

## ENGRD 2700 F14 Basic Engineering Probability and Statistics

### Homework 1

- You may complete this homework either on your own or in pairs. No groups of 3 or more! If you complete it with a partner, hand in a **single** homework with **both** of your names etc. on it.
- Be sure to **clearly** write your name, Net ID and the section you *attend* on your homework so we can record your score and return your homework in recitations, e.g., Shane Henderson, sgh9, Friday 8.40-9.55.
- Remember to staple your homework pages together.
- There is no need to use Word or similar software (Latex) to prepare your homework. Handwritten homework with printouts as needed is just fine, but see the next point.
- Statistics is partly about communication. Write clearly and concisely. Points will be taken off if your answers are garbled, difficult to read or disorganized.
- You may use any software package you like.
- Provide evidence for each of your answers, but there is no need to print every piece of working that you complete. If a calculation involves only very minor computation then explain the computation you performed and give the results. If a calculation involves more complicated steps on many many records then hand in the calculations and formulas for the first few records only. (To show formulas in Excel, click the “Show Formulas” checkbox under Excel Preferences → View or type CTRL + ‘(Left of 1 key).)

The data for the first part of this homework consists of data from the 2013 National Football League (NFL) Season. There are two Excel spreadsheets that you will need to download: NFLGames.xls and NFLTeams.xls. I assume no prior knowledge of football for this problem set.

The fields (column names) for NFLGames.xls are as follows:

1. Week. There are 17 weeks in the NFL season. Each team plays once a week, but also gets one of the 17 weeks off. As a result, each team plays 16 games.
2. Home Team ID. Each game is hosted by one of the two teams that plays. The team that hosts the game is called the home team. This column gives the unique team ID of the home team.
3. Away Team ID. The team that travels to play the home team is called the away team. This column gives the away team ID
4. Home Team Score. The number of points scored by the home team
5. Away Team Score. The number of points scored by the away team
6. Day of the Week. The day of the week during which the game was played.
7. Time. Column which tells whether the game was played during the day or night.
8. Field Surface. The type of field that the games was played on.
9. Temperature. The outside temperature at the start of the game.
10. Wind. The wind speed at the start of the game.

The fields (column names) for NFLTeams.xls relate team IDs to team names as follows:

1. Team ID. Unique ID for each team.
2. Team Name. The name of the NFL team that corresponds to each team ID.

## Data Visualization and Analysis Questions

(20 points)

1. (4 points) First, we will want to use NFLTeams.xls to figure out the names of the teams that played in each of the games. In NFLGames.xls, add two blank columns next to Away Team ID. Title the first “Home Team Name” and the second “Away Team Name” You can do this by right-clicking column header D above Home Team Score and choosing insert. Next, we will want to “look up” the team name corresponding to each team ID in the table NFLTeams.xls. This can be done using the Excel command VLOOKUP. Provide a screen shot of the final product with formulas showing in each cell (There is an explanation at the top of the homework that shows how to do this).

To use VLOOKUP, click on the first blank cell under Home Team Name and type “=VLOOKUP(a,b,c,d)” where the inputs to VLOOKUP have the following interpretations:

- a: Column whose entries you want to lookup
- b: All of the columns of the look-up table (Here, you will want to use the columns from NFLTeams.xls. It is okay that these columns are in a different spreadsheet.)
- c: Column number that we want to output (indices start at 1 and we want to output the team name)
- d: FALSE (we want an exact match)

Obviously, this is not exactly what you should type, I am giving you some hints on how to use VLOOKUP and there is plenty more guidance in the Excel documentation or at an office hour!

2. **Analyzing Home Field Advantage:** In the following problems, we will investigate whether there is an advantage to playing at home.
  - (a) (2 points) Calculate the mean and standard deviation for the scores of the home team and the away team.
  - (b) (3 points) Provide a histogram for the points scored by the home team. Compare this to a histogram of the points scored for the away team. Comment on the shapes of both histograms.
  - (c) (1 points) What percentage of games were won by the home team?
  - (d) (2 points) Based on this analysis does there appear to be a “home-field advantage”? What another analysis might be useful?
3. **Analyzing the effect of weather:** In this part we see if weather conditions affect the number of points scored.
  - (a) First, delete the rows that have a temperature of 0 degrees as these entries correspond to games played in domes (A downside of my data source). You can do this quickly using the Filter button on the Data tab in Excel. First, highlight all rows and columns by clicking on the gray cell above row 1 and left of column A. Then, click Filter, little arrows should appear next to the column names. You can click on these arrows to filter each of the columns. To unfilter, click the filter button again.
  - (b) (2 points) Plot the total points scored in each game versus the temperature. Comment on the relationship between the temperature and the number of points scored.
  - (c) (2 points) Repeat part b) but replace temperature with wind.
4. (4 points) Create a time series for the San Francisco 49ers showing this team’s scoring from week to week.

## Counting and Probability Questions

1. (a) How many distinct nine-letter words can be made by permuting the letters in “PEPPERONI”? (For this problem, count gibberish such as “EPPERONIP” as a word.)

There are  $9!$  ways to permute the letters in the word “PEPPERONI”, but not all of these permutations result in distinct words. The letter “P” appears three times, and we can ignore the order

in which each “P” appears within the permuted word. The same is true for the letter “E”. Since there are  $3!$  ways to permute 3 copies of the letter “P”, and  $2!$  ways to permute 2 copies of the letter “E”, we can create

$$\frac{9!}{3!2!} = 30,240 \text{ words.}$$

- (b) How many distinct ways are there to arrange the letters “PEPPERONI” in a circle?

If we attempt to build necklaces out of the 30,240 words from part (a), there will be duplicates. For instance, necklaces built from “PEPPERONI” and “IPEPPERON” are identical, as each one can be rotated to obtain the other. We will thus obtain 9 copies of each necklace, and so there are

$$\frac{30,240}{9} = 3,360$$

distinct necklaces that can be built.

2. A man has  $n$  keys, of which only one unlocks his car. He picks a key uniformly at random among the keys that he has not already tried. This means that if there  $k \leq n$  keys remaining, the probability of choosing each one is  $\frac{1}{k}$ . Compute the probability that he succeeds on his  $i^{\text{th}}$  try.

The order in which he selects keys to unlock his car is equally likely to be any of the  $n!$  possible sequences. To find the probability that he succeeds on his  $i^{\text{th}}$  try, we need to find the number of sequences such that the desired key appears in the  $i^{\text{th}}$  position. This number is  $(n-1)!$  — the number of ways in which the remaining  $n-1$  keys can be arranged in the sequence. Thus he succeeds on his  $i^{\text{th}}$  try with probability

$$\frac{(n-1)!}{n!} = \frac{1}{n}.$$

3. Find the probability of drawing a five-card hand containing two clubs, two spades, and one diamond from a standard deck of playing cards.

We found in lecture that there are  $\binom{52}{5}$  possible poker hands, and so it remains to find the number of five-card hands containing two clubs, two spades, and one diamond.

To find this number, we break the problem up into parts. There are  $\binom{13}{2}$  ways to build a two-card hand from a 13-card deck containing only clubs (or a deck containing only spades). Similarly, there are  $\binom{13}{1}$  possible one-card hands from a deck containing only diamonds. Using the product rule, there are

$$\binom{13}{2} \binom{13}{2} \binom{13}{1}$$

possible hands containing two clubs, two spades, and one diamond. The probability of drawing such a hand is

$$\frac{\binom{13}{2} \binom{13}{2} \binom{13}{1}}{\binom{52}{5}} \approx 0.030.$$

4. Consider a lottery in which we purchase a ticket with five distinct numbers from the set  $\{1, 2, \dots, 39\}$ . During the drawing, five of these numbers are randomly selected (again, uniformly at random), which we’ll call the *winning numbers* associated with that drawing.

For  $i = 0, 1, \dots, 5$ , let  $p_i$  be the probability that the ticket contains  $i$  winning numbers. Order does not matter, you just have to match numbers. Compute  $p_5$  and  $p_4$ .

There are  $\binom{39}{5}$  possible outcomes for the drawing, each of which is equally likely. Since there is only one way for the ticket to have five winning numbers, it follows that

$$p_5 = \frac{1}{\binom{39}{5}} = \frac{1}{575,757} \approx 1.74 \times 10^{-6}$$

There are  $\binom{5}{4}$  ways to select 4 numbers on the ticket to be winning numbers. The remaining number must be selected from the pool of 34 numbers not found on the ticket, which can be done in  $\binom{34}{1}$  ways. Thus

$$p_4 = \frac{\binom{5}{4} \binom{34}{1}}{\binom{39}{5}} = \frac{170}{575,757} \approx 2.95 \times 10^{-4}.$$

5. Consider a single elimination, bracket style tournament similar to the NCAA basketball tournament: In the first round, every team plays a game against another team, winners move on to play other winners and losers go home. This means that the numbers of teams left is cut in half after every round. The tournament continues until there is only one team left, at which point this team is deemed the champion. Instead of 64 teams as in the NCAA tournament, assume that we have  $2^{1000}$  teams. How many total games will be played in the tournament?

In every game played there is exactly one loser. Therefore, to count the number of games played, we just have to count the number of losers. Every team loses except for the champion. This means there are  $2^{1000} - 1$  losers and so there are  $2^{1000} - 1$  games.