Regression and Survival Analysis – Quiz

Probability & Statistics

Question 1

In simple linear regression, what does the slope coefficient represent?

A. The predicted value when $X = 0$
B. The change in Y for a one-unit change in X
C. The correlation between X and Y
D. The variance of the residuals

Question 1

In simple linear regression, what does the slope coefficient represent?

A. The predicted value when $X = 0$
B. The change in Y for a one-unit change in X
C. The correlation between X and Y
D. The variance of the residuals

Answer: B
The slope coefficient represents the expected change in Y for each one-unit increase in X, holding other variables constant.

Question 2

What does R-squared measure?

A. The slope of the regression line
B. The proportion of variance in Y explained by the model
C. The significance of the coefficients
D. The number of predictors in the model

Question 2

What does R-squared measure?

A. The slope of the regression line
B. The proportion of variance in Y explained by the model
C. The significance of the coefficients
D. The number of predictors in the model

Answer: B
R-squared (coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables.

Question 3

In logistic regression, an odds ratio of 2.5 means:

A. The probability increases by 2.5
B. The odds of the outcome are 2.5 times higher for a one-unit increase in X
C. The log-odds increase by 2.5
D. The outcome is 2.5 times more likely than not

Question 3

In logistic regression, an odds ratio of 2.5 means:

A. The probability increases by 2.5
B. The odds of the outcome are 2.5 times higher for a one-unit increase in X
C. The log-odds increase by 2.5
D. The outcome is 2.5 times more likely than not

Answer: B
An odds ratio of 2.5 indicates that for each one-unit increase in the predictor, the odds of the outcome occurring are 2.5 times higher.

Question 4

What is censoring in survival analysis?

A. Removing outliers from the dataset
B. When the event of interest is not observed during the study period
C. Adjusting for confounding variables
D. Transforming the outcome variable

Question 4

What is censoring in survival analysis?

A. Removing outliers from the dataset
B. When the event of interest is not observed during the study period
C. Adjusting for confounding variables
D. Transforming the outcome variable

Answer: B
Censoring occurs when we have incomplete information about survival time - the subject was alive at last follow-up but we don't know when (or if) the event occurred.

Question 5

The Kaplan-Meier estimator is used to:

A. Calculate regression coefficients
B. Estimate the survival function from observed survival times
C. Test for normality of residuals
D. Compare means between groups

Question 5

The Kaplan-Meier estimator is used to:

A. Calculate regression coefficients
B. Estimate the survival function from observed survival times
C. Test for normality of residuals
D. Compare means between groups

Answer: B
The Kaplan-Meier method is a non-parametric estimator of the survival function that accounts for censored observations.

Question 6

In the Cox proportional hazards model, a hazard ratio of 0.5 indicates:

A. The hazard is 50% lower compared to the reference group
B. The hazard is 50% higher compared to the reference group
C. The survival probability is 0.5
D. The median survival time is 0.5 years

Question 6

In the Cox proportional hazards model, a hazard ratio of 0.5 indicates:

A. The hazard is 50% lower compared to the reference group
B. The hazard is 50% higher compared to the reference group
C. The survival probability is 0.5
D. The median survival time is 0.5 years

Answer: A
A hazard ratio of 0.5 means the hazard (instantaneous risk) is 50% lower in the treatment group compared to the reference group.

Question 7

Multicollinearity in regression causes problems because:

A. It makes the model run slower
B. It inflates standard errors and makes coefficients unstable
C. It always reduces R-squared
D. It violates the assumption of independence

Question 7

Multicollinearity in regression causes problems because:

A. It makes the model run slower
B. It inflates standard errors and makes coefficients unstable
C. It always reduces R-squared
D. It violates the assumption of independence

Answer: B
When predictors are highly correlated, it becomes difficult to separate their individual effects. This inflates standard errors and makes coefficient estimates unstable.

Question 8

AUC (Area Under the ROC Curve) measures:

A. The percentage of correct predictions
B. The model's ability to discriminate between classes
C. The variance explained by the model
D. The number of false positives

Question 8

AUC (Area Under the ROC Curve) measures:

A. The percentage of correct predictions
B. The model's ability to discriminate between classes
C. The variance explained by the model
D. The number of false positives

Answer: B
AUC measures the model's ability to distinguish between positive and negative classes. An AUC of 1.0 means perfect discrimination; 0.5 means no better than random.

Question 9

What assumption is unique to Cox proportional hazards regression?

A. Normality of residuals
B. Homoscedasticity
C. Proportional hazards over time
D. Independence of observations

Question 9

What assumption is unique to Cox proportional hazards regression?

A. Normality of residuals
B. Homoscedasticity
C. Proportional hazards over time
D. Independence of observations

Answer: C
The proportional hazards assumption states that the hazard ratio between groups remains constant over time. This can be tested using Schoenfeld residuals.

Question 10

When is logistic regression preferred over linear regression?

A. When the outcome is continuous
B. When the outcome is binary (0/1)
C. When predictors are categorical
D. When sample size is small

Question 10

When is logistic regression preferred over linear regression?

A. When the outcome is continuous
B. When the outcome is binary (0/1)
C. When predictors are categorical
D. When sample size is small

Answer: B
Logistic regression is used when the outcome is binary (e.g., success/failure, yes/no). Linear regression would predict values outside [0,1] which are invalid probabilities.