SAGE reference

Encyclopedia of Measurement and Statistics

Inferential Statistics

Contributors: Jonna M. Kulikowich & Maeghan N. Edwards
Edited by: Neil J. Salkind
Book Title: Encyclopedia of Measurement and Statistics
Chapter Title: "Inferential Statistics"
Pub. Date: 2007
Access Date: March 12, 2016
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412916110
Online ISBN: 9781412952644
DOI: http://dx.doi.org/10.4135/9781412952644.n214

Print pages: 458-460

Inferential statistics allow researchers to make generalizations about how well results from samples match those for populations. Because samples are parts of populations, samples do not include all of the population information. Thus, no inference can be perfect because samples cannot represent completely their parent populations.

Imagine that a population is defined as all students enrolled in U.S. schools. Suppose researchers want to study how well students in that population enjoy school. A survey is developed, and researchers prepare to collect response data. Researchers estimate the target population consists of 3 million students. Because the population encompasses the entire United States, the team of researchers must include all geographic areas from Maine to Hawaii. However, actually collecting data for all U.S. students is too massive, so a sample is selected. Ideally, this sample would be *randomly* selected, which means that all U.S. students have the same chance of being selected. This makes the sample *representative* of the population. Essentially, this means that although the sample may represent only a small percentage of the students who comprise the complete population, that sample is assumed to reflect the characteristics of all U.S. students, including those not selected.

Parameters, Statistics, and Probability

A key component of inferential statistics is the degree to which error in estimating population values from sample values can be minimized. Probability theory, a branch of mathematics, plays a central role in inferential statistics. Probability theory serves as the backdrop for two important inferential statistics procedures. One is estimation, and the other is hypothesis testing. *Estimation* focuses on the degree to which sample values indicate true population values. For populations, computed values such as means, standard deviations, and variances are called *parameters*. For samples, those values are called *statistics*. Therefore, questions of estimation address the degree that statistics are equivalent to parameters.

Hypothesis testing pertains to investigators' attempts to answer specific research questions based on theoretical premises. For instance, when researchers want to determine the relationship between two variables—“Is there a relationship between Scholastic Aptitude Test scores and college grade point average?” or “Do males and females differ in their reading ability?”—hypothesis testing converts the research questions into predictive statements so that they can be subjected to empirical testing. Before we describe procedural steps used in hypothesis testing, let's take a very brief look at the history of probability theory in inferential statistics.

Probability Theory and Inferential Statistics

In his classic text titled *Probability, Statistics, and Truth*, Richard von Mises places the onset of probability theory in statistics in the early 1900s. Although properties of distributions, such as those of the normal curve, had been deduced mathematically by the early 1800s, there was limited research on the degree to which the normal curve reflected phenomena observed in the real world. In the early 20th century, the use of probability and the normal curve became important in fields such as agriculture, genetics, and medicine. At this time, R.A. Fisher, a British statistician, introduced the term *likelihood*. This term essentially means probability. Sample data could yield likelihoods of responses that are then compared to what is expected for the population based on properties of mathematical distributions such as the normal curve.

Estimation and Hypothesis Testing Procedures

To illustrate the procedures of estimation and hypothesis testing in inferential statistics, consider the following situation. Suppose high school principals in one district want to answer the following questions about their students' performance on a high-stakes assessment:

1. Did our students perform the same as students nationwide with a mean of 120?
2. Did our students perform differently from students attending a neighboring school district?
3. Did our current students perform better than district students who took the test 4 years ago?

There are four basic inferential statistics steps needed to address any one of the questions. First, researchers must translate each research question into a pair of hypotheses. The first hypothesis, the *null hypothesis*, addresses the question as if the expected answer were "no." The second hypothesis, the *alternative* or *researcher's hypothesis*, addresses the question as if the expected answer were "yes."

In the second step, researchers choose the statistical technique that can help them address each specific question, along with an acceptable error rate with which they justify their conclusion. This error rate is referred to as the *alpha level*. The alpha level (symbolized as α-level) is the degree of *Type I error*. Type I error represents the probability that the null hypothesis will be rejected when the null hypothesis is true for the population. Usually, the alpha level for the social sciences is preset at .05. What this means is that for every 100 times the research question is addressed by a unique sample of data drawn randomly from the same population, there are 5 times when the results are purely attributable to chance.

In the third step of hypothesis testing, researchers actually compute the statistical values given their sample data. Computer programs such as the Statistical Package for the Social Sciences (SPSS) report estimates of the alpha level given the size of the statistical value computed as well as the sample size. Logically, the larger the sample size, the less error in inferring what the population value is based on the sample estimate. This should make sense because larger sample sizes mean that more information about the population is available to the researchers.

In the last step of hypothesis testing, researchers make a decision and state a conclusion. The decision is made in reference to the null hypothesis. Researchers either reject or fail to reject the null hypothesis. If they fail to reject the null hypothesis, it means the answer to the initial research question was "no." If they reject the null hypothesis, then they are actually supporting the alternative hypothesis. This means that the answer to the research question is "yes." The conclusion essentially restates the decision but in less statistical terms. It is usually in their conclusion that researchers also use the word *significant*. This means that if researchers rejected the null hypothesis, then they believe there is a strong likelihood that a result or relationship exists for the population as it does for the sample. We will now take a look at the four steps in answering the principals' research questions.

To answer Question 1, the researcher would set up the following hypotheses:

$$H_0 : \mu_1 = 120$$
$$H_a : \mu_1 \neq 120$$

Although the notation may be unfamiliar to some readers, the statistical symbols are easy to interpret. The symbol *H*0 stands for the null hypothesis, whereas the symbol *Ha* stands for the alternative hypothesis. For both hypotheses, the Greek letter μ (mu) is used because the high school principals are interested in an average or a mean. All hypotheses must be written to reference parameters, not statistics. When researchers state their conclusion, it is a generalization or inference from the sample results to what is expected for the population. Parameters are symbolized with letters from the Greek alphabet. Therefore, all null and alternative hypotheses should be written in notation form with letters like μ (for means), ρ (rho, for correlation coefficients), or β (beta, for regression coefficients).

In Step 2, the researchers state that a sample mean will be computed for the school district. The alpha level or Type I error level is set at .05. In Step 3, the computed sample mean value is compared to a *critical value* using the mathematical properties of the normal curve. The critical value depends on the sample size and the alpha level, and it is determined based on the theoretical properties of the normal curve. These properties have been derived from mathematical calculations using a formula from calculus. There are infinitely many critical values, just as there are infinitely many points that represent the score continuum under the normal curve. Mathematicians typically present critical values in appendix tables. Researchers look at their computed sample value and compare it to the critical value given their specific sample size and set alpha level.

In Step 4 of hypothesis testing, if the mean computed for the sample is larger than the critical value, then the researchers reject the null hypothesis. The conclusion would then be drawn that the school district's test score mean is not equal to the nationwide population with its mean of 120.

For Questions 2 and 3, the pairs of hypotheses would be presented in slightly different notational form. For Question 2, two sample means are compared. The first mean represents one school district, and the second mean represents the neighboring school district. Therefore, these two school districts provide two samples of data. We want to know if the two samples represent the same population or different ones. Therefore, the following pair of hypotheses is recorded:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

The school principals are still interested in means, so μs are used to represent population parameters; however, the subscripts 1 and 2 beside the parameters indicate that there are two samples of means. As was done for the first research question, statisticians would proceed with Steps 2 through 4 of the basic set of hypothesis testing procedures. They would state which statistic should be used and then set the alpha level (i.e., Step 2). This time, a *t*-test value for independent samples would be computed to compare means (i.e., Step 3). The *t*-test value would be compared to a critical value, and finally, a decision and conclusion about the significance of the results would be made (i.e., Step 4).

It is important to note that for Questions 1 and 2, the investigators used an equality sign when stating the hypotheses. This is because the hypotheses are considered *nondirectional*. This means that if we reject the null hypothesis for the first research question, then we are inferring only that the mean is different from 120, not whether it is below or above 120. Likewise, for Question 2, if we reject the null hypothesis, then we know only that the two sample means likely represent two different populations. We are not testing which population assumedly has the greater mean.

For Question 3, however, the principals are interested in the direction of the results. Thus, these *directional* hypotheses are written with inequality signs:

$$H_0 : \mu_1 \leq \mu_2$$
$$H_a : \mu_1 > \mu_2$$

where Group 1 consists of the population from which scores from the school are drawn this year, and Group 2 consists of the population from which scores from this school were drawn 4 years ago. If the null hypothesis is rejected, then the researchers will infer that the current average for their school district is significantly greater than the average reported 4 years ago.

Controversies Surrounding Inferential Statistics

There are contemporary debates surrounding the use of inferential statistics. One debate pertains to statistical power. We mentioned that there is always error in inferential statistics, and Type I error is one example. Statisticians must simultaneously deal with *Type II error*, given every null hypothesis tested. Type II error occurs when researchers fail to reject the null hypothesis, and the null hypothesis was wrong. Statistical power is based on Type II error given the simple formula $1 - \beta$, where β stands for the probability of Type II error. Thus, *statistical power* reflects the degree to which a decision is made without error.

Currently, many investigators think that stating the decision and conclusion is insufficient in inferential statistics because statistical power is not addressed directly. What these researchers know is that as sample size increases, there is a greater likelihood of rejecting the null hypothesis. However, the statistical value may be very small and meaningless given researchers' interests. With a large school district, the test mean might be only 122, and the null hypothesis presented for Question 1 could be rejected. Practically speaking, 122 and 120 are not that different when comparing standardized test scores in schools.

Therefore, many researchers also compute effect sizes. *Effect sizes* tell the direction and magnitude of differences between means in standard deviation units. As stated, as sample size increases, it is easier to reject a null hypothesis. Likewise, in some fields, researchers may not be able to sample large numbers of participants. Consequently, they may not be able to reject the null hypothesis, yet they may be able to compute large effect sizes. For that reason, many journal editors recommend reporting the level of statistical significance *and* the effect size.

- null hypothesis

- critical value
- inferential statistics
- normal curve
- sampling
- hypothesis testing
- sample size

Jonna M. Kulikowich and Maeghan N. Edwards
http://dx.doi.org/10.4135/9781412952644.n214
See also

- Hypothesis and Hypothesis Testing
- Significance Level

Further Reading
Heiman, G. W. (2006). Basic statistics for the behavioral sciences. Boston: Houghton Mifflin.
von Mises, R. (1957). Probability, statistics, and truth. New York: Dover.

- http://www.socialresearchmethods.net/kb/statinf.htm Inferential statistics