1. I'm teaching my nephew to count by 2's. He's just little, so he hasn't gotten any farther than 2, 4, 6. Compute the mean, variance, and standard deviation of these three numbers.

The mean is $\frac{2+4+6}{3} = 4$. The variance is $\frac{(2-4)^2+(4-4)^2+(6-4)^2}{3-1} = 4$. The standard deviation is the square root of the variance, or 2.

2. Different types of writing can sometimes be distinguished by the lengths of the words used. A student interested in this fact wants to study the lengths of words used by Tom Clancy in his novels. She opens a Clancy novel at random and records the lengths of each of the first 250 words on the page. What is the population in this study? What is the sample? What is the variable measured?

(This is problem 3.37 from the textbook.) The population is all of the words in all of Clancy's novels. The sample is the first 250 words on the selected page. The variable measured is word length.

3. Peter Porker's Pig Emporium sells hams (mmmm, ham) by mail order. The hams vary in size, with a mean weight of 6 pounds and standard deviation of 0.65 pounds.

(a) If these weights were expressed in kilograms (1 pound = .45 kg), what would the mean and standard deviation be?

(b) When the emporium ships these hams, the box and packing materials add 0.8 kilograms. What are the mean and standard deviation of the weights of the shipped boxes, expressed in kilograms?

a) Both the mean and s.d. are multiplied by .45, so the mean in kg is $.45 \cdot 6 = 2.7$ and the s.d. in kg is $.45 \cdot .65 = .2925$.

b) Adding 0.8 kg increases the mean by 0.8, to $2.7 + 0.8 = 3.5$. The s.d. doesn't change from .2925(moving all the data up 0.8 doesn't affect the spread).

4. Suppose that the diameters of the trees in a Minnesota forest are approximately normally distributed, with mean 10.4 inches and standard deviation 4.7 inches.

(a) About what fraction of the trees are less than 10.4 inches in diameter?

(b) About what fraction of the trees are greater than 5 inches in diameter?

(c) About what fraction of the trees are between 5 and 10.4 inches in diameter?

(d) Paul Bunyan chopped down 2667 trees this morning before breakfast. Assuming that he chose the trees at random, about how many of these would you expect to be between 5 and 10.4 inches in diameter?

a) 10.4 is the mean. Since the the normal distribution is symmetric, half the data are below the mean. (Another way of saying the same thing is that 10.4 is 0 s.d.'s from the mean (the Z-score is $\frac{10.4-10.4}{4.7} = 0$), and Table A tells us that the proportion of the data to the left of 0 is 0.5.)

b) 5 is 5.4 below the mean, which is 1.15 s.d.'s (in other words, the Z-score is $\frac{5-10.4}{4.7} = -1.15$). By Table A, .1251 is the fraction of the data less than 5, so $1 - .1251 = .8749$ is the fraction greater than 5.

c) This is the fraction less than 10.4 inches, minus the fraction less than 5 inches, or $.5 - .1251 = .3749$.

d) $.3749 \cdot 2667 = 999.9$, so we expect about 1000 of them to be between 5 and 10.4 inches.

5. Are good grades in high school associated with family togetherness? A simple random sample of 142 high-school students was asked how many meals per week (x) their families ate together. Their responses produced a mean of $\bar{x} = 3.78$ meals per week, with a standard deviation of

$s_x = 2.2$. Researchers then matched these responses against the students' grade point averages y, which had mean $\bar{y} = 3.15$ and standard deviation $s_y = 0.48$. The scatterplot appeared to be reasonably linear, so they created a regression line. The correlation r was 0.5, and no apparent pattern emerged in the residuals plot.

(a) Compute the equation of the regression line $\hat{y} = a + bx$.

(b) Interpret the slope b in this context.

(c) Little Polly Perfect eats three meals a day (so 21 meals a week) with her family. What does our regression line predict her GPA will be? Keeping in mind that GPA is on a scale of 0 to 4, how do you explain this answer?

(d) (EXTRA CREDIT) Upon hearing of this study, a counselor recommended that parents who want to improve the grades their children earn should get the family to eat together more often. Do you agree with this interpretation? Explain.

a) The slope b is $\mathbf{r \cdot \frac{s_y}{s_x} = 0.5 \cdot \frac{0.48}{2.2} = 0.11}$. The intercept a is $\mathbf{\bar{y} - b\bar{x} = 3.15 - 0.11 \cdot 3.78 = 2.73}$. So the regression line is $\mathbf{\hat{y} = 2.73 + 0.11x}$.

b) For every extra meal per week that a student eats with the family, the student's predicted GPA will rise 0.11 points.

c) Plug $\mathbf{x = 21}$ into the regression equation: $\mathbf{\hat{y} = 2.73 + 0.11 \cdot 21 = 5.04}$. What has probably happened is that the number of meals that Polly eats per week with her family is so far outside the usual range of the data that the regression line can't be used to extrapolate her GPA.

d) Disagree, because association isn't causation! Could there be lurking variables?