Tema1. Modelo Lineal General.

1. Si $X = (X_1, X_2, X_3, X_4)^t$ tiene distribución normal con vector de medias $\mu = (2, 1, -1, -3)^t$ y matriz de covarianzas

$$V = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 3 & 0 \\ -1 & 1 & 0 & 2 \end{pmatrix}$$

Halla:

   a) La distribución marginal de los vectores $(X_2, X_1, X_3)^t$ y $(X_1, X_4)^t$.

   b) La distribución condicional de $(X_1, X_4)^t$ dado $X_2 = x_2$, $X_3 = x_3$.

   c) La distribución de $Z = 2X_1 - 6X_3 + 4X_4$ y la del vector $(Z_1, Z_2)^t$ siendo $Z_1 = X_1 - 3X_4 + 4X_2$ y $Z_2 = X_3 + 2X_2 - X_1 + 2X_4$.

2. Un vector aleatorio $(X, Y)^t$ tiene función de densidad conjunta

$$f(x,y) = \frac{1}{\sigma_x \sigma_y \sqrt{1-\rho^2}(2\pi)} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho \frac{(x-\mu_x)}{\sigma_x} \frac{(y-\mu_y)}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right] \right\}$$

Demuestra que $(X, Y)^t$ sigue una distribución normal bidimensional. Halla su vector de medias y su matriz de covarianzas. ¿Qué representa el parámetro ρ?

3. En las condiciones del problema anterior, prueba que si $\mu_x = \mu_y = 0$ entonces $\mathrm{Cov}(X^2, Y^2) = 2(\mathrm{Cov}(X, Y))^2$.

4. a) Sea $Y \sim N_n(\mathbf{0}, V)$. Prueba que los momentos de orden 3 de Y son nulos.

   b) Sea $X \sim N(0, \sigma^2)$. Prueba que $E[X^4] = 3(\sigma^2)^2$.

   Indicación: Utiliza en ambos casos la función generatriz de momentos. En el apartado a), podemos suponer que Y es tridimensional. Los momentos de orden 3 serán de la forma $E[Y_i^3]$, $E[Y_i^2 Y_j]$ o bien $E[Y_i Y_j Y_k]$.

5. Halla la media y varianza de una distribución chi-cuadrado con n grados de libertad y coeficiente de descentralización μ^*.

6. Dado un vector aleatorio n dimensional Y y la matriz simétrica A, la variable aleatoria $Y^t A Y$ será denominada forma cuadrática asociada a Y. Los siguientes resultados nos relacionan la distribución y los momentos de $Y^t A Y$ con los de Y. Supongamos que Y tiene vector de medias mu y matriz de covarianzas V. Entonces se verifica:

   a) $E[Y^t A Y] = \mathrm{tr}(AV) + \mu^t A \mu$

   Si además $Y \sim N_n(\mu, V)$, entonces

   b) $\mathrm{Var}[Y^t A Y] = 2\,\mathrm{tr}((AV)^2) + 4\mu^t A V A \mu$.

   c) $\mathrm{Cov}[Y, Y^t A Y] = 2VA\mu$.

7. Sean $Q_1 \sim \chi^2(n_1)$, $Q_2 \sim \chi^2(n_2)$ con $n_1 > n_2$ y sea $Q = Q_1 - Q_2$ independiente de Q_2. Prueba que $Q \sim \chi^2(n_1 - n_2)$. (Indicación: Utiliza la función generatriz de momentos)

8. Si Y es un vector aleatorio n dimensional con distribución $N_n(\mu, V)$,con V no singular, prueba que $Y^t V^{-1} Y$ tiene distribución $\chi^2(n, \lambda)$, siendo $\lambda = \frac{1}{2}\mu^t V^{-1} \mu$.

9. Sea Y un vector aleatorio n dimensional con distribución $N_n(\mathbf{0}, I_n)$ y sea X una matriz de orden $n \times p$ (con $n > p$) y rango p.

a) ¿Cuál es la distribución de $Y^t X(X^t X)^{-1} X^t Y$?

b) ¿Y la distribución de $Y^t(I_n - X(X^t X)^{-1} X^t)Y$?

c) Halla $E[Y^t(I_n - X(X^t X)^{-1} X^t)Y]$

10. Sea X una matriz como en el problema anterior y sea Y un vector aleatorio n dimensional con distribución $N_n(X\beta, I_n)$, donde β es un vector $p \times 1$.

a) ¿Cuál es la distribución de $Y^t X(X^t X)^{-1} X^t Y = Q_1$?

b) ¿Y la distribución de $Y^t(I_n - X(X^t X)^{-1} X^t)Y = Q_2$?

c) Prueba que Q_1 y Q_2 son independientes.

d) Halla la distribución de $u = \dfrac{n-p}{p}\dfrac{Q_1}{Q_2}$

11. Sea $Y = (Y_1, \ldots, Y_n)^t$ un vector aleatorio n dimensional con distribución $N_n(\mathbf{0}, I_n)$. Denotaremos por $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$

a) Halla la distribución de $Q_1 = n\bar{Y}^2$.

b) Halla la distribución de $Q_2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$.

c) Prueba que Q_1 y Q_2 son independientes.

d) ¿Qué distribución relacionarías con Q_1/Q_2? Razona la respuesta.

12. Sean $Y_1, \ldots, Y_n$ variables aleatorias i.i.d. con distribución común $N(\mu, \sigma^2)$.

a) ¿Satisface $Y = (Y_1, \ldots, Y_n)^t$ un modelo lineal normal?, ¿de rango completo o no completo?

b) Obtén las distribuciones de $\bar{Y}$ y $(n-1)S^2/\sigma^2$, siendo $S^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$. Prueba que ambas variables son independientes.

13. Sean $X_1, \ldots, X_{n_1}$ variables aleatorias i.i.d. con distribución común $N(\mu_1, \sigma^2)$ e $Y_1, \ldots, Y_{n_2}$ variables aleatorias i.i.d. con distribución común $N(\mu_2, \sigma^2)$. Además ambos conjuntos de variables aleatorias son independientes entre si. Denotemos

$$\bar{X} = n_1^{-1}\sum_{i=1}^{n_1} X_i\,, \quad \bar{Y} = n_2^{-1}\sum_{i=1}^{n_2} Y_i\,, \quad S_1^2 = (n_1-1)^{-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2\,, \quad S_2^2 = (n_2-1)^{-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$

a) ¿Satisface $Y = (X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2})^t$ un modelo lineal normal?, ¿de rango completo o no completo?

b) Obtén las distribución de Q_1/σ^2, siendo $Q_1 = n_1\bar{X}^2 + n_2\bar{Y}^2$.

c) Obtén las distribución de Q_2/σ^2, siendo $Q_2 = (n_1-1)S_1^2 + (n_2-1)S_2^2$.

d) ¿Qué distribución relacionarías con Q_1/Q_2? Razona la respuesta.

Tema 2. Modelo lineal de rango completo.

1. Sea $Y = X\beta + \mathcal{E}$ un modelo lineal donde $Y = (Y_1, \ldots, Y_n)^t$ y $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_n)$ son variables aleatorias n-dimensionales, X es una matriz de orden $n \times p$ y $\beta = (\beta_1, \ldots, \beta_p)^t \in \mathbb{R}^p$. Supondremos $r(X) = p$ y que $E[\mathcal{E}] = 0$ y $\text{Cov}[\mathcal{E}] = \sigma^2 I_n$, es decir, se trata de un modelo lineal de rango completo básico. Denotaremos por $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^t = (X^tX)^{-1}X^tY$, al estimador de mínimos cuadrados de β. Si definimos los *valores predichos*, *predicciones* o *valores ajustados* como $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)^t = X\hat{\beta}$ y los *residuos* como $e = (e_1, \ldots, e_n)^t = Y - \widehat{Y}$:

   a) Obtén $E[\hat{Y}]$, $\text{Cov}[\hat{Y}]$, $E[e]$ y $\text{Cov}[e]$.

   b) Prueba que $\sum_{i=1}^n \text{Var}[\hat{Y}_i] = \sigma^2 p$.

   c) Prueba que $\sum_{i=1}^n \hat{Y}_i e_i = 0$.

   d) Prueba que $e^te = Y^tY - \hat{\beta}^tX^tX\hat{\beta}$

2. Sea $Y = X\beta + \mathcal{E}$ un modelo lineal básico de rango completo, $r(X) = p$, con la siguiente partición $Y = X_1\gamma_1 + X_2\gamma_2 + \mathcal{E}$, donde γ_1 es un vector $r \times 1$ y X_1 una matriz $n \times r$ ($r < p$). Prueba que $\hat{\gamma}_1 = (X_1^tX_1)^{-1}X_1^tY$ no es un estimador insesgado de γ_1. ¿Bajo qué condiciones será insesgado?

3. **(Diseño ortogonal)** Consideremos un modelo lineal de rango completo $Y = X\beta + \mathcal{E}$ como en el ejercicio anterior. Partimos X en la forma $X = (X_1, \ldots, X_p) = (W, X_p)$,donde $X_1, \ldots, X_p$ son las columnas de X.

   a) Probar que
   $$|X^tX| = |W^tW|(X_p^tX_p - X_p^tW(W^tW)^{-1}W^tX_p)$$

   b) Deducir que
   $$\frac{|W^tW|}{|X^tX|} \geq \frac{1}{X_p^tX_p}$$
   y como consecuencia, probar que $\text{Var}[\hat{\beta}_p] \geq \sigma^2(X_p^tX_p)^{-1}$ y se da la igualdad si y sólo si $X_p^tX_j = 0$ $(j = 1, \ldots, p-1)$

   c) Demuestra que $\hat{\beta}_1, \ldots, \hat{\beta}_p$ son incorrelados si y sólo si $X_1, \ldots, X_p$ son ortogonales (i.e. $X_i^tX_j = 0$, para todo $i, j = 1, \ldots, p$, $i \neq j$)

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

   a) las del primer tipo tienen de media θ.

   b) las del segundo tipo tienen de media $\theta + \Phi$.

   c) las del tercer tipo tienen de media $\theta - 2\Phi$

   Todas las observaciones están sujetas a errores incorrelados de media 0 y varianza constante.

   Si se efectúan m observaciones de tipo a), m observaciones de tipo b) y n observaciones de tipo c), halla los estimadores de mínimos cuadrados de θ y Φ. Prueba que dichos estimadores son incorrelados si y sólo si $m = 2n$.

5. **(Intervalos de confianza simultáneos para β_i: Método de Bonferroni)** Sea $Y = X\beta + \mathcal{E}$, donde X es una matriz $n \times p$ de rango $p(< n)$ y $\mathcal{E} \sim \mathcal{N}_n(0, \sigma^2 I_n)$. Sean
   $$\Lambda_i = \left[\hat{\beta}_i - t_{\alpha/(2p)}(n-p)\sqrt{c_{ii}\tilde{\sigma}^2}, \hat{\beta}_i + t_{\alpha/(2p)}(n-p)\sqrt{c_{ii}\tilde{\sigma}^2}\right] \qquad i = 1, \ldots, p$$
   donde $t_\gamma(n-p)$ es el cuantil de orden $1-\gamma$ de una distribución $t(n-p)$, $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^t = (X^tX)^{-1}X^tY$, $\tilde{\sigma}^2 = (Y - X\hat{\beta})^t(Y - X\hat{\beta})/(n-p)$ y $(X^tX)^{-1} = (c_{ij})_{i,j=1,\ldots,p}$. Probar que
   $$P(\beta \in \Lambda_1 \times \ldots \times \Lambda_p) \geq 1 - \alpha.$$

6. **(Región de confianza para β)** Sea $Y = X\beta + \mathcal{E}$, donde X es una matriz $n \times p$ de rango $p(< n)$ y $\mathcal{E} \sim \mathcal{N}_n(0, \sigma^2 I_n)$. Probar que

$$1 - \alpha = P\left((\beta - \hat{\beta})^t X^t X(\beta - \hat{\beta}) \leq p\tilde{\sigma}^2 F_\alpha(p, n-p)\right)$$

donde $F_\alpha(p, n-p)$ es el cuantil de orden $1 - \alpha$ de una distribución $F(p, n-p)$, $\hat{\beta} = (X^t X)^{-1} X^t Y$ y $\tilde{\sigma}^2 = (Y - X\hat{\beta})^t (Y - X\hat{\beta})/(n-p)$.

7. Sea $Y = X\beta + \mathcal{E}$ un modelo lineal básico de rango completo, $r(X) = p$, e $Y = Z\gamma + \mathcal{E}$ una reparametrización suya, es decir $Z = XU$ siendo U no singular. Expresa $\hat{\gamma}$ en términos de $\hat{\beta}$.

8. Sea $Y_{11}, \ldots, Y_{1n_1}$ una muestra aleatoria simple extraída de la población $N(\mu_1, \sigma^2)$, e $Y_{21}, \ldots, Y_{2n_2}$ una muestra aleatoria simple extraída de la población $N(\mu_2, \sigma^2)$. Supongamos que ambas muestras son independientes.

   a) Obtén intervalos de confianza para μ_1, μ_2 y $\mu_1 - \mu_2$.

   b) Obtén la tabla de análisis de la varianza para el test $H_0 : \mu_1 = \mu_2$.

9. **(Modelo de regresión lineal)** Considera el siguiente modelo lineal de rango completo

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \mathcal{E}_i \quad (i = 1, \ldots, n)$$

   a) Prueba que $\sum_{i=1}^n e_i = 0$.

   b) Prueba que si $\bar{Y} := n^{-1} \sum_{i=1}^n Y_i$ entonces

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

10. En el modelo del ejercicio anterior definimos el *coeficiente de correlación múltiple* R como la correlación entre Y e $\hat{Y}$, es decir,

$$R = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\left(\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2\right)^{1/2}}$$

siendo $\bar{\hat{Y}} = n^{-1} \sum_{i=1}^n \hat{Y}_i$. Probar que

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{\hat{\beta}^t X^t Y - n\bar{Y}^2}{Y^t Y - n\bar{Y}^2}$$

11. Sea $Y_i = \beta_0 + \beta_1 x_i + \mathcal{E}_i$ $(i = 1, \ldots, n)$, donde $E[\mathcal{E}] = 0$ y $\mathrm{Cov}[\mathcal{E}] = \sigma^2 I_n$. Encontrar los estimadores de mínimos cuadrados de β_0 y β_1. Probar que son incorrelados si y sólo si $\bar{x} = 0$.

12. Consideramos el modelo lineal

$$Y_i = \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \mathcal{E}_i \qquad i = 1, \ldots, n$$

siendo $(a_1, \ldots, a_n)^t$, $(b_1, \ldots, b_n)^t$ y $(c_1, \ldots, c_n)^t$ linealmente independientes y $\mathcal{E} \sim \mathcal{N}_n(0, \sigma^2 I_n)$.

   a) Escribir las ecuaciones normales del modelo.

   b) Hallar intervalos de confianza al 95% para β_i, $\;i = 1, 2, 3$ y σ^2.

   c) Escribir la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \beta_1 = \beta_3 = 0$.

d) Escribir la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : 2\beta_1 - \beta_2 = \beta_3 - \beta_2 = 0$.

13. Hallar $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ y $\hat{\sigma}^2$, asumiendo que los datos

y	1	5	0	4	4	-1
x_1	1	2	1	3	3	3
x_2	1	1	2	1	2	3

se ajustan al modelo de regresión lineal $Y_i = \beta_1 + \beta_2 b_i + \beta_3 c_i + \mathcal{E}_i$. ¿Con estos datos, tenemos razones suficientes para rechazar la hipótesis del apartado c) del ejercicio anterior?

14. **(Modelo de regresión polinómica)** Consideramos el modelo lineal

$$Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \cdots + \beta_p x_j^p + \mathcal{E}_j \qquad j = 1, \ldots, n, \quad p + 1 < n$$

donde x_j $j = 1, \ldots, n$ son distintos $\mathcal{E} \sim \mathcal{N}_n(0, \sigma^2 I_n)$

a) Comprobar que se trata de un modelo lineal de rango completo y hallar las ecuaciones normales para estimar β.

b) Escribir las tablas de análisis de la varianza para contrastar la hipótesis $H_0 : \beta_p = 0$ en los modelos correspondientes a $p = 1$ (regresión lineal), $p = 2$ (regresión por polinomios cuadráticos) y $p = 3$ (regresión por polinomios cúbicos).

c) En el caso $p = 2$, escribir la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \beta_0 + 2\beta_1 - \beta_2 = 2\beta_1 - \beta_2 = 0$.

15. Se consideran los siguientes datos

y	0.5	3.0	1.6	1.3	0.2	0.5	0.1	1.2
x	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5

Supondremos que se ajustan a un modelo polinómico de grado 3 como el del ejercicio anterior.

a) Estimar los valores de $\beta_0, \beta_1, \beta_2, \beta_3$ y σ^2.

b) Hallar intervalos de confianza al 95% para los parámetros del apartado anterior.

c) ¿Podemos aceptar la hipótesis de que los datos se ajustan en realidad a un polinomio de grado 2?

Tema 3. Modelo lineal de rango no completo.

1. Considera el modelo lineal de rango no completo
$$Y = X\beta + \mathcal{E}$$
siendo X de orden $n \times p$ y $r(X) = k < p$ y $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I_n)$. Supón que $\Psi_1 = \lambda_1^t \beta$ y $\Psi_2 = \lambda_2^t \beta$ son funciones lineales estimables y que $\hat{\Psi}_1$ y $\hat{\Psi}_2$ son sus estimadores lineales insesgados de mínima varianza. Halla $\mathrm{Cov}(\hat{\Psi}_1, \hat{\Psi}_2)$.

2. Para el modelo
$$y_{ij} = \mu + \tau_i + \mathcal{E}_{ij} \qquad j = 1, \ldots, n_i, \quad i = 1, 2$$
   a) Halla las ecuaciones normales y dos funciones estimables linealmente independientes.
   b) Obtén los estimadores de las funciones del apartado a), su varianza y su covarianza.
   c) Reparametriza de modo adecuado el modelo anterior, para obtener un modelo lineal de rango completo.
   d) ¿Es estimable la hipótesis $H_0 : \tau_1 = \tau_2$? En caso de respuesta afirmativa, construye la tabla de análisis de la varianza.

3. Considera el modelo $Y = X\beta + \mathcal{E}$ que viene dado por las ecuaciones
$$\begin{array}{rcccccccc} y_1 = & \beta_1 & + & \beta_2 & + & \beta_3 & + & \mathcal{E}_1 \\ y_2 = & \beta_1 & & & + & \beta_3 & + & \mathcal{E}_2 \\ y_3 = & & & \beta_2 & & & + & \mathcal{E}_3 \\ y_4 = & 2\beta_1 & - & 3\beta_2 & + & 2\beta_3 & + & \mathcal{E}_4 \end{array}$$
   a) Halla las ecuaciones normales del modelo y dos funciones estimables.
   b) Prueba que $\beta_1 + \beta_3 - 2\beta_2$ también es estimable y halla su estimador lineal insesgado de mínima varianza.
   c) Halla dos reparametrizaciones distintas del modelo y prueba que el estimador de $\beta_1 - \beta_2 + \beta_3$ es el mismo en ambas.
   d) Calcula las varianzas de los estimadores de $\beta_1 + \beta_3 - 2\beta_2$ y $\beta_1 + 2\beta_2 + \beta_3$ así como la covarianza de ambos estimadores.
   e) Halla la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \beta_1 = \beta_3$ si es que dicha hipótesis es estimable.

4. Para el modelo
$$y_{ij} = \mu + \tau_i + \delta_j + \mathcal{E}_{ij} \qquad i = 1, \ldots, n \quad j = 1, \ldots, m$$
prueba que $\sum_{j=1}^m a_j \delta_j$ es una función lineal estimable si y sólo si $\sum_{j=1}^m a_j = 0$.

5. Supón que en un modelo como el del ejercicio anterior para $n = 3$ y $m = 4$, tenemos los siguientes valores:

y_{11}	y_{12}	y_{13}	y_{14}	y_{21}	y_{22}	y_{23}	y_{24}	y_{31}	y_{32}	y_{33}	y_{34}
6.2	2.0	5.8	6.3	5.1	5.3	5.8	5.4	6.1	6.2	5.7	5.9

   a) Contrasta la hipótesis $H_0 : \tau_1 = \tau_2 = \tau_3$ a un nivel de significación de 0.05.
   b) Halla un intervalo de confianza al 95% para el parámetro $\Psi = \tau_1 + 3\tau_2 - 4\tau_3$.

6. En el modelo lineal de rango no completo, sean $\beta_1, \beta_2, \ldots, \beta_q$ $(q \leq k)$ un subconjunto de los elementos del vector de parámetros β. Supón que $\sum_{i=1}^{q} c_i\beta_i$ es estimable para cada conjunto de constantes c_i tales que $\sum_{i=1}^{q} c_i = 0$. Prueba que la hipótesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q$ es una hipótesis estimable y halla la tabla de análisis de la varianza para contrastarla.

7. Sea b un vector de constantes tales que $E[b^tY] = 0$; llamaremos a b^tY un estimador insesgado de cero. Prueba que un estimador insesgado de cero es incorrelado con el estimador lineal insesgado de mínima varianza de cualquier función estimable.

8. Sea $Y = X\beta + \mathcal{E}$ un modelo lineal de rango no completo con $r(X) = k$ y $\mathcal{E} \sim \mathcal{N}_n(0, \sigma^2 I_n)$. Si $\hat{\beta}_1$ y $\hat{\beta}_2$ son dos soluciones de las ecuaciones normales, prueba que $(Y - X\hat{\beta}_1)^t(Y - X\hat{\beta}_1) = (Y - X\hat{\beta}_2)^t(Y - X\hat{\beta}_2)$ y además
$$S^2 = \frac{1}{n-k}(Y - X\hat{\beta}_2)^t(Y - X\hat{\beta}_2)$$
es un estimador insesgado de σ^2.

9. En las condiciones del ejercicio anterior, prueba
   a) $(n-k)S^2/\sigma^2 \sim \chi^2(n-k)$
   b) $(\hat{\beta} - \beta)^t(X^tX)(\hat{\beta} - \beta)/\sigma^2 \sim \chi^2(k)$
   c) Ambas variables son independientes.