NJU Statistics and Data Analysis in Earth Science
Spring 2026

Mid-Term Examination

True-False Questions

1.	In regression analysis confidence intervals are calculated via reference to the *F* distribution.	**F**
2.	A scalar is a single element of a field or range of variation that pertains to a variable or vector.	**T**
3.	Ternary charts are used to display proportional data.	**T**
4.	By removing noise and redundancy in the data, dimensionality reduction represents a strategy that can be used to mitigate the effects of dimensionality's "curse".	**T**
5.	The lines of all bivariate linear regression models pass through the bivariate mean.	**T**
6.	Discrimination is the process of placing an object or set of observations into one or more a priori-defined groups based on their similarity or dissimilarity relations with other group members	**F**
7.	Multiple regression specifies an *m*-dimensional geometric plane that passes through the multivariate mean.	**T**
8.	Cluster diagrams (dendrograms) can be used to display the structure similarities in a dataset without inducing distortion.	**F**
9.	Most descriptive statistical quantities (e.g., variance, covariance, correlation) can be calculated using matrix equations.	**T**
10.	In a standard data matrix the variables are represented by the matrix's rows.	**F**
11.	A median can be calculated for any variable.	**F**
12.	The term "regression" in regression analysis refers to the tendency of children to have the characteristics of their parents.	**F**
13.	In the regression equation $y = 94.67 - 2.68x$ the coefficient -2.68 is the value of the slope.	**T**
14.	Bar charts are used to visualize patterns and/or trends in categorical data.	**T**
15.	It is appropriate to use it to predict values that lie beyond the range of your data.	**F**
16.	In analyzing data using scatterplots it is appropriate to ignore the geometric relations between variable vectors.	**F**
17.	A variable is a characteristic that may differ from one individual to another.	**T**
18.	Eigenvectors are vectors whose span remains constant despite linear transformation of their coordinate system.	**T**
19.	A variable's variance is the average of the deviations from its arithmetic mean.	**F**
20.	Ordinary least squares (OLS) bivariate regression specifies lines that minimize variation in the dependent variable.	**T**
21.	A variable range is the middle number of an ordered dataset.	**F**
22.	In ANOVA the SSQ_{Model} is calculated used the deviation from the regression line.	**F**
23.	High correlation = high multicolinearity	**T**
24.	A variable's coefficient of variation is the ratio between its mean and standard deviation.	**T**

25.	Parametric statistical tests assume variables have unequal variances.	**F**

Matching Questions

26. **Match the following multivariate data-analysis methods (left) with their type classification (right).**

a. **Canonical variates analysis (CVA) (I)**
b. **Singular value decomposition (SVD) (II)**
c. **Principal components analysis (PCA) (II)**
d. **Multidimensional scaling (II)**
e. **Linear discriminant analysis (LDA) (I)**

i. Classification method
ii. Pooled-sample method

27. **Match the following equations to their appropriate regressions (below). Note each category may fit more than a single equation.**

i.
$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon_i$$

ii.
$$y_i = \alpha + \beta_1 x_i + \epsilon_i$$

iii.
$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2^2 + \beta_3 x_3^3 + \epsilon_i$$

iv.
$$\begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{1_i} & \sum_{i=1}^n x_{2_i} \\ \sum_{i=1}^n x_{1_i} & \sum_{i=1}^n x_{1_i}^2 & \sum_{i=1}^n x_{1_i} x_{2_i} \\ \sum_{i=1}^n x_{2_i} & \sum_{i=1}^n x_{1_i} x_{2_i} & \sum_{i=1}^n x_{2_i}^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1_i} y_i \\ \sum_{i=1}^n x_{2_i} y_i \end{pmatrix}$$

v.
$$\begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{pmatrix} \cdot \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix}^{-1}$$

a. **Ordinary least-squares linear regression (II)**
b. **Curvilinear least-squares linear regression (III & V)**
c. **Multiple least squares linear regression (I & IV)**

28. Match the following *Q*-mode distances (left) to their appropriate equation.

i.
$$\sum_{i=0}^{n} (x_{ij} - x_{ik})^2$$

ii.
$$\frac{\sum_{i=1}^n x_{ik} x_{jk}}{\sqrt{\sum_{i=1}^n x_{ik}^2 \cdot \sum_{i=1}^n x_{jk}^2}}$$

iii.
$$\sum_{i=0}^{n} \sqrt{(x_{ij} - x_{ik})^2}$$

iv.
$$\frac{\frac{\sum_{k=1}^p \left| x_{ik} - x_{jk} \right|}{Range_k}}{p}$$

v.
$$\left[\bar{X}_i - \bar{X}_j\right]' \cdot \left[s_{ij}^2\right] \cdot \left[\bar{X}_i - \bar{X}_j\right]$$

a. **Cosine θ "distance" (II)**
b. **Euclidean distance (III)**
c. **Gower distance (IV)**
d. **Squared Euclidean distance (I)**
e. **Mahalanobis distance (V)**

29. Match the following matrices with the descriptive matrix terms (below). Note each matrix may fit more than a single descriptive category.

i.

5	1	2	2
1	10	4	3
7	9	6	8
1	9	1	8

ii.

4.9	9.6	8.3	5.1	2.9
9.9	2.7	1.2	10.0	5.3
1.8	3.2	8.0	7.4	10.9
2.5	8.4	8.2	9.5	8.9

iii.

9.1	8.7	4.5	4.1
8.7	7.0	7.1	1.7
4.5	7.1	3.9	3.6
4.1	1.7	3.6	5.6

iv.

9	1	2	1
2	2	10	2
7	6	2	1
6	7	2	9

v.

10.0	5.0	6.2	9.3
5.0	9.0	4.1	2.1
4.1	6.2	2.0	3.6
3.6	2.1	9.3	7.3

a. **Square matrix (I, III, IV, V)**
b. **Rectangular matrix (II)**
c. **Symmetric matrix (III)**
d. **Asymmetric matrix (I, II, IV, V)**
e. **4 x 4 matrix (I, III, IV, V)**

30. Match the data types (left) with their correct examples right).

a. **Interval data (IV)**	i. cat, dog, bunny, sheep, guinea pig
b. **Mixed-mode data (V)**	ii. 1.56mm, 2.91mm, 4.39mm, 4.05mm, 1.39mm
c. **Nominal data 9 (I)**	iii. small, long, medium, small, long
d. **Ordinal data (III)**	iv. 0.0°C, -273°C, 357°C, 100°C, -196°C
e. **Ratio data (II)**	v. horse, 15,5 mm, 5, ladder, 10.5 cm^3

31. Match the values (left) with the definitions in the context of a normal distribution.

a. **Range (IV)**

b. **3σ (III)**

c. **2σ (V)**

d. **$\bar{x}$ (II)**

e. **σ (I)**

i. 68.26%

ii. $\frac{\sum_{i=1}^{N} x_i}{N}$

iii. 99.72%

iv. max(x)-min(x

v. 95.44%

32. Match the expressions (left) to their proper places in the OLS ANOVA table (right).

a. MS_A/MS_B **(III)**

b. $\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ **(IV)**

c. $\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$ **(I)**

d. $SS_A/(m-1)$ **(II)**

e. $\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$ **(V)**

Source of Variation	Sum of Squares	Dof	Mean Squares	F Test
Due to Model	i.		ii.	iii.
Unexplained (Error)	iv.			
Total Variation	v.			

33. Write the value of the number present each of the listed matrix cells (right)

Matrix A

23.77	24.56	7.39	11.33	15.37	13.30	6.34	18.06	5.02	17.21
15.46	6.11	10.63	25.65	10.99	5.97	10.16	12.85	20.93	16.85
25.16	25.86	18.05	5.90	25.21	25.85	1.48	5.05	10.13	18.77
5.03	10.58	25.80	9.32	7.39	5.79	5.83	3.02	2.85	7.60
16.42	20.05	11.70	6.16	25.42	25.06	19.01	23.94	6.89	21.18
20.77	3.59	2.52	8.60	1.52	9.90	13.61	8.02	17.99	11.81
16.27	24.25	23.59	14.70	15.76	16.95	5.82	19.98	9.05	14.50
9.36	12.49	8.44	8.34	17.22	12.09	20.34	21.67	7.23	19.90
15.33	25.80	1.35	23.41	10.42	21.81	17.82	5.63	5.85	17.79
8.13	17.96	1.02	5.90	25.99	7.74	2.45	3.82	18.22	19.39

a. **12.85**

b. **10.13**

c. **5.97**

d. **21.67**

e. **10.42**

i. A_{82}

ii. A_{39}

iii. A_{62}

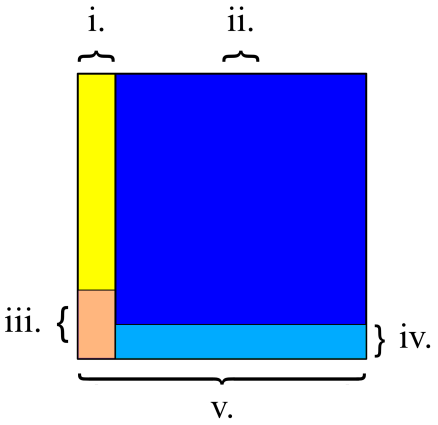
iv. A_{88}

v. A_{95}

34. Match the following parts of Bayes Equation (left) with their correct expressions (right)

a.	$P(E \mid \neg H)$ **(III)**	i.	Likelihood
b.	$P(H)$ **(V)**	ii.	Posterior probability
c.	$P(H \mid E)$ **(II)**	iii.	Probability of seeing evidence the hypothesis is false
d.	$P(E \mid H)$ **(I)**	iv.	Probability of the hypothesis being false
e.	$P(\neg H)$ **(IV)**	v.	Prior probability

35. **Match the regions on the diagram below to their appropriate expressions in Bayes equation.**



a. $P(E \mid \neg H)$ **(IV)**

b. $P(E \mid H)$ **(III)**

c. $P(H)$ **(I)**

d. #👤 **(V)**

e. $P(\neg H)$ **(II)**

Multiple-Choice Questions

36. Identify the statements that are true of an analysis of variance (ANOVA) test for the significance of a linear regression. Choose all options that apply.

a. This test's null hypothesis (H_0) is that the regression model accounts for no more of the variance of the observed data than does the sum of deviations from the regression model.
b. The test's measure of variation due to the model is found by using the regression equation to project the observed data points onto the regression line and assessing the squared variation of these projected points from the sample mean.
c. The test's measure of variation due to error about the regression model is found by summing the squared distances between the observed data points and their projected positions along the regression model.
d. In a bivariate regression problem, the mean squares due to the model will equal the sum of squares due to the models
e. All of the above.

37. Identify the statements that are true of the correlation index Choose all options that apply.

    a. Use of correlation coefficient assumes that the variables are all referenced to the same unit and all have similar value magnitudes.
    b. The correlation coefficient is an unbounded index of angular association between two variables.
    c. A correlation coefficient of 0.80 indicates a scatterplot of the variables will exhibit a linear trend with a positive slope.
    d. All of the above.
    e. None of the above.

38. Identify the problems inherent in analyzing compositional data (e.g., ratios, percentages) Choose all options that apply.

    a. Variable values for each sample constrained to sum to a constant and common value.
    b. Calculation or compositional values applies in inconsistent transformation to the data, which results in a larger change in the magnitudes of some variable values than others.
    c. Loss of variable independence.
    d. Compositional transformation alters the structure of geometric relations among the variables
    e. All of the above.

39. Identify the statements that are true of regression analysis. Choose all options that apply.

    a. Regression represents the study of the manner in which one or more variables can be used to predict the values of another variable.
    b. The term regression originally referred to attempts to predict the extent to which the oceans withdraw from the land in earth history.
    c. All regression lines are constrained to pass through the bivariate or multivariate mean of the dataset.
    d. All of the above.
    e. None of the above.

40. Identify the characteristics of a distance matrix. Choose all options that apply.

    a. The matrix is square.
    b. All the values of the matrix are positive.
    c. The trace (A_{kk}) of the matrix contains only 1s.
    d. All of the above.
    e. None of the above.

41. Identify the characteristics of a covariance matrix. Choose all options that apply.

    a. Number of rows = number of columns
    b. The matrix is asymmetric
    c. The trace (A_{kk}) of the matrix contains only positive values.
    d. All of the above.
    e. None of the above.

42. Identify the statements that are true of scatterplot charts. Choose all options that apply.

    a. Scatterplots are used to visualize geometric and/or distributional relations among two variables.
    b. Continuous or discontinuous variable can be plotted on either axis of a scatterplot.
    c. The fact that the two axes of a scatterplot are always drawn at right angles (90°) to one another can produce a misleading impression of relations between the variables.
    d. All of the above.
    e. None of the above.

43. Identify the purpose of calculating the mean of a population or sample. Choose all options that apply.

    a. As a basis for comparing different populations, samples or groups.

b. As a basis for estimating the variability of a population, sample or group.
c. As a basis for transforming variables whose measurements units force their values into non-comparable magnitudes.
d. In order to locate the position of a population or sample along the number line.
e. None of the above.

44. Identify the statements that are true of both histograms and spindle/density charts. Choose all options that apply.

a. Both are used to display frequency counts over a series of bin intervals created via subdivision of a continuous variable axis.
b. Both are restricted to having the variable axis displayed as the ordinate (horizontal) axis.
c. Both are able to display variation in multiple variables simultaneously.
d. All of the above.
e. None of the above.

45. Identify the statements that are true of dendrogram charts. Choose all options that apply.

a. Dendrograms always reflect accurate estimates of similarity/dissimilarity relations between groups variables and/or samples.
b. Dendrograms are most commonly used to represent non-heirarchical relations among variables and/or samples.
c. It's an easy matter to assess the significance of variable and/or sample clustering patterns in dendrograms.
d. All of the above.
e. None of the above.

46. Identify the statements that are true of a curvilinear regression. Choose all options that apply.

a. A curvilinear regression estimates a non-linear model of variable between the variables.
b. Geometrically, the reason a curvilinear regression line is curved is that its regression model incorporates a series of exponential functions (e.g., $y = x^2, y = x^3$) into regression equation.
c. A curvilinear regression incorporates the same error-minimization model as on ordinary least squares (OLS) regression.
d. All of the above.
e. None of the above.

47. Identify the statements that are true of pie charts. Choose all options that apply.

a. Pie charts can be used to conveniently display relations between any set of data no matter how large or small.
b. Pie charts can be used to display multivariate data.
c. Pie charts take up more space than other types of charts because, whereas the chart itself is. round, it usually requires a square space to be displayed. This results in much empty space (usually).
d. All of the above.
e. None of the above.

48. Identify the statements that are true of ordinary least squares (OLS) bivariate regression analysis. Choose all options that apply.

a. Used to predict the nature of geometric relations between variables.
b. Requires both variables to be measured in the same units.
c. Only capable of modelling linear, straight-line relations between variables.
d. All of the above.
e. None of the above.

49. Identify the statements that are true of the covariance index. Choose all options that apply.

    a. Use of covariance index assumes that you want to use all the information contained in the values of all the variables.
    b. The covariance is a bounded measure of the joint variation between two non-standardized variables.
    c. The covariance index is restricted to telling us whether the trend of the paired data (variable 1 vs variable 2) is positive or negative
    d. All of the above.
    e. None of the above.

50. According to linear classifier performance studies, classifiers performance peaks at what dimensionality? Choose all options that apply.

    a. 2
    b. 4
    c. 5
    d. 10
    e. 20

51. Identify the non-parametric statistical test strategies from the following list. Choose all options that apply.

    a. Bootstrap test
    b. Monte Carlo simulation test
    c. Analysis of variance (ANOVA) test
    d. Permutation test
    e. All of the above

52. Identify the statements that are true of frequency histograms. Choose all options that apply.

    a. The number line of a histogram represent a discontinuous, interval scale.
    b. A histogram's shape is controlled to a large extent by number and size of the bins used to accumulate data.
    c. The abscissa (vertical) axis of a histogram is most often used to display frequency counts.
    d. All of the above.
    e. None of the above.

53. Identify the statements that are true of the residuals test for the appropriateness of a regression model. Choose all options that apply.

    a. The regression residuals are calculated as the squared distance between the observed data points and the sample mean.
    b. If the regression model fits the data a scatterplot of the regression residual distances should be very close to the regression model.
    c. If the regression model fits the data the slope of the regression-residual regression should be identical to the slope of the actual regression.
    d. All of the above.
    e. None of the above.

54. Identify the true characteristics of principal components analysis (PCA). Choose all options that apply.

    a. Able to be used to reduce dataset dimensionality.
    b. Provides a low-dimensional representation of sample similarity structure that enables the identification of subordinate sample clusters.
    c. Facilitates association between PC axes (= eigenvectors) and dominant variables.
    d. All of the above.
    e. None of the above.

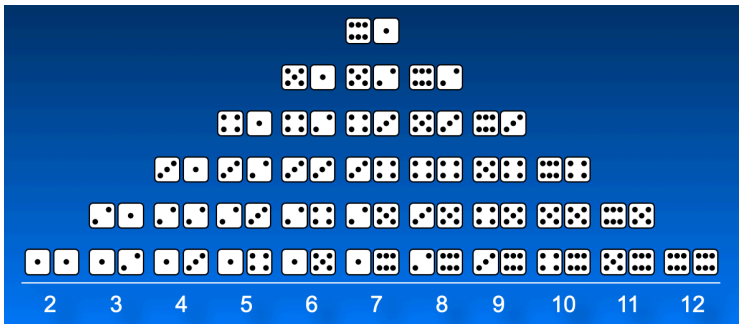
55. Identify the characteristics of a vector Choose all options that apply.

a. Length
b. Width
c. Direction (orientation)
d. All of the above.
e. None of the above.

56. Which of the following constitute reliable guides to the process of statistical testing that will help you avoid error. Choose all options that apply.

a. Always assume your data conform to a normal distribution.
b. So long as a sample can be obtained it may be assumed to be a fair representation of a population of interest.
c. Frequentist, parametric statistical tests are the most powerful statistical tests.
d. All of the above.
e. None of the above.

57. Given the image at right, select the probability of throwing a 5. Choose all options that apply.



a. 0.1667
b. 0.1389
c. 0.3333
d. 0.4444
e. All of the above.

58. Which statements are true of statistical hypothesis testing. Choose all options that apply.

a. Statistical tests are used to evaluate the probability of the null hypothesis (H_0) being true.
b. A p-value of 0.95 means there is a 95% chance the alternative hypothesis is true.
c. Acceptance of the null hypothesis means the observation, or observed value, can be accounted for by a simple appeal to random sampling error.
d. Statistical tests are always reference to the p= 0.95, or 95%, critical value
e. All of the above.

59. Which statements are true of the normal distribution. Choose all options that apply.

a. The expected distribution of observations influenced by many independent factors.
b. Mean = Mode = Median
c. The expected distribution of the means of a series of random samples collected from a single population.
d. All of the above.
e. None of the above.

60. Identify which of the following quantities represent a measure of variability. Choose all options that apply.

a. Analysis of Variance (ANOVA)
b. Coefficient of variation
c. Percentile
d. Standard deviation
e. All of the above

61. Identify the statements that are true of major axis (MA) regression analysis. Choose all options that apply.

a. The slope of this regression model minimizes the squared deviations of observed independent variable values from the linear model.

b. Estimation of the y-intercept for this regression model is accomplished in the same way as for OLS regression.
c. The error-minimization aspect of this regression model is identical to that used in principal component analysis (PCA).
d. All of the above.
e. None of the above.

62. Identify the advantages of displaying your data in the PCA space. Choose all options that apply.

a. The first few principal components summarize more information than any single variable.
b. Low dimensional ordinations in the PCA space will always provide accurate summary of clustering in your dataset.
c. All principal components are oriented orthogonal to one another.
d. The set or principal components are aligned with the directions of maximum independent variation within the pooled dataset.
e. All of the above.

63. The two-sample Students *t*-distribution is used for which of the following. Choose all options that apply.

a. Testing whether the means of two samples were collected from populations with the same mean and standard deviation.
b. Estimating the values of a normal distribution associated with specific probabilities.
c. Estimating the confidence values of a linear regression slope.
d. All of the above.
e. None of the above.

	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7
X1	-0.164	0.422	0.644	-0.092	-0.224	0.415	0.386
X2	-0.141	0.447	-0.714	-0.050	-0.394	0.066	0.330
X3	0.173	0.257	-0.129	0.631	0.606	0.279	0.211
X4	-0.170	0.650	0.147	0.213	-0.035	-0.402	-0.565
X5	-0.546	-0.135	0.105	0.164	0.161	-0.597	0.513
X6	-0.768	-0.133	-0.149	-0.061	0.207	0.465	-0.327
X7	-0.073	-0.313	0.067	0.718	-0.598	0.108	-0.090

64. Given the eigenvector loading matrix (right) identify the five variables whose vectors align most strongly with the second principal component.

a. X4
b. X2
c. X1
d. X7
e. X3

65. Identify the alternative criteria that be used to assign unknown samples to groups following a linear discriminant analysis Choose all options that apply.

a. Projected coordinate position falls within a group's 95% confidence ellipsoid.
b. Minimum linear distance of group centroid
c. Permutation test
d. *n* Nearest neighbor analysis
e. All of the above.

66. Identify the statements that are true of a multiple regression. Choose all options that apply.

a. Whereas a linear regression fits a line (or a vector) to the data, for a five variable problem a multiple regression fits a hyperplane (plane of > three dimensions) to the data.
b. In a multiple regression a set of independent variables are used to predict the value of a single dependent variable.
c. Because of their computational complexity, multiple regression problems can only be solved by matrix methods.
d. All of the above.
e. None of the above.

67. Identify the problem(s) inherent in attempting to analyze high-dimension data. Choose all options that apply.

    a. Need to collect large samples.
    b. Issues concerning the degree to which the sample is representative of the population of interest.
    c. Issues associated with multicolinearity among variables.
    d. All of the above.
    e. None of the above.

68. Identify the statements that are true of ordinary standardized major axis (SMA) regression analysis. Choose all options that apply.

    a. The slope of this regression is estimated as the ratio between the two variable's standard deviations.
    b. Estimation of the y-intercept for this regression model is accomplished in the same way as for OLS regression.
    c. This regression fits a linear model to the data by minimizing the area of a set of right triangles between the observed data and the linear model.
    d. All of the above.
    e. None of the above.

69. Identify the order of the data operations performed during a canonical correlation analysis, Choose all options that apply.

    i. Scaling of the preliminary ordination space by $1/\sqrt{eigenvalues}$.
    ii. Calculation of group centroids in the processed ordination space.
    iii. PCA analysis of pooled-data covariance/correlation matrix.
    iv. Use of processed ordination space eigenvalues to project data into the ordination space.

    a. i - ii - iii - iv
    b. ii - i - iii - iv
    c. iii - i - ii - iv
    d. iii - iv - ii - i
    e. None of the above.

70. Identify the conditions under which it would be appropriate to choose the major axis (MA) model for a regression analysis. Choose all options that apply.

    a. Interested in predicting the values of one variable in terms of the pattern of data point in another variable.
    b. Interested in minimizing variation of the observed data from both variables simultaneously.
    c. Is particularly appropriate for use in the analysis of variables of different types, measured in different units, and/or whose values have grossly different intrinsic magnitudes.
    d. All of the above.
    e. None of the above.

71. Identify the assumptions of a parametric statistical test. Choose all options that apply.

    a. Samples must have been obtained randomly with equiprobable selection potential.
    b. The test of question is either referenced to a known statistical distribution (e.g., *t*-Distribution, *F*-Distribution) or assumed to reference the normal distribution
    c. In the comparison of variables, each is assumed to have equivalent variances.
    d. In the comparison of multiple group, unless the test specifically requires an estimate of the group variances be supplied, they are assumed to be equivalent.
    e. All of the above.

72. Given the two matrices at right, what will be the size of the product matrix ($A.B$) Choose all options that apply.

Matrix A

2	8
8	7
10	6

Matrix B

9	1	7	2	10	2
5	6	3	3	9	5
7	1	3	9	9	9

a. 2 x 3
b. 3 x 3
c. 2 x 6
d. 6 x 6
e. None of the above.

73. In applying the bootstrap procedure to Hotelling's T^2 test for group separation which assumption(s) can be relaxed. Choose all options that apply.

a. Specimens have been selected randomly and equiprobably.
b. Group variance-covariance matrices are equal.
c. All groups exhibited multivariate normal variable distributions.
d. No mis-classifications are present in the group samples.
e. All of the above.

74. Identify the type of data whose central point would be appropriate to estimate using the geometric mean. Choose all options that apply.

a. When variables are non-linear and multiplicative.
b. When variables are ratios.
c. When variables are linear and additive.
d. When variables have the same scale.
e. When variables have different scales

75. Identify the advantage(s) of singular value decomposition (SVD). Choose all options that apply.

a. Generalized, data-driven, dimensionality reduction
b. Supports eigenanalysis of non-square data matrices
c. Efficient for analyzing large datasets
d. All of the above.
e. None of the above.