Research Methods in Political Science I

7. Linear Regression (2)

Yuki Yanai

School of Law and Graduate School of Law

November 18, 2015

KOBE UNIVERSITY

Today's Menu

Expression with Vectors and Matrices

- y_i: the response of the i-th observation
- n: the sample size (the number of observations: $i = 1, 2, \dots, n$)
- Linear predictor: $X_i\beta = \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$
- k: the number of predictors including the constant
- X: the predictor matrix
- X_i: the i-th row of X
- Constant term: $X_{i1} = 1$ for all i
- β: the coefficient vector
- Error: $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$
- All vectors are column vectors unless specified otherwise
- a^T: a row vector, which is the transpose of a column vector a

Representation of Linear Regression Models

- Representation 1

$$\begin{aligned} y_i &= X_i\beta + \varepsilon_i \\ &= \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i, \ \text{for} \ i = 1, 2, \ldots, n \end{aligned}$$

- Representation 2

$$\begin{aligned} y_i &\sim \mathrm{N}(X_i\beta, \sigma^2), \ \text{for} \ i = 1, 2, \ldots, n \\ &\text{or} \\ y &\sim \mathrm{N}(X\beta, \sigma^2 I) \end{aligned}$$

We estimate $\hat{\beta}$ and $\hat{\sigma}$ by some statistical methods

Decomposing the Linear Regression Models

- Decompose the model into two parts
  1. Stochastic component (Normal):
  $$y_i \sim \mathrm{N}(\mu_i, \sigma^2)$$
  2. Systematic component (Linear):
  $$\mu_i = X_i \beta$$
- Hence, the models are called **Normal-Linear** Models

Estimating the Parameters of Linear Regression

- Use the function `lm()`
- To view the estimation results
  1. Use the function `summarize()`
  2. Or use `arm::display()`
- See the course website for more information

Ordinary Least Square Estimates (OLSE) of β (1)

- Linear regression model: $y = X\beta + \varepsilon, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$
- Find β that minimizes the error: find β that minimizes the square sum of the error, which is

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - X_i\beta)^2.$$

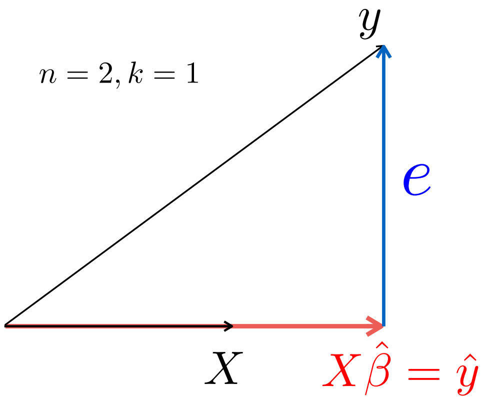
- But we can't observe β because it is a parameter. So we try to find $\hat{\beta}$ that minimizes the square sum of the residuals, which is

$$\sum_{i=1}^{n} e_i^2 = e^T e = (y - X\hat{\beta})^T (y - X\hat{\beta}).$$

$\rightarrow$ We wanna find $\hat{\beta}$ that minimizes the distance between y and $X\hat{\beta}$.

OLSE of β (2)

- Residual vector: $e = y - X\hat{\beta} = y - \hat{y}$
- Distance between y and $X\hat{\beta}$ is shortest when X and e are orthogonal to each other



OLSE of β (3)

- Orthogonality condition: $X_m^T e = 0$ for all m, $\quad m = 1, 2, \ldots, k$
  Thus,

$$
\begin{aligned}
X^T e = \mathbf{0} \quad &\Longleftrightarrow \quad X^T(y - X\hat{\beta}) = \mathbf{0} \\
&\Longleftrightarrow \quad X^T X\hat{\beta} = X^T y \qquad \text{Normal Equation} \\
&\Longleftrightarrow \quad (X^T X)^{-1}(X^T X)\hat{\beta} = (X^T X)^{-1} X^T y \\
&\Longleftrightarrow \quad I\hat{\beta} = (X^T X)^{-1} X^T y \\
&\Longleftrightarrow \quad \hat{\beta} = (X^T X)^{-1} X^T y
\end{aligned}
$$

- $\hat{\beta}$ above is the OLSE of β if $(X^T X)^{-1}$ exists
- OLSE of β is a linear function of y

Standard Errors of the Coefficients (1)

- Standard error (SE) shows the uncertainty of the estimate
- Used to calculate the confidence intervals of the estimate
- In R, shown as coef.se by `arm::display()`
- Values in $\hat{\beta} \pm 2\text{se}$ are thought to be consistent with the data
- The statistic $\frac{\hat{\beta}-\beta}{\text{se}}$ follows the t distribution with $n-k$ degrees of freedom (Approximately normal if n is large enough, but it is always easy to use the t distribution in R)

SEs of the Coefficients (2)

- Uncertainty of the estimates correlates
- The variance-covariance matrix of the estimated coefficients is $V_\beta \hat{\sigma}^2$, where $V_\beta = (X^T X)^{-1}$
  - Diagonal elements: variance of each coefficient estimate
  - Off-diagonal elements: covariance of corresponding two coefficient estimates

Statistical Significance

Setting the significance level at 0.05 (5%)

- If the range $\hat{\beta}_i \pm 2\mathrm{se}$ doesn't have 0 inside, the direction the predictor's effect is thought to be clear (negative or positive)
- Then, the effect is thought to be "**statistically** significant"
- Statistical significance tells nothing about the size of the effect
- In applied works (including political science research), we must present the **effect size** or ***substantive* significance**
- No theoretical justification for the significance level of 0.05
- "$p < 0.05$" does **not** mean a good result

p values (1)

- In R, displayed in the "`Pr(>|t|)`" column by `summary()`
- p is the probability that we obtain the data we have in hand **if the effect of the predictor is 0 (or if the null hypothesis is true)**
    - It is **not** the probability that the null is correct
    - It is **not** the probability that the alternative hypothesis is wrong

p values (2)

- Inference using p values

  the p value, which is calculated assuming the null to be true, is small
  $\rightarrow$ the probability of getting the data we are analyzing is low, but we did get them
  $\rightarrow$ the null hypothesis should be wrong
  $\rightarrow$ reject the null hypothesis

- p is **not** the significance level: p values are calculated with data, the significance level is arbitrarily set without data

- For Japanese speakers: p 値を「有意確率」と呼ばないで!!!

Confidence Intervals (1)

- Let $t_{\alpha,n-k}$ denote the 100α percentile of the t distribution with $n-k$ degrees of freedom
- the 95 percent confidence interval (CI) of $\hat{\beta}$ is:

$$[\hat{\beta} - t_{.025,n-k} \cdot \text{se},\ \ \hat{\beta} + t_{.975,n-k} \cdot \text{se}]$$

Confidence Intervals (3)

- 95% CI: If we infinitely repeat
  1. generating data by a specific data generation process, and
  2. analyzing the generated data and getting the 95% CI for each dataset,

  95 percent of the 95% confidence intervals contain the parameter value (the true value of β)
- "The probability that the CI contains the true value is 95%": This statement is **WRONG!**
- **The probability that any given CI contains the true value is either 0 or 1**

Let's run some simulations! (see the course website)

Next Week

Linear Regression (3)

- Assumptions of linear regression and regression diagnostics
- Variable transformation
  - Linear transformation
  - Standardization
  - Centering
  - Logarithmic transformation
- How to present the results
  - What to report
  - How to report