

Relación 2

Ejercicio 1

Los datos del fichero Ejercicio 1.sf3 son relativos a diferentes variables de coches.

- a. **Ajustar un modelo de regresión múltiple, siendo la variable respuesta *millas por galón* (inversa del consumo) y las variables regresoras: *cilindros*, *caballos*, *peso* y *precio*.**

- b. **¿Son todas las variables significativas (prueba T)?**

Como todos los Valores-P de las variables independientes son menores que α (0,05), existen evidencias suficientes como para **rechazar H_0** , resultando representativas.

A su vez, como el Valor-P de la variable dependiente (respuesta) también es menor que α , existen evidencias suficientes como para **rechazar H_0** , lo que se traduce en que el modelo tiene un término constante distinto de 0.

- c. **¿Cuál es el coeficiente de determinación?**

~~En StatGraphics, el coeficiente de determinación aparece como «R cuadrada», en este caso, el coeficiente de determinación es 0,758682 ó un 75,8682 %.~~

Es R-Cuadrado en este caso.

- d. **¿Qué conclusiones se btienen de la tabla ANOVA (contraste conjunto F)?**

Como el Valor-P = 0 < α = 0,05 (por defecto), existen evidencias suficientes como para **rechazar H_0** .

Por tanto, **existe un a_i** significativo.

- e. **Analizar los residuos utilizando un estudio descriptivo de los mismos y los gráficos de residuos. ¿Se verifican las hipótesis del modelo (homocedasticidad y normalidad)?**

Guardar los residuos > Análisis de 1 variable

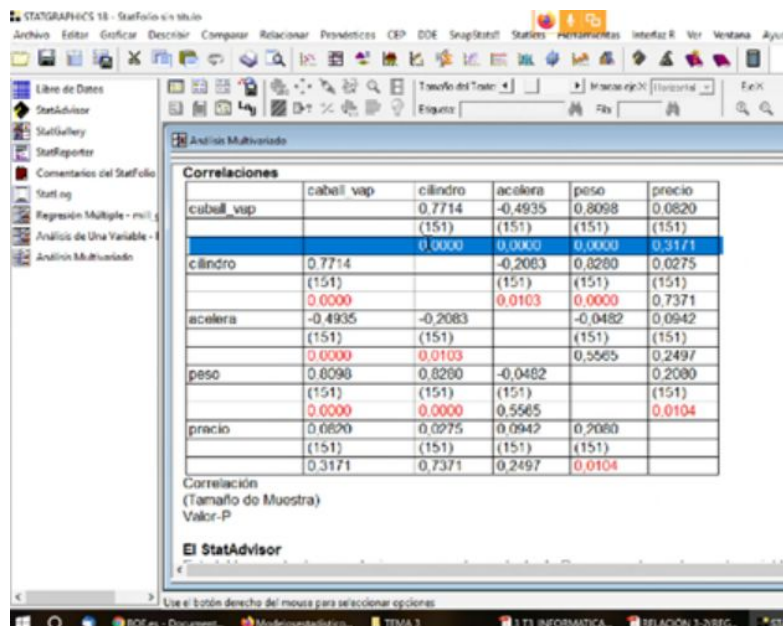
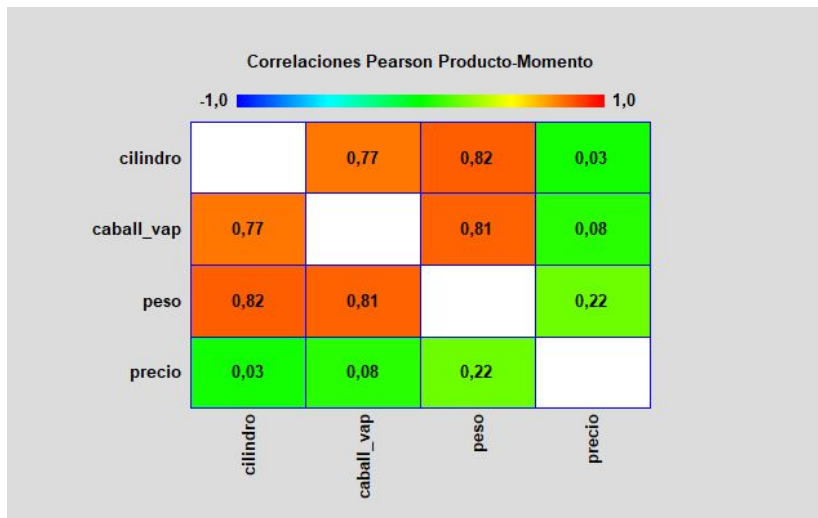
Como P: 0,0705 > 0,05 sigue una distribución y se rechaza H_0 .

¿Mejorará el modelo si se introduce la variable *aceleración*?

Añadir Aceleración al modelo multivariado >
Ahora se usa el modelo ajustado (R_Cuadrado) >
Se añade a la «pizarra» de antes una a más

Resulta no significativa la variable introducida y el modelo, además caballos se vuelve no-significativa. Se complica el modelo para nada.

f. Analizar la hipótesis multicolineal.



Esta matriz es la misma que la coloreada, pero con más datos (saber H_0).

Ejercicio 2

Se quiere ajustar un modelo que permita estimar los gastos en alimentación de una familia en base a la información proporcionada los ingresos mensuales y el número de miembros de la familia. Para ello se recoge una muestra aleatoria simple de tamaño 15 cuyos resultados son los que están recogidos en el fichero Ejercicio 2.sf3.

Ajuste el modelo.

Puede observarse una tendencia al alza, similar a una recta, por lo que la regresión lineal sería correcta.

Con el valor de R-Cuadrada se puede explicar un 94,964 % de la variabilidad en Gasto.

Con el valor de R-Cuadrado puede observarse que se ajusta bastante bien a una recta (94,1246 %), por lo que la ecuación del modelo ajustado es:

$$\text{GASTO} = -0,160458 + 0,148727 \cdot \text{INGRESO} + 0,0769152 \cdot \text{TAMAÑO}$$

Ejercicio 3

El fichero Ejercicio 3.sf3 contiene datos relativos a partidos de la liga ACB de baloncesto. Los datos son de 62 jugadores al azar del total y han sido obtenidos de la Base de Datos Histórica de la Liga 1989 - 1990. En base a esta muestra se desea estudiar si existe una relación lineal entre la variable *puntos por partido* (*punt x part*) que es capaz de anotar un jugador de baloncesto respecto a las siguientes variables: La *altura del jugador* (*altura*), los *minutos que juega el partido* (*min x part*), los *balones que pierde por partido* (*bp x part*), las *faltas personales cometidas* (*fp x part*), el *porcentaje en tiros de campo por partido* (*por_tc*).

En base a estos datos:

1. Ajustar un modelo de regresión sin excluir ninguna variable e interpretar los resultados, ¿es un buen ajuste?

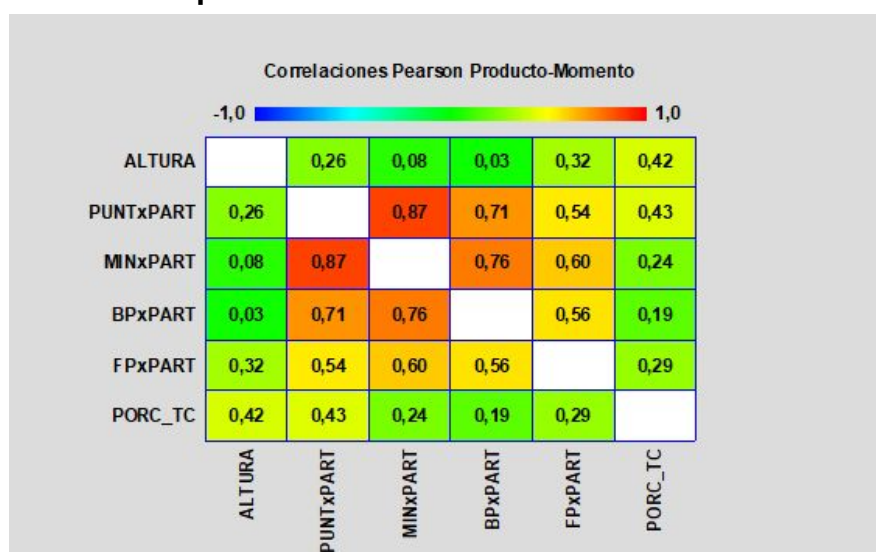
Se rechaza H_0 porque al menos hay un parámetro $a_1 \dots a_5$ que es distinto de 0.

Por tanto, el ANOVA es significativo.

Las variables BP x PART y FP x PART no son significativas (mayores a 0,05) por tanto podrían eliminarse.

Sí, es un buen ajuste. Si vemos el valor de R-cuadrada, con nuestro modelo se explica un 83,3646 % de la variabilidad usando las 5 variables, que es un porcentaje muy decente.

2. Analizar la hipótesis de multinealidad.



Estos pares son los que cumplen que su P-Valor < 0,05, por lo que tienen correlaciones significativamente diferentes de cero.

- ALTURA y PUNTxPART
- ALTURA y MINxPART
- ALTURA y FPxPART
- PUNTxPART y MINxPART
- PUNTxPART y PORC_TC
- MINxPART y FPxPART
- BPxPART y FPxPART

3. En el modelo ajustado, ¿cuáles son las observaciones atípicas / influyentes?

<i>Influencia</i>	<i>Distancia de Mahalanobis</i>	<i>DFITS</i>
0,118223	7,0608	1,00849
0,16857	11,1812	1,23638
0,129458	7,939	-0,708305
0,319522	27,1897	-0,243992
0,145626	9,24323	-0,729793
0,20162	14,1686	0,628084
0,0714904	3,63608	-0,677797