Introduction to regression

Paul Schrimpf

UBC
Economics 326

January 23, 2018

Review of last week

- Expectations and conditional expectations
  - Linear
  - Iterated expectations
- Asymptotics – using large sample distribution to approximate finite sample distribution of estimators
  - LLN: sample moments converge in probability to population moments,

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} g(x_i)}_{\text{sample moment}} \xrightarrow{p} \underbrace{\mathrm{E}[g(x)]}_{\text{population moment}}$$

  - CLT: centered and scaled sample moments converge in distribution to population moments

$$\underbrace{\sqrt{n}}_{\text{“scaling”}} \left(\frac{1}{n}\sum_{i=1}^{n} g(x_i) \underbrace{- \mathrm{E}[g(x)]}_{\text{“centering”}} \right) \xrightarrow{d} N\left(0, \mathrm{Var}(g(x))\right)$$

  - Using CLT to calculate p-values

Part I

Definition and interpretation of regression

References

- Main texts:
  - Angrist and Pischke (2014) chapter 2
  - Wooldridge (2013) chapter 2
  - Stock and Watson (2009) chapter 4-5
- More advanced:
  - Angrist and Pischke (2009) chapter 3 up to and including section 3.1.2 (pages 27-40)
  - Bierens (2012)
  - Abbring (2001) chapter 3
  - Baltagi (2002) chapter 3
  - Linton (2017) chapters 16-20, 22
- More introductory:
  - Diez, Barr, and Cetinkaya-Rundel (2012) chapter 7

Section 1

Motivation

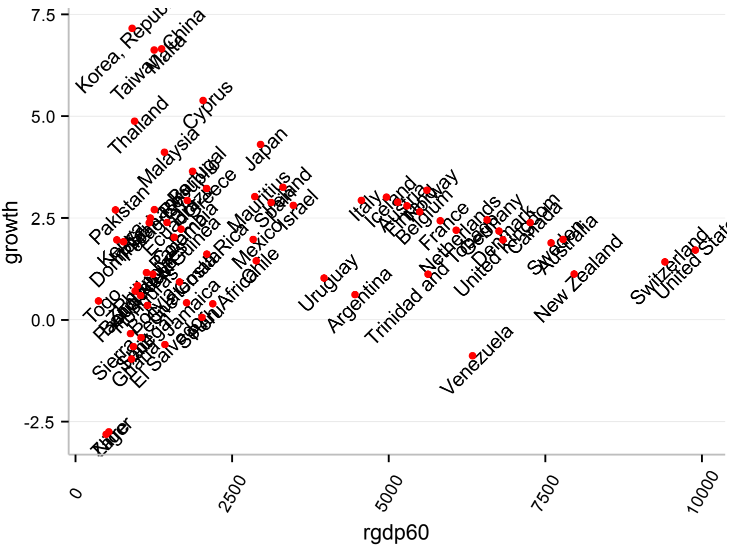
General problem

- Often interested in relationship between two (or more) variables, e.g.
  - Wages and education
  - Minimum wage and unemployment
  - Price, quantity, and product characteristics
- Usually have:
  1. Variable to be explained (dependent variable)
  2. Explanatory variable(s) or independent variables or covariates

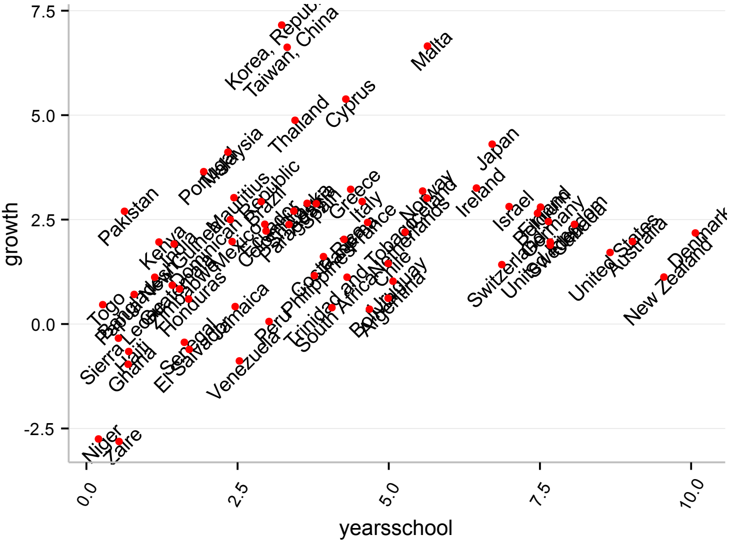
Dependent	Independent
Wage	Education
Unemployment	Minimum wage
Quantity	Price and product characteristics
Y	X

- For now agnostic about causality, but $\mathrm{E}[Y|X]$ usually is not causal

Example: Growth and GDP



Years of schooling in 1960 and growth



Section 2

Conditional expectation function

Conditional expectation function

- One way to describe relation between two variables is a function,
$$Y = h(X)$$
- Most relationships in data are not deterministic, so look at average relationship,
$$\begin{aligned} Y = & \underbrace{\mathrm{E}[Y|X]}_{\equiv h(X)} + \underbrace{(Y - \mathrm{E}[Y|X])}_{\equiv \epsilon} \\ = & \mathrm{E}[Y|X] + \epsilon \end{aligned}$$
- Note that $\mathrm{E}[\epsilon] = 0$ (by definition of ϵ and iterated expectations)
- $\mathrm{E}[Y|X]$ can be any function, in particular, it need not be linear

Conditional expectation function

- Unrestricted $\mathrm{E}[Y|X]$ hard to work with
  - Hard to estimate
  - Hard to communicate if X a vector (cannot draw graphs)
- Instead use linear regression
  - Easier to estimate and communicate
  - Tight connection to $\mathrm{E}[Y|X]$

Section 3

Population regression

Population regression

- The bivariate population regression of Y on X is

$$(\beta_0, \beta_1) = \arg\min_{b_0, b_1} \mathrm{E}[(Y - b_0 - b_1 X)^2]$$

  i.e. β_0 and β_1 are the slope and intercept that minimize the expected square error of $Y - (\beta_0 + \beta_1 X)$
- Calculating β_0 and β_1:
  - First order conditions:

$$\begin{aligned}[b_0] : 0 = & \frac{\partial}{\partial b_0} \mathrm{E}[(Y - b_0 - b_1 X)^2] \\ = & \mathrm{E}\left[\frac{\partial}{\partial b_0}(Y - b_0 - b_1 X)^2\right] \\ = & \mathrm{E}\left[-2(Y - \beta_0 - \beta_1 X)\right] \end{aligned} \quad (1)$$

Population regression

and

$$
\begin{aligned}
[b_1]: 0 =& \frac{\partial}{\partial b_1} \mathrm{E}[(Y - b_0 - b_1 X)^2] \\
=& \mathrm{E}\left[\frac{\partial}{\partial b_1}(Y - b_0 - b_1 X)^2\right] \\
=& \mathrm{E}\left[-2(Y - \beta_0 - \beta_1 X)X\right] \qquad (2)
\end{aligned}
$$

- (1) rearranged gives $\beta_0 = \mathrm{E}[Y] - \beta_1 \mathrm{E}[X]$
- Substituting into (2)

$$
\begin{aligned}
0 =& \mathrm{E}\left[X(-Y + \mathrm{E}[Y] - \beta_1 \mathrm{E}[X] + \beta_1 X)\right] \\
=& \mathrm{E}\left[X(-Y + \mathrm{E}[Y])\right] + \beta_1 \mathrm{E}\left[X(X - \mathrm{E}[X])\right] \\
=& -\mathrm{Cov}(X, Y) + \beta_1 \mathrm{Var}(X) \\
\beta_1 =& \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}
\end{aligned}
$$

- $\beta_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$, $\beta_0 = \mathrm{E}[Y] - \beta_1 \mathrm{E}[X]$

Population regression approximates $\mathrm{E}[Y|X]$

Lemma

The population regression is the minimal mean square error linear approximation to the conditional expectation function, i.e.

$$\underbrace{\arg\min_{b_0,b_1} \mathrm{E}\left[(Y-(b_0+b_1X))^2\right]}_{\text{population regression}} = \arg\min_{b_0,b_1} \underbrace{\mathrm{E}_X\left[(\mathrm{E}[Y|X]-(b_0+b_1X))^2\right]}_{\text{MSE of linear approximation to } \mathrm{E}[Y|X]}$$

Corollary

If $\mathrm{E}[Y|X] = c + mX$, then the population regression of Y on X equals $\mathrm{E}[Y|X]$, i.e. $\beta_0 = c$ and $\beta_1 = m$

Proof

Proof.

- Let b_0^*, b_1^* be minimizers of MSE of approximation to $\mathrm{E}[Y|X]$
- Same steps as in population regression formula gives

$$0 = \mathrm{E}\left[-2(\mathrm{E}[Y|X] - b_0^* - b_1^* X)\right]$$

and

$$0 = \mathrm{E}\left[-2(\mathrm{E}[Y|X] - b_0^* - b_1^* X)X\right]$$

- Rearranging and combining,

$$b_0^* = \mathrm{E}[\mathrm{E}[Y|X]] - b_1^* \mathrm{E}[X] = \mathrm{E}[Y] - b_1^* \mathrm{E}[X]$$

and

$$\begin{aligned} 0 =& \mathrm{E}\left[X(-\mathrm{E}[Y|X] + \mathrm{E}[Y] + b_1^* \mathrm{E}[X] - b_1^* X)\right] \\ =& \mathrm{E}\left[X(-\mathrm{E}[Y|X] + \mathrm{E}[Y])\right] + b_1^* \mathrm{E}\left[X(X - \mathrm{E}[X])\right] \\ =& -\mathrm{Cov}(X, Y) + b_1^* \mathrm{Var}(X) \\ b_1^* =& \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \end{aligned}$$

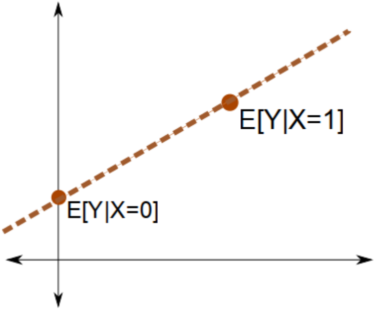
□

Regression interpretation

- Regression = best linear approximation to $\mathrm{E}[Y|X]$
- $\beta_0 \approx \mathrm{E}[Y|X=0]$
- $\beta_1 \approx \frac{d}{dx}\mathrm{E}[Y|X] \approx$ change in average Y per unit change in X
- Not necessarily a causal relationship (usually not)
- Always can be viewed as description of data

Regression with binary X

- Suppose X is binary (i.e. can only be 0 or 1)
- We know $\beta_0 + \beta_1 X =$ best linear approximation to $\mathrm{E}[Y|X]$
- X only takes two values, so can draw line connecting $\mathrm{E}[Y|X = 0]$ and $\mathrm{E}[Y|X = 1]$, so $\beta_0 + \beta_1 X = \mathrm{E}[Y|X]$
    - $\beta_0 = \mathrm{E}[Y|X = 0]$
    - $\beta_0 + \beta_1 = \mathrm{E}[Y|X = 1]$



Section 4

Sample regression

Sample regression

- Have sample of observations: $\{(y_i, x_i)\}_{i=1}^n$
- The sample regression (or when unambiguous just "regression") of Y on X is

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

i.e. $\hat{\beta}_0$ and $\hat{\beta}_1$ are the slope and intercept that minimize the sum of squared errors, $(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$

  - Same as population regression but with sample average instead of expectation

- Same calculation as for population regression would show

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(X)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Sample regression

- Sample regression is an estimator for the population regression
- Given an estimator we should ask:
    - Unbiased?
    - Variance?
    - Consistent?
    - Asymptotically normal?
- We will address these questions in the next week or two

Section 5

Regression in R

Regression in R

```r
require(datasets) ## some datasets included with R
stateDF <- data.frame(state.x77)
summary(stateDF) ## summary statistics of data

## Sample regression function
regress <- function(y, x) {
  beta <- vector(length=2)
  beta[2] <- cov(x,y)/var(x)
  beta[1] <- mean(y) - beta[2]*mean(x)
  return(beta)
}

## Regress life expectancy on income
beta <- regress(stateDF[,"Life.Exp"],stateDF$Income)
beta

## builtin regression
lm(Life.Exp ~ Income, data=stateDF)
## more detailed output
summary(lm(Life.Exp ~ Income, data=stateDF))
```

https://bitbucket.org/paulschrimpf/econ326/src/master/notes/03/regress.R?at=master

Part II

Properties of regression

Section 6

Fitted value and residuals

Fitted values and residuals

- Fitted values:
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
- Residuals:
$$\hat{\epsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = y_i - \hat{y}_i$$
$$y_i = \hat{y}_i + \hat{\epsilon}_i$$
- Sample mean of residuals $= 0$
  - First order condition for $\hat{\beta}_0$,
$$0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$
$$0 = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i$$
- Sample covariance of x and $\hat{\epsilon} = 0$

Fitted values and residuals

- First order condition for $\hat{\beta}_1$,

$$0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i x_i$$

Fitted values and residuals

- Sample mean of $\hat{y}_i = \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n y_i = & \frac{1}{n}\sum_{i=1}^n \hat{y}_i + \hat{\epsilon}_i \\
= & \frac{1}{n}\sum_{i=1}^n \hat{y}_i \\
= & \frac{1}{n}\sum_{i=1}^n \hat{\beta}_0 + \hat{\beta}_1 x_i \\
= & \hat{\beta}_0 + \hat{\beta}_1 \bar{x}
\end{aligned}
$$

Fitted values and residuals

- Sample covariance of y and $\hat{\epsilon}$ = sample variance of $\hat{\epsilon}$:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} y_i(\hat{\epsilon}_i - \bar{\hat{\epsilon}}) =& \frac{1}{n}\sum_{i=1}^{n} y_i\hat{\epsilon}_i \\
=& \frac{1}{n}\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\epsilon}_i)\hat{\epsilon}_i \\
=& \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i + \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_i\hat{\epsilon}_i + \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 \\
=& \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2
\end{aligned}
$$

R^2

- Decompose y_i

$$y_i = \hat{y}_i + \hat{\epsilon}_i$$

- Total sum of squares = explained sum of squares + sum of squared residuals

$$\underbrace{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2}_{SST} = \underbrace{\frac{1}{n}\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{SSE} + \underbrace{\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2}_{SSR}$$

- R-squared: fraction of sample variation in y that is explained by x

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = o\widehat{\text{Corr}}(y, \hat{y})$$

  - $0 \leq R^2 \leq 1$
  - If all data on regression line, then $R^2 = 1$
  - Magnitude of R^2 does not have direct bearing on economic importance of a regression

Section 7

Statistical properties

Unbiased

- $\mathrm{E}[\hat{\beta}] = ?$
- Assume:

SLR.1 (linear model) $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

SLR.2 (independence) $\{(x_i, y_i)\}_{i=1}^n$ is independent random sample

SLR.3 (rank condition) $\widehat{\mathrm{Var}}(X) > 0$

SLR.4 (exogeneity) $\mathrm{E}[\epsilon|X] = 0$

- Then, $\mathrm{E}[\hat{\beta}_1] = \beta_1$ and $\mathrm{E}[\hat{\beta}_0] = \beta_0$

Variance

- $\text{Var}(\hat{\beta})$?
- Assume SLR.1-4 and

SLR.5 (homoskedasticity) $\text{Var}(\epsilon|X) = \sigma^2$

- Then,

$$\text{Var}(\hat{\beta}_1|\{x_i\}_{i=1}^n) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\text{Var}(\hat{\beta}_0|\{x_i\}_{i=1}^n) = \frac{\sigma^2 \frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Distribution with normal errors

- Assume SLR.1-SLR.5 and

SLR.6 (normality) $\epsilon_i|x_i \sim N(0, \sigma^2)$

- Then $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$, and

$$\hat{\beta}_1|\{x_i\}_{i=1}^n \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

- Even without assuming normality, the central limit theorem implies $\hat{\beta}$ is asymptotically normal (details in a later lecture)

Summary

- Simple linear regression model assumptions:

  SLR.1 (linear model) $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

  SLR.2 (independence) $\{(x_i, y_i)\}_{i=1}^n$ is independent random sample

  SLR.3 (rank condition) $\widehat{\mathrm{Var}}(X) > 0$

  SLR.4 (exogeneity) $\mathrm{E}[\epsilon|X] = 0$

  SLR.5 (homoskedasticity) $\mathrm{Var}(\epsilon|X) = \sigma^2$

  SLR.6 (normality) $\epsilon_i|x_i \sim N(0, \sigma^2)$

- $\hat{\beta}$ unbiased if SLR.1-SLR.4
- If also SLR.5, then $\mathrm{Var}(\hat{\beta}_1|\{x_i\}_{i=1}^n) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$
- If also SLR.6, then $\hat{\beta}_1|\{x_i\}_{i=1}^n \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$

Discussion of assumptions

SLR.1 Having a linear model makes it easier to state the other assumptions, but we could instead start by saying let $\beta_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ and $\beta_0 = \mathrm{E}[Y] - \beta_1 \mathrm{E}[X]$ be the population regression coefficients and define $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$

Discussion of assumptions

SLR.2 Independent observations is a good assumption for data from a simple random sample

- Common situations where it fails in economics are when we have a time series of observations, e.g. $\{(x_t, y_t)\}_{t=1}^n$ could be unemployment and GDP of Canada for many different years; and clustering, e.g. the data could be students test scores and hours studying and our sample consists of randomly chosen courses or schools–students in the same course would not be independent, but across different courses they might be.
- Still have $\mathrm{E}[\hat{\beta}_1] = \beta_1$ with non-independent observations as long as $\mathrm{E}[\epsilon_i | x_1, ..., x_n] = 0$
- The variance of $\hat{\beta}_1$ will change with non-independent observations
- Simulation code

Discussion of assumptions

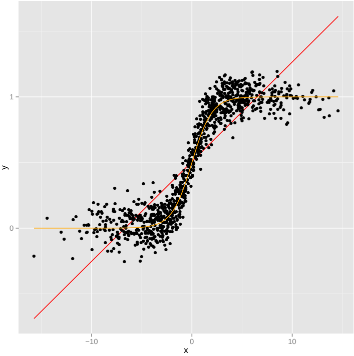
SLR.3 If $\widehat{\text{Var}}(X) = 0$, then $\hat{\beta}_1$ involves dividing by 0

- If there is no variation in X, then we cannot see how Y is related to X

Discussion of assumptions

SLR.4 To think about mean independence of ϵ from x we should have a model motivating the regression

- If the model we want is just a population regression, then automatically $\mathrm{E}[\epsilon X] = 0$, and $\mathrm{E}[\epsilon|X] = 0$ if the conditional expectation function is linear; if conditional expectation nonlinear maybe still a useful approximation



Code

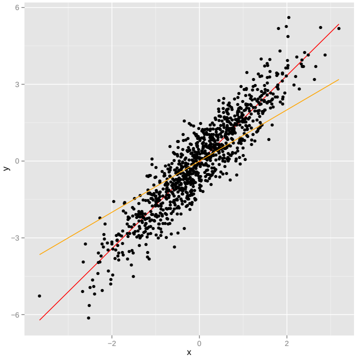
Discussion of assumptions

SLR.4 To think about mean independence of ϵ from x we should have a model motivating the regression

- If the model we want is anything else, then maybe $\mathrm{E}[\epsilon X] \neq 0$ (and $\mathrm{E}[\epsilon|X] \neq 0$), e.g.
  - Demand curve

    $$p_i = \beta_0 + \beta_1 q_i + \epsilon_i$$

    $\epsilon_i =$ everything that affects price other than quantity. q_i determined in equilibrium implies $\mathrm{E}[\epsilon_i|q_i] \neq 0$
  - $\mathrm{E}[\hat{\beta}_1] \neq \beta_1$ and $\hat{\beta}_1$ does not tell us what we want

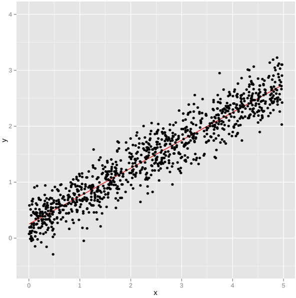


Code

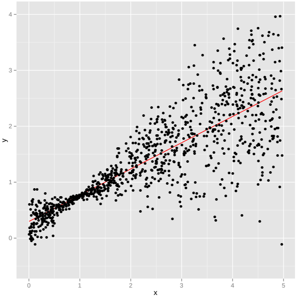
Discussion of assumptions

SLR.5 Homoskedasticity: variance of ϵ does not depend on X

Homoskedastic



Heteroskedastic



Code

- Heteroskedasticity is when $\text{Var}(\epsilon|X)$ varies with X
- If there is heteroskedasticity, the variance of $\hat{\beta}_1$ is different, but we can fix it

Discussion of assumptions

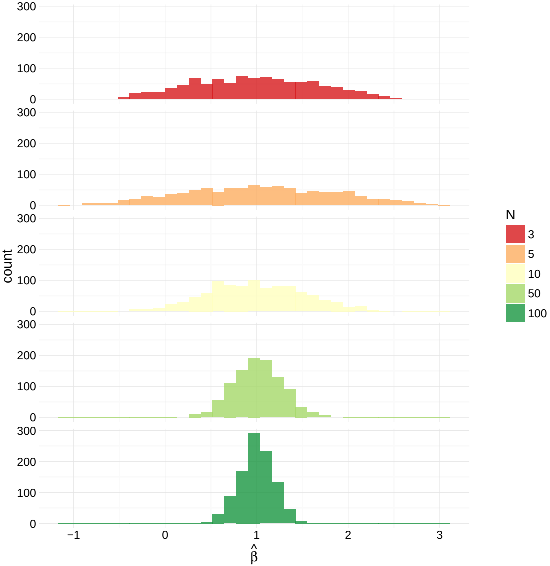
- "robust standard errors" / "heteroscedasticity-consistent (HC) standard errors" / "Eicker–Huber–White standard errors"

Discussion of assumptions

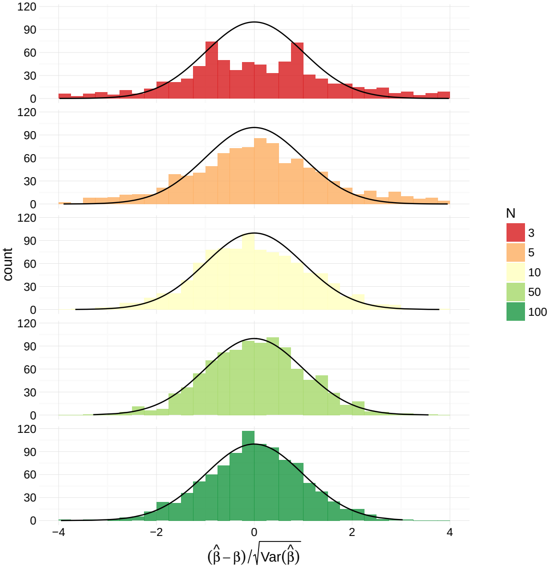
SLR.6 If $\epsilon_i | x_i \sim N$, then $\hat{\beta}_1 \sim N$

- What if ϵ_i not normally distributed?
- We will see that $\hat{\beta}_1$ still asymptotically normal
- Simulation

Discussion of assumptions



Discussion of assumptions



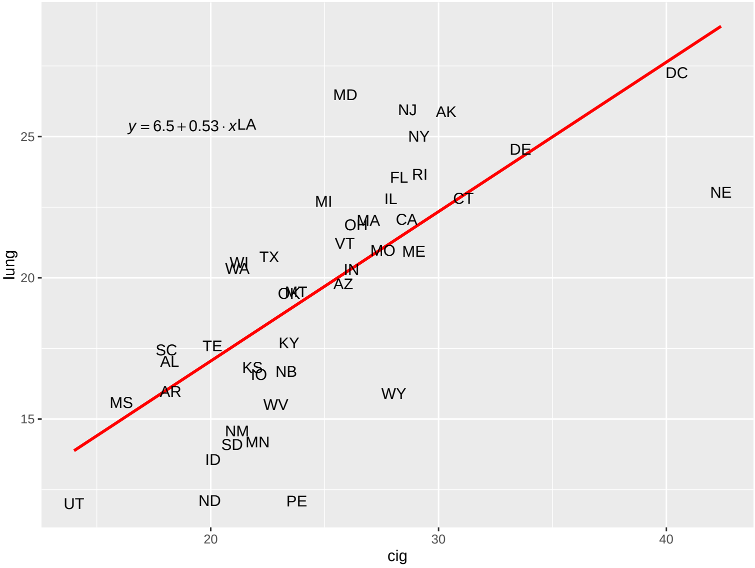
Section 8

Examples

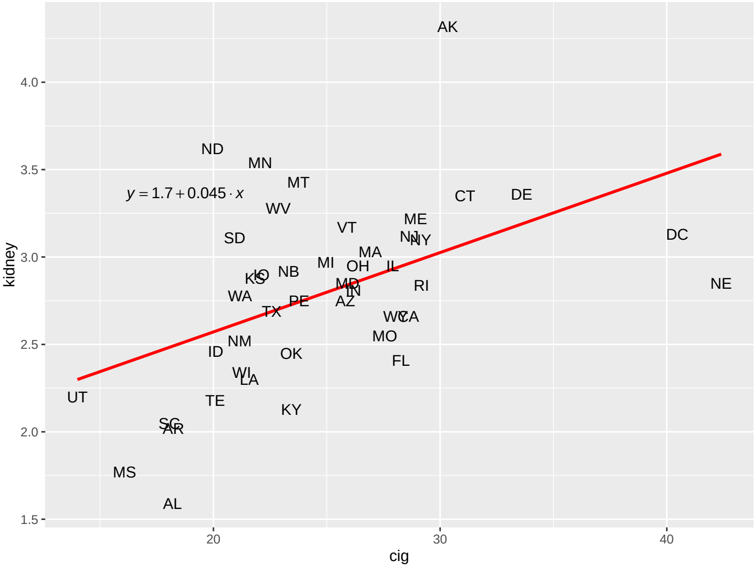
Example: smoking and cancer

- Data on per capita number of cigarettes sold and death rates per thousand from cancer for U.S. states in 1960
- `http://lib.stat.cmu.edu/DASL/Datafiles/cigcancerdat.html`
- Death rates from: lung cancer, kidney cancer, bladder cancer, and leukemia Code

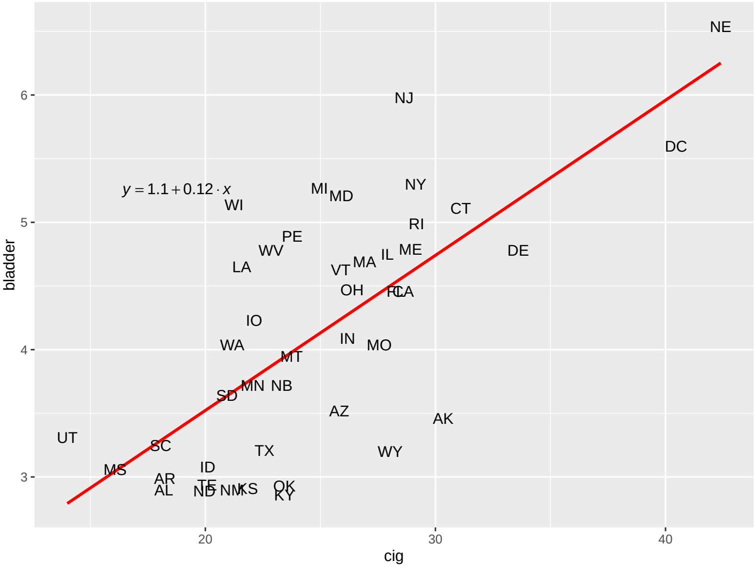
Smoking and lung cancer



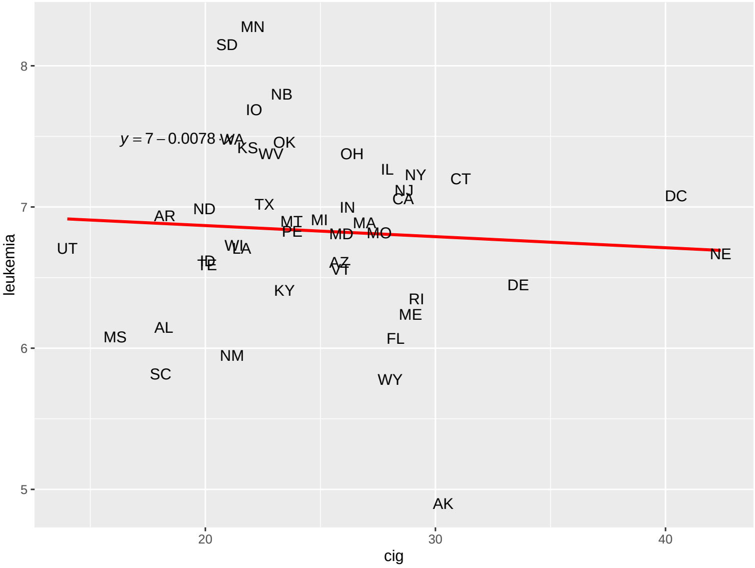
Smoking and kidney cancer



Smoking and bladder cancer



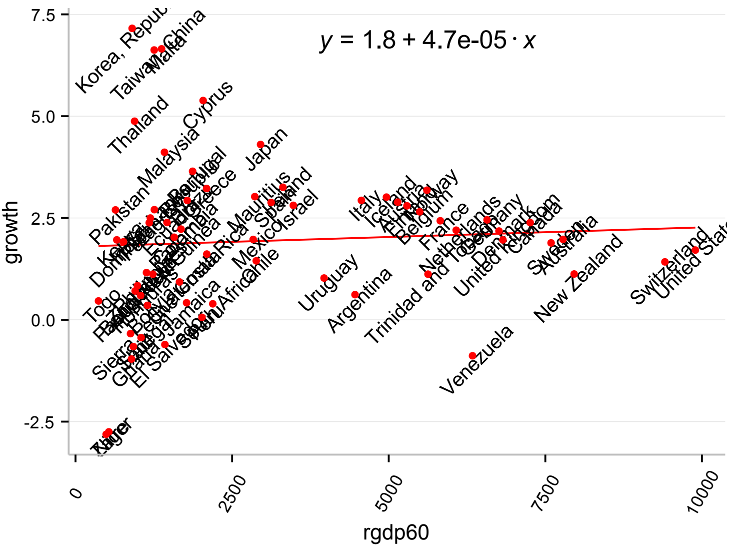
Smoking and leukemia



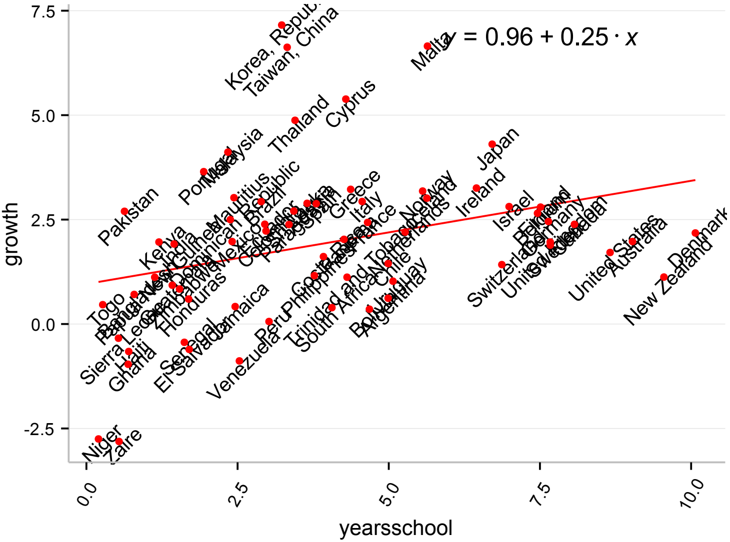
Example: convergence in growth

- Data on average growth rate from 1960-1995 for 65 countries along with GDP in 1960, average years of schooling in 1960, and other variables
- From `http://wps.aw.com/aw_stock_ie_2/50/13016/3332253.cw/index.html`, originally used in Beck, Levine, and Loayza (2000)
- Question: has there been in convergence, i.e. did poorer countries in 1960 grow faster and catch-up?
- Code

GDP in 1960 and growth



Years of schooling in 1960 and growth



- Things look different 1995-2014
- Code to download and recreate results using updated growth data through 2014 from the World Bank

Section 9

Inference

Inference with normal errors

- Regression estimates depend on samples, which are random, so the regression estimates are random
    - Some regressions will randomly look "interesting" due to chance
- Logic of hypothesis testing: figure out probability of getting an interesting regression estimate due solely to change
- Null hypothesis, H_0 : the regression is uninteresting, usually $\beta_1 = 0$

Inference with normal errors

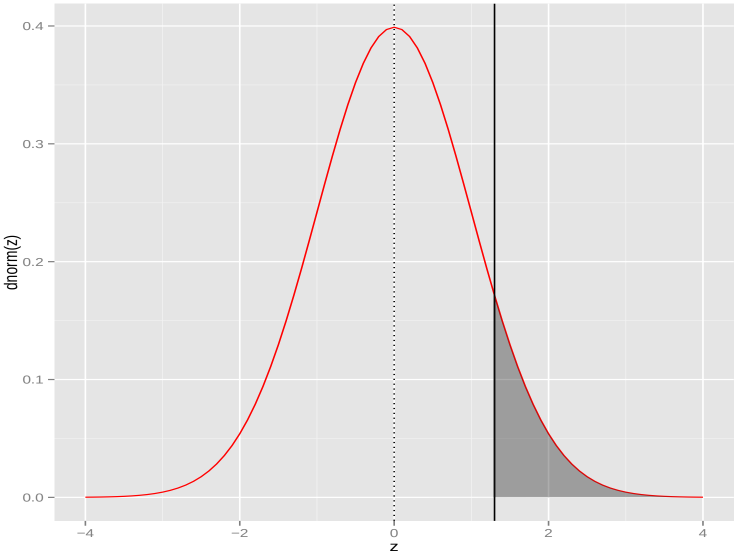
- With assumptions SR.1-SR.6 and under $H_0 : \beta_1 = \beta_1^*$, we know

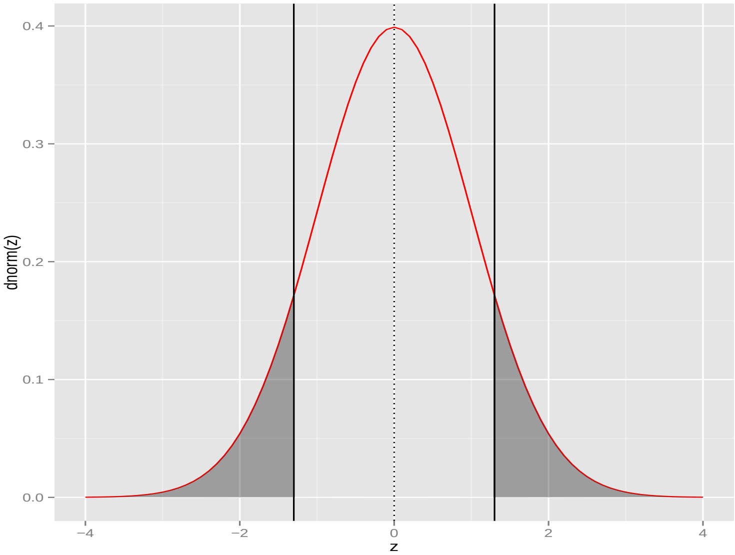
$$\hat{\beta} \sim N\left(\beta_1^*, \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

or equivalently,

$$t \equiv \frac{\hat{\beta} - \beta_1^*}{\sigma_\epsilon / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \sim N(0,1)$$

- P-value: the probability of getting a regression estimate as or more "interesting" than the one we have
  - As or more interesting = as far or further away from β_1^*
  - If we are only interested when $\hat{\beta}_1$ is on one side of β_1^*, then we have a one sided alternative, e.g. $H_a : \beta_1 > \beta_1^*$
  - If we are equally interested in either direction, then $H_a : \beta_1 \neq \beta_1^*$


0.4
0.3
0.2
0.1
0.0
dnorm(z)
−4
−2
0
2
4
z


0.4
0.3
0.2
0.1
0.0
dnorm(z)
−4
−2
0
2
4
z

Inference with normal errors

- One-sided p-value: $p = \Phi(-|t|) = 1 - \Phi(|t|)$
- Two-sided p-value: $p = 2\Phi(-|t|) = 2(1 - \Phi(|t|))$
- Interpretation:
  - The probability of getting an estimate as strange as the one we have if the null hypothesis is true.
  - It is *not* about the probability of β_1 being any particular value. β_1 is not a random variable. It is some unknown number. The data is what is random. In particular, the p-value is *not* the probability that that H_0 is false given the data.
- Hypothesis testing: we must make a decision (usually reject or fail to reject H_0)
  - Choose significance level α (usually 0.05 or 0.10)
  - Construct procedure such that if H_0 is true, we will incorrectly reject with probability α
  - Reject null if p-value less than α

Smoking and cancer

	Model 1	Model 2	Model 3	Model 4
(Intercept)	1.09^{*}	6.47^{**}	1.66^{***}	7.03^{***}
	(0.48)	(2.14)	(0.32)	(0.45)
cig	0.12^{***}	0.53^{***}	0.05^{***}	-0.01
	(0.02)	(0.08)	(0.01)	(0.02)
R^2	0.50	0.49	0.24	0.00
Adj. R^2	0.48	0.47	0.22	-0.02
Num. obs.	44	44	44	44
RMSE	0.69	3.07	0.46	0.64

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Smoking and cancer

Growth and GDP

	Model 1	Model 2
(Intercept)	1.80***	0.96*
	(0.38)	(0.42)
rgdp60	0.00	
	(0.00)	
yearsschool		0.25**
		(0.09)
R^2	0.00	0.11
Adj. R^2	-0.01	0.10
Num. obs.	65	65
RMSE	1.91	1.80

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Growth and GDP and education in 1960

Caution: multiple testing

- We just looked at 6 regressions, if $H_0 : \beta_1 = 0$ is true in all of them the probability that correctly fail to reject all 6 null hypotheses with a 5% test is $0.95^6 = 0.74$ (assuming the 6 tests are independent)
- A quarter of the time if we look at 6 regressions, we will randomly find at least significant relationship; if we look at 14 regressions the probability that we incorrectly reject a null is more than 0.5

Caution: economic significance $\neq$ statistical significance

Estimating σ_ϵ^2

- Recall that $\mathrm{Var}(\hat{\beta}|x_1, ..., x_n) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma_\epsilon^2}{n\widehat{\mathrm{Var}}(x)}$
- σ_ϵ^2 unknown
- We estimate σ_ϵ^2 using the residuals,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-2} \sum_{i=1}^n \underbrace{\hat{\epsilon}_i^2}_{=(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$$

- If SLR.1-SLR.5, $\mathrm{E}[\hat{\sigma}_\epsilon^2] = \sigma_\epsilon^2$
- Using $\frac{1}{n-2}$ instead of $\frac{1}{n}$ makes $\hat{\sigma}_\epsilon^2$ unbiased
  - $\hat{\epsilon}_i$ depends on 2 estimated parameters, $\hat{\beta}_0$ and $\hat{\beta}_1$, so only $n-2$ degrees of freedom
- Estimate $\mathrm{Var}(\hat{\beta}_1|x_1, ..., x_n)$ by

$$\widehat{\mathrm{Var}}(\hat{\beta}_1|x_1, ..., x_n) = \frac{\hat{\sigma}_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{1}{n-2}\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Estimating σ_ϵ^2

- Standard error of $\hat{\beta}_1$ is $\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 | x_1, ..., x_n)}$
- If SLR.1-SLR.6, t-statistic with estimated $\widehat{\text{Var}}(\hat{\beta}_1 | x_1, ..., x_n)$ has a $t(n-2)$ distribution instead of $N(0,1)$

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 | x_1, ..., x_n)}} \sim t(n-2)$$

Confidence intervals

- $\hat{\beta}_1$ is random
- $\widehat{\text{Var}}(\hat{\beta}_1)$, p-values, and hypthesis tests are ways of expressing how random is $\hat{\beta}_1$
- Confidence intervals are another
- A $1-\alpha$ confidence interval, $CI_{1-\alpha} = [LB_{1-\alpha}, UB_{1-\alpha}]$ is an interval estimator for β_1 such that

$$\mathrm{P}(\beta_1 \in CI_{1-\alpha} = 1 - \alpha)$$

($CI_{1-\alpha}$ is random; β_1 is not)

- Recall: if SLR.1-SLR.6, then

$$\hat{\beta}_1 \sim N\left(\beta_1, \text{Var}(\hat{\beta}_1)\right)$$

Confidence intervals

- Implies

$$P\left(\hat{\beta}_1 < \beta_1 + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(\alpha/2)\right) = \alpha/2$$
$$P\left(\hat{\beta}_1 - \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(\alpha/2) < \beta_1\right) = \alpha/2$$

and

$$P\left(\hat{\beta}_1 > \beta_1 + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(1-\alpha/2)\right) = \alpha/2$$
$$P\left(\hat{\beta}_1 - \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(1-\alpha/2) > \beta_1\right) = \alpha/2$$

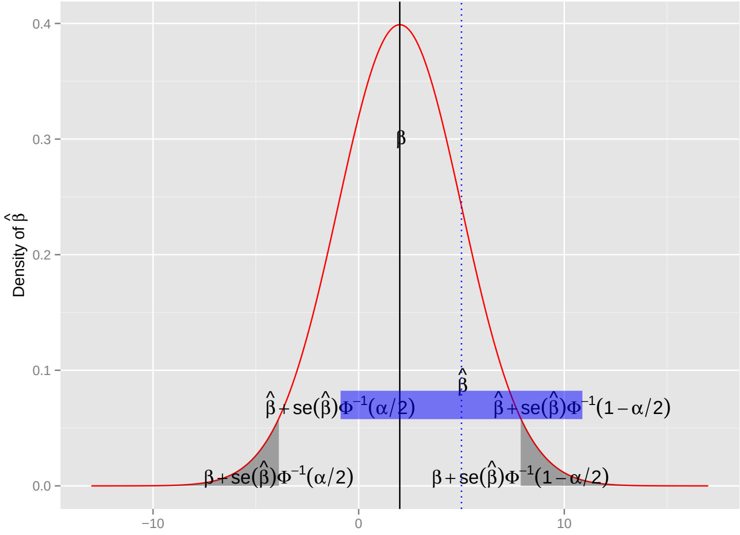
Confidence intervals

so

$$
\begin{aligned}
&\mathrm{P}\begin{pmatrix} \hat{\beta}_1 + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(\alpha/2) < \beta_1 \\ \beta_1 < \hat{\beta} + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(1-\alpha/2)\end{pmatrix} = \\
=&1 - \mathrm{P}\left(\hat{\beta}_1 + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(\alpha/2) < \beta_1\right) - \\
&- \mathrm{P}\left(\hat{\beta}_1 + \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(1-\alpha/2) > \beta_1\right) \\
=&1-\alpha
\end{aligned}
$$

- For $\alpha = 0.05$, $\Phi^{-1}(0.025) \approx -1.96$, $\Phi^{-1}(0.975) \approx 1.96$
- For $\alpha = 0.1$, $\Phi^{-1}(0.05) \approx -1.64$

Confidence intervals



Confidence intervals

- $1-\alpha$ confidence interval

$$\hat{\beta}_1 \pm \sqrt{\mathrm{Var}(\hat{\beta}_1)}\Phi^{-1}(\alpha/2)$$

- With estimated $\hat{\sigma}_\epsilon^2$, use t distribution instead of normal

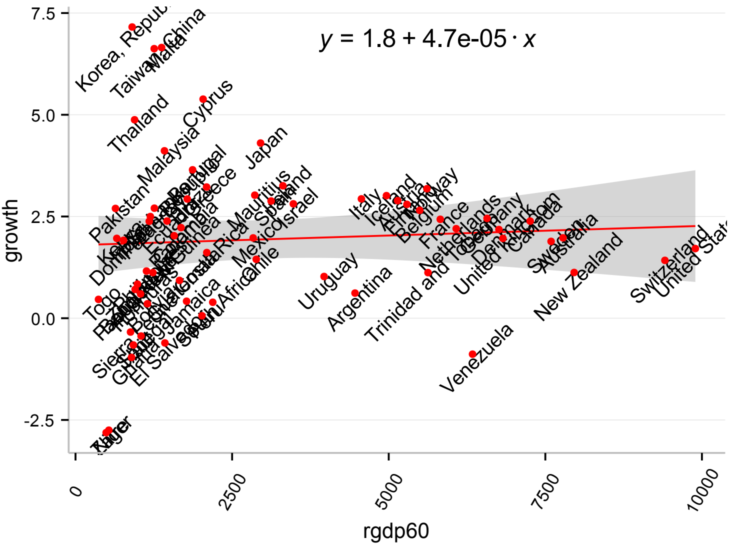
$$\hat{\beta}_1 \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}F_{t,n-2}^{-1}(\alpha/2)$$

$F_{t,n-2}^{-1}$ = inverse CDF of $t(n-2)$ distribution

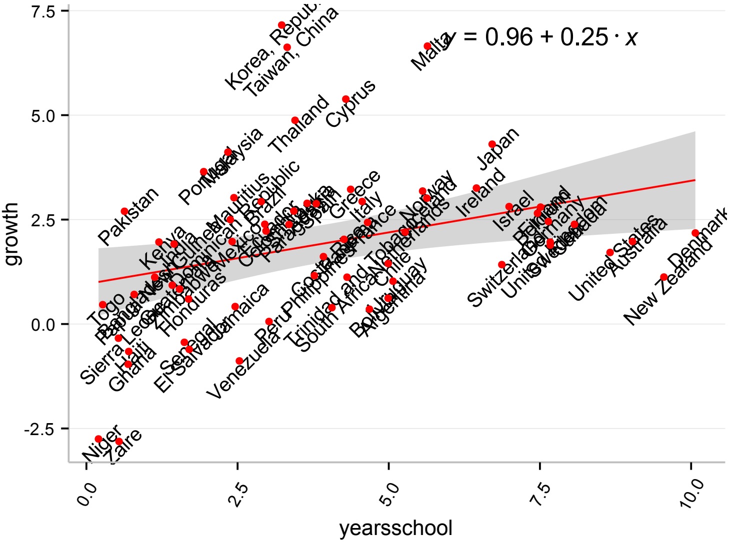
$F_{t,n-2}(1-\alpha/2)$

	$n-2$					
$\alpha/2$	5	10	20	50	100	∞
0.025	2.57	2.23	2.09	2.01	1.98	1.96
0.05	2.02	1.81	1.72	1.68	1.66	1.64

Example: GDP in 1960 and growth



Example: Years of schooling in 1960 and growth



Section 10

Efficiency

Gauss-Markov theorem

- The sample regression estimator,

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

  also called Ordinary Least Squares (OLS) is not the only unbiased estimator of

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

- Gauss-Markov theorem: if SLR.1-SLR.5, then OLS is the Best Linear Unbiased Estimator
  - Linear means linear in y, $\hat{\beta}_1 = \sum_{i=1}^n c_i y_i$ with $c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$
  - Unbiased means $\mathrm{E}[\hat{\beta}_1] = \beta_1$
  - Best means that among all linear unbiased estimators, OLS has the smallest variance

Proof: setup

- Let $\tilde{\beta}_1$ be a linear unbiased estimator of β_1
  - Linear: $\tilde{\beta}_1 = \sum_{i=1}^n c_i y_i$
  - Unbiased: $\mathrm{E}[\tilde{\beta}_1 | x_1, ..., x_n] = \beta_1$ (for all possible β_0, β_1)
- We will show that

$$\mathrm{Var}(\tilde{\beta}_1 | x_1, ..., x_n) \geq \mathrm{Var}(\hat{\beta}_1 | x_1, ..., x_n)$$

Proof: outline

1. Show that $\sum_{i=1}^n c_i = 0$ and $\sum_{i=1}^n c_i x_i = 1$
2. Show $\mathrm{Cov}(\tilde{\beta}_1, \hat{\beta}_1 | x_1, ..., x_n) = \mathrm{Var}(\hat{\beta}_1 | x_1, ..., x_n)$
3. Show $\mathrm{Var}(\tilde{\beta}_1 | x_1, ..., x_n) \geq \mathrm{Var}(\hat{\beta}_1 | x_1, ..., x_n)$
4. Show $\mathrm{Var}(\tilde{\beta}_1 | x_1, ..., x_n) = \mathrm{Var}(\hat{\beta}_1 | x_1, ..., x_n)$ only if $\tilde{\beta}_1 = \hat{\beta}_1$

[0] We will go over the proof in class. See Marmer's slides or Wooldridge (2013) 3A for details

References

Abbring, Jaap. 2001. "An Introduction to Econometrics: Lecture notes." URL http://jabbring.home.xs4all.nl/courses/b44old/lect210.pdf.

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Baltagi, BH. 2002. *Econometrics*. Springer, New York. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001086635&title=Econometrics.

References

Beck, T., R. Levine, and N. Loayza. 2000. "Finance and the Sources of Growth." *Journal of financial economics* 58 (1):261–300. URL http://www.sciencedirect.com/science/article/pii/S0304405X00000726.

Bierens, Herman J. 2012. "The Two-Variable Linear Regression Model." URL http://personal.psu.edu/hxb11/LINREG2.PDF.

Diez, David M, Christopher D Barr, and Mine Cetinkaya-Rundel. 2012. *OpenIntro Statistics*. OpenIntro. URL http://www.openintro.org/stat/textbook.php.

Linton, Oliver B. 2017. *Probability, Statistics and Econometrics*. Academic Press. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001868500&title=Probability%2C%20statistics%20and%20econometrics.

References

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.