STAT5101: Foundations of Data Science

Assignment 1

Academic year 18/19, First term

Deadline: During Class, Sep 26 (WED), 2018.

1. When do Americans decide what to make for dinner? An online survey indicated the following:

When Americans decide what to make for dinner	%
At the last minute	37
Plan in advance	25
That day	37
Don't know	1

   a. Construct a bar chart, a pie chart, and a Pareto diagram.

   b. Which graphical method do you like the best to portray these data? provide one good feature of the selected method.

2. The following data represent the battery life, in shots, for three pixel digital cameras:

30	18	14	17	38	46	26	35	38	12	11

   a. Place the data into an ordered array.

   b. Construct a stem-and-leaf display.

   c. For the ordered array and the stem-and-leaf display, which method provides more information? Discuss.

   d. Compute the mean, minimum, maximum, median, first quartile, and third quartile using R. (Hint: apply R commands "mean" and "summary" on the vector)

   e. Compute the mode, variance, standard deviation, range, and coefficient of variation.

   f. Form the box-and-whisker plot.

3. One operation of a mill is to cut pieces of steel into parts that will later be used as the frame for front seats in an automotive plant. The steel is cut with a diamond saw and requires the resulting parts to be within $\pm$ 0.005 inch of the length specified by the automobile company. The data in the file **STEEL.xls** comes from a sample of 100 steel parts. The measurement reported is the difference in inches between the actual length of the steel part, as measured by a laser measurement device, and the specified length of the steel part For example, the first observation, -0.002, represents a steel part that is 0.002 inch shorter than the specified length.

   a. Construct the frequency distribution and the percentage distribution. (use 6 categories between -0.00350 and 0.00550)

   b. Plot the percentage histogram.

   c. Plot the percentage polygon.

   d. Plot the cumulative percentage ploygon.

4. The data in the file **BATTERIES2.XLS** represent the price and cold-cranking amps (which denotes the starting current the battery can deliver) of automobile batteries.

   a. Compute the coefficient of correlation r.
   b. What conclusions can you reach about the relationship between the cold-cranking amps and the price?

5. Where people turn to for news is different for various age groups. A study indicated where different age groups primarily get their news.

	Age Group		
Media	Under 36	36-50	50+
Local TV	107	119	133
National TV	73	102	127
Radio	75	97	109
Local newspaper	52	79	107
Internet	95	83	76

Construct a side-by-side bar chart for three age groups.