2024/07/24 - Part A

Applied Statistics Exam

--> Each question may have one or more correct answers. Indicate **all** correct answers.
--> The response to each of the 10 questions is considered valid if and only if all correct answers are indicated correctly.

The use of R is allowed.
Consulting the R help (using the ? command) and course materials is allowed.

The use of other sources is not allowed, particularly searching for solutions to questions on the internet or communicating with others during the exam.

* Questo modulo registrerà il tuo nome, inserire il nome.

1

What is the meaning of the term "heteroscedasticity"? (1 punto)

- [ ] The errors are not linearly independent
- [ ] The errors have non-zero mean
- [ ] The variance of the dependent variable is not constant and the errors are not linearly independent of one another
- [ ] The variance of the errors is not constant

2

The relationship between number of beers consumed (x) and blood alcohol content (y) measured in grams per deciliter (g/dL) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$\hat{y} = -0.0127 + 0.0180 * x$

The above equation implies that: (1 punto)

- [ ] The slope of 0.0180 suggests that there is no relationship between beer consumption and blood alcohol content
- [ ] The slope of 0.0180 suggests that consuming one beer will result in a blood alcohol content of 0.0180 grams per deciliter (g/dL)
- [ ] For each beer consumed, the predicted blood alcohol content increases by 0.0180 grams per deciliter (g/dL)
- [ ] For each beer consumed, the predicted blood alcohol content decreases by 0.0180 grams per deciliter (g/dL)

3

The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

y^ = -0.0127 + 0.0180 * x

Suppose that the legal limit to drive is a blood alcohol content of 0.08. If Ricky consumed 5 beers the model would predict that he would be:
(1 punto)

- [ ] 0.09 above the legal limit
- [ ] 0.0733 above the legal limit
- [ ] 0.0027 below the legal limit
- [ ] 0.0027 above the legal limit

4

Data was collected on two variables x and y and a OLS regression line was fitted to the data. The fitted line is represented by equation (1) below.
We know that in (x=5, y=6), the residual equals -0.21. Then: (1 punto)

$\hat{y} = -2.29 + \hat{\beta} \cdot x$

- [ ] $\hat{\beta} = 1.658$
- [ ] $\hat{\beta} = 1.50$
- [ ] $\hat{\beta} = 1.90$
- [ ] None of these answers is correct.
- [ ] $\hat{\beta} = 1.70$

5

Suppose (Xn) is a Markov Chain with 3 states and transition probability matrix as inidicated in Figure.

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Indicate the true statements: (1 punto)

- [ ] {Xn} is irreducible.
- [ ] None of these answers is correct.
- [ ] {Xn} has an absorbing state.
- [ ] {Xn} is recurrent.
- [ ] {Xn} does not admit a stationary probability distribution.

6

In relation to Kriging models, indicate the true statement(s): (1 punto)

- [ ] Ordinary Kriging is used if the mean value changes over space
- [ ] Simple Kriging is used when the mean value mu(s) of the process is unknown and constant
- [ ] Ordinary Kriging is used when the mean value mu(s) of the process is unknown and constant
- [ ] Universal Kriging is used if the mean value changes over space

7

What is Cook's distance used for? (1 punto)

- [ ] As an alternative to the p-value of the t-test for assessing the coefficients' significance in OLS models.
- [ ] Assessing the presence of auto-correlation.
- [ ] None of these answers.
- [ ] Identifying influential observations in OLS models.
- [ ] Determining if the overall regression model is significant.

8

The Moran's I statistic: (1 punto)

- [ ] Is an indicator of global spatial autocorrelation.
- [ ] Depends from the total number of observations.
- [ ] Is independent from the spatial weights matrix.
- [ ] Corresponds to the slope of the regression line in the Moran scatter plot.

9

Consider two linear regression models for the same target variable,, M1 and M2, both fitted with OLS on the same dataset and both featuring the intercept along with two continuous regressors which, however, are different in the two models.

The likelihood computed for model M1 is 0.06, while for model M2, it is 0.05. (1 punto)

- [ ] It is not possible to know which model is the best in terms of AIC because some parameters values are missing.
- [ ] In terms of AIC, M1 and M2 are equally good.
- [ ] In terms of AIC, the best model is M1.
- [ ] In terms of AIC, the best model is M2.

10

In what type of data distribution is the difference between the mean and median most informative? (1 punto)

- [ ] Skewed distribution.
- [ ] Uniform distribution.
- [ ] Symmetrical distribution.
- [ ] There is almost never a high difference between the mean and median.