{n-1}\sum_i (\bar{x}_i - \bar{x})^2 = 0{,}0632$$

$$Var(\hat{\mu}) = (1-f)\frac{\hat{S}_{BC}^2}{nM} = \left(1 - \frac{6}{1000}\right)\frac{0{,}0632}{120} \Longrightarrow e(\bar{x}) = 0{,}0229$$

$n_C = nM = 120 \longrightarrow$ Es pot suposar normalitat:

$$\mu \in \left[\bar{x} \pm z_{\propto/2} e(\bar{x})\right] = (1{,}23 \pm 1{,}64 * 0{,}0229) = (1{,}1925;\ 1{,}2675)$$

c) Estimeu també el contingut total de nicotina per paquet.

$$\hat{P} = M\hat{\mu} = 24{,}6 \text{ mg/paquet}$$

d) Calculeu l'error absolut de mostreig comès en l'estimació del total de nicotina per paquet.

$$Var(\hat{P}) = Var(M\hat{\mu}) = M^2 Var(\hat{\mu}) \Longrightarrow e(\hat{P}) = 0{,}4576$$

e) Trobeu una estimació per interval al 90% per al paràmetre anterior. Interpreteu el resultat.

$$P \in \left[\hat{P} \pm z_{\propto/2} e(\hat{P})\right] = (24{,}6 \pm 1{,}64 * 0{,}4576) = (23{,}8495;\ 25{,}3505)$$

f) Construïu la taula ANOVA per la mostra.

Font de Variació	*Suma de quadrats*	*Graus de llibertat*	*Quadrats Mitjans*
Entre conglomerats (BC)	0,316	5	0,0632
Intraconglomerats (WC)	13,956	114	0,1224
Total	$\mathbf{14,272}$	$\mathbf{119}$	$\mathbf{0,1199}$

g) Estimeu el coeficient de correlació intraconglomerats.

$$\hat{S}_0^2 = \frac{(N-1)\hat{S}_{BC}^2 + N(M-1)\hat{S}_{WC}^2}{NM-1} = 0{,}1194$$

$$\hat{\delta}_0 = \frac{\hat{S}_{BC}^2 - \hat{S}_0^2}{(M-1)\hat{S}_0^2} = -0{,}0248$$

h) Expliqueu si en aquest cas el mostreig per conglomerats és més eficient que el mostreig aleatori simple.

$\hat{\delta}_0 < 0 \longrightarrow$ És millor el mostreig per conglomerats que el m.a.s., ja que la variabilitat intraconglomerats és molt gran (alta heterogeneïtat dels conglomerats), en comparació amb la variabilitat entre conglomerats (que són força homogenis entre si).

i) Calculeu l'efecte de disseny i interpreteu el resultat.

$$Efecte\ disseny = 1 + (M-1)\hat{\delta} = 0{,}529124$$

$n_C = n_a\left[1 + (M-1)\hat{\delta}\right] = 52{,}9124\%\, n_a \longrightarrow$ A igual precisió a obtindre pels estimadors, la mida mostral per conglomerats és el $52{,}9124\%$ d'una mostra per m.a.s.

j) Retornant a l'apartat *b)*, suposem que l'interval que hem obtingut ens sembla poc precís, i que volguéssim estimar la mitjana amb una precisió (error relatiu) de l'1%. A partir de la informació que ens ha proporcionat la mostra amb la que hem treballat, calculeu quina seria la mida mostral necessària per aconseguir la millora esmentada.

$$e_r(\bar{x}) = 0{,}01 \rightarrow Var(\bar{x}) = [\bar{x} * e_r(\bar{x})]^2 = 0{,}0001513$$

$$Var(\bar{x}) = (1-f)\frac{\hat{S}_{BC}^2}{nM} \Longrightarrow n = 20{,}17 \rightarrow n = 21$$

5.2 EXERCICIS PROPOSATS

1. Un laboratori farmacèutic fabrica uns comprimits de vitamines. Per tal de dur a terme un control de qualitat, es desitja estimar el contingut mitjà de vitamina C dels comprimits, que es presenten en caixes de 20. Del total de 1.000 caixes fabricades en una dia, es decideix extraure una mostra de 5 caixes, i examinar el contingut en vitamina C de cadascun dels seus comprimits. Les mitjanes de les caixes (en mg) són: $275, 290, 280, 285, 290$, i la quasivariància mostral val 125.

a) Estimeu el contingut mitjà en vitamina C d'aquests comprimits.

b) Doneu un interval de confiança al 95% per al paràmetre anterior.

c) Construïu la taula ANOVA per la mostra.

d) Estimeu el coeficient de correlació intraconglomerats.

e) Què és millor en aquest cas, el mostreig per conglomerats o l'aleatori simple?

f) Quin és l'efecte de disseny?

g) Suposem que la mida de la mostra aleatòria simple per la precisió que volem garantir en l'estimador és de 100. Valoreu la conveniència d'usar el m.a.s. o el mostreig per conglomerats, en referència al nombre de caixes de comprimits que no resulten comercialitzables després de dur a terme el control de qualitat.

2. Es desitja dur a terme un control de qualitat sobre els productes rebuts en una comanda de 1.400 lots formats per 25 articles cadascun. Amb aquest objectiu, s'extrau una mostra sense reposició de 35 lots i s'observa que 16 no tenen articles defectuosos, 12 en tenen un i 7 en tenen dos.

a) Estimeu la proporció i el nombre total de productes defectuosos en la comanda.

b) Trobeu els errors absoluts i relatius de mostreig.

c) Realitzeu l'estimació per interval al 99%.

3. Una editorial de premsa estrangera està estudiant la possibilitat de traduir al català la revista *Fashion*. Per tal de saber quants punts de venda inicial podrien tenir, subcontracta una empresa de premsa nacional per tal que els seus repartidors enquestin els propietaris dels quioscs on reparteixen els seus diaris, preguntant-los si estarien interessats a comprar la revista *Fashion*.

El nombre total de quioscs és de 30.000 i cada repartidor s'encarrega d'atendre 30 quioscs propers. Per tal d'aprofitar els desplaçaments, es fa el següent mostreig: s'escull una mostra de 600 quioscs agrupats en 20 conglomerats de mida 30; és a dir, s'encarrega a 20 repartidors que facin l'enquesta en els seus 30 quioscs. Els repartidors s'escullen per mostreig aleatori simple. Els resultats mostren que, a cada conglomerat, el nombre de quioscs interessats a comprar la revista són:

$$\{20, 15, 12, 25, 28, 14, 16, 9, 22, 25, 23, 9, 16, 17, 24, 11, 18, 21, 16, 17\}$$

a) Estimeu el nombre total de quioscs de la població que estarien interessats a comprar la revista. Feu l'estimació puntual i per interval (al 95% de confiança).

b) Estimeu també per punt i per interval la proporció de quioscs interessats.

c) Si s'hagués estimat la proporció a partir d'un mostreig aleatori simple, s'hagués guanyat precisió? Fes els càlculs (si la quasivariància mostral és de 0,63) i raona el perquè.

d) Ordeneu la informació anterior en forma de la taula ANOVA de la mostra.

e) Per tal d'assolir un error absolut de 0,1 per a l'estimació de la proporció, calculeu la mostra que seria necessària per tal de dur a terme aquesta enquesta a partir d'un mostreig per conglomerats.

4. Una fàbrica produeix galetes dietètiques i desitja realitzar un control de qualitat per a estimar-ne el contingut calòric mitjà. El producte es presenta en caixes de 30 unitats; del total de 1.200 caixes que es produeixen en un dia, s'extrau una mostra de 5 caixes i es determina el contingut calòric de cada una d'elles. Les mitjanes de les caixes (en calories) són: 12, 11, 9, 13 i 14 i la quasivariància mostral val 12.

a) Estimeu el contingut mitjà en calories d'aquestes galetes.

b) Elaboreu un interval de confiança al 98% per al paràmetre anterior.

c) Construïu la taula ANOVA per la mostra.

d) Estimeu el coeficient de correlació intraconglomerats.

e) Què és millor en aquest cas, el mostreig per conglomerats o l'aleatori simple?

f) Quin és l'efecte de disseny?

g) Si per mostreig aleatori simple, calculem que cal prendre una mostra de 220 galetes, quina serà la quantitat de paquets a escollir si es realitza un mostreig per conglomerats?

5. Un club d'esports té 40.000 socis. Es vol saber quants socis volen un determinat canvi en els estatuts del club. Com que es té una llista dels socis classificats per criteris geogràfics, se'n decideix enquestar 800, agrupats en conglomerats de 10 socis que viuen relativament a prop.

Un cop seleccionats els conglomerats que formaran part de la mostra, es veu que el nombre de socis favorables a introduir canvis a cada conglomerat és Y_i (on Y_i és una variable que pren valors entre 0 i 10, ambdós inclosos).

Els resultats són els següents:

$$\sum_{i=1}^{n} Y_i = 370; \ \sum_{i=1}^{n} Y_i^2 = 2.536$$

a) Estimeu la proporció de socis que són favorables als canvis.

Indicació: $\sum_i Y_i = \sum_i (\sum_j A_{ij})$; $\hat{p}_i = \frac{\sum_j A_{ij}}{M}$

b) Estimeu el nombre total de socis que són favorables als canvis.

c) Construïu l'interval de confiança al 95% per a la proporció i el total.

d) Els directius creuen que hi ha prou socis a favor i tiraran endavant el canvi als estatuts si el total de socis que volen els canvis és de 22.000. Què els podeu dir?

e) Elaboreu la taula ANOVA per la mostra.

f) Calculeu el coeficient de correlació intraconglomerats i comenteu-ne els resultats.

6. Un centre comercial vol premiar als seus clients més fidels; per això, es decideix regalar una cistella de Nadal amb un assortiment de productes per a les festes. La companyia es vol assegurar que aquest obsequi vagi a parar realment als millors clients; per tant, la direcció de màrqueting considera que el criteri òptim per decidir els afortunats és realitzar una enquesta. L'objectiu d'aquesta és conèixer els clients que realitzen la totalitat de les compres en el centre. Com que els clients es tenen classificats per codi postal de residència, s'opta per enquestar-ne a 1.200, agrupats en conglomerats de 25. El centre comercial compta amb 25.000 clients que han demanat el carnet per a poder tenir descomptes i participar en ofertes especials.

Aplicant mostreig aleatori simple sense reposició, se seleccionen 48 conglomerats i, a cadascun, s'observa el nombre de clients que no realitzen la totalitat de les compres en el centre comercial (x_i):

$$\sum_{i=1}^{48} x_i = 625; \ \sum_{i=1}^{48} x_i^2 = 9.426$$

a) Estimeu la proporció de socis que no realitzen la totalitat de les compres en el centre comercial.

b) Estimeu el nombre total de socis que no realitzen la totalitat de les compres en el centre comercial.

c) Calculeu la quasivariància mostral entre conglomerats.

d) Els directius de màrqueting creuen que aquest estudi els permetrà decidir si han de fer més publicitat del centre. En faran en el cas que la proporció de gent que no realitza totes les seves compres en el centre comercial sigui molt elevada. En aquest cas, quin és l'interval de confiança al 95% per a la proporció de socis que no realitzen la totalitat de les compres en el centre?

e) Es considerarà que és necessari dur a terme una nova campanya de publicitat del centre en el cas que el nombre de socis que també compren en altres centres sigui superior als 14.000 clients. S'haurà de fer una nova campanya publicitària?

7. Al llarg d'un estudi, l'investigador treballa amb una població estructurada en 105 conglomerats de 25 individus:

a) Observa que la variable d'interès dels elements de la població no és independent amb la de la resta d'elements del mateix conglomerat, sinó que es produeix una correlació negativa entre les mateixes. En aquesta situació, dedueix que el mostreig per conglomerats serà més precís que un aleatori simple amb la mateixa mida mostral. Esteu d'acord amb aquesta conclusió? Justifiqueu la resposta.

b) Observa que les mitjanes i variàncies dels diferents conglomerats són molt similars. En aquesta situació, dedueix que el mostreig per conglomerats serà més precís que un d'aleatori estratificat, amb afixació proporcional. Esteu d'acord amb aquesta conclusió? Justifiqueu la resposta.

8. Un estudiant està valorant a quina residència d'estudiants vol viure. El seu criteri per a considerar que una residència és adequada és que un mínim de 3 de cada 4 estudiants de la residència tinguin una mitjana en les seves notes igual o superior a 3, en una escala del 0 al 5.

Per a avaluar la residència X, que té 100 habitacions amb 4 estudiants cadascuna, duu a terme un mostreig per conglomerats consistent en escollir aleatòriament 5 habitacions i enquestar-ne als quatre ocupants. Les notes dels estudiants enquestats són les següents:

	1	2	3	4	5
1	3,08	2,36	2,00	3,00	2,68
2	2,60	3,04	2,56	2,88	1,92
3	3,44	3,28	2,52	3,44	3,28
4	3,04	2,68	1,88	3,64	3,20

Creieu que la residència és adequada, segons els seus criteris?

9. Un pagès disposa d'un terreny de $30.000\ \text{km}^2$ en una zona boscosa dedicada al cultiu de pins. Es divideix l'esmentada regió en 150 zones d'igual superficie per tal d'estudiar les diferents varietats de bolets que hi creixen. A partir d'una mostra aleatòria de 20 de les zones, es fa un recompte dels diferents tipus de bolets que s'hi troba:

$$\{5, 12, 18, 3, 14, 9, 16, 7, 10, 8, 16, 12, 13, 7, 11, 8, 4, 13, 10, 16\}$$

Elaboreu una estimació puntual i per interval (al 95% de confiança) del nombre mitjà de varietats de bolets que podem esperar trobar en una de les zones del terreny.

10. Una colònia de formigues està formada per 220 conglomerats de 45 unitats cadascun. S'extreu una mostra aleatòria de 5 conglomerats per a l'estudi i es calcula la proporció de formigues que compleixen determinada característica en relació a la seva mida: $\{0{,}92;\ 0{,}88;\ 0{,}89;\ 0{,}95;\ 0{,}91\}$.

a) Estimeu el nombre total de formigues que compleixen la característica.

b) Calculeu l'error absolut i relatiu de l'estimador anterior.

c) Estimeu el coeficient de correlació intraconglomerats, si la quasivariància mostral és de 0,82.

11. Sobre una població de 1.000 llars formades per parelles amb dos fills majors de 16 anys, se selecciona una mostra de 10 vivendes, a les quals es pregunta sobre el seu consum d'un determinat producte. Si $A_{ij} = 1$ quan l'individu j de la llar i compra el producte, s'obtenen els següents resultats: $\sum_j A_{ij} = \{2, 2, 4, 3, 1, 1, 3, 4, 0, 3\}$.

a) Estimeu la proporció de persones que no compren el producte.

b) Estimeu les quasivariàncies total, entre conglomerats i intraconglomerats.

c) Elaboreu la taula ANOVA per la mostra.

d) Estimeu la correlació intraconglomerats. Comenteu els resultats.

e) Quin nombre de conglomerats cal seleccionar si volem que l'error relatiu en l'estimació de la proporció sigui de l'10%? Compareu el resultat amb la mida mostral necessària en cas de mostreig aleatori simple.

5.3 SOLUCIONS

1. a) $\bar{x} = 284$ mg

b) $\mu \in (278{,}23; 289{,}77)$

c)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre conglomerats (BC)	3400	4	850
Intraconglomerats (WC)	8975	95	94,474
Total	**12375**	**99**	**125**

d) $\hat{\delta} = 0{,}2857$

e) El mostreig aleatori simple és més precís.

f) $\left[1 + (m-1)\,\hat{\delta}\,\right] = 6{,}4283$

g) $n_{mas} = 100 \rightarrow n_{mc} = 643,\ n = 33$. El m.a.s. implica que 100 caixes de comprimits resulten no comercialitzables, mentre que el mostreig per conglomerats redueix aquesta xifra a 33 caixes.

2. a) $\hat{p} = 2{,}9714\%;\ \hat{A} = 1040$ peces

b) $e(\hat{p}) = 0{,}0052;\ e_r(\hat{p}) = 17{,}5266\%$

$e(\hat{A}) = 182{,}275;\ e_r(\hat{A}) = 17{,}5266\%$

c) $p \in (1{,}6272\%;\ 4{,}3156\%);\ A \in (569{,}73;\ 1510{,}27)$

3. a) $\hat{A} = 17.900$ quioscs

$e(\hat{A}) = 1216{,}5 \rightarrow A \in (15.515{,}66;\ 20.284{,}34)$

b) $\hat{p} = 59{,}6667\%$

$e(\hat{p}) = 0{,}0405 \rightarrow p \in (51{,}7189\%;\ 67{,}6145\%)$

c) $\hat{\delta} = 0{,}0206 \rightarrow$ El mostreig aleatori simple és més precís que el mostreig per conglomerats.

d)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre conglomerats (BC)	19,1267	19	1,0667
Intraconglomerats (WC)	358,2433	580	0,6177
Total	**377,37**	**599**	**0,63**

e) $n_a = 24{,}41 \rightarrow n_a = 25;\ n_c = 40 \rightarrow n = 2$

4. a) $\hat{\mu} = 11{,}8$

b) $\mu \in (9{,}8;\ 13{,}8)$

c)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre conglomerats (BC)	444	4	111
Intraconglomerats (WC)	1344	145	9,269
Total	**1788**	**149**	**12**

d) $\hat{\delta} = 0{,}2679 \rightarrow$ El mostreig aleatori simple és millor que el mostreig per conglomerats.

f) $Ef.Disseny = 8{,}769$

g) $n_c = 1929{,}18 \rightarrow n_c = 1930 \Longrightarrow n = 65$

5. a) $\hat{p} = 46{,}25\%$

b) $\hat{A} = 18.500$ socis

c) $p \in (39{,}241\%;\;\; 53{,}259\%)$

$A \in (15.697;\;\; 21.304)$

e)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre conglomerats (BC)	82,475	79	1,044
Intraconglomerats (WC)	116,4	720	90,1617
Total	**198,875**	**799**	**0,2489**

f) $\hat{\delta} = 0{,}3549$ → El mostreig aleatori simple és millor que el mostreig per conglomerats.

6. a) $\hat{p} = 52{,}0833\%$

b) $\hat{A} = 13.020{,}83$ clients

c) $S_{BC}^2 = 1{,}0962$

d) $p \in (46{,}3035\%;\;\; 57{,}8631\%)$

e) $A \in (11576{,}31;\;\; 14.465{,}35) \rightarrow$ No cal fer campanya publicitària.

8. $\propto = 5\% \rightarrow$ Residència adequada: $p \in (23{,}8412\%;\;\; 76{,}1588\%)$

$\propto = 10\% \rightarrow$ Residència no adequada: $p \in (28{,}112\%;\;\; 71{,}888\%)$

9. $\hat{\mu} = 10{,}6;\;\; \mu \in (8{,}8619;\;\; 12{,}3381)$

10. a) $\hat{A} = 9.009$ unitats

b) $e(\hat{A}) = 119{,}864$

c) $\hat{\delta} = -0{,}0218$

11. a) $\hat{p} = 42{,}5\%$

b) $S^2 = 0{,}2506;\;\; S_{BC}^2 = 0{,}4472;\;\; S_{WC}^2 = 0{,}1916$

c)

Font de variació	*Suma de Quadrats*	*Graus de llibertat*	*Quadrats mitjans*
Entre conglomerats (BC)	4,025	9	0,4472
Intraconglomerats (WC)	5,748	30	0,1916
Total	**9,7734**	**39**	**0,2506**

d) $\hat{\delta} = 0{,}2502$ → El mostreig aleatori simple és millor que el mostreig per conglomerats.

e) $n_a = 90{,}99 \rightarrow n_a = 91;\;\; n_c = 161 \rightarrow n = 41$

Annex A. EXERCICIS EN SPSS I R

EXERCICIS EN SPSS I R

A.1 MOSTREIG EN SPSS I R

Com extraure mostres en SPSS?

A continuació, s'expliquen els procediments necessaris per a l'extracció de mostres aleatòries simples, estratificades o per conglomerats amb el programa estadístic SPSS. El procediment és força simple, degut a l'existència d'un assistent per al mostreig, que es troba als menús mitjançant la ruta *Analizar / Muestras Complejas / Seleccionar una muestra*.

Malgrat que l'assistent permet, sense excessives dificultats, el disseny de mostres complexes (per exemple, polietàpiques), tot seguit es detallen els procediments per a l'extracció de mostres simples (m.a.s., m.a.e. i mostreig per conglomerats).

Un cop s'accedeix a l'assistent, els passos a seguir són els següents:

1. A la pantalla inicial, es pregunta per les opcions relatives a l'arxiu del pla de mostreig. El primer cop que s'extrau una mostra, cal crear un pla de mostreig (opció *Diseñar una muestra* i seleccionar una carpeta on guardar l'arxiu d'SPSS que recollirà els detalls de la mostra).

   Si, en canvi, es vol tornar a extreure una mostra segons les especificacions guardades en un pla de mostreig existent, cal escollir l'opció *Editar un diseño muestral* i seleccionar l'arxiu creat anteriorment.
2. A continuació (secció *Variables del diseño*), es poden escollir les variables d'estratificació (en el cas de mostreig estratificat) o les definitòries dels conglomerats (per al mostreig per conglomerats). En cas de mostreig aleatori simple, no caldrà fer cap acció en aquest moment.
3. Al pas *Método de muestreo*, s'escull el mètode que es vol usar per a extraure la mostra: *Muestreo aleatorio simple*, en tots els casos (és obvi en el m.a.s., però també en el m.a.e., on les unitats mostrejades de cada estrat s'escullen aleatòriament, i en el mostreig per conglomerats, on els conglomerats que formaran la mostra també s'extrauen segons aquest mètode).
4. La secció *Tamaño de la muestra* serveix per a especificar la mida mostral desitjada.
   a. En el cas de mostreig aleatori simple, s'especifica n al camp *Valor*.
   b. En el mostreig aleatori estratificat, s'escull l'opció *Valor* si es tracta d'una afixació uniforme o l'opció *Valores desiguales para los estratos* en la resta d'afixacions. En el darrer cas, la finestra que apareix en seleccionar el botó *Definir* servirà per a introduir les mides mostrals estratals.
   c. Per al mostreig per conglomerats, al camp *Valor*, s'especificarà el nombre de conglomerats que han d'integrar la mostra.
5. A *Variables de resultado* es pot seleccionar si es vol guardar altres variables derivades del mostreig. Habitualment, en dissenys mostrals simples, no és necessari.
6. A continuació, es passa a *Opciones de selección*, on caldrà especificar el valor llavor en base al qual s'extraurà la mostra.

Aquesta opció permet garantir l'obtenció de la mateixa seqüència de nombres aleatoris quan es treballa amb un programa informàtic. Com que l'extracció d'una mostra es basa en la generació d'*n* nombres aleatoris en una població de mida *N*, si no s'especifica un valor llavor abans d'extraure la mostra d'interès, resultarà impossible replicar l'operació obtenint la mateixa mostra, ja que cada cop que n'extraiem una, aquesta serà diferent.

L'SPSS té l'inconvenient que l'arxiu corresponent al pla de mostreig no emmagatzema el valor llavor. D'aquesta manera, cada cop que es vol treballar en base a un pla de mostreig prèviament dissenyat (es crea a *1.* i s'emmagatzema al pas *7.*), caldrà especificar en aquest pas el valor llavor que s'utilitza per a la mostra en qüestió.

7. Finalment, la secció *Archivos de resultados* permetrà escollir on es guardarà la base de dades mostral (i sota quin nom) i el darrer pas (*Finalización del Asistente de Muestreo*), si es vol guardar les especificacions al pla de mostreig creat o enganxar-ho a la sintaxi.

En finalitzar l'Assistent tal com es detalla, s'obté una base de dades de mida *n*, que inclou la mostra, amb les opcions especificades.

Com extraure mostres en R?

Per a dur a terme l'extracció d'una mostra aleatòria simple, estratificada o per conglomerats mitjançant el paquet estadístic R, cal tindre en compte les quatre funcions següents:

1. *Funció* `set seed()`*, l'especificació d'un valor llavor:*

   Aquesta funció serveix per a establir un valor llavor en l'extracció de la mostra que garanteixi l'obtenció de la mateixa mostra cada cop que executem el procediment.

   La única opció rellevant per a aquesta funció és especificar una xifra en l'interval ±2.147.483.647, que serà el valor llavor a especificar cada cop que vulguem treballar amb les mateixes dades mostrals.

2. *Funció* `order()`*, per a ordenar la base de dades mostral:*

   Quan l'objectiu sigui l'extracció d'una mostra aleatòria estratificada o per conglomerats, l'R requereix que la base de dades poblacional estigui ordenada segons la variable definitòria dels estrats o dels conglomerats.

   L'ús de la funció `order()` en aquest context es fa de la següent manera:

   ```
   base_dades <- base_dades[order(base_dades$variable),]
   ```

   on *base_dades* és el nom de la base de dades poblacional amb què es treballa i *variable*, el nom de la variable que defineix els estrats o conglomerats en base als quals es vol extreure la mostra.

3. *L'extracció de la mostra:*

   Les funcions necessàries per a extraure la mostra no s'inclouen als paquets base d'R. Així, caldrà instal·lar el paquet `sampling` i carregar-lo.

   a) *Funcions* `strata()` *i* `getdata()`*, per al mostreig aleatori simple i estratificat.*

   La sintaxi bàsica de les funcions és la següent:

```
nom = strata(base_dades, stratanames=c("variable"),
             size=n, method="srswor", description=T)
mostra <- getdata(base_dades, nom)
```

   on cal substituir:

   - *nom* (variable que guarda un vector que indica les observacions que s'inclouran a la mostra) pel nom que es vulgui per aquesta variable;
   - *base_dades*, pel nom utilitzat per anomenar la base de dades poblacional;
   - *variable*, pel nom de la variable que defineix l'estratificació de la mostra. Per a mostres aleatòries simples, s'utilitza l'opció: `stratanames=NULL`;
   - *n*, per la mida mostral. En el cas de les mostres aleatòries estratificades, *n* s'ha de substituir per un vector que inclogui les mides mostrals estratals desitjades: $c(n_1, n_2, ..., n_k)$;
   - *mostra*, pel nom que se li vol donar a la base de dades corresponent a la mostra resultant.

   El resultat de l'execució de l'anterior sintaxi serà que la mostra aleatòria simple o estratificada quedarà emmagatzemada sota el nom *mostra*, una base de dades de mida *n* amb les mateixes variables incloses a la *base_ dades* poblacional.

   b) *Funcions* `cluster()` *i* `getdata()`*, per al mostreig per conglomerats.*

   La sintaxi bàsica de les funcions necessàries és la següent:

```
nom = cluster(base_dades, clustername=c("variable"),
              size=n, method="srswor", description=T)
mostra <- getdata(base_dades, nom)
```

   on la interpretació de la majoria d'opcions és la mateixa que en el cas anterior, amb les excepcions següents:

   - *variable*, cal substituir-ho pel nom de la variable que defineix els conglomerats;
   - *n*, per la mida mostral, és a dir, el nombre de conglomerats a incloure a la mostra.

   De forma similar, la mostra per conglomerats de la mida desitjada quedarà emmagatzemada sota el nom *mostra*.

A.2 EXERCICIS PROPOSATS

1. La base de dades *museu.xls* conté informació poblacional sobre els *N* visitants d'un museu durant un any. Per a cada visitant, s'ha enregistrat l'hora d'entrada i la de sortida, i també el dia de la setmana en què ha dut a terme la visita. Les variables contingudes en el fitxer són *ID* (identificador de la persona), *durada* (durada de la visita en minuts), *diaset* (dia de la visita; 1 dilluns, 2 dimarts, 3 dimecres, 4 dijous, 5 divendres, 6 dissabte, 7 diumenge) i *obra* (declaren quina obra els ha agradat més d'un total de 5; Obra n.1, Obra n.2, Obra n.3, Obra n.4, Obra n.5).

Podeu resoldre aquest exercici usant l'SPSS o l'R. Mostreu els resultats amb tres decimals.

a) Estudi poblacional. Responeu les preguntes següents:
   i. Nombre de persones que han visitat el museu durant l'any.
   ii. Durada mitjana de la visita al museu.
   iii. Percentatge de persones que hi són més de 60 minuts.
   iv. Els tres paràmetres anteriors, segons el dia de la setmana.

b) Extracció d'una mostra aleatòria simple d'individus:
   i. Calculeu la mida mostral necessària per a garantir que l'estimació que s'obtindrà de la mostra tingui un error relatiu inferior a l'1% (tant per l'estimació de la durada mitjana com per l'estimació de la proporció d'estades de més d'una hora).

      Per a l'estimació de la proporció, cal dir que disposem d'estudis previs que situen la proporció de visites de més d'una hora al voltant del 82,57%.

      Indicació: Calculeu les mides mostrals necessàries per a l'estimació de la mitjana i per a la proporció, i escolliu aquella n que garanteixi que s'obtindrà un error relatiu inferior o igual a l'esmentat en l'estimació d'ambdós paràmetres.
   ii. Preneu la mostra aleatòria simple, utilitzant el valor llavor 5476.

c) Estudi de la mostra aleatòria simple. Responeu les preguntes següents:
   i. Nombre d'individus mostrejats.
   ii. Fracció de mostreig en percentatge.
   iii. Calculeu els estimadors mostrals pels paràmetres recollits a la taula següent.
   iv. Calculeu també l'error mostral i l'interval de confiança (IC) al 98%.
   v. Els errors relatius obtinguts en les tres estimacions anteriors són els previstos? Justifiqueu la resposta.

	Resultat poblacional	**Resultat mostral**	**Error mostral**	**Error relatiu**	**IC - Límit inferior, al 98%**	**IC - Límit superior, al 98%**
Durada mitjana de la visita						
Proporció de visites de durada superior a 1 hora						
Proporció de visites que prefereixen la Obra 3						

d) Extracció d'una mostra aleatòria estratificada d'individus:
   i. La mida mostral de la mostra aleatòria estratificada serà la mateixa que la de la m.a.s. Realitzeu l'afixació proporcional, considerant una estratificació segons el dia de la setmana en què es visita el museu.

      Indicació: La mostra de visitants a prendre en dilluns resulta de 1,199 individus. Per a facilitar els càlculs posteriors, arrodonim a una mida mostral per dilluns de 2 persones (en comptes d'arrodonir a 1, com correspondria).
   ii. Torneu a la base de dades poblacional i preneu la mostra aleatòria estratificada, utilitzant el valor llavor 5476.

e) Estudi de la mostra aleatòria estratificada. Per cada dia de la setmana (també totals) i només pels individus seleccionats a la mostra aleatòria estratificada, responeu les preguntes següents:
   i. La durada mitjana de la visita.
   ii. Els errors mostral i relatiu.
   iii. L'interval de confiança al 98%.
   iv. Escriviu els valors poblacionals que heu calculat a l'apartat *a)*.
   v. Observeu si els paràmetres poblacionals estan realment continguts als IC que heu trobat. Escriviu Sí / No a la columna corresponent.
   vi. Completeu la filera de totals per la mostra m.a.e.
   vii. Comenteu els resultats, centrant el vostre interès en els intervals de confiança resultants i els errors relatius obtinguts (es compleix l'objectiu fixat en referència a l'error relatiu en calcular la mida mostral?, de què està en funció l'amplitud dels intervals de confiança?,...).

	Durada mitjana estimada	**Error mostral**	**Error relatiu**	**Límit inferior IC (98%)**	**Límit superior IC (98%)**	**Durada mitjana poblacional**	**IC conté la mitjana pobl.?**
Dilluns							
Dimarts							
Dimecres							
Dijous							
Divendres							
Dissabte							
Diumenge							
TOTAL m.a.e.							

f) Estudi de la mostra aleatòria estratificada. Per a cada dia de la setmana (també totals) i només pels individus seleccionats a la mostra aleatòria estratificada, responeu les preguntes següents:
   i. El % de visitants que tarden més de 60 minuts a visitar el museu.
   ii. Els errors mostral i relatiu.
   iii. L'interval de confiança al *98%*.
   iv. Escriviu els valors poblacionals que heu calculat a l'apartat *a)*.
   v. Observeu si els paràmetres poblacionals estan realment continguts als IC que heu trobat. Escriviu Sí / No a la columna corresponent.
   vi. Completeu la filera de totals per la mostra m.a.e.
   vii. Comenteu els resultats, centrant el vostre interès en els intervals de confiança resultants i els errors relatius obtinguts (es compleix l'objectiu fixat en referència a l'error relatiu en calcular la mida mostral?, de què està en funció l'amplitud dels intervals de confiança?,...).

	% estimat d'estades $> 1h$	Error mostral	Error relatiu	Límit inferior IC (98%)	Límit superior IC (98%)	% pobl. d'estades $> 1h$	IC conté el % pobl.?
Dilluns							
Dimarts							
Dimecres							
Dijous							
Divendres							
Dissabte							
Diumenge							
TOTAL m.a.e.							

g) Estudi de la mostra aleatòria estratificada. Calculeu la proporció de visites que prefereixen l'obra 3 (també l'error mostral i l'IC al 98%).

	Resultat poblacional	Resultat mostral	Error mostral	Error relatiu	IC - Límit inferior, al 98%	IC - Límit superior, al 98%
Proporció de visites que prefereixen la O3						

h) Comparació d'eficiències segons el mètode de mostreig. Per tal de comparar els resultats del m.a.s. i del m.a.e., responeu les següents preguntes:

i. Quin mètode de mostreig és més precís per a estimar la mitjana de la durada de les visites? I per la proporció de visites que superen una hora de durada? Com ho deduïu? Per què creieu que és així?

ii. Quin mètode de mostreig és més precís per a estimar la proporció dels que prefereixen l'obra O3? Per què creieu que és així?

iii. Escriviu una conclusió general.

Opcional: *Utilitzant el paquet estadístic R, **r**epetiu l'exercici utilitzant el valor llavor 1753 en l'extracció de les mostres. Les conclusions que obteniu són les mateixes?*

2. L'arxiu *salaris.xls* conté informació poblacional relativa als 34.030 individus assalariats i treballant a jornada completa que viuen a una ciutat nord-americana (són dades de l'any 2002).

La base de dades inclou les variables *ID* (identificació de l'individu), *salari* (salari anual net, en $) i *nacionalitat* (una variable fictícia, que pren el valor 1 si l'individu va néixer i actualment resideix al mateix estat, valor 2 si l'individu és nascut a EE.UU. però no al mateix estat on viu actualment, valor 3 si es tracta d'un immigrant nascut en un país de la UE, Canadà o Japó, i valor 4, per als immigrants no nascuts en aquests països).

Podeu resoldre aquest exercici usant l'SPSS o l'R. Mostreu els resultats amb tres decimals.

a) En base a les dades poblacionals, calculeu els següents paràmetres poblacionals, tant pel total de la població com en funció de la nacionalitat de l'individu: salari mitjà i proporció de treballadors amb sou superior als 40.000$ anuals.

A partir d'aquest moment, farem l'estudi basant-nos en diferents tipus de mostres.

Indicació: Utilitzeu el valor llavor 9812 (en R) o 7689(en SPSS) per a l'extracció de totes les mostres.

b) Mostra aleatòria simple de mida $n = 2500$. Estimeu el salari mitjà de la població assalariada a jornada completa i la proporció de treballadors amb salari superior als 40.000\$. Obteniu un interval de confiança al 95% per ambdues estimacions.

c) Mostra aleatòria estratificada de mida $n = 2500$, amb afixació proporcional, essent la nacionalitat de l'individu la variable d'estratificació. Estimeu el salari mitjà de la població assalariada a jornada completa i la proporció de treballadors amb salari superior als 40.000\$ (tant pel total de la mostra com pels 4 estrats). Obteniu un interval de confiança al 95% per totes les estimacions.

d) Mostra aleatòria estratificada de mida $n = 2500$, amb afixació de variància mínima, essent la nacionalitat de l'individu la variable d'estratificació. Estimeu el salari mitjà de la població assalariada a jornada completa i la proporció de treballadors amb salari superior als 40.000\$ (tant pel total de la mostra com pels 4 estrats). Obteniu un interval de confiança al 95% per totes les estimacions.

e) Compareu els resultats obtinguts en els apartats anteriors. En particular, comenteu si els IC calculats inclouen els paràmetres poblacionals, compareu els errors relatius de les estimacions i extraieu una conclusió sobre quin és el mètode de mostreig més adequat en cada cas. Observeu alguna contradicció?

3. Una empresa es dedica a organitzar exàmens de xinès arreu del món, per tal de certificar un nivell mínim de domini d'aquesta llengua. A Catalunya, examinen un total de 650 alumnes cada any, en una única convocatòria.

Es vol estudiar la nota mitjana i la proporció d'aprovats (la nota per aprovar és, igual que aquí, un 5) durant l'any 2009 (darreres dades de què es disposa). L'investigador es planteja un mostreig sistemàtic d'1 en 25, ja que els estudiants estan asseguts en vint-i-cinc columnes en fer l'examen (tal com es mostra a la taula de l'arxiu notes.xls, al full *població*).

Per a dur a terme l'estudi, imagineu que l'investigador es pot trobar en tres situacions diferents:

a) Per a fer l'examen, els alumnes es numeren (i, per tant, estan asseguts) per ordre alfabètic segons els seus cognoms.

   i. Justifiqueu si és raonable que l'investigador faci el supòsit de que el coeficient de correlació intramostral val zero ($\rho_w \approx 0$).

   Per a fer el mostreig, s'extreu un nombre aleatori entre 1 i 25 i resulta la xifra 3. Per tal de dur a terme l'estudi, l'investigador demana les notes dels següents estudiants: $3,\ 28,\ 53,\ 78,\ \ldots, 603$ i 628; l'arxiu *notes.xls*, al full *a)*, inclou aquestes 26 notes.

   ii. A partir d'aquestes dades i sota el supòsit fet a *i)*, ompliu la següent taula:

	Estimació	**Error mostral**	**Error relatiu**	**IC - Límit inferior, al 98%**	**IC - Límit superior, al 98%**
Nota mitjana de l'examen					
Proporció d'aprovats					

b) Suposem ara que els alumnes s'asseuen per fileres a partir d'una ordenació que fa l'empresa organitzadora de l'examen. L'ordre que aquesta segueix és creixent segons el nombre d'anys que l'alumne porta estudiant xinès (dada que obtenen a partir del formulari d'inscripció a l'examen).

   i. Justifiqueu si és raonable que l'investigador faci el supòsit que el coeficient de correlació estratal val zero ($\rho_{St} \approx 0$).

   En base a l'anterior informació, l'investigador decideix equiparar el mostreig sistemàtic amb l'aleatori estratificat i, per això, considera que la població (els estudiants de l'examen) està distribuïda d'una forma diferent: la nova estructura es mostra al full *m.a.e.* de l'arxiu *notes.xls*. Per tal d'extraure la mostra, genera dos nombres aleatoris entre 1 i 50 mitjançant un programa informàtic: el 7 i el 28. Finalment, demana a l'empresa les notes dels estudiants amb ID 7, 28, 57, 78, 107, …, 557, 578, 607 i 628; l'arxiu *notes.xls*, al full *b)*, inclou les 26 notes.

   ii. A partir d'aquestes dades, ompliu la següent taula:

   Indicació: Recordeu que una proporció és la mitjana d'una variable binària; per tant, es poden aplicar les mateixes fórmules que per al càlcul de la mitjana d'una variable no binària.

	Estimació	**Error mostral**	**Error relatiu**	**IC - Límit inferior, al 98%**	**IC - Límit superior, al 98%**
Nota mitjana de l'examen					
Proporció d'aprovats					

c) Finalment, suposem que els alumnes estan ordenats segons l'acadèmia a partir de la qual s'apunten a l'examen, ja que només s'hi pot accedir a partir de determinades escoles d'idiomes.

   i. Justifiqueu si és raonable que l'investigador no pugui fer el supòsit de $\rho_w \approx 0$ ni de $\rho_{St} \approx 0$, si es troba en aquest cas.

   En base als criteris d'ordenació dels alumnes per part de l'entitat organitzadora, l'investigador es decanta per utilitzar el mètode de les mostres interpenetrants, per $t = 2$. De nou, extreu dos nombres aleatoris entre 1 i 25: el 5 i el 18. Per tal d'eliminar qualsevol possible efecte que tingui l'ordenació dels estudiants segons l'acadèmia de xinès on estudiaven, decideix extreure les observacions que s'inclouran a la mostra de forma aleatòria: genera 13 xifres aleatòries entre 1 i 26, que seran les observacions mostrals per la columna 5; i 13 nombres més, per la columna 18. Finalment, la mostra resultant està formada per les següents observacions: 30, 43, 55, 93, 118, 143, 155, 205, 218, 230, 280, 318, 330, 343, 405, 443, 468, 480, 505, 518, 543, 555, 605, 618, 630 i 643.

   Així, demana a l'empresa les notes dels estudiants en qüestió per tal de realitzar l'anàlisi; l'arxiu *notes.xls*, al full *c)*, inclou les 26 notes.

ii. A partir d'aquestes dades, ompliu la següent taula:

Indicació: Recordeu que una proporció és la mitjana d'una variable binària; per tant, es poden aplicar les mateixes fórmules per al càlcul de la mitjana d'una variable, independentment que sigui binària o no.

	Estimació	**Error mostral**	**Error relatiu**	**IC - Límit inferior, al 98%**	**IC - Límit superior, al 98%**
Nota mitjana de l'examen					
Proporció d'aprovats					

d) Valoreu els resultats obtinguts a l'apartat *c)*, en comparació amb els de les dues seccions anteriors. És raonable la pèrdua de precisió que s'observa, o es tracta d'un resultat casual?

4. Una empresa de tecnologia ha de satisfer una comanda mensual d'un determinat tipus de xip informàtic a diferents empreses. La comanda consisteix en 50 lots de 20 xips cadascun.

Un equip de treballadors del Departament de Qualitat s'encarrega, mensualment, d'avaluar si s'estan complint les característiques tècniques que ha de satisfer la comanda. Per això, realitzen un mostreig de les peces per conglomerats, per tal de certificar si, amb una certesa del 98%, es compleix el següent:

- La longitud mitjana dels xips és superior a 24′5 mm.
- La proporció d'unitats de menys de 24′5 mm no és superior al 25%.

Per a realitzar el mostreig en base al qual es controla la qualitat d'aquesta comanda, l'equip selecciona aleatòriament 6 lots d'unitats. D'aquesta manera, disposen de 120 unitats en base a les quals realitzar les proves pertinents per a veure si es compleixen els requisits contractuals.

L'arxiu d'Excel *qualitat.xls* recull les dades que haureu d'utilitzar. El full *població* inclou una representació de la comanda (50 lots de 20 peces cadascun); i al full *mostra*, hi podeu trobar les longituds de les peces incloses als 6 lots seleccionats aleatòriament aquest mes (es tracta dels següents: 4, 6, 10, 25, 45 i 49).

a) En base a l'anterior informació, comproveu si la comanda d'aquest mes compleix els dos requisits especificats.

b) Calculeu el coeficient de correlació intraconglomerats per ambdues estimacions (la mitjana i la proporció) i expliqueu quina implicació té sobre la mida mostral que requeriria el m.a.s. per a assolir la mateixa precisió.

c) Calculeu el nombre de conglomerats que seria necessari incloure a la mostra si es vol obtindre una estimació de la mitjana poblacional amb un error de precisió de 0,06 i, alhora, una estimació de la proporció d'unitats de menys de 24′5 mm amb un error de 0,035.

Annex B. FORMULARIS

FORMULARIS

B.1 FORMULARI DE MOSTREIG PROBABILÍSTIC I ESTIMADORS

Paràmetres i estimadors

	Paràmetre	*Estimador no esbiaixat*	*Variància de l'estimador*
Mitjana	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$Var(\bar{x}) = \frac{\sigma_x^2}{n}$
Total	$T = \sum_{i=1}^{N} x_i$	$\hat{T} = \sum_{i=1}^{n} x_i$	$Var(\hat{T}) = N^2 \frac{\sigma_x^2}{n}$
Proporció de classe	$p = \frac{1}{N}\sum_{i=1}^{N} A_i$	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} A_i$	$Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$
Total de classe	$A = \sum_{i=1}^{N} A_i$	$\hat{A} = \sum_{i=1}^{n} A_i$	$Var(\hat{A}) = N^2 \frac{\hat{p}(1-\hat{p})}{n}$

on $A_i = \begin{cases} 1, si\ i \in A \\ 0, si\ i \notin A \end{cases}$

Característiques dels estimadors

Esperança → $E(\hat{\theta}) = \begin{cases} \sum t\ P(\hat{\theta} = t)\text{, per distribucions discretes} \\ \int_{-\infty}^{+\infty} t\ f(t)dt\text{, per distribucions contínues} \end{cases}$

Biaix → $B(\hat{\theta}) = E(\hat{\theta}) - \hat{\theta}$

Variància → $Var(\hat{\theta}) = E\left[\hat{\theta} - E(\hat{\theta})\right]^2 = E(\hat{\theta}^2) - E(\hat{\theta})^2$

Error Quadràtic Mitjà (EQM) → $EQM(\hat{\theta}) = Var(\hat{\theta}) + B(\hat{\theta})^2$

Tipus d'errors

Error absolut → $e(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sigma(\hat{\theta})$

Error relatiu → $e_r(\hat{\theta}) = \frac{\sigma(\hat{\theta})}{\hat{\theta}}$

Marge d'error absolut → $e_\propto(\hat{\theta})\ \ tq\ \ P\{\theta \in [\hat{\theta} \pm e_\propto(\hat{\theta})]\} = 1-\propto$

Marge d'error relatiu → $e_{r\propto}(\hat{\theta})\ \ tq\ \ P\left\{\left|\frac{\hat{\theta}-\theta}{\hat{\theta}}\right| \leq e_{r\propto}(\hat{\theta})\right\} = 1-\propto$

Intervals de confiança

a) $\hat{\theta} \sim Normal,\ \sigma^2$ és coneguda → $\theta \in [\hat{\theta} \pm z_{\propto/2}\, \sigma(\hat{\theta})]$

b) $\hat{\theta} \sim Normal,\ \sigma^2$ desconeguda → $\theta \in [\hat{\theta} \pm t_{n-1;\propto/2}\, \hat{\sigma}(\hat{\theta})]$

c) $\hat{\theta} \nsim Normal$ → $\theta \in \left[\hat{\theta} \pm \frac{\sigma(\hat{\theta})}{\sqrt{\propto}}\right]$

B.2 FORMULARI DE MOSTREIG ALEATORI SIMPLE

Paràmetres i estimadors

	Estimador no esbiaixat	*Variància de l'estimador*
Mitjana	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$Var(\bar{x}) = (1-f)\frac{S_x^2}{n}$
Total	$\hat{T} = \sum_{i=1}^{n} x_i = N\bar{x}$	$Var(\hat{T}) = N^2(1-f)\frac{S_x^2}{n}$
Proporció de classe	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} A_i$	$Var(\hat{p}) = \frac{1}{n-1}(1-f)\,\hat{p}(1-\hat{p})$
Total de classe	$\hat{A} = \sum_{i=1}^{n} A_i = N\hat{p}$	$Var(\hat{A}) = N^2\frac{1}{n-1}(1-f)\,\hat{p}(1-\hat{p})$

on: - $f = \frac{n}{N}$ és la fracció de mostreig

- $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ és la quasivariància mostral

Mida mostral, per poblacions finites i infinites

		e	e_r	$e_\propto$	$e_{r\propto}$
N finit	Mitjana	$n = \dfrac{N\,S_x^2}{N\,e^2 + S_x^2}$	$n = \dfrac{N\ {S_x^2}/{\mu}}{N\,{e_r}^2 + {S_x^2}/{\mu}}$	$n = \dfrac{N\,z_{\propto/2}^2\,S_x^2}{N\,e_\propto^2 + z_{\propto/2}^2\,S_x^2}$	$n = \dfrac{N\,z_{\propto/2}^2\ {S_x^2}/{\mu}}{N\,{e_{r\propto}}^2 + z_{\propto/2}^2\ {S_x^2}/{\mu}}$
	Total	$n = \dfrac{N^2\,S_x^2}{e^2 + N\,S_x^2}$		$n = \dfrac{N^2\,z_{\propto/2}^2\,S_x^2}{e_\propto^2 + N\,z_{\propto/2}^2\,S_x^2}$	
	Proporció	$n = \dfrac{S_x^2}{e^2}$	$n = \dfrac{N\,(1-p)}{p\,(N-1)\,{e_r}^2 + (1-p)}$	$n = \dfrac{z_{\propto/2}^2\,S_x^2}{e_\propto^2}$	$n = \dfrac{z_{\propto/2}^2\,N\,(1-p)}{p\,(N-1)\,{e_{r\propto}}^2 + z_{\propto/2}^2\,(1-p)}$
	Total de classe	$n = \dfrac{p(1-p)}{e^2}$		$n = \dfrac{z_{\propto/2}^2\,p(1-p)}{e_\propto^2}$	
$N \to \infty$	Mitjana	$n \to \dfrac{S_x^2}{e^2}$	$n \to \dfrac{S_x^2}{\mu^2\,{e_r}^2}$	$n \to \dfrac{z_{\propto/2}^2\,S_x^2}{e_\propto^2}$	$n \to \dfrac{z_{\propto/2}^2\,S_x^2}{\mu^2\,{e_r}^2}$
	Proporció	$n \to \dfrac{p(1-p)}{e^2}$	$n \to \dfrac{(1-p)}{p\,{e_r}^2}$	$n \to \dfrac{z_{\propto/2}^2\,p(1-p)}{e_\propto^2}$	$n \to \dfrac{z_{\propto/2}^2\,(1-p)}{p\,{e_{r\propto}}^2}$

B.3 FORMULARI DE MOSTREIG ALEATORI ESTRATIFICAT

Paràmetres i estimadors

	Estimador no esbiaixat	*Variància de l'estimador*
Mitjana	$\bar{x} = \sum_{h=1}^{L} W_h \, \bar{x}_h$	$Var(\bar{x}) = \sum_{h=1}^{L} W_h^2 \, (1 - f_h) \frac{S_h^2}{n_h}$
Total	$\hat{T} = \sum_{i=1}^{n} N_h \, \bar{x}_h = N\bar{x}$	$Var(\hat{T}) = \sum_{h=1}^{L} N_h^2 \, (1 - f_h) \frac{S_h^2}{n_h}$
Proporció de classe	$\hat{p} = \sum_{h=1}^{L} W_h \, \hat{p}_h$	$Var(\hat{p}) = \sum_{h=1}^{L} W_h^2 \, (1 - f_h) \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$
Total de classe	$\hat{A} = \sum_{h=1}^{L} N_h \, \hat{p}_h = N\hat{p}$	$Var(\hat{A}) = \sum_{h=1}^{L} N_h^2 \, (1 - f_h) \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$

on: - $f_h = \frac{n}{N}$ és la fracció de mostreig estratal

- $W_h = \frac{N_h}{N}$ representa el pes de l'estrat *h* en la població

- $S_h^2 = \begin{cases} \frac{n_h}{n_h - 1} \sigma_h^2 & \text{és la quasivariància mostral estratal d'una variable} \\ \frac{n_h}{n_h - 1} \, \hat{p}_h(1 - \hat{p}_h) & \text{és la quasivariància mostral d'una variable binària} \end{cases}$

i $\sigma_h^2 = \frac{1}{n_h} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$.

Afixacions de mostres autoponderades

Tipus d'afixació	n_h
Uniforme	$n_h = k = \frac{n}{L}$
Proporcional	$n_h = n \, W_h$
Mínima variància	$n_h = n \frac{N_h \, S_h}{\sum_{h=1}^{L} N_h \, S_h} = n \frac{W_h \, S_h}{\sum_{h=1}^{L} W_h \, S_h}$
Òptima	$n_h = n \frac{N_h \, S_h / \sqrt{C_h}}{\sum_{h=1}^{L} \left(N_h \, S_h / \sqrt{C_h} \right)}$

Mida mostral, per poblacions finites

	Paràmetre poblacional	Afixació proporcional	Afixació de mínima variància	Afixació òptima	Sense especificar el tipus d'afixació
e	Mitjana	$n = \frac{\sum_{h=1}^{L} W_h\, S_h^2}{e^2 + \sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} W_h\, S_h\right)^2}{e^2 + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} \frac{W_h\, S_h}{\sqrt{C_h}}\right)\left(\sum_{h=1}^{L} W_h\, S_h \sqrt{C_h}\right)}{e^2 + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\sum_{h=1}^{L} \frac{W_h^2}{w_h}\, S_h^2}{e^2 + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$
	Total	$n = \frac{N\ \sum_{h=1}^{L} N_h\, S_h^2}{e^2 + \sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} N_h\, S_h\right)^2}{e^2 + \frac{1}{N}\sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} \frac{N_h\, S_h}{\sqrt{C_h}}\right)\left(\sum_{h=1}^{L} N\, S_h \sqrt{C_h}\right)}{e^2 + \frac{1}{N}\sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\sum_{h=1}^{L} \frac{N_h^2}{w_h}\, S_h^2}{e^2 + \sum_{h=1}^{L} N_h\, S_h^2}$
$e_\propto$	Mitjana	$n = \frac{\sum_{h=1}^{L} W_h\, S_h^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} W_h\, S_h\right)^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} \frac{W_h\, S_h}{\sqrt{C_h}}\right)\left(\sum_{h=1}^{L} W_h\, S_h \sqrt{C_h}\right)}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$	$n = \frac{\sum_{h=1}^{L} \frac{W_h^2}{w_h}\, S_h^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} W_h\, S_h^2}$
	Total	$n = \frac{N \sum_{h=1}^{L} N_h\, S_h^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} N_h\, S_h\right)^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\left(\sum_{h=1}^{L} \frac{N_h\, S_h}{\sqrt{C_h}}\right)\left(\sum_{h=1}^{L} N_h\, S_h \sqrt{C_h}\right)}{\frac{e_\propto^2}{z_{\propto/2}^2} + \frac{1}{N}\sum_{h=1}^{L} N_h\, S_h^2}$	$n = \frac{\sum_{h=1}^{L} \frac{N_h^2}{w_h}\, S_h^2}{\frac{e_\propto^2}{z_{\propto/2}^2} + \sum_{h=1}^{L} N_h\, S_h^2}$

B.4 FORMULARI DE MOSTREIG SISTEMÀTIC

Paràmetres i estimadors

	Estimador no esbiaixat	*Variància de l'estimador*
Mitjana	$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \bar{x}_j$	$Var(\hat{\mu}) = (1-f)\frac{S_{BS}^2}{n}$ $= \sigma^2 - \frac{n-1}{n} S_{WS}^2$
Total	$\hat{T} = \frac{N}{n}\sum_{i=1}^{n} x_{ij} = N\bar{x}_j$	$Var(\hat{T}) = N^2 Var(\hat{\mu})$ $= N(N-1)S^2 - N(N-k)S_{WS}^2$
Proporció de classe	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} A_{ij} = \hat{p}_j$	$Var(\hat{p}) = \frac{n}{N}\sum_j (\hat{p}_j - \hat{p})^2$ $= p(1-p) - \frac{1}{k}\sum_j \left[\hat{p}_j(1-\hat{p}_j)\right]$
Total de classe	$\hat{A} = \frac{N}{n}\sum_{i=1}^{n} A_{ij} = N\hat{p}_j$	$Var(\hat{A}) = N^2 Var(\hat{p})$ $= N^2\left\{p(1-p) - \frac{1}{k}\sum_j\left[\hat{p}_j(1-\hat{p}_j)\right]\right\}$

on: - $f = \frac{n}{N}$ és la fracció de mostreig

- $\sigma^2 = \frac{1}{N}\sum_{i,j} (x_{ij} - \bar{x}_j)^2$

- $S^2 = \begin{cases} \frac{1}{N-1}\sum_{i,j}(x_{ij}-\mu)^2 \\ \frac{N}{N-1}\, p(1-p) \end{cases}$ és la quasivariància poblacional,

de manera que es demostra: $(N-1)S^2 = (N-k)S_{WS}^2 + (k-1)S_{BS}^2$

- $S_{BS}^2 = \begin{cases} \frac{n}{k-1}\sum_j(\bar{x}_j-\mu)^2 \\ \frac{n}{k-1}\sum_j(\hat{p}_j-p)^2 \end{cases}$ és la quasivariància intermostral

- $S_{WS}^2 = \begin{cases} \frac{1}{N-k}\sum_{i,j}(x_{ij}-\bar{x}_j)^2 \\ \frac{1}{N-k}\sum_j \hat{p}_j(1-\hat{p}_j) \end{cases}$ és la quasivariància intramostral

Coeficients de correlació

a) Coeficient de correlació intramostral (ρ_w):

$$\left.\begin{matrix} Var(\hat{\mu}) \\ Var(\hat{p}) \end{matrix}\right\} = \frac{N-1}{N}\frac{S^2}{n}\left[1+(n-1)\rho_w\right]$$

b) Coeficient de correlació estratal (ρ_{St}):

$$\left.\begin{matrix} Var(\hat{\mu}) \\ Var(\hat{p}) \end{matrix}\right\} = (1-f)\frac{S_{WSt}^2}{n}\left[1+(n-1)\rho_{St}\right]$$

Estimació real de les variàncies dels estimadors

a) Supòsit $\rho_w \approx 0$: ús de les fórmules pròpies del m.a.s.
b) Supòsit $\rho_{St} \approx 0$: ús de les fórmules pròpies del m.a.e., per exemple:

$$Var(\hat{\mu}) = \frac{1-f}{n^2}\sum_{h=1}^{n/2}(x_{h1} - x_{h2})^2$$

c) No es pot fer cap dels supòsits anteriors: ús de mostres interpenetrants.

$$\hat{\mu} = \bar{x}_C = \frac{1}{t}\sum\nolimits_i \bar{x}_i$$

$$Var(\hat{\mu}) = \frac{1}{t(t-1)}\sum_{i=1}^{t}\left(\bar{x}_i^2 - \bar{x}_C^2\right) = \{t = 2\} \Rightarrow \frac{(\bar{x}_1 - \bar{x}_1)^2}{4}$$

B.5 FORMULARI DE MOSTREIG PER CONGLOMERATS

Paràmetres i estimadors

	Estimador no esbiaixat	*Variància de l'estimador*
Mitjana	$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{M}\sum_{j=1}^{n} x_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_i = \bar{x}$	$Var(\hat{\mu}) = (1-f)\frac{S_B^2}{nM}$
Total	$\hat{T} = NM\bar{x}$	$Var(\hat{T}) = N^2M^2\,(1-f)\frac{S_B^2}{nM}$
Proporció de classe	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{M}\sum_{j=1}^{n} A_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_i$	$Var(\hat{p}) = (1-f)\frac{S_B^2}{nM}$
Total de classe	$\hat{A} = NM\hat{p}$	$Var(\hat{A}) = N^2M^2(1-f)\frac{S_B^2}{nM}$

on: - $f = \frac{n}{N}$ és la fracció de mostreig

- $S^2 = \frac{1}{nM-1}\sum_{i=1}^{n}\sum_{j=1}^{M}(x_{ij} - \hat{\mu})^2$, és la quasivariància mostral no esbiaixada per mostres grans ($n > 50$).
  En el cas de mostres petites, l'estimació anterior és esbiaixada i cal estimar S^2 de la següent manera: $\hat{S}_0^2 = \frac{N-1}{NM-1}\hat{S}_B^2 + \frac{N(M-1)}{NM-1}\hat{S}_W^2$.

- $S_W^2 = \begin{cases} \frac{1}{NM-n}\sum_{i=1}^{n}\sum_{j=1}^{M}(x_{ij}-\bar{x}_i)^2 \\ \frac{1}{NM-n}\sum_{i=1}^{n}\sum_{j=1}^{M}(A_{ij}-\hat{p}_i)^2 \end{cases}$ és la quasivariància mostral dins els conglomerats
- $S_B^2 = \begin{cases} \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{M}(\bar{x}_i-\bar{x})^2 \\ \frac{M}{N-1}\sum_{i=1}^{n}(\hat{p}_j-\hat{p})^2 \end{cases}$ és la quasivariància mostral entre conglomerats

Coeficient de correlació intraconglomerats

Coeficient de correlació intraconglomerats (δ):

$$\left.\begin{matrix} Var(\hat{\mu}) \\ Var(\hat{p}) \end{matrix}\right\} = (1-f)\frac{S^2}{nM}[1+(M-1)\delta]$$

Annex C. TAULES ESTADÍSTIQUES

TAULES ESTADÍSTIQUES

C.1 DISTRIBUCIÓ NORMAL ESTÀNDARD

z	**0**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998

C.2 DISTRIBUCIÓ T-STUDENT

Graus de llibertat	Probabilitat acumulada						
v	$0,8$	$0,9$	$0,95$	$0,975$	$0,99$	$0,995$	$0,999$
1	1,376	3,078	6,314	12,706	31,821	63,656	318,289
2	1,061	1,886	2,92	4,303	6,965	9,925	22,328
3	0,978	1,638	2,353	3,182	4,541	5,841	10,214
4	0,941	1,533	2,132	2,776	3,747	4,604	7,173
5	0,92	1,476	2,015	2,571	3,365	4,032	5,894
6	0,906	1,44	1,943	2,447	3,143	3,707	5,208
7	0,896	1,415	1,895	2,365	2,998	3,499	4,785
8	0,889	1,397	1,86	2,306	2,896	3,355	4,501
9	0,883	1,383	1,833	2,262	2,821	3,25	4,297
10	0,879	1,372	1,812	2,228	2,764	3,169	4,144
11	0,876	1,363	1,796	2,201	2,718	3,106	4,025
12	0,873	1,356	1,782	2,179	2,681	3,055	3,93
13	0,87	1,35	1,771	2,16	2,65	3,012	3,852
14	0,868	1,345	1,761	2,145	2,624	2,977	3,787
15	0,866	1,341	1,753	2,131	2,602	2,947	3,733
16	0,865	1,337	1,746	2,12	2,583	2,921	3,686
17	0,863	1,333	1,74	2,11	2,567	2,898	3,646
18	0,862	1,33	1,734	2,101	2,552	2,878	3,61
19	0,861	1,328	1,729	2,093	2,539	2,861	3,579
20	0,86	1,325	1,725	2,086	2,528	2,845	3,552
21	0,859	1,323	1,721	2,08	2,518	2,831	3,527
22	0,858	1,321	1,717	2,074	2,508	2,819	3,505
23	0,858	1,319	1,714	2,069	2,5	2,807	3,485
24	0,857	1,318	1,711	2,064	2,492	2,797	3,467
25	0,856	1,316	1,708	2,06	2,485	2,787	3,45
26	0,856	1,315	1,706	2,056	2,479	2,779	3,435
27	0,855	1,314	1,703	2,052	2,473	2,771	3,421
28	0,855	1,313	1,701	2,048	2,467	2,763	3,408
29	0,854	1,311	1,699	2,045	2,462	2,756	3,396
30	0,854	1,31	1,697	2,042	2,457	2,75	3,385
35	0,852	1,306	1,69	2,03	2,438	2,724	3,34
40	0,851	1,303	1,684	2,021	2,423	2,704	3,307
45	0,85	1,301	1,679	2,014	2,412	2,69	3,281
50	0,849	1,299	1,676	2,009	2,403	2,678	3,261
55	0,848	1,297	1,673	2,004	2,396	2,668	3,245
60	0,848	1,296	1,671	2	2,39	2,66	3,232
70	0,847	1,294	1,667	1,994	2,381	2,648	3,211
80	0,846	1,292	1,664	1,99	2,374	2,639	3,195
90	0,846	1,291	1,662	1,987	2,368	2,632	3,183
100	0,845	1,29	1,66	1,984	2,364	2,626	3,174
120	0,845	1,289	1,658	1,98	2,358	2,617	3,16
1000	0,842	1,282	1,646	1,962	2,33	2,581	3,098

BIBLIOGRAFIA

- ARDILLY, P. i TILLÉ, Y. (2006): *Sampling Methods: Exercices and Solutions*. Springer, New York.
- AZORÍN, F. (1972): *Curso de muestreo y aplicaciones*. Aguilar, Madrid.
- AZORÍN, F. i SÁNCHEZ-CRESPO, J.L. (1986): *Métodos y aplicaciones del muestreo*. Alianza, Madrid.
- COCHRAN, W. G. (1984): *Técnicas de muestreo*. Compañía Editorial Continental, México.
- FERNÁNDEZ, F. R. i MAYOR, J.A. (1995): *Muestreo en poblaciones finitas: curso básico*. EUB, Barcelona.
- HANSEN, M. H., HURWITZ, N. i MADOW, W.G. (1953): *Sample survey methods and theory*. Wiley, New York.
- KISH, L. (1979): *Muestreo de encuestas*. Trillas, México.
- LOHR, S. (2000): *Muestreo: Diseño y análisis*. International Thomson Editores. México.
- MARTÍNEZ, V. C. (2004): *Diseño de encuestas de opinión*. Ra-ma. Madrid.
- PÉREZ, C. (1999): *Técnicas de muestreo estadístico.* Ra-ma. Madrid.
- PÉREZ, C. (2005): *Muestreo estadístico. Conceptos y problemas resueltos*. Pearson-Prentice Hall.
- RODRÍGUEZ, J. (1991): *Métodos de muestreo*. Centro de Investigaciones Sociológicas, Madrid.
- RODRÍGUEZ, J. (1993): *Métodos de muestreo: casos prácticos*. Centro de Investigaciones Sociológicas, Madrid.