<div align="center">
University of Chicago
Booth School of Business
41000: Business Statistics, Autumn 2021: Homework Assignment 2. Due in Week 5
</div>

Problem 1: Bayes Gold and Silver Coins

A chest has two drawers. It is known that one drawer has 3 gold coins and no silver coins. The other drawer is known to contain 1 gold coin and 2 silver coins.

You don't know which drawer is which. You randomly select a drawer and without looking inside you pull out a coin. It is gold.

(a) Show that the probability that the remaining two coins in the drawer are gold is 75%.

Problem 2: The Monty Hall Problem.

This problem is named after the host of the long-running TV show, *Let's Make a Deal.*

A contestant is given a choice of 3 doors. There is a prize (a car, say) behind one of the doors and something worthless behind the other two doors (say two goats).

After the contestant chooses a door Monty opens one of the other two doors, revealing a goat.

  (a) The contestant has the choice of switching doors. Is it advantageous to switch doors or not?

There is a clear discuss about Monty Hall problem on Wikipedia

Problem 3: Descriptive Statistics in R

Download the `superbowl1.txt` and `derby.csv` datasets from the course web-page. The Superbowl contains data 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.

    1. Plot the `spread` and `outcome` variables. Calculate means, standard deviations, covariances, correlations and alpha and beta's.

    2. Use a `boxplot` to compare the favourites' score versus the underdog.

    3. 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 4: 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.

```
install.packages("quantmod")
library(quantmod)
getSymbols("BRK-A", from = "1990-01-01")
```

Problem 5: 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 6: 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 7: 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%.

	SP Index	Hedge Funds
2008	-37.0%	-23.9%
2009	26.6%	15.9%
2010	15.1%	8.5%
2011	2.1%	-1.9%
2012	16.0%	6.5%
2013	32.3%	11.8%
2014	13.6%	5.6%
2015	1.4%	1.7%
cumulative	65.7%	21.9%

(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?