Tarea 1 Modelos Lineales

Pablo Borrego Ramos

Marzo 2024

1

1.1 Enunciado

Supongamos que $\mathbf{Y}$ se ajusta a un modelo lineal básico de rango completo, es decir, $E[\mathbf{Y}] = \mathbf{X}\beta$ y $Cov[\mathbf{Y}] = \sigma^2 I_n$ siendo $\mathbf{X}$ una matriz de orden $n \times p$ con $p < n$ y $r(\mathbf{X}) = p$. Supongamos que escribimos $\mathbf{X}\beta = \mathbf{X}_1\delta_1 + \mathbf{X}_2\delta_2$ donde $\delta_1 \in \mathbb{R}^r$ y $\delta_2 \in \mathbb{R}^{p-r}$, $\mathbf{X}_1$ es una matriz $n \times r$ y $\mathbf{X}_2$ es una matriz $n \times (p-r)$.
Probar que $\hat{\delta_1} = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}$ y $\hat{\delta_2} = (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}$ son estimadores insesgados para δ_1 y δ_2 respectivamente si y solo si $\mathbf{X}_1^t\mathbf{X}_2 = 0$, en cuyo caso $Cov[\hat{\delta_1}, \hat{\delta_2}] = 0$

1.2 Respuesta

$\Rightarrow$
Si $E[\hat{\delta_1}] = \delta_1$ entonces $(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t E[\mathbf{Y}] = \delta_1$ simplificando la parte izquierda de la igualdad llegamos a que $(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}\beta = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t(\mathbf{X}_1\delta_1+\mathbf{X}_2\delta_2) = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_1)\delta_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_2)\delta_2 = \mathbf{I}_r\delta_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_2)\delta_2 = \delta_1$ realizando un procedimiento similar con la igualdad $E[\hat{\delta_2}] = \delta_2$ obtenemos $\mathbf{I}_{p-r}\delta_2 + (\mathbf{X}_2^t\mathbf{X}_2)^{-1}(\mathbf{X}_2^t\mathbf{X}_1)\delta_1 = \delta_2$ llegando a un sistema de ecuaciones de la forma:

$$(\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_2)\delta_1 = 0$$

$$(\mathbf{X}_2^t\mathbf{X}_2)^{-1}(\mathbf{X}_2^t\mathbf{X}_1)\delta_2 = 0$$

Como $r(\mathbf{X}) = p$ entonces $\mathbf{X}_1$ o $\mathbf{X}_2$ debe de ser de rango máximo, y por lo tanto $(\mathbf{X}_1^t\mathbf{X}_1)^{-1}$ o $(\mathbf{X}_2^t\mathbf{X}_2)^{-1}$ es no nulo, por lo tanto para que se cumpla el sistema de ecuaciones debe de ocurrir que $(\mathbf{X}_1^t\mathbf{X}_2) = 0$ o $(\mathbf{X}_2^t\mathbf{X}_1) = 0$ en el primer caso ya habríamos terminado, en el segundo si $(\mathbf{X}_2^t\mathbf{X}_1) = 0$ entonces $(\mathbf{X}_2^t\mathbf{X}_1)^t = (\mathbf{X}_1^t\mathbf{X}_2) = 0$. Ahora veamos que $Cov[\hat{\delta_1}, \hat{\delta_2}] = 0$:

$Cov[\hat{\delta_1}, \hat{\delta_2}] = Cov[(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}, (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}] = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t Cov[\mathbf{X}, \mathbf{X}]\mathbf{X}_2((\mathbf{X}_2^t\mathbf{X}_2)^{-1})^t =$

$(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\sigma^2\mathbf{I}_n\mathbf{X}_2((\mathbf{X}_2^t\mathbf{X}_2)^{-1})^t = \sigma^2(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t 0_n\mathbf{X}_2((\mathbf{X}_2^t\mathbf{X}_2)^{-1})^t = 0$ donde la penúltima igualdad se debe a que $(\mathbf{X}_1^t\mathbf{X}_2) = 0$

$\Leftarrow$

Si $(\mathbf{X}_1^t\mathbf{X}_2) = 0$ veamos que $E[\hat{\delta_1}] = \delta_1$:
$E[\hat{\delta_1}] = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}E[\mathbf{Y}] = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}\beta = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t(\mathbf{X}_1\delta_1+\mathbf{X}_2\delta_2) = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_1)\delta_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_2)\delta_2 = \mathbf{I}_r\delta_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}(\mathbf{X}_1^t\mathbf{X}_2)\delta_2 = \mathbf{I}_r\delta_1 + 0_{p-r}\delta_2 = \delta_1$ donde la penúltima desigualdad se debe a que estamos suponiendo que $(\mathbf{X}_1^t\mathbf{X}_2) = 0$

2

2.1 Enunciado

Supongamos que $\mathbf{Y}$ se ajusta a un modelo lineal básico de rango completo, es decir, $E[\mathbf{Y}] = \mathbf{X}\beta$ y $Cov[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ siendo $\mathbf{X}$ una matriz de orden $n \times p$ con $p < n$ y $r(\mathbf{X}) = p$. Denotaremos por $\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$ al estimador de mínimos cuadrados de β. Si $\psi = \lambda^t\beta, \lambda \in \mathbb{R}$ es una función lineal de δ, consideramos el estimador $\hat{\psi} = \lambda^t\hat{\beta}$.

a) Probar que $\hat{\psi}$ es un estimador insesgado para ψ y calcular su varianza

b) Demostrar que $\hat{\psi}$ es el estimador insesgado de mínima varianza de ψ, en el sentido que si $T = c^t\mathbf{X}, c \in \mathbb{R}^n$ es otro estimador lineal insesgado de ψ, entonces $Var[\hat{\psi}] \leq Var[T]$ y se da la igualdad si y solo si $T = \hat{\psi}$

2.2 Respuesta

a)

1. $E[\hat{\psi}] = E[\lambda^t\hat{\beta}] = \lambda^t E[\hat{\beta}] = \lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t E[\mathbf{Y}] = \lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\beta = \lambda^t\beta = \psi$
2. $Var[\hat{\psi}] = \sum_{i=1}^n Var[\hat{\psi}_i] = tr(Cov[\hat{\psi}]) = tr(Cov[\lambda^t\hat{\beta}]) = tr(\lambda^t Cov[\hat{\beta}]\lambda) = tr(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t Cov[\mathbf{Y}]((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)^t\lambda) = tr(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\sigma^2\mathbf{I}_n((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)^t\lambda) = tr(\lambda^t\sigma^2(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}((\mathbf{X}^t\mathbf{X})^{-1})^t\lambda) = \sigma^2 tr(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda)$

b)

Sea $T = c^t\mathbf{Y}$ podemos reescribir c de la forma $c = (\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t + b)^t$ entonces para que T sea insesgado debe de ocurrir que $\psi = E[T] = E[(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t + b)\mathbf{Y}] = E[\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}] + E[b\mathbf{Y}] = E[\hat{\psi}] + b\mathbf{X}\beta = \psi + b\mathbf{X}\beta$ por lo tanto para

que T sea insesgado debe de ocurrir que $b\mathbf{X} = 0$. Teniendo esto en cuenta podemos calcular la matriz de Covarianzas del siguiente modo: $Cov[T] = c \cdot Cov[Y] \cdot c^t = \sigma^2 \cdot c \cdot c^t = \sigma^2(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t + b)(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t + b)^t = \sigma^2(\lambda(\mathbf{X}^t\mathbf{X})^{-1}\lambda^t + \lambda^t(\mathbf{X}^t\mathbf{X})^{-1}(b\mathbf{X})^t + b\mathbf{X}\lambda^t(\mathbf{X}^t\mathbf{X})^{-1} + b \cdot b^t) = \sigma^2(\lambda^t(\mathbf{X}^t\mathbf{X})^{-1} + b \cdot b^t)$.
De este modo si $b = (b_1, ...b_n)$ tenemos que: $ECM(T) = tr(Cov[T]) = \sigma^2 tr(\lambda(\mathbf{X}^t\mathbf{X})^{-1}\lambda^t) + \sigma^2 tr(b \cdot b^t) = ECM(\hat{\psi}) + \sigma^2 \sum_{i=1}^n b_i^2$. Concluimos que $ECM(T) \geq ECM(\hat{\psi})$ ($ECM(\hat{\psi}) = Var[\hat{\psi}] = \sigma^2 tr(\lambda(\mathbf{X}^t\mathbf{X})^{-1}\lambda^t)$ se debe a que ψ es insesgado y a el apartado a) de este ejercicio) y se dará la igualdad si y solo si $\sum_{i=1}^n b_i^2 = 0$, es decir, $b = 0$

3

3.1 Enunciado

Supongamos que $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ donde $\mathbf{X}$ es una matriz de orden $n \times p$ ($p < n$) y $r(\mathbf{X}) = p$, es decir, $\mathbf{Y}$ se ajusta a un modelo lineal normal de rango completo. Denotaremos por $\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$, al estimador de mínimos cuadrados y de máxima verosimilitud de β. Si $\psi = \lambda^t\beta$, $\lambda \in \mathbb{R}^p$ es una función lineal de β, consideramos el estimador $\hat{\psi} = \lambda^t\hat{\beta}$.

a) Probar que $\hat{\psi} \sim \mathcal{N}(\psi, \sigma^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda)$

b) Deducir un intervalo de confianza para ψ al nivel de confianza $1 - \alpha$ ($0 < \alpha < 1$).

3.2 Respuesta

a) Primero estudiemos la distribución que sigue $\hat{\beta}$, como $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ y $\hat{\beta}$ es una transformación lineal de $\mathbf{Y}$, la distribución resultante también será normal, como esta distribución queda determinada por la media y la varianza basta calcular:

$$E[\hat{\beta}] = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\beta = \beta$$

$$Var[\hat{\beta}] = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\sigma^2\mathbf{I}_n\mathbf{X}((\mathbf{X}^t\mathbf{X})^{-1})^t = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}$$

Por tanto $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$. Como $\hat{\psi} = \lambda^t\hat{\beta}$ aplicando el mismo razonamiento calculamos la media y la varianza de $\hat{\psi}$:

$$E[\hat{\psi}] = \lambda^t E[\hat{\beta}] = \lambda^t\beta = \psi$$

$$Var[\hat{\psi}] = \sigma^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda = \sigma^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda$$

Y por tanto conluimos que: $\hat{\psi} \sim \mathcal{N}(\psi, \sigma^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda)$.

b) Por el apartado anterior sabemos que $\hat{\psi} \sim \mathcal{N}(\psi, \sigma^2 \lambda^t (\mathbf{X}^t\mathbf{X})^{-1}\lambda)$. Tipificando obtenemos que

$$\frac{\hat{\psi} - \psi}{\sqrt{\sigma^2 \lambda^t (\mathbf{X}^t\mathbf{X})^{-1}\lambda)}} \sim \mathcal{N}(0,1)$$

Por otro lado, por los apartados b) y c) de la Proposición 7 del Tema 2 sabemos que $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$ y dicha variable es independiente de $\hat{\beta}$ y por lo tanto de cualquier combinación lineal suya, es decir $\hat{\psi}$. De todo ello se deduce que

$$\frac{\frac{\hat{\psi}-\psi}{\sqrt{\sigma^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda)}}}{\sqrt{\frac{(n-p)\hat{\sigma}^2}{(n-p)\sigma^2}}} = \frac{\hat{\psi} - \psi}{\sqrt{\hat{\sigma}^2 \lambda^t (\mathbf{X}^t\mathbf{X})^{-1}\lambda)}} \sim t(n-p)$$

En consecuencia, si $t_{n-p,\alpha/2}$ es el cuantil de orden $1-\alpha/2$ de la distribución $t(n-p)$ se verifica que $P(-t_{n-p,\alpha/2} \leq \frac{\hat{\psi}-\psi}{\sqrt{\hat{\sigma}^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda}} \leq t_{n-p,\alpha/2}) = 1-\alpha$. Despejando ψ obtenemos que $P(\hat{\psi} - \sqrt{\hat{\sigma}^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda} t_{n-p,\alpha/2} \leq \psi \leq \hat{\psi} + \sqrt{\hat{\sigma}^2\lambda^t(\mathbf{X}^t\mathbf{X})^{-1}\lambda} t_{n-p,\alpha/2}) = 1-\alpha$ y el ejercicio está completo.

4

4.1 Enunciado

Para estimar dos parámetros θ y Φ se realizan observaciones de tres tipos:

a) m observaciones $Y_{11}, ..., Y_{1m}$ con media θ

b) m observaciones $Y_{21}, ..., Y_{2m}$ con media $\theta + \Phi$

c) n observaciones $Y_{31}, ..., Y_{3n}$ con media $\theta - 2\Phi$

Todas las observaciones están sujetas a errores incorrelados de media 0 y tienen varianza σ^2. Hallar los estimadores de mínimos cuadrados de θ y Φ. Probar que dichos estimadores son incorrelados si y sólo si $m = 2n$.

4.2 Respuesta

Sabemos que $E[Y_{1i}] = \theta \, \forall i \in \{1, ..m\}$, $E[Y_{1i}] = \theta + \Phi \, \forall i \in \{1, ..m\}$ y $E[Y_{3i}] = \theta - 2\Phi \, \forall i \in \{1, ..n\}$. Construimos el vector $\mathbf{Y} = (Y_{11}, ..., Y_{1m}, Y_{21}, ..., Y_{2m}, Y_{31}, ..., Y_{3n})$ que nos permite expresar las igualdades anteriores de forma matricial:

$$E[\mathbf{Y}] = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -2 \\ \vdots & \vdots \\ 1 & -2 \end{pmatrix} \begin{pmatrix} \theta \\ \Phi \end{pmatrix} \tag{1}$$

Por otra parte sabemos que existen unos errores $\varepsilon = (\varepsilon_{11}, ..., \varepsilon_{1m}, \varepsilon_{21}, ..., \varepsilon_{2m}, \varepsilon_{31}, ..., \varepsilon_{3n})$ que cumplen que $E[\varepsilon] = 0$ y $Cov[\varepsilon_{ij}, \varepsilon_{tk}] = \sigma^2\delta_{ij,tk}$. Como $r(\mathbf{X}) = 2$ nos encontramos ante un modelo lineal de rango completo donde $E[\mathbf{Y}] = \mathbf{X}\beta + \varepsilon$

Para obtener el EMC de θ y Φ basta encontrar el EMC de β. Por la Proposición 1 del Tema 2 sabemos que el EMC de β es de la forma $\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$.
Tras realizar los cálculos (los he hecho en sucio) llegamos a que la matriz $\hat{\beta}$ es de la forma:

$$\begin{pmatrix} \hat{\theta} \\ \hat{\Phi} \end{pmatrix} = \frac{1}{m^2 + 13nm} \begin{pmatrix} (m+4n)\mathbf{Y}_1 + 6n\mathbf{Y}_2 + 3m\mathbf{Y}_3 \\ (-m+2n)\mathbf{Y}_1 + (m+3n)\mathbf{Y}_2 + -5m\mathbf{Y}_3 \end{pmatrix} \tag{2}$$

Donde $\mathbf{Y}_1 = \sum_{i=1}^m Y_{1i}$, $\mathbf{Y}_2 = \sum_{i=1}^m Y_{2i}$, $\mathbf{Y}_3 = \sum_{i=1}^n Y_{3i}$.
Además por el apartado a) de la Proposición 1 del Tema 2 sabemos que $(\hat{\beta}, \hat{\Phi})$ es un estimador insesgado. Veamos que su matriz de covarianzas es diagonal si y solo si $m = 2n$:

$$Cov[(\hat{\theta}, \hat{\Phi})] = \begin{pmatrix} Cov[\hat{\theta}] & Cov[\hat{\theta}, \hat{\Phi}] \\ Cov[\hat{\Phi}, \hat{\theta}] & Cov[\hat{\Phi}] \end{pmatrix} \tag{3}$$

Pero también sabemos que

$$Cov[(\hat{\theta}, \hat{\Phi})] = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1} = \frac{1}{m^2 + 13nm} \begin{pmatrix} m+4n & -m+2n \\ -m+2n & 2m+n \end{pmatrix} \tag{4}$$

Entonces $Cov[(\hat{\theta}, \hat{\Phi})] = 0 \iff \frac{1}{m^2+13nm}(-m+2n) = 0 \iff -m+2n = 0 \iff m = 2n$ Por lo tanto los estimadores $\hat{\theta}, \hat{\Phi}$ son insesgados si y solo si $m = 2n$

5

5.1 Enunciado

Supongamos que $\mathbf{Y}$ se ajusta a un modelo lineal básico de rango completo, es decir, $E[\mathbf{Y}] = \mathbf{X}\beta$ y $Cov[\mathbf{Y}] = \sigma^2 I_n$ siendo $\mathbf{X}$ una matriz de orden $n \times p$

con $p < n$ y $r(\mathbf{X}) = p$. Si definimos el vector *valores ajustados* como $\hat{\mathbf{Y}} = (\hat{Y}_1, .., \hat{Y}_n)^t = \mathbf{X}\hat{\beta}$ y el vector de *residuos* como $\hat{e} = (e_1, .., e_n)^t = \mathbf{Y} - \hat{\mathbf{Y}}$. Se pide:

a) Calcular $E[\hat{\mathbf{Y}}]$, $Cov[\hat{\mathbf{Y}}]$, $E[\hat{e}]$ y $Cov[\hat{e}]$

b) Probar que $\sum_{i=1}^n Var[\hat{\mathbf{Y}}_i] = \sigma^2 p$ y que $\sum_{i=1}^n \hat{\mathbf{Y}}_i e_i = 0$

5.2 Respuesta

a)

1. $E[\hat{\mathbf{Y}}] = E[\mathbf{X}\hat{\beta}] = \mathbf{X}E[\hat{\beta}] = \mathbf{X}\beta$ donde $\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$
2. $Cov[\hat{\mathbf{Y}}] = Cov[\mathbf{X}\hat{\beta}] = \mathbf{X}Cov[\hat{\beta}]\mathbf{X}^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t Cov[\mathbf{Y}]\mathbf{X}^t((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)^t = \sigma^2\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$
3. $E[\hat{e}] = E[\mathbf{Y} - \hat{\mathbf{Y}}] = E[\mathbf{Y}] - E[\hat{\mathbf{Y}}] = 0$
4. $Cov[\hat{e}] = E[(\hat{e} - E[\hat{e}])(\hat{e} - E[\hat{e}])^t] = E[ee^t] = E[SEC]$ y por la Proposición 4 del Tema 2 sabemos que $E[SEC] = (n-p)\sigma^2$ por lo tanto $Cov[\hat{e}] = (n-p)\sigma^2$.

b) $\sum_{i=1}^n Var[\hat{\mathbf{Y}}_i] = tr(Cov[\hat{\mathbf{Y}}]) = tr(\sigma^2\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)$ pero por el Lema 3 del Tema 2 sabemos que $\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ es idempotente:

$$(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$$

y ademas es simétrica, por lo tanto por la propiedad d) de las matrices itempotentes del Tema 1 tenemos que $r(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = tr(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)$, por la propiedad b) del rango del Tema 1 sabemos que $r(\mathbf{X}) = r(\mathbf{X}^t\mathbf{X}) = r((\mathbf{X}^t\mathbf{X})^{-1})$ por otra parte $\mathbf{X}$ es no singular y por tanto invertible, por la propiedad a) del mismo apartado sabemos que $r(A) = r(BA) = r(AC)$ con B y C matrices no singulares, por lo tanto como $r(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = r(\mathbf{X}) = p$, concluimos que $tr(\sigma^2\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = \sigma^2 tr(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = \sigma^2 r(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = \sigma^2 p$

Para demostrar que $\sum_{i=1}^n \hat{\mathbf{Y}}_i e_i = 0$ debemos de tener en cuenta que $\hat{\mathbf{Y}} = H\mathbf{Y}$ y que $\hat{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I}_n - H)\hat{\mathbf{Y}}$ donde $H = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, como sabemos que H y $\mathbf{I}_n - H$ generan subespacios sobre $\mathbb{R}^n$ ortogonales entre si concluimos que $(\mathbf{I}_n - H)\hat{\mathbf{Y}}$ y $H\mathbf{Y}$ son vectores ortogonales.

6

6.1 Enunciado

Sea $\mathbf{Y} \sim \mathcal{N}(\mu, \mathbf{V})$

a) Si $\mathbf{Z} = \mathbf{A}\mathbf{Y} + b$ con $\mathbf{A}$ matriz $k \times n$ de rango k $(k \leq n)$ y $b \in \mathbb{R}^k$, entonces $\mathbf{Z} \sim \mathcal{N}(\mathbf{A}\mu + b, \mathbf{A}\mathbf{V}\mathbf{A}^t)$.

b) Probar que un vector aleatorio $\mathbf{Y} = (Y_1, .., Y_n)^t \sim \mu, \mathbf{V}$ si y solo si para todo $\lambda \in \mathbb{R}^n$, $\lambda^t\mathbf{Y} \sim \mathcal{N}(\lambda^t\mu, \lambda^t\mathbf{V}\lambda)$

6.2 Respuesta

(a partir de ahora la transpuesta será denotada con el símbolo T en vez de t)

a) Sabemos que una v.a sigue una distribución normal cuando su función de momentos es de la forma: $M_X(t) = exp(t^T\mu + \frac{1}{2}t^TVt)$ donde V es la matriz de covarianzas. Por otra parte la función de momentos de una v.a cualquiera es de la forma $M_R(t) = E[exp(t^Tx)]$, Veamos que $\mathbf{Z}$ tiene una función de momentos de la primera forma:

$M_Z(t) = E[exp(t^T(\mathbf{A}x+b))] = E[exp(t^T(\mathbf{A}x))exp(t^Tb)] = exp(t^Tb)E[exp(t^TAx)] = exp(t^Tb)M_Y(\mathbf{A}^Tt) = exp(t^T(\mathbf{A}\mu + b) + \frac{1}{2}t^T\mathbf{A}\mathbf{V}\mathbf{A}^Tt)$
Concluimos que $\mathbf{Z}$ sigue una distribución normal

b)
$\Rightarrow$

Una v.a también queda determinada por su función característica, en el caso de una v.a con distribución normal, esta es: $\phi_Y(t) = exp(it\mu - \frac{1}{2}t^T\mathbf{V}t)$, por otra parte la función característica de $\lambda^T\mathbf{Y} = E[exp(it(\lambda^T\mathbf{V}))] = E[exp(i(t\lambda^T)\mathbf{V})] = exp(i(\lambda t)^T\mu - \frac{1}{2}(\lambda t)^T)\mathbf{V}\lambda t) = exp(it(\lambda^T\mu) - \frac{1}{2}t(\lambda^T\mathbf{V}\lambda)t$ Por la unicidad de la función característica concluimos que $\lambda^t\mathbf{Y} \sim \mathcal{N}(\lambda^t\mu, \lambda^t\mathbf{V}\lambda^t)$ Otra forma más cómoda de verlo es reducirnos al caso a), tomando $b = 0$ y $k = 1$

$\Leftarrow$

$\phi_Y(\beta) = E[exp(i\beta^T\mathbf{Y})]$ tomando $\beta = \lambda t$ tenemos $= E[exp(i(t\lambda)^T\mathbf{Y})] = E[exp(it(\lambda^T\mathbf{Y}))] = exp(i(\lambda t)^T\mu - \frac{1}{2}(\lambda t)^T\mathbf{V}\lambda t) = exp(i(\beta)^T\mu - \frac{1}{2}(\beta)^T\mathbf{V}\beta)$ y por la unicidad de la función caracteristica concluimos que $\mathbf{Y} \sim \mathcal{N}(\mu, \mathbf{V})$.