Poli 30D Political Inquiry

Regression

Shane Xinyang Xuan
ShaneXuan.com

November 10, 2016

Contact Information

Shane Xinyang Xuan
xxuan@ucsd.edu
We have someone to help you every day!

Professor Desposato	M	1330-1500 (Latin American Center)
Shane Xuan	Tu	1600-1800 (SSB332)
Cameron Sells	W	1000-1200 (SSB352)
Kelly Matush	Th	1500-1700 (SSB343)
Julia Clark	F	1200-1400 (SSB326)

Supplemental Materials
Our class oriented
ShaneXuan.com
UCLA SPSS starter kit
www.ats.ucla.edu/stat/spss/sk/modules_sk.htm
Princeton data analysis
http://dss.princeton.edu/training/

Announcement

Second SPSS lab on 11/9 – 11/10 at ERC 117 (same as our last lab)!

Teaching Evaluation

Speed

- Take too many questions $\rightsquigarrow$ too many examples
- Take too many questions $\rightsquigarrow$ slow down the progress
- Tentative solution: We will finish the slides first. Then I will take questions. And you are not required to stay if you understand the materials.

Teaching Evaluation

Speed

- Take too many questions $\rightsquigarrow$ too many examples
- Take too many questions $\rightsquigarrow$ slow down the progress
- Tentative solution: We will finish the slides first. Then I will take questions. And you are not required to stay if you understand the materials.

Should be clear on what should be tested

- Tentative solution: Not going to happen.

Teaching Evaluation

Speed

- Take too many questions $\rightsquigarrow$ too many examples
- Take too many questions $\rightsquigarrow$ slow down the progress
- Tentative solution: We will finish the slides first. Then I will take questions. And you are not required to stay if you understand the materials.

Should be clear on what should be tested

- Tentative solution: Not going to happen.
- TAs see the exam at the same time as the students do. I simply do not know if a particular question will be on the exam or not.

Teaching Evaluation

Speed

- Take too many questions $\rightsquigarrow$ too many examples
- Take too many questions $\rightsquigarrow$ slow down the progress
- Tentative solution: We will finish the slides first. Then I will take questions. And you are not required to stay if you understand the materials.

Should be clear on what should be tested

- Tentative solution: Not going to happen.
- TAs see the exam at the same time as the students do. I simply do not know if a particular question will be on the exam or not.
- Also, the point of learning is not for exams.

Teaching Evaluation

Speed

- Take too many questions $\rightsquigarrow$ too many examples
- Take too many questions $\rightsquigarrow$ slow down the progress
- Tentative solution: We will finish the slides first. Then I will take questions. And you are not required to stay if you understand the materials.

Should be clear on what should be tested

- Tentative solution: Not going to happen.
- TAs see the exam at the same time as the students do. I simply do not know if a particular question will be on the exam or not.
- Also, the point of learning is not for exams.
- Moreover, the professor and TAs have made it clear that homework assignments are the best study guide for the exam.

Quiz

We are half way into the quarter. I want you to evaluate your own performance in the class. So here is our quiz question:

Quiz

We are half way into the quarter. I want you to evaluate your own performance in the class. So here is our quiz question:

(1) If you can assign yourself a participation score (on a scale of 1-10), what will it be?

(2) Convince me why do you think so?

Again, here is the template:

LAST NAME, FIRST NAME
EMAIL

ANSWER

Wrap up for controlled comparison

The following table calculates the column percentage

FEELINGS ABOUT PORNOGRAPHY LAWS * attendance at religious services * RESPONDENTS SEX Crosstabulation

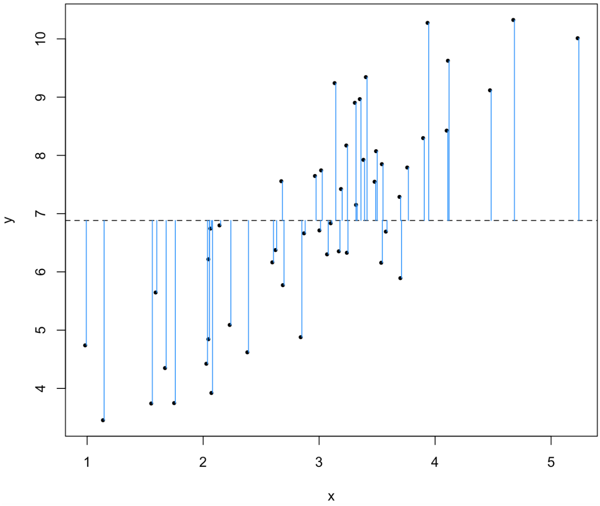
RESPONDENTS SEX				attendance at religious services 1.00 often	2.00 sometimes	3.00 infrequently	Total
1 MALE	FEELINGS ABOUT PORNOGRAPHY LAWS	1 ILLEGAL TO ALL	Count	99	45	90	234
			% within attendance at religious services	52.9%	26.8%	19.1%	28.3%
		2 ILLEGAL UNDER 18	Count	82	115	355	552
			% within attendance at religious services	43.9%	68.5%	75.4%	66.8%
		3 LEGAL	Count	6	8	26	40
			% within attendance at religious services	3.2%	4.8%	5.5%	4.8%
	Total		Count	187	168	471	826
			% within attendance at religious services	100.0%	100.0%	100.0%	100.0%
2 FEMALE	FEELINGS ABOUT PORNOGRAPHY LAWS	1 ILLEGAL TO ALL	Count	191	106	150	447
			% within attendance at religious services	69.2%	46.5%	35.6%	48.3%
		2 ILLEGAL UNDER 18	Count	76	115	252	443
			% within attendance at religious services	27.5%	50.4%	59.9%	47.9%
		3 LEGAL	Count	9	7	19	35
			% within attendance at religious services	3.3%	3.1%	4.5%	3.8%
	Total		Count	276	228	421	925
			% within attendance at religious services	100.0%	100.0%	100.0%	100.0%

Making Regression Make Sense

- We have been primarily working on conceptualization and operationalization in the first half of the quarter. Today we will talk about inference.
- I will first give you some intuition for regressions.

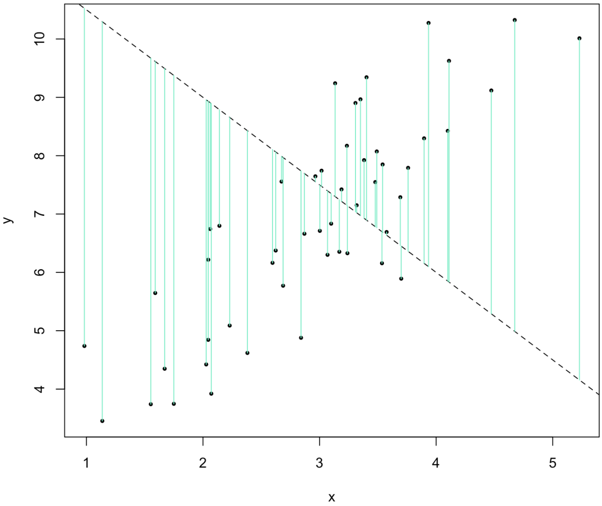
Regression: Examples!

Figure: Data points



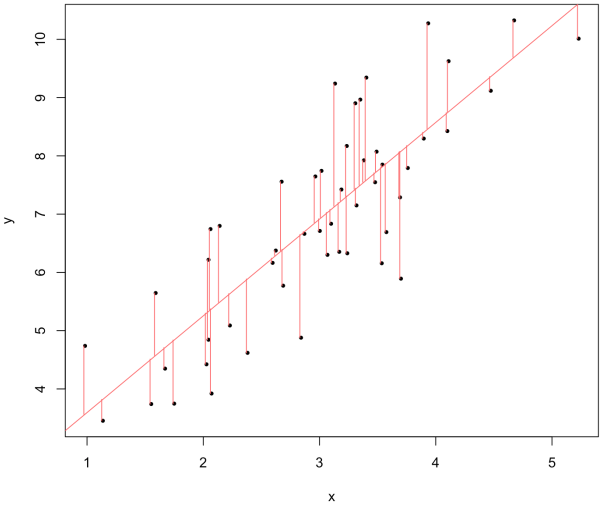
Regression: Examples!

Figure: Bad fit



Regression: Examples!

Figure: Good fit



Model

- Population

$$y_i = \alpha + \beta x_i$$

Model

- Population

$$y_i = \alpha + \beta x_i$$

- Estimation

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i + \hat{e}$$

Model

- Population

$$y_i = \alpha + \beta x_i$$

- Estimation

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i + \hat{e}$$

- Regression Coefficient is calculated by

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

Model

- Population

$$y_i = \alpha + \beta x_i$$

- Estimation

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i + \hat{e}$$

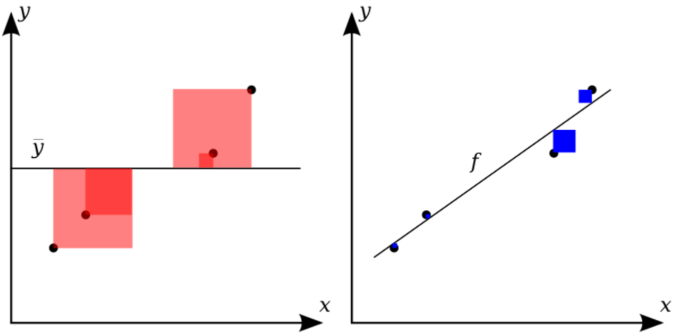
- Regression Coefficient is calculated by

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

- R^2 is calculated by

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Intuition



$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

- Red squares represent the squared residuals wrt the average
- Blue squares represent the squared residuals wrt the 'best fit'
- Interpret R^2

R and R^2

Suppose $R = 0.96$, and $R^2 = 0.92$

R and R^2

Suppose $R = 0.96$, and $R^2 = 0.92$

- Variation in X explains 92% variation in Y
- There is a strong, positive, linear relationship between X and Y

R and R^2

Suppose $R = 0.96$, and $R^2 = 0.92$

- Variation in X explains 92% variation in Y
- There is a strong, positive, linear relationship between X and Y

It's possible that R is negative. But R^2 is always positive.

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$
2. Multiply to get $(\bar{y} - y)^2, (\bar{x} - x)^2, (\bar{y} - y)(\bar{x} - x)$

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$
2. Multiply to get $(\bar{y} - y)^2, (\bar{x} - x)^2, (\bar{y} - y)(\bar{x} - x)$
3. Sum over what you obtained from step 2

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$
2. Multiply to get $(\bar{y} - y)^2, (\bar{x} - x)^2, (\bar{y} - y)(\bar{x} - x)$
3. Sum over what you obtained from step 2
4. You thus obtain $\hat{\beta}$

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$
2. Multiply to get $(\bar{y} - y)^2, (\bar{x} - x)^2, (\bar{y} - y)(\bar{x} - x)$
3. Sum over what you obtained from step 2
4. You thus obtain $\hat{\beta}$
5. Use $y_i - \hat{\beta} x_i = \alpha_i$ to get α_i's

Calculate by hand

Calculate $\hat{\beta}$ (coefficient), and $\hat{\alpha}$ (constant) by hand:

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

X	18	20	22	24	24
Y	5	5	6	6	6

1. Calculate $\bar{y}, \bar{x}, \bar{y} - y, \bar{x} - x$
2. Multiply to get $(\bar{y} - y)^2, (\bar{x} - x)^2, (\bar{y} - y)(\bar{x} - x)$
3. Sum over what you obtained from step 2
4. You thus obtain $\hat{\beta}$
5. Use $y_i - \hat{\beta} x_i = \alpha_i$ to get α_i's
6. Average α_i to get $\hat{\alpha}$

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

1. Follow the previous slide to get $\hat{\beta}, \hat{\alpha}$

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

1. Follow the previous slide to get $\hat{\beta}, \hat{\alpha}$
2. Calculate $\hat{y}_i$ using $\hat{y}_i = \hat{\beta}x_i + \hat{\alpha}$

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

1. Follow the previous slide to get $\hat{\beta}, \hat{\alpha}$
2. Calculate $\hat{y}_i$ using $\hat{y}_i = \hat{\beta}x_i + \hat{\alpha}$
3. Calculate the residuals $(y_i - \hat{y}_i)$ $\rightsquigarrow$ square $\rightsquigarrow$ sum

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

1. Follow the previous slide to get $\hat{\beta}, \hat{\alpha}$
2. Calculate $\hat{y}_i$ using $\hat{y}_i = \hat{\beta}x_i + \hat{\alpha}$
3. Calculate the residuals $(y_i - \hat{y}_i)$ $\rightsquigarrow$ square $\rightsquigarrow$ sum
4. You have $\sum_i (y_i - \bar{y})^2$ from previous

Calculate by hand (cont.)

Calculate R^2 by hand

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

X	18	20	22	24	24
Y	5	5	6	6	6

1. Follow the previous slide to get $\hat{\beta}, \hat{\alpha}$
2. Calculate $\hat{y}_i$ using $\hat{y}_i = \hat{\beta} x_i + \hat{\alpha}$
3. Calculate the residuals $(y_i - \hat{y}_i)$ $\rightsquigarrow$ square $\rightsquigarrow$ sum
4. You have $\sum_i (y_i - \bar{y})^2$ from previous
5. You obtain R^2 (yes!)

Calculate by hand (cont.)

Let's go through the example in detail to make sure you understand it! Download the example from my website (`https://shanexuan.com/teaching/`).

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i + \hat{e}$$

$$\hat{\beta} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

$$\mathsf{R}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Application

Suppose we have the model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_0 + \varepsilon$$

$\rightsquigarrow$ A 1-unit change in X_1 is associated with a β_1-unit change in Y, all else equal.

$\rightsquigarrow$ A 1-unit change in X_2 is associated with a β_2-unit change in Y, all else equal.