**Ministry of Higher Education & Scientific Research**
**Salahaddin University-Erbil**
**College of Administration and Economics**
**Statistics & Informatics Department**
**High Diploma**

**QUESTIONS BANK OF MULTIVARIATE STATISTICAL ANALYSIS**

**Q1/ Are the following statements TRUE or FALSE? Correct the false statements?**

1) In multivariate statistical analysis, the word "multivariate" indicates that in a single analysis two dependent variables are simultaneously used.
2) $X_{jk}$ = The $j^{th}$ item for the measurement of the $k^{th}$ variable.
3) Comparing the means of the weight, length, and blood sugar level variables for people in Iraq, France, Japan, and Iran is an example of analysis of variance (ANOVA).
4) $(ABC)' = A'\,B'C'$
5) When $A^{-1} = A'$ , A is called Idempotent matrix.
6) If $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 5 & 9 \\ 4 & 9 & -1 \end{bmatrix}$ and let $\boldsymbol{C} = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 1 & 4 \\ 3 & 1 & 4 \end{bmatrix}$, then $tr(C^{-1}AC)$ =5
7) Let A = $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 9 \\ 0 & 9 & -1 \end{bmatrix}$ and $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$. Then, quadratic form $Q(\underline{X}) = x_1{}^2 + 9x_1x_2 + 5x_2{}^2 + 2x_3{}^2$
8) If $\underline{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $A = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$, then $Q(\underline{X})$ is p.s.d.
9) If $A = \begin{bmatrix} -7 & 0 \\ 0 & 7 \end{bmatrix}$ , then $Q(\underline{X})$ is p.d.
10) If $\underline{X}^{(1)}, \underline{X}^{(2)}$are dependent, then $E\left(\underline{X}^{(1)} - \underline{\mu}^{(1)}\right)\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right)' = \mathbf{0}$
11) In multivariate statistical analysis, the word "multivariate" indicates that two or more dependent variables are simultaneously used in a single analysis.
12) Correlation for multiple regression predictions always infers causation.
13) The higher the multicollinearity, the smaller the standard error for the regression coefficients.
14) Data errors and observations which represent extreme magnitudes on variables are both causes of outliers.

**Q2/ For the following data that are normally distributed ($\underline{\boldsymbol{X}} \sim \boldsymbol{N}(\underline{\boldsymbol{\mu}}, \Sigma)$), find X matrix, mean vector, var-covar matrix, and prove or disprove that $\underline{\boldsymbol{X}}^{(1)}$ and $\underline{\boldsymbol{X}}^{(2)}$ are independent, if:**

$$\underline{X}^{(1)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \underline{X}^{(2)} = [x_3]$$

| | | | | | |
|---|---|---|---|---|---|
| ***x1*** | 1 | 2 | 3 | 4 | 5 |
| ***x2*** | 5 | 4 | 3 | 2 | 1 |
| ***x3*** | -2 | -2 | -2 | -2 | -2 |

**Q3/** Let $\underline{X} \sim N(\underline{\mu}, \Sigma)$, where $\underline{X} = \begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$ , $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

If $\underline{X}^{(1)}$and $\underline{X}^{(2)}$ are independent. Show that $\Sigma_{12} = 0$

**Q4/ If the quadratic form $Q(\underline{X}) = 2x_1{}^2 + 4x_3{}^2 + 2x_1x_2 + 6x_2x_3 + 5x_4{}^2 + 7x_1x_3$**

Find:

1. The matrix A,
2. What is the classification of $Q(\underline{X})$ and why? (p.d., p.s.d, n.d., n.s.d)

**Q5/ Multiple choices:** Select the best answer for the following statements:

❖ __ used when a numerical predictor has a curvilinear relationship with the response.

*a. Simple regression* *b. Multiple regression* *c. Quadratic regression*

*d. all of them* *e. None of them*

❖ ____ used to check the assumptions of the regression model.

*a. Stepwise regression* *b. $R^2$ Adjusted* *c. Correlation*

*d. all of them* *e. None of them*

❖ ____ worst kind of outlier, can totally reverse the direction of association between x and y.

*a. Residual plots* *b. independent variable* *c. dependent variable*

*d. bar chart* *e. None of them*

❖ __ problem that can occur when the information provided by several predictors overlaps.

*a. Cause and effect* *b. Logistic regression* *c. Factor analysis*

*d. Error* *e. None of them*

**Q6/ Fill the blanks** for the following statements:

- ✓ The most popular rotation strategy in factor analysis is __________________.
- ✓ __________________ is a correlational technique that allows you to evaluate the relationship between two variables with the effects of a third removed from both of them.
- ✓ The number of possible discriminant functions in a discriminant analysis is limited to ___________ or to _______________________, whichever is less.
- ✓ In the formula for the Pearson correlation coefficient, the calculation of variability can be found in __________________________

**Q7/** For every vector $\underline{\alpha}$ if $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$are independent. Let $X_i$ be any Component of $\underline{X}^{(1)}$, Show That for all linear Combination $\underline{\alpha}\ \underline{X}^{(2)}$ Which is Minimize the Variance $\left(X_i - \underline{\alpha X}^{(2)}\right)$.

**Q8/** If we have three Variables (p=3), $R^2_{1.23} = \frac{{\rho_{12}}^2+{\rho_{13}}^2-2\rho_{12}\rho_{13}\rho_{23}}{(1-{\rho_{23}}^2)}$ , and $\rho^2{}_{13.2} = \left(\frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{1-{\rho_{12}}^2}\sqrt{1-{\rho_{23}}^2}}\right)^2$.

Prove that $1 - R^2_{1.23} = (1-\rho^2_{12})(1-\rho^2_{13.2})$