Chen et al., Int J Nurs Clin Pract 2023, 10: 383
https://doi.org/10.15344/2394-4978/2023/383

International Journal of
Nursing & Clinical Practices

Research Article Open Access

Common Statistical Concepts and Methods Used in Nursing Research and Scholarly Projects

Yea-Jyh Chen[1], and Ching-Yu Cheng[2*]

[1]*College of Health and Human Services, School of Nursing, University of North Carolina at Wilmington, Wilmington, NC, USA*

[2]*College of Nursing, Chang Gung University of Science and Technology - Chiayi Campus, Taiwan*

Abstract

Analyzing and comprehending statistics is challenging and stressful for most clinical nurses. In current nursing education the integration of statistical curriculum and pedagogical design is limited. This limitation reflects a focus on lecture-textbook learning over applied data analysis and practical exercises. The extent to which nursing students comprehend statistics and their competence in selecting appropriate statistical methods is crucial to conducting research and/or scholarly projects. Statistical methods are essential tools in clinical nursing research to analyze data, draw meaningful conclusions, and establish empirical evidence to inform practice. In this paper, fundamental statistical concepts and their interactions are discussed and highlighted based on common techniques utilized in nursing research: the t-tests, ANOVA, correlation analysis, regression, and nonparametric analysis. A table summarizing each method provides an overview of its purpose, assumptions, data types of dependent and independent variables, sample questions that can be addressed, and important notes when applying these techniques. Careful consideration should be given to the selection and evaluation of statistical methods to prevent flawed results.

Publication History:

Received: November 21, 2023
Accepted: December 07, 2023
Published: December 09, 2023

Keywords:

Nursing research, Parametric/nonparametric test, Statistical concepts, Statistical and Clinical significance

Introduction

Methodological procedures applied to nursing research or scholarly projects involve selecting statistical techniques and analyzing data to answer research questions, evaluate effects of interventions, and draw conclusions from the results. With the growing importance placed on evidence-based nursing and practice, clinical nurses and nursing students must possess a working understanding of statistical methodology and interpretation. However, in real-life practice statistical techniques and applications often fall outside the scope of clinical nursing, causing discomfort or distress to many nurses who seldom use statistics in day-to-day clinical work [1]. Furthermore, lack of statistical knowledge and/or limited availability of statistical courses in current nursing education programs hinders nurses' capacity to perform analytical analysis and evaluate quantitative results effectively [2,3]. Concerns have been raised about how reliable and confident nurses' understanding and interpretation of scientific evidence are absent a foundational basis in statistical concepts and applications despite the well-known principle that nursing care is built upon evidence-based practice.

This paper aims to provide nurses with a basic understanding of commonly used statistical methods to facilitate selection of the right tools for their research goals.

Statistical Concepts and Their Associations

Prior to performing statistical analysis, it is crucial to fully understand several statistical definitions, concepts, and their associations including power and effect size, statistical versus clinical significance, and statistical assumptions for parametric/non-parametric tests. These concepts, which are individually distinct but closely interrelated, are often referred to but may not be clearly understood or applied.

Power and Effect Size

Statistical power and effect size both closely influence the validity of data results and interpretation. Statistical power is the ability of a statistical test to truly detect an effect or difference in a study population and to minimize the risk of Type II errors (false negatives). It essentially quantifies the probability of correctly rejecting a null hypothesis that is false [4]. It is a predetermined value estimated before conducting a project. Often, it requires a large sample size and well-designed intervention protocol to attain a high or optimal power. Conversely, effect size measures the magnitude or strength of a relationship or difference in a statistical analysis. It provides a practical assessment of the significance of an observed effect on the outcome measure regardless of sample size [4,5]. A larger effect size typically leads to higher statistical power, as larger effects make it easier to detect differences/relationships with greater sensitivity.

A larger effect size combined with a high power increases the ability to detect larger differences between groups or stronger associations between variables; that is, it is more likely to show statistical significance. In contrast, a smaller effect size requires a larger sample size to achieve the same level of power meaning that studies with low power are less likely to detect small, but potentially clinically significant, effects. While a study may have sufficient power to detect an effect, a small effect size may have limited practical or clinical significance. Because of the previously mentioned conditions, calculating power and effect size is essential to design studies that yield valid and clinically relevant results in nursing research and clinical projects. A balance between attaining adequate power (≥ 80%) [6] and ensuring that the observed effect size is large enough is expected in implementing interventional research to detect statistical significance and be practically meaningful.

***Corresponding Author:** Dr. Ching-Yu Cheng, College of Nursing, Chang Gung University of Science and Technology - Chiayi Campus, 2, Chiapu Rd, West Sec. Putz, Chiayi, Taiwan; Telephone: +886-53628800 ext. 2603; E-mail: cycheng@mail.cgust.edu.tw

Citation: Chen YJ, Cheng CY (2023) Common Statistical Concepts and Methods Used in Nursing Research and Scholarly Projects. Int J Nurs Clin Pract 10: 383. doi: https://doi.org/10.15344/2394-4978/2023/383

Citation: Chen YJ, Cheng CY (2023) Common Statistical Concepts and Methods Used in Nursing Research and Scholarly Projects. Int J Nurs Clin Pract 10: 383.
doi: https://doi.org/10.15344/2394-4978/2023/383

Statistical versus Clinical Significance

Statistical significance and clinical significance are two distinct concepts in research and scholarly data analysis. Statistical significance is determined by a p-value when comparing the difference of the observed effect or testing a real relationship from the target population-based null hypothesis [7]. Significant results are greatly influenced by statistical power, sample size, pre-set alpha value, and the statistical methods chosen [4,8]. Clinical significance assesses whether an intervention effect is important or relevant in real-world practice [4,7]. Its focus is more on whether statistically significant results have meaningfully impact patient outcomes and are therefore valuable for nursing scholars and providers to inform clinical practice change [9].

Statistical significance is routinely reported while testing quantitative hypotheses whereas clinical significance is not described in parallel. These two types of significance are equally important, especially when the primary objective is to improve individuals' health and well-being. Thus, statistical and clinical significance should be distinguished and interpreted correctly to ensure that test findings are both reliable and meaningful for practical application and not relying solely on p-values [7,8,10].

Interactions between statistical and clinical significance are highlighted below:

- A result with statistical significance indicates that the observed effect is unlikely made by chance, but it does not necessarily imply practical importance.
- With a large sample size statistical significance can be found with a small effect size; however, it may or may not have much clinical relevance.
- Data results with a large effect size may show both statistical and clinical significance.
- Both statistical and clinical significance are essential; a statistically significant result may not be practically significant if a small effect size exists, or meaningful consequences are absent.

Statistical Assumptions for Parametric/Non-parametric Tests

In general, three assumptions need to be met for many statistical tests: normality; homogeneity of variance (or homoscedasticity); and independence (Table 1) [10,11]. Normality indicates that the scores on individual variables are normally distributed with a bell-shaped curve in each of the groups, whereas independence requires data collected from independent groups [10]. When variances of the dependent variable for groups are equal, the assumption of homogeneity of variance is met [4,5]. The distribution of data must be tested before choosing appropriate statistical tests for analysis. Parametric tests are used only when normality is met [10-12]. If the data are not normally distributed, non-parametric tests are selected [10-12]. Several ways can be used to assess the distribution of data including graphical tests (histograms, Q-Q [quantile-quantile] plots), and statistical tests (skewness/kurtosis, Shapiro-Wilk test, and Kolmogorov-Smirnov test) [4,10]. These normality tests are available in most statistical software platforms, e.g., SPSS, SAS, R, JASP, etc. Relying on a single test, such as visualizing a histogram alone, may be insufficient to assess data normality. Performing multiple normality checks is therefore suggested as a more reliable assessment [13], especially when two assessment results are incompatible.

Descriptive and Inferential Statistical Methods

Descriptive versus inferential statistics are two types of statistical foundations. Descriptive statistics are regularly used in both quantitative and qualitative data analyses. It is the must-have, basic element for each research study and scholarly project to articulate and summarize data variables, for instance, sample demographic characteristics. Data description is commonly presented by frequency, percentage, mean, standard deviation, central tendency, and variability. Inferential statistics, also referred to as the methods for hypothesis testing, applies various parametric/non-parametric test techniques to detect the tested outcomes [10,12].

The popular parametric tests include t-tests, analysis of variance (ANOVA), correlation analysis, and predictive regression. Non-parametric tests are distribution-free such as the Mann-Whitney U test, Wilcoxon Signed Rank test, the Kruskal-Wallis test, chi-square test and Spearman's correlation. Both descriptive and inferential statistics, particularly t-test and chi-square test, are frequently used in previous and recent literature conducted by nursing professionals including clinical nurses and nursing doctoral students [14,15]. The rationale for employing statistical methods in testing hypotheses includes group comparison, understanding relationships, and detecting predictors related to health outcomes. Table 1 provides a summary of frequently applied statistical techniques, along with associated principles and sample questions. This facilitates nurse investigators in selecting an appropriate statistical method for data analysis.

Comparison

Independent t-test

The independent t-test is commonly employed to assess the significance of differences between the means of two independent groups calculated using the t-statistic. This method is applicable in clinical nursing when evaluating interventions, treatments, or therapies by comparing outcomes in two distinct groups of patients.

Mann-Whitney U test

The Mann-Whitney U test is a non-parametric statistical alternative to the independent t-test. This technique is frequently utilized to compare two independent groups, particularly when assessing ordinal or non-normally distributed variables such as a Likert scale score between an intervention and control group. While employing the test, the U statistic is computed and determined if there are statistically significant differences between the two groups.

Paired t-test

The paired t-test, also known as the dependent t-test, is utilized to evaluate changes within the same group or paired/matched data before and after an intervention or treatment. Clinical nurses commonly employ this test when assessing the efficacy of interventions, measuring variables such as pain levels or vital signs before and after a medical procedure or therapy. The t-statistic is calculated and reported from one group or associated individuals at two distinct time points.

Wilcoxon Signed Rank test

The Wilcoxon Signed Rank test is a non-parametric method equivalent to the paired t-test when dealing with non-normally distributed data or ordinal variables. To effectively apply this test differences between the paired, related data points are ranked and calculated in the W test statistic.

Citation: Chen YJ, Cheng CY (2023) Common Statistical Concepts and Methods Used in Nursing Research and Scholarly Projects. Int J Nurs Clin Pract 10: 383. doi: https://doi.org/10.15344/2394-4978/2023/383

Table 1: Summary of assumptions, data types, and sample questions for frequently used statistical methods.

Test		Assumption			Data		Application Notes	Sample Questions
		ND	EV	I	DV	IV		
Two groups comparison	Paired samples t- test (*t*)	✓	✓	✓	continuous	binary	• compare means of the same person at 2 different times • compare means of the dependent/matched pairs on a DV • use Wilcoxon Signed test for non-normal distributed DV	Does women's prenatal depression score differ from their postpartum depression score?
	Wilcoxon Signed rank test (*W*)			✓	continuous	ordinal	• one IV with two or more dependent/matched groups	Does the median score of concern for job security differ between husband and wife?
	Independent t-test (*t*)	✓	✓	✓	continuous	binary	• make sure that group data are independent or not paired • cannot count within-person variation • use the Mann-Whitney test or Wilcoxon Rank Sum test if data are not normally distributed	Does medication error differ between hospitals A and B?
	Mann-Whitney U test (*U*)			✓	ordinal	binary	• shapes of distributions for DVs for two populations are the same. • used for data that are not paired	Does the median score of menstrual knowledge differ for younger and older women?
Three or more groups comparison	One-way ANOVA (*F*)	✓	✓	✓	continuous	categorical	• check differences among three or more groups use post-hoc test to see subgroup comparisons • use the Kruskal-Wallis test for non-normal distributed • if data are matched, use repeated ANOVA • if with equal sample size, robust enough to violate variance of homogeneity • if the sample size is large, robust to violate normality	Does the mean score on the Life Satisfaction Scale differ by marital status (married, single, divorced)?
	Kruskal-Wallis test (*H*)			✓	continuous	categorical	• one IV with three or more groups • compare medians of three or more unpaired groups • if the total sample size is less than 7, the Kruskal-Wallis test will always give a $p \geq .05$	Do the medians of depression scores differ among those who take placebos, those who take low doses of drug A, and those who take high doses of drug A?
Correlation	Pearson correlation (*r*)	✓		✓	2 continuous (X & Y)		• correlations between two variables or sets of variables • for every X value, Y has equal variance • if X values were controlled, use linear regression • use Spearman correlation for non-normal distributed data • if both variables are binary, use X^2 test • if one continuous variable and one true dichotomous variable, use point-biserial correlation; if artificial dichotomous, use biserial correlation.	What is the relationship between the scores of midterm and final tests of students in the statistics class?
	Spearman correlation (rho)			✓	2 ordinal		• relationship between non-normal distributed or proportional/ ranked data • a straightforward way to bring outliers under control	What is the relationship between the ranks of midterm test and final test of students in the statistics class?
	Chi-square test crosstabs (X^2)			✓	binary/categorical		• two categorical variables (one IV & one DV) • variables with two or more levels	Does poor or good sleep quality differ by employment status?
Regression	Multiple linear regression (R^2)	✓		✓	continuous	continuous	• two or more interval and/or categorical IV (need to be dummy coded) • use logistic regression analysis for categorical DV, especially polytomous variable • use hierarchical regression to remove the effect of covariates	How well can moral injury and resilience scores predict compassion fatigue scores?
	Logistic regression (*OR*)			✓	continuous	categorical categorical/	• one kind of nonlinearity regression • use two or more levels of DV • with multiple IVs • if with covariates, may use sequential logistic regression • if Y is survival time, use proportional hazards regression	Can blood pressure (high or low) be predicted by patients' readiness for operation and anxiety score?

Note. ND: normal distribution; EV: equal variance; I: independent; DV: dependent variable; IV: independent variable. Quick memorizing: one-way—one factor (independent variable), two-way—two factors (two independent variables), and so forth. Univariate—one dependent variable, multivariate—two or more dependent variables. Paired—same subject measures two times.

One-way ANOVA

One-way analysis of variance (ANOVA) is an extension of the independent student t-test to examine significant differences among three or more groups. This method is particularly relevant when investigating effectiveness of interventions or treatments by comparing means across multiple groups of patients. When reporting results, the F statistic with its associated p-value is evaluated to determine whether a post-hoc analysis is necessary to understand which specific group comparisons differ significantly.

Kruskal-Wallis test

The Kruskal-Wallis test, a non-parametric statistical technique analogous to one-way ANOVA, compares differences amongst multiple (> 3) groups on a dependent variable when the assumptions of normality are violated or when a small sample size is involved. It is used to calculate a H statistic and assess the efficacy of an intervention or treatment by comparing ranked data across multiple groups of patients. If the H statistic shows statistical significance the Mann-Whitney U test can then be employed to compare between-group differences.

Relationship

Pearson correlation

Pearson product-moment correlation is a widely used method for examining the direction and magnitude of linear associations or dependencies between two continuous variables, such as the correlation between patient age and recovery time after a surgical procedure. To effectively report Pearson correlation, the correlation coefficient (*r*) along with significant p-value measures the strength of both variables of interest ranging from -1 to 1. Whether a positive ($r > 0$) or negative ($r < 0$) relationship exists and what the degree of the bivariate relationship is, is the major information to be interpreted and determined. According to Cohen's guidelines [6], the magnitudes of association between two variables are classified as small (r = .10), moderate (r = .30), and large (r = .50). Applying correlation results, nurse clinicians can identify variables potentially associated with patients' outcomes and therefore can focus more on certain factors to mitigate adverse influences and strengthen positive outcomes for the enhancement of healthcare practices [16].

Spearman correlation

Spearman's rank correlation, a non-parametric test similar to Pearson correlation, serves to assess relationships between two non-linear or ranked variables. However, it is particularly suited for situations where the linearity assumption is not met or when ranked variables are measured other than on a continuous scale, such as ranked pain intensity (e.g., very painful, moderately painful, not so much pain, no pain at all) or medication dosage. Like Pearson correlation, the Spearman's correlation coefficient (*rho*) is calculated and interpreted in terms of direction and strength of bivariate associations.

Chi-square test of Independence

Chi-square is a nonparametric test that is particularly relevant when analyzing data with categorical outcomes or when assessing the relationships or dependencies between categorical and nominal predictor variables, such as evaluating the association between patient outcomes (e.g., improved, stable, worsened) and different treatment modalities (e.g., music therapy or not). The chi-square test is not appropriate to use if the observed data are paired such as pre-test and post-test. To effectively employ the Chi-square test, expected and observed categorical data are calculated and compared whether or not there is a significant association between the expected and observed values. The χ^2 statistic, along with the associated p-value, is obtained using a contingency table (crosstabs) to assess the statistical significance of the observed data.

Prediction

Multiple regression

Multiple regression is a sophisticated statistical technique commonly utilized to explore complex relationships between a dependent variable and multiple independent variables. This method is particularly applicable when seeking to understand how several predictor variables collectively influence a continuous outcome; for example, predicting patient recovery time based on factors of age, medication dosage, and comorbidities. To effectively report the regression results, the percentage of variance in the dependent variable is explained by the linear model of all the selected predictors (R^2) [5]. The regression standardized coefficient (parameter *βeta*) and significance of each predictor are interpreted. Multiple regression provides insight into the relative importance of each predictor to facilitate more comprehensive patient care planning.

Logistic regression

Logistic regression is used to examine the odds of risk occurrence in a binary clinical situation; for instance, survival (yes/no) or depression (depressed/not depressed). This method is particularly relevant when aiming to understand how multiple factors influence the likelihood of a specific binary outcome. The odds ratio (*OR*) and significance of each predictor are interpreted and reported effectively. Logistic regression is beneficial to identify the likelihood of potential risk factors and negative effects that contribute to patient outcomes and to guide decisions regarding intervention strategies and patient management.

Conclusion

In the field of nursing research and scholarly projects, statistical methods are indispensable tools for extracting meaningful insights from collected clinical data and translating them for use as a guide to evidence-based nursing and practice. This paper addresses several fundamental but important statistical concepts and conceptual interactions in alignment with commonly used statistical techniques in nursing research. Each method serves a unique purpose in answering researchable questions, making informed decisions, and advancing the professional nursing field. Whether nurses are comparing treatment outcomes, assessing associations between variables, or predicting patient outcomes, a solid grasp of these statistical concepts and tools is essential. In alignment with the recommendations of a global panel of biostatisticians [17], the approach that incorporates applied statistics teaching and training into nursing education empowers nursing students and clinical nurses, fostering confidence in their statistical comprehension and utilization. This in turn contributes to the body of knowledge driving improvements in healthcare and ultimately enhances patient care and nursing practice.

Competing Interests

The author declare that they have no competing interests.

Refrences

1. Mathew L, Aktan NM (2014) Nurisng student attitudes toward statistics. J Nurs Educ 53: 233-237.
2. Hayat MJ, Eckardt P, Higgins M, Kim M, Schmiege SJ (2013) Teaching statistics to nursing students: an expert panel consensus. J Nurs Educ 52: 330-334.

3. Kuerban A (2017) Biostatistic course in postmaster doctor of nursing practice programs. J Nurs Educ Pract 7: 26-31.
4. Corty EW (2016) Using and interpreting statistics. (3rd edition). New York: Worth, USA.
5. Plichta SB, Kelvin E (2013) Munro's statistical methods for health care research. (6th edition). Philadelphia: Wolters Kluwer, Lippincott Williams & Wilkins, USA.
6. Cohen J (1992) A power primer. Psychol Bull 112:155-159.
7. Sharma H (2021) Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Saudi J. Anaesth 15: 431-434.
8. Tenny S, Abdelgawad I (2023) Statistical Significance. Treasure Island (FL): StatPearls (Internet).
9. Cron SG (2020) The role of statistical analysis in modern nursing research. Res Nurs Health. 43: 301.
10. Bettany-Saltikov J, Whittaker VJ (2014) Selecting the most appropriate inferential statistical test for your quantitative research study. J Clin Nurs 23:1520-1531.
11. Nayak BK, Hazra A (2011) How to choose the right statistical test? Indian J Ophthalmol 59: 85-86.
12. Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M (2019) Selection of appropriate statistical methods for data analysis. Ann Card Anaesth 22: 297-301.
13. Kim HY (2013) Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 38: 52-54.
14. Kuerban A (2018) Statistic methods used in scholarly projects of doctor of nursing practice graduates. Int J Nurs Clin Pract 5: 276.
15. Zellner K, Boerst CJ, Tabb W (2007) Statistics used in current nursing research. J Nurs Educ 46: 55-59.
16. Milligan RA, Frick KD, Pugh LC (2011) Understanding correlation analysis. Am Nurse Today 6: 16-17.
17. Hayat MJ, Higgins M, Schwartz TA, Staggs VS (2015) Statistical challenges in nursing education and research. Nurse Educ 40: 21-25.