

Stats 314, Data Analysis #2

Cody Malick

malickc@oregonstate.edu

October 16, 2016

Part I

a

The dataset with the smallest standard deviation is 'iii.' The reason being all the numbers are the same. The standard deviation is a quantity calculated to indicate the extent of deviation for the entire dataset. If they're all the same number, then the standard deviation of the dataset is 0.

b

The set with the largest standard deviation will be the one with the most values far away from the mean. In this case, 'i' would have the largest standard deviation because of the numbers '7' and '11.' Seven and eleven are the only differences between the two datasets. Because these two values 'deviate' from mean more than the values in dataset 'ii,' then 'i' has a larger standard deviation.

Part II

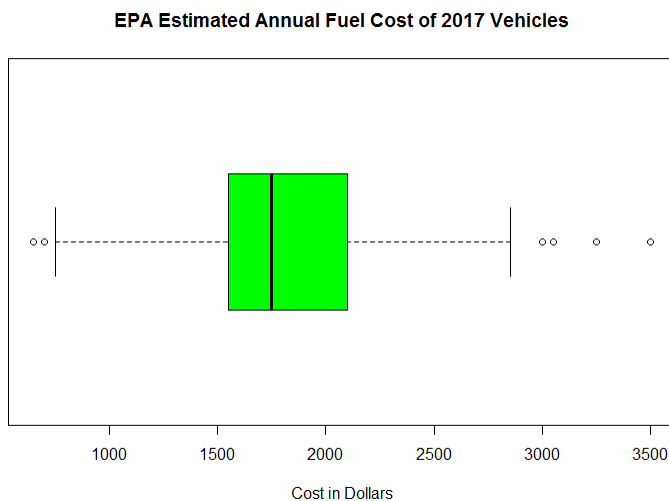
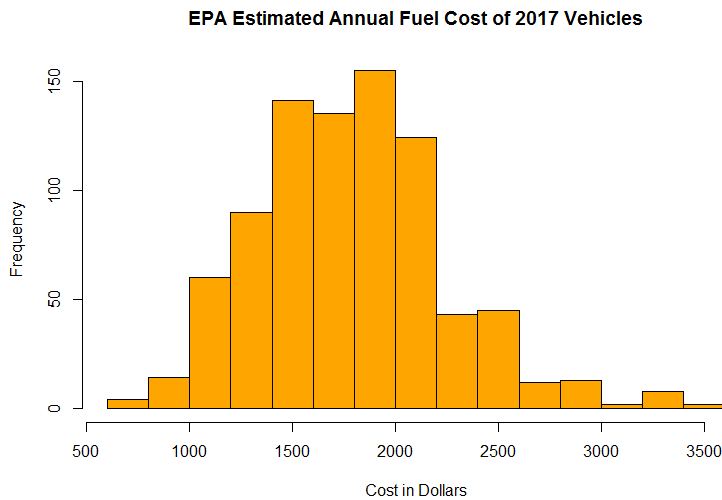
Scheme B would be better because even numbers of both barley populations get an even amount of water from the river. The main issue with scheme B is that two of type 1 barley could get all the land by the river, encouraging growth.

Scheme B would create two groups of even barley groups. One with both types getting water, and one with both types not getting water. Scheme B, as stated above, would be susceptible to an uneven distribution of samples with and without water.

Part III

a

[illegible]



I feel that the stem and leaf plot is not appropriate for this type of data set. The histogram does an excellent job showing where the data is most concentrated, as well as being visually intuitive. The only issue being it's not readily clear what the exact median is. The box plot, on the other hand, while not being entirely intuitive visually, immediately gives some clear measurements about what parts of the data we care about, and what parts are irrelevant. I would say in this case that the histogram is the best choice.

b

No, because there are too many values to make the stem and leaf plot useful.

c

It represents the area between the first and third quartiles.

d

Mean	Standard Deviation	Minimum	1st Quartile	median	3rd quartile	maximum	IQR
1815	459.428	650	1550	1750	2100	3500	550

e

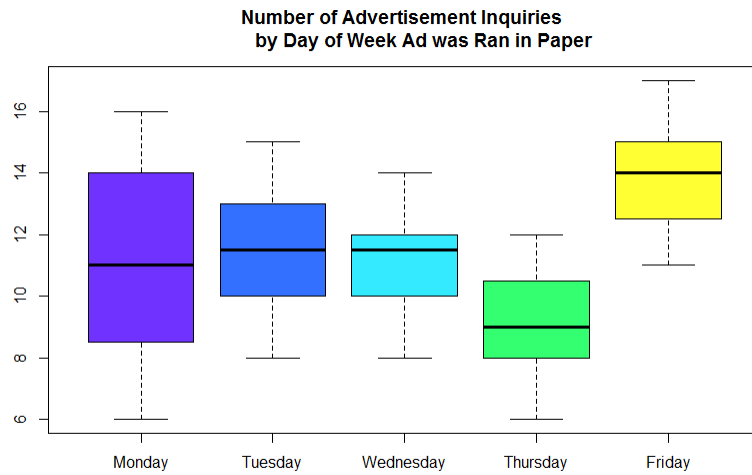
The data has a single mode, with a positively skewed shape. There were a few outliers as we approach the maximum around 3000 to 3500. The center is the median at 1750.

f

In this case, either the mean (1815) or the median (1750) could be argued to be great metrics. The reason for this is that the distribution of the data is fairly grouped around the median/mean. They both land in the second quartile, and are near the top of the mode.

Part IV

a



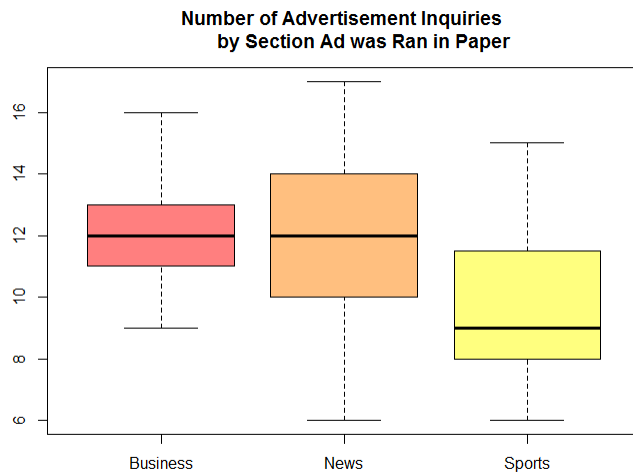
b

Day	Inquires.Min	Inquiries.1st Quartile	Inquiries.Median	Standard Deviation
Monday	6.00	8.75	11.00	3.315
Tuesday	8.00	10.00	11.50	2.067
Wednesday	8.00	10.00	11.50	1.696
Thursday	6.00	8.00	9.00	1.758
Friday	11.00	12.75	14.00	1.781

c

At first glance the visually intuitive answer is that Friday is the best day to run an add in the paper. The reason for this being that the IRQ is very high compared to any of the other days. The majority of the data lies in a high range. Looking at the median for that day, it also places much, much higher than the other days. Looking at the data in our table, it also indicates that friday is a better day because of the median and the minimum.

d



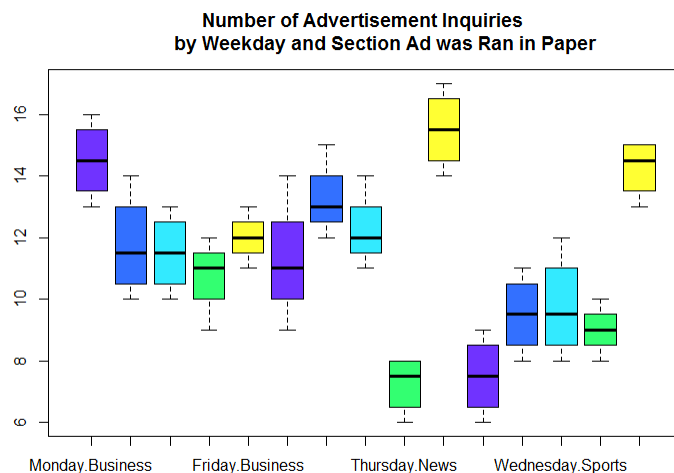
e

Section	Inquires.Min	Inquires.1st Quartile	Inquires.Median	Standard Deviation
Business	9.00	11.00	12.00	1.744
News	6.00	10.50	12.00	3.059
Sports	10.00	11.25	15.00	2.575

f

At first glance, Business looks like a fairly solid section to run adds in. The reason for this being that the distance between the first and third quartiles is fairly small, meaning that most of the ads run that day lie in a fairly concentrated range. The standard deviation for Business is fairly low. Looking at News, however, shows that there are days where News will run more inquiries than business, but also the possibility of less inquiries. Sports, on the other hand, runs less inquiries in general.

g



h

Day	Section	Inquires
Monday	Business	14.50
Tuesday	Business	11.75
Wednesday	Business	11.50
Thursday	Business	10.75
Friday	Business	12.00
Monday	News	11.25
Tuesday	News	13.25
Wednesday	News	12.25
Thursday	News	7.25
Friday	News	15.50
Monday	Sports	7.50
Tuesday	Sports	9.50
Wednesday	Sports	9.75
Thursday	Sports	9.00
Friday	Sports	14.25

i

Overall, immediately visible, Friday is one of the best days two out of three times. Friday, followed by Monday, would be the best two days to place ads overall. The other days vary throughout the week, but Monday and Friday end up having the highest average medians.