Assignment 2 Solutions

1. Correlation is only defined for two quantitative variables. Since eye colour and hair colour are categorical variables, this is not a valid value of r.

2. As the age of a university student increases, the amount of time until graduation tends to decrease, so these two variables are **negatively** associated. The sign of the correlation coefficient should be **negative**. So, $r = 0.75$ is not a reasonable value of r.

3. We typically do not have a perfect (positive or negative) relationship between two variables – that is, a correlation of $+1$ or -1. However, in this case, it is true that **every time** the number of minutes watched of a movie increases by a certain number of minutes, the number of minutes remaining will decrease by the same number of minutes. So $r = -1$ is a reasonable value of r.

4. As the age of a vehicle increases, the number of miles driven on the vehicle will also tend to increase, so a positive correlation is reasonable. Furthermore, there is likely a fairly strong relationship between these two variables, so a magnitude of 0.75 is also reasonable.

5. As the number of classes missed increases, exam score will tend to decrease, so a negative relationship between the two variables seems reasonable. However, it is not reasonable that **every time** number of classes missed increases by a certain amount, exam score will decrease by the same fixed amount. Therefore, $r = -1$ is not a reasonable value of r.

6. Correlation is unaffected by a change in units of either X or Y, so the correlation would still be 0.54.

7. We first identify the explanatory variable X as height of Christmas tree and the response variable Y as number of presents. We are given $\overline{x} = 6.3$, $s_x = 1.2$, $\overline{y} = 5.8$, and $s_y = 3.2$. We are also told that the slope $b_1 = 1.3$. Since $b_1 = r\dfrac{s_y}{s_x}$, we can solve for r by plugging in the values of b_1, s_y, and s_x:

$$\begin{aligned} 1.3 &= r\frac{3.2}{1.2} \\ \Rightarrow r &= 0.4875 \end{aligned}$$

8. Since amount of melatonin taken is being used to attempt to explain the variation in amount of time to fall asleep, the explanatory variable is "melatonin dose" and the response variable is "minutes to fall asleep".

9. We are given that 82.19% of the variation in time to fall asleep is explained by its regression on melatonin dose. So $r^2 = 0.8219$. The correlation between melatonin dose and time to fall asleep is

$$r = -\sqrt{r^2} = -\sqrt{0.8219} = -0.91$$

Note that we take the negative square root of r^2, as the relationship between melatonin dose and time to fall asleep is negative. The sign of the correlation must be the same as the sign of the slope of the least-squares regression line (the slope is -0.986).

10. The residual is

$$y - \hat{y} = 15 - (19.059 - 0.986(5)) = 15 - 14.129 = 0.871$$

The positive residual tells us that the point for this day falls above the least-squares regression line, i.e., the time to fall asleep for this day is higher than we would have predicted it to be, based on the melatonin dose.

11. In general, the correct interpretation for the slope is "for every increase in 1 unit of X, we predict that Y will increase/decrease by the value of b_1". Since the explanatory variable X is "melatonin dose", the response variable Y is "minutes to fall asleep", and the slope $b_1 = -0.986$, the correct interpretation for the slope is "for every increase in 1 mg of melatonin, we predict it will take 0.986 minutes less minutes to fall asleep."

12. Statement (A) states the slope lies between -1 and 1, inclusive. In fact, the correlation lies between -1 and 1, so statement (A) is false.

Statement (B) states that the larger the value of the slope, the stronger the linear relationship between the variables. This is not true, as it possible to have a large slope for the regression line, but the points don't fall very close to the line (indicating a weak relationship). Similarly, it is possible to have a low value of the slope, but the points do fall very close to the line (indicating a strong relationship). So statement (B) is false.

Statement (C) states that the slope always has the same sign as the correlation. This is true.

Statement (D) states that the square of the slope is equal to the fraction of variation in Y that that is explained by its regression on X. In fact, this is true about the square of the **correlation**.