Foundations in Biostatistics and Epidemiology

Session 3: Statistical inference and hypothesis testing

The following provides a brief summary of the material that is included in the session modules.

Part I: Terminology

- Null hypothesis: H_0
    - Typically a statement of no treatment effect
    - Assumed true until evidence suggests otherwise
- Alternative: H_A
    - Reject null hypothesis in favor of alternative hypothesis
    - Often two-sided
- Decisions and errors in hypothesis testing

		TRUTH	
		Association	No Association
STUDY	Reject Null	Correct	Type I Error *False positive*
	Fail to Reject Null	Type II Error *False negative*	Correct

- Significance level: α
    - Probability of a Type I error
    - Probability of a false positive
- Power: 1-beta ($1-\beta$)
    - Probability of detecting a true treatment effect
    - Power = (1- probability of a false negative)
      = (1-probability of Type II error)
      = ($1-\beta$) = probability of a true positive
- P-value
    - The probability of obtaining a difference at least as extreme as that obtained, provided the two groups are really equal (null hypothesis is true)
    - The probability that an observed difference in outcome is due to chance alone.
- Statistical Significance:
    - If the p-value of the calculated statistic is less than the alpha set in advance by the researcher (usually 0.05), then we can reject the null hypothesis and conclude that the groups are different.
    - Not all statistically significant differences are clinically significant
- Confidence interval
    - An interval estimate consisting of a range of values (with a lower and upper bound) constructed to have a specific probability (the confidence) of including the population parameter with repeated sampling.

- Holding the alpha and standard deviation constant: Increasing the sample size decreases the width of the confidence interval (tighter CI), and vice versa

Part II: Types of hypotheses

- Superiority:
  - A trial with the primary objective of showing that response to the investigational product is superior to a comparative agent (active or placebo control).
  - Note: Failure to show a significant difference is NOT the same as proving equivalence.
- Equivalence:
  - A trial with the primary objective of showing that the response to two or more treatments differs by an amount which is clinically unimportant.
  - This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence margin (δ) of clinically acceptable differences.
  - Statistical analysis is generally based on the use of confidence intervals (CI).
  - For equivalence trials, two-sided confidence intervals should be used.
  - Equivalence is inferred when the entire confidence interval falls within the equivalence margins ($-\delta$ to $+\delta$).
- Non-inferiority:
  - A trial with the primary objective of showing that the response to the investigational product is not clinically inferior to a comparative agent (active or placebo control)
  - Statistical analysis is generally based on the use of confidence intervals. A one-sided confidence interval should be used for non-inferiority trials.

Part III: Statistical hypothesis testing methods

- Normal distribution characteristics
  - Symmetrical about its mean μ (mirror image)
  - Mean, median, and mode are all equal (bell shaped)
  - The area between -1 standard deviation and +1 standard deviation from the mean is approximately 68% of total area under the curve
    - ± 2 standard deviation is about 95% of the total area under the curve
    - ± 3 standard deviation is about 99.7% of the total area under the curve

- Hypothesis testing methods
  - Sampling from a normal distribution or sample sizes large (>30)
    - T-test: Compare 2 independent means
    - Paired t-test: Compare 2 dependent means
    - ANOVA: Compare 3 or more independent means
    - Repeated Measures ANOVA: Compare 3 or more dependent means
    - Linear Regression
      - Describes the relationship between an explanatory variable (independent) and a <u>continuous outcome</u> variable (dependent)

        - Independent variable: categorical or continuous
        - How well does variable X predict variable Y?
        - Multiple linear regression – used to include multiple independent variables
    - Correlation
        - Describes the strength of a linear relationship between two continuous variables
        - Pearson correlation coefficient (r)
        - $-1 \leq r \leq 1$
- Nonparametric tests should typically be used if:
    - the variable (or a transformed version) does not have an approximately normal distribution
    - the distribution is unknown and cannot rely on large sample (>30) theory

Parametric	Nonparametric
One-sample t-test	Sign Test (ordinal data)
Paired t-test	Signed-Rank Test
t-test: 2 independent samples	Mann-Whitney Test Wilcoxon Rank Sum Test
Pearson Correlation	Spearman Correlation
ANOVA	Kruskal-Wallis 1-way ANOVA

- Compare proportions between two or more populations:
    - If the groups are independent – a general chi-square is appropriate
    - If the groups are dependent – a McNemar chi-square is appropriate
    - If expected cell counts are too small – a Fisher's Exact test is appropriate
- Logistic regression: Used to predict dichotomous outcome from an explanatory (independent) variable
    - Independent variable: categorical or continuous
    - Modeling concept similar to linear regression
    - Interpretation deals with log odds
    - Multiple logistic regression used to include multiple independent variables
- Time-to-event data
    - Kaplan-Meier method: estimate survival distributions
    - Log-rank test: Nonparametric method for statistically comparing survival distributions
    - Cox proportional hazards model: regression model for time-to-event outcome data

- Summary:

Outcome Variable	Explanatory Variable: Categorical	Explanatory Variable: Continuous
Categorical	Chi-square, Logistic Regression	Logistic Regression
Continuous	ANOVA, t-test, Linear Regression	Linear Regression/ Correlation