BALAJI INSTITUTE OF I.T AND MANAGEMENT(BIMK) KADAPA

SEMESTER-1 INTERNAL-2

STATISTICS FOR MANAGERS (SM)

ICET Code : BIMK

Also download softcopy at http://www.bimkadapa.in/materials.html

Name of the Faculty: HIMMAT T

Units covered: 2.5 to 5 units

E-Mail: himmatbimk@gmail.com

(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

to get the posteior distribution.

* As a formal theorem, Bayes theorem is valid in all common intepretations of probability.

Binomial Distribution :-

The binomial distribution also known as "Bernouli Distribution" & associated with the name of a Swiss mathematician James Beenouli also known as Jacques & Jakob (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 processes in business and the social sciences as well as other areas.

Mathematical Distribution :-

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

$$\boxed{p(r) = n_{c_r}\, p^r\, q^{n-r}}$$

This probability distribution is called the "Binomial Probability distribution".

where, $p=$ ~~pobitit~~ probability of success in a single trail.

$q = 1-P$, n = no. of trails.

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

Obtaining Co-efficients of the Binomial :-

For obtaining co-efficients 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 $n_{c_1} q^{n-1} p$.
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 co-efficient of the preceeding 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

$$(q+p)^n = q^n + n_{c_1} q^{n-1} p + n_{c_2} q^{n-2} p^2 + \cdots + n_{c_r} q^{n-r} p^r + \cdots + p^n$$

where,

$1, n_{c_1}, n_{c_2}$ --- are called the binomial (distribution) co-efficients.

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' increases 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 mean 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 particulare outcome becomes smaller.

4) If 'n' is the 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:-

* The outcome of results of each trail in the process are characterised as one of two types of possible outcomes. In otherwords they are attributes.

* The possibility of outcomes of any trail does not change and is independent of the results of perious trails.

1) A fair coin is tossed thrice. find the probability of getting.

i) Exactly 2 heads..

ii) Atleast 2 heads.

Sol:- Binomial distribution.

$$p(r) = n_{C_r}\, p^r\, q^{n-r}$$

$p = \frac{1}{2}$ i.e., probability of a getting a success case.

$$q = 1 - p = 1 - \frac{1}{2} = \frac{2-1}{2} = \frac{1}{2}.$$

i) Exactly 2 heads :- $r = 2$heads

$$p(r) = n_{C_r}\, p^r\, q^{n-r}$$

$$p(2H) = 3_{C_2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{3-2}$$

$$= 3_{C_2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1$$

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

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

$$p(2H) = \frac{3}{8}.$$

ii) Atleast 2 heads :- $r = $(2heads, 3heads)

$$p(2H) = 3_{C_2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{3-2}$$

$$= 3_{C_2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1$$

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

$$p(2H) = \frac{3}{8}$$

$p(3H) = {}^3C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{3-3}$

$= \frac{3\times2\times1}{1\times2\times3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0$

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

$P(3H) = \frac{1}{8}$.

$\therefore 2H + 3H = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$

2) 4 coins tossed simultaneously what is the probability of getting (i) No heads.

(ii) No tails :

(iii) 2 heads only (or) exactly 2 heads.

Sol:- (i) No heads :- $r=0$.

$P(0) = 4_{C0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$

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

$= 1\times1\times\frac{1}{16}$

$p(0) = 0.0625$.

(ii) No Tails :- $r=0$

$p(0) = 4_{C0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$

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

$= 1\times1\times\frac{1}{16}$

$= \frac{1}{16}$

$p(0) = 0.0625$.

(iii) 2 heads only :-

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

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

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

Note :-

* whenever mean, standard deviation and variance are given in the binomial distribution we can consider as

mean = 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 & q values.

Sol:- Mean, np = 20

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

variance, $npq = (4)^2$
$= 16$ i.e., $\sqrt{npq} = 4$ Squaring on bothsides

$(\sqrt{npq})^2 = (4)^2$

$npq = 16$

(q)
probability = $\frac{\text{Variance}}{\text{mean}}$. $= \frac{npq}{np} = \frac{16}{20} = \frac{4}{5}$ $\therefore q = \frac{4}{5}$

we know that $p+q=1$

$p = 1-q$

$= 1-\frac{4}{5} = \frac{5-4}{5} = \frac{1}{5}$

Substitute $p=\frac{1}{5}$ in mean.

$np = 20$

$n\left(\frac{1}{5}\right) = 20. \Rightarrow \frac{n}{5} = 20 \Rightarrow n = 20 \times 5 \Rightarrow n = 100.$

ance

(Standard deviation

$\sqrt{npq} = 4$.

Squaring on bothsides) cancel.

∴ The values of n, p & q is 100, $\frac{1}{5}$, $\frac{4}{5}$.

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

Sol: Mean, $np = 6$.

variance, $npq = 4$.

$\sqrt{npq} = \sqrt{4} = 2$.

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

$q = \frac{2}{3}$.

$p = 1-q = 1 - \frac{2}{3} = \frac{3-2}{3} = \frac{1}{3}$

Substitute $p = \frac{1}{3}$ in mean

$np = 6$

$n(\frac{1}{3}) = 6$

$\frac{n}{3} = 6 \Rightarrow n = 18$.

∴ The values of n, p & q is 18, $\frac{1}{3}$, $\frac{4}{6}$, $\frac{2}{3}$.

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. = $\frac{\text{no. of items then even no. existed}}{\text{Total no. of cases}}$

$$p = \frac{3}{6}$$

$$p = \frac{1}{2}.$$

q = probability of getting a failure case

$$q = 1-p = 1-\frac{1}{2} = \frac{1}{2}.$$

$$q = \frac{1}{2}.$$

(i) 4 success cases i.e. $r=4$

$$p(r) = p(4) = {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{5-4}$$

$$= \frac{5\times\cancel{4\times3\times2}}{1\times\cancel{2\times3\times4}} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1$$

$$= 5\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)$$

$$= 5\times\frac{1}{16}\times\frac{1}{2}$$

$$= \frac{5}{32}$$

$$p(4) = 0.156.$$

(ii) Atleast 4 success cases i.e. $r \neq 0,1,2,3,4$. $r = 4,5$.

$$P(4) = {}^5C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4}$$

$$= \frac{5\times\cancel{4\times3\times2}}{1\times\cancel{2\times3\times4}}\left(\frac{1}{16}\right)\left(\frac{1}{2}\right)$$

$$= 5\left(\frac{1}{32}\right)$$

$$= \frac{5}{32}.$$

$$p(4) = 0.156.$$

$$p(5) = 5_{c_5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{5-5}.$$

$$= \frac{5\times4\times3\times2\times1}{1\times2\times3\times4\times5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0$$

$$= 1\times\frac{1}{32}\times1$$

$$=\frac{1}{32}$$

$$p(5) = 0.031$$

$$p(4)+p(5) = 0.156 + 0.031 = \underline{\underline{0.187}}$$

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 distribution $(q+p)^n$. 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. $\boxed{p(r) = N\times n_{c_r}\, p^r q^{n-r}.}$

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.

Solution Here, $N = 160$

$n = 4$

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

$p = \frac{1}{2}$, $Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$.

No. of Heads	Expected frequency.
0	10
1	40
2	60
3	40
4	10

$r=0$ $p(0) = N \times {}^{n}C_{r}\, p^{r} q^{n-r} \Rightarrow 160 \times 4c_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$

$= 160 \times 4c_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4$

$= 160 \times 1 \times 1 \times \frac{1}{2^4}$

$= 160 \times \frac{1}{16}$

$p(0) = 10.$

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

$= 160 \times 4_{C_1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1}$

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

$= \overset{40}{\cancel{160}} \times \overset{2}{\cancel{4}} \times \frac{1}{\cancel{2}_{1}} \times \frac{1}{\cancel{8}\cancel{4}_{1}} = 40.$

$p(1) = 40.$

Subject :
Title of the test case :
Case study No. :
Date :
Page No. :

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

$$= 160 \times 4_{C_2} \left(\frac{1}{2}\right)^2 . \left(\frac{1}{2}\right)^{4-2}$$

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

$$> 160 \times 6 \times \frac{1}{4} \times \frac{1}{4}$$

$$= \overset{10}{\cancel{160}} \times \frac{6}{\cancel{16}_1}$$

$$P(2) = 60.$$

$$P(3) = 160 \times 4_{C_3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{4-3}$$

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

$$= \overset{40}{\cancel{160}} \times \overset{2}{\cancel{4}} \times \frac{1}{\cancel{8}_{\cancel{4}_1}} \times \frac{1}{\cancel{2}_1}$$

$$P(3) = 40$$

$$P(4) = 160 \times 4_{C_4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4}$$

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

$$> 160 \times \frac{1}{24} \times \left(\frac{1}{2}\right)^0$$

$$= \overset{10}{\cancel{160}} \times \frac{1}{\cancel{16}_1} \times 1$$

$$P(4) = 10$$

2) Fit a binomial distribution from the following data.

x : 0 1 2 3 4
f : 28 62 46 10 4

50% m

x : 0 1 2 3 4
f : 28 62 46 10 4

fx : 0 62 92 30 16 = 200

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

we know that, mean $np = 4/3$, but $n = 4$

$4p = 4/3$ $q = 1-p$

$p = \frac{\cancel{4} \times \cancel{4}}{3 \times \cancel{4}}$ $q = 1 - \frac{1}{3}$

$p = \frac{1}{3}$ $q = \frac{2}{3}$.

If $r = 0$

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

$= 150 \times 4_{c_0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{4-0}$

$= 150 \times 1 \times 1 \times \frac{16}{81}$

$= 150 \times \frac{16}{81}$

$= \frac{2400}{81}$

$= 29.62$.

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

$= 150 \times 4_{c_1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{4-1}$

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

$= 150 \times 4 \times \frac{1}{3} \times \frac{8}{27}$

$= 59.25$.

$p(2) = 150 \times 4_{c_2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{4-2}$

$= 150 \times \frac{\overset{2}{\cancel{4}} \times 3}{1 \times \cancel{2}_1} \left(\frac{1}{9}\right) \left(\frac{2}{3}\right)^2$

$= 150 \times 6 \times \frac{1}{9} \times \frac{4}{9}$

$= 44.44$.

$$P(3) = 150\times {}^4c_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{4-3}$$

$$= 150\times \frac{4\times3\times2}{1\times2\times3}\left(\frac{1}{27}\right)\left(\frac{2}{3}\right)^1$$

$$= 150\times \frac{\cancel{24}^{4}}{\cancel{6}_{1}}\times\frac{1}{27}\left(\frac{2}{3}\right)$$

$$= 150\times4\times\frac{1}{27}\times\frac{2}{3}$$

$$= 150\times4\times\frac{2}{27\times3}$$

$$= 14\cdot81.$$

$$P(4) = 150\times {}^4c_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^{4-4}$$

$$= 150\times \frac{\cancel{4\times3\times2\times1}}{\cancel{1\times2\times3\times4}}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^0$$

$$= 150\times1\times\frac{1}{81}\times1$$

$$= 1\cdot851.$$

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 behaviour 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".

Mathematical Definition:-

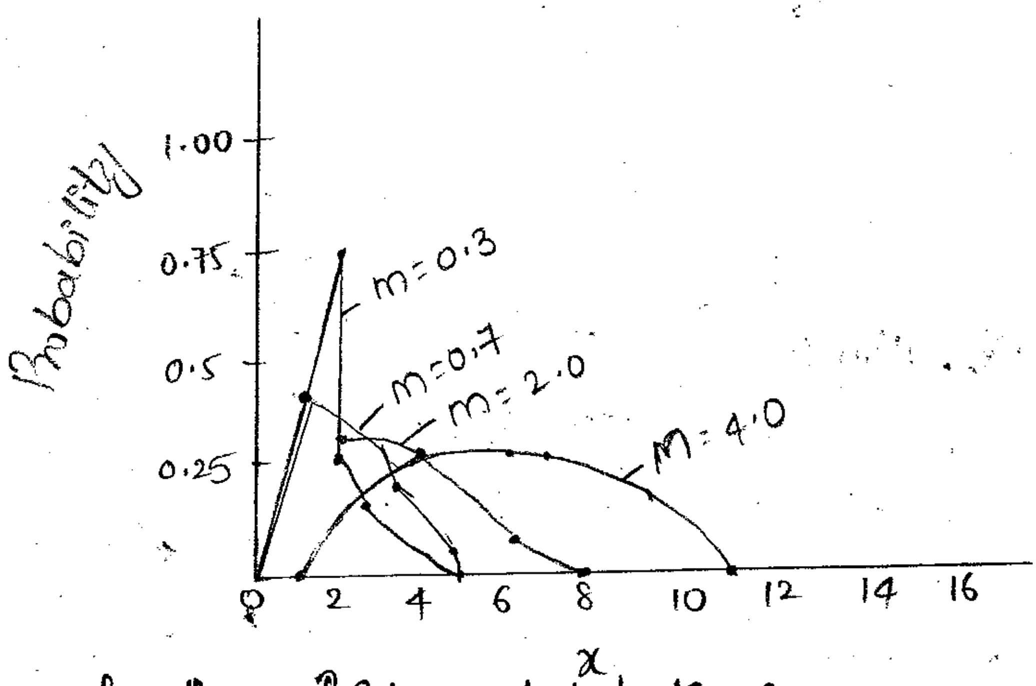
The poisson distribution $p(r)= \frac{e^{-m} m^{r}}{r!}$ m

where, $r=0,1,2,3,4,$ ----

$e = 2.7183$ the base of natural logarithms

m = mean of the poisson distribution.

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



Role of the Poisson 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 radio active substance.

* In Insurance problems to count the no. of cesuelities

* In waiting time problems to count the no. of incoming telephone calls (or) incoming customers.

* No. of traffic arrivals such as trucks at terminals, aeroplanes at airports ships and so fourth.

* In deteming in 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 and

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

Characters of Poison Distribution :-

Discrete distribution :-
Like binomial distribution it is also a discrete probability i.e., occurances 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 distributions.

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

Properties :-

- * The experiments results in outcomes that can be classified as successes (or) failures.
- * The average no. of success (m) that occur in a specified region is known.
- * The probability that a success will occur is proportional to the size of the region.
- * The probability that a success will occur is an extremely small region is virtually zero.
- * It is discrete probability distribution where the random variable x assumes the infinite set of values 0, 1, 2, ---.
- * Mean = m = parameter of the distribution, variance $(\sigma)^2 = m$, S.D $(\sigma) = \sqrt{m}$, Skewness $= \frac{1}{\sqrt{m}}$ & kurtosis $= \frac{1}{m}$.
- * The mode of poisson distribution is that value x which occurs with largest probability it may have either one of two models. If 'm' is not an integer, the mode is the integral value between m-1 & m. If, however m is an integer, then there are two modes which are m-1 & m.
- * If x & y be two independent poisson variates with parameters m_1 & m_2 respectively, then their sum x+y is also a poisson variate with parameter $m_1 + m_2$.
- * The first, second and third new movements are respectively m, m^2+m, m^3+3m^2+m.

1) 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%

$= \frac{2}{100} = 0.02$,

$q = 1 - P = 1 - 0.02 = 0.98$

Mean $= np$, here $n = 100$

$= 100 \times 0.02$

Mean $= 2$

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

i) No defective items $(r=0)$:-

$$p(0) = \frac{e^{-2} \cdot 2^0}{0!} = \frac{0.135 \times 1}{1!} = 0.135.$$

ii) One defective screws :-

$$p(1) = \frac{e^{-2} \cdot 2^1}{1!} = \frac{0.135 \times 2}{1} = 0.270.$$

iii) Two (or) More defective items :- $(r=2)$

$$= 1 - [(p(0) + p(1))]$$

$$= 1 - [0.135 + 0.27] = 1 - 0.405 = 0.595$$

2) 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 houses 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 houses in the fire accident during a year $\frac{1}{1000}$.

$$\text{mean} = np$$

$$= 2000 \times \frac{1}{1000}$$

$$\text{Mean} = 2$$

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

i) Probability of getting exactly 5 houses in a fire accident during a year, $r=5$.

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

$$= \frac{0.135\ (32)}{5 \times 4 \times 3 \times 2 \times 1}$$

$$= \frac{4.32}{120}$$

$$p(5) = 0.036$$

Fitting a poison Distribution:-

The process of fitting a poison distribution is very simple. we have just obtain the value of 'm'. i.e., the average occurance and calculate the frequency of '0' success. The other frequencies can be very easily calculated as follows

$N(P_0) = Ne^{-m}$

$N(P_1) = N(P_0) \times \frac{m}{1}$

$N(P_2) = N(P_1) \times \frac{m}{2}$

$N(P_3) = N(P_2) \times \frac{m}{3}$ etc.

i) 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 mistake occur (f) :	211	90	19	5	0

Sol: Here $N = 325$ $(211+90+19+5+0)$

$fx = 0 \quad 90 \quad 38 \quad 15 \quad 0$

$\Sigma fx = 143$

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

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

$NP(0) = N \times e^{-m} = 325 \times 0.644$

$= 209.3.$

$NP(1) = NP(0) \times \frac{m}{1} = 209.3 \times \frac{0.44}{1}$

$= 92.09$

$NP(2) = NP(1) \times \frac{m}{2} = 92.09 \times \frac{\cancel{m}}{2}(0.44) = 92.09 \times \frac{0.44}{2}$

$= 92.09 \times 0.22$

$= 20.25$

$NP(3) = NP(2) \times \frac{m}{3} = 20.25 \times \frac{0.44}{3} = 20.25 \times 0.146 = 2.9.$

$NP(4) = NP(3) \times \frac{m}{4} = 2.9 \times \frac{0.44}{4} = 2.9 \times 0.11$

$= 0.319.$

Assumed (or) Success cases	Expected cases.
0	209.3
1	92.09
2	20.25
3	2.9
4	0.3
	324.9

$\cong 325.$

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:

$m = \frac{\Sigma fx}{N} = \frac{145}{330} = 0.439$

$NP(0) = N \times e^{-m} = 330 \times e^{-0.439} = 330 \times 0.6447 = 212.75.$

$NP(1) = NP(0) \times \frac{m}{1} = 212.75 \times \frac{0.439}{1} = 212.75 \times 0.439$

$= 93.39.$

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

$NP(2) = NP(1) \times \frac{m}{2} = \frac{0.439}{2} = 0.2195 = \underline{\frac{m}{2}}$

$= 93.39 \times \frac{m}{2} = 93.39 \times 0.2195$

$= 20.499.$

$NP(3) = NP(2) \times \frac{m}{3} = 20.499 \times \frac{0.439}{3} = 20.499 \times 0.146$

$= 2.992.$

$NP(4) = NP(3) \times \frac{m}{4} = 2.992 \times \frac{0.439}{4} = 2.992 \times 0.109$

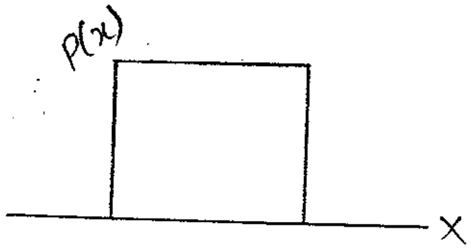
$= 0.326$

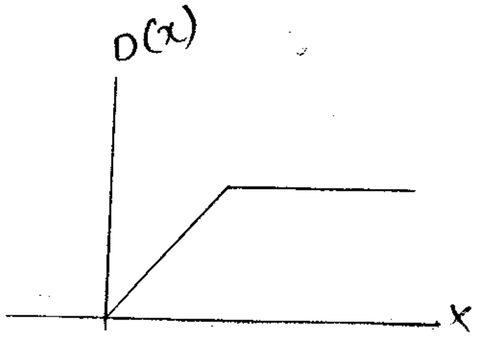
Success cases	Expected cases.
0	212.75
1	93.39
2	20.499
3	2.992
4	0.326
	329.957

≈330.

Uniform Distribution :-

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





The probability density function and cummulative distribution function for a continuous uniform distribution on the interval $[a,b]$ are

$$p(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \le x \le b \end{cases} \rightarrow ①$$

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

Mean and S.D of a union Distribution:-

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

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

Probabilities in a uniform Distribution:-

The following equation is used to determining the probabilities of 'x' for a uniform distribution between a & b.

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

Normal distribution:-

The normal distribution was first described by "Abraham Demoivre" as the limiting from of the binomial model in 1733. Normal distribution was rediscovered by Gauss in 1809 & by Leplace in 1812.

The normal distribution also called "the normal probability distribution".

Mathematical definition:-

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

x = value of the continuous random variable.

m = mean of the normal random variable.

e = mathematical constant approximated by 2.7183.

π = mathematical constant approximated by 3.1416.

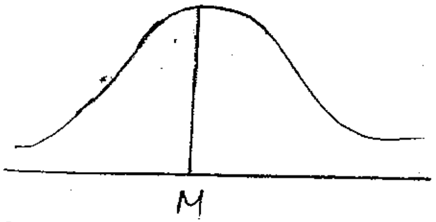
$(\sqrt{2\pi} = 2.5066)$.

Graph of Normal Distribution:-

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

* Normal distribution is a limiting case of binomial distribution when (i) $n \to \infty$

(ii) Neither p or q is very small



* Normal Distribution is a limiting case of poison distribution when its mean 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 socelled control limit theorem.
* Account to this theorem as the sample size 'n' increase the distribution of mean, $\overline{n}$ 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 are

- * The normal curve is 'bell-shaped' & symmetrical in it's appearance. If the curves were folded along it's vertical axis, the two halves would coincide.
- * The height of the normal curve is at it's 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 inflection i.e., the points where the change in curvature occur are $\bar{x} \pm \sigma$
- * As distinguished from binomal and passion distributed where the variable discrete the variable distributed acount to the normal curve is a continuous one.
- * The 1^{st} & 3^{rd} variables are equidistant from the median.
- * The mean deviation is 4th of mole preciously 0.7979 of the S.D.
- * The area under the normal curve distributed as follows
- * Mean $\pm 1\sigma$ covers 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 99.73% area.

UNIT - IV

TESTING OF HYPOTHESIS

Introduction :-

The term hypothesis derives from the Greek "hypotithenai" meaning "to put under" (or) "to suppose".

Hyphothesis is a tentative conjecture explaning 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 'H_0'.

Alternative Hypothesis :-

It is a statement of what a statistical hypothesis test is set up to establish. It is denoted by 'H_1'.

Procedure for testing of hypothesis :-

The following are various steps in testing a statistical hypothesis.

- * Assume Null hypothesis : H_0
- * Alternative hypothesis H_1, helps us to decide whether we have to use a single tailed (or) two tailed test.

3. **Level of Significance :-**

choose appropriate level of significance (α) depending on the permissible risk the α is fixed in advance before sample is drawn.

4. **Test statistics :-**

Compute the test statistic,

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

5 * **Inference :-**

We compare the computed value of z in step(4) with significant value (tabulated value) z_α = 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 & 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.

$$E(x) = np \quad \text{and} \quad v(x) = npq$$

$$\text{where, } q = 1-p.$$

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".

$$p = \frac{x}{n}$$

$$\therefore E(p) = E\left(\frac{x}{n}\right)$$

$$= \frac{1}{n} \cdot E(x)$$

$$= \frac{1}{\cancel{n}} \cdot \cancel{n}p$$

$$= p$$

Thus, the sample proportion p is unbaised estimate of population proportion 'p'.

Also $V(p) = V\left(\frac{x}{n}\right)$

$$= \frac{1}{n^2} \cdot V(x)$$

$$= \frac{npQ}{n^2}$$

$$= \frac{pQ}{n}$$

standard error $S.E(p) = \sqrt{\frac{pQ}{n}}$

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

$$z = \frac{\cancel{p}p -}{\sqrt{\frac{pQ}{n}}} \; p - p$$

Note :- The limit for P at level of level of significance α are given by

$$p \pm z\alpha\sqrt{\frac{pq}{n}}$$

1) In a sample of 1000 people in karnataka 540 are rice eaters and rest are wheat eaters can we assume both rice & 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 $p = \frac{x}{n}$

$$= \frac{540}{1000}$$

$$p = 0.54$$

Null hypothesis, (H0) :- Both rice and wheat eaters are equally popular in the state

$$H_0 : P = 0.5$$

$$P = 0.5 \text{ and } Q = 1-P = 1-0.5 = 0.5$$

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

Test statistic :- Under H0 test statistic is given by

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$$z = \frac{0.54-0.5}{\sqrt{\frac{0.5\times0.5}{1000}}} = \frac{0.04}{0.0138}$$

$$z = 2.532.$$

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.

2) 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 $p = \frac{x}{n}$

$= \frac{200}{700}$

$p = 0.286$

$q = 1 - p = 1 - 0.286 = 0.714$.

Hence, standard error $SE(p)$ is given by

$$SE(p) = \sqrt{\frac{pq}{n}}$$

$$= \sqrt{\frac{0.286 \times 0.714}{700}}$$

$$= 0.017.$$

i) 95% confidence limits for p are given by

$$p \pm Z_\alpha \sqrt{\frac{pq}{n}}$$

5% loss significant value is 1.96 (Z_α)

$\Rightarrow p \pm 1.96\sqrt{\frac{pq}{n}} = 0.286 \pm 1.96 \times 0.017$

$= 0.286 \pm 0.033$

$= (0.253, 0.319)$

(ii) 99% confidence limits for p are given by $p \pm z_\alpha\sqrt{\frac{pq}{n}}$

1% loss significant value is 2.58 (Z_α)

$p \pm 2.58\sqrt{\frac{pq}{n}} = 0.286 \pm 2.58 \times 0.017$

$= 0.286 \pm 0.044)$

$= (0.242, 0.33)$

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 samples.
- * 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

$$P_1 = \frac{x_1}{n_1}, \quad P_2 = \frac{x_2}{n_2}$$

if P_1 and P_2 are population proportions,

$E(P_1)=P_1$, $E(P_2)=P_2$

$V(P_1)=\frac{P_1Q_1}{n_1}$, $V(P_2)=\frac{P_2Q_2}{n_2}$

Under $H_0 \Rightarrow P_1=P_2=P$, $Q_1=Q_2=Q$. then the test statistic will becomes

$$Z=\frac{p_1-p_2}{\sqrt{PQ\left(1/n_1+1/n_2\right)}} \sim N(0,1)$$

1) 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 are same against that they are not at 5% loss.

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

$\Rightarrow p_1=\frac{x_1}{n_1}=\frac{200}{400}=0.5$

$n_2=600$, $x_2=325$

$\Rightarrow p_2=\frac{x_2}{n_2}=\frac{325}{600}=0.54$.

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, $H_1 \Rightarrow P_1 \neq P_2$.

Test statistics:-

Since samples are large, the test statistic under H_0 is $Z = \frac{P_1 - P_2}{\sqrt{PQ(1/n_1 + 1/n_2)}}$;

where $P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2}$

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

$= 0.524$.

$Q = 1 - P = 1 - 0.524 = 0.476$.

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

$|Z| = \frac{0.04}{\sqrt{0.524 \times 0.476 \left(\frac{10}{2400} \right)}}$

$|Z| = \frac{0.04}{0.0323}$

$|Z| = 1.269$.

Conclusion:-

Since $Z = 1.269$ which is less than 1.96 significant value at 5% loss.

Hence, H_0 may be accepted.

2) 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 $n_1 = 1000$, $n_2 = 1200$

$x_1 = 800$, $x_2 = 800$

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

Null hypothesis, $H_0 : P_1 = P_2$

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

Alternate hypothesis: $H_1 ; P_1 \neq P_2$.

Test statistics is given by under H_0 is

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

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

$$\therefore Z = \frac{0.8 - 0.67}{\sqrt{16/22 \times \frac{6}{22}\left(1/1000 + 1/1200\right)}}$$

$$= \frac{0.13}{0.019} = 6.842.$$

Conclusion:- Alternative 5% loss 1.96 we found evidance against H_0; Hence, we reject H_0.

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 σ^2 then the sample mean is distributed normally with mean μ and variance $\frac{\sigma^2}{n}$; 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 a finite population with size N, the corresponding limits are

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

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

1) 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 ($\alpha = 5\%$ loc).

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 statistics is given by

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

Conclusion:- We have found evidence against null hypothesis H_0. so, it can be reasonably regarded that the given sample is from the said population at 5%.

2) 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?

Sol:- Given, $n = 100$, $N = 2000$, $\bar{x} = 0.354$

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 $\bar{x}_1$ be the mean of a random sample of size n_1 from a population with mean μ_1 and variance σ_1^2 and $\bar{x}_2$ be the mean of a random sample of size n_2 from a population mean μ_2 and variance σ_2^2. Then, sample sizes n_1 and n_2 are large then

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

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

1) In a random sample of 500, the 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 $z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$

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

$$= \frac{5}{0.018}$$

$$= 277.77$$

Reject H_0.

Assumptions for students t-test :-

The following assumptions are made in the students t-test.

* The parent population from which the sample drawn is normal.
* The population, observations are independent, 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 difference 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.

students t:-

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

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

where, $\bar{x} = \frac{\Sigma x}{n}$

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

Test for Single Mean:-

1) A machine is designed to produce insulating washers for electrical devices of an average thickness of 0.025 cm. A random sample of 10 washers was found to have an average thickness of 0.024 cm with a standard deviation of 0.002 cm. Test the significance of the deviation.

Sol:- we are given, $n=10$, $\bar{x} = 0.024$ cm, $s = 0.002$ cm

$\mu = 0.025$ cm.

Null hypothesis :-

$H_0 : \mu = 0.025$ cm, i.e., there is no significant deviation between sample mean $\bar{x} = 0.024$ and population when $\mu = 0.025$.

Alternative hypothesis :-

$H_1 ; \mu \neq 0.025$ cm

Under Ho, the test statistics is

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

$$\frac{-0.001 \times 3}{0.002} = -1.5.$$

Tabulated value of $t_{0.05}$ for 9 degrees of freedom = 1.833

Since, $|t| < 1.833$ is not significant between sample mean and population mean (is not significant).

2) Certain pesticide is packed into 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: $H_0 = \mu = 50$ kgs.

i.e., the average packing is 50 kgs.

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

x: 50 49 52 44 45 48 46 45 49 45.

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
$\Sigma x =$ 473		$\Sigma X^2 = 64.1$

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

Standard deviation = $\sqrt{\frac{\Sigma x^2}{n}}$ $(\because x-\bar{x} = X)$

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

$= \frac{64.1}{10}$

$s^2 = 6.41$

The test statistic is $t = \frac{\bar{x} - \mu}{\sqrt{s^2 / n-1}}$

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

$= \frac{-2.7}{\sqrt{0.712}} = \frac{-2.7}{0.8438}$

$= -3.2.$

Tabulated value of $t_{0.05}$ for 9 degress of freedom $= 1.833$.

Since, calculated $|t|$ is greater than tabulated t, it is significant. Hence, H_0 is rejected.

2) 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)

Sol: Null hypothesis : $H_0 : \mu = 100$

i.e., the assumption of a population of IQ is 100.

x	$X = (x-\bar{x})$	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
		$\Sigma X^2 = 1{,}833.6$

Mean = $\frac{972}{10}$

$\bar{x} = 97.2$

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

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

$= \frac{1833.6}{10}$

$s^2 = 183.36$

The test statistic is $t = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}}$

$$= \frac{97.2 - 100}{\sqrt{183.36/9}}$$

$$= \frac{-2.8}{\sqrt{20.373}}$$

$$= \frac{-2.8}{4.513}$$

$$t = -0.62$$

$$|t| = 0.62$$

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, H_0 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 $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2, \ldots y_{n_2}$ be two independent random samples from the given normal populations.

we set up the null hypothesis $H_0 = \mu_x = \mu_y$ under the H_0 the test statistic is

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

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

$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1}$, $s_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2}$, $s^2 = \frac{n_1 s^2 + n_2 s_2^2}{n_1+n_2-\cancel{2}\,2}$

(follows students t-distribution with n_1+n_2-2 dof)

1) 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.

50|34 In the usual notations we are given

$n_1 = n_2 = 25$, $\bar{x} = 200$, $\bar{y} = 250$, $s_1 = 20$, $s_2 = 25$

Null hypothesis $H_0 = \mu_1 = \mu_2$ i.e, both the machines are equally efficient.

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

Under the H_0 the test by statistics is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{x_1}+\frac{1}{x_2}\right)}} \quad \text{where} \quad s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

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

$$s^2 = \frac{25625}{48}$$

$$\therefore s^2 = 533.85$$

$$t = \frac{200-250}{\sqrt{533.85\left(\frac{1}{25}+\frac{1}{25}\right)}}$$

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

$$= \frac{-50}{\sqrt{42.708}} = \frac{-50}{6.535} = -7.65$$

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

Since, calculated $|t| >$ tabulated t, it is highly significant. Hence, H_0 is rejected and we conclude that both the machines are not equally efficient at 5% level of significance.

2) 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?

Soln: In the usual notations we are given

$n_1 = 9$, $\bar{x} = 196.42$, $\Sigma(x-\bar{x})^2 = 26.94$.

$n_2 = 7$, $\bar{y} = 198.82$, $\Sigma(y-\bar{y})^2 = 18.73$

Null hypothesis: The samples 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 statistics is.

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{x_1}+\frac{1}{x_2}\right)}}$$

we have $s^2 = \dfrac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n_1+n_2-2}$

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

$$S^2 = \frac{26.94 + 18.73}{9+7-2} = \frac{45.67}{14} = 3.26.$$

$$= \frac{-2.40}{\sqrt{3.26 \times 0.254}}$$

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

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

Tabulated $t_{0.05}$ for 14 dof is 1.761

Since, calculated |t| is greater than tabulated t, it is significant. Hence, Ho is rejected

3) Two different types of drugs A and B were tried 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 are 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)

F-distribution (F-test)

F-distribution was introduced by G.W. Snedecor. The F-test is named in honour of the great

Statisfaction R.A. Fisher.

F-Test for two sample standard deviations :-

Let $x_1, x_2, ---- x_n$ be a random sample of size n_1 from the first population with variance σ_1^2 and $y_1, y_2, ----y_n$ be a random sample of size n_2 from the second normal population with variance σ_2^2. Obviously the two samples are independent.

We set up the null hypothesis as

$$H_0 = \sigma_1^2 = \sigma_2^2 = \sigma^2$$

i.e., population variances are same.

under H_0, the test statistic is

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1) \text{ where } S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1}$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1}$$

follows F-distribution with (n_1-1, n_2-1) d.f.

Assumption in F-test :-

The F-test is based on the following assumptions.

- * **Normality** :- Values in each group are normally distributed.
- * **Homogenity** :- The variance with in each group should be equal for all groups $(\sigma_1^2 = \sigma_2^2 = ---- = \sigma_k^2)$
- * **Independence of Error** :- It states that the error should be independent for each value.

Applications 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 $H_0 = \mu_1 = \mu_2 ---- = \mu_k$ for k normal populations.

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

Method - 1 - 20 16 26 27 23 32.

Method - 2 - 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?

Solu: We set up null hypothesis as $H_0 : \sigma_1^2 = \sigma_2^2$

i.e., there is no significant difference between the variances of the time distribution by the workers in performing a job by 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
$\Sigma x =$ 134		$\Sigma(x-\bar{x})^2 =$ 81.34

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
$\Sigma y =$ 241		$\Sigma(y-\bar{y})^2 =$ 133.72

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

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

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

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

Since, $s_2^2 > s_1^2$, under H_0; the test statistic is

$$F = \frac{s_2^2}{s_1^2} \sim F(n_2-1, n_1-1)$$

$$F = \frac{22.28}{16.26} = 1.37$$

Tabulated $F_{0.05}(6,5) = 4.95$.

Since, calculated F is less than tabulated F, it is not significant. Hence, H_0 may be accepted at 5% level of significance.

2) 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 3.8mm. Compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance.

Solution Given data.

$n_1 = 22$ $n_2 = 16$

$s_1 = 2.9mm$, $s_2 = 3.8mm$

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 firms B have the some variability.

we have $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$

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

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

$= \frac{185.02}{21}$ $= \frac{231.04}{15}$

$= 8.810$ $= 15.402$

Since, $S_2^2 > S_1^2$ under H_0 the test statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{15.402}{8.810} = 1.748.$$

which follows F distribution with (15, 21)

Tabulated $F_{0.05}(15, 21) = 2.20$.

Since, calculated F is less than the tabulated F, it is not significant at 5% level of significance. Hence, H_0 is accepted.

*** Design of Experiments :-**

An experimental design is a plan and a structure to test hypothesis is which the recorder either controls & 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 :-

* Comparison * Randomization, * Blocking.

* Replication. * 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.

Analysis of variance (ANOVA) :-

* Analysis of variance was developed by R.A. fishner.

* Analysis of variance the significance of the difference between 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, socialogy, 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 analysing the variance to which a response is subject into it's 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 $V_1 = V_2 = V_3 = \dots\dots V_n$.
* 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 with in 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.

$$T = \Sigma x_1 + \Sigma x_2 + ---- + \Sigma x_n.$$

* calculate the value of connection factor $\left(\frac{T^2}{N}\right)$.

where, N = Total no. of variables.

* calculate the value of $SST = \Sigma x_1^2 + \Sigma x_2^2 + ---- + \Sigma x_n^2 - \frac{T^2}{N}$

(sum of squares for variance of total).

* Calculate the value of $SSB = \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n} + --- + \frac{(\Sigma x_n)^2}{n} - \frac{T^2}{N}$

(sum of squares for variance between the samples].

* Find out the value of $SSW = SST - SSB$.

(sum of squares for variance between with in the samples]

* Draw the ANOVA Table.

* Finally, F-ratio may be worked out as,

$$F\text{-ratio} = \frac{MSB}{MSW}.$$

MSB = Mean square between samples.

MSW = Mean square with in samples.

1) 4 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 machines.

Soln: Let us take the null hypothesis that the machines do not differ significantly in performance,

i.e., $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

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
$\Sigma x_1 = 40$	$\Sigma x_1^2 = 410$	$\Sigma x_2 = 28$	$\Sigma x_2^2 = 216$	$\Sigma x_3 = 53$	$\Sigma x_3^2 = 745$	$\Sigma x_4 = 90$	$\Sigma x_4^2 = 2038$

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

$$= 40 + 28 + 53 + 90$$

$$T = 211.$$

correction factor $= \frac{T^2}{N} = \frac{(211)^2}{16} = \frac{44521}{16} = 2782.56$

Sum of squares for variance of Total (SST) =

$$\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2 - \frac{T^2}{N}$$

$= 410+216+745+ 2036 - 2782.56$:

$= 3409 - 2782.56$

$= 626.44$.

Sum of squares for variance between the sample (SSB) =

$$= \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} + \frac{(\Sigma x_4)^2}{n_4} - \frac{T^2}{N}$$

$$= \frac{(40)^2}{4} + \frac{(28)^2}{4} + \frac{(53)^2}{4} + \frac{(90)^2}{4} - 2782.56$$

$$= \frac{1600}{4} + \frac{784}{4} + \frac{2809}{4} + \frac{8100}{4} - 2782.56$$

$$= 400 + 196 + 702.25 + 202.5 - 2782.56.$$

$$= 540.69$$

Sum of squares for variance with in the samples (SSW) =

$SST - SSB$

$= 626.44 - 540.69$

$= 85.75$

N = Total no. of variables (or) sample values.

K = Number of variables types (sample (or) variables)

ANOVA Table.

①	②	③	④ = 2/3
Source of Variation	Sum of Squares	Degree of freedom	Mean Square
Between samples	540.69	3 (K-1)	180.23
within samples	85.75	12 (N-K)	7.15.

$$F - \text{ratio} = \frac{MSB}{MSW} = \frac{180.23}{7.15} = 25.207.$$

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 a significant difference in the performance of the four machines.

2) A random sample is selected from each of 3 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 2 cases are existed.

* ANOVA technique is context of 2-way design when

repeated values are not there.

ANOVA is context of 2-way design when repeated values are there.

The following steps are involved.

- * Use the coding device.
- * calculate the sum of normal, squares of the individual variables.
- * Calculate the sum of individual sum of the variables

$$T = \Sigma x_1 + \Sigma x_2 + \text{---} + \Sigma x_n$$

- * calculate the value of correction factor $\left(\frac{T^2}{N}\right)$.

where, N = Total no of variables.

- * calculate the value of $SST = \Sigma x_1^2 + \Sigma x_2^2 + \text{---} + \Sigma x_n^2 - \frac{T^2}{N}$
- * (calculate the value of $SSB = \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \text{---} + \frac{(\Sigma x_n)^2}{n} - \frac{T^2}{N}$
- * 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 concerning 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.

* Findout the value of sum of squares of deviations for residual (or) error variance [SSE] = SST - (SSC+SSR).

* Draw the ANOVA Table.

* Test statistic $F = \frac{\text{MSB (columns)}}{\text{MSR}}, \frac{\text{MSB (rows)}}{\text{MSR}}$

MSR = Mean square residual.

1) The following table gives the number of refrigrators sold by 4. salesman in 3 months may, June & July.

	Salesman			
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 4 salesman? Is there a significant difference in the the sales made during different months?

Solution Let us take the null hypothesis that there is no significant difference between sales made by the four salesmen during different months.

∴ The given data are coded by substracting 40 from each observation calculations for a 2-criterion month & sales man.

Subject :

Title of the test case :

Case study No. :

Date :

Page No. :

Sales man

Month	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2	Row sum
May	10	100	0	0	8	64	−1	1	17
June	6	36	8	64	10	100	5	25	29
July	−1	1	4	16	0	0	−1	1	2
	$\Sigma x_1 =$ 15	$\Sigma x_1^2 =$ 137	$\Sigma x_2 =$ 12	$\Sigma x_2^2 =$ 80	$\Sigma x_3 =$ 18	$\Sigma x_3^2 =$ 164	$\Sigma x_4 = 3$	$\Sigma x_4^2 =$ 27	48

T = sum of all observations = 48

correction factor $= \frac{T^2}{N} = \frac{(48)^2}{12} = 192$

SSC = sum of squares between sales men (columns)

$= \frac{(15)^2}{3} + \frac{(12)^2}{3} + \frac{(18)^2}{3} + \frac{(3)^2}{3} - 192$

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

SSR = sum of squares between months (rows)

$= \frac{(17)^2}{4} + \frac{(29)^2}{4} + \frac{(2)^2}{4} - 192$

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

$= 91.5.$

$SST = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2 - C.F$

$= 137 + 80 + 164 + 27 - 192$

$= 216.$

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

Degrees of freedom $c-1 = 4-1 = 3$

$r-1 = 3-1 = 2$

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

The ANOVA Table:-

S.No	Sources of variation	Sum of squares	d.f	Mean squares	Variance ratio.
1.	Between salesmen	42	3	14	$F = \frac{14}{13.75} = 1.018$
2.	Between months.	91.5	2	45.75	$F = \frac{45.75}{13.75} = 3.327$
3.	Residual Error	82.5	6	13.75	$F = \frac{13.75}{13.75} = 1.00$

Conclusion:-

*1 The table value of $F = 4.75$ for $df_1 = 3$, $df_2 = 6$ & $\alpha = 0.05$, 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 is table value, the null hypothesis is accepted.

* Perform ANOVA and decide whether the mean productivity is same (or) differs among workers.

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

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%

UNIT - V

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."

χ^2 - test (chy-square test) :-

The χ^2 test was first used by "karl pearson" in the year 1900. The χ^2 describes the magnitude of the descrepency between theory & observation.

χ^2 -Square Distribution :-

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

$\therefore \left(\frac{x-u}{\sigma}\right)^2$ is a chy-square variate with 1 degree of freedom (dof).

Applications of χ^2-Test :-

chy-square distribution has a number of applications some of which are enumerated below:

1) chy-square test of goodness of fit.
2) chy-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.

Chy-square test for single sample standard deviation :-

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

$\sigma^2 = \sigma_0^2$ (say) or not

σ_0^2 = 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 $H_0 = \sigma^2 = \sigma_0^2$

under the Ho, test statistic is

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

where, s^2= variance sample

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

n = sample size.

s = standard deviation.

σ = Expected S.D.

σ^2 = Expected variance.

1) Weights in kg of 10 students are given below.

38, 40, 45, 53, 47, 43, 55, 48, 52, 49.

Can we say that variance of distribution of weights of all students from which the above sample of 10 students was drawn is equal to 20.

Sol^n: We set up the null hypothesis as $H_0 = \sigma^2 = 20$.

Calculation of sample variance.

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

$$\bar{X} = \frac{\Sigma x}{x} = \frac{470}{10} = 47.$$

Under the test statistic is $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{\Sigma(x-\bar{x})^2}{\sigma^2} = \frac{280}{20} = 14$.

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, Ho may be accepted.

2) A Ramdom 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_0 = The population standard deviation.

We are given $n=20$ and $s=6$.

under H_0: the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma^2} = \frac{20\times36}{81} = 8.89$$

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.

chy-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 decrepency between experimental values and the theoretical values obtained under some theory of hypothesis. This test known as χ^2- Test of goodness of fit.

we set up the null hypothesis as there is no significant difference between the observed. (Experimental) and the theoretical (hypothetical) values.

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

*1 Compute the expected frequencies $E_1, E_2, --- E_n$ Corresponding to the observed frequencies $O_1, O_2, --- O_n$. Under some theory of hypthesis.

*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 $\frac{(O-E)^2}{E}$

*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 fis $(n-1)$ 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-singnificant at the required level of significance and we may conclude that there is a good correspondance between theory & 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.

1) The number of automobile accidents per 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 given

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

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 $\frac{100}{10} = 10.$

Week	observed No. of accidents (O)	Expected No. of accidents (E)	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
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.

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

$$= 26.6.$$

dof = 10-1=9, Tabulated $\chi^2_{0.05}$ for 9 dof = 16.919.

Since, calculated value $\chi^2 = 26.6$ is greater than the tabulated value 16.919, it is significant and null hypothesis is rejected at 5% level of significance.

2) 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 yellow, 290 round and green, and 119 wrinkled and green. Find the chy-square value and examine the correspondance between the theory and the experiment.

Sol: We set up the null hypothesis as,

H_0: it is assumed that the theoretical values correspond to the experiment values.

Total no. of observed plans: 891+316+290+119=1616.

Expected Frequencies :-

Round & yellow → $\frac{9}{16} \times 1616 = 909$.

wrinkled & yellow → $\frac{3}{16} \times 1616 = 303$.

Round & Green → $\frac{3}{16} \times 1616 = 303$.

wrinkled & green → $\frac{1}{16} \times 1616 = 101$.

procedure is same $\chi^2 = 4.6799$.

dof = 4-1=3, Tabulated $\chi^2_{0.05}$ for 3dof = 7.80.

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

Chy-square test for independence of attributes:-

Suppose that the given population consisting of N items is divided into 'r' mutually disjoint (Exclusive) & exhaustive classes $A_1, A_2, --- A_r$, with respect to the attribute 'A'.

Similarly, let us suppose that the same population is divided into 's' mutually disjoint & exhaustive classes $B_1, B_2, ---- B_s$, with respect to the another attribute 'B'.

We set up null hypothesis as the two attributes A and B are independent.

If $(A_i B_j)_0$ denote the expected frequency of (A_i, B_j) then:

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

$$i = 1,2,\text{----}r$$
$$j = 1,2,\text{-----}s.$$

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 goodyness of fit, the statistic is

$$\chi^2 = \sum_i \sum_j \left[\frac{(A_iB_j) - (A_iB_j)_0}{(A_iB_j)_0} \right]^2$$

follows χ^2-distribution with $(r-1) \times (s-1)$ dof

r = rows value.

s = columns value.

1) 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.

	A1	A2	
	Infection	No infection	Total
B_1 administering the drug	144 (A_1B_1)	312 (A_2, B_1)	456
B_2 without administering the drug	192 (A_1B_2)	72 (A_2, B_2)	264
Total	336	384	720

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

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	$(O-E)^2$
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

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

$$= 4733.44[0.0047 + 0.0081 + 0.0041 + 0.0071]$$

$$= 4733.44 \times 0.0240$$

$$= 113.60256.$$

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 & we conclude that the drug is certainly effective in containing typhoid.

8) Data on the hair colour and the eye colour are given in the table. calculate the χ^2 value. Determine the association between the hair colour and the eye colour.

		Fair	Brown	Black	Total.
Eye colour	Blue	15	20	5	40
	Grey	20	20	10	50
	Brown	25	20	15	60
	Total	60	60	30	150.

Sol: (iii) We set up null hypothesis as the 2 attributes association between hair colour & eye colour on same.

Under Ho the expected frequency are

(i) $E(15) = \frac{40\times60}{150} = 160.$ (v) $E(20) = \frac{50\times60}{150} = 20$

(ii) $E(20) = \frac{50\times60}{150} = 20.$ (vi) $E(20) = \frac{60\times60}{150} = 24$

(iii) $E(25) = \frac{60\times60}{150} = 24$ (vii) $E(5) = \frac{40\times30}{150} = 8$

(iv) $E(20) = \frac{40\times60}{150} = 16$ (viii) $E(10) = \frac{50\times30}{150} = 10$

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

Compute χ^2 is

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

≃ 3.65.

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

$= 3.65$

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

degree of freedom (dof) = $r-1 \times s-1$

$= 3-1 \times 3-1$

$= 2 \times 2$

$= 4$

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 3.65, So, the project is rejected at 4 degree of freedom: with 5% level of significance.

* According to rates correction ⇒ (2x2)

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

Sign test for paired data :-

The sign test is the oldest of all non-parametric procedures and it was introduced by "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 : $p = 1/2$ that

$$z = \frac{|r - n/2|}{\sqrt{n/4}}$$

where, r = no. of positive signs.

n = Total no. of signs [Except '0'].

Single data (or) ordinary sign test :-

1) The following data in turns are the amounts of sulphur 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 $\mu_0 = 21.5$

Sol: Let $x_1, x_2, ---- x_n$ be the values of the sample size 'n' we want to test $H_0 = \mu = \mu_0$.

compare each of the 'n' values of the sample with μ_0. 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 differences when compare with $\mu = 21.5$.

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

r = no. of positive signs = 16

we want to test $H_0 : \mu_0 = 21.5$

under H_0, the test statistic is $\left(z_1 = \frac{|r - n/2|}{\sqrt{n/4}} \right) = \frac{|16-20|}{\sqrt{10}}$

critical value of z at 5% is 1.96.

we accept the H_0.

Paired Data:

1) 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 1 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, the 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
without 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	−	−	−	−	−	0	−	+	−	−	−	+	−

[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 '0']

under H_0 test statistic is $Z = \dfrac{\left|r - \frac{n}{2}\right|}{\sqrt{n/4}}$

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

$$= \frac{\left| 2 - \frac{12}{2} \right|}{\sqrt{12/4}} = \frac{|2-6|}{\sqrt{3}} = \frac{4}{\sqrt{3}} = \frac{4}{1.73} = 2.31.$$

Since, calculated value $Z = 2.31$ is greater than the critical value $Z = 1.96$ at 5% level of significance they H_0 is rejected.

Code: 17E00105

MBA & MBA (Finance) I Semester Regular & Supplementary Examinations December/January 2018/19

STATISTICS FOR MANAGERS

(For students admitted in 2017 & 2018 only)

Time: 3 hours Max. Marks: 60

SECTION – A

(Answer the following: (05 X 10 = 50 Marks)

*****

1 Discuss the application of dispersion measures for business decision making.

OR

2 A security analyst studied hundred companies and obtained the following data for the year 1997:

Dividend declared (%)	0-8	8-16	16-24	24-32	32-40
Number of companies	15	30	40	10	5

Calculate the standard deviation of the dividend declared.

3 Obtain the lines of regression from the following data:

x:	16	12	10	14	18
y:	19	11	15	18	17

OR

4 Define regression. Explain its properties and applications.

5 What are the properties of Poisson distribution?

OR

6 Potassium blood levels in healthy humans are normally distributed with a mean of 17.0 mg/100 ml, and standard deviation of 1.0 mg/100 ml. Elevated levels of potassium indicate an electrolyte balance problem, caused by Addison's disease. However, a test for potassium level should not cause too many "false positives". What level of potassium should we use so that only 2.5% of healthy individuals are classified as "abnormally high"?

7 Explain the steps involved in testing the hypothesis. What are the possible errors that may occur while testing the hypothesis?

OR

8 Explain the different types of ANOVA. What are the steps involved in carrying out ANOVA?

9 How chi-square is calculated? Explain any two of its applications.

OR

10 Explain the different types of non parametric tests.

SECTION – B

(Compulsory question, 01 X 10 = 10 Marks)

11 **Case Study:**

The following are the details of sales effected by three sales persons in three door-to-door campaigns.

Sales person	Sales in door – to – door campaign			
A	8	9	5	10
B	7	6	6	9
C	6	6	7	5

Construct an ANOVA table and find out whether there is any significant difference in the performance of the sales persons.

*****

Code: 14E00105

MBA I Semester Supplementary Examinations December/January 2018/19

BUSINESS STATISTICS

(For students admitted in 2014 (LC), 2015 & 2016 only)

Time: 3 hours Max. Marks: 60

SECTION – A

(Answer the following: (05 X 10 = 50 Marks)

*****

1 Enumerate the methods of measuring dispersion and state the characteristics of a good measure of dispersion.

OR

2 The coefficient of variation of wages of male workers and female workers are 55% and 70% respectively, while the standard deviations are 22.0 and 15.4 respectively. Calculate the overall average wages of all workers given that 80% of the workers are male.

3 State the properties of Karl Pearson's coefficient of correlation and explain how would you internet the value of r with suitable example.

OR

4 Find the coefficient of correlation by Karl Pearson's method from the following table.

X	6	2	10	4	8
Y	9	11	?	8	7

Arithmetic means of X and Y are 6 and 8 respectively.

5 State the important characteristics and properties of binomial distribution. Under what conditions can a binomial distribution be applied?

OR

6 The following table shows the distribution of number of faulty units produced in a single shift in a factory. The data is for 400 shifts.

No. of faults	0	1	2	3	4
No. of shifts	138	161	69	27	5

Fit a Poisson distribution to the data.

7 In a large city A, 20% of the random sample of 1000 school children had defective eye sight. In another large city B, 15% of a random sample of 2000 children had the same defect. Is this difference between two proportions significant? Obtain 95% confidence limits for the difference in the population proportions.

OR

8 Two random samples were drawn from two normal population and their values are:

A	66	67	75	76	82	84	88	90	92		
B	64	66	74	78	82	85	87	92	93	95	97

Test whether the two population have the same variance at 5% level of significance.

(F = 4.30 at 5% level for v_1 = 10 and v_2 = 8)

Contd. in page 2

9 The number of car accidents in a city was found as 20, 17, 12, 6, 7, 15, 8, 5, 16 and 14 per month. Use Chi-square test to check whether these frequencies are in agreement with the belief that occurrence of accidents was the same during the 10 month period. Test at 5% level of significance.

OR

10 In a survey of 200 girls of which 40% were intelligent, 30% had uneducated fathers, while 20% of the unintelligent girls had educated fathers. Do these figures support the hypothesis that educated fathers have intelligent girls? Test at 5% level of significance. (Table value of χ^2 = 3.84)

SECTION – B

(Compulsory question, 01 X 10 = 10 Marks)

11 **Case Study:**

In a certain factory production can be accomplished by four different workers on 5 different types of machines. A sample study, in context of a two-way design without repeated values, is being made with two-fold objectives of examining whether the four workers differ from with respect to mean productivity and whether the mean productivity is the same for the 5 different machines. The researcher involved in this study reports while analyzing the data as under.

(i) Sum of squares for variance between machines = 35.2
(ii) Sum of squares for variance between work man = 53.8
(iii) Sum of square for total variance = 174.2

Set up ANOVA table for the given information and draw the inference about variances at 5% level of significance (Table value F = 2.53)

*****

Code: 17E00105

MBA I Semester Regular Examinations December/January 2017/2018

STATISTICS FOR MANAGERS

(For students admitted in 2017 only)

Time: 3 hours Max. Marks: 60

SECTION – A

(Answer the following: (05 X 10 = 50 Marks)

*****

1 Explain the measures of central tendency for business decision making.

OR

2 Find standard deviation from the following data:

Values	5	10	15	20	25	30	35
Frequency	2	7	11	15	18	4	1

3 Calculate the coefficient of correlation of the following data:

x	2	3	4	5	6
y	7	9	10	14	15

OR

4 Compute rank correlation from the following table.

x	415	434	420	430	424	428
y	330	332	328	331	327	325

5 Difference between binomial and Poisson distribution.

OR

6 Fit a Poisson distribution to the following data and find out theoretical or expected frequencies.

x	0	1	2	3	4	5	6	7
f	48	72	99	73	43	20	8	2

7 Test the significance difference between simple mean and the population mean.

OR

8 The average hourly wage of a sample of 150 workers in plant A is Rs. 256 with a standard deviation of Rs. 1.08. Average wage of a sample of 200 workers in plant B Rs. 2.87 with a standard deviation of Rs. 1.28 can be applicant safely. Assume that the hourly wages paid by plant B is higher than plant A.

9 Tests are made on the proportion of defective costing produced by five different molds. If there were 14 defectives among 100 costing made with mold – I. 33 defectives among 200 costings made with mold – II. 21 defective among 180 costings made with mold – III. 17 defectives among 120 costings made with mold – IV and 25 defectives among 150 costings made with mold – V. Use the 0.01 level of significance to test whether the true proportion of defective is the same for each mold.

OR

10 The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur at equal frequency in the directory.

SECTION – B

(Compulsory question, 01 X 10 = 10 Marks)

11 Case Study:

The 3 samples given below have been obtained from a normal population with equal variance. Test the hypothesis that sample means are equal.

A	8	10	7	14	11
B	7	5	10	9	9
C	12	9	13	12	14

*****

MBA I Semester Regular & Supplementary Examinations December/January 2016/2017

BUSINESS STATISTICS

(For students admitted in 2014, 2015 & 2016 only)

Time: 3 hours Max. Marks: 60

All questions carry equal marks

(Statistical tables is permitted in the examination hall)

*****

SECTION – A

Answer the following: (05 X 10 = 50 Marks)

1 What are the various methods of measuring dispersion? Explain each one with suitable examples.

OR

2 (a) Calculate mean for the following frequency distribution.

Value	10	27	28	34	55	38	52	40	45	57
Frequency	5	6	8	9	6	5	7	4	3	5

(b) The monthly salaries of employees (in thousand rupees) is given in the following table. Compute the median salary of the employees.

Monthly salaries of employees (in thousand rupees)										
Employee	1	2	3	4	5	6	7	8	9	10
Salary	120	35	132	128	148	136	138	151	153	150

3 (a) Define and distinguish between correlation and regression.

(b) Elaborate the utility of regression analysis.

OR

4 The sales revenue and advertisement expenses of a company for the past 10 months is given in the following table. Calculate the Karl Pearson's coefficient of correlation between sales and advertisement.

Sales and advertisement expenses for 10 months (in Rs. 1000's)										
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct
Advertisement expenses	10	11	12	13	11	10	9	10	11	14
Sales	110	120	115	128	137	145	150	130	120	115

5 What is binomial distribution? What are the main assumptions of a binomial distribution? Define mean and standard deviation in a binomial distribution.

OR

6 The retail price of a 5 kg bag of white cement of a company varies from Rs. 200 per bag to Rs. 230 per bag. Assuming that these prices are uniformly distributed, (i) Compute mean, variance and standard deviation of prices of this distribution. (ii) if a price is randomly selected, what is the probability that this price is in between Rs. 210 to Rs. 225? (iii) Compute the probability that this price is less than or equal to Rs. 227.

7 Define and briefly explain the following terms:

(a) Independent variable.
(b) Treatment variable.
(c) Classification variable.
(d) Experimental units.
(e) Dependent variables.

OR

8 A firm allows its employees to pursue additional income-earning activities such as consultancy, tuitions, etc. in their out-of-office hours. The average weekly earnings through these additional income earning activities is Rs. 5000 per month per employee. A new HR manager who has recently joined the firm feels that this amount may have changed. For verifying his doubt, he has taken a random sample of 45 employees. The sample mean is computed as Rs. 5500 and the sample standard deviation is computed as Rs. 1000. Use $\alpha = 0.10$ to rest whether the additional average income has changed in the population.

Contd. in page 2

9 What is χ^2 – distribution? What is its importance in business decision making?

OR

10 A company is trying to improve the work efficiency of its employees. It has organized a special training programme for all employees. In order to assess the effectiveness of the training programme, the company has selected 10 employees randomly and administered a well-structured questionnaire. The scores obtained by the employees are given below.

S.No	Before training	After training
1	30	35
2	32	34
3	37	31
4	34	33
5	36	33
6	33	37
7	29	37
8	33	42
9	30	40
10	32	43

At 95% confidence level, examine whether the training programme has improve the efficiency of employees.

SECTION – B

(Compulsory Question) 01 X 10 = 10 Marks

11 **Case study:**

A company organized a training programme for three categories of officers: sales managers, zonal managers and regional managers. The company also considered the educational level of the employees. Based on their qualifications, officers were also divided into three categories: graduate, post graduate and doctorate. The company wants to ascertain the effectiveness of the training programme on employees across designation and educational levels. The scores obtained from randomly selected employees across different categories are given below.

		Designation		
		Sales managers	Zonal managers	Regional managers
Qualification	Graduate	30	34	38
		40	40	39
		42	42	40
		33	45	42
	Post Graduate	35	36	40
		39	38	43
		41	42	41
		39	43	32
	Doctorate	34	44	30
		38	45	28
		39	37	32
		35	38	29

Employ a two-way analysis of variance to determine whether there is significant difference in effects. Take α = 0.05.

*****

Code: 14E00105

MBA I Semester Regular & Supplementary Examinations December/January 2015/2016

BUSINESS STATISTICS

(For students admitted in 2014 & 2015 only)

Time: 3 hours Max. Marks: 60

All questions carry equal marks

*****

SECTION – A

Answer the following: (05 X 10 = 50 Marks)

1 What is the concept of coefficient of variation? What is the application of coefficient variation in business decision making?

OR

2 (a) Find the mean, median and mode for the following set of numbers:

(i) 3, 5, 2, 6, 5, 9, 5, 2, 8 and 6.

(ii) 51.6, 48.7, 50.3, 49.5 and 48.9.

(b) From the following data, find the first and third quartiles:

Serial No.	1	2	3	4	5	6	7	8
Daily wages (in hundred rupees)	15	20	34	45	52	63	71	82

3 What are the assumptions of regression analysis? Distinguish between correlation and regression.

OR

4 Determine the line of regression for the following data, taking:

(a) X as the independent variable and Y as the dependent variable.

(b) Y as the independent variable and X as the dependent variable.

($\alpha = 0.05$)

X	12	21	28	25	32	42	43	39	55
Y	14	22	12	28	35	37	32	44	49

5 Define probability. Explain the concept of marginal probability, union probability, joint probability and conditional probability.

OR

6 In a toy manufacturing company, three machines namely, A, B and C, are employed to manufacture toys. Machines A, B and C manufacture 20%, 30% and 50% of the total toys, respectively. A quality control officer examined the machines and found that A, B and C produce 2%, 3% and 5% defectives of the total output. A toy is selected at random and is found to be defective. What are the probabilities that this toy came from machine A, B and C respectively.

7 What is hypothesis? Discuss the hypothesis testing procedure.

OR

8 Modern bicycles has conducted a survey among 100 randomly selected men and 120 randomly selected women. As per the findings, 25 men and 35 women say that the size of the wheel is a very important factor in purchasing a bicycle. On the basis of this data, can the company claim that a significantly higher proportion of women when compared to men believe that the size of wheels is a very important factor. Take 95% as the confidence level.

9 (a) What is the χ^2 goodness-of-fit test and what are its applications in decision making?

(b) Under what circumstances is the χ^2 test of independence used?

Contdֵ. in Page 2

OR

10 A vice president (sales) of a garment company wants to determine whether the sales of the company's brand of jeans is independent of age group. He has appointed a marketing researcher for this purpose. This marketing researcher has taken a random sample of 703 consumers who have purchased jeans. The researcher conducted survey for three brands of the jeans, namely brand 1, brand 2 and brand 3. The researcher has also divided the age groups into four groups: 15 to 25, 26 to 2, 26 to 45 and 46 to 55. The observations of the researcher are provided in the following table:

Brand Age	Brand 1	Brand 2	Brand 3	Row Total
15 to 25	65	75	72	212
26 to 35	60	40	64	164
36 to 45	45	52	50	147
46 to 55	55	65	60	180
Column total	225	232	246	703

Determine whether brand preference is independent of age group. Use α = 0.05.

SECTION – B

(Compulsory Question) 01 X 10 = 10 Marks

11 **Case study:**

A dealer of a motor cycle company believes that there is a positive relationship between the number of salespeople employed and the increase in the sales of bikes. Data for 14 randomly selected weeks are given in the following table:

Weeks	No. of salespeople employed	Sales (in units)
1	17	34
2	14	39
3	25	60
4	40	80
5	15	38
6	18	50
7	13	35
8	11	25
9	27	51
10	12	29
11	38	89
12	36	85
13	41	90
14	28	63

Questions:

(a) Develop a regression model to predict sales from the number of salespeople employed.

(b) Predict sales when number of sales people employed are 100.

*****

Code: 14E00105

MBA I Semester Supplementary Examinations August 2015

BUSINESS STATISTICS

(For students admitted in 2014 only)

Time: 3 hours Max. Marks: 60

Issue of T, F, χ^2, Z values tables at 5% level of significance are permitted in the examination hall

All questions carry equal marks

*****

SECTION – A

Answer the following: (05 X 10 = 50 Marks)

1 Brief out various measures of dispersion.

(OR)

2 Calculate standard deviation from the following data:

Class interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	5	7	16	27	39	53	18	45

3 Calculate correlation coefficient between X and Y series:

X	78	89	96	69	59	79	68	61
Y	125	137	156	112	107	136	123	105

(OR)

4 Discuss the concept and advantages of regression analysis.

5 What is probability? Brief out the significance of probability in business applications.

(OR)

6 Fit a binomial distribution by using direct method for the following data:

X	0	1	2	3	4	5	6	7	8
Frequency	17	64	140	210	132	75	45	56	80

7 Distinguish the features and purpose of ANOVA one and two way classification.

(OR)

8 An IQ test was conducted to 5 persons before and after they were trained. The results are given below.

Candidates	I	II	III	IV	V
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

9 Calculate Chi-Square test from the data given below:

Observed frequency	60	75	50	92	46	74	86	48	94	85
Expected frequency	68	72	43	103	63	35	94	32	75	93

(OR)

10 Brief out the Non-Parametric methods of statistics.

SECTION – B

(Compulsory Question) 01 X 10 = 10 Marks

11 **Problem**:

Analyze one way classification from the following data:

1	10	10	45	44	8	13	41	43
2	29	30	10	8	33	27	12	10
3	37	33	26	27	32	36	27	30
4	39	40	31	32	42	42	32	32

*****

Code: 14E00105

MBA I Semester Regular Examinations February/March 2015

BUSINESS STATISTICS

(For students admitted in 2014 only)

Time: 3 hours Max. Marks: 60

All questions carry equal marks

Use of T, F, χ^2 and Z value tables at 5% level of significance are permitted.

*****

SECTION – A

Answer the following: (05 X 10 = 50 Marks)

1 What are the measures of central tendency? Explain the need and advantages of central tendency.

(OR)

2 Calculate standard deviation from the following data:

Class interval	0-5	5-10	10-15	15-20	20-25	25-30	30-35	40-45
Frequency	9	16	12	26	14	12	6	5

3 Define regression analysis. Brief out the significance and types of regression analysis.

(OR)

4 Calculate rank correlation coefficient between X and Y series.

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

5 Fit a Poisson distribution for the following data by using recurrence relation model.

No. of deaths	0	1	2	3	4	5	6	7
Frequency	305	365	210	80	28	9	2	1

(OR)

6 Explain different theories of probability.

7 What is meant by hypothesis? Explain various tests for testing hypothesis.

(OR)

8 The time taken by workers in performing a job by Method-I and Method-II are given below.

Method – I	20	16	26	27	23	22	25
Method – II	27	33	42	35	32	34	38

9 What is Chi-square test? Explain the features and applications of chi-square test.

(OR)

10 The following table shows the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Using Chi-square test whether the digits may be taken to occur equally frequently in the directory at 5% level.

SECTION – B

(Compulsory Question) 01 X 10 = 10 Marks

11 Case study:

A stenographer claims that she can take the dictation at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 words with a standard deviation of 15 words? Use Z-test at 5% level of significance.

*****