Chapter 4

TESTS OF HYPOTHESES

4.1 Basic Definitions

The testing of statistical hypotheses is perhaps the most important area of decision theory. First, let us define precisely what we mean by a statistical hypothesis.

***A statistical hypothesis is assumption or statement, which may or may not be true, concerning one or more populations*.**

The truth of falsity of a statistical hypothesis is never known with absolute certainty unless we examine the entire population. This, of course, would be impractical in most situations.

Instead, we take a random sample from the population of interest and use the information contained in this sample to decide whether the hypothesis is likely to be true or false. It is important to understand that the rejection of a hypothesis is to conclude that it is false. While the acceptance of a hypothesis merely implies that we have no evidence to believe otherwise. Because of this terminology, the statistician or experimenter should always state at his hypothesis that which he hopes to reject. If he is interested in a new cold vaccine, he should assume that it is no better than the vaccine now on the market and then set out to reject this contention. Similarly, to prove that one teaching technique is superior to another, we test the hypothesis that there is no difference in the two techniques.

Hypothesis that we formulate with the hope of rejecting are called ***null hypotheses and are denoted by H_o***. The ***rejection*** of H_o leads to the ***acceptance*** of an **alternative hypothesis**, denoted by $\mathbf{H_a}$. A null hypothesis will always be stated so as to specify an exact value of the population parameter, whereas the alternative hypothesis allow for the possibility of one or several values.

Hence, if H_o is the null hypothesis: $p = 0.5$ for a binomial experiment, the alternative hypothesis H_o might be $p = 0.7$, $p < 0.5$, $p > 0.5$, or $p \neq 0.5$

4.2 Type I and Type II Errors

The decision procedure just described could lead to either of two wrong conclusions.

A type I error is committed if we reject the null hypothesis when it is true.

A type II error is committed if we accept the null hypothesis when it is false. The ***probability of committing a type I error*** is called the **level of significance** of the test and is denoted by **α**.

	H_0 **is true**	H_0 **is false**
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

In order for any tests of hypotheses or rules of decision to be good they must be designed so as to minimize errors of decision. This is not a simple matter since, for a given sample size, an attempt to decrease one type of error is accompanied in general by an increase in the other type of error. In practice one type of error may be more serious than the other, and so a compromise should be reached in favor of a limitation of the more serious error. The only way to reduce both types of error is to increase the sample size, which may or may not be possible.

4.3 One Side and Two-sided Tests

A test of any statistical hypothesis where the alternative hypothesis is expressed by mean of a less than symbol (**<**) or greater than symbol (**>**), is called a one-tailed test, since the entire critical region lies in one tail of the distribution of the test statistic. In a sense, the symbol points in the direction where the critical region lies, e.g.

$$H_o : \mu = \mu_o \qquad \textbf{against} \qquad H_a : \mu > \mu_o$$

or

$$H_o : \mu = \mu_o \qquad \textbf{against} \qquad H_a : \mu > \mu_o$$

A test of any statistical hypothesis where the alternative is written with a not-equals sign (≠) is called a two tailed (sided), test, since the critical region is split into two equal parts, one in each tail of the distribution of the test statistic.

$$H_o : \mu = \mu_o \qquad \textbf{against} \qquad H_a : \mu \neq \mu_o$$

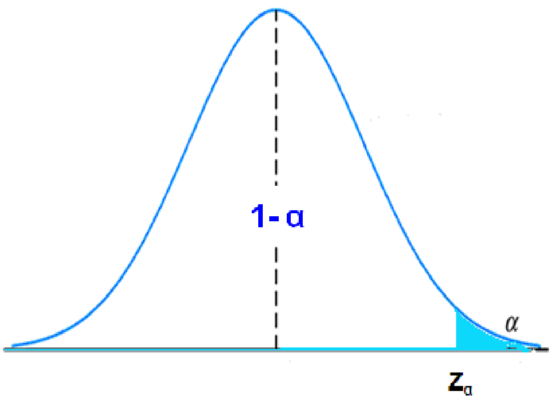


Fig: **4.1(a)** *Critical region for the alternative hypothesis* $H_a : \mu > \mu_o$.

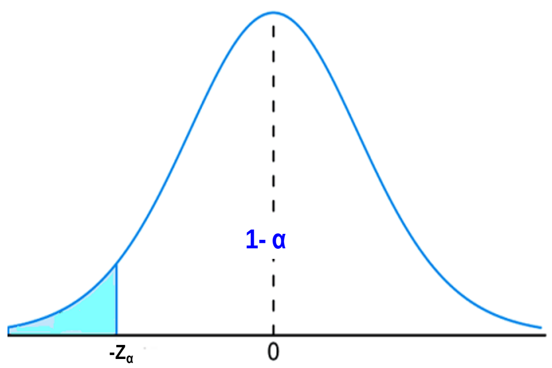


Fig: **4.1(b)** *Critical region for the alternative hypothesis* $\mathbf{H_a : \mu < \mu_o}$.

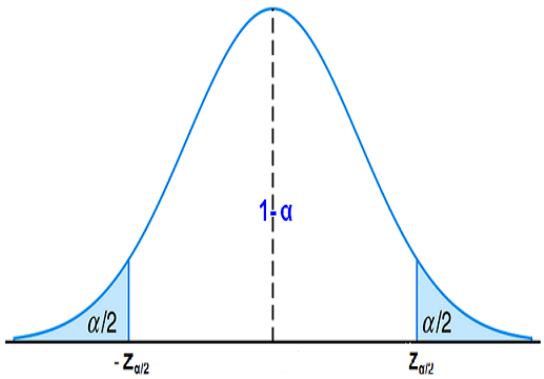


Fig: **4.1(c)** *Critical region for the alternative hypothesis* $\mathbf{H_a : \mu \neq \mu_o}$.

For example, in testing a new drug, one set up the hypothesis that it is no bettthan similar drugs now on the market and tests this against the alternative hypothesis that the new drug is superior. Such an alternative hypothesis will result in a one-tailed test with the critical region in the right tail. However, if we wish to compare a new teaching technique with the conventional classroom procedure, the alternative hypothesis should allow for the new approach to be either inferior or superior to the conventional procedure. Hence the test is two-tailed with the critical region divided equally so as to fall in the extreme left and right tails of the distribution of our statistic.

4.4 Tests Concerning The Population Mean μ

Case 1: σ is known (or n is large)

Consider the problem of testing the hypothesis that the mean μ of a population, with known variance σ^2, equals a specified value μ_o against the alternative that the mean is not equal to μ_o; that is, we shall test

$$\mathbf{H_o : \mu = \mu_o}, \qquad \textbf{agains} \qquad \mathbf{H_a : \mu \neq \mu_o}.$$

The appropriate statistic on which we base our decision is the random variable $\mathbf{\overline{X}}$.

From the ***Central Limit Theorem*** we already know that the sampling distribution of $\mathbf{\overline{X}}$ is approximately normally distributed with mean $\mu_{\bar{x}} = \mu$ and variance $\sigma^2_{\bar{x}} = \sigma^2/n$, where μ and σ^2 are the mean and variance of the population from which we select a random sample of size n.

It is convenient to standardize $\bar{X}$ and formally involve the ***standard normal*** random variable *Z*, where

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

We know that ***under*** **H_0**, that is, if **$\mu = \mu_0$**,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

and hence the expression

$$P(-z_{\alpha/2} < \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

can be used to write an appropriate non rejection region. We should keep in mind that, formally, the critical region is designed to control **α**, the ***probability of type I error***. Thus, given a computed value $\bar{X}$, the formal test involves rejecting **H_0** if the computed ***test statistic z falls in the critical region*** described as:

$$z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < -z_{\alpha/2} \quad \text{or} \quad z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

If **$-z_{\alpha/2} < z < z_{\alpha/2}$**, **do not reject H_0**. **Rejection of H_0**, of course, implies **acceptance of the alternative hypothesis** $H_a: \mu \neq \mu_o$. With this definition of the critical region, it should be clear that there will be probability α of rejecting **H_0** (falling into the critical region) when, indeed, $\mu = \mu_o$.

Tests of one-sided hypotheses on the mean involve the same statistic described in the two-sided case. The difference, of course, is that the critical region is only in one tail of the standard normal distribution. For example, suppose that we seek to test

$$H_o: \mu = \mu_o \qquad \textbf{against} \qquad H_a: \mu > \mu_o$$

The signal that favors **H_a** comes from ***large values*** of z. Thus, rejection of **H_0** results when the computed **$z > z_\alpha$**. Obviously, if the alternative is $H_a: \mu < \mu_o$, the critical region is entirely in the lower tail and thus rejection results from **$z < -z_\alpha$**. Although in a one-sided testing case the null hypothesis can be written as **$H_0 : \mu \leq \mu_0$** or

H_0: $\mu \geq \mu_0$, it is usually written as **H_0: $\mu = \mu_0$.**

The test procedure just described is equivalent to find a 100(1-α) % confidence interval for μ and do not reject H_o if μ_o lies in the interval. If μ_o lies outside the interval, we reject H_o in favor of the alternative hypothesis H_a.

The steps for testing a hypothesis about a mean of a population with known variance σ^2 (or when n > 30) against various alternative hypothesis may be summarized as follows:

1- **H_o** : $\mu = \mu_o$.

2- H_a : Alternatives are $\mu < \mu_o$, $\mu > \mu_o$, or $\mu \neq \mu_o$

3- Choose a **level of significance** equate to **α**.

4- **Critical region**.

   $Z < -Z_\alpha$ for the alternative: $\mu < \mu_o$ (see Fig. 4.1a),

   $Z > Z_\alpha$ for the alternative: $\mu > \mu_o$ (see Fig. 4.1b),

   $|Z| > Z_{\alpha/2}$ for the alternative $\mu \neq \mu_o$ (see Fig.4.1c)

   where **Z** has a standard normal distribution.

5- **Compute** $\bar{x}$ from a random sample of size n, and then find

$$Z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$$

   (If σ is unknown but n > 30, replace σ by S in the above formula).

6. **Conclusion: Reject H_o** if **Z** falls in the critical region, otherwise, **do not reject H_o.**

Example 4.1

It has been found from experience that the mean breaking strength of a particular brand of thread is 9.50 ounces with a standard deviation of 1.4 ounces. Recently a sample of 36 pieces of thread showed a mean breaking strength of 8.94 ounces. Can one conclude, at a significance level 0.05, that the thread has become inferior?

Solution

A stepwise solution is as follows:

1. The null hypothesis is: H_o : $\mu = 9.5$ ounces, against.
2. The alternative hypothesis is: H_a: μ , 9.5.
3. Significance level: $\alpha = 0.05$.
4. Clearly σ is known and n > 30, so we can use the normal distribution. Thus $Z_\alpha = Z_{0.05} = 1.645$ and the critical region is: $Z < -1.645$.
5. Computations: **$\bar{x} = 8.94$, $\sigma = 1.4$ & n = 36, hence**

$$Z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} = \frac{8.94 - 9.5}{1.4/\sqrt{36}} = -2.4$$

6. Conclusion; Z = -2.4 falls in the critical region, i.e. reject H_o at 0.05 level of significance and we conclude that the thread has become inferior.

The P-value

The preselection of a significance level **α** has its roots in the philosophy that the maximum risk of making a type I error should be controlled. Over a number of generations of statistical analysis, it had become customary to choose **α** as 0.05 or 0.01 and select the critical region accordingly. However, this approach does not account for values of test statistics that are "close" to the critical region. Moreover, if other researchers want to apply the results of your study using a different value of **α** then they must compute a new rejection region before reaching a decision concerning $\mathbf{H_a}$ and $\mathbf{H_0}$. The ***P*-value** approach has been adopted extensively by users of applied statistics. The approach is designed to give the user an alternative (in terms of a probability) to a mere **"reject"** or **"do not reject"** conclusion. The ***P*-value** computation also gives the user important information when the *z*-value falls well ***into the ordinary critical region***.

Use of the ***P*-value** approach as an aid in decision-making is quite natural, and nearly all computer packages that provide hypothesis-testing computation print out *P*-values along with values of the appropriate test statistic. The following is a formal definition of a *P*-value.

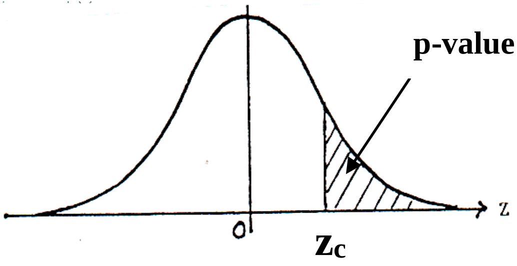
Definition: A *P*-value is the lowest level (of significance) at which the observed value of the test statistic is significant.

For the Z-test concerning the population mean **μ,** the **P-value** is calculated as:

$$\mathbf{P-value} = \begin{cases} \mathbf{P(Z > |z_c|)} & \textbf{for one-sided test} \\ \mathbf{2P(Z > |z_c|)} & \textbf{for two-sided test} \end{cases}$$

where $\mathbf{Z_c}$ is the calculated test statistic,

$$\mathbf{Z_c = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}}$$



The following decision rule yields results that will always agree with the testing procedures

1. If the **p-value $\leq \alpha$**, then **reject H_0.**
2. If the **p-value $> \alpha$**, then **fail to reject H_0.**

More precisely, reject H_0 for any significance level greater than the calculated p-value. For illustration, in example 7.1, since the calculated Z statistic is z_c = -2.923 and the test is two sided, then

$$\text{p-value} = 2P(Z > |3.25|) = 0.001$$

Recall that the **p-value** for a test is the smallest value of **α** for which the null hypothesis can be rejected. Since, in Example 4.1, our p-value is .001, we know that we could have chosen an **α** value as small as .001 and still have rejected the null hypothesis. If we had chosen an smaller than .001, we would not have been able to reject the null hypothesis.

Case 2 (*σ is unknown and n is small) :*

Many problems arise in which we are testing hypothesis concerning the population mean when σ^2 is unknown, and it is impractical to take a large sample. For n < 30 we must base our decision criterion on the t-distribution with v = n-1 degrees of freedom. The procedure is the same as for large samples except that we use t values in place of z values and replace σ by its estimate S. Consider the following one-tailed test:

$$\mathbf{H_o : \mu = \mu_o}, \qquad \textbf{against} \qquad \mathbf{H_a : \mu < \mu_o}.$$

The critical region will fall in the left tail of the t distribution. For a level of significance equal to α, we find a single critical value $-t_\alpha$ such that $T < -t_\alpha$ represent the critical region and $T > -t_\alpha$ constitutes the acceptance region, where

$$\mathbf{T = \frac{\bar{X} - \mu_o}{S / \sqrt{n}}}$$

To test a hypothesis about a mean of a population with unknown variance σ^2 against various alternative hypothesis, based on a sample of size n < 30, we proceed by the following steps:

1. $H_o : \mu = \mu_o$
2. H_a : Alternatives are $\mu < \mu_o$; $\mu > \mu_o$ or $\mu \neq \mu o$
3. Choose a level of significance equal to α.
4. Critical region: $t < -t_{\alpha/2}$ for the alternatives $\mu \neq \mu_o$

   where **t** has a t-distribution with (n-1) degrees of freedom.
5. Compute $\bar{x}$ and S from a random sample of size n, and then find

$$t_c = \frac{\bar{x} - \mu_o}{S/\sqrt{n}}.$$

Conclusion: Reject H_o if T falls in the critical region; otherwise accept H_o.

For the t-tests, the same procedure is carried out with finding the t-distribution $F(|t_c|)$ instead of the normal distribution $\Phi(|z_c|)$.

(a) $P < \alpha \Rightarrow$ reject H_0 **(b)** $P > \alpha \Rightarrow$ do not reject H_0

Example 4.2

The manufacturer of a power supply is interested in the mean of output voltage. He has tested 12 units, chosen at random, with the following results

5.34 4.90 5.07 5.25 5.65 5.45 5.25 5.35 4.86 5.54 5.44 4.90.

Test the hypothesis that the true mean voltage does not equal 5. Use α **= 0.05.**

Solution

1. **H_o** : $\mu = 5$ ml , against.
2. **H_a** : $\mu \neq 5$.
3. **α** = 0.05.
4. Clearly $\boldsymbol{\sigma^2}$ is unknown and n < 30, then we have to use the t-distribution with d.f. = ν = n-1=11, thus $t_{\alpha/2} = t_{0.025} = 2.201$. Hence the **critical region** is t < -2.201or t > 2.201.
5. **Computations:**

$$\bar{X} = \frac{1}{n}\sum x_i = 5.25, \quad s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2} = 0.2642 \text{ and } T_c = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} = 3.28$$

6. **Conclusion:**

   T = 3.28 > 2.201, i.e. T_c falls in the critical region and therefore reject H_o at 0.05 significance level. This means that the mean of output voltage does not equal 5.

   p-value = 2 P(T > 3.28) = 0.007 and therefore we reject H_0 at any significance level greater than 0.00072.

   Note that the 95% C.I. for μ is (5.08214, 5.41786) which does not include the value of $\boldsymbol{\mu_0}$.

4.5 Tests Concerning Two Populations

Suppose that we have two distinct populations, the first with mean $\boldsymbol{\mu_1}$, and variance $\boldsymbol{\sigma_1^2}$, and the second with mean $\boldsymbol{\mu_2}$ and variance $\boldsymbol{\sigma_2^2}$. Let $\bar{x}_1$ and $\bar{x}_2$ represent the means of random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from the first and second populations respectively.

We may wish to test hypothesis about the means of the two populations. We are essentially testing the hypothesis that the difference between the two populations means **(μ_1-μ_2)**, equals a specified value **d_o** i.e.

$$\mathbf{H_o : \mu_1 - \mu_2 = d_o}$$

The alternative hypothesis may be one of the usual alternatives;

$$\mathbf{H_o : \mu_1 - \mu_2 < d_o} \quad \textbf{or}$$

$$\mathbf{H_a : \mu_1 - \mu_2 > d_o} \quad \textbf{or}$$

$$\mathbf{H_a : \mu_1 - \mu_2 \neq d_o}$$

when $d_o = 0$ we are testing the null hypothesis that the two population means are equal. As we know the particular procedure used in carrying out a test of hypothesis between two means will depend on the distribution of the population involved, the size of the independent samples selected and the knowledge about the population variances.

Case 1: σ_1 & σ_2 are known (or n_1 & n_2 are large)

Consider first the problem of testing the null hypothesis H_o that the difference between two population means, **μ_1 -μ_2**, equals a specified value **d_o** when the variances **σ_1^2** and **σ_2^2** are known and one can assume that the two populations are approximately normally distributed. In the case of large samples, the assumption of normally may be relaxed and, if necessary, **S_1^2** and **S_2^2** may be substituted for **σ_1^2** and **σ_2^2** respectively.

Our decision to accept or reject H_o is based on the Z value corresponding to the value assumed by the random variable $(\bar{X}_1 - \bar{X}_2)$.

First, we select two independent random samples, one from each population of size n_1 and compute the difference, $\bar{X}_1$ **and** $\bar{X}_2$ of the sample means. From the sampling distribution of $\bar{X}_1 - \bar{X}_2$ we know that

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_o}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(or when σ_1 & σ_2 are unknown but n_1 and n_2 are large enough; we replace σ_1 & σ_2 by S_1 & S_2 respectively) has a standard normal distribution when H_o is true. Hence for a two-tailed test at the level of significance, the critical region is $Z < -z_{\alpha/2}$ and $Z > z_{\alpha/2}$. In the case of a one-tailed test for which the alternative hypothesis is μ_1-$\mu_2 < d_o$, the critical region of size α is $Z < -z_\alpha$. Similarly, for the alternative
μ_1-$\mu_2 > d_o$, the critical region of size α is $Z > z_\alpha$.

To test a hypothesis about the difference between two population means when σ_1^2 and σ_2^2 are known, we proceed by the same basic six steps as outlined above.

Example 4.3

Two types of chemical solutions, A and B, were tested for their pH (degree of A acidity of the solution). Analysis of 40 samples of A showed a mean pH of 7.55 with a standard deviation of 0.06. Analysis of 50 samples of B showed a mean pH of 7.48 with a standard deviation of 0.08. Test the hypothesis that the two types of solutions have different pH values using level of significance of 0.01.

Solution

Following the six-step procedure, we have

1. $H_o : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ (i.e. $d_o = 0$)
2. $H_a : \mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$
3. $\alpha = 0.01$
4. $Z_{\alpha/2} = Z_{0.005} = 2.58$. Hence the critical region is $Z< -2.58$ or $Z > 2.58$.
5. **Computations:** $n_1 = 40$, $\bar{x}_1 = 7.55$, $S_1 = 0.06$

   $n_2 = 50$, $\bar{x}_2 = 7.48$, $S_2 = 0.08$

Thus

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_o}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = \frac{0.07 - 0}{\sqrt{0.0036/40 + 0.0064/50}} = 4.74$$

6. **Conclusion:** Reject H_o and conclude that the two types of solutions have different pH values.

Case 2: σ_1 & σ_2 are unknown and n_1 & n_2 are small

Suppose that σ_1^2 and σ_2^2 are unknown and the sample sizes are small. In this case our decision to accept or reject on the null hypothesis H_o that $\mu_1 - \mu_2 = d_o$ is based on the t distribution with (n_1+n_2-2) degrees of freedom, provided that and the populations are approximately normally distributed.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

The procedures is the same as for large samples except that Z statistic is replaced by

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_o}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

which is known as the common (or pooled) variance. Hence, for a two tailed test at the level of significance α, the critical region is $T < -t_{\alpha/2}$ and $T > t_{\alpha/2}$. For the alternative $\mu_1 - \mu_2 < d_o$, the critical region is $T < -t_\alpha$ and for alternative $\mu_1 - \mu_2 > d_o$, the critical region is $T > t_\alpha$.

Example 4.4

A sample of 9 plants of the same variety was grown in one type of soil in a greenhouse, and after a fixed time they were removed and dried. Their dry weights were 27.5, 22.3, 24.7, 26.1, 26.5, 20.0, 31.0, 25.3, 28.6 g. A further sample of all similar plants was grown in identical conditions but in another type of soil. Their weights were 31.8, 30.3, 26.4, 24.2, 27.8, 29.1, 25.5, 28.9, 30.0, 24.9, 31.7 g. Do the two soil types have different effects on the plants?

Solution

1- $H_o : \mu_1 = \mu_2$ against,

2- $H_a : \mu_1 \neq \mu_2$.

3- Choose $\alpha = 0.05$

4- It is a two sided test, in which, σ_1 & σ_2 are unknown and n_1 & n_2 are small, thus the critical region is given using the t- distribution, with d.f. = $v = n_1 + n_2 - 2 = 18$, $t_{\alpha/2}=t_{0.025}=2.1$. Hence the critical region is $T < -2.1$ or $T > 2.1$.

5- **Computations:** From the data given above, the following results can be easily computed as before.

$$n_1 = 9, \bar{x}_1 = 25.78, \quad s_1 = 3.27,$$
$$n_2 = 11, \bar{x}_2 = 28.24, \quad s_2 = 2.68.$$

Hence

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 8.74 \Rightarrow S_p = 2.957$$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - d_o}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -1.85, \quad \text{p-value} = 0.081$$

7. **Conclusion:** Do not reject H_o at 0.05 level of significance and we conclude that the two types of soil have the same effects on the plants.

Case 3: Paired Comparisons

In sampling from two populations it sometimes happens that measurements are taken on individuals prior to and following treatment. By examining the difference of all pairs of measurements we hope to draw a conclusion about the effectiveness of the treatment.

On the other hand, sometimes, the extraneous factors cause a significant difference in means whereas there was no difference in effects we are trying to measure. Suppose we wish to compare the mean viscosity of two chemicals X and Y. If the average of n samples of chemical X is compared to the average of n samples of chemical Y part of the difference may be due to temperature, humidity, ext. and to chemical effect. It might be that the chemicals cause a difference, this difference is obscured by other factors such as different temperature or any other extraneous effects. To obviate this difficulty, observations are taken in pairs:

$$(\mathbf{x_1, y_1}),\ (\mathbf{x_2, y_2}),\ \ldots,\ (\mathbf{x_n, y_n})$$

each pair being observed under the same experimental conditions. Let

$$\mathbf{d_1 = x_1 - y_1,\ d_2 = x_2 - y_2, \ldots, d_n = x_n - y_n.}$$

It is assumed that the differences between each pair are independent normal variates with unknown mean μ_d and unknown variance σ_d^2. The mean of the population of differences, μ_d, will be equal to the difference of the two population means, $\mu_1 = \mu_2$. Therefore, the problem of testing the null hypothesis $Ho_: \mu_1 - \mu_2 = d_o$ is equivalent to testing $\mu_d = d_o$ were d_o is some specified value. The alternative hypothesis, H_o, might be any one of the following:

$$\boldsymbol{\mu_d < d_o\,,\ \mu_d > d_o\ \text{ or }\ \mu_d \neq d_o}$$

By pairing observations, we have essentially reduced a two-sample problem to a single random sample of difference and in doing so we can test the null hypothesis that $\mu_d = d_o$ by using the methods described before. The decision whether to accept or reject H_o is, therefore, based on the statistic

$$\mathbf{t = \frac{\bar{d} - d_o}{S_d / \sqrt{n}}}$$

which has a t-distribution with $\boldsymbol{\nu = n\text{-}1}$ degrees of freedom, and the critical regions, corresponding to the various alternative hypothesis, are set up as before.

Example 4.5

Nine adults agreed to test the efficiency of a new diet program to decrease the weight. Their weights (pounds) were measured before and after the program and found to be as follows:

Adults :	1	2	3	4	5	6	7	8	9
Before :	132	139	126	114	122	132	142	119	126
After :	124	141	118	116	114	132	145	123	121

Test the efficiency of this program for decreasing the weight at level $\alpha = 0.01$?

Solution

1. It is required to test the null hypothesis that the diet is not effective, $H_o:\mu_1 = \mu_2$ against:
2. The alternative $H_a : \mu_1 > \mu_2$, that is effective.
3. $\alpha = 0.01$
4. d.f. $v = n-1 = 8$, $t_\alpha = t_{0.01} = 2.9$

   $\therefore$ critical region is $T > 2.9$.
5. **Computations:** The differences are:

   d_i : 8 -2 8 -2 8 0 -3 -4 5

   Hence

   $$\bar{d} = \frac{1}{n}\sum d_i = 18/9 = 2,$$

   $$S_d^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2 = 26.75 \Rightarrow S_d = 5.17$$

   $$\therefore T = \frac{\bar{d} - d_o}{S_d/\sqrt{n}} = \frac{2-0}{5.17/\sqrt{9}} = 1.16$$
6. **Conclusion:** Don't reject H_o at 0.01 level of significance i.e. the diet program is not effective.

4.6 Tests Concerning the Population Variance

Null hypothesis: $H_o : \sigma^2 = \sigma_0^2$

Test statistic : $\chi_c^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

Significance level: α

Alternative hypothesis	Rejection	Region	P-Value

$H_a : \sigma^2 > \sigma_0^2$	**$\chi_c^2 > \chi^2_{\alpha, n-1}$**	**(One sided)**	**$P = P(\chi^2 > \chi_c^2)$**
$H_a : \sigma^2 < \sigma_0^2$	**$\chi_c^2 < \chi^2_{1-\alpha, n-1}$**	**(One sided)**	**$P = P(\chi^2 < \chi_c^2)$**
$H_a : \sigma^2 \neq \sigma_0^2$	**$\chi_c^2 > \chi^2_{\alpha/2, n-1}$ or $\chi_c^2 < \chi^2_{1-\alpha/2, n-1}$**	**(Two sided)**	**$P = 2P(\chi^2 > \chi_c^2)$ or $P = 2P(\chi^2 < \chi_c^2)$**

Example 4.6

Suppose that the thickness of a part used in a semiconductor is its critical dimension and that measurements of the thickness of a random sample of 18 such parts have variance 0.68, where the measurements are in thousands of an inch. The process is considered to be under control if the variance of the thickness is not greater than 0.36. Test the hypotheses that the process used in the manufacture of these parts is under control at 0.05 level of significance.

Solution

We wish to test $H_0 : \sigma^2 = 0.36$ against $H_a : \sigma^2 > 0.36$ (one sided test).

Assuming that measurements constitute a random sample from a normal population, then under H_0, $\chi_c^2 = \dfrac{(n-1)\,S^2}{0.36} \sim X^2_{n-1}$ so reject H_0 if $\chi_c^2 > \chi^2_{\alpha, n-1} = \chi^2_{0.05,17} = 27.587$.

Now, $\chi_c^2 = \dfrac{(17)\,(0.68)}{0.36} = 32.11 > 27.587$, hence the null hypothesis must be rejected and the process used in the manufacture of the parts must be adjusted.

The P-value is then given by:

$$P = P(\chi^2 > \chi_c^2) = P(\chi^2 > 32.11) = 0.0146 < \alpha \quad (\text{Reject } H_o)$$

4 .7 Tests For the Equality of Two Variances

Null hypothesis: $H_o : \sigma_1^2 = \sigma_2^2$, **Alternative hypothesis** $H_a : \sigma_1^2 \neq \sigma_2^2$

Test statistic : $F = \dfrac{S_1^2}{S_2^2}$

Significance level: α

Alternative hypothesis	Rejection	Region	P-Value
$H_a : \sigma_1^2 > \sigma_2^2$	**$F_c > F_{\alpha,\, n_1-1,\, n_2-1}$**	**(One sided)**	**$P = P(F > F_c)$**
$H_a : \sigma_1^2 < \sigma_2^2$	**$F_c < F_{1-\alpha,\, n_1-1,\, n_2-1}$**	**(One sided)**	**$P = P(F < F_c)$**

$H_a : \sigma_1^2 \neq \sigma_2^2$	$F_c > F_{\frac{\alpha}{2},\, n_1-1,\, n_2-1}$ or $F_c < F_{1-\frac{\alpha}{2},\, n_1-1,\, n_2-1}$	**(Two sided)**	$P = 2P(F > F_c)$ **or** $P = 2P(F < F_c)$

Example 4.7

Test the hypothesis that the variance of the two distributions, in example 4.4, are equal. Use $\alpha = 0.02$.

Solution

We wish to test $\mathbf{H_0 : \sigma_1 = \sigma_2}$ against $\mathbf{H_a : \sigma_1 \neq \sigma_2}$ (Two sided test).

Assuming that measurements constitute random samples from normal populations, then under H_0, $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$ so reject H_0 if

$$F_c = \frac{S_1^2}{S_2^2} > F_{\frac{\alpha}{2}; n_1-1, n_2-1} = F_{.02;8,10} = 4.13 \ .$$

Now, $F_c = \frac{(3.27)^2}{(2.68)^2} = 1.49 < 4.13$, **p-value = 0.542**, hence the null hypothesis cannot be rejected and we conclude that the two populations are homogeneous.

<+><+><+><+><+><+><+><+>

EXERCISES

[1] The life in hours of a battery is known to be approximately normally distributed, with standard deviation $\boldsymbol{\sigma} = 1.25$ hours. A random sample of 10 batteries has a mean life of $\bar{\mathbf{x}}$ **=40.5** hours.

a- Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.05$.

b- What is the P-value for the test in part (a)?

[2] It has been found from experience that the mean breaking strength of a particular brand of thread is 9.50 ounces with a standard deviation of 1.4 ounces. Recently a sample of 36 pieces of thread showed a mean breaking strength of 8.94 ounces. Can one conclude, at a significance level 0.05, that the thread has become inferior?

[3] The yield of a chemical process is being studied. From previous experience yield is known to be normally distributed and $\sigma = 3$. The past five days of plant operation have resulted in the following percent yields: 91.6, 88.75, 90.8, 89.95, and 91.3.

a- Find a 95% confidence interval on the true mean yield.

b- Is there evidence that the mean yield is not 90%? Use $\alpha = 0.05$

c- What is the P-value for this test?

[4] The following data are the oxygen uptakes (milliliters) during incubation of a random sample of 15 suspensions: 14.0 14.1 14.5 13.2 11.2 14.0 14.1 12.2 11.1 13.7 13.2 16.0 12.8 14.4 12.9. Do these data provide sufficient evidence at the .05 level of significance that the population mean is not 12 ml?

[5] Test the hypothesis that the average weight of boxes dog food is 10 ounces if the weights of a random sample of 10 boxes are 10.2 9.7 10.1 10.3 10.1 9.8 9.9 10.4 10.3 9.8 ounces. Use a 0.01 level of significance and assume the distribution of weights to be approximately normal.

[6] The mean time taken for mice to die when injected with 1000 leukemia cells is known to be 12.5 days. When the injection does was doubled in a sample of 10 mice, the survival time were 11.7, 10.5 11.2 12.9 12.7 10.3 10.4 10.9 11.3 10.6. If the survival times are normally distributed do the results suggest that the increased injection does has decreased survival ship?

[7] In population the weight of babies at birth is normally distributed with mean 3200 g. A group of fifteen babies in which a low protein level was, detected in mothers during pregnancy gave the following weights at birth 2305, 2908, 2577, 2275, 2941, 3680, 2997, 2403, 3089, 3176, 3280, 3337, 2847, 2927, 2751. Determine if mothers with a low protein level give birth to babies having a lower birth weight, than those from the whole population.

[8] Can we conclude that the mean maximum voluntary ventilation value for apparently healthy college seniors is not 110 liters per minute (let $\alpha = .01$)? A sample of 20 yielded the following values: 132 33 91 108 67 169 54 203 190 133 69 30 187 21 63 166 84 110 157 138.

[9] The following are the systolic blood pressures (mm Hg) of 12 patients undergoing drug therapy for hypertension: 183 152 178 157 194 163 144 114 178 152 118 158. Can we conclude on the basis of these data that the population mean is less than 165? Let $\alpha = .05$.

[10] Researcher wished to know if they could conclude that two populations of infants differ with respect to mean age at which they walked alone. The following data (ages in months) were collected

Sample from population A: 9.5, 10.5, 9.0, 9.75, 10.0, 13.0, 10.0, 13.5, 10.0, 9.5, 10.0, 9.75

Sample from population B: 12.5, 9.5, 13.5, 13.75, 12.0, 13.75, 12.5, 9.5, 12.0, 13.5, 12.0, 12.0

What should the researchers conclude? Let $\alpha = .05$.

[11] Does sensory deprivation have an effect on a person's alpha-wave frequency? Twenty volunteer subjects were random divided into two groups. Subjects in group A were subjected to a 10-day period of sensory deprivation, while subjects in group B served as controls. At the end of the experimental period the alpha-wave frequency component of subjects electroencephalograms were measured.

The results were as follows:

Group A: 10.2, 9.5, 10.1, 10.0, 9.8, 10.9, 11.4, 10.8, 9.7, 10.4

Group B: 11.0, 11.2, 10.1, 11.4, 11.7, 11.2, 10.8, 11.6, 10.9. 10.7

Let $\alpha = .05$.

[12] Two methods for producing gasoline from crude oil are investigated. The yields (%) of both processes are assumed to be normally distributed. The following yield (%) data have been obtained from the pilot plant

Process	Sample size	Sample Mean	Sample variance
A	15	25.8	0.64
B	12	21.2	0.81

a- Find a 98% confidence interval for the true mean yield of process A.

b- Test at the 0.02 significant level, whether it is reasonable to assume that the two populations of yields (%) of both processes A and B have equal variances.

c- Test at the 0.05 level of significance whether the difference between the true means yield of processes A and B is significant (state clearly H_0, H_a, decision and conclusion).

[13] An investigation of repair times for two kinds of photocopying equipment obtained these data values:

Equipment type	Number of repair jobs	Sample Mean	Sample variance
1	25	84.2	19.4
2	20	91.6	18.2

All times are in minutes

a- Test at the 0.01 level of significant whether it is reasonable to assume that the populations of the two kinds of photocopying equipment have equal variances.

b- Test at the 0.02 significant level, whether the difference between these two sample means is significant.

[14] The following table gives the weight (in kg) of 6 women before and after the fasting month of Ramadan:

Women	1	2	3	4	5	6
Weight before Ramadan	70.0	71.5	72.1	56.8	63.7	59.5
Weight after Ramadan	71.2	72.2	72.1	60.1	60.1	59.0

Do the data provide evidence that the average weight of women tend to increase during the month of Ramadan? Use a 0.01 level of significance.

[15] The following are the average weekly losses of man-hours due to accidents in 9 industrial plants before and after a certain safety program were put into operation.

Before the safety program	45	73	46	124	33	57	83	34	17
After the safety program	44	65	44	119	35	56	80	31	11

Use the 0.05 level of significance to test whether the safety program is effective. State, clearly, H_0 and H_a , your decision and the conclusion.

[20] Answer (T/F) true or false:

a- **The** alternative hypothesis is stated in terms of a sample statistic.

b- a large P-value indicates strong evidence against H_o.

c- If there is sufficient evidence to reject H_o at α=0.10, then there is sufficient evidence to reject it also at α=0.05.

d- If the population mean is known, there is no reason to run a hypothesis test on the population mean .

e- The P-value is usually chosen before an experiment is conducted.

f- A well-planned test of significance should result in a statement either that H_o is true or that it is false.

<+><+><+><+><+><+><+><+>