(17E00105) STATISTICS FOR MANAGERS

The objective of this course is to familiarize the students with the statistical techniques popularly used in managerial decision making. It also aims at developing the computational skill of the students relevant for statistical analysis.

1.Introduction of statistics – Nature & Significance of Statistics to Business, , Measures of Central Tendency- Arithmetic – Weighted mean – Median, Mode – Geometric mean and Harmonic mean – Measures of Dispersion, range, quartile deviation, mean deviation, standard deviation, coefficient of variation – Application of measures of central tendency and dispersion for business decision making.

2. Correlation: Introduction, Significance and types of correlation – Measures of correlation – Co-efficient of correlation. Regression analysis – Meaning and utility of regression analysis – Comparison between correlation and regression – Properties of regression coefficients-Rank Correlation.

3. Probability – Meaning and definition of probability – Significance of probability in business application – Theory of probability –Addition and multiplication – Conditional laws of probability – Binominal – Poisson – Uniform – Normal and exponential distributions.

4. Testing of Hypothesis- Hypothesis testing: One sample and Two sample tests for means and proportions of large samples (z-test), One sample and Two sample tests for means of small samples (t-test), F-test for two sample standard deviations. ANOVA one and two way .

5. Non-Parametric Methods: Chi-square test for single sample standard deviation. Chi-square tests for independence of attributes - Sign test for paired data.

Textbooks:

- Statistical Methods, Gupta S.P., S.Chand. Publications

References:

- Statistics for Management, Richard I Levin, David S.Rubin, Pearson,
- Business Statistics, J.K.Sharma, Vikas house publications house Pvt Ltd
- Complete Business Statistics, Amir D. Aezel, Jayavel, TMH,
- Statistics for Management, P.N.Arora, S.Arora, S.Chand
- Statistics for Management , Lerin, Pearson Company, New Delhi.
- Business Statistics for Contemporary decision making, Black Ken, New age publishers.
- Business Statistics, Gupta S.C & Indra Gupta, Himalaya Publishing House, Mumbai

(17E00105) STATISTICS FOR MANAGERS

The objective of this course is to familiarize the students with the statistical techniques popularly used in managerial decision making. It also aims at developing the computational skill of the students relevant for statistical analysis.

1.Introduction of statistics – Nature & Significance of Statistics to Business, , Measures of Central Tendency- Arithmetic – Weighted mean – Median, Mode – Geometric mean and Harmonic mean – Measures of Dispersion, range, quartile deviation, mean deviation, standard deviation, coefficient of variation – Application of measures of central tendency and dispersion for business decision making.

2. Correlation: Introduction, Significance and types of correlation – Measures of correlation – Co-efficient of correlation. Regression analysis – Meaning and utility of regression analysis – Comparison between correlation and regression – Properties of regression coefficients-Rank Correlation.

3. Probability – Meaning and definition of probability – Significance of probability in business application – Theory of probability –Addition and multiplication – Conditional laws of probability – Binominal – Poisson – Uniform – Normal and exponential distributions.

4. Testing of Hypothesis- Hypothesis testing: One sample and Two sample tests for means and proportions of large samples (z-test), One sample and Two sample tests for means of small samples (t-test), F-test for two sample standard deviations. ANOVA one and two way .

5. Non-Parametric Methods: Chi-square test for single sample standard deviation. Chi-square tests for independence of attributes - Sign test for paired data.

Textbooks:

- Statistical Methods, Gupta S.P., S.Chand. Publications

References:

- Statistics for Management, Richard I Levin, David S.Rubin, Pearson,
- Business Statistics, J.K.Sharma, Vikas house publications house Pvt Ltd
- Complete Business Statistics, Amir D. Aezel, Jayavel, TMH,
- Statistics for Management, P.N.Arora, S.Arora, S.Chand
- Statistics for Management , Lerin, Pearson Company, New Delhi.
- Business Statistics for Contemporary decision making, Black Ken, New age publishers.
- Business Statistics, Gupta S.C & Indra Gupta, Himalaya Publishing House, Mumbai

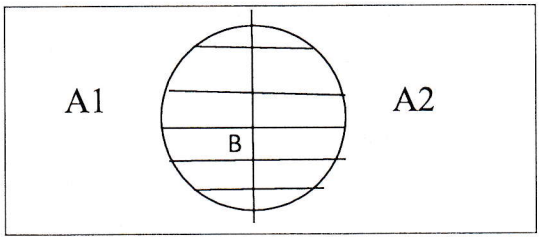
5. CONDITIONAL LAWS OF PROBABILITY :

5.1 BAYE'S THEOREM: The probability is known in different names, posterior probability, revised probability and Inverse probability. This has been introduced by "**Thomas Baye's**" an English mathematician been in this work known as Bayesian Decision theory published in 1763. This theory consists of finding the probability of an event taking into account of a given sample information.

Baye's theorem is a means for qualifying uncertainty. Based on the probability theory, the theorem defines a rule for refining a hypothesis by factoring in additional evidence and back ground information, and leads to a number representing the degree of probability that the hypothesis is true.

Thus a sample of 3 defective items out of 100 might be used to estimate the probability that a machine is (event A) not working properly (event B).

It is to be denoted that the Bayesian probability is based on the formula of conditional probability where A_1 & A_2 are two events which are mutually exclusive & exhaustive & B is a simple event which intersects each of the A events as shown in the Venn diagram to the right.



This is called posterior probability because; it is calculated after information is taken in to account. This is called revised probability as it is determined by revising the prior probabilities in the light of the additional information gathered. Further this is called inverse probability also, as it consists of finding the probability of a problem.

However, the Bayesian or the posterior probabilities are always conditional probabilities which are calculated for every events as follows.

5.2 Mutually Exclusive Events: If an event E can only occur in combination with one of the mutually exclusive events E_1, E_2 ----En than

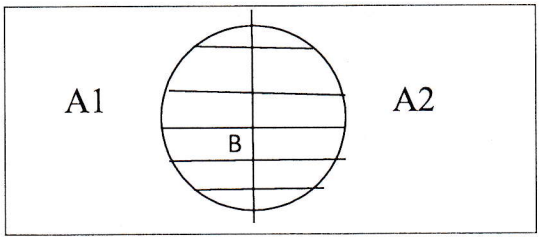
5. CONDITIONAL LAWS OF PROBABILITY :

5.1 BAYE'S THEOREM: The probability is known in different names, posterior probability, revised probability and Inverse probability. This has been introduced by "**Thomas Baye's**" an English mathematician been in this work known as Bayesian Decision theory published in 1763. This theory consists of finding the probability of an event taking into account of a given sample information.

Baye's theorem is a means for qualifying uncertainty. Based on the probability theory, the theorem defines a rule for refining a hypothesis by factoring in additional evidence and back ground information, and leads to a number representing the degree of probability that the hypothesis is true.

Thus a sample of 3 defective items out of 100 might be used to estimate the probability that a machine is (event A) not working properly (event B).

It is to be denoted that the Bayesian probability is based on the formula of conditional probability where A_1 & A_2 are two events which are mutually exclusive & exhaustive & B is a simple event which intersects each of the A events as shown in the Venn diagram to the right.



This is called posterior probability because; it is calculated after information is taken in to account. This is called revised probability as it is determined by revising the prior probabilities in the light of the additional information gathered. Further this is called inverse probability also, as it consists of finding the probability of a problem.

However, the Bayesian or the posterior probabilities are always conditional probabilities which are calculated for every events as follows.

5.2 Mutually Exclusive Events: If an event E can only occur in combination with one of the mutually exclusive events E_1, E_2 ----En than

$$P(E_K|E) = \frac{[P(E_K)][P(E|E_K)]}{\sum_{i=1}^{n} P(E_1)\ P(E|E_1)}$$

: where K=1, 2,-------n

Mutually Exclusive & Exhaustive Events :-If A_1, A_2 are two Mutually Exclusive and Exhaustive events

$$P(A_1|B) = \frac{P(A_1)\ P(B|A_1)}{P(A_1)\ P(B|A_1) + P(A_2)\ P(B|A_2)}$$

$$P(A_2|B) = \frac{P(A_2)\ P(B|A_2)}{P(A_1)\ P(B|A_1) + P(A_2)\ P(B|A_2)}$$

Q)Assumed that a factory has 2 machines past records shows that machine 1 produces 30% of the items of the output and machine 2 produces 70% of the items from the output further 5% of items produced by machine 1 for defective and only 1% produced by machine 2 for defective. If a defective item is drawn and random. What is the probability that the defective items produced by machine 1 (or) machine 2.

Sol: Let a_1: items produced by machine 1

a_2: items produced by machine 2

b: defective items produced by either 1 (or) 2 machines.

Probability of the items produced by machine 1

$P(A_1) = 30\% = 30/100 = 0.3$

Probability of the items produced by machine 2

$P(A_2) = 70\% = 70/100 = 0.7$

Probability of the defective items in machine 1

$P(B/A_1) = 5\% = 5/100 = 0.05$

Probability of the defective items in machine 2

$P(B/A2) = 1\ \% = 1/100 = 0.01$

Probability of the defective items produced by machine 1

$$P\left(\frac{A_1}{B}\right) = \frac{P(A_1) \cdot P(B/A_1)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2)}$$

$$= \frac{0.3 \text{ X}0.05}{0.3\text{X } 0.05 + 0.7\text{X } 0.01}$$

$$= \frac{0.015}{0.015 + 0.007}$$

$$= \frac{0.015}{0.022} = 0.68$$

Probability of defective items produced by machine 2

$$P\left(\frac{A_2}{B}\right) = \frac{P(A_2)\ P(B/A_2)}{P(A_1) \cdot P(B/A_1) + P(A_2)\ P(B/A_2)}$$

$$= \frac{0.7 \times 0.01}{0.3\text{X } 0.05 + 0.7\text{X } 0.01}$$

$$= \frac{0.007}{0.015 + 0.07}$$

$$= \frac{0.007}{0.022} = 0.32$$

P (A2/B) =0.32

Q)In a Bolt factory machine a_1, machine a_2 and machine a_3 manufactures respectively 25%, 35% and 40% of the total of their output 5, 4, 2 percentages are defective bolts produced by the machines. A bolt is drawn at a random from the product is found to defective. What is the probability that it was manufactured by machine "3"?

Sol: - P (A_1) =25%

= 25/100 = 0.25

P (A_2) = 35%

= 35/100 =035

P (A_3) = 40%

= 40/100 =0.40

$P(B/A_1) = 5\%$

$= 5/100 = 0.05$

$P(B/A_2) = 4\%$

$= 4/100 = 0.04$

$P(B/A_3) = 2\%$

$= 2/100 = 0.02$

$$P\left(\frac{A3}{B}\right) = \frac{P(A3) \cdot P(B/A3)}{P(A1) \cdot P(B/A1) + P(A2) \cdot P(B/A2) + P(A3) \cdot P(B/A3)}$$

$$= \frac{0.40 \times 0.02}{0.25x0.05+0.35x0.04+0.40x0.02}$$

$$= \frac{0.08}{0.0125+0.014+0.08}$$

$$= \frac{0.08}{0.0345}$$

P (A3/B) = 0.231

$$P\left(\frac{A2}{B}\right) = \frac{P(A2) \cdot P(B/A2)}{P(A1) \cdot P(B/A1) + P(A2) \cdot P(B/A2) + P(A3) \cdot P(B/A3)}$$

$$= \frac{0.35 \times 0.04}{0.25x0.05+0.35x0.04+0.40x0.02}$$

$$= \frac{0.014}{0.0345}$$

$= 0.4057$

$$P\left(\frac{A1}{B}\right) = \frac{P(A1) \cdot P(B/A1)}{P(A1) \cdot P(B/A1) + P(A2) P(B/A2) + P(A3) \cdot P(B/A3)}$$

$$= \frac{0.25 \times 0.05}{0.25x0.05+0.35x0.04+0.40x0.02}$$

$$= \frac{0.0125}{0.0345}$$

P (A1/B) = 0.3623

5.3 Needs of Baye's Theorem:-

- The sample space is portioned in to a set of mutually exclusive events (A_1, A_2....An).
- within the sample space, there exists on event B, for which P (B) > 0
- The analytical goal is to compute a conditional probability of the form P (A_K/B)
- At least one of the two sets of probabilities described below :

i) P ($A_k \cap B$) for each A_k

ii) P (A_K) and P (B/A_k) for each A_k.

Features:-

1. Through it deals with a conditional probability; its interpretation is different from that of the general conditional probability theorem.
2. Very useful to decision making.
3. The nations of priors and 'posterior' in 'Bayes' theorem are relative to a given sample outcome.

Applications:-

1. The theorem still prescribes multiplying the prior distributing by the likelihood function and them normalizing, to get the posterior distribution.
2. As a formal theorem, Bayes theorem is valid in all common interpretations of probability.

Problems:-

Q) Assumed that a factory has 2 machines past records. shows that machine 1 produces 30 % of the items of the output and machine 2 produces 70% of the items from the output further 5% of items produced by machine 1 for defective and only 1% produced by machine 2 for defective. If a defective item is drawn and random. what is the probability that the defective items produced by machine 1 (or) machine '2'

Sol: - Let a_1=items produced by machine1

a_2= items produced by machine2

b= defective items produced by either 1 (or) 2 machines

Probability of the items produced by machine1

P (A_1) =30%

$= 30/100 = 0.3$

Probability of the items produced by machine 2

$P(A_2) = 70\% = 70/100 = 0.7$

The probability of the defective items in machine1

$P(B/A_1) = 5\% = 5/100 = 0.05$

Probability of the defective items in machine 2

$P(B/A_2) = 1\% = 1/100 = 0.01$

Probability of the defective items produced by machine 1

$$P\left(\frac{A_1}{B}\right) = \frac{P(A_1) \cdot P(B/A_1)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2)}$$

$$= \frac{0.3 \times 0.05}{0.3 \times 0.05 + 0.7 \times 0.01}$$

$$= \frac{0.015}{0.015 + 0.007}$$

$$= \frac{0.015}{0.022}$$

$P(A1/B) = 0.68$

Probability of defective items produced by machine 2

$$P\left(\frac{A_2}{B}\right) = \frac{P(A_2) \cdot P(B/A_2)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2)}$$

$$= \frac{0.7 \times 0.01}{0.3 \times 0.05 + 0.7 \times 0.01}$$

$$= \frac{0.007}{0.015 + 0.07} = \frac{0.007}{0.022} = 0.32$$

$P(A2/B) = 0.32$

Q) In a Bolt factory machine a_1, machine a_2 and machine a_3 manufactures respectively 25%, 35% and 40% of the total of their output 5,4,2 percentages are defective bolts produced by the machines. A bolt is drawn at a random from

the product is found to defective. What is the probability that it was manufactured by machine 3.

Sol: - $P(A_1) = 25\%$

$= 25/100 = 0.25$

$P(A_2) = 35\%$

$= 35/100 = 0.35$

$P(A_3) = 40\%$

$= 40/100 = 0.40$

$P(B/A_1) = 5\%$

$5/100 = 0.05$

$P(B/A_2) = 4\% = 4/100 = 0.04$

$P(B/A_3) = 2\% = 2/100 = 0.02$

$$P\left(\frac{A_3}{B}\right) = \frac{P(A_3) \cdot P(B/A_3)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3)}$$

$$= \frac{0.40 \times 0.02}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= \frac{0.08}{0.0125 + 0.014 + 0.08}$$

$$= \frac{0.008}{0.0345}$$

$= 0.231$

$$P\left(\frac{A_2}{B}\right) = \frac{P(A_2) \cdot P(B/A_2)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3)}$$

$$= \frac{0.35 \times 0.04}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= \frac{0.014}{0.0345}$$

$= 0.4057$

$$P\left(\frac{A_1}{B}\right) = \frac{P(A_1) \cdot P(B/A_1)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3)}$$

$$= \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= \frac{0.0125}{0.0345}$$

$$= 0.3623$$

6. Binomial Distribution:-

The binomial distribution also known as "Bernoulli Distribution" is associated with the name of a Swiss mathematician James Bernoulli also known as Jacques or Jacob (1654-1705). Binomial distribution is a probability distribution expressing the probability of one size of dichotomous alternatives. i.e. success or failure.

The distribution has been used to describe a wide variety of processor in business and the social sciences as well as other areas.

Mathematical Distribution:-

If an Event 'E' has probability 'p' of accounting in each of 'n' independent trails and that of failure in any i.e., q=1-P then the probability that it will occur exactly 'r' times in 'n' trails is given by

$$f(r) = {}^{n}C_{r}\, p^{r}\, q^{n-r}$$

This probability distribution is called the "Binomial probability Distribution "

Where, P= probability of success in a single trait.

q= 1-P; n=no. of trails

r = no. of success of 'n' trails.

Obtaining Coefficients of the Binomial:-

For obtaining coefficients from the binomial expansion the following rules may be remembered.

To find the terms of the expansion of $(q+p)^n$.

1. The first term is q^n.
2. The second term is n_c, q^{n-1} P.
3. In each succeeding term the power of q is reduced by '1' and the power of a 'p' is increased by 1.
4. The co-efficient of any term is found by multiplying the coefficient of the preceding term by the power of 'q' in that preceding term, and dividing the products so obtained by one more than the power of p in that preceding term. When we expand $(q+p)^n$, we get

$(q+p)^n = q^n + {}^nC_1 q^{n-1} p + {}^nC_2 q^{n-2} p^2 + \cdots + {}^nC_r q^{n-r} p^r + \cdots + p^n$ where $1, {}^nC_1, {}^nC_2 \cdots$ are called binomial coefficient

Properties of Binomial distribution:-

1. The shape and location of binomial distribution changes as a 'p' changes for a given 'n' or as 'n' changes for a given 'p'. As 'p' increase for a fixed 'n', the binomial distribution shifts to the right.
2. The mode of the binomial distribution is equal to the value of x which has the largest probability.
3. As 'n' increases for a fixed 'p', the binomial distribution moves to the right, hattens & spreads out. The means of the binomial distribution 'np', obviously increases as 'n' increases with 'p' held constant. For large 'n' there are more possible outcomes of a binomial experiment and the probability associates with any particular outcome become smaller.

4. If 'n' is large and if neither 'p' nor 'q' is too close to zero. The binomial distribution can be closely approximated by a normal distribution with standardized variable given by

$$Z = \frac{x - np}{\sqrt{npq}}$$

.The approximation becomes better with increase 'n'.

Importance :-

The binomial probability distribution is a discrete probability distribution that useful in describing an enormous variety of real life events.

The binomial distribution can be used when :-

1. The outcome of results of each trail in the processor is characterized as one of two types of possible outcomes. In other words they are attributes.
2. The possibility of outcomes of any trail does not change and is independent of the results of previous trails.

1. A fair coin is cost thrice (3times) find the probability of getting.

   i Exactly 2 heads

   ii At least heads

Sol: Binomial distribution

$$P(r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

P=1/2 i.e probability of a getting a success case.

q= 1-P = 1-1/2 =1/2

i Exactly 2 heads, r=2+1

$$P(r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

$$P(2H) = {}^{3}C_{2} \left(\frac{1}{2}\right)^{2} \left(\frac{1}{2}\right)^{1}$$

$$= \frac{3\times 2}{1\times 2} \left(\frac{1}{2}\right)^{2} \left(\frac{1}{2}\right)^{1}$$

$$= 3\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = 3/8.$$

ii At least 2 heads $r = (2H, 3H)$

$$P(3H) = {}^{3}C_{3} \left(\frac{1}{2}\right)^{3} \left(\frac{1}{2}\right)^{3-3}$$

$$= \frac{3\times 2\times 1}{1\times 2\times 3} \left(\frac{1}{2}\right)^{3} \left(\frac{1}{2}\right)^{0}$$

$$= 1\left(\frac{1}{2}\right)^{3} (1) = 1/8 \quad \therefore 2H + 3H = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}.$$

2)4 coins cost simultaneously what is the probability of getting i) No heads ii) No trails iii) 2 heads only (or) exactly 2 heads.

Sol: - i) no heads, r =0

$$P(0) = {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$$

$$= 1 \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^4$$

$$= 1 \times 1 \times \frac{1}{2^4}$$

$$= 1 \times \frac{1}{16} = 0.0625$$

ii) No trails

$$P(0) = {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$$

$$= 1 \cdot \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4$$

$$= 1 \times 1 \times \frac{1}{2^4}$$

$$= 1 \times 1 = \frac{1}{16} = 1/16 = 0.0625$$

iii)2 heads only

$$P(2H) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2}$$

$$= \frac{4 \times 3}{1 \times 2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$$

$$= 6\left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = 6 \times \frac{1}{16} = \frac{6}{16} = 0.375$$

Note:-

Whenever mean standard deviation and variance are given in the binomial distribution we can consider as

Means =np

Standard deviation = $\sqrt{npq}$

Variance = $(\sqrt{npq})^2$

1) The mean of a binomial distribution is 20 and standard deviation is 4. Find n,p and q values.

Sol:- mean, np=20

Standard deviation, $\sqrt{npq}=4$

Variance npq $=(4)^2=16$

Probability = variance/mean

$= 16/20$

$q = 4/5$

$p = 1-q$

$p = 1-4/5$

$p = 1/5$

Substitute P = 1/5 is mean

$np = 20$

$n(1/5) = 20$

$n/5 = 20$

$n = 20x5$

$n = 100$

Standard deviation

$npq = 4$

Squaring on both sides

$npq = 16$

$n\,(1/5)\,(4/5) = 16$

$4n/25 = 16$

$4n = 400$

$n = 100.$

3. The mean of a binomial distribution is 6. and variance is 4. Find n,p,q values.

Sol:- mean, np -6

$npq = 4$

Variance npq =4

Probability = $\frac{\text{variance}}{\text{mean}} = \frac{npq}{np} = \frac{4}{6}$

$q = 2/3$

$P = 1-q$

$P = 1-2/3$

$P = 3-2/3 = 1/3$

Substitute p =1/3 in mean

$np = 6$

$n(1/3) = 6$

$n/3 = 6$

$n = 18.$

3. A dye is thrown 5 times if getting an even no. is a success. What is the probability of getting i)4 success cases ii) at least 4 success cases.

Sol: n= no. of times a dye is thrown =5

P = probability of getting a even no. = $\frac{\text{no. of times then even no. Existed}}{\text{Total no. of cases}}$

$P = 3/6$

$P = ½$

Q = probability of getting a failure case,

$Q = 1 - P$

$= 1 - ½$

$Q = ½$

(i) 4 success cases

$P(r) = p(4) = {}^{5}C_{4}(1/2)^{4}(1/2)^{5-4}$

$= \frac{5x4x3x2}{1x2x3x4}(1/2)^4 (1/2)^1$

$= 5 (1/2)^4 (1/2)$

$= 5 \times 1/16 \times ½$

$= 5/32$

$= 0.156$

(ii) At least 4 success cases

$P(4) = 5/4 (1/2)^4 (1/2)^{5-4}$

$= \frac{5x4x3x2}{1x2x3x4}(1/2)^4 (1/2)^1$

$= 5 (1/16) (1/2)$

$= 5/32$

$= 0.156$

$P(5) = {}^{5}C_{2} \left(\frac{1}{2}\right)^{5} \left(\frac{1}{2}\right)^{5-5}$

$$= \frac{5x4x3x2x1}{1x2x3x4x5} (1/2)^5 (1/2)^0$$

$= 1x1/32 \times 1$

$= 1/32$

$= 0.031$

$P(4) + P(5) = 0.156+0.031 = 0.187.$

Fitting a binomial distribution:-

When a binomial distribution is to be fitted to observe data, the following procedure is adopted.

- Determine the values of p & q. if one of these values is known the other can be found out by the simple relationship p=1-q & q=1-P. when P & q are equal the distribution is symmetrical, for P & q may be interchanged without alternating the value of any terms & consequently terms equidistant from the two ends of the series are equal.
- Expand the binomial $(q+p)^n$. The power 'n' is equal to one less than the number of terms in the expanded binomial thus when two coins are tossed (n=2) there will be three terms in the binomial.
- Multiply each term of the expanded binomial by N (frequency) in order to obtain the expected frequency in each category.

1) 4 coins are tossed 160 times and the following results are obtained.

No. of heads: 0 1 2 3 4

Frequency : 17 52 54 31 6

Fit a binomial distribution under the assumption the coins are unbiased

Sol:- Here N=160

n =4

r = 0, 1,2,3,4 (success cases)

P= ½ q =1-P

Q=1-1/2

No. of Heads	Expected frequently
0	10
1	40
2	60
3	160
4	26.6 (or) 27

Now r = 0

$P(0) = N \times {}^{n}C_{r}\, p^{r} q^{n-r}$

$= 160 \times {}^{4}C_{0} (1/2)^{0} q^{4-0}$

$= 160 \times 1 \times 1 \times (1/2)^{4}$

$= 160 \times 1/16$

$= 10$

$P(1) = N \times {}^{n}C_{r}\, p^{r} q^{n-r}$

$= 160 \times {}^{4}C_{1} \left(\frac{1}{2}\right)^{1} \left(\frac{1}{2}\right)^{4-1}$

$= 160 \times 4 \times (1/2) \times (1/2)^{3}$

$= 160 \times 4 \times 1/2 \times 1/8$

$= 40$

$P(2) = N \times {}^{n}C_{r}\, p^{r} q^{n-r}$

$= 160 \times {}^{4}C_{2} \left(\frac{1}{2}\right)^{2} \left(\frac{1}{2}\right)^{4-2}$

$= 160 \times 4x3/2x1 \ (0.25) \ (1/4) \ (1/2)^{2}$

$= 160 \times 6x \ ¼ \times ¼$

$= 60$

$P(3) = 160 \times {}^{4}C_{3} \left(\frac{1}{2}\right)^{3} \left(\frac{1}{2}\right)^{4-3}$

$= 160 \times \frac{4 \times 3 \times 2}{3 \times 2 \times 1} \left(\frac{1}{8}\right) \left(\frac{1}{2}\right)^{1}$

$= 160 \times 16 \times 1/8 \times ½$

$= 160$

$P(4) = 160 \times {}^{4}C_{4} \left(\frac{1}{2}\right)^{4} \left(\frac{1}{2}\right)^{4-4}$

$= 160 \times \frac{4 \times 3 \times 2 \times 1}{1 \times 2 \times 3 \times 4} \left(\frac{1}{2}\right)^{4} \left(\frac{1}{2}\right)^{0}$

= 160 x 1 x 1/16 x 1

=26.6

1. Fit A Binomial Distribution From The Following Data

X :	0	1	2	3	4
Y :	28	62	46	10	4

SOL :

x:	0	1	2	3	4
f:	28	62	46	10	4
f_x:	0	62	92	30	16 = 200

Mean $\bar{x} = \frac{\Sigma fx}{N}$ = 200/150 = 4/3

We know that, mean np = 4/3, but n=4

4p = 4/3 q = 1-p

P =4/3x4 q = 1-1/3

P = 1/3 q = 2/3

If r=0

$P(0) = N \times {}^{n}C_{r} p^{r} q^{n-r}$

$= 150 \times {}^{4}C_{0} \left(\frac{1}{3}\right)^{0} \left(\frac{2}{3}\right)^{4-0}$

=150x1x1x16/81

=150x16/81

=2400/81 = 29.62

$P(1) = N \times {}^{n}C_{r} p^{r} q^{n-r}$

$= 150 \times {}^{n}C_{1} \times \left(\frac{1}{3}\right)^{1} \left(\frac{2}{3}\right)^{4-1}$

$=150x4/1X(1/3)X(2/3)^3$

=150X4X1/3X8/27 =59.25

P (2) = $150 \times {}^4C_2 \times \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{4-2}$

=150x4X3/2X1 X 1/9 X $(2/3)^2$

=150X12/2X1/9X4/9

=150X6X1/9X4/9

=44.44

P (3) = $150 \times {}^4C_3 \times \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{4-3}$

=150x4X3X2/3X2X1 X $(1/3)^3$ $(2/3)^1$

=150X 24/6 X 1/27 (2/3)

=150X4X1/27X 2/3

=150X4X1/27X2/3

=14.81

P (4) = $150 \times {}^4C_4 \times \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^{4-4}$

=150x4X3X2X1/1X2X3X4 $(1/3)^4$ $(2/3)^0$

=150X 24/24 $(1/3)^4$ $(2/3)^0$

=150X1X1/81X1

=1.851

7. Poisson Distribution:-

Poisson distribution is a discrete probability distribution and is very widely used in statistical work. It was developed by French mathematician simeon Denis Poisson (1781-1840) in 1837.

Poisson distribution may be expected in cases where the chance of any individual event being a success is small. The distribution is used to describe the behavior of rare events such as the no. of accidents on road, no. of printing mistakes in a book etc. and has been called "The law of improbable events"

7.1 Mathematical Definition:-

The Poisson distribution

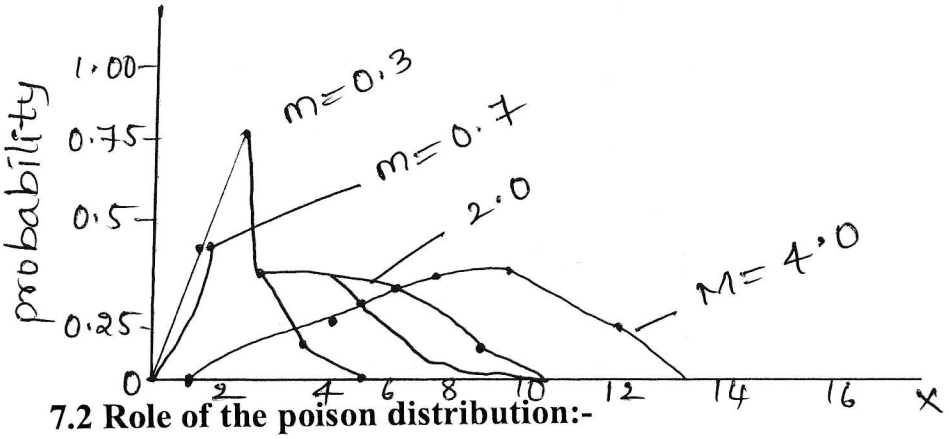
$$P(r) = \frac{e^{-m} m^{r}}{r!}$$

Where r=0, 1, 2, 3, 4.....

E = 2.7183 (base of natural logarithms)

m = the mean of the Poisson distribution

The poison distribution is a discrete distribution with a single parameter 'm'. As 'm' increases the distribution shifts to the right.



7.2 Role of the poison distribution:-

- It is used in quality control statistics to count the no. of defects of an item.
- In biology to count the no. of bacteria.
- In physics to count the no. of practices emitted from a radioactive substance.
- In insurance problems to count the no. of causalities.
- In waiting time problems to count the no. of incoming telephones calls or incoming customers.
- No. of traffic arrivals such as trucks at terminals, aero planes at airports, ships and so forth.
- In determining the no. of deaths in a distinct in a given period, say, a year, by a rare disease.
- The no. of typographical errors per page in typed material ; no. of deaths as a result of good accidents etc;
- In problems dealing with the inspection of manufactured products with the probability that any one piece is defective is very small and the lots are very large arc.

- To model the distribution of the no. of persons joining a queue to receive a service or purchase of a product.

7.3 Characteristics of poison Distribution:-

Discrete Distribution:-

Like binomial distribution it is also a discrete probability i.e., occurrence can be described by a random variable.

Main parameter:-The main parameter is mean (m) which is equal to np i.e. m=np

Form: It is a positively skewed distribution.

No upper limit:- There is no upper limit with the No. of occurrences of an event during a specified time periods.

Properties:-

1. The experiments results in outcomes that can be classified as successor or failures.
2. The average no of success (m) that occur in a specified region is known.
3. The Probability that a success will occur is proportional to the size of the region
4. The probability that a success will occur is an extremely small region is virtually zero.
5. It is discrete probability distribution where the random variable x assumes the infinite set of values 0, 1,2 .....
6. Mean = m= parameter of the distribution, variance ( )=m, s.w ( ) = m, skewers = 1/m & kurtosis =1/m.
7. The mode of Poisson distribution is that value x which occurs with largest probability. It may have either one or two models. If 'm' is not an integer; the mode is the integral value b/w m-1& m.If however m is an integer, then there are two modes which are m-1 and m.
8. It X & Y be two independent Poisson varieties with parameters m_1& m_2 respectively, then their sum x+y is also a Poisson variate with parameter m_1+m_2.
9. The first, second and third new movements are respectively m, m^2+m,m^3+3m^2+m.

Q)It is given that 2% of screws manufactured by a company are defective use Poisson distribution to find the probability that a packet contains 100 screws.

i) No. defective items (or) screws
ii) One defective screws
iii) Two (or) more defective screws.

Sol:- P = probability of getting the defective items =2%

=2/100 =0.02
q = 1-P
q = 1-0.02
q = 0.98

Mean = np here n=100

=100 x 0.02

Mean = 2

$$P(r) = \frac{e^{-m} m^{r}}{r!}$$

i) No defective items (r=0) :-
P (0) = $e^{-2} \cdot 2^{0} / 0!$
= 0.135X1/1!
=0.135

ii) No defective screws:-
P(1) = $e^{-2} \cdot 2! / 1!$
= 0.135x2/1
=0.270

iii) Two or more defective items:- (r=2)
= $1 - [P(0) + P(1)] \Rightarrow 1 - [0.135 + 0.27]$
=1-0.405
=0.595

Q)Suppose on an average one house in 1000 in certain district has a fire during a year. If there are 2000 houses in the district what is the probability that exactly 5 house will have a fire during the year.

Sol: - Total no. of houses in a district n =2000

P = Probability of getting 1 house in 1000 house in the fire accidental during a year 1/1000

Mean = np

=2000x 1/1000

Mean = 2

Poisson distribution, $P(r) = \frac{e^{-m} m^r}{r!}$

i) Probability of getting exactly 5 houses in a fire accident during a year .

r =5

$$P(5) = \frac{e^{-2} (2)^5}{5!}$$

$$= \frac{0.135\ (32)}{5x4x3x2x1}$$

$$= 4.32/120$$

$$P(5) = 0.036$$

Putting a poison distribution:-

The process of fitting a poison distribution is very simple. We have just obtained the value of 'm'. i.e., the average occurrence and calculate the frequency of 'o' success. The other frequency can be very easily calculated as follows.

$N(P_o) = Ne^{-m}$

$N(P_1) = N(P_o) \times m/1$

$N(P_2) = N(P_1) = m/2$

$N(P_3) = N(P_2) = m/3$, Etc.

1. The following mistakes for a page were observed in a book. No. of mistakes per page.

No. of mistakes per page (x)	:	0	1	2	3	4
No. of times the mistakes occur (f) :		211	90	19	5	0

Here N = 325 (2110 +90+19+5)

F_x = 0 90 38 15 0

$\sum f_x$ = 143

$$\text{Mean, } M = \frac{\sum fx}{N} = \frac{143}{325} = 0.44$$

$$e^{-m} = e^{-0.44}$$
$$= 0.644$$

NP (o) = Nxe^{-m}
= 325x0.644
= 09.3

NP (1) = NP (0) x m/1
= 209.3 x0.44/1
= 92.09

NP (2) = NP (1) x m/2
= 92.09 x m/2 (0.44)
= 92.09 x 0.22
= 20.25

NP (3) = NP (2) x m/3
= 20.25 x 0.44/3
= 20.25 x 0.146
= 2.9

NP (4) = NP (3) x m/4
= 2.9 x 0.44/4
= 2.9 x 0.11
= 0.319

Assumed (or) Success cases	Excepted cases
0	209.3
1	92.09
2	20.25
3	2.9

4 0.3

324.9

S 325

2. The no. of Defects per unit in a sample of 330 units of manufacturing product was found by the following

No. of Sackets	:	0	1	2	3	4
No. of units	:	214	92	20	3	1

Fit a poison Distribution to the data under the test for goodness

Sol: - $\frac{\sum fx}{N} = \frac{145}{330} = 0.439$

$NP(0) = N \times e^{-m}$

$= 330 \times e^{-0.439}$

$= 330 \times 0.6447$

$= 212.75$

NP (1) $= NP(0) \times m/1$

$= 212.75 \times 0.439/1$

$= 212.75 \times 0.439$

$= 93.39$

NP (2) $= NP(1) \times m/2$

$= 0.439/2$

$= 0.2195$

$= 93.39 \times 0.2195$

$= 20.499$

NP (3) $= NP(2) \times m/3$

$= 20.499 \times 0.439/3$

$= 20.499 \times 0.146$

$= 2.992$

NP (4) $= NP(3) \times m/4$

$= 2.992 \times 0.439/4$

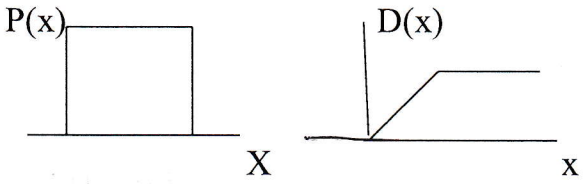
$= 2.992 \times 0.109$

$= 0.326$

Assumed (or) Success cases	Excepted cases
0	212.75
1	93.39
2	20.499
3	2.992
4	0.326
	329.957
Σ	330

8. Uniform Distribution:-

A uniform distribution sometimes also known as a recentagular distribution is a distribution that has constant probability.



The probability density function and cumulative distribution function for a continuous uniform distribution on the internal (a,b) are

$$P(x) = \begin{cases} 0 & \text{for } x<a \\ 1/b-a & \text{for } a \le x \le b \end{cases} \quad \text{------ 1}$$

$$D(x) = \begin{cases} 0 & \text{for } x<a \\ \frac{x_2-x_1}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x>b \end{cases} \quad \text{------ 2}$$

Mean and S.D of a union Distribution :-

$$\text{Mean } \mu = \frac{a+b}{2}$$

$$\text{S.D } (\sigma) = \sqrt{\frac{b-a}{12}}$$

Probabilities in a uniform Distribution:-

The following equation is used to deterring the probabilities of 'x' for a uniform distribution b/w a & b

$$P(x) = \frac{x_2 - x_1}{b-a} \quad ; a \le x_1 \le x_2 \le b$$

9. Normal Distribution:-

The normal distribution was first described by Abraham Demoire as the limiting form of the binomial model in 1733 .Normal distribution was rediscovered by gauss in 1809 & by replace in 1812.

The normal distribution also called the normal probability distribution.

Mathematical definition:-

$$\text{The normal distribution } p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)}{2\sigma^2}}$$

X = value of the continuous random variable

μ= mean of the normal random variable

e = mathematical constant approximated by 2.7183

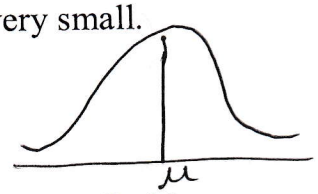
π = mathematical constant approximated by 3.1416

$\sqrt{2\pi} = 2.5066$)

Graph of Normal Distribution:-

- The normal distribution can have different shapes depending on different values of M & but there is one & only normal distribution for any given pair of values for μ & σ

- Normal distribution is a limiting case of binomial distribution when i) n ii)neither p or q is very small.



- Normal distribution is a limiting case of poisson distribution when its means m is large.
- The mean informally distributed population lies at the centre of its normal curve.
- The two tails of the normal probability distribution extent infinitely and never too the horizontal axis.

Importance:-

- The normal distribution has the remarkable property stated in the so called control limit theorem.
- According to this theorem as the sample size 'n' increase the distribution of mean, n of a random sample taken from practically any population approaches a normal distribution.
- As 'n' becomes large the normal distribution saves as a good approximation of many discrete distributions.
- In theoretical statistics many problems can be solved.
- The normal distribution has numerous mathematical properties which make it popular and comparatively easy to manipulate.
- The normal distribution is used extensively in statistical quality control in industry in setting up of control limits.

Significance :-

- The approximate of fit a distribution of measurement under certain conditions.
- The approximate the binomial distribution and other discrete of continuous probability distributions under suitable conditions.
- The approximate the distribution of means & certain other quantities calculated from samples, especially large samples.

Properties:-

- The normal curve is 'bell –shaped' & symmetrical in its appearance. If the curves were folded along its vertical axis, the two halves would coincide.
- The height of the normal curve is at its maximum at the mean.
- There is one maximum point of the normal curve which occurs at the mean. The height of the curve declines as we go in either direction form the mean.
- Since there is only one maximum point, the normal curve is unimodel, i.e., it has only one mode.
- The points of infection. i.e., the points where the change in curvature occur are
- As distinguished from binomial and passion distributed where the variable discrete. The variable distributed a/c to the normal curve is a continuous one.
- The 1st & 3rd variables are equidistant from the median
- The mean deviation is 4th or more preciously 0.7979 of the S.D
- The area under the normal curve distributed as follows.
- Mean $\pm$ σcovers 68.27%, area -34.135 % area will lie on either side of the mean.
- Mean $\pm 2\sigma$covers 95.45 % area.
- Mean $\pm 3\sigma$covers 97.73 % area.

Prepared By
M.NAVANEETH KUMAR REDDY
B-Tech, MBA
ASSISSTANT PROFESSOR
BALAJI INSTITUTE OF IT AND MANAGEMENT

(17E00105) STATISTICS FOR MANAGERS

The objective of this course is to familiarize the students with the statistical techniques popularly used in managerial decision making. It also aims at developing the computational skill of the students relevant for statistical analysis.

1.Introduction of statistics – Nature & Significance of Statistics to Business, , Measures of Central Tendency- Arithmetic – Weighted mean – Median, Mode – Geometric mean and Harmonic mean – Measures of Dispersion, range, quartile deviation, mean deviation, standard deviation, coefficient of variation – Application of measures of central tendency and dispersion for business decision making.

2. Correlation: Introduction, Significance and types of correlation – Measures of correlation – Co-efficient of correlation. Regression analysis – Meaning and utility of regression analysis – Comparison between correlation and regression – Properties of regression coefficients-Rank Correlation.

3. Probability – Meaning and definition of probability – Significance of probability in business application – Theory of probability –Addition and multiplication – Conditional laws of probability – Binominal – Poisson – Uniform – Normal and exponential distributions.

4. Testing of Hypothesis- Hypothesis testing: One sample and Two sample tests for means and proportions of large samples (z-test), One sample and Two sample tests for means of small samples (t-test), F-test for two sample standard deviations. ANOVA one and two way .

5. Non-Parametric Methods: Chi-square test for single sample standard deviation. Chi-square tests for independence of attributes - Sign test for paired data.

Textbooks:

- Statistical Methods, Gupta S.P., S.Chand. Publications

References:

- Statistics for Management, Richard I Levin, David S.Rubin, Pearson,
- Business Statistics, J.K.Sharma, Vikas house publications house Pvt Ltd
- Complete Business Statistics, Amir D. Aezel, Jayavel, TMH,
- Statistics for Management, P.N.Arora, S.Arora, S.Chand
- Statistics for Management , Lerin, Pearson Company, New Delhi.
- Business Statistics for Contemporary decision making, Black Ken, New age publishers.
- Business Statistics, Gupta S.C & Indra Gupta, Himalaya Publishing House, Mumbai

UNIT –4
TESTING OF HYPOTHESIS

1. HYPOTHESIS TESTING: ONE SAMPLE AND TWO SAMPLE TESTS FOR MEANS AND PROPORTIONS OF LARGE SAMPLES (Z-TEST):

Introduction :-The term hypothesis derives from the Greek ' Hypotithenai' meaning " to put under " or " to suppose".

Hypothesis is a tentative conjecture explaining an observation, phenomenon, or scientific problem that can be tested by further observation, investigation, and / or experimentation.

According to prof. Morris Hamburg, A Hypothesis in statistics is simply a quantitative statement about population.

Statistical Hypothesis: - A statement about population in terms of population parameter is known as a statistical hypothesis and denoted by 'H'.

Test of Hypothesis: - A test of a hypothesis is a two action decision problem after the experimental sample values have been obtained, the two actions being the acceptance or rejection of the hypothesis under consideration.

Null Hypothesis:-It is a statement which is believed to be true or it is used as a basis for argument but has not been proved it is denoted by 'Ho'.

Alternative Hypothesis:-It is a statement of what a statistical hypothesis test is set up to established. It is denoted by 'H_1'.

Procedure for testing of hypothesis:-

The following are various steps in testing a statistical hypothesis.

1. Assume Null hypothesis : H_o
2. Alternative hypothesis H_1, helps us to decide whether we have to use as single-tailed or two tailed test.
3. Level of significance:-Choose appropriate level of significance (α) depending on the permissible risk α is fixed in advance before sample is drawn.
4. Test statistic :- Compute the test statistic,

$$Z = \frac{t - E(t)}{S\,\varepsilon(t)} \quad N(0,1)$$

5. Inference:-We compare the computed value of Z in step (4) with significant value (tabulated value) $Z\alpha$ = at the given level of significance 'α'.

If $|Z| < Z\alpha$ we can say it is not significant i.e., the sample data do not provide us sufficient evidence against null hypothesis when may be accepted.

If $|Z| > Z_\alpha$, if the computed value of test statistics is more than the critical or significant value, then we say the null hypothesis is rejected.

Advantages:-

- Determine the focus of direction for a research effort.
- Development of a hypothesis forces the researcher to clearly state the purpose of the research activity.
- Determine what variables will not be considered in a study, as well as those that will be considered.

Disadvantages:-

- This type of tests should not be used in a mechanical fashion.
- This test do not explain the reason as to why does difference – exist.
- Statistical inferences based on the significance tests can't be said to be entirely correct evidences concerning the truth of the hypothesis.

Significance test for single proportion:-Since Sample Size 'n' is large and 'x' is number of successes in 'n' independent trails with constant probability 'p' of success for each trail

E (x) = np and V (x) = npq; where q = 1-P

It has been proved that if 'n' is large binomial distribution tends to normal distribution.

If sample size 'n' is large (i.e. $n \geq 30$) then the number of persons possessing attribute called proportion of success.

$$P = x/n$$

$$E(P) = E(x/n)$$

$$= 1/n\, E(x)$$

$$= 1/n.\, np$$

$$= P$$

Thus the sample proportion p is unbiased estimate of population proportion 'P'

$$\text{Also } V(P) = V(x/n)$$

$$= 1/n^v \,.\, V(x)$$

$$= \frac{npQ}{n^v}$$

$$= \frac{PQ}{n}$$

Standard Error $S.E(P) = \sqrt{\frac{PQ}{n}}$

then $Z = \frac{P - E(p)}{S.E(P)}$

$$Z = \frac{P - P}{\sqrt{\frac{PQ}{n}}}$$

Note :-The limit for P at level of significance α are given by $P \pm Z_\alpha \sqrt{PQ/n}$

Q)In a sample of 1000 people in Karnataka 540 are rice eaters and rest are wheat eaters can we assume both rice and wheat eaters are equally popular in this state at 1% level of significance ?

Sol:- Given sample n = 1000

Let no. of rice eaters X = 540

Proportion of rice eaters P = x/n

= 540/1000

P = 0.54

Null Hypothesis, Ho: - Both rice and wheat eaters are equally popular in the state

Ho: P =0.5

P = 0.5 and Q = 1-P = 0.5

Alternative Hypothesis, H1: - $P \neq 0.5$

Test static: - under Ho test statistic is given by

$$Z = \frac{P - P}{\sqrt{\frac{PQ}{n}}}$$

$$Z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = \frac{0.04}{0.0138}$$

$Z = 2.532$

Significant value at 1% level for two tailed test is 2.58

Conclusion:-Calculated value is less than significant value at 1% level of significance.
Hence accept null hypothesis.

Q)A random sample of 700 units from a large consignment showed that 200 were damaged. Find i) 95 % ii) 99 % confidence limits for the proportion of damaged units in the consignment.
Sol:- Given random sample n = 700, X =200
Proportion of damaged units $P = x/n = 200/700 = 0.286$
$q = 1-P = 1-0.286 = 0.714$

Hence Standard error SE (p) is given

$$SE(P) = \sqrt{pq/n}$$
$$= \sqrt{\frac{0.286 \times 0.714}{700}}$$
$$= 0.017$$

i) 95 % confidence limits for P are given by $P + Z_\alpha\sqrt{pq/n}$
5% loss significant value is 1.96 (Z_α)

$$= P + 1.96\sqrt{pq/n} = 0.286 + 1.96 \times 0.017$$
$$= 0.286 + 0.033$$
$$= (0.253, 319)$$

ii) 99 % confidence limits for P are given by $P + Z_\alpha\sqrt{pq/n}$
1 % loss significant value is 2.58 (Z_α)

$$= P + 2.58\sqrt{pq/n} = 0.286 + 2.58 \times 0.017$$
$$= 0.286 + 0.044$$
$$= (0.242, 0.33)$$

Applications of Z –test:-

- Hypothesis testing for one mean of one sample
- Hypothesis testing for difference between means of two samples.
- Hypothesis testing for one proportion of one sample
- Hypothesis testing for two proportions of two sample
- Hypothesis testing for two Standard deviation of two samples.

Significance test for difference of proportions :-Since sample sizes n_1 and n_2 are large with x_1 and x_2 individuals possessing attributes we have

$$P_1 = x_1/n_1 \qquad P_2 = x_2/n_2$$

If P_1 and P_2 are population proportions,

$E(P_1) = P_1$ $\quad E(P_2) = P_2$

$V(P_1) = \frac{P_1 Q_1}{n_1}$, $V(p2) = \frac{P_2 Q_2}{n2}$

Under Ho = $P_1 = P_2 = P$, $Q_1 = Q_2 = Q$ then the test statistic will becomes

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Q)Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour the proposal is same against that they are not at 5 % loss.

Given, data $n_1 = 400$, $x_1 = 200$

$$P_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5$$

$n_2 = 600$, $x_2 = 325 \Rightarrow P_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.54$

Null Hypothesis, $H_0 \Rightarrow P_1 = P_2 = P$.

Assumption of null hypothesis is there is no significant difference between the opinion of men and women as per as proposal of flyover.

Alternative Hypothesis, H1 ; P1 ≠P2

Test statistics:-Since Samples are large. The test statistic under Ho is

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

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

$$= \frac{400 \times 0.5 + 600 \times 0.54}{400 + 600}$$

$$= 0.524$$

$Q = 1 - P = 1 - 0.524 = 0.476$ $\quad |Z| = \frac{|0.5 - 0.54|}{\sqrt{0.524 \times 0.476\left(\frac{1}{400} + \frac{1}{600}\right)}}$

$$|Z| = \frac{0.04}{\sqrt{0.524 \times 0.476\left(\frac{10}{24000}\right)}}$$

Conclusion:-Since Z = 1.269 which is less than 1.96 Significant value at 5 % loss

Hence Ho may be accepted

Q) In a survey 800 persons out of 1000 are found tea drinkers before increase excise duty. After increase excise duty 800 persons tea drinkers out of 1200. Using standard error of proportion, state whether there is a significant decrease in the consumption of tea after the increase of excise duty?

Sol: Given data $n_1 = 1000$ $n_2 = 1200$

$x_1 = 800$ $x_2 = 800$

$$P_1 = \frac{800}{1000} = 0.8, \quad P_2 = \frac{800}{1200} = 0.67$$

Null hypothesis, Ho : $P_1 = P_2$

Assume that there is no significant different in the consumption of tea before and after increase in excise duty.

Alternate hypothesis: H1; $P1 \neq P2$

Test statistics is given by under Ho is

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{16}{22}, \quad Q = 1 - P = 1 - \frac{16}{22} = \frac{6}{22}$$

$$\therefore Z = \frac{0.8 - 0.67}{\sqrt{16/22 \times 6/22\,(1/1000 + 1/1200)}} = \frac{0.13}{0.019} = 6.842$$

Conclusion:-Alternative 5 % loss 1.96 we found evidence against Ho : Hence we reject Ho.

Testing for means:-

In this section we will discuss the sampling of variables for example height, weight, income, age of a group of persons. These sampling variables each number of population provides the value of the variable.

Test of Significance for single mean :- If x_i, i= 1,2,3 .....n is a random sample of size 'n' from a normal population with mean 'μ' and variance then the

sample mean is distributed normally with mean μ and variance ; however this result hold even in a random sampling from non – normal population provided the same size 'n' large.

Thus for large samples, the standard normal variate corresponding to $\bar{x}$ is

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

In a random sampling from a large population if sampling from finite population with size N, the corresponding limits are

$$\bar{x} \pm 1.96 \sqrt{\frac{N-n}{N-1}} \times \frac{\sigma}{\sqrt{n}} \quad \text{and}$$

$$\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \text{are 95\% \& 99\% confidence limits.}$$

Q)A sample of 400 male students is found to have a mean height of 67.47 inches can it be reasonably regarded as a sample from a large population, with mean height 67.39 inches and standard deviation 1.3 inches (α= 5 % loss)

Sol :- $n = 400, \sigma = 1.3, \mu = 67.3, \bar{x} = 67.47$

under null Hypothesis $H_0 : \mu = 67.39$

Alternative Hypothesis $H_1 : \mu > 67.39$

Test statistic is given by,

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{67.47 - 67.39}{1.3 / \sqrt{400}} = 1.23.$$

Conclusion:-We have found evidence against null hypothesis Ho. So it can be reasonably regarded that the given sample is from the said population at 5 %.

Q)A random sample of 100 articles selected from a batch of 2000 articles shows that the average diameter of the article 0.354 with standard deviation is 0.048. find 95% confidence intervals for the average of this batch of 2000 articles ?

Given n =100, N =2000, x = 0.354

Standard deviation = 0.048

Standard error $SE(\bar{x}) = \sqrt{\frac{N-n}{N-1}} \times \frac{\sigma}{\sqrt{n}}$

$$= \sqrt{\frac{2000-100}{2000-1}} \times \frac{0.048}{\sqrt{100}}$$

$$SE(\bar{x}) = 0.00468$$

95 % confidence limits for the μ are given by

$$\bar{x} \pm 1.96 \sqrt{\frac{N-n}{N-1}} \times \frac{\sigma}{\sqrt{n}}$$

$$= 0.354 \pm 1.96\,(0.00468)$$

$$= (0.3448,\ 0.3632)$$

Test of Significance for difference of means :-Let x_1 be the mean of a random sample of size n, from a population with mean μ, and variance and be the mean of a random sample of size n2 from a population mean μ_2 and variance . Then sample sizes n_1 and n_2 are large

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

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

Q) In a sample of 500, then mean is found to be 20. In another sample of 400, the mean is 15. Is the two samples drawn independently from same population with Sd 4 ?

Sol :- $n_1 = 500,\ n_2 = 400,\ \bar{x} = 20,\ \bar{x}_2 = 15,\ \sigma = 4$.

under null Hypothesis, $H_0 : \mu_1 = \mu_2$.

Alternative Hypothesis, $H_1 : \mu_1 \neq \mu_2$.

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

$$z = \frac{20 - 15}{4\sqrt{1/500 + 1/400}}$$

$$= 5 / 0.018 = 277.77$$

∴ Reject H_0.

2. ONE SAMPLE AND TWO SAMPLE TESTS FOR MEANS OF SMALL SAMPLES (T-TEST):

Assumptions for students t –test :-The following assumptions are made in the students' t –test.

- The present population from which the sample drawn is normal

- The population, observations are independence, i.e., the given sample is random
- The standard sample deviation is unknown.

Applications of t-distribution:-The t-distribution has a number of applications in statistics, of which we shall discuss some of them

- T-test for significance of single mean, population variance being unknown
- T-test for significance of different between two sample means, the population variances being equal but unknown
- T-test for significance of an observed sample correlation co-efficient

T-Test:-

The greatest contribution to the theory of small samples was made by " Sir William Sealy Gossett".

Gossett published his discovery in 1905 under the pen name 'student' and it is popularly known as t-test or student t – distribution or student's distribution.

Student's t :-

If $x_1, x_2 \ldots x_n$ is a random sample of size 'n' from a normal population with mean 'μ' and variance 'σ^2' the students t – statistic is defined as

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}}, \quad \bar{x} = \frac{\Sigma x}{n}$$

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

Test for single mean:-

Q) A machine is designed to produce insulating warners for electrical devices of an average thickness of 0.025 cm. A random sample of 10 warners was found to have an average thickness of 0.024 cm with a standard deviation of 0.02 cm. Test the Significance of the deviation.

A) We are given n = 10, x = 0.024 cm S = 0.002 Cm μ = 0.025 cm

Null hypothesis:-

Ho: μ = 0.025 cm, i.e., there is no significant deviation between sample mean $\bar{x}$=0.024 and population when μ= 0.025

Alternative hypothesis:- $H_1 ; \mu \neq 0.025$ cm

under Ho', The test statistic is,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \Rightarrow \frac{0.024 - 0.025}{0.002/\sqrt{10-1}} \Rightarrow \frac{-0.001 \times 3}{0.002}$$

Tabulated value of $t_{0.05}$ for a degree of freedom = 1.833

Since $|t| < 1.833$ is not significant between sample mean and population mean (is no/ significant).

Q) Certain pesticide is packed in to bags by a machine. A random sample of 10 bags is drawn and their contents are found to weight as follows.
50, 49, 52, 44, 45, 48, 46, 45, 49, 45 test if the average packing can be taken to be 50 kg.

Sol: - Null hypothesis : Ho = μ= 50 kgs i.e., the average packing is 50 kgs.

Alternative hypothesis: H_1: $\mu \neq 50$ kgs

x	$x(x-\bar{x})$	x^2
50	2.7	7.29
49	1.7	2.89
52	4.7	22.09
44	-3.3	10.89
45	-2.3	5.29
48	0.7	0.49
46	-1.3	1.69
45	-2.3	5.29
49	1.7	2.89
45	-2.3	5.29
473		64.1

Mean, $\frac{473}{10} \Rightarrow \bar{x} = 47.3$

Standard deviation, $= \sqrt{\frac{\Sigma x^2}{n}}$

Variance, $(s^2) = \frac{\Sigma x^2}{n} = \frac{64.1}{10}$

$s^2 = 6.41$

Test statistic, $t = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}}$

$= \frac{47.3 - 50}{\sqrt{6.41/9}}$

$= \frac{-2.7}{\sqrt{0.712}} \Rightarrow$

$\Leftarrow \frac{-2.7}{0.8438}$

$= -3.2$

$\therefore$ T.V of $t_{0.05}$ for dom $= 1.833$

$|t|$ is $> t$, Ho is Rejected.

Q)A random samples of 10 boys had the following IQ's 70,120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean IQ of 100 (Ans: 0.62)

Null Hypothesis: Ho : $\mu = 100$. i.e, The Assumption of a population of IQ is 100.

Mean, $= \frac{972}{10}$, $\bar{x} = 97.2$

x	$(x-\bar{x})$	x^2
70	−27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	−9.2	84.64
83	−14.2	201.64
95	−2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
		1,833.6

$$S.D = \sqrt{\frac{\Sigma x^2}{n}} = \frac{1833.6}{10}$$

$$S^2 = 183.36$$

Test statistic, $t = \frac{\bar{x}-\mu}{\sqrt{s^2/n-1}} = \frac{97.2-100}{\sqrt{183.36/9}} = \frac{-2.8}{\sqrt{20.373}} = -2.8/4.513.$

$t = -0.62 \Rightarrow |t| = 0.62$

T.V $t_{0.05}$ 9 dom 1.833. $|t| > t$,

Hence Ho is Rejected.

Tabulated value of $t_{0.05}$ for 9 degrees of freedom 1.833 since calculated t is greater than tabulated t. it is significant Hence Ho is rejected

T- test for difference of means :-

Suppose we want to test if two independent samples have been drawn from the two normal populations having the same means.

Let $x_1, x_2 \ldots\ldots x_{n1}$ and $y_1, y_2 \ldots\ldots y_{n2}$ be two independent random samples from the given normal populations.

we set up the null hypothesis Ho $=\mu_x = \mu_y$ under the Ho the test statistic is

$$|t| = \left| \frac{\bar{x}-\bar{y}}{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \right| \text{ is } t_{n_1+n_2} = 2$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

Where, $\bar{x} = \frac{\Sigma x}{n_1}$, $\bar{y} = \frac{\Sigma y}{n_2}$

$$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1}$$

$$s_2^2 = \Sigma(y-\bar{y})^2/n_2$$

Q) The average number of articles produced by two machines per day are 200 and 250 with standard deviations 20 and 25 respectively on the basis of records of 25 days production. Can you regard both the machine equally efficient at 5 % level of significance?

Sol:- In the usual notations we are given

$n_1 = n_2 = 25$, x =200, y = 250. S_1 =20, S_2 =25

Null hypothesis: Ho = $\mu_1 = \mu_2$ i.e., both the machines are equally efficient.

Alternative hypothesis : $H_1: \mu_1 \neq \mu_2$.

under the Ho the test statistic is, $t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{x_1}+\frac{1}{x_2}\right)}}$

Where, $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} \Rightarrow t = \frac{200-250}{\sqrt{533.85\left(\frac{1}{25}+\frac{1}{25}\right)}}$

$$= \frac{25\times400 + 25\times625}{25+25-2}$$

$$= \frac{25625}{48} = 533.85$$

$$= \frac{-50}{\sqrt{533.85\times0.08}} = \frac{-50}{\sqrt{42.708}} = \frac{-50}{6.535} = -7.65$$

Tabulated $t_{0.05}$ value for 48 = 1.67

Since calculated t > tabulated t, it is highly significant. Hence Ho is rejected and we conclude that both the machine are not equally efficient at 5 % level of significance.

Q) The means of 2 random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from the mean are 26.94 & 18.73 respectively can the samples be considered to have been drawn from the same normal population?

Sol:- In the usual notations we are given

$n_1=9, n_2=7, \bar{x}=196.42, \bar{y}=198.82, \Sigma(x-\bar{x})^2=26.94, \Sigma(y-\bar{y})^2=18.73.$

Null Hypothesis:- The sample have been drawn from the same normal populations. i.e $H_0: \mu_1=\mu_2$.

Alternative Hypothesis:- $H_1: \mu_1 \neq \mu_2$ under the H_0, The test statistic is $t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $s^2 = \frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n_1+n_2-2}$

$$t = \frac{196.42-198.82}{\sqrt{3.26\left(\frac{1}{9}+\frac{1}{7}\right)}}$$

$$s^2 = \frac{26.94+18.73}{9+7-2}$$

$$= \frac{-240}{\sqrt{3.26\times0.254}}$$

$$= \frac{45.67}{14}$$

$$= 3.26$$

$$= \frac{-2.40}{\sqrt{0.828}} = \frac{-2.40}{0.9099} = -2.64$$

T.V is 14 dof is 1.761

$|t| > t$

H_0 is Rejected.

Q)Two different types of drugs A and B were tired on certain patients for increasing weight, 5 persons were given drug A and 7 persons were given drug B. The increase in weight in pounds is given below.

Drug A : 8 12 13 9 3 - -

Drug B : 10 8 12 15 6 8 11

Do the two drugs differ significantly with regard to their effect in increasing weight

(Ans : - 0.501) Given, $n_1 = 5$, $n_2 = 7$. let us consider, Drug A observations are x_1, x_2, x_3, x_4, x_5, Drug B $y_1, y_2, y_3, y_4, y_5, y_6, y_7$

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
8	−1	1	10	0	0
12	3	9	8	−2	4
13	4	16	12	2	4
9	0	0	15	5	25
3	−6	36	6	−4	16
		62	8	−2	4
			11	1	1
					54

$\bar{x} = \frac{45}{5} = 9$, $\bar{y} = \frac{70}{7} = 10$.

$s_1^2 = \frac{62}{5} = 12.4$, $s_2^2 = \frac{54}{7} = 7.71$

$s^2 = \frac{5(12.4) + 7(7.71)}{5+7-2}$

$= \frac{62+54}{12-2} = \frac{116}{10} = 11.6$

$t = \left| \frac{9-10}{11.6\left(\frac{1}{5}+\frac{1}{7}\right)} \right|$

$= \left| \frac{-1}{11.6(0.2+0.14)} \right|$

$\Rightarrow \frac{1}{11.6(0.34)} = \frac{1}{3.9214} \Rightarrow t = 0.25$

Dof $= 5+7-2 = 10$. We can assume that 5% level of significance, T.V value 10 DOF is, 1.812

H_0 is Accepted.

3. F-TEST FOR TWO SAMPLE STANDARD DEVIATIONS:

F- Distribution (f –test) :-

F-Distribution was introduced by G.W Snedecor. The f –test is named in honour of the great satisfaction R.A. fisher.

F- Test for two sample standard deviations:- Let $x_1, x_2, \ldots x_n$ be a random sample of size n_1 from the first population with varience and $y_1, y_2, \ldots y_n$ be a random sample of size n_2 from the second normal population with varience , obviously the two samples are independent. We set up the null hypothesis as

i.e., population variances are same

Under Ho, the test statistic is $F = \frac{s_1^2}{s_2^2} \sim F(n_1-1, n_2-1)$ $s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1}$

$s_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1}$ F-distribution, with (n_1-1, n_2-1) dof

Assumption in F –test: - The F-test is based on the following assumptions

Normality: - values in each group are normally distributed.

Homogenity :- The varience with in each group should be equal for all groups

$(\sigma_1^2 = \sigma_2^2 = \cdots, \sigma_k^2)$

Independence of Error :- it states that the error should be independent for each value.

Applicants of F –test:-

- F-test for testing the significance of an observed sample multiple correlation
- F-test for testing the significance of an observed sample correlation ratio.
- F- test for testing the linearity of regression.
- F- test for testing the equality of several population means, i.e., for testing Ho $=\mu_1 =\mu_2 \ldots.\mu_K$ for K normal populations.

Q)Time taken by workers in performing a job by method 1 and method 2 is given below

Method I	20	16	26	27	23	32 22
Method II	27	33	42	35	32	34 38

Do the data show that the variance of time distribution from population. From which these samples are drawn do not differ significantly?

Sol:- we set up null hypothesis as Ho: = i.e., there is no significant difference b/w the varience of the time distribution by the workers in performing a job method I and Method II

Method–I

x	$x-\bar{x}$	$(x-\bar{x})^2$
20	-2.3	5.29
16	-6.3	39.69
26	3.7	13.69
27	4.7	22.09
23	0.7	0.49
22	-0.3	0.09
134		81.34

$\bar{x} = \frac{134}{6} = 22.3$

$\bar{y} = \frac{241}{7} = 34.4$

Method II

y	$(y-\bar{y})$	$(y-\bar{y})^2$
27	-7.4	54.76
33	-1.4	1.96
42	7.6	57.76
35	0.6	0.36
32	-2.4	5.76
34	-0.4	0.16
38	3.6	12.96
241		133.72

$$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1} = \frac{81.34}{6-1} = \frac{81.34}{5} = 16.26$$

$$s_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1} = \frac{133.72}{7-1} = \frac{133.72}{6} = 22.28$$

Since $s_2^2 > s_1^2$ under Ho The test statistic is, $F = \frac{s_2^2}{s_1^2}$ as $F(n_2-1, n_1-1)$

$F = 22.28 / 16.26 = 1.37$, Tabulated $F_{0.05}(6,5) = 4.95$

Since calculated F is less than tabulated F, it is not significant. Hence Ho maybe accepted at 5% level of significance

Q)It is known that the mean diameters of rivets produced by 2 firms A and B practically the same but the standard deviations may differ. For 22 rivets produced by firm A, the Standard deviation is 2.9mm while for 16 rivets manufactured by firm B, the standard deviation is 3.8 mm. compute the statistic you would use to test whether the products of firms A have the same variability as those of firms B and test its significance.

Given $n_1 = 22$, $n_2 = 16$, $s_1 = 2.9mm$, $s_2 = 3.8mm$.

We set up the null hypothesis as, $H_0: \sigma_1^2 \geq \sigma_2^2$ i.e, the products of both the firms A and firm B. have the same variability.

We have, $s_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ $\quad s_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$

$= \frac{22 \times (2.9)^2}{22-1}$ $\quad = \frac{16 \times (3.8)^2}{16-1}$

$= \frac{22 \times 8.41}{21}$ $= \frac{185.02}{21}$ $\quad = \frac{16 \times 14.44}{15}$

$= 8.810$ $\quad = \frac{231.04}{15}$

$= 15.402$

Since $s_2^2 > s_1^2$ under Ho the test statistic is,

$F = 1.748$, at $(15, 21)$ Then T.V at $0.05 = 2.20.(15,21)$

Since calculated F is less than the tabulated f, it is not significant at 5 % level of significance hence Ho is accepted.

Design of Experiments: - An experimental design is a paln and a structure to test hypothesis is which the recorder either controls or manipulates one or more variables it contains independent and dependent variables.

Independent variables: - work shift, gender of employee, region type of machine, quality of tire.

Dependent variable: - A Dependent variable is the response to the different levels of the independent variables.

Principles of experimental design:-

1) Comparison 2) Randomization 3) Blocking
4) Replication 5) factorial experiments

Procedure in effective design of experiment :-

1. Select problem
2. Determining dependent variables
3. Determining independent variables
4. Determining number of levels of independent variables
5. Determining possible contributions
6. Determining number of observations
7. Randomization
8. Meet ethical and legal requirements
9. Mathematical model
10. Data collection
11. Data reduction
12. Data verification

4. ANOVA ONE AND TWO WAY:

Analysis of Variance (ANOVA):-

- Analysis of Variance was developed by R.A fisher
- Analysis of Variance, the significance of the difference b/w the means of two samples can be judged through either Z- test or t –test, but the difficulty arises when we used ANOVA.
- ANOVA is useful in the fields of economics, biology, education, psychology, sociology, and business and in research of several other disciplines
- ANOVA is essentially a procedure for testing the difference among different groups of data for homogeneity.
- ANOVA is a method of analyzing the variance to which a response is subject into its various components corresponding to various sources of variation.

Assumptions of ANOVA:-

- It is assumed that the universe from which the different samples are drawn for study is normally distributed.

- It is assumed that there is no significant difference amongst the variances of the different universes from which the samples have been drawn.
- It start with null hypothesis that $v_1=v_2= v_3.....v_n$.
- It is assumed that the critical values of the variance ratio (F) is estimated at different levels of significance, Ex : 5% or 1% etc.

Applications of ANOVA:-

- We can explain various varieties of seeds of fertilizers or soils differ significantly so that a policy decision could be taken with help of 'ANOVA'.
- Various types of drugs manufactured for curing a specific disease may be studied and judged.
- A manager of a big concern can analyze the performance of various sales man.

Analysis of Variance for one –way classification:-

Under the one way ANOVA, we consider only one factor. We determine if there are differences within that factor.

The technique involves the following steps:-

- Calculate sum of normal, squares of the individual variables.
- Calculate the sum of individual sum of the variables
  $T = \sum x_1 + \sum x_2 + ...+\sum x_n$
- Calculate the value of connection factor (T^2/N) where N = total no.of variables
- Calculate the value of SST = $\sum x_1^2 + \sum x_2^2+ .....+\sum x_n^2 - T^2/N$ (sum of squares for variance of total)
- Calculate the value of SSB =

  (Sum of squares for variance b/w the samples)
- Find out the value of SSW = SST -SSB (sum of squares for variance b/w with the samples)
- Draw the ANOVA table.
- Finally, F –ratio may be worked out as, F – ratio $= \frac{MSB}{MSW}$

  MSB = means square b/w samples

  MSW= means square with in samples.

Q)Four machines A,B,C, D are used to produce a certain kind of cotton fabrics. Samples of size 4 with each unit as 100 square meters are selected from the outputs of the machines at random, and the number of flows in each 100 square meters are counted, with the following result.

A	B	C	D
8	6	14	20
9	8	12	22
11	10	18	25
12	4	9	23

Do you think that there is a significant difference in the performance of the four machine

Sol:- let us take null hypothesis that the machines do not differ significantly in performance,

i.e., Ho = $\mu_1 = \mu_2 = \mu_3 = \mu_4$

x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2
8	64	6	36	14	196	20	400
9	81	8	64	12	144	22	484
11	121	10	100	18	324	25	625
12	144	4	16	9	81	23	529
40	410	28	216	53	745	90	2038

$T = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \Sigma x_4$
$= 40 + 28 + 53 + 90$
$= 211$

$C.F = \frac{T^2}{N} = \frac{44521}{16}$
$= 2782.56$

Sum of squares for variance of (SST) $= \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2 - \frac{T^2}{N}$
$= 410 + 216 + 745 + 2038 - 2782.56$
$= 3409 - 2782.56 = 626.44$

Sum of squares for variance (SSB) $\Rightarrow \frac{\Sigma x_1^2}{n_1} + \frac{\Sigma x_2^2}{n_2} + \frac{\Sigma x_3^2}{n_3} + \frac{\Sigma x_4^2}{n_4} - \frac{T^2}{N}$
$= \frac{(40)^2}{4} + \frac{(28)^2}{4} + \frac{(53)^2}{4} + \frac{(90)^2}{4} - 2782.56$
$= 400 + 196 + 702.25 + 202.5 - 2782.56$
$= 540.69$

Sum of squares of with Samples (SSW) = SST − SSB.
$= 626.44 - 540.69$
$= 85.75$

N = Total No. of Variables (or) Sample.

K = Number of variables types.

F-Ratio $= \frac{MSB}{MSW} = \frac{180.23}{7.15} = 25.20$

ANOVA Table

S.O.V	S.S	DF	M.Sq
B/W Sam	540.69	3(K-1)	180.23
Within Sam	85.75	12(N-K)	7.15

The table value for $F_{(3,12)}$ at 1% level of significance is 5.95. The calculated value of 'F' is greater than the table value.Hence , we reject the null hypothesis and conclude that there is 9 significant difference in the performance of the four machines.

Q)A random sample is selected from each of three makes of rope and their breaking strength are measured, with the following results:

x_1	x_2	x_3
70	100	60
72	110	65
75	108	57
80	112	84
83	113	87
—	120	73
—	107	—

Test, whether the breaking strength of the ropes differ significantly.

Analysis of Variance for two-way classification:

The way ANOVA techniques is used when the data are classified on the basis of two factors.

For example : the agriculture output may be classified on the basis of different varieties of seeds and also on the basis of different varieties of fertilizers used .

In this 2 way classification two cases are existed

1. ANOVA technique is context of 2-way design when repeated values are not there
2. ANOVA is context of 2 way design when repeated values are there

The following steps are involved

- Use the coding device
- Calculate sum of normal squares of the individual variables
- Calculate sum of normal squares of the individual variables
- Calculate the sum of individual sum of the variables $T= \sum x_1+\sum x_2+\text{-------}+\sum x_n$
- Calculate the value of correction factor (T^2/N) where N =total number of variables
- Calculate the value of $SST= \sum x_1^2+\sum x_2^2+\text{------}+\sum x_n^2-T^2/N$
- Calculate the value of $SSB= (\Sigma x_1)^2/n_1+\frac{(\Sigma x_2)^2}{n_2}+\frac{(\Sigma x_3)^3}{n_2}+\text{---}+\frac{(\Sigma x_n)^2}{n}-\frac{T^2}{N}$
- Find out the value of SSW=SST-SSB

- Take the total of different columns and they obtain the square of each column total and divide such squared values of each column by the number of items in the concerniery column and take the total of the result thus obtained. Finally, subtract the correction factor from this total to obtain the sum of square of deviations for variances between columns (SSC)
- Calculate SSR value
- Find out the value of sum of squares of deviations for residual or error variance (SSE)=SST-(SSC+SSR)
- Draw the ANOVA table
- Test statistic F= $\frac{MSB\ (columns)}{MSR}$, $\frac{MSB\ (rows)}{MSR.}$
- MSR = Mean Square residual

Q)The following table gives the number of refrigerators sold by 4 salesman in 3 months may, june, july.

Sales man

Month	A	B	C	D
May	50	40	48	39
June	46	48	50	45
July	39	44	40	39

Is there a significant difference in the sales made by the four salesman? is their a significant difference in the sales made during different months ?

Sol : let us take the null hypothesis that there is no significant difference between sales made by the four salesman during different months
The given data are coded by subtracting 40 from each observation calculation for a 2-criterion month and sales man

Sales man

Month	A	B	C	D
may	50	40	48	39
June	46	48	50	45
July	39	44	40	39

month	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2	Row sum
	salesman								
may	10	100	0	0	8	64	−1	1	17
june	6	36	8	64	10	100	5	25	19
july	−1	1	4	16	0	0	−1	1	2
	15	137	12	80	18	164	3	27	48

T = sum of all observations = 48. $C.F = \frac{(48)^2}{12} = 192$.

$SSC = (75+48+108+3) - 192 = 42$

$SSR = (72.25 + 210.25 + 1) - 192 = 91.5$

$SST = 137 + 80 + 164 + 27 - 192 = 216$.

$SSE = SST - (SSC + SSR) = 216 - (42 + 91.5) = 82.5$

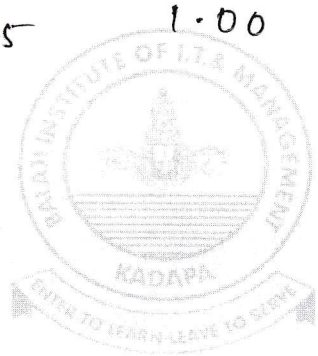
$DOF = C - 1 = 4 - 1 = 3$

$r - 1 = 3 - 1 = 2$

$(C-1)(r-1) = 3 \times 2 = 6$.

Table:—

S.O.V	S.S	dof	MS	V. Ratio
B/W salesman	42	3	14	1.018
B/W mon	91.5	2	45.75	3.327
Resedual error	82.5	6	13.75	1.00



Conclusion :

1. The table value of F = 4.75 for $df_1 = 3, df_2 = 6$ and $\alpha = 0.05$, since the calculated F= 1.018 is less than table value the null hypothesis is accepted .
2. The table value of F=5.14 for $df_1 = 2$,$df_2 = 6$ and $\alpha = 0.05$, since the calculated value of F=3.327 is less than table value, the null hypothesis is accepted

Q)Perform ANOVA and decide whether the mean productivity is same our differs among workers

Machine Type

Workers	A	B	C	D
1	40	36	48	38
2	52	44	52	42
3	35	38	45	36
4	48	32	45	34
5	40	40	50	40

Test significance levels at 5%

Sol: let us consider the null hypothesis, that there is no significant difference in the mean productivity of workers in the machine types. H_0 :- $x_1 = x_2 = x_3 = x_4$ and $y_1 = y_2 = y_3 = y_4 = y_5$.

We set up Alternative hypothesis :- $x_1 \neq x_2 \neq x_3 \neq x_4$, $y_1 \neq y_2 \neq y_3 \neq y_4 \neq y_5$.

Machine types

Workers	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2	R.S
1	0	0	-4	16	64	-2	4	16	2
2	12	144	4	16	144	2	4	16	30
3	-5	25	-2	4	25	-4	16	36	-6
4	8	64	-8	64	25	-6	36	16	-1
5	0	0	0	0	100	0	0	0	10

$\Sigma x_1 = 15$, $\Sigma x_1^2 = 233$, $\Sigma x_2 = -10$, $\Sigma x_2^2 = 100$, $\Sigma x_3 = 40$, $\Sigma x_3^2 = 356$, $\Sigma x_4 = -10$, $\Sigma x_4^2 = 60$.

$T = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \Sigma x_4$

$= 15 - 10 + 40 - 10 = 35$.

$C.F = \frac{1225}{20} = 61.25 \Rightarrow SST = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2 - \frac{T^2}{N}$

$= 233 + 100 + 358 + 60 - 61.25$

$= 689.75$

$SSC = \frac{(\Sigma x_1)^2}{\text{No. of cv}} + \frac{(\Sigma x_2)^2}{\text{No. of cv}} + \frac{(\Sigma x_3)^2}{\text{No. of cv}} + \frac{(\Sigma x_4)^2}{\text{No. cv}} - \frac{T^2}{N}$

Prepared By
M.NAVANEETH KUMAR REDDY
B-Tech, MBA
ASSISSTANT PROFESSOR
BALAJI INSTITUTE OF IT AND MANAGEMENT

$= 45+20+320+20-61.25 = 343.75$

$SSR = 1+225+9+0.25+25-61.25 = 199.$

$SSE = SST - (SSC + SSR)$

$= 689.75 - 542.75$

$= 147$

ANNOVA Table:-

Source of Information	Sum of squares	DOF	mean squares
SSC	343.75	$(C-1)=4-1=3$	$msc = \frac{SSC}{C-1} = 343.75/3 = 114.58.$
SSR	199	$5-1=4$	$msR = 199/4 = 49.75$
SSE	147	$(C-1)\times(R-1) = 12$	$msE = 147/12 = 12.25.$

F- Ratio of colu $= \frac{msc}{msE} = 9.35.$

F-ratio of rows $= \frac{msR}{msE} = \frac{49.75}{12.25} = 4.06.$

T.V:- For columns, Tabulated value (3,12) DOf at 5%. is 3.49

For rows, Tabulated value (4,12) at 5%. is 3.26.

Conclusion:- columns: The tabulated value is less when compare Hence project is Rejected.

Rows:- The tabulated value is less when compared to the calculated value so, The project is Rejected.

(17E00105) STATISTICS FOR MANAGERS

The objective of this course is to familiarize the students with the statistical techniques popularly used in managerial decision making. It also aims at developing the computational skill of the students relevant for statistical analysis.

1.Introduction of statistics – Nature & Significance of Statistics to Business, , Measures of Central Tendency- Arithmetic – Weighted mean – Median, Mode – Geometric mean and Harmonic mean – Measures of Dispersion, range, quartile deviation, mean deviation, standard deviation, coefficient of variation – Application of measures of central tendency and dispersion for business decision making.

2. Correlation: Introduction, Significance and types of correlation – Measures of correlation – Co-efficient of correlation. Regression analysis – Meaning and utility of regression analysis – Comparison between correlation and regression – Properties of regression coefficients-Rank Correlation.

3. Probability – Meaning and definition of probability – Significance of probability in business application – Theory of probability –Addition and multiplication – Conditional laws of probability – Binominal – Poisson – Uniform – Normal and exponential distributions.

4. Testing of Hypothesis- Hypothesis testing: One sample and Two sample tests for means and proportions of large samples (z-test), One sample and Two sample tests for means of small samples (t-test), F-test for two sample standard deviations. ANOVA one and two way .

5. Non-Parametric Methods: Chi-square test for single sample standard deviation. Chi-square tests for independence of attributes - Sign test for paired data.

Textbooks:

- Statistical Methods, Gupta S.P., S.Chand. Publications

References:

- Statistics for Management, Richard I Levin, David S.Rubin, Pearson,
- Business Statistics, J.K.Sharma, Vikas house publications house Pvt Ltd
- Complete Business Statistics, Amir D. Aezel, Jayavel, TMH,
- Statistics for Management, P.N.Arora, S.Arora, S.Chand
- Statistics for Management , Lerin, Pearson Company, New Delhi.
- Business Statistics for Contemporary decision making, Black Ken, New age publishers.
- Business Statistics, Gupta S.C & Indra Gupta, Himalaya Publishing House, Mumbai

UNIT-5
NON-PARAMETRIC METHODS

Non-Parametric Methods:-

Practical data, to estimate the parameters such as mean, variance etc and use the standard tests, they are known as parametric tests.

The practical data may be non-normal (or) it may me not possible to estimate the parameters of the data. The test which are used for such situations are called non-parametric tests.

1.Chi-Square test for single sample standard deviation:

χ^2 – test (chi-square test) :-

The χ^2 test was first used by karl-pearson in the year 1900. The χ^2 describes the magnitude of the discrepency b/w theory and observation.

χ^2- square distribution :-

The square of a standard normal variate is called a chi-square variate with 1 degree of freedom (dof). Thus if x is a random variable following normal distribution with mean 'μ' and standard deviation 'σ ' then $\left(\frac{x-\mu}{\sigma}\right)$ is a standard normal variate.

- $\left[\frac{x-\mu}{\sigma}\right]^2$ is a chi-square variate with 1 degree of freedom (dof).

Applications of Chi-square – test :-

Chi-square distribution has a number of applicable, some of which are enumerated below:

1. Chi-square test of goodness of fit.
2. Chi-square test for independence of attributes.
3. To test if the population has a specified value of the variable σ^2

Conditions for applying χ^2 –Test :-

- N, the total number of frequencies should be reasonably large, say greater than 50.
- The sample observations should be independent.
- No theoretical cell frequency should be small.
- The given distribution should not be replaced by relative frequencies of proportions but data should be given in original units.

Chi-Square test for single sample standard deviation :-

Suppose we want to test if the given normal population has a specified variance.

$\sigma^2 = \sigma_o^2$ (say) or not

σ_o^2 =specified value

If $x_1, x_2 \ldots\ldots x_n$ is a random sample of size 'n' from the given population.

We set up null hypothesis as $H0 = \sigma^2 = \sigma_o^2$

under the H0, test statistic is

$$\chi^2 = ns^2/\sigma^2 \text{ follows } \chi^2\text{-distribution with } (n-1) \text{ dof}$$

Where S^2 = variance sample

$$1/n \Sigma (x-\bar{x})$$

n = sample size

s = standard deviation

σ = Expected S.D

σ^2 = Expected Varience.

1) Weights in Kg of 10 students are given below 38,40,45,53,47,43,55,48,52,49. Can we say that varience of distribution of weights of all students from which the above sample of 10 students was drawn; is equal to 20.

Sol :- we set up the null hypothesis as $H_o=\sigma^2=20$.

Calculation of sample varience.

X	38	40	45	53	47	43	55	48	52	49
x- ẍ	-9	-7	-2	6	0	-4	8	1	5	2
$(x-x)^2$	81	49	4	36	0	16	64	1	25	4

$$\ddot{x} = \{\Sigma x / x\} = 470/10 = 47$$

under the test statistic is

$\chi^2 = ns^2/\sigma^2 = \Sigma(x-\ddot{x})^2 / \sigma^2 = 280/20 = 14$

which follows χ^2 distribution with dof (10-1)=9.

Tabulated of χ^2 at 9 dof is 16.919. since calculated value of χ^2 is less than the tabulated value of for 9 dof at 5% level of significance. it is not significant ; Hence H_o may be accepted.

2. A Random sample of size 20 form a population gives the sample standard deviation of 6. Test the Hypothesis that the population S.D is 9.

Sol:- we set up the null the hypothesis as

H_o = the population standard deviation

we are given n=20 and S=6

under H_o : the test statistic is

$$\chi^2 = ns^2/\sigma^2 = 20x36/81 = 8.89$$

and it follows χ^2- distribution (20-1) = 19 dof

Tabulated value of χ^2 for 19 dof =30.144

since calculated value is less than the tabulated value; it is not significant. Hence null hypothesis that the population standard deviation is 9 may be accepted at 5% level of significance.

Chi-Square test of goodness of fit :-

We are given a set of observed frequencies obtained under some experiment and we want to test if the experimental results support a particular hypothesis or theory.

Karl Pearson in 1900, developed a test for testing the significance of the discrepancy b/w Experimental values and the theoretical values obtained under some theory or hypothesis. This test known as χ^2-test of goodness of fit.

We set up the null hypothesis as there is no significant difference b/w the observed. (Experimental) and the theoretical (hypothetical) values.

Steps for consumption of χ^2 and drawing the conclusions :-

1. Compute the expected frequencies E_1, E_2 .....En corresponding to the observed frequencies O_1,O_2.....On under some theory or hypothesis.
2. Compute the deviations (O-E) for each frequency and then square them to obtain $(O-E)^2$
3. Divide the square of the deviations $(O-E)^2$ by the corresponding expected frequency to obtain $(0-E)^2/E$
4. Add the values obtained in step (3) to compute $\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$
5. Look at the tabulated values of χ^2 for (n-1)

dof at certain level of significance, usually 5% or 1% from the table of significant values of χ^2.

6. If calculated value of χ^2 is the less than the tabulated value, then it is said to be non-significant at the required level of significance and we may conclude that there is good correspondence b/w theory and experiment.
7. If calculated value of χ^2 is greater than the tabulated value, it is said to be significant and we may conclude that the experiment does not support the theory.

Q)The number of automobile accidents for week in a certain community were as follows.

12,8,20,2,14,10,15,6,9,4

Are these frequencies in agreement with the belief that accident conditions were the same during this 10-week period.

Sol:- we set up the null hypothesis as the given frequencies are consistent with the belief that the accident conditions were same during the 10-week period. Since the total number of accidents over the 10-weeks are :

12+8+20+2+14+10+15+6+9+4=100

Under the null hypothesis, these accidents should be uniformly distributed over the 10-week period and hence the expected number of accidents for each of the 10 weeks are 100/10=10

Week	Observed no.of accidents (O)	Expected No. of accidents (E)	(0-E)	$(0-E)^2$	$(0-E)^2/E$
1	12	10	2	4	0.4
2	8	10	-2	4	0.4
3	20	10	10	100	10
4	2	10	-8	64	6.4
5	14	10	4	16	1.6
6	10	10	0	0	0
7	15	10	5	25	2.5
8	6	10	-4	16	1.6
9	9	10	-1	1	0.1
10	4	10	-6	36	3.6
					26.6

					26.6

$$\chi^2 = \varepsilon\left[\frac{(O-E)^2}{E}\right] = 26.6$$

dof =10-1=9, Tabulated $\chi^2_{0.05}$ for 9 dof = 16.919 since calculated value χ^2 =26.6 is greater than the tabulated value 16.919, it is significant and null hypothesis is rejected at 5% level of significance

3.In a mendelian experiment on breeding for types of plants are expected to occur in the proportion of 9:3:3:1. The observed frequencies are 891 round and yellow, 316 wrinkled and green. Find the chi-square value and examine the correspondence b/w the theory and the experiment.

Sol:- we set up the null hypothesis as,
H_o : it is assumed that the theoretical values correspond to the experiment values. correspondance of the values.

Total No. of observed plans : 891+316+290+119=1616
Expected frequencies :

Round & Yellow = 9/16 x 1616 = 909
Wrinkiled & yellow = 3/16 x 1616 =303
Round & Green =3/16 x 1616 =303
Wrinkiled and green = 1/16x1616 =101

Procedure is same χ^2=4.6799
dof =4-1=3, tabulated $\chi^2_{0.05}$ for 3 dof =7.80.

since calculated value of χ^2=4.6799 is less than the tabulated value 7.80, it is not significant and null hypothesis is accepted at 5% level of significance.

2. Chi-Square test for independence of attributes:-

Suppose that the given population consisting of N items is divided into 'r' mutually disjoint (Exclusion) and Exhaustive classes $A_1, A_2 \ldots Ar$, with respect to the attribute 'A'.

Similarly let us suppose that the same population is divided in to 'S' mutually disjoint and exhaustive classes $B_1, B_2, \ldots B_s$; with respect to the another attribute 'B'.

We set up null hypothesis as the two attributes A and B are independent if $(A_i B_j)_o$ denote the expected frequency of (A_i, B_j) then

$$(A_i B_j) = \frac{(A_i)(B_j)}{N}$$

$i = 1,2 \ldots\ldots r$

$j = 1,2 \ldots\ldots s$

i.e; the expected frequency for any cell frequency can be obtained on multiplying the row totals and column totals in which the frequency occurs and dividing the product by the total frequency 'N'.

Applying χ^2 –test of goodness of fit, the statistic is $\chi^2 = \sum_i \sum_j \left[\frac{(A_i B_j) - (A_i B_j)_o}{(A_i B_j)_o}\right]^2$ follows χ^2 -distribution with (r-1) x (S-1) dof

r = rows value

s = columns value

Q.A certain drug was administrated to 456 males out of a total 720 in a certain locality to test its efficiency against typhoid. The incidence of typhoid is shown below. Find out the effectiveness of the drug against the disease

	A_1	A_2	
B1 Administering the drug	Infective	No infective	total
	144 (A1 B1)	312 (A2,B1)	456
B2 without Administering the drug	192 (A1 B2)	72 (A2,B2)	264
Total	336	384	720

Sol:- we set up the null hypothesis as the two attributes incidence of typhoid and the 'administration of the drug' are independent. In other words, the drug is not effective against the disease.

Under Ho, the expected frequencies are,

$$E(144) = \frac{336 \times 456}{720} = 212.8$$

$$E(192) = \frac{336 \times 264}{720} = 123.2$$

$$E(312) = \frac{384 \times 456}{720} = 243.2$$

$$E(72) = \frac{384 \times 264}{720} = 140.8$$

Computation of χ^2:-

Observed Frequency 'O'	Expected Frequency 'E'	(O-E)	$(0-E)^2$
144	212.8	-68.8	4733.44
192	123.2	68.8	4733.44
312	243.2	68.8	4733.44
72	140.8	-68.8	4733.44

$$\chi^2 = \sum \left[\frac{(O-E)^2}{E}\right] = 4733.44\left[\frac{1}{212.8} + \frac{1}{123.2} + \frac{1}{243.2} + \frac{1}{140.8}\right]$$

$$= 4733.44\,[X]\,0.0240 = 113.60256$$

Follows χ^2-dof = (r-1) (s-1) = (2-1) (2-1) =1

Tabulated value of $\chi^2_{0.05} = 3.841$

since calculated value of χ^2 is very much greater than tabulated value, it is highly significant. Hence the null hypothesis is rejected at 5% level of significance and we conclude that the drug is certainly effective in controlling typhoid.

Q) Data on hair colour and the eye colour are given in the table. Calculate the value determine the association between the hair colour and the eye colour.

		Fair	Brown	Black	Total
	Blue	15	20	5	40
Eye	Grey	20	20	10	50

colour					
	Brown	25	20	15	60
	Total	60	60	30	150

Sol:- we set up null hypothesis as the 2 attributes association b/w hair colour and Eye colour on same under Ho the expected frequency are.

i $E(15) = \frac{40 \times 60}{150} = 16$

ii $E(20) = \frac{50 \times 60}{150} = 20$

iii $E(25) = \frac{60 \times 60}{150} = 24$

iv, $E(20) = \frac{40 \times 60}{150} = 16$

v, $E(20) = \frac{50 \times 60}{150} = 20$

vi, $E(20) = \frac{60 \times 60}{150} = 24$

vii, $E(5) = \frac{40 \times 30}{150} = 8$

viii, $E(10) = \frac{50 \times 30}{150} = 10$

ix, $E(15) = \frac{60 \times 30}{150} = 12$

Compute :-

Observed Frequency (o)	Expected Frequency (E)	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
15	16	-1	1	0.0625
20	20	0	0	0
25	24	1	1	0.0417
20	16	4	16	1

20	20	0	0	0
20	24	-4	16	0.6667
5	8	3	9	1.125
10	10	0	0	0
15	12	3	9	0.7500
				3.6459

Test statistic = $\chi^2 = \sum \left[\frac{(O-E)^2}{E} \right] = 3.65$

$=3.65$

We can assume that level of significance is 5% i.e 0.05

degree of freedom (dof) = $0.05 \Rightarrow 3-1 \times (3-1) = 3-1 \times (3-1) = 2 \times 2 = 4$

Tabulated values at 4 degree of freedom with 5% level of significance is 9.488

Conclusion :-Here table value is high when compared to calculated value (3.65) i.e., 9.488 high than 36.5. so the project is adjusted at 4 degree of freedom with 5 % level of significance.

- According to rates correction $\Rightarrow$ (2×2)

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

3.Sign test for paired data :-

The sign test is the oldest of all Non-Parametric procedures and it was introducedby Arbuthnott (1710).

The sign test gets its name from the fact that it uses plus and minus signs rather than quantitative measurements as its data.

It is particularly useful where quantitative measurement is impossible or infeasible.

Applying Z-test statistic to test the null hypothesis is Ho : P= ½ that

where r = no.of positive signs

n = total no.of signs (Except 'o').

Single data (or) ordinary sign test:-

The following data in turns are the amounts of sulpher oxides emitted by a large industrial plant, in 40 days sample values.

17,15,20,29,19,18,22,25,27,9,24,20,17,6,24,14,15,23,24,26,19,23,28,19,16,22, 24,17,20,13,19,10,23,18,31,13,20,17,24,14

Test the null hypothesis $\mu_o = 21.5$

Sol:- Let $x_1, x_2, \ldots\ldots x_n$ be the values of the sample size 'n' we want to test $Ho = \mu = \mu_o$

compare each of the 'n' values of the sample with μ_o. if the difference is positive write '+' sign, if it is negative write '-' sign. If the difference is zero ignore all such values and read just the sample size 'n'

Sign of difference when compare with $\mu = 21.5$

−, −, −, +, −, −, +, +, +, −, +, −, −, −, +, −, −, +, +, −, +, +, −, −, +, +, −, −, −, −, −, +, +, −, −

r = no. of positive signs = 16 −, +, −

We want to test Ho : $\mu o = 21.5$

under Ho, the test statistic is $z = \frac{|\bar{x} - \mu_2|}{\sqrt{n/4}} = \frac{|16 - 20|}{\sqrt{10}}$

critical value of z at 5 % is 1.96.

we accept the Ho.

Paired data :-

The nutritionist and medical doctors are always believed that vitamin c is highly effective in reducing the incidents of cold. To test this belief, a random sample of 13 persons is selected and they are given large daily doses of vitamin c under medical supervision over a period of one year. The number of persons who catch cold during the year is recorded and a comparison is made with the number of cold contacted by each such person daily the previous year. This comparison is recorded as follows, along with the sign of the change.

Observations	1	2	3	4	5	6	7	8	9	10	11	12	13

With vitamin c	2	1	0	1	3	2	3	5	1	4	4	3	4
With out vitamin c	7	5	2	3	8	2	4	4	3	7	6	2	10

Using the sign test at $\alpha = 0.05$ level of significance test whether vitamin c is effective in reducing the cold.

Sol:- Let us table the null hypothesis that large is no difference in the number of cold contacted with or without vitamin 'c'

Without vitamin C	7	5	2	3	8	2	4	4	3	7	6	2	10
With vitamin C	2	1	0	1	3	2	3	5	1	4	4	3	4
Sign	-	-	-	-	-	o	-	+	-	-	-	+	-

(To compare with '2' one first one is the bigger value at the time taken the sign is '-')

r = no. of positive signs =2

n = 12 (total signs Except 'o')

Under Ho test statistic is $z = \dfrac{|r - \frac{n}{2}|}{\sqrt{n/4}} = \dfrac{|2 - \frac{12}{2}|}{\sqrt{12/4}} = \dfrac{|12-6|}{4/\sqrt{3}} = 4/1.73 = 2.31$

Since calculated value z =2.31 is greater than the critical value z =1.96 at 5 % level of significance they Ho is rejected.

Prepared By
M.NAVANEETH KUMAR REDDY
B-Tech, MBA
ASSISSTANT PROFESSOR
BALAJI INSTITUTE OF IT AND MANAGEMENT,KADAPA