Take Test: Quiz 2

Test Information

Description

Instructions

Multiple Attempts This test allows 3 attempts. This is attempt number 1.

Force Completion This test can be saved and resumed later.

Question Completion Status:

QUESTION 1

1 points Saved

Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Gender (1 for Female and 0 for Male), X4 = Interaction between GPA and IQ (i.e., GPA*IQ), and X5 = Interaction between GPA and Gender (i.e., GPA*Gender). The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and the intercept b0 and the five coefficients (associated with X1 to X5) are

- b0 = 50; b1 = 20; b2 = 0.07;
- b3 = 35; b4 = 0.01; b5= -10.

Predict the salary of a female with IQ of 100 and a GPA of 3. Provide your prediction using the same unit as the response variable, salary in thousands of dollar, that is, if the actual salary is $10,000, you should input it as 10.

125

QUESTION 2

1 points Saved

Continue with the GPA and Salary Data. Which of the following are correct.

- [ ] 1. For a fixed valueof IQ and GPA, males earn more on average than females provided that their GPA < 3.5.
- [ ] 2. For a fixed valueof IQ and GPA, females earn less on average than males.
- [ ] 3. For a fixed valueof IQ and GPA, females earn more on average than males.
- [ ] 4. For a fixed valueof IQ and GPA, females earn more on average than males provided that their GPA > 3.5.
- [x] 5. For a fixed valueof IQ and GPA, females earn more on average than males provided that their GPA < 3.5.
- [x] 6. For a fixed valueof IQ and GPA, males earn more on average than females provided that their GPA > 3.5.

QUESTION 3

2 points Saved

Continue with the GPA and Salary Data. If we change the unit of IQ to 100 (i.e., IQ score 100 is recorded as IQ=1, and IQ score 150 is recorded as 1.5), then

- The new LS estimate for the intercept b0 is equal to 50 ;
- The new LS estimate for the intercept b2 is equal to 7 ;
- The new LS estimate for the intercept b3 is equal to 35 ;
- The new LS estimate for the intercept b4 is equal to 1 ;

QUESTION 4

2 points Saved

Continue with the GPA and Salary Data. Based on the LS coefficients, given IQ and GPA, the prediction for the salary of a **female** student can be written as the following linear combination

w0 + w1*GPA + w2*IQ + w3*(GPA*IQ),

where w0= 85 , w1= 10 , w2= 0.07 and w3= 0.01 .

Note: the IQ is still in the original scale as in Problem 1, i.e., IQ score 100 is recorded as 100 and score 150 is recorded as 150.

QUESTION 5

2 points Saved

Continue with the GPA and Salary Data. Based on the LS coefficients, given IQ and GPA, the prediction for the salary of a **male** student can be written as the following linear combination

w0 + w1*GPA + w2*IQ + w3*(GPA*IQ),

where w0= 50 , w1= 20 , w2= 0.07 and w3= 0.01 .

Note: the IQ is still in the original scale as in Problem 1, i.e., IQ score 100 is recorded as 100 and score 150 is recorded as 150.

QUESTION 6

3 points Saved

Continue with the GPA and Salary Data. Suppose we code X3 = Gender as 1 for Male and 0 for Female. Let a0 denote the intercept and the a1 to a5 to denote the five coefficients associated with X1 to X5, respectively. Then

- a0 = 85 ; a1 = 10 ; a2 = 0.07 ;
- a3 = -35 ; a4 = 0.01 ; a5 = 10 .

Note: IQ is in the original scale as in Problem 1, i.e., IQ score 100 is recorded as X2=100 and score 150 is recorded as X2=150.

QUESTION 7

1 points Saved

Continue with the GPA and Salary Data. What does it mean if the LS coefficient for X1=GPA is significant?

- [ ] 1. The absolute value of the correlation between Y and X1 is large
- [x] 2. The conditional correlation between Y and X1, given other four predictors, is not zero
- [ ] 3. The correlation between Y and X1 is not zero
- [x] 4. The absolute value of the conditional correlation between Y and X1, given other four predictors, is large

QUESTION 8

1 points Saved

Continue with the GPA and Salary Data. Suppose we add one more predictor to the model. Which of the following statements are true?

- [x] 1. R-square will increase
- [ ] 2. Sigma-hat will be smaller
- [x] 3. RSS will decrease
- [ ] 4. Adjusted R-square will increase

QUESTION 9

1 points Saved

Suppose you want to fit a multiple linear regression model on a data set with n=100 observations. You mistakenly duplicate the data set, that is, you end up fitting a multiple linear regression model on n=200 observations (each of the 100 original samples appear twice). Which of the following estimates stay unchanged?

- [ ] 1. Sigma-hat
- [ ] 2. RSS
- [x] 3. R-square
- [x] 4. LS coefficients

QUESTION 10

1 points Saved

Suppose we are fitting a linear regression model with n samples and p predictors (including the intercept). Circle correct statements.

- [x] 1. The LS coefficients are not unique if a non-intercept predictor takes a constant value
- [ ] 2. The LS coefficients are not unique if p is bigger than or equal to n.
- [ ] 3. Tne LS coefficients are not unique if the response variable Y takes a constant value
- [ ] 4. The LS coefficients are unique only if the predictors are uncorrelated with each other.
- [x] 5. The LS coefficients are not unique if columns of the n-by-p design matrix X are not linear independent.

QUESTION 11

4 points Saved

The following questions are related to a subset of the Iowa Housing Data, data for last semester's project 1.

- Load the training data (train_housing_quiz2.csv). The data has 1460 rows and 81 columns.
- The last column "SalePrice" records the sale price for each house. The cheapest price is 35 thousand dollars (**round your answer to the nearest integer**, e.g., if the cheapest house was sold for 2,400 dollars, your answer would be ``2"), and the most expensive one is sold for 755 thousand dollars. The median housing price is 163 thousand dollars. 728 houses are sold for more than half-million dollars.
- The column "HouseStyle" records the house style, which is a categorical variable taking 8 different values.
- The column "GarageCars" records size of garage in car capacity for each house. 5 houses have more than 3 car garages.

QUESTION 12

1 points Saved

Continue with the Iowa Housing Training Data. We will fit a linear regression model with five predictors that are highly correlated with the housing price: OverallQual, X1stFlrSF, GrLivArea, GarageCars, GarageArea.

- Since the evaluation is based on log of the sale price, we change SalePrice to log-scale.
- Predictors that measure amounts (such as price, size, volumn, etc) are usually right-skewed: its density plot has a long right tail, i.e., there are some very large values. Suggest to apply square-root or log-transformation on those predictors. For this dataset, we'll change X1stFlrSF, GrLivArea, and GarageArea to log-scale.
- Fit a linear regression model to predict the log(SalePrice+1) using predictors: OverallQual, log(X1stFlrSF+1), log(GrLivArea+1), GarageCars, log(GarageArea+1). Treat all predictors as numerical variables.
- **The R-square of this model is equal to** 0.80 **(round to the 2nd digit after the decimal point, e.g., 0.56.).**
- The median OverallQual is 6, the median X1stFlrSF is 1087, the median GrLivArea is 1464, the median GarageCars is 2, and the median GarageArea is 480. Use the fitted model to predict the SalePrice for a house with the above median features. The prediction is equal to 1692 hundred dollars (round to the nearest hundred, i.e., the number you fill in the blank is an integer.)

QUESTION 13

1.5 points Saved

Continue with the multiple linear regression model that predicts log(SalePrice+1) using predictors: OverallQual, log(X1stFlrSF+1), log(GrLivArea+1), GarageCars, log(GarageArea+1). Now, refit the model treating GarageCars as a categorical variable with 5 levels.

- Is the R-square of this new model guaranteed to be bigger than the old linear model that treats GarageCars as a numerical variable? No (fill in the blank with "**Yes**" or "**No**".)
- The new model implies 5 different linear regression models for houses with GarageCars = 0, 1, 2, 3, 4:
  log(SalePrice+1) = b0 + b1*OverallQual + b2*log(X1stFlrSF+1) + b3*log(GrLivArea+1) + b4*log(GarageArea+1)

  In particular, b0 = 7.1 for GarageCars = 1 (**Round to the first digit after the decimal point**)
- Use F-test to compare the two models, the old one that treats GarageCars as a numerical variable and the new one that treats GarageCars as a categorical variable. What the p-value of this test? 0.02 (**Round to the 2nd digit after the decimal point**, e.g., 0.07.)

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers | Close Window | Save and Submit