Center: a typical value that summarizes the data set in a single number

mean = average = $\frac{\sum x}{n}$ (best measure of center for symmetric distributions)
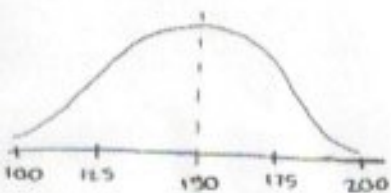
median = 50% value, the middle value in the data set; half of the values are
    greater than the median, and half are less than the median
      (best measure of center for slowed distributions)

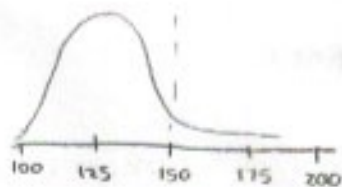mode = single most common data value, most frequent value

Stock market = moving averages

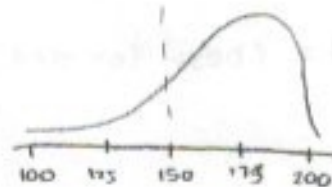Shape: visual description of data

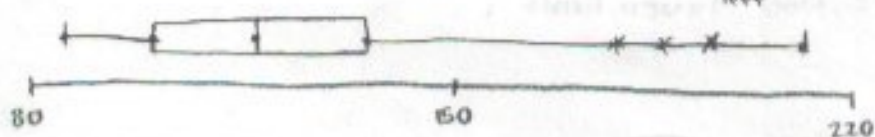symmetric   skew right   skew left

Tools: Histogram, Box Plots, Dot Plots $(u/n)\wedge 0.5$

Stock market: candlestic Charts   skew stansnc

                                          $\pm 2 \cdot \sqrt{\frac{6}{n}}$ n = count
                                                                    190

Bin width: (10-12 bars) $\frac{range}{bin}$ highly   nearly   Adjust whiskers, lowest
                                                         skew   symmetric highly to highest outlier
                                                         left   skew
                                                                                  right
                                                    -0.43 :71   0   :21 +0.43   ιι
Box Plot (Box and whisker)   * Modified Box Plot
                                                    slight   slight skew   does not consider
                                                    skew   right   outlier
                                                    left   min
                                                                  max   (individually graphed)
   80   150   220   Q1
                                                    Q3
                                                    median
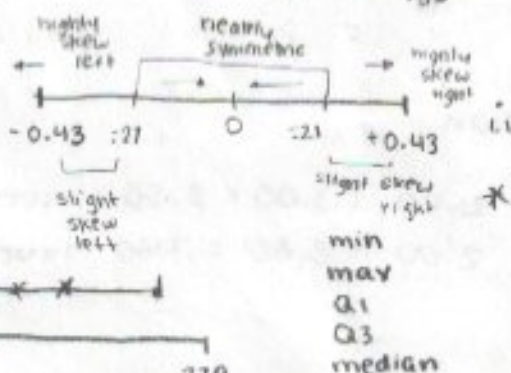
<u>Spread</u>: a measure that describes how dispersed the data set is

Standard deviation = (best measure of spread for symmetrical distributions)

$\left[\begin{array}{l}\text{range} = \text{maximum} - \text{minimum} \\ \text{interquartile range} = Q_3 - Q_1, \text{ where } \begin{array}{l} Q_1 = \text{1st Quartile value } (25\%) \\ Q_3 = \text{3rd Quartile value } (75\%)\end{array}\end{array}\right]$ IQR

(best for skewed distribution)

Stock market = Beta

<u>Outliers</u>: unusual values within the data set

Standard deviation method = (best for symmetric distribution)

IQR method = (best for skewed distributions)

<u>Standard deviation method</u>:

mean	standard deviation
2.1 $\pm$	$(2 - 1.11)$

$S_x = 1.11$

$\bar{X} = 2.10$

$2.1 + 2.22 \rightarrow 4.32$	upper limit
$2.1 - 2.22 = -0.12$	lower limit

4 outliers = 5,5,5,5

<u>IQR method</u>:

median :

IQR : 2.00
$\times 1.5$
$\boxed{3.00}$

$2.00 + 3.00 = 5.00$	upper limit
$2.00 - 3.00 = 1.00$	lower limit

Univariate Data Analysis

Victor Katona, P3
October 5, 2017

<u>Center:</u> a typical value that summarizes the data set in a single number

Mean = Average = $\frac{\Sigma x}{n}$ → sum (Best measure of center for symmetric distribution)

Median = 50% value, the middle value in the data set; half of the values are greater than the median, and half are less than the median. (Best measure of center for skewed distribution.)

Mode = Single most common data value "Most Frequent Value"
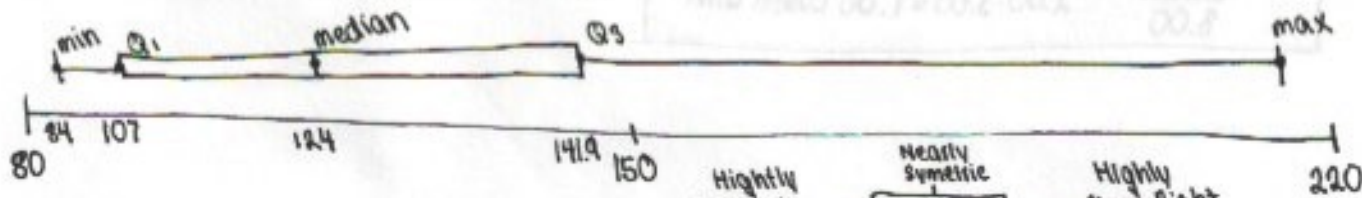
↳ Stock Market: Moving Averages

<u>Shape:</u>

100 140   180   200

Skew left

Skew Right
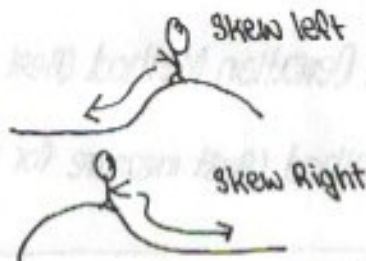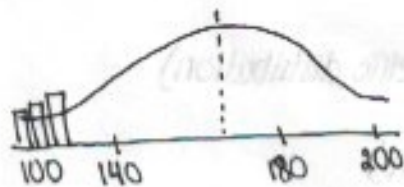
Symetric? or Skew?
↓
Left/Right

Tools: Histogram, Box Plots, Dot Plots,
Stock Market: Candlestick Charts

↑ sample size
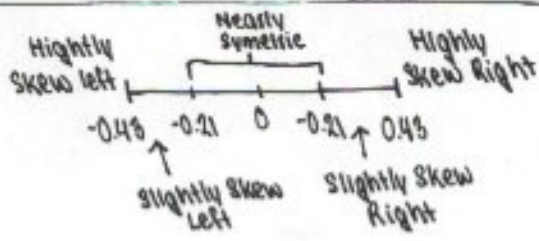$\pm 2 \cdot \sqrt{\frac{6}{n}}$ n=count
                                    180
$((6/n)^{0.5}) \cdot 2$

Bin Width: The number of bins.

min Q₁   median   Q₃   max
84   107   124   141.9 150
80   220

Plot: Min, Q₁, Median, Q₃, Max

Highly Skew left   Nearly Symetric   Highly Skew Right
-0.43   -0.21   0   -0.21   0.43
         ↑slightly skew Left   slightly Skew Right↑

Spread: a measure that describes how "dispursed" the data set is.

       Standard Deviation: (Best measure of spread for symetric distribution)

Best measure   Range = Max - Min

for skewed   Intercartile Range (IQR): $Q_3 - Q_1$ where $Q_3 =$ Quantitive Value (75%)

distribution $Q_1 =$ Quantitive Value (25%)
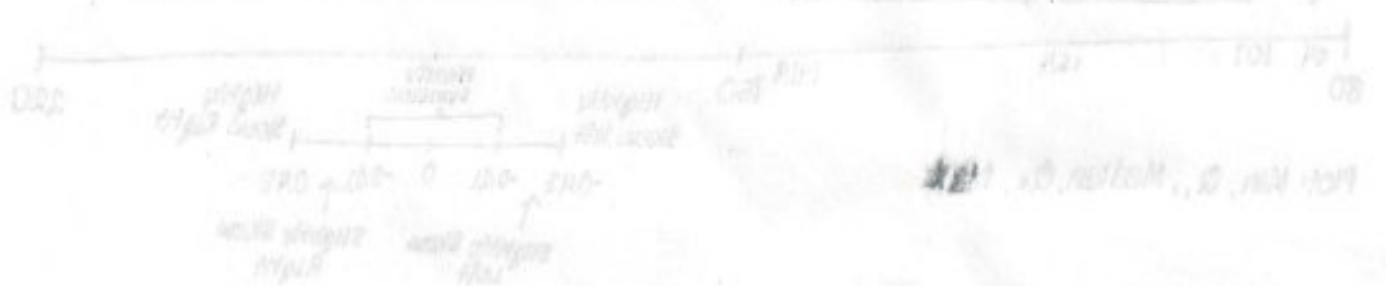
        Stock Market: Beta

Outliers: unusual values within the data set.

       Standard Deviation Method (Best measure for symetric distribution)

       IQR Method (Best measure for skewed distribution.)

$S_x = 1.11$ Mean $2 \cdot S_x$

           $2.1 \pm 2(1.11) = 2.1 + 2.22 = 4.32$

                      $= 2.1 - 2.22 = \overline{0.12}$

$IQR = 2.00$ Median

  $\times 1.5$ $2.00 + 3.00 = 5.00$ Uper Limit

  $\overline{3.00}$ $2.00 - 3.00 = \overline{-}1.00$ Lower Limit

Hayley Lunn
Per. 1 (Winner!)

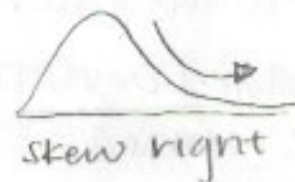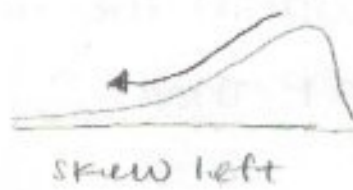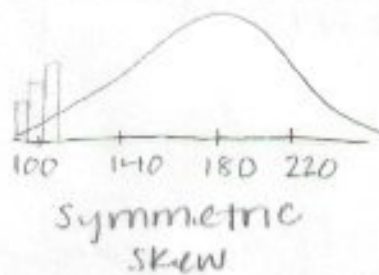center: A "typical" value that summarizes the data set in a single number.

Mean = Average = $\frac{\Sigma x}{n}$ ← symbol for sum (Best measure of center for symmetric distributions)

Median = 50% value, the middle value in the data set; half of the values are greater than the median, and half are less than the median. (Best measure of center for skewed distributions)

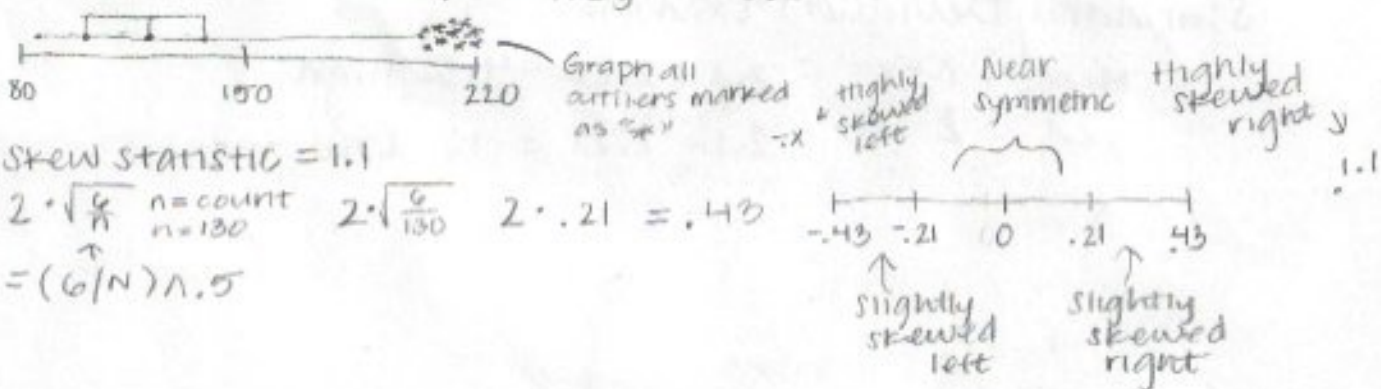Mode = single most "common" data value

Moving averages

Shape

100   140   180   220
Symmetric
skew

Skew left

Skew right

Tools: Histogram, Box plots, Dot plots
Stock Market: Candlestick charts
Bin Width 10 Q_0 Min Q, Q_2 Median Q_3, Q_4 Max   Box Plot

80   150   220   Graph all outliers marked as "*"

Skew statistic = 1.1

$\pm 2 \cdot \sqrt{\frac{6}{n}}$ n = count n = 130 $2 \cdot \sqrt{\frac{6}{130}}$ $2 \cdot .21 = .43$

$= (6/N)^{.5}$

Highly skewed left   Near symmetric   Highly skewed right

-.43 -.21   0   .21   .43

↑ slightly skewed left   ↑ slightly skewed right
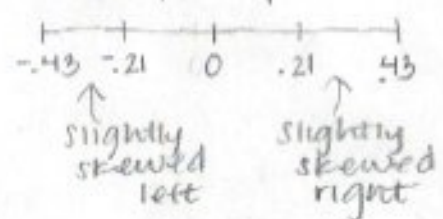
spread: A measure that describes how "dispursed" the data set is.
(how spread out)

Best measures for skewed distribution Standard Deviation

$$\left\{\begin{array}{l} \text{Range} = \text{Max} - \text{Min} \\ \text{Interquartile Range} = Q_3 - Q_1 \end{array}\right.$$

where $Q_3 = $ 3rd quartile value (75%)
$Q_1 = $ 1st quartile value (25%)

Stock Market: Beta

Outliers: unusual values within the data set

Standard Deviation Method

IQR Method

- $Q_1 - Q_3 \cdot 1.5$
- upper limit = Median + answer $(Q_1 - Q_3 \cdot 1.5)$
- lower limit = Median - Answer

Standard Deviation Example

Mean + 2.5x $= 2.1 + 2.22 = 4.32$ High
2.1   2·1.11 $2.1 - 2.22 = .12$ Low
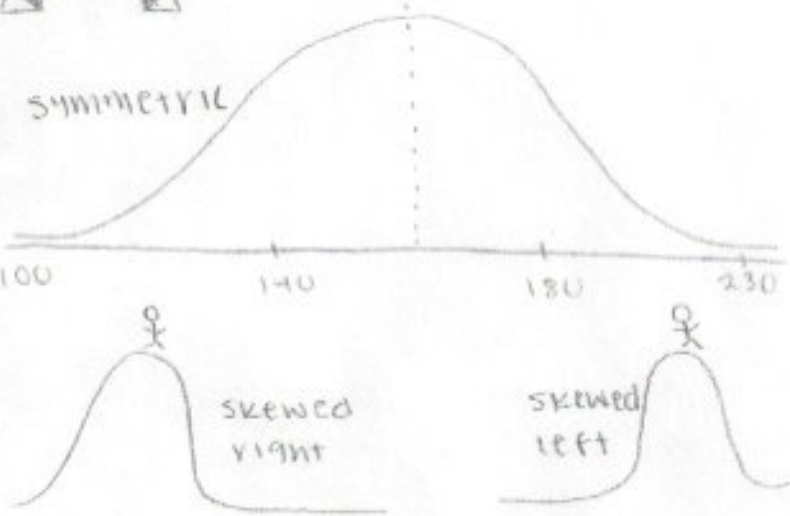
UNIVARIATE DATA ANALYSIS

Center

- A typical value that summarizes the data set in a single number
- Mean: Average $= \frac{\sum x}{n}$ (best for symmetric)
- Median: 50% value, the middle value in the data set; half the values are greater than the median, half are less than the median. (Best measure of center for skewed distribution)
- Mode: single most "common" data value
  - ↳ stock market: moving averages

$$\pm 2 \sqrt{\frac{6}{n}} \quad n = \text{count} = .43$$

$$\downarrow \quad 130$$

$$(6/n)^{.5}$$

Spread:

- a measure that describes how "dispersed" the data set is

Standard Deviation (best measure of spread for symmetric dist.)

Range: Max - Min

Interquartile Range: $Q_3 - Q_1$

Q_3: 3rd quartile
Q_1: 1st quartile

stock market: Beta

Shape

symmetric

100   140   180   230

skewed right   skewed left

Tools: Histogram, Box Plots, Dot Plots, stock market candle-stick charts

Bin width: 10-12 bars

84   107   124

80

Modified Box Plot ↓
141.9

Outliers

- unusual values within the data set

Standard Deviation Method

Mean $2.1 + 2.22$

$2.1 \pm 2.22 = 4.32$ upper limit

$2.1 - 2.22 = -.12$ lower limit

IQR Method

IQR   34.93
   ×   1.5
   52.40

Median   upper
$124 + 52.40 = 176.4$
$124 - 52.40 = 71.6$
            lower

176.2   188   197   207   218

150   220