Math 148 - Elementary Bio Stats

Solutions for Spring 2016 Lecture Quiz 02

This quiz is worth a total of 8 points.

1. The formula for the r.m.s. error of the regression line is

$$s_{est} \;:=\; \sqrt{\frac{1}{n}\sum_j (y_j - \hat{y}_j)^2} \;=\; \sqrt{1-r^2}\cdot \mathrm{SD}_y.$$

The factor $\sqrt{1-r^2}$ is between 0 and 1. This means that s_{est} will never exceed SD_y. Explain why.

Hint: Interpret both SD_y and s_{est} as spreads about estimates for y.

Solution to problem 1:

SD_y is the spread of the scatterdiagram about the horizontal line $y = \bar{y}$ whose (constant) y-value $\bar{y}$ is the best possible estimate for y if you use no information about x.

s_{est} is the spread of the scatterdiagram about the regression line whose y-values are, for $x \approx x_0$, the best possible estimate $\bar{y}|_{x=x_0}$ for y if the additional information that $x = x_0$ is used.

The scatterdiagram is clustered more tightly about the better estimates $\bar{y}|_{x=x_0}$ than about the "one shoe fits all" estimate $\bar{y}$ for its y-values. ■.