

Exercicis resolts per a l'assignatura d'Estadística II



Teresa Corbella i Domènech
Coia Domingo i Vernis

PUBLICACIONS DE LA UNIVERSITAT ROVIRA I VIRGILI

Av. Catalunya, 35 · 43002 Tarragona

Tel. 977 558 474 · publicacions@urv.cat

www.publicacions.urv.cat



1a edició: octubre de 2025

ISBN (PDF): 978-84-1365-200-9

DOI: 10.17345/9788413652009



Cita el llibre.



Consulta el llibre a la nostra web.



Llibre sota una llicència Creative Commons BY-NC-SA.

Publicacions de la Universitat Rovira i Virgili és membre de la Unión de Editoriales Universitarias Españolas i de la Xarxa Vives, fet que garanteix la difusió i comercialització de les seves publicacions a nivell nacional i internacional.

Exercicis resolts
per a l'assignatura d'Estadística II

Teresa Corbella i Domènech
Coia Domingo i Vernis



Tarragona, 2025



Dedicat al professor Juan Manuel López-Rey Laurens (Tarragona, 27 de novembre de 1953-13 d'abril de 2018), company, amic i mestre, que va ser professor d'Estadística durant molts anys i fins al final de la seva vida a la Facultat d'Economia i Empresa de la Universitat Rovira i Virgili.

Aquest llibre ha volgut recuperar alguns dels seus enunciats, que s'han conservat com a imatge d'un text antic, picats amb una màquina d'escriure mecànica i en la llengua original, i que s'han traduït al català i resolt per a aquest llibre. Qualsevol possible error és responsabilitat de les autores.

Continguts

EL PROGRAMA.....	9
Càlcul de probabilitats en una distribució uniforme contínua	
Càlcul de l'esperança en una distribució uniforme contínua amb canvi d'origen	
RIFA PER A LA BIBLIOTECA.....	15
Càlcul de probabilitats d'una loteria	
Construcció d'una variable aleatòria	
Càlcul de l'esperança i de la variància d'una variable aleatòria discreta	
Propietat de l'esperança i la variància: canvi d'origen	
DEMANDA DIÀRIA D'UN PRODUCTE (<i>exercici del professor J.M. López-Rey Laurens</i>)..	23
Càlcul de probabilitats a partir d'una funció de densitat	
Càlcul de l'esperança matemàtica per a una variable aleatòria contínua	
Càlcul de l'esperança matemàtica per a una variable aleatòria discreta	
DETERGENT PER A RENTADORES (<i>exercici del professor J.M. López-Rey Laurens</i>) ...	29
Càlcul de la constant de normalització d'una funció de densitat	
Càlcul de probabilitats a partir d'una funció de densitat	
Càlcul de la variància d'una variable aleatòria contínua	
TORRE DE BABEL	35
Distribució binomial	
Propietat additiva de la distribució binomial	
Teorema de Laplace - De Moivre	
Aproximació de la distribució binomial en la distribució normal amb correcció per continuïtat	
ABRAÇADES.....	45
Distribució de Poisson	
Distribució exponencial	
Càlcul d'una mediana	
AVETS A LA FIRA DE SANTA LLÚCIA.....	53
Distribució normal	
Propietat additiva de la distribució normal	
Construcció i transformació de variables aleatòries	

MONA DE PASQUA	61
Propietat additiva de la distribució normal	
Distribució de Poisson	
Distribució exponencial	
Falta de memòria de la distribució exponencial (propietat)	
PERES PER SOPAR.....	67
Distribució binomial	
Propietat additiva de la distribució normal	
JOC DE CURSES DE CAVALLS	75
Teorema central del límit	
Distribució binomial	
Distribució uniforme discreta (el cas del dau)	
VIDEOJOCs (<i>exercici del professor J.M. López-Rey Laurens</i>)	83
Distribució uniforme	
Teorema central del límit	
Suma de normals	
REVETLLA DE SANT JOAN.....	89
Propietats de l'esperança i la variància	
Teorema central del límit	
Desigualtat de Txevixev	
A CAL PERRUQUER.....	95
Interpretació d'un interval de confiança d'una mitjana poblacional (mostra gran)	
Càlcul d'un interval de confiança d'una mitjana poblacional (mostra gran)	
Contrast d'hipòtesis unilateral d'una mitjana poblacional (mostra gran)	
Estimacions puntuals de la mitjana i de la variància poblacionals	
MATRÀS (<i>exercici del professor J.M. López-Rey Laurens</i>)	103
Càlcul d'un interval de confiança d'una mitjana poblacional (mostra petita)	
Contrast d'hipòtesis unilateral d'una mitjana poblacional (mostra petita)	

MISSATGES DE TELÈFON.....	109
Càlcul d'interval de confiança de la diferència de mitjanes poblacionals (mostres grans)	
Contrast d'hipòtesis sobre la diferència de mitjanes poblacionals (mostres grans)	
Tipus d'error en un contrast d'hipòtesis	
Efecte del canvi del nivell de confiança	
BARRA DE SUSHI	119
Contrast d'hipòtesis bilateral de la diferència de mitjanes poblacionals (mostres petites)	
Càlcul de l'interval de confiança de la diferència de mitjanes poblacionals (mostres petites)	
Càlcul d'interval de confiança d'una mitjana poblacional (mostra petita)	
Càlcul d'un valor esperat	
XOCOLATA DESFETA	129
Contrast d'hipòtesis unilateral d'una proporció poblacional	
Càlcul d'un interval de confiança d'una proporció poblacional	
Efecte de la grandària mostral en l'amplitud dels intervals de confiança	
CANVIS SOCIALS	137
Contrast d'hipòtesis bilateral de la diferència de dues proporcions poblacionals	
Error de tipus II en un contrast d'hipòtesis	
CERVESES SENSE ALCOHOL (<i>exercici del professor J.M. López-Rey Laurens</i>)	143
Prova d'independència de χ^2 entre dues variables categòriques	
REFERÈNCIES A ALTRES LLIBRES.....	147
ÍNDIX PER TEMES	149

El programa

```
Function Check
    //check "who am I";
    // Check "what am I";
End psycho program.
Function My name is
    name =: HAL;
# Include HAL= HAL.
If name=HAL
    compute a=b+1;
else
    through the astronaut
end if.
//print all.
```

Amb l'ajut d'un programa informàtic s'ha generat una variable aleatòria (X) que segueix una distribució uniforme contínua entre 50 i 60.

Sigui $Y = X + 2$. Trobeu:

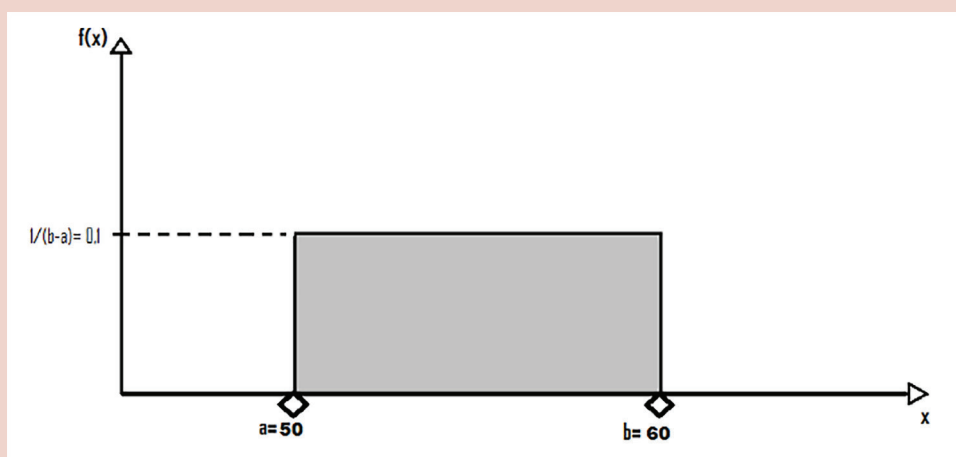
- a) La probabilitat que X prengui un valor estrictament inferior a 65.
- b) La probabilitat que X prengui un valor estrictament superior a 58.
- c) La probabilitat que X prengui un valor estrictament superior a 65.
- d) La probabilitat que X prengui un valor entre 51 i 57, no inclosos.
- e) La probabilitat que X prengui un valor entre 45 i 55, no inclosos.
- f) L'esperança de Y .

Uniforme contínua

$X \sim U(a; b)$, on a és el valor mínim que pren la variable —en aquest exercici, 50— i b el valor màxim —en aquest exercici, 60. Aquesta variable aleatòria queda representada per la funció de densitat $f(x)$:

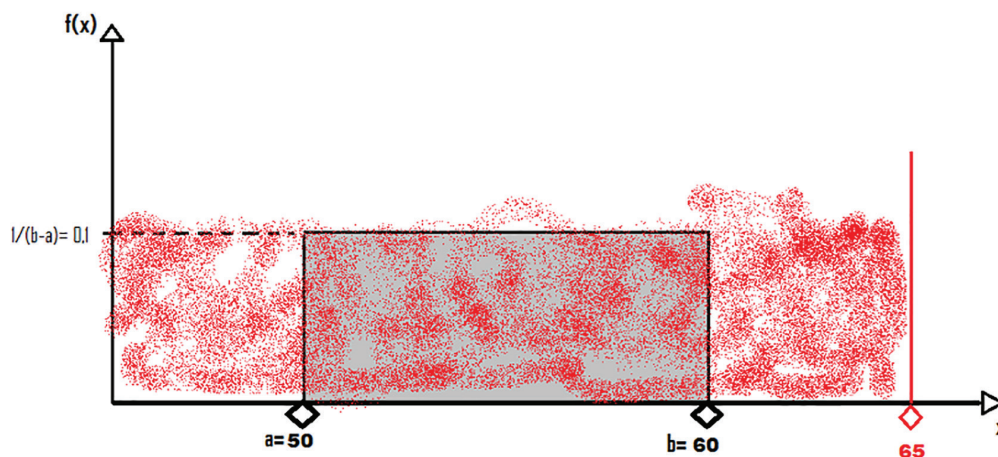
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{altrament} \end{cases}$$

L'àrea per sota de la funció de densitat entre dos punts representa la probabilitat que X prengui un valor en aquell interval. L'àrea per sota de la corba entre 50 i 60 (rectangle gris) és d'1 ($0,1 \cdot 10 = 1$).



Resolució de la pregunta a

Es demana el resultat de $P(X < 65)$.



L'àrea vermella representa l'àrea que compleix $P(X < 65)$, que inclou tot el rectangle gris que representa la variable X . Per tant, sempre que l'ordinador generi un valor de X , aquest valor sempre serà inferior a 65. Es tracta d'un esdeveniment segur i la probabilitat d'un esdeveniment segur és 1; per tant, $P(X < 65) = 1$.

La probabilitat que X prengui un valor inferior a 65 és igual a 1.

Resolució de la pregunta b

Es demana el resultat de $P(X > 58)$.

Opció a

Es pot obtenir directament el resultat mitjançant el complementari de la funció de distribució acumulada:

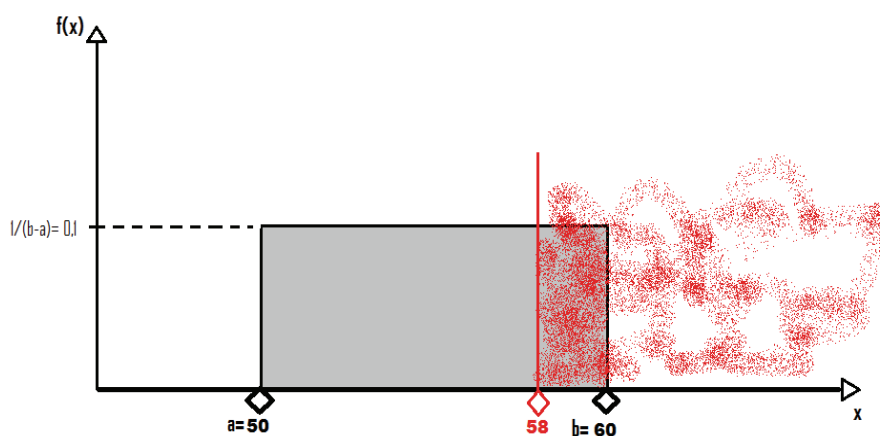
$$P(X > 58) = 1 - P(X \leq 58) = 1 - F(58) = 1 - \frac{58-50}{60-50} = 1 - 0,8 = 0,2$$

La funció de distribució acumulada d'una variable aleatòria uniforme contínua X és:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases}$$

Opció b

La probabilitat demanada es correspon amb l'àrea que queda a sota de la funció de densitat entre 58 i 60; és a dir, en el rectangle gris i vermell. Aquest rectangle té una alçada de 0,1 i una base de 2 ($60 - 58$); per tant, la seva àrea és $2 \cdot 0,1 = 0,2$.



La probabilitat que X prengui un valor superior a 58 és igual a 0,2.

Resolució de la pregunta c

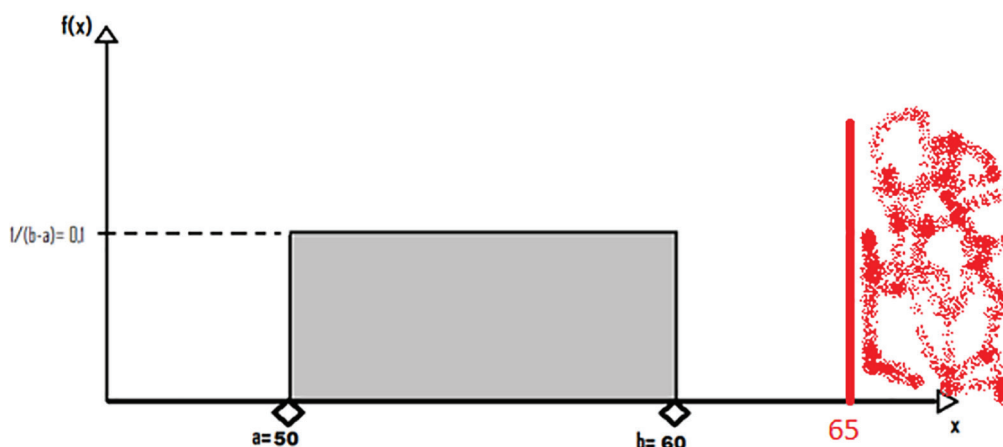
Es demana el resultat de $P(X > 65)$, atès que la variable és contínua $P(X > 65) = P(X \geq 65)$.

Opció a

A l'apartat a s'ha obtingut $P(X < 65) = 1$. El complementari de $P(X < 65)$ és $P(X \geq 65)$; per tant, $P(X \geq 65) = 1 - P(X < 65) = 1 - 1 = 0$.

Opció b

La probabilitat que es busca es correspon amb la intersecció entre l'àrea grisa i la vermella. La intersecció és buida, atès que les àrees, com es pot veure a la gràfica, estan disjunts i, per tant, $P(X > 65) = 0$.



En aquest cas, l'àrea vermella no té elements en comú amb el rectangle gris. Si l'ordinador genera un valor entre 50 i 60, aquest valor no serà mai superior a 65. Obtenir un element superior a 65 és un esdeveniment impossible, la probabilitat és 0.

La probabilitat que X prengui un valor superior a 65 és igual a 0.

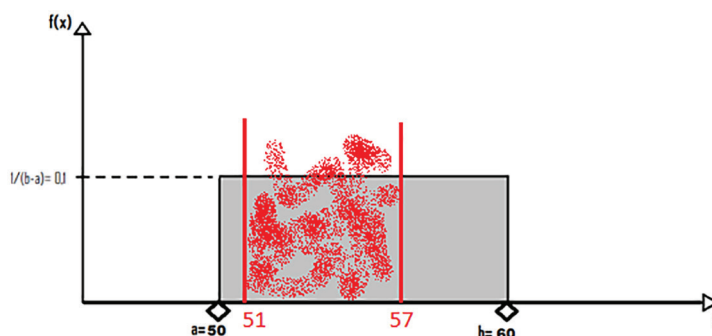
Resolució de la pregunta d

Es demana $P(51 < X < 57)$.

Opció a

La probabilitat d'un interval es pot trobar com la diferència entre els valors de les funcions acumulades en aquells punts:

$$P(51 < X < 57) = F(57) - F(51) = \frac{57-50}{60-50} - \frac{51-50}{60-50} = \frac{57-51}{60-50} = 0,6$$

Opció b

Una altra manera d'arribar a la solució: aquesta probabilitat es correspon amb l'àrea que queda a sota de la funció de densitat entre 51 i 57; és a dir, amb el rectangle gris i vermell. Aquest rectangle té una alçària de 0,1 i una base de 6 ($57 - 51$); per tant, la seva àrea és $6 \cdot 0,1 = 0,6$.

La probabilitat que X prengui un valor entre 51 i 57 és igual a 0,6.

Resolució de la pregunta e

Es demana $P(45 < X < 55)$.

Opció a

Com abans, la probabilitat d'un interval es pot trobar com la diferència entre els valors de les funcions acumulades en aquells punts. Tanmateix, en aquest cas cal fer-ho per parts, atès que la funció de distribució acumulada és diferent per sota i per sobre de 50:

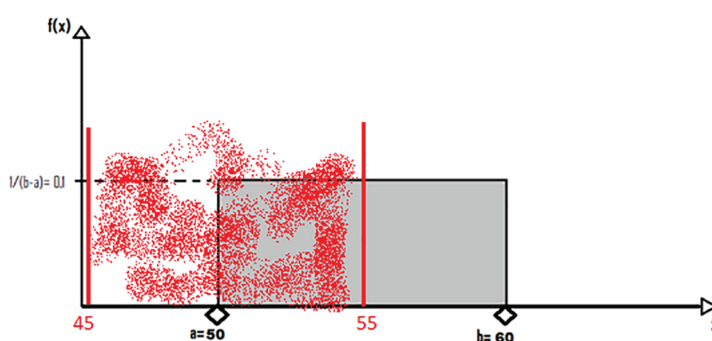
$$P(45 < X < 55) = P(45 < X < 50) + P(50 < X < 55)$$

$$P(45 < X < 50) = F(50) - F(45) = 0 - 0 = 0$$

$$P(50 < X < 55) = F(55) - F(50) = \frac{55-50}{60-50} - 0 = 0,5$$

Finalment:

$$P(45 < X < 55) = P(45 < X < 50) + P(50 < X < 55) = 0 + 0,5 = 0,5$$

Opció b

Com a la pregunta b, l'àrea buscada és la grisa i vermella (la intersecció entre la zona vermella i la grisa). Aquesta àrea fa 0,1 d'alçària i 5 d'amplada; per tant, el resultat és $0,1 \cdot 5 = 0,5$.

La probabilitat que X prengui un valor entre 45 i 55 és igual a la probabilitat que prengui un valor inferior a 55, que és igual a 0,5.

Resolució de la pregunta f

Es demana $E[Y]$.

Opció a

És pot calcular l'esperança matemàtica de Y a partir de la l'esperança matemàtica de X aplicant-hi les propietats de l'esperança.

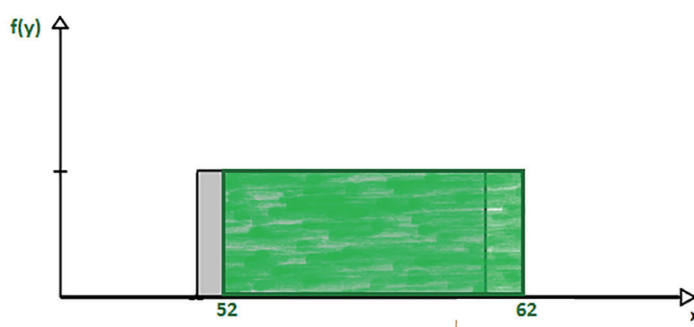
$$E[Y] = E[X + 2] = E[X] + 2 = \frac{50+60}{2} + 2 = 57$$

L'esperança matemàtica d'una variable aleatòria uniforme contínua X és $E[X] = \frac{a+b}{2}$.

Opció b

Atès que $Y = X + 2$ i X segueix una distribució uniforme contínua entre 50 i 60, Y segueix una distribució uniforme contínua entre 52 i 62.

$$E[Y] = \frac{52+62}{2} = \frac{114}{2} = 57$$



L'esperança de Y és igual a 57.

Rifa per a la biblioteca



Sala de lectura de la Biblioteca Riccardiana (Florència),
de Sailko, sota llicència CCBY 3.0.

Al poble s'organitza una rifa per recollir diners en benefici de la biblioteca local. Se sorteja una botifarra blanca artesana valorada en 7 euros, gentilesa de la Sra. Maria, la cansaladera; una garrafa de 2 litres d'oli d'oliva verge valorada en 11 euros i una ampolla de vi valorada en 15 euros (productes oferts pel sindicat del poble). Es posen a la venda 120 tiquets a 1 euro cada un. El sorteig té lloc el diumenge següent després de missa: s'extreuen 3 boles seqüencialment (sense reposició) d'una urna amb 120 boles; la primera bola correspon a la botifarra, la segona a l'oli i la tercera al vi.

En Joan ha comprat 3 tiquets. Trobeu:

- a) La probabilitat que en Joan guanyi la botifarra.
- b) La probabilitat que en Joan guanyi la botifarra, l'oli i el vi.
- c) La probabilitat que en Joan guanyi la botifarra i només acabi guanyant la botifarra.
- d) L'esperança i la variància dels guanys monetaris d'en Joan abans del sorteig.
- e) L'Elisa també té 3 tiquets, però en el seu cas els hi han regalat. Com és l'esperança i la variància dels guanys monetaris de l'Elisa en comparació amb les d'en Joan?

Resolució de la pregunta a

En la primera extracció hi ha 120 tiquets a l'urna i tots tenen la mateixa probabilitat que els triïn. Per aquest motiu, la probabilitat es pot calcular dividint el nombre d'elements favorables (numerador) pel nombre d'elements possibles (denominador):

$$\frac{\text{nombre de casos favorables (nombre d'elements que generen èxit)}}{\text{nombre de casos possibles (nombre d'elements d'\Omega)}}$$

En Joan té 3 tiquets que li són favorables dels 120 possibles:

$$\frac{3}{120} = 0,025$$

La probabilitat que en Joan guanyi la botifarra és igual a 0,025.

Resolució de la pregunta b

En Joan ha guanyat la botifarra. En la segona extracció queden 119 tiquets a l'urna. D'aquests 119 tiquets, 2 són favorables a en Joan. La probabilitat que en Joan guanyi l'oli havent guanyat la botifarra és:

$$\frac{2}{119} = 0,0168$$

De la mateixa manera, la probabilitat que en Joan guanyi l'ampolla de vi en la tercera ronda havent guanyat els altres dos premis és:

$$\frac{1}{118} = 0,0085$$

La probabilitat de la seqüència dels 3 esdeveniments (en Joan ho guanya tot) és el producte de les probabilitats condicionades:

$$\frac{3}{120} \cdot \frac{2}{119} \cdot \frac{1}{118} = 0,000\,003\,56$$

La probabilitat que en Joan guanyi tots tres productes és igual a 0,000 003 56.

Resolució de la pregunta c

En la primera extracció en Joan guanya la botifarra amb una probabilitat igual a 0,025.

En la segona extracció queden 119 tiquets a l'urna, dels quals 117 li són desfavorables. La probabilitat que en Joan no guanyi l'oli havent guanyat la botifarra és:

$$\frac{117}{119} = 0,9832$$

De la mateixa manera, en la tercera extracció queden 118 butlletes a l'urna, de les quals 116 li són desfavorables. La probabilitat que tampoc no guanyi el vi és:

$$\frac{116}{118} = 0,9831$$

Finalment, la probabilitat d'aquesta seqüència d'esdeveniments és el producte de les probabilitats condicionades:

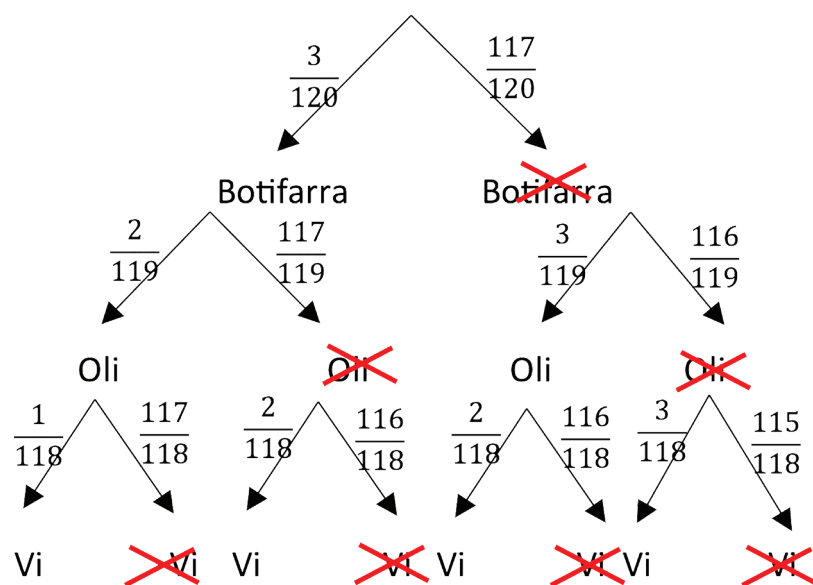
$$\frac{3}{120} \cdot \frac{117}{119} \cdot \frac{116}{118} = 0,0242$$

La probabilitat que en Joan guanyi la botifarra i només la botifarra és igual a 0,0242.

Resolució de la pregunta d

L'arbre següent explicita les 8 seqüències d'esdeveniments. En cada pas, sobre la fletxa s'indica la probabilitat de l'esdeveniment individual en aquell moment de la seqüència (tenint en compte el que ha passat fins aleshores). La probabilitat de cada seqüència d'esdeveniments es calcula com el producte de les probabilitats que hi ha a sobre de cada fletxa de la seqüència.

Així, per exemple, la probabilitat de guanyar els tres productes (pregunta b), el camí que hi ha més a l'esquerra, és el producte de les tres probabilitats que hi ha a sobre de les fletxes ($\frac{3}{120} \cdot \frac{2}{119} \cdot \frac{1}{118}$). La probabilitat de guanyar la botifarra, guanyar l'oli i no guanyar el vi és: $\frac{3}{120} \cdot \frac{2}{119} \cdot \frac{117}{118}$, i així successivament.



D'altra banda, els guanys monetaris d'aquests esdeveniments són la suma del valor monetari de cada producte que s'ha guanyat menys els cost del bitllets adquirits. Així, per exemple, el benefici monetari de guanyar els 3 productes és $7 + 11 + 15 - 3 = 30$ €. Mentre que el de guanyar només la botifarra i l'oli és igual a $7 + 11 - 3 = 15$ €.

La taula següent recull els guanys monetaris de cada seqüència d'esdeveniments, que anomenem X , així com la probabilitat d'ocurrència d'aquests esdeveniments ($P(X = x)$).

Premis	X	$P(X = x)$
Tots	30	0,00000356
Botifarra i oli	15	0,00041661
Botifarra i vi	19	0,00041661
Botifarra	4	0,02416322
Oli i vi	23	0,00041661
Oli	8	0,02416322
Vi	12	0,02416322
Cap	-3	0,92625694

L'esperança dels guanys $E[X]$ és la suma dels productes entre el valor monetari i les probabilitats d'ocurrència de cada seqüència d'esdeveniments. En el cas d'una variable aleatòria discreta: $E[X] = \sum_{x \in R_X} x \cdot P(X = x)$

$$\begin{aligned}
E[X] &= 30 \cdot 0,00000356 \\
&+ 15 \cdot 0,00041661 \\
&+ 19 \cdot 0,00041661 \\
&+ 4 \cdot 0,02416322 \\
&+ 23 \cdot 0,00041661 \\
&+ 8 \cdot 0,02416322 \\
&+ 12 \cdot 0,02416322 \\
&+ (-3) \cdot 0,92625694 \\
&= -2,175
\end{aligned}$$

L'esperança matemàtica és negativa i igual a $-2,175$ €.

La variància és: $Var(X) = E[X^2] - E[X]^2$, on $E[X^2]$ és l'esperança de X^2 i $E[X]^2$ és l'esperança de X , calculada anteriorment, al quadrat.

En el cas d'una variable aleatòria discreta: $E[X^2] = \sum_{x \in R_X} x^2 \cdot P(X = x)$

Per calcular la variància primer es calcula l'esperança dels guanys al quadrat:

$$\begin{aligned}
E[X^2] &= 30^2 \cdot 0,00000356 \\
&+ 15^2 \cdot 0,00041661 \\
&+ 19^2 \cdot 0,00041661 \\
&+ 4^2 \cdot 0,02416322 \\
&+ 23^2 \cdot 0,00041661 \\
&+ 8^2 \cdot 0,02416322 \\
&+ 12^2 \cdot 0,02416322 \\
&+ (-3)^2 \cdot 0,92625694 \\
&= 14,2166
\end{aligned}$$

A aquest resultat, s'hi resta l'esperança al quadrat:

$$Var[X] = 14,2166 - (-2,175)^2 = 9,486$$

La variància dels guanys és igual a $9,486$ €².

Resolució de la pregunta e

Sigui Y els guanys monetaris de l'Elisa. L'única diferència entre els guanys de l'Elisa i els d'en Joan són els 3 € corresponents a la despesa pels bitllets que ha pagat en Joan. Independentment dels premis que guanyi, l'Elisa sempre té 3 euros més de guanys:

$$Y = X + 3$$

A priori, les probabilitats que l'Elisa guanyi premis són idèntiques a les d'en Joan, atès que tenen el mateix nombre de bitllets. La taula de probabilitats dels guanys monetaris de l'Elisa és:

Premis	Y	$P(Y = y)$
Tots	33	0,00000356
Botifarra i oli	18	0,00041661
Botifarra i vi	22	0,00041661
Botifarra	7	0,02416322
Oli i vi	26	0,00041661
Oli	11	0,02416322
Vi	15	0,02416322
Cap	0	0,92625694

A partir d'aquesta taula es pot calcular l'esperança i la variància de Y de la mateixa manera que s'ha fet per a X . Tanmateix, no cal, atès que Y és igual a X més una constant (canvi d'origen). Tenint en compte les propietats de l'esperança i la variància es pot deduir l'esperança i la variància de Y a partir de les de X .

Un canvi d'origen en una variable aleatòria X (sumar o restar una constant k a la variable) modifica l'esperança de la nova variable aleatòria (Y) en el valor de la constant. En canvi, la variància de la nova variable no queda afectada pel canvi d'origen.

$$\begin{aligned}\text{Sigui } Y &= X + 3 \\ E[Y] &= E[X] + 3 \\ \text{Var}[Y] &= \text{Var}[X]\end{aligned}$$

L'esperança de Y és igual a l'esperança de X més 3.

$$E[Y] = E[X] + 3 = -2,175 + 3 = 0,825$$

L'esperança matemàtica dels guanys monetaris de l'Elisa és igual a 0,825 €.

La variància de Y és igual a la variància de X .

La variància dels guanys de l'Elisa és igual a 9,486 €².

Comentaris

En aquest exercici es construeix una variable aleatòria pas a pas (els guanys d'una rifa en què s'han adquirit els bitllets), a fi de poder calcular l'esperança i la variància. A partir d'aquesta primera variable aleatòria es construeix una segona variable aleatòria fent un canvi d'origen i es calcula l'esperança i la variància de la nova variable aplicant-hi propietats, sense que calgui refer tots els càlculs.

Demanda diària d'un producte

En un establecimiento comercial, la demanda diaria (en kilos) de un determinado artículo puede considerarse como una variable aleatoria con la siguiente función densidad:

$$f_d(x) = \begin{cases} 1/8 & \text{para } 0 < x \leq 4 \\ \frac{12-x}{64} & \text{para } 4 < x \leq 12 \\ 0 & \text{en los demás casos} \end{cases}$$

Los beneficios diarios por la venta de tal producto dependen de la demanda según la siguiente distribución:

- 5 unidades monetarias (u. m.) si la demanda es inferior a 2 quilograms.
- 5 u. m. si la demanda está comprendida entre 2 y 4 quilograms.
- 10 u. m. si la demanda está comprendida entre 4 y 8 quilograms.
- 15 u. m. si la demanda está comprendida entre 8 y 12 quilograms.

Calcular:

- a) La probabilidad de que en un día cualquiera la demanda sea superior a 10 quilograms.
- b) La probabilidad de que la demanda sea inferior a 10 quilograms.
- c) La demanda diaria esperada.
- d) El beneficio diario esperado.

En un establiment comercial, la demanda diària (en quilos) d'un determinat article pot considerar-se com una variable aleatòria amb la funció de densitat següent:

$$f_d(x) = \begin{cases} 1/8 & \text{per a } 0 < x \leq 4 \\ \frac{12-x}{64} & \text{per a } 4 < x \leq 12 \\ 0 & \text{en la resta de casos} \end{cases}$$

Els beneficis diaris per la venda del producte depenen de la demanda segons la distribució següent:

−5 unitats monetàries (u. m.) si la demanda és inferior a 2 quilograms.

5 u. m. si la demanda és entre 2 i 4 quilograms.

10 u. m. si la demanda és entre 4 i 8 quilograms.

15 u. m. si la demanda és entre 8 i 12 quilograms.

Calculeu:

- La probabilitat que en un dia qualsevol la demanda sigui superior a 10 quilograms.
- La probabilitat que la demanda sigui inferior a 10 quilograms.
- La demanda diària esperada.
- El benefici diari esperat.

Resolució de la pregunta a

La probabilitat que un dia qualsevol la demanda sigui superior a 10 quilograms es calcula integrant la funció de densitat entre els valors demanats que pertanyin al rang; en aquest cas, entre 10 i 12, atès que el màxim del rang és 12.

$$\begin{aligned} P(X > 10) &= \int_{10}^{+\infty} f_d(x) \, dx \\ &= \int_{10}^{12} \frac{12-x}{64} \, dx \\ &= \frac{1}{64} \left[12x - \frac{x^2}{2} \right]_{10}^{12} \\ &= \frac{1}{64} \left[\left(12 \cdot 12 - \frac{12^2}{2} \right) - \left(12 \cdot 10 - \frac{10^2}{2} \right) \right] = \frac{2}{64} = 0,03125 \end{aligned}$$

La probabilitat que en un dia qualsevol la demanda sigui superior a 10 quilograms és igual a 0,03125.

Resolució de la pregunta b

De la mateixa manera, la probabilitat que la demanda sigui inferior a 10 quilograms es calcula integrant la funció de densitat entre els valors demanats que pertanyin al rang; en aquest cas, entre 0 i 10, atès que el mínim del rang és 0 (opció a). Altrament, també es pot calcular a partir del resultat de l'apartat a (opció b).

Opció a

Per a una variable aleatòria contínua
 $P(X \leq 10) = P(X < 10)$, atès que $P(X=10)=0$.

$$\begin{aligned}
 P(X \leq 10) &= \int_{-\infty}^{10} f_d(x) dx \\
 &= \int_0^4 \frac{1}{8} dx + \int_4^{10} \frac{12-x}{64} dx \\
 &= \left[\frac{1}{8}x \right]_0^4 + \frac{1}{64} \left[12x - \frac{x^2}{2} \right]_4^{10} \\
 &= \left[\frac{4}{8} - 0 \right] + \frac{1}{64} \left[\left(12 \cdot 10 - \frac{10^2}{2} \right) - \left(12 \cdot 4 - \frac{4^2}{2} \right) \right] = \frac{1}{2} + \frac{30}{64} = 0,96875
 \end{aligned}$$

Opció b

La demanda diària sempre se situa entre 0 quilograms i 12 quilograms. Per tant, la probabilitat que la demanda estigui entre 0 quilograms i 12 quilograms és igual a 1. Si la probabilitat que la demanda sigui superior a 10 quilograms és igual a 0,03125, la probabilitat que sigui inferior a 10 quilograms és el seu complementari.

$$P(X \leq 10) = 1 - P(X > 10) = 1 - 0,03125 = 0,96875$$

La probabilitat que en un dia qualsevol la demanda sigui inferior a 10 quilograms és igual a 0,96875.

Resolució de la pregunta c

Per calcular la demanda diària esperada, cal calcular l'esperança matemàtica de la variable.

L'esperança matemàtica d'una variable aleatòria contínua és $E[X] = \int_{-\infty}^{+\infty} x \cdot f_d(x) dx$

$$\begin{aligned} E[X] &= \int_0^4 \frac{x}{8} dx + \int_4^{12} x \cdot \frac{(12-x)}{64} dx \\ E[X] &= \frac{1}{8} \left[\frac{x^2}{2} \right]_0^4 + \frac{1}{64} \left[\frac{12 \cdot x^2}{2} - \frac{x^3}{3} \right]_4^{12} \\ E[X] &= \frac{1}{8} \left[\frac{4^2}{2} - 0 \right] + \frac{1}{64} \left[\left(\frac{12 \cdot 12^2}{2} - \frac{12^3}{3} \right) - \left(\frac{12 \cdot 4^2}{2} - \frac{4^3}{3} \right) \right] \\ E[X] &= 1 + \frac{213}{64} = 4,33 \end{aligned}$$

La demanda diària esperada és de 4,33 quilograms.

Resolució de la pregunta d

Sigui Y el benefici diari esperat. Y pot prendre 4 valors:

$$R_Y = \{-5, 5, 10, 15\}$$

Per calcular el benefici esperat cal conèixer la funció de quantia, és a dir, la probabilitat d'ocurrència de cada un d'aquests beneficis.

Probabilitat de tenir un benefici negatiu i igual a -5 u. m.:

$$P(Y = -5) = P(X \leq 2) = \int_0^2 \frac{1}{8} dx = \left[\frac{x}{8} \right]_0^2 = \frac{1}{4}$$

Probabilitat de tenir un benefici positiu de 5 u. m.:

$$P(Y = 5) = P(2 < X \leq 4) = \int_2^4 \frac{1}{8} dx = \left[\frac{x}{8} \right]_2^4 = \left[\frac{4}{8} - \frac{2}{8} \right] = \frac{1}{4}$$

Probabilitat de tenir un benefici de 10 u. m.:

$$P(Y = 10) = P(4 < X \leq 8)$$

$$P(4 < X \leq 8) = \int_4^8 \frac{12-x}{64} dx = \frac{1}{64} \left[12x - \frac{x^2}{2} \right]_4^8 = \frac{1}{64} \left[\left(12 \cdot 8 - \frac{8^2}{2} \right) - \left(12 \cdot 4 - \frac{4^2}{2} \right) \right] = \frac{3}{8}$$

Probabilitat de tenir un benefici de 15 u. m.:

$$P(Y = 15) = P(8 < X \leq 12) = 1 - P(X \leq 8),$$

$$\text{on } P(X \leq 8) = P(X \leq 2) + P(2 < X \leq 4) + P(4 < X \leq 8) = \frac{1}{4} + \frac{1}{4} + \frac{3}{8} = \frac{7}{8}.$$

$$\text{Per tant, } P(8 < X \leq 12) = 1 - \frac{7}{8} = \frac{1}{8}.$$

Amb aquesta informació es pot calcular el benefici esperat $E[Y]$.

L'esperança matemàtica en el cas d'una variable aleatòria discreta és la suma dels productes entre els valors que pren la variable i la probabilitat d'ocurrència d'aquests valors.

$$E[Y] = \sum_{\forall y \in R_Y} y \cdot P(Y = y)$$

$$E[Y] = (-5) \cdot \frac{1}{4} + 5 \cdot \frac{1}{4} + 10 \cdot \frac{3}{8} + 15 \cdot \frac{1}{8} = 5,625$$

El benefici diari esperat és de 5,625 u. m.

Detergent per a rentadores

Tras múltiples estudios, la dirección técnica de una empresa del sector de los detergentes para lavadoras ha llegado a la conclusión de que la producción diaria (en miles de unidades) de una de sus factorías es una variable aleatoria no negativa cuyo comportamiento sigue aproximadamente la siguiente función de densidad:

$$f_d(x) = \frac{6}{A} x(25 - x)$$

- a) Si el límite físico de la producción diaria es de 25000 unidades, obtener el valor del parámetro A para que cumpla los requisitos exigibles a una verdadera función de densidad.
- b) ¿Cuál es la probabilidad de que la producción en un día se sitúe entre las 15000 y las 20000 unidades?
- c) Calcular la varianza de la producción diaria.

Després de diversos estudis, la direcció tècnica d'una empresa del sector dels detergents per a rentadores ha arribat a la conclusió que la producció diària (en milers d'unitats) d'una de les fàbriques és una variable aleatòria no negativa que segueix aproximadament la funció de densitat següent:

$$f_d(x) = \frac{6}{A} x(25 - x)$$

- a) Si el límit físic de la producció diària és de 25000 unitats, calculeu el valor del paràmetre A perquè compleixi els requisits exigibles a una veritable funció de densitat.
- b) Quina probabilitat hi ha que la producció en un dia se situï entre les 15000 i les 20000 unitats?
- c) Calculeu la variància de la producció diària.

Resolució de la pregunta a

Sigui X la producció diària en milers d'unitats d'una de les fàbriques. El límit de la producció diària d'aquella fàbrica és de 25000 unitats. Així, X és una variable que pren valors entre 0, atès que la producció no pot ser negativa, i 25.

Per tant, el rang de la variable és: $R_x = [0; 25]$

La producció diària queda descrita per la funció de densitat:

$$f_d(x) = \begin{cases} \frac{6}{A}x(25-x) & \text{per a } 0 \leq x \leq 25 \\ 0 & \text{altrament} \end{cases}$$

Si es compleix que:

$$\begin{aligned} i. & \quad f_d(x) \geq 0 \quad \forall x \in \mathbb{R} \\ ii. & \quad \int_{-\infty}^{+\infty} f_d(x) dx = 1 \end{aligned}$$

$$i. \quad \text{Atès que } 0 \leq x \leq 25, \text{ la funció } \frac{6}{A}x(25-x) \geq 0 \text{ si } A \geq 0.$$

$$ii. \quad \int_{-\infty}^{\infty} f_d(x) dx = 1$$

$$\Rightarrow \int_0^{25} \frac{6}{A}x \cdot (25-x) dx = 1$$

$$\Rightarrow \frac{6}{A} \int_0^{25} 25x - x^2 dx = 1$$

$$\Rightarrow \frac{6}{A} \cdot \left[\frac{25}{2}x^2 - \frac{1}{3}x^3 \right]_0^{25} = 1$$

$$\Rightarrow A = 6 \cdot \left(\frac{25^3}{2} - \frac{25^3}{3} \right) = 25^3 = 15625$$

Finalment, el paràmetre $A = 25^3 = 15625$ i la producció diària queda descrita per la funció de densitat:

$$f_d(x) = \begin{cases} \frac{6}{25^3} x(25 - x) & \text{per a } 0 \leq x \leq 25 \\ 0 & \text{altrament} \end{cases}$$

Resolució de la pregunta b

Aquesta pregunta es pot resoldre directament amb la funció de densitat (opció a). Alternativament, es pot construir primer la funció de distribució acumulada de la variable i després respondre a la pregunta mitjançant la funció de distribució acumulada (opció b).

Opció a

Atès que X està expressada en milers d'unitats, per trobar aquesta probabilitat cal calcular la integral de la funció de densitat entre 15 i 20.

$$\begin{aligned} P(15 < X < 20) &= \int_{15}^{20} \frac{6}{25^3} x \cdot (25 - x) dx \\ &= \frac{6}{25^3} \left[\frac{25}{2} x^2 - \frac{1}{3} x^3 \right]_{15}^{20} \\ &= \frac{6}{25^3} \left[\left(\frac{25}{2} 20^2 - \frac{1}{3} 20^3 \right) - \left(\frac{25}{2} 15^2 - \frac{1}{3} 15^3 \right) \right] \\ &= \frac{3}{25^2} (20^2 - 15^2) + \frac{2}{25^3} (15^3 - 20^3) = 0,248 \end{aligned}$$

Opció b

En aquest cas, primer es construeix la funció de distribució acumulada i, després, s'obté la probabilitat que la variable prengui un valor en l'interval $[15; 20]$.

El rang de la variable és $R_x = [0; 25]$. Per tant, la probabilitat acumulada fins a 0, el límit inferior del rang, és nul·la i la probabilitat acumulada a partir de 25, el límit superior del rang, és igual a 1. La funció de distribució acumulada per a $x \in [0; 25]$ serà la que resulti de resoldre:

$$\begin{aligned}
 P(X < x) &= \int_0^x \frac{6}{25^3} t \cdot (25 - t) dt \\
 &= \frac{6}{25^3} \left[\frac{25}{2} t^2 - \frac{1}{3} t^3 \right]_0^x \\
 &= \frac{3}{25^2} x^2 - \frac{2}{25^3} x^3
 \end{aligned}$$

La funció de distribució acumulada és:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{3}{25^2} x^2 - \frac{2}{25^3} x^3 & \text{si } 0 \leq x \leq 25 \\ 1 & \text{si } x > 25 \end{cases}$$

A partir de la funció de distribució acumulada, la probabilitat que la variable prengui un valor en l'interval es calcula com la diferència entre la probabilitat que prengui un valor inferior al límit superior de l'interval menys la probabilitat que prengui un valor inferior al límit inferior de l'interval.

$$\begin{aligned}
 P(15 < X < 20) &= F_X(20) - F_X(15) \\
 &= \left(\frac{3}{25^2} 20^2 - \frac{2}{25^3} 20^3 \right) - \left(\frac{3}{25^2} 15^2 - \frac{2}{25^3} 15^3 \right) = 0,248
 \end{aligned}$$

La probabilitat que en un dia la producció se situï entre les 15000 i les 20000 unitats és igual a 0,248.

Resolució de la pregunta c

La variància es pot calcular a partir de la fórmula: $Var[X] = E[X^2] - E[X]^2$. $E[X]$ és l'esperança de la variable. En el cas d'una variable aleatòria contínua és:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_d(x) dx$$

$E[X^2]$ és l'esperança de la variable al quadrat. En el cas d'una variable aleatòria contínua és:

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f_d(x) dx$$

Càlcul de l'esperança de la variable:

$$E[X] = \int_{-\infty}^0 0x \, dx + \int_0^{25} x \cdot \left(\frac{6}{25^3} x(25-x) \right) dx + \int_{25}^{+\infty} 0x \, dx$$

$$E[X] = \int_0^{25} \frac{6}{25^2} x^2 - \frac{6}{25^3} x^3 \, dx$$

$$E[X] = \left[\frac{2}{25^2} x^3 - \frac{3}{2 \cdot 25^3} x^4 \right]_0^{25} = 50 - 37,5 = 12,5$$

Càlcul de l'esperança de la variable al quadrat:

$$E[X^2] = \int_{-\infty}^0 0x^2 \, dx + \int_0^{25} x^2 \cdot \left(\frac{6}{25^3} x(25-x) \right) dx + \int_{25}^{+\infty} 0x^2 \, dx$$

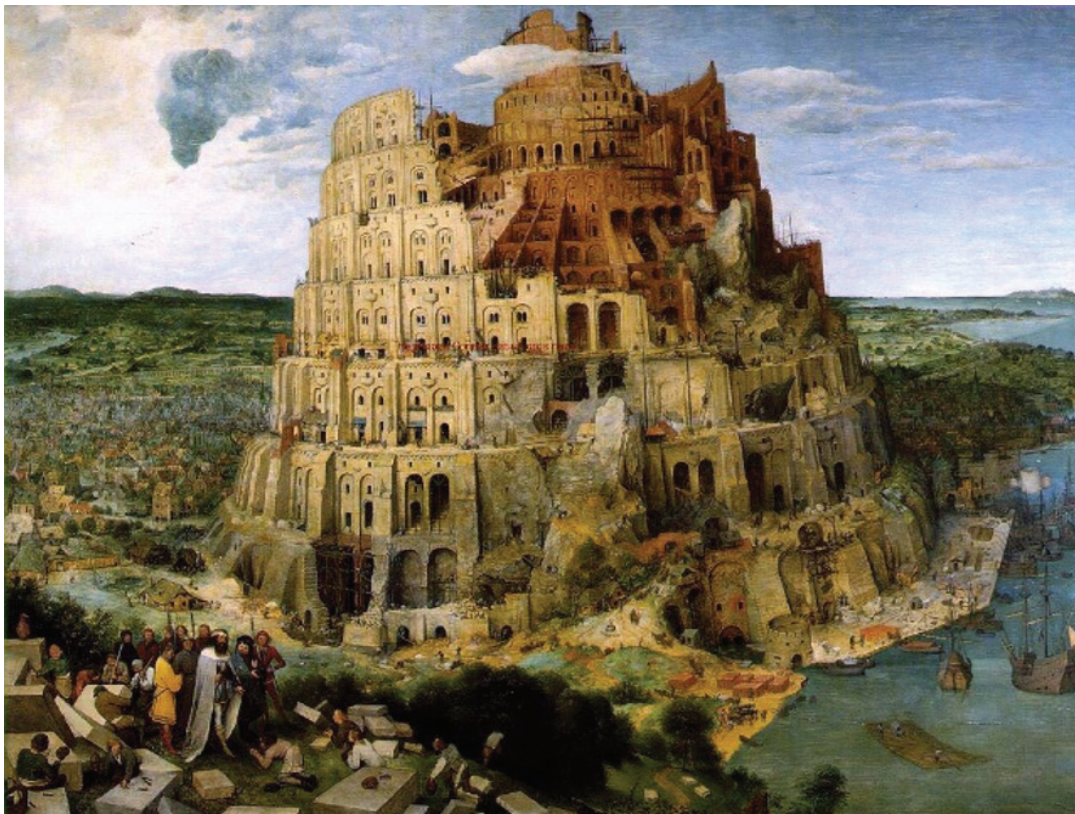
$$E[X^2] = \int_0^{25} \frac{6}{25^2} x^3 - \frac{6}{25^3} x^4 \, dx$$

$$E[X^2] = \left[\frac{3}{2 \cdot 25^2} x^4 - \frac{6}{5 \cdot 25^3} x^5 \right]_0^{25} = \frac{3 \cdot 25^2}{2} - \frac{6 \cdot 25^2}{5} = 187,5$$

Finalment, la variància és $Var[X] = 187,5 - 12,5^2 = 31,25$

La variància de la variable és de 31,25 (milers d'unitats)²;
és a dir, 31250000 unitats².

Torre de Babel



Torre de Babel, de Pieter Brueghel el Vell (1563),
sota llicència CC1.0.

Aquest estiu, en Pere, un noi molt simpàtic amb un somriure encisador i que parla cinc idiomes, ha estat contractat per un restaurant de platja per aconseguir que els turistes hi entrin a dinar. En Pere convenç d'entrar a dinar 1 de cada 3 turistes (o grup) amb qui parla. Treballa cada migdia i té acordat amb el seu cap que després de parlar amb 24 turistes ja pot passar a dinar de franc. A més, per cada turista (o grup) que entra, en Pere rep 5 euros.

- Quina és la probabilitat que en Pere aconsegueixi que entrin exactament 10 turistes (o grups) un dia qualsevol?
- Quina és la probabilitat que aconsegueixi que entrin com a mínim la meitat dels turistes (o grups) amb qui parla un dia qualsevol?
- Quins són els guanys diaris esperats d'en Pere? Quina és la variància d'aquests guanys?
- Aquest estiu, en Pere treballa 87 dies. Trobeu la probabilitat que al llarg de l'estiu guanyi, com a mínim, 3500 euros per aquesta feina. Supposeu que el nombre de turistes que aconsegueix que entrin en un dia és independent del nombre de turistes que entra qualsevol altre dia.

Per resoldre l'exercici es considera que els turistes (o grups de turistes) amb qui parla en Pere són independents entre ells.

Resolució de les preguntes a i b

Sigui X una variable que descriu el nombre de turistes que en Pere aconsegueix que entrin a dinar en un dia. Cal conèixer la distribució d'aquesta variable per poder respondre a les primeres preguntes.

En el supòsit que els turistes (o grups de turistes) són independents els uns dels altres, X es pot modelitzar mitjançant una llei binomial, atès que és un comptatge d'èxits entre $n = 24$ esdeveniments que poden succeir o no succeir amb una probabilitat d'èxit $p = \frac{1}{3}$ constant.

$$X \sim B\left(n = 24, \quad p = \frac{1}{3}\right)$$

Pregunta a

La probabilitat que un dia qualsevol entrin exactament 10 dels 24 turistes és:

$$P(X = 10) = \binom{24}{10} \cdot 0,33^{10} \cdot 0,67^{14} = \frac{24!}{10! \cdot 14!} \cdot 0,33^{10} \cdot 0,67^{14} = 0,1138$$

La funció de quantia de la distribució binomial per als valors que pertanyen al rang és:

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Recordatori: $\binom{n}{x} = \frac{n!}{x! (n-x)!}$

La probabilitat que un dia qualsevol entrin exactament 10 dels 24 turistes és igual a 0,1138.

Pregunta b

La probabilitat que un dia qualsevol entrin, com a mínim, la meitat dels turistes amb qui en Pere parla és la probabilitat que entrin 12 turistes (o grups) o més.

$$P(X \geq 12) = P(X = 12) + P(X = 13) + \dots + P(X = 24)$$

També es pot calcular aquesta probabilitat mitjançant el complementari, tot i que en aquest cas concret ambdues opcions són molt llargues.

$$P(X \geq 12) = 1 - P(X < 12) = 1 - P(X \leq 11)$$

$$P(X \geq 12) = 1 - [P(X = 0) + P(X = 1) + \dots + P(X = 11)]$$

La funció de distribució acumulada de la distribució binomial per als valors que pertanyen al rang no té una formulació compacta i es calcula com una suma de funcions de quantia. Tanmateix, molts programes informàtics en donen el resultat directament.

X	P(X=x)	P(X≤x)
0	0,000059403192	0,000059403192
1	0,000712838305	0,000772241497
2	0,004098820252	0,004871061749
3	0,015029007592	0,019900069341
4	0,039451144929	0,059351214270
5	0,078902289858	0,138253504129
6	0,124928625609	0,263182129738
7	0,160622518640	0,423804648378
8	0,170661426055	0,594466074434
9	0,151699045383	0,746165119816
10	0,113774284037	0,859939403853
11	0,072401817114	0,932341220968
12	0,039217650937	0,971558871905
13	0,018100454279	0,989659326183
14	0,007110892752	0,996770218935
15	0,002370297584	0,999140516520
16	0,000666646196	0,999807162715

17	0,000156857928	0,999964020643
18	0,000030500153	0,999994520796
19	0,000004815814	0,999999336610
20	0,000000601977	0,999999938586
21	0,000000057331	0,999999995918
22	0,000000003909	0,999999999827
23	0,000000000170	0,999999999996
24	0,000000000004	1

Taula amb els resultats numèrics per a cada un dels valors del rang de la funció de quantia $P(X=x)$ i de la funció de distribució acumulada $P(X \leq x)$.

$$P(X \geq 12) = 1 - 0,9323 = 0,0677$$

La probabilitat que un dia qualsevol entrin com a mínim la meitat dels turistes amb qui en Pere parla és igual a 0,0677.

Resolució de la pregunta c

Es demanen els guanys diaris esperats d'en Pere i la variància d'aquests guanys. Cal doncs crear aquesta nova variable.

Sigui G els guanys diaris d'en Pere fruit de la seva feina al restaurant. Atès que per cada turista o grup que entra rep 5 € i que el nombre de turistes que entren al restaurant un dia qualsevol gràcies a en Pere queda recollit per la variable X , G és X multiplicat per 5.

$$G = 5 \cdot X$$

Ara cal buscar l'esperança i la variància de G . Amb aquesta fi, primer es buscarà l'esperança i la variància de X i, després, s'obtinran l'esperança i la variància de G aplicant-hi les propietats del canvi d'escala.

Pas 1: l'esperança i la variància de la variable X , sabent que $X \sim B(n = 24; p = 1/3)$.

En una distribució binomial de paràmetres n i p , l'esperança és igual a n per la probabilitat d'èxit (p) i la variància és igual a n pel producte de la probabilitat d'èxit i de la probabilitat de fracàs ($1-p=q$).

$$E[X] = n \cdot p$$

$$Var[X] = n \cdot p \cdot q$$

El nombre mitjà de clients (o grups de clients) que entren al restaurant per en Pere és:

$$E[X] = 24 \cdot \frac{1}{3} = 8 \text{ clients}$$

La variància del nombre de clients que entren al restaurant gràcies a en Pere és:

$$Var[X] = 24 \cdot \frac{1}{3} \cdot \frac{2}{3} = 5,33 \text{ clients}^2$$

De mitjana, entren 8 clients al dia; la variància és igual a 5,33 clients².

Pas 2: l'esperança i la variància de la variable G , sabent que $G = 5 \cdot X$

Un canvi d'escala en una variable aleatòria X (multiplicar una variable per una constant k) modifica l'esperança de la nova variable aleatòria (G) en el valor de la constant i modifica la variància de la nova variable en el valor de la constant al quadrat.

$$\begin{aligned} \text{Si } G &= k \cdot X \\ E[G] &= k \cdot E[X] \\ Var[G] &= k^2 \cdot Var[X] \end{aligned}$$

L'esperança dels guanys diaris és:

$$E[G] = 5 \cdot E[X] = 5 \cdot 8 = 40 \text{ €}$$

La variància dels guanys diaris és:

$$Var[G] = 5^2 \cdot Var[X] = 5^2 \cdot 5,33 = 133,33 \text{ €}^2$$

Els guanys esperats diaris d'en Pere són de 40 €, amb una variància igual a 133,33 €².

Resolució de la pregunta d

Per respondre a aquesta pregunta es pot optar per diferents vies.

Opció a

Primer es construeix una variable que descriu el nombre de turistes que en Pere aconsegueix que entrin en tot l'estiu i, després, es calcula la probabilitat que al llarg de tot l'estiu entrin com a mínim 700 clients, atès que aquesta quantitat de clients li reportaria uns guanys de 3500 euros. La variable nombre de turistes que en Pere aconsegueix que entrin al llarg de l'estiu es pot construir com una binomial (**opció a1**) o bé aproximar-la a una normal mitjançant el teorema de Laplace-de Moivre (**opció a2**).

Opció b

Es construeix una variable que descriu els guanys d'en Pere de tot l'estiu i es calcula la probabilitat que els guanys assoleixin com a mínim 3500 euros.

Opció a1

Sigui W el nombre de turistes que en Pere aconsegueix que entrin al llarg de l'estiu. W és la suma dels turistes que entren durant els 87 dies:

$$W = \sum_{i=1}^{87} X_i = X_1 + X_2 + \dots + X_{87}$$

A partir de la propietat additiva de la distribució binomial, es pot deduir que W també segueix una llei binomial:

$$W \sim B\left(n = 24 \cdot 87 = 2088, \quad p = \frac{1}{3}\right)$$

Propietat additiva de la distribució binomial

La suma de dues o més distribucions binomials, cadascuna amb la seva grandària (n_i) però amb la mateixa probabilitat d'èxit (p), genera una distribució binomial amb una n igual a la suma de totes les n_i i la p inalterada.

$$\left. \begin{array}{l} X_1 \sim B(n_1, p) \\ X_2 \sim B(n_2, p) \\ \dots \\ X_m \sim B(n_m, p) \end{array} \right\} \sum_{i=1}^m X_i \sim B\left(\sum_{i=1}^m n_i, p\right)$$

Per aconseguir 3500 euros o més, cal que en tot l'estiu entrin com a mínim $3500/5 = 700$ turistes.

$$P(W \geq 700) = 1 - P(W < 700) = 1 - [P(W = 0) + P(W = 1) + \dots + P(W = 699)] \\ \Rightarrow P(W \geq 700) = 0,4345$$

[resultat 1: calculat amb un ordinador]

La probabilitat que al llarg de l'estiu en Pere guanyi com a mínim 3 500 euros per aquesta feina és igual a 0,4345. [resultat 1]

Opció a2

Una opció tradicionalment més ràpida és aproximar la distribució de W a una distribució normal mitjançant el resultat del teorema de Laplace-de Moivre. Aquest teorema es pot considerar un antecessor particular del teorema central del límit.

El teorema de Laplace-de Moivre estableix que ateses $X_1, X_2, \dots, X_n$ variables aleatòries independents entre elles que segueixen una distribució de Bernoulli de paràmetre p , la suma d'aquestes variables (S_n) convergeix cap a una distribució normal quan n tendeix cap a l'infinit.

En concret, la suma $S_n = X_1 + X_2 + \dots + X_n$ es correspon amb una distribució binomial de paràmetres n i p : $S_n \sim B(n, p)$ que convergeix cap a una distribució normal de paràmetres $\mu = n \cdot p$ i $\sigma^2 = n \cdot p \cdot q$.

$$W \approx N\left(\mu = 2088 \cdot \frac{1}{3} = 696, \quad \sigma^2 = 2088 \cdot \frac{1}{3} \cdot \frac{2}{3} = 464\right)$$

Atès que la distribució normal és contínua, mentre que la distribució binomial és discreta, per calcular $P(W \geq 700)$ és adequat aplicar-hi una correcció per continuïtat i calcular $P(W \geq 699,5)$.

La correcció per continuïtat reparteix l'espai que hi ha entre 699 i 700 en dos. Cal tenir present que en la distribució binomial $P(W \geq 700) = 1 - P(W \leq 699)$, atès que $P(699 < W < 700) = 0$, mentre que en la distribució normal $P(W \geq 700) \neq 1 - P(W \leq 699)$, atès que $P(699 < W < 700) \neq 0$.

$$P(W \geq 699,5) = P\left(Z \geq \frac{699,5 - 696}{\sqrt{464}}\right) = P(Z \geq 0,1625) = 0,4355$$

[resultat 2]

El resultat calculat mitjançant aquesta aproximació estableix que la probabilitat que al llarg de l'estiu en Pere guanyi com a mínim 3500 euros per aquesta feina és igual a 0,4355. [resultat 2]

Es pot recalcar que sense la correcció per continuïtat (és a dir, calculant directament $P(W \geq 700)$ mitjançant l'aproximació a la normal), l'aproximació hauria estat clarament pitjor i igual a 0,4263 [resultat sense correcció per continuïtat]. Les aproximacions eren el mètode emprat antigament, quan no es feien servir ordinadors.

Opció b

Sigui IE els ingressos de tot l'estiu d'en Pere en concepte del nombre de turistes que entren al restaurant. IE és la suma dels ingressos de cada dia, que se suposa que són independents els uns dels altres.

$$IE = G_1 + G_2 + \dots + G_{87}$$

El teorema central del límit permet deduir que IE s'aproxima a una distribució normal amb una mitjana igual a la suma de mitjanes i una variància igual a la suma de variàncies. En aquest cas, a diferència del teorema de Laplace-de Moivre, no és important la distribució d'origen de la variable, mentre sigui una suma de variables aleatòries independents i idènticament distribuïdes.

$$\mu_{IE} = 40 + 40 + \dots + 40 = 40 \cdot 87 = 3480$$

$$\sigma_{IE}^2 = 133,33 + 133,33 + \dots + 133,33 = 133,33 \cdot 87 = 11600$$

$$IE \approx N(\mu_{IE} = 3480, \sigma_{IE}^2 = 11600)$$

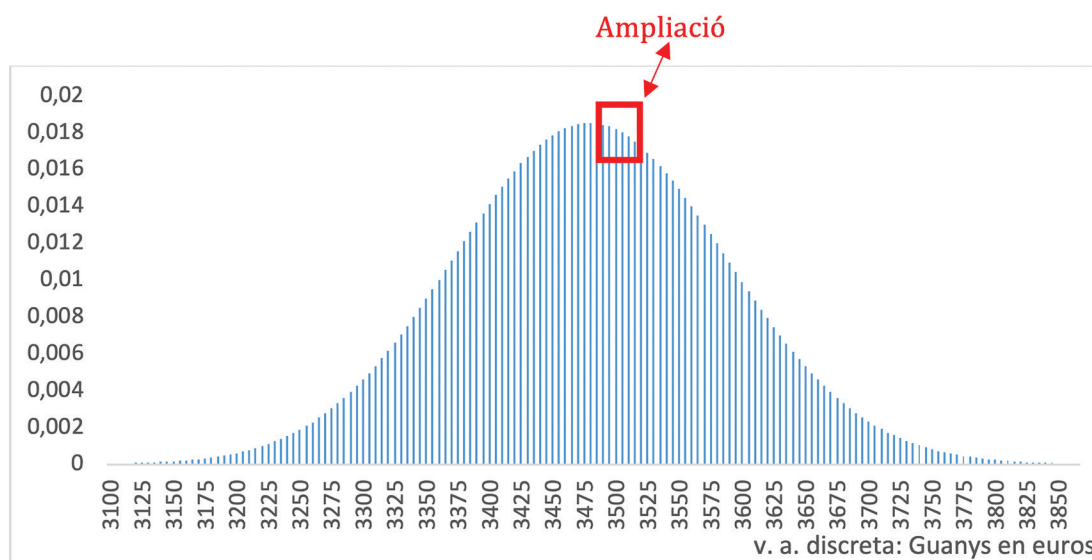
Es busca que els ingressos de la temporada siguin, com a mínim, de 3500 €. Atenció! Si es calcula:

$$P(IE \geq 3500) = P(Z \geq \frac{3500-3480}{\sqrt{11600}}) = 0,4263$$

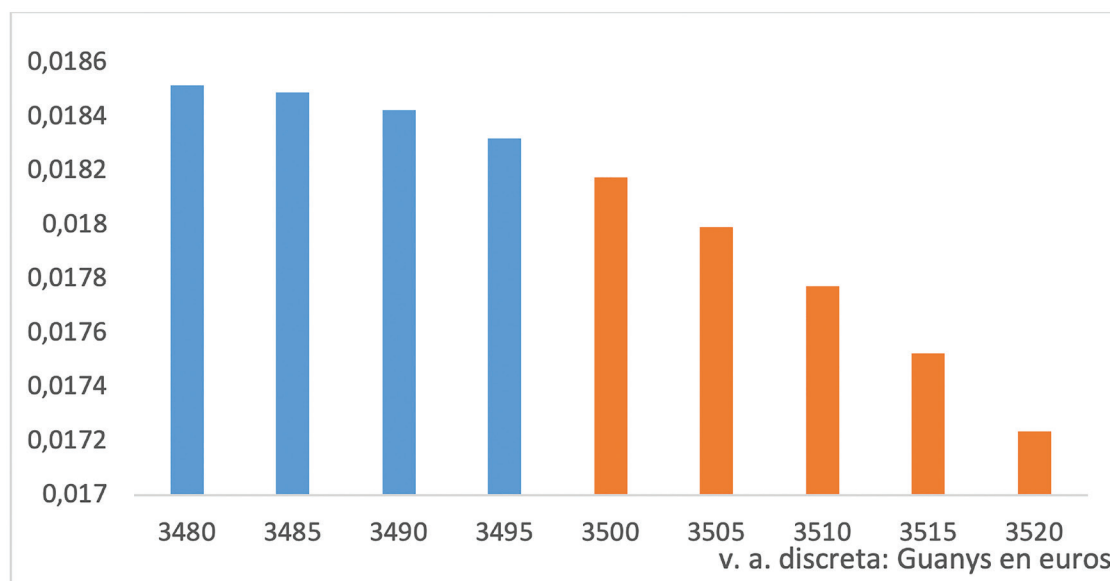
[resultat sense correcció per continuïtat]

s'obté una aproximació barroera del resultat.

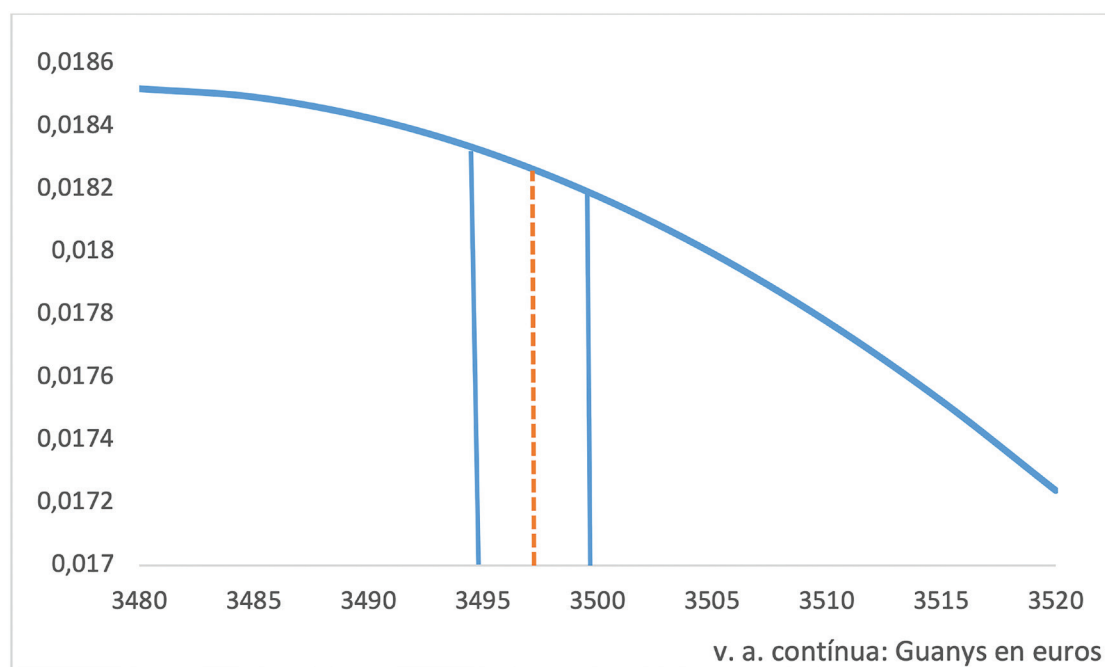
Com ja s'ha comentat, això és fruit del fet que la variable original és una variable aleatòria discreta, mentre que la distribució normal representa una variable aleatòria contínua. En fer l'aproximació, es tracta una variable aleatòria discreta com si fos contínua.



Gràfica 1: Distribució dels possibles guanys de la temporada.



Gràfica 2: Ampliació del rectangle vermell de la gràfica 1.



Gràfica 3: Aproximació a la normal del rectangle vermell de la gràfica 1.

Aquest resultat és millorable fent una correcció per continuïtat, que consisteix a dividir l'espai que queda entre les dues barres blaves en dues parts iguals.

Atès que un turista genera 5 euros, els salts van de 5 en 5. La correcció per continuïtat divideix aquest salt que va de 4495 a 4500 en dues parts iguals. Una aproximació millorada és:

$$P(IE \geq 3497,5) = P\left(Z \geq \frac{3497,5 - 3480}{\sqrt{11599,7}}\right) = 0,4355 \quad [\text{resultat 2}]$$

La probabilitat que al llarg de l'estiu en Pere guanyi com a mínim 3500 euros per aquesta feina és aproximadament igual a 0,4355. [resultat 2]

Comentari de l'apartat d

Aquest exercici permet veure les diferències entre el resultat del càlcul emprant la distribució exacta [resultat 1] o una aproximació a la normal [resultat 2]. Òbviament, el resultat més precís és el de la distribució original [resultat 1], en aquest cas la binomial. També s'explica la problemàtica de passar d'una variable aleatòria discreta a una de contínua i la solució, sempre aproximada, que ofereix la correcció per continuïtat [resultat 2]. El resultat depèn de la tria que es faci: emprar una distribució exacta o aproximada i, en aquest darrer cas, fer una correcció per continuïtat o no fer-la.

Abraçades



Joves de diferents parts del món ofereixen abraçades gratuïtes als vianants prop d'una estació de metro. Panamà, de Yoiska Muñoz, sota llicència CC BY-SA 4.0.

Els anys abans la pandèmia de COVID-19 es va posar de moda fer abraçades: hi havia llibres que n'explicaven els grans beneficis terapèutics, tasses amb dibuixos de dònuts feliçment abraçats i amics que van canviar els clàssics petonets per una efusiva abraçada, fins i tot hi havia activistes que, sense avisar prèviament, abraçaven desconeguts pel carrer.

És l'any 2017, a la Maria l'acaba d'abraçar un senyor força ben plantat que en sa vida havia vist. Ja feia setmanes que no li passava; ha constatat que li passa una mitjana

d'un cop l'any i comença a pensar a modelitzar el nombre d'abraçades que li fan els estranys en un període determinat mitjançant una distribució de Poisson.

- a) Expliqueu en quins supòsits és raonable aquesta modelització.
- b) Trobeu la probabilitat que en un any la Maria sigui abraçada exactament dos cops per desconeguts.
- c) Trobeu la probabilitat que en un any la Maria sigui abraçada com a mínim un cop per desconeguts.
- d) Trobeu la probabilitat que en dos anys la Maria sigui abraçada exactament quatre vegades per desconeguts.
- e) Trobeu la probabilitat que en un mes la Maria sigui abraçada exactament un cop per desconeguts.
- f) De mitjana, quantes vegades l'abracen en un mes? Quina és la desviació estàndard de la variable "nombre d'abraçades de desconeguts mensual que li fan a la Maria".
- g) De mitjana, quants mesos passen entre dues abraçades de desconeguts? Quina és la desviació estàndard de la variable "mesos que transcorren entre dues abraçades de desconeguts consecutives"?
- h) De mediana, quants mesos passen entre dues abraçades de desconeguts?
- i) Acaben d'abraçar la Maria, quina és la probabilitat que hagi d'esperar com a mínim 2 anys fins a la següent abraçada d'un desconegut?

Resposta a la pregunta a

Una distribució de Poisson serveix per modelitzar fenòmens aleatoris dicotòmics (èxit-fracàs) que succeeixen en un espai continu (el temps, per exemple). En aquest context, la distribució modelitza una variable que es defineix com a nombre d'èxits (nombre d'abraçades) en un lapse de temps determinat (per exemple, en un any o en un mes).

Perquè el model pugui ajustar-se bé, és una condició necessària que els èxits siguin independents entre ells, en aquest cas que els activistes siguin independents els uns dels altres. A tall de contraexemple, si un dia sortissin a abraçar 5 activistes conjuntament i abracessin el mateix grup de persones l'un darrere l'altre, no seria adequat utilitzar una distribució de Poisson per modelitzar el nombre d'abraçades. També cal que el ritme dels èxits sigui constant en el temps o, dit d'una altra manera, que no es detectin uns períodes en què augmenta el nombre d'abraçades i altres en què disminueix. Segons això, si per exemple es detectés que hi ha estacionalitat en el nombre d'abraçades, de manera que el nombre augmentés a la primavera i disminuís a la tardor, tampoc no seria adequat modelitzar el nombre d'abraçades anual mitjançant una distribució de Poisson.

Finalment, cal dir que tradicionalment la distribució de Poisson s'ha emprat per modelitzar fenòmens rars, és a dir, poc freqüents. Si de sobte els estranys fessin abraçades a totes hores, és a dir, si la freqüència fos elevada, es podria emprar la distribució normal per modelitzar el fenomen. Això s'entén millor apel·lant als orígens de la distribució. Formalment, la distribució de Poisson es descriu com a límit de la distribució binomial quan la n tendeix a l'infinit (un espai continu) i la probabilitat d'ocurrència tendeix a 0 (un fenomen rar). En canvi, la distribució binomial tendeix cap a una distribució normal quan la n tendeix cap a l'infinit i la probabilitat no és extrema (pren valors al voltant de 0,5), un fenomen que té tantes possibilitats de passar com de no passar. Així, la distribució de Poisson tendeix cap a la normal quan un fenomen deixa de ser rar.

Resolució de la pregunta b

Com s'ha dit, s'utilitza la distribució de Poisson per modelitzar el nombre d'abraçades en un lapse de temps determinat. S'especifica el lapse de temps tenint en compte la pregunta (per any).

Es defineix una variable (X_{any}) com el nombre d'abraçades per any. Aquesta variable es modelitza mitjançant una distribució de Poisson de paràmetre λ_{any} , on λ_{any} és el nombre mitjà d'abraçades per any, que, en aquest exercici, és igual a 1.

$$X_{any} \sim P(\lambda_{any} = 1)$$

La funció de quantia de la distribució de Poisson és .

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \forall x \in R_X$$

Es busca la probabilitat que en un any abracin la Maria exactament dues vegades:

$$P(X_{any} = 2) = \frac{e^{-1} 1^2}{2!} = \frac{e^{-1}}{2} = 0,1839$$

La probabilitat que en un any uns desconeguts abracin la Maria exactament dues vegades és igual a 0,1839.

Resolució de la pregunta c

En aquest cas es busca la probabilitat que abracin la Maria com a mínim un cop en un any. El lapse de temps continua essent un any i, per tant, es torna a treballar amb X_{any} .

$$\begin{aligned}
 & P(X_{any} \geq 1) \\
 &= P(X_{any} = 1) + P(X_{any} = 2) + \dots \\
 &= 1 - P(X_{any} = 0) \\
 &= 1 - \frac{e^{-1} 1^0}{0!} = 1 - e^{-1} \\
 &= 1 - 0,3679 = 0,6321
 \end{aligned}$$

La probabilitat que en un any abracin la Maria com a mínim una vegada és igual a 0,6321.

Resolució de la pregunta d

Per respondre a aquest apartat es defineix una segona variable que descriu el nombre d'abraçades en dos anys (X_{2anys}). En aplicació de la propietat additiva de la funció de distribució de Poisson, es dedueix que X_{2anys} segueix una distribució de Poisson de paràmetre $\lambda_{2anys} = \lambda_{any} + \lambda_{any} = 2$.

$$X_{2anys} \sim P(\lambda_{2anys} = 2)$$

La propietat additiva, també coneguda com a *propietat reproductiva*, estableix que la suma de variables aleatòries independents entre elles que segueixen una distribució de Poisson també segueix una distribució de Poisson. El paràmetre (λ) de la variable suma s'obté sumant els paràmetres de les distribucions d'origen.

Un cop especificada la distribució de probabilitat, ja es pot respondre a la pregunta sobre quina és la probabilitat que en dos anys uns desconeguts abracin la Maria exactament quatre cops.

$$P(X_{2anys} = 4) = \frac{e^{-2} \cdot 2^4}{4!} = \frac{e^{-2} \cdot 16}{4 \cdot 3 \cdot 2 \cdot 1} = \frac{2}{3} \cdot e^{-2} = 0,0902$$

La probabilitat que en dos anys uns desconeguts abracin la Maria quatre cops és igual a 0,0902.

Resolució de la pregunta e

Per respondre a aquest apartat, es defineix una tercera variable que descriu el nombre d'abraçades per mes (X_{mes}). Atès que la funció de distribució de Poisson és infinitament divisible, X_{mes} segueix una distribució de Poisson de paràmetre $\lambda_{mes} = \frac{1}{12} = 0,083$.

$$X_{mes} \sim P(\lambda_{mes} = \frac{1}{12})$$

Es diu que una funció de distribució de probabilitat és infinitament divisible si es pot expressar com la funció de distribució de probabilitat de la suma d'un nombre qualsevol de variables aleatòries independents i idènticament distribuïdes (iid). En la resolució d'aquest exercici s'ha acceptat implícitament que tots els mesos tenen la mateixa distribució, tot i que, com que no tenen exactament el mateix nombre de dies, es podria qüestionar el supòsit.

Un cop especificada la distribució de probabilitat, ja es pot respondre a la pregunta sobre quina és la probabilitat que en un mes qualsevol abracin la Maria exactament una vegada:

$$P(X_{mes} = 1) = \frac{e^{-1/12} \cdot (1/12)^1}{1!} = e^{-0,08\hat{3}} \cdot 0,083\hat{3} = 0,0767$$

La probabilitat que en un mes abracin la Maria un i només un cop és igual a 0,0767.

Resolució de la pregunta f

En aquesta pregunta cal calcular l'esperança i la desviació estàndard de la variable X_{mes} .

L'esperança d'una variable que segueix una distribució de Poisson de paràmetre λ és igual a λ , la variància també és igual a λ i la desviació estàndard és $+\sqrt{\lambda}$.

$$\text{Atès que, } X_{mes} \sim P(\lambda_{mes} = \frac{1}{12})$$

$$E[X_{mes}] = \frac{1}{12} = 0,0833 \text{ abraçades}$$

$$DE[X_{mes}] = + \sqrt{\frac{1}{12}} = 0,2887 \text{ abraçades}$$

La Maria és abraçada una mitjana de 0,0833 vegades al mes per desconeguts. La desviació estàndard de la variable és igual a 0,2887 abraçades.

Resolució de la pregunta g

Per respondre a aquesta pregunta, cal modelitzar el temps que transcorre entre dues abraçades (en lloc del nombre d'abraçades en un lapse de temps).

El temps entre dos èxits consecutius —el nombre dels quals en un lapse de temps T (any, mes...) segueix una distribució de Poisson— segueix una distribució exponencial. Aquest temps es mesura en T unitats. Les dues distribucions tenen el mateix paràmetre poblacional λ_T .

Atès que es demana expressar el temps entre dues abraçades en mesos, es parteix de la distribució nombre d'abraçades en mesos X_{mes} . Ja s'ha vist que

$$X_{mes} \sim P(\lambda_{mes} = \frac{1}{12})$$

Sigui Y_{mes} el temps en mesos entre dues abraçades consecutives:

$$Y_{mes} \sim \exp(\lambda_{mes} = \frac{1}{12})$$

El nombre mitjà de mesos que passen és igual a l'esperança de la distribució entre dues abraçades consecutives.

L'esperança d'una variable que segueix una distribució exponencial de paràmetre λ és igual a $\frac{1}{\lambda}$, la variància és igual a $\frac{1}{\lambda^2}$ i la desviació estàndard és $+\sqrt{\frac{1}{\lambda^2}}$, és a dir $\frac{1}{\lambda}$.

$$E[Y_{mes}] = \frac{1}{1/12} = 12$$

$$DE[Y_{mes}] = + \sqrt{\frac{1}{1/144}} = 12$$

Passen una mitjana de 12 mesos entre dues abraçades. La desviació estàndard de la variable és igual a 12 mesos.

Resolució de la pregunta b

Per respondre a aquesta pregunta cal trobar el nombre de mesos y_{Me} , de manera que:

$$P(Y_{mes} \leq y_{Me}) = P(Y_{mes} > y_{Me}) = 0,5$$

La funció de distribució acumulada de la distribució exponencial és

$$P(X \leq x) = 1 - e^{-\lambda \cdot x} \quad \forall x > 0$$

Per tant,

$$\begin{aligned} P(Y_{mes} \leq y_{Me}) &= 1 - e^{-\frac{1}{12}y_{Me}} \\ \Rightarrow 1 - e^{-\frac{1}{12}y_{Me}} &= 0,5 \\ \Rightarrow e^{-\frac{1}{12}y_{Me}} &= 0,5 \\ \Rightarrow \ln\left(e^{-\frac{1}{12}y_{Me}}\right) &= \ln(0,5) \\ \Rightarrow -\frac{1}{12}y_{Me} &= \ln(0,5) \\ \Rightarrow y_{Me} &= -12 \cdot \ln(0,5) = 8,3178 \text{ mesos} \end{aligned}$$

La mediana del temps que passa entre dues abraçades és igual a 8,32 mesos.

Resolució de la pregunta i

Per respondre a aquesta pregunta es pot optar tant per emprar una distribució exponencial (opció a) com per emprar una distribució de Poisson (opció b). En el primer cas es calcula la probabilitat que el temps sigui superior a 2 anys i en el segon, que no li facin cap abraçada en dos anys.

Opció a

Atès que en aquesta pregunta la unitat de mesura del temps són anys, es pot crear la variable Y_{any} : “temps en anys entre dues abraçades consecutives”.

$$Y_{any} \sim \exp(\lambda_{any} = 1)$$

$$\begin{aligned} P(Y_{any} > 2) &= 1 - P(Y_{any} \leq 2) \\ &= 1 - (1 - e^{-2}) \\ &= e^{-2} = 0,1353 \end{aligned}$$

Opció b

Una altra manera de resoldre aquesta pregunta és mitjançant la variable X_{2anys} : “nombre d’abraçades de desconeguts que rep la Maria en 2 anys” (variable emprada en l’apartat d).

$$X_{2anys} \sim P(\lambda_{2anys} = 2)$$

El fet que hagi d’esperar com a mínim 2 anys fins a l’abraçada següent implica que en dos anys el nombre d’abraçades és igual a zero.

$$P(X_{2anys} = 0) = \frac{e^{-2} \cdot 2^0}{0!} = e^{-2} = 0,1353$$

La probabilitat que la Maria hagi d’esperar dos anys o més fins a l’abraçada següent és igual a 0,1353.

Comentaris

Aquest exercici sobre la distribució de Poisson i la distribució exponencial permet veure que per calcular la probabilitat d’ocurrència d’un fenomen en un lapse de temps determinat cal definir la funció de probabilitat per a aquell temps. No és correcte sumar resultats de Poisson quan el temps varia. Això s’emplifica comparant els apartats b i d. En l’apartat b es calcula la probabilitat que es produeixin dos èxits en un any (el resultat és 0,1839), en l’apartat d es demana la probabilitat de quatre èxits en dos anys. Es podria pensar que el resultat és semblant, atès que la mitjana és de dos èxits per any; tanmateix, el resultat és totalment diferent (0,0902).

D’altra banda, també es veu com a vegades es pot resoldre una mateixa pregunta mitjançant diferents distribucions de forma exacta (apartat i), atès que la distribució de Poisson i la distribució exponencial estan relacionades.

Avets a la fira de Santa Llúcia



Plantació d'avets nadalencs, de Howard Mattinson, sota llicència CC BY-SA 2.0.

El pes d'un avet de mida mitjana (sense el test) en una fira nadalenca segueix una llei normal amb una mitjana de 16 quilograms i una variància de 9 quilograms². El test de terrissa de la Bisbal de l'Empordà pesa 1,2 quilograms.

- a) En Pere, pare de 2 nens i que va carregat de paquets, escull un avet a l'atzar.
 - i. Trobeu la probabilitat que l'avet (sense el test) pesi més de 20 quilograms.
 - ii. Trobeu la probabilitat que l'avet amb el test inclòs pesi més de 20 quilograms.
- b) La senyora Llúcia té encarregats 4 avets mitjans sense test (un per a la casa de cada un dels seus nets). Suposant que els 4 avets hagin estat escollits a

l'atzar, trobeu la probabilitat que l'encàrrec de la senyora Llúcia pesi més de 75 quilograms.

- c) El transportista de la senyora Llúcia li cobra per dur els avets a casa dels seus nets 55 cèntims per quilogram. Trobeu la probabilitat que la senyora Llúcia pagui menys de 30 euros al transportista.

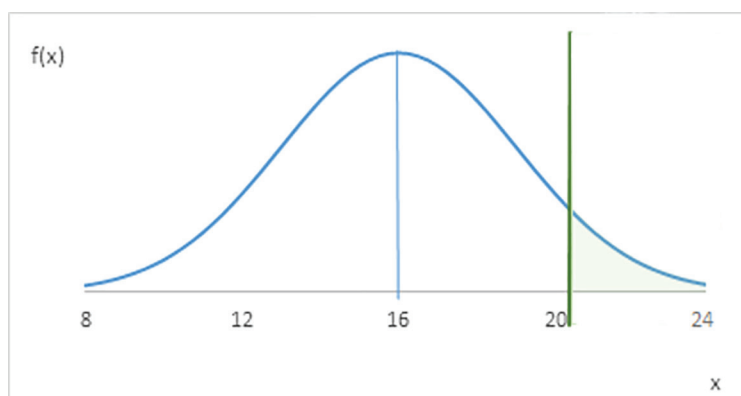
Resolució de la pregunta a.i: probabilitat que l'avet pesi més de 20 quilograms

Sigui X el pes en quilograms d'un avet de mida mitjana.

$$X \sim N(\mu_X = 16; \sigma_X^2 = 9)$$

X segueix una distribució normal amb una mitjana de $\mu_X = 16$ i una variància de $\sigma_X^2 = 9$. Es busca $P(X > 20)$, probabilitat que correspon a l'àrea verda de la gràfica.

Gràfica 1: Àrea corresponent a $P(X > 20)$ en la representació gràfica de $N(\mu_X = 16; \sigma_X^2 = 9)$



Molts programes donen el resultat de la funció de distribució acumulada $P(X \leq x)$ i, per tant, es pot buscar el resultat de l'àrea verda mitjançant el complementari. En la gràfica 1, $P(X \leq 20)$ és l'àrea blanca situada a sota de la corba

$$P(X > 20) = 1 - P(X \leq 20) = 1 - 0,9088 = 0,0912$$

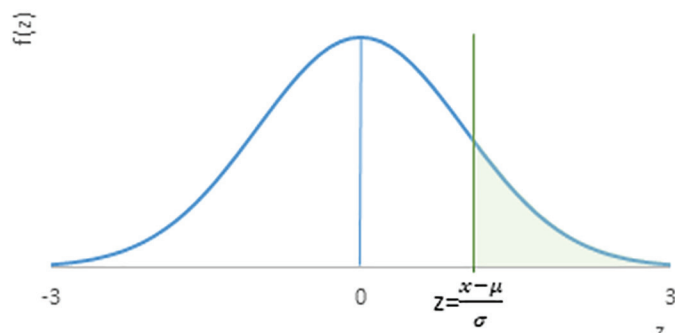
També es pot buscar l'àrea en la normal estàndard (Z), és a dir, una normal amb una mitjana 0 i una variància 1: $Z \sim N(\mu_Z = 0; \sigma_Z^2 = 1)$. Per això, cal trobar un valor z de manera que $P(Z > z) = P(X > 20)$, tal com es mostra en la gràfica 2.

Sigui $X \sim N(\mu_X; \sigma_X^2)$ i sigui $Z \sim N(\mu_Z = 0; \sigma_Z^2 = 1)$

$$P(X \leq x) = P\left(Z \leq \frac{x - \mu_X}{\sqrt{\sigma_X^2}}\right) = P\left(Z \leq \frac{x - \mu_X}{\sigma_X}\right)$$

Per trobar el valor de z que compleix $P(Z > z) = P(X > 20)$, cal tipificar la variable restant la mitjana i dividint per la desviació estàndard.

Gràfica 2: Àrea corresponent a $P(Z > z) = P(X > 20)$ en la representació gràfica de $N(\mu_X = 0; \sigma_X^2 = 1)$



$$P(X > 20) = P(Z > \frac{20-16}{3}) = P(Z > \frac{4}{3}) = P(Z > 1, \hat{3}) = 1 - P(Z \leq 1, \hat{3}) = 0,0912$$

La probabilitat que l'avet d'en Pere pesi més de 20 quilograms és igual a 0,09.

Resolució de la pregunta a.ii: probabilitat que l'avet amb el test inclòs pesi més de 20 quilograms

Sigui XT el pes en quilograms d'un avet de mida mitjana amb un test.

$$XT = \text{pes de l'avet} + \text{pes del test} = X + 1,2$$

A partir de les propietats de l'esperança i de la variància, es dedueix que:

$$E[XT] = E[X + 1,2] = E[X] + E[1,2] = E[X] + 1,2 = 16 + 1,2 = 17,2$$

$$\text{Var}[X] = \text{Var}[X + 1,2] = \text{Var}[X] + \text{Var}[1,2] = \text{Var}[X] + 0 = 9$$

Atès que XT és una normal amb un canvi d'origen, XT segueix una normal:

$$XT \sim N(\mu_{XT} = 17,2; \sigma_{XT}^2 = 9)$$

Si a una variable aleatòria que segueix una distribució normal s'hi suma una constant (k), es genera una nova variable aleatòria que també segueix una distribució normal. El valor de l'esperança de la nova variable haurà variat. En canvi, no s'en modifica la variància.

A partir d'aquí, aquest apartat es resol igual que l'anterior: $P(XT > 20) = 0,1753$.

La probabilitat que l'avet amb test d'en Pere pesi més de 20 quilograms és igual a 0,18.

Resolució de la pregunta b

Per respondre a aquesta pregunta, primer cal construir una variable que representi el pes dels 4 avets. Després cal buscar quina distribució segueix la nova variable.

Sigui Y el pes en quilograms dels 4 avets. Y és la suma del pes dels avets de cadascun dels nets; cada avet pot tenir un pes diferent.

$$Y = X_1 + X_2 + X_3 + X_4$$

X_1 és el pes de l'avet del primer dels 4 nets, X_2 és el pes de l'avet del segon, X_3 és el pes de l'avet del tercer i X_4 és el pes de l'avet del quart.

La suma de distribucions normals segueix una distribució normal. La mitjana de la nova variable és igual a la suma de mitjanes. La variància de la nova variable és igual a la suma de variàncies, si les distribucions d'origen són independents entre elles. Aquest és un cas particular de la propietat additiva de la normal.

Per tant, la propietat additiva de la normal permet saber que la variable Y segueix una normal. Se suposa que els pesos dels avets són independents entre ells.

$$E(Y) = 16 + 16 + 16 + 16 = 64 \text{ quilograms}$$

$$\text{Var}(Y) = 9 + 9 + 9 + 9 = 36 \text{ quilograms}$$

Finalment:

$$Y \sim N(\mu_Y = 64; \sigma_Y^2 = 36)$$

Un cop deduïda la distribució de Y , es planteja la pregunta:

$$P(Y > 75) = 0,0334$$

La probabilitat que l'encàrrec de la senyora Llúcia pesi més de 75 quilograms és igual a 0,0334.

Resolució de la pregunta c

En aquesta ocasió, cal construir una variable que representi el cost del transport dels avets. Després cal buscar quina distribució segueix aquesta variable.

Sigui T el cost en euros del transport dels 4 avets. Aquest cost s'obté multiplicant el pes dels avets en quilograms per 0,55. Es tracta d'un canvi d'escala.

$$T = 0,55 \cdot Y$$

La variable T es construeix a partir de la variable Y , que segueix una distribució normal, i s'hi aplica un canvi d'escala.

Si una variable aleatòria que segueix una distribució normal es multiplica per una constant (k), genera una nova variable aleatòria que també segueix una distribució normal. L'esperança i la variància de la nova variable hauran variat: l'esperança de la nova variable serà igual a l'esperança de l'antiga variable multiplicada per k i la variància de la nova variable serà igual a la variància de l'antiga variable multiplicada per k^2 .

Així doncs, la variable T segueix una distribució normal. L'esperança i la variància de T són:

$$\begin{aligned} E[T] &= E[0,55 \cdot Y] \\ &= 0,55 \cdot E[Y] \\ &= 0,55 \cdot 64 \\ &= 35,2 \text{ €} \end{aligned}$$

$$\begin{aligned} Var[T] &= Var[0,55 \cdot Y] \\ &= 0,55^2 \cdot Var[Y] \\ &= 0,3025 \cdot 36 \\ &= 10,89 \text{ €}^2 \end{aligned}$$

Finalment:

$$T \sim N(\mu_T = 35,2; \sigma_T^2 = 10,89)$$

Un cop deduïda la distribució de T , es planteja la pregunta:

$$P(T < 30) = 0,0575$$

La probabilitat que el cost del transport de l'encàrrec de la senyora Llúcia valgui menys de 30 € és igual a 0,0575.

Comentaris

Aquest exercici va bé per distingir entre la suma de variables aleatòries i un canvi d'escala (apartats b i c, respectivament).

En el primer apartat es treballa amb la variable “pes d'un avet”. Aquesta és la variable que es defineix a l'enunciat i segueix una distribució normal. A partir d'aquesta variable inicial es crea la variable “pes d'un avet amb test”, mitjançant un canvi d'origen (s'hi suma una constant).

En el segon apartat es considera la variable “pes d’un conjunt de 4 avets”. Es tracta d’una suma de variables aleatòries independents. Per què una suma de variables aleatòries i no $4X$? Perquè no tots els avets pesen el mateix. Si es fes servir $4X$ voldria dir que són 4 avets idèntics (tots amb el mateix pes). Només indicant un valor tindríem el pes de tots 4 avets. En canvi, indicant $X_1 + X_2 + X_3 + X_4$ es dona la possibilitat que no hi hagi cap avet igual. Això sí, estan tots extrets d’una mateixa població d’avets que segueix una distribució normal amb una mitjana de 16 quilograms i una variància 9 quilograms². La variable Y es distribuirà d’acord amb una normal, ja que es construeix a partir de la suma de distribucions normals. En el tercer apartat es valora el cost del transport d’aquests 4 avets, recollit en la variable T . Es dona el cost per quilo. Per tant, si es coneix quant pesen en conjunt els 4 avets (variable Y), se’n pot deduir el cost del transport. Es tracta d’un canvi d’escala (es multipliquen els valors de la variable per una constant).

Exercici resolt equivalent

Taronges

El pes d’una malla de taronges de suc segueix una llei normal amb una mitjana de 2 quilograms i una variància de 0,01 quilograms².

- a) En Jaume agafa una malla de la pila a l’atzar. Trobeu la probabilitat que:
 - la malla pesi més de 2,1 quilograms.
 - la malla pesi menys de 2,1 quilograms.
 - la malla pesi més d’1,9 quilograms.
 - la malla pesi menys d’1,8 quilograms.
 - la malla pesi entre 1,8 quilograms i 2,1 quilograms.
- b) La Berta agafa 3 malles a l’atzar. Trobeu la probabilitat que la Berta compri menys de 5,8 quilograms de taronges de suc.
- c) El preu per quilogram de les taronges és d’1,6 €. Trobeu la probabilitat que la Berta pagui més de 10 € per les taronges.
- d) A més de les taronges, la Berta compra 2 rajoles de xocolata negra del 85 % a 2,15 € cada una, una barra de pa per 1 € i un tall de formatge suís curat per 4,80 €. Treu un bitllet de 20 € per pagar. Quina és la probabilitat que hi hagi d’afegir diners?

Solució

a)

	Probabilitat
més de 2,1 quilograms	0,1587
menys de 2,1 quilograms	0,8413
més d'1,9 quilograms	0,8413
menys d'1,8 quilograms	0,0228
entre 1,8 i 2,1 quilograms	0,8186

b) 0,1241

c) 0,0745

d) 0,1395

Mona de Pasqua



Mona de Pasqua (fotografia retallada), de Wamito, sota llicència CC1.0.

Un pastisser elabora mones de 20 cm de diàmetre. El pes de les mones de mantega sense guarniments segueix aproximadament una distribució normal amb una mitjana de 900 grams i una variància 2500 grams². Per decorar-les, posa 3 ous de xocolata petits amb pollets al damunt de la mona. El pes de cada ou segueix aproximadament una distribució normal amb una mitjana igual a 40 grams i una desviació estàndard igual a 6 grams.

- a) El pastisser ven les mones a 28 €/quilograms. Trobeu la probabilitat que una mona valgui més de 30 euros.

El pastisser té un ajudant, l'Enric, que sap fer la millor de les cremes angleses però que és una mica sapastre. Una mitjana d'un cop cada dues setmanes a l'Enric li cau un pastís a terra. Mitjançant una distribució de Poisson, calculeu:

- b) Probabilitat que a l'Enric li caigui com a màxim un pastís en dues setmanes.
 c) Probabilitat que li caiguin exactament dos pastissos en una setmana.
 d) Probabilitat que passin com a mínim quatre setmanes sense que li caigui cap pastís.
 e) Probabilitat que passin com a mínim quatre setmanes sense que li caigui cap pastís, sabent que ja n'ha passat una.
 f) De mitjana, cada quantes setmanes li cau un pastís?

Resolució de la pregunta a

Sigui X el pes en grams d'una mona decorada amb ous. El pes de la mona decorada és igual al pes d'una mona sense decorar més el pes de cada un dels tres ous, que poden ser diferents.

$$X = \text{pes mona} + \text{pes } \text{🍪}_1 + \text{pes } \text{🍪}_2 + \text{pes } \text{🍪}_3$$

Atès que tant el pes de la mona sense decorar com el pes de cada ou segueix una distribució normal, la suma també segueix una distribució normal (propietat additiva de la distribució normal). S'assumeix que els pesos del diferents elements són independents entre ells.

La suma de distribucions normals independents entre elles segueix una distribució normal. La mitjana de la nova variable és igual a la suma de les mitjanes i la variància de la nova variable és igual a la suma de les variàncies.

$$X \sim N(900 + 3 \cdot 40 = 1020; 2500 + 3 \cdot 36 = 2608)$$

El preu P d'una mona decorada és igual al pes en grams multiplicat per 0,028 € (o igual al pes en quilograms multiplicat per 28 €).

$$P = 0,028 \cdot X$$

Això és un canvi d'escala de la variable; la distribució de P continua essent normal, tot i que la mitjana i la variància hauran canviat:

$$P \sim N(\mu = 0,028 \cdot 1020; \sigma^2 = 0,028^2 \cdot 2608)$$

$$P \sim N(\mu = 28,56; \sigma^2 = 2,045)$$

La probabilitat demanada és: $P(P > 30) = \left(Z > \frac{30-28,56}{1,43}\right) = 0,157$

La probabilitat que una mona valgui més de 30 € és igual a 0,157.

Resolució de la pregunta b

Sigui X_b el nombre de pastissos que cauen en dues setmanes. X_b segueix una distribució de Poisson de paràmetre λ_b , que és igual al nombre mitjà de pastissos que cauen en dues setmanes, que és 1:

$$X_b \sim P(\lambda_b = 1)$$

La funció de quantia de la distribució de Poisson és $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$ per a tot $x \in R_X$.

La probabilitat que en dues setmanes li caigui com a màxim 1 pastís cal calcular-la com la suma de la probabilitat que no en caigui cap més la probabilitat que en caigui exactament 1, atès que la funció de distribució acumulada de Poisson no té una formulació sintètica i compacta.

$$P(X_b \leq 1) = P(X_b = 0) + P(X_b = 1) = \frac{e^{-1} \cdot 1^0}{0!} + \frac{e^{-1} \cdot 1^1}{1!} = 0,7358$$

La probabilitat que en dues setmanes a l'Enric li caigui com a màxim 1 pastís és de 0,7358.

Resolució de la pregunta c

Per respondre a aquesta pregunta cal crear la variable “nombre de pastissos que cauen en 1 setmana”. Aquesta variable segueix una distribució de Poisson, atès que la distribució de Poisson és infinitament divisible. El paràmetre λ_c de la nova variable és igual al nombre mitjà de pastissos que cauen en una setmana:

Sigui X_c el nombre de pastissos que cauen en 1 setmana. X_c segueix una distribució de Poisson de paràmetre $\lambda_c = 0,5$:

$$X_c \sim P(\lambda_c = 0,5)$$

Probabilitat que caiguin dos pastissos en una setmana: $P(X_c = 2) = \frac{e^{-0,5} \cdot 0,5^2}{2!} = 0,0758$.

La probabilitat que a l'Enric li caiguin dos pastissos a terra en una setmana és igual a 0,0758.

Resolució de la pregunta d

Per respondre a aquesta pregunta es pot crear una variable que descriu el nombre de pastissos que han caigut en 4 setmanes i, amb aquesta variable, calcular la probabilitat que en 4 setmanes no en caigui cap (**opció a**). Altrament, es pot crear una variable que modelitzi el temps que ha transcorregut, en setmanes per exemple, entre la caiguda consecutiva de dos pastissos i, amb aquesta variable, calcular la probabilitat que aquest temps sigui com a mínim de 4 setmanes (**opció b**).

El lapse de temps que transcorre entre dos esdeveniments consecutius que segueixen una distribució de Poisson de paràmetre λ segueix una distribució exponencial del mateix paràmetre λ . La unitat de mesura entre les dues distribucions ha de ser la mateixa (en setmanes, per exemple).

Opció a

Sigui X_d el nombre de pastissos que cauen a terra en 4 setmanes. X_d segueix una distribució de Poisson per la propietat additiva de la distribució de Poisson. El paràmetre λ_d de la nova distribució es correspon amb el nombre mitjà de pastissos que cauen en 4 setmanes.

$$X_d \sim P(\lambda_d = 2)$$

Per trobar la probabilitat que passin com a mínim 4 setmanes sense que li caigui cap pastís, cal calcular la probabilitat que la nova variable prengui un valor 0: $P(X_d = 0) = \frac{e^{-2} \cdot 2^0}{0!} = 0,1353$.

La probabilitat que passin com a mínim 4 setmanes sense que a l'Enric li caigui cap pastís a terra és igual a 0,1353.

Opció b

Sigui Y_{setmana} el temps en setmanes entre la caiguda consecutiva de dos pastissos. Y_{setmana} segueix una distribució exponencial de paràmetre igual a la mitjana del nombre de pastissos que cauen per setmana ($\lambda_c = 0,5$).

$$Y_{\text{setmana}} \sim \exp(\lambda_c = 0,5)$$

La funció de distribució acumulada de l'exponencial és $P(X \leq x) = 1 - e^{-\lambda \cdot x}$ per a tot $x \in \mathbb{R}_+$.

La probabilitat que passin com a mínim 4 setmanes sense que li caigui cap pastís es pot calcular a partir del complementari:

$$P(Y_{\text{setmana}} > 4) = 1 - P(Y_{\text{setmana}} \leq 4) = 1 - (1 - e^{-0,5 \cdot 4}) = e^{-2} = 0,1353$$

La probabilitat que passin com a mínim 4 setmanes sense que a l'Enric li caigui cap pastís és igual a 0,1353.

Resolució de la pregunta e

En aquest cas es pregunta sobre una probabilitat condicionada. Es demana la probabilitat que passin com a mínim quatre setmanes sense que li caigui cap pastís condicionada pel fet que ja ha passat una setmana: $P(Y_{\text{setmana}} > 4 \mid Y_{\text{setmana}} > 1)$

La funció de distribució exponencial té una propietat anomenada "manca de memòria". Aquesta propietat estableix que $P(Y_{\text{setmana}} > b \mid Y_{\text{setmana}} > a) = P(Y_{\text{setmana}} > b - a)$ per a $b > a$.

S'oblida doncs la setmana que ja ha transcorregut i es calcula directament la probabilitat de les setmanes que falten per passar.

$$\begin{aligned} P(Y_{\text{setmana}} > 4 \mid Y_{\text{setmana}} > 1) &= P(Y_{\text{setmana}} > 4 - 1) = P(Y_{\text{setmana}} > 3) \\ &= 1 - (1 - e^{-0,5 \cdot 3}) = e^{-1,5} = 0,2231 \end{aligned}$$

La probabilitat que passin com a mínim quatre setmanes sense que li caigui cap pastís sabent que ja n'ha passat una és igual a 0,2231.

Resolució de la pregunta f

L'esperança de la variable Y_{setmana} retorna el nombre mitjà de setmanes entre la caiguda de dos pastissos.

L'esperança d'una variable que segueix una distribució exponencial de paràmetre λ és igual a $\frac{1}{\lambda}$.

$$E[Y_{setmana}] = \frac{1}{0,5} = 2$$

Passen una mitjana de dues setmanes entre la caiguda de dos pastissos.

Comentaris

En la segona part d'aquest enunciat s'hi han aplicat dues propietats de la distribució de Poisson molt útils, atès que permeten modificar el lapse de temps entre dos "èxits" pel qual està definida la distribució de Poisson inicial.

També s'ha exemplificat la propietat "manca de memòria" de la distribució exponencial. Aquesta propietat estableix que si es busca la probabilitat que un "èxit" es produeixi abans d'un moment fixat i ja ha passat un cert temps des de l'inici del còmput del temps, aleshores es pot reduir el lapse de temps i calcular només la probabilitat del temps que manca.

Peres per sopar



The fruits of America: containing richly colored figures, and full description of all the choicest varieties cultivated in the United States, de Charles Mason Hovey (1848), sota llicència CC1.0.

En Pau ha enviat d'urgència el seu fill Joan a la fruiteria a comprar 10 peres per sopar. En Joan, un adolescent al qual tant li és tot, agafa de mala gana les 10 peres d'una pila, sense ni mirar-se-les. El 30 % de les peres de la pila són massa verdes per poder-les menjar. Les peres es venen a 2,6 euros el quilogram. El pes d'una pera de les que ja estan seleccionades per vendre-les segueix aproximadament una distribució normal amb una mitjana igual a 110 grams i una desviació estàndard igual a 11 grams.

- a) Trobeu la probabilitat que exactament 4 peres siguin madures.
- b) Trobeu la probabilitat que com a màxim 4 peres siguin verdes.
- c) Quina quantitat mitjana de peres es pot esperar que siguin madures?
- d) En Pau li ha donat 3 euros al seu fill. Trobeu la probabilitat que en Joan no porti prou diners per pagar les peres.

Resolució de les preguntes a, b i c

Sigui X_M el nombre de peres madures entre les 10 peres que agafa en Joan. En Joan agafa primer 1 pera, la probabilitat que sigui verda és de 0,3, atès que el 30 % són verdes; de la mateixa manera, la probabilitat que sigui madura és de 0,7. En Joan repeteix aquesta operació 10 cops. Es considera (1) que el fet que agafi una pera verda/madura no condiciona com serà la següent i (2) que el nombre de peres de la pila és tan gran que agafar-ne una no afecta la probabilitat que la pera que s'ha agafat després sigui verda o madura. Així, la selecció de peres per part d'en Joan es pot veure com una extracció de 10 peres independents entre elles amb una probabilitat d'èxit (pera madura) sempre igual a 0,7. En aquestes condicions, X es pot modelitzar mitjançant una distribució binomial.

$$X_M \sim B(n = 10, p = 0,7)$$

Resolució de la pregunta a

Probabilitat que exactament 4 peres siguin madures:

$$\begin{aligned} P(X_M = 4) &= \binom{10}{4} \cdot 0,7^4 \cdot 0,3^6 \\ &= \frac{10!}{4! \cdot 6!} \cdot 0,7^4 \cdot 0,3^6 \\ &= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{4 \cdot 3 \cdot 2 \cdot 1 \cdot 6!} \cdot 0,7^4 \cdot 0,3^6 = 210 \cdot 0,7^4 \cdot 0,3^6 = 0,0368 \end{aligned}$$

La probabilitat que exactament 4 peres siguin madures és igual a 0,0368.

Resolució de la pregunta b

Aquesta pregunta es pot respondre o bé comptant el nombre de peres madures (**opció a**) o bé comptant el nombre de peres verdes (**opció b**).

Opció a

Si com a màxim 4 peres són verdes, com a mínim 6 peres són madures.

$$\begin{aligned}
 P(X_M \geq 6) &= P(X_M = 6) + P(X_M = 7) + P(X_M = 8) + P(X_M = 9) + P(X_M = 10) \\
 &= \binom{10}{6} \cdot 0,7^6 \cdot 0,3^4 && \text{on } \binom{10}{6} = \frac{10!}{6! \cdot 4!} = 210 \\
 &+ \binom{10}{7} \cdot 0,7^7 \cdot 0,3^3 && \text{on } \binom{10}{7} = \frac{10!}{7! \cdot 3!} = 120 \\
 &+ \binom{10}{8} \cdot 0,7^8 \cdot 0,3^2 && \text{on } \binom{10}{8} = \frac{10!}{8! \cdot 2!} = 45 \\
 &+ \binom{10}{9} \cdot 0,7^9 \cdot 0,3^1 && \text{on } \binom{10}{9} = \frac{10!}{9! \cdot 1!} = 10 \\
 &+ \binom{10}{10} \cdot 0,7^{10} \cdot 0,3^0 && \text{on } \binom{10}{10} = \frac{10!}{10! \cdot 0!} = 1 \\
 &= 0,8497
 \end{aligned}$$

Opció b

Sigui X_V el nombre de peres verdes que agafa en Joan.

$$X_V \sim B(n=10, p=0,3)$$

Probabilitat que com a màxim 4 peres siguin verdes:

$$\begin{aligned}
 P(X_V \leq 4) &= P(X_V = 0) + P(X_V = 1) + P(X_V = 2) + P(X_V = 3) + P(X_V = 4) \\
 &= \binom{10}{0} \cdot 0,3^0 \cdot 0,7^{10} && \text{on } \binom{10}{0} = \frac{10!}{0! \cdot 10!} = 1 \\
 &+ \binom{10}{1} \cdot 0,3^1 \cdot 0,7^9 && \text{on } \binom{10}{1} = \frac{10!}{1! \cdot 9!} = 10 \\
 &+ \binom{10}{2} \cdot 0,3^2 \cdot 0,7^8 && \text{on } \binom{10}{2} = \frac{10!}{2! \cdot 8!} = 45 \\
 &+ \binom{10}{3} \cdot 0,3^3 \cdot 0,7^7 && \text{on } \binom{10}{3} = \frac{10!}{3! \cdot 7!} = 120 \\
 &+ \binom{10}{4} \cdot 0,3^4 \cdot 0,7^6 && \text{on } \binom{10}{4} = \frac{10!}{4! \cdot 6!} = 210 \\
 &= 0,8497
 \end{aligned}$$

La probabilitat que com a màxim 4 peres siguin verdes és igual a 0,8497.

Resolució de la pregunta c

L'esperança de la binomial és $n \cdot p$. Atès que s'estan comptant les peres madures, es treballa amb X_M .

$$E[X_M] = 10 \cdot 0,7 = 7 \text{ peres}$$

S'espera que una mitjana de 7 peres siguin madures.

Resolució de la pregunta d

Sigui G el pes d'una pera expressat en grams. Tal com s'indica a l'enunciat de l'exercici:

$$G \sim N(\mu_G = 110; \sigma_G^2 = 11^2 = 121)$$

Per resoldre aquest problema es pot construir primer una variable que modelitzi el pes de 10 peres, després construir una segona variable que sigui el preu d'un paquet amb 10 peres i, finalment, calcular la probabilitat que el preu d'un paquet de 10 peres superi els 3 euros (**opció a**). Una altra opció és construir primer una variable que sigui el preu d'una pera i, a partir d'aquesta, construir la segona variable, que és el preu de 10 peres per calcular la probabilitat que el preu d'un paquet de 10 peres superi els 3 euros (**opció b**).

Opció a

Sigui BG el pes d'una bossa de 10 peres en grams:

$$BG = G_1 + G_2 + \dots + G_9 + G_{10}$$

on G_1 és el pes en grams de la primera pera que el Joan agafa a l'atzar, G_2 és el pes en grams de la segona pera que el Joan agafa a l'atzar i així fins a la darrera. Atès que BG és la suma de 10 normals, per la propietat additiva de la distribució normal es pot deduir que BG segueix una distribució normal amb una mitjana igual a la suma d'esperances i, en el supòsit que el pes d'una pera és independent del pes de les altres peres, una variància igual a la suma de variàncies.

$$BG \sim N(\mu_{BG} = 1100; \sigma_{BG}^2 = 1210)$$

El preu de les peres és de 2,6 €/quilograms. Atès que tenim una variable expressada en grams i el preu s'expressa en quilograms, hi ha dues alternatives: o bé s'expressa el preu en euros per gram (alternativa i) o bé es construeix una variable que sigui el pes de 10 peres en quilograms (alternativa ii). La primera alternativa és més senzilla; tanmateix, es desenvolupen les dues.

Alternativa i:

El preu de les peres en euros per gram és:

$$\frac{2,6}{1000} = 0,0026 \text{ €/g}$$

Sigui BP el preu d'una bossa de 10 peres en euros. El preu és el resultat de multiplicar el pes pel preu, de manera que es pot construir BP multiplicant el pes en grams pel preu en euros/gram.

$$BP = 0,0026 \cdot BG$$

BP segueix una distribució normal, atès que per construir BP s'aplica un canvi d'escala a BG . L'esperança de BP és igual a l'esperança de BG multiplicada per la constant (0,0026) i la variància de BP és igual a la variància de BG multiplicada per la constant al quadrat ($0,0026^2$).

$$BP \sim N(\mu_{BP} = 0,0026 \cdot 1100; \sigma_{BP}^2 = 0,0026^2 \cdot 1210)$$

$$BP \sim N(\mu_{BP} = 2,86; \sigma_{BP}^2 = 0,0081796)$$

Un cop coneguda la distribució de BP , la probabilitat que en Joan no porti prou diners és igual a la probabilitat que el preu de les 10 peres sigui superior a 3 euros:

$$P(BP > 3) = 0,0608$$

Alternativa ii:

Sigui K el pes d'una pera expressada en quilograms.

$$K = \frac{G}{1000}$$

Es tracta d'un canvi d'escala a G .

$$K \sim N\left(\mu_K = \frac{110}{1000} = 0,11; \sigma_K^2 = \frac{121}{1000^2} = 0,000121\right)$$

A partir d'aquí es construeix BK , que és el pes en quilograms d'una bossa de 10 peres:

$$BK = K_1 + K_2 + \dots + K_{10}$$

on K_1 és el pes en quilograms de la primera pera que en Joan agafa a l'atzar, K_2 és el pes en quilograms de la segona pera que el Joan agafa a l'atzar i així fins a la darrera pera. Es torna a tractar d'una suma de 10 normals independents entre elles i, per tant, per la propietat additiva de la normal:

$$BK \sim N(\mu_{BK} = 1,1; \sigma_{BK}^2 = 0,00121)$$

A partir d'aquesta darrera variable, el preu de 10 peres en euros (BP) es calcula com el pes en quilograms pel preu en quilograms.

$$BP = 2,6 \cdot BK$$

BP segueix una distribució normal, atès que per construir BP s'aplica un canvi d'escala a BK . L'esperança de BP és igual a l'esperança de BK multiplicada per la cons-

tant (2,6) i la variància de BP és igual a la variància de BG multiplicada per la constant al quadrat ($2,6^2$).

$$BP \sim N(\mu_{BP} = 2,6 \cdot 1,1; \sigma_{BP}^2 = 2,6^2 \cdot 0,00121)$$

$$BP \sim N(\mu_{BP} = 2,86; \sigma_{BP}^2 = 0,0081796)$$

Lògicament, s'obté la mateixa distribució. A partir d'aquí, com abans, resta calcular:

$$P(BP > 3) = 0,0608$$

Opció b

En aquest cas, primer es descriu el preu en euros d'una sola pera. Sigui P el preu en euros d'una pera; aquest preu és el resultat de multiplicar el preu (una constant) pel pes (una variable aleatòria que segueix una distribució normal). Es tracta, doncs, d'un canvi d'escala aplicat a una distribució normal. Atès que es disposa del preu en euros i del pes en grams, com abans o bé es passa tot a grams o bé es passa tot a quilograms.

Alternativa i: $P = 0,0026 \cdot G$

Alternativa ii: $P = 2,6 \cdot K$

En tots dos casos s'arriba a:

$$P \sim N(\mu_P = 0,286; \sigma_P^2 = 0,00081796)$$

Ara cal construir la variable BP com la suma del cost de cada pera:

$$BP = P_1 + P_2 + \dots + P_{10},$$

on P_1 és el preu en euros de la primera pera que agafa a l'atzar, P_2 és el preu en euros de la segona pera que el Joan agafa a l'atzar i així fins a la darrera. Atès que BP és la suma de 10 normals independents entre elles, per la propietat additiva de la normal es pot deduir que BP segueix una distribució normal amb una mitjana igual a la suma d'esperances i una variància igual a la suma de variàncies.

$$BP \sim N(\mu_{BP} = 2,86; \sigma_{BP}^2 = 0,0081796)$$

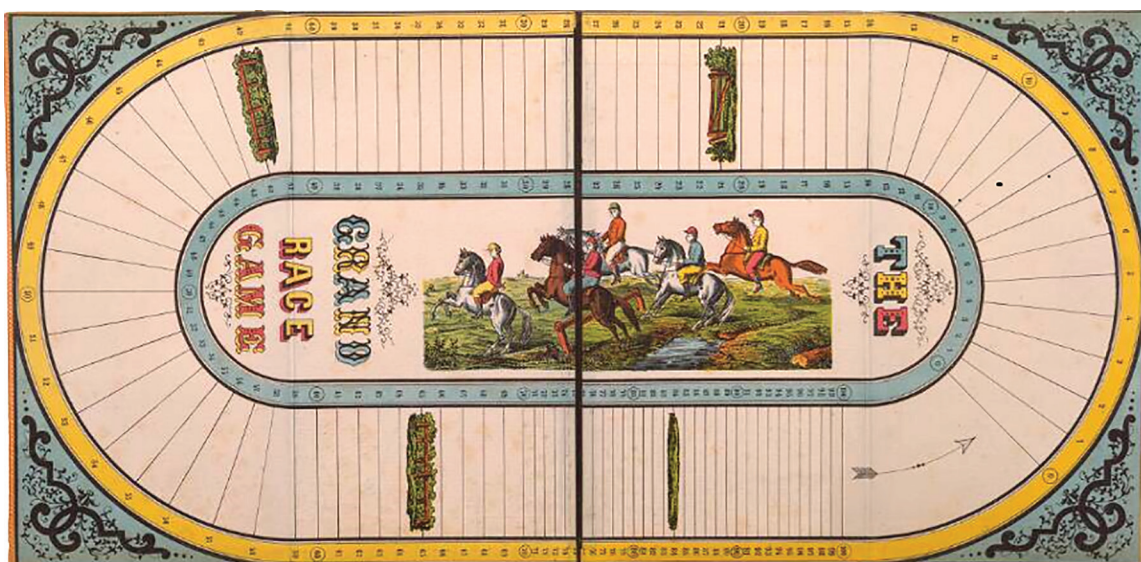
A partir d'aquí, com abans, resta calcular $P(BP > 3) = 0,0608$.

La probabilitat que en Joan no porti prou diners per pagar les peres és de 0,0608.

Comentaris

En aquest exercici s'exemplifica (apartat b) com en una binomial, l'esdeveniment que s'escull com a èxit o fracàs no és rellevant, sinó la formulació que es fa del problema en base a aquesta elecció. També s'exemplifica com en la construcció d'una variable es poden prendre diferents camins (apartat d).

Joc de curses de cavalls



The Grand Race Game (1890), de la Biblioteca Bodleiana de l'Oxford University, sota llicència CC BY 4.0.

En un joc de taula de curses de cavalls, en una partida cada jugador (una cursa) tira el dau 35 vegades. La meta és a la casella 120, quan els cavalls arriben a la meta no s'aturen i els jugadors continuen tirant el dau fins a esgotar les 35 tirades (la suma de les 35 tirades és el que finalment avança el cavall d'un jugador).¹ D'altra banda, s'estima que un jugador triga una mitjana de 24 segons a fer una jugada (tirada i moviment), amb una desviació estàndard de 6 segons.

- a) Sabent que la meta és a la casella 120, trobeu la probabilitat que un cavall no arribi a la meta.

¹ Aquest reglament, molt simple, no correspon al de *The Grand Race Game*.

- b) Trobeu la probabilitat que tots els cavalls arribin a la meta en una partida amb 4 jugadors.
- c) De mitjana, quants cavalls arriben a la meta en una partida de 4 jugadors?
- d) Trobeu la probabilitat que una partida de 4 jugadors duri més d'una hora.

Resolució de l'apartat a

Sigui W el nombre de caselles que avança el cavall d'un jugador qualsevol en una partida. Per respondre a la pregunta, cal trobar primer la distribució de probabilitat de la variable W . Amb aquesta fi, es busca (1) la distribució de probabilitat del que avança un cavall qualsevol en una tirada (anomenem X aquesta variable) i (2) la distribució de probabilitat del que avança un cavall qualsevol en 35 tirades (variable W). Quan es disposa de la distribució de W , es busca la probabilitat que la variable W sigui inferior a 120.

Pas 1: distribució i característiques de la variable *nombre de caselles que avança un cavall en una tirada*

Sigui X el nombre de caselles que avança un jugador en una tirada.

X és una variable aleatòria discreta que pot prendre qualsevol dels valors enters entre 1 i 6:

$$\text{Rang de } X: R_X = \{1, 2, 3, 4, 5, 6\}$$

Es considera que és un dau equilibrat (no trucat) i, per tant, que tots els valors tenen la mateixa probabilitat d'ocurrència (equiprobabilitat). Aquestes característiques (nombre de valors del rang finit i equiprobables) són les que corresponen a una distribució uniforme discreta:

$$X \sim U[1; 6]$$

Una variable aleatòria discreta segueix una distribució uniforme discreta si tots els valors que pertanyen al rang —que és numerable— tenen la mateixa probabilitat d'ocurrència.

La funció de quantia de X és:

$$f_q(X) = P(X = x) = \begin{cases} \frac{1}{6} & \forall x \in R_X \\ 0 & \forall x \notin R_X \end{cases}$$

L'esperança de X és:

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5 \text{ caselles}$$

La variància de X és: $Var[X] = E[X^2] - E[X]^2$

$$E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15,167 \text{ caselles}^2$$

$$Var[X] = 15,167 - 3,5^2 = 2,9167 \text{ caselles}^2$$

La desviació estàndard de X és:

$$DE[X] = \sqrt{Var[X]} = 1,708 \text{ caselles}$$

Pas 2: distribució i característiques de la variable *nombre de caselles que avança un cavall en 35 tirades*

Sigui W el nombre total de caselles que avança un cavall en una partida (la suma de les 35 tirades). W és la suma del resultat de cada una de les 35 tirades; els resultats de les tirades poden ser diferents i són independents entre ells (el resultat d'una tirada no condiciona el d'una altra tirada).

$$W = X_1 + X_2 + X_3 + \dots + X_{35}$$

És possible deduir que W segueix aproximadament una distribució normal gràcies al teorema central del límit (TCL), atès que el nombre de tirades es considera prou gran i les tirades són independents i idènticament distribuïdes.

$$W \approx N(\mu_W; \sigma_W^2)$$

El teorema central del límit estableix que la suma de n variables aleatòries que tenen esperança i variància, que són independents entre elles i que estan idènticament distribuïdes tendeix cap a una distribució normal quan n tendeix cap a l'infinit.

Quan es té una suma de n variables aleatòries independents i idènticament distribuïdes amb esperança i variància i n és prou gran (es considera prou gran si $n > 30$), la distribució de la variable suma es pot aproximar mitjançant una distribució normal. L'esperança i la variància de la variable nova seran, respectivament, la suma d'esperances i la suma de variàncies; és a dir, n vegades l'esperança de la variable original i n vegades la variància de la variable original.

$$\begin{aligned}E[W] &= E[X_1 + X_2 + \dots + X_{35}] \\&= E[X_1] + E[X_2] + \dots + E[X_{35}] \\&= 35 \cdot E[X_i] = 35 \cdot 3,5 = 122,5 \text{ caselles}\end{aligned}$$

$$\begin{aligned}\text{Var}[W] &= \text{Var}[X_1 + X_2 + \dots + X_{35}] \\&= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_{35}] \\&= 35 \cdot \text{Var}[X_i] = 35 \cdot 2,9167 \\&= 102,0833 \text{ caselles}^2\end{aligned}$$

Finalment:

$$W \approx N(\mu_W = 122,5; \sigma_W^2 = 102,0833)$$

Nota: el símbol $\approx$ indica que segueix **aproximadament** la distribució.

Pas 3: càlcul de la probabilitat que un cavall no arribi a la meta

Atès que la meta és a la casella 120, un cavall no arriba a la meta quan després de 35 tirades no avança com a mínim 120 caselles.

$$P(W < 120) = P\left(Z < \frac{120 - 122,5}{\sqrt{102,08}}\right) = P(Z < -0,25) = 0,4023$$

La probabilitat que un cavall qualsevol no arribi a la meta és aproximadament igual a 0,4023.

Resolució de l'apartat b

Com s'ha vist, la probabilitat que un cavall qualsevol no arribi a la meta és igual a 0,4023. Per tant, el seu complementari $1 - 0,4023 = 0,5977$ és la probabilitat que un cavall qualsevol pel seu compte arribi a la meta.

En una partida hi ha 4 cavalls, el fet que un cavall arribi a la meta és independent del que facin la resta de cavalls, atès que el que avança cada cavall depèn de les seves pròpies tirades de daus. D'altra banda, tots tenen la mateixa probabilitat $p = 0,5977$ d'arribar a la meta.

Sigui C el nombre de cavalls que arriben a la meta en una partida amb 4 jugadors. C es pot modelitzar mitjançant una distribució binomial, atès que es tracta d'un comptatge finit d'èxits independents entre ells i amb una probabilitat constant.

$$C \sim B(n = 4, p = 0,5977)$$

L'esdeveniment "tots 4 cavalls arriben a la meta" es correspon amb el valor 4 de la variable. Per tant, la probabilitat d'aquest esdeveniment es correspon amb $P(C = 4)$.

$$P(C = 4) = \binom{4}{0} \cdot 0,5977^4 \cdot 0,4023^0 = 0,5977^4 = 0,1276$$

La probabilitat que tots 4 cavalls arribin a la meta és igual a 0,1276.

Resolució de l'apartat c

El nombre mitjà de cavalls que arriben a la meta en una partida qualsevol es correspon amb l'esperança de la variable C .

En una distribució binomial de paràmetres n i p , l'esperança és igual a n per la probabilitat d'èxit (p).

$$E[C] = n \cdot p = 4 \cdot 0,5977 = 2,3908$$

Arriben a la meta una mitjana de 2,3908 cavalls.

Resolució de l'apartat d

Sigui Y el temps en segons que triga un jugador qualsevol a fer una jugada. L'esperança de Y és de 24 segons i la variància, de 36 segons². Es desconeix la distribució de Y .

Sigui T el temps en segons que es triga a completar una partida de 4 jugadors. T és el temps que es triga a fer 140 jugades (4 jugadors per 35 jugades cada un).

$$T = Y_1 + Y_2 + \dots + Y_{140}$$

Si s'accepta que el temps que es triga a fer una jugada és independent del temps que es triga a fer la resta de jugades, aleshores és possible deduir la distribució aproximada de T gràcies, com en l'apartat a, al teorema central del límit.

$$T \approx N(\mu_T = 3360; \sigma_T^2 = 5040)$$

Atès que:

$$E[T] = 24 + 24 + \dots + 24 = 140 \cdot 24 = 3360 \text{ segons}$$

$$\text{Var}[T] = 36 + 36 + \dots + 36 = 140 \cdot 36 = 5040 \text{ segons}^2$$

$$DE[T] = \sqrt{5040} = 70,99 \text{ segons}$$

Si una partida dura més d'una hora, dura més de 3600 segons.

$$P(T > 3600) = P\left(Z > \frac{3600 - 3360}{70,99}\right) = P(Z > 3,38) = 0,0004$$

La probabilitat que una partida de 4 jugadors duri més d'una hora és molt baixa i aproximadament igual a 0,0004.

Comentaris

Aquest exercici exemplifica l'aplicació del teorema central del límit (apartats a i d). En els dos apartats es treballa amb una suma de variables aleatòries. En l'apartat a, la distribució d'origen és coneguda (una uniforme discreta); en l'apartat d, és desconeguda. Això no té cap incidència en l'aplicació del teorema.

Diferències entre aquest exercici i l'exercici "Avets a la fira de Santa Llúcia": En l'exercici "Avets a la fira de Santa Llúcia", també s'hi feien sumes de variables aleatòries, però s'hi aplicava la propietat additiva de la normal.

- La distribució inicial era coneguda i seguia una distribució normal.
- No era important el nombre de variables que se sumessin (se'n podien sumar 2 o moltes més).
- Les variables podien tenir distribucions amb paràmetres (μ i σ^2) diferents entre elles.
- El resultat no era una aproximació, era una normal perfecta.

Exercici equivalent al del "Joc de curses de cavalls"

En un joc de taula sobre viatges de marciàns (Υ) que van cap a Venus, en una partida (un viatge) cada jugador tira el dau 40 vegades, la suma de les 40 tirades és el que finalment avança el marcià d'un jugador. S'estima que un jugador triga una mitjana de 22 segons a fer una jugada (tirada i moviment), amb una desviació estàndard de 5 segons.

- a) Venus és a la casella 150. Trobeu la probabilitat que un marcià no arribi a Venus.
- b) Trobeu la probabilitat que cap dels marciàns no arribi a Venus en una partida amb 4 jugadors.
- c) De mitjana, quants marciàns arriben a Venus en una partida de 4 jugadors?
- d) Trobeu la probabilitat que una partida de 4 jugadors duri més d'una hora.

Solució:

a) $W \approx N(\mu_W = 140; \sigma_W^2 = 116,668)$ $P(W < 150) = 0,823$

b) $M \sim B(n = 4, p = 0,177)$ $P(M = 0) = 0,458$

c) $E[M] = 4 \cdot 0,177 = 0,709$

d) $T \approx N(\mu_T = 3520; \sigma_T^2 = 4000)$ $P(T > 3600) = 0,103$

Videojocs

Dos empresas lanzan a principios de año un nuevo videojuego cada una —VX, VY—. El departamento de juguetería de unos grandes almacenes comprueba que el número de unidades vendidas por semana se distribuye uniformemente:

para el videojuego $X = U [20, 30]$

para el videojuego $Y = U [25, 35]$

a) Hallar una buena aproximación de la distribución de las siguientes variables:

n.º de unidades vendidas por año de VX

n.º de unidades vendidas por año de VY

Año = 52 semanas

Cada videojuego X tiene un PVP* = 2000 ptas.

Cada videojuego Y tiene un PVP = 1500 ptas.

b) Hallar una buena aproximación de la distribución de la variable: Ingresos anuales por venta de los dos tipos de videojuegos.

c) Si el margen bruto en la venta de los videojuegos es igual al 20 % del precio de venta, ¿cuál es la probabilidad de que este margen supere 1000000 ptas. al cabo del año?

* PVP: precio de venta al público

Dues empreses treuen a principis d'any un nou videojoc cadascuna —VX, VY. El departament de joguines d'uns grans magatzems comprova que el nombre d'unitats venudes per setmana es distribueix uniformement:

per al videojoc $X = U [20, 30]$

per al videojoc $Y = U [25, 35]$

- a) Trobeu una bona aproximació de la distribució de les variables següents:
 nombre d'unitats venudes per any de VX.
 nombre d'unitats venudes per any de VY.
 Any = 52 setmanes
 Cada videojoc X té un PVP* = 2000 ptes.
 Cada videojoc Y té un PVP* = 1500 ptes.
- b) Trobeu una bona aproximació de la distribució de la variable: ingressos anuals per venda dels dos tipus de videojocs.
- c) Si el marge brut en la venda dels videojocs és igual al 20 % del preu de venda, quina probabilitat hi ha que aquest marge superi 1000000 ptes. al cap de l'any?
- *PVP: preu de venda al públic

Resolució de l'apartat a

Sigui X_i el nombre de videojocs de tipus X que es venen a la setmana i sigui XA el nombre de videojocs de tipus X que es venen en un any.

$$XA = X_1 + X_2 + \dots + X_{52}$$

Paral·lelament, sigui Y_i el nombre de videojocs de tipus Y que es venen a la setmana i sigui YA el nombre de videojocs de tipus Y que es venen en un any.

$$YA = Y_1 + Y_2 + \dots + Y_{52}$$

Se suposa que el volum de venda és independent entre setmanes.

La suma de n variables aleatòries independents entre elles i idènticament distribuïdes amb la mateixa esperança i variància tendeix cap a una distribució normal amb una esperança igual a la suma d'esperances i una variància igual a la suma de variàncies quan n tendeix cap a l'infinit (teorema central del límit, TCL). Atès que $n = 52$, es considera prou gran per aplicar-hi el TCL. Així, XA i YA es poden aproximar cada una a una distribució normal.

$$XA \approx N(\mu_{XA} = 52 \cdot \mu_X; \sigma_{XA}^2 = 52 \cdot \sigma_X^2)$$

$$YA \approx N(\mu_{YA} = 52 \cdot \mu_Y; \sigma_{YA}^2 = 52 \cdot \sigma_Y^2)$$

D'altra banda, X és una variable aleatòria discreta que segueix una distribució uniforme entre 20 i 30. La taula de probabilitat següent descriu els valors que pren la variable X i la seva funció de quantia.

X	20	21	22	23	24	25	26	27	28	29	30
$P(X=x)$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$	$1/11$

$$E[X] = 20 \cdot 1/11 + 21 \cdot 1/11 + \dots + 30 \cdot 1/11 = 25$$

$$Var[X] = E[X^2] - E[X]^2$$

$$E[X^2] = 20^2 \cdot 1/11 + 21^2 \cdot 1/11 + \dots + 30^2 \cdot 1/11 = 635$$

$$Var[X] = 635 - 25^2 = 10$$

Finalment, $XA \approx N(\mu_{XA} = 52 \cdot 25 = 1300; \sigma_{XA}^2 = 52 \cdot 10 = 520)$.

De la mateixa manera, Y és una variable aleatòria discreta que segueix una distribució uniforme entre 25 i 35. En concret, la variable Y és idèntica a la variable X , però desplaçada 5 unitats:

$$Y = X + 5$$

Per tant:

$$E[Y] = E[X] + 5 = 30$$

$$Var[Y] = Var[X] = 10$$

Finalment, $YA \approx N(\mu_{YA} = 52 \cdot 30 = 1560; \sigma_{YA}^2 = 52 \cdot 10 = 520)$

Resolució de l'apartat b

Sigui IX els ingressos anuals pel nombre de videojocs de tipus X que s'han venut per 2000 pessetes cada un.

$$IX = 2000 \cdot XA$$

Es tracta d'un canvi d'escala de la variable XA .

Per tant:

$$E[IX] = 2000 \cdot E[XA] = 2600000 \text{ ptes}$$

$$Var[IX] = 2000^2 \cdot 520 = 2080000000 \text{ ptes}^2$$

$$IX \approx N(\mu_{IX} = 2600000; \sigma_{IX}^2 = 2080000000)$$

De la mateixa manera, sigui IY els ingressos anuals pel nombre de videojocs de tipus Y que s'ha venut per 1500 ptes cada un.

$$IY = 1500 \cdot YA$$

$$E[IY] = 1500 \cdot E[YA] = 2340000 \text{ ptes}$$

$$\text{Var}[IX] = 1500^2 \cdot 520 = 1170000000 \text{ ptes}^2$$

$$IY \approx N(\mu_{IY} = 2\,340\,000; \sigma_{IY}^2 = 1170000000)$$

Sigui IT els ingressos totals anuals en pessetes per la venda de videojocs de qual-sevol dels dos tipus. Aquests ingressos són la suma dels ingressos anuals per les vendes del videojoc de tipus X i dels ingressos anuals per les vendes del videojoc de tipus Y .

$$IT = IX + IY$$

Atès que es tracta de la suma de dues distribucions aproximadament normals, IT també segueix una distribució aproximadament normal (propietat additiva de la distribució normal). L'esperança de la nova variable serà igual a la suma d'esperances ($2600000 + 2340000 = 4940000$) i la variància de la nova variable serà igual a la suma de variàncies ($2080000000 + 1170000000 = 3250000000$), en el supòsit —que s'assumeix— que les vendes de tots dos videojocs són independents.

$$IT \approx N(\mu_{IT} = 4940000; \sigma_{IT}^2 = 3250000000)$$

Resolució de l'apartat c

Atès que el marge brut és el 20 % del preu de venda, aquest 20 % superarà el 1000000 de pessetes si els ingressos superen els 5000000 de pessetes (**opció a**). Una altra opció és construir una variable (MB) que sigui el marge brut dels ingressos i després calcular la probabilitat que el marge brut superi el milió de pessetes (**opció b**).

Opció a

Cal trobar la probabilitat que els ingressos superin 5 milions de pessetes.

$$P(IT > 5000000) = P\left(Z > \frac{5000000 - 4940000}{\sqrt{3250000000}} = 1,0525\right) = 0,1463$$

La probabilitat que el marge brut superi el milió de pessetes al cap de l'any és igual a 0,1463.

Opció b

Primer es construeix la variable MB (marge brut dels ingressos), que és el 20 % dels ingressos mitjançant un canvi d'escala.

$$MB = 0,2 \cdot IT$$

Atès que IT segueix una distribució aproximadament normal, MB també segueix una distribució aproximadament normal. L'esperança de MB és igual a l'esperança de IT per 0,2 i la variància de MB és igual a la variància de IT per $(0,2)^2$.

$$MB \approx N(\mu_{MB} = 0,2 \cdot 4940000; \sigma_{MB}^2 = 0,04 \cdot 3250000000)$$

$$MB \approx N(\mu_{MB} = 988\,000; \sigma_{MB}^2 = 130000000)$$

Ara cal trobar la probabilitat que aquest marge superi el milió de pessetes.

$$P(MB > 1000000) = P\left(Z > \frac{1000000 - 988000}{\sqrt{130000000}} = 1,0525\right) = 0,1463$$

La probabilitat que el marge brut superi el milió de pessetes al cap de l'any és igual a 0,1463.

Revetlla de Sant Joan



Nit de Sant Joan (Tarragona, 2018), de Nicolas Vigier, sota llicència CC0 1.0.

Una botiga d'alimentació ofereix per a la revetlla de Sant Joan un lot amb una coca artesana, dues ampolles de cava i una malla de peretes de Sant Joan. Les coques pesen de mitjana 600 grams (desviació estàndard igual a 120 grams). Les malles de peretes pesen de mitjana 1 quilograms (desviació estàndard igual a 0,08 quilograms). Totes les ampolles de cava pesen exactament el mateix, 1,45 quilograms.

- a) Calculeu el pes mitjà esperat d'un lot, així com la desviació estàndard d'aquest pes. Doneu el resultat en quilograms i en grams.

- b) L'encarregat de la botiga omple una estanteria sencera amb 44 lots. Trobeu la probabilitat que l'estanteria hagi d'aguantar més de 200 quilograms.
- c) Tot i que l'estanteria pot aguantar aquest pes, és recomanable no sobrepassar els 190 kilograms, a fi d'evitar que s'arquegi abans d'hora. Quin és el nombre de lots màxim que es pot posar a l'estanteria a fi que la probabilitat que els lots sobrepassin els 190 quilograms no superi el 0,05?
- d) L'advocat que té el despatx en el primer pis del mateix edifici ha decidit comprar un lot per a cada un dels seus 3 treballadors, que ell mateix puja fins al despatx. Estimeu la probabilitat que el pes dels 3 lots superi els 15 quilograms.

Resolució de la pregunta a

Sigui X_{LK} el pes d'un lot en quilograms i X_{LG} el pes d'un lot en grams.
El pes esperat d'un lot és igual a la suma de pesos mitjans.

L'esperança d'un suma de variables aleatòries és igual a la suma d'esperances.

$$E[X_{LK}] = 0,6 + 1 + 2 \cdot 1,45 = 4,5 \text{ kg}$$

$$E[X_{LG}] = 1000 \cdot E[X_{LK}] = 4500 \text{ g}$$

Per calcular la desviació estàndard d'un lot, primer se'n calcula la variància i després la desviació estàndard. En la mesura que els pesos dels elements que componen el lot siguin independents entre ells, es pot calcular la variància del pes del lot com a suma de les variàncies.

La variància d'un suma de variables aleatòries independents entre elles és igual a la suma de variàncies.

$$\text{Var}[X_{LK}] = 0,12^2 + 0,08^2 + 2^2 \cdot 0 = 0,0208 \text{ kg}^2$$

$$\text{DE}[X_{LK}] = \sqrt{0,0208} = 0,1442 \text{ kg}$$

La variància (i la desviació estàndard) del pes de les ampolles de cava és igual a 0, atès que totes són idèntiques entre elles. El pes d'una ampolla de cava es pot considerar com una constant.

La variància (i la desviació estàndard) d'una constant és nul·la.

Per passar la variància de quilos a grams, la multipliquem per mil al quadrat.

$$\text{Var}[X_{LG}] = 1000^2 \cdot 0,0208 = 20\,800 \text{ g}^2$$

$$\text{DE}[X_{LG}] = \sqrt{20800} = 144,2 \text{ g}$$

En un canvi d'unitats de mesura (canvi d'escala), la variància queda afectada pel factor de multiplicació al quadrat.

El pes mitjà esperat d'un lot és de 4,5 quilograms (4500 grams) i la desviació estàndard és de 0,1442 quilograms (144,2 grams).

Resolució de la pregunta b

Sigui X_{EK} el pes de 44 lots en quilograms. Se suposa que el pes de cada lot és independent del pes dels altres.

En aquest supòsit, mitjançant el teorema central del límit, es pot deduir que la distribució de X_{EK} segueix aproximadament una distribució normal amb una mitjana igual a la suma de mitjanes i una variància igual a la suma de variàncies.

$$X_{EK} \approx N(\mu_{X_{EK}} = 44 \cdot 4,5; \sigma_{X_{EK}}^2 = 44 \cdot 0,0208)$$

$$X_{EK} \approx N(\mu_{X_{EK}} = 198; \sigma_{X_{EK}}^2 = 0,9152)$$

Un cop trobada la distribució de la variable, només resta trobar la probabilitat que el seu valor superi els 200 quilograms.

$$P(X_{EK} > 200) = 0,0183$$

La probabilitat que el pes dels 44 lots superi els 200 quilograms és de 0,0183.

Resolució de la pregunta c

Sigui X_{nK} el pes de n lots en quilograms.

En aquesta pregunta es demana trobar n de manera que la probabilitat que el pes dels n lots superi els 190 quilograms sigui com a màxim 0,05, és a dir:

$$P(X_{nK} > 190) \leq 0,05$$

Si n és prou gran i s'assumeix que el pes de cada lot és independent del pes dels altres lots, aleshores el teorema central del límit permet deduir que la distribució de

X_{nK} segueix aproximadament una distribució normal amb una mitjana igual a la suma de mitjanes i una variància igual a la suma de variàncies.

$$X_{nK} \approx N(\mu_{X_{EK}} = n \cdot 4,5; \sigma_{X_{EK}}^2 = n \cdot 0,0208)$$

Un cop es coneix la distribució de X_{nK} , es pot tipificar: $Z_{nK} = \frac{X_{nK} - 4,5 \cdot n}{\sqrt{0,0208 \cdot n}}$; de manera que $Z_{nK} \approx N(0; 1)$.

Per tant,

$$\begin{aligned} P(X_{nK} > 190) &\leq 0,05 \\ \Rightarrow P(Z > \frac{190 - 4,5 \cdot n}{\sqrt{0,0208 \cdot n}}) &\leq 0,05 \end{aligned}$$

Atès que $P(Z > 1,645) = 0,05$

$$\begin{aligned} \Rightarrow \frac{190 - 4,5 \cdot n}{\sqrt{0,0208 \cdot n}} &= 1,645 \\ \Rightarrow 4,5 \cdot n + 1,645 \cdot \sqrt{0,0208} \cdot \sqrt{n} - 190 &= 0 \end{aligned}$$

Si $\sqrt{n} = x$

$$\Rightarrow 4,5 \cdot x^2 + 0,23725 \cdot x - 190 = 0$$

Les dues solucions d'aquesta equació són:

$$\begin{aligned} \frac{-0,23725 - \sqrt{0,23725^2 + 4 \cdot 4,5 \cdot 190}}{9} &= -6,52 \rightarrow (-6,52)^2 = 42,57 \\ \frac{-0,23725 + \sqrt{0,23725^2 + 4 \cdot 4,5 \cdot 190}}{9} &= 6,47 \rightarrow 6,47^2 = 41,88 \end{aligned}$$

La solució que es busca és un valor enter, atès que representa el nombre de lots. La condició que s'ha de complir és un màxim: pot ser inferior a 190 quilograms, però no els pot superar. L'arrodoniment s'ha de fer doncs a la baixa: 41. D'altra banda, 41 és un nombre de lots suficient perquè l'aproximació a la normal pugui ser assumida.

El nombre màxim de lots per garantir que la probabilitat de no excedir els 190 quilograms sigui com a màxim de 0,05 és 41.

Comprovació:

$n = 41$: sigui X_{41K} el pes de 41 lots en quilograms

$$X_{41K} \approx N(\mu_{41K} = 41 \cdot 4,5 = 184,5; \sigma_{X_{41K}}^2 = 41 \cdot 0,0208 = 0,8528)$$

$$P(X_{41K} > 190) = 1,29 \cdot 10^{-9} \leq 0,05$$

$n = 42$: sigui X_{42K} el pes de 42 lots en quilograms

$$X_{42K} \approx N(\mu_{42K} = 42 \cdot 4,5 = 189; \sigma_{X_{42K}}^2 = 42 \cdot 0,0208 = 0,8736)$$

$$P(X_{42K} > 190) = 0,1423 \not\leq 0,05$$

Resolució de la pregunta d

Sigui X_{3LK} el pes de 3 lots en quilograms. S'assumeix que el pes de cada lot és independent del pes de la resta de lots.

Es pot calcular el pes mitjà esperat de 3 lots i la variància a partir de la suma d'esperances i de la suma de variàncies, respectivament.

$$E[X_{3LK}] = 4,5 + 4,5 + 4,5 = 13,5 \text{ kg}$$

$$\text{Var}[X_{3LK}] = 0,0208 + \dots + 0,0208 = 0,0624 \text{ kg}^2$$

Malauradament, no es pot deduir la distribució de la variable X_{3LK} , atès que es desconeix la distribució del pes d'un lot (X_{LK}) i només se sumen tres variables. Per tant, no es pot calcular la probabilitat exacta que el lot superi els 15 quilograms. Tanmateix, atès que es coneix l'esperança i la variància del pes dels tres lots, es pot delimitar aquesta probabilitat mitjançant la desigualtat de Tchebixev.

Desigualtat de Txevixev: $P(|X - E[X]| > \varepsilon) < \frac{\text{Var}[X]}{\varepsilon^2}$
on ε és un valor tan petit com es vulgui.

Pafnuti Lvóvitx Txevixev (1821-1894)



Segell de la URSS de 1946.

De Post of USSR - <<http://kolekzioner.net/modules/smartsection/item.php?itemid=211>>, Domini públic,
<<https://commons.wikimedia.org/w/index.php?curid=12818442>>

Es busca:

$$P(X_{3LK} > 15)$$

$$\Rightarrow P(X_{3LK} - 13,5 > 15 - 13,5) \Rightarrow P(X_{3LK} - 13,5 > 1,5)$$

Sigui $\varepsilon=1,5$, la desigualtat de Txevixev permet establir que:

$$P(|X_{3LK} - 13,5| > 1,5) < \frac{0,0624}{1,5^2}$$

$$\Rightarrow P(|X_{3LK} - 13,5| > 1,5) < 0,0277$$

La desigualtat delimita la probabilitat de les dues cues; en aquest cas, o bé que el lot pesi més de 15 quilograms o bé que pesi menys de 12 quilograms.

Atenció: atès que se'n desconeix la distribució, no es pot distribuir la probabilitat entre les dues cues. Pot ser que tots els lots pesin més de 15 quilograms, que tots pesin menys de 12 quilograms o qualsevol altra combinació.

L'única cosa que se sap del cert és que si la probabilitat de les dues cues és inferior a 0,0277, la d'una cua, també.

La probabilitat que els tres lots pesin més de 15 quilograms és inferior a 0,0277.

A cal perruquer



Façana d'una perruqueria a Lavapiés (Madrid), de Tamorlan, sota llicència CC BY-SA 3.0.

El Manolo vol obrir una nova perruqueria on només s'utilitzin productes ecològics, no testats en animals. Preveu que el seu públic tindrà una mitjana de 45 anys. Tanmateix, encarrega un estudi a una consultora sobre les característiques de la seva futura clientela. La consultora selecciona per mostreig aleatori simple una mostra de 84 clients potencials i n'estudia diverses característiques.

Amb un nivell de confiança del 99 %, la consultora diu al Manolo que l'edat mitjana de la seva futura clientela estarà situada entre 48,7 i 55,3 anys.

- a) Creieu que els resultats de l'estudi corroboren l'opinió del Manolo sobre l'edat mitjana de la clientela potencial?
- b) En vista d'aquests resultats, deduiu la mitjana i la variància mostral de la variable edat que ha emprat la consultora.
- c) Proposeu una estimació puntual de la mitjana i la variància poblacionals amb les dades mostrals.
- d) Plantegeu un contrast d'hipòtesi paramètrica per convèncer el Manolo que l'edat mitjana de la seva clientela és superior a 45 anys per a un nivell de significació $\alpha = 0,01$.
- e) Ja fa tres mesos que la perruqueria és oberta i té força èxit. El Manolo ha fet una fitxa de cada client on se'n recull la data de naixement; així, si el dia de l'aniversari van a la perruqueria, sempre els fa un detall. A partir de la informació de les fitxes, que és un cens dels clients, un amic li ha calculat l'edat mitjana i la variància de l'edat de la seva clientela. El paper amb els resultats s'ha mullat i ja només s'hi pot veure que la variància de l'edat dels clients és de 138 anys². Amb aquesta informació addicional i les dades inicials dels clients potencials, torneu a calcular l'interval de confiança de la mitjana poblacional per a un nivell de confiança del 90 %.

Resolució de la pregunta a

L'edat mitjana dels clients prevista pel Manolo era de 45 anys, valor que no està inclòs en l'interval de confiança calculat per la consultora.

Per tant:

Els resultats de la consultora no corroboren l'opinió d'en Manolo.

Resolució de la pregunta b

Sigui X l'edat en anys dels clients del Manolo. Per calcular l'interval de confiança de la mitjana poblacional, se suposa que la consultora ha emprat la fórmula següent, que és habitual per a aquest interval quan les mostres són prou grans i les variàncies poblacionals són desconegudes:

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n-1}},$$

on $\bar{X}$ és la mitjana mostral, S la desviació estàndard mostral, n la grandària mostral i $z_{\alpha/2}$ el valor en una distribució normal estàndard que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2}$, on α és el nivell de significació expressat en tant per u (també és el valor complementari del nivell de confiança expressat en tant per u).

En aquests tipus d'interval, la mitjana mostral queda al centre, atès que el límit inferior es calcula com la mitjana mostral menys una expressió coneguda com el marge d'error o precisió, i el superior com la mitjana mostral més la mateixa expressió.

Per tant, per trobar la mitjana mostral només cal trobar el valor situat a la meitat de l'interval. Aquest valor és:

$$X_{\text{meitat interval}} = \frac{\text{Límit inferior} + \text{Límit superior}}{2}$$

$$\Rightarrow \bar{X} = \frac{48,7 + 55,3}{2} = 52 \text{ anys}$$

La mitjana mostral és igual a 52 anys.

Un cop deduïda la mitjana mostral, només queda com a incògnita la desviació estàndard mostral:

$$52 - 2,58 \frac{S}{\sqrt{84 - 1}} = 48,7 ,$$

on 52 és la mitjana mostral, S és el valor que es busca (la desviació estàndard mostral), 84 és la grandària mostral i 2,58 és el valor de la distribució normal estàndard que deixa a la dreta una probabilitat igual a 0,005.

$$\Rightarrow S = \frac{(52 - 48,7) \cdot \sqrt{84 - 1}}{2,58}$$

$$\Rightarrow S = \frac{3,3 \cdot \sqrt{84 - 1}}{2,58}$$

$$\Rightarrow S = 11,67 \text{ anys}$$

Aquest resultat també es pot obtenir a partir del límit superior, atès que la diferència entre la mitjana mostral i el límit inferior és la mateixa que la diferència entre la mitjana mostral i el límit superior (en aquest cas, 3,3 anys):

$$52 + 2,58 \frac{S}{\sqrt{84 - 1}} = 55,3$$

$$\Rightarrow S = \frac{(55,7 - 52) \cdot \sqrt{84 - 1}}{2,58}$$

$$\Rightarrow S = \frac{3,3 \cdot \sqrt{84 - 1}}{2,58} = 11,67 \text{ anys}$$

S és la desviació estàndard mostral i el seu quadrat és la variància mostral:

$$S^2 = 11,67^2 = 136,23 \text{ anys}^2$$

La variància mostral és igual a 136,23 anys².

Resolució de la pregunta c

La mitjana mostral és un bon estimador de la mitjana poblacional. En canvi, la variància mostral no ho és.

La variància mostral és un estimador esbiaixat de la variància poblacional. L'estimador no esbiaixat de la variància poblacional és la quasivariància.

Sigui S^2 la variància mostral i sigui S^{2*} la quasivariància mostral.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$$S^{2*} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Per tant,

$$n \cdot S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = (n-1) \cdot S^{2*}$$

$$\Rightarrow S^2 = \frac{n-1}{n} S^{2*} \quad i \quad S^{2*} = \frac{n}{n-1} S^2$$

$$S^{2*} = \frac{84}{83} \cdot 136,23 = 137,87 \text{ anys}^2$$

Les estimacions puntuals de la mitjana i la variància poblacionals són 52 anys i 137,87 anys², respectivament.

Resolució de la pregunta d

Es vol valorar si l'edat mitjana de la població potencial de clients (μ) de la perruqueria ecològica és igual a 45 anys o si és superior.

Amb aquesta fi, es planteja un contrast d'hipòtesi paramètrica en quatre passos. En el primer es plantegen la hipòtesi que s'ha de contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió del contrast.

Pas 1: plantejament de les hipòtesis

Es formula aquest enunciat en forma de dues hipòtesis sobre el paràmetre poblacional (en aquest problema, la mitjana poblacional μ): la hipòtesi nul·la (H_0) i l'alternativa (H_1).

La hipòtesi nul·la es planteja en forma d'igualtat i, posteriorment, s'utilitza en el càlcul de l'estadístic de prova (pas 2). La hipòtesi alternativa complementa la nul·la, a fi de recollir l'enunciat (que l'edat mitjana és superior a 45 anys).

$$H_0: \mu = 45$$

$$H_1: \mu > 45$$

Pas 2: càlcul de l'estadístic de prova (EP)

Es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa.

De què depèn l'estadístic de prova?

- Del problema: Què es contrasta? Es contrasta una mitjana poblacional.
- De les característiques de l'estudi: s'ha pres una mostra per mostreig aleatori simple que té la grandària $n = 84$, que es considera prou gran per emprar estadístics de prova la distribució del quals pot ser aproximada a la llei normal. La variància poblacional és desconeguda.

Estadístic de prova:

$$EP = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} \sim N(0; 1)$$

on μ_0 és la mitjana poblacional suposant que la hipòtesi nul·la és certa: $\mu_0 = 45$.

$$EP = \frac{52 - 45}{\frac{11,67}{\sqrt{84 - 1}}} = 5,46$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació (opció a), o bé buscant el p-valor (opció b).

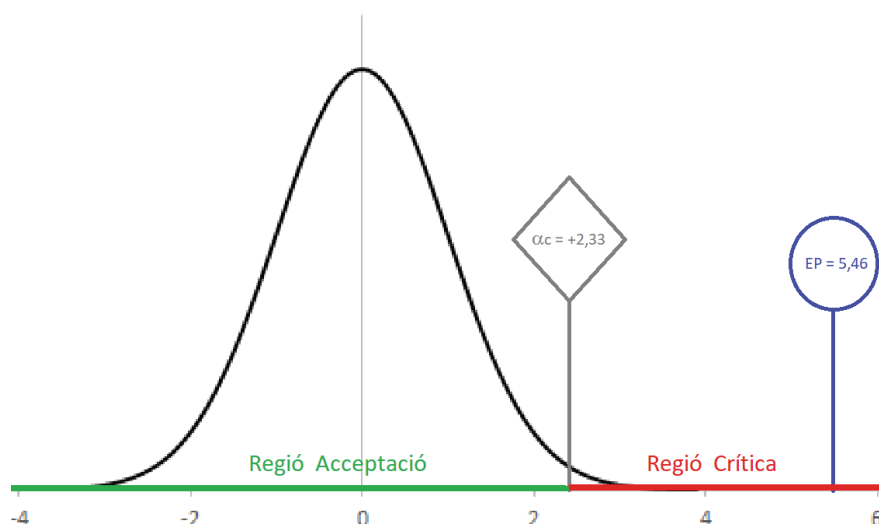
Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova, suposant que la hipòtesi nul·la és verdadera, es divideix en dues regions. Posteriorment, es comprova en quina de les regions es troba l'estadístic de prova calculat amb les dades mostrals.

Informacions necessàries per trobar la zona crítica i la zona d'acceptació:

- Distribució de l'estadístic de prova: en aquest cas, una distribució normal estàndard.
- Nivell de significació: $\alpha = 0,01$.
- Signe de la hipòtesi alternativa: $>$ (és un contrast unilateral dret); la regió crítica és a la zona dreta de la distribució.

Amb aquesta informació es busca a la distribució normal estàndard un únic valor crític positiu (α_c), tal que $P(Z > \alpha_c) = \alpha$, atès que és un contrast unilateral dret: $\alpha_c = 2,33$.



L'estadístic de prova ($EP = 5,46$) és a la regió crítica (la zona vermella de la gràfica).

$$EP \in \text{regió crítica}$$

Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha calculat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la sigui certa.

Informació necessària per calcular el p-valor:

- Distribució de l'estadístic de prova: en aquest cas, una normal estàndard.
- Signe de la hipòtesi alternativa: $>$ (és un contrast unilateral dret).

Atès que és un contrast unilateral dret, el p-valor és el valor de l'àrea que queda a la dreta de l'estadístic de prova (5,46).

$$P(X > 5,46) = 0,000\ 000\ 02$$

$$p\text{-valor} < 0,0001 < \alpha$$

Aquest valor, una vegada calculat, es compara amb el nivell de significació. En aquest cas, el p-valor és inferior al nivell de significació ($p\text{-valor} < \alpha$).

Pas 4: conclusió

Atès que l'estadístic de prova pertany a la regió crítica ($EP \in \text{regió crítica}$) o bé atès que el p-valor és inferior al nivell de significació (α), es conclou que per a un nivell de significació $\alpha = 0,01$ hi ha evidència suficient per rebutjar la hipòtesi nul·la i, per tant, es conclou que l'edat mitjana de la clientela potencial de la nova perruqueria d'en Manolo és superior a 45 anys.

Resolució de la pregunta e

La nova informació és sobre el valor de la variància poblacional σ^2 . Abans la informació provenia d'una mostra, ara en canvi la variància s'ha calculat amb la informació que conté la fitxa de cada un dels clients. Es té, doncs, informació de tota la població de clients (cens).

Informació necessària per seleccionar la fórmula de l'interval de confiança:

- De què es fa l'interval? En aquest cas, d'una mitjana poblacional (μ).
- De les característiques de l'estudi: s'ha pres per mostreig aleatori simple una mostra prou gran ($n = 84$) per aproximar la distribució de l'estadístic de prova a una normal.
- De la informació disponible: es coneix la variància poblacional.

L'interval adequat és:

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

on $\bar{X}$ és la mitjana mostral, σ la desviació estàndard poblacional, n la grandària mostral i $z_{\alpha/2}$ el valor en una distribució de normal estàndard que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2}$.

La desviació estàndard és l'arrel quadrada positiva de la variància: $\sigma = \sqrt{138} = 11,75$ anys.

Límit inferior:

$$52 - 1,645 \cdot \frac{11,75}{\sqrt{84}} = 49,9$$

Límit superior:

$$52 + 1,645 \cdot \frac{11,75}{\sqrt{84}} = 54,1$$

Per a un nivell de confiança $1 - \alpha = 0,90$, l'edat mitjana dels clients de la perruqueria ha d'estar situada dins de l'interval calculat, és a dir, pren un valor situat entre 49,9 anys i 54,1 anys.

Comentaris

Aquest exercici exemplifica com es pot donar resposta a preguntes concretes sobre paràmetres poblacionals tant amb un interval de confiança (apartat a) com amb un contrast d'hipòtesi (apartat d). Cal recordar que cap de les dues tècniques treballa amb una certesa absoluta.

Matràs

Los inspectores del Ministerio de Industria utilizan, para comprobar el buen funcionamiento de los surtidores de gasolina, un matraz (botella) homologado de 8000 litros. La inspección de 21 surtidores situados en distintas gasolineras ha dado los siguientes resultados:

7933	7912	7930	7908	7905	8005	7953	7942	7941	7592	7549
7521	7490	7447	7937	8002	7973	7965	7958	7802	7683	

- Legalmente se admite un error del 1 %. Hallar el volumen mínimo que debe verterse en el matraz para superar favorablemente la inspección.
- ¿Qué surtidores de la muestra incumplen las especificaciones legales? (Táchese lo que proceda.)

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	

- Calcular la media y la desviación estándar de la muestra.

Suponemos que la variable “Volumen de gasolina vertido en el matraz por cada surtidor” es normal. A partir de la muestra de 21 observaciones:

- Hallar el intervalo de confianza para la esperanza de esa variable. Grado de confianza: 95 %.
- A partir de la muestra obtenida, debemos decidir si hay fraude generalizado. Plantear y resolver el único contraste unilateral que tiene sentido. (Nivel de significación: 1 % y 5 %.)

Els inspectors del Ministeri d'Indústria utilitzen un matràs (ampolla) homologat de 8000 litres per comprovar el bon funcionament dels assortidors de gasolina. La inspecció de 21 assortidors situats en diferents gasolineres ha donat els resultats següents:

7933	7912	7930	7908	7905	8005	7953	7942	7941	7592	7549
7521	7.490	7447	7937	8002	7973	7965	7958	7802	7683	

- a) Legalment s'admet un error de l'1%. Trobeu el volum mínim que s'ha d'abocar al matràs per superar favorablement la inspecció.
- b) Quins assortidors de la mostra incompleixen les especificacions legals? (Ratlleu el que calgui.)

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	

- c) Calculeu la mitjana i la desviació estàndard de la mostra.

Suposem que la variable "Volum de gasolina abocat en el matràs per cada assortidor" és normal. A partir de la mostra de 21 observacions.

- d) Trobeu l'interval de confiança per a l'esperança d'aquesta variable. Nivell de confiança: 95 %.
- e) A partir de la mostra obtinguda, hem de decidir si hi ha frau generalitzat. Plantegeu i resoleu l'únic contrast unilateral que té sentit. (Nivell de significació: 1 % i 5 %.)

Resolució de la pregunta a

Atès que només s'admet un error màxim d'un 1 %, cal que hi hagi com a mínim el 99 % dels 8000 litres en els matrassos homologats. Si n'hi posen de més, no estafen ningú i, per tant, només se'n comprova el límit inferior.

$$8000 \cdot 0,99 = 7920 \text{ litres}$$

El volum mínim per superar favorablement la inspecció és de 7920 litres per matràs.

Resolució de la pregunta b

Tots els assortidors dels quals s'hagin extret menys de 7920 litres incompleixen les especificacions legals. Aquests assortidors són:

2	4	5	10	11
12	13	14	20	21

Resolució de la pregunta c

Sigui X la variable volum de benzina que conté cada matràs en litres.

Mitjana aritmètica de la mostra: $\bar{x} = \frac{7933 + \dots + 7683}{21} = 7826,10$ litres

La mitjana és de 7826,10 litres.

Variància mostral: $S^2 = a_2 - a_1^2$
 a_1^2 és la mitjana (7826 litres) al quadrat.

$$a_2 = \frac{7933^2 + \dots + 7683^2}{21} = 61282109,5$$

$$S^2 = 61282109,5 - 7826,10^2 = 34342,85 \text{ litres}^2$$

$$S = \sqrt{34342,85} = 185,32 \text{ litres}$$

La desviació estàndard és igual a 185,32 litres.

Resolució de la pregunta d

Atès que la mostra és reduïda però les dades provenen d'una distribució normal i que la variància poblacional és desconeguda, per calcular l'interval de confiança s'utilitza la fórmula següent:

$$\bar{X} \pm t_{\alpha/2; n-1} \frac{S}{\sqrt{n-1}}$$

$t_{\alpha/2; n-1}$ és el valor en una distribució de la t de Student amb $n-1$ graus de llibertat, que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2} = 0,025$

$$t_{0,025; 20} = 2,09$$

$$\text{Límit inferior: } 7826 - 2,09 \frac{185,32}{\sqrt{21-1}} = 7739,66 \text{ litres}$$

$$\text{Límit superior: } 7826 + 2,09 \frac{185,32}{\sqrt{21-1}} = 7912,53 \text{ litres}$$

Per a un nivell de confiança $1 - \alpha = 0,95$, el volum mitjà de benzina als matrassos està situat entre 7740 i 7913 litres.

Resolució de la pregunta e

Pas 1: plantejament de les hipòtesis

Es tracta de plantejar un contrast que permeti, en cas de frau, trobar-ne una evidència clara. Es pot dir que es produeix frau de manera generalitzada si de mitjana no s'arriba a la quantitat mínima que requereixen les especificacions legals, que és de 7920 litres. Aquesta circumstància es recull en la hipòtesi alternativa. Així, si finalment es rebutja la hipòtesi nul·la és perquè s'haurà trobat una evidència que la mitjana és més baixa.

$$H_0: \mu = 7920$$

$$H_1: \mu < 7920$$

Pas 2: càlcul de l'estadístic de prova (EP)

Es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa, és a dir, substituint a la fórmula de l'estadístic de prova el valor del paràmetre poblacional pel que s'estableix en la hipòtesi nul·la. La fórmula de l'estadístic de prova correspon al cas d'una mostra petita, la població d'origen és normal i la variància poblacional és desconeguda.

$$EP = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} \sim t_{n-1} \text{ graus de llibertat}$$

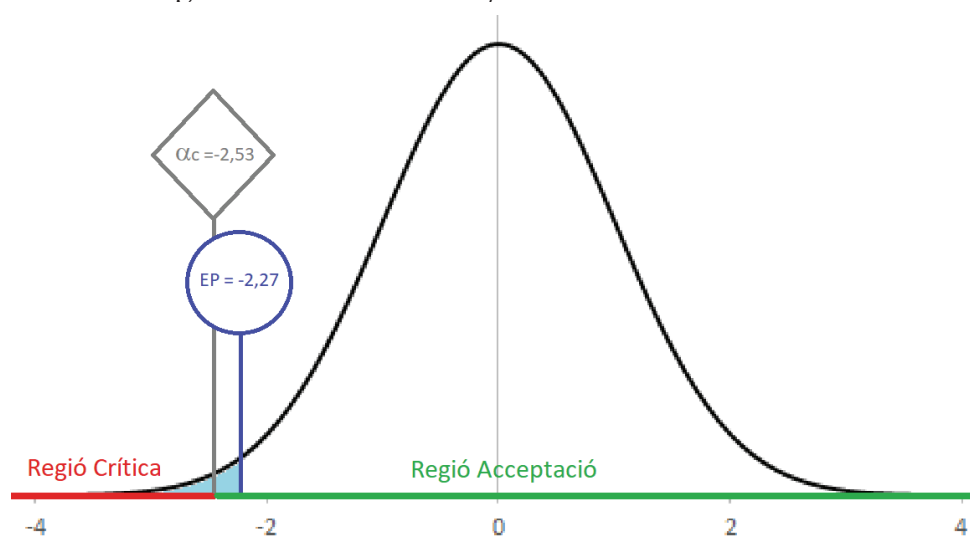
$$EP = \frac{7826 - 7920}{\frac{185,32}{\sqrt{21-1}}} = -2,27$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova suposant que la hipòtesi nul·la és verdadera es divideix en dues regions. Atès que la hipòtesi alternativa conté $<$ (contrast unilateral esquerre) per a cada nivell de significació, cal buscar un sol valor crític (α_c), que separa per l'esquerra una cua que conté la probabilitat equivalent al nivell de significació (α), que serà un valor negatiu en una distribució de la t de Student amb 20 graus de llibertat.

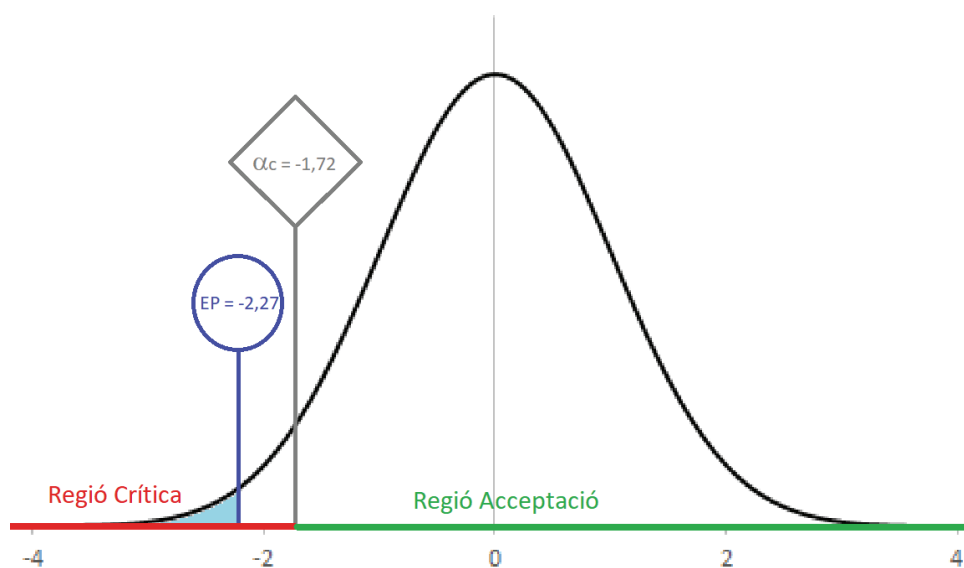
- Nivell de significació $\alpha = 0,01$: $\alpha_c = -2,53$



L'estadístic de prova és a la regió d'acceptació (la zona verda de la gràfica de sobre).

EP $\in$ Regió d'acceptació

- Nivell de significació $\alpha = 0,05$: $\alpha_c = -1,72$



L'estadístic de prova és a la regió crítica (la zona vermella de la gràfica de sobre).

EP $\in$ Regió crítica

Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha calculat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la sigui certa.

Atès que és un contrast unilateral esquerre, el p-valor és el valor de l'àrea que queda a l'esquerra de l'estadístic de prova (l'àrea en blau de les gràfiques).

$$P(X < -2,27) = \text{p-valor} = 0,0173$$

Un cop s'ha calculat aquest valor, es compara amb el nivell de significació. En aquest exercici el p-valor és superior al nivell de significació si $\alpha = 0,01$ i és inferior al nivell de significació si $\alpha = 0,05$.

Pas 4: conclusió

- ♦ Nivell de significació $\alpha = 0,01$
Atès que $EP \in$ regió d'acceptació o bé atès que $\text{p-valor} > \alpha$, es conclou que per a un nivell de significació $\alpha = 0,01$ no hi ha evidència suficient per rebutjar la hipòtesi nul·la, és a dir per concloure que s'incomplixen les especificacions legals.
- ♦ Nivell de significació $\alpha = 0,05$
Atès que $EP \in$ regió crítica o bé atès que $\text{p-valor} < \alpha$, es conclou que per a un nivell de significació $\alpha = 0,05$ hi ha evidència suficient per rebutjar la hipòtesi nul·la, és a dir es creu que s'incomplixen les especificacions legals.

Comentaris

Aquest exercici va bé per mostrar que, a vegades, la conclusió d'un contrast no queda prou clara i depèn del nivell de significació. Això succeeix quan el p-valor és proper als nivells de significació habituals. En canvi, quan el p-valor pren un valor extremadament baix o un valor molt elevat, el resultat és més clar —tot i que mai no és amb certesa.

Missatges de telèfon



The illustration shows a doctor in a white coat and tie, sitting at a desk with a telephone. He is looking towards a woman who is also on the telephone. The woman is seated and looking at the doctor. The background shows a medical office with shelves of bottles and a window.

THE DOCTOR AND THE TELEPHONE

Many lives are saved each year because the doctor is reached promptly by the Bell Telephone. From the information he receives by telephone he can give directions for emergency measures, and provide himself with the necessary medicines and instruments.

Consultations with specialists are easily arranged by means of the Bell Long Distance Service.

Call Contract Department,
The Mountain States Telephone & Telegraph Co.
KING COBBS, DIV. COM'L SUPT.

The Mountain States Telephone & Telegraph Co. published this advertisement in *The Meridian Times*, "The Doctor and the Telephone" in 1911.

Un segle després de la publicació d'aquell anunci del telèfon, es fa un estudi sobre l'ús de la missatgeria telefònica. Amb aquesta fi, s'elegeixen, per mostreig aleatori simple, dues mostres independents, una corresponent a nois i l'altra a noies, als quals se'ls pregunta sobre el nombre de missatges que envien al dia des del seu telèfon mòbil. Amb les dades d'aquest estudi, que són a la taula, responeu a les preguntes que segueixen.

	Nois ($\nearrow$)	Noies (+)
n	80	84
Mitjana	37	44
Quasivariància	280	320

- Considerem que sobre la base d'aquestes dades es pot dir que els nois envien de mitjana menys missatges que les noies? Realitzeu el contrast d'hipòtesi paramètrica adient per comprovar aquest fet per a un nivell de significació $\alpha = 0,01$.
- En vista de la vostra conclusió, quin tipus d'error podríeu estar cometent?
- A partir de quin nivell de significació canviaria la conclusió del contrast?
- Trobeu l'interval de confiança de la diferència del nombre de missatges diaris que envien des del mòbil els nois i les noies. Calculeu intervals per a un nivell de confiança $1 - \alpha = 0,99$; per a un nivell de confiança $1 - \alpha = 0,95$ i per a un nivell de confiança $1 - \alpha = 0,90$. Compareu els tres resultats.

Resolució de la pregunta a

Sigui $X_{\nearrow}$ una variable aleatòria que descriu el nombre de missatges que envien els nois i sigui X_{+} una variable aleatòria que descriu el nombre de missatges que envien les noies. Es valora si el nombre mitjà de missatges que envien els nois ($\mu_{\nearrow}$) és inferior al nombre mitjà de missatges que envien les noies (μ_{+}).

Amb aquesta fi es planteja un contrast d'hipòtesi paramètrica en quatre passos. En el primer es plantegen la hipòtesi que s'ha de contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió.

Pas 1: plantejament de les hipòtesis

Es formula aquest enunciat mitjançant dues hipòtesis sobre la diferència de mitjanes poblacionals: la hipòtesi nul·la (H_0) i l'alternativa (H_1).

La hipòtesi nul·la es planteja en forma d'igualtat i, posteriorment, s'utilitza per calcular l'estadístic de prova (pas 2). La hipòtesi alternativa complementa la nul·la, a fi de recollir l'enunciat; en aquest problema, si el valor de la mitjana dels nois és inferior al de les noies².

² En aquest exercici, la hipòtesi alternativa recull una desigualtat. En aquests casos també és habitual (i correcte) escriure la hipòtesi nul·la com la desigualtat no estricta (sempre s'hi inclou també la igualtat) que complementa la hipòtesi alternativa. En aquest exemple seria $H_0: \mu_{\nearrow} \geq \mu_{+}$. Tanmateix, en aquest llibre la hipòtesi nul·la s'escriu sempre com una igualtat.

$$H_0: \mu_{\neg} = \mu_{+} \Leftrightarrow \mu_{\neg} - \mu_{+} = 0$$

$$H_1: \mu_{\neg} < \mu_{+} \Leftrightarrow \mu_{\neg} - \mu_{+} < 0$$

Pas 2: càlcul de l'estadístic de prova (EP)

Es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa.

L'estadístic de prova depèn:

- Del problema: Què es contrasta? En aquest cas, la diferència entre dues mitjanes poblacionals.
- De les característiques de l'estudi: s'ha pres per mostreig aleatori simple dues mostres independents prou grans (80 i 84) per aproximar la distribució de la diferència de mitjanes a una distribució normal i les variàncies poblacionals són desconegudes.

$$\text{Estadístic de prova: } EP = \frac{(\overline{X}_{\neg} - \overline{X}_{+}) - (\mu_{\neg} - \mu_{+})}{\sqrt{\left(\frac{S_{\neg}^{*2}}{n_{\neg}} + \frac{S_{+}^{*2}}{n_{+}}\right)}} \sim N(0,1)$$

Atenció, atès que es tracta d'un contrast unilateral (la hipòtesi alternativa és <), l'ordre en el qual s'introdueixen les mitjanes és important. Ha de ser l'ordre recollit pel sistema d'hipòtesis.

- $n_{\neg}$ i n_{+} són les grandàries mostrals.
- $\overline{X}_{\neg}$ i $\overline{X}_{+}$ són les mitjanes mostrals.
- $\mu_{\neg} - \mu_{+}$ suposant que la hipòtesi nul·la és certa, és igual a 0.
- $S_{\neg}^{*2}$ i S_{+}^{*2} són les quasivariàncies mostrals.

$$EP = \frac{(37 - 44) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{280}{80} + \frac{320}{84}\right)}}$$

$$EP = -2,59$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova, suposant que la hipòtesi nul·la és verdadera, es divideix en dues regions. Posteriorment, es comprova en quina de les regions es troba l'estadístic de prova que s'ha calculat amb les dades mostrals.

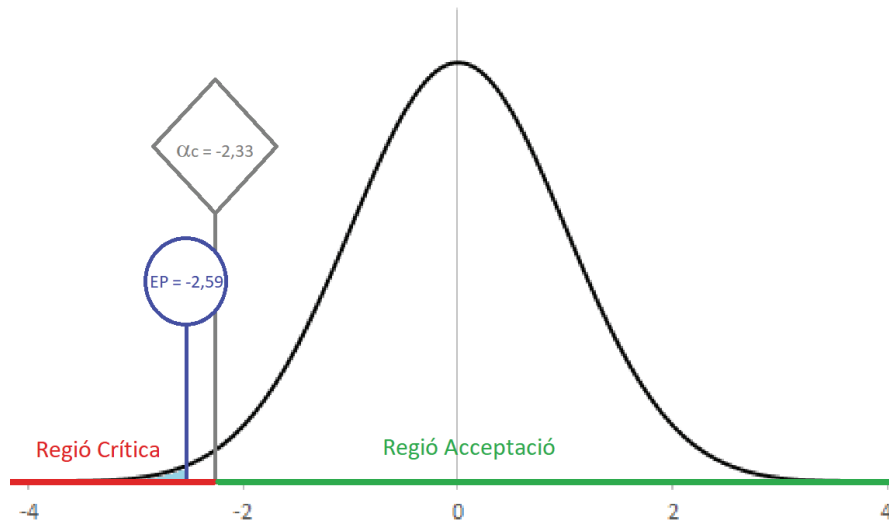
La regió crítica és una zona de baixa probabilitat, definida pel nivell de significació (α): amb la hipòtesi que les mitjanes poblacionals són iguals (H_0), és poc probable agafar per atzar mostres amb mitjanes mostrals molt diferents que generin un estadístic de prova que estigui situat en aquella zona. Si l'estadístic de prova és a la zona crítica, atès que és molt poc probable si la hipòtesi nul·la és certa, es conclourà que la hipòtesi nul·la no és correcta. S'estableix com a llindar de poc probable el nivell de significació.

La regió d'acceptació és l'altra part, on es considera raonable que s'obtinguin a l'atzar mostres que generin un estadístic de prova que se situï en aquella part si la hipòtesi nul·la és certa. Si l'estadístic de prova és a la zona d'acceptació es conclou que no s'ha trobat evidència suficient per rebatre la hipòtesi nul·la i, per tant, res no indica que els homes de mitjana enviïn menys missatges que les dones.

Per dur a terme aquesta operació es necessiten 3 peces d'informació:

- Distribució de l'estadístic de prova: en aquest cas, una distribució normal estàndard.
- Nivell de significació: $\alpha = 0,01$.
- Signe de la hipòtesi alternativa: $<$ (és un contrast unilateral esquerre); la regió crítica és a la part esquerra de la distribució.

Amb aquesta informació es busca el valor crític (α_c), que és negatiu, atès que és un contrast unilateral esquerre, en una distribució normal estàndard ($\alpha_c = -2,33$).



L'estadístic de prova ($EP = -2,59$) és a la regió crítica, la zona vermella d'aquesta gràfica.

$$EP \in \text{Regió crítica}$$

Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha calculat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la sigui certa i tenint present l'alternativa que indica de quin fet es busca l'evidència.

Per dur a terme aquesta operació es necessiten 2 peces d'informació:

- Distribució de l'estadístic de prova: en aquest cas, una normal estàndard.
- Signe de la hipòtesi alternativa: $<$ (és un contrast unilateral esquerre).

Atès que és un contrast unilateral esquerre, el p-valor és el valor de l'àrea que queda a l'esquerra de l'estadístic de prova ($P(X < -2,59)$); és l'àrea blava de la gràfica anterior.

$$p\text{-valor} = 0,005 < \alpha$$

Un cop s'ha calculat aquest valor, es compara amb el nivell de significació. En aquest cas, el p-valor és inferior al nivell de significació ($p\text{-valor} < \alpha$).

Pas 4: conclusió

Atès que $EP \in$ regió crítica o bé atès que $p\text{-valor} < \alpha$, es conclou que per a un nivell de significació $\alpha = 0,01$ hi ha evidència suficient per rebutjar la hipòtesi nul·la i, per tant, el nombre mitjà de missatges que envien les noies és superior al nombre mitjà de missatges que envien els nois.

Resolució de la pregunta b

En aquest estudi s'ha rebutjat que la hipòtesi nul·la sigui certa; tanmateix, pot ser que la conclusió de l'estudi sigui errònia. L'error que seria possible cometre és rebutjar la hipòtesi nul·la quan és certa. Aquest error és conegut com a *error de tipus I* o *error de primera espècie*, i, a priori, es pot produir amb una probabilitat igual al nivell de significació ($\alpha = 0,01$) si la hipòtesi nul·la és certa. Cal recalcar, però, que un cop s'ha fet el contrast, l'investigador desconeix si l'error s'ha produït o no s'ha produït; és un fet que ja ha succeït i, per tant, no és adequat dir que és probable que succeeixi. Es pot parlar, per exemple, del risc d'haver comès l'error.

Resolució de la pregunta c

La conclusió del contrast ha estat rebutjar la hipòtesi nul·la, atès que el p-valor ha estat inferior al nivell de significació $\alpha = 0,01$. En canvi, si es pren un nivell de significació α que sigui inferior al p-valor, aleshores la conclusió del contrast varia. Per tant, a partir de nivells de significació inferiors a 0,005 es modifica la conclusió del contrast.

Resolució de la pregunta d

Si dues mitjanes són iguals, la seva diferència és nul·la. Així, per valorar si hi ha diferències entre dues mitjanes es pot calcular l'interval de confiança de la diferència de mitjanes poblacionals i comprovar si l'interval conté el valor zero. En cas que el contingui, el valor zero és un valor plausible de la diferència entre mitjanes poblacionals. La conclusió seria que no s'han detectat diferències entre les mitjanes poblacionals amb el nivell de confiança escollit.

Per escollir la fórmula de l'interval de confiança, cal conèixer:

- De què es fa l'interval? En aquest cas, de la diferència entre dues mitjanes poblacionals.
- De les característiques de l'estudi: s'ha pres per mostreig aleatori simple dues mostres independents prou grans (80 i 84) per aproximar la distribució de l'estadístic de prova a una normal. Es desconeixen les variàncies poblacionals.

L'interval adequat és:

$$\overline{X}_+ - \overline{X}_\gamma \pm z_{\alpha/2} \sqrt{\left(\frac{S_+^{*2}}{n_+} + \frac{S_\gamma^{*2}}{n_\gamma}\right)}$$

Atès que és l'interval d'una diferència i que entre nois i noies no hi ha un ordre natural, no té importància si per fer la diferència es fa la mitjana de les noies menys la dels nois o a l'inrevés. Per comoditat, s'opta per posar primer el més gran i treballar amb un valor positiu de la diferència. Si es fa a l'inrevés, els resultats dels intervals canvien de signe.

$z_{\alpha/2}$ és el valor en una distribució de la normal estàndard que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2}$. Cal calcular tres intervals diferents per a 3 nivells de confiança.

Per a un nivell de confiança $1 - \alpha = 0,99$:

$$\frac{\alpha}{2} = 0,005$$

$$z_{\alpha/2} = 2,58$$

$$\text{Límit inferior: } (44 - 37) - 2,58 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 0,04$$

$$\text{Límit superior: } (44 - 37) + 2,58 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 13,96$$

L'interval de confiança de la diferència en el nombre mitjà de missatges que envien les noies i els nois per a un nivell de confiança $1 - \alpha = 0,99$ és $[0,04; 13,96]$.

Per a un nivell de confiança $1 - \alpha = 0,95$:

$$\frac{\alpha}{2} = 0,025$$

$$z_{\alpha/2} = 1,96$$

$$\text{Límit inferior: } (44 - 37) - 1,96 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 1,70$$

$$\text{Límit superior: } (44 - 37) + 1,96 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 12,30$$

L'interval de confiança de la diferència en el nombre mitjà de missatges que envien les noies i els nois per a un nivell de confiança $1 - \alpha = 0,95$ és $[1,70; 12,30]$.

Per a un nivell de confiança $1 - \alpha = 0,90$:

$$\frac{\alpha}{2} = 0,05$$

$$z_{\alpha/2} = 1,645$$

$$\text{Límit inferior: } (44 - 37) - 1,645 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 2,55$$

$$\text{Límit superior: } (44 - 37) + 1,645 \sqrt{\left(\frac{320}{84} + \frac{280}{80}\right)} = 11,45$$

L'interval de confiança de la diferència en el nombre mitjà de missatges que envien les noies i els nois per a un nivell de confiança $1 - \alpha = 0,90$ és $[2,55; 11,45]$.

Cap dels tres intervals calculats conté el valor zero; per tant, en els tres casos es conclou que hi ha diferències entre les mitjanes del nombre de missatges que envien els nois i les noies. Com es pot veure, com més alt és el nivell de confiança, més gran és l'amplitud dels intervals de confiança.

Comentaris

En el càlcul d'un interval de confiança cal trobar un equilibri entre el nivell de confiança, és a dir, que els resultats generin prou confiança per prendre una decisió, i la precisió del resultat. La precisió fa referència a l'amplitud de l'interval. Ateses les dades d'un estudi, com més precisió (un interval de confiança més estret), més baix és el nivell de confiança ($1 - \alpha$). Inversament, si es pren un nivell de confiança més alt, l'interval resultant té un rang més gran (és més ampli).

D'altra banda, a l'hora de fer un contrast cal tenir present que els resultats no són mai segurs, que hi ha la possibilitat de cometre errors. Un possible error és rebutjar la hipòtesi nul·la quan és certa —error de tipus I— (apartat *b*). Si la hipòtesi nul·la és certa, l'error de tipus I o error de primera espècie es comet amb una probabilitat igual al nivell de significació (α). Per tant, en el disseny d'un estudi, per reduir la probabilitat potencial de cometre aquest error es podria optar per considerar un nivell de significació molt i molt petit. Per què no es fa? No es fa perquè com més petit és el nivell de significació, més fàcil és creure que la hipòtesi nul·la és certa tot i no ser-ho. No rebutjar la hipòtesi nul·la quan és falsa es coneix com a *error de tipus II* o *error de segona espècie*. El nivell de significació es fixa a priori sense saber si la hipòtesi nul·la és certa o falsa i, per tant, és un valor de compromís entre el risc de cometre un error o un altre (rebutjar la hipòtesi nul·la quan és certa o no rebutjar-la quan és falsa).

Finalment, cal recordar que, quan es comparen magnituds, cal detectar les diferències rellevants per a la presa de decisions. En aquest sentit, cal distingir les diferències estadísticament significatives de les diferències econòmicament rellevants.

Barra de sushi



Yo! Sushi, de Dave Spellman, sota llicència CC BY 2.0.

En dos restaurants japonesos en els quals hi ha una cinta sobre la qual circulen platets de menjar que els clients van agafant, es prenen dues mostres aleatòries de 9 i 11 clients, respectivament. De cada client es recull el nombre total de platets que agafa. Les dades mostrals són les següents:

Restaurant A	Restaurant B
12	8
14	11
9	5
10	9
7	6
7	12
9	9
12	6
19	10
	5
	7

- Suposant que el nombre de platets que els clients agafen en cada restaurant segueix una distribució normal amb variàncies poblacionals iguals però desconegudes, contrasteu mitjançant un contrast d'hipòtesi paramètrica si el nombre mitjà de platets que els clients agafen és igual en tots dos restaurants, per a un nivell de significació $\alpha = 0,05$.
- Responen a la mateixa pregunta en els mateixos supòsits mitjançant el càlcul d'un interval de confiança per a un nivell de confiança $1 - \alpha = 0,95$.
- Calculeu els intervals de confiança del nombre mitjà de platets que els clients agafen en cada un dels restaurants per separat per a un nivell de confiança $1 - \alpha = 0,95$ i compareu-los.
- El propietari del restaurant A té fixat un preu per al bufet lliure (que inclou una beguda) de 13,5 euros. El cost mitjà d'un platet, tenint en compte totes les despeses, és d'1,1 euros i el de la beguda és de 30 cèntims. S'estima que el 25 % dels clients demanen una segona beguda, que reporta al restaurant un benefici net d'1,7 euros, i un 30 % demana un cafè, que comporta un marge de benefici d'1,2 euros. Feu una estimació del benefici esperat d'un dia qualsevol en què s'atenen 40 comensals.

Resolució de la pregunta a

Sigui X_A una variable aleatòria que descriu el nombre de platets que agafa un client en el restaurant A i sigui X_B una variable aleatòria que descriu el nombre de platets que agafa un client en el restaurant B. Es valora si el nombre mitjà de platets que agafen els clients en el restaurant A (μ_A) és igual al nombre mitjà de platets que agafen els clients en el restaurant B (μ_B) o si, al contrari, són diferents.

Amb aquesta fi, es planteja un contrast d'hipòtesi paramètrica en quatre passos. En el primer es plantegen la hipòtesi que cal contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la

credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió.

Pas 1: plantejament de les hipòtesis

Es formula l'enunciat mitjançant dues hipòtesis sobre els paràmetres poblacionals (en aquest problema, les mitjanes poblacionals μ_A i μ_B). La hipòtesi nul·la (H_0) es planteja en forma d'igualtat i, posteriorment, s'implementa en el càlcul de l'estadístic de prova (pas 2). La hipòtesi alternativa (H_1) complementa la hipòtesi nul·la a fi de recollir l'enunciat; en aquest cas, que les mitjanes poblacionals són diferents.

$$H_0: \mu_A = \mu_B \Leftrightarrow \mu_A - \mu_B = 0$$

$$H_1: \mu_A \neq \mu_B \Leftrightarrow \mu_A - \mu_B \neq 0$$

Pas 2: càlcul de l'estadístic de prova (EP)

Es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa.

L'estadístic de prova depèn:

- De què es contrasta: es contrasta la diferència entre dues mitjanes poblacionals.
- De les característiques de l'estudi: s'han pres dues mostres petites i independents per mostreig aleatori simple ($n_A = 9$ i $n_B = 11$), provinents de poblacions que segueixen distribucions normals amb variàncies poblacionals iguals però desconegudes.

Estadístic de prova:

$$EP = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) \cdot \left(\frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}\right)}} \sim t_{n_A + n_B - 2} \text{ graus de llibertat}$$

- n_A i n_B són les grandàries mostrals.
- $\overline{X}_A$ i $\overline{X}_B$ són les mitjanes mostrals. En aquest exercici, les estimacions puntuals de les mitjanes són: $\overline{x}_A = \frac{\sum_{i=1}^n x_i}{n} = \frac{99}{9} = 11$ platets i $\overline{x}_B = \frac{\sum_{i=1}^n x_i}{n} = \frac{88}{11} = 8$ platets.
- $\mu_A - \mu_B$, suposant que la hipòtesi nul·la és certa i és igual a 0.

- S_A^2 i S_B^2 són les variàncies mostrals: en aquest exercici, les variàncies mostrals són:

$$S_A^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{X}_A^2 = \frac{1205}{9} - 11^2 = 12,89 \text{ platets}^2$$

$$\text{i } S_B^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{X}_B^2 = \frac{762}{11} - 8^2 = 5,27 \text{ platets}^2$$

$$EP = \frac{(11 - 8) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{9} + \frac{1}{11}\right) \cdot \left(\frac{9 \cdot 12,89 + 11 \cdot 5,27}{9 + 11 - 2}\right)}} \sim t_{18 \text{ graus de llibertat}}$$

$$EP = 2,15$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova suposant que la hipòtesi nul·la és verdadera es divideix en dues regions: la regió crítica i la regió d'acceptació. Posteriorment, es comprova en quina de les regions es troba l'estadístic de prova que s'ha calculat amb les dades mostrals.

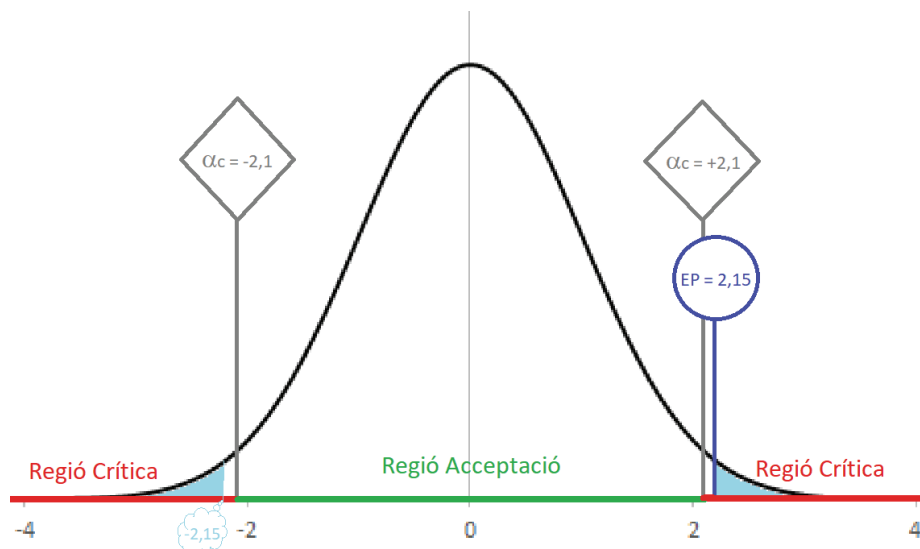
La regió crítica és una zona de baixa probabilitat, definida pel nivell de significació (α): amb la hipòtesi que les mitjanes poblacionals són iguals (H_0), és poc probable agafar per atzar mostres amb mitjanes mostrals molt diferents que generin un estadístic de prova que estigui situat en aquella zona. Si l'estadístic de prova és a la zona crítica, atès que és molt poc probable si la hipòtesi nul·la és certa, es conclourà que la hipòtesi nul·la no és correcta. S'estableix com a llindar de poc probable el nivell de significació.

La regió d'acceptació és l'altra part, la part on es considera raonable que si la hipòtesi nul·la és certa s'obtinguin a l'atzar mostres que generin un estadístic de prova que se situï en aquella part. Si l'estadístic de prova és a la zona d'acceptació, es conclou que no s'ha trobat evidència suficient per rebatre la hipòtesi nul·la i, per tant, és una hipòtesi creïble.

Per dur a terme aquesta operació, es necessiten 3 peces d'informació:

- Distribució de l'estadístic de prova: en aquest cas, una t de Student amb 18 graus de llibertat.
- Nivell de significació: $\alpha = 0,05$.
- Signe de la hipòtesi alternativa: $\neq$ (és un contrast bilateral); la regió crítica està formada per les dues cues de la distribució. La probabilitat de les dues cues considerades conjuntament és igual a $\alpha = 0,05$.

Amb aquesta informació es busquen els valors crítics $(-\alpha_c, \alpha_c)$, que en són dos, atès que és un contrast bilateral, en una t de Student amb 18 graus de llibertat: $|\alpha_c| = 2,1$.



L'estadístic de prova (2,15) és a la regió crítica, la zona vermella d'aquesta gràfica.

$$EP \in \text{Regió crítica}$$

Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha trobat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la és certa.

Per dur a terme aquesta operació, es necessiten 2 peces d'informació:

- Distribució de l'estadístic de prova: en aquest cas, una t de Student amb 18 graus de llibertat.
- Signe de la hipòtesi alternativa: $\neq$ (contrast bilateral).

Atès que és un contrast bilateral, el p-valor és la suma de les àrees de les dues cues: el valor de l'àrea que queda a la dreta de l'estadístic de prova (+2,15) més el valor de l'àrea que queda a l'esquerra de l'estadístic de prova en valor negatiu (-2,15). És la suma de les dues àrees blaves de la gràfica anterior.

$$p\text{-valor} = 0,046 < \alpha$$

El p-valor es compara amb el nivell de significació. En aquest cas és inferior al nivell de significació ($p\text{-valor} < \alpha$), cosa que significa que es considera poc probable haver

seleccionat a l'atzar aquesta mostra suposant que la hipòtesi nul·la és certa i, per tant, la hipòtesi nul·la no és creïble.

Pas 4: conclusió

Atès que $EP \in$ regió crítica o bé atès que $p\text{-valor} < \alpha$, es conclou que per a un nivell de significació $\alpha = 0,05$ hi ha evidència suficient per rebutjar la hipòtesi nul·la i, per tant, es considera que el nombre mitjà de platets que agafen els clients en un restaurant i en l'altre és diferent.

Resolució de la pregunta b

Per comprovar si hi ha diferències entre ambdues mitjanes, també es pot calcular l'interval de confiança de la diferència de mitjanes poblacionals i comprovar si l'interval conté el valor zero. Si l'interval conté el zero, aleshores el zero és un valor possible de la diferència entre les mitjanes poblacionals per a aquest nivell de confiança. Una diferència igual a zero indica que no hi ha diferències. En canvi, si l'interval no conté el zero, es pot concloure que es detecten diferències entre les dues mitjanes per a aquest nivell de confiança.

Per escollir la fórmula de l'interval de confiança cal conèixer:

- De què es fa l'interval? Es contrasta la diferència entre dues mitjanes poblacionals.
- De les característiques de l'estudi: s'han pres dues mostres petites per mostreig aleatori simple provinents de poblacions que segueixen distribucions normals amb variàncies poblacionals iguals però desconegudes.

L'interval adequat és:

$$\bar{X}_A - \bar{X}_B \pm t_{\alpha/2; n_A+n_B-2} \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) \sqrt{\left(\frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}\right)}}$$

$t_{\alpha/2; n_A+n_B-2}$ és el valor en una distribució de la t de Student amb $n - 1$ graus de llibertat, que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2} = 0,025$.

Límit inferior:

$$(11 - 8) - 2,1 \sqrt{\left(\frac{1}{9} + \frac{1}{11}\right) \sqrt{\frac{9 \cdot 12,89 + 11 \cdot 5,17}{9 + 11 - 2}}} = 0,064$$

Límit superior:

$$(11 - 8) + 2,1 \sqrt{\left(\frac{1}{9} + \frac{1}{11}\right)} \sqrt{\frac{9 \cdot 12,89 + 11 \cdot 5,17}{9 + 11 - 2}} = 5,936$$

L'interval de confiança per a un nivell de confiança $1 - \alpha = 0,95$ és $[0,064; 5,936]$. Atès que aquest interval no conté el valor zero, es conclou que sobre la base d'aquest resultat s'han detectat diferències en el nombre mitjà de platets que agafen els clients en un restaurant i en l'altre.

Resolució de la pregunta c

En aquest apartat es calcula primer l'interval de confiança de la mitjana poblacional de plats que els clients agafen en el restaurant A i després es fa el mateix per al restaurant B. Per als dos restaurants les característiques són les mateixes, atès que les dues mostres són petites i les poblacions d'origen segueixen una distribució normal.

L'interval adequat és:

$$\bar{X} \pm t_{\alpha/2; n-1} \frac{S}{\sqrt{n-1}}$$

$t_{\alpha/2; n-1}$ és el valor en una distribució de la t de Student amb $n-1$ graus de llibertat, que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2}$. En el cas del restaurant A, $t_{0,025; 8} = 2,306$, i en el cas del restaurant B, $t_{0,025; 10} = 2,228$.

Límit inferior del restaurant A:

$$11 - 2,306 \frac{3,59}{\sqrt{9-1}} = 8,073$$

Límit superior del restaurant A:

$$11 + 2,306 \frac{3,59}{\sqrt{9-1}} = 13,927$$

Per a un nivell de confiança $1 - \alpha = 0,95$, la mitjana poblacional de platets que els clients agafen en el restaurant A està situat entre 8,1 i 13,9 platets.

Límit inferior del restaurant B:

$$8 - 2,228 \frac{2,30}{\sqrt{11-1}} = 6,382$$

Límit superior del restaurant B:

$$8 + 2,228 \frac{2,30}{\sqrt{11-1}} = 9,618$$

Per a un nivell de confiança $1 - \alpha = 0,95$, la mitjana poblacional de platets que els clients agafen en el restaurant B està situat entre 6,4 i 9,6 platets.

Aquests dos intervals s'encavallen. En vista d'aquests resultats, és possible que el nombre mitjà de platets per client en el restaurant A sigui, per exemple, 9 platets, atès que aquest valor pertany a l'interval calculat. De la mateixa manera, és possible que el nombre mitjà de platets per client en el restaurant B sigui, per exemple, 9 platets, atès que aquest valor pertany a l'interval calculat. Així, sobre la base d'aquesta anàlisi, no és possible descartar que el nombre mitjà de platets que agafa un client en un restaurant i en l'altre sigui el mateix.

Aquest resultat sembla que contradiu el dels apartats *a* i *b*. Tanmateix, no ho fa. Quan es tracten els dos restaurants per separat, atès que les mostres són reduïdes, l'anàlisi no té gaire potència i no hi detecta diferències, quan en realitat sí que n'hi ha (resultat dels apartats *a* i *b*).

Quan les mostres són petites, la variància de l'estimador mitjana mostral és gran (en comparació amb la d'una mostra gran), cosa que implica que les estimacions són menys precises. A conseqüència d'aquesta imprecisió, és més difícil detectar-hi diferències. Tanmateix, si quan es calculen els intervals per separat tot i tenir mostres petites s'hi detecten diferències (hi ha diferències si els intervals estan separats, és a dir, que no s'encavallen gens), aleshores si es calcula l'interval de la diferència per al mateix nivell de significació també s'hi detectaran diferències. La inversa, en canvi, com s'ha vist, no és certa.

Resolució de la pregunta *d*

Per a cada client, el benefici que el propietari espera (ingressos – despeses) és:

$$[13,5 + (1,7 * 0,25 + 0 * 0,75) + (1,2 * 0,3 + 0 * 0,7)] - [1,1 * E[X_A] + 0,3]$$

Ingressos: 13,5 € és l'ingrés generat pel preu del menú. L'ingrés en concepte de segona beguda és d'1,7 €, però només un de cada quatre clients demanen una segona beguda; l'altre 75% dels clients no en demanen cap, per la qual cosa es generen uns ingressos nuls per aquest concepte. Es tracta d'una variable dicotòmica. De la mateixa manera, només un 30 % dels clients prenen cafè, concepte pel qual es generen uns ingressos nets d'1,2 €.

Despeses: inclou el cost estimat de cada platet pel nombre esperat de platets demanats i el cost de la beguda.

Atès que la mitjana mostral és un bon estimador de la mitjana poblacional i que $E[X_A] = \mu_A$, per trobar una estimació del benefici esperat per client se substitueix $E[X_A]$ per la mitjana mostral $\bar{x}_A = 11$.

$$(13,5 + 1,7 * 0,25 + 1,2 * 0,3) - (1,1 * 11 + 0,3) = 1,885$$

Si té 40 clients, el benefici esperat és:

$$1,885 \cdot 40 = 75,4 \text{ euros}$$

L'estimació puntual que s'obté a partir d'aquesta mostra del benefici esperat si entren 40 comensals al restaurant A és de 75,4 euros.

Comentaris

S'ha exemplificat com es poden emprar procediments diferents per comparar dues mitjanes (per exemple, interval de confiança o contrast d'hipòtesis de la diferència de mitjanes poblacionals). Els resultats poden variar segons el procediment (vegeu l'apartat c), atès que alguns procediments tenen menys potència que altres. S'anomena potència d'un contrast la capacitat de rebutjar la hipòtesi nul·la quan és falsa. Cal tenir present que la conclusió d'un contrast és forta quan es rebutja la hipòtesi nul·la (s'ha trobat prou evidències per rebutjar-la). La conclusió de no rebutjar és molt més dèbil (les conclusions potencials són asimètriques). Finalment, cal recordar que els resultats mai no tenen una certesa absoluta.

Xocolata desfeta



La Casa Amatller, en el local situat al passeig de Gràcia de Barcelona, ofereix, entre les seves especialitats, una xocolata a la pedra desfeta que demanen el 20 % dels comensals per berenar. Recentment, una revista d'oci ha fet un article sobre el local, amb una atenció especial a la xocolata desfeta. Per valorar l'impacte d'aquesta publicitat, s'ha pres una mostra aleatòria de 88 clients que berenaven i s'ha comprovat que 22 han demanat xocolata desfeta.³

- Contrasteu mitjançant un contrast d'hipòtesi paramètrica si la proporció de clients que demana xocolata desfeta ha augmentat després que s'hagi publicat l'article per a un nivell de significació $\alpha = 0,05$.
- Calculeu l'interval de confiança de la proporció poblacional de clients que demana xocolata desfeta per a un nivell de confiança $1 - \alpha = 0,99$.

³ La Casa Amatller existeix, està situada al passeig de Gràcia i ofereix xocolata desfeta, però totes les dades d'aquest enunciat són fictícies.

- d) Calculeu l'amplitud de l'interval anterior i l'amplitud d'un interval calculat suposant que la mostra és el doble de gran (176 clients), però mantenint inalterada la proporció mostral. Comproveu que l'amplitud de l'interval es redueix.
- c) Trobeu la grandària mostral necessària per reduir a la meitat l'amplitud d'un interval de confiança d'una proporció poblacional quan la mostra inicial és gran.

Resolució de la pregunta a

Es valora si la proporció poblacional π de clients que demana xocolata desfeta després que s'hagi publicat l'article es manté en el 20 % o si ha augmentat.

Amb aquesta fi, es planteja un contrast d'hipòtesi paramètrica en quatre passos. En el primer es plantegen la hipòtesi que s'ha de contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió.

Pas 1: plantejament de les hipòtesis

Es transcriu aquest enunciat en forma de dues hipòtesis sobre el paràmetre poblacional (en aquest problema, la proporció poblacional π): la hipòtesi nul·la (H_0) i l'alternativa (H_1).

La hipòtesi nul·la es planteja en forma d'igualtat i, posteriorment, s'utilitza per calcular l'estadístic de prova (pas 2). La hipòtesi alternativa complementa la nul·la, a fi de recollir l'enunciat; en aquest exercici que la proporció poblacional ha augmentat.

$$H_0: \pi = 0,2$$

$$H_1: \pi > 0,2$$

Pas 2: càlcul de l'estadístic de prova (EP)

Es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa.

L'estadístic de prova depèn:

- Del problema: Què es contrasta? Es contrasta una proporció poblacional.
- De les característiques de l'estudi: s'ha pres una mostra per mostreig aleatori simple d'una grandària $n = 88$, que es considera prou gran per emprar estadístics de prova que es poden aproximar a la llei normal.

Estadístic de prova:

$$EP = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1 - \pi_0)}{n}}} \sim N(0,1)$$

$$EP = \frac{0,25 - 0,2}{\sqrt{\frac{0,2 \cdot 0,8}{88}}} = 1,17$$

p és la proporció mostral de clients que demanen xocolata desfeta després que s'hagi publicat l'article: $p = \frac{22}{88} = 0,25$.

π_0 és la proporció poblacional suposant que la hipòtesi nul·la és certa: $\pi_0 = 0,2$.

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova suposant que la hipòtesi nul·la és verdadera es divideix en dues regions. Posteriorment, es comprova en quina de les regions es troba l'estadístic de prova que s'ha calculat amb les dades mostrals.

Per dur a terme aquesta operació, es necessiten 3 peces d'informació:

- Distribució de l'estadístic de prova: distribució normal estàndard.
- Nivell de significació: $\alpha = 0,05$.
- Signe de la hipòtesi alternativa: $>$ (contrast unilateral dret); la regió crítica és a la zona dreta de la distribució.

Amb aquesta informació, es busca el valor crític (α_c) en una distribució normal estàndard: $\alpha_c = 1,645$. És positiu perquè és un contrast unilateral dret.

Atès que $1,17 < 1,645$, l'estadístic de prova és a la regió d'acceptació.

$EP \in \text{Regió d'acceptació}$

Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha calculat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la és certa.

Per dur a terme aquesta operació, es necessiten dues peces d'informació:

- Distribució de l'estadístic de prova: distribució normal estàndard.
- Signe de la hipòtesi alternativa: $>$ (contrast unilateral dret).

Atès que és un contrast unilateral dret, el p-valor és el valor de l'àrea que queda a la dreta de l'estadístic de prova.

$$p\text{-valor} = 0,1205 > \alpha$$

Una vegada calculat, el p-valor es compara amb el nivell de significació. En aquest cas, el p-valor és superior al nivell de significació ($p\text{-valor} > \alpha$). Això significa que els resultats obtinguts no contradiuen la hipòtesi nul·la i, per tant, no fan pensar que sigui falsa.

Pas 4: conclusió

Atès que $EP \in$ regió d'acceptació o bé atès que $p\text{-valor} > \alpha$, es conclou que per a un nivell de significació $\alpha = 0,05$ no hi ha evidència suficient per rebutjar la hipòtesi nul·la i, per tant, no es pot dir que el consum de xocolata desfeta hagi augmentat.

Resolució de la pregunta b

Es calcula l'interval de confiança d'una proporció poblacional i, posteriorment, se'n calcula l'amplitud. L'amplitud de l'interval és la diferència entre el límit superior de l'interval i el límit inferior de l'interval.

Per escollir la fórmula de l'interval de confiança cal conèixer:

- De què es fa l'interval? En aquest cas, d'una proporció poblacional.
- De les característiques de l'estudi: s'ha pres una mostra prou gran per mostreig aleatori simple per aproximar la distribució de la proporció mostral a una distribució normal.

L'interval adequat és:

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

$z_{\alpha/2}$ és el valor en una distribució normal estàndard, que deixa a la dreta una probabilitat igual a $\frac{\alpha}{2}$.

$\sqrt{\frac{p \cdot (1-p)}{n}}$ és l'error estàndard estimat. *Error estàndard* és el nom que s'utilitza per anomenar la desviació estàndard de l'estimador (proporció mostral: $\hat{p}$) del paràmetre poblacional (proporció poblacional: π). Atès que es desconeix π , per estimar l'error estàndard també s'utilitza la proporció mostral estimada (en aquest cas, $p = 0,25$).

$$0,25 - 2,58 \sqrt{\frac{0,25 \cdot 0,75}{88}} = 0,1311$$

$$0,25 + 2,58 \sqrt{\frac{0,25 \cdot 0,75}{88}} = 0,3689$$

L'interval de confiança per a un nivell de confiança $1 - \alpha = 0,99$ de la proporció poblacional és $[0,1311; 0,3689]$.

Per tant, amb una confiança del 99 % es pot dir que no s'ha detectat cap augment de la demanda de xocolata després que s'hagi publicat l'article, ja que l'interval de la proporció de clients que demana xocolata (13,11 %–36,89 %) inclou com a valor el 20 %, que és el valor de la proporció de clients que demanava xocolata abans de la campanya de publicitat.

Resolució de la pregunta c

L'amplitud de l'interval és la diferència entre el límit superior i l'inferior:

$$0,3689 - 0,1311 = 0,2378$$

L'amplitud de l'interval (en tant per u) quan el nivell de confiança és del 99 % i la grandària mostral $n = 88$ és de 0,2378.

Quan $n_{gran} = 176$, el nou interval és:

$$0,25 - 2,58 \sqrt{\frac{0,25 \cdot 0,75}{176}} = 0,1659$$

$$0,25 + 2,58 \sqrt{\frac{0,25 \cdot 0,75}{176}} = 0,3341$$

L'amplitud de l'interval (en tant per u) quan el nivell de confiança és del 99 % i la grandària mostral $n_{gran} = 176$ és de $0,3341 - 0,1659 = 0,1681$.

Si s'augmenta la grandària mostral es redueix l'error estàndard, atès que la n és al divisor. En conseqüència, quan s'augmenta la grandària mostral es redueix l'amplitud de l'interval per un mateix nivell de confiança i, per tant, augmenta la precisió de l'estimació.

Resolució de la pregunta d

Sigui a_1 l'amplitud d'un primer interval de confiança d'una proporció poblacional π per a un nivell de confiança $1 - \alpha$ quan la grandària de la mostra és n_1 .

Sigui a_2 l'amplitud d'un segon interval de confiança de la proporció poblacional π per al mateix nivell de confiança $1 - \alpha$ quan la grandària de la mostra és n_2 .

Sigui $a_1 = 2 \cdot a_2$, és a dir, el primer interval és el doble d'ample que el segon, atès que l'amplitud s'ha reduït a la meitat.

L'amplitud del primer interval (a_1) és:

$$\begin{aligned} a_1 &= \left(p + z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} \right) - \left(p - z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} \right) \\ \Rightarrow a_1 &= p + z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} - p + z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} \\ \Rightarrow a_1 &= 2 \cdot z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} \end{aligned}$$

De la mateixa manera, l'amplitud del segon interval és:

$$\Rightarrow a_2 = 2 \cdot z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_2}}$$

atès que $a_1 = 2 \cdot a_2 \Rightarrow a_2 = \frac{a_1}{2}$

$$\Rightarrow a_2 = \frac{2 \cdot z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}}}{2} = z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}}$$

Per tant:

$$\begin{aligned}
2 \cdot z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_2}} &= z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n_1}} \\
\Rightarrow 2 \cdot \sqrt{\frac{p \cdot (1-p)}{n_2}} &= \sqrt{\frac{p \cdot (1-p)}{n_1}} \\
\Rightarrow \left(2 \cdot \sqrt{\frac{p \cdot (1-p)}{n_2}} \right)^2 &= \left(\sqrt{\frac{p \cdot (1-p)}{n_1}} \right)^2 \\
\Rightarrow 4 \cdot \frac{p \cdot (1-p)}{n_2} &= \frac{p \cdot (1-p)}{n_1} \\
\Rightarrow 4 \cdot n_1 &= n_2
\end{aligned}$$

La mostra del segon interval és 4 vegades més gran que la del primer interval.

Per reduir a la meitat l'amplitud de l'interval, cal quadruplicar la grandària mostral.

Comentaris

La precisió d'un interval de confiança depèn de tres elements: el nivell de confiança $1 - \alpha$ (com més alt és el nivell de confiança, més gran és l'amplitud i, per tant, la precisió és més petita —vegeu l'exercici de “Missatges de telèfon”), de la grandària mostral (com més gran la mostra, menys amplitud té l'interval i, per tant, major precisió) i de la variabilitat de la variable d'estudi, que és un element extern (com més variabilitat, més amplitud i, per tant, menys precisió). L'investigador pot modificar la grandària mostral i/o el nivell de confiança per canviar el nivell de precisió.

Canvis socials



Barcelona, febrer de 2021. El passeig de Gràcia durant les manifestacions amb motiu de l'arrest de Pablo Hasél.

Correspon sovint als joves promoure els canvis de paradigma social. S'ha passat una enquesta a 80 persones d'entre 18 i 26 anys (joves) i a 106 persones d'entre 45 i 55 anys en que se'ls demana si estaven a favor d'implementar un nou impost per lluitar contra

el canvi climàtic. Entre els joves, hi han respost afirmativament 34 persones; quant a les persones al voltant dels cinquanta anys, les respostes afirmatives han estat 40.

- a) Mitjançant un contrast d'hipòtesi paramètrica, per a un nivell de significació de 0,05 valoreu si hi ha diferències d'opinió entre aquests dos grups de població sobre la implementació del nou impost.
- b) En vista de la conclusió, quin tipus d'error es podria estar cometent?

Resolució de la pregunta a

Es valora si l'opinió dels joves coincideix amb la de la generació que té prop de 50 anys. Amb aquesta fi, es fa un contrast per valorar si la proporció poblacional de joves (π_j) que està a favor de l'impost és igual a la proporció poblacional de persones vora la cinquantena (π_{50}) que està a favor de l'impost.

Es planteja, doncs, un contrast d'hipòtesi paramètrica en quatre passos. En el primer es plantegen la hipòtesi que s'ha de contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió.

Pas 1: plantejament de les hipòtesis

Es planteja una hipòtesi que suposa que les proporcions poblacionals que són favorables a l'impost són iguals, és a dir, que no hi ha diferències entre els dos grups d'edat. Tots els càlculs es fan assumint aquesta hipòtesi com a certa. Aquesta és la hipòtesi nul·la (H_0). Complementant la H_0 també es planteja la hipòtesi alternativa (H_1), que estableix que les proporcions poblacionals són diferents.

$$H_0: \pi_j = \pi_{50} \Leftrightarrow \pi_j - \pi_{50} = 0$$

$$H_1: \pi_j \neq \pi_{50} \Leftrightarrow \pi_j - \pi_{50} \neq 0$$

Pas 2: càlcul de l'estadístic de prova (EP)

Com ja s'ha comentat, es calcula l'estadístic de prova suposant que la hipòtesi nul·la és certa. L'estadístic de prova que es calcula utilitza l'aproximació a la distribució normal de la diferència mostral de proporcions, atès que són mostres grans i independents.

$$EP = \frac{p_j - p_{50}}{\sqrt{p \cdot (1 - p) \cdot \left(\frac{1}{n_j} + \frac{1}{n_{50}}\right)}} \sim N(0,1)$$

p_j és la proporció mostral de joves que estan a favor de l'impost: $p_j = \frac{34}{80} = 0,425$.
 p_{50} és la proporció mostral de persones al voltant de la cinquantena que estan a favor de l'impost: $p_{50} = \frac{40}{106} = 0,3774$.

p és la proporció de persones que es declaren a favor del nou impost en el conjunt dels dos grups d'edat:

$$p = \frac{p_j \cdot n_j + p_{50} \cdot n_{50}}{n_j + n_{50}}$$

$$p = \frac{0,425 \cdot 80 + 0,3774 \cdot 106}{80 + 106} = \frac{34 + 40}{186} = 0,3978$$

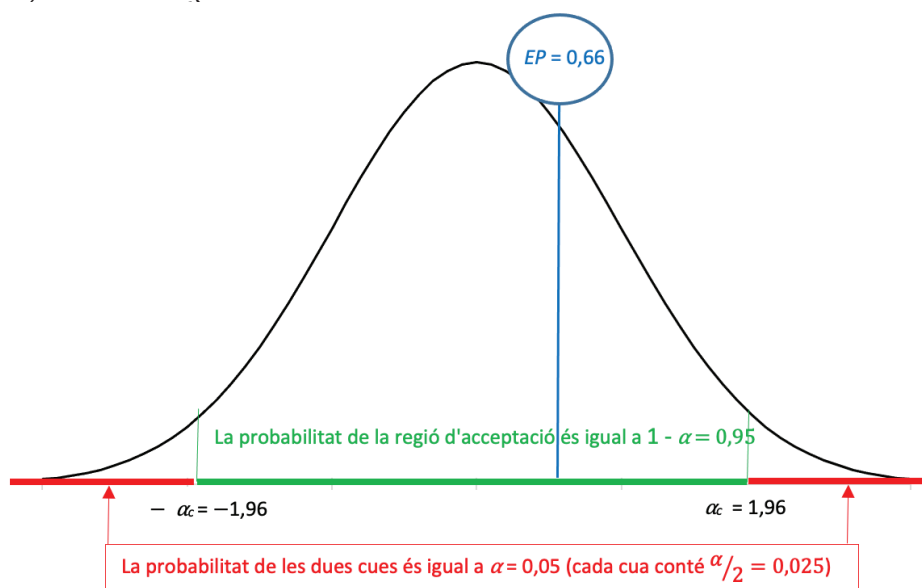
$$EP = \frac{0,425 - 0,3774}{\sqrt{0,3978 \cdot (1 - 0,3978) \left(\frac{1}{80} + \frac{1}{106} \right)}} = 0,6572$$

$$EP = 0,66$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de l'estadístic de prova suposant que la hipòtesi nul·la és verdadera es divideix en dues regions. Atès que la hipòtesi alternativa és bilateral ($\neq$), la regió crítica està formada per les dues cues i la regió d'acceptació, per la part central, amb una probabilitat centrada ($1 - \alpha = 0,95$). Hi ha dos valors crítics, un de positiu ($\alpha_c = 1,96$) i un de negati



La regió crítica és la zona en vermell i regió d'acceptació és la zona en verd. Com es pot veure en la gràfica, l'estadístic de prova és a la regió d'acceptació.

$$EP \in \text{Regió d'acceptació}$$

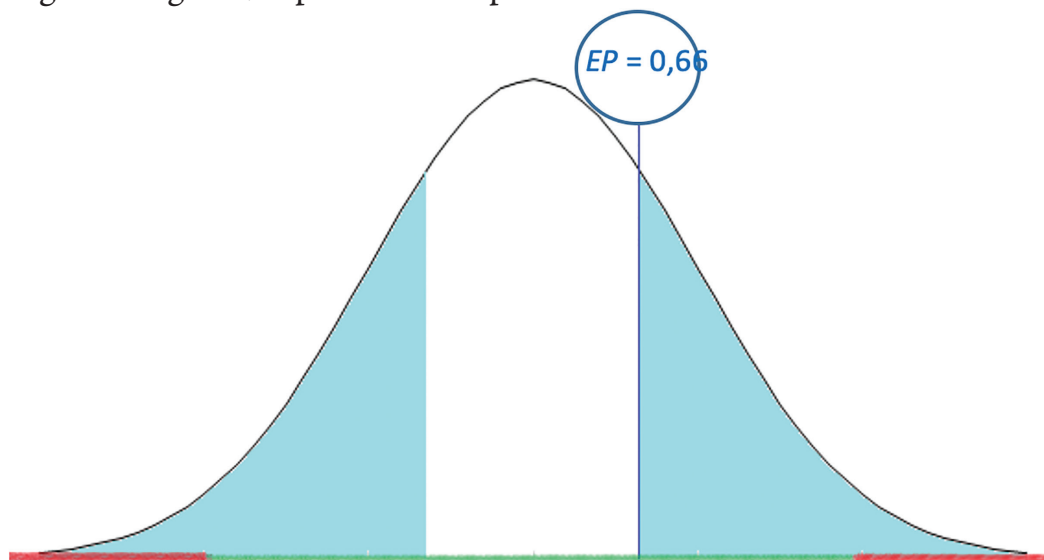
Opció b: p-valor

El p-valor és la probabilitat d'observar un estadístic de prova com el que s'ha calculat o un d'encara més atípic, en el supòsit que la hipòtesi nul·la és certa.

Atès que és un contrast bilateral, el p-valor és la suma de les dues cues de la distribució normal estàndard: el valor de l'àrea que hi ha sota la corba que queda a la dreta de l'estadístic de prova i el valor de la cua simètrica $[P(Z > |EP|)]$.

$$P(Z > 0,66) + P(Z < -0,66) = 0,5110$$

En la gràfica següent, el p-valor correspon a la zona de color blau clar.



Per comparar els dos procediments, també es marquen la regió crítica i la regió d'acceptació, que s'han explicat en l'opció a de la resolució. La regió crítica i la d'acceptació són en vermell i verd, respectivament.

$$p\text{-valor} = 0,51 > \alpha$$

Aquest valor es compara amb el nivell de significació. En aquest cas, el p-valor és molt superior al nivell de significació si $\alpha = 0,05$.

Pas 4: conclusió

Atès que $EP \in$ regió d'acceptació o bé atès que $p\text{-valor} > \alpha$, es conclou que per a un nivell de significació $\alpha = 0,05$ no hi ha evidència suficient per rebutjar la hipòtesi nul·la, és a dir, l'opinió de les persones joves respecte a la introducció del nou impost no és diferent de la de les persones que ronden els 50 anys.

Resolució de la pregunta b

El possible error que es pot estar cometent és no rebutjar la hipòtesi nul·la, quan en realitat és falsa. És a dir, no es detecten diferències quan en realitat sí que n'hi ha. Aquest error és conegut com a *error de tipus II*, *error β* o *error de segona espècie*.

Cervezes sense alcohol

Se lleva a cabo una encuesta (muestra aleatoria simple) sobre las preferencias de los consumidores respecto a tres marcas de cerveza sin alcohol en dos comunidades autónomas.

La información muestral es la siguiente:

	CA1	CA2
Marca A	44	50
Marca B	50	74
Marca C	66	76

A partir de la información anterior y con el nivel de significación $\alpha = 0.05$: ¿La preferencia por una determinada marca depende de la comunidad autónoma?

Explicar con detalle el contenido de H_0 y H_1 ; definir el estadístico de prueba y enunciar sus características, y efectuar el contraste.

Es duu a terme una enquesta (mostra aleatòria simple) sobre les preferències dels consumidors respecte a tres marques de cervesa sense alcohol en dues comunitats autònomes.

La informació mostral és la següent:

	CA1	CA2
Marca A	44	50
Marca B	50	74
Marca C	66	76

A partir de la informació anterior i amb el nivell de significació $\alpha = 0,05$, la preferència per una determinada marca depèn de la comunitat autònoma?

Expliqueu el contingut de H_0 i H_1 , definiu l'estadístic de prova i enumereu-ne les característiques, i realitzeu el contrast.

Resolució

Per valorar si les preferències per les diferents marques de cervesa estan relacionades amb la comunitat autònoma dels consumidors, una possibilitat és realitzar un contrast d'independència entre les dues variables categòriques, que es coneix com a *prova d'independència de χ^2* . Aquestes variables són, d'una banda, la comunitat autònoma, que pren dos valors (CA1 i CA2), i, de l'altra, les marques de cervesa, que prenen tres valors (marca A, marca B i marca C). Amb aquesta fi, es calcula la taula de valors esperats en cas que siguin independents i es realitza el contrast d'independència comparant els valors observats amb els esperats. Si els valors observats i els valors esperats en cas que siguin independents són massa diferents, s'assumeix que no són independents, és a dir, que les preferències depenen de la comunitat autònoma.

Es planteja un contrast d'hipòtesis —en aquest cas no és paramètric, perquè no es refereix a paràmetres concrets de la distribució poblacional— en quatre passos. En el primer es plantegen la hipòtesi que s'ha de contrastar, hipòtesi nul·la (H_0), i la hipòtesi alternativa (H_1). En el segon es calcula l'estadístic de prova. En el tercer es valora la credibilitat de l'estadístic de prova suposant que la hipòtesi nul·la és certa. Finalment, en el quart pas s'arriba a la conclusió.

Pas 1: plantejament de les hipòtesis

En aquest cas, la hipòtesi nul·la (H_0) és que les dues variables són independents entre elles, mentre que l'alternativa (H_1) és que no són independents entre elles.

H_0 : les variables són independents entre elles
 H_1 : les variables no són independents entre elles

Pas 2: càlcul estadístic de prova (EP)

Es calcula l'estadístic de prova:

$$EP = \sum_{\forall i} \sum_{\forall j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(I-1) \cdot (J-1)} \text{ graus de llibertat}$$

O_{ij} és el valor observat de la cel·la de la i -èsima fila i la j -èsima columna; són els valors mostrals que facilita l'enunciat.

E_{ij} és el valor esperat de la cel·la de la i -èsima fila i la j -èsima columna en cas que siguin independents: $E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$, on $n_{i.}$ és el nombre total d'observacions de la fila i , $n_{.j}$ és el nombre total d'observacions de la columna j i n és el nombre total d'observacions de la taula.

I és el nombre de files (hi ha 3 files) i J és el total de columnes (hi ha 2 columnes).

$\chi^2_{(I-1) \cdot (J-1)}$ graus de llibertat és la distribució de la khi-quadrat amb $(3 - 1) \cdot (2 - 1)$ graus de llibertat que segueix l'estadístic de prova.

La taula de valors observats és:

	CA1	CA2	Total
Marca A	44	50	94
Marca B	50	74	124
Marca C	66	76	142
Total	160	200	360

La taula de valors esperats en cas que siguin independents és:

	CA1	CA2
Marca A	$\frac{94 \cdot 160}{360} = 41,78$	$\frac{94 \cdot 200}{360} = 52,22$
Marca B	$\frac{124 \cdot 160}{360} = 55,11$	$\frac{124 \cdot 200}{360} = 68,89$
Marca C	$\frac{142 \cdot 160}{360} = 63,11$	$\frac{142 \cdot 200}{360} = 78,89$

Es pot observar que tots els valors esperats són superiors a 5 (condició necessària per dur a terme el contrast).

$$EP = \frac{(44 - 41,78)^2}{41,78} + \dots + \frac{(76 - 78,89)^2}{78,89}$$

$$EP = 1,3040$$

Pas 3: aquest pas es pot dur a terme o bé buscant la regió crítica i la regió d'acceptació, o bé buscant el p-valor.

Opció a: regió crítica i regió d'acceptació

La distribució de la χ^2 no és simètrica i pren només valors positius. En cas d'independència (hipòtesi recollida per H_0), els valors observats i esperats serien iguals i, per tant, si es calculés l'estadístic de prova amb les dades de tota la població el resultat seria igual

a 0. Fruit de l'atzar en prendre una mostra, els valors observats poden ser lleugerament diferents dels esperats. Si els valors són molt diferents, aleshores es pot pensar que potser la hipòtesi nul·la no és correcta. L'estadístic de prova recull aquestes diferències per al conjunt de la taula: com més gran sigui, menys creïble és la hipòtesi d'independència. Per aquest motiu, la regió crítica només és a la cua dreta de la distribució, per a valors alts de l'estadístic de prova. En aquesta prova, el valor crític per a un nivell de significació $\alpha = 0,05$ és doncs un únic valor que deixa a l'esquerra una probabilitat acumulada de 0,95 i a la dreta de 0,05, cosa que denota que les diferències entre les dues taules són grans. El valor crític en una χ^2 amb 2 graus de llibertat és igual a $\alpha_c = 5,9915$. La regió per sota d'aquest valor és la regió d'acceptació; per sobre d'aquest valor hi ha la regió crítica. L'estadístic de prova és a la regió d'acceptació.

$$EP < \alpha_c$$

Opció b: p-valor

En aquest contrast, el p-valor és igual a l'àrea que queda per sobre de l'estadístic de prova en una distribució de la probabilitat χ^2 amb 2 graus de llibertat.

$$p\text{-valor}=0,5210 > \alpha$$

Pas 4: conclusió

Atès que $EP \in$ regió d'acceptació o bé atès que $p\text{-valor} > \alpha$, es conclou que per a un nivell de significació $\alpha = 0,05$ no hi ha evidència suficient per rebutjar la hipòtesi nul·la i, per tant, no hi ha cap evidència per dir que la preferència per una marca determinada depèn de la comunitat autònoma.

Referències a altres llibres

Atès que es tracta d'un llibre d'exercicis fet per practicar, no s'hi ha inclòs cap apartat de bibliografia. Tanmateix, s'esmenten dues referències a altres llibres d'exercicis fets a la unitat d'estadística del Departament d'Economia de la URV. En el més antic dels dos, hi va participar el professor Juan Manuel López-Rey Laurens.

ALLEPÚS, J.; CORBELLÀ, T.; DOMINGO, C.; LÓPEZ-REY, J. M.; MANJÓN, M. (2002): *Exercicis d'inferència estadística*. Col·lecció Eina 22, Cossetània Edicions.

DOMINGO, C. (coord.) (2006): *Exercicis d'estadística empresarial*. Col·lecció Eina 36, Cossetània Edicions.

Índex per temes

ESPERANÇA MATEMÀTICA, VARIÀNCIA I MEDIANA D'UNA VARIABLE ALEATÒRIA: CÀLCUL I PROPIETATS

- El programa
- Rifa per a la biblioteca (variable aleatòria discreta)
- Demanda diària d'un producte (variable aleatòria discreta i contínua)
- Detergent per a rentadores (variable aleatòria contínua)
- Torre de Babel
- Abraçades (variable aleatòria contínua)
- Avets a la fira de Santa Llúcia
- Barra de sushi

DESIGUALTAT DE TXEVIXV Revetlla de Sant Joan

DISTRIBUCIÓ UNIFORME El programa Joc de curses de cavalls Videojocs

DISTRIBUCIÓ BINOMIAL Torre de Babel Peres per sopar Joc de curses de cavalls

Distribució de Poisson i distribució exponencial Abraçades Mona de Pasqua

Distribució normal Avets a la fira de Santa Llúcia (propietat additiva) Mona de Pasqua (propietat additiva) Peres per sopar (propietat additiva) Videojocs (propietat additiva i teorema central del límit) Torre de Babel (teorema central del límit) Joc de curses de cavalls (teorema central del límit) Revetlla de Sant Joan (teorema central del límit)

INTERVALS DE CONFIANÇA

A cal perruquer (una mitjana poblacional, mostra gran)

Matràs (una mitjana poblacional, mostra petita)

Missatges de telèfon (diferència de mitjanes poblacionals amb canvis del nivell de confiança, mostres grans)

Barra de sushi (una mitjana poblacional i diferència de mitjanes poblacionals, mostres petites)

Xocolata desfeta (una proporció poblacional amb canvis de la grandària mostral)

Canvis socials (dues proporcions poblacionals)

CONTRAST D'HIPÒTESIS

A cal perruquer (una mitjana poblacional, mostra gran)

Matràs (una mitjana poblacional, mostra petita)

Missatges de telèfon (diferència de mitjanes poblacionals, mostres grans, i tipus d'errors)

Barra de sushi (diferència de mitjanes poblacionals, mostres petites)

Xocolata desfeta (una proporció poblacional)

Canvis socials (dues proporcions poblacionals)

Cervesa sense alcohol (contrast d'independència de dues variables categòriques)

Aquesta publicació ofereix una selecció d'exercicis resolts dissenyats per afavorir la consolidació dels continguts de l'assignatura d'Estadística II. Amb un enfocament pràctic i rigorós, inclou problemes de probabilitats, distribucions i inferència estadística, resolts pas a pas. Tot i que ficticis, els enunciats es presenten en contextos versemblants i propers que en faciliten la comprensió i n'il·lustren l'aplicació pràctica. Les solucions, acompanyades de gràfics, taules i explicacions detallades, constitueixen un recurs de gran utilitat per a l'aprenentatge i per a l'aprofundiment en els conceptes fonamentals de l'estadística inferencial.