Logarithms in Regression Analysis with Asiaphoria and Dummy Variables

Lecture 6

Reading: "Logarithms in Regression Analysis with Asiaphoria," Section 7.8

1

Outline for Today

- Using natural logarithm transformations to straighten scatter plots
  - *Interpreting* coefficients when x and/or y are logged
  - Asiaphoria and Penn World Tables data
    - The power and convenience of natural logs
- Regression when the x-variable is a dummy
  - Combine everything: ln(y) and x is a dummy

2

2017 World Happiness Report (WHR)

- "Happiness is increasingly considered the proper measure of social progress and the goal of public policy. In June 2016, the OECD committed itself 'to redefine the growth narrative to put people's well-being at the centre of governments' efforts.' In a recent speech, the head of the UN Development Program (UNDP) spoke against what she called the 'tyranny of GDP,' arguing that what matters is the quality of growth." p. 3

http://worldhappiness.report/wp-content/uploads/sites/2/2017/03/HR17.pdf

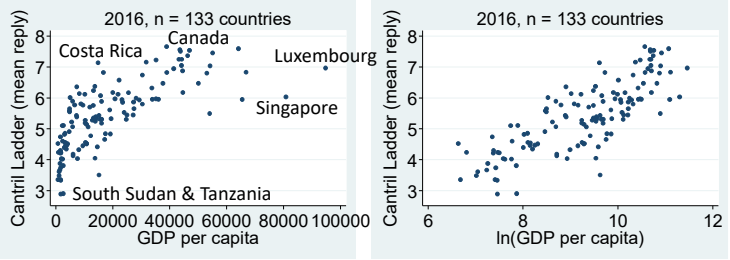
3

Worldwide Happiness Survey

- Huge annual survey in 150+ countries, each w/ ~1,000 respondents, measuring happiness
  - Cantril ladder: "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?"

4

Straightening a Scatter Plot



Variable	Mean	S.D.	Median	Min.	Max.	25th Per.	75th Per.
Cantril Ladder	5.41	1.13	5.43	2.89	7.66	4.52	6.12
GDP per capita	18,950	18,521	13,178	760	94,774	4,731	27,453
Ln(GDP per capita)	9.28	1.19	9.49	6.63	11.46	8.46	10.22

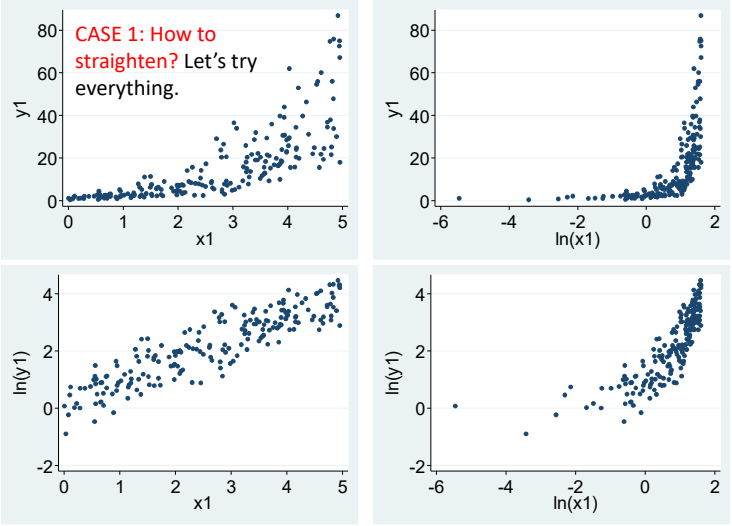
"GDP per capita is in terms of Purchasing Power Parity (PPP) adjusted to constant 2011 international dollars, taken from the World Development Indicators (WDI) released by the World Bank in August 2016." p. 17 of 2017 WHR

5

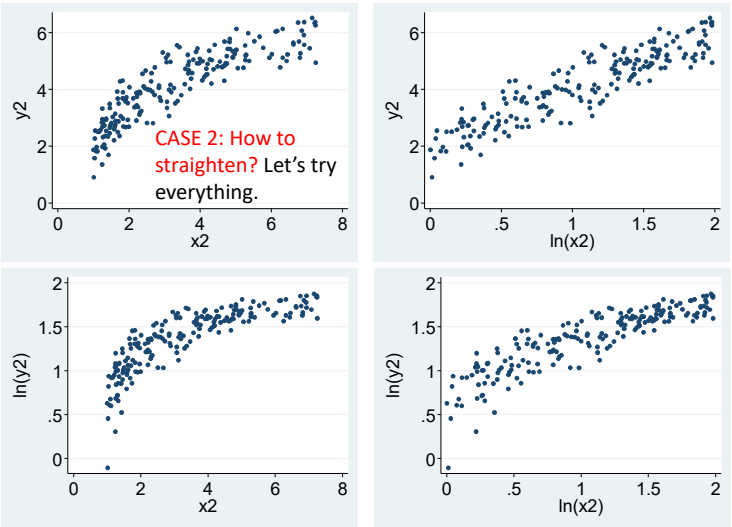
What to Log?

- Given a curved relationship, how do we know whether to take the natural log of x, the natural log of y, or the natural log of both?
  - A quick rule of thumb: apply the natural log to the positively skewed variable
  - You can think about which variable would have diminishing returns (applying the natural log to it)
  - Of course, there's always trial-by-error, so long as you know how to use software (which you do)

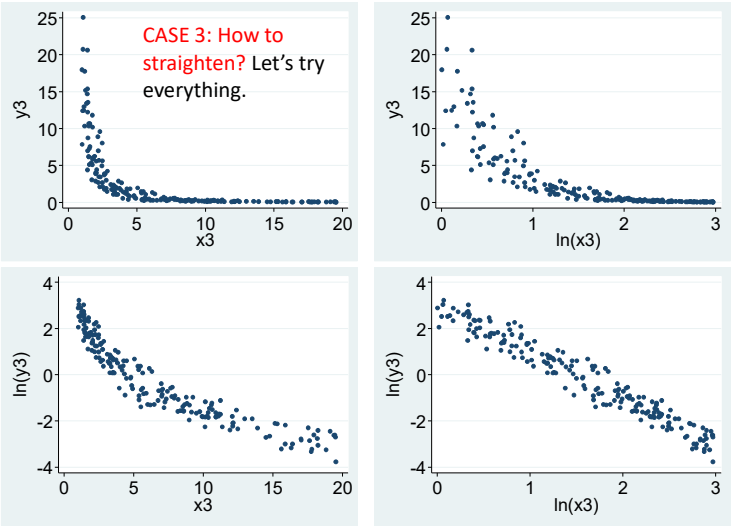
6


CASE 1: How to straighten? Let's try everything.
y1
x1
ln(x1)
ln(y1)

7


CASE 2: How to straighten? Let's try everything.
y2
x2
ln(x2)
ln(y2)

8


CASE 3: How to straighten? Let's try everything.
y3
x3
ln(x3)
ln(y3)

9

Functional Form

- Some common *functional forms* in economics:
  - Linear: $\hat{Y} = a + bX$
  - Log-log (constant elasticity): $\widehat{\ln(Y)} = a + b\ln(X)$
  - Semi-log (log-lin): $\widehat{\ln(Y)} = a + bX$
  - Lin-log: $\hat{Y} = a + b\ln(X)$
- Next, review first two in a demand context and then summarize all four
  - "Logarithms in Regression Analysis with Asiaphoria"

10

Review: Elasticity of Demand

- Recall ECO101/102: $elasticity = \varepsilon = \eta = \frac{\%\,\Delta Q^D}{\%\,\Delta P}$
- With calculus: $= \frac{100*\frac{dQ}{Q}}{100*\frac{dP}{P}} = \frac{dQ}{dP}\frac{P}{Q}$

See Question 8 in the Diagnostic Quiz in the Prerequisite Review

Table III. Own and Cross Price Elasticities

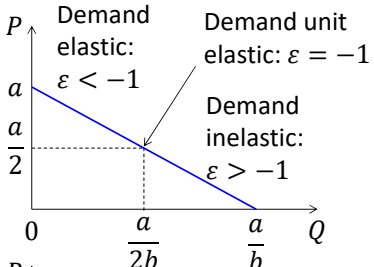
		with respect to the price of		
		Cottonelle	Charmin	Angel Soft
Elasticity of the demand for	Cottonelle	-3.304	0.737	0.621
	Charmin	0.242	-2.292	0.262
	Angel Soft	0.765	1.132	-4.066

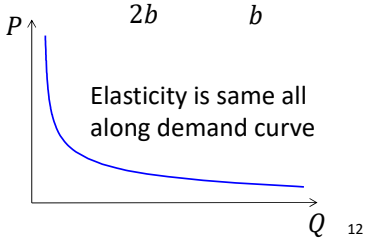
Hausman and Leonard (2002) "The Competitive Effects of a New Product Introduction: A Case Study" p. 251 https://onlinelibrary.wiley.com/doi/epdf/10.1111/1467-6451.00176

11

Demand Specifications

- Linear specifications of demand:
  $Q = \frac{a}{b} - \frac{1}{b}P$
  - Graph: $P = a - bQ$
  - Elasticity: $\varepsilon = -\frac{P}{a-P}$
- Constant elasticity function form:
  $\ln(Q) = a - b\ln(P)$
  - Elasticity $\varepsilon = -b$

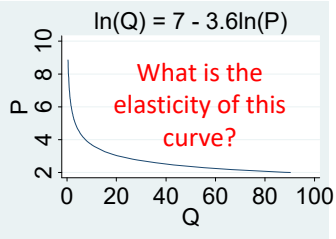




12

Constant Elasticity (log-log; log-linear)

$$\ln(Q) = a - b\ln(P)$$
$$\exp(\ln(Q)) = \exp(a - b\ln(P))$$
$$Q = \exp(a)\exp(-b\ln(P))$$
$$Q = \exp(a)\exp\left(\ln\left(P^{-b}\right)\right)$$
$$Q = e^{a}P^{-b}$$



$$elasticity = \frac{\%\,\Delta Q}{\%\,\Delta P} = \frac{100*\frac{dQ}{Q}}{100*\frac{dP}{P}} = \frac{dQ}{dP}\frac{P}{Q}$$

$$elasticity = \frac{dQ}{dP}\frac{P}{Q} = e^{a}(-b)P^{-b-1}\frac{P}{e^{a}P^{-b}} = -b$$

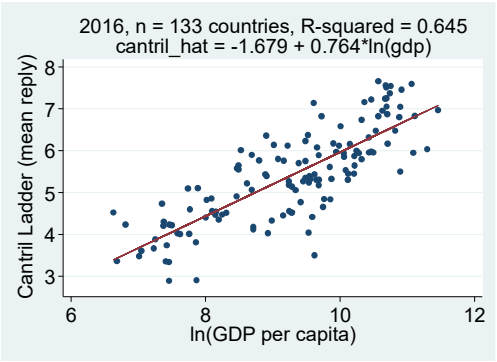
13

Regression Result	**What is b_1?**
$\hat{y}_i = b_0 + b_1 x_i$	Slope: a one unit Δ in x is associated with a b_1 unit Δ in y on average
$\widehat{\ln(y_i)} = b_0 + b_1\ln(x_i)$	Elasticity: a one percent Δ in x is associated with approximately a b_1 percent Δ in y on average
$\widehat{\ln(y_i)} = b_0 + b_1 x_i$	"Semi-elasticity": a one unit Δ in x is associated with approximately a 100*b_1 percent Δ in y on average
$\hat{y}_i = b_0 + b_1\ln(x_i)$	(no name): a one percent Δ in x is associated with approximately a $b_1/100$ unit Δ in y on average

Note well: A full interpretation must specify the context, give variable descriptions, be clear on causality, and provide the specific units of measurement of any non-logged variables.

14

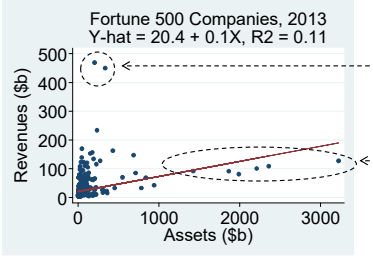
Interpret the Results



Do these data have summary values (Section 19.4)?

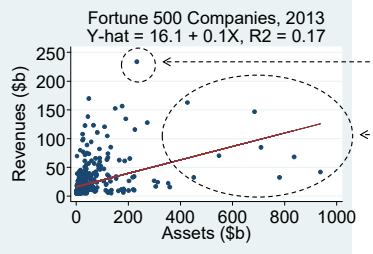
For countries with GDP per capita that is 10% higher, we observe mean happiness, which is measured on a 10-point Cantril ladder scale, that is approximately 0.08 points higher on average.

15



Wal-Mart & Exxon Mobile

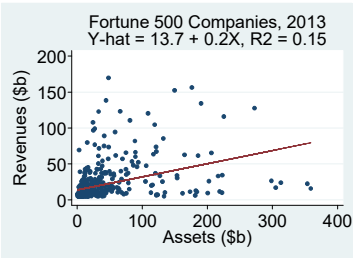
Fannie Mae, J.P. Morgan Chase, Bank of America, Freddie Mac, Citigroup, Wells Fargo



Chevron

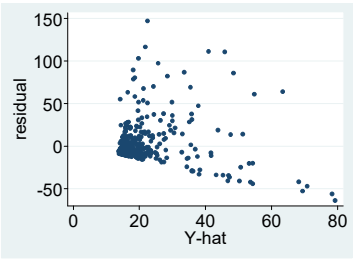
Goldman Sachs, MetLife, Morgan Stanley, Prudential Financial, GE, AIG, TIAA-CREF, Berkshire Hathaway

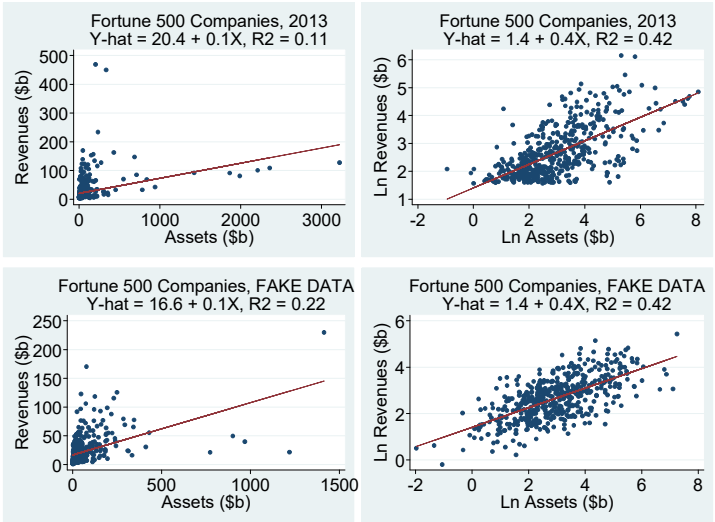
http://money.cnn.com/magazines/fortune/fortune500/2013/full_list/



Clearly the problem is not outliers as it may have first appeared.

Instead, the problem is that the functional form is not linear so linear descriptive techniques (OLS, R^2) are NOT appropriate.

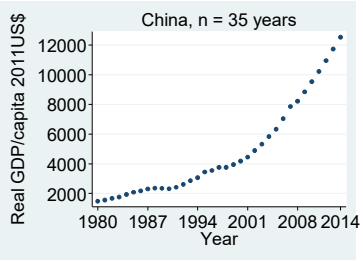


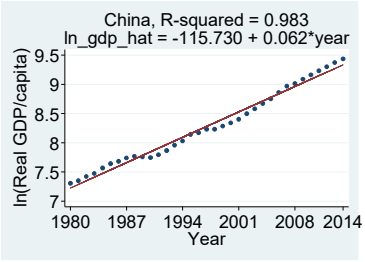


Penn World Tables (PWT) 9.0

- High-quality publicly available data, which are periodically updated
  - Pritchett and Summers (2014) "Asiaphoria Meets Regression to the Mean" use PWT 8.0
  - It contains country-level GDP measures – more than one variable measuring GDP depending on researcher's purpose – for each year
    - Are these data cross-sectional, time series or panel?
  - Can use these data to compute growth rates

Getting from GDP levels to growth



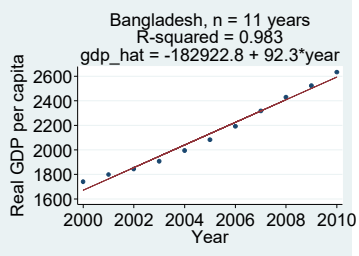


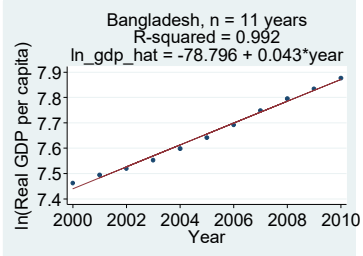
How to interpret 0.062?

How to interpret 0.983?

"Real GDP/capita" is rgdpna/pop from PWT 9.0 (DACM): rgdpna is "Real GDP at constant 2011 national prices (in mil. 2011US$)" and pop is "Population (in millions)."

Sometimes ln() is just convenient





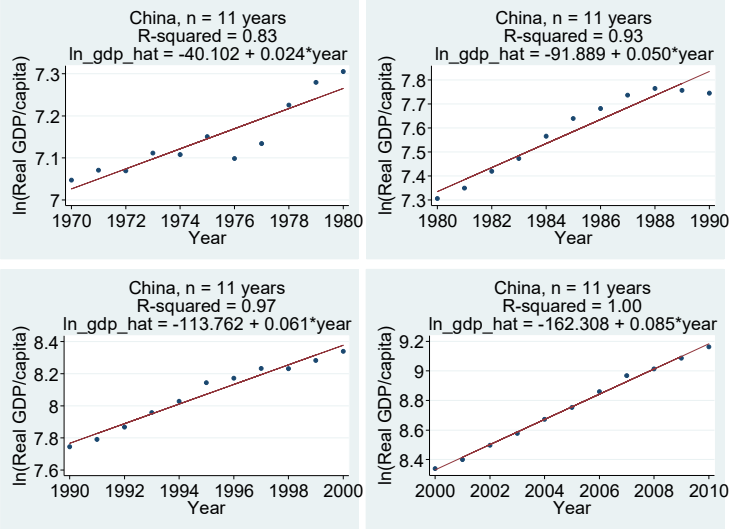
How to interpret 92.3?

How to interpret 0.043?

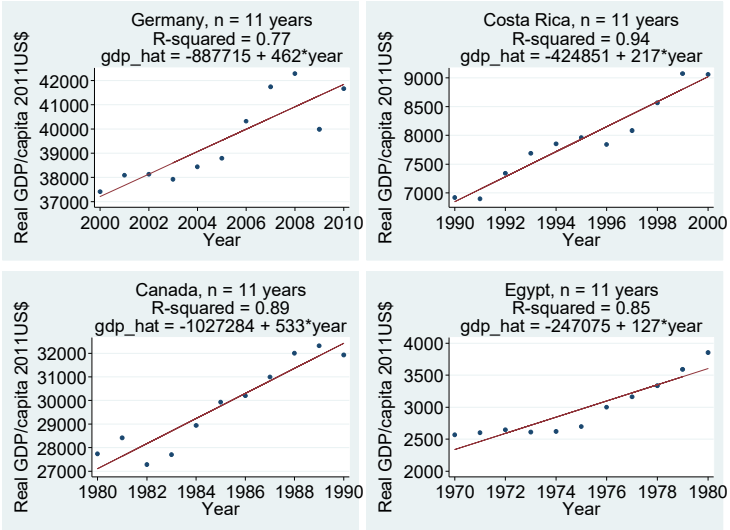
Pritchett and Summers (2014)

- One goal of the paper is to assess how well past growth rates predict future growth rates using a cross-section of countries
- To obtain the data on GDP growth rates, they run many regressions and retrieve the OLS coefficients: these populate the variables that measure growth rates in each decade
  - Growth *rates* allow cross-country comparisons

22



Repeat for each of 141 other countries to get growth rates? 23



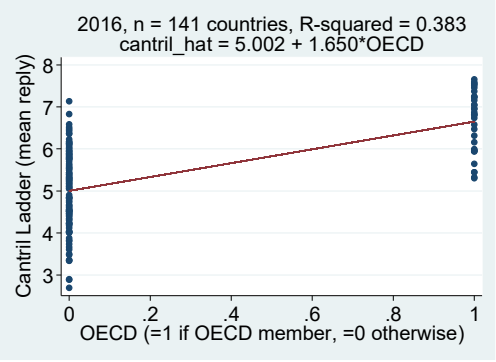
Are growth *levels* comparable across countries? Over time? 24

Regression when x is a dummy

- Recall that a dummy variable (also called an indicator variable or a fixed effect) takes only two possible values: 0 and 1
- Codes categorical information so we can use methods (e.g. OLS, correlation, mean, s.d., etc.) usually reserved for interval variables
  - E.g. we can have a dummy variable for sex named female that equals 1 for females and 0 otherwise

25

Happier in the OECD?



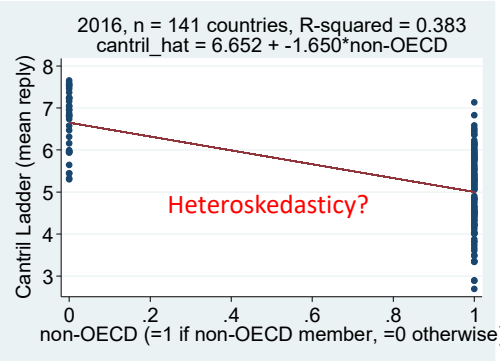
How to interpret 5.002?

Are these data observational or experimental?

How to interpret 1.650?

26

What if reverse definition of dummy?

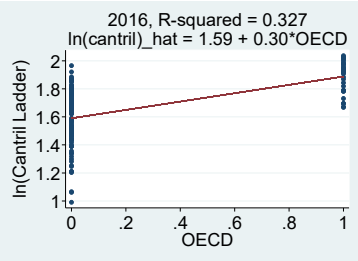


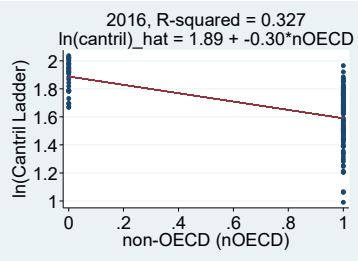
How to interpret 6.652?

How to interpret -1.650?

27

Natural Logs and Dummies





(6.652 – 5.002)/5.002 = 0.33 (5.002 – 6.652)/6.652 = –0.25

Remember: Interpretations with logs are approximate.

Cantril Ladder	Mean	S.D.	Median	Min.	Max.
34 OECD countries	6.652	0.730	6.824	5.303	7.660
107 non-OECD countries	5.002	0.950	5.100	2.693	7.136

28