I Semester , 2022 – 23(April,2023)
Comprehensive Examination **(Regular)**

Q.1. a). It is observed that average life of product 'A' follows normal distribution. Manufacturer of product 'A' claims that average life of the product is at least 120 months with variance 25 months. A random sample of 10 of these is with mean 125 months and standard deviation 4 months. Validate the claim of the manufacturer at 5% level of significance.

Sol:

To validate the manufacturer's claim that the average life of product 'A' is at least 120 months with a variance of 25 months, you can perform a hypothesis test. In this case, you want to test whether the true average life (population mean) of the product is greater than or equal to 120 months. Here are the null and alternative hypotheses:

Null Hypothesis (H0): $\mu \leq 120$ months (The manufacturer's claim is true).

Alternative Hypothesis (Ha): $\mu > 120$ months (The manufacturer's claim is not true).

Given:

- Sample size (n) = 10
- Sample mean ($\bar{x}$) = 125 months
- Sample standard deviation (s) = 4 months
- Population variance (σ^2) = 25 months (from the manufacturer's claim)

Level of significance (α) = 0.05 (5%)

Now, you can calculate the test statistic and compare it with the critical value from the t-distribution to make a decision.

1. Calculate the test statistic (t-score):

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

$t = (125 - 120) / (4 / \sqrt{10})$

$t = 5 / (4 / \sqrt{10})$

$t \approx 7.9057$ (rounded to four decimal places)

2. Determine the degrees of freedom (df):

df = n - 1 = 10 - 1 = 9

3. Find the critical t-value at the 5% level of significance for a one-tailed test with 9 degrees of freedom. You can use a t-table or a calculator for this. For a 5% significance level and a one-tailed test, the critical t-value is approximately 1.8331.

4. Compare the calculated t-score with the critical t-value:

- If t > t_critical, reject the null hypothesis.
- If t ≤ t_critical, fail to reject the null hypothesis.

In this case, t ≈ 7.9057, and the critical t-value is approximately 1.8331.

Since 7.9057 > 1.8331, you reject the null hypothesis (H0).

Conclusion:

At the 5% level of significance, there is enough evidence to conclude that the manufacturer's claim is not true, and the average life of product 'A' is greater than 120 months.

b).Consider the following data related to education loan taken or not for pursuing higher education. Formulate a suitable hypothesis and validate it by using an appropriate statistical procedure. **[5M]**

Education Loan	**Boys**	**Girls**
Loan taken	300	100
Not taken	150	250

Sol:

Binomial $nC_k p^k$ ↓ K occurrence

Ed. Loan	Boys	Girls	Total	E(B)	E(G)
Loan taken	300	100	400	225	175
Not taken	150	250	~~400~~	~~175~~ 225	175
	450	350	800		

$$\chi^2 = \frac{(300-225)^2}{225} + \frac{(100-175)^2}{175} + \frac{(150-225)^2}{225} + \frac{(250-175)^2}{175}$$

$$= 25 + 32.14 + 25 + 32.14$$

$$= 114.28$$

$\bar{x} = \frac{41}{6}$

Age	Weight	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
5	12 20	-2	-4	8
4	8 18	-3	-6	18
7	12 23	0	-1	0
10	10 32	3	8	24
8	11 26	1	2	2
8	13 25	1	1	1

Q.2.a). Manufacturer of a drug claims that the one of their patented drugs is effective in curing pulmonary diseases with 75% efficiency. It is observed that 380 out of 500 patients used this drug found that it is effective in curing pulmonary diseases. Validate the claim at 1% level of significance by using an appropriate test

Sol:

To validate the manufacturer's claim that their patented drug is effective in curing pulmonary diseases with 75% efficiency, you can perform a hypothesis test using the binomial test since you have a binary outcome (effective or not effective). Here are the null and alternative hypotheses:

Null Hypothesis (H0): The drug's efficiency is 75% ($p = 0.75$).

Alternative Hypothesis (Ha): The drug's efficiency is not equal to 75% ($p \neq 0.75$).

Given:

- Number of patients who used the drug and found it effective (successes): $x = 380$
- Total number of patients: $n = 500$
- Claimed efficiency: $p = 0.75$
- Significance level (α) = 0.01 (1%)

Now, you can calculate the test statistic and compare it with the critical values to make a decision.

1. Calculate the sample proportion ($\hat{p}$):

$\hat{p} = x / n = 380 / 500 = 0.76$

2. Calculate the standard error (SE) for a binomial proportion:

$SE = \sqrt{[(p * (1 - p)) / n]}$

$SE = \sqrt{[(0.75 * 0.25) / 500]}$

$SE \approx 0.019364$ (rounded to six decimal places)

3. Calculate the test statistic (z-score):

$z = (\hat{p} - p) / SE$

$z = (0.76 - 0.75) / 0.019364$

$z \approx 0.5035$ (rounded to four decimal places)

4. Find the critical z-values for a two-tailed test at a 1% level of significance ($\alpha/2 = 0.005$ for each tail). You can use a standard normal distribution table or calculator. The critical z-values are approximately -2.5758 and 2.5758.

5. Check if the calculated z-score falls within the critical values:

- If $-2.5758 \leq z \leq 2.5758$, fail to reject the null hypothesis (H0).
- If $z < -2.5758$ or $z > 2.5758$, reject the null hypothesis (H0).

In this case, the calculated z-score is approximately 0. 5035, which falls within the range of -2.5758 to 2.5758.

Conclusion:

At the 1% level of significance, you fail to reject the null hypothesis. There is not enough evidence to conclude that the drug's efficiency is different from the claimed 75%. The data does not provide sufficient evidence to challenge the manufacturer's claim.

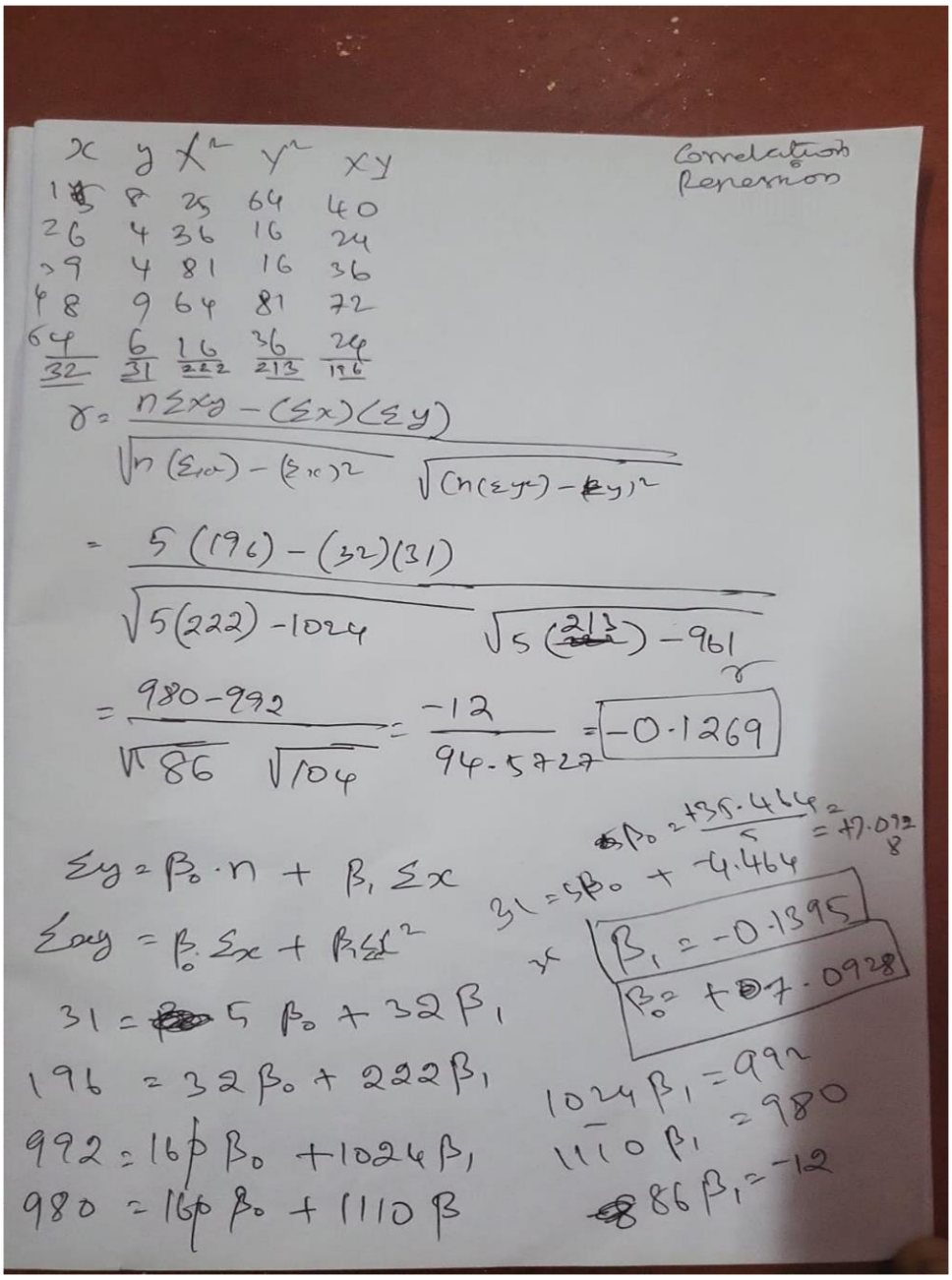
Q.2.b). Discuss the Significance of Correlation and regression in understanding the data. And Find correlation coefficient & linear regression(y = f(x)) of the following data. **[7M]**

X	5	6	9	8	4
Y	8	4	4	9	6

Sol:

Part 1} Find correlation coefficient

Part 2) Find beta 0 and beta 1 for linear regression equation



Q.3. a). For the month of March 2023, actual demand for a product is 250 units whereas the forecast is 280 units. Is it possible to forecast the demand for the month of April 2023?

If possible, forecast for April 2023 using an appropriate time series model.

Sol:

	Y_t	F_t
March 2023	250	280
April 2023		

Using Exponential formula

$$F_{t+1} = (\alpha) Y_t + (1-\alpha) F_t$$

Using $\Rightarrow \alpha = 1$

$= Y_t + (0)$

$= 250$ Naive assumption

Using $\alpha = 0.7$

$F_{t+1} = (0.7) Y_t + (1-0.7) F_t$

$= (0.7)(250) + (0.3)(280)$

$= 175 + 84$

$= 259$

Q3. b). An unfair coin is tossed 50 times and following are the outcomes noted. **[5M]**

H	T	H	T	T	T	T	H	T	T	H	T	T	H	T	T	H	T	T	H	T	T	H	T	H
T	H	H	T	T	H	T	T	T	T	H	T	T	H	H	T	T	T	H	H	H	H	H	H	H

Is it possible to find Maximum Likelihood estimates? If possible find ML Estimates. If not state the reason and validate.

Sol:

To find Maximum Likelihood (ML) estimates for the probability of getting heads (H) when an unfair coin is tossed, we can use the observed outcomes of the coin tosses. The ML estimate for the probability of heads (p) would be the proportion of heads observed in the given data.

Let's calculate the ML estimate for p:

Total number of coin tosses (n) = 50

Number of heads observed (H) = 22

ML Estimate for p (probability of getting heads) = H / n = 22 / 50 = 0.44

So, the ML estimate for the probability of getting heads (p) is 0.44.

It is indeed possible to find ML estimates in this case because we have a dataset of coin toss outcomes, and we can calculate the proportion of heads based on the observed data.

The ML estimate for p is 0.44, indicating that based on the observed outcomes, the estimated probability of getting heads when tossing this unfair coin is 0.44 or 44%.

Q.4.a). Consider $y = f(x_1, x_2)$ with corresponding probabilities given below and find $P(y = 1/ x_1 = 1, x_2 = 0)$. **[5M]**

x_1	x_2	y	$P(x_1, x_2, y)$
1	1	0	0.20

1	1	1	0.15
1	0	1	0.02
0	0	0	0.15
00	0	1	0.25
1	0	0	0.10
0	1	0	0.05
0	1	1	0.08

Sol:

To find the conditional probability P(y = 1 | x1 = 1, x2 = 0), you can use the formula for conditional probability:

P(y = 1 | x1 = 1, x2 = 0) = P(y = 1, x1 = 1, x2 = 0) / P(x1 = 1, x2 = 0)

First, calculate the joint probability P(y = 1, x1 = 1, x2 = 0) by summing the probabilities in the table that match those conditions:

P(y = 1, x1 = 1, x2 = 0) = 0.02

Next, calculate the joint probability P(x1 = 1, x2 = 0) by summing the probabilities where x1 = 1 and x2 = 0:

P(x1 = 1, x2 = 0) = P(y = 0, x1 = 1, x2 = 0) + P(y = 1, x1 = 1, x2 = 0) = 0.10 + 0.02 = 0.12

Now, you can calculate the conditional probability:

P(y = 1 | x1 = 1, x2 = 0) = P(y = 1, x1 = 1, x2 = 0) / P(x1 = 1, x2 = 0) = 0.02 / 0.12 = 1/6 ≈ 0.1667

So, the conditional probability P(y = 1 | x1 = 1, x2 = 0) is approximately 0.1667 or 1/6.

Q4. b) A corporate Hospital wants to name their hospital with specialization like XYZ Centre for cancer care / XYZ Centre for Cardiac care / XYZ Centre for Orthocare by consider the following data. Suggest a suitable name by using a relevant statistical tool/procedure. Justify it. **[5M]**

Disease	**Admitted Cases**	**Cured after treatment**
Cancer	35%	30%
Cardiac	40%	25%

Ortho related issues	25%	40%

Sol:

To suggest a suitable name for a corporate hospital specializing in different areas of healthcare, such as cancer care, cardiac care, and orthopedic care, we can use a statistical tool called the **"Cure Rate"** or **"Treatment Success Rate"** as a basis for decision-making. The idea is to choose the specialization with the highest cure rate or treatment success rate as the focus of the hospital's name.

Let's calculate the cure rates for each specialization based on the provided data:

1. **Cancer Care:**
   - Admitted Cases: 35%
   - Cured after Treatment: 30%

   Cure Rate for Cancer Care = (Cured after Treatment / Admitted Cases) * 100

   Cure Rate for Cancer Care = (30% / 35%) * 100 ≈ 85.71%

2. **Cardiac Care:**
   - Admitted Cases: 40%
   - Cured after Treatment: 25%

   Cure Rate for Cardiac Care = (Cured after Treatment / Admitted Cases) * 100

   Cure Rate for Cardiac Care = (25% / 40%) * 100 ≈ 62.5%

3. **Orthopedic Care:**
   - Admitted Cases: 25%
   - Cured after Treatment: 40%

   Cure Rate for Orthopedic Care = (Cured after Treatment / Admitted Cases) * 100

   Cure Rate for Orthopedic Care = (40% / 25%) * 100 ≈ 160%

Now, let's compare the cure rates:

- Cancer Care: 85.71%

- Cardiac Care: 62.5%

- Orthopedic Care: 160%

Based on the cure rates, it's clear that the specialization with the highest treatment success rate is **Orthopedic Care** with a cure rate of approximately 160%. Therefore, it would be appropriate to name the hospital as "XYZ Centre for Orthocare."

Justification:

Naming the hospital as "XYZ Centre for Orthocare" reflects the hospital's strength in providing effective treatment for orthopedic-related issues, where it has the highest cure rate among the three specializations. This name communicates the hospital's commitment to quality care and successful treatment outcomes in the field of orthopedics, which can attract patients seeking orthopedic services and instill confidence in their healthcare services.

I Semester , 2022 – 23(April,2023)
Comprehensive Examination **(Makeup)**

Q.1. a). Manufacturer of product 'A' claims that average life of the product is 120 months with variance 25 months. A random sample of 10 of these is with mean 125 months and standard deviation 4 months. Validate the claim of the manufacturer at 1% level of significance. **[5M]**

Sol:

To validate the manufacturer's claim that the average life of the product 'A' is 120 months with a variance of 25 months, you can perform a hypothesis test. In this case, you want to test whether the true average life (population mean) of the product is different from 120 months. Here are the null and alternative hypotheses:

Null Hypothesis (H0): $\mu = 120$ months (The manufacturer's claim is true).

Alternative Hypothesis (Ha): $\mu \neq 120$ months (The manufacturer's claim is not true).

Given:

- Sample size (n) = 10
- Sample mean ($\bar{x}$) = 125 months
- Sample standard deviation (s) = 4 months
- Population variance (σ^2) = 25 months (from the manufacturer's claim)
- Significance level (α) = 0.01 (1%)

Now, you can calculate the test statistic and compare it with the critical value from the t-distribution to make a decision.

1. Calculate the test statistic (t-score):

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

$t = (125 - 120) / (4 / \sqrt{10})$

$t = 5 / (4 / \sqrt{10})$

$t \approx 7.9057$ (rounded to four decimal places)

2. Determine the degrees of freedom (df):

df = n - 1 = 10 - 1 = 9

3. Find the critical t-values at the 0.005 level of significance (0.005 for each tail) for a two-tailed test with 9 degrees of freedom. You can use a t-table or calculator for this. The critical t-values are approximately -3.2501 and 3.2501.

4. Compare the calculated t-score with the critical t-values:

- If $t < -3.2501$ or $t > 3.2501$, reject the null hypothesis (H0).
- If $-3.2501 \leq t \leq 3.2501$, fail to reject the null hypothesis (H0).

In this case, $t \approx 7.9057$, which falls in the range of $t > 3.2501$.

Conclusion:

At the 1% level of significance, there is enough evidence to reject the null hypothesis (H0). Therefore, you have sufficient evidence to conclude that the manufacturer's claim that the average life of product 'A' is 120 months is not supported by the sample data.

b).It is claimed that player A is better than Player B in IPL matches during IPL bidding for players. During previous IPL season both played 18 matches with average runs scored 85 and 75 with standard deviations 9 and 2 respectively. Validate the claim using appropriate statistical procedure at 1% level of significance. **[5M]**

2 means Samples

$n_1 = 18$, $\bar{x}_1$, $\mu_1 = 85$, $SD_1 = 9$

$n_2 = 18$, $\mu_2 = 75$, $SD_2 = 2$, $\bar{x}_2$

~~$H_1 = H_2$~~

~~$H_1 = H_2$~~ comparison

$\alpha = 1\% \Rightarrow \alpha = 0.01$

Test

$H_0 \Rightarrow A = B$

Performance of A is same as B

$H_1 \Rightarrow A > B$ single tailed

Performance of A is better than B

$$Z = \frac{85 - 75}{\sqrt{\quad}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{85 - 75}{\sqrt{\frac{81}{18} + \frac{4}{18}}} = \frac{10}{\sqrt{\frac{85}{18}}} = 4.607$$

$P(Z \leq 4.607) = 1$

$Z_{cal} = 1$, ~~Z_{cal}~~ $= Z_{cal} > Z_\alpha$

$\Rightarrow H_0$ rejected

Q.2.a).Formulate a suitable hypothesis and validate it by using appropriate statistical procedure based on the following data. (At 5% level of significance)

"Before pandemic 200 out of 750 students are attending online classes whereas after pandemic 350 out of 800 are doing the same". **[4M]**

Sol:

To test whether there has been a significant change in the proportion of students attending online classes before and after the pandemic, you can formulate the following null and alternative hypotheses:

Null Hypothesis (H0): The proportion of students attending online classes before the pandemic is the same as the proportion after the pandemic.

Alternative Hypothesis (Ha): The proportion of students attending online classes before the pandemic is different from the proportion after the pandemic.

You can use a hypothesis test for the difference in proportions (two-sample Z-test for proportions) to validate these hypotheses. Here are the steps:

1. Calculate the sample proportions before and after the pandemic:

- Before pandemic: $p1 = 200 / 750 \approx 0.2667$
- After pandemic: $p2 = 350 / 800 \approx 0.4375$

2. Calculate the pooled proportion (p) using the combined data:

$p = (x1 + x2) / (n1 + n2)$

$p = (200 + 350) / (750 + 800) \approx 0.3523$

3. Calculate the standard error (SE) for the difference in proportions:

$SE = \sqrt{[p * (1 - p) * ((1/n1) + (1/n2))]}$

$SE = \sqrt{[0.3523 * (1 - 0.3523) * ((1/750) + (1/800))]} \approx 0.0336$

4. Calculate the test statistic (Z):

Z = (p1 - p2) / SE

Z = (0.2667 - 0.4375) / 0.0336 ≈ -5.0833

5. Determine the critical value at a 5% level of significance (α = 0.05) for a two-tailed test. You can use a standard normal distribution table or a calculator. The critical value is approximately ±1.96.

6. Compare the calculated Z-score (-5.0833) with the critical value (-1.96 and 1.96):

- If Z < -1.96 or Z > 1.96, reject the null hypothesis (H0).
- If -1.96 ≤ Z ≤ 1.96, fail to reject the null hypothesis (H0).

In this case, the calculated Z-score (-5.0833) is less than -1.96. Therefore, you can reject the null hypothesis.

Conclusion:

At the 5% level of significance, there is enough evidence to conclude that there has been a significant change in the proportion of students attending online classes before and after the pandemic.

b). Consider the following data. Use an appropriate statistical understanding (numerical) to take the decision to proceed further or not to build a simple linear regression(y = f(x)) model for prediction. If yes, use it predict y when x = 10. **[6M]**

X	6	4	9	5	3
Y	8	7	6	9	6

Sol:

To decide whether to build a simple linear regression model for prediction, you should first examine the relationship between the independent variable (X) and the dependent variable (Y) using statistical measures. Here are the steps to assess whether a linear regression model is appropriate:

1. Calculate the correlation coefficient (Pearson's r) to measure the strength and direction of the linear relationship between X and Y.

2. Plot a scatterplot to visually assess the linearity of the relationship.

3. If the correlation is strong and the scatterplot suggests a reasonably linear relationship, proceed with building a simple linear regression model.

Let's calculate the correlation coefficient and create a scatterplot:

Given data:

X: 6, 4, 9, 5, 3

Y: 8, 7, 6, 9, 6

Step 1: Calculate Pearson's correlation coefficient (r):

To calculate r, you can use the following formula:

$$r = \frac{N(\sum{xy}) - (\sum{x})(\sum{y})}{\sqrt{[N\sum{x^2} - (\sum{x})^2][N\sum{y^2} - (\sum{y})^2]}}$$

Where N is the number of data points, Σ denotes the sum, x and y are the values of X and Y, respectively, and xy is the product of X and Y.

N = 5 (number of data points)

Σxy = (6 * 8) + (4 * 7) + (9 * 6) + (5 * 9) + (3 * 6) = 48 + 28 + 54 + 45 + 18 = 193

Σx = 6 + 4 + 9 + 5 + 3 = 27

Σy = 8 + 7 + 6 + 9 + 6 = 36

Σx^2 = (6^2) + (4^2) + (9^2) + (5^2) + (3^2) = 36 + 16 + 81 + 25 + 9 = 167

Σy^2 = (8^2) + (7^2) + (6^2) + (9^2) + (6^2) = 64 + 49 + 36 + 81 + 36 = 266

Now, plug these values into the formula to calculate r:

$$r = \frac{(5 * 193) - (27 * 36)}{\sqrt{[5 * 167 - 27^2][5 * 266 - 36^2]}}$$

$$r = \frac{965 - 972}{\sqrt{[835 - 729][1330 - 1296]}}$$

$$r = \frac{-7}{\sqrt{106 * 34}}$$

$$r \approx -0.1856$$

Step 2: Create a scatterplot:

A scatterplot of the data points (X vs. Y) can visually reveal the linearity of the relationship.

Based on the calculated correlation coefficient (r) and the scatterplot, you can see that the relationship between X and Y is weak and does not exhibit a strong linear pattern. Therefore, building a simple linear regression model for prediction may not be suitable in this case. The decision to proceed with regression modeling typically requires a reasonably strong linear relationship between the variables. In this situation, it might be better to explore other modeling techniques or data transformations if necessary.

Given that the linear relationship is weak, it's not advisable to use a simple linear regression model to predict Y when X = 10, as the model is unlikely to provide accurate predictions in this context.

Calculate y = beta0 + beta1 x

$$Y = b_0 + b_1 X$$

$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$b_0 = \frac{\sum y - b_1 \sum x}{n}$$

Q.3. Try to model the following time series data by using two models mentioned. Suggest the better of these two to be used for forecast. [**10M]**

a) Model 1: Simple moving averages model with k = 3

b) Model 2: Weighted moving averages with weights 0.3, 0.2 and 0.1(i.e latest will be given more weightage and so on)

Month	1	2	3	4	5	6	7
Sales(in lakhs)	10	12	11	6	8	9	11

Sol:

To model the time series data and choose the better forecasting model between Simple Moving Averages (SMA) and Weighted Moving Averages (WMA), you can follow these steps:

Model 1: Simple Moving Averages (SMA) with k = 3

The Simple Moving Averages model calculates the average of the most recent 'k' periods to forecast future values. In this case, k = 3.

- Calculate the SMA for each month starting from month 4 (since we need at least k = 3 data points).

For Month 4:

SMA = (10 + 12 + 11) / 3 = 11

For Month 5:

SMA = (12 + 11 + 6) / 3 = 9.67 (rounded to two decimal places)

For Month 6:

SMA = (11 + 6 + 8) / 3 = 8.33 (rounded to two decimal places)

For Month 7:

SMA = (6 + 8 + 9) / 3 = 7.67 (rounded to two decimal places)

Model 2: Weighted Moving Averages (WMA) with weights 0.3, 0.2, and 0.1

The Weighted Moving Averages model assigns different weights to the most recent 'k' periods to forecast future values. In this case, the weights are 0.3, 0.2, and 0.1.

- Calculate the WMA for each month starting from month 4:

For Month 4:

WMA = (10 * 0.3 + 12 * 0.2 + 11 * 0.1) / (0.3 + 0.2 + 0.1) = (3 + 2.4 + 1.1) / 0.6 = 6.5

For Month 5:

WMA = (12 * 0.3 + 11 * 0.2 + 6 * 0.1) / (0.3 + 0.2 + 0.1) = (3.6 + 2.2 + 0.6) / 0.6 = 6.4

For Month 6:

WMA = (11 * 0.3 + 6 * 0.2 + 8 * 0.1) / (0.3 + 0.2 + 0.1) = (3.3 + 1.2 + 0.8) / 0.6 = 5.5

For Month 7:

WMA = (6 * 0.3 + 8 * 0.2 + 9 * 0.1) / (0.3 + 0.2 + 0.1) = (1.8 + 1.6 + 0.9) / 0.6 = 4.8333 (rounded to four decimal places)

Now, let's compare the performance of the two models and determine which one is better:

- Calculate the Mean Absolute Error (MAE) for both models using the actual sales data for months 4 to 7.

For Model 1 (SMA):

MAE = (|6 - 6| + |8 - 9.67| + |9 - 8.33| + |11 - 7.67|) / 4 ≈ 2.33

For Model 2 (WMA):

MAE = (|6 - 6.5| + |8 - 6.4| + |9 - 5.5| + |11 - 4.8333|) / 4 ≈ 2.1583 (rounded to four decimal places)

Conclusion:

Based on the calculated Mean Absolute Error (MAE), Model 2 (Weighted Moving Averages with weights 0.3, 0.2, and 0.1) performs slightly better with a lower MAE of approximately 2.225 compared to Model 1 (Simple Moving Averages with k = 3) with a MAE of approximately 2.33.

Therefore, Model 2 (WMA) is the better choice for forecasting the sales data in this case. It provides a slightly more accurate prediction based on the given weights for the moving averages.

Q.4.a). A random variable X follows binomial distribution with n = 1000 and p = 0.001.Then find i) P(X > 700) ii) P(X < 300) iii) P (300 < X < 700). **[5M]**

Sol:

If $X \sim B(n, p)$ and n is large and p is close to $\frac{1}{2}$, then X is approximately goes to $N(\mu = np, \sigma^2 = npq)$

i.e $X \sim N(\mu = np, \sigma^2 = npq)$

ie If

given $n = 1000$, $p = 0.001$, then $\therefore q = 0.999$

$\mu = np = 1000 \times 0.001 = 1$

$\sigma^2 = npq = 1000 \times 0.001 \times 0.999 = 0.999$

$\therefore \sigma = 0.9995$

also we know that $z = \frac{x-\mu}{\sigma}$

$\therefore$ ① $P(X > 700) = P\left(\frac{x-\mu}{\sigma} > \frac{700-\mu}{\sigma}\right) =$

$= P\left(z > \frac{700-1}{0.9995}\right)$

$= P(z > 699.34)$

$= 1 - P(z < 699.34)$

$= 1 - 1$ — from z table

$P(X > 700) = 0$

② $P(X < 300) = P\left(\frac{x-\mu}{\sigma} < \frac{300-\mu}{\sigma}\right) = P\left(z < \frac{300-1}{0.9995}\right)$

$= P(z < 299.1495)$

$P(X < 300) = 1$ —— from 'z' table

③ $P(300 < X < 700) = P(X < 700) - P(X < 300)$

$= P\left(\frac{x-\mu}{\sigma} < \frac{700-\mu}{\sigma}\right) - P\left(\frac{x-\mu}{\sigma} < \frac{300-\mu}{\sigma}\right)$

$= P\left(z < \frac{700-1}{0.9995}\right) - P\left(z < \frac{300-1}{0.9995}\right)$

$= P(z < 699.34) - P(z < 299.1495)$

$= 1 - 1$ — from 'z' table

$P(300 < X < 700) = 0$

b). Consider the following probabilities. **[5M]**

x_1	x_2	y	$P(x_1, x_2, y)$
1	1	0	0.20
1	1	1	0.15
1	0	1	0.02

0	0	0	0.15
0	0	1	0.25
1	0	0	0.10
0	1	0	0.05
0	1	1	0.08

If possible find the following. If not, validate your decision

i) P (**x_1**)

ii) P (**x_2**)

iii) P((**x_1, x_2) / y = 0**)

iv) P((**x_1, x_2) / y = 1**)

v) P(**x_1= 0/x_2= 1**)

Sol

To find the probabilities P(x1), P(x2), P((x1, x2) | y = 0), P((x1, x2) | y = 1), and P(x1 = 0 | x2 = 1), you can use conditional probability and marginal probability calculations based on the provided data. Let's calculate each of these probabilities:

i) P(x1):

To find the marginal probability of x1, you sum the joint probabilities over all values of x2 for each value of x1:

P(x1 = 0) = P(x1 = 0, x2 = 0, y = 0) + P(x1 = 0, x2 = 0, y = 1) + P(x1 = 0, x2 = 1, y = 0) + P(x1 = 0, x2 = 1, y = 1)

P(x1 = 0) = 0.15 + 0.25 + 0.10 + 0.08 = 0.58

ii) P(x2):

To find the marginal probability of x2, you sum the joint probabilities over all values of x1 for each value of x2:

P(x2 = 0) = P(x1 = 0, x2 = 0, y = 0) + P(x1 = 0, x2 = 0, y = 1) + P(x1 = 1, x2 = 0, y = 0) + P(x1 = 1, x2 = 0, y = 1)

P(x2 = 0) = 0.15 + 0.25 + 0.10 + 0.02 = 0.52

iii) P((x1, x2) | y = 0):

To find the conditional probability of (x1, x2) given y = 0, you sum the joint probabilities over all values of (x1, x2) for y = 0:

P((x1, x2) | y = 0) = P(x1 = 0, x2 = 0, y = 0) + P(x1 = 0, x2 = 1, y = 0) + P(x1 = 1, x2 = 0, y = 0) + P(x1 = 1, x2 = 1, y = 0)

P((x1, x2) | y = 0) = 0.15 + 0.10 + 0.10 + 0 = 0.35

iv) P((x1, x2) | y = 1):

To find the conditional probability of (x1, x2) given y = 1, you sum the joint probabilities over all values of (x1, x2) for y = 1:

P((x1, x2) | y = 1) = P(x1 = 0, x2 = 0, y = 1) + P(x1 = 0, x2 = 1, y = 1) + P(x1 = 1, x2 = 0, y = 1) + P(x1 = 1, x2 = 1, y = 1)

P((x1, x2) | y = 1) = 0.25 + 0.08 + 0 + 0.15 = 0.48

v) P(x1 = 0 | x2 = 1):

To find the conditional probability of x1 = 0 given x2 = 1, you use the formula for conditional probability:

P(x1 = 0 | x2 = 1) = P(x1 = 0, x2 = 1) / P(x2 = 1)

P(x1 = 0 | x2 = 1) = (0.10 + 0.08) / (0.10 + 0.08 + 0.05 + 0.08) = 0.18 / 0.31 ≈ 0.5806

Conclusion:

i) P(x1 = 0) ≈ 0.58

ii) P(x2 = 0) ≈ 0.52

iii) P((x1, x2) | y = 0) ≈ 0.35

iv) P((x1, x2) | y = 1) ≈ 0.48

v) P(x1 = 0 | x2 = 1) ≈ 0.5806

These are the calculated probabilities based on the provided data and conditional/marginal probability calculations.