|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PM592: Regression Analysis for Data Science** | | | Name: |  |  |
| **HW5** |  |  |  |  |  |
| *Confounding, Interaction* | | | | |  |

**Instructions**

* Answer questions directly within this document.
* Upload to Blackboard by the due date & time.
* Clearly indicate your answers to all questions.
* If a question requires analysis, attach all relevant output to this document in the appropriate area. Do not attach superfluous output.
* There are 3 questions and 30 points possible.

Researchers at Nittany University were interested in factors that influenced the satisfaction of individuals taking group exercise classes. They recruited 90 individuals who decided to enroll in one of three classes: 1) cardio, 2) strength, 3) flexibility. Individuals took group exercise classes every other day for two weeks. The data they collected is located in the “gx.csv” file.

The researchers were interested in the following:

* Do the rate of perceived exertion, instructor encouragement, participant control, and perceived competence relate to intrinsic satisfaction with the workouts?
* Do these effects vary based on the type of class the participant is engaged in?

**Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Meaning** | **Coding** |
| classtype | Type of class (randomization condition) | 1 = Cardio 2 = Strength 3 = Flexibility |
| age | Age of participant (years) collected at baseline |  |
| bmi | Body mass index of participant collected at baseline |  |
| rpe | Mean of rate of perceived exertion across all workouts | 1-10 scale, higher score represents more exertion |
| encourage | Encouragement scale: “The instructor encouraged me” | 1-7 scale, 1 = strongly disagree, 7 = strongly agree |
| control | Control scale: “The instructor made me do things their way” | 1-7 scale, 1 = strongly agree, 7 = strongly disagree |
| perc\_comp | Perceived Competence scale: “I believe I completed the exercises today the way they should be done” | 1-7 scale, 1 = strongly disagree, 7 = strongly agree |
| satisfaction | Satisfaction scale: a measure of how much intrinsic satisfaction participants had from the program | 6-21 scale, 6 = highly dissatisfied, 21 = highly satisfied |

Note: it may be helpful to convert “classtype” to a factor variable before you begin your analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question 1** |  |  |  |  | [8 points] |  |

Perform a preliminary set of multivariable linear regressions to address the research questions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Model 1  Estimate (SE) Unadjusted | Model 2  Estimate (SE) Age-Adjusted | Model 3  Estimate (SE) BMI-Adjusted | Model 4  Estimate (SE) Age & BMI Adjusted |
| Intercept |  |  |  |  |
| Class Type |  |  |  |  |
| Perceived Exertion |  |  |  |  |
| Encouragement |  |  |  |  |
| Control |  |  |  |  |
| Competence |  |  |  |  |

|  |  |
| --- | --- |
|  | 1a. [4 points] Construct a preliminary table of parameter estimates for 4 models: 1) unadjusted, 2) age-adjusted, 3) bmi-adjusted, and 4) age & bmi adjusted. Use the above table as a template. Note: you will have to figure out how to present the estimates in the table for “class type”, clearly conveying information about the reference group. |

|  |  |
| --- | --- |
|  | 1b. [2 point] For each independent variable in the table, state whether age and BMI appear to confound the relationship between that variable and satisfaction score, and why. |

|  |  |
| --- | --- |
|  | 1c. [2 points] Based on your answer to (1b), which model do you feel comfortable proceeding with? Justify your answer. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question 2** |  |  |  |  | [13 points] |  |

Add complexity to your model by testing whether model effects vary by class type.

|  |  |
| --- | --- |
|  | 2a. [2 points] Do any of the main independent variables (in the table in Question 1) interact with classtype in their association with satisfaction? Provide the p-values you used to test these interactions. |

|  |  |
| --- | --- |
|  | 2b. [3 points] Based on your analyses so far, and considering interactions and confounding (you may reassess confounding after making a decision on 2a), decide on your preliminary final model—the model that best describes the researchers’’ questions. Provide the parameter estimates, standard errors, and p-values of the coefficients. |

|  |  |
| --- | --- |
|  | 2c. [5 points] For the variables that have significant interaction terms, describe the nature of how class type interacts these variables, providing the stratum-specific estimates of the relationships between that variable and satisfaction. Provide a plot that illustrates the interaction. |

|  |  |
| --- | --- |
|  | 2d. [3 points] Evaluate your model assumptions of linear regression. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question 3** |  |  |  |  | [9 points] |  |

Provide a conclusion.

|  |  |
| --- | --- |
|  | 3a. [7 points] Write a concluding methods/results paragraph. Keep your conclusion brief (only your first 400 words will be graded), providing only the most important aspects of your methods and results. Your conclusion should explain:   * The research question you attempted to address * The steps you took to evaluate the research question, including your modeling approach * How you addressed confounding and interactions and what you found * Convince the reader you have a good model with respect to the assumptions and potential outliers * Provide an interpretation of coefficients in your final model (with relevant p-values), keeping in mind how this interpretation relates to the research question |

|  |  |
| --- | --- |
|  | 3b. [2 points] How much of the variation in satisfaction scores is explained by your model? |