International Journal of Faith Community Nursing

Volume 2
Issue 3 *Special Research Issue*

Article 2

December 2016

Common Statistical Tests and Interpretation in Nursing Research

Maria E. Main
Western Kentucky University

Veletta L. Ogaz
Western Kentucky University

Common Statistical Tests and Interpretation in Nursing Research

Evidence-based practice (EBP) requires nurses to appraise current literature for delivering quality health care to patients (Melnyk & Fineout-Overholt, 2015). In 2014, the Faith Community Nurse Network made an organizational commitment to strengthen research infrastructure, create partnerships with nursing programs, and increase awareness of EBP among faith community nurses (Wessman, 2014). As part of meeting this goal, faith community nurses need a basic understanding of common statistical tests and their interpretation to aid in the appraisal of nursing research. The purpose of this article is to review basic statistical concepts, define common statistical tests, and interpret the results of common statistical tests.

Basic Statistical Concepts and Application

What is the Research Design?

The research design of the study guides the choice of an appropriate statistical test. Is the research design descriptive in nature, comparing differences in groups, or examining relationships among variables? Common statistical tests that measure differences in groups are independent samples *t*-test, paired sample *t*-tests, and analysis of variance. Two common statistical tests that measure relationships are the Pearson product moment correlation and chi-square.

The statistical analysis of research includes both descriptive and inferential statistics. Descriptive statistics are used to summarize and organize data including measures of central tendency and measures of variability. Inferential statistics are utilized to infer results on target populations from a sample population. Hypotheses' testing is used in this inferential process and the null hypotheses are there is no difference in the groups or there is no relationship in the variables. The alternative hypotheses are there is a difference in the groups or a relationship in variables (Gravetter & Wallnau, 2012). For example, the researcher desires to know if patient education on Type II diabetes makes a difference in patients' knowledge of Type II diabetes. In this instance, the null hypothesis is patient education does not change the knowledge level of the participants. The alternative hypothesis is that patients' knowledge of Type II diabetes improves following patient education.

Each inferential statistical test is limited in use to specific levels of measurement for independent and/or dependent variables. Two broad levels of measurement are categorical data including nominal and ordinal data and continuous data including interval and ratio data (Gravetter & Wallnau, 2012). Each level of measurement has attributes that are common to the level of measurement and contribute to the designation of categorical or continuous data (see Table 1).

Table 1
Categorical and Continuous Summary of Attributes of the Levels of Measurement

Level of Measurement	Categorical or Continuous	Exhaustive and Exclusive	Rank Ordered	Equal Intervals	Absolute Zero
Nominal	Categorical	X			
Ordinal	Categorical	X	X		
Interval	Continuous	X	X	X	
Ratio	Continuous	X	X	X	X

***P*-value and Effect Size**

A *p*-value or significance level indicates the probability that a result is obtained by chance. In nursing research, the most common significance levels are 0.05 or 0.01, which indicate a 5% or 1% chance, respectively of rejecting the null hypothesis when it is true. A smaller *p*-value of .01 as compared to a *p*-value of .05 will decrease the chances of rejecting the null hypothesis when it is true. When a *p*-value is less than or equal to the significance level designated by the researcher should have rejected the null hypothesis and reported a difference in the groups or a relationship among the variables (Gravetter & Wallnau, 2012).

While a significant *p*-value indicates statistical significance, effect size denotes the relative magnitude of the differences or the relationship (Gravetter & Wallnau, 2012). Effect size is more useful for clinical practice as it indicates clinical significance and importance. There are many different measures of effect size, which correspond to the statistical test utilized (Cumming, 2012). Effect size calculators are available online and the reader may calculate effect sizes if the researcher did not calculate the value.

Common Statistics that Compare Groups

Independent Samples *t*-test

The independent samples *t*-test can be employed when comparing two independent groups on a continuous dependent variable. As the name of the test indicates, the groups must be independent with different participants in each group and the dependent variable must be continuous (Gravetter & Wallnau, 2012). The researcher asks in a hypothetical study of caregivers in community dwelling older adults, the researcher asks is there a significant difference in the mean depression scores for employed and unemployed caregivers? The dependent variable is depression scores and the researcher designated the significance level of $p < .01$. Differences were noted in mean depression scores of employed caregivers (M =13.42, SD =10.3) and unemployed caregivers (M =19.61, SD = 10.3) with a value of t (108) = 3.53, p =.007. The significance (*p*-value) of 0.007 indicates the difference in the mean depressions scores between the employed and unemployed caregivers is statistically significant. Very small differences between groups in large sample sizes, may be statistically significant but have no clinical importance (Cumming, 2012). While there is statistical significance in the difference of the mean depression scores is it clinically important? The calculation of an effect size from the information the researcher has provided may be completed. A moderate effect size of .60 is noted and is clinically important as unemployed caregivers report higher levels of depression.

Paired Samples *t*-test

A paired samples *t*-test compares two sets of data from one group of people on a continuous dependent variable (Gravetter & Wallnau, 2012). Commonly, this statistical test is utilized pre and post-intervention. The researcher asks is there a significant change in participants' knowledge of cardiovascular disease following a cardiovascular disease educational intervention. A statistically significant increase from pretest knowledge (M = 15.60, SD = 2.55) to posttest knowledge (M = 17.10, SD = 3.55) is found (***t*** (55) = -5.837, p = .0001). A p-value of .0001 indicates statistical significance; however, small effect size of .162 is noted indicating minor clinical significance.

One-Way Analysis of Variance (ANOVA)

A one-way analysis of variance (ANOVA) compares two or more independent groups or conditions to investigate the presence of differences between groups on a continuous variable (Gravetter & Wallnau, 2012). The researcher asks in adults with chronic back pain is there a difference in mean pain scores across three treatment groups of: (a) yoga and standard exercise, (b) meditation and standard exercise, and (c) standard exercise only. The researcher designated a significance level of $p<.05$. In this example there is a statistically significant difference in mean pain scores for three treatment groups: ***F*** (2, 432) = 4.36, p = .002. The next question would be which group(s) are different? A post-hoc comparison indicated that the difference in the mean scores is between the (a) yoga and standard exercise group and (c) the standard exercise group only. Although the results reached statistical significance, the difference in the mean scores is small. The effect size is .02 indicating a small effect size.

Common Statistical Tests that Measure Relationships

Pearson Product-Moment Correlation

The Pearson product-moment correlation coefficient measures the relationship between two continuous variables. The Pearson product moment correlation is reported as the statistic ***r*** and ranges from -1.00 to 1.00. Positive r values indicate that as the value of one variable (x) increases the associated values of the corresponding value (y) increase. Negative r values indicate that there is a an inverse (negative) relationship as the value of one x increases the value of y decreases. The strength of the relationship can be described in accordance with the r value, weak (± .00 to <.30), moderate (± .30 to .50), and strong (> ± .50) (Gravetter & Wallnau, 2012). In a hypothetical example, the researcher examines the relationship in primigravida women between postpartum depression and maternal perception of fatigue at four weeks postpartum. The continuous variables examined are depression scores and maternal fatigue scores. Maternal perception of fatigue is strongly related to depression (r = .58, p < .001). The results of a statistically significant correlation can be examined for clinical significance by calculating the coefficient of determination

(r^2) resulting in an effect size of .34 which is considered a moderate effect size (Gravetter & Wallnau, 2012).

Chi-Square Test

A chi-square test (χ^2) is a statistical test that examines the relationship in variables measured at the categorical level. The χ^2 test compares the frequency of data observed with the expected frequencies of the data expected if there is no relationship between the variables resulting in a Pearson's Chi-Square (Gravetter & Wallnau, 2012). The researcher asks are males more likely to drop out of cardiovascular rehabilitation than females? A chi-square test indicated no significant relationship between gender and dropping out of cardiovascular rehabilitation χ^2 (1, n = 438 = .37, p = .56, *phi* = -.03). In a significant finding, the phi coefficient can indicate effect size; however, in the hypothetical example the findings were not significant.

Conclusion

Evidence-based practice in faith community nursing, like other areas of nursing practice, is vital to the provision of high quality health care. The purpose of this article is to review basic statistical concepts and the use of selected common statistical tests. Particularly important is the ability to examine research for the appropriate statistical test use and interpretation. Knowledge of statistical concepts and common statistical tests assist in the appraisal of nursing research for evidence-based practice.

References

Cumming, G. (2012). *Understanding the new statistics effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Gravetter, F. J. & Wallnau, L. B. (2012). *Statistics for the behavioral health sciences (9th ed).* Belmont, CA: Wadsworth.

Melnyk & Fineout-Overholt (2015). Making the case for evidence-based practice and cultivating a spirt of inquiry. In B. Melnyk & E. Fineout-Overholt (Eds.), *Evidence-based practice in nursing and healthcare* (pp. 3-23).

Wessman, J. P.(2014). Faith community nurse network of the greater twin cities white paper on research emphasis within the network. Retreived from http://www.fcnntc.org/uploads/3/8/0/0/38002891/white_paper_final_2.pdf