AP STATS Chapter 1 Notes

Friday Sept 11

Exploring Data

Individual- objects described by a set of data (what is on the x-axis)

Variable – characteristic of the individual

Categorical variable- places individuals in groups – non numerical

Quantitative variable- numerical values, one can average this data

Distributions- what values the variables take

Bar Graph is a graph that represents categorical data. The bars can be in any order and they do not touch.

Dot Plot is a graph that uses dots to show each piece of data

Enrollment in Introductory Courses at Union University

Read pages 4-10 and do problems 1-6 and finish getting to know you activity
Monday Sept 14

Graphs are the first steps in looking at data. It gives a visual of the data.

S – Shape

Symmetrical and Skewed Distributions

- Symmetrical
- Skewed right
- Skewed left

O-Outliers Data that appears to fall outside of the overall pattern
C-Center  The median of the data

S-Spread  The range (high – low)

**SOCS** use this acronym to remember what needs to be done for each graph description

Ways to organize data

Dot plot

Stem plot (stem and leaf plot or split stem)

Data set: 12, 13, 21, 27, 33, 34, 35, 37, 40, 40, 41.

<table>
<thead>
<tr>
<th>stem</th>
<th>leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 3</td>
</tr>
<tr>
<td>2</td>
<td>1 7</td>
</tr>
<tr>
<td>3</td>
<td>3 4 5 7</td>
</tr>
<tr>
<td>4</td>
<td>0 0 1</td>
</tr>
</tbody>
</table>

**Split stem**

<table>
<thead>
<tr>
<th>stem</th>
<th>leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>5, 11</td>
</tr>
<tr>
<td>4</td>
<td>0, 6, 6, 79</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

First part of stem has leaf parts between 0 and 4

The second part of the stem has leaf parts 5 and 9

Displaying quantitative variables

Histogram - similar to a bar graph, except the bars touch each other and the x-axis is done in **equal** intervals.
One way to get equal intervals is to take the range and divide into equal intervals. You choose how many intervals. You should have at least 5.

(One major error on graph, they did not put in a break in data on the x-axis. This needs to be included.)

Show how to use calculator to make a histogram and input data (page 21)

Percentile

The pth percentile of a distribution is the percent of observations that fall below

Relative cumulative frequency graph (ogive)

The graph starts at 0% and ends at 100%.

Look at graphs page 30
Time plots

Time always goes on the x-axis, showing change over time.

Look for patterns and deviations from the pattern

Trend- long term upward or downward movement over time

Seasonal variation- pattern that repeats itself at regular intervals

Homework Read pages 11-27 and work problems 8,9,10,15,16, and 20

Quiz over section 1.1 on Wednesday
Tuesday Sept 15

Answer any questions from homework assignment.

Discuss getting to know you activity

Dangers of bread activity

Submit a summary of the variables contained in the article, answering: What are the variables?
Are the variables categorical? quantitative? continuous? Discrete?

Please give as many variables as possible.
Wednesday Sept 16

Quiz over section 1.1

Homework start when done with quix

Read pages 27-34 problems 23,24,28
Thursday Sept 17

Describing distributions with numbers

Measuring center

Median – Middle of the data

Mean- average (add everything up and divide)

Using calculator, put data in a list and do one variable statistics.

\[ x = \bar{x} \]

Mean is sensitive to the influence of extremes

The mean is pulled in the direction of the extremes.

Mean is NOT a resistant measure

Median- Put all numbers in order smallest to largest

Find the middle number. If between two numbers, then average the two numbers to find the median.

Median is resistant to extremes
If mean and median are the same, then the data is symmetrical.

If mean is greater than the median, then the data is skewed right.

If the mean is less than the median, then the data is skewed left.

Measuring spread

**Range** - High – low (difference) this will tell us the spread of variability.

Box plots (5 number summary)

```
Min   Quartile 1   Median   Quartile 3   Max
     (middle of lower)   (Middle of upper)
```

25% of all data falls into each of the categories.

Make sure you place a scale below the box plot.

*(Show an example on the inspire)*
A box and whisker plot allows us to see how each 25% of the data is distributive. We are normally concerned with the middle 50%.

Inter quartile range (IQR)

\[ IQR = Q_3 - Q_1 \]

Test for outliers

1.5(IQR) First find this value

Then \( Q_1 - (1.5IQR) \) – if any data points are lower than this number, they are outliers

\[ Q_3 + (1.5IQR) \] – if any data points are higher than this number, they are outliers

Modified box plot – same as a box plot, except outliers are noted as points instead of part of the whisker

Show how to use calculator.

Show how to do double box plot

Homework Read pages 37-46 problems 31,34,36,39
Friday Sept 18

Measuring spread

Standard deviation

How far observations fall from the mean

Smaller standard deviation, data is clustered close to center

Larger standard deviation, data is more spread out

**VARIANCE** - $S^2$ Average of the squares of the distance it is from the mean

$$S^2 = \frac{\sum_{k=1}^{n}(x_k - \bar{x})^2}{n - 1}$$

**NEED TO KNOW**

******Standard deviation is ******

Standard deviation

$$S = \sqrt{\frac{\sum_{k=1}^{n}(x_k - \bar{x})^2}{n - 1}}$$

You will never do this by hand. It is done with the calculator using a list of data and one variable statistics!!!
Sum of \((x_i - x)\) is zero

Why \(n-1\) (degrees of freedom)

Because the sum of \((x_i - x)\) has to equal zero, one number has no freedom. The rest can be anything so the degrees of freedom is \(n - 1\).

Properties of standard deviation

\(S\) is used for spread when mean is chosen as center

(range is used for spread when median is chosen)

\(S = 0\) when there is no spread.

Example data set, 5,5,5,5,5

Is \(s \neq 0\) then \(s > 0\) (can never be negative)

\(S\) is not resistant which means outliers influence the spread.

When using skewed data, 5 number summary is better choice than mean and standard deviation.

When symmetric or normal use mean and standard deviation

Homework Read 49-52 and do problems 40, 41, 43
Monday September 21

Changing unit of measurement

Linear transformations

Changing units- original variable x

New variable $x_{\text{new}}$

$$x_{\text{new}} = a + bx$$

$a$ = constant- moves (shifts) whole graph left or right

$b$ – multiply by a positive constant changes the size of the measurement (affects $\uparrow$)

Look at problem page 53 – 54

Put data in calculator page 55

List 1 data

list 2 multiply L1 by 110% (1.1)

Find mean and standard deviation for both lists. (use one variable statistics)

List 3 put L1 + 200,000 or L! + .2

Find mean and standard deviation

What is the original mean and standard deviation, then L2, then L3

What happens, how affected?
Adding a constant to each observation does not change the spread (range or standard deviation)

Linear transformations – do not change the shape of the distribution

Multiply by b – mean/median x b, spread and standard deviation or IQR x b

Adding by a – mean/median add a, spread and standard deviation stays the same

Homework read pages 53-55 do problems 44-46

Tuesday September 22

Comparing distributions

Classwork on comparing graphs and matching histograms and box plots

Homework read pages 56-61 problems 48,49

Wednesday September 23

Decisions through data

Class work answer all questions and then read pages 64-66 and do problems 60,63,66,67

Hand out AP Set for chapter 1
Thursday September 24

AP set Due – Discuss how to do

Go over applet from book on graphs

Homework Review worksheet

Friday September 25

Chapter 1 exam