Anàlisi de Dades I
AC2 –
Entrega

CORRECCIÓ PAC 2 (Consultor: Antoni Cosculluela Mas)

CORRECCIÓ PAC2 (Consultor: Antoni Cosculluela Mas):

Instruccions pel lliurament de l'activitat:

Pel que fa al context, veure l'enunciat de la PAC1.

Sabem que a la població la distribució de la variable OPINIO (veure PAC1) segueix una llei normal N (μ=3, σ=1,1) la variable EDAT una llei normal de una llei normal N (μ=40, σ=4) i la variable DESPSESA una llei normal N (μ=70 , σ=15)

Enunciat

1. Quin percentatge de subjectes d'aquesta població tindrà una puntuació superior a 4 punts a la variable OPINIO?

   X: OPINIO μ=3 σ=1,1

   Tenim que obtenir: P (X > 4)

   En primer lloc tenim que estandarditzar la puntuació directa de 4 punts

$$Z = \frac{X-\mu}{\sigma} = \frac{\mathbf{4-3}}{\mathbf{1{,}1}} = \mathbf{0{,}91}$$

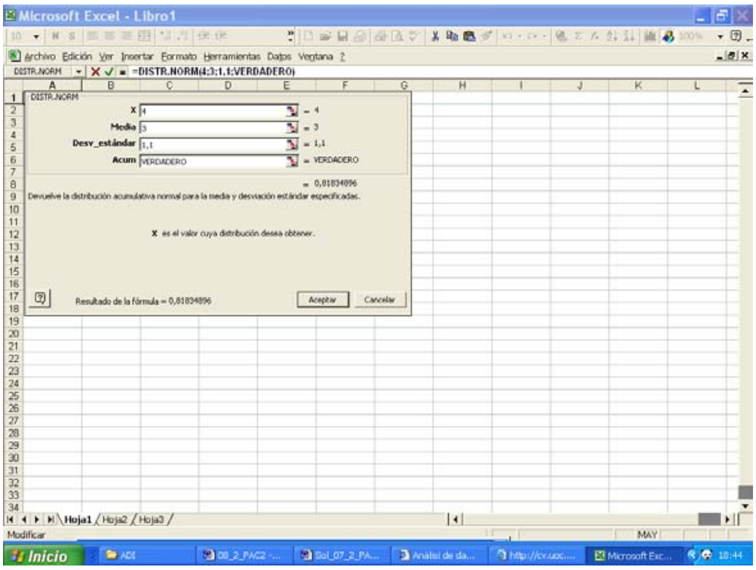
   Llavors sabem que P (Z > 0,91)

   i busquem a la taula de les àrees sota la corba normal estàndard (Taula 3 de l'Annex) :

   P (Z > 0,91) = 0,1814

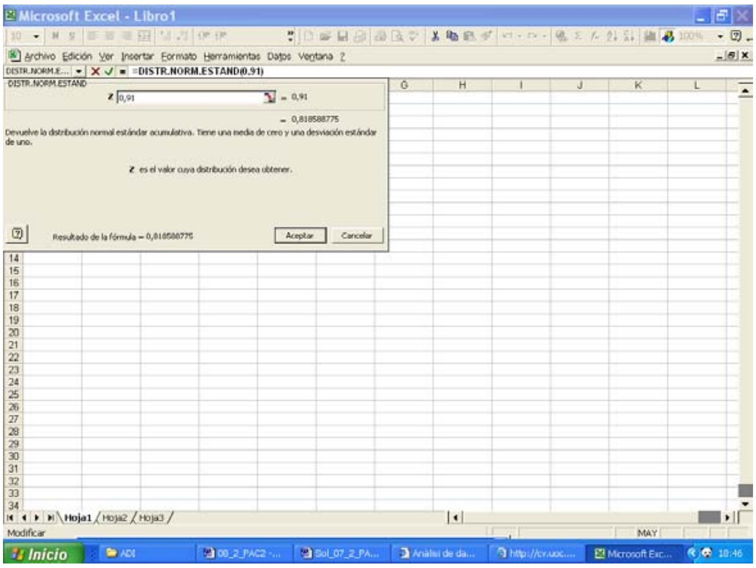
   Aquest resultat està expressat en probabilitats o proporcions, i com ens ho demanen en percentatges, només tindrem que multiplicar-lo per 100.

   També ho podíem obtenir amb la funció de l'Excel "DISTR.NORM":



que com veiem ens dona: P (X < 4) = 0,8183

I, per tant, P (X > 4) = 1 – 0,8183 = 0,1817 o amb la funció "DISTR.NORM.ESTAND" amb Z = 0,91, que ens dona el mateix resultat.



Així, un 18% de subjectes d'aquesta població tindrà una puntuació superior a 4 punts a la variable OPINIO.

2. Quina és la probabilitat de que un subjecte triat a l'atzar d'aquesta població tingui una puntuació entre 2 punts i 5 punts a la variable OPINIO?

   Tenim que obtenir: P (2 < X < 5)

   En primer lloc tenim que estandarditzar les puntuacions directes de 2 i 5

$$Z = \frac{X-\mu}{\sigma} = \frac{2-3}{1{,}1} = -0{,}91 \qquad\qquad Z = \frac{X-\mu}{\sigma} = \frac{5-3}{1{,}1} = 1{,}82$$

   Llavors sabem que P(-0,91 < Z < 1,82) = P (Z < 1,82) - P(Z < -0,91)

   i busquem a la taula de les àrees sota la corba normal estàndard (Taula 3 de l'Annex), o amb l'excel:

   P (Z < 1,82) = 0,9656 P (Z < -0,91) = 0,1814

   I, per tant, P (Z < 1,82) - P (Z < -0,91) = 0,9656 – 0,1814 = 0,7841

   Així, la probabilitat de que un subjecte triat a l'atzar d'aquesta població tingui una puntuació entre 2 punts i 5 punts a la variable OPINIO és de 0,7841.

3. Un subjecte ha obtingut les següents puntuacions:

   OPINIO – 4,

   EDAT – 35,

   DESPESA – 60,

   En quina de les tres variables ha obtingut una puntuació més gran amb relació a la població de referència?

   (SUGGERIMENT: calcula el percentil de cada puntuació)

   En primer lloc podem obtenir la puntuació estandarditzada d'aquest subjecte per cada una de les tres variables:

   OPINIO: $$z = \frac{x-\mu}{\sigma} = \frac{4-3}{1{,}1} = 0{,}91$$

   EDAT: $$z = \frac{x-\mu}{\sigma} = \frac{35-40}{4} = -1{,}25$$

   DESPESA: $$z = \frac{x-\mu}{\sigma} = \frac{60-70}{15} = -0{,}67$$

Ara, mitjançant les taules de la llei normal podem obtenir el percentil que correspon a cada puntuació estandarditzada:

OPINIO: $z = 0{,}91 \rightarrow p = 0{,}8186$, o sigui Percentil 82

EDAT: $z = -1{,}25 \rightarrow p = 0{,}1056$, o sigui Percentil 11

DESPESA: $z = -0{,}67 \rightarrow p = 0{,}2514$, o sigui Percentil 25

Per tant, en la variable OPINIO és on el subjecte ha obtingut una puntuació més gran en relació a la població de referència.

4. Entre quins valors de la variable OPINIO es troba el 95% central dels subjectes d'aquesta població?.

Tenim que obtenir en les taules de la llei normal, les puntuacions estandarditzada (z) que deixen un 95% central entre elles. Així, en el límit superior tindrem la puntuació estandarditzada que deixa per sota 0,975 $\left(1-\frac{1-0{,}95}{2}\right)$, i en el límit inferior la que deixa per sota un 0,025 $\left(\frac{1-0{,}95}{2}\right)$:

Així tenim que: p = 0,975 → z = 1,96 i p = 0,025 → z = -1,96

Ara podem obtenir les puntuacions de la variable OPINIO que corresponen a aquestes dues puntuacions estandarditzades.

$$1{,}96 = \frac{x-3}{1{,}1} \rightarrow x = 3 + 1{,}96 \times 1{,}1 = 5{,}16$$

$$-1{,}96 = \frac{x-3}{1{,}1} \rightarrow x = 3 - 1{,}96 \times 1{,}1 = 0{,}84$$

Per tant, en la variable OPINIO, el 95% central dels subjectes d'aquesta població es troba entre 0,84 I 5 punts (ja que l'escala de la variable té com a límit superior els 5 punts).

5. Quins són els tres quartils de la distribució de la variable OPINIO?. 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 OPINIO que corresponen a aquesta puntuació estandarditzada.

$$-0{,}67 = \frac{x-3}{1{,}1} \rightarrow x = 3 - 0{,}67 \times 1{,}1 = 2{,}26$$

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 3

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 = 0,25 → z = 0,67

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

$$0{,}67 = \frac{x-3}{1{,}1} \rightarrow x = 3 + 0{,}67 \times 1{,}1 = 3{,}74$$

Per tant, el quartil 1 de la variable OPINIO és el valor 2,26, el quartil 2 el 3 i el quartil 3 el 3,74.

El 50% centrals de subjectes d'aquesta població es trobaran, en la variable OPINIO, entre 2,26 i 3,74 (entre el quartil 1 i el 3).

6. Si es considera que el 15% dels subjectes amb puntuacions més altes es poden considerar amb un nivell alt amb "OPINIO", quina es la puntuació que determinarà el punt de tall que permet classificar als subjectes com a "subjectes amb nivell alt a la variable OPINIO"?.

   Per obtenir el 15% dels subjectes amb puntuacions més altes, tenim que buscar en les taules de la llei normal, la puntuació estandarditzada (z) que deixa per sota un 85% de la distribució (p = 1 - 0,15).

   Així tenim que: p = 0,85 → z = 1,03

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

$$1{,}03 = \frac{x-3}{1{,}1} \rightarrow x = 3 + 1{,}03 \times 1{,}1 = 4{,}13$$

   Per tant, la puntuació que determinarà el punt de tall que permet classificar als subjectes com a "subjectes amb nivell alt a la variable OPINIO serà els 4,13 punts.

7. La probabilitat de que al triar un subjecte de la mostra sigui home es 0,5. Si tenim una mostra de 5 subjectes, calcular:

   Aquest cas s'ajusta a una distribució binomial amb paràmetres:

   n = 5 i π = 0,5

   a. La probabilitat de que els 5 subjectes siguin dones

      Busquem p (X = 5):

$$p(X=5) = \binom{5}{5} 0{,}5^5 (1-0{,}5)^0 = 0{,}5^5 = 0{,}03125$$

      Així, la probabilitat de que els 5 subjectes siguin dones és de 0,03

   b. La probabilitat de que, como mínim, 3 subjectes siguin homes

      Busquem p (X ≥ 3):

$$p(X \geq 3) = \sum_{x=3}^{x=5} \binom{5}{x} 0{,}5^x (1-0{,}5)^{5-x} = \binom{5}{3} 0{,}5^3 (1-0{,}5)^2 + \binom{5}{4} 0{,}5^4 (1-0{,}5) + \binom{5}{5} 0{,}5^5 =$$

$$0{,}3125 + 0{,}15625 + 0{,}03125 = 0{,}5$$

      Així, la probabilitat de que, com mínim, 3 subjectes siguin homes és de 0,5

8. La probabilitat de que al triar un subjecte de la mostra estigui cassat – e parella es 0,6. Es trien 3 subjectes.

   a. Troba la distribució de la variable aleatòria "nombre de subjectes triats No casats".

      Els valors que pot tenir aquesta variable aleatòria son: 0, 1, 2 i 3

      La variable segueix una llei o distribució binomial amb paràmetres:

      n = 3 i π = 0,4 (probabilitat de que el subjecte triat no estigui casat).

      Per tant, la distribució de la variable serà:

$$p(X=0) = \binom{3}{0} 0{,}4^0 (1-0{,}4)^3 = 0{,}216$$

$$p(X=1) = \binom{3}{1} 0{,}4^1 (1-0{,}4)^2 = 0{,}432$$

$$p(X=2) = \binom{3}{2} 0{,}4^2 (1-0{,}4)^1 = 0{,}288$$

$$p(X=3) = \binom{3}{3} 0{,}4^3 (1-0{,}4)^0 = 0{,}064$$

Així, la distribució de la variable aleatòria "nombre de subjectes triats No casats" serà:

X	f(X)
0	0,216
1	0,432
2	0,288
3	0,064

b. Troba la esperanza matemàtica y la variancia d'aquesta variable aleatòria.

Esperança matemàtica: $E(X) = \mu_x = n.\pi = \mathbf{3 \times 0{,}4 = 1{,}2}$

Variància: $Var(X) = \sigma_x^2 = n.\pi.(\mathbf{1} - \pi) = \mathbf{3 \times 0{,}4 \times 0{,}6 = 0{,}72}$

c. Troba la Distribució acumulada.

La Distribució acumulada de la variable aleatòria "nombre de subjectes triats no casats" serà:

X	F(X)
0	0,216
1	0,648
2	0,936
3	1