STA 674

Regression Analysis And Design Of Experiments

Fitting Multiple Linear Regression Models – Lecture 3

STA 674, RADOE:
Fitting Multiple Linear Regression Models

- Last time, we covered inference.
- This time, we cover sources of error, preparing to use ANOVA as a tool in multiple linear regression significance testing.

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Introduction

- So far we have considered inference (confidence intervals and significance tests) for individual parameters. However, we can also conduct inference for a model as a whole.

Question

IS OUR LINEAR REGRESSION MODEL USEFUL?

- Do the set of predictor variables in a linear regression model help to predict the response or do we do almost as well by assuming a constant mean?

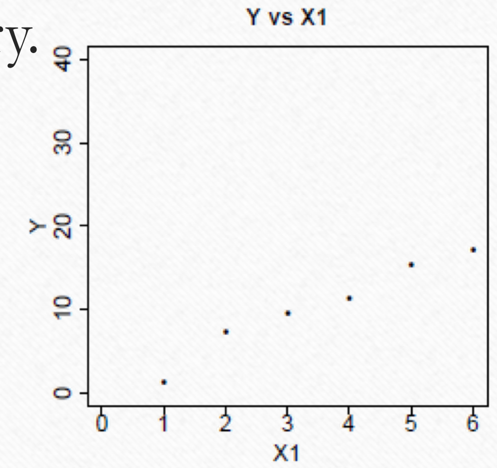
STA 674, RADOE:
Fitting Multiple Linear Regression Models

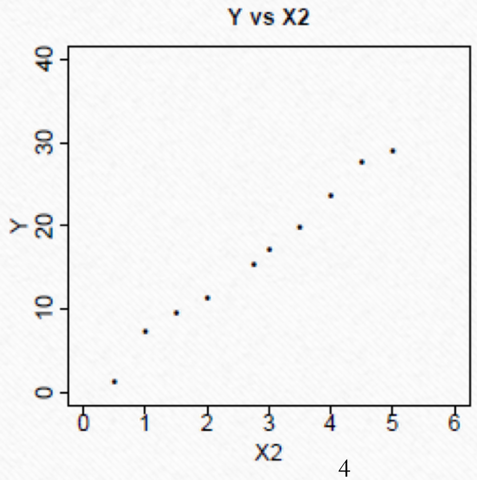
Exercise:

SLR of each predictor shows a strong positive association, but MLR using both predictors does not...because in MLR we use inference when other predictor(s) is constant

- The following plots illustrate the relationship between a response, y, and two predictors, x_1 and x_2 and the table provides information from the multiple regression model. Write a brief summary.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.27432	0.72568	-0.38	0.7166
x1	1	1.55851	2.19343	0.71	0.5004
x2	1	2.82973	4.39419	0.64	0.5401





STA 674, RADOE:
Fitting Multiple Linear Regression Models

SSE

- The basis for assessing how well a model is the data is the sum of squares error (SSE):

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

- If the SSE is small then it means that the points, on average, lie close to the fitted values. If the SSE is big then it means that the points, on average, lie far from the fitted values.

STA 674, RADOE:
Fitting Multiple Linear Regression Models

SST

- To judge if the SSE is big or small, we can compare it with another sum of squares – called the sum of squares total (SST):

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

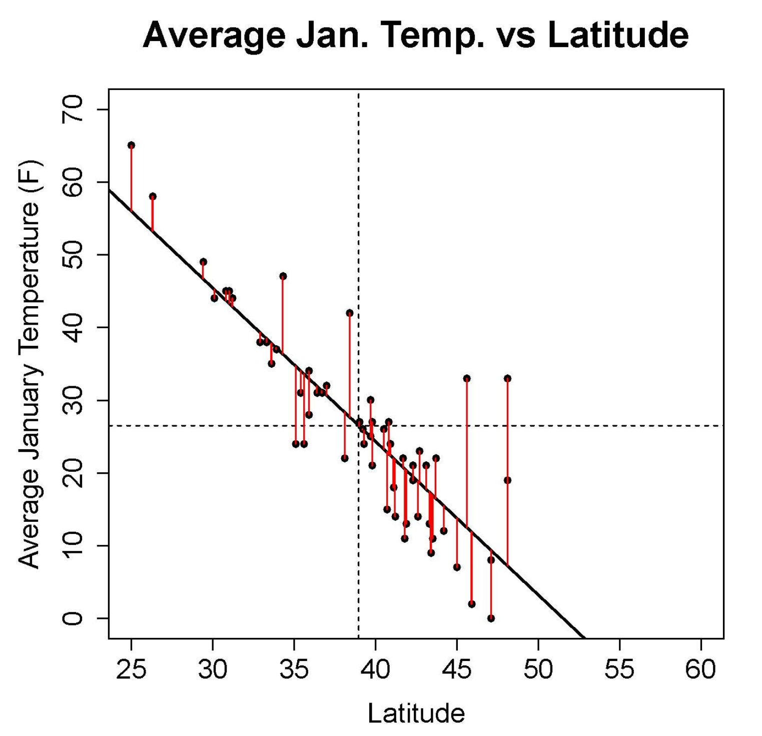
- This sum of squares measures the average (squared) distance between the observed responses and their mean – the best guess if we were to ignore all of the predictor variables.
- If SSE is small compared to SST then the linear regression model has reduced the residual errors by a lot.
- If SSE is big compared to SST then it means that the residuals errors have decreased only a little.

STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: US temperature data - ***SSE***

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

y hat = mean of regression line
y i = data point

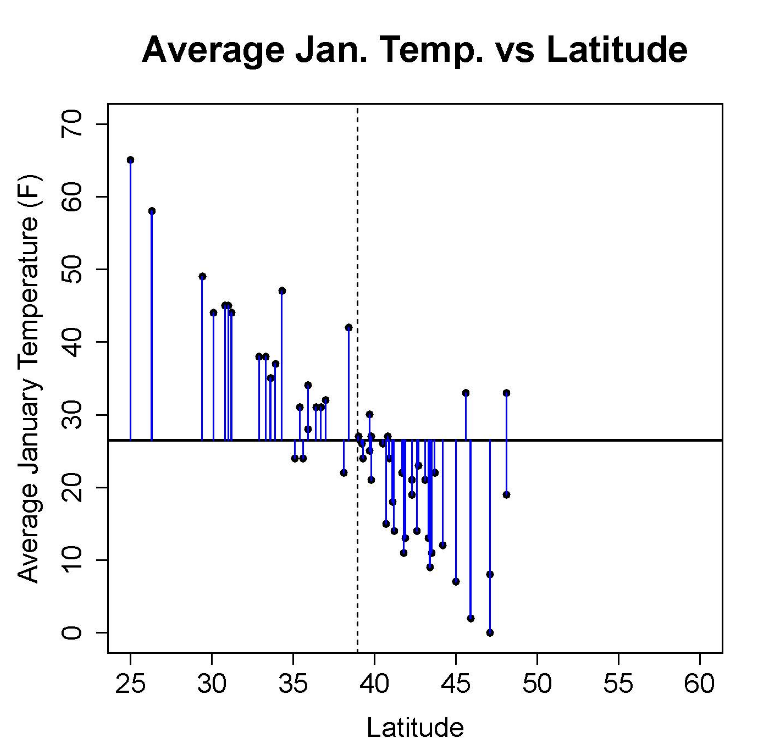


STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: US temperature data - ***SST***

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

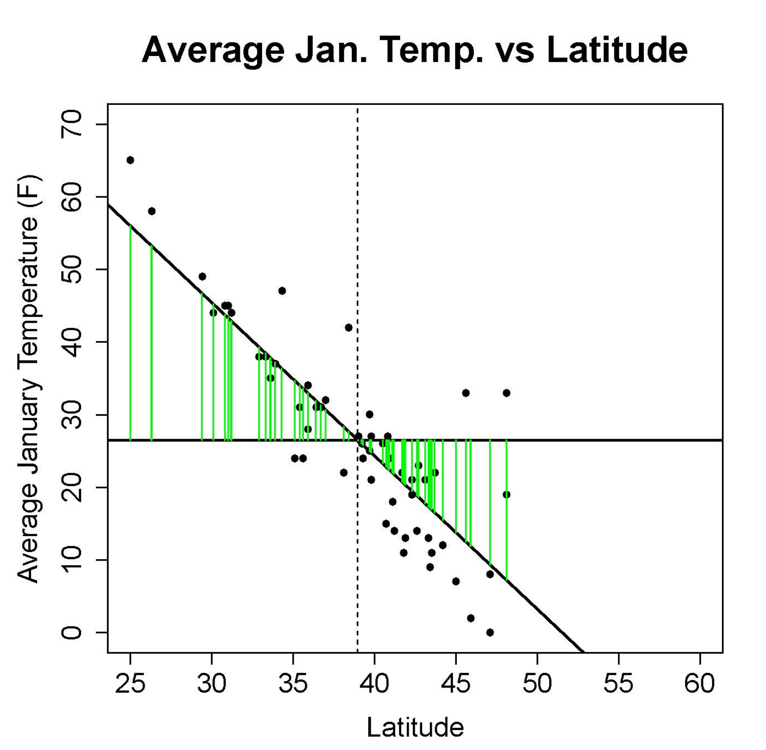
y bar = mean of all data points
y i = data points



STA 674, RADOE:
Fitting Multiple Linear Regression Models

Example: US temperature data - *SSR*

- The difference between the SSE and SST is the ***sum of squares for the regression***: $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$



STA 674, RADOE:
Fitting Multiple Linear Regression Models

Sum of Squares:

Regression	$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	SSR
		$+$
Error	$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	SSE
		$=$
Total	$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$	SST