

Chapters 4 and 5: Quantitative Data I

Data Set:

- COC Math 140 Survey Results (Spring 2010) - All Variables
- <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. The basics. For **work**, find the following:

a) Histogram



On the left, we have Minitab's default histogram. This is a bad graph (we'll explore this more next time) because some people entered higher precision than others. We need to smooth out the graph...

If you double-click on one of the bins (in Minitab), you will pull up the "edit bins" dialog box. From there, click on the binning tab, and you can alter the number of bins. I chose 10 to get the second histogram.

b) Describe the shape of the distribution.

Skew right, 3 modes (one at 0, one at about 20, and one at 40), there do not appear to be gaps/outliers

c) Median = 20

d) First Quartile = 1

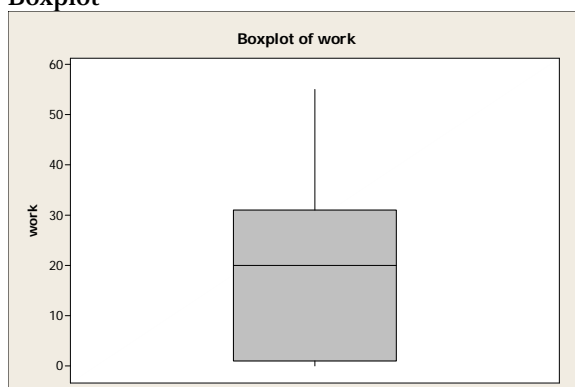
e) Third Quartile = 31

f) Range = 55

g) IQR = 30

h) Five Number Summary {0, 1, 20, 31, 55}

i) Boxplot

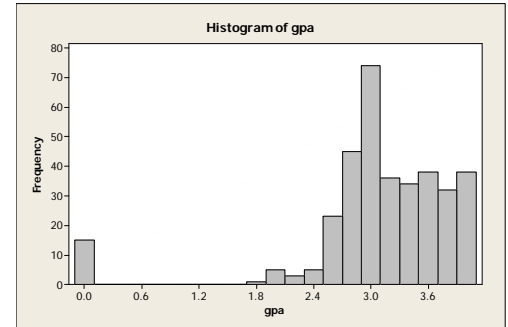


- Which is best to measure the center for the data set? Median or mode? Why?
Median. This particular variable has three modes, making it hard to use as a **center**.
- Which measure of spread is best? Range or IQR? Why?
IQR. The range is very sensitive to outliers, and this variable has outliers as evidenced by the boxplot.
- Plot a histogram for **gpa**. If we are studying GPA at COC, why might we want to remove observations that reported 0 for GPA? Remove them and re-plot the histogram. How does the shape of the distribution change?
*Note: To sort the data by GPA, go to the Data Menu > Sort. Select **all** the variables to sort. Under "By column", select GPA. To remove an entire row, left click on the row number to select the entire row. Then right click, and select "Delete cells".*

The first histogram is shown at the top right.

Students who reported 0 for GPA are likely new students to COC; they do not yet have a GPA. Including their results biases our analysis to the low end.

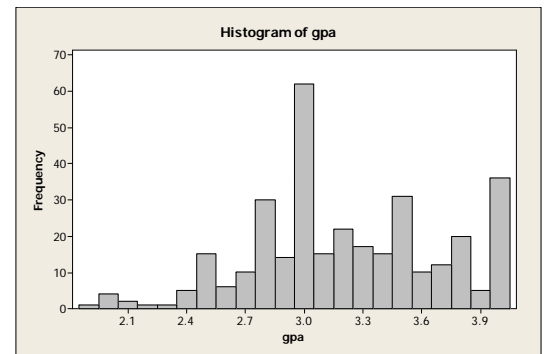
Note: When we remove results, we also run the risk of deleting real data. It is important to include analyses WITH and WITHOUT the outliers.



The histogram with 0 GPA removed is shown at the right.

We no longer see the enormous gap.

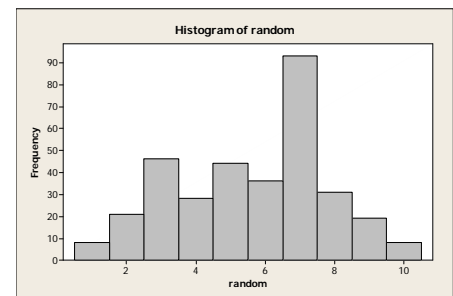
NOTE: for the rest of these questions, I left the values removed.



- Provide an appropriate graph for **random**. Is asking people to generate a number at random truly going to generate random numbers, or are some of the numbers more likely to be chosen?

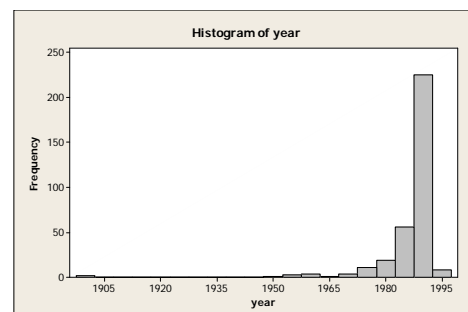
NO. If we were generating equally likely numbers, then we would have seen a uniform distribution (flat).

Instead, we can see that people are much more likely to choose the number 7.



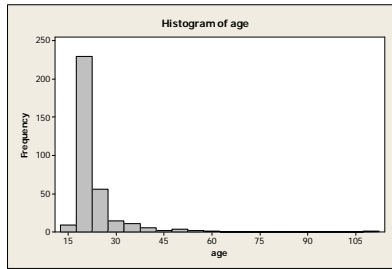
- Scavenger Hunt!!!! Find a variable that is... (include the variable name and an appropriate graph; answers may vary)
 - skew left

YEAR



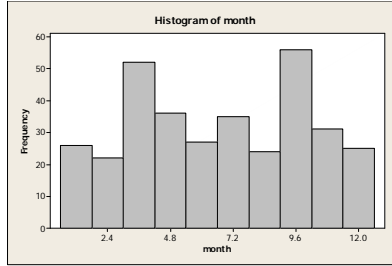
b) skew right

AGE



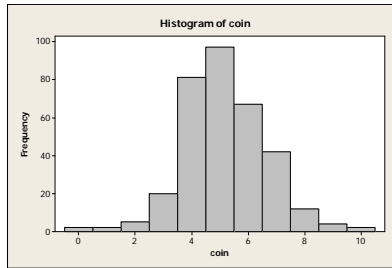
c) bimodal

MONTH



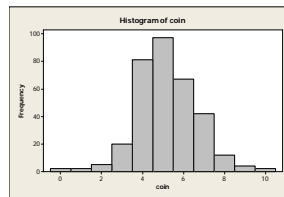
d) unimodal

COIN



e) symmetric

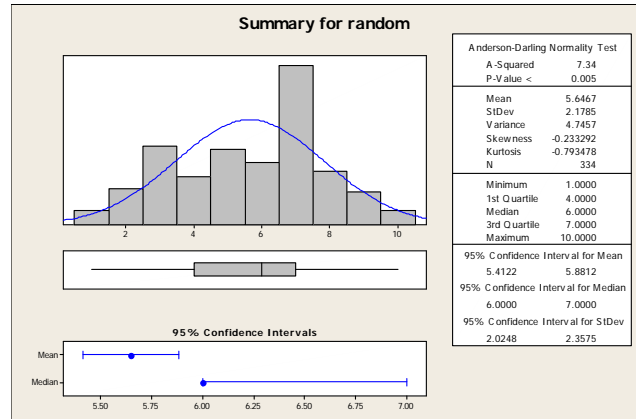
COIN



f) has no outliers

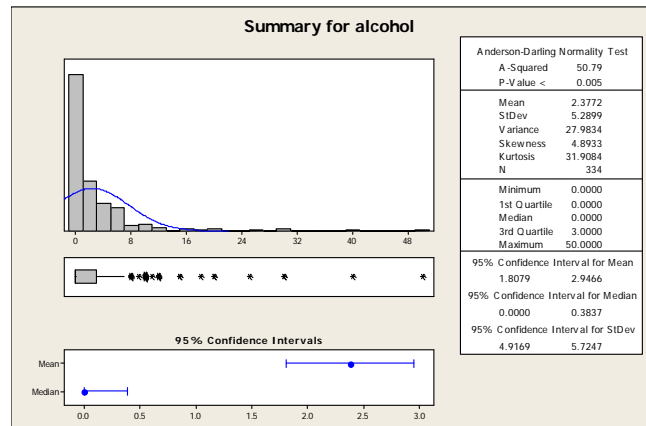
RANDOM

Note that I include the box plot so we can verify no outliers there too...



g) has outliers

ALCOHOL

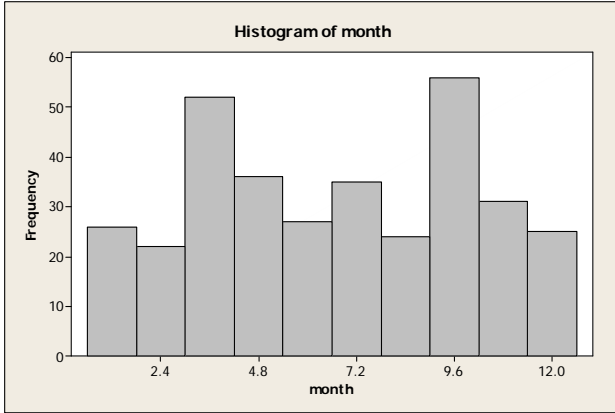


7. Months of Birth

a) Why might we expect the histogram of month of birth to be roughly uniform?

There is no real reason to expect any month to come up more often than another (holidays in all months, etc).

b) Plot and examine a histogram for **month**. Why might our histogram not be uniform? *Hint: Count the bins... how many should there have been?*



There are only 10 bins in our graph. This lumps 2 months together twice (if you hold your mouse over a bar in Minitab, you can see the bin boundaries for that bar).

Note: even changing the number of bins to 12 doesn't fix things because Minitab cuts the bins at weird locations. We can fix it by marking the bins breaks ourselves, or we can treat the data as categorical and graph it as a bar chart.