

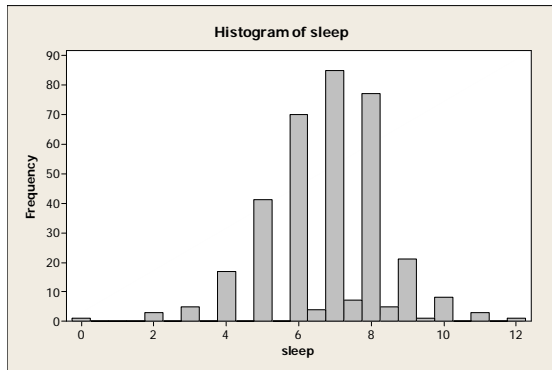
Chapters 4 and 5: Quantitative Data II

Data Set:

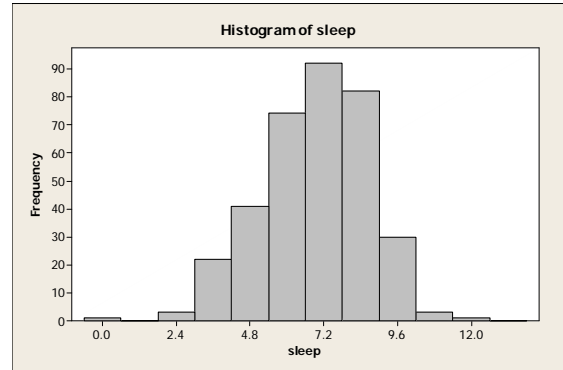
- COC Math 140 Survey Results
- <http://www.canyons.edu/faculty/morrowa/140/datasets/>
- Focus on the quantitative variables only
- Note: Questions are online at <http://classes.thegradekeeper.com/math140.php> (if you need a reminder).

1. Changing Bin Size. Adjust Minitab's number of bins for the variable **sleep** to get a good histogram. How many bins did you use? Include your new histogram. Describe why this new histogram is better.

BAD HISTOGRAM:



BETTER HISTOGRAM (12 bins):



The first histogram is bad because it hides the true pattern of the data. Some people entered half hours, which is what causes the low bars between the high bars (whole hours). The new histogram smooths the gaps, showing us the trend in the data.

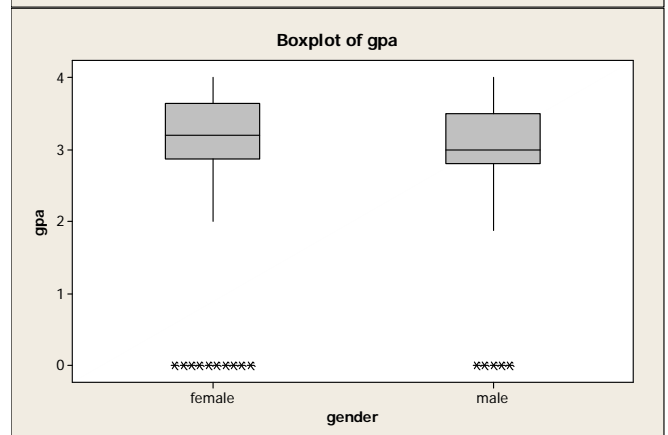
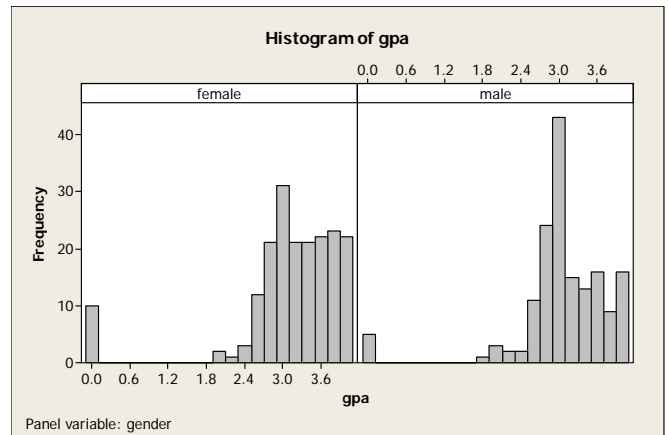
2. Side-By-Side Histograms/Boxplots. We wish to compare **gpa** of males vs. females.

- Create side-by-side histograms for GPA of males and females.
- Create side-by-side boxplots for GPA of males and females.
- Does there appear to be a difference in male/female gpa?

First, they are similar in shape. They are both skew left. They are both unimodal (if we ignore GPAs of 0.00). They both have a gap separating the 0 GPAs. They both show outliers at 0.

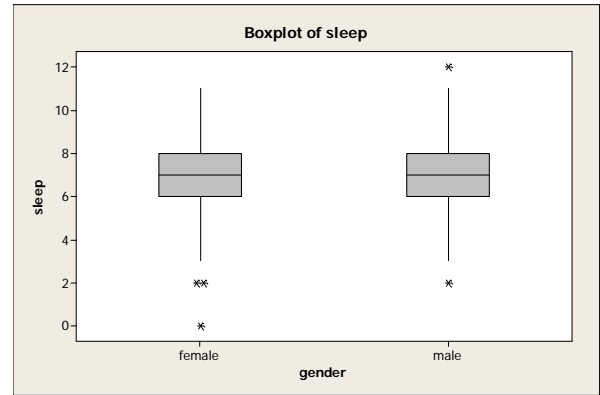
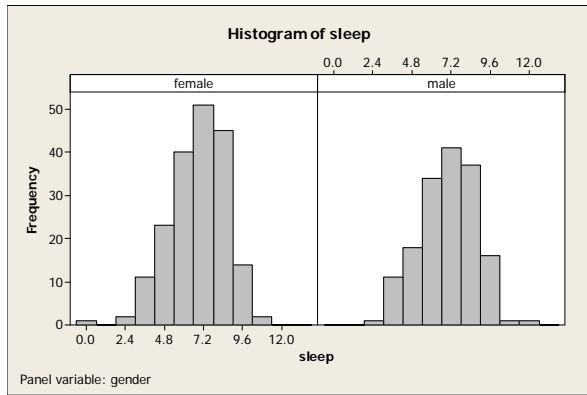
However, they have different centers and spreads. Female's have a slightly higher GPA in that they have a higher mean and median. Female GPAs are more spread out, measured both by the IQR and the SD.

Because the data is skew, the median and IQR are best to measure center and spread.



Variable	gender	Mean	StDev	Median	IQR
gpa	female	3.1015	0.8679	3.2000	0.7750
gpa	male	3.0397	0.7183	3.0000	0.7000

3. Compare **sleep** habits for males vs. females. (Include appropriate graphs, statistics, and descriptions.)



Note: I adjusted bin size on sleep.

Both distributions appear unimodal, roughly symmetric (very slight skew left for females). Females have outliers on the low end, but males have outliers on both the high end and the low end.

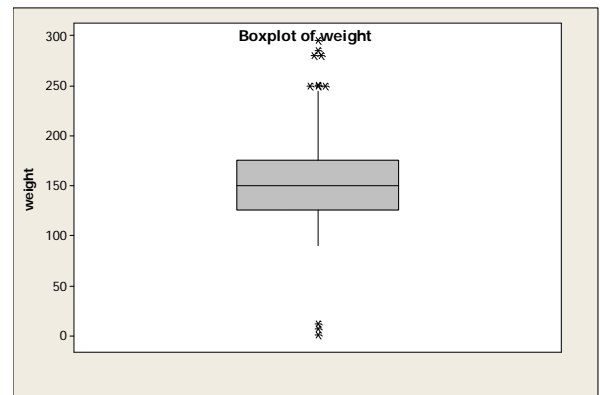
The centers (means and medians) are roughly the same.
Spread (based on IQR) is approximately the same for both.

Variable	gender	Mean	StDev	Median	IQR
sleep	female	6.748	1.615	7.000	2.000
	male	6.838	1.607	7.000	2.000

It appears that males and females get the same amount of sleep.

4. The basics. For **weight** of math 140 students, find the following:

- Examine the weights of math 140 students. Create a boxplot.
- Are there any weight values that warrant removing? Why? (Remove rows that contain weight values that you think are mistakes).

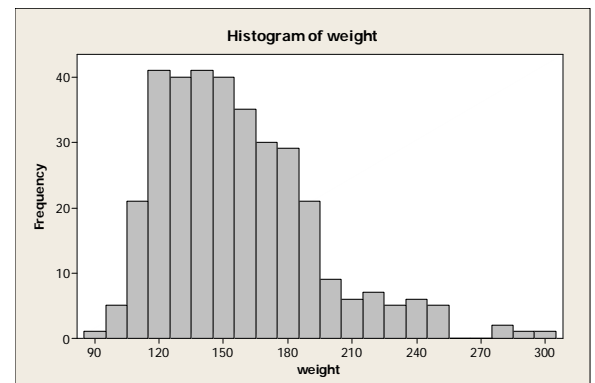


The weights on the high end are high, but reasonable. On the low end, though, I will remove any weights less than 50 pounds. I removed the following rows:
Row 1, 43, 252

Note: The rest of my answers have those rows removed. Also note: As you remove rows, the data shifts. If I delete row 1, then what used to be row 43 is now 42.

- Create a histogram. Describe the histogram.

Skew right. Almost looks bimodal (one mode at 120 and one at 140 – probably females/males—could find more by graphing them separately – next time). Gap on the high end. Potential outliers above 270.



- For the remaining data, find the

- Mean = 154.97
- Median = 150
- Standard deviation = 36.04

e) How do the mean and median compare? Why?

They are pretty close. The mean is higher because the data is right skew. Skew data pulls the mean in the direction of the tail.

5. Geometrically speaking, the mean acts like what for a histogram?

The center of balance/gravity.

6. When is the mean the best measure of center?

Symmetric data.

7. Complete the definition: The standard deviation is approximately the ____.

average distance a typical value is from the mean.

8. When is the IQR best to measure spread? When is the range best? When is the standard deviation best?

IQR: best for skew

Range: Never

SD: best for symmetric