


Guion Práctica 3. Análisis de datos y regresión Estadística - Grado en Ingeniería Informática

Ejercicio 1. Descarga del Campus Virtual el archivo `DatosGapminder.csv`. Este archivo ha sido obtenido de la web <http://www.gapminder.org/>. Importa los datos a  y comprueba que todo ha funcionado correctamente con el comando `head`. Se pide:

- a) Resumen gráfico de la variable PCs.
- b) Varianza y desviación típica de la variable PCs.
- c) Cuantiles de orden 0.3, 0.6 y 0.9 de la variable PCs.
- d) Tipificar la variable PCs. Comprobar que la variable tipificada tiene media 0 y varianza 1.
- e) Realizar el mismo análisis descrito en los pasos anteriores para la variable GDP.

Ejercicio 2. Estudia la dependencia de la variable GDP sobre la variable PCs, es decir, ¿podríamos explicar el comportamiento de la variable GDP sobre la variable PCs?

- a) Representa un diagrama de dispersión de GDP sobre PCs.
- b) Estudia la relación entre el GDP y los PCs mediante la covarianza y el coeficiente de correlación.
- c) Ajusta la recta de regresión de la variable del GDP sobre PCs. Representa el modelo ajustado en el diagrama de dispersión obtenido en el apartado a).
- d) Obtén el coeficiente de determinación e interpreta el valor obtenido.
- e) Representa los residuos del ajuste frente a la variable explicativa. ¿Crees que la variabilidad de los residuos presenta una varianza que no depende de la variable explicativa?
- f) Calcula predicciones para la variable GDP sabiendo que la variable PCs toma los valores $x_0 = 5$, $x_1 = 25$ y $x_2 = 75$.
- g) Calcula y representa de nuevo la recta de regresión de GDP sobre PCs, pero sólo teniendo en cuenta los datos de los países de la Unión Europea. Para facilitar la tarea de selección, a continuación se ofrece una lista que los contiene:

```
> UE=c("Austria", "Belgium", "Bulgaria", "Croatia", "Republic of Cyprus",  
+ "Czech Republic", "Denmark", "Estonia", "Finland", "France", "Germany",  
+ "Greece", "Hungary", "Ireland", "Italy", "Latvia", "Lithuania",  
+ "Luxembourg", "Malta", "Netherlands", "Poland", "Portugal", "Romania",  
+ "Slovakia", "Slovenia", "Spain", "Sweden")
```

Ejercicio 3. Opcional

En este ejercicio volveremos a trabajar con los datos del geyser *Old Faithful* que están guardados en la variable `geyser` de la librería `MASS`. Recuerda que tienes que cargar dicha librería mediante el comando `library(MASS)`.

En este estudio la variable respuesta es el tiempo de espera entre erupciones (`waiting`) y la variable explicativa el tiempo de duración de la erupción (`duration`)

- a) Representa un diagrama de dispersión de ambas variables.
- b) Estudia la relación lineal entre las dos variables mediante la covarianza y el coeficiente de correlación.
- c) Ajusta la recta de regresión de Y sobre X . Representa el modelo ajustado en el diagrama de dispersión obtenido en el apartado a).
- d) Obtén el coeficiente de determinación e interpreta el valor obtenido.
- e) Representa los residuos del ajuste frente a la variable explicativa. ¿Crees que la variabilidad de los residuos presenta una varianza que no depende de la variable explicativa?
- f) Calcula las predicciones para la variable Y sabiendo que la variable X vale $x_0 = 4.3$ y $x_1 = 2.3$. ¿Sería razonable hacer una predicción para $x_2 = 7$?