RESEARCH METHODS FOR CLINICAL INVESTIGATORS
Session 3:

Comparing Means or Proportions: How should I analyze my data?

Veronica G. Brown, PhD, MSPH

Objectives

At the end of the presentation, the audience will be able to:

- Define the differences between a mean and proportion
- Determine the use of a mean vs. proportion in statistical analysis
  - Ascertain if there is a difference between two group proportions using hypothesis testing
- Interpret the results from hypothesis testing

Mean

Average value of a sample

X= Number of observations in the sample with a certain characteristic

n= Total number of observations in the sample

Σ= Sum

$$\mathbf{x = \Sigma x / n}$$

Proportion

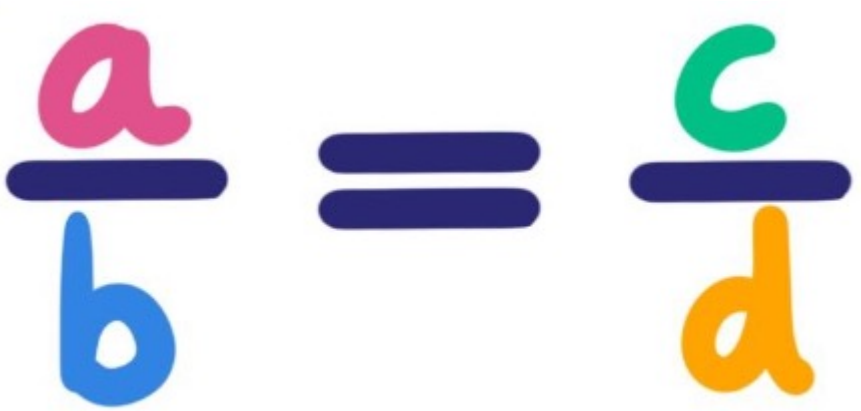
Amount of the sample that shares a commonality relative to its whole

X= Number of observations in the sample with a certain characteristic

n= Total number of observations in the sample

P= X / n

Proportion



Essentially, a proportion is a ratio

Hypothesis Testing

Statistical procedure used to make an assumption about the parameters of a population or its distribution

Null (H_0): Null hypothesis states that the difference is not statistically significant

Alternative (H_A): Alternative hypothesis states that the difference is statistically significant

Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press

Hypothesis Testing cont'd

- Significance Tests
  - Used to decide whether to **Reject** or **Fail to Reject (Accept)** the Null Hypothesis
  - Involves calculation of a test statistic, which is compared with a critical value obtained from statistical tables

Baker, Charlotte (2014). Introduction to Epidemiology. Virginia Tech University

Hypothesis Testing cont'd

- The critical value is set by the significance level of the test

α (alpha): Usually set to .05 → If there is <5% chance of a result as extreme as the sample result, the Null hypothesis is rejected (Statistical significance)

- The significance level is the *chance* of rejecting the Null hypothesis

"The chance of rejecting the Null hypothesis when it is in fact true"

Baker, Charlotte (2014). Introduction to Epidemiology. Virginia Tech University

Hypothesis Testing cont'd

Significance Tests

1. What is the probability that a relationship exists between two variables?
   a. What type of variables are included in the data (Continuous vs. Categorical)?

Continuous: Numerical values within a specified interval, i.e. Height, Weight, BMI, etc.

Categorical: Division of nominal values into groups and specifies two or more categories but no intrinsic ordering, i.e. Binary variables (Yes/No), (Dead/Alive) **or** Multiple variable (Eye Color: Blue, Brown, Black, Green)

Hypothesis Testing cont'd

Significance Tests

2. If there is a relationship, how strong is the relationship?

Types of Tests

*T-tests: Continuous variables

*ANOVA: Continuous variables

Chi-square (X^2) or Fisher's Exact: Categorical variables

*Assumption: Distribution of the sample means are normally distributed

Hypothesis Testing cont'd

T-tests (Continuous Variables): Comparing means between two groups

1. One Sample T-Tests: Compare a sample mean to a given value
    1. Ex. Mean Age, BMI, Systolic Blood Pressure

2. Two Sample T-Tests: Compare a single group with a before and after **OR** Compare two different groups by a specific variable
    - Ex. Mean Age or BMI among men vs. women

Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press

Hypothesis Testing cont'd

- ANOVA (Continuous Variables): Comparing means between multiple groups
- Determines there is a difference **between** the groups, not what group is different
- Ex. Mean Age or BMI for 3 study groups taking different medications for blood pressure

1. Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press
2. Baker, Charlotte (2014). Biostatistics. Virginia Tech University

Hypothesis Testing cont'd

- Chi-square (Categorical Variables): Examination of the relationship between two or more variables

- Comparison of observed VS. expected distributions

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press

Hypothesis Testing cont'd

- P-values (Threshold for determining if results are statistically significant or not)

"Quantifies the probability of observing a difference or association as larger than actually observed if the Null hypothesis were true"

Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press

Hypothesis Testing cont'd

P-values

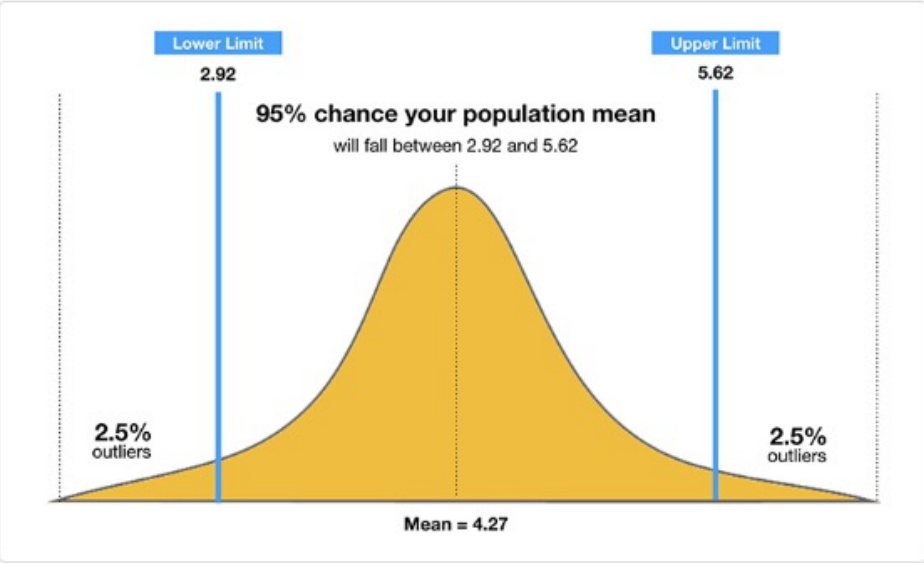
Symbols	Phrase	p-value
NS	Not Significant	P>0.05
=	Significant	P=0.05
*	Significant	P<0.05
**	Highly Significant	P<0.01
***	Extremely Significant	P<0.001

Motulsky, Harvey. (2014). *Intuitive Biostatistics: Nonmathematical Guide to Statistical Thinking*. Oxford University Press

Hypothesis Testing cont'd

Significance Tests → Measures of Effect

1. Confidence Intervals (CI)- Calculated interval of values that contain the true value of the population parameter



Gordis, Leon. (2018). *Epidemiology*. Saunders Elsevier.

Hypothesis Testing cont'd

Confidence Intervals (CI)

Calculated parameter for Measures of Association

1. Odds Ratio (OR): Case-Control
2. Relative Risks (RR): Cohort
3. Hazard Ratio (HR): Randomized Clinical Trials

Key questions to be asked every time you evaluate an association:

1.) Could the association have been observed by chance?

2.) Could the association be due to bias?

3.) Could other variables account for the observed relationship?

1. Gordis, Leon. (2018). *Epidemiology*. Saunders Elsevier.
2. Baker, Charlotte (2014). Introduction to Epidemiology. Virginia Tech University

Hypothesis Testing cont'd

- Confidence Intervals (CI)
  - 95% CI- Implies that the level of significance in the study is alpha=0.05 (p-value)
  - Answers the question of whether the null hypothesis fits inside the interval or not

Essentially:

No statistical association when the CI **includes** 1

Gordis, Leon. (2018). *Epidemiology*. Saunders Elsevier.

Hypothesis Testing cont'd

Ex. *(Adjusted OR: 0.83, 95% CI 0.67-1.03) [1]

Article Summary: Anticoagulation with heparin **did not** increase the likelihood of survival to hospital discharge or medical support for respiratory adverse events among patients diagnosed with COVID-19 vs those who received the standard thromboprophylaxis

1. Bradbury, C., McVerry, B., et al. (2021). Therapeutic Anticoagulation with Heparin in Critically Ill Patients with COVID-19. *N England J Med*. 385(9): 777-789. doi: 10.1056/NEJMoa2103417

Summary

- Mean: Average value of a sample
- Proportion: Amount of the sample that shares a commonality relative to its whole
- Hypothesis Testing: Statistical procedure to make an assumption about the parameters of a population or its distribution
  - Significance Tests
  - P-values
  - Confidence Intervals (CIs)

Summary

Hypothesis Testing:

- Reject the **Null Hypothesis:** p-value ≤ 0.05 statistically significant

- Accept the **Null Hypothesis:** p-value ≥ 0.05 is <u>not</u> statistically significant

Bottom Line: Do we reject or accept the **Null Hypothesis?**