AP Statistics – Chapter 5A Test

(All MC and subquestions are 3pts, except where noted – TOTAL 88 POINTS)

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1) Researchers studying growth patterns of children collect data on the heights of fathers and sons. The correlation between the fathers' heights and the heights of their 16 year–old sons is most likely to be...

    A) near 0
    B) somewhat greater than 1.0
    C) exactly +1.0
    D) near –1.0
    (E) near +0.7

1)   **E**

2) The auto insurance industry crashed some test vehicles into a cement barrier at speeds of 5 to 25 mph to investigate the amount of damage to the cars. They found a correlation of $r = 0.60$ between speed (MPH) and damage ($). If the speed at which a car hit the barrier is 1.5 standard deviations above the mean speed, we expect the damage to be _____ the mean damage.

    A) equal to
    B) 0.60 SD above
    C) 1.5 SD above
    D) 0.36 SD above
    (E) 0.90 SD above

$.6 \times 1.5 = .9 \, SD$

2)   **E**

3) The correlation between X and Y is $r = 0.35$. If we double each X value, decrease each Y by 0.20, and interchange the variables (put X on the Y–axis and vice versa), the new correlation

    A) is 0.50
    B) cannot be determined.
    (C) is 0.35
    D) is 0.70
    E) is 0.90

3)   **C**

4) The correlation between a family's weekly income and the amount they spend on restaurant meals is found to be $r = 0.30$. Which must be true?

    I. Families tend to spend about 30% of their incomes in restaurants.
    II. In general, the higher the income, the more the family spends in restaurants.
    III. The line of best fit passes through 30% of the (income, restaurant$) data points.

    A) III only
    B) I, II, and III
    C) I only
    D) II and III only
    (E) II only

4)   **E**

5) A medical researcher finds that the more overweight a person is, the higher his pulse rate tends to be. In fact, the model suggests that 12–pound differences in weight are associated with differences in pulse rate of 4 beats per minute. Which is true?

    I. The correlation between pulse rate and weight is 0.33.
    II. If you lose 6 pounds, your pulse rate will slow down 2 beats per minute.
    III. A positive residual means a person's pulse rate is higher than the model predicts.

    A) II only
    B) none
    C) II and III only
    D) I only
    (E) III only

5)   **E**

6) A least squares line of regression has been fitted to a scatterplot; the model's residuals plot is shown. Which is true?   6) __D__
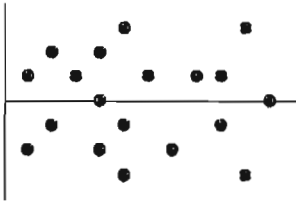
A) None of these.
B) The linear model is poor because some residuals are large.
C) The linear model is poor because the correlation is near 0.
✓D) The linear model is appropriate.
E) A curved model would be better.

7) All but one of these statements contain a mistake. Which could be true?   7) __D__
A) There is a high correlation (1.09) between height of a corn stalk and its age in weeks.
B) The correlation between a football player's weight and the position he plays is 0.54.
C) The correlation between a car's length and its fuel efficiency is 0.71 miles per gallon. NO UNITS
(D) The correlation between the amount of fertilizer used and the yield of beans is 0.42.
E) There is a correlation of 0.63 between gender and political party.

8) Residuals are...   8) __C__
A) none of these
B) possible models not explored by the researcher.
✓C) the difference between observed responses and values predicted by the model.
D) variation in the data that is explained by the model.
E) data collected from individuals that is not consistent with the rest of the group.

9) It's easy to measure the circumference of a tree's trunk, but not so easy to measure its height. Foresters developed a model for ponderosa pines that they use to predict the tree's height (in feet) from the circumference of its trunk (in inches): $\ln \hat{h} = -1.2 + 1.4(\ln C)$. A lumberjack finds a tree with a circumference of 60"; how tall does this model estimate the tree to be?   9) __B__

A) 5'   B) 93'   C) 11'   D) 83'   E) 19'

$\ln \hat{h} = -1.2 + 1.4 \cdot \ln(60)$ $e^{\ln \hat{h}} \approx e^{4.53}$ $\hat{h} \approx 92.95$

10) If the point in the upper right corner of this scatterplot is removed from the data set, then what will happen to the slope of the line of best fit (b) and to the correlation (r)?   10) __B__

Takeout
slope will go down ↓
r ↑ strengthened
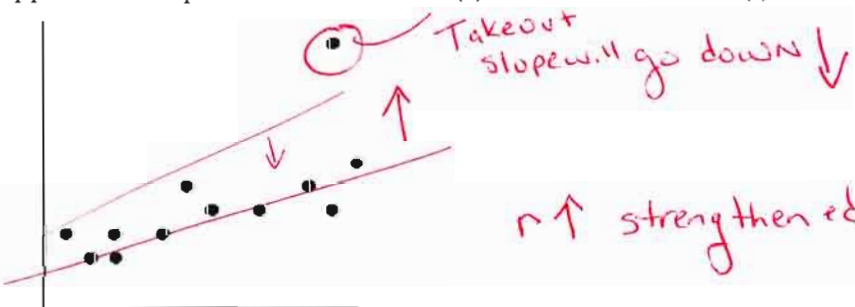
A) b will increase, and r will decrease.
B) b will decrease, and r will increase.
C) both will remain the same.
D) both will decrease.
E) both will increase.

2

$$\left(\sqrt{\hat{str}}\right)^2 12 + 20\left(\frac{1}{2}\right) = (22)^2$$

11) The model $\sqrt{\hat{str}} = 12 + 20dia$ can be used to predict the breaking strength of a rope (in pounds) from its diameter (in inches). According to this model, how much force should a rope one–half inch in diameter be able to withstand?

11) __B__

    A) 16 lbs   B) 484 lbs   C) 256 lbs   D) 22 lbs   E) 4.7 lbs

12) Which is true?

12) __A__

I. Random scatter in the residuals indicates a model with high predictive power. ✗

II. If two variables are very strongly associated, then the correlation between them will be near +1.0 or –1.0. ✗   linear

III. The higher the correlation between two variables the more likely the association is based in cause and effect. ✗

    A) none

    B) I and II only

    C) II only

    D) I only

    E) I, II, and III

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

13) **Assembly line** Your new job at *Panasony* is to do the final assembly of camcorders. As you learn how, you get faster. The company tells you that you will qualify for a raise if after 13 weeks your assembly time averages under 20 minutes. The data shows your average assembly time during each of your first 10 weeks.

13) _____

Week	Time(min)
1	43
2	39
3	35
4	33
5	32
6	30
7	30
8	28
9	26
10	25

a. Which is the explanatory variable?___Weeks WORKED___ (2pts)

b. What is the correlation between these variables? ___r = -.97___ (3pts)

c. You want to predict whether or not you will qualify for that raise. Would it be appropriate to use a linear model? Explain. (5pts)

NO. THE RESIDUAL PLOT SHOWS A DISTINCT CURVE, SO A LINEAR MODEL IS INAPPROPRIATE FOR THIS MODEL.

ALSO, PREDICTIONS ABOUT WHAT WILL HAPPEN IN 3 WEEKS IN THE FUTURE ARE LIKELY TO BE UNRELIABLE.

4

14) **Gas mileage** An important factor in the amount of gasoline a car uses is the size of the engine. Called "displacement", engine size measures the volume of the cylinders in cubic inches. The regression analysis is shown.

14) _____

```
Dependent variable is:                    MPG
89 total cases of which 0 are missing
R squared = 60.9%    R squared (adjusted) = 60.0%
s = 3.056  with (89) - 2 = 87  degrees of freedom
```

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	696.744	1	696.744	74.6
Residual	448.236	48	9.33826	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	34.9799	1.231	28.4	Š 0.0001
Eng. Displcmt	-0.066196	0.0077	-8.64	Š 0.0001

a. How many cars were included in this analysis? ___ 89 CARS ___ (2pts)

b. skipped

c. A car you are thinking of buying is available with two different size engines, 190 cubic inches or 240 cubic inches. How much difference might this make in your gas mileage? (Show your work – 5pts)

$$\widehat{MPG} = 34.9799 - .066196 \, (ENG \; DISPLCMT)$$

$$\widehat{MPG} = 34.9799 - .066196 \, (190) = 22.40266 \quad \text{(1PT)}$$

$$\widehat{MPG} = 34.9799 - .066196 \, (240) = 19.09286 \quad \text{(1PT)}$$

$$\overline{}$$

$$3.3098 \; mpg$$

The 190 cubic inch engine is about 3.3 mpg better than the 240 cubic inch engine

UNITS
1 pt

2pts

5

15) **Crawling** Researchers at the University of Denver Infant Study Center investigated whether   15) _____
babies take longer to learn to crawl in cold months (when they are often bundled in clothes that
restrict their movement) than in warmer months. The study sought an association between
babies' first crawling age (in weeks) and the average temperature during the month they first
try to crawl (about 6 months after birth). Between 1988 and 1991 parents reported the birth
month and age at which their child was first able to creep or crawl a distance of four feet in one
minute. Data were collected on 208 boys and 206 girls. The graph below plots average crawling
ages (in weeks) against the mean temperatures when the babies were 6 months old. The
researchers found a correlation of $r = -0.70$ and the their line of best fit was

$$AvAge = 36 - 0.08AvTemp.$$
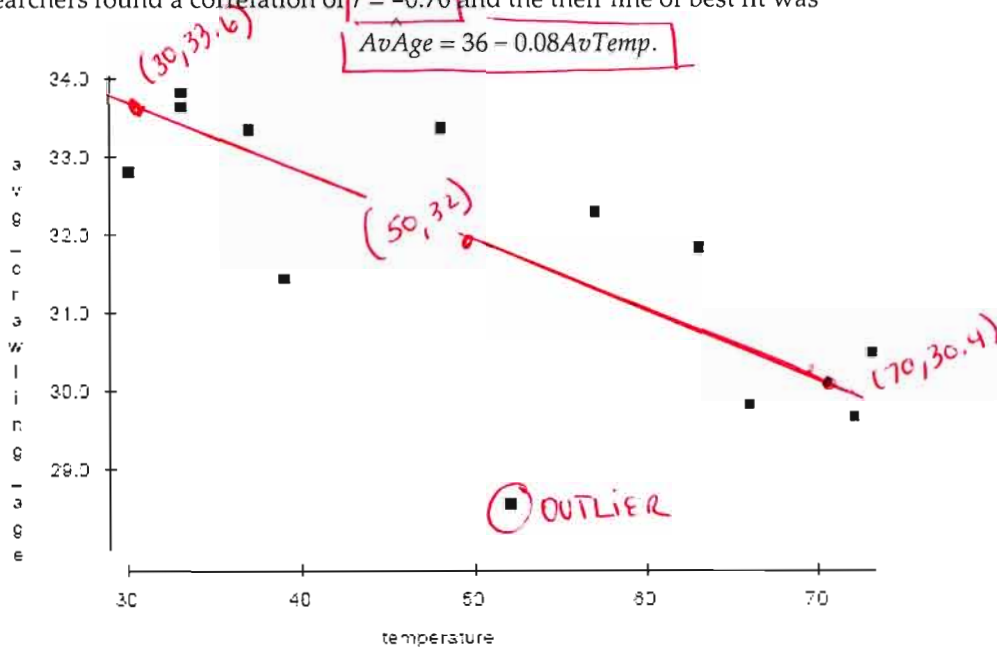
a. Draw the line of best fit on the graph. (Show your method clearly.)

plot   2 points $(30, 33.6)$ $(50, 32)$ $(70, 30.4)$
                    $(40, 32.8)$ $(60, 31.2)$

b. Describe the association in context (4pts)

Need 4 of 5

"The association is linear, moderately strong, Negative, with 1 outlier.

CONTEXT: Children seem to crawl earlier when the temperature
is higher (+1 though there was on unusually early age observed
for a temperature about 50°)

c. Explain (in context) what the slope of the line means (4pts)

The model suggests, on average, babies crawl about .8 weeks
earlier for every 10° higher the temperature is.

d. Explain (in context) what the y-intercept of the line means (4pts)

3pts

The model predicts that at a temperature of 0° babies
would crawl at about 36 weeks old

1pt

ALTHOUGH THIS MAY NOT[6] MEAN MUCH AS NO DATA WERE
COLLECTED AT SUCH COLD TEMPERATURES.

e. Explain (in context) what r^2 means (4pts) — 2pts ←

ABOUT 49 % OF THE VARIABILITY IN CRAWLING AGE CAN BE EXPLAINED BY VARIATIONS IN TEMPERATURE

f. In this context, what does a negative residual indicate? (4pts)

A NEGATIVE RESIDUAL WOULD ~~MXXXX~~ INDICATE THAT BABIES CRAWLED AT A YOUNGER AGE THAN THE MODEL PREDICTED.
3pts   1pt

16) **Penicillin** Doctors studying how the human body assimilates medication inject some patients with penicillin, and then monitor the concentration of the drug (in units/cc) in the patients' blood for seven hours. The data are shown in the scatterplot. First they tried to fit a linear model. The regression analysis and residuals plot are shown.
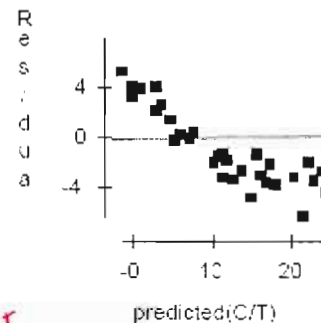
16) _____

Dependent variable is: Concentration
No Selector
R squared = 90.8% R squared (adjusted) = 90.6%
s = 3.472 with 43 - 2 = 41 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4900.55	1	4900.55	407
Residual	494.199	41	12.0536	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	40.3266	1.295	31.1	≤ 0.0001
Time	-5.95956	0.2956	-20.2	≤ 0.0001

Res: dua   predicted(C/T)

a. Find the correlation between time and concentration. __−.953__
1pt   2pts
√.908

b. Using this model, estimate what the concentration of penicillin will be after 4 hours. (show work)

$\overset{\wedge}{Concentration}$ = 40.3266 − 5.95956 (4) = 16.48836

About 16.5 UNITS/CC penicillin

c. skipped.

Now the researchers try a new model, using the re-expression log(*Concentration*). Examine the regression analysis and the residuals plot below.

Dependent variable is: LogCnn
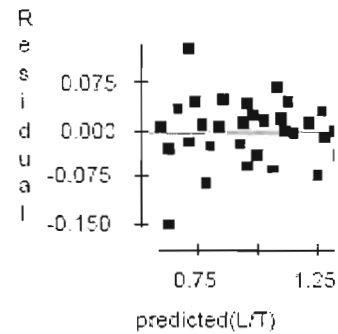No Selector
R squared = 98.0%   R squared (adjusted) = 98.0%
s = 0.0451 with 43 - 2 = 41 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4.11395	1	4.11395	2022
Residual	0.083412	41	0.002034	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.80184	0.0168	107	≤ 0.0001
Time	-0.172672	0.0038	-45.0	≤ 0.0001

Residual plot: predicted(L/T) on x-axis (0.75, 1.25); Residual on y-axis (0.075, 0.000, -0.075, -0.150)

d. Explain why you think this model is better than the original linear model.

The residuals show a random pattern with no curvatures

e. Using this new model, estimate the concentration of penicillin after 4 hours.(show work)

$$\log \hat{Cnn} = 1.80184 - .172672 (4)$$

$$\log_{10} \hat{Cnn} = 1.1111152$$

$$\hat{Cnn} = 12.9167127$$

About 12.9 UNITS/cc penicillin

8

AP Statistics – Chapter 5B Test

(All MC and subquestions are 3pts, except where noted – TOTAL 88 POINTS)

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1) It takes a while for new factory workers to master a complex assembly process. During the first month   1) ___ E ___
new employees work, the company tracks the number of days they have been on the job and the length
of time it takes them to complete an assembly. The correlation is most likely to be
    A) near +0.6
    B) near 0
    C) exactly +1.0
    D) exactly –1.0
    (E) near –0.6

2) For families who live in apartments the correlation between the family's income and the amount of rent   2) ___ E ___
they pay is $r = 0.60$. Which is true?
   ✓ I. In general, families with higher incomes pay more in rent.
   II. On average, families spend 60% of their income on rent. ✗
   III. The regression line passes through 60% of the (income$, rent$) data points. ✗
    A) I, II, and III
    B) I and III only
    C) II only
    D) I and II only
    (E) I only
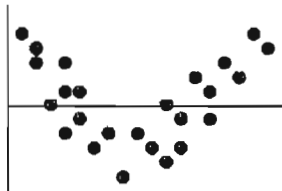
3) Variables X and Y have $r = 0.40$. If we decrease each X value by 0.1, double each Y value, and then   3) ___ E ___
interchange them (put X on the Y– axis and vice versa) the new correlation will be
    A) 0.80   B) 0.15   C) –0.40   D) 0.60   (E) 0.40

4) The residuals plot for a linear model is shown. Which is true?   4) ___ B ___

    A) The linear model is okay because the association between the two variables is fairly strong.
    (B) The linear model is no good because of the curve in the residuals.
    C) The linear model is no good because the correlation is near 0.
    D) The linear model is no good because some residuals are large.
    E) The linear model is okay because approximately the same number of points are above the line as
       below it.

5) A regression model examining the amount of weight a football player can bench press found that 10 cm   5) ___ D ___
differences in chest size are associated with 8 kg differences in weight pressed. Which is true?
   I. The correlation between chest size and weight pressed is $r = 0.80$ ✗
  H   II. As a player gets stronger and presses more weight his chest will get bigger. ✗
   III. A positive residual indicates that the player pressed more weight than predicted.
    A) none   B) I only   C) I and III   (D) III only   E) I and II

1

6) When using midterm exam scores to predict a student's final grade in a class, the student would prefer to have a   6) ___D___
   A) negative residual, because that means the student's final grade is lower than we would predict with the model.
   B) residual equal to zero, because that means the student's final grade is exactly what we would predict with the model.
   C) positive residual, because that means the student's final grade is lower than we would predict with the model.
   D) positive residual, because that means the student's final grade is higher than we would predict with the model.
   E) negative residual, because that means the student's final grade is higher than we would predict with the model.

7) This regression analysis examines the relationship between the <u>number of years of formal education a</u> person has and their <u>annual income</u>. According to this model, about how much more money do people who finish a 4-year college program earn each year, on average, than those with only a 2-year degree?   7) ___D___

Dependent variable is **Income**
R-squared = 25.8%
s = 3888 with 57 degrees of freedom

Variable	Coefficient	s.e. of Coeff
Constant	3984.45	6600
Education	2668.45	600.1

4 YR $\hat{y} = 3984.45 + 2668.45\,(4)$
      $= 14,658$

2 YR $\hat{y} = 9321.4$

OR $2\,(2668.45)$

A) \$2006   B) \$9321   C) \$2710   D) \$5337   E) \$7968

Tip
Put $\hat{y}$ in y1
Set + b_1 set to
Set Auto to Table
Go to Table
Set x's
+

8) A correlation of zero between two quantitative variables means that   8) ___E___
   A) re-expressing the data will guarantee a linear association between the two variables.
   B) there is no association between the two variables.
   C) we have done something wrong in our calculation of r.
   D) none of these
   E) there is no linear association between the two variables.

9) If the point in the upper left corner of the scatterplot is removed, what will happen to the correlation (r) and the slope of the line of best fit (b)?   9) ___C___
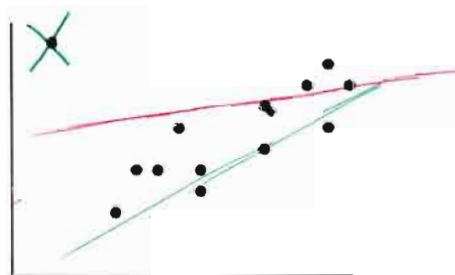
Slope increases

R increases

A) Both will decrease.
B) r will increase and b will decrease.
C) Both will increase.
D) r will decrease and b will increase.
E) They will not change.

2

10) A regression analysis of students' college grade point averages (GPAs) and their high school GPAs found $R^2 = 0.311$. Which of these is true?

    10) **C**

I. High school GPA accounts for 31.1% of college GPA.
II. 31.1% of college GPAs can be correctly predicted with this model.
III. 31.1% of the variance in college GPA can be accounted for by the model. ✓

    A) I and II   B) II only   **C) III only**   D) None   E) I only

11) All but one of the statements below contain a mistake. Which one could be true?

    11) **B**

    A) The correlation between gender and age is –0.171.
    B) The correlation between weight and length of foot is 0.488.
    C) The correlation between height and weight is 0.568 inches per pound. *NO UNITS*
    D) If the correlation between blood alcohol level and reaction time is 0.73, then the correlation between reaction time and blood alcohol level is –0.73.
    E) The correlation between the breed of a dog and its weight is 0.435. *MUST BE QUANTITATIVE*

12) The model $\sqrt{\widehat{distance}} = 3.30 + 0.235 \times (speed)$ can be used to predict the stopping distance (in feet) for a car traveling at a specific speed (in mph). According to this model, about how much distance will a car going 65 mph need to stop?

    12) **C**

    A) 27.0 feet   B) 729.0 feet   **C) 345.0 feet**   D) 18.6 feet   E) 4.3 feet

$$\sqrt{distance} = 3.30 + .235(65)$$

$$\left(\sqrt{distance}\right)^2 = (18.575)^2$$

$$distance \approx 345$$

3

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

13) **Subaru costs** Data collected from internet ads for 1999 Subarus were used to create a model to   13) _____
estimate
the asking price of the car based on the number of miles it had been driven. The model has $r^2 =$
0.47
and equation $\hat{Price} = 15{,}327 - 0.11(Miles)$. The plot of residuals versus the predicted price is
shown.
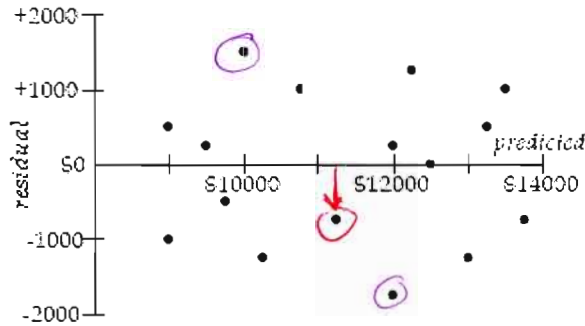
a. Do you think you could make accurate estimates of Subaru prices with this model? Explain

2pts — ESTIMATES OF PRICE WILL ONLY BE MODERATELY ACCURATE.

1pt — USING THE MODEL, MILEAGE EXPLAINS ONLY
47% OF THE VARIABILITY IN PRICE

+1 — AND SOME OF THE RESIDUALS OFF BY $2,000

b. Interpret the slope of the line (n context).

1pt → SLOPE = -.11/mile

2pts → The model predicts that for every additional mile
the car has been driven the price of the
Car would decrease by 11¢'s on average.

c. One of the cars in the data set had been driven 42,000 miles. How much was the owner
asking for it? (Show work.)
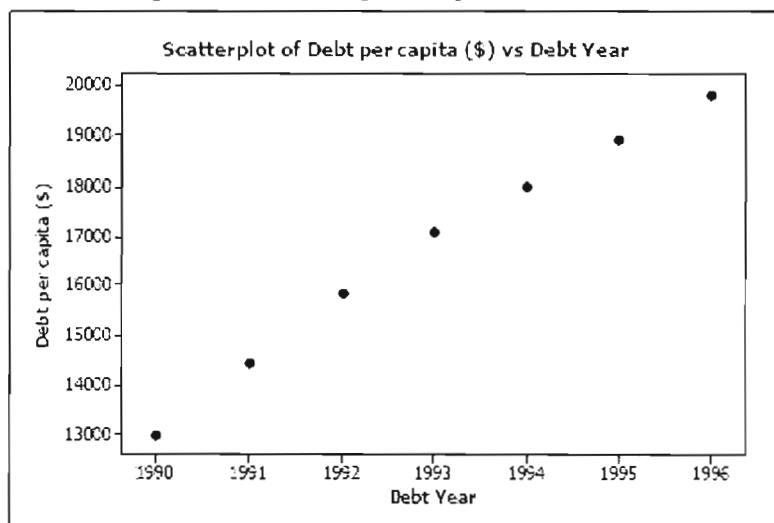
$\hat{Price} = 15{,}327 - .11(42000) = \$10{,}707$ ← 1pt

Residual = -1,000 ← 1pt

ASKING PRICE WAS $11,707 ← 1pt

4

14) **Personal debt** According to *The World Almanac and Book of Facts 2004*, the debt per capita for the years 1990–2001 gives the following scatterplot: 14) _____

Scatterplot of Debt per capita ($) vs Debt Year

Regression output gives the equation of the regression line as

$\hat{Debt}$ = –2,231,226 + 1128(Year) with R^2 = 98.8%. $\sqrt{.988}$

a. What is the response variable?   **DEBT PER CAPITA**

b. What is the correlation coefficient   **.994**

c. Explain in context what the slope of the line means.   **SLOPE IS $1,128/yr ← 1pt**

ON AVERAGE, DEBT PER CAPITA INCREASES ← 2PTS **ABOUT $1,128 PER YEAR.**

d. Explain in context what R^2 = 98.8% means.

ABOUT 98.8% OF THE VARIABILITY IN DEBT PER CAPITA IS EXPLAINED BY THE MODEL

1 p+   **2PTS**

e. You decide to take a look at a residuals plot before making any predictions. Based on the following residuals plot, does linear regression seem appropriate for these data? Explain
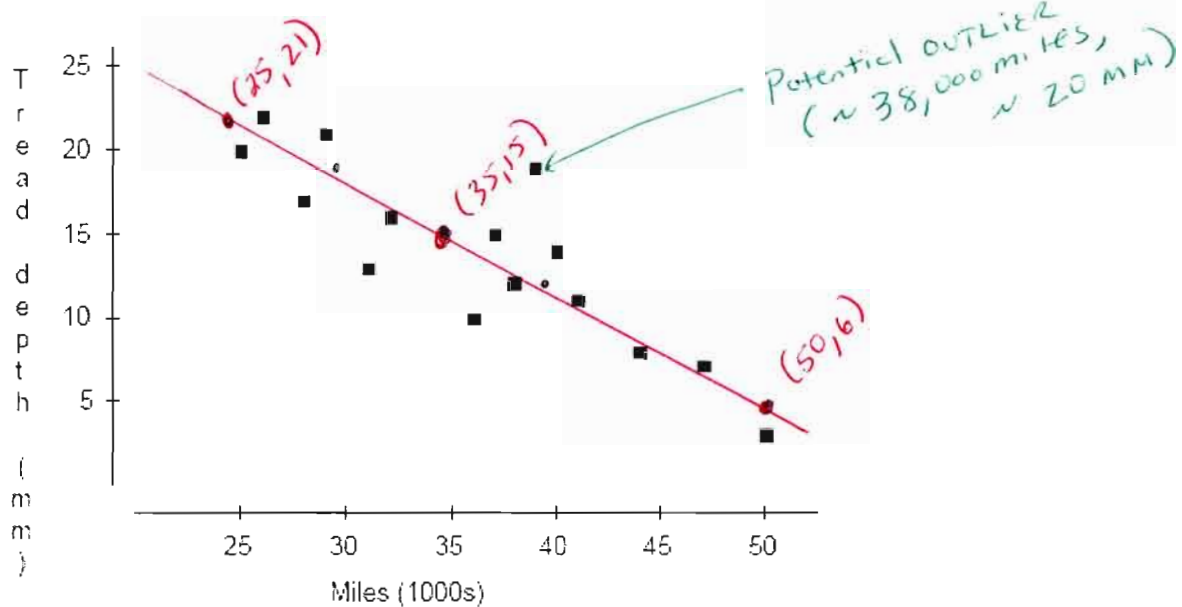

Residuals Versus the Fitted Values
(response is Debt per capita ($))
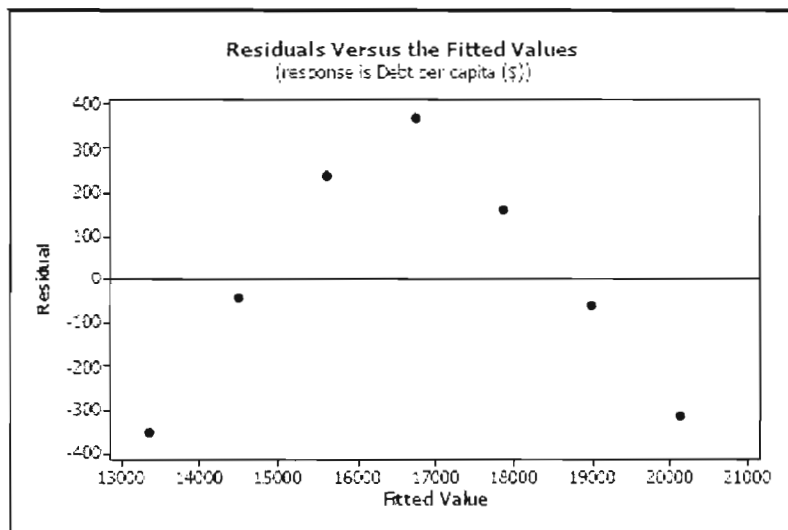
NO, THE LINEAR MODEL IS NOT APPROPRIATE

BECAUSE THERE IS A DEFINITE CURVE IN THE RESIDUAL PLOT WHICH WAS NOT OBVIOUS IN THE ORIGINAL SCATTER PLOT.

5

15) **Taxi tires** A taxi company monitoring the safety of its cabs kept track of the number of miles tires had been driven (in thousands) and the depth of the tread remaining (in mm). Their data are displayed in the scatterplot. They found the equation of the least squares regression line to be $\widehat{tread} = 36 - 0.6 miles$, with $r^2 = 0.74$.

15) _____

Tread depth (mm) vs Miles (1000s)

Handwritten annotations on graph: (25, 21), (35, 15), (50, 6), Potential OUTLIER (~38,000 miles, ~20 MM)

a. Draw the line of best fit on the graph. (Show your method clearly.)

PLOT 2 POINTS (25, 21) (35, 15) (45, 9)
(30, 18) (40, 12) (50, 6)

b. What is the explanatory variable?

The number of miles, TIRES HAD BEEN DRIVEN,
_____2 PT_____   I PT

SIGN IPT

c. The correlation $r = -.86$ √.74
2PTS

d. Describe the association in context. ← CONTEXT

CONTEXT IPT
2pts for elements →

THE ASSOCIATION BETWEEN NUMBER OF MILES TIRES DRIVEN (1000's) AND THE TIRE TREAD DEPTH (IN mm) IS A MODERATELY STRONG, NEGATIVE LINEAR RELATIONSHIP WITH 1 POSSIBLE OUTLIER (~38,000 miles, 20mm)

e. Explain (in context) what the slope of the line means.

IIPT →

The model predicts that for every additional 1,000 miles the tires are driven, the depth of the tire tread decreases by about .6 mm. /2pts/

f. Explain (in context) what the y-intercept of the line means.

The model predicts that brand New tires (number of miles is 0) have tread averaging 36mm deep.

g. Explain (in context) what r^2 means.

EACH PART IPT →

r^2 means that 74% of the variability in tread depth is explained by the variations in the number of miles the tires have been driven

h. In this context, what does a negative residual mean?

IPT →

Residuals EQUAL THE OBSERVED TREAD DEPTH MINUS THE PREDICTED DEPTH. A NEGATIVE RESIDUAL MEANS THE TIRE TREAD IS WEARING OUT FASTER THAN PREDICTED OR THE OBSERVED AMOUNT OF TREAD DEPTH IS LESS THAN THE PREDICTED AMOUNT OF TREAD DEPTH

2PTS

6

16) A forester would like to know how big a maple tree might be at age 50 years. She gathers data from some trees that have been cut down, and plots the diameters (in inches) of the trees against their ages (in years). First she makes a linear model. The scatterplot and residuals plot are shown.

16) _____
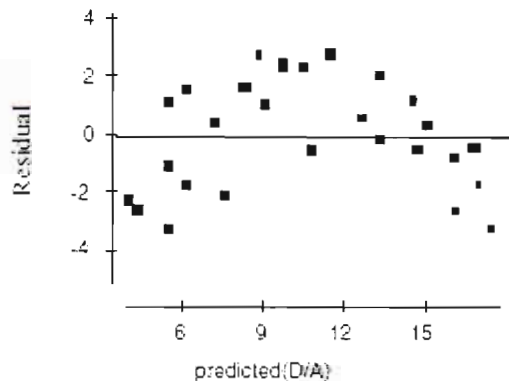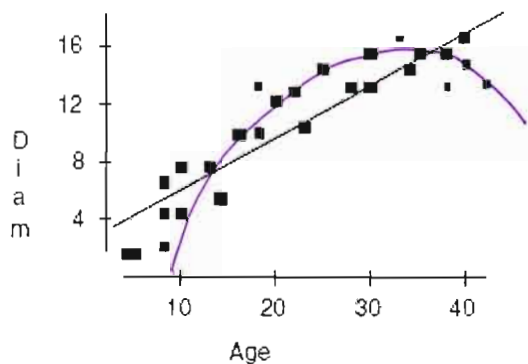
a. Describe the association shown in the scatterplot.

1pt context

THE ASSOCIATION BETWEEN AGE OF MAPLE TREES AND TREE

*2pts* → DIAMETER IS MODERATELY STRONG; POSITIVE; AND CURVED (NOT LINEAR)

b. Do you think the linear model is appropriate? Explain.

NO, THE SCATTER PLOT SHOWS A SLIGHT CURVE, THE RESIDUAL PLOT SHOWS A CURVED PATTERN WHICH MAKE US CONCLUDE THAT A LINEAR MODEL IS NOT APPROPRIATE.

3pts Residual Plot

c. skipped

Now she re-expresses the data, using the logarithm of age to try to predict the diameter of the tree. Here are the regression analysis and the residuals plot.

Dependent variable is: Diam
R squared = 84.3%

Variable	Coefficient	s.e. of Coeff
Constant	- 8.60770	1.681
Log(Age)	15.0701	1.299
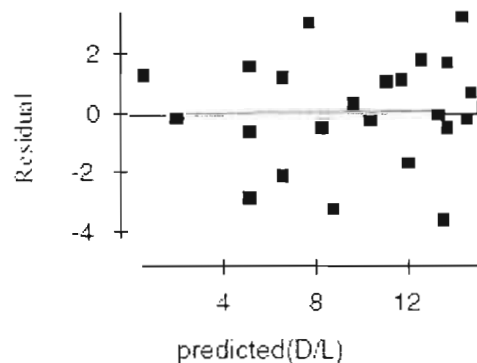
d. Explain why you think this is a better model.

The residual plot for the transformed data shows No obvious pattern, The residuals are randomly spread out

e. Using this model, predict the diameter of a maple tree at age 50 years.

$$\widehat{diam} = -8.60770 + 15.0701 \cdot \log(50)$$ ← 1pt

$$\widehat{diam} = 16.996$$ ← 1pt

The diameter is about 17 inches ← 1pt

7