Practice Assignment: Interaction Terms

1) Which of the following types of variables can be included in a multiple linear regression model?
   a) Categorical variables
   b) Continuous variables
   c) Interaction terms
   d) All of the above

Questions 2 and 3: Consider a multiple linear regression model that predicts infection risk in a hospital using the explanatory variables of average length of patient's stay, frequency of giving X-rays, and U.S. regions (East and West).

2) How many indicator variables are needed for this model?
   a) 4
   b) 2
   c) 1
   d) 3

3) How would you interpret the regression coefficient in general for the U.S. East region, assuming that the reference category is the West region?
   a) The difference in average infection risk for the U.S. East region compared to the West region, assuming length of stay and frequency of X-rays is held constant
   b) The difference in average length of patient's stay for the U.S. East region compared to the West region, assuming frequency of X-rays is held constant
   c) The difference in average infection risk for the U.S. East region compared to the West region
   d) The difference in effect of length of stay on the average infection risk for the U.S. East region compared to the West region

4) What is an interaction effect?
   a) The effect of the response variable on a quantitative explanatory variable
   b) The correlation between two quantitative explanatory variables
   c) The interpretation of how much an explanatory variable will increase/decrease when holding another explanatory variable constant

d) The effect on the response variable of one explanatory variable that depends on the value of another explanatory variable

Questions 5–7: You will need to use the dataset "LeadMoss." The dataset is contained in spreadsheet DCMP_STAT_17C_leadmoss.

This dataset is from a study of atmospheric pollution on the slopes of the Blue Ridge Mountains in Tennessee. The dataset contains the levels of lead found in 70 fern moss specimens (in micrograms of lead per gram of moss tissue) collected from the mountain slopes, along with the elevation of the moss specimens (in feet) and the direction (east/west) of the slope face.

The first 10 observations are provided in the following table:

LeadLevels	***Elevation_Feet***	***Direction***
3.475	2000	west
3.359	2000	west
3.877	2000	west
4	2500	west
3.618	2500	west
2.644	2500	west
4.575	3000	west
8.6	3000	west
6.952	3000	east
4.349	3000	east

5) Given the regression coefficients in the following table, write the multiple linear regression model with an interaction term to predict levels of lead found in the moss using elevation and direction. The reference group is *Direction* = west.

[Continued on the next page.]

Parameter	**Estimate**	**Standard Error**	**t-value**	**Pr > |t|**
Intercept	2.385	5.393	0.44	0.6598

Parameter	Estimate	Standard Error	t-value	Pr > |t|
***Direction* = east**	3.201	7.670	0.42	0.6777
Elevation_Feet	0.002	0.002	0.84	0.4014
***Elevation_Feet* * *Direction* = east**	-0.001	0.003	-0.44	0.6629

Part A: Write the full equation for the model.

Part B: Write out the equation for a model with *Direction* = west.

Part C: Write out the equation for a model with *Direction* = east. Simplify.

6) What is the change in average lead levels for every one-foot increase in elevation for moss specimens on the east slope?

a) 0.001

b) -0.001

c) 0.002

d) 0.003

7) What is the interpretation of the elevation regression coefficient for when the direction is west?