

Assessment

Introduction To Statistics (April 2023)

Theory section

1) Define

a) Correlation vs Covariance

Ans -> **Correlation** shows both direction and strength of the relationship between two variables. Its value is always between -1 and +1.

- Close to +1 : Strong positive relation
- Close to -1 : Strong negative relation
- Close to 0 : weak or no linear relation

Covariance means how two values (variables) move together.

- Positive covariance : Both increase or decrease together.
- Negative covariance : One increase while the other decreases.

Difference → Correlation shows direction and strength of the relationship (between -1 and +1) while covariance only shows whether variables move together or opposite without showing strength.

b) Null and Alternative Hypothesis

Ans. → Null Hypothesis (H_0) means there is no effect or no difference between two groups.

→ Alternative Hypothesis (H_1) means there is an effect or a difference between two groups.

c) p-value and significance level

p-value :

- The p-value tells how likely our result is if the null hypothesis is true (Probability between 0 and 1).
- Small p-value → our result is unusual → we reject H_0
- Large p-value → our result is normal → we do not reject H_0

Significance level :

- This is the cutoff value we choose usually 0.05 (5%).
- If $p \leq \alpha$ (alpha sign) → We reject Null Hypothesis (H_0)
- If $p > \alpha$ → do not reject Null Hypothesis (H_0)

Practical Section

2) The stock prices of Company A and B over 20 days are:

- Company A: 45, 47, 48, 50, 52, 53, 55, 56, 58, 60, 62, 64, 65, 67, 69, 70, 72, 74, 76, 77
- Company B: 52, 54, 55, 57, 59, 60, 61, 62, 64, 66, 67, 69, 71, 73, 74, 76, 78, 80, 82, 83

→ Tasks:

Que : Calculate Covariance and Pearson Correlation between A and B.

Ans → Covariance :

Mean = Sum of number / Number of values

Mean of A :

$$\text{Sum of A} = 45 + 47 + 48 + 50 + 52 + 53 + 55 + 56 + 58 + 60 + 62 + 64 + 65 + 67 + 69 + 70 + 74 + 76 + 77 \rightarrow 1220$$

Total number = 20

$$\text{Mean of A} \rightarrow 1220 / 20 = 61$$

Mean of B

$$\text{Sum of B} = 52 + 54 + 55 + 57 + 59 + 60 + 61 + 62 + 64 + 66 + 67 + 69 + 71 + 73 + 74 + 76 + 78 + 80 + 82 + 83 \rightarrow 1343$$

$$\text{Mean of B} \rightarrow 1343 / 20 = 67.15$$

$$\text{Cov}(X,Y) = (\text{Sum of } [(X_i - \text{mean of } A) * (Y_i - \text{mean of } B)]) / (n - 1)$$

$$\text{Cov}(X,Y) = 1853 / (20 - 1)$$

$$\text{Covariance}(X,Y) = 97.53$$

Correlation :

$$\text{Standard deviation (sd) of } A = \sqrt{(\text{sum of } (X_i - \text{mean})^2 / (n-1))}$$

$$= \sqrt{1936} / (20-1)$$

$$= \sqrt{1936} / 19$$

$$= \sqrt{101.89}$$

$$\text{SD of } A = 10.25$$

$$\text{Standard deviation (sd) of } B = \sqrt{(\text{sum of } (Y_i - \text{mean})^2 / (n-1))}$$

$$= \sqrt{1778.55} / (20-1)$$

$$= \sqrt{93.60}$$

$$\text{SD of } B = 9.67$$

$$\text{Correlation (r)} = \text{covariance} / (\text{SD of } A * \text{SD of } B)$$

$$r = 97.53 / (10.25 * 10.49)$$

$$r = 97.53 / 107.42$$

$$r = 0.9986$$

Que : Interpret the nature of their relationship

- The covariance is positive (97.53) which means when the price of Company A increases, the price of Company B also tends to increase.
- The correlation is 0.99 which is very close to +1.
- This shows a very strong positive relationship between stock prices of both companies.

3) A two-sample z-test is conducted to compare exam scores:

- **Group 1: [85, 88, 92, 87, 90]**
 - **Group 2: [78, 80, 75, 82, 76]**
-
- **Test at 5% significance level.**
 - **Assume population std. dev = 5 for both groups.**
 - **Perform the test and state your conclusion.**

Ans → $n_1 = 5$, $n_2 = 5$, $\mu_1 = 5$, $\mu_2 = 5$ (parameters)

Standard Deviation (sd) = 5 for both Group1 or Group2

Group 1 mean : $85 + 88 + 92 + 87 + 90 / 5 \rightarrow 442 / 5 \rightarrow 88.4$

Group 2 mean: $78 + 80 + 75 + 82 + 76 / 5 \rightarrow 391 / 5 \rightarrow 78.2$

Calculation of the Z-Test statistic

$$Z = (\text{group1 mean} - \text{group mean2}) / \sqrt{((\mu_1^2) / n) + ((\mu_2^2) / n)}$$

$$Z = (88.4 - 78.2) / \sqrt{((5^2) / 5) + (5^2) / 5}$$

$$Z = 10.2 / \sqrt{5 + 5}$$

$$Z = 10.2 / \sqrt{10}$$

$$z \sim 3.2255$$

4) A survey was conducted to check ice cream preference by gender:

	Chocolate	Vanilla	Strawberry
Male	20	15	10
Female	25	20	30

- Perform a Chi-Square Test for Independence.
- State H_0 and H_1 . Interpret your result.

Ans → The Chi-Square Test for independence checks if there is a relationship between two categorical variables (Gender and Ice cream preference)

- Null Hypothesis (H_0) : Gender and ice cream preference are independent.
- Alternative Hypothesis (H_1) : Gender and ice cream preference are dependent.

Preference	Chocolate(C1)	Vanilla(C2)	Strawberry(C3)	Row Total (rt)
Male(R1)	20	15	10	45
Female(R2)	25	20	30	75
Col total (ct)	45	35	40	N = 120

Step 2 : Degrees of freedom and Critical value

Degree of freedom (df) : $df = (\text{Row} - 1) * (\text{Col} - 1) = (2-1)*(3-1) = 2$

Critical value (χ^2_{critical}) : For df = 2 and mu = 0.05, the critical value is 5.5991

Step 3 : Calculation of expected frequencies (E)

$$E = (rt * ct) / N$$

- Male , chocolate(E11) : $(45 * 45) / 120 = 2025 / 120 \rightarrow 16.88$ (E11 means row 1 col 1)
- Male , Vanilla(E12) : $(45*35) / 120 = 1575 / 120 \rightarrow 13.125$
- M, Strawberry(E13) : $(45 * 10) / 120 = 1800 / 120 \rightarrow 15.00$
- F, Chocolate(E21) : $(75 * 45) / 120 = 3375 / 120 \rightarrow 28.123$
- F, Vanilla(E22) : $(75 * 35) / 120 = 2625 / 120 \rightarrow 21.88$
- F, Strawberry(E23) : $(75 * 40) / 120 = 3000 / 120 \rightarrow 25.00$

O means observation O1 = R1C1 (Male , chocolate)

O2 = R1C2 (M , Vanilla) , O3 = R1C3 (M , Strawberry) O4 = R2C1 , O5 = R2C2 , O6 = R2C

Step 4 : Calculation of the test Statistic (χ^2)

$$\chi^2 = \text{sum of } ((O - E)^2 / E)$$

Cell	O (Observation)	E (E11,E12 etc)	(O - E)	$(O - E)^2$	$(O - E)^2 / E$
M, Choco	20	16.875	3.125	9.7344	0.5767
M , Vanilla	15	13.125	1.875	3.4969	0.2663
M, Straw	10	15.00	-5.000	25.00	1.6667
F , Choco	25	28.125	-3.125	9.7969	0.3483
F , Vanilla	20	21.875	-1.875	3.5354	0.1615

F , Straw	30	25.00	5.000	25.00	1.000
Total $(O - E)^2 / E \rightarrow 4.0295$					

$$X^2 = 4.0295$$

Interpretation and Conclusion

→ Decision : The calculated χ^2 value (4.0295) is less than critical value (5.991). We fail to reject the null Hypothesis (H_0).