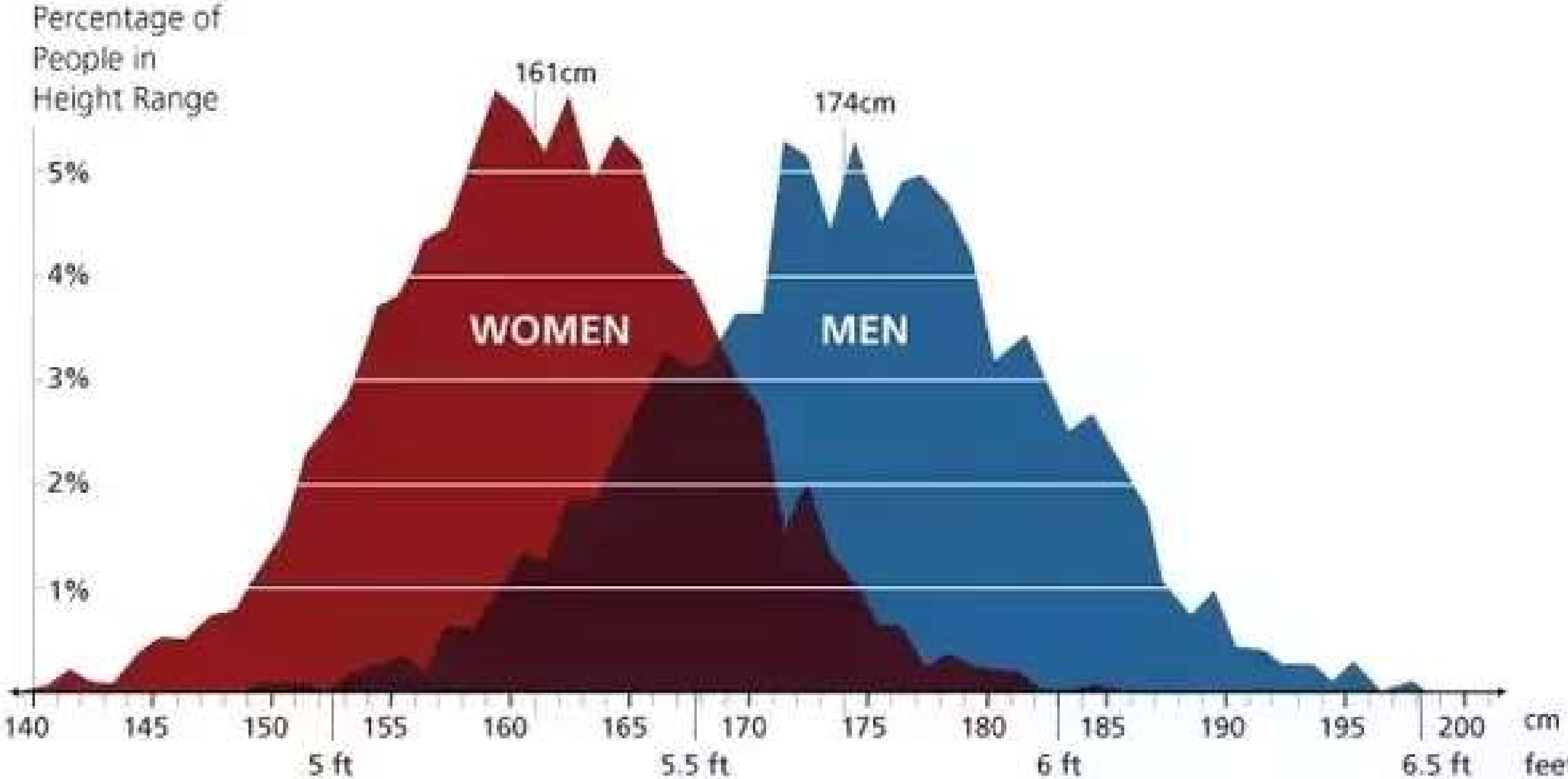
Quantitative Skills & Reasoning – Math 1001

Dr. Bob Brown, Jr.
Dean Emeritus
Professor Emeritus
East Georgia State College

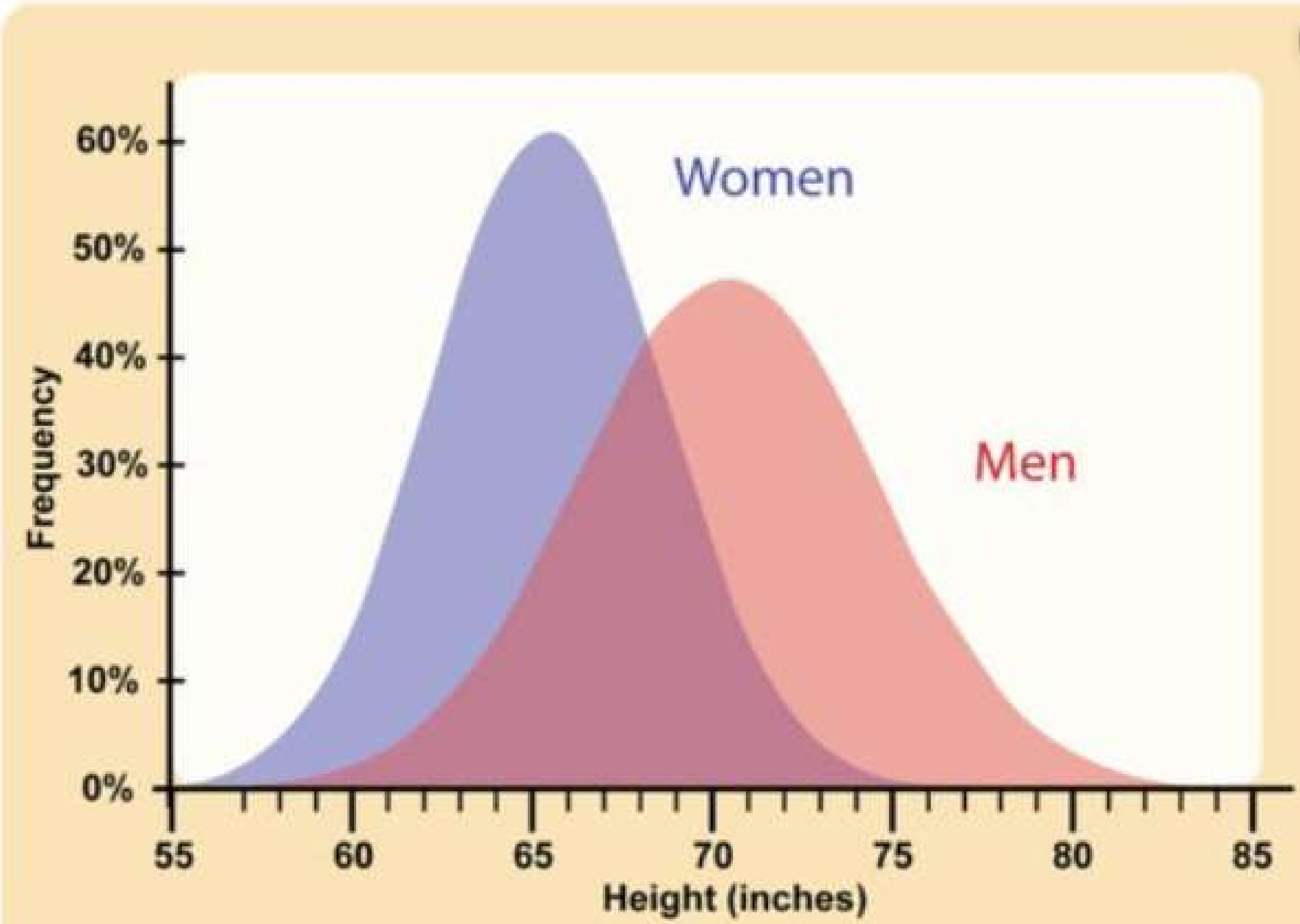
Data Analysis Unit
Measure of Center and Variation
pp 262-270 in textbook

Height of Adult Women and Men

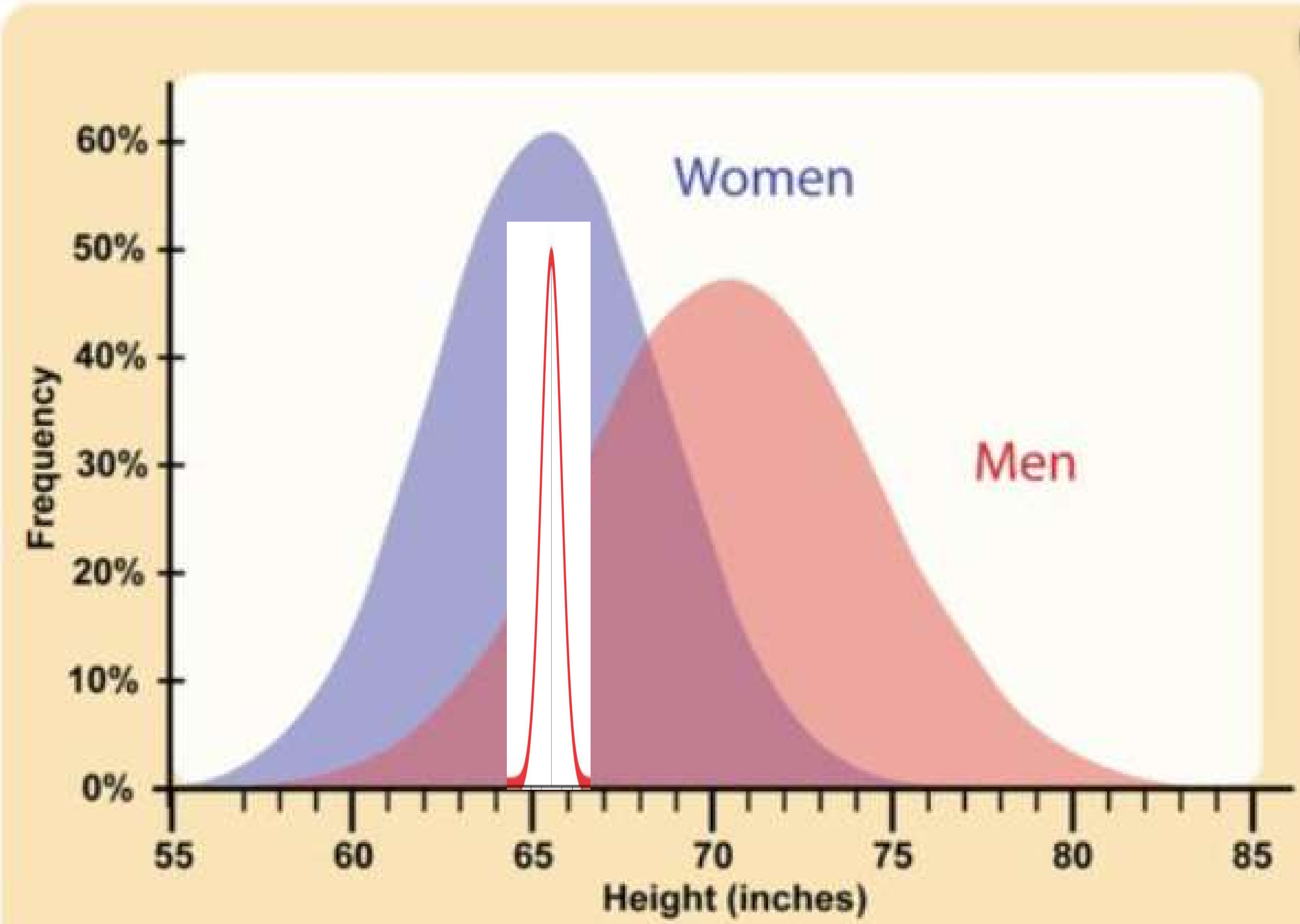
Within-group variation and between-group overlap are significant



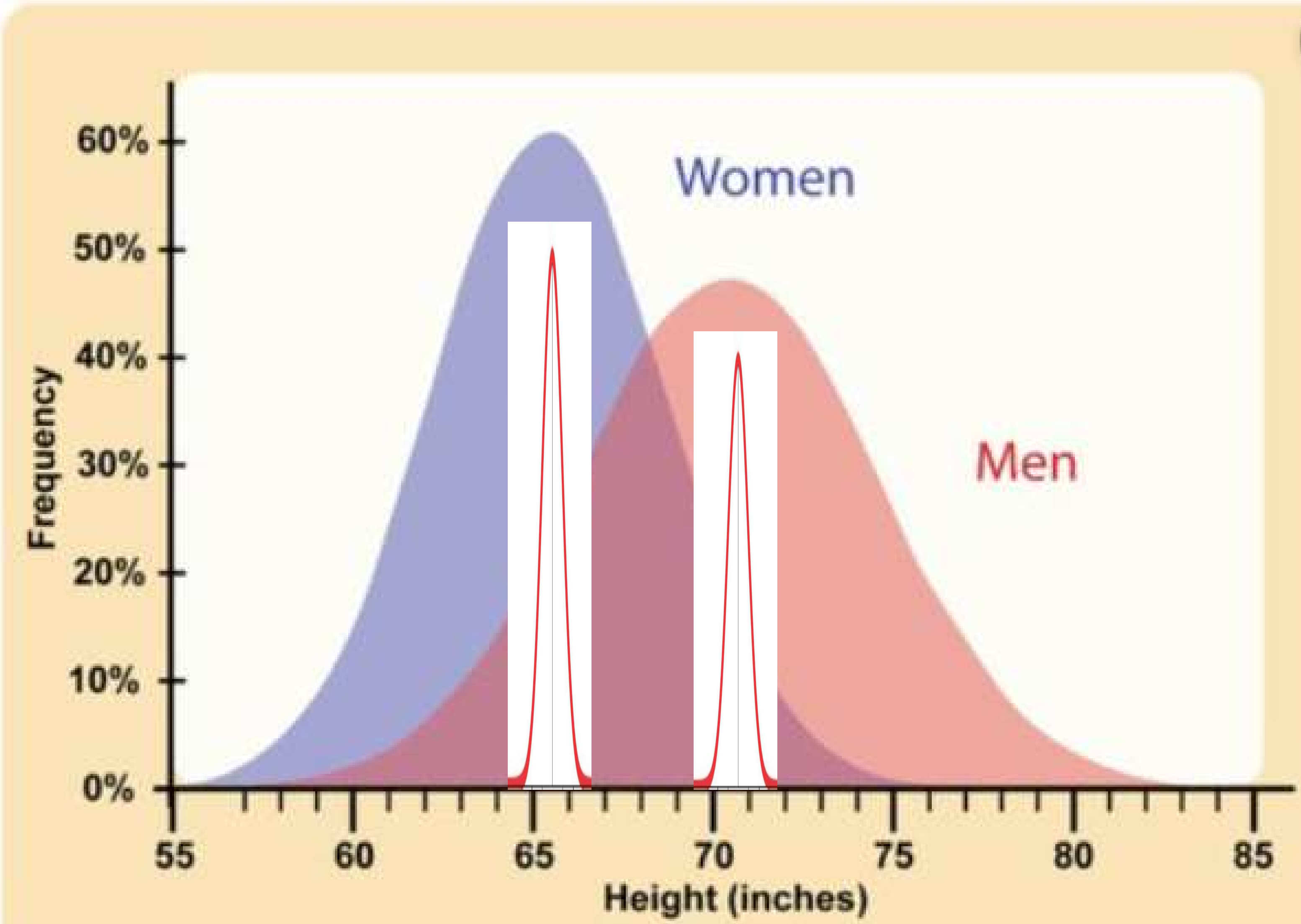
Data from U.S. CDC, adults ages 18-86 in 2007



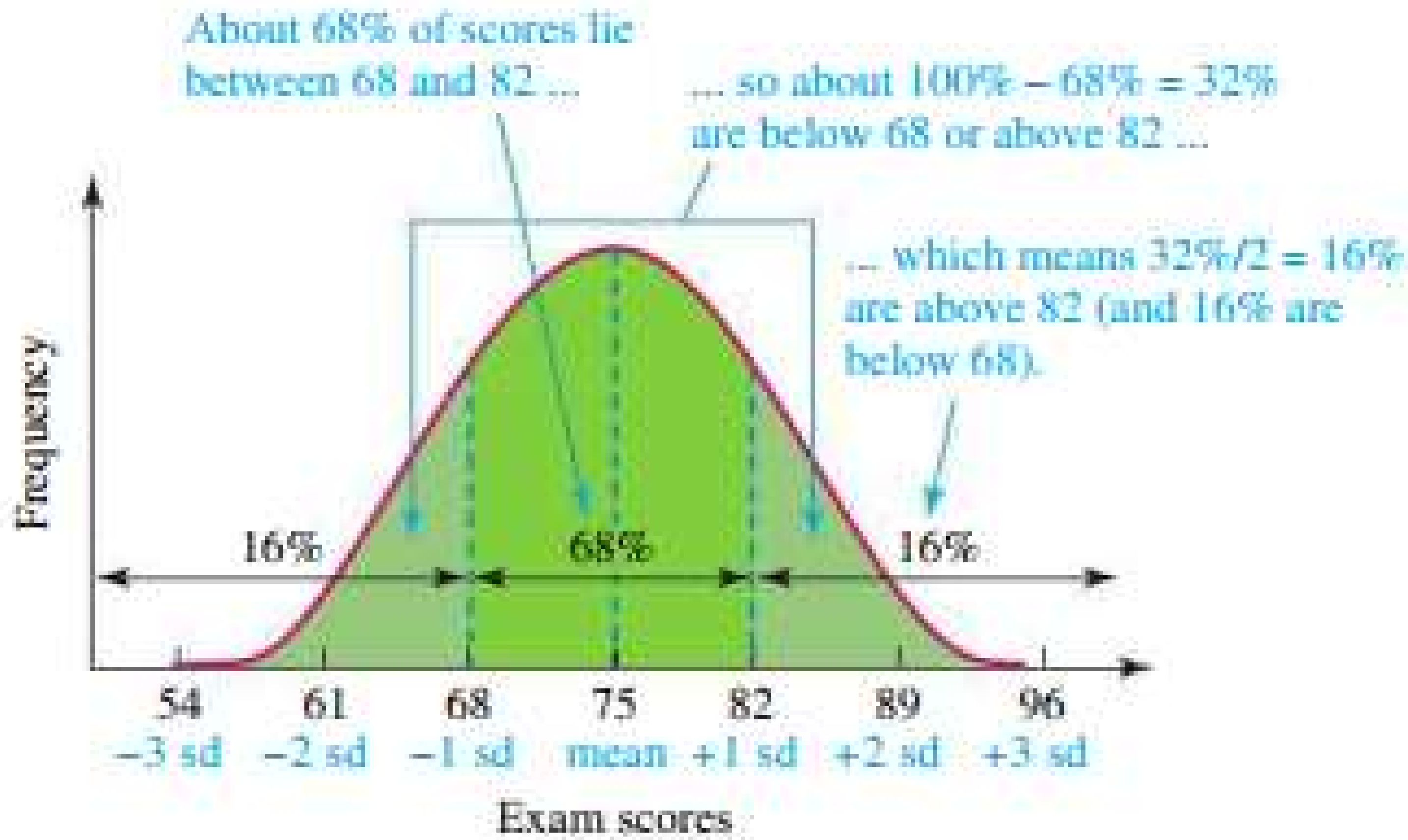
Women
Men
60%
50%
40%
30%
20%
10%
0%
Frequency
55
60
65
70
75
80
85
Height (inches)



Frequency
60%
50%
40%
30%
20%
10%
0%
Women
Men
55
60
65
70
75
80
85
Height (inches)



Women
Men
Frequency
60%
50%
40%
30%
20%
10%
0%
55
60
65
70
75
80
85
Height (inches)


About 68% of scores lie
between 68 and 82 ...
... so about 100% – 68% = 32%
are below 68 or above 82 ...
... which means 32%/2 = 16%
are above 82 (and 16% are
below 68).
Frequency
16%
68%
16%
54
61
68
75
82
89
96
–3 sd
–2 sd
–1 sd
mean
+1 sd
+2 sd
+3 sd
Exam scores

Measures of Center Tendency

The distribution of a variable (or data set) refers to the way its values are spread over all possible values. A distribution can be shown visually with a table or graph.

Mean

The **arithmetic mean**, is what we most commonly call the “average”. It is defined as follows

$$\text{mean} = \frac{\text{sum of all values}}{\text{total number of values}}$$

Median

The **median** is the middle value when the dataset is sorted in numerical order (or halfway between the two middle values if the number of values is even).

Mode

The **mode** is the most common value (or group of values) in a distribution.

Outlier

An **outlier** is a data value that is much higher or much lower than almost all other values. Outliers almost always affects the mean of a dataset.

Range

The **range** is the difference between the maximum value and the minimum value of the dataset.

Standard Deviation

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

A few important characteristics:

- Standard deviation is always positive. Standard deviation will be zero if all the data values are equal, and will get larger as the data spreads out.
- Standard deviation has the same units as the original data.
- Standard deviation, like the mean, can be highly influenced by outliers.

Standard Deviation (cont.)

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

Standard deviation can be written symbolically using the following formula

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

s = standard deviation
x_i = individual data value
$\bar{x}$ = mean
n = total number of data values
$\sum$ = summation or sum of

Examples

For the following dataset of T-shirts sold per week by a student who started his own online T-shirt business, find the mean, median, and mode:

n = 12

Mean ($\overline{x}$) = 7

Median = 6

Mode = 3

Range = 9

S.D. = 4.11

T-Shirts Sold per Week	Frequency
3	5
6	2
9	1
12	4

Range = 12-3 = 9

Mode → *3, 3, 3, 3, 3*, **6, 6**, 9, 12, 12, 12, 12

Min

Median = (6+6)/2 =6

Max

Mean & Standard Deviation

3, 3, 3, 3, 3, 6, 6, 9, 12, 12, 12, 12

Mean ($\bar{x}$) = (3+3+3+3+3+6+6+9+12+12+12+12)/12 = 7

$$\textbf{S.D.} = \sqrt{\begin{array}{c}[(3-7)^2+(3-7)^2+(3-7)^2+(3-7)^2+(3-7)^2+(6-7)^2 \\ +(6-7)^2+(9-7)^2+(12-7)^2+(12-7)^2+(12-7)^2 \\ +(12-7)^2]/11\end{array}}$$

= 4.11

Examples

For the following dataset of T-shirts sold per week by a student who started his own online T-shirt business, find the mean, median, and mode:

n = 12

Mean ($\bar{x}$) = 7

Median = 6

Mode = 3

Range = 9

S.D. = 4.11

T-Shirts Sold per Week	Frequency
3	5
6	2
9	1
12	4

Examples

For the following dataset of contract offers, find the mean, median, mode, range, and standard deviation:

\$50,000   \$80,000   \$100,000   \$90,000   \$10,000,000

Put in ascending order (n = 5)

\$50,000   \$80,000   ***\$90,000***   \$100,000   \$10,000,000

Mean = (50,000 + 80,000 + 90,000 + 100,000 + 10,000,000)/5

= \$2,064,000

Range = \$10,000,000 - \$50,000 = \$9,950,000

For the following dataset of contract offers, find the mean, median, mode, range, and standard deviation:

$50,000 $80,000 ***<u>$90,000</u>*** $100,000 $10,000,000

Mean = $2,064,000

$(50000 - 2064000)^2$ = $4.056196\ X\ 10^{12}$

$(80000 - 2064000)^2$ = $3.936256\ X\ 10^{12}$

$(90000 - 2064000)^2$ = $3.896676\ X\ 10^{12}$

$(100000 - 2064000)^2$ = $3.857296\ X\ 10^{12}$

$(10000000 - 2064000)^2$ = $6.2980096\ X\ 10^{13}$

Sum =$7.872652\ X\ 10^{13}$ Sum/4 = $1.968163\ X\ 10^{13}$

$$S_x = \sqrt{19.68163\ X\ 10^{12}} = 4{,}436{,}398.31$$

Examples

For the following dataset of contract offers, find the mean, median, mode, range, and standard deviation:

$50,000 $80,000 $100,000 $90,000 $10,000,000

n = 5 ***outlier: $10,000,000***

Mean ($\overline{x}$) = $2,064,000

Median = $90,000

Mode = none

Range = $9,950,000

S.D. = $4,436,398.31

Examples

For the following dataset of gallons of gasoline purchased by 28 drivers, find the mean, median, and mode:

7, 4, 18, 4, 9, 8, 8, 7, 6, 2, 9, 5, 9, 12, 4, 14, 15, 7, 10, 2, 3, 11, 4, 4, 9, 12, 5, 3

Sorted

2, 2, 3, 3, **4, 4, 4, 4, 4**, 5, 5, 6, 7, **7, 7**, 8, 8, 9, 9, 9, 9, 10, 11, 12, 12, 14, 15, 18

Mode 4 Median = (7+7)/2 = 7 Range = 18-2 = 16

Put numbers in calculator to find mean and Standard Deviation

Check You Work – Input numbers and check them

Examples

For the following dataset of gallons of gasoline purchased by 28 drivers, find the mean, median, and mode:

7, 4, 18, 4, 9, 8, 8, 7, 6, 2, 9, 5, 9, 12, 4, 14, 15, 7, 10, 2, 3, 11, 4, 4, 9, 12, 5, 3

n = 28

Mean ($\overline{x}$) = 7.54

Median = 7

Mode = 4

Range = 16

S.D. = 4.10

Examples – What about categorical data?

For the following dataset of vehicle colors:

n = 20
Mean ($\bar{x}$) = none
Median = none
Mode = Green
Range = none
S.D. = none

Color	Frequency
Blue	3
Green	5
Red	4
White	3
Black	2
Grey	3