MA 542: Homework 3 Solutions

1. (a) The 99% PI for Y_{new} when $X_{new} = 2$ is (12.75,23.65). Interpretation: With confidence coefficient .99, we predict that the number of ampules found to be broken for the next shipment with two times transfers will be somewhere between 12.75 and 23.65.

   (b) The 99% confidence band boundary when $X_{new} = 2$ is (15.44,20.96) and when $X_{new} = 4$ is (20.03,32.37). Apparently, they are wider than the corresponding CIs in part (a). It should be. In order to maintain an overall confidence level, confidence band would be more conservative than confidence intervals.

2. (a) The 98% PI for Y_{new} when $X_{new} = 30$ is (220.869,238.393).

   (b) The 98% confidence band boundary when $X_{new} = 30$ is (226.949,232.313). It is wider than the interval in part (a), as it should be, since it is a simultaneous boundary for all X_h.

3. (a) Set up the ANOVA table.

```
> anova(SLR)
Analysis of Variance Table

Response: Y
           Df Sum Sq Mean Sq F value   Pr(>F)
X           1  3.588  3.5878  9.2402 0.002917 **
Residuals 118 45.818  0.3883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

   (b) MSR estimates $\sigma^2 + \beta_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2$, MSE estimates σ^2. When $\beta_1 = 0$, MSR and MSE estiamte the same quantity σ^2.

   (c) We test $H_0 : \beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. We reject H_0 in favor of H_a if the p-value is less than 0.01. Since the p-value is $0.002917 < 0.01$, we reject H_0 in favor of H_a and conclude $\beta_1 \neq 0$.

   (d) The absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model is the sum of squares of regression or SSR = 3.587846. The relative reduction in the variation of Y when X is introduced into the regression model to the variation of Y is 0.07262044. This latter measure is known as the coefficient of determination.

   (e) The correlation coefficient between Y (grade point average) and X (ACT test score) is 0.2694818.

   (f) R^2 gives us the fraction of the total variability in Y that is explained by the regression model or X. On the other hand, r provides the strength of linear association between X and Y. Here, R^2 provides the more clear-cut interpretation.

4. (a) Set up the ANOVA table.

```
> anova(SLR)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X          1  160.0   160.0  72.727 2.749e-05 ***
Residuals  8   17.6     2.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

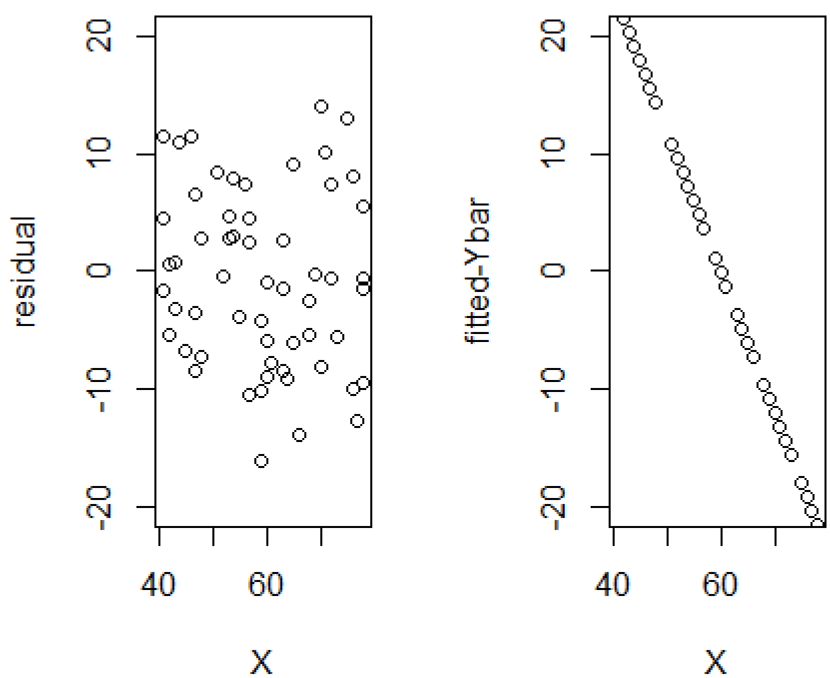
Sum of Square and degree of freedom are additive.

(b) We test $H_0 : \beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. We reject H_0 in favor of H_a if the p-value is less than 0.05. Since the p-value is $2.749 \times 10^{-5} < 0.05$, we reject H_0 in favor of H_a and conclude there is a linear association between X and Y.

(c) $t^* = 8.528, F^* = 72.727$. Hence, $\sqrt{F^*} = t^*$.

(d) $R^2 = SSR/SSTO = 0.9009$, $r = cor(X, Y) = 0.949158$. 90.09% of the variation in Y is accounted for by introducing X into the regression model.

5. (a) Two graphs:



By the graphs, SSR appears to be the larger component of SSTO. Thus, $R^2 > 1/2$.

(b) Set up the ANOVA table.

```
> anova(SLR)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X          1 11627.5 11627.5  174.06 < 2.2e-16 ***
Residuals 58  3874.4    66.8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(c) We test $H_0 : \beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. We reject H_0 in favor of H_a if the p-value is less than 0.05. Since the p-value is $2.2 \times 10^{-16} < 0.05$, we reject H_0 in favor of H_a and conclude $\beta_1 \neq 0$.

(d) $1 - R^2 = SSE/SSTO = 0.2499$, which is relatively small.

(e) $R^2 = 1 - SSE/SSTO = 0.7501$, $r = cor(X, Y) = -0.8661$.

6. $\sigma^2\{\hat{Y}_h\} = \sigma^2(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}) \to 0$ if $X_h = \bar{X}$. $\sigma^2\{pred\} = \sigma^2(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}) \geq \sigma^2 > 0$.