Analysing Politics
Lecture 15: Multivariate data analysis: linear regression

Jos Elkink

University College Dublin

March 1, 2010

1. Introduction
2. Scatterplots
3. Correlation and regression

Outline

Typology of methods

	Descriptive	Inferential
Univariate		
Multivariate		

Refresher

- Levels of measurement
  - Dichotomous
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Refresher

- Levels of measurement
  - Dichotomous
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Histogram / frequency table

Refresher

- Levels of measurement
  - Dichotomous
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Histogram / frequency table
- Measures of central tendency
  - Mean
  - Median
  - Mode

Refresher

- Levels of measurement
  - Dichotomous
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Histogram / frequency table
- Measures of central tendency
  - Mean
  - Median
  - Mode
- Variation
  - Variance
  - Standard deviation

Outline

Scatterplot

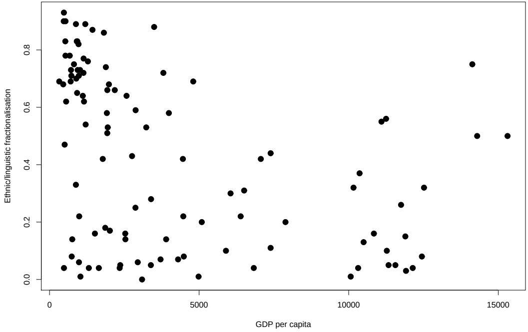

Figure: Ethnic/linguistic fractionalisation by GDP per capita, 1980

Scatterplot

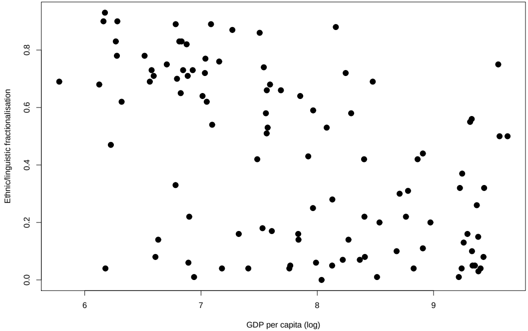

Figure: Ethnic/linguistic fractionalisation by GDP per capita (logged), 1980

Regression line

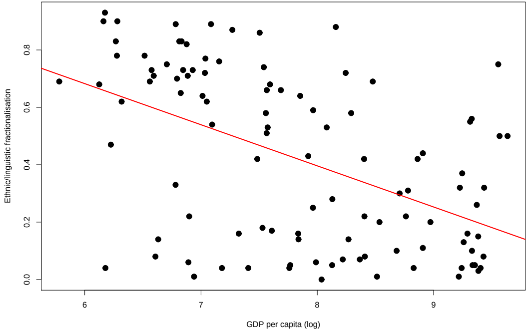

Figure: Ethnic/linguistic fractionalisation by GDP per capita (logged), 1980

Regression line

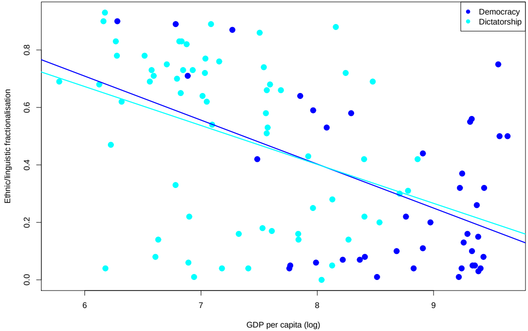

Figure: Ethnic/linguistic fractionalisation by GDP per capita (logged) and regime type, 1980

Outline

Variance

Variance estimator:

$$Var(x) = s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

Covariance

Covariance estimator:

$$Cov(x, y) = s^2_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

Pearson correlation

Correlation estimator:

$$Cor(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

```
> cor(aclp$ELF60, log(aclp$LEVEL), use="complete.obs")
[1] -0.5038412
```

Covariance: example

	A	B	C
	1	2	6
	4	8	1
	2	3	6
	5	10	2
	4	9	1
	6	10	0
	3	2	5
	2	1	4
Total	27	45	25
Mean	3.375	5.625	3.125

Covariance: example

	A	$(A - \bar{A})$	B	$(B - \bar{B})$
	1	-2.375	2	-3.625
	4	0.625	8	2.375
	2	-1.375	3	-2.625
	5	1.625	10	4.375
	4	0.625	9	3.375
	6	2.625	10	4.375
	3	-0.375	2	-3.625
	2	-1.375	1	-4.625
Total	27	0	45	0

Covariance: example

	A	$(A-\bar{A})$	B	$(B-\bar{B})$	$(A-\bar{A})(B-\bar{B})$
	1	-2.375	2	-3.625	8.609
	4	0.625	8	2.375	1.484
	2	-1.375	3	-2.625	3.609
	5	1.625	10	4.375	7.109
	4	0.625	9	3.375	2.109
	6	2.625	10	4.375	11.484
	3	-0.375	2	-3.625	1.359
	2	-1.375	1	-4.625	6.359
Total	27	0	45	0	42.125

$$Cov(A,B) = \frac{\sum_{i=1}^{N}(A_i-\bar{A})(B_i-\bar{B})}{N-1} = \frac{42.125}{8-1} = 6.018$$

Correlation

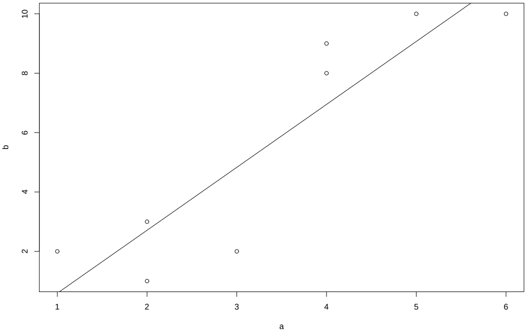

Figure: A vs B

Correlation

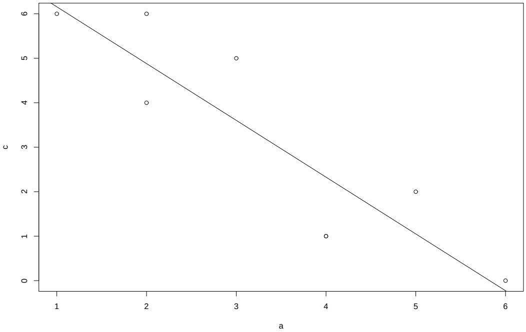

Figure: A vs C

Simple linear regression

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

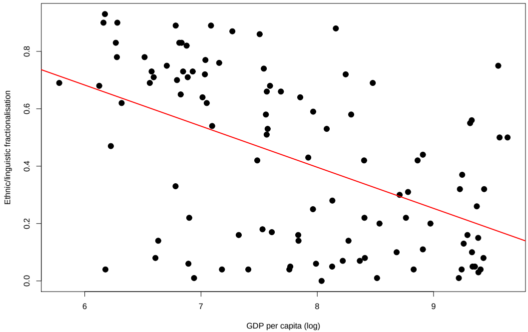

Figure: Ethnic/linguistic fractionalisation by GDP per capita (logged), 1980

Linear regression visualised

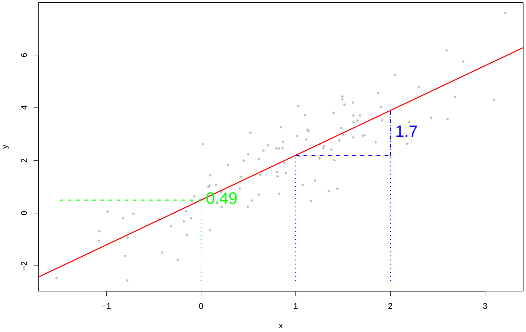

Figure: Linear regression.

Linear regression visualised

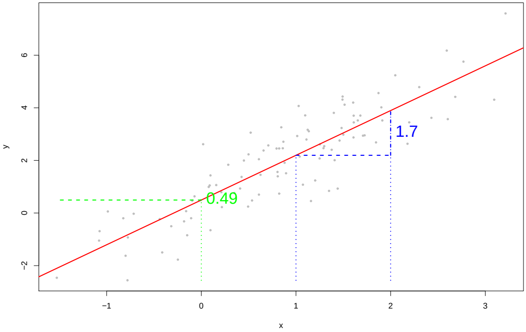

Figure: Linear regression.