Lab 4: Subgroup-specific effects and omnibus tests

Data: mri.csv (see the mri.pdf file for data dictionary/useful information).

Practical objective: To practice extraction of subgroup-specific effects and implementing testing in Stata.

Scientific objective: To investigate the association between smoking and FEV in adults.

Noteworthy commands: Below is a list of Stata commands that will be helpful for this lab.

- `regress`
- `lincom`
- `test`
- `testparm`

Exercises: Below is a description of the scientific question that we will use the MRI study to answer, along with accompanying exercises (we will go through them individually, in small groups, and/or together as appropriate and as time permits).

Scientific Question: We seek to understand how smoking history is associated with lung function. Key features of a patient's smoking history include: (1) the extent of prior smoking history and (2) whether or not—and for how long—a patient has stopped smoking. These variables are represented in the data set by the names packyrs and yrsquit, respectively. Therefore, consider the following model:

$$\mathbf{E}[\text{FEV} \mid \text{packyrs}, \text{yrsquit}] = \beta_0 + \beta_1 \text{ packyrs} + \beta_2 \text{ yrsquit} \qquad [1]$$

Exercise 1: Provide plain-language interpretations for each coefficient of Model [1], and estimate its parameters in Stata using OLS.

Exercise 2: Use Model [1] to test whether there is an overall association between smoking history and mean FEV.

Exercise 3: Consider the following subgroups: (1) individuals with no smoking history, and (2) current smokers with a smoking history of 30 pack-years. Would you expect Model [1] to produce a confidence interval for the mean FEV that is meaningfully smaller, meaningfully larger, or about the same width in subgroup (2) as compared to subgroup (1)? Justify (and verify) your answer.

In reality, we may reasonably suspect that the association between extent of smoking and FEV may be modified by the time since quitting smoking, and that age may be an important confounder to account for. Therefore, consider the following model:

$$\mathbf{E}[\text{FEV} \mid \text{packyrs}, \text{yrsquit}, \text{age}] = \beta_0 + \beta_1 \text{ packyrs} + \beta_2 \text{ yrsquit} + \beta_3 \text{ yrsquit} \times \text{packyrs} + \beta_4 \text{ age} \qquad [2]$$

Exercise 4: Provide plain-language, literal interpretations for each coefficient of Model [2]. Which are scientifically meaningful? Estimate the parameters of Model [2] in Stata using OLS.

Exercise 5: Use Model [2] to test whether there is an overall association between extent of prior smoking and mean FEV.

Exercise 6: Use Model [2] to test whether there is an overall association between time since quitting smoking and mean FEV.

Exercise 7: Use Model [2] to test whether there is an association between overall smoking history and mean FEV.

Exercise 8: Consider the following subgroups: (1) 80-year old individuals with no smoking history, and (2) 80-year old current smokers with a smoking history of 30 pack-years. Would you expect Model [2] to produce a confidence interval for the mean FEV that is meaningfully smaller, meaningfully larger, or about the same width in subgroup (2) as compared to subgroup (1)? Justify your answer with exploratory techniques. Verify your answer.

Exercise 9: Use Model [2] to determine a point estimate and 95% confidence interval for the age-adjusted association between extent of prior smoking and mean FEV among individuals who report not having smoked in 20 years.