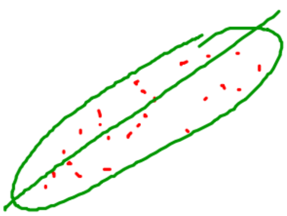
Sec 12.1

Inference For Regression

- Is there a linear association between 2 quantitative variables in a population?

- First, need to know if there is a linear relationship between 2 quantitative variables from a random sample ...

Given 2 (x,y) variables, is there a linear relationship?



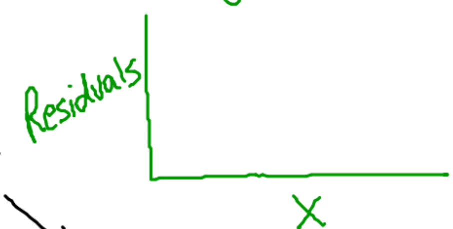
Scatterplot

↓

Correlation (LinReg $a+bx$)

↓

Residual Plot



No Pattern / Random	Pattern / Not Random
Linear Model ☺	No Linear Model ☹
$\hat{y} = a + bx$	

LINEAR REGRESSION T TEST

This test is used to determine if there is a linear relationship between 2 quantitative variables in a population

Child development researchers explored the relationship between the crying of infants 4 to 10 days old and their later IQ scores. The number of peaks in the most active 20 seconds of crying were counted and recorded. The tables contain data from a random sample of 36 infants.

Cry	10	12	9	16	18	15	12	20	16	33	20	16
IQ	87	97	103	106	109	114	119	132	136	159	90	100

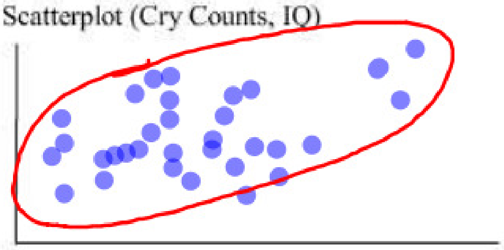
Cry	23	27	15	21	12	15	17	17	19	13	18	18
IQ	103	108	112	114	120	133	141	94	103	104	109	112

Cry	16	19	22	30	12	12	14	10	23	9	16	31
IQ	118	120	135	155	94	103	106	109	113	119	124	135

Do these data provide convincing evidence of a linear relationship between crying counts and IQ in the population of infants 4 to 10 days old?

DETERMINE IF THERE IS A LINEAR RELATIONSHIP FROM THE SAMPLE

1) Scatterplot (Cry Counts, IQ)

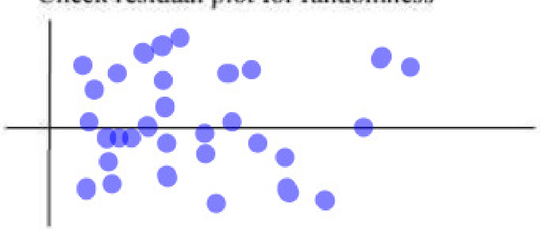


2) Calculate r and LSRL

$r = .5455$ (moderately weak positive linear relationship)

$\hat{y} = 87.91 + 1.55x \rightarrow \widehat{IQ} = 87.91 + 1.55(\text{Cry Count})$

3) Check residual plot for randomness



} Random

PERFORM TEST:

P) STATE POPULATION PARAMETER:

β = true slope of population regression line determined by crying counts and IQ

H) STATE HYPOTHESES:

H_o: $\beta = 0$

H_a: $\beta \neq 0$

A) VERIFY ASSUMPTIONS REQUIRED FOR TEST:

Linear- moderately strong linear relationship exists with a random residual plot

Independent- IQs independent and N > 10(36) > 360 infants

Normal (Residuals)- NPP of residuals linear and $n > 30$

Equal variance- residuals equally scattered around $x = 0$

Random- random sample taken

T) PERFORM TEST:

a) Using Formula:

$$t = \frac{b}{SE_b} \text{ with } df = n - 2 \text{ where } SE_b = \frac{\sqrt{\frac{\Sigma(y_i - \hat{y})^2}{n-2}}}{\sqrt{\Sigma(x_i - \bar{x})^2}}$$

b) Using Minitab Output?

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	87.9055	8.934	10.22	0.0000
Crycount	1.5517	0.4094	3.79	0.0005
s = 14.25	R-Sq = 29.8%		R-Sq (adj) = 28.5%	

c) Using Calculator:

STAT → TESTS → Lin Reg T Test → $t = 3.79$, $P\text{-}value = .0005$

S) STATE CONCLUSION:

At $\alpha = .05$, there is strong evidence to reject H_0 and conclude a linear relationship exists between crying and IQ in the population of infants

CONFIDENCE INTERVAL

A 99% confidence interval for the true population slope can be found using:

a) Formula:

$$CI = b \pm t^* SE_b = 1.55 \pm (2.750)(.4094) = (.43, 2.66)$$

Table B $\qquad SE_b = \dfrac{b}{t}$

b) Calculator?

STAT → TESTS → Lin Reg T Interval = (.43, 2.66)

We are 99% confident that for every cry count, IQ increases between .43 and 2.66 points

Calculator Note:

$H_0: \beta = 0 \qquad H_o: \rho = 0$

rho (population correlation)

If $\beta = 0$, $\rho = 0$
$\beta > 0$, $\rho > 0$
$\beta < 0$, $\rho < 0$

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

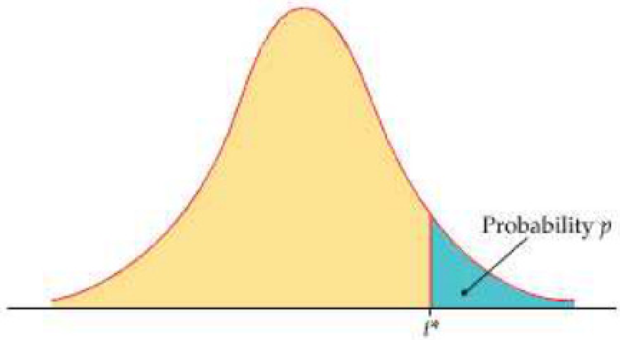


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Sec 12.2

Making Predictions

Linear Model

$\hat{y} = a + bx$

Exponential Model

$\hat{y} = ab^x$

Power Model

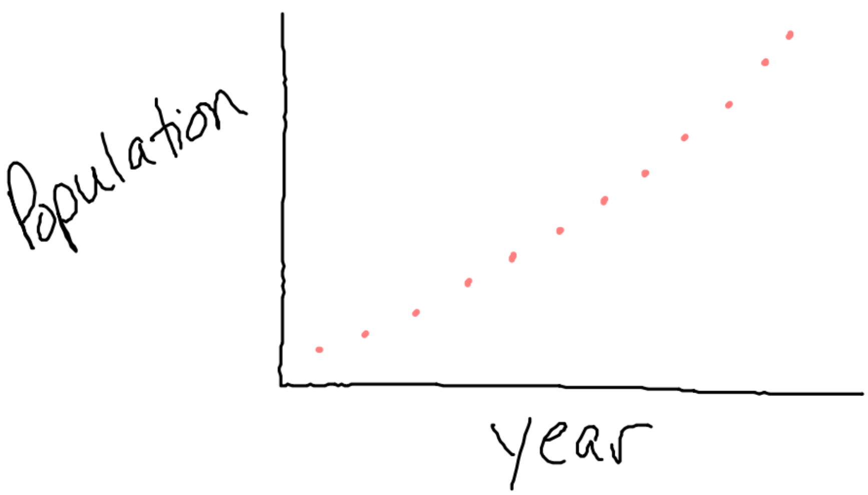
$\hat{y} = ax^b$

Find a model to predict population in 2010:

Year (x)	US Pop (y)
1920	105.7
1930	122.8
1940	131.7
1950	151.3
1960	179.3
1970	203.3
1980	226.5
1990	248.7
2000	281.4

Begin with a linear model ...

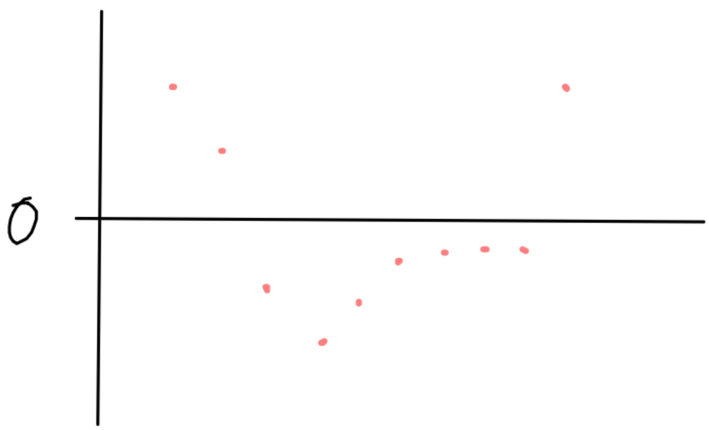
1) Scatterplot (L_1, L_2)



2) Calculate Correlation (L_1, L_2)

$r = .9928$

3) Create residual plot $(L_1, RESID)$



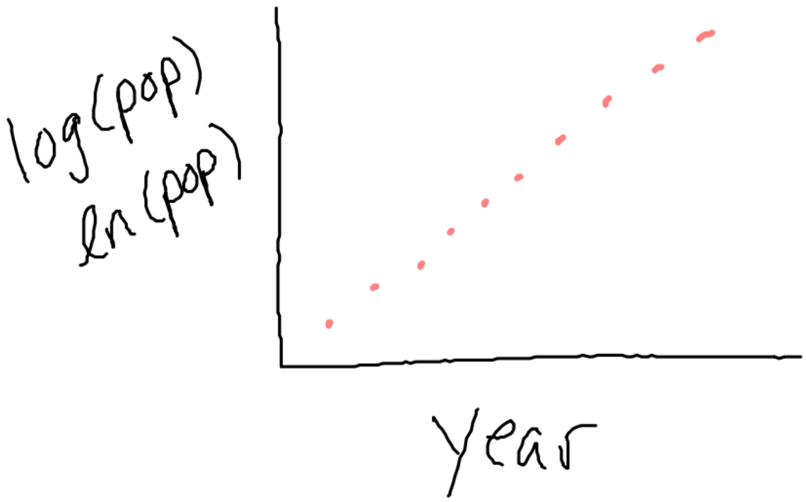
NOT random...
linear model
no good :(

See If An Exponential Model "Works" . . .

1) Create new list

L_1	L_2	L_3	
X	Y	log Y	$\Big\}$ Either
		ln Y	

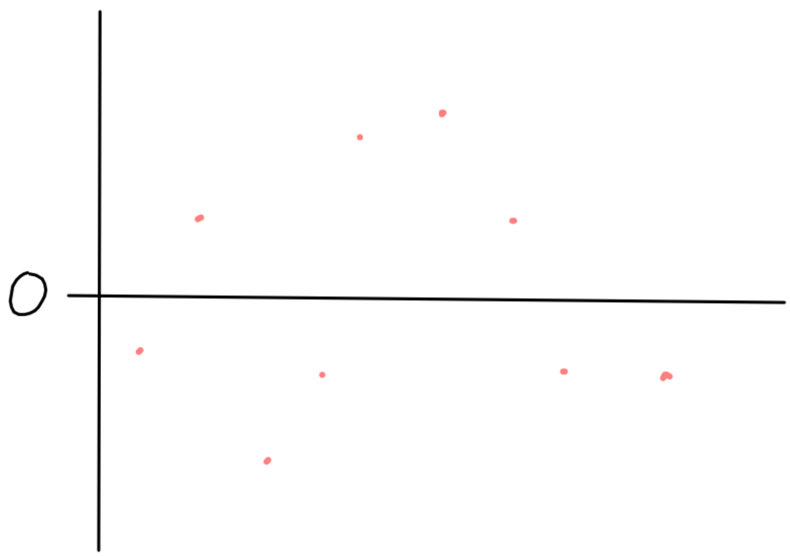
2) Scatterplot (L_1, L_3)



★3) Linear Regression (L_1, L_3)

r = .9976

4) Create New Residual Plot (L_1, RESID)



Random ☺ exponential model appropriate

5) Write Exponential Model

$$y = ab^x$$

a) Transform Linear Equation

$\log y$

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

Really

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

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

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

$\ln y$

$$\hat{y} = -19.06 + .0124x$$

Really

$$\ln \hat{y} = -19.06 + .0124x$$

$$\hat{y} = e^{-19.06 + .0124x}$$

$$\hat{y} = (e^{-19.06})(e^{.0124x})$$

$$\hat{Pop} = (5.3 \times 10^{-9})(1.01)^{year}$$

b) Use Calculator?

$\boxed{STAT} \to CALC \to ExpReg(L_1, L_2)$

$$\hat{Pop} = (5.29 \times 10^{-9})(1.01)^{Year}$$

Making Predictions - Use model to predict population in 2010 > 309.3M

1) Plug'n Chug (Use log/ln form w/out rounding)

a) $\widehat{\log pop} = -\underline{8.276051195} + \underline{.0053661533}(\cancel{year}\ 2010)$

$\widehat{\log pop} = 2.509916943$

$\widehat{pop} = 10^{2.509916943} = 323.5M$

b) $\widehat{\ln pop} = -19.05631211 + .0123560246(\cancel{\text{years}}\ 2010)$

$\widehat{\ln pop} = 5.779297336$

$\widehat{pop} = e^{5.779297336} = 323.5\text{ M}$

2) Using Calculator

a) Store LSRL when calculating

CALC $\rightarrow$ LinReg L_1, L_3, Y_1

b) Define Y_2

$Y_2 = 10 \wedge Y_1$ $\quad Y_2 = e \wedge Y_1$

c) Use $\boxed{\text{VARS}}$ to enter X-value

$Y_2(2010) = 323.5M$

2) Using Calculator

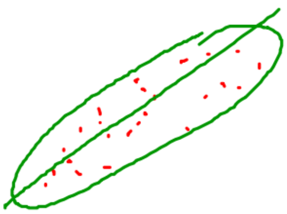
a) Store ExpReg when Calculating

$\text{CALC} \rightarrow \text{ExpReg } L_1, L_2, Y_1$

b) Use $\boxed{\text{VARS}}$ to enter X-value

$Y_1(2010) = 323.5\text{M}$

Given 2 (x,y) variables, is there an exponential relationship?



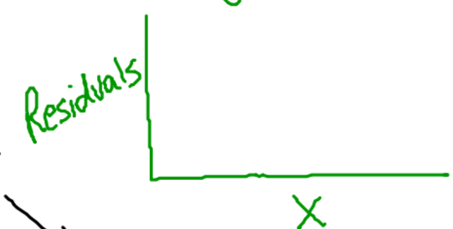
Scattersplot (x, log y)

↓

Correlation (LinReg a+bx)

↓

Residual Plot

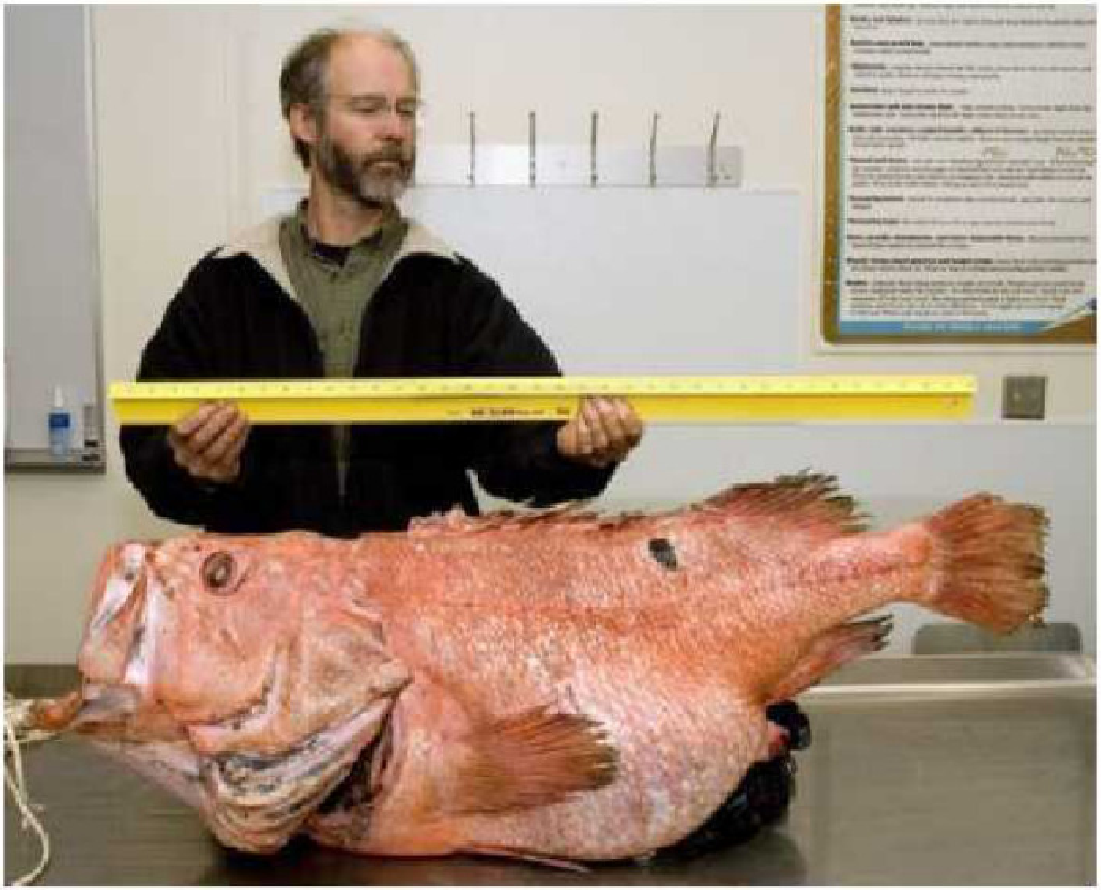


No Pattern / Random	Pattern / Not Random
Exponential Model ☺	No Exponential Model ☹
$\hat{y} = ab^x$	

Sec 12.2 cont

(Power Regression)

ROCKFISH



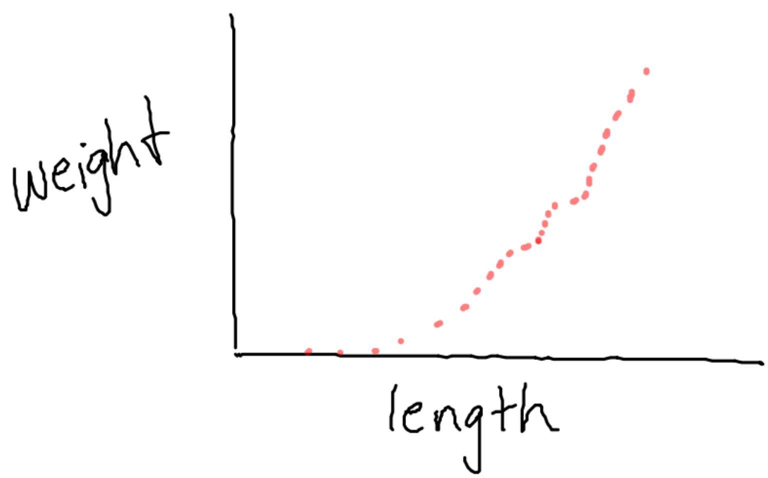
L_1 L_2

Age (yr)	Length (cm)	Weight (g)	Age (yr)	Length (cm)	Weight (g)
1	5.2	2	11	28.2	318
2	8.5	8	12	29.6	371
3	11.5	21	13	30.8	455
4	14.3	38	14	32.0	504
5	16.8	69	15	33.0	518
6	19.2	117	16	34.0	537
7	21.3	148	17	34.9	651
8	23.3	190	18	36.4	719
9	25.0	264	19	37.1	726
10	26.7	293	20	37.7	810

Find a model to predict the weight of a rockfish given its length

Begin with a linear model ...

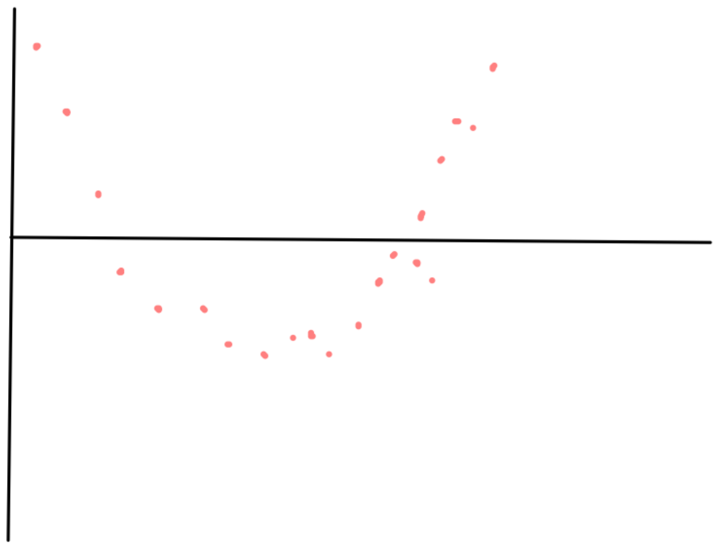
1) Scatterplot (L_1, L_2)



2) Calculate Correlation

$r = .9461$

3) Create Residual Plot (L_1, RESID)



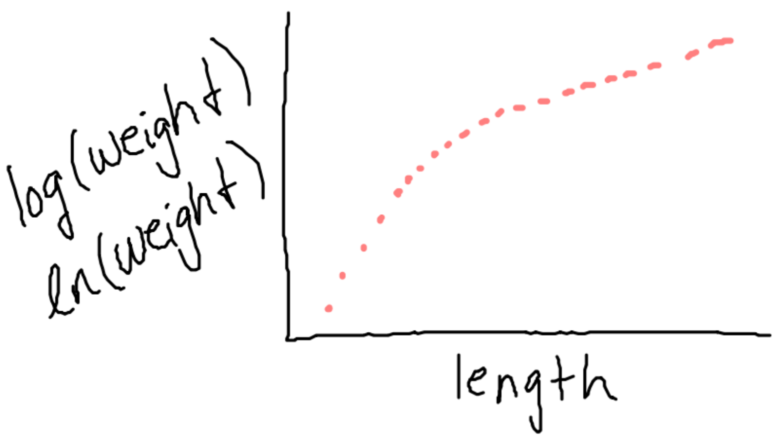
Pattern...
Linear Model
No Good :(

See If An Exponential Model Works . . .

1) Create new list

L_1	L_2	L_3	
x	y	$\log y$	Either
		$\ln y$	

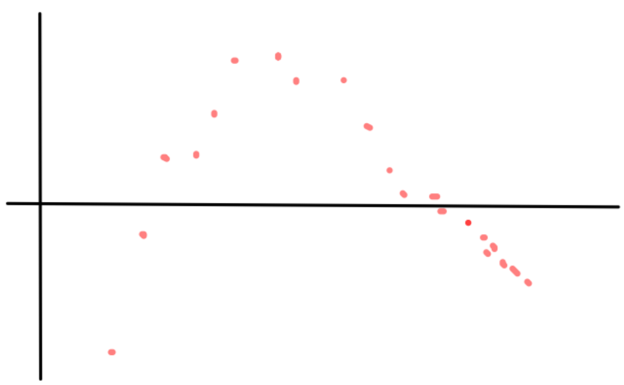
2) Scattesplot (L_1 L_3)



3) Calculate Correlation (L_1 L_3)

r = .963

4) Create Residual Plot (L_1, RESID)



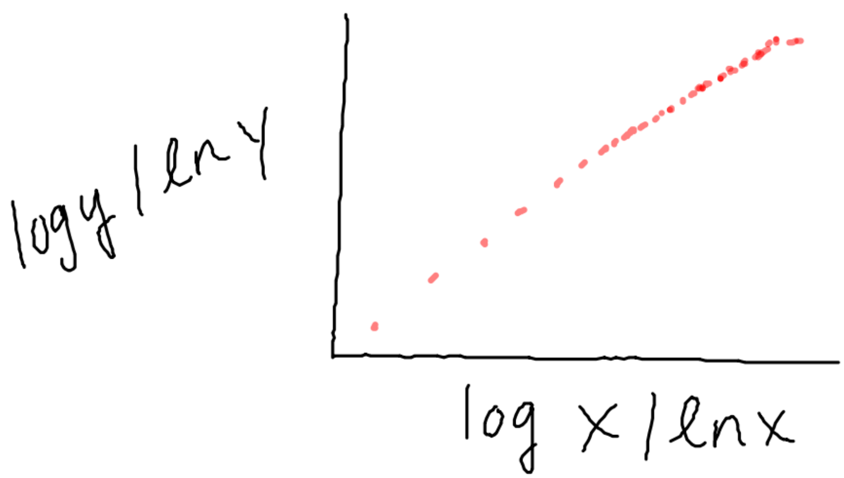
Pattern...
Exponential Model
No Good :(

See If A Power Model Works . . .

1) Create another new list

$\underline{L_1}$	$\underline{L_2}$	$\underline{L_3}$	$\underline{L_4}$	
X	Y	log Y	log X	Either
		ln Y	ln X	

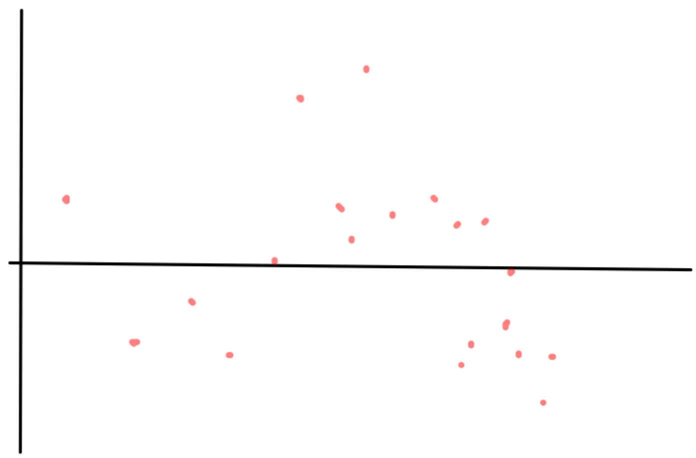
2) Scatterplot (L4, L3)



3) Calculate Correlation (L4, L3)

$r = .999$

4) Create Residual Plot (L4, RESID)



Random... Power Model Best

5) Write Power Model ($y = ax^b$)

log $\swarrow$ $\searrow$ ln

$\hat{y} = -1.8994 + 3.0494x$

Really $\downarrow$

$\log\hat{y} = -1.8994 + 3.0494(\log x)$

$\hat{y} = \left(10^{-1.8994}\right)\left(x^{3.0494}\right)$

$\hat{y} = -4.3735 + 3.0494x$

$\downarrow$ Really

$\ln\hat{y} = -4.3735 + 3.0494(\ln x)$

$\hat{y} = \left(e^{-4.3735}\right)\left(x^{3.0494}\right)$

$\widehat{\text{Weight}} = (.0126)(\text{length})^{3.0494}$

Making Predictions - Find weight if length is 24cm

1) Plug 'n Chug (Use log/ln form without rounding)

a) $\log \hat{y} = -1.8994 + 3.0494(\log \overset{24}{\cancel{x}}) = 2.30941616$

$\widehat{\text{weight}} = 10^{2.30941616} = 204\text{g}$

b) $\widehat{\ln y} = -4.3735 + 3.0494(\ln \overset{24}{\cancel{x}}) = 5.31765735$

$\widehat{\text{weight}} = e^{5.31765735} = 204\text{g}$

2) Using calculator

a) Let $Y_1 = LSRL(L_4, L_3)$

$\boxed{X,T,\theta}$

b) Define $Y_2 = (10 \wedge a)(X \wedge b)$ or $(e \wedge a)(X \wedge b)$

$\boxed{VARS} \to$ 5: Statistics $\to$ EQ

c) $\boxed{VARS} \to$ Y-VARS $\to$ Function $\to Y_2(24) = 204$ g

Notes

1) Using $\boxed{\text{CALC}}$ → PwrReg L_1, L_2 may not work

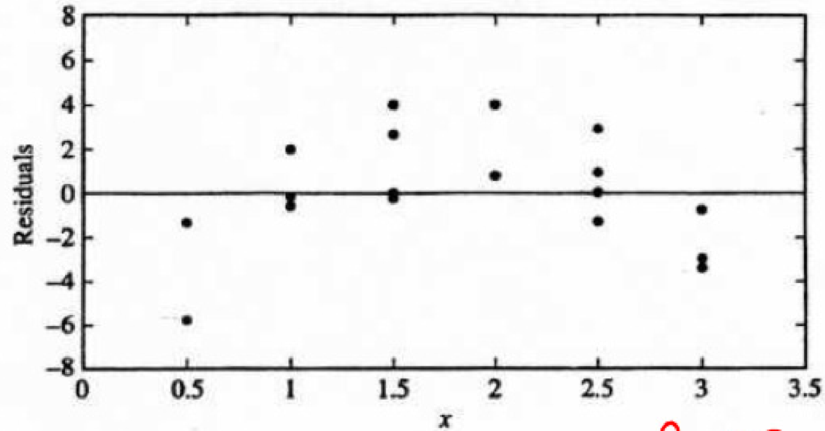
overflow error?

2) <u>Not</u> a lot of these calculations on AP Exam...

See MC Example (next page)

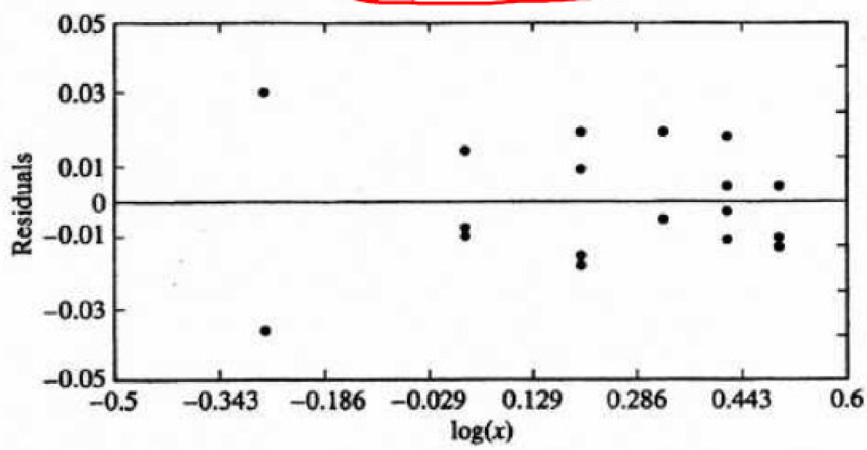
28. Two measures x and y were taken on 18 subjects. The first of two regressions, Regression I, yielded $\hat{y} = 24.5 + 16.1x$ and had the following residual plot.

Linear



Power

The second regression, Regression II, yielded $\widehat{\log(y)} = 1.6 + 0.51\log(x)$ and had the following residual plot.



More Random

Which of the following conclusions is best supported by the evidence above?

~~(A)~~ There is a linear relationship between x and y, and Regression I yields a better fit.

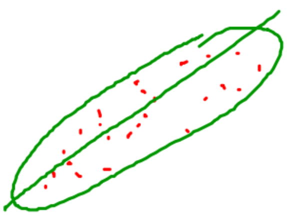
~~(B)~~ There is a linear relationship between x and y, and Regression II yields a better fit.

~~(C)~~ There is a negative correlation between x and y.

~~(D)~~ There is a nonlinear relationship between x and y, and Regression I yields a better fit.

(E) There is a nonlinear relationship between x and y, and Regression II yields a better fit.

Given 2 (x,y) variables, is there a power relationship?



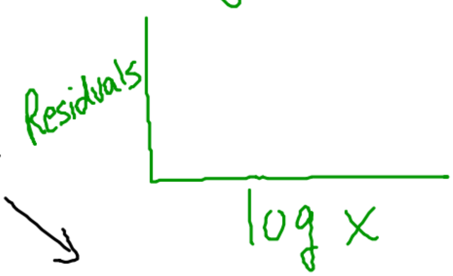
Scatterplot (log x, log y)

↓

Correlation (LinReg a+bx)

↓

Residual Plot



No Pattern / Random

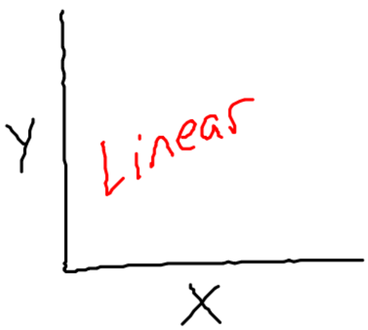
Power Model :)

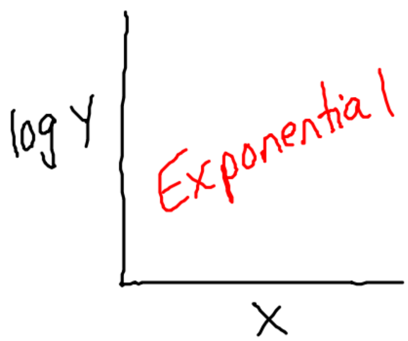
$$\hat{y} = ax^b$$

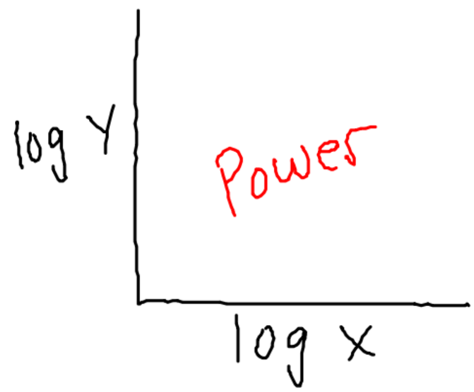
Pattern / Not Random

No Power Model :(

Review

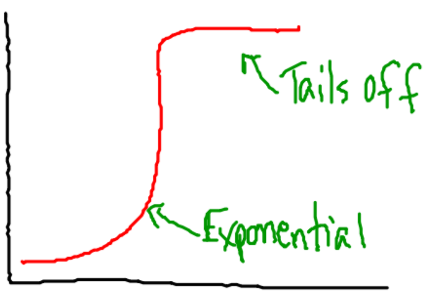






Note

There are other models!



Logistic Regression

$$\hat{y} = \frac{c}{1 + ae^{-bx}}$$