**Scatterplots, Correlation**

|  |  |
| --- | --- |
| Scatterplots | Correlation |
| Displays a relationship between 2 **quantitative** variables measured on the same individuals  Each axis gets a variable | Changing units, Switching x and y axis does NOT affect the correlation  Outliers have a large affect on the correlation |

1. True or false. You could draw a scatter plot comparing the following variables:

a) person’s weight vs. person’s height

b) test scores vs. number of hours studied

c) favorite music type vs. hours of sleep on a given night

d) male height vs. female height in this course

e) the distribution of shoe size at this school

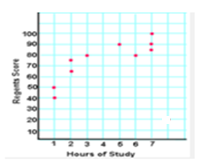
2. In part “a” in number 1 , would there be an explanatory and a response variable or is it just a relationship?

3. In part “b” in number 1, would there be an explanatory and a response variable or is it just a relationship?

4. Ben finds a positive correlation between the ml of soda his classmates drink per week and their calorie

intake for the week. If he changed the soda measurements to Liters, what would happen to the

correlation?

5. The following scatterplot shows test scores and hours of study for 11 students at a local middle school

1. Describe the overall pattern (direction, strength, form)
2. If a student studied for 8 hours, but failed the Regents test with a 40, what would happen to the correlation

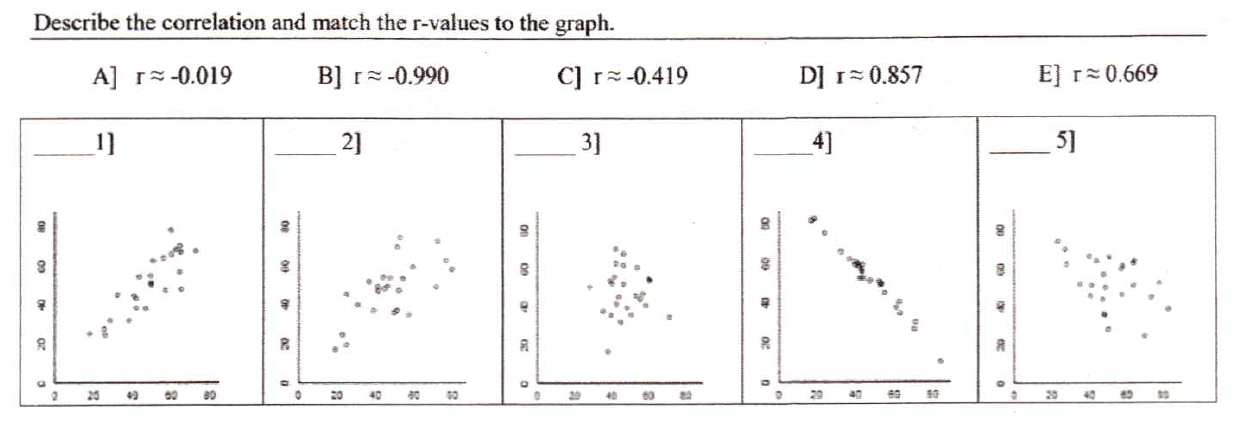
6. A scatterplot was created showing the heights and weights of students at Hillgrove. If we changed the

height from feet to inches and the weights from pounds to grams, which of the following would be affected:

a. Slope b. Y-intercept c. (r) correlation coefficient d. r2

7. A study showed that SAT Verbal scores were positively associated with the GPA’s of first year liberal arts

majors. What does this tell us?



8.

**Correlation/Regression (r and r2)**

|  |  |
| --- | --- |
| Correlation | Regression |
| r is always between -1 and 1  If r is closer to 0, the weaker the correlation  r2 tells what percentage of observed data is explained by the regression equation | Has a slope that has same sign as the correlation  Has a y-intercept that is not always statistically valuable  Can be used to predict values whenever the correlation is strong |

9. Sandy believes there is a link between how much TV a person watches (explanatory) and how much fat

food they eat (response). She gathers data on the number of times per week people eat fast food and the

number of hours of TV they watch per week. She finds out the following information:

a = 0.5, b = 2, r = 0.03 a. What is the regression equation?

1. What does the slope and y-intercept mean in this situation?
2. What does r tell you in this situation?
3. Should she use her regression to predict how much fast food people will eat in week based on the number of hours they watch TV?

10. Mult. Choice. The correlation between the height of fathers and heights of their (grownup) sons is r = 0.62.

This tells us that:

1. taller than average fathers tend to have taller than average sons
2. taller than average father tend to have shorter than average sons
3. sons are, on the average, taller than their fathers
4. 52% of all sons are taller than their fathers
5. There is no connection between heights of fathers and sons

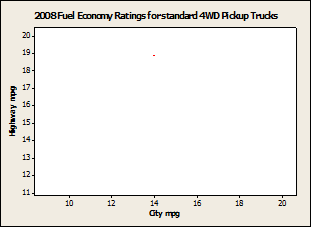
11. If r = 0.23, what is r2? What would r2 mean in this case? Is our regression equation a good predictor? Why

or why not?

12. The 2008 EPA fuel economy ratings for both highway and city driving are given for 10 randomly selected

standard pickup trucks with 4 wheel drive

|  |  |
| --- | --- |
| City | Hwy |
| 4 | 12 |
| 5 | 14 |
| 8 | 13 |
| 11 | 15 |
| 12 | 17 |
| 12.5 | 19 |
| 13 | 16 |
| 14 | 18 |
| 15 | 19 |
| 20 | 20 |

1. Draw the scatterplot
2. Find the regression equation, r, and r2
3. What does the slope tell you about the data?
4. Is the equation a good fit for the data? Why or Why not?
5. Predict how many Hwy mpgs for a truck that gets 10 mpg in the city.

**Residuals** (observed – predicted) – Review WS 4.5

13. Beatrice finds a regression equation of y =- -1.3x + 34. Her actual data value (observed) when x = 3 is 30.

What is the predicted data value when x = 3? What is the residual when x = 3? Does the equation

over/under predict the value for x = 3?

14. Can the frequency of cricket chirps be used to predict outdoor temperature?

|  |  |
| --- | --- |
| Chirps per minute | Outdoor temp (degrees F) |
| 20.0 | 88.6 |
| 16.0 | 71.6 |
| 19.8 | 93.3 |
| 18.4 | 84.3 |
| 17.1 | 80.6 |
| 15.5 | 75.2 |
| 14.7 | 69.7 |
| 17.1 | 82.0 |
| 15.4 | 69.4 |
| 16.2 | 83.3 |
| 15.0 | 79.6 |
| 17.2 | 82.6 |
| 16.0 | 80.6 |
| 17.0 | 83.5 |
| 14.4 | 76.3 |

1. What is the regression line for this data? Interpret the slope in

context of the problem.

1. Is the y-intercept meaningful in this situation? Why or why not?
2. What would you predict the outdoor temperature to be if a cricket

chirps 10 times per minute?

1. How many chirps would there be if it was 68 degrees outside?
2. Find r and r2 . Explain what this tells us.
3. Find the TOTAL residual value and determine if the regression

equation under/over predicts the outdoor temp.

1. Can we conclude that the higher the number of chirps per minute,

the higher the Full residual problem

**Lurking/Confounded Variables**

15. A recent study showed that long-distance runners tend to live longer.

a. Can we conclude that running long-distances increases lifespan? Why or why not?

b. What is a lurking variable in this situation?

c. If there are several lurking variables that cause long distance runners to live longer, what do we call

this situation?

16. There is a positive correlation between the numbers of times per week people go out to eat and the

number of times per month they go to the movies.

a. Can we say that the more people go out to eat the more they go to the movies?

b. What is a lurking variable in this situation?