Homework Assignment 2

Nick Polson
41000: Business Statistics
Booth School of Business

Due in Week 5

Problem 1: Descriptive Statistics in R

Download the superbowl1.txt and derby.csv datasets from the course web-page. The Superbowl contains data on the outcome and the spread of all previous Superbowls. The outcome is defined as the difference in scores of the favourite minus the underdog. The spread is the bookmakers’ prediction of the outcome before the game begins.

The Derby dataset consists of all of the results on the Kentucky Derby which is run on the first Saturday in May every year at Churchill Downs racetrack.

Install R and RStudio and answer the following questions:

• For the Superbowl data.
  (a) Plot the spread and outcome variables.
      Calculate means, standard deviations, covariances, correlations and alpha and beta’s.
  (b) Use a boxplot to compare the favourites’ score versus the underdog.
  (c) Does this data look normally distributed?

• For the Kentucky Derby data.
  (a) Plot a histogram of the winning speeds and times of the horses.
      Why is there a long right-hand tail to the distribution of times?
  (b) Can you identify the outlying horse with the best winning time?
Problem 2: YahooFinance Berkshire Hathaway

Using the R script on the course webpage download daily return data in Warren Buffett’s firm Berkshire Hathaway (ticker symbol: BRK-A) from 1990 to the present. Using R or RStudio analyze this data in the following way:

(a) Plot the Historical Price Performance
(b) Calculate the Daily returns. Plot a histogram of the returns. Comment on the distribution that you obtain.
(c) Use the summary command to provide statistical data summaries. Interpret your findings.

Hint: you will find the following R commands useful.

You will need to install the quantmod package

# Install the package first
install.packages("quantmod")
library(quantmod)
# get data of BRK.A
gtSymbols("BRK-A", from = "1990-01-01")
# convert data type
BRK.A = get("BRK-A")
# print out first several lines
head(BRK.A)
Problem 3: AB Testing

SimCity 5 is one of Electronic Arts (EA's) most popular video games. As EA prepared to release the new version, they released a promotional offer to drive more pre-orders. The offer was displayed on their webpage as a banner across the top of the pre-order page. They decided to test some other options to see what design or layout would drive more revenue.

The control removed the promotional offer from the page altogether. The test lead to some very surprising results. With a sample size of 1000 visitors, of the 500 which got the promotional offer they found 143 people wanted to purchase the games and of the half that got the control they found that 199 wanted to buy the new version of SimCity.

Test at the 1% level whether EA should provide a promotional offer or not
Problem 4: Hypothesis Testing

In a recent article it was claimed that “96% of Americans under the age of 50” spent more than three hours a day on Facebook.

To test this hypothesis, a survey of 418 people under the age of 50 were taken and it was found that 401 used Facebook for more than three hours a day.

Test the hypothesis at the 5% level that the claim of 96% is correct.
Problem 5: Paired T-Test

The following table shows the outcome of eight years of a ten year bet that Warren Buffett placed with Protege Partners, a New York hedge fund. Buffett claimed that a simple index fund would beat a portfolio strategy (fund-of-funds) picked by Protege over a ten year time frame. At Buffett’s shareholder meeting, he provided an update of the current state of the bet. The bundle of hedge funds picked by Protege had returned 21.9% in the eight years through 2015 and the S&P500 index fund had soared 65.7%.

<table>
<thead>
<tr>
<th></th>
<th>SP Index</th>
<th>Hedge Funds</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>-37.0%</td>
<td>-23.9%</td>
</tr>
<tr>
<td>2009</td>
<td>26.6%</td>
<td>15.9%</td>
</tr>
<tr>
<td>2010</td>
<td>15.1%</td>
<td>8.5%</td>
</tr>
<tr>
<td>2011</td>
<td>2.1%</td>
<td>-1.9%</td>
</tr>
<tr>
<td>2012</td>
<td>16.0%</td>
<td>6.5%</td>
</tr>
<tr>
<td>2013</td>
<td>32.3%</td>
<td>11.8%</td>
</tr>
<tr>
<td>2014</td>
<td>13.6%</td>
<td>5.6%</td>
</tr>
<tr>
<td>2015</td>
<td>1.4%</td>
<td>1.7%</td>
</tr>
<tr>
<td>cumulative</td>
<td>65.7%</td>
<td>21.9%</td>
</tr>
</tbody>
</table>

(a) Use a paired $t$-test to assess the statistical significance between the two return strategies

(b) How likely is Buffett to win his bet in two years?

[Aside: the outcome of the bet can be found here: https://www.aei.org/publication/warren-buffett-wins-1m-bet-made-a-decade-ago-that-the-sp-500-stock-index]