(2½ Hours)

[Total Marks: 75]

N. B.: (1) **All** questions are **compulsory**.
(2) Make **suitable assumptions** wherever necessary and **state the assumptions** made.
(3) Answers to the **same question** must be **written together**.
(4) Numbers to the **right** indicate **marks**.
(5) Draw **neat labelled diagrams** wherever **necessary**.
(6) Use of **Non-programmable** calculators is **allowed**.

Marks have to be given for steps. Answers are given below the question.

1. Attempt *any three* of the following: **15**

a. The following table gives the heights of 100 students at XYZ College. Find the mean height of the students.

Height (in)	No. of Students
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
	100

Ans 67.45 in

b. During one year the ratio of milk prices per quart to bread prices per loaf was 3.00, whereas during the next year the ratio was 2.00.

i. Find the arithmetic mean of these ratios for the 2-year period.
ii. Find the arithmetic mean of the ratios of bread prices to milk prices for the 2-year period.
iii. Discuss the advisability of using the arithmetic mean for averaging ratios.
iv. Discuss the suitability of the geometric mean for averaging ratios.

Ans
i. 2.5
ii. 0.417
iii. We would expect the mean ratio of milk to bread prices to be the reciprocal of the mean ratio of bread to milk prices if the mean is an appropriate average. However, 1/0.417 = 2.40 ≠ 2:50. This shows that the arithmetic mean is a poor average to use for ratios.
iv.

$$\text{Geometric mean of ratios of milk to bread prices} = \sqrt{(3.00)(2.00)} = \sqrt{6.00}$$
$$\text{Geometric mean of ratios of bread to milk prices} = \sqrt{(0.333)(0.500)} = \sqrt{0.0167} = 1/\sqrt{6.00}$$

Since these averages are reciprocals, our conclusion is that the geometric mean is more suitable than the arithmetic mean for averaging ratios for this type of problem.

c. Two variables, X and Y, assume the values $X_1 = 2$, $X_2 = -5$, $X_3 = 4$, $X_4 = -8$ and $Y_1 = -3$, $Y_2 = -8$, $Y_3 = 10$, $Y_4 = 6$, respectively. Calculate:

$$i. \sum XY, \quad ii. \sum X \sum Y, \quad iii. \sum XY^2, \quad iv. \sum X^2, \quad v. \sum (X - Y)(X + Y)$$

Ans 26, -35, -190, 109, -100

d. On a final examination in statistics, the mean grade of a group of 150 students was 78 and the standard deviation was 8.0. In algebra, however, the mean final grade of the group was 73 and the standard deviation was 7.6. In which subject was there the greater (i) absolute dispersion and (ii) relative dispersion?

Ans (i) Statistics (ii) Algebra

e. State and explain the properties of standard deviation.

Ans (i) The standard deviation can be defined as

$$s = \sqrt{\frac{\sum_{j=1}^{N} (X_j - a)^2}{N}}$$

where a is an average besides the arithmetic mean. Of all such standard deviations, the minimum is that for which $a = \bar{X}$

(ii) Suppose that two sets consisting of N_1 and N_2 numbers (or two frequency distributions with total frequencies N_1 and N_2) have variances given by s_1^2 and s_2^2, respectively, and have the *same* mean $\bar{X}$. Then the *combined*, or *pooled*, *variance* of both sets (or both frequency distributions) is given by

$$s^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2}$$

This is a weighted arithmetic mean of the variances. This result can be generalized to three or more sets.

(iii) Chebyshev's theorem states that for $k > 1$, there is at least $(1 - (1/k^2)) \times 100\%$ of the probability distribution for any variable within k standard deviations of the mean. In particular, when $k = 2$, there is at least $(1 - (1/2^2)) \times 100\%$ or 75% of the data in the interval $(\bar{x} - 2S, \bar{x} + 2S)$, when $k = 3$ there is at least $(1 - (1/3^2)) \times 100\%$ or 89% of the data in the interval $(\bar{x} - 3S, \bar{x} + 3S)$, and when $k = 4$ there is at least $(1 - (1/4^2)) \times 100\%$ or 93.75% of the data in the interval $(\bar{x} - 4S, \bar{x} + 4S)$.

(iv)

f. For a group of 200 candidates, the mean arid standard deviation of scores were found to be 40 and 15 respectively. Later on, it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the corrected figures.

Ans 39.955, 14.97

2. Attempt *<u>any three</u>* of the following: **15**

a. Find the (i) first, (ii) second, (iii) third and (iv) fourth moments about the mean of the set 2, 3, 7, 8, 10.

Ans 0, 9.2,–3.6 , 122

b. In a frequency distribution1 the co-efficient of skewness based upon the quartiles is 0·6. If the sum of the upper and lower quartiles is 100 and median is 38, find the value of the upper and the lower quartiles.

Ans 70,30

c. In a survey of 500 adults were asked the three-part question (1) Do you own a cell phone, (2) Do you own an ipod, and (3) Do you have an internet connection? The results of the survey were as follows (no one answered no to all three parts):

cell phone 329	cell phone and ipod 83
ipod 186	cell phone and internet connection 217
internet connection 295	ipod and internet connection 63

Find the probability that:
(i) answered yes to all three parts, (ii) had a cell phone but not an internet connection, (iii) had an ipod but not a cell phone, (iv) had an internet connection but not an ipod, (v) had a cell phone or an internet connection but not an ipod and, (vi) had a cell phone but not an ipod or an internet connection.

Ans 0.106, 0.224, 0.206, 0.464, 0.628, 0.164 (Even if students have found number it is fine (53, 112, 103, 232, 628, 82)

d. One bag contains 4 white balls and 2 black balls; another contains 3 white balls and 5 black balls. If one ball is drawn from each bag, find the probability that (i) both are white, (ii) both are black, and (iii) one is white and one is black.

Ans ¼. 5/24, 13/24

e. Assume that the heights of 3000 male students at a university are normally distributed with mean 68.0 inches (in) and standard deviation 3.0 in. If 80 samples consisting of 25 students each are obtained, what would be the expected mean and standard deviation of the resulting sampling distribution of means if the sampling were done (i) with replacement and (ii) without replacement? Give the interpretation of the result.

Ans 0.6, 0.59759 which is only very slightly less than 0.6 in and can therefore, for all practical purposes, be considered the same as in sampling with replacement.
Thus, we would expect the experimental sampling distribution of means to be approximately normally distributed with mean 68.0 in and standard deviation 0.6 in.

f. Five hundred ball bearings have a mean weight of 5.02 grams (g) and a standard deviation of 0.30 g. Find the probability that a random sample of 100 ball bearings chosen from this group will have a combined weight of (i) between 496 and 500 g and (ii) more than 510 g. (Use the table of area under normal curve from 0 to z).

Ans 0.2164, 0.0015

3. Attempt *any three* of the following: **15**

a. In measuring reaction time, a psychologist estimates that the standard deviation is 0.05 seconds (s). How large a sample of measurements must he take in order to be (i) 95% and (ii) 99% confident that the error of his estimate will not exceed 0.01 s?

Ans 97 or larger, 167 or larger.

b. A measurement was recorded as 216.480 grams (g) with a probable error of 0.272 g. What are the 95% confidence limits for the measurement?

Ans 216:480 ± 0:790 g.

c. A sample poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favor of a particular candidate. Find the (a) 95%, (b) 99%, and (c) 99.73% confidence limits for the proportion of all the voters in favor of this candidate.

Ans 0.55 ± 0.10, 0.55 ± 0.13, 0.55 ± 0.15

d. Explain Type I and Type II errors and Level of Significance.

Ans **Type I and Type II errors:**
If we reject a hypothesis when it should be accepted, we say that a Type I error has been made. If, on the other hand, we accept a hypothesis when it should be rejected, we say that a Type II error has been made. In either case, a wrong decision or error in judgment has occurred.
In order for decision rules (or tests of hypotheses) to be good, they must be designed so as to minimize errors of decision. This is not a simple matter, because for any given sample size, an attempt to decrease one type of error is generally accompanied by an increase in the other type of error. In practice, one type of error may be more serious than the other, and so a compromise should be reached in favor of limiting the more serious error. The only way to reduce both types of error is to increase the sample size, which may or may not be possible.
Level of Significance:

In testing a given hypothesis, the maximum probability with which we would be willing to risk a Type I error is called the level of significance, or significance level, of the test. This probability, often denoted by α, is generally specified before any samples are drawn so that the results obtained will not influence our choice.

In practice, a significance level of 0.05 or 0.01 is customary, although other values are used. If, for example, the 0.05 (or 5%) significance level is chosen in designing a decision rule, then there are about 5 chances in 100 that we would reject the hypothesis when it should be accepted; that is, we are about 95% confident that we have made the right decision. In such case we say that the hypothesis has been rejected at the 0.05 significance level, which means that the hypothesis has a 0.05 probability of being wrong.

e. The breaking strengths of cables produced by a manufacturer have a mean of 1800 pounds (lb) and a standard deviation of 100 lb. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850 lb. Can we support the claim at the 0.01 significance level?

Ans 3.55> 2.33, The claim should be supported.

f. Two groups, A and B, consist of 100 people each who have a disease. A serum is given to group A but not to group B (which is called the control); otherwise, the two groups are treated identically. It is found that in groups A and B, 75 and 65 people, respectively, recover from the disease. At significance levels of (a) 0.01, (b) 0.05, and (c) 0.10, test the hypothesis that the serum helps cure the disease. Compute the p-value and show that p-value>0.01, p-value4\>0.05, but p-value<0.10.

Ans Using a one-tailed test at the 0.01 significance level, we would reject hypothesis H_0 only if the z score were greater than 2.33. Since the z score is only 1.54, we must conclude that the results are due to chance at this level of significance.

Using a one-tailed test at the 0.05 significance level, we would reject H_0 only if the z score were greater than 1.645. Hence we must conclude that the results are due to chance at this level also.

If a one-tailed test at the 0.10 significance level were used, we would reject H_0 only if the z score were greater than 1.28. Since this condition is satisfied, we conclude that the serum is effective at the 0.10 level.

4. Attempt *<u>any three</u>* of the following: 15

a. A random sample of 10 boys had the following I.Q.s :

70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

Do these data support the assumption of a population mean I.Q. of 100? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Ans Degrees of freedom = 9, Calculated value of t = 0.62, Tabulated value of $t_{0.5}$ for 9 d.f. at 5% significance level is for two tailed test is 2.262. Since calculated value is less than the tabulated value, we may conclude that the data are consistent with the assumption of mean I.Q. of 100 in the population.

b. Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test the hypothesis that the true variances are equal. against the alternative that they are not. at the 10% level. [Assume that P ($F_{10,8} \geq 3.35$) = 0.05 and P ($F_{8,10} \geq 3.07$) = 0.05

Ans **Solution.** We want to test *Null Hypothesis*, $H_0 : \sigma_X^2 = \sigma_Y^2$.
against the *Alternative Hypothesis*, $H_1 : \sigma_X^2 \neq \sigma_Y^2$ (Two-tailed).

We are given :

$$n_1 = 11, n_2 = 9, s_X = 0{\cdot}8 \text{ and } s_Y = 0{\cdot}5.$$

Under the null hypothesis, $H_0 : \sigma_X = \sigma_Y$, the statistic

$$F = \frac{s_X^2}{s_Y^2}$$

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

Now $\qquad n_1 s_X^2 = (n_1 - 1) S_X^2$

$$\therefore \qquad S_X^2 = \left(\frac{n_1}{n_1 - 1}\right) s_X^2 = \left(\frac{11}{10}\right) \times (0{\cdot}8)^2 = 0{\cdot}704$$

Similarly, $$S_Y^2 = \left(\frac{n_2}{n_2 - 1}\right) s_Y^2 = \left(\frac{9}{8}\right) \times (0{\cdot}5)^2 = 0{\cdot}28125$$

$$\therefore \qquad F = \frac{0{\cdot}704}{0{\cdot}28125} = 2{\cdot}5$$

The significant values of F for two tailed test at level of significance $\alpha = 0{\cdot}10$ are :

$$\left.\begin{aligned} F &> F_{10,8}(\alpha/2) = F_{10,8}(0{\cdot}05) \\ \text{and } F &< F_{10,8}(1 - \alpha/2) = F_{10,8}(0{\cdot}95) \end{aligned}\right\} \quad \ldots(*)$$

We are given the tabulated (significant) values :

$$P[F_{10,8} \geq 3{\cdot}35] = 0{\cdot}05 \quad \Rightarrow \quad F_{10,8}(0{\cdot}05) = 3{\cdot}35 \quad \ldots(**)$$

Also $\qquad P[F_{8,10} \geq 3{\cdot}07] = 0{\cdot}05 \quad \Rightarrow \quad P\left[\frac{1}{F_{8,10}} \leq \frac{1}{3{\cdot}07}\right] = 0{\cdot}05$

$$\Rightarrow \quad P[F_{10,8} \leq 0{\cdot}326] = 0{\cdot}05 \quad \Rightarrow \quad P[F_{10,8} \geq 0{\cdot}326] = 0{\cdot}95 \ldots(***)$$

Hence from (*), (**) and (***), the critical values for testing $H_0 : \sigma_X^2 = \sigma_Y^2$, against $H_1 : \sigma_X^2 \neq \sigma_Y^2$ at level of significance $\alpha = 0{\cdot}10$ are given by :

$$F > 3{\cdot}35 \quad \text{and} \quad F < 0{\cdot}326 = 0{\cdot}33$$

Since, the calculated value of F (=2·5) lies between 0·33 and 3·35, it is not significant and hence null hypothesis of equality of population variances may be accepted at level of significance $\alpha = 0{\cdot}10$.

c. The standard deviation of the heights of 16 male students chosen at random in a school of 1000 male students is 2.40 in. Find the (i) 95% and (ii) 99% confidence limits of the standard deviation for all male students at the school.

Ans i.

The 95% confidence limits are given by $s\sqrt{N}/\chi_{.975}$ and $s\sqrt{N}/\chi_{.025}$.

For $\nu = 16 - 1 = 15$ degrees of freedom, $\chi^2_{.975} = 27.5$ (or $\chi_{.975} = 5.24$) and $\chi^2_{.025} = 6.26$ (or $\chi_{.025} = 2.50$). Then the 95% confidence limits are $2.40\sqrt{16}/5.24$ and $2.40\sqrt{16}/2.50$ (i.e., 1.83 and 3.84 in). Thus we can be 95% confident that the population standard deviation lies between 1.83 and 3.84 in.

ii.

The 99% confidence limits are given by $s\sqrt{N}/\chi_{.995}$ and $s/\sqrt{N}/\chi_{.005}$.

For $\nu = 16 - 1 = 15$ degrees of freedom, $\chi^2_{.995} = 32.8$ (or $\chi_{.995} = 5.73$) and $\chi^2_{.005} = 4.60$ (or $\chi_{.005} = 2.14$). Then the 99% confidence limits are $2.40\sqrt{16}/5.73$ and $2.40\sqrt{16}/2.14$ (i.e., 1.68 and 4.49 in). Thus we can be 99% confident that the population standard deviation lies between 1.68 and 4.49 in.

d. Calculate the chi-square value for the following data.

Colour	Red	Green	Yellow
Observed Frequency	12	16	20
Expected Frequency	16	8	15

Ans

$$X^2 = \sum \frac{(O-E)^2}{E}$$

First, lets calculate $(O - E)^2$ for each color.

Red color = $(O - E)^2$ = $(12 - 16)^2$ = 16

Green color = $(O - E)^2$ = $(16 - 8)^2$ = 64

Yellow color = $(O - E)^2$ = $(20 - 15)^2$ =15

Chi-Square value for Red color = $\frac{(12-16)^2}{16}$ = 1

Chi-Square value for Green color = $\frac{(16-8)^2}{8}$ = 8

Chi-Square value for Yellow color = $\frac{(20-15)^2}{15}$ = 1

So, the chi-square value for the given data is = 1 + 8 + 1 = 10

Chi Square value = 10

e. Acme Toy Company prints baseball cards. The company claims that 30% of the cards are rookies, 60% veterans but not All-Stars, and 10% are veteran All-Stars.
Suppose a random sample of 100 cards has 50 rookies, 45 veterans, and 5 All-Stars. Is this consistent with Acme's claim? Use a 0.05 level of significance. (Use chi-square goodness of fit). Given P(χ^2 > 19.58) = 0.0001

Ans

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.
  - Null hypothesis: The proportion of rookies, veterans, and All-Stars is 30%, 60% and 10%, respectively.
  - Alternative hypothesis: At least one of the proportions in the null hypothesis is false.
- **Formulate an analysis plan**. For this analysis, the significance level is 0.05. Using sample data, we will conduct a chi-square goodness of fit test of the null hypothesis.

Analyze sample data. Applying the chi-square goodness of fit test to sample data, we compute the degrees of freedom, the expected frequency counts, and the chi-square test statistic. Based on the chi-square statistic and the degrees of freedom, we determine the P-value.

DF = k - 1 = 3 - 1 = 2 $(E_i) = n * p_i$

$(E_1) = 100 * 0.30 = 30$

$$(E_2) = 100 * 0.60 = 60$$
$$(E_3) = 100 * 0.10 = 10 \quad X^2 = \Sigma [(O_i - E_i)^2 / E_i]$$
$$X^2 = [(50 - 30)^2 / 30] + [(45 - 60)^2 / 60] + [(5 - 10)^2 / 10]$$
$$X^2 = (400 / 30) + (225 / 60) + (25 / 10) = 13.33 + 3.75 + 2.50 = 19.58$$

where DF is the degrees of freedom, k is the number of levels of the categorical variable, n is the number of observations in the sample, E_i is the expected frequency count for level i, O_i is the observed frequency count for level i, and X^2 is the chi-square test statistic.

The P-value is the probability that a chi-square statistic having 2 degrees of freedom is more extreme than 19.58.

$P(X^2 > 19.58) = 0.0001$

Interpret results. Since the P-value (0.0001) is less than the significance level (0.05), we cannot accept the null hypothesis.

f. A survey of 320 families with 5 children each revealed the following distribution:

Boys	5	4	3	2	1	0
Girls	0	1	2	3	4	5
No of families	14	56	110	88	40	12

Is this result consistent with the hypothesis that male and female births are equally probable?

Ans **Solution.** Let us set up the *null hypothesis* that *the data are consistent with the hypothesis of equal probability for male and female births.* Then under the null hypothesis :

$$p = \text{Probability of male birth} = \frac{1}{2} = q$$

$$p(r) = \text{Probability of 'r' male births in a family of 5}$$

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

The frequency of r male births is given by :

$$f(r) = N.\ p(r) = 320 \times \binom{5}{r} \times \left(\frac{1}{2}\right)^5$$

$$= 10 \times \binom{5}{r} \qquad \ldots(*)$$

Substituting $r = 0, 1, 2, 3, 4$ successively in (*), we get the expected frequencies as follows :

$f(0) = 10 \times 1 = 10,\qquad f(1) = 10 \times {}^5C_1 = 50$

$f(2) = 10 \times {}^5C_2 = 100,\qquad f(3) = 10 \times {}^5C_3 = 100$

$f(4) = 10 \times {}^5C_4 = 50,\qquad f(5) = 10 \times {}^5C_5 = 10$

CALCULATIONS FOR χ^2

Observed Frequencies (O)	*Expected Frequencies (E)*	$(O-E)^2$	$(O-E)^2/E$
14	10	16	1·6000
56	50	36	0·7200
110	100	100	1·0000
88	100	144	1·4400
40	50	100	2·0000
12	10	4	0·4000
Total 320	320		7·1600

$$\therefore \qquad \chi^2 = \sum \left[\frac{(O-E)^2}{E}\right] = 7{\cdot}16$$

Tabulated $\chi^2_{0.05}$ for $6 - 1 = 5$ d.f. is 11·07.

Calculated value of χ^2 is less than the tabulated value, it is not significant at 5% level of significance and hence the null hypothesis of equal probability for male and female births may be accepted.

5. Attempt *any three* of the following: **15**

a. Fit an exponential curve of the form $Y = ab^x$ to the following data:

x	1	2	3	4	5	6	7	8
Y	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

Ans $Y = 0.6821(1.38)^x$

b. The weights of a calf taken at weekly intervals are given below. Fit a straight line using the method of least squares and calculate the average rate of growth, per week.

Age (X)	1	2	3	4	5	6	7	8	9	10
Weight (Y)	52.5	58.7	65.0	70.2	75.4	81.1	87.2	95.5	102.2	108.4

Ans $Y = 45.74 + 6.16X$

c. Fit a second-degree parabola to the following data taking X as the independent variable:

X	1	2	3	4	5	6	7	8	9
Y	2	6	7	8	10	11	11	10	9

Ans $Y = -1 + 3.55X - 0.27X^2$

d. Find the coefficient of linear correlation between the variables X and Y presented in Table below:

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Ans 0.977

e. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible:

Variance of $X = 9$, Regression equations: $8X - 10Y + 66 = 0$, $40X - 18Y = 214$

Find (i) Mean values of X and Y. (ii) the correlation coefficient between X and Y. and (iii) the standard deviation of Y.

Ans **Solution** (*i*) Since both the lines of regression pass through the point $(\overline{X}, \overline{Y})$, we have $8\overline{X} - 10\overline{Y} + 66 = 0$, and $40\overline{X} - 18\overline{Y} = 214$.

Solving, we get $\overline{X} = 13$, $\overline{Y} = 17$.

(*ii*) Let $8X - 10Y + 66 = 0$ and $40X - 18Y = 214$ be the lines of regression of Y on X and X on Y respectively. These equations can be put in the form :

$$Y = \frac{8}{10}X + \frac{66}{10} \text{ and } X = \frac{18}{40}Y + \frac{214}{40}$$

$\therefore$ b_{YX} = Regression coefficient of Y on $X = \frac{8}{10} = \frac{4}{5}$

and b_{XY} = Regression coefficient of X on $Y = \frac{18}{40} = \frac{9}{20}$

Hence $r^2 = b_{YX} \cdot b_{XY} = \frac{4}{5} \cdot \frac{9}{20} = \frac{9}{25}$

But since both the regression coefficients are positive, we take $r = +0{\cdot}6$

(*iii*) We have $b_{YX} = r \cdot \frac{\sigma_Y}{\sigma_X} \Rightarrow \frac{4}{5} = \frac{3}{5} \times \frac{\sigma_Y}{3}$ $[\because \sigma_X^2 = 9 \text{ (Given)}]$

Hence $\sigma_Y = 4$

f. Find the equations of lines of regression for the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Obtain the estimate of X for $Y = 70$.

Ans **Solution.** Let $U = X - 68$ and $V = Y - 69$, then

$\bar{U} = 0, \bar{V} = 0, \sigma_U^2 = 4.5, \sigma_V^2 = 5.5$, Cov $(U, V) = 3$ and $r(U, V) = 0.6$

Since correlation coefficient is independent of change of origin, we get

$$r = r(X, Y) = r(U, V) = 0.6$$

We know that if $U = \dfrac{X - a}{h}$, $V = \dfrac{Y - b}{k}$, then

$$\bar{X} = a + h\bar{U}, \bar{Y} = b + k\bar{V}, \sigma_X = h\sigma_U \text{ and } \sigma_Y = k\,\sigma_Y$$

In our case $h = k = 1$, $a = 68$ and $b = 69$.

Thus $\quad \bar{X} = 68 + 0 = 68, \bar{Y} = 69 + 0 = 69$

$$\sigma_X = \sigma_U = \sqrt{4.5} = 2.12 \text{ and } \sigma_Y = \sigma_V = \sqrt{5.5} = 2.35$$

Equation of line of regression of Y on X is

$$Y - \bar{Y} = r\frac{\sigma_Y}{\sigma_X}(X - \bar{X})$$

i.e., $Y = 69 + 0.6 \times \dfrac{2.35}{2.12}(X - 68) \Rightarrow Y = 0.665X + 23.78$

Equation of line of regression of X on Y is

$$X - \bar{X} = r\frac{\sigma_X}{\sigma_Y}(Y - \bar{Y})$$

$\Rightarrow \quad X = 68 + 0.6 \times \dfrac{2.12}{2.35}(Y - 69)$ *i.e.,* $X = 0.54Y + 30.74$

To estimate X for given Y, we use the line of regression of X on Y. If $Y = 70$, estimated value of X is given by

$$\hat{X} = 0.54 \times 70 + 30.74 = 68.54,$$

where $\hat{X}$ is estimate of X.