R.m.s error of the Regression Line Practice

1. A statistical analysis was made of the midterm and final scores in large course, with the following results:

$$\text{average midterm score} \approx 50, \quad \text{SD} \approx 25$$
$$\text{average final score} \approx 55, \quad \text{SD} \approx 15, \quad r \approx 0.60$$

   The scatter diagram was football-shaped. For each student, the final score was predicted from the midterm using the regression line.

   (a) For about 32% of the students, the prediction for the final score was off by more than _____ points. Options: 6,9,12,15,25.

   (b) Predict the final score for a student whose midterm score was 80.

   (c) This prediction is likely to be off by _____ points or so. Options: 6,9,12,15,25.

2. Use the data in problem 1 to answer the following question.

   (a) About what percentage of students scored over 80 on the final?

   (b) Of the students who scored 80 on the midterm, about what percentage scored over 80 on the final?

   Explain your answers.

3. The freshmen in a large university are required to take a battery of aptitude tests. Students who score high on the mathematics text also tend to score high on the physics test. On both tests, the average score is 60, the SDs are the same too. The scatter diagram is football-shaped. Of the students who scored about 75 on the mathematics test:

   (i) just about half scored over 75 on the physics test.

   (ii) more than half scored over 75 on the physics test.

   (iii) less than half scored over 75 on the physics test.

   Choose one option and explain.

Answers follow on the next page.

Answers.

1. (a) r.m.s. error $= \sqrt{1 - r^2} \times$ SD of final scores $= 12$.

   (b) 65.8: this student is $30/25 = 1.2$ SDs above average on the midterm, and should be above average on the final by $r \times 1.2 = .72$ SDs; that is 10.8 points. (The SD on the final is 15 points.)

   (c) 12, the r.m.s. error: see part (a). It is OK to use the r.m.s. error inside a strip because the diagram is football-shaped.

2. (a)

$$\text{Percent} \approx \text{shaded area} \approx 5\%$$
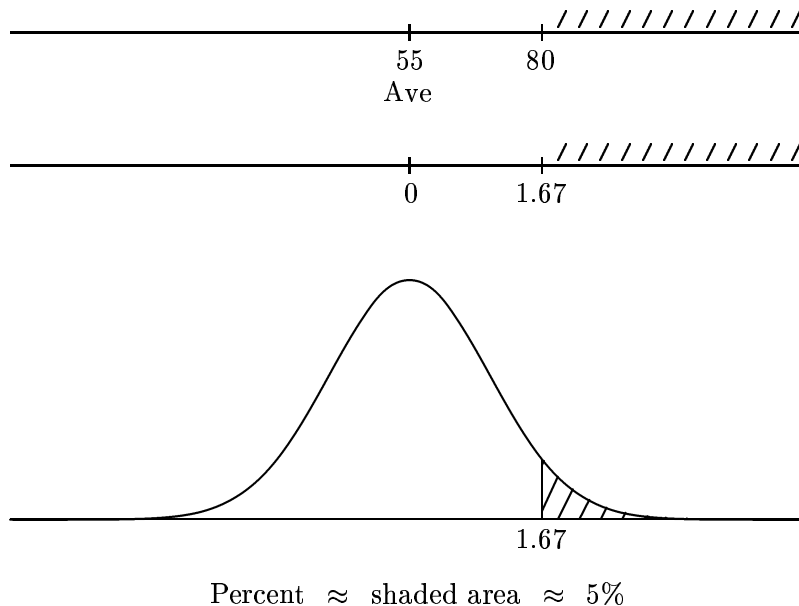
   (b) new average final score $= 55 + (1.2 \times 0.60 \times 15) = 65.8$ (since 80 on the midterm in standard units is $(80 - 50)/25$),

   new SD $= \sqrt{1 - 0.6^2} \times 15 = \sqrt{0.64} \times 15 = 0.8 \times 15 = 12$,

   $(80 - 65.8)/12 \approx 1.18$

$$\text{Percent} \approx \text{shaded area} \approx 12\%$$
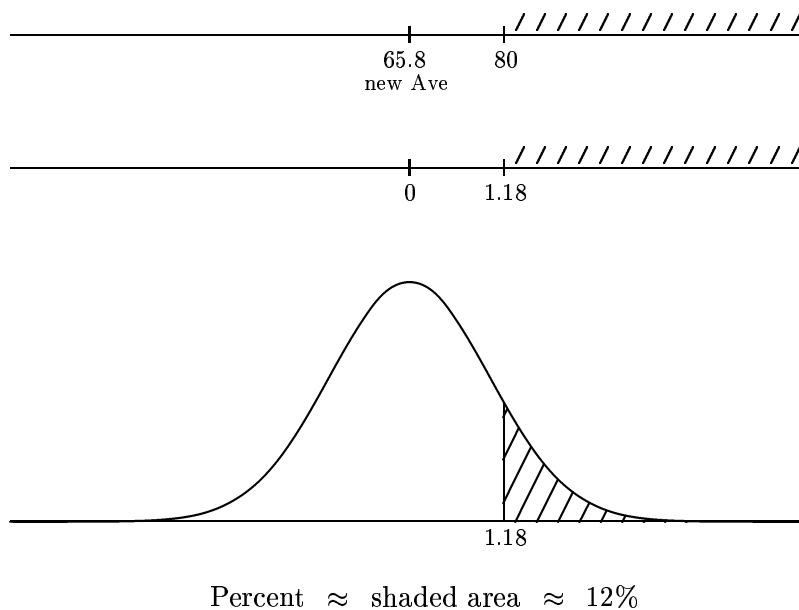
3. Option (iii) is right: regression effect.