Problemas de Programación.

1. Considera un modelo lineal de rango completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n \times 1$, $\mathbf{X}$ $n \times p$, $r(\mathbf{X}) = p$, $\boldsymbol{\beta}$ $p \times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{I_n})$. Realiza un programa que calcule las estimaciones puntuales de $\beta_1, \ldots, \beta_p$ y de σ^2, así como una estimación del error cuadrático medio (salvo para σ^2) e intervalos de confianza para dichos parámetros a un nivel $1 - \alpha$ determinado.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `conf`, nivel de confianza de los intervalos.

   Salida: Tabla/matriz en cuyas columnas se presenten, por este orden, las estimaciones de σ^2, $\beta_1, \ldots, \beta_p$, las estimaciones de los errores cuadráticos medios, los extremos inferiores de los intervalos y los extremos superiores de los mismos. Los nombres de las filas deben referirse al parámetro que contienen y los de las columnas a lo que en cada una se proporciona.

2. Considera un modelo lineal de rango completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n \times 1$, $\mathbf{X}$ $n \times p$, $r(\mathbf{X}) = p$, $\boldsymbol{\beta}$ $p \times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{I_n})$. Realiza un programa que calcule las estimaciones puntuales de $\boldsymbol{\lambda}_1^t\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_k^t\boldsymbol{\beta}$ y de σ^2, así como una estimación del error cuadrático medio (salvo para σ^2) e intervalos de confianza para dichos parámetros a un nivel $1 - \alpha$ determinado.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `u`, matriz cuyas filas o columnas son $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k$; `conf`, nivel de confianza de los intervalos.

   Salida: Tabla/matriz en cuyas columnas se presenten, por este orden, las estimaciones de σ^2, $\boldsymbol{\lambda}_1^t\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_k^t\boldsymbol{\beta}$, las estimaciones de los errores cuadráticos medios, los extremos inferiores de los intervalos y los extremos superiores de los mismos. Los nombres de las filas deben referirse al parámetro que contienen y los de las columnas a lo que en cada una se proporciona.

   NOTA MÁXIMA: 7

3. Considera un modelo lineal de rango completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n \times 1$, $\mathbf{X}$ $n \times p$, $r(\mathbf{X}) = p$, $\boldsymbol{\beta}$ $p \times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{I_n})$. Realiza un programa que calcule la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \beta_{i_1} = \ldots = \beta_{i_k} = 0$, con $k < p$, y que calcule el p-valor del correspondiente test.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `hip`, vector conteniendo los índices de los parámetros considerados en H_0.

   Salida: Tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después debe aparecer el p-valor del correspondiente test.

   NOTA MÁXIMA: 8

4. Considera un modelo lineal de rango completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n \times 1$, $\mathbf{X}$ $n \times p$, $r(\mathbf{X}) = p$, $\boldsymbol{\beta}$ $p \times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{I_n})$. Realiza un programa que calcule la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \boldsymbol{\lambda}_1^t\boldsymbol{\beta} = \ldots = \boldsymbol{\lambda}_k^t\boldsymbol{\beta} = 0$, con $k < p$ y $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k$ linealmente independientes. Dicho programa debe calcular también el p-valor del correspondiente test.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `hip`, matriz cuyas filas sean los vectores $\boldsymbol{\lambda}_i^t$, $i = 1, \ldots, k$ considerados en H_0.

   Salida: Tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después debe aparecer el p-valor del correspondiente test.

   Indicación: Haz un programa auxiliar que complete la matriz `hip` a una matriz invertible. Con dicha matriz, tal como se ha visto en clase, este problema se reduce al anterior.

   NOTA MÁXIMA: 9

5. Considera un modelo lineal de rango no completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n\times 1$, $\mathbf{X}$ $n\times p$, $r(\mathbf{X}) = k < p$, $\boldsymbol{\beta}$ $p\times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I_n})$. Crea un programa, cuyo único argumento será la matriz $\mathbf{X}$, que calcule una reparametrización U de este modelo, así como la matriz $\mathbf{Z}$ del modelo ya reparametrizado. Realiza un programa que compruebe si una función $\boldsymbol{\lambda}^t\boldsymbol{\beta}$ es lineal estimable. En caso afirmativo, haciendo uso del primero de los programas, calcula la estimación puntual de $\boldsymbol{\lambda}^t\boldsymbol{\beta}$, el error cuadrático medio a la hora de estimar, y un intervalo de confianza al nivel $1-\alpha$ para dicho parámetro.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `lambda`, vector asociado a la función; `conf`, nivel de confianza de los intervalos.

   Salida: Si $\boldsymbol{\lambda}^t\boldsymbol{\beta}$ no es f.l.e. frase "La función que has introducido no es estimable en este modelo". Si dicha función es estimable, matriz de una fila que contenga, por este orden, la estimación de $\boldsymbol{\lambda}^t\boldsymbol{\beta}$, la estimación del error cuadrático medio, el extremo inferior y el superior del intervalo. Los nombres de las columnas deben referirse a lo que en cada una se proporciona.

   Indicación: Es posible utilizar algún programa correspondiente a modelos lineales de rango completo.

   NOTA MÁXIMA: 8

6. Considera un modelo lineal de rango no completo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, con $\mathbf{Y}$ $n \times 1$, $\mathbf{X}$ $n \times p$, $r(\mathbf{X}) = k < p$, $\boldsymbol{\beta}$ $p \times 1$, $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I_n})$. Realiza un programa que compruebe si la hipótesis $H_0 : \boldsymbol{\lambda}_1^t\boldsymbol{\beta} = \ldots = \boldsymbol{\lambda}_r^t\boldsymbol{\beta} = 0$, con $r < k$ y $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$ linealmente independientes, es estimable y, en caso afirmativo, que calcule la tabla de análisis de la varianza para contrastar dicha hipótesis . El programa debe calcular también el p-valor del correspondiente test.

   Argumentos: `x`, matriz del modelo; `y`, vector de observaciones; `hip`, matriz cuyas filas sean los vectores $\boldsymbol{\lambda}_i^t$, $i = 1, \ldots, r$ considerados en H_0.

   Salida: Si la hipótesis es estimable, la salida será la tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después debe aparecer el p-valor del correspondiente test. Si la hipótesis no es estimable, la salida deber ser la frase "La hipótesis introducida no es estimable en este modelo".

   Indicación: Utiliza el programa de reparametrización elaborado en el ejercicio anterior y también el programa del ejercicio 4 para un modelo de rango completo.

   NOTA MÁXIMA: 10

7. Considera un modelo lineal con un factor; diseño completamente aleatorizado.

$$Y_{ij} = \mu + \tau_i + \mathcal{E}_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, n_i,\ N = \sum_{i=1}^{r} n_i,\ \boldsymbol{\mathcal{E}} = (\mathcal{E}_{ij}) \sim \mathcal{N}_N(0, \sigma^2\mathbf{I}_N)$$

   Realiza un programa que calcule los estimadores puntuales, con sus errores cuadráticos medios, e intervalos de confianza a un nivel $1-\alpha$ determinado para σ^2 y $\mu + \tau_i$, $i = 1, \ldots, r$.

   Argumentos: `y`, vector de observaciones; `f`, factor; `conf`, nivel de confianza para los intervalos.

   Salida: Tabla en cuyas columnas se presenten, por este orden, las estimaciones de σ^2, $\mu + \tau_1, \ldots, \mu + \tau_r$, las estimaciones de los errores cuadráticos medios, los extremos inferiores de los intervalos y los extremos superiores de los mismos. Los nombres de las filas deben referirse al parámetro que contienen y los de las columnas a lo que en cada una se proporciona.

   Indicación: Los niveles del factor se obtienen con el comando `level`.

   NOTA MÁXIMA: 8

8. Considera un modelo lineal con un factor; diseño completamente aleatorizado.

$$Y_{ij} = \mu + \tau_i + \mathcal{E}_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, n_i,\ N = \sum_{i=1}^{r} n_i,\ \boldsymbol{\mathcal{E}} = (\mathcal{E}_{ij}) \sim \mathcal{N}_N(0, \sigma^2\mathbf{I}_N)$$

   Realiza un programa que calcule la tabla de análisis de la varianza para contrastar la hipótesis $H_0 : \tau_1 = \ldots = \tau_r$. El programa debe calcular también el p-valor del correspondiente test.

Argumentos: `y`, vector de observaciones; `f`, factor.

Salida: Tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después debe aparecer el p-valor del correspondiente test.

NOTA MÁXIMA: 8

9. Considera un modelo lineal con un factor; diseño completamente aleatorizado.

$$Y_{ij} = \mu + \tau_i + \mathcal{E}_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, n_i,\ N = \sum_{i=1}^{r} n_i,\ \boldsymbol{\mathcal{E}} = (\mathcal{E}_{ij}) \sim \mathcal{N}_N(0, \sigma^2 \mathbf{I}_N)$$

Realiza un programa para llevar a cabo las comparaciones múltiples en dicho modelo por el procedimiento LSD a un nivel determinado $1 - \alpha$.

Argumentos: `y`, vector de observaciones; `f`, factor; `alfa`, nivel de significación para las comparaciones.

Salida: Tabla conteniendo en sus dos primeras columnas los extremos inferior y superior de los intervalos LSD al nivel 1-`alfa` para todos los contrastes de la forma $\tau_i - \tau_j$. En la tercera columna debe aparecer la decisión de aceptación o rechazo al nivel `alfa`. Los nombres de las filas deben ser "i-j". Los de las columnas "Ext.Inf.", "Ext.Sup", "Dec.".

NOTA MÁXIMA: 9

10. Idem por el método de Bonferroni.

NOTA MÁXIMA: 9

11. Idem por el método de Tukey. Si el modelo es no balanceado toma como n la media geométrica de $n_1, \ldots, n_r$.

NOTA MÁXIMA: 9

12. Idem por el método de Scheffé.

NOTA MÁXIMA: 9

13. Considera un modelo lineal con un factor; diseño de bloques al azar.

$$Y_{ij} = \mu + \tau_i + \delta_j + \varepsilon_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, s,\ N = rs,\ \varepsilon = (\varepsilon_{ij}) \sim \mathcal{N}_N(0, \sigma^2 I_N)$$

Realiza un programa que calcule los estimadores puntuales, con sus errores cuadráticos medios, e intervalos de confianza a un nivel $1 - \alpha$ determinado para σ^2 y $\mu + \tau_i + \delta_j$, $i = 1, \ldots, r$ $j = 1, \ldots, s$.

Argumentos: `y`, vector de observaciones; `f`, factor; `b`, bloques; `conf`, nivel de confianza para los intervalos.

Salida: Tabla en cuyas columnas se presenten, por este orden, las estimaciones de σ^2, $\mu + \tau_1 + \delta_1, \ldots, \mu + \tau_r + \delta_s$, las estimaciones de los errores cuadráticos medios, los extremos inferiores de los intervalos y los extremos superiores de los mismos. Los nombres de las filas deben referirse al parámetro que contienen y los de las columnas a lo que en cada una se proporciona.

Indicación: A efectos de programación, los bloques `b` deben considerarse como un segundo factor.

NOTA MÁXIMA: 8

14. Considera un modelo lineal con un factor; diseño de bloques al azar.

$$Y_{ij} = \mu + \tau_i + \delta_j + \varepsilon_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, s,\ N = rs,\ \varepsilon = (\varepsilon_{ij}) \sim \mathcal{N}_N(0, \sigma^2 I_N)$$

Realiza un programa que calcule la tabla de análisis de la varianza para contrastar las hipótesis $H_0 : \tau_1 = \ldots = \tau_r$ y $H_0' : \delta_1 = \ldots = \delta_s$. El programa debe calcular también los p-valores de los correspondientes tests.

Argumentos: `y`, vector de observaciones; `f`, factor; `b`, bloques; `conf`, nivel de confianza para los intervalos.

Salida: Tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después deben aparecer los p-valores de los correspondientes tests.

NOTA MÁXIMA: 8

15. Considera un modelo lineal con un factor; diseño de bloques al azar.

$$Y_{ij} = \mu + \tau_i + \delta_j + \varepsilon_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, s,\ N = rs,\ \varepsilon = (\varepsilon_{ij}) \sim \mathcal{N}_N(0, \sigma^2 I_N)$$

Realiza un programa para llevar a cabo las comparaciones múltiples en dicho modelo por el procedimiento LSD a un nivel determinado $1-\alpha$.

Argumentos: `y`, vector de observaciones; `f`, factor; `b`, bloques; `alfa`, nivel de significación para las comparaciones.

Salida: Tabla conteniendo en sus dos primeras columnas los extremos inferior y superior de los intervalos LSD al nivel 1-`alfa` para todos los contrastes de la forma $\tau_i - \tau_j$. En la tercera columna debe aparecer la decisión de aceptación o rechazo al nivel `alfa`. Los nombres de las filas deben ser "i-j". Los de las columnas "Ext.Inf.", "Ext.Sup", "Dec.". Tras una línea en blanco, debe aparecer una tabla similar a la anterior para los contrastes $\delta_i - \delta_j$.

NOTA MÁXIMA: 10

16. Idem por el método de Bonferroni.

NOTA MÁXIMA: 10

17. Idem por el método de Tukey.

NOTA MÁXIMA: 10

18. Idem por el método de Scheffé.

NOTA MÁXIMA: 10

19. Considera un modelo lineal con dos factores, clasificación cruzada e interacción.

$$Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \varepsilon_{ijk},\ i = 1, \ldots, r,\ j = 1, \ldots, s,\ k = 1, \ldots, n,\ N = rsn,\ \varepsilon = (\varepsilon_{ijk}) \sim \mathcal{N}_N(\mathbf{0}, \sigma^\mathbf{2}\mathbf{I_N})$$

Realiza un programa que calcule la tabla de análisis de la varianza para contrastar las hipótesis H_0 : "el modelo no presenta interacción" y que calcule también el p-valor del correspondiente test.

Argumentos: `y`, vector de observaciones; `f1`, factor 1; `f2`, factor 2.

Salida: Tabla de análisis de la varianza, lo más parecida posible a la vista en clase, con los nombres de las filas y de las columnas. Tras la tabla debe haber una línea en blanco y después debe aparecer el p-valor del test de "interacción".

NOTA MÁXIMA: 8

Nota importante: En caso de que el usuario del programa cometa algún error a la hora de introducir los datos, las salidas de los programas deberán dar un mensaje de error que aporte la mayor información posible a dicho usuario. (Es imposible predecir lo que el usuario de un programa informático puede llegar a hacer, pero podemos controlar los errores que a nuestro juicio sean más frecuentes. Por ejemplo, la matriz no es de rango completo, el número de filas de $\mathbf{X}$ no coincide con la longitud del vector $\mathbf{Y}$, las longitudes del vector de observaciones $\mathbf{Y}$ y del factor o los bloques son distintas, etc).