Math 2311

Bekki George – bekki@math.uh.edu

Office Hours: MW 11am to 12:45pm in 639 PGH

Online Thursdays 4-5:30pm

And by appointment

Class webpage:

http://www.math.uh.edu/~bekki/Math2311.html

★ Discussion Board ★
on CASA

This week:

- Population
- Sample
- Mean
- Median
- Mode → what use for categorical data (can't find mean or median of cat. data)
- Five number summary = min, Q1, Q2, Q3, max
- IQR = Q3 - Q1
- Variance
- Standard Deviation
- Some graphs

center: Mean, Median, Mode

spread: IQR, Variance, Standard Deviation

range = max - min

Categorical	Quantitative
bar	histogram
pie	stem/leaf plots
	dot plots
	box plots
	ogive

Some review examples:

1. Which of the following is quantitative data?
   a. Hair color
   b. Letter grade for a class
   c. Rating of movie on a scale of 1 to 5
   d. Numerical grade on a test
   e. None of these

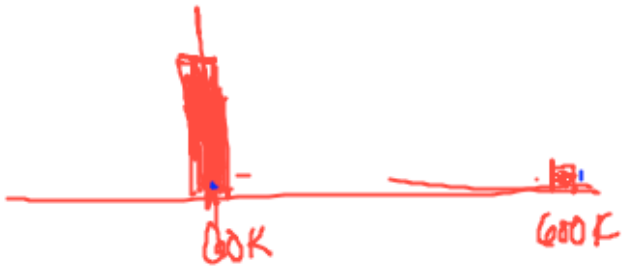
⇐ "iffg" could be Quant/discrete ... ,85,86,87, ... ,100

2. In #1, the quantitative data is
   a. Discrete
   b. Continuous

Ex. of continuous: time, temp., height, distance

3. Suppose we were looking at salaries for a small company. Most employees make the same amount per year but the CEO makes 10 times that amount. Which is larger:
   a. mean
   b. median

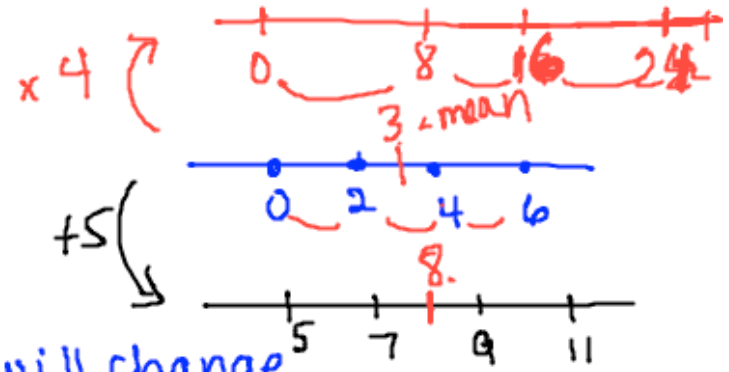
Ex: 15 ppl. make $60K
1 CEO makes $600K



⇐ mean is not a good measure of center

4. Which of the following lists would have the largest variance?
   a. 3, 3, 3, 3, 3
   b. 1, 2, 3, 4, 5
   c. 0, 2, 4, 6, 8
   d. 1, 3, 7, 10, 22

more spread out



5. Which of the following lists would have 0 variance?
   a. 3, 3, 3, 3, 3
   b. 1, 2, 3, 4, 5
   c. 0, 2, 4, 6, 8
   d. 1, 3, 7, 10, 22

= no variance

mean will change

6. If a constant value is added to all data in a sample, the new variance
   a. will be unchanged
   b. will be equal to the old variance plus the constant value.
   c. will be equal to the old variance plus the square of the constant value.
   d. none of these

↑ st. dev. no change in spread

7. Variance and standard deviation are measures of
   a. spread
   b. center
   c. dispersion
   d. spread and dispersion mean the same thing

Variance = average squared distance from mean

st. dev = $\sqrt{\text{variance}}$

Calculator

TI - 83 or 84

STAT - EDIT

- enter list into $L_1, L_2, \ldots$

R - Studio

download R first

then download R-Studio

★ how to make a list ★

name = c( )

↑ list goes here

Box plot.

Data : 5, 10, 12, 7, 8, 16, 4, 9, 11, 7, 7, 6, 5, 13, 12, 9, 1

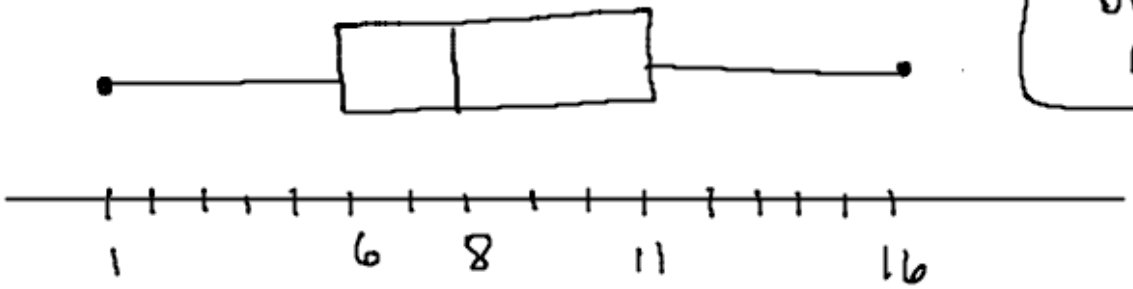
R Studio

```
> data1=c(5,10,12,7,8,16,4,9,11,7,7,6,5,13,12,9,1)
> sort(data1)
 [1]  1  4  5  5  6  7  7  7  8  9  9 10 11 12 12 13
[17] 16
>
```

fivenum(data1)

1 6 8 11 16

min Q1 Q2 Q3 max



Outliers

IQR = 11 − 6 = 5

(1.5)IQR = 7.5

Q1 − 7.5 = −1.5

Q3 + 7.5 = 18.5

[−1.5, 18.5]

outside this is an outlier

> boxplot(no

```
12  \newcommand{\note}[1]{\textbf{\textcolor{NoteColor}{#1}}}
13  \newcommand{\str}{{\color{Head}\rule{\textwidth}{.05cm}}}
14  \newcommand{\mytitle}[1]{\begin{center}{\huge\bfseries\tex
    tcolor{Head}{#1}}\end{center}}
```

1:1 TeX

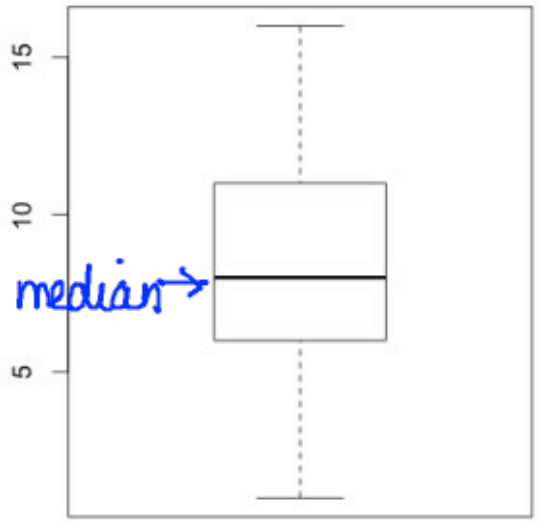
b	numeric[9]
bodyb	lm[12]
bodymod	lm[12]
cans	numeric[15]

Console ~/Desktop/UH Online Files/ semesters/spring13/3339/R_data/

```
R version 2.15.1 (2012-06-22) -- "Roasted Marshmallows"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/Desktop/UH Online Files/
semesters/spring13/3339/R_data/.RData]

> data1=c(5,10,12,7,8,16,4,9,11,7,7,6,5,13,12,9,1)
> sort(data1)
 [1]  1  4  5  5  6  7  7  7  8  9  9 10 11 12 12 13
[17] 16
> fivenum(data1)
[1]  1  6  8 11 16
> boxplot(data1)
>
```

Files Plots Packages Help

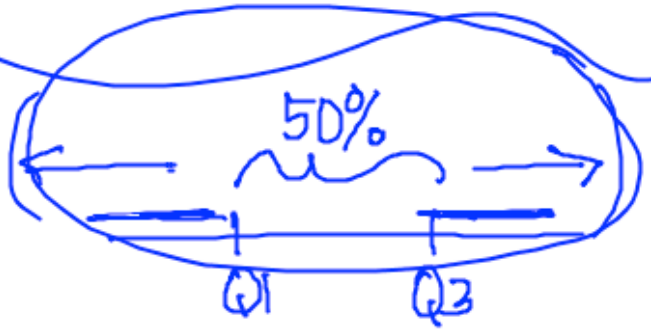
Zoom Export Clear All



2. The table below records the lengths of reign (in years) of some of the rulers of England and Great Britain.

William I	21	Richard I	10	Edward III	50	Edward IV	22
William II	13	John	17	Richard II	22	Edward V	0
Henry I	35	Henry III	56	Henry IV	13	Richard III	2
Stephen	19	Edward I	35	Henry V	9	Henry VII	24
Henry II	35	Edward II	20	Henry VI	39	Henry VIII	38

a. What is the mean of the data?
b. What is the median of the data?
c. What is the mode of the data?
d. Which is the better description of the center of this data and why?



1.4 ex 2

Outliers

① find IQR (IQR = Q3 - Q1)

② mult. the IQR by 1.5

③ subtract this from Q1 & add to Q3

9. A data set has only positive values. If the largest value of a data set is doubled, which of the following is not true?

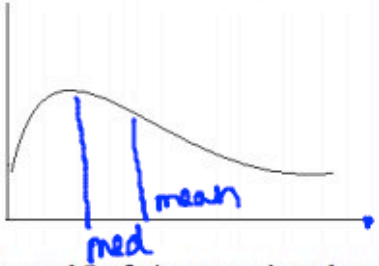
A. The mean increases. T
B. The range increases. T
C. The interquartile range increases. F
D. The standard deviation increases. T
E. All of these are true.

orig. 1, 2, 3, 5, 7

new 1, 2, 3, 5, 14

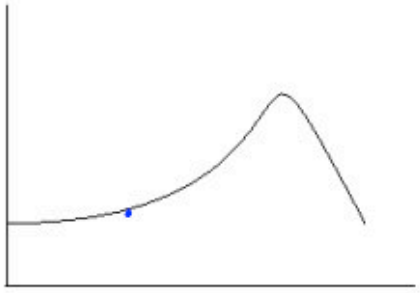
1.5

- Skewed Right (or positive skew) – longer tail on the right side. The mean will be larger than the median in a skewed right distribution.



med < mean

- Skewed Left (or negative skew) – longer tail on the left side. The mean will be smaller than the median in a skewed left distribution.



med > mean

How to be successful:

1. Read book
2. watch videos
3. Keep up w/ homework and quizzes
4. check discussion board & class page
5. ask questions (on discussion board and in problem session)