Tema 4. Proves d'hipòtesis en una població

1. Un fabricant de refrescos de cola sap que ha d'omplir les seves llaunes amb un contingut mitjà de 33cl. Si les omple amb menys quantitat pot ser multat per frau. Aquesta setmana un inspector ha passat per la planta d'envasament i ha recollit dades sobre el contingut de 25 llaunes. En les 25 llaunes que ha recollit l'inspector, ha resultat un contingut mitjà de 31.95cl i una desviació tipus de 2.25cl. Podem dir que el fabricant de refrescos omple les llaunes seguint la normativa o fa trampa i en posa menys? Es pot suposar que els continguts de les llaunes es distribueixen seguint una distribució normal.

   (a) Quina seria en aquest cas la hipòtesi nul·la? I l'alternativa?

   (b) Es pot afirmar amb un nivell de significació del 5% que el fabricant fa trampa? Quina seria la regió d'acceptació i la regió de rebuig en aquest cas?

   (c) Calcula un IC del 95% per al veritable contingut mitjà de les llaunes. Inclou el valor de H_0? Què vol dir això?

2. En un procés industrial una màquina necessita de mitjana 8 minuts per produir una peça d'acer. La distribució dels temps es modela amb la distribució normal. Per disminuir el temps de producció es van incorporar unes millores en el disseny de la màquina. Posteriorment, es va fer una prova estadística per comprovar si efectivament en mitjana el temps era inferior a 8 minuts. Els valors obtinguts en una mostra de peces han estat:

   $$7.3 \quad 8.5 \quad 7.4 \quad 6.4 \quad 6.2 \quad 8.2 \quad 7.3$$

   Podem acceptar que les millores han reduït el temps mitjà de producció d'una peça?

   (a) Quina seria en aquest cas la hipòtesi nul·la? I l'alternativa?

   (b) Pots afirmar, amb un nivell de significació $\alpha = 0.05$ que la mitjana de temps de produir una peça és inferior a 8 minuts?

   (c) Calcula l'error de tipus II, per a la hipòtesi alternativa que la mitjana és $\mu = 7$ i suposant que $\sigma = 2$.

3. El Servei de Parcs i Jardins d'un municipi ha d'encarregar la compra d'un lot d'arbustos pels parterres de la ciutat. Està interessat en uns arbustos que, una vegada crescuts, tinguin una alçada semblant per obtenir un aspecte estètic i evitar el cost de podar-los sovint. Una empresa de jardineria a la que s'ha adreçat el Servei els ha garantit que disposa d'una espècie d'arbustos que tenen una alçada amb una desviació tipus de 2 cm. Els tècnics del Servei han demanat a l'empresa si els hi poden enviar les alçades d'un mostra d'arbustos per poder comprovar si efectivament la desviació tipus és 2 cm o bé han de descartar aquesta xifra. L'empresa els ha enviat les alçades de 10 plantes:

   $$82.6, 82.2, 85.8, 85.6, 86.5, 83.2, 88.4, 84.2, 86.5, 87.9$$

   Respon a les qüestions seguents per ajudar al Servei de Parcs i Jardins a prendre una decisió per encarregar la compra o no del lot a aquesta empresa.

   (a) Està justificat adoptar la distribució normal per descriure l'alçada dels arbustos?

   (b) Quines són les hipòtesis a contrastar?

   (c) Quina funció de discrepància (o estadístic de contrast) és adequat per resoldre aquest contrast d'hipòtesis?

   (d) Amb un nivell de significació del 5%, quina és la regió de crítica?

   (e) Quina decisió hauria de prendre el Servei de Parcs i Jardins?

4. Els editors d'un diari universitari d'un gran campus mantenen que almenys un 75% dels estudiants estan a favor de les qualificacions numèriques més que de les literals: S/A/N/E/H. Per obtenir informació al respecte, un deganat selecciona una mostra de 50 estudiants i observa que 39 diuen estar a favor de les qualificacions numèriques. Són aquestes dades concordants amb l'opinió del editors del diari? Per donar-hi resposta soluciona els apartats següents:

   (a) Quines són les hipòtesis a contrastar?

(b) Quina funció de discrepància (o estadístic de contrast) és adequat per resoldre aquest contrast d'hipòtesis?
(c) Quina és la regió crítica amb un nivell de significació del 5%?
(d) Quina decisió prendrà el deganat si vol complaure als estudiants?

5. Les dades històriques indiquen que el nivell d'acidesa mitjana (pH) de la pluja en una determinada regió industrial és 5.2. Per contrastar si recentment s'ha produït algun canvi en aquest valor de referència es va mesurar l'acidesa de 12 tempestes de pluja durant el passat any, i es van obtenir les dades següents:

$$6.1, 5.4, 4.8, 5.8, 6.6, 5.3, 6.1, 4.4, 3.9, 6.8, 6.5, 6.3$$

Són aquestes dades prou concloents per poder afirmar, amb un nivell de significació del 5%, que l'acidesa de la pluja ha canviat respecte el valor històric? Suposa normalitat.

6. Una companyia d'aigües afirma que en mitjana el consum diari per domicili en una certa zona residencial és superior a 600 litres. Per contrastar aquesta afirmació, es selecciona una mostra aleatòria de 25 cases de la zona. La mitjana mostral dels consum va ser 656 litres i la desviació típica de 32 litres. Està en concordança això amb el que afirma la companyia? Suposa normalitat i pren un nivell de significació del 5%.

7. Les bibliotecàries de la Facultat de Matemàtiques i Estadística (FME)afirmen que, històricament, la mitjana de persones que visiten la biblioteca en un dia és de 300. Amb l'inici de les classes del grau d'Estadística a la FME, hi ha més estudiants a la facultat. Les bibliotecàries sostenen que això ha fet variar l'afluència a la biblioteca. Per comprovar-ho, es recullen el número de visitants de la biblioteca durant 10 dies i s'obtenen les dades següents:

$$301, 278, 315, 350, 189, 456, 323, 269, 254, 336$$

Podem afirmar, amb un nivell de significació del 5%, que ha variat l'afluència de visitants a la biblioteca? Suposa normalitat de les dades.

(a) Quina és la hipòtesi a contrastar?
(b) Quina és la regió de rebuig en aquest cas? I la d'acceptació?
(c) És cert el que sostenen les bibliotecàries?

8. La web dels Transports Metropolitans de Barcelona afirma que la durada mitjana del trajecte de metro entre Sants-Estació i Palau Reial és de 12 minuts. Hi ha un cert grup d'estudiants que fa el trajecte cada dia que no hi està d'acord i sosté que la durada mitjana és superior a 12 minuts. Per això, durant dues setmanes han apuntat la durada d'aquest trajecte obtenint els valors:

$$12.5, 12.7, 13, 12.8, 11.7, 11.8, 12, 11.2, 12.8, 13, 12, 14, 11.7, 12.1$$

Tenen raó els estudiants? Suposa normalitat de les dades i un nivell de significació del 5%.

9. Quina és la grandària mínima mostral que es necessita per contrastar $H_0 : \mu = 7$ vs $H_1 : \mu < 7$ si la variància de X és 57 i volem que la probabilitat de rebutjar H_0 quan no ho havíem de fer sigui com a màxim de 0.01? A més, volem que si $\mu = 5$, la probabilitat d'acceptar H_0 quan no ho havíem de fer sigui com a màxim de 0.05.

10. La nota mitjana d'un alumne en el primer curs de carrera va ser un 6.3. Per millorar el seu rendiment, va decidir apuntar-se a una acadèmia. Les notes que va obtenir durant el segon any van ser:

$$8, 7.5, 5.2, 6, 6, 5.2, 5, 5.5$$

Podem afirmar que la mitjana de les seves notes, que suposem que segueix una distribució normal, va passar a ser superior a 6?

(a) Quina és la hipòtesi a contrastar?
(b) Quina és la regió d'acceptació en aquest cas?
(c) Amb un nivell de significació del 5%, podem afirmar que la mitjana de notes és superior a 6?

1. (a) Plantegem el contrast d'hipòtesi unilateral de la manera següent:

$$\begin{cases} H_0 : \mu = 33 \\ H_1 : \mu < 33 \end{cases}$$

   (b) Per saber si el fabricant fa trampa calculem el valor de l'estadístic de contrast i el comparem amb el valor teòric sota la hipòtesi nul·la. Com que suposem normalitat de les dades i el valor de la variància és desconegut utilitzem el següent estadístic.

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

   que es distribueix com una t-Student amb $n-1$ graus de llibertat.

   La regió crítica quedarà determinada per

$$t_{ob} < -t_{n-1,\alpha}$$

   Per una banda tenim que el valor observat de l'estadístic de contrast és

$$t_{ob} = \frac{31.95 - 33}{2.25/\sqrt{25}} = -\frac{7}{3} \simeq -2.333$$

   Per l'altra banda, mirant la taula de la t-Student, veiem que el valor crític de l'estadístic és

$$t_{24,0.05} = 1.711$$

   Llavors la regió crítica és $(-\infty, -1.711]$ i la regió d'acceptació és $[-1.711, \infty)$.
   Com que el valor observat cau a la regió crítica, podem dir que la mitjana del contingut de les llaunes és significativament menor que 33cl i, per tant, que el fabricant fa trampa.

   (c) Com que la variància és desconeguda, l'interval de confiança per la mitjana queda determinat per

$$\left(\bar{x} - t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{x} + t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$

   Substituint cada terme pel seu valor corresponent tenim

$$\left(31.95 - 2.064\frac{2.25}{\sqrt{25}}, 31.95 - 2.064\frac{2.25}{\sqrt{25}}\right) = (31.021, 32.878)$$

   Com que $\mu_0 = 33 \notin [31.0212, 32.8788]$, sabem que amb una probabilitat d'un 95% la mitjana (real) del contingut de les llaunes produides pel fabricant no es troba a l'interval calculat.

2. (a) En aquest cas, el contrast d'hipòtesi és:

$$\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu < 8 \end{cases}$$

   (b) Com que les nostres dades són normals i la variància poblacional és desconeguda, utilitzem l'estadístic de contrast

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

   que es distribueix com una t-Student amb $n-1$ graus de llibertat.

Amb les dades del problema, tenim que $\bar{X} = 7.328$, $\mu_0 = 8$, $S = 0.844$ i $n = 7$.
Calculem el valor de l'estadístic de contrast observat

$$t_{obs} = \frac{7.328 - 8}{0.844/\sqrt{7}} = -2.106.$$

Com que estem contrastant la hipòtesi $\mu = \mu_0$ vs $\mu < \mu_0$, la regió de rebuig és

$$R = \left\{ \bar{X} : \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -t_{n-1,\alpha} \right\}.$$

Amb les taules de la t-Student calculem el valor crític de l'estadístic $-t_{n-1}(0.05)$

$$-t_{6,0.05} = -1.943$$

i obtenim que la regió de rebuig és $R = (-\infty, -1.943)$ i la regió d'acceptació $A = [-1.943, \infty)$. Com que el valor observat (-2.1065) pertany a la regió crítica, rebutgem la hipòtesi nul·la i podem afirmar que la mitjana del temps de produir una peça després de les millores introduides és inferior a 8 minuts.

(c) L'error de tipus II és la probabilitat d'acceptar H_0 quan H_0 és falsa. En el nostre cas, com que $\mu = 7$, volem calcular $P\,(\text{acceptar } H_0 | \mu = 7)$. Com que es tracta d'un contrast unilateral, prenem com a estadístic de discrepància

$$d = \frac{\bar{x} - 8}{2/\sqrt{7}}.$$

Amb un nivell de significació del 0.05 i tenint en compte que estem contrastant $\mu = \mu_0$ vs $\mu < \mu_0$ amb σ coneguda, la regio de rebuig és

$$R = \left\{ \bar{x} : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{0.05} \right\} = \left\{ \bar{x} : \bar{x} < \mu_0 - z_{0.05} \frac{\sigma}{\sqrt{n}} \right\} = \left\{ \bar{x} : \bar{x} < 8 - 1.64 \frac{2}{\sqrt{7}} \right\} = \{ \bar{x} : \bar{x} < 6.76 \}.$$

Per tant, la probabilitat de cometre un error de tipus II amb aquest contrast quan $\mu = 7$ és

$$P(\text{error tipus II}) = P\,(\text{acceptar } H_0 \mid \mu = 7) = P\,(d > -z_{0.05} \mid \mu = 7) =$$
$$= P\,(\bar{x} > 6.76 \mid \mu = 7) = P\left(\frac{\bar{x} - 7}{2\sqrt{7}} > \frac{6.76 - 7}{2\sqrt{7}} \right) = P\,(z > -0.371) = 0.625.$$

3. (a) Sí, ja que les variables biomètriques com l'alçada d'una certa espècie d'arbust (pes dels homes d'una certa població és un altre exemple) acostumen a distribuir-se de forma normal.

(b) El test d'hipòtesi queda plantejat de la següent manera

$$\begin{cases} H_0 : \sigma^2 = 4 \\ H_1 : \sigma^2 \neq 4 \end{cases}$$

que és un test bilateral esquerra.

(c) Com que la distribució es suposa normal, la funció de discrepància que utilitzarem és

$$U_0 = \frac{(n-1)S^2}{\sigma_0^2} \sim \mathcal{X}^2_{n-1}$$

(d) El test d'hipòtesi plantejat és bilateral, per tant la regió crítica quedarà definida per $\left(\mathcal{X}^2_{n-1,\frac{\alpha}{2}}, \mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}\right)$, és a dir si el valor observat de l'estadístic, u_{ob}, pertany a aquest interval podrem rebutjar la hipòtesi nul·la, o el que és el mateix, acceptar l'alternativa.

Calculem els valors observats i els comparem amb els valors teòrics (trobats a la taula de la khi-quadrat).

$$u_{ob} = \frac{9 \cdot 4.679}{4} = 10.527$$

$$\mathcal{X}^2_{9,0.025} = 2.70 \quad \text{i} \quad \mathcal{X}^2_{9,0.975} = 19.02$$

Com $10.527 \in (2.70, 19.02)$ no tenim evidències suficients per poder rebutjar la hipòtesi nul·la

(e) El Servei de Parcs i Jardins hauria de comprar el lot d'arbustos.

4. Les dades del problema són: $p_0 = 0.75$, $n = 50$ i $\hat{p} = \frac{39}{50} = 0.78$

(a) El test d'hipòtesi que ens hem de plantejar és el següent

$$\begin{cases} H_0 : p = 0.75 \\ H_1 : p > 0.75 \end{cases}$$

que és un test unilateral dreta.

(b) L'estadístic de contrast per a poder contrastar el test plantejat és

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$$

Per a que aquesta aproximació sigui vàlida és necessari que la mida de la mostra sigui suficientment gran. En general el que s'acostuma a comprobar és el següent:

$$np_0 \geq 5 \ \text{i} \ n(1-p_0) \geq 5$$

que al nostre cas els valors que prenen són

$$np_0 = 37.5 \geq 5 \quad \text{i} \quad n(1-p_0) = 12.5 \geq 5$$

(c) Com el test és unilateral la regió crítica queda determinada per $z_{ob} \geq Z_\alpha$, és a dir que l'interval d'acceptació és $(-\infty, Z_\alpha)$ o el que és el mateix, si $z_{ob} \in (-\infty, Z_\alpha)$ aleshores no podrem rebutjar la hipòtesi nul·la.

Calculant el valor observat de léstadístic, z_{ob} amb les dades obtingudes tenim

$$z_{ob} = \frac{0.78 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{50}}} = 0.49$$

Mirant a les taules de la normal estàndard obtenim els valors crítics necessaris

$$Z_{0.05} = 1.645$$

Per tant, com $0.49 \in (-\infty, 1.645)$ no podem rebutjar la hipòtesi nul·la.

(d) El deganat no hauria de canviar qualificacions literals per les numèriques.

5.

6.

7. (a) El contrast d'hipòtesi en aquest cas és:

$$\begin{cases} H_0 : \mu = 300 \\ H_1 : \mu \neq 300 \end{cases}$$

(b) La regió crítica per aquest contrast queda determinada per

$$|t_{ob}| > t_{n-1,\alpha/2} \iff \left[t_{n-1,\alpha/2}, +\infty\right)$$

Amb la taula de la t-Student, obtenim que el valor crític de l'estadístic és

$$t_{9,0.025} = 2.262$$

Llavors la regió crítica és $[2.262, +\infty)$ i la regió d'acceptació és $(-\infty, 2.262)$.

(c) En aquest cas, com que les nostres dades tenen una distribució normal i no coneixem la variància, l'estadístic de contrast és

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

que es distribueix com una t-Student amb $n-1$ graus de llibertat.

En el cas de la nostra mostra, tenim que $\mu_0 = 300$, $n = 10$, $S = 70.131$ i $\bar{X} = 307.1$ Per tant, l'estadístic val

$$t_{ob} = \frac{307.1 - 300}{70.131/\sqrt{10}} = 0.32$$

Com que el valor observat (0.32) cau a la regió d'acceptació trobada en l'apartat anterior ($0.32 < 2.26$), podem dir que la mitjana d'afluència a la biblioteca és la mateixa independentment dels alumnes d'estadística. Per tant, les bibliotecàries estaven equivocades.

8.

9.

10. (a) El contrast d'hipòtesi en aquest cas és

$$\begin{cases} H_0 : \mu = 6 \\ H_1 : \mu > 6 \end{cases}$$

(b) Com que estem contrastant la hipòtesi $\mu = \mu_0$ vs $\mu > \mu_0$, la regió d'acceptació és

$$A = \left\{ \bar{X} : \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq t_{n-1}(\alpha) \right\}.$$

(c) Calculem el valor de l'estadístic de contrast i el comparem amb el valor tèoric sota la hipòtesi nul·la. Com que suposem normalitat de les dades i el valor de la variància és desconegut, utilitzem l'estadístic

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

que es distribueix com una t-Student amb $n-1$ graus de llibertat.

Amb les dades del problema, tenim que $\bar{X} = 6.05$, $\mu_0 = 6$, $S = 1.119$ i $n = 8$. Per tant, el valor del nostre estadístic de contrast és

$$t_{ob} = \frac{6.05 - 6}{1.119/\sqrt{8}} = 0.126.$$

Amb les taules de la t-Student calculem el valor crític de l'estadístic $t_{n-1,0.05}$

$$t_{7,0.05} = 1.895$$

i obtenim que la regió d'acceptació és $A = (-\infty, 1.895]$.

Com que el valor observat (0.12) pertany a la regió d'acceptació, no podem rebutjar la hipòtesi nul·la i concloem que el fet d'apuntar-se a una acadèmia no ha millorat la mitjana de l'alumne.