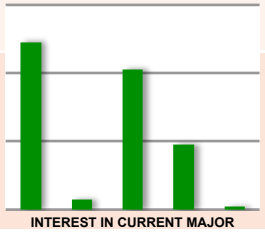
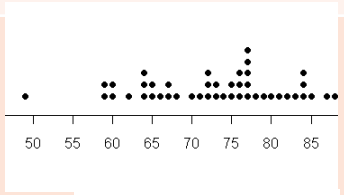
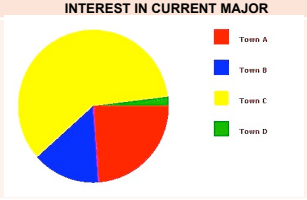
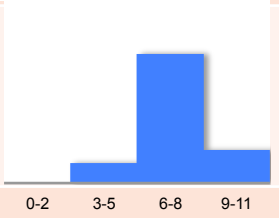
COMM215

First the Foundation, then Innovation

LESSON TWO
DESCRIPTIVE STATISTICS

SAMIE L.S. LY

- **Qualitative and Quantitative Data**
- **Describe Central Tendency**
- **Measures of Variation**
- **Percentiles, Quartiles, and Box and Whiskers.**
- **Midterm Preparation**

QUALITATIVE DATA	QUANTITATIVE DATA
FREQUENCY DISTRIBUTION	
RELATIVE FREQUENCY PERCENT FREQUENCY DISTRIBUTION	
BAR CHART 	DOT PLOT 
PIE CHART 	HISTOGRAM 
	CUMULATIVE DISTRIBUTION

Bowerman, et al. (2017) pp. 55

DESCRIBING QUALITATIVE DATA

- FREQUENCY DISTRIBUTION
- RELATIVE FREQUENCY DISTRIBUTION
- PERCENT FREQUENCY DISTRIBUTION
- BAR CHARTS, PIE CHARTS

Bowerman, et al. (2017) pp. 55

INTEREST IN YOUR MAJOR?

FREQUENCY DISTRIBUTION is

a tabular summary of data showing
the number of items in each of several non-overlapping classes.

INTEREST IN MAJOR	FREQUENCY
WELL-PAYED JOB	49
PROMOTION	3
GRAD SCHOOL	41
NO CLUE	19
FREE	1
	113

Bowerman, et al. (2017) pp. 55

RELATIVE FREQUENCY

FREQUENCY ÷ TOTAL N

41 ÷ 113

= 0.363

INTEREST IN MAJOR	FREQUENCY	RELATIVE FREQUENCY
WELL-PAYED JOB	49	
PROMOTION	3	0.027
GRAD SCHOOL	41	0.363
NO CLUE	19	
FREE	1	0.009
TOTAL	113	

Bowerman, et al. (2017) pp. 55

RELATIVE FREQUENCY

FREQUENCY ÷ TOTAL N

INTEREST IN MAJOR	FREQUENCY	RELATIVE FREQUENCY
WELL-PAYED JOB	49	
PROMOTION	3	0.027
GRAD SCHOOL	41	0.363
NO CLUE	19	
FREE	1	0.009
TOTAL	113	

Bowerman, et al. (2017) pp. 55

INTEREST IN YOUR MAJOR?

PERCENT FREQUENCY: RELATIVE FREQUENCY X 100

INTEREST IN MAJOR	FREQUENCY	RELATIVE FREQUENCY	PERCENT FREQUENCY
WELL-PAYED JOB	49	0.434	43.363
PROMOTION	3	0.027	2.655
GRAD SCHOOL	41	0.363	
NO CLUE	19	0.168	16.814
FREE	1	0.009	0.885
TOTAL	113	1.000	100.000

INTEREST IN YOUR MAJOR?

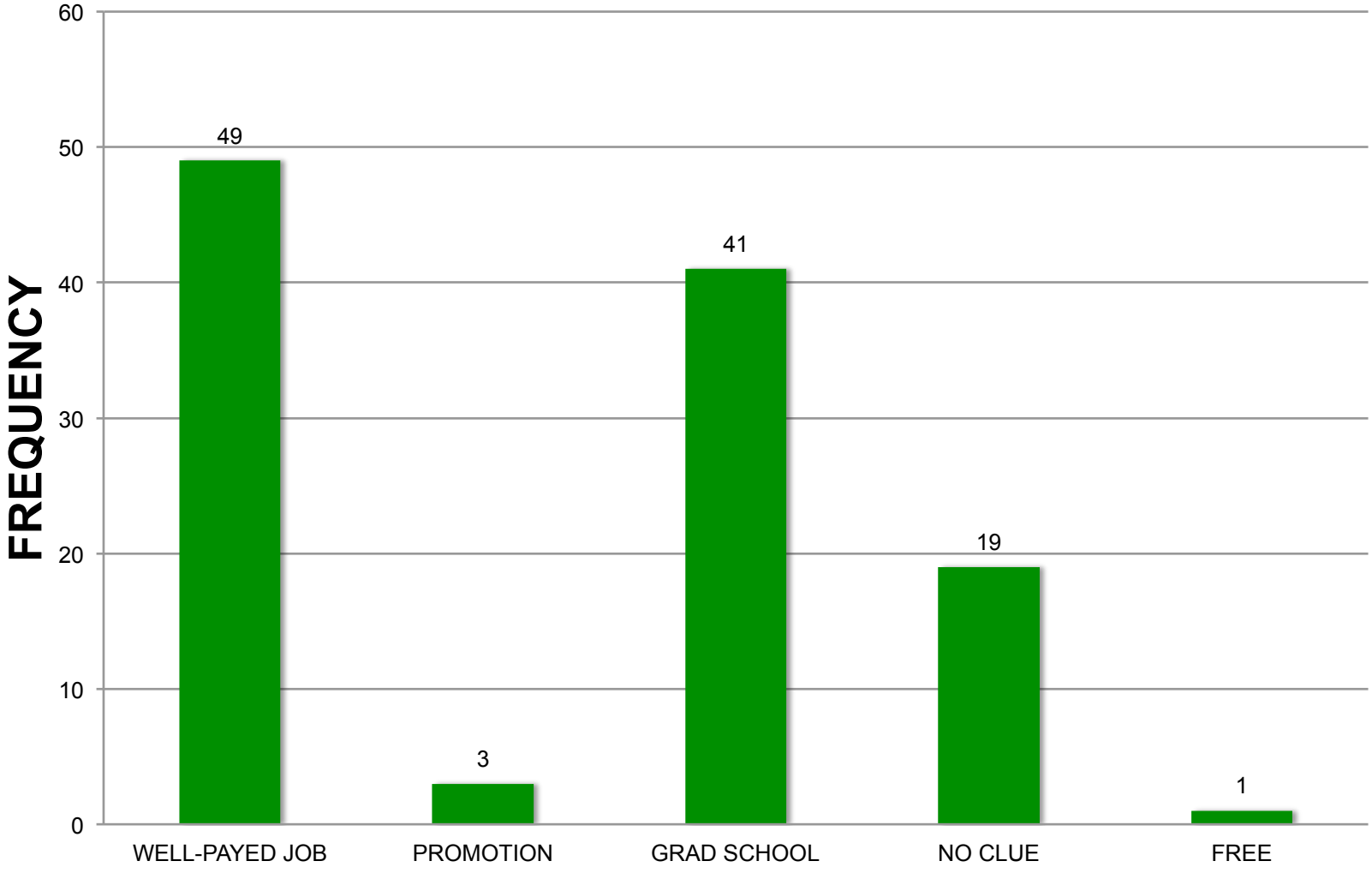
PERCENT FREQUENCY: RELATIVE FREQUENCY X 100

0.434 x 100 = 43.363

INTEREST IN MAJOR	FREQUENCY	RELATIVE FREQUENCY	PERCENT FREQUENCY
WELL-PAYED JOB	49	0.434	43.363
PROMOTION	3	0.027	2.655
GRAD SCHOOL	41	0.363	
NO CLUE	19	0.168	16.814
FREE	1	0.009	0.885
TOTAL	113	1.000	100.000

BAR GRAPH

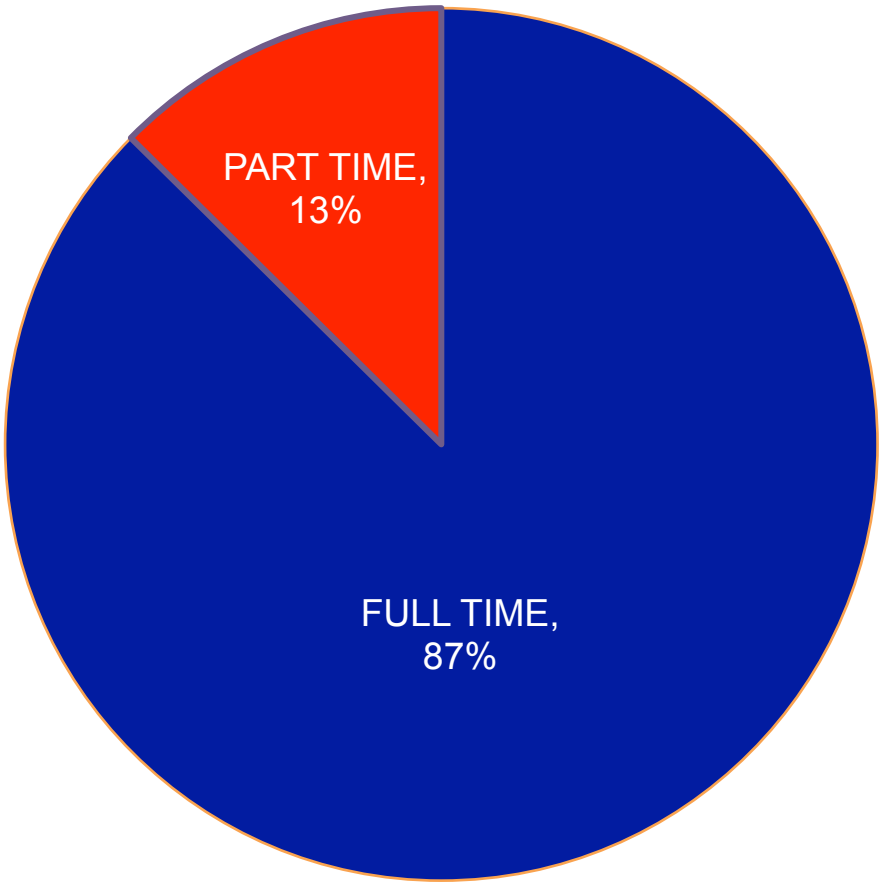
Sample Size: **120** STUDENTS



Bowerman, et al. (2017) pp. 56

COMM215
First the Foundation, then Innovation

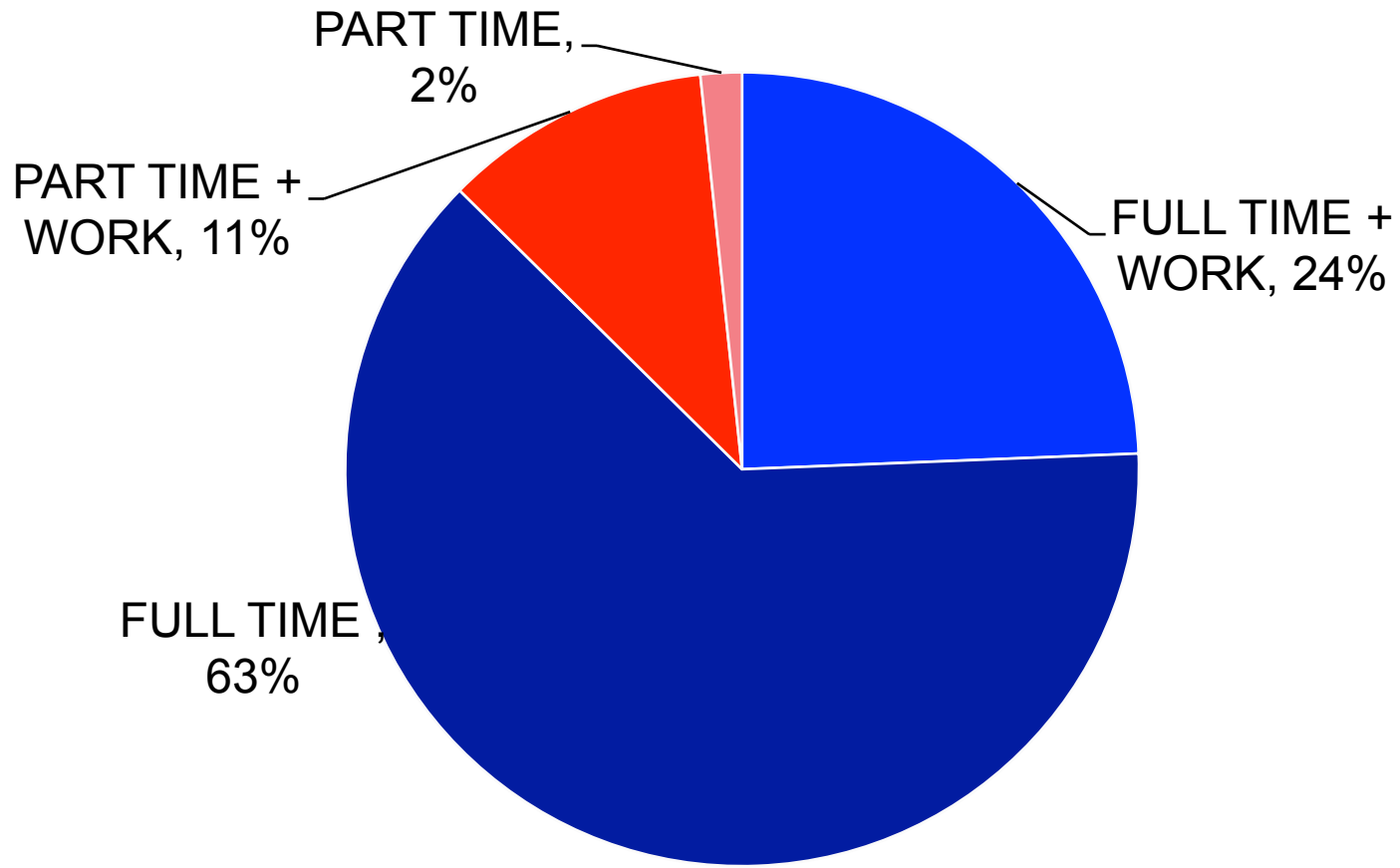
UNIVERSITY STATUS



Bowerman, et al. (2017) pp. 58

UNIVERSITY STATUS + WORK

Sample size: **120** STUDENTS



Bowerman, et al. (2017) pp. 58

SUMMARIZING QUALITATIVE DATA

Qualitative Data

- Grouped by specific categories.
- Used either the NOMINAL scale or the ORDINAL scale of measurement.

- NOMINAL: 1: Male, 2: Female
- ORDINAL: Excellent, Very Good, Good, Poor

Bowerman, et al. (2017) pp. 55

SUMMARIZING QUANTITATIVE DATA

- **FREQUENCY DISTRIBUTION**
  - Relative Frequency
  - Percent Frequency
- **CUMULATIVE DISTRIBUTION**
  - Cumulative Relative Frequency
  - Cumulative Percent Frequency
- **HISTOGRAM**

Bowerman, et al. (2017) pp. 61

BUILDING A HISTOGRAM

0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20

[ ] [ ]
[ ] [ ]
[ ] [ ]
[ ]

NUMBER OF CLASSES (K)

Possible hours of sleep in a day: 0 to 20.

***RECOMMENDED TO USE 5 TO 20 CLASSES**

GOAL: use enough classes to show a variation in data

WIDTH OF CLASSES

$$\frac{\text{[LARGEST VALUE-SMALLEST VALUE]}}{\text{NUMBER OF CLASSES}}$$

$$\frac{20 - 0}{7} = 2.85$$

Bring this number to the next integer

$$2^K > n$$

Bowerman, et al. (2017) pp. 62

SLEEP TIME

NUMBER OF CLASSES

I CHOSE TO HAVE 7 CLASSES,

[0-2], [3-5], [6-8], [9-11] , [12-14], [15-17],[18-20]

CLASS LIMITS

Each Item belongs to one class and not the other.

2 HOURS OF SLEEP FALLS INTO [0-2]

3 HOURS OF SLEEP FALLS INTO [3-5]

SLEEP TIME

1+8+93 = 102

SLEEP TIME PER DAY	FREQUENCY	RELATIVE FREQUENCY	PERCENT FREQUENCY
0-2	1	0.008	0.847
3-5	8	0.068	6.780
6-8	93	0.788	78.814
9-11	16	0.136	13.559
TOTAL	118	1	100

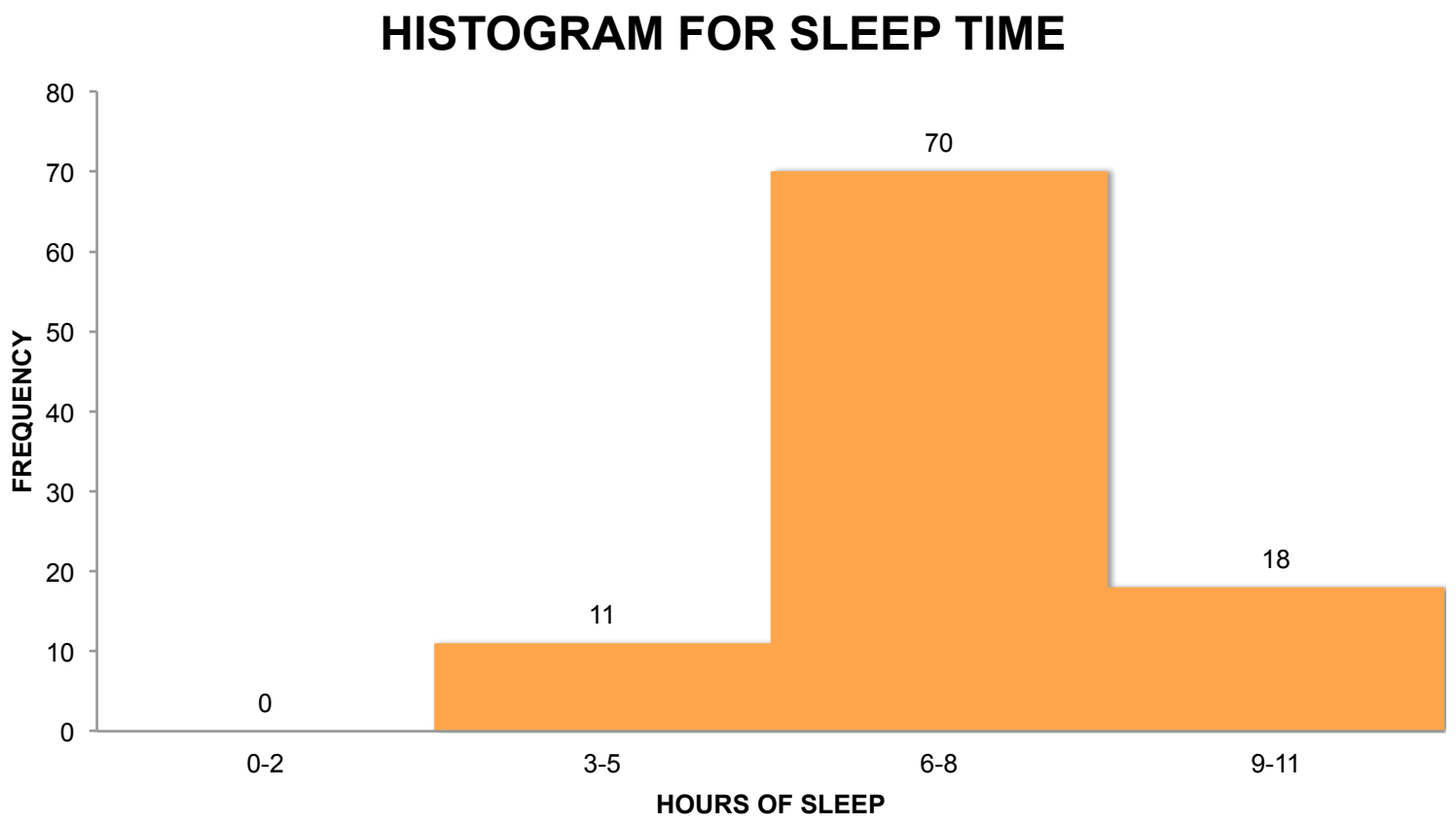
SLEEP TIME PER DAY	CUMULATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY	CUMULATIVE PERCENT FREQUENCY
LESS THAN OR EQUAL TO 2	1	0.008	0.847
LESS THAN OR EQUAL TO 5	9		7.627
LESS THAN OR EQUAL TO 8			
LESS THAN OR EQUAL TO 11	118	1.000	100.000

SLEEP TIME

SLEEP TIME PER DAY	FREQUENCY	RELATIVE FREQUENCY
0-2	1	0.008
3-5	8	0.068
6-8	93	0.788
9-11	16	0.136
TOTAL	118	1

SLEEP TIME PER DAY	CUMULATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY	CUMULATIVE PERCENT FREQUENCY
LESS THAN OR EQUAL TO 2	1	0.008	0.847
LESS THAN OR EQUAL TO 5	9		7.627
LESS THAN OR EQUAL TO 8			
LESS THAN OR EQUAL TO 11	118	1.000	100.000

SLEEP TIME

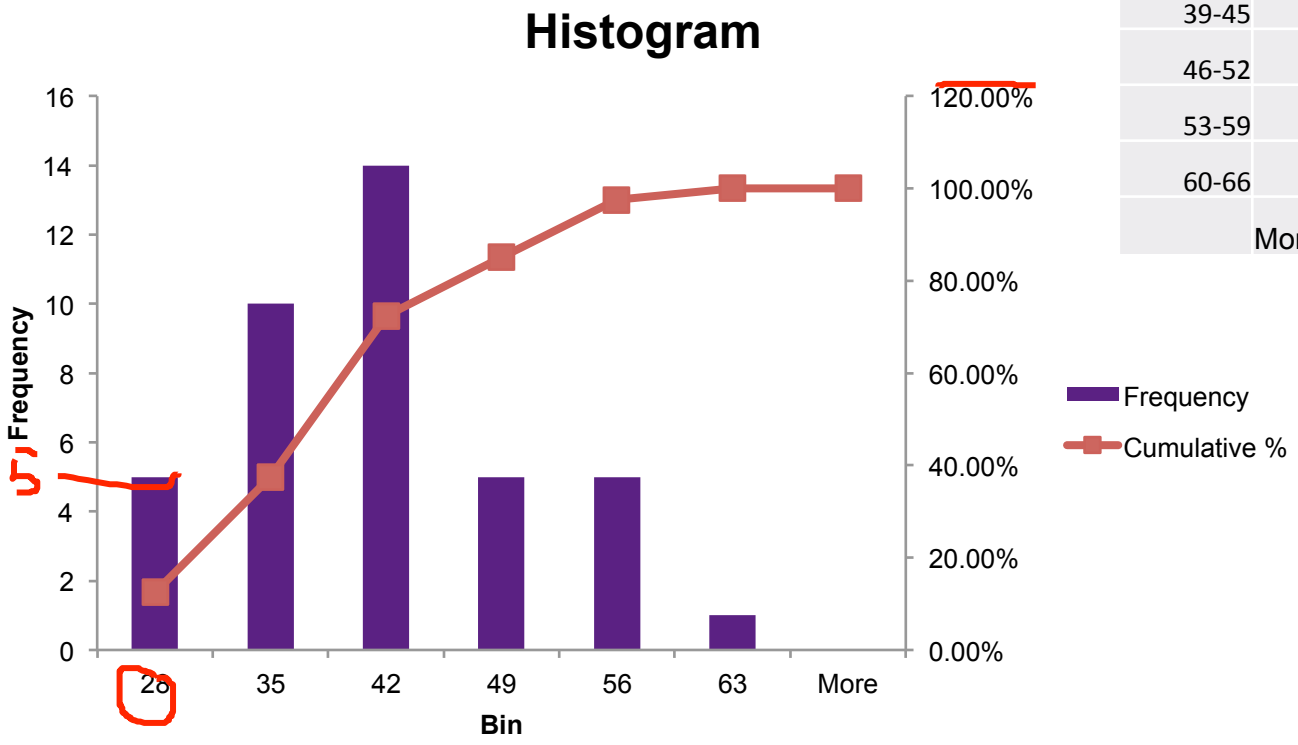


COMM215
First the Foundation, then Innovation

OGIVE

A graph of a cumulative distribution

Width	Midpoint	Frequency	Cumulative %
25-31	28	5	12.50%
32-38	35	10	37.50%
39-45	42	14	72.50%
46-52	49	5	85.00%
53-59	56	5	97.50%
60-66	63	1	100.00%
	More	0	100.00%



Bowerman, et al. (2017) pp. 71

STEM AND LEAF DISPLAY

A way to summarize number distributions

72	48	
54	59	
9	50	
9	51	
9	52	6
321	53	7
986	54	5
7775544 3221	55	556
977	56	024789
96220	57	23466679
766544 422211000	58	00034444556678889
9998887 7643333210	59	01223355567788899
743211110	60	0001223344445677789
887655544322110	61	22224444555566777789
999988875544322110	62	00112223334444667777889999
98665543333321000	63	00112234555668889999
9888877665543332211110	64	0000011222455666 89
985544200	65	12334677789
99866422220	66	00478
77664	67	00136
9331	68	8
8210	69	
	70	
9	71	
	72	
3	73	

Bowerman, et al. (2017) pp. 76

DATA SET:

25, 26, 30, 31, 32

STEM AND LEAF DISPLAY

LEAF UNIT

2	**5 6**
3	**0 1 2**

LEAF UNIT = **1**

DATA SET: **25** x 1 **26** x 1

LEAF **UNIT** = **10**

25 x 10 **26** x 10

DATA SET: 250, 260

LEAF UNIT = **0.1**

25 x 0.1 **26** x 0.1

DATA SET: 2.5, 2.6

Bowerman, et al. (2017) pp. 76

PROBLEM # 2.1

A large investment firm on Wall Street wants to review the distribution of ages of its stockbrokers. The firm believes that this information can be useful in developing plans to recruit new brokers. The ages of a sample of 40 brokers are shown here.

46	28	51	34	29	40	38	33	41	52
53	40	50	33	36	41	−25	38	37	41
36	50	46	33	− 61	48	32	28	30	49
41	37	26	39	35	39	46	26	31	35

a. Draw a stem and leaf display
b. Draw a histogram

STEM AND LEAF

CHOOSE YOUR UNIT

LEAF UNIT: 1

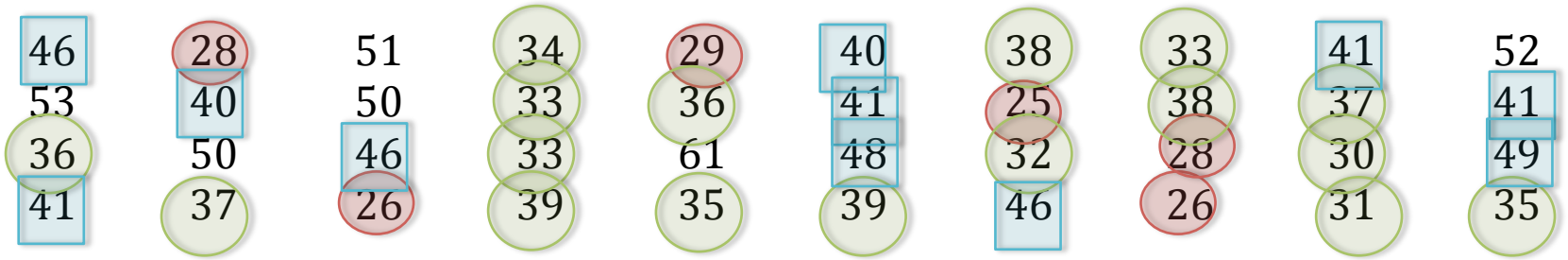
2| 5 6 6 8 8 9

3| 0 1 2 3 3 3 4 5 5 6 6 7 7 8 8 9 9

4| 0 0 1 1 1 1 6 6 6 8 9

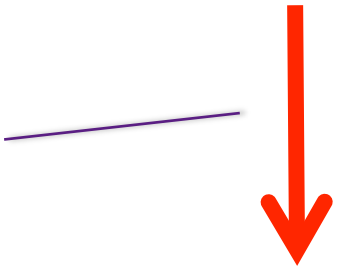
5| 0 0 1 2 3

6| 1



LEAF UNIT

DATA SET



Sometimes…You will have to cut the numbers out.

STEM & LEAF DISPLAY

DATA SET: 356, 451, 423

LEAF UNIT: 10

3|6

4|2 5

Round the data!
356 becomes 360

COMM215
First the Foundation, then Innovation

HISTOGRAM PROBLEM # 2.1

NUMBER OF CLASSES

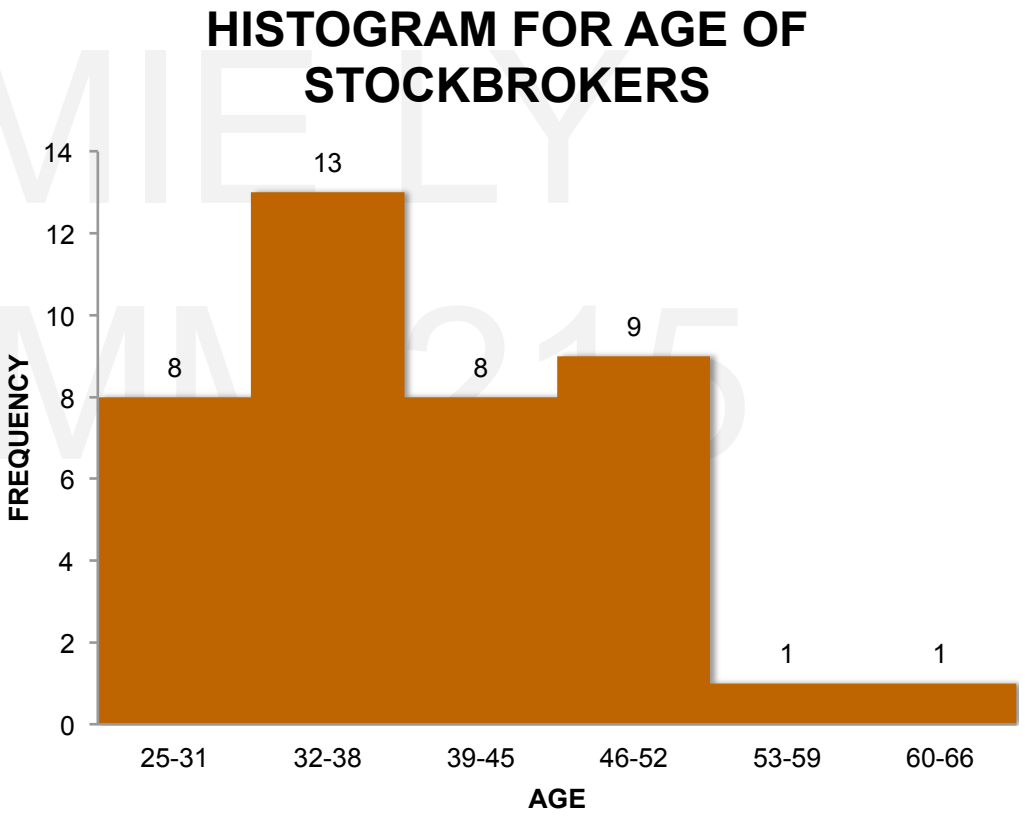
Largest value = 61 ,
Smallest value = 25

6 classes

[61-25]/6 = 6

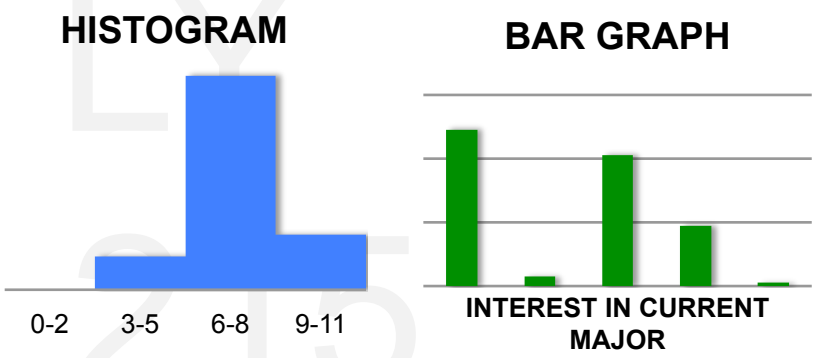
CLASS WIDTH

[25-31], [32-38],
[39-45], [46-52],
[53-59], [60-66]



PROBLEM # 2.2

What is the difference between a histogram and a bar chart? For what type of data would each be appropriate?



A histogram graphically displays

- class intervals
- class frequencies.

A bar chart displays

- frequencies for a set of categories or classes.

Histograms are appropriate for quantitative data

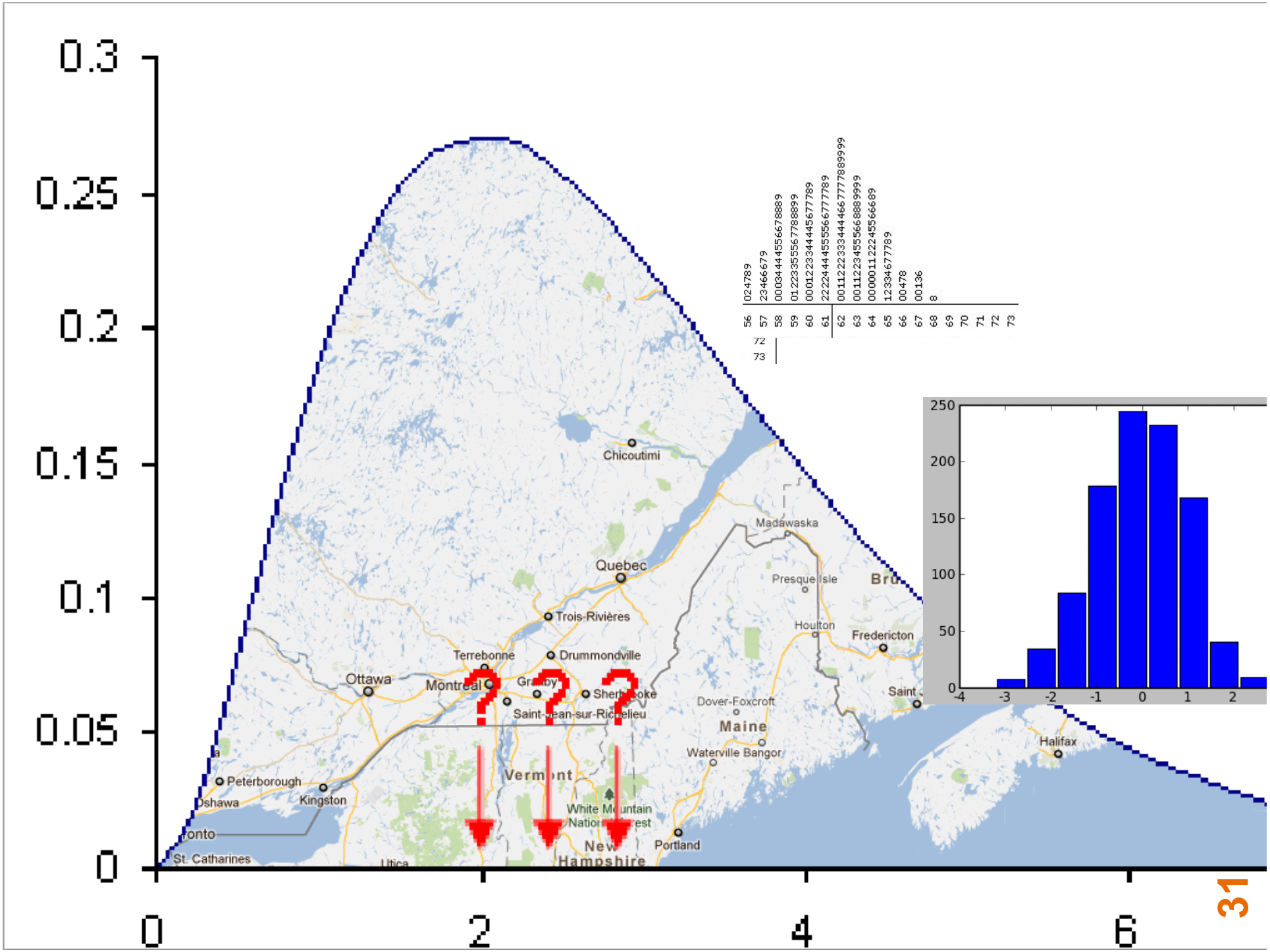
Bar charts are better for qualitative data.

- **Describe Central Tendency**
- **Measures of Variation**
- **Percentiles, Quartiles, and Box and Whiskers.**
- **Midterm Preparation**

Bowerman, et al. (2017) Chapter 3

COMM215
First the Foundation, then Innovation


MONTREAL
Where is Montreal Located?
How far from other cities is Montreal?
Chicoutimi
Quebec
Sudbury
North Bay
Trois-Rivières
Terrebonne
Drummondville
Granby
Sherbrooke
Saint-Jean-sur-Richelieu
Madawaska
Presque Isle
New Brunswick
Houlton
Fredericton
Moncton
Saint John
Nova Scotia
Halifax
Dover-Foxcroft
Maine
Waterville Bangor
Vermont
Barrie
Kawartha Lakes
Peterborough
Kingston
Markham
Oshawa
Mississauga
Toronto
Cambridge
St. Catharines
Rochester
Utica
Hamilton
Sarnia
London
Norfolk
Buffalo
New York
Albany
Manchester
Haverhill
Lowell
Boston
Massachusetts
Detroit
Erie
Cleveland
Lorain
Akron
Mansfield
Youngstown
Canton
Pittsburgh
Pennsylvania
Altoona
Scranton
Bridgeport
New Haven
Bedford
New York
Allentown
Edison
Brentwood
Middletown
Rhode Island
Reading
Trenton
Ohio
Columbus
Lancaster
Morgantown
Philadelphia
Wilmington
Maryland
New Jersey
Parkersburg
Columbia
Dover
Rockville
Washington
Delaware
West Virginia
Harrisonburg
Huntington


0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
024789
23466679
000344445566788889
012233555677888999
000122334444556777789
222244445555667777789
001122233344446677778889999
001122345555668889999
000001122245566689
12334677789
00478
00136
8
56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
72
73
250
200
150
100
50
0
-4 -3 -2 -1 0 1 2
Chicoutimi
Quebec
Trois-Rivières
Terrebonne
Drummondville
Ottawa
Montreal
Granby
Sherbrooke
Saint-Jean-sur-Richelieu
Madawaska
Presque Isle
Houlton
Fredericton
Dover-Foxcroft
Maine
Waterville Bangor
Halifax
Peterborough
Kingston
Oshawa
Vermont
White Mountain
New Hampshire
Portland
St. Catharines
Utica
31

MEASURES OF LOCATION

- MEAN
- MEDIAN
- MODE

Bowerman, et al. (2017) pp.135

PROBLEM # 2.5

Consider the following sample of n = 7 measurements: 5, 7, 4, 5, 20, 6, 2.

a. Calculate the mean of this sample.

b. Calculate the median of this sample.

c. Eliminate the last measurement (2) and calculate the median, n= 6

d. What is the mode?

$\frac{5+6}{2} = 5.5$

2, 4, 5, 5, 6, 7, 20

MEAN…(AVERAGE)

EXTREME numbers can inflate the mean.

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Population Mean

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

Sample Mean

Bowerman, et al. (2017) pp.136

PROBLEM # 2.5

Consider the following sample of n = 7 measurements: 5, 7, 4, 5, 20, 6, 2.

a. Calculate the mean of this sample.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5 + 7 + 4 + 5 + 20 + 6 + 2}{7} = 7$$

ARRANGE NUMBERS in ORDER

MEDIAN- MIDDLE VALUE

ODD NUMBER OF N

Median is the middle value

n = 5

5/2 = 2.5

The median will be the 3rd value.

ROUND UP!

EVEN NUMBER OF N

Median is the average of the two middle values

n = 6

6/2 = 3

Take the 3rd and 4th value and make an average.

Bowerman, et al. (2017) pp.138

PROBLEM # 2.5

Consider the following sample of n = 7 measurements: 5, 7, 4, 5, 20, 6, 2.

b. Calculate the median of this sample.

c. Eliminate the last measurement (2) and calculate the median, n= 6

2, 4, 5, 5, 6, 7, 20

Median's position:

7/2 = 3.5 ROUND UP

at the 4th position

Median (m) = 5

4, 5, 5, 6, 7, 20

Median's position:

6/2 = 3

Average of 3rd and 4th position.

Median (m) = (5+6)/2 = 5.5

PROBLEM # 2.5

Consider the following sample of n = 7 measurements: 5, 7, 4, 5, 20, 6, 2.

d. What is the mode?

MODE: The **Most Repeated Number** in your sample

Mode = 5

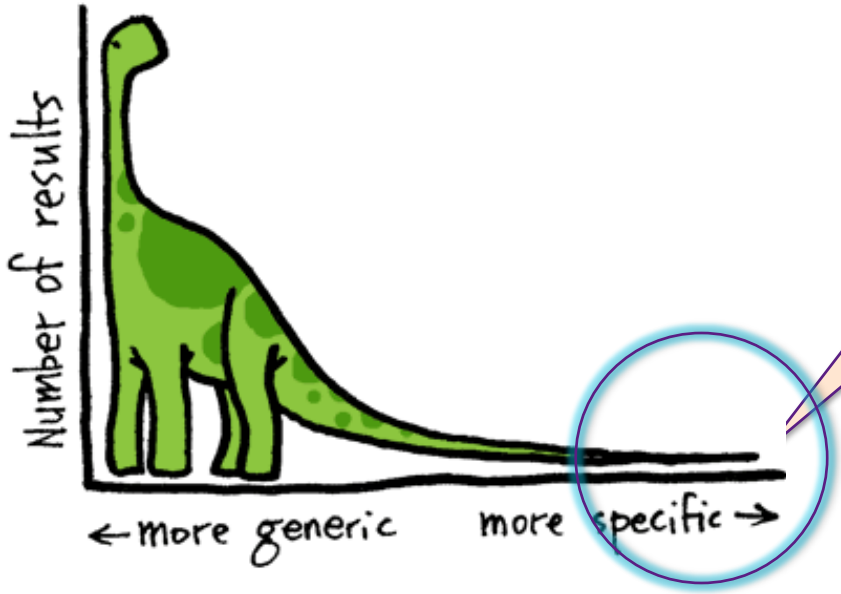
NEW Data Set: 5, 7, 4, 5, 20, 6, 2,7

Which are the modes? 5 & 7

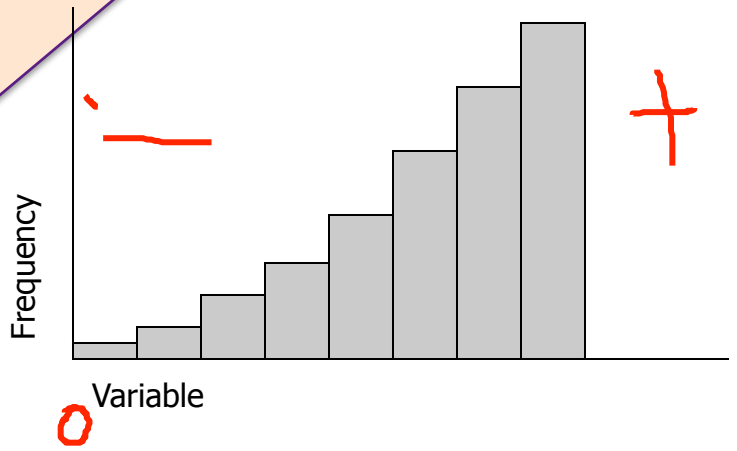
SHAPES OF HISTOGRAMS...

Skewness

A skewed histogram is one with a long tail extending to either the right or the left:



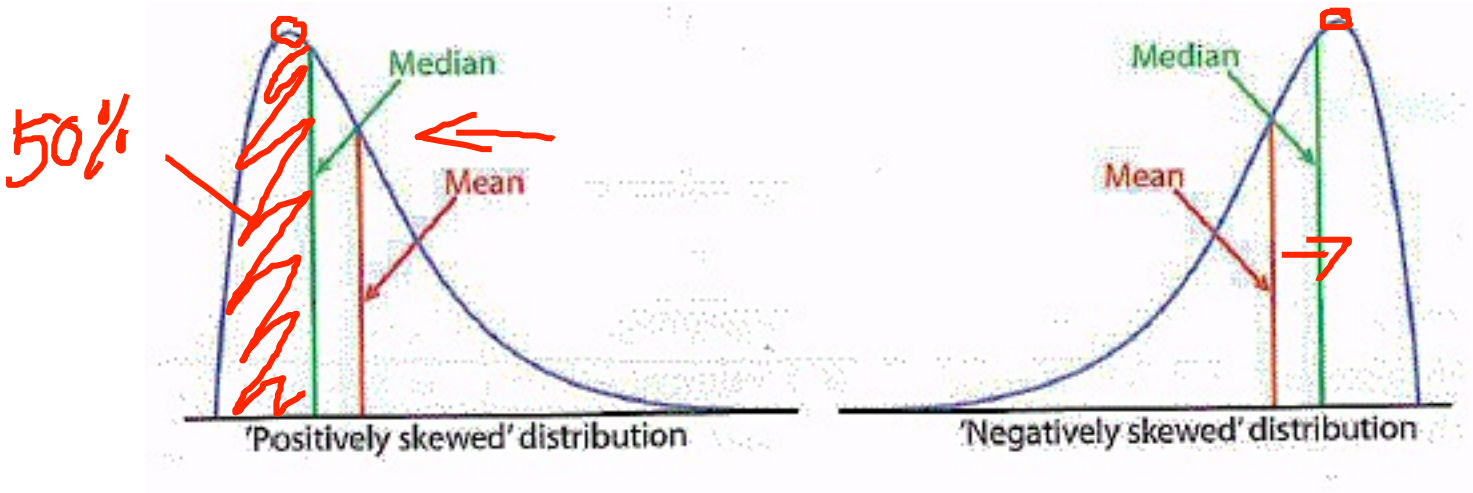
WHERE THE TAIL GOES, CONSIDER IT THE SIDE.



Negatively Skewed

Bowerman, et al. (2017) pp.139

SKEWNESS

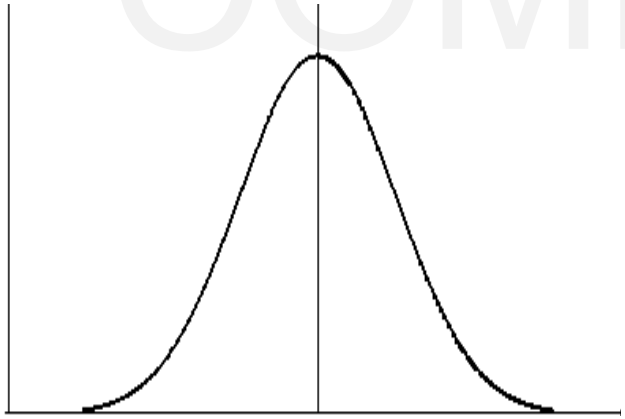


Bowerman, et al. (2017) pp.139

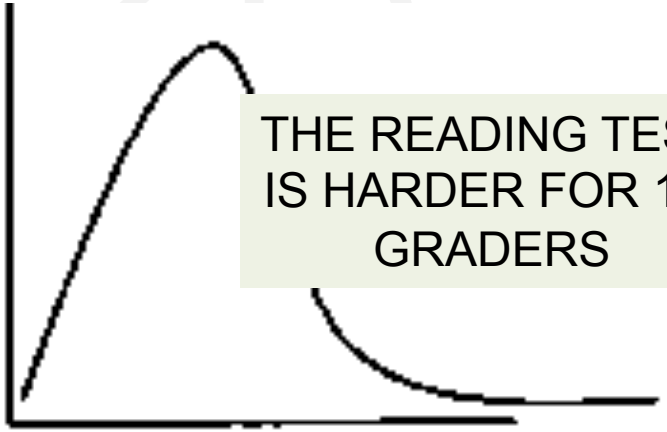
PROBLEM # 2.7

A reading test with 120 possible points yields a bell-shaped distribution with scores ranging from 5 to 50 on a large sample of third graders. If the same test were administered to first graders, what would we expect the form of the frequency distribution to be?

3RD GRADERS



1ST GRADERS



THE READING TEST IS HARDER FOR 1ST GRADERS

Measures of Variation

Percentiles, Quartiles, and Box and Whiskers.

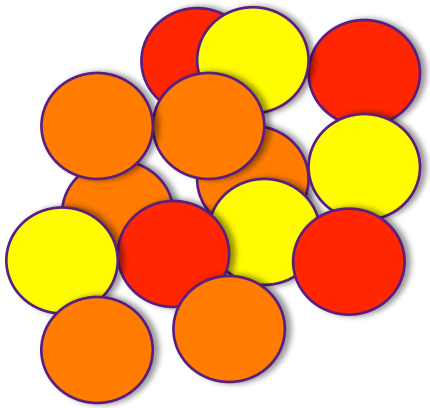
Midterm Preparation

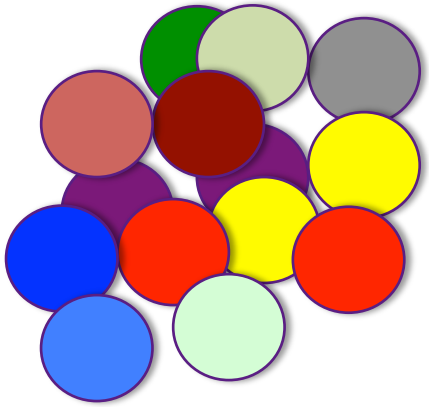
MEASURES OF VARIATION

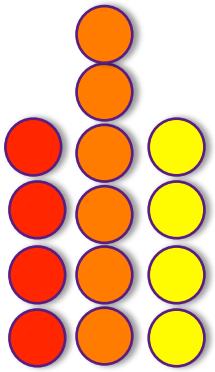
- RANGE
- VARIANCE
- STANDARD DEVIATION
- COEFFICIENT OF VARIATION

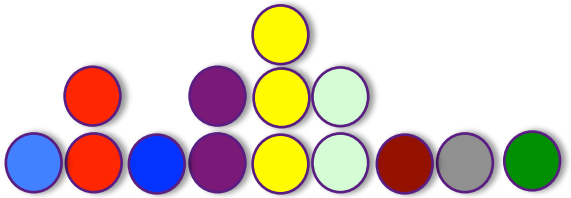
Bowerman, et al. (2017) pp.145

VARIABILITY

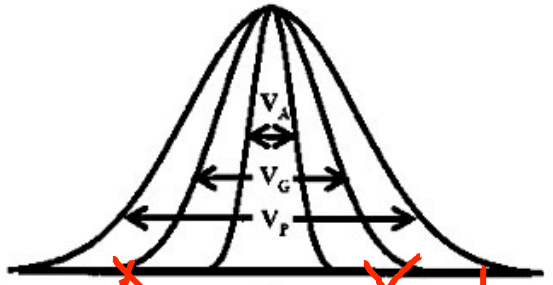








VARIANCE



$(x_1 - \bar{x})^2$ 10 100

$(x_2 - \bar{x})^2$ -30 900

$(x_3 - \bar{x})^2$ 20 400

1400

50 80 90 100

x_2 $\bar{x}$ x_1 x_3

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - (\sum x_i)^2 / n}{n-1}$$

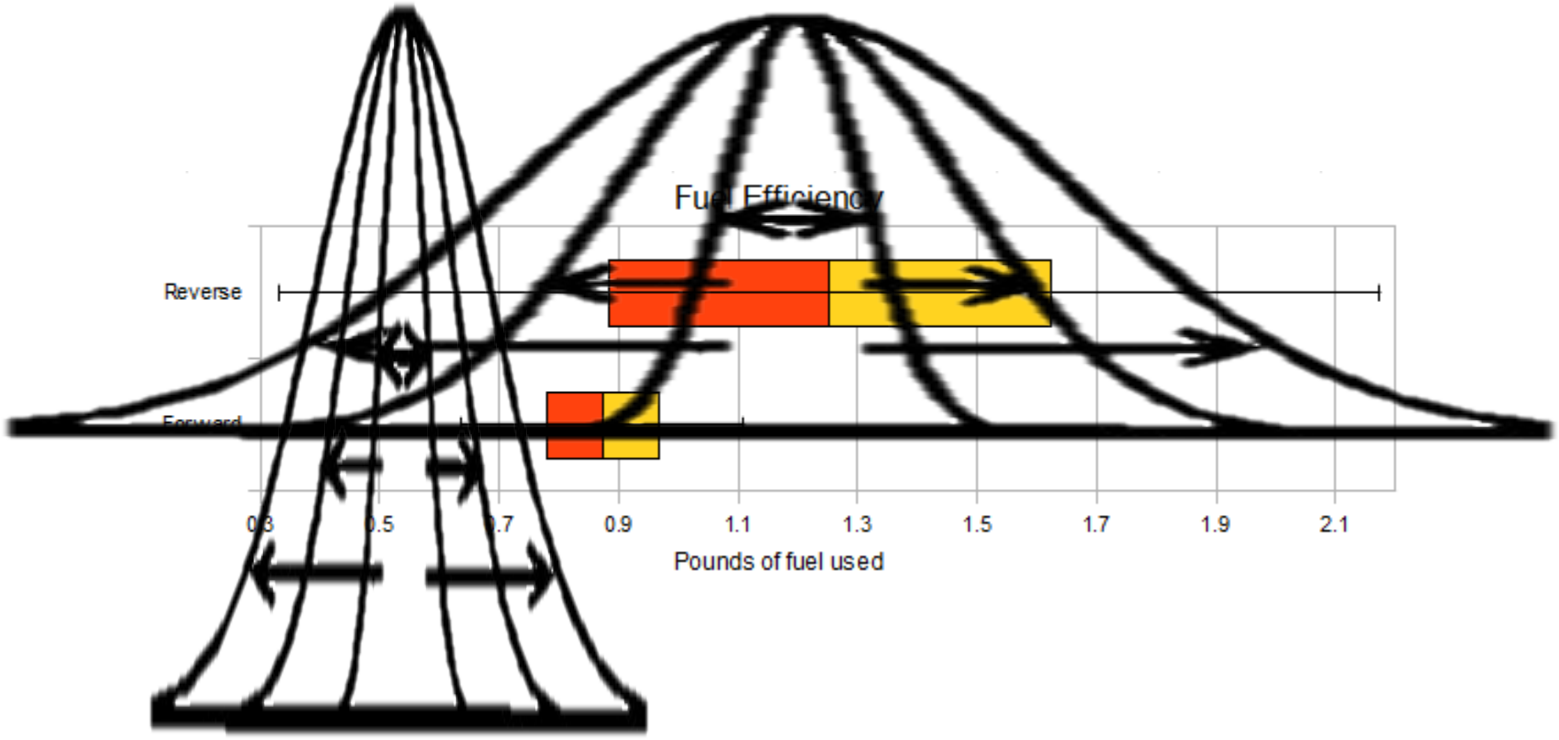
3 − 1

466.67

700

$$\sigma^2 = \frac{}{N}$$

VARIABILITY



PROBLEM # 2.8

Data set: 0, 7, 12, 5 , 33, 14, 8 , 0 , 9 , 22

c. Calculate the standard deviation

STANDARD DEVIATION

is measured in the SAME UNITS as the data,

therefore easier to interpret.

Population

$$\sigma = \sqrt{\sigma^2}$$

Sample

$$s = \sqrt{s^2}$$

How large is the Standard Deviation from the mean.

Coefficient of Variation:

$$\left(\frac{Standard\ Deviation}{Mean} \times 100\right)\%$$

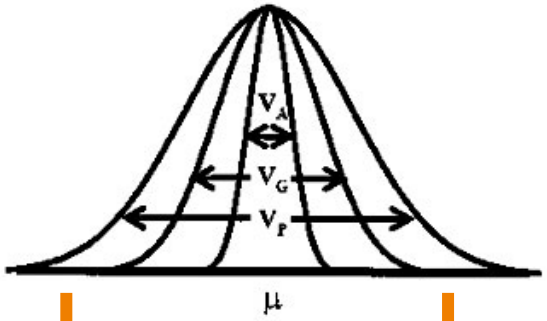
10g

80g

Bowerman, et al. (2017) pp.146

COMM215 First the Foundation, then Innovation

VARIANCE



$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - (\sum x_i)^2 / n}{n-1}$$

PAGE 105
Groebner et al. (2014)

PROBLEM # 2.8

Data set: 0, 7, 12, 5 , 33, 14, 8 , 0 , 9 , 22

c. Calculate the standard deviation

PROBLEM # 2.9

Answer the following questions about variability of data sets:

a. What is the primary disadvantage of using the range to compare the variability of data sets?

b. Can the variance of a data set ever be negative? Explain.

c. Can the variance ever be smaller than the standard deviation? Explain.

PROBLEM # 2.9

BOTH
RANGE of 14

	1	1
	2	2
	2	4
	4	11
	5	12
	6	13
	7	14
	14	14
	15	15
Variance	**25.94**	**31.28**
Std Dev	**5.09**	**5.59**

BOTH
Sample size:
n=9

a. What is the primary disadvantage of using the range to compare the variability of data sets?

The two data sets can

- have the same range
- be vastly different with respect to the data variation.

The range is greatly affected by extreme values.

Greatly affected by EXTREME numbers

PROBLEM # 2.9

b. Can the variance of a data set ever be negative? Explain.

The variance can never be negative due to the squared deviations from the mean.

The square of any number , positive or negative will become positive.

The variance is always positive

PROBLEM # 2.9

c. Can the variance ever be smaller than the standard deviation? Explain.

The variance is usually greater than standard deviation.

HOWEVER, if the data are between 0 and 1, the variance will be smaller than the standard deviation.

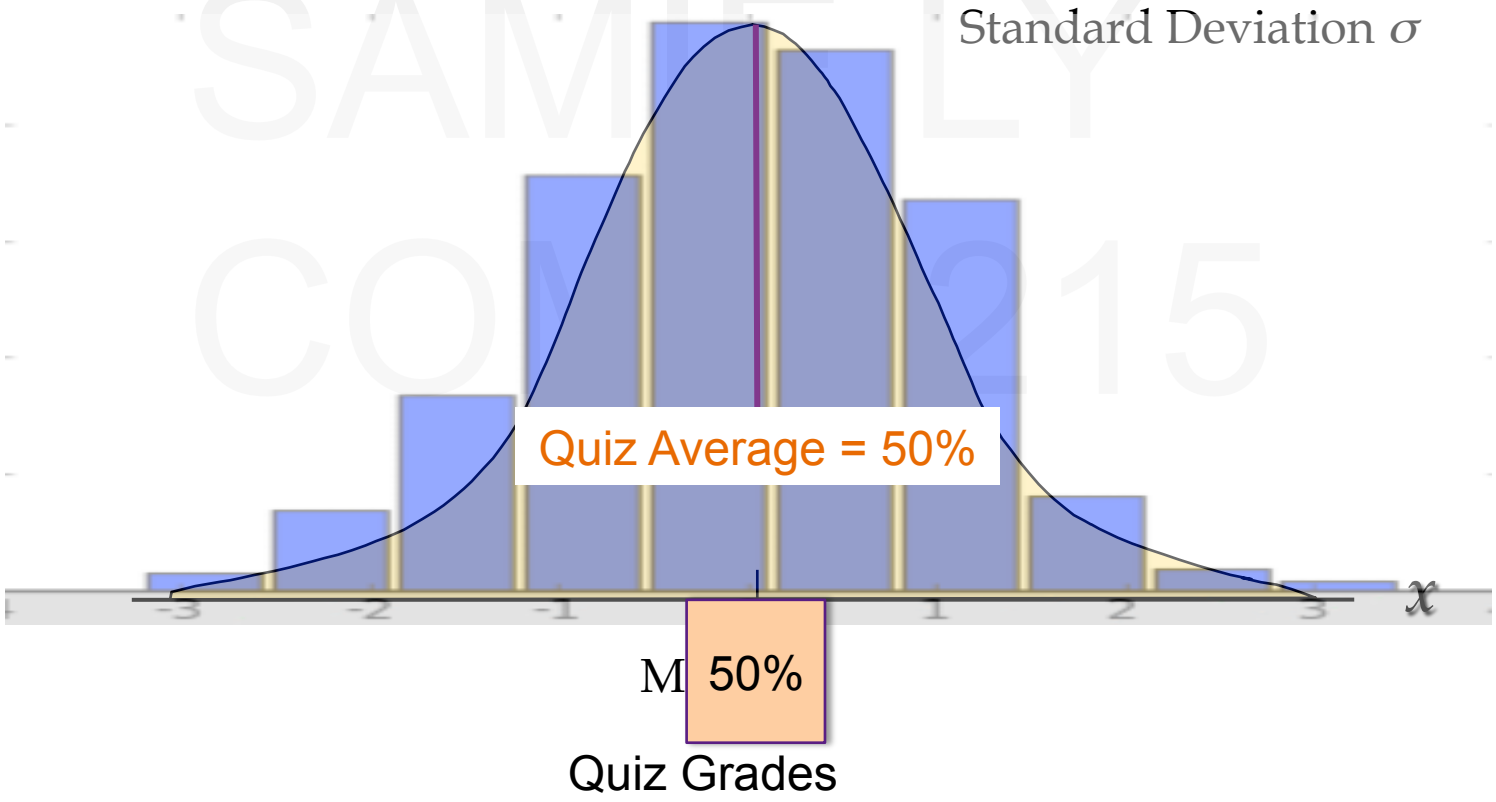
Try this!

$s^2 = 4$, what is the standard deviation?

$s^2 = 0.8$, what is the standard deviation?

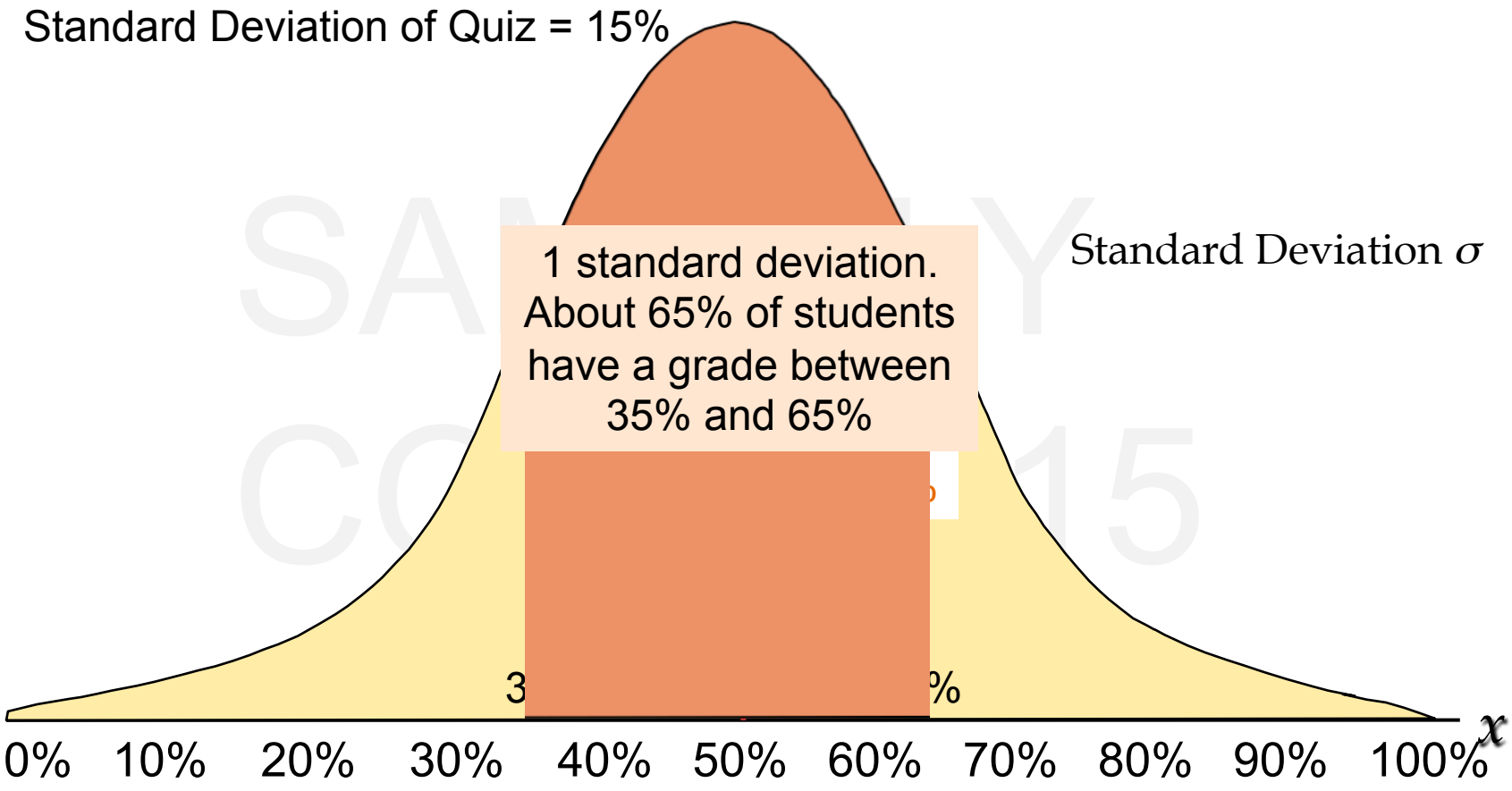
ESTIMATING DATA

- **EMPIRICAL RULE**
- **CHEBYCHEV'S THEOREM**



Bowerman, et al. (2017) pp.148

ESTIMATING DATA



Quiz Grades

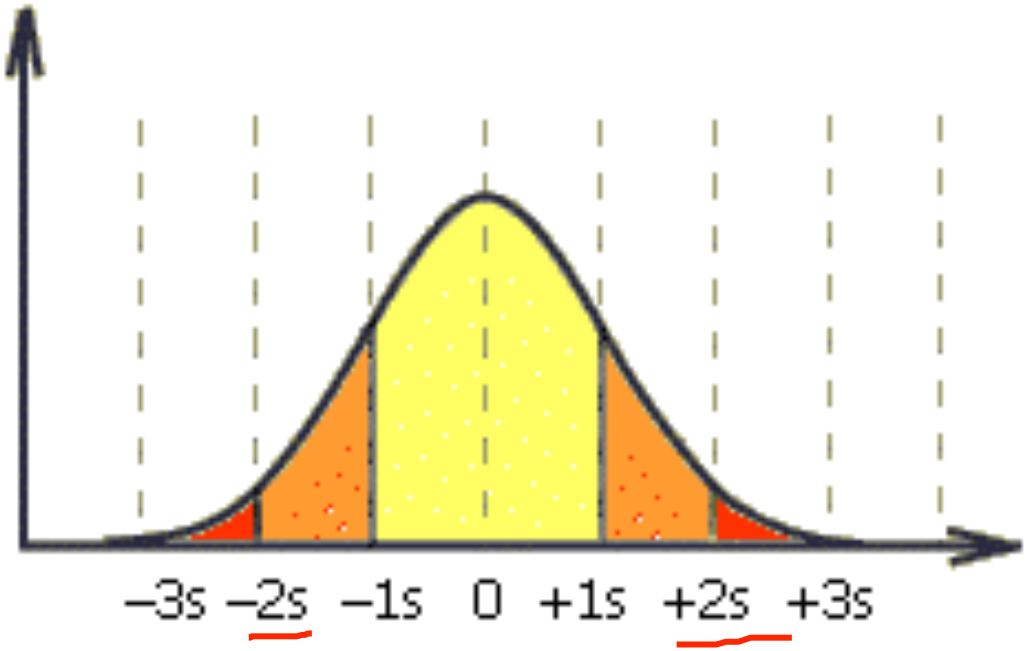
THE EMPIRICAL RULE

Bell-Shaped Data!

Approximately _____% of all observations fall

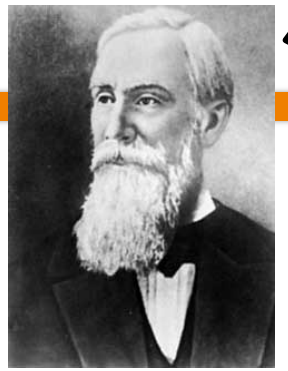
Within _____ standard deviation of the mean.

Percent(%) of observations	Standard deviation



Bowerman, et al. (2017) pp.148

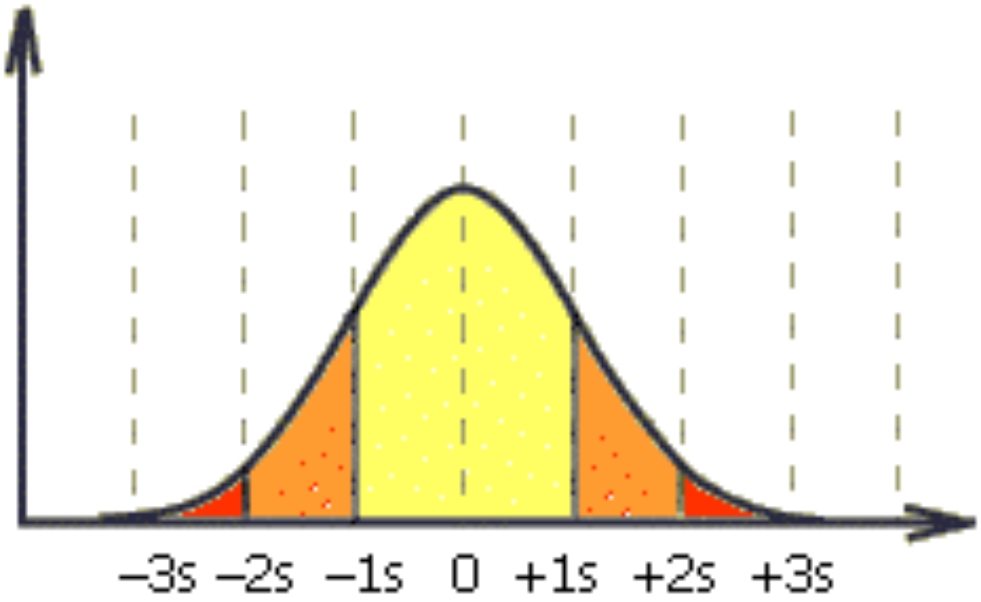
CHEBYCHEV'S RULE



AT LEAST _____% of all observations fall

Within _____ standard deviation of the mean.

Percent(%) of observations	Standard deviation



Bowerman, et al. (2017) pp.151

PROBLEM # 2.11

The annual salaries of the employees of a chain of computer stores produced a positively skewed histogram. The mean and standard deviation are $ 28,000 and $3,000 respectively. What can you say about the salaries at this chain?

NOTE: The histogram is not bell shaped, therefore you cannot use Empirical Rule. You must use Chebychev's theorem instead.

PROBLEM # 2.11

At least 75% of the salaries lie WITHIN 2 STANDARD DEVIATION

28,000 – 2 (3,000) = 22,000$

28,000 + 2 (3,000) = 34,000$

At least 88.9% of the salaries lie WITHIN 3 STANDARD DEVIATION

28,000 – 3(3,000) = 19,000$

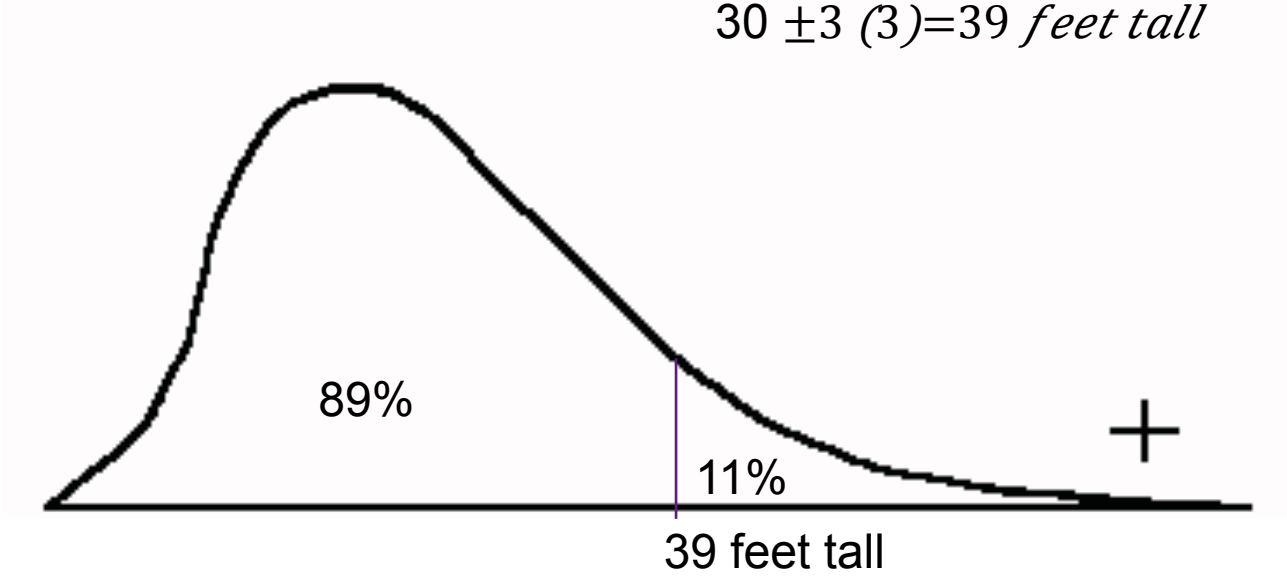
28,000 + 3(3,000) = 37,000$

PROBLEM # 2.12

Land purchase decision. A buyer for a lumber company must decide whether to buy a piece of land containing 5,000 pine trees. If 1,000 of the trees are at least 40 feet tall, the buyer will purchase the land; otherwise, he won't. The owner of the land reports that the height of the trees has a mean of 30 feet and a standard deviation of 3 feet. Based on this information, what is the buyer's decision?

PROBLEM # 2.12

Using Chebychev's rule, at 3 standard deviation



Will not buy because it is not 20% =1000/5000

OUTLIERS USING Z-SCORE

Z-score = 1, at 1 standard deviation from the mean

Z-score = 2, at 2 standard deviation from the mean

Z-score = 3, at 3 standard deviation from the mean

$$z = \frac{x - \overline{x}}{s}$$

Any value outside of 3 standard deviation is considered an outlier.

Bowerman, et al. (2017) pp.152

PAGE 53
ADCOCK et AL (2015)

Percentiles, Quartiles, and Box and Whiskers.

Midterm Preparation

Bowerman, et al. (2017) pp.155

PERCENTILES, QUARTILES, AND BOX AND WHISKERS

- PERCENTILE
- INTERQUARTILE RANGE (IQR)
- BOX PLOTS
- FIVE NUMBER SUMMARY

Bowerman, et al. (2017) pp.155

COMM215
First the Foundation, then Innovation

PERCENTILES

Arrange the dataset by ascending order (smallest-largest)

Compute i

$$i = (p/100)n$$

(a) If it is not an integer, ROUND UP

(b) If it is an integer, the pth percentile is the average of the values in position i and i +1

QUARTILES are 25^{th}, 50^{th}, 75^{th} percentiles.

Bowerman, et al. (2017) pp.155

COMM215 First the Foundation, then Innovation

PROBLEM # 2.8

Data set: 0, 7, 12, 5 , 33, 14, 8 , 0 , 9 , 22

a. Calculate the $50^{th,}$ 25^{th} percentiles of the data above.

b. Find the interquartile range

PROBLEM # 2.8

$$i = (p/100)n$$

Data set: 0, 7, 12, 5 , 33, 14, 8 , 0 , 9 , 22

a. Calculate the 50th, 25th percentiles of the data above.

Order Data set: 0, 0, 5, 7, 8, 9, 12, 14, 22, 33.

i = (50/100) * 10 = 5

= $\frac{(5^{th} \text{ value} + 6^{th} \text{ value})}{2}$

= $\frac{(8+9)}{2}$ = 8.5

The 50th percentile value = 8.5

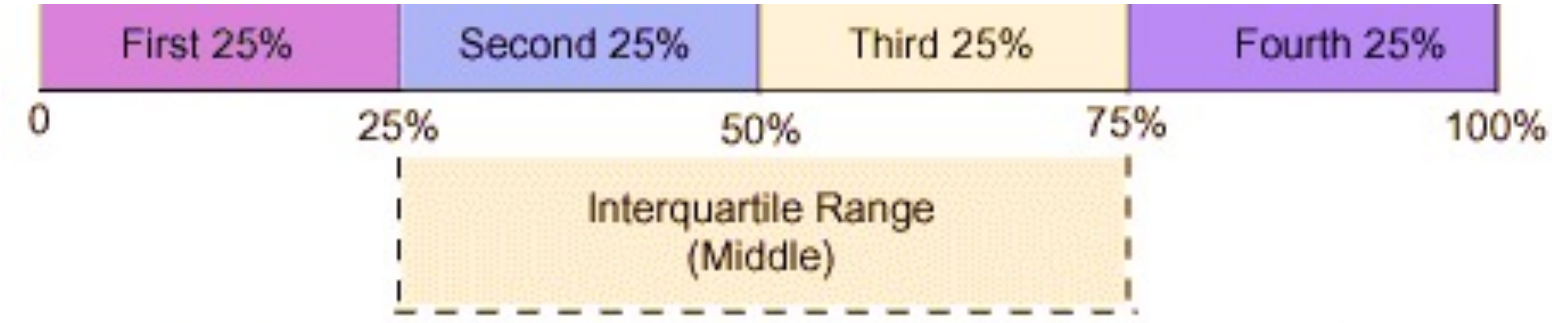
i = (25/100) * 10 = 2.5 Round up!

= 3rd value

The 25th percentile value = 5

INTERQUARTILE RANGE (IQR)

Q3 (75th percentile) – Q1 (25th percentile) = IQR



The largest number – smallest number in your sample is considered the RANGE. Example: 20-2 = 18.

Bowerman, et al. (2017) pp.157

PROBLEM # 2.8

$$i = (p/100)n$$

Data set: 0, 7, 12, 5 , 33, 14, 8 , 0 , 9 , 22

b. Find the interquartile range

Order Data set: 0, 0, 5, 7, 8, 9, 12, 14, 22, 33.

IQR = 9

i = (75/100) * 10 = 7.5 Round up!

= 8th value

The 75th percentile value = 14

BOX PLOT & WHISKERS

Smallest observation

Largest observation

First Quartile (Q1)

Second Quartile (Q2)

Third Quartile (Q3)

Interquartile Range (IQR)

Upper Fences, Lower Fences (Inner and Outer)

Whiskers

Outliers

PAGE 59

ADCOCK et Al (2011)

Bowerman, et al. (2017) pp.157

BOX PLOT

Upper Inner Fence, Lower Inner Fence

- 1.5(IQR) below Q1
- 1.5(IQR) above Q3

Upper Outer Fence, Lower Outer Fence

- 3(IQR) below Q1
- 3(IQR) above Q3

Whiskers

- Smallest and Largest values WITHIN THE LIMITS

Outliers

For Box Plot: Outside the Limits

For Empirical/Chebychev's Rule: Outside the 3rd Std Dev

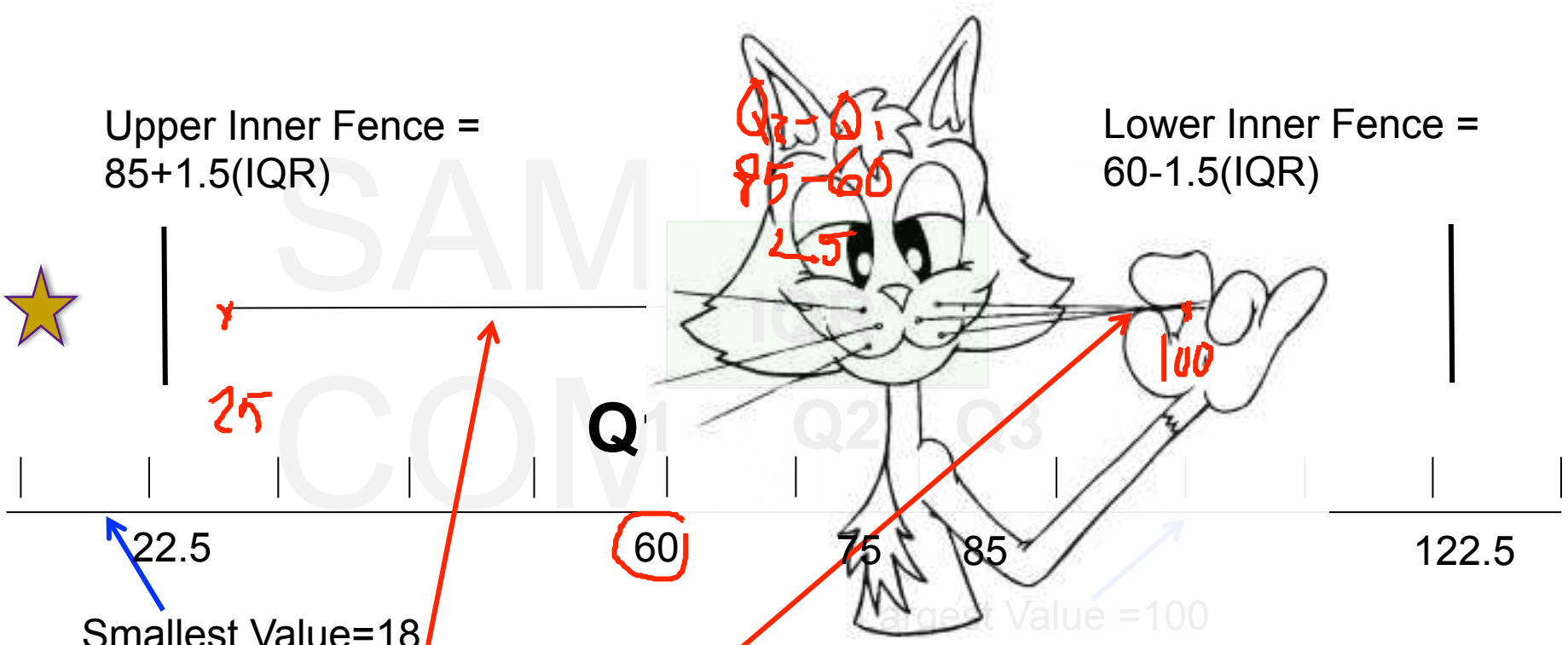
Bowerman, et al. (2017) pp.157

PROBLEM # 2.13

Suppose a data set consisting of exam scores has a lower quartile $Q_L = 60$, a median $m = 75$, and an upper quartile $Q_U = 85$.

The scores on the exam range from 18 to 100. Without having the actual scores available to you, construct as much of a box plot as possible.

Partial Data: 18, 25….100


Upper Inner Fence =
85+1.5(IQR)
Lower Inner Fence =
60-1.5(IQR)
Q3-Q1
85-60
25
25
100
Q
22.5
60
75
85
122.5
Smallest Value=18
Whiskers
OUTLIERS

FIVE NUMBER SUMMARY

Smallest observation

Largest observation

First Quartile (Q1)

Second Quartile (Q2)

Third Quartile (Q3)

Interquartile Range (IQR)

Upper Fences, Lower Fences (Inner and Outer)

Whiskers

Outliers

Bowerman, et al. (2017) pp.156

Midterm Preparation

THE CONTEXT

One of the major measures of the quality of service provided by any organization is _____.

The speed

with which it responds to customer complaints.

The data below give

the number of waiting days

Between the receipt and resolution of a complaint

HOX PLOT PROBLEM

A cellphone software company was receiving complaints about the installation of cellphone software.

The experimental unit to the data is: Waiting Days

72	20	29	23	30	27	25	32	4	29
29	31	27	60	13	51	36	5	26	1
12	22	14	28	35	26	9	34	3	1

HOX PLOT PROBLEM

72	20	29	23	30	27	25	32	4	29
29	31	27	60	13	51	36	5	26	1
12	22	14	28	35	26	9	34	3	1

Construct a box plot for the data set and comment on the nature of the plot including identifying outliers, if any.

HOX PLOT PROBLEM

72	20	29	23	30	27	25	32	4	29
29	31	27	60	13	51	36	5	26	1
12	22	14	28	35	26	9	34	3	1

Compute the proportion of observations that are within one, two and three standard deviations of the mean. Does the empirical rule apply for this data set? Identify outliers, if any.

HOX PLOT PROBLEM

72	20	29	23	30	27	25	32	4	29
29	31	27	60	13	51	36	5	26	1
12	22	14	28	35	26	9	34	3	1

To improve services, management wants to grant a special offer to customers whose complaints take unnecessarily longer periods to resolve. However, it is desirable that no more than 25% of the customers who lodge a complaint would receive the special offer. Approximately, what number of waiting days between the receipt and resolution of a complaint may be used for the cutoff?

PREVIEW ON LESSON 3

REFERENCES

Keller, G. (2012). *Statistics for management and economics.* Mason: Cengage Learning.

McClave, J. T., Benson, G. P., & Sincich, T. (2008). *Statistics for Business and Economics.* New Jersey: Prentice Hall.

Weiers, R. M. (2011). *Introduction to Business Statistics.* Mason: Cengage Learning.

(GMAC), F. t. (Ed.). (2005). *GMAT -Quantitative Review.* Oxford, UK: Blackwell.