

Project #3: Simple linear regression. CAPM.

Milica Cudina

2022-10-03

Before you start working on this project, watch the following two videos about simple linear regression courtesy of OpenCourse Statistics:

[Ideas of fitting a line](#)

[Fitting a least-squares regression line](#)

Next, keep this tutorial on *simple linear regression* in R handy

[Simple Linear Regression in R](#)

Problem #1 (20 points)

Download the data set with information about the cricket chirps per second and the temperature available in Canvas.

(5 points) What plot would you provide to visualize the association between the frequency of cricket chirps and the temperature? Plot it and comment on at least **three** relevant features of the data.

(2 points) If you were to provide a single value quantifying the *strength* of the relationship, what would it be? Calculate that value.

(5 points) Which trend curve (if any) would you use for the association you visualized?

(3 points) Imagine that you are in the situation where someone tells you they hear 18 chirps per second, at which temperature would you say it was?

(5 points) If you had been asked to provide the same temperature “prediction” for the given value of 10 chirps per second, what would you say?

Problem 2. (40 points)

The data set `mfm.csv` is available in Canvas. Download it and import it into R. Identify the names of the columns.

The immediately relevant columns are `mktrf`, `rf`, and `wmt`. They contain the daily excess return of the market, the Treasury return (a proxy for the risk-free interest rate), and the daily returns of *Walmart*, respectively.

(10 points) In the context of CAPM, provide an appropriate visualization of the relationship between excess returns of the market and the excess returns of *Walmart*. Comment on at least three features of the plot you obtain.

(20 points) Estimate the daily *beta* of *Walmart* based on the above information. More precisely, provide the entire `summary` of your model and, then, emphasize the estimated value of the *beta*.

(10 points) Go through the following tutorial on the meaning and construction of **prediction intervals** in R.

[Prediction intervals](#)

On 2022/08/31, the excess return of the market was -0.0074 , and the Treasury return was 0.00008 . By CAPM, what is the **prediction interval** of the return of *Walmart* on 2022/08/31?

Your instructor calculated the *Walmart* return for that date as 0.0005283816 . Do you have any comments/conclusions that you would like to share?

Problem 3. (20 points)

Read the following articles.

[An article on familiarity bias.](#)

[Familiarity breeds investment.](#)

[Wikipedia: Overconfidence effect.](#)

[Wikipedia: Disposition effect.](#)

[Investopedia: Herd behavior.](#)

(10 points) Do the above behaviors agree with the assumptions of CAPM? Why? Justify your answer.

(10 points) Which of the above behaviors would result in a *systematic bias*, i.e., which of the above behaviors would “push” the security prices away from their fundamental value? Substantiate your response.

Problem 5. (10 points)

Consult the following Wikipedia page:

[Fama-French](#)

Explain in a short paragraph what modifications to the CAPM model result in the Fama-French model. Which statistical tool would you have to use to estimate the parameters in this model? Look back at the data set `mf.m.csv`. Do the names of the columns look familiar now?

Problem 6. (10 points)

Read the following article:

[Beyond Multi-Factor Models](#)

Write a paragraph or two discussing the pros and cons of CAPM (and its kin, i.e., multi-factor models).