Testing of Hypothesis

A statistical hypothesis is an assertion or conjecture concerning population.

The truth or 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 data contained in this sample to provide evidence that other supports or does not support the hypothesis. Evidence from sample that is inconsistent with the stated hypothesis leads too a rejection of the hypothesis, where as evidence supporting the hypothesis leads to its acceptance.

Null Hypothesis and Alternative Hypothesis

The hypothesis which is to be tested is called *null hypothesis* and is denoted by H_0. The rejection of H_0 leads to the acceptance of an *alternative hypothesis*, which is denoted by H_1. A null hypothesis concerning a population parameter will always be stated so as to specify an exact value of the parameter, where as the alternative hypothesis allows for the possibility of several values. Hence, if H_0 is the null hypothesis $p = 0.5$ for a binomial population, the alternative hypothesis H_1 would be one of the following:

$$p > 0.5, \quad p < 0.5, \quad p \neq 0.5.$$

To illustrate the concepts used in testing a statistical hypothesis about a population, consider the following example:

$$Refer\ example : 10.2.$$

Here we are testing $H_0 : p = 1/4$ against $H_1 : p > 1/4$. Here test statistic is X, the number of individuals in our test group of 20 who receive protection from the new vaccine for a period of atleast 2 years. Definitely, X can assume values from 0 to 20. Here, the critical region is $X > 8$. $X \geq 8$ determine the acceptance region. 8 is called critical value. Thus if $x > 8$, we reject H_0 and if $x \leq 8$, we accept H_0.

type I error $=$ (Reject $H_0|H_0$)
type II error $=$ (Accept $H_0|H_0$ is false)

In testing any statistical hypothesis, there are four possible situations that determine whether our decision is correct or not.

	H_0 is true	H_0 is false
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

P(Type I error) $=$ level of significance, α.
In the above example,

$$\begin{aligned}\alpha &= P\,(Type\ I\ error)\\ &= P\,(X > 8\ when\ p = 1/4)\\ &= \sum_{x=9}^{20} b(x : 20, 1/4)\\ &= 0.0409\end{aligned}$$

Some times α is called the size of the test. This means that $H_0 : p = 1/4$, is being tested at the $\alpha = 0.0409$ level of significance.

The probability of committing a type II error, denoted by β, is impossible to compute unless we have a specific alternative hypothesis. If the alternative hypothesis is say, $H_1 : p = 1/2$, then

$$\begin{aligned}\beta &= P\,(type\ II\ error)\\ &= P\,(X \le 8\ when\ p = 1/2)\\ &= \sum_{x=0}^{8} b(x : 20, 1/2)\\ &= 0.2517.\end{aligned}$$

Ideally, we like to use a test procedure for which the *type I* and *type II* error probabilities are both small. A reduction in β is possible by increasing the size of the critical region. *For example*, assume that our critical value is 7. Then in testing $H_0 : p = 1/4$ against $H_1 : p = 1/2$, we have $\alpha = \sum_{x=8}^{20} b(x; 20, 1/4) = 0.1018$ and $\beta = \sum_{x=0}^{7} b(x; 20, 1/2) = 0.1316$.

For a fixed sample size, a decrease in the probability of one error will usually result in an increase in the probability of the other error. Fortunately, the probability of committing both types of error can be reduced by increasing the sample size.

For example, consider the same problem using a random sample of 100 individuals. Assume that the critical value is 36.
Here $X \to N(np, \sqrt{npq})$.

$$\alpha = P(type\ I\ error) = P(X > 36\ when\ p = 1/4) = P(Z > 2.66) = 0.0039$$

Similarly,

$$\beta = P(type\ II\ error) = P(X \le 36\ when\ p = 1/2) = P(Z < -2.7) = 0.0035$$

Example:- Consider the null hypothesis that the average weight of male students in a certain college is $68kg$ against the alternative hypothesis that it is unequal to $68kg$. *i.e*, we wish to test

$$H_0 : \mu = 68$$

$$H_1 : \mu \neq 68 \quad i.e., \mu > 68\ or\ \mu < 68$$

Let the test statistic be the sample mean. Assume that a random sample of size 36 is taken. Suppose that the critical region is $\overline{x} < 67$ and $\overline{x} > 69$ (arbitrarily chosen). Then the acceptance region is $67 \le \overline{x} \le 69$.

$$\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Let $\sigma = 3.6$.

$$\begin{aligned}\alpha &= P\left(\overline{X} < 67|\mu = 68\right) + P\left(\overline{X} > 69|\mu = 68\right)\\ &= P\,(Z < -1.67) + P\,(Z > 1.67)\\ &= 2P(Z < -1.67)\\ &= 0.0950\end{aligned}$$

Thus 9.5% of all samples of size 36 would lead us to reject $H_0(i.e.,\ \mu = 68kg)$ when, in fact, it is true. To reduce α, we have a choice of increasing the sample size or widening the acceptance region. Suppose that we increase the sample size to $n = 64$. Then $\sigma_{\overline{X}} = 0.45$. Hence

$$\begin{aligned}\alpha &= P\,(Z < -2.22) + P\,(Z > 2.22)\\ &= 2P(Z < -2.22)\\ &= 0.0264\end{aligned}$$

The reduction in α is not sufficient by itself to guarantee a good testing procedure. We must evaluate β for various alternative hypothesis that we should be accepted. Let H_1 be either $\mu = 66$ or $\mu = 70$. *For example,* let it be $\mu = 70$. Then,

$$\begin{aligned} \beta &= P\left(67 \leq \overline{X} \leq 69 \; when \; \mu = 70\right) \\ &= P\left(-6.67Z < -2.22\right) \\ &= 0.0132 \end{aligned}$$

For all positive values of $\mu < 66$ or $\mu > 70$, the value of β will be even smaller when $n = 64$, and consequently there would be little chance of accepting H_0 when it is false. The probability of committing *type II* error increases rapidly when the true value of μ approaches, but is not equal to, the hypothesized value. *For example,* if $H_0 : \mu = 68$ against $H_1 : \mu = 68.5$, then $\beta = P(67 \leq \overline{X} \leq 69 \; when \; \mu = 68.5) = 0.8861$.

Important Properties of a Test of Hypothesis

1. The type I and type II error are related. A decrease in the probability of one generally results in an increase in the probability of the other.
2. The size of the critical region, and therefore the probability of committing a type I error, can always be reduced by adjusting the critical value.
3. An increase in the sample size n will reduce α and β simultaneously.
4. If the null hypothesis is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The grater the distance between the true value and the hypothesized value, the smaller β will be.

The *power* of the test is the probability of rejecting H_0 given that a specific alternative is true. *i.e.,* $power = 1 - \beta$.

One-tailed and Two-tailed Tests

A test of any statistical hypothesis, where the alternative is one-sided, such as $\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{array}$ or $\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{array}$ is called a *one tailed test.*

Generally, the critical region for the alternative hypothesis $\theta > \theta_0$lies in the right tail of the distribution of the test statistic, while the critical region for the alternative hypothesis $\theta < \theta_0$ lies entirely in the left tail.

A test of any statistical hypothesis where the alternative is two sided, such as $\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array}$ is called a *two-tailed test,* since the critical region is split into two parts, often having equal probabilities placed in each tail of the distribution of the test statistic.

Important Steps in Testing

1. State the null and alternative hypothesis. **Null Hypothesis Ho :** **Alternative Hypothesis H1 :**
2. Choose a fixed significance level α.
3. Choose an appropriate test statistic and establish the critical region based on α.
4. From the computed test statistic, reject H_0 if the test statistic is in the critical region. Otherwise, do not reject.
5. Draw scientific or engineering conclusions.

Test Concerning a Single $\boxed{\text{Mean}}$ $\boxed{(\sigma^2 \text{ is known})}$

Let the hypothesis be $\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0. \end{array}$

$\overline{X}$ is the estimate of μ. For large n, $\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

$\Rightarrow Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to N(0,1)$.

Let α be the level of significance under H_0,

i.e., if $\mu = \mu_0$, $\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \to N(0,1)$.

If for a particular sample, $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > Z_{\alpha/2}$ or $z < -Z_{\alpha/2}$, we reject H_0 and if $-Z_{\alpha/2} < z < Z_{\alpha/2}$, accept H_0. Or equivalently, if $\overline{x} > \mu_0 + Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ or $\overline{x} < \mu_0 - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$, we reject H_0.

Tests of one sided hypotheses on the mean involve the same statistic described in the two sided case. The difference is that the critical region is only in one tail of the standard normal distribution. *For example,* if we seek to test $\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0, \end{array}$ the rejection of H_0 results when $z > Z_\alpha$. If $H_1 : \mu < \mu_0$, rejection of H_0 when $z < -Z_\alpha$.

Note: The testing of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at a significance level α is equivalent to computing a $(1-\alpha)100\%$ confidence interval for μ and rejecting H_0 if μ_0 is not inside the confidence interval. If μ_0 is inside the confidence interval, we accept the hypothesis.

Example 1: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population *S.D* of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

$$H_0 : \mu = 70 \ years$$

$$H_1 : \mu > 70 \ years$$

$$\alpha = 0.05,$$

Critical region $z > 1.645$, where $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$.

$$z = \frac{71.8 - 70}{8.9/10} = 2.01 > 1.645.$$

Hence, the decision is rejected H_0 and conclude that the mean life span today is greater than 70 years.

Example 2: A manufacturer of sports equipments has developed a new synthetic fishing line that he claims has a mean breaking strength of 8 kg with a *S.D* of 0.05 kg. Test the hypothesis that $\mu = 8kg$ against the alternative that $\mu \neq 8kg$, if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kg. Use a 0.01 level of significance.

$$H_0 : \mu = 8$$

$$H_1 : \mu \neq 8$$

$$\alpha = 0.01,$$

Critical region: $z < -2.575$ and $z > 2.575$ where $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$.

Here $z = -2.83 < -2.575$.

Hence reject H_0 and conclude that the average breaking strength is not equal to 8 but it, in fact, less than 8 kg.

Choice of Sample Size for Testing Means

In most practical circumstances, the sample size is usually made to achieve good power for a fixed α and fixed specific alternative.

Suppose we wish to test the hypothesis $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$ with a significance level α when the variance σ^2 is known. Assume that the critical region is $\overline{x} > a$. For a specific alternative, say $\mu = \mu_0 + \delta$, we have,

$$\begin{aligned}
\beta &= P\left(\overline{X} > a | \mu = \mu_0 + \delta\right) \\
&= P\left(\frac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} < \frac{a - (\mu_0 + \delta)}{\sigma/\sqrt{n}}\right) \\
i.e., \quad \beta &= P\left(Z < \frac{a - \mu_0}{\sigma/\sqrt{n}} - \frac{\delta}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z < Z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z < -Z_\beta\right)
\end{aligned}$$

and hence,

$n = \dfrac{(Z_\alpha + Z_\beta)^2\sigma^2}{\delta^2}$, which is also true when $H_1 : \mu < \mu_0$.

In case of two-tailed test we obtain the power $1-\beta$ for a specified alternative when $n \approx \dfrac{(Z_\alpha + Z_\beta)^2\sigma^2}{\delta^2}$.

Example: Suppose that we wish to test $H_0 : \mu = 68\ kg$, $H_1 : \mu > 68\ kg$ for the weight of male students at a certain college using $\alpha = 0.05$ level of significance when it is known that $\sigma = 5$. Find the sample size required if the power of our test is to be 0.95 when the true mean is 69 kg.

$$\alpha = \beta = 0/05, \quad Z_\alpha = Z_\beta = 1.645, \quad \delta = 1, \quad n = 270.6$$

Therefore, 271 observations are required if the test is to reject the null hypothesis 95% of the time when in fact μ is as large as 69 *kgms*.

Test Concerning a Single Mean (~~σ^2 is Known~~) unk

Let $X_1, X_2, \ldots, X_n$ be a random sample taken from a normal population with mean μ and unknown variance σ^2. Then $\dfrac{X - \mu}{S/\sqrt{n}} \to t_{n-1}$. For the two-sided hypothesis, $\begin{matrix} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{matrix}$, rejection of H_0 at significance level α results when a computed $t-$statistic $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} > t_{\alpha/2,n-1}$ or $t < -t_{\alpha/2,n-1}$. For $H_1 : \mu > \mu_0$, rejection results when $t > t_{\alpha,n-1}$ and for $H_1 : \mu < \mu_0$,, critical region is given by $t < -t_{\alpha,n-1}$.

Example: An institute has published figures on the annual number on kilowatt-hours expended by various home appliances. It is claimed that a vacuum cleaner expends an average of 46 kilowatt-hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners expend an average of 42 kilowatt-hours per year with a S.D. of 11.9 kilowatt-hours, does this suggest at the 0.05 level of significance that vacuum cleaners expend, on the average, less than 46 kilowatt-hours annually? Assume the population of kilowatt-hours to be normal.

$\begin{matrix} H_0 : \mu = 46 \\ H_1 : \mu < 46 \end{matrix}$ $\quad \alpha = 0.05$. Critical region $t < -1.796$ where $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ with 11 $d.o.f.$

$$\overline{x} = 42, \quad s == 11.9, \quad n = 12.$$

Here $t = -1.16$. Hence do not reject H_0, and conclude that the average number of kilowatt-hours expended annually by home vacuum cleaners is not significantly less than 46.

$\boxed{\text{Test on Single Proportion}}$

Consider the hypotheses $\begin{matrix} H_0 : p = p_0 \\ H_1 : p > p_0. \end{matrix}$ The test statistic is X, the number of successes in n trials when n is large. If p_0 is neither close to zero or close to one, then $X \to N(npq, \sqrt{npq})$, so that in this case $Z = \dfrac{X - np}{\sqrt{npq}} \to N(0, 1)$. Under H_0, the value of Z, say $Z_0 = \dfrac{x - np_0}{\sqrt{np_0q_0}}$. For the one-tailed test at the $\alpha-$level significance, the critical region is $Z_0 > Z_\alpha$, or $Z_0 < -Z_\alpha$ according to $H_1 : p > p_0$ or $p < p_0$ respectively. Similarly for $H_1 : p \neq p_0$, the critical region is given by $Z_0 > Z_{\alpha/2}$ and $Z_0 < -Z_{\alpha/2}$. *Example:* A commonly prescribed drug for relieving nervous tension is believed to be only 60% defective. Experimental results with a new drug administrated to a random sample of $\boxed{100}$ adults who were suffering from nervous tension show that 70 received relief. Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed? Use a 0.05 level of significance.

$$\begin{matrix} H_0 : p = 0.6 \\ H_1 : p > 0.6 \end{matrix} \quad \alpha = 0.05, \quad \boxed{n = 100} \text{ Critical region is } Z > 1.645$$

Sample Size > 30
Use Z disbn

$$Z_0 = \frac{x - np_0}{\sqrt{np_0q_0}} = \frac{70 - 100 \times 0.6}{\sqrt{100 \times 0.6 \times 0.4}} = 2.04 > 1.645.$$

Hence reject H_0 and conclude that the new drug superior.

$\boxed{\text{Test Concerning Variances}}$

Consider $\begin{matrix} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{matrix}$. Assume that the distribution of the population being sampled normal, the statistic for testing $\sigma^2 = \sigma_0^2$ is given by $\boxed{\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}.}$

If H_0 is true, χ^2 is a value of the *Chi*$-$squared distribution with $n - 1$ *d.o.f.* Hence, for a two-tailed test at the $\alpha-$level of significance, the critical region is $\chi^2 < \chi^2_{1-\alpha/2}$ and $\chi^2 > \chi^2_{\alpha/2}$. For the one-sided alternative $\sigma^2 < \sigma_0^2$, the critical region is $\chi^2 < \chi^2_{1-\alpha}$, and for the one sided alternative $\sigma^2 > \sigma_0^2$, the critical region is $\chi^2 > \chi^2_\alpha$.

Example: A manufacturer of car batteries claims that the life of his batteries is approximately normally distributed with a S.D. equal to 0.9 years. If a random sample of 10 of these batteries has a S.D. of 1.2 years, do you think that $\sigma > 0.9$ years? Use 0.05 level of significance.

$$\begin{matrix} H_0 : \sigma^2 = 0.81 \\ H_1 : \sigma^2 > 0.81 \end{matrix} \;, \alpha = 0.05, \quad n = 10, \quad s = 1.2.$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \times 1.2^2}{0.81} = \frac{9 \times 1.44}{0.81} = 16.$$

$\chi^2_{0.05} = 16.919$. Hence $\chi^2 < \chi^2_{0.05}$, so we do not reject H_0. Hence $\sigma = 0.9$

$\boxed{\text{Assignment}}$

1. Test $\begin{matrix} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{matrix}$ [Population variances are known and are σ_1^2 and σ_2^2 respectively, large sample tests].

2. Test $\begin{matrix} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{matrix}$ [Populations from which the samples are taken are approximately normal].

It two random samples of size n_1 and n_2 are taken from a normal population having means μ_1 and μ_2 and having common variance, then

$$t = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n_1+n_2-2}$$

where, $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$.