Module 5: Data Preparation and Analysis

Preparing Data

After data collection, the researcher must prepare the data to be analyzed. Organizing the data correctly can save a lot of time and prevent mistakes. Most researchers choose to use a database or statistical analysis program (e.g. Microsoft Excel, SPSS) that they can format to fit their needs and organize their data effectively. Once the data has been entered, it is crucial that the researcher check the data for accuracy. This can be accomplished by spot-checking a random assortment of participant data groups, but this method is not as effective as re-entering the data a second time and searching for discrepancies. This method is particularly easy to do when using numerical data because the researcher can simply use the database program to sum the columns of the spreadsheet and then look for differences in the totals. One of the best methods of checking for accuracy is to use a specialized computer program that cross-checks double-entered data for discrepancies.(1)

Descriptive Statistics

Descriptive statistics describe but do not draw conclusions about the data. Each descriptive statistic summarizes multiple discrete data points using a single number. They can tell the researcher the central tendency of the variable, meaning the average score of a participant on a given study measure. The researcher can also determine the distribution of scores on a given study measure, or the range in which scores appear. Additionally, descriptive statistics can be used to tell the researcher the frequency with which certain responses or scores arise on a given study measure. For example, in the Module 1 example about the effectiveness of corrective lenses on economic productivity, the researcher might observe that the average dollars-per-week of a person with corrected vision is $500, whereas the average DPW for a person without corrected vision is $450. This amount of information is not enough information to conclude that vision correction affects economic productivity. Inferential statistics are necessary to draw conclusions of this kind. Descriptive statistics might also tell the researcher that the distribution of DPW is $351-$640 for the whole sample and that the average DPW is $445 for the sample.(2)

Correlation

Correlation is one of the most often used (and most often *mis*used) kinds of descriptive statistics. It is perhaps best described as “a single number that describes the degree of relationship between two variables.”(3) If two variables tend to be “correlated,” that means that a participant’s score on one variable tends to vary with a score on the other. For example, people’s height and shoe size tend to be positively correlated. This means that for the most part, if a person is tall, they are likely to have a large shoe size, and conversely, if they are short, they are likely to have a smaller shoe size. Correlation can also be negative. For example, warmer temperatures outside may be negatively correlated with the number of hot chocolates sold at a local coffee shop. This is to say that as the temperature goes up, hot chocolate sales tend to go

down. Although causality may seem to be implied in this situation, it is important to note that on a statistical level, **correlation does not imply causation**. A good researcher knows that there is no way to assess from *correlation alone* that a causal relationship exists between two variables. In order to assert that "X caused Y," a study should be experimental, with control groups and random sampling procedures. Determining causation is a difficult thing to do, and it is a common mistake to assert a cause-and-effect relationship when the study methodology does not support this assertion.

Inferential Statistics

Inferential statistics allow the researcher to begin making inferences about the hypothesis based on the data collected. This means that, while applying inferential statistics to data, the researcher is coming to conclusions about the population at large. Inferential statistics seek to generalize beyond the data in the study to find patterns that ostensibly exist in the target population. This course will not address the specific types of inferential statistics available to the researcher, but a succinct and very useful summary of them, complete with step-by-step examples and helpful descriptions, is available here.(4)

Statistical Significance

Researchers cannot simply conclude that there is a difference between two groups in a well-constructed study. This difference must be due to the manipulation of the independent variable. No matter how well a researcher designs the study, there always exists a degree of error in the results. This error may be due to individual differences within and between experimental groups, or the error may be due to systematic differences within the researcher's sample. Irrespective of its source, this error acts as "noise" in the data and affects participants' scores on study measures even though it is not the variable of interest. Statistical significance is aimed at determining the probability that the observed result of a study was due to the influence of something other than chance. A result is "statistically significant" at a certain level. For example, a result might be significant at $p<.05$. "P" represents the probability that the result was due to chance, and .05 represents a 5% probability that the result was due to chance. Therefore, in a well-run study, $p<.05$ means that inferential statistical analysis has indicated that the observed results have over a 95% probability of being due to the influence of the independent variable. The 5% cutoff is generally thought of as the standard for most scientific research. Note that it is theoretically impossible to ever be entirely certain that one's results are not due to chance, as the nature of science is one of observing trends and testing hypotheses, not immutable proof.(5)

Go To Module 6: The Importance of Research >>

Footnotes

(1) Trochim, W. M. K. "Data Preparation" *Research Methods Knowledge Base 2nd Edition.* Accessed 2/24/09.

(2) Trochim, W. M. K. "Descriptive Statistics" *Research Methods Knowledge Base 2nd Edition.* Accessed 2/24/09.

(3) Trochim, W. M. K. "Descriptive Statistics" *Research Methods Knowledge Base 2nd Edition.* Accessed 2/24/09.

(4) Trochim, W. M. K. "Inferential Statistics" *Research Methods Knowledge Base 2nd Edition.* Accessed 2/24/09.

(5) Pelham, B. W.; Blanton, H. *Conducting Research in Psychology: Measuring the Weight of Smoke, 3rd Edition.* Wadsworth Publishing (February 27, 2006).

Unite For Sight
International Headquarters
234 Church Street, 15th Floor
New Haven, CT 06510
United States of America

Unite For Sight is a 501(c)(3) nonprofit organization.

Telephone: +1 (203) 404-4900
Email: ufs@uniteforsight.org

First Name

Last Name

Email Address

How did you hear about Unite For Sight?

Submit