STA 674

Regression Analysis And Design Of Experiments

Fitting Simple Linear Regression Models – Lecture 3

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Last time: introduced to LSE—least squares estimation.
- This time we will talk about where that estimation comes from and why it is useful.

Use statistics of linear model (b0 and b1) to make inference (process) about the population parameters (beta0 and beta1)

STA 674, RADOE:
Fitting Simple Linear Regression Models

Some Notation

- **Population parameters**: Population parameters are denoted with Greek letters—β_0 and β_1 are the true intercept and slope—and are numbers that describe the population.
- **Parameter estimates**: Least squares estimates are denoted with Latin letters—b_0 and b_1 are the least squares estimates of the intercept and slope—and are numbers calculated from the sample (statistics).

STA 674, RADOE:
Fitting Simple Linear Regression Models

Some Notation

- **Fitted values**: Given estimates b_0 and b_1, the fitted value for observation i is:

$$\hat{y}_i = b_0 + b_1 x_i$$

- **Residual**: the residual:

$$\hat{e}_i = y_i - \hat{y}_i$$

is an estimate of the error for observation i.

STA 674, RADOE:
Fitting Simple Linear Regression Models

Least Squares Estimation – why?

BLUE - best linear unbiased estimate

- **Unique**: LS estimates, b_0 and b_1, for a given set of data are unique.
- (Could be) **Computed by hand**:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

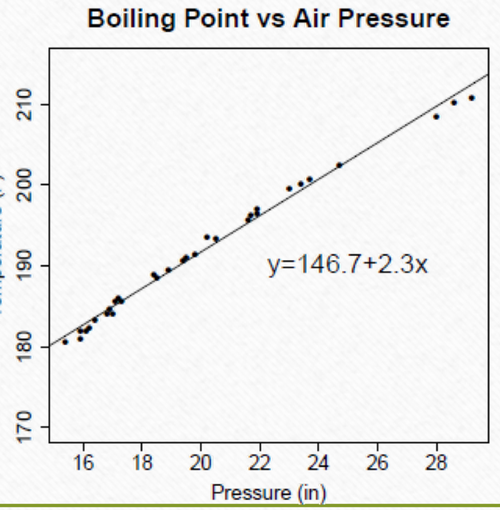
$$b_0 = \bar{y} - b_1\bar{x}$$

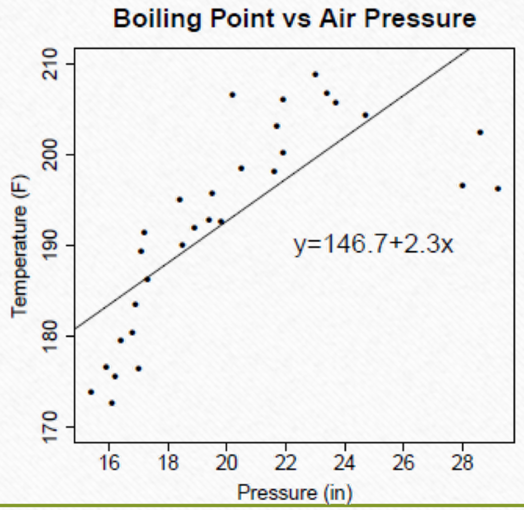
- **Optimal**: provided we assume the distribution of the errors is normal

STA 674, RADOE:
Fitting Simple Linear Regression Models

Least Squares Estimation – one more thing …

- Look at the two plots—the estimates, b_0 and b_1, for both are the same (146.7 and 2.3, respectively.)
- What's different?





STA 674, RADOE:

Fitting Simple Linear Regression Models

Least squares estimation – the error variance

The full model of y_i is:

y i hat...estimate of linear model

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

all variability is in ei

The (usually, but not here) implicit assumptions in linear regression are that:

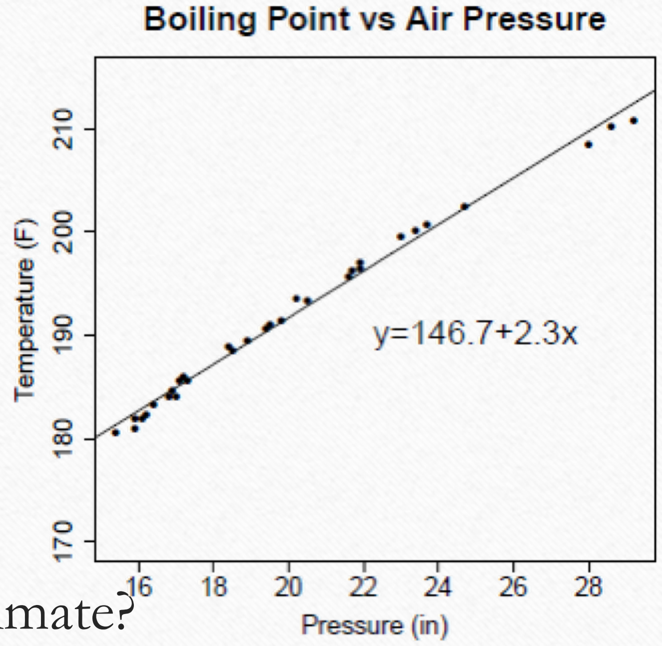
1. the expected value of the $e_1, e_2, \dots, e_n$ is 0.
2. the variance of the $e_1, e_2, \dots, e_n$ is σ_e^2
3. $e_1, e_2, \dots, e_n$ are normally distributed
4. $e_1, e_2, \dots, e_n$ are independent

STA 674, RADOE:
Fitting Simple Linear Regression Models

Least squares estimation – the error variance

The least squares estimate (LSE) of σ_e^2 is:

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



Interpret the slope, intercept of the LS line.

What would be your estimate of the boiling temperature at 22 inches of air pressure?

$s_e = \sqrt{s_e^2} \approx 0.8$°F. What would that tell you about your estimate?