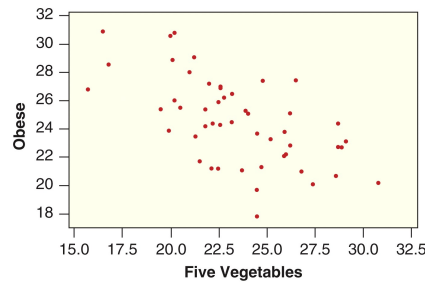


Name: White Solutions

1. The graph below plots the percent of obese people in each of the 50 US states against the percent of people in the state that consume at least 5 servings of fruit or vegetables per day.



- (a) (3 pts) Is the relationship described by the scatter plot for these two variables positive or negative? *Explain.*

The relation is **negative** because as the percentage of people consuming at least 5 servings of fruit/vegetables **increases**, the percentage of obese people **decreases** (so the cloud of points is trending down from left to right).

- (b) (3 pts) Which of $-1, -0.6, -0.02, 0.5, 0.98$ is most likely to be the correlation for the data above? *Explain briefly.*

The correlation is most likely to be -0.6 (among the numbers offered)... Because the correlation is negative (so 0.98 and 0.5 are out), and while the data doesn't lie on (or nearly lie on) a straight line, so -1 is out, there is a clear downward trend so -0.02 (which would imply essentially no linear trend) is also out.

2. In 2005, the *Educational Testing Service* observed the average Math SAT score for each of the 50 states and the percentage of high school seniors in the state who took the test. The correlation between these two variables was -0.84 .

- (a) (3 pts) **True** or **False**: test scores tend to be lower in states where a higher percentage of students take the test. *Explain your answer briefly.*

True. This is what negative correlation between means. If the correlation between x and y is negative, then y (average test score) will tend to decrease when x (percentage of population taking the test) increases.

Comment: The negative correlation makes sense. As the proportion of students taking the test increases, the wider the range of abilities is likely to be, so the lower the average score on the test is likely to be.

- (b) (3 pts) In Connecticut the average Math SAT score in 2005 was 517, and in Iowa it was 608. Do the data show that schools in Iowa do a better job teaching math than schools in Connecticut, or is there another possible explanation? *Justify your answer.*

While it is **possible** that math education is better in Iowa than in Connecticut, These data do **not** show that. It is possible for example that the percentage of students who take the test in Connecticut is higher than the percentage in Iowa, and it is this (possible) fact that accounts for the lower average in Connecticut.

Comment: You have to take all the data (and statistics) into account, not just the averages in Connecticut and Iowa.

3. In a study of the stability of IQ scores, a large group of individuals is tested at age 18 and again at age 45. The following results are obtained.

age 18: average score ≈ 100 , SD ≈ 15

age 45: average score ≈ 100 , SD ≈ 18 , $r \approx 0.8$

- (a) (3 pts) Estimate the average score at age 45 of all the people who scored 120 at age 18.

Using the regression equation, the estimated average score at age 45 of people who scored 120 at age 18 is

$$100 + 0.8 \cdot \frac{18}{15}(120 - 100) = 119.2.$$

(b) (3 pts) Jane scored 115 on the test at age 18. **True** or **False**: Assuming that the data was *homoskedastic*, Jane's score on the test at age 45 is likely to be in the *approximate* range 114 ± 11 . *Justify your answer.*

True: *The predicted average score for people who scored 115 at age 18 is*

$$100 + \frac{0.8 \cdot 18}{15}(115 - 100) = 114.4$$

and the R.M.S. error of regression in this case is

$$\sqrt{1 - 0.8^2} \times 18 = 10.8,$$

so the predicted score at 45 of an individual (Jane) who scored 115 at age 18 falls in the range

$$114.4 \pm 10.8 \approx 114 \pm 11.$$

(c) (2 pts) The researchers observed that individuals who scored high at age 18 scored lower at age 45 on average, while individuals who scored low at age 18 improved their score at age 45 on average. To explain this phenomenon, they suggested that people who did well on the test at 18 were less concerned with their performance the second time around, while people who did not score high at 18, took the test at 45 more seriously.

Do you agree? Explain briefly.

No. *This phenomenon is explained by the **regression effect**, not the motivation or lack thereof at age 45 of the subjects in the study.*