



Feature Article

A Statistical Primer: Understanding Descriptive and Inferential Statistics

Gillian Byrne
Information Services Librarian
Queen Elizabeth II Library
Memorial University of Newfoundland
St. John's, NL, Canada
Email: gbyrne@mun.ca

Received: 13 December 2006

Accepted: 08 February 2007

© 2007 Byrne. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

As libraries and librarians move more towards evidence-based decision making, the data being generated in libraries is growing. Understanding the basics of statistical analysis is crucial for evidence-based practice (EBP), in order to correctly design and analyze research as well as to evaluate the research of others. This article covers the fundamentals of descriptive and inferential statistics, from hypothesis construction to sampling to common statistical techniques including chi-square, correlation, and analysis of variance (ANOVA).

Introduction

Much of the research done by librarians, from bibliometrics to surveys to usability testing, requires the measurement of certain factors. This measurement results in numbers, or data, being collected, which must then be analyzed using quantitative research methods. A basic understanding of statistical techniques is essential to properly designing research, as well as accurately evaluating the research of others.

This paper will introduce basic statistical principles, such as hypothesis construction and sampling, as well as descriptive and inferential statistical techniques. Descriptive statistics describe, or summarize, data, while inferential statistics use methods to infer conclusions about a population from a sample.

In order to illustrate the techniques being

	Great Job	Lousy Job
If you accept the job	Have a great experience	Waste time & effort
If you decline the job	Waste an opportunity	Avoid wasting time & effort

Figure 1. Illustration of Type I & II errors.

described here, an example of a fictional article will be used. Entitled *Perceptions of Evidence-Based Practice: A Survey of Canadian Librarians*, this article uses various quantitative methods to determine how Canadian librarians feel about Evidence-based Practice (EBP). It is important to note that this article, and the statistics derived from it, is entirely fictional.

Hypothesis

Hypotheses can be defined as “untested statements that specify a relationship between two or more variables” (Nardi 36). In social sciences research, hypotheses are often phrased as research questions. In plain language, hypotheses are statements of what you want to prove (or disprove) in your study. Many hypotheses can be constructed for a single research study, as you can see from the example in Fig. 1.

In research, two hypotheses are constructed for each research question. The first is the null hypothesis. The null hypothesis (represented as H_0) assumes no relationship between variables; thus it is usually phrased as “this has no affect on this”. The alternative hypothesis (represented as H_1) is simply stating the opposite, that “this has an affect on this.” The null hypothesis is generally the one constructed for scientific research.

Type I & II Errors

Anytime you make a decision in life, there is a possibility of two things going wrong. Take the example of a job offer. If you

decide to take the job and it turned out to be lousy, you would have wasted a lot of time and energy. However, if you decided to pass on the job and it was great, you would have wasted an opportunity. It's best illustrated by a two by two box (Fig. 1).

It is obvious that, despite thorough research about the position (speaking to people that work there, interview process, etc.), it is possible to come to the wrong conclusion about the job. The same possibility occurs in research. If your research concludes that there is a relationship between variables when in fact there is no relationship (i.e., you've incorrectly assumed the alternative hypothesis is proven), this is a Type I error. If your research concludes that there is no relationship between the variables when in fact there is (i.e., you've incorrectly assumed the null hypothesis is proven), this is a Type II error. Another way to think of Type I & II errors is as false positives and false negatives. Type I error is a false positive, like concluding the job is great when it's lousy. A Type II error is a false negative; concluding the job is lousy when it's great.

Type I errors are considered by researchers to be more dangerous. This is because concluding there is a relationship between variables when there is not can lead to more extreme consequences. A drug trial illustrates this well. Concluding falsely that a drug can help could lead to the drug being put on the market without being beneficial to the public. A Type II error would lead to a promising drug being left off the market,

which while serious, isn't considered as dire. To help remember this, think of the conservative nature of science. Inaction (and possibly more testing) is less dangerous than action. Thus, disproving the null hypothesis, which supposes no relationship, is preferred to proving the alternative hypothesis.

There are many safety features built in to research methodology which help minimize the possibility of committing both errors, including sampling techniques and statistical significance, both of which you will learn about later.

Dependent and Independent Variables

Understanding hypotheses help you determine which variables are dependent and which are independent (why this is important will be revealed a bit later). Essentially it works like this: the dependent variable (DV) is what you are measuring, while the independent variable (IV) is the cause, or predictor, of what is being measured.

In experimental research (research done in controlled conditions like a lab), there is usually only one hypothesis, and determining the variables are relatively simple. For example, in drug trials, the dosage is the independent variable (what the researcher is manipulating) while the effects are dependent variables (what the researcher is measuring).

In non-experimental research (research which takes place in the 'real world', such as survey research), determining your dependent variable(s) is less straightforward. The same variable can be considered independent for one hypothesis while dependent for another. An example – you might hypothesize that hours spent in the library (independent variable) are a predictor of grade point average (dependent variable). You might also hypothesize that

major (independent variable) affects how much time students spend in the library (dependent variable). Thus, your hypothesis construction dictates your dependent and independent variables.

A final variable to be aware of in quantitative research is the confounding variable (CV). Also known as lurking variables, a confounding variable is an unacknowledged factor in an experiment which might affect the relationship between the other variables. The classic example of a confounding example affecting an assumption of a relationship is that murder rates and ice cream purchased are highly correlated (when murder rates go up, so does the purchase of ice cream?). What is the relationship? There isn't one; both variables are affected by a third, unacknowledged variable: hot weather.

Population, Samples & Sampling

Although it is possible to study an entire population (censuses are examples of this), in research samples are normally drawn from the population to make experiments feasible. The results of the study are then generalized to the population. Obviously, it is important to choose your sample wisely!

Population

This might seem obvious, but the first step is to carefully determine the characteristics of the population about which you wish to learn. For example, if your research involves your university, it is worthwhile to investigate the basic demographic features of the institution; i.e., what is the percentage of undergraduate students vs. graduate students? Males vs. females? If you think these are groups you would like to compare in your study, you must ensure they are properly represented in your sample.

Sampling Techniques

Probability Sampling

Probability sampling means that each member of the population has an equal chance of being selected for the survey. There are several flavors of probability sampling; the common characteristic being that in order to perform probability sampling you must be able to identify all members of your population

Random sampling is the most basic form of probability sampling. It involves identifying every member of a population (often by assigning each a number), and then selecting sample subjects by randomly choosing numbers. This is often done by computer programs.

Stratified random sampling ensures the sample matches the population on characteristics important to a study. Using the example of a university, you might separate your population into graduate students and undergraduate students, and then randomly sample each group separately. This will ensure that if your university has 70% undergraduates and 30% graduates, your sample will have a similar ratio.

Cluster sampling is used when a population is spread over a large geographic region. For example, if you are studying librarians who work at public libraries in Canada, you might randomly sample 50 libraries, and then randomly sample the librarians within those libraries.

Non-probability Sampling

Simply put, this is any sampling technique that does not involve random sampling. Often samples are not random because in some research it is easier to perform *convenience sampling* (surveying those who volunteer, for example). Also, sometimes the population from which the sample is to be taken cannot be easily identified. A common strategy employed by libraries is to

use patron records to derive random samples. This is probability sampling only if the population is library users; if the population is an entire institution or city, it is no longer random. With non-probability samples, you can only generalize to those who participated, not to a population.

Sample Size

Sample size is also extremely important to be able to accurately generalize to a population. Generally, the bigger the sample, the better. The *Central Limit Theorem* states that the larger the sample, the more likely the distribution of the means will be normal, and therefore population characteristics can more accurately be predicted. Some other things to keep in mind:

- If you want to compare groups with each other (for example, majors), you will need at least 5 subjects in each group to do many statistical analyses.
- Poor response rate can severely compromise a study, if surveys are involved. Depending on the distribution method, response rate can be as low as 10% (ideally you want a response rate over 70%) (Weisberg 119). Ensure your sample size is large enough to still provide accurate results with a poor response rate.

There is no magic formula to determine the proper sample size – it depends on the complexity of your research, how homogenous the population is, and time and human resources you have available to compile and analyze data.

Descriptive Statistics

Once you have performed your research and gathered data, you need to perform

A clear understanding of librarians' perceptions of EBP is necessary to inform the development of systems to support EBP in librarianship.

The following research questions were posed:

1. What are the perceptions of librarians of EBP?
2. Does institution type the librarian works at affect perception?
3. Does length of service of the librarian affect perception?

What are the hypotheses?

There are three being provided. Here is a rephrasing of number 3:

H_0 = "Length of service of librarians has no affect on the perception of EBP"

H_1 = "Length of service of librarians affects the perception of EBP"

What are the Type I & II error possibilities?

		The real situation (in the population)	
		H_0 is true	H_1 is true
Result of Research (from sample):	H_0 is proven (length of service doesn't affect perception)	No error	Type II error
	H_1 is proven (length of service does affect perception)	Type I error	No error

What are the dependent and independent variables?

The researchers are attempting to determine whether length of service can predict perception of EBP, or to rephrase, is perception of EBP dependant on length of service. Therefore:

Dependent variables: perception of EBP

Independent variable: length of service

Table 1. Examples of hypotheses.

data analysis. Choosing the appropriate statistical method for the data is crucial. The bad news is, this means you have to know a

whole lot about your data – is it nominal, ordinal or ratio? Is it normally distributed? Let's start from the very beginning.

Levels of Measurement

Nominal variables are measured at the most basic level. They are discrete levels of measurement where a number represents a category (i.e., 1 = male; 2 = female), but these numbers do not imply order and mathematical calculations cannot be performed on them. You could just as easily say, 1 = male and 36,000 = female - this doesn't mean that females are 35,999 times bigger or better than males! Nominal variables are of the least use statistically.

Ordinal variables are also discrete categories, but there is an order to the categories; they increase and decrease at regular intervals. A

good example is a Likert scale: 1 = very poor; 2 = poor; 3 = average, etc. In this example, you can state 1 is 'less' or 'smaller' or 'worse' than 2. The disadvantage of ordinal variables is that you cannot measure in between the values. You do not know how much worse 1 is than 2.

Ratio (sometimes known as scale, continuous or interval) variables are the most robust, statistically, of variable types. Ratio variables have natural order, and the distance between the points in the same. Think of pounds on a scale. You know that

The sampling frame was the database of all librarians (defined as those who hold an MLS) who were members of the Canadian Library Association in March 2005. A total of 5,683 librarians were on the list. The list was divided up by type of library worked at (academic, public, school, special, and other / not stated). A proportional random sample of 210 was then selected. This ensured that even at a return rate of 40% a final sample size of 150 would be achieved.

Is this a random sample?

On first glance, yes. However, this is only a true random sample if all librarians in Canada belonged to the Canadian Library Association. The design of this study means that the results can only be generalized to Canadian Library Association members, not to Canadian librarians.

What sampling technique is used?

This survey used stratified random sampling to ensure that all types of librarians would be represented, as illustrated in the chart below. Please remember that all values in this table are for demonstration purposes and do not accurately reflect reality.

	Academic Librarians	Public Librarians	School Librarians	Special Librarians	Other / Not Stated	Totals
Real Proportion	1136 (20%)	2273 (40%)	568 (10%)	582 (15%)	582 (15%)	5683
Sample Size	42 (20%)	84 (40%)	21 (10%)	31 (15%)	31 (15%)	210

Table 2. Examples of sampling.

100 is lighter than 101. You also know that 101 is 1 pound heavier than 100. Finally the scale is continuous; it is possible to weigh 100.58 pounds. The power of the ratio variable is important to keep in mind for your study. For example, rather than asking subjects to tick off an age category in a box, you can ask them to fill in their age. This gives you the freedom to keep it as a ratio variable, or to round the ages up into

appropriate ordinal values.

Measures of Central Tendency

The theory of normal distribution tells us that, if you tested an entire population, the result (parameter) would look like a bell curve, with the majority of values grouped in the middle. A good example of this would be scores on test.

Selection of variables used in the study

Variable Name	Variable Label	Values
TYPE	Type of library worked at	1 = academic, 2 = public...
LENGTH	Length of service	
INCOME	Income of respondent	1 = under 30,000, 2 = 31,000-40,000...
AGE	Age of respondent	
EBP_AWARE	Answer to the question I have heard of EBP	1 = yes, 2 = no
EBP_SCORE	Score on the EBP Perceptions Test	

What level of measurement is TYPE?

TYPE is a nominal measurement. The numbers represent types of libraries, but no mathematical calculations can be performed on them. EBP_AWARE is also a nominal measurement.

What level of measurement is LENGTH?

Because there are no values set for LENGTH it is a ratio variable. Each librarian's length of service will be entered in years. EBP_SCORE and AGE are also ratio variables.

What level of measurement is INCOME?

INCOME is an ordinal variable. It has numbers representing categories, but there is a clear ranking. Librarians in category one earn less money than librarians in category two.

Table 3. Examples of variables.

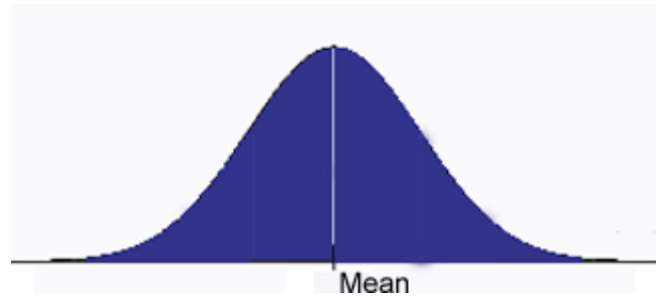


Figure 2. Normal distribution of a bell curve.

However, when moving from parameters to statistics, there is the probability that the results will not reflect the population, and thus not be normally distributed. Measures of Central Tendency provide you with information about how your results are grouped. There are three measures, and which one to use depends on what level of measurement the variable is.

Mean (represented by M or μ) is the most commonly referred to measure of central tendency. It is the average measure, where each value is added, and then the sum divided by the number of cases. However, it should be quite clear that the mean cannot be used with nominal and ordinal variables. Imagine again a Likert scale. The mean value might be 2.36, but what does that tell you? That the average respondent falls somewhere closer to “I found this difficult” than “I have no opinion”?

Median (represented by Mdn) is the measure commonly used with ordinal data. The median is the halfway point of the data. To calculate simply order your values from lowest to highest and see at what value half the data is below, and half is above. The median is also an extremely valuable measure for ratio data when there are outliers (think how the average income variable would be skewed in a town with one multimillionaire). This is because median is not affected by how far away from the middle values are, just the quantity

of them. The median for 2, 2, 3, 4, 4 is 3; the median for 2, 2, 3, 4, 10 is also 3.

Mode is often used with nominal data (though it can also be calculated for other variable types). It is simply the most frequently occurring value in a dataset. An example of when this would be an appropriate measure is for major. The average major makes no sense, nor does the halfway point major, but the most frequently occurring major does.

Measures of Spread

Measures of central tendency reveal much about data, but not the whole story. You also need to know how the values are spread across the spectrum. Measures of spread will tell us whether the values are clustered around the mean or more spread out. Think of test scores; one group might all score 70, while another group's score might range from 60-90. In this case, it is possible that the mean, median and mode would be the same, but we can see the distribution is quite different. There are three main statistical methods for determining spread.

Range is the most basic measure; it is calculated simply by subtracting the lowest score from the highest score. However, this is not the most accurate method as the range can be skewed by outlier values (a very high or very low score).

Interquartile range is less likely to be distorted by outliers, as it is calculated by ordering the sample from highest to lowest, then dividing the sample into four equal quarters (percentiles). The median is then calculated for each quartile. Subtracting the median of the first quartile from the third quartile obtains the interquartile range.

Standard deviation (represented by *SD* or σ) is the most sophisticated measure of spread, and a widely used statistical concept.

Statistical software will easily calculate standard deviation, so the formula will not be covered here. Because standard deviation relies on calculations of the mean it can only be used with continuous variables. A standard deviation score of 0 indicates that there is no variation of values. The higher the standard deviation, the larger the spread.

Bivariate Analysis

At heart of all research is an interest in determining relationships between variables.

Characteristics of the variable AGE

Age of Respondent		
N		210
Mean		44.05
Median		43
Mode		33
Std Deviation		12.77
Range		38
Percentiles	25	33
	50	43
	75	56.50

What does this tell us about the central tendencies of the data?

The average age of librarian respondent to this survey is 44.05. Half of the librarians were over 43, while other half were under 43. The most commonly occurring age was 33.

What does this tell about the spread of the data?

We can tell something about spread simply by looking at the difference between mean, median and mode. The fact that the mean is slightly higher than the median and much higher than the mode indicates that there are some older respondents skewing the data.

The range indicates that there are 38 years between oldest and youngest respondent. This large value could be due to the outliers at the upper end of the scale. However, the large standard deviation also indicates a wide spread of values. This is not surprising, as logically in any profession, there is likely to be a wide variety of ages.

Table 4. An example of measures of central tendency and measures of spread.

There are many statistical methods for exploring those relationships, which ones to choose are often dependent on the type of variables with which you are working (nominal, ordinal or ratio). It is also important to understand statistical significance (the extent to which the relationship can be generalized to the population) and effect size (the strength of the relationship) with bivariate analysis techniques.

Statistical Significance

Comprehending inferential statistics requires a clear understanding of what is meant by statistical significance. For something to be statistically significant, it is unlikely to have occurred by chance (remember that every time you are dealing with a sample you are taking the chance that your results will not reflect the population). Another way of putting it is that significance tests denote how large the possibility is that you are committing a Type I error. Significance tests are affected by the strength of relationship between variables and the size of the sample. Common levels of significance (represented by *alpha*, or α) are 5%, 1% and 0.1%; if $\alpha = .01$, you are stating that there is a one in one thousand chance this happened by coincidence.

Cross Tabulation

What is a cross tabulation?

Essentially a cross tabulation (*cross tab*) is a table in which each cell represents a unique combination of values. This allows you to visually analyze whether one variable's distribution is contingent on another's.

When would you use a cross tabulation?

Cross tabulations can be used to show relationships between two nominal variables, nominal and ordinal variables, or two ordinal variables. It can be used with

ratio data, as long as the variable has a limited number of values. Limitations of the cross tabulation

Cross tabulations provide you with a visual view of comparative data, but because they display simple values and percentages, there is no way to gauge whether any differences in the distribution are statistically significant.

Chi-Square

What is a chi-square?

A chi-square is a test which looks at each cell in a cross tabulation and measures the difference between what was observed and what would be expected in the general population. It is used to evaluate whether there is a relationship between the values in the rows and columns of a cross tab, and the likelihood that any differences can be put down to chance.

When would you use a chi-square?

Chi-square is one of the most important statistics when you are assessing the relationship between ordinal and/or nominal measures.

Are there limitations of using chi-square?

Chi-square cannot be used if any cell has an expected frequency of zero, or a negative integer. It can be affected by low frequencies in cells; if many of your cells have a frequency of less than 5, the chi-square test might be compromised.

How do I know if the relationship is statistically significant?

The chi-square test provides a significance value called a *p-value*. The p-value is compared to α , which can be set at different levels. If $\alpha = .05$, then a p score less than .05 indicates statistical significant differences, a p score greater than .05 means that there is no statistical difference.

Cross tabulation of type of library and I have heard the
term evidence-based practice

	Yes	No	Total
Academic Library			
Count	30	12	42
Percentage	71.42%	28.58%	100%
Public Library			
Count	54	30	84
Percentage	64.28%	35.72%	100%
School Library			
Count	9	12	21
Percentage	42.86%	57.14%	100%
Special Library			
Count	22	10	31
Percentage	70.96%	29.04%	100%
Other/Not Stated			
Count	20	11	31
Percentage	64.51%	35.49%	100%
Total Count	100	110	210

What does this table tell us?

This table allows us to see the numbers of librarians who have heard of the term Evidence-based Practice broken down by type of library worked at. As you can see, there are some differences between the groups; a smaller percentage of school librarians have heard of EBP (42.86%, N = 9) than other type of librarians. There is no indication from this table, however, if that difference is statistically significant.

Table 5. Example of cross tabulation.

T-test

What is a t-test?

A t-test compares the means between two values. It tests whether any differences in the means are statistically significant or can be explained by chance.

When do you use a t-test?

T-tests are normally used when comparing differences between two groups (i.e., undergraduates versus graduates) or in a before and after situation (student achievement before versus after library

instruction). A t-test involves means, therefore the dependent variable (the variable you are attempting to measure) must be a ratio variable. The independent variable is nominal or ordinal.

Limitations of the t-test

A t-test can only be used to analyze the means of two groups. For more than two groups, use ANOVA.

How do I know if the relationship is statistically significant?

Like the chi-square test, the t-test provides a significance value called a p-value, and is presented the same way.

Correlation Coefficients

What are correlation coefficients?

Correlation coefficients measure the strength of association between two variables, and reveal whether the correlation is negative or positive. A negative relationship means that when one variable

increases the other decreases (e.g., drinking alcohol and reaction time). A positive relationship means that when one variable increases so does the other (e.g., study time and test scores). Correlation scores range from -1 (strong negative correlation) to 1 (strong positive correlation). The closer the figure is to zero, the weaker the association, regardless whether it is a negative or positive integer.

A chi-square statistic was then performed to determine if type of library worked at affected whether librarians had heard the term evidence-based practice. As you can see by the table below, $p > .05$, therefore there is no statistical difference in distribution of awareness of EBP based on the type of library worked at.

	Value	Df	Sig.
Chi-Square	16.955	4	.990

Why use a chi-square?

A chi-square is the statistic being used here because the relationship between two ordinal variables (type of library worked at and awareness of the term EBP) is being explored.

What does *value* mean?

It is simply the mathematical calculation of the chi-square. It is used to then derive the p-value, or significance.

What does *df* mean?

Df stands for degrees of freedom. Degrees of freedom is the number of values that can vary in the estimation of a parameter. It is calculated for the chi-square statistic by looking at the cross tabulation and multiplying the number of rows minus one by the number of columns minus one $(r-1) \times (c-1)$. In this case, if we look back to Fig. 4, we can see that we have a two by five table. Thus, $(2-1) \times (5-1) = 4$.

What does *sig.* mean?

Sig. stands for significance level, or p-level. In this case $p = .990$. As this number is larger than .05, the null hypothesis is proven. There is no statistically relationship between type of library and awareness of EBP, despite the differences in percentages we saw in Table 5.

Table 6. Example of chi-square.

When should you use correlation coefficients?

Correlation coefficients should be used whenever you want to test the strength of a relationship. There are many tests to measure correlation; which one to use depends on what variables you are examining. A few are listed below:

Nominal variables: *Phi*, *Cramer's V*, *Lambda*, *Goodman and Kruskal's Tau*

Ordinal variables: *Gamma*, *Sommers D*, *Spearman's Rho*

Ratio variables: *Pearson r*

Limitations of correlation coefficients

Correlation does not indicate causality.

Simply because there is a relationship between two variables does not mean that one causes the other. Keep in mind correlation only looks at the relationship between two variables; there may be others affecting the relationship (remember the confounding variable!). Correlation

coefficients can also be skewed by outlier values.

How do I know if the relationship is statistically significant?

Correlation scores range from -1 (strong negative correlation) to 1 (strong positive correlation). The closer the figure is to zero, the weaker the association, regardless of whether it is a negative or positive integer.

Analysis of Variance (ANOVA)

What is ANOVA?

Like the t-test, ANOVA compares means, but can be used to compare more than two groups. ANOVA looks at the differences between categories to see if they are larger or smaller than those within categories.

When should you use ANOVA?

The dependent variable in ANOVA must be ratio. The independent variable can be

An independent samples t-test was performed to determine if there was a statistical difference between genders on the Evidence-based Practice test. As the table below illustrates, there was a significant difference in performance between males and females, $t(19) = -.398$ $p < .05$

	Value	df	Sig.
T-test	-.398	19	.049

Why use a t-test?

A t-test is used for these variables because we are comparing the mean of one variable (EPB Test Score, a ratio variable) between 2 groups (sex, a nominal variable). An independent samples t-test is used here because the groups being compared are mutually exclusive - male and female.

How is the t-test interpreted?

The t-test value, degrees of freedom, and significance values can be interpreted in precisely the same way as the chi-square in Fig. 5. The significance value of .049 is less than .05, therefore it can be stated that the null hypothesis is disproved; there is a statistical significant difference between the performance of male librarians and the performance of female librarians on the EBP Perceptions Test.

Table 7. Example of a t-test.

A Pearson r correlation was performed to determine if there was a relationship between age and score on the EBP test instrument. The correlation revealed that the two were significantly related, $r=+.638$, $n=210$, $p<.05$.

Why was a Pearson r correlation performed?

A Pearson r was done because both variables involved, Age and EBP Perceptions Test score, are ratio variables.

What does the r value tell us?

The r is correlation score. Remember that correlation scores range from $+/-1$ to 0 . Therefore, a score of $+.638$ reveals that there is a strong positive correlation between age and EBP score.

The fact that it is positive means that when one variable increases so does the other – the older the librarian, the higher they scored on the EBP test instrument.

Table 8. Example of a Pearson r correlation.

nominal or ordinal, but must be composed of mutually exclusive groups

Limitations of ANOVA

ANOVA measures whether there are significant differences between three or more groups, but it does not illustrate where the significance lies – there could be differences between all groups or only two. There are tests called *post hoc comparisons* which can be performed to determine where significance lies, however.

How do I know if the relationship is statistically significant?

An ANOVA uses an *f-test* to determine if there is a difference between the means of groups. The *f-test* can be used to calculate a *p-score*, which is analyzed in the same way as chi-squares and *t-tests*.

Statistical Significance and Effect Size Measures

Significance tests have a couple of weaknesses. One is the fairly arbitrary value at which statistical significance is said to have occurred. Why is $\alpha = .051$ not a significant finding while $\alpha = .049$ is? The

second disadvantage is that significance tests do not give an indication of the strength of a relationship, merely that it exists. A smaller significance value could be the result of a larger sample rather than a strong relationship. This is where effect sizes come in. Effect sizes are tests which gauge the strength of a relationship. There are many different effect size indices; which to use depends on the statistical test being performed.

Multivariate Analysis

Any in-depth discussion of multivariate analysis is beyond the scope of a paper entitled “Statistical Primer”; however, here is a brief introduction.

Multivariate analysis looks at the relationship between more than two variables, for example length of service and type of librarian might together be predictors of perception of EBP. Using bivariate statistical methods, it is not possible to see the relationship between two independent variables as well as their effect on the dependent variable. There are several multivariate statistical methods. Here are two of the most common.

For the EBP Test Instrument Score, the analysis of variance (ANOVA) revealed that there was not a significant difference in performance $F(3, 47)=3.43$, $p<.05$ between types of librarians. The critical value (.245) for the scores was obtained the F distribution table using $df_{\text{between}}=4$ and $df_{\text{within}}=16$.

Why was an ANOVA performed?

An ANOVA was the appropriate statistical technique because the dependent variable (EBP Test score) is continuous, while the independent variable (type of library worked at) is nominal and composed of several groups.

What does this tell us?

The F test score was calculated at 3.43. This score was used in conjunction with the degrees of freedom (because we are comparing several groups, there are two degrees of freedom scores, one for between the groups (4) and one for within the groups (16) to calculate the p -score. $P = .245$, which is greater than .05. Therefore there is no difference in performance on the test based on the type of library worked at.

Table 9. Example of ANOVA.

Statistical Test	Effect Size Measure	Comments
Chi-square	<i>phi</i>	Phi tests return a value between zero (no relationship) and one (perfect relationship).
T-test	<i>Cohen's d</i>	Cohen's d results are interpreted as 0.2 being a small effect, 0.5 a medium and 0.8 a large effect size. (Cohen 157)
ANOVA	<i>Eta squared</i>	Eta square values range between zero and one, and can be interpreted like phi and Cohen's d .

Table 10. Statistical tests and effect size measures.

Multivariate analysis of variance (MANOVA) is an ANOVA which analyses several dependent variables. It can be interpreted in much the same way as ANOVA tests. MANOVA has advantages over doing multiple ANOVA tests, including reducing the potential for Type I errors (concluding that there is a relationship when there is not). Conversely, MANOVA tests can also reveal relationships not apparent in ANOVA tests. *Multiple linear regression* examines "the relationship between one 'effect' variable,

called the dependent or outcome variable, and one or more predictors, also called independent variables" (Muijs 168). It is designed to work with continuous variables, though there are different techniques available for analyzing other variable types. While performing and analyzing regressions are complicated, they are valuable tools for examining the relationship between many variables. It is important to note that, like other inferential statistical techniques, values are created that provide the statistical significance of the relationships.

Conclusion

This paper is not intended to produce statistical experts. Rather, it is a guide to understanding the basic principles and techniques common in library and related research. Most statistical software packages, such as SPSS or SAS, will effortlessly perform statistics, so it is far more important that as a researcher you know a) how to select an appropriate sample; b) know what statistical technique is appropriate in which situations; and c) be able to interpret results correctly. There are a few things you can do to make yourself more comfortable with statistics. One is to purchase a basic quantitative methods textbook. Look for one that comes with a CD of sample data sets. Running through the exercises in the textbook will provide you with valuable practice in performing and analyzing statistics. There are several textbooks available in the library field, although any social science quantitative methods texts

would be useful. The second thing you can do is to read the research literature in your field. If you know the topic well, it is easier to evaluate and interpret results.

Works cited

- Cohen, J. "A Power Primer." *Psychological Bulletin* 112 (1992): 155-159.
- Muijs, Daniel. *Doing Quantitative Research in Education with SPSS*. London: SAGE, 2004.
- Nardi, Peter M. *Doing Survey Research: A Guide to Quantitative Methods*. Boston: Allyn and Bacon, 2003.
- Weisberg, Herbert F., Jon A. Krosnick, and Bruce D. Bowen. *An Introduction to Survey Research, Polling, and Data Analysis*. 3rd ed. Thousand Oaks, Calif.: Sage Publications, 1996.