The Ochsner Journal 10:213–216, 2010

Statistics: A Brief Overview

Ryan Winters, MD,* Andrew Winters, BS,† Ronald G. Amedee, MD, FACS‡

**Department of Otolaryngology—Head and Neck Surgery, Tulane University School of Medicine, New Orleans, LA*
†Department of Mathematics, Florida State University, Tallahassee, FL
‡Department of Otolaryngology—Head and Neck Surgery, Ochsner Clinic Foundation, New Orleans, LA

ABSTRACT

The Accreditation Council for Graduate Medical Education sets forth a number of required educational topics that must be addressed in residency and fellowship programs. We sought to provide a primer on some of the important basic statistical concepts to consider when examining the medical literature. It is not essential to understand the exact workings and methodology of every statistical test encountered, but it is necessary to understand selected concepts such as parametric and nonparametric tests, correlation, and numerical versus categorical data. This working knowledge will allow you to spot obvious irregularities in statistical analyses that you encounter.

INTRODUCTION

The Accreditation Council for Graduate Medical Education sets forth a number of required educational topics that must be addressed in residency and fellowship programs. One of these topics is biostatistics, and for many trainees this is a confusing subject. We sought to provide a primer on some of the important basic statistical concepts to consider when examining the medical literature.

We put a great deal of faith in numbers. The underlying concept of evidence-based medicine requires ''evidence'' of a treatment's efficacy or benefit, or of the detriment of a side effect or complication; this evidence comes in numerical form. As our profession relies increasingly on mathematics, we must equip ourselves with a basic grasp of statistical methods in order to interpret the literature.

Address correspondence to:
Ryan Winters, MD
1430 Tulane Avenue, SL-59
New Orleans, LA 70112
Tel: (504) 988-5454
Fax: (504) 988-7846
Email: rwinters@tulane.edu

Key Words: ACGME resident education, mathematics, statistics

At their most basic, statistics are studies of population samples, and the goal is to apply the results from these samples to whole populations. This deceptively simple concept requires several important considerations, such as the design of the study, selection of the study sample, and choice of statistical test. Aberrations in any of these will yield erroneous conclusions about significance or applicability of the results. In the words of Homer Simpson, ''*People can come up with statistics to prove anything…14% of people know that*.'' If, when reading an article, the research question asked seems straightforward, the data collection seems uncomplicated, but the authors have used a statistical test that you have never heard of, be cautious of the conclusions presented.

DEFINITIONS

Mean

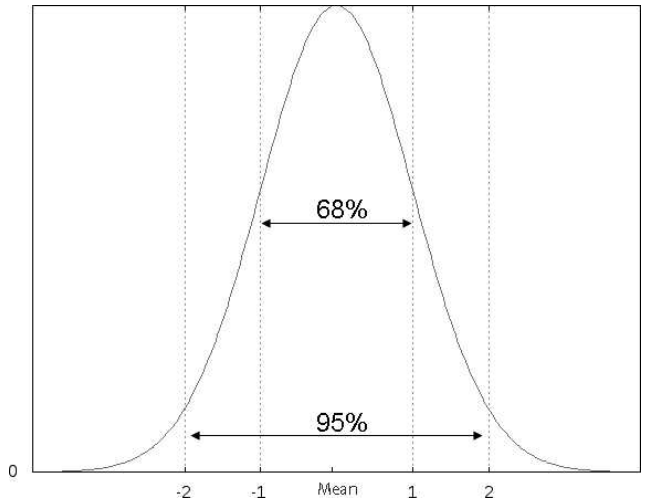
The mean is the arithmetic average of all the values in a set of data. It is often used to approximate the central tendency of a set of data but is very susceptible to outliers. For example, the mean income of Omaha, Nebraska, would likely be inordinately high because of Warren Buffett.

Median

The median is the middle value of a data set when the set is ordered chronologically. It is not influenced by outliers but does not give any indication of the actual values of the numbers in a data set; it is not an average. It is therefore useful for small, highly skewed sets of numbers.

Distribution

When the values of a data set are plotted on a graph, the shape of the resultant curve defines the distribution. The *normal* (also called Gaussian) distribution is the classic bell curve, and in this distribution the mean = median (Figure). This allows several important mathematical assumptions to be made, allowing the use of statistical tests that are very sensitive. One of Fisher's assumptions states that the larger a sample size, the more closely the distribution will approximate a normal distribution. This is one



Normal (Gaussian) distribution. Mean = 0 in this example. The 1 SD on either side of the mean encompasses ~68% of all values under the curve, while 2 SD encompasses 95% of all values under the curve.

mathematical reason why studies with large sample sizes are better.

Standard Deviation

The standard deviation (σ or SD) denotes how far away from the mean an individual value lies. For a normally distributed data set, approximately 68% of the values will lie within 1 SD of the mean, and about 95% of the values will lie within 2 SDs (Figure).

P Value

The *P* value represents the probability that the observed outcome was the result of chance. Arbitrarily in the scientific and medical communities, $P < .05$ has been chosen as the cutoff for ''statistically significant.'' This means that there is a less than 5% possibility (1 chance in 20) that the observed result was from chance alone. A *P* value outside the significant range could indicate one of two things. Either there is no ''real'' difference between the 2 sets of data, or the sample size was too small to detect a difference between the 2 sets, even if it exists. You cannot tell from the *P* value which of these possibilities is at fault, however.

Confidence Interval

A confidence interval (CI) is a range of values within which it is fairly certain the true value lies. This is based on the idea that if you were to repeat the exact same study on random groups of subjects multiple times, you would not get the exact same results each time. You would have a range. For the purposes of interpreting the medical literature, 95% CI is talked about frequently. What this represents mathematically, for a given statistical result, is that we can be 95% certain that the true value lies within the range denoted by the CI (ie, within 2 SD). The narrower the range of the CI, the closer any observed value is to the true value, and the more precise the result. Confidence intervals can be positive or negative numbers, and this has no implications. If the 95% CI includes zero as well as values of practical importance, however, then the result is not statistically significant, regardless of the *P* value.[1] This applies to inferential statistics and is slightly different for odds ratios, which are not discussed in this review. Mathematically, zero means there is no difference between the sample and the true population value. Therefore, the possibility of no difference between the 2 groups has not been excluded, despite $P < .05$. It is also paramount to remember that such calculations serve only to determine the mathematical validity of results; they do not determine the clinical utility of said results. Determining such utility is a fascinating topic unto itself and is well beyond this brief overview of statistics.

TYPES OF DATA

Numbers are assigned to many different things in the annals of research, but they fall broadly into 2 categories, each amenable to different statistical tests.

Numerical Data

The number itself has relevance. Examples of numerical data are things that are directly measured as a number such as CD4 counts, low-density lipoprotein levels, blood alcohol content, and tumor size.

Categorical Data

Numbers that are assigned to nonnumerical values of interest represent categorical data. Examples of categorical data include the city of residence or the sex of study participants.

TYPES OF STATISTICAL TESTS

There are 2 broad categories of statistical tests.

Parametric Tests

Parametric tests assume that sample data come from a set with a particular distribution, typically from a normal distribution. Generally, this requires a large sample size. Because the distribution is known (mathematically, because the shape of the curve is known), parametric tests are able to examine absolute differences between individual values in a sample and are more powerful. They are able to identify smaller differences than are nonparametric tests and should be used whenever possible.

Table 1. For Each Parametric Statistical Test There Is an Analogous Nonparametric Test

Parametric	Nonparametric
Chi-square test[a]	Fisher exact test
Paired Student *t* test	Wilcoxon signed rank test
Unpaired Student *t* test	Mann-Whitney *U* test
ANOVA by sum of squares	ANOVA by rank
Pearson product moment coefficient	Spearman rank correlation coefficient

ANOVA: analysis of variance.

[a] Chi-square is a nonparametric test. Some authors propose thinking of it as parametric, as it works with the sample distribution, mathematically speaking.

Table 2. Hypothetical Contingency Table Comparing Estrogen Receptor/Progesterone Receptor (ER/PR) Status With Metastasis[a]

Status	ER/PR+	ER/PR−	Totals
Metastasis+	a	b	a + b
Metastasis−	c	d	c + d
Totals	a + c	b + d	a + b + c + d

[a] To form a contingency table, each category must have a yes/no or positive/negative (+/−) answer.

Nonparametric Tests

Nonparametric tests make no such assumptions about the distribution of originating data and therefore must ignore absolute values of data points and focus instead on ordinal properties (eg, which is smallest, which is most common). It is more difficult to demonstrate statistical significance with a nonparametric test (ie, the difference between the 2 groups must be larger) than with a parametric test.

For each parametric statistical test there is a nonparametric analog to be used when conditions for parametric analysis cannot be met (Table 1).

TESTS

Chi-square

Mathematically speaking, the chi-square (χ^2) test is a nonparametric test. Practically, it requires large sample sizes and should not be used when numbers are less than 20. The chi-square test is used with categorical data, and actual tally numbers must be used, not percentages or means. It tests the distribution of 2 or more independent data sets compared with a theoretical distribution. The more alike the distributions are, the more related they are determined to be, and the larger the chi-square value. A value of $\chi^2 = 0$ implies there is no relationship between the samples.

Fisher Exact Test

Analogous to the chi-square test, the Fisher exact test is a nonparametric test for categorical data but can be used in situations in which the chi-square test cannot, such as with small sample sizes. This test is used when comparing 2 data sets that create a contingency table (Table 2) and tests the association (contingency) between the 2 criteria. This is observed when each data set has a ''yes/no'' answer, such as tumor cells present or absent, blood cultures positive or negative, breast cancer specimens that are estrogen receptor/progesterone receptor positive or negative, and so forth.

Student *t* Test

The Student *t* test is likely the most widely used test for statistical significance and is a parametric test suitable for either numerical or categorical data. The test compares the means of 2 data sets to determine if they are equal; if they are, then no difference exists between the sets. It exists as both a *paired* and *unpaired* test. A paired test means that the same thing was measured on each subject twice. For example, you measured each subject's heart rate, administered a beta blocker, then measured each person's heart rate again and want to compare the difference before and after administration of the drug. If this was not done, use an unpaired test.

Wilcoxon Signed Rank Test and Mann-Whitney *U* Test

The Wilcoxon signed rank test and Mann-Whitney *U* test are nonparametric analogs of the paired and unpaired *t* tests, respectively, in many situations. There are specific scenarios in which other tests are used as nonparametric analogs of the Student *t* test, such as when examining survival time or when examining categorical data, which are beyond the scope of this review. If these scenarios are encountered, it is advisable to examine closely the references listed in the study to determine the reason for the choice of test used. These tests analyze whether the median of the 2 data sets is equal or if the sample sets are drawn from the same population. As with all nonparametric tests, they have less power than the parametric counterpart (Student *t* test) but can be used with small samples or nonnormally distributed data.

Analysis of Variance

Analysis of variance (ANOVA or F test) is a generalization of the Student *t* test (or Wilcoxon or Mann-Whitney *U* test) when 3 or more data sets are being

compared. There are both parametric and nonparametric analyses of variance referred to as ANOVA by sum of squares or ANOVA by rank, respectively.

Pearson Product Moment Correlation Coefficient

Perhaps the most misused and misunderstood of all statistical analyses is the Pearson product moment correlation coefficient test. Fundamentally, correlation is a departure from independence of 2 or more variables. This can be in the form of any relationship, positive or negative. Correlation does not equal causation, nor does it imply causation; it merely records the fact that 2 or more variables are not completely independent of one another. Pearson coefficient (r) is a parametric test that can be used with numerical or categorical data, but it is only meaningful under very select circumstances. For it to be a valid measure, all of the following 4 criteria must be met. (1) The data must be normally distributed. (2) The 2 data sets must be independent of one another. One value should not automatically vary with the other. For example, number of drinks consumed and blood alcohol level are not independent of one another; you must drink alcohol to change your blood alcohol level. (3) Only a single pair of measurements should be made on each subject. (4) Every r value calculated should be accompanied by a P value and/or CI.[2]

Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient test (r_s or ρ) is the nonparametric counterpart to the Pearson coefficient and is a good option when all of the criteria cannot be met to calculate a meaningful r value and numerical data are being examined. As a nonparametric test, the degree of departure from independence will have to be greater to reach significance using this test than it would with the Pearson test. It is also more laborious to calculate, although in the modern era of statistical software this is largely irrelevant.

Regression Analysis

Regression quantifies the numerical relationship between 2 variables that are correlated. Essentially, it creates an equation wherein if one variable is known, the other can be estimated. This process, called *extrapolation*, should be done cautiously, and likewise interpreted cautiously. The numerous reasons for this are beyond the scope of this introduction.

DISCUSSION

When examining the statistics section of an article, it is helpful to have a systematic approach. It is not essential to understand the exact workings and methodology of every statistical test encountered, but it is necessary to understand selected concepts such as parametric and nonparametric tests, correlation, and numerical versus categorical data. This working knowledge will allow you to spot obvious irregularities in statistical analyses that you encounter. Greenhalgh[2,3] proposes asking the following questions, among others, as a first pass of any article: (1) Were the 2 groups evaluated for comparability at baseline? (2) Does the test chosen reflect the type of data presented (parametric vs nonparametric, categorical vs numerical)? (3) Have the data been analyzed according to the original study protocol? (4) If an obscure test was used (essentially any test not mentioned in this review), was an explanation and a reference provided?

For further reading, an excellent series of articles from the *Canadian Medical Association Journal*[4-7] delves into greater depths and explores some additional concepts of statistics. A little solid foundational knowledge and a systematic approach are invaluable in using the medical literature and help make the idea of ''statistics'' a little less daunting.

REFERENCES

1. Blume J, Peipert JF. What your statistician never told you about *P*-values. *J Am Assoc Gynecol Laparosc*. 2003;10(4):439-444.
2. Greenhalgh T. How to read a paper: statistics for the non-statistician, I: different types of data need different statistical tests [erratum appears in *BMJ*. 1997;315(7109):675]. *BMJ*. 1997;315:364-366.
3. Greenhalgh T. How to read a paper: statistics for the non-statistician, II: ''significant'' relations and their pitfalls. *BMJ*. 1997;315:422-425.
4. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 1. hypothesis testing. *CMAJ*. 1995;152(1):27-32.
5. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 2. interpreting study results: confidence intervals. *CMAJ*. 1995;152(2):169-173.
6. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 3. assessing the effects of treatment: measures of association. *CMAJ*. 1995;152(3):351-357.
7. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 4. correlation and regression. *CMAJ*. 1995;152(4):497-504.