STA 674

Regression Analysis And Design Of Experiments

Fitting Multiple Linear Regression Models – Lecture 4

STA 674, RADOE:
Fitting Multiple Linear Regression Models

- Last time, we covered sources of squared error: regression versus error (adding to the total.)
- This time, we discuss ANOVA for multiple linear regression.

SSE error -- y-yhat
SSR regression (model) -- yhat-ybar
SST total -- y-ybar
SSE + SSR = SST

STA 674, RADOE:

Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- Consider the example of predicting the output of a nuclear power plant based on the year in which it was built and the cost.

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			

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- Consider the example of predicting the output of a nuclear power plant based on the year in which it was built and the cost.

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			

$$\frac{SSE}{SST} = \frac{736573}{1111564} = 0.663$$

Is this small enough to conclude that the model is useful?

STA 674, RADOE:
Fitting Multiple Linear Regression Models

ANOVA Tables – Degrees of Freedom (DF)

	SS	DF
Regression	$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	K = number of predictors
Error	$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$n - K - 1$ = number of data points minus number of predictors minus 1
Total	$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$	$n - 1$ = number of data points minus 1

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- Consider the example of predicting the output of a nuclear power plant based on the year in which it was built and the cost.

from formulas

normalized by df

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			

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Mean Squares

- Means squares are computed by dividing the sums of squares by their degrees of freedom:

$$MSR = \frac{SSR}{K} \text{ and } MSE = \frac{SSE}{n - K - 1}$$

F-Statistic

- The F-statistic is the ratio of MSR to MSE:

$$F = \frac{MSR}{MSE}$$

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- Analysis of Variance – MSR's and F statistic

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			

STA 674, RADOE:

Fitting Simple Linear Regression Models

ANOVA tables—the *F*-test

- **Hypotheses**: $H_0: \beta_1 = \beta_2 = \cdots = \beta_K = 0$ vs. $H_a: \beta_k \neq 0$ for some k

  all predictor Bk equal to 0

  at least one predictor Bk not equal to 0

- **Test Statistic**: $F = \frac{MSR}{MSE}$

- **Rejection Rule**: Reject H_0 at the α level of significance if $F > F_{\alpha,K,n-K-1}$. Otherwise, we fail to reject H_0.

  level of significance

  numerator df

  denominator df

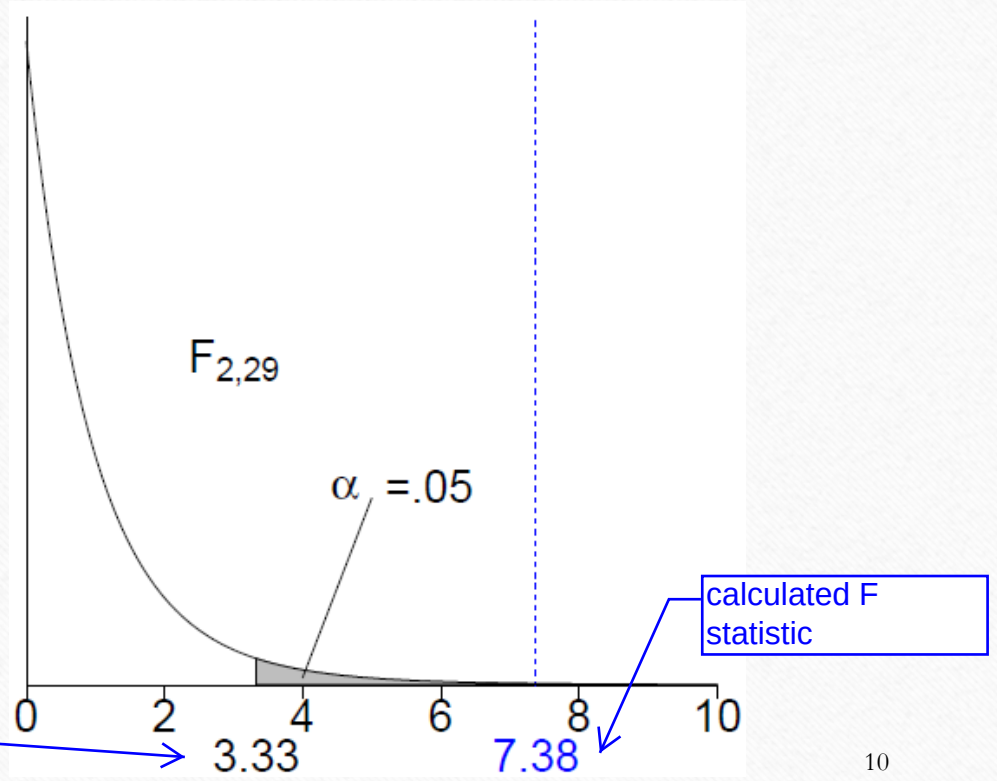
- ***P*-value**: $P\left(F_{K,n-K-1} > F\right)$ all are single tail test...F test is always a right tail test

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- If H_0 is true, then F is distributed according to an F-distribution with 2 and 29 degrees of freedom:

Two wasy to determine significance...
1. compare critical F value with F statistic
2. compare p value with alpha



STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: Nuclear Reactor Data

- Analysis of Variance – *P*-value

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			