Practice Assignment: Indicator Variable

In this practice assignment, you will be using spreadsheet DCMP_STAT_17A_PulseRate, which was used in Preview Assignment 17.A, to practice interpreting prediction models with indicator variables.

For easy reference, here is the background on the data again:

Students in an introductory statistics class at The University of Queensland participated in a simple experiment.[1] The students took their own pulse rates. They were then asked to flip a coin. If the coin came up heads, they were to run in place for one minute. Otherwise, they sat for one minute. Afterward, everyone took their pulse rates again. The pulse rates and other physiological and lifestyle data were recorded. There are a total of 110 observations and 11 variables. The variables in the dataset are:

Variable	**Description**
ID	Identification number
Height	Height in centimeters (cm)
Weight	Weight in kilograms (kg)
Age	Age in years
Sex	Male/female
Smokes	Are you a regular smoker? (yes/no)
Alcohol	Are you a regular drinker? (yes/no)
Exercise	What is your frequency of exercise? (low, moderate, high)
GroupAssignment	Whether the student ran or sat between the first and second pulse measurements
Pulse1	First pulse measurement (rate per minute)
Pulse2	Second pulse measurement (rate per minute)
Year	Year of class (1993–1998)

Questions 1–7: We are interested in building a prediction model that would allow us to predict the first pulse measurement based on weight and alcohol use.

1) How many levels of the categorical variable *Alcohol* are there? What are the levels?

2) We need to create an indicator variable for the categorical variable *Alcohol*. The reference group should be *Alcohol* = no. What would be the value of each level for the indicator variable?

[1] Wilson, R. J. (n.d.). *Pulse rates before and after exercise*. StatSci.org. http://www.statsci.org/data/oz/ms212.html

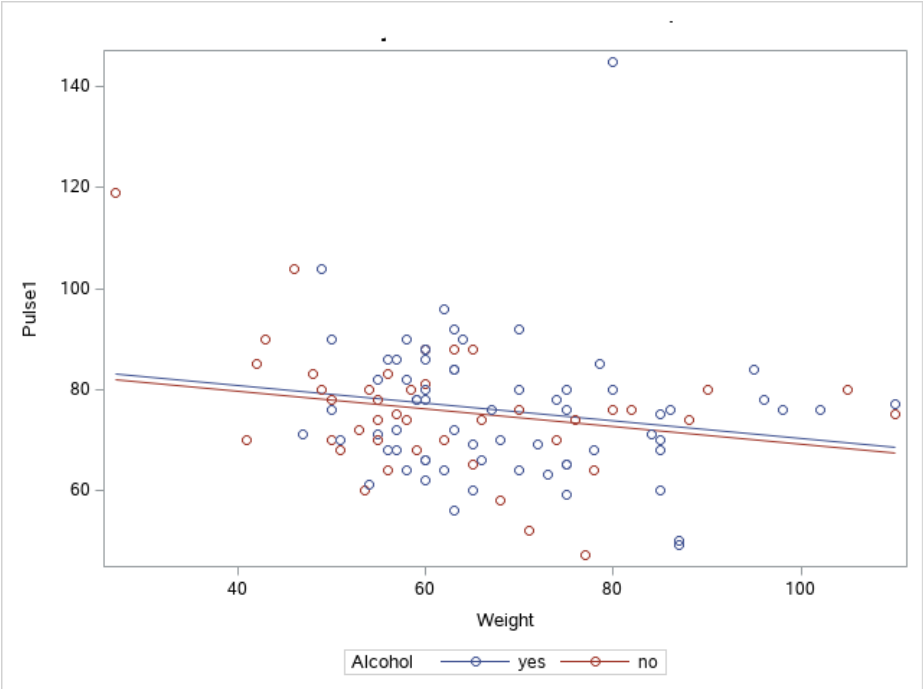
3) Create an indicator variable for *Alcohol* in the following sample dataset of 10 observations. Fill in the appropriate values for the indicator variable.

ID	***Alcohol*** **(Are you a regular drinker?)**	**Indicator variable**
2	Yes	
3	Yes	
6	Yes	
10	Yes	
11	Yes	
18	Yes	
22	No	
26	No	
29	Yes	
30	Yes	

4) Using the following output from a multiple linear regression model, write the multiple linear regression equation for predicting the initial pulse rate (*Pulse1*) using the explanatory variables of math and reading test scores and gender of the student.

Coefficients	**Estimate**
Intercept	86.776
Weight	−0.176
Alcohol indicator variable (1 = yes)	1.011

5) Using the estimated linear model, what is the interpretation of the slope for the explanatory variable of the indicator variable *Alcohol* (1 = yes) in the context of the dataset?

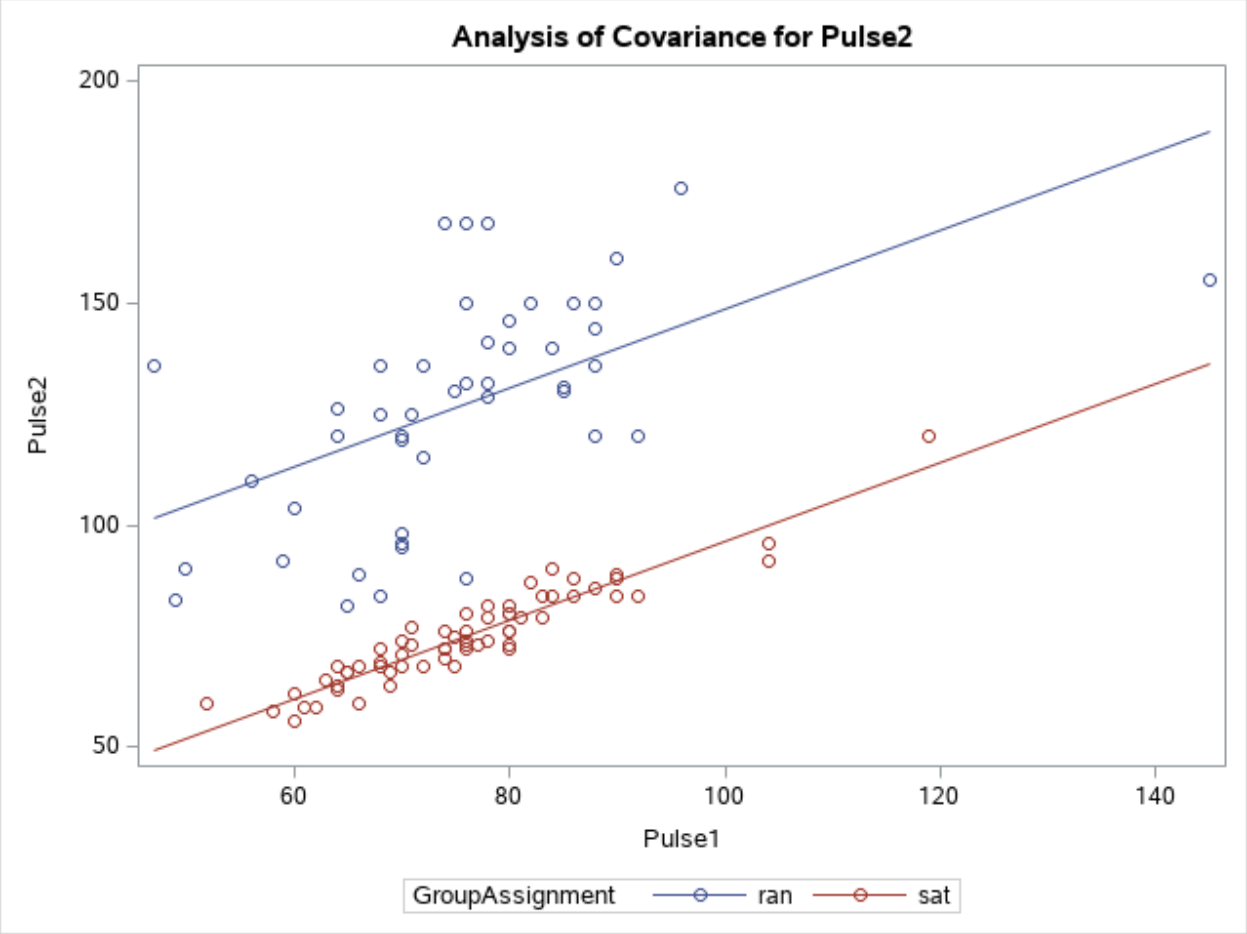


6) Recall that the students in an introductory statistics class at The University of Queensland participated in a simple experiment.[2] The students took their own pulse rates. They were then asked to flip a coin. If the coin came up heads, they were to run in place for one minute. Otherwise, they sat for one minute. Then everyone took their pulse rates again. The pulse rates and other physiological and lifestyle data are included in the dataset. Suppose we wanted to model the differences in the pulse rates between the two groups, controlling for their initial pulse rates. This type of analysis is called Analysis of Covariance (ANCOVA).

   We are looking for significant differences in the expected pulse rates after one minute (*Pulse2*) for the two experimental groups run vs. sit, **controlling for** the students' baseline pulse rates. Basically, you are looking for a group assignment difference *controlling for* or *statistically adjusting for* the students' baseline pulse rates (*Pulse1*). In this type of analysis, you are usually interested in looking for the difference in the indicator variable while controlling for a continuous **covariate**.

   The ANCOVA results are provided in the following figure and table:

[2] Wilson, R. J. (n.d.). *Pulse rates before and after exercise*. StatSci.org. http://www.statsci.org/data/oz/ms212.html



Parameter	Estimate
Intercept	7.54680337
Pulse1	0.88733028
GroupAssignment = ran	52.34616648

Write the multiple linear regression equation for estimating the pulse rate after one minute (*Pulse2*) using the explanatory variables of *Pulse1* and *GroupAssignment*. Round coefficients to two decimal places.

7) What is the expected difference in students' pulse rates after one minute between students who sat for one minute and students who ran for one minute, controlling for their initial pulse rates?