

ECO220Y

Regression Diagnostics

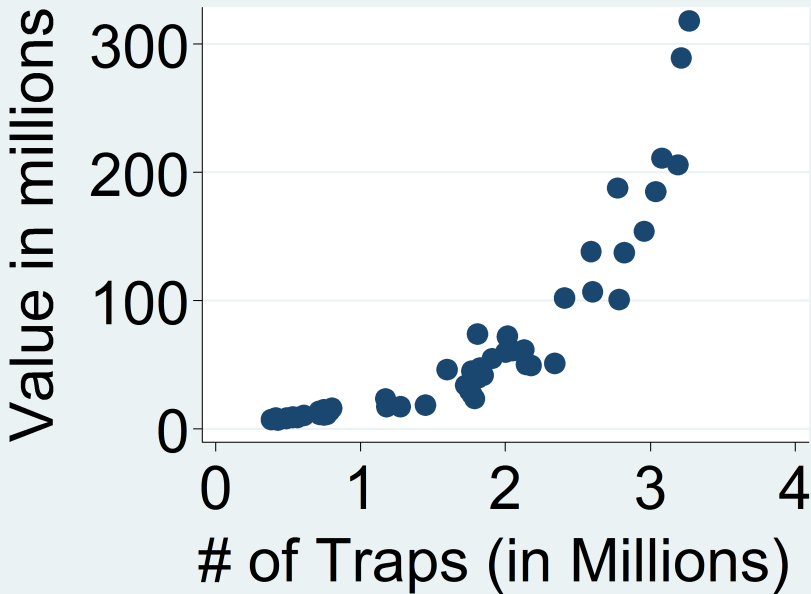
Readings: Chapter 19

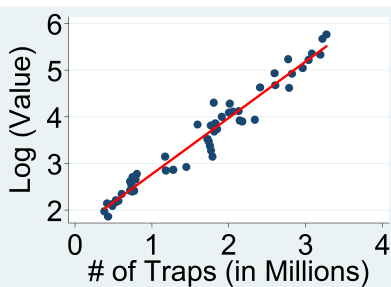
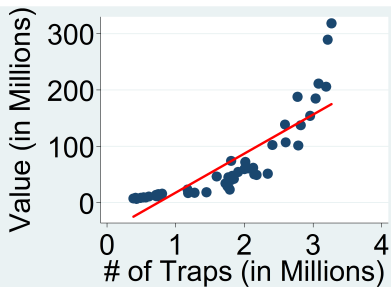
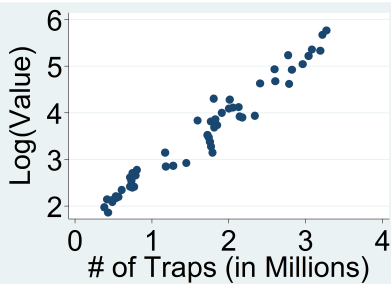
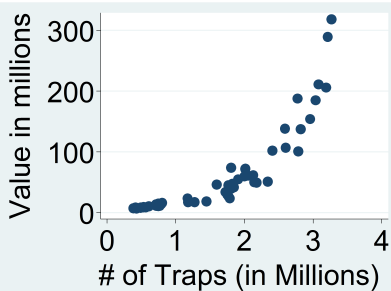
Spring 2012

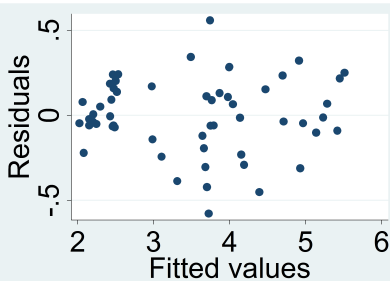
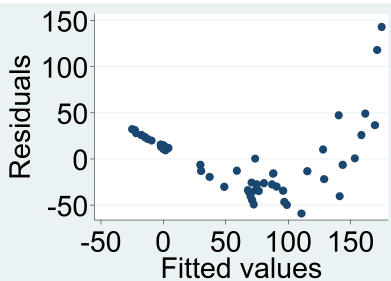
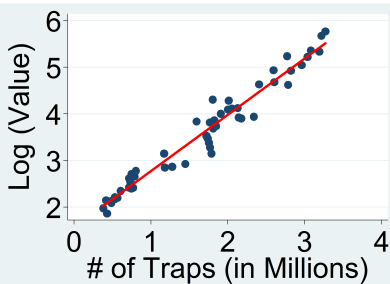
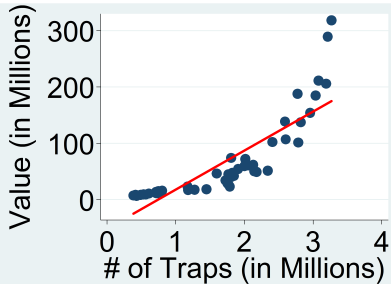
Lecture 20 Part 2

Lobster Industry

- Chapter 19, Ex. 33-36
- Lobster Industry in Maine
- Annual data from 1950 through 2005
- The value of the annual lobster catch (“value”), the number of lobster traps (“traps”), the number of licensed lobster fishers (“fisher”), the price of lobster (“price”)



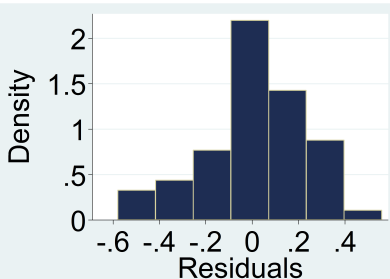
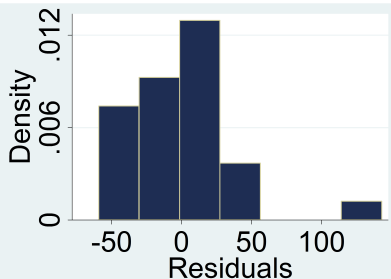
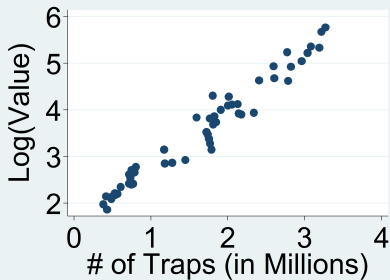
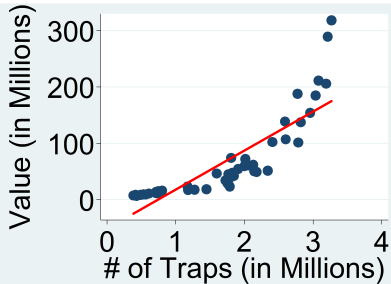




Misspecified Model \Rightarrow Heteroscedasticity

- Consequences:
 - ▶ Does not bias the coefficients' estimates
 - ▶ Does not bias the R^2
 - ▶ Does bias standard errors of coefficient estimates
- Correction:
 - ▶ Use robust standard errors
 - ▶ Robust to heteroscedasticity
 - ▶ Implemented in Stata by typing "robust"

Heteroscedasticity: affects point estimate? Affects interval estimate?
Affects t-stat? Affects p -value?



Source	SS	df	MS	Number of obs = 56		
Model	62.7565362	1	62.7565362	F(1, 54) = 1275.33	Prob > F = 0.0000	
Residual	2.65723489	54	.049208053	R-squared = 0.9594	Adj R-squared = 0.9586	
Total	65.4137711	55	1.18934129	Root MSE = .22183		

logvalue	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
trap	1.208941	.0338527	35.71	0.000	1.14107	1.276812
_cons	1.560603	.062362	25.02	0.000	1.435574	1.685631

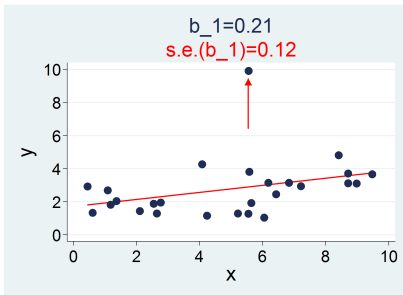
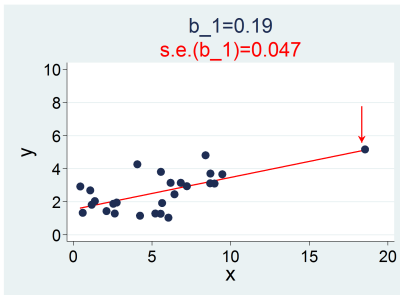
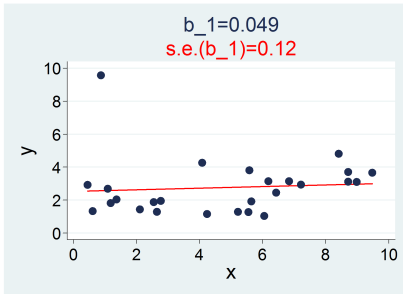
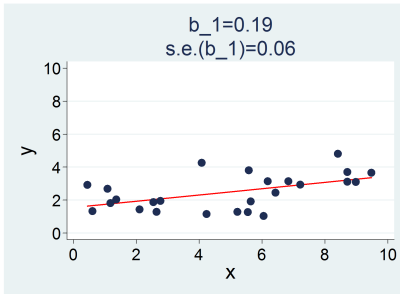
$$\text{Ln}(\text{Value})\text{-hat} = 1.56 + 1.21 * \text{Traps}$$

(0.062) (0.034)

- Interpretation of the slope?
- Can we infer causal relationship?
- Can we infer linear relationship between the value of the catch and the number of traps?

Outliers

- Outliers are observations substantially different from the bulk of data.
- Can be easily spotted on the scatter diagram.
- Origin:
 - ▶ Incorrect data entry
 - ▶ Confusing question
 - ▶ Non-sampling errors
 - ▶ **Valid data point - influential observation (next slide)**
- Should be handled with care



Finding and Dealing with Outliers

- Describe data with statistics and graphs:
 - ▶ Histograms, scatter diagrams
 - ▶ Check carefully variables' definitions (Ex: age never <0)
- If keep observation that is an outlier:
 - ▶ Give reason the observation is valid
- If drop observation that is an outlier:
 - ▶ Give reason the observation is invalid

Should you automatically drop outliers?