STA 674

Regression Analysis And Design Of Experiments

Fitting Simple Linear Regression Models – Lecture 4

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Last time: finally got to the LS estimates of b_0, b_1, and s_e^2.
- Before we get to the useful (interval) estimates of these, we need to do a cursory review of sampling distributions.

STA 674, RADOE:
Fitting Simple Linear Regression Models

Sampling distribution – what?

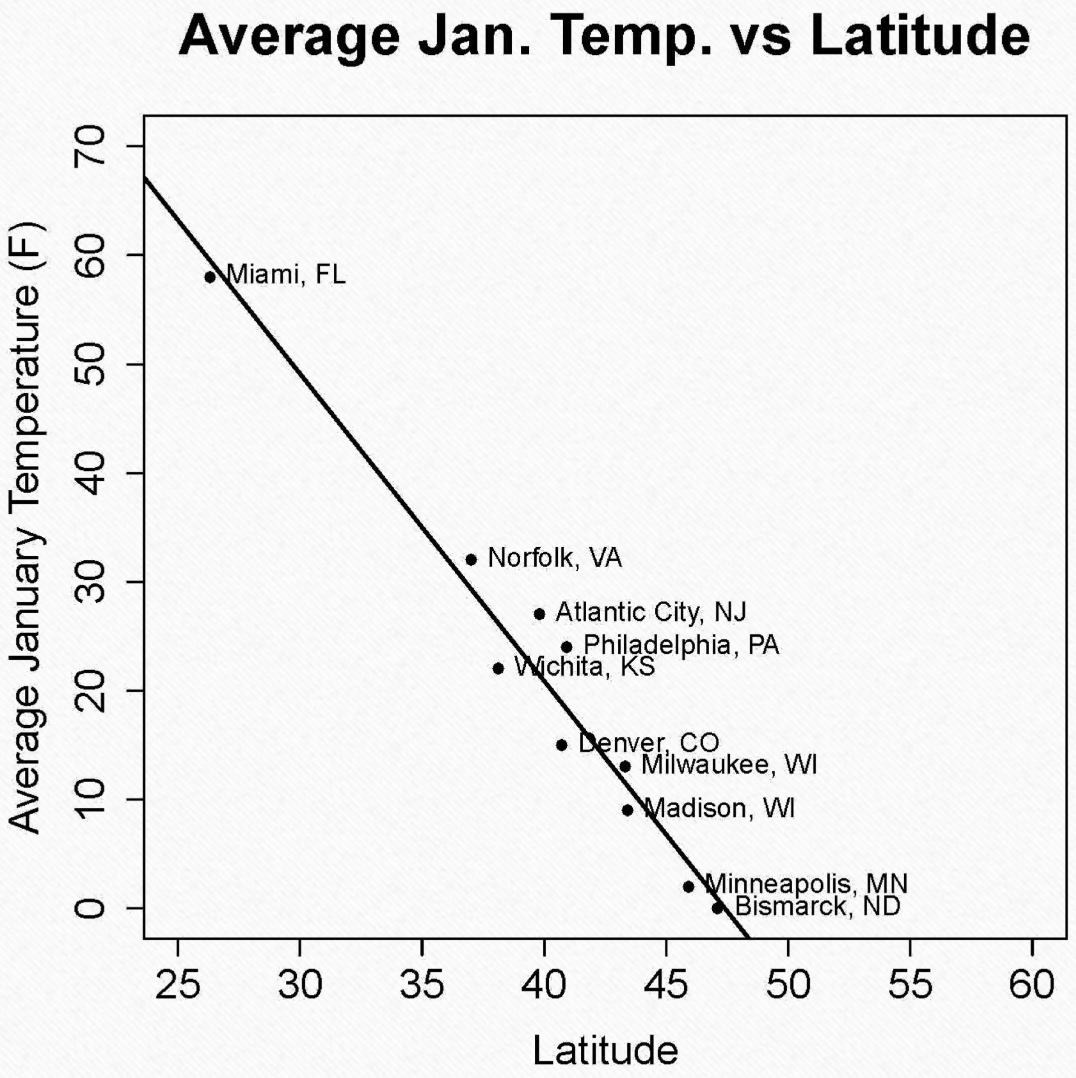
- Fact: If we repeat an experiment then we will get different estimates of the parameters because the errors will be different.
- For example:
  - If each of us repeated Hooker's (boiling temperature versus pressure) experiment then we'd all get slightly different estimates of b_0 and b_1.
- Definition: The **sampling distribution** of a parameter estimate is the distribution of the values we would get from collecting many, many data sets and computing estimates for each

STA 674, RADOE:
Fitting Simple Linear Regression Models

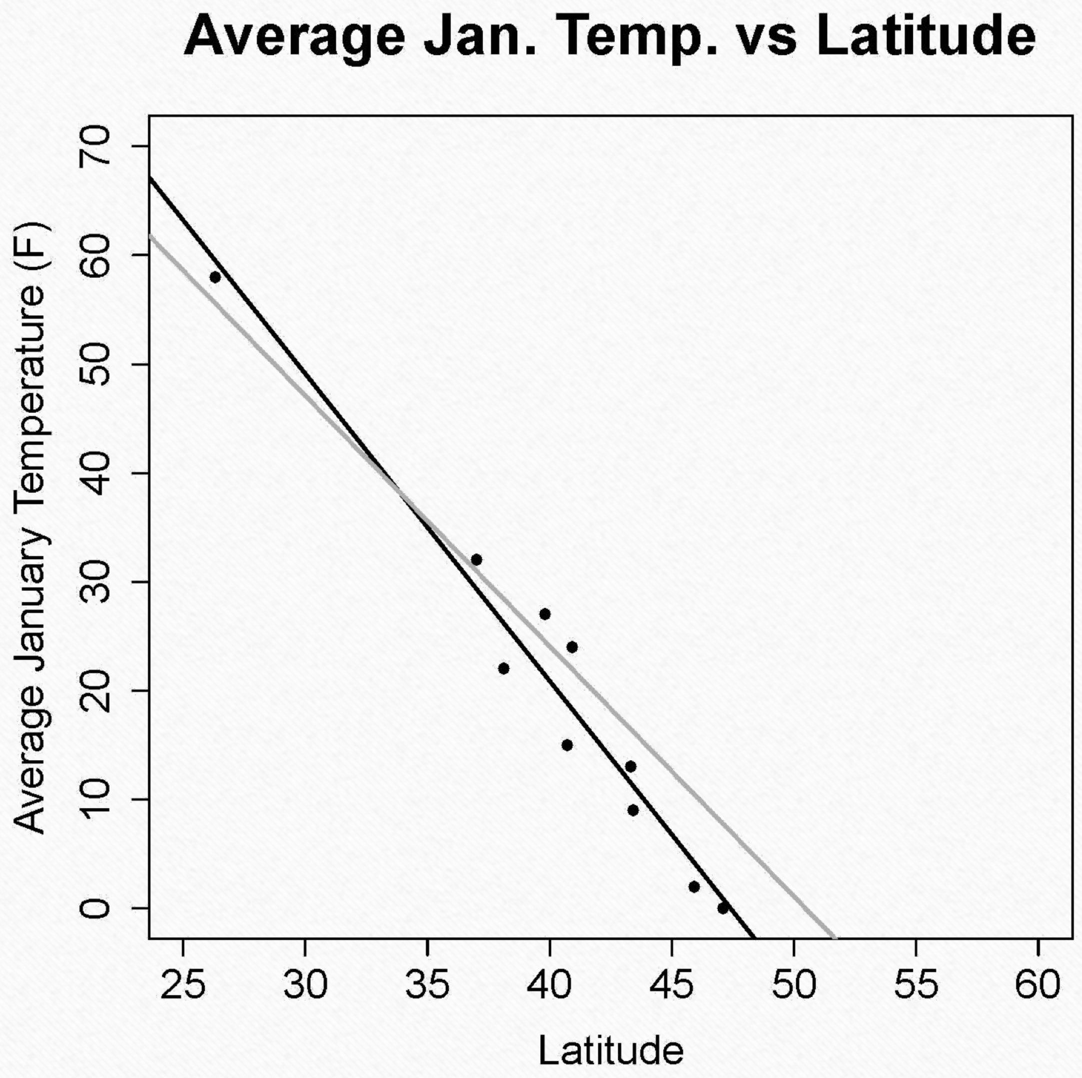
Sampling distribution – smaller example

- The original US temperature data contain 56 cities.
- The 10 that I chose formed one possible sample. There are over 3 billion possible samples of 10 cities that I could have chose.
- What would the results have been if I had chosen a different sample of 10 cities?

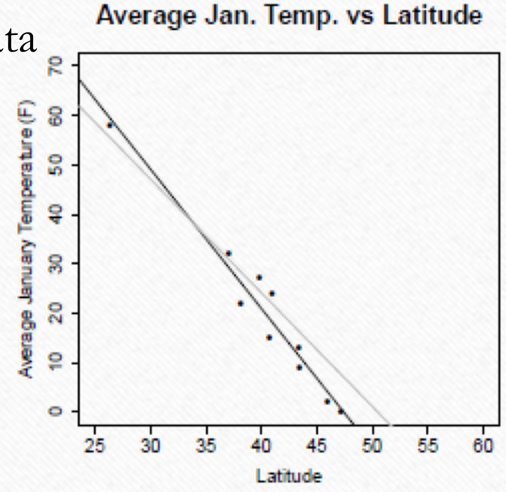
US Temperature Data

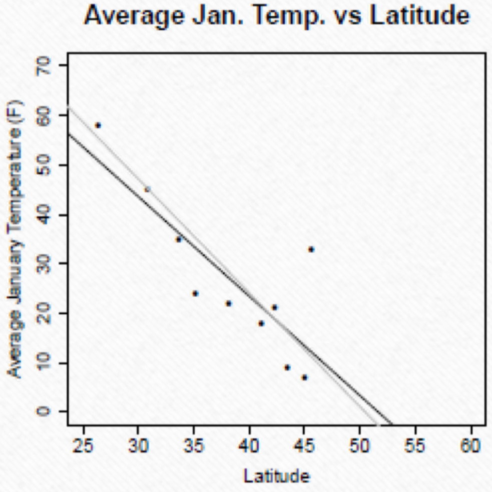


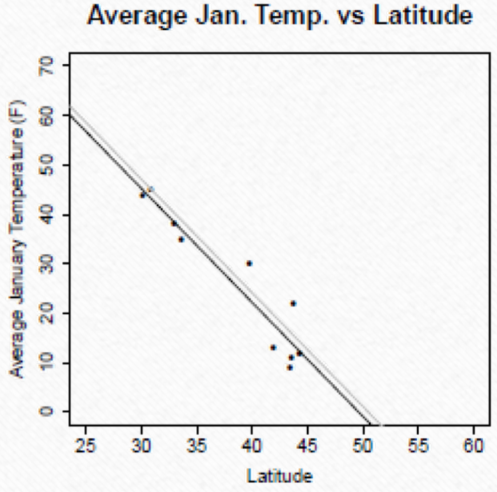
US Temperature Data

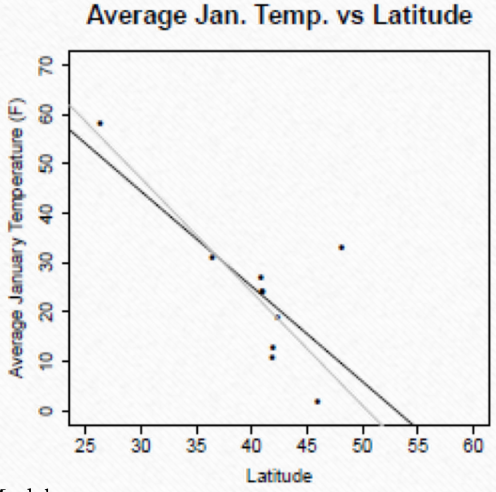


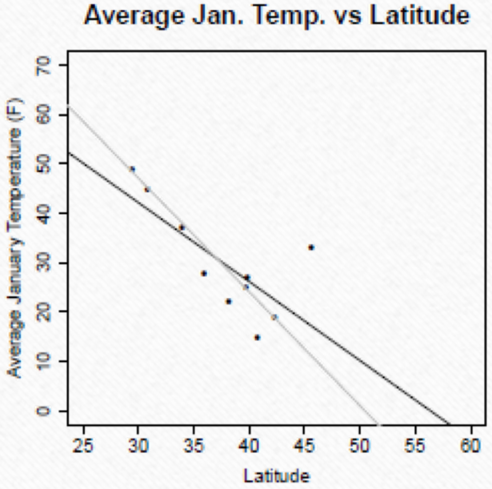
US Temperature Data

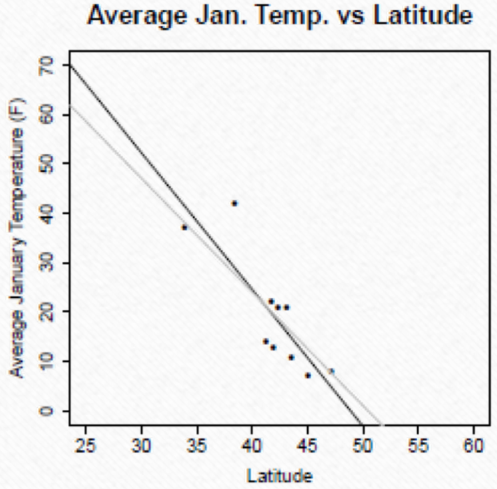




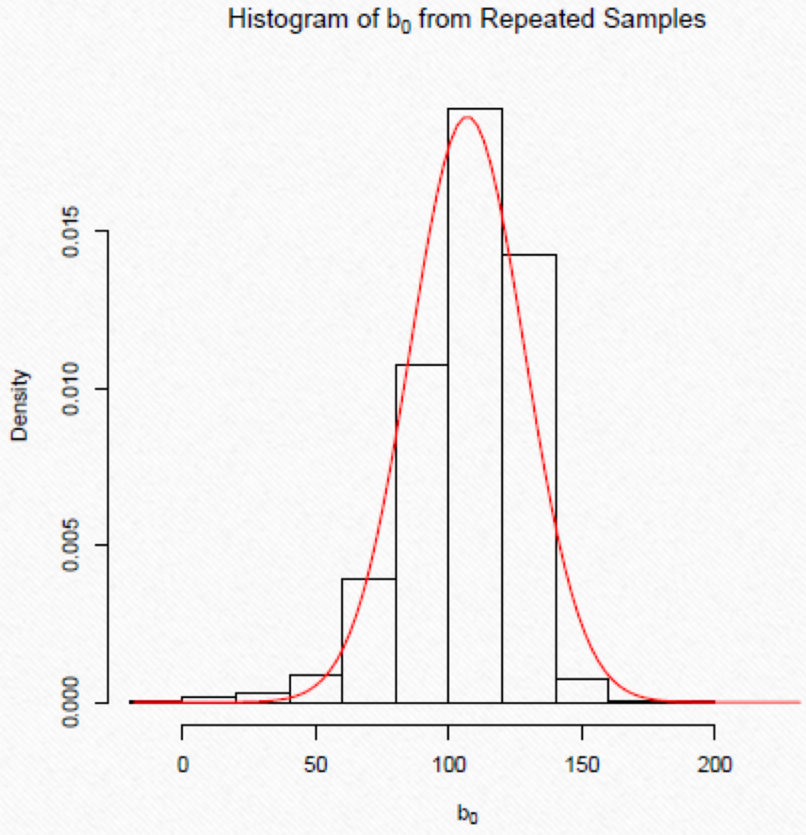


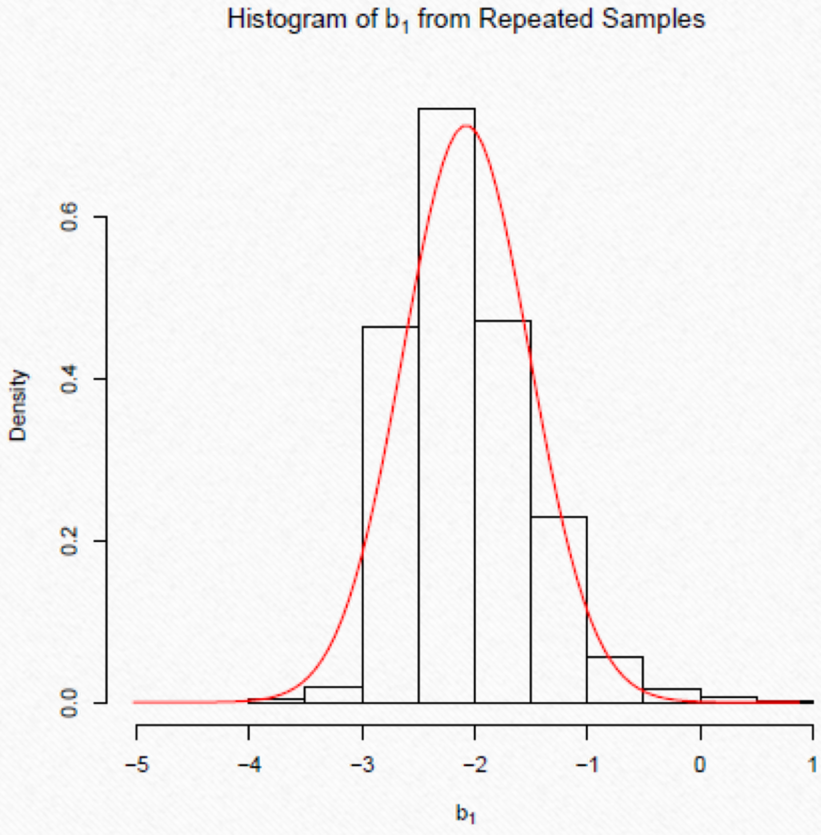






US Temperature Data:
Variation in LS estimates
for 1000 samples of 10 cities





STA 674, RADOE:

Fitting Simple Linear Regression Models

Least squares estimation – sampling distributions of the LS estimates

The assumptions of the regression model:

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

Are that:

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

IF these assumptions are satisfied **THEN** …

STA 674, RADOE:
Fitting Simple Linear Regression Models

Least squares estimation – sampling distributions of the LS estimates

The assumptions of the regression model:

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

Are that:

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

IF these assumptions are satisfied **THEN** …

$$b_0 \sim \text{Normal}\left(\beta_0, \sigma_e^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right)\right)$$

$$b_1 \sim \text{Normal}\left(\beta_1, \frac{\sigma_e^2}{(n-1)s_x^2}\right)$$