# 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) How many observations are in the sample data?
**Observations Total = (71 corrected total + 1 Degree's of Freedom)** *= 72*

(2) Write out the null and alternate hypotheses for the t-test for Beta1. The hypothesis for Beta1 is meant to be an implication of whether there is a linear relationship between the dependent and independent variable.
$H_0: B_1 = 0 \quad vs. \quad H_1: B_1 \neq 0$

(3) Compute the t- statistic for Beta1.
$T\text{-}Statistic = \frac{2.18604}{0.41043} = 5.3262$

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

$$R\_squared = \frac{2126.00904}{2756.36857} = .7713$$

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

$$Adjusted\ R\text{-}Squared = 1 - \frac{\frac{630.35953}{67}}{\frac{2756.36857}{71}} = .7576$$

(6) Write out the null and alternate hypotheses for the Overall F-test.

$$H_0 : B_1 = B_2 = B_3 = B_4 = 0 \quad vs. \quad H_a : B_i \neq 0,\ for\ at\ least\ one\ value\ of\ i$$

(7) Compute the F-statistic for the Overall F-test.

$$F\text{-}Statistic = \frac{531.5023}{9.40835} = 56.4926$$

**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 |

(8) 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.

***Model 1 =*** $\boldsymbol{y = B_0 + B_1(X_1) + B_2(X_2) + B_3(X_3) + B_4(X_4)}$
***Model 2 =*** $\boldsymbol{y = B_0 + B_1(X_1) + B_2(X_2) + B_3(X_3) + B_4(X_4) + B_5(X_5) + B_6(X_6)}$

***Two models are nested if both of the models contain the same terms and one has at least one additional term, therefore Model 1 is nested in Model 2.***

***Model 1 would be the reduced model, and Model 2 is the full model***

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

$\boldsymbol{H_0 : B_5 = B_6 = 0}$
$\boldsymbol{H_a : B_i \neq 0}$***, for at least one value of i***

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

***F-Statistic =*** $\boldsymbol{\frac{\frac{(630.35953-572.60911)}{2}}{\frac{572.60911}{65}} = 3.2777}$

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

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

***Model 1 AIC =*** $\boldsymbol{72 * \ln(\frac{630.35953}{72}) + 2 * 5 = 166.2129}$

***Model 2 AIC =*** $\boldsymbol{72 * \ln(\frac{572.60911}{72}) + 2 * 7 = 163.2946}$

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

***Model 1 BIC =*** $\boldsymbol{72 * \ln(\frac{630.35953}{72}) + 5 * \ln(72) = 177.5962}$

***Model 2 BIC =*** $\boldsymbol{72 * \ln(\frac{572.60911}{72}) + 7 * \ln(72) = 179.2313}$

(13) Compute the Mallow's Cp values for both Model 1 and Model 2.

***Model 1*** $\boldsymbol{C_p = (\frac{630.35953}{9.40835}) - 72 + 2 * 5 = 5}$

***Model 2*** $\boldsymbol{C_p = (\frac{572.60911}{8.80937}) - 72 + 2 * 7 = 7}$

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

$$\textit{T-Statistic } I = \frac{11.33027}{1.99409} = 5.6819$$

$$\textit{T-Statistic } X_1 = \frac{2.18604}{0.41043} = 5.3262$$

$$\textit{T-Statistic } X_2 = \frac{8.27430}{2.33906} = 3.5374$$

$$\textit{T-Statistic } X_3 = \frac{0.49182}{0.26473} = 1.8578$$

$$\textit{T-Statistic } X_4 = \frac{-0.49356}{2.29431} = -0.2151$$

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

$$\textit{Model 1 } MST = \left(\frac{2126.00904}{4}\right) = 531.5022$$

$$\textit{Model 1 } MSE = \left(\frac{630.35953}{67}\right) = 9.4083$$

$$\textit{Model 2 } MST = \left(\frac{2183.75946}{6}\right) = 363.9599$$

$$\textit{Model 2 } MSE = \left(\frac{572.60911}{65}\right) = 8.8093$$

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

$$\textit{Model 1 } RMSE = \sqrt{9.408351194 = 3.067303}$$

$$\textit{Model 2 } RMSE = \sqrt{8.809370923 = 2.968058}$$