# Statistics Literacy

By Dr Elif Bilgic

June 6 2023; HSED Residency Week

Assistant Professor & Education Scientist

e-mail: bilgice@mcmaster.ca

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Educational Experience

- **BSc**, Anatomy and Cell Biology, McGill University
- **PhD**, Surgical Education

The Steinberg-Bernstein Centre for MINIMALLY INVASIVE SURGERY at McGill University

- **Postdoctoral fellow**

McGill | STEINBERG CENTRE FOR SIMULATION AND INTERACTIVE LEARNING / CENTRE DE SIMULATION ET D'APPRENTISSAGE INTERACTIF STEINBERG

- **Postdoctoral fellow**

SAILS LAB

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

McMaster University | Pediatrics

MERIT MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Session Goals

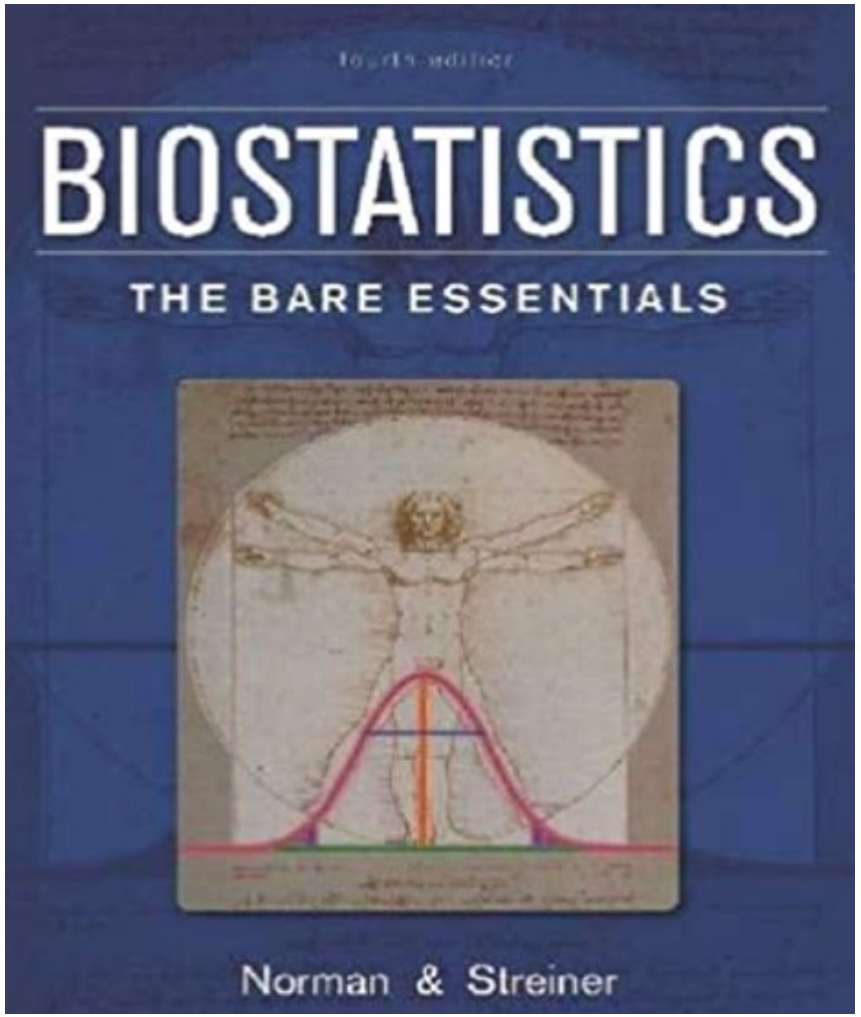
- Using an example, describe
  - Study designs
  - Variable types
  - Descriptive and inferential analysis

- Practical approach to choosing a statistical test

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Main References*




**A Clinician-Educator's Roadmap to Choosing and Interpreting Statistical Tests**

*Donna M. Windish, MD, MPH,[1] Marie Diener-West, PhD[2]*

[1]Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA; [2]Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

MedEdPORTAL® | The AAMC Journal of Teaching and Learning Resources

Original Publication

**A Guide to Basic Statistics for Educational Research**

Donna M. Windish, MD, MPH*

*Most figures/information from these 3 references

# Why does statistical literacy matter?

- Help design your research
- Help write your research
- Help read the literature better

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Approach to choosing a statistical test

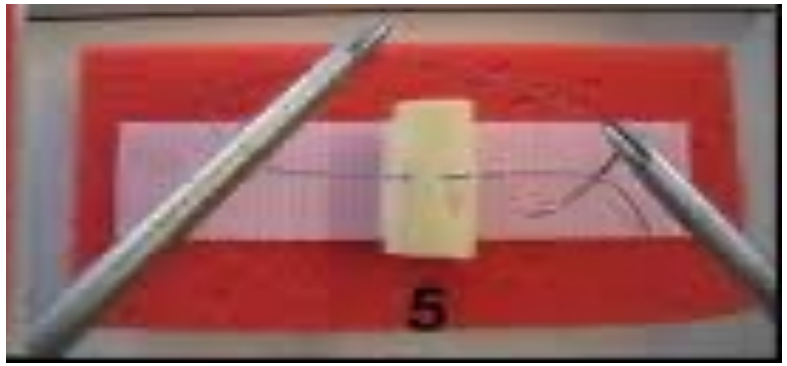
1. What is/are the Research Question(s) we are trying to answer
2. What is the Research Design
3. What is/are the Outcome Variable(s) (type)
4. How is/are the outcome variable(s) Distributed?

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Example of an educational intervention

- 1 month curriculum to improve 1st year residents'
  - Suturing skills (technical)
  - Confidence
  - Professionalism
  - Pass rate

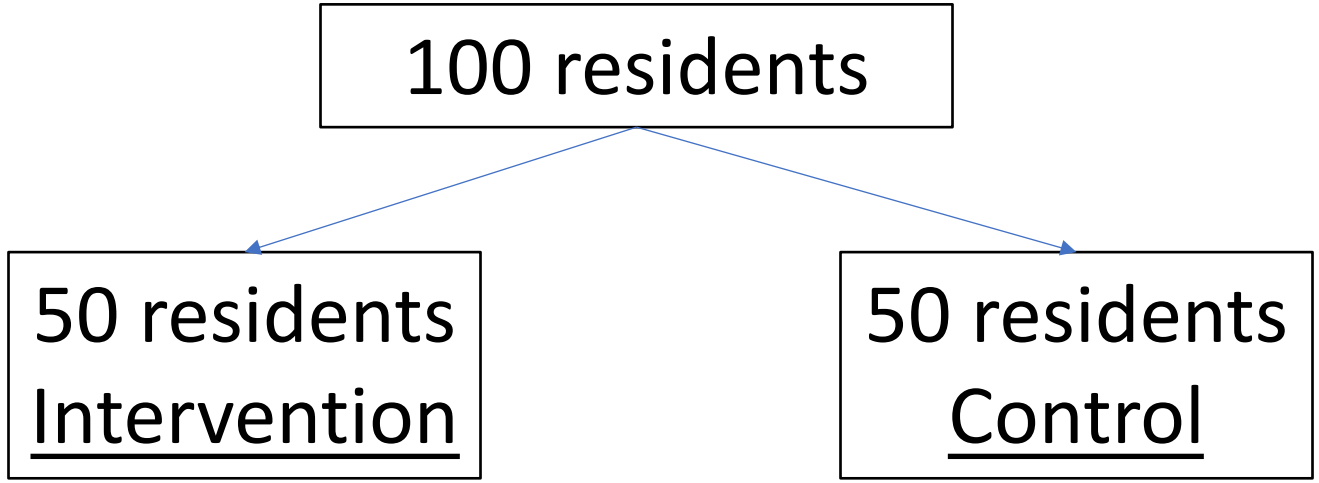



McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Example of an educational intervention

- Randomized Controlled Trial




McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# 1) Research Question (RQ)

Do residents in control and intervention groups differ in the mean number of suturing maneuvers that they perform correctly?

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# 2) Research Design: Observational or Experimental

Observational: 'Examine groups at one or more points in time without an intervention' (e.g., case-control, cross-sectional (e.g., survey), and cohort studies (follow residents over time))

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# 2) Research Design: Observational or Experimental

Experimental: ‘Allocate interventions to one or more groups and make comparisons across groups to assess differences in outcomes’

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# 2) Research Design: Subjects (data) Paired or Unpaired

Why is it important: If paired, results obtained for each resident during different measurements are more likely to be highly correlated than the results of 2 randomly selected residents

# 2) Research Design: Subjects (data) **Paired**

- Pre-post design, each resident is assessed using the same tools, at <u>different points in time</u>
- Each participant serves as their own comparison

- Assess differences in resident skills before and after intervention/control

Same residents in 1 group

Baseline skill

Intervention

Post-intervention skill

# 2) Research Design: Subjects (data) Unpaired

- Measurements from independent/unrelated groups

Post-intervention skill

Group 1

VERSUS

Post-intervention skill

Group 2

# 2) Research Design: Observational or Experimental? Back to our Example

- Randomized Controlled Trial

- Is our example study Observational or Experimental?

# 2) Research Design: Observational or Experimental? Back to our Example

- Randomized Controlled Trial

- EXPERIMENTAL

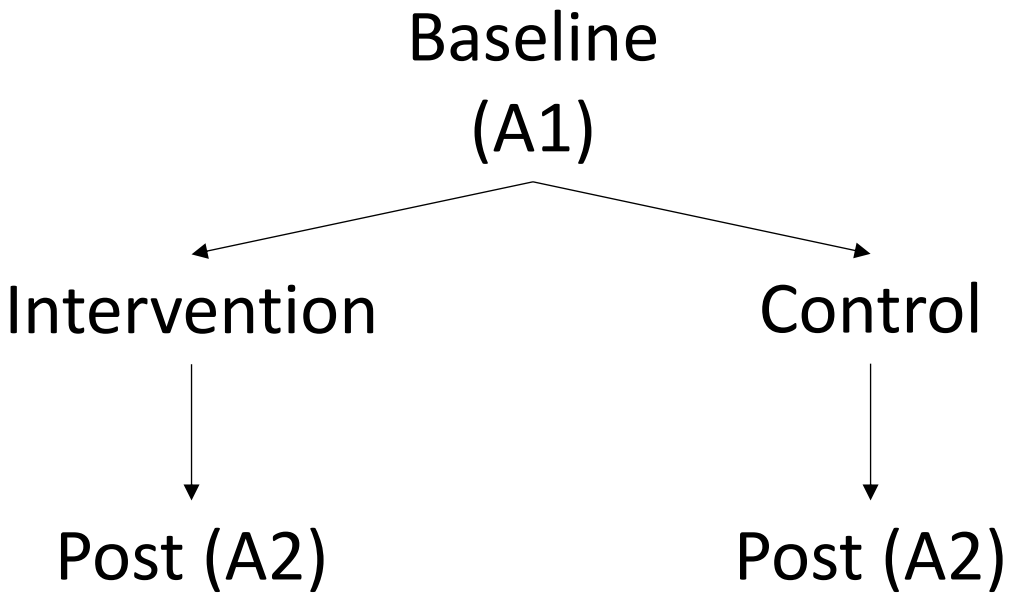
# 2) Research Design: Paired or Unpaired? Back to our Example

- Randomized Controlled Trial

- Does our example study include Paired or Unpaired data, or both?

# 2) Research Design: Paired or Unpaired? Back to our Example

- Randomized Controlled Trial
- Has both paired and unpaired data
  - Depends on each research question



# 2) Research Design: Paired or Unpaired? Back to our Example

- Randomized Controlled Trial
- Has both paired and unpaired data
  - Depends on each research question

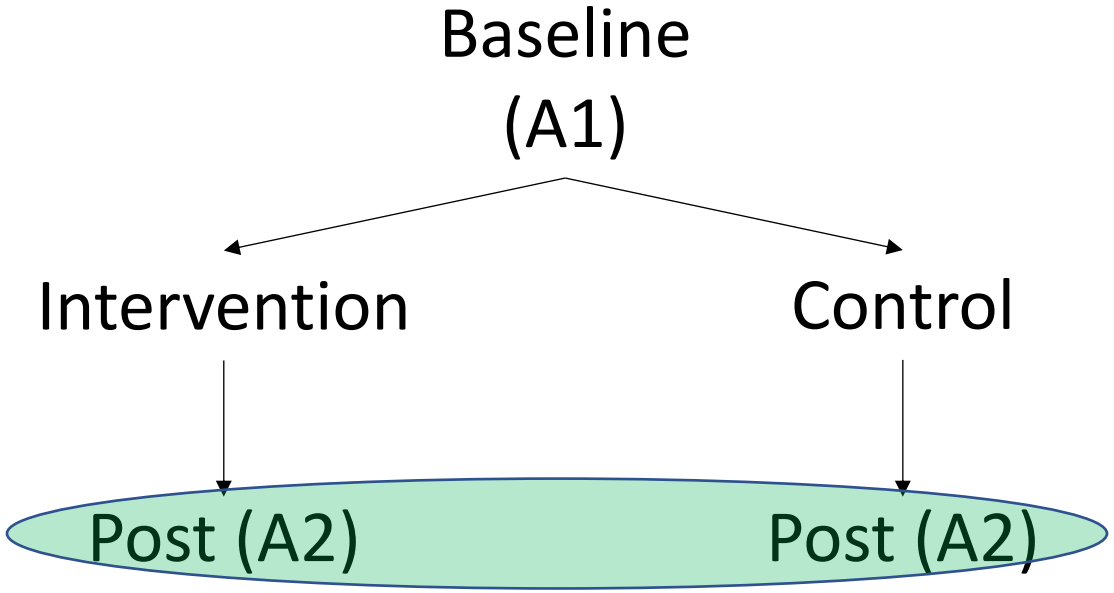
**RQ1:** Do residents in control and intervention groups differ in the mean number of suturing maneuvers that they perform correctly?

Paired or Unpaired?

# 2) Research Design: Paired or Unpaired? Back to our Example

Do residents in control and intervention groups differ in the mean number of suturing maneuvers that they perform correctly?

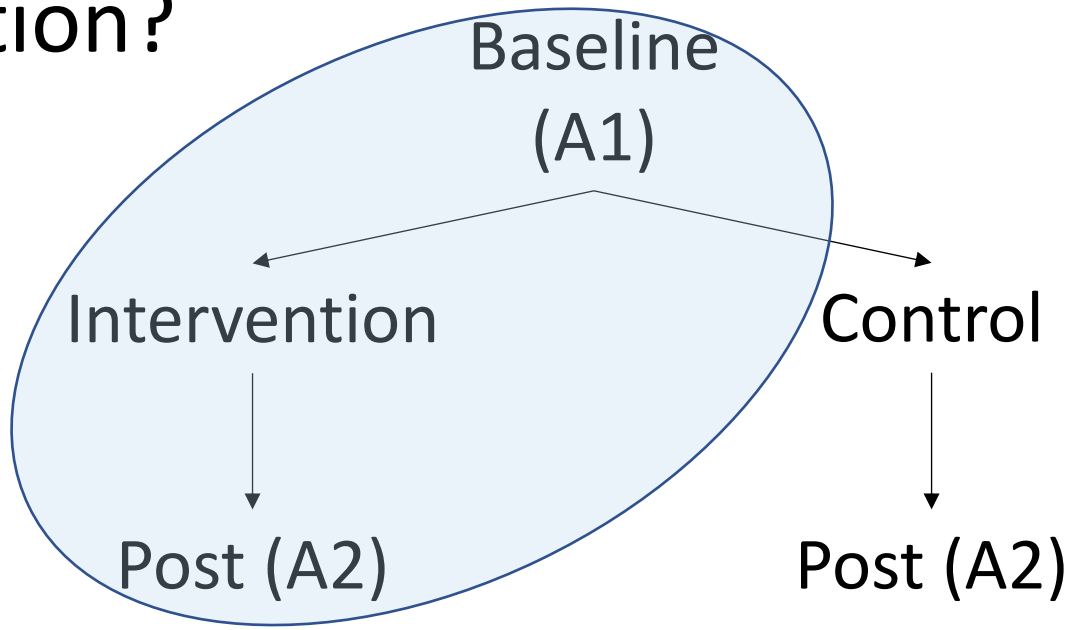
- Unpaired groups just residents in a program, so no pairing
- Are they different at the end of the curriculum



# 2) Research Design: Paired or Unpaired? Back to our Example

Paired data

- **RQ1 modified:** Do performance of residents in the intervention group change from baseline to post-intervention?



# 3) Outcome Variables

**Variable:** What is being observed or measured

# 3) Outcome Variables

**Dependent variable:** The outcome of interest, should change in response to an intervention

**Independent/Covariate variable:** What we are manipulating, or the intervention

E.g., Test whether changes in room temperature have an effect on knowledge test scores.

IV

DV

# 3) Outcome Variables

**Quantitative Variables** (numerical)

- **Continuous:** 'Have no gaps in the values (e.g., age)' – may take any value within a defined range
- **Discrete:** 'Have gaps (e.g., the number of study participants, number of admissions to the hospital, number of missing teeth)'

# 3) Outcome Variables

**Qualitative Variables** (describe certain attributes: categorical)

- **Dichotomous:** Categorical variable with 2 possible values (e.g., pass vs fail)

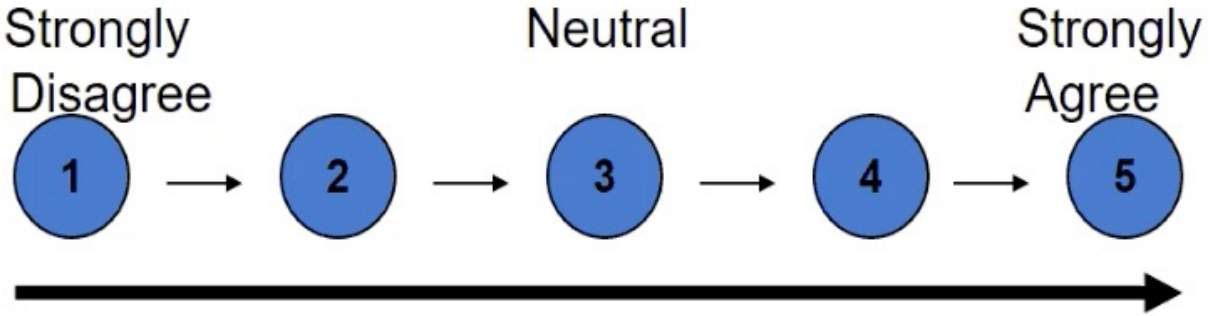
Pass  Fail

- **Nominal:** 'Descriptive and cannot be ordered; classifies data into categories (e.g., college major, school location)'
  - Challenging statistics
  - Try to make it dichotomous

Ontario  Quebec

Nova Scotia  Manitoba

# 3) Outcome Variables




**Qualitative Variables** (describe certain attributes: categorical)

- **Ordinal:** 'Have an implicit ranking associated with them (e.g., Likert scale of 1-5)', and differences between categories cannot be considered as equal
  - *If have 5-7 or more potential points in the scale, can consider ordinal as continuous
  - Stats for continuous is easier to understand and perform
  - You can dichotomize ordinal scales (e.g., who received a score from 1-2, and who from 3-5)
  - **Interval variable:** If distance between values is constant/equal

# 3) Outcome Variables
# Back to Our Example

- The mean number of suturing maneuvers residents perform correctly

- What type is our outcome variable?
  - Continuous, discrete, dichotomous, nominal, ordinal

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# 3) Outcome Variables
# Back to Our Example

- The mean number of suturing maneuvers residents perform correctly

- CONTINUOUS

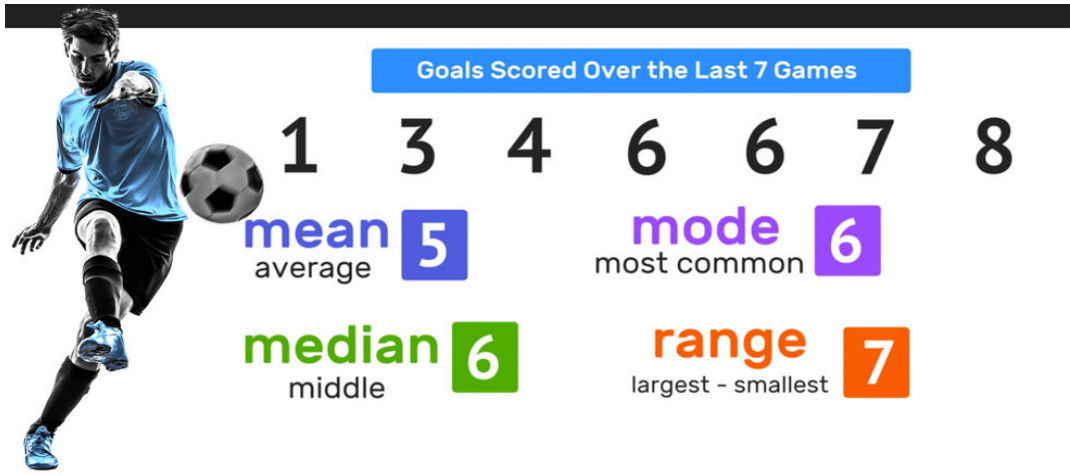
# 4) Distribution of Outcome Variable Exploratory Data Analysis (Descriptive Statistics)

- **'Concerned with the presentation, organization, and summarization of data'**
  - Are there any mistakes in data entry
  - Outliers?
  - Distribution of data
    - E.g., bell-shaped
  - Choose the summary measure
  - Choose parametric vs non-parametric stats, based on distribution

# Descriptive Statistics with Summary Measures

**Measures of Central Tendency (typical value for the data)**

- **Mean:** Sum of the value for each subject ÷ total sample size
- **Median:** Value where half of data points fall above and half below
- **Mode:** Most frequently occurring category




McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Descriptive Statistics with Summary Measures

**Measures of Dispersion (how closely the data cluster around the measure of central tendency)**

- **Range:** The difference between the highest and lowest values
- **Interquartile range:** Difference between lower and upper quartile, and comprises of middle 50% of data (from the median)
- **Standard deviation:** How dispersed the data is in relation to the mean

# Descriptive Statistics with Summary Measures

## Skewness and Kurtosis

- **Skew:** Symmetry of the curve
- **Kurtosis:** How flat or peaked the curve is

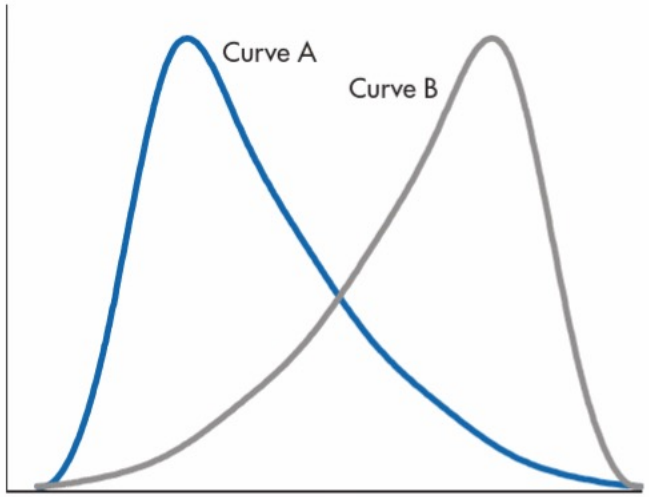



**FIGURE 3–4**
Two curves, one with positive and one with negative *skew*.

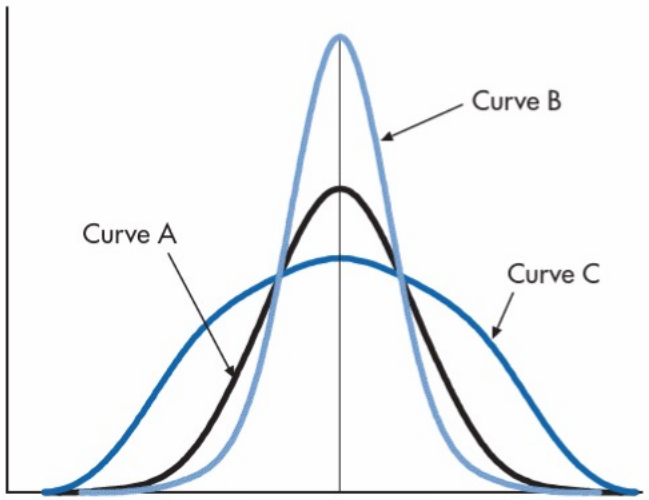



**FIGURE 3–5**
Three distributions differing in terms of *kurtosis*.

# 4) Distribution of Outcome Variable
# Exploratory Data Analysis (Descriptive Statistics)

- Plotting out your data

Eyeball test

- **Histogram** - No spaces between bars
- **Bar chart** - There are spaces between bars

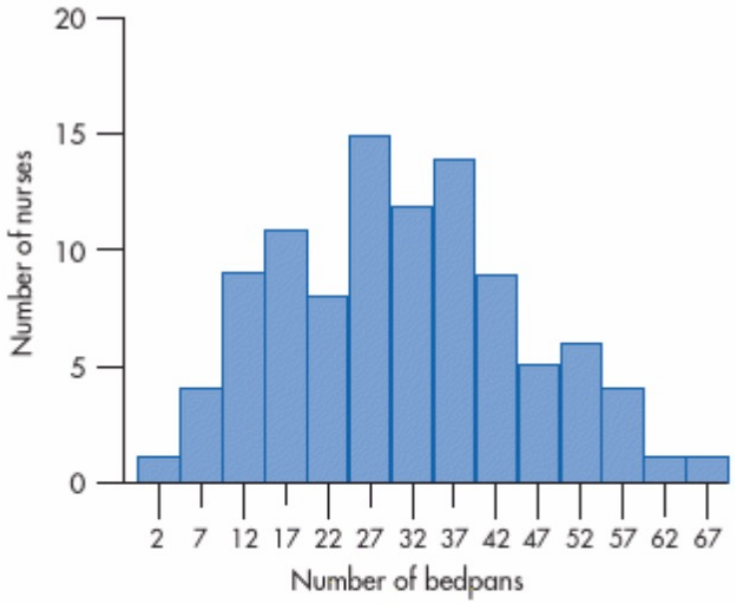



**FIGURE 2–5**
Histogram showing the number of bedpans emptied during the past month by each of 100 nursing students.

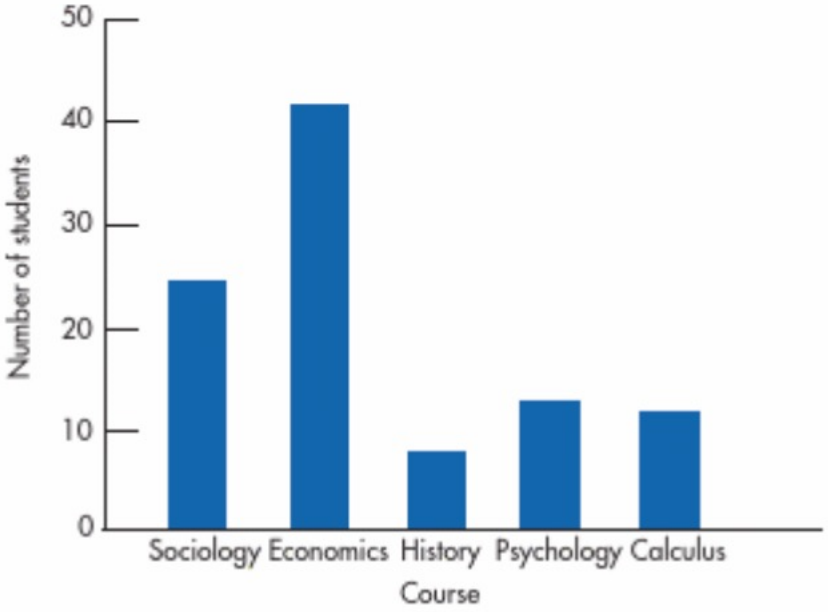


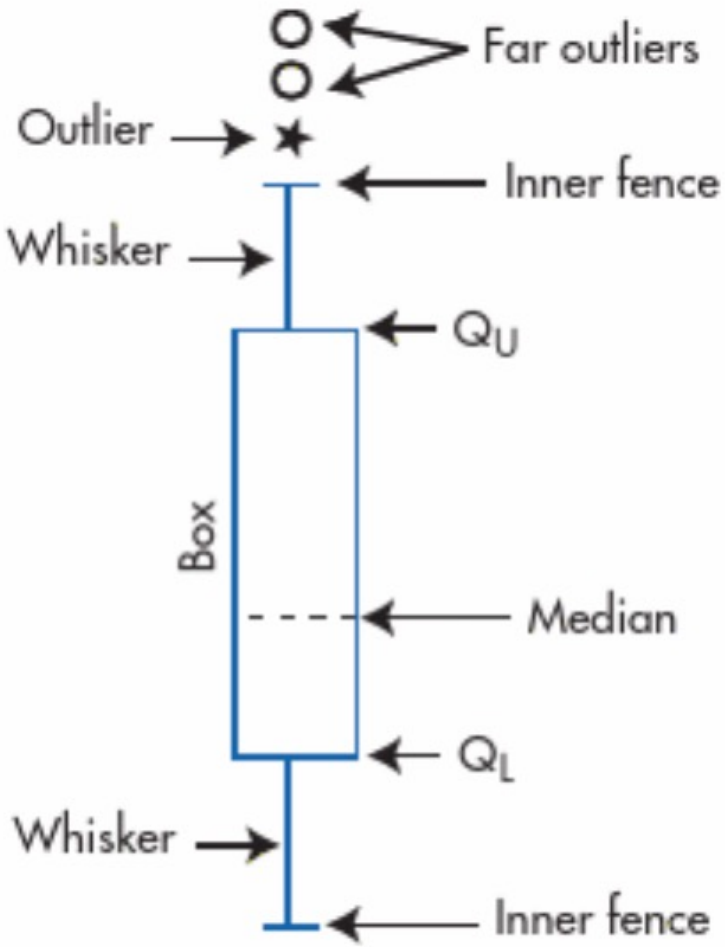

**FIGURE 2–1**
Bar chart of the five least popular courses

# Descriptive Statistics with Summary Measures

- **Box plot** – lower part 25th percentile, top 75th percentile. If normal distribution, line should be in the middle (50th percentile)

**FIGURE 3–7**

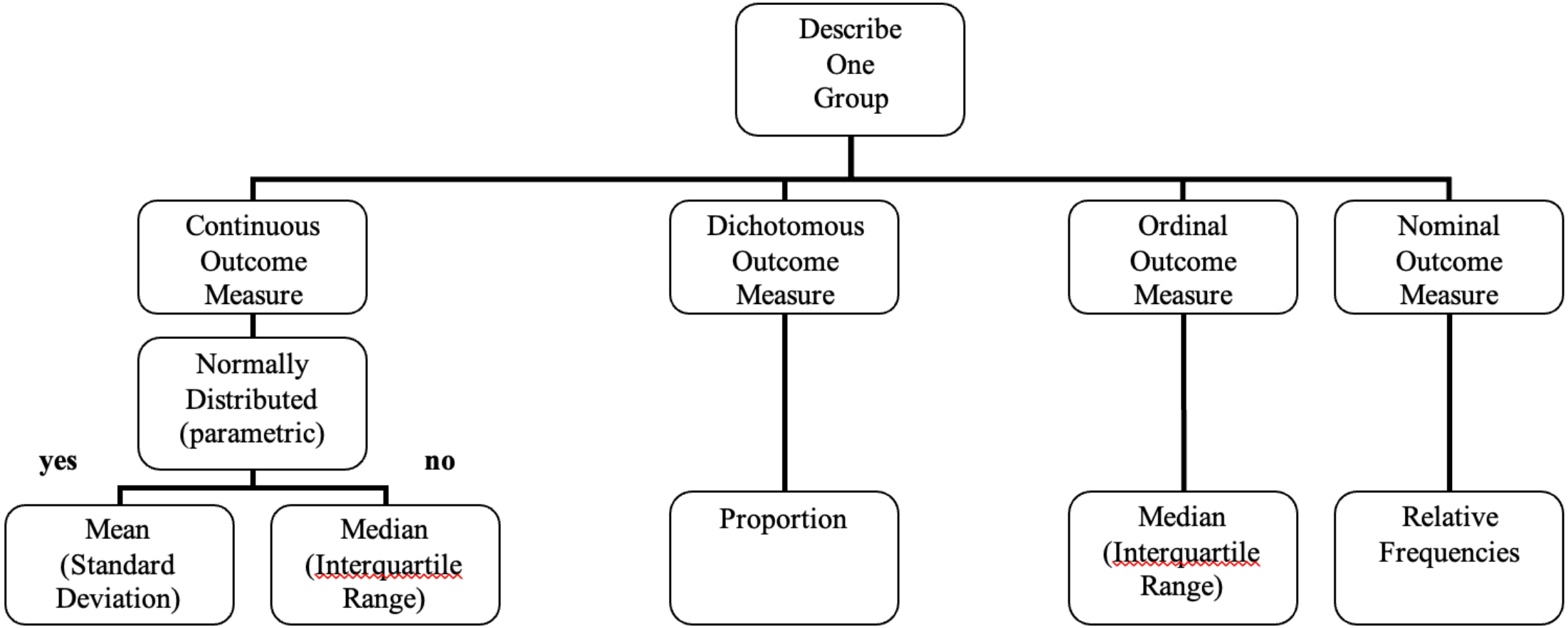
Anatomy of a box plot.



# 4) Distribution of Outcome Variable
# Back to Our Example

- Lets assume that the distribution of the mean number of suturing maneuvers for each group appears to be <u>normally distributed</u>

# Descriptive Statistics with Summary Measures




McMaster University | Pediatrics

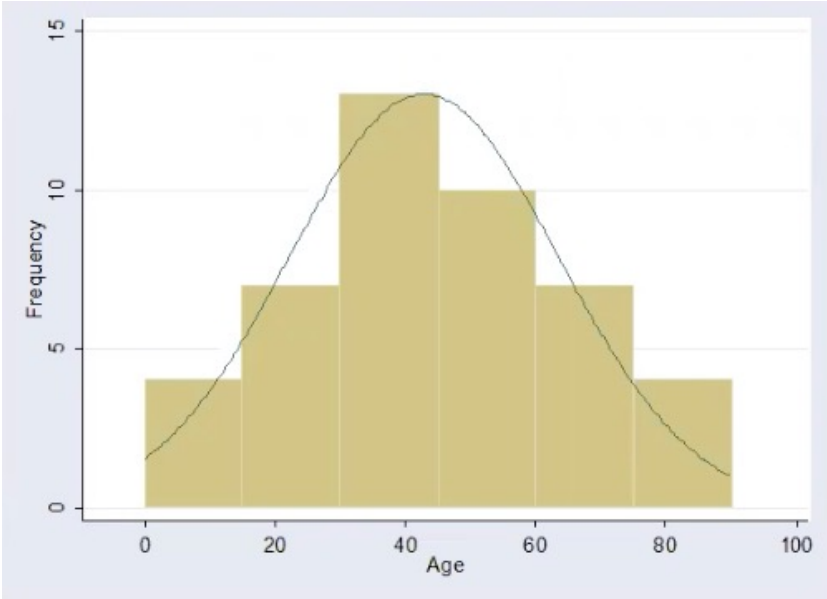
MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Confirmatory Data Analysis (Inferential Statistics)

- **'Allow us to generalize from our sample of data to a larger group of subjects'**
- **'Used to determine the likelihood that a conclusion based on the analysis of data from the sample is true'**
  - We are Testing a Hypothesis
  - Assess strength of evidence
  - Predict outcomes
  - Make conclusions about a specific population
  - Basing all of this through our sample data

# Types of Statistical Tests: Parametric

- Used when evaluating <u>continuous variables</u> with <u>normal distribution</u>
  - *sometimes ordinal variables as well



# Types of Statistical Tests: Nonparametric

- Used when sample size is small, or data is not normally distributed (skewed)
- Used when evaluating continuous or ordinal variables
  - *Ordinal variables are usually analyzed using nonparametric tests

- **More conservative, but important to use when parametric considerations do not hold

# Hypothesis Testing – Null Hypothesis

- Need statement of null hypothesis
  - Statement of no effect or association

- Back to our Example
  - Residents in both groups do not differ in mean number of suturing maneuvers performed correctly after the curriculum

# Hypothesis Testing – Null Hypothesis

- Two types of errors can occur when making conclusions regarding the null hypothesis: Type I error and Type II error.


Rejecting the null hypothesis when the null hypothesis is true

Accepting the null hypothesis when it is false

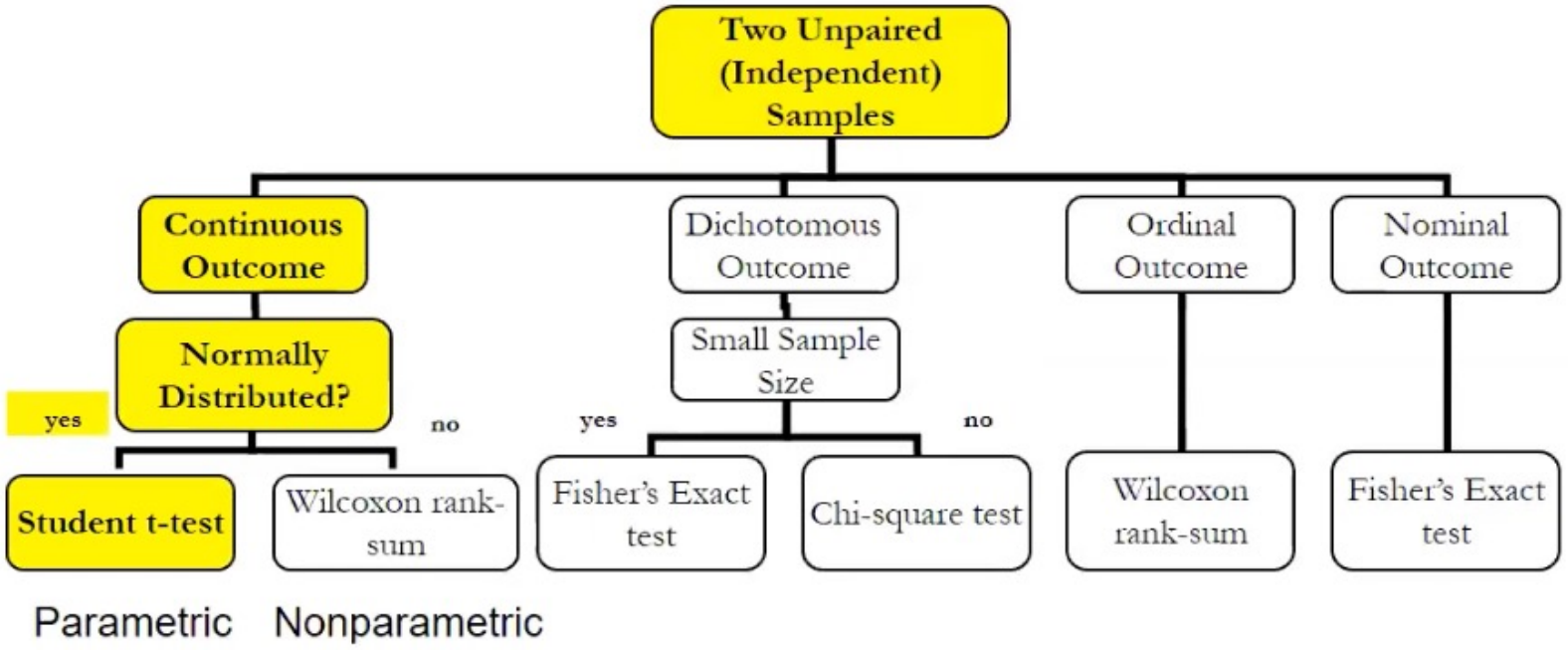
# Hypothesis Testing- Level of Statistical Significance

- **P-value:** Probability of an observed result assuming that the null hypothesis is true
  - Decide if p-value is statistically, and hopefully educationally, significant (p-value <0.05)

- If no statistical significance observed; null hypothesis is true (no real difference exists) or sample size not large enough (insufficient power)

# Back to Our Example

- RQ1: Do residents in control and intervention groups differ in the mean number of suturing maneuvers that they perform correctly?
- We have 2 groups, unpaired data, outcome variable is continuous, normally distributed



# Back to Our Example- Results

- Mean (SD) of intervention: 4(0.9) and control 3(0.9)
- Student t-test: $p<0.0001$

- Students in the intervention group performed statistically significantly more suturing skills vs controls

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Back to Our Example - RQ2

- Confidence

- Is there a difference in resident's confidence level in suturing before and after the curriculum?

McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Back to Our Example - RQ2

- Only looking at the intervention group's baseline (pre) vs post-curriculum levels

- Paired vs Unpaired?

# Back to Our Example - RQ2

- Only looking at the intervention group's baseline (pre) vs post-curriculum levels => PAIRED group

# Back to Our Example - RQ2

- Outcome measured using a 4-point Likert scale

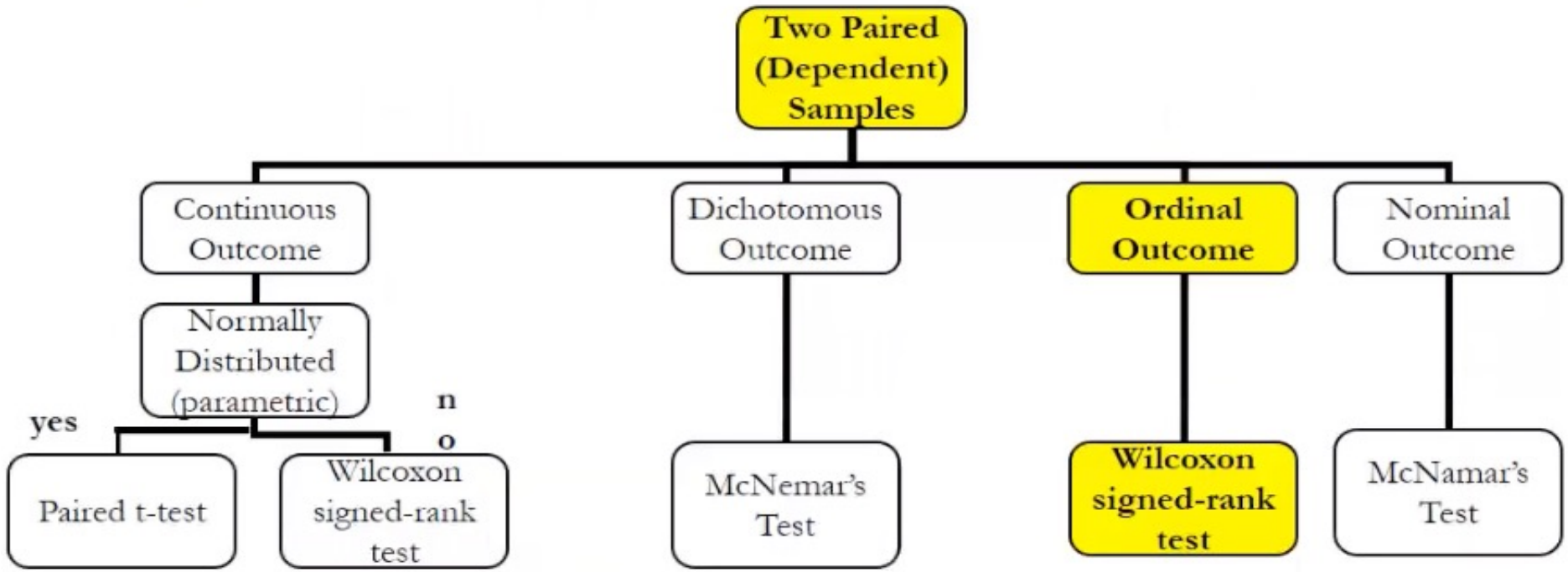
- What type is our outcome variable?
  - Continuous, discrete, dichotomous, nominal, ordinal

# Back to Our Example - RQ2

- Outcome measured using a 4-point Likert scale => ORDINAL variable

# Back to Our Example - RQ2

- Paired group
- Ordinal variable



# Back to Our Example - RQ2 Results

- Since nonparametric, used median (IQR)
- $p<0.05$, the intervention was statistically significantly successful at improving confidence

# Back to Our Example - RQ3

- Professionalism skills
- Do residents in the 2 groups differ in their overall professionalism score post curriculum?
- *went back to looking at the 2 groups, rather than just 1
- Paired vs Unpaired

# Back to Our Example - RQ3

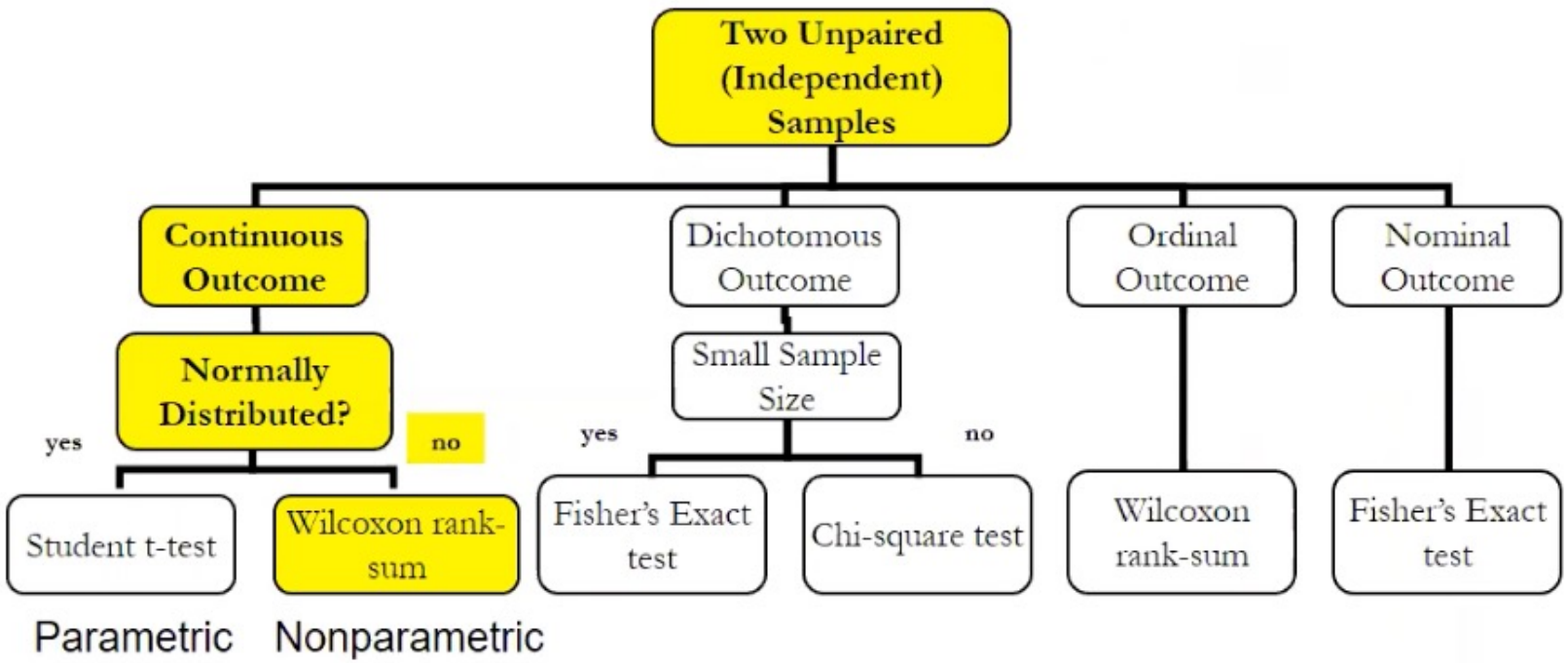
- Professionalism score is a SUM of 20-item assessment tool, each item rated on a 5-point Likert scale (1 needs improvement to 5 excellent)
  - Lowest score is 20 and highest score is 100

- Using small ordinal scales, but totaling them to get a bigger number (20-100). Huge range, so can consider as a continuous outcome

# Back to Our Example - RQ3

- Unpaired
- Continuous
- Distribution => skewed

=> Nonparametric



# Back to Our Example - RQ3 - Results

- Median (IQR), P>0.06
- Not statistically significant

- The curriculum did not statistically significantly improve professionalism

# Back to Our Example – RQ4

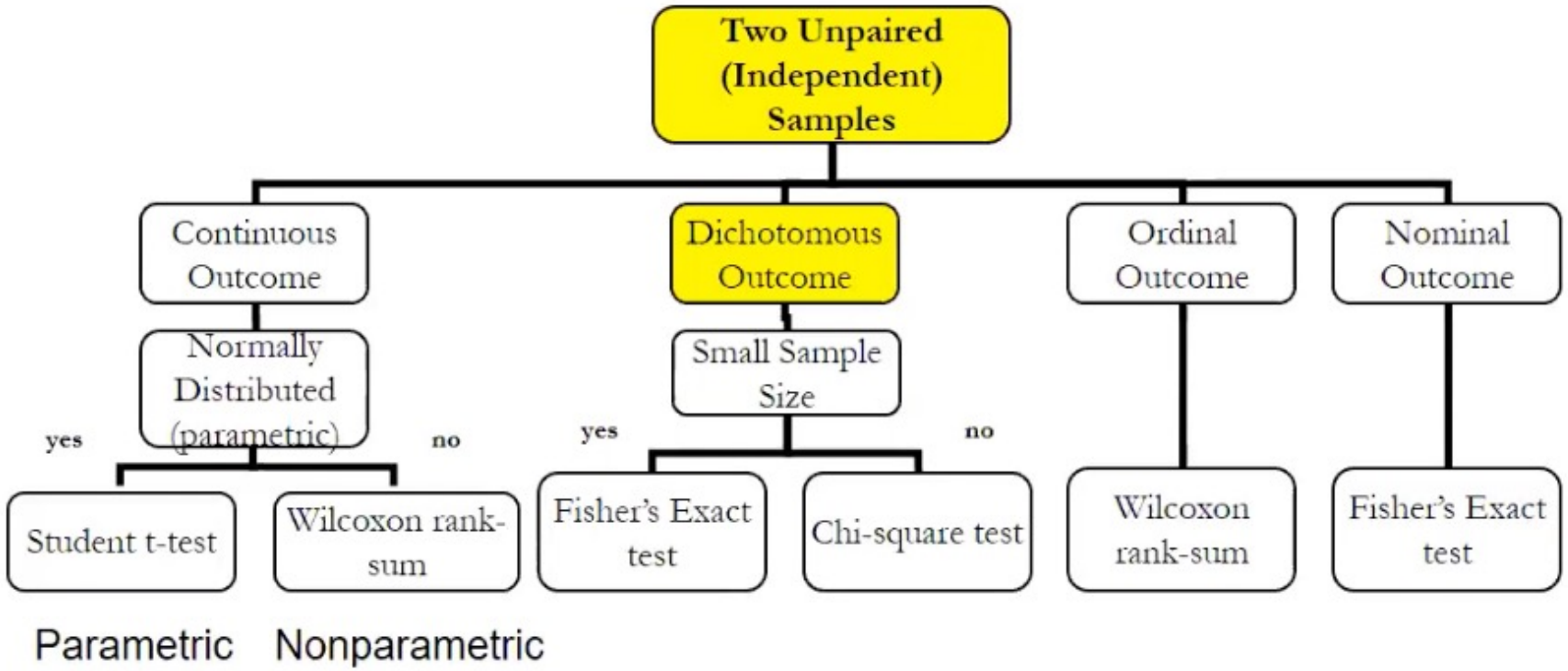
- Pass rate
- Do residents in the 2 groups differ in their overall pass rates post curriculum?
- RCT, looking at the 2 groups, rather than just 1
- Paired vs Unpaired

# Back to Our Example – RQ4

- Pass / Fail is a dichotomous outcome

# Back to Our Example – RQ4

- Need to determine sample size, will determine the test to use
  - Differences in these 2 tests: Number of observations



# Back to Our Example – RQ4

- **Chi-square test (Pearson)**— 'statistical test used to compare two unpaired samples where the outcome is dichotomous or nominal and the sample size is large (>30)'.
  - Looking at comparison between proportions

| | Passed Exam | Failed Exam |
|---|---|---|
| New Curriculum | 58 | 2 |
| Old Curriculum | 50 | 10 |

2 by 2 table

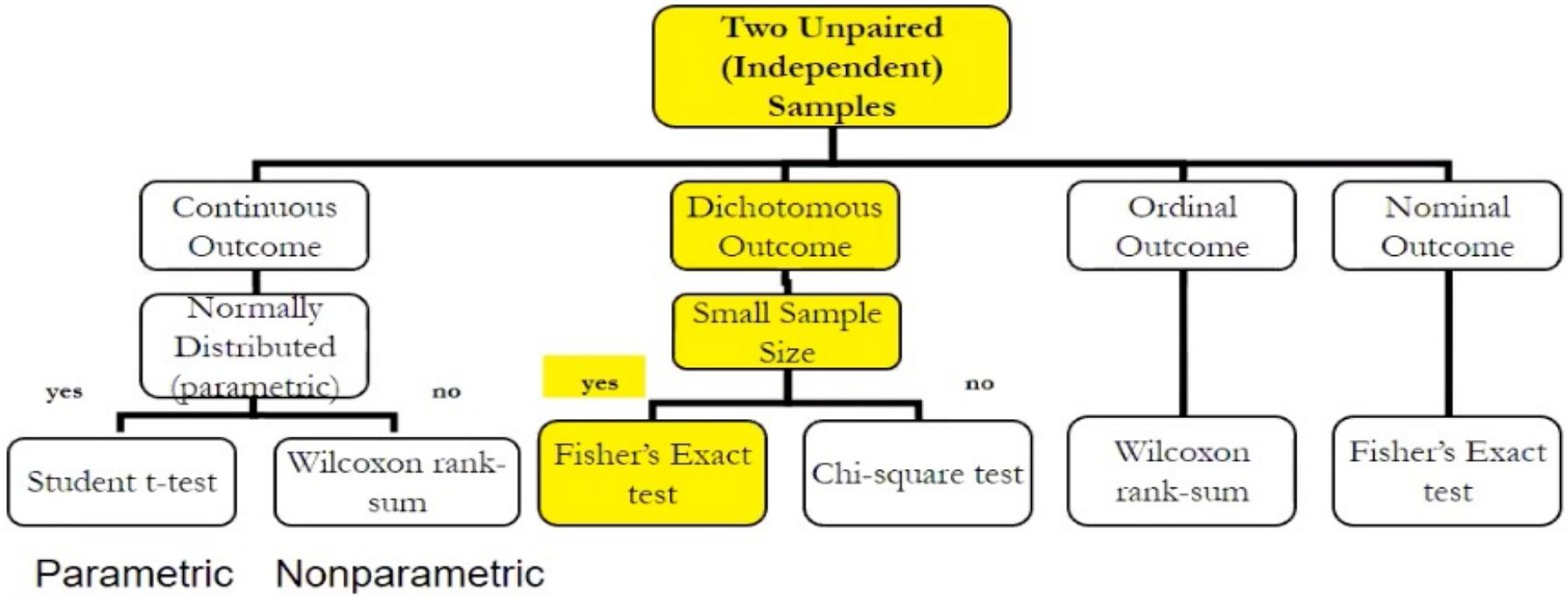
# Back to Our Example – RQ4

- **Fisher Exact Test**—'statistical test used to compare two unpaired samples where the outcome is dichotomous or nominal and the sample size is small (observations are rare)'.
  - Looking at comparison between proportions
  - Less than 5 in at least 1 cell

| | Passed Exam | Failed Exam |
|---|---|---|
| New Curriculum | 58 | 2 |
| Old Curriculum | 50 | 10 |

# Back to Our Example – RQ4

- Results: Using number and percentage;
  - Intervention led to a statistically significant pass rate than control
  - e.g., for intervention 58 (97%) vs control 50 (83%), p<0.01



# GROUP ACTIVITY

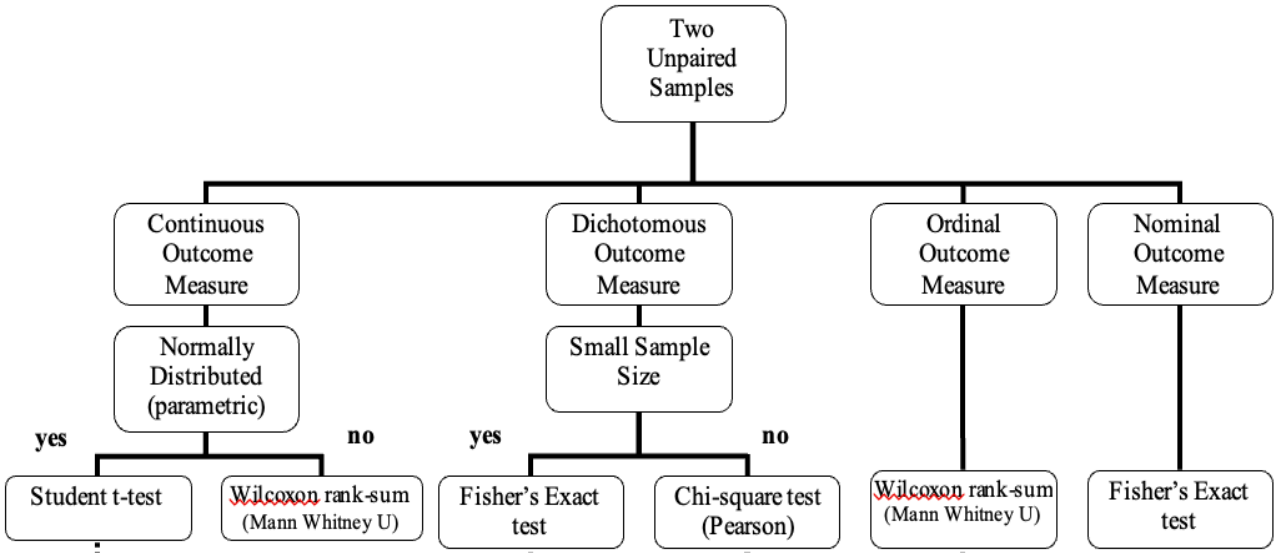
- **Case**: Communication (giving bad news) curriculum for pediatric residents

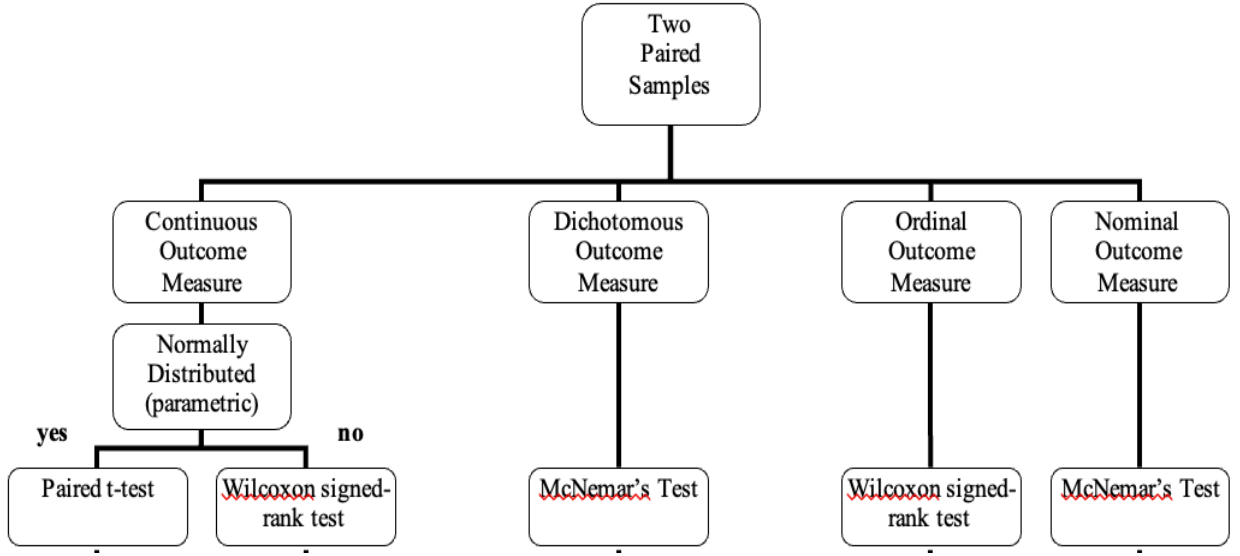
- **Study Aim** – An intervention for residents aimed at improving communication skills

- **Study Design**:  Intervention vs Control groups
  - 50 residents in each group
  - 5-item assessment tool, using 5-point Likert scale, Summation score of the 5 items
- **Null Hypothesis:** The residents in the intervention group will not improve their overall communication scores before vs after the intervention

# Please Answer

- What is the Research Question we are trying to answer
- What is the Research Design (obs/exp? and paired/unpaired?)
- What is the Outcome Variable (type)
- How is the outcome variable Distributed? [Assume not-normal]
- What is the most appropriate test?



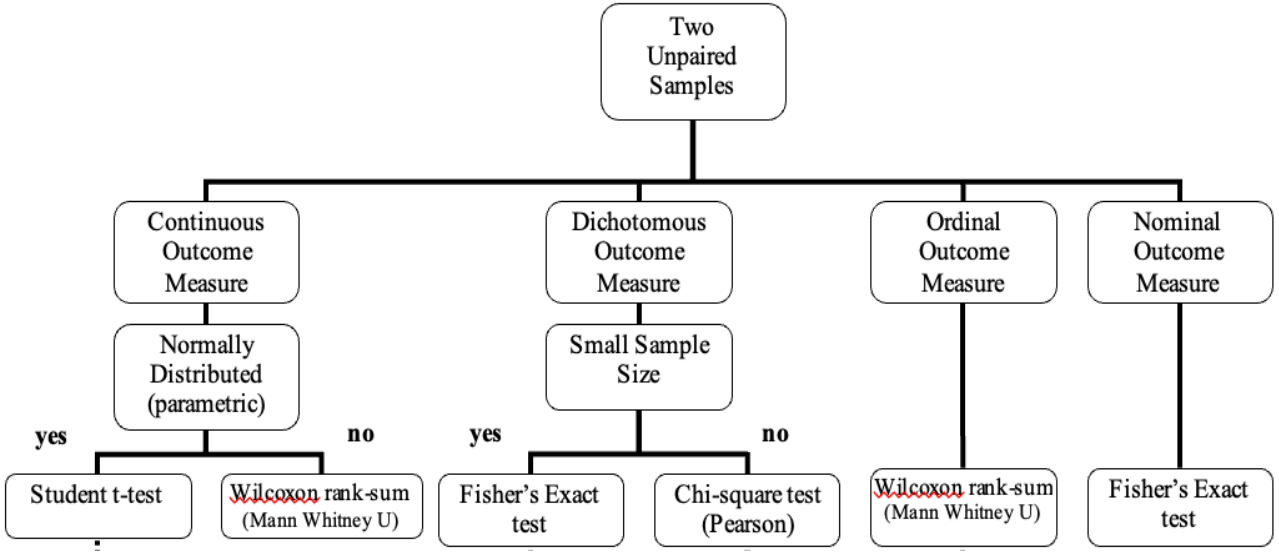
- **Case**: Communication (giving bad news) curriculum for pediatric residents

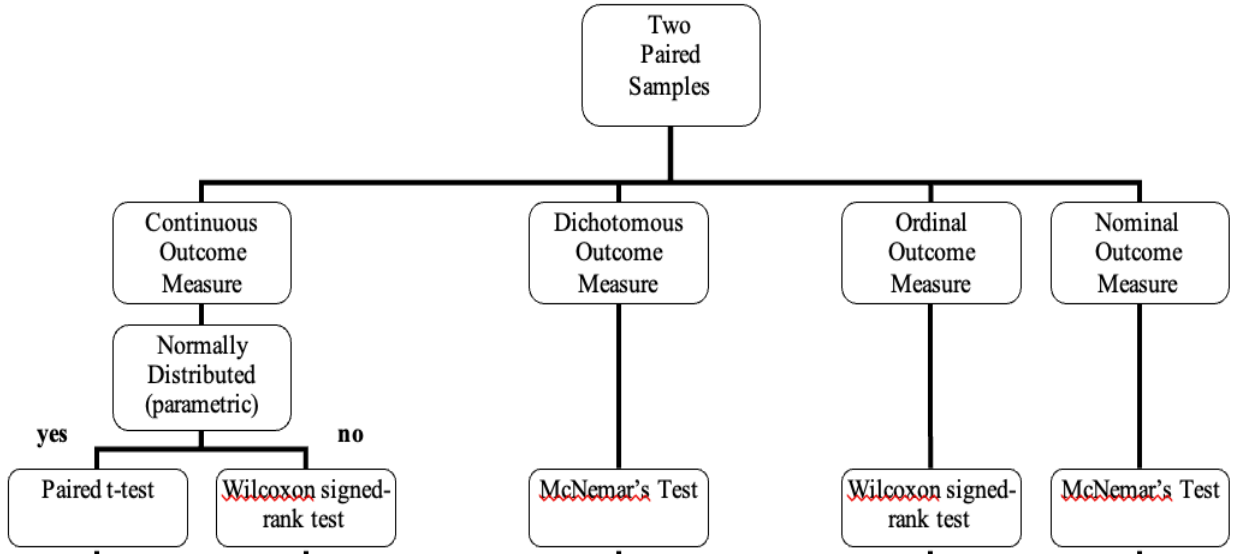
- **Study Aim** – An intervention for residents aimed at improving communication skills

- **Study Design**: Intervention vs Control groups
  - 50 residents in each group
  - 5-item assessment tool, using 5-point Likert scale, Summation score of the 5 items

- **Null Hypothesis:** The residents in the intervention group will not improve their overall communication scores before vs after the intervention

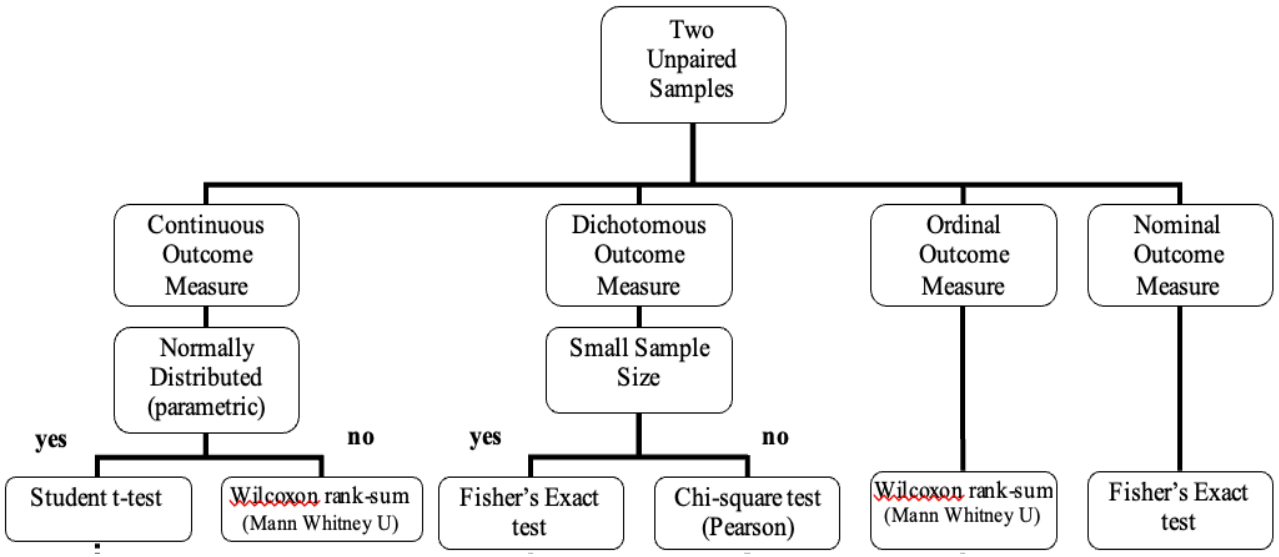
- What is the Research Question we are trying to answer
- What is the Research Design (obs/exp? and paired/unpaired?)
- What is the Outcome Variable (type)
- How is the outcome variable Distributed? [Assume not-normal]
- What is the most appropriate test?

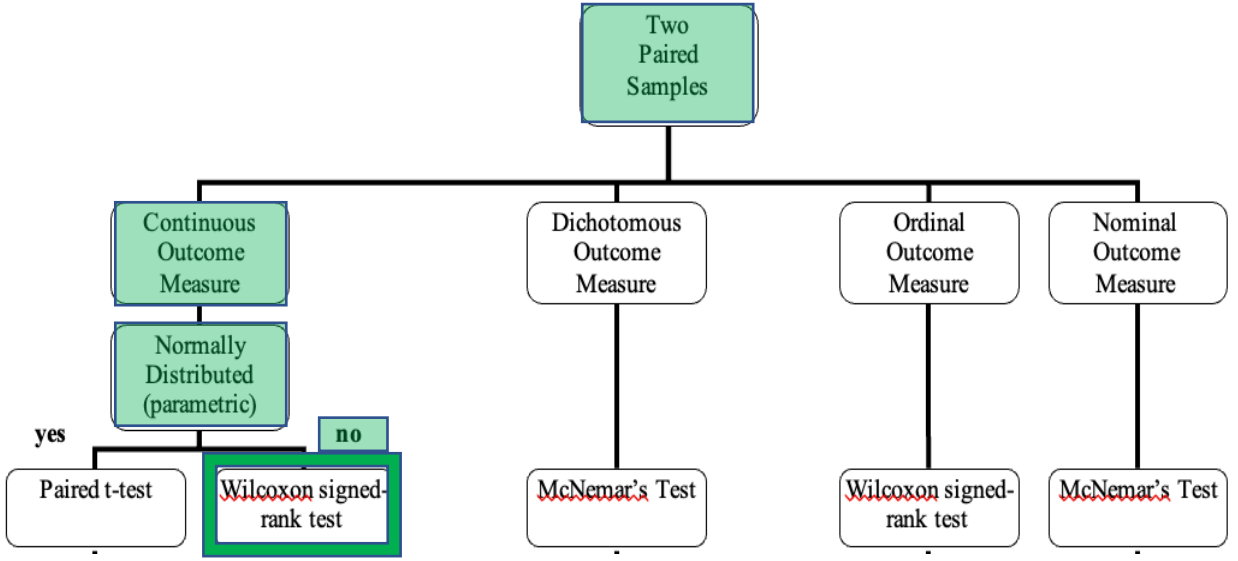


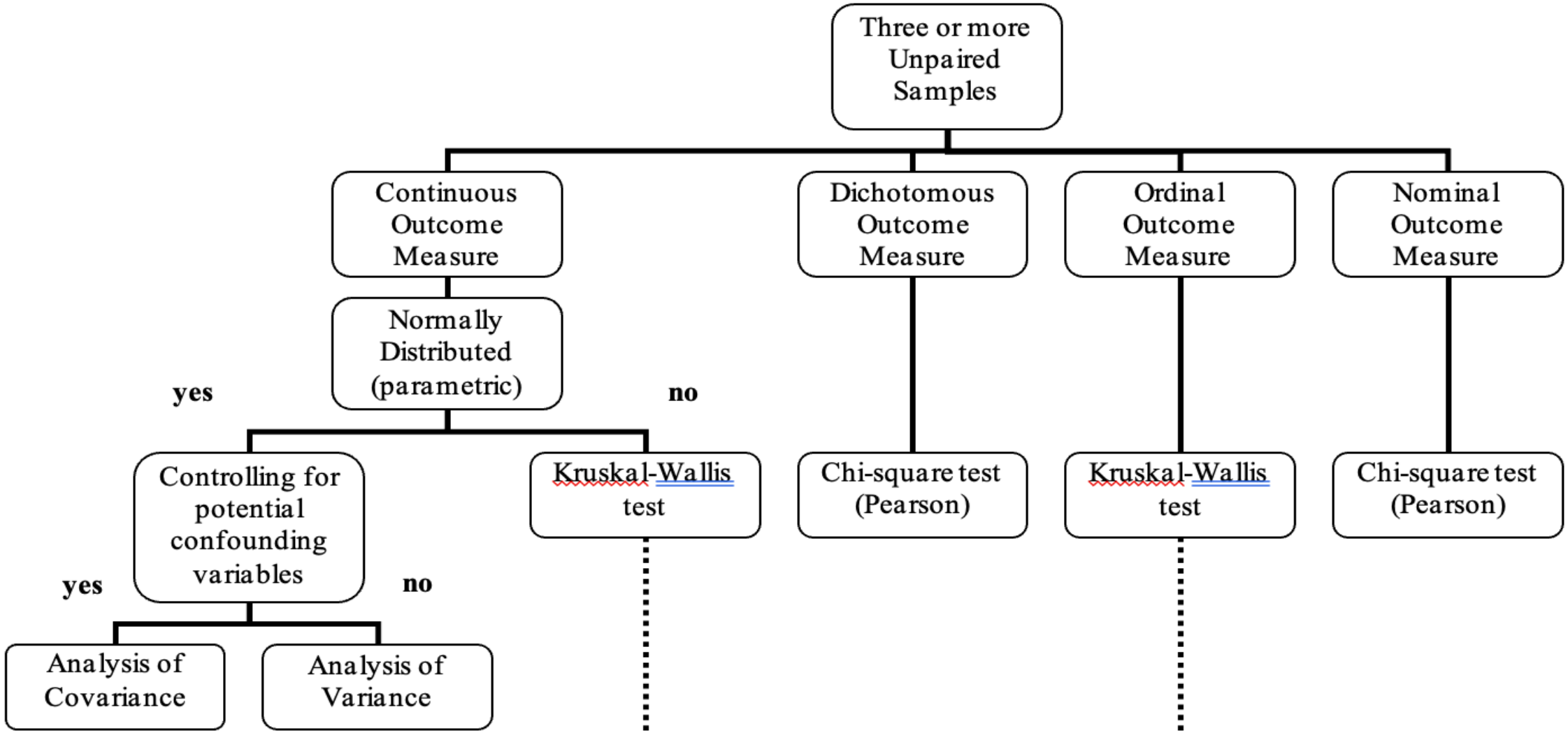
# Please Answer

- Do residents in the intervention group improve their communication skills from pre to post-intervention?
- Experimental, paired
- Continuous variable
- [Assume not-normal distribution]
- Wilcoxon signed-rank test



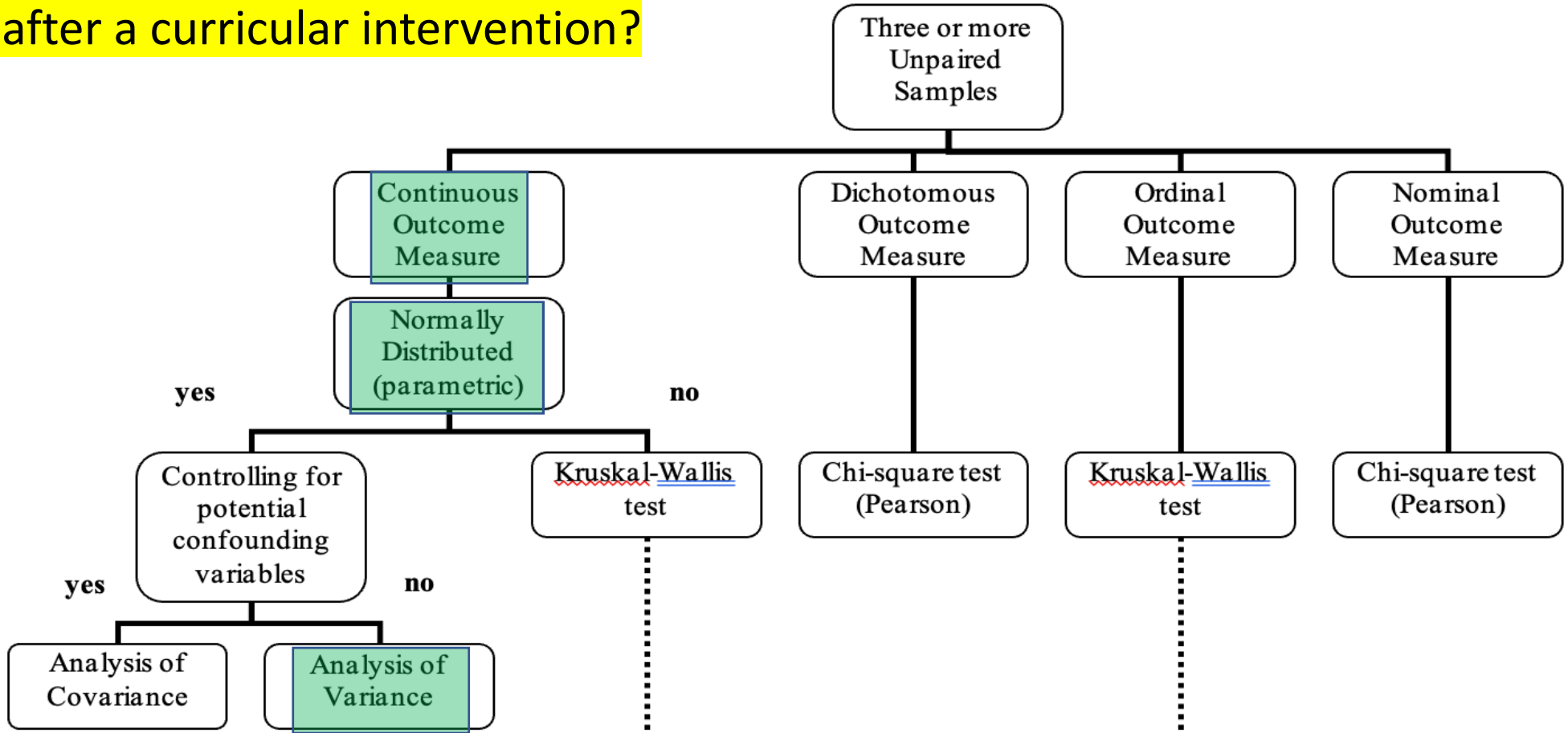
# 3+ samples, 1 outcome: Unpaired data



# 3+ samples, 1 outcome
# Unpaired data

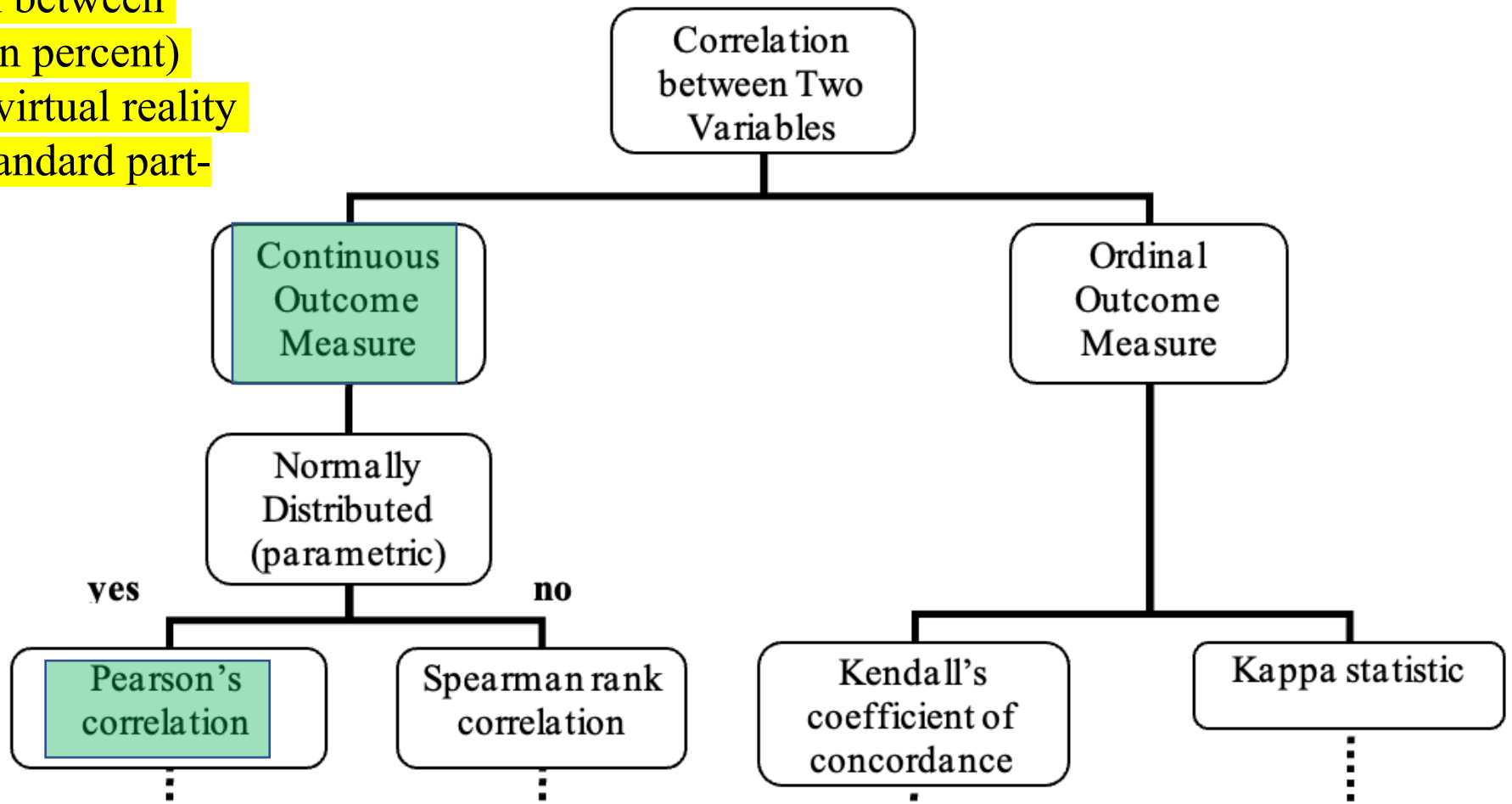
RQ: What is the difference in the sum communication skills performance on an OSCE exam of pediatric residents at three different institutions after a curricular intervention?

- **Case**: Communication (giving bad news) curriculum for pediatric residents
- **Study Aim** – An intervention for residents aimed at improving communication skills
- **Study Design**: Intervention vs Control groups
  - 50 residents in each group
  - 5-item assessment tool, using 5-point Likert scale, Summation score of the 5 items



# Correlation between Two Variables

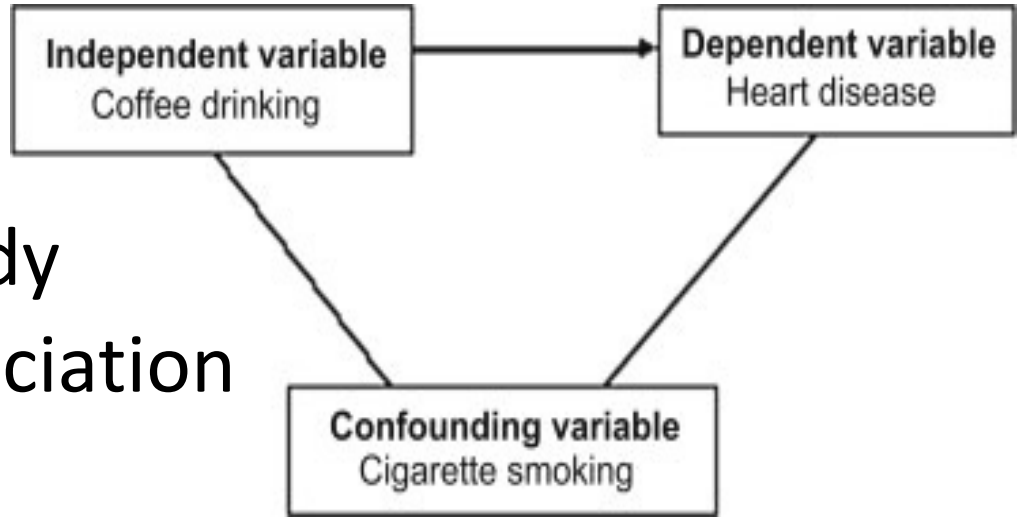
Determine the association between resident suturing scores (in percent) using a newly developed virtual reality simulation task and the standard part-task trainer.



# Regression

- Describe the association between 1 dependent variable and 1+ independent variables

- Adjusts for confounding variables
  - Related to the variable(s) in a study
  - May mask or falsely show an association
  - E.g., age, gender, comorbidities



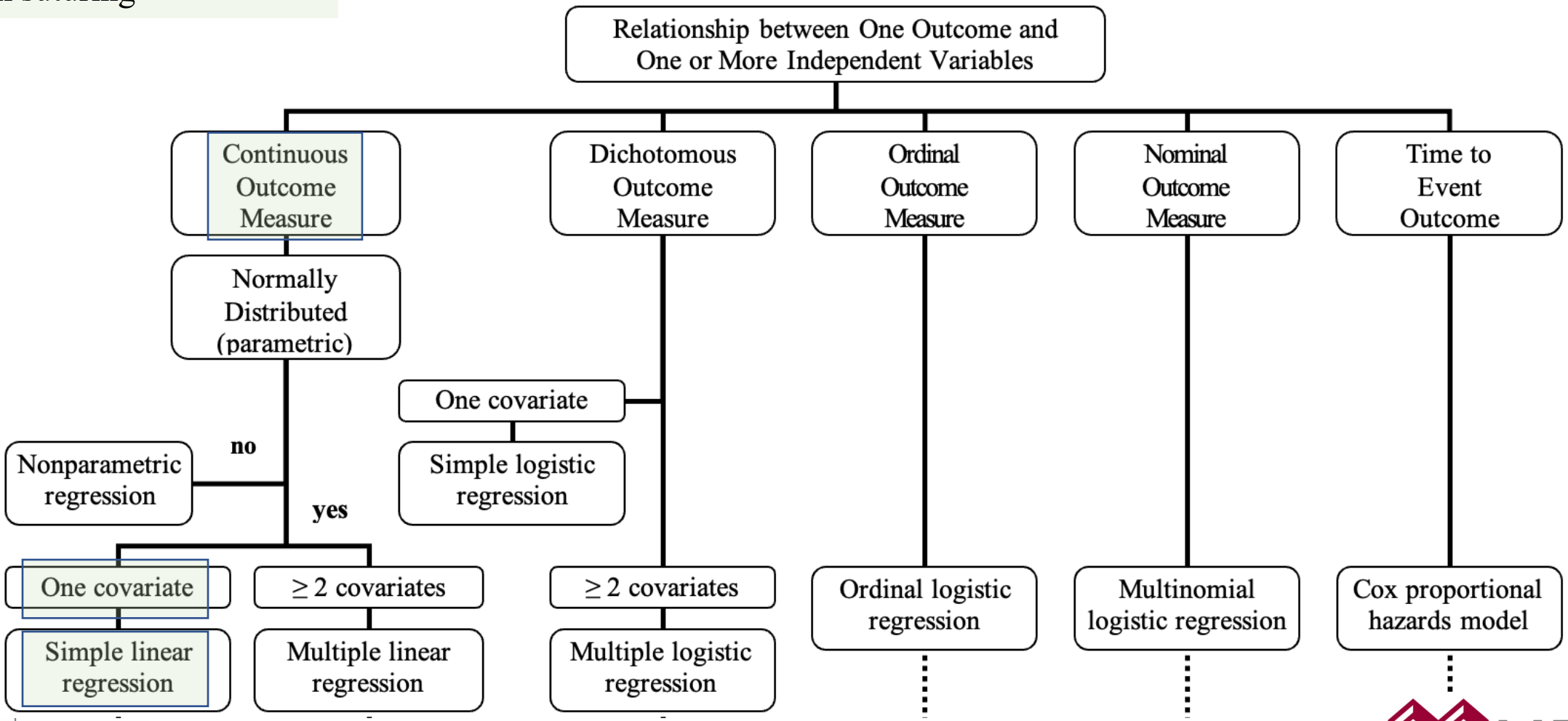
# Regression Analyses

- Deciding factors to control for

- Look at what factors are significant (univariate/simple regression)
- Control factors known to be confounding
- Control for factors where you assume would be confounding

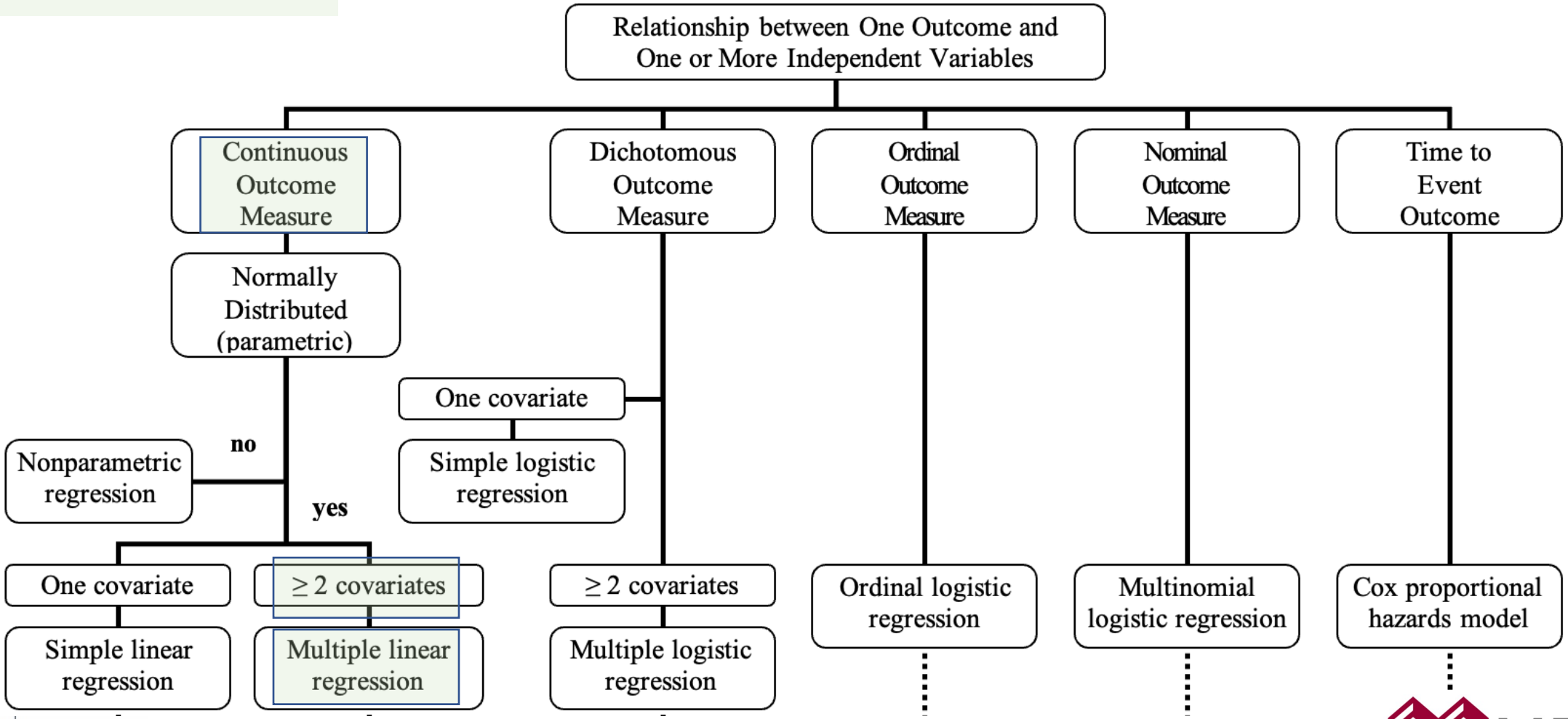
Determine the effect of using a specific type of suture on the length of time it takes residents to perform suturing

# Regression Analyses



Determine the effect of gender, age, and PGY level on the amount of time needed for suturing

# Regression Analyses




McMaster University | Pediatrics

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY

# Final Activity: Case Study

Surgical Innovation
18(1) 48–54

DOI: 10.1177/1553350610389826
http://sri.sagepub.com
SAGE

## GOALS-Incisional Hernia: A Valid Assessment of Simulated Laparoscopic Incisional Hernia Repair

Marilou Vaillancourt, MD,[1] Iman Ghaderi, MD,[1] Pepa Kaneva, MSc,[1] Melina Vassiliou, MD,[1] Nicoleta Kolozsvari, MD,[1] Ivan George, PhD,[2] F. Erica Sutton, MD,[2] F. Jacob Seagull, MD,[2] Adrian E. Park, MD,[2] Gerald M. Fried, MD,[1] and Liane S. Feldman, MD[1]

- The Global Operative Assessment of Laparoscopic Skills (GOALS) is a valid and reliable measure of basic, non-procedure specific laparoscopic skills. GOALS-incisional hernia (GOALS-IH) was developed to evaluate performance of laparoscopic incisional hernia repair (LIHR). **The purpose of this study was to assess the validity and reliability of GOALS-IH during LIHR simulation**. GOALS-IH assesses 7 domains with a maximum score of 35. A total of **12 experienced surgeons and 10 novices** performed LIHR on the Surgical Abdominal Wall simulator. Performance was assessed by a trained observer and by self-assessment using GOALS-IH, basic GOALS and a visual analog scale (VAS) for overall competence.

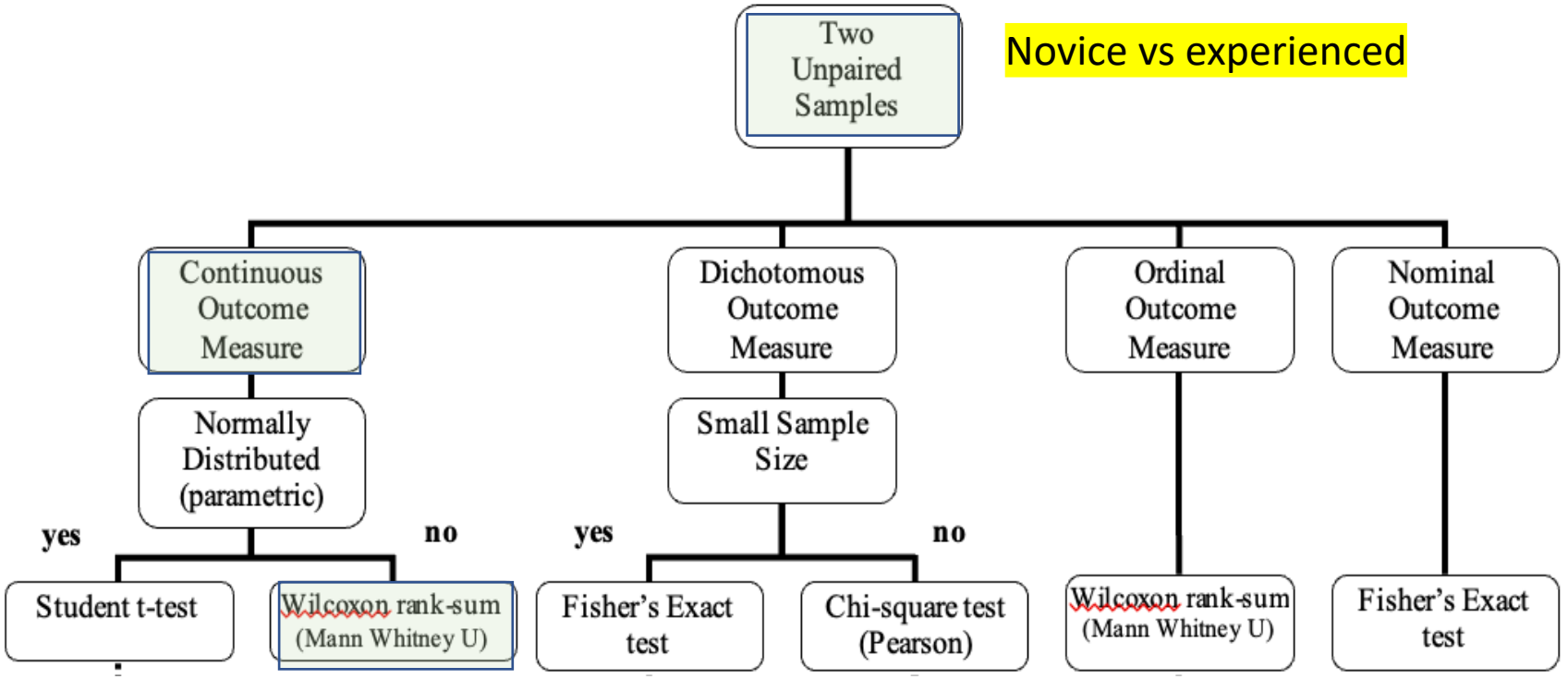
# Final Activity: Case Study

"Wilcoxon nonparametric test was used to compare case experience (total laparoscopic and LIHRs) between the novice and experienced groups."

What can we tell about the data for the RQ that led to this statistical test?

# Final Activity: Case Study

"Wilcoxon nonparametric test was used to compare case experience (total laparoscopic and LIHRs) between the novice and experienced groups."



# Summary and Conclusion

- Use the approach to choose a statistical test

1. What is/are the Research Question(s) we are trying to answer
2. What is the Research Design
3. What is/are the Outcome Variable(s) (type)
4. How is/are the outcome variable(s) Distributed?

- Diagrams will help
- Do not worry about memorizing tests

# Examples of Methodology related Electives Outside HSED (pre-approved)

HRM 702: Introduction to Biostatistics

Basic statistical concepts and techniques as they apply to analysis and presentation of data in biostatistical and epidemiology practice. The course covers: graphical presentation of data, elementary probability, descriptive statistics, probability distributions, and introduces hypothesis testing using parametric and non-parametric methods. Specific techniques covered include z-tests, t-tests, ANOVA, contingency tables, regression, and correlation. This course is offered in-person in the Fall term.

REHAB 774: Quantitative Research Methods

This course provides learners with an introduction to quantitative research methods commonly used in rehabilitation practice. It is intended to prepare learners to understand the foundations of clinical research, engage in critical evaluation of research, explore opportunities for research in their clinical practice, and enhance their skills as research collaborators. The course emphasizes the development of knowledge related to the formulation of research questions; reporting guidelines; specific observational designs as well as randomized control trials, systematic reviews and meta-analyses. This course is offered online in the Spring/Summer term.

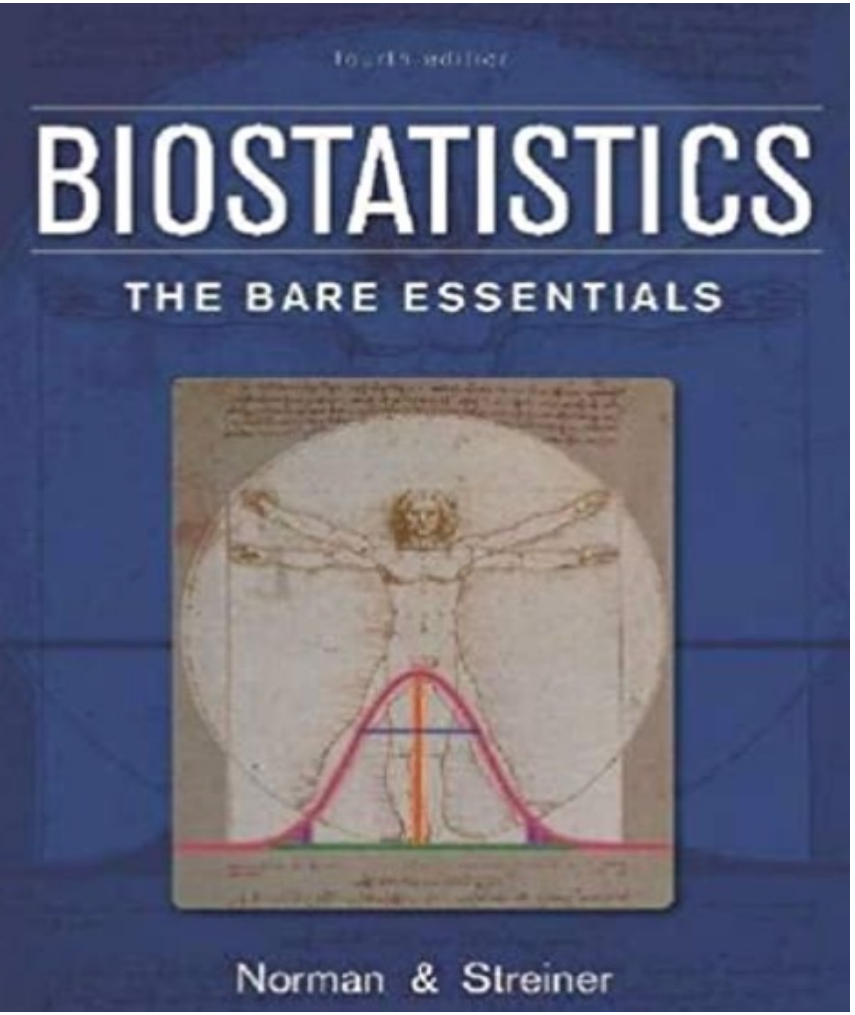
REHAB 772: Introduction to Qualitative Research

This is a graduate course for health professionals who have had little or no exposure to qualitative research and wish to acquire an introductory knowledge of qualitative research philosophy, methods, and methodological process. The course emphasizes the development of knowledge related to the philosophical and theoretical foundations of qualitative research, major qualitative research approaches, primary techniques for gathering data, and data management, analysis, and interpretation. Overall, this course encourages students to think more critically their assumptions, positionality, and experiences; as well as to see multiple interpretations and constructions of reality in relation to research. This course aims to help students understand the nature of lived experience and the importance of thinking and acting in critically reflexive ways. This course is offered online in Spring/Summer term.

NURS 745: Qualitative Health Research Methods

This course introduces learners to theoretical traditions and corresponding methods of qualitative research using health and health care research as examples. Specific topics covered include: theoretical paradigms of qualitative research, types of research questions best answered by qualitative methods, sampling objectives and procedures, methods of data collection, methods of analysis and interpretation, ethical issues, and responsibilities of qualitative researchers. Criteria for evaluating qualitative research will be discussed and applied to specific research studies. Learners will gain “hands on” experience using qualitative methods through in-class and take-home exercises. This course is offered in-person in the Winter term.

# References




**A Clinician-Educator's Roadmap to Choosing and Interpreting Statistical Tests**

*Donna M. Windish, MD, MPH,[1] Marie Diener-West, PhD[2]*

[1]Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA; [2]Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

MedEdPORTAL® | The AAMC Journal of Teaching and Learning Resources

Original Publication

**A Guide to Basic Statistics for Educational Research**

Donna M. Windish, MD, MPH*

https://www.youtube.com/watch?v=qwfd8cf3_UY
https://www.karger.com/Article/Fulltext/323136
https://learning.eupati.eu/mod/book/view.php?id=362&chapterid=388
https://www.sciencedirect.com/topics/nursing-and-health-professions/confounding-variable
https://www.simplypsychology.org/confidence-interval.html
https://www.mashupmath.com/blog/mean-median-mode-range-guide

# Statistics Literacy

By Dr Elif Bilgic
June 6 2023; HSED Residency Week
Assistant Professor & Education Scientist
e-mail: bilgice@mcmaster.ca

McMaster University | Pediatrics

# THANK YOU

# ANY QUESTIONS/ COMMENTS

MERIT
MCMASTER EDUCATION RESEARCH, INNOVATION & THEORY