Regression Practice: Chapter 10

1. A very large, on-line statistics class had a midterm, of which the results were not very encouraging. The class averaged 65 with an SD of 12.
   So, the instructor kindly offered a retake of that midterm. Unfortunately, the results of the midterm were pretty much the same: they averaged 65 with an SD of 12.
   The correlation between the first test and retest was 0.8; the scatter diagram was football-shaped.

   (a) A few, very successful students, got 95 on the first test. Estimate the average scores of all those students on the retake.

   (b) On the retake, there were a few students who got a score of 89. For all of those students, estimate their average on the first test.

      i. The results of parts 1 (a) and 1 (b) seem incompatible. Explain that apparent contradiction.

      ii. It seems that those who did really well on the first test had a lower—but still good—score on the retest, but that those who did really well on the retest, on average, had not done so well on the first test. How is this possible? (It would seem, at first glance, that students in those elite groups tended to score lower on the other test. That's a difficulty: a student cannot score lower on *both* tests.)

2. True or false, and explain:

   (a) With a positive correlation coefficient, the slope of the SD line here ends up as +1.

   (b) Whenever the slope of the SD line is +1, the correlation must be perfect.

   (c) Since the correlation coefficient is positive and the scatter diagram is football-shaped, the regression line slopes upward and does not contain all the points of the scatter diagram.

   (d) The slope of the regression line is $m = 0.8$.

   (e) Because the correlation coefficient is positive, a student who did well on the first test will generally do somewhat better on the retake.

   (f) Because the correlation coeffficient is positive, the higher a student scores on the first test, the higher the prediction of that student's score on the retest.

   (g) A student who scored 15 points higher than his friend on the first test would be predicted to score 12 points higher on the retest.

Answers

1. (a) 89.

   (b) About 84.

      i. It's the regression fallacy.

      ii. Indeed the students who did very well on the first test scored somewhat lower on the retry. But, among all students scoring as well as that on the retry, only a small part of them were those who did excellently on the first test. So, as a whole, the students with good scores on the retry had done worse on the first test; the small part of them aforementioned is not *every student*.
      The key is the distinction between "on average," and "a student." Take all students who got 89 on the retake. They averaged 84 on the first test, but a student who got 95 on the first one was one of the 89's on the retest, and his first test was higher, going against the prediction. The reason: this student was not the typical 89 on the retake, he was a relatively rare case and an exception.

2. (a) True. It's (SD of y)/(SD of x) $= 12/12 = 1$, and $+1$ because r is positive.

   (b) False Whenever $r = +1$ or $r = -1$, the correlation will be perfect.

   (c) True. Positive r means the regression line is upward sloping; and football-shaped means not all points of the scatter diagram will lie on the line.

   (d) True. The slope m of the regression line equals r times the slope of the SD line, which was $+1$, giving $0.8 \times +1 = +0.8$.

   (e) False. The correlation is indeed positive, but regression is toward the mean (regression to mediocrity).

   (f) True. The line slopes up. A larger value of x will produce a larger value of y.

   (g) True. An increase of 15/12=1.25 SDs in x will be associated with an increase of $r \times 1.25 = 0.8 \times 1.25 = 1$ SD in y, which is 12 points.