Unit-4

2 mark questions

1 What is Hypothesis? Give an Example.

→ A hypothesis is an assumption that is made based on some evidence. In other words, hypothesis are tentative explanations of a principle operating in nature.

Example, The field of business, decision makers are continually attempting to find answers to questions

2 What are the types of hypothesis?

→ ① Research hypothesis
② Statistical hypothesis
③ Substantive hypothesis

3. What are Type-1 error? Give an example.

→ A Type-1 error is committed by rejecting a true null hypothesis. With a Type-1 error, the null hypothesis is true, but the business researcher decides that it is not.

Example, if a manager fires an employee because some evidence indicates that she is stealing from the company and if she really is not stealing from the company, then the manager has committed a Type I error

4 What are null and alternate hypothesis? Give example.

→ The null hypothesis states that the "null" condition exists; that is there is nothing new happening, the old theory is still true, the old standard is correct and the system is in control.

The alternative hypothesis, on the other hand, states that the new theory is true, there are new standards, the system is out of control, and/or something is happening.

Example, suppose flour packaged by a manufacturer is sold by weight; and a particular size of package is supposed to average 40 ounces. Suppose the manufacturer wants to test to determine whether their packaging process is out of control as determined by the weight of the flour packages. The null hypothesis for this experiment is that the average weight of the flour packages is 40 ounces (no problem). The alternative hypothesis is that the average is not 40 ounces (process is out of control)

5. What are Type-II error? Give example

→ A Type-II error is committed when a business researcher fails to reject a false null hypothesis. In this case, the null hypothesis is false, but a decision is made to not reject it

Example, Suppose in the business world an employee is stealing from the company. A manager sees some evidence that the stealing is occuring but lacks enough evidence to conclude that the employee is stealing from the company. The manager decides not to fire the employee based on theft. The manager has committed a Type-II error

6. Show the relationship between α, β and power

→ α → The probability of committing a Type I error is called alpha (α) or level of significance. Alpha equals to the area under the curve that is in the rejection region beyond the critical value(s). The value of alpha is always set before the experiment or study is undertaken. The common values of alpha are .05, .01, .10 and .001

β → The probability of committing a Type-II error is beta (β). Unlike alpha, beta is not usually state at the beginning of the hypothesis testing procedure. Actually, beta occurs only when the null hypothesis is not true. the computation of beta varies with the many possible alternative parameters that might occur power → It is equal to $1-\beta$, is the probability of statistical test rejecting the null hypothesis when the null hypothesis is false.

7 What are One Sample t Test? List its uses

→ The one sample t test examines whether the mean of a population is statistically different from a known or hypothesized value. The one sample t test is a parametric test.

Uses :- • Statistical difference between a mean and a known or hypothesized value of the mean in the population.
• Statistical difference between a change score and zero

8 What the t test statistics formula for one sample t test?

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

$df = n-1$

9 State the hypothesis One Sample t Test?

→ The null hypothesis (H_0) and (two-tailed) alternative hypothesis (H_1) of the one sample T test can be expressed as:

$H_0 : \mu = \mu_0$ ("The population mean is equal to the proposed population mean")

$H_1 : \mu \neq \mu_0$ ("The population mean is not equal to the proposed population mean")

where μ is the "true" population mean and μ_0 is the proposed value of the population mean.

10 What are Paired Samples t Test? List its uses

→ The Paired Samples t Test compares the means of two measurements taken from the same individual, object, or related unit. These "paired" measurements

Uses - • Statistical difference between two time points
• Statistical difference between two conditions
• Statistical difference between two measurements
• Statistical difference between a matched pair

11 What the t Test statistics formula for Paired Sample t Test

→ $$t = \frac{\bar{d} - D}{\frac{S_d}{\sqrt{n}}}$$

$df = n - 1$

12

12 State the hypothesis Paired Sample t Test.

→ The hypothesis can be expressed in two different ways that express the same idea and are mathematically equivalent

$H_0 : \mu_1 = \mu_2$ ("The paired population mean are equal")

$H_1 : \mu_1 \neq \mu_2$ ("The paired population mean are not equal")

OR

$H_0 : \mu_1 - \mu_2 = 0$ ("The difference between the paired population means is equal to 0")

$H_1 : \mu_1 - \mu_2 \neq 0$ ("The difference between the paired population means is not 0")

where, μ_1 is the population mean of variable 1

and

μ_2 is the population mean of variable 2

13 What are the independent sample t test? List its uses

→ The independent sample t Test compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different. The independent sample t Test is a parametric test.

Use - • Statistical difference between the means of two groups
• Statistical difference between the means of two interventions
• Statistical difference between the means of two change scores

14 State the hypothesis independent sample t Test
→ The null hypothesis (H_0) and alternative hypothesis (H_1) of the independent samples t Test can be expressed in two different but equivalent ways.

$H_0: \mu_1 = \mu_2$ ("The two population means are equal")

$H_1: \mu_1 \neq \mu_2$ ("The two population means are not equal")

OR

$H_0: \mu_1 - \mu_2 = 0$ ("The difference between the two population mean is equal to 0")

$H_1: \mu_1 - \mu_2 \neq 0$ ("The difference between the two population mean is not 0")

where, μ_1 and μ_2 are the population mean for group 1 and group 2 respectively.

15 What are chi-square test of independence? List its uses.
→ The chi-square test of independence determines whether there is an association between categorical variables. It is a non parametric test

Uses - • Statistical independence or association between two or more categorical variable

16 Write the t test statistic formula for chi-square test of independence.
→ $$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

17 What are ONE WAY ANOVA? List its uses

→ One way ANOVA compares the means of two or more independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different. One-way ANOVA is a parametric test.

Uses = • Statistical difference among the means of two or more groups
• Statistical difference among the means of two or more interventions
• Statistical difference among the means of two or more change scores.

18 What are the formula of MSC, MSE and F statistics for ONE WAY ANOVA.

→ $MSC = \frac{SSC}{df_c}$ $\quad MSE = \frac{SSE}{df_E}$ $\quad F = \frac{MSC}{MSE}$

19 Write the formula of SSC, SSE and SST for ONE WAY ANOVA

→ $SSC = \sum_{j=1}^{c} n_j (\bar{x}_j - \bar{x})^2$ $\quad SSE = \sum_{i=1}^{n_i} \sum_{j=1}^{c} (x_{ij} - \bar{x}_j)^2$

$SST = \sum_{i=1}^{n_i} \sum_{j=1}^{c} (x_{ij} - \bar{x})^2$ OR $SST = SSC + SSE$

20. Write the Z-formula for On Sample Proportion Test

→ $$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

21 Write the Z formula for two sample proportion test

→ $$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}}$$

22 Define correlation Analysis and regression analysis
→ Correlation analysis is a statistical technique to study the degree and direction of relationship between two or more variables.
Regression analysis is a statistical tool to study the nature and extent of functional relationship between two or more variables and to estimate the unknown values of dependent variable from the known values of independent variable.

23 What are the uses of correlation?
→ ① Correlation analysis helps in deriving precisely the degree and the direction of such relationship.
② The effect of correlation is to reduce the range of uncertainity of our prediction. The prediction based on correlation analysis will be more reliable and near to reality.
③ The measure of coefficient of correlation is a relative measure of change.

Q4. List the types of correlation.
→ ① Positive and Negative
② Simple, Partial and Multiple.
③ Linear and Non-Linear.

Q5 What are Positive Correlation? Give Example.
→ If both the variables vary in the same direction, correlation is said to be positive. It means if one variable is increasing, the other on an average is also increasing or if one variable is decreasing, the other on an average is also decreasing.
Example, The correlation between heights and weights of a group of persons is a positive correlation.

26 What are negative correlation? Give Example.

→ If both the variables vary in opposite direction, the correlation is said to be negative. It means if one variable increases, but the other variable decreases, or if one variable decreases, but the other variable increases.

Example, the correlation between the price of a product and its demand is a negative correlation.

27 What are Simple correlation? Give Example

→ When only two variables are studied, it is a case of simple correlation

Example, when one studies relationship between the marks secured by student and the attendence of student in class, it is a problem of simple correlation

28 What are Partial correlation? Give Example

→ In case of partial correlation one studies three or more variables but considers only two variables to be influencing each other and the effect of other influencing variables being held constant

Example, when one studies relationship between the marks secured by student and the attendence of students in class, the other variable influencing such as effective teaching of teacher, use of teaching aid like computer, smart board etc are assumed to be constant

29 What are Multiple Correlation? Give Example

→ When three or more variables are studied, it is a case of multiple correlation.

Example, when one studies relationship between the marks secured by student and the attendence of student in class, the other variable influencing such as effective teaching of teacher, use of teaching aid like computer, smart board etc.

Then if study covers the relationship between student marks, attendance of students, effectiveness of teacher, use of teaching aids etc. it is a case of multiple correlation.

30 What are zero correlation? Give example
→ Actually it is not a type of correlation but still it is called as zero or no correlation. When we don't find any relationship between the variables then, it is said to be zero correlation. It means a change in value of one variable doesn't influence or change the value of other variable
Example, the correlation between weight of person and intelligence is a zero or no correlation.

31 What are Linear correlation? Give Example
→ If the amount of change in one variable bears a constant ratio to the amount of change in the other variable, then correlation is said to be linear.
Example, if it is assumed that to produce on unit of finished product we need 10 units of raw materials then subsequently to produce 2 units of finished product we need the double of the one unit

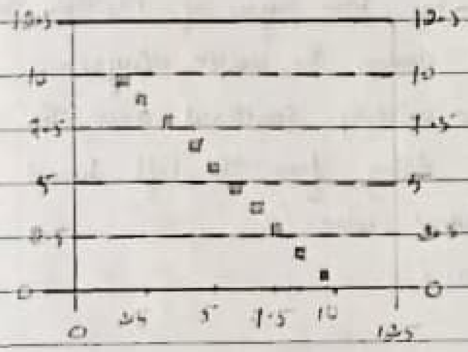
32 What are Non-Linear correlation? Give Example.
→ If the amount of change in one variable does not bear a constant ratio to the amount of change to the other variable, then correlation is said to be non-linear
Example, if we double the amount of advertisement expenditure, then sales volume would not necessarily be doubled

33 What do you mean by Perfect Positive Correlation? Write its condition and draw the scatter diagram
→ In this case, the points will form on a straight line rising from the lower left hand corner to the upper right hand corner
Condition = $r = +1$

34 What do you mean by perfect negative correlation? Write its condition and draw the scatter diagram.

→ In this case, the points will form on a straight line declining from the upper left hand corner to the lower right hand corner

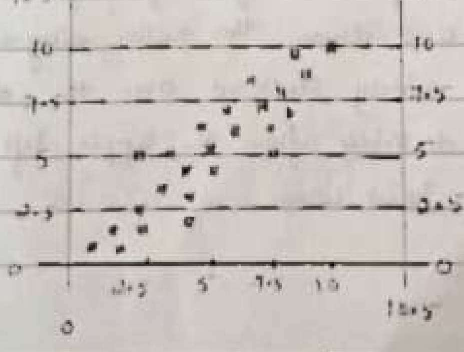
condition = $r = -1$



35 What do you mean by High Degree of Positive Correlation? Write its condition and draw the scatter diagram.

→ In this case, the plotted points fall in a narrow band, wherein points show a rising tendency from the lower left hand corner to the upper right hand corner

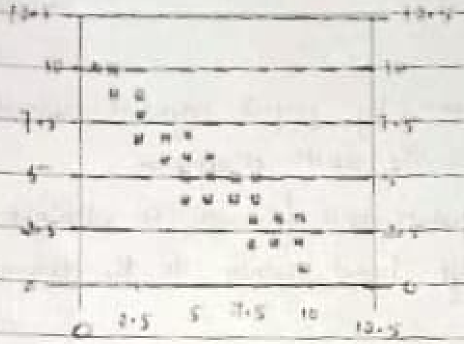
condition = $r >= 0.5$



36 What do you mean by High Degree of Negative? Write its condition and draw the scatter diagram

→ In this case, the plotted points fall in a narrow band, wherein points show a declining tendency from upper left hand corner to the lower right hand corner

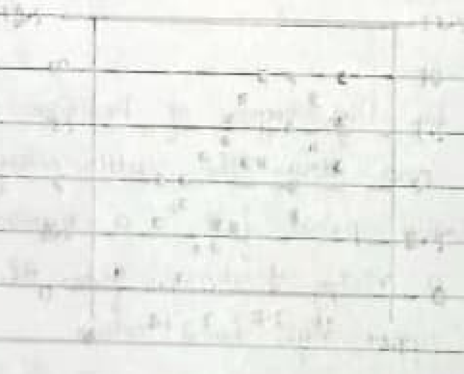
Condition - $r >= -0.5$



37 What do you mean by Low Degree of Positive Correlation? Write its condition and draw the scatter diagram

→ If the points are widely scattered over the diagrams, wherein points are rising from the left hand corner to the upper right hand corner

Condition, $r < 0.5$

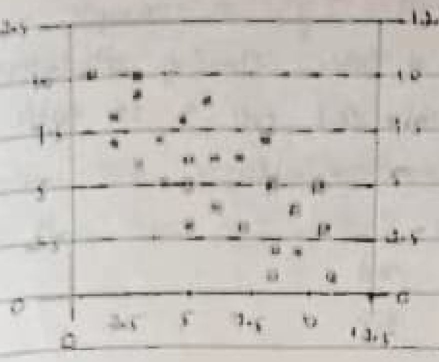


38 What do you mean by Low Degree of Negative Correlation? Write its condition and draw the scatter diagram

→ If the points are widely scattered over the diagrams, wherein points are declining from the upper left hand corner to the lower right hand corner

Condition - $r < -0.5$

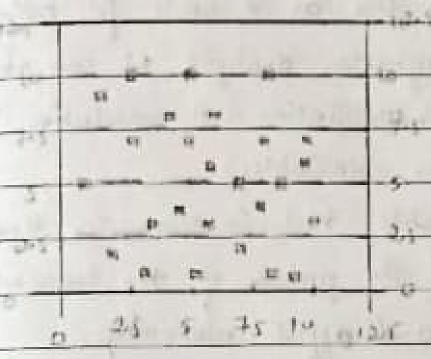
Date 04/10/23
Page 109



39 What do you mean by Zero (No) Correlation? Write its condition and draw the scatter diagram.

→ When plotted points are scattered over the graph haphazardly, then it indicate that there is no correlation or zero correlation between two variables

condition, $r = 0$



40 Write the formula to calculate Karl Pearson's coefficient of correlation method using direct method

→

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

41 Write the formula to calculate Karl Pearson's product moment coefficient of correlation

→

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

42. What is a correlation matrix ? give a example

→ A correlation matrix is a table showing correlation coefficients between variables Each cell in the table shows the correlation between two variables

Example.

	y	x1	x2
y	1	0.6	0.5
x1	0.6	1	0.7
x2	0.5	0.7	1

43 What are the uses of regression analysis?

→ ① It provides estimates of values of the dependent variables from values of independent variables

② It is used to obtain a measure of the error involved in using the regression line as a basis for estimation

③ With the help of regression analysis, we can obtain a measure of degree of association or correlation that exists between the two variables.

④ It is highly valuable tool in economics and business research, since most of the problem of the economic analysis are based on cause and effect relationship

44. Write the formula for regression equation of X on Y and Y on X.

→ X on Y ⇒ $x = a + bY$

will change to

$(x-\bar{x}) = b_{xy}(y-\bar{y})$

Y on X ⇒ $Y = a + bX$

will change to

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

Long Answer Questions (THREE, FOUR or FIVE and SIX Marks Questions)

1. Explain three types of Hypotheses with example

→ The three types of hypothesis are -
- Research hypotheses
- Statistical hypotheses
- Substantive hypotheses

Research hypotheses → Research hypotheses are most nearly like hypotheses defined earlier) A research hypotheses is a statement of what the researcher believes will be the outcome of an experiment or study. Before studies are undertaken, business researcher often have some idea or theory based on experience or previous work as to how the study will turn out. These ideas, theories, or notions established before an experiment or study is conducted are research hypotheses

Example = • Older workers are more loyal to a company
• The price of scrap metal is a good indicator of the industrial production index six month later

Statistical hypotheses → In statistics, a hypothesis is a statement or assumption about a population parameter that we want to test using sample data. There are two types of hypotheses: null hypotheses (H_0) and alternative hypotheses (H_1). The null hypotheses often represents a default or no-effect assumption, while the alternative hypothesis suggests a difference, effect or relationship

Example, • The average height of the plant species is equal to 50cm $H_0: \mu = 50$
• The average hight of the plant species is not equal to 50cm $H_1: \mu \neq 50$

4 A random of sample size 20 is taken resulting in sample mean of 25.51 and a sample standard deviation of 2.1933. Assume data is normally distributed use this information and $\alpha = 0.05$ to test the following hypothesis

$H_0 : \mu = 25$ pounds

$H_1 : \mu \neq 25$ pounds

→ $\mu = 25.00$ $s = 2.1933$ $\alpha = 0.05$ $n = 20$ $df = n-1 = 19$

$\bar{x} = 25.51$

This is 2 tailed test. So alpha (α) must be split

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

$t_{0.025, 19} = \pm 2.093$

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

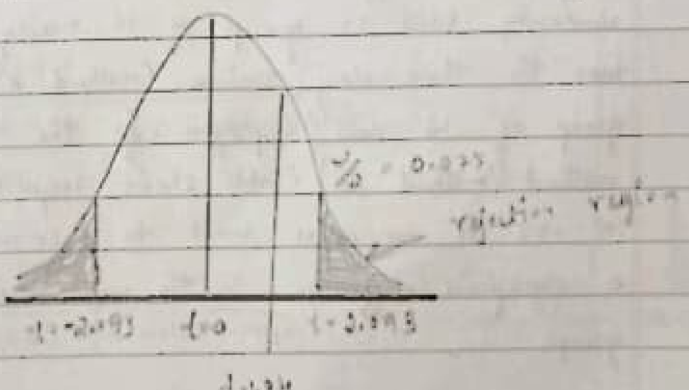
$$= \frac{25.51 - 25.00}{\frac{2.1933}{\sqrt{20}}}$$

$$= \frac{0.51}{\frac{2.1933}{4.472135955}}$$

$$= \frac{0.51}{0.490436789}$$

$= 1.039889362$

Because the observed t value is 1.04 the null hypothesis is not rejected



Substantive hypotheses → Substantive hypotheses are statements about the relationships between variables or the expected outcomes of a study

Example. • There is a significant difference in the mean scores of two groups

• There is a linear relationship between a dependent variable and one or more independent variables

2 Explain Type I and Type II errors with example

→ Refer 2 marks questions 3 and 5

3. Imagine a company wants to test the claim that their batteries last more than 40 hours. Using a simple random sample of 15 batteries yielded a mean of 44.9 hours, with a standard deviation of 8.9 hours. Test this claim using a significance level of 0.05

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

$\bar{x} = 44.9$ $\mu = 40$ $n = 15$ $df = n-1 = 14$ $s = 8.9$

$t = \frac{44.9 - 40}{\frac{8.9}{\sqrt{15}}}$

$= \frac{4.9}{\frac{8.9}{3.872983346}}$

$= \frac{4.9}{2.297970119}$

$= \underline{2.132316674}$

$t_{0.05, 14} = 1.761$

$2.13 > 1.761$

Reject Null Hypothesis

5 A random sample of size 20 is taken, resulting in a sample mean of 16.45 and a sample standard deviation of 3.59. Assume x is normally distributed and use this information on $\alpha = 0.05$ to test the following hypotheses

$H_0: \mu = 16 \quad H_1: \mu \neq 16$

→ $\bar{x} = 16.45 \quad n = 20 \quad df = n-1 = 19 \quad \mu = 16 \quad S = 3.59 \quad \alpha = 0.05$

two tailed

$\alpha = 0.05/2$

$\underline{0.025}$

$\therefore t_{0.025, 19} = 2.093$

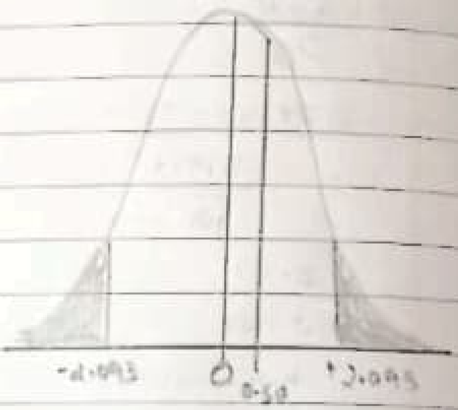
$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{16.45 - 16}{\frac{3.59}{\sqrt{20}}}$$

$$= \frac{0.45}{\frac{3.59}{4.472135955}}$$

$$= \frac{0.45}{0.802748403}$$

$$= \underline{0.560574144}$$



Since the observed t value (0.56) is lower than the table value the null hypothesis is not rejected

6 To test the difference in the two methods, the manager randomly select one group of 15 newly hired employees to take the three-day seminar (method - A) and a second group of 12 new employees for the two-day DVD method (method B). Table shows required data. Using $\alpha = .05$, the management want to determine whether there is a significant difference in the mean scores of the two groups

Method A	Method B
$\bar{x}_1 = 47.73$	$\bar{x}_2 = 56.5$
$s_1^2 = 19.495$	$s_2^2 = 18.273$
$n_1 = 15$	$n_2 = 12$

→ $\alpha = 0.05$

$df = n_1 + n_2 - 2 = 15 + 12 - 2 = \underline{25}$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$H_0 : \mu_1 - \mu_2 = 0 \qquad H_a : \mu_1 - \mu_2 \neq 0$

it is a two tailed test so $\alpha/2 = 0.05/2 = 0.025$

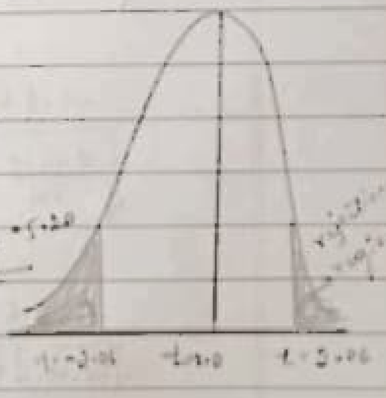
$t_{0.025,\,25} = \pm 2.060$

$$t = \frac{(47.73 - 56.5) - (0)}{\sqrt{\frac{(19.495)(15-1) + (18.273)(12-1)}{15+12-2}}\sqrt{\frac{1}{15} + \frac{1}{12}}}$$

$$= \frac{-8.77 - 0}{\sqrt{\frac{(19.495)(14) + (18.273)(11)}{25}}\sqrt{0.066666666 + 0.083333333}}$$

$$= \frac{-8.77}{\sqrt{\frac{272.93 + 201.003}{25}}\sqrt{0.149999999}}$$

$$= \frac{-8.77}{\sqrt{\frac{473.933}{25}}\sqrt{0.149999999}}$$



∴ The null hypothesis is rejected as observed t value is less than the critical t value

$$= \frac{-8.77}{\sqrt{18.95732}\sqrt{0.149999999}}$$

$$= \frac{-8.77}{4.354000459 \times 0.387298333} = \frac{-8.77}{1.68629712} = \underline{-5.20074422}$$

7 Use the data given and the eight step process to test the following hypothesis. Use 1% level of significance

$H_0 : \mu_1 - \mu_2 = 0$ $\quad$ $H_a : \mu_1 - \mu_2 < 0$

Sample 1	Sample 2
$n_1 = 8$	$n_2 = 11$
$\bar{x}_1 = 24.56$	$\bar{x}_2 = 26.42$
$s_1^2 = 12.4$	$s_2^2 = 15.8$

→ $\alpha = 0.01$

$df = n_1 + n_2 - 2 = 8 + 11 - 2 = 17$

two-tailed test so $0.01/2 = 0.005$

$t_{0.005, 17} = 2.898$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(24.56 - 26.42) - 0}{\sqrt{\frac{12.4(8-1) + 15.8(11-1)}{8 + 11 - 2}} \sqrt{\frac{1}{8} + \frac{1}{11}}}$$

$$= \frac{-1.86}{\sqrt{\frac{12.4(7) + 15.8(10)}{17}} \sqrt{0.125 + 0.09090909}}$$

$$= \frac{-1.86}{\sqrt{\frac{86.8 + 158}{17}} \sqrt{1.375000014}}$$

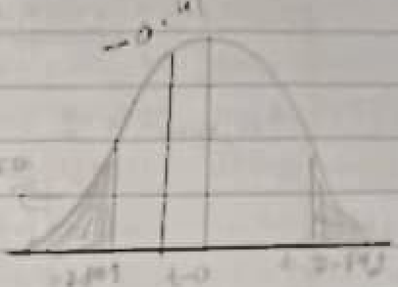
$$= \frac{-1.86}{\sqrt{\frac{244.8}{17}} \sqrt{1.375000014}}$$

$$= \frac{-1.86}{\sqrt{14.4} \sqrt{1.375000014}}$$

$= \frac{-1.86}{3.7947331922 \times 1.1726039446}$

$= \frac{-1.86}{4.449719113}$

$= -0.418003912$



The null hypotheses is not rejected because the observed t value is greater than table value

8 To test this we may recruit a simple random sample of 20 college basketball players and measure each of their max vertical jump. Then, we may have each player use the training program for one month and then measure their max vertical jump again at the end of the month. To determine whether the training program increase max vertical jump, we will perform a paired samples t-test at significance level $\alpha = 0.05$, sample mean of the differences is -0.95 and sample standard deviation of the differences is 1.317

→ $\bar{d} = -0.95$ $s_d = 1.317$

$H_0 : D = 0 \; (\mu_1 - \mu_0) = 0$ (no. difference)

$H_a : D < 0 \; (\mu_1 - \mu_0) < 0$ (difference of population mean)

$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$

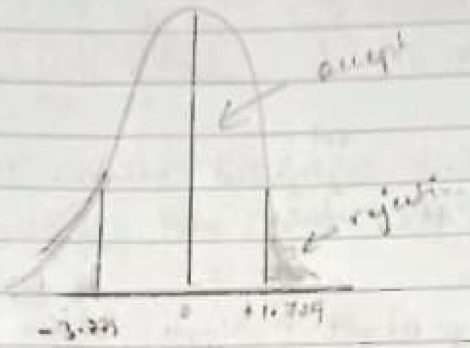
$= \frac{-0.95 - 0}{\frac{1.317}{\sqrt{20}}}$

$= \frac{-0.95}{\frac{1.317}{4.472135955}}$

$$= \frac{-0.95}{0.294449012}$$

$= -3.225914318$

$t_{19,\,0.05} = +1.729$ (table value)



This sample has not enough evidence to reject the hypotheses. So null hypotheses is accepted

9 Suppose a stock market investor is interested in determining whether there is a significant difference in the P/E (Price to earnings) ratio for companies from one year to the next. Assume $\alpha = .01$ Assume that difference in P/E ratios are normally distributed in the population. $n = 9$

Company	Year 1 P/E Ratio	Year 2 P/E Ratio
1	8.9	12.7
2	38.1	45.4
3	43.0	10.0
4	34.0	27.2
5	34.5	22.8
6	15.2	24.1
7	20.3	32.3
8	19.4	40.1
9	61.9	106.5

→ $n = 9$ $\alpha = 0.01$ $df = n-1 = 8$

$H_0 : D = 0$ $t_{\alpha/2} = t_{0.01/2} = t_{0.005}$

$H_a : D \neq 0$ $t_{0.005,8} = \pm 3.355$

Company	Year 1	Year 2	d	$d-\bar{d}$	$(d-\bar{d})^2$
1	8.9	12.7	-3.8	1.23333333	1.52111111
2	38.1	45.4	-7.3	-2.26666667	5.137777779
3	43.0	10.0	33	38.03333333	1446.534444
4	34.0	27.2	6.8	11.83333333	140.0277777
5	34.5	22.8	11.7	16.73333333	280.0044443
6	15.2	24.1	-8.9	-3.86666667	14.95111111
7	20.3	32.3	-12	-6.96666667	48.53444445
8	19.9	40.1	-20.2	-15.16666667	230.0277779
9	61.9	106.5	-44.6	-39.56666667	1565.521111
			-45.3		3732.259999

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-45.3}{9} = -5.033333333$$

$$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{3732.259999}{8}} = \sqrt{466.5324999} = 21.59936341$$

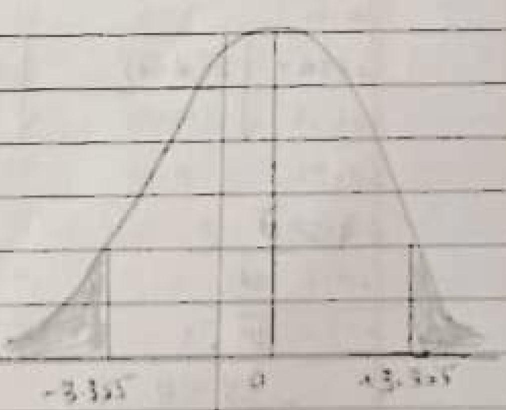
$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}}$$

$$= \frac{-5.033333333 - 0}{\frac{21.59936341}{\sqrt{9}}}$$

$$= \frac{-5.033333333}{\frac{21.59936341}{3}}$$

$$= \frac{-5.033333333}{7.199787803}$$

$$= -0.699094677$$



the sample has no enough evidence to reject the null hypothesis. Hence it is accepted.

10 Use the data given and a 1% level of significance to test the following hypotheses. Assume the differences are normally distributed in the population.

$H_0 : D = 0$ $\qquad$ $H_a : D > 0$

Pair	Sample 1	Sample 2
1	38	22
2	27	28
3	30	21
4	41	38
5	36	38

→

$n = 5 \quad \alpha = 0.01 \Rightarrow 0.01/2 = 0.005 \quad df = 5-1 = 4$

$H_0 : D = 0$

$H_a : D \neq 0$ $\qquad$ $t_{.005,4} = \pm 4.604$

Pair	Sample 1	Sample 2	d	$d - \bar{d}$	$(d-\bar{d})^2$
1	38	22	16	11	121
2	27	28	-1	-6	36
3	30	21	9	4	16
4	41	38	3	-2	4
5	36	38	-2	-7	49
			25		226

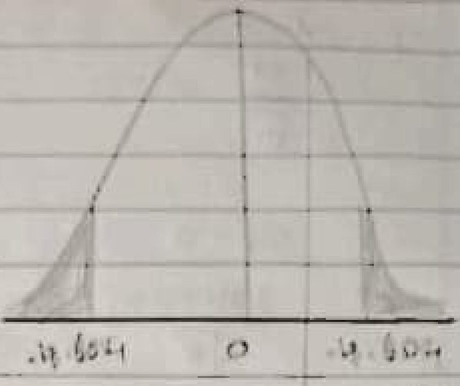
$$\bar{d} = \frac{\Sigma d}{n} = \frac{25}{5} = 5$$

$$sd = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{226}{4}} = \sqrt{56.5} = 7.516648189$$

$$t = \frac{\bar{d} - D}{\frac{sd}{\sqrt{n}}}$$

$$= \frac{5-0}{\frac{7.516648189}{\sqrt{5}}} = \frac{5}{\frac{7.516648189}{2.236067977}} = \frac{5}{3.361547263}$$

$$= 1.487410293$$



The sample has no enough evidence to reject the null hypothesis. Hence it is accepted.

11 Use the data given to test the following hypotheses. Assume the difference are normally distributed in the population

$H_0 : D = 0$ $\quad\quad$ $H_a : D \neq 0$

Individual	Before	After
1	107	102
2	99	98
3	110	100
4	113	108
5	96	89

→ $n = 5$ $\quad$ $\alpha = 0.05$ $\quad$ $df = 5 - 1 = 4$ $\quad$ $\alpha/2 = 0.05/2 = 0.025$

$H_0 : D = 0$ $\quad\quad$ $H_a : D \neq 0$

$t_{0.025, 4} = 2.776$

Individual	Before	After	d	$d - \bar{d}$	$(d - \bar{d})^2$
1	107	102	5	-0.6	0.36
2	99	98	1	-4.6	21.16
3	110	100	10	4.4	19.36
4	113	108	5	-0.6	0.36
5	96	89	7	1.4	1.96
			28		43.2

$\bar{d} = \frac{\Sigma d}{n} \quad \frac{28}{5} = 5.6$

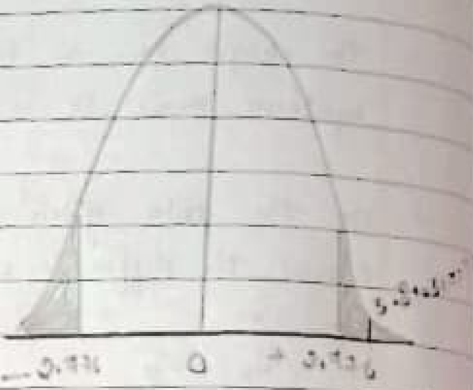
$sd = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{43.2}{4}} = \sqrt{10.8} = 3.286335345$

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$= \frac{5.6 - 0}{\frac{3.286335345}{\sqrt{5}}}$$

$$= \frac{5.6}{\frac{3.286335345}{2.236067977}}$$

$$= \frac{5.6}{1.469693846}$$



$= \underline{3.810317377}$

∴ The sample has an enough evidence to reject the null hypothesis. Hence it is rejected.

12 Suppose a store manager wants to find out whether the results of this consumer survey apply to customers of supermarkets in her city. To do so, she interviews 207 randomly selected consumers as they leave supermarkets in various parts of the city. Now the manager can use a chisquare test to determine whether the observed frequencies of responses from this survey are the same as the frequencies that would be expected on the basis of the national survey. ($\alpha = .05$)

Results of a Local Survey of Consumer Satisfaction

Response	Frequency (f_o)
Excellent	21
Pretty good	109
Only fair	62
Poor	15

Expected %

Excellent	8%
Pretty good	47%
Only fair	34%
Poor	11%

→

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

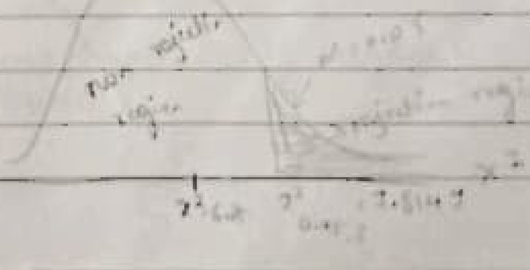
$\alpha = 0.05$ $k = 4$ $df = k-1 = \underline{3}$ $n = 207$

$\chi^2_{0.05,\,3} = \underline{7.815}$ (table)

Response	Expected proportion	Expected frequency (f_e) (proportion x sample total)
Excellent	0.08	(0.08)(207) = 16.56
Pretty good	0.47	(0.47)(207) = 97.29
fair	0.34	(0.34)(207) = 70.38
poor	0.11	(0.11)(207) = 22.77
		207

Response	f_o	f_e	$(f_o - f_e)^2 / f_e$
Excellent	21	16.56	1.190434783
Pretty good	109	97.29	1.409436736
fair	62	70.38	0.997789144
poor	15	22.77	2.651422925
			6.249053588

The observed value of chisquare of 6.25 is not greater than the critical table value of 7.8147. therefore the store manager will not reject the null hypothesis



13 Use chi-square test to determine whether the observed frequencies are distributed the same as the expected frequencies ($\alpha = 0.05$)

category	f_o	f_e
1	53	68
2	37	42
3	32	33
4	28	22
5	18	10
6	15	8

→

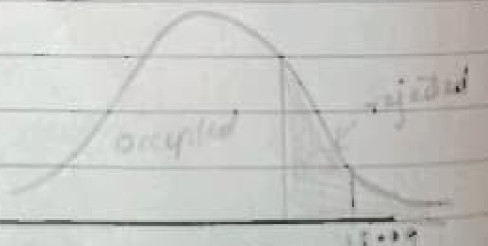
$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$\alpha = 0.05$ $k = 6$ $df = k-1 = 5$ $n = 183$

$$\chi^2_{0.05, 5} = 12.592$$

category	f_o	f_e	$(f_o - f_e)^2/f_e$
1	53	68	3.308823529
2	37	42	0.595238095
3	32	33	0.03030303
4	28	22	1.636363636
5	18	10	6.4
6	15	8	6.125
			18.09572829

The observed value of chisquare is 18.09 it is greater than the critical value which is 12.592. Therefore the null hypothesis is rejected



14 Use chi square test to determine whether the observed frequencies represent a uniform distribution ($\alpha = 0.01$)

category	f_o
1	19
2	17
3	14
4	18
5	19
6	21
7	18
8	18

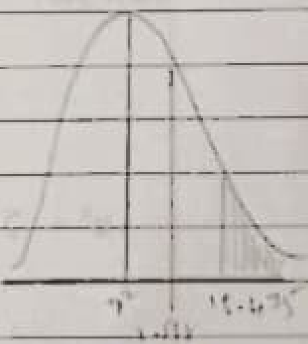
→

$$f_e = \frac{\Sigma f_o}{n} = \frac{144}{8} = 18$$

category 1 $k = 8$ $df = k-1 = 7$ $n = 144$.

$$\chi^2_{0.01, 7} = 18.475$$

category	f_o	f_e	$(f_o - f_e)^2/f_e$
1	19	18	0.055555555
2	17	18	0.055555555
3	14	18	0.888888888
4	18	18	0
5	19	18	0.055555555
6	21	18	0.5
7	18	18	0
8	18	18	0
			1.555555553



The observed value of chi-square is 1.555 is lower than critical table value which is 18.475

∴ Hence the null hypotheses is accepted

15 Dairies would like to know whether the sales of milk are distributed uniformly over a year so they can plan for milk production and storage. A uniform distribution means that the frequencies are the same in all categories. In this situation, the producers are attempting to determine whether the amounts of milk sold are the same for each month of the year. They ascertain the number of gallons of milk sold by sampling one large supermarket each month during a year, obtaining the following data. Use $\alpha = .01$ to test whether the data fit a uniform distribution (using chi-square test)

Month	Gallons	Month	Gallons
January	1610	August	1350
February	1585	September	1495
March	1649	October	1564
April	1590	November	1602
May	1540	December	1655
June	1397	Total	18,447
July	1410		

→

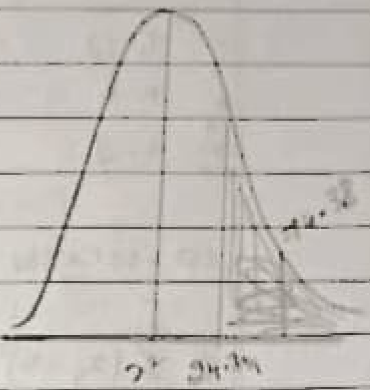
$$f_e = \frac{\Sigma f_o}{n} = \frac{18447}{12} = 1537.25$$

$\alpha = 0.01$ $k = 12$ $df = k-1 = 11$

$$\chi^2_{0.01, 11} = 24.725$$

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right]$$

Month	Gallons (f.)	χ^2
January	1610	3.442876891
February	1555	1.483208652
March	1649	8.12363799
April	1590	1.81009107 2
May	1540	0.00491949 9
June	1397	12.79561717
July	1410	10.53346073
August	1350	22.80862742
September	1495	1.161205074
October	1564	0.465482192
November	1602	2.731527078
December	1655	9.027054806
	15,447	74.38770852



The observed χ^2 value 74.32 is greater than critical χ^2 table value i.e. 24.725

∴ The null hypothesis is rejected. The monthly figures for milk sales are not uniformly distributed

16 Calculate one way ANOVA table for following

1	2	3
2	5	3
1	3	4
3	6	5
3	4	5
2	5	3
1		5

→

1	2	3
2	5	3
1	3	4
3	6	5
3	4	5
2	5	3
1		5

T_j	$T_1 = 12$	$T_2 = 23$	$T_3 = 25$	$T = 60$
n_j	$n_1 = 6$	$n_2 = 5$	$n_3 = 6$	$n = 17$
$\bar{x}_j$	$\bar{x}_1 = 2$	$\bar{x}_2 = 4.6$	$\bar{x}_3 = 4.166666667$	$\bar{x} = 3.529411765$

SST = SSC + SSE

$$\sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{x})^2 + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$SSC = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{x})^2$$

$$= 6(2-3.529411765)^2 + 5(4.6-3.529411765)^2 + 6(4.166666667 - 3.529411765)^2$$

$$= 6(1.529411765)^2 + 5(1.070588235)^2 + 6(0.637254902)^2$$

$$= (6 \times 2.339100347) + (5 \times 1.146159169) + (6 \times 0.40609381)$$

$$= 14.03460208 + 5.732295845 + 2.43656286$$

$$= \underline{\underline{22.20346079}}$$

$$SSE = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$= \left((2-2)^2 + (1-2)^2 + (3-2)^2 + (3-2)^2 + (2-2)^2 + (1-2)^2\right) + \left((5-4.6)^2 + (3-4.6)^2 + (6-4.6)^2 + (4-4.6)^2 - (5-4.6)^2\right) + \left((3-4.166666667)^2 + (4-4.166666667)^2 + (5-4.166666667)^2 + (5-4.166666667)^2 + (3-4.166666667)^2 + (5-4.166666667)^2\right)$$

$$= \left((0)^2 + (-1)^2 + (1)^2 + (1)^2 + (0)^2 + (-1)^2\right) + \left((0.4)^2 + (-1.6)^2 + (1.4)^2 + (-0.6)^2 + (0.4)^2\right) + \left((-1.166666667)^2 + (-0.166666667)^2 + (0.833333333)^2 + (0.833333333)^2 + (-1.166666667)^2 + (0.833333333)^2\right)$$

$= (0+1+1+1+0+1) + (0.16+2.56+1.96+0.36+0.16) +$
$(1.36111112 + 0.02777777 + 0.69444443 +$
$0.19444443 + 1.36111112 + 0.69444443)$
$= 4 + 5.2 + 4.8333333$
$= \underline{14.0333333}$

$SST = SSC + SSE$
$= 22.20346079 + 14.0333333$
$= \underline{36.23679412}$

$df_c = C-1$ $\quad$ $df_e = N-C$ $\quad$ $df_T = N-1$
$= 3-1$ $\quad$ $= 17-3$ $\quad$ $= 17-1$
$= \underline{2}$ $\quad$ $= \underline{14}$ $\quad$ $= \underline{16}$

$MSC = \frac{SSC}{df_c}$ $\quad$ $MSE = \frac{SSE}{df_e}$ $\quad$ $F = \frac{MSC}{MSE}$

$= \frac{22.20346079}{2}$ $\quad$ $= \frac{14.0333333}{14}$ $\quad$ $= \frac{11.1017304}{1.002380952}$

$= \underline{11.1017304}$ $\quad$ $= \underline{1.002380952}$ $\quad$ $= \underline{11.0753605}$

Source of varience	df	SS	MS	F
Between	2	22.20346079	11.1017304	11.0753605
Error	14	14.0333333	1.002380952	
Total	16	36.23679412		

17 A company has three manufacturing plants, and company officials want to determine whether there is a difference in the average age of workers at the three locations. The following data are the ages of the five randomly selected workers at each plant. Perform a one-way ANOVA to determine whether there is a significance difference in the mean ages of the workers at the three plants. Use $\alpha = .01$ and note that the sample sizes are equal.

How I didn't take full value after point so it ...

Plant (Employee Ages)

1	2	3
29	32	25
27	33	24
30	31	24
27	34	25
28	30	26

→

	1	2	3	
	29	32	25	
	27	33	24	
	30	31	24	
	27	34	25	
	28	30	26	
T_j	$T_1 = 141$	$T_2 = 160$	$T_3 = 123$	$T = 424$
n_j	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n = 15$
$\bar{x}_j$	$\bar{x}_1 = 28.2$	$\bar{x}_2 = 32$	$\bar{x}_3 = 24.6$	$\bar{\bar{x}} = 28.26$

SST = SSC + SSE

$$\sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{\bar{x}})^2 = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{\bar{x}})^2 + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$SSC = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{\bar{x}})^2$$

$$= 5(28.2-28.26)^2 + 5(32-28.26)^2 + 5(24.6-28.26)^2$$

$$= 5(-0.06)^2 + 5(3.74)^2 + 5(-3.66)^2$$

$$= 5(0.0036) + 5(13.9876) + 5(13.3956)$$

$$= 0.018 + 69.938 + 66.978$$

$$= \underline{136.934}$$

$$SSE = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$= \left((29-28.2)^2 + (27-28.2)^2 + (30-28.2)^2 + (27-28.2)^2 + (28-28.2)^2\right)$$
$$+ \left((32-32)^2 + (33-32)^2 + (31-32)^2 + (34-32)^2 + (30-32)^2\right) +$$
$$\left((25-24.6)^2 + (24-24.6)^2 + (24-24.6)^2 + (25-24.6)^2 + (25-24.6)^2\right)$$

$= \left((0.8)^2 + (-1.2)^2 + (1.8)^2 + (-1.2)^2 + (-0.2)^2\right) + \left((0)^2 + (1)^2 + (-1)^2 + (2)^2 + (-2)^2\right) + \left((0.4)^2 + (-0.6)^2 + (-0.6)^2 + (0.4)^2 + (0.4)^2\right)$

$= (0.64 + 1.44 + 3.24 + 1.44 + 0.04) + (0 + 1 + 1 + 4 + 4) + (0.16 + 0.36 + 0.36 + 0.16 + 0.16)$

$= 6.8 + 10 + 1.2$

$= \underline{18.1}$

$SST = SSC + SSE$

$= 136.434 + 18.1$

$= \underline{154.934}$

$df_c = c - 1$	$df_e = N - c$	$df_T = N - 1$
$= 3 - 1$	$= 15 - 3$	$= 15 - 1$
$= \underline{2}$	$= \underline{12}$	$= \underline{14}$

$MSC = \dfrac{SSC}{df_c}$	$MSE = \dfrac{SSE}{df_e}$	$F = \dfrac{MSC}{MSE}$
$= \dfrac{136.934}{2}$	$= \dfrac{18.1}{12}$	$= \dfrac{68.467}{1.508}$
$= \underline{68.467}$	$= \underline{1.508}$	$= \underline{45.644}$

Source of varience	df	SS	MS	F
Between	2	136.934	68.467	45.644
Error	12	18.1	1.508	
Total	14	155.034		

$F > F_{df_c, df_e, \alpha} \; (F_{2,12,0.01})$

$45.644 > 6.93$

∴ the null hypothesis is rejected because the observed value is greater than the critical table value

18 Construct one way ANOVA table for following data

Machine operated

1	2	3	4
6.33	6.26	6.44	6.29
6.26	6.36	6.38	6.23
6.31	6.23	6.58	6.19
6.29	6.27	6.54	6.21
6.40	6.19	6.56	
	6.50	6.34	
	6.19	6.58	
	6.22		

→

	1	2	3	4	
	6.33	6.26	6.44	6.29	
	6.26	6.36	6.38	6.23	
	6.31	6.23	6.58	6.19	
	6.29	6.27	6.54	6.21	
	6.40	6.19	6.56		
		6.50	6.34		
		6.19	6.58		
		6.22			
T_j	$T_1 = 31.59$	$T_2 = 50.22$	$T_3 = 45.42$	$T_4 = 24.92$	$T = 152.15$
n_j	$n_1 = 5$	$n_2 = 8$	$n_3 = 7$	$n_4 = 4$	$n = 24$
$\bar{x}_j$	$\bar{x}_1 = 6.318$	$\bar{x}_2 = 6.2775$	$\bar{x}_3 = 6.488571429$	$\bar{x}_4 = 6.24$	$\bar{\bar{x}} = 6.3958333$

SST = SSC + SSE

$$\sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{\bar{x}})^2 = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{\bar{x}})^2 + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$SSC = \sum_{j=1}^{c} n_j(\bar{x}_j-\bar{\bar{x}})^2$$

$$= 5(6.318-6.3958333)^2 + 8(6.2775-6.3958333)^2 + 7(6.488571429-6.3958333)^2 + 4(6.24-6.3958333)^2$$

$$= 5(0.0778333)^2 + 8(-0.1183333)^2 + 7(0.09273809)^2 + 4(-6.1558333)^2$$

$= (5 \times 0.006048027) + (8 \times 0.014002776) + (7 \times 0.008600355)$
$+ (4 \times 0.024284026)$
$= 0.030240135 + 0.112022208 + 0.060202485 + 0.097136104$
$= 0.299650932$

$$SSE = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

$= \big((6.33-6.318)^2 + (6.28-6.318)^2 + (6.31-6.318)^2 + (6.29-6.318)^2$
$+ (6.40-6.318)^2\big) + \big((6.26-6.2775)^2 + (6.36-6.2775)^2 +$
$(6.23-6.2775)^2 + (6.27-6.2775)^2 + (6.19-6.2775)^2 +$
$(6.50-6.2775)^2 + (6.19-6.2775)^2 + (6.22-6.2775)^2\big) +$
$\big((6.44-6.488571429)^2 + (6.38-6.488571429)^2 + (6.58-$
$6.488571429)^2 + (6.54-6.488571429)^2 + (6.56-6.488571429)^2$
$+ (6.34-6.488571429)^2 + (6.58-6.488571429)^2\big) + \big((6.29-$
$6.24)^2 + (6.23-6.24)^2 + (6.19-6.24)^2 + (6.25-6.24)^2\big)$
$= \big((0.012)^2 + (-0.038)^2 + (-0.008)^2 + (-0.028)^2 + (0.082)^2\big) +$
$\big((-0.0175)^2 + (0.0825)^2 + (-0.0475)^2 + (-0.0075)^2 + (-0.0875)^2$
$+ (0.2225)^2 + (-0.0875)^2 + (-0.0575)^2\big) + \big((-0.048571429)^2 +$
$(-0.108571429)^2 + (0.091428571)^2 + (0.051428571)^2 +$
$(0.071428571)^2 + (-0.148571429)^2 + (0.091428571)^2\big) +$
$\big((0.05)^2 + (-0.01)^2 + (-0.05)^2 + (0.01)^2\big)$
$= (0.000144 + 0.001444 + 0.000064 + 0.000784 + 0.006724) +$
$(0.00030625 + 0.00680625 + 0.00225625 + 0.00005625 +$
$0.00765625 + 0.04950625 + 0.00765625 + 0.00330625) +$
$(0.002359183 + 0.011787755 + 0.008359183 +$
$0.002644897 + 0.00510204 + 0.022073469 + 0.008359183) +$
$+ (0.0025 + 0.0001 + 0.0025 + 0.0001)$
$= 0.00916 + 0.07424375 + 0.060685527 + 0.0052$
$= 0.149289277$

$SST = SSC + SSE$
$= 0.299650932 + 0.149289277$
$= 0.448940209$

$df_1 = c-1$ $\quad$ $df_e = n-c$ $\quad$ $df_t = n-1$

$= 4-1$ $\quad$ $= 24-4$ $\quad$ $= 24-1$

$= 3$ $\quad$ $= 20$ $\quad$ $= 23$

$MSC = \frac{SSC}{df_c}$ $\quad$ $MSE = \frac{SSE}{df_e}$

$= \frac{0.299650932}{3}$ $\quad$ $= \frac{0.149289277}{20}$

$= 0.099883644$ $\quad$ $= 0.007464463$

$F = \frac{MSC}{MSE}$

$= \frac{0.099883644}{0.007464463}$

$= 13.381223$

Source of variance	df	SS	MS	F
Between	3	0.299650932	0.099883644	13.381223
Error	20	0.149289277	0.007464463	
Total	23	0.448940209		

19 A survey of the morning beverage market shows that the primary breakfast beverage for 17% of Americans is milk. A milk producer in Wisconsin where milk is plentiful, believes the figure is higher for Wisconsin. To test this idea, she contacts a random sample of 550 Wisconsin residents and asks which primary beverages they consumed for breakfast that day. Suppose 115 replied that milk was the primary beverage. Using a level of significance of .05, test the idea that the milk figure is higher for Wisconsin.

→

$H_0 : p = 0.17$

$H_a : p > 0.17$

$n = 550$ $\quad$ $x = 115$ $\quad$ $\hat{p} = x/n = 0.20909090909$ $\quad$ $p = 0.17$

$\alpha = 0.05$ $\quad$ $1-\alpha = 0.95$

in our table we have probability only till 0.5 to find 0.95 subtract 0.5 from 0.95 you get. now look at the table and find the probability of 0.45 it is between 1.64 and 1.65 take the average it is 1.645

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} \qquad \hat{p} = \frac{115}{550} = .209$$

$$= \frac{.209 - .17}{\sqrt{\frac{(.17)(.83)}{550}}}$$

$$= \frac{.039}{.016}$$

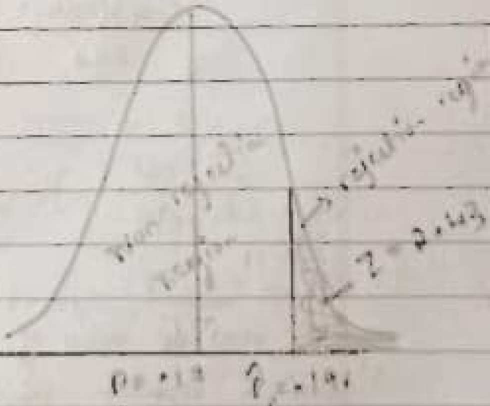
$$= \underline{\underline{2.44}}$$

Because $z = 2.44$ is beyond $Z_{.05} = 1.645$ in the rejection region, the milk producer rejects the null hypotheses

of the proportion of residents who drink milk for breakfast is higher in Wisconsin than in other parts of the United States

$$Z_{.05} = \frac{\hat{P}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$1.645 = \frac{\hat{P}_c - .17}{\sqrt{\frac{(0.17)(.83)}{550}}}$$



$$\hat{P}_c = .17 + 1.645\sqrt{\frac{(.17)(.83)}{550}}$$

$$\hat{P}_c = .17 + .026 = \underline{\underline{.196}}$$

20 A manufacturer believes exactly 8% of its products contain at least one minor flaw. Suppose a company researcher wants to test this belief. The null and alternative hypotheses are

$H_0 : p = .08$

$H_a : p \neq .08$

The business researcher randomly selects a sample of 200 products, inspects each item for flaws, and determines that 33 items have at least one minor flaw. Calculating the sample proportion ($\alpha = .10$)

→ $n = 200$ $\quad x = 33$ $\quad \hat{p} = x/n = 0.165$ $\quad p = .08$

$\alpha = .10$ two tailed so $\alpha/2 = 0.05$ $\quad 1 - \alpha = 1 - 0.05 = 0.95$

In our table we have the probability only till 0.5 to find 0.95 subtract 0.5 from 0.95 you get 0.45. now look at the table and find the probability of 0.45 it is between 1.64 and 1.65 take the average which is ± 1.645

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$= \frac{.165 - .080}{\sqrt{\frac{(.08)(.092)}{200}}}$$

$$= \frac{.085}{.019}$$

$$= 4.43$$

The observed value of Z is in the rejection region, so the business researcher rejects the null hypotheses

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\bar{p}\,\bar{q})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.475543478 - 0.449382716) - (0)}{\sqrt{(0.461836998 \times 0.538163002)\left(\frac{1}{368} + \frac{1}{405}\right)}}$$

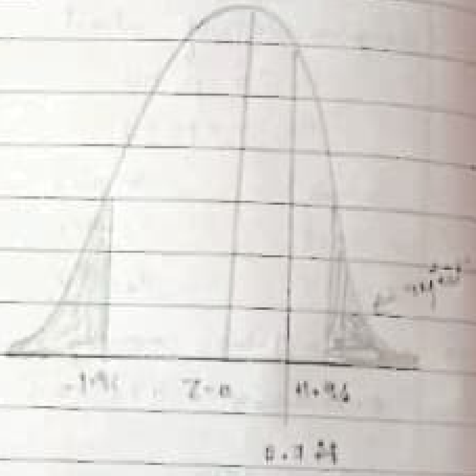
$$= \frac{0.026160762}{\sqrt{0.248543585\,(0.002717391 + 0.002469135)}}$$

$$= \frac{0.026160762}{\sqrt{0.248543585 \times 0.005186526}}$$

$$= \frac{0.026160762}{\sqrt{0.0012890477}}$$

$$= \frac{0.026160762}{0.035903718}$$

$$= \underline{0.728636562}$$



The table value (1.96) is greater than the observed value (0.728). Hence the null hypothesis is not rejected. It is accepted.

Q2 Using the given sample information, test following hypothesis. Note that x is the number in the sample having the characteristics of interests.

$H_0 : P_1 - P_2 = 0$ $\qquad$ $H_a : P_1 - P_2 > 0$

Sample 1 $\qquad$ Sample 2

$n_1 = 649$ $\qquad$ $n_2 = 558$

$\hat{P}_1 = .38$ $\qquad$ $\hat{P}_2 = .25$ $\qquad$ Let $\alpha = .10$

One tailed

$\alpha = .10$

$Z_{.10} = 1 - 0.10 = 0.9 - 0.5 = 0.4 = 1.28$

$$\bar{P} = \frac{n_1 \hat{P}_1 + n_2 \hat{P}_2}{n_1 + n_2}$$

$$= \frac{(649 \times 0.38) + (558 \times 0.25)}{649 + 558}$$

$$= \frac{246.62 + 139.5}{1207}$$

$$= \frac{386.12}{1207}$$

$$= \underline{0.31990058}$$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{(\bar{p} \cdot \bar{q}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$= \frac{(.38 - .25) - 0}{\sqrt{(0.31990058 \times 0.68009942) \left(\frac{1}{649} + \frac{1}{558} \right)}}$$

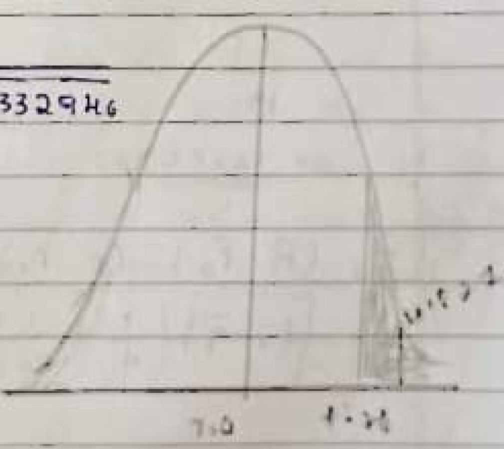
$$= \frac{0.13}{\sqrt{0.217564198 (0.001540832 + 0.001792114)}}$$

$$= \frac{0.13}{\sqrt{0.217564198 \times 0.003332946}}$$

$$= \frac{0.13}{\sqrt{0.000725129}}$$

$$= \frac{0.13}{0.026928232}$$

$$= \underline{4.827646909}$$



The table value 1.28 is lesser than observed value 4.827. Hence the null hypothesis is rejected.

Date 27/10/23
Page 140

23 Suppose you decide to test this result by taking a survey of your own and identify female entrepreneurs by gross sales. You interview 100 female entrepreneurs with gross sales of less than $100,000 and 24 of them define sales/profit as success. You then interview 95 female entrepreneurs with gross sales of $100000 to $500000 and 39 cite sales/profit as a definition of success. Use this information to determine whether there is a significant difference in the proportion of the two groups that define success as sales/profit. Use $\alpha = .01$

→

$n_1 = 100$ $n_2 = 95$

$x_1 = 24$ $x_2 = 39$

$H_0 : p_1 - p_2 = 0$ $\alpha = .01$

$H_a : p_1 - p_2 \neq 0$

$\alpha/2 = 0.005$

$Z_{0.005} = 1 - 0.005 = 0.995 - 0.5 = 0.495 = \underline{\underline{2.575}}$

$\hat{P}_1 = \frac{24}{100} = \underline{\underline{0.24}}$ $\hat{P}_2 = \frac{39}{95} = 0.410526315$

$\bar{P} = \frac{x_1 + x_2}{n_1 + n_2}$

$= \frac{24 + 39}{100 + 95}$

$= \frac{63}{195}$

$= \underline{\underline{0.323076923}}$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{(\bar{P} \cdot \bar{q})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.24 - 0.410526315) - 0}{\sqrt{(0.323076923 \times 0.676923077)\left(\frac{1}{100} + \frac{1}{95}\right)}}$$

Consumers	CEOs
$n_1 = 755$	$n_2 = 616$
$\hat{p}_1 = .57$	$\hat{p}_2 = .50$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$= \frac{(755 \times .57) + (616 \times .50)}{755 + 616}$$

$$= \frac{430.35 + 308}{755 + 616}$$

$$= \frac{738.35}{1371}$$

$$= \underline{0.538548504}$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\bar{p} \cdot \bar{q})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.57 - 0.50) - (0)}{\sqrt{(0.538548504 \times 0.461451496)\left(\frac{1}{755} + \frac{1}{616}\right)}}$$

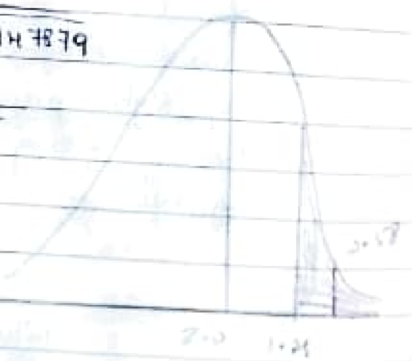
$$= \frac{0.07}{\sqrt{0.248514012(0.001324503 + 0.001623376)}}$$

$$= \frac{0.07}{\sqrt{0.248514012 \times 0.002947879}}$$

$$= \frac{0.07}{\sqrt{0.000732589}}$$

$$= \frac{0.07}{0.027066385}$$

$$= \underline{2.586233739}$$



Since the table value is lesser than the observed z value
Hence the null hypothesis is rejected

25 Explain the types of correlation with example

26 Explain types of correlation with respect to correlation coefficient with condition and suitable scatter diagram

→ refer 2 marks question from 25 to 29

27 From the following information find the correlation coefficient between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method (Direct Method)

Firm	1	2	3	4	5	6	7	8	9	10
Advertisement Exp. (Rs. In Lakhs)	11	13	14	16	16	15	15	14	13	13
Sales Volume (Rs. In Lakhs)	50	50	55	60	65	65	65	60	60	50

→

X	Y	$x = X - \bar{X}$	x^2	$y = Y - \bar{Y}$	y^2	xy
11	50	-3	9	-8	64	24
13	50	-1	1	-8	64	8
14	55	0	0	-3	9	0
16	60	2	4	2	4	4
16	65	2	4	7	49	14
15	65	1	1	7	49	7
15	65	1	1	7	49	7
14	60	0	0	2	4	0
13	60	-1	1	2	4	-2
13	50	-1	1	-8	64	8
140	580		22		360	70

$\bar{X} = \frac{\Sigma X}{N}$ $\quad$ $\bar{Y} = \frac{\Sigma Y}{N}$

$= \frac{140}{10}$ $\quad$ $= \frac{580}{10}$

$= 14$ $\quad$ $= 58$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{70}{\sqrt{22 \times 360}}$$

$$= \frac{70}{\sqrt{7920}}$$

$$= \frac{70}{88.99}$$

$$= 0.7866$$

$\therefore$ This is high degree positive correlation

28 Calculate the Karl Pearson's product moment of coefficient correlation

Day	Interest x	Futures Index y
1	7.43	221
2	7.48	222
3	8.00	226
4	7.75	225
5	7.60	224

→

x	y	xy	x^2	y^2
7.43	221	1642.03	55.2049	48841
7.48	222	1660.56	55.9504	49284
8.00	226	1808	64	51076
7.75	225	1743.75	60.0625	50625
7.60	224	1702.4	57.76	50176
38.26	1118	8556.74	292.9778	250002

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2 \times (n\Sigma y^2 - (\Sigma y)^2)}}$$

$$= \frac{(5 \times 8556.74) - (38.26 \times 1118)}{\sqrt{((5 \times 292.9778) - (38.26)^2) \times ((5 \times 250002) - (1118)^2}}$$

$$= \frac{42783.7 - 42774.68}{\sqrt{(1464.889 - 1463.8276) \times (1250010 - 1249924)}}$$

$$= \frac{9.02}{\sqrt{1.0614 \times 86}}$$

$$= \frac{9.02}{\sqrt{91.2804}}$$

$$= \frac{9.02}{9.554077663}$$

$$= \underline{\underline{0.944099505}}$$

22. Calculate the Karl Pearson's product moment of coefficient of correlation.

Student	1	2	3	4	5	6
Statistics (x)	7	4	6	9	3	8
Mathematics (y)	8	5	4	8	3	6

→

x	y	xy	x^2	y^2
7	8	56	49	64
4	5	20	16	25
6	4	24	36	16
9	8	72	81	64
3	3	9	9	9
8	6	48	64	36
37	34	229	255	214

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2 \times (n\Sigma y^2 - (\Sigma y)^2)}}$$

$$= \frac{(6 \times 229) - (37 \times 34)}{\sqrt{((6 \times 255) - (37)^2) \times ((6 \times 214) - (34)^2)}}$$

$$= \frac{1374 - 1258}{\sqrt{(1530 - 1369) * (1284 - 1156)}}$$

$$= \frac{116}{\sqrt{161 * 128}}$$

$$= \frac{116}{\sqrt{20608}}$$

$$= \frac{116}{143.5548676}$$

$$= \underline{0.808053408}$$

30 Determine the karl Pearson's product moment of coefficient of correlation

x	4	6	7	11	14	17	21
y	18	12	13	8	7	7	4

→

x	y	xy	x^2	y^2
4	18	72	16	324
6	12	72	36	144
7	13	91	49	169
11	8	88	121	64
14	7	98	196	49
17	7	119	289	49
21	4	84	441	16
80	69	624	1148	815

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2 * (n\Sigma y^2 - (\Sigma y)^2)}}$$

$$= \frac{(7 \times 624) - (80 \times 69)}{\sqrt{((7 \times 1148) - (80)^2) * ((7 \times 815) - (69)^2)}}$$

$$= \frac{4368 - 5220}{\sqrt{(8036 - 6400) \times (5705 - 4761)}}$$

$$= \frac{-852}{\sqrt{1636 \times 944}}$$

$$= \frac{-852}{\sqrt{1544384}}$$

$$= \frac{-852}{1242.732473}$$

$$= \underline{-0.685578601}$$

There exists a low degree of negative correlation between x and y

3) Determine the Karl Pearson's product moment of coefficient of correlation

x	158	296	87	110	436
y	349	510	301	322	550

→

x	y	xy	x^2	y^2
158	349	55142	24964	121801
296	510	150960	87616	260100
87	301	26187	7569	90601
110	322	35420	12100	103684
436	550	239800	190096	302500
1087	2032	507509	322345	878686

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{(n\Sigma x^2 - (\Sigma x)^2) \times (n\Sigma y^2 - (\Sigma y)^2)}}$$

$$= \frac{(5 \times 507509) - (1087 \times 2032)}{\sqrt{(5 \times 322345 - (1087)^2) \times (5 \times 878686 - (2032)^2)}}$$

$$= \frac{2537545 - 2208784}{\sqrt{(1611725 - 1181569) \times (4393430 - 4129024)}}$$

$$= \frac{328761}{\sqrt{430151 \times 264406}}$$

$$= \frac{328761}{\sqrt{113735827300}}$$

$$= \frac{328761}{337247.4274}$$

$$= \underline{0.974836198}$$

High degree positive correlation between x and y.

32. Find the two regression equation of x on y and y on x from the following data

x	10	12	16	11	15	14	20	22
y	15	18	23	14	20	17	25	28

→

x	y	xy	x^2	y^2
10	15	150	100	225
12	18	216	144	324
16	23	368	256	529
11	14	154	121	196
15	20	300	225	400
14	17	238	196	289
20	25	500	400	625
22	28	616	484	784
120	160	2542	1926	3372

$\bar{x} = \frac{\Sigma x}{N} = \frac{120}{8} = \underline{15}$ $\qquad$ $\bar{y} = \frac{160}{8} = \underline{20}$

X on Y

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$$

$$= \frac{(8\times 2542) - (120\times 160)}{(8\times 3372) - (160)^2}$$

$$= \frac{20336 - 19200}{26976 - 25600}$$

$$= \frac{1136}{1376}$$

$$= 0.825581395$$

$(x - \bar{x}) = b_{xy}(y - \bar{y})$

$(x - 15) = 0.825581395(y - 20)$

$x - 15 = 0.825581395y - 16.5116279$

$x = 0.825581395y - 16.5116279 + 15$

$x = 0.825581395y - 1.5116279$

Thus, we got the value $a = -1.5116279$ $b = 0.825581395$

Y on X

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{(8\times 2542) - (120\times 160)}{(8\times 1926) - (120)^2}$$

$$= \frac{20336 - 19200}{15408 - 14400}$$

$$= \frac{1136}{1008}$$

$$= 1.126984127$$

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$y-10 = -1.34(x-6)$

$y-10 = -1.34x + 8.04$

$y = -1.34x + 8.04 + 10$

$y = -1.34x + 18.04$

Thus, we got the value $a = 18.04$, $b = -1.34$

Find the regression equation of x on y and predict the value of x when y is 4

x	3	6	5	4	4	6	7	5
y	3	2	3	5	3	6	6	4

→

x	y	x^2	y^2	xy
3	3	9	9	9
6	2	36	4	12
5	3	25	9	15
4	5	16	25	20
4	3	16	9	12
6	6	36	36	36
7	6	49	36	42
5	4	25	16	20
40	32	212	144	166

$\bar{x} = \frac{40}{8} = 5$ $\qquad$ $\bar{y} = \frac{32}{8} = 4$

$$b_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{(8 \times 166) - (40 \times 32)}{(8 \times 144) - (32)^2}$$

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$y-20 = 1.126984127(x-15)$

$y-20 = 1.126984127x - 16.9047619$

$y = 1.126984127x - 16.90476191 + 20$

$y = 1.126984127x + 3.095238095$

Thus, we get the value $a = 3.095238095$, $b = 1.126984$

33 Compute the regression equation of y on x from the following data

x	2	4	5	6	8	11
y	18	12	10	8	7	5

→

x	y	x^2	y^2	xy
2	18	4	324	36
4	12	16	144	48
5	10	25	100	50
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
36	60	266	706	293

$\bar{x} = \frac{36}{6} = 6$

$\bar{y} = \frac{60}{6} = 10$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{(6\times 293) - (36\times 60)}{(6\times 266) - (36)^2}$$

$$= \frac{1758 - 2160}{1596 - 1296}$$

$$= \frac{-402}{300}$$

$$= -1.34$$

$$= \frac{1328 - 1280}{1152 - 1024}$$

$$= \frac{48}{128}$$

$$= \underline{\underline{0.375}}$$

$(x - \bar{x}) = b_{xy} (y - \bar{y})$

$x - 5 = 0.375 (y - 4)$

$x - 5 = 0.375 y - 1.5$

$x = 0.375 y - 1.5 + 5$

$x = 0.375 y + 3.5$

Here the value of y is given as 9

$x = 0.375 \times 9 + 3.5$

$x = 3.375 + 3.5$

$x = \underline{\underline{6.875}}$