Chapter 2 Textbook exercises

Solutions to even-numbered questions
Statistics and statistical programming
Northwestern University
MTS 525

Aaron Shaw

September 24, 2020

All questions taken from the *OpenIntro Statistics* textbook, 4th edition, Chapter 2.

2.12

The median seems to be around 80. The mean would be slightly lower than the median because the distribution has a long left tail (is left skewed).

2.16

(a) The distribution is right skewed with potential outliers on the positive (high) end, therefore the median and the IQR are preferable measures of center and spread because they are robust to outliers.

(b) The distribution is somewhat symmetric and has few, if any, extreme observations, therefore the mean and the standard deviation are preferable measures of center and spread.

(c) The distribution would be right skewed. There would be some students who did not consume any alcohol, but this is the minimum since students cannot consume fewer than 0 drinks. There would be a few students who consume *many* more drinks than their peers, giving the distribution a long right tail. Due to the skew, the median and IQR would be preferable measures of center and spread.

(d) The distribution would be right skewed. Most employees would make something on the order of the median salary, but we would anticipate upper management makes much more. The distribution would have a long right tail, and the median and the IQR would be preferable measures of center or spread.

2.20

(a) The distribution of percentage of population that is Hispanic is extremely right skewed with majority of counties with less than 10% Hispanic residents. However there are a few counties that have more than 90% Hispanic population. It might be preferable, in certain analyses, to use the log-transformed values since this distribution would be much less skewed.

(b) The map reveals that counties with higher proportions of Hispanic residents are clustered along the Southwest border, all of New Mexico, a large swath of Southwest Texas, the bottom two-thirds of California, and in Southern Florida. In the map all counties with more than 40% of Hispanic residents are indicated by the darker shading, so it is impossible to discern the how high Hispanic percentages go. The histogram reveals that there are counties with over 90% Hispanic residents. The histogram is also useful for estimating measures of center and spread.

(c) Both visualizations are useful and a preference for one over the other most likely depends on the context in which you plan to use it. The textbook authors seem to prefer the map, so if you chose that one you can rejoice in having anticipated the authors' preferences?

2.30

(a) This distribution would most likely be symmetric, resulting in equal values of the mean ($\bar{x}$) and the median.

(b) This distribution would most likely be left-skewed (have a long tail of values towards zero) since the mean would be pulled down towards zero more than the median, resulting in a fraction where $\frac{\bar{x}}{median} < 1$.

(c) This distribution would most likely be right-skewed (have a long tail of values far from zero) since the mean would be pulled up away from zero more than the median, resulting in a fraction where $\frac{\bar{x}}{median} > 1$.