9.8 Descriptive Statistics

Mean, Median and Mode

The average on the test was an 84 - MEAN

The average test score puts you in the middle of the class - MEDIAN

The average American student starts college at 18- MODE

Mean, Median and Mode

76

The **mean** of a list of n numbers $\{x_1, x_2, \ldots, x_n\}$ is:

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

the **mean** is strongly effected by <u>outliers</u>

The **median** of a list of n numbers $\{x_1, x_2, \ldots, x_n\}$ arranged in order (either ascending or descending) is:

- The middle number is n is odd
- The mean of the two middle numbers if n is even

Median is <u>resistant</u> meaning it is not strongly effected by outliers

The **mode** of a list of numbers is the number that appears most frequently.

Five number Summary

Range = maximum - minimum

Quartiles split the data into **fourths**

First Quartile (Q_1) = the median of the lower half of the data
Second Quartile = the median

Third Quartile (Q_3) = the median of the upper half of the data

Interquartile Range (IQR) measures the spread between Q_1 and Q_3

$$IQR = Q_3 - Q_1$$

Five number summary = {minimum, Q_1, median, Q_3, maximum}

Find the five number summary for the male and female life expectancies in South American nations and compare.

males: {59.0, 60.5, 61.5, 66.7, 67.9, 68.5, 69.0, 70.3, 71.4, 71.9, 72.1, 72.6}

females: {66.2, 66.7, 67.7, 72.8, 74.3, 74.4, 74.6, 76.5, 76.6, 78.8, 79.0, 79.4}

MEDIAN: 68.75

$Q_1 = 64.1$

$Q3 = 71.65$

$\{59, 64.1, 68.75, 71.65, 72.6\}$

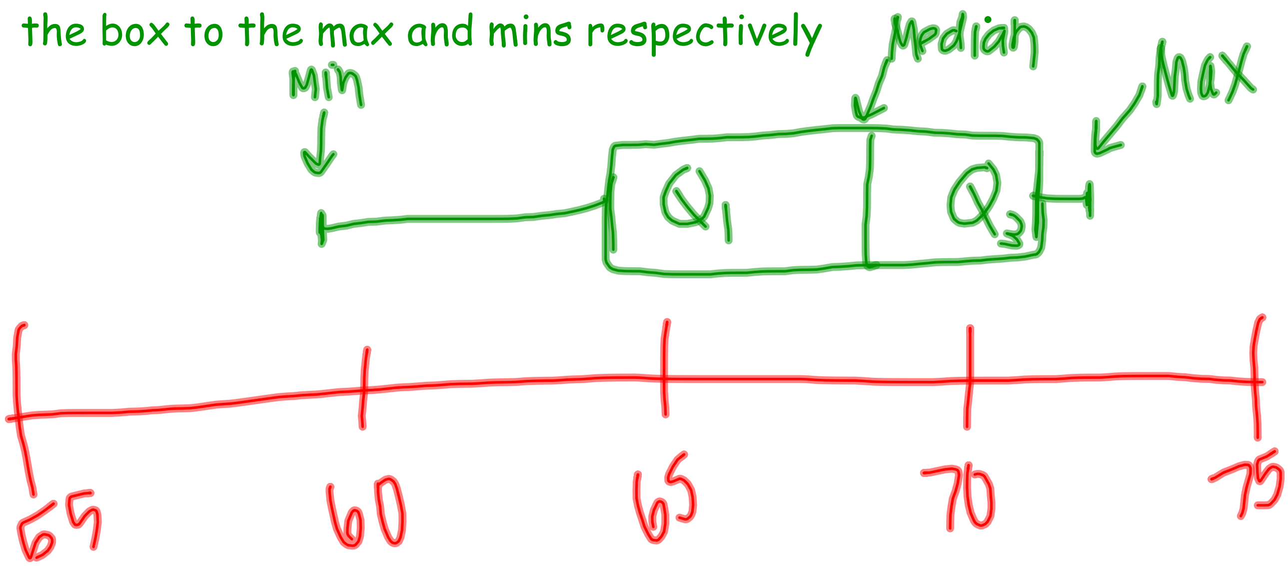
Range: $72.6 - 59 = 13.6$

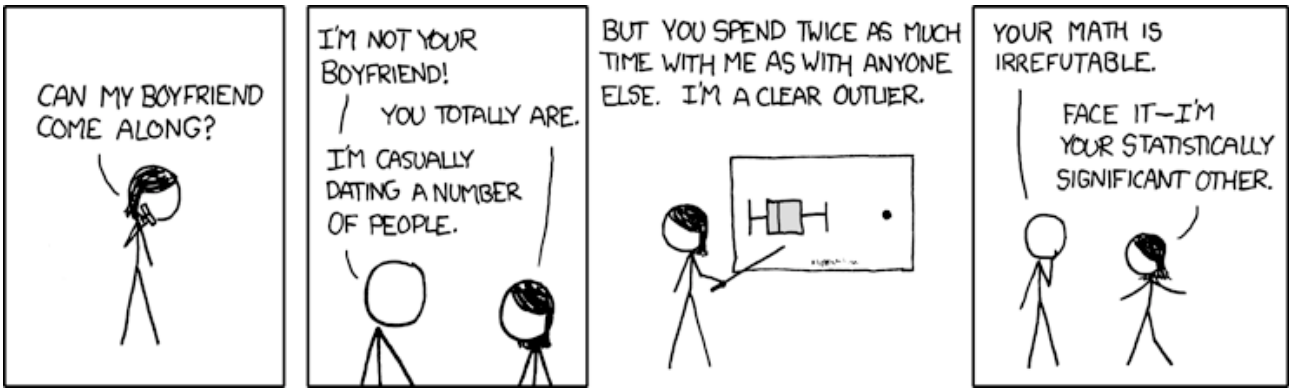
IQR: $Q_3 - Q_1 = 71.65 - 64.1 = 7.55$

A **box plot** (sometimes called box and whisker plot) is a graph that depicts the five number summary of a data set.

To Construct: $\{59, \boxed{64.1, 68.75, 71.65,} 72.6\}$

1. Draw a rectangular box from Q_1 to Q_3 with a vertical line for the median
2. Draw line segments (whiskers) that extend from the end of the box to the max and mins respectively





Box and Whisker plots allow us to get a good visual of outliers: a number that makes one of the whiskers noticeably longer than the box:

RULE OF THUMB: a number is considered an <u>outlier</u> if it is more than 1.5 X IQR below Q_1 or above Q_3

Is 61 an outlier in Roger Maris's home run data? $\boxed{\text{yes}}$

Five number summary = {5, 11, 19.5, 30.5, 61}

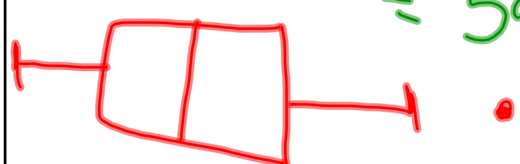
$61 > 1.5 \underline{\times IQR} + Q_3$

$IQR = Q_3 - Q_1$

$30.5 - 11 = 19.5$

$1.5(19.5) + 30.5$

$= 59.75$



Variance and Standard deviation

78

Measures variability

The $\boxed{\text{standard deviation}}$ of the numbers $\{x_1, x_2, \ldots, x_n\}$ is

Sigma (lowercase)

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

Sigma (uppercase)

where $\bar{X}$ denotes the mean. The variance is σ^2 the square of the standard deviation.

By hand this can be tedious- luckily we can do this in our calculator.

Weights in grams of 30 loon chicks

79.5	87.5	88.5	89.2	91.6	84.5	82.1	82.3	85.1	89.8
84.0	84.8	88.2	88.2	82.9	89.8	89.2	94.1	88.0	91.1
91.8	87.0	87.7	88.0	85.4	94.4	91.3	86.3	85.7	86.0

Standard deviation: $\sigma = 3.4$

Variance: $\sigma^2 = 11.56$

Mean: $\bar{x} = 87.5$

68-95-99.7 Rule

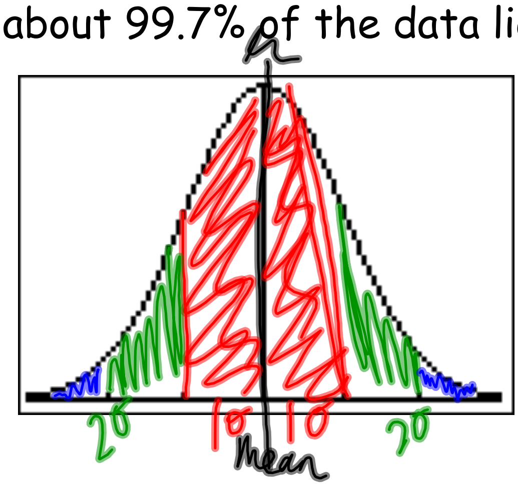
79

If the data for a population are normally distributed with mean μ and standard deviation σ then,

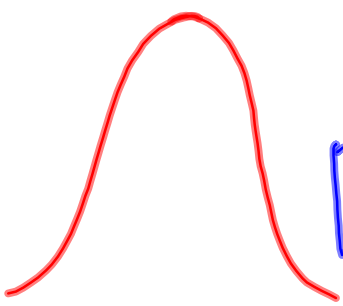
about 68% of the data lie between $\mu - 1\sigma$ and $\mu + 1\sigma$

about 95% of the data lie between $\mu - 2\sigma$ and $\mu + 2\sigma$

about 99.7% of the data lie between $\mu - 3\sigma$ and $\mu + 3\sigma$



would a chick(bird) weighing 95 grams be in the top 2.5%



$\boxed{\text{YES}}$

$100-2.5=\boxed{97.5\%}$

$1\sigma = 68\%$

$2\sigma = 95\%$

$3\sigma = 99.7\%$

$\mu = 87.5$

$\sigma = 3.5$

$87.5+2(3.5)=94.5$