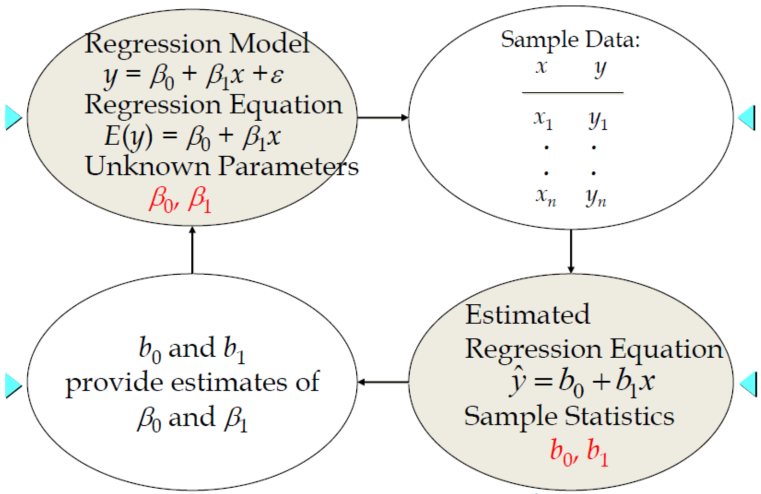
STAC67: Regression Analysis

Lecture 3

Sohee Kang

Jan. 20, 2021

Parameter Estimation



$\hat{\beta}_0, \hat{\beta}_1$

- Maximum Likelihood Estimation For $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$

$$f(y; \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

- Likelihood Function:

$$L(\beta_0, \beta_1, \sigma^2) = \Pi_{i=1}^{n} f(y_i; \beta_0, \beta_1, \sigma^2)$$

Parameter Estimation

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\sum_{i=1}^{n} -\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

Log-likelihood Function:

$$\ln L(\beta_0, \beta_1, \sigma^2) = K - n \ln \sigma - \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}$$

- Maximizing Likelihood function w.r.t β_0, β_1 is equivalent to minimizing,

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

.

Find m.l.e for β_0, β_1, and σ^2. Homework ☆

$$Q = \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i)^2$$

$$\frac{dQ}{dB_0} = -2 \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i) = 0$$

$$nB_0 = \sum Y_i - B_1 \sum X_i$$

$$\hat{B}_0 = \bar{Y} - B_1 \bar{X} \quad \underline{MLE}$$

$$\frac{dQ}{dB_1} = -2 \sum (Y_i - B_0 - B_1 X_i) X_i = 0$$

$$\sum Y_i X_i - (\bar{Y} - B_1 \bar{X}) X_i - B_1 X_i^2 = 0$$

$$\sum Y_i X_i - \bar{Y} X_i + B_1 \bar{X} X_i - B_1 X_i^2 = 0$$

$$B_1 = \frac{\sum Y_i X_i - \bar{Y} X_i}{\sum X_i^2 - X_i \bar{X}}$$

$$= \frac{\sum (Y_i - \bar{Y}) X_i}{\sum (X_i - \bar{X}) X_i}$$

$$= \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{\sum (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}} //$$

$$\ln L(B_0, B_1, \sigma^2)$$

$$= K - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i)^2$$

$$= S(B_0, B_1, \sigma^2$$

$$\frac{dS}{} = \frac{-n}{} + \frac{1}{} \sum^{n} (Y_i - B_0 - B_1 X_i)^2$$

$\partial \sigma \qquad \sigma \qquad \sigma^3 \quad i=1$

$= 0$

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$= \frac{\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n}$$

Parameter Estimation

- Method of Least Squres * Doesn't need assumption of distribution to optimize

Simple linear model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- **Goal**: Find the best estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$ given the data.

- What does it mean, "best"?
- Least squares: best by criterion

same as mle method.

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

- Find the $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimizes the criterion Q.
  1. Write the normal equations (derivatives of Q set to 0)
  2. Find the solution of the normal equations

Least Squre Estimators

① $\dfrac{dQ}{dB_0} = -2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i) = 0$

$nB_0 = \sum_{i=1}^{n} y_i - B_1 \sum_{i=1}^{n} x_i \Rightarrow \hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}$

② $\dfrac{dQ}{dB_1} = -2\sum(y_i - B_0 - B_1 x_i)x_i = 0$

$\xrightarrow{\hat{B}_0}$ $\hat{B}_1 = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

$= \dfrac{S_{xy}}{S_{xx}}$

Thus the estimated regression line is $\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$

Least Square Estimators

- Other Criteria

Why square the residuals?

we could use least absolute deviations estimates, minimizing

$$Q_1(\beta_0, \beta_1) = \sum_{i=1}^{n} |(y_i - \beta_0 - \beta_1 x_i)|$$

We have to use linear ~~pro~~ programming to find the least absolute deviations estimations

- **Convenience**
- **Optimality**

Least square estimates are BLVE

Best Linear Unbiased Estimator

BLUE: ε_i are uncorrelated with constant variance.

Gauss-Markov Theorem

- **Theorem 1**

Consider the simple linear model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

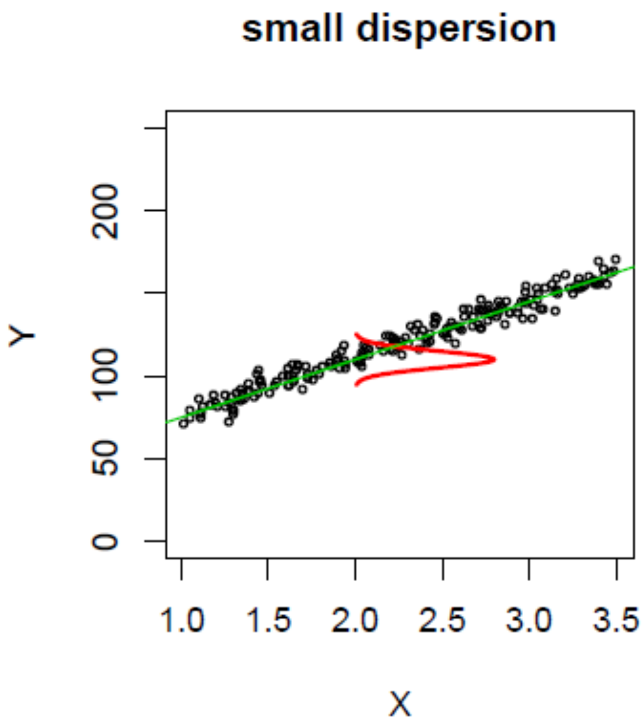
Suppose that the following assumptions (called Gauss-Markov asumptions) conerning the random erros are safisfied:

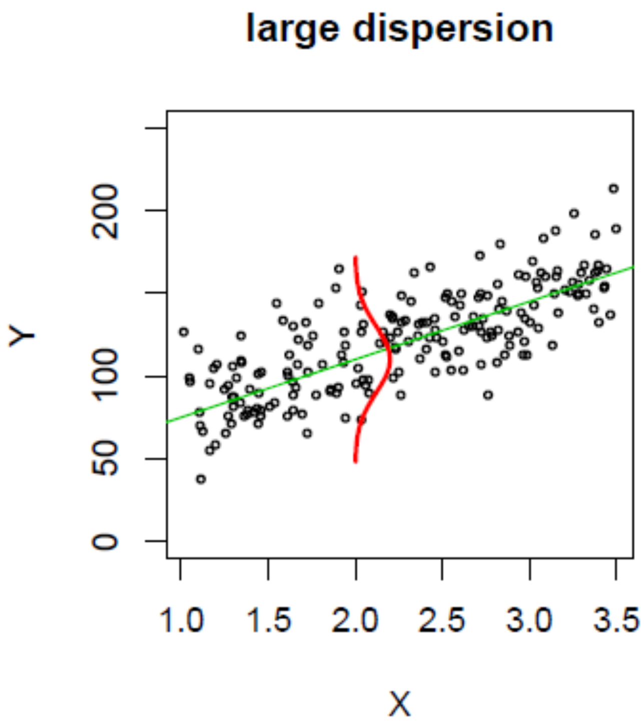
- Mean zero: $E(\epsilon_i) = 0$
- Constant variance: $Var(\epsilon_i) = \sigma^2$
- Uncorrelated: $Cov(\epsilon_i, \epsilon_j) = 0, \quad i \neq j$

Then the least squre estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased and have minimum variance among all unbiased linear estimators (BLUE).

The interpretation of σ^2

- The variance, σ^2 controls the disperson of Y around $\beta_0 + \beta_1 X$





Fitted values and Residuals

- **Regression equation** or **fitted regression line**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

, where $\hat{Y}$ is the estimated mean of the response variable at level X of the explanatory.

- For each observation, we can compute the fitted value:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, \ldots, n$$

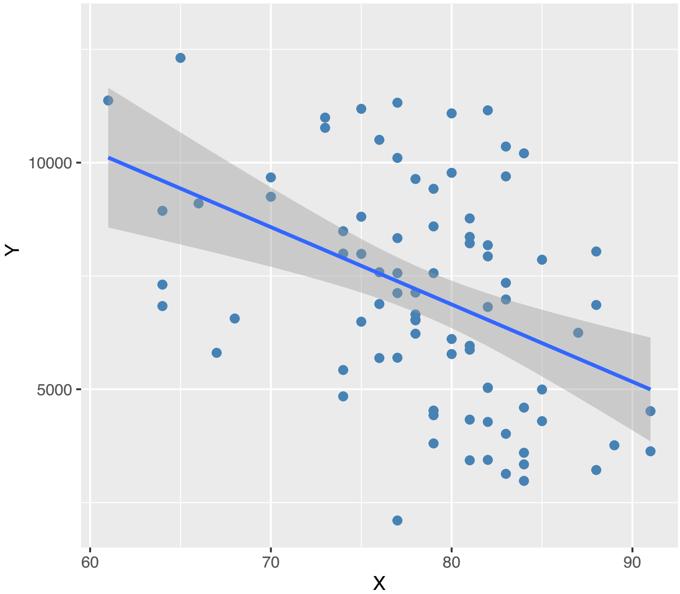
- The vertical distance from the observed y_i to the fitted value is called: residual

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i), \quad i = 1, \ldots, n$$

The residuals can be thought of as predicted (observed) valus of the unknown error, $\epsilon_1, \ldots, \epsilon_n$.

Example: Crime rate

A criminologists studying the relationship between level of education and crime rate in medium-sized U.S. counties collected from a random sample of 84 counties; X is the percentage of individuals having at least a high-school diploma, and Y is the crime rate (crimes reported per 100,000 residents) last year.



Example (Continued)

```
attach(Crime)
fit = lm(Y~X, data=Crime)
coefficients(fit)
```

```
## (Intercept)           X
##  20517.5999   -170.5752
```

```
c(mean(Y), mean(X))
```

```
## [1] 7111.20238   78.59524
```

```
c(sum((Y-mean(Y))*(X- mean(X))), sum((X-mean(X))^2))
```

```
## [1] -547928.119    3212.238
```

```
c(Y[10], X[10])
```

```
## [1] 7932   82
```

Exercise

1. Let's compute the estimated regression coefficients by hands:

   b_0 : $= \bar{Y} + b_1 \bar{X} = 7111.20238 - 170.58 \cdot 78.59524 \approx 20517.6$

   b_1 : $\dfrac{Sxy}{Sxx} = \dfrac{-547928.199}{3212.238} \approx -170.58$

2. Obtain the point estiamtes of the following:
   - the difference in the mean crime rate for two counties whose high-school graduation rates differ by one percentage points.

     -170.58

   - the mean crime rate last year in counties with high school graduation percentage $X = 80$.

     $E(Y|X=80) = -170.58 \cdot 80 + 20517.6$

   - $\epsilon_{10} = e_{10} = 7932 - (20517.6 - 17058 \cdot 82)$

Properties of the fitted regression line

$E(\hat{Y} \mid x = \bar{x}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \beta_1 \bar{x} + \beta_1 \bar{x} = \bar{Y}$

- The least squre line always passes through the point $(\bar{X}, \bar{Y})$.
- The estimated slope $\hat{\beta}_1$ always has the same sign as the sample correlation between X and Y. $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = r \cdot \frac{s_y}{s_x}$
- the sum of residuals are 0.

  (i) $\sum e_i = 0$

  $\sum Y_i - (\beta_0 + \beta_1 X_i) = 0$

  $\Rightarrow \sum Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) = \sum e_i = 0$

  (ii) $\sum e_i X_i = 0$ $\quad \sum (Y_i - (\beta_0 + \beta_1 X_i)) X_i = 0$

  $\Rightarrow \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) X_i = \sum e_i X_i = 0$

- The sum of squares of the e_i's is called: Residual sum of Squres or Sum of Squared Errors (SSE). $SSE = \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2 = \sum e_i^2$

Properties of the fitted regression line

- An unbiased estimator of σ^2 is:

Accept : $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{SSE}{n-2} = MSE$

- Result:

① $\sum_{i=1}^{n} (x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0$

② $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \qquad \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0$

③ $\sum_{i=1}^{n} e_i \hat{y}_i = 0 \qquad \sum_{i=1}^{n} e_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 \sum_{i=1}^{n} e_i + \hat{\beta}_1 \sum_{i=1}^{n} e_i x_i = 0$

Properties of Least Squares Estiamtes

- $$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{\sum(X_i - \bar{X})Y_i}{\sum(X_i - \bar{X})^2}$$

- The constants k_i have seveal interesting properties:
  1. $\sum_{i=1}^{n} k_i =$
  2. $\sum_{i=1}^{n} k_i X_i = \frac{\sum(X_i - \bar{X})X_i}{S_{xx}}$
  3. $\sum_{i=1}^{n} k_i^2 =$