HW 1

For tips on using Minitab, look at “A Quicker Introduction to Minitab”, on the course website. (Click on “Minitab”).

Please submit the relevant output and graphs produced by Minitab. You can copy and paste Minitab graphs as well as output from the Minitab session window into a Word document.

1) A group of college students believes that herbal tea has remarkable restorative powers. To test this belief, they make weekly visits to a local nursing home, visiting with the residents and serving them herbal tea. The nursing home staff reports that after several months most of the residents are more cheerful and healthy. A skeptical sociologist commends the students for their good deeds but scoffs at the idea that herbal tea helped the residents. It’s all confounding, says the sociologist. Identify the treatment and response variable in this informal study. Then explain what other variable is confounded with the treatment.

2) During World War II, many economists, mathematicians and statisticians were members of Columbia University’s Statistics Research Group, which did high level consulting work for the armed forces. As part of this group’s work, statistician Abraham Wald was asked where to place armor on planes. It seemed obvious to the aircraft engineers that armor was needed at the places most frequently hit as found in a large sample of battle proven airplanes. After studying the bullet holes of a sample of returning planes, Wald’s conclusion was to place armor where bullet holes were least frequently found in these planes. What was wrong with the engineers’ sampling design that made Wald come up with this seemingly odd conclusion? Hint: Are the engineers seeing a random sample of all planes?

3) The file CelebrityEarnings.MTW contains the personal earnings in 2007-2008 for the Forbes Top 10 Most Powerful Celebrities. Use Stat → Basic Statistics → Display Descriptive Statistics to compute the mean, median and interquartile range, for the earnings. You will need to obtain the interquartile range by hand as the difference between the third quartile (Q3, the 75’th percentile) and the first quartile (Q1, the 25’th percentile). Why do you think that the mean is greater than the median? To help answer this, look back at the data set, and also create a histogram, using Graph → Histogram → Simple. Next, remove the two largest observations, for J.K. Rowling and Oprah Winfrey, by replacing their earnings in the Minitab spreadsheet by * (a missing value). Recompute the mean, median and interquartile range for the modified data set. Compare with the corresponding values for the full data set. What changed and what stayed essentially the same? Why?
4) For the 2019 Major League Baseball salaries (BaseballSalaries2019.mwx), compute the log salaries using Calc → Calculator → Store Result in variable: LogSalary, Expression: log(Salary) → OK. Which player had the highest salary? Which player had the highest log salary? Explain why this is not a coincidence. Next, plot a histogram of the log salaries. Describe any patterns you see in this histogram. Which data points would you need to remove to get a nearly symmetric histogram of log salaries?