Lesson 1: Introduction to Hypothesis Testing

Hypothesis testing is the process of evaluating the accuracy of a statement regarding a population parameter given sample information.

A **hypothesis** is a statement about the value of a population parameter developed for the purpose of testing a theory.

Example

We think (hypothesize) that the average points scored in each game by a basketball player throughout his career is greater than 30. To test the validity of our statement:

- We would need to get some sample information.
- Conduct a hypothesis test on the sample information.
- Only then will we be able to comment on the accuracy of the statement pertaining to the population parameter.

THE NULL HYPOTHESIS

The **null hypothesis (H_0)** generally represents the status quo, and is the hypothesis that we are *interested in rejecting.*

This hypothesis will not be rejected unless the sample data provides sufficient evidence to reject it.

- Null hypotheses regarding the mean of the population can be stated in the following ways:

In the basketball player's example that we just described, the null hypotheses would be that the player's average score is less than, or equal to 30 points. Confirmation of our belief (that his average score is greater than 30) requires rejection of the null hypothesis.

THE ALTERNATE HYPOTHESIS

The **alternate hypothesis (H_a)** is essentially the statement whose *validity we are trying to evaluate.*

- The alternate hypothesis is the statement that will only be accepted if the sample data provides convincing evidence of its truth.
- It is the conclusion of the test if the null hypothesis is rejected.

Alternate hypotheses can be stated as:

In our example, the alternate hypothesis is that the player's scoring average is greater than 30 points. Recall that we are trying to evaluate the validity of the statement that his scoring average is greater than 30 points.

THE SIGNIFICANCE LEVEL

A hypothesis test is always conducted at a particular **level of significance (α).**
The level of significance represents the chance that we are willing to take that the conclusion from the test might be wrong.

A HYPOTHESIS TEST

A hypothesis test involves the comparison of a sample's test statistic to a critical value.

- **Test Statistic**

- **Critical Value**

ONE-TAILED HYPOTHESIS TESTS

Under **one-tailed tests**, we assess whether the value of a population parameter is either greater than or less than a given hypothesized value. Hypotheses for one-tailed tests can be stated as:

1. H0: $\mu \le \mu 0$ versus Ha: $\mu > \mu 0$

2. H_0: $\mu \ge \mu_0$ versus Ha: $\mu < \mu_0$

Example

Suppose that the basketball player's average score in a sample of 49 games is 36 points with a standard deviation of 9 points. Determine the accuracy of the statement that his career scoring average is greater than 30 points. Use the 5% level of significance.

REJECTION RULES

The following rejection rules apply when trying to determine whether a population mean is *greater* than the hypothesized value.

- Reject H0 when:
Test statistic > positive critical value

- Fail to reject H0 when:
Test statistic ≤ positive critical value

Example

For the same data that we used in the previous example, assume that the sampling distribution has a mean of 28.5. Evaluate the validity of the statement that the player's scoring average over his career is less than 30 points at the 5% level of significance.

REJECTION RULES

The following rejection rules apply when trying to determine whether a population mean is less than the hypothesized value.

- Reject H_0 when:
  Test statistic < negative critical value

- Fail to reject H0 when:
  Test statistic ≥ negative critical value

TWO-TAILED HYPOTHESIS TESTS

Under **two-tailed tests**, we assess whether the value of the population parameter is simply *different from* a given hypothesized value.

The hypotheses for two-tailed tests are stated as:

Two-tailed hypotheses tests have 2 rejection regions.

Example

A two-tailed test will be used to determine if the player's scoring average is simply different from, or not equal to 30. His scores can differ from 30 in two ways – by being less than 30 or by being more than 30; hence the two-tailed test. Given that over a sample of 49 games, the player averaged 33 points with a standard deviation of 9 points, test whether his career scoring average is different from 30 at the 5% significance level.

REJECTION RULES FOR TWO-TAILED HYPOTHESIS TESTS

The following rejection rules apply when trying to determine whether a population mean is *different from* the hypothesized value.

- Reject H_0 when:
  Test statistic < lower critical value
  Test statistic > upper critical value

- Fail to reject H_0 when:
  Lower critical value ≤ test statistic ≤ upper critical value

TESTING ERRORS

There are two types of errors that can be made when conducting a hypothesis test:

Type I error: Rejecting the null hypothesis when it is actually true.

- The significance level (α) represents the probability of making a Type I error.
- A significance level of 5% means that there is a 5% chance of rejecting the null when it is actually true.
- A significance level is required in any hypothesis test to compute critical values against which the test statistic is compared.

TESTING ERRORS

Type II error: Failing to reject the null hypothesis when it is actually false.

If we were to fail to reject the null hypothesis given the lack of overwhelming evidence in favor of the alternate, we risk a Type II error *failing to reject the null hypothesis when it is false.*

- Sample size and the choice of significance level (probability of Type I error) together determine the probability of a Type II error.
- Measuring the risk of Type II errors is very difficult.

The power of a test is the probability of *correctly* rejecting the null hypothesis when it is false.

- Power of a test = 1 - P(Type II error)

TESTING ERRORS

- The higher the power of the test, the better it is for purposes of hypothesis testing.
- Decreasing the significance level reduces the probability of Type I error. However, reducing the significance level means shrinking the rejection region, and inflating the "fail to reject the null region". This increases the probability of failing to reject a false null hypothesis (Type II error) and reduces the power of the test.
- The power of the test can only be increased by reducing the probability of a Type II error.
  - This can only be accomplished by reducing the "fail to reject the null region", which is equivalent to increasing the size of the rejection region and increasing the probability of a Type I error.
  - Basically, an increase in the power of a test comes at the cost of increasing the probability of a Type I error.
- The only way to decrease the probability of a Type II error given the significance level (probability of Type I error) is to increase the sample size.

Relation Between Confidence Intervals and Hypothesis Tests

A **confidence interval** states the range of values within which a population parameter is estimated to lie.

If the confidence interval is calculated at a 95% level of confidence, there is a 95% chance that the relevant population parameter lies within the interval.

Confidence intervals and hypothesis test are linked by critical values.
In a confidence interval, we state that the population parameter lies within the interval, which represents the “fail-to-reject- the-null region” with a particular degree of confidence ($1-\alpha$).

In a hypothesis test, we examine whether the population parameter lies in the rejection region, or outside the interval at a particular level of significance (α).

Example

Construct a 95% confidence interval for the basketball player's career scoring average if, over a sample of 49 games, he averaged 31 points with a standard deviation of 9 points. Use this confidence interval to determine whether the player's career average is different from 33 points.

The **p-value** is the smallest level of significance at which the null hypothesis can be rejected. It represents the probability of obtaining a critical value that would lead to rejection of the null hypothesis.

Rejection rules using p-values:

- If the p-value is lower than the required level of significance, we reject the null hypothesis.
- If the p-value is greater than the required level of significance, we fail to reject the null hypothesis.