Research Methods in Political Science I

6. Linear Regression (1)

Yuki Yanai

School of Law and Graduate School of Law

November 11, 2015

KOBE UNIVERSITY

Today's Menu

What Is Linear Regression?

Linear regression

Method to summarize how the mean of the outcome variable (response) changes with the value defined by a linear function of explanatory variables (predictors).

What Does "Linear" Mean?

- A function $f(.)$ is **linear** if it satisfies the following conditions.

  Additivity（加法性） $f(x+y) = f(x) + f(y), \quad \forall x, \forall y$

  Homogeneity（斉次性） $f(kx) = kf(x), \quad \forall x, \forall k$

- The rate of change is constant.

- In a graph with the horizontal axis for x and the vertical for $f(x)$, a linear function is represented by a straight line.

Response and Predictor Variables

- **Response variable** (応答変数): variable to be explained

  Other names: dependent v (従属変数), outcome v (結果変数), explained v (被説明変数), regressand, etc.

- **Predictor v's** (予測変数): cause of the change in response v.

  other names: independent v (独立変数), explanatory v. (説明変数), regressor, etc.

- causal relationship between the explanatory and response variables is **assumed** in order to conduct linear regression: we can't verify it

- We regress y **(response) on** x **(predictors)**

Predictors: Explanatory and Control Variables

- General distinction
  - Explanatory v's: main causes
  - Controls: causes other than the main ones
- Statistical (or mathematical) difference: None
- Need not distinguish explanatory variables from controls

Dummy variables

Dummy variables: (usually) an indicator for a specific attribute

- Female dummy: assign 1 to people with the “female” attribute, 0 otherwise
- Male dummy: assign 1 to people with the “male” attribute, 0 otherwise
- Variable with 1 for female and 2 for male?

  We don’t tend to call it a “sex (or gender) dummy”: everybody has gender: Simply call it a “sex (or gender) variable”

Simple Regression and Multiple Regression

- Simple regression (単回帰): only one predictor (without any controls)
- Multiple regression (重回帰): two or more predictors (including controls)
- Regression: include both simple and multiple regressions

Model 1

Explain the vote share in the HR election by the experience of MP

- Response: vote share (%)
- Predictor: MP experience: 1 for the incumbent and ex-MPs, 0 otherwise
- Estimation result:

$$\text{Vote share} = 14 + 31 \cdot \text{Experience} + \text{error}$$

- Predicted values:

$$\widehat{\text{Vote share}} = 14 + 31 \cdot \text{Experience}$$

Data source: Asano and Yanai (2013)

Predicted Values and the Estimated Coefficients

- Predicted value: **the mean of the response conditional on the predictor**
- Predicted values are represented by "^" (hat)
- The predicted value of Model 1: the mean of the vote share conditional on the MP experience

$$\widehat{\text{Vote share}} = 14 + 31 \cdot \text{Experience}$$

  - without experience: $\widehat{\text{Vote share}} = 14 + 31 \cdot 0 = 14$
  - with experience: $\widehat{\text{Vote share}} = 14 + 31 \cdot 1 = 45$
- Estimated coefficient: $31 = 45 - 14$: the difference in the mean vote share between the candidates with experience and those without

Visual Display of Model 1

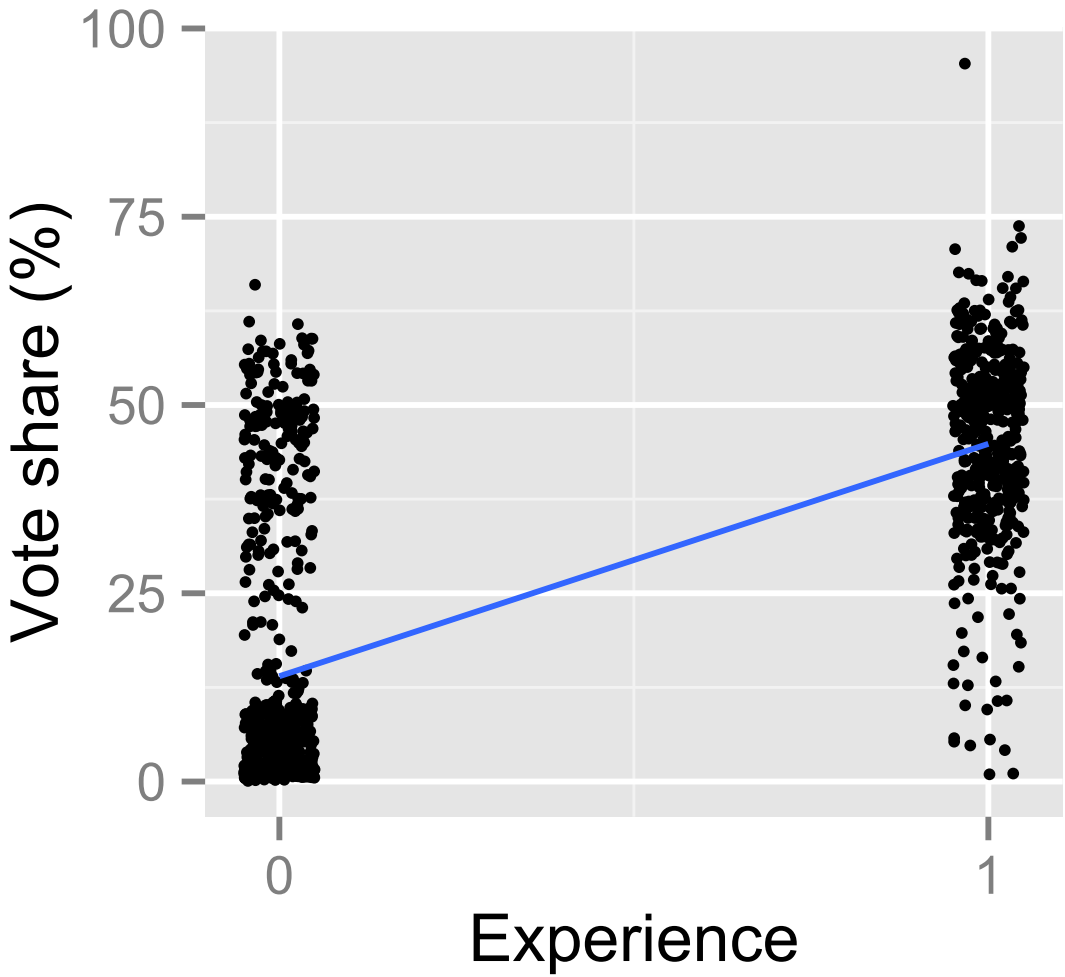

Figure: Explaining the vote share by the MP Experience

Model 2

Explain the vote share in the HR election by the electoral expenditure

- Response: vote share (%)
- Predictor: electoral expenditure (million yen)
- Estimation result:

$$\text{Vote share} = 7.7 + 3.1 \cdot \text{Expenditure} + \text{error}$$

- Points on the regression line (next slide):
  the predicted vote share conditional on the amount of electoral spending

Visual Display of Model 2

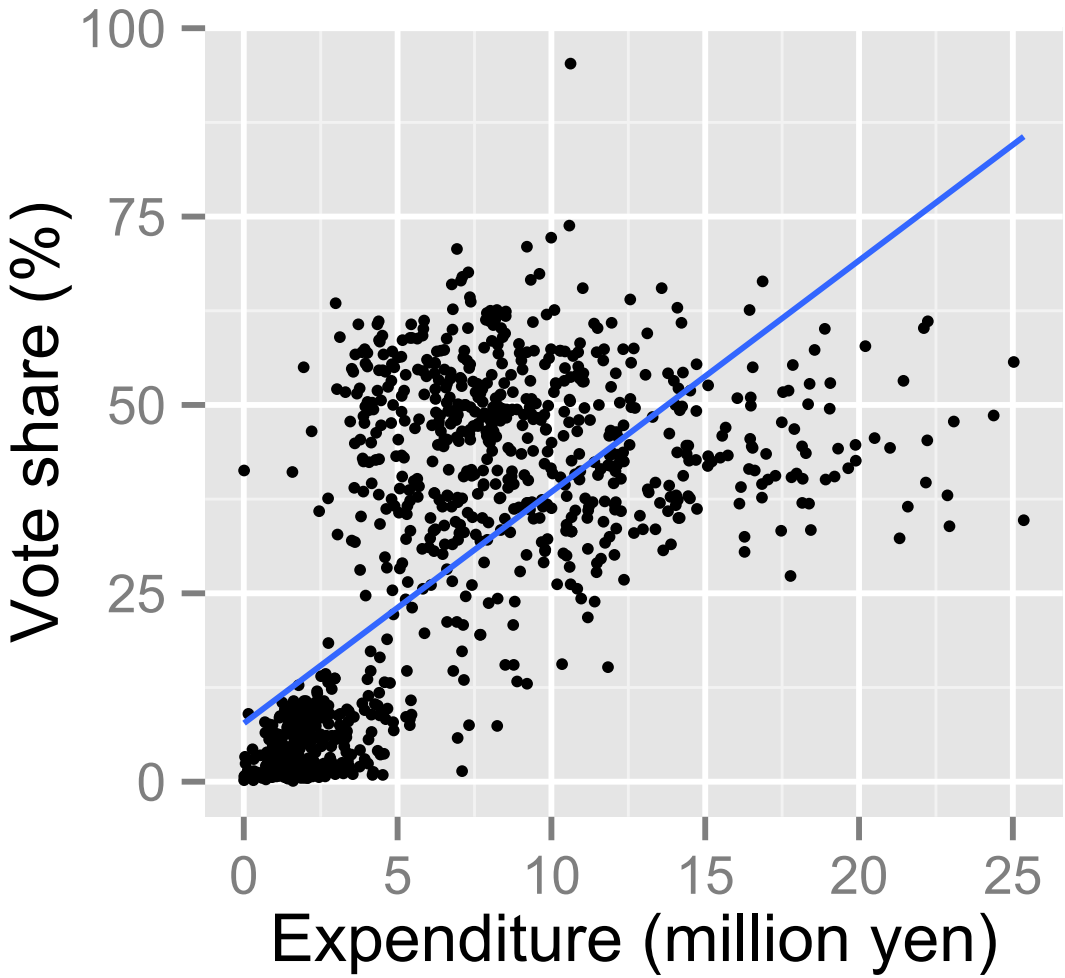

Figure: Explaining the vote share by the electoral expenditure

Meaning of the Estimated Coefficients (1)

$$\text{Vote share} = 7.7 + 3.1 \cdot \text{Expenditure} + \text{error}$$

- Coefficient of Expenditure (3.1): If we compare two candidates one of whom spent more than the other by 1 unit (in this example, 1 million yen), the person who spent more gets the higher vote share than the other by 3.1 percentage point **on average**
  - By increasing the expenditure by a million yen, we **expect** that the vote share will increase by 3.1 point
  - By increasing the expenditure by 10 million yen, we **expect** that the vote share will increase by 31 point

Meaning of the Estimated Coefficients (2)

$$\text{Vote share} = 7.7 + 3.1 \cdot \text{Expenditure} + \text{error}$$

- Coefficient of Intercept (7.7): the mean vote share for the candidates who spent 0 yen for the election
  - Nobody spent 0 yen! (or everybody's expenditure is positive)
  - Need to transform some variables to make the regression intercept "meaningful"

Model 3

Explain the vote share by MP experience *and* the electoral expenditure

- Response: vote share (%)
- Predictor 1: MP experience, 0 or 1
- Predictor 2: electoral expenditure (million yen)
- Estimation result:

$$\text{Voteshare} = 7.9 + 18.1 \cdot \text{Experience} + 1.9 \cdot \text{Expenditure} + \text{error}$$

- Two parallel regression lines (next slide): the model restricts the coefficient of the expenditure)

Visual Display of Model 3

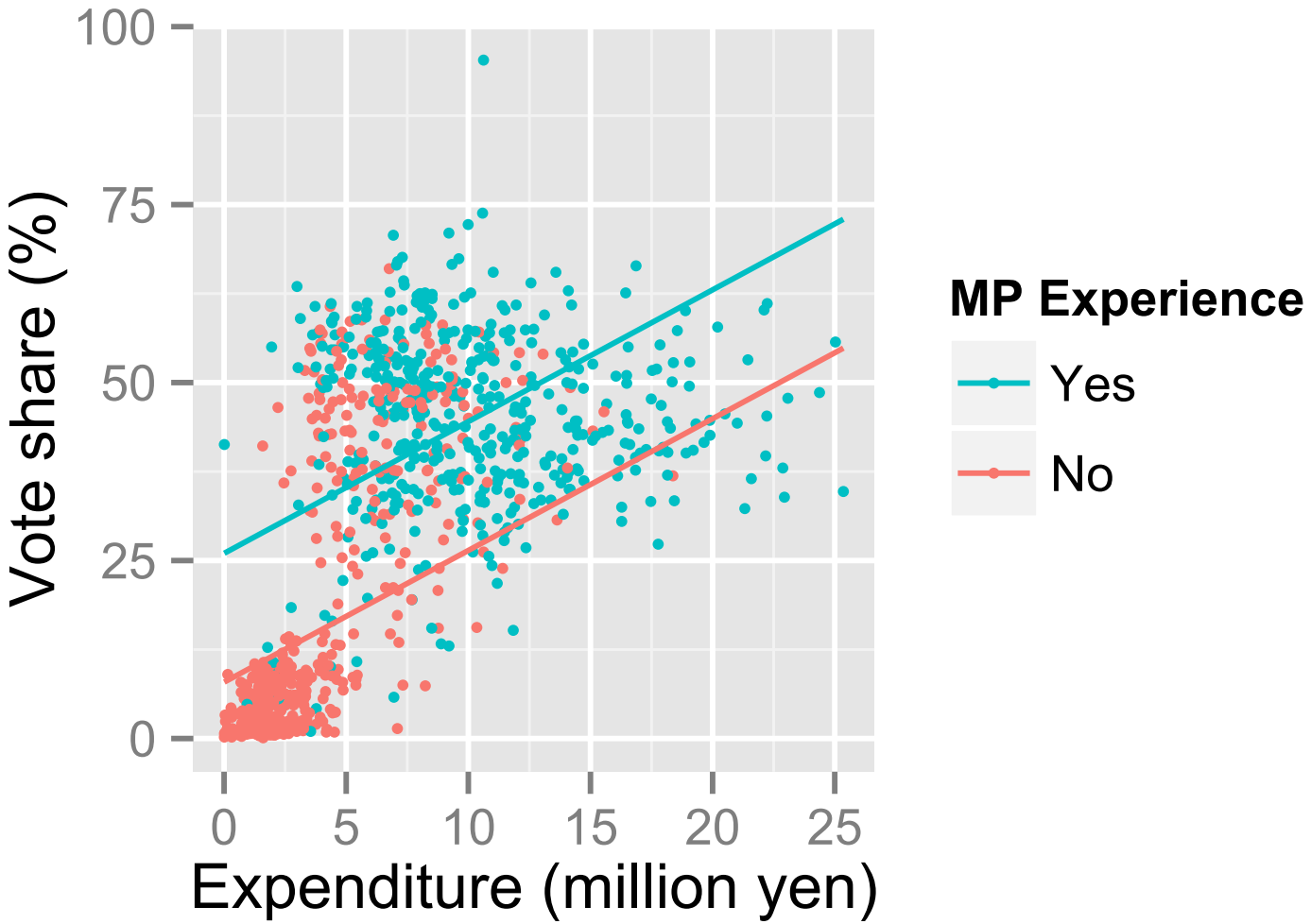

Figure: Explaining vote share by experience and expenditure

Interpreting the Result of Model 3 (1)

$$\text{Voteshare} = 7.9 + 18.1 \cdot \text{Experience} + 1.9 \cdot \text{Expenditure} + \text{error}$$

- Intercept (7.9): the predicted vote share of the candidates with no prior MP experience *and* 0-yen spending
- Coefficient of Experience (18.1): Difference in the predicted vote share between candidates with and without prior MP experience, if the amount of spending is exactly same
  - **Given the same amount of expenditure**, the candidates with experience get 18.1 point higher vote share **on average** than those without

Interpreting the Result of Model 3 (1)

$$\text{Voteshare} = 7.9 + 18.1 \cdot \text{Experience} + 1.9 \cdot \text{Expenditure} + \text{error}$$

- Coefficient of Expenditure (1.9): Difference in the predicted vote share between two (groups of) candidates one of whom spent more than the other by 1 unit of expenditure, if both have prior experience or both do not
  - **Given the same experience**, the vote share increases by 1.9 point **on average** by every additional 1 million yen

Regression Coefficients of Multiple Regression: *ceteris paribus*...

- Coefficient of each predictor represents how many units of change we predict in the response if we increase the predictor by 1 unit, **other things equal**
- Is it possible to always keep "other things equal"?

No!!!

- E.g. "age" and "age squared" as predictors
- E.g. Model with interaction terms

Model 4

- Model 3: the slopes of the regression lines are same for two different groups
- Model 4: "free" the slopes
  $\rightarrow$ take into account the interaction between the MP experience and the expenditure
  - Response: vote share (%)
  - Predictor 1: MP experience, 0 or 1
  - Predictor 2: expenditure (million yen)
  - Predictor 3: MP experience $\times$ expenditure
  - Estimation result:

$$\begin{aligned} \text{Voteshare} &= -2.1 + 45.9 \cdot \text{Experience} + 4.9 \cdot \text{Expenditure} \\ &\quad -4.8 \cdot \text{Experience} \cdot \text{Expenditure} + \text{error} \end{aligned}$$

Visual Display of Model 4

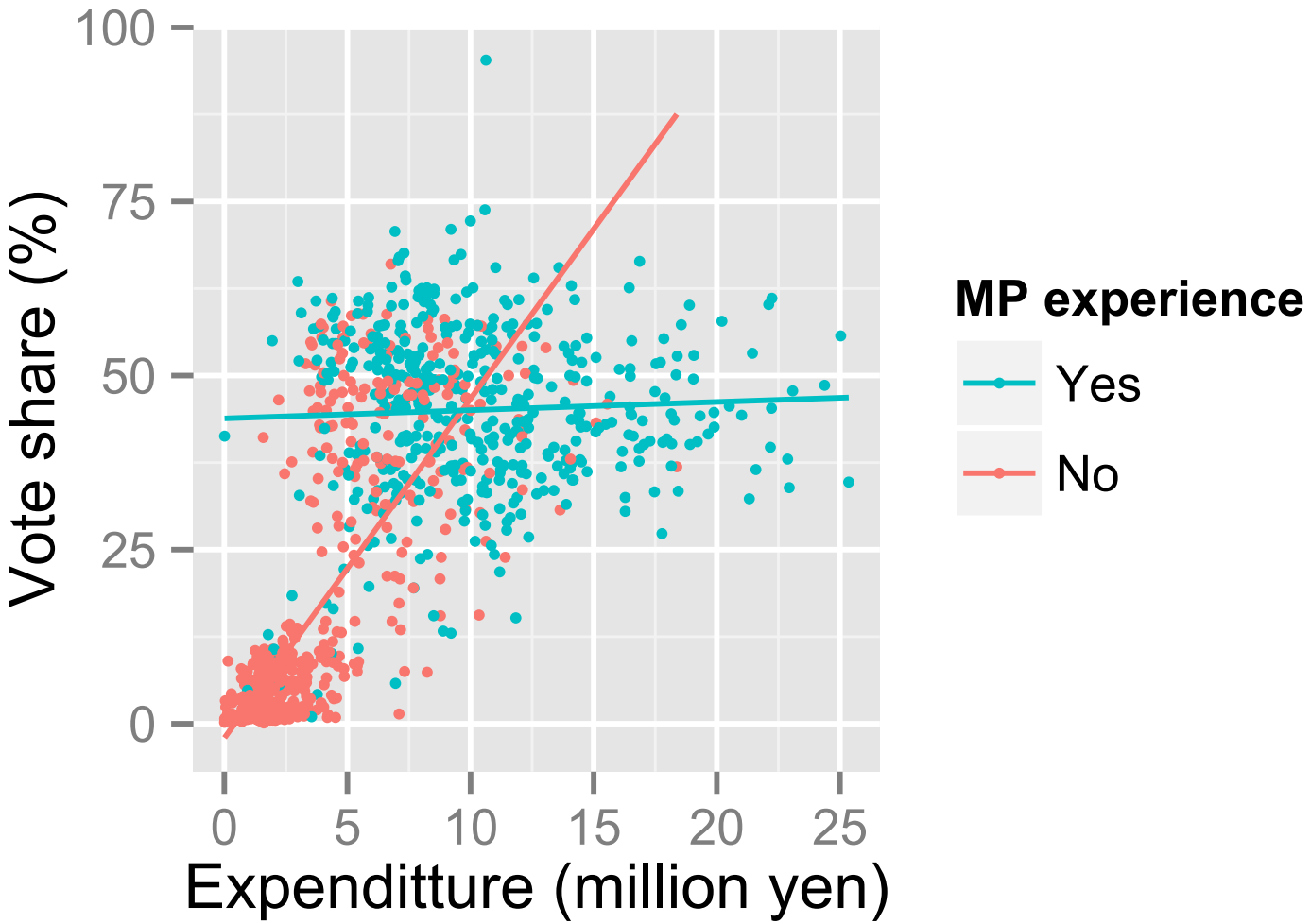

Figure: Explaining vote share by experience and expenditure

Interpreting the Result of Model 4 (1)

$$\begin{aligned}\text{Vote share} = & -2.1 + 45.9 \cdot \text{Experience} + 4.9 \cdot \text{Expenditure} \\ & -4.8 \cdot \text{Experience} \cdot \text{Expenditure} + \text{error}\end{aligned}$$

- Intercept: Predicted vote share for the candidates without prior experience and 0 spending (negative vote share???)
- Coefficient of Experience: Difference in the predicted vote share between the candidates with and without prior experience, **among the candidates who spent 0 yen**

Interpreting the Result of Model 4 (2)

$$\begin{aligned}\text{Vote share} &= -2.1 + 45.9 \cdot \text{Experience} + 4.9 \cdot \text{Expenditure} \\ &\quad -4.8 \cdot \text{Experience} \cdot \text{Expenditure} + \text{error}\end{aligned}$$

- Coefficient of Expenditure: Difference in the predicted vote share between the candidates who spent 1 unit more than the other, **among the candidates without prior experience**
- Coefficient of the interaction: Difference of the slopes (that is, difference in the effect of expenditure) for the candidates with and without experience

Careful interpretation is required for the models with interaction terms!

Interpreting the Result of Model 4 (3)

$$\begin{aligned}\text{Vote share} &= -2.1 + 45.9 \cdot \text{Experience} + 4.9 \cdot \text{Expenditure} \\ &\quad -4.8 \cdot \text{Experience} \cdot \text{Expenditure} + \text{error}\end{aligned}$$

1. Candidates without prior MP experience

$$\begin{aligned}\widehat{\text{Vote share}} &= -2.1 + 45.9 \cdot 0 + 4.9 \cdot \text{Expenditure} \\ &\quad -4.8 \cdot 0 \cdot \text{Expenditure} \\ &= -2.1 + 4.9 \cdot \text{Expenditure}\end{aligned}$$

2. Candidates with prior MP experience:

$$\begin{aligned}\widehat{\text{Vote share}} &= -2.1 + 45.9 \cdot 1 + 4.9 \cdot \text{Expenditure} \\ &\quad -4.8 \cdot 1 \cdot \text{Expenditure} \\ &= -2.1 + 45.9 + 4.9 \cdot \text{Expenditure} - 4.8 \cdot \text{Expenditure} \\ &= 43.8 + 0.1 \cdot \text{Expenditure}\end{aligned}$$