4.2 Exercise **MATH 241** **THOMPSON**

1. Complete parts (a) through (c) by hand for the data below.

x	3	4	5	7	8
y	5	8	10	17	20

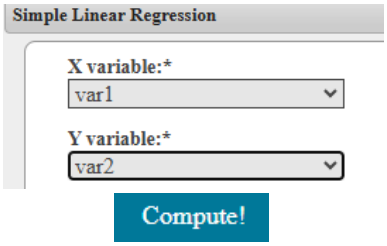
,

STATCRUNCH in new window

STAT – REGRESSION – SIMPLE LINEAR

(a) By hand, draw a scatter diagram treating x as the explanatory variable and y as the response variable. Cho

var1	var2
3	5
4	8
5	10
7	17
8	20



Simple linear regression results:
Dependent Variable: var2
Independent Variable: var1
var2 = -4.3255814 + 3.0232558 var1
Sample size: 5
R (correlation coefficient) = 0.99749465
R-sq = 0.99499558
Estimate of error standard deviation: 0.51338669

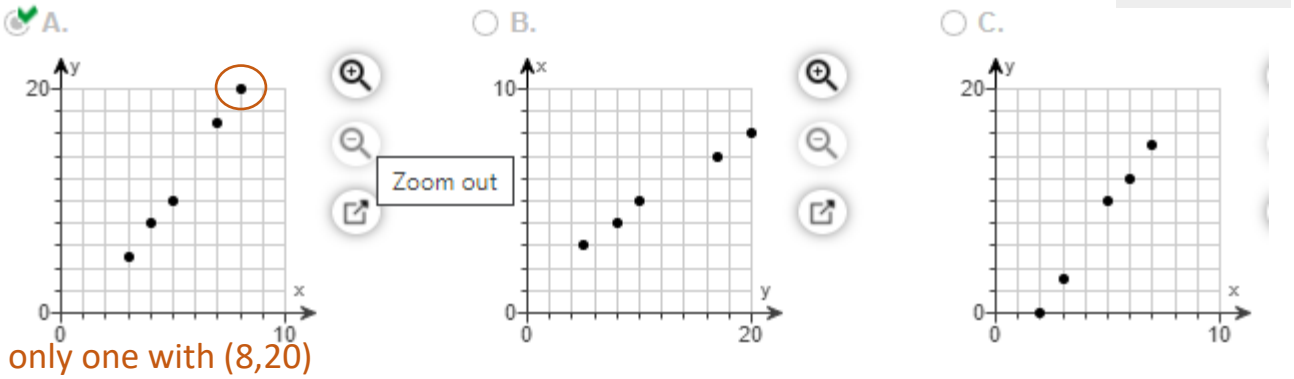
var1 is x and var2 is y

var2 = -4.3255814 + 3.0232558 var1

y = -4 + 3x

put in order **y = 3x – 4**

click right arrow to see graph



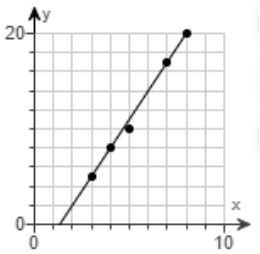
only one with (8,20)

(b) Find the equation of the line containing the points (3,5) and (8,20).

$y = 3x + (-4)$

(c) Graph the line found in part (b) on the scatter diagram. Choose the correct graph below.

same scatter plot with best fit line



2. An engineer wants to determine how the weight of a car, x in pounds, affects gas mileage, y in miles per gallon. From the data collected from cars ranging in weight from 2500 to 3500 pounds, it is determined that the least-squares regression line is $\hat{y} = -0.0062x + 41.4640$. Perdict the gas mileage of a Ford Mustang if it weighs 3300 lb.

y = -0.0062(3300) + 41.464

- A. 62 mpg
- B. Cannot be determined
- C. 25 mpg
- D. 21 mpg ✔
- E. 18 mpg

3. Use your calculator to complete parts (a) and (b) for the data below.

x	3	4	5	7	8
y	6	9	11	18	21

STATCRUNCH in new window

STAT – REGRESSION – SIMPLE LINEAR

(a) Determine the least-squares regression line.

StatCrunch ▾ Applets ▾ Edit ▾

Row	var1	var2
1	3	6
2	4	9
3	5	11
4	7	18
5	8	21

Simple linear regression results:
Dependent Variable: var2
Independent Variable: var1
var2 = -3.3255814 + 3.0232558 var1
Sample size: 5
R (correlation coefficient) = 0.99749465
R-sq = 0.99499558

y = C+mx then put in y = mx + b form

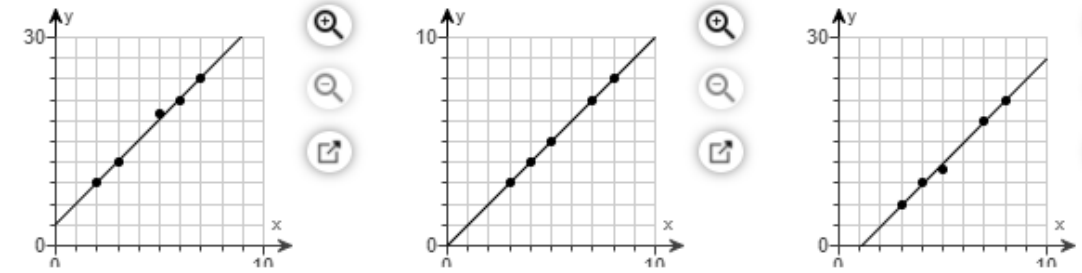
$\hat{y} = 3.023x + (-3.326)$

(Round to three decimal places as needed.)

Click arrow to get graph

(b) Graph the least-squares regression line on the scatter diagram. Choose the correct graph below.

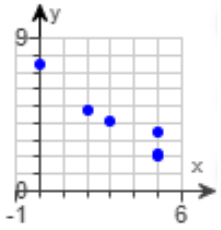
- A.
- B.
- C. ✔

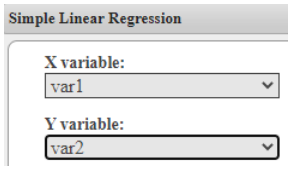


4. The scatter diagram for the data set below is shown. Determine the least-squares regression line.

STAT – REGRESSION – SIMPLE LINEAR

x	0	2	3	5	5	5
y	7.5	4.8	4.1	2.1	3.5	2.2





$\hat{y} = -.927x + 7.122$

(Round to three decimal places as needed.)

Click arrow to get graph

< >

5. A data set is given below.

(a) Draw a scatter diagram. Comment on the type of relation that appears to exist between x and y.

(b) Given that $\bar{x} = 3.5000$, $s_x = 2.5884$, $\bar{y} = 3.9833$, $s_y = 2.0114$, and $r = -0.9508$, determine the least-squares regression line

(c) Graph the least-squares regression line on the scatter diagram drawn in part (a).

x	0	1	3	5	6	6
y	5.7	6.4	5.2	2.7	1.8	2.1

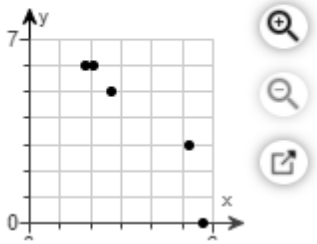
STATCRUNCH – STAT – REGRESSION – SIMPLE LINEAR

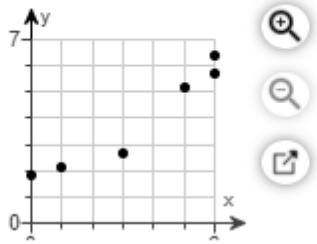
(a) Choose the correct graph below.

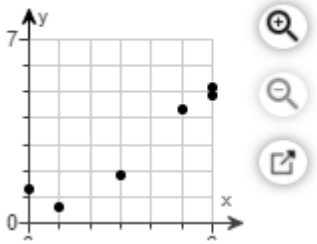
CLICK ARROW FOR GRAPH

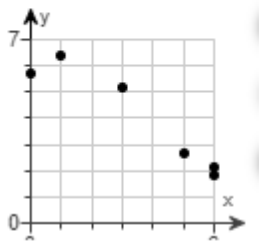
< >

A. B. C. D. ✓









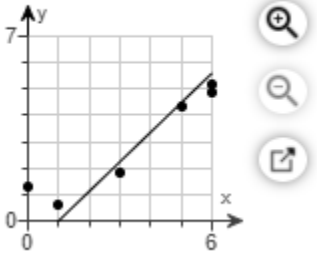
There appears to be a linear, negative relationship.

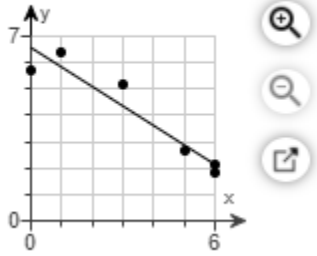
STAT – REGRESSION – SIMPLE LINEAR

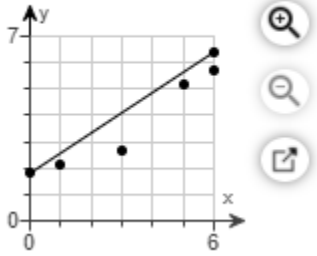
(b) $\hat{y} = -.739x + (6.569)$

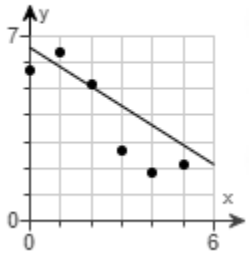
(Round to three decimal places as needed.)

A. B. ✓ C. D.

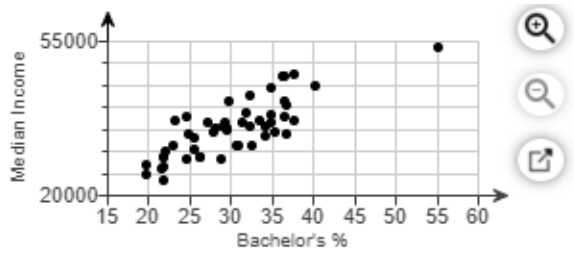








6. The least-squares regression equation is $\hat{y} = 780.1x + 11{,}816$ where y is the median income and x is the percentage of 25 years and older with at least a bachelor's degree in the region. The scatter diagram indicates a linear relation between the two variables with a correlation coefficient of 0.8048. Complete parts (a) through (d).



y = 780.1(30) + 11816

(a) Predict the median income of a region in which 30% of adults 25 years and older have at least a bachelor's degree.

$ 35219 (Round to the nearest dollar as needed.)

(b) In a particular region, 29.6 percent of adults 25 years and older have at least a bachelor's degree. The median income in this region is $38,370. Is this income higher than what you would expect? Why?

y = 780.1(29.6) + 11816

This is higher than expected because the expected income is $ 34907
(Round to the nearest dollar as needed.)

(c) Interpret the slope. Select the correct choice below and fill in the answer box to complete your choice.
(Type an integer or decimal. Do not round.)

- A. For a median income of $0, the percent of adults with a bachelor's degree is ____ %.
- B. For 0% of adults having a bachelor's degree, the median income is predicted to be $ ____ .
- ✔ C. For every percent increase in adults having at least a bachelor's degree, the median income increases by $ 780.1 , on average.

(d) Explain why it does not make sense to interpret the y-intercept. Choose the correct answer below.

- A. It does not make sense to interpret the y-intercept because an x-value of 0 does not make sense.
- ✔ B. It does not make sense to interpret the y-intercept because an x-value of 0 is outside the scope of the model.

7. Is there a relation between the age difference between husband/wives and the percent of a country that is literate? Researchers found the least-squares regression between age difference (husband age minus wife age), y, and literacy rate (percent of the population that is literate), x, is $\hat{y} = -0.0469x + 6.9$. The model applied for $20 \le x \le 100$. Complete parts (a) through (e) below.

(a) Interpret the slope. Select the correct choice below and fill in the answer box to complete your choice.

For every unit increase in literacy rate, the age difference falls by .0469 units, on average.
(Type an integer or decimal. Do not round.)

(b) Does it make sense to interpret the y-intercept? Explain. Choose the correct answer below.

- A. Yes—it makes sense to interpret the y-intercept because an x-value of 0 is within the realm of possibilities.
- B. No—it does not make sense to interpret the y-intercept because an x-value of 0 is impossible.
- C. No—it does not make sense to interpret the y-intercept because a y-value of 0 is outside the scope of the model.
- ✔ D. No—it does not make sense to interpret the y-intercept because an x-value of 0 is outside the scope of the model.

(c) Predict the age difference between husband/wife in a country where the literacy rate is 27 percent.

5.6 years (Round to one decimal place as needed.) y = -0.0469(27) + 6.9

(d) Would it make sense to use this model to predict the age difference between husband/wife in a country where the literacy rate is 12%?

- A. No—it does not make sense because a y-value of 12 is outside the scope of the model.
- ✔ B. No—it does not make sense because an x-value of 12 is outside the scope of the model.

(e) The literacy rate in a country is 97% and the age difference between husbands and wives is 2 years. Is this age difference above or below the average age difference among all countries whose literacy rate is 97%? Select the correct choice below and fill in the answer box to complete your choice.
(Round to one decimal place as needed.)

$y = -0.0469(97) + 6.9$

A. Above—the average age difference among all countries whose literacy rate is 97% is ____ years.

B. Below—the average age difference among all countries whose literacy rate is 97% is 2.4 years.

8. A pediatrician wants to determine the relation that exists between a child's height, x, and head circumference, y. She randomly selects 11 children from her practice, measures their heights and head circumferences, and obtains the accompanying data. Complete parts (a) through (g) below.

Click the icon to view the children's data.

(a) Find the least-squares regression line treating height as the explanatory variable and head circumference as the response variable.

STATCRUNCH – STAT – REGRESSION – SIMPLE LINEAR

$\hat{y} = .155x + (13.3)$ *make sure to read blue for all decimal places

(Round the slope to three decimal places and round the constant to one decimal place as needed.)

(b) Interpret the slope and y-intercept, if appropriate.

First interpret the slope. Select the correct choice below and, if necessary, fill in the answer box to complete your choice.

A. For a height of 0 inches, the head circumference is predicted to be ____ in.
(Round to three decimal places as needed.)

B. For every inch increase in height, the head circumference increases by .155 in., on average.
(Round to three decimal places as needed.)

Interpret the y-intercept, if appropriate. Select the correct choice below and, if necessary, fill in the answer box to complete your choice.

E. It is not appropriate to interpret the y-intercept.

(c) Use the regression equation to predict the head circumference of a child who is 24.25 inches tall.

$\hat{y} =$ 17.06 in.

Plug in 24.25 in for x into the equation: y=0.155(24.25)+133=19.06

(Round to two decimal places as needed.)

(d) Compute the residual based on the observed head circumference of the 24.25-inch-tall child in the table. Is the head circumference of this child above or below the value predicted by the regression model?

Get the y value when x = 24.25 in table (16.9) then subtract $y - \hat{y}$

The residual for this observation is −.16, meaning that the head circumference of this child is below the value predicted by the regression model.

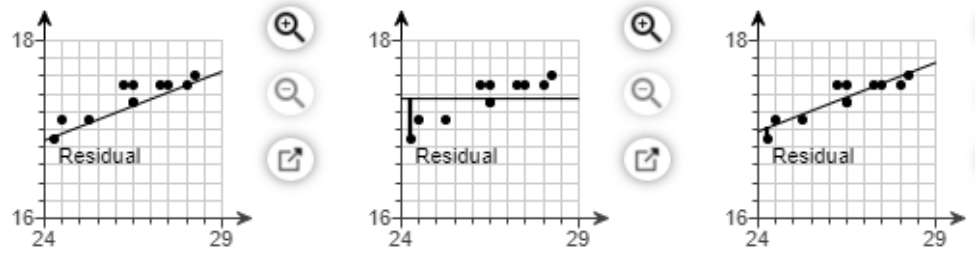
16.9-17.06= -.16 BELOW because it is negative

(Round to two decimal places as needed.)

(e) Draw the least-squares regression line on the scatter diagram of the data and label the residual from part (d). Choose the correct graph below.

Height (inches), x	Head Circumference (inches), y
27.75	17.6
24.75	17.2
25.75	17.3
26	17.6
24.25	16.9
28	17.8
26.5	17.4
27	17.6
26.5	17.4
26.5	17.6
27.75	17.6

A. B. C.



Click right arrow in the bottom to get graph

(f) Notice that two children are 26.5 inches tall. One has a head circumference of 17.3 inches; the other has a head circumference of 17.5 inches. How can this be?

- A. The only explanation is that the difference was caused by measurement error.
- B. For children with a height of 26.5 inches, head circumferences vary.

(g) Would it be reasonable to use the least-squares regression line to predict the head circumference of a child who was 32 inches tall? Why?

- A. Yes—the calculated model can be used for any child's height.
- B. No—this height is not possible.
- C. Yes—this height is possible and within the scope of the model.
- D. No—this height is outside the scope of the model.

9. The data below represent the number of days absent, x, and the final grade, y, for a sample of college students at a large univers
Complete parts (a) through (e) below.

No. of absences, x	0	1	2	3	4	5	6	7	8	9
Final grade, y	89.1	86.2	83.2	80.6	77.6	73.1	63.4	67.9	65.0	62.0

(a) Find the least-squares regression line treating the number of absences, x, as the explanatory variable and the final grade, y, as the response variable.

STATCRUNCH – STAT – REGRESSION – SIMPLE LINEAR

$\hat{y} =$ −3.181 x + (89.125)
(Round to three decimal places as needed.)

(b) Interpret the slope and y-intercept, if appropriate.

Interpret the slope. Select the correct choice below and, if necessary, fill in the answer box to complete your choice.
(Round to three decimal places as needed.)

- D. For every day absent, the final grade falls by 3.181 , on average.
- E. It is not appropriate to interpret the slope.

Interpret the y-intercept. Select the correct choice below and, if necessary, fill in the answer box to complete your choice.
(Round to three decimal places as needed.)

- A. For every unit change in the final grade, the number of days absent falls by ____ days, on average.
- B. For every day absent, the final grade falls by ____ , on average.
- C. For zero days absent, the final score is predicted to be 89.125 .

(c) Predict the final grade for a student who misses five class periods and compute the residual. Is the observed final grade above or below average for this number of absences?

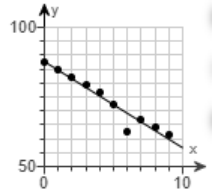
The predicted final grade is 73.2 . This observation has a residual of −.1 , which indicates that the final grade is below average.
(Round to one decimal place as needed.)

Predicted grade – plug 5 into equation

73.1 from the table at 5 then subtract 73.1 – 73.2 $y - \hat{y} = -0.1$

(d) Draw the least-squares regression line on the scatter diagram of the data. Choose the correct graph below.

B.



. No—15 missed class periods is outside the scope of the model.

< >

10. The data in the table represent the number of licensed drivers in various age groups and the number o fatal accidents within the age group by gender. Complete parts (a) to (c) below.

Click the icon to view the data table.

(a) Find the least-squares regression line for males treating the number of licensed drivers as the explanatory variable, x, and the number of fatal crashes, y, as the response variable. Repeat this procedure for females.

Find the least-squares regression line for males.

$\hat{y}$ = .343 x + 998 x is males y is crashes

(Round the slope to three decimal places and round the constant to the nearest integer as needed.)

Find the least-squares regression line for females.

$\hat{y}$ = .104 x + 517 x is females y is crashes

(Round the slope to three decimal places and round the constant to the nearest integer as needed.)

(b) Interpret the slope of the least-squares regression line for each gender, if appropriate. How might an insurance company use this information?

A. If the number of male licensed drivers increases by 1 (thousand), then the number of fatal crashes increases by .343, on average.
(Round to three decimal places as needed.)

A. If the number of female licensed drivers increases by 1 (thousand), then the number of fatal crashes increases by .104, on average.
(Round to three decimal places as needed.)

The slope of the regression line for males is greater than that for females. This means that males tend to be involved in more fatal crashes than females. An insurance company may use this information to argue for higher rates for male customers.

(c) Was the number of fatal accidents for 16 to 20 year old males above or below average? Was the number of fatal accidents for 21 to 24 year old males above or below average? Was the number of fatal accidents for males greater than 74 years old above or below average? How might an insurance company use this information? Does the same relationship hold for females?

Plug in the x from table into the equation and Compare the two

The number of fatal accidents for 16 to 20 year old males was above average. The number of fatal accidents for 21 to 24 year old males was above average. The number of fatal accidents for males greater than 74 years old was below average.

Ex: 16 to 20 y=0.343(6242)+998 = 3139 The average of 3139 is above the actual of 5180

An insurance company could use it to argue for higher rates for younger drivers and lower rates for older drivers.

Does the same relationship hold for females?

Yes

11. Explain what each point on the least-squares regression line represents.

Choose the correct answer below.

- A. Each point on the least-squares regression line represents the y-value of the data set at that corresponding value of x.
- B. Each point on the least-squares regression line represents the y-values that would be considered ideal at that corresponding value of x.
- ✔ C. Each point on the least-squares regression line represents the predicted y-value at the corresponding value of x.

12. If the linear correlation between two variables is negative, what can be said about the slope of the regression line?

Choose the correct answer below.

- A. More information is needed
- B. Positive
- ✔ C. Negative