

107 18 18
A Q2A ESTADÍSTICA
AFC 180A


FONAMENTS D'ESTADÍSTICA APLICADA

Quadern de pràctiques amb Minitab

Albert Ferrer i Biosca
Antoni Guillamon i Grabolosa
Margarida Mitjana i Riera
Joan J. Rodríguez i Jordana
Carles Serrat i Piè

Professors de la Universitat Politècnica de Catalunya
Departament de Matemàtica Aplicada I
Secció Escola Universitària Politècnica de Barcelona

Barcelona 1997

 **UPC**
Escola Universitària
Politécnica de Barcelona
BIBLIOTECA

Pròleg

Amb aquest quadern pretenem introduir l'alumne en l'ús dels paquets estadístics donant les pautes més elementals del programa d'anàlisi estadística Minitab, així com indicacions per a la seva utilització en els temes vinculats als cursos d'Estadística dels estudis d'Arquitectura Tècnica i Enginyeria Tècnica en Topografia que s'imparteixen a l'Escola Universitària Politècnica de Barcelona. El seu contingut està pensat també per ser utilitzat en les sessions pràctiques d'aquestes assignatures.

L'ús, cada vegada més estès, de les noves tecnologies a tots els camps, en particular a l'ensenyament, planteja nous reptes i obra noves possibilitats. És obvi que la utilització de paquets estadístics permet el plantejament i la resolució ràpida de nous problemes amb gran nombre de dades i amb tècniques sofisticades d'anàlisi. Tots aquests aspectes són difícils de resoldre amb tècniques manuals.

El plantejament que proposem en aquestes pràctiques és fer servir el programa Minitab per a l'aprenentatge de l'Estadística, no tant com instrument d'aplicació pràctica i per a obtenir resultats fàcilment, sinó com a eina que permet completar l'actuació docent i, en conseqüència, millorar la comprensió dels conceptes estadístics que els plans d'estudis contemplen.

Per tal d'aconseguir l'objectiu docent, cada pràctica està organitzada de la manera següent: **Objectius** -conceptes principals que la pràctica tracta-. **Dades** -breu descripció dels conjunts de dades que intervenen en la pràctica-. **Desenvolupament** -sessió guiada, si fa falta organitzada en forma de problemes, en què es van tractant el diferents conceptes-. I **Exercicis** -problemes oberts que es poden resoldre de manera molt semblant a com s'ha desenvolupat la pràctica-.

Els temes que es tracten en les pràctiques, d'acord amb els objectius d'una assignatura bàsica d'Estadística són:

- **Pràctica 0: Primeres passes amb Minitab.** Serveix d'introducció general al programa; s'hi tracten aspectes genèrics de funcionament, de gestió de finestres, d'edició de comandes, d'utilització de menús, de gràfics... És un primer recorregut pel programa.
- **Pràctica 1: Estadística descriptiva i distribucions bidimensionals.** Es treballen els càlculs bàsics de l'estadística descriptiva i la seva interpretació i els elements que intervenen en la regressió entre dues variables.
- **Pràctica 2: Exercicis de probabilitat i distribucions.** Aquí es pretén, d'una banda, l'ús del Minitab per a la simulació de processos aleatoris i, d'una altra, il·lustrar les distribucions de probabilitat i l'estimació de la probabilitat d'un esdeveniment, tant en el cas discret com en el continu. També s'aprofita per a tractar conceptes com ara el Teorema de Txebixev o l'aproximació de la distribució binomial per la distribució Normal.
- **Pràctica 3: Estadística inductiva.** L'objectiu en aquesta pràctica és il·lustrar, mitjançant simulació, la distribució mostral dels estadístics mitjana mostral i el vinculat a la desviació tipus mostral. Després, això permetrà tractar l'estimació de la mitjana per intervals de confiança, el prendre decisions basades en contrastos d'hipòtesi sobre la mitjana d'una distribució i l'estimació de proporcions per intervals de confiança.

Barcelona, Octubre de 1997.

Els autors.

Índex

Pràctica 0: Primeres passes amb Minitab	1
Pràctica 1: Estadística descriptiva i distribucions bidimensionals	7
Pràctica 2: Exercicis de probabilitat i distribucions	12
Pràctica 3: Estadística inductiva	21
Bibliografia	29

PRÀCTICA 0: PRIMERES PASSES AMB MINITAB

OBJECTIUS

L'objectiu d'aquesta pràctica inicial és el d'introduir-vos en el funcionament general del programa d'anàlisi estadística de dades MINITAB. Per tant es tractarà de fer diverses actuacions que ens permetin fer un primer recorregut pel programa, a la vegada que n'anem comentant el seu funcionament.

Aquesta documentació parteix de la base que l'usuari està familiaritzat amb l'entorn Windows 3.*. És a dir, que se suposa que se sap moure entre finestres, amb menús desplegable, quadres de diàleg, botons ... i que es coneix, més o menys, la mecànica de Copiar (Ctrl+C), tallar (Ctrl+X) i enganxar (Ctrl+V).

DADES

El fitxer inicial de dades amb què comença la pràctica és el *notes.mtw*. Aquest fitxer conté dues úniques variables C1 i C2 que contenen les notes que 150 alumnes han obtingut en dos exàmens parcials.

DESENVOLUPAMENT

1. Poseu en marxa el programa MINITAB a partir de la icona corresponent del Grup de programes *Minitab 11 for Windows*.

Una vegada en funcionament observem que l'àrea de treball conté quatre finestres (ara dues a la vista i unes altres dues minimitzades) que són: *Session*, *Data*, *History* i *Info*, el contingut de les quals ara passem a descriure.

- *Session*: Contindrà tot el contingut del desenvolupament de la sessió de treball, tant l'encàrrec d'ordres com els resultats de les mateixes. També s'anomena la *Sortida* o *Output*. És editable i per tant hi podem posar comentaris a les anàlisis que anem realitzant. Podem accedir a aquesta finestra amb el ratolí, amb l'opció de menú *Window* ▸ *Session* o simplement amb la combinació de tecles Ctrl+M.
- *Data*: És la taula que conté les dades. Les diferents variables estan organitzades en forma de columnes i cada fila representa una observació. La navegació per la taula és l'habitual i podem accedir a aquesta finestra

amb el ratolí, amb l'opció de menú *Window* ▷ *Data* o simplement amb la combinació de tecles **Ctrl+D**.

- *History*: És la finestra on quedaran enregistrades totes les ordres que anem fent. Podem accedir a aquesta finestra amb el ratolí, amb l'opció de menú *Window* ▷ *History* o simplement amb la combinació de tecles **Ctrl+H**.
- *Info*: És la finestra on tenim constància del nombre de variables que tenim definides, amb quantes observacions i quin nom tenen, les constants que estem utilitzant (que s'identificaran per K1, K2, ... amb el valor actualitzat i el seu nom, i les matrius que tinguem definides en aquell moment. Accedim a aquesta finestra amb el ratolí, amb l'opció de menú *Window* ▷ *Info* o simplement amb la combinació de tecles **Ctrl+I**.

Les opcions que trobem en la barra de menús i en la barra de botons depenen de la finestra que tiguem activa en aquells moments.

2. Comproveu l'estat actual de les quatre finestres esmentades.
3. Carregueu el fitxer *notes.mtw* del subdirectori *F:\soft\minitab\data* fent:
> *File* ▷ *Open Worksheet* ▷ *notes.mtw*
i comproveu el contingut de les finestres.
4. En el menú *Editor* observeu les opcions *Enable/Disable Command Language* i *Make Output Read Only/Editable*. Aquestes opcions operen amb caracter d'interruptor i ens interessa que a la finestra *Session* hi puguem entrar comandes (presentarà el símbol del punt indicatiu MTB >) i que sigui editable.
Deixeu el menú *Editor* amb les opcions *Disable Command Language* i *Make Output Read Only*. A la finestra *Session* hi ha la línia de comandes que s'identifica pel punt indicatiu MTB >, i, immediatament al damunt, la línia d'edició de text. Escriviu-hi com a text el nom de la pràctica: **PRÀCTICA 0: PRIMERES PASSES AMB MINITAB**.
Amb l'opció
> *Editor* ▷ *Select Fonts* ▷ *Select Comment Font...*
podeu canviar el tipus i estil de lletra que esteu fent servir per als vostres comentaris. Escolliu lletra Arial, Cursiva de 14 punts.
5. A la finestra de les dades, identifiqueu el contingut de les dues variables i navegueu amb les tecles del cursor, amb les tecles **Inicio** i **Fin**, i amb la combinació **Ctrl+Inicio** i **Ctrl+Fin**.
6. Poseu nom a cada variable: **Primer** per a la variable C1 i **Segon** per a la variable C2. Només cal situar-se a la casella immediatament inferior a la del nom de la variable i escriure el text corresponent (i acabar l'edició amb la tecla **Return**).

A partir d'aquest moment ens podem referir des del punt indicatiu a aquestes variables per C1 i C2 o 'Primer' i 'Segon', respectivament. (Atenció: No us deixeu la cometa quan hi feu referència.)

7. Ara mateix tenim 150 dades. Afegiu, al final, les notes corresponents a 10 estudiants més, que han obtingut les notes

2.3, 4.4, 5.7, 0.8, 6.5, 8.3, 4.1, 4.8, 2.7 i 7.2;
3.9, 5.9, 6.7, 1.5, 8.5, 1.0, 4.7, 8.0, 4.8 i 8.8

respectivament a cadascuna de les proves.

8. Amb l'opció

> *File* ▸ *Save Worksheet As...*

gardeu les noves dades a un fitxer de nom *notesb.mtw* en el directori temporal *C:\temp* i poseu una petita descripció amb l'opció *Description....* (Observeu que les finestres *Session*, *History* i *Info* s'han anat actualitzant de manera automàtica.)

9. Després de repassar les notes anteriors ens adonem que el sisè estudiant que hem afegit havia obtingut un 10.0 i no 1.0 en el segon parcial. Esmeneu la dada en qüestió i guardeu amb l'opció

> *File* ▸ *Save Worksheet*.

10. Anem a obtenir una estadística descriptiva dels valors del primer parcial. Amb l'opció

> *Stat* ▸ *Basic Statistics* ▸ *Descriptive Statistics...* ▸

amb el cursor a *Variables*: podem escriure C1 o bé Primer o bé fer doble click sobre C1 Primer de la columna de l'esquerra i OK. A la finestra *Session* obtenim el resultat demanat (Atenció: la StDev que ha calculat és la desviació tipus corregida).

11. Feu una anàlisi descriptiva de les dues variables C1 i C2 (Podeu seleccionar simultàniament les dues variables, o bé indicar-les per C1-C2) que mostri també un histograma de les dades, fent servir l'opció

> *Stat* ▸ *Basic Statistics* ▸ *Descriptive Statistics* ▸ *Graphs...*

S'obtenen dues finestres gràfiques que mostren els histogrames respectius i la finestra *Session* mostra els valors numèrics resultants. Comenteu els resultats numèrics i gràfics obtinguts.

12. Amb el primer bloc d'opcions del menú *Window*, organitzeu diferents presentacions de les finestres que teniu en aquests moments.

Amb l'opció *Close All Graphics* del menú *Window*, tanqueu les dues finestres gràfiques que hem generat. Repartiu l'àrea de treball entre les finestres *Session* i *Data*.

13. A la variable C3, guardeu-hi la mitjana aritmètica de les variables C1 i C2. Per a fer-ho, feu servir l'opció

> *Calc* ▷ *Calculator* ▷

Store result in variable: C3

Expression: $(C1+C2)/2$ o bé $(\text{Primer}+\text{Segon})/2$

Observeu l'ordre que s'ha executat en la finestra *Session*.

Anomeneu *Global* a aquesta nova variable C3.

14. Observeu que la variable *Global* està expressada amb dos decimals. Arrodoniu a un sol decimal els valors de la nota global. Amb l'opció

> *Help* ▷ *Contents* ▷ *Glossary*

consulteu l'ajuda sobre les funcions que comencen per R.

A la finestra *Session*, assigneu a C4 el resultat de l'arrodoniment de la variable C3 i anomeu *Nota* la nova variable.

Canvieu la fórmula anterior per tal que mostri realment un decimal. A la finestra *Session* intercaleu algunes línies de comentari sobre el que esteu fent.

15. Amb l'opció

> *Stat* ▷ *Basic Statistics* ▷ *Descriptive Statistics* ▷ *Graphs...*

elaboreu un *Boxplot* de la variable *Nota*. S'identifica algun valor atípic (*outlier*)? Tanqueu el gràfic.

16. Els 160 alumnes que tenim, de fet, pertanyen a dos grups diferents, l'1 i el 2. Definiu una variable que es digui *Grup* i assigneu als 80 primers el valor 1 i als restants el valor 2. Per això utilitzeu l'opció

> *Calc* ▷ *Make Patterned Data* ▷ *Simple Set of Numbers* ▷

Store patterned data in: C5

From first value: 1

To last value: 2

In steps of: 1

List each value: 80

Guardeu les dades.

17. Compareu els estadístics per separat dels grups 1 i 2 i feu un *Boxplot* per a cada categoria, amb l'opció

> *Stat* ▷ *Basic Statistics* ▷ *Descriptive Statistics* ▷

By variable: C5 o bé *Grup*

Què en podeu dir comparativament dels dos grups? I dels valors atípics? Comenteu-ho a la finestra *Session*.

18. Suposem que la nota global és el resultat de fer la mitjana ponderada dels dos exàmens d'acord amb la fórmula

$$Global = \frac{K1 * Primer + K2 * Segon}{K1 + K2}$$

on $K1$ i $K2$ són dues constants que fixem oportunament.

A la finestra *Session*, assigneu amb l'ordre `Let K1=2` el valor 2 a la constant $K1$ i, de manera anàloga, assigneu el valor 3 a la constant $K2$ (Observeu els valors de les constants a la finestra *Info*). Recalculeu les variables *Global* i *Nota*. Dibueixeu un histograma amb l'opció

> *Graph* ▸ *Histogram...* ▸

Graph 1 : Nota

i amb el botó *Annotation* ▸ *Title* poseu un títol personalitzat al gràfic. No tanqueu el gràfic resultant.

19. Suposem que us interessa *simular* l'histograma de la variable *Nota* per a diferents valors de les constants $K1$ i $K2$. Com que és el que acabeu de fer per als valors 2 i 3, respectivament, només cal que a la finestra *Session* copieu (amb `Ctrl+C`) les línies corresponents a les ordres que voleu executar i que les enganxeu (`Ctrl+V`) a la línia de comandes. Una vegada enganxades, editeu els valors que us interessin i `Return` per acabar. El procés el podreu repetir tantes vegades com us convingui.

Repetiu l'apartat anterior pels valors $K1=3$ i $K2=10$, i per a $K1=10$ i $K2=3$.

Compareu el gràfics resultants. Una cop feta la comparació tanqueu les finestres gràfiques.

20. Les variables i les constants es poden veure a la finestra *Session* amb la comanda `print ...` i es poden eliminar amb la comanda `erase ...`

Visualitzeu els valors de les constants i d'algunes variables. Elimineu les constants actuals.

21. Per a acabar, anem a definir una variable qualitativa, que anomenarem *Qualific*:
> *Manip* ▸ *Code* ▸ *Numeric to Text...* ▸

```
Code data from columns: C4
Into columns:          C6
Original values:      New
0:4.9 :                Suspens
5:6.5 :                Aprovat
6.5:8.5 :              Notable
8.5:9.75 :             Excel.lent
9.75:10 :              Matr. Honor
```

Observeu que l'etiqueta que identifica aquesta variable ja indica que es tracta d'una variable de tipus Text. Enregistreu les dades.

Podeu fer ara un Diagrama de Sectors amb la variable Qualific:

> *Graph* ▷ *Pie Chart...* ▷

```
Chart data in: C6
Title:         Distribució de Notes
```

Si feu doble click sobre el gràfic apareixen les barres d'icones gràfiques per a editar-lo. Canvieu algun color i modifiqueu la posició d'alguna etiqueta, d'un sector circular ... per tal de fer-lo més atractiu.

Tanqueu el gràfic. Guardeu les dades i, després d'aquesta passejada pel MINITAB, passeu a la Pràctica 1. En sortir del programa per l'opció

> *File* ▷ *Exit*

aquest ens permetrà de guarda la finestra *Session* en format de text, per a futures anàlisis. Mentre no abandonem MINITAB, els resultats es van acumulant en aquesta finestra.

PRÀCTICA 1: ESTADÍSTICA DESCRIPTIVA I DISTRIBUCIONS BIDIMENSIONALS

OBJECTIUS

Els objectius són que l'alumne aprengui a fer els càlculs elementals d'estadística descriptiva, les representacions gràfiques més habituals de les dades d'una mostra, així com regressions entre variables.

DADES

El fitxer que s'utilitza és *lacramoc.mtw*, que conté 4 columnes i 40 files cadascuna, amb dades demogràfiques sobre les comarques catalanes. Les dades de cada columna corresponen, respectivament, al Nom, la Superfície, Població al 1900 i Població al 1986 de cada comarca. El fitxer *lamron.mtw* és necessari per fer l'Exercici 1, i conté una columna amb 500 dades.

DESENVOLUPAMENT

L'entrada a MINITAB ens presenta les finestres de *Session* i *Data*. A part, observeu que hi ha minimitzades les de *History* i *Info*, en general menys utilitzades.

1. Carregueu el fitxer *lacramoc.mtw* fent:
> *File* ▾ *Open Worksheet* ▾ *lacramoc.mtw*
2. Un error en la manipulació ha fet que ens deixéssim d'introduir la **Fila 14**.
Inserteu entre les files 13 i 14 actuals les dades (1182.46, 27217, 40677).
Cal posar-se damunt l'actual Fila 14 i fer:
> *Editor* ▾ *Insert Row*
Llavors, substituïu els asteriscs per les dades corresponents.
3. Calculeu les mesures de tendència central i de dispersió de les 3 variables, amb l'opció
> *Stat* ▾ *Basic Statistics* ▾ *Descriptive statistics*
4. Feu un *boxplot* de les dades de la columna C2:
> *Graph* ▾ *Boxplot* ▾
Abans de prémer *OK*, observeu que tinguem les opcions *IQ Range* i *Outlier*

symbols. D'aquesta manera ens farà el *boxplot* tal com estem acostumats (vegeu p.24 del llibre) i ens marcarà les dades atípiques, generalment amb un asterisc. Podeu demanar també uns resums gràfics d'estadística descriptiva, que inclouen, a part dels *boxplot*, histogrames i altres gràfics:
> *Stat* ▸ *Basic Statistics* ▸ *Descriptive Statistics* ▸ *Graph* ▸

Variables: C2

i seleccioneu després els que us convinguin.

5. Feu el mateix amb la columna C3 i fixeuvos que ara sí que surten dades atípiques.
6. Constrasteu les dades del *boxplot* amb els estadístics que ja havíeu calculat abans. Amb
> *Window* ▸ *Session*
canvieu de finestra i, dins de *Session*, feu, per exemple, el *boxplot* de C4.
7. Un cop establert aquest primer contacte, ens disposem a efectuar una regressió entre les columnes C3 i C4. Si només ens interessa el coeficient de correlació, fem:
> *Stat* ▸ *Basic Statistics* ▸ *Correlation* ▸

Variables: C3 C4

8. Se'n pot deduir una correlació significativa?
9. Si volem veure gràficament la relació, anem a
> *Stat* ▸ *Regression* ▸ *Fitted Line Plot*

Response (Y): C4

Response (X): C3

Sense sortir d'aquest quadre, aneu a *Options*, on podeu escollir:

- Si voleu transformar logarítmicament alguna de les variables (*Transformation*).
- Si voleu fixar un nivell de confiança en les estimacions (*Confidence level*).
- Si voleu que mostri unes bandes al voltant de la recta de regressió, sota el nivell de confiança que haguem escollit. Això indica que amb una probabilitat del 95% els valors reals (és a dir, els valors reals de la variable C4, fixat un valor de la variable C3) cauran dins de la banda. També permet apreciar fàcilment les dades atípiques. (*Display Options*, *Display confidence bands*).

Maximitzeu la finestra gràfica per a veure millor el núvol de punts. Si us interessa guardar les gràfiques, ho podeu fer en un format especial *.mgf:

> *File* ▸ *Save Window As...*

10. Què diríeu que ha d'ocórrer amb les bandes si canviem el 95% per un 90% o per un 99%, respectivament? Després de pensar-ho, contrasteu la vostra opinió.

11. Suposeu que ara ens interessa treballar amb els logaritmes de les dades. Creeu una columna amb els logaritmes de C3 i una altra amb els de C4, i situeu-les, respectivament, a C5 i C6. Hi ha diverses maneres de dur-ho a terme:

> *Calc* ▸ *Mathematical Expressions* ▸

```
Variable (new or modified): C5
Expression:                  Loge(C3)
```

o bé:

> *Calc* ▸ *Functions* ▸

```
Input Column: C3
Result in:    C5
☉ Natural Log
```

És millor, però, utilitzar l'opció:

> *Calc* ▸ *Calculator*

i fer l'operació que convingui, com si es tractés d'una calculadora.

12. Un cop creades C5 i C6, estudeu-ne la correlació. Ha canviat respecte la que teníem entre C3 i C4? Doneu-ne una explicació.

13. Finalment, esborreu les dues darreres columnes:

> *Manip* ▸ *Erase variables* ▸

```
Columns ...to erase:: C5 C6
```

Salveu finalment el full de treball:

> *File* ▸ *Save Worksheet*

14. Elimineu el Barcelonès i torneu a fer la regressió dels apartats 7 a 9 per tal comparar els resultats obtinguts.

15. Calculeu la recta de regressió i el coeficient de correlació per a les mateixes variables anteriors, però ara utilitzant només les comarques que tenen menys de 100000 habitants. Per això, cal crear columnes amb les dades que ens interessin fent:

> *Manip* ▸ *Copy Columnn* ▸

```
Copy from columns: C1 C2 C3 C4
To columns:       C5 C6 C7 C8
```

Use Rows ▷

```
⊙ Use rows with column:      C4  
equal to (eg, -4,5 -2:3 14): 0:100000
```

Així, a les columnes C5, C6, C7 i C8, hi queden emmagatzemades les dades que a la columna C4 tenien un nombre menor que 100000. Busqueu, doncs, la correlació entre les columnes C7 i C8, ara.

16. Després d'efectuar la regressió, què s'observa quant a correlacions respecte l'apartat 9? Quina explicació hi trobeu?

EXERCICIS

A continuació, proposem una sèrie de exercicis per completar l'ús de les comandes relatives als dos primers temes del curs. El primer d'ells, es presenta d'una manera mixta entre el que s'entén per "pràctica guiada" i el que s'entén per "exercici proposat".

1. Considereu les dades del fitxer lamron.mtw, extreieu-ne una mostra de la columna C1 que contingui 10 elements i situeu-la a C2:
> Calc ▷ Random Data ▷ Sample from columns ▷

```
Sample:          10  
rows from column(s): C1  
Store samples in: C2
```

Observeu que hi ha una opció *Sample with replacement*. Si l'escollim, un cop treu una dada, la torna a posar entre les escollibles i, per tant, podem tenir dades repetides. En aquest cas, no l'escollirem.

- (a) Creeu a continuació 4 mostres més de 10 elements, 5 de 50 i 5 de 100. (Podeu fer servir la mecànica de Copiar (Ctrl+C), tallar (Ctrl+X) i enganxar (Ctrl+V).)
- (b) Preneu, per exemple, les 5 mostres de 10 elements. De cada una d'elles, calculeu la mitjana i copieu-les a les caselles d'una nova columna. Considereu la nova columna com a mostra i calculeu-ne la mitjana \bar{X}_{10} i la desviació tipus S_{10} .
- (c) Feu el mateix amb les mostres de 50 i de 100 elements i completeu la Taula 1.
- (d) Feu un histograma de cada una de les 3 mostres. Cal seguir l'opció
> Graph ▷ Histogram

	$n = 10$	$n = 50$	$n = 100$
X_n			
S_n			
$S_n \cdot \sqrt{n}$			

Table 1: Mitjanes de les mostres de 10, 50 i 100 elements.

- (e) Comenteu l'efecte del factor \sqrt{n} que apareix a la Taula 1. Si haguessiu de decidir-ho, quina creuríeu que ha de ser la mitjana i la desviació tipus de la població de la qual s'han extret aquestes mostres?
- (f) Graveu els càlculs efectuats, batejant el fitxer amb les 8 primeres lletres del vostre cognom (li direm genèricament fitxer `personal.mtw`). Després de fer-ho, ajunteu les dades amb les 3 o 4 companys més i podreu efectuar un experiment de més envergadura, per exemple, amb les columnes de 100 elements, repetint el mateix que s'ha fet des de l'apartat 1b fins a l'??.
2. Un constructor observa que el preu per peça d'un cert material canvia segons la quantitat de peces que compra de cop. No sap exactament quin criteri segueix el proveïdor, però pensa que el pot deduir a partir de les dades que té acumulades, i que apareixen a la Taula 2. Sospita, per experiències anteriors,

$n =$ Peces	10	30	50	100	175
$p =$ Preu total	150	260	325	500	670

Table 2: Preus p de comandes de n peces d'un determinat material.

que la relació entre n i p pot ser quadràtica, però com que no n'està segur, fa una regressió lineal, una de quadràtica i una de cúbica.

- (a) Comproveu els resultats de les regressions lineal i quadràtica amb les obtingudes a classe.
- (b) Observeu què passa amb els coeficients d'ordre més alt de la regressió cúbica. Hi trobeu alguna explicació?
3. Feu l'Exercici 5, de la Lliçó 2 (pàg. 52) del llibre.

PRÀCTICA 2: EXERCICIS DE PROBABILITAT I DISTRIBUCIONS

OBJECTIUS

Amb la segona pràctica amb MINITAB, es pretén, per una banda, fer servir les possibilitats que proporciona un programa d'estadística per a simular processos aleatoris. Un altre objectiu és la utilització del programa per a il·lustrar les distribucions de probabilitat i fer estimacions de la probabilitat d'un succés, tant en el cas discret com en el continu. També s'aprofita per a aprofundir en la comprensió d'alguns conceptes relacionats amb els temes, com el Teorema de Txebixev o l'aproximació de la binomial per la normal. Hi ha exercicis amb generació de dades per a simular un experiment aleatori i d'altres en els quals es parteix de dades emmagatzemades en fitxers externs.

DADES

El fitxer `peces.mtw` conté les dades d'una remesa de peces que han arribat a un magatzem. A la columna C1 hi ha els valors 1, 2 o 3, depenent de l'empresa que l'ha fabricat i a la columna C2, un 0 si la peça és defectuosa i un 1 en cas contrari.

Per a treballar amb distribucions de probabilitat, utilitzarem el fitxer `sacsci.mtw`, que a la columna C1 hi té els pesos de 1000 sacs de ciment.

El fitxer `KHI2.MAC` és una MACRO que automatitza la repetició d'una sèrie de comandes MINITAB.

El fitxer `Poisson.mtw` conté la simulació del nombre d'errades per pàgina en les proves d'edició d'un llibre.

És important recordar que si a la finestra *Session* hi volem entrar comandes (símbol del punt indicatiu `MTB >`) i que, a més, sigui editable, cal tenir el menú *Window* amb les opcions *Disable Command Language* i *Make Output Read Only*.

DESENVOLUPAMENT

Problema 1. Probabilitat. Simulació de tirades d'un dau.

1. Comenceu una sessió nova de MINITAB. Introduïu a les dues primeres colum-

nes C1 i C2, els valors que s'obtenen en realitzar l'experiència aleatòria (1,2, ...6) i la probabilitat de cada succés (0.166667), respectivament.

2. A la columna C3 volem simular que hem tirat el dau 200 vegades amb l'opció
> *Calc* ▸ *Random Data* ▸ *Discrete*

```
Generate:          200
Store in columns:  C3
Values in:         C1
Probabilities in : C2
```

3. Per a trobar les freqüències absoluta i relativa del succés "treure un 5", posem per cas, es pot fer servir la comanda
> *Stat* ▸ *Tables* ▸ *Tally*

```
Variables: C3
Display:   Counts
           Percents
```

Compareu-la amb la probabilitat que té el succés "treure un 5".

4. Feu l'histograma de freqüències
> *Graph* ▸ *Histogram...* per comprovar la relació de les freqüències de les dades generades amb la probabilitat.
5. Genereu ara 400 i 800 tirades de dau, observeu els tres histogrames de freqüència i comproveu que la freqüència relativa del succés "treure un 5" tendeix a la probabilitat del succés.
6. Torneu a fer ara el mateix procés però suposeu que el dau està trucat i que és més probable treure 5 que 3, en canvi la probabilitat de treure 1,2,4, o 6 és la mateixa. Per exemple, es pot escollir

$$P(\{1\}) = P(\{2\}) = P(\{4\}) = P(\{6\}) = 0.1, P(\{5\}) = 0.4 \text{ i } P(\{3\}) = 0.2.$$

Justifiqueu que l'anterior assignació de probabilitats és correcta i trieu d'altres opcions.

Problema 2. Probabilitat condicionada.

1. Carregueu el fitxer *peces.mtw*, que conté a la columna C1, els valors 1,2 o 3, depenent de l'empresa que ha fabricat unes peces i, a la columna C2, un 0 si la peça és defectuosa i un 1 en cas contrari. Volem obtenir la freqüència relativa del succés "una peça escollida a l'atzar és defectuosa". La comanda utilitzada a l'apartat 3 del Problema 1 pot ser útil per a resoldre la qüestió.

2. Suposem ara que s'ha escollit una peça a l'atzar i resulta ser defectuosa. Per tal d'estimar la probabilitat que té d'haver estat fabricada per l'empresa 2, podem fer la següent operació:

> Manip ▷ Copy Columns ▷

```
Copy from columns: C1
To columns:       C3
```

Use Rows ▷

```
⊙ Use rows with column: C2
equal to (eg, -4,5 -2:3 14): 0
```

Feu servir procediments semblants als de l'apartat anterior per estimar aquesta probabilitat condicionada.

3. Si es tria una altra peça i ha estat fabricada per l'empresa 2, estimeu la probabilitat que té de ser defectuosa.
4. Compareu el resultat obtingut a l'apartat 3 amb el que s'obtindria aplicant el Teorema de Bayes amb les dades de l'apartat 2 i les freqüències dels esdeveniments D (ja calculada a l'apartat 1) i E_i , on $E_i =$ "la peça ha estat fabricada per l'empresa i " i $D =$ "la peça és defectuosa".

Problema 3. Treball amb una mostra d'una variable aleatòria que segueix una distribució normal.

1. Carregueu el fitxer `sacsci.mtw`. Comproveu que les dades que s'han donat s'aproximen molt a una distribució normal fent
> Stat ▷ Basic Statistics ▷ Descriptive Statistics... ▷ Graphs ▷ Histogram of data with normal curve. Trobeu la mitjana i la desviació tipus d'aquesta mostra. (Recordeu que la desviació tipus que calcula el programa és la corregida.)
2. Suposem que els sacs de menys de 158 Kg no s'accepten. La següent comanda permet posar 0 i 1 a cada sac segons que sigui acceptable o no.
> Manip ▷ Code ▷ Numeric to Numeric...

```
Code data from columns: C1
Into columns:          C2
Original values:       158:250
New:                   1
Original values:       100:157.999
New:                   0
```

Trobeu el percentatge de sacs acceptats.

3. Feu servir la distribució normal que té implementada el programa MINITAB per a calcular el percentatge de sacs acceptats si el pes dels sacs segueix una distribució normal de paràmetres $\mu = 169.96\text{Kg}$ i $\sigma = 10.39$, que són els calculats a l'apartat 1. Aquest procés es realitza a través del menú
> *Calc* ▷ *Probability Distributions* ▷ *Normal...*

Cumulative probability
Mean : 169.96
Standar deviation: 10.39
 Input constant : 158
Optional storage: K1

Notem que l'opció

Cumulative probability

calcula $P(X \leq x_0)$, on x_0 és el valor que s'escriu a *Input constant*. Si es vol guardar aquest resultat en una constant, diguem-li *K1*, a l'apartat *Optional storage* cal escriure *K1*

S'obté el mateix resultat en escriure

MTB> CDF 158;

SUBC> Normal 169.96 10.39.

a la finestra *Session*. El resultat apareix escrit a la mateixa finestra. Si es vol guardar aquest resultat en una constant, *K1*, la primera instrucció ha de ser

MTB> CDF 158 K1;

En qualsevol cas, interpreteu el resultat obtingut i doneu la resposta correcta.

4. L'opció
> *Calc* ▷ *Probability Distributions* ▷ *Normal...*

Inverse cumulative probability

calcula $P(X \leq x_0) = p_0$, on p_0 és el valor que s'escriu a *Input constant*. Igualment el valor de x_0 es pot guardar en una constant, com abans. Feu-la servir per a trobar a partir de quin pes s'han d'acceptar els sacs si volem rebutjar-ne el 5%.

5. Trobeu quina probabilitat té un sac dels que estem estudiant, de pesar menys de 168 Kg.
6. Compteu el percentatge de sacs del fitxer *sacsci.mtw* que pesen entre 149.92 Kg i 190 Kg. Comproveu que es compleix el Teorema de Txebixev.
7. Suposem que es trien 8 sacs del fitxer *sacsci.mtw*. Quina és la probabilitat que n'hi hagi més d'un de NO acceptable? Es tracta d'una distribució

Binomial, de paràmetres $n = 8$ i $p = 0.133$ (que és la probabilitat que té un sac de ser no acceptable i que s'ha trobat a l'apartat 2). Les comandes
 > Calc ▷ Probability Distributions ▷ Binomial...

⊙ Cumulative probability
 Number of trials: 8
 Probability of success: 0.133
 ⊙ Input constant: 1
 Optional storage: (en blanc)

calculen $P(X \leq 1)$ quan X es una variable aleatòria $B(8,0.133)$. Feu-les servir per a determinar la probabilitat de què hi hagi més d'un sac de poc pes en un grup de vuit sacs.

8. Les comandes

> Calc ▷ Random Data ▷ Sample from columns

Sample: 8 rows
 from columns: C1
 Store samples in: C4
 Sample with replacement: (en blanc)

es poden fer servir per a simular que heu fet una extracció de 8 sacs i comptar quants se'n rebutgen, és a dir estudiar els valors que pren la variable aleatòria $X =$ "nombre de sacs rebutjats". Repetiu aquest procés 20 vegades (copiar i enganxar) i trobeu la freqüència relativa de $X > 1$. Compareu la resposta amb el resultat obtingut a l'apartat 7.

9. Suposem ara que s'encarreguen 35 sacs d'aquest ciment. Calculeu la probabilitat que hi ha de rebre'n exactament 10 que pesin menys de 168 Kg. Tenim una llei binomial amb $n = 35$ i $p = 0.43$ (el resultat trobat a l'apartat 5), llavors tal com s'ha calculat a l'apartat 7 però fent servir

⊙ Probability

en lloc de Cumulative probability (perquè?), s'obté

$$P(X = 10) = 0.0313.$$

10. Ara aproximem amb una llei normal, fent servir la correcció del mig punt. Tindrem

$$\mu = np = 35 * 0.43 = 15.05$$

$$\sigma^2 = np(1 - p) = 35 * 0.43 * (1 - 0.43) = 8.58 \rightarrow \sigma = 2.93$$

Llavors $P(X = 10) = P(9.5 < X < 10.5) = 0.0311$.

Problema 4. Variables aleatòries. Simulació d'una variable χ^2 .

En aquest problema es tracta de generar una mostra de 100 observacions o realitzacions d'una variable aleatòria que segueix una distribució χ^2 amb 9 graus de llibertat. Després es fa un estudi descriptiu de la mostra i es comparen els valors dels estadístics mostrals amb els corresponents paràmetres poblacionals obtinguts a partir de la distribució de probabilitat.

La forma de generar les realitzacions de la variable aleatòria χ^2 és la següent. Si escollim un nombre n gran ($n=1000$ en el nostre cas) de dígit (nombres de 0 a 9) aleatòriament, cal esperar que la freqüència d'ocurrència de cada un dels dígit, si realment han estat escollits aleatòriament, sigui $n/10$ ($E_i = 100, i = 0, 1, \dots, 9$ en el nostre cas). Si anomenem O_i a les freqüències observades efectivament, la variable aleatòria

$$\sum_{i=0}^9 \frac{(O_i - E_i)^2}{E_i}$$

que pren valors diferents cada vegada que es repeteix el procés, segueix una distribució de probabilitat χ^2 amb 9 graus de llibertat.

1. Ompliu les 10 primeres files de la columna C1 amb els 10 dígit de 0 a 9. Podeu fer-ho a mà, però també podeu utilitzar les següents opcions de menú:
> *Calc* ▸ *Make Patterned Data* ▸ *Simple Set of Numbers*

```
Store Patterned Data in:  C1
From First Value:         0
To Last Value:            9
In Steps of:              1
List Each Value:         1 times
List the Whole Sequence:  1 times
```

2. Ompliu les 10 primeres files de la columna C2 amb les freqüències esperades, $E_i = 100$, de cada un dels dígit si se'n trien 1000 a l'atzar. Podeu fer-ho a mà, però també podeu utilitzar les següents opcions de menú:
> *Calc* ▸ *Make Patterned Data* ▸ *Simple Set of Numbers*

```
Store Patterned Data in:  C2
From First Value:         100
To Last Value:            100
In Steps of:              1
List Each Value:         10 times
List the Whole Sequence:  1 times
```

3. Feu la tria de 1000 dígit a l'atzar i emmagatzemeu-los a la columna C3 amb les següents opcions de menú

> Calc ▸ Random Data ▸ Sample from Columns

```
Sample :                1000 rows
From columns :          C1
Store Samples in :     C3
Sample with replacement: (marqueu la creu)
```

4. Calculeu la freqüència amb què ha sortit cada dígit i guardeu aquests valors observats O_i a la columna C4. L'opció de menú

> Stat ▸ Tables ▸ Tally

no permet emmagatzemar els resultats en una columna. Per a això, aneu a la línia de comandes i feu

```
MTB> Tally C3 ;
```

```
SUBC> Store C1 C4 .
```

que col·locarà les categories 0-9 a la columna C1 (on ja hi eren) i la freqüència corresponent a la columna C4.

5. Feu el càlcul de

$$\sum_{i=0}^9 \frac{(O_i - E_i)^2}{E_i}$$

i guardeu el resultat a la primera fila de la columna C5. A la línia de comandes feu

```
MTB> Let C5(1) = Sum((C4-C2)**2/C2)
```

6. Ara tenim una realització d'una variable aleatòria χ_9^2 a la primera fila de la columna C5 i volem repetir el procés, des de la tria de les 1000 xifres, 99 vegades per a tenir 100 realitzacions d'aquesta variable aleatòria a la columna C5. Per a fer-ho automàticament, executarem la MACRO KHI2.MAC des de la línia de comandes fent

```
MTB> %F:\soft\minitab\data\KHI2.MAC
```

Aquesta MACRO és un petit programa que conté les següents comandes

```
DO K1=2:100
```

```
SAMPLE 1000 C1 C3;
```

```
REPLACE.
```

```
TALLY C3;
```

```
STORE C1 C4.
```

```
LET C5(K1) = SUM((C4-C2)**2/C2)
```

```
ENDDO
```

7. Feu un estudi descriptiu i dibuixeu l'histograma de la columna C5 amb les opcions de menú

> Stat ▸ Basic Statistics ▸ Descriptive Statistics...

Compareu la mitjana i la desviació tipus de la mostra obtingudes en l'estudi descriptiu, amb la mitjana i la desviació tipus d'una variable χ^2 amb 9 graus de llibertat. Compareu l'histograma de la mostra amb la gràfica de la funció de densitat d'una variable χ^2 amb 9 graus de llibertat, que es mostra a la Figura 1.

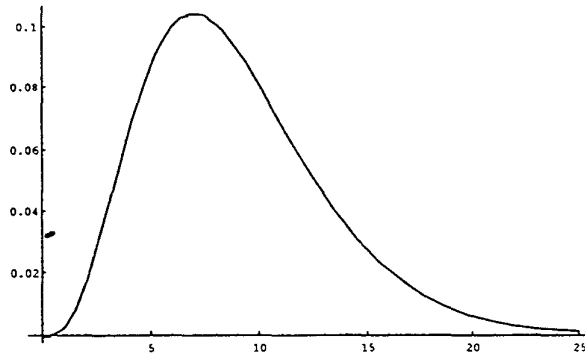


Figura 1: Gràfica de la funció de densitat χ^2 amb 9 graus de llibertat.

8. Compareu les quartils Q1 i Q3, obtingudes en l'estudi descriptiu de la mostra, amb els valors χ_a^2 i χ_b^2 de la variable X que segueix una llei χ^2 amb 9 graus de llibertat, tals que $P(X < \chi_a^2) = P(X > \chi_b^2) = 0.25$. Per a això feu

```
> Calc ▸ Probability Distributions ▸ Chisquare
  ⊙ Inverse cumulative probability
  Degrees of freedom:          9
  ⊙ Input Constant:           0.25
  Optional Storage:           K1
```

```
> Calc ▸ Probability Distributions ▸ Chisquare
```

```
  ⊙ Inverse cumulative probability
  Degrees of freedom:          9
  ⊙ Input Constant:           0.75
  Optional Storage:           K2
```

```
MTB>Print K1 K2
```

EXERCICIS

- El temps de combustió d'una espelma es considera que segueix una distribució normal $N(5.25 \text{ min}; 45 \text{ seg})$. Trobeu la probabilitat que té una espelma de cremar-se en

- menys de 4.35 minuts.
 - entre 4.26 i 5.67 minuts.
 - més de 5 minuts
2. Una persona tira una ametlla enlaire, i li va a parar a la boca amb una probabilitat de $2/3$. Suposem que les tirades són independents. Quina probabilitat hi ha que en 6 tirades
- les encerti totes 6.
 - n'encerti menys de 4.
 - n'encerti com a mínim 3.
 - n'encerti com a màxim 4.
3. En les proves d'edició d'un llibre de 250 pàgines, s'ha trobat un nombre d'errades per pàgina recollit a la columna C1 del fitxer `Poisson.mtw`.
- Feu un estudi descriptiu d'aquesta mostra i dibuixeu-ne l'histograma.
 - Quin nombre exacte de pàgines hi ha amb 2 errades?
 - Quin nombre exacte de pàgines hi ha amb més de 2 errades?
 - Quin nombre exacte de pàgines hi ha amb 1,2 o 3 errades?
 - Responen les mateixes preguntes però donant el nombre aproximat calculat a partir d'una distribució de Poisson amb paràmetre λ igual a la mitjana obtinguda en l'estudi descriptiu del primer apartat. Es pot dir que la mostra del fitxer `Poisson.mtw` ho és d'una variable aleatòria poissoniana?
4. Problema 12 de la pàgina 137 del llibre.

PRÀCTICA 3: ESTADÍSTICA INDUCTIVA

OBJECTIUS

Amb aquesta pràctica ens proposem els següents objectius:

- Il·lustrar, mitjançant simulació, la distribució mostral dels estadístics mitjana mostral i el vinculat a la desviació tipus mostral, $\frac{ns_n^2}{\sigma^2}$, així com les seves propietats.
- Estimar per intervals de confiança la mitjana d'una mostra d'una distribució Normal, tenint en compte dues possibilitats: en un cas coneixerem la desviació tipus de la població i en l'altre, aquest paràmetre serà desconegut.
- Decidir contrastos d'hipòtesi sobre la mitjana poblacional d'una distribució. També, com abans, suposarem en un cas que coneixem la desviació tipus de la població i en l'altre, que aquest paràmetre no el coneixem.
- Estimar proporcions per intervals de confiança i estudiar la relació entre la grandària de la mostra i l'error màxim permès en l'estimació.

DADES

Les macros `est10.mac`, `est25.mac`, `est100.mac` i `est400.mac`, que estan agrupades en la macro `estn.mac`, permetran el mostreig automàtic, per a diferents grandàries mostrals, d'una distribució Normal.

Les macros `intconfz.mac` i `intconfn.mac` calculen intervals de confiança i valoren la seva qualitat.

El fitxer `opinio.mtw` mostra a la columna C1 l'opinió favorable (codificat amb el valor 1) o desfavorable (codificat amb el valor 0) sobre el *pay-per-view* de les transmissions esportives d'un municipi de 5000 persones.

DESENVOLUPAMENT

Problema 1. Distribució mostral de \bar{X}_n i de $\frac{ns_n^2}{\sigma^2}$.

1. Comenceu una sessió nova de MINITAB. Etiqueteu la columna C1 amb el nom Mostra i genereu-hi una mostra de 10 observacions d'una variable Normal de mitjana 215 i desviació tipus 10, utilitzant l'opció
> Calc ▷ Random Data ▷ Normal...

```
Generate:          10
Store in columns:  C1
Mean:              215
Standard deviation: 10
```

2. Etiqueteu les columnes C2, C3, C4, C5, C6 amb els nom Mtjn10, MtjMm10, Stdevc10, Stdev10, Khi9 i assigneu a la primera fila de cadascuna els valors corresponents de les variables aleatòries Mitjana, Mitjana dels valors Màxim i Mínim, Desviació tipus corregida, Desviació tipus i $\frac{ns_n^2}{\sigma^2}$, respectivament, sobre la mostra generada. Des de la finestra *Session* podeu utilitzar ordres com ara: MTB> LET C2(1)=MEAN(C1) ...

3. A l'apartat anterior hem conseguit només una observació de les variables aleatòries esmentades. Com que ens interessa poder veure com és la distribució d'aquestes variables, convé repetir el procés anterior un suficient nombre de vegades per a poder il·lustrar la seva distribució.

La macro est10.mac permet obtenir 500 observacions de les esmentades variables per a una grandària mostral de 10 individus. Vegeu la pàgina 18 per recordar com es carreguen les macros. Executeu-la i observeu el resultat.

4. Anàlogament, les macros est25.mac, est100.mac i est400.mac fan el mateix per a grandàries de 25, 100 i 400 individus. Les 4 macros anteriors s'executen successivament amb la macro estn.mac. Obtingueu 500 observacions, de cadascuna de les variables esmentades, per a cadascuna de les grandàries mostrals, executant la macro estn.mac.
5. Deseu el *Worksheet* actual amb el nom estn.mtw.
6. Obtingueu simultàniament els estadístics bàsics de les variables Mitjanes mostrals Mtjn10, Mtjn25, Mtjn100 i Mtjn400, amb els corresponents histogrames i l'ajustament Normal. Verifiqueu gràficament la Normalitat dels histogrames obtinguts.
7. Observeu les mitjanes i desviacions tipus corregides que s'han obtingut en el resum de resultats. Comproveu que s'ajusten als paràmetres μ i $\frac{\sigma}{\sqrt{n}}$, respectivament.

Tanqueu les finestres gràfiques.

8. Per a il·lustrar la variació de la desviació tipus de les corresponents distribucions, refeu els histogrames anteriors utilitzant l'opció
> *Graph* ▸ *Histogram...* i escollint successivament
> *Frame* ▸ *Multiple Graphs...*

- Each graph on a separate page :
- Same X and same Y :

i

> *Options...*

- Density :
- MidPoint :
- Automatic :

Tanqueu les finestres gràfiques.

9. Refeu els tres apartats anteriors per a les variables aleatòries $MtjMm10$, $MtjMm25$, $MtjMm100$ i $MtjMm400$, i analitzeu la normalitat dels histogrames resultants així com el comportament dels estadístics mitjana i desviació tipus.
10. Feu l'Estadística Descriptiva de les variables aleatòries $Khi9$, $Khi25$, $Khi99$ i $Khi399$. Comproveu que els estadístics resultants corresponents a la mitjana i a la desviació tipus corregida s'ajusten als descrits a la teoria. Obtingueu els histogrames de les variables $Khi9$ i $Khi24$. Feu el mateix per a les variables $Khi99$ i $Khi399$.

Problema 2. Interval de confiança per a la mitjana (σ coneguda)

1. Suposem que la distribució del diàmetre dels cargols produïts per una màquina segueix una distribució Normal de paràmetres $\mu = 12mm$ i amb una desviació tipus de $0.5mm$. Genereu 1000 observacions d'aquesta variable i guardeu-les a la columna C1. Guardeu les dades en el fitxer `intconf.mtw`.
2. Justifiqueu la hipòtesi de Normalitat de la variable a partir del corresponent histograma amb l'ajustament Normal i dels estadístics obtinguts.
3. Suposem que desconeixem el valor real de μ i que el volem estimar a partir de la mateixa producció, però sabem que σ és igual a $0.5mm$. Per motius econòmics resulta que només podem prendre periòdicament una mostra de grandària 25.

Obtingueu una mostra de 25 cargols d'entre tota la producció i guardeu-la a la columna C2.

A partir d'aquesta mostra volem calcular un interval de confiança del 95% per a la mitjana poblacional. Per a fer-ho seguim els següents passos:

a) Emmagatzemeu a la constant K1 el valor de l'estadístic Z_α per al nivell de confiança esmentat. Visualitzeu el valor obtingut.

b) Calculeu l'extrem esquerre de l'interval de confiança i guardeu-lo a la casella C3(1) mitjançant la comanda:

```
MTB> LET C3(1)=MEAN(C2)-K1*0.5/SQRT(25)
```

c) Feu el mateix per a l'extrem dret de l'interval i guardeu-lo a la casella C4(1).

d) Verifiqueu els resultats obtinguts als dos apartats anteriors utilitzant la comanda:

```
> Stat > Basic Statistics > 1-Sample Z...
```

Variables :	C2
Confidence Interval :	
Level :	95
Sigma :	0.5

e) Com que coneixem el valor real de la mitjana poblacional ($\mu = 12mm$) podem valorar l'interval de confiança proposat, assignant un 1 a la casella C5(1) si aquest valor real pertany a l'interval, i un 0 en cas contrari. Per a fer-ho utilitzeu l'operador lògic I, que en MINITAB s'expressa pel símbol &, entre les condicions $C3(1) < 12$ i $12 < C4(1)$ i la comanda d'assignació:

```
MTB> LET C5(1)=(C3(1) < 12) & (12 < C4(1))
```

4. Repetiu l'apartat anterior per a una confiança del 99%, utilitzant la constant K2 i les columnes C6, C7 i C8.
5. Per tal d'il·lustrar el paper que juga el nivell de confiança, repetirem els càlculs dels anteriors intervals de confiança 100 vegades i farem un recompte del nombre d'uns obtinguts. Aquests valors haurien d'estar a l'entorn del 95 i 99, respectivament.

La macro `intconfz.mac` calcula els 100 valors dels corresponents intervals de confiança de forma automàtica. Executeu la macro i calculeu el nombre de 1's obtinguts. Comproveu que s'ajusten als esperats. Salveu les dades obtingudes.

Problema 3. Intervals de confiança per a la mitjana (σ desconeguda)

1. Aquest problema és la continuació de l'anterior, però en aquest cas no suposarem coneguda la desviació tipus poblacional (σ). Carregueu, si s'escau, el fitxer `intconf.mtw`.

2. Com abans, obtingueu a C2 una mostra de 25 cargols del total de la producció.
3. A partir d'aquesta mostra, ens continua interessant calcular un interval de confiança del 95% per a la mitjana poblacional. Com que no coneixem la desviació tipus de la població, calculeu-ne a K3 una estimació puntual no esbiaixada.
A diferència del problema anterior, la distribució mostral de referència és una *t*-Student de 24 graus de llibertat. Procediu d'acord amb els passos:
4. Emmagatzemeu a la constant K1 el valor del corresponent estadístic t_α per al nivell de confiança esmentat. Visualitzeu el valor obtingut.
5. Calculeu l'extrem esquerre de l'interval de confiança i guardeu-lo a la casella C9(1) mitjançant la comanda:
MTB> LET C9(1)=MEAN(C2)-K1*K3/SQRT(25)
6. Repetiu la resta de passos del problema anterior, utilitzant les variables i constants corresponents, la comanda
> Stat ▸ Basic Statistics ▸ 1-Sample t...

Variables :	C2
Confidence Interval :	
Level :	95

i la macro intconft.mac

Problema 4. Estimació de proporcions.

1. Recupereu el fitxer opinio.mtw. Simuleu a les columnes C2, C3 i C4, respectivament, una enquesta d'opinió de grandària 100, 400 i 1600, sobre el conjunt del municipi.
2. Ens interessa un interval de confiança al 95% per a l'estimació de la proporció real d'opinió favorable. Considerarem el tres supòsits descrits a la teoria:
 - a) El cas de màxima indeterminació ($p = q = 0.5$). En aquest és suficient utilitzar la comanda
> Stat ▸ Basic Statistics... ▸ 1-Sample Z i introduir els paràmetres corresponents.
Calculeu l'interval de confiança resultant en aquest cas.
 - b) El cas que utilitzem una proporció "a priori" coneguda. En aquest és suficient utilitzar la comanda
> Stat ▸ Basic Statistics... ▸ 1-Sample Z calculant previamente el corresponent valor de la desviació tipus mostral.
Calculeu l'interval de confiança resultant en aquest cas.

c) El cas que utilitzem per a l'estimació de la desviació tipus mostral la proporció observada a la mostra. Aquest mateix procés és el que calcula la comanda

> *Stat* ▷ *Basic Statistics...* ▷ *1-Sample t.*

Calculeu l'interval de confiança resultant en aquest cas.

d) Comproveu que els tres anteriors intervals de confiança són els descrits a la teoria.

3. Repetiu l'apartat anterior per a les grandàries mostrals 400 i 1600, i deduiu la relació que hi ha entre ells pel que fa a les amplituds dels intervals.
4. Quina ha de ser la grandària de la mostra perquè l'error en l'estimació no superi, en tots els supòsits, el 4%, amb un nivell de confiança del 95%? Obtingueu a C5 una mostra d'aquesta grandària, estimeu els corresponents intervals de confiança i comproveu que s'ajusten a l'error prefixat.

EXERCICIS

1. La secció de control de qualitat d'una empresa de fabricació de formigó per a constructores mostreja aleatòriament 30 cubes, i de les provetes fetes amb el corresponent formigó en calcula la resistència a compressió. Els resultats obtinguts, en Kg/cm^2 , són els següents:

206	171	203	234	226	193	191	217	203	197
216	217	227	213	209	223	203	206	211	206
184	207	203	211	209	241	200	188	205	190

- Sabent que la distribució de la resistència és Normal, calculeu un interval de confiança per a l'estimació de la mitjana de tota la producció, amb un nivell de significació del 5%
 - El cap de producció ens informa que totes les cubetes de la mostra transportaven formigó de resistència mitjana $215 Kg/cm^2$. Treballant amb un nivell de confiança del 99%, estudeu si està d'acord aquesta afirmació amb les dades observades.
2. Es mesura la distància entre dues ciutats amb un aparell que ofereix lectures en Km que segueixen una distribució Normal de desviació tipus $0.75 Km$. El resultat de fer 10 lectures és:

215.2	216.3	215.0	217.0	214.9
215.4	216.1	215.7	216.2	216.8

- Determineu un interval de confiança per a la distància real entre les ciutats amb una confiança del 95%.

- Quina és la probabilitat que la distància real estigui fora de l'interval resultant de l'apartat anterior.

En un mapa de carreteres trobem que la distància en qüestió és de 216.3 Km.

Amb un nivell de significació del 5%,

- feu un contrast d'hipòtesi per a decidir si la distància és efectivament 216.3 Km o no ho és,
- decideu si hi ha motius suficients per a rebutjar la mesura del mapa a favor de pensar que la mesura real és més petita que la que aquest proposa.

3. Es realitza un estudi per saber la proporció d'estudiants d'arquitectura tècnica que troben feina en finalitzar els seus estudis. D'una mostra aleatòria de 144 estudiants, 115 troben feina.

- Feu una columna que contingui, codificat amb 0's i 1's, les dades de l'estudi.
- Trobeu l'interval de confiança del 95% per a l'estimació de la proporció d'estudiants que troben feina en acabar els seus estudis.
- Quina ha de ser la grandària de la mostra perquè l'error en l'estimació no superi el 4%?

4. Problema 15 de la pàgina 194 del llibre.

Bibliografia

Ferrer, A. i altres. *Fonaments d'Estadística Aplicada.*
Ed. Els autors. Barcelona, 1995

Clarke, G.M. i Cooke, D. *A basic course in Statistics.*
Ed. Edward Arnold. London, 1989

Peña, D. *Estadística. Modelos y métodos (Vol. I i II).*
Ed. Alianza Universidad Textos. Madrid, 1992

Ras, A. *Estadística aplicada per a enginyeria.*
Ed. Edicions UPC. Barcelona, 1993

MINITAB *Reference Manual.*
Ed. Minitab Inc. State College, PA, USA, 1996
