Basics of statistics for primary care research

Timothy C Guetterman

ABSTRACT

The purpose of this article is to provide an accessible introduction to foundational statistical procedures and present the steps of data analysis to address research questions and meet standards for scientific rigour. It is aimed at individuals new to research with less familiarity with statistics, or anyone interested in reviewing basic statistics. After examining a brief overview of foundational statistical techniques, for example, differences between descriptive and inferential statistics, the article illustrates 10 steps in conducting statistical analysis with examples of each. The following are the general steps for statistical analysis: (1) formulate a hypothesis, (2) select an appropriate statistical test, (3) conduct a power analysis, (4) prepare data for analysis, (5) start with descriptive statistics, (6) check assumptions of tests, (7) run the analysis, (8) examine the statistical model, (9) report the results and (10) evaluate threats to validity of the statistical analysis. Researchers in family medicine and community health can follow specific steps to ensure a systematic and rigorous analysis.

Family Medicine, University of Michigan, Michigan Medicine, Ann Arbor, Michigan, USA

Correspondence to
Dr Timothy C Guetterman;
tguetter@med.umich.edu

Investigators in family medicine and community health often employ quantitative research to address aims that examine trends, relationships among variables or comparisons of groups (Fetters, 2019, this issue). Quantitative research involves collecting structured or closed-ended data, typically in the form of numbers, and analysing that numeric data to address research questions and test hypotheses. Research hypotheses provide a proposition about the expected outcome of research that may be assessed using a variety of methodologies, while statistical hypotheses are specific statements about propositions that can only be tested statistically. Statistical analysis requires a series of steps beginning with formulating hypotheses and selecting appropriate statistical tests. After preparing data for analysis, researchers then proceed with the actual statistical analysis and finally report and interpret the results.

Family medicine and community health researchers often limit their analyses to descriptive statistics—reporting frequencies, means and standard deviation (SD). While sometimes an appropriate stopping point, researchers may be missing opportunities for more advanced analyses. For example, knowing that patients have favourable attitudes about a treatment may be important and can be addressed with descriptive statistics. On the other hand, finding that attitudes are different (or not) between men and women and that difference is statistically significant may give even more actionable information to healthcare professionals. The latter question, about differences, can be addressed through inferential statistical tests. The purpose of this article is to provide an accessible introduction to foundational statistical procedures and present the steps of data analysis to address research questions and meet standards for scientific rigour. It is aimed at individuals new to research with less familiarity with statistics and may be helpful information when reading research or conducting peer review.

FOUNDATIONAL STATISTICAL TECHNIQUES

Statistical analysis is a method of aggregating numeric data and drawing inferences about variables. Statistical procedures may be broadly classified into (1) statistics that describe data—descriptive statistics; and (2) statistics that make inferences about more general situations beyond the actual data set—inferential statistics.

Descriptive statistics

Descriptive statistics aggregate data that are grouped into variables to examine typical values and the spread of values for each variable in a data set. Statistics summarising typical values are referred to as measures of central tendency and include the mean, median and mode. The spread of values is represented through measures of variability, including the variance, SD and range. Together, descriptive statistics provide indicators of the distribution of data, or the frequency of values through the data set as in a histogram plot. Table 1 summarises commonly used descriptive statistics. For consistency, I use the terms independent variable and dependent variable,

Table 1 Descriptive statistics

Statistic	Statistic	Description of calculation	Intent
Measures of central tendency	Mean	Total of values divided by the number of values.	Describe all responses with the average value.
	Median	Arrange all values in order and determine the halfway point.	Determine the middle value among all values, which is important when dealing with extreme outliers.
	Mode	Examine all values and determine which one appears most frequently.	Describe the most common value.
Measures of variability	Variance	Calculate the difference of each value from the mean, square this difference score, sum all of the squared difference scores and divide by the number of values minus 1.	Provide an indicator of spread.
	Standard deviation	Square root of variance.	Give an indicator of spread by reporting on average how much values differ from the mean.
	Range	The difference between the maximum and minimum value.	Give a very general indicator of spread.
	Frequencies	Count the number of occurrences of each value.	Provide a distribution of how many times each value occurs.

but in some fields and types of research such as correlational studies the preferred terms may be predictor and outcome variable. An *independent variable* influences, affects or predicts a *dependent variable*.

Inferential statistics: comparing groups with t tests and ANOVA

Inferential statistics are another broad category of techniques that go beyond describing a data set. Inferential statistics can help researchers draw conclusions from a sample to a population.[1] We can use inferential statistics to examine differences among groups and the relationships among variables. Table 2 presents a menu of common, fundamental inferential tests. Remember that even more complex statistics rely on these as a foundation.

The t test is used to compare two group means by determining whether group differences are likely to have occurred randomly by chance or systematically indicating a real difference. Two common forms are the independent samples t test, which compares means of two unrelated groups, such as means for a treatment group relative to a control group, and the paired samples t test, which compares means of related groups, such as the pretest and

post-test scores for the same individuals before and after a treatment. A t test is essentially determining whether the difference in means between groups is larger than the variability within the groups themselves.

Another fundamental set of inferential statistics falls under the general linear model and includes analysis of variance (ANOVA), correlation and regression. To determine whether group means are different, use the t test or the ANOVA. Note that the t test is limited to two groups, but the ANOVA is applicable to two or more groups. For example, an ANOVA could examine whether a primary outcome measure—dependent variable—is significantly different for groups assigned to one of three different interventions. The ANOVA result comes in an *F* statistic along with a p value or confidence interval (CI), which tells whether there is some significant difference among groups. We then need to use other statistics (eg, planned comparisons or a Bonferroni comparison, to give two possibilities) to determine which of those groups are significantly different from one another. Planned comparisons are established before conducting the analysis to contrast the groups, while other tests like the Bonferroni comparison are conducted post-hoc (ie, after analysis).

Examining relationships using correlation and regression

The general linear model contains two other major methods of analysis, correlation and regression. Correlation reveals whether values between two variables tend to systematically change together. Correlation analysis has three general outcomes: (1) the two variables rise and fall together; (2) as values in one variable rise, the other falls; and (3) the two variables do not appear to be systematically related. To make those determinations, we use the correlation coefficient (r) and related p value or CI. First, use the p value or CI, as compared with established significance criteria (eg, $p < 0.05$), to determine whether a

Table 2 Inferential statistics

Statistic	Intent
t tests	Compare groups to examine whether means between two groups are statistically significant.
Analysis of variance	Compare groups to examine whether means among two or more groups are statistically significant.
Correlation	Examine whether there is a relationship or association between two or more variables.
Regression	Examine how one or more variables predict another variable.

relationship is even statistically significant. If it is not, stop as there is no point in looking at the coefficients. If so, move to the correlation coefficient.

A correlation coefficient provides two very important pieces of information—the strength and direction of the relationship. An r statistic can range from −1.0 to +1.0. Strength is determined by how close the value is to −1.0 or 1.0. Either extreme indicates a perfect relationship, while a value of 0 indicates no relationship. Cohen provides guidance for interpretations: 0.1 is a weak correlation, 0.3 is a medium correlation and 0.5 is a large correlation.[1 2] These interpretations must be considered in the context of the study and relative to the literature. The valence (+ or −) coefficient reveals the direction of the relationship. A negative correlation means one value rises, while the other tends to fall, and a positive coefficient means that the values of the two variables tend to rise and fall together.

Regression adds an additional layer beyond correlation that allows predicting one value from another. Assume we are trying to predict a dependent variable (Y) from an independent variable (X). Simple linear regression gives an equation ($Y = b_0 + b_1X$) for a line that we can use to predict one value from another. The three major components of that prediction are the constant (ie, the intercept represented by b_0), the systematic explanation of variation (b_1), and the error, which is a residual value not accounted for in the equation[3] but available as part of our regression output. To assess a regression model (ie, model fit), examine key pieces of the regression output: (1) *F* statistic and its significance to determine whether the model systematically accounts for variance in the dependent variable; (2) the r square value for a measure of how much variance in the dependent variable is accounted for by the model; (3) the significance of coefficients for each independent variable in the model; and (4) residuals to examine random error in the model. Other factors, such as outliers, are potentially important (see Field[4]).

The aforementioned inferential tests are foundational to many other advanced statistics that are beyond the scope of this article. Inferential tests rely on foundational assumptions, including that data are normally distributed, observations are independent, and generally that our dependent or outcome variable is continuous. When data do not meet these assumptions, we turn to non-parametric statistics (see Field[4]).

A brief history of foundational statistics

Prominent statisticians Karl Pearson and Ronald A Fisher developed and popularised many of the basic statistics that remain a foundation for statistics today. Fisher's ideas formed the basis of null hypothesis significance testing that sets a criterion for confidence or probability of an event.[4] Among his contributions, Fisher also developed the ANOVA. Pearson's correlation coefficient provides a way to examine whether two variables are related. The correlation coefficient is denoted by r for a relationship between two variables or R for relationships among more than two variables as in multiple correlation or regression.[4] William Gosset developed the t distribution and later the t test as a way to examine whether two values of means were statistically different.[5]

Statistical software

While the aforementioned statistics can be calculated manually, researchers typically use statistical software that process data, calculate statistics and p values, and supply a summary output from the analysis. However, the programs still require an informed researcher to run the correct analysis and interpret the output. Several available programs include SAS, Stata, SPSS and R. Try using the programs through a demonstration or trial period before deciding which one to use. It also helps to know or have access to others using the program should you have questions.

Example study

The remainder of this article presents steps in statistical analysis that apply to many techniques. A recently published study on communication skills to break bad news to a patient with cancer provides an exemplar to illustrate these steps.[6] In that study, the team examined the validity of a competence assessment of communication skills, hypothesising that after receiving training, post-test scores would be statistically improved from pretest scores on the same measure. Another analysis was to examine pretest sensitisation, tested through a hypothesis that a group randomly assigned to receive a pretest and post-test would not be significantly different from a post-test-only group. To test the hypotheses, Guetterman *et al*[6] examined whether mean differences were statistically significant by applying t tests and ANOVA.

STEPS IN STATISTICAL ANALYSIS

Statistical analysis might be considered in 10 related steps. These steps assume necessary background activities, such as conducting literature review and writing clear research question or aims, are already complete.

Step 1. Formulate a hypothesis to test

In statistical analysis, we test hypotheses. Therefore, it is necessary to formulate hypotheses that are testable. A hypothesis is specific, detailed and congruent with statistical procedures. A null hypothesis gives a prediction and typically uses words like 'no difference' or 'no association'.[7] For example, we may hypothesise that group means on a certain measure are not significantly different and test that with an ANOVA or t-test. For example, in the exemplar study, one of the hypotheses was 'MPathic-VR scores will improve (decreased score reflects better performance) from the preseminar test to the postseminar test based on exposure to the [breaking bad news] BBN intervention' (p508), which was tested with a t test.[6] Hypotheses about relationships among variables could be tested with correlation and regression. Ultimately,

hypotheses are driven by the purpose or aims of a study and further subdivide the purpose or aims into aspects that are specific and testable. When forming hypotheses, a concern is that having too many dependent variables leads to multiple tests of the same data set. This concern, called multiple comparisons or multiplicity, can inflate the likelihood of finding a significant relationship when none exists. Conducting fewer tests and adjusting the p value are ways to mitigate the concern.

Step 2. Select a test to run based on research questions or hypotheses

The statistical test must match the intended hypothesis and research question. Descriptive statistics allow us to examine trends limited to typical values, spread of values and distributions of data. ANOVAs and t tests are methods to test whether means are statistically different among groups and what those differences are. In the exemplar study, the authors used paired samples t-tests for pre–post scores with the same individuals and independent t tests for differences among groups.[6]

Correlation is a method to examine whether two or more variables are related to one another, and regression extends that idea by allowing us to fit a line to make predictions about one variable based on a linear relationship to another. These statistical tests alone do not determine cause and effect, but merely associations. Causal inferences can only be made with certain research designs (eg, experiments) and perhaps with advanced statistical techniques (eg, propensity score analysis). Table 3 provides guidance for determining which statistical test to use.

Step 3. Conduct a power analysis to determine a sample size

Before conducting analysis, we need to ensure that we will have an adequate sample size to detect an effect. Sample size relates to the concept of power. For example, to detect a small effect, a larger sample is needed. Larger sample sizes can thus detect a smaller effect. Sample size is determined through a power analysis. The determination of sample size is never a simple percent of the population, but a calculated number based on the planned statistical tests, significance level and effect size.[8] I recommend

Table 3 Choosing and interpreting statistics for studies common in primary care

I want to	Statistical choice	Independent variable	Dependent variable	How to interpret
Examine trends or distributions.	Descriptive statistics	Categorical or continuous	Categorical or continuous	Report the statistic as is to describe the data set.
Compare group means.	t tests	Categorical with two levels (ie, two groups)	Continuous	Examine the t statistic and significance level. If significant, clearly report which group mean is higher, along with the effect size.
Compare group means.	Analysis of variance	Categorical with two or more levels (ie, two or more groups)	Continuous	Examine the F statistic and significance level. If significant, clearly report which group means are significantly different and how (eg, which are higher), along with the effect size.
Examine whether variables are associated.	Correlation	Continuous	Continuous	Examine the r statistic and significance level. If significant, describe whether a positive or negative correlation and its strength.
Gain a detailed understanding of the association of variables and use one or more variables to predict another.	Regression	Continuous or categorical, may have more than one independent variable in multiple regression	Continuous	Examine the F statistic and significance level. If significant, examine the R square for how much variance the model accounts for. Determine whether each regression coefficient is significant; if significant, discuss the coefficients.

using G*Power for basic power calculations, although many other options are available. In the exemplar study, the authors did not report their power analysis prior to conducting the study, but they gave a post-hoc power analysis of the actual power based on their sample size and the effect size detected.[6]

Step 4. Prepare data for analysis

Data often need cleaning and other preparation before conducting analysis. Problems requiring cleaning include values outside of an acceptable range and missing values. Any particular value could be wrong because of a data entry error or data collection problem. Visually inspecting data can reveal anomalies. For example, an age value of 200 is clearly an error, or a value of 9 on a 1–5 Likert-type scale is an error. An easy way to start inspecting data is to sort each variable by ascending values and then descending values to look for atypical values. Then, try to correct the problem by determining what the value should be. Missing values are a more complicated problem because a concern is why the value is missing. A few missing values at random is not necessarily a concern, but a pattern of missing values (eg, individuals from a specific ethnic group tend to skip a certain question) indicates a systematic missingness that could indicate a problem with the data collection instrument. Descriptive statistics are an additional way to check for errors and ensure data are ready for analysis. While not discussed in the communication assessment exemplar, the authors did prepare data for analysis and report missing values in their descriptive statistics.

Step 5. Always start with descriptive statistics

Before running inferential statistics, it is critical to first describe the data. Obtaining descriptive statistics is a way to check whether data are ready for further analysis. Descriptive statistics give a general sense of trends and can illuminate errors by reviewing frequencies, minimums and maximums that can indicate values outside of the accepted range. Descriptive statistics are also an important step to check whether we meet assumptions for statistical tests. In a quantitative study, descriptive statistics also inform the first table of the results that reports information about the sample, as seen in table 2 of the exemplar study.[6]

Step 6. Check assumptions of statistical tests

All statistical tests rely on foundational assumptions. Although some tests are more robust to violations, checking assumptions indicates whether the test is likely to be valid for a particular data set. Foundational parametric statistics (eg, t tests, ANOVA, correlation, regression) assume independent observations and a normal linear distribution of data. In the exemplar study, the authors noted 'Data from both groups met normality assumptions, based on the Shapiro–Wilk test' (p508), and gave the statistics in addition to noting specific assumptions for the independent t tests around equality of variances.[6]

Step 7. Run the analysis

Conducting the analysis involves running whatever tests were planned. Statistics may be calculated manually or using software like SPSS, Stata, SAS or R. Statistical software provides an output with key tests statistics, p values that indicate whether a result is likely systematic or random, and indicators of fit. In the exemplar study, the authors noted they used SPSS V.22.[6]

Step 8. Examine how well the statistical model fits

The first step involves examining whether the statistical model was significant or a good fit. For t tests, ANOVAs, correlation and regression, first examine an overall test of significance. For a t test, if the t statistic is not statistically significant (eg, p>0.05 or a CI crossing 0), we can conclude no significant difference between groups. The communication assessment exemplar reports significance of the t tests along with measures such as equality of variance.

For an ANOVA, if the F statistic is not statistically significant (eg, p>0.05 or a CI crossing 0), we can conclude no significant difference between groups and stop because there is no point in further examining what groups may be different. If the F statistic is significant in an ANOVA, we can then use contrasts or post-hoc tests to examine what is different. For a correlation test, if the r value is not statistically significant (eg, p>0.05 or a CI crossing 0), we can stop because there is no point in looking at the magnitude or direction of the coefficient. If it is significant, we can proceed to interpret the r. Finally, for a regression, we can examine the F statistic as an omnibus test and its significance. If it is not significant, we can stop. If it is significant, then examine the p value of each independent variable and residuals.

Step 9. Report the results of statistical analysis

When writing statistical results, always start with descriptive statistics and note whether assumptions for tests were met. When reporting inferential statistical tests, give the statistic itself (eg, a F statistic), the measure of significance (p value or CI), the effect size and a brief written interpretation of the statistical test. The interpretation, for example, could note that an intervention was not significantly different from the control or that it was associated with improvement that was statistically significant. For example, the exemplar study gives the pre–post means along standard error, t statistic, p value and an interpretation that postseminar means were lower, along with a reminder to the reader that lower is better.[6]

When writing for a journal, follow the journal's style. Many styles italicise non-Greek statistics (eg, the p value), but follow the particular instructions given. Remember a p value can never be 0 even though some statistical programs round the p to 0. In that case, most styles prefer to report as p<0.001.

Table 4 Threats to statistical conclusion validity

Threat	Description
Low statistical power (see step 3)	The sample size is not adequate to detect an effect.
Violated assumptions of statistical tests (see step 6)	The data violate assumptions needed for the test, such as normality.
Fishing and error rates	Repeated tests of the same data (eg, multiple comparisons) increase chances of errors in conclusions.
Unreliability of measures	Error in measurement or instruments can artificially inflate or decrease apparent relationships among variables.
Restricted range	Statistics can be biased by limited outcome values (eg, high/low only) or floor or ceiling effects in which participants scores are clustered around high or low values.
Unreliability of treatment implementation	In experiments, unstandardised or inconsistent implementation affects conclusions about correlation.
Extraneous variance in an experiment	The setting of a study can introduce error.
Heterogeneity of units	As participants differ within conditions, standard deviation can increase and introduce error, making it harder to detect effects.
Inaccurate effect size estimation	Outliers or incorrect effect size calculations (eg, a continuous measure for a dichotomous dependent variable) can skew measures of effect.

Step 10. Evaluate threats to statistical conclusion validity

Shadish *et al*[9] provide nine threats to statistical conclusion validity in drawing inferences about the relationship between two variables; the threats can broadly apply to many statistical analyses. Although it helps to consider and anticipate these threats when designing a research study, some only arise after data collection and analysis. Threats to statistical conclusion validity appear in table 4.[9] Pertinent threats can be dealt with to the extent possible (eg, if assumptions were not met, select another test) and should be discussed as limitations in the research report. For example, in the exemplar study, the authors noted the sample size as a limitation but reported that a post-hoc power analysis found adequate power.[6]

Resources

Key resources to learn more about statistics include Field[4] and Salkind[10] for foundational information. For advanced statistics, Hair *et al*[11] and Tabachnick and Fidell[12] provide detailed information on multivariate statistics. Finally, the University of California Los

Angeles Institute for Digital Research and Education (stats.idre.ucla.edu/other/annotatedoutput/) provides annotated output from Stata, SAS, Stata and MPlus for many statistical tests to help researchers read the output and understand what it means.

CONCLUSION

Researchers in family medicine and community health often conduct statistical analyses to address research questions. Following specific steps ensures a systematic and rigorous analysis. Knowledge of these essential statistical procedures will equip family medicine and community health researchers with interpreting literature, reviewing literature and conducting appropriate statistical analysis of their quantitative data.

Nevertheless, I gently remind you that the steps are interrelated, and statistics is not only a consideration at the end of data collection. When designing a quantitative study, investigators should remember that statistics is based on distributions, meaning statistics works with aggregated numerical data and relies on variance within that data to test statistical hypotheses about group differences, relationships or trends. Statistics provides a broad view, based on these distributions, which brings implications at the early design phase. In designing a quantitative study, the nature of statistics generally suggests a larger number of participants in the research (ie, a larger n) to have adequate power to detect statistical significance and draw valid conclusions. Therefore, it will likely be helpful for researchers to include a biostatistician as early as possible in the research team when designing a study.

REFERENCES

1. Trochim WM. Research methods knowledge base, 2006. Available: http://www.socialresearchmethods.net/kb
2. Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
3. Darlington RB, Hayes AF. *Regression analysis and linear models: concepts, applications, and implementation*. New York, NY: Guilford Publications, 2016.
4. Field A. *Discovering statistics using IBM SPSS statistics*. 4th edn. London, UK: SAGE, 2013.
5. Guinness BJF. Gosset, Fisher, and small samples. *Statistical Science* 1987;2:45–52.
6. Guetterman TC, Kron FW, Campbell TC, *et al*. Initial construct validity evidence of a virtual human application for competency assessment

in breaking bad news to a cancer patient. *Adv Med Educ Pract* 2017;8:505–12.

7. Creswell JW, Guetterman TC. *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. 6th edn. Boston, MA: Pearson, 2019.

8. Murnane RJ, Willett JB. *Methods matter: improving causal inference in educational and social science research*. New York, NY: Oxford University, 2010.

9. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin, 2002.

10. Salkind NJ. *Statistics for people who (think they) hate statistics. thousand oaks*. CA: Sage, 2016.

11. Hair J, Black W, Babin B, *et al*. *Multivariate data analysis*. 8th edn. Upper Saddle River, NJ: Centgage, 2018.

12. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 6th edn. Boston, MA: Allyn & Bacon/Pearson Education, 2012.