(6) Weibull Distribution is given by

$$f(x) = \frac{\alpha}{C} x^{\alpha-1} e^{-\frac{x^\alpha}{C}}, \quad x > 0,\ c > 0$$

where C is a scale parameter and α is a shape parameter.

This distribution is used for

(1) variation in the fatigue resistance of steel and its elastic limits.

(2) variation of length of service of radio service equipment.

Exercise 11.7

Find the mean and variance for the following distributions

1. Rectangular distribution **Ans.** $\frac{1}{2}, \frac{1}{12}$

2. Uniform distribution $f(x) = \frac{1}{n}$, $x = 1, 2, \ldots n$ **Ans.** $\frac{1}{2}(n+1), \frac{1}{12}(n^2-1)$

3. Geometric distribution $p(r) = 2^r$, $r = 1, 2, 3 \ldots$

4. Exponential distribution $p(x) = \lambda e^{-\lambda x}$ **Ans.** $\frac{1}{\lambda}, \frac{1}{\lambda}$

SAMPLING OF VARIABLES

11.30 POPULATION (Universe)

Before giving the notion of sampling, we will first define *population*. The group of individuals under study is called *population* or *universe*. It may be finite or infinite.

11.31 SAMPLING

A part selected from the population is called *a sample*. The process of selection of a sample is called sampling. A *Random sample* is one in which each member of population has an equal chance of being included in it. There are ${}^NC_n$ different samples of size n that can be picked up from a population of size N.

11.32 PARAMETERS AND STATISTICS

The statistical constants of the population such as mean (μ), standard deviation (σ) are called parameters. Parameters are denoted by Greek letters.

The mean ($\bar{x}$), standard deviation $|S|$ of a sample are known as statistics. Statistics are denoted by Roman letters.

Symbols for Population and Samples

Characteristic	Population	Sample
	Parameter	Statistic
Symbols	population size = N	sample size = n
	population mean = μ	sample mean = $\bar{x}$
	population standard deviation = σ	sample standard deviation = s
	population proportion = p	sample proportion = $\bar{p}$

11.33 AIMS OF A SAMPLE

The population parameters are not known generally. Then the sample characteristics are utilised to approximately determine or estimate of the population. Thus, static is an estimate of the parameter. To what extent can we depend on the sample estimates?

The estimate of mean and standard deviation of the population is a primary purpose of all scientific experimentation. The logic of the sampling theory is the logic of *induction*. In induction, we pass from a particular (sample) to general (population). This type of generalization here is known as *statistical inference*. The conclusion in the sampling studies are based not on certainties but on probabilities.

11.34 TYPES OF SAMPLING

Following types of sampling are common:

(1) Purposive sampling (2) Random sampling

(3) Stratified sampling (4) Systematic sampling

11.35 SAMPLING DISTRIBUTION

From a population a number of samples are drawn of equal size n. Find out the mean of each sample. The means of samples are not equal. The means with their respective frequencies are grouped. The frequency distribution so formed is known as *sampling distribution of the mean*. Similarly, sampling distribution of standard deviation we can have.

11.36. STANDARD ERROR (S.E.) is the standard deviation of the sampling distribution. For assessing the difference between the expected value and observed value, standard error is used. Reciprocal of standard error is known as *precision*.

11.37 SAMPLING DISTRIBUTION OF MEANS FROM INFINITE POPULATION

Let the population be infinitely large and having a population mean of μ and a population variance of σ^2. If x is a random variable denoting the measurement of the characteristic, then

Expected value of x, $E(x) = \mu$

Variance of x, $\text{Var}(x) = \sigma^2$

The sample mean $\bar{x}$ is the sum of n random variables, *viz.*, $x_1, x_2, \ldots, x_n$, each being divided by n. Here, $x_1, x_2, \ldots, x_n$ are independent random variables from the infinitely large population.

$$\therefore \quad E(x_1) = \mu \quad \text{and} \quad Var(x_1) = \sigma^2$$

$$E(x_2) = \mu \quad \text{and} \quad Var(x_2) = \sigma^2 \text{ and so on}$$

Finally

$$E(\bar{x}) = E\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= \frac{1}{n} E x_1 + \frac{1}{n} E(x_2) + \ldots + \frac{1}{n} E(x_n)$$

$$= \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu$$

$$= \mu$$

and

$$Var(\bar{x}) = Var\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= Var\left(\frac{x_1}{n}\right) + Var\left(\frac{x_2}{n}\right) + \ldots + Var\left(\frac{x_n}{n}\right)$$

$$= \frac{1}{n^2} Var(x_1) + \frac{1}{n^2} Var(x_2) + \ldots + \frac{1}{n^2} Var(x_n)$$

$$= \frac{1}{n^2}\cdot\sigma^2 + \frac{1}{n^2}\cdot\sigma^2 + \ldots + \frac{1}{n^2}\cdot\sigma^2$$

$$= \frac{n\,\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

The expected value of the sample mean is the same as population mean. The variance of the sample mean is the variance of the population divided by the sample size.

The average value of the sample tends to true population mean. If sample size (n) is increased then variance of $\bar{x}$, $\left(\frac{\sigma^2}{n}\right)$ gets reduced, by taking large value of n, the variance $\left(\frac{\sigma^2}{n}\right)$ of $\bar{x}$ can be made as small as desired. The standard deviation $\left(\frac{\sigma}{\sqrt{n}}\right)$ of $\bar{x}$ is also called standard error of the mean. It is denoted by $\sigma_{\bar{x}}$.

Sampling with Replacement

When the sampling is done with replacement, so that the population is back to the same form before the next sample member is picked up. We have

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}. \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Sampling without replacement from Finite population

When a sample is picked up without replacement from a finite population, the probability distribution of second random variable depends on the outcome of the first pick up. n sample members do not remain independent. Now we have

$$E(\bar{x}) = \mu$$

and
$$Var(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\cdot\frac{N-n}{N-1} \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$= \frac{\sigma}{\sqrt{n}} \text{ app.} \qquad \left(\text{if } \frac{n}{N} \text{ is very small}\right)$$

Sampling from Normal Population

If $x \sim N(\mu, \sigma^2)$ then it follows that $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

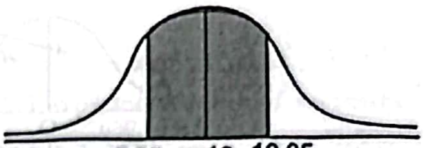
Example 71. *The diameter of a component produced on a semi-automatic machine is known to be distributed normally with a mean of 10 mm and a standard deviation of 0.1 mm. If we pick up a random sample of size 5, what is the probability that the same mean will be between 9.95 and 10.05 mm?*

Solution. Let x be a random variable representing the diameter of one component picked up at random.

Here $x \sim N(10, 0.01)$, Therefore, $\bar{x} \sim N\left(10, \frac{0.01}{5}\right)$ $\qquad \left[\bar{x} = N\left(\bar{x}, \frac{\sigma^2}{n}\right)\right]$

$$Pr\{9.95 \le \bar{x} \le 10.05\} = 2 \times Pr\{10 \le \bar{x} \le 10.05\} \qquad \left\{z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right\}$$

$$= 2 \times Pr\left\{\frac{10-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{10.05-\mu}{\frac{\sigma}{\sqrt{n}}}\right\}$$



$$= 2 \times Pr\left\{0 \le z \le \frac{10.05-10}{\frac{0.1}{\sqrt{5}}}\right\}$$

$$= 2 \times Pr\{0 \le z \le 1.12\}$$

$$= 2 \times 0.3686$$

$$= 0.7372 \qquad \textbf{Ans.}$$

Similar Question

A sample of size 25 is picked up at random from a population which is normally distributed with a mean 100 and a variance of 36. Calculate (*a*) $Pr\{\bar{x} \le 99\}$, (*b*) $Pr\{98 \le \bar{x} \le 100\}$

Ans. (*a*) 0.2023 (*b*) 0.4522

11.38 SAMPLING DISTRIBUTION OF THE VARIANCE

We use a sample statistic called the sample variance to estimate the population variance. The sample variance is usually denoted by s^2

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

11.39 TESTING A HYPOTHESIS

On the basis of sample information, we make certain decisions about the population. In taking such decisions we make certain assumptions. These assumptions are known as *statistical hypothesis*. There hypothesis are tested. Assuming the hypothesis correct we calculate the probability of getting the observed sample. If this probability is less than a certain assigned value, the hypothesis is to be rejected.

11.40 NULL HYPOTHESIS (H_0)

Null hypothesis is based for analysing the problem. Null hypothesis is the *hypothesis of no difference*. Thus, we shall persume that there is no significant difference between the observed value and expected value. Then, we shall test whether this hypothesis is satisfied by the data or not. If the hypothesis is not approved the difference is considered to be significant. If hypothesis is approved then the difference would be described as due to sampling fluctuation. Null hypothesis is denoted by H_0.

11.41 ERRORS

In sampling theory to draw valid inferences about the population parameter on the basis of the sample results.

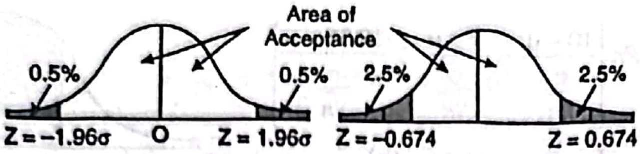
We decide to accept or to reject the lot after examining a sample from it. As such, we are liable to commit the following two types of errors.

Type 1 Error. If H_0 is rejected while it should have been accepted.

Type II Error. If H_0 is accepted while it should have been rejected.

11.42 LEVEL OF SIGNIFICANCE

There are two critical regions which cover 5% and 1% areas of the normal curve. The shaded portions are the critical regions.



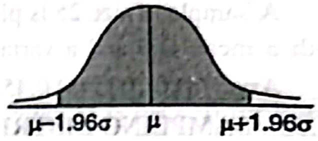
1% level of significance 5% level of significance

Thus, the probability of the value of the variate falling in the critical region is the level of significance. If the variate falls in the critical area, the hypothesis is to be rejected.

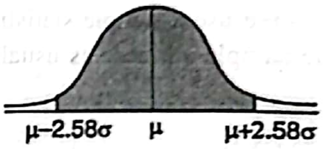
11.43 TEST OF SIGNIFICANCE

The tests which enables us to decide whether to accept or to reject the null hypothesis is called the tests of significance. If the difference between the sample values and the population values are so large (lies in critical area), it is to be rejected.

11.44 CONFIDENCE LIMITS



$\mu - 1.96\,\sigma, \mu + 1.96\,\sigma$ are 95% confidence limits as the area between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 95%. If a sample statistics lies in the interval $\mu - 1.96\,\sigma$, $\mu + 1.96\sigma$, we call 95% confidence interval.



Similarly, $\mu - 2.58\,\sigma$, $\mu + 2.58\,\sigma$ is 99% confidence limits as the area between $\mu - 2.58\,\sigma$ and $\mu + 2.58\,\sigma$ is 99%. The numbers 1.96, 2.58 are called confidence coefficients.

11.45 TEST OF SIGNIFICANCE OF LARGE SAMPLES ($N > 30$)

Normal distribution is the limiting case of Binomial distribution when n is large enough. For normal distribution 5% of the items lie outside $\mu \pm 1.96\,\sigma$ while only 1% of the items lie outside $\mu \pm 2.586\,\sigma$.

$$z = \frac{x - \mu}{\sigma}$$

where z is the standard normal variate and x is the observed number of successes.

First we find the value of z. Test of significance depends upon the value of z.

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

(*b*) If $|z| > 1.96$, difference is significant at 5% level of significance.

(*ii*) (*a*) If $|z| < 2.58$, difference between the observed and expected number of successes is not significant at1% level of significance.

(*b*) If $|z| > 2.58$, difference is significant at 1% level of significance.

Example 72. *A cubical die was thrown 9,000 times and 1 or 6 was obtained 3120 times. Can the deviation from expected value lie due to fluctuations of sampling?*

Solution. Let us consider the hypothesis that the die is an unbiased one and hence the probability of obtaining 1 or 6 $= \frac{2}{6} = \frac{1}{3}$ *i.e.,* $P = \frac{1}{3}$, $q = \frac{2}{3}$

The expected value of the number of successes $= np = 9000 \times \frac{1}{3} = 3000$

Also $\quad \sigma = \text{S.D.} = \sqrt{npq} = \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{2000} = 44.72$

$$3\sigma = 3 \times 44.72 = 134.16$$

Actual number of successes = 3120

Difference between the actual number of successes and expected number of successes

$$= 3120 - 3000 = 120 \text{ which is } < 3\sigma$$

Hence, the hypothesis is correct and the deviation is due to fluctuations of sampling due to random causes.

Ans.

11.46 SAMPLING DISTRIBUTION OF THE PROPORTION

A simple sample of n items is drawn from the population. It is same as a series of n independent trials with the probability P of success. The probabilities of 0, 1, 2, ..., n success are the terms in the binomial expansion of $(q+p)^n$.

Here mean = np and standard deviation = $\sqrt{npq}$.

Let us consider the proportion of successes, then

(*a*) Mean proportion of successes = $\frac{np}{n} = p$

(*b*) Standard deviation (standard error) of proportion of successes = $\frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$

(*c*) Precision of the proportion of success = $\frac{1}{\text{S.E.}} = \sqrt{\frac{n}{pq}}$.

Example 73. *A group of scientific mens reported 1705 sons and 1527 daughters. Do these figures conform to the hypothesis that the sex ratio is $\frac{1}{2}$.*

Solution. The total number of observation = 1705 + 1527 = 3232

The number of sons = 1705

Therefore, the observed male ratio = $\frac{1705}{3232} = 0.5175$

On the given hypothesis the male ratio = 0.5000

Thus, the difference between the observed ratio and theoretical ratio

$$= 0.5275 - 0.5000$$
$$= 0.0275$$

The standard deviation of the proportion $= s = \sqrt{\frac{pq}{n}} = \sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{3232}} = 0.0088$

The difference is more than 3 times of standard deviation.

Hence, it can be definitely said that the figures given do not conform to the given hypothesis.

11.47 ESTIMATION OF THE PARAMETERS OF THE POPULATION

The mean, standard deviation etc. of the population are known as parameters. They are denoted by μ and σ. Their estimates are based on the sample values. The mean and standard deviation of a sample are denoted by $\bar{x}$ and s respectively. Thus, a static is an estimate of the parameter. There are two types of estimates.

(*i*) *Point estimation*: An estimate of a population parameter given by a single number is called a point estimation of the parameter. For example,

$$(\text{S.D.})^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

(*ii*) *Interval estimation*: An interval in which population parameter may be expected to lie with a given degree of confidence. The intervals are

(*i*) $\bar{x} - \sigma_s$ to $\bar{x} + \sigma_s$ (68.27% confidence level)

(*ii*) $\bar{x} - 2\sigma_s$ to $\bar{x} + 2\sigma_s$ (95.45% confidence level)

(*iii*) $\bar{x} - 3\sigma_s$ to $\bar{x} + 3\sigma_s$ (99.73% confidence level)

$\bar{x}$ and σ_s are the mean and S.D. of the sample.

Similarly, $\bar{x} \pm 1.96\,\sigma_s$, $\bar{x} \pm 2.58\,\sigma_s$ are 95% and 99% confidence of limits for μ.

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ and $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ are also the intervals as $\sigma_s = \frac{\sigma}{\sqrt{n}}$.

11.48 COMPARISON OF LARGE SAMPLES

Let two large samples of size n_1, n_2 be drawn from two populations of proportions of attributes A's as P_1, P_2 respectively.

(*i*) *Hypothesis*: As regards the attribute A, the two populations are similar. On combining the two samples

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

where p is the common proportion of attributes.

Let e_1, e_2 be the standard errors in the two samples, then

$$e_1^2 = \frac{pq}{n_1} \text{ and } e_2^2 = \frac{pq}{n_2}$$

If e be the standard error of the combined samples, then

$$e = P_1^2 + P_2^2 = \frac{pq}{n_1} + \frac{pq}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

$$z = \frac{P_1 - P_2}{e}$$

1. If $z > 3$, the difference between P_1 and P_2 is significant.
2. If $z < 2$, the difference may be due to fluctuations of sampling.
3. If $2 < z < 3$, the difference is significant at 5% level of significance.

(*ii*) *Hypothesis*. In the two populations, the proportions of attribute A are not the same, then standard error e of the difference $p_1 - p_2$ is

$$e^2 = p_1 + p_2$$

$$= \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}, \quad z = \frac{P_1 - P_2}{e} < 3,$$

differenceisduetofluctuationsofsamples.

Example 74. *In a sample of 600 men from a certain city, 450 are found smokers. In another sample of 900 men from another city, 450 are smokers. Do the data indicate that the cities are significantly different with respect to the habit of smoking among men.*

Solution. $n_1 = 600$ men, Number of smokers $= 450$, $P_1 = \frac{450}{600} = 0.75$

$n_2 = 900$ men, Number of smokers $= 450$, $P_2 = \frac{450}{900} = 0.50$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{600 \times 0.75 + 900 \times 0.5}{600 + 900} = \frac{900}{1500} = 0.60$$

$$q = 1 - P = 1 - 0.6 = 0.4$$

$$e^2 = P_1^2 + P_2^2 = Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$$e^2 = 0.6 \times 0.4\left(\frac{1}{600} + \frac{1}{900}\right) = 0.000667$$

$$e = 0.02582$$

$$z = \frac{P_1 - P_2}{e} = \frac{0.75 - 0.50}{0.02582} = 9.682$$

$z > 3$ so that the difference is significant. **Ans.**

Example 75. *One type of aircraft 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. $n_1 = 100$ flights, Number of troubled flights $= 5$, $P_1 = \frac{5}{100} = \frac{1}{20}$

$n_2 = 200$ flights, Number of troubled flights $= 7$, $P_2 = \frac{7}{200}$

$$e^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2} = \frac{0.05 \times 0.95}{100} + \frac{0.035 \times 0.965}{200}$$

$$= 0.000475 + 0.0001689 = 0.0006439$$

$$e = 0.0254$$

$$z = \frac{0.05 - 0.035}{0.0254} = 0.59$$

$z < 1$, Difference is not significant. **Ans.**

11.49 THE *t* DISTRIBUTION (For small sample)

The student distribution is used to test the significance of

(*i*) The mean of a small sample.

(*ii*) The difference between the means of two small samples or to compare two small samples.

(*iii*) The correlation coefficient.

Let $x_1, x_2, x_3, \ldots, x_n$ be the members of random sample drawn from a normal population with mean μ. If $\bar{x}$ be the mean of the sample then

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{where} \quad s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

Example 76. *A machine which produces mica insulating washers for use in electric device to turn out washers having a thickness of 10 mm. A sample of 10 washers has an average thickness 9.52 mm with a standard deviation of 0.6 mm. Find out t.*

Solution. $\bar{x} = 9.52$, $M = 10$, $S' = 0.6$, $n = 10$

$$t = \frac{\bar{x}-M}{\frac{s}{\sqrt{n}}} = \frac{9.52-10}{\frac{0.6}{\sqrt{10}}} = \frac{0.48\sqrt{10}}{0.6} = -\frac{4}{5}\sqrt{10}$$

$$= -0.8 \times 3.16 = -2.528$$ **Ans.**

Example 77. *Compute the students t for the following values in a sample of eight: −4, −2, −2, 0, 2, 2, 3, 3 taking the mean of universe to be zero.* (*A.M.I.E., Summer 1995*)

Solution. $\mu = 0$

S.No.	x	$x-\bar{x} = \left(x-\frac{1}{4}\right)$	$(x-\bar{x})^2 = \left(x-\frac{1}{4}\right)^2$
1	−4	$-\frac{17}{4}$	$\frac{289}{16}$
2	−2	$-\frac{9}{4}$	$\frac{81}{16}$
3	−2	$-\frac{9}{4}$	$\frac{81}{16}$
4	0	$-\frac{1}{4}$	$\frac{1}{16}$
5	2	$\frac{7}{4}$	$\frac{49}{16}$
6	2	$\frac{7}{4}$	$\frac{49}{16}$
7	3	$\frac{11}{4}$	$\frac{121}{16}$
8	3	$\frac{11}{4}$	$\frac{121}{16}$
$n = 8$	$\Sigma x = 2$		$\Sigma(x-\bar{x})^2 = \frac{792}{16}$

$$\bar{x} = \frac{2}{8} = \frac{1}{4}$$

$$\text{S.D.} = s = \sqrt{\frac{(x-\bar{x})^2}{n-1}} = \sqrt{\frac{792}{16\times 7}} = \sqrt{7.07} = 2.66$$

$$t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{1}{4}-0}{\frac{2.66}{\sqrt{8}}} = \frac{\sqrt{8}}{4\,(2.66)} = \frac{2.83}{10.64} = 0.266$$ **Ans.**

11.50 WORKING RULE

To calculate significance of sample mean at 5% level.

Calculate $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$ and compare it to the value of t with $(n-1)$ degrees of freedom at 5% level, obtained from the table. Let this tabulated value of t be t_1.

If $t < t_1$, then we accept the hypothesis *i.e.*, we say that the sample is drawn from the population.

If $t > t_1$, we compare it with the tabulated value of t at 1% level of significance for $(n-1)$ degrees of freedom. Denote it by t_2. If $t_1 < t < t_2$ then we say that the value of t is significant.

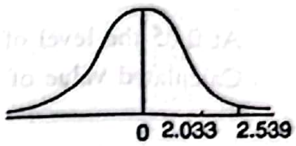
If $t > t_1$, we reject the hypothesis and the sample is not drawn from the population.

Example 78. *A manufacturer intends that his electric bulbs have a life of 1000 hours. He tests a sample of 20 bulbs, drawn at random from a batch and discovers that the mean life of the sample bulbs is 990 hours with a s.d. of 22 hours. Does this signify that the batch is not up to the standard?*

[Given: *The table value of t at 1% level is significance with 19 degrees of freedom is 2.539*]

Solution. $\bar{x} = 990$, $\sigma = 22$, $x = 1000$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1000 - 990}{\frac{22}{\sqrt{20}}}$$



$$= \frac{10\sqrt{20}}{22} = \frac{22.36}{11} = 2.033$$

Since the calculated value of t (2.032) is less than the value of t (2.539) from the table. Hence, it is not correct to say that this batch is not upto this standard. **Ans.**

Example 79. *Ten individuals are chosen at random from a population and their heights are found to be in inches 63, 63, 64, 65, 66, 69, 70, 70, 71. Discuss the suggestion that the Mean height of universe is 65.*

For 9 degree of freedom t at 5% level of significance = 2.262.

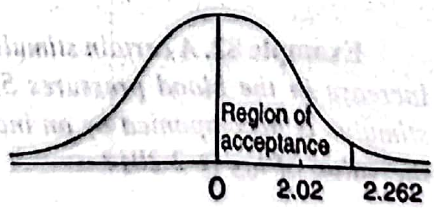
Solution.

x	$x - 67$	$(x-67)^2$
63	-4	16
63	-4	16
64	-3	9
65	-2	4
66	-1	1
69	$+2$	4
69	$+2$	4
70	$+3$	9
70	$+3$	9
71	$+4$	16
$\Sigma x = 670$		$\Sigma(x-\bar{x})^2 = 88$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{670}{10} = 67, \; s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{88}{9}} = 3.13$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{67 - 65}{\frac{3.13}{\sqrt{10}}}$$

$$= \frac{2\sqrt{10}}{3.13} = 2.02$$



$$2.02 < 2.262$$

Calculated value of t (2.02) is less than the table value of t (2.262). The hypothesis is accepted the mean height of universe is 65 inches. **Ans.**

Example 80. *The mean life time of sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. The company claims that the average life of the bulbs produced by it is 1600 hours. Using the level of significance of 0.05, is the claim acceptable?*

Solution. $\bar{x} = 1570,\ S = 120,\ n = 100,\ \mu = 1600$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{1570-1600}{\frac{120}{\sqrt{100}}} = \frac{1570-1600}{12} = 2.5$$

At 0.05 the level of significance, $t = 1.96$.

Calculated value of $t >$ Table value of t.

$$2.5 > 1.96$$

Hence the claim is to be rejected. **Ans.**

Example 81. *A sample of 6 persons in an office revealed an average daily smoking of 10, 12, 8, 9, 16, 5 cigarettes. The average level of smoking in the whole office has to be estimated at 90% level of confidence.*

t = 2.015 for 5 degree of freedom.

Solution.

x	$x - 10$	$(x-10)^2$
10	0	0
12	2	4
8	-2	4
9	-1	1
16	$+6$	36
5	-5	25
Total	0	$\Sigma(x-10)^2 = 70$

$$\text{Mean} = a + \frac{\Sigma fd}{\Sigma f} = 10 + \frac{0}{6} = 10$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{70}{5}} = 3.74$$

At 90% level of confidence, $t = \pm 2.015$.

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \quad \text{or} \quad \pm 2.015 = \frac{10-\mu}{\frac{3.74}{\sqrt{6}}}$$

or $$\mu = 2.015 \times \frac{3.74}{\sqrt{6}} + 10 = 6.92,\ 13.08$$ **Ans.**

Example 82. *A certain stimulus administered to each of 12 patients resulted in the following increase in the blood pressures 5, 2, 8, –1, 3, 0, 6, –2, 1, 5, 0, 4. Can it be calculated that stimulus is accompanied by an increase in blood pressure given that for 11 degrees of freedom the value of $t_{0.5}$ is 2.201?* (A.M.I.E., Summer 1998)

Solution.

$$\bar{x} = \frac{5+2+8-1+3+0+6-2+1+5+0+4}{12}$$

$$= \frac{31}{12} = 2.583 = 2.6 \text{ approx.}$$

x	$x-2.6$	$(x-2.6)^2$
5	2.4	5.76
2	−0.6	0.36
8	5.4	29.16
−1	−3.6	12.96
3	0.4	0.16
0	−2.6	6.76
6	3.4	11.56
−2	−4.6	21.16
1	−1.6	2.56
5	2.4	5.76
0	−2.6	6.76
4	1.4	1.96
$\Sigma x = 12$	$\Sigma(x-2.6)$	$\Sigma(x-2.6)^2 = 104.92$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{104.92}{12-1} = 9.54$$

$$s = 3.08$$

Assuming that the stimulas will not be accompanied by increase in blood pressure, *i.e.*, the mean of increase in blood pressure for the population is zero, we have

$$t = \frac{\bar{x}-\mu}{s}\sqrt{n} = \frac{2.6-0}{3.08}\sqrt{12} = \frac{2.6}{3.08}\times 3.464 = 2.924$$

As the computed value of t, *i.e.*, 2.924 is greater than $t_{0.05}$, *i.e.*, 2.201 we find that our assumption is wrong and we conclude that as a result of the stimulus blood pressure will increase. **Ans.**

Example 83. *A fertiliser mixing machine is set to give 12 kg of nitrate for quintal bag of fertiliser. Ten 100 kg bags are examined. The percentages of nitrate per bag are as follows:*

11, 14, 13, 12, 13, 12, 13, 14, 11, 12

Is there any reason to believe that the machine is defective? Value of t for 9 degrees of freedom is 2.262. (A.M.I.E., Winter 1997)

Solution.

The calculation of $\bar{x}$ and s is given in the following table:

x	$d = x-12$	d^2
11	−1	1
14	2	4
13	1	1
12	0	0
13	1	1

12	0	0
13	1	1
14	2	4
11	–1	1
12	0	0
$\Sigma x = 125$	$\Sigma d = 5$	$\Sigma d^2 = 13$

$$\mu = 12 \quad \text{kg}, \quad n = 10, \bar{x} = \frac{\Sigma x}{n} = \frac{125}{10} = 12.5$$

$$s^2 = \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2 = \frac{13}{10} - \left(\frac{5}{10}\right)^2 = \frac{13}{10} - \frac{1}{4} = \frac{21}{20} = \frac{105}{100}$$

$$s = 1.024$$

Value of t for 9 degrees of freedom = 2.262

Also
$$t = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

$$= \frac{12.5 - 12}{1.024}\sqrt{10} = 1.54$$

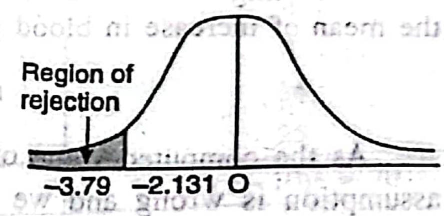
Since the value of t is less than 2.262, there in no reason to believe that machine is defective. **Ans.**

Example 84. *A random sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviationn from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% and 99% confidence limits of the mean of the population.*

$$\gamma = 15, \quad \alpha = 0.05, \; t = 2.131$$

$$\alpha = 0.01, \; t = 2.947$$

Solution. $\mu = 56$, $n = 16$, $\bar{x} = 53$, $\Sigma(x - \bar{x})^2 = 150$



$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{150}{15} = 10$$

$$s = \sqrt{10}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{53 - 56}{\frac{\sqrt{10}}{\sqrt{16}}} = \frac{-3 \times 4}{\sqrt{10}} = -3.79$$

$$t = 3.79$$

$3.79 > 2.131$ and also $3.79 \; 2.947$.

Thus, the sample cannot be regarded as taken from the population. **Ans.**

11.51 TESTING FOR DIFFERENCE BETWEEN MEANS OF TWO SMALL SAMPLES

Let the mean and variance of the first population be μ_1 and σ_1^2 and $\mu_2 \cdot \sigma_2^2$ be the mean and variance of the second population.

Let $\bar{x}_1$ be the mean of small sample of size n_1 from first population and $\bar{x}_2$ the mean of a sample of size n_2 from second population.

We know that

$$E(\bar{x}_1) = \mu_1 \text{ and } Var(\bar{x}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{x}_2) = \mu_2 \text{ and } Var(\bar{x}_2) = \frac{\sigma_2^2}{n_2}$$

If the samples are independent, then $\bar{x}_1$ and $\bar{x}_2$ are also independent.

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2 \text{ and } Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{x}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

then

$$(\bar{x}_1 - \bar{x}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}}$$

If the population is the same then

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (\mu_1 - \mu_2 = \mu_1 - \mu_1 = 0)$$

Example 85. *Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in ounces):*

Sample 1: 9 11 13 11 15 9 12 14

Sample 2: 10 12 10 14 9 8 10

Is the difference between the means of the sample significant?

[Given for V = 13, $t_{0.05}$ = 2.16] (A.M.I.E., Winter 1996)

Solution.

Assumed mean of x = 12, Assumed mean of y = 10

x	$(x-12)$	$(x-12)^2$	y	$(y-10)$	$(y-10)^2$
9	−3	9	10	0	0
11	−1	1	12	2	4
13	1	1	10	0	0
11	−1	1	14	4	16
15	3	9	9	−1	1
9	−3	9	8	−2	4
12	0	0	10	0	0
14	2	4	–	–	–
94	−2	34	73	3	25

$$\bar{x} = \frac{\Sigma x}{n} = \frac{94}{8} = 11.75$$

$$\sigma_x^2 = \frac{\Sigma(x-12)^2}{n} - \left(\Sigma\frac{(x-12)}{n}\right)^2 = \frac{34}{8} - \left(\frac{-2}{8}\right)^2 = 4.1875$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{73}{7} = 10.43$$

$$\sigma_y^2 = \frac{\Sigma(y-10)^2}{n} - \left[\frac{\Sigma(y-10)}{n}\right]^2 = \frac{25}{7} - \left(\frac{3}{7}\right)^2 = 3.438$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}} = \sqrt{\frac{34+25}{8+7-2}} = \sqrt{\frac{59}{13}} = \sqrt{4.54} = 2.13$$

$$t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{11.75-10.43}{2.13\sqrt{\frac{1}{8}+\frac{1}{7}}} = \frac{1.32}{2.13\sqrt{0.268}} = \frac{1.32}{2.13\times 0.518}$$

$$= \frac{1.32}{1.103} = 1.12$$

The 5% value of t for 13 degree of freedom is given to be 2.16. Since calculated value of t is 1.12 is less than 2.16, the difference between the means of samples is not significant. **Ans.**

Exercise 11.8

1. A random sample of six steel beams has mean compressive strength of 58.392 psi (pounds per square inch) with a standard deviation of $s = 648$ psi. Test the null hypothesis $H_0 = \mu = 58{,}000$ psi against the alternative hypothesis $H_1 : \mu > 58{,}000$ psi at 5% level of significance (value for t at 5 degree of freedom and 5% significance level is 2.0157). Here μ denotes the population mean. *(A.M.I.E., Summer 2000)*

2. A certain cubical die was thrown 96 times and shows 2 upwards 184 times. Is the die biased? **Ans.** die is biased.

3. In a sample of 100 residents of a colony 60 are found to be wheat eaters and 40 rice eaters. Can we assume that both food articles are equally popular?

4. Out of 400 children, 150 are found to be under weight. Assuming the conditions of simple sampling, estimate the percentage of children who are underweight in, and assign limits within which the percentage probably lies. **Ans.** 37.5% approx. Limits = 37.5 ± 3 (2.4)

5. 500 eggs are taken at random from a large consignment, and 50 are found to be bad. Estimate the percentage of bad eggs in the consignment and assign limits within which the percentage probably lies. **Ans.** 10%, 10 ± 3.9

6. A machine puts out 16 imprefect articles in a sample of 500. After the machine is repaired, puts out 3 imprefect articles in a batch of 100. Has the machine been improved? **Ans.** The machine has not been improved.

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? **Ans.** $z = 0.37$, Difference between proportions is 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? **Ans.** $z = 2.5$, not hidden at 5% level of significance.

9. One thousand articles from a factory are examined and found to be three percent defective. Fifteen hundred similar articles from a second factory are found to be only 2 percent defective. Can it reasonably be concluded that the product of the first factory is inferior to the second? **Ans.** It cannot be reasonable concluded that the product of the first factory is inferior to that of the second.

10. A manufacturing company claims 90% assurance that the capacitors manufactured by them will show a tolerance of better than 5%. The capacitors are packaged and sold in lots of 10. Show that about 26% of his customers ought to complain that capacitors do not reach the specified standard.
11. An experiment was conducted on nine individuals. The experiment showed that due to smoking, the pulse rate increased in the following order:

$$5,\ 3,\ 4,\ -1,\ 2,\ -3,\ 4,\ 3,\ 1.$$

Can you maintain that smoking leads to an increase in the pulse rate?
(t for 8 d.f. at 5% level of significance = 2.31). **Ans.** Yes.
12. Nine patients to whom a certain drink was administered registered the following in blood pressure:

$$7, 3, -1, 4, -3, 5, 6, -4, 1.$$

Show that the data do not indicate that the drink was responsible for these increments.
13. A machine has produced washers having a thickness of 0.50 mm. To determine whether the machine is in proper working order, a sample of 10 washers is chosen for which the mean thickness is 0.53 mm. and the standard deviation is 0.03 mm. Test the hypothesis that the machine is in proper working order using a level of significance (*a*) 0.05 (*b*) 0.01.

Ans. (*a*) The machine is not in proper working order at 0.05 level of significance.
(*b*) The machine is in proper working order at 0.01 level of significance.
14. Eleven school boys were given a test in drawing. They were given a months further tuition and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefitted by extra coaching.

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I Test	23	20	19	21	18	20	18	17	23	16	19
Marks II Test	24	19	22	18	20	22	20	20	23	20	17

Ans. t = 1.48, The value of t is not significant at 5% level of significance. (*i.e.*, the test, i.e., the students) no evidence that the students have benefitted by extra coaching.
15. Two horses *A* and *B* were tested according to the time (in seconds) to run a particular race with the following results:

Horse A	28	30	32	33	33	29	39
Horse B	29	30	30	24	27	29	

Test whether you can discriminate between two horses? **Ans.** Yes with 75% confidence.

11.52 THE CHI-SQUARE DISTRIBUTION

Chi-square is a measure of actual divergence of the observed and expected frequencies. If f_0 is the observed frequency and f_e the expected frequency of a class interval, then x^2 is defined as

$$x^2 = \Sigma\frac{(f_0 - f_e)^2}{f_e}$$

11.53 DEGREE OF FREEDOM (*df*)

The degree of freedom refers to the number of "independent constrants" in a set of data. We shall illustrate this concept with example. There is a 2×2 association table and the actual frequencies as under:

Let the two attributes *A* and *B* be independent.

$$\text{Expected frequency of } (AB) = \frac{30 \times 60}{100} = 18$$

After finding the frequency of (AB), the expected frequencies of the remaining three classes are automatically fixed.

$$\text{Expected frequency of } (\alpha B) = 60 - 18 = 42$$

$$\text{Expected frequency of } (A\,\beta) = 30 - 18 = 12$$

$$\text{Expected frequency of } (\alpha\,\beta) = 70 - 42 = 28$$

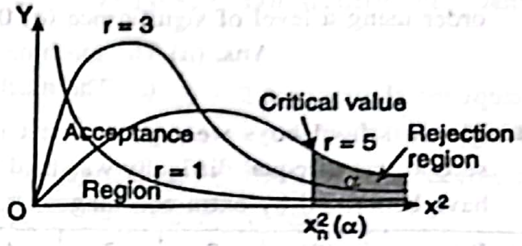
It means that only one choice is fixing of frequency of AB is independent choice. Frequencies of the remaining three classes depend on the frequency of (AB). It means, we have only one degree of freedom.

$$\text{Degree of freedom} = (r-1)(c-1)$$

where r is the number of rows and c is the number of columns.

11.54 x^2 – CURVE

Let $x_1, x_2, \ldots, x_n$ be n standard variates with mean zero and S.D. unity. Then x^2-distribution has the $x_1^2, x_2^2, x_3^2, \ldots, x_n^2$ random variates.



Equation of the x^2-curve is

$$y = y_0\, e^{-\frac{x^2}{2}}\, (x^2)^{\frac{r-1}{2}}, \quad r = n - 1$$

where r is the degree of freedom.

Since this equation does not have any parameter. So it can be used for every problem of chi-square. $x_n^2(\alpha)$ denote the value of chi-square for n degree of freedom such that the area to the right of this point is α.

11.55 GOODNESS OF FIT

The value of x^2 is used to find the divergence of the observed frequency from the expected frequency.

If the value of P is high the fit is said to be good. It means that there no significant divergence between observed and expected data.

If the curve of the expected frequency is super imposed on the curve of observed frequencies there would not be much divergence between the two. The fit would be good. If the value of P is small, the fit is said to be poor.

11.56 STEPS FOR TESTING

(*i*) First calculate the value of x^2.

(*ii*) From the table read the value of x^2 for a given degree of freedom.

(*iii*) Find out the probability P corresponding to the calculated values of x^2.

(*iv*) If $P > 0.05$, the value is not significant and it is a good fit.

(*v*) If $P < 0.05$ the deviations of are significant.

Example 86. *The following table is given*

		Eye colour in sons		
		not light	*light*	
Eye colour in fathers	*Not light*	230	148	378
	light	251	471	622
		381	619	1000

Test whether the colour of the son's eyes is associated with that of the fathers.

Given: *value of x^2 is 3.84 for 1 degree of freedom.*

Solution.

Hypothesis: Let the eye colour of sons and the eye colour of fathers independent.

		Eye colour in sons	
		not light	*light*
Eye colour in fathers	Not light	$\frac{378 \times 381}{1000} = 144$	$\frac{378 \times 619}{1000} = 234$
	light	$\frac{622 \times 381}{1000} = 237$	$\frac{622 \times 619}{1000} = 385$

$$X^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$X^2 = \frac{(230-144)^2}{144} + \frac{(148-234)^2}{234} + \frac{(151-237)^2}{237} + \frac{(471-385)^2}{385}$$

$$= (86)^2\left[\frac{1}{144} + \frac{1}{234} + \frac{1}{237} + \frac{1}{385}\right] = 133.37$$

The degree of freedom $= (c-1)(r-1) = (2-1)(2-1) = 1$

The value of x^2 at 5% level of significance for 1 degree of freedom is 3.841 and the calculated value is 133.37

$$133.37 > 3.841$$

This leads to the conclusion that the hypothesis is wrong and the colour of son's eyes is associated with that of the fathers to a great extent. **Ans.**

Example 87. *From the following table, showing the number of plants having certain characters, test the hypothesis that the flower colour is independent of flatness of leaf.*

	Flat leaves	*Curled leaves*	*Total*
White Flowers	99	36	135
Red Flowers	20	5	25
Total	119	41	160

Solution. Null Hypothesis: The flower colour is dependent of flatness of leaf.

The following table shows the theoritical frequencies.

	Flat leaves	*Curled leaves*	*Total*
White flowers	$\frac{135 \times 119}{160} = 100$	$\frac{135 \times 41}{160} = 35$	135
Red flowers	$\frac{25 \times 119}{160} = 19$	$\frac{25 \times 41}{160} = 6$	25
Total	119	41	160

$$x^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$x^2 = \frac{(99-100)^2}{100} + \frac{(36-35)^2}{35} + \frac{(20-19)^2}{19} + \frac{(5-6)^2}{6}$$

$$x^2 = \frac{1}{100} + \frac{1}{35} + \frac{1}{19} + \frac{1}{6} = 0.2579$$

Degree of freedom = $(r-1)(c-1) = (2-1)(2-1) = 1$

We have $\chi^2 = 0.0158$ at 0.1 level of significance.

$$0.2579 > 0.0158$$

This leads to the conclusion that the hypothesis is wrong and the flower colour is independent of flatness of leaf at the 0.1 level of significance. **Ans.**

Example 88. *The following table gives the number of aircraft accidents that occurs during various days of the week. Find whether the accidents are uniformly distributed over the week.*

Days	*Sun.*	*Mon.*	*Tues.*	*Wed.*	*Thurs.*	*Fri.*	*Sat.*
No. of accidents	*14*	*16*	*8*	*12*	*11*	*9*	*14*

Given: *The values of chi-square significant at 5, 6, 7, d.f. are respectively 11.07, 12.59, 14.07 at the 5% level of significance.*

Solution. Null Hypothesis: The accidents are uniformly distributed over the week.

Expected frequencies of the accidents are given below:

Days	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Total
No. of accidents	12	12	12	12	12	12	12	84

$$\chi^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$= \frac{1}{12}[4+16+16+0+1+9+4] = \frac{50}{12} = 4.17$$

The number of degrees of freedom = Number of observations – Number of independent constants = 7 – 1 = 6.

The tabulated $\chi^2_{0.05}$ for 6 d.f. = 12.59

Since the calculated χ^2 is much less than the tabulated value, we accept the null hypothesis. Hence, the accidents are uniformly distributed over the week. **Ans.**

Example 89. *A set of five similar coins is tossed 320 times and the result is*

No. of heads	*0*	*1*	*2*	*3*	*4*	*5*
Frequency	*6*	*27*	*72*	*112*	*71*	*32*

Test the hypothesis that the data follow a binomial distribution. **(Warangal 1995)**

Solution. P (Head) $= \frac{1}{2}$, $q = 1 - \frac{1}{2} = \frac{1}{2}$

Theoretical frequencies are

$$P(0H) = q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}, \text{ Frequency of 0 head} = \frac{320}{32} = 10$$

$$P(1H) = {}^5C_1\, pq^4 = {}^5C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{5}{32}, \text{ Frequency of 1 head} = \frac{5}{32} \times 320 = 50$$

$$P(2H) = {}^5C_2\, p^2 q^3 = 10\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}, \text{ Frequency of 2 heads} = \frac{10}{32} \times 320 = 100$$

$$P(3H) = {}^5C_3\, p^3 q^2 = 10\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}, \text{ Frequency of 3 heads} = \frac{10}{32} \times 320 = 100$$