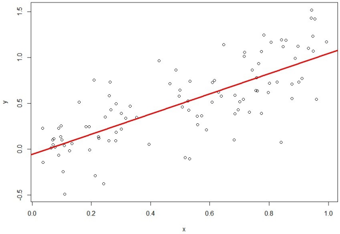
Stat GR5205 Lecture 2

Jingchen Liu

Department of Statistics
Columbia University

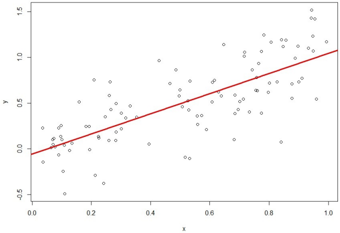
Least squares estimator



- Fitting a straight line $y = \beta_0 + \beta_1 x$
- Least squares estimate

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Least squares estimator



- Fitting a straight line $y = \beta_0 + \beta_1 x$
- Least squares estimate

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Least squares estimator for simple linear regression

- The slope

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- 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

- The slope

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- 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

- The slope

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- 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

- The slope

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

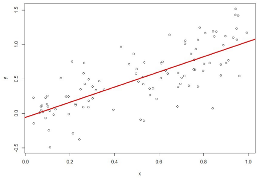
- Slope

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

- The fitted regression line

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

On the decomposition of the y



- The fitted values – predictable

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

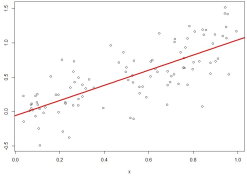
- The residuals – unpredictable

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

- The decomposition

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

On the decomposition of the y



- The fitted values – predictable

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

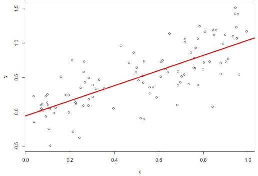
- The residuals – unpredictable

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

- The decomposition

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

On the decomposition of the y



- The fitted values – predictable

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- The residuals – unpredictable

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

- The decomposition

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

On the properties of the fitted regression line

- $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$
- $\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} \hat{y}_i \hat{\varepsilon}_i = 0$

On the properties of the fitted regression line

- $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$
- $\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} \hat{y}_i \hat{\varepsilon}_i = 0$

On the properties of the fitted regression line

- $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$
- $\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} \hat{y}_i \hat{\varepsilon}_i = 0$

On the properties of the fitted regression line

- $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$
- $\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$
- $\sum_{i=1}^{n} \hat{y}_i \hat{\varepsilon}_i = 0$

Expectation

- Expectation

$$E(X) = \int xf(x)dx$$

- Linear operator

$$E(aX + bY) = aE(x) + bE(Y)$$

- Long run average

$$\frac{X_1 + ... + X_n}{n} \to E(x) \text{ as } n \to \infty \text{ as most surely}$$

Expectation

- Expectation

$$E(X) = \int xf(x)dx$$

- Linear operator

$$E(aX + bY) = aE(x) + bE(Y)$$

- Long run average

$$\frac{X_1 + ... + X_n}{n} \to E(x) \text{ as } n \to \infty \text{ as most surely}$$

Expectation

- Expectation

$$E(X) = \int xf(x)dx$$

- Linear operator

$$E(aX + bY) = aE(x) + bE(Y)$$

- Long run average

$$\frac{X_1 + ... + X_n}{n} \to E(x) \text{ as } n \to \infty \text{ as most surely}$$

Expectation

- Variance: $Var(X) = E\{X - E(X)\}^2$
- Covariance: $Cov(X, Y) = E\{X - E(X)\}\{Y - E(Y)\}$

Expectation

- Variance: $Var(X) = E\{X - E(X)\}^2$
- Covariance: $Cov(X, Y) = E\{X - E(X)\}\{Y - E(Y)\}$

Probability model

- The model setup

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

where $E(\varepsilon_i | x_i) = 0$.

- The conditional expectation

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i.$$

- The term "regression."

Probability model

- The model setup

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

where $E(\varepsilon_i | x_i) = 0$.

- The conditional expectation

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i.$$

- The term "regression."

Probability model

- The model setup

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

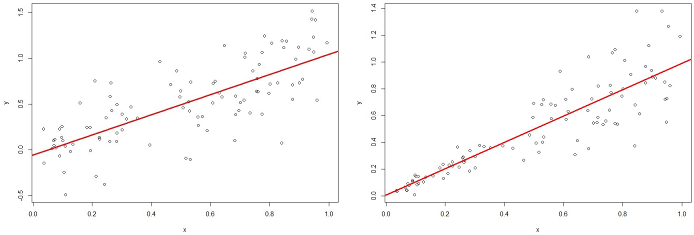
where $E(\varepsilon_i | x_i) = 0$.

- The conditional expectation

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i.$$

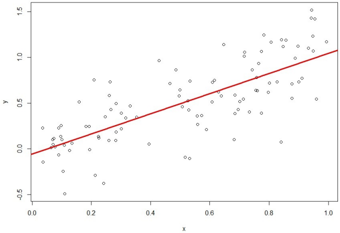
- The term "regression."

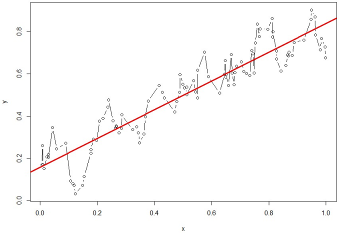
Additional assumption



- Common variance $Var(\varepsilon_i|x_i) = \sigma^2$
- Quantifying the uncertainty

Additional assumption





- Uncorrelated errors $E(\varepsilon_i \varepsilon_j | x_i, x_j) = 0$
- Quantifying the uncertainty

Scope of statistical inference

- Point estimate (Frequentist distribution)
- Interval estimate
- Hypothesis testing

Scope of statistical inference

- Point estimate (Frequentist distribution)
- Interval estimate
- Hypothesis testing

Scope of statistical inference

- Point estimate (Frequentist distribution)
- Interval estimate
- Hypothesis testing

The four assumptions of linear model

- Linearity
- Equal variance
- Independence
- Normality

The four assumptions of linear model

- Linearity
- Equal variance
- Independence
- Normality

The four assumptions of linear model

- Linearity
- Equal variance
- Independence
- Normality

The four assumptions of linear model

- Linearity
- Equal variance
- Independence
- Normality

The four assumptions of linear model

- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- ε_i's are independently and identically distributed as $N(0, \sigma^2)$

The parameters

- Regression coefficients: β_0 and β_1
- Variance: σ^2

The parameters

- Regression coefficients: β_0 and β_1
- Variance: σ^2

The sampling distribution

- Sampling distribution
- On the sampling distribution of $(\hat{\beta}_0, \hat{\beta}_1)$.

Additional assumption

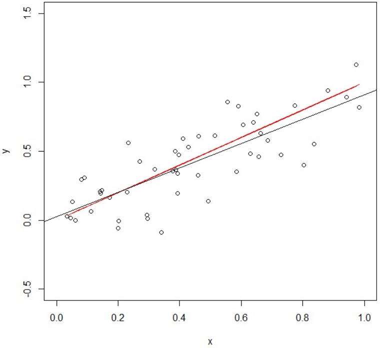


Figure: $y = x + N(0, 0.04)$, $n = 50$, data 1

Additional assumption

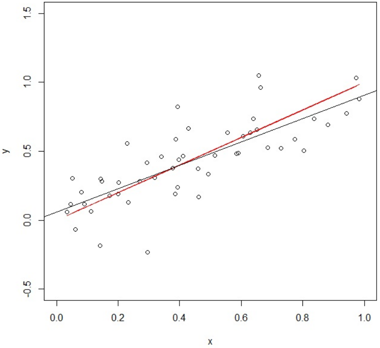


Figure: $y = x + N(0, 0.04)$, $n = 50$, data 2

Additional assumption

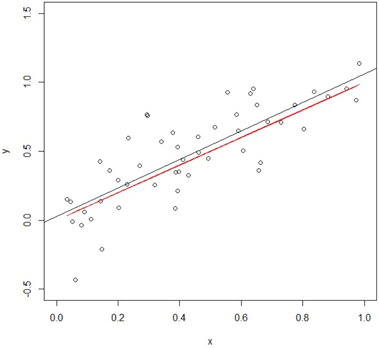


Figure: $y = x + N(0, 0.04)$, $n = 50$, data 3

Additional assumption

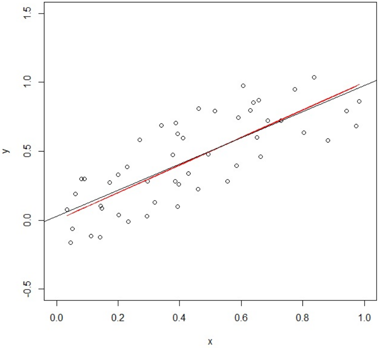


Figure: $y = x + N(0, 0.04)$, $n = 50$, data 4

Additional assumption

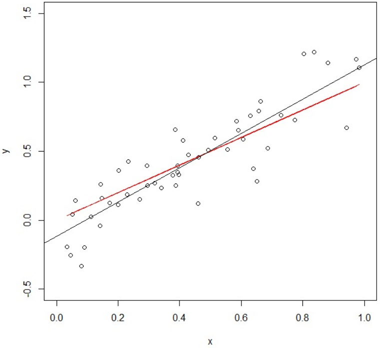


Figure: $y = x + N(0, 0.04)$, $n = 50$, data 5

Frequentist distribution

- The slope:

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Frequentist distribution

- The slope:

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

- The intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Expectation

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

Probabilistic properties of the least squares estimate

- A note on variance calculation
- The variances are

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

Probabilistic properties of the least squares estimate

- A note on variance calculation
- The variances are

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

About normal distribution

- Standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

About normal distribution

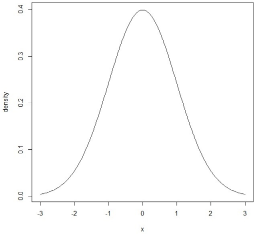


Figure: Density function of standard normal distribution

About normal (Gaussian) distribution

- The most natural distribution
- Stable distribution
- If Z_1 and Z_2 are independent normal random variables, then $Z_1 + Z_2$ is also a normal random variable.
- The distribution of $Z_1 + Z_2$ is ...

About normal (Gaussian) distribution

- $Z_1, \ldots, Z_n$ are independent normal random variables with means $\mu_1, \ldots, \mu_n$ and variances $\sigma_1^2, \ldots, \sigma_n^2$, then

$$c_1 Z_1 + \ldots + c_n Z_n \text{ is...}$$

About normal (Gaussian) distribution

- The χ^2 distribution with degrees of freedom n

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

Confidence interval

- About confidence interval
- The confidence interval of β_0 and β_1.

Inference with unknown σ^2

- The noise level

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

- Point estimate

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

- Distribution of σ^2

Inference with unknown σ^2

- The noise level

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

- Point estimate

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

- Distribution of σ^2

Inference with unknown σ^2

- The noise level

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

- Point estimate

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

- Distribution of σ^2

Confidence interval

- The confidence interval of β_0 and β_1 with σ^2 unknown.