(1) Dado un modelo lineal general $\overrightarrow{\mathbf{Y}} = \mathbf{X}\overrightarrow{\beta} + \overrightarrow{\varepsilon}$, comprobar que el modelo ajustado $\widehat{\mathbf{Y}}$ es invariante por el cambio de escala $\mathbf{X} = c\mathbf{U}$, con $c \in \mathbb{R}$.

(2) Dado el modelo lineal $Y_i = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2) + \varepsilon_i$ para $i = 1, 2, 3$, con $x_1 = -1$, $x_2 = 0$ y $x_3 = 1$.

- (a) Encontrar las estimaciones por mínimos cuadrados de los coeficientes β_0, β_1 y β_2.
- (b) ¿Cómo son las estimaciones de β_0 y β_1 si $\beta_2 = 0$?
- (c) ¿Y las de β_0 y β_2 cuando $\beta_1 = 0$?

(3) En un modelo económico $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t$ con $\varepsilon_t \sim N(0, \sigma^2)$ e independientes para $t = 1, ..., n$, siendo n par. Sabiendo que los coeficientes están relacionados con un parámetro económico básico α tal que

$$\beta_1 + \beta_2 = \alpha \text{ y } \beta_1 + \beta_3 = -\alpha$$

- (a) Formular el modelo lineal correspondiente al índice económico Y.
- (b) Suponiendo que $x_{1t} = 1 + 2(-1)^t$, $x_{2t} = 1 - (-1)^t$, $x_{3t} = (-1)^t$ para $t = 1, ..., n$. Obtener el estimador lineal insesgado de α, y expresar su varianza.
- (c) Establecer el contraste de igualdad entre β_1 y β_3.

(4) Para analizar el efecto de un factor sobre la regresión, se recogen m observaciones de Y para cada valor de x_i en cada nivel j. Suponiendo que $x_i = i$ para $i = 1, 2$ en los dos grupos de la población ($j = 1, 2$), todas las observaciones incorreladas entre sí y con varianzas iguales, tal que $\sum_{k=1}^{m} y_{ijk} = (i + 2(j-1))m$.

- (a) Formular el modelo lineal.
- (b) Estimar las regresiones en cada nivel del factor.
- (c) Analizar si el factor provoca una diferencia significativa entre las pendientes de las regresiones (bajo la suposición de normalidad).

(5) Sea $\overrightarrow{\mathbf{Y}} = \mathbf{X}\overrightarrow{\beta} + \overrightarrow{\varepsilon}$ un modelo lineal con matriz de diseño de rango no completo, $rango(\mathbf{X}) = k \le p$. Sea $\theta = \overrightarrow{\lambda}^t \overrightarrow{\beta}$ una función estimable:

- (a) Comprobar que si $\widehat{\beta} = \mathbf{P}\mathbf{X}^t\overrightarrow{\mathbf{Y}}$ es un estimador de $\overrightarrow{\beta}$, entonces $\widehat{\beta} = \mathbf{P}\mathbf{X}^t\overrightarrow{\mathbf{Y}} + \left(\mathbf{P}\mathbf{X}^t\mathbf{X} - \mathbf{I}_p\right)\overrightarrow{\mathbf{Z}}$ es otro estimador, para todo vector p-dimensional $\overrightarrow{\mathbf{Z}}$.
- (b) Probar que el estimador $\widehat{\theta}$ de mínimos cuadrados de θ es único.
- (c) Obtener la media y varianza del estimador de mínimos cuadrados.

(6) Sea $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ para $i = 1, ..., k$ y $j = 1, ..., n_i$, siendo ε_{ij} variables aleatorias independientes con $E(\varepsilon_{ij}) = 0$ y $Var(\varepsilon_{ij}) = \sigma^2$ para todo i, j.

- (a) Expresarlo como modelo lineal de rango no completo y obtener las ecuaciones normales en términos de $(\mu, \alpha_1, ..., \alpha_k)$.
- (b) Probar que $\theta = \sum_{i=1}^{k} c_i \alpha_i$ es estimable si y sólo si $\sum_{i=1}^{k} c_i = 0$.
- (c) Encontrar una solución de las ecuaciones normales y un estimador para $\theta = \alpha_1 - \alpha_2$.

(7) Sea Y_1,...,Y_n una muestra aleatoria de una variable poblacional con distribución normal $Y \sim N(\log \beta, \sigma^2)$, siendo σ^2 conocido. Encontrar el estimador de máxima verosimilitud de β.

(8) Sea $\overrightarrow{\mathbf{Y}} = \mathbf{X}\overrightarrow{\beta} + \overrightarrow{\varepsilon}$ un modelo lineal general. Asumiendo que todas las variables predictoras del modelo están centradas, comprobar que el coeficiente de determinación es el cuadrado del coseno entre el vector respuesta $\overrightarrow{\mathbf{Y}}$ y el vector del modelo estimado $\widehat{\mathbf{Y}}$.

(9) Un estudio sobre el peso de los recién nacidos analiza su relación con un conjunto de variables de tipo socioeconómico, antropométrico y gestacional en parejas sanas. Los datos registrados en una muestra de 100 observaciones se encuentran en la base *peso*.

- (a) Analizar el modelo de regresión para el peso de un recién nacido a través de las semanas de gestación, peso de la madre y del padre y altura de la madre y del padre (el modelo ajustado, medida de bondad y contrastes del modelo).
- (b) Realizar el diagnóstico gráfico de los residuos.
- (c) Contrastar las condiciones iniciales del análisis de regresión múltiple.
- (d) Estimar con un 95 %, el peso medio del recién nacido a las 39 semanas de gestación cuando la madre pesa 62kg y mide 168cm y el padre pesa 78kg y mide 170cm. Así mismo, predecir el peso a las 36 semanas de gestación cuando la madre pesa 74kg y mide 160cm y el padre pesa 88kg y mide 185cm.
- (e) Estudiar el modelo de regresión para el peso de un recién nacido a través de las variables más significativas entre todas las registradas en el experimento.

(10) Los estudiantes de una clase participaron en un sencillo experimento en el que cada estudiante registró su estatura, peso, sexo, preferencia para fumar, nivel de actividad usual y pulso en reposo. Luego, todos lanzaron una moneda y aquellos que obtuvieron cara corrieron durante un minuto, tras el que se registró de nuevo el pulso a toda la clase. La base de datos *pulso* contiene los datos de este experimento realizado sobre 92 estudiantes, y sus variables se describen en la siguiente tabla:

Nombre	Descripción
pulse1	Primera tasa de pulso
pulse2	Segunda tasa de pulso pasado un minuto
ran	Corrió un minuto =1, no corrió un minuto =2
smokes	Fuma regularmente =1, no fuma regularmente =2
sex	Masculino =1, femenino=2
weight	Peso del estudiante
height	Altura del estudiante
activity	Nivel usual de actividad física: 1=ligero, 2=moderado y 3=mucho

- (a) Crear una variable con el incremento del pulso observado en el experimento transcurrido el minuto en el que han corrido los que obtuvieron cara. Incluir esta nueva variable en la tabla de datos *pulso*.
- (b) Analizar los efectos principales de los diferentes tipos de factores contemplados en el experimento sobre el incremento del pulso: si corrió un minuto o no (ran), si fuma o no (smokes), si es hombre o mujer (sex) y grado de actividad habitual (activity).
- (c) Seleccionar los términos más relevantes para el estudio del diferencias en el aumento del pulso, y analizar el modelo resultante.
- (d) Discutir si puede identificarse algún grupo o grupos, mediante los factores registrados, con peor condiciones en cuanto al ritmo cardiaco medio.
- (e) ¿Son fiables las respuestas a los apartados anteriores?

(11) Para analizar el rendimiento de un proceso químico en un experimento de laboratorio, se evalúa el efecto de los niveles de temperatura y de PH bajo los que se realiza. El experimento se realizó en 5 ocasiones bajo cada uno de los niveles de PH y de temperatura en las que se registraron el rendimiento del proceso. Los tres niveles utilizados de PH fueron básico, neutro y ácido, y para temperaturas de 30ºC, 35ºC y 40ºC. Las observaciones del experimento se encuentran en el conjunto de datos *ph*.

- (a) Analizar los efectos principales de ambos factores, nivel de *pH* y la temperatura *Temp*, sobre el rendimiento del proceso químico.
- (b) Determinar si la interacción entre ambos factores provoca diferencias del rendimiento medio del proceso químico.
- (c) Discutir las representaciones gráficas del análisis de los residuos y realizar los contrastes adecuados para la validez de las conclusiones de este modelo ANOVA.