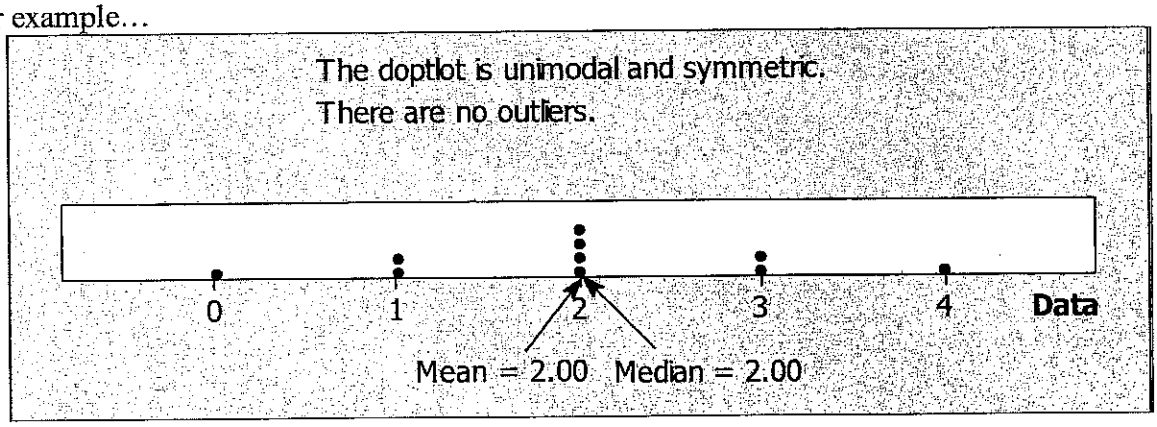
Introduction to Statistics and Data Analysis
Measures of Typical Value (Center)

Part I. For each data set given below, do the following:

- Construct a dotplot of the data.
- Describe the shape of the dotplot.
- Compute the mean and mark its place on the number line of your dotplot.
- Compute the median and mark its place on the number line of your dotplot.

For example…



Data Set #1: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Data Set #2: 1, 1, 1, 2, 3, 8, 9, 10, 10, 10

Data Set #3: 0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 3

Data Set #4: 0, 17, 18, 18, 19, 19, 19, 19, 20, 20

Part II. Answer each of the following questions…

A. When are the mean and the median equal?

B. When is the mean greater than the median?

C. When is the mean less than the median?

D. When is a "typical" value of the data set best represented by the mean? By the median?

Solution To Measures Of Typical Value (Center)

Part I.

Data Set #1: Dotplot has no mode, but is symmetric. There are no outliers.
Mean = 5.5, Median = 5.5

Data Set #2: Dotplot is bimodal and symmetric. There are no outliers.
Mean = 5.5, Median = 5.5

Data Set #3: Dotplot is unimodal and skewed to the right. There are no outliers.
Mean = 0.8, Median = 0.5

Data Set #4: Dotplot is unimodal and skewed to the left. There is one outlier (0).
Mean = 16.9, Median = 19.0

Part II.

A. When the data is symmetric, mean = median
When the data is approximately symmetric, mean $\approx$ median

B. When data is skewed to the right, mean > median
If the data is skewed right with outliers, mean >> median

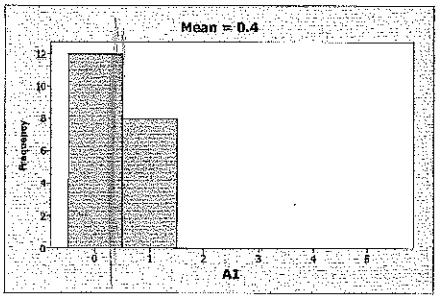
C. When data is skewed to the left, mean < median
If the data is skewed left with outliers, mean << median

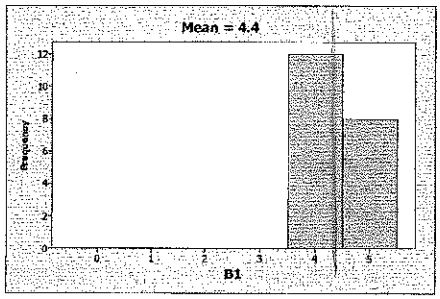
D. Mean is preferred to represent typical value if the data is approx symmetric
Mean is computed using ALL values in the data, so it is more reliable

Median is preferred if the data is skewed (especially if outliers are present)
Median is not affected by outliers

Introduction to Statistics and Data Analysis
Measuring Spread (Variability)

Part I. For each of the four pairs of histograms below, choose the statement from the box at the right that best describes the situation. HINT: Mark the value of the mean on the x-axis and consider where the data lies relative to the mean.

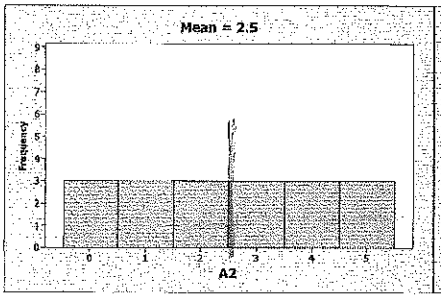


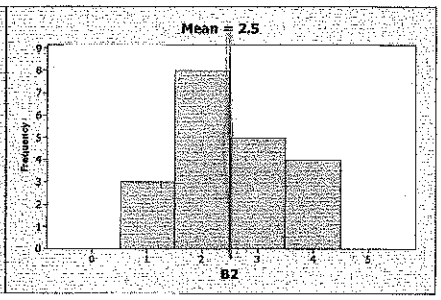


Pair #1
A has larger std dev

B has larger std dev

Both graphs have the same std dev

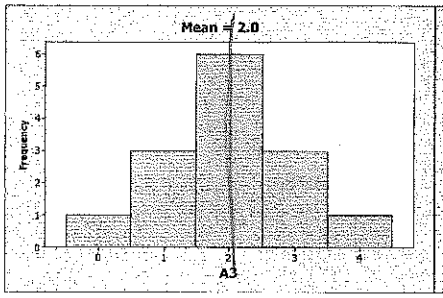


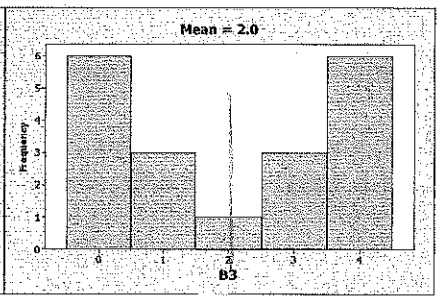


Pair #2
A has larger std dev

B has larger std dev

Both graphs have the same std dev

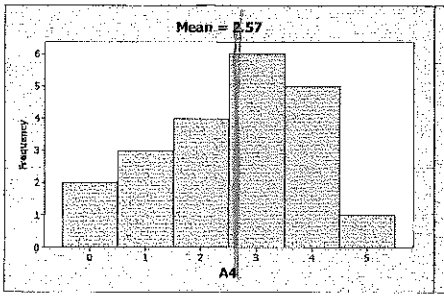


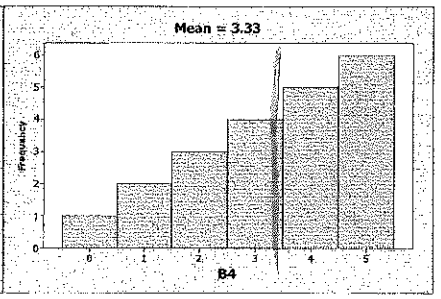


Pair #3
A has larger std dev

B has larger std dev

Both graphs have the same std dev





Pair #4
A has larger std dev

B has larger std dev

Both graphs have the same std dev

Part II. The twenty-two properties on a Monopoly board, and their respective rents, are listed in the table below.

PROPERTY	RENT
Mediterranean Ave.	2
Baltic Ave.	4
Oriental Ave.	6
Vermont Ave.	6
Connecticut Ave.	8
States Ave.	10
St. Charles Place	10
Virginia Ave.	12
St. James Place	14
Tennessee Ave.	14
New York Ave.	16

Kentucky Ave.	18
Indiana Ave.	18
Illinois Ave.	20
Atlantic Ave.	22
Ventnor Ave.	22
Marvin Gardens	24
Pacific Ave.	26
No. Carolina Ave.	26
Pennsylvania Ave.	28
Park Place	35
Boardwalk	40

Enter the rents into a list in your calculator.

1. Construct a histogram of the rents. Describe the shape of the distribution. Plot

2. Compute the mean and standard deviation for the rents. Since this data set is the entire population of Monopoly rents, be sure to obtain the popn std dev.
Calc → 1-Var Stats

Mean = _____ Std Dev = _____

3. Why is it appropriate to apply the empirical rule to this set of data?

4. We expect ≈95% of the rents to be between _____ and _____.

Approximately _____% of the rents are expected to be between 7.5 and 27.1.

5. Answer the following:

A. What is the z-score for Baltic Ave.'s rent? Z-Score = _____

B. Is this value considered "unusual"? Explain.

C. Is Boardwalk's rent "unusual"? Explain.

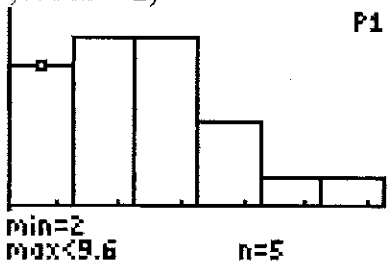
Solution To Measuring Spread (Variability)

Part I.

Pair #1 – both graphs have the SAME std dev
Pair #2 – A has the larger std dev
Pair #3 – B has the larger std dev
Pair #4 – similar std dev's, but B is slightly larger

Part II.

1. Histogram (bucket width = 7.6, Xmin = 2)



Unimodal (mound-shaped), Skewed to the Right

2. Mean = $\bar{x}$ = 17.3, Popn Std Dev = σ_X = 9.8

3. The data is approximately "**mound-shaped**", so empirical rule is appropriate.

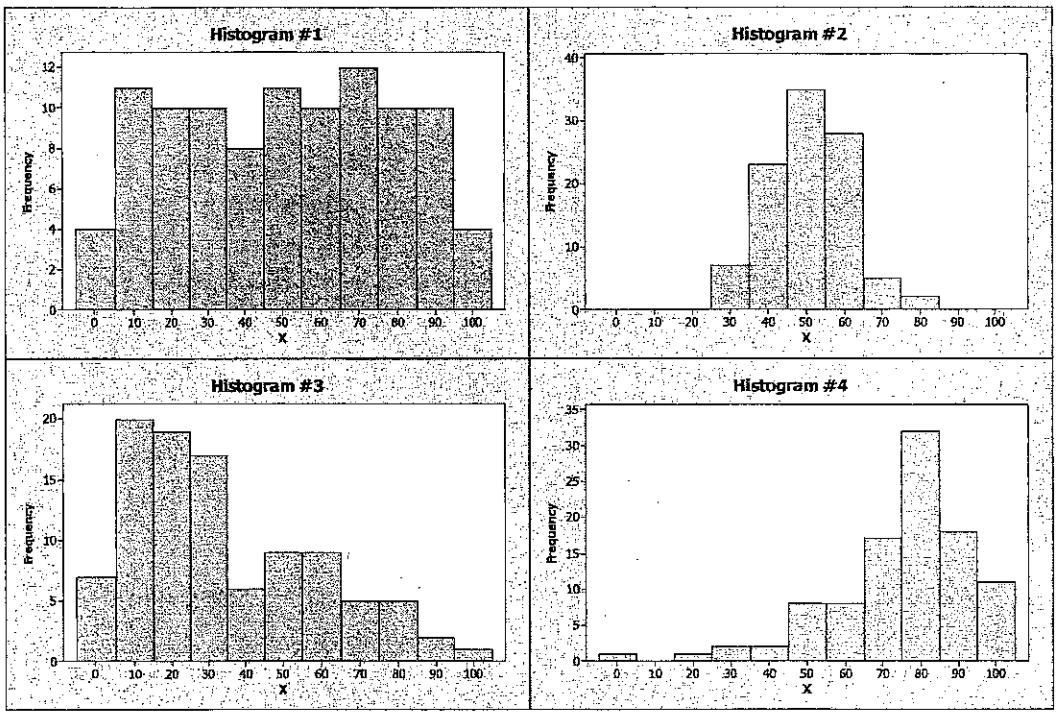
4. Expect ≈95% of rents between **-2.3** and **36.9**
   Approximately **68%** of rents expected between 7.5 and 27.1

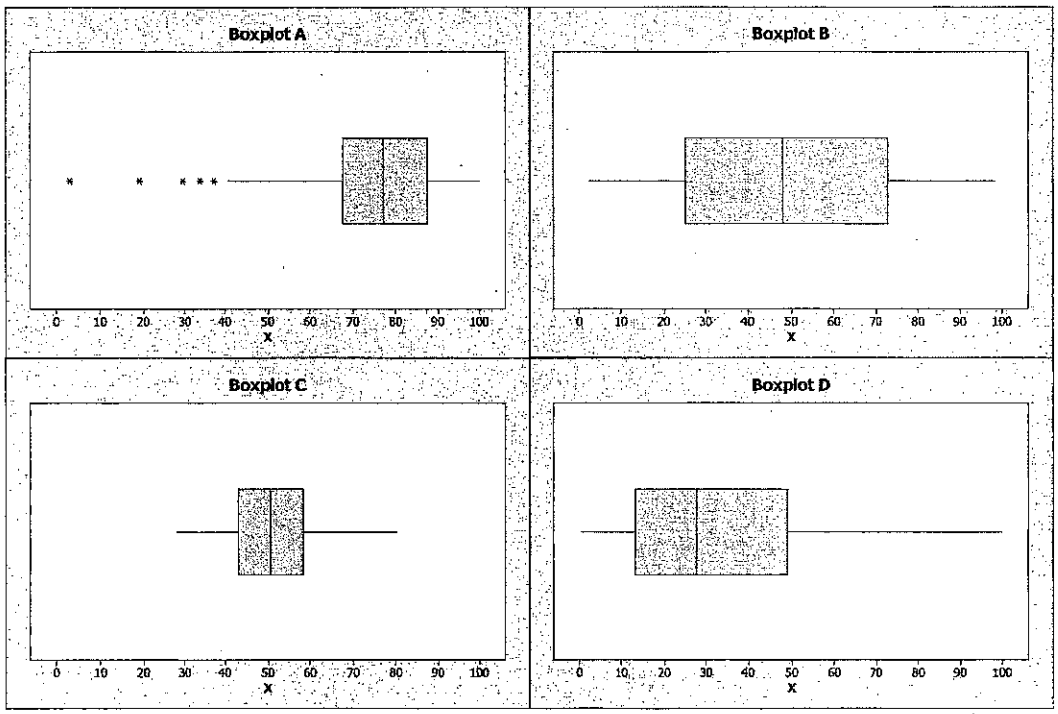
5. Baltic Ave z-score = -1.36
   -1.36 is not considered unusual (z-score within -2 and +2)

6. Boardwalk z-score = 2.32
   Boardwalk has an unusually large rent (z-score > 2)

Introduction to Statistics and Data Analysis
Quartles, IQR, and Boxplots

Part I. Match the histogram to the corresponding boxplot.





Part II. The table below contains the ages at which Oscar winning actresses won their awards ("Ages of Oscar-Winning Best Actors and Actresses", Richard Brown and Gretchen Davis, *Mathematics Teacher* magazine).

50	44	35	80	26	28	41	21	61	38	49	33	74
30	33	41	31	35	41	42	37	26	34	34	35	26
61	60	34	24	30	37	31	27	39	34	26	25	33

**Enter the data into a list in your calculator. **

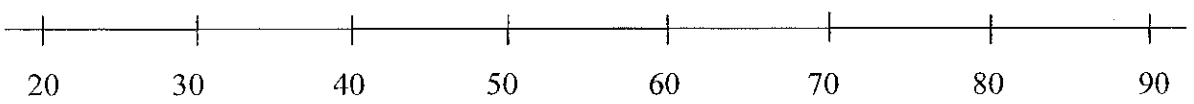
1. Obtain the five-number summary. (Choose Calc, then 1-Var Statistics)

2. Compute the IQR.

3. Compute the missing values below:

For the actresses, values below _____ OR above _____ are considered to be outliers.

4. a. Draw a *modified* boxplot below.
b. Determine whether the distribution is symmetric or skewed.
c. Are there outliers?
d. Determine an appropriate typical value and measure of variability.



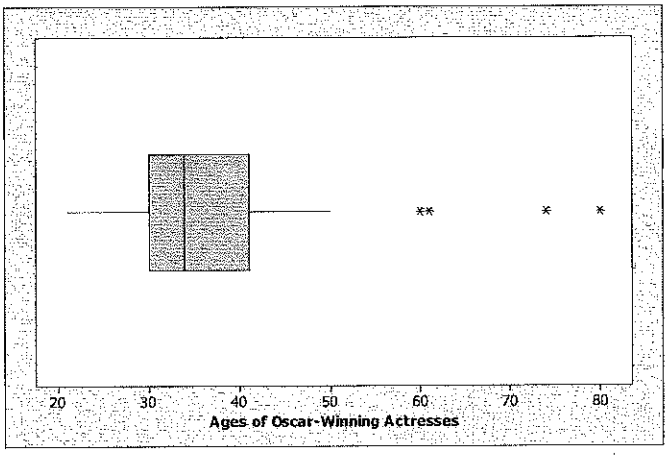
Solution To Quartiles, IQR and Boxplot

Part I

1 = B (fairly symmetric, large spread)
2 = C (fairly symmetric, small spread)
3 = D (very skewed to the right)
4 = A (slightly skewed to the left)

Part II

1. Min = 21, Q1 = 30, Med = 34, Q3 = 41, Max = 80
2. IQR = 41 – 30 = 11
3. Values less than **13.5** OR greater than **57.5** are outliers
   $< 30 - 1.5(11) \rightarrow < 13.5$
   $> 41 + 1.5(11) \rightarrow > 57.5$
4.



Dist from box to largest obs > Dist from box to smallest obs
So data is skewed to the right
The values 60, 61, 74 and 80 are outliers
Typical Value = Median = 34 and Meas of Spread = IQR = 11