Batlle Casellas, Àlex

1. Sigui Y una variable aleatòria Gamma amb funció de densitat:

$$f(y;\beta) = \frac{\beta^{18.7}}{\Gamma(18.7)} \cdot y^{17.7} \cdot e^{-\beta \cdot y}$$

amb $y \geq 0$ i el paràmetre $\beta > 0$ desconegut.

A més teniu $\boldsymbol{y}$, una mostra aleatòria de Y, al fitxer "Batlle Casellas, Àlex E1 ap1.csv" que hi ha al campus virtual.

Trobeu la funció de versemblança $\mathcal{L}(\beta; \boldsymbol{y})$ i també la funció de log-versemblança $\ell(\beta; \boldsymbol{y})$ que va millor per fer els càlculs.

Càlculs:

Logaritme de la densitat: $\log f(y;\beta) = -\log\Gamma(18.7) + 18.7 \cdot log\beta + 17.7 \cdot \log y - \beta \cdot y$

La grandària de la mostra és 58.

Funció de versemblança:$\mathcal{L}(\beta; \boldsymbol{y}) = \prod_{i=1}^{58} f(y_i;\beta) = \prod_{i=1}^{58} \left(\frac{\beta^{18.7}}{\Gamma(18.7)} \cdot y_i^{17.7} \cdot e^{-\beta \cdot y_i} \right)$

Funció de log-versemblança: $\ell(\beta; \boldsymbol{y}) = \sum_{i=1}^{58} \log f(y_i;\beta) = \sum_{i=1}^{58} (-\log\Gamma(18.7) + 18.7 \cdot log\beta + 17.7 \cdot \log y_i - \beta \cdot y_i) =$
$= 58 \cdot \left(-\log\Gamma(18.7) + 18.7 \cdot log\beta + 17.7 \cdot \overline{\log \boldsymbol{y}} - \beta \cdot \overline{\boldsymbol{y}}\right) = 58 \cdot (-35.523 + 18.7 \cdot log\beta + 17.7 \cdot -0.098032 - \beta \cdot 0.93379)$

$\overline{\boldsymbol{y}} = 0.93379$ i $\overline{\log \boldsymbol{y}} = -0.098032$ per tant

$\ell(\beta; \boldsymbol{y}) = -2160.9 + 1084.6 \cdot log\beta - 54.16 \cdot \beta$ i $\mathcal{L}(\beta; \boldsymbol{y}) = e^{-2160.9 + 1084.6 \cdot log\beta - 54.16 \cdot \beta}$

(a) Quin dels dos valors és més versemblant, $\beta = 18$ o $\beta = 19$?

Contesteu:

1)	El valor de $\mathcal{L}(18; \boldsymbol{y})$ ha donat: $e^{-2160.9+1084.6 \cdot log(18) - 54.16 \cdot 18} = \mathbf{0.39512}$
2)	El valor de $\ell(19; \boldsymbol{y})$ ha donat: $-2160.9 + 1084.6 \cdot log(19) - 54.16 \cdot 19 = \mathbf{3.5527}$
3)	Quin dels dos valor de β, $\{18, 19\}$ és més versemblant: **b)** a) $\beta = 18$ b) $\beta = 19$ c) No es pot respondre.

(a) Calculeu $\hat{\beta}$, que és l'estimació pel mètode de màxima versemblança del paràmetre β.

$\frac{\partial}{\partial\beta}\ell(\beta; \boldsymbol{y}) = \frac{\partial}{\partial\beta}\ell(-2160.9 + 1084.6 \cdot log\beta - 54.16 \cdot \beta) = \frac{1084.6}{\beta} - 54.16$

$\frac{\partial^2}{\partial\beta^2}\ell(\beta; \boldsymbol{y}) = \frac{\partial}{\partial\beta}\ell\left(\frac{1084.6}{\beta} - 54.16\right) = -\frac{1084.6}{\beta^2} < 0$

Contesteu:

4)	El valor de $\hat{\beta}$ ha donat: $\frac{1084.6}{\hat{\beta}} - 54.16 = 0 \Rightarrow \hat{\beta} = \frac{1084.6}{54.16} = \mathbf{20.026}$
5)	$\mathcal{L}\left(\hat{\beta}; \boldsymbol{y}\right) = e^{-2160.9+1084.6 \cdot log(20.026) - 54.16 \cdot 20.026} = \mathbf{152.38}$
6)	$\frac{\partial^2}{\partial\beta^2}\ell\left(\hat{\beta}; \boldsymbol{y}\right) = -\frac{1084.6}{20.026^2} = \mathbf{-2.7045}$

(b) Trobeu $IC_{95\%}(\beta)$, l'interval de confiança asimptòtic de β, amb una confiança del 95% i dues cues iguals.

Asimptòticament $\hat{\beta} \sim N\left(\beta, S_{\hat{\beta}}^2\right)$ amb $S_{\hat{\beta}}^2 = -E\left[\frac{\partial^2}{\partial\beta^2}\ell\left(\hat{\beta}; \boldsymbol{y}\right)\right]^{-1} = -\frac{1}{-2.7045} \Rightarrow S_{\hat{\beta}} = 0.60807$, per tant l'interval de confiança serà $\left(\hat{\beta} - z_{97.5\%}S_{\hat{\beta}}, \hat{\beta} - z_{97.5\%}S_{\hat{\beta}}\right)$

Contesteu:

7)	L'extrem inferior de $IC_{95\%}(\beta)$ és: $20.026 - 1.96 \cdot 0.60807 = \mathbf{18.834}$
8)	L'extrem superior de $IC_{95\%}(\beta)$ és: $20.026 + 1.96 \cdot 0.60807 = \mathbf{21.218}$
9)	La variància asimptòtica de $\hat{\beta}$ és: $0.60807^2 = \mathbf{0.36975}$

(c) Utilitzant el test de raó de versemblança contrasteu les hipòtesis:

$$H_0: \quad \beta = 18$$
$$H_1: \quad \beta \neq 18$$

$T_{calc} = 2 \cdot \left(\ell\left(\hat{\beta}; \boldsymbol{y}\right) - \ell(18; \boldsymbol{y})\right) = 2 \cdot (\log(152.38) - \log(0.39512))$ de distribució asimptòtica χ_1^2 amb el que el $p_{valor} = \Pr\left(\chi_1^2 > T_{calc}\right)$

Contesteu:

10)	L'estadístic de contrast del test a donat: **11.91**
11)	El p_{valor}del test ha donat: **0.00055838**
12)	Escolliu l'opció correcte: **b)** a) No n'estem segurs però acceptem que $\beta \neq 18$ b) Estem segurs que $\beta \neq 18$ c) No n'estem segurs però acceptem que $\beta = 18$ d) Estem segurs que $\beta = 18$ e) Cap de les anteriors

2. S'han fet dos tasts per estudiar la acidesa de 4 genotips de tomàquets (Bodar, Durinta, Montgri i Pebrot) però codificats G495, G465, G537 i G436, per no influir en els resultats de les valoracions, ja que els tastadors no sabien a quin genotip corresponia cada codi.
El 1r tast el va fer un panell 20 especialistes, amb els tomàquets presentats en trossos sencers,
el 2n tast el va fer un altre panell, també de 20, però amb els tomàquets triturats.
En els dos casos cada tastador va valorar els 4 genotips. Els resultats els teniu al fitxer "Batlle Casellas, Àlex E1 ap2.csv", que també està al campus virtual, cada fila correspon a les puntuacions d'un tastador.
Quan sigui necessari suposeu que la variable aleatòria $acidesa_{i,j} \sim N\left(\mu_{i,j}, \sigma^2\right)$ on i indica el genotip i $j \in (sencer, triturat)$, i utilitzeu un nivell de significació del 1%.

(a) Per a cadascuna de les 8 combinacions de genotip i panell (forma), calculeu-ne les mitjanes i les desviacions tipus.

Forma	Sencer				Triturat			
Genotip	G495	G465	G537	G436	G495	G465	G537	G436
Mitjana	6.385	7.55	6.91	7.12	3.095	5.65	4.5	6.91
Desv.Tipus	1.9513	1.7698	2.507	2.2571	1.4544	2.7018	2.1046	2.507

Contesteu:

13)	El genotip que en trossos sencers ha obtingut la valoració més alta és: **b) G465** a) G495 b) G465 c) G537 d) G436
14)	La valoració més baixa dels genotips triturats és: **3.095**
15)	El genotip que triturat ha obtingut una valoració més dispersa és: **b) G465** a) G495 b) G465 c) G537 d) G436

(b) Per la variable aleatòria acidesa del genotip G495 en trossos sencers, calculeu $IC_{99\%}(\mu)$ i $IC_{99\%}(\sigma)$, intervals de confiança de dues cues iguals.
En aquesta situació l'acidesa és una variable aleatòria $Y \sim N\left(\mu, \sigma^2\right)$ amb una mostra $\boldsymbol{y}$ de grandària 20, $\overline{\boldsymbol{y}} = 6.385$ i $S = 1.9513$. Per tant $IC_{99\%}(\mu) = \left(\overline{\boldsymbol{y}} - t_{n-1,99.5\%}\frac{S}{\sqrt{n}}, \overline{\boldsymbol{y}} + t_{n-1,99.5\%}\frac{S}{\sqrt{n}}\right)$ i $IC_{99\%}\left(\sigma^2\right) = \left(\frac{(n-1)S^2}{\chi_{n-1,99.5\%}}, \frac{(n-1)S^2}{\chi_{n-1,0.5\%}}\right)$

Contesteu:

16)	L'extrem superior de $IC_{99\%}(\mu)$ és: $6.385 + 2.8609\frac{1.9513}{\sqrt{20}} = \mathbf{7.6333}$
17)	L'extrem inferior de $IC_{99\%}(\mu)$ és: $6.385 - 2.8609\frac{1.9513}{\sqrt{20}} = \mathbf{5.1367}$
18)	La longitud de $IC_{99\%}(\sigma)$ és: $\sqrt{19 \cdot 1.9513^2}\left(\frac{1}{\sqrt{6.844}} - \frac{1}{\sqrt{38.582}}\right) = \mathbf{1.8819}$

(c) En el cas del genotip G436, plantegeu i contrasteu el test adequat per saber si els valors esperats de l'acidesa, en les dues formes (sencers,triturats), són iguals o no.

En aquest cas tenim les variables $Y_1 \sim N\left(\mu_1, \sigma^2\right)$ a "sencer", i $Y_2 \sim N\left(\mu_2, \sigma^2\right)$ a "triturat" (tenen la mateixa variància per l'enunciat). Tenim dues mostres independents, $\boldsymbol{y}_1$ i $\boldsymbol{y}_2$, de grandària 20 cadascuna i $\overline{\boldsymbol{y}}_1 = 7.12$, $\overline{\boldsymbol{y}}_2 = 6.91$, $S_1 = 2.2571$ i $S_2 = 2.507$. Per tant $T_{calc} = \frac{\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2}{S_C \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ de distribució $t_{n_1+n_2-2}$, on $S_c^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

Nota: segons l'ordre en que fem la diferència $\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2$, T_{calc} pot canviar de signe. En ser de dues cues el $p_{valor} = 2 \cdot \Pr\left(t_{n_1+n_2-2} < -\left|T_{calc}\right|\right)$

El resultat de $S_c = \sqrt{\frac{2.7018^2 + 2.1046^2}{2}} = 2.3853$.

Contesteu:

19)	El càlcul de l'estadístic de contrast ha donat: $\frac{7.12-6.91}{2.3853\sqrt{\frac{2}{20}}} = \mathbf{0.2784}$ pot ser $(\pm)$
20)	El p_{valor} del test és: $2 \cdot \Pr\left(t_{38} < -\lvert 0.2784\rvert\right) = \mathbf{0.78221}$
21)	Escolliu l'opció correcte: **b)** a) No n'estem segurs però acceptem són diferents. b) No n'estem segurs però acceptem que són iguals. c) Estem segurs que són diferents. d) Estem segurs que són iguals e) Cap de les anteriors

(d) En el cas dels genotips G465 i G537 triturats, plantegeu i contrasteu el test adequat per saber si els valors esperats de l'acidesa, en els dos genotips, són iguals o no.

Ara tenim tenim les variables $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ a G465, i $Y_2 \sim N\left(\mu_2, \sigma^2\right)$ a G537 . Tenim dues mostres que no són independents, són aparellades pel tastador, $\boldsymbol{y}_1$ i $\boldsymbol{y}_2$, de grandària 20 cadascuna i $\overline{\boldsymbol{y}}_1 = 5.65$, $\overline{\boldsymbol{y}}_2 = 4.5$ i $S_{\boldsymbol{y}_1-\boldsymbol{y}_2} = 3.193$. Per tant $T_{calc} = \frac{\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2}{S_{\boldsymbol{y}_1-\boldsymbol{y}_2}\sqrt{\frac{1}{n}}}$ de distribució t_{n-1}. Nota: segons l'ordre en que fem la diferència $\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2$, T_{calc} pot canviar de signe. En ser de dues cues el $p_{valor} = 2 \cdot \Pr\left(t_{n-1} < -\left|T_{calc}\right|\right)$.

Contesteu:

22)	El càlcul de l'estadístic de contrast ha donat: $\frac{5.65-4.5}{3.193\sqrt{\frac{1}{20}}} = \mathbf{1.6107}$ pot ser $(\pm)$
23)	El p_{valor} del test és: $2 \cdot \Pr\left(t_{19} < -\lvert 1.6107\rvert\right) = \mathbf{0.12373}$
24)	Escolliu l'opció correcte: **c)** a) No n'estem segurs però acceptem són diferents. b) Estem segurs que són diferents. c) No n'estem segurs però acceptem que són iguals. d) Estem segurs que són iguals e) Cap de les anteriors

Al qüestionari d'Atenea escriviu les respostes numèriques amb 5 (o més) xifres significants.

Si convé es pot utilitzar la notació científica, per exemple: $1.2345e-9$