

Chapter One of *Regression Analysis*: Overview¹
STA302 Fall 2016

¹See last slide for copyright information.

Simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where

x_1, \dots, x_n are observed, known constants.

$\epsilon_1, \dots, \epsilon_n$ are random variables satisfying the *Gauss-Markov conditions*.

$$E(\epsilon_i) = 0$$

$$\text{Var}(\epsilon_i) = \sigma^2$$

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j.$$

β_0, β_1 and σ^2 are unknown constants with $\sigma^2 > 0$.

Least Squares

Background that's not in the text

- The random variable y has a distribution that depends on the parameter θ .
- How can we estimate θ from data y_1, \dots, y_n ?
- The expected value $E(y)$ is a function of θ .
- Write it $E_\theta(y)$.
- Estimate θ by the value that gets the observed data values as close as possible to their expected values.
- Minimize

$$\mathcal{S} = \sum_{i=1}^n (y_i - E_\theta(y_i))^2$$

over all θ .

- The value of θ that minimizes \mathcal{S} is the *least squares estimate*.

Least squares regression

$$\text{Minimize } \mathcal{S} = \sum_{i=1}^n (y_i - E_{\theta}(y_i))^2$$

- Model equation is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- $E(y_i) = \beta_0 + \beta_1 x_i$
- Minimize $\mathcal{S} = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ over $\theta = (\beta_0, \beta_1)$.
- Take partial derivatives, set to zero, solve two equations in two unknowns.
- Least squares estimate of β_0 is b_0 . Least squares estimate of β_1 is b_1 .

Vocabulary and concepts

A preview of almost the entire course

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

- Linear regression means linear in the β parameters.
- Polynomial regression $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$, etc.
- Centered model $y_i = (\beta_0 + \beta_1 \bar{x}) + \beta_1 (x_i - \bar{x}) + \epsilon_i$
- Predicted value $\hat{y}_i = b_0 + b_1 x_i$
- Residual $e_i = y_i - \hat{y}_i$
- Plotting residuals (p.5) to diagnose problems with the model.
- Gauss-Markov conditions.
- Measure of model fit R^2
- Mean and variance of b_0 and b_1 .
- Confidence intervals and tests.
- Predicting future observations.

Copyright Information

This slide show was prepared by **Jerry Brunner**, Department of Statistical Sciences, University of Toronto. It is licensed under a **Creative Commons Attribution - ShareAlike 3.0 Unported License**. Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/302f16>