Andrew Williams

Professor Elliot

PTC 604

19 Nov. 2006

Assignment 10: Reliability and Statistical Decision-Making, Part 1

What is the NHST?

NHST is null hypothesis significance testing, which is the testing to determine if the difference between the population and sample means is due to chance or if the difference is based on the relationship of the study's variables. The hypotheses that are tested by NHST tests are the null and the alternative hypothesis. The null hypothesis states that there is no difference between the sample mean and a population mean, except for chance. The alternative hypothesis is the opposite of the null and states that there is a difference, that is not due to chance, between the sample mean and the population mean (Bernard 529). Generally, the null hypothesis states the opposite of what the study is attempting to prove and the alternative hypothesis states the direction of the difference of the expected observations (Norusis 247).

In NHST, the probability (p) value measures the level of significance, which customarily is set at .05. In recent years, the NHST test has been evaluated to be somewhat limited because whether or not a p value is detected and the size of the p value if it is detected is in direct correlation to the sample size. NHST may therefore result in tests that lack sufficient power to return significant p value levels. As a result, the NHST should include measures other than just the p value to ensure adequate data is obtained against which null hypothesis and alternative hypotheses may be evaluated (Rosnow and Rosenthal 274). In effect NHST provides a method for ruling out chance and building evidence in favor of either the null or the alternative hypothesis.

<u>What is a confidence interval, and why do we select the 95% confidence interval as a standard of judgment that can be used to refute the null hypothesis?</u>

The confidence interval, also referred to as the margin of error, represents the range of values that contains the expected value, expressed in terms of a percentage, usually 95%. A 95% confidence level indicates that if samples of the same size are drawn repeatedly from a population, and a confidence interval is calculated from each sample, then 95 out of 100 intervals should contain the population mean (Norusis 244). If the mean or a proportion from a sample is likely to occur more than 5% of the time, then the null hypothesis cannot be rejected (Bernard 530).

The customary two-tailed level of significance used in behavioral and social sciences is .05. The confidence interval is $(1-\alpha)100$, where $\alpha = p = .05$. Wider intervals indicate lower precision, narrower intervals indicate greater precision.

In addition to the convenience factor of using a 95% confidence level—"most statistical tables show 5% alpha values"--95% has become the standard as the confidence level "only by custom" (Bernard 530). The level of confidence can be customized relevant to the research study's needs.

<u>What is a Type I error and how is it to be avoided?</u>

A Type I error, known as alpha (α), the significance level and the p value, is rejection of the null hypothesis when it is true and should not be rejected and in effect, claims a relationship between variables that does not exist. Type I errors are considered a mistake of gullibility in the failure to claim a relationship that exists (Rosnow and Rosenthal 279).

If used as the sole criterion, the p value may be low enough to justify rejecting the null hypothesis; however, a p value alone may not give enough information to demonstrate the true strength of the effect of variable relationships. We can reduce the probability of a Type I error by raising the confidence interval; however, the probability of a Type II error is then increased. In addition to the confidence level, we can determine the significance of variable relationships and their effect relative to the size of the sample population.

The t-test compares two means to determine if the means of two independent groups differ on some variable. The formula for the t-test is:

$t = x\ bar_1 - x\ bar_2 / \sqrt{\sigma^2}(1/N_1 + 1/N_2)$

where x bar 1 and x bar 2 are the means of the two sample populations, sigma squared is the variance of the parent population and N_1 and N_2 are the sample populations (Bernard 548) .

In the t-test result, it is important to know the statistical significance of the result, which reveals the practical importance of the relationship between variables. To find the significance of t, we perform a one-tailed test if we are only interested in the size of the difference between the means, or a two-tailed test to if we need to know the direction. We determine the degrees of freedom to be allowed by using the formula $(N_1 - N_2) - 2$, where

N_1 is the first sample size and N_2 is the second sample size. Using the number of degrees of freedom (df), the value of t, and the tail type, we locate the significance value in the t-Distribution table. We then compare our t-value to the standard p value of .05 at the df level to find out if the relationship of the means are statistically significant. A value above .05 means that the value of the relationship is statistically significant and the null hypothesis should not be rejected.

We can also look at the relationship between variables to measure how closely they are correlated to help further define the strength of variable relationships. The Pearson's r correlation coefficient is a formula for calculating the relationship measured by linearity. Pearson's r measures how much of the time changes in one variable correspond with equivalent changes in other variables (Bernard 581). Also called the product-moment correlation because the distances from the means, or z scores, are referred to as moments, and are multiplied to form products. The formula to calculate Pearson's r is:

$$r_{xy} = Z_x - Z_y / N$$

where the linear correlation between variables X and Y equals the sum of the products of the z scores of the variables divided by the number of pairs of X and Y scores (Rosnow and Rosenthal 257). However, simply because variables show a high positive correlation does not mean that causation can be inferred. A third antecedent or intervening variable may affect the behavior of the highly correlated variables. While correlation figures alone will not support causation theories, used with probability significance testing and confidence intervals they can help us to determine the significance of variable relationships and aid in reducing Type I errors.

What is a Type II error, and how is it to be avoided?

A Type II error is failing to reject the null hypothesis when it is false and should be rejected. Type II errors are considered as mistakes of blindness to a relationship that actually does exist and is the opposite of Type I errors. In avoiding Type II errors, the p value, or confidence interval, can be made narrower to include more cases under the distribution, for example, lowered from a .05 level to a .01 level. At the same time, the effect and sample sizes must be taken into consideration. A correlation level may be significant, but its practical importance can be measured only in relation to its accompanying sample size. The significance of r at the lowered p level for the sample size should be ascertained to show the true strength of variable correlation.

WORKS CITED

Bernard, Russell H. Social Research Methods. Thousand Oaks, CA: Sage Publications, Inc. 2000.

Norusis, Marija. SPSS 14.0 Guide to Data Analysis. Upper Saddle River, NJ: Prentice Hall, 2006.

Rosnow, Ralph L. and Robert Rosenthal. Beginning Behavioral Research: A Conceptual Primer. Prentice Hall.