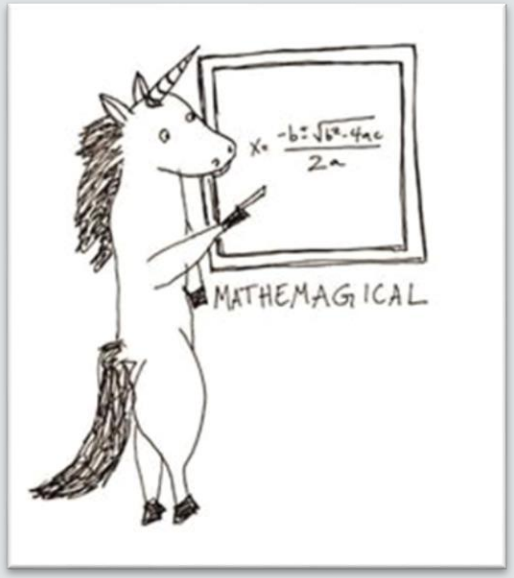
RESEARCH ESSENTIALS SEMINAR

Evaluating Your Data:
Types of Data and Basic Tests of Association

L. H. Domenico, PhD, RN, CARN
ldomenico@ufl.edu

Learning Objectives

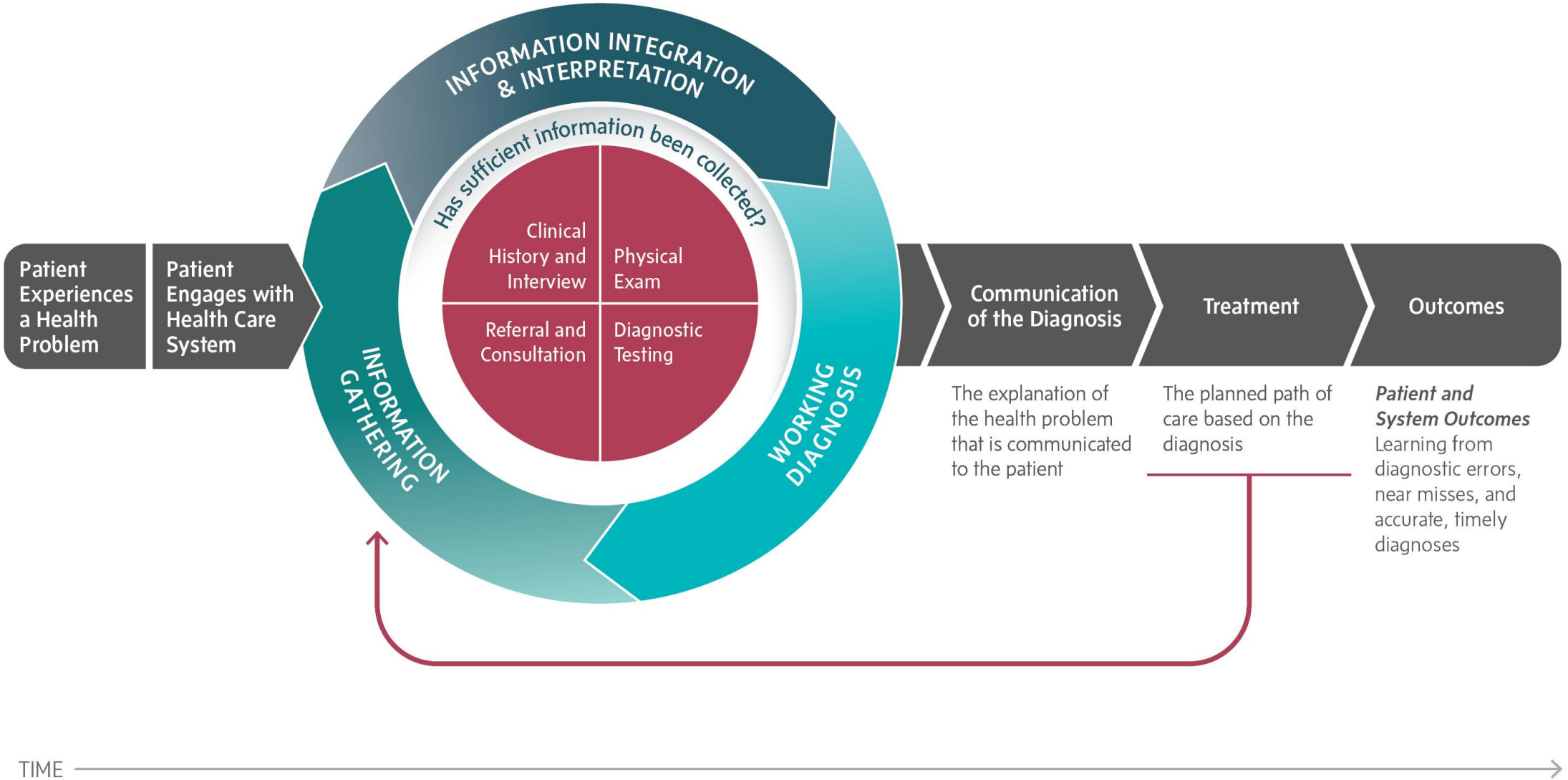
1. Describe the different *types* of data commonly encountered within medical research
2. Critically examine the theoretical principles guiding selected *univariate* and *multivariate* statistical approaches
3. Identify key factors that influence the selection of appropriate statistical approaches for addressing relevant research questions



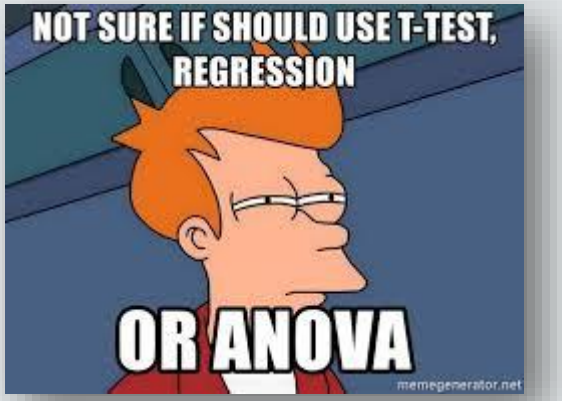
It’s about finding the right instrument to address the problem at hand!

It's about using established protocols

The Diagnostic Process



Choosing the Right Statistic: The Decision Making Process



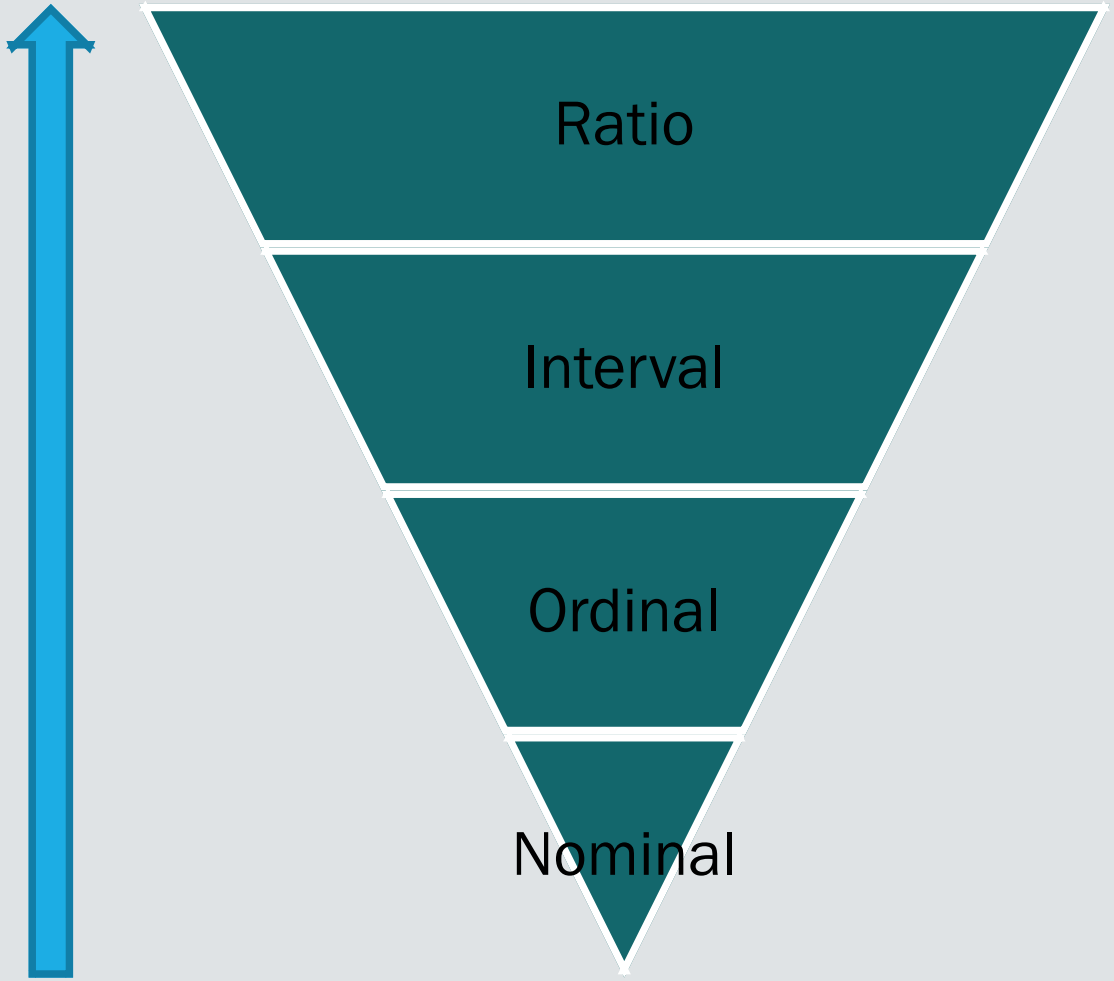
1. What type of data will your measures render?
   - *Identify the nature of each of your variables*
2. What questions do you want to address?
   - *Is there a relationship between age and level of optimism?*
   - *Are older people more optimistic than younger people?*
3. How many Independent (predictor) Variables and Dependent (outcome) Variables will you have?
4. Decide whether a parametric or non-parametric statistical technique is appropriate

Today's Topics

1. Identifying the nature of your variables
   - *Levels of measurement*
   - *Descriptive statistics*
2. Determining the type of question you want to answer
   - *Preliminary analysis (univariate statistics)*
   - *Univariate statistics*
   - *Multivariate statistics*
     - Statistical techniques for explore relationships among variables
     - Statistical techniques to compare groups
3. Deciding whether a parametric or non-parametric statistical technique is appropriate
   - *What to do if your data are not normally distributed?*

1. IDENTIFYING THE NATURE OF YOUR VARIABLES

Comparison of Measurement Levels



- At each successive measurement level, there is more information, and greater analytic flexibility.
- If you start with ratio or interval measures, you can collapse information to a lower-level measure, but the reverse is not true.
- Higher-level scales are usually (though not always) preferred.

Order of Statistical Analyses

1
- Univariate
- Descriptive Statistics

2
- Bivariate
- Relationships/associations

3+
- Multivariate
- Relationships btw 3 or more variables

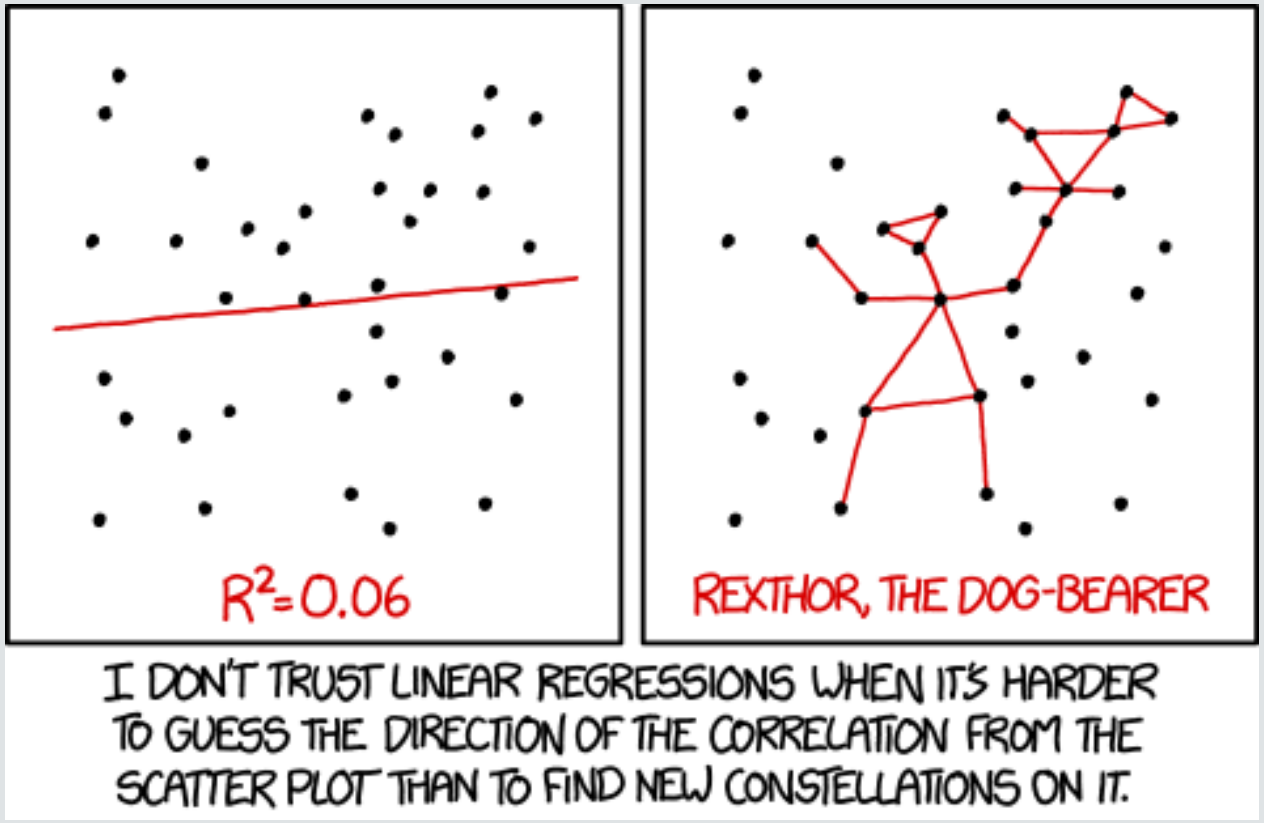
Descriptive Statistics

- Used to:
  - *Describe the characteristics of your sample*
  - *Check for any violation of the assumptions underlying the statistical techniques that you will use to address your research questions*
  - *Address specific research questions*
- Categorical variables
  - *Frequencies, Percent*
- Continuous variables
  - *Mean, Median, Standard deviation, Distribution (Skewness & Kurtosis)*

2. DETERMINING THE TYPE OF QUESTION YOU WANT TO ANSWER

Statistical Techniques to Explore Relationships Among Variables

- Correlation
- Partial correlation
- Multiple regression
- Logistic regression
- Factor analysis
- Chi-square



Correlation (Pearson product moment correlation coefficient (*r*))

- Purpose
  - *To describe the strength and direction of the linear relationship between two variables*
- Sample research question
  - *Is there a relationship between age and optimism scores? Does optimism increase with age?*
- Type of variables needed
  - *Two continuous variables*

Partial Correlation

- Purpose
  - *Similar to Pearson product-moment correlation, except that it allows you to control for an additional variable*
- Sample research question
  - *After controlling for the effects of socially desirable responding, is there still a significant relationship between optimism and life satisfaction scores?*
- Type of variables needed
  - *Three continuous variables*

Multiple Regression

- Purpose
  - *A family of techniques used to explore the relationship between one continuous DV and a number of IVs or predictors*
- Sample research question
  - *How much of the variance in life satisfaction scores can be explained by the following variables: self-esteem, optimism and perceived control?*
  - *Which of these variables is a better predictor of life satisfaction?*
- Type of variables needed
  - *One continuous DV*
  - *Two or more continuous IVs*

Logistic Regression

- Purpose
  - *A family of techniques used to explore the relationship between one or more **categorical** DV and two or more categorical and/or continuous IVs or predictors*
- Sample research question
  - *What factors predict the likelihood that respondents would report that they had a problem with their sleep?*
- Type of variables needed
  - *Two or more continuous or categorical predictor IVs*
  - *One categorical (dichotomous) DV (e.g. problem with sleep: Yes/No)*

Factor Analysis

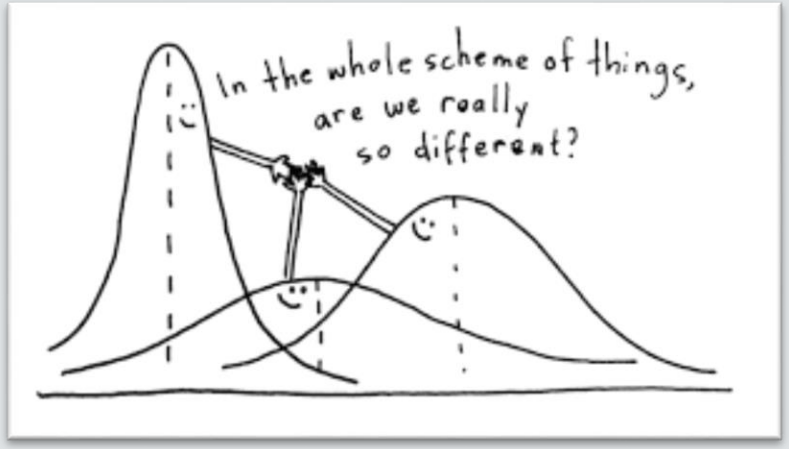
- Purpose
- Sample research question
  - *What is the underlying structure of the items that make up the Positive and Negative Affect Scale?*
  - *How many factors are involved*
- Type of variables needed
  - *Set of related continuous variables (e.g. items of the Positive and Negative Affect Scale*

Chi-square (test for independence)

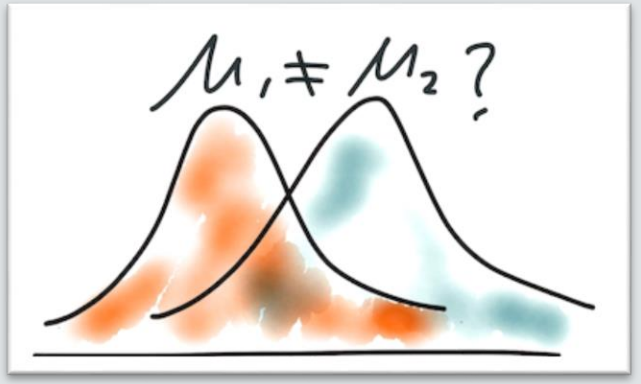
- Purpose
  - To explore the relationship between two categorical variables. Compares the observed frequencies or proportions of cases that occur in each of the categories with the values that would be expected if there was not association between the two variables being measured
- Sample research question
  - *What is the relationship between gender and dropout rates from therapy?*
- Type of variables needed
  - *Two categorical variables, with two or more categories in each (e.g. Gender- Male/Female; Drop out- yes/no)One categorical*

Statistical Techniques to Compare Groups

- T-tests
- One-way analysis of variance (ANOVA)
- Two-way between-groups ANOVA
- Multivariate analysis of variance (MANOVA)
- Analysis of covariance (ANCOVA)
- Chi-square



T-Tests



Independent-samples t-test

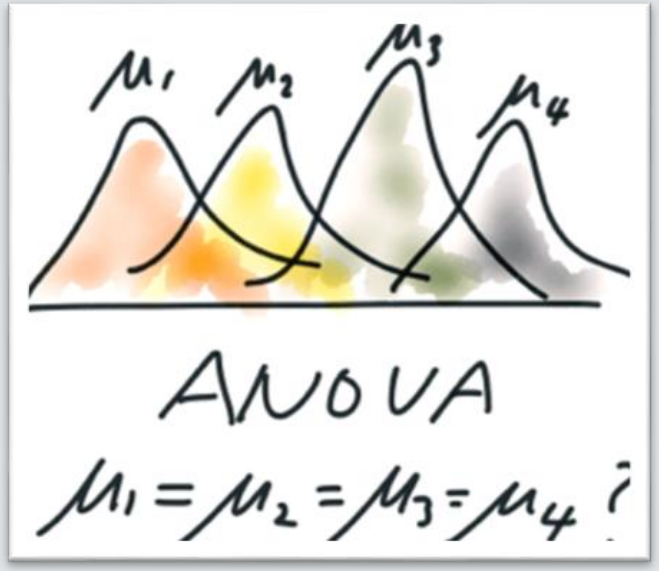
- Purpose
  - *To compare the mean scores of two different groups of people or conditions*
- Sample research question
  - *Are males more optimistic than females?*
- Type of variables needed
  - *One categorical IV with only two groups (e.g. Sex: Males/Females)*
  - *One continuous DV*

Paired-samples t-test

- Purpose
  - *To compare the mean scores for the same group of people on two different occasions, or when you have matched pairs*
- Sample research question
  - *Does ten weeks of meditation training result in a decrease in participants' level of anxiety?*
- Type of variables needed
  - *One categorical IV (e.g. Time: Time 1; Time 2)*
  - *One continuous DV measured on two different occasions or under different conditions*

One-Way Analysis of Variance (ANOVA)

- Purpose
  - *To determine if there are significant differences in the mean scores on the DV across three or more groups.*
- Sample research question
  - *Is there a difference in optimism scores for people who are under 30, between 31-49 and 50 years and over?*
- Type of variables needed
  - *One categorical IV with three or more groups (e.g. Age: under 30; between 31-49 and 50 plus)*
  - *One continuous DV*



Two-Way Between-Groups ANOVA

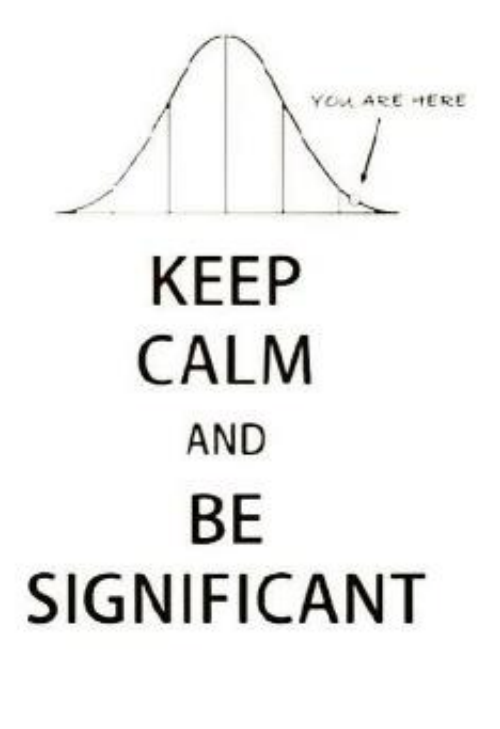
- Purpose
  - *To simultaneously test for the effect of each of your IVs on the DV and also identifies any interaction effect*
- Sample research question
  - *What is the effect of age and gender on optimism scores?*
  - *Does gender moderate the relationship between age and optimism?*
- Type of variables needed
  - *Two categorical IVs (e.g.* ***Sex****: males/females;* ***age group****: under 30; 31-49; 50+)*
  - *One continuous DV*

Multivariate Analysis of Variance (MANOVA)

- Purpose
  - *To compare two or more groups in terms of their means on a group of DVs. Tests the null hypothesis that the population means on a set of DVs do not vary across different levels of a factor or grouping variable.*
- Sample research question
  - *Do males and females differ in terms of overall well-being?*
  - *Are males healthier than females in terms of their general physical and psychological health (operationalized as anxiety, depression levels and perceived stress)?*
- Type of variables needed
  - *One categorical IV*
  - *Two or more continuous DVs that are related (e.g. negative affect, positive affect, perceived stress)*

Analysis of Covariance (ANCOVA)

- Purpose
  - *Used when you have a two-group pre-test/post-test design. The scores on the pre-test are treated as a covariate to control for pre-existing differences between the groups.*
- Sample research question
  - *Is there a significant difference in the Fear of Statistics Test scores for participants in the math skills group and the confidence-building group, while controlling for their pre-test scores on this test?*
- Type of variables needed
  - *One categorical IV with two or more levels (e.g. group 1/group 2)*
  - *One continuous DV (e.g. Fear of Statistics Test scores at Time 2)*
  - *One or more continuous covariates (e.g. Fear of Statistics Test scores at Time 1)*



Chi-square (goodness of fit)

- Purpose
  - *The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.*
- Sample research question
  - *Are males more likely to drop out of therapy than females?*
- Type of variables needed
  - *One categorical IV with two groups (e.g. Sex: Males/Females)*
  - *One continuous DV*

3. DECIDING WHETHER A PARAMETRIC OR NON-PARAMETRIC STATISTICAL TECHNIQUE IS APPROPRIATE

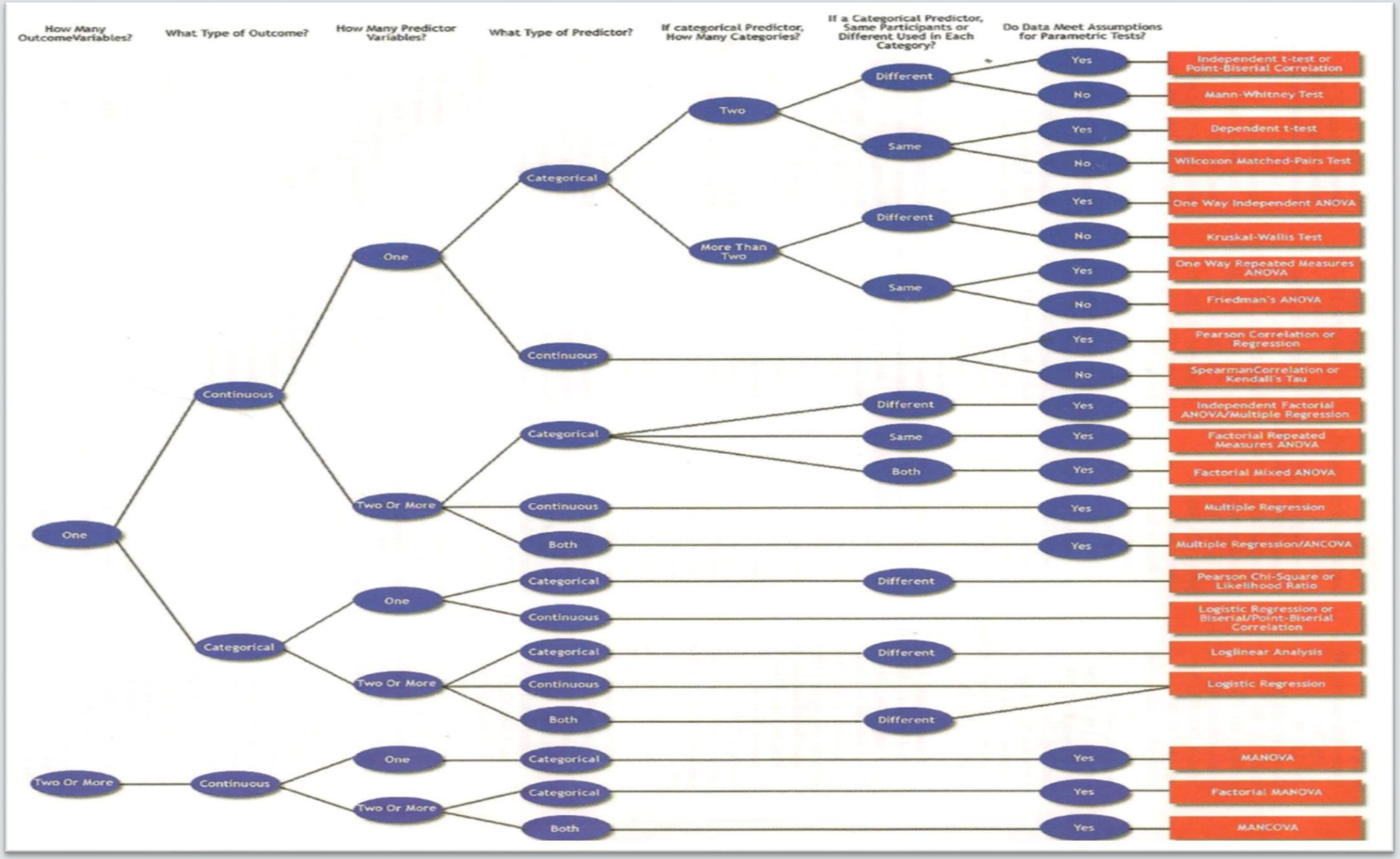
Parametric vs. Non-Parametric Statistics

- ***Parametric*** comes from ***parameter***, or characteristic of a population
- Parametric tests make assumptions about the population from which the sample has been drawn
  - *This often includes assumptions about the shape of the population distribution (e.g. normally distributed)*
- Non-parametric tests do not have such stringent requirements and do not make assumptions about the underlying population distribution

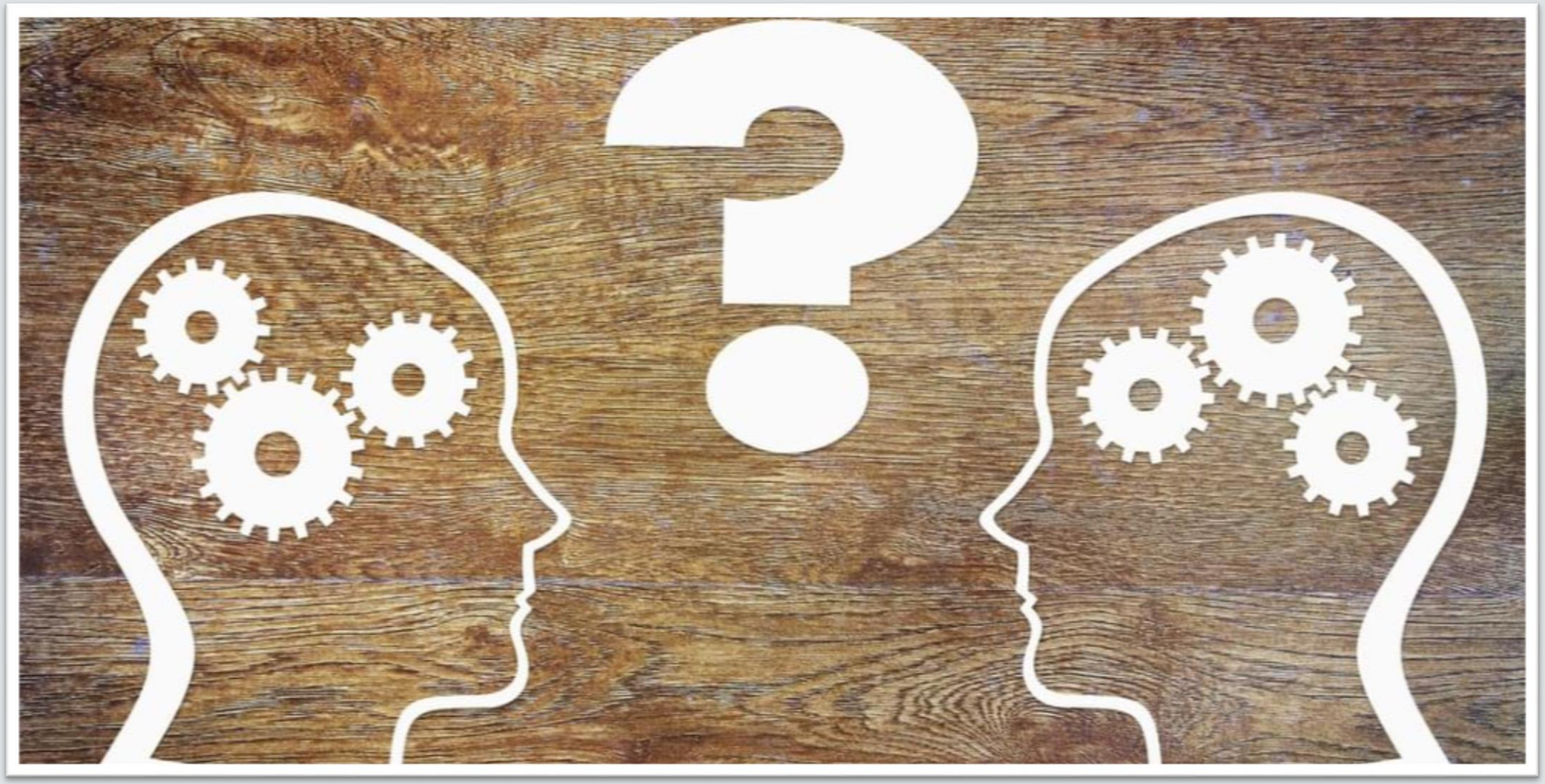
- Advantages:
  - *Ideal when you have data that are measured on nominal (categorical) and ordinal (ranked) scales*
  - *Accommodate very small samples*
  - *Data do not meet the stringent assumptions of parametric techniques*

- Disadvantages:
  - *Tend to be less sensitive than their more powerful parametric cousins*
    - May fail to detect differences between groups that actually exist

Purpose	Sample Question	Non-Parametric Statistic	Parametric Counterpart
Exploring Relationships	Is there a relationship between age and optimism scores?	Spearman's Rank Order Correlation (rho)	Pearson product-moment correlation coefficient (*r*)
	What is the relationship between gender and dropout rates from therapy?	Chi-square	
Comparing Groups	Is there a significant difference in the mean self-esteem scores for males and females?	Mann-Whitney U Test	Independent samples t-test
	Is there a change in participants' anxiety scores from Time 1 to Time 2?	Wilcoxon Signed Rank Test	Paired samples t-test
	Is there a difference in optimism scores for people who are under 35 years, 36-49 years and 50+ years?	Kruskal-Wallis Test	One-way between groups ANOVA
	Is there a change in participant's anxiety scores from Time 1, Time 2 and Time 3?	Friedman Test	Two-way repeated measures ANOVA
	Are males more likely to drop out of therapy than females?	Chi-square	



Fields, A. (2017). *Discovering Statistics Using IBM SPSS Statistics* (J. Seaman Ed. 5th ed.). Thousand Oaks, CA: Sage Publications Ltd.




thank you
danke
謝謝
ngiyabonga
شكراً جزيلاً
рахмат
teşekkür ederim
спасибо
Баярлалаа
gracias
dank je
tapadh leat
obrigado
bedankt
dziękuję
merci
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
arigatô
dakujem
grazie
takk
tack
mauruuru
hvala
sagolun
faafetai lava
kiitos
dankie
vinaka
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
дякую
djiere dieuf
tau
akun
dankon
aciû
köszönöm
gràcie
bayarlalaa
nanni
nandri
dhanyavad
enkosi
sobodi
dèkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
rahmet
tanemirt
ありがとう
chnorakaloutioun
gratias ago
grâcies
sulpáy
trugarez
мерси
shukriya
merci
dhanyavadagalu
diolch
xièxie
감사합니다
ευχαριστώ
তোমাকে ধন্যবাদ

References

Fields, A. (2017). *Discovering Statistics Using IBM SPSS Statistics* (J. Seaman Ed. 5th ed.). Thousand Oaks, CA: Sage Publications Ltd.

Pallant, J. (2015). *SPSS Survival Manual: A step by step guide to data analysis using SPSS* (6th ed.). New York, NY: McGraw-Hill.

Polit, D. F. (2010). *Statistics and Data Analysis for Nursing Research 2nd Edition*. Upper Saddle River, NJ: Pearson.