Sec 4.1

Determine $x \to y$ Relationship

↓

Scatterplot

↓

Numerical Summaries

(r, r^2, LSRL)

↓

Residual Plot

No Pattern/Random ↙ ↘ Pattern

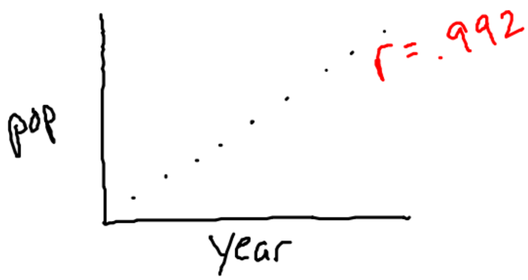
Linear $\hat{y} = a + bx$

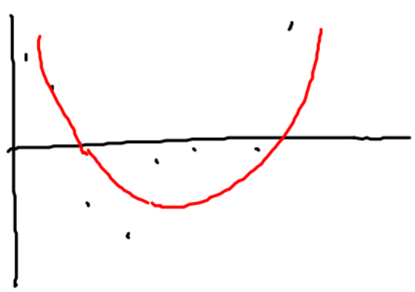
Exponential $\hat{y} = ab^x$

Power $\hat{y} = ax^b$

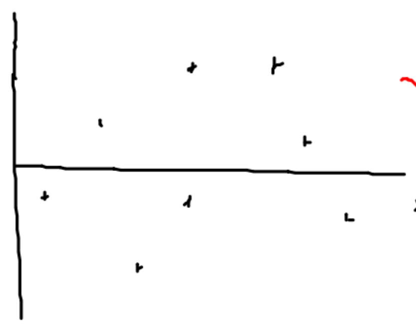
Ex US Population (Ex 4.11, p. 213)

L_1	L_2	L_3
year (1920)	pop	log y





Pattern ☹



No pattern ☺

Write exponential model:

a) Using Back Transformations

$$\hat{y} = -8.28 + .005x$$

$$\log_{10} \hat{y} = -8.28 + .005x$$

$$\hat{y} = 10^{-8.28 + .005x}$$

$$\hat{y} = (10^{-8.28})(10^{.005x})$$

$$\hat{Pop} = \underbrace{(10^{-8.28})}_{a}\underbrace{(10^{.005})}_{b}{}^{year}$$

Write exponential model:

a) Using Back Transformations

$\hat{y} = -8.28 + .005x$ > L_1, L_3

$\log_{10}\hat{y} = -8.28 + .005x$

$\hat{y} = 10^{-8.28 + .005x}$

$\hat{y} = (10^{-8.28})(10^{.005x})$

$$\hat{Pop} = \underbrace{(10^{-8.28})}_{a}\underbrace{(10^{.005})}_{b}{}^{year}$$

b) Using Calculator

$\boxed{\text{CALC}} \to$ ExpReg L_1, L_2, Y_1 } May Cause "Overflow Error"

$$\hat{y} = (.00000000529)(1.01)^x$$

$$\widehat{Pop} = (.00000000529)(1.01)^{year}$$

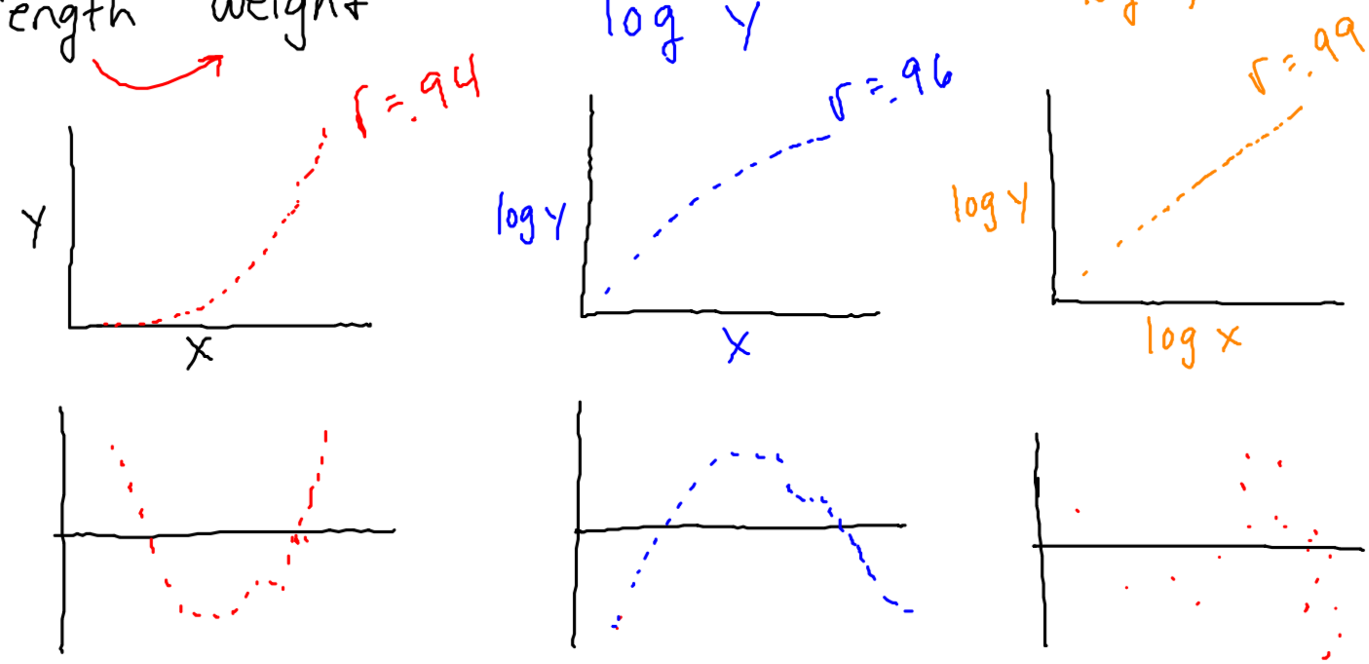
Exponential Regression Test

$L_2 = 105.7,\ 122.8,\ 131.7,\ 151.3,\ 179.3,\ \ldots\ 248.7,\ 281.4$

1.16   1.07   1.19   1.13

Power Regression (Rockfish, P.216)

L_1	L_2	L_3	L_4
length	weight	log Y	log X



Linear Exponential Power

Write Power Equation

a) Use backward transformations

$$\hat{y} = -1.899 + 3.049x$$

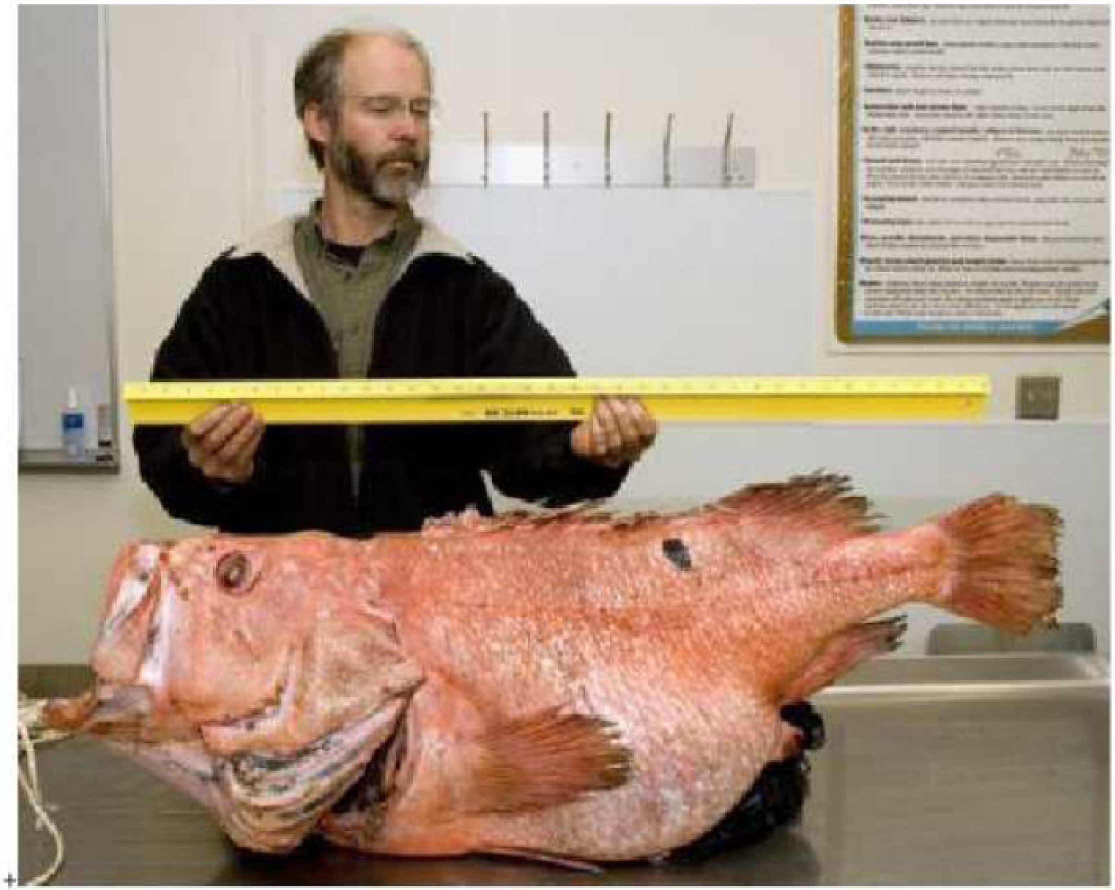
$$\log \hat{y} = -1.899 + 3.049(\log x)$$

$$\hat{y} = 10^{-1.899} x^{3.049}$$

$$\widehat{\text{Weight}} = (10^{-1.899})(\text{length})^{3.049}$$

b) Use Calculator

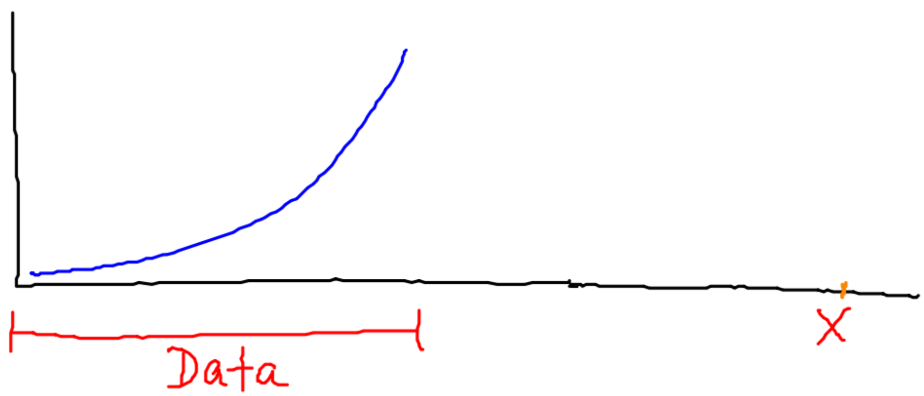
$\boxed{\text{CALC}} \rightarrow$ PwrReg L_1, L_2, Y_1 } Overflow Error?



Sec 4.2

Cautions About Correlation / Regression

1) Avoid Extrapolation

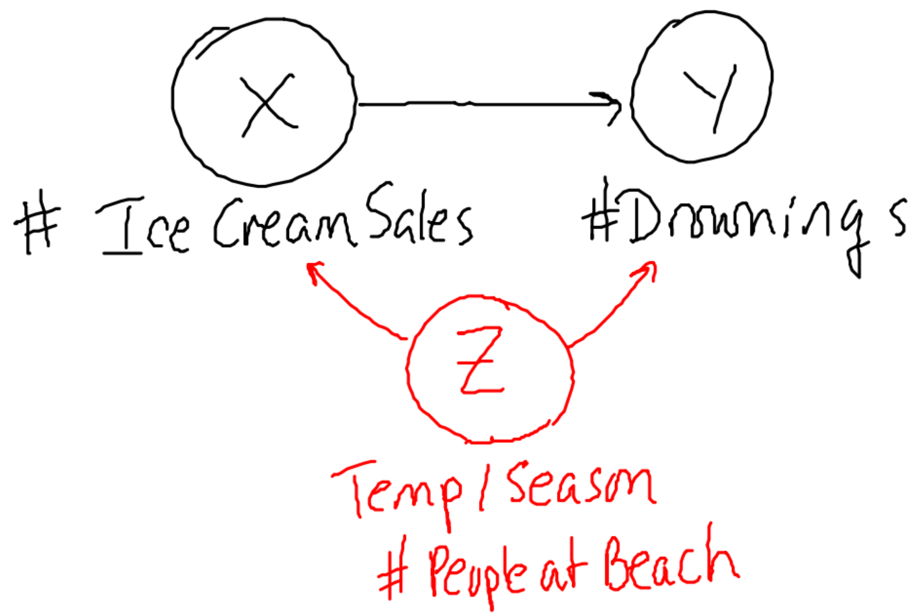


2) Lurking Variables

Variable(s) that is/are not an explanatory / response variable BUT may influence either/both

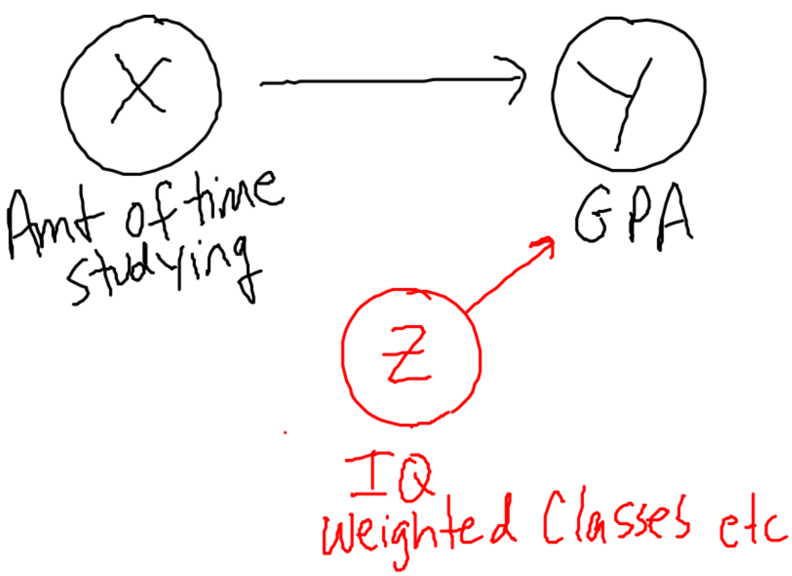
a) Common Response

Lurking variable which impacts both x and y



b) Confounding

Lurking variable which affects the response variable (y)



3) Strong Correlation $\neq$ Causation

- Experiments
- Correlational Studies

Sec 4.3

Relations In Categorical Data (%)

	Student Smokes	Student Not Smokes	
Both Parents Smoke	400	1380	1780
One Parent Smokes	416	1823	2239
Neither Smokes	188	1168	1356
	1004	4371	5375

Marginal Distributions

% of students smoke $= \frac{1004}{5375} = 18.68\%$

* Conditional Distributions

1) % of students smoke when both parents smoke $= \frac{400}{1780} = 22.47\%$
2) % of students smoke when neither parent smoke $= \frac{188}{1356} = 13.86\%$

Conclusion

- When both parents, student smoking increases from 14% to 22%

- Percent Increase $= \frac{8}{14} = 57\%$

Ex Cocaine Use

$1 \rightarrow 3 \rightarrow$ 200% increase

Simpson's Paradox

- Lurking variables change/reverse relationship

Surgery	A	B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

$$\text{Died} \mid \text{Hospital A} = \frac{63}{2100} = 3\%$$

$$\text{Died} \mid \text{Hospital B} = \frac{16}{800} = 2\%$$

Good Condition

	A	B
Died	6	8
Survived	594	592
	600	600

Poor Condition

	A	B
Died	57	8
Survived	1443	192

$$\text{Died} \mid \text{Hospital A} = \frac{6}{600} = 1\%$$

$$\text{Died} \mid \text{Hospital B} = \frac{8}{600} = 1.3\%$$

30% increase

Independence

Marginal Distribution = Conditional Distribution

Ex

	Y	N
M	5	5
F	5	5

$$P(\text{Male}) = \frac{10}{20} = 50\%$$

$$P(\text{Male} \mid \text{Yes}) = \frac{5}{10} = 50\%$$

Given That