
No documents are allowed. The use of a calculator is forbidden.

Questions displaying the symbol ♣ may have one or several correct answers. The others have exactly one correct answer. There are 5 questions.

Marking: For multiple-answer questions, a perfect answer is worth 1 point, and ticking every box is worth 0 points (wrong answers are therefore penalised). For single-answer questions, the correct answer is worth 1 point, a wrong one 0.

Answers must be given exclusively at the end of the paper.

Question 1 ♣ A statistical test is carried out and the p-value obtained is 0.024. Which statements are correct?

- ☐ A H_0 is rejected at the 10% type I error level.
- ☐ B H_0 is rejected at the 1% type I error level.
- ☐ C If the test is carried out on another independent sample with the same distribution, the p-value will be identical.
- ☐ D If the sample size is increased and the p-value recomputed, the latter will be smaller than 2.4%.
- ☐ E H_0 is rejected at the 5% type I error level.

Question 2 ♣ Consider a statistical test whose critical region, for a 5% type I error level, is $\{|T| > 2.0\}$. From a sample of size n , one observes $T = 2.4$. Which statement(s) is (are) necessarily true?

- ☐ A At the 5% level, the null hypothesis is rejected.
- ☐ B The observed p-value is smaller than 0.05.
- ☐ C If the sample size is increased, the value of T will decrease.
- ☐ D At the 10% level, the null hypothesis is rejected.
- ☐ E At the 1% level, the null hypothesis is rejected.

Question 3 Two variables are recorded for n individuals: their blood group and their region of origin. Which tool seems the most suitable to study a possible association between these two variables?

- ☐ A A chi-square test
- ☐ B A representation using boxplots
- ☐ C Computing Pearson's correlation coefficient
- ☐ D An analysis of variance test
- ☐ E A scatter plot

Question 4 Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, \dots, n$. What is the formula of the least squares estimator $\hat{\beta}_1$ of the slope β_1 ?

- | | |
|---|---|
| ☐ A $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$ | ☐ C $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$ |
| ☐ B $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}) \sum_{i=1}^n (Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$ | ☐ D $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}) \sum_{i=1}^n (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$ |

Question 5 ♣ In a linear regression model of the form $Y = X\beta + \epsilon$, what do the assumptions $\mathbb{E}(\epsilon) = 0$ and $V(\epsilon) = \sigma^2 I_n$ imply?

- ☐ A The OLS estimator of β , if it exists, is unbiased
- ☐ B ϵ and Y are uncorrelated
- ☐ C The components of the vector ϵ are pairwise uncorrelated.
- ☐ D The empirical mean of ϵ is zero
- ☐ E The vector Y has size n
- ☐ F The matrix $X'X$ is invertible

No documents are allowed. The use of a calculator is forbidden.

Questions displaying the symbol ♣ may have one or several correct answers. The others have exactly one correct answer. There are 5 questions.

Marking: For multiple-answer questions, a perfect answer is worth 1 point, and ticking every box is worth 0 points (wrong answers are therefore penalised). For single-answer questions, the correct answer is worth 1 point, a wrong one 0.

Answers must be given exclusively at the end of the paper.

Question 1 For a two-sided test, let F denote the cumulative distribution function of the statistic T under H_0 . One observes $T = 1.8$. Assume that $F(1.8) = 0.96$. What is the p-value of the test?

- ☐ A 0.04 ☐ B 0.08 ☐ C 0.96 ☐ D 1.8 ☐ E 0.92 ☐ F 0.48

Question 2 ♣ A two-sided test uses a statistic T whose distribution under H_0 is continuous. For a 5% type I error level, the critical region is $\{|T| > 1.96\}$. One observes $T = 1.2$. Which statement(s) is (are) necessarily true?

- ☐ A H_0 is not rejected at the 10% level.
☐ B H_0 is not rejected at the 5% level.
☐ C H_0 is not rejected at the 1% level.
☐ D The p-value is smaller than 0.05.

Question 3 ♣ Two variables are recorded for n high-school graduates: their final average mark and the honours grade they were awarded. The link between these two variables is obvious, since it is deterministic. What would one obtain with the usual statistical tools? Assume that, for each honours grade, at least two individuals with different marks are observed.

- ☐ A Pearson's correlation between the two variables equals 1
☐ B An analysis of variance test necessarily leads to the alternative hypothesis
☐ C The between-group variance is strictly positive
☐ D The chi-square distance equals 0
☐ E The within-group variance is strictly positive

Question 4 What is the ordinary least squares estimator of β ?

- ☐ A $\mathbb{E}((XX')^{-1}XY)$
☐ B $(X'X)^{-1}X'Y$
☐ C $\mathbb{E}((X'X)^{-1}XY)$
☐ D $(XX')^{-1}XY$

Question 5 ♣ In a linear regression model, let X denote the design matrix, of size (n, p) , and β the regression coefficient, of size p . What does the assumption $\text{rk}(X) = p$ imply?

- ☐ A The OLS estimator of β has a closed form.
☐ B The matrix X is invertible.
☐ C The OLS estimator of β is unbiased.
☐ D $p \leq n$.
☐ E The regression coefficient β , of size p , is identifiable.

