

Citadel Statistics Course

Jeffrey R. Russell

Homework 4

1. Use the data set strategies.xls for this problem. The daily returns on two different trading strategies are given under strategy1 and strategy2.
 - a. Test each series to see if they are iid. State your conclusions.
 - b. Test each series to see if they are Normally distributed. State your conclusions.
 - c. What model would you suggest using for each return series. Be very explicit here.
 - d. Do the strategies returns appear to be correlated?
 - e. Test the null hypothesis at the 5% level that the two returns have the same expected value.
 - f. What's the p-value?
 - g. Would you reject at the 1% level? Explain why or why not.
 - h. What is the smallest level at which you can reject the null?
 - i. What's the p-value?
 - j. Would you reject at the 1% level? Explain why or why not.
 - k. What is the smallest level at which you can reject the null?
2. Use the data set mfunfs.xls. Consider the "drifus" and "fidel" funds.
 - a. Test the null that each series is iid (perform two separate tests, one for each series).
 - b. Test the null that each series is Normally distributed (perform two separate tests, one for each series).
 - c. What do you conclude about the iid Normal model for the two return series. Is it a good idea based on the tests? Why or why not?
3. Consider the following simple linear regression model $Y = 2 + 4X + \varepsilon$ and $\varepsilon \sim N(0, 4)$ (treat this as the true model).
 - a. What is $E(Y|X=0)$?
 - b. What is $E(Y|X=1)$?
 - c. What is $V(Y|X)$?
4. Use the data set censuswage.xls from the data directory. This data set contains wages per hour and years of education for a sample of workers from the 1990 US census.

Run a regression of wages on education (ie put wages on the left side and education on the right side).

- a. Interpret the regression slope coefficient estimate in words. What does it tell you?
- b. What is the expected wage of a high school graduate (in their measurement this corresponds to years of education equal to 10)?

- c. Report the SST, SSE, and SSR. Calculate the R-squared of the regression using these numbers (show your work).
 - d. Use the SSE to calculate the estimate of the standard deviation of ε (show your work).
5. Use the data set capm.xls for this problem. It contains monthly observations for a set of stocks. AT&T, GE, GAP, HD (Home Depot), IBM, and GM as well as a proxy for the market portfolio returns, and a proxy for the risk free rate given by the 30-day T-Bill rate. The returns cover the period January 1982 to December 2001.
 - a. Choose your favorite stock among those in the worksheet and use the whole span of data. Construct what are called the excess returns for your stock by subtracting the risk-free rate from that for your stock. In other words, if your stock is GE construct a column of data containing $r_t^{excess GE} = r_t^{GE} - r_t^{risk\ free}$. Do the same for the market return by forming $r_t^{excess\ Market} = r_t^{Market} - r_t^{risk\ free}$. Now estimate a measure of how your stock co-moves with the market return called the beta in the Capital Asset Pricing Model (CAPM) by estimating the following regression:

$$r_t^{excess\ Your\ Stock} = \beta_0 + \beta_1 r_t^{Excess\ Market} + \varepsilon_t$$
 Report the parameter estimate for β_1 .
 - b. What do your estimates say about the relationship between excess returns on your stock and the market excess return? Explain.
 - c. If the market excess return is 1%, what is the expected return on your asset?
 - d. Find the covariance between the excess returns on your stock and the excess returns on the market. Find the variance of the excess market returns. Use these numbers to find the slope coefficient estimate b_1 (show your work).
 - e. Find the mean of both excess return series. Use this to find the estimate of the slope parameter b_0 (show your work).
6. Consider the following simple linear regression model $Y = 2 + 4X + \varepsilon$ and $\varepsilon \sim N(0, 4)$ (treat this as the true model).
 - a. Compute a 95% prediction interval for Y given X=1.
 - b. Compute a 95% prediction interval for the average of two Y values given that both have X=1.
 - c. What is the distribution of Y if X=2?
 - d. What is $\Pr(Y < 8 | X = 2)$?
 - e. What is the distribution of Y if X=2.5?
 - f. If $X_1 = 2$ and $X_2 = 2.5$ find $\Pr(Y_2 > Y_1)$
7. Use the dataset censuswage.xls. This dataset contains wages per hour and years of education for a sample of workers from the 1990 US census.

Run a regression of wages on education (ie put wages on the left side and education on the right side).

- a. Use the data to find an estimate for the covariance between b_0 and b_1 .
- b. Test the null hypothesis that the true slope parameter is zero. What do you learn from this test?
- c. My high school teacher once told our class that each additional year of education translates into one dollar per hour more in wages. Is the data consistent with this claim? Test this hypothesis.
- d. What is the p-value for this test?
- e. Build a 95% confidence interval for the true slope parameter.