Statistics Notes

https://www.youtube.com/watch?v=9FtHB7V14Fo&list=PL5102DFDC6790F3D0

Basic terms

Term	Explanation
Data set	A collection of numerical results from measurements or experiments.
Variable	The item or characteristic that is being measured.
Mean (average)	The sum of all values divided by the number of values.
Median	The middle value when all results are arranged in order.
Range	The difference between the highest and lowest values.
Standard deviation (SD)	A measure of how spread out the values are; larger values indicate more variation.

When the mean is about twice as large as the standard deviation, it shows that most values are close to the average, with few extreme results, meaning the data is consistent.

Quarter Data

Term	Full Name	Meaning	What It Shows
Q1	First Quartile	Value below which 25% of data fall	Lower boundary of the middle 50%
Q2	Second Quartile (Median)	Middle value (50% below, 50% above)	Central point of data
Q3	Third Quartile	Value below which 75% of data fall	Upper boundary of the middle 50%
IQR	Interquartile Range	Q3 − Q1	Spread of the middle 50% of results

Term	Definition	Purpose
Frequency	The number of times a particular value or category appears in a data set	Shows how often each response or outcome occurs
Relative Frequency	The proportion of the total that a category represents	Allows comparison between groups of different sizes
Percentage Frequency	The relative frequency expressed as a percentage	Makes results easier to interpret and compare
Cumulative Frequency	The running total of frequencies up to each category	Shows how many observations fall below a given value
Cumulative Percentage	The running total of percentage frequencies	Shows the percentage of data up to a certain point

Comparing two groups

Term	Explanation
Difference	The result of subtracting one group's mean from another.
Null hypothesis (H_0)	The assumption that there is no difference between the groups.
Alternative hypothesis (H_1)	The statement suggesting there is a real difference between the groups.
Two-tailed test	A test that checks for any difference, in either direction.
One-tailed test	A test that checks for a difference in one specific direction only.

Term	Meaning	Explanation
μ_1 (mu 1)	Population mean for group 1	Represents the true average value for the first population being tested.
μ_2 (mu 2)	Population mean for group 2	Represents the true average value for the second population being tested.
t	Test statistic	Indicates how far the difference between the sample means is from zero, in terms of standard error.
p	p-value	Shows the probability of obtaining a test result as extreme as, or more extreme than, the one observed if the null hypothesis is true.
df	Degrees of freedom	The number of independent values that can vary in a statistical calculation after certain constraints are applied.
F-test	Test for equality of variances	Used to check whether two samples have similar or different levels of variation. Helps decide which type of t-test to use.

Term	Meaning
Criterion	The rule used to decide whether to reject or not reject the null hypothesis.
Level of significance	The probability of wrongly rejecting the null hypothesis when it is true.
Sample	The subset of data collected from the population for analysis.
Population	The entire group that the research aims to draw conclusions about.
Test statistic	The calculated value from the sample used to test the hypothesis.
Alpha (α)	The chosen threshold for the level of significance, often 0.05.
Alpha over 2 (α/2)	Half the significance level, used when performing a two-tailed test.
Graphs	Visual displays of data patterns such as bar charts or histograms.
UCB (Upper Class Boundary)	The highest value included in a class interval for grouped data.
Class Mark	The midpoint of a class interval, found by averaging its lower and upper boundaries.
Relative Frequency	The proportion of observations in a class compared with the total number of observations.