CORRECCION DE LOS EJERCICIOS

CAPITULO 3

1. La función de densidad conjunta de la muestra es

$$L(x_1, ..., x_n/\alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta \sum x_i} \prod_{i=1}^{n} x_i^{\alpha-1}$$

$$\log(L(x_1, .., x_n/\alpha, \beta)) = n\alpha \log(\beta) - n\log(\Gamma(\alpha)) - n\bar{(}x)\beta + (\alpha - 1)\sum_{i=1}^{N} \log(x_i)$$

Las condiciones necesarias para que el par $(\hat{\alpha}, \hat{\beta})$ sea un punto de máximo son

$$\frac{\partial L(.)}{\partial \alpha} = 0 \iff \log(\hat{\beta}) - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$$

y

$$\frac{\partial L(.)}{\partial \beta} = 0 \iff \frac{\hat{\alpha}}{\hat{\beta}} = \bar{x}$$

Recordemos ahora el Principio de Invarianza. Si $T : \Omega_1 \to \Omega_2$ es una transformación **biyectiva** y $\hat{\theta}$ el estimador de M.V. de θ, entonces el estimador de M.V. de $T(\theta)$ es $T(\hat{\theta})$.
Para aplicar esto aquí hacemos $\Omega_1 = \Omega_2 = \{(\alpha, \beta)\alpha > 0, \beta > 0\}$ y definimos la transformación T como $T(\alpha, \beta) = (\alpha, \frac{\alpha}{\beta})$
Notando que el Jacobiano de T es no nulo en Ω_1 se deduce que T es biyectiva [1]. Con esto

$$\frac{\hat{\alpha}}{\hat{\beta}} = \bar{x}$$

Como E(X)=$\frac{\alpha}{\beta}$ y $E(\bar{x}) = \frac{\alpha}{\beta}$, el estimador es insesgado.
Además

$$Var(\bar{x}) = \frac{1}{n}Var(X) = \frac{1}{n}\frac{\alpha}{\beta^2}$$

Entonces $\lim_{n\to+\infty} Var(\bar{x}) = 0$. Con esto y la desigualdad de Chebichef

$$I\!P(|\bar{x} - \frac{\alpha}{\beta}| > \epsilon) \leq \frac{Var(\bar{x})}{\epsilon^2} \quad \forall \epsilon > 0$$

[1] Teorema oe ta FunciöW Inversa

se tiene que $\bar{x}$ es consistente.
Como $E(|\bar{x}-\frac{\alpha}{\beta}|^2)=Var(\bar{x})$, $\bar{x}$ converge en media cuadrática.

2. $L(x_1,...x_n/\theta)=\theta^n\prod_{j=1}^n x_j^{\theta-1}$ entonces, $\log(L(x_1,...x_n/\theta))=n\log(\theta)-n+\sum_{j=1}^n\log(x_j)$ de donde $\hat{\theta}=\frac{n}{\sum_{j=1}^n -\log(x_j)}$
Notemos ahora que $z_j=-\log(x_j)$ tiene una densidad Exponencial [2] de parámetro θ, y que calcular la esperanza de $\hat{\theta}$ se reduce a calcular $E(\frac{1}{Z})$ con Z la suma de n v.a. Exponenciales de parámetro θ. Es fácil mostrar, usando funciones características, que Z tiene una densidad Gamma(n,θ). Así

$$E(\frac{1}{Z})=\frac{1}{\Gamma(n)}\int_0^\infty Z^{-1}\theta^n Z^{n-1}e^{-\theta}dZ$$

usando la propiedad $\Gamma(n)=n\Gamma(n-1)$, se obtiene

$$\frac{\theta^2}{n(n-1)\Gamma(n-2)}\int_0^\infty \theta^{n-2}z^{(n-2)-1}e^{-\theta z}dz$$

de donde $E(Z^{-1})=\frac{n-2\theta}{n(n-2)}$, ya que la integral anterior corresponde a la Esperanza de una densidad Gamma(n-2,θ) y por lo tanto vale $\frac{n-2}{\theta}$. Con esto, E($\hat{\theta}$) $=\frac{n-2}{n-1}\theta$ i.e. $\hat{\theta}$ es sesgado. (Subestima el valor de θ). Notando que $\lim_{n\to\infty}\frac{n-2}{n-1}=1$, se deduce que el estimador es asintoóticamente insesgado.
Además, $Var(\hat{\theta})=E(\hat{\theta}^2-E(\hat{\theta}))^2=E(\hat{\theta}^2-\theta)^2$ y, haciendo un cálculo análogo al anterior, resulta que $E(\hat{\theta}^2)=\frac{n\theta}{n(n-1)}$. De esto se deduce que $Var(\hat{\theta})=\theta(\frac{n^2}{n(n-1)}-1)$. Con esto, $\lim_{n\to+\infty}Var(\hat{\theta})=0$ y , por lo tanto, convergente en media cuadrática. Esto implica (ver ejercicio 1) que $\hat{\theta}$ es consistente.

3. La función de probabilidad de Y puede escribirse como $f(y)=\theta^y(1-\theta)^{1-y}\quad y\in\{0,1\}$
La función de verosimilitud es: $L(y_1,...y_n/\theta)=\theta^{n\bar{y}}(1-\theta)^{n-n\bar{y}}$
La densidad a posteriori de θ es proporcional a: $L(y_1,...y_n,\theta)\theta^a(1-\theta)^{1-b}=\theta^{a+n\bar{y}}(1-\theta)^{1-b+n-n\bar{y}}$, que se reconoce como una $Beta(a+n\bar{y},b+n-n\bar{y})$. Como estamos usando una función de pérdida cuadrática $\hat{\theta}=E(\theta/y_1,...y_n)=\frac{a+n\bar{y}}{a+b+n}$, la esperanza de θ calculada con la f.d. a posteriori.
Como $E(\bar{y})=\theta$, se deduce que $\hat{\theta}$ es sesgado, pero $E(\hat{\theta})\longrightarrow\theta$, es decir es asintóticamente insesgado. Además $var(\bar{y})=\theta(1-\theta)/n$, es decir $Var(\hat{\theta})\longrightarrow 0$. Luego $\hat{\theta}$ es convergente en media cuadrática y es consistente.

[2]Usar wl Teorema oel Cambio oe tariables · @alcular oirectamente $\mathbb{P}(-\log(x_j)<C)$

4. a) $I\!P(SI) = I\!P(Q)I\!P(SI/Q) + I\!P(Q')I\!P(SI/Q') \Longrightarrow p = \theta\pi + (1-\theta)(1-\pi)$
b) Sea Y el número de personas que contestan SI; $Y \sim Bin(n,p)$.
El EMV de p es $\hat{p} = y/n \Longrightarrow \hat{\pi} = \dfrac{1-\hat{p}-\theta}{1-2\theta}$
$E(\hat{\pi}) = \dfrac{1-p-\theta}{1-2\theta} = \pi$ y $Var(\hat{\pi}) = \dfrac{p(1-p)}{n(1-2\theta)^2} = \dfrac{\pi(1-\pi)}{n} + \dfrac{\theta(1-\theta)}{n(1-2\theta)^2}$
c) El estimador $\hat{\pi}$ es insesgado y su varianza converge a 0, luego es consistente.
La varianza de $\hat{\pi}$ es máxima cuando $\theta = 0.5$; de toda manera $\hat{\pi}$ no esta definido en esta caso.

5. a) $E(\theta - T(\underline{X}))^2 = E(\theta^2 + T(\underline{X})^2 - 2\theta T(\underline{X}))$
$E(\theta T(\underline{X})) = E(E(\theta T(\underline{X})/\underline{X})) = E(T(\underline{X})E(\theta/\underline{X}))$
Pero $E(\theta/\underline{X}) = T(\underline{X})$ dado que es un estimador de Bayes bajo pérdida cuadrática.
Luego $E(\theta T(\underline{X})) = E(T(\underline{X})^2)$
Por otro lado $E(\theta T(\underline{X})) = E(E(\theta T(\underline{X})/\theta)) = E(\theta E(T(\underline{X})/\theta))$
Pero $E(T(\underline{X})/\theta) = \theta$ dado que $T(\underline{X})$ es insesgado, luego $E(\theta T(\underline{X})) = E(\theta^2)$
De aquí $E[(\theta - T(\underline{X}))^2] = 0$
b) $E[(\theta - T(\underline{X}))^2] = E[E(\theta - \bar{X})^2/\underline{X}] = E[Var(\bar{x}/\underline{X}] = 1/n$. De la parte a), se concluye que $\bar{X}$ no puede ser un estimador de Bayes.

6. a) $I\!P(y_i = 1) = \theta \Longrightarrow y_i \sim Bernoulli(\theta)$
b) $\hat{\theta} = \bar{y}$
c) $E(\hat{\theta}) = \theta$, $Var(\hat{\theta}) = \dfrac{\theta(1-\theta)}{n}$
d) Del ejercicio 3, se deduce que el estimador de Bayes para π_1 es $\hat{\theta}_1 = \dfrac{\alpha + n\bar{y}}{\alpha+\beta+n}$.
Para π_2, se tiene una distribución a posteriori $\xi_2(\theta/y_1, ...y_n) \propto (1-\theta)^{1+n-n\bar{y}}\theta^{n\bar{y}}$; lo que nos da una distribución $Beta(n\bar{y}+1, n-n\bar{y}+2)$ y el estimador de Bayes para π_2 es $\hat{\theta}_2 = \dfrac{n\bar{y}+1}{n+3}$.
$Var(\hat{\theta}_1) = \dfrac{n\theta(1-\theta)}{(\alpha+\beta+n)^2}$ y $Var(\hat{\theta}_2) = \dfrac{n\theta(1-\theta)}{(n+3)^2}$.
e) $\bar{y} = 0.4$; $\hat{\theta}_1 = 6/14 = 0.4286$ y $\hat{\theta}_2 = 5/13 = 0.38646$
$Var(\hat{\theta}_1) = 0.0510\ \theta(1-\theta)$ y $Var(\hat{\theta}_2) = 0.0592\ \theta(1-\theta)$.

7. a) $S^{*2} = \dfrac{\sum(Y_i^* - Y^*)^2}{n(n-1)} = S^{*2} = \dfrac{1}{n(n-1)}(\sum Y_i^* - (1/n)(\sum Y_i^*)^2) = S^{*2} = \dfrac{\sum(X_i - \bar{X}_n)^2}{n(n-1)}$
b) $(Y^* - \theta)/S^*$ sigue una Student a n-1 g.l.

8. a) $\xi(\theta_i/x) = \dfrac{f(x/\theta_i)\pi(\theta_i)}{\sum_i f(x/\theta_i)\pi(\theta_i)} \quad \forall i \in \Theta$
$E(L(\theta,\delta)/x) = \sum_{\theta_i} L(\theta_i,\delta)\xi(\theta_i/x) = c\sum_{\theta_i \neq \delta} xi(\theta_i/x) = c(1 - \xi(\delta/x))$
b) $Min_{\delta\in\Theta} E(L(\theta,\delta)/x) \Longleftrightarrow Min_{\theta_i\in\Theta} c(1-\xi(\theta_i/x)) \Longleftrightarrow Max_i \xi(\theta_i/x)$

c) Si $\pi(\theta_i) = \frac{1}{N} \Longrightarrow \xi(\theta_i/x)$ es proporcional a $f(x/\theta_i)/N$. Lo que equivale a maximizar la verosimilitud $f(x/\theta_i)$.

9. a) $I\!P(X=x) = 1 \Longrightarrow \sum_{x=0}^{\infty} \frac{a_x\theta^x}{h(\theta)} = 1 \Longrightarrow h(\theta) = \sum_{x=0}^{\infty} a_x\theta^x$

$\Longrightarrow h'(\theta) = \sum_{x=0}^{\infty} x a_x \theta^{x-1} = \frac{1}{\theta}\sum_{x=0}^{\infty} x a_x\theta^x$

b) $f_n(\underline{x}/\theta) = \prod_i^n a_{x_i}\frac{\theta^{\sum x_i}}{(h(\theta))^n}$, luego $\lambda_n(\underline{x}/\theta) = log f_n(\underline{x}/\theta) = \sum log a_{x_i} + \sum x_i log\theta - n log h(\theta)$

$\lambda_n'(\underline{x}/\theta) = \frac{\sum x_i}{\theta} - n\frac{h'(\theta)}{h(\theta)} = 0$. El EMV $\hat\theta$ de θ cumple: $\bar{x}_n = \hat\theta\frac{h'(\hat\theta)}{h(\hat\theta}$

c) $m_1(\theta) = E(X) = \sum x a_x\theta^x/h(\theta)$, luego $E(X) = \theta\frac{h'(\theta)}{h(\theta)}$ y $\bar{x}_n = \hat\theta\frac{h'(\hat\theta)}{h(\hat\theta)}$

d) Si $X \sim Bin(N,p)$, $I\!P(X=x) = a_x\theta^x/h(\theta)$, con $a_x = \begin{pmatrix} N \\ x \end{pmatrix}$ para $x = 0,1,...N$; $\theta = \frac{p}{1-p}$; $h(\theta) = (1+\theta)^N$

El EMV de θ cumple: $\bar{x} = \frac{N\hat\theta}{1+\hat\theta} = \hat{p}$

Si $X \sim \mathcal{P}(\lambda)$, $I\!P(X=x) = \lambda^x e^{-\lambda}/x!$, luego tomando $a_x = 1/x!$ y $\theta = \lambda$, se obtiene $\hat\lambda = \bar{x}$.

10. a) $E(T) = \theta \iff \sum_{i=1}^N \lambda_i(b_i+\theta) = \theta \iff \sum_{i=1}^N \lambda_i = 1$ y $\sum_{i=1}^N \lambda_i b_i = 0$ i.e. T debe ser una combinación convexa de los T_i.

b) Resulta cómodo usar notación vectorial. Sean $\underline{T} = (T_1 \ldots T_N)^t$ y $\underline{\lambda} = (\lambda_1 \ldots \lambda_N)^t$. Así $T = \underline{\lambda}^t\underline{T}$. entonces $Var(T) = \underline{\lambda}^t\Gamma\underline{\lambda}$, en que $\Gamma_{ij} = Cov(T_i,T_j)$; Γ es la matriz de varianza-covarianza de los T_i. El problema se puede plantear como $\min_{\underline{\lambda}} \underline{\lambda}^t\Gamma\underline{\lambda}$, con la restricción $\sum_{i=1}^N \lambda_i = 1$. Así es posible calcular los coeficientes óptimos si la matriz de varianza covarianza es conocida.

c) Si los T_i son no correlacionados, la matriz de varianzas-covarianzas es diagonal. Llamemos σ_i^2 al iésimo elemento de la diagonal de Γ ($\sigma_i^2 = Var(T_i)$). El problema de minimización anterior se escribe como $\min_{\underline{\lambda}} \sum_{i=1}^N \lambda_i^2\sigma_i^2$ sujeto a $\sum_{i=1}^N \lambda_i = 1$. Si se usan Multiplicadores de Lagrange, el problema se transforma en

$$\min_{(\lambda,\alpha)} L(\underline{\lambda},\alpha) = \sum_{i=1}^N \lambda_i^2\sigma_i^2 - \alpha(\sum_{i=1}^N \lambda_i - 1)$$

La condición de primer orden es : $\frac{\partial L(.)}{\partial\lambda_k} = 0 \iff 2\lambda_k = \frac{\alpha}{\sigma_k^2}\ \forall k$

Sumando esta igualdad de k=1 hasta N se obtiene el valor del multiplicador $\alpha = \frac{2}{\sum_{k=1}^N \frac{1}{\sigma_k^2}}$ y usando esto se obtiene el valor de λ_k: $\lambda_k = \frac{\alpha}{2\sigma_k^2} = \frac{1}{\sum_{i=1}^N \frac{\sigma_k^2}{\sigma_i^2}}$

d) Usaremos lo anterior. Es claro que los s_i^2 son estimadores insesgados de σ^2. Además,

como $\dfrac{(n_i-1)s_i^2}{\sigma^2}$ tiene distribución χ_{n_i-1}, se deduce que $var(s_i^2) = \dfrac{2\sigma^4}{n_i-1}$
Con esto $\dfrac{var(s_k^2)}{var(s_i^2)} = \dfrac{n_i-1}{n_k-1}$
$\sum_{i=1}^N \dfrac{var(s_k^2)}{var(s_i^2)} = \dfrac{\sum_{i=1}^N n_i - N}{n_k-1}$ $\lambda_i = \dfrac{(n_k-1)}{\sum_{i=1}^N n_i - N}$, que son exactamente los coeficientes del estimador S^2.

CAPITULO 4

1. a) El intervalo para θ es $I = [\bar{x} - 1.96\sigma/\sqrt{n}, \bar{x} + 1.96\sigma/\sqrt{n}]$ y tiene un largo igual a $3.96\sigma/\sqrt{n} \Longrightarrow \sqrt{n} > 392 \Longrightarrow n > 153664$
b) $\mathbb{P}(a \leq \theta \leq b) = 1-\alpha$, con $L = b - a = \sigma/5$
$\mathbb{P}(\frac{\sqrt{n}(\bar{x}-b)}{\sigma} \leq \frac{\sqrt{n}(\bar{x}-\theta)}{\sigma} \leq \frac{\sqrt{n}(\bar{x}-a)}{\sigma} = 1-\alpha \Longrightarrow \frac{\sqrt{n}(b-a)}{\sigma} = \frac{\sqrt{n}}{5}$

n	10	20	30	100
I	[-0.316, 0.316]	[-0.447, 0.447]	[-0.548, 0.548]	[-1.0, 1.0]
$1-\alpha$	0.43	0.52	0.55	0.76

c) Si σ^2 es desconocido, se toma $L = s/5$ y el estadistico t_{n-1}.

n	10	20	30	100
I	[-0.316, 0.316]	[-0.447, 0.447]	[-0.548, 0.548]	[-1.0, 1.0]
$1-\alpha$	0.25	0.35	0.40	0.70

Cuando n aumenta, el nivel de confianza aumenta también en ambas casos. Pero e el segundo caso el nivel es siempre más pequeño para un n dado.

2. $[36.08, 43.92]$

3. $\bar{x} = 247.3$, $s^2 = (1/n)\sum(x_i - \bar{x})^2 = 2.01$.
a) $\bar{x} \sim \mathcal{N}(\mu, 0.15) \Longrightarrow \mathbb{P}(\mid \dfrac{(\bar{x}-\mu)}{\sqrt{0.15}} \mid \leq 1.96) = 0.95$; el intervalo de confianza para μ es igual a $[\bar{x} - 1.96\sqrt{0.15}, \bar{x} + 1.96\sqrt{0.15}] = [246.54, 248.06]$
b) $\dfrac{\bar{x}-\mu}{s_n/3} \sim t_9$; luego $\mathbb{P}(\mid \dfrac{(\bar{x}-\mu)}{s_n/3} \mid \leq 2.26) = 0.95$; el intervalo de confianza para μ es igual a $[\bar{x} - 2.26 s_n/3, \bar{x} + 2.26 s_n/3] = [246.23, 248.36]$
Se observara que hacer la aproximación normal daría un intervalo más pequeño.
c) $ns^2/\sigma^2 \sim \chi_9^2$.
$\mathbb{P}(2.7 \leq ns^2/\sigma^2 \leq 19.02) = 0.95$; luego se obtiene el intervalo de nivel de confianza a 0.95%: $[ns^2/19.02, ns^2/2.7] = [1.057, 7.44]$.

4. $I\!P(a \leq \theta \leq b) = I\!P(\frac{n\bar{x}-nb}{\sqrt{n\theta(1-\theta)}} \leq \frac{n\bar{x}-n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{n\bar{x}-na}{\sqrt{n\theta(1-\theta)}})$
$= I\!P(-1.96 \leq \frac{n\bar{x}-n\theta}{\sqrt{n\theta(1-\theta)}} \leq 1.96) = 0.95$
$\Longrightarrow \frac{(n\bar{x}-n\theta)^2}{n\theta(1-\theta)} \leq (1.96)^2$
La inecuación que define el intervalo de confianza es entonces:

$$(3.84n + n^2)\theta^2 - (3.84n + 2n^2\bar{x})\theta + (n\bar{x})^2 \leq 0$$

5. Si t_o es tal que $I\!P(-t_o < t < t_o) = 0.95$, como n se supone grande, $t_o = 1.96$ y el intervalo es $[Y^* - 1.96S^*, Y^* + 1.96S^*]$.

6. a) $\sigma_2 = k\sigma_1$, entonces se puede construir un estadístico que no depende de σ_1^2 y σ_2^2:

$$\frac{\bar{x}_1 - \bar{x}_2 - \mu_1 - \mu_2}{\sqrt{(\frac{n_1\hat{\sigma}_1^2+n_2\hat{\sigma}_2^2/k^2}{n_1+n_2-2})(\frac{k^2n_1+n_2}{n_1n_2})}} \sim t_{n_1+n_2-2}$$

El intervalo es entonces igual a:

$$[\bar{x}_1-\bar{x}_2-t^{\alpha}_{n_1+n_2-2}\sqrt{(\frac{n_1\hat{\sigma}_1^2+n_2\hat{\sigma}_2^2/k^2}{n_1+n_2-2})(\frac{k^2n_1+n_2}{n_1n_2})}, \bar{x}_1-\bar{x}_2+t^{\alpha}_{n_1+n_2-2}\sqrt{(\frac{n_1\hat{\sigma}_1^2+n_2\hat{\sigma}_2^2/k^2}{n_1+n_2-2})(\frac{k^2n_1+n_2}{n_1n_2})}]$$

b) Las cotas del intervalo convergen en probabilidad hacia $\mu_1 - \mu_2$.
c) El método de los momentos produce un estimador de k igual a $\hat{k} = \frac{\sigma_2}{\sigma_1}$.

7. a) $C_a(x) = \begin{cases} \text{[x - a, 0]} & \text{si } x < -a \\ \text{[x-a, x+a]} & \text{si } -a < x < a \\ \text{[0, x+a]} & \text{si } x > a \end{cases}$

$I\!P(\mu \in C_a(x)/\mu) = I\!P(-a < \mu < 0/\mu = 0 \wedge x < -a)I\!P(x < -a) + I\!P(x - a < \mu < x + a/\mu = 0 \wedge -a < x < a)I\!P(-a < x < a) + I\!P(0 < \mu < x + a/\mu = 0 \wedge x > a)I\!P(x > a) = I\!P(x < -a) + I\!P(-a < x < a) + I\!P(x > a) = 1$

b) $I\!P(\mu \in C_a(x))/x) = \begin{cases} I\!P(\mu \in [x-a, 0]) & \text{si } x < -a \\ I\!P(\mu \in [x-a, x+a]) & \text{si } -a < x < a \\ I\!P(\mu \in [0, x+a]) & \text{si } x > a \end{cases}$

$$x \sim \mathcal{N}(\mu, 1) \Longrightarrow \begin{cases} I\!P(\mu \in [x-a, 0]) = I\!P(x - \mu < a) = 0.95 & \text{si } x < -a \\ I\!P(\mu \in [x-a, x+a]) = I\!P(\mid x - \mu \mid< a) = 0.95 & \text{si } -a < x < a \\ I\!P(\mu \in [0, x+a]) = I\!P(x - \mu > -a) = 0.95 & \text{si } x > a \end{cases}$$

c) Si la distribución a priori de $\mu = 1$, $\forall \mu$, la distribución a posteriori de μ es $\xi(\mu/x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{1}{2}(x-\mu)^2\} \Longrightarrow \mu \sim \mathcal{N}(x, 1)$.
d) Como $\mu \sim \mathcal{N}(x, 1) \Longrightarrow \mu - x \sim \mathcal{N}(0, 1)$

$$\mathbb{P}(\mu \in C_a(x))/x) = \begin{cases} \mathbb{P}(\mu \in [x-a, 0]) & \text{si } x < -a \\ \mathbb{P}(\mu \in [x-a, x+a]) & \text{si } -a < x < a \\ \mathbb{P}(\mu \in [0, x+a]) & \text{si } x > a \end{cases}$$

$$\mathbb{P}(\mu \in C_a(x))/x) = \begin{cases} \mathbb{P}(\mu - x \in [-a, -x]) & \text{si } x < -a \\ \mathbb{P}(\mu - x \in [-a, +a]) & \text{si } -a < x < a \\ \mathbb{P}(\mu - x \in [-x, +a]) & \text{si } x > a \end{cases}$$

$$\mathbb{P}(\mu \in C_a(x))/x) = \begin{cases} \Phi(-x) - \Phi(-a) & \text{si } x < -a \\ \Phi(a) - \Phi(-a) & \text{si } -a < x < a \\ \Phi(a) - \Phi(-x) & \text{si } x > a \end{cases}$$

e) Si $a = 1.65$, $\Phi(a) = 0.95$, $\Phi(-a) = 0.05$ y $\Phi(-x) > 0.95$ si $x < -a$, luego $\mathbb{P}(\mu \in C_a(x))/x) \begin{cases} > & 0.90 \quad \text{si } x < -a \text{o } x > a \\ = & 0.90 \quad \text{si } -a < x < a \end{cases}$

$\lim_{a\to\infty} \Phi(-a) = 0$, $\lim_{a\to\infty} \Phi(a) = 1$; $\Phi(-x) = 1$, $six < -a$ y $\Phi(x) = 1$, $six > a$. Luego $\lim_{a\to\infty} \mathbb{P}(\mu \in C_a(x)/x) = 1$.

CAPITULO 5

1. a) $X \to P(\lambda)$, donde X es el número de pasas en una empanada.
La distribución de probabilidad de una Poisson es $f_{x_i} = e^{-\lambda}\lambda^{x_i}/x_i!$.
Luego, la verosimiliutd está dada por $L(\underline{x}, \lambda) = e^{-nl} l^{\Sigma x_i} \prod(1/x_i!)$
Tomando el logaritmo de L y derivando con respecto al parámetro tenemos:

$$\frac{\partial L}{\partial l} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0 \Longrightarrow \hat{l} = \bar{x}$$

b) (i) Sea $p = \mathbb{P}$(una cierta pasa esté en una empanada). La probabilidad de tener x pasas en una empanada está dada por una distribución binomial pues definimos una variable aleatoria Y_i como valiendo 1 si hay una pasa, y cerro si no la hay; luego, en n empanadas hay $\sum_{i=1}^{n} Y_i$ empanadas, v.a. que sigue una $Binomial(n, p)$.
Esta distribución se puede aproximar por una Poisson (ver curso de Probabilidades).
(ii) Se puede probar la hipótesis nula: $H_0 : X \sim \mathcal{P}(l)$ con un test χ^2.
El estadístico utilizado es, por construcción: $Q = \sum_{i=1}^{k} \dfrac{(f_i - np_i)^2}{np_i}$, donde $n = 20$ (el tamaño de la muestra) y $k = 7$ (el número de clases en que están repartidos los datos).
Los grados de libertad de la χ^2 están dados por $k - 1 - 1$ pues hay un parámetro estimado.
Haciendo los cálculos se obtiene $Q = 0.449$. Con un nivel de significación del 5%, obtenemos la región crítica $Q \geq 11.070$. Como $0.449 < 11.070$, no se rechaza la hipótesis nula.

c) Para $H_0 = 3.5$ vs. $H_1 = 2.5$, se favorece al cocinero, pero para H_0' vs. H_1', se favorece a los alumnos. d) Para el test H_0, tenemos: $P(rechazar H_0/H_0 cierto) = 0.05$.
Del Lema de Neymann-Pearson tenemos que la región crítica está caracterizada por $\bar{x} < a$.
Por otro lado, $\sum x_i \sim \mathcal{P}(nl)$ donde $nl = 70$ bajo H_0.
Luego, $z = \dfrac{\sum x_i - E(\sum x_i)}{Var(\sum x_i)} \sim \mathcal{N}(0,1)$.
y $I\!P(recharzar H_0/H_0 cierta) = I\!P(z < \dfrac{a - nl}{\sqrt{nl}}/H_0) = 0.05$.
Luego, utilizando la tabla de la distribución Normal, se obtiene que $a = 56.195$ y la región crítica está dada por $\sum x_i < 56.195$. En la muestra $\sum x_i = 60$, por lo que no se rechaza H_0.
Calculemos $\beta = I\!P(recharzar H_0/H_1 cierta) = I\!P(\sum x_i < 56.195/l = 2.5) =$
$I\!P(z < \dfrac{56.195 - nl}{\sqrt{nl}}/l) = 0.5)$
Bajo H_1, $\sum x_i \sim \mathcal{P}(50)$; luego, $\beta = 0.81$.
e) Para H_0' procediendo de manera análoga se obtiene:
$0.05 = P(\sum x_i > a/l = 2.5) = P(z > \dfrac{a - 50}{\sqrt{50}}) \Longrightarrow a = 61.667)$
y la región crítica $\{\sum x_i > 61.667\}$.
Considerando la muestra concluimos que no se rechaza H_0'.
f) Se resumen los resultados anteriores en la tabla siguiente:

$\sum x_i$	$]-\infty, 56.195]$	$]56.195, 61.667]$	$]61.667, +\infty]$
$H_0 : \lambda = 3.5$	no son aceptables	son aceptables	son aceptables
$H_0' : \lambda = 2.5$	no son aceptables	no son aceptables	son aceeptables

Estamos ante una zona de contradicción: [56.195, 61.667]. Para suprimirla se propone dos maneras que se dejan propuestas:
(i) Fijar n y escoger un α que elimine la zona de contradicción.
(ii) Fijar un α y escoger un n

2. Se puede resolver de dos maneras:
a) Un test de Student:
Sea $D_i = p_i' - p_i$ las diferencias de pesos después y antes del matrimonio. Para responder, resolvemos la siguiente hipótesis: $H_0 : E(D) = 0$ vs. $H_1 : E(D) > 0$.
Suponiendo la normalidad tenemos: $\dfrac{\bar{D} - E(D)}{s/\sqrt{n-1}} \sim t_{n-1}$.
Luego, bajo H_0, se tiene $\dfrac{\bar{D}}{s/\sqrt{n-1}} \sim t_{n-1}$ con s^2 la varianza empírica.
$I\!P(t_9 > 1.83) = 0.05$. Luego, la región crítica estará dada por $\bar{D} > \dfrac{1.83s}{3}$.
Luego, de los datos se concluye:

- Para los hombres, $\bar{D} = 2.58$ y $s = 2.65$; luego, se rechaza H_0.
- Para las mujeres, $\bar{D} = -0.8$ y $s = 1.99$; luego, no se rechaza H_0.

Por tanto, hay evidencia para afirmar que los hombres suben de peso, las mujeres no.
b) Un test de Wilcoxon:
Sobre el rango de las diferencias positivas:

Hombres	$\lvert D_i \rvert$	5	0.5	4	3.5	2.5	2	3	1	1.5	6
	rango de $\lvert D_i \rvert$	9	1	8	7	5	4	6	2	3	19
	signo	$+$	$-$	$+$	$+$	$+$	$-$	$+$	$-$	$+$	$+$
Mujeres	$\lvert D_i \rvert$	2	1	3	1	2	2	2	3	3	1
	rango de $\lvert D_i \rvert$	5.5	2	9	2	5.5	5.5	5.5	9	9	3
	signo	$+$	$-$	$-$	$-$	$+$	$-$	$-$	$-$	$+$	$-$

Cuando hay empates, se da el rango promedio.
Luego:

$$W^+_{Hombres} = 48 \quad y \quad W^-_{Hombres} = 7 \qquad W^+_{Mujeres} = 20 \quad y \quad W^-_{Mujeres} = 35$$

donde $W^+ = \sum_{i=1}^n i\, \epsilon_i$, donde ϵ_i vale 1 si la diferencia es positiva, y 0 si no.
Para H_0, $I\!P(\epsilon_i^+) = \frac{1}{2}$ y ϵ_i^+ son independientes.
Luego, $E(W^+) = n(n+1)/4$ y $Var(W^+) = n(n+1)(2n+1)/24$. Por lo cual: $I\!P(\frac{W^+ - 27.5}{9.8} \geq 1.65) = 0.05$, de donde se obtiene como región crítica $W^+ \geq 43.7$.
Como $W^+_{Hombres} = 48$, se rechaza H_0. Para las mujeres no se puede usar esta aproximación ya que hay empates.

3. a). Sean $n_{i,j}$ las frecuencias de la tabla. Entonces: $n_{\bullet 1} = \sum_{i=1}^2 n_{i1} \sim B(n, p_1')$ y $n_{2\bullet} = \sum_{i=1}^2 n_{2j} \sim B(n, p_2)$.
Entonces se tiene la aproximación: $d = \dfrac{n_{2\bullet} - n_{\bullet 1}}{n} \sim \mathcal{N}(p_2 - p_1', \dfrac{p_2(1-p_2) + p_1'(1-p_1')}{n})$.
Estimando la varianza con $\hat{p_2} = \frac{n_{2\bullet}}{n}$ y $\hat{p_1'} = \frac{n_{\bullet 1}}{n}$, se obtiene bajo H_0 que $d \sim \mathcal{N}(0, 0.00244)$.
Luego, $I\!P(\frac{|d|}{0.049} \geq 1.96) = 0.05$ implica una región crítica de la forma $|d| \geq 0.096$.
En la muestra, $d = -0.05$, por lo que no se rechaza H_0.
Para H_0', se deja propuesto.
b) Se calcula el χ^2 de contingencia: $Q = \sum \dfrac{(n_{ij} - n_{i\bullet} n_{\bullet j}/n)^2}{n_{i\bullet} n_{\bullet j}/n} = 97.306$
Bajo la hipótesis de independencia, Q sigue una χ^2_1. Como $P(Q > 3.84) = 0.05$, se rechaza la independencia, lo que concuerda con los resultados de la parte anterior.

4. a) Si $0 < y < \phi$, entonces $P(Y_n) \leq y = (\frac{y}{\phi})^n$ pues la muestra es aleatoria, los X_i siguen una $U(0, \phi)$ y se calcula la distribución del máximo.

Análogamente, si $y \geq \phi$, $P(Y_n \leq y) = 1$ pues $Y_n \leq \phi$ por definición de los X_i.

Por lo tanto, $\pi(\phi) = \begin{cases} I\!P(Y_n \leq 1.5) = 1 & \text{si } \phi \leq 1.5 \\ I\!P(Y_n \leq 11.5) = (\frac{1.5}{\phi})^n & \text{si } \phi > 1.5 \end{cases}$

b) El tamaño del test es $\alpha = sup_{\phi\geq 2}(\frac{1.5}{\phi})^n = (\frac{1.5}{2})^n = (\frac{3}{4})^n$.

5. a) Para cualquier valor de p, $\pi(p) = P(Y \geq 7) + P(Y \leq 1)$ por propiedad de la probabilidad.
Para saber la distribución de Y hacemos lo siguiente:
Sea Y_i una variable que vale 1 si la pieza i es defectuosa, y 0 sino. Esta v.a. sigue una distribución de Bernoulli de parámetro $P(Y_i = 1) = I\!P$(la pieza i sea defectuosa)$=p$.
Luego, $Y = \sum_{i=1}^{20} Y_i \sim B(20, p)$.
Para $p = 0$, $I\!P(Y \geq 7) = 0$ y $I\!P(Y \leq 1) = 1$. Por lo tanto, $\pi(0) = 1$.
Para $p = 0.1$, de una tabla de distribución binomial, concluimos:
$P(Y \geq 7) = .0020 + .003 + .001 + .000 = .00214$ y $P(Y \leq 1) = .1216 + .2701 = .3917$.
Luego, $\pi(0.1) = .3941$
Similarmente, $\pi(0.2) = 0.1558$
Nota: Si la muestra fuese de mayor tamaño, se puede utilizar la aproximación normal: $\mathcal{N}(p, p(1-p))$. Se puede escoger la varianza máxima, esto es. cuando $p = \frac{1}{2}$.
b) Puesto que H_0 es una hipótesis simple, el tamaño del test α se obtiene evaluando la función de potencia en el punto especificado por $H_)$. Luego, $\alpha = \pi(0.2) = 0.1558$.

6. Como la hipótesis nula es simple, el tamaño del test está dado por $\alpha = P(rechazar H_0/\mu = \mu_0)$. Bajo H_0 la v.a. $Z = \sqrt{n}(\bar{X}_n - \mu_0) \sim \mathcal{N}(0,1)$. Luego, como el tamaño de la muestra es 25, $\alpha = I\!P(|\bar{X}_n - \mu_0| \geq 0) = P(|Z| \geq 5c) = 2[1 - \Phi(5c)] = 0.05$ en donde $\Phi(5c) = 0.975$, lo que implica que $c = 0.392$.

7. a) Las condiciones aquí son diferentes de las del Lema de Neyman-Pearson: en vez de fijar el valor de $\alpha(\delta)$ y minimizar $\beta(\delta)$, debemos fijar aquí el valor de $\beta(\delta)$ y minimizar $\alpha(\delta)$. Sin embargo, la misma demostración dada para el Lema de Neyman-Pearson muestra que el procedimiento óptimo es rechazar H_0 si $\dfrac{f_1(x)}{f_0(x)} > k$, donde k se escoge tal que $\beta(\delta) = I\!P(\text{No rechazar } H_0/H_1) = I\!P(\dfrac{f_1(x)}{f_0(x)} < k/H_1) = 0.05$.
Haciendo los cálculos, que son simples, se obtiene $\log \dfrac{f_1(x)}{f_0(x)} = \dfrac{3}{2} n\bar{x}_n - (constantes)$.
Luego, la razón de verosimilitud será más grande que k si y solo si $\bar{x}_n > k'$, con lo cual se ha hallado un procedimiento óptimo para rechazar H_0, escogiendo k' tal que: $I\!P(\bar{X}_n < k'/H_1) = 0.05$.
Para calcular k', notemos que si H_1 es cierta, entonces $Z = \sqrt{n}(\bar{x}_n - 5.0) \sim \mathcal{N}(0,1)$.
Luego, $I\!P(\bar{x}_n < k'/H_1) = \Phi[\sqrt{n}(k' - 5.0)] = 0.05$
en donde $k' = 5.0 - \dfrac{1.645}{\sqrt{n}}$.

b) Para $n = 4$, se rechaza H_0 si y solo si $\bar{x}_n > 4.1775$ (basta reemplazar $n = 4$ arriba). Por lo tanto, $\alpha(\delta) = \mathbb{P}(\text{recharzar } H_0/H_0 \text{ cierta}) = \mathbb{P}(\bar{x}_n > 4.1775/H_0)$.
Bajo H_0, $\bar{x}_n \sim \mathcal{N}(3.5, 0.25)$. Por lo tanto, $Z = 2(\bar{x}_n - 3.5) \sim \mathcal{N}(0,1)$. Luego, $\alpha(\delta) = P(Z > 2(4.1775 - 3.5)) = 0.0877$.

8. a) Bajo H_0, $X \sim U(0,1)$. Luego es imposible obtener un valor de X más grande que 1 bajo H_0; sin embargo, esto es posible bajo H_1. Por lo tanto, si una solución del test rechaza H_0 sólo cuando $x > 1$, $\alpha(\delta) = 0$ y $\beta(\delta) = P(X < 1/H_1) = \frac{1}{2}$.
b) Para tener $\alpha(\delta) = 0$, en la región crítica sólo podemos incluir los puntos tales que tienen probabilidad nula bajo H_0. Por tanto, sólo los puntos $x > 1$ deben ser considerados. Para minimizar $\beta(\delta)$ podríamos escoger un conjunto que maximiza la probabilidad bajo H_1. Los puntos $x > 1$ pueden ser usados en esta región crítica.

9. Aplicamos el teorema que minimiza la combinación lineal de los errores con $a = b = 1$. El procedimiento óptimo que rechaza H_0 cuando $\dfrac{f_1(x)}{f_0(x)} > 1$.
En este caso, $\log \dfrac{f_1}{f_0} = y \;\; \log(\dfrac{l_1}{l_0}) - n(l_1 - l_0)$.
Puesto que $l_1 > l_0$, $\dfrac{f_1}{f_0} > 1$ si y solo si $\bar{x}_n > \dfrac{l_1 - l_0}{\log l_1 - \log l_0}$.

10. La probabilidad de rechazar H_0 es 0.05. Por lo tanto, el valor de la función de potencia en todo valor de ϕ es 0.05.

11. Cambiamos el parámetro de ϕ a $\zeta = -\phi$. En términos de este nuevo parámetro, las hipótesis se reescriben: $H_0 : \zeta \leq -\phi_0$ vs $H_1 : \zeta > -\phi_0$.
Sea $g_n(x/\zeta) = f_n(x/-\zeta)$. Si $\zeta_1 < \zeta_2$, entonces $\phi_1 = -\zeta_1 > \phi_2 = -\zeta_2$. Por lo tanto, la razón $\dfrac{g_n(x/\zeta_2)}{g_n(x/\zeta_1)}$ será una función decreciente de $r(X)$. Se sigue que esta razón será una función creciente de $s(X) = -r(X)$. Así, en términos de ζ podemos aplicar el teorema correspondiente, y por lo tanto se rechaza H_0 cuando $s(X) \geq c'$ y éste procedimiento es UMP. Pero, $s(X) \geq c'$ si y solo si $r(X) \leq c$, donde $c = -c'$. Por lo tanto, el test que rechaza H_0 cuando $T \leq c$ es un test UMP. Si c es escogido como se afirma en el enunciado, entonces del mismo teorema se sigue que el tamaño del test es α_0.

12. Es fácil verificar que un test que rechaza H_0 cuando $\sum_{i=1}^n X_i \leq c$ será un test UMP. Cuando $n = 10$ y $l = 1$, $\sum_{i=1}^n X_i \sim Poisson(10)$ y $\alpha_0 = \mathbb{P}(\sum_{i=1}^n X_i \leq c/l = 1)$. De una tabla de Poisson se tiene:

c	0	1	2	3	4
α	0.0000	0.0005	0.0028	0.0104	0.0293

Para valores más grandes de c, $\alpha_0 > 0.03$.

13. H_0 es una hipótesis simple. Por el Lema de Neyman-Pearson, el test que maximiza la función de potencia en un valor particular $\phi_1 > 0$ será rechazado si $\dfrac{f(x/\phi = \phi_1)}{f(x/\phi = 0)} > c$, donde c es escogido tal que la probabilidad que esta desigualdad será satisfecha cuando $\phi = 0$ es α_0. Aquí, $\dfrac{f(x/\phi = \phi_1)}{f(x/\phi = 0)} > c$ siy solo si $(1-c)x^2 + 2c\phi_1 x > c\phi_1^2 - (1-c)$. Para cada valor de ϕ_1, el valor de c se escoge tal que el conjunto de los puntos que satisfacen esta desigualdad tienen probabilidad α_0 cuando $\phi = 0$. Para dos valores diferentes de ϕ_1, estos dos conjuntos serán diferentes. Por lo tanto, diferentes procedimientos de test maximizarán la potencia con dos valores diferentes de ϕ_1. Se conclue que no existe un test UMP.

14. El test UMP rechazará H_0 cuando $\bar{X}_n \geq c$, donde $I\!P(\bar{X}_n \geq c/\mu = 0) = I\!P(\sqrt{n}\bar{X}_n \geq \sqrt{n}c/\mu = 0) = 0.025$. Sin embargo, cuando $\mu = 0$, $Z = \sqrt{n}\bar{X}_n \sim \mathcal{N}(0,1)$. Por tanto, $I\!P(Z \geq 1.96/\mu = 0) = 0.025$, de donde $c = \dfrac{1.96}{\sqrt{n}}$.
Cuando $\mu = 0.5$, la v.a. $Z = \sqrt{n}(\bar{X}_n - 0.5) \sim \mathcal{N}(0.1)$. Por lo tanto, $\pi(0.5/\delta^*) = I\!P(\bar{X}_n \geq 1.96n^{-1/2}/\mu = 0.5) = I\!P(Z \geq 196 - 0.5n^{-1/2}) = I\!P(z \leq 0.5n^{-1/2} - 1.96)$ por la simetría de la normal.
Pero, $\Phi(1.282) = 0.9$. Por lo tanto, $\pi(0.5/\delta^*) \geq 0.9$ si y solo si $n \geq 42.042$. Así, una muestra de tamaño $n = 43$ es requerida. Puesto que la función de potencia es una función estrictamente creciente de μ, entonces se tendra $\pi(0.5/\delta^*) \geq 0.9$ para $\mu > 0.5$.
Cuando $\mu = -0.1$, la v.a. $Z = \sqrt{n}(\bar{X}_n + 0.1) \sim \mathcal{N}(0,1)$. Por lo tanto, $\pi(-0.1/\delta^*) = 1 - Phi(1.96 + 0.1n^{1/2}$. Pero como $\Phi(3.10) = 0.999$, $\pi(-0.1/\delta^*) \leq 0.001$ si y solo si $n \geq 129.96$. Así, una muestra de tamaño $n = 130$ se requiere para que $\pi(-0.1/\delta^*) \leq 0.001$. Como la función de potencia es estrictamente creciente como función de μ, entonces $\pi(\mu/\delta^*) \leq 0.001$ para $\mu < -0.1$.

15. Sea $\alpha_0 = 0.05$ y sea $c_1 = 3\alpha_0^{1/n}$ (Justifique la elección de esta constante). Sea $c_2 = 3$. Entonces $\pi(\phi/\delta) = I\!P(T \leq 3\alpha_0^{1/n}/\phi) + I\!P(T \geq 3/\phi)$.
Puesto que $I\!P(T \geq 3/\phi) = 0$ para $\phi \geq 3$, entonces: $\pi(\phi/\delta) = [\dfrac{3\alpha_0^{1/n}}{\phi}]^n + 1 - (\dfrac{3}{\phi})^n > \alpha_0$.

CAPITULO 7

1. a) Se observa un R^2 relativamente alto, R=cor(y,$\hat{y}$)=0.85. El F de Fisher a 3 y 196 g.l. muestra que se rechaza $H_o : E(y) = CONSTANTE$. Sin embargo, sólo el Atraso Medico tienen influencia significativa sobre el tiempo de espera, y el número de medicos en menor grado, como lo muestran las probabildades de las T de Student que tienen 196 g.l.

b)

$$F = \frac{R^2/3}{(1-R^2)/196}$$

$$\hat{\sigma}^2 = \sum \hat{\epsilon}_i^2/196$$

$$\hat{\sigma}^2 = \sum y_i^2(1-R^2)/196 = 64.286$$

Si se agrega una variable explicativa, el coeficiente de correlación multiple R^2 no puede disminuir, dado que el subespacio W se amplia.

c) $\dfrac{\hat{\beta}_j - \beta_j}{\sigma_{\beta_j}} \sim t_{196} \Longrightarrow I\!\!P(|t_{196}| < 1.96) = 0.95$

Luego, el intervalo de confianza simetrico a 95% para el coeficiente del ATRASO MEDICO es: $[0.09 - 1.96 \times 0.01, 0.09 + 1.96 \times 0.01] = [0.0704, 0.1096]$

d) La predicción del tiempo de espera para este nuevo paciente es:

$$\hat{y}_o = 22.00 + 0.09 \times 100 - 0.02 \times 0 - 1.61 \times 4 = 24.56$$

El intervalo de confianza a 95% es igual a: $[\hat{y}_o - 1.96\hat{\sigma}_o, \hat{y}_o + 1.96\hat{\sigma}_o]$,
con la varianza del error de predicción igual a: $\hat{\sigma}_o^2 = \hat{\sigma}^2(1 + x_o^t(X^tX)^{-1}x_o)$

2. a) El estimador de M.V. de β es: $\hat{\beta} = (X^tX)^{-1}X^tY$
Como aquí X^tX es de la forma diagonal por bloque:

$$\hat{\beta} = (X_1^tX_1)^{-1}X_1^tY + (X_2^tX_2)^{-1}X_2^tY$$

Pero $(X_1^tX_1)^{-1}X_1^tY$ es el estimador de M.V. del modelo $Y = X_1\alpha_1 + \epsilon$ y $(X_2^tX_2)^{-1}X_2^tY$ es el estimador de M.V. del modelo $Y = X_2\alpha_2 + \epsilon$.
b) En el caso de $X_1^tX_2 \neq 0$, el estimador $\hat{\beta}$ para β es sesgado.

3. a) El coeficiente de correlación multiple no es muy elevado, sin embargo los coeficientes son significativos individual y globalmente (se rechaza las hipotesis $H_o : \alpha = 0$ y $H_o : \beta = 0$ y $H_o : \alpha = \beta = 0$).
b) $R^2 = \frac{Var(\hat{y})}{Var(y)}$; luego $\sum_i e_i^2 = nVar(y)(1-R^2)$.
$\hat{\sigma}^2 = \frac{\sum_i e_i^2}{n-2} = \frac{nVar(y)(1-R^2)}{n-2} = 7.687$
c) Se denota una tendencia creciente de los residuos, lo que indica un sesgo en la estimación: falta algunas variables explicativas en el modelo.
d) La predicción para y_o se hace a traves del modelo:
$\hat{y}_o = 7.05 + 0.79 * 6.50 = 12.185$
$Var(\hat{y}_o) = \frac{\sum x_i^2 - 2x_o\sum x_i + nx_o^2}{n\sum x_i^2 - (\sum x_i)^2} = 0.0264.$
e) Como $\sum_i x_iz_i = 0$ y $\sum z_i = 0$, se tiene que $X_1^tX_2 = 0$. Luego del problema 1 se deduce que $\hat{\beta}_0 = 7.05$ $\hat{\beta}_1 = 0.79$ y $\hat{\beta}_2 = 1.00$.

4. El modelo a ajustar es tal que: $Y_i = \beta_j + \epsilon_i \quad$ si $t_i \in I_j$.
Sean las variables indicadoras z_j definidas sobre los intervalos I_j por:

$$z_{ji} = \begin{cases} 1 & \text{si } t_i \in I_j \\ 0 & \text{si } t_i \notin I_j \end{cases}$$

El modelo lineal se escribe entonces: $Y_i = \sum_{j=1}^{k} \beta_j z_{ji} + e_i$.
b) El criterio de los minimos cuadrados se escribe:
$Q = \sum_i (Y_i - f(t_i))^2 = \sum_j \sum_{i \in I_j} (Y_i - \beta_j)^2 \Longrightarrow \frac{dQ}{d\beta_j} = 2\sum_{i \in I_j}(Y_i - \beta_j) = 0$
Se obtiene $\hat{f(t)} = \hat{\beta}_j = \sum_{i \in I_j} Y_i/n = \bar{Y}_j$ si $t \in I_j$.
c) Como $\bar{Y}_j \sim \mathcal{N}(\beta_j, \sigma^2/n_j)$, se obtiene un intervalo de confianza para β_j a 95% con: $[\bar{Y}_j - 1.96\sigma/\sqrt{n_j}, \bar{Y}_j + 1.96\sigma/\sqrt{n_j}]$.
Como $\int_{a_0}^{a_K} \hat{f(t)}dt = \sum_1^K (a_j - a_{j-1})\bar{Y}_j$ y los $\bar{Y}_j$ son independientes, $\sum_1^K (a_j - a_{j-1})\bar{Y}_j \sim \mathcal{N}(\int_{a_0}^{a_K} f(t)dt, \sigma^2 \sum_1^K \frac{(a_j - a_{j-1})^2}{n_j})$; se obtiene entonces un intervalo de confianza a 95% para $\int_{a_0}^{a_K} f(t)dt$ con: $[\sum_1^K (a_j - a_{j-1})\bar{Y}_j - 1.96\sigma\sqrt{\sum \frac{(a_j - a_{j-1})^2}{n_j}}), \sum_1^K (a_j - a_{j-1})\bar{Y}_j + 1.96\sigma\sqrt{\sum \frac{(a_j - a_{j-1})^2}{n_j}})]$

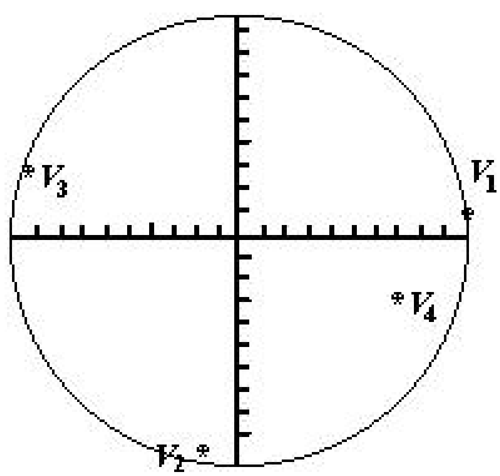
5. Como $W = H_a \oplus \Delta_a$ con $H_a \perp \Delta_a$, $P_W(Y) = P_{H_a}(Y) + P_{\Delta_a}(Y)$. Además H_a pertenece a W, luego $P_{\Delta_a}(Y) = P_{\Delta_a}(\hat{Y}) = (\frac{a^t\hat{Y}}{a^t a})a$. Luego, $\hat{Y} = Y^* - (\frac{a^t\hat{Y}}{a^t a})a$.
b) Como $H_a \perp \Delta_a$, $Var(b^t\hat{y}) = Var(b^t y^*) + Var((\frac{a^t\hat{y}}{a^t a})b^t a)$.
Pero $Var((\frac{a^t\hat{y}}{a^t a})b^t a) = (\frac{b^t a}{a^t a})^2 Var(a^t\hat{y})$.
Por otro lado $Var(\hat{y}) = \sigma^2 P_W$, y como $a \in W$, $P_W(a) = a$.
Luego, $Var(b^t\hat{y}) = Var(b^t y^*) + \sigma^2 \frac{(b^t a)^2}{a^t a}$.
c) $\sqrt{\sum \epsilon_i^{*2}}$ es la norma de la proyección de un vector normal $\mathcal{N}_n(0, \sigma^2 I_n)$ sobre el subespacio ortogonal a H_a en $\mathbb{R}^n$, luego $\frac{\sum_{i=1}^n \epsilon_1^{*2}}{\sigma^2} \sim \chi^2_{n-p}$, siendo n-p la dimensión de $H_a^\perp$.
d) La distribución de $\frac{\sum y_i^{*2}/p}{\sum \epsilon_i^{*2}/(n-p)}$ es una F de Fisher a p y n-p g.l., dado que Y^* pertenece a $H_{\mathbf{1}_n}$ que tiene dimensión p y los residuos al ortogonal de $H_{\mathbf{1}_n}$, que tiene dimensión n-p.
Si las variables son centradas, W es ortogonal al vector $\mathbf{1}_n$ luego $\hat{Y}$ lo es también; si $a = \mathbf{1}_n$ entonces $W = H_{\mathbf{1}_n}$, e $\hat{Y} = Y^*$.

CAPITULO 8

1. a) La suma de cada columna de X es nula. $V = \frac{1}{6}\begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}$

b) $I_0 = tr(VM) = tr(V) = 2$
c) $det(V - lI) = 0 = -l(l-1)^2$, de donde $l_3 = 0$ y $l_1 = l_2 = 1$
d) Un vector $\underline{u}$ asociado a l_3 está dado por $\underline{u} = (1, 1, 1)^t$.
e) El subespacio propio asociado a $l = 1$ es de dimensión 2 (¿Por qué?).
Una base ortonormal esta dada por $v_1 = (1/\sqrt{2}, 0, -1/\sqrt{2})^t$ y $v_2 = (1/\sqrt{6}, -2/\sqrt{6}, 1/\sqrt{6})^t$.

2. a) $V_1 + V_2 + V_3 = 100$, luego la representación del conjunto de puntos es plana.
En el gráfico se ve una cierta tendancia a separarse las tres clases.
b) Las coordenadas de las variables estan dadas por las correlaciones de la tabla 8.9.
Las variables se encuentran en la circunferencia del círculo.
La primera componente principal depende más bien de V_1 y V_3 con V_1 y V_3 opuestos.
La segunda componente depende fuertamente de V_2.



c) Las coordenadas de V_4 son dadas por sus correlaciones con las componentes principales (0.69, -0.29).
d) La matriz X es de rango incompleto, luego la matriz X^tX no es invertible.
e) $R^2 = (0.69)^2 + (-0.29)^2 = 0.5$
f) $V_4 = \beta_o + \beta_1 c_1 + \beta_2 c_2$
$\hat{\beta}_o$ es igual a la media de V_4, es decir $\hat{\beta}_o = 242.5$
$\hat{\beta}_1$ es la proyección de V_4 sobre C_1, es decir $\hat{\beta}_1 = \frac{<V_4, C_1>}{|||C_1||} = \sigma_{V_4} cor(V_4, C_1) = 0.69 \times 57.73 = 39.86$

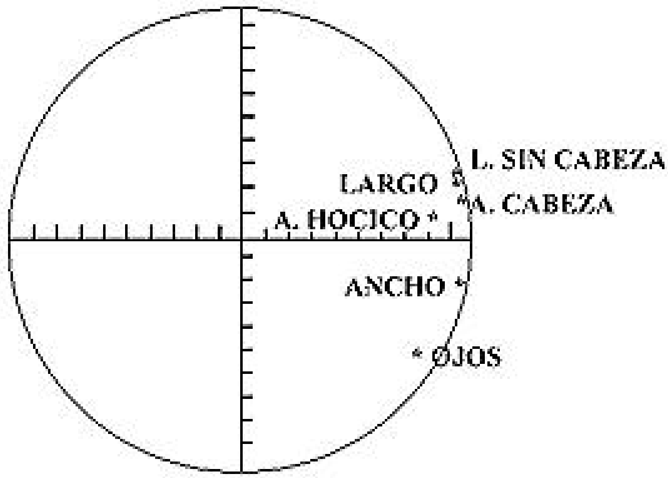
$\hat{\beta}_2 = \frac{<V_4,C_2>}{|||C_2||} = \sigma_{V_4} cor(V_4, C_2) = -0.29 \times 57.73 = -16.92$

3. a) $V_i = P_i V P_i^t$, con $i = 1, 2$ pues se proyectan las columnas de V, esto es, $P_i X$, con $i = 1, 2$. b) $W_1 \bot_{V^{-1}} W_2 \Longleftrightarrow (V^{-1}P_i)^t = V^{-1}P_i$ para $i = 1, 2$
$\Longleftrightarrow VP_i^t = P_i V$, $i = 1, 2$
$\Longleftrightarrow P_i V P_i^t = P_i V$, $i = 1, 2$
$\Longleftrightarrow V_1 + V_2 = P_1 V P_1^t + P_2 V P_2^t = P_1 V + P_2 V = P_1 V + (I - P_1)V = V$
pues $P_2 = I - P_1$ ya que $I\!R = W_1 \bigoplus W_2$. c) $V\underline{u} = l\underline{u}$. Luego, si $W_1 \bot_{V^{-1}} W_2$, entonces $P_i V \underline{u} = l P_i u$ con $i = 1, 2$ de donde $V_i \underline{u} = l P_i \underline{u}$ con $i = 1, 2$, de donde necesariamente $\underline{u} \in W_1 \bigcup W_2$.

4. a) $S : E \rightarrow F$. Luego, usando la propiedad de S,
$[S(\underline{x}_1 - \underline{x}_2)]^t N [S(\underline{x}_1 - \underline{x}_2] = (\underline{x}_1 - \underline{x}_2)^t M (\underline{x}_1 - \underline{x}_2)$ para cualquieras $\underline{x}_1, \underline{x}_2 \in E$, de donde $S^t S = I$.
b) Notemos que $S\underline{x} = -A\underline{x} + (I - A)\underline{x} = I\underline{x} - 2A\underline{x}$, de donde $S = I - 2A$. Además, $Ker(S) = \{\underline{x} : (I - 2A)\underline{x} = 0\} = \{\underline{x} : A\underline{x} = \frac{1}{2}\underline{x}\}$.
Como A es proyector, sus valores propios son 0 y 1. Entonces, $KerS = \{\underline{0}\}$, lo que implica que S es invertible.
c) S es simétrica dado que I y A lo son. Luego, $SS^t = S^2 = (I - 2A)(I - 2A) = I$, entonces S es ortogonal.
d) $S\underline{x} = \underline{x} \Longleftrightarrow \underline{x} \in E_2$. Como hay simetría de los puntos de $calM$, $SVS = V$. Luego, si $\underline{u} \in E_2$, $V\underline{u} = SVS\underline{u} = SV\underline{u}$. Como $S(V\underline{u}) = V\underline{u}$, entonces $V\underline{u} \in E_2$.

5. a) $I_0 = \sum_i m_i \underline{x}_i' M \underline{x}_i$
b) $I_0 = tr(I_0) = tr(\sum_i m_i \underline{x}_i' M \underline{x}_i) = tr(\sum_i m_i \underline{x}_i \underline{x}_i' M) = tr(VM)$.
c) $I_H = \sum_i m_i d^2(\underline{x}_i, H) = \sum_i m_i \|\underline{a}_i\|^2 = \sum_i m_i \underline{u}^t M \underline{x}_i \underline{x}_i' M \underline{u} = \underline{u}^t M [\sum_i m_i \underline{x}_i \underline{x}_i'] M \underline{u} = \underline{u}^t MVM\underline{u}$.

6. **PARTE 1**
a) La primera componente principal tiene una varianza mucho mayor que las otras (82% la primera en comparación de 8% la segunda en la tabla 8.11).
b) El pez 1 es grande en general; el 5 y el 8 pequeños, pero el 8 tiene ojos chicos. El pez 11 es de tamaño mediano con ojos grandes.
c) Las coordenadas sobre la primera componente principal C_1 son los coeficientes de correlación de las variables con C_1:



Se observa que todas las variables son altamente correlacionadas positivamente con la primera componente principal.

d) Las coordenadas de C_1 y C_2 se obtienen de las correlaciones anteriores: para el ANCHO es $Cor(ANCHO, C_1)/\sqrt{\lambda_1}) = 0.417$.

C_1 = 0.428 * LARGO + 0.425 * L. SIN CABEZA + 0.433 * A. CABEZA + 0.417 * ANCHO + 0.37 * A. HOCICO + 0.37 * OJOS

C_2= 0.322 * LARGO + 0.376 * L. SIN CABEZA + 0.189 * A. CABEZA - 0.308 * ANCHO + 0.103 * A. HOCICO - 0.783 * OJOS

Se observa que en C_1 las 6 variables tienen casi el mismo peso, pero en C_2 los pesos son diferentes y hay dos variables con signo negativo. Se puede decir que C_1 es un factor tamaño del pez.

e) Para el PESO: 0.98 sobre C_1 y 0.0 sobre C_2, para la RADIOACTIVIDAD: -0.41 sobre C_1 y 0.23 sobre C_2

f) $R^2 = 0.98^2 + 0 = 0.96$

PARTE 2

a) $X = \begin{pmatrix} 1 & C_{11} & C_{21} \\ 1 & C_{12} & C_{22} \\ . & .. & .. \\ 1 & C_{123} & C_{223} \end{pmatrix}$

Observando que las componentes principales están centradas y de varianza iguales a los valores propios se obtiene:

$X^tX = 23\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4.88 & 0 \\ 0 & 0 & 0.49 \end{pmatrix} \qquad (X^tX)^{-1} = 1/23\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.205 & 0 \\ 0 & 0 & 2.041 \end{pmatrix}$

$$\text{b) } X^tY = \begin{pmatrix} \sum Y_i \\ \sum C_{1i}Y_i \\ \sum C_{2i}Y_i \end{pmatrix} = n\begin{pmatrix} \bar{Y} \\ cov(C_1, Y) \\ cov(C_2, Y) \end{pmatrix} = 23\begin{pmatrix} 27.22 \\ -14.91 \\ 2.65 \end{pmatrix}$$

En efecto $\sum C_{1i}Y_i$ y $\sum C_{2i}Y_i$ son las covarianzas dado que C_1 y C_2 son centradas.
c) $\hat{\beta}_o = \bar{Y} = 27.22$, $\hat{\beta}_1 = -3.057$ y $\hat{\beta}_2 = 5.409$
d) $R^2 = (-0.41^2) + 0.23^2 = 0.221$
$\hat{\sigma}^2 = \sum \hat{\epsilon}_i^2/20 = 23var(Y)(1 - R^2)/20 = 23 * (16.47)^2(1 - 0.221)/20 = 243$
$var(\hat{\beta}_o) = 243 * 1.0 = 243$, $var(\hat{\beta}_1) = 243 * 0.205 = 49.81$, $var(\hat{\beta}_2) = 243 * 2.041 = 496$
e) La matriz $(X^tX)^{-1})$ muestra que $\hat{\beta}_1$ y $\hat{\beta}_2$ tienen un coeficiente de correlación nulo.
$\hat{\beta}_1 = -3.057$ y $\hat{\beta}_2 = 5.409$, luego se obtiene los intervalos:

$$\beta_1 \in [-3.057 - 2\sqrt{49.81}, -3.057 + 2\sqrt{49.81}] = [-17.17, 11.06]$$

$$\beta_2 \in [5.409 - 2\sqrt{496}, 5.409 + 2\sqrt{496}] = [-16.86, 27.68]$$

Lo que muestre que los coeficientes no son muy significativos.
f) Bajo H_o, $T = \hat{\beta}_1/\sqrt{49,81} \sim t_{20}$. Como la probabilidad de encontar un valor de T mayor que 0.433 y menor que -0.433 es mayor que 0.05, no se puede rechazar H_o.

PARTE 3

a) $\sum \epsilon_i^2 = \sum_{i\in AC=1}(y_i - \beta_o - \beta_1)^2 + \sum_{i\in AC=2}(y_i - \beta_o - \beta_2)^2 + \sum_{i\in AC=3}(y_i - \beta_o - \beta_3)^2$
$\frac{dQ}{\beta_1} = -2\sum_{i\in AC=1}(y_i - \beta_o - \beta_1) = 0$
$\frac{dQ}{\beta_2} = -2\sum_{i\in AC=2}(y_i - \beta_o - \beta_2) = 0$
$\frac{dQ}{\beta_3} = -2\sum_{i\in AC=3}(y_i - \beta_o - \beta_3) = 0$
$\hat{\beta}_1 = \bar{y}_1 - \beta_o = \bar{y}_1 - \bar{y} = 15.25 - 27.22 = -11.97$
$\hat{\beta}_2 = \bar{y}_2 - \beta_o = \bar{y}_2 - \bar{y} = 33.50 - 27.22 = 6.28$
$\hat{\beta}_3 = \bar{y}_3 - \beta_o = \bar{y}_3 - \bar{y} = 33.71 - 27.22 = 6.49$
b) Se tiene $H_o : \beta_3 - \beta_2 = 0$.
Ahora bien bajo el supuesto de normalidad y de independencia de las observaciones $\bar{y}_3 - \bar{y}_2 = \hat{\beta}_3 - \hat{\beta}_2 \sim \mathcal{N}(\beta_3 - \beta_2, \tau^2)$ en donde $\tau^2 = \frac{\sigma_2^2}{n_2} + \frac{\sigma_3^2}{n_3}$, con σ_j^2 y n_j la varianza

y el efectivo en el acuario j. Si se hace el supuesto de igualdad de las dos varianzas, $\sigma_2^2 = \sigma_3^2 = \sigma^2$, un estimador insesgado de σ^2 es dado por $S^2 = (n_2 S_2^2 + n_3 S_3^2)/(n_2 + n_3 - 2) = 346.42$ en que $S_2^2 = 147.14$ y $S_3^2 = 470.46$ son las varianzas empiricas sesgadas de σ_2^2 y σ_3^2. Entonces

$$t = \frac{\bar{x}_3 - \bar{x}_2}{\sqrt{(\frac{n_3 S_3^2 + n_2 S_2^2}{n_2 + n_3 - 2})(\frac{n_2 + n_3}{n_2 n_3})}}$$

sigue una distribución de t de Student a $n_2 + n_3 - 2 = 13$ grados de libertad.
La región critica se define entonces como:

$$I\!P(|\bar{x}_3 - \bar{x}_2| > t_\alpha \sqrt{(\frac{n_2 S_2^2 + n_3 S_3^2}{n_2 + n_3 - 2})(\frac{n_2 + n_3}{n_2 n_3})})$$

en donde t_α es tal que

$$I\!P(|t_{n_2+n_3-2|} > t_\alpha) = \alpha$$

Aquí se hizo el supuesto de igualdad de las varianzas y de independencia de las observaciones. En la muestra se obtuvo $t = \frac{0.21}{18.61 \times 0.268} = 0.042$, por lo cual no se rechaza la hipótesis H_o de igualdad de los parametros β_2 y β_3.
c) C_1= 0.428 * LARGO + 0.425 * L. SIN CABEZA + 0.433 * A. CABEZA + 0.417 * ANCHO + 0.37 * A. HOCICO + 0.37 * OJOS
C_2= 0.322 * LARGO + 0.376 * L. SIN CABEZA + 0.189 * A. CABEZA - 0.308 * ANCHO + 0.103 * A. HOCICO - 0.783 * OJOS
Ojo, hay que corregir los valores debido al centramiento y reducción de las 6 variables iniciales:
El nuevo pez toma los valores: LARGO: (180-190.17)/17.99=-0.565, LARGO SIN CABEZA: (152-170.70)/15.69=-1.192, ANCHO CABEZA: (40-42.78)/4.8=-0.579, ANCHO: (38-39.3)/4.57=-0.284, ANCHO HOCICO: (15-13.57)/2.54=0.563, DIAMETRO OJOS: (12-9.74)/0.96=2.354.
$C_1 = 0.428 * -0.565 + ... + 0.37 * 2.354 = -0.038$
$C_2 = 0.322 * -0.565 + ... - 0.783 * 2.354 = -2.437$
Su RADIOACTIVIDAD puede estimar a partir del modelo definido en 2.4: 27.22 + 3.057 * 0.038 - 5.409 * 2.437 = 14.15
d) $\xi(\mu/x_1, ..., x_23) \propto \theta exp(-\theta\mu)\mu^{23} exp(-23 \times 27.22\mu)$, que corresponde a una Gamma(24, 626+θ).
e) Para función de pérdida cuadrática, el estimador de Bayes de μ es la esperanza de la destribución a posteriori. Aquí $\hat{\mu} = \frac{24}{626+\theta}$.

PANORAMA DE LAS TECNICAS MULTIDIMENSIONALES

OBJETIVOS	METODOS
DESCRIPCION	Análisis en Componentes Principales Análisis de Correspondencias Múltiples Clasificación Automática Métodos de Escalonamiento multidimensional
ASOCIACION ROLES SIMETRICOS DE LAS VARIABLES	Análisis en Componentes Principales Análisis de Correspondencias Múltiples Modelo Log-lineal Análisis Canónico
ASOCIACION VARIABLES A EXPLICAR CUALITATIVAS	Análisis Discriminante Modelo Logit Segmentación por Arbol de Decisión Anánlisis de Correspondencias sobre Tablas Juxtapuestas
ASOCIACION VARIABLES A EXPLI-CAR CUANTITATIVAS	Regresión Lineal Anova, Manova Segmentación por Arbol de Decisión

BIBLIOGRAFIA

BENJAMIN, Probabilidad y Estadística en Ingeniería Civil, Mc Graw Hill Latinoamericana (1981).

CRAMER, H., Mathematical Methods of Statistics, Princeton University Press, 1961 (existe una versión en Castellano).

DEGROOT, M., PROBABILIDAD y ESTADISTICA, Addison-Wesley, 1987.

GILBERT, Estadítica, Interamericana (1981).

KENDALL, M.G., STAUART, A., The Advanced Theory of Statistics, Lossey-Bass (1966).

LEBART L. et al., Traitement des Données Statistiques, Dunod 1979.

MILLER G., Beyond ANOVA, Basics of Applied Statistics, Wiley 1986.

MOOD, GRAYBILL, Introducción a la Estadistica Matemática, Aguilar.

MOSTELLER, TUKEY, Data Analysis and Regression, Addison-Wesley.

ZUWAYLIF F., Estadistica General Aplicada, Fundo Educativo Interamericano 1977.