Test Corrections

Problem 2

b) $\widehat{biomass} = \hat{\beta}_0$
c) $\widehat{biomass} = \hat{\beta}_0 + \hat{\beta}_1 pH + \hat{\beta}_2 K$

Problem 3

b) Observation 65
c) Observation 9
d) Observation 53

Problem 4

a) Observation 66
b) Observation 65
d) Observation 53

Problem 6b)

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_A: at least one $\beta_i \neq 0$; where $i = 1$ or 2 or 3

P-value $= 1.669 \times 10^{-12}$

Null hypothesis:
Ineffective model, none of the predictors has any linear relationship with the response variable

Alternative hypothesis:
one or more of the predictors is effective (linear relationship) in the model

since the p-value, $1.669 \times 10^{-12} < \alpha = 0.05$, we reject the null hypothesis and conclude that there is indeed linear association for at least one predictor and response.

7a)

$H_0: \beta_2 = \beta_3 = 0$

H_A: at least one $\beta_i \neq 0$, $i = 2$ and/or 3

Null hypothesis:
Ineffective model, none of the categorical predictors has any linear relationship with the response variable

Alternative hypothesis:
one or more of the categorical predictors is effective (linear relationship) in the model

7b)

From problem 6.b, the reduced model would be

$$mpg = \beta_0 + \epsilon$$

From problem 7a, the reduced model would be

$$mpg = \beta_0 + \beta_1 \text{ weight} + \epsilon$$

we need to assess whether assumptions were met for the model $mpg = \beta_0 + \beta_1$ weight (Model 1)

In the studentized Residuals vs. Fitted Values, we noticed that linearity may not be met because it seems like there is more points above the ∅ line. It also doesn't seem to meet variance because it has a megaphone shape.

In the normal quantile plot, we notice the points near the tails of the line are far from the line, drifting away. So normality isn't met either.

Because these main assumptions were not met, the tests in problem 7a is not reliable.

The reduced model from 6b is

$$mpg = \beta_0 + \epsilon$$

where $\hat{\beta}_0 = \bar{y}$

so mpg is a function of the overall mean
sin $\hat{\beta}_0 = \bar{y}$, the variability is accounted for so it is more
reliable than 7a.

we know that the error sum of squares for $mpg = \beta_0 + \epsilon$

is $\sum(\text{observed} - \text{fitted})^2 = \sum(y_i - \bar{y})^2$

The variability is small (reduce)