Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study

Scot H Simpson

There are three kinds of lies: lies, damned lies, and statistics.
– Mark Twain[1]

INTRODUCTION

Statistics represent an essential part of a study because, regardless of the study design, investigators need to summarize the collected information for interpretation and presentation to others. It is therefore important for us to heed Mr Twain's concern when creating the data analysis plan. In fact, even before data collection begins, we need to have a clear analysis plan that will guide us from the initial stages of summarizing and describing the data through to testing our hypotheses.

The purpose of this article is to help you create a data analysis plan for a quantitative study. For those interested in conducting qualitative research, previous articles in this Research Primer series have provided information on the design and analysis of such studies.[2,3] Information in the current article is divided into 3 main sections: an overview of terms and concepts used in data analysis, a review of common methods used to summarize study data, and a process to help identify relevant statistical tests. My intention here is to introduce the main elements of data analysis and provide a place for you to start when planning this part of your study. Biostatistical experts, textbooks, statistical software packages, and other resources can certainly add more breadth and depth to this topic when you need additional information and advice.

TERMS AND CONCEPTS USED IN DATA ANALYSIS

When analyzing information from a quantitative study, we are often dealing with numbers; therefore, it is important to begin with an understanding of the source of the numbers. Let us start with the term *variable*, which defines a specific item of information collected in a study. Examples of variables include age, sex or gender, ethnicity, exercise frequency, weight, treatment group, and blood glucose. Each variable will have a group of categories, which are referred to as *values*, to help describe the characteristic of an individual study participant. For example, the variable "sex" would have values of "male" and "female".

Although variables can be defined or grouped in various ways, I will focus on 2 methods at this introductory stage. First, variables can be defined according to the level of measurement. The categories in a *nominal* variable are names, for example, male and female for the variable "sex"; white, Aboriginal, black, Latin American, South Asian, and East Asian for the variable "ethnicity"; and intervention and control for the variable "treatment group". Nominal variables with only 2 categories are also referred to as *dichotomous* variables because the study group can be divided into 2 subgroups based on information in the variable. For example, a study sample can be split into 2 groups (patients receiving the intervention and controls) using the dichotomous variable "treatment group". An *ordinal* variable implies that the categories can be placed in a meaningful order, as would be the case for exercise frequency (never, sometimes, often, or always). Nominal-level and ordinal-level variables are also referred to as *categorical* variables, because each category in the variable can be completely separated from the others. The categories for an *interval* variable can be placed in a meaningful order, with the interval between consecutive categories also having meaning. Age, weight, and blood glucose can be considered as interval variables, but also as *ratio* variables, because the ratio between values has meaning (e.g., a 15-year-old is half the age of a 30-year-old). Interval-level and ratio-level variables are also referred to as *continuous* variables because of the underlying continuity among categories.

As we progress through the levels of measurement from nominal to ratio variables, we gather more information about the study participant. The amount of information that a variable provides will become important in the analysis stage, because we lose information when variables are reduced or aggregated—a common practice that is not recommended.[4] For example, if age is reduced from a ratio-level variable (measured in years) to an ordinal variable (categories of < 65 and ≥ 65 years) we lose the ability to make comparisons across the entire age range and introduce error into the data analysis.[4]

A second method of defining variables is to consider them as either dependent or independent. As the terms imply, the value of a *dependent* variable depends on the value of other variables, whereas the value of an *independent* variable does not rely on other variables. In addition, an investigator can influence the value of an independent variable, such as treatment-group assignment. Independent variables are also referred to as *predictors* because we can use information from these variables to predict the value of a dependent variable. Building on the group of variables listed in the first paragraph of this section, blood glucose could be considered a dependent variable, because its value may depend on values of the independent variables age, sex, ethnicity, exercise frequency, weight, and treatment group.

Statistics are mathematical formulae that are used to organize and interpret the information that is collected through variables. There are 2 general categories of statistics, descriptive and inferential. *Descriptive* statistics are used to describe the collected information, such as the range of values, their average, and the most common category. Knowledge gained from descriptive statistics helps investigators learn more about the study sample. *Inferential* statistics are used to make comparisons and draw conclusions from the study data. Knowledge gained from inferential statistics allows investigators to make inferences and generalize beyond their study sample to other groups.

Before we move on to specific descriptive and inferential statistics, there are 2 more definitions to review. *Parametric* statistics are generally used when values in an interval-level or ratio-level variable are normally distributed (i.e., the entire group of values has a bell-shaped curve when plotted by frequency). These statistics are used because we can define parameters of the data, such as the centre and width of the normally distributed curve. In contrast, interval-level and ratio-level variables with values that are not normally distributed, as well as nominal-level and ordinal-level variables, are generally analyzed using *nonparametric* statistics.

METHODS FOR SUMMARIZING STUDY DATA: DESCRIPTIVE STATISTICS

The first step in a data analysis plan is to describe the data collected in the study. This can be done using figures to give a visual presentation of the data and statistics to generate numeric descriptions of the data.

Selection of an appropriate figure to represent a particular set of data depends on the measurement level of the variable. Data for nominal-level and ordinal-level variables may be interpreted using a *pie graph* or *bar graph*. Both options allow us to examine the relative number of participants within each category (by reporting the percentages within each category), whereas a bar graph can also be used to examine absolute numbers. For example, we could create a pie graph to illustrate the proportions of men and women in a study sample and a bar graph to illustrate the number of people who report exercising at each level of frequency (never, sometimes, often, or always).

Interval-level and ratio-level variables may also be interpreted using a pie graph or bar graph; however, these types of variables often have too many categories for such graphs to provide meaningful information. Instead, these variables may be better interpreted using a *histogram*. Unlike a bar graph, which displays the frequency for each distinct category, a histogram displays the frequency within a range of continuous categories. Information from this type of figure allows us to determine whether the data are normally distributed. In addition to pie graphs, bar graphs, and histograms, many other types of figures are available for the visual representation of data. Interested readers can find additional types of figures in the books recommended in the "Further Readings" section.

Figures are also useful for visualizing comparisons between variables or between subgroups within a variable (for example, the distribution of blood glucose according to sex). *Box plots* are useful for summarizing information for a variable that does not follow a normal distribution. The lower and upper limits of the box identify the interquartile range (or 25th and 75th percentiles), while the midline indicates the median value (or 50th percentile). *Scatter plots* provide information on how the categories for one continuous variable relate to categories in a second variable; they are often helpful in the analysis of correlations.

In addition to using figures to present a visual description of the data, investigators can use statistics to provide a numeric description. Regardless of the measurement level, we can find the *mode* by identifying the most frequent category within a variable. When summarizing nominal-level and ordinal-level variables, the simplest method is to report the proportion of participants within each category.

The choice of the most appropriate descriptive statistic for interval-level and ratio-level variables will depend on how the values are distributed. If the values are normally distributed, we can summarize the information using the parametric statistics of mean and standard deviation. The *mean* is the arithmetic average of all values within the variable, and the *standard deviation* tells us how widely the values are dispersed around the mean. When values of interval-level and ratio-level variables are not normally distributed, or we are summarizing information from an ordinal-level variable, it may be more appropriate to use the nonparametric statistics of median and range. The first step in identifying these descriptive statistics is to arrange study participants according to the variable categories from lowest value to highest value. The *range* is used to report the lowest and highest values. The *median* or 50th percentile is located by dividing the number of participants into 2 groups, such that half (50%) of the participants have values above the median and the other half (50%) have values below the median. Similarly, the 25th percentile is the value with 25% of the participants having values below and 75% of the participants having values above, and the 75th percentile is the value with 75% of participants having values below and 25% of participants having values above. Together, the 25th and 75th percentiles define the *interquartile range*.

PROCESS TO IDENTIFY RELEVANT STATISTICAL TESTS: INFERENTIAL STATISTICS

One caveat about the information provided in this section: selecting the most appropriate inferential statistic for a specific study should be a combination of following these suggestions, seeking advice from experts, and discussing with your co-investigators. My intention here is to give you a place to start a conversation with your colleagues about the options available as you develop your data analysis plan.

There are 3 key questions to consider when selecting an appropriate inferential statistic for a study: What is the research question? What is the study design? and What is the level of measurement? It is important for investigators to carefully consider these questions when developing the study protocol and creating the analysis plan. The figures that accompany these questions show decision trees that will help you to narrow down the list of inferential statistics that would be relevant to a particular study. Appendix 1 provides brief definitions of the inferential statistics named in these figures. Additional information, such as the formulae for various inferential statistics, can be obtained from textbooks, statistical software packages, and biostatisticians.

What Is the Research Question?

The first step in identifying relevant inferential statistics for a study is to consider the type of research question being asked. You can find more details about the different types of research questions in a previous article in this Research Primer series that covered questions and hypotheses.[5] A *relational* question seeks information about the relationship among variables; in this situation, investigators will be interested in determining whether there is an *association* (Figure 1). A *causal* question seeks information about the effect of an intervention on an outcome; in this situation, the investigator will be interested in determining whether there is a *difference* (Figure 2).

What Is the Study Design?

When considering a question of association, investigators will be interested in measuring the relationship between variables (Figure 1). A study designed to determine whether there is consensus among different raters will be measuring *agreement.* For example, an investigator may be interested in determining

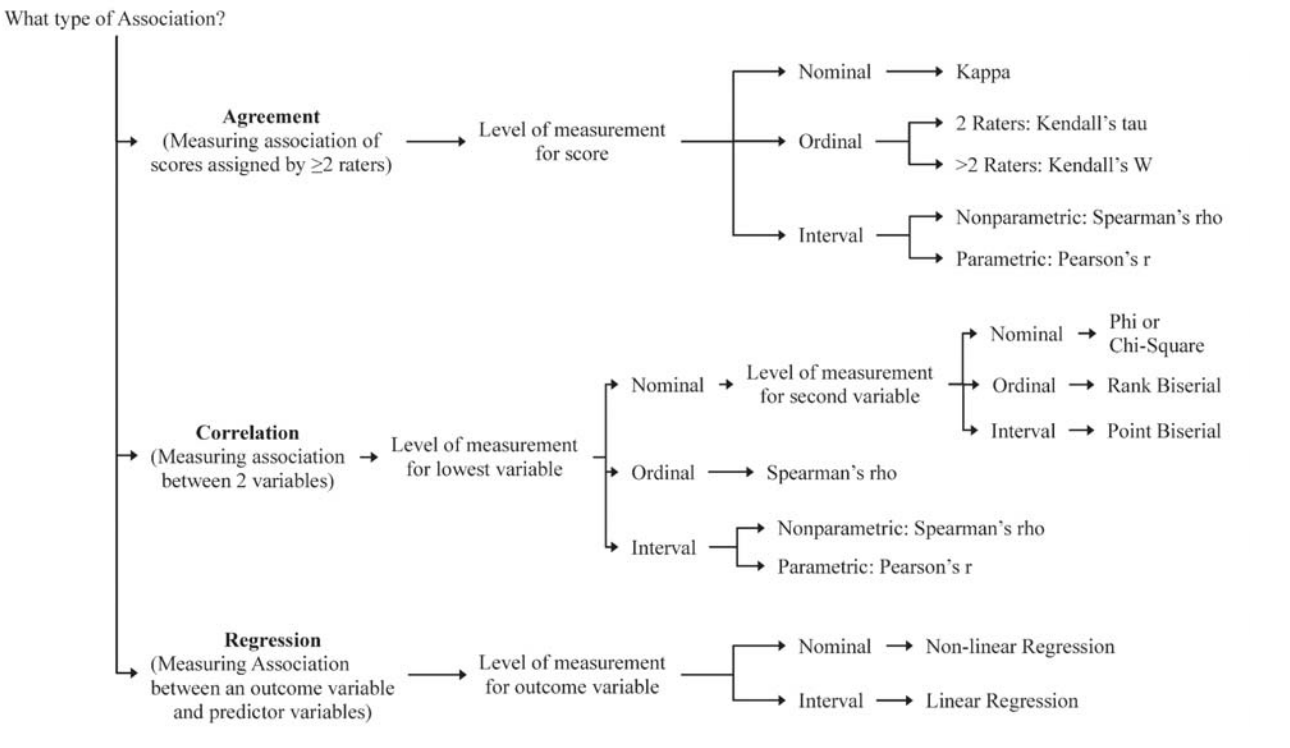
What type of Association?
Agreement
(Measuring association of scores assigned by ≥2 raters)
Level of measurement for score
Nominal
Kappa
Ordinal
2 Raters: Kendall's tau
>2 Raters: Kendall's W
Interval
Nonparametric: Spearman's rho
Parametric: Pearson's r
Correlation
(Measuring association between 2 variables)
Level of measurement for lowest variable
Nominal
Level of measurement for second variable
Nominal
Phi or Chi-Square
Ordinal
Rank Biserial
Interval
Point Biserial
Ordinal
Spearman's rho
Interval
Nonparametric: Spearman's rho
Parametric: Pearson's r
Regression
(Measuring Association between an outcome variable and predictor variables)
Level of measurement for outcome variable
Nominal
Non-linear Regression
Interval
Linear Regression

Figure 1. Decision tree to identify inferential statistics for an association.

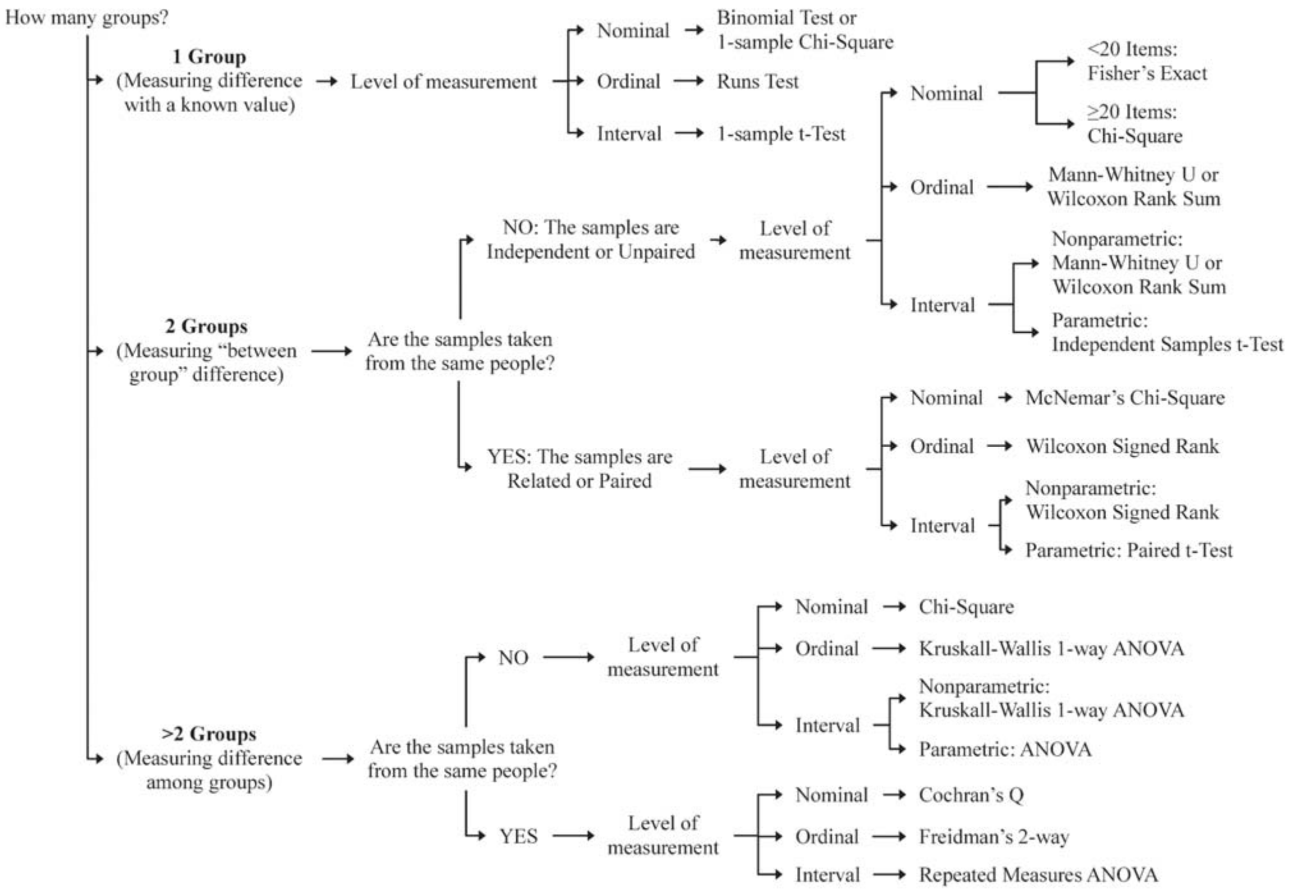
How many groups?
1 Group
(Measuring difference with a known value)
Level of measurement
Nominal
Binomial Test or 1-sample Chi-Square
Ordinal
Runs Test
Interval
1-sample t-Test
2 Groups
(Measuring "between group" difference)
Are the samples taken from the same people?
NO: The samples are Independent or Unpaired
Level of measurement
Nominal
<20 Items: Fisher's Exact
≥20 Items: Chi-Square
Ordinal
Mann-Whitney U or Wilcoxon Rank Sum
Interval
Nonparametric: Mann-Whitney U or Wilcoxon Rank Sum
Parametric: Independent Samples t-Test
YES: The samples are Related or Paired
Level of measurement
Nominal
McNemar's Chi-Square
Ordinal
Wilcoxon Signed Rank
Interval
Nonparametric: Wilcoxon Signed Rank
Parametric: Paired t-Test
>2 Groups
(Measuring difference among groups)
Are the samples taken from the same people?
NO
Level of measurement
Nominal
Chi-Square
Ordinal
Kruskall-Wallis 1-way ANOVA
Interval
Nonparametric: Kruskall-Wallis 1-way ANOVA
Parametric: ANOVA
YES
Level of measurement
Nominal
Cochran's Q
Ordinal
Freidman's 2-way
Interval
Repeated Measures ANOVA

Figure 2. Decision tree to identify inferential statistics for measuring a difference.

whether 2 raters, using the same assessment tool, arrive at the same score. *Correlation* analyses examine the strength of a relationship or connection between 2 variables, like age and blood glucose. *Regression* analyses also examine the strength of a relationship or connection; however, in this type of analysis, one variable is considered an outcome (or dependent variable) and the other variable is considered a predictor (or independent variable). Regression analyses often consider the influence of multiple predictors on an outcome at the same time. For example, an investigator may be interested in examining the association between a treatment and blood glucose, while also considering other factors, like age, sex, ethnicity, exercise frequency, and weight.

When considering a question of difference, investigators must first determine how many groups they will be comparing. In some cases, investigators may be interested in comparing the characteristic of one group with that of an external reference group. For example, is the mean age of study participants similar to the mean age of all people in the target group? If more than one group is involved, then investigators must also determine whether there is an underlying connection between the sets of values (or *samples*) to be compared. Samples are considered *independent* or *unpaired* when the information is taken from different groups. For example, we could use an unpaired *t* test to compare the mean age between 2 independent samples, such as the intervention and control groups in a study. Samples are considered *related* or *paired* if the information is taken from the same group of people, for example, measurement of blood glucose at the beginning and end of a study. Because blood glucose is measured in the same people at both time points, we could use a paired *t* test to determine whether there has been a significant change in blood glucose.

What Is the Level of Measurement?

As described in the first section of this article, variables can be grouped according to the level of measurement (nominal, ordinal, or interval). In most cases, the independent variable in an inferential statistic will be nominal; therefore, investigators need to know the level of measurement for the dependent variable before they can select the relevant inferential statistic. Two exceptions to this consideration are correlation analyses and regression analyses (Figure 1). Because a correlation analysis measures the strength of association between 2 variables, we need to consider the level of measurement for both variables. Regression analyses can consider multiple independent variables, often with a variety of measurement levels. However, for these analyses, investigators still need to consider the level of measurement for the dependent variable.

Selection of inferential statistics to test interval-level variables must include consideration of how the data are distributed. An underlying assumption for parametric tests is that the data approximate a normal distribution. When the data are not normally distributed, information derived from a parametric test may be wrong.[6] When the assumption of normality is violated (for example, when the data are skewed), then investigators should use a nonparametric test. If the data are normally distributed, then investigators can use a parametric test.

ADDITIONAL CONSIDERATIONS

What Is the Level of Significance?

An inferential statistic is used to calculate a *p* value, the probability of obtaining the observed data by chance. Investigators can then compare this *p* value against a prespecified level of significance, which is often chosen to be 0.05. This level of significance represents a 1 in 20 chance that the observation is wrong, which is considered an acceptable level of error.

What Are the Most Commonly Used Statistics?

In 1983, Emerson and Colditz[7] reported the first review of statistics used in original research articles published in the *New England Journal of Medicine*. This review of statistics used in the journal was updated in 1989 and 2005,[8] and this type of analysis has been replicated in many other journals.[9-13] Collectively, these reviews have identified 2 important observations. First, the overall sophistication of statistical methodology used and reported in studies has grown over time, with survival analyses and multivariable regression analyses becoming much more common. The second observation is that, despite this trend, 1 in 4 articles describe no statistical methods or report only simple descriptive statistics. When inferential statistics are used, the most common are *t* tests, contingency table tests (for example, χ^2 test and Fisher exact test), and simple correlation and regression analyses. This information is important for educators, investigators, reviewers, and readers because it suggests that a good foundational knowledge of descriptive statistics and common inferential statistics will enable us to correctly evaluate the majority of research articles.[11-13] However, to fully take advantage of all research published in high-impact journals, we need to become acquainted with some of the more complex methods, such as multivariable regression analyses.[8,13]

What Are Some Additional Resources?

As an investigator and Associate Editor with *CJHP*, I have often relied on the advice of colleagues to help create my own analysis plans and review the plans of others. Biostatisticians have a wealth of knowledge in the field of statistical analysis and can provide advice on the correct selection, application, and interpretation of these methods. Colleagues who have "been there and done that" with their own data analysis plans are also valuable sources of information. Identify these individuals and consult with them early and often as you develop your analysis plan.

Another important resource to consider when creating your analysis plan is textbooks. Numerous statistical textbooks are available, differing in levels of complexity and scope. The titles listed in the "Further Reading" section are just a few suggestions. I encourage interested readers to look through these and other books to find resources that best fit their needs. However, one crucial book that I highly recommend to anyone wanting to be an investigator or peer reviewer is Lang and Secic's *How to Report Statistics in Medicine* (see "Further Reading"). As the title implies, this book covers a wide range of statistics used in medical research and provides numerous examples of how to correctly report the results.

CONCLUSIONS

When it comes to creating an analysis plan for your project, I recommend following the sage advice of Douglas Adams in *The Hitchhiker's Guide to the Galaxy*: Don't panic![14] Begin with simple methods to summarize and visualize your data, then use the key questions and decision trees provided in this article to identify relevant statistical tests. Information in this article will give you and your co-investigators a place to start discussing the elements necessary for developing an analysis plan. But do not stop there! Use advice from biostatisticians and more experienced colleagues, as well as information in textbooks, to help create your analysis plan and choose the most appropriate statistics for your study. Making careful, informed decisions about the statistics to use in your study should reduce the risk of confirming Mr Twain's concern.

References

1. Twain M (Kiskis MJ, editor). *Mark Twain's own autobiography: the chapters from the North American review.* 2nd ed. Madison (WI): University of Wisconsin Press; 2010. 318 p.
2. Austin Z, Sutton J. Qualitative research: getting started. *Can J Hosp Pharm.* 2014;67(6):436-40.
3. Sutton J, Austin Z. Qualitative research: data collection, analysis, and management. *Can J Hosp Pharm*. 2015;68(3):226-31.
4. Dawson NV, Weiss R. Dichotomizing continuous variables in statistical analysis: a practice to avoid. *Med Decis Making.* 2012;32(2):225-6.
5. Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. *Can J Hosp Pharm*. 2014;67(1):31-4.
6. Harwell MR. Choosing between parametric and nonparametric tests. *J Couns Dev.* 1988;67(1):35-8.
7. Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med.* 1983;309(12):709-13.
8. Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med.* 2005;353(18):1977-9.
9. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 1. Hypothesis testing. *CMAJ.* 1995;152(1):27-32.

10. Goldin J, Zhu W, Sayre JW. A review of the statistical analysis used in papers published in *Clinical Radiology* and *British Journal of Radiology*. *Clin Radiol.* 1996;51(1):47-50.
11. Reed JF 3rd, Salen P, Bagher P. Methodological and statistical techniques: what do residents really need to know about statistics? *J Med Syst.* 2003;27(3):233-8.
12. Hellems MA, Gurka MJ, Hayden GF. Statistical literacy for readers of *Pediatrics*: a moving target. *Pediatrics.* 2007;119(6):1083-8.
13. Taback N, Krzyzanowska MK. A survey of abstracts of high-impact clinical journals indicated most statistical methods presented are summary statistics. *J Clin Epidemiol.* 2008;61(3):277-81.
14. Adams D. *The hitchhiker's guide to the galaxy.* London (UK): Pan Books; 1979.

Further Reading

Devor J, Peck R. *Statistics: the exploration and analysis of data*. 7th ed. Boston (MA): Brooks/Cole Cengage Learning; 2012.

Lang TA, Secic M. *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers.* 2nd ed. Philadelphia (PA): American College of Physicians; 2006.

Mendenhall W, Beaver RJ, Beaver BM. *Introduction to probability and statistics.* 13th ed. Belmont (CA): Brooks/Cole Cengage Learning; 2009.

Norman GR, Streiner DL. *PDQ statistics.* 3rd ed. Hamilton (ON): B.C. Decker; 2003.

Plichta SB, Kelvin E. *Munro's statistical methods for health care research.* 6th ed. Philadelphia (PA): Wolters Kluwer Health/Lippincott, Williams & Wilkins; 2013.

Scot H Simpson, BSP, PharmD, MSc, is Professor and Associate Dean, Research and Graduate Studies, Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta. He is also an Associate Editor with the *CJHP*.

Competing interests: None declared.

Address correspondence to:
Scot H Simpson
Faculty of Pharmacy and Pharmaceutical Sciences
3126 Dentistry/Pharmacy
University of Alberta
Edmonton AB T6G 2N8

e-mail: scot@ualberta.ca

Appendix 1. Glossary of statistical terms* (part 1 of 2)

ANOVA (analysis of variance): Parametric statistic used to compare the means of 3 or more groups that are defined by 1 or more variables.

- 1-way ANOVA: Uses 1 variable to define the groups for comparing means. This is similar to the Student *t* test when comparing the means of 2 groups.
- Kruskall–Wallis 1-way ANOVA: Nonparametric alternative for the 1-way ANOVA. Used to determine the difference in medians between 3 or more groups.
- *n*-way ANOVA: Uses 2 or more variables to define groups when comparing means. Also called a “between-subjects factorial ANOVA”.
- Repeated-measures ANOVA: A method for analyzing whether the means of 3 or more measures from the same group of participants are different.
- Freidman ANOVA: Nonparametric alternative for the repeated-measures ANOVA. It is often used to compare rankings and preferences that are measured 3 or more times.

Binomial test: Used to determine whether the observed proportion is significantly different from a known or hypothesized proportion. The variable is dichotomous (nominal-level data with 2 options).

Biserial correlation (rank or point): Correlation technique when one of the variables is dichotomous (or measured at the nominal level).

Chi-square (χ^2) test: Nonparametric test used to determine whether a statistically significant association exists between rows and columns in a contingency table.

- Fisher exact: Variation of chi-square that accounts for cell counts < 5.
- McNemar: Variation of chi-square that tests statistical significance of changes in 2 paired measurements of dichotomous variables.
- Cochran Q: An extension of the McNemar test that provides a method for testing for differences between 3 or more matched sets of frequencies or proportions. Often used as a measure of heterogeneity in meta-analyses.

Descriptive statistics: Numeric or graphic summaries (or descriptions) of a variable.

Inferential statistics: Measures the difference between 2 variables or subgroups of a variable. Allows the investigator to make inferences about another group on the basis of information generated from the study data.

Kappa (κ): Measures the degree of nonrandom agreement between observers or measurements for the same nominal-level variable.

Kendall tau (τ): Nonparametric alternative for the Spearman correlation. Used when measuring the relationship between 2 ranked (or ordinal-level data) variables.

Mann–Whitney *U* test: Nonparametric alternative for the independent *t* test. One variable is dichotomous (e.g., group A versus group B) and the other variable is either ordinal or interval.

Pearson correlation: Parametric test used to determine whether an association exists between 2 variables measured at the interval or ratio level.

Phi (ϕ): Used when both variables in a correlation analysis are dichotomous.

Runs test: Used to determine whether a series of data occurs from a random process.

Spearman rank correlation: Nonparametric alternative for the Pearson correlation coefficient. Used when the assumptions for Pearson correlation are violated (e.g., data are not normally distributed) or one of the variables is measured at the ordinal level.

***t* test:** Parametric statistical test for comparing the means of 2 independent groups.

- 1-sample: Used to determine whether the mean of a sample is significantly different from a known or hypothesized value.

continued on page 317

Appendix 1. Glossary of statistical terms* (part 2 of 2)

- Independent-samples *t* test (also referred to as the Student *t* test): Used when the independent variable is a nominal-level variable that identifies 2 groups and the dependent variable is an interval-level variable.
- Paired: Used to compare 2 pairs of scores between 2 groups (e.g., baseline and follow-up blood pressure in the intervention and control groups).

Wilcoxon rank–sum test: Nonparametric alternative to the independent *t* test based solely on the order in which observations from the 2 samples fall. Similar to the Mann–Whitney *U* test.

Wilcoxon signed-rank test: Nonparametric alternative to the paired *t* test. The differences between matched pairs are computed and ranked. This test compares the sum of the negative differences and the sum of the positive differences.

**Sources*

Lang TA, Secic M. *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers.* 2nd ed. Philadelphia (PA): American College of Physicians; 2006.

Norman GR, Streiner DL. *PDQ statistics.* 3rd ed. Hamilton (ON): B.C. Decker; 2003.

Plichta SB, Kelvin E. *Munro's statistical methods for health care research*. 6th ed. Philadelphia (PA): Wolters Kluwer Health/ Lippincott, Williams & Wilkins; 2013.

This article is the 12th in the *CJHP* Research Primer Series, an initiative of the *CJHP* Editorial Board and the CSHP Research Committee. The planned 2-year series is intended to appeal to relatively inexperienced researchers, with the goal of building research capacity among practising pharmacists. The articles, presenting simple but rigorous guidance to encourage and support novice researchers, are being solicited from authors with appropriate expertise.

Previous articles in this series:

Bond CM. The research jigsaw: how to get started. *Can J Hosp Pharm*. 2014;67(1):28-30.

Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. *Can J Hosp Pharm*. 2014;67(1):31-4.

Loewen P. Ethical issues in pharmacy practice research: an introductory guide. *Can J Hosp Pharm.* 2014;67(2):133-7.

Tsuyuki RT. Designing pharmacy practice research trials. *Can J Hosp Pharm*. 2014;67(3):226-9.

Bresee LC. An introduction to developing surveys for pharmacy practice research. *Can J Hosp Pharm*. 2014;67(4):286-91.

Gamble JM. An introduction to the fundamentals of cohort and case–control studies. *Can J Hosp Pharm*. 2014;67(5):366-72.

Austin Z, Sutton J. Qualitative research: getting started. *Can J Hosp Pharm*. 2014;67(6):436-40.

Houle S. An introduction to the fundamentals of randomized controlled trials in pharmacy research. *Can J Hosp Pharm*. 2014; 68(1):28-32.

Charrois TL. Systematic reviews: What do you need to know to get started? *Can J Hosp Pharm*. 2014;68(2):144-8.

Sutton J, Austin Z. Qualitative research: data collection, analysis, and management. *Can J Hosp Pharm*. 2014;68(3):226-31.

Cadarette SM, Wong L. An introduction to health care administrative data. *Can J Hosp Pharm.* 2014;68(3):232-7.