**Linear “OLS” Regression**

**Error Metrics (10 points)**

Assume two models were developed using 50 data points (n=50). The models and their respective ANOVA tables are given below.

**Model 1:** Let’s consider the regression model, which we will refer to as Model 1, given by

Y = 10,000 + 150\*X1 + 25\*X1^2 + 60\*X2 (M1).



**Model 2:** Now let’s consider an alternate regression model, which we will refer to as Model 2, given by

Y = 9,750 + 145\*X1 + 75\*X2 (M2).



Assignment: Research the formula for the following metrics. Also, determine how to interpret the results.

* Adjusted R-Squared
* Akaike Information Criteria (AIC)
* Akaike Information Criteria with correction for sample size (AICC)
* Schwarz Information Criteria (BIC, SBC)

Using the information given for M1 and M2, calculate (by hand or using EXCEL) the **Adjusted R-Squared, AIC, SBC,** and **AICC** for both models (M1 and M2). Simple, right? OK, here's the catch. You need to calculate the values \*AND\* comment on them (1 or 2 sentences should suffice). I just need to know that you know what these metrics are and how to use them. For example, you might say,

* “Using Adjusted R-Squared, I would prefer Model M1 because ….”
* “Using AIC, I would prefer Model M2 because ….”

**Solution:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Calculations** | **Formulas** | **Model 1** | **Model 2** |
| No of Data Points (n) | n | 50 | 50 |
| # of predictors terms in Model | p | 3 | 2 |
| # of parameters in Model | k or p+1 | 4 | 3 |
| Sum of Squares from the Regression | SSR | 750 | 725 |
| Sum of Squares for the Error | SSE | 250 | 275 |
| Total Sum of Squares | SST | 1000 | 1000 |
|  |  |  |  |
| **• Adjusted R-Squared** | 1-((n-1)/(n-p-1))\*(SSE/SST) | 0.733695652 | 0.713297872 |
| **• Akaike Information Criteria (AIC)** | n\*ln(SSE/n)+2\*(p+1) | 88.47189562 | 91.23740461 |
| **• Akaike Information Criteria with  correction for sample size (AICC)** | AIC + (2(p+1+2)(p+1+3)/(n-p+1-3)) | 90.42538399 | 92.60104098 |
| **• Schwarz Information Criteria (BIC, SBC)** | n\*ln(SSE/n)+ (p+1) ln(n) | 96.11998764 | 96.97347363 |

* When comparting two models using Adjusted R-Squared, Model with highest Adjusted R-Squared value should be selected. Using Adjusted R-Squared, I would prefer Model1.
* When comparting two models using AIC, Model with smaller AIC value is preferred. Using AIC, I would prefer Model1.
* When comparting two models using AICC, Model with smaller AICC value is preferred Using AICC, I would prefer Model1.
* When comparting two models using BIC, Model with smaller BIC value is preferred. Using BIC, I would prefer Model1.

From all four model statistics, model 1 is clearly better choice than model 2.