Jumping into Statistics: Introduction to Study Design and Statistical Analysis for Medical Research Using JMP Pro Statistical Software

FALL 2022/SPRING 2023

DR. TERRIE VASILOPOULOS, DR. CYNDI GARVAN,
& DR. PENNY REYNOLDS

Meet the Instructors

TERRIE VASILOPOULOS, PHD

Research Associate Professor in Anesthesiology and Orthopaedics and Rehabilitation

CYNTHIA GARVAN, PHD

Research Professor in Anesthesiology

PENNY REYNOLDS, PHD

Research Assistant Professor in Anesthesiology and Veterinary Medicine

Course Objectives

- Review fundamentals of study design and research methodology
- Understand how to choose best statistical test for your research question
- Practice basic statistical analysis use JMP Pro Software

Course Topics

- Life Cycle of Research and Asking a Good Research Question
- Choosing the Right Study Design for Your Research
- Clinical Trial Design
- Populations, Samples, and Hypothesis Testing in Medical Research
- Introduction to Data Types
- Best Practices in Data Collection and Database Management: Getting Started with SAS JMP Pro
- Summarizing and Visualizing Data
- Statistical Methods and How to Choose Them
- Risk Assessment Methods
- Introduction to Regression and Correlation
- Time-to-Event (Survival) Analysis
- Methods for Clinical Diagnostic Testing

Statistical Methods and How to Choose Them

12/14/2022

Learning Objectives

Participants will be able to:

1) Explain criteria in choosing a statistical method

2) Distinguish between parametric and non-parametric methods

3) Identify correct methods to compare groups

4) Conduct group comparisons in SAS JMP Pro

Why is this topic important?

The **GOAL** of statistical analysis is to understand something about populations using sample data.

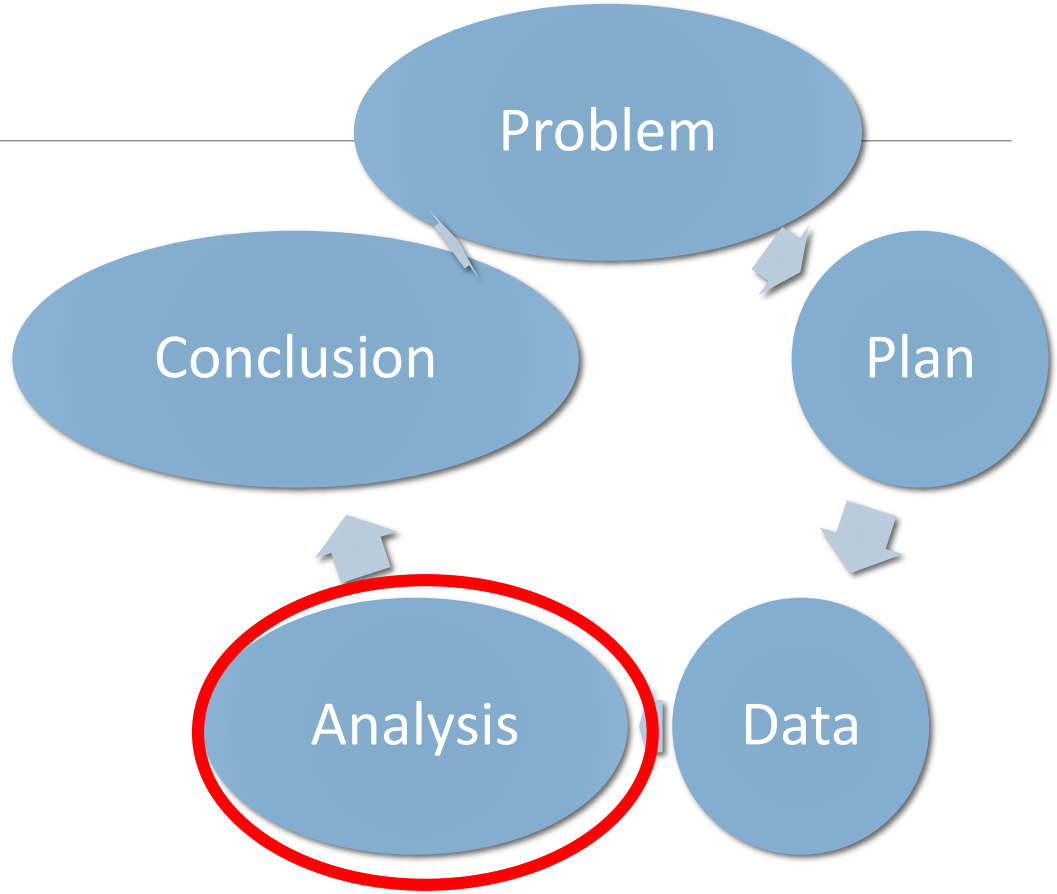
We need to make sure our choice of statistical method is correct for this understanding.

Overview
Statistical Methods

1. Introduction: What are the Criteria for Choosing a Statistical Method?
2. Assumptions for Parametric and Nonparametric Statistical Methods
3. Bivariate Methods to Analyze the Relationship Between Two Variables
    a. Chi-squared and Fisher's exact test
    b. T-test and Wilcoxon rank sum test
    c. ANOVA and Kruskal-Wallis test
    d. Pearson correlation and Spearman correlation

The Life Cycle of Research

In this lecture we will focus on statistical methods based on tests of hypotheses.



Translating the Research Question to a Test of Hypothesis

Translating the research question into a testable hypothesis is nontrivial.

Statisticians practice their craft for years and still find the translation of the research question into a test of hypothesis daunting.

Translation Example

In patients undergoing elective total knee arthroplasty (TKA), what is the effect of general anesthesia on post-operative pain compared to spinal anesthesia?

Step one in translation is to recognize

- The variables in the research question
- The data type of each variable

In this example, there are two variables:

1) Type of anesthesia (general versus spinal) *Categorical/Nominal*
2) Outcome (amount of post-operative pain medication taken) *Continuous*

Translation Example

In patients undergoing elective total knee arthroplasty (TKA), what is the effect of general anesthesia on post-operative pain compared to spinal anesthesia?

Step two : Formulate the null and alternative hypotheses.

- The null hypothesis is set up to be the “no difference” or “no relationship” hypothesis.
- In this step we need to recognize the population parameters which underlie the research question.
- In this example, mean amount of pain medication taken (μ) is of interest.

HYPOTHESIS TESTING OUTCOMES

Research	Reality: The Null Hypothesis Is True	Reality: The Alternative Hypothesis is True
The Null Hypothesis Is True	Accurate $1 - \alpha$ 	Type II Error β 
The Alternative Hypothesis is True	Type I Error α 	Accurate $1 - \beta$ 

Translation Example

Thus the hypothesis corresponding to: In patients undergoing elective total knee arthroplasty (TKA), what is the effect of general anesthesia on post-operative pain compared to spinal anesthesia?

Looks like:

In words	In symbols
Null: There **IS NO** difference in post-operative mean pain between the group of patients who receive general anesthesia and the group of patients who receive spinal anesthesia Alternative: There **IS** a difference in post-operative mean pain between the group of patients who receive general anesthesia and the group of patients who receive spinal anesthesia	$H_0: \mu_{general} = \mu_{spinal}$ $H_a: \mu_{general} \neq \mu_{spinal}$ Where μ is mean amount of post-operative pain medication taken

Choosing the Statistical Method

The choice of statistical method will depend on the number of variables in the research question.

Univariate methods: there is just one variable in the research question (example, paired t-test).

Bivariate methods: there are just two variables in the research question (example, independent samples t-test).

Multivariable methods: there are two or more variables predicting a ***single outcome*** variable in the research question (example, regression and logistic regression).

Mutlivariate methods: there are two or more ***outcome*** variables in the research question (example, longitudinal data analysis methods, repeated measures ANOVA, MANOVA, PCA, etc.).

In this lecture we will look at the bivariate statistical methods.

Choosing the Statistical Method

In addition to the number of variables in the research question, the choice of statistical method will also depend on data type. For the purposes of this lecture we will use the following data types to help us determine the correct choice of statistical method:

1) Binary data (i.e., two groups)
2) Categorical data (i.e., three or more groups or categories)
3) Ordinal data (i.e., values of categorical data can be meaningfully ordered)
4) Continuous data (i.e., interval or ratio data)

Translating the Research Question to a Test of Hypothesis and Choosing the Correct Statistical Method for Analysis

Translating the Research Question to a Test of Hypothesis

Step 1: Recognize how many variables there are in the research question. Identify the data type of each variable.

Step 2: Formulate the null and alternative hypothesis. The null hypothesis is set up to be the “no difference” or “no relationship” hypothesis. In this step we need to recognize the population parameter or characteristic which underlie the research question.

Choosing the Correct Statistical Method for Analysis

Depends on number of variables and data type of variables.

Parametric and Nonparametric Tests

Parametric versus Nonparametric Tests

Statistical tests are classified as **parametric** or **nonparametric** tests.

Parametric tests have a population parameter (such as μ) in the null and alternative hypotheses statements. This is why they are called "parametric tests."

Nonparametric tests have a population characteristic in the null and alternative hypotheses statements – such as location.

Warning!

There is no such thing as “non-parametric data” (even though this term appears in numerous journal articles).

Assumptions for Parametric and Nonparametric Tests

Assumptions for all statistical tests:

1) The sample is fairly representative of the population of interest
2) The data types of variables are appropriate for the test
3) Observations are independent (unless using a statistical method that allows for correlated measures such as methods to analyze repeated measures data)

Parametric versus Nonparametric Tests

For parametric tests there may be additional assumptions. For example:

1) The sample size is sufficiently large
2) The data are normally distributed

Advantages of nonparametric tests:

1) Have fewer assumptions compared to parametric tests
2) Can be used to analyze both ordinal and count data
3) Can be used even if data are very skewed
4) Can be used if there are extreme outliers which prevent a transformation to more normally distributed data
5) Are valid even when sample size is very small
6) Do not require normality of data

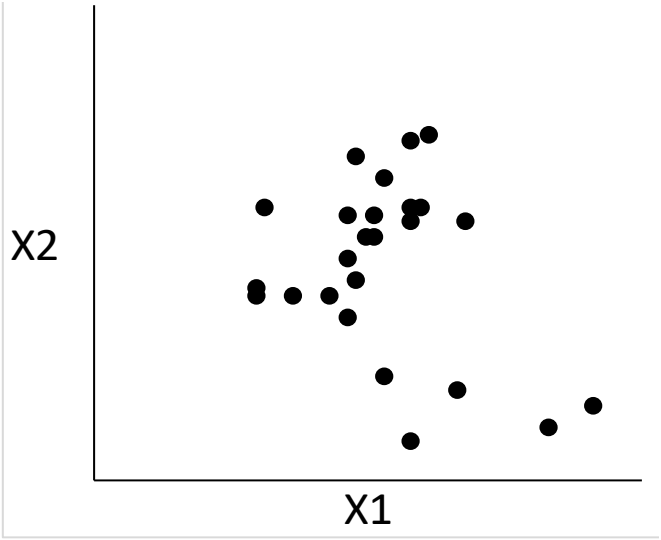
The Independent Observations Assumption

Unless you are using a statistical method which allows for two or more observations from a subject (e.g., Longitudinal Data Analysis methods, Hierarchal Linear Modeling, etc.), you may not include more than one observation for a statistical analysis.

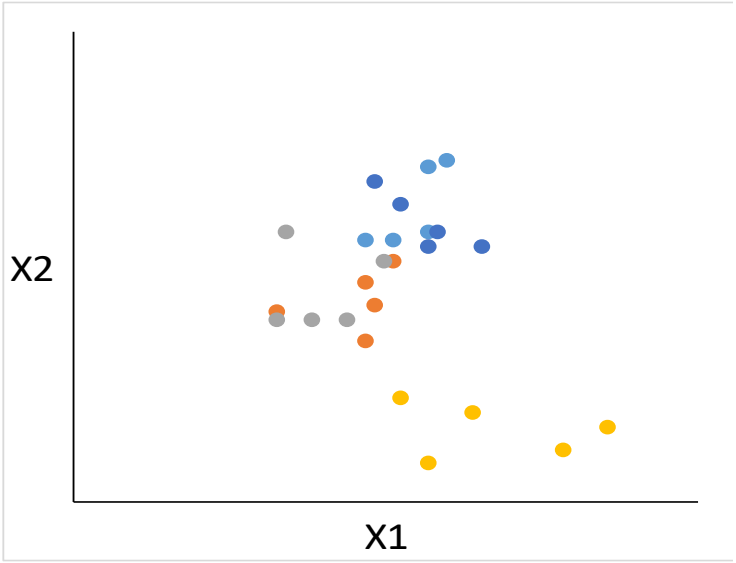
For example, if you are studying strength of hand grip, the data collected on a person's left hand is NOT a separate observation from the data collected on a person's right hand. The left hand data is NOT independent of the right hand data.

Warning Don't Do this! Example of Multiple Observations per Subject

In this data set, there were 25 observations on 5 subjects. Pooling of the data (which violates the test assumption) results in an apparent moderate negative relationship between X_1 and X_2: **r = -0.41**



Separating out observations by subject (each subject's data is represented in a different color) shows that there is no real relationship between X1 and X2. The correlation of -0.41 is an artifact of subject differences.



Having a sample which fairly represents the population of interest (i.e., a random sample) is a key assumption for the validity of every statistical test.

KEY POINT

Statistical tests are valid only when their assumptions are met. Both parametric and nonparametric tests have assumptions.

KEY POINT

Grip strength is measured on a patient's left hand and a patient's right hand. This is an example of independent data. True or False?

Your Turn

False: Two measurements made on the same individual are correlated (or dependent data).

Your Turn

Bivariate Methods to Analyze the Relationship Between Two Variables

Summary of Statistical Methods for Bivariate Analysis

In this lecture we focus on statistical methods for analyzing the relationship between two variables.

Let X = “Variable 1” and Y = “Variable 2”.

The X variable be a binary data type variable, a categorical data type variable, an ordinal data type variable, or a continuous data type variable. The Y variable be a binary data type variable, a categorical data type variable, an ordinal data type variable, or a continuous data type variable.

The choice of statistical method for analyzing the relationship between X and Y will depend on the data types of both variables.

Statistical Methods to Analyze the Relationship Between Two Variables

Bivariate Method		Data Type of Y Variable			
		Binary	***Categorical***	***Ordinal***	***Continuous***
Data Type of X Variable	***Binary***	Chi-squared test or Fisher's Exact test	Chi-squared test or Fisher's Exact test	Wilcoxon rank sum test	T-test or Wilcoxon rank sum test
	Categorical		Chi-squared test or Fisher's Exact test	Kruskal-Wallis test	ANOVA or Kruskal-Wallis test
	Ordinal			Spearman correlation	Spearman correlation
	Continuous				Pearson correlation or Spearman correlation

The correct statistical method to analyze the relationship between X and Y can be found by cross-tabbing the data type of each variable.

The BIS monitor algorithms are powered by deep mathematics (which is proprietary and therefore unknowable). However, anesthesiologists are able to use BIS monitors. The same is true for statistical methods. There is deep mathematics behind each method which pHD level statisticians learn and understand.
Clinicians do not need to understand the underlying mathematics of stats methods, but they need to know how to correctly apply methods and interpret results – just like anesthesiologists need to know correctly use a BIS monitor.

Linking to what you know

Back to the Translation Example

In patients undergoing elective total knee arthroplasty (TKA), what is the effect of general anesthesia on post-operative pain compared to spinal anesthesia?

In words	In symbols
Null: There **IS NO** difference in post-operative mean pain between the group of patients who receive general anesthesia and the group of patients who receive spinal anesthesia Alternative: There **IS** a difference in post-operative mean pain between the group of patients who receive general anesthesia and the group of patients who receive spinal anesthesia	$H_0: \mu_{general} = \mu_{spinal}$ $H_a: \mu_{general} \neq \mu_{spinal}$ Where μ is mean amount of post-operative pain medication taken

What is/are correct methods for analyzing these data?

In patients undergoing elective total knee arthroplasty (TKA), what is the effect of general anesthesia on post-operative pain compared to spinal anesthesia?

Let X = “Variable 1” and Y = “Variable 2”.

X = General anaesthesia type (Spinal versus General)

Y = Amount of pain medication taken

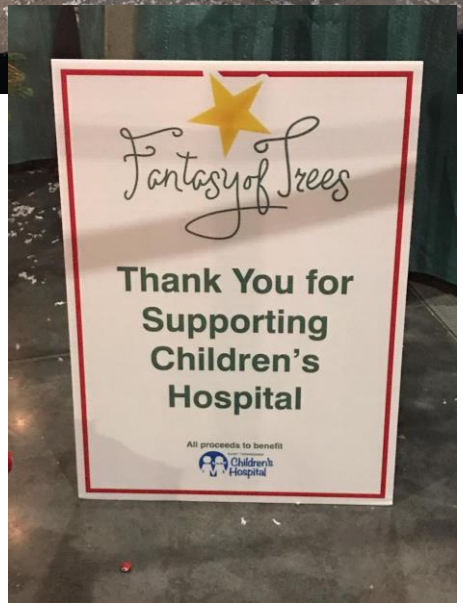
Statistical Methods to Analyze the Relationship Between Two Variables

Bivariate Method		Data Type of Y Variable			
		Binary	***Categorical***	***Ordinal***	***Continuous***
Data Type of X Variable	***Binary***	Chi-squared test or Fisher's Exact test	Chi-squared test or Fisher's Exact test	Wilcoxon rank sum test	T-test or Wilcoxon rank sum test
	Categorical		Chi-squared test or Fisher's Exact test	Kruskal-Wallis test	ANOVA or Kruskal-Wallis test
	Ordinal			Spearman correlation	Spearman correlation
	Continuous				Pearson correlation or Spearman correlation

The correct statistical method to analyze the relationship between X and Y can be found by cross-tabbing the data type of each variable.

Questions?


WARM
WINTER
WISHES


Fantasy of Trees
Thank You for Supporting Children's Hospital
All proceeds to benefit
Children's Hospital

Summary Tips

- Memorize the grid!

JMP Pro!

https://software.ufl.edu/