

Perspective

Statistical Analysis for Nurse Anesthesiology: What's the Word with All These Numbers?

Thomas Diller, Ph.D., CRNA¹ 

¹ Nurse Anesthesia, University of Tennessee at Chattanooga

Keywords: CRNA, statistical analysis, evidence-based practice, Nurse Anesthesiology, methodology, descriptive statistics, parametric statistics, nonparametric statistics, predictive statistics

<https://doi.org/10.63524/jnae.2882579>

Journal of Nurse Anesthesia Education

The American Association of Colleges of Nursing established educational requirements that every Certified Registered Nurse Anesthesiologist must complete. Whether you are an aspiring nurse anesthesiologist or an experienced practitioner, there are vital concepts to understand evidence-based practice. This educational article provides an overview of key statistical concepts for analyzing healthcare research. It is not imperative that every healthcare provider know how to perform each statistical test they encounter, but it is essential to know why this analysis was chosen, how it answers the research question, what the results infer, and how strong the evidence is to support these claims. A working knowledge of statistical analysis will give nurse anesthesiologists the tools to spot irregularities, skewed data, or inadequate evidence to support generalization to a larger population.

STATISTICAL ANALYSIS FOR NURSE ANESTHESIOLOGY

*The Essentials of Doctoral Education for Advanced Nursing Practice*¹ outlines the clinical scholarship and analytical methods for evidence-based practice. Whether you are a practicing Certified Registered Nurse Anesthesiologist (CRNA) or an aspiring CRNA in nurse anesthesiology school, statistical analysis and translating research into evidence-based practice is expected and vital for practice. There are many research designs for collecting and analyzing data. Commonly seen designs include quantitative (e.g., randomized control trials or cohort studies), systematic or meta-analysis, quality improvement projects, qualitative (e.g., case studies, observational, interviews, lived-experience), or a mixed-method approach. This educational and perspective article provides an introduction to statistical methodologies in quantitative nurse anesthesia research and a stepwise approach for critical analysis of the data.

INTRODUCTION TO DATA COLLECTION AND ANALYSIS

After a researcher has developed their research question, what they hope to discover or answer, a research design is created to tackle the clinical concern. The research question will not only drive the design but also the statistical analysis of the collected data. The first step in reviewing a research article is to identify the study's research question and how it is addressed. The goal of many research studies is to have evidence that can be generalized to a larger population and support a particular clinical intervention or

finding. The next step in reviewing a study would be to determine whether the evidence is strong enough to support the study's inference, practice change, or recommendations. Understanding the purpose of the statistical analysis will assist in this process.

DESCRIPTIVE STATISTICS

After the research design is developed and implemented, the researchers will have collected data on their participants. The first step in data analysis is to make sense of all the data and numbers collected. That is where descriptive statistics come into play. Descriptive statistics are the basic numbers that reveal the average (*mean*), the most frequent response (*mode*), and whether the sample is a normal distribution or skewed in any way (e.g., outliers) (See [Table 1](#)).

There are different types or levels of data that can be collected. The lower levels of variables consist of nominal and ordinal scales; higher levels are interval and ratio scales. Definition and examples of these different types of variables are illustrated in [Table 2](#).² The higher the level of measurement, the more statistical analysis options are available. Therefore, interval or ratio-level variables are often collected. In addition, the researcher will designate an independent and dependent variable. The independent variable is the hypothesized cause of, or influence on, the outcome, and the dependent variable is the outcome of interest.^{2,3} For example, studying the efficacy of an adductor canal block (independent variable) on postoperative pain scores (dependent variable) and time in the PACU (dependent variable).

Table 1. Descriptive Statistics

Terms	Definition	Example
Mean	Arithmetic average: data values are summed & divided by the total (N).	Sample ages: 26, 27, 27, 28, 29, 30, 31 <i>Mean</i> age = 28.3
Median	The score that divides the distribution into 2 equal halves.	Sample ages: 26, 27, 27, 28, 29, 30, 31 <i>Median</i> age = 28
Mode	Highest frequency value or score.	Sample ages: 26, 27, 27, 28, 29, 30, 31 <i>Mode</i> age = 27
Standard deviation	An index that conveys how much, on average, scores in a distribution vary. Calculated by subtracting the mean from each person's original score.	Sample ages: 26, 27, 27, 28, 29, 30, 31 <i>Mean</i> age: 28.3 <i>Standard Deviation</i> = 1.8
P value	Measures the probability of obtaining a value if there is no effect or no differences in your groups (e.g., null hypothesis is true).	Common <i>P</i> value is .05. Represents a 5% chance that the observed data could have occurred due to random variation.
Confidence Interval	Designated range of values within which the parameter has a percentage probability of lying.	Expressed as 95% CI or 99% CI and calculated based on the sample <i>mean</i> , estimated standard error of the mean (<i>SEM</i>), and the value corresponding to the area from the theoretical distribution for the desired CI %.
Positively skewed	Instead of a normal distribution, the longer tail trails off to the right of the sample.	Fewer people with high values, such as income
Negatively skewed	Instead of a normal distribution, the longer tail trails off to the left of the sample.	Fewer people with lower values, such as age at death.

Table 2. Level of Variables

Level of Data/ Variable	Definition	Example
Nominal	Lowest form of measurement; numbers are used simply as labels to name categories	Gender, CRNA or MD, Practicing or Retired
Ordinal	Used to designate ordering on an attribute, but does not indicate distance between values	Pain Scale: 0 equals no pain, and 10 equals the worst pain ever. How would you rate your pain?
Interval	Uses numbers to designate ordering on an attribute and conveys information about the amount & distance between values; <i>mean</i> values can be calculated	Temperature, heart rate, respiratory rate, blood pressure
Ratio	Uses numbers to designate ordering, conveys information about amounts, distances are equal, and there is an absolute zero; the <i>mean</i> values can be calculated	Medication dose (e.g., milligrams, volume), time

CENTRAL TENDENCY THEOREM

Vital to understanding descriptive statistics is understanding the *Central Tendency Theorem* and the terms associated with it. The *mean* is the arithmetic average of all values in a dataset.^{2,4} The *median* is the value that is in the middle of the dataset.^{2,4} The *mode* is the value that occurs most frequently.^{2,4} These values assist the researcher in understanding the distribution of the data and determining whether the dataset is normally distributed or skewed positively or negatively.

A normal distribution is critical to understanding the *Central Tendency Theorem*. If the dataset follows a normal distribution (e.g., bell-curved or Gaussian), the *mean* value lies in the middle, and the distribution is symmetric.² Therefore, normal distribution allows the researcher to make statistical assumptions and to use parametric tests to infer findings to a larger population. The researcher will use values of *Standard Deviation* (SD), the distance a particular value lies away from the *mean*, *Confidence Interval* (CI), the

certainty that the true value lies within a particular range (e.g., 95% CI), and *P* value, which is whether the value is statistically significant.²

The *P* value is the probability that the results occurred by chance. A common acceptable limit is a *P* value of .05.⁵ This indicates that the probability of these values occurring by chance alone is less than 5%. If a *P* value is less than .05 (e.g., $P < .05$), the possibility of this value occurring by chance is low; therefore, the *mean* scores are significantly different from each other. If the *P* value is greater than .05, there is no statistical difference between the values.^{2,3,5}

INFERENTIAL STATISTICS

After organizing the dataset and running descriptive statistics, the next step is to determine which type of inferential statistics would be appropriate for the collected data. Inferential statistics are the conclusions the researcher wants to obtain from the sample's data. Inferential statistics relies on the laws of probability and is used to test hypotheses

Table 3. Parametric Versus Nonparametric Tests

Parametric	Nonparametric
Paired (dependent) Student's <i>t</i> test	Fisher's exact test
Unpaired (independent) Student's <i>t</i> test	Wilcoxon signed-rank test
ANOVA	Mann-Whitney U test
Pearson coefficient	ANOVA by rank
Correlations	Spearman's rank correlation coefficient

(e.g., relationship between variables) in the research design.^{2,3} Hypothesis testing is based on the researcher's hypothesis (e.g., the difference that is posited to occur) compared to the null hypothesis (e.g., there is no difference between the 2 samples).²

PARAMETRIC STATISTICS

Parametric statistics involves estimating a population parameter (see [Table 3](#)).^{2,3,6-8} Typically, this type of analysis assumes the dependent variable is normally distributed and requires a large sample size.² When reviewing a study, the descriptive statistics should be reported, along with whether the sample is normally distributed. These initial analyses of the data set are performed to test assumptions for performing parametric statistics. If assumptions are not met, then nonparametric analysis is required.

The assumptions for parametric statistics include a random sample, normality, and equal variances (e.g., the 2 samples are similar). The test of normality is assessed using the Kolmogorov-Smirnov test if the sample size is greater than 50, or the Shapiro-Wilk test if the sample size is less than 50. The test for equality of variance is assessed by a Levene's *F* test. Lastly, Cohen *d* can be used to calculate an effect size. Cohen *d* is the difference between the *mean* scores divided by the pooled standard deviation. A large effect size would yield a Cohen *d* of 0.8, while a small effect size would yield a 0.2.

DEPENDENT AND INDEPENDENT T TESTS

The cornerstone of clinical research is comparing 2 groups to determine if there is a difference between them. A dependent *t* test analyzes the difference between the same group of participants before and after an intervention (e.g., pre- and post-test).^{2,3,6,7} An independent *t* test analyzes the difference between 2 independent groups (e.g., control group and experimental group).^{2,3} To perform this analysis, assumptions must be met, including that the variable is an interval or ratio level, that 2 independent groups are being compared, and that the samples are similar in terms of variance and distribution.^{2,3} These assumptions are tested via test for normality (e.g., Shapiro-Wilk Test or Kolmogorov-Smirnov Test) and Equality of Variance (e.g., Levene's *F* test).^{2,3}

To illustrate these principles, let's look at a couple of articles from nurse anesthesiology research. Conn⁹ studied the impact of mandatory substance use disorder education on peer perception of impairment in CRNAs. The study

design compared participants who received education to those who did not to determine whether there was a statistically significant difference in their perception towards impaired CRNAs. The research question is whether education impacts CRNAs' perception of impaired CRNAs. To answer this question, the researchers surveyed a group of CRNAs without education and a group with education and compared their results. While the *mean* scores were more positive after education (*mean* of 62.44; *SD* of 7.124) than among those who did not (*mean* of 64.17; *SD* of 6.919), the difference was not statistically significant ($P = .133$).⁹

An example of a dependent *t* test is Allen et al¹⁰'s study of the effectiveness of low-fidelity simulation training on self-efficacy and performance related to scalpel-bougie-tube surgical cricothyrotomy among Resident Registered Nurse Anesthesiologists. The research question is whether the simulation training would improve performance. The research design used a pre-post-test, and a dependent or paired *t* test was used to compare the results. Allen et al¹⁰ reported a statistically significant improvement in self-efficacy (3.13 to 4.5 out of 5, $P < .001$), mean completion time (103.5 seconds to 55.9 seconds, $P < .001$), and checklist scores (5.5 to 9.1 out of 10, $P < .001$) (See [Table 4](#)).

ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

If the researcher wants to compare more than 2 group *mean* scores or study the same sample at 3 or more intervals, a One-way Analysis of Variance (ANOVA) is used.^{2,3,6,7} It is important to note that an ANOVA alone cannot tell the researcher which specific groups were statistically significantly different from each other.² A post hoc analysis is required to determine which specific groups are different from each other (e.g., Tukey Post Hoc analysis). The assumptions for this analysis are similar to those for the *t* test: random sampling, normal distribution, equal variances, and equal group sizes.^{2,3}

Kruse et al¹¹ studied second victim distress among CRNAs. The research question consisted of what impact second victim distress had on absenteeism and turnover. Further analysis of the data consisted of determining differences across groups (e.g., gender, age, and practice type). Since the researchers are looking at multiple groups (e.g., more than 2), an ANOVA test is most appropriate. Kruse et al¹¹ reveal no statistical significance in second victim distress among the groups compared (see [Table 5](#)).

Table 4. Dependent *t* Test Results^{a,b}

I am that confident that I:	Strongly Disagree n (%)	Disagree n (%)	Neutral n (%)	Agree n (%)	Strongly Agree n (%)	P value ^d n (%)
Could identify CICO event ^c	0 (0)	0 (0)	1 (10)	8 (80)	1 (10)	.26
Pre:	0 (0)	0 (0)	0 (0)	7 (70)	3 (30)	
Post:						
Know Equipment Required:	0 (0)	4 (40)	3 (30)	1 (10)	2 (20)	.012
Pre:	0 (0)	0 (0)	0 (0)	2 (20)	8 (80)	
Post:						
Know Landmarks:	0 (0)	2 (20)	2 (20)	5 (50)	1 (10)	.014
Pre:	0 (0)	0 (0)	0 (0)	5 (50)	5 (50)	
Post:						
Know Appropriate Steps:	0 (0)	5 (50)	3 (30)	2 (20)	0 (0)	.007
Pre:	0 (0)	0 (0)	0 (0)	2 (20)	8 (80)	
Post:						
Possess Technical Skills:	1 (10)	3 (30)	3 (30)	3 (30)	1 (10)	.011
Pre:	0 (0)	0 (0)	0 (0)	0 (0)	4 (40)	
Post:						
Could Remain Calm:	1 (10)	6 (60)	1 (10)	1 (10)	1 (10)	.011
Pre:	0 (0)	2 (20)	1 (10)	3 (30)	5 (50)	
Post:						

Abbreviations: CICO: cannot intubate, cannot oxygenate

^aNonparametric analysis was used on this study^bReference: Allen et al¹⁰^cSelf-Efficacy Pre-and Post-education and simulation-based intervention^dP value: significance level**Table 5. ANOVA Comparison of Groups^a**

Groups	Second victim distress: Mean score	P value
Gender:	.56	.365
Male:	.89	
Female		
Age:	.67	.233
20-39	.36	
40-49	.65	
50-59	.42	
≥ 60		
Work environment:	.55	.845
Level 1 trauma	.44	
Level 2 trauma:	.61	
Level 3-5 Trauma	.47	
Does not apply/unknown		

^aReference: Kruse et al¹¹

NONPARAMETRIC STATISTICS

Nonparametric statistics are used when the outcomes are not measured on an interval or ratio level, or when the sample collected does not meet the assumptions for parametric statistics.^{2,3,6,7} Therefore, the sample is either not normally distributed (e.g., skewed or containing outliers), is too small, or the variances between the samples are unequal. Each parametric test has an equivalent nonparametric analysis that can be performed (See [Table 3](#)).

CHI-SQUARE, MANN-WHITNEY U TEST, AND WILCOXON SIGNED-RANK TEST

A *Chi-squared* test will test the relationship between 2 categorical variables.^{2,3,6,7} It is also called a *Pearson's chi-*

square test. One dependent variable is measured at the dichotomous level (e.g., 2 categorical independent groups such as regional versus general anesthesia), and one independent variable has 3 or more independent groups (e.g., low, medium, or high).^{2,3} The *Chi-square* test then analyzes differences in proportions across groups.

The *Mann-Whitney U* test is the nonparametric alternative to the independent *t* test. This analysis is appropriate if the assumptions for an *independent t* test were not met (e.g., not normal distribution, ordinal dependent variable).^{2,3} The *Mann-Whitney U* test determines whether there are differences between groups based on the *mean ranks* (e.g., *median* differences). The Wilcoxon Signed-Rank Test is the nonparametric alternative to the *dependent t* test.

Table 6. Pearson's r Test

Coefficient value	Strength of Association
$0.1 < r < 0.3$	Small correlation
$0.3 < r < 0.5$	Medium/Moderate correlation
$ r > 0.5$	Large/Strong correlation

Table 7. Pearson's r Correlational Analysis^a

Variables	Correlations	NCE Score	P value
Grit overall average score	Pearson Correlation	-0.097	.455
Perseverance of effort average score	Spearman Correlation	-0.269	.035
Consistency of interests average score	Pearson Correlation	0.037	.776

^aReference: Collins et al¹³

An example of nonparametric analysis is a study by Villarreal et al¹² on intubation techniques in a patient in the prone position. The research question was which intubation technique provided better results when intubating a patient in the prone position. The research design consisted of a comparison of reintubation on a manikin in the prone position using a video laryngoscope or intubation with a laryngeal mask airway (LMA) and a fiberoptic scope. Since all variables measured deviated strongly from normality (e.g., assumptions were not met for parametric analysis), a *Mann-Whitney U* test was used to compare the 2 groups. Villarreal et al¹² reported statistically significant lower median times using the Glidescope (73.5 seconds versus 130 seconds; $P < .001$) compared to LMA and fiberoptic scope.

PREDICTIVE STATISTICS: CORRELATION AND SIMPLE LINEAR REGRESSION

Correlation and simple linear regression are predictive statistical analysis methods. These methods assess the linear relationship between 2 continuous variables to predict the value of a dependent variable based on the value of an independent variable.^{2,3,6,7} For example, a researcher wanted to determine whether there was a relationship between cholesterol levels and sedentary lifestyle. The first step would be to create a scatter plot graph that visually portrays a linear regression equation. Next, the researcher must check the assumptions for this analysis. An independence of observation (e.g., *Durbin-Watson* test) and a Normality and Equality of Variance will assess for outliers. If the assumptions are met, a Pearson's r test can determine whether the correlation is small, medium, or strong (See [Table 6](#)).

The last example from nurse anesthesia literature comes from Collins et al¹³'s study regarding GRIT as a predictor of

nurse anesthesiology student success. The research question was whether "grit" was predictive of National Certification Examination (NCE) results. The research design consisted of students taking a reliable and validated instrument that measures a student's grit (e.g., perseverance and effort) and comparing these findings to NCE results (e.g., first-time passing). The researchers reported a statistically significant correlation between "perseverance of effort" and nurse anesthesiology success, defined as first-time passing NCE (See [Table 7](#)).

CONCLUSION

It is fitting that we end on grit because it takes perseverance and effort to perform research and understand the statistical analysis behind the results. When developing evidence-based practice DNP Projects, it is helpful to have a systematic approach to analyzing research. Understanding the methods of research design, variables being studied and collected, and how these outcomes are being analyzed will enable the nurse anesthesiology student and CRNA to identify the strengths and limitations. A stepwise approach recommended is to identify: 1) the research questions, 2) what type of design is attempting to answer these questions, 3) what type of variables/outcomes are being collected, 4) what statistical tests are being used to analyze this data, and are they appropriate, and 5) is there enough evidence to support the researchers claims or inferences. By implementing these steps and carefully examining the evidence, CRNAs can continue to be leaders in their healthcare settings and establish evidence-based practices for their patients.

Submitted: December 16, 2025 EST. Accepted: February 10, 2026 EST. Published: February 16, 2026 EST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. American Association of Colleges of Nursing. The Essentials of Doctoral Education for Advanced Nursing Practice. Accessed December 10, 2025. <https://www.aacnnursing.org/Portals/0/PDFs/Publications/DNPEssentials.pdf>
2. Polit DF. *Statistics and Data Analysis for Nursing Research*. 2nd ed. Pearson Education Inc; 2010.
3. Laerd Statistics. Features overview. Accessed December 15, 2025. <https://statistics.laerd.com/features-overview.php>
4. Winters R, Winters A, Amedee RG. Statistics: a brief overview. *Ochsner J*. 2010;10:213-216.
5. Cohen J. Statistical power analysis. *Curr Dir Psychol Sci*. 1992;1(3):98-101. doi:[10.1111/1467-8721.ep10768783](https://doi.org/10.1111/1467-8721.ep10768783)
6. Mishra P, Pandey C, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. *Ann Card Anaesth*. 2019;22(3):297-301. doi:[10.4103/aca.ACA_248_18](https://doi.org/10.4103/aca.ACA_248_18)
7. Illescas A, Zhong H, Cozowicz C, et al. Health services research in anesthesia: a brief overview of common methodologies. *Anesthesia and Analgesia*. 2022;134(3):540-547. doi:[10.1213/ANE.0000000000005884](https://doi.org/10.1213/ANE.0000000000005884)
8. Turner DP, Deng H. Overview of common statistical tests and their assumptions. *Headache*. 2020;60(5):826-832. doi:[10.1111/head.13789](https://doi.org/10.1111/head.13789)
9. Vice O'Con K. The impact of mandatory substance use disorder education on peer perception of impairment in CRNAs. *AANA J*. 2025;93(4):289-293. doi:[10.70278/AANAJ.0000001031](https://doi.org/10.70278/AANAJ.0000001031)
10. Allen D, Bulick H, Bowman Dalley C, Crowell N, Miller C, McAuliffe M. Emergency surgical front-of-neck airway access: effectiveness of a didactic and simulation-based training intervention to improve performance and self-efficacy among Resident Registered Nurse Anesthetists. *AANA J*. 2025;93(3):213-220. doi:[10.70278/AANAJ.0000001020](https://doi.org/10.70278/AANAJ.0000001020)
11. Kruse JA, Podojil-Kostecki P, Smth B. Living with the aftermath: the second victim experience among Certified Registered Nurse Anesthetists. *AANA J*. 2024;92(3):173-180.
12. Villarreal A, Ruiz de Somocurcio A, Schauweker C, Fisher R, Cornelius B. Endotracheal tube replacement in the prone position with intubating laryngeal mask airway/flexible fiberoptic scope combination or Glidescope® video laryngoscope: a manikin-based study. *AANA J*. 2025;5(93):339-344. doi:[10.70278/AANAJ.0000001001](https://doi.org/10.70278/AANAJ.0000001001)
13. Collins S, Tracy A. GRIT as a predictor of nurse anesthesiology student success. *AANA J*. 2025;4(93):295-300. doi:[10.70278/AANAJ.0000001022](https://doi.org/10.70278/AANAJ.0000001022)