Week 10 Regression Interaction Clinic: Your instructor will supply 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 particular 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.

The research goal related to trying to develop a product evaluation process that is not biased by characteristics of the evaluators. The research question asks whether self-ratings of being a music enthusiast related to the star evaluation of the earbuds. The researchers also want to be able to choose product evaluators whose ratings of a product are unaffected by whether they are music lovers. So the second research question asks about the interaction between audiophile knowledge and music enthusiasm: Do audiophiles as a group provide more stable ratings of a product?

Phase 1: 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. Run summary statistics, histograms, a correlation matrix, and a pairs plot on the data.

3. 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 don't overwrite the original variables. If you use scale(), finish the job by coercing the result with as.numeric() to simplify a later step. Add a comment saying why we center the IVs?

4. Create a main effects only regression model with the scaled IVs as the predictors. Examine the output.

5. Create an interaction model with the scaled IVs as the predictors. Examine the output. Make sure to examine the significance of the interaction term first.

6. Use anova() to compare the main effects only model with the interaction model. Interpret the results.

7. Optionally, also use modelCompare() from the lmSupport package to compare the two models.

Phase 2 – Interpreting the Interaction: Now let's break open the model to see how the interaction works.

8. Create a small data frame with columns for the scaled versions of IV1 and IV2. Create four rows with different combinations by crossing a low value of IV1 with high and low values of IV2, and then a high value of IV1 with high and low values of IV2.

9. Use the predict command to create four predictions of the DV. Add a comment interpreting the pattern of results.

10. Next, create two separate data sets, doing a median split on IV2. So one data set should contain just the rows where IV2 is above the median and another dataset where IV2 is below the median.

11. Run two separate simple regression models where IV1 predicts DV.

12. Create a scatterplot of IV1 versus DV. Then use abline() with the reg= option to plot a different colored regression line for the two different models from the previous step.

13. Using all the evidence you have, interpret the overall results of the study. What rater characteristics are connected with more stable ratings of a product?

Phase 3 – Bonus Examination of Country of Residence: So far, we have not included GRP in any of our analysis. Are U.S. and Canadian raters different? Does it matter?

14. Run t-tests using GRP as the grouping variable. Examine and interpret mean differences on DV, IV1, and IV2.

15. Create a simple linear (main) effects only regression model predicting DV from IV1 (centered), IV2 (centered), and GRP. Should GRP be centered before doing the analysis?

16. Create a model that includes the IV1 * IV2 interaction and just the main effects. Interpret the results.

17. Feel free to experiment with more complex models that contain more of the possible interaction effects. What do you learn from these additional models?