4-1996

Basics of research (Part 6): Quantitative data analysis

Cheryl Thompson
University of Nebraska Medical Center, cbthompson@unmc.edu

Robert Schwartz
University of Pittsburgh

Eric Davis
Strong Memorial Hospital

Edward A. Panacek
University of California, Davis

Follow this and additional works at: https://digitalcommons.unmc.edu/con_articles

Part of the Nursing Commons

Basics of Research (Part 6): Quantitative Data Analysis

Cheryl Bagley Thompson, PhD, RN, CS,[1] Robert Schwartz, MD, MPH,[2] Eric Davis, MD,[3] Edward A. Panacek, MD[4]

1. University of Utah College of Nursing, Salt Lake City, Utah

2. Department of Emergency Medicine, Center for Injury Research and Control, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

3. Department of Emergency Medicine, Strong Memorial Hospital, Rochester, New York

4. Division of Emergency Medicine and Clinical Toxicology, University of California, Davis, Medical Center, Sacramento, California

Address for correspondence and reprints: Cheryl Bagley Thompson, PhD, RN, CS, University of Utah College of Nursing, 25 South Medical Drive, Salt Lake City, UT 84112

Introduction

You can do it early or you can do it late, but eventually all investigators using quantitative methods have to deal with statistical analysis. The purpose of this part of the Basics of Research series is to try to take some of the fear out of approaching your data analysis by providing an introduction to basic statistical concepts and the process of data analysis. To help accomplish this, definitions will be provided for uncommon concepts and examples will be used to demonstrate key points.

The discussion will proceed through the chronologic steps used for a statistical analysis: planning, preparation, and statistical concepts. The final section will include a discussion on interpreting your statistical analysis.

Planning

The most important part of any research project is the planning process. The more complete the planning, the smoother the project generally runs. This includes data collection and analysis. The development of a statistical analysis should not be delayed until after the data is in hand. Rather, the investigator should have a specific plan for data analysis before initiating the study.

Investigators uncomfortable with statistical analysis should consult a statistician early in the planning phase. A statistician will help the investigator determine what statistical analysis is most appropriate for answering the research question considering the type of data collected. Statisticians, however, can be intimidating to some individuals. This should not be the case. The investigator

should consider the statistician an important consultant and collaborator and should avail himself or herself of the statistician's expertise.

To help diminish the stress of a statistical consultation, the investigator should prepare a list of questions before meeting with the statistician. In creating the list of questions the investigator needs to start with the research question.[1] If the investigator has an idea of what statistics to use, the questions for the statistician are related to whether the proposed analyses are appropriate and what other statistics should be considered. If the investigator has no idea of what statistics to use, the first question should be, "What statistics are appropriate for the research questions being addressed?".

The investigator should take advantage of the meeting with the statistician to find out why the analysis is appropriate and to increase his or her knowledge of statistics. The investigator needs to be able to defend his or her choice of statistics at presentations and within publications. Saying merely that the statistician said so is not sufficient. Consequently, this understanding of the analysis is essential.

Several advantages result from having a plan for data analysis before beginning the study. The most obvious is that the investigator is not left perplexed about what to do with all the data now in the computer. A plan speeds the process of data analysis. If a computer program will be used, the commands for the analysis can even be written before data collection is complete. In this case, as soon as all the data are entered, the investigator runs the predetermined pro-

grams and the analysis is ready for interpretation.

The second advantage of planning the statistical analysis before the study is an increase in scientific integrity. The investigator who has a plan ahead of time is less likely to bend the analysis to suit the purpose. A plan also prevents the process of doing repeated analyses until something is found that is statistically significant. A post hoc (after the fact) approach to statistical analysis is inappropriate and increases the chance of making a type I statistical error (see following section on hypothesis testing for discussion of type I errors). If post hoc analyses are used, a technique such as Bonferroni adjustment is needed to decrease the chance of a type 1 error.[2]

Before beginning a study, the investigator also should know what computer hardware and what statistical and data entry software are available for data entry and data analysis. The investigator should know the capabilities of the statistics program to be used and should determine whether his or her computer is powerful enough to run the planned analyses. The investigator should not plan for analyses that he or she has no way of conducting. The investigator also should spend time during the early phases of the project becoming familiar with the statistical package to be used. Data analysis will proceed more smoothly if the investigator does not need to stop and ask for technical assistance.

Data Management
Data Storage
Data, once collected, should be stored carefully to prevent damage or violations of subject privacy. Having a plan for organizing the data before data collection begins is also advisable. A commonly used method is manila file folders. Data should be sorted into appropriate file folders and placed into a locked file cabinet. If data is not to be entered into a computer file but will be tabulated by hand, it is advisable to make a complete copy of the data and then store the complete copy at another site.

Security of data remains an issue even after data is stored within a computer file. The investigator must ensure that unauthorized access to the files is not possible. Data files need to be stored in a computer that can only be accessed by the research team or must be placed into files with password protection. In addition, the investigator must have backups in case of catastrophic loss. Backups should be made at frequent intervals during data entry to prevent loss caused by power failure or other mechanical difficulty. The investigator needs to mark the disk(s) carefully so that the most recent disk is used for the next data entry session. Data can be inadvertently lost if some subjects are entered on one disk and the next set of subjects added to a backup that does not contain a complete set of data.

Data backups should be made at the end of each day and can consist of the entire computer system or of individual data files for the study. Backups can be made onto tapes or onto floppy disks. The best backup plan will provide for storage of backup tapes at a site remote from the location of the original data. This plan will protect against major loss of data in case of fire or other disaster that might destroy data residing in a desk, as well as on the computer itself.

Data entry
Data can be entered into a computer file for analysis using one of several data entry methods. Although not common with air medical research, data can be collected on forms that are automatically read into a computer data file. Because of the expense involved, this is not a common method for data entry.

A second method entails the development of a computerized data entry form that resembles the original form onto which the data is hand written by the data collector or subject. This method is user friendly because the screen resembles the paper form and generally is easy to understand. In addition, the individual entering data can see that the data they are entering matches the variable name on the screen. The disadvantages of this method are that data entry takes longer because the enter, tab, or arrow key must be pressed with each data element and time is needed to create and debug the input form.

Data also can be directly entered onto a spreadsheet. With spreadsheet data entry the names of the variables are placed across the top row and the individual subjects are listed down the side. This data format takes less time to create than a data input screen but still requires the enter, tab, or arrow key to be pressed after each data element. Most current spreadsheets allow for a variety of simple statistics to be calculated without using a statistical analysis program. Investigators needing a sophisticated statistical analysis may need more power than is available with a spreadsheet. However, statistical programs such as SPSS for windows provide a spreadsheet-type data entry format that then allows complex data analysis using SPSS.[3]

A final method for data entry is entering data directly into a flat text file. This type of data entry is demonstrated in Table 1. The main advantage of flat files is the speed of data entry. A competent typist can enter data quickly because few if any characters need to be typed between variables. Spaces are left occasionally to allow the eye to follow columns of values, but the spaces are much fewer than needed to enter data into a spreadsheet. The disadvantage of this method of data entry is that it is easier to make a mistake. Numbers once entered are not associated directly with a variable name. Consequently, finding errors in data entry may be more difficult.

Data coding is the process of assigning a numerical value to a qualitative response. For example, an investigator may ask for the professional background of a transport team member. The responses may be physician, nurse, paramedic, respiratory therapist, pilot, or other. To facilitate the statistical analysis and to save space in the data file, the investigator may assign a numerical value to each possible response such that all physicians would be 0, all nurses 1, all paramedics 2, all respiratory therapists 3, all pilots 4, and all others 5. Although numbers are easier to deal with, the investigator must be careful that he or she does not attribute ordinal or ratio level characteristics to data that were originally nominal in nature. For example, a correlation requires that the level of measurement for both of the variables be at the interval or ratio level of measurement. Coding profession as a numerical value does not then allow the investigator to do correlations between profession and another variable. The data remain at

the nominal level and thus are inappropriate for inclusion in a correlation analysis.

Data transformation is a variation of data coding and is the process of changing the numerical representation of a quantitative value to another value. Data can be changed to reflect a new measurement scale or to make alterations in the distribution of the data. For example, data on neonate weight may be obtained in grams. However, if older subjects also are included in the dataset, the neonate weights may need to be transformed into kilograms from the original grams. In the case of family income level an uneven distribution of subjects is usually found at the low end of the scale. Economists often use a logarithmic transformation so that the data are more evenly distributed. If the income data are reduced to their logarithmic value, the high incomes are brought closer to the lower end of the scale and provide a distribution closer to a normal curve.

Physically coding the data can be accomplished by one of several methods. Some researchers like to evaluate each data collection form before data entry and place the appropriate code next to the response on the form. Others prefer to put the appropriate code on the data collection form before distribution so that the code can be read from the form at the time of data entry. A final way to code data is to have a separate code book and to use this as a reference during data entry into the computer. The first method requires two passes through the data collection forms. Consequently, one of the later two approaches often is preferred. If the investigator does not wish the subject to be aware of the coding scheme or does not have room for it on the data collection form, the last option may be the best.

Data cleaning

Data entered into the computer will inevitably contain mistakes. Data cleaning is the process of trying to find errors and to correct them before data analysis. Several techniques can be used for data cleaning.

Programs are available that allow the investigator to enter the data twice. The computer then checks for inconsistencies between the two sets of data and notifies the individual that a mistake has

Table 1
Flat Text File Data Entry

101120195	3455	11122012
102120395	3545	11021121
103120695	4443	01011022
104121095	3343	10112222

Table 2
Misaligned Data Indicating Data Entry Error

101120195	3455	11122012
1022120395	3545	11021121*
103120695	4443	01011022
104121095	3343	10112222

*The longer line is because an extra 2 was entered with the subject number of 102 in the first three columns.

Figure 1

Population

Group 1	Group 2
p o l x p o l x p o l x	p o l x p o l x p o l x
p o l x p o l x p o l x	p o l x p o l x p o l x
p o l x p o l x p o l x	p o l x p o l x p o l x

Sample

Group 1	Group 2
p x p l	p o x l
p o p x	p p x l
p p l o	o o x l

been made. The primary disadvantage of this method is the increased time needed for data entry. Another disadvantage is that commonly the same mistake will be made twice, thus concealing the mistake.

Another method for cleaning data as it is entered is by using controls in the data entry program. If a form is constructed for use with data entry, value checking rules can be created. Once a rule is created, the computer will check all data entered into that field to make sure that it meets the established criteria.

Other methods of data cleaning include a variety of ways for examining the raw data after computer entry. The first approach is to use the appearance of the data. One of the most common errors with data entry is skipping variables or double entry of a value. If periodic spaces are planned into data entry, either of these errors will cause the data to become misaligned (Table 2).

The method of data cleaning that catches most errors is the process of reading the data back. With this method one investigator reads the data elements from the data entry form and the second checks the values entered by reading a copy of the data file. Inconsistencies between what the data should be and the data that was entered become obvious and errors can be corrected. Some investigators elect not to read the data back

believing that their data entry was adequate and that obvious errors can be found with other methods.

A final method for data checking is to obtain a frequency distribution for all variables. The investigator can check for obvious errors by comparing the data obtained with the data that should have been obtained. For example, if gender is coded as 0 for males and 1 for females, no 2's or 3's should appear for the value of gender. If the frequency distribution shows an erroneous value, the investigator can search the data set for the erroneous value, find the subject number for the subject where the error was made, find the correct value on the original data collection form, and reenter the data.

Statistical Analysis
Overview
Understanding the concepts of a statistical analysis before actually running any statistical tests is extremely important. On a basic level, statistics (specifically inferential statistics) try to determine whether a difference exists between measured values and how confident you can be that the difference is, in fact, a real difference. The measured values for a study are derived from a sample, which is some subset of a population.[4] Statistical tests can be used to determine whether the sample is representative of the total

Table 3	
Yearly Income	
Ratio Level Data	**Ordinal Level Data**
4590	<10,000
11,230	10,000-20,000
25,600	20,001-30,000
33,775	30,001-40,000

Data from the ratio level column can be reclassified as a value from the ordinal level column but not vice versa.

population. This information tells the investigator how much confidence he or she should have in the results.

For example, in Figure 1, the total population has 9 x's, 9 o's, 9 p's, and 9 l's in both group 1 and group 2. However, the sample from group 1 overrepresents the number of p's, showing that there are twice as many p's in group 1 as in group 2. This sample would not be representative of the entire population, and an appropriate statistical analysis may help to demonstrate that the observed differences are not present in the population.

If it were feasible to measure the entire population, the errors introduced by sampling would be eliminated. However, for a variety of reasons we cannot measure the entire sample. In air medical research we commonly measure samples of patients. For example, in a study of intubated patients we cannot have measurements on every patient intubated in the entire country, but we can take samples of patients intubated by one or maybe more programs. Consequently, we must rely on statistical analyses to help determine whether what is observed in a sample is representative of the population.

Level of Measurement
Level of measurement refers to the amount of information contained within the data element. Data elements are measured at the nominal, ordinal, interval, or ratio (continuous) level. Nominal level data (or categorical) refers to data that can only be put into groups. For example, the demographic data elements of race and religion are nominal level. This means that the values consist of categories such as white, black, Asian, and other. With nominal level data no category is better than another and the difference between categories cannot be deter-

mined. For example, neither Catholic nor Protestant is better than the other, and it is not known whether Catholic is closer to Protestant or to Jewish. The only interpretation possible is that two data elements are or are not the same.

A specific subset of nominal level data is dichotomous data. Dichotomous data are data that are nominal but have only two possible categories. A common example is mortality. The values for mortality are either live or die. Other dichotomous variables are things that can be measured as yes or no, on or off, or present or absent. Dichotomous variables possess characteristics beyond those of other nominal level data, but such a discussion is beyond the scope of this paper.

Ordinal level data are one step up from nominal data. As the name implies, ordinal data have an inherent order. Data values such as never, sometimes, often, and always have order. An individual would interpret sometimes as being more frequently than never, and always as more frequent than often. However, the difference in magnitude is not known with ordinal level data. It cannot be said of ordinal level data that always is twice as frequent as often or that the distance between never and sometimes is the same as the distance between sometimes and often. The only interpretation available is that of which is greater or which is less.

At the interval level of measurement distances between data elements can be determined. Temperature is the most common variable measured at the interval level. With temperature the difference between 40° and 50° is the same as the difference between 50° and 60°. However, interval level data have no true zero. Consequently, multiplication is not allowed with interval level data. This

means that you cannot say that 10° is twice as hot as 5°. Because there is no true zero, a reference point does not exist. Zero degrees Fahrenheit or Celsius are arbitrary numbers that do not relate to the amount of temperature present. In contrast, temperature measured in Kelvin has a true zero (absolute zero), in which the speed of molecular movement serves as the reference point. Unlike Fahrenheit or Celsius, at 20° Kelvin molecules are moving twice as fast as at 10° Kelvin.

Data elements that have the characteristic of a true zero are measured at the ratio level of measurement. Ratio level data are measured at the highest level of measurement and contain the greatest amount of information. Ratio level data can be transformed by addition, subtraction, multiplication, or division and their relative values are not altered.

Data elements can always be reduced in their level of measurement but can never be increased. For example, if income is measured at the ratio level, the data elements can be reduced to ordinal level data by the creation of categories (Table 3). However, data measured at the ordinal level of measurement cannot be increased to the ratio level of measurement. If we ask a subject what category their income fits into, we can never determine from the raw data their exact income level. Consequently, data should always be collected at the highest level of measurement possible and converted at the time of data analysis if a lower level of measurement is desired. The only case when this is not recommended is when there is strong reason to believe that a subject will not be truthful if they are required to provide exact data or when the data cannot be expected to be accurately measured at the higher level of measurement.

Computer Programs
Because of the availability and ease of use of current statistical computer software packages, most researchers use a computer to perform their data analysis. Hand calculation of all statistics is possible but is very tedious, time-consuming, and open for calculation errors for all but the simplest of statistics and consequently is not recommended.

Computer programs are available for

Table 4

Frequency Distribution

Gender	Number (%)
Male	154 (43.4)
Female	201 (56.6)

Table 5

Bivariate Frequency Distribution

Gender	Professional Role			
	Paramedic	Nurse	Physician	Respiratory Therapist
Male	74	62	12	6
Female	53	142	5	1

Figure 2

Frequency Distribution (Histogram)
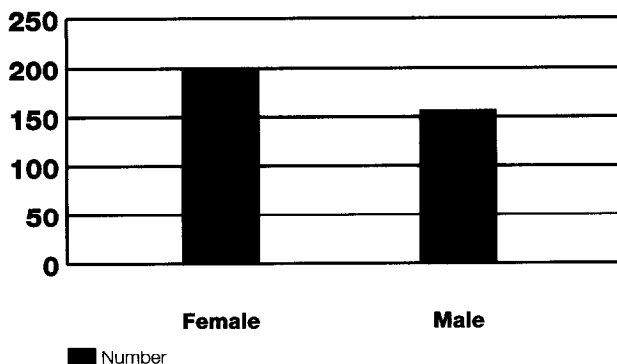
both mainframe and personal computers. Personal computers are generally easier to use but the software may be expensive for an individual to purchase. However, use of an agency mainframe computer for a small data analysis may be cheaper than purchasing a program for one time use on an investigator's personal computer.

Some of the more common statistical packages include SPSS, SAS, and Minitab.[3] In addition, as mentioned earlier, most current spreadsheets and database applications provide options for simple and moderately complex statistical analyses. The investigator should determine before the study what statistical packages are available and make arrangements to purchase a package if none are available. Checking the capabilities of your computer before purchasing a computer program is highly recommended because most statistical packages re-

quire a fair amount of random access memory (RAM) and hard disk space. If you do not have a computer or software for analysis available, you may be able to contract with a statistician or other individual to run your analyses for you.

Descriptive Statistics
Descriptive statistics are numbers that summarize your data with the purpose of describing what occurred in your sample. In contrast, inferential statistics are numbers that allow the investigator to determine whether differences exist between two or more samples and whether these differences are likely to be present in the population of interest. Descriptive statistics are not meant to be used to draw conclusions. However, descriptive statistics can be used to compare samples from one study to another. Descriptive statistics also help the researcher de-

tect characteristics of the sample that may influence their conclusions. For example, if a sample of air medical personnel included 400 women and only 20 men, the investigator would need to be careful about generalizing the findings to male air transport personnel.

Frequency distributions are often the first analysis to be done on a data set. Frequency distributions are a valuable method for describing nominal or ordinal level data (discreet data). Because discreet data only characterize the quantity within categories, a frequency distribution adequately describes nominal or ordinal level data. Frequency distributions also can help to detect data entry errors.

A frequency distribution consists of a description of the number of subjects selecting each possible option and may include the percentage of the sample that this number represents. For example, for the variable gender, a frequency distribution would list how many males and how many females were in the sample. A frequency distribution can be shown using numerical values or using graphical techniques (Table 4 and Figure 2).

Frequency distributions are often univariate (one variable only) but may be bivariate (including two variables). A bivariate frequency distribution is often presented as a table with the name and values of one variable across the top and the name and values of the second variable down the left side. See Table 5 for an example of a bivariate frequency distribution. Multivariate frequency distributions describing greater than two variables at one time are possible but become more complex and beyond the scope of this article.

Measures of central tendency are statistics that describe where the middle of the sample lies. The lowest level measure of central tendency is the mode. The mode is the value most frequently occurring within the dataset. A mode can be used with all levels of measurement and is the primary measure of central tendency available for nominal level data. When examining the number of patients with trauma, cardiac, or other medical problems, the category having the most patients represents the mode. If evaluating the ages of a sample, the age represented by the most subjects is the mode.

Although the mode provides helpful

Table 6	
Mode for Interval or Ratio Level Data	
Age	Frequency
18	5
23	1
31	1
40	3
42	4
43	2
45	1
46	2
54	1
65	1

Mode = 18

Table 7		
Median Example		
	Sample 1	Sample 2
	18	28
	20	34
	21	36
	36	42
	45	45
	46	55
	52	60
	53	62
	54	68
Mean	**38.3**	**47.8**
Median	**45**	**45**

Table 8		
Range and Standard Deviation Example		
	Sample 1	Sample 2
	19	33
	21	42
	32	48
	44	51
	55	53
	69	54
	71	56
	72	60
	76	62
Range	**57**	**29**
Mean	**51**	**51**
Standard Deviation	**22.7**	**9.0**

Figure 3

Normal Distribution (Bell Curve)
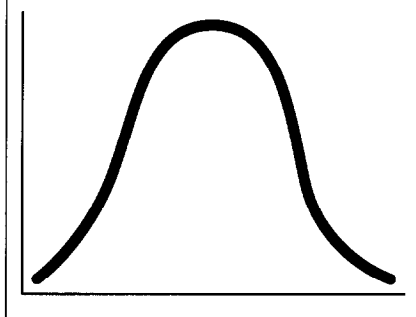
information for nominal or ordinal level data with only a few categories, the mode may be of little value with interval or ratio level data. In Table 6 age 18 is the most common age; however, the sample is in general much older than that. Overall the sample consists mostly of individuals in their 40s, although not many subjects have the exact same age in that small range.

The next measure of central tendency is median. The median is the value that is in the exact middle of the sample. The median is the point at which half of the subjects lie above this value and half of the subjects lie below it. For example, the ages of the nine subjects in sample one of Table 7 are arranged in numerical order. The fifth age, 45, is in the exact middle of the sample. Consequently age

45 is the median. The median is a better measure of central tendency because it is not influenced by an accidental grouping of values away from the true center of the data. However, the median cannot be determined for nominal level data because no order is present within the data.

The mean (or average) is the most commonly seen measure of central tendency. The mean is calculated by adding up the value for all subjects and dividing by the total number of subjects (n). For sample one in Table 7, the mean is 38.3.

The mean is more sensitive to outliers and more influenced by the distribution of the values than is the median. In Table 7 two data sets are demonstrated. Both have the same median but have very different means. Consequently, different pieces of information are available with the two measures and one or both may be relevant to your research.

The use of means to describe a dataset should be limited to interval and ratio level data. Nominal level data do not have a true numerical value, so it is not possible to compute a mean. Although ordinal data might have numerical values, the intervals between the values do not have to be the same, therefore the mean would be difficult to interpret. If one takes the three ordinal measures of 1, 2, and 3, the relative value between 1 and 2 may not be the same as the relative value between 2 and 3. If these intervals are uneven, the true average may be higher or lower than 2, depending on which interval actually is greatest.

Measures of central tendency provide

information on where most data lie. However, these measures do not inform the reader of how far apart the data are or their variability from one subject to the next. The distribution of the data can be described by the width of the distribution (range and/or variance) and the location of the middle of the distribution (central). In Table 8 the two sets of data have the same mean, but the variance, or spread of the values, is greatly dissimilar.

One possible distribution is the normal distribution. A normal distribution is typically described as being bell shaped and has a middle that is exactly in the center of the distribution. In addition, the tails (sides) of the distribution are symmetric or have the exact same shape (Figure 3). A normal distribution of data in the population is a common assumption for inferential statistics.

One measure of variability is the range. Range is the difference between the greatest value and the smallest value. For the example in Table 8 the range in sample one is 57 and in sample two is 29. The range informs the reader that one set of data is more spread out (more variance) than the other.

The range, like the mean of a sample, is very sensitive to outliers, or measurements that are greatly different than the rest of the sample. In Table 9 the range is wide mostly because of the influence of only two values, 18 and 74. The rest of the values are clustered between 30 and 50, so a range of 57 could be considered misleading.

A measure of variability that mini-

mizes the effects of outliers is standard deviation. A standard deviation is a mathematical calculation of the variance of all the measurements in a sample. The standard deviation can be viewed as the average distance from the mean that each of the values lies. The mathematical equations for calculating the standard deviation can be found in the reference list[3] and is easily performed by basic statistical software programs.

Hypothesis Testing

Hypothesis testing is what most of statistical analysis for individual research projects really is, the method for determining the probability of an observed event occurring only by chance. If chance was not the cause of an event, something else must be (e.g., your treatment had an effect on the outcome). By convention, the null hypothesis is always the hypothesis that is tested. The null hypothesis states that there is no difference between the groups that you are comparing. If you have a control and a treatment group, the null hypothesis states that the dependent variable (the outcome–for example, mortality) is the same for both the control and treatment groups.

An example of hypothesis testing would be when evaluating the effect of a drug for controlling hypertension. If one group of subjects received the investigational drug and the second received a placebo (sugar capsule), the investigator could examine blood pressure as the outcome variable. The blood pressure of an individual is related to a variety of factors so will not be constant over time. Blood pressure for a given individual may vary naturally (or only by chance) 10 or more mm Hg from time one to time two. So, if the treatment group has a 10 mm Hg drop in blood pressure, the investigator needs to ask is this due to normal variation or is the drop due to the effect of the drug. Hypothesis testing assists the investigator in answering this question.

When doing hypothesis testing, one can make one of two types of errors. One can conclude that there is a difference between the two groups when there actually is no difference or one can conclude that no difference exists between the two groups when there actually is a difference (fail to measure a difference). The first error is called a type I (or alpha)

error and the second is called a type II (or beta) error.

Type I and Type II errors are important to the investigator because he or she can make decisions in the study design and data analysis that will decrease one or the other of these errors. This is particularly important because the investigator cannot know whether he or she has made either of these errors when interpreting the data. An additional difficulty is that as the chance of a type I error decreases the chance of a type II error increases and vice versa.[5]

Type I errors reflect overconfidence in a finding of differences between two groups. The frequency of type I errors is directly related to the alpha set by the investigator. Alpha is the statistical term for the p value that the investigator sets as being significant. For example, if the investigator sets his or her alpha at .05, all p values obtained below .05 will be considered statistically significant. This means that with the results obtained there is only a 1 in 20 chance that the results would have been obtained if there was no difference between the groups in the population.

The investigator is free to choose whatever alpha (p value) appears appropriate for a particular study. The higher the p value, the easier it is to obtain statistical significance; however, the greater the chance that a type I error has been made. Consequently, with a high alpha the investigator may claim that the treatment had an effect when in reality there was no effect and the differences that resulted were actually seen as a result of chance.

Type II errors occur when the investigator did not find a difference between two or more groups that was actually present. The frequency of type II errors is inversely related to the frequency of type I errors and is inversely related to the alpha set by the investigator. The lower the alpha, the higher the probability of a type II error. The chance of making a type II error is referred to as beta. Power is the inverse of beta and is the chance of finding a difference that is actually there. Power is directly related to the number of subjects in the sample.[4]

Because of the influence that the alpha level has on both type I and type II errors, this value must be set carefully.

Table 9

Example of Wide Range

18
30
33
40
42
43
50
74

Range = 56

An alpha that is too high may result in statistical significance when no actual difference was present in the data. However, if the alpha is set too low, real differences in the data may remain undetected. The most common values for alpha are .05 and .01.[3]

The investigator should determine the appropriate alpha on the basis of the consequences of making either a type I or a type II error. If the investigator is examining a new treatment for a fatal disease, he or she may wish to find anything that might actually help patients. Consequently, the investigator would rather make a type I error, say something worked when it really didn't, and set the alpha at .05 or even .10. However, if a new drug for a minor disease had potentially harmful side effects, the investigator would rather say that it was ineffective even if the drug actually was effective (type II error). In this case, the investigator would set the alpha at a lower level (.01). Other reasons for adjusting the alpha related to the specific type of statistical analysis are beyond the scope of this article.

Inferential Statistics

Inferential statistics are based on probability theory[3] and the process of hypothesis testing as described above. In contrast to descriptive statistics, inferential statistics are calculated with the purpose of generalizing the findings to the entire population of interest. The investigator wants to know whether differences between groups found in his or her sample are unique to this sample (because of chance) or are due to real differences between the population represented by group 1 and the population represented by group 2 (or however many groups are involved). Inferential statistics are fre-

Chi-square Example*

Intubation Success

Neuromuscular Blockade

Frequency Percent Row Pct Col Pct	0	1	Total
0	46 6.40 71.88 10.98	18 2.50 28.13 6.00	64 8.90
1	373 51.88 56.95 89.02	282 39.22 43.05 94.00	655 91.10
Total	419 58.28	300 41.72	719 100.00

Frequency Missing = 3

Statistics for Table of Success by Neuromuscular Blockade

Statistic	DF	Value	Prob
Chi-square	1	5.344	0.021

Effective sample size = 719
Frequency missing = 3

*This analysis is for demonstration purposes only.

quently used to examine the relationship between two or more variables.

Inferential statistics can be classified as either parametric or nonparametric. Nonparametric statistics are most commonly used for variables at the nominal or ordinal level of measurement, which basically means that they are used for variables that do not have a normal distribution. Statistical significance is calculated using information contained only in the sample (rather than the population) and may use measures of central tendency appropriate for nominal or ordinal level data (i.e., the median rather than the mean).

One of the more common nonparametric statistics is chi-square. Chi-square is used when both the independent and dependent variables are nominal or ordinal level measures. A chi-square can be used to evaluate the data obtained in a contingency table (bivariate frequency distribution, see Table 5). The chi-square

procedure calculates the expected number of observations in each cell of the table and compares them with the number of observations actually occurring in each cell. The greater the deviation of the observed frequencies from the expected frequencies, the greater the chance for statistical significance.

In Table 10 a statistically significant relationship between intubation success and use of neuromuscular blockade exists (even though the relationship is not perfect). The chi-square is 5.3 and is statistically significant at the .05 level because the p value of the chi-square is 0.02, well below the .05 cutoff level.

Other nonparametric statistics used with health sciences research include Fisher's exact test, the Mann-Whitney U test, and the Kruskal-Wallis test. For further information on these procedures, consult a nonparametric statistical reference.[6]

Parametric statistics are the most

common approach to inferential statistical analysis.[3] Parametric statistics require that the variables be measured at the interval or ratio level. Use of parametric statistics also relies on other assumptions such as the expectation that values for a given variable will be normally distributed (Figure 3) in the population.

The most common statistic used to describe the relationship (the correlation) between two variables is a Pearson's r. A Pearson's r is a descriptive statistic when used only to describe a relationship. A Pearson's r is an inferential statistic when used to infer a relationship in the population.

The Pearson's r requires that both variables be measured at least at the interval level of measurement. Consequently, a correlation between role and age is not appropriate, even if role is assigned a numerical value (1 = nurse, 2 = physician).

The Pearson's r ranges from -1.0 to

Figure 4

Perfect Positive Relationship
(r = +1.0)
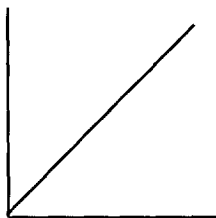
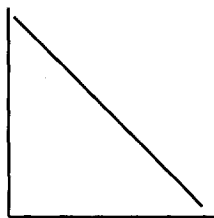**Perfect Negative Relationship**
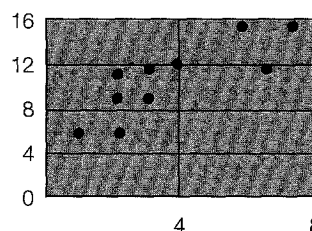**(r = -1.0)**

Figure 5

Less than Perfect Correlation
(r = +.82)

+1.0. A +1.0 indicates a perfect direct (positive) relationship and a -1.0 indicates a perfect inverse (negative) relationship. A Pearson's r of +1.0 or -1.0 are both represented as a straight line when displayed graphically (Figure 4). The further the data are from a perfect relationship, the more the points spread out from a straight line (Figure 5).

Calculation of the correlation between two interval or ratio level variables (Pearson's r) is a common inferential statistic. A Pearson's r is considered an inferential statistic when the investigator is using the value to determine the strength of a relationship in the population (rather than in the sample) and to determine whether the relationship in the population is greater than zero. (Less commonly the r found in the sample can be compared with other possible r's in the population. A discussion of this analysis is beyond the scope of this article.)

Table 11 is an example of a correlation used as an inferential statistic. This analysis is examining the relationship between scene time and flight time. The analysis of 609 subjects indicates that a weak positive (r = +.37) relationship exists between scene time and flight time. The p value indicates that this is a statistically significant relationship (p = .0001).

A t test is a frequently used parametric statistic. The t test is used to evaluate differences in means between two groups (a dichotomous independent variable). When using a t test, the dependent variable must be measured at the interval or ratio level.

A common use of t tests is to determine whether differences exist between an experimental and control group. For example, in an investigation of transport times a t test can be used to determine

Table 11

Correlation Analysis (Pearson's r)

	Scene Time	Flight Time	
Scene time	1.00000	0.37142	Pearson's r
	0.0	0.0001	p value
	792	609	No. of subjects
Flight time	0.37142	1.00000	
	0.0001	0.0	
	609	647	

This analysis is for demonstration purposes only.

Table 12

T Test Procedure

Dependent Variable: Flight Time

Diagnosis	N	Mean	SD	SE
Nontrauma	220	25.29	31.64	2.13
Trauma	426	18.75	25.73	1.24

T	DF	Prob>|T|
2.82	644.0	0.0049

whether a difference exists in flight time for trauma and nontrauma patients. In Table 12, as in many standard tables for a t test, the mean (flight time) of both groups is listed. After the mean is the t value for the analysis (t = 2.82), the degrees of freedom (df = 644), and the p value (p = 0.0049). The t is the actual statistic that is calculated, degrees of freedom represent the conditions under which the t is calculated (related to numbers of subjects, number of groups, and the statistic), and the p value is the probability of whether the differences seen in the two flight times (25.29 minutes and 18.75 minutes) is present because of a true difference (in population means) or because of chance (seen only in this sample, not the population). In the example given in Table 12 a true effect of the diag-

nosis on flight times exists because a statistically significant difference exists between the two flight times. This interpretation is made because the p value is < .05, the alpha selected for this investigation.

Some investigators prefer to give an exact p rather than indicating whether the p is less than or greater than the chosen alpha. This approach to statistical significance has become more common as computer programs have begun to provide an exact p value. Previously, researchers were required to use statistical tables to determine whether the t value provided by the computer was greater than or less than the threshold for a given alpha. Consequently, an exact p value could not be determined. Because alpha should be determined before data

analysis and is selected on the basis of the desired chance for making a type I or type II error, the investigator should interpret the obtained statistic (the t in the case of a t test) in relationship to the selected alpha. Providing an exact p value may obscure the importance of the concept of alpha, particularly to readers with less statistical knowledge. However, some investigators do argue that providing an exact p allows the reader to select their own alpha and to interpret the data in a manner the reader, rather than the author, chooses.

An analysis of variance (ANOVA) is slightly more complex than a t test but is based on the same mathematical principles. In fact, when an ANOVA is calculated for a two-group independent variable, the results are exactly the same as the results obtained with a t test. Although ANOVA can be used with two groups, it is most commonly used for independent variables that have three or more groups. Again, the dependent variable must be measured at the interval or ratio level.

Questions commonly analyzed with an ANOVA also are group difference questions. For example, is there a difference in systolic blood pressure for the control group, the group with drug A, the group with drug B, or the group with both drugs A and B? In this example the independent variable has four possible values.

The data printout obtained for an ANOVA is much more complex than that for a t test. Explanation of the meaning of sum of squares and other statistical concepts is beyond the scope of this article. The interested reader can refer to the reference list for further information on this topic.[7] This article will concentrate on the information necessary to make a basic interpretation of the results.

In the case of an ANOVA the statistic calculated is an F, rather than a t (although F is the square of t if a two-group ANOVA is done). The F statistic is the value compared against the table (or used in calculations) for determination of the significance level (p value). In the example in Table 13, the F is 7.97, resulting in a p value of .0049. The p is interpreted exactly the same as for a t test; p is either more than or less than the selected alpha. In Table 13 p is again less than .05 so is interpreted as statistically significant.

A statistically significant p value for an ANOVA tells the researcher that there are group differences present in the data. The analysis does not indicate which groups are different. Thus, in an analysis of four treatment groups, if group four has a much greater blood pressure than any of the other three groups, the investigator cannot assume that this specific difference is statistically significant. Further analysis (post hoc analysis) is needed to determine the nature of the differences.

Post hoc analysis is beyond the scope of this article. The reader may wish to consult a statistician once statistically significant findings are obtained in a primary analysis or may wish to consult other references on the topic.[7]

The final statistical procedure to be discussed in this paper is regression analysis. Simple regression analysis is used for a single independent variable. Multiple regression analysis is used for multiple independent variables. Regression analysis is computationally related to ANOVA but is used for independent variables that are measured at the interval or ratio level rather than at the nominal or ordinal level. Interval or ratio level independent variables can be broken down into ordinal categories and analyzed using ANOVA (i.e., ages 20 to 39 = 1, ages 40 to 59 = 2, ages 60 to 79 = 3). However, because a great deal of information is lost with this approach, regression analysis is preferred for interval or ratio level independent variables.

An example of a question to be analyzed by regression analysis is whether weight, height, smoking status, and group (ran or didn't run immediately before taking pulse) influence pulse rate. This analysis is presented in Table 14. As with an ANOVA, many of the complex concepts in the table will be left for more in-depth discussions.[7] The important numbers to note are the R^2 (0.43), the F statistic (16.575) and the p value

Table 14

Regression Analysis

Model: Model 1
Dependent variable: Pulse

Analysis of Variance

Source	DF	Sum of squares	Mean square	F value	Prob> F
Model	4	11499.95240	2874.98810	16.575	0.0001
Error	87	15090.04760	173.44882		
C Total	91	26590.00000			

Root MSE	13.17000	R-square	0.4325	
Dep mean	80.00000	Adj R-sq	0.4064	
C.V.	16.46250			

Parameter Estimates

Variable	DF	Parameter estimate	Standard error	T for H0: Parameter = 0	Prob > ITI
Intercep	1	180.152470	33.71931619	5.343	0.0001
Weight	1	-0.169399	0.09710304	-1.745	0.0846
Height	1	-0.512214	0.62052319	-0.825	0.4114
Smokes	1	-2.085548	3.09064040	-0.675	0.5016
Ran	1	-22.740362	2.91323638	-7.806	0.0001

*This analysis is for demonstration purposes only.

(0.0001). Although some computer printouts provide an R^2 for an ANOVA, R^2 is traditionally associated with regression analysis. R^2 is the percentage of variance in the dependent variable explained by the independent variable. Another way to say this is that the R^2 is the percentage of overlap between the independent and dependent variables. By convention a lowercase r2 is used when referring to a single independent variable and an uppercase R^2 is used for multiple independent variables.[3] The F statistic is the same as for ANOVA, as is the p value. The R^2 is the square of the R that would be obtained if a correlation analysis was computed for the independent and dependent variables.

In the example in Table 14 the R^2 is 0.4325, indicating that about 43% of the pulse can be explained by weight, height, smoking status, and whether the subject ran. The p value is less than .05, indicating that this is a statistically significant degree of explanation. Although not strictly true, an increase in the R^2 tends to increase the chance of statistical significance.

The common statistical analyses discussed in this article are summarized in Table 15. The reader should consult other resources for further information on advanced statistics.[7] For example, ANOVA also can be used with multiple independent variables (multifactor ANOVA, two-way ANOVA). Canonical correlation and multivariate analysis of variance (MANOVA) are designed to analyze questions involving multiple independent and dependent variables.

Interpretation

The preceding discussion focused on the mechanics of statistical analysis. This section will focus more on the fuzzier aspects of statistical analysis, statistical reporting and interpretation.

Although limited somewhat by journal requirements, authors have some leeway in how they present the results of their statistical analysis. This article has provided several tables depicting a standard format for results reporting. The reader is advised to examine several journals for statistical output formats to find one that is most appealing and understandable (unless restricted by journal).

The author also may elect to place results only within the text of the manuscript. Although tables often are easier to read, simple results may not merit the space required for a table or figure. When reporting findings in text, the author must report the names of the independent and dependent variables, the statistical procedure used, the statistic calculated (t or F), the value of the statistic, appropriate degrees of freedom (from the computer print out), and the p value (in relation to alpha or actual value). An example for a narrative report of the t test from Table 12 would be as follows: A t test was used for the analysis of research question one. There was a statistically significant relationship between diagnosis and flight time for air transport ($t = 2.82$, $df = 644$, $p < .05$)

A final topic for discussion is the interpretation of statistical output. Knowledge of more than the numbers obtained from the computer is needed to adequately interpret a statistical analysis. The first consideration is clinical versus statistical significance. Not all results that achieve statistical significance have clinical significance. For example, in a study of the effect of drug A on chronic hypertension a statistically significant difference between the control and experimental group was found. However, the average systolic blood pressure for the control group was 172 and the average systolic blood pressure for the experimental group was 168. Although this drop in blood pressure was statistically significant, such a small drop in blood pressure has no clinical significance to the patient. In addition, a larger drop still might be considered clinically insignificant if the cost of the drug was extreme in comparison to a moderate drop in blood pressure. Statistically significant results that

Common Statistical Techniques

Statistical Test	Independent Variable	Dependent Variable	Comments
Nonparametric			
chi-square	Nominal	Nominal	Need > 5 expected subjects/cell
Fisher's exact	Nominal	Nominal	
Mann-Whitney U	Dichotomous	Ordinal	
Kruskal-Wallis	Nominal	Ordinal	3 or more values for independent variable
Parametric			
Pearson's r (correlation)	Interval/ratio	Interval/ratio	
T test	Dichotomous	Interval/ratio	
ANOVA	Nominal/ordinal	Interval/ratio	
MANOVA	Nominal/ordinal	Interval/ratio	Multiple dependent variables
Simple regression	Interval/ratio	Interval/ratio	Single independent variable
Multiple regression	Interval/ratio	Interval/ratio	Multiple independent variables
Canonical correlation	Interval/ratio	Interval/ratio	Multiple independent and dependent variables

do not have clinical significance are more common in studies with large sample sizes. For this reason, researchers often decrease alpha in the case of a large sample so as to diminish this problem (and decrease the chance of a type I error).

On a very rare occasion a study might have clinically significant findings without achieving statistical significance. This most often occurs in the event of small sample sizes. For example, a pilot study of 10 patients, 5 in the control and 5 in the experimental group, might show a large difference in the means of the systolic blood pressure for the two groups. However, large variability in systolic blood pressure between subjects in each group (large standard deviation)

and a small sample size may preclude finding statistical significance. The researcher still may wish to consider the results clinically significant and worthy of further investigation.

A final caution about the interpretation of statistical analyses is that statistics can be misleading. A reader should look to a report of the study to find evidence that many of the decisions that are important to the analysis are made before the study, not after the results are obtained. For example, a researcher should not wait until after obtaining the p value before selecting an alpha.

Other methods for manipulating an analysis of which the reader should be aware include subdividing or combining

groups within the sample to maximize the results. At times there are good reasons to alter group designations (too few subjects per cell), but these possibilities should be discussed before the analysis and not after.

Finally, statistical significance can be found if the investigator looks hard enough. As discussed previously, the p value relates to the probability of finding results only by chance. If an alpha is set at .05 and 20 analyses are computed, at least one of the analyses is likely to be statistically significant merely by chance (5% chance). Because research journals often are more interested in studies that demonstrate statistical significance, an author may be tempted to do multiple exploratory analyses merely to find something that is publishable. In this case, being able to link the review of literature to the research questions and results becomes very important. Such links help to demonstrate that the questions were well thought out in advance and not merely the result of creative data analysis.

Conclusion
We would like to stress that conducting and interpreting a statistical analysis are not impossible. Understanding a few basic concepts will provide a background helpful for interpreting a wide range of statistics even if the details of the math are not understood. The researcher also is reminded to seek out the assistance of a statistician early in the process of research. A statistician can provide direction, education, and reassurance as one grows in statistical sophistication.

References
1. Panacek EA, Thompson C. Basics of research (Part 1). Why conduct clinical research and how to get started? Air Med J 1995;14:33-6.
2. Shott S. Statistics for health professionals. Philadelphia: Saunders, 1990.
3. Polit DF, Hungler BP. Nursing research: principles and methods. 5th ed. Philadelphia: JB Lippincott, 1995.
4. Thompson CB, Panacek EA, Davis E. Basics of research (Part 4). Research study design (Part 2). Air Med J 1995;14:222-31.
5. Polit DF. Data analysis and statistics for nursing research. Stamford, CT: Appleton & Lange, 1996.
6. Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1988.
7. Cohen J, Cohen P. Applied multiple regression/correlation analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlabaum Associates, 1983.