HYPOTHESIS

(Dr. N. N. Mahto, Visthapit Mahvidyalaya, Balidih)

DEFINITION OF HYPOTHESIS

A hypothesis is a tentative assertion or a formal statement of theory (testable or refutable) that shows how two or more variables are expected to relate to one another [1]. It could also be a formal version of a speculation that is usually based on a theory. Therefore, a statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true. This is why hypotheses are more specific than theories. Multiple hypotheses may relate to one theory. However, hypotheses result from the reasoning done in the conceptual framework.

Hypotheses can take the form of a simple proposition of an expected outcome, or can assert the existence of a relationship. For instance, a simple proposition might be that one production system, based on a particular technology is more profitable than another production system based on another technology

On the other hand, a hypothesis of a relationship could be that in the demand for pork in Nigeria, the per capita consumption of pork is affected by price of pork, the price of other substitutes (meat or fish), per capita income, religious affiliation, and ethnic background.

TYPES OF STATISTICAL HYPOTHESES

There are two types of statistical hypotheses;

1. **Null hypothesis**: The null hypothesis, denoted by H_0, is usually the hypothesis that sample observations result purely from chance. The null hypothesis, always states that the treatment has no effect (no change, no difference). According to the null hypothesis, the population mean after treatment is the same as it was before treatment. The α-level establishes a criterion, or "cut-off", for making a decision about the null hypothesis. The alpha level also determines the risk of a Type I error.

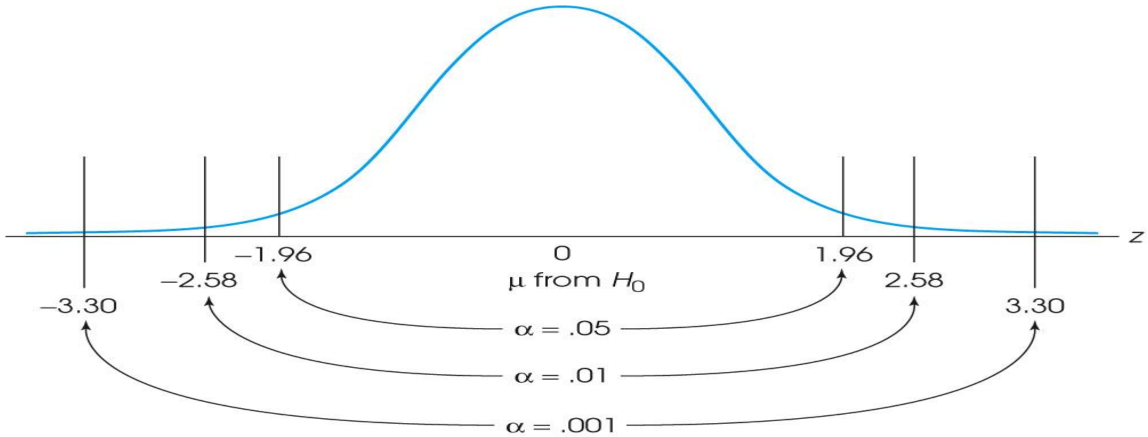

Figure 1

The locations of the critical region boundaries for three different levels of significance: α = .05, α = .01, and α = .001 are shown in Figure 1. The critical region consists of outcomes that are very unlikely to occur if the null hypothesis is true. That is, the critical region is defined by sample means that are almost impossible to obtain if the treatment has no effect. This means that these samples have a probability (p) that is less than the alpha level.

2. **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.

HYPOTHESIS TESTING

Hypothesis testing is a technique which helps to determine whether a specific treatment has an effect on the individuals in a population[]. It is a formal procedures used by statisticians to accept or reject 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.

For instance, 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

$$H_0: P = 0.5$$
$$H_a: P \neq 0.5$$

Suppose we flipped the coin 30 times, resulting in 20 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.

In another example,

- Statements about characteristics of populations, denoted H:
  - H: normal distribution, $\mu = 28; \sigma = 13$
  - H: $N(28,13)$
- The hypothesis actually tested is called the *null hypothesis*, H_0
  - E.g., $H_0 : \mu = 100$
- The other hypothesis, assumed true if the null is false, is the *alternative hypothesis*, H_1
  - E.g., $H_1 : \mu \neq 100$

The general goal of a hypothesis test is to rule out chance (sampling error) as a plausible explanation for the results from a research study.

The basic experimental situation for hypothesis testing

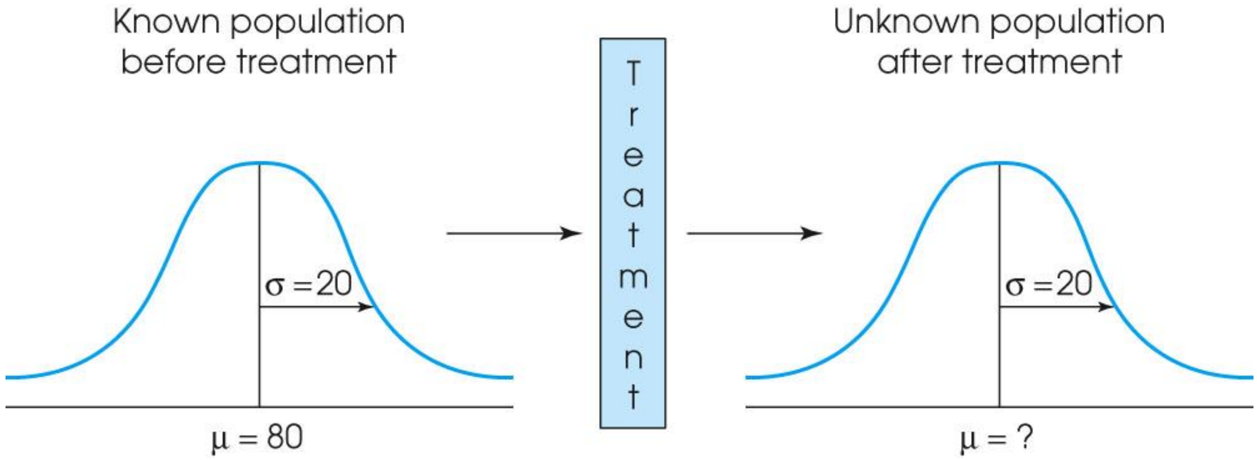

Figure 1

From Fig. 1, it is assumed that the parameterμ is known for the population before treatment. The purpose of the experiment is to determine whether or not the treatment has an effect on the population mean.If the individuals in the sample are noticeably different from the individuals in the original population, we have evidence that the treatment has an effect.

However, it is also possible that the difference between the sample and the population is simply sampling error

Also, from Fig. 2, the entire population receives the treatment and then a sample is selected from the treated population. In the actual research study, a sample is selected from the original population and the treatment is administered to the sample. From either perspective, the result is a treated sample that represents the treated population.

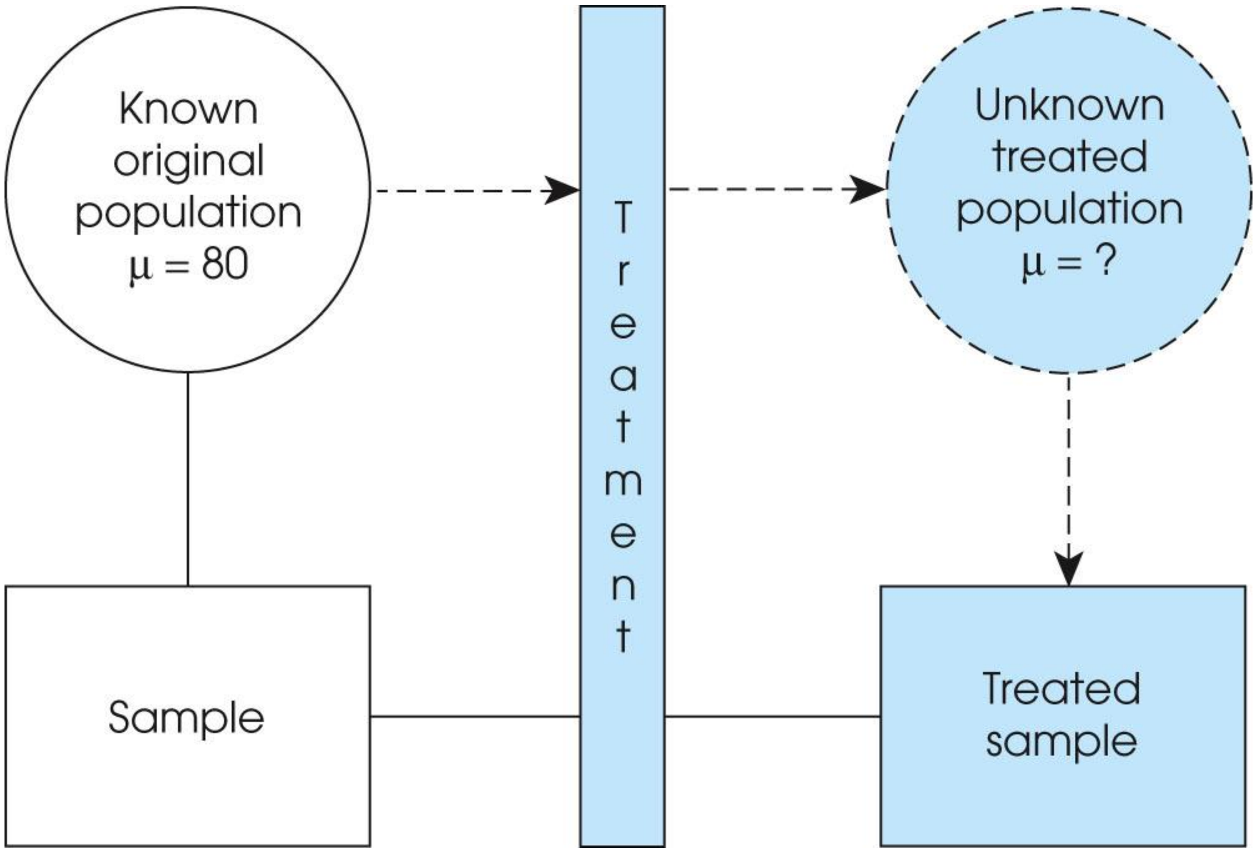

Figure 2

The purpose of the hypothesis test is to decide between two explanations:

i. The difference between the sample and the population can be explained by sampling error (there does not appear to be a treatment effect)

ii. The difference between the sample and the population is too large to be explained by sampling error (there does appear to be a treatment effect).

STEPS IN HYPOTHESIS TESTING

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 [2].

a. **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.

b. **<u>Formulate an analysis plan</u>**: The analysis plan describes how to use sample data to evaluate the null hypothesis. The evaluation often focuses around a single test statistic.

c. **<u>Analyze sample data</u>**: Find the value of the test statistic (mean score, proportion, t statistic, z-score, etc.) described in the analysis plan.

d. **<u>Interpret results</u>**: 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.

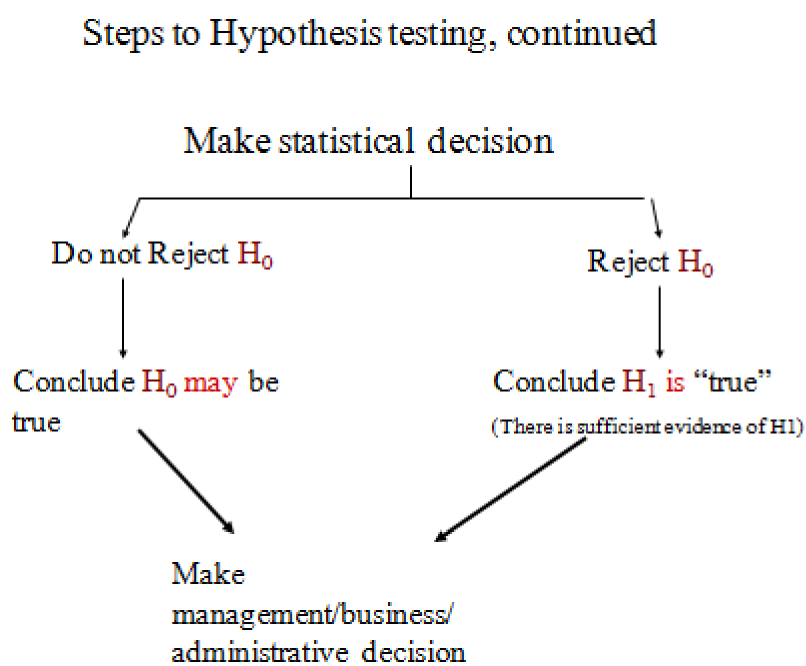

Figure 3

<u>DECISION ERRORS</u>

Two types of errors can result from a hypothesis test.

- **Type I error**: A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by α.

- **Type II error**: A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by β. The probability of *not* committing a Type II error is called the **Power** of the test.

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 (for instance, Fig. 4), the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance.

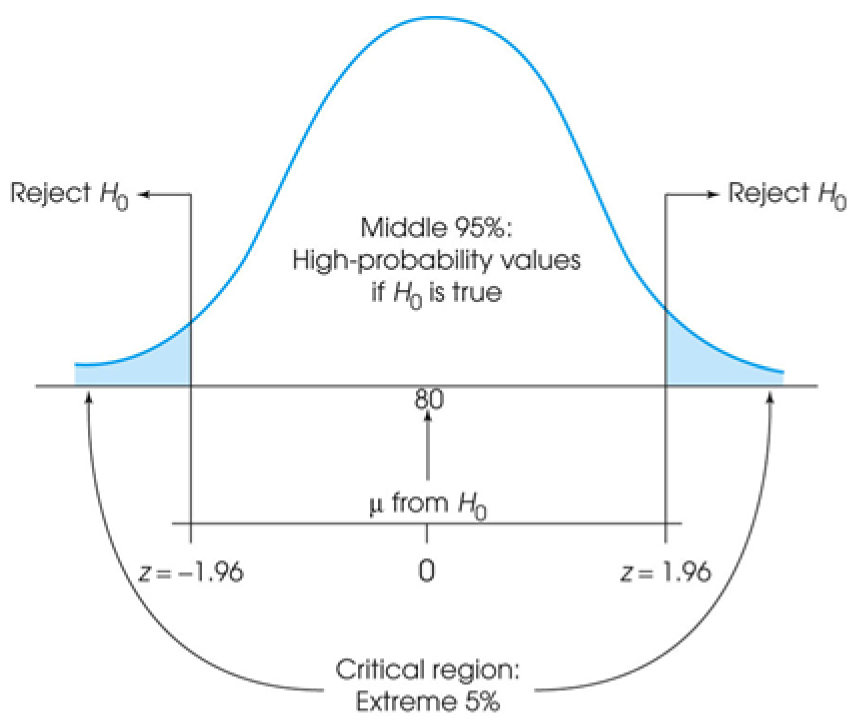

Figure 4

APPLICATIONS OF SELECTED HYPOTHESIS TESTING PROCEDURE

Hypothesis Test for Regression Slope

This describes how to conduct a hypothesis test to determine whether there is a significant linear relationship between an independent variable *X* and a dependent variable *Y*. The test focuses on the slope of the regression line;

$$Y = B_0 + B_1X$$

Where B_0 is a constant, B_1 is the slope (also called the regression coefficient), X is the value of the independent variable, and Y is the value of the dependent variable.

If we find that the slope of the regression line is significantly different from zero, we will conclude that there is a significant relationship between the independent and dependent variables.

This test is appropriate when the following conditions are met;

- The dependent variable *Y* has a linear relationship to the independent variable *X*.
- For each value of X, the probability distribution of Y has the same standard deviation σ.
- In a simpler form, the data for both the X and Y value must be quantitative.
- For any given value of X,
    - The Y values are independent.
    - The Y values are roughly normally distributed (i.e., symmetric and unimodal). A littleskewness is okay if the sample size is large.

The test procedure consists of four steps:

1. **State the Hypotheses**

If there is a significant linear relationship between the independent variable *X* and the dependent variable *Y*, the slope will *not* be equal to zero.

$$H_0: B_1 = 0$$
$$H_a: B_1 \neq 0$$

The null hypothesis states that the slope is equal to zero, and the alternative hypothesis states that the slope is not equal to zero.

2. **Formulate an Analysis Plan**

The analysis plan describes how to use sample data to accept or reject the null hypothesis. The plan should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
- Test method. Use a linear regression test to determine **whether the slope of the regression line differs significantly from zero.**

State the Hypotheses

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

Table 3 shows three sets of null and alternative hypotheses. Each makes a statement about the difference *d* between the mean of one population μ_1 and the mean of another population μ_2. (In the table, the symbol ≠ means "not equal to ".)

Table 3

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu_1 - \mu_2 = d$	$\mu_1 - \mu_2 \neq d$	2
2	$\mu_1 - \mu_2 \geq d$	$\mu_1 - \mu_2 < d$	1
3	$\mu_1 - \mu_2 \leq d$	$\mu_1 - \mu_2 > d$	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses (Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis.

When the null hypothesis states that there is no difference between the two population means (i.e., $d = 0$), the null and alternative hypothesis are often stated in the following form.

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

Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.

- Test method. Use the two-sample t-test to determine whether the difference between means found in the sample is significantly different from the hypothesized difference between means.

Analyze Sample Data

Using sample data, the following can be computed:

- **Standard error** (SE) of the sampling distribution.

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)]$$

  where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, and n_2 is the size of sample 2.

- The degrees of freedom (DF) is:

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$

  If DF does not compute to an integer, round it off to the nearest whole number. Some texts suggest that the degrees of freedom can be approximated by the smaller of $n_1 - 1$ and $n_2 - 1$; but the above formula gives better results.

- The test statistic is a t statistic (t) defined by the following equation.

$$t = [(x_1 - x_2) - d] / SE$$

where x_1 is the mean of sample 1, x_2 is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

- The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a t statistic, use the t Distribution Calculator to assess the probability associated with the t statistic, having the degrees of freedom computed above.

Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

Now, let us use hypothetical examples to illustrate how to conduct a hypothesis test of a difference between mean scores. The first problem involves a two-tailed test; the second problem, a one-tailed test.

Problem 1: Two-Tailed Test

Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Similoluwa and Mrs. Juliet. After the assignment, Mrs. Similoluwa had 30 students, and Mrs. Juliet had 25 students.

At the end of the year, each class took the same standardized test. Mrs. Similoluwa's students had an average test score of 78, with a standard deviation of 10; and Mrs. Juliet’s students had an average test score of 85, with a standard deviation of 15.

Test the hypothesis that Mrs. Similoluwa and Mrs. Juliet are equally effective teachers. Use a 0.10 level of significance. (Assume that student performance is approximately normal.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

$$\text{Null hypothesis: } \mu_1 - \mu_2 = 0$$
$$\text{Alternative hypothesis: } \mu_1 - \mu_2 \neq 0$$

  Note that these hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the difference between sample means is too big or if it is too small.

- **Formulate an analysis plan**. For this analysis, the significance level is 0.10. Using sample data, we will conduct a two-sample t-test of the null hypothesis.

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)]$$
$$SE = sqrt[(10^2/30) + (15^2/25] = sqrt(3.33 + 9) = sqrt(12.33) = 3.51$$
$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$
$$DF = (10^2/30 + 15^2/25)^2 / \{ [(10^2 / 30)^2 / (29)] + [(15^2 / 25)^2 / (24)] \}$$
$$DF = (3.33 + 9)^2 / \{ [(3.33)^2 / (29)] + [(9)^2 / (24)] \} = 152.03 / (0.382 + 3.375) =$$
$$152.03/3.757 = 40.47$$
$$t = [(x_1 - x_2) - d] / SE = [(78 - 85) - 0] / 3.51 = -7/3.51 = -1.99$$

  where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, x_1 is the mean of sample 1, x_2 is the mean of sample 2, d is the hypothesized difference between the population means, and SE is the standard error.

  Since we have a two-tailed test, the P-value is the probability that a t statistic having 40 degrees of freedom is more extreme than -1.99; that is, less than -1.99 or greater than 1.99.

  We use the t Distribution Calculator to find P (t < -1.99) = 0.027, and P (t > 1.99) = 0.027. Thus, the P-value = 0.027 + 0.027 = 0.054.

- **Interpret results**. Since the P-value (0.054) is less than the significance level (0.10), we cannot accept the null hypothesis.

Problem 2: One-Tailed Test

The Acme Company has developed a new battery. The engineer in charge claims that the new battery will operate continuously for *at least* 7 minutes longer than the old battery.

To test the claim, the company selects a simple random sample of 100 new batteries and 100 old batteries. The old batteries run continuously for 190 minutes with a standard deviation of 20 minutes; the new batteries, 200 minutes with a standard deviation of 40 minutes.

Test the engineer's claim that the new batteries run at least 7 minutes longer than the old. Use a 0.05 level of significance. (Assume that there are no outliers in either sample.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

  Null hypothesis: $\mu_1 - \mu_2 >= 7$
  Alternative hypothesis: $\mu_1 - \mu_2 < 7$

  Note that these hypotheses constitute a one-tailed test. The null hypothesis will be rejected if the mean difference between sample means is too small.

- **Formulate an analysis plan**. For this analysis, the significance level is 0.05. Using sample data, we will conduct a two-sample t-test of the null hypothesis.

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)]$$
$$SE = sqrt[(40^2/100) + (20^2/100] = sqrt(16 + 4) = 4.472$$
$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$
$$DF = (40^2/100 + 20^2/100)^2 / \{ [(40^2 / 100)^2 / (99)] + [(20^2 / 100)^2 / (99)] \}$$
$$DF = (20)^2 / \{ [(16)^2 / (99)] + [(2)^2 / (99)] \} = 400 / (2.586 + 0.162) = 145.56$$
$$t = [(x_1 - x_2) - d] / SE = [(200 - 190) - 7] / 4.472 = 3/4.472 = 0.67$$

where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, x_1 is the mean of sample 1, x_2 is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

Here is the logic of the analysis: Given the alternative hypothesis ($\mu_1 - \mu_2 < 7$), we want to know whether the observed difference in sample means is small enough (i.e., sufficiently less than 7) to cause us to reject the null hypothesis.

The observed difference in sample means (10) produced a t statistic of 0.67. We use the t Distribution Calculator to find $P(t \leq 0.67) = 0.75$.

This means we would expect to find an observed difference in sample means of 10 or less in 75% of our samples, if the true difference were actually 7. Therefore, the P-value in this analysis is 0.75.

- **Interpret results**. Since the P-value (0.75) is greater than the significance level (0.05), we cannot reject the null hypothesis.