Bio-instrumentation & Biostatistics

Dr. Bibha Verma
PG Department of Zoology
Maharaja College Ara

PG Sem II, CC-6, Unit -III

3.4 Hypothesis Testing

Introduction

A **statistical hypothesis** is an assumption about a population parameter. This assumption may or may not be true. **Hypothesis testing** refers to the formal procedures used by statisticians to accept or reject statistical hypotheses.

Statistical Hypotheses

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

There are two types of statistical hypotheses.

- **Null hypothesis**. The null hypothesis, denoted by H_o, is usually the hypothesis that sample observations result purely from chance.
- **Alternative hypothesis**. The alternative hypothesis, denoted by H_1 or H_a, is the hypothesis that sample observations are influenced by some non-random cause.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

H_o: P = 0.5

H_a: P ≠ 0.5

Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably not fair and balanced.

Hypothesis Tests

Statisticians follow a formal process to determine whether to reject a null hypothesis, based on sample data. This process, called hypothesis testing, consists of four steps.

State the hypotheses. This involves stating the null and alternative hypotheses. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false.

Formulate an analysis plan. The analysis plan describes how to use sample data to evaluate the null hypothesis. The evaluation often focuses around a single test statistic.

Analyze sample data. Find the value of the test statistic (mean score, proportion, t statistic, z-score, etc.) described in the analysis plan.

Interpret results. Apply the decision rule described in the analysis plan. If the value of the test statistic is unlikely, based on the null hypothesis, reject the null hypothesis.

Errors in Statistical Tests

The interpretation of a statistical hypothesis test is probabilistic.

That means that the evidence of the test may suggest an outcome and be mistaken.

For example, if alpha was 5%, it suggests that (at most) 1 time in 20 that the null hypothesis would be mistakenly rejected or failed to be rejected because of the statistical noise in the data sample.

Given a small p-value (reject the null hypothesis) either means that the null hypothesis false (we got it right) or it is true and some rare and unlikely event has been observed (we made a mistake). If this type of error is made, it is called a false positive. We falsely believe the rejection of the null hypothesis.

Alternately, given a large p-value (fail to reject the null hypothesis), it may mean that the null hypothesis is true (we got it right) or that the null hypothesis is false and some unlikely event occurred (we made a mistake). If this type of error is made, it is called a false negative. We falsely believe the null hypothesis or assumption of the statistical test.

Each of these two types of error has a specific name.

Type I Error: The incorrect rejection of a true null hypothesis or a false positive.

Type II Error: The incorrect failure of rejection of a false null hypothesis or a false negative.

All statistical hypothesis tests have a chance of making either of these types of errors. False findings or false discoveries are more than possible; they are probable.

Ideally, we want to choose a significance level that minimizes the likelihood of one of these errors. E.g. a very small significance level. Although significance levels such as 0.05 and 0.01 are common in many fields of science, harder sciences, such as physics, are more aggressive.

It is common to use a significance level of 3 * 10^-7 or 0.0000003, often referred to as 5-sigma. This means that the finding was due to chance with a probability of 1 in 3.5 million independent repeats of the experiments. To use a threshold like this may require a much large data sample.

Nevertheless, these types of errors are always present and must be kept in mind when presenting and interpreting the results of statistical tests. It is also a reason why it is important to have findings independently verified.

Decision Rules

The analysis plan includes decision rules for rejecting the null hypothesis. In practice, statisticians describe these decision rules in two ways - with reference to a P-value or with reference to a region of acceptance.

P-value. The strength of evidence in support of a null hypothesis is measured by the P-value. Suppose the test statistic is equal to S. The P-value is the probability of observing a test statistic as extreme as S, assuming the null hypothesis is true. If the P-value is less than the significance level, we reject the null hypothesis.

Region of acceptance. The region of acceptance is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is not rejected. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

The set of values outside the region of acceptance is called the region of rejection. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance.

These approaches are equivalent. Some statistics texts use the P-value approach; others use the region of acceptance approach. On this website, we tend to use the region of acceptance approach.

Use and importance

Statistics are helpful in analyzing most collections of data. This is equally true of hypothesis testing which can justify conclusions even when no scientific theory exists.

Real world applications of hypothesis testing include:

Testing whether more men than women suffer from nightmares

Establishing authorship of documents

Evaluating the effect of the full moon on behavior

Determining the range at which a bat can detect an insect by echo

Deciding whether hospital carpeting results in more infections

Selecting the best means to stop smoking

Checking whether bumper stickers reflect car owner behavior

Testing the claims of handwriting analysts

Statistical hypothesis testing plays an important role in the whole of statistics and in statistical inference.

Significance testing has been the favored statistical tool in some experimental social sciences (over 90% of articles in the Journal of Applied Psychology during the early 1990s).

References :

https://stattrek.com/hypothesis-test/hypothesis-testing.aspx

https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

https://www.investopedia.com/terms/h/hypothesistesting.asp

https://machinelearningmastery.com/statistical-hypothesis-tests/