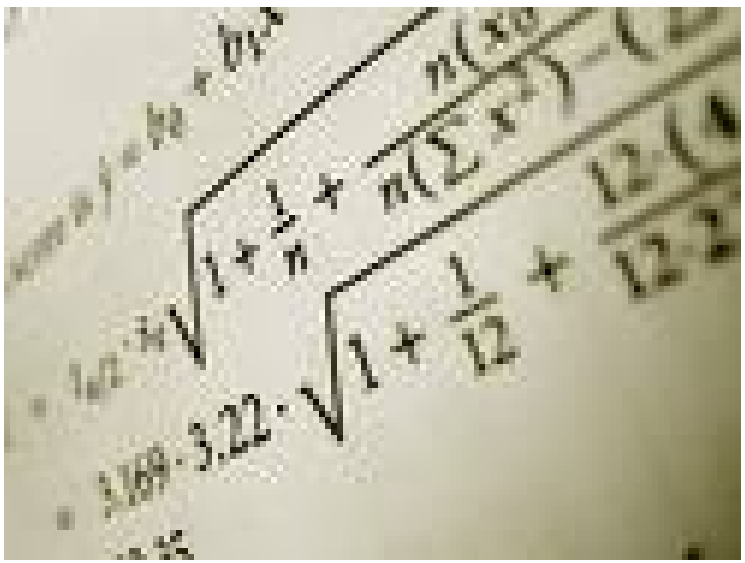
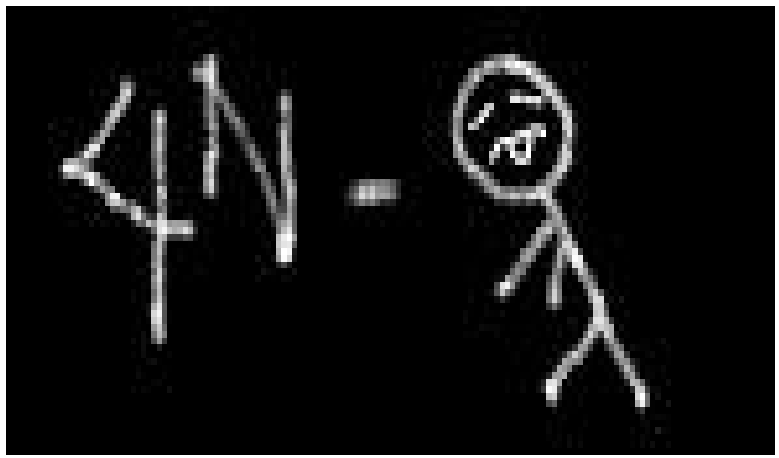
Statistics in Forensics

January 2009





What can STATS in forensics tell us?

- Allows us to quantify events
- Allows us to measure relationships in the data
- Allows us to make meaningful comparisons between groups.
- Gives us a means of testing hypotheses.
- Allows us to predict the probability of future outcomes
- Allows us to draw objective conclusions from data.

Two basic types of STATS

- Descriptive Statistics (data base)
  - Descriptive statistics give us a way to summarize and describe our data but do not allow us to make a conclusion related to our hypothesis.
- Variables
- distributions

- Inferential Statistics
  - As the name suggests inferential statistics attempt to make an inference about our data.
- Populations
- Probability
- Sampling
- Matching

Similarities vs differences in statistical tendencies

- **Stats that show how different units seem similar.**
- **This statistic is often called a measure of statistical tendency.**
- **Mean**
- **Median**
- **Mode**

- **Stats that show how different units seem to differ.**
- **This statistic is often called a measure of statistical variability**
- **Range**
- **Variance**
- **Standard deviation**

Measure of Central Tendency

Describe where the data points cluster

- Mean
- Median
- Mode

Mean

Mean= average

Found by taking the sum of the numbers and then diving by how many numbers you added together

Example: 3, 4, 5, 6, 7 (total amount of numbers=5)

3+4+5+6+7= 25

25/ 5 = 5

Median

- When numbers are arranged in numerical order, the MIDDLE number is the median.
- Ex: 3,6,2,5,7,
- Arrange in order: 2,3,5,6,7
- The middle number is the 5
- The median is 5

Mode

- The number that occurs the most frequently is the MODE
- Ex: 2,2,2,4,5,6,7,7,7,7,8
- 7 is the number that occurs the most frequently (the most times)
- The mode is 7

Measures of Variability

Describe the dispersion of the values

Range

Variance

Standard deviation

Range

- The difference between the largest and smallest scores
- Example: 2,3,4,6,8,10
- 10-2=8
- The range is 8

Variance

- Measure how spread out a distribution is.
- The variance is computed as the average squared deviation of each number from its mean. For example, for the numbers 1, 2, and 3, the **mean is 2** and the variance is:

$$\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = 0.667$$

Standard deviation

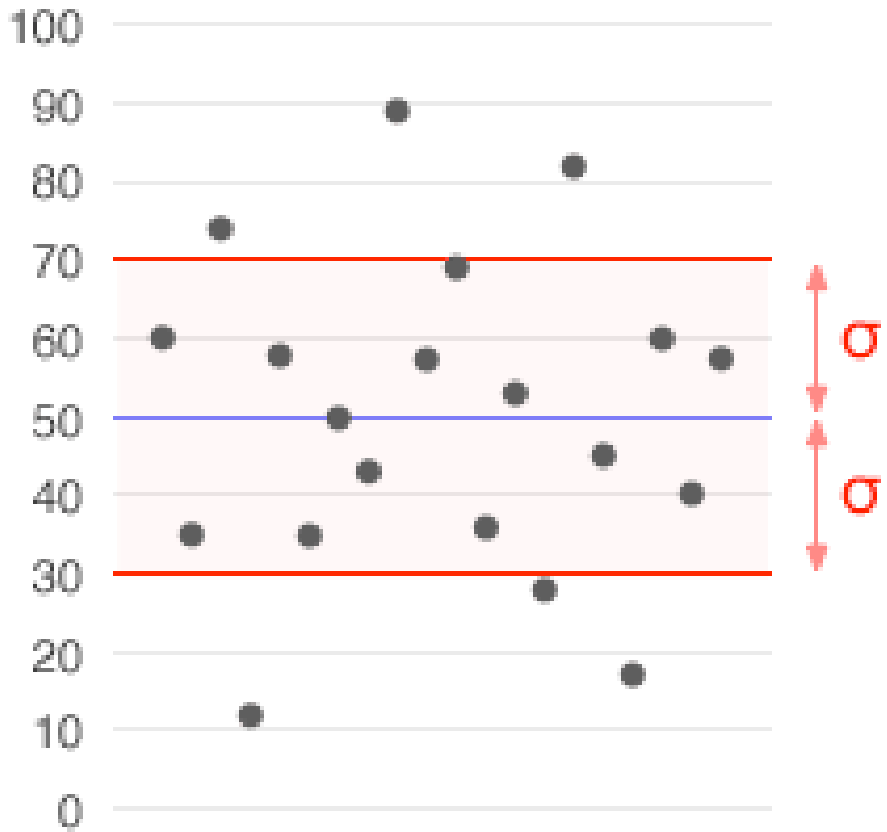
- Is like the mean of the mean…
- Or the average of the average…
- The standard deviation formula is very simple: it is the square root of the variance.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

Standard deviation (σ)

- In probability and statistics, the **standard deviation** of a collection of numbers is a measure of the dispersion of the numbers from their expected (mean) value.
- The standard deviation is usually denoted with the letter σ (lowercase sigma).
- It is defined as the root-mean-square (RMS) deviation of the values from their mean, or as the square root of the variance.
- The standard deviation remains the most common measure of statistical dispersion, measuring how widely spread the values in a data set are.

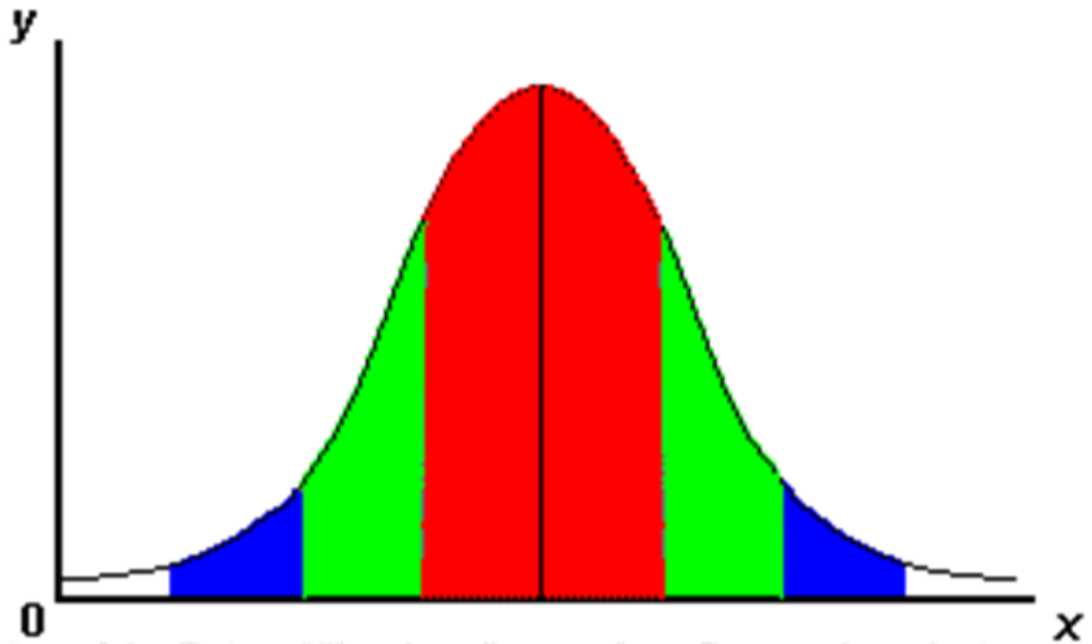
Deviation from the mean



- A data set with a mean of 50 (shown in blue) and a standard deviation (σ) of 20. (red lines)

Standard deviation graph

- Let's say a group of students take the SAT test and score an average of 500 in reading. The red columns would represent one deviation away from the average. This accounts for 68% (34% on either side) of all scores.
- The green columns represent two deviations away from the average. This accounts for 27% more (13.5% more on either side)
- The blue columns represent three deviations away from the average. This account for an additional 4% (or 2% on each side)



From the SAT College Board Testing Site:

- Mean=The mean is the arithmetic average.
- (σ) =The standard deviation (SD) is a measure of the variability\of a set of scores. If test scores cluster tightly around the mean score, as they do when the group tested is relatively homogeneous, the SD is smaller than it would be with a more diverse group and a greater scatter of scores around the mean.

Almost done!
Let's try one example of variance and standard deviation!!

- Suppose we wished to find the standard deviation of the data set consisting of the values 3, 7, 7, and 19.
- It takes FIVE steps to complete.
- Step ONE: find the mean (average)

Almost done!
Let's try one example of variance and standard deviation!!

- Suppose we wished to find the standard deviation of the data set consisting of the values 3, 7, 7, and 19.
- It takes FIVE steps to complete.
- Step ONE: find the mean (average)

$$\frac{3+7+7+19}{4} = 9.$$

Step 2

- find the deviation of each number from the mean

**This is where you subtract each number from the mean

Step 2

- find the deviation of each number from the mean

**This is where you subtract each number from the mean

$$3 - 9 = -6$$

$$7 - 9 = -2$$

$$7 - 9 = -2$$

$$19 - 9 = 10.$$

Step 3

- square each of the deviations, which amplifies large deviations and makes negative values positive

Step 3

- square each of the deviations, which amplifies large deviations and makes negative values positive

$$(-6)^2 = 36$$
$$(-2)^2 = 4$$
$$(-2)^2 = 4$$
$$10^2 = 100.$$

Step 4

- find the mean of those squared deviations
- This is easy…just find the average!

Step 4

- find the mean of those squared deviations
- This is easy…just find the average!
- This is the variance= σ^2

$$\frac{36+4+4+100}{4}=36.$$

Step 5

- Find the standard deviation by squaring the variance!

$$\sqrt{36} = 6$$

- So, the standard deviation of the set is 6

Statistics is the branch of mathematics that is usually employed to quantify "confidence".