Week 2 - AYU - Individual

Son Nguyen

2023-04-05

### Type 1: Parameter Estimators

**Problem 1** (Course 120 Study Note 120-82-94 Question 11)

An automobile insurance company wants to use gender (, if female and ) and traffic penalty point () to predict the number of claims (). The observed values of these variables for a sample of six motorists are given by:

| Motorist | 1 | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | 0 | 1 | 1 | 1 |
|  | 0 | 1 | 2 | 0 | 1 | 2 |
|  | 1 | 0 | 2 | 1 | 3 | 5 |

You are using the following model:

You have determine

Determine .

1. -0.25
2. 0.25
3. 1.25
4. 2.00
5. 4.25

**Problem 2** (Course 120 Study Note 120-10-90 Question 19)

You are performing a multiple regression of the form:

You have obtain the following data

|  |  |  |
| --- | --- | --- |
| 1 | -1 | -1 |
| 2 | 1 | -1 |
| 3 | -1 | 1 |
| 4 | 1 | 1 |

Determine

1. 3.5
2. 4.0
3. 4.5
4. 5.0
5. 5.5

### Type 2: Inference (Testing, prediction intervals…)

#### F-test, and reduced models

**Problem 3** (CAS Exam ST Spring 2016 Question 22)

The following two models were fit to 18 observations:

Model 1:   
Model 2:

The result of the regression are:

| Model Number | Error Sum of Squares | Regression Sum of Squares |
| --- | --- | --- |
| 1 | 102 | 23 |
| 2 | 78 | 47 |

Calculate the value of the F-statistics used to test the hypothesis that

1. Less than 1.30
2. At least 1.30, but less than 1.40
3. At least 1.40, but less than 1.50
4. At least 1.50, but less than 1.60
5. At least 1.60

**Problem 4** (SOA Course)

A professor ran an experiment in three sections of a psychology courses to show that the more digits in a number, the more difficult it is to remember. The following variables were used in a multiple regression:

$x\_1 = $ number of digits in the number if student was in section 1, 0 otherwise if student was in section 2, 0 otherwise $y = $ percentage of students correctly remembering the number

You are given

* A total of 42 students participated in the study
* The regression equation
* was to fit the data and resulted in
* A second regression equation was to fit to the data and resulted

Determine the F statistic used to test whether class section is significant variable.

1. 5.4
2. 7.3
3. 7.7
4. 7.9
5. 8.3

**Problem 5** (SOA Course)

You are determining the relationship of salary () to experience () for both men () and woman (). You fit the model

You are given

| Sources | Sum of Squares | Degree of Freedom |
| --- | --- | --- |
| Regression | 330.0117 | 3 |
| Error | 12.8156 | 7 |

You also fit the below model to the same observations.

You are given

| Sources | Sum of Squares | Degree of Freedom |
| --- | --- | --- |
| Regression | 315.0992 | 1 |
| Error | 27.7281 | 9 |

Determine the F-ratio to test whether the linear rellationship between salary and experience is identical for men and women.

1. 0.6
2. 2.0
3. 3.5
4. 4.1
5. 6.2

**Problem 6** (SRM - Sample Question 19)

The regression model is being investigated. The following maximized log-likelihoods are obtained:

* Using only the intercept term: –1126.91
* Using only the intercept term, , and : –1122.41
* Using all four terms: –1121.91

The null hypothesis is being tested at the 5% significance level using the likelihood ratio test.

Determine which of the following is true.

1. The test statistic is equal to 1 and the hypothesis cannot be rejected.
2. The test statistic is equal to 9 and the hypothesis cannot be rejected
3. The test statistic is equal to 10 and the hypothesis cannot be rejected.
4. The test statistic is equal to 9 and the hypothesis should be rejected.
5. The test statistic is equal to 10 and the hypothesis should be rejected.

**Problem 7** (SRM - Sample Question 24)

Sarah performs a regression of the return on a mutual fund () on four predictors plus an intercept. She uses monthly returns over 105 months. Her software calculates the statistic for the regression as , but then it quits working before it calculates the value of . While she waits on hold with the help desk, she tries to calculate from the F-statistic.

Determine which of the following statements about the attempted calculation is true.

1. There is insufficient information, but it could be calculated if she had the value of the residual sum of squares (RSS).
2. There is insufficient information, but it could be calculated if she had the value of the total sum of squares (TSS) and RSS.

**Problem 8** (SRM - Sample Question 27)

Trevor is modeling monthly incurred dental claims. Trevor has 48 monthly claims observations and three potential predictors: - Number of weekdays in the month - Number of weekend days in the month - Average number of insured members during the month

Trevor obtained the following results from a linear regression:

|  | Coefficient | Standard Error | Stat | p-value |
| --- | --- | --- | --- | --- |
| Intercept | –45,765,767.76 | 20,441,816.55 | –2.24 | 0.0303 |
| Number of weekdays | 513,280.76 | 233,143.23 | 2.20 | 0.0330 |
| Number of weekend days | 280,148.46 | 483,001.55 | 0.58 | 0.5649 |
| Average number of members | 38.64 | 6.42 | 6.01 | 0.0000 |

Determine which of the following variables should be dropped, using a 5% significance level.

I. Intercept  
II. Number of weekdays  
III. Number of weekend days  
IV. Number of members.

1. I only
2. II only
3. III only
4. IV only
5. None should be dropped from the model

**Problem 9** (SRM - Sample Question 45)

Trish runs a regression on a data set of n observations. She then calculates a 95% confidence interval on y for a given set of predictors. She also calculates a 95% prediction interval on y for the same set of predictors. Determine which of the following must be true.

I.   
II.   
III.

1. None
2. I and II only
3. I and III only
4. II and III only
5. The correct answer is not given by (A), (B), (C), or (D).

### Type 3: Using MLM Models

**Problem 10** (SRM - Sample Question 49)

The actuarial student committee of a large firm has collected data on exam scores. A generalized linear model where the target is the exam score on a 0-10 scale is constructed using a log link, resulting in the following estimated coefficients

| Predictor Variables | Coefficient |
| --- | --- |
| Intercept | -0.1 |
| Study Time (in units of 100 hours) | 0.5 |
| Attempt (1 for first attempt, else 0) | 0.5 |
| Master’s degree (1 for Yes, 0 for No) | –0.1 |
| Interaction of Attempt and Master’s degree | 0.2 |

The company is about to offer a job to an applicant who has a Master’s degree and for whom the exam would be a first attempt. It would like to offer half of the study time that will result in an expected exam score of 6.0. Calculate the amount of study time that the company should offer. (A) 123 hours (B) 126 hours (C) 129 hours (D) 132 hours (E) 135 hours

### Type 4: Collinearity

**Problem 11** (CAS Exam)

You are given the following information related to the following model

( is linearly regressed on )

| Sources | Degree of Freedom | Sum of Squares |
| --- | --- | --- |
| Regression | 2 | 3.73 |
| Error | 503 | 3.04 |

| Sources | Degree of Freedom | Sum of Squares |
| --- | --- | --- |
| Regression | 2 | 24.74 |
| Error | 503 | 226.34 |

| Sources | Degree of Freedom | Sum of Squares |
| --- | --- | --- |
| Regression | 2 | 214,258 |
| Error | 503 | 85,883 |

Calculate the variance inflation factor (VIF) for the variable which exhibits the greatest collinearity in the original model.

1. Less than 1.0
2. At least 1.0, but less than 2.0
3. At least 2.0, but less than 3.0
4. At least 3.0, but less than 4.0
5. At least 4.0

**Problem 12** (CAS Exam)

A least squares regression is fit to a dataset. The results are as follows.

|  | Fitted | Standard Error |
| --- | --- | --- |
|  | 1.00 | 0.05 |
|  | 2.50 | 0.15 |
|  | -0.50 | 0.20 |
|  | 2.00 | 0.80 |

A second model is fit to the same data using only the third independent variable.

|  | Fitted | Standard Error |
| --- | --- | --- |
|  | 1.25 | 0.08 |
|  | 2.20 | 0.75 |

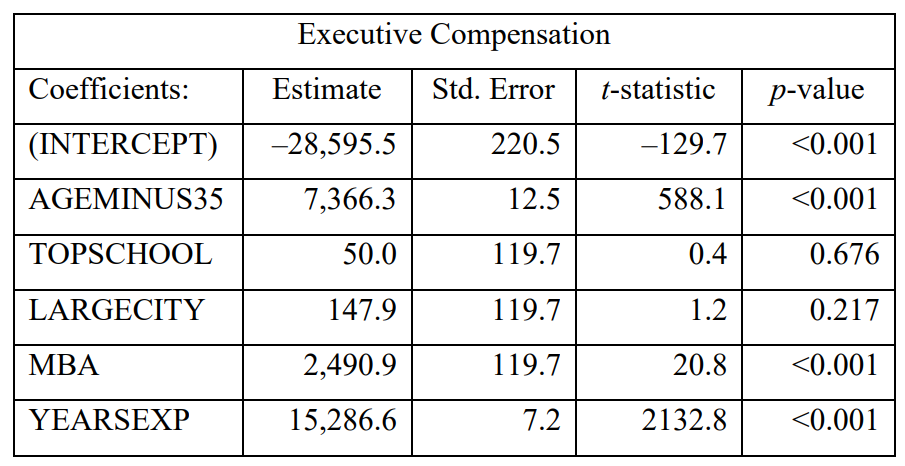
Calculate the VIF of .

1. Less than 1.0
2. At least 1.0, but less than 1.05
3. At least 1.05, but less than 1.10
4. At least 1.10, but less than 1.15
5. At least 1.15

### Type 5: Variable selection, AIC, BIC

**Problem 13** (SRM - Sample Question 54) For a regression model of executive compensation, you are given:

1. The following statistics:



1. The acceptable significance level is .

Determine which variable or variables should be removed first prior to rerunning the model.

1. (INTERCEPT)
2. AGEMINUS35, MBA, and YEARSEXP
3. TOPSCHOOL
4. TOPSCHOOL and LARGECITY
5. YEARSEXP

**Problem 14** (SRM - Sample Question 61)

A linear model has been fit to a dataset containing six predictor variables, F, G, H, I, J, and K. Determine which of the following statements regarding using Akaike information criterion (AIC) or Bayesian information criterion (BIC) to select an optimal set of predictor variables for this linear model is/are true.

I. AIC and BIC provide a direct estimate of the test error.  
II. When choosing between the subsets {F, G, H} and {I, J, K}, AIC and BIC will always select the same subset.  
III. For large sample sizes (n > 7), the number of variables selected by BIC will be less than or equal to the number selected by AIC.

1. None
2. I and II only
3. I and III only
4. II and III only
5. The correct answer is not given by (A), (B), (C), or (D).

**Problem 15** (SOA Course Note) You apply all possible regression model to a set of five observations with three explanatory variables. You have the follows.

| Model | Variables | RSS |
| --- | --- | --- |
| I | None | 10.24 |
| II |  | 5.85 |
| III |  | 8.45 |
| IV |  | 6.15 |
| V |  | 5.12 |
| VI |  | 4.35 |
| VII |  | 1.72 |
| VIII |  | 1.07 |

Using Adjusted determine the best model.

1. Model I
2. Model II
3. Model IV
4. Model VII
5. Model VIII