|  |  |  |  |
| --- | --- | --- | --- |
| **PM592: Regression Analysis for Health Data Science** |  |  |  |
| **Lab 7 – Advanced Variable Coding**  **Data Needed:** *ex\_sos* | | | |

**This lab is devoted entirely to the exercise.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Lab 7 Exercises** |  |  |  |  |  |  |

|  |  |
| --- | --- |
| Objective(s): | Determine the functional form of independent variables, build an association model from start to finish |
| Datasets Required: | ex\_sos |

This data set re-creates the findings from the Sources of Strength suicide prevention intervention. Baseline data was acquired on students participating in one of the intervention schools (n=1000). These students answered surveys that included questions such as:

* Trusted adults (TATOT): Name up to 7 adults at school who you trust and feel that you can talk to openly. This variable is the sum of the number of adults they named.
* Maladaptive help-seeking (M\_HELP): A series of questions asking them about their help-seeking attitudes. This survey consisted of 6 questions that were combined into a composite score. Higher scores reflected students were more maladaptive with regard to help-seeking for problems—that is, they viewed seeking help for themselves and/or their friends more unfavorably.
* Age, in years (AGE)
* Ethnicity (ETH): classified as white (W), black (B), or other (O).
* Gender (GENDER): classified as 1=female, 0=male

Researchers were interested in whether more adult connections at school improved students’ attitudes toward seeking help. Additionally, they were interested in whether this effect was even stronger for students who named no adults vs. naming any adults. They wanted to adjust for age, ethnicity, and gender.

1. Begin this analysis by producing the following descriptive information:
   1. To make the scale of m\_help more interpretable, create a variable called m\_help.z which is the z-score for m\_help.
   2. Summary statistics for all variables in the data set. Ensure that the ranges make sense and examine any outliers.
   3. Does anything concern you about the distributions of these variables?
2. Run ggpairs() on all variables in the data set.
   1. Does anything concern you about the associations among all these variables?
   2. Do any variables appear to be collinear?
3. Examine the effect of trusted adults – keep in mind the research question: whether maladaptive attitudes are even higher for those who named no adults.
   1. Examine the mean of m\_help.z for each value of tatot.
   2. Produce a graph of m\_help.z by tatot that includes the mean of m\_help.z for each value of tatot (this is in the lab R file).
   3. Does the difference in mean m\_help.z associated with 1 vs 0 trusted adults appear greater than the difference associated with 2 vs 1 trusted adults?

(Questions 3d-3j are optional for advanced modeling practice)

* 1. Create the following two variables as a set to describe the effect of trusted adults:
  2. What is the baseline value for this variable set?
  3. What will represent?
  4. What will represent?
  5. Fit the regression model of m\_help.z on noad and tatot1. What is your output?
  6. Use the Extra Sums of Squares test to determine whether this coding scheme fits better than a model in which trusted adults is treated completely linear (tatot).
  7. Given this information, how will you model trusted adults?

1. Examine the effect of age
   1. Examine the mean of m\_help.z for each value of age.
   2. Produce a graph of m\_help.z by age that includes a locally-weighted “loess” smoothed line.
   3. Does age appear to have a linear effect on m\_help.z? If not, explain what the relationship might be.
   4. Fit the regression model of m\_help.z on however you decided to code age. What is your output?
   5. Given this information, how will you model age?
2. Examine the effect of gender
   1. Examine the mean of m\_help.z for each value of gender.
   2. Produce a graph or boxplot of m\_help.z by gender.
   3. Determine whether m\_help.z varies by gender using whichever method you like.
3. Examine the effect of ethnicity
   1. Examine the mean of m\_help.z for each value of ethnicity.
   2. Produce a graph or boxplot of m\_help.z by ethnicity.
   3. Do you think that the “other” ethnicity category should be combined with any other group? Consider the sample size in that category, and how similar it is to other groups.
   4. Determine whether m\_help.z varies by ethnicity using whichever method you like.
4. Combined model
   1. Include all covariates of interest with the effect for trusted adults in one model.
   2. Establish your “preliminary final model” – a model you’re content with evaluating
5. Model diagnostics
   1. Examine the assumptions of linear regression. Are these assumptions met?
   2. Examine influential observations. Which observations stand out as being influential?
   3. Perform a sensitivity analysis by excluding the influential observations. How different is your model and do your conclusions change?
6. Wrap-Up
   1. Report the means of variables, unadjusted (univariable) parameter estimates, and adjusted (multivariable) parameter estimates in a table.
   2. Write a concluding sentence or two about the effect of trusted adults on maladaptive help-seeking.