

## Probabilidad y Estadística (C)

### Clases Prácticas 24: Tests de hipótesis (iii).

1. Este ejercicio se encuentra resuelto en el libro *An introduction to statistical learning*, de James, G., Witten, D., Hastie, T. y Tibshirani, R. (2013).

El conjunto de datos *Advertising.csv* contiene las ventas de un producto en 200 mercados, junto con el presupuesto publicitario en TV.

- (a) Haga un diagrama de dispersión. ¿Hay relación entre el presupuesto en publicidad y las ventas? En caso afirmativo, qué características tiene esta relación.
  - (b) Proponer un modelo de regresión lineal que tenga como respuesta a la variable ventas y como covariable a la variable TV. Definir las variables aleatorias involucradas en el modelo. Estimar los coeficientes de la recta de regresión y la varianza de los errores.
  - (c) ¿Cuál es el error promedio que se comete al predecir las ventas de un mercado utilizando este modelo?
  - (d) Dar intervalos de confianza para la pendiente y para la ordenada al origen. Interpretar los resultados.
  - (e) ¿Encuentra evidencia en estos datos de que la inversión en publicidad influye en las ventas? Describir el test de hipótesis que permite responder esta pregunta. ¿Cómo se relaciona con los intervalos de confianza del ítem anterior?
  - (f) ¿Qué proporción de la variabilidad en las ventas queda explicada por la inversión en publicidad?
  - (g) Si en un mercado se invierten 250000 pesos en publicidad, ¿cómo puede predecir las ventas en ese mercado?
  - (h) En el conjunto de datos *Advertising2.txt* se han agregado datos sobre 10 nuevos mercados. Estimar los coeficientes y analizar el ajuste. ¿Qué observa? ¿Cómo puede solucionarse?
2. Utilizando el conjunto de datos *Auto* de la librería *ISLR* de R, realizar la regresión lineal simple respondiendo las siguientes preguntas.
    - (a) Plantear un modelo de regresión lineal simple que permita explicar la variable *mpg* (milles per gallon) como respuesta de *horsepower*.
      - i. ¿Hay una relación entre estas variables?
      - ii. ¿Cuán fuerte es esta relación? ¿Es positiva o negativa?
      - iii. ¿Cuál es el valor predicho para *mpg* asociado con una potencia de 98?
    - (b) Realizar un gráfico de dispersión para las variables y en el mismo graficar la recta de regresión ajustada. ¿Nota algún inconveniente en el gráfico? ¿Cómo podría resolverlo?