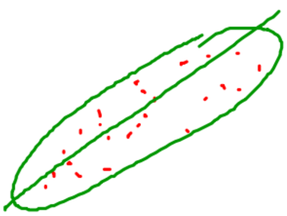
Sec 12.1

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?



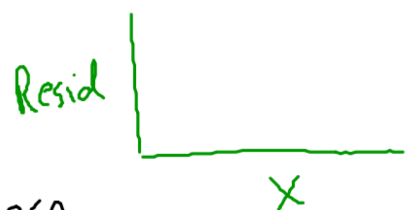
Scatterplot (Form, Direction, Strength)

↓

Numerical Summaries (LSRL, r, r^2) } LinReg

↓

Residual Plot



No Pattern → Linear Model

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

Pattern → Exponential / Power Model

$$\hat{y} = ab^x \qquad \hat{y} = ax^b$$

Ex Crying and IQ (P. 752)

L_1 (x)	L_2 (y)
Crying	IQ

$$\widehat{IQ} = 91.27 + \boxed{1.49}(\text{Crying})$$

sample slope

Lin Reg T Test

P β = the true slope of population regression line determined by [crying and IQ]

H $H_0: \beta = 0$ $H_a: \beta \neq 0$

A

Random ✓

Normal Samp Dist ✓

Independent ($N > 10(38) > 380$ ✓)

Linear - No pattern in residual plot ✓

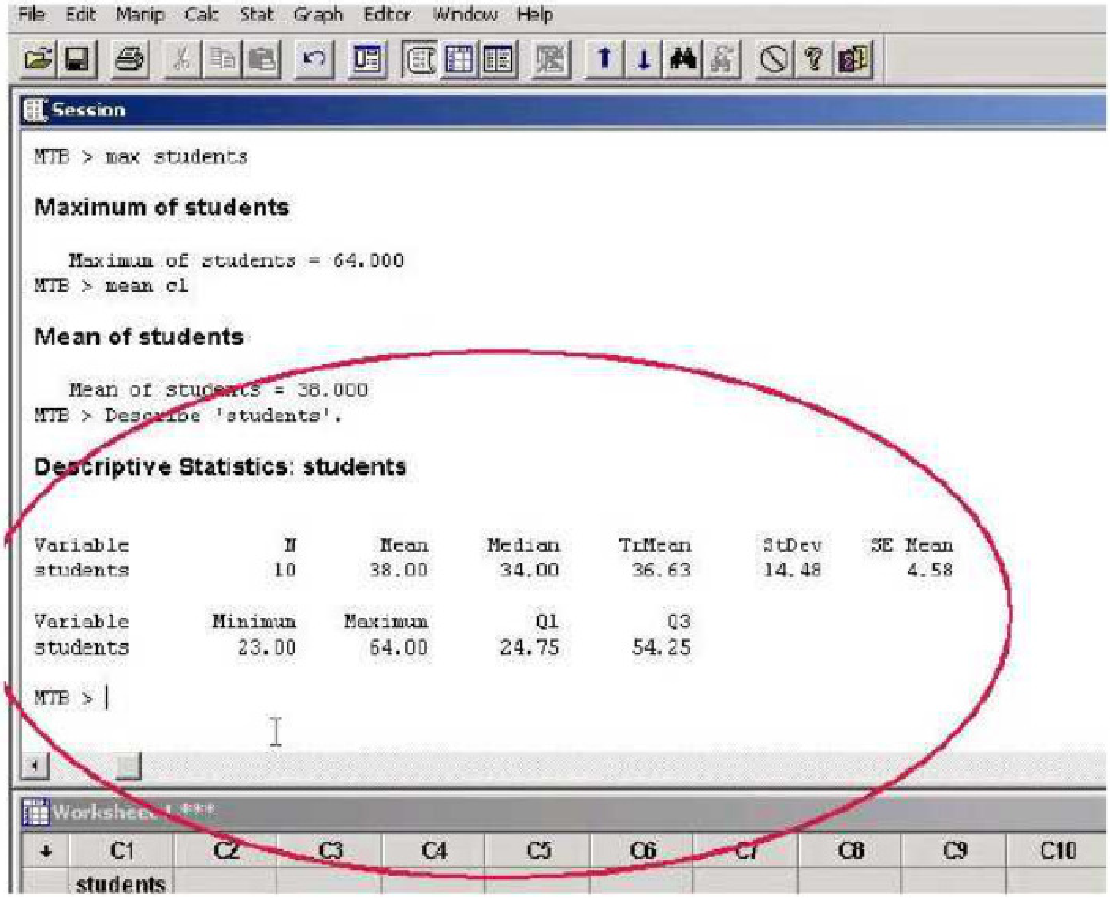
T 1) Formula / Minitab Output

$$t = \frac{b}{SE_b} = \frac{1.4929}{.4870} = 3.07$$

$$df = n - 2 = 38 - 2 = 36$$

Table B ($p < .0025$), tcdf = $p = .002$

MINITAB OUTPUT




```
MTB > max students

Maximum of students

   Maximum of students = 64.000
MTB > mean c1

Mean of students

   Mean of students = 38.000
MTB > Describe 'students'.

Descriptive Statistics: students

Variable          N     Mean   Median   TrMean    StDev   SE Mean
students         10    38.00    34.00    36.63    14.48      4.58

Variable    Minimum  Maximum       Q1       Q3
students      23.00    64.00    24.75    54.25

MTB >
```

Regression Analysis: Cost versus Income

The regression equation is
Cost = 439 + 0.511 Income

	Predictor	Coef	SE Coef	T	P
a	Constant	438.525	3.341	131.25	0.000
b	Income	0.51145	0.02325	22.00	0.000

SE b

S = 12.2225   R-Sq = 91.0%   R-Sq(adj) = 90.8%

2) Calculator

$\boxed{\text{STAT}} \rightarrow$ TESTS $\rightarrow$ LinRegTTest

$t = 3.065$, $p = .004$

S There is enough evidence ($\alpha = .05$) to reject H_0 and conclude that there is a linear relationship between crying and IQ in the population of all infants (4-10 days old)

Confidence Interval (for true slope)

P Same

A Same

I 1) Using Formula

$$CI = b \pm t^* SE_b$$

$$90\%\ CI = 1.4929 \pm 1.697(.4870)$$

$$= (.67, 2.32)$$

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^*.

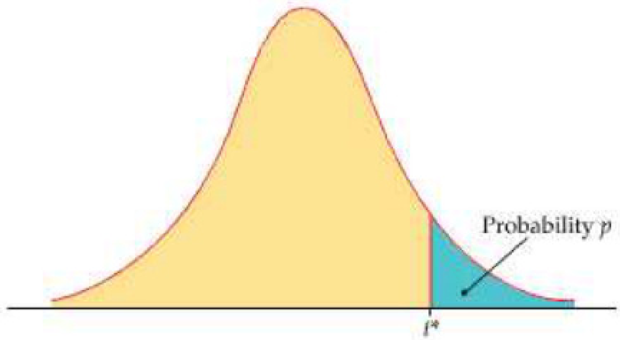


TABLE D

t distribution critical values

df	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											

S I am 90% confident that IQ increases .67 and 2.32 points for each additional peak in crying for all infants (4-10 days old)

Calculator Note

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

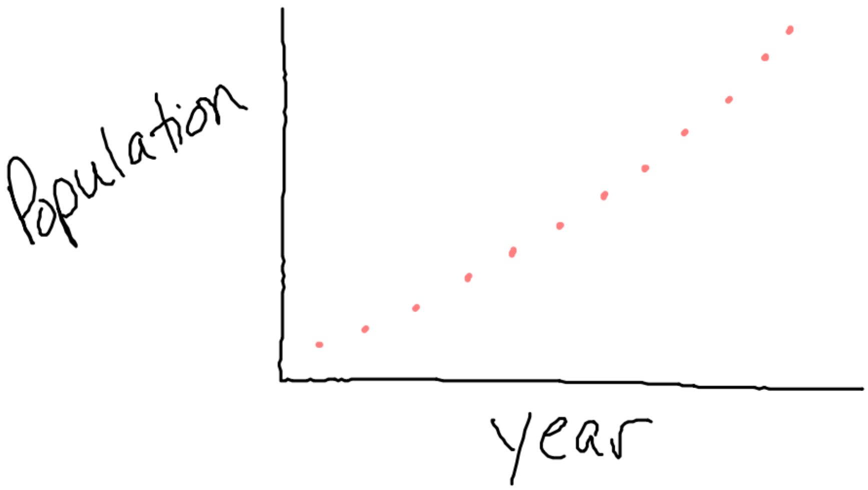
rho (population correlation)

Sec 12.2

Exponential Regression

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

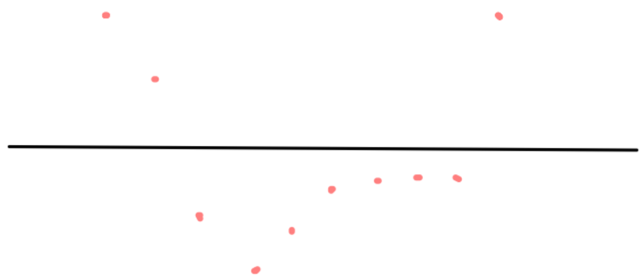
1) Scatterplot (L_1, L_2)



2) Linear Regression (L_1, L_2)

$r = .9928$

3) Create residual plot (L_1, RESID)

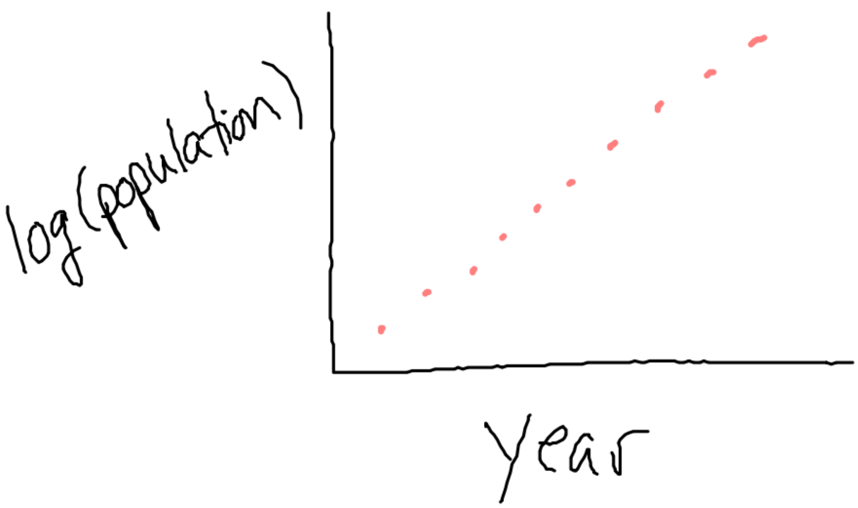


Pattern...
linear model
no good

4) Create new list

$\underline{L_1}$ (x) $\underline{L_2}$ (y) $\underline{L_3}$ (log y)

5) Scatterplot (L_1, L_3)



6) Linear Regression (L_1, L_3)

r = .9976

7) Create New Residual Plot (L_1, RESID)

No Pattern...
exponential
model
appropriate

8) Write Exponential Model $(y = ab^x)$

$\text{LinReg}(L_1, L_3)$

$\hat{y} = -8.27 + .005x$

really

$\log \hat{y} = -8.27 + .005x$

$\hat{y} = (10^{-8.27})(10^{.005})^x$

$\hat{Pop} = (10^{-8.27})(10^{.005})^{year}$

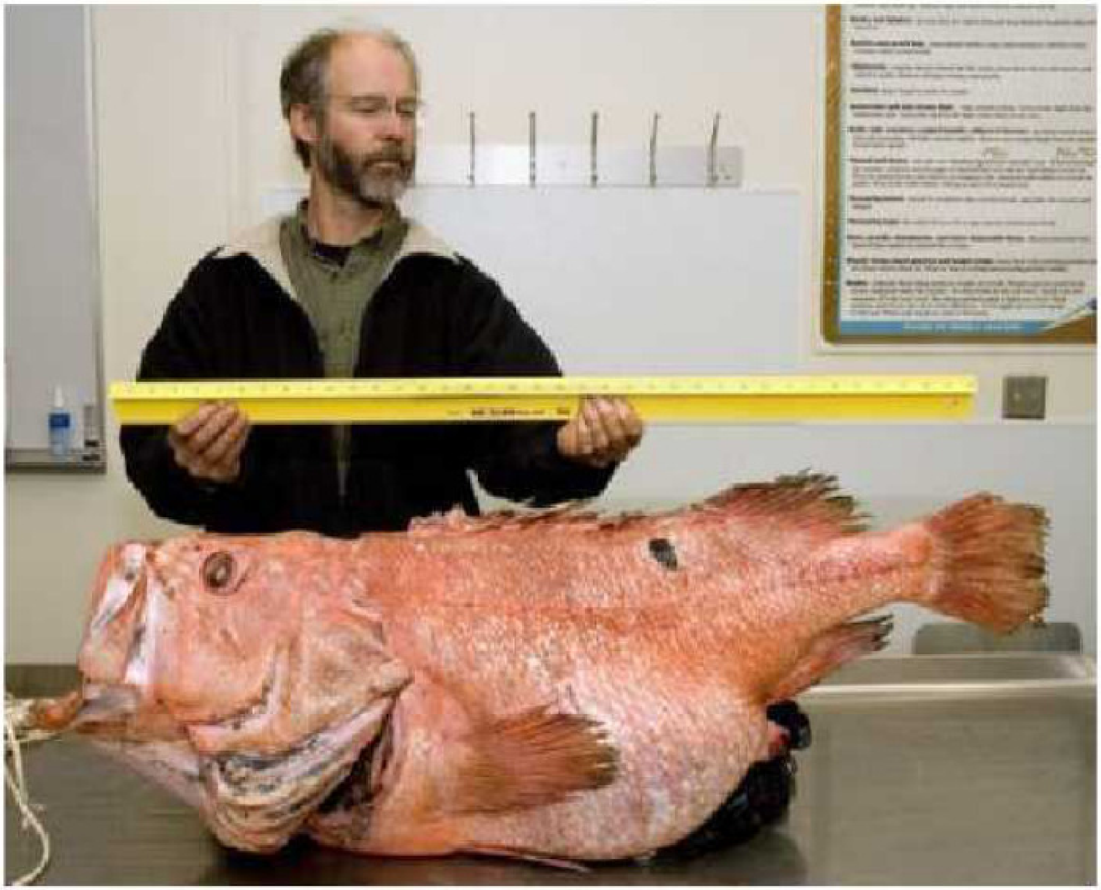
$\text{ExpReg}(L_1, L_2)$

$\hat{y} = (5.29 \times 10^{-9})(1.01)^x$

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

Power Regression

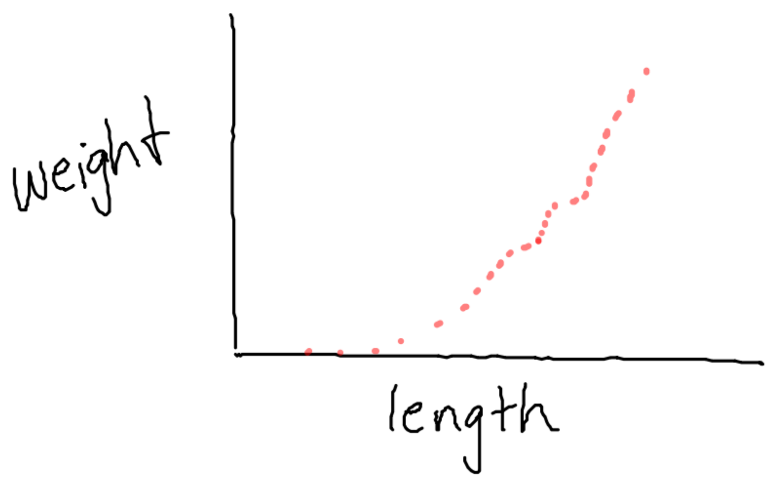
ROCKFISH



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

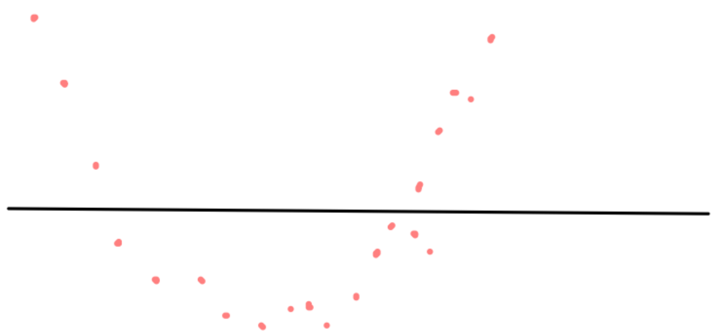
1) Scatterplot (L_1, L_2)



2) LinReg (L_1, L_2)

$r = .946$

3) Create Residual Plot (L_1, RESID)

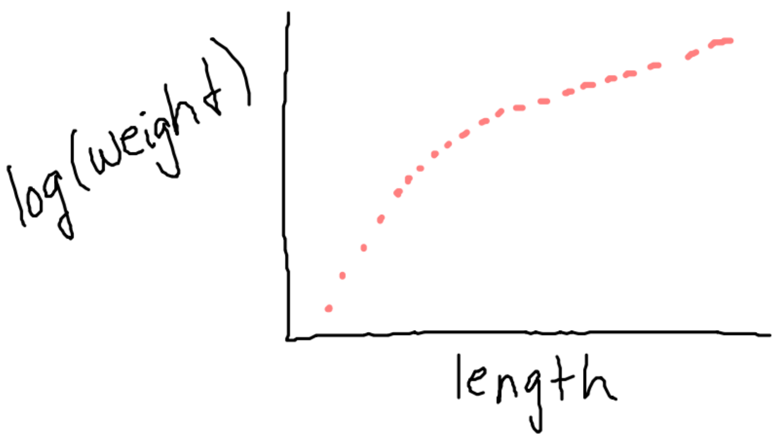


Pattern...
Linear Model
No Good :(

4) Create a new list

L_1	L_2	L_3
X	Y	log Y

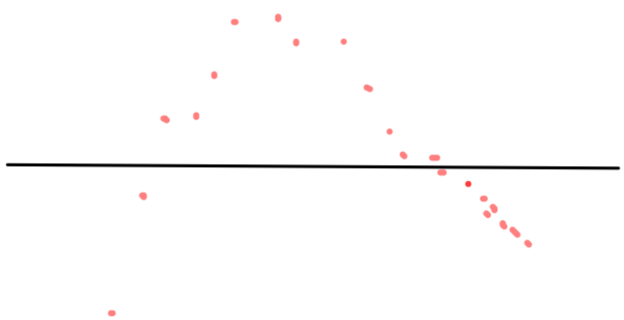
5) Scatterplot (L_1, L_3)



6) LinReg (L_1, L_3)

$r = .963$

7) Create Residual Plot (L_1, RESID)

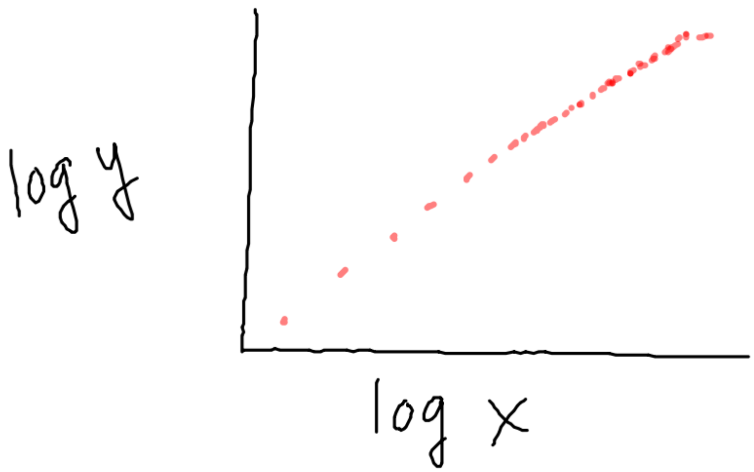


Pattern...
Exponential Model
No Good :(

8) Create another new list

L_1	L_2	L_3	L_4
X	Y	log Y	log X

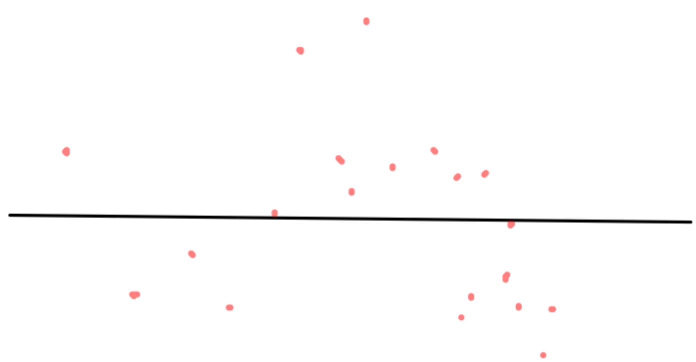
9) Scatterplot (L_4, L_3)



10) LinReg (L_4, L_3)

$r = .999$

11) Create Residual Plot (L4, RESID)



Random...
Power Model
Best

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

LinReg (L_4, L_3)

$\hat{y} = -1.89 + 3.04x$

Really

$\log \hat{y} = -1.89 + 3.04(\log x)$

$\hat{y} = (10^{-1.89})(x)^{3.04}$

$\widehat{\text{Weight}} = (10^{-1.89})(\text{length})^{3.04}$

PwrReg (L_1, L_2)

$\hat{y} = (.0126)(x)^{3.04}$

$\widehat{\text{Weight}} = (.0126)(\text{length})^{3.04}$

What is the relationship between per capita income in a country and child mortality? On this page is computer output for three different regression models examining this relationship for countries in Central and South America. Child mortality is measured in deaths before age 5 per 1000 children born, and income is measured in U.S. dollars per person. Questions about these data are on the next page. All logarithms are base 10.

Linear

I. Child mortality *versus* Income

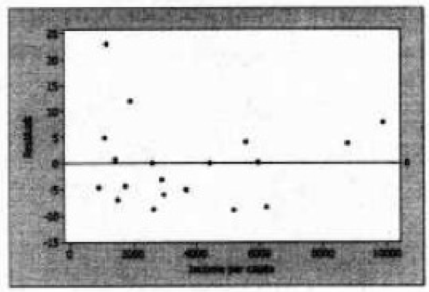
Predictor	Coef	SE Coef	T	P
Constant	34.149	3.397	10.05	0.000
Income per capita	-0.0027295	0.0007530	-3.62	0.002

S = 8.35744 R-Sq = 43.6% R-Sq(adj) = 40.3%

Y

r=.66

X



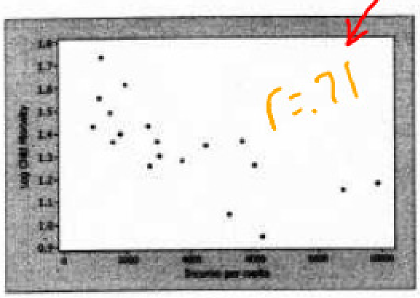
Exponential

II. Log child mortality *versus* Income

Predictor	Coef	SE Coef	T	P
Constant	1.53434	0.05580	27.50	0.000
Income per capita	-0.00005198	0.00001237	-4.20	0.001

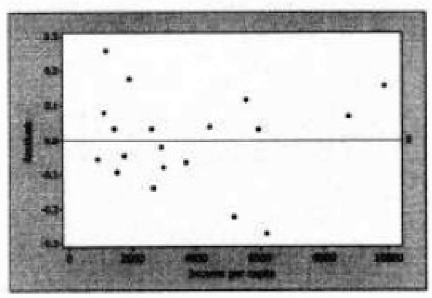
S = 0.137280 R-Sq = 51.0% R-Sq(adj) = 48.1%

log Y



r=.71

X



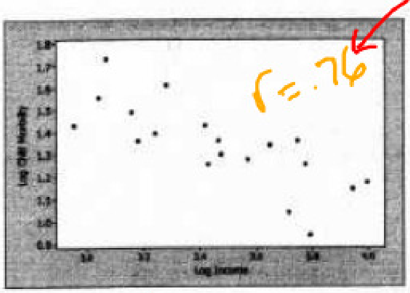
Power

III. Log child mortality *versus* Log income

Predictor	Coef	SE Coef	T	P
Constant	2.9649	0.3299	8.99	0.000
Log Income	-0.46824	0.09476	-4.94	0.000

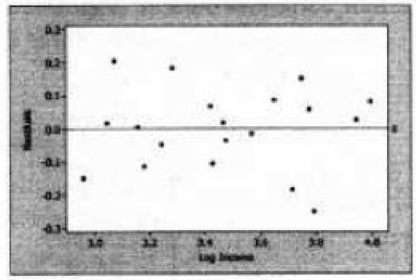
S = 0.125578   R-Sq = 59.0%   R-Sq(adj) = 56.5%

log Y



r = .76

log x



Most Random

1. Explain why the information provided suggests that a linear model may not be appropriate for describing the relationship between Child mortality and Income in these countries.

a) Scatterplot shows curved relationship

b) Residual plot shows slightly U-shaped pattern

2. Would an exponential model or a power model provide a better description of this relationship? Use the information provided to justify your answer.

Correlation stronger

residual plot more random