Self Assessment 1

1. The estimators of the linear regression model are derived by:

a. Maximizing the sum of squared differences between observed and expected values of the response variable.

b. Minimizing the sum of absolute differences between observed and expected values of the response variable.

c. Maximizing the sum of absolute differences between observed and expected values of the response variable.

d. Minimizing the sum of squared differences between observed and expected values of the response variable.

**Ans: D**

2. For a sales equation given in terms of marketing expense and channel as:

Sales = 10000 + 5 x marketing expense + 5000 x offline

where the channel has 2 levels: ‘online’, ‘offline’ and reference level is ‘online’. The marketing expense varies from a range of $1,000 USD to $10,000 USD. What will be the sales value in USD for marketing expense = 15,000 USD and channel = ‘offline’

a. 85000 USD

b. 90000 USD

c. 75000 USD

d. Can’t say

**Ans: D** Linear regression should be used only within the range of data given and is not valid for extrapolation

3. Given a linear model, 𝑴𝑷𝑮=𝜷𝟎+ 𝜷𝒘𝑾𝑻+𝜷𝑯𝑯𝑷A screenshot of a cell phone

Description automatically generated

What is the R2 for the given model and whether this model is a good fit for the data:

a. 0.82, Yes

b. 0.82, No

c. 0.18, Yes

d. 0.18, No

**Ans: A** RSS (Regression sum of squares) = 6641.5 + 35.4 = 6676.9

TSS (Total sum of squares) = 6641.5 + 35.4 + 1430.6 = 8107.5

4. Transformation of variable help to eliminate which of the following problems in linear regression?

a. Outliers

b. Correlation of error terms

c. Non-constant variance of error term

d. Collinearity

**Ans: C** Reason: Only feasible option from given choices. It is also used to convert from non-linear to linear patterns.

5. Which of the following plots help us check if residuals have any non-linear patterns?

a. Residual vs. Fitted values plot

b. Q-Q plot

c. Scale-Location plot

d. Residuals vs. Leverage plot

**Ans: A.** A Residual vs. Fitted values plot helps visually detect if the residuals are randomly distributed across fitted values(in case of linear) or organize in a pattern (in case of non-linear)

6. Problem of multicollinearity in a linear model occurs when:

a. Two or more explanatory variables are perfectly correlated with one another

b. The explanatory variables are highly correlated with the error term

c. The explanatory variables are highly correlated with the dependent variable

d. Two or more explanatory variables are highly correlated with one another

**Ans: D** Reason: Multicollinearity refers to explanatory variables being positively correlated among themselves causing problems in model interpretability

7. For the categorical variable, coffee, there are three sizes (small, medium, large). If we create two dummy variables and regress them against the amount spent on a coffee, which scenario will result in the maximum average amount spent on coffee?

a) When we consider the base case as small.

b) When we consider the base case as medium

c) When we consider the base case as Large

d) None of the above

**Ans: D** Explanation: All of them will yield approximately the same average value.

8. Which model has the best 𝑅2 for given a dataset?

a) Log-Log model

b) Log-Linear model

c) Linear-Log model

d) Depends on the dataset

Ans: D

9. How many dummy variables will be created for a categorical variable with M levels?

a) M

b) M-1

c) M+1

d) Depends on the categorical variable

Ans: B

10. Which of the following is not a true statement about the intention behind log transformation?

a) To make the variance of model more constant

b) To make a distribution of the model more log-normal

c) To achieve a more linear model

d) To increase R-Squared of the model

Ans: B, To make a distribution of the model more normal