Test 2 Corrections

1. a- This is correct because in the MLR equation for biomass, pH has an estimate of 414.9021 while K has an estimate of -1.0095. Since for every increase in pH, biomass changes more than for every increase in K, this answer is correct.

b- This isn't correct because it's the opposite of a. K doesn't have a bigger effect on biomass when compared to pH. So, this answer isn't correct

c- This is not correct since there's nothing that really shows that pH's effect on biomass depends on K. pH has a very low p-value and so does K, which means that most likely they don't have dependency on each other or are kind of related to each other. They're also not an interaction term and can act independently when one is constant.

d- This is correct since Snows Marsh has a positive estimate. If the location isn't Smith Island or Snows Marsh, both Smith Island and Snows Marsh get a value of zero. This also means that Oak Island is the location in that situation. Since Oak Island isn't in the MLR equation, that most likely means it has an estimate of zero. Considering this, since Snows Marsh has an estimate of 58.1814, that should mean Snows Marsh tends to have higher grass than Oak Island.

3. a - Observation 66 is likely not most unusual since it's within the boundaries of the leverage lines on the graph. Leverage shows how far away independent variable (predictor) values are from all of the other independent variable values. pH and K are both predictors, so leverage is what you should use to determine if a point is unusual in this case. Observation 66 doesn't have a particularly high leverage, so it's most likely not unusual in this case.

b - Observation 65 is likely most unusual since it's beyond the more solid line for leverage. Observation 65 has an extremely high leverage point. Like I said above, leverage is used to determine if predictors are unusual. That's why observation 65 is likely most unusual.

c - Observation 9 is likely most unusual due to the same reasons as observation 65. It's beyond the solid dashed leverage line. High leverage means a point is likely unusual for predictors.

d - Observation 53 is likely most unusual due to the same reasons as observations 65 and 9. Observation 53 is beyond the solid dashed line for leverage. High leverage means a point is likely unusual for predictors.

4. (a-) Observation 66 is likely most unusual. This time we look at the green hard dashed lines which is the studentized residual extreme point cutoff. You use studentized residuals to determine if a response variable is likely unusual or not. Observation 66 is beyond the extreme cut off, meaning it is likely unusual with respect to biomass (response).

(b-) Observation 65 is likely most unusual due to the same reason as 66. It's beyond the extreme cut off for studentized residuals. That means with respect to biomass (response), observation 65 is likely unusual due to the extreme studentized residual.

c- Observation 9 is not likely most unusual as it is between the dotted green lines that represent moderate cutoffs for studentized residuals. Observation 9 is not even close to having an extreme studentized residuals meaning that most likely it isn't unusual in respect to observed biomass (response).

(d-) Observation 53 is likely most unusual due to the same reasons as 66 and 65. It's beyond the extreme studentized residual cutoff. That means in respect to observed biomass (response), observation 53 is likely most unusual due to the extreme studentized residual point.

5. a- Drinking alcohol has a negative impact on average life expectancy. When the value of smoker is held constant, your life expectancy is estimated to go down by 3.2656 years with every extra drink you have per day. according to the MLR model.

c - The average difference between lifespans of smokers and non-smokers is 23.4392 years. That means that non-smokers have an average life expectancy that is 23.4392 years longer than the life expectancy of a smoker according to this MLR model.

6. b - I: $H_0: \beta_1, \beta_2,$ and $\beta_3 = 0$

H_A: at least one of the predictors $\beta_i \neq 0$

II: p-value = 1.669e-12

conclusion: Since p-value of 1.669e-12 is $< \alpha = .05$, we reject H_0. There is significant evidence that at least one of the predictors β_i is not equal to 0.

7. a- $H_0: \beta_2 = \beta_3 = 0$

H_A: at least one of β_2 or β_3 doesn't $= 0$

b- Problem 6b is more reliable ? Will come by during office hours to get this explained didn't have time earlier due to practice + work for other classes

8. You could look at R^2 and R^2_{adj} and compare Model 3 and this new model with displacement. You could see if these values increased or decreased with the addition of displacement. I would use Model 3 to compare to the new model since it had the highest R^2 and R^2_{adj} values compared to the other models. Another way you could statistically support this decision is by looking at residual/normal plots and seeing if there's patterns there after adding displacement. If there's no patterns in those plots, then you're one step closer to proving that adding displacement is valid. You can also compare how the resid and normal plots of Model 3 look like compared to the new model. If the plots look better than Model 3 that's another reason to say adding displacement is valid. I would compare the plots to Model 3 because it has good looking plots plus the 2 R^2 values are high. You can also look at leverage, std resids, and stud resids and plot them to see if there's outliers when adding the new variable. You can compare to Model 3 as well to see if there's some improvements. Also, if p-values go down when compared to Model 3 when adding displacement, that's another good sign to say that adding displacement is valid.