For each statement, circle True or False. If False, explain why or make a small change to correct it.

1. True or False:

   If the correlation coefficient r is equal to 0.99, it follows that a linear fit will be much better than a nonlinear fit.

Solution: False. The correlation coefficient measures strength of a linear fit, but does not distinguish between linear and nonlinear fits. See class notes for example of data with r close to one, but a clear nonlinear relationship between x and y.

2. True or False:

   The least squares regression line minimizes the sum of the residuals.

Solution: False. Minimizes the sum of the squared residuals, not the sum of the residuals.

3. True or False:

   If the points in a scatter-plot of two quantitative variables lie exactly on the line $y = 3 - 2x$, then the correlation coefficient r must be exactly 1.

Solution: False. If the points are on a straight line, then $r = 1$ or $r = -1$. (Actually there are two other possibilities. If the points are on a line with slope 0, then $r = 0$. If the points are on a vertical line, then r is undefined.)

The slope here is negative, so $r = -1$.

4. True or False:

   If a plot of residuals versus fitted values shows a wedge pattern of small residuals for small fitted values and large residuals for large fitted values, this indicates a good fit that meets the model assumptions.

Solution: False. This wedge pattern of residuals indicates that the variance depends on x, contrary to a model assumption. There is heteroskedasticity in the data.

5. True or False:

   If the regression line for y on x is $\hat{y} = b_0 + b_1 x$, we can find the regression line for x on y simply by solving the previous equation for x.

Solution: False. Regression of y on x is different from regression of x on y. Thinking geometrically, the first minimizes the sum of squared vertical distances between points and the line while the second minimizes the sum of squared horizontal distances. We can also verify this algebraically.

$$y = b_0 + b_1 x \qquad \text{implies that} \qquad x = -\frac{b_0}{b_1} + \frac{1}{b_1} \times y$$

The slope of this line is

$$\begin{aligned} \frac{1}{b_1} &= \frac{1}{r\frac{s_y}{s_x}} \\ &= \frac{1}{r} \times \frac{s_x}{s_y} \end{aligned}$$

But the slope of the regression of x on y would be $r \times s_x/s_y$. Both lines go through $(\bar{x}, \bar{y})$, but they are different lines as their slopes differ.x

Statements 6–10 refer to the following situation. The correlation coefficient of the adult heights (measured in inches) of father's heights and son's heights is 0.5.

For each statement, circle True or False. If False, explain why or make a small change to correct it.

6. True or False:

The correlation coefficient would be 2.54 times as large if the heights had been measured in centimeters. (There are 2.54 cm per inch.)

Solution: False. The correlation coefficient is unitless. Linear unit changes do not change r.

7. True or False:

Heights of fathers and heights of sons are positively associated.

Solution: True. The sign of the correlation coefficient measures the direction of association.

8. True or False:

If a father's height is one standard deviation of the mean, we predict his son's height to also be one standard deviation above the mean.

Solution: False. If the father's height is z standard deviations above the mean, we predict the son's height to be rz standard deviations above the mean. Here, this is only 0.5 standard deviations.

9. True or False:

A nonlinear relationship will be better than a linear relationship because the linear relationship is fairly weak.

Solution: False. Weak linear relationships are not necessarily improved by a nonlinear fit. Data can be scattered widely around a straight line. Sometimes weka linear fits are adequate.

10. True or False:

If the standard deviations of heights of fathers and of sons are each four inches, the slope of a regression line for predicting a son's height from a father's height will be 0.5.

Solution: True. The slope of the regression line is $r \times s_y/s_x$, or $0.5 \times 4/4 = 0.5$.