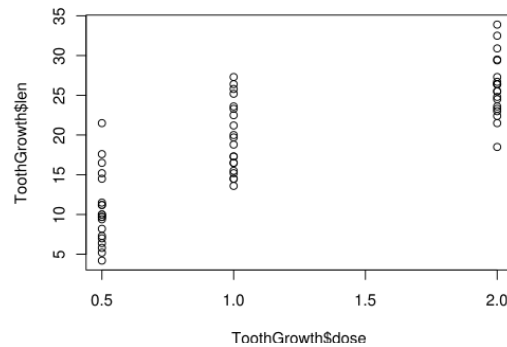




Quiz: Regression

1. The length of odontoblasts (tooth cells) in each of 10 guinea pigs at each of three dose levels of vitamin C (0.5, 1, and 2 mg) is measured. The scatter plot of the data is shown below.



The equation for the regression line computed from the data is

$$y \approx 10x + 7$$

where x is dose level in mg and y is odontoblast length. The correlation coefficient for the data is $r \approx 0.8$.

- a) Sketch the regression line in the above plot.

The line contains these points:

x	0.5	1.0	1.5	2.0
y	12	17	22	27

- b) Estimate the odontoblast length if the dose level is 1.5 mg.

If $x = 1.5$, then $y = 22$ is the estimate (the point on the line).

- c) The RMS error for the regression line is 4.5. Among guinea pigs that receive a vitamin C dose level of 2 mg, 68% have odontoblast length in what range?

If $x = 2$, then $y = 27$ is the prediction. Since we want a range for 68% of the points (with $x = 2$), we use

$$27 \pm 1 \times \text{RMS} = 27 \pm 4.5.$$

- d) Among guinea pigs that receive a vitamin C dose level of 1.5 mg, 95% have odontoblast length in what range?

If $x = 1.5$, then $y = 22$. Since we want a range for 95% of the points (with $x = 1.5$), we use

$$22 \pm 2 \times \text{RMS} = 27 \pm 4.5 = 22 \pm 9.$$

- e) One measurement has dose = 1.0 mg and length = 14.0. Calculate the length predicted by the regression line then calculate the residual error for that value: $y - (\text{predicted value})$.

The predicted value is 17 if $x = 1$. So, the error is $14 - 17 = -3$.

2. For men age 25–34, the relationship between education (years of schooling completed) and systolic blood pressure can be summarized as follows.

average education ≈ 13 years,	sd ≈ 3 years
average blood pressure ≈ 119 mm	sd ≈ 12 mm

The correlation coefficient is $r \approx -0.1$.

a) Sketch the regression line and write down its equation.

The equation is

$$y - 119 = -0.1 \frac{12}{3} (x - 13)$$
$$y - 119 = -0.4 (x - 13)$$

This is a line that contains the point $(13, 119)$ and has slope -0.4 .

b) Predict the blood pressure of a man with 20 years of education.

If $x = 20$, then $y = 119 - 0.4(20 - 13) = 119 - 0.4 \times 7 = 119 - 2.8 = 116.2$. Our prediction is the point on the line.

c) Estimate the average blood pressure among all subjects with 20 years of education.

Numerically, our answer is 116.2 again. The reason is that the *average* of the y -values for dots with $x = 20$ is the point on the line (which we calculated in b).

d) One subject had 20 years of education, and his blood pressure was 118 mm. True or false and explain: compared to other men at his educational level, his blood pressure was a bit on the high side.

His z score is $z = (118 - 116.2)/\text{RMS}$ since RMS is standard deviations for the normal curves that move along the regression line. The value is:

$$\text{RMS} = \text{sd}_y \sqrt{1 - r^2} = 12 \sqrt{1 - (-0.1)^2} \approx 12.$$

So, the subject's z score is $z \approx (118 - 116.2)/12 = 1.8/12 = 0.15$. That is close to the center of the normal curve. So, his blood pressure is about average.

e) Suggest reasons for the sign (± 1) and magnitude (0.1) of r .

The correlation is negative. That means subjects with more education had lower blood pressure. Perhaps that is because they have better jobs and, consequently, better health care.

The correlation is close to zero, however. That means education level is not a good predictor of blood pressure. Diet and exercise would have much stronger correlations.