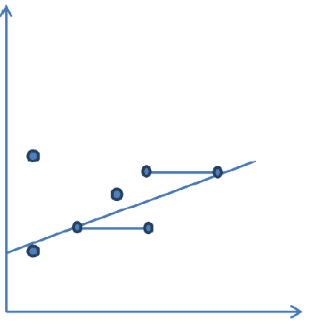
**Subject Code & Semester: AML 5201, Even Semester 2025**

**Subject Name: Advanced Applications of Probability & Statistics**

**Midterm Practice Questions**

1. In which of the following output and input variable combinations can a simple linear regression model not be built?
2. Output: blood pressure, input: weight
3. Output: sales, input: advertisement budget
4. Output: education level, input: income
5. Output: income, input: education level

Correct Answer: c

1. Which one of the following plots shows the correct residual magnitude in simple linear regression based on ordinary least squares estimation?
2. 
3. A diagram of a graph

   AI-generated content may be incorrect.
4. A diagram of a line graph

   AI-generated content may be incorrect.
5. A diagram of a line graph

   AI-generated content may be incorrect.

Correct Answer: b

1. Suppose we have the following simple linear regression model for estimating sales (in 1000s of units) using advertisement budget (in 10000s of Rupees): 

An increase of Rupees 10000 in the advertisement budget will result in an approximate increase in sales by

1. 0.005 units
2. 50 units
3. 5000 units
4. 5 units

Correct Answer: d

1. In a simple linear regression model, a categorical predictor with 5 levels will result in \_\_ dummy variables.
2. 4
3. 5
4. 6
5. 0

Correct Answer: a

1. Suppose that in a simple linear regression model, we have a categorical predictor called *heating* that has 3 levels: (1) electric (2) gas (3) hotair. This results in the creation of 2 dummy variables *heatinggas* and *heatinghotair*. What will be the values of the 2 dummy variables for a house that is electric heated?
2. *heatinggas* = 1*, heatinghotair* = 1
3. *heatinggas* = 1*, heatinghotair* = 0
4. *heatinggas* = 0*, heatinghotair* = 1
5. *heatinggas* = 0*, heatinghotair* = 0

Correct Answer: d

1. For which one of the following quantities associated with a simple linear regression model can a confidence interval be calculated?
2. 
3. 
4. 
5. RSS

Correct Answer: c

1. Suppose we have a simple linear regression model for estimating sales (in 100s of units) using advertisement budget (in 10000s of Rupees) that results in the coefficient estimate  with a 95% confidence interval [0.04104, 0.06106]. We can be 95% confident that an increase of Rupees 10000 in the advertisement budget will result in an approximate increase in sales between
2. 41 to 61 units
3. 4 to 6 units
4. 410 to 610 units
5. 4104 to 6106 units

Correct Answer: b

1. Suppose we build a linear regression model for predicting house price based on square feet area. An R2 value of 60% means that the remaining 40% of the variance in the house price is due to
2. Noise
3. Other predictors missing from the model
4. Nonlinearity
5. All the above

Correct Answer: d

1. Suppose we have the following log-transformed linear regression model for predicting blood pressure BP (in mm Hg) using height (in inches):



This model predicts that an increase of 1 inch in height will result in approximately

1. a relative decrease in BP by 2.5%
2. a relative increase in BP by 25%
3. a relative increase in BP by 2.5%
4. an increase in BP by 0.025 mm Hg

Correct Answer: c

1. An MLRM for predicting house price (in Dollars) as a function of living area (in square feet) and age (in years) results in 

For the same living area, if a house is a year older, then the predicted change in that house’s price is

1. An increase of 224.751 Dollars
2. A decrease of 111.277 Dollars
3. An increase of 111.277 Dollars
4. A decrease of 224.751 Dollars

Correct Answer: d

1. Suppose we used the lm() function in R for predicting house price (in Dollars) as a function of living area (in square feet) and whether the property is a new construction or not with feature name new taking values yes or no. Which of the following is a possible result in R:
2. 
3. 
4. 
5. 

Correct Answer: a

1. An MLRM for predicting house price (in Dollars) as a function of living area (in square feet) and whether the property is a centrally air-conditioned or not results in 

For the same living area, a centrally air-conditioned house is \_\_\_\_ expensive as one which is not centrally air-conditioned.

1. less
2. more
3. equally
4. not

Correct Answer: b

1. An MLRM for predicting house price (in Dollars) as a function of living area (in square feet) and whether the property is a centrally air-conditioned or not results in the following:

A close-up of a computer error

AI-generated content may be incorrect.

Suppose for a centrally air-conditioned house, we increase the living area by 1 square feet. With 95% confidence, we can expect the house price to increase by an amount between

1. 106.76-3x2.75 and 106.76-3x2.75 Dollars
2. 106.76-1.96x38.822 and 106.76-1.96x38.822 Dollars
3. 106.76-1.96x2.75 and 106.76+1.96x2.75 Dollars
4. 106.76-1.96x4.75 and 106.76-1.96x4.75 Dollars

Correct Answer: c

1. Which of the following is not a related metric/test for assessing the accuracy of the coefficient estimates of an MLRM:
2. Hypothesis test
3. R2 value
4. Standard error
5. Confidence intervals

Correct Answer: b

1. Following is the result of fitting an MLRM for predicting an athlete’s heptathlon score based on their performance in 200m, 800m, high jump, and javelin events: A white background with black numbers

   AI-generated content may be incorrect.

Which event is most likely not useful for predicting an athlete’s heptathlon score?

1. 200m
2. 800m
3. High jump
4. Javelin

Correct Answer: d

1. Which of the following is not a possible value of the correlation coefficient?
2. -0.9
3. 0
4. 0.5
5. 1.5

Correct Answer: h

1. In Mumbai, temperature is measured in degrees Centigrade and NO2 concentration in ppb (parts per billion), whereas in Boston, temperature is measured in degrees Fahrenheit and NO2 concentration in ppm (parts per million). Both cities report their respective correlation between temperature and NO2 concentration as 0.6. Which of the following is true:
2. Boston really has the higher correlation, because Fahrenheit temperatures are higher than Centigrade
3. Mumbai really has the higher correlation, because ppb values are bigger than ppm values
4. Both cities have the same correlation, because correlation is independent of the units of measurement
5. We do not know which city has the higher correlation

Correct Answer: c

1. Which of the following is not necessarily true in a linear regression model built with correlated predictors?
2. Coefficient estimates will change dramatically depending on which correlated predictors are included or not in building the model
3. Coefficient estimates for predictors with known strong relationships with the response will not be accurate
4. The R2 value of the model will be close to zero
5. Standard errors of the coefficient estimates will be (relatively) large

Correct Answer: a

1. The following VIF values inside the parentheses indicate which feature is most likely to be collinear with respect to the rest?

Weight (13.6), Abdo (4.8), Neck (2.6), Thigh (4.4), Hip (10.7)

1. Weight
2. Hip
3. Neck
4. Thigh

Correct Answer: a

1. Suppose we are interested in predicting the average credit card balance of an individual using income (in thousands of dollars) and another predictor “student” indicating whether the individual is a student or not. An interaction-based MLRM results in the following least squares lines:

A graph of income and balance

AI-generated content may be incorrect.

Which of the following statement is true?

1. The slope for students is greater than the slope for non-students
2. Students typically tend to have a lower average credit card balance than non-students
3. For a thousand dollar increase in salary, credit card balance tends to increase more for non-students than for students
4. The intercept for students is smaller than the intercept for non-students

Correct Answer: c

1. Suppose we want to predict starting salary after graduation (in thousands of dollars) using the following predictors:

(1) GPA (2) IQ (3) Gender (female and male) (4) Interaction between GPA and IQ (5) Interaction between GPA and Gender.

The results of fitting a linear regression model are:



The predicted salary (in thousands of dollars) of a female with IQ 110 and GPA 4 is

1. 142.1
2. 1.421
3. 142100
4. 182.1

Correct Answer: a

1. Suppose we want to predict house price (in dollars) using living area (in square feet) and type of heating (categorical with 3 levels electric, hot air, and hot water/steam) as predictors. An MLRM with interaction has the following coefficient estimates:

A white background with black text

AI-generated content may be incorrect.

Which one of the following will happen for a one square foot increase in living area for a hot air-heated house?

1. Increase the house price by approximately 94.916 dollars
2. Increase the house price by approximately 121.171 dollars
3. Decrease the house price by approximately 14.331 dollars
4. Decrease the house price by approximately 80.585 dollars

Correct Answer: b

1. Suppose we want to model credit-score using income, gender (categorical with 2 levels), and education level (categorical with 4 levels). An MLRM with interaction between gender and education level will result in \_\_ coefficient estimates.
2. 11
3. 8
4. 10
5. 9

Correct Answer: d