

## 9.1 Scatter Plots and Linear Correlation

### Answers

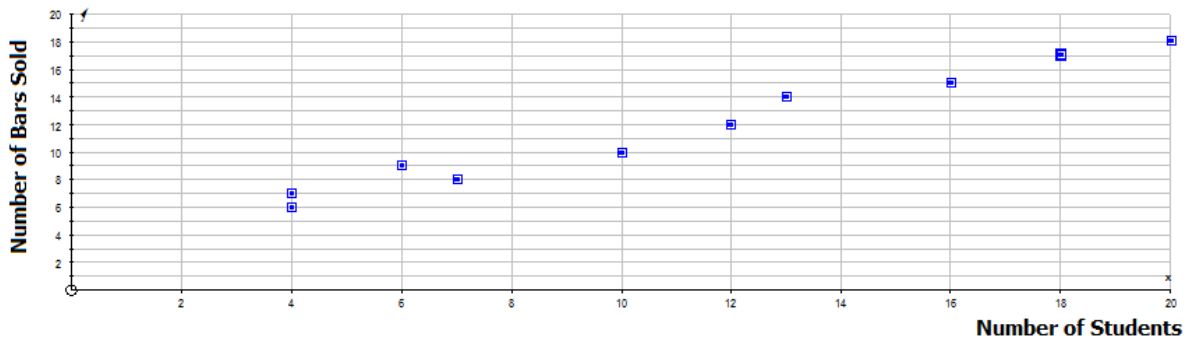
1. A high school psychologist wants to conduct a survey to answer the question: "Is there a relationship between a student's athletic ability and his/her popularity?"

A newly wed wants to surprise her husband with a turkey dinner for supper. She checks her cookbook to find out the time needed to cook turkey which weighs 13 pounds. In the cookbook she finds the following chart:

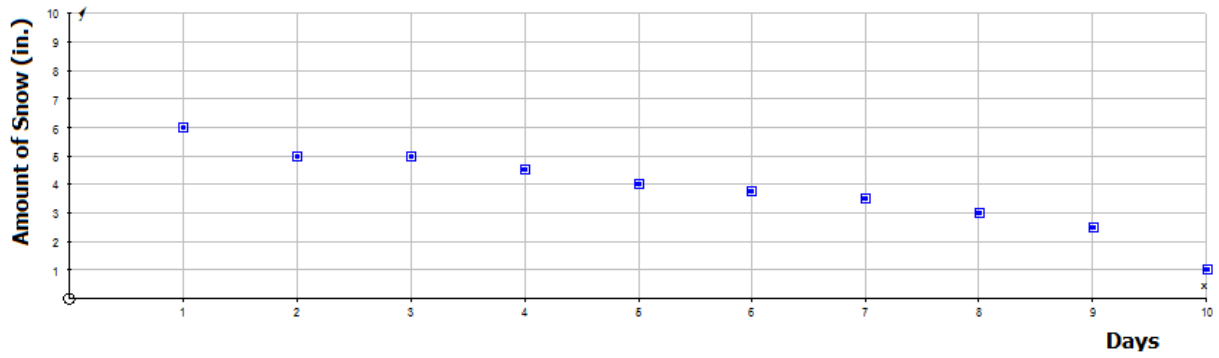
Weight (Pounds)	Cooking Time (Hours)
6	4.0
8	4.5
11	5.0
15	6.0
19	7.25
23	8.0

She will use this data to determine a rule to use to find out the time needed to cook a turkey given the weight of the turkey.

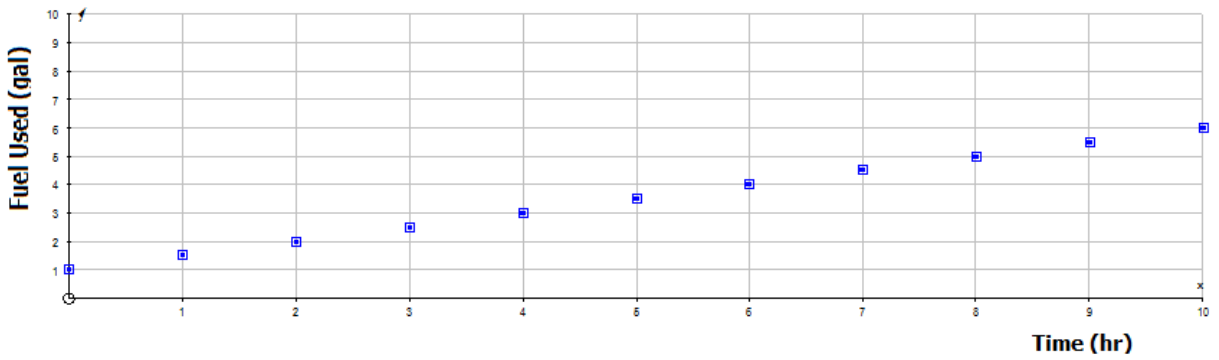
2. a) Positive Correlation



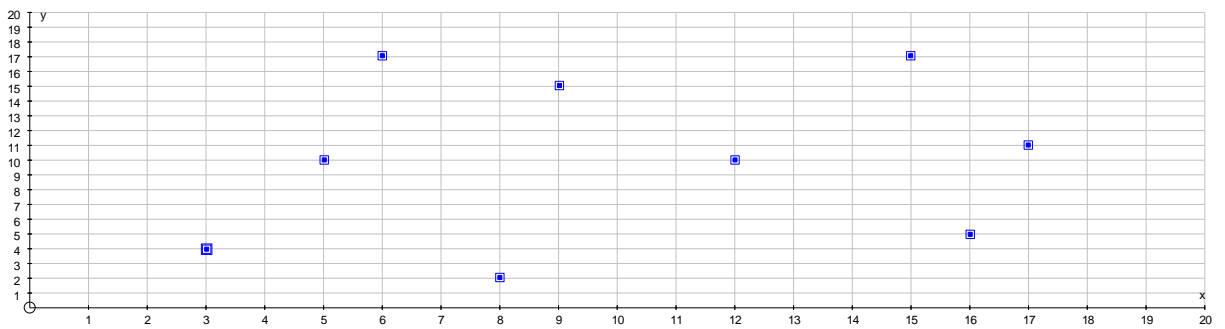
b) Negative Correlation



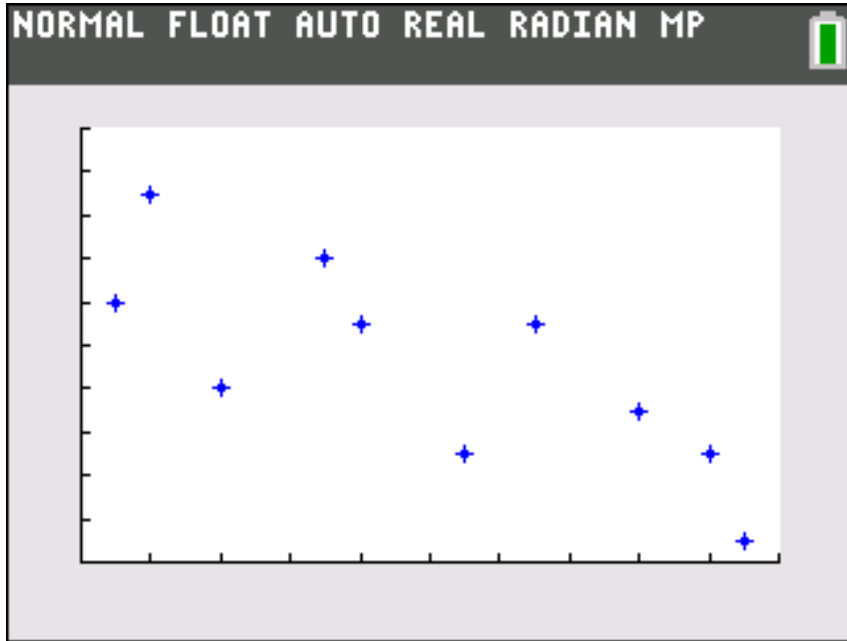
c) Perfect Correlation



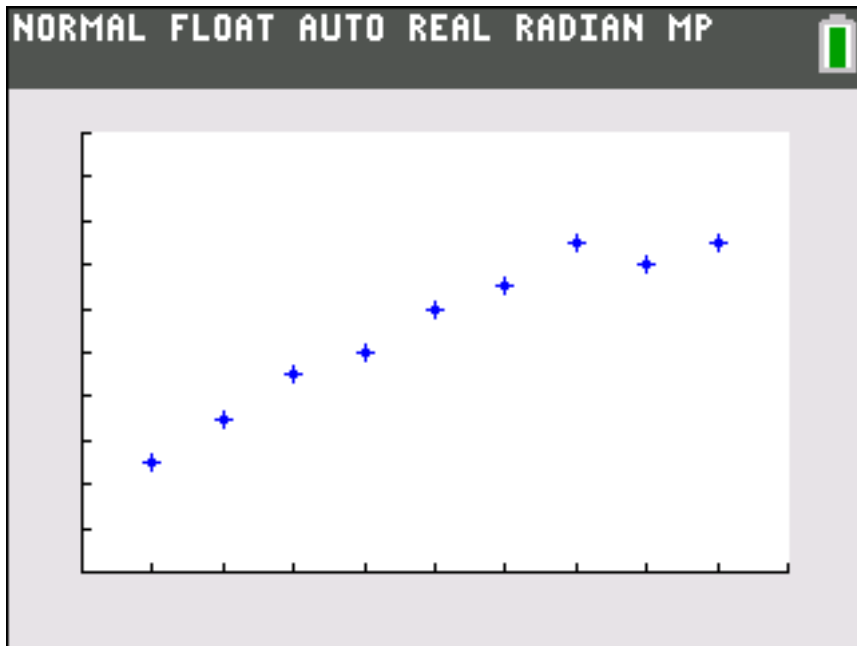
d) Zero Correlation



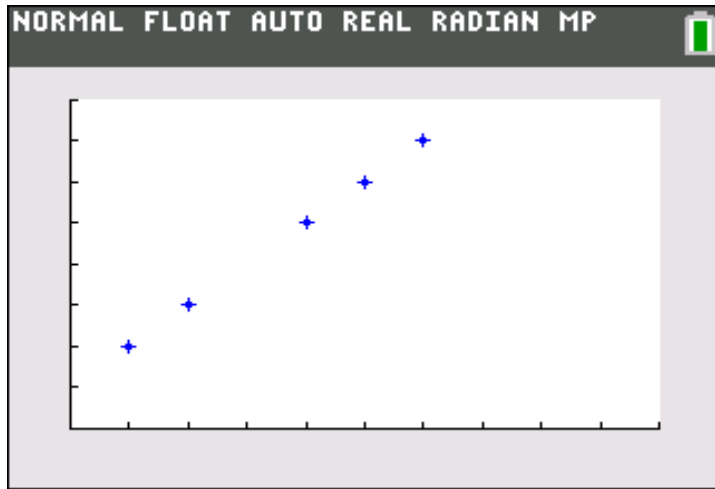
3. Weak Correlation



Strong Correlation

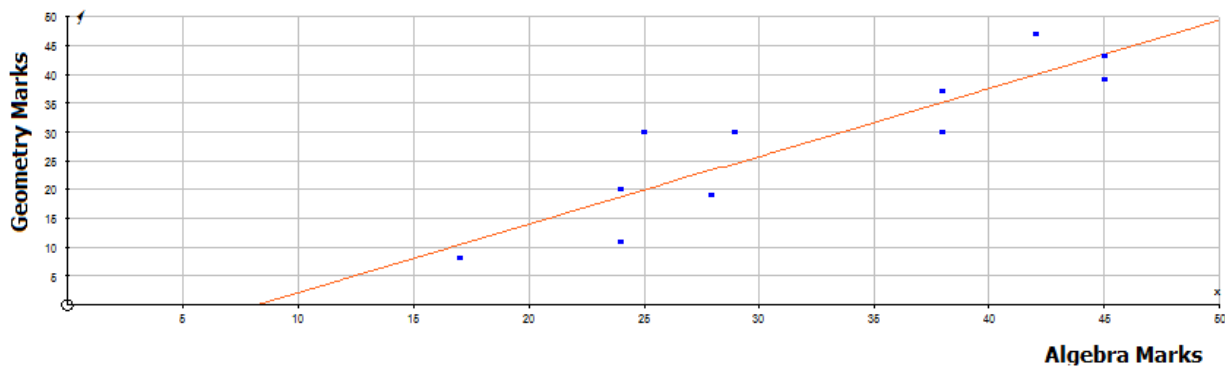


4. The correlation coefficient is a measure of the linear relationship between two variables.
5. a) This is an example of perfect correlation. The value of  $r$  is one.



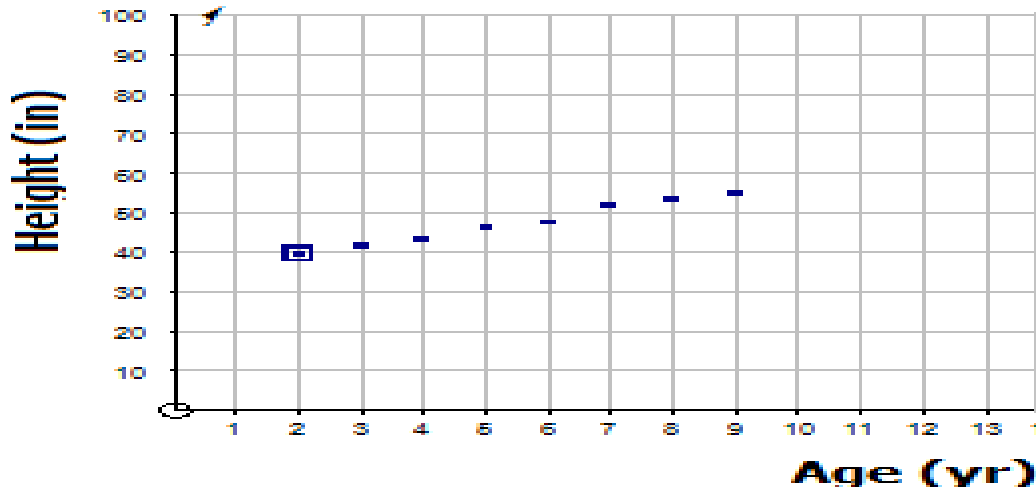
- b) The value of the Pearson correlation coefficient is 1.0.
6. a) The Pearson correlation coefficient is 0.568.
- b) The percentage of the variance  $r^2$  in the scores of Quiz 2 associated with the variance in the scores of Quiz 1 is 32.3%.
- c) The correlation coefficient of the two quizzes is positive but its value of 0.568 indicates that the linear relationship between the two is not a strong relationship. The value of  $r^2$  indicates that only 32.3% of the variance of Quiz 1 is shared with Quiz 2.
7. The three factors that affect the magnitude and accuracy are linearity, homogeneity of the group and the sample size.
8. a) A positive association
- b) No association
- c) A positive association
9. a) A scatterplot would be a good visual representation to determine what type of relationship exists between blood pressure and age.
- b) A scatterplot would not be a good visual representation because opinions are not quantitative data.
- c) A scatterplot would be a good visual representation to determine the strength and direction of the correlation.

10. The numbers -1.9 and 2.6 could not be values of the correlation coefficient since the values range from -1.0 to +1.0. A perfect correlation has a value of positive 1 or of negative 1.
11. The value -0.8 implies a stronger linear relationship. The value -0.8 is closer to -1.0 than 0.6 is to +1.0.
12. The correlation coefficient indicates a linear relationship between two variables. A correlation coefficient of zero means there is not a linear relationship. It does not mean that there is not a relationship between the variables. The relationship between the variables that resulted in a curvilinear relationship may be quadratic.
13. The figure insert is not displayed in the concept.
14. The correlation coefficient is 0.857
15. A scatterplot is a visual representation of two variables which are both quantitative. The scatterplot consists of plotted points that allow you to see the distribution of the data and to interpret the relationship between the variables.
16. a) The following scatterplot represents the marks of 11 students in an Algebra test and in a Geometry test. What mark would a student who made 37 in the Algebra test make in the Geometry test?



It was appropriate to insert a line of best fit for predicting.

- b) The following scatterplot represents the height of girls at ages 2, 3, 4, 5, 6, 7, 8 and 9 years of age. Describe the shape, trend and strength of the relationship.



No line of best fit will be inserted because the height of a girl does not depend upon her age.

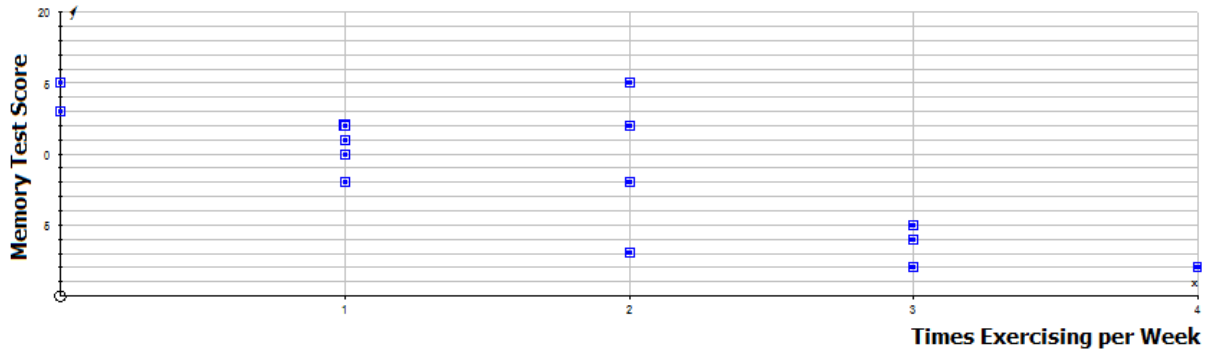
17. The correlation coefficient would not change if all the weights were converted to ounces.

18. a) The correlation should be described as being **STRONG** not as high.  
 b) A correlation value cannot be greater than one.  
 c) The value of the correlation coefficient does not have units. It is a value only.

## 9.2 Least-Squares Regression

### Answers

1. a)



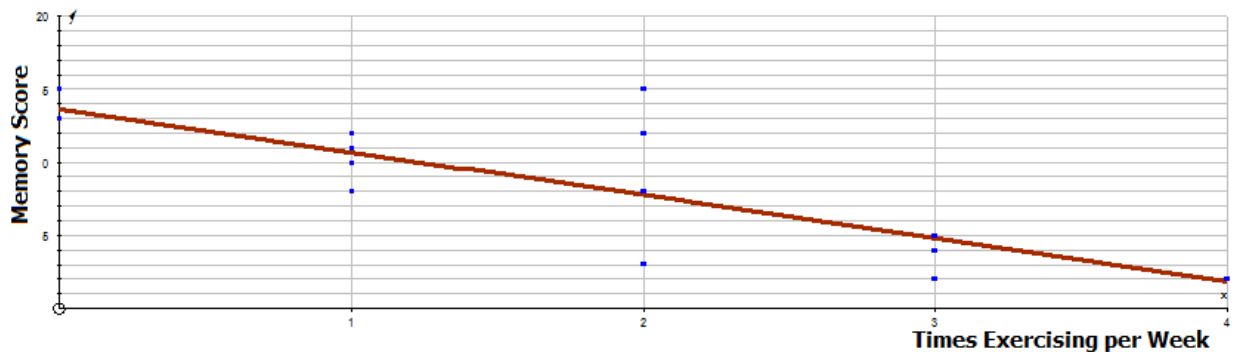
b) This appears to be a linear relationship as no curve or arc is visible.

c) The regression model to use is LinReg ( $a+bx$ ).

d) The regression coefficient is  $-2.95$ . This means that a change of 1 in the number times exercising per week is associated with a change of  $-2.95$  units of the memory score.

e) The regression equation is  $y = -2.95x + 13.65$

f)

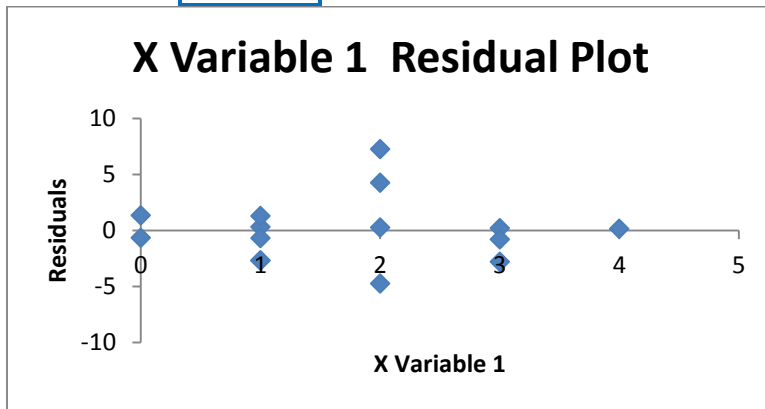


g) The predicted memory test score for a person who exercises three times a week is  $4.8 \approx 5$ .

h) No, a data transformation is not necessary because there is no visible curve in the scatterplot.

i)

Residuals
1.35
-4.75
4.25
0.3
0.2
-2.7
7.25
-0.65
-2.8
-0.8
0.15
-2.7
-0.7
1.3
0.25



j) A transformation of the data is not necessary because the residual values are not extremely far from zero. Therefore the linear regression equation is very suitable for this dataset.

2.
  - a) An estimate of the weight is 230 pounds.
  - b) The slope of the regression line is 7. For each inch of height increase the weight increases 7 pounds.
  - c) If the man is 70 inches tall his weight is 241 pounds.
  - d) The residual is – 48.



3. The units for are           . The units for are           . I am taking this to mean: The units for y are dollars. The units for x are years.
4. a) The slope of the regression line is 7/15 or 0.467  
 b) The estimated systolic blood pressure is 128.
5. The regression equation is  $y = 2x + 5$

X	Y	y
4	15	13
4	11	13
7	19	19
10	21	25
10	29	25

6. a) The average GPA would be 2.9 or approximately 3.  
 b) For a change of 1 in the GPA, the SAT score increases by a factor of 0.00365.  
 c) The estimated difference in GPAs is 0.2.  
 d) The SAT score of 200 would result in an average GPA of 1.3. If this were included in the dataset, the regression line equation would change slightly but not enough to alter any of the GPA estimations. Thus, the test score can be kept as part of the dataset.
7. The range of x-values to use is (2, 4, 6). The first three data points (2, 28); (4, 33) and (6, 45) result in a regression equation that gives the y value corresponding to the x value.
8. The slope of the line is 1.556.
9. The estimated height of the wife is 64 inches.
10. This means that a regression line is not the best fit for the dataset.
11. a) The regression equation is  $y = .052x + 287.12$  .  
 b) The length at 2 kg is 287.22 cm.  
     The length at 3 kg is 287.28 cm.  
     The length at 5 kg is 287.38 cm.  
     The length at 105 kg is 292.58 cm.  
 c) The length of the spring under no load is 287.12 cm.  
 d) The constant in Hooke’s law is the slope of the line.

12. The regression line  $y = 1 + 4x$  is better for this data since  $\sum(y - \hat{y})^2$  is 4.
13. a) The answers to this part will vary.  
b) The answers to this part will vary
14. a) A correlation of -0.45 is a moderate to weak negative correlation. Therefore the x-variable and the y-variable relate by a weak negative correlation.  
b) The slope is negative since the correlation is negative  
c) The slope is -0.20 when  $x = 45$  and  $y = -2.5$
15. a) True  
b) True  
c) False
16. a)  $\hat{Y} = 29.98 + 0.576x_1$   
b) 7.73  
c) The null hypothesis is that there is no correlation between the son's height and the dad's height. The slope is not zero as coefficient for dad/height is 0.57568.  
d) 30.56

### 9.3 Inferences about Regression

---

#### Answers

1. The predictor, or  $X$ , variable is the number of years of formal education. The reasoning behind this decision is that we are trying to determine and predict the financial benefits of further education (measured from the annual salary) by using the number of years of formal education
2. Yes. With an  $r$ -value of 0.67, these two variables appear to be moderately to strongly correlated. The nature of the relationship is a moderately strong, positive correlation.
3.  $y = 0.782x + 3.842$
4. \$14 399
5.
  - a)  $H_0 : \beta = 0, H_a : \beta \neq 0$
  - b) The critical values are set at  $t = \pm 1.98$ .
  - c)  $S_b = 0.08, t = 9.8$
  - d) Reject the null hypothesis since the test statistic value ( $t = 9.8$ ) is greater than the critical value of 1.98.
6. 16.8%
7. 4.65%
8.  $S_{\hat{y}} = 3.14$   
 $CI_{95} = (10.13, 22.57)$
9. Answers will be vary.
10. Reject the null hypothesis since the test statistic value ( $t = 9.8$ ) is greater than the critical value of 2.508.

## 9.4 Multiple Regression

---

### Answers

- Three; Test 1, Test 2, and Test 3
- The regression coefficient is 0.5560. This means that every 0.5560% change in Test 2 is associated with a 1.000% change in the final exam (assuming everything else is constant)
- $Y=0.0506X_1+-.5560X_2+0.2128X_3+10.7562$
- $R^2=0.4707$ . So 47.07% of the variance in the final exam can be accounted for by the combination of predictor variables.
- $F=13.621$  with  $p=0.000$ . Therefore the probability of observing an R value have occurred by chance is not significant is very small (close to 0.000)
- Test 2 since  $t = 3.885$  is greater than the critical value (because of the low p-value of 0.003)
- No because these two variables do not have test statistics that exceed the critical values.
- i) 3.8; ii) 1.6
  - 0.2
- 899, residual = 14
  - $F(2, 46) = 97.27$ ;  $p(T < 97.27) = 0.9999$
- $y=98.6 - 2.16X_1 + 0.134X_2$
  - 98.6452; initial or base line temperature.
  - As  $X_1$  increases, January temperature would decrease.
  - No.  $F(53, 3) = 0.1619$ ;  $p(T < 0.1619) = 0.5640$
- 

Set	Weight (kg)	Age	Years	Percent Life	
	$x_1$	$x_2$	$x_3$	$x_4$	$\hat{y}$
1	65	30	15	50	15.33
2	65	50	15	30	17.91
3	65	50	25	50	19.71
4	65	50	35	70	21.51

12.  $\hat{Y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$  ; response variable = college freshman year GPA, explanatory variables = total SAT scores ( $x_1$ ), high school GPA ( $x_2$ ), and high school rank ( $x_3$ ).

13. Answers will vary.

14. **There is no regression equation**

15. a)  $\hat{Y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

b)  $\beta_1$  is the correlation coefficient of  $X_1$  when the model fits a multiple regression linear equation;  $b_1$  is the beta coefficient for the least squares estimate.