Stat 571 - HW12 - Nathanael Fillmore

1. (a) True, $-1.3 = b = rs_x/s_y$ and s_x, s_y are positive.
   (b) False, he must have made a mistake, since $r \leq 1$.
   (c) False, it is evidence of nonlinearity but not necessarily of nonconstant variance.
2. (a) Constant variance. (Size of residuals changes depending on X.)
   (b) Linearity. (Up-down pattern to the residuals.)
   (c) Normality. (Too few residuals near 0.)
   (d) Normality. (Outlier.)
3. (a) I.e., in the best fit of $Y = a + bX$, what is the coefficient b? Recall (regression slide 10)

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Here $\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = 0.2535$ and $\sum_{i=1}^n (X_i - \bar{X})^2 = 100.210$ so $b = 0.2535/100.210 = 0.002529688$. Recall (p. 495) that

$$SE = \sqrt{\frac{MS_{residual}}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

$$MS_{residual} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 + b \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-2}$$

So here

$$MS_{residual} = \frac{0.001024 + 0.002529688 \cdot 0.2535}{21-2} = 8.76461e-05$$

$$SE = \sqrt{\frac{8.76461e-05}{100.210}} = 0.0009352135$$

   (b) I.e., $H_0 : b = 0$. From page 497, $t = \frac{b - \beta_0}{SE_b}$ follows a t distribution with $n-2$ degrees of freedom. Here $\beta_0 = 0$, and

$$t = \frac{0.002529688 - 0}{0.0009352135} = 2.704931$$

Note

```
> pt(2.704931, df=21-2, lower.tail=FALSE)
[1] 0.007019554
```

So the hypothesis $H_0 : b = 0$ has strong evidence against it (two-sided t-test, $p = 0.007019554, t = 2.704931, df = 19$).

   (c) Recall (p. 496) that the CI is $b - t_{\alpha(2),df} SE_b < \beta < b + t_{\alpha(2),df} SE_b$. Here, $t_{97.5\%,19}$ is:

```
> qt(0.975, 19)
[1] 2.093024
```

so the CI is:

$$0.0005722637 = 0.002529688 - 2.093024 \cdot 0.0009352135 < \beta < 0.002529688 + 2.093024 \cdot 0.0009352135 = 0.004487112$$

   (d) Assumptions are that the deviations of relative growth rate from $a + b$(mean fleck duration) are independent and normally distributed with mean 0 and a common standard deviation sigma.

(e) Plot residuals and look for patterns as found in Question 2 above.

4. (a) The regression equation is $Y = a + bX$. Here $a = 16.7693, b = 1.2010$. So Feeding $= 16.7693 + 1.2010 *$NonFeeding $= 16.7693 + 1.2010 * 80.0 = 112.8493$.

   (b) Recall $b - tSE < \beta < b + tSE$. Here $SE = 0.1124$, $t = 2.306004$ (0.975 quantile of t distribution with 10-2 df), and $b = 1.2010$, so the CI is:

   $$0.9418052 = 1.2010 - 2.306004 \cdot 0.1124 < \beta < 1.2010 + 2.306004 \cdot 0.1124 = 1.460195$$

   (c) The prediction interval will be wider, because it also accounts for the sampling variance.

   (d) (1), because the prediction intervals are tightest near the mean, because if X is near the mean $\bar{X}$, then $X - \bar{X}$ is small.

   (e) The correlation is $sign \cdot \sqrt{R^2} = sign \cdot 0.9666954$. We know that *sign* is positive because the b coefficient is 1.2010, which is positive. So the correlation is 0.9666954.

   (f) False, even if the correlation coefficient is cloase to one, a nonlinear model could fit better.