BRIDGE Patient to Investigator Training Optional Module
Overview of Statistical Analysis Summary Document

What is Statistics? Statistics is a branch of math focused on the collection, analysis, interpretation and presentation of numerical data or information.

How do you use Statistics in a Research study?

- It is used to **summarize** the data
- Data is summarized in **same way across all studies**
- Presents **quantitative** data
- Helps the reader to **draw conclusions**

Example of Statistics in Research:
All studies use percentages or proportions to describe the demographics of individuals enrolled in the study:
For example:

- "**50% or half of the participants were women"**
- **"66% or two-thirds were white"**
- **"75% or three-fourths were over 65**"

Statistical Analysis = Quantitative Data

In a research study:

- survey answers,
- the number, race, age of participants,
- patients' physical traits (weight, and height),
- patients' performance on clinical tests (like lung function or exercise endurance test) and
- even an individual's perceptions of his/her quality of life...

... **can ALL be *measured*** and **represented with *numbers* –**this is ***quantitative data.***

Mean, Median and Mode

When reviewing and evaluating results from a study—it is often helpful to take all the results/measurements—the full set of numbers (**quantitative data**)—and ***find the center***. The easiest way to find this center is through ***Mean, Median and Mode:***

- **Mean**: The mean is the average of a set of numbers
- **Median**: The middle of a sorted list of numbers is the median. To find the median, place the numbers in value order and find the middle number. When there are two middle numbers, the average of those two numbers is the median.
- **Mode**: The mode is the number which occurs most often in a set of numbers.

***Statistical* significance** means the results of the study are *not* **likely to have occurred by chance** but are instead **likely to be attributable to a specific cause or association.** In other words, statistical significance is the **measurement of how confident we are that a difference or relationship exists between two variables** *(such as the treatment group versus the placebo group}* and is not just a result of chance.

What is a *P* value? The way significance is reported statistically. *P* value measures how likely the results could have just occurred from chance. The lower a *p* value , the more "real" the results (not a result of coincidence). **The *results* of a study are *statistically significant* when *p value is ≤ 0.05***.

Statistical tests (Chi-Square Test, T-Tests, Analysis of Variance, Regression Analysis) are used to *compare two or more groups of data/information (and p values are typically used to determine whether the differences observed are significant).*

Statistical Significance versus Clinical Significance

Statistical Significance	Clinical Significance
Signifies that results are ***unlikely to be from chance or coincidence***	Signifies that results can have ***an impact on health or health care***
Measures the ***reliability*** of the results	Measures the ***importance*** of the results
"There was a **difference between the two groups studied that did NOT occur from random chance"**	"What we learned from the **difference in results in the two groups studied is important and has practical applications**—we should consider using this in the future"
Statistically significant results may be **useful to researchers** to suggest a new direction or type of study or medication to study.	***Clinically significant results*** may be **useful to doctors and other clinicians, health systems, insurers, and researchers.**

Correlation/Causation

- ***Correlation*** is a measure of how things are related. In statistics, it is a statistical calculation of if and how much two paired variables are related.
- Two variables can have a "positive" correlation—both variables move in the same direction—increasing or decreasing together, in parallel. ***Example:*** *Tall people have bigger shoe sizes.*
- Or two variables can have a "negative" correlation—one variable increases while the other decreases. ***Example***: Students with increased absences have lower grades.
- Correlation works for *quantitative data*—data that can be *measured and summarized with numbers*.

- ***Causation*** means one variable or event is directly responsible for another event or variable.
- In statistics, **causation is extremely hard to prove**. ***Example:***
  - A study compares individuals who drink alcohol and those who don't.
  - The drinkers' group has a higher rate of pancreatic cancer than the non-drinkers.
  - But we cannot say if the one variable, drinking, is the cause of the cancer. There could be something else that caused the drinkers to be different from the non-drinkers.
  - For example, many people who are smokers, also drink alcohol. So did the drinkers' smoking cause the higher risk of cancer?