**Exam I**

*MAT 135*

*Introduction to Statistics*

*Total Points :* 40

Instructions: Follow the instructions for each part of the Exam. The first part is handwritten while the last three parts will be done entirely in Excel. Turn in and upload all of your work to D2L when you are done. All data sets referenced can be found on *D2L > Course Materials > Content > Exams > Exam I.*

**Part 1: Definitions and Concepts**

Instructions: Provide a short answer for each of the following questions. Clearly explain each answer. You should be able to fit your answer on the exam itself, but if you need extra room, take out a sheet of paper. Be sure to label each problem and write your name at the top.

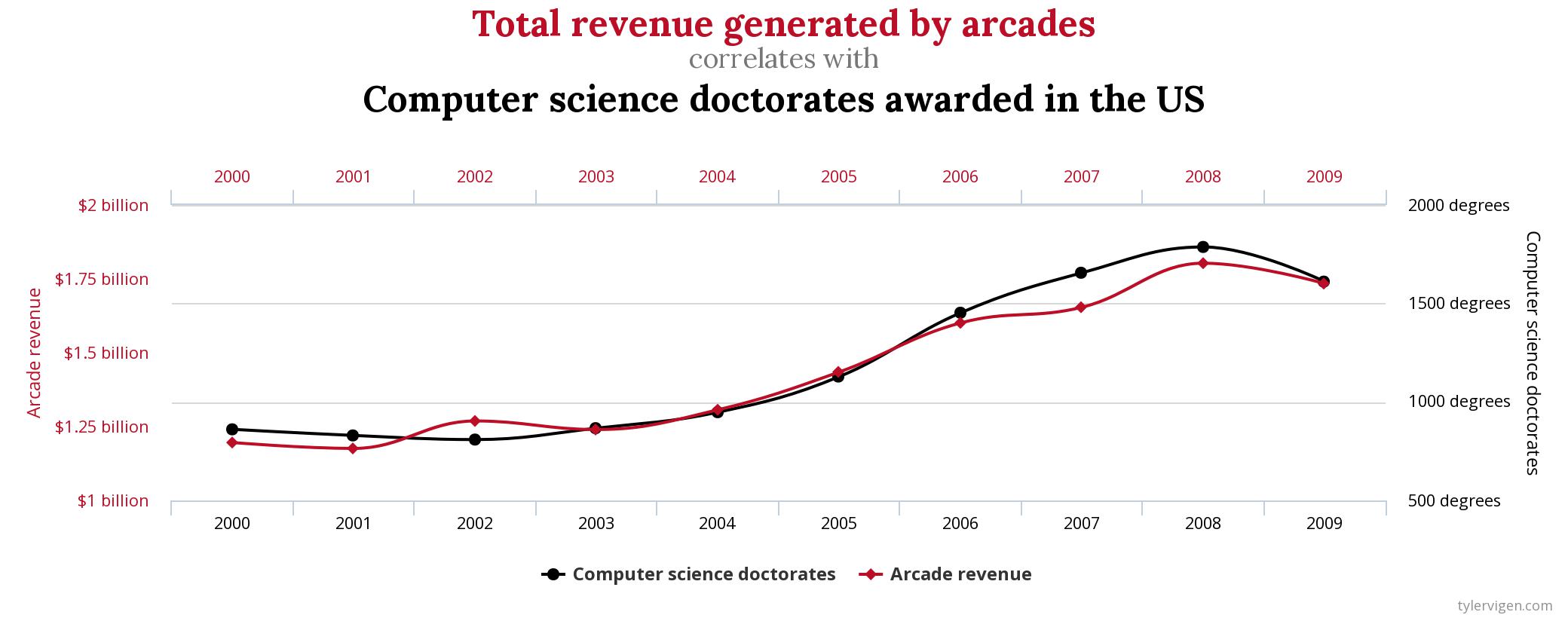
1. (3 pts) You want to estimate the average weight of the entire population living in Longmont, CO. In order to do so, you stand outside the Planet Fitness gym and ask every 10th person that enters the building how much they weigh. You collect a sample of data with 16 observations.

i. Identify the *population* and the *sample* in this problem.

ii. What type of sampling method is used in this study?

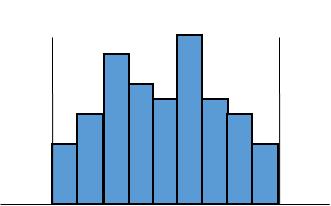
iii. Identify all potential sources of bias in this statistical survey. If you see a potential for bias, explain how the bias could be removed.

2. (2 pts) A statistician approaches you on the street (*as they often do)* and shows you the following graph,

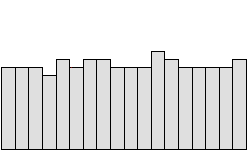


The statistician then begins ranting at you, "*See! A correlation of 0.92! Near perfect! The arcade industry is doomed unless more people get computer science doctorates*!" What is wrong with his argument?

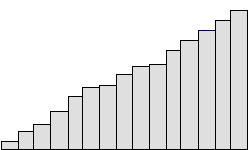
3. (*5 pts*) Match the distribution shape with its type.

 A. a. Skewed Right

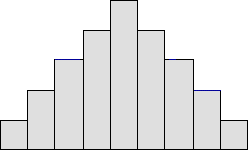
B. b. Skewed Left



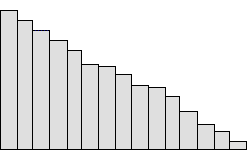
C. c. Bimodal



D. d. Normal



E. e. Uniform



**Part 2 : Frequency Distributions and Histograms**

Data Set 1: Corn Yields

Instructions: The first data set is a sample of the average corn yield during the harvest season in the United States. The sample is measured in bushels per planted acre and was taken over the years of 1975 to 2014. Navigate to the datasets on D2L and download the dataset "CORN\_YIELDS*\_YOURNAME.xls*". Save the spreadsheet and replace your name where appropriate. Complete the following problems.

4. (5 pts) Using 10 classes and a class width of 14, construct a frequency distribution for this data set.

i. Create columns for lower class limits, upper class limits and frequencies. Fill in the cells appropriately.

ii. Using the SUM function, verify the sum of frequencies in your frequency table add up to the

total number of observations in the dataset. Save the result in a cell and label it.

iii. Add a new column to your table and construct a relative frequency distribution.

iv. Add a new column to your table and construct a cumulative relative frequency distribution.

5. (2 pts) Using the results from the first problem, construct histograms for the relative frequency distribution and the cumulative relative frequency distribution.

i. Using the appropriate histogram, estimate the median of the dataset, i.e. the harvest yield such that 50% of harvests will be less than this yield.

Data Set 2: Robert De Niro Movie Ratings

Instructions: Save the previous data set and upload it to D2L to the Exam I dropbox. Navigate back to the folder on D2L containing the exam datasets and open the dataset "*DE\_NIRO\_YOURNAME.xls*". Save the spreadsheet and replace your name where appropriate.

The second dataset consists of aggregate review scores for Robert De Niro movies taken from the website *Rotten Tomatoes*. Reviews from multiple media outlets are taken and averaged together to generate the score on *Rotten Tomatoes*. The movies are then assigned a score between 0 and 100. Complete the following problems regarding this dataset.

6. (5 pts) Construct a relative frequency distribution for this dataset by following the steps in this problem:

a. Find the minimum and maximum observations. Save and label the results in a cell.

b. Find the total number of observations. Save and label the results in a cell.

c. Find the range of the dataset. Save and label the results in a cell.

d. Using 8 classes, calculate the class width. Round up the result to the nearest whole number. Save and label the result in a cell.

e. Create columns for the lower class limits, upper class limits and frequencies. Fill in the cells appropriately.

f. Add a new column and construct a relative frequency distribution\

7. (3 pts) Using the results from problem #3, create a pie chart of Robert De Niro's movie career.

4

i. Based on your pie chart, what is the most frequent range of ratings on Rotten Tomatoes for Robert De Niro movies? Save and label your answer in a cell.

**Part 3: Lines Graphs, Scatter Plots and Correlation**

Data Set 3: Big Bang Theory Viewers

Instructions: Save the previous data set and upload the spreadsheet to the Exam I dropbox on D2L. The third data set lists the number of viewers tuned into to the Big Bang Theory and the corresponding air date of the episode. Go to D2L and open the Big Bang Theory data set. The file name should be, "*BIG\_BANG\_THEORY\_YOURNAME.xls*". Save the spreadsheet and replace your name where appropriate. Complete the following problem.

8. (5 pts) Using this dataset, create a line graph. Put the number of viewers on the vertical axis and the air date of the episode on the horizontal axis. Format the axes appropriately.

a. Based on your graph, identify the date of peak viewership for the Big Bang Theory. Sav e and label your answer in a cell.

Data Set 4: United States GDP

Instructions: Save the previous data set and upload the spreadsheet to the Exam I dropbox on D2L. The fourth and final data set is a bivariate dataset of the US Gross GDP (measured in billions of dollars) versus the year. GO to D2L and open the US GDP vs Year data set. The file name will be, "*US\_GDP\_VS\_YEAR\_YOURNAME.xls*" Save the spreadsheet and replace your name where appropriate. Complete the following problems.

9. (5 pts) Construct a scatter plot of the US GDP and the year.

a. Which variable is the independent variable and which variable is the dependent variable? Justify your answer. Save your answer in a cell

b. Based on the scatter plot, what type of correlation do you observe in this data set? Save your answer in a cell.

c. Insert the line of best fit on the graph along with R-squared.

10. (5 pts) Determine if the correlation in this dataset is significant by following the steps in this problem:

a. Calculate *r*, the correlation coefficient. Save your answer in a cell.

b. Calculate the number of observations in the dataset.

c. Refer to the table of critical values on the projector up front to find the critical cut-off point for significant correlations. Save your answer in a cell.

NOTE: If you can't find the exact sample size on the chart, get as close as you can without going over!

d. Is there a significant correlation in this dataset? Save your answer in a cell.

e. Based on your answer to part d, can we use the line of best fit to make predictions? If so, what should the US GDP be in 2019? Save your answer in a cell.

**THE END**

Instructions: Be sure to save your final spreadsheet and upload it to D2L! Turn in the rest of your exam when you are finished.