Laura Frau
http://gemma.atipic.net

Anàlisi de Dades I
AC3 –
Entrega

CORRECCIÓ PAC 3 (Consultor: Antoni Cosculluela Mas)

CORRECCIÓ PAC3 (Consultor: Antoni Cosculluela Mas):

Enunciat A

Veure enunciat i dades de la PAC 1.

1. Representeu els diagrames de caixa i de tiges i fulles de la variable DESPESA.

   Diagrama de caixa:

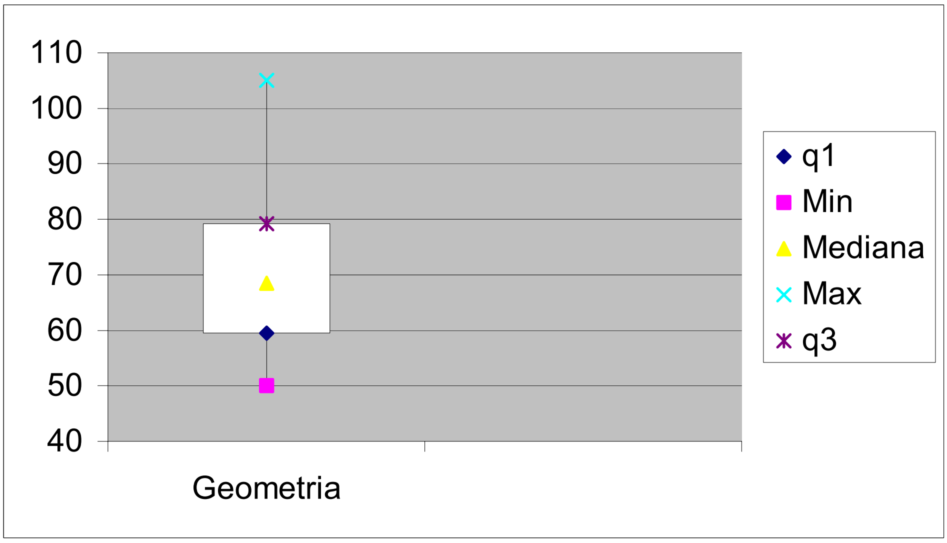

Diagrama de tija i fulles:

5	00668
6	00345
7	25689
8	014
10	05

Enunciat B

Amb el mateix context, veure l'enunciat de la PAC1, i seguim amb les condicions de la PAC2. Sabem que en la població, la distribució de la variable DESPESA segueix una llei normal N (μ=70, σ=15)

2. Quina es la probabilitat de que un subjecte triat a l'atzar tingui una despesa entre 55 i 95 Euros (variable DESPESA)

   Tenim que obtenir: P (55 < X < 95)

   En primer lloc tenim que estandarditzar les puntuacions directes de 55 i 95

$$Z = \frac{X-\mu}{\sigma} = \frac{55-70}{15} = -1$$

$$Z = \frac{X-\mu}{\sigma} = \frac{95-70}{15} = 1{,}67$$

Llavors sabem que P(-1 < Z < 1,67) = P (Z < 1,67) - P(Z < -1) i busquem a la taula de les àrees sota la corba normal estàndard (Taula 3 de l'Annex) :

P (Z < 1,67) = 0,9525

P (Z < -1) = 0,1587

I, per tant, P (Z < 1) - P (Z < -1) = 0,9525 – 0,1587 = 0,7938

Així, la probabilitat de que un subjecte triat a l'atzar tingui una despesa entre 51 i 95 Euros és de 0,7938.

3. Quins són els tres quartils de la distribució de la variable DESPESA?. Entre quins valors de la variable es troba el 50% central dels subjectes d'aquesta població?.

Per obtenir el quartil 1 tenim que buscar en les taules de la llei normal, la puntuació estandarditzada (z) que deixa per sota un 25% de la distribució (p=0,25).

Així tenim que: p = 0,25 → z = -0,67

Ara podem obtenir la puntuació de la variable DESPESA que corresponen a aquesta puntuació estandarditzada.

$$-0{,}67 = \frac{x-70}{15} \rightarrow x = 70 - 0{,}67 \times 15 = 59{,}95$$

El quartil 2 correspon a la mediana i coincideix amb la mitjana aritmètica (ja que la distribució normal és perfectament simètrica).

Així el quartil 2 és igual a 70

Per obtenir el quartil 3 tenim que buscar en les taules de la llei normal, la puntuació estandarditzada (z) que deixa per sota un 75% de la distribució (p=0,75).

Així tenim que: p = 0,75 → z = 0,67

Ara podem obtenir la puntuació de la variable DESPESA que corresponen a aquesta puntuació estandarditzada.

$$0{,}67 = \frac{x-70}{15} \rightarrow x = 70 + 0{,}67 \times 15 = 80{,}05$$

Per tant, el quartil 1 de la variable DESPESA és el valor 59,95, el quartil 2 el 70 i el quartil 3 el 80,05

El 50% centrals de subjectes d'aquesta població es trobaran, en la variable DESPESA, entre 59,95 i 80,05 (entre el quartil 1 i el 3).

4. Si agafem mostres de grandària 100 subjectes. Entre quins valors de la variable DESPESA es trobaran el 90% central de la distribució de las mitjanes observades en mostres de grandària100, extretes de la població?

La mitjana de la distribució mostral de mitjanes de la variable DESPESA serà 70 (igual a la mitjana poblacional), i la desviació estàndard d'aquesta distribució mostral de mitjanes serà: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1{,}5$

Per altre banda, tenim que obtenir en les taules de la llei normal, les puntuacions estandarditzada (z) que deixen un 90% central entre elles. Així, en el límit superior tindrem la puntuació estandarditzada que deixa per sota 0,95 $\left(1 - \frac{1-0{,}90}{2}\right)$, i en el límit inferior la que deixa per sota un 0,05 $\left(\frac{1-0{,}90}{2}\right)$:

Així tenim que: p = 0,95 → z = 1,64 i p = 0,05 → z = -1,64

Ara podem obtenir els valors de la distribució mostral de mitjanes de la variable DESPESA que corresponen a aquestes dues puntuacions estandarditzades.

$$1{,}64 = \frac{\bar{x}-70}{1{,}5} \rightarrow \bar{x} = 70 + 1{,}64 \times 1{,}5 = 72{,}46$$

$$-1{,}64 = \frac{\bar{x}-70}{1{,}5} \rightarrow \bar{x} = 70 - 1{,}64 \times 1{,}5 = 67{,}54$$

Per tant, en la distribució mostral de mitjanes de la variable DESPESA, el 90% central dels valors es troben entre 72,46 i 67,54 punts.

5. Si agafem mostres de grandària 100 subjectes. Quina serà la probabilitat que triant a l'atzar una d'aquestes mostres, la despesa mitjana superi els 72 Euros (variable DESPESA)

   Tenim que obtenir $p(\bar{x} > 72)$

   $$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{72 - 70}{1,5} = 1,33 \rightarrow p(z > 1,33) = 0,09$$

   Si agafem mostres de grandària 100 subjectes, la probabilitat que triant a l'atzar una d'aquestes mostres, la seva mitjana superi els 72 Euros a la variable DESPESA és de 0,09.

Enunciat C

En una segona fase es tria una mostra de 500 subjectes. S'ha obtingut una puntuació mitjana en la variable DESPESA de 71 Euros. Sabem que la variància de la població (σ2) es de 225 Euros2.

6. Estimar la mitjana poblacional amb un nivell de confiança del 95% (α=5%), i amb un nivell de confiança del 90% (α=10%). Canvien els límits del interval de confiança? Per que?

   Hem de fer una estimació per interval de la mitjana poblacional de la variable DESPESA, coneguda la desviació estàndard de la població.

   Tenim que calcular l'error típic de la mitjana: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{225}}{\sqrt{500}} = 0,67$

   Calculem el marge d'error com z* per l'error típic: $z * \sigma_{\bar{x}}$

   Per un nivell de confiança del 95% (α=5%), $z * \sigma_{\bar{x}} = 1,96 \times 0,67 = 1,31$

   Per un nivell de confiança del 90% (α=10%), $z * \sigma_{\bar{x}} = \mathbf{1,64 \times 0,67 = 1,10}$

   Tenim així que l'interval de confiança és la mitjana més menys el marge d'error: $\bar{x} \pm z * \sigma_{\bar{x}}$

   Per un nivell de confiança del 95% (α=5%):

   $$71 \pm 1,31 \rightarrow 69,69 \Leftrightarrow 72,31$$

   Per un nivell de confiança del 90% (α=10%):

   $$\mathbf{71 \pm 1,10 \rightarrow 69,90 \Leftrightarrow 72,10}$$

   Els límits de l'interval de confiança canvien, així amb un nivell de confiança del 95% aquest límits son més amplis que amb un nivell de confiança del 90%. Si augmentem el nivell de confiança, disminueix la precisió de l'estimació.

7. Quants subjectes hauria de tenir la mostra per a poder fer una estimació per interval, amb un nivell de confiança del 95% i una precisió de 1 Euro?

   Mida de la mostra: $1,96^2 \frac{225}{1} = 864$

   La mostra per a poder fer una estimació per interval, amb un nivell de confiança del 95% i una precisió de 1 punt, ha de tenir 864 subjectes

Enunciat D

L'anterior mostra se ha triat entre els nivells de formació mencionats a la PAC1. Els que han estudiat FP representen el 40% del total de la població.

8. Calcular la probabilitat de que triant a l'atzar aquesta mostra, menys del 30% hagin estudiat FP

   p (FP) = π = 0,4

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{\mathbf{0{,}4 \times 0{,}6}}{\mathbf{500}}} = \mathbf{0{,}02}$$

   Busquem (p < 0,30):

$$Z = \frac{p-\pi}{\sigma_P} = \frac{\mathbf{0{,}3 - 0{,}4}}{\mathbf{0{,}02}} = \mathbf{-5} \to p(z < \mathbf{-5}) \approx \mathbf{0}$$

   La probabilitat de que triant a l'atzar aquesta mostra, menys del 30% hagin estudiat FP és pràcticament de 0.

9. Calcular la probabilitat de que, en les mateixes condicions, els que hagin estudiat FP es trobin entre el 30% i el 50%.

   p (FP) = π = 0,4

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{\mathbf{0{,}4 \times 0{,}6}}{\mathbf{500}}} = \mathbf{0{,}02}$$

   Busquem p (0,30 < p < 0,50):

$$Z = \frac{p-\pi}{\sigma_P} = \frac{\mathbf{0{,}3 - 0{,}4}}{\mathbf{0{,}02}} = \mathbf{-5} \to p(z < \mathbf{-5}) \approx \mathbf{0}$$

$$Z = \frac{p-\pi}{\sigma_P} = \frac{\mathbf{0{,}5 - 0{,}4}}{\mathbf{0{,}02}} = \mathbf{5} \to p(z < \mathbf{5}) \approx \mathbf{1}$$

   p (0,30 < p < 0,50) $\approx$ 1

   La probabilitat de que, en les mateixes condicions, els que hagin estudiat FP es trobin entre el 30% i el 50% és pràcticament 1.