



Module 2: Analyze Your Data

Why Statistics

- In any science project you are going to compare two groups
- Comparing the before and after you did an experiment
- Or just comparing two groups to see if there is a difference or not
- Statistics is the way to measure each group and compare two sets of data

MEAN

- The mean is the statistical term for the simple average. To calculate the mean add up all the values for a group and then divide by the total number of items in that group.
- The mean is one of several ways of looking at a group of numbers or data to analyze and find meaningful information.

Popcorn Mean

- In the popcorn data for brand A there were five samples that had values of:
86, 82, 84, 78, and 91, for a total of 421.
- Taking the total value and dividing by the total number of samples yields:
 $421 / 5 = 84.2$ This is the mean value.
- No sample actually had 84.2 kernels, but the mean is the average of the numbers.

MEDIAN

- The median is another statistical tool to analyze a group of numbers to find meaningful information.
- The median is the value that has half the numbers above it and half the numbers below it when the numbers are arranged in order.
- For a group of numbers that has an odd number of values, it is the middle value.

Popcorn Median

- For an even group of numbers it is the mathematical calculation halfway between the two values that are closest to the middle.
- In the popcorn data for brand A the middle value is 84, because two values are above it and two values are below it.
- For brand B: 85, 88, 96, 101 the median is halfway between 88 and 96 = 92.

MODE

- The mode is the value that appears most frequently.
- For small data sets there might not be a mode, if no number is repeated.
- The mode is more useful in very large data sets where one particular number may appear more frequently than all the rest.

Popcorn Mode

- For popcorn brand A there is not mode.
- For this data set:
86, 82, 84, 78, 91, 88, 85, 96, 101, 82, 83, 85,
82, 81, 80, 94, 97, 79, 93, 78, 87, 89, 99
The mode is 82 because it appears three times, which is more than any other number.

Range

- Range is a measure of how big your data set is
- The difference between the highest value and the lowest value

- For this data set:

86, 82, 84, 78, 91, 88, 85, 96, 101, 82, 83, 85,
82, 81, 80, 94, 97, 79, 93, 78, 87, 89, 99

the range is $101 - 78 = 23$

Variance

- Variance is a way of describing numerically how much variation there is for the data points from the mean or average of the group
- To calculate the variance, calculate the difference between each value and the mean, (which could be positive or negative), then square each value, add them all up, and divide by either N or $n-1$

Variance

- For the whole population or for large samples of over 30 values, N = total number of values
- For small sample sizes, less than 30, $n-1$ is equal to the number of values minus one

Variance Popcorn # A

Value	Difference (Value - Mean)	Difference ²
86	1.8	3.24
82	2.2	4.84
84	0.2	0.04
78	6.2	38.44
<u>91</u>	6.8	<u>46.24</u>
		92.8

$$421 / 5 = 84.2 = \text{Mean}$$

Variance of Popcorn # A

- The sum of the differences squared is 92.80
- Since this sample size is five which is less than 30, we use the $n-1$, or $5 - 1 = 4$
- 92.80 divided by $4 = 23.2$
- Variance is 23.2

Standard deviation

- Standard deviation is another way to express how much each of the values differs from the mean of the group
- Standard deviation is defined as the square root of the variance
- For Popcorn sample A it would be
the square root of $23.2 = 4.81$
- Standard deviation for this group is 4.81

Parametric Statistics

- The statistics that we have been talking about, such as, mean and standard deviation are known as parametric statistics.
- Parametric statistics come from the standard normal distribution.