





Q1.

Some researchers argue that using social media makes people feel worse about themselves. Prof. Vitak surveyed students at Michigan State University about their Facebook use. Using her survey results your job is to propose an analysis of the data that would test the relationship between Facebook use and self-esteem. The data is stored as a large csv file ([here](#)  ) , the first row contains shortened variable names, the codebook ([here](#)  ) explains what these variables mean. Your write up should include the following:

a) A clear statement of the research question

Hypothesis: increased Facebook use decreases self-esteem

Question: Do students who use Facebook more than 60 minutes (codebook 19) report lower self satisfaction in a survey (codebook 17)?

b) Description of the study design

- **What is the population? The sample?**
 - The population is university students.
 - The sample is MSU students.

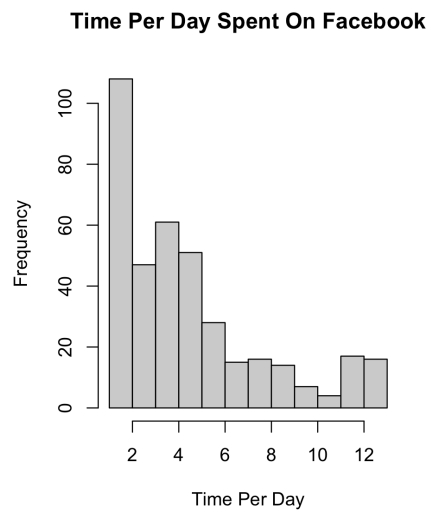
- **What kind of study is it?**
 - Observational

- **What will be your independent and dependent variables?**
 - Independent variable: minutes spent on Facebook per day
 - Dependent variable: lower reported self worth

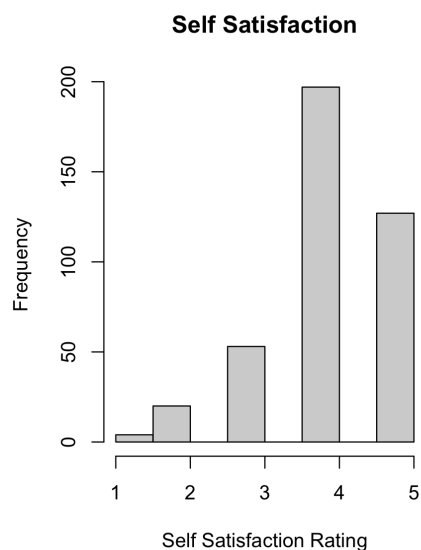
- **What scale of measurement is used for your variables?**
 - The independent variable uses intervals, or increments of 15 minutes.
 - The dependent variable uses an ordinal scale of 1-4.

c) An exploratory analysis of your main variables

- **How many observations are in your data set? Are there missing values?**
 - IV: 384 observations out of 437 responses, 53 missing
 - DV: 401 observations out of 437 responses, 36 missing
- **Create a histogram for each of your main variables. What is the distribution of each of them?**
 - The independent variable has a right skewed distribution



- The dependent variable has a left skewed distribution



- **Compute descriptive statistics for your main variables (e.g., means, standard deviations)?**

- IV:

mean	4.856770833
median	4
mode	2
variance	10.84366161
std dev	3.292971547
sample	384

- DV:

mean	4.054862843
median	4
mode	4
variance	0.7369825436
std dev	0.8584768743
sample	401

d) A description of what can be learned from this analysis.

- **What do you expect to find and how will it address the research question?**
 - If we look at the descriptive statistics for time spent on Facebook daily, we see that the median lies at around 45 minutes to an hour. However, if we look at the histogram, we do see that there are other reports, resulting in a right skewed graph. Checking the standard deviation, we see that it is around 3 intervals, or around 30 to 45 minutes worth of difference between reported numbers. The mode is 2, standing for 15-30 minutes.
 - The median and mode for self-satisfaction lies at 4, which registers as “Agree” with the statement that the student is satisfied with themselves as a whole. The standard deviation is 0.86, indicating that there is less than a full interval’s worth of deviation from each data point.
 - If self reported satisfaction decreases as time spent on Facebook increases, then the hypothesis of increased Facebook time decreasing self-esteem would be correct. The research question would be answered in the affirmative. For fun, I sorted the data by increasing satisfaction and graphed it on the x-axis against the time spent on Facebook. After looking at the trendline, the R^2 indicates that the response variable cannot be explained by the predictor variable at all.

Comparing Satisfaction Against Time on Facebook

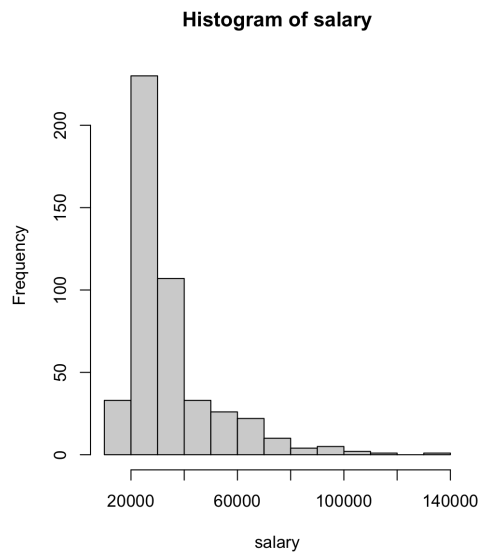


- **Are there any limitations with this analysis (e.g. bias)? How might these limitations affect your results?**
 - Initially, I looked at "I feel that I'm a person of worth, at least on an equal plane with others," but determined that since it was a judgment based on relativity, it would be difficult to use on equal footing. I chose self-satisfaction as my variable since self-esteem and self-satisfaction are closely related; one is in regards to what one thinks of themselves, and the other is a measure of how poorly or well one thinks of themselves.
 - The statement I decided on was "On the whole, I am satisfied with myself."
 - Students were to respond to this statement with "Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree."
 - However, there are a few issues with this. First, self-reported statistics are subjective and prone to misinterpretation. Second, this statement was posed well after numerous questions about self-worth, which may have primed the student to reply a certain way. Lastly, an ordinal scale is nondescript; students may list other issues associated with their self-satisfaction that are unrelated to social media. For example; someone may be dissatisfied with themselves because they have future goals that are yet to be achieved and wish to challenge themselves further. The lack of specificity with survey questions leads to varying reasons behind answers that are not readily obvious.

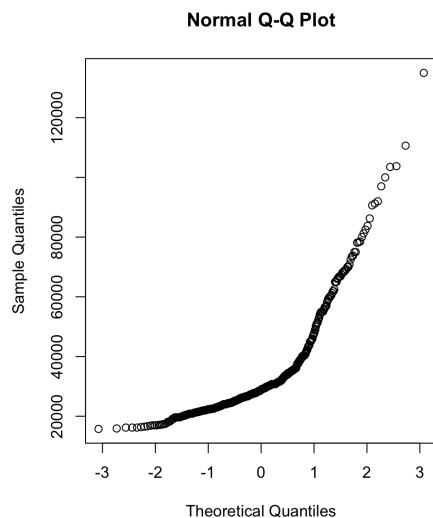
Q2)

Employee_data.csv gives a sample of salaries of students who graduated from the MIM program. Using this data please answer the three questions listed below. Write up your answers in a Word document, include all your work (requested graphs, R code). To complete this assignment you will want to download the data [Employee_data.csv](#).

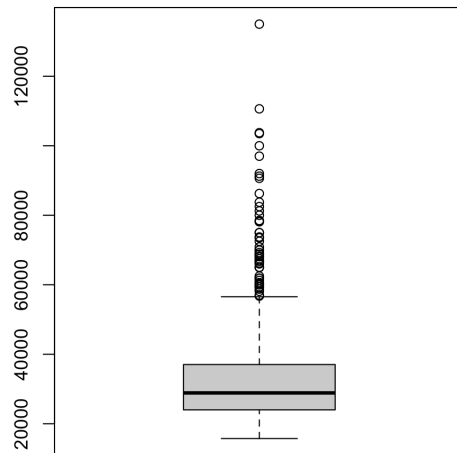
- **a) Create a histogram of employee salary (column "salary").**



- **b) Create a normal Q-Q plot of employee salary.**



- **c) Create a box plot of employee salary.**



- **b) Is employee salary normally distributed? Explain, why or why not.**
 - If we refer to the histogram generated for the salaries, we see that employee salary is right skewed, with higher frequencies on the lower end of the x values. After generating the Q-Q plot, this becomes even more obvious as the values do not fall along the diagonal line. Finally, the box plot shows that the median is far to one side, and there is a very long tail on the other.
- **e) Assuming that employee salary is normally distributed, what is the probability of a MIM graduate in our *sample* earning greater than \$60,000 per year?**
 - Since the mean is around 34000 and we are looking for earnings greater than 60000, the parameter "lower.tail" will be false here. When this is run, R returns a value of 0.06705832, so a 6.71% probability that a graduate will earn more than 60000 a year.
- **f) Assuming that employee salary is normally distributed, what is the probability of a MIM graduate in our *sample* earning between \$25,000 and \$40,000 per year?**
 - To see the probability between two points, we subtract the pnorms. This results in a value of 0.3374943, or a 33.75% probability that a graduate will earn between 25000 and 40000 a year. Of course, if we look at our distribution, it is clearly right skewed and not normally distributed, so this probability is not true.