**IST 772 Week 10 Breakout**: **Interactions in Multiple Regression**

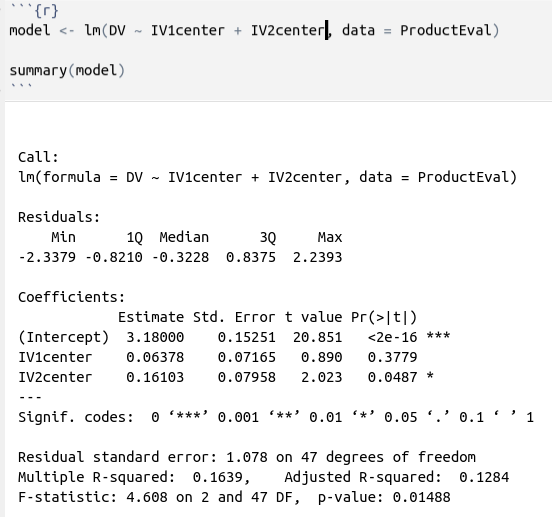
There is a csv file that you can read into R and give the name ProductEval. This dataset contains 50 observations, n=25 from the U.S. (GRP==1) and n=25 from Canada (GRP==2). Each research participant provided a star rating (from 1 to 5 stars) of one brand of earbuds, while listening to a brief musical selection. Each participant provided a self-rating **(IV1)** of on a 10-point scale (1=not at all, 10=totally) on whether they were a music enthusiast. Each participant also took a test of audiophile knowledge **(IV2)** with 20 true-false questions, where the score shows the number of correct responses for each person. (An audiophile is a person who loves listening to high-quality sound from a stereo or home theater.)

Read in the dataset and examine it. Each row contains a **star rating (DV)** of the earbuds, the **self-rated music enthusiasm of the rater (IV1)**, the **audiophile knowledge test (IV2)** and the country of residence for the rater (GRP).

1. Place the Week10interaction.csv file on your computer. Read it in using:  
   library(readr)  
   ProductEval <- read\_csv("Week10interaction.csv").
2. For the 2 independent variables and the dependent variable:
3. Run summary statistics and comment on the results.
4. Look at the histograms and comment on the shape of the distributions.
5. Center IV1 and IV2 by subtracting the mean of each variable from itself or using the scale() function. As a good data management practice, create new variables on the dataset to hold the centered versions of the IVs so that you do not overwrite the original variables. If you use scale(), finish the job by coercing the result with as.numeric() to simplify a later step.

Ex. ProductEval$IV1center <- as.numeric(scale(ProductEval$IV1, center = TRUE, scale = FALSE))

1. Create a main effects only regression model with the centered IVs as the predictors. Report on the output.



1. Create an interaction model with the centered IVs as the predictors. Report on the output. Make sure to examine the significance of the interaction term first.
2. **In the Discussion in Blackboard:**

**a. Report on the results of the main effects model in #4.**

This output is showing the results of a multiple linear regression model with two independent variables (IV1center and IV2center) predicting a dependent variable (DV).

The intercept (3.18) is the estimated value of the dependent variable when both independent variables are equal to zero.

The regression coefficients indicate the estimated change in the dependent variable for each one-unit change in the corresponding independent variable, while holding all other variables constant.

IV1center has a coefficient of 0.06378, indicating that a one-unit increase in IV1center is associated with a 0.06 unit increase in the dependent variable, although this effect is not statistically significant (p = 0.3779).

IV2center has a coefficient of 0.16103, indicating that a one-unit increase in IV2center is associated with a 0.16 unit increase in the dependent variable, and this effect is statistically significant (p = 0.0487).

The overall model fit is evaluated using the F-statistic, which compares the model to a null model with no predictors. The F-statistic of 4.608 with a corresponding p-value of 0.01488 indicates that the model is a significant improvement over the null model, and that at least one of the independent variables is a significant predictor of the dependent variable.

The multiple R-squared value of 0.1639 indicates that the independent variables in the model explain 16.39% of the variability in the dependent variable. The adjusted R-squared value of 0.1284 adjusts for the number of predictors in the model and is a more conservative estimate of the model's explanatory power.

**b. Report on the results of the interaction model in #5.**

The output shows that the model is significant at the 5% level (p-value = 0.005489), and the adjusted R-squared is 0.1884, indicating that about 19% of the variation in the dependent variable is explained by the model.

The intercept coefficient of 3.32879 is statistically significant (p-value < 0.001) and represents the expected value of the dependent variable when both IV1 and IV2 are zero.

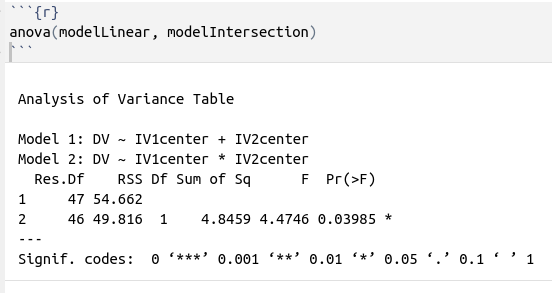
The main effect coefficient for IV1 is 0.05274, but it is not statistically significant (p-value = 0.4507), indicating that IV1 alone does not have a significant effect on the dependent variable.

Similarly, the main effect coefficient for IV2 is 0.10900, but it is not statistically significant (p-value = 0.1831), indicating that IV2 alone does not have a significant effect on the dependent variable.

However, the interaction term coefficient (-0.05165) is statistically significant (p-value = 0.0398), indicating that there is a significant interaction effect between IV1 and IV2 on the dependent variable. This implies that the effect of IV1 on the dependent variable is not constant across all levels of IV2, and vice versa.

In summary, this interaction model suggests that the effect of IV1 on the dependent variable depends on the level of IV2, and vice versa. The model as a whole is significant, but the main effects of IV1 and IV2 are not significant individually, only the interaction term is significant.

**c. Run an anova() to compare the main effects only model (#4) with the interaction model (#5). Give an interpretation of the results.**

  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
The F-statistic of 4.4746 and the corresponding p-value of 0.03985 indicate that the interaction model (Model 2) is a better fit for the data than the main effects only model (Model 1) at the 5% level of significance.

In other words, the interaction between IV1 and IV2 is a significant predictor of the dependent variable, and including this interaction term in the model results in a better fit to the data compared to the main effects only model.