STA 674

Regression Analysis And Design Of Experiments

Fitting Simple Linear Regression Models – Lecture 1

STA 674, RADOE:
Fitting Simple Linear Regression Models

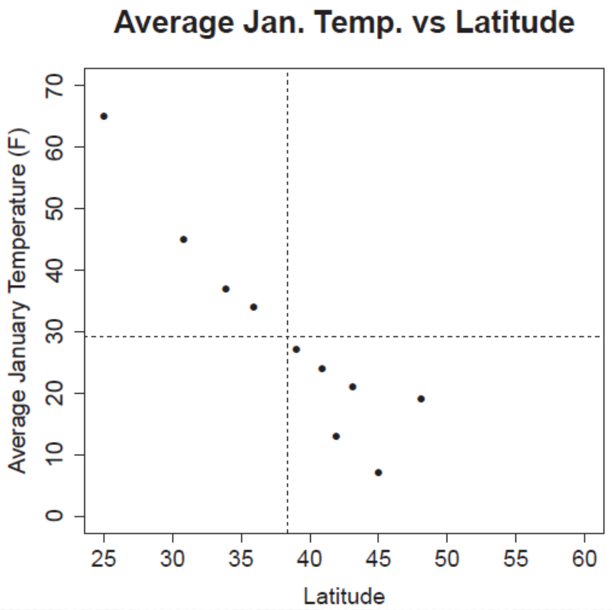
- Where does it fit in?
- What is it?
- Where next?

STA 674: Fitting Simple Linear Regression Models
Where does it fit in?

Regression and correlation provide the basis for studying relationships between two or more variables and for answering questions about associations which arise in all areas of quantitative research.

STA 674: Fitting Simple Linear Regression Models
What is it?

- First of all, you can't get very far if lines are mysterious (or merely "fuzzy", math-wise,) so I've posted a help file on Canvas linear_function_help.pdf.
- Next, consider
  our sample of cities:
- Is there a linear
  function that goes
  through all the points?



STA 674: Fitting Simple Linear Regression Models

- Residual Error: we will account for the fact that the points do not lie exactly along a straight line by incorporating errors (disturbances) into the linear model.
- For the i^{th} observation we will have:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

STA 674: Fitting Simple Linear Regression Models

- **Stochastic (or random, or probabilistic) model**: we will account for the fact that the points do not lie exactly along a straight line by incorporating errors (disturbances) into the linear model.
- For the i^{th} observation we will have:

$$y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 x_i} + \boxed{e_i}$$

ei =
amount of error in y...distance from linear model line (B0+B1x) in vertical direction

STA 674: Fitting Simple Linear Regression Models

- Remembering that $y_i = \beta_0 + \beta_1 x_i + e_i$, the value

$$\hat{y}_i = \mu_{y_i|x_i} = \beta_0 + \beta_1 x_i$$

Mu yi ! xi = estimated value at yi given value at xi

is called the **fitted value** or **conditional mean** of y given x. It represents the value of y_i given x_i – the value we would expect if all the points lay precisely on the line.

- What's the difference between y_i and $\hat{y}_i$ (literally and conceptually)?

error between predicted and actual value = residual error = ei

STA 674: Fitting Simple Linear Regression Models

- **Residual Error**: the value

$$e_i = y_i - \hat{y}_i = y_i - (\beta_0 - \beta_1 x_i)$$

is called the *residual error*. It is the difference between the observed value for the i^{th} observation, y_i, and the value we expected, $\mu_{y_i|x_i}$.

STA 674: Fitting Simple Linear Regression Models

Sources of error:

- error (missing components in the model) other variables affecting prediction; i.e., humidity affecting termperature, etc
- sampling error (variation between units)
- experimental error (measurement error) instrument error

The presence of error is why we need statistics.

STA 674: Fitting Simple Linear Regression Models

Questions to consider:

1. Estimation: What are the best guesses (estimates) of β_0 and β_1?

2. Inference:

- Confidence intervals: How would the estimates change if we collected new data?
- Significance testing: Can we conclude that $\beta_1 > 0$, $\beta_1 < 0$, $\beta_1 = 0$?

3. Prediction:

Given a new value of x:

- What is the most plausible value of y?
- What range of values is plausible?