

Statistical significance: sample size essay



Statistical significance within a study determines the confidence the researcher can have in the differences found between two or more variables of that study. In other words, this demonstrates how sure a researcher can be that the relationship found between two variables really does exist.

Statistical significance can be tested using various methods, and the significance level at which a variable is detected gives a view of how thoroughly the study has been conducted (Gardenier & Resnik, 2002). This thoroughness often has much to do with the sample size used within an experiment, and in order to produce accurate research, one should avoid certain pitfalls connected with sample sizes (Breugh, 2003). When a small sample size is used in a given experiment, the numbers that the researcher has to work with are so small that even small portions of that sample will represent larger significances than would be the case with a larger population (Lowry, 2007).

A look at two extreme examples will give an idea of how this works. Within a population of two persons, just one of them will represent 50% of the group. However, in a population of 1000 people, 50% is represented by no fewer than 500 people, while one person represents only .01 percent of the population. In a sample containing two persons, statistical analysis cannot be expected to detect differences below 50%. In fact, such an experiment is able only to detect differences (or connections between variables) on a spectrum containing only the proportions represented by 0%, 50% and 100%.

No other proportion is possible (Lowry, 2007; StatPac, 2007). On the other hand, in a sample containing 1000 persons or participants, statistical

analysis has the ability to detect deviations or connections at a much higher significance level because one person's deviation represents such a small portion of the population as .01%. Within a population of 500, that same one person would represent .02 of the population, which is twice .

01%—and any statistical analysis done on this sample would no longer have the ability to detect significance within a .01 significance level. Therefore, though precisely the same protocol might be followed in all the hypothetical experiments cited above, it is the sample size which ends up determining the ability of the analysis to detect changes at various significance levels.

Difference between a statistical and practical significance In a large sample size, it is possible to detect very small deviations within the data or very small differences between groups studied.

Though it is often useful to make such accurate detections, it is also possible that the very small differences detected will not be very significant in a practical way at all. One example has been given in possible the difference in IQ scores between males and females in a hypothetical test (StatPac, 2007). The ability of the test to detect minute differences between the scores of two groups will become higher once the number of persons within the samples increases. However, if for example the test is able to detect a two point difference between the scores received by males and females, this outcome demonstrates the accuracy of the test but gives results that are insignificant. This insignificance comes from the fact that a two-point difference in IQ makes very little difference in the real world, as a person with an IQ of 110 is likely to be as capable as a person with an IQ of 112.

In such a case, it would be more practical to detect a much larger deviation in scores, as that would be much more significant in a practical and useful way (Gardenier & Resnik, 2002; StatPac, 2007). What a statistically significant result denotes is how well the particular variable can be detected by the statistical methods used to analyze it (Gardenier & Resnik, 2002). If a particular relationship exists within a population, yet can only be detected within a small portion of that population, then the result becomes statistically significant at a smaller level or within a smaller confidence interval. Statistical significance speaks about a particular variable and the strength of its relationship to another within a certain population. It is usually left up to the researcher to interpret the value of that relationship and how the knowledge gained from the finding can be of value to the area of research in which the study takes place.

Meta-analysis Meta-analysis was designed as a method of reducing the threats to validity that often arise as a result of small sample sizes (Davies & Crombie, n. d; Gardenier & Resnik, 2002). When sample sizes used for a particular experiment are too small, it becomes possible for errors to enter the data and cause it to become skewed or biased. Meta-analysis involves the survey and investigation of data from a number of related studies. Such analysis is usually advantageous in its ability to produce more accurate data.

One of the problems that arise when conducting a review of studies comes from the methods chosen to analyze data. The usual methods of integrating research that has been previously done often prove unable to cope with the growing amounts of research with which some researchers have to deal.

Meta-analysis helps eliminate this problem. It also delves into the quality of

<https://assignbuster.com/statistical-significance-sample-size-essay/>

the research being evaluated, in order to reduce the problem of citing research without proper examination of the conclusions and the methods used to reach these. It also prescribes methods for researchers to weigh adequately all the evidence whether it is for or against their own preconceived ideas or preferences, thereby reducing the bias of research.

Problems with internal validity arise as a result of such practices as non-randomization, small sample size, discontinuation of the studies by participants (drop-out), the occurrence of significant historical events during a study, lack of control groups, and the problem of extreme results versus the regression effect toward the mean (Losh, 2002). In order to improve the internal validity of research, meta-analysis covers a wide array of studies that serve to combat each of these problems in the following ways. Because meta-analysis deals with a large number of individual studies, problems regarding small sample size can be diminished as the number of participants within the study now becomes the aggregate of all those who participated in the individual studies. As a result, meta-analyses “ have more power to detect small but clinically significant effects” (Davies & Crombie, n. d.).

Biases in the data that arise from non-randomization and problems with lack of control groups can also be diminished because of the practices of meta-analysis experts in choosing carefully which studies to include in their research. When conducting this type of research, it becomes crucial to choose primary research that is “ a complete, unbiased collection of original, high-quality studies that examine the same [...] question” (Davies & Crombie, n. d.).

Researchers who adhere to this practice scrutinize the methodologies of the different studies and remove those that contain major control and randomization flaws. The large number of studies used in meta-analysis also combats the problems or biases that may arise from such phenomena as regression toward the mean. When studies are done (or tests taken) it is often the case that a small percentage of participants score exceptionally high or low. It is often the case, too, that when/if retakes of these studies are done, these same exceptional scorers either increase or decrease their scores, taking them closer to the mean.

With a large body of studies taken in meta-analysis, the effects of these exceptions and regressions can be evened out, so that the study gives a more accurate and statistically valid picture of the problem/issue being examined. As external validity is related to the ability to generalize results across populations, though similar studies must be chosen for meta-analyses, the researcher may be careful to include ones that contain a wide variety of subject types. This will reduce the effects of population sensitization (familiarity with the processes of the test) as well as the likelihood of certain subject types to be (artificially) more inclined to one outcome or another based on the demographic of that particular group. The more inclusive the criteria for the participants, the more widely generalizable will the meta-analytic study become (Davies & Crombie, n.

d.). My Research Training According to McBurney and White (2003), an experiment can be defined as “ a research procedure in which the scientist must select subjects for different conditions from pre-existing groups” (qtd. in Marks, 2006). Some experiments appear to fall into the category of quasi-

experiment, and two of the main contributors to this are lack of randomization and internal validity threats in the event of limited testing variables. If the design for experiments lack detail, several other areas may exist in which internal and external validity might be compromised.

Such details that may be lacking include sample size and protocol, which do have a direct effect on the validity of any experiment. The structure of the hypothesis is also very important (Cortina, 2002). It is necessary that hypotheses be both testable and falsifiable, and this involves the creation of both hypotheses and null hypotheses (Baroudi & Orlikowski, 1989). However, once one gets beyond that point, an experiment may still fail to be a true one if nothing is done to effect the randomization of subjects.

It is important to specify how the initial sample of participants will be selected, making sure to choose a method that will not cause a bias within the population. In some cases, experimenters may choose on purpose to assign participants to groups based on their differences as a method of somehow testing those differences. This would represent a quasi-experiment. What would also make it necessary to classify such a design as a quasi-experiment is the possibility that the experimenter makes no allowance for the possibility (or rather, the probability) that participants differ according to other variables that have not been made a part of the experiment (Shaver, 2005).

Whenever prior information is known about existing differences between the testing and control groups, researchers are expected to put into place statistical controls for these differences (Losh, 2002; Marks, 2006). The lack

of a treatment protocol in the design of any experiment is also a major factor contributing to the classification of such an intervention as a quasi-experiment (Gardenier & Resnik, 2002). The lack of protocol will most likely lead to differences in the experience of each of the groups being studied. It is usually desirable in an experiment that each group follow identical methods of treatment or intervention. However, even with a detailed protocol, threats exist in the form of negligence in overseeing the experiment or in the form of subjects who (intentionally or unintentionally) deviate from the procedure (Shaver, 2005).

Therefore, with no detailed protocol designed for an experiment, the dangers of having dissimilar interventions for each group are compounded even further (Chevalier, 2003; Marks, 2006). Another threat to the experimental nature of this intervention regarding the protocol is the necessity that the procedures be representative of the program that the researcher is interested in. That is, whatever procedures make up the protocol should have some amount of resemblance to the actual activity or phenomenon being studied. Other considerations such as duration and costs of the experiment also have an effect on their external validity (Chevalier, 2003). With no specified protocol, the external validity of the experiment might be affected (Gardenier & Resnik, 2002).

In order for an experiment to qualify as a true one, the size of an experimenter's testing group must be large enough to enable good statistical analysis (Breaugh, 2003). Failure to ensure this poses a problem for more than one reason. Randomization cannot be said to have properly taken place without having a large enough population to choose from (Losh, <https://assignbuster.com/statistical-significance-sample-size-essay/>

2002). Furthermore, in order to have an internally valid experiment, sample size must be adequate in order to increase the precision of estimates and predictions.

Finally, consideration must be given to the possibility that participants might drop out of this experiment. Therefore, the researcher needs to specify a large enough sample size for his testing groups so that his experiment might remain valid if some should drop out. Without this specification, the experiment would just be a quasi-experiment.