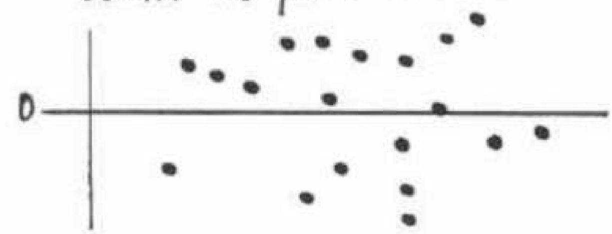
14d) P β = slope of population regression line determined by <time at table and calorie consumption>

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

A L Scatterplot moderately linear ($r = -.64$) with no pattern in residual plot



I #Toddlers > 10(20) > 200 Toddlers

N NPP of residuals linear

E Residuals equally scattered around $y = 0$

R Random sample used

T From MiniTab:

$t = -3.62$ $p = 2(.001) = .002$

S At $\alpha = .01$, there is strong evidence to reject H_0 and conclude a linear relationship exists between the time spent at a table and caloric consumption for all toddlers

AP STATISTICS
(Section 12.1)

Sarah's parents are concerned that she seems short for her age. Their doctor has recorded Sarah's age and height on six consecutive office visits. These data were entered into a Minitab worksheet and Sarah's height was regressed on her age. Here is part of the computer output.

Predictor		Coef	StDev	t-ratio	P
Constant	a	71.950	1.053	*	*
Age	b	0.38333	0.02041	*	*
			SE_b		
s = 0.3873			R-sq = 98.9%		R-sq (adj) = 98.6%
			(r = .9944)		

1. Suppose we want to conduct a test to determine whether Sarah's age is useful in predicting her height. Define the parameter of interest; write an appropriate null and alternative hypothesis for such a test.

β = the true slope of population regression line determined by Sarah's height and age

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

2. What is the equation of the least-squares regression line?

$\widehat{\text{Height}} = 71.950 + .3833\,(\text{Age})$ > Words!

3. The t statistic for testing H_0 has been left out. Find t.

$\boxed{t = \frac{b}{SE_b}} = \frac{.3833}{.02041} = 18.78$

4. How many degrees of freedom does t have?

$df = n - 2 = 6 - 2 = 4$

5. Use Table B or your calculator to find the P-value.

$P < .0005 \times 2 < .001$ tcdf (18.78, 100, 4) = .00002 × 2 = .00004

6. Write your conclusion in plain language (at $\alpha = .05$).

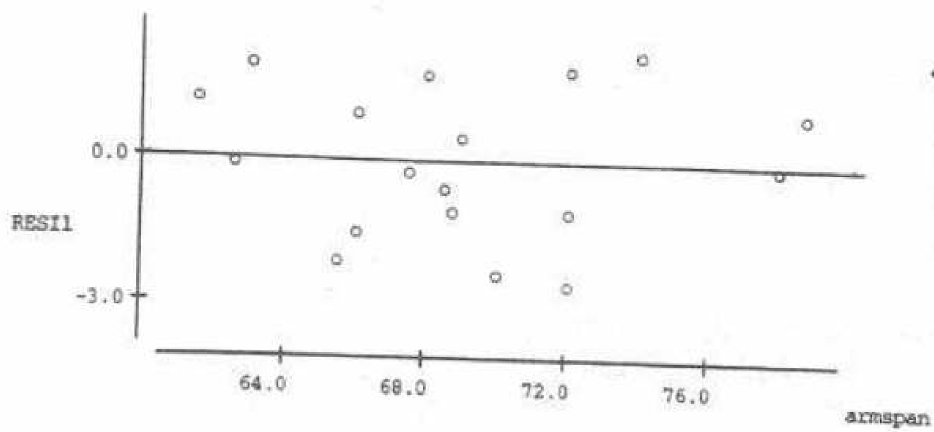
At $\alpha = .05$, there is strong evidence to reject H_0 and conclude a linear relationship exists between Sarah's height and age to a point (< 18)

Ideal Proportions

Once upon a time, a random sample of students at Indiana Univerity made measurements of their arm span and height. They entered their results into a Minitab worksheet, requested least squares regression of height on arm span (both in inches) and obtained the following output:

Predictor	Coef	StDev	t-ratio	P
Constant	a 11.547	5.600	2.06	0.056
Arm Span	b 0.84042	0.08091	10.39	0.000
		SE_b	t	P
s = 1.613		R-sq = 87.1% ($r = .9332$)		R-sq (adj) = 86.3%

A residual plot for the data looks like this:



$n = 18$

1. Determine the equation of the least squares regression line. > sample

$$\widehat{Height} = 11.547 + .84042\ (Arm\ Span)$$

2. Construct a 95% confidence interval for the true slope determined by arm span and height for all students at IU (use PAIS):

P β = true slope of population regression line determined by arm span and height for all IU students

A L Strong linear association with no pattern in residual plot

I # IU students > 18 (10) > 180

N NPP of residuals unknown

E Residuals equally scattered around $y = 0$

R Random sample used

I $\quad 95\%\ CI = b \pm t^* SE_b$

$$= .84042 \pm (2.120)(.08091)$$

$$= (.67, 1.01)$$

S $\quad$ I am 95% confident that height increases between .67" and 1.01" for each 1" increase in arm span for all IU students

$\rightarrow df = 12-2 = 10$

As part of a class project at a large university, Amber selected a random sample of 12 students in her major field of study. All students in the sample were asked to report their number of hours spent studying for the final exam and their score on the final exam. A regression analysis on the data produced the following partial computer output.

Predictor	Coef	SE Coef	T	P
Constant	62.328	4.570	13.64	0.000
Study Hours	2.697	0.745	3.62	0.005
S = 5.505	R-sq = 56.7%			

Amber wants to compute a 95 percent confidence interval for the slope of the least squares regression line in the population of all students in her major field of study. Assuming that conditions for inference are satisfied, which of the following gives the [margin of error] for the confidence interval?

(A) $(\cancel{2.228})(0.745)$

(B) $(\cancel{2.228})\left(\frac{0.745}{\sqrt{12}}\right)$

(C) $(\cancel{2.228})(5.505)$

(D) $(\cancel{2.228})\left(\frac{5.505}{\sqrt{12}}\right)$

(E) $(\cancel{2.228})(2.697)$

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

$(2.228)(.745)$

AP STATISTICS

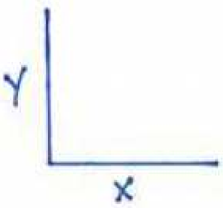
(Exponential Models)

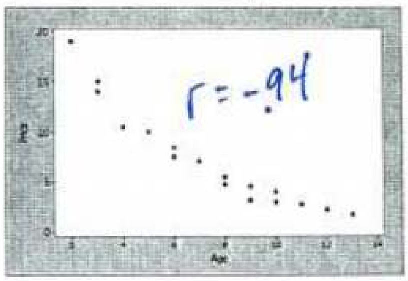
Brandon is shopping for a used car and collects data on age (in years) and price (in 1000s of dollars) for Ford Taurus sedans on a used-car Web site. Below are computer outputs for two different regression models: Price *vs* Age and Log (Price) *vs* Age. All logarithms are base 10.

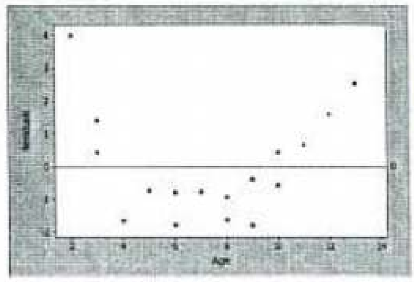
I. Price *versus* Age

Predictor	Coef	SE Coef	T	P
Constant	17.870	1.030	17.35	0.000
Age	-1.4300	0.1276	-11.21	0.000

S = 1.68336 R-Sq = 89.3% R-Sq(adj) = 88.6%





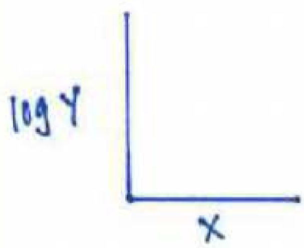


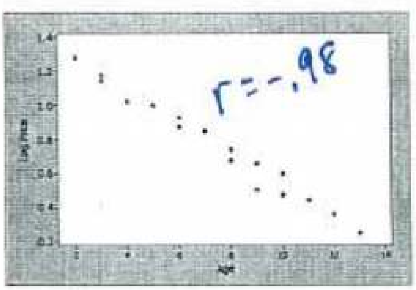
Not Random

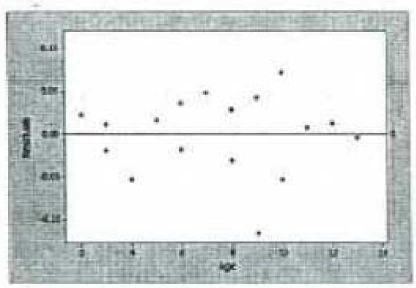
II. Log Price *versus* Age

Predictor	Coef	SE Coef	T	P
Constant	1.43723	0.02881	49.89	0.000
Age	-0.090652	0.003569	-25.40	0.000

S = 0.0470892 R-Sq = 97.7% R-Sq(adj) = 97.6%







Random

1. Explain how the information provided suggests that an exponential model would describe the relationship between car age and price better than a linear model would.

(x, log y) produced a stronger correlation and more evenly scattered residual plot ✓

Exponential ↗

2. Write the equation of the exponential model in the form of $\hat{y} = ab^x$

$$\log \hat{y} = 1.43723 + {}^{-}.090652\,x$$

$$\hat{y} = \left(\underline{10^{1.43723}}\right)\left(\underline{10^{-.090652}}\right)^x$$

$$\widehat{\text{Price}} = (27.3671)(.8116)^{\text{year}}$$

3. Use the exponential model to predict the price of a 5-year old Ford Taurus.

$$\log \hat{y} = 1.43723 - .090652\,x$$

$$\log \hat{y} = 1.43723 - .090652(5)$$

$$\log \hat{y} = .98397$$

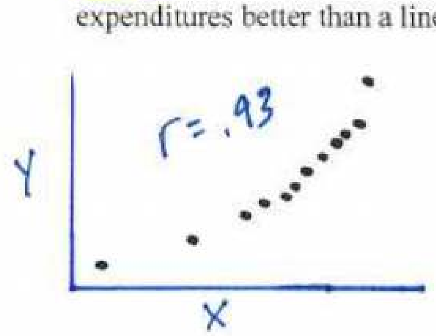
$$\widehat{\text{Price}} = 10^{.98397} = 9.637 \text{ thousand or } \$9637$$

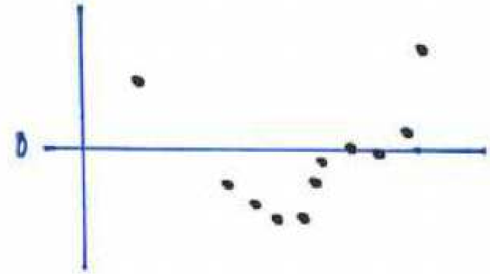
According to data from the U.S. Health Care Financing Administration, the national expenditures for drugs (in billions of dollars) for selected years from 1970 to 1997 are as follows:

X	Year	70	80	85	87	89	90	91	92	93	94	95	97
Y	Spent	8.8	21.6	37.1	43.2	50.6	59.9	65.6	71.2	75	77.7	83.4	108.9

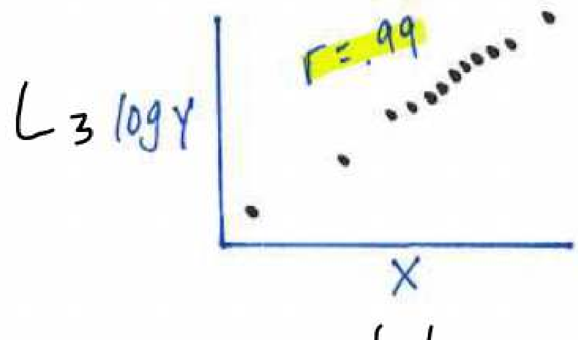
→ 2019

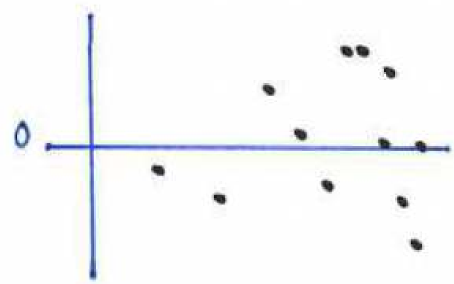
4. Justify why an exponential model would describe the relationship between year and expenditures better than a linear model.





Residual plot not evenly scattered





Residual more evenly scattered

5. Write the equation of the exponential model in the form of $\hat{y} = ab^x$

$\log \widehat{\text{spent}} = -1.865 + .0402(\text{year})$

$\widehat{\text{spent}} = (10^{-1.865})(10^{.0402})^{\text{year}}$

$\widehat{\text{spent}} = (.0136)(1.0969)^{\text{year}}$

ExpReg (L_1, L_2)

$\widehat{\text{Spent}} = (.0136)(1.0969)^{\text{year}}$

6. Predict the national drug expenditures for this year. Do you have confidence in this result? Why or why not?

For 2017, Use 117 → \$690 billion

No Confidence → extrapolation

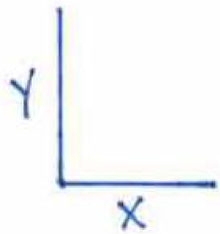
12.2 AP Statistics Name:

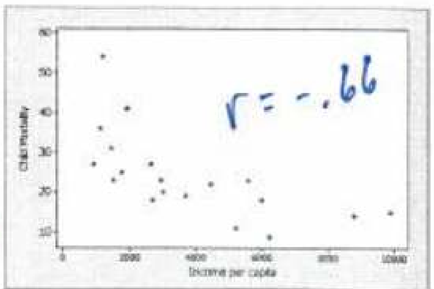
What is the relationship between per capita income in a country and child mortality? On this page is computer output for three different regression models examining this relationship for countries in Central and South America. Child mortality is measured in deaths before age 5 per 1000 children born,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

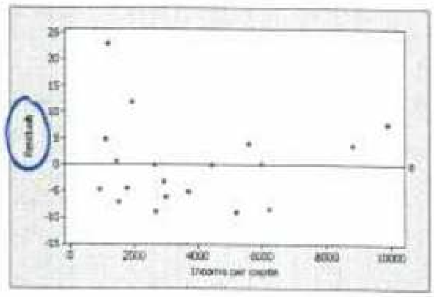
I. Child mortality *versus* Income

Predictor	Coef	SE Coef	T	P
Constant	34.1[illegible]	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%







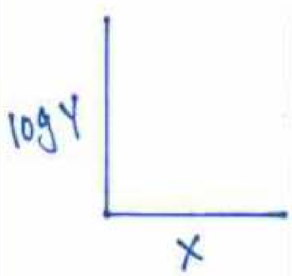
Not evenly scattered

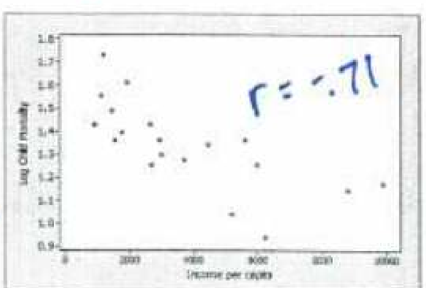
II. Log child mortality *versus* Income

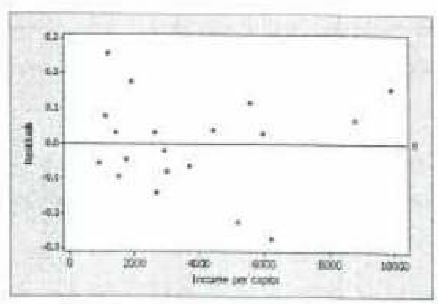
Exp

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%







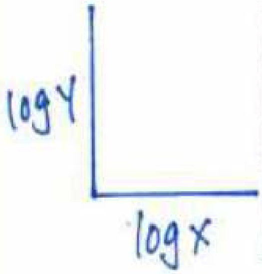
More evenly scattered

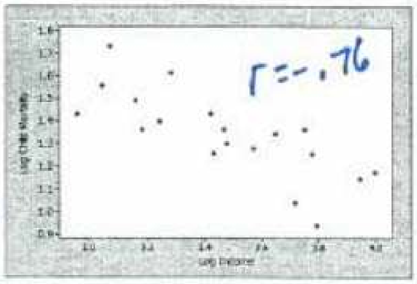
III. Log child mortality *versus* Log income

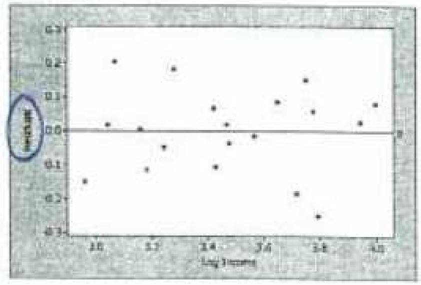
Power

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%







Most evenly scattered

1. Explain why the information provided suggests that a power model would better describe the relationship between Child mortality and Income in these countries than a linear model or exponential model would.

(log x, log y) produced a stronger correlation and most randomly scattered residual plot

power

2. Write the equation of the power model in the form $\hat{y} = ax^b$

$\widehat{\log y} = 2.9649 - .46824 (\log x)$ ← \$1300

$\hat{y} = (10^{2.9649})(x)^{-.46824}$

$\widehat{\text{Mortality}} = (922.35)(\text{Income})^{-.46824}$

3. Use the model to predict the mortality rate in a country with an income of \$1,300 per person. } US is \$26,000

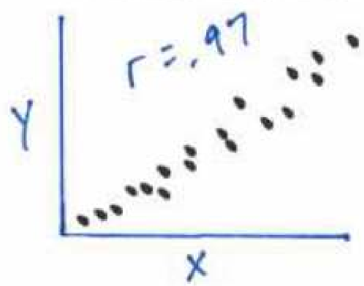
$\widehat{\text{Mortality}} = (922.35)(1300)^{-.46824}$

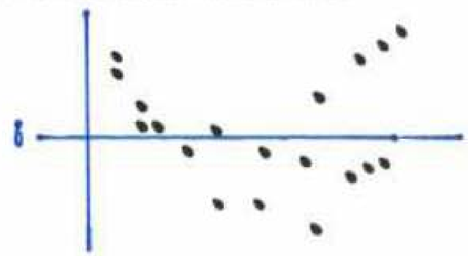
$= 32.12$ deaths / 1000 children

Foresters are interested in predicting the amount of usable lumber they can harvest from various tree species. The following data have been collected on the diameter of Ponderosa trees and the yield in board feet (a piece of lumber 12 inches by 12 inches by 1 inch).

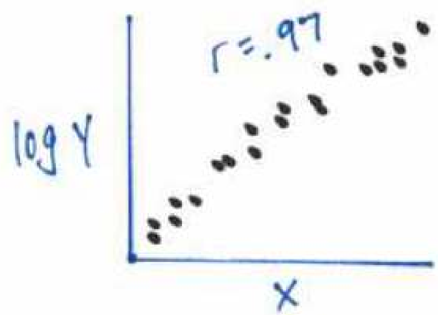
X	D	36	28	28	41	19	32	22	38	25	17	31	20	25	19	39	33	17	37	23	39
Y	BF	192	113	88	294	28	123	51	252	56	16	141	32	86	21	231	187	22	205	57	265

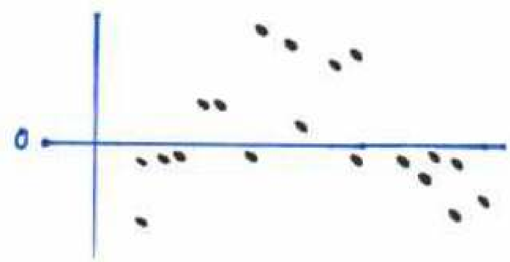
4. Justify why a power model would describe the relationship between diameter and board feet better than either a linear or exponential would.



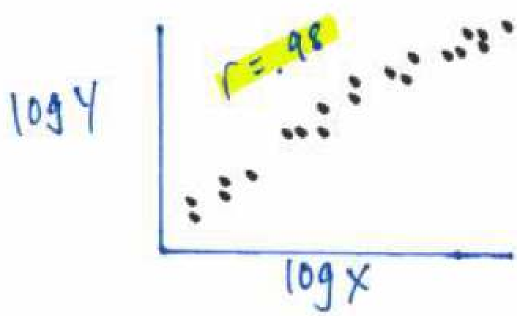


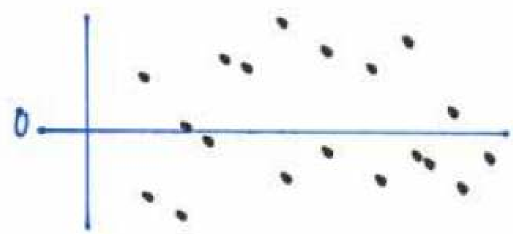
Not evenly scattered





Not evenly scatterd





Most evenly scattered

5. Write the equation of the power model in $\hat{y} = ax^b$

$\widehat{\log BF} = -2.5691 + 3.1366 (\log \text{Diameter})$

$\widehat{\text{Board Feet}} = (10^{-2.5691})(\text{Diam})^{3.1366}$

$\widehat{\text{Board Feet}} = (.0026)(\text{Diameter})^{3.1366}$

PwrReg (L_1, L_2):

$\widehat{BF} = (.0026)(\text{Diam})^{3.1366}$