

Stat GR5205 Lecture 3

Jingchen Liu

Department of Statistics
Columbia University

Least squares estimator for simple linear regression

- Slope

$$\hat{\beta}_1 = \rho_{x,y} \frac{s_y}{s_x}.$$

- The fitted regression line

$$(x - \bar{x}) = \rho_{x,y} \frac{s_y}{s_x} (y - \bar{y})$$

Least squares estimator for simple linear regression

- Slope

$$\hat{\beta}_1 = \rho_{x,y} \frac{s_y}{s_x}.$$

- The fitted regression line

$$(x - \bar{x}) = \rho_{x,y} \frac{s_y}{s_x} (y - \bar{y})$$

The sampling distribution

► $E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1$

► The variances are

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

► $\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \text{Var}(\bar{x}\hat{\beta}_1)$. Derive $\text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$

► Normality: simulation illustration

The sampling distribution

- ▶ $E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1$

- ▶ The variances are

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

- ▶ $\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \text{Var}(\bar{x}\hat{\beta}_1)$. Derive $\text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$

- ▶ Normality: simulation illustration

The sampling distribution

► $E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1$

► The variances are

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

► $\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \text{Var}(\bar{x}\hat{\beta}_1)$. Derive $\text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$

► Normality: simulation illustration

The sampling distribution

- ▶ $E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1$

- ▶ The variances are

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

- ▶ $\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \text{Var}(\bar{x}\hat{\beta}_1)$. Derive $\text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$

- ▶ Normality: simulation illustration

Confidence interval

- ▶ About confidence interval
- ▶ A confidence interval of β_0 and β_1 .

$$[\beta_i - 1.96 \times SD(\beta_i), \quad \beta_i + 1.96 \times SD(\beta_i)]$$

Confidence interval

- ▶ What if σ^2 unknown?

Inference with unknown σ^2

- ▶ The noise level

$$\sigma^2 = \text{Var}(\varepsilon_i) = E(y_i - \beta_0 - \beta_1 x_i)^2$$

- ▶ Point estimate

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Inference with unknown σ^2

- ▶ The noise level

$$\sigma^2 = \text{Var}(\varepsilon_i) = E(y_i - \beta_0 - \beta_1 x_i)^2$$

- ▶ Point estimate

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

About normal (Gaussian) distribution

- ▶ Z_1, \dots, Z_n are independent and identically distributed $N(0, 1)$
- ▶ The χ^2 distribution with degrees of freedom n

$$Z_1^2 + \dots + Z_n^2$$

Inference with unknown σ^2

- Distribution of $\hat{\sigma}^2$

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$$

Derive sample $(n-1)s_x^2/\sigma^2 \sim \chi_{n-1}^2$ and \bar{x} is independent of s_x^2 .

- About χ^2 -distribution
- $\hat{\sigma}^2$ is independent of $(\hat{\beta}_0, \hat{\beta}_1)$.
- Confidence interval of $\hat{\beta}_i$ when σ^2 is unknown.

Inference with unknown σ^2

- Distribution of $\hat{\sigma}^2$

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$$

Derive sample $(n-1)s_x^2/\sigma^2 \sim \chi_{n-1}^2$ and \bar{x} is independent of s_x^2 .

- About χ^2 -distribution
- $\hat{\sigma}^2$ is independent of $(\hat{\beta}_0, \hat{\beta}_1)$.
- Confidence interval of $\hat{\beta}_i$ when σ^2 is unknown.

Inference with unknown σ^2

- Distribution of $\hat{\sigma}^2$

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$$

Derive sample $(n-1)s_x^2/\sigma^2 \sim \chi_{n-1}^2$ and \bar{x} is independent of s_x^2 .

- About χ^2 -distribution
- $\hat{\sigma}^2$ is independent of $(\hat{\beta}_0, \hat{\beta}_1)$.
- Confidence interval of $\hat{\beta}_i$ when σ^2 is unknown.

Inference with unknown σ^2

- Distribution of $\hat{\sigma}^2$

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$$

Derive sample $(n-1)s_x^2/\sigma^2 \sim \chi_{n-1}^2$ and \bar{x} is independent of s_x^2 .

- About χ^2 -distribution
- $\hat{\sigma}^2$ is independent of $(\hat{\beta}_0, \hat{\beta}_1)$.
- Confidence interval of $\hat{\beta}_i$ when σ^2 is unknown.

On the Student's t -distribution

- ▶ $Z \sim N(0, 1)$ and $\gamma^2 \sim \frac{\chi_v^2}{v}$
- ▶ $T = \frac{Z}{\gamma} \sim t_v$
- ▶ About t -distribution

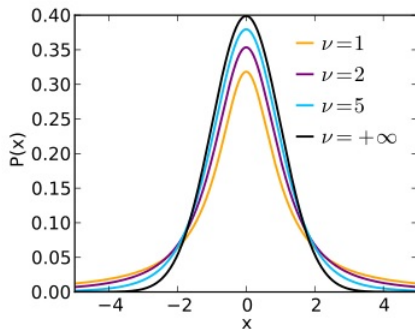
On the Student's t -distribution

- ▶ $Z \sim N(0, 1)$ and $\gamma^2 \sim \frac{\chi_v^2}{v}$
- ▶ $T = \frac{Z}{\gamma} \sim t_v$
- ▶ About t -distribution

On the Student's t -distribution

- ▶ $Z \sim N(0, 1)$ and $\gamma^2 \sim \frac{\chi_v^2}{v}$
- ▶ $T = \frac{Z}{\gamma} \sim t_v$
- ▶ About t -distribution

On the Student's t -distribution



On the Student's t -distribution

- ▶ Normal approximation of t -distribution
- ▶ Rationale

On the Student's t -distribution

- ▶ Normal approximation of t -distribution
- ▶ Rationale

Prediction

- ▶ $\hat{\mu}_x \triangleq \widehat{E(y|x)} = \hat{\beta}_0 + \hat{\beta}_1 x$
- ▶ The sampling distribution of $\hat{\mu}_x$
- ▶ Confidence interval

Prediction

- ▶ $\hat{\mu}_x \triangleq \widehat{E(y|x)} = \hat{\beta}_0 + \hat{\beta}_1 x$
- ▶ The sampling distribution of $\hat{\mu}_x$
- ▶ Confidence interval

Prediction

- ▶ $\hat{\mu}_x \triangleq \widehat{E(y|x)} = \hat{\beta}_0 + \hat{\beta}_1 x$
- ▶ The sampling distribution of $\hat{\mu}_x$
- ▶ Confidence interval

Prediction of a new observation

- ▶ Prediction of a new observation when parameters are known
 - ▶ When σ^2 is known
 - ▶ When σ^2 is unknown

Prediction of a new observation

- ▶ Prediction of a new observation when parameters are known
- ▶ When σ^2 is known
- ▶ When σ^2 is unknown

Prediction of a new observation

- ▶ Prediction of a new observation when parameters are known
- ▶ When σ^2 is known
- ▶ When σ^2 is unknown

Simultaneous confidence band

- ▶ Simultaneous confidence band versus usual confidence interval
- ▶ The confidence band

$$\hat{\mu}_x \pm \lambda SD(\hat{\mu}_x)$$

where

$$\lambda^2 = 2F(1 - \alpha; 2, n - 2)$$

Simultaneous confidence band

- ▶ Simultaneous confidence band versus usual confidence interval
- ▶ The confidence band

$$\hat{\mu}_x \pm \lambda SD(\hat{\mu}_x)$$

where

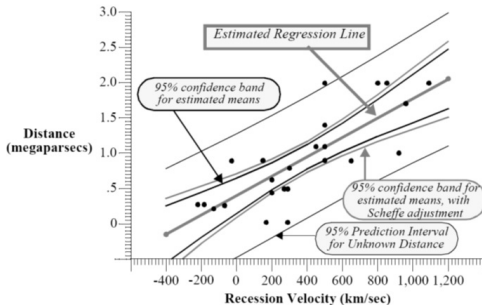
$$\lambda^2 = 2F(1 - \alpha; 2, n - 2)$$

Comparing the intervals

Display 7.11

p. 189

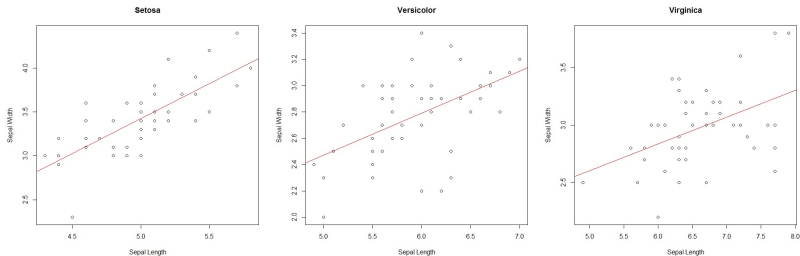
The 95% confidence band on the population regression line, the 95% confidence interval band for single mean estimates, and a 95% prediction interval band for the Big Bang example



Some descriptive statistics – the Iris data



Some descriptive statistics – the Iris data



Some descriptive statistics

- ▶ The coefficient of determination (a.k.a. the R^2)

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

- ▶ Another representation for the simple linear regression

$$R^2 = \rho^2$$

- ▶ Derive analysis of variance of simple linear regression

Some descriptive statistics

- ▶ The coefficient of determination (a.k.a. the R^2)

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

- ▶ Another representation for the simple linear regression

$$R^2 = \rho^2$$

- ▶ Derive analysis of variance of simple linear regression

Some descriptive statistics

- ▶ The coefficient of determination (a.k.a. the R^2)

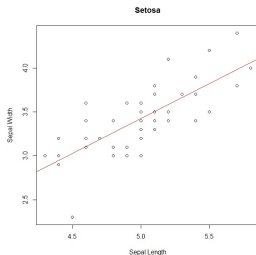
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- ▶ Another representation for the simple linear regression

$$R^2 = \rho^2$$

- ▶ Derive analysis of variance of simple linear regression

Hypothesis testing – the Iris data



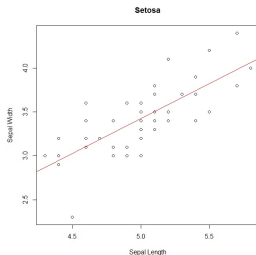
- The fitted regression

$$y = -0.57 + 0.80x$$

and $\hat{\sigma} = 0.26$.

- Question: whether plants with longer sepal tend to have wider sepal?

Hypothesis testing – the Iris data



- The fitted regression

$$y = -0.57 + 0.80x$$

and $\hat{\sigma} = 0.26$.

- Question: whether plants with longer sepal tend to have wider sepal?

Hypothesis testing

- ▶ Testing of hypotheses

- ▶ Basic setting

1. Two families of models without overlap: the null hypothesis (H_0) and the alternative hypothesis (H_1)
2. Which family were the data sampled from?

Hypothesis testing

- ▶ Testing of hypotheses
- ▶ Basic setting
 1. Two families of models without overlap: the null hypothesis (H_0) and the alternative hypothesis (H_1)
 2. Which family were the data sampled from?

Hypothesis testing

- ▶ Testing of hypotheses
- ▶ Basic setting
 1. Two families of models without overlap: the null hypothesis (H_0) and the alternative hypothesis (H_1)
 2. Which family were the data sampled from?

Hypothesis testing

- ▶ Testing of hypotheses
- ▶ Basic setting
 1. Two families of models without overlap: the null hypothesis (H_0) and the alternative hypothesis (H_1)
 2. Which family were the data sampled from?

Hypothesis testing

- ▶ The decision: reject the null or do not reject the null
- ▶ Two types of errors
 - ▶ Type I error
 - ▶ Type II error

Hypothesis testing

- ▶ The decision: reject the null or do not reject the null
- ▶ Two types of errors
 - ▶ Type I error
 - ▶ Type II error

Hypothesis testing

- ▶ The decision: reject the null or do not reject the null
- ▶ Two types of errors
 - ▶ Type I error
 - ▶ Type II error

Hypothesis testing

- ▶ The decision: reject the null or do not reject the null
- ▶ Two types of errors
 - ▶ Type I error
 - ▶ Type II error

Hypothesis testing

- ▶ The rationale: assuming the null hypothesis and trying to reach contradiction
- ▶ Rejection region
- ▶ Test statistic: differentiating the null and the alternative

Hypothesis testing

- ▶ The rationale: assuming the null hypothesis and trying to reach contradiction
- ▶ Rejection region
- ▶ Test statistic: differentiating the null and the alternative

Hypothesis testing

- ▶ The rationale: assuming the null hypothesis and trying to reach contradiction
- ▶ Rejection region
- ▶ Test statistic: differentiating the null and the alternative

Hypothesis testing

- ▶ Formulation

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0$$

- ▶ $\hat{\beta}_1 = 0.80$ and $SD(\hat{\beta}_1) = 0.10$.
- ▶ The sampling distribution of β_1

Hypothesis testing

- ▶ Formulation

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0$$

- ▶ $\hat{\beta}_1 = 0.80$ and $SD(\hat{\beta}_1) = 0.10$.
- ▶ The sampling distribution of β_1

Hypothesis testing

- ▶ Formulation

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0$$

- ▶ $\hat{\beta}_1 = 0.80$ and $SD(\hat{\beta}_1) = 0.10$.
- ▶ The sampling distribution of β_1

Hypothesis testing – reference distribution

- ▶ Reference distribution
- ▶ The Z -test
- ▶ The t -test
- ▶ The p -value

Hypothesis testing – reference distribution

- ▶ Reference distribution
- ▶ The Z -test
- ▶ The t -test
- ▶ The p -value

Hypothesis testing – reference distribution

- ▶ Reference distribution
- ▶ The Z -test
- ▶ The t -test
- ▶ The p -value

Hypothesis testing – reference distribution

- ▶ Reference distribution
- ▶ The Z -test
- ▶ The t -test
- ▶ The p -value

Robustness

- ▶ About robustness: appropriateness in the absence of the assumptions
- ▶ Linearity
- ▶ Independence and equal variance
- ▶ Normality

Robustness

- ▶ About robustness: appropriateness in the absence of the assumptions
- ▶ Linearity
- ▶ Independence and equal variance
- ▶ Normality

Robustness

- ▶ About robustness: appropriateness in the absence of the assumptions
- ▶ Linearity
- ▶ Independence and equal variance
- ▶ Normality

Robustness

- ▶ About robustness: appropriateness in the absence of the assumptions
- ▶ Linearity
- ▶ Independence and equal variance
- ▶ Normality

Some hypothetical examples

