Exercicis de Probabilitat i Estadística amb l'ajuda de MINITAB

Santiago Forcada, Immaculada Gálvez, Josep Gibergans,
Maria José Jiménez, Víctor Mañosa, Enric Monsó.

Graus d'enginyeria industrial
Escola d'Enginyeria de Terrassa

Índex

NOTA INICIAL

El que teniu a les mans és una col·lecció d'exercicis corresponents a l'assignatura de Probabilitat i Estadística dels graus d'enginyeria industrial de l'Escola d'Enginyeria de Terrassa.

Molts professors i professores del Departament de Matemàtica Aplicada III els han anat compilant al llarg dels anys. La novetat d'aquest recull és la introducció de diverses anotacions per poder resoldre els exercicis amb l'ajut de MINITAB.

Si detecteu alguna errada us estarem agraïts si ens el comuniqueu a qualsevol dels professors i professores de l'assignatura.

Santiago Forcada, Immaculada Gálvez, Josep Gibergans,
Maria José Jiménez, Víctor Mañosa i Enric Monsó.
Departament de Matemàtica Aplicada III
Escola d'Enginyeria de Terrassa

1 ESTADÍSTICA DESCRIPTIVA

1.1 Exercicis

1. Els preus de 21 pisos en venda en una determinada zona venen donats per les següents dades (en milers d'euros): Preu= {213, 222, 240, 155, 261, 251, 221, 212, 235, 241, 253, 240, 279, 223, 231, 242, 234, 252, 235, 243, 248}

   (a) Feu el diagrama de tija i fulles.

   (b) Feu la següent la següent taula de freqüències (al vostre full).

Classe	Freq. Abs.	Freq. Rel.	Freq. Rel. Ac.
Preu$\leq$210			
Preu$\in$(210,220]			
Preu$\in$(220,230]			
Preu$\in$(230,240]			
Preu$\in$(240,250]			
Preu$\in$(250,260]			
Preu$\in$(260,270]			
Preu$\in$(270,280]			

   Quin és el percentatge dels pisos arriba, com a molt, a un preu de 240.000 Euros?

   (c) Indiqueu quins són els valors dels quartils Q_1 i Q_3, la mediana, i el rang interquartil·lic.

   (d) Feu el Diagrama de Caixa, i indiqueu quines dades poden ser considerades anomalies moderades, o extremes.

   (e) Feu servir la calculadora per determinar la *mitjana* $\bar{X}$ i la *desviació típica* S_{n-1}.

2. A una sabateria de Barcelona s'ha fet un estudi sobre les vendes de sabates per senyors distribuïdes per talles durant el mes passat. Els resultats s'han recollit a una taula però s'han esborrat accidentalment. Disposem per tant només de les dades següents:

Talla (X)	Freqüència	Freqüència acumulada	Freqüència relativa	Freqüència relativa acumulada
37				0.01
38		25	0.015	
39		70		
40	234			
41	366			
42	229			
43				0.97
44				
45			0.01	

   (a) Completeu les dades que falten a la taula anterior indicant quins raonaments heu fet.

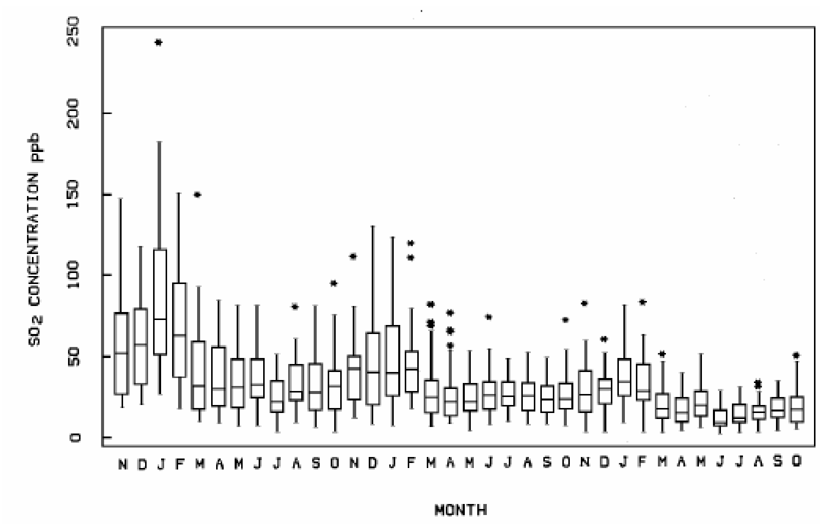
   (b) Representeu el polígon de freqüències absolutes.

   (c) Trobeu la mitjana, mediana i moda.

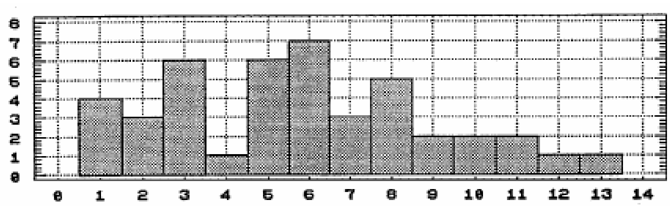
   (d) Trobeu la desviació típica, la variancia i el coeficient de variació.

3. Les dades referides al nombre d'alertes ambientals produïdes l'any 2003 a 20 països són: $\{93, 105, 116, 110, 109, 87, 117, 104, 116, 131, 111, 90, 92, 102, 105, 110, 92, 116, 99, 117\}$.
   (a) Construïu la taula de freqüències *relatives acumulades*, prenent classes de longitud 10, començant per 80 i acabant per 140. Quin percentatge de països superen les 110 alertes ambientals?
   (b) Feu l'histograma.
   (c) Feu el diagrama de tija i fulles. Indiqueu els quartils, la mediana, i el rang interquartil·lic.
   (d) Feu el Diagrama de Caixa, i indiqueu quines dades poden ser considerades anomalies moderades, o extremes.
   (e) Feu servir la calculadora per determinar la *mitjana* $\bar{X}$ i la *desviació típica* S_{n-1}.

4. En el gràfic següent, cada BoxPlot correspon a les mesures diàries, preses al llarg d'un mes, de concentració (en ppb) de SO_2 en l'atmosfera de la ciutat de Nova York des d'el novembre de 1969 fins l'octubre de 1972.
   (a) En quins mesos la concentració de SO_2 ha estat més variable.
   (b) En quins mesos, més de la meitat dels dies s'ha mesurat una concentració superior a 50 ppb?
   (c) Com s'explica que en els mesos de març i abril de 1971 es considerin dades anòmales concentracions que als mesos de gener i febrer de 1970 són considerades del tot normals?
   (d) L'efecte de l'estacionalitat té una influència important en la concentració de $S0_2$?
   (e) Quina conclusió podem extreure de la comparació dels resultats de l'evolució de la concentració de $S0_2$?



5. El següent diagrama de barres correspon a la representació gràfica de la taula de freqüències de la variable que comptabilitza el número d'avaries diàries en una planta industrial comptabilitzades al llarg de n dies.



(a) Quin és el número de dies n?

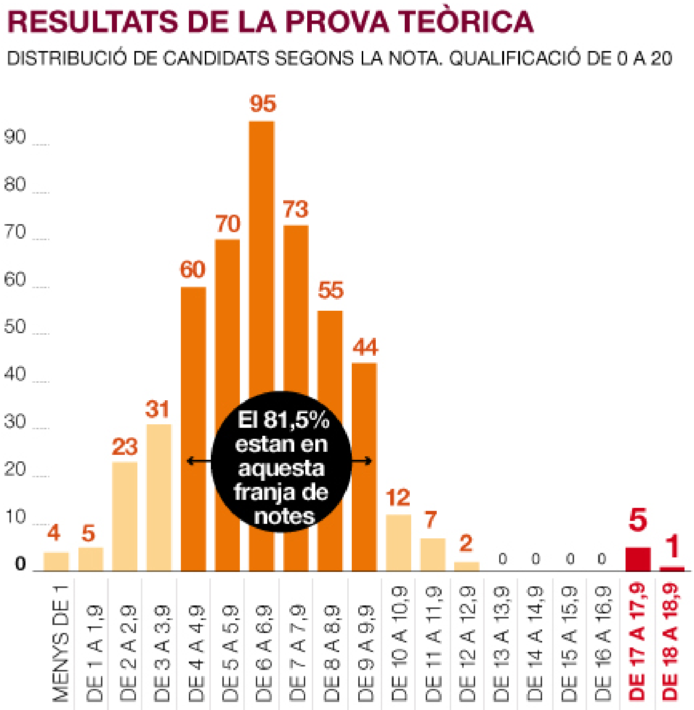
(b) Construïu la taula de freqüències relatives acumulades.

(c) Calculeu els quartils Q_1, Q_2 i Q_3.

(d) Construïu el Boxplot.

(e) Quin és el percentatge de dies en els que s'han produït més de 8 avaries?

6. En una ocasió un diari* va publicar aquest "histograma" amb els resultats de les notes d'un concurs públic (les notes van de 0 a 20). Es demana:



(a) Calculeu els quartils, la mediana, el rang i el rang interquartílic.

(b) Feu el diagrama de caixa i indiqueu quines dades poden ser considerades anomalies moderades, o extremes.

Indicació: considereu que totes les dades estan representades per la seva marca de classe: $0.5, 1.5, 2.5, ..., 19.5$*, i les classes són en realitat intervals semioberts* $[k, k+1)$.

7. (a) Tenim un fitxer amb els sous dels 500 empleats de l'empresa TRD. Sabem que la mitjana dels sous mensuals és de 2000 euros amb una variància de 400. Digueu si són certes o falses les següents afirmacions i expliqueu breument el vostre raonament:

   (a1) Si augmentem tots els sous un 10%, aleshores la mitjana serà de 2200 euros.

   (a2) Si augmentem tots els sous un 10%, aleshores la variància serà de 440.

   (a3) Si augmentem tots els sous un 10%, aleshores la moda no variarà.

   (a4) Si decidim un augment lineal de 100 euros, aleshores la variància serà superior a 400.

   (a5) Si decidim augmentar 300 euros només el sous del 100 empleats que més cobren, aleshores la mediana no variarà.

*El Periódico, 3/10/2014.

(b) El treballador A té un sou de 2000 euros a una empresa on la mitjana dels sous és de 1800 euros i la desviació típica de 200 euros. El treballador B té un sou de 2400 euros a una empresa on la mitjana dels sous és de 2100 euros i la desviació típica de 400 euros. Si els graus de responsabilitat dins de les empreses són similars, quin dels dos està més ben pagat?

(c) A l'empresa TRD2 agafem una mostra de 400 empleats. La mitjana dels seus sous és de 1800 euros i la desviació típica de 100 euros. Demostreu que més del 75% dels treballadors tenen un sou entre 1600 i 2000 euros.

8. D'una determinada mostra sabem que la seva mitjana és $\bar{x} = 40.5$, i la seva desviació típica és $s = 3.1$. Per quin dels següents valors del paràmetre δ podem afirmar que més del 80% de les observacions estan a l'interval $[40.5 - \delta\,,\, 40.5 + \delta]$:

   (A) $\delta = 2.236$, (B) $\delta = 6.932$, (C) $\delta = 2.48$, (D) cap dels anteriors.

9. Es vol construir tubs cilindrics per envasar un lot de N boles d'acer. S'ha calculat la mitjana i la variança de les longituds dels radis d'aquestes boles, que expressats en mil·limetres són: $\bar{X} = 30$ i $S^2 = 9$. Calculeu el radi mínim que han de tenir els tubs per tal que es puguin envasar el 75% de les boles.

10. Durant 20 dies s'ha fet un estudi del promig de tares que contenen les peces de camiseria que produïm, obtenint les dades següents:

$$\{9.6, 7.1, 3.5, 4.4, 0.5, 1.3, 3.1, 4.3, 5.5, 6.0, 15.0, 8.5, 9.6, 6.3, 6.1, 4.5, 4.4, 5.3, 5.6, 5.9\}.$$

   a) Feu el diagrama de tija i fulles. Indiqueu els quartils, la mediana, i el rang interquartil.lic.

   b) Feu el Diagrama de Caixa, i indiqueu quines dades poden ser considerades anomalies moderades, o extremes.

1.2 Solucions

1. (a)

15	**5**					
16						
17						
18						
19						
20						
21	**3**	**2**				
22	**2**	**1**	**3**			
23	**5**	**1**	**4**	**5**		
24	**0**	**1**	**0**	**2**	**3**	**8**
25	**1**	**3**	**2**			
26	**1**					
27	**9**					

(b)

Classe	Freq. Abs.	Freq. Rel.	Freq. Rel. Ac.
Preu$\leq$210	**1**	**0.0476**	**0.0476**
Preu$\in$(210,220]	**2**	**0.0952**	**0.1428**
Preu$\in$(220,230]	**3**	**0.1429**	**0.2857**
Preu$\in$(230,240]	**6**	**0.2857**	**0.5714**
Preu$\in$(240,250]	**4**	**0.1904**	**0.7618**
Preu$\in$(250,260]	**3**	**0.1429**	**0.9047**
Preu$\in$(260,270]	**1**	**0.0476**	**0.9523**
Preu$\in$(270,280]	**1**	**0.0476**	**0.9999$\simeq$ 1**

El percentatge de pisos els preus dels quals arriba, com a molt, a 240 és el 57.14%.

(c)

							f	f_{ac}
15	**5**						1	1
16							0	1
17							0	1
18							0	1
19							0	1
20							0	1
21	**3**	**2**					2	3
22	**2**	**1**	**3**				3	6
23	**5**	**1**	**4**	**5**			4	10
24	**0**	**1**	**0**	**2**	**3**	**8**	6	16
25	**1**	**3**	**2**				3	19
26	**1**						1	20
27	**9**						1	21

El Q_1 està en la posició 6, la mediana en la posició 11 i el Q_3 en la posició 16. Directament del diagrama de tija i fulles tenim: $Q_1 = 223$, $M = 240$, $Q_3 = 248$.

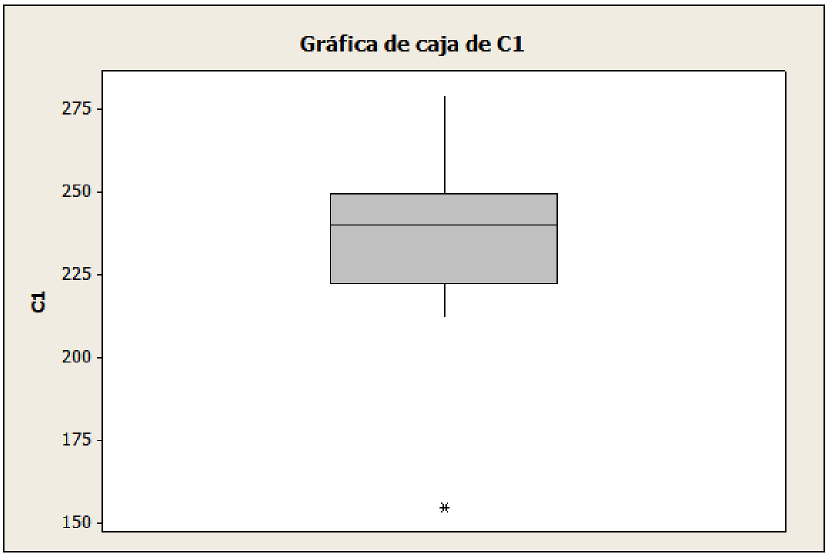
(d) Donat que el $RIQ = Q_3 - Q_1 = 248 - 223 = 25$ les fronteres de la zona de normalitat vénen donandes per

$$\begin{aligned} f_1 &= Q_1 - 1.5RIQ = 185.5 \\ f_3 &= Q_3 + 1.5RIQ = 285.5 \\ F_1 &= Q_1 - 3RIQ = 148 \\ F_3 &= Q_3 + 3RIQ = 323 \end{aligned}$$

Observem que les fronteres de la zona d'anormalitat extrema estan fora del rang dels valors observats i no les tindrem en compte. Pel que fa a les últimes dades normals estadísticament (els adjunts) tenim que

$$\begin{aligned} \text{adj}_1 &= \text{mínima obs. a } [f_1, Q_1] = 212 \\ \text{adj}_3 &= \text{màxima obs. a } [Q_3, f_3] = 279 \end{aligned}$$

Per tant



Observem que no tenim cap anomalia estadística extrema i una de moderada 155.

(e) Una vella calculadora Casio $fx - 180P$ de començament dels anys 80 dóna: $\bar{X} = 234.8095$, $S = S_{n-1} = \sigma_{n-1} = 24.2768$, que coincideix amb el resultats de Minitab.

2.

(a) Per iniciar el procés podeu fixar-vos en el següent: Si diem $X = f_a(37)$, $Y = f_a(38)$ i N el total de sabates. Aleshores

$$\begin{cases} X + Y = 25 \\ X/N = 0.01 \\ Y/N = 0.015 \end{cases}$$

Per tant dividint les darreres equacions obtenim

$$\frac{Y}{X} = 1.5 \Rightarrow Y = 1.5\,X.$$

Substituint ara aquesta expressió a la primera equació tenim

$$X + 1.5X = 2.5X = 25 \Rightarrow X = 10,\ Y = 15 \text{ i } N = 1000.$$

A partir d'aquí la resta surt directament.

Talla (X)	Freqüència	Freqüència acumulada	Freqüència relativa	Freqüència relativa acumulada
37	**10**	**10**	**0.01**	0.01
38	**15**	25	0.015	**0.025**
39	**45**	70	**0.045**	**0.07**
40	234	**304**	**0.234**	**0.304**
41	366	**670**	**0.366**	**0.67**
42	229	**899**	**0.229**	**0.899**
43	**71**	**970**	**0.071**	0.97
44	**20**	**990**	**0.02**	**0.99**
45	**10**	**1000**	0.01	**1**

(b) Trivial.

(c) La mediana s'obté de la següent manera:

$$M = \frac{\text{Pos}_{500} + \text{Pos}_{501}}{2} = \frac{41 + 41}{2} = 41.$$

Com que les dades vénen agregades, tenim:

$$\bar{X} = \sum_{\text{classes}} x_i \,\text{fr}(x_i) = 37 \cdot 0.01 + 38 \cdot 0.015 + \cdots + 45 \cdot 0.01 = 41.062.$$

La moda és 41

(d) Com que les dades vénen agregades tenim:

$$S_{n-1}^2 = \frac{\sum_{\text{classes}} x_i^2 \,\text{f}_\text{i}}{n-1} - \frac{n}{n-1}\left(\bar{X}\right)^2.$$

On f_i aquí és freqüència absoluta. Per tant,

$$\sum_{\text{classes}} x_i^2 \,\text{fi} = 37^2 \cdot 10 + \cdots 45^2 \cdot 10 = 1687646.$$

Per tant:

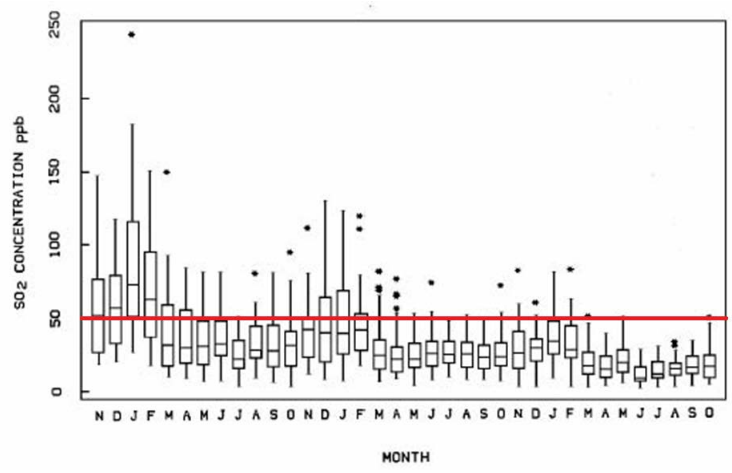
$$S_{n-1}^2 = \frac{1687646}{999} - \frac{1000}{999}(41.062)^2 = \mathbf{1.559715} \Rightarrow S_{n-1} = \mathbf{1.248885503}.$$

El $CV = \frac{S_{n-1}}{\bar{X}} = \frac{1.248885503}{41.062} = \mathbf{0.0304146}$, el que indica que les dades presenten una variació del 3% respecte de la mitjana.

3. La filosofia és la mateixa que la de l'exercici 1. Fer el diagrama de tija i fulles i a partir d'aquí, obtenir el boxplot.

4. (a) Els quatre primers mesos de l'estudi, per exemple.

   (b) Mireu en el grafic següent els mesos en els que la mediana supera 50ppb.



(c) Perquè les anomalíes estadístiques que detecta el Boxplot ho són només respecte del seu propi grup de dades.

(d) Si, observeu les petites oscil·lacions.

(e) Hi ha una tendència a una reducció i una menor dispersió.

5. (a) $n = 43$.

(b) Directament del gràfic de l'enunciat tenim

Avaries	Freq.	Freq. Relativa	Freq. Acumulada	Freq. Relativa Acumulada	Freq. Relativa Acumulada 100%
0	0	0,00	0	0,00	0,00
1	4	0,09	4	0,09	9,30
2	3	0,07	7	0,16	16,28
3	6	0,14	13	0,30	30,23
4	1	0,02	14	0,33	32,56
5	6	0,14	20	0,47	46,51
6	7	0,16	27	0,63	62,79
7	3	0,07	30	0,70	69,77
8	5	0,12	35	0,81	81,40
9	2	0,05	37	0,86	86,05
10	2	0,05	39	0,91	90,70
11	2	0,05	41	0,95	95,35
12	1	0,02	42	0,98	97,67
13	1	0,02	43	1,00	100,00
Total:	43				

(c) El Q_1 és la mitjana de la posició 11 i 12 , la mediana està en la posició 22 i el Q_3 és la mitjana de la posició 32 i 33. Directament de la columna de freqüencia relativa acumulada tenim: $Q_1 = 3$, $M = 6$, $Q_3 = 8$.

(d) Donat que el $RIQ = Q_3 - Q_1 = 5$ les fronteres de la zona de normalitat vénen donandes per

$$\begin{aligned} f_1 &= Q_1 - 1.5RIQ = -4.5 \\ f_3 &= Q_3 + 1.5RIQ = 15.5 \\ F_1 &= Q_1 - 3RIQ = -12 \\ F_3 &= Q_3 + 3RIQ = 23 \end{aligned}$$

Observem que totes les fronteres estan fora del rang dels valors observats i no les tindrem en compte. Totes les dades són estadísticament normals i els adjunts, doncs, són les observacions mínimes i màximes respectivament.

$$\begin{aligned} \text{adj}_1 &= 1 \\ \text{adj}_3 &= 13 \end{aligned}$$

(e) el percentatge de dies amb més de 8 avaries és el complementari de dies amb núm. menor o igual a 8 avaries, és a dir $(100 - 81)\% = 19\%$.

6. (a) Disposem d'un conjunt $X = \{x_1, x_2, \cdots, x_{487}\}$ de $n = 487$ dades, per tant la mediana correspondrà a la dada en la posició $\frac{n+1}{2} = 244$, és a dir $M = Q_2 = x_{244} = 6.5$.

Similarment s'obté $Q_1 = 4.5$ i $Q_3 = 8.5$, per tant el el rang interquartílic és $RIQ = 8.5 - 4.5 = 4$.

Per últim, com que estem suposant les dades centrades en cada interval, el rang és $18.5 - 0.5 = 18$.

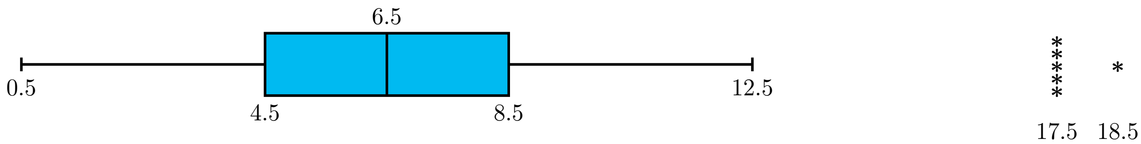
(b) Per dibuixar el diagrama de caixa necessitem calcular els següents nombres:

$$f_1 = Q_1 - 3/2 \cdot RIQ = -1.5 \qquad f_3 = Q_3 + 3/2 \cdot RIQ = 14.5$$

$$F_1 = Q_1 - 3 \cdot RIQ = -7.5 \qquad F_3 = Q_3 + 3 \cdot RIQ = 20.5$$

$$\text{adj}_1 = \text{mínima obs. a } ([f_1, Q_1]) = 0.5 \qquad \text{adj}_3 = \text{màxima obs. a } ([Q_3, f_3]) = 12.5.$$

Aleshores, el diagrama de caixa és:



Per tant, *hi ha* 6 anomalies moderades (5 amb valor 17.5 i una amb valor 18.5), i *no hi ha* anomalies extremes.

7. (a) Per respondre als apartats (a1)–(a3) primer hem de tenir clar què vol dir augmentar en un 10% els sous. Si el sou inicial era x_i ara es cobra

$$y_i = x_i + x_i \frac{10}{100} = x_i \cdot 1.1$$

Aixídoncs, denotarem per Y els nous sous i X els antics, assumint la relació $Y = X \cdot 1.1$.

(a1)

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} x_i \, 1.1 = 1.1 \, \frac{1}{n} \sum_{i=1}^{n} x_i = 1.1 \bar{X}.$$

Per tant,

$$\bar{Y} = 1.1 \cdot 2000 = 2200.$$

CERT.

(a2)

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(1.1x_i - 1.1\bar{X})^2 = (1.1)^2\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 = (1.1)^2 S_X^2.$$

Per tant,

$$S_Y^2 = (1.1)^2 \cdot 400 = 484.$$

FALS.

(a3) En cas que la moda existeixi, la nova moda serà l'anterior multiplicada per 1.1, per tant varia: **FALS.**

(a4) Si augmentem els sous en 100, tenim: $Y = X + 100$, per tant

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - 100) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{n\,100}{n} = \bar{X} + 100.$$

Aleshores,

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i + 100 - \bar{X} - 100)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 = S_X^2.$$

Per tant,

$$S_Y^2 = 400$$

FALS.

(a5) No varia, perquè la mediana és la mitjana de les observacions 250 i 251 i per tant no es veu afectada. **CERT.**

Recordeu que en general, si $Y = aX + b$ aleshores $\bar{Y} = a\bar{X} + b$ i $S_Y^2 = a^2 S_X^2$.

(b) Fixem-nos que

$$\begin{aligned}
&\bar{X}_A = 1800\\
&S_A = 200 \Rightarrow X_A = 2000 = \bar{X}_A + S_A\\
&\bar{X}_B = 2100\\
&S_B = 400 \Rightarrow X_B = 2400 < \bar{X}_B + S_B
\end{aligned}$$

En termes absoluts la resposta és obvia: 2400 euros és més que 2000 euros. Ara bé, en termes relatius, és a dir el context de la seva propia empresa, el treballador A està en una franja més alta respecte la mitjana que el segon.

(c) La desigualtat de Txebixev garanteix que a $[\bar{X} - k\,S, \bar{X} + k\,S]$, hi ha més del $(1 - 1/k^2)\cdot 100\%$ de les observacions. Per $k = 2$ tenim doncs un 75% de les observacions. Si $\bar{X} = 1800$ i $S = 100$ tenim que a $[\bar{X} - 2\,S, \bar{X} + 2\,S] = [1600, 2000]$, hi ha, efectivament un 75% de les observacions.

8. La desigualtat de Txebixev ens garanteix que l'interval $[\overline{x} - k\,s_x\,,\, \overline{x} + k\,s_x]$ conté com a mínim el $(1 - \frac{1}{k^2})$ per cent de les dades de la mostra.

   Imposant $1 - \frac{1}{k^2} = 0.8$ obtenim $k = \sqrt{5}$, i per tant $\delta = k\,s_x = \sqrt{5}\cdot 3.1 = 6.932$.

   Per tant l'opció correcta és (B).

9. Aplicarem la desigualtat de Txebixev segons la qual a l'interval $[\bar{X} - k\,S, \bar{X} + k\,S]$, hi ha més del $(1 - 1/k^2)\cdot 100\%$ de les observacions.

   Si trobem k de manera que més del 75% de les dades estiguin a $[\bar{X} - k\,S, \bar{X} + k\,S]$, podrem agafar com a radi $\bar{X} + k\,S$. Imposant $1 - 1/k^2 = 0.75 \Rightarrow k = 2$. Per tant, $\bar{X} + 2\,S = 30 + 2\cdot 3 = 36$.

10. La filosofia és la mateixa que la de l'exercici 1. Fer el diagrama de tija i fulles i a partir d'aquí, obtenir el boxplot.

    Algunes dades per obtenir el Boxplot són:

    El Q_1 és la mitjana de la posició 5 i 6 , la mediana és la mitjana de la posició 10 i 11 i el Q_3 és la mitjana de la posició 15 i 16: $Q_1 = 4.35$, $M = 5.55$, $Q_3 = 6.7$.

    Donat que el $RIQ = Q_3 - Q_1 = 2.35$ les fronteres de la zona de normalitat vénen donandes per

$$\begin{aligned}
f_1 &= Q_1 - 1.5RIQ = 0.825 \\
f_3 &= Q_3 + 1.5RIQ = 10.225 \\
F_1 &= Q_1 - 3RIQ = -2.7 \\
F_3 &= Q_3 + 3RIQ = 13.75
\end{aligned}$$

$$\begin{aligned}
\text{adj}_1 &= \text{mínima obs. a } [f_1, Q_1] = 1.3 \\
\text{adj}_3 &= \text{màxima obs. a } [Q_3, f_3] = 9.6
\end{aligned}$$

    Hi ha una anomalia moderada a 0.5 i una d'extrema a 15.

2 REGRESSIÓ LINEAL

2.1 Exercicis

1. Les dades següents són les mesures de la velocitat de l'aire i el coeficient d'evaporació de gotes de combustible en una turbina de propulsió:

X: velocitat de l'aire (cm/s)	Y: coeficient d'evaporacio (mm^2/s)
20	0.18
60	0.37
100	0.35
140	0.78
180	0.56
220	0.75
260	1.18
300	1.36
340	1.17
380	1.65

   (a) Representeu el diagrama de dispersió.
   (b) Trobeu la recta de regressió.
   (c) És raonable suposar que hi ha una relació lineal entre els coeficients d'evaporació la velocitat de l'aire.
   (d) Estima el coeficient d'evaporació d'una goteta quan la velocitat de l'aire és 190cm/s.

2. S'ha fet un estudi per veure si existeix qualque relació lineal entre el temps (X) i el nombre de missatges spams rebuts en una compte de correu (Y) durant X hores. Durant 20 dies s'han comptabilitzat el nombre d'hores que estava engegada la compte de correu i el nombre de spams rebuts i s'han obtingut els resultats següents:

$$\sum x_i = 102, \sum y_i = 292, \sum x_i^2 = 588, \sum y_i^2 = 4526, \sum x_i y_i = 1620.$$

   (a) Determineu la recta de regressió del nombre de missatges spams rebuts en funció del temps.
   (b) Calculeu el coeficient de correlació i el coeficient de determinació.

3. S'ha fet un estudi de la pèrdua de pes (és a dir, la *corrosió*) d'un determinat aliatge en funció del temps que ha estat submergit en aigua de mar.

   Amb els 50 parells de valors observats s'ajusta una recta de regressió que expressa la pèrdua de pes, y, de l'aliatge en funció del temps en dies, x, d'exposició a l'aigua de mar. L'equació de regressió de mínims quadrats és $y = 3.67 + 1.77x$. Sabem que les variàncies de x i de y són respectivament 2.064 i 6.8. Es demana:

   (a) Calculeu els coeficients de correlació i de determinació. És un bon ajust?
   (b) Interpreteu correctament el fet que $y(9) = 19.6$.
   (c) Quina és la velocitat de corrosió (expressada en grams per dia)?.

4. Es fan uns estudis de resistència a la tensió d'un tipus de fibra, en funció del PH de la solució on es tinta. La variable Y recull les dades de resistència i la variable X les dades del PH. Després de molts experiments obtenim: $\bar{X} = 1.5$, $\bar{Y} = 9.6$, $S_X = 0.7$, $S_Y = 4.5$, i $S_{XY} = -3.1$. Es demana:

   a) Es pot afirmar que hi ha correlació lineal entre "la resistència a la tensió" i el PH? Raoneu la resposta.

   b) Calculeu la recta de regressió de mínims quadrat, per a predir la "resistència a la tensió"en funció del PH. Feu una predicció de resistència de la fibra, per un PH: $x = 1$.

5. Es fan uns estudis sobre contaminació d'un riu. La variable Y recull les dades de *mortaldat de peixos*, (peixos recollits en un dia en un tram d' 1Km), i la variable X les dades del PH. Després de molts experiments obtenim: $\bar{X} = 5.5$, $\bar{Y} = 40$, $S_X = 0.5$, $S_Y = 1$, i $S_{XY} = -0.49$. Es demana:

   (a) Es pot afirmar que s'observa una correlació lineal alta entre les dades de de *mortaldat de peixos* i el PH? Raoneu la resposta.

   (b) Calculeu la recta de regressió de mínims quadrat, per a predir la *mortaldat de peixos* en funció del PH. Feu una predicció de mortaldat per un PH: $x = 4$.

   (c) Per quin PH la mortaldat és nul.la? Aquest resultat no té gaire sentit, (penseu que el PH de l' NH_3 és 10). Doneu una explicació de com pot produïr-se aquest resultat.

6. Dues variables X i Y tenen associades les rectes de regressió $Y = 0.89\,X + 3.82$ i $Y = 0.76\,X + 4.29$. Es demana:

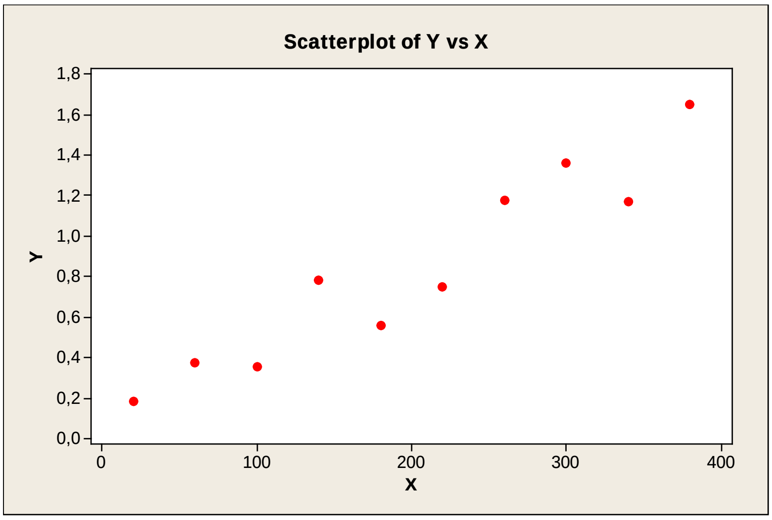
   (a) Calculeu què val $\bar{X}$, $\bar{Y}$.

   (b) Calcula el coeficient de determinació de R^2.

   (c) Quin signe tindrà el coeficient de correlació lineal?

2.2 Solucions

1.



(a)

(b) Considerem la següent taula:

	X	Y	X^2	Y^2	XY
	20	0,18	400	0,0324	3,6
	60	0,37	3600	0,1369	22,2
	100	0,35	10000	0,1225	35
	140	0,78	19600	0,6084	109,2
	180	0,56	32400	0,3136	100,8
	220	0,75	48400	0,5625	165
	260	1,18	67600	1,3924	306,8
	300	1,36	90000	1,8496	408
	340	1,17	115600	1,3689	397,8
	380	1,65	144400	2,7225	627
SUMA	2000	8,35	532000	9,1097	2175,4

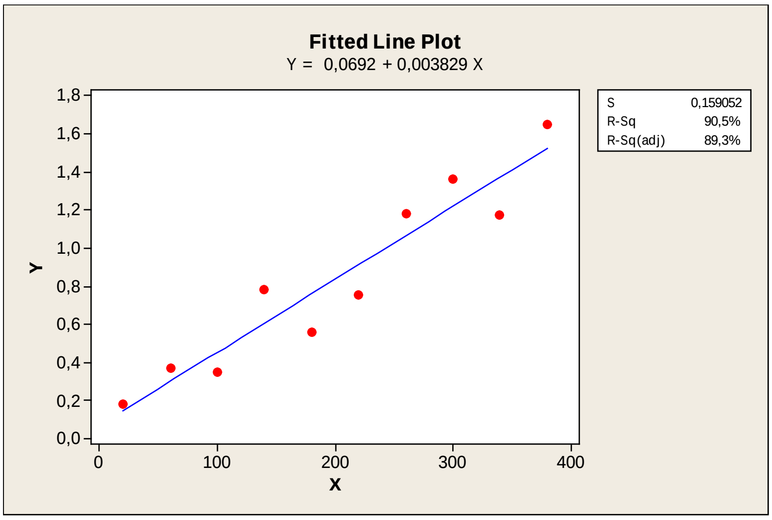
D'aquí obtenim:

$$
\begin{aligned}
n &= 10 \\
\bar{X} &= \frac{\sum_{i=1} x_i}{n} = \frac{2000}{10} = 200 \\
\bar{Y} &= \frac{\sum_{i=1} y_i}{n} = \frac{8.35}{10} = 0.835 \\
S_{XY} &= \frac{\sum_{i=1} x_i y_i}{n} - \bar{X}\bar{Y} = \frac{2175.5}{10} - 200 \cdot 0.835 = 50.54 \\
S_X^2 &= \frac{\sum_{i=1} x_i^2}{n} - \bar{X}^2 = \frac{53200}{10} - 200^2 = 13200 \\
S_Y^2 &= \frac{\sum_{i=1} y_i^2}{n} - \bar{Y}^2 = \frac{9.1097}{10} - 0.835^2 = 0.213745
\end{aligned}
$$

Per tant, obtenim

$$
\begin{aligned}
y &= \frac{S_{XY}}{S_X^2}\left(x - \bar{X}\right) + \bar{Y} = \frac{50.54}{13200}(x - 200) + 0.835 \\
&= 0.00383(x - 200) + 0.835 = 0.00383x + 0.069.
\end{aligned}
$$

La recta de regressió és $y = 0.00383x + 0.069$.



(c) El coeficient de determinació és

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2} = \frac{50.54^2}{13200 \cdot 0.213745} = 0.9053,$$

la qual cosa es pot interpretar com un 90.53% de correlació lineal, i per tant si que es pot afirmar.

(d) $y(190) = 0.7967$.

2. De les dades del problema tenim:

$$n = 20$$

$$\bar{X} = \frac{\sum_{i=1} x_i}{n} = 5.1$$

$$\bar{Y} = \frac{\sum_{i=1} y_i}{n} = 14.6$$

$$S_{XY} = \frac{\sum_{i=1} x_i y_i}{n} - \bar{X}\bar{Y} = 6.54$$

$$S_X^2 = \frac{\sum_{i=1} x_i^2}{n} - \bar{X}^2 = 3.39$$

$$S_Y^2 = \frac{\sum_{i=1} y_i^2}{n} - \bar{Y}^2 = 13.14$$

Per tant,

(a) La recta és

$$\begin{aligned} y &= \tfrac{S_{XY}}{S_X^2}\left(x - \bar{X}\right) + \bar{Y} = \tfrac{6.54}{3.39}(x - 5.1) + 14.6 \\ &= 1.9292(x - 5.1) + 14.6 \end{aligned}$$

(b) El coeficient de determinació és:

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2} = \frac{6.54^2}{3.39 \cdot 13.14} = 0.96019.$$

Com que S_{XY} és positiva el coeficient de correlació lineal és

$$R = +\sqrt{R^2} = +0.97989.$$

3. (a) Observem que el coeficient de correlació lineal i el de determinació són, respectivament:

$$r = \frac{S_{XY}}{S_x\, S_y} \text{ i } R = \frac{S_{XY}^2}{S_X^2\, S_Y^2}.$$

Observem que l'expressió de la recta de mínims quadrats és

$$y = \frac{S_{XY}}{S_x^2}\,(x - \bar{x}) + \bar{y} = 1.77x + 3.67,$$

per tant:

$$\frac{S_{XY}}{S_x^2} = \frac{S_{XY}}{2.064} = 1.77 \Rightarrow S_{XY} = 3.65328.$$

Finalment, doncs, el coeficient de determinació és

$$R^2 = \frac{S_{XY}^2}{S_X^2\, S_Y^2} = \frac{(3.65328)^2}{2.064 \cdot 6.8} = 0.9509.$$

$$r = \frac{S_{XY}}{S_x\, S_y} = \frac{3.65328}{\sqrt{2.064}\sqrt{6.8}} = 0.9752$$

Noteu que el signe de r és positiu perquè la covariança ha sortit positiva.

Efectivament, és un bon ajust. El coeficient de determinació indica que hi ha un 95.09% de correlació lineal.

(b) El fet que $y(9) = 19.6$ indica que en mitjana la pèrdua de pes (corrosió) d'una mostra d'aliatge submergida 9 dies és de 19.6 grams. És molt important entendre que és una estimació en mitjana, i que no vol dir que cada vegada que submergin una mostra durant 9 dies es perdran exactament 19.6 grams. Aquesta darrera interpretació és errònia.

(c) La velocitat de corrosió (expressada en grams per dia) és exactament la derivada de la corrosió, que es correspon amb el pendent de la recta, per tant és 1.77.

4. (a) De les dades obtenim que el coeficient de determinació és:

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2} = 0.9685059210$$

que és alt i per tant hi ha correlació lineal. Com que S_{XY} és negativa el coeficient de correlació lineal és

$$R = -\sqrt{R^2} = -0.9841269842.$$

(b) La recta de regressió és

$$y \; = \tfrac{S_{XY}}{S_X^2}\left(x - \bar{X}\right) + \bar{Y} = -6.326530612\,x + 1908979592$$

i $y(1) = 12.76326531$.

5. Aquest exercici és igual que l'anterior. (a) $R^2 = 0.9604$ alt i per tant hi ha correlació lineal; $R = -0.98$. (b) $y = -1.96\,x + 50.78$; $y(4) = 42.94$ (c) Igualant $y = -1.96\,x + 50.78 = 0$ obtenim un PH 25.9081 la qual cosa és un absurd (el pH, en les dissolucions aquoses, té una escala que va del 0 al 14). És molt probable que hi hagi un comportament òptim per la mortalitat i que superat aquest PH la mortalitat empitjori. Això correspon a un mínim, i per tant a un comportanent no lineal. És doncs probable que fora del rang de dades que conformen l'estudi el comportament sigui fortament no-lineal.

6. Les dues rectes de regressió són

$$y = \frac{S_{XY}}{S_X^2}\left(x - \bar{X}\right) + \bar{Y},$$

que serveix per a predir Y en funció de X perquè, parlant en termes informals, minimitza la suma dels quadrats dels "errors verticals", i

$$y = \frac{S_Y^2}{S_{XY}}\left(x - \bar{X}\right) + \bar{Y},$$

que serveix per a predir X en funció de Y perquè minimitza la suma dels quadrats dels "errors horitzontals".

(a) Les dues rectes de regressió passen pel punt $(\bar{X}, \bar{Y})$. Buscant el punt de tall de les dues rectes

$$\begin{cases} Y = 0.89\,X + 3.82, \\ Y = 0.76\,X + 4.29, \end{cases}$$

obtenim que aquest punt és $(\bar{X}, \bar{Y}) = (3.6154, 7.0377)$.

(b) El coeficient de determinació és

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2} = \frac{\frac{S_{XY}}{S_X^2}}{\frac{S_Y^2}{S_{XY}}}.$$

Per tant és el quocient del pendent les dues rectes. Ara bé, no sabem quina de les dues rectes donades $Y = 0.89\,X + 3.82$ i $Y = 0.76\,X + 4.29$ té pendent $\frac{S_{XY}}{S_X^2}$ i quina té pendent $\frac{S_Y^2}{S_{XY}}$... Doncs bé, és igual ja que si considerèssim

$$R^2 = \frac{0.89}{0.76} = 1.17...$$

és a dir, que obtindriem un número major que 1 la qual cosa no pot ser (recordeu que $R^2 \in [0, 1]$)! Per tant:

$$R^2 = \frac{0.76}{0.89} = 0.8539.$$

(c) El signe tindrà del coeficient de correlació lineal serà positiu ja que S_{XY} és positiva perquè els pendents de les rectes ho són.

3 PROBABILITAT ELEMENTAL

3.1 Exercicis

1. Una empresa substitueix el reductor d'engranatges d'una turbina cada cert període de temps. L'experència permet assegurar que la probabilitat que el reductor d'engranatges estigui aquest període sense fallar és de 0.8. Solament es poden produir dos tipus de falla. Està comprovat que les del primer tipus es produeixen el triple de cops que les del segon i que no es poden produir els dos tipus de falla de manera simultània. Quines són les probabilitats de cada tipus de falla?

2. Es disposa de quatre mostres de diferents tipus de fibra. Només un dels quatre tipus té la resistència que es necessita, però no se sap de quin tipus és. La resistència es determina mitjançant proves destructives. Si realitzem les proves seleccionant les mostres en ordre aleatori i sense reposició, quina és la probabilitat que siguin necessaries com a mínim tres proves per a detectar quin tipus de fibra és l'adequat?

3. Si B és un esdeveniment amb probabilitat 0.3, aleshores calculeu les següents probabilitats:
   (a) $P(B|B)$
   (b) $P(\overline{B}|B)$
   (c) $P(B|B \cup \overline{B})$
   (d) $P(B \cap \overline{B} \mid B)$

4. Quan un DVD falla, en un 60% dels casos es detecten funcionaments defectuosos del motor, i en un 50% dels casos hi ha funcionaments defectuosos dels lectors. Tenint en compte que en un 65% de les vegades que el DVD falla es detecta un o altre mal funcionament, quina és la probabilitat que en obrir un DVD defectuós a l'atzar ens trobem que:
   (a) El motor funcioni correctament.
   (b) Tant el motor com els lectors funcionen correctament.
   (c) El motor funciona incorrectament però el lector funciona bé. *Indicació: feu-vos un diagrama de Venn.*

5. Tenim barrejats 200 components fabricats en dues línies de producció A i B. Sabem que la quantitat d'articles correctes i defectuosos per cada línia es:

	Art. Correctes	Art. Defectuosos
Línia A	90	10
Línia B	80	20

   Es demana:
   (a) Quina probabilitat assignaríeu al succés {*s'escull un article a l'atzar i resulta que prové de la línea A*}?
   (b) Quina probabilitat assignaríeu al succés {*s'escull un article a l'atzar i resulta que és defectuós*}?
   (c) Quina probabilitat assignaríeu al succés {*s'escull un article a l'atzar i resulta que prové de la línea A i és defectuós*}?

(d) S'escull un artícle a l'atzar i resulta que és defectuós. Quina és la probabilitat que provingui de la línea A?

6. En una xarxa d'aeroports, la probabiltat que un vol regular surti puntual és 0.83, la que arribi puntual és 0.82, i la que surti i arribi puntual és 0.78. Calculeu la probabilitat que: (a) Arribi puntual si surt puntual i (b) Hagi sortit puntual si ha arribat puntual.

7. Els estudis experimentals posen en evidència que quan s'avaria una màquina del nostre taller:

   - El 50% de les vegades està avariat el dispositiu hidràulic (DH, en endavant).
   - El 70% de les vegades s'ha avariat o el DH o la targeta integrada (TI, en endavant).
   - $P(\text{avaria en TI}) = 0.3$.

   Es demana: Determineu si els successos *"quan s'avaria la màquina, hi ha una avaria al DH"* i *"quan s'avaria la màquina, hi ha una avaria a la TI"* són o no independents.

8. Un nou dispositiu de frenat dissenyat per impedir que els cotxes derrapin conté una quantitat considerable de peces electròniques, hidràuliques i mecàniques. Tot el sistema global es pot descomposar en tres subsistemes que operen en sèrie i de forma independent: un sistema electrònic, un sistema hidràulic i un actuador mecànic. En un cert tipus de frenada les confiabilitats de cada una d'aquestes unitats són aproximadament 0.995, 0.993 i 0.994 respectivament. Estimeu la confiabilitat del sistema global.

9. Un aparcament té dos compartiments; en el primer hi ha quatre cotxes i en el segon cinc. En el primer compartiment hi ha dos cotxes espatllats i en el segon només hi ha un cotxe que funcioni. a) Quina probabilitat d'escollir un cotxe que funcioni té una persona que escull primer una compartiment i després un cotxe?. b) Si el cotxe escollit funciona, quina és la probabilitat que el compartiment escollit hagi estat el primer?.

10. Una empresa compra el 80% del que necessita a un proveïdor que subministra un 1% dels articles defectuosos. La resta de les compres es fan a un segon proveïdor que subministra el 2% defectuós. Quina és la probabilitat que una peça escollida a l'atzar en el magatzem de l'empresa sigui defectuosa?. Sabent que la peça és defectuosa, quina és la probabilitat que procedeixi del primer proveïdor?.

11. En un determinat procés industrial s'ha de fer unes determinades anàlisis. Aquesta labor es pot realitzar mitjançant dos procediments: A i B. Normalment A s'utilitza el doble de cops que B. El 75% de les vegades que es fa servir A el temps necessari per a realitzar les anàlisis és com a màxim de tres hores. En canvi fent servir B, el temps necessari és de tres hores o menys només en el 60% dels cops. Si es fa una anàlisi i el temps empleat ha estat de tres hores i dos quarts, quina és la probabilitat que el mètode utilitzat hagi estat el B?.

12. Una anàlisi química té per objectiu detectar la presència de l'element A en un determinat producte. En dur a terme l'anàlisi, la probabilitat de detectar A en un producte que realment conté aquest element és 0.8. La probabilitat de no detectar A en un producte que no el conté és 0.9. La probabilitat que un producte qualsevol contingui A és 0.4. Si es realitzen tres anàlisis independents i en dos d'ells es detecta A, quina és la probabilitat que A sigui realment present?.

13. Els clients de dues consultories A i B provenen de la Indústria Mecànica (M), la Indústria Electrònica (E) i la Indústria Química (Q). Les consultories no comparteixen clients. Se sap també que:

   - El 30% dels clients són de la consultoria B.

- La probabilitat que *"si un client és de la consultoria A, sigui de la Indústria Mecànica"* és del 0.5.
- La probabilitat que un client *"sigui de la consultoria A i de la Indústria Electrònica"* és de 0.14.
- La probabilitat que un client *"sigui de la consultoria B i de la Indústria Química"* és de 0.24.
- La probabilitat que *"un client sigui de la consultoria B i de la Indústria Mecànica"és de 0.03"*.

Es demana:

(a) Quina és la probabilitat que un client *"sigui de la Indústria Mecànica"*?

(b) Els successos *"ser de la consultoria A"* i *"ser de la Indústria Mecànica"*, són successos independents? Per què?

(c) Quina és la probabilitat que un client *"sigui de A i de la Indústria Química"*?; Quina és la probabilitat que un client *"sigui de la Indústria Química"*?; Quina és la probabilitat *"ser de consultoria A, si és de la Indústria Química"*?

(d) Quina és la probabilitat que un client, que *"se sap que NO és de la Indústria Electrònica, sigui de la consultoria B"*?

3.2 Solucions

1. La idea és que la probabilitat del primer tipus de falla és $3x$ si la del segon tipus de falla és x. Com que *"la probabilitat que el reductor d'engranatges estigui aquest període* sense fallar*"* és 0.8, la probabilitat que *"es produeixi una falla"* és 0.2.

   D'altra banda, a frase *"Solament es poden produir dos tipus de falla"* indica que els esdeveniments fallada de primer tipus i de segon tipus són disjunts (i per tant la probabilitat de la unió és la suma), per tant:
$$x + 3x = 0.2 \Rightarrow 4x = 0.2 \Rightarrow x = 0.05 \Rightarrow 3x = 0.15.$$

   Les probabilitats són doncs 0.05 i 0.15 respectivament.

2. La probabilitat és 0.5.

3. (a) $P(B|B) = \dfrac{P(B \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1.$

   (b) $P(\overline{B}|B) = \dfrac{P(\overline{B} \cap B)}{P(B)} = \dfrac{P(\emptyset)}{P(B)} = \dfrac{0}{P(B)} = 0.$

   (c) $P(B|B \cup \overline{B}) = P(B|\Omega) = \dfrac{P(B \cap \Omega)}{P(\Omega)} = \dfrac{P(B)}{1} = P(B) = 0.3.$ v $P(B \cap \overline{B} \mid B) = P(\emptyset|B) = \dfrac{P(\emptyset \cap B)}{P(B)} = \dfrac{P(\emptyset)}{P(B)} = \dfrac{0}{P(B)} = 0.$

4. Si anomenem $A =$"defecte al motor" i $B =$"defecte al lector", les dades del problema indiquen que $P(A) = 0.6$, $P(B) = 0.5$ i $P(A \cup B) = 0.65$. Aleshores

   (a) $P(A^c) = 1 - P(A) = 0.4.$

   (b) $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.65 = 0.35.$

   (c) Observem que $P(A \cap B^c) = P(A) - P(A \cap B)$. Com que
$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.6 + 0.5 - 0.65 = 0.45,$$
   tenim $P(A \cap B^c) = P(A) - P(A \cap B) = 0.6 - 0.45 = 0.15.$

5. (a) A té 100 artícles d'un total de 200, per tant $P(A) = 100/200 = 0.5$.

   (b) Hi ha 30 artícles defectuosos d'un total de 200, per tant $P(\text{def}) = 30/200 = 0.15$.

   (c) Hi ha 10 artícles defectuosos fets a la línia A, per tant: $P(A \cap \text{def}) = 10/200 = 0.05$.

   (d)
$$P(A|\text{def}) = \frac{P(A \cap \text{def})}{P(\text{def})} = \frac{0.05}{0.15} = 0.3333.$$

6. (a) 0.939759; (b) 0.9512195.

7. Les dades són: $P(av.DH) = 0.5$, $P(\{av.DH\} \cup \{av.TI\}) = 0.7$, $P(av.TI) = 0.3$. Per determinar si són o no independents els successos que ens demanen aplicarem si és cert o no que
$$P(\{av.DH\} \cap \{av.TI\}) = P(av.DH) \cdot P(av.TI)?$$

Observem que

$$P(\{av.\,DH\}\cap\{av.\,TI\}) = P(av.\,DH)+P(av.\,TI)-P(\{av.\,DH\}\cup\{av.\,TI\}) = 0.5+0.3-0.7 = 0.1.$$

Per altra banda

$$P(av.\,DH)\cdot P(av.\,TI) = 0.5\cdot 0.3 = 0.15.$$

Per tant NO són independents.

8. La confiabilitat total és la probabilitat que funcioni cadascún dels dispositius (treballen en sèrie). Això es la intersecció i es pot calcular com el producte degut a la independència 0.982106790.

9. (a) 0.35; (b) 0.714286.

10. (a) 0.012 (b) 0.6667.

11. 0.4444.

12. Les dades indiquen que $P(A) = 0.4$, $P(+|A) = 0.8$, $P(-|A) = 0.2$, i per tant $P(A^c) = 0.6$, $P(-|A) = 0.2$, $P(+|A^c) = 0.1$.

    Farem servir la notació $\{++-\}$ per designar que s'han produit dues deteccions i una no-detecció *sense* tenir en compte l'ordre i la notació $(+,+,-)$ (*tenint en compte l'ordre!*)

    Usant la fórmula de Bayes, tenim:

$$P(A|\{++-\}) = \frac{P(A\cap\{++-\})}{P(\{++-\})} = \frac{P(\{++-\}|A)\,P(A)}{P(\{++-\}|A)\,P(A)+P(\{++-\}|A^c)\,P(A^c)}.$$

    Per calcular $P(\{++-\}|A)$ (i anàlogament $P(\{++-\}|A^c)$), primer observem que

$$\{++-\} = (+,+,-)\cup(+,-,+)\cup(-,+,+)$$

    i que aquesta unió és disjunta, per tant

$$P(\{++-\}|A) = P((+,+,-)|A)+P((+,-,+)|A)+P((-,+,+)|A) = 3P((+,+,-)|A).$$

    Ja que els successos $(+,+,-)$, $(+,-,+)$, i $(-,+,+)$ tenen la mateix probabilitat. Ara observem que

$$(+,+,-|A) = \{+|A\}\cap\{+|A\}\cap\{-|A\},$$

    per tant fent servir la independència dels diferents anàlisis tenim:

$$P(\{++-\}|A) = 3P((+,+,-)|A) = 3\cdot P(+|A)\cdot P(+|A)\cdot P(-|A) = 3\cdot 0.8\cdot 0.8\cdot 0.2 = 0.384.$$

    Anàlogament

$$P(\{++-\}|A^c) = 3\cdot P(+|A^c)\cdot P(+|A^c)\cdot P(-|A^c) = 3\cdot 0.1\cdot 0.1\cdot 0.9 = 0.027.$$

    Per tant:

$$\begin{aligned} P(A|\{++-\}) &= \frac{P(\{++-\}|A)\,P(A)}{P(\{++-\}|A)\,P(A)+P(\{++-\}|A^c)\,P(A^c)} = \frac{0.384\cdot 0.4}{0.384\cdot 0.4+0.027\cdot 0.6} \\ &= 0.9045936396. \end{aligned}$$

13. A partir de l'enunciat, tenim les següents probabilitats:

$$P(B) = 0.3, \quad P(M|A) = 0.5, \quad P(A \cap E) = 0.14$$
$$P(B \cap Q) = 0.24, \quad P(B \cap M) = 0.03$$

(a) A partir del Teorema de la Probabilitat Total:

$$P(M) = P(A \cap M) + P(B \cap M)$$

Ens cal calcular $P(A \cap M)$. Donat que $P(M|A) = 0.5$ i que $P(A) = 1 - P(B) = 1 - 0.3 = 0.7$, tenim que $P(A \cap M) = P(A) \cdot P(M|A) = 0.7 \cdot 0.5 = 0.35$. Així doncs,

$$P(M) = P(A \cap M) + P(B \cap M) = 0.35 + 0.03 = 0.38.$$

(b) No són successos independents ja que:

$$P(A \cap M) \neq P(A) \cdot P(M)$$

ja que $P(A \cap M) = 0.35$ i $P(A) \cdot P(M) = 0.7 \cdot 0.38 = 0.266$.

(c) Observem que

$$P(A) = P(A \cap E) + P(A \cap M) + P(A \cap Q) \Rightarrow P(A \cap Q) = 0.7 - (0.14 + 0.35) = 0.21$$

de manera que

$$P(Q) = P(A \cap Q) + P(B \cap Q) = 0.24 + 0.21 = 0.45.$$

Per tant, a partir de la definició de probabilitat condicionada:

$$P(A|Q) = \frac{P(A \cap Q)}{P(Q)} = \frac{0.21}{0.45} = 0.467.$$

(d) Ara es demana la probabilitat $P(B|\overline{E})$. A partir de la definició de probabilitat condicionada, es té:

$$P(B|\overline{E}) = \frac{P(B \cap \overline{E})}{P(\overline{E})}$$

on $\overline{E} = E^c$ denota el complementari del succés E. Com que $\overline{E}$ és la unió disjunta de M i Q tenim:

$$P(B \cap \overline{E}) = P(B \cap M) + P(B \cap Q) = 0.03 + 0.24 = 0.27$$

i

$$P(\overline{E}) = P(M) + P(Q) = 0.38 + 0.45 = 0.83.$$

Per tant:

$$P(B|\overline{E}) = \frac{P(B \cap \overline{E})}{P(\overline{E})} = \frac{0.27}{0.83} = 0.325$$

4 MODELS DISCRETS

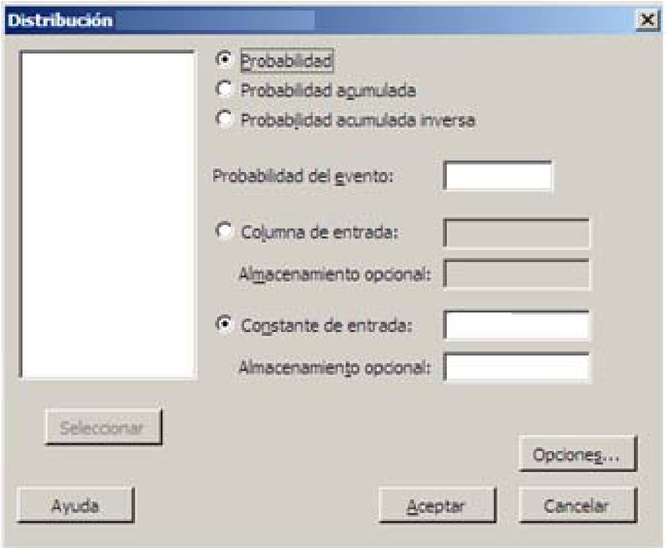
4.1 Exercicis

1. La distribucio de probabilitat d'una variable aleatòria discreta X és $P(X = 1) = 0.1$, $P(X = 2) = 0.2$, $P(X = 3) = 0.3$, $P(X = 4) = 0.25$, $P(X = 5) = 0.15$. Calculeu:
   (a) La probabilitat que al dur a terme l'experiment en el que es mesura X s'obtingui un valor mes gran que 1 sense superar 4.
   (b) $E(X)$ i $V(X)$
   (c) La funcio de distribucio $F_X(x)$.
2. La funció de distribució de la variable aleatòria discreta X és

$$F(X) = \begin{cases} 0 & \text{si } x < 5 \\ 0.1 & \text{si } 5 \leq x < 7 \\ 0.25 & \text{si } 7 \leq x < 9 \\ 0.65 & \text{si } 9 \leq x < 10 \\ 0.95 & \text{si } 10 \leq x < 15 \\ 1 & \text{si } x \geq 15 \end{cases}$$

   Calculeu
   (a) La probabilitat que X prengui un valor superior a 8.
   (b) La distribució de probabilitat de X.
3. Enviem un codi format per 15 bits. La probabilitat d'enviar un bit erròni és del 20%. Calculeu la probabilitat dels següents successos. *Feu servir les taules de la distribució binomial als apartats (b) i (d).*
   (a) No es rep cap bit erroni.
   (b) Es rep un nombre menor o igual que 5 bits erronis.
   (c) Es reben exactament 5 bits erronis.
   (d) Es rep un nombre major o igual que 5 bits erronis.
4. La probabilitat que un cert dispositiu tingui una durada de mes de 500 hores és 0.6. Seleccionem a l'atzar 10 d'aquests dispositius. Es demana:
   (a) Calculeu el número esperat de dispostius d'aquest grup que superaran les 500 h de vida, i la probabilitat que *cap* d'ells duri més de 500 h
   (b) Quina és la probabilitat que com a mínim 3 d'ells superin les 500 hores de durada.
   (c) Es realitza una seqüència d'inspeccions. Tenint en compte que la probabilitat de trobar un dispositiu que superari les 500 h de vida al k-èssim intent és de 0.0384, digueu quin és el valor d'aquest k.

Als següents exercicis, si un cop plantejats els càlculs aquests són molt feixucs, podeu usar Minitab com a calculadora de probabilitats fent `Calc`$\rightarrow$ `Distribuciones de probabilidad` (i escollint la distribució adient), on trobareu una pantalla similar a aquesta:



Segons sigui la distribució escollida, la finestra pot variar una mica. Fixeu-vos que heu d'escollir si voleu calcular una probabilitat, una probabilitat acumulada o bé una inversa. Altres dades que es demanen seran els paràmetres de la distribució i el valor d'entrada (aquell sobre el que voleu calcular la probabilitat), que el podeu donar en el full de càlcul (columna de entrada) o directament a la finestra (constante de entrada).

A les següents preguntes escolliu quin dels següents models és l'adequat per a respondre el problema d'entre els models Binomial, Geomètric, Binomial negatiu, Hipergeomètrica, Poisson.

5. Un examen tipus test consta de deu qüestions. Cada pregunta presenta quatre possibles respostes, de les quals només una és correcta. a) Quina probabilitat d'aprovar l'examen té un estudiant que contesta totes les preguntes a l'atzar? b) Quin és el número esperat de respostes correctes?

6. S'envia un codi format per 16 bits. La probabilitat d'enviar un bit erroni, que és independent per cada bit, és del 25%. Calculeu la probabilitat dels següents successos. a) No es rep cap bit erroni. b) Es rep un nombre menor o igual que 4 bits erronis. c) Es reben exactament 4 bits erronis. d) Es rep un nombre estrictament major que 4 bits erronis.

7. En el context de l'exercici anterior quina és la probabilitat que el primer bits erroni enviat sigui el cinquè?

8. Tenim un lot de $N = 5000$ unitats. N'inspeccionem 300 i si d'aquestes no n'hi ha més de 10 de defectuoses donarem el lot per bo. a) Si un lot conté un 2% d'unitats defectuoses quina és la probabilitat que, amb aquesta regla de decisió, el donem per bo? b) I si conté el 5% defectuos?

9. En un procés de fabricació la probabilitat que una peça sigui defectuosa és 0.03. a) Quina és la probabilitat que la primera peça defectuosa s'obtingui més enllà de la cinquena peça? b) Quina és la probabilitat que les tres primeres peces defectuoses s'obtinguin més enllà de la cinquena peça?

10. Quan un sistema falla s'emet un senyal de so que es repeteix cada minut. La probabilitat que el centre de control capti el senyal de so quan aquell s'emet és 0.85, de forma independent per cada emissió. a) Quina és la probabilitat que la falla no es detecti fins el tercer avís? b) Quants senyals s'hauran d'emetre en mitjana teòrica perquè el centre de control detecti la falla?

11. Un procés de fabricació té 100 comandes en espera de ser servides. Cada comanda necessita un component que es compra a un altre proveïdor. Normalment aquests components presenten un percentatge d'un 2% d'unitats defectuoses. L'estat de cada component és independent del dels altres. a) Si al magatzem hi ha 100 components quina és la probabilitat que es puguin servir totes les comandes pendents sense haver de demanar més components? b) Contesteu a la pregunta anterior en el supòsit que el magatzem disposi actualment de 102 components. d) Contesteu el mateix que abans però ara per a 105 components en el magatzem.

12. El número de trucades que rep una centraleta, durant el matí, segueix una llei de Poisson. Es sap que en un període cinc minuts es reben, en mitjana, 3 trucades. a) Quin és el número esperat de trucades en un quart d'hora? b) Quina és la probabilitat que en un quart d'hora es rebin més de 10 trucades?

13. El número de clients que arriben, al llarg d'un dia, a un determinat taller de reparacions segueix una distribució de Poisson de paràmetre $\lambda = 5$: a) Quants clients s'espera per dia, en mitjana? b) Quina és la probabilitat que al llarg d'un dia no hi vagi cap client?

14. En una partida d'una determinada peça de roba s'ha observat un cert tipus de defecte. En mitjana es presenten tres defectes per m^2. a) Quin és el número de defectes esperat en una peça de $10m^2$? b) Quina és la probabilitat que en una peça de $10m^2$ hi hagi més de 50 defectes?

15. Un procés produeix un 4% d'unitats defectuoses. Les unitats produïdes es disposen en caixes de 60 unitats. a) Quina és la probabilitat que una caixa contingui més de dues unitats defectuoses? b) Si s'agafen 10 caixes a l'atzar, quina és la probabilitat que més del 50% de les caixes contingui més de dues unitats defectuoses?

16. Un inspector està buscant soldadures defectuoses en una canonada de conducció d'aigües lineal en la que cada trenta metres hi ha una soldadura. La probabilitat que té cada soldadura de ser defectuosa de 0.05. Suposem que els defectes a les soldadures són independents. L'inspector està decidit a continuar la inspecció fins a trobar tres soldadures defectuoses. Quina és la probabilitat que l'inspector hagi de caminar més d'un quilòmetre?

4.2 Solucions

1. (a) $P(1 < X \leq 4) = P(X=2) + P(X=3) + P(X=4) = 0.2 + 0.3 + 0.25 = 0.75$

   (b) $\mu = E(X) = \sum_{x_i} x_i\ P(X = x_i) = 1 \cdot 0.1 + \ldots + 5 \cdot 0.15 = 3.15,$

   $\sigma^2 = V(X) = E(X^2) - \mu^2 = \sum_{x_i} x_i^2\ P(X = x_i) - \mu^2 = 1^2 \cdot 0.1 + \ldots + 5^2 \cdot 0.15 - (3.15)^2 = 1.4275.$

   (c)

$$F(X) = \begin{cases} 0 & \text{si } x < 1 \\ 0.1 & \text{si } 1 \leq x < 2 \\ 0.3 & \text{si } 2 \leq x < 3 \\ 0.6 & \text{si } 3 \leq x < 4 \\ 0.85 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$$

2. (a) $P(X > 8) = P(8 < X \leq 15) = P(X \leq 15) - P(X \leq 8) = F(15) - F(8) = 1 - 0.25 = 0.75.$

   (b) $P(X = 5) = 0.1$, $P(X = 7) = 0.15$, $P(X = 9) = 0.4$, $P(X = 10) = 0.3$, $P(X = 15) = 0.05$

3. Com que es tracta de la repetició $n = 15$ vegades de forma independent de la transimissió d'un bit i la probabilitat que sigui erroni és $p = 0.2$, la variable que fa el recompte de respostes encertades és una variable binomial, $X \sim \text{Bin}(15, 0.2)$.

   (a)

$$P(X = 0) = \binom{15}{0} 0.2^0\, 0.8^{15} = \frac{15!}{0!15!} 0.2^0\, 0.8^{15} = 1 \cdot 1 \cdot 0.03518437 = 0.03518437 \simeq 0.03518.$$

   (b) Mirarem directament les taules, buscant $n = 15$, $p = 0.2$.

c. *n* = 15

		p							
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
	0	.860	.463	.206	.035	.013	.005	.000	.000
	1	.990	.829	.549	.167	.080	.035	.005	.000
	2	1.000	.964	.816	.398	.236	.127	.027	.004
	3	1.000	.995	.944	.648	.461	.297	.091	.018
	4	1.000	.999	.987	.836	.686	.515	.217	.059
	5	1.000	1.000	.998	.939	.852	.722	.402	.151
	6	1.000	1.000	1.000	.982	.943	.869	.610	.304
x	7	1.000	1.000	1.000	.996	.983	.950	.787	.500
	8	1.000	1.000	1.000	.999	.996	.985	.905	.696
	9	1.000	1.000	1.000	1.000	.999	.996	.966	.849
	10	1.000	1.000	1.000	1.000	1.000	.999	.991	.941
	11	1.000	1.000	1.000	1.000	1.000	1.000	.998	.982
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

   Tenint en compte que les taules indiquen $P(X \leq k)$ obtenim

$$P(X \leq 5) = 0.939.$$

(c)

$$P(X=5)=\left(\begin{array}{c}15\\5\end{array}\right)0.2^5\,0.8^{10}=\tfrac{15!}{5!10!}\,0.2^5\,0.8^{10}=\tfrac{15\cdot14\cdot13\cdot12\cdot11}{5\cdot4\cdot3\cdot2\cdot1}\,0.2^5\,0.8^{10}=$$
$$=0.1031822943\simeq0.1032.$$

(d) Tenint en compte que les taules indiquen $P(X\leq k)$ i que la variable és discreta, obtenim:

$$P(X\geq5)=1-P(X<5)=1-P(X\leq4)=1-0.836=0.164.$$

4. Considerem $X=$ *nombre de dispositius amb durada de mes de 500 hores*, la probabilitat que un dispositiu tingui una durada de mes de 500 hores és $p=0.6$ i seleccionem de forma aleatòria $n=10$ dispositius. La variable aleatòria discreta X segueix una distribució binomial de paràmetres $n=10$ i $p=0.6$.

(a) El valor esperat d'una variable aleatòria X que segueix una distribució binomial de paràmetres n i p és $E(X)=n\cdot p$. En el nostre cas serà $E(X)=10\cdot0.6=6$. La funció de probabilitat d'una distribució binomial de paràmetres n i p ve donada per:

$$P(X=k)=\frac{n!}{k!(n-k)!}\,p^k(1-p)^{n-k}$$

Així doncs,

$$P(X=0)=\frac{10!}{0!(10-0)!}\,0.6^0 0.4^{10}=0.0001$$

(b) Ara ens demanen:

$$P(X\geq3)=1-P(X\leq2)=1-[P(X=0)+P(X=1)+P(X=2)]$$
$$=1-[0.00010+0.00157+0.01062]=0.98771$$

Per descomptat, aquest càlcul també es pot fer fent servir les taules.

(c) En aquest apartat considerem que inspeccionem dispositius fins trobar el primer amb durada superior a les 500 hores (èxit). la variable aleat òria $Y=$ *nombre de fracassos fins a trobar el primer èxit*, segueix una distribució geomètrica de paràmetre $p=0.6$. Si

$$P(Y=k-1)=0.4^{k-1}\cdot0.6=0.0384$$

hem de resoldre l'equació anterior per trobar el valor de k. Observem que:

$$0.4^{k-1}=\frac{0.0384}{0.6}$$

prenem logaritmes i aïllem k:

$$(k-1)\ln(0.4)=\ln\left(\frac{0.0384}{0.6}\right)\quad\Rightarrow\quad k=1+\frac{\ln\left(\frac{0.0384}{0.6}\right)}{ln(0.4)}=4.$$

5. *Farem servir distribucions binomials.* Com que es tracta de la repetició $n=10$ vegades de forma independent d'una pregunta i la probabilitat d'encertar és $p=0.25$, la variable que fa el recompte de respostes encertades és una variable binomial, $X\sim\mathrm{Bin}(10,0.25)$.

(a)

$$P(\text{Aprovar})=P(X\geq5)=\sum_{k=5}^{10}P(X=k)=\sum_{k=5}^{10}\left(\begin{array}{c}10\\k\end{array}\right)0.25^k0.75^{10-k}=0.0781269\simeq0.0781.$$

Aquest càlcul és tediós i es pot fer de diverses maneres:

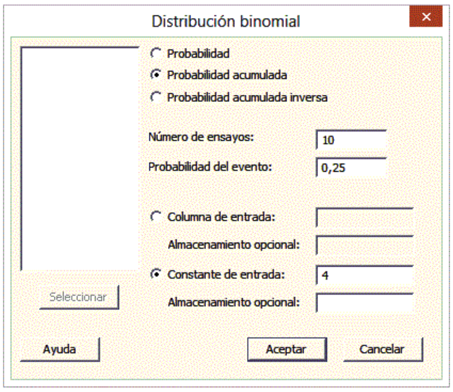
- A mà...si disposeu de l'energia vital suficient.
- Amb taules, buscant $n = 10$, $p = 0.25$, i tenint en compte que l es taules indiquen $P(X \leq k)$ i que X és una variable discreta que pren valors enters des de 0 a 10:

$$P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - 0.922 = 0.078.$$

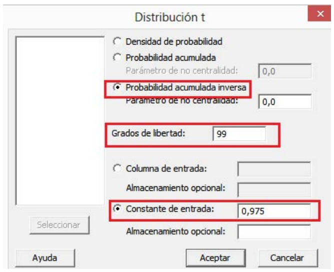
- Amb MINITAB, tenint en compte que X és una variable discreta que pren valors enters des de 0 a 10:

$$P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$$

Aleshores seguint els passos que s'han indicat anteriorment introduint:



amb la qual cosa obtenim:



Per tant:

$$P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - 0,921873 = 0,078127.$$

- Amb MAPLE amb la instrucció

```
> sum(binomial(10,k)· 0.25^k· 0.75^(10-5),k=5..10);
```

(b) $E(X) = np = 10 \cdot 0.25 = 2.5.$

6. *Farem servir distribucions binomials.* La variable que fa el recompte de bits erronis és una variable binomial, $X \sim \text{Bin}(16, 0.25)$.

(a) $P(X = 0) = \left(\begin{array}{c} 16 \\ 0 \end{array} \right) 0.25^0 0.75^{16} = \frac{16!}{0!16!} \cdot 1 \cdot 0.0100226 = 0.0100226 \simeq 0.0100.$

(b)

$$P(X \leq 4) = \sum_{k=0}^{4} P(X = k) = \sum_{k=0}^{4} \binom{16}{k} 0.25^k 0.75^{16-k} = 0.6301861752 \simeq 0.6302$$

Podeu usar MINITAB o bé MAPLE amb la instrucció

```
> sum(binomial(16,k)· 0.25^k· 0.75^(16-5),k=0..4);
```

(c) $P(X = 4) = \binom{16}{4} 0.25^4 \cdot 0.75^{12} = \frac{16!}{4!12!} 0.25^4 \cdot 0.75^{12} = 0.2251990652 \simeq 0.2252$

(d) $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.6302 = 0.3698$.

7. *Farem servir la distribució geomètrica.* $p = 0.25$, la variable Y que el número d'assajos fins a obtenir el primer bit erroni és una variable geomètrica. Per tant:

$$P(Y = 5) = (1-p)^{k-1} p = 0.75^4 \cdot 0.25 = 0.0791016 \simeq 0.0791.$$

8. *Farem servir la distribució hipergeomètrica.* Tenim un conjunt total format per $N = 5000$ unitats.

(a) Si hi ha un 2% d'unitats defectuoses aleshores en el conjunt total hi ha $K = 100$ unitats defectuoses. si X és la variable que compta la quantitat d'unitats defectuoses d'una mostra aleatòria de $n = 300$ unitats aleshores té una distribució hipergeomètrica, per tant fent servir que

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

Fent servir MINITAB o MAPLE obtenim

$$P(X \leq 10) = \sum_{k=0}^{10} \frac{\binom{100}{k} \cdot \binom{4900}{300-k}}{\binom{5000}{300}} = 0.9639492966 \simeq 0.9639.$$

Com que 300 és petit respecte de 5000 podem aproximar la variable X usant una variable binomial $X \simeq \text{Bin}(300, 0.02)$, per tant

$$P(X \leq 10) \simeq \sum_{k=0}^{10} \binom{300}{k} 0.02^k 0.98^{300-k} = 0.9590379409 \simeq 0.9590,$$

la qual cosa representa un error del 0.5095% respecte del resultat real calculat anteriorment.

(b) Si conté 5% d'unitats defectuoses, $K = 250$ i per tant

$$P(X \leq 10) = \sum_{k=0}^{10} \frac{\binom{250}{k} \cdot \binom{4750}{300-k}}{\binom{5000}{300}} = 0.1050317314 \simeq 0.1050.$$

9. *Farem servir distribucions geomètriques a l'apartat (a) i binomial negativa al (b).*

(a) Si Y compta el nombre d'assajos fins a obtenir la primera peça defectuosa, és una variable geomètrica amb $p = 0.03$:

$$\begin{aligned}
&P(Y > 5) = 1 - P(Y \leq 5) = 1 - \left(\sum_{i=1}^{5} P(Y = k)\right) = \\
&= 1 - \left(0.97^0 \cdot 0.03 + 0.97^1 \cdot 0.03 + \cdots + 0.97^4 \cdot 0.03\right) = \\
&= 1 - 0.03\left(1 + 0.97 + 0.97^2 + 0.97^3 + 0.97^4\right) = 1 - 0.03 \cdot 4.70886581 = 0.858734025 \simeq 0.8587.
\end{aligned}$$

(b) La variable X que compta el número d'assajos realitzats fins a obtenir 3 peces defectuoses és una variable binomial negativa. Fent servir que

$$P(X = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r,$$

obtenim que

$$\begin{aligned}
&P(X > 5) = 1 - P(X \leq 5) = 1 - \left(\sum_{i=3}^{5} P(X = k)\right) = \\
&= 1 - \left(\binom{2}{2} 0.97^0 0.03^3 + \binom{3}{2} 0.97^1 0.03^3 \binom{4}{2} 0.97^2 0.03^3\right) = 0.9997420042 \simeq 0.9997
\end{aligned}$$

10. *Farem servir la distribució geomètrica.* Si X compta el nombre d'assajos fins captar el so, és una variable geomètrica amb $p = 0.85$:

(a) $P(X = 3) = (1 - 0.85)^2\, 0.85 = 0.019125 \simeq 0.0191$.

(b) $E(X) = 1/p = 1/0.85 = 1.17647$.

11. *Farem servir distribucions binomials.* (a) 0.13262 (b) 0.66575 (c) 0.98076.

12. *Farem servir distribucions Poisson.* Considerem

$$\lambda = \frac{3 \text{ trucades}}{5 \text{ minuts}} = 0.6 \text{ trucades/minut},$$

.

(a) En 15 minuts esperem $15 \cdot 0.6 = 9$ trucades.

(b) La variable que compta el número de trucades en un quart d'hora és una variable Poisson $P(9)$ i per tant:

$$\begin{aligned}
&P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{k=0}^{10} \mathrm{e}^{-9}\frac{9^k}{k!} = \\
&= 1 - \mathrm{e}^{-9}\sum_{k=0}^{10} \frac{9^k}{k!} = 1 - 0.7059883204 = 0.2940116796 \simeq 0.2940.
\end{aligned}$$

El càlcul el podeu fer amb MINITAB o amb MAPLE.

13. *Farem servir distribucions Poisson.* Si C és la variable que compta els clients al llarg d'un dia $C \sim P(5)$, per tant: (a) $E(C) = 5$ i (b) $P(C = 0) = \mathrm{e}^{-5}5^0/0! = \mathrm{e}^{-5} = 0.0067379$.

14. *Farem servir distribucions Poisson.* La variable que compta els defectes cada $10m^2$ és una variable Poisson $X \sim P(30)$, fent servir MINITAB o MAPLE obtenim que

$$\begin{aligned}
&P(X > 50) = 1 - P(X \leq 50) = 1 - \textstyle\sum_{k=0}^{50} \mathrm{e}^{-30}\frac{30^k}{k!} = \\
&= 1 - \mathrm{e}^{-30}\textstyle\sum_{k=0}^{50} \frac{30^k}{k!} 1 - 0.9997019870 = 0.00029800130 \simeq 0.0003.
\end{aligned}$$

15. *Farem servir distribucions binomials.* (a) 0.4324 (b) 0.2255.

16. *Farem servir la distribució binomial negativa.* L'inspector caminarà més d'un kilòmetre si la 3a falla es troba més enllà de l'intent número 34 ja que $1000m/30m = 33.3333$. Si X és la variable que compta el número d'inspeccions fins a obtenir 3 defectes, aleshores X té una distribució binomial negativa

$$P(X \geq 34) = 1 - P(X \leq 33) = 1 - (P(X=3) + P(X=4) + \cdots + P(X=33)) =$$
$$= 1 - \left(\sum_{k=3}^{33} \binom{k-1}{2} 0.95^{k-3} 0.05^3 \right) = 1 - 0.2271931304 = 0.7728068696 \simeq 0.7728.$$

5 MODELS CONTINUS I DISTRIBUCIONS MOSTRALS

5.1 Exercicis

1. La funció de densitat de probabilitat de la variable aleatòria X és

$$f_X(x) = \begin{cases} 2(1-x) & \text{si } 0 < x < 1, \\ 0 & \text{altrament.} \end{cases}$$

   Calculeu:

   (a) La probabilitat que al dur a terme l'experiment en que es mesura X s'obtingui un valor superior a $2/3$.

   (b) $E(X)$ i $V(X)$

   (c) La funció de distribució $F_X(x)$ de X.

2. La funció de distribució de la variable aleatòria X és

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - \mathrm{e}^{-x} & \text{si } x \geq 0. \end{cases}$$

   (a) Calculeu $P(X \leq 2)$.

   (b) Calculeu la funció de densitat de la variable X.

3. El temps de durada, T de la bateria d'un dispositiu electrònic segueix una distribució exponencial amb temps de vida teòric de 100 setmanes. Es a dir

$$T \sim \exp(\lambda), \text{ tal que } f_T(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

   amb $\lambda = 1/100$.

   (a) Quina és la probabilitat que la bateria duri menys de 50 setmanes?

   (b) Quina és la probabilitat que duri més de 100 setmanes?

   (c) Quina és la probabilitat que duri més de 100 setmanes si sabem que ha durat més de 50?

   Investigueu què és la propietat de carència de memòria de les variables exponencials. Hi ha alguna altra variable aleatòria que tinguin aquesta propietat?

4. Fabriquem resistències de 10Ω. Les resistències no sempre surten igual, però podem pensar que segueixen una distribució normal tal que la mitjana de la nostra producció és de 10Ω, amb una desviació típica de 1Ω.

   (a) Calculeu la probabilitat que si agafem una resistencia a l'atzar, aquesta tingui una resistència major que 9.5Ω.

   (b) Calculeu la probabilitat que si agafem una resistencia a l'atzar aquesta tingui una resistència entre 8Ω i 9.5Ω.

(c) † Si els lots de resistències contenen 100 unitats, quina és la probabilitat que a l'escollir un lot a l'atzar, la mitjana de les resistències sigui menor de 9.5Ω?

5. Comprem unes lents per a fabricar lectors òptics.

   (a) Si assumim que la mitjana del diàmetre de les lents es pot modelitzar com una V.A. normal amb mitjana $\mu = 0.25$, i desviació típica $\sigma = 0.03$. Quina proporció de lents té un diàmetre entre $[0.24, 0.26]$.

   (b) Canviem el métode de fabricació. Assumim que la mitjana del diàmetre de les lents és $\mu = 0.25$ i que la variable continua sent normal. Quin valor hauria de tenir la desviació típica σ, per tal que la proporció de lents amb un diàmetre entre $[0.24, 0.26]$ sigui el 90%

6. L'amplada d'una peça de duralumini es distribueix normalment amb $\mu = 0.9000$ i $\sigma = 0.0030$. Els límits d'especificacio (límits dintre dels quals una peca no es considera defectuosa) són 0.9000 ± 0.0050.

   (a) Quin percentatge de peces resultara defectuós?

   (b) Si es vol que no es produeixi més d'un 1% de peces defectuoses, fins a quin valor s'hauria de reduïr, com a mínim, la variança σ^2 de la variable amplada?

7. La longitud d'un estoig de cintes magnètiques mollejat per injecció, té una distribució normal amb mitjana 90.5 mm i desviacio tipica 0.1 mm.

   (a) Quina es la probabilitat que la longitud d'un estoig sobrepassi els limits d'especificació 90 ± 0.3 mm?

   (b) Si la desviació típica es manté a 0.1, a quin valor s'ha d'ajustar la mitjana del procés per tal que el màxim número d'estoigs tinguin longitud compresa entre els límits d'especificació?

   (c) Si es desestimen els estoigs amb longituds fora de l'interval d'especificacions, quin es el rendiment del procés per al valor de la mitjana determinat a l'apartat anterior?

   (d) El procés s'ajusta a la mitjana de l'apartat (b) i es mesuren les longituds de deu estoigs produïts de forma independent. Quina es la probabilitat que les deu longituds es trobin entre 89.7 i 90.3 mm?

   (e) En les condicions de l'apartat anterior, quin numero d'estoigs, del total de deu, esperem que presentin una longitud dintre dels límits d'especificació?

8. En un procés es fabriquen tubs de 15 cm mitjancant la connexió de dos trams de 5 i 10 cm respectivament. Els trams es fabriquen en processos independents i les seves longituds L_1 i L_2 són variables normals amb $\mu_{L_1} = 5$, $\sigma_{L_1} = 0.245$, $\mu_{L_2} = 10$, i $\sigma_{L_2} = 0.316$. El tub final només és utilitzable si la seva longitud, L és tal que $14 \leq L \leq 16$. Calculeu el percentatge de tubs que resulten no utilitzables en aquest procés.

9. Es fabriquen independentment uns determinats cilindres i uns determinats pistons. Cada pistó s'ha d'ajustar dintre d'un cilindre. L'assignacio d'un pistó a un cilindre es fa a l'atzar. Es sap que tant el diàmetre dels pistons com el dels cilindres segueixen una distribució normal. El diàmetre extern del pistó té mitjana 99.7 mm i desviació típica de 0.15 mm, mentre que la mitjana del diàmetre intern del cilindre es de 100.2 mm i la seva desviació típica de 0.20 mm.

   (a) Si s'escull a l'atzar un pistó i un cilindre, quina es la probabilitat que el pistó no entri dintre el cilindre?

† Feu-ho un cop hagueu estudiat la propietat reproductiva de les variables normals.

(b) Si es considera que per tenir un ajust acceptable la diferència entre el diàmetre interior del cilindre i el diàmetre exterior del pistó no pot ser superior a 0.5 mm, quina serà la proporció de parells de pistons i cilindres que ajustaran correctament?

(c) Quina hauria de ser la tolerància màxima per la diferència entre diàmetres per tal de tenir un ajust de com a mínim el 95%?

10. El procés d'emplenament automàtic d'ampolles d'un determinat tipus de beguda es fa mitjancant l'abocament independent de dos compostos liquids A i B, que es barregen per donar el producte desitjat. El volum X_A de A abocat, en centímetres cúbics, segueix una distribució normal amb $\mu_{X_A} = 20$ i $\sigma_{X_A} = 2.5$ mentre que el volum, X_B de B abocat, també en centímetres cúbics, segueix una distribució normal amb $\mu_{X_B} = 10$ i $\sigma_{X_A} = 1.5$. Una ampolla plena es considera correcta sempre que el volum abocat de A sigui a l'interval $[15, 25]$ i el de B no sigui superior a 12. En cas que algun d'aquests dos requeriments falli, l'ampolla es considera defectuosa. Les ampolles un cop plenes s'emmagatzemen en caixes de 8 unitats.

    (a) Cada ampolla té una capacitat maxima de 38 centímetres cúbics. Independentment de si una ampolla resulta defectuosa o no, es vol saber si és molt frequent el fet que la quantitat total abocada superi els $38\,\mathrm{cm}^3$. Calculeu la proporció d'ampolles per les que no hi ha vessament de liquid per excés.

    (b) Calculeu la proporció d'unitats defectuoses que genera aquest procés.

    (c) Quina es la probabilitat que en una caixa de 8 unitats, emplenades de forma independent, es trobi més d'una ampolla defectuosa?

    (d) Si el procés permet ajustar la desviació típica de la quantitat de A abocada, calculeu fins a quin valor s'hauria de reduïr el valor de σ_{X_A} sense modificar X_B per tal que la proporció d'unitats defectuoses generada pel procés es reduís a 0.10.

11. La probabilitat d'efectuar erròniament una operació concreta en un procés de producció d'una unitat AE35 és de 0.015. En aquest procés de producció l'operació es repeteix de forma independent 1000 vegades. Calculeu *aproximadament*, la probabilitat d'efectuar menys de 20 operacions errònies en el procés de producció de la peça AE35.

12. La proporció de xips defectuosos en la planta experimental de producció d'industries ACME és de 29%. Si a la producció "en fase experimental"es formen lots de 1000 xips fabricats de forma independent, quina és la probabilitat que en un lot hi hagi més de 250 xips defectuosos?

13. Volem fer una inspecció ambiental. Prenem mostres de l'aigua d'un riu a la sortida d'una empresa que estem auditant. Suposem que el nombre de partícules per cm^3 d'un determinat residu tòxic és pot modelitzar com una variable de Poisson amb mitjana 1000. Prenem $1\,\mathrm{cm}^3$ d'aquesta aigua, quina és la probabilitat de trobar com a mínim 950 partícules?

14. En una partida de cable de fibra òptica observem un cert tipus de defecte que es presenta de forma aleatòria i independent. En mitjana es presenten tres defectes per metre. Quina és la probabilitat que en un tram de 3m hi hagi més de 10 defectes?

15. Per fabricar un teixit que recobrirà un enginy espacial, s'usa una fibra sintètica. La resistència a la tensió d'un fil d'aquesta fibra té una distribució $\mathrm{N}(75.5, 3.5)$ (les xifres venen donades en unitats de tensió u.t.). Agafem una mostra aleatòria de 9 espècimens de fibra. Quina és la probabilitat que la mitjana de la resistència a la tensió a la mostra sigui major que 75, 75u.t.?

16. Fabriquem resistències de 10Ω. Les resistències no sempre surten igual. Desconeixem com modelitzar la distribució aleatòria dels valors de la resistència, però la mitja la nostra producció és de 10Ω, amb una desviació típica de 1Ω. Si els lots de resistències contenen 100 unitats. Quina

és la probabilitat aproximada que si escollim un lot al atzar, la mitjana de les resistències sigui menor de $9,8\Omega$?

17. Donada una població X de *dades normals* amb mitjana μ i desviació típica σ, es pren una mostra $\{X_1,\ldots,X_{20}\}$ i es calcula la mitjana $\overline{X}$ i la desviació típica mostral S. Per estudiar la variable aleatòria $\dfrac{X-\mu}{S/\sqrt{20}}$ haurem de consultar les taules de la distribució...

    (A) T-Student amb 20 graus de llibertat. (B) χ^2 amb 19 graus de llibertat. (C) $N(0,1)$
    (D) T-Student amb 19 graus de llibertat. (E) χ^2 amb 20 graus de llibertat.

18. Donada una població, es pren una mostra $X_1,\ldots,X_{20}$. Un estimador sense biaix del paràmetre σ^2 és

    (A) $\frac{\sum_{i=1}^{20}(X_i-\bar{X})^2}{20}$ (B) $\sqrt{\sum_{i=1}^{20}\left|X_i-\bar{X}\right|^2}$ (C) $\frac{\sum_{i=1}^{20}(X_i-\bar{X})^2}{19}$ (D) La variable χ^2_{20} (E) Cap de les anteriors.

5.2 Solucions

1. (a) $P(X > 2/3) = \int_{2/3}^{\infty} f_X(x)dx = \int_{2/3}^{1} 2(1-x)dx = \left(2x - x^2\right|_{2/3}^{1} = 1/9$

   (b)

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx = \int_0^1 2x(1-x)dx = \left(x^2 - \frac{2x^3}{3}\right|_0^1 = 1/3$$

   i

$$V(X) = \int_{-\infty}^{\infty} x^2\, f_X(x)dx \; - \; (E(X))^2 = \int_0^1 2x^2(1-x)dx \; - \; \left(\frac{1}{3}\right)^2 =$$
$$= \left(\frac{2x^3}{3} - \frac{-x^4}{2}\right|_0^1 - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

   (c) $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$ per tant fàcilment tenim:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 2x - x^2 & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

2. (a) $P(X \leq 2) = F_X(2) = 0.86466$.

   (b) $f_X(x) = F'_X(x)$ per tant:

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \mathrm{e}^{-x} & \text{si } x \geq 0 \end{cases}$$

3. (a)

$$P(T < 50) = \int_0^{50} \frac{1}{100}\mathrm{e}^{-t/100}dt = -\left(\mathrm{e}^{-t/100}\right|_0^{50} = -(\mathrm{e}^{-0.5} - 1) = 1 - \mathrm{e}^{-0.5} \simeq 0.3934$$

   (b)

$$P(T > 100) = \int_{100}^{+\infty} \frac{1}{100}\mathrm{e}^{-t/100}dt = -\left(\mathrm{e}^{-t/100}\right|_{100}^{+\infty} = -(\lim_{t\to+\infty} \mathrm{e}^{-t/100} - \mathrm{e}^{-1}) = \mathrm{e}^{-1} \simeq 0.3678$$

   (c)

$$P(T > 100|T > 50) = \frac{P(T > 100 \cap T > 50)}{P(T > 50)} = \frac{P(T > 100)}{P(T > 50)} =$$
$$= \frac{\mathrm{e}^{-1}}{1 - (1 - \mathrm{e}^{-0.5})} = \frac{\mathrm{e}^{-1}}{\mathrm{e}^{-0.5}} = \mathrm{e}^{-0.5} \simeq 0.6065$$

   Fixem-nos que també $P(T > 50) = \mathrm{e}^{-0.5}$ i que per tant, $P(T > 100|T > 50) = P(T > 50)$ la qual cosa és sorprenent (penseu si això també ens passa a nosaltres, les persones!).

   Aquesta és una propietat general, és coneguda com la *propietat de carència de memòria* de les variables exponencials,

$$P(X > s + t|X > t) = P(X > s) \text{ per tot } s, t > 0. \tag{1}$$

   D'altra banda la propietat de carència de memòria és una característica única d'aquestes variables, ja que tota variable contínua‡ que satisfà l'equació (1) és una variable exponencial per algun paràmetre $\lambda > 0$. Busqueu, si voleu, alguna referència a on es demostri això.

‡La distribució geomètrica, que és discreta, té una propietat anàloga: Si $X \sim G(p)$ aleshores $P(X \geq s + t|X > t) = P(X \geq s)$ per tot $s, t \in \mathbb{N}$.

A partir d'ara denotem $Z \sim \mathrm{N}(0,1)$ i Φ la seva funció de distribució.

4. Considerem $R \sim \mathrm{N}(10, 1^2)$.

   (a) Tipificant tenim:

   $$P(R > 9.5) = P\left(\frac{R-10}{1} > \frac{9.5-10}{1}\right) = P(Z > -0.5) = 1 - \Phi(-0.5) \simeq 0.6915$$

   (b)

   $$P(8 < R < 9.5) = P\left(\frac{8-10}{1} < \frac{R-10}{1} < \frac{9.5-10}{1}\right) =$$
   $$= P(-2 < Z < -0.5) = \Phi(-0.5) - \Phi(-2) \simeq 0.2858$$

   (c) Aquí no ens pregunten pel comportament de R sinó pel de la variable aleatòria $\bar{R}$, hem de fer servir una conseqüència de la propietat reproductiva de les variables normals segons la qual si $R \sim \mathrm{N}(\mu, \sigma^2)$ aleshores $\bar{R} \sim \mathrm{N}(\mu, (\sigma/\sqrt{n})^2)$ (vegeu l'equació (3) més avall). Per tant $\bar{R} \sim \mathrm{N}(10, 0.1^2)$, i tipificant tenim:

   $$P(\bar{R} < 9.5) = P\left(\frac{\bar{R}-10}{0.1} < \frac{9.5-10}{0.1}\right) = P(Z < -5) = \Phi(-5) \simeq 0.$$

   Observeu que $P(R < 9.5) = 0.3085$. És clar, $\bar{R}$ té menys dispersió que R!

5. Considerem $D \sim \mathrm{N}(0.25, 0.03^2)$.

   (a) Tipificant tenim:

   $$P(0.24 < D < 0.26) = P\left(\frac{0.24-0.25}{0.03} < \frac{D-0.25}{0.03} < \frac{0.26-0.25}{0.03}\right) =$$
   $$\simeq P(-0.33 < Z < 0.33) = \Phi(0.33) - \Phi(-0.33) \simeq 0.2586$$

   (b)

   $$P(0.24 < D < 0.26) = P\left(\frac{0.24-0.25}{\sigma} < \frac{D-0.25}{\sigma} < \frac{0.26-0.25}{\sigma}\right) =$$
   $$\simeq P\left(\frac{-0.01}{\sigma} < Z < \frac{0.01}{\sigma}\right) = \Phi\left(\frac{0.01}{\sigma}\right) - \Phi\left(\frac{-0.01}{\sigma}\right)$$

   Per tant
   $$P(0.24 < D < 0.26) = \Phi\left(\frac{0.01}{\sigma}\right) - \Phi\left(\frac{-0.01}{\sigma}\right) = 0.9$$

   Per la simetria de Φ respecte a l'eix d'ordenades obtenim que $\Phi\left(\frac{0.01}{\sigma}\right) = 0.95$, o equivalentment $\Phi\left(\frac{-0.01}{\sigma}\right) = 0.05$, i per tant mirant a les taules

   $$\frac{0.01}{\sigma} \simeq 1.645 \text{ i per tant } \sigma = 6.079 \cdot 10^{-3}.$$

6. Considerem $A \sim \mathrm{N}(0.9, 0.003^2)$.

   (a) Tipificant obtenim:

   $$P(\text{defectuós}) = P(A < 0.895 \cup A > 0.905) = 1 - P(0.895 < A < 0.905) =$$
   $$\simeq 1 - P(-1.67 < Z < 1.67) = 1 - (\Phi(1.67) - \Phi(-1.67)) \simeq$$
   $$\simeq 1 - (0.9525 - 0.0475) = 0.0969$$

(b)

$$P(\text{defectuós}) = 1 - P(0.895 < A < 0.905) = 1 - P\left(\frac{0.895-0.9}{\sigma} < Z < \frac{0.905-0.9}{\sigma}\right) =$$

$$= 1 - P\left(\frac{-0.005}{\sigma} < Z < \frac{0.005}{\sigma}\right) \leq 0.01.$$

Per tant,

$$P\left(\frac{-0.005}{\sigma} < Z < \frac{0.005}{\sigma}\right) = \Phi\left(\frac{0.005}{\sigma}\right) - \Phi\left(\frac{-0.005}{\sigma}\right) \geq 0.99,$$

i per la simetria de Φ tenim,

$$\Phi\left(\frac{0.005}{\sigma}\right) \geq 0.995,$$

i mirant les taules obtenim:

$$\frac{0.005}{\sigma} \simeq 2.58 \text{ i per tant } \sigma = 1.938 \cdot 10^{-3}.$$

7. Considerem $L \sim \mathrm{N}(90.5, 0.1^2)$

(a) Tipificant obtenim $P(\text{sobrepassi}) = 1 - P(89.7 < L < 90.3) \simeq 0.9773$, molt alta.

(b) Si feu un esboç de la gràfica de la funció de densitat veureu que convé "centrar"el procés a $\mu = 90$.

(c) El rendiment vol dir la proporció de correctes, si ara prenem $L \sim \mathrm{N}(90, 0.1^2)$, i tipifiquem tenim $P(89.7 < L < 90.3) \simeq 0.9973$.

(d) Considerem $L \sim \mathrm{N}(90, 0.1^2)$. Prendre 10 mesures de forma independent, vol dir considerar 10 variables independents $L_1, \ldots, L_{10}$ amb distribució $\mathrm{N}(90, 0.1^2)$.

Recordem que quan tenim n variables independents $X_1, \ldots, X_n$, i I és un interval[§]*, aleshores*

$$P(X_1 \in I \cap X_2 \in I \cap \cdots \cap X_n \in I) = P(X_1 \in I) \cdots P(X_2 \in I) \cdots P(X_n \in I)$$

Per tant, si $I = [89.7, 90.3]$,

$$P(L_1 \in I \cap L_2 \in I \cap \cdots \cap L_{10} \in I) = P(L_1 \in I) \cdots P(L_2 \in I) \cdots P(L_{10} \in I) =$$

$$= P(L \in I)^{10} = 0.9973^{10} \simeq 0.9733$$

(e) Fixem-nos que Y és el número d'estoigs dins els límits d'especificació, aleshores $Y \sim \mathrm{Bin}(10, 0.9973)$, i.e. $n = 10$ i $p = 0.9973$. Per tant,

$$E(Y) = n\,p = 10 \cdot 0.9973 = 9.973.$$

A partir d'ara, hem de tenir en compte la propietat reproductiva de les variables normals que diu:

Si $X = c_1 X_1 + \cdots + c_n X_n$ on $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ són variables independents, aleshores

$$X \sim \mathrm{N}(\mu, \sigma^2) \text{ amb } \mu = c_1\,\mu_1 + \cdots + c_n\mu_n, \text{ i } \sigma^2 = c_1^2\,\sigma_1^2 + \cdots + c_n^2\sigma_n^2. \tag{2}$$

[§] de fet qualsevol conjunt mesurable de $\mathbb{R}$.

Una conseqüència important d'aquesta propietat és

Si $\bar{X} = (X_1 + \cdots + X_n)/n$ *on* $X_i \sim \mathrm{N}(\mu, \sigma^2)$ *són variables independents, aleshores*

$$\bar{X} \sim \mathrm{N}(\mu, (\sigma/\sqrt{n})^2). \tag{3}$$

O, equivalentment,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

8. Si $L_1 \sim \mathrm{N}(5, 0.245^2)$ i $L_2 \sim \mathrm{N}(10, 0.316^2)$ són variables independents, aleshores per la propietat reproductiva, vegeu equació (2), tenim $L \sim \mathrm{N}(\mu, \sigma^2)$, on

$$\mu = 5 + 10 = 15 \;\text{ i }\; \sigma = \sqrt{0.245^2 + 0.316^2} \simeq 0.3999.$$

   Considerem, doncs, $L \sim \mathrm{N}(15, 0.3999^2)$.

$$P(\text{no utilitzable}) = 1 - P(14 \leq L \leq 16) = 1 - P\left(\frac{14 - 15}{0.3999} \leq Z \leq \frac{16 - 15}{0.3999}\right) =$$

$$= 1 - P(-2.5009 \leq Z \leq 2.5009) \simeq 1 - (\Phi(2.50) - \Phi(-2.50)) \simeq 0.0124$$

9. Siguin $P \sim \mathrm{N}(99.7, 0.15^2)$ i $C \sim \mathrm{N}(100.2, 0.2^2)$ les varaiables que mesuren els diàmetres dels pistons i dels cilindres, respectivament.

   (a) $P(\text{no entri}) = P(C < P) = P(C - P < 0)$. Considerem, doncs, la variable $Y = C - P \sim \mathrm{N}(\mu, \sigma^2)$, on

$$\mu = 100.2 - 99.7 = 0.5 \;\text{ i }\; \sigma = \sqrt{(1)^2(0.2)^2 + (-1)^2(0.15)^2} = \sqrt{(0.2)^2 + (0.15)^2} = 0.25.$$

   Per tant

$$P(C - P < 0) = P(Y < 0) = P\left(\frac{Y - 0.5}{0.25} < \frac{0 - 0.5}{0.25}\right) = P(Z < -2) = 0.0228$$

   (b)

$$P(\text{acceptables}) = P(0 < C - P < 0.5) = P(0 < Y < 0.5) =$$

$$= P\left(\frac{0 - 0.5}{0.25} < \frac{Y - 0.5}{0.25} < \frac{0.5 - 0.5}{0.25}\right) = P(-2 < Z < 0) = \Phi(0) - \Phi(-2) \simeq 0.4772$$

   (c) Busquem la tolerància ε tal que $P(0 < C - P < \varepsilon) \geq 0.95$. Tipificant

$$P\left(-2 < Z < \frac{\varepsilon - 0.5}{0.25}\right) \geq 0.95$$

   per tant

$$\Phi\left(\frac{\varepsilon - 0.5}{0.25}\right) - \Phi(-2) \geq 0.95 \;\Rightarrow\; \Phi\left(\frac{\varepsilon - 0.5}{0.25}\right) \geq 0.95 + 0.0228 = 0.9728$$

   Mirant les taules tenim:

$$\frac{\varepsilon - 0.5}{0.25} \simeq 1.93 \;\Rightarrow\; \varepsilon \simeq 0.9825$$

10. Considerem $X_A \sim \mathrm{N}(20, 2.5^2)$ i $X_B \sim \mathrm{N}(10, 1.5^2)$. L'ampolla es considera correctament omplerta si $15 \leq X_A \leq 25$ i $X_B \leq 22$.

(a) Observem que $X = X_A + X_B \sim \mathrm{N}(\mu, \sigma^2)$ on

$$\mu = 20 + 10 = 30 \;\text{ i }\; \sigma = \sqrt{2.5^2 + 1.5^2} \simeq 2.9155.$$

Per tant

$$P(X_A + X_B \leq 38) = P\left(\frac{X_A + X_B - 30}{2.9155} \leq \frac{38 - 30}{2.9155}\right) \simeq P(Z \leq 2.74) \simeq 0.9969$$

(b) Farem servir que els processos d'abocament dels dos líquids són independents

$$P(\text{defectuoses}) = 1 - P(\text{correctes}) = 1 - P(15 \leq X_A \leq 25 \cap X_B \leq 22) =$$

$$= 1 - P(15 \leq X_a \leq 25)\, P(X_B \leq 22).$$

Calculem ara, tipificant

$$P(15 \leq X_a \leq 25) \simeq 0.9545 \;\text{ i }\; P(X_B \leq 22) \simeq 0.9082.$$

Per tant,

$$P(\text{defectuoses}) \simeq 1 - 0.9545 \cdot 0.9082 = 0.1331$$

(c) La variable X que mesura el número d'ampolles defectuoses en una caixa de 8 unitats té distribució $\mathrm{Bin}(8, 0.1331)$. Per tant,

$$P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) + P(X = 1) =$$

$$= 1 - \binom{8}{0} 0.1331^0\, 0.8669^{10} - \binom{8}{1} 0.1331^1\, 0.8669^9 = 1 - 0.23971214 - 0.294434752 \simeq 0.2895$$

(d) Imposem ara

$$P(\text{defectuoses}) = 1 - P(15 \leq X_A \leq 25)\, P(X_B \leq 22) = 0.1.$$

Recordem que $P(X_B \leq 22) \simeq 0.9082$. Per tant

$$1 - P(15 \leq X_A \leq 25) \cdot 0.9082 = 0.1 \;\Rightarrow\; P(15 \leq X_A \leq 25) = \frac{0.9}{0.9082} \simeq 0.9910.$$

Per altra banda

$$P(15 \leq X_A \leq 25) = P\left(\frac{15 - 20}{\sigma} \leq Z \leq \frac{25 - 20}{\sigma}\right) = \Phi\left(\frac{5}{\sigma}\right) - \Phi\left(\frac{-5}{\sigma}\right) = 0.9910.$$

Fent servir la simetria de Φ obtenim $\Phi\left(\frac{5}{\sigma}\right) \simeq 0.9955$, i consultant les taules obtenim

$$\frac{5}{\sigma} \simeq 2.61 \;\Rightarrow\; \sigma \simeq 1.9157$$

Per fer els seguents exercicis heu de fer servir que les variables binomials i Poisson es poden aproximar per variables normals. Resumint, tenim[¶] :

[¶]La següent informació la podreu trobar a la Secció 4-7 del llibre D.C. Montgomery & G.C. Runger, *Probabilidad y Estadística aplicadas a la Ingeniería*, Mc. Graw-Hill 1994. També a la Secció 4.3 del llibre J. Devore, *Probabilidad y Estadística para Ingeniería y Ciencias*, 4a ed. International Thomson 1998.

Sigui X és una variable aleatòria amb distribució $\mathrm{Bin}(n,p)$*, aleshores si n és prou gran i la funció de probabilitat de X no està massa esbiaxada, tenim que* $X \simeq \mathrm{N}(np, \sqrt{np(1-p)})$*, o equivalentment:*

$$\frac{X-np}{\sqrt{np(1-p)}} \simeq Z \sim \mathrm{N}(0,1). \tag{4}$$

I també

Sigui X és una variable aleatòria amb distribució $P(\lambda)$, aleshores si λ és prou gran s'obté $X \simeq \mathrm{N}(\lambda, \sqrt{\lambda})$*, o equivalentment:*

$$\frac{X-\lambda}{\sqrt{\lambda}} \simeq Z \sim \mathrm{N}(0,1). \tag{5}$$

En alguns llibre trobareu com a *regla pràctica*, que això ho podrem fer servir si n és gran, $np > 5$, i $n(1-p) > 5$ en el cas Binomial, i $\lambda > 5$ en el cas Poisson.

També cal tenir en compte una tècnica molt senzilla coneguda com a *correcció per continuïtat* o *correcció de Yates* que sol millorar els resultats de l'aproximació de les variables discretes per contínues.

Per exemple: sigui X una variable discreta que pren valors enters, i volem calcular $P(X \leq k)$, per $k \in \mathbb{Z}$. Si Y és una variable contínua que l'aproxima, aleshores

$$P(X \leq k) = P(X \leq k+0.5) \simeq P(Y \leq k+0.5).$$

De la mateixa manera podem fer servir:

$$\begin{array}{l} P(X < k) = P(X \leq k-1) = P(X \leq k-1+0.5 =)P(X \leq k-0.5) \simeq P(Y \leq k-0.5), \\ P(k \leq X) = P(k-0.5 \leq X) \simeq P(k-0.5 \leq Y), \\ P(k_1 \leq X \leq k_2) \simeq P(k_1 - 0.5 \leq N \leq k_2 + 0.5), \end{array}$$

on k, k_1 i $k_2 \in \mathbb{Z}$.

Per exemple, si X té distribució $\mathrm{Bin}(100, 0.5)$, podem aproximar-la amb una variable Y amb distribució normal $\mathrm{N}(50, 5)$. Així, si volem calcular $P(X \leq 40)$, podem fer el següent:

$$P(X \leq 40) = P(X \leq 40.5) \simeq P(Y \leq 40.5) = P(Z \leq (40.5-50)/5) = P(Z \leq -1.9) \simeq 0.0287.$$

11. Si X compta el número d'operacions errònies $X \sim \mathrm{Bin}(1000, 0.015)$, per tant

$$P(X < 20) = \sum_{k=0}^{19} \binom{1000}{k} 0.015^k \, 0,985^{1000-k}.$$

Un camí, doncs, feixuc. Provem d'aproximar X amb una variable normal: com que $np = 15 > 5$ i $n(1-p) = 985 > 5$, considerem $X \simeq Y$ on $Y \sim \mathrm{N}(np, \sqrt{np(1-p)}) = \mathrm{N}(15, 3.8438^2)$. Aplicant doncs l'equació (4) i la correcció de Yates:

$$P(X < 20) = P(X \leq 19) = P(X < 19.5) \simeq P(Y < 19.5) = P\left(\frac{Y-15}{3.8438} < \frac{19.5-15}{3.8438}\right) \simeq$$

$$\simeq P(Z < 1.1707) \simeq 0.879.$$

12. Si X compta el número de xips defectuosos, aleshores $X \sim \text{Bin}(1000, 0.29)$, per tant:

$$P(X > 250) = 1 - P(X \leq 250) = 1 - \sum_{k=0}^{250} \binom{1000}{k} 0.29^k \, 0.71^{1000-k}.$$

Provem d'aproximar X amb una variable normal: com que $np = 290 > 5$ i $n(1-p) = 710 > 5$, considerem $X \simeq Y$ on $Y \sim \text{N}(np, \sqrt{np(1-p)}) = \text{N}(290, 14.3492^2)$. Aplicant doncs l'equació (4) i la correcció de Yates:

$$P(X > 250) = 1 - P(X \leq 250) = 1 - P(X \leq 250.5) \simeq$$
$$\simeq 1 - P(Y < 250.5) = P\left(\frac{Y - 290}{14.3492} < \frac{250.5 - 290}{14.3492}\right) \simeq$$
$$\simeq 1 - P(Z < -2.7876) \simeq 1 - 0.0026 = 0.9974.$$

13. Si X compta el número de partícules, aleshores $X \sim \text{P}(1000)$, per tant:

$$P(X \geq 950) = 1 - P(X \leq 949) = 1 - \sum_{k=0}^{949} \mathrm{e}^{-1000} \frac{1000^k}{k!}.$$

Aproximarem X amb una variable normal $Y \sim \text{N}(1000, \sqrt{1000}^2)$: com que $\lambda = 1000 > 5$ podem aplicar l'equació (5) i la correcció de Yates, obtenint

$$P(X \geq 950) = 1 - P(X \leq 949) = 1 - P(X \leq 949.5) \simeq 1 - P(Y < 949.5) =$$
$$= 1 - P\left(\frac{Y - 1000}{\sqrt{1000}} < \frac{945.5 - 1000}{\sqrt{1000}}\right) \simeq 1 - P(Z < -1.5969) \simeq$$
$$\simeq 1 - P(Z < -1.60) = 1 - 0.0548 = 0.9454.$$

14. Si D mesura els defectes cada 10m^2, aleshores $D \sim \text{P}(9) \simeq \text{N}(9, 3^2)$. Per tant

$$P(D > 10) = 1 - P(D \leq 10) = 1 - P(D \leq 10.5) \simeq 1 - P(Z \leq 0.5) \simeq 1 - 0.6915 = 0.3085.$$

Per fer els següents exercicis cal que recordeu, que *si* $\bar{X} = (X_1 + \cdots + X_n)/n$ *on* $X_i \sim \text{N}(\mu, \sigma^2)$ *són variables independents, aleshores*

$$\bar{X} \sim \text{N}(\mu, (\sigma/\sqrt{n})^2) \;\Rightarrow\; \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{N}(0, 1). \tag{6}$$

Això és, de fet més general: *Si* $X_1, X_2, \cdots, X_n$ *és una mostra aleatòria d'una variable* X *amb qualsevol distribució, amb mitjana* μ *i desviació típica* σ*, aleshores: (a)* $E(\bar{X}) = \mu_{\bar{X}} = \mu$*; (b)* $V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n$*; (c) la desviació típica de* $\bar{X}$ *és* $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

Però principalment cal que tingueu present aquesta versió del *Teorema del límit central*[‖], que anomenarem TCL d'ara en endavant:

Si $X_1, X_2, \cdots, X_n$ *és una mostra aleatòria d'una variable* X *amb qualsevol distribució (per exemple desconeguda), amb mitjana* μ *i desviació típica* σ*. Aleshores, si* n *és prou gran,* $\bar{X} \simeq \text{N}\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$.

[‖] La següent informació la podreu trobar a la Secció 6-5 del llibre d'en Montgomery & Runger *op. cit.*, o a la Secció 5.4 del llibre J. Devore, *op. cit.*

15. Considerem que la resisteència ve donada per la variable $R \sim \mathrm{N}(75.5, 3.5^2)$, és a dir $\mu = 75.5$ i $\sigma = 3.5$. Si considerem mostres de tamany $n = 9$ aleshores

$$\bar{R} \sim \mathrm{N}\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) = \mathrm{N}(75.5, 1.1667^2).$$

Per tant:

$$P(\bar{R} > 75.75) = P\left(\frac{\bar{R} - 75.75}{1.1667} > \frac{75.75 - 75.5}{1.1667}\right) = P(Z > 0.2143) =$$
$$= 1 - P(Z \leq 0.2143) = 1 - 0.58317 = 0.4168$$

16. Sigui R la variable que mesura la resistència. Desconeixem la seva distribució, però sabem que $\mu = E(X) = 10$ i $\sigma = \sqrt{V(X)} = 1$. Si considerem que la mostra de tamany $n = 100$ és prou gran, pel TCL obtenim

$$\bar{X} \simeq \mathrm{N}\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) = \mathrm{N}(10, 0.1^2).$$

Per tant,

$$P(\bar{R} < 9.8) = P\left(\frac{\bar{R} - 10}{0.1} < \frac{9.8 - 10}{0.1}\right) \simeq P(Z < -2) = 0.0228.$$

17. (D) T-Student amb 19 graus de llibertat.

18. L'estimador sense biaix del paràmetre σ^2 és $\frac{\sum_{i=1}^{20}(X_i - \bar{X})^2}{19}$, que es sol denotar com a S^2_{n-1}, S^2, $\widetilde{S}$ i anomenar variança mostral corregida. La resposta és (C).

6 INTERVALS DE CONFIANÇA I CONTRAST D'HIPÒTESIS

6.1 Exercicis

6.1.1 Intervals de confiança.

1. Un fabricant de fibres sintètiques vol estimar la tensió mitjana de ruptura d' una fibra. Dissenya un experiment en el que s'observen les tensions de ruptura de setze fils del procés, seleccionats aleatòriament. Les tensions són 20.8, 20.6, 21.0, 20.9, 19.9, 20.2, 19.8, 19.6, 20.9, 21.1, 20.4, 20.6, 19.7, 19.6, 20.3, 20.7.

   (a) Comproveu que si és possible suposar que la tensió de ruptura està distribuïda normalment amb desviació típica 0.45, llavors $(20.1608, 20.6018)$ és un interval de confiança del 95% per al valor mitjà, μ, de la tensió de ruptura de la fibra.

   (b) Comproveu que si no es pot suposar coneguda la desviació típica, l'interval del 95% per μ és $(20.1025, 20.66)$

   (c) Compareu els resultats.

   (d) Comproveu que $(0.38669, 0.80959)$ és un interval de confiança del 95% per σ.

   (e) Quin error es comet, com a màxim, en l'estimació de μ, per una confiança del 95%, al prendre $\mu = 20.3813$, si es considera σ coneguda i igual a 0.45? I si no es pot considerar σ coneguda?

   (f) Quin error es comet, com a màxim, en l'estimació de σ al prendre $\sigma = 0.45$, per una confiança del 95%?

   (g) Quin tamany de mostra mínim és necessari per tal de poder estimar μ amb un error màxim de $\varepsilon = 0.1$ amb una confiança del 95%?

2. En el gruix de les làmines de plàstic produïdes per una certa màquina es detecta una variació aleatòria. Per tal d'esbrinar els límits d'aquesta oscil.lació, es seleccionen cada dia, de forma aleatòria, dotze làmines i se les mesura el gruix. Les dades obtingudes són 12.6, 11.9, 12.3, 12.8, 11.8, 11.7, 12.4, 12.1, 12.3, 12.0, 12.5, 12.9.

   (a) Si aquest gruix es una variable que es pot suposar normal, comproveu que l'interval de confiança del 99% per a la desviació típica desconeguda del gruix és $(0.247762, 0.794357)$.

   (b) Si no és acceptable una desviació típica superior a 0.9 mm, té el fabricant, en base a aquesta evidència, prou raons per a preocupar-se?

   (c) Comproveu que l'interval de confiança del 95% per a que s'obté a partir d'aquesta mostra μ és $(12.0295, 12.5205)$.

   (d) Quin error es comet, com a màxim, suposant $\mu = 12.275$? I suposant $\mu = 12.5$?

3. Es pot suposar, en base a l'experiència, que la resistència a la tensió X, per a una determinada fibra sintètica es distribueix de forma aproximadament normal. Es selecciona una mostra aleatòria de setze trossos de fibra i es determina, per a cada cas, la seva resistència. La mitjana i la desviació típica mostrals, per a aquesta mostra, són $\overline{X} = 49.86$ i $S = 1.66$ respectivament. En base a aquesta informació es calcula, $(48.9754, 50.7446)$, l'interval de confiança del 95% per $E(X)$.

(a) Hi ha evidències estadístiques que permetin assumir que $E(X) = 50$ amb una confiança del 95% ?

(b) Si la resposta és afirmativa, quin és l'error màxim comès, amb probabilitat 0.95, al fer $E(X) = 50$?

4. La durada mitjana d'un determinat tipus de pneumàtic s'estima en uns 35759.6 Km, amb una desviació típica de 59.8 Km.

   (a) Si aquests valors s'han calculat en base a una mostra de 10 pneumàtics, comproveu que els intervals de confiança del 95% per a aquestes dues característiques són $(35716.8, 35802.4)$, i $(41.1325, 109.171)$ respectivament.

   (b) Si aquests valors s'han calculat en base a una mostra de 100 pneumàtics, comproveu que els intervals de confiança del 95 % són $(35747.7, 35771.5)$, i $(52.5048, 69.4682)$ respectivament.

5. Una màquina està ajustada de manera que la quantitat de líquid que expulsa es distribueix aproximadament segons una llei normal amb desviació estàndar igual a 0.15 decilitres. Una mostra de 36 expulsions ha donat una mitjana de 2.25 decilitres.

   (a) Comproveu que $(2.201, 2.299)$ és un interval del 95% per a la quantitat mitjana teòrica, μ, expulsada.

   (b) Quin hauria de ser el tamany de la mostra si es vol assegurar amb una confiança del 95% un error inferior als 0.02 decilitres?

6. En una mostra de cent peces, escollida a l'atzar d'entre tota la producció, hi ha 55 que presenten un determinat tipus de defecte. Determineu els límits de confiança del 95%, per a la proporció, p, de peces defectuoses en el total de la producció.

7. Quin serà el tamany de mostra apropiat per a estimar la proporció d'individus que tenen una determinada característica en una població concreta, de manera que l'error comès sigui com a màxim 0.01 amb una confiança del 90%? Respon la mateixa pregunta però ara sabent que s'ha fet una prova pilot i en una mostra de 100 individus n'hi havia quinze que presentaven la característica anterior.

8. El consum X d'una determinada màquina es comporta pràcticament com una variable normal. Per estimar el consum mitjà teòric es mesura el consum de 77 màquines que operen de forma independent obtenint-se $\overline{X} = 22.5468$. i $S^2 = 14.3094$. En base a aquests resultats, el departament d'anàlisi emet un informe on diu: es pot assegurar que $\mu = 22.5468$ amb un error de ± 0.85855 i una fiabilitat del 95% i que $\sigma = 3.78277$ amb un error màxim de 0.5175 i una fiabilitat del 95%. Expliqueu el significat exacte d'aquestes afirmacions.

9. En la situació de l'exercici anterior s'han fet una serie de proves i finalment s'ha determinat que σ és exactament 3.75. Tenint en compte aquesta informació, calculeu quin és el tamany de mostra més petit que es pot fer servir per tal d'estimar μ amb un error màxim de 0.2 per una confiança de 95%.

10. La resistència d'un determinat cable, en ohms, és una variable normal i es sap que les dues màquines que produeixen el cable presenten la mateixa variabilitat. Per tal de comprovar si el cables produïts per dues màquines diferents presenten les mateixes resistències mitjanes es prenen dues mostres, una de cada màquina, obtenint-se els resultats següents: 0.136, 0.141, 0.140, 0.142, 0.139, 0.145, 0.141 per a la primera màquina i 0.136, 0.135, 0.143, 0.138, 0.140 per a la segona.

(a) Comproveu que l'interval de confiança del 95% per a la diferència de mitjanes que donen aquestes mostres és $I_{0.95}(\mu_1 - \mu_2) = (-0.00167, 0.00602)$

(b) Hi ha realment diferència significativa?

(c) És acceptable en aquest cas suposar que les variàncies són iguals sabent que l'interval del 95% pel quocient de desviacions típiques és $I_{0.95}(\sigma_1/\sigma_2) = (0.0804272, 4.60632)$?

6.1.2 Contrast d'hipòtesis.

1. El director de recursos humans d'una empresa està interessat en l'eficàcia d'un curs de seguretat en el treball dissenyat per tal que els treballadors tinguin cura de seguir les normes de seguretat establertes

   (a) Quines hipòtesis estarà contrastant aquesta persona si comet un error de tipus I al concloure equivocadament que el curs sobre seguretat no és eficaç?

   (b) Quines hipòtesis estarà contrastant aquesta persona si comet un error de tipus II al concloure equivocadament que el curs de seguretat és eficaç?

2. Una gran empresa manufacturera ha estat acusada de discriminació pel que fa a la contractació.

   (a) Quines hipòtesis s'estaran contrastant si un jurat comet un error de tipus I al trobar que l'empresa és culpable?

   (b) Quines hipòtesis s'estaran contrastant si un jurat comet un error de tipus II al trobar que l'empresa és culpable?

3. La variable $X \sim \mathrm{N}(\mu, \sigma^2)$ correspon a la resistència en psi d'un determinat aliatge. Establiu el contrast d'hipòtesis adequat per contrastar que μ supera els 155 psi i decidiu quina decisió es prendrà a partir de la mostra de X següent:

   105, 97, 245, 163, 207, 134, 218, 199, 160, 196, 221, 154, 228, 131, 180, 178, 157, 151, 175, 201, 183, 153, 174, 154, 190, 76, 101, 142, 149, 200, 186, 174, 199, 115, 193, 167, 171, 163, 87, 176, 121, 120, 181, 160, 194, 184, 165, 145, 160, 150, 181, 168, 158, 208, 133, 135, 172, 171, 237, 170, 180, 167, 176, 158, 156, 229, 158, 148, 150, 118, 143, 141, 110, 133, 123, 146, 169, 158, 135, 149.

   Nota: Aquesta és una bona ocasió per fer servir Minitab. Podeu fer l'assistent de contrast d'hipòtesis per a obtenir informació**. Només cal fer:

   `Asistente`→ `Pruebas de hipótesis`→ `Comparar una muestra con un objetivo`→ `t de 1 muestra`.

4. El temps de reparació d'una font d'alimentació és una variable distribuïda normalment amb $\mu = 3$ hores i $\sigma = 0.6$ hores. Últimament s'han modificat alguns elements d'aquest model de font amb la intenció de facilitar les reparacions i poder estalviar temps de reparació.

   (a) Plantejeu el contrast d'hipòtesis que s'haurà de dur a terme per tal verificar que les modificacions realitzades han tingut èxit.

   (b) Quin és l'estadístic de prova?

   (c) Quina és la regió crítica per $\alpha = 0.05$ i tamany de mostra $n = 25$?

** També podeu fer el següent, tot i que obtindreu molta menys informació: `Estadísticas`→ `Estadística básica`→ $1t$ `t de 1 muestra`.

(d) Si els temps observats en una mostra de mida $n = 10$, en hores, han estat 3.04, 2.74, 2.28, 3.2, 2.8, 2.52, 2.99, 2.19, 2.45, 1.89, podem concloure que les modificacions han estat un èxit?

5. El rendiment d'un procés químic és una variable $X \sim N(\mu, \sigma^2)$. La mitjana teòrica del rendiment no pot ser inferior al 90%, de manera que periòdicament es fan proves de control. L'última prova s'ha fet en base a una mostra de tamany $n = 5$ per contrastar les hipòtesis amb una confiança del 95%.

$$\begin{cases} H_0 : \mu = 90, \\ H_1 : \mu < 90. \end{cases}$$

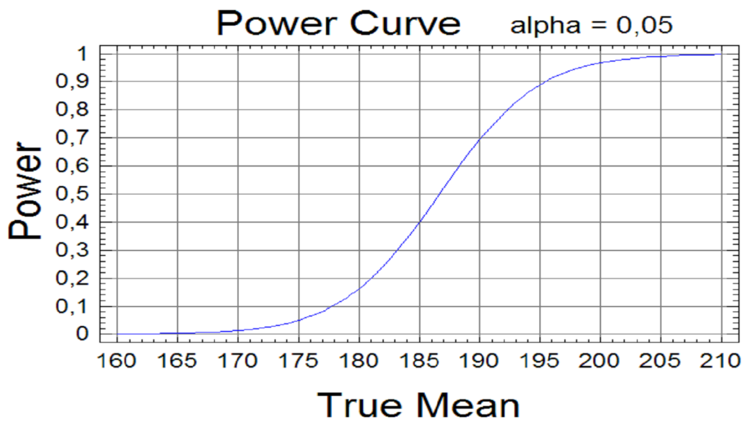
   (a) Quin és l'estadístic de prova?

   (b) Hi ha motiu per preocupar-se si s'obté $t = -4.6041$ que correspon a un *valor p* ($p-$value) de 0.005?

6. S'està desenvolupant una màquina per tallar automàticament barres d'acer. La longitud de les barres és aproximadament normal Es volen contrastar les hipòtesis $H_0 : \mu = 175$ contra $H_1 : \mu > 175$ amb una mostra de tamany $n = 10$.

   (a) Digueu quin és l'estadístic de prova i quina la regió crítica per a $\alpha = 0.05$.

   (b) La corba de potència del test és



   Quin és el risc de segona espècie si $\mu = 190$ mm?.

   (c) A quina conclusió es pot arribar si $\overline{X} = 190$ mm i $S = 20.1$?

   (d) Quina regió crítica s'haurà d'agafar si es vol que el risc de primera espècie sigui 0.01?

7. La quantitat de calor, en calories per gram, que despren una determinada combinació de productes es distribueix aproximadament segons una llei normal. S'ha comprovat que $\sigma = 2$ i es vol contrastar les hipòtesis $H_0 : \mu = 100$ contra $H_1 : \mu \neq 100$ amb una mostra de tamany $n = 9$. Per fer-ho es fixa la regió d'acceptació $|Z| = \left| \dfrac{\overline{X} - 100}{2} \sqrt{9} \right| < 2.25$.

   (a) Comproveu que la regió d'acceptació és la definida per $98.5 \leq \overline{X} \leq 101.5$.És a dir, rebutjar H_0 si la mitjana mostral és més petita que 98.5 o més gran que 101.5.

   (b) Trobeu el risc de primera espècie del test.

   (c) Calculeu el risc de segona espècie si el valor real de la quantitat de calor despresa té una mitjana teòrica de 103. Calculeu-lo també quan la mitjana teòrica és 105.

8. Tenim les variables $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ i $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$. Plantegeu el contrast d'igualtat de σ_1 i σ_2 amb $P(\text{rebutjar } H_0 | \sigma_1 = \sigma_2) = 0.05$, a partir de l'observació de dues mostres independents de X_1 i X_2 de mides n_1 i n_2 respectivament, i digueu a quina conclusió s'arriba quan:

   (a) $n_1 = 11, n_2 = 16$ i s'obté $S_1 = 9.7$ i $S_2 = 15.9$.

   (b) $n_1 = 100, n_2 = 100$ i s'obté $S_1 = 10.6$ i $S_2 = 15.9$.

9. Volem saber si podem suposar que els cilindres que ens subministren dos proveïdors diferents tenen les mateixes especificacions. Suposarem que els dos proveïdors fabriquen els cilindres amb un diàmetre que varia segons una distribució normal. Per poder decidir, amb una confiança del 95%, prenem una mostra de 30 cilindres de cadascun dels proveïdors, obtenint: $\overline{X}_1 = 5.0951$, $S_1 = 0.087$, $\overline{X}_2 = 4.8817$, $S_2 = 0.053$. Es demana:

   (a) Quin contrast d'hipòtesi plantejaríeu per decidir sobre la igualtat, o no, de les *variancies*? Quin estadístic de contrast faríeu servir? Descriviu quina és la regió crítica.

   (b) Respecte de l'anterior contrast, amb un ordinador s'ha obtingut que l'estadístic de contrast adequat val 2.69. Quina decisió preneu?

   (c) En coherència amb l'apartat anterior, quin contrast d'hipòtesi per a comparar les *mitjanes* escolliríeu. Sense calcular-lo, digueu quin estadístic de contrast faríeu servir? Deixeu indicada quina és la regió crítica?

   (d) Respecte de l'anterior contrast, amb un ordinador hem obtingut que el valor p val $\simeq 0.000$. Quina decisió preneu?

10. A partir de les dades presentades i els estadístics calculats a l'exercici 10 d'intervals de confiança, feu servir Minitab per a aplicar el contrast d'hipòtesis que considereu adequat per comparar les mitjana de resistència del cable fabricat per les dues màquines.

6.2 Solucions

6.2.1 Intervals de confiança.

1. Com que treballarem amb una confiança del 95%, tenim $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1-\alpha/2 = 0.975$. D'altra banda, $n = 16$ i amb una calculadora fàcilment obtenim: $\overline{X} = 20.38125$; $S = 0.52309177$ (i per tant $S^2 = 0.273625$).

   (a) Com que σ és coneguda, tenint en compte que $z_{0.975} = 1.96$, l'interval és (l, u), on:

   $$l = \overline{X} - z_{0.975}\frac{\sigma}{\sqrt{n}} = 20.38125 - 1.96\,\frac{0.45}{\sqrt{16}} = 20.16075 \simeq 20.1608,$$
   $$u = \overline{X} + z_{0.975}\frac{\sigma}{\sqrt{n}} = 20.38125 + 1.96\,\frac{0.45}{\sqrt{16}} = 20.60175 \simeq 20.6018.$$

   (b) Com que σ és desconeguda en aquest cas, tenint en compte que $t_{15,0.975} = 2.1314$, l'interval és (l, u), on:

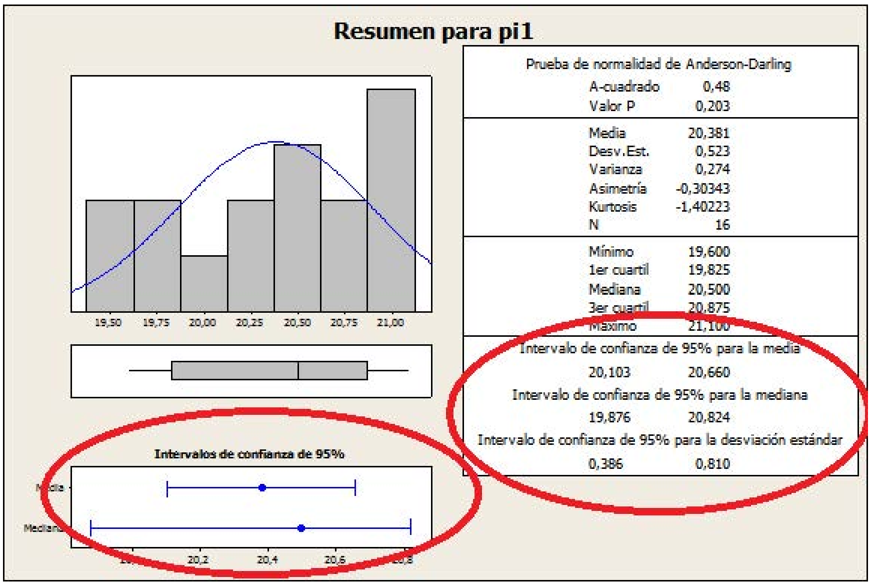
   $$l = \overline{X} - t_{15,0.975}\frac{S}{\sqrt{n}} = 20.38125 - 2.1314\,\frac{0.52309177}{\sqrt{16}} = 20.1025,$$
   $$u = \overline{X} + t_{15,0.975}\frac{S}{\sqrt{n}} = 20.38125 + 2.1314\,\frac{0.52309177}{\sqrt{16}} = 20.65998 \simeq 20.66.$$

   Potser és un bon moment per a recordar com obtenir intervals de confiança amb Minitab. Una manera és fer: `Estadísticas`$\rightarrow$ `Estadística básica`$\rightarrow$ 1t `t de 1 muestra`.

T de una muestra: pi1

Variable	N	Media	Desv.Est.	Error estándar de la media	IC de 95%
pi1	16	20,381	0,523	0,131	(20,103; 20,660)

Una altra manera de fer és: `Estadísticas`$\to$ `Estadística básica`$\to$ `Resumen gráfico`.



(c) L'interval de l'apartat (b) és menys precís. Tot i que no sempre serà així (al cap i a la fi depèn del valor de S), és força raonable que passi això ja que hi ha més incertesa, i d'altre banda sempre $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$.

(d) Tenint en compte que $\chi^2_{15,0.9025} = 6.262$ i $\chi^2_{15,0.975} = 27.488$, l'interval és (l,u), on:

$$l = \sqrt{\frac{(n-1)S^2}{\chi^2_{15,0.975}}} = \sqrt{\frac{15 \cdot 0.273625}{27.488}} = 0.3866946506 \simeq 0.38669,$$
$$u = \sqrt{\frac{(n-1)S^2}{\chi^2_{15,0.025}}} = \sqrt{\frac{15 \cdot 0.273625}{6.262}} = 0.8095934487 \simeq 0.80959.$$

(e) Els errors són

$$z_{0.975}\frac{\sigma}{\sqrt{n}} = 1.96\,\frac{0.45}{\sqrt{16}} = 0.2205, \text{ i } t_{15,0.975}\frac{\sigma}{\sqrt{n}} = 2.1314\,\frac{0.52309177}{\sqrt{16}} = 0.27873,$$

respectivament.

(f) Observem que $|0.45-l| = 0.06331$ i $|0.45-u| = 0.359593$, per tant $\max(|0.45-l|, |0.45-u|) = 0.359593$.

(g) Si suposem $\sigma = 0.45$ tenim

$$z_{0.975}\frac{\sigma}{\sqrt{n}} < 0.1 \Rightarrow 1.96\frac{0.45}{\sqrt{n}} < 0.1 \Rightarrow 1.96\frac{0.45}{0.1} < \sqrt{n} \Rightarrow 77.79... \simeq \left(1.96\frac{0.45}{0.1}\right)^2 < n.$$

Per tant $n \geq 78$.

2. Tenim $n = 12$ i amb una calculadora fàcilment trobem: $\overline{X} = 12.275$; $S = 0.386417108$ (i per tant $S^2 = 0.149318181$).

(a) Com que treballarem amb una confiança del 99%, tenim $1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow 1-\alpha/2 = 0.995$. Tenint en compte que $\chi^2_{11,0.005} = 2.603$ i $\chi^2_{11,0.995} = 26.757$, l'interval és

(l,u), on:

$$l = \sqrt{\frac{(n-1)S^2}{\chi^2_{11,0.995}}} = \sqrt{\frac{11 \cdot 0.149318181}{26.757}} = 0.2477615898 \simeq 0.247762,$$
$$u = \sqrt{\frac{(n-1)S^2}{\chi^2_{11,0.005}}} = \sqrt{\frac{11 \cdot 0.149318181}{2.603}} = 0.7943567758 \simeq 0.794357.$$

(b) No, al contrari. Com que $u = 0.794323 < 0.9$ no hi ha evidències estadístiques per poder afirmar que $\sigma \geq 0.9$.

(c) Com que treballarem amb una confiança del 95%, tenim $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1-\alpha/2 = 0.975$. Tenint en compte que $t_{11,0.975} = 2.201$, obtenim:

$$l = \overline{X} - t_{11,0.975}\frac{S}{\sqrt{n}} = 20.38125 - 2.201\,\frac{0.386417108}{\sqrt{12}} = 12.02948063 \simeq 12.0295,$$
$$u = \overline{X} + t_{11,0.975}\frac{S}{\sqrt{n}} = 20.38125 + 2.201\,\frac{0.386417108}{\sqrt{12}} = 12.52051937 \simeq 12.5205.$$

(d) L'error màxim si suposem $\mu = 12.275$, és

$$t_{11,0.975}\frac{S}{\sqrt{n}} = 2.201\frac{0.386417108}{\sqrt{12}} = 0.2455193726 \simeq 0.2455.$$

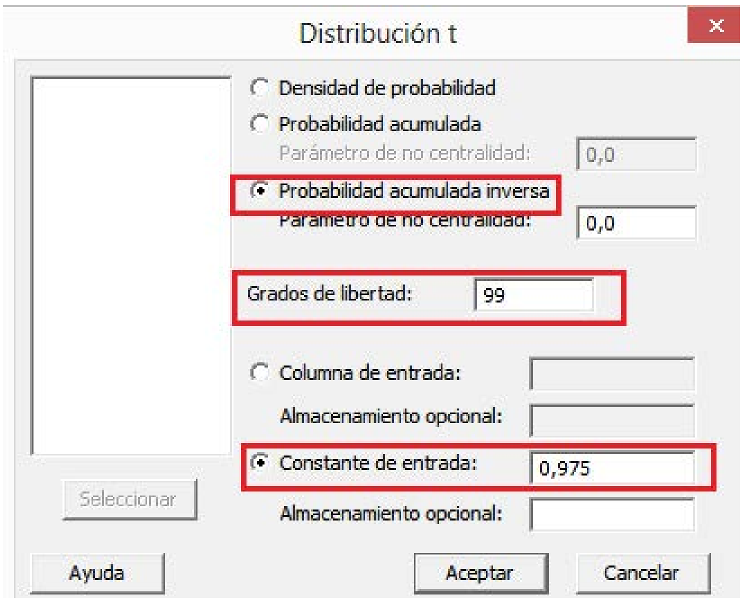
Si suposem $\mu = 12.5$, l'error és $\max(|l-12.5|, |u-12.5|) = 0.4705$.

3. (a) Deixem clar que la probabilitat que $E(X) = 50$ és zero. Ara bé com que $50 \in (48.9754, 50.7446)$, podem dir que no hi ha evidències estadístiques per afirmar el contrari, i per tant, es pot assumir que $E(X) = 50$ amb la confiança indicada.

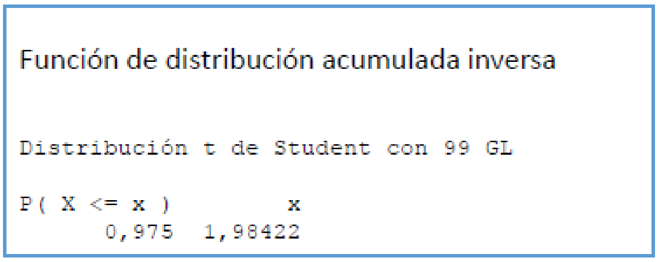
   (b) L'error màxim és $\max(|l-50|, |u-50|) = \max(|48.9754-50|, |50.7446-50|) = 1,0246$.

4. Tant a l'apartat (a) com (b) heu de procedir igual que a l'apartat (b) de l'exercici 1, ja que no ens han donat σ, les dades de l'enunciat corresponen a $\overline{X} = 35759.6$ i $S = 59.8$. Recordeu que heu de treballar, doncs, amb les distribucions T de Student i χ^2 amb 9 graus de llibertat a l'apartat (a) i 99 graus de llibertat a l'apartat (b), per trobar els intervals per μ i σ respectivament.

   Com que no teniu la taula per T_{99} haureu de fer servir T_{100} com a aproximació. És a dir $t_{99,0.975} \simeq t_{100,0.975} = 1.9840$. També podeu calcular el nombre exacte amb Minitab fent: `Calc`→ `Distribuciones de probabilidad`→ `t...`, omplim el menú següent



Obtenint $t_{99,0.975} = 1,98422$.


Función de distribución acumulada inversa

Distribución t de Student con 99 GL

```
P( X <= x )        x
      0,975  1,98422
```

5. (a) Aquí podem fer servir que $\sigma = 0.15$, tenint en compte que $z_{0.975} = 1.96$, l'interval és (l, u), on:

$$l = \overline{X} - z_{0.975}\frac{\sigma}{\sqrt{n}} = 2.25 - 1.96\,\frac{0.15}{\sqrt{36}} \simeq 2.201,$$
$$u = \overline{X} + z_{0.975}\frac{\sigma}{\sqrt{n}} = 2.25 + 1.96\,\frac{0.15}{\sqrt{36}} \simeq 2.299.$$

   (b) Imposem

$$z_{0.975}\frac{\sigma}{\sqrt{n}} = 1.96\,\frac{0.15}{\sqrt{n}} < 0.02.$$

   Aïllant, obtenim $n > 216.09$, per tant:

$$n \geq 217.$$

6. Com a d'altres casos $1 - \alpha/2 = 0.975$. De les dades, tenim $\widehat{p} = 55/100 = 0.55$. L'interval de confiança és:

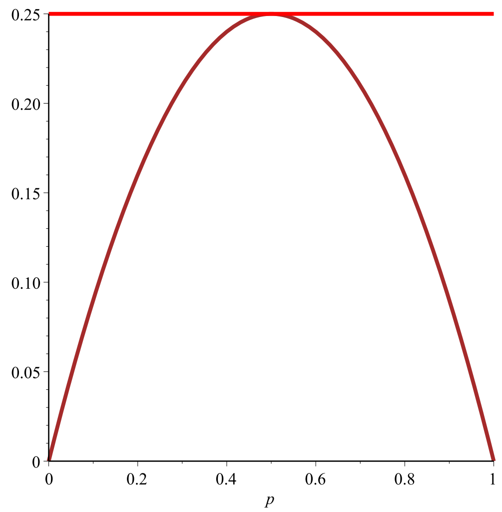
$$l = \widehat{p} - z_{0.975}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.55 - 1.96\sqrt{\frac{0.55\cdot 0.45}{100}} = 0.4524912312,$$
$$u = \widehat{p} + z_{0.975}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.55 + 1.96\sqrt{\frac{0.55\cdot 0.45}{100}} = 0.6475087688.$$

7. (a) Aquí la confiança és del 90%. Per tant $1 - \alpha/2 = 0.95$. Tindrem, doncs en compte que $z_{0.95} \simeq 1.645$. Aquí no coneixem $\widehat{p}$, per tant si imposem:

$$\text{Error } = z_{0.95}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.645\,\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} < 0.01$$

   no podem aïllar n!

   Acotarem, doncs, el número (desconegut) $\widehat{p}(1-\widehat{p})$. És fàcil veure que la paràbola $y(p) = p(1-p)$ té un màxim a $p = 1/2$ que val $y(1/2) = 1/4$.



Per tant, si imposem

$$\text{Error } = z_{0.95}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.645\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq 1.645\sqrt{\frac{0.25}{n}} < 0.01,$$

i aïllem obtenim $n > 6765.06...$, per tant $n \geq 6766$.

(b) Si fem servir la informació de la prova pilot i considerem $\widehat{p} \simeq 0.15$, tenim:

$$\text{Error } = z_{0.95}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.645\sqrt{\frac{0.15 \cdot 0.85}{n}} < 0.01,$$

obtenint $n \geq 3451$.

8. Aquestes afirmacions són una altra manera d'expressar el que s'ha obtingut quan s'han calculat els corresponents intervals de confiança en base al estadístics mostrals $\overline{X} = 22.5468$ i $S = 3.78277$.

9. Imposem

$$\text{Error } = z_{0.975}\frac{\sigma}{\sqrt{n}} = 1.96\frac{3.75}{\sqrt{n}} < 0.2,$$

obtenint:

$$n \geq 1351.$$

10. Les dades de les mostres són:

Mostra 1: $n_1 = 7$, $\overline{X}_1 = 0.14057$, $S_1 = 0.00276$.

Mostra 2: $n_2 = 5$, $\overline{X}_2 = 0.13840$, $S_2 = 0.00321$.

(a) Observem que σ_1 i σ_2 són desconegudes, però ens diuen que presenten la mateixa variabilitat, per tant podem assumir que $\sigma_1 = \sigma_2$, i per tant, farem servir l'interval de confiança

$$\overline{X}_1 - \overline{X}_2 \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

on

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}.$$

Calculem ara cada element de la fórmula: $\overline{X}_1 - \overline{X}_2 = 0.00217$, $t_{10,0.975} = 2.2281$, $S_p^2 = \dfrac{6 \cdot (0.00276)^2 + 4 \cdot (0.00321)^2}{10} = 8.6922 \cdot 10^{-6}$ i per tant $S_p = 0.00294825$. Per tant:

$$\overline{X}_1 - \overline{X}_2 \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} =$$

$$= 0.00217 \pm 2.2281 \cdot 0.00294825 \cdot \sqrt{\frac{1}{7} + \frac{1}{5}} = 0.00217 \pm 0.003846414958,$$

per tant:

$$\text{I.C.}\,(\mu_1 - \mu_2) \simeq (-0.00167, 0.00602)\,.$$

(b) No, donat que l'interval conté el 0, no tenim evidències per afirmar que hi hagi una diferència significativa.

Tot i que en aquesta part de la llista no demanem contrastar hipòtesis, mireu si us plau, la resolució de l'exercici 10 de contrast d'hipòtesis.

(c) És acceptable, donat que $1 \in \text{I.C.}(\sigma_1/\sigma_2)$.

6.2.2 Contrast d'hipòtesis.

1. Recordem que l'error de tipus I és rebutjar H_0 quan és certa, i que l'error de tipus II és "acceptar" H_0 quan és falsa. Per tant:

   (a) Si concloure que no és eficaç és un error de tipus I vol dir que estem acceptant H_1, per tant:

   $$\begin{cases} H_0 : \text{eficaç}, \\ H_1 : \text{no eficaç}. \end{cases}$$

   (b) Ídem.

2. (a)

   $$\begin{cases} H_0 : \text{innocent}, \\ H_1 : \text{culpable}. \end{cases}$$

   (b)

   $$\begin{cases} H_0 : \text{culpable}, \\ H_1 : \text{innocent}. \end{cases}$$

3. Volem contrastar

   $$\begin{cases} H_0 : \mu = 155, \\ H_1 : \mu > 155. \end{cases}$$

   Prenem $\alpha = 0.05$. L'estadístic de contrast és $t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} = \sqrt{n}\,\dfrac{\overline{X} - \mu_0}{S}$, on $\mu_0 = 155$. La regió de rebuig de la hipòtesi nul·la és $t > t_{n-1,1-\alpha}$, ja que és un contrast unilateral.

   Introduint les dades a una vella calculadora obtenim: $n = 80$, $\overline{X} = 162.6625$, $S = 33.77323626$, per tant

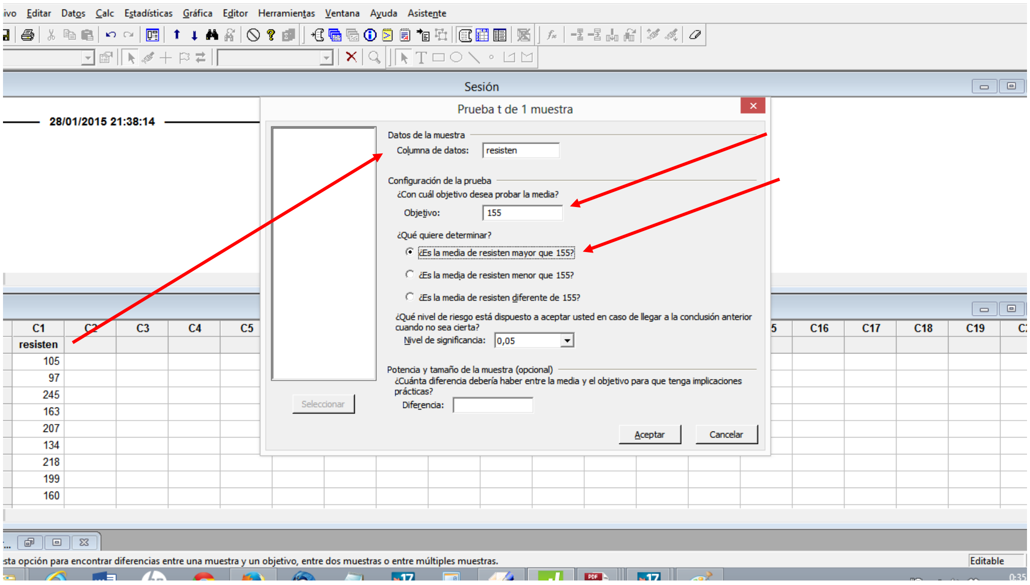
   $$t = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S} = 2.029.$$

El valor $t_{79,0.95}$ no surt a les taules, però $t_{70,0.95} = 1.6669$ i $t_{80,0.95} = 1.6641$. El valor que busquem està entre aquests dos valors. Com que clarament $t = 2.029 > t_{79,0.95}$, tenim evidències estadístiques per a rebutjar H_0, i acceptar que $\mu > 155$.

Potser aquesta sigui una bona ocasió per fer servir Minitab. Primerament podeu fer l'assistent de contrast d'hipòtesis per a obtenir informació. Només cal fer:

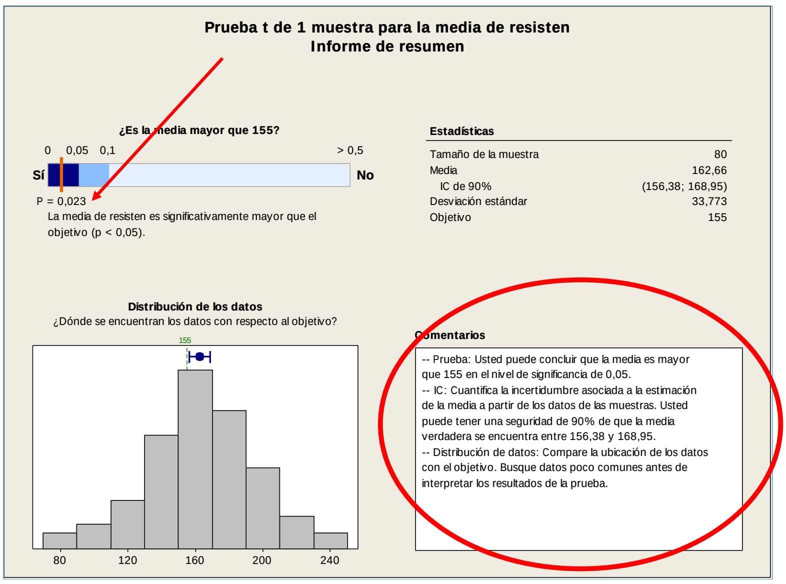
`Asistente`$\rightarrow$ `Pruebas de hipótesis`$\rightarrow$ `Comparar una muestra con un objetivo`$\rightarrow$ t `de 1 muestra`

Aleshores omplirem aquest menú:

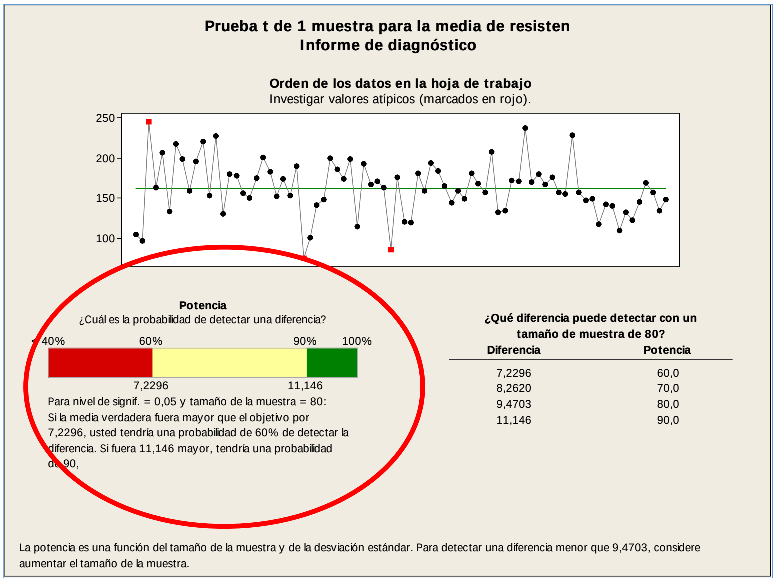


obtenint tres pantalles. A la primera pantalla observem el resultat *"Usted puede concluir que la media es mayor que 155 en el nivel de significancia de 0,05"*. També veiem un resum dels estadístics i el *valor p*:

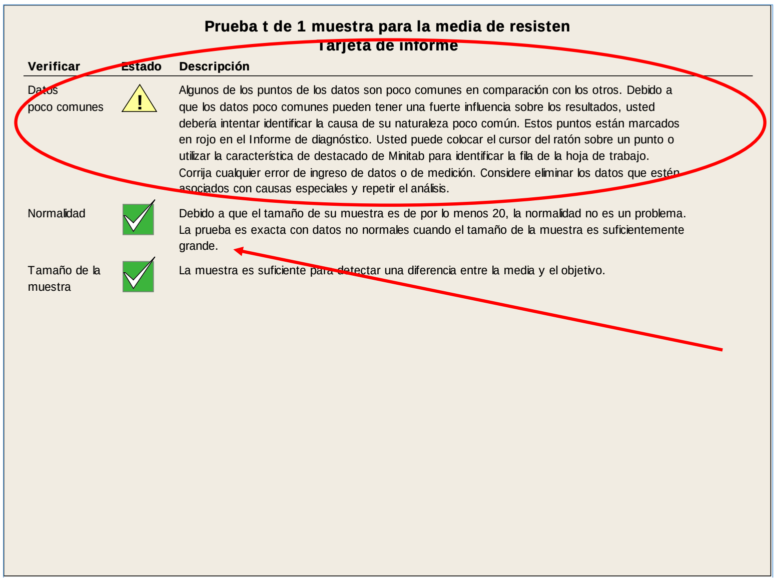
$$p = 0.023 < 0.05 = \alpha.$$



A la segona pantalla es mostren que alguns valors de la mostra són anòmals. Potser l'investigador haurà de valorar d'incloure'ls, o no, a la mostra i tornar, o no, a repetir el contrast. També ens mostra la potència, és a dir, el valor de $1-\beta$, és a dir la probabilitat de trobar que la diferència és significativa, quan realment ho és. En aquest cas és del 60%.



A la tercera pantalla se'ns torna a informar de l'existència de valors anòmals i també que no ens hem de preocupar per si podem assumir o no que les dades són normals.



4. (a) Volem contrastar

$$\begin{cases} H_0 : \mu = 3, \\ H_1 : \mu < 3. \end{cases}$$

(b) Compte! com que s'han modificat elements per tal de reduir el valor de μ, no sembla sensat assumir que σ es manté constant. Per tant, pensarem que és desconeguda i treballarem amb l'estadístic de contrast

$$t = \frac{\overline{X}-3}{S/\sqrt{n}} = \sqrt{n}\,\frac{\overline{X}-3}{S}.$$

(c) Quan $\alpha = 0.05$ i $n = 25$, la zona de rebuig és $t < -t_{n-1,1-\alpha} = -t_{24,0.95} = -1.7109$.

(d) Tenim $n = 10$, $\overline{X} = 2.61$, $S = 0.417372735$, i

$$t = \sqrt{10}\,\frac{2.61 - 3}{S = 0.417372735} = -2.95488464 < -1.8113 = -t_{9,0.95},$$

per tant rebutgem H_0, i considerem $\mu < 3$. Hi ha evidències estadístiques per afirmar que els canvis han estat efectius.

5. (a) L'estadístic de prova (o de contrast) és

$$t = \frac{\overline{X} - 90}{S/\sqrt{n}} = \sqrt{n}\,\frac{\overline{X} - 90}{S}.$$

La regió crítica és $t < t_{4,0.05} = -t_{4,0.95} = -2.1319$.

(b) Recordem la "regla" del valor p:

$$\begin{array}{l} p \leq \alpha \Rightarrow \text{ hi ha evidències per rebutjar } H_0 \text{ (rebutgem } H_0\text{)}, \\ p > \alpha \Rightarrow \text{ no hi ha evidències per rebutjar } H_0 \text{ (``acceptem'' } H_0\text{)}. \end{array}$$

Si $p = 0.005$ hi ha evidències estadístiques per rebutjar H_0 ja que $p = 0.005 < 0.05 = \alpha$ i, per tant, per afirmar que $\mu < 90$. En conseqüència hi ha raons per preocupar-se ja que volíem que el rendiment no fos inferior a 90%. Noteu que convé no confondre $\mu = 90\%$ i el 95% de confiança. Veurem que la decisió serà la mateixa per valors més alts de la confiança. Observeu que sempre que fixem una $\alpha > 0.005$ rebutjarem H_0. En altres paraules, que sempre que treballem amb una confiança inferior a 99,95% rebutjarem H_0. Per tant hi ha raons per preocupar-se, ja que normalment es treballa amb confiances del 90%, 95%, 99%. Per totes aquestes confiances rebutjaríem H_0. Quan treballem amb una confiança superior a 99,95% no rebutjarem H_0, però no perquè aquesta sigui certa, sino per que cal que la diferència entre $\bar{X}$ i 90 sigui excessivament gran per tal que el contrast la consideri significativa.

6. Volem contrastar

$$\begin{cases} H_0 : \mu = 175, \\ H_1 : \mu > 175. \end{cases}$$

(a) L'estadístic de contrast és

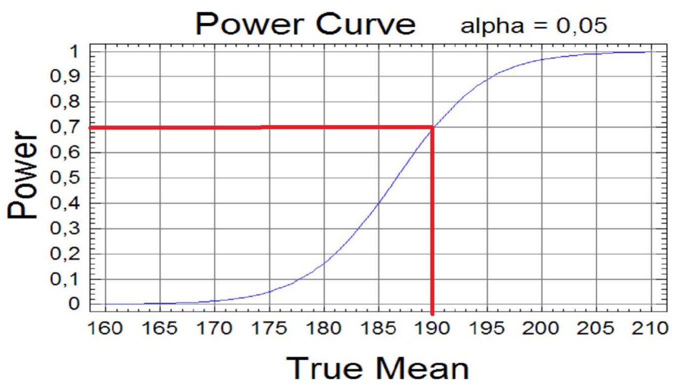
$$t = \frac{\overline{X} - 175}{S/\sqrt{10}}.$$

La regió crítica si $\alpha = 0.05$ i és $t > t_{n-1,1-\alpha} = t_{9,0.95} = 1.8331$.

(b) La corba de potència del test és

$$\Pi(\mu) = 1 - \beta(\mu) = P(\text{rebutjar } H_0 | \text{el valor real de la mitjana és } \mu).$$

Recordem que $\beta = P(\text{escollir } H_0 | H_1 \text{ és certa})$, i que per tant $1-\beta = P(\text{escollir } H_1 | H_1 \text{ és certa})$. Observem que en aquest cas $\Pi(190) = 1 - \beta(190) = 0.7$, per tant $\beta(190) = 0.3$.



(c) L'estadístic de contrast és
$$t = \frac{190 - 175}{20.1/\sqrt{10}} \simeq 2.3599.$$
Com que $t = 2.3599 > t_{n-1,1-\alpha} = t_{9,0.95} = 1.8331$. Rebutgem, doncs, H_0, i considerem $\mu > 175$.

(d) Si es vol $\alpha = 0.01$ aleshores com que $t = 2.3599 < t_{n-1,1-\alpha} = t_{9,0.99} = 2.8214$, amb aquest nivell de confiança la diferència entre 190 i 175 no és prou significativa, i no tenim prou evidències estadístiques com per a rebutjar H_0.

7. (a) Fixem-nos que en aquest cas el paràmetre σ és conegut. L'estadístic de contrast és
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \sqrt{n}\,\frac{\overline{X} - \mu_0}{\sigma},$$
on $\mu_0 = 100$, i la regió d'acceptació de H_0 és $|Z| < 2.25$. Per tant
$$|Z| < 2.25 \Rightarrow -2.25 < \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < 2.25,$$
aïllant $\overline{X}$, obtenim que la regió d'acceptació de H_0 és
$$\mu_0 - 2.25\,\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + 2.25\,\frac{\sigma}{\sqrt{n}},$$
que en el nostre cas, com que $\mu_0 = 100$, $\sigma = 2$ i $n = 9$, dóna, efectivament
$$98.5 < \overline{X} < 101.5.$$

(b) El risc de primera espècie és $\alpha = P(\text{escollir } H_1|H_0 \text{ és certa})$. Observem que com que el contrast és bilateral, la regió d'acceptació de H_0 és $|Z| < z_{1-\alpha/2}$. Per tant
$$z_{1-\alpha/2} = 2.25.$$
Mirant les taules de la distribució normal estàndard tenim doncs que
$$1 - \frac{\alpha}{2} = 0.9878 \text{ per tant } \alpha = 0.244.$$

(c) El risc de segona espècie és $\beta = P(\text{escollir } H_0|H_1 \text{ és certa})$. Per tant, si $\mu = 103$, tenim:
$$\beta = P(\text{escollir } H_0|H_1 \text{ és certa}) = P(98.5 < \overline{X} < 101.5|\mu = 103).$$
Recordeu que $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, per tant tipificant, és a dir: considerant $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$, obtenim:
$$\beta = P\left(\frac{98.5 - 103}{2/3} < \frac{\overline{X} - 103}{2/3} < \frac{101.5 - 103}{2/3}\right) = P(-6.75 < Z < -2.25) =$$
$$= \Phi(-2.25) - \Phi(-6.75) \simeq 0.0122 - 0 = 0.0122.$$

(d) Anàlogament, quan $\mu = 105$ obtenim
$$\beta = P(\text{escollir } H_0|H_1 \text{ és certa}) = P(98.5 < \overline{X} < 101.5|\mu = 105) =$$
$$= P\left(\frac{98.5 - 105}{2/3} < \frac{\overline{X} - 105}{2/3} < \frac{101.5 - 105}{2/3}\right) = P(-9.75 < Z < -5.25) =$$
$$= \Phi(-5.25) - \Phi(-9.75) \simeq 0.$$
Òbviament, quan el valor real de la μ s'allunya del valor que volem contrastar μ_0, la probabilitat d'acceptar erròniament H_0 disminueix.

8. Primerament, notem que $P(\text{rebutjar } H_0|\sigma_1 = \sigma_2) = 0.05$, vol dir que $\alpha = 0.05$.

   (a) El contrast d'igualtat de desviacions típiques és

$$\begin{cases} H_0: & \sigma_1/\sigma_2 & = & 1 \\ H_1: & \sigma_1/\sigma_2 & \neq & 1 \end{cases}$$

   L'estadístic de prova és $F = \dfrac{S_1^2}{S_2^2}$ que quan H_0 és certa segueix una distribució de Fisher $F_{n_1-1,n_2-1} = F_{10,15}$. La regió crítica o de rebuig per al nostre contrast bilateral és $F < f_{10,15,0.025}$ i $F > f_{10,15,0.975}$. De les taules, obtenim:

$$f_{10,15,0.975} = 3.06 \text{ i } f_{10,15,0.025} = \frac{1}{f_{15,10,0.975}} = \frac{1}{3.06} = 0.2841.$$

   Com que

$$F = \frac{9.7^2}{15.9^2} = 0.3722 \in (0.2841, 3.06),$$

   no tenim evidències estadístiques per rebutjar H_0.

   (b) Ara l'estadístic de prova, quan H_0 és certa, segueix una distribució de Fisher $F_{99,99}$. La regió crítica o de rebuig per al nostre contrast bilateral és $F < f_{99,99,0.025}$ i $F > f_{99,99,0.975}$. Com que no tenim a les taules 99 graus de llibertat, farem servir la més propera, 120 graus de llibertat:

$$f_{99,99,0.975} \simeq f_{120,120,0.975} = 1.43 \text{ i } f_{99,99,0.025} = \frac{1}{f_{99,99,0.975}} \simeq \frac{1}{f_{120,120,0.975}} = \frac{1}{1.43} = 0.6993.$$

   Com que

$$F = \frac{10.6^2}{15.9^2} = 0.4444 \notin (0.6993, 1.43),$$

   per tant, amb aquest nivell de confiança, rebutgem H_0.

   Observeu que això no contradiu el resultat de l'apartat (a). A un cert nivell de confiança, la diferència entre les desviacions típiques mostrals no és prou significativa i a un altre nivell si.

   Observeu també que podeu calcular el nombre exacte de $f_{99,99,0.975}$ amb Minitab fent: `Calc`→ `Distribuciones de probabilidad`→ `F...`, obtenint

$$f_{99,99,0.975} = 1,48623 \text{ i } f_{99,99,0.975} = 0.672842.$$

   Recordeu d'activar l'opció "`Probabilidad acumulada inversa`". En qualsevol cas, amb els valors exactes la decisió és la mateixa.

9. (a) El contrast d'igualtat de variàncies és

$$\begin{cases} H_0: & \sigma_1^2/\sigma_2^2 & = & 1 \\ H_1: & \sigma_1^2/\sigma_2^2 & \neq & 1 \end{cases}$$

   L'estadístic de prova és $F = \dfrac{S_1^2}{S_2^2}$ que quan H_0 és certa segueix una distribució de Fisher $F_{n_1-1,n_2-1} = F_{29,29}$. La regió crítica o de rebuig per al nostre contrast bilateral és $F < f_{29,29,\alpha/2}$ i $F > f_{29,29,1-\alpha/2}$. Per a $\alpha = 0.05$ tenim, directament de les taules,

$$f_{29,29,0.975} = 2.1 \text{ i } f_{29,29,0.025} = \frac{1}{2.1} = 0.4761904762.$$

(b) Com que $F = 2.69 > f_{29,29,0.975} = 2.1$ es rebutja H_0 amb risc α, és a dir, es pot afirmar que les variàncies de les dues poblacions analitzades són diferents.

(c) A la vista de les mitjanes de les dues mostres, el contrast per a la igualtat de mitjanes més adient és:

$$\begin{cases} H_0: & \mu_1 - \mu_2 & = & 0 \\ H_1: & \mu_1 - \mu_2 & > & 0 \end{cases}$$

amb $\sigma_1 \neq \sigma_2$ i desconegudes. L'estadístic de contrast és

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

que si H_0 és certa segueix una distribució aproximada T-Student amb k graus de llibertat on k és l'enter més proper a

$$k = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$$

La regió crítica és $t > t_{k,0.95}$.

(d) Que l'ordinador digui $p = 0.000$ vol dir $p < 10^{-3}$. Com que $\alpha = 0.05 > p$ hem de rebutjar la hipòtesi nul·la del contrast de l'apartat anterior, ja que en la pràctica això implica que l'estadístic de contrast està a la regió crítica $t > t_{k,0.95}$.

10. Com que només necessitem el valor p per decidir, per demanar a Minitab que faci el contrast

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0, \\ H_1: \mu_1 - \mu_2 \neq 0. \end{cases}$$

només cal fer: `Estadísticas`$\rightarrow$ `Estadística básica`$\rightarrow$ $2t$. `t de 2 muestras`, obtenint:

Resultados para: Hoja de trabajo 3

Prueba T e IC de dos muestras: pi10_1; pi10_2

```
T de dos muestras para pi10_1 vs. pi10_2

                                 Error
                              estándar
                                 de la
         N    Media  Desv.Est.   media
pi10_1   7  0,14057    0,00276  0,0010
pi10_2   5  0,13840    0,00321  0,0014


Diferencia = mu (pi10_1) - mu (pi10_2)
Estimado de la diferencia:  0,00217
IC de 95% para la diferencia:  (-0,00167; 0,00602)
Prueba T de diferencia = 0 (vs. no =): Valor T = 1,26  Valor P = 0,237  GL = 10
Ambos utilizan Desv.Est. agrupada = 0,0029
```

Com que el *valor p* és $p = 0.237 > 0.05 = \alpha$, això vol dir que l'estadístic de contrast està en la regió d'acceptació de H_0, és a dir que no tenim evidències per a rebutjar que $\mu_1 = \mu_2$, és a dir que la diferència no és significativa. Noteu que aquests resultats s'han obtingut sense assumir que $\sigma_1 = \sigma_2$, el que és més realista.

Evidentment, si anem a l'assistent obtindrem molta més informació (vegeu, per exemple, l'exercici 3). Heu de fer:

`Asistente`$\rightarrow$ `Pruebas de hipótesis`$\rightarrow$ `Comparar dos muestras entre sí`$\rightarrow$ `t de 2 muestras`.