Stat 573
4-2-15

Measures of variability or dispersion

Data 1	Data 2
60	79
80	80
100	81

R = the range = largest value − smallest

$R = 100 - 60$	$R = 81 - 79$
$= 40$	$= 2$

S^2 = sample variance

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$S = \sqrt{S^2}$ = sample standard deviation

MAD = mean absolute deviation

$$= \sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{n}$$

Combining measures of location & dispersion

(3)

City A	City B
$\bar{x} = \$200{,}000$	$\bar{x} = \$500{,}000$
$s = \$10{,}000$	$s = \$20{,}000$

$CV = \frac{s}{\bar{x}}$ = Coefficient of variation

$CV = \frac{10{,}000}{200{,}000} = .05 = 5\%$ | $\frac{20{,}000}{500{,}000} = .04 = 4\%$

(4)

Your score on a standardized test was 600

$(\bar{x} = 500, s = 50)$ $\quad z = \frac{600-500}{50} = 2$

Your friend scores 620 the next year.

$(\bar{x} = 550, s = 40)$ $\quad z = \frac{620-550}{40} = 1.75$

Standardization creates z-scores

$$z = \frac{x_i - \bar{x}}{s}$$

Box plot (John Tukey)

Data set:

3	2
4	
5	5
6	566
7	145
8	5

Locate Q_1, Q_2, Q_3

↑ median

To find any percentile, compute the index

$i = (n+1)p$

If i is a whole number, take the i^{th} value on your ordered list.

If i is a fraction, average the closest 2 items

$n = 9$

Q_1: $i = (9+1)(.25) = 2.5$

$\frac{55+65}{2} = 60 = Q_1$

Q_2: $i = (9+1)(.5) = 5$ $\quad Q_2 = 66$

Q_3: $i = (9+1)(.75) = 7.5$

$\frac{74+75}{2} = 74.5 = Q_3$

Fences: IQR = interquartile range = $Q_3 - Q_1$

$= 74.5 - 60 = 14.5$

(7)

lower fence = LF

$= Q_1 - 1.5 \cdot IQR$

$= 60 - 1.5(14.5) = 38.25$

upper fence = UF

$= Q_3 + 1.5\,IQR$

$= 74.5 + 1.5(14.5) = 96.25$

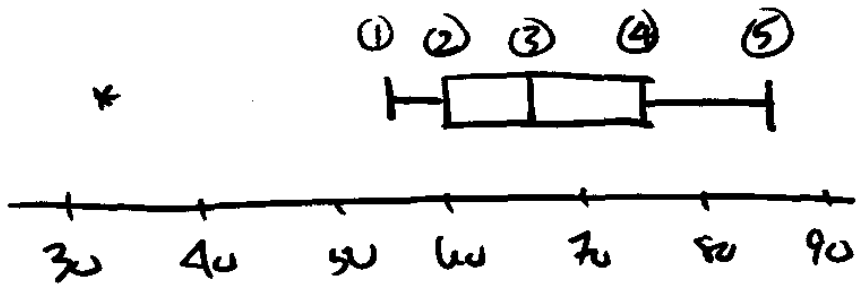
The fences determine boundaries

$\Rightarrow$ 32 is an outlier

(8)

Collect 5 pieces of info

① smallest non-outlier = 55
② $Q_1 = 60$
③ $Q_2 = 66$
④ $Q_3 = 74.5$
⑤ largest non-outlier = 85



HW #1 is attached

Due 4/9/15

24. CO_2 Emissions The following data represent the carbon dioxide emissions per capita (total carbon dioxide emissions, in tons, divided by total population) for the countries of Western Europe in 2004.

(a) Determine and interpret the quartiles.
(b) Is the observation corresponding to Luxembourg, 6.81, an outlier?

2.34	1.34	3.86	1.40	2.08	2.39	2.68
2.64	3.44	2.07	1.67	1.61	6.81	2.21
2.09	1.64	2.87	2.38	1.47	1.53	
1.44	3.65	2.12	5.22	2.67	1.01	

Source: Carbon Dioxide Information Analysis Center

12. CO_2 Emissions Revisited The following data represent the carbon dioxide emissions per capita (total carbon dioxide emissions, in tons, divided by total population) for the countries of Western Europe in 2004. In Problem 24 from Section 3.4, you found the quartiles.

(a) Construct a boxplot.
(b) Use the boxplot and quartiles to describe the shape of the distribution.

2.34	1.34	3.86	1.40	2.08
2.64	3.44	2.07	1.67	1.61
2.09	1.64	2.87	2.38	1.47
1.44	3.65	2.12	5.22	2.67
2.68	2.39	6.81	1.53	1.01
2.21				

Source: Carbon Dioxide Information Analysis Center

31. Which Car Would You Buy? Suppose that you are in the market to purchase a car. With gas prices on the rise, you have narrowed it down to two choices and will let gas mileage be the deciding factor. You decide to conduct a little experiment in which you put 10 gallons of gas in the car and drive it on a closed track until it runs out gas. You conduct this experiment 15 times on each car and record the number of miles driven.

Car 1				
228	223	178	220	220
233	233	271	219	223
217	214	189	236	248

Car 2				
277	164	326	215	259
217	321	263	160	257
239	230	183	217	230

Describe each data set. That is, determine the shape, center, and spread. Which car would you buy and why?