Solució Control 4 Aplicacions Estadístiques. Data: 23/05/2011

Problema 1 *Suposem que un* 60 % *dels universitaris opinen que són millors les pizzes de salami que les de roquefort.*

a) Quina és la probabilitat que més del 70 % *dels components d'una mostra de 200 universitaris siguin d'aquesta opinió?*

b) Quina és la probabilitat que menys del 50 % *dels components d'una mostra de 100 universitaris siguin d'aquesta opinió?*

c) Si la proporció d'universitaris que opinen que les pizzes de salami són millors que les de roquefort és un valor desconegut p, quin és el tamany mínim de la mostra que ens 'assegura' (amb probabilitat superior al 95 %*) que l'error comès en estimar p a partir de la proporció mostral és inferior a $0{,}01$? (Suposau que la mostra està formada per més de 30 persones).*

Solució

Tenim que la *proporció poblacional* d'universitaris que opinen que les pizzes de salami són millors que les de roquefort és $p = 0{,}6$.

a) $n = 200$, per tant la *proporció mostral* d'universitaris que opinen que les pizzes de salami són millors que les de roquefort és una variable aleatòria amb distribució:

$$\hat{p}_X \sim N(p, \frac{p(1-p)}{n}) = N(0{,}6, \frac{0{,}6(1-0{,}6)}{200}) = N(0{,}6, 0{,}0012) = N(0{,}6, 0{,}0346^2)$$

Ens demanen calcular $P(\hat{p}_X > 0{,}7)$:

$$P(\hat{p}_X > 0{,}7) = 1-P(\hat{p}_X \leq 0{,}7) = 1-F_{\hat{p}_X}(0{,}7) = 1-F_Z(\frac{0{,}7-0{,}6}{0{,}0346}) = 1-F_Z(2{,}89) = (\text{taula normal}) = 1-0{,}9981 = 0{,}0019$$

b) $n = 100$, per tant la *proporció mostral* d'universitaris que opinen que les pizzes de salami són millors que les de roquefort és una variable aleatòria amb distribució:

$$\hat{p}_X \sim N(p, \frac{p(1-p)}{n}) = N(0{,}6, \frac{0{,}6(1-0{,}6)}{100}) = N(0{,}6, 0{,}0024) = N(0{,}6, 0{,}049^2)$$

Ens demanen calcular $P(\hat{p}_X < 0{,}5)$:

$$P(\hat{p}_X < 0{,}5) = F_{\hat{p}_X}(0{,}5) = F_Z(\frac{0{,}5-0{,}6}{0{,}049}) = F_Z(-2{,}04) = 1-F_Z(2{,}04) = (\text{taula normal}) = 1-0{,}9793 = 0{,}0207$$

c) Ens demanen calcular el valor mínim de n que garanteix $P(|\hat{p}_X - p| < 0{,}01) > 0{,}95$

Com que suposam $n > 30$, la distribució de valors de $\hat{p}_X$ segueix una llei normal: $\hat{p}_X \sim N(p, \frac{p(1-p)}{n})$.

$$\begin{aligned} P(|\hat{p}_X - p| < 0{,}01) &= P(-0{,}01 < \hat{p}_X - p < 0{,}01) = P(p-0{,}01 < \hat{p}_X < p+0{,}01) = F_{\hat{p}_X}(p+0{,}01) - F_{\hat{p}_X}(p-0{,}01) = \\ &= F_Z(\tfrac{p+0{,}01-p}{\sqrt{p(1-p)/n}}) - F_Z(\tfrac{p-0{,}01-p}{\sqrt{p(1-p)/n}}) = F_Z(\tfrac{0{,}01}{\sqrt{p(1-p)/n}}) - F_Z(\tfrac{-0{,}01}{\sqrt{p(1-p)/n}}) = \\ &= F_Z(\tfrac{0{,}01}{\sqrt{p(1-p)/n}}) - (1-F_Z(\tfrac{0{,}01}{\sqrt{p(1-p)/n}})) = 2F_Z(\tfrac{0{,}01}{\sqrt{p(1-p)/n}}) - 1 \end{aligned}$$

Per tant:

$$2F_Z(\frac{0{,}01}{\sqrt{p(1-p)/n}}) - 1 > 0{,}95 \qquad \Longrightarrow \qquad F_Z(\frac{0{,}01}{\sqrt{p(1-p)/n}}) > \frac{1+0{,}95}{2} = 0{,}975$$

Mirant les taules de la normal:

$$\frac{0{,}01}{\sqrt{p(1-p)/n}} \geq 1{,}96$$

D'on deduïm:

$$n \geq \frac{1{,}96^2}{0{,}01^2} p(1-p)$$

Com que desconeixem p triam el valor de n que garanteix que la condició se cumplirà sigui quin sigui p. El màxim de la funció $\frac{1{,}96^2}{0{,}01^2}p(1-p)$ es troba per a $p = 0{,}5$, per tant el valor de n triat és:

$$n \geq \frac{1{,}96^2}{0{,}01^2} 0{,}5(1-0{,}5) = 9604$$

Problema 2 *Un conductor fa habitualment el trajecte entre Bunyola i la UIB i assegura que tarda una mitjana de 9 minuts en fer el trajecte. Per comprovar si es cumpleix l'afirmació d'aquest conductor es pren una mostra dels seus 7 darrers trajectes i s'obtenen els següents temps (en minuts):*

$$10{,}5 \quad 7{,}3 \quad 15{,}1 \quad 8{,}9 \quad 9{,}6 \quad 11{,}7 \quad 12{,}5$$

Suposant que el temps que tarda el conductor en fer el trajecte segueix una distribució normal feu un contrast d'hipòtesis per confirmar o rebutjar l'afirmació del conductor amb un nivell de significació del 10 %. *Justificau les hipòtesis utilitzades i calculau el p-valor del contrast.*

Solució

Volem fer un contrast d'hipòtesis per al valor de la *mitjana poblacional* μ. La mitjana, variància i desviació típica *mostrals* dels 7 trajectes són:

$$\bar{x} = \tfrac{10{,}5+7{,}3+\cdots+12{,}5}{7} = 10{,}8$$

$$\hat{s}_X^2 = \tfrac{7}{6}\left(\tfrac{10{,}5^2+7{,}3^2+\cdots+12{,}5^2}{7} - 10{,}8^2\right) = 6{,}6$$

$$\hat{s}_X = \sqrt{\hat{s}_X^2} = \sqrt{6{,}6} = 2{,}57$$

La hipòtesi nul.la del problema serà $H_0 : \mu = 9$. Com que $\bar{x} = 10{,}8 > 9$, prenim com a hipòtesi alternativa $H_1 : \mu > 9$.

Es tracta per tant d'un *contrast unilateral per la dreta*. Ens demanen un nivell de significació $\alpha = 0{,}1$, per tant el valor límit M entre les regions crítica i d'acceptació ha de cumplir:

$$P(\bar{X} > M) = 0{,}1$$

La mitjana mostral és una variable aleatòria que segueix una llei de Student (ja que σ és desconegut i $n < 30$):

$$\frac{\bar{X}-\mu}{\hat{s}_X/\sqrt{n}} \sim t_{n-1} \qquad \Longrightarrow \qquad \frac{\bar{X}-9}{2{,}57/\sqrt{7}} \sim t_6$$

Per tant:

$$P(\bar{X} > M) = 1 - P(\bar{X} \leq M) = 1 - P(\frac{\bar{X}-9}{2{,}57/\sqrt{7}} \leq \frac{M-9}{2{,}57/\sqrt{7}}) = 1 - P(t_6 \leq \frac{M-9}{2{,}57/\sqrt{7}})$$

Com s'ha de cumplir $P(\bar{X} > M) = 0{,}1$, llavors

$$1 - P(t_6 \leq \frac{M-9}{2{,}57/\sqrt{7}}) = 0{,}1 \qquad \longrightarrow \qquad P(t_6 \leq \frac{M-9}{2{,}57/\sqrt{7}}) = 0{,}9$$

Mirant la taula de la v.a. Student trobam $P(t_6 \leq 1{,}44) = 0{,}9$. Per tant deduïm:

$$\frac{M-9}{2{,}57/\sqrt{7}} = 1{,}44 \qquad \longrightarrow \qquad M = 9 + \frac{2{,}57}{\sqrt{7}} \cdot 1{,}44 = 10{,}398$$

Com que el valor mostral trobat està dins la zona crítica ($\bar{x} = 10{,}8 > 10{,}398$), llavors **rebutjam la hipòtesi nul.la** i acceptam com a correcta la hipòtesi alternativa $\mu > 9$.

El *p-valor* del contrast és:

$$\begin{aligned}\text{p-valor} &= P(\bar{X} > 10{,}8) = 1 - P(\bar{X} \leq 10{,}8) = 1 - P(\tfrac{\bar{X}-9}{2{,}57/\sqrt{7}} \leq \tfrac{10{,}8-9}{2{,}57/\sqrt{7}}) = \\ &= 1 - P(t_6 \leq 1{,}853) = (\text{taula Student}) \approx 1 - 0{,}95 = 0{,}05\end{aligned}$$