Probabilitat i Estadística 2

Temes Previs

GCED

Curs 2019-2020

Index

Funció de densitat d'una mostra aleatòria I

Donada Y una variable aleatòria, que depèn del paràmetre θ, que pot ser un vector $(\theta_1, \ldots, \theta_K)$, la seva funció de densitat (o de probabilitat si és discreta) és:

$$f_Y\left(y; \theta\right) \qquad \textit{y és la variable de la funció}$$

Per una mostra aleatòria (repliques independents) de grandària N, $\{Y_1, Y_2, \ldots, Y_N\}$ la funció de densitat de la mostra és

$$f_{\{Y_1, Y_2, \ldots, Y_N\}}\left(y_1, y_2, \ldots, y_N; \theta\right) = \prod_{i=1}^{N} f_{Y_i}\left(y_i; \theta\right)$$

Funció de densitat d'una mostra aleatòria II

Exemple d'una mostra d'una *Poisson* (λ)

Per a cada Y_i la seva funció (en aquest cas de probabilitat) és $f_{Y_i}(y_i;\lambda) = e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}$ per tant la funció (de probabilitat) de la mostra es

$$f_Y(y;\lambda) = \prod_{i=1}^{N} e^{-\lambda}\frac{\lambda^{y_i}}{y_i!} = e^{-N\lambda}\frac{\lambda^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{n} y_i!}$$

Exemple: En un punt d'una carretera amb poc tràfic comptem el nombre de vehicles que han passat durant 10 minuts, si segueix una distribució Poisson d'esperança 7 tindrem $f_{Y_i}(y_i) = e^{-7}\frac{7^{y_i}}{y_i!}$.
Si fem el recompte 5 vegades, per la mostra $\{Y_1,\ldots,Y_5\}$ tindrem $f_Y(y) = e^{-5\cdot 7}\frac{7^{\sum_{i=1}^{5} y_i}}{\prod_{i=1}^{5} y_i!}$

Funció de densitat d'un model I

Si les variables $Y = \{Y_1, Y_2, \ldots, Y_N\}$ a més del paràmetre θ depenen dels valors de les variables explicatives X,quan $\{X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\}$ (les X_i poden ser múltiples). La funció de densitat de cada variable és

$$f_{Y_i|X_i}(y_i; x_i, \theta) \qquad y_i \text{ és la variable de la funció}$$

La funció de densitat del model és:

$$f_{Y|X}(y; x, \theta) = \prod_{i=1}^{N} f_{Y_i|X_i}(y_i; x_i, \theta)$$

Funció de densitat d'un model II

Exemple: Recta de regressió

$E[Y] = \beta_0 + \beta_1 X$ donats els valors $\{X_1 = x_1, \ldots, X_N = x_N\}$ i sabent que la distribució de les Y_i és $N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$
Aleshores la funció de densitat de cada Y_i és

$$f_{Y_i|X_i}\left(y_i; x_i, \beta, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right)^2}$$

per tant la funció de densitat del model és

$$f_{Y|X}\left(y; x, \beta, \sigma\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right)^2}$$

com a funció depèn de les variables $y = (y_1, \ldots, y_N)$

Funció de densitat d'un model III

Continuació exemple de la recta de regressió

Exemple: Si sabem que en els estudiants (sexe masculí, edat entre 18 i 26 anys) la funció de densitat del seu pes és

$$Pes \sim N\left(-33 + 0.62 \cdot Alçada, 8^2\right)$$

Per un estudiant d'alçada 170 tindrem

$$f_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi 8^2}} e^{-\frac{1}{2}\left(\frac{y_i - (-33+0.62\cdot 170)}{8}\right)^2} = 0.0499 e^{-\frac{1}{128}(y_i - 72.4)^2}$$

i per 5 estudiants d'alçades $x = \{170, 188, 178, 192, 184\}$ tindrem

$$f_Y(y) = (0.0499)^5\, e^{-\frac{1}{128}\sum_{i=1}^{5}(y_i - (-33+0.62x_i))^2}$$

Funció de versemblança d'una mostra

Canviem el paper de les Y_i i dels paràmetres θ. Suposem:

- Tenim els valors experimentals de les $Y_i = y_i$
- Desconeixem el valor dels paràmetres θ

Aleshores per una mostra tindrem que $f_Y(y;\theta) = \prod_{i=1}^{N} f_{Y_i}(y_i;\theta)$ és una funció de θ, ja que les y_i són conegudes i l'anomenarem funció de versemblança. Al seu logaritme serà molt útil i l'anomenarem funció de log-versemblança

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} f_{Y_i}(y_i;\theta)$$
$$\ell(\theta; y) = \log \mathcal{L}(\theta; y) = \sum_{i=1}^{N} \log f_{Y_i}(y_i;\theta)$$

Funció de versemblança d'un model

També canviem el paper de les Y_i i dels paràmetres θ. Suposem:

- Tenim els valors experimentals de les $Y_i = y_i$ i de les $X_i = x_i$
- Desconeixem el valor dels paràmetres θ

Aleshores per un model tindrem que $f_{Y|X}(y; x, \theta) = \prod_{i=1}^{N} f_{Y_i|X_i}(y_i; x_i, \theta)$ és una funció de θ, ja que les y_i i les x_i són conegudes i també l'anomenarem funció de versemblança, i al seu logaritme, de log-versemblança

$$\mathcal{L}(\theta; y, x) = \prod_{i=1}^{N} f_{Y_i|X_i}(y_i; x_i, \theta)$$
$$\ell(\theta; y, x) = \log \mathcal{L}(\theta; y, x) = \sum_{i=1}^{N} \log f_{Y_i|X_i}(y_i; x_i, \theta)$$

Exemples de funcions de versemblança i de log versemblança I

Mostra d'una *Poisson* (L)

- $\mathcal{L}(\lambda; y) = e^{-N\lambda} \frac{\lambda^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{n} y_i!}$
- $\ell(\lambda; y) = -N\lambda + \log(\lambda) \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \log(y_i!)$

Si de l'exemple del nombre de vehicles no coneixem el valor de λ, però hem obtingut 5 recomptes $y = \{11, 6, 8, 5, 8\}$ aleshores com que $\sum_{i=1}^{5} y_i = 38$ i $\frac{1}{\prod_{i=1}^{n} y_i!} = 1.78 \cdot 10^{-22}$

- $\mathcal{L}(\lambda) = e^{-5\lambda} 1.78 \cdot 10^{-22} \lambda^{38}$
- $\ell(\lambda) = -5\lambda + 38 \log(\lambda) - 50.078$

Exemples de funcions de versemblança i de log versemblança II

Mostra d'una exponencial $f_Y(y_i;\alpha) = \alpha e^{-\alpha y_i}$

- $\mathcal{L}(\alpha; y) = \alpha^N e^{-\alpha \sum_{i=1}^N y_i}$
- $\ell(\alpha; y) = N \log(\alpha) - \alpha \sum_{i=1}^N y_i$

Si el temps de vida d'una bombeta té distribució exponencial de α desconeguda però hem mesurat els dies que han durat 10 bombetes i hem obtingut

$y = \{70, 822, 33, 344, 265, 1374, 481, 2494, 901, 909\}$ com que $\sum_{i=1}^{10} y_i = 7693$

- $\mathcal{L}(\alpha) = \alpha^{10} e^{-\alpha 7693}$
- $\ell(\alpha) = 10 \log(\alpha) - 7693\alpha$

Exemples de funcions de versemblança i de log versemblança III

Mostra d'una binomial $f_Y(y_i; p, m) = \binom{m}{y_i} p^{y_i} (1-p)^{m-y_i}$

- $\mathcal{L}(p; y, m) = \prod_{i=1}^{N} \binom{m}{y_i} \cdot p^{\sum_{i=1}^{N} y_i} (1-p)^{N \cdot m - \sum_{i=1}^{N} y_i}$
- $\ell(p; y, m) = \sum_{i=1}^{N} \log \binom{m}{y_i} + \log p \sum_{i=1}^{N} y_i + \log(1-p) \left(N \cdot m - \sum_{i=1}^{N} y_i\right)$

No sabem quina és la probabilitat (p) de germinació d'unes llavors, però n'hem comptat el nombre de germinades de 3 safates de 144 alvèols/safata. Hem obtingut $y = (106, 112, 124)$, per tant

- $\mathcal{L}(p) = \binom{144}{106}\binom{144}{112}\binom{144}{124} \cdot p^{342} (1-p)^{90}$
- $\ell(p) = 209.94 + \log(p)\, 342 + \log(1-p)\, 90$

Exemples de funcions de versemblança i de log versemblança IV

Mostra d'una normal $f_Y(y_i;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$

- $\mathcal{L}(\mu,\sigma;y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2}\sum_{i=1}^N\left(\frac{y_i-\mu}{\sigma}\right)^2}$
- $\ell(\mu,\sigma;y) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^N (y_i-\mu)^2$

Si pensem que les alçades de les estudiants (entre 18 i 26 anys) tenen distribució normal, de μ i σ desconegudes, però n'hem mesurat 12 i hem obtingut $y = \{160, 168, 173, 164, 168, 162, 170, 155, 171, 156, 169, 175\}$ per tant $\sum_{i=1}^{12} y_i = 1991$ i $\sum_{i=1}^{12} y_i^2 = 330805$ aleshores

- $\mathcal{L}(\mu,\sigma) = \sigma^{-12}(2\pi)^{-6}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{12}(y_i-\mu)^2}$
- $\ell(\mu,\sigma) = -4.798 - 6\log\sigma - \frac{330805-2\cdot1991\mu+12\mu^2}{2\sigma^2}$

Exemples de funcions de versemblança i de log versemblança V

Model recta de regressió

- $\mathcal{L}\left(\beta,\sigma;y,x\right)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2}\sum_{i=1}^N\left(\frac{y_i-(\beta_0+\beta_1 x_i)}{\sigma}\right)^2}$
- $\ell\left(\beta,\sigma;y,x\right)=-\frac{N}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^N\left(y_i-\left(\beta_0+\beta_1 x_i\right)\right)^2$

Si en l'exemple del model de regressió els 5 estudiants d'alçades $x=\{170,188,178,192,184\}$ tenen pesos $y=\{74,90,75,86,80\}$ però les β són desconegudes, fent els càlculs $\sum_{i=1}^5 x_i=912$, $\sum_{i=1}^5 x_i^2=166648$, $\sum_{i=1}^5 x_i y_i=74082$, $\sum_{i=1}^5 y_i=405$ i $\sum_{i=1}^5 y_i^2=32997$:

- $\mathcal{L}\left(\beta,\sigma\right)=0.0101\sigma^{-5}e^{-\frac{32997-2(\beta_0 405+\beta_1 74082)+5\beta_0^2+2\beta_0\beta_1 912+\beta_1^2 166648}{2\sigma^2}}$
- $\ell\left(\beta,\sigma\right)=-4.595-5\log\sigma-$ $-\frac{1}{2\sigma^2}\left(32997-2\left(\beta_0 405+\beta_1 74082\right)+5\beta_0^2+2\beta_0\beta_1 912+\beta_1^2 166648\right)$

Estimació per màxima versemblança I

Significat de la funció de densitat o de probabilitat, coneixent θ

Tant si és d'una variable com d'una mostra aleatòria, o com d'un model, la funció $f_Y(y;\theta)$

- és sempre positiva
- $\int f_Y(y;\theta)\,\mathrm{d}y = 1$ pel que en les discretes $f_Y(y;\theta) \leq 1$ però en les contínues pot ser $f_Y(y;\theta) > 1$

El que ens indica és que com més gran és la funció, més gran és la probabilitat d'obtenir $\boldsymbol{y}$

Estimació per màxima versemblança II

Significat de la funció de versemblança quan no coneixem θ però hem obtingut $\boldsymbol{y}$

El que ens dona $\mathcal{L}(\theta; \boldsymbol{y})$ és per cada possible valor de θ, quina seria la probabilitat d'obtenir la mostra que hem obtingut.

Per tant, si per exemple $\mathcal{L}(\theta_1; \boldsymbol{y}) > \mathcal{L}(\theta_2; \boldsymbol{y})$ tindríem que $\boldsymbol{y}$ és més fàcil (probable) d'obtenir si $\theta = \theta_1$ que si $\theta = \theta_2$, és a dir θ és desconegut però seria més creïble o versemblant que $\theta = \theta_1$ que no pas $\theta = \theta_2$.

Estimació per màxima versemblança III

Mètode de màxima versemblança

Consistirà en estimar el paràmetre θ amb el valor $\hat{\theta}$ que sigui més versemblant, és a dir el $\hat{\theta} = \max_\theta \mathcal{L}(\theta; \boldsymbol{y})$.

En general aquest màxim existeix i és únic, però hi ha situacions (estranyes) en que no existeix i/o no és únic.

Estimació per màxima versemblança IV

Càlcul de l'estimador màxim versemblant

Com que és trobar $\max_\theta \mathcal{L}(\theta; \boldsymbol{y})$ i $\mathcal{L}(\theta; \boldsymbol{y})$ normalment és diferenciable, consistirà en resoldre les equacions

$$\frac{\partial}{\partial \theta_i}\mathcal{L}(\theta; \boldsymbol{y}) = 0 \quad o\ bé \quad \frac{\partial}{\partial \theta_i}\ell(\theta; \boldsymbol{y}) = 0$$

ja que el logaritme és creixent, per tant el màxim s'aconsegueix al mateix valor, i a més acostuma a tenir expressions més simples ja que en lloc de productes tindrem sumes.

És convenient comprovar que és màxim mirant que $\frac{\partial^2}{\partial \theta}\mathcal{L}(\theta; \boldsymbol{y})$ o que $\frac{\partial^2}{\partial \theta}\ell(\theta; \boldsymbol{y})$ és negativa si θ és univariant, o que és una matriu definida negativa si $\theta = (\theta_1, \ldots, \theta_K)$

Estimació per màxima versemblança V

Propietats importants dels estimadors màxim versemblants

- Invariància: Si podem reparametritzar la variable aleatòria amb el canvi $\theta = h(\alpha)$, podem obtenir $\hat{\theta}$ i també $\hat{\alpha}$, però compleixen que $\hat{\theta} = h(\hat{\alpha})$.
- $\hat{\theta}$ és asimptòticament no esbiaixat, $E\left[\hat{\theta}\right] \xrightarrow[n\to\infty]{} \theta$. Però en molts casos és no esbiaixat $E\left[\hat{\theta}\right] = \theta$
- La distribució asimptòtica de $\hat{\theta}$ és $N\left(\theta, \sigma^2_{\hat{\theta}}\right)$ amb $-\left(\frac{\partial^2}{\partial\theta}\ell\left(\hat{\theta}; \mathbf{y}\right)\right)^{-1} \xrightarrow[n\to\infty]{} \sigma^2_{\hat{\theta}}$

Exercicis de màxima versemblança I

Exercici 1 (exemple de la Poisson)

En un punt d'una carretera amb poc tràfic comptem el nombre de vehicles que passen durant 10 minuts, considerem que el recompte segueix una distribució $Poisson\,(\lambda)$ d'esperança λ desconeguda. S'han fet 5 recomptes i s'ha obtingut la mostra $\boldsymbol{y} = \{11, 6, 8, 5, 8\}$.

(a) Trobeu l'estimador màxim versemblant de λ i comproveu que és únic.

(a) Doneu la distribució asimptòtica de $\hat{\lambda}$.

1 - Solució (a)

Ho fem amb $\ell(\lambda)$ ja que és més còmode que amb la funció de versemblança.

- $\ell(\lambda; y) = -N\lambda + \log(\lambda)\sum_{i=1}^{N} y_i - \sum_{i=1}^{N}\log(y_i!)$

Si no coneixem el valor de $\lambda \geq 0$, però hem obtingut 5 recomptes $y = \{11, 6, 8, 5, 8\}$ aleshores com que $\sum_{i=1}^{5} y_i = 38$ i $-\sum_{i=1}^{N}\log(y_i!) = \log\left(\frac{1}{\prod_{i=1}^{n} y_i!}\right) = \log\left(1.78 \cdot 10^{-22}\right) = -50.078$

- $\ell(\lambda) = -5\lambda + 38\log(\lambda) - 50.078$
- $\frac{\partial}{\partial\lambda}\ell(\lambda) = -5 + \frac{38}{\lambda} = 0 \Rightarrow \hat{\lambda} = \frac{38}{5} = 7.6 = \overline{\boldsymbol{y}}$

A més $\frac{\partial^2}{\partial\lambda^2}\ell(\lambda) = -\frac{38}{\lambda^2} < 0\ \forall\lambda$, amb el que $\hat{\lambda} = 7.6$ és un màxim i a més és únic.

1-Solució (b)

$-\left(\frac{\partial^2}{\partial\lambda^2}\ell\left(\hat{\lambda}\right)\right)^{-1} = \frac{\hat{\lambda}^2}{38} = \frac{7.6^2}{38} = 1.52$. Per tant la distribució asimptòtica de $\hat{\lambda}$ és $N(\lambda, 1.52) = N\left(\lambda, 1.233^2\right)$

Exercici 2 Exemple bombetes

Suposem que T, el temps de vida d'una bombeta, té distribució exponencial $f_T(t) = \alpha e^{-\alpha t}$ amb α desconeguda. Hem mesurat el temps que han durat 10 bombetes, els resultats obtinguts són $\boldsymbol{t} = \{70, 822, 33, 344, 265, 1374, 481, 2494, 901, 909\}$.

(a) Trobeu l'estimador màxim versemblant de α i comproveu que és únic.

(b) Doneu la distribució asimptòtica de $\hat{\alpha}$.

2- Solució (a)

- $\ell(\alpha;\boldsymbol{t}) = N\log(\alpha) - \alpha\sum_{i=1}^{N} t_i$

$\boldsymbol{t} = \{70, 822, 33, 344, 265, 1374, 481, 2494, 901, 909\}$ com que $\sum_{i=1}^{10} t_i = 7693$

- $\ell(\alpha) = 10\log(\alpha) - 7693\alpha$
- $\frac{\partial}{\partial\alpha}\ell(\alpha) = \frac{10}{\alpha} - 7693 = 0 \Rightarrow \hat{\alpha} = \frac{7693}{10} = 769.3 = \overline{\boldsymbol{t}}$

A més $\frac{\partial^2}{\partial\alpha^2}\ell(\alpha) = -\frac{10}{\alpha^2} < 0\ \forall\alpha$, amb el que $\hat{\alpha} = 769.3$ és un màxim i a més és únic.

2-Solució (b)

$-\left(\frac{\partial^2}{\partial\alpha^2}\ell(\hat{\alpha})\right)^{-1} = \frac{\hat{\alpha}^2}{10} = \frac{769.3^2}{10} = 59182.25$. Per tant la distribució asimptòtica de $\hat{\alpha}$ és $N(\alpha, 1.52) = N\left(\lambda, 243.27^2\right)$

Exercici 3 Exemple binomial

No sabem quina és la probabilitat (p) de germinació d'unes llavors, però n'hem comptat el nombre de germinades de 3 safates de 144 alvèols/safata i hem obtingut $\boldsymbol{y} = (106, 112, 124)$

(a) Trobeu l'estimador màxim versemblant de p i comproveu que és únic.

(a) Doneu la distribució asimptòtica de $\hat{p}$.

3-Solució (a)

- $\ell(p; y, m) = \sum_{i=1}^{N} \log \binom{m}{y_i} + \log p \sum_{i=1}^{N} y_i + \log(1-p)\left(N \cdot m - \sum_{i=1}^{N} y_i\right)$

Hem obtingut $\boldsymbol{y} = (106, 112, 124)$, per tant $\sum_{i=1}^{3} y_i = 342$, $3 \cdot 144 - 342 = 90$ i $\log\left(\binom{144}{106}\binom{144}{112}\binom{144}{124}\right) = 209.94$

- $\ell(p) = 209.94 + \log(p)\,342 + \log(1-p)\,90$
- $\frac{\partial}{\partial p}\ell(p) = \frac{342}{p} - \frac{90}{1-p} = 0 \Rightarrow (1-p)\,342 - p90 = 0 \Rightarrow \hat{p} = \frac{342}{342+90} = 0.792 = \frac{\overline{y}}{144}$

A més $\frac{\partial^2}{\partial p^2}\ell(p) = -\frac{342}{p^2} - \frac{90}{(1-p)^2} < 0\ \forall p \in (0,1)$, amb el que $\hat{p} = 0.792$ és un màxim i a més és únic.

3-Solució (b)

$-\left(\frac{\partial^2}{\partial \alpha^2}\ell(\hat{p})\right)^{-1} = \frac{1}{\frac{342}{\hat{p}^2} + \frac{90}{(1-\hat{p})^2}} = 0.0004259$. Per tant la distribució asimptòtica de $\hat{p}$ és $N(p, 0.0003818) = N(p, 0.0195)$

Exercici 4. Exemple Normal

Sigui H l'alçada de les estudiants (entre 18 i 26 anys), considerem $H \sim N\left(\mu, \sigma^2\right)$ però μ i σ són desconegudes. N'hem mesurat 12 i hem obtingut $\boldsymbol{h} = \{160, 168, 173, 164, 168, 162, 170, 155, 171, 156, 169, 175\}$.

(a) Trobeu l'estimador màxim versemblant de μ i σ. Comproveu que $(\hat{\mu}, \hat{\sigma})$ és únic.

(a) Doneu la distribució asimptòtica de $(\hat{\mu}, \hat{\sigma})$.

4-Solució (a)

- $\ell(\mu,\sigma;\boldsymbol{h}) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(h_i-\mu\right)^2$

De les alçades que hem obtingut podem calcular $\sum_{i=1}^{12} h_i = 1991$ i $\sum_{i=1}^{12} h_i^2 = 330805$ per tant

- $\ell(\mu,\sigma) = -4.798 - 6\log\sigma - \frac{330805-2\cdot1991\mu+12\mu^2}{2\sigma^2}$
- $\frac{\partial}{\partial\mu}\ell(\mu,\sigma) = -\frac{-1991+12\mu}{\sigma^2} = 0 \Rightarrow \hat{\mu} = \frac{1991}{12} = 165.92 = \overline{\boldsymbol{h}}$
- $\frac{\partial}{\partial\sigma}\ell(\mu,\sigma) = -\frac{6}{\sigma} + \frac{330805-2\cdot1991\mu+12\mu^2}{\sigma^3} = 0 \Rightarrow \hat{\sigma}^2 = \frac{330805-2\cdot1991\hat{\mu}+12\hat{\mu}^2}{6} = 77.486$ i $\hat{\sigma} = 8.803$

A més

- $\frac{\partial^2}{\partial\mu^2}\ell(\mu,\sigma) = -\frac{14}{\sigma^2}$, $\frac{\partial^2}{\partial\mu\partial\sigma}\ell(\mu,\sigma) = 2\frac{12\mu-1991}{\sigma^2}$ i $\frac{\partial^2}{\partial\sigma^2}\ell(\mu,\sigma) = \frac{6}{\sigma^2} - 3\frac{330805-2\cdot1991\mu+12\mu^2}{\sigma^4}$
- $\frac{\partial^2}{\partial\mu^2}\ell(\hat{\mu},\hat{\sigma}) = -0.1549$, $\frac{\partial^2}{\partial\mu\partial\sigma}\ell(\hat{\mu},\hat{\sigma}) = 0$ i $\frac{\partial^2}{\partial\sigma^2}\ell(\hat{\mu},\hat{\sigma}) = -0.1549$
- amb el que $(\hat{\mu},\hat{\sigma}) = (165.92, 8.803)$ és un màxim únic.

4-Solució (b)

$-\left(\frac{\partial^2}{\partial\theta^2}\ell\left(\hat{\mu},\hat{\sigma}\right)\right)^{-1} = \begin{pmatrix} 0.1549 & 0 \\ 0 & 0.1549 \end{pmatrix}^{-1} =$
$\begin{pmatrix} 6.457 & 0 \\ 0 & 6.457 \end{pmatrix}$. Per tant la distribució asimptòtica de $(\hat{\mu},\hat{\sigma})$ és
$N\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \begin{pmatrix} 6.457 & 0 \\ 0 & 6.457 \end{pmatrix}\right) =$
$N\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \begin{pmatrix} 2.541^2 & 0 \\ 0 & 2.541^2 \end{pmatrix}\right)$

Exercici 5. Model regressió pes alçada

Sabem que la distribució del pes (P) d'estudiants (sexe masculí, edat entre 18 i 26 anys) depèn de la seva alçada (H) i és $P \sim N\left(\beta_0 + \beta_1 \cdot H, \sigma^2\right)$. Tenim l'alçada de 5 estudiants $\boldsymbol{h} = \{170, 188, 178, 192, 184\}$ i els seus pesos han donat respectivament $\boldsymbol{p} = \{74, 90, 75, 86, 80\}$.

(a) Trobeu l'estimador màxim versemblant de β. Comproveu que $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1\right)$ és únic.

(b) Doneu la distribució asimptòtica de $\left(\hat{\beta}_0, \hat{\beta}_1\right)$, que pot dependre de σ.

(c) Trobeu l'estimador de β per mínims quadrats. Es a dir, trobar el mínim de $\sum_{i=1}^{5} \left(p_i - (\beta_0 + \beta_1 h_i)\right)^2$.

5-Solució (a)

- $\ell(\beta,\sigma;\boldsymbol{p},\boldsymbol{h}) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(p_i - (\beta_0+\beta_1 h_i)\right)^2$

Alguns càlculs de les dades obtingudes són $\sum_{i=1}^{5} h_i = 912$, $\sum_{i=1}^{5} h_i^2 = 166648$, $\sum_{i=1}^{5} h_i p_i = 74082$, $\sum_{i=1}^{5} p_i = 405$ i $\sum_{i=1}^{5} p_i^2 = 32997$:

- $\ell(\beta,\sigma) = -4.595 - 5\log\sigma - \left(\frac{32997-2(\beta_0 405+\beta_1 74082)+5\beta_0^2+2\beta_0\beta_1 912+\beta_1^2 166648}{2\sigma^2}\right)$

- $$\begin{cases} \frac{\partial}{\partial\beta_0}\ell(\beta,\sigma) = -\left(\frac{-2\cdot 405+10\beta_0+2\beta_1 912}{2\sigma^2}\right) = 0 \Rightarrow -5b_0 + 912b_1 = 405 \\ \frac{\partial}{\partial\beta_1}\ell(\beta,\sigma) = -\left(\frac{-2\cdot 74082+2\cdot 166648\beta_1+2\beta_0 912}{2\sigma^2}\right) = 0 \Rightarrow -912b_0 - 166648b_1 = 74 \end{cases}$$

  $$\begin{cases} \hat\beta_0 = -47.0214 \\ \hat\beta_1 = 0.70187 \end{cases} \quad \text{és un màxim únic.}$$

5-Solució (b)

- $\frac{\partial^2}{\partial\beta^2}\ell(\beta,\sigma) = -\frac{1}{\sigma^2}\begin{pmatrix} 5 & 912 \\ 912 & 166648 \end{pmatrix}$

$$-\left(\frac{\partial^2}{\partial\beta^2}\ell(\beta,\sigma)\right)^{-1}=\sigma^2\begin{pmatrix}5 & 912\\ 912 & 166648\end{pmatrix}^{-1}=$$

$\sigma^2\begin{pmatrix}111.396 & -0.6096\\ -0.6096 & 0.003342\end{pmatrix}$. Per tant la distribució asimptòtica de

$\hat\beta$ és $N\left(\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},\sigma^2\begin{pmatrix}111.396 & -0.6096\\ -0.6096 & 0.003342\end{pmatrix}\right)$

$\hat\beta_0\sim N\left(\beta_0,10.55^2\sigma^2\right)$ i $\hat\beta_1\sim N\left(\beta_1,0.0578^2\sigma^2\right)$

-

5-Solució (c)

$SQ=\sum_{i=1}^5\left(p_i-(\beta_0+\beta_1 h_i)\right)^2$

- $\left\{\begin{array}{l}\frac{\partial}{\partial\beta_0}SQ=2\sum_{i=1}^5\left(p_i-(\beta_0+\beta_1 h_i)\right)=0\\ \frac{\partial}{\partial\beta_1}SQ=2\sum_{i=1}^5 h_i\left(p_i-(\beta_0+\beta_1 h_i)\right)=0\end{array}\right\}\Rightarrow$

  $\left\{\begin{array}{l}-5b_0+912b_1=405\\ -912b_0-166648b_1=74082\end{array}\right\}$ Les mateixes equacions que en l'apartat (a) per tant les estimacions són les mateixes.

Resultats normals I

- Donada una v.a. Y a $\mathbb{R}$, es diu que els seus resultats "normals" són el conjunt (a, b) amb $\Pr(Y \in (a, b)) = p_0$ on p_0 és una probabilitat predeterminada alta, també anomenada nivell de confiança, per exemple 95%, 99%...
- Els conjunts $(-\infty, a]$ i $[b, \infty)$ són els resultats "estranys",
    - generalment es demana que $\Pr(Y \in (-\infty, a]) = \Pr(Y \in [b, \infty))$,
    - també es pot planejar que una de les cues tingui probabilitat 0, és a dir que els resultats normals siguin de la forma $[a, \infty)$ o bé $(-\infty, b]$.

Al conjunt dels resultats normal d'una població se l'anomena **regió, o interval, de predicció**.

Resultats normals II

Exercicis

1. Si $Y \sim N\left(\mu, \sigma^2\right)$ amb μ i σ^2 coneguts, trobeu-ne els resultats normals al 99% on els valors estranys són dues cues equiprobables.
2. Si $Y \sim N\left(\mu, \sigma^2\right)$ amb μ i σ^2 desconeguts, però tenim una m.a.s. $\{Y_1, ..., Y_{10}\}$ de la que s'ha obtingut $\overline{Y} = m$ i $S^2 = v$, trobeu-ne (aproximadament, "interval de predicció") els resultats normals al 99%, on els valors estranys són dues cues equiprobables.
3. En la mateixa situació de l'apartat (2), trobeu els resultats normals al 99% de dues cues iguals: de $\overline{Y}$ suposant μ conegut, i també de S_Y^2 però suposant σ^2 conegut.

6-Solució (1)

$Y \sim N\left(\mu, \sigma^2\right) \Rightarrow z = \frac{Y-\mu}{\sigma} \sim N(0,1)$.

Els resultats normals de z són $(-z_0, z_0)$ amb $\Pr\left(z \in (-z_0, z_0)\right) = 0.99 \Rightarrow F_z(z_0) = 0.995 \Rightarrow z_0 = F_z^{-1}(0.995) = 2.5758$

Com que $0.99 = \Pr\left(\frac{Y-\mu}{\sigma} \in (-2.5758, 2.5758)\right) = \Pr\left(Y \in \left(\frac{\mu-2.5758}{\sigma}, \frac{\mu+2.5758}{\sigma}\right)\right) \Rightarrow$

$$(a, b) = (\mu - 2.5758\sigma, \mu + 2.5758\sigma)$$

6-Solució (2)

$Y \sim N\left(\mu, \sigma^2\right)$, $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ com que una futura Y i la mostra $\{Y_1, ..., Y_{10}\}$ són ind. $\Rightarrow Y - \overline{Y} \sim N\left(0, \left(1 + \frac{1}{n}\right)\sigma^2\right) \Rightarrow$

$$\frac{Y-\overline{Y}}{\sigma\sqrt{1+\frac{1}{n}}} \sim N(0,1) \Rightarrow \frac{\frac{Y-\overline{Y}}{\sigma\sqrt{1+\frac{1}{n}}}}{\frac{S}{\sigma}} = \frac{Y-\overline{Y}}{S\sqrt{1+\frac{1}{n}}} \sim t_{n-1}.$$

Els resultats normals de la t_{10-1} són $(-t_0, t_0)$ amb

$\Pr\left(t_9 \in (-t_0, t_0)\right) = 0.99$ amb $\Pr\left(t_9 \in (-t_0, t_0)\right) = 0.99 \Rightarrow$

$F_{t_9}(t_0) = 0.995 \Rightarrow t_0 = F_{t_0}^{-1}(0.995) = 3.2498$. Com que

$0.99 = \Pr\left(\frac{Y-m}{\sqrt{1.1v}} \in (-3.2498, 3.2498)\right) =$

$\Pr\left(Y \in \left(m - 3.2498\sqrt{1.1v}, m + 3.2498\sqrt{1.1v}\right)\right) \Rightarrow$

$(a, b) = \left(m - 3.4085\sqrt{v}, m + 3.4085\sqrt{v}\right)$

6-Solució (3)

$\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{Y}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ amb $n = 10$ però si no coneixem σ^2/n i en lloc seu utilitzem v/n tindrem $\frac{\overline{Y}-\mu}{\sqrt{v/n}} \sim t_{n-1}$ els resultats normals de la t_{n-1} són

$t_{n-1,0.005} = F^{-1}_{t_{n-1}}(0.01/2) = -3.2500$ i

$t_{n-1,0.995} = F^{-1}_{t_{n-1}}(1 - 0.01/2) = 3.2500$ per tant

$(a,b) = \left(\mu - 3.25\sqrt{\frac{v}{10}}, \mu + 3.25\sqrt{\frac{v}{10}}\right)$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ amb $n = 10$, els valors normals de la χ^2_{n-1} són

$\chi^2_{n-1,0.005} = F^{-1}_{\chi^2_{n-1}}(0.01/2) = 1.7349$ i

$\chi^2_{n-1,0.995} = F^{-1}_{\chi^2_{n-1}}(1 - 0.01/2) = 23.5893$ per tant

$$(a,b) = \left(\frac{1.7349\sigma^2}{9}, \frac{23.5893\sigma^2}{9}\right)$$

Intervals de confiança de nivell de confiança $1-\alpha$ I

D'una població Y de la que desconeixem el valor d'un paràmetre θ però

- tenim dades experimentals, $\boldsymbol{y}$.
- $T(\boldsymbol{Y},\theta)$ un estadístic adequat.

Donat el nivell de confiança, $1-\alpha$, anomenarem **interval de confiança** de nivell $1-\alpha$ a $IC_{1-\alpha}(\theta)$ a:

- El conjunt de valors $\tilde{\theta}$ pels que:

  el valor obtingut de $T\left(\boldsymbol{y},\tilde{\theta}\right)=T_{calc}$ és un resultat "normal de probabilitat $1-\alpha$" de T.

Intervals de confiança de nivell de confiança $1-\alpha$ II

Exemple

- Tenim la variable aleatòria $Y \sim N\left(\mu, \sigma^2\right)$ amb μ i σ desconeguts.
- Tenim $\boldsymbol{y}$ una m.a.s de grandària $n = 12$, $\overline{\boldsymbol{y}} = 123$ i $S_n^2 = \frac{\sum_{i=1}^n y_i}{n-1} = 6.75$

1. Calcular $IC_{95\%}\left(\mu\right)$
2. Calcular $IC_{95\%}\left(\sigma\right)$ o Calcular $IC_{95\%}\left(\sigma^2\right)$

Intervals de confiança de nivell de confiança $1-\alpha$ III

$IC_{95\%}(\mu)$

- $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{Y}-\mu}{\frac{\sigma^2}{n}} \sim N(0,1) \Rightarrow T(\boldsymbol{Y},\mu) = \frac{\overline{Y}-\mu}{\sqrt{\frac{S^2}{n}}} = \frac{\overline{Y}-\mu}{\sqrt{\frac{\sigma^2}{n}}\sqrt{\frac{S^2}{\sigma^2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = t_{n-1}$
- Els resultats normals del 95% de dues cues de t_{12-1} són $(t_{11,0.025}, t_{11,0.975}) = (-2.2, 2.2)$
- El resultat obtingut $T_{calc} = \frac{123-\mu}{\sqrt{\frac{6.75}{12}}} = \frac{123-\mu}{0.75}$ serà normal si $\frac{123-\mu}{0.75} \in (-2.2, 2.2) \Rightarrow$

$$IC_{95\%}(\mu) = (123 - 2.2 \cdot 0.75, 123 + 2.2 \cdot 0.75) = (121.35, 124.65)$$

Intervals de confiança de nivell de confiança $1-\alpha$ IV

$IC_{95\%}\left(\sigma^2\right)$

- $T\left(\boldsymbol{Y},\sigma^2\right)=\frac{(n-1)S_n^2}{\sigma^2}\sim\chi^2_{n-1}$
- Els resultats normals del 95% de dues cues de χ^2_{12-1} són $\left(\chi^2_{11,0.025},\chi^2_{11,0.975}\right)=(3.816,21.92)$
- Per que el resultat obtingut sigui normal $T_{calc}=\frac{11\cdot 6.75}{\sigma^2}=\frac{74.25}{\sigma^2}\in(3.816,21.92)\Rightarrow$

$$IC_{95\%}\left(\sigma^2\right)=\left(\frac{74.25}{21.92},\frac{74.25}{3.816}\right)=(3.387,19.459)$$

$$IC_{95\%}\left(\sigma\right)=\left(\sqrt{3.387},\sqrt{19.459}\right)=(1.840,4.411)$$

Conceptes dels tests d'hipòtesis I

Definició

- Test d'hipòtesis és un procediment per contestar, utilitzant resultats experimentals, una pregunta sobre els paràmetres d'una població.
- Plantejarem dues hipòtesis disjuntes i seguint el procediment escollirem una de les dues hipòtesis.
- No tractarem les dues hipòtesis de forma igual.

Conceptes dels tests d'hipòtesis II

Hipòtesis del test

- H_0 serà la **hipòtesi nul·la**. Ens ha de permetre fer càlculs i és la que acceptaríem si no podem demostrar el contrari.
- H_1 serà la **hipòtesi alternativa**. És la que volem demostrar, només l'acceptarem si n'estem segurs, es a dir si la probabilitat d'acceptar-la erròniament és més petita que un llindar (probabilitat petita com 5%, 1%, ...) que hem pre-establert i que anomenem nivell de significació α.

Conceptes dels tests d'hipòtesis III

Procediment (introductori)

- Fixem α, el nivell de significació.
- De la pregunta planegem les dues hipòtesis.
- De les dades experimentals calculem un estadístic T, que anomenarem **estadístic de contrast**.
- Suposant que la hipòtesi nul·la és certa, amb probabilitat $1-\alpha$ calculem la regió de resultats normals de l'estadístic de contrast, que anomenarem **regió d'acceptació**, el complementari l'anomenarem **regió de rebuig**.
- Decidirem en funció de a quina regió surti T_{calc}, l'estadístic de contrast calculat.

Conceptes dels tests d'hipòtesis IV

Forma de decidir

Depenen de si l'estadístic de contrast calculat està a la regió d'acceptació o a la regió de rebuig escollirem una hipòtesi o l'altre:

- Si està a la regió d'acceptació: escollirem H_0 que és la que hem utilitzat per fer els càlculs, direm que el test és **no significatiu**, que de fet vol dir que no n'estem segurs, que no hem pogut demostrar el contrari.
- Si està a la regió rebuig: escollirem H_1, de fet rebutgem la que hem utilitzat per fer els càlculH_0, direm que el test és **significatiu**, que de fet vol dir que n'estem segurs ja que la $\Pr(error) < \alpha$.

Conceptes dels tests d'hipòtesis V

Tipus d'error

En prendre la decisió de quina hipòtesi del test escollim, ens podem trobar en una d'aquestes situacions, encara que no sabem quina és la certa:

	Escollim H_0	Escollim H_1
H_0 és correcte	Correcte $_{(\Pr=1-\alpha)}$	Error: Tipus I $_{(\Pr=\alpha)}$
H_1 és correcte	Error: Tipus II $_{(\Pr=\beta)}$	Correcte $_{(\Pr=1-\beta)}$

Conceptes dels tests d'hipòtesis VI

Probabilitats de les decisions

$\Pr\left(\textit{Error I}\right) = \Pr\left(\textit{Rebutjar } H_0 | H_0 \textit{ certa}\right)$ és la **significació del test** i per construcció $\leq \alpha$ (nivell de significació establert per nosaltres).

- $\Pr\left(\textit{Escollir } H_0 | H_0 \textit{ certa}\right) = 1 - \Pr\left(\textit{Rebutjar } H_0 | H_0 \textit{ certa}\right) \geq 1 - \alpha$, $1 - \alpha$ s'anomena nivell de confiança.

$\Pr\left(\textit{Error II}\right) = \Pr\left(\textit{Rebutjar } H_1 | H_1 \textit{ certa}\right) \equiv \beta$, voldríem que fos petita però en general serà gran.

- $\Pr\left(\textit{Escollir } H_1 | H_1 \textit{ certa}\right) = 1 - \Pr\left(\textit{Rebutjar } H_1 | H_1 \textit{ certa}\right) \equiv 1 - \beta$, s'anomena la potència del test i voldríem que fos el més gran possible.

Per augmentar la potència (disminuir Error II) necessitem més dades experimentals o escollir un estadístic de contrast més adequat.

Conceptes dels tests d'hipòtesis VII

Procediment de decisió utilitzant el p_{valor}, serà la habitual

En el procediment introductori dèiem:

- Suposant que la hipòtesi nul·la és certa, amb probabilitat $1-\alpha$ calculem la regió de resultats normals de l'estadístic de contrast, que anomenarem **regió d'acceptació**, el complementari l'anomenarem **regió de rebuig**.

Però la regió d'acceptació la podem calcular per qualsevol probabilitat, el que farem és trobar quina probabilitat fa que el T_{calc} surti just a la frontera entre la regió d'acceptació i la de rebuig. En aquest cas a la $\Pr(\textit{regió rebuig})$ l'anomenarem $\boldsymbol{p_{valor}}$ del test.

Podem pensar el p_{valor} com

la probabilitat d'error quan rebutgem H_0

Conceptes dels tests d'hipòtesis VIII

Com decidir utilitzant el p_{valor}

- Si $p_{valor} \geq \alpha$ és que $T_{calc} \in \{Resultats\ normal\ al\ 1-\alpha\} \Rightarrow$

  No significatiu ja que

  $\{Resultats\ normal\ al\ 1-p_{valor}\} \subset \{Resultats\ normal\ al\ 1-\alpha\}$
- Si $p_{valor} < \alpha$ és que $T_{calc} \notin \{Resultats\ normal\ al\ 1-\alpha\} \Rightarrow$

  Significatiu ja que

  $\{Resultats\ normal\ al\ 1-p_{valor}\} \supset \{Resultats\ normal\ al\ 1-\alpha\}$
  - A més sabem que $\Pr(Error\ I) \leq p_{valor} < \alpha$,

    altrament només sabíem que $\Pr(Error\ I) \leq \alpha$

Conceptes dels tests d'hipòtesis IX

Exemples, algunes preguntes sobre l'exercici 7)

S'ha fet un seguiment de iogurts en la fase de comercialització, per comparar l'efecte de la temperatura de la fermentació.

Individus: Unitats de iogurt, de cada temperatura de fermentació, analitzades alguns dies després:

Dues variables explicatives:

- Dies: transcorreguts entre la fermentació i l'anàlisi del iogurt.
- Grups: Dos temperatures de la fermentació 42° o 43.5°.

Tres variables resposta mesurades durant l'experiència:

- pH: del iogurt.
- Strep.: concentració de *Streptococcus salivarius thermophilus*
- Lactob.: concentració de *Lactobacillus delbrueckii bulgaricus*

Temperatura 42°, el dia de la fermentació. Ex. 7.a)

Suposant que $pH \sim N\left(\mu, \sigma^2\right)$, amb $\alpha = 0.05$ dieu si

el seu valor esperat és 4.4? o no?

- Els valors obtinguts són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ $n = 6$ i calculant obtenim $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- Hipòtesis:

  H_0 : $\mu_{pH} = 4.4$, ens permet fer càlculs ja que conté el " $=$ ".
  Si l'escollim no n'estarem segurs (No significatiu).

  H_1 : $\mu_{pH} \neq 4.4$, no ens permet fer càlculs ja que no conté el " $=$ ".
  Si l'escollim sí n'estarem segurs (Significatiu).
- Estadístic: $\frac{\overline{pH} - \mu_{H_0}}{\sqrt{\frac{\sigma^2}{n}}} \sim t_{n-1}$

Resolució:

- De les dades obtenim $n = 6$, $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- $T_{calc} = \frac{\overline{pH} - \mu_{H_0}}{\sqrt{\frac{\sigma^2}{n}}} = \frac{4.45 - 4.4}{\sqrt{\frac{0.00048}{6}}} = 5,031$
- Forma de la regió d'acceptació, dues cues iguals $(-\delta, \delta)$
- La regió d'acceptació amb el p_{valor}serà $(-5,031, 5,031)$
- Per simetria $p_{valor} = 2\Pr(t_5 > 5,031) = 0.003997 < 0.05 \Rightarrow$ ***Significatiu***

Conclusió:

$$\mu_{pH} \neq 4.4, \text{ n'estem "molt segurs"}, \Pr(Error) \leq 0.003997$$

Temperatura 42°, el dia de la fermentació, diferent pregunta. Ex. 7.b)

Suposant que $pH \sim N(\mu, \sigma)$, amb $\alpha = 0.05$ dieu si

el seu valor esperat és inferior a 4.4? o no?

- Els valors obtinguts són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ $n = 6$ i calculant obtenim $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- Hipòtesis:

  H_0 : $\mu_{pH} \geq 4.4$, ens permet fer càlculs ja que conté el " = ".
  Si l'escollim no n'estarem segurs (No significatiu).

  H_1 : $\mu_{pH} < 4.4$, no ens permet fer càlculs ja que no conté el " = ".
  Si l'escollim sí n'estarem segurs (Significatiu).
- Estadístic: $\frac{\overline{pH}-\mu_{H_0}}{\sqrt{\frac{\sigma^2}{n}}} \sim t_{n-1}$

Resolució: és d'una cua

- De les dades obtenim $n = 6$, $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- $T_{calc} = \frac{\overline{pH} - \mu_{H_0}}{\sqrt{\frac{\sigma^2}{n}}} = \frac{4.45 - 4.4}{\sqrt{\frac{0.00048}{6}}} = 5,031$
- Forma de la regió de rebuig, una cua, $(-\infty, \delta)$ ja que suposant H_0 els valors estranys només serà la cua negativa, si es compleix H_1 aleshores $\frac{(<4.4)-(4.4)}{\sqrt{\frac{\sigma^2}{n}}} < 0$.
- La regió de rebuig amb el p_{valor}serà $(-\infty, 5,031)$
- Per tant $p_{valor} = \Pr(t_5 < 5,031) = 0.998 > 0.05 \Rightarrow$ ***No significatiu***

Conclusió:

Acceptem que$\mu_{pH} \geq 4.4$, però no n'estem "segurs"

Temperatura 42°, el dia de la fermentació, pregunta sobr σ. Ex. 7.c)

Suposant que $pH \sim N(\mu, \sigma)$, amb $\alpha = 0.05$ dieu si

la seva desviació tipus és 0.02? o no?

- Els valors obtinguts són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ $n = 6$ i calculant obtenim $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- Hipòtesis:

  H_0 : $\sigma_{pH} = 0.02$, ens permet fer càlculs ja que conté el " $=$ ".
  Si l'escollim no n'estarem segurs (No significatiu).

  H_1 : $\sigma_{pH} \neq 0.02$, no ens permet fer càlculs ja que no conté el " $=$ ".
  Si l'escollim sí n'estarem segurs (Significatiu).
- Estadístic: $\frac{(n-1)S^2_{pH}}{\sigma^2_{pH}} \sim \chi^2_{n-1}$

Resolució: és d'una cua

- De les dades obtenim $n = 6$, $\overline{pH} = 4.45$ i$S^2_{pH} = 0.00048$.
- $T_{calc} = \frac{(n-1)S^2_{pH}}{\sigma^2_{pH} de\ H_0} = \frac{(6-1)0.00048}{0.02^2} = 6$
- Forma de la regió d'acceptació és (a, b), dues cues equiprobables de rebuig.
- La regió d'acceptació amb el p_{valor}será $(a, 6)$, o bé $(6, b)$, com que $\Pr\left(\chi^2_5 < 6\right) = 0.694 > 0.5$ només pot ser de la forma $(a, 6)$.
- Per simetria $p_{valor} = 2\Pr\left(\chi^2_5 > 6\right) = 2\left(1 - 0.694\right) = 0.612 > 0.05 \Rightarrow$ ***No significatiu***

Conclusió:

Acceptem que$\boldsymbol{\sigma_{pH} = 0.02}$, però no n'estem "segurs"

Tests de comparació de dues poblacions normals I

Tenim dues poblacions normals $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$ i $Z \sim N\left(\mu_Z, \sigma_Z^2\right)$ amb tots els paràmetres desconeguts.

Tenim també les m.a.s. $\boldsymbol{y} = (y_1, \ldots, y_{n_Y})$ i $\boldsymbol{z} = (z_1, \ldots, z_{n_Z})$

Algunes de les preguntes que es fan són:

1. $\mu_Y \neq \mu_Z$? o $\mu_Y > \mu_Z$?
    - suposant que $\boldsymbol{y}$ i $\boldsymbol{z}$ són independents i a més $\sigma_Y = \sigma_Z = \sigma$.
    - suposant que $\boldsymbol{y}$ i $\boldsymbol{z}$ són independents però pot ser que $\sigma_Y \neq \sigma_Z$.
    - suposant que $\boldsymbol{y}$ i $\boldsymbol{z}$ són dades aparellades.
2. $\sigma_Y \neq \sigma_Z$? o $\sigma_Y > \sigma_Z$? suposant que $\boldsymbol{y}$ i $\boldsymbol{z}$ són independents.

Tests de comparació de dues poblacions normals II

$\mu_Y \neq \mu_Z$?, suposant que y i z són independents i a més $\sigma_Y = \sigma_Z = \sigma$.

H_0 : $\mu_Y = \mu_Z \Leftrightarrow \mu_Y - \mu_Z = 0$

H_1 : $\mu_Y \neq \mu_Z \Leftrightarrow \mu_Y - \mu_Z \neq 0$

Estadístic:

- $\overline{Y} - \overline{Z} \sim N\left(\mu_Y - \mu_Z, \frac{\sigma^2}{n_Y} + \frac{\sigma^2}{n_Z}\right) =_{per\ H_0}= N\left(0, \left(\frac{1}{n_Y} + \frac{1}{n_Z}\right)\sigma^2\right)$
- $\hat{\sigma}^2 = \left\{\begin{array}{c} S_Y^2 \sim \sigma^2 \frac{\chi^2_{n_Y-1}}{n_Y-1} \\ S_Z^2 \sim \sigma^2 \frac{\chi^2_{n_Z-1}}{n_Z-1} \end{array}\right\} \Rightarrow$ *com que per independents* $\chi^2_{gl_1} + \chi^2_{gl_2} = \chi^2_{gl_1+gl_2}$ $\Rightarrow$

  $\hat{\sigma}^2 =_{variància\ conjunta}= S_C^2 = \frac{(n_Y-1)S_Y^2+(n_Y-1)S_Z^2}{(n_Y-1)+(n_Z-1)} \sim \sigma^2 \frac{\chi^2_{n_Y+n_Z-2}}{n_Y+n_Z-2}$
- $T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\left(\frac{1}{n_Y}+\frac{1}{n_Z}\right)S_C^2}} \sim t_{n_Y+n_Z-2}$

Tests de comparació de dues poblacions normals III

$\mu_Y \neq \mu_Z$?, suposant que y i z són independents però pot ser que $\sigma_Y \neq \sigma_Z$.

$$H_0: \mu_Y = \mu_Z \Leftrightarrow \mu_Y - \mu_Z = 0$$

$$H_1: \mu_Y \neq \mu_Z \Leftrightarrow \mu_Y - \mu_Z \neq 0$$

Estadístic:

- $\overline{Y} - \overline{Z} \sim N\left(\mu_Y - \mu_Z, \frac{\sigma_Y^2}{n_Y} + \frac{\sigma_Z^2}{n_Z}\right) =_{per\ H_0} = N\left(0, \frac{\sigma_Y^2}{n_Y} + \frac{\sigma_Z^2}{n_Z}\right)$
- $\left\{ \begin{array}{l} \hat{\sigma}_Y^2 = S_Y^2 \sim \sigma_Y^2 \frac{\chi^2_{n_Y - 1}}{n_Y - 1} \\ \hat{\sigma}_Z^2 = S_Z^2 \sim \sigma_Z^2 \frac{\chi^2_{n_Z - 1}}{n_Z - 1} \end{array} \right\} \Rightarrow$
- $T = \frac{\overline{Y} - \overline{Z}}{\sqrt{\left(\frac{S_Y^2}{n_Y} + \frac{S_Z^2}{n_Z}\right)}} aprox. \sim t_{gle}$ on $\underset{g.l.\ efectius}{gle} = \frac{\left(\frac{S_Y^2}{n_Y} + \frac{S_Z^2}{n_Z}\right)^2}{\frac{S_Y^4}{n_Y^2 (n_Y - 1)} + \frac{S_Z^4}{n_Z^2 (n_Z - 1)}}$

Tests de comparació de dues poblacions normals IV

$\mu_Y \neq \mu_Z$?, suposant que $\boldsymbol{y}$ i $\boldsymbol{z}$ són aparellades per tant $n_Y = n_Z = n$.

H_0 : $\mu_Y = \mu_Z \Leftrightarrow \mu_Y - \mu_Z = 0$

H_1 : $\mu_Y \neq \mu_Z \Leftrightarrow \mu_Y - \mu_Z \neq 0$

Estadístic: $\boldsymbol{y} - \boldsymbol{z} = \{y_1 - z_1, \ldots, y_n - z_n\}$ com si només tinguéssim la població $Y - Z$

- $\overline{Y - Z} = \overline{Y} - \overline{Z} \sim N\left(\mu_Y - \mu_Z, \frac{\sigma^2_{Y-Z}}{n}\right) =_{per\ H_0} = N\left(0, \frac{\sigma^2_{Y-Z}}{n}\right)$
- $\hat{\sigma}^2_{Y-Z} = S^2_{Y-Z} \sim \sigma^2_{Y-Z}\frac{\chi^2_{n-1}}{n-1}$
- $T = \frac{\overline{Y} - \overline{Z}}{\sqrt{\frac{S^2_{Y-Z}}{n}}} \sim t_{n-1}$

Tests de comparació de dues poblacions normals V

$\sigma_Y^2 \neq \sigma_Z^2$?, suposant que y i z són independents.

$H_0 : \sigma_Y^2 = \sigma_Z^2 \Leftrightarrow \frac{\sigma_Y^2}{\sigma_Z^2} = 1$

$H_1 : \sigma_Y^2 \neq \sigma_Z^2 \Leftrightarrow \frac{\sigma_Y^2}{\sigma_Z^2} \neq 1$

Estadístic:

- $\hat{\sigma}_Y^2 = S_Y^2 \sim \sigma_Y^2 \frac{\chi^2_{n_Y - 1}}{n_Y - 1}$
- $\hat{\sigma}_Z^2 = S_Z^2 \sim \sigma_Z^2 \frac{\chi^2_{n_Z - 1}}{n_Z - 1}$
- $T = \frac{s_Y^2/\sigma_Y^2}{s_Z^2/\sigma_Z^2} =_{per\ H_0} = \frac{S_Y^2}{S_Z^2} \sim F\left(n_Y - 1, n_Z - 1\right)$

Tests de comparació de dues poblacions normals VI

Exemples, algunes preguntes sobre l'exercici 8)

És el mateix exemple anterior:

S'ha fet un seguiment de iogurts en la fase de comercialització, per comparar l'efecte de la temperatura de la fermentació.

Individus: Unitats de iogurt, de cada temperatura de fermentació, analitzades alguns dies després:

Dues variables explicatives:

- *Dies: transcorreguts entre la fermentació i l'anàlisi del iogurt.*
- *Grups: Dos temperatures de la fermentació 42° o 43.5°.*

Una de les tres variables resposta mesurades durant l'experiència:

- *pH: del iogurt.*

Però ara tindrem els resultats del *pH* de les dues fermentacions.

El dia de la fermentació, les dues temperatures. Ex. 8.a) I

Suposant que $pH|T \sim N\left(\mu_T, \sigma_T^2\right)$, amb $\alpha = 0.05$ dieu si

Les dues esperances són iguals? o no?

Els valors obtinguts a $T = 42°$ són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ i per $T = 43.5°$ són $\{4.38, 4.37, 4.33, 4.34, 4.31, 4.39\}$.

H_0 : $\mu_{42^\circ} = \mu_{43.5^\circ}$, ens permet fer càlculs ja que conté el " $=$ ".

H_1 : $\mu_{42^\circ} \neq \mu_{43.5^\circ}$, no ens permet fer càlculs ja que no conté el " $=$ ".

$$\textit{suposant } \sigma_{42^\circ}=\sigma_{43.5^\circ}: \quad T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\left(\frac{1}{n_Y}+\frac{1}{n_Z}\right)S_C^2}} \sim t_{n_Y+n_Z-2}$$

$$\textit{suposant } \sigma_{42^\circ}\neq\sigma_{43.5^\circ}: \quad T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\left(\frac{S_Y^2}{n_Y}+\frac{S_Z^2}{n_Z}\right)}} \textit{aprox.} \sim t_{gle}$$

El dia de la fermentació, les dues temperatures. Ex. 8.a) II

Resolució suposant $\sigma_{42^\circ} = \sigma_{43.5^\circ}$:

- *De les dades obtenim* $n = 6$, *per les dues* T, $\overline{pH}_{42^\circ} = 4.45$, $\overline{pH}_{43.5^\circ} = 4.3533$, $S^2_{42^\circ} = 0.00048$ *i* $S^2_{43.5^\circ} = 0.000987$.
- $S^2_C = \frac{6 \cdot 0.00048 + 6 \cdot 0.000987}{6+6} = 0.000733$ *i* $T_{calc} = \frac{4.45 - 4.3533}{\sqrt{\left(\frac{1}{6}+\frac{1}{6}\right)0.000733}} = 6.1828 \sim t_{12-2}$
- *Forma de la regió d'acceptació, dues cues iguals* $(-\delta, \delta)$
- *La regió d'acceptació amb el* p_{valor} *serà* $(-6.1828, 6.1828)$
- *Per simetria* $p_{valor} = 2\,\Pr\left(t_{10} > 6.1828\right) = 0.000104 < 0.05 \Rightarrow$ *Significatiu*

Conclusió:

$\mu_{42^\circ} \neq \mu_{43.5^\circ}$, ***n'estem "molt segurs"***, $\Pr\left(Error\right) \leq 0.000104$

El dia de la fermentació, les dues temperatures. Ex. 8.a) III

Resolució suposant $\sigma_{42^\circ} \neq \sigma_{43.5^\circ}$:

- Els mateixos càlculs i forma de la regió d'acceptació del cas anterior.
- $T_{calc} = \dfrac{4.45-4.3533}{\sqrt{\left(\frac{0.00048}{6}+\frac{0.000987}{6}\right)}} = 6.1828 \sim aprox.t_{gle}$

$$gle = \frac{\left(\frac{S_Y^2}{n_Y}+\frac{S_Z^2}{n_Z}\right)^2}{\frac{S_Y^4}{n_Y^2\,(n_Y-1)}+\frac{S_Z^4}{n_Z^2\,(n_Z-1)}} = 8.934$$

Nota: Surt el mateix valor de la T_{calc} si la n és la mateixa pels 2 casos.

Sempre $gle \leq n_1 + n_2 - 2$, es compleix la igualtat quan $S_Y^2 = S_Z^2$ i va disminuint a mesura que les variàncies són més diferents.

- *Per simetria* $p_{valor} = 2\,\Pr\left(t_{8.934} > 6.1828\right) = 0.000167 < 0.05 \Rightarrow$ *Significatiu*

Conclusió:

$\mu_{42^\circ} \neq \mu_{43.5^\circ}$, ***n'estem "molt segurs"***, $\Pr\left(Error\right) \leq 0.000167$

El dia de la fermentació, les dues temperatures. Ex. 8.b) I

Suposant que $pH|T \sim N\left(\mu_T, \sigma_T^2\right)$, amb $\alpha = 0.05$ dieu si

Podem assegurar que el pH és més alt al fermentar a 42°?

Els valors obtinguts a $T = 42°$ són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ i per $T = 43.5°$ són $\{4.38, 4.37, 4.33, 4.34, 4.31, 4.39\}$.

H_0 : $\mu_{42°} \leq \mu_{43.5°}$, ens permet fer càlculs ja que conté el " $=$ ".

H_1 : $\mu_{42°} > \mu_{43.5°}$, no ens permet fer càlculs ja que no conté el " $=$ ".És la volem escollir només si n'estem segurs.

$$\textit{suposant } \sigma_{42°}=\sigma_{43.5°}: \quad T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\left(\frac{1}{n_Y}+\frac{1}{n_Z}\right)S_C^2}} \sim t_{n_Y+n_Z-2}$$

$$\textit{suposant } \sigma_{42°}\neq\sigma_{43.5°}: \quad T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\left(\frac{S_Y^2}{n_Y}+\frac{S_Z^2}{n_Z}\right)}} aprox. \sim t_{gle}$$

El dia de la fermentació, les dues temperatures. Ex. 8.b) II

Resolució suposant $\sigma_{42^\circ} = \sigma_{43.5^\circ}$:

- *De les dades obtenim* $n = 6$, *per les dues* T, $\overline{pH}_{42^\circ} = 4.45$, $\overline{pH}_{43.5^\circ} = 4.3533$, $S^2_{42^\circ} = 0.00048$ *i* $S^2_{43.5^\circ} = 0.000987$.
- $S_C^2 = \frac{6 \cdot 0.00048 + 6 \cdot 0.000987}{6+6} = 0.000733$ *i* $T_{calc} = \frac{4.45 - 4.3533}{\sqrt{\left(\frac{1}{6}+\frac{1}{6}\right)0.000733}} = 6.1828 \sim t_{12-2}$
- *Forma de la regió d'acceptació* $(-\infty, \delta)$, *una cua de rebuig* (δ, ∞)
- *La regió d'acceptació amb el* p_{valor} *serà* $(-\infty, 6.1828)$
- *Per tant* $p_{valor} = \Pr(t_{10} > 6.1828) = 5.18 \cdot e - 5 < 0.05 \Rightarrow$ *Significatiu*

Conclusió:

$\mu_{42^\circ} > \mu_{43.5^\circ}$, ***n'estem "molt segurs"***, $\mathbf{Pr}\,(\mathit{Error}) \leq \mathbf{5.18e - 5}$

El dia de la fermentació, les dues temperatures. Ex. 8.b) III

Resolució suposant $\sigma_{42^\circ} \neq \sigma_{43.5^\circ}$:

- Els mateixos càlculs i forma de la regió d'acceptació del cas anterior.
- $T_{calc} = \frac{4.45-4.3533}{\sqrt{\left(\frac{0.00048}{6}+\frac{0.000987}{6}\right)}} = 6.1828 \sim aprox.t_{gle}$

$$gle = \frac{\left(\frac{S_Y^2}{n_Y}+\frac{S_Z^2}{n_Z}\right)^2}{\frac{S_Y^4}{n_Y^2(n_Y-1)}+\frac{S_Z^4}{n_Z^2(n_Z-1)}} = 8.934$$

Nota: Surt el mateix valor de la T_{calc} si la n és la mateixa pels 2 casos.

Sempre $gle \leq n_1 + n_2 - 2$, es compleix la igualtat quan $S_Y^2 = S_Z^2$ i va disminuint a mesura que les variàncies són més diferents.

- *Per tant* $p_{valor} = \Pr\left(t_{8.934} > 6.183\right) = 8.36e-5 < 0.05 \Rightarrow$ *Significatiu*

Conclusió:

$\mu_{42^\circ} > \mu_{43.5^\circ}$, ***n'estem "molt segurs", Pr (Error) $\leq 8.36e-5$***

El dia de la fermentació, les dues temperatures. Ex. 8.c) I

Suposant que $pH|T \sim N\left(\mu_T, \sigma^2_T\right)$, amb $\alpha = 0.05$ dieu si

Les variàncies del pH a les dues temperatures

són iguals? o no?

Els valors per $T = 42^\circ$ són $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$
i per $T = 43.5^\circ$ són $\{4.38, 4.37, 4.33, 4.34, 4.31, 4.39\}$.

Hipòtesis:

H_0 : $\sigma^2_{42^\circ} = \sigma^2_{43.5^\circ}$, ens permet fer càlculs ja que conté el " $=$ ".

H_1 : $\sigma^2_{42^\circ} \neq \sigma^2_{43.5^\circ}$, no ens permet fer càlculs ja que no conté el " $=$ ".

Estadístic: $T = \frac{S^2_Y}{S^2_Z} \sim F\left(n_Y - 1, n_Z - 1\right)$

El dia de la fermentació, les dues temperatures. Ex. 8.c) II

Resolució :

- *Tenim* $n = 6$, $S^2_{42^\circ} = 0.00048$ *i* $S^2_{43.5^\circ} = 0.000987$.
- $T_{calc} = \frac{0.00048}{0.000987} = 0.4865 \sim F(6-1, 6-1)$
- *Forma de la regió d'acceptació, dues cues de rebuig* (a, b).
- *La regió d'acceptació amb el* p_{valor} *a la frontera pot ser* $(a, 0.4865)$ *o bé* $(0.4865, b)$.
  Com que $\Pr(F(5,5) < 0.4865) = 0.224 < 0.5$ *serà la* $(0.4865, b)$.
- *Per simetria* $p_{valor} = 2 \cdot \Pr(F(5,5) < 0.4865) = 0.448 > 0.05 \Rightarrow$ ***No Significatiu***

Conclusió:

Acceptem $\boldsymbol{\sigma^2_{42^\circ} = \sigma^2_{43.5^\circ}}$***, però no n'estem "segurs"***

Exemples, exercici 9)

S'ha fet un tast per conèixer les preferències dels consumidors respecte a dos vins A i B, cada tastador puntua els dos vins. Els resultats obtinguts són:

Consum.	1	2	3	4	5	6	7	8	9	10
A	5.3	4.1	5	5.7	5.8	7.2	3.4	5	6.2	5.8
B	4.7	5	6.2	5.6	6.4	7.9	4.4	5.5	6.5	5.7

(a) Contrasteu amb $\alpha = 0.05$ el test adient per contestar la pregunta:

Els consumidors prefereixen algun dels dos vins?

(b) Dieu quines suposicions heu fet per poder contrastar el test.

9-Solució

És un exemple de dades aparellades, en que l'estadístic de contrast és: $T = \frac{\overline{Y}-\overline{Z}}{\sqrt{\frac{S^2_{Y-Z}}{n}}} \sim t_{n-1}$.

En aquest cas $E[Y] = \mu_A$, $E[Z] = \mu_B$, $n = 10$,

$\boldsymbol{y} - \boldsymbol{z} = 0.6, -0.9, -1.2, 0.1, -0.6, -0.7, -1.0, -0.5, -0.3, 0.1 \Rightarrow$

$\overline{Y} - \overline{Z} = \overline{Y - Z} = -0.44$ i $S^2_{Y-Z} = 0.3204$

9-Solució (a)

- Les hipòtesis són $H_0 : \mu_A = \mu_B$ vs $H_1 : \mu_A \neq \mu_B$
- $T_{calc} = \frac{\overline{Y}-\overline{Z}}{\sqrt{\frac{S^2_{Y-Z}}{n}}} = \frac{-0.44}{\sqrt{\frac{0.3204}{10}}} = -2.458 \sim t_{10-1}$
- La forma de la regió d'acceptació és $[-\delta, \delta]$ per tant posant T_{calc}ala frontera queda $[-2.458, 2.458]$
- Per simetria el $p_{valor} = 2\Pr(t_9 < -2.458) = 0.03628$

Conclusió:

$$\mu_A \neq \mu_B, \textit{ n'estem ``segurs''}, \Pr(\textit{Error}) \leq 0.03628$$

9-Solució (b)

L'única suposició que es necessita és que
$Y - Z \sim N\left(\mu_{Y-Z}, \sigma^2_{Y-Z}\right)$.
No cal que Y i Z siguin normals, però si ho són la diferència $Y - Z$ també ho és.

Test de raó de versemblança I

Definició models encaixats

- Si tenim dos models estadístics $m_A(\boldsymbol{y}, \theta_A)$ i $m_B(\boldsymbol{y}, \theta_2)$.

Direm que són encaixats, $m_A(\boldsymbol{y}, \theta_A) \subset m_B(\boldsymbol{y}, \theta_B)$, si $\forall \theta_A = (\theta_{A,1}, \ldots, \theta_{A,k_A})$ sempre hi ha uns valors de θ_B, $\tilde{\theta}_B = \left(\tilde{\theta}_{B,1}, \ldots, \tilde{\theta}_{B,k_B}\right)$, de forma que $m_A(\boldsymbol{y}, \theta_A) = m_B\left(\boldsymbol{y}, \tilde{\theta}_B\right)$

Test de raó de versemblança II

Exemple de factors encaixats

En l'enunciat de l'exercici 8) tenim els pH en dues temperatures de fermentació. Podem considerar:

- $m_A\left(\theta_A, pH\right) = pH|T \sim N\left(\mu, \sigma^2\right)$ independentment de la temperatura o sigui $\theta_A = (\mu, \sigma)$
- $m_B\left(\theta_B, pH\right) = pH|T \sim N\left(\mu_T, \sigma_T^2\right)$ o sigui $\theta_B = \left(\mu_{42^\circ}, \mu_{43.5^\circ}, \sigma_{42^\circ}, \sigma_{43.5^\circ}\right)$
- Si agafem $\tilde{\theta}_B = \left(\tilde{\mu}_{42^\circ} = \mu, \tilde{\mu}_{43.5^\circ} = \mu, \tilde{\sigma}_{42^\circ} = \sigma, \tilde{\sigma}_{43.5^\circ} = \sigma\right)$ aleshores $m_A\left(\theta_A, pH\right) = m_B\left(\tilde{\theta}_B, pH\right)$

Test de raó de versemblança III

Test de raó de versemblança

Tenim els dos models encaixats $m_A\left(\boldsymbol{y},\theta_A\right) \subset m_B\left(\boldsymbol{y},\theta_B\right)$.

Volem comprovar que el model B , que té més paràmetres: $k_B > k_A$, millora el model A, és a dir si B és més versemblant que A.

Hipòtesis:

- H_0 $m_B\left(\boldsymbol{y},\theta_B\right)$ es tant versemblant com $m_A\left(\boldsymbol{y},\theta_A\right)$
- H_1 $m_B\left(\boldsymbol{y},\theta_B\right)$ es més versemblant que $m_A\left(\boldsymbol{y},\theta_A\right)$

Estadístic:

- $T_{calc} = 2\left(\ell_B\left(\hat{\theta}_B,\boldsymbol{y}\right) - \ell_A\left(\hat{\theta}_A,\boldsymbol{y}\right)\right) \sim \chi^2_{k_B-k_A}$ asimptòticament.

$p_{valor} = \Pr\left(\chi^2_{k_B-k_A} > T_{calc}\right)$

Exemples de test de raó de versemblança *lrt* I

Exercici 10) En l'exercici 1) teníem:

En un punt d'una carretera amb poc tràfic comptem el nombre de vehicles que passen durant 10 minuts, considerem que el recompte segueix una distribució Poisson (λ) *d'esperança* λ *desconeguda. S'han fet 5 recomptes i s'ha obtingut la mostra* $\boldsymbol{y} = \{11, 6, 8, 5, 8\}$.

Aquests recomptes s'havien fet en dies sense pluja. S'ha fet també recomptes en dies de pluja, obtenint-se $\boldsymbol{z} = \{8, 11, 7, 10\}$.

(a) Estimeu per màxima versemblança el nombre esperat de vehicles els dies de pluja.

(b) Utilitzant el test de raó de versemblança, amb $\alpha = 0.05$, dieu si passen el mateix nombre de cotxes tant els dies de pluja com els dies que no plou.

Exemples de test de raó de versemblança *lrt* II

10) - Solució: càlculs amb la *Poisson*

Per una *Poisson* (λ) amb la mostra $\boldsymbol{y}$, tenim:

- La funció log-versemblança és:
  $\ell(\lambda; y) = -N\lambda + \log(\lambda)\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \log(y_i!)$
- L'estimador màxim versemblant de λ és $\hat{\lambda} = \overline{\boldsymbol{y}}$

Per tant $\ell\left(\hat{\lambda}; y\right) = -N\overline{\boldsymbol{y}} + \log(\overline{\boldsymbol{y}})\, N\overline{\boldsymbol{y}} - \sum_{i=1}^{N} \log(y_i!)$

Ja havíem vist que el recompte dels dies sense pluja és $Y \sim$ *Poisson* (λ_Y) amb $\hat{\lambda}_Y = \overline{\boldsymbol{y}} = 7.6$

Exemples de test de raó de versemblança *lrt* III

10) - Solució (a)

El recompte dels dies amb pluja és $Z \sim Poisson\left(\lambda_Z\right)$ amb $\hat{\lambda}_Z = \overline{z} = 9$

Exemples de test de raó de versemblança *lrt* IV

10) - Solució (b): test $H_0 : \lambda_Y = \lambda_Z$ vs $H_1 : \lambda_Y \neq \lambda_Z$

Tenim els models encaixats:

m_A: Considera que, tant si plou com si no, tenim la mateixa *Poisson* (λ_U). La seva mostra és la unió $\boldsymbol{u} = \boldsymbol{y} \cup \boldsymbol{z} \Rightarrow$ $N_U = 9$, $\hat{\lambda}_U = \overline{u} = 8.22$ i
$\ell\left(\hat{\lambda}_U; \boldsymbol{u}\right) = -9 \cdot 8.22 + \log(8.22)\, 9 \cdot 8.22 - \sum_{i=1}^{9} \log(u_i!) = -19.90854$

m_B: Considera que si no plou tenim *Poisson* (λ_Y) amb la mostra $\boldsymbol{y}$ i si plou és *Poisson* (λ_Z) amb la mostra $\boldsymbol{z}$. Sabem que $\hat{\lambda}_Y = \overline{\boldsymbol{y}} = 7.6$ i que $\hat{\lambda}_Z = \overline{\boldsymbol{z}} = 9 \Rightarrow$
$\ell = -5 \cdot 7.6 - 4 \cdot 9 + \log(7.6)\, 5 \cdot 7.6 + \log(9)\, 4 \cdot 9 - \sum_{i=1}^{9} \log(u_i!) = -19.64502$

Exemples de test de raó de versemblança *lrt* V

10) - Solució (b) continuació

L'estadístic de contrast calculat serà:

$2\left(-19.64502-(-19.90854)\right)=0.52704\sim\chi^2_{2-1}$ per tant:

$p_{valor}=\Pr\left(\chi^2_1>0.52704\right)=0.467854>0.05$ No significatiu

No podem assegurar que el nombre de vehicles que passen sigui diferent segons si plou o no

Exemples de test de raó de versemblança *lrt* VI

Exercici 11) En l'exercici 3) teníem:

No sabem quina és la probabilitat (p) *de germinació d'unes llavors, però n'hem comptat el nombre de germinades de* 3 *safates de* 144 *alvèols/safata i hem obtingut* $\mathbf{y} = (106, 112, 124)$.

Tractàvem les tres safates com si les llavors fossin les mateixes, però sabem que abans de l'experiència s'havien conservat de formes diferents que anomenarem $\mathbf{c} = (C1, C2, C3)$.

(a) Per a cada tipus de conservació estimeu (m.v) $\hat{p} = (\hat{p}_1, \hat{p}_2, \hat{p}_3)$

(b) Utilitzant *lrt* contesteu la pregunta:

El tipus de conservació té algun efecte sobre les p_i*?*

(c) Si fos el cas, dieu quines diferències hi ha entre les p_i.

Exemples de test de raó de versemblança *lrt* VII

11) - Solució: càlculs amb la *Binomial*

Per una *Binomial* (m,p) amb la mostra $\boldsymbol{y}$, tenim:

- La funció log-versemblança és:
  $\ell = \sum_{i=1}^{N} \log \binom{m}{y_i} + \log p \sum_{i=1}^{N} y_i + \log(1-p)\left(N \cdot m - \sum_{i=1}^{N} y_i\right)$
- L'estimador màxim versemblant és $\hat{p} = \frac{\overline{\boldsymbol{y}}}{m}$

Per tant $\ell = \sum_{i=1}^{N} \log \binom{m}{y_i} + \log \frac{\overline{y}}{m} N\overline{y} + \log\left(1 - \frac{\overline{y}}{m}\right)\left(N \cdot m - N\frac{\overline{y}}{m}\right)$

Ja havíem vist que el recompte considerant que no hi ha efecte de la conservació és $Y \sim Binomial\,(p; \boldsymbol{y}, 144)$ amb $\hat{p}_Y = \frac{\overline{\boldsymbol{y}}}{3} = 0.792$ i que $\ell(p) = 209.94 + \log(p)\,342 + \log(1-p)\,90$ amb el que $\ell(\hat{p}) = 209.94 + \log(\hat{p})\,342 + \log(1-\hat{p})\,90 = -11.13231$

Exemples de test de raó de versemblança *lrt* VIII

11) - Solució (a):

$\forall C_i,\ y_i \sim Binomial\,(m, p_i) \Rightarrow \hat{p}_i = \frac{\overline{y}_i}{m}$, com que per cada C_i només tenim una dada y_i i $m = 144$ aleshores:

$$\hat{p}_1 = \frac{106}{144} = 0.736$$

$$\hat{p}_2 = \frac{112}{144} = 0.778$$

$$\hat{p}_3 = \frac{124}{144} = 0.861$$

Exemples de test de raó de versemblança *lrt* IX

11) - Solució (b): test $H_0 : p_1 = p_2 = p_3$ vs $H_1 : \cancel{p_1 = p_2 = p_3}$

Nota: $\cancel{p_1 = p_2 = p_3} \Leftrightarrow \exists i, j : p_i \neq p_j$

Tenim els models encaixats:

m_A: Considera que en totes les conservacions tenim la mateixa $Binomial\,(m, p_0)$. La seva mostra $(106, 112, 124)$ i ja sabem $\ell\,(\hat{p}_0) = -11.1323$

m_B: Considera que la conservació té efecte, $\forall i,\ y_i \sim Binomial\,(m, p_i)$ aleshores:

$$\ell\,(\hat{p}_1, \hat{p}_2, \hat{p}_3; y, m) =$$

$$\sum_{i=1}^{3} \log \binom{m}{y_i} + \sum_{i=1}^{3} y_i \log \hat{p}_i + \sum_{i=1}^{3} (m - y_i) \log (1 - \hat{p}_i) =$$

$$= 209.94 + \sum_{i=1}^{3} y_i \log \frac{y_i}{m} + \sum_{i=1}^{3} (m - y_i) \log \left(1 - \frac{y_i}{m\,i}\right) =$$

$$= 209.94 - 79.1648 - 138.2367 = -7.4622$$

Exemples de test de raó de versemblança *lrt* X

11) - Solució (b) continuació

Estadístic calculat:

- $T = 2\left(\ell_c - \ell_0\right) = 2\left(-7.4622 - (-11.1323)\right) = 7.3403$

$p_{valor} = \Pr\left(\chi^2_{3-1} > 7.3403\right) = 0.0255$

Significatiu: $\cancel{p_1 = p_2 = p_3}$ **n'estem "segurs" ja que** $p_{valor} = 0.0255 < 0.05$

Exemples de test de raó de versemblança *lrt* XI

11) - Solució (c): 3 tests $H_0 : p_i = p_j$ vs $H_1 : p_i \neq p_j$

$H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$ és comparar els models (p_{12}, p_{12}, p_3) vs (p_1, p_2, p_3)

Calculem $\ell(\hat{p}_{12}, \hat{p}_{12}, \hat{p}_3) = -7.8022$ ja sabem que $\ell(\hat{p}_1, \hat{p}_2, \hat{p}_3) = -7.4622 \Rightarrow$

$T_{12} = 0.68$, $p_{valor} = \Pr\left(\chi_1^2 > 0.68\right) = 0.41 > 0.05$

No significatiu: no detectem diferències entre p_1 i p_2

$H_0 : p_1 = p_3$ vs $H_1 : p_1 \neq p_3$ és comparar els models (p_{13}, p_2, p_{13}) vs (p_1, p_2, p_3)

$T_{13} = 7.09$, $p_{valor} = 0.0077 < 0.05$

Significatiu: segur que $p_1 \neq p_3$

$H_0 : p_2 = p_3$ vs $H_1 : p_2 \neq p_3$ és comparar els models (p_1, p_{23}, p_{23}) vs (p_1, p_2, p_3)

$T_{23} = 3.40$, $p_{valor} = 0.065 > 0.05$

No significatiu: no detectem diferències entre p_2 i p_3