Exercises Day 3

PSY8003

Matthias Mittner

spring 2022

Exercise 1: Interactions

Do experts overrate their expertise?

This example is adapted from https://sites.trinity.edu/osl/data-sets-and-activities/regression-activities

The data is published in the following article:

> Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. Psychological Science, 26, 1295-1303.

Background

Valuing expertise is important for modern life. When people have a problem, they need to know who to turn to for a solution to their problem. For example, when people get sick, they know that a doctor is an expert in the field of medicine and can help them get better. In general, experts simply know more about a topic than do non-experts. However, experts may be vulnerable to a particular problem of knowing so much. They may have the illusion that they know more about a topic than they actually do. This particular type of overconfidence is called overclaiming. Essentially, overclaiming occurs when people claim that they know something that is impossible to know, such as claiming to know the capital of Sharambia (a country that doesn't actually exist).

To test if experts are susceptible to overclaiming, Atir, Rosenzweig, and Dunning (2015) recruited 202 individuals from an online participant pool. They first asked participants to complete either a measure of self-perceived knowledge, or an overclaiming task (to test for a possible order effect, half of the participants completed the measure of perceived knowledge first, whereas the other half completed the overclaiming task first). The self-perceived knowledge questionnaire asked people to indicate their level of knowledge in the area of personal finance. The overclaiming task asked participants to indicate how much they knew about 15 terms related to personal finance (e.g., home equity). Included in the 15 items were three terms that do not actually exist (e.g., annualized credit). Thus, overclaiming occurred when participants said that they were knowledgeable about the non-existent terms. Finally, participants completed a test of financial literacy called the FINRA. Whereas the earlier questionnaires measured self-perceived knowledge, the FINRA measured actual knowledge.

More background is available from the original article (file `Atir2015.pdf`).

The following variables are in the file `atir.dta`:

- `id`: subject identifier
- `order_of_tasks`: which task came first? (categorical, either 1 or 2)
- `self_perceived_knowledge`: self-rated knowledge in the area of personal finance (1 = not knowledgeable at all, 7 = extremely knowledgeable)
- `overclaiming_proportion`: ranges from 0-1 and measures how often subjects said they were knowledgable about a non-existent topic (see paper for details)
- `accuracy`: continuous variable quantifying accuracy of detecting non-existent topics: "Accuracy was obtained by subtracting the averaged false alarm rate from the averaged hit rate (i.e., the proportion of real items about which each participant claimed knowledge, averaged across all six potential cutoff points)"
- `FINRA_score`: objective measure of expertise in personal finance (number of correct answers, ranges from 0-5)

Exercises

1. You want to examine the relationship between self-perceived knowledge and overclaiming. You also want to take into account the accuracy with which participants responded during the overclaiming task (that is the ability of people to distinguish between the 12 real terms and the 3 fake terms). Conduct an analysis that uses both self-perceived knowledge and accuracy to predict overclaiming.

2. You next want to determine whether there is an order effect (based on whether participants completed the self-perceived knowledge measure first, or the overclaiming task first. Compare the mean level of overclaiming based on the order of the tasks.

3. If you found a significant difference in overclaiming in the analysis above, re-perform the analysis from #1 to check to see if the relationship between self-perceived knowledge and overclaiming changes, when taking into account the order of the tasks.

    - run a linear regression model that includes `order_of_task` as a dummy-variable and model the main effect of `order_of_task` and the `order_of_task` x `self_perceived_knowledge` interaction
    - is the main effect of `self_perceived_knowledge` still present?

- is there an interaction effect? If it is, what does it mean?

4. You next want to determine if the self-perceived knowledge still predicts overclaiming while accounting for the variance due to genuine expertise, as measured by the FINRA.

   - First, find the mean and standard deviation for scores on the FINRA.
   - Then, re-perform the analysis from #1, but this time include scores on the FINRA as an additional predictor variable.

5. Based on this final model, perform an analysis of the residuals.

Exercise 2: Nonlinear regression

This data is taken from https://osf.io/mc5gu/. Load the dataset `explorepenguin.dta`. There are many variables in this file. We will focus on only a few of them.

`age` is given as year of birth such that the analysis can be updated depending on when it is run. - make a new variable `age.years` which contains the age in years for each subject - make an additional variable `age.dec` that contains the age in decades for each subject (1 unit = 10 years)

`nwsize` is a measure of the size of a persons social network (see https://osf.io/2rm5b for details). - make a scatter plot of network-size as a function of `age.dec` - fit a linear regression line and report (and interpret) the result - plot the predicted against the residual values; what do you see?

Fit a fractional polynomial model to the relationship between `age` and network-size.

- fit the model using the appropriate function (`mfp` in R, `fp` in Stata)
- now that you found the powers of this best model (let's call them `p1` and `p2`), create two new variables in the dataset that are transformed versions of `age.dec` using the powers from the FP model

  - $\texttt{age.dec.p1} = \texttt{age.dec}^{p1}$
  - $\texttt{age.dec.p2} = \texttt{age.dec}^{p2}$

- fit the polynomial regression model that uses those powers

  - this fit will give you coefficients for the `intercept` and the powers of the transformed `age.dec` variables
  - use these coefficients to create a new variable that contains the "predicted" values from the FP model
  - $\texttt{predicted} = \texttt{intercept\_coef} + \texttt{coef1} \times \texttt{age.dec.p1} + \texttt{coef2} \times \texttt{age.dec.p2}$

- plot the `age.dec` variable against this `predicted` variable to observe the non-linear trend of the model
  - if possible, overlay this prediction onto the original dataset (scatter plot)
- to see the independent effect of the two powers, you can also plot the separate parts of the model:
  - create two variables `predict_p1` and `predict_p2` in the same way as above with
    - $\texttt{predict\_p1} = \texttt{intercept\_coef} + \texttt{coef1} \times \texttt{age.dec.p1}$
    - $\texttt{predict\_p2} = \texttt{intercept\_coef} + \texttt{coef2} \times \texttt{age.dec.p2}$
  - create plots of the two functions to see the effect of each individual coefficient
  - can you give an interpretation?
- Finally, plot the prediction vs. the residuals; what do you see?

Exercise 2: Splines

Fit a few regression splines to the data, varying the number of knots and the order. Which setting of the parameters do you think provides a good fit to the data?

- Instead of using Stata or R, you can use the simplified version of this webapp: https://ipsuit.shinyapps.io/splinedemo/
- you see here a plot of the dataset discussed in this exercise and you can play around with the number knots and the degree of the regression splines
- all necessary output is provided in the table
- make sure to select the correct dataset (`explorepenguin_share`) from the drop-down menu