**Estadística I, 2024. Ejercicio de correlación y regresión, parte B**

Tenemos un conjunto de datos (<https://www.uv.es/mperea/PRIM_SE.jasp>), que contiene información de las notas en el primer cuatrimestre de un grupo de estudiantes en una carrera de ciencias, así como información de las notas promedio en instituto de matemáticas, ciencias y lengua. También tenemos las notas de selectivo (PAU) de matemáticas y lengua, así como el sexo y si llevan gafas.

La idea (global) es ver cuán predictivas son las notas del instituto cuando el alumnado llega a la universidad en su primer cuatrimestre y ha de examinarse.

Correlación

*--¿Qué relación aprecias (signo y fuerza) entre las notas de primer cuatrimestre (Y) y las notas en selectivo de matemáticas (X)? Efectúa el gráfico adecuado y calcula el índice adecuado.*

*--Pensemos que tenemos una matriz de correlaciones con las siguientes variables: notas promedio de instituto en matemática, notas promedio de instituto en lengua, notas de la PAU en matemáticas y notas de la PAU en lengua. ¿En qué caso la relación es más fuerte? ¿Y más débil? ¿Pensáis que hubiera sido posible encontrar relaciones negativas entre estas variables?*

*--Queremos saber si la proporción de personas con gafas es mayor en chicas que en chicos (o en otras palabras, si hay relación entre sexo y el llevar o no gafas). Efectúa la tabla de contingencia y calcula el índice apropiado.*

Regresión simple

*--¿Cuál es la ecuación de la recta de regresión de las notas de primer cuatrimestre (Y) y a partir de las notas de la PAU en matemáticas (X) (en puntuaciones directas)? ¿Qué quiere decir la ordenada en el origen y la pendiente en este caso?*

*--Idem en puntuaciones típicas. ¿El valor de la pendiente coincide necesariamente con…?*

*--¿Qué porcentaje de varianza de las notas de primer cuatrimestre (Y) explica la recta de regresión?*

*--¿Qué puntuación en notas de primer cuatrimestre predirías a la persona de la primera fila del archivo con la recta de regresión?*

Regresión múltiple

*--¿Cuál es la ecuación de regresión de las notas de primer cuatrimestre (Y) y a partir de los predictores nota de la PAU en matemáticas y nota de la PAU en lengua? ¿Qué quieren decir los componentes de la ecuación de regresión en puntuaciones directas?*

*--¿Qué porcentaje de varianza de las notas de primer cuatrimestre explica la ecuación de regresión de la pregunta anterior?*

*--¿Cuál es el mejor predictor de la ecuación anterior y por qué?*

*--¿Crees que ha habido problemas de colinealidad entre los predictores? (recordad: mirar el VIF [Variance Inflation Factor], en el caso ideal sería 1.0)*

*--Pensemos que hiciéramos la regresión “stepwise” (paso a paso). ¿Qué ecuación resultante tendríamos? ¿Por qué?*