

UNIVERSIDAD NACIONAL DE ROSARIO

Facultad de Ciencias Económicas y Estadística

Maestría en Estadística Aplicada

Curso de nivelación

"Métodos Estadísticos. Análisis de regresión"

Práctica

2019

1. Los productores de caña de azúcar están interesados en la relación entre **la superficie de tierras cosechadas (hectáreas)** y **la producción de caña de azúcar (en toneladas)** de esta superficie. Para dar respuesta a la inquietud de los productores se analizó la cosecha del año 2014 de 14 departamentos productores de caña de azúcar del norte argentino y se relevaron los siguientes datos:

| Departamento | Sup (Ha) | Producción |
|--------------|----------|------------|
| 1 | 13638 | 940000 |
| 2 | 6151 | 460000 |
| 3 | 5828 | 440000 |
| 4 | 931 | 65000 |
| 5 | 12222 | 830000 |
| 6 | 5302 | 380000 |
| 7 | 11979 | 860000 |

| Departamento | Sup (Ha) | Producción |
|--------------|----------|------------|
| 8 | 8175 | 590000 |
| 9 | 13679 | 1020000 |
| 10 | 8296 | 585000 |
| 11 | 13396 | 1020000 |
| 12 | 3238 | 200000 |
| 13 | 16633 | 1130000 |
| 14 | 7244 | 570000 |

Datos en archivo: Datos Ejercicio_01.xls

Sintaxis R para este ejercicio: Ejercicio_01.r

- 1.1. Calcular las medidas descriptivas para cada variable.
- 1.2. ¿Cómo es la relación existente entre las hectáreas cosechadas y las toneladas de producción de azúcar? ¿Cuál sería la variable explicativa y cuál la respuesta?
- 1.3. Ajustar un modelo de regresión lineal simple y graficarlo junto a los datos.
- 1.4. Interpretar los coeficientes de regresión estimados en términos del problema.
- 1.5. Construir el cuadro del Anova.
- 1.6. Probar la significación de la regresión utilizando la estadística F y la estadística t. Comparar los resultados.
- 1.7. Construir un intervalo del 95% de confianza para β_1 e interpretar.
- 1.8. ¿Qué valor toma el coeficiente de determinación R^2 ? Interpretar.
- 1.9. ¿Cuál será la producción media esperada de azúcar para 5302 hectáreas cosechadas? Agregar un intervalo de confianza para su estimación.
- 1.10. Un grupo de productores de un determinado departamento están interesados en predecir cuál será la producción de azúcar al final de este año. Esperan cosechar 7244 hectáreas. Realizar la estimación puntual y por intervalo. Compararlos con el valor observado.
- 1.11. ¿Podemos utilizar este modelo para predecir la producción media esperada de azúcar para 20000 hectáreas cosechadas?
- 1.12. Estudiar el cumplimiento de los supuestos.

2. En un análisis de control de calidad de un producto enlatado se desea conocer la relación entre **el tiempo de exposición del producto a 150°C de calor** y **el número de bacterias sobrevivientes**. Para ello se realizó un experimento en 12 unidades seleccionadas al azar de la línea de producción. A cada unidad se la expuso a 150°C por un determinado tiempo (en minutos) y se registró la cantidad de bacterias sobrevivientes. Se registró, además, el orden en el cuál las unidades fueron analizadas.

Datos en archivo: Datos Ejercicio_02.csv

Sintaxis R para este ejercicio: Ejercicio_02.r

- 2.1. Mediante un diagrama de dispersión analizar la relación entre estas dos variables. ¿Qué tipo de relación se observa?
- 2.2. Ajustar un modelo de regresión lineal simple. Interpretar los coeficientes estimados.
- 2.3. ¿Hay evidencia estadística suficiente para concluir que la variable tiempo de exposición del producto a 150°C aporta significativamente a la explicación del número de bacterias sobrevivientes? Utilizar la estadística F.
- 2.4. Construir un intervalo del 95% de confianza para la pendiente del modelo. Interpretar.
- 2.5. Obtener la cantidad media esperada de bacterias sobrevivientes cuando exposición a 150°C es de 11 minutos. Acompañar con un intervalo de confianza del 95%.
- 2.6. Estudiar el cumplimiento de los supuestos.

3. Una consultora de nuestra ciudad está estudiando cómo estimar la cantidad de supervisores que se necesitan para coordinar un determinado conjunto de operarios en empresas metalúrgicas del cordón industrial de Rosario. Para ello seleccionó al azar 27 empresas de esa región (con cantidades de operarios que se desean considerar en el estudio) y relevó las siguientes variables:

Cantidad de supervisores en la empresa

Cantidad de operarios supervisados.

Datos en archivo: Datos Ejercicio_03.txt

Sintaxis R para este ejercicio: Ejercicio_03.r

- 3.1. Realizar el diagrama de dispersión y postular un modelo de regresión lineal.
- 3.2. Estimar el modelo.
- 3.3. ¿En cuánto cambia la cantidad media de supervisores, a medida que la cantidad de operarios aumenta en uno? ¿Y a medida que la cantidad de operarios aumenta en 100?
- 3.4. ¿Existe regresión? Utilizar la estadística T.
- 3.5. ¿Qué porcentaje de la variabilidad total observada es explicada por la regresión?
- 3.6. El gerente de una empresa metalúrgica supone que para el próximo año va a contar con 999 operarios y desea estimar la cantidad de supervisores que va a necesitar para coordinarlos. Acompañar la predicción con un intervalo del 95%.
- 3.7. Estudiar el cumplimiento de los supuestos.

4. Se desea explicar **el salario de los profesores de matemática de la Universidad (Y)** a través de las siguientes variables:

X₁: calidad de las publicaciones (puntaje que varía de 0 a 10).

X₂: años de experiencia.

X₃: índice de buen desempeño en sus funciones (puntaje que varía de 0 a 10).

Para ello, se realizó un estudio en donde se incluyeron 24 profesores universitarios seleccionados al azar.

Datos en archivo: Datos Ejercicio_04.xls

Sintaxis R para este ejercicio: Ejercicio_04.r

- 4.1. Evaluar la relación bivariada entre las variables relevadas.
- 4.2. Postular un modelo de regresión lineal que incluya las tres variables regresoras mencionadas. Escribirlo también en notación matricial.

- 4.3. Estimar el modelo postulado e interpretar los coeficientes estimados.
- 4.4. Construir un intervalo de confianza para β_1 del 95%. Interpretar
- 4.5. Construir la tabla ANOVA y realizar el test de regresión.
- 4.6. Realizar los test parciales con la estadística t-student.
- 4.7. Realizar los test parciales con la estadística F. Comparar resultados con los del punto anterior.
- 4.8. Calcular el R^2 y el R^2 ajustado. Interpretar.
- 4.9. ¿Cuál es el salario promedio para un profesor que tiene 20 años de experiencia, un puntaje de calidad de las publicaciones de 5.3 y un índice de buen desempeño en sus funciones de 6.4? Comparar con el valor observado para el profesor #2. Acompañar esta estimación con un intervalo de confianza del 95% para la respuesta media. Interpretar.
- 4.10. Realizar el análisis de residuos.

5. Una empresa que vende notebooks y que también brinda el servicio técnico de las mismas está interesada en estimar **el tiempo en minutos que demora un técnico en realizar el servicio técnico** (Y) en función de las siguientes variables:

X₁: cantidad de desperfectos a ser reparados.

X₂: antigüedad de la notebook (en meses).

X₃: experiencia del técnico que realiza el servicio técnico (en años).

Para ello, relevó la información necesaria a partir de 18 servicios técnicos brindados en el mes.

Se ajustó un modelo de regresión lineal múltiple que incluye las tres variables explicativas consideradas y se obtuvieron las siguientes salidas. La impresora no funcionaba correctamente y hubo datos que no se imprimieron, apareciendo en su lugar ###:

| Análisis de la varianza | | | | |
|---------------------------------|----|----------------------------|-------------------------|---------|
| Fuente | DF | Suma de cuadrados | Cuadrado de la media | F-Valor |
| Modelo | 3 | 16486 | ### | ### |
| Error | 14 | 17.92475 | ### | |
| Total corregido | 17 | 16504 | | |
| R-cuadrado | | 0.9989 | R-Cuad Adj 0.9987 | |
| Estimadores de los coeficientes | | | | |
| Variable | DF | Estimador del parámetro | Error estándar | Valor t |
| Intercept | 1 | ### | 1.11989 | -11.92 |
| X1 | 1 | 14.93147 | ### | ### |
| X2 | 1 | 0.46610 | ### | ### |
| X3 | 1 | 0.18831 | ### | ### |

| Matriz $(X'X)^{-1}$ | | | | |
|---------------------|-----------|---------|---------|---------|
| | Intercept | X1 | X2 | X3 |
| Intercept | 0.9796 | -0.0619 | -0.0144 | -0.0889 |
| X1 | -0.0619 | 0.0137 | 0.0004 | -0.0017 |
| X2 | -0.0144 | 0.0004 | 0.0007 | -0.0005 |
| X3 | -0.0889 | -0.0017 | -0.0005 | 0.0269 |

- 5.1. Estimar el modelo que se ajustó e interpretar los coeficientes de regresión estimados.
- 5.2. Realizar el test de regresión.
- 5.3. El gerente de la empresa sospecha que para las notebooks con una antigüedad determinada y con una cierta cantidad de desperfectos, los años de experiencia del técnico es una variable muy relevante en la explicación de la duración promedio del service. Según el modelo ajustado, ¿Las sospechas del gerente son ciertas?
- 5.4. Construir un intervalo de confianza para el coeficiente de regresión que mide el cambio en la duración promedio del service a medida que la cantidad de desperfectos aumenta en una unidad, mientras la antigüedad de la notebook y los años de experiencia del técnico permanecen constantes en el modelo. Interpretar.
- 5.5. ¿La “antigüedad de la notebook” aporta significativamente en la explicación de la duración promedio del service cuando las otras 2 variables están en el modelo? Utilizar la estadística F, obteniendo la suma de cuadrados parcial correspondiente mediante la siguiente salida:

| Análisis de la varianza | | | | | |
|-------------------------|----|-------------------|----------------------|---------|--------|
| Fuente | DF | Suma de cuadrados | Cuadrado de la media | F-Valor | Pr > F |
| Modelo | 2 | 16193 | 8096.61998 | 390.81 | <.0001 |
| Error | 15 | 310.76004 | 20.71734 | | |
| Total corregido | 17 | 16504 | | | |

| Estimadores del parámetro | | | | | |
|---------------------------|----|-------------------------|----------------|---------|---------|
| Variable | DF | Estimador del parámetro | Error estándar | Valor t | Pr > t |
| Intercept | 1 | -4.31616 | 3.81144 | -1.13 | 0.2752 |
| X1 | 1 | 14.70977 | 0.52883 | 27.82 | <.0001 |
| X3 | 1 | 0.53055 | 0.74047 | 0.72 | 0.4847 |

- 5.6. Se necesita predecir la duración del próximo service de una notebook comprada hace 19 meses con 4 desperfectos. El service lo realizará el técnico Juan Cruz que cuenta con 2.5 años de experiencia.
- 5.7. Utilizando los datos que se encuentran en el archivo “[Datos Ejercicio_05.xls](#)” realizar el análisis de residuos.

6. En una industria se evalúa la relación de las **ventas anuales de una empresa** (en miles de pesos) con las siguientes variables:

X₁: **inversión en publicidad** (en miles de pesos)

X₂: **volumen de producción no defectuosa** (en miles de unidades)

X₃: **costos de producción** (en miles de pesos)

Para ello, se registran los valores de estas variables durante los 12 últimos años y se estima el siguiente modelo:

$$\hat{y} = 186 + 5 X_1 + 10 X_2 - 8 X_3$$

$$R^2 = 0.876$$

$$s(b_1) = 0.3 \quad s(b_2) = 4.7 \quad s(b_3) = 0.2$$

- 6.1. Interpretar los coeficientes en término del problema.

- 6.2. Obtener un intervalo de confianza del 95% para el coeficiente de regresión correspondiente a X_3 .
- 6.3. ¿Las tres variables en forma conjunta contribuyen en forma significativa a la explicación de las ventas anuales promedio?
- 6.4. ¿Los costos de producción contribuyen en la explicación de las ventas anuales promedio cuando se considera la inversión en publicidad y el volumen de la producción no defectuosa?
- 6.5. ¿La inversión en publicidad contribuye en la explicación de las ventas anuales promedio cuando se considera el volumen de la producción no defectuosa y los costos de producción?
- 6.6. ¿El volumen de la producción no defectuosa contribuye en la explicación de las ventas anuales promedio, cuando se considera la inversión en publicidad y los costos de producción?
- 6.7. Para el siguiente año se espera que la inversión en publicidad sea de \$5500, el volumen de producción no defectuosa sea de 13600 unidades y que los costos de producción se encuentren en \$8.700. ¿A cuánto ascenderán las ventas promedio en el próximo año?

7. Una ONG de defensa al consumidor desea estimar la **emisión promedio de monóxido de carbono (CO)** de los cigarrillos de una marca en particular a partir del **peso** de los mismos y de los **niveles de alquitrán y nicotina**. Para ello, realizó un estudio en donde incluyó 24 cigarrillos de esa marca y en primer lugar midió el peso de cada uno. Luego, utilizando un método estándar, una máquina construida para tal fin “fuma” el cigarrillo hasta un punto cercano a la punta del filtro y mide los niveles de monóxido de carbono (mg/cig), alquitrán (mg/cig) y nicotina (mg/cig) contenidos en el humo del cigarrillo.

Datos en archivo Excel: Datos Ejercicio_07.xls

Sintaxis para este ejercicio: Ejercicio_07.r

- 7.1. Realizar el gráfico de dispersión matricial. ¿Qué observa?
- 7.2. Plantear un modelo de regresión lineal múltiple que estime la emisión promedio de monóxido de carbono de los cigarrillos a partir del peso y de los niveles de alquitrán y nicotina.
- 7.3. Estimar el modelo planteado e interpretar los coeficientes.
- 7.4. Evaluar si las tres variables explicativas consideradas contribuyen en la explicación de la emisión promedio de monóxido de carbono de los cigarrillos.
- 7.5. Realizar los test parciales con la estadística T. ¿Qué observa?
- 7.6. Evaluar la multicolinealidad. ¿Qué sugieren los resultados?
- 7.7. Plantear y estimar el modelo sin considerar la variable “nivel de alquitrán”.
- 7.8. Realizar el análisis de residuos para este último modelo.

8. En un estudio de control de calidad, un ingeniero mecánico desea conocer cómo influye la velocidad máxima que adquiere el motor de una herramienta en la vida útil de la misma y si esa relación difiere según el tipo de herramienta. Para ello, se obtiene la siguiente información de 20 máquinas:

Y: **vida útil de la herramienta (en años)**

X₁: **velocidad máxima que adquiere el motor de la herramienta (en rpm: revoluciones por minuto)**

X₂: **tipo de herramienta.** (A o B).

Datos: Datos Ejercicio_08.csv

Sintaxis para R: Ejercicio_08.R

8.1 Realice un gráfico de dispersión entre la velocidad máxima del motor y la vida útil de la herramienta. Realice nuevamente el gráfico según tipo de herramienta. ¿Qué observa?

8.2 Plantee un modelo de regresión lineal aditivo que relacione la vida útil de la herramienta con la velocidad del motor y que tenga en cuenta el tipo de herramienta. ¿Cómo debe ingresarse la variable tipo de herramienta al modelo si se considera al tipo A de máquina como el de referencia? Escriba el modelo para cada tipo de herramienta. También plantee el modelo en forma matricial.

8.3 Realice el test de regresión.

8.4 ¿El tipo de herramienta aporta significativamente en la explicación de la vida útil promedio de la herramienta cuando la velocidad máxima del motor se considera en el modelo?

8.5 ¿La velocidad máxima del motor aporta significativamente en la explicación de la vida útil promedio de la herramienta cuando el tipo de herramienta se considera en el modelo?

8.6 Estime los parámetros del modelo resultante e interprete.

8.7 Al volver a inspeccionar el gráfico de dispersión un ayudante del ingeniero supone que las rectas de regresión para cada tipo de herramienta podrían diferir también en la pendiente. Para evaluar esta suposición plantee un modelo de regresión lineal no aditivo. Realice el test para evaluar coincidencia, si rechaza esta hipótesis realice el test de paralelismo.

8.8 Según todo lo analizado concluya.

9. Una investigación realizada en Inglaterra consistió en comparar 3 variedades de trigo (SL, LAN y OST) con el objeto de conocer si había relación entre **el rendimiento** de las mismas (**en libras por parcela**) y **tiempo transcurridos hasta la aparición de las primeras espigas (en días)**. Se sospecha que cuánto más rápido salen las espigas el rendimiento de la parcela es superior. Para ello, se analizaron 22 parcelas.

Datos: Datos Ejercicio_09.csv

Sintaxis para R: Ejercicio_09.R

9.1 Postule un único modelo para probar si el rendimiento promedio es el mismo en las tres poblaciones considerando a la variedad OST como la categoría de referencia. Escriba la matriz X y la ecuación para cada grupo.

9.2 Pruebe si las tres rectas son paralelas. Use las sumas de cuadrados secuenciales. ¿Cómo queda el modelo según lo concluido en este punto?

9.3 Si no rechaza la hipótesis anterior pruebe coincidencia. Pruébela por modelos reducidos, indicando cuáles es el modelo completo y el modelo reducido.

9.4 Escriba el modelo ajustado para cada variedad.

10. Una pequeña empresa familiar contrató a un analista externo de recursos humanos para realizar un análisis de los sueldos de sus 20 empleados (sin importar la jerarquía de los mismos) con el objetivo de actualizar las escalas salariales. Para ello, el analista recabó información diversa sobre cada empleado. Las variables que se analizan son:

Y: **último sueldo percibido del empleado (en pesos)**

X: antigüedad del empleado en la compañía (en años).

Datos: Datos Ejercicio_10.csv

Sintaxis para R: Ejercicio_10.R

- 10.1 Realice un gráfico de dispersión entre la antigüedad del empleado en la compañía y el último sueldo percibido. ¿Qué observa?
- 10.2 Postule un modelo de regresión lineal y gráfiquelo junto a los datos. ¿Qué observa? ¿Resulta adecuado postular este modelo? Realice el análisis de residuos
- 10.3 Postule y estime un modelo de regresión por tramos (piecewise) con el punto de quiebre en 10 años.
- 10.4 ¿Cómo queda el modelo estimado en cada tramo? Interprete.
- 10.5 Obtenga el valor estimado del sueldo promedio de un empleado que tiene 10 años de antigüedad con los dos modelos definidos en cada tramo. ¿Qué observa?

11. Se quiere determinar la relación entre los componentes de la mezcla del cemento y la cantidad de calor emanado por el cemento durante el secado.

Para ello, se seleccionó una muestra de 13 mezclas y se registraron los siguientes datos:

Y: **calor emanado en una mezcla de cemento (en kcal)**

X₁: **cantidad de la sustancia A (aluminato tricálcico)**

X₂: **cantidad de la sustancia B (silicato tricálcico)**

X₃: **cantidad de la sustancia C (aluminoferrito tetracálcico)**

X₄: **cantidad de la sustancia D (silicato dicálcico)**

Datos: cement (paquete MASS)

Sintaxis para R: Ejercicio_11b.R

Se desea realizar una selección de variables para este estudio mediante diferentes métodos:

- 11.1 todas las regresiones posibles.
- 11.2 selección desde adelante (Fin= 5%).
- 11.3 selección desde atrás (Fout= 5%).
- 11.4 selección paso a paso (Fin= 5% Fout= 10%).
- 11.5 método de selección paso a paso (Fin= 1% Fout= 10%). ¿Qué puede ocurrir cuando se elige otro nivel de significación?

12. Se desea encontrar el mejor modelo que explique la calidad de un determinado vino fabricado en una bodega boutique de la ciudad de Cafayate. Para ello, se tomaron 52 muestras del vino de diferentes vasijas de roble. Para cada una de ellas, un enólogo experto determinó **la calidad del vino** en escala de 1 a 100 (Y) y otro enólogo determinó las siguientes 6 características:

X₁: **aroma** (escala de 1 a 100)

X₂: **cuerpo** (escala de 1 a 100)

X₃: **claridad** (escala de 1 a 100)

X₄: **sabor** (escala de 1 a 100)

X₅: **fuerza** (escala de 1 a 100)

X₆: **color** (escala de 1 a 100)

Aplique el método de selección paso a paso utilizando la información de la siguiente tabla:

| Variables incluidas en el modelo | SCRm | SCE |
|-------------------------------------|---------|----------|
| x1 | 0.72 | 12312.00 |
| x2 | 1055.78 | 11257.00 |
| x3 | 2622.56 | 9689.95 |
| x4 | 88.54 | 12224.00 |
| x5 | 4236.61 | 8075.90 |
| x6 | 3065.98 | 9246.53 |
| x1 x2 | 1057.19 | 11255.00 |
| x1 x3 | 2635.97 | 9676.55 |
| x1 x4 | 88.73 | 12224.00 |
| x1 x5 | 4555.85 | 7756.66 |
| x1 x6 | 3065.98 | 9246.53 |
| x2 x3 | 3045.25 | 9267.16 |
| x2 x4 | 1123.24 | 11189.00 |
| x2 x5 | 5360.29 | 6952.23 |
| x2 x6 | 3278.81 | 9033.71 |
| x3 x4 | 2916.45 | 9396.06 |
| x3 x5 | 4677.36 | 7635.15 |
| x3 x6 | 4223.99 | 8088.52 |
| x4 x5 | 4706.04 | 7606.47 |
| x4 x6 | 3066.00 | 9246.51 |
| x5 x6 | 5325.49 | 6987.02 |
| x1 x2 x3 | 3058.42 | 9254.09 |
| x1 x2 x4 | 1123.97 | 11189.00 |
| x1 x2 x5 | 5670.97 | 6641.54 |
| x1 x2 x6 | 3278.86 | 9033.65 |
| x1 x3 x4 | 2941.76 | 9370.75 |
| x1 x3 x5 | 4853.59 | 7458.92 |
| x1 x3 x6 | 4228.67 | 8083.85 |
| x1 x4 x5 | 5104.12 | 7208.39 |
| x1 x4 x6 | 3066.01 | 9246.51 |
| x1 x5 x6 | 5538.72 | 6773.79 |
| x2 x3 x4 | 3278.60 | 9033.92 |
| x2 x3 x5 | 5495.90 | 6816.61 |
| x2 x3 x6 | 4319.84 | 7992.77 |
| x2 x4 x5 | 5904.74 | 6407.77 |
| x2 x4 x6 | 3279.02 | 9033.50 |
| x2 x5 x6 | 5847.81 | 6664.70 |
| x3 x4 x5 | 5588.69 | 6723.82 |
| x3 x4 x6 | 4518.93 | 7793.58 |
| x3 x5 x6 | 5546.75 | 6765.77 |
| x4 x5 x6 | 5791.53 | 6520.99 |
| x1 x2 x3 x4 | 3302.08 | 9010.44 |
| x1 x2 x3 x5 | 5727.97 | 6584.54 |
| x1 x2 x3 x6 | 4324.89 | 7987.62 |
| x1 x2 x4 x5 | 6299.10 | 6013.41 |
| x1 x2 x4 x6 | 3279.06 | 9033.45 |
| x1 x2 x5 x6 | 6083.99 | 6228.52 |
| x1 x3 x4 x5 | 5778.23 | 6534.28 |
| x1 x3 x4 x6 | 4531.40 | 7781.11 |
| x1 x3 x5 x6 | 5680.95 | 6631.57 |
| x1 x4 x5 x6 | 6069.84 | 6242.67 |
| x2 x3 x4 x5 | 6321.28 | 5991.23 |
| x2 x3 x4 x6 | 4585.19 | 7727.32 |
| x2 x3 x5 x6 | 5939.84 | 6372.68 |
| x2 x4 x5 x6 | 6369.82 | 5942.69 |
| x3 x4 x5 x6 | 6240.10 | 5972.42 |
| x1 x2 x3 x4 x5 | 6564.71 | 5747.81 |
| x1 x2 x3 x4 x6 | 4597.71 | 7714.80 |
| x1 x2 x3 x5 x6 | 6121.38 | 6191.14 |

| | | |
|-------------------|---------|---------|
| x1 x2 x4 x5 x6 | 6488.03 | 5824.48 |
| x1 x3 x4 x5 x6 | 6488.03 | 5824.48 |
| x2 x3 x4 x5 x6 | 6700.44 | 5612.08 |
| x1 x2 x3 x4 x5 x6 | 6895.46 | 5717.06 |

13. Se desarrolló un estudio que estima el personal necesario para realizar instalaciones eléctricas en edificios estatales considerando 7 variables explicativas. La variable respuesta se definió como “**horas hombre mensuales**”.

Datos: Datos Ejercicio_13.csv

Sintaxis para R: Ejercicio_13.R

- 13.1 Ajuste el modelo con todas las variables y realice el análisis de los residuos.
 - 13.2 Evaluar si hay observaciones influyentes.
 - 13.3 Con el mismo conjunto de datos se desea ajustar otro modelo. Utilice la técnica paso a paso y estime el modelo resultante ($F_{in}=5\%$, $F_{out}=10\%$).
 - 13.4 Evalúe la existencia de multicolinealidad. En caso de existir proponga como solución un nuevo modelo.
 - 13.5 Realice el análisis de residuos con este nuevo modelo.
 - 13.6 Considerando el nuevo modelo ajustado ¿identifica observaciones atípicas respecto a la variable respuesta?
 - 13.7 Considerando el nuevo modelo ajustado ¿identifica observaciones atípicas respecto a conjunto de variables explicativas?
- En el caso de haber identificado observaciones atípicas, estudie la influencia de las mismas.