NAGARJUNA COLLEGE OF ENGINEERING AND TECHNOLOGY

(An autonomous institution under VTU)

Mudugurki Village, Venkatagirikote Post, Devanahalli Taluk, Bengaluru – 562164

NAGARJUNA
COLLEGE OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF MATHEMATICS

MATHEMATICSFORCOMPUTERSCIENCE

(COURSE CODE 22MATS31)

MODULE-3

STATISTICALINFERENCE-1

STATISTICAL INFERENCE-1

Introduction, sampling distribution, standard error, testing of hypothesis, levels of significance,test of significances, confidence limits, simple sampling of attributes, test of significance forlargesamples, comparisonoflargesamples.Sampling variables, central limit theorem and confidence limit for unknown mean. Test ofSignificance for means of two large samples

Introduction:

Population:

An aggregate of objects under study is called Population or Universe.

Example: The weights of all students in a particular college.

Sample:

A small section selected from the population is called a sample. i.e. a subset of the population.

Sampling:

The process of drawing sample is called sampling. A sample must be a random selection. Therefore sampling is also known as Random sampling.

Parameters:

The statistical constants of the population such as mean (μ), standard deviation (σ) etc. are called parameters.

Statistic:

The statistical constants for the sample drawn from population such as mean $(\bar{x})$, standarddeviation(s) etc. are called statistic.

Statistical Inference:

The fundamental object of the sampling is to get as much information as possible of the whole population by examining only a part of it i.e., sample with the minimum effort, cost, and time.

The process of pass the information from a sample to population is known as statistical inference.

Example: In a shop we assess the quality of rice or any other commodity by taking only a handful of it from the bag.

Sampling Distribution:

Consider all possible samples of size n which are drawn from a given population. If we group different means of these samples according to their frequencies, then the frequency distribution so formed is called **sampling distribution**. Similarly, we can have sampling distribution of standard deviation etc.

If we replace the previous sample while drawing each sample from population, then the sampling is called as sampling with replacement. Otherwise, sampling is called sampling without replacement.

Standard error:

The standard deviation of the sampling distribution is called standard error(S.E). Thus the S.E of the sampling distribution of means is called the standard error of means.

The S.E is used to assess the difference between the expected and observed values.

The reciprocal of the S.E is called **precision.**

If n $\geq$ 30, a sample is called **large** otherwise **small**.

Simple sampling:

A random sampling in which trials are independent and probability of success is a constant is called simple sampling.

Testing a Hypothesis:

(i)Statistical hypothesis:

We make certain assumptions about the population to reach decisions about populations on the basis of sample information which may or may not be true are called **statistical hypothesis.**

A process for deciding whether to accept or reject the hypothesis is called testing a hypothesis.

(ii) Errors:

If a hypothesis is rejected while it should have been accepted, then we say that a **Type-Ierror** has been committed.

If a hypothesis is accepted while it should have been rejected, then we say that a **Type-IIerror** has been committed.

(iii) Null Hypothesis:

The hypothesis formulated for the purpose of rejecting it, under the assumptions that is true, is called the **null hypothesis** and is denoted by H_0.

To test whether there is a relationship between two variates, we takeH_0 that there is no relationship.

Level of significance:

The probability level below which we reject the hypothesis is known as the **level of significance.**

The region in which a sample value falling is rejected is called **critical region.**

Generally, we take two critical regionswhich covers 5% and 1% areas of thenormal curve. The shaded portion in the fig.1 corresponds to 5% level ofsignificance. Thus the probability ofthe value of the variate falling in the critical region is the **level of significance.**

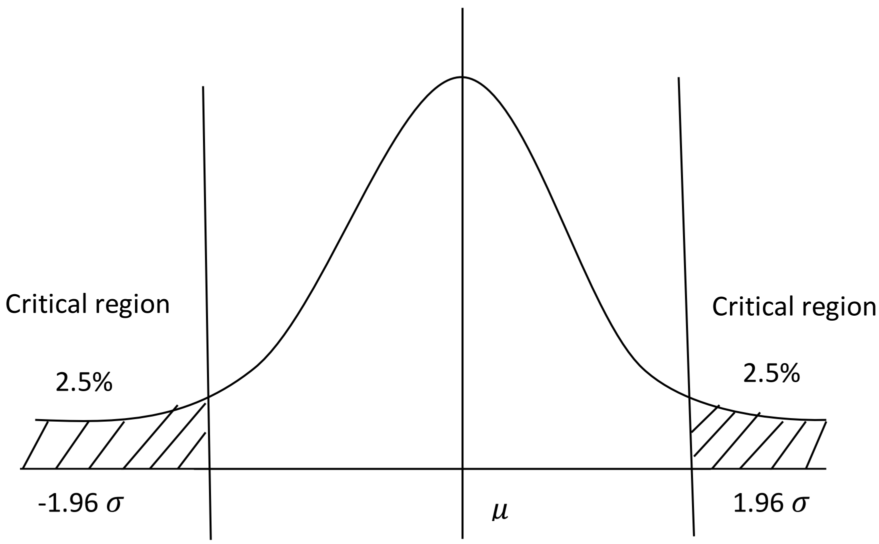

Fig.1

One – tailed test and two – tailed test(single – tail test and double – tail test):

Depending on the nature of the problem, we use these test to estimate the significance of a result.

In one – tailed test, only the area on the right of an ordinate is taken into consideration.

In two – tailed test, the area of both tails of the curve representing the sampling distribution are taken into consideration.

For Example: To test whether a coin is biased (unfair) or not, two – tailed test should be used, since a biased coin gives more number of heads than tails (corresponds to right tail) or more number of tails than heads (corresponds to left tail only).

Test of significance:

The process which enables us to decide to accept or reject the hypothesis is called the test of significance.

Confidence Limits:

Suppose the sampling distribution of a statistic S is normal with mean μ and standard deviationσ. The sample statistic S can be expected to lie in the interval $(\mu - 1.96\sigma,\ \mu + 1.96\sigma)$ for 95% times. i.e. we can be confident of finding μ in the interval $(S - 1.96\sigma,\ S + 1.96\sigma)$ in 95% cases. The interval $(S - 1.96\sigma,\ S + 1.96\sigma)$ is called 95% confidence interval for estimation of μ.

The ends of this interval i.e., $S \pm 1.96\sigma$ are called 95% confidence limits for S. Similarly, $S \pm 2.58\sigma$ are 99% confidence limits. The numbers 1.96, 2.58 etc. are called confidence coefficients.

Simple sampling of Attributes:

The sampling of attributes is the selection of samples from a population whose members possess the attribute K or not K. The presence of K is called a success and its absence a failure.

Drawing a simple sample of n items is same as a series of n independent trials with the same probability p of success. The probability of 0, 1, 2, ……., n successes are the terms of binomial expansion of$(q+p)^n$where p+q=1. Therefore the mean of success in a sample of n items is $\bar{x} = np$ and the standard error is$S = \sqrt{npq}$.

If we consider the **proportion of successes**, then

(i) Mean proportion of successes = $\frac{np}{n} = p$.

(ii) Standard error of proportion of successes = $\sqrt{n.\frac{p}{n}.\frac{q}{n}} = \sqrt{\frac{pq}{n}}$and

(iii) Precision of the proportion of successes = $\sqrt{\frac{n}{pq}}$, which varies as$\sqrt{n}$, since p and q are constants.

Test of significance for large samples:

As n is large, the binomial distribution tends to normal distribution. Therefore in sampling distribution of number of successes, the mean μ and standard deviation σ are np and $\sqrt{npq}$ respectively.

For normal distribution, only 5% of the members lies outside $\mu \pm 1.96\sigma$ while only 1% of the members liesoutside$\mu \pm 2.58\sigma$.

If x be the observed number of successes in the sample and z is the standard normal variate, then $z = \frac{x-\mu}{\sigma}$i.e. $z = \frac{x-np}{\sqrt{npq}}$.

Thus we have the following test of significance

(i)If $|z| < 1.96$, difference between the observed and expected number of successes is not significant. (Hypothesis accepted at 5% level of significance).

(ii) If $|z| > 1.96$, difference is significant at 5% level of significance. (Hypothesis rejected at 5% level of significance).

(iii) If $|z| > 2.58$, difference is significant at 1% level of significance. (Hypothesis rejected at 1% level of significance).

Problems on Testing of Hypothesis:

1. A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased(fair coin) at 5% level of significance.

Solution:

Suppose the coin is unbiased(H_0). Then probability of getting the head in a toss is $= \frac{1}{2}$.

$\therefore \; q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}, \; n = 400,$ and$x =$ Observed value of successes = 216.

$\therefore$ Mean = expected number of successes = np = 400× $\frac{1}{2}$ = 200.

$\therefore$Standard deviation of simple sampling = $\sqrt{npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{100} = 10.$

$\therefore \; z = \frac{x-np}{\sqrt{npq}} = \frac{216-200}{10} = 1.6 < 1.96.$

$\therefore$ The difference between the observed and expected number of success is not significant.

$\therefore$The hypothesis is accepted at 5% level of significance.

$\therefore$The coin is unbiased at 5% level of significance.

2. A die is tossed 960 times and it falls with 5 upwards 184 times. Is the die biased?

Solution:

Suppose the die is unbiased(H_0). Then$p = \frac{1}{6}, \; q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}, \; n = 960.$

$\therefore np = 960 \times \frac{1}{6} = 160, \quad \sqrt{npq} = \sqrt{960 \times \frac{1}{6} \times \frac{5}{6}} = 11.5, \; and \;\; x = 184.$

$$\therefore \quad z = \frac{x - np}{\sqrt{npq}} = \frac{184 - 160}{11.55} = 2.08 > 1.96.$$

$\therefore$The hypothesis is rejected at 5% level of significance.$\therefore$ The die is biased.

3. A die was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing do the data indicate an unbiased die.

Solution:

Suppose the die is unbiased (H_0). Then the probability of throwing 5 or 6 with the die is

$p = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. $\therefore q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$, $n = 9000$, $np = 9000 \times \frac{1}{3} = 3000$, $x = 3240$.

$\therefore \sqrt{npq} = \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}} = 44.72$.

$\therefore z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - 3000}{44.72} = 5.4 > 2.58$.

∴The hypothesis has to be rejected at 1% level of significance.

∴ The die is biased.

Problems on Proportion:

4. In a locality containing 18000 families, a sample of 840 families was selected at random of these 840 families, 206 families were found to have a day income Rs.250 or less. It is desired to estimate how many out of 18,000 families have a day income of Rs.250 or less. Within what limits would you place your estimate.

Solution:

Here n = 840,$p = \frac{206}{840} = \frac{103}{420} = 0.2452$. $\therefore q = 1 - p = 1 - \frac{103}{420} = 0.7548$.

∴ Mean of proportion of successes = $\frac{np}{n} = p = 0.2452$ and Standard error of the proportion of successes = $\sqrt{n.\frac{p}{n}.\frac{q}{n}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{103}{420} \times \frac{317}{420} \times \frac{1}{840}} = 0.015 = 1.5\%$.

$\therefore \mu = p = 0.2452 = 24.52\%$ and $\sigma = SE = 0.015 = 1.5\%$.

∴24.52% to be the estimate of families having day income of Rs.250 or less in the locality. The probable limits are $\mu \pm 2.58\sigma \approx p \pm 3\sigma = 0.2452 \pm 3(0.015) = 0.2452 \pm 0.045$ $= 0.2002$ and 0.2902 i.e. 20% and 29% approximately.

∴0.2002(18000) = 3603 and 0.2902(18000) = 5223 i.e. 3603 families and 5223 families approximately.[2.58≈3 and 3 is corresponding to 99.73% of confidence level]

5. A sample of 1000 days is taken from meteorological records of a certain district and 120 of them are found to be foggy what are the probable limits to the percentage of foggy days in the district?

Solution:

Here n = 1000,$p = \frac{120}{1000} = 0.12$and q = 1 – p = 1 – 0.12 = 0.88.

∴ Mean of the proportion of successes is$\mu = \frac{np}{n} = p = 0.12$ and

Standard error of the proportion of successes is $\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.12\times0.88}{1000}} = 0.0102.$

∴The probable limits of foggy days = $\mu \pm 3\sigma = p \pm 3\sigma = 0.12 \pm 3(0.0102)$

$= 0.0894\ and\ 0.1508 = 8.94\%\ and\ 15.08\%$.

6. A random sample of 500 apples was taken from a large consignment and 65 were found to be bad. Estimate the proportion of the bad apples in the consignment as well as the standard error of the estimate. Deduce that the percentage of bad apples in the consignment almost certainly lies between 8.5 and 17.5.

Solution:

Here n = 500, $p = \frac{65}{500} = 0.13$ and q = 1 – p = 1 – 0.13 = 0.87

∴ Mean of the proportion of successes is $\mu = \frac{np}{n} = p = 0.13\ i.e. 13\%$

∴ The proportion of the bad apples in a sample is 13%

Standard error of the proportion of the bad apples is $\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13\times0.87}{500}} = 0.015$

∴ The probable limit of the bad apple in the consignment = $\mu \pm 3\sigma = p \pm 3\sigma$

$= 0.13 \pm 3(0.015) = 0.13 \pm 0.045$= 0.085 and 0.175= 8.5% and 17.5%.

Comparison of large samples:

Consider two large samples of sizes $n_1,\ n_2$ are taken from two populations giving proportions of attributes A's as $p_1,\ p_2$respectively.

(i) Suppose the populations are similar as regards the attribute A, we combine two samples to estimate the common value of proportion of A's in the populations which is given by

$p = \frac{n_1p_1+n_2p_2}{n_1+n_2}$.

If e_1, e_2 be the standard errors in the two samples, then ${e_1}^2 = \frac{pq}{n_1}$ and ${e_2}^2 = \frac{pq}{n_2}$where q = 1 – p.

If e be the standard error of the difference between p_1 and p_2, then

$e^2 = {e_1}^2 + {e_2}^2 = \frac{pq}{n_1} + \frac{pq}{n_2} = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right).\quad \therefore\quad z = \frac{p_1-p_2}{e}$.

If z> 3, the difference between p_1 and p_2 is real one.

If z< 2, the difference may be due to fluctuations of simple sampling.

If zlies between 2 and 3, then the difference is significant at 5% level of significance.

(ii) If the proportions of A's are not the same in the two populations from which the samples are drawn, but p_1 and p_2 are the true values of proportions then standard error e of the difference $p_1 \sim p_2$ is given by $e^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$.

If $z = \frac{p_1 \sim p_2}{e} < 3$, the difference could have arisen due to fluctuations of simple sampling.

Problems:

7. In a city A 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?

Solution:

Here $n_1 = 900,\ n_2 = 1600,\ p_1 = \frac{20}{100} = 0.2,\ p_2 = \frac{18.5}{100} = 0.185.$

$\therefore p_1 - p_2 = \frac{20}{100} - \frac{18.5}{100} = 0.015.$ Assume that $p_1 = p_2 (H_0)$.
i.e., there is no significant difference between two sample proportions.

$\therefore\ \ p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{900(0.2)+1600(0.185)}{900+1600} = 0.19$ and q = 1 – p = 1 – 0.19 = 0.81.

$\therefore\ \ e^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 0.19(0.81)\left(\frac{1}{900} + \frac{1}{1600}\right) = 0.000267.\ \therefore e = 0.016.$

$\therefore z = \frac{p_1 - p_2}{e} = \frac{0.015}{0.016} = 0.93 < 1.$

As z <1, the difference between the proportions is not significant.

8. In two large populations there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Solution:

Here$n_1 = 1200,\ n_2 = 900, p_1 = \frac{30}{100} = 0.3,\ p_2 = \frac{25}{100} = 0.25.$

$\therefore\ \ p_1 - p_2 = 0.3 - 0.25 = 0.05.$ Assume that $p_1 \neq p_2$.

$\therefore\ \ q_1 = 1 - p_1 = 1 - 0.3 = 0.7, q_2 = 1 - p_2 = 1 - 0.25 = 0.75.$

$\therefore\ \ e^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.3(0.7)}{1200} + \frac{0.25(0.75)}{900} = 0.0003833.\ \ \therefore\ \ e = 0.0195.$

$\therefore\ \ z = \frac{p_1 - p_2}{e} = \frac{0.05}{0.0195} = 2.5 < 3. \therefore\ \ z < 3.$

∴The difference could have arisen due to fluctuations of simple sampling.
Hence it is unlikely that the real difference will be hidden.

9. A machine produces 16 imperfect articles in a sample of 500. After machine overhauled, it produces 3 imperfect article in a batch of 100. Has the machine been improved?

Solution:

Here $n_1 = 500,\ n_2 = 100,\ p_1 = \frac{16}{500} = 0.032,\ p_2 = \frac{3}{100} = 0.03.$

$\therefore\ p_1 - p_2 = 0.032 - 0.03 = 0.002.$ Assume that $p_1 = p_2$.

$\therefore\ p = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{500(0.032)+100(0.03)}{500+100} = 0.032.\ \ \therefore\ q = 1 - p = 1 - 0.032 = 0.968.$

$\therefore\ e^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 0.032(0.968)\left(\frac{1}{500} + \frac{1}{100}\right) = 0.0003717.\ \ \therefore\ e = 0.019 \cong 0.02.$

$\therefore\ z = \frac{p_1-p_2}{e} = \frac{0.002}{0.02} = 0.1.\ \ \therefore\ z < 1$

∴The difference between the proportions is not significant.
Hence the machine has not improved due to overhauling.

10. One type of air craft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significant difference in the two types of aircrafts so far as engine defects are concerned?

Solution:

Here$n_1 = 100,\ n_2 = 200,\ p_1 = \frac{5}{100} = 0.05,\ p_2 = \frac{7}{200} = 0.035.$

$\therefore\ p_1 - p_2 = 0.05 - 0.035 = 0.015.$ Assume that $p_1 = p_2$.

$\therefore\ p = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{100(0.05)+200(0.035)}{100+200} = 0.04.\ \ \therefore\ q = 1 - p = 1 - 0.04 = 0.96.$

$\therefore\ e^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 0.04(0.96)\left(\frac{1}{100} + \frac{1}{200}\right) = 0.000576.\ \ \therefore\ e = 0.024.$

$\therefore\ z = \frac{p_1-p_2}{e} = \frac{0.015}{0.024} = 0.62.\ \ \therefore z < 1.$

Hence the difference between the two types of aircrafts is not significant.

Sampling of Variables:

Consider the sampling of variables such as weight, height, etc. Each member of the population gives a value of the variable, and the population is a frequency distribution of the variable. Thus a random sample of size n from the population is same as selecting n values of the variables from those of the distribution.

Sampling distribution of the mean:

If a population is distributed normally with mean$\mu$and standard deviation σ, then the means of all positive random samples of size n, are also distributed normally with meanμ and standard error$\frac{\sigma}{\sqrt{n}}$.

Central limit Theorem:

If the variable X has a non-normal distribution with mean μ and standard deviation σ, then the limiting distribution of$z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ $as\ n \to \infty$, is the standard normal distribution (i.e. with mean 0 and Standard deviation = 1).

This theorem holds good for a sample of 25 or more which is regarded as large.

Thus if the population is normal, the sampling distribution of the mean is also normal with mean μ and standard error$\frac{\sigma}{\sqrt{n}}$, while for large samples the same result holds even if the distribution of the population is non-normal.

Confidence limits for unknown mean:

If the mean μ of a random sample of size n drawn from a population is not known, then there will be a large values of μ for which observed mean $\bar{x}$ of the sample is not significant at any assigned level of probability.

If $\bar{x}$is not significant at 5% level of probability, then 95% confidence limits for the mean of the population corresponding to given sample are$\bar{x} \pm 1.96\sigma/\sqrt{n}$.

Similarly 99% confidence limits for μ are$\bar{x} \pm 2.58\sigma/\sqrt{n}$.

Problems:

1. A sample of 900 members is found to have a mean of 3.4cm.Can it be reasonably regarded as a truly random sample from a large population with mean 3.25cm and standard deviation 1.61cm.

Solution:

Here n=900,$\bar{x}$ = 3.4cm,μ = 3.25cm, σ= 1.61cm.
Suppose the sample can be regarded as a truly random sample.

$$\therefore \quad z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{3.4-3.25}{\left(\frac{1.61}{\sqrt{900}}\right)} = 2.8. \therefore \quad z > 1.96.$$

∴The deviation of the sample mean from the mean of population is significant at 5% level of significance. Hence it cannot be regarded as a random sample.

2. A sample of 400 items is taken from a normal population whose mean is 4 and variance is 4. If the sample mean is 4.45,can the samples be regarded as a simple sample?

Solution:

Here n=400, $\bar{x}$ = 4.45, μ = 4,$\sigma^2 = 4$ $\therefore \sigma = 2$.

Suppose the sample can be regarded as a simple sample.

$\therefore z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{4.45-4}{\left(\frac{2}{\sqrt{400}}\right)} = 4.5 > 1.96.$

∴The deviation of the sample mean from the mean of population is significant at 5% level of significance. Hence it cannot be regarded as a simple sample.

3. A normal population has a mean 0.1 and a standard deviation of 2.1. Find the probability that the mean of simple sample of 900 members will be a negative?
[Use $P(1.428)$= 0.4236]

Solution:

Here n = 900, μ = 0.1,σ = 2.1, $\bar{x}$ is negative.

$\therefore \quad z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{\bar{x}-0.1}{\left(\frac{2.1}{\sqrt{900}}\right)} = \frac{\bar{x}-0.1}{0.07} = \frac{\bar{x}}{0.07} - \frac{0.1}{0.07} = \frac{\bar{x}}{0.07} - 1.428.$

$\therefore \quad z < -1.428 \because \bar{x}$ is negative.

$\therefore \quad p(z < -1.428) = p(z > 1.428) = 0.5 - P(1.428) = 0.5 - 0.4236 = 0.0764.$

4. The mean of a certain normal population is equal to the standard error of the mean of the samples of 100 from that distribution. Find the probability that the mean of the sample of 25 from the distribution will be negative?[Use $P(0.5)$= 0.1915]

Solution:

Letμbe the mean and σbe the standard deviation of the distribution.

Given n = 100,μ = standard error of the sample mean = $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$.

Also for sample size n = 25, we have$z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{\bar{x}-\frac{\sigma}{10}}{\left(\frac{\sigma}{\sqrt{25}}\right)} = \frac{\bar{x}-\frac{\sigma}{10}}{\left(\frac{\sigma}{5}\right)} = \frac{5\bar{x}}{\sigma} - \frac{1}{2}$.

$\therefore \quad z < -\frac{1}{2} \therefore \because$ given $\bar{x}$ is negative.

$\therefore\ p\left(z < -\frac{1}{2}\right) = p\left(z > \frac{1}{2}\right) = 0.5 - P\left(z = \frac{1}{2}\right) = 0.5 - 0.1915 = 0.3085.$

Test of Significance for Means of two Large samples:

(i) Suppose two random samples of sizes n_1 and n_2 have been drawn from the same population with standard deviation σ. We wish to test the difference between the sample means $\bar{x}_1$ and $\bar{x}_2$ is significant or is merely due to fluctuations of sampling.

If the samples are independent, then the standard error e of the difference of their means is given by $e^2 = {e_1}^2 + {e_2}^2$, where $e_1 = \frac{\sigma}{\sqrt{n_1}}$, $e_2 = \frac{\sigma}{\sqrt{n_2}}$ are the standard errors of the means of the two samples. $\therefore\ e = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

Hence $z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ is normally distributed with mean zero and standard deviation is 1.

Test of Significance (n_1, n_2 being large):

If z > 1.96, then the difference is significant at 5% level of significance.

If z > 3, then the samples has not been drawn from the same population or the sampling is not simple.

(ii) If the samples are drawn from different populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, then the standard error of their mean is given by $e = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$.

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

Assuming that the two populations have the same mean (i.e., $\mu_1 = \mu_2$), the difference of the means 0 and standard deviation is e. The same procedure of test of significance is applied.

Problems:

1. The means of simple samples of sizes 1000 and 2000 are 67.5cm and 68cm, respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5cm.

Solution:

Here $n_1 = 1000,\ n_2 = 2000, \bar{x}_1 = 67.5,\ \bar{x}_2 = 68$

On the hypothesis, that the samples are drawn from the same population of standard deviation $\sigma = 2.5$, we get

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{67.5-68}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = -5.14 \therefore |z| = 5.14 > 3$$

∴The samples cannot be regarded as drawn from the same population.

2. A sample of 100 bulbs produced by company A showed a mean life of 1190 hours and a standard deviation of 90 hours. Also a sample of 75 bulbs produced by a company B showed a mean life of 1230 hours and a standard deviation of 120 hours. Is there a difference between the mean life times of the bulbs produced by the two companies at (i) 5% level of significance (ii) 1%level of significance.

Solution:

Here $n_1 = 100,\ \sigma_1 = 90,\ \bar{x}_1 = 1190,\ n_2 = 75,\ \sigma_2 = 120,\ \bar{x}_2 = 1230.$

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{1190 - 1230}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}} = -2.42. \quad \therefore |z| = 2.42$$

$\therefore\ \ |z| > 1.96\ \ and\ \ |z| < 2.58$

∴(i) There is a difference between the mean time of the bulbs at 5% level of significance and

(ii) not at 1% level of significance.

3. A sample height of 6400 soldiers has a mean of 67.85 inches and a standard deviation of 2.56 inches while a simple sample of heights of 1600 sailors has a mean of 68.55 inches and a standard deviation of 2.52 inches. Do the data indicate the sailors are on the average taller than soldiers?

Solution:

Here$n_1 = 6400,\ \sigma_1 = 2.56,\ \bar{x}_1 = 67.85,\ n_2 = 1600,\ \sigma_2 = 2.52,\ \bar{x}_2 = 68.55.$

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}} = -9.9.$$

$\therefore |z| = 9.9 > 3$. This is highly significant.

Hence the data indicates that the sailors are on the average taller than soldiers

4. The means of two large samples of 1000 and 2000 members are 168.75cms and 170cms, respectively. Can the samples be regarded as drawn from the same population of standard deviation 6.25cms.

Solution:

Here $n_1 = 1000,\ n_2 = 2000,\ \bar{x}_1 = 168.75,\ \bar{x}_2 = 170,\ \sigma = 6.25.$

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{168.75 - 170}{6.25\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.043.$$

$\therefore\ \ |z| = 5.04 > 3.$ ∴Samples cannot be regarded as drawn from the same sample.