**PREDICT 410-DL\_SEC58**

**Assignment 4: Statistical Inference in Linear Regression**

**Model 1:** Let’s consider the following SAS output for a regression model which we will refer to as Model 1.

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 4 | 2126.00904 | 531.50226 |  | <.0001 |
| **Error** | 67 | 630.35953 | 9.40835 |  |  |
| **Corrected Total** | 71 | 2756.36857 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 3.06730 | **R-Square** |  |
| **Dependent Mean** | 37.26901 | **Adj R-Sq** |  |
| **Coeff Var** | 8.23017 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 11.33027 | 1.99409 | 5.68 | <.0001 |
| **X1** | 1 | 2.18604 | 0.41043 |  | <.0001 |
| **X2** | 1 | 8.27430 | 2.33906 | 3.54 | 0.0007 |
| **X3** | 1 | 0.49182 | 0.26473 | 1.86 | 0.0676 |
| **X4** | 1 | -0.49356 | 2.29431 | -0.22 | 0.8303 |

| **Number in Model** | **C(p)** | **R-Square** | **AIC** | **BIC** | **Variables in Model** |
| --- | --- | --- | --- | --- | --- |
| **4** | 5.0000 | 0.7713 | 166.2129 | 168.9481 | X1 X2 X3 X4 |

1. (5 points) How many observations are in the sample data?

There are 72 observations in the sample data.

The reported 'Corrected Total' degrees of freedom is 71, which is equal to N – 1, where N is the number of observations.

1. (5 points) Write out the null and alternate hypotheses for the t-test for Beta1.

The null hypothesis is that is equal to zero, with the alternate hypothesis that is not equal to zero.

1. (5 points) Compute the t- statistic for Beta1.

where is the estimated coefficient, and is the standard error of that estimate

therefore,

1. (5 points) Compute the R-Squared value for Model 1.

where is the sum of squares for regression, is the sum of squares for residuals and is the total sum of squares

therefore,

1. (5 points) Compute the Adjusted R-Squared value for Model 1.

where,

is the number of observations

is the total number of parameters in the model

is the number of model parameters in the model excluding any intercept term

is the sum of squares for residuals

is the total sum of squares

therefore,

1. (5 points) Write out the null and alternate hypotheses for the Overall F-test.

The null hypothesis is that all non-constant coefficients are equal to zero, with the alternate hypothesis that at least one of the non-constant coefficients are non-zero.

1. (5 points) Compute the F-statistic for the Overall F-test.

where,

is the number of observations

is the total number of parameters in the model

is the number of model parameters in the model excluding any intercept term

is the sum of squares for regression

is the sum of squares for residuals

therefore,

**Model 2:** Now let’s consider the following SAS output for an alternate regression model which we will refer to as Model 2.

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 2183.75946 | 363.95991 | 41.32 | <.0001 |
| **Error** | 65 | 572.60911 | 8.80937 |  |  |
| **Corrected Total** | 71 | 2756.36857 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2.96806 | **R-Square** | 0.7923 |
| **Dependent Mean** | 37.26901 | **Adj R-Sq** | 0.7731 |
| **Coeff Var** | 7.96388 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 14.39017 | 2.89157 | 4.98 | <.0001 |
| **X1** | 1 | 1.97132 | 0.43653 | 4.52 | <.0001 |
| **X2** | 1 | 9.13895 | 2.30071 | 3.97 | 0.0002 |
| **X3** | 1 | 0.56485 | 0.26266 | 2.15 | 0.0352 |
| **X4** | 1 | 0.33371 | 2.42131 | 0.14 | 0.8908 |
| **X5** | 1 | 1.90698 | 0.76459 | 2.49 | 0.0152 |
| **X6** | 1 | -1.04330 | 0.64759 | -1.61 | 0.1120 |

| **Number in Model** | **C(p)** | **R-Square** | **AIC** | **BIC** | **Variables in Model** |
| --- | --- | --- | --- | --- | --- |
| **6** | 7.0000 | 0.7923 | 163.2947 | 166.7792 | X1 X2 X3 X4 X5 X6 |

1. (5 points) Now let’s consider Model 1 and Model 2 as a pair of models. Does Model 1 nest Model 2 or does Model 2 nest Model 1? Explain.

Two linear models are nested if one model (the restricted model) can be obtained from the other model (the full model), by setting some parameters of the other model to zero.

The two models can be represented as,

Notice that the predictor variables in Model 1 are a subset of predictor variables in Model 2. Therefore, it can be said that Model 1 (the reduced model) nests within Model 2 (the full model).

1. (5 points) Write out the null and alternate hypotheses for a nested F-test using Model 1 and Model 2.

The null hypothesis is that all additional terms in the full model are equal to zero, with the alternate hypothesis that at least one of the additional terms in the full model are non-zero.

where,

is the number of parameters in the reduced model

is the number of additional parameters in the full model

1. (5 points) Compute the F-statistic for a nested F-test using Model 1 and Model 2.

where,

is the number of observations

is the number of parameters in the full model

is the number of parameters in the restricted model

is the sum of squares for residuals for the full model

is the sum of squares for residuals for the restricted model

therefore,

**Here are some additional questions to help you understand other parts of the SAS output.**

1. (0 points) Compute the AIC values for both Model 1 and Model 2.

Model 1

72 \* log(630.35953/72) + (2\*5)

166.2129

Model 2

72 \* log(572.60911/72) + (2\*7)

163.2947

1. (0 points) Compute the BIC values for both Model 1 and Model 2.

Model 1

72 \* log(630.35953/72) + (2\*(5+2)\*((72\*9.40835)/(630.35953))) - (2\*((72\*9.40835)/(630.35953))\*\*2)

= 168.9481

Model 2

72 \* log(572.60911/72) + (2\*(7+2)\*((72\*8.80937)/(572.60911))) - (2\*((72\*8.80937)/(572.60911))\*\*2)

= 166.7792

1. (0 points) Compute the Mallow’s Cp values for both Model 1 and Model 2.

Model 1

(630.35953/9.40835) + (2\*5) - 72

= 5

Model 2

(572.60911/8.80937) + (2\*7) – 72

= 7

1. (0 points) Verify the t-statistics for the remaining coefficients in Model 1.

Intercept

11.33027/1.99409

= 5.6819

X1

2.18604/0.41043

= 5.3262

X2

8.27430/2.33906

= 3.5374

X3

0.49182/0.26473

= 1.8578

X4

-0.49356/2.29431

= -0.2151

1. (0 points) Verify the Mean Square values for Model 1 and Model 2.

Model 1

2126.00904/4

= 531.5023

Model 2

2183.75946/6

= 363.9599

1. (0 points) Verify the Root MSE values for Model 1 and Model 2.

Model 1

sqrt(9.40835)

= 3.0673

Model 2

sqrt(8.80937)

= 2.9681