Multiple Linear Regression

Cesar Acosta Ph.D.

Department of Industrial and Systems Engineering
University of Southern California

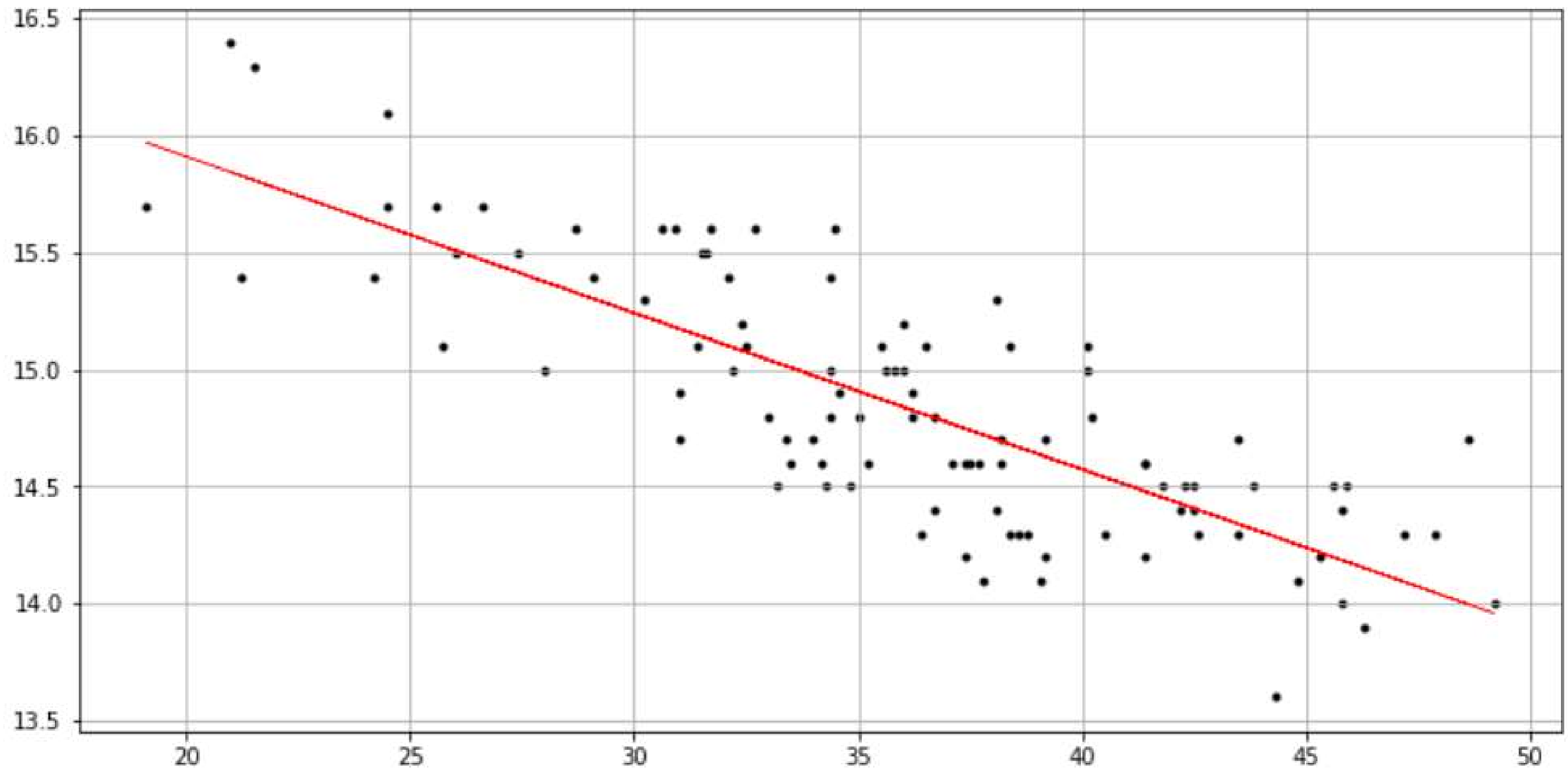
REGRESSION ANALYSIS

Regression analysis is useful to find a relationship between a response and a set of predictors

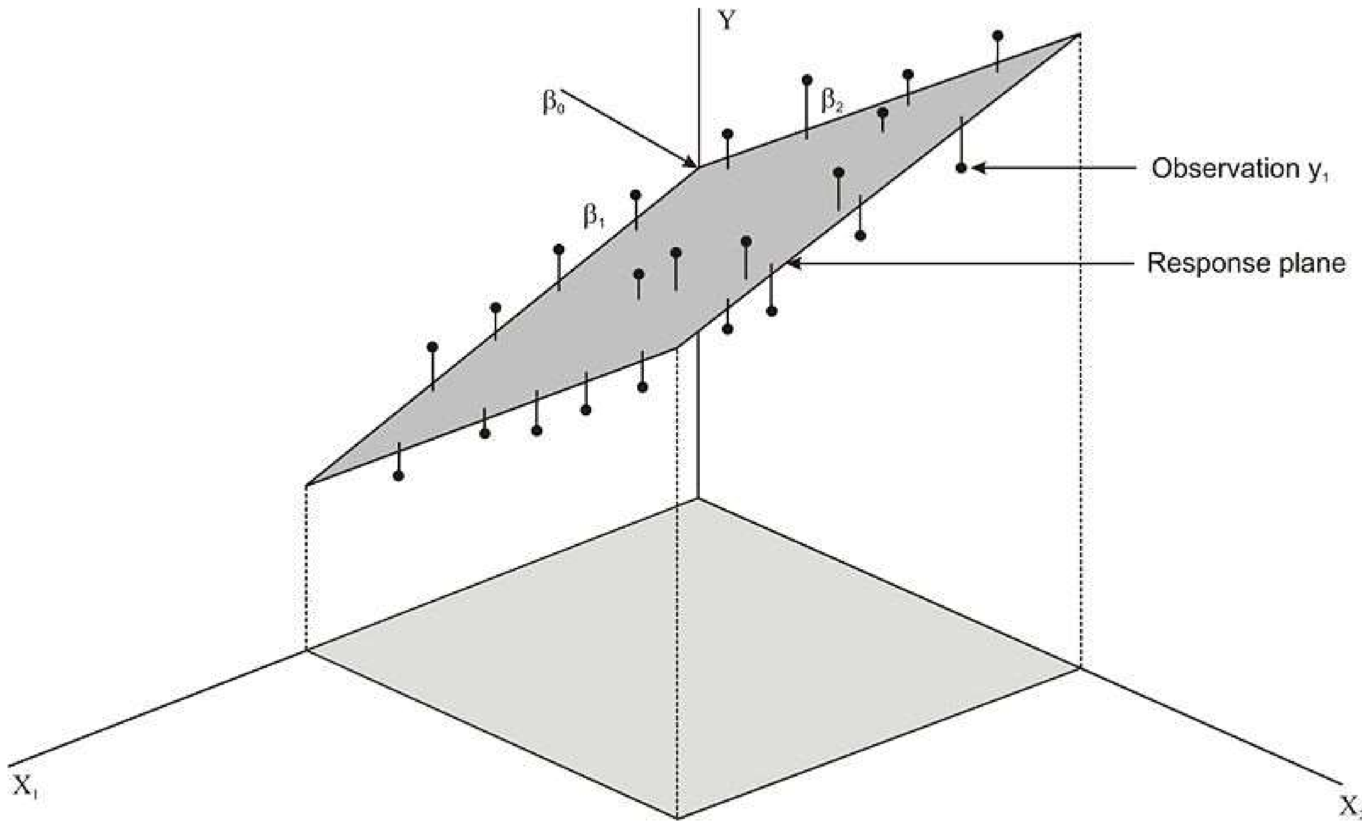
Two Regression Models

- Simple linear regression (SLR)
- Multiple linear regression (MLR)

OLS line – one predictor



2 predictors

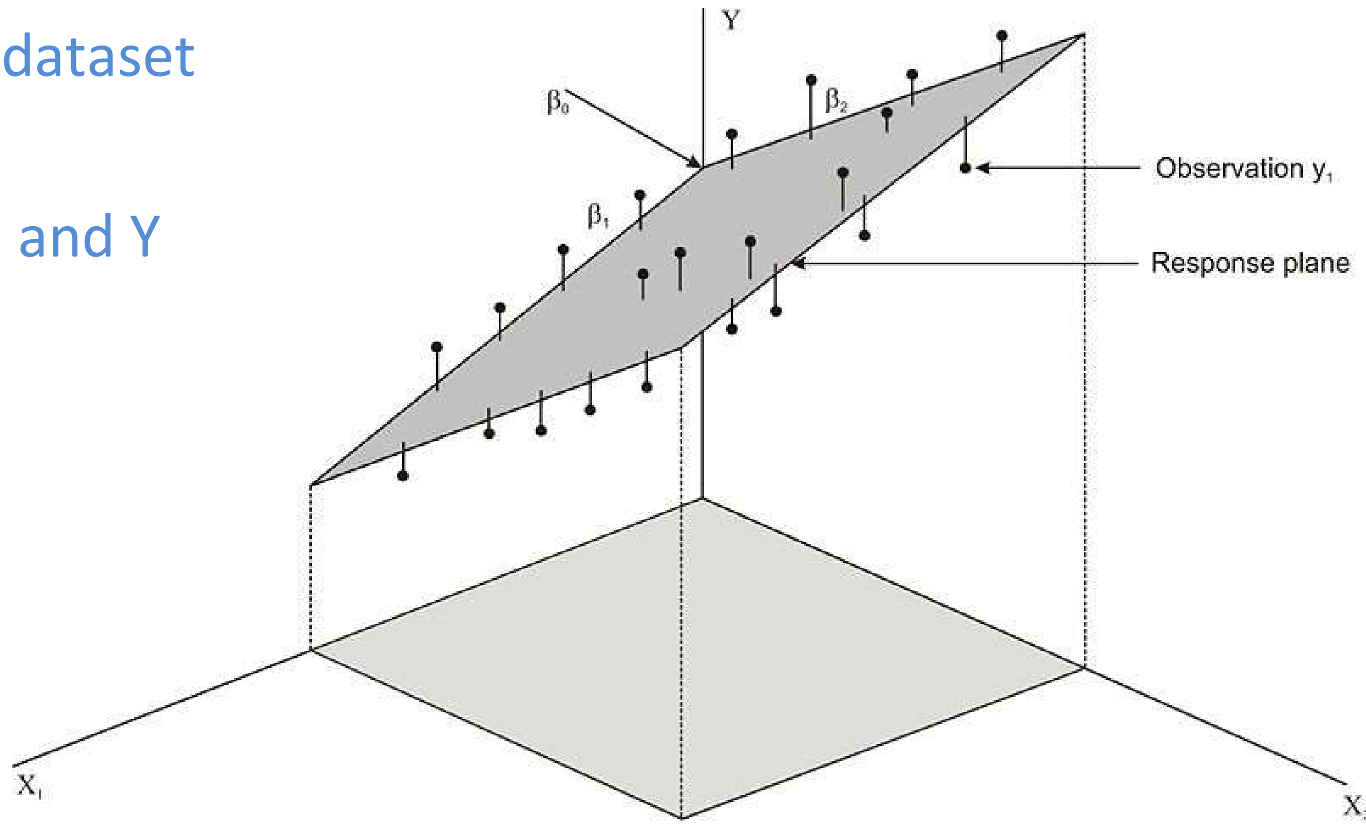


OLS plane

Closest plane to dataset

Average relation

between (X1,X2) and Y



Multiple Linear Regression

Consider predictors $X_1, X_2, \ldots, X_p$

Regression plane $$E[Y] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Multiple Linear Regression

Consider predictors X_1, X_2,…,X_p

Regression plane $$E[Y] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Fitted plane $$\hat{Y} = b_0 + b_1 X_1 + \cdots + b_p X_p$$

Model performance

How good is the regression model?

Model performance

How good is the regression model?

- *How well the model fits the data?*
- *How well the model predicts the data?*

Model performance

How good is the regression model?

- *How well the model fits the data?*
- *How well the model predicts the data?*

Model performance

How good is the regression model?

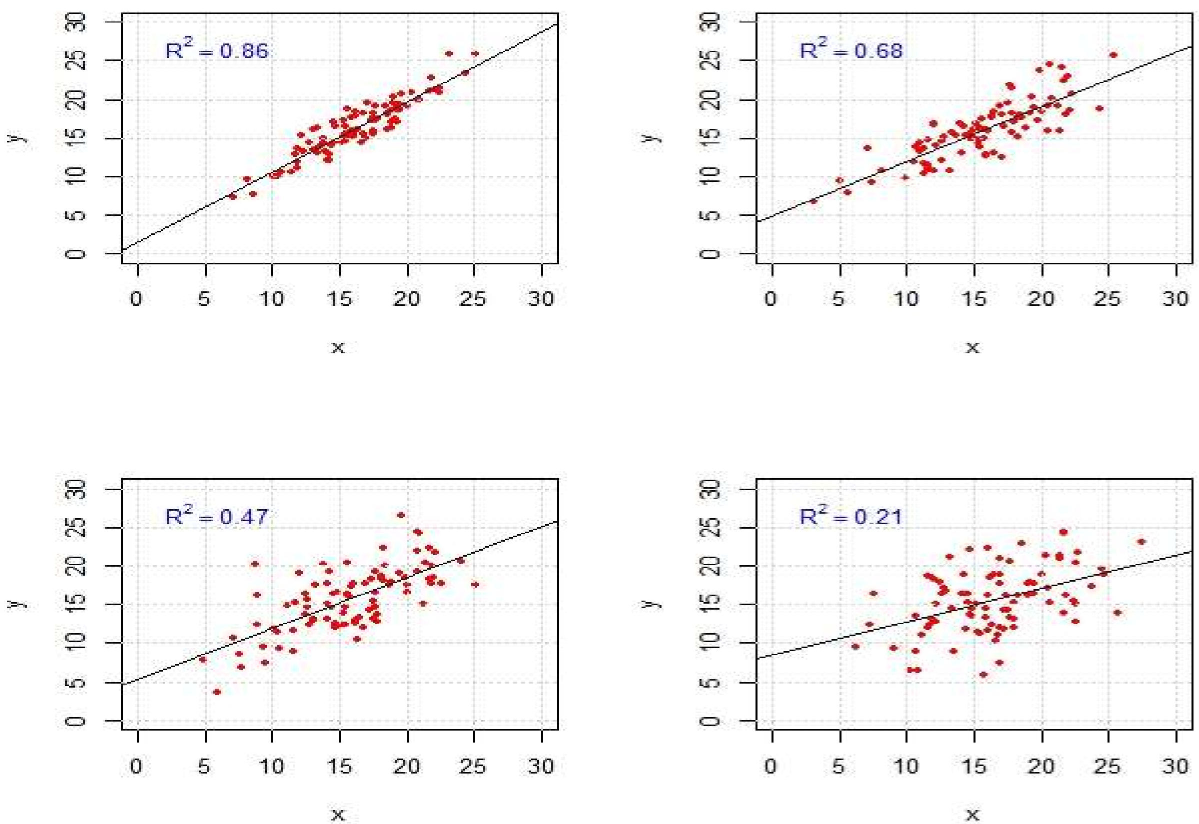
- *How well the model **fits** the data?* SSE R^2

- *How well the model **predicts** the data?*

Model adequacy

R^2 is the proportion of the variation in Y that is explained by $X_1, X_2, \ldots, X_p$

One predictor

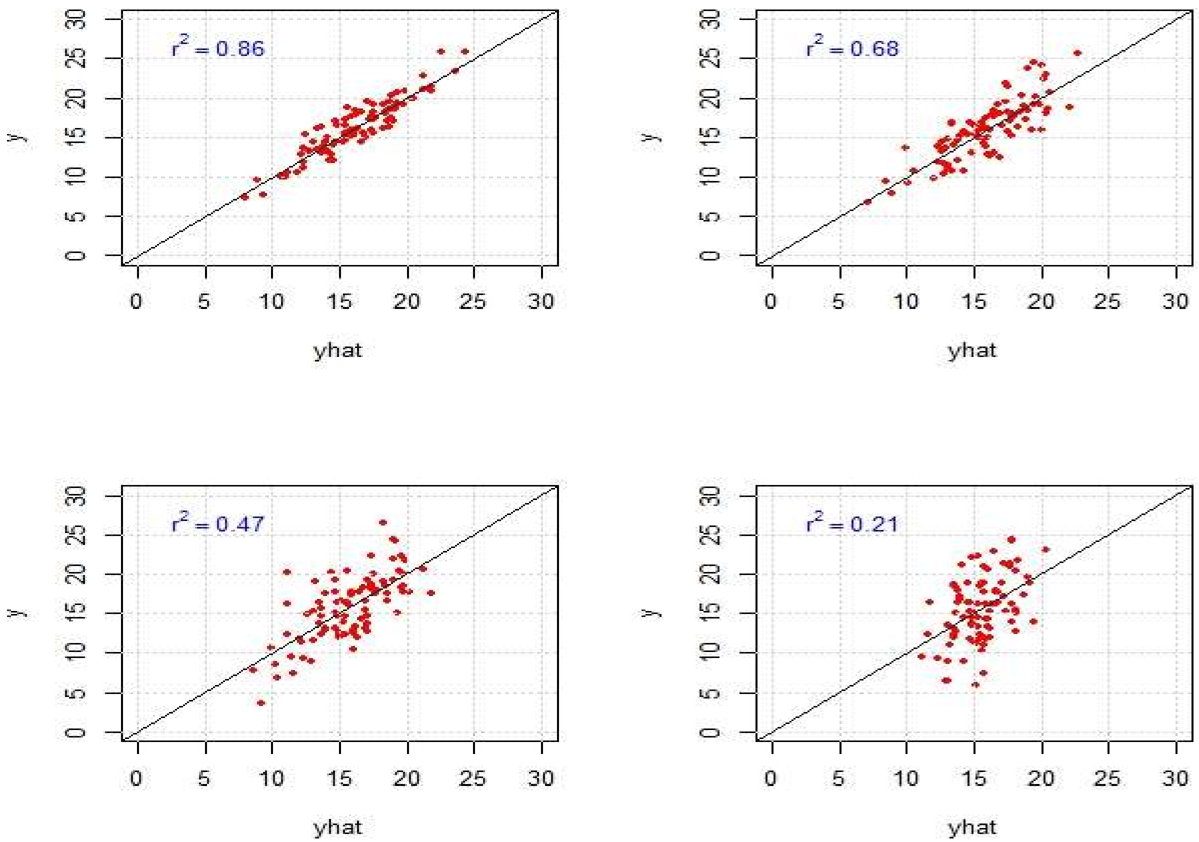


R-squared vs correlation (y, yhat)

If r is the correlation between y and yhat

$$R^2 = r^2$$

y vs. yhat plot



MULTIPLE LINEAR REGRESSION

We say that p increases

when more predictors (features)

are added to the model

MULTIPLE LINEAR REGRESSION

We say that *p* decreases

when some predictors (features)

are removed from the model

MULTIPLE LINEAR REGRESSION

Given a set of *p* predictors

- What is the best predictor?
- What is the best set of predictors?`

MULTIPLE LINEAR REGRESSION

Fact 1

If p increases, SSE decreases, always

MULTIPLE LINEAR REGRESSION

Fact 1

If p increases, SSE decreases, always

Therefore,

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

if p increases, R^2 increases

MULTIPLE LINEAR REGRESSION

R^2 useful to compare models

with the same number *p* of predictors

MULTIPLE LINEAR REGRESSION

R^2 useful to compare models

with the same number p of predictors

R^2 not useful to compare models

with different number p of predictors

MULTIPLE LINEAR REGRESSION

How to compare models with different number of predictors?

MULTIPLE LINEAR REGRESSION

How to compare models with different number of predictors?

- Adjusted R^2
- AIC (Information Criteria)

MULTIPLE LINEAR REGRESSION

How to compare models with different number of predictors?

- Adjusted R^2 (larger is better)
- AIC (smaller is better)

MULTIPLE LINEAR REGRESSION

Fact 2

If *p* increases, MSE may increase or decrease

$$MSE = \frac{SSE}{n-p-1}$$

MULTIPLE LINEAR REGRESSION

Fact 2

If *p* increases, MSE may increase or decrease

$$MSE = \frac{SSE \downarrow}{n - p - 1 \downarrow}$$

MULTIPLE LINEAR REGRESSION

Fact 2

If p increases, MSE may increase or decrease

Therefore, if p increases

$$\text{adjusted-}R^2 = 1 - \frac{MSE}{MST}$$

adjusted-R^2 may increase or decrease

MULTIPLE LINEAR REGRESSION

Fact 2

If *p* increases, MSE may increase or decrease

Therefore, if p increases

$$\text{adjusted-}R^2 = 1 - \frac{MSE}{MST}$$

select number of predictors s.t. adj-R^2 is smallest

MULTIPLE LINEAR REGRESSION

Adj-R^2 useful to compare models

with different number of predictors

MS formulas for Simple Linear Regression

$$SSTotal = \sum_{i=1}^{n} (y - \overline{y})^2 \qquad MSTotal = \frac{SSTotal}{n-1}$$

sample variance Y

$$SSR = \sum_{i=1}^{n} (\hat{y} - \overline{y})^2 \qquad MSR = \frac{SSR}{1}$$

$$SSE = \sum_{i=1}^{n} (y - \hat{y})^2 \qquad MSE = \frac{SSE}{n-2}$$

MS formulas for Multiple Linear Regression

$$SSTotal = \sum_{i=1}^{n}(y_i - \overline{y})^2 \qquad MSTotal = \frac{SSTotal}{n-1}$$

sample variance Y

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \qquad MSR = \frac{SSR}{p}$$

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

ANOVA DECOMPOSITION, *R-squared*

$$SSTotal = SSE + SSR$$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

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

ANOVA DECOMPOSITION, *R-squared*

$$SSTotal = SSE + SSR$$

$$1 = \frac{SSE}{SST} + \boxed{\frac{SSR}{SST}}$$

$$1 = \frac{SSE}{SST} + \boxed{R^2}$$

ANOVA DECOMPOSITION, *R-squared*

$$SSTotal = SSE + SSR$$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$\boxed{1 = \frac{SSE}{SST} + R^2}$$

R-squared vs adj *R-squared*

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

$$adj\ R^2 = 1 - \frac{MSE}{MST}$$

R-squared vs adj *R-squared - interpretation*

- *$100R^2$ is the percentage of variation in Y that is explained by the model*

- *Adjusted R^2 has no interpretation*

R-squared *vs* adj *R-squared*

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

$$adj\ R^2 = 1 - \frac{MSE}{MST}$$

adj *R-squared relation with R-squared*

$$adj\ R^2 = 1 - \frac{MSE}{MST}$$

$$= 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}}$$

adj *R-squared relation with R-squared*

$$adj\ R^2 = 1 - \frac{MSE}{MST}$$

$$= 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}}$$

$$= 1 - \frac{n-1}{n-p-1}\frac{SSE}{SST}$$

adj *R-squared relation with R-squared*

$$adj\ R^2 = 1 - \frac{n-1}{n-p-1}\ \frac{SSE}{SST}$$

$$1 - adj\ R^2 = \frac{n-1}{n-p-1}\ \frac{SSE}{SST}$$

adj *R-squared relation with R-squared*

$$adj\ R^2 = 1 - \frac{n-1}{n-p-1}\frac{SSE}{SST}$$

$$1 - adj\ R^2 = \frac{n-1}{n-p-1}\frac{SSE}{SST}$$

$$1 - adj\ R^2 = \frac{n-1}{n-p-1}(1 - R^2)$$

adj *R-squared relation with R-squared*

$$1 - adj\, R^2 = \frac{n-1}{n-p-1}(1-R^2)$$

$$adj\, R^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

adj *R-squared relation with R-squared*

$$1 - adj\, R^2 = \frac{n-1}{n-p-1}(1 - R^2)$$

$$\boxed{adj\, R^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2)}$$

MULTIPLE LINEAR REGRESSION

Akaike Information Criteria

Measures the lost of information by fitting a model from a sample.

$$AIC = n \log\left(\frac{SSE}{n}\right) + 2p$$

MULTIPLE LINEAR REGRESSION

As the number *p* of predictors increases

SSE decreases

$$AIC \;=\; n \log\left(\frac{SSE}{n}\right) + 2p$$

AIC may increase or decrease