MODULE-3

STATISTICAL INFERENCE

A population in statistical language is a collection of individuals or their attributes, rather than the individuals themselves. A population is finite or infinite, according as the number of elements is finite or infinite. A finite subset of a population is called sample and the process of selection of such samples is called sampling. The basic objective of the theory of sampling is to draw inference about the population using the information of the sample. If x_i be the value of x for the i^{th} member of the sample, then $(x_1, x_2, \dots x_n)$ are referred to as sample observations. Statistical measures such as mean and variance calculated on the basis of the population values of x are called parameters and sample observations are called statistics.

The probability distribution of the statistic that would be obtained if the number of samples, each of same size were infinitely large is called the sampling distribution of the statistics. The standard deviation of the sampling distribution of a statistic is called the standard error of the statistic.

The confidence intervals

Let $x_1, x_2, \dots x_n$ be a random sample from a distribution with mean μ and standard deviation σ. Then the sample mean $\bar{x}$ has approximately a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The 95% confidence interval for the population mean μ is $\left[\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right), \ \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$ or

$$\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right).$$

The interval width, $W = 2(1.96)\left(\frac{\sigma}{\sqrt{n}}\right)$

The 99% confidence interval for the population mean μ is $\left[\bar{x} - 2.58\left(\frac{\sigma}{\sqrt{n}}\right), \ \bar{x} \ \ 2.58\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

The 90% confidence interval for the population mean μ is $\left[\bar{x} - 1.645\left(\frac{\sigma}{\sqrt{n}}\right), \ \bar{x} + 1.645\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

In generally, a confidence interval for the mean μ of a normal population when the value of σ is known, is given by

$$\left[\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \ \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

The sample size necessary for the confidence interval to have a width W is $n = \left(2 z_{\alpha/2} \frac{\sigma}{W}\right)^2$

If n is sufficiently large, the confidence interval for μ is

$$\left[\bar{x} - Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right), \bar{x} + Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right]$$

Problems

1. Consider a study of preferred height for an experiential keyboard with large forearm wrist support. A sample of $n=31$ trained typists was selected, and the preferred keyboard height was determined for each typist. Assuming that the preferred height is normally distributed with $\sigma = 2$cm and the sample average preferred height was $\bar{x} = 80$. Obtain a confidence interval to the average preferred height for the population mean μ.

Ans: Given $n = 31$, $\sigma = 2$ and $\bar{x} = 80$.

For 95% confidence level, the confidence interval for μ is $\left[\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

$$= \left[80 - 1.96\left(\frac{2}{\sqrt{31}}\right), 80 + 1.96\left(\frac{2}{\sqrt{31}}\right)\right]$$

$$= [79.3, 80.7]$$

Confidence interval for a population proportion.

Let p denote the proportion of successes in a population. A random sample of n individuals is to be selected and x is the number of successes in the sample. x can be regarded as a binomial random variable with $E(x) = np$ and $\sigma_x = \sqrt{npq}$ where $q = 1-p$. If $np \geq 10$ and $nq \geq 10$, x has approximately a normal distribution. Then the confidence interval for the population proportion p is

$$\left[\hat{p} - z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right),\ \hat{p} + z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)\right],$$ where $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1-\hat{p}$

Problems

1. In 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette. Let p denote the long run proportion of all such trials that would result in ignition. Find the confidence interval for p with a confidence level 95%.

Ans: Given $n = 48$, $x = 16$

$$\hat{p} = \frac{x}{n} = \frac{16}{48} = 0.333$$

$$\therefore \hat{q} = 1 - \hat{p} = 1 - 0.333 = 0.667$$

Since confidence level is 95%, $z_{\alpha/2} = 1.96$

$\therefore$ The confidence interval is

$$\left[\hat{p} - z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right),\ \hat{p} + z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)\right]$$

$$= \left[0.333 - 1.96\left(\sqrt{\frac{0.333 \times 0.667}{48}}\right), 0.333 + 1.96\left(\sqrt{\frac{0.333 \times 0.667}{48}}\right)\right]$$

$= [0.333 - 0.133, 0.333 + 0.133]$

$= [0.2, 0.466]$

2. In a study of a particular wafer inspection process, 356 dies were examined by an inspection probe and 201 of these passed the probe. Assuming a stable process, calculate a 95% confidence interval for the proportion of all dies that pass the probe.

Ans: Given $n = 356$, $x = 201$

$$\hat{p} = \frac{x}{n} = \frac{201}{356} = 0.565$$

$$\therefore \hat{q} = 1 - \hat{p} = 0.435$$

Since confidence level is 95%, $z_{\alpha/2} = 1.96$

$\therefore$ The confidence interval is

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

$$= \left[0.565 - 1.96\sqrt{\frac{0.565 \times 0.435}{356}}, 0.565 + 1.96\sqrt{\frac{0.565 \times 0.435}{356}}\right]$$

$$= [0.514, 0.616]$$

Scanned with CamScanner

Test of Hypothesis

A statistical hypothesis or hypothesis is a statement about a population parameter. In any hypothesis testing problem, there are two contradictory hypotheses under consideration. The null hypothesis denoted by H_0, is the claim that is initially assumed to be true. The alternate hypothesis denoted by H_a (or H_1) is the statement that is contradictory to H_0. The null hypothesis will be rejected in favour of the alternate hypothesis only if sample evidence suggests that H_0 is false. Otherwise, the null hypothesis H_0 is accepted. If the null hypothesis is rejected, then the alternate hypothesis is accepted.

Large sample test for population mean:

A test of hypothesis is a method for using sample data to decide whether the null hypothesis should be rejected or accepted.

The null hypothesis statement is $H_0 : \mu = \mu_0$

The possible alternate hypothesis are

$$H_a : \mu \neq \mu_0 \quad \text{(two tailed test)}$$

$$\left.\begin{array}{l} H_a : \mu > \mu_0 \\ H_a : \mu < \mu_0 \end{array}\right\} \text{(one tailed test)}$$

A test statistic is computed after stating the null hypothesis. The test statistics is used to test whether the null hypothesis H_0 should be accepted or rejected

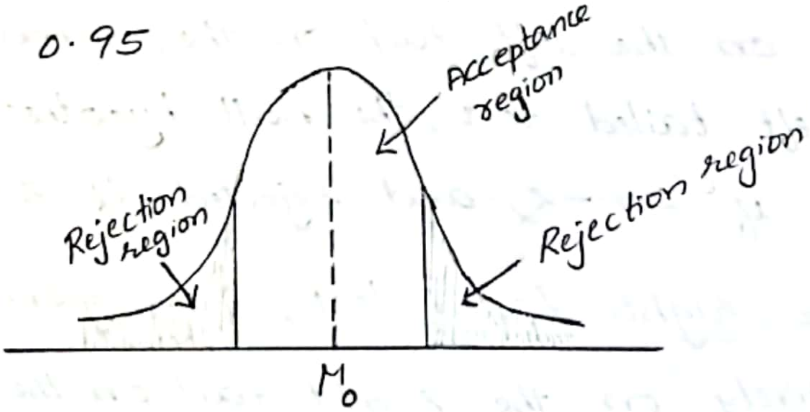
The <u>level of significance</u> is the maximum probability of rejecting a null hypothesis when it is true and is denoted by α. The level of significance may be taken as 1% or 5% or 10%.
If the level of significance is 5%, it means that there is 95% confidence of making a correct decision. When no level of significance is mentioned, it is taken as $\alpha = 0.05$

Two tailed test

Two tailed test is $H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$, where μ is the parameter to be tested and μ_0 is the hypothesised value.

In a two tailed test if the significance level is 5%, the probability of the rejection region is 0.05. It is equally split on both sides of the curve as 0.025. The region of acceptance in this case is 0.95



If the population standard deviation σ is known and n denotes the size of the sample, the test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

If σ is not known, $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$, where s is the standard deviation of the sample.

In two tailed test, Null hypothesis H_0 is accepted if

$$|z| = \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| \leq \text{table value of } z_{\alpha/2}$$

and rejected if $|z| = \left| \frac{x - \mu}{\sigma/\sqrt{n}} \right| > \text{table value of } z_{\alpha/2}$

One tailed test

One tailed test is either

$H_0 : \mu = \mu_0$ or $H_0 : \mu = \mu_0$

$H_a : \mu < \mu_0$ $H_a : \mu > \mu_0$

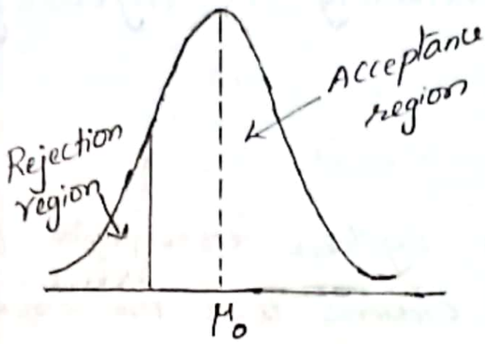
where μ is the parameter to be tested and μ_0 is the hypothesised value.

In the left tailed test, the region of rejection lies entirely on the left tail on the normal curve.

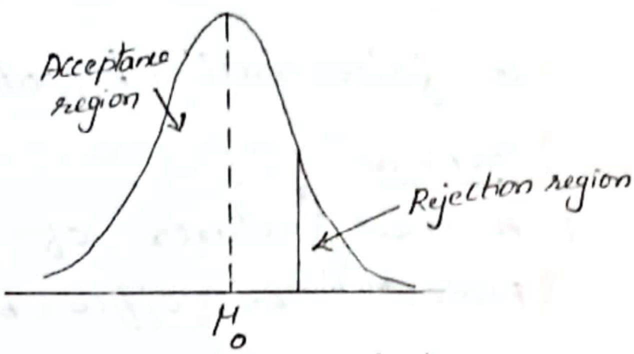
$\therefore$ In left tailed test, the null hypothesis H_0 is accepted if $z > -z_\alpha$ and rejected if $z \leq -z_\alpha$.

In the right tailed test, the region of rejection lies entirely on the right tail on the normal curve

$\therefore$ In right tailed test, the null hypothesis H_0 is accepted if $z < z_\alpha$ and rejected if $z \geq z_\alpha$.



Left tailed test



Right tailed test

Procedure for Hypothesis testing.

1. Stating of hypothesis : construct the null hypothesis H_0 and alternate hypothesis H_a.

2. Identification of test statistic :

   The test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ or $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

3. specify the level of significance
4. Determine the value of the test statistic
5. check whether the null hypothesis is accepted or rejected.

The critical values of z are

Critical value (z_α)	Level of significance		
	1%	5%	10%
Two tailed test	$z_{\alpha/2} = 2.58$	$z_{\alpha/2} = 1.96$	$z_{\alpha/2} = 1.645$
one tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$

Type I and Type II error

Rejecting a null hypothesis when it is true is called type I error. Accepting a null hypothesis when it is false is called a type II error.

<u>Power of a test</u> is the probability of rejecting a false null hypothesis.

Problems

1. A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system - activation temperature is 130°. A sample of $n=9$ systems, when tested, yields a sample average activation temperature of 131.08°F. If the distribution of activation times is normal with standard deviation 1.5°F, does the data contradicts the manufacturers claim at significance level $\alpha = 0.01$?

Ans: Given, population mean $\mu = 130$
population standard deviation, $\sigma = 1.5$
sample size $n = 9$
Sample mean $\bar{x} = 131.08$

Null Hypothesis $H_0 : \mu = 130$
Alternate hypothesis $H_1 : \mu \neq 130$

Test statistic $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{131.08 - 130}{1.5/\sqrt{9}}$

$= \dfrac{1.08}{0.5} = 2.16$

For a two tailed test, at significance level $\alpha = 0.01$,
$Z_{\alpha/2} = 2.58$

$\therefore |Z| = 2.16 < Z_{\alpha/2}$

Since $|Z| < Z_{\alpha/2}$, $\therefore H_0$ is accepted

$\therefore$ The data does not give strong support to the claim that the true average differs from the design value of 130.

Large sample test for population proportion.

Let p denote the proportion of individuals or objects in a population who posses a specified property. If an individual or object with the property is labelled a success the p is the population proportion of successes. Tests concerning p will be based on a random sample of size n from the population. The number of successes in the sample x follows a binomial distribution. The estimator $\hat{P} = \frac{x}{n}$ has approximately a normal distribution and its standard deviation is $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$.

If $E(\hat{P}) = P_0$, then $\sigma_{\hat{p}} = \sqrt{\frac{P_0(1-P_0)}{n}}$

$$\boxed{\text{The test statistic } Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}}$$

has approximately a standard normal distribution.

A test on statistical hypothesis, two tailed test is

Null hypothesis, $H_0 : P = P_0$

Alternate hypothesis, $H_a : P \neq P_0$

H_0 is accepted if $|z| \leq z_{\alpha/2}$ and

H_0 is rejected if $|z| > z_{\alpha/2}$.

In a left tailed test, Null hypothesis $H_0 : P = P_0$

Alternate hypothesis $H_a : P < P_0$

H_0 is accepted if $z > -z_\alpha$ and rejected if $z \leq -z_\alpha$.

In a right tailed test, Null hypothesis $H_0 : P = P_0$

Alternate hypothesis $H_a : P > P_0$

H_0 is accepted if $z < z_\alpha$ and rejected if $z \geq z_\alpha$.

1. A random sample of 150 recent donations at a certain blood bank reveals that 82 were type A blood. Does this suggest that the actual percentage of type A donations differ from 40%, the percentage of the population having type A blood? carry out a test of the appropriate hypothesis using a significance level of 0.01. would your conclusion have been different if a significance level of 0.05 has been used?

Ans: Given $n = 150$, $x = 82$, $P_0 = 40\% = 0.4$

$$\hat{P} = \frac{x}{n} = \frac{82}{150} = 0.5467$$

Null hypothesis $H_0 : P = 0.4$

Alternate hypothesis $H_a : P \neq 0.4$

Test statistic, $z = \dfrac{\hat{P} - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}} = \dfrac{0.5467 - 0.4}{\sqrt{\dfrac{0.4(1-0.4)}{150}}}$

$$= \frac{0.1467}{0.04} = 3.667$$

At $\alpha = 0.01$, $z_{\alpha/2} = 2.58$

$\therefore |z| > z_{\alpha/2}$. So H_0 is rejected

At $\alpha = 0.05$, $z_{\alpha/2} = 1.96$

$\therefore |z| > z_{\alpha/2}$. So H_0 is rejected

2. If 20 people were attacked by a disease and only 18 survived, will you reject the hypothesis that the survival rate if attacked by this disease is 85%.

in favour of the hypothesis that is more at 5% level?

Ans: Given $n = 20$, $x = 18$, $P_0 = 0.85$

$$\hat{p} = \frac{x}{n} = \frac{18}{20} = 0.9$$

Null hypothesis $H_0 : P = 0.85$

Alternate hypothesis $H_a : P > 0.85$

Test statistic, $$z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.9 - 0.85}{\sqrt{\frac{0.85(1-0.85)}{20}}}$$

$$= \frac{0.05}{0.07968} = 0.627$$

At $\alpha = 0.05$, for one tailed test $z_\alpha = 1.645$

$\therefore$ $z < z_\alpha$. So H_0 is accepted

Hence the proportion of the survival is 0.85.

Large sample test for difference between two population means (Equality of means)

Let $\bar{x}_1$ be the mean of an independent random sample of size n_1 from a population with mean μ_1 and variance σ_1^2. Let $\bar{x}_2$ be the mean of an independent random sample of size n_2 from a population with mean μ_2 and variance σ_2^2. Then the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(*) If $\sigma_1 = \sigma_2 = \sigma$, then the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(*) If σ_1 and σ_2 are not known, then the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$, where s_1^2 and s_2^2 are sample variance.

(*) If $\sigma_1 = \sigma_2$ and σ is not known, the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Note

1. Here for two tailed test, null hypothesis $H_0: \mu_1 = \mu_2$

and alternate hypothesis $H_a: \mu_1 \neq \mu_2$

H_0 is accepted if $|z| \leq z_{\alpha/2}$ and rejected if $|z| > z_{\alpha/2}$

2. In left tailed test, null hypothesis $H_0: \mu_1 = \mu_2$ and alternate hypothesis $H_a: \mu_1 < \mu_2$

H_0 is accepted if $z > -z_\alpha$ and rejected if $z \le -z_\alpha$

3. In right tailed test, null hypothesis $H_0: \mu_1 = \mu_2$ and alternate hypothesis $H_a: \mu_1 > \mu_2$

H_0 is accepted if $z < z_\alpha$ and rejected if $z \ge z_\alpha$

Problems

1. Analysis of a random sample consisting of $m = 20$ specimens of cold rolled steel to determine yield strengths resulted in a sample average strength of $\overline{x}_1 = 29.8$. A second random sample of $n = 25$ two sided galvanized steel specimens gave a sample average strength of $\overline{x}_2 = 34.7$. Assume that the two yield strength distributions are normal with $\sigma_1 = 4$ and $\sigma_2 = 5$. Does the data indicate that the corresponding true average yield strengths μ_1 and μ_2 are different? Test at significance level $\alpha = 0.01$

Ans: Given $n_1 = 20$, $n_2 = 25$, $\overline{x}_1 = 29.8$, $\overline{x}_2 = 34.7$

$\sigma_1 = 4$, $\sigma_2 = 5$ $\alpha = 0.01$

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternate hypothesis $H_a: \mu_1 \neq \mu_2$

Test statistic $z = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \dfrac{29.8 - 34.7}{\sqrt{\frac{16}{20} + \frac{25}{25}}} = \dfrac{-4.9}{\sqrt{1.8}}$

$= -3.65$

when $\alpha = 0.01$, $Z_{\alpha/2} = 2.58$

$\therefore |z| > Z_{\alpha/2}$

So H_0 is rejected

2. Samples of students were drawn from two universities and from their weights in kg, the SD are calculated. Make a large sample test for the significance of the difference between the means.

	Mean	SD	Size of the sample.
university A	55	10	400
university B	57	15	100

Ans: Given $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $S_1 = 10$, $S_2 = 15$, $n_1 = 400$, $n_2 = 100$

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_a : \mu_1 \neq \mu_2$

Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{55 - 57}{\sqrt{\dfrac{100}{400} + \dfrac{225}{100}}}$

$= -1.26$

At $\alpha = 0.05$, $Z_{\alpha/2} = 1.96$

$\therefore |z| < Z_{\alpha/2}$

So H_0 is accepted.

Large sample test for difference between two population proportions (Equality of proportions)

Consider a sample of size n_1, selected from one population and independently a sample of size n_2 selected from other population. Let x_1 denote the number of successes in the first sample and x_2 denote the number of successes in the second sample. Independence of the two samples implies that x_1 and x_2 are independent. Let P_1 be the proportion of success in first population and P_2 be the proportion of success in second population. If x_1 follows binomial distribution with parameters n_1 and P_1, and x_2 follows binomial distribution with parameters n_2 and P_2, then $\hat{P}_1 = \frac{x_1}{n_1}$ and $\hat{P}_2 = \frac{x_2}{n_2}$ follows approximate normal distribution.

(*) The test statistic $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where

$$\hat{p} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1+n_2} \text{ and } \hat{q} = 1-\hat{p}.$$

(*) If p denote the common value of P_1 and P_2, the test statistic is $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $q = 1-p$.

Note

1) For two tailed test null hypothesis $H_0 : P_1 = P_2$

and alternate hypothesis $H_a : P_1 \neq P_2$

Null hypothesis is accepted if $|z| \leq z_{\alpha/2}$ and rejected if $|z| > z_{\alpha/2}$.

(2) For left tailed test null hypothesis $H_0 : P_1 = P_2$

and alternate hypothesis $H_1 : P_1 < P_2$

H_0 is accepted if $z > -z_\alpha$ and rejected if $z \leq -z_\alpha$

(3) For right tailed test null hypothesis $H_0 : P_1 = P_2$

and alternate hypothesis $H_a : P_1 > P_2$

H_0 is accepted if $z < z_\alpha$ and rejected if $z \geq z_\alpha$

Problems

1. Out of 549 study, participants who regularly used aspirin after being diagnosed with colorectal cancer, there were 81 colorectal cancer specific deaths, whereas among 730 similarly diagnosed individuals who did not subsequently use aspirin, there were 141 colorectal cancer specific deaths. Does this data suggest that the regular use of aspirin after diagnosis will decrease the incidence rate of colorectal cancer specific deaths? Let's test the approximate hypothesis using a significance level of 0.05.

Ans: Given $n_1 = 549$, $x_1 = 81$, $\hat{P}_1 = \frac{x_1}{n_1} = \frac{81}{549} = 0.1475$

$n_2 = 730$, $x_2 = 141$, $\hat{P}_2 = \frac{x_2}{n_2} = \frac{141}{730} = 0.1932$

$\alpha = 0.05$

$$\hat{P} = \frac{n_1 \hat{P_1} + n_2 \hat{P_2}}{n_1 + n_2} = \frac{81 + 141}{549 + 730} = 0.1736$$

$$\therefore \hat{q} = 1 - \hat{P} = 1 - 0.1736 = 0.8264$$

Null hypothesis $H_0 : P_1 = P_2$

Alternate hypothesis $H_a : P_1 < P_2$ (left tailed test)

Test statistic $z = \dfrac{\hat{P_1} - \hat{P_2}}{\sqrt{\hat{P}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{0.1475 - 0.1932}{\sqrt{(0.1736)(0.8264)\left(\frac{1}{549} + \frac{1}{730}\right)}}$$

$$= -2.136$$

At $\alpha = 0.05$, $z_\alpha = 1.645$

$\therefore z < -z_\alpha$. So H_0 is rejected

$\therefore$ Use of aspirin reduces the rate of colorectal cancer specific deaths.

Scanned with CamScanner

Small sample t-tests for single mean of normal population

If the population standard deviation is not known and the size of the sample is less than or equal to 30, we use a t-test. The t-distribution was established by fisher in year 1920. It is also known as students t-distribution.

Let $x_1, x_2, \ldots x_n$ be the members of a random sample drawn from a normal population with mean μ and variance s^2. If s^2 is not given, then $s^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$

The test statistic is $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$, which follows t-distribution with $(n-1)$ degrees of freedom.

The rejection region for the t-test differs from z-test only in that at critical value $t_{\alpha,n-1}$ replaces the z critical value z_α.

Note

1) In two tailed test, null hypothesis $H_0: \mu = \mu_0$ and alternate hypothesis $H_a: \mu \neq \mu_0$

H_0 is accepted if $|t| \leq$ table value of $t_{\alpha/2, n-1}$

2) In left tailed test, null hypothesis $H_0: \mu = \mu_0$ and alternate hypothesis $H_a: \mu < \mu_0$

H_0 is accepted if $t > -t_{\alpha,n-1}$ and rejected if $t \leq -t_{\alpha,n-1}$

3) In right tailed test, null hypothesis $H_0 : \mu = \mu_0$

and alternate hypothesis $H_a : \mu > \mu_0$

H_0 is accepted if $t < t_{\alpha, n-1}$ and rejected if $t \geq t_{\alpha, n-1}$

Problems

1. Glycerol is a major by product of ethanol fermentation in wine production and contributes to the sweetness, body and fullness of wine. Consider the following observations on glycerol concentration for samples of standard quality white wines: 2.67, 4.62, 4.14, 3.81, 3.83. Suppose the desired concentration value is 4. Does the sample data suggest that true average concentration is something other than the desired value?

Ans: Given $n = 5$, $\mu = 4$

$$\bar{x} = \frac{2.67 + 4.62 + 4.14 + 3.81 + 3.83}{5}$$

$$= 3.814$$

$$S^2 = \frac{1}{n-1} \sum_{n=1}^{5} (x_i - \bar{x})^2$$

$$= \frac{1}{4}\left[(-1.144)^2 + (0.806)^2 + (0.326)^2 + (0.004)^2 + (0.016)^2\right]$$

$$= \frac{1}{4}\left[2.06492\right] = 0.51623$$

$$S = 0.718$$

Null hypothesis $H_0 : \mu = 4$

Alternate hypothesis $H_a : \mu \neq 4$

Test statistic $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{3.814 - 4}{0.718/\sqrt{5}} = \dfrac{-0.186}{0.3211}$

$$= -0.58$$

On level of significance $\alpha = 0.05$, $\alpha/2 = 0.025$

$t_{\alpha/2, n-1} = t_{0.025, 4} = 2.776$

$\therefore |t| < t_{\alpha/2, n-1}$. So H_0 is accepted.

2. A random blood sample for test of fasting sugar for 10 boys gave the following data in mg/dl 70, 120, 110, 101, 88, 83, 95, 107, 100, 98. Does this support the assumption of population mean of 100 mg/dl? Find a reasonable range in which most of the mean fasting sugar test of the 10 boys lie.

Ans: Given $n = 10$, $\mu = 100$

$$\bar{x} = \frac{70+120+110+101+88+83+95+107+100+98}{10}$$

$$= 97.2$$

$$S^2 = \frac{1}{n-1} \sum_{n=1}^{10} (x_i - \bar{x})^2$$

$$= \frac{1}{9}\Big[(-27.2)^2 + (22.8)^2 + (12.8)^2 + (3.8)^2 + (-9.2)^2 + (-14.2)^2 + (-2.2)^2 + (9.8)^2 + (2.8)^2 + (0.8)^2\Big]$$

$$= \frac{1}{9}(1833.6) = 203.73$$

$\therefore S = 14.27$

Null hypothesis $H_0 : \mu = 100$

Alternate hypothesis $H_a : \mu \neq 100$

Test statistic $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{97.2 - 100}{14.27/\sqrt{10}} = \dfrac{-2.8}{4.5126}$

$= -0.6204$

Level of significance $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$t_{\alpha/2, n-1} = t_{0.025, 9} = 2.262$

$\therefore |t| < t_{\alpha/2, n-1}$

so H_0 is accepted

3. A machinist is expected to make engine parts with axle diameter of 1.75 cm. A random sample of 10 parts shows a mean diameter 1.85 cm with a SD of 0.1 cm. On the basis of this sample, would you say that the work of the machinist is inferior? Test at 5% significance level and 1% significance level.

Ans: Given $\mu = 1.75$ cm. $n = 10$, $\bar{x} = 1.85$ $s = 0.1$

Null hypothesis $H_0 : \mu = 1.75$

Alternate hypothesis $H_a : \mu < 1.75$ (left tailed test)

Test statistic $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{1.85 - 1.75}{0.1/\sqrt{10}} = 3.16$

when $\alpha = 0.05$

$\therefore t_{\alpha, n-1} = t_{0.05, 9} = 1.833$

$\therefore t > -t_{\alpha, n-1}$. So H_0 is accepted

when $\alpha = 0.01$

$\therefore t_{\alpha, n-1} = t_{0.01, 9} = 2.821$

$\therefore t > -t_{\alpha, n-1}$. so H_0 is accepted

Small sample t-tests for difference between two population means (Equality of means)

Let $\bar{x}_1$ be the mean of an independent random sample of size n_1 from a population with mean μ_1 and variance s_1^2. Let $\bar{x}_2$ be the mean of an independent random sample of size n_2 from a population with mean μ_2 and variance s_2^2. $\bar{x}_1$ follows a normal distribution with mean μ_1 and SD s_1^2/n_1 and $\bar{x}_2$ follows a normal distribution with mean μ_2 and SD s_2^2/n_2.

Here the null hypothesis $H_0 : \mu_1 = \mu_2$

The test statistic, $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$.

(*) If s_1 and s_2 are not given, then

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2\right]$$

The degrees of freedom is (n_1+n_2-2)

Conditions for acceptance and rejection is same as that of t-test for single mean.

Problems

1. Two salesman A and B work in a certain district, from a sampling survey conducted by the head office, the following results were obtained. State whether there is any significant difference in the average sales between the two salesman.

	A	B
No. of sales	20	18
Average sales	170	203
SD	20	25

Ans: Given $n_1 = 20, \ \bar{x}_1 = 170, \ s_1 = 20$

$n_2 = 18, \ \bar{x}_2 = 203, \ s_2 = 25$

degrees of freedom $= n_1 + n_2 - 2 = 20 + 18 - 2 = 36$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{20(20)^2 + 18(25)^2}{36} = 534.72$$

$\therefore S = 23.124$

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_a : \mu_1 \neq \mu_2$

Test statistic $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$= \frac{170 - 203}{23.124\sqrt{\frac{1}{20} + \frac{1}{18}}}$$

$$= \frac{-33}{7.5129} = -4.39$$

At $\alpha = 0.05$, $\alpha/2 = 0.025$

$\therefore t_{\alpha/2, n_1+n_2-2} = t_{0.025, 36} = 2.028$

$\therefore |t| > t_{\alpha/2, n_1+n_2-2}$. So H_0 is rejected

$\therefore$ There is significant difference in the average sales between the two salesman.