Fuentes

ST-314 Homework 1
Due Monday, April 14th at 5:00 PM

Please download this document. Then, complete your solutions and upload as PDF or Word Document. No other formats will be accepted.

Typing or entering answers by hand is accepted as long as solutions are presented neatly and the document is uploaded as PDF. Give the solutions in the space provided.

Instructions

- The homework is due on Monday, Apr. 14th at 5:00 PM. The homework MUST be uploaded to Blackboard as PDF or Word document. No other formats will be accepted.
- The homework is worth 25 points.
- Late homeworks will not be accepted under ANY circumstances.
- You will have 3 attempts to successfully upload your homework. Only your last successful attempt will be graded.
- You may work in groups of 2-3 people, but must submit individual solutions.
- You must provide complete answers in order to receive full credit.
- Please, use the space assigned to provide solutions to the problems.
- Failing to follow any of these instructions may result in a deduction of points from your total score.

Scanners

If you need to scan your homework, eScanners are located in the 2nd floor Copy Center, 1st, 3rd and 5th floors. These scanners allow you to scan documents (color, gray, b/w) in searchable PDF or quick PDF and send them to an email address or store on a flash drive. It is recommended that you scan no more than 17 pages at a time to prevent the eScanners from freezing. Your mailbox will also need to have sufficient storage space for your documents or they will be lost when sent. There is no charge for the service at this time.

Problem 1. (8 Points)
A study was done of the strength of high-performance concrete obtained by using superplastiziers and certain binders. In particular, the following data was gathered on the flexural strength (a measure of the ability to resist failure in bending), which are measured in megapascals (MPa), for a sample of 28 beams made of high-performance concrete.

5.4	7.2	7.3	6.3	8.2	6.9	7.0	7.6	6.8	6.4	8.3	6.3	7.9	9	8.2
8.7	7.8	9.7	8.5	7.7	9.7	7.8	7.7	11.3	10.3	11.8	10.7	13.7	14.1	

In a short ***paragraph*** include the following:

Stem and Leaf Display:

5 | 4
6 | 33489
7 | 023677889
8 | 22357
9 | 077
10 | 37
11 | 38
12 |
13 | 7
14 | 1

Given the stem and leaf plot,The data seems to have a modal distribution around 7 MPa. The data is right skewed. In addition, 13.7 and 14.1 MPa could be considered outliers in the data. Here is a descriptive table outlining some important features in the data:

Mean	Median(2^{nd} Quartile)	1^{st} Quartile	3^{rd} Quartile	Interquartile Range: IQR()	Lower Bound	Upper Bound	Outliers
8.562	7.900	7.2	9.70	2.5	5.4	12	13.7, 14.1

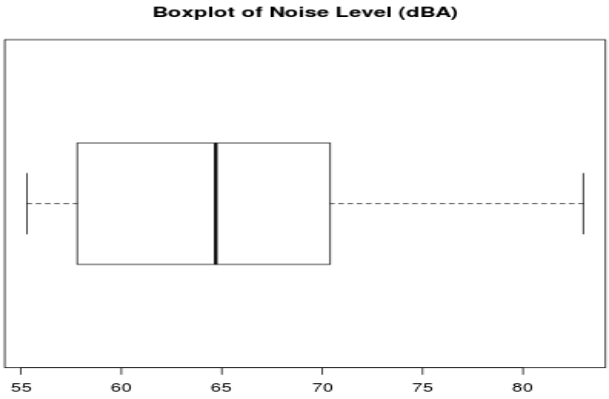
In this sample it makes sense to use the median of the data as a mark for the middle of the data sset. The median appears to be closer to the peak of the data from looking at the stem and leaf plot. In addition, the median is more resistant to outliers which are present in the data. The proportion of beams which would meet specifications to be bridge supports are:

Answer = Number higher than 7.0 Mpa / Total. Answer = 23 / 29 = 0.793 = 79.3%

Considering that a bridge collapse could be catastrophic, I don't think that these beams should be used as bridge supports. 79.3% is not a high enough guarantee of safety for such a key part of a bridge.

Problem 2. (9 Points)
A sample of 77 individuals working at a particular construction site office was selected at random. The noise level (dBA) experienced by each individual was determined. See exercise 63 on page 46 to view the data.

a. Use the R code and instructions under the Data Sets and Code tab to obtain descriptive statistics and graphical displays.
   Min. 1st Qu. Median Mean 3rd Qu. Max.
   55.30 57.80 64.70 64.89 70.40 83.00
b. Choose one of the graphical displays you feel best summarizes the data.



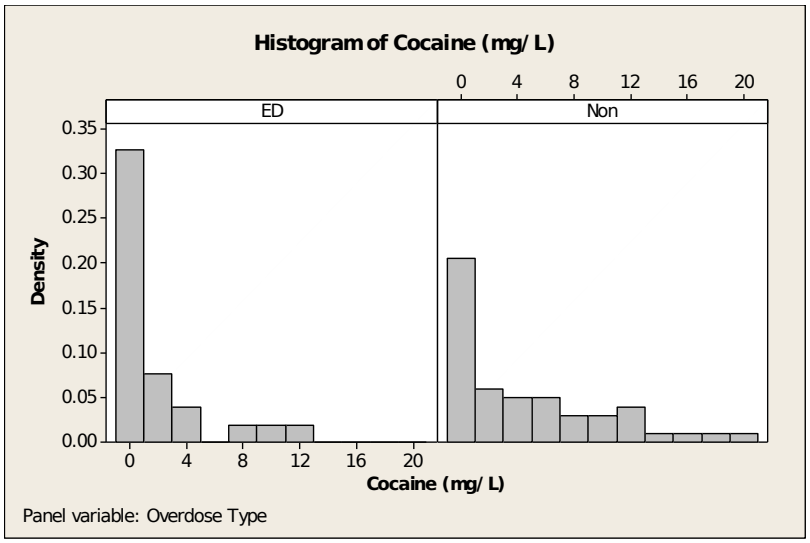
c. Give a table that includes the mean, median, standard deviation, quartiles and interquartile range.

Median	Median (Second Quartile)	Mean	Standard Deviation	1st Quartile	3rd Quartile	Interquartile Range
64.70	64.70	64.89	7.802671	57.80	70.40	12.6

d. Use parts (b) and (c) to thoroughly describe the data in the context of the problem.
   The noise experienced by workers at the construction seems to center on 65 Dba. It appears that most workers (those centered between the 1st and 3rd quartile) heard noise on the lower end of the spectrum captured by the data. The data is right skewed meaning that a smaller minority of workers are experiencing very loud sounds.
e. Given the shape of the data which measure, the mean or the median, would be a more appropriate to represent the center of the data? Explain your reasoning.
   I think that the median would be a more apropriate representation of the center of the data. The median is more resistant to outliers being present in the data.

Problem 3. (8 Points)

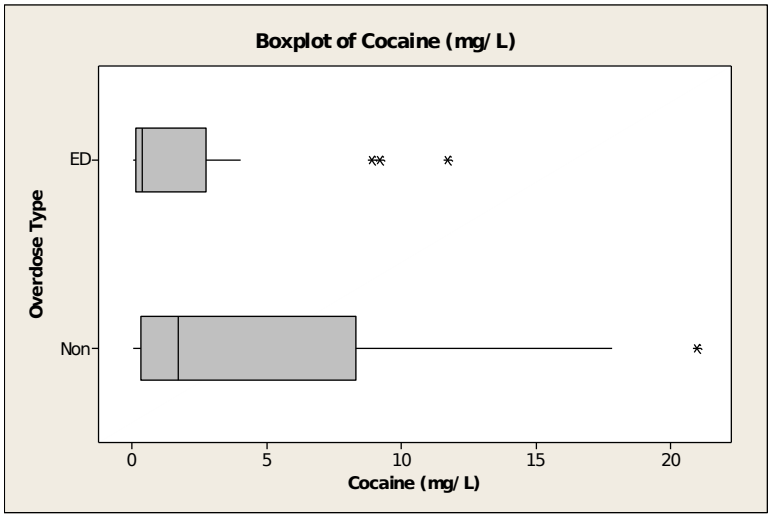
Blood cocaine concentration (mg/L) was determined both for a sample of individuals who had died from cocaine induced excited delirium (ED) and for a sample of those who had died from cocaine overdose without excited delirium (Non); survival time for people in both groups was at most six hours. Data obtained from Exercise 01-59 in text page 45.



a. Describe the data distributions from the histograms of both groups. Include description of the shape, center and spread. Compare the two distributions in the context of the situation.

The distribution of people who died from cocaine induced excited delirium is highly unimodal with a slight right skew. The data has a center of around 0 mg/L. The distribution of the graph of people who died without cocaine induced excited delirium is also unimodal with a center of around 0 mg/L. However, the data has a large right skew.

This means that the majority of people who die because of cocaine induced excited delirium die, on average from a much lower dose of cocaine than those without the condition. People who do not experience the condition seem to be able to tolerate much larger doses of cocaine without dying.



b. Describe the data distributions of both groups. Be sure to point out distinct features of the plot. Include description of the shape, center and spread. Compare the two distributions in the context of the situation.

The data for the ED spread is highly concentrated around the median with a slight right skew. In addition, there are several outliers. The interquartile range is much smaller than the spread for non-ED data. In addition, both the ED data and the non-ED data have a much smaller range between the 1st and 2nd quartiles than between the 2nd and 3rd quartiles. The spread of non-ED data is also significantly more right skewed than the spread of ED data.

The boxplots show that people non-ED cocaine overdose had a wider variety of amounts of the drug in their bodies when they died than ED-related deaths. In addition, the non-ED deaths could tolerate significantly more cocaine in their bodies on average before they died than people that died from ED.

c. The histograms and side-by-side boxplots display the same data. What are the advantages and disadvantages of each display? Specifically, what features are easier to see in each plot? Which plot do you prefer when comparing the two groups? Why?

In the histograms, it is easier to see the concentration of the data. It is easier to get an idea of how many people fit into each category of the data. In the boxplots, import information about the data is clearly labeled. Outliers and quartiles are clearly visible. The downside of the boxplot is that it is hard to get an idea of exactly how many observations there were in each dataset. However, at face-value I prefer a box-plot to a histogram because it is easier to quickly pull important information about a dataset from a box-plot.