STA 674

Regression Analysis And Design Of Experiments

Fitting Multiple Linear Regression Models – Lecture 5

STA 674, RADOE:
Fitting Multiple Linear Regression Models

- Last time, we discovered the ANOVA table for multiple linear regression.
- This time, we look at some more examples and talk about some *derived values* used more and less in different fields.

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

Source	DF	Sum of Squares	Mean Square	F value	Prob > F
Model	2	374991	187495	7.38	0.0026
Error	29	736573	25399		
Total	31	1111564			

Root MSE 159.37083 **R-Square** 0.3374

Dependent Mean 825.37500 **Adj R-Sq** 0.2917

Coeff Var 19.30890

y bar = mean of all data points

RMSE divided by y bar times 100

STA 674, RADOE:
Fitting Multiple Linear Regression Models

ANOVA Tables – Derived Values I

- Root Mean Squared Error ($RMSE$)

$$s_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-K-1}} = \sqrt{\frac{SSE}{n-K-1}} = \sqrt{MSE}$$

- $RMSE$ is an estimate of σ_e

STA 674, RADOE:
Fitting Multiple Linear Regression Models

ANOVA Tables – Derived Values II

- ***R*-squared** (R^2)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

This always goes up as paramters are added

(model includes all predictors included in the model)

R-squared measures the proportion of variance in the data that is explained by the model.

- **Adjusted *R*-squared** ($\text{Adj}R^2$)

$$\text{Adj}R^2 = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)}$$

- The adjusted R-squared measures the proportion of variance in the data that is explained by the model discounted for the number of parameters.

best to compare different models

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

Source	DF	Sum of Squares	Mean Square	F value	Prob > F
Model	2	374991	187495	7.38	0.0026
Error	29	736573	25399		
Total	31	1111564			

Root MSE 159.37083 **R-Square** 0.3374

Dependent Mean 825.37500 **Adj R-Sq** 0.2917

Coeff Var 19.30890

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Exercise: topic-starting example

The following tables present the ANOVA for the model of y as a function of x_1 and x_2 described in an earlier lecture. Use this information to summarize the model.

Source	DF	Sum of Squares	Mean Square	F value	Prob > F
Model	2	727.72772	363.86386	336.14	<0.0001
Error	7	7.57728	1.08247		
Total	9	735.30500			

Root MSE 1.04042 **R-Square** 0.9897

Dependent Mean 16.15000 **Adj R-Sq** 0.9868

Coeff Var 6.44222