

Class 2 – Problem set

Due on Canvas on or before Sep 13

Submit a single document with all responses in it

You have only one task to complete: Develop and assess a multivariate regression model explaining job satisfaction (JobSat) in the GSS data

- 1) Choose four variables of interest from the GSS data (do not choose any of Age, Female, or JobPride). These four variables will be IVs used to predict JobSat.
 - a) At least one of them must be “new” (i.e., not already cleaned by me, so you must clean it for use).
 - i) Look through the codebook for a variable of interest
 - ii) Use the cleaning code from session #1 to assign missing values the value of NA
 - iii) Make any other changes you deem required (e.g., simplify categories, reverse direction, etc.)
 - b) Submit your cleaning code.
- 2) After choosing variables, and before conducting any analysis, make predictions. For each of your four IVs, write a paragraph explaining what you think you will find, and why you think so (i.e., four paragraphs total). Before writing these paragraphs, look at 3c below. Your prediction needs to imply an effect size and provide a rationale for it. [This may feel hard because it’s more precise. You may be guessing; the explanation may be thin. That’s fine; this area isn’t your domain of expertise. Just make your best reasoned guess and give a halfway credible justification for it. I want you to practice doing it, not worry about the accuracy of your prediction.]
- 3) State ONE of your four hypotheses in three ways
 - a) In the traditional NHST format of null hypothesis and alternative hypothesis
 - b) In the conventional format used in published articles (look at AMJ and JAP for examples, if you are unsure)
 - c) In the format suggesting by Cumming (2014) in Step 1 of his eight-step strategy
- 4) State the remaining three hypotheses in the Cumming format only
- 5) Create a “Table 1” summary of your data. (No need to interpret it)
- 6) Run the model. Include your R code and output
- 7) Explain whether each of your hypotheses were supported. Refer to p-values, confidence intervals, and effect sizes
- 8) Conduct a diagnostic analysis of your model. In each of the areas below, include R code, output, an explanation of whether the situation is good or bad, and how you know
 - a) Linear relationships between IVs and DV

- b) Multicollinearity
 - c) Normality of residuals
 - d) Presence of heteroskedasticity
 - e) Influential cases
- 9) Based on the results from #8, what can you conclude about your model? If you were going to continue with this research and model, would you need to change any of your analysis? Any of your conclusions? If so, why and in what way?
- 10) Reflect on your findings in a five-paragraph mini-discussion section: One paragraph for each of your hypotheses, plus one addressing the most important implications of your data and/or your results.