



## Problem 1

Which of the following statements are supported by the R output for this model? (Circle all that apply)

- ☒ (a) The effect of changing pH level of the soil has a greater impact on the biomass than changing the potassium level, given the location of the grass is the same.
- ☐ (b) The effect of changing the potassium level of the soil has a greater impact on the biomass than changing the pH level, given the location of the grass is the same.
- ☐ (c) The effect of changing the pH level of the soil on the biomass depends on the potassium level of the soil.
- ☒ (d) Comparing grass from soil with matching pH levels and matching potassium levels, grass from the Snows Marsh tends to have higher biomass than grass from Oak Island.

## Problem 2

Which of the following represents a valid reduced model if we want to determine whether or not to include the categorical predictor for location? (Circle all that apply)

- ☐ (a)  $\widehat{biomass} = \hat{\beta}_0 + \hat{\beta}_1 pH + \hat{\beta}_2 K + \hat{\beta}_4 LocationSM + \epsilon$
- ☐ (b)  $\widehat{biomass} = \hat{\beta}_0 + \hat{\beta}_1 pH + \hat{\beta}_2 K + \hat{\beta}_4 LocationSM + \epsilon$



6a) I would choose the third model since the residual plot for both the first and second model has a curved pattern. The residuals seem to be more randomly spread out with no obvious pattern. This makes model 3 more reliable.

Likewise, the adjusted  $TR^2$  for model 3 is the highest and it is comparable to the first two model on all other criteria.

7a)

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

There is no relationship between transmission type and fuel consumption.

$$H_A: \text{Some } \beta_2, \beta_3 \neq 0$$

There is a relationship between transmission type and fuel consumption.

7b) The reliability of statistical model depends upon whether the assumptions for fitting the model are met.

The test conducted in problem 6b is a  $t$ -test. Since the residual plot shows a random distribution of residuals and the quantile plot do not show any serious departure from normality, I would say the test conducted in problem 6b is more reliable.

On the other hand, the test in problem 7a is an  $F$ -test. So we need to look at the residual and quantile plots for both model 1 and model 3. Since the residual plot for model 1 violates the linearity assumption and quantile plot shows more deviation than that for model 3, the test in 7a is less reliable.