Assignment 5:
Simple linear regression

Due in class of week 8.

1. Keynes versus Buffett

This problem compares the investment records of two famous figures in finance: John Maynard Keynes and Warren Buffett. The keynes.txt data records the annual returns for the portfolio Keynes managed for King’s College of Cambridge University in the 1930s. The buffett.txt data records the performance of Berkshire Hathaway from 1965 to 2012.

(a) The market model assumes that annual returns (in percent terms) of a portfolio are linearly related to returns on the market, i.e. \( Y = a + Xb + \epsilon \), where \( Y \) is the portfolio return, \( X \) is the market return for the same period, and \( \epsilon \) is an independent noise random variable representing random fluctuations around the linear trend; that is, \( E(\epsilon) = 0 \). For each data set, determine values of \( a \) and \( b \) that minimize \( E[(a + Xb - Y)^2] \) where the expectation is taken with respect to the empirical distributions defined by the provided Keynes and Buffett datasets. That is, find the best linear predictor of \( Y \) (portfolio return) using \( X \) (market return). Who is the better investor based on these findings?

(b) Find the best linear predictor using the \( \text{lm}() \) command in R. The estimated coefficients should be similar to (but not quite exactly the same as) the best linear predictor you computed above. That is, run a linear regression of portfolio return on market return for the Keynes and Buffett data sets. You can report your results using the output of the \( \text{summary}() \) command. Make scatter plots with overlaid fitted lines to visualize your work.

(c) What would you predict for the Keynes and Buffett portfolio returns when the market is up 10% versus when it is down 10%? Provide 95% prediction intervals for your point predictions. Which portfolio would you prefer and why?

(d) Let \( \mu_m \) denote the mean market return, meaning that \( E(X) = \mu_m \) where \( X \) is the random variable representing market return. Express the expected value of the Keynes and Buffett portfolio returns in terms of \( \mu_m \), using the relation \( Y = a + Xb \). In light of this calculation, how can we interpret the roll of \( a \) in the market model?

(e) Let \( \sigma_m \) denote the standard deviation of the market returns, meaning that \( \text{V}(X) = \sigma_m^2 \) where \( X \) is the random variable representing market return. Express the variance of Keynes and Buffett’s portfolio returns in terms of \( \sigma_m \) using the relation \( Y = a + Xb \). In light of this calculation, how can we interpret the roll of \( b \) in the market model?
2. **Moneyball lite**

We run a regression trying to understand the impact of OBP (on base percentage) on RPG (runs per game). The results for each team in the American League are shown below.

|               | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | -8.247   | 1.921      | -4.294  | 0.00104  ** |
| OBP[League == "American"] | 38.817   | 5.506      | 7.050   | 1.34e-05 *** |

Residual standard error: 0.2099 on 12 degrees of freedom. R-squared: 0.8055.

(a) Verify the numbers in this table in R, using the `baseball.csv` data from the class web site. Use the `subset` option in the `lm()` command to restrict the analysis to the American League teams.

(b) Give an approximate 90% confidence interval for the slope.

(c) Test the hypothesis that OBP is unrelated to RPG at the 10% level.

(d) What is the correlation between RPG and OBP?

(e) Give the plug-in 95% prediction interval for RPG when OBP is 0.400.

(f) Test the hypothesis that the true slope is exactly 30 at the 5% level. Count as extreme only values greater than the null hypothesis value of 30.