

Stat 528 - Homework 1 Solutions

1.2

The individuals of the set of data are the students: Advani, Barton, etc.

The data contain the values of students, major, points and grade. So the variables are Major, Points and Grade. Major is categorical. Points is quantitative. Grade is categorical.

1.10

Since the proportions of dishwashers that need a service call for each brand are different and comparable, we could use the proportion to measure its reliability.

$2942 / 13376 = 0.219$, about 22% of Brand A dishwashers needed a service call, while $192 / 480 = 0.4$, 40% of Brand B dishwashers did.

1.18

The overall shape is skewed to the right with the center around 4. This means that there are more short words (3-4 letters) than quite long words (>10 letters) in Shakespeare's plays. We would expect the distribution of other authors to be similar, because short words, such as "the", "is", "as" and "you", etc, are common.

1.24

There are no obvious differences between the two distributions. All of the data are scattered between 90 to 140, and most data are centered between 110 and 115.

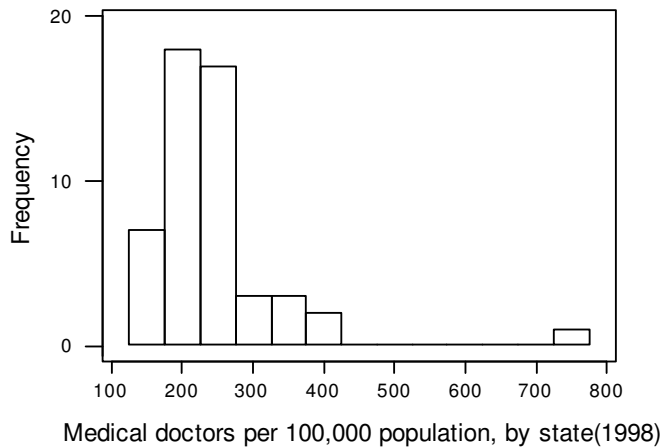
Yes, both centers are close together.

```
      9 8
2    10 2
77   10 9
2210 11 0224
      11 79
3    12 3
9    12
      13 0
6    13 0
```

1.28

(a) The populations of the states vary widely. It should be expected that there are many more doctors in New York or California than in Wyoming. Health care is better described in terms of the number of residents who are to be served by each doctor.

(b)



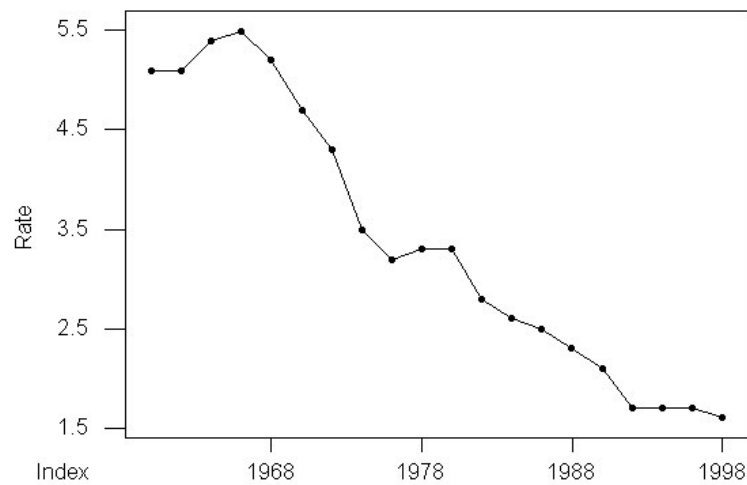
The distribution is skewed to the right with a single peak around 200-250. Most states have 150-350 medical doctors per 100,000 population. The one that stands out is D.C., with 737 medical doctors. It's an outlier, because it does not fit the overall pattern of the data. This is most likely due to the fact that D.C. is a city, like any other major cities, there are many doctors working in the areas so the number of medical doctors per 100,000 population could be much more than that of any other states which typically include urban and rural area. Perhaps it also has to do with that there are overall large hospitals in D.C., but few people actually live in D.C.

1.38

(a) Plot below. Deaths are generally decreasing as the years go by.

(b) No, the decline is fairly steady from the mid 1960s to 1990. So lowering speed limits does not necessarily explain the decrease in the death rate.

(c) No. It is more informative to study the changes over time rather than consider the distribution of deaths overall, since there is a clear trend over time and histogram could certainly miss this information.

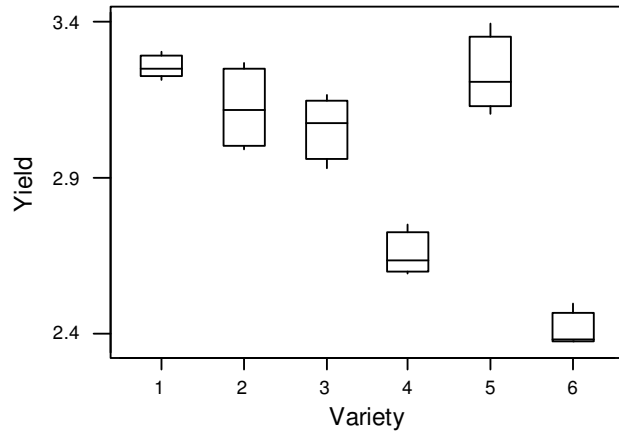


Note: To make a time series plot with the year in index, click Options... button in Time Series Plot in Minitab, type "1960" in Index on the top, and type "1960:1998/2" in Index in the middle. That's it.

In case you couldn't print out the dots on the plot after copying the graph to a word file, you can try two ways: first, print the graph directly from the Minitab; second, copy the graph to Paint (you can find it from start-all programs-accessories-paint), and then copy it from Paint to a word file.

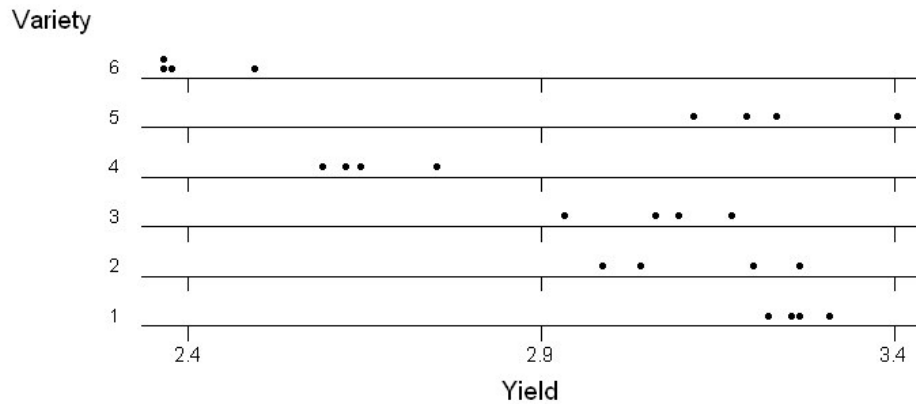
Special Problem:

(a)



(b)

Dotplot for Yield



In case you couldn't print out the dots on the plot after copying the graph to a word file, you can try two ways: first, print the graph directly from the Minitab; second, cope the graph to Paint (you can find it from start-all programs-accessories-paint), and then copy it from Paint to a word file.

(c)

The yield varies with the variable of Variety. So it appears that a typical value of yield for each variety, which could be measured by the median yield, depends on variety. Especially, variety 4 and 6 have yields much lower than other varieties.