1.6   Practical 1

Dataset required: `mental.dta`

Introduction

The purpose of today's session is to perform a full analysis of the mental activity data trial introduced earlier We will use a series of normal error models, and compare their respective fits using F-tests.

The data are from a three-arm placebo-controlled randomised trial assessing the effects of injection with morphine and heroin on mental activity. Seventy two people were randomized to one of three arms (placebo, morphine, heroin), and mental activity was assessed by the sum of scores on seven items of a questionnaire, with a high score indicating high mental activity.

The dataset contains three variables:

   i **treat** - Randomised arm: 1=placebo; 2=morphine; 3=heroin

  ii **prement** - Mental activity score before injection

 iii **mentact** - Mental activity score 2 hours after injection with drug.

The questions that the analysis should address are as follows.

Aims

  1 Do the mean post-injection mental activity scores differ among groups?

  2 What effect does adjustment for pre-injection mental activity score have on your conclusions?

  3 Are the effects of the drugs different for subjects with different levels of pre-injection mental activity?

Analysis

  1 Read the data into Stata and examine. Use tables, histograms and plots to look at the distributions of the pre- and post-injection scores.

    Discuss: Discuss with your colleagues (in your Breakout room if online) what can be concluded about the appropriateness of linear regression models for post-injection scores on the basis of these preliminary explorations.

  2 Use Stata's `lowess` command to plot the post-injection scores against the pre-injection scores with a lowess smoother, separately in each of the three groups. Explore what happens if you change the bandwidth. The basic form of the command is along the lines of the following:

```
.  lowess mentact prement if treat==1, bw(0.8)
```

  3 Write down in algebraic form the linear predictors that correspond to each of the models described in Table 1 below. Carefully define each term that you use in the various models.

Note that the distributional assumption in each model is $Y_i \sim N(\mu_i, \sigma^2)$ where Y_i is the post-injection mental activity score. In each model the link function is the identity $\eta_i = \mu_i$.

Table 1

#	Terms fitted	Linear predictor
1	Overall mean	
2	Drugs	
3	Pre-inj (Pre)	
4	Drugs + Pre	
5	Drugs + Pre + (Drugs-by-Pre interaction)	

4 Fit each linear regression model in Stata, using the `regress` command, to complete Table 2.

Table 2

#	Terms fitted	RSS	Residual df
1	Overall mean		
2	Drugs		
3	Pre-inj (Pre)		
4	Drugs + Pre		
5	Drugs + Pre + (Drugs-by-Pre interaction)		

5 Use Stata to calculate fitted values for model 5. Display these for each trial arm in a graph. Compute the intercepts and slopes of the three fitted lines from the parameters estimated in the model.

Table 3

Trial arm	Fitted Intercept	Fitted Slope
Placebo		
Morphine		
Heroin		

6 Pen & paper exercise: use the results in Table 2 to carry out formal comparisons of

   (a) the fits of models 3 and 4

   (b) the fits of models 4 and 5

Note: it is good practice to use a calculator for these calculations, rather than your phone or computer software such as Excel or Stata.

Confirm your results using Stata's `test` command.

Discuss: What do you conclude from these tests? Why can an analogous test not be used to compare the fits of models 2 and 3?

7 Use the `glm` command to fit model 4. In what way do the results differ from those with `regress`?

8 Use results from models 2 and 4 to complete Table 4.

Table 4

	Mean	Mean difference from placebo	SE	Adjusted difference from placebo	SE
Placebo		0	-	0	-
Morphine					
Heroin					

9 Briefly summarise your conclusions regarding the effects of morphine and heroin on mental activity scores in this trial.

Working together with one or more of your colleagues (in your Breakout Room if online), write a paragraph to summarise your conclusions from these analyses. You should try to answer each of the aims listed above. If online, one of you should post your group's paragraph in the Zoom chat.

10 If you have time, explore how to obtain an estimate and confidence interval for the difference in mean post-treatment scores between the morphine and heroin groups using model 4.

There are two ways to do this:

(a) use the post-estimation command `lincom`, or

(b) re-parameterise the model using the "bx." syntax to replace "i.", where x is the number of the group you wish to make the baseline.