The data of this exercise is provided in Examdata4.tsv. All numerical values are rounded to 4 decimal places. Which of the following statements is ***not*** correct?

Table 4.2 Critical Points for the Q–Q Plot Correlation Coefficient Test for Normality

Sample size	Significance levels α		
n	.01	.05	.10
5	.8299	.8788	.9032
10	.8801	.9198	.9351
15	.9126	.9389	.9503
20	.9269	.9508	.9604
25	.9410	.9591	.9665
30	.9479	.9652	.9715
35	.9538	.9682	.9740
40	.9599	.9726	.9771
45	.9632	.9749	.9792
50	.9671	.9768	.9809
55	.9695	.9787	.9822
60	.9720	.9801	.9836
75	.9771	.9838	.9866
100	.9822	.9873	.9895
150	.9879	.9913	.9928
200	.9905	.9931	.9942
300	.9935	.9953	.9960

Choose one answer

- The correlation coefficient r_Q for the Q-Q plot of X_2 is 0.9978.
- X_1 is normally distributed with significance level 0.05.
- X_2 is normally distributed with significance level 0.05.
- X_3 is normally distributed with significance level 0.1.

Assume that we have data that can be classified into 2 groups. Three classification methods (Method #1, Method #2 and Method #3) are applied to this data set. Their performance are evaluated according to their AUC (area under the curve) values, i.e.

Methods	AUC
Method #1	0.8
Method #2	0.85
Method #3	0.90

Which method performs the best?

Choose one answer

- Method #3
- Method #1
- Method #2

The data of this exercise is provided in Examdata5a.tsv and Examdata5b.tsv. We want to test the equality of their covariance matrices with significant level 0.05, assume that data are normally distributed. Null hypothesis is that their covariance matrices are equal, what is the correct conclusion based on the Box's M-test.

Choose one answer

- Reject the null hypothesis: their covariance matrices are not the same.
- Accept null hypothesis: their covariance matrices are the same.

Let $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ be a random vector with properties $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{X}) = \begin{bmatrix} 2 & -0.5 & 0.5 \\ -2 & 2 & 0 \\ 3 & 0 & 6 \end{bmatrix}$. Further assume that X_1 is categorical. Which of the following statements is correct?

Choose one answer

- X_2 *and* X_3 are uncorrelated.
- X_2 *and* X_3 are independent.
- The standard deviation of X_1 is 2.
- There is a positive correlation between X_1 *and* X_2.

The data of this exercise is provided in Examdata3.tsv. All numerical values in the following are rounded to 4 decimal places. PCA analysis is done without standardization. Which of the following statements is ***not*** correct?

Choose one answer

- ○ The correlation coefficient between the second principal component and X_3 is negative.
- ○ The first principal component is clearly dominant, since it explains more than 95% of the total variance.
- ○ The first principal component is $0.6831X_1 + 0.5102X_2 + 0.5223X_3$
- ○ The sample mean vector is $(4.7254, 4.4776, 3.7032)'$

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{1000}$ be a random sample from the normal distribution $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Which of the following statements is correct?

Choose one answer

- The distribution of $(\overline{\boldsymbol{X}} - \boldsymbol{\mu})' \cdot \boldsymbol{S}^{-1} \cdot (\overline{\boldsymbol{X}} - \boldsymbol{\mu})$ is approximately χ^2_3
- $\overline{\boldsymbol{X}}$ is distributed as $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $1000 \cdot (\overline{\boldsymbol{X}} - \boldsymbol{\mu})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\overline{\boldsymbol{X}} - \boldsymbol{\mu})$ is distributed as χ^2_4
- The distribution of $(\boldsymbol{X}_1 - \boldsymbol{\mu}) \cdot \boldsymbol{\Sigma} \cdot (\boldsymbol{X}_1 - \boldsymbol{\mu})'$ is χ^2_3.