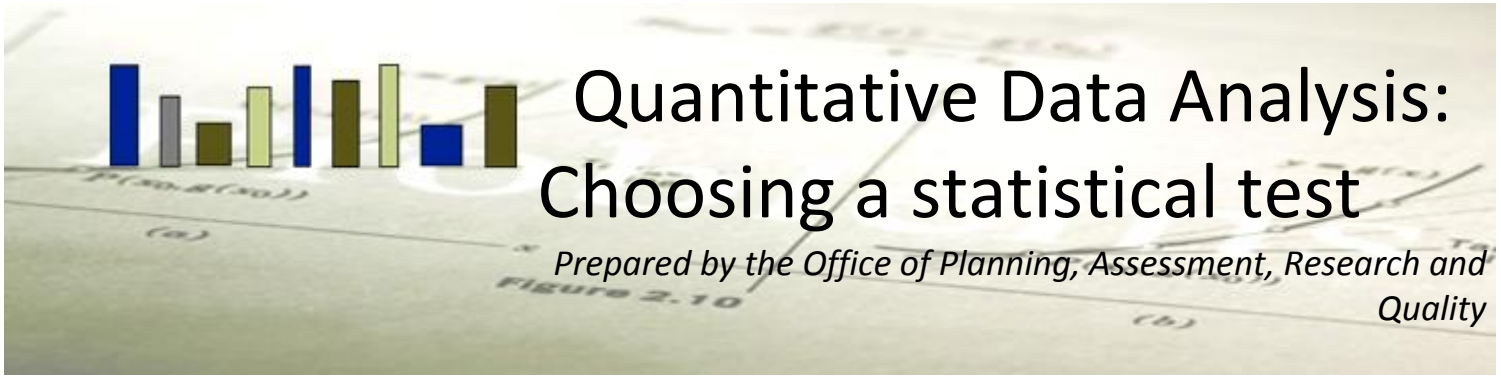


Quantitative Data Analysis: Choosing a statistical test

Prepared by the Office of Planning, Assessment, Research and Quality

To help choose which type of quantitative data analysis to use either before or after data has been collected.

Before beginning this step in the research process, it is important to know the following information about the project:

- What is/are your specific research question(s)?
- What types of data will you collect -- nominal, ordinal, or ratio? (See Glossary for definitions).
- What is/are the projected size(s) of your sample and groups?
- What are your independent and dependent variables?

Once you have this information you are ready to move through the document.

For questions, please contact:

Susan Greene
Institutional Planner – Institutional Effectiveness
Office of Planning, Assessment, Research & Quality
116 Bowman Hall | University of Wisconsin-Stout
P: 715.232.1638 | F: 715.232.5406 | greenes@uwstout.edu | www.uwstout.edu/parq

To Consider Before Choosing an Analysis

Before going through a selection table, items to identify include:

Independent Variable (IV)

Variable that is either manipulated by the researcher or that won't change due to other variable. Can be thought of as either the cause of change in the dependent variable, or impacts the dependent variable.

Examples:
Demographics such as gender, year in school; experimental/ control group; time (pre/post)

Dependent Variable (DV)

Variable whose change depends on change in another variable (IV). Can be thought of as the "effect" due to independent variable "cause"; the impacted variable. The researcher does not manipulate this variable.

Examples:
satisfaction rating, course grade, retention in program, anxiety score, calorie intake, test score

Level of data

- Nominal – examples: gender, ethnicity
- Ordinal – examples: ranking preference, age categories
- Interval – examples: Likert rating scale, test score

Research Question(s)

Research Questions are the reason why collect data – what specifically do you want to know?

Examples:
- Who answered my survey?
- How satisfied are my clients?
- Does overall satisfaction differ by gender?
- Do test scores change after a reading intervention is given?

Tip: If survey project, before you use these decision tabs create a codebook that
-- maps each of the survey questions to the relevant Research Question
-- identifies the level of data for <u>each</u> survey question
-- identifies the independent and dependent variables for each Research Question

Quantitative Data Analysis Decision Guide Home Page

What do I want to know?

For additional information on items to identify before selecting an analysis, see *Things to Consider* tab

Describe the sample – *How can I summarize my data?*

- Central tendency
  - Mean, median, mode
- Dispersion around the central tendency
  - Standard deviation, range
- Distribution of responses
  - Frequency/ percentage of responses

To learn more about **Describe click here** or go to the *Describe* page

Compare the responses of the sample – *How did the data differ across groups?*

- Was my sample demographics similar to underlying population characteristics?
- Do groups within my data differ on a measure?
- Do the individual's responses differ across the measures?

To learn more about **Compare click here** *or* go to the *Compare Home* page

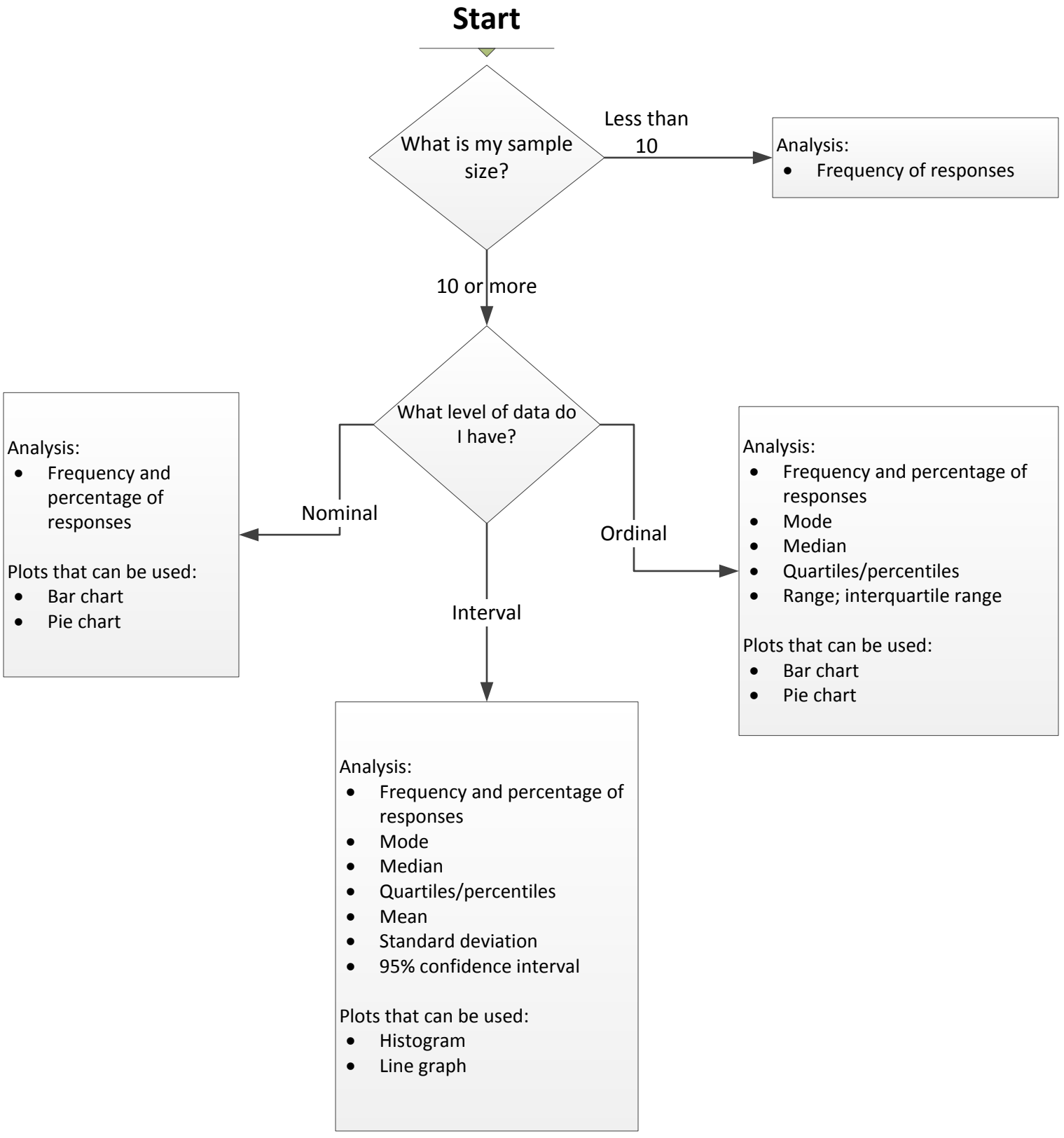
Make **predict**ions based on the responses of the sample – *How can I summarize the relationship between measures?*

- Is there a relationship between responses on 2 measures?
- How well can I predict an outcome based on the measures?

To learn more about **Predict click here** or go to the *Predict* page

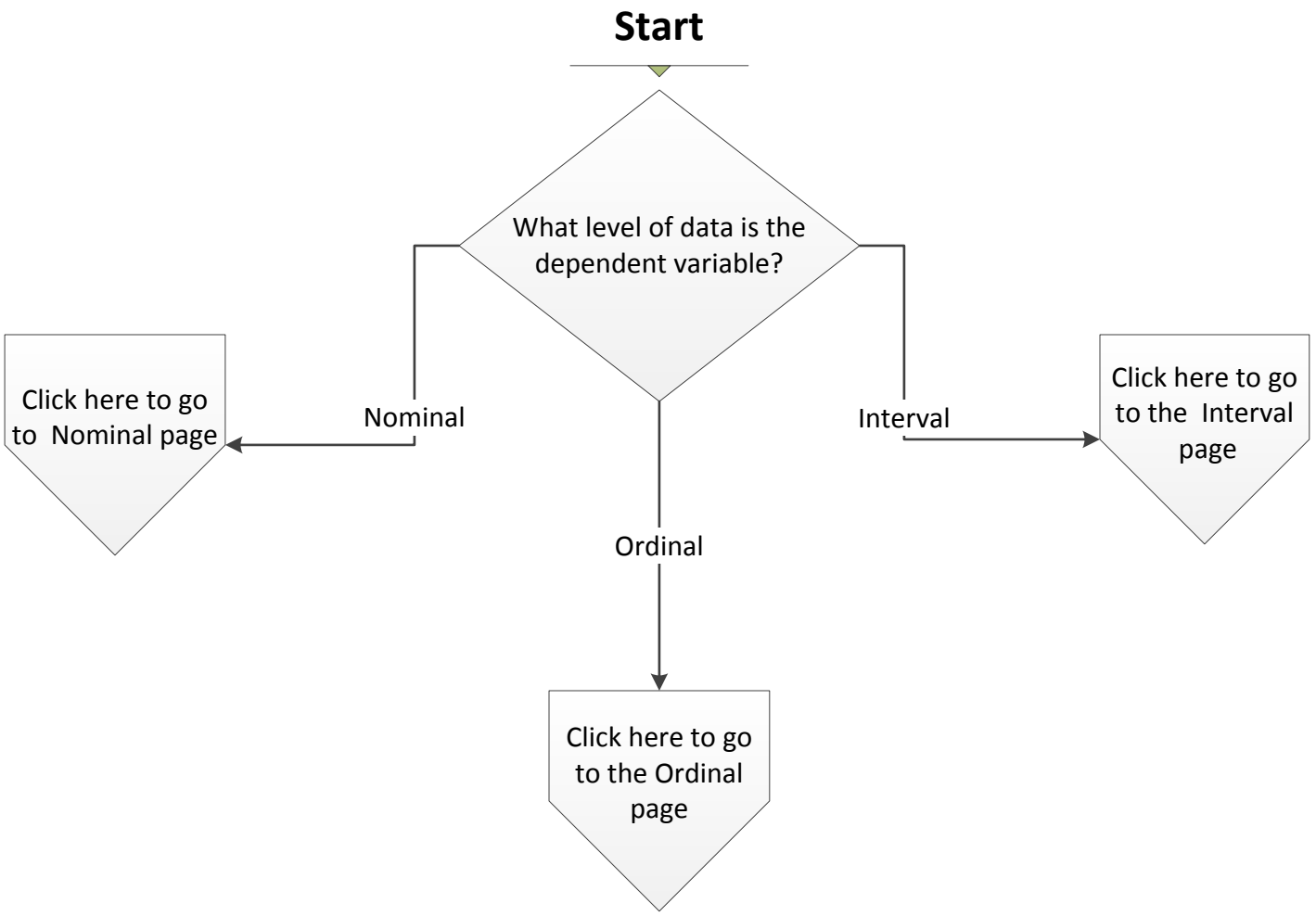
Describe

How can I summarize the data?

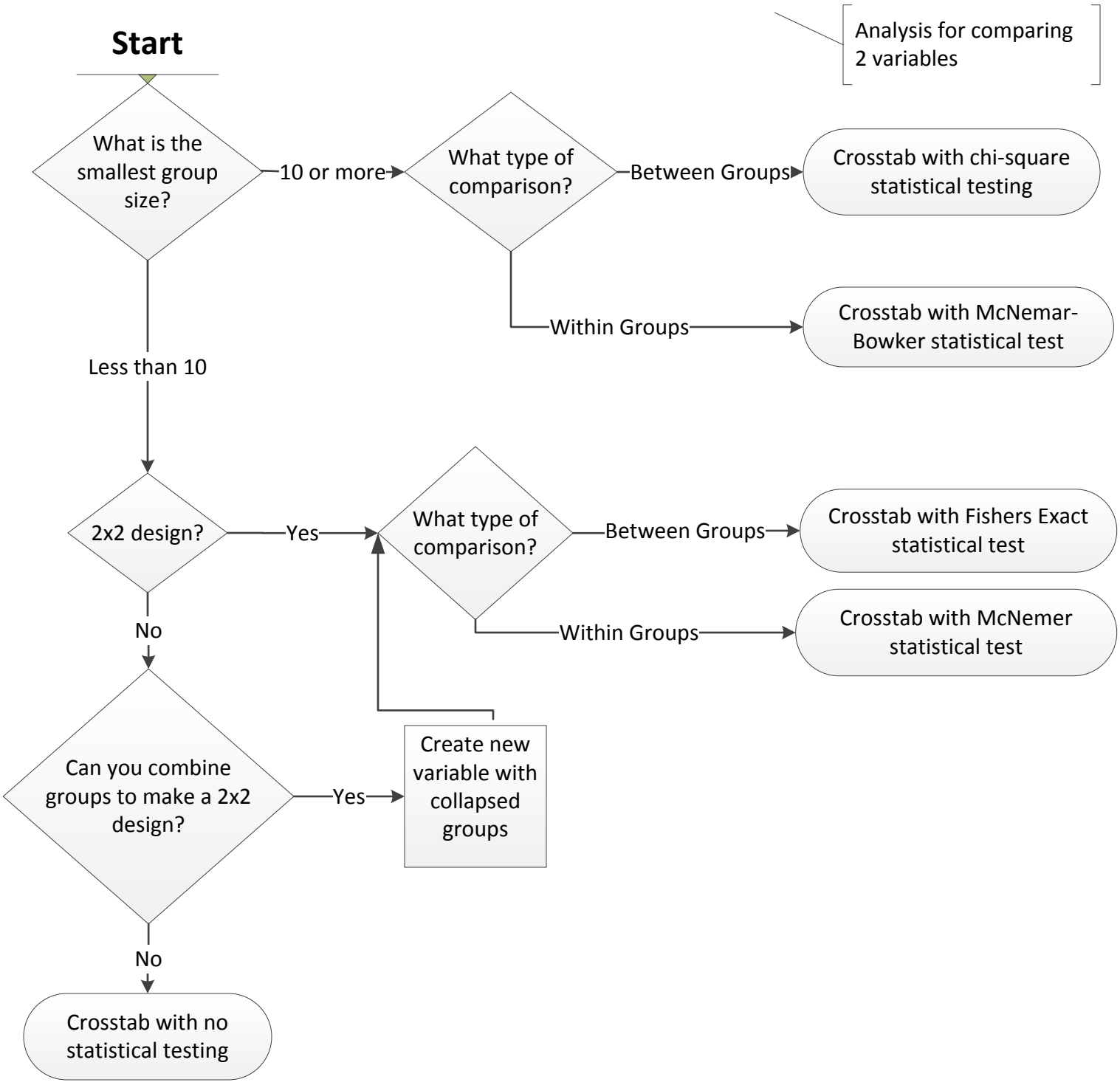


Compare

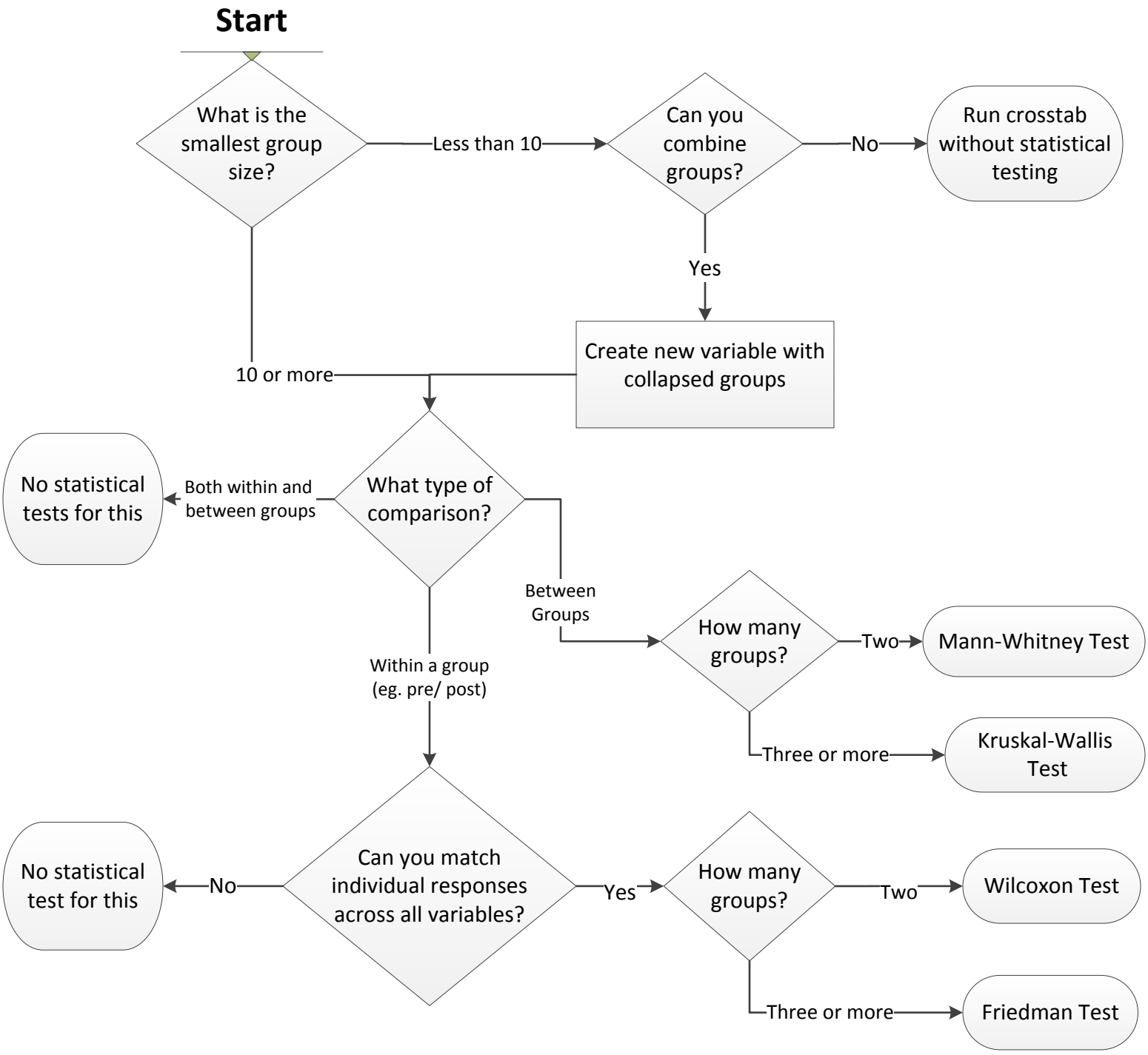
How did the data differ across groups?



Compare: Nominal Data

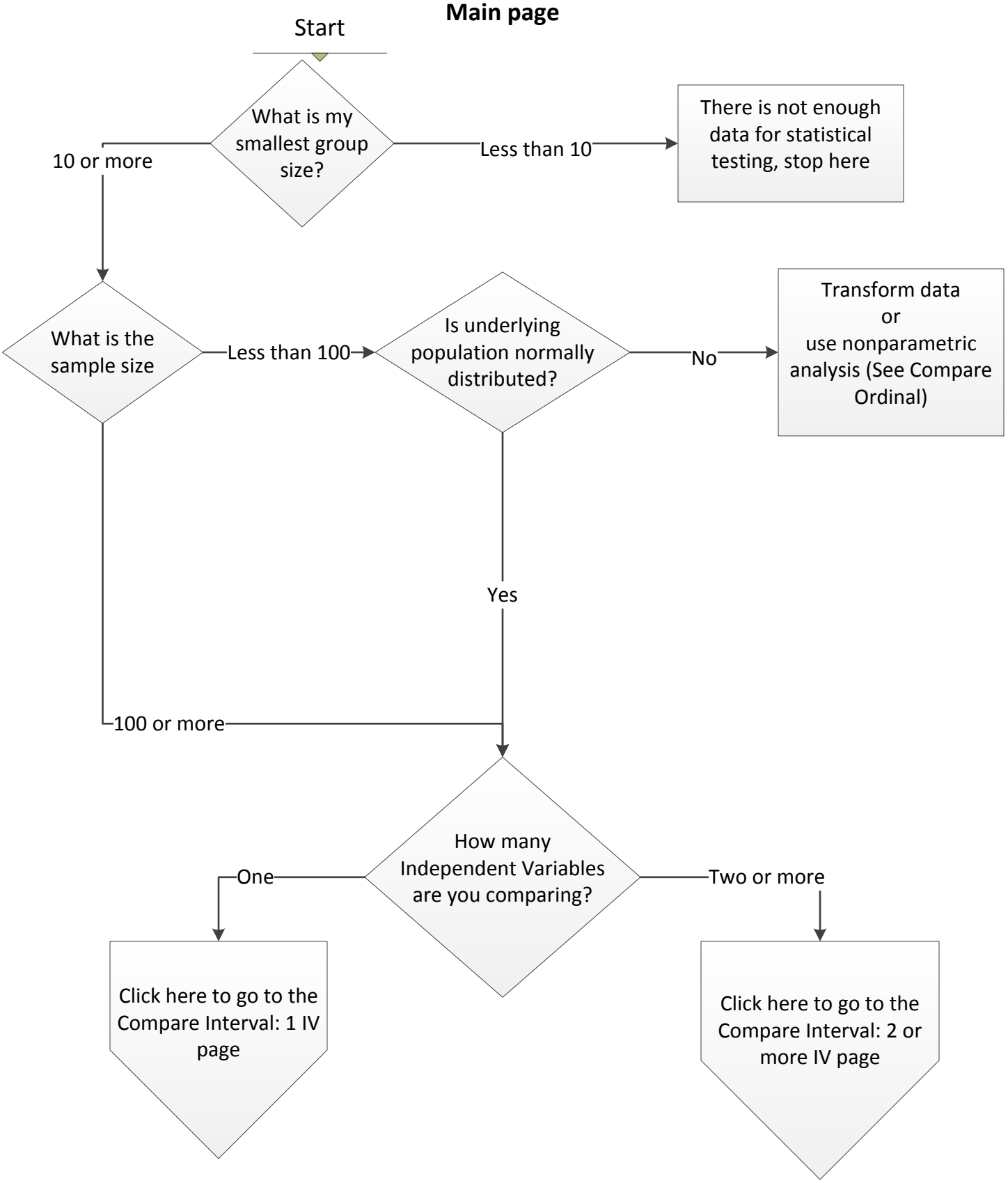


Compare: Ordinal Data



Compare: Interval Data

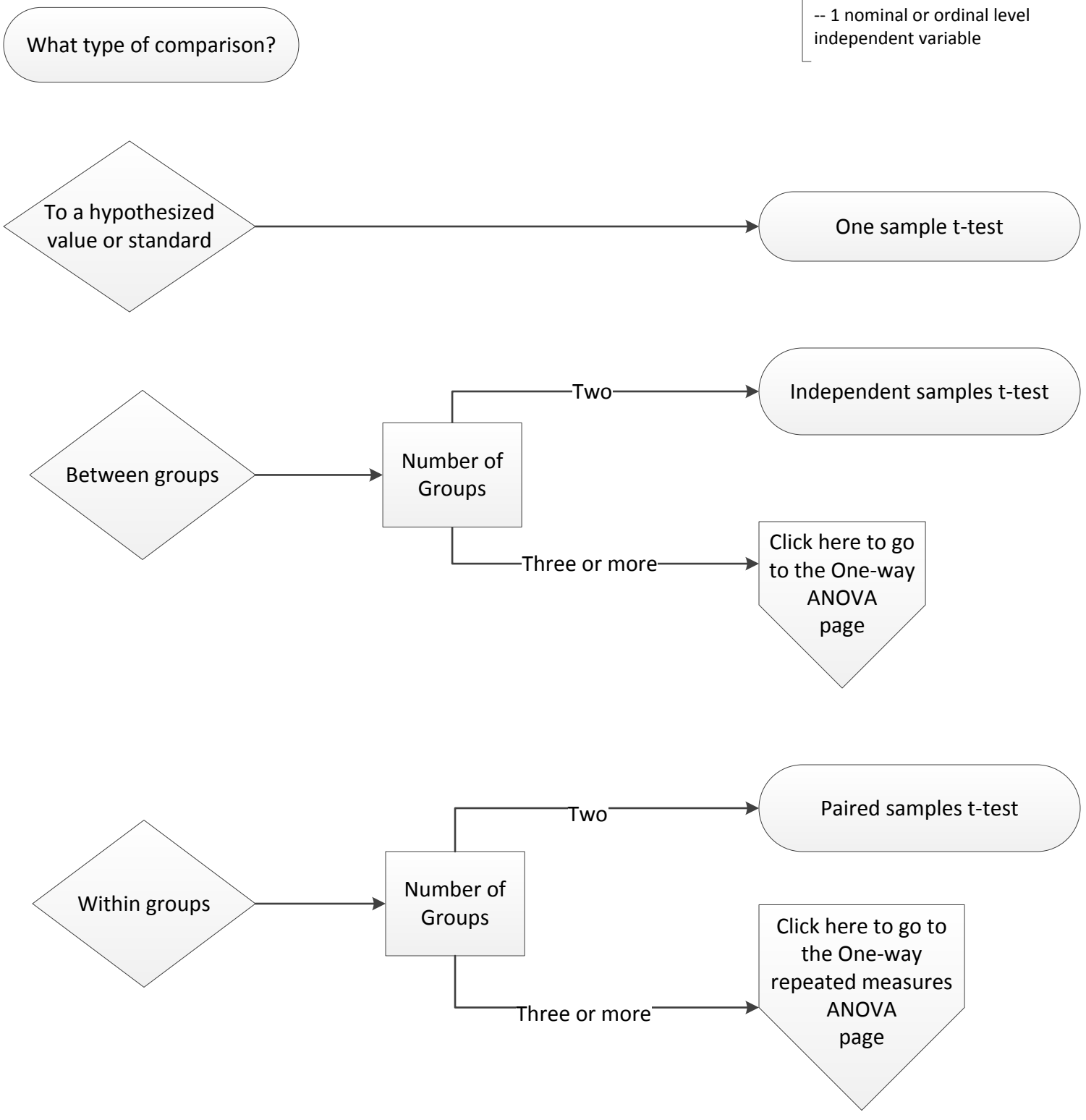
Main page



Compare: Interval Data

One Independent Variable, One Dependent Variable

Requires:
-- 1 interval level dependent variable and
-- 1 nominal or ordinal level independent variable



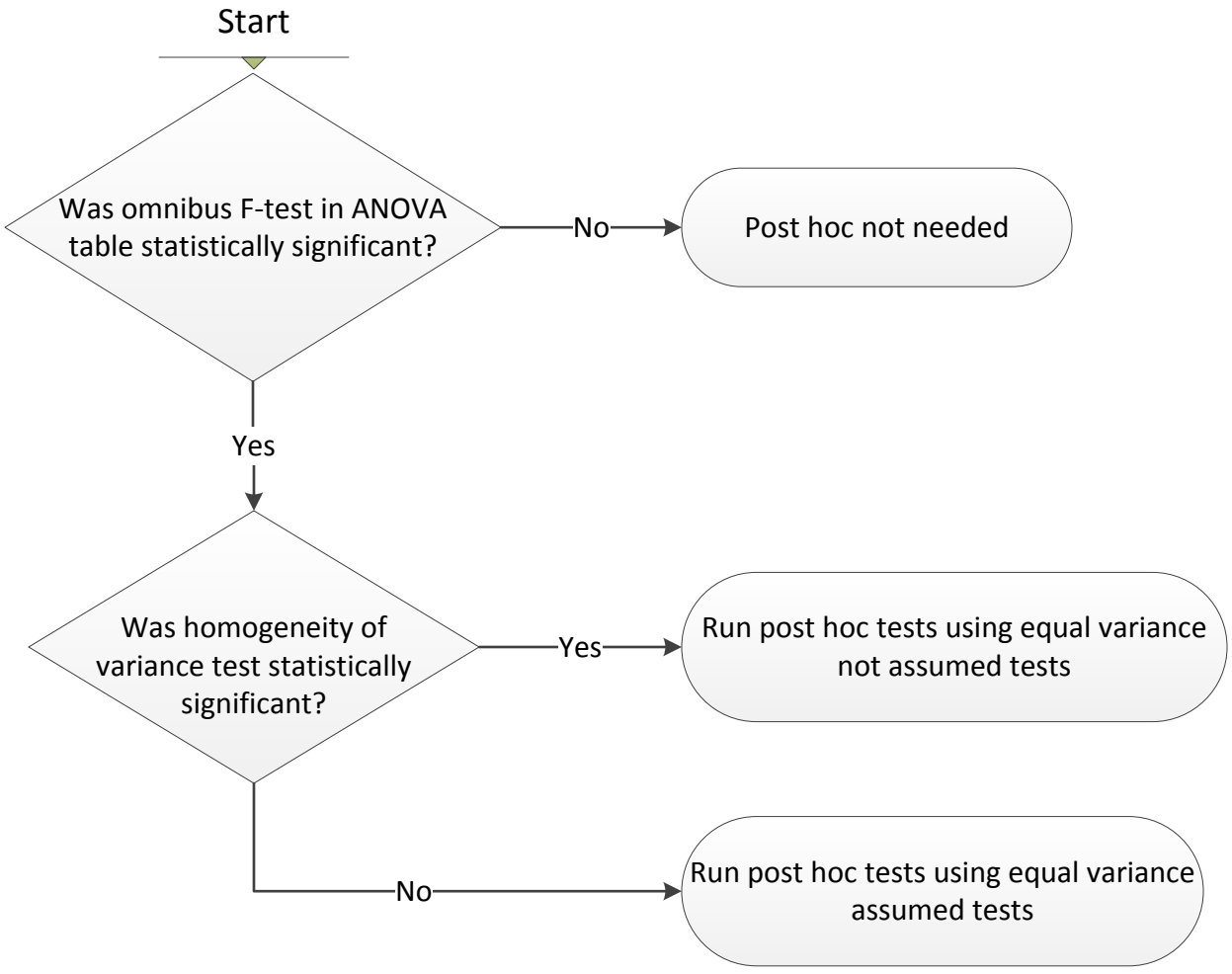
One-way ANOVA Analysis

Requires:
-- 1 Interval dependent variable
-- 1 nominal independent variable with 3 or more groups

Note: ensure that assumptions from Compare Interval Home Page are met prior to using this analysis

Perform one-way ANOVA with:
- Descriptive statistics
- Test for homogeneity of variance
- Estimate of effect sizes

Interpret results



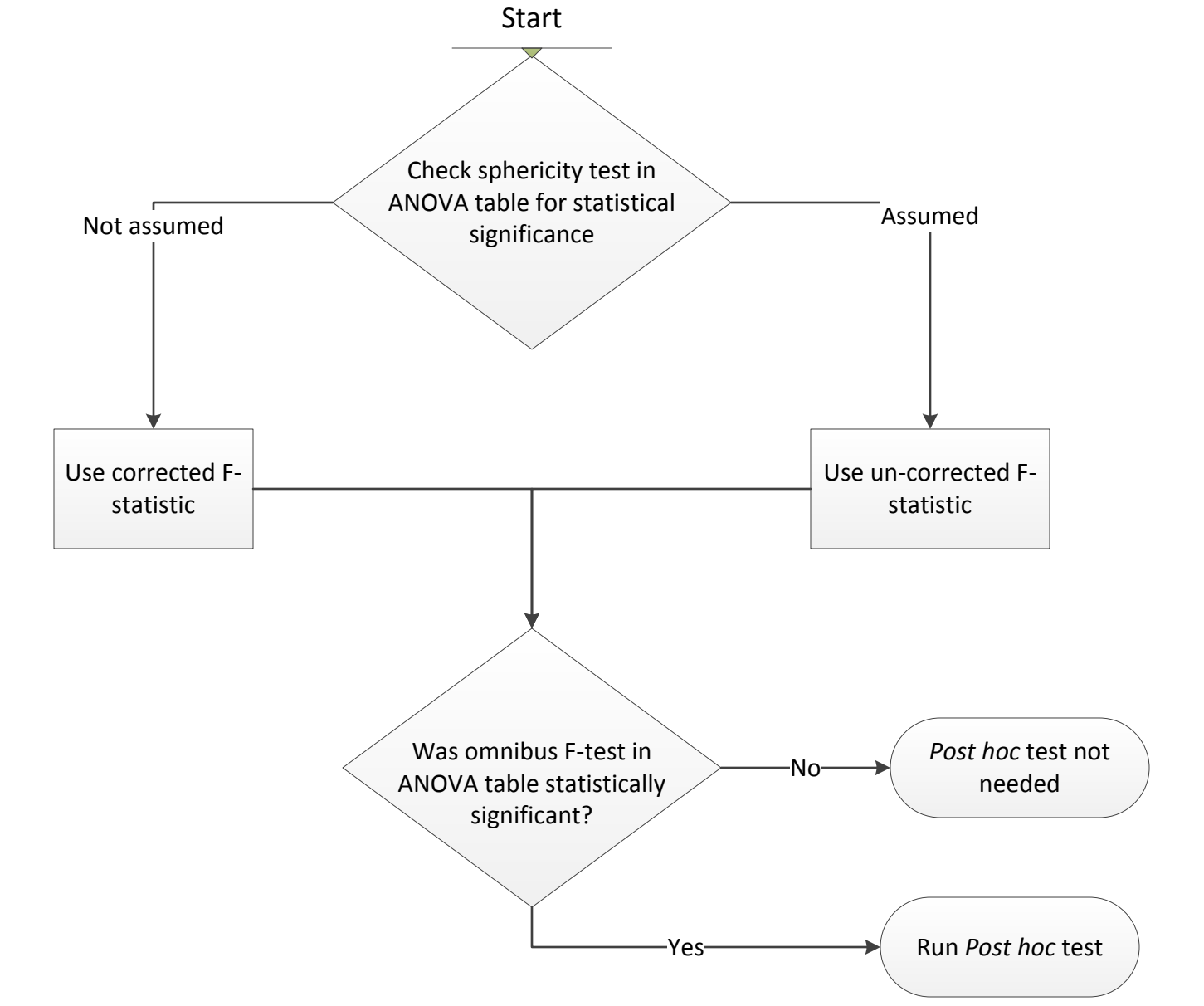
One-way Repeated Measures ANOVA

Requires:
-- 1 Interval dependent variable with matched measures across all of the repeats
-- 1 nominal or ordinal independent variable with 3 or more repeated measurements

Note: ensure that assumptions from Compare Interval Home Page are met prior to using this analysis

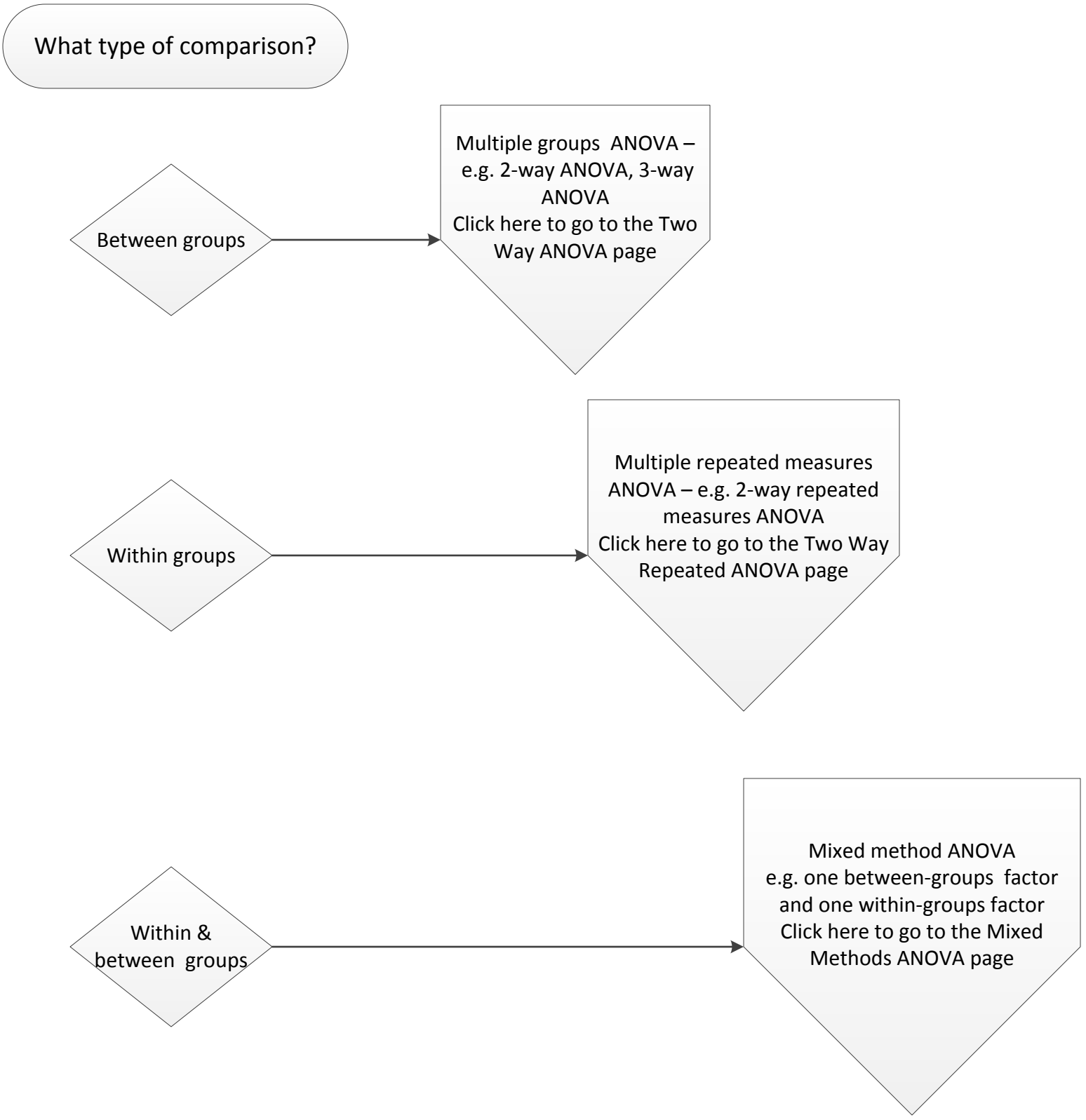
Perform one-way repeated measures ANOVA with:
- Test for Sphericity
- Estimate of effect size

Interpret results



Compare: Interval Data

Two or More Independent Variables, with one Dependant Variable

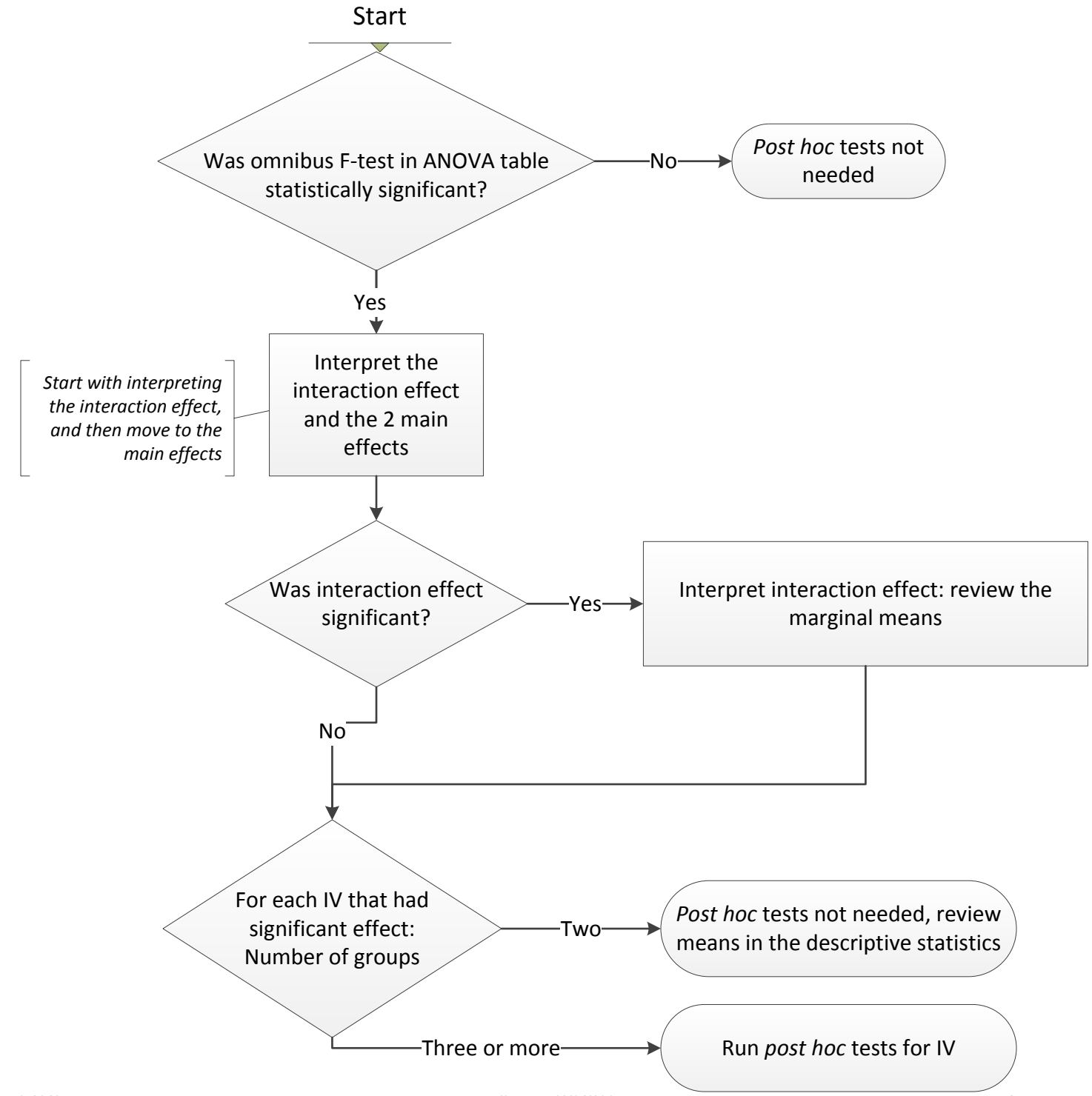


Two-way ANOVA Analysis

Requires:
- 2 nominal or ordinal independent variables (IV) with 2 or more groups each and at least 20 data points of the dependent variable per grouping cell
- 1 interval dependent variable (DV)
- Minimum of 20 data points of the dependent variable per grouping cell

Perform two-way ANOVA with:
- Descriptive statistics
- Tables and plots for marginal means
- Estimate of effect sizes

Interpret results



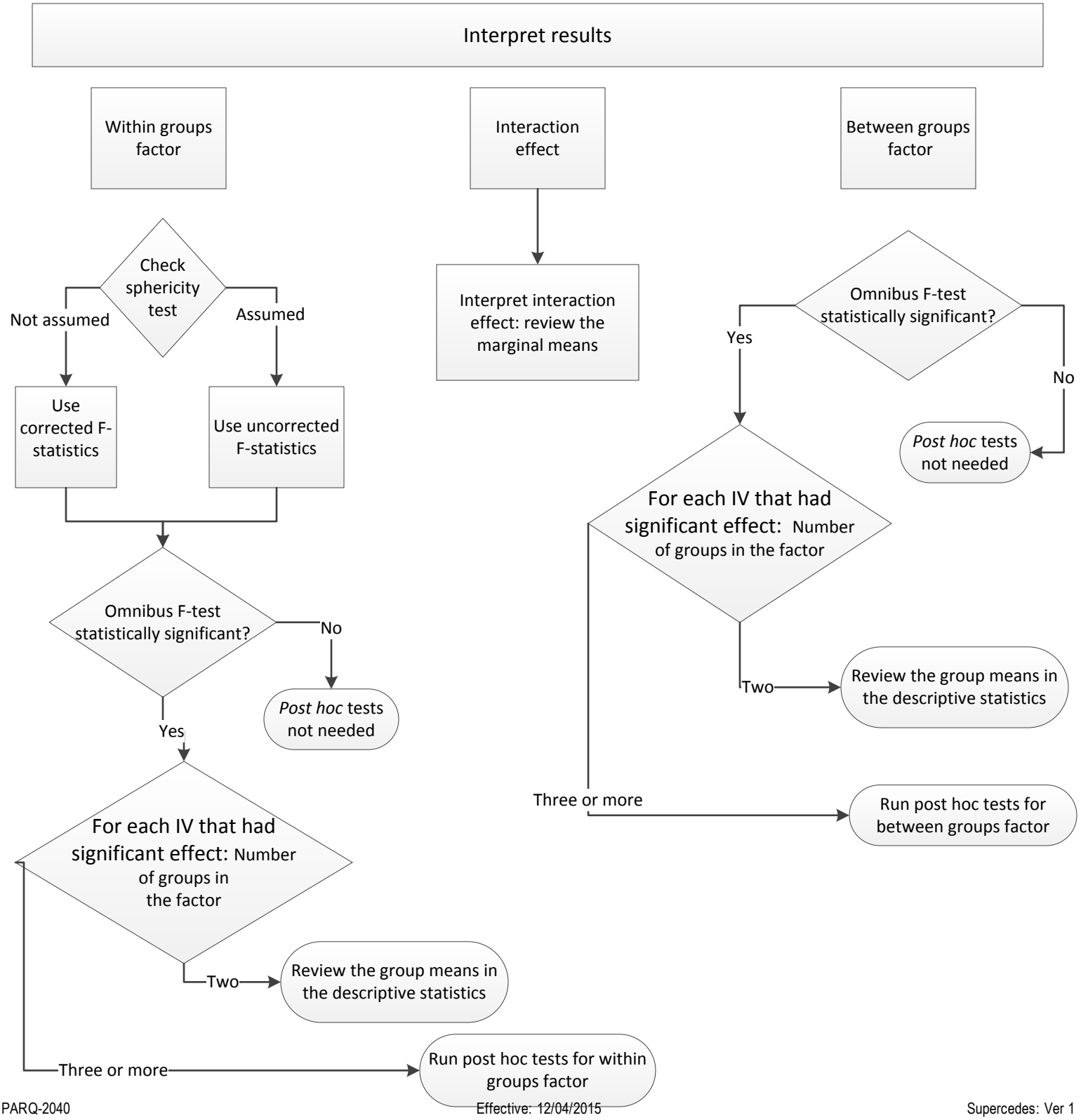
Mixed Methods ANOVA Analysis

Perform mixed methods ANOVA with:

- Descriptive statistics
- Test for sphericity

Requires:

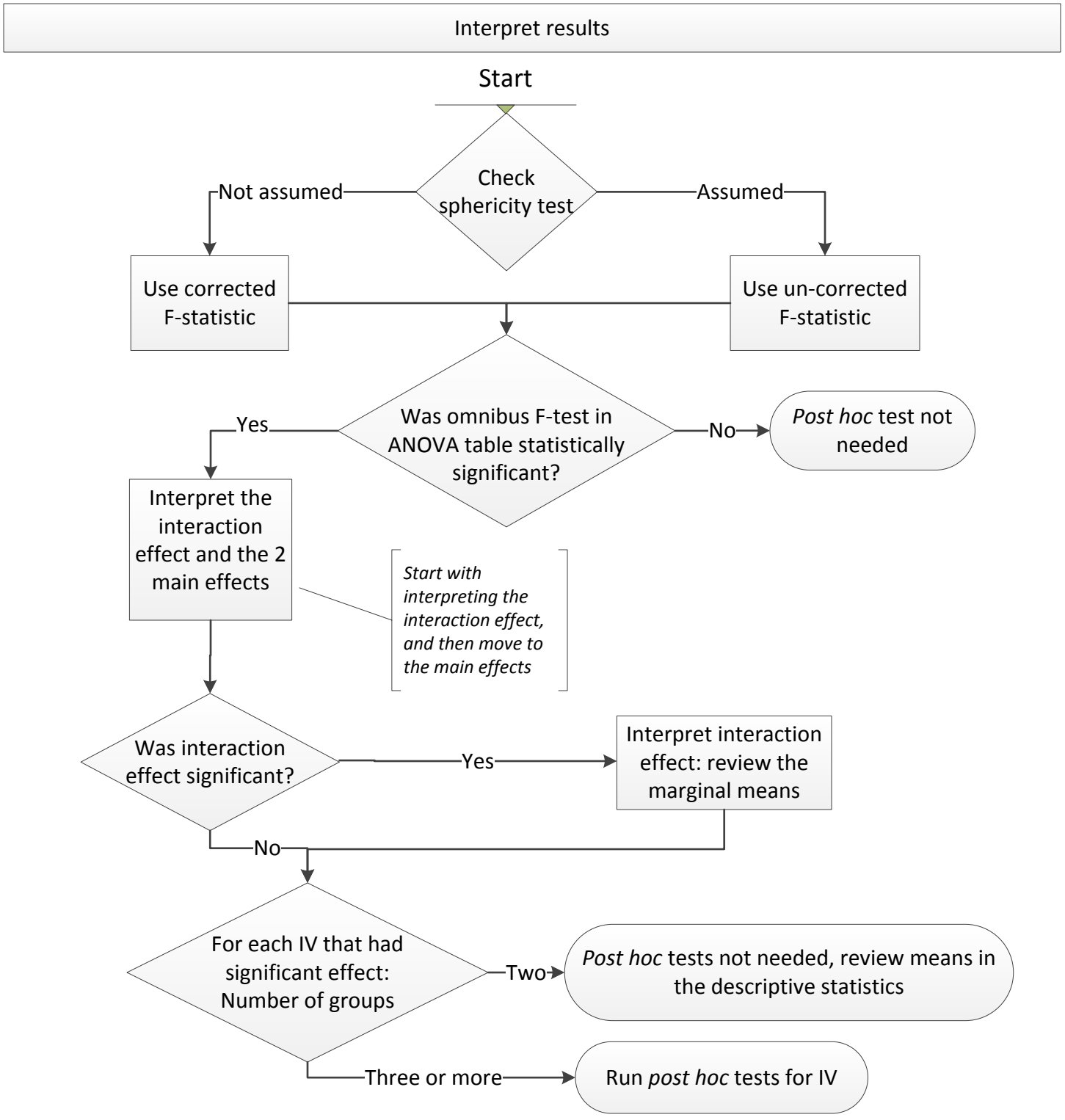
- 1 or more nominal independent variables (IV) with 2 or more groups [between groups factor]
- 1 nominal independent variable (IV) with 2 or more repeats [within groups factor]
- 1 interval dependent variable (DV)
- Minimum of 20 data points of the dependent variable per grouping cell



Two-way Repeated Measures ANOVA

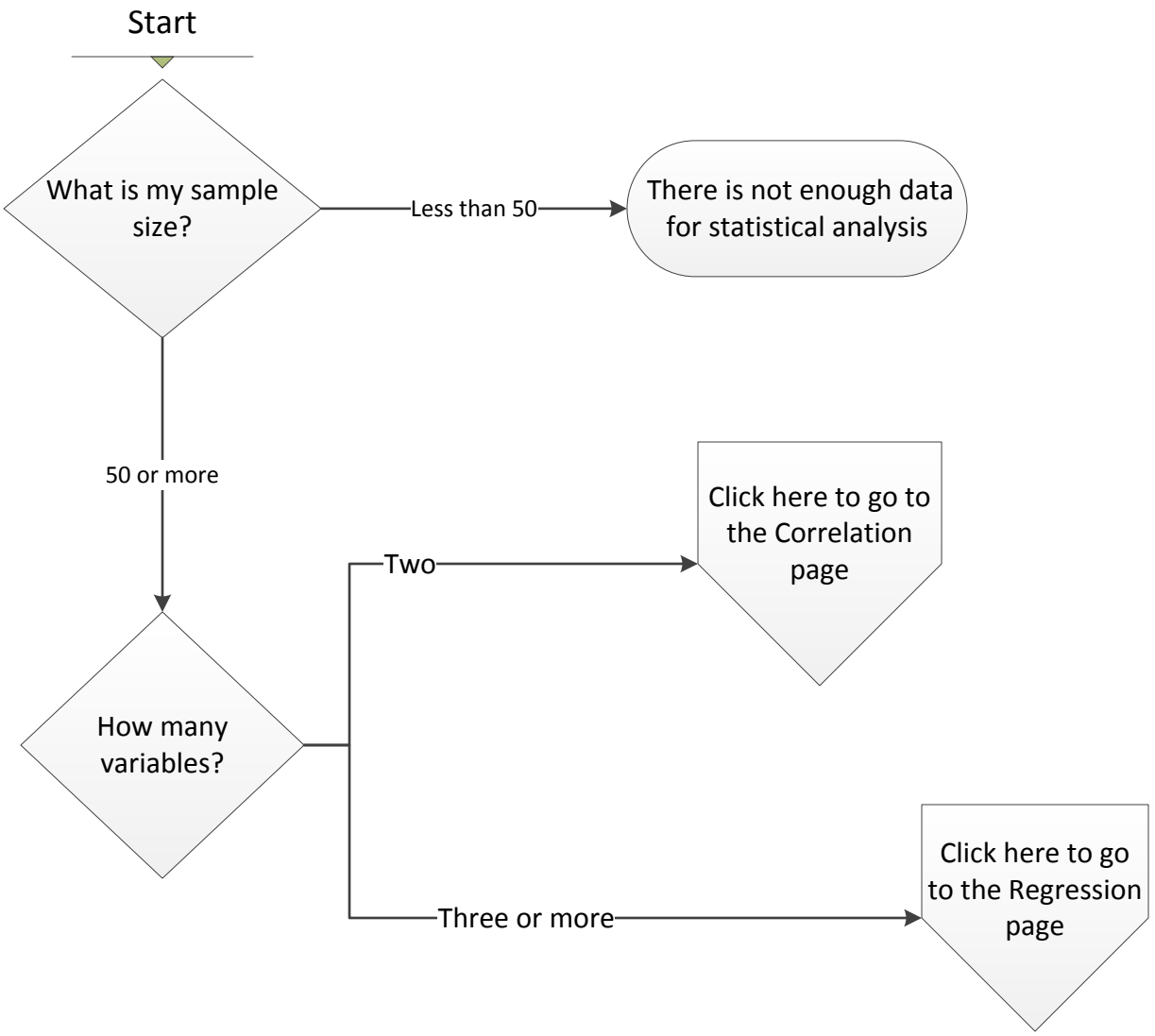
Requires:
- 1 Interval dependent variable (DV) with matched measures across all of the repeats
- 2 nominal or ordinal independent variable (IV's) with 2 or more repeated measurements. Most commonly the two IV's are time and condition
- Minimum of 20 data points of the dependent variable per grouping cell

Perform two-way repeated measures ANOVA with:
- Test for Sphericity
- Estimate of effect size



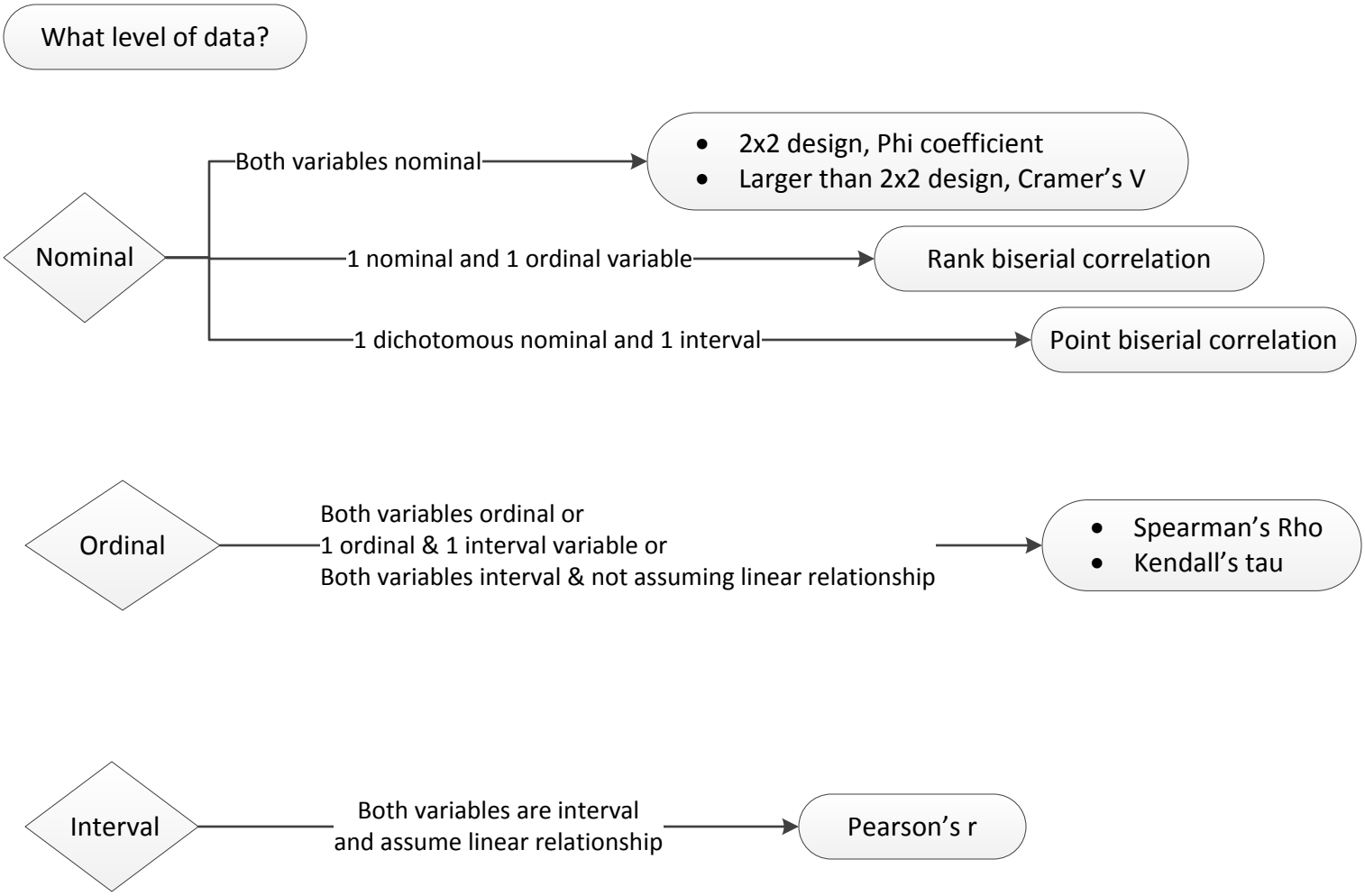
Predict

How can I summarize the relationship between variables?



Correlation

Testing for relationship between two variables

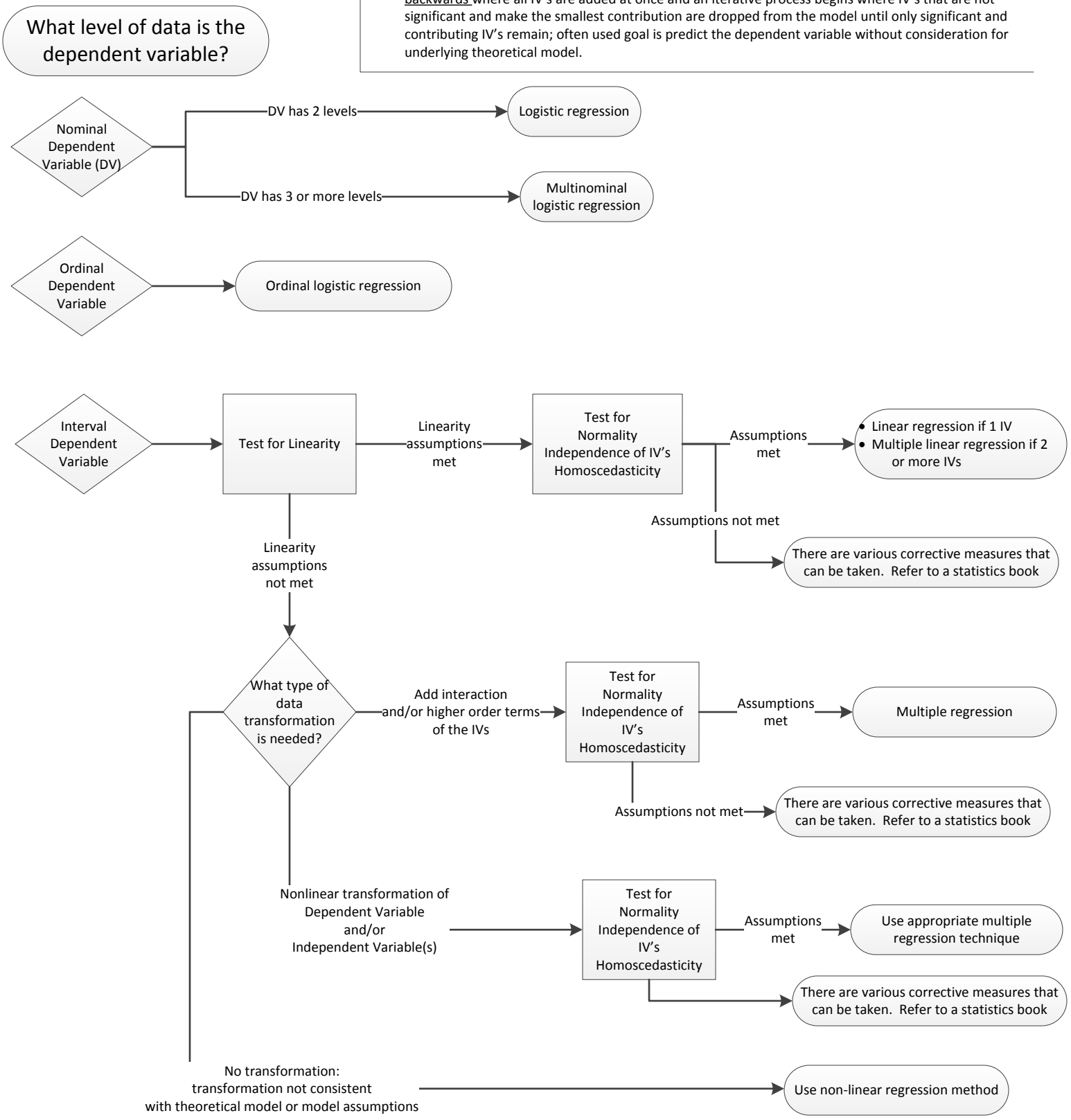


Regression

Considerations:
Minimum sample size for regression is best estimated using power analysis prior to collecting data. See https://www.uwstout.edu/parq/intranet/upload/Methods-for-determining-random-sample-size.pdf

There are 3 approaches to performing regression, depending on the research question

- Simultaneous method, where all of the independent variables (IV's) are treated together and at the same time; used when no theoretical basis for one or a group of IV's to be prior to another in the model.
- Hierarchal method, where groups of independent variables are entered cumulatively according to a hierarchy specified by the theory or logic of the research; used when there is a theoretical basis for one or a group of IV's to be prior to another in the model.
- Stepwise method, where the "best" set of independent variables are selected posteriori by the software – forward, where the model sequentially adds IV's until R^2 no longer increases; and the backwards where all IV's are added at once and an iterative process begins where IV's that are not significant and make the smallest contribution are dropped from the model until only significant and contributing IV's remain; often used goal is predict the dependent variable without consideration for underlying theoretical model.



Glossary

Bar chart **- a graph using parallel bars of varying lengths to illustrate frequency of responses, for example number of responses per year in school, per satisfaction level, etc.**

Between groups **– design where the comparison is between mutually exclusive groups. For example, comparing responses of males and females. *Comparing you to me.***

Dependent variable **- Variable whose change depends on change in another variable (IV). Can be thought of as the "effect" due to independent variable "cause"; the impacted variable. The researcher does not manipulate this variable. Examples: satisfaction rating, course grade, retention in program, anxiety score, calorie intake, test score.**

Frequency **- This number represents a count of the number of respondents that chose a specific answer for a question.**

Group **– all the possible responses in a variable. For example, if gender was asked as male/female, then there were 2 groups.**

Group size **– the number of respondents in the group. For example, if you had data from 15 respondents and there were 10 males and 5 females, the then group size of the males was 10.**

Histogram **- a graph of a frequency distribution in which rectangles with bases on the horizontal axis are given widths equal to the class intervals and heights equal to the corresponding frequencies**

Independent variable **- Variable that is either manipulated by the researcher or that won't change due to other variable. Can be thought of as either the cause of change in the dependent variable, or impacts the dependent variable. Examples: demographics such as gender, year in school; experimental/control group; time (pre/post).**

Interaction effect **- This tests to see if there was a differential effect on the dependent variable depending on which set of groups the person belonged to. For example, was there different effect on average income for the gender groups based on their minority status?**

Interval or ratio data **- data where the numbering of responses indicates both relative and absolute strength/value of responses. Therefore, the difference between two values is a meaningful measurement. For example, Likert-type rating scales can be considered interval data; age in years is ratio data.**

Glossary

Level of data **– the structure and nature of the data collected; level of data determines what type of analysis can be used.**

Line graph **- Line graphs compare two variables. Each variable is plotted along an axis . A line graph has a vertical axis and a horizontal axis. So, for example, if you wanted to graph the cost of tuition over time, you could put time along the horizontal, or x-axis, and tuition cost along the vertical, or y-axis.**

Main effect **- The effect of an independent variable on a dependent variable often explored after a regression analysis or ANOVA was performed.**

Marginal mean**- In a design with two factors, the marginal means for one factor are the means for that factor averaged across all levels of the other factor.**

Mean **- The sum of a set of values divided by the total number of values, which is also known as arithmetic average.**

Median **- This figure is the value that separates the higher half of a sample from the lower half. The valid data is sorted in ascending order, and if there is an odd number of data points, the median is the middle number; however if there is an even number of data points, the median is the average of the middle two numbers.**

Measure **- quantitative information that can be communicated by a set of scores.**

Mode **- The number or value that appears most frequently in a distribution of numbers. There may be multiple modes.**

Nominal data **- data where the values assigned to responses are mutually exclusive, but the values have no order. Gender is an example of nominal data – males can be assigned the value 1 and females the value 2 or vice versa and it would not impact the analysis results or interpretation.**

Normally distributed **- Quantitative data that when graphed resembles a bell-shaped curve. The data is symmetrically clustered around the mean so that the mean, median, and mode are approximately the same, and 95% of the sample is within two standard deviations below and above the mean.**

Ordinal data **- data where the numbering of the responses indicates the relative order but does not indicate the absolute strength/value of the responses. For example, class level – the coding of freshman, sophomore, junior, and senior from 1 to 4 indicates relative rank but the absolute difference between the ranks may not have the same meaning. Simple arithmetic operations are not meaningfully applied to ordinal data.**

Pie chart **- a graphic representation of quantitative information by means of a circle divided into sectors, in which the relative sizes of the areas (or central angles) of the sectors correspond to the relative sizes or proportions of the quantities.**

Population **- The entire group of individuals from which a sample may be selected.**

Glossary

Quartile/Percentile **- These figures represent the range of data broken down by percentiles. The lower quartile is the 25th percentile where 75% of the scores are above this number; the middle quartile is the median; the highest quartile is the 75th percentile where 25% of the scores are above this number.**

Range **- The range is a measure of data dispersion. It is the distance between the lowest number and the highest number in a distribution of numbers. For example, if the lowest person scored 50 on a test and the highest person scored 95, the range is said to be from 50 to 95.**

Sample **- A subset of participants from the population of interest from which data is collected.**

Sample Size ("N") **- The total sample size represents the number of people who were in the sample or were asked a question.**

Standard deviation **- The standard deviation is a measure of dispersion that describes the average distance from the mean in a distribution of data. A distribution that has a relatively small standard deviation is associated with less variability among the data, whereas a distribution that has a relatively large standard deviation is associated with more variability among the data. Stated differently, the numbers in a distribution with a relatively small standard deviation are clustered more closely around the mean than numbers in a distribution with a relatively large standard deviation.**

Statistical significance **- A statistical test to determine the probability that the observed relationship between variables or difference between means in a sample occurred by chance, and that the observed result is actually representative of the population. The test statistic that represents statistical significance is the p value. A lower p value indicates that there is a smaller probability that the resulting relationship or difference was due to chance. For example $p < .05$ indicates that there is a less than five percent chance that the observed result was due to error, but $p < .01$ indicates that there is a less than one percent chance that the observed result was due to error. See** http://www.statsoft.com/textbook/elementary-concepts-in-statistics/.

Within groups**: design where a respondent's responses are compared to themselves, either are more than one point in time (pre/post), across survey questions, or across other measures.** ***Comparing me to me*****.**

95% confidence interval **- Confidence intervals with a 95% confidence level are most common and indicate that 95% of samples would contain the statistic if hundreds of samples were randomly drawn from the population.**

2x2 design **– comparing 2 variables where each variable has 2 groups. For example, compare gender and under/upperclassman – the design is (male or female) compared to (under classman or upper classman)**