

Práctica 5.- Recta de regresión

1. En el archivo 'datos.dat' se encuentran las alturas y pesos de 1000 personas. Se pide:
 - a) Construir un diagrama de dispersión o nube de puntos de las variables altura y peso.
 - b) Calcular la recta de regresión de peso respecto a altura.
 - c) Calcular el coeficiente de correlación r y el coeficiente de determinación R^2 y, explica que información se deduce de ellos. ¿Es bueno el ajuste dado por dicha recta?
 - d) Usar la recta de regresión del apartado anterior para predecir el peso de una persona con altura 1.8.
 - e) Obtener los residuos y los valores ajustados por el modelo dado por la recta de regresión del apartado anterior y comprobar la relación entre las varianzas de las variables, el modelo y los residuos.
2. En el fichero 'tiempos.csv' se encuentran datos sobre el tiempo que han tardado tres algoritmos en ordenar diferentes listas de números. Para cada experimento se ha almacenado el algoritmo usado (entre 0 y 2), el máximo número de veces que se ha repetido algún elemento de la lista, la longitud de la lista y el tiempo necesitado.
 - a) Construir un diagrama de dispersión o nube de puntos de las variables Tiempo y Longitud. ¿Parece correcto intentar calcular la recta de regresión de Tiempo respecto a Longitud?
 - b) Se quiere calcular aproximadamente para cada algoritmo el orden de magnitud del tiempo necesario para ordenar una lista en función de la longitud de la misma. Para ello, para cada algoritmo crea dos nuevas variables que contengan los logaritmos de los tiempos y longitudes correspondientes. Hallar la recta de regresión de los logaritmos de los tiempos respecto a los logaritmos de las longitudes. Interpreta los resultados
 - c) Para el primer algoritmo, calcular la recta de regresión del tiempo respecto al cuadrado de las longitudes. Calcular el coeficiente de correlación r y el coeficiente de determinación R^2 y, explica que información se deduce de ellos. ¿Es bueno el ajuste dado por dicha recta?
 - d) Usar la recta de regresión del apartado anterior para predecir el tiempo que tardaría este algoritmo en ordenar dos listas de 3000 y 10000 elementos (experimentalmente se obtuvieron tiempos de 0.623 y 7.61).
 - e) Obtener los residuos y los valores ajustados por el modelo dado por la recta de regresión del apartado anterior y comprobar la relación entre las varianzas de las variables, el modelo y los residuos.
 - f) Repite el estudio del apartado anterior para el tercer algoritmo