Exercise for SPSS Syntax Workshop
Nov 5, 2014

1. Create a random sample of 200 students, who are nested within 10 classes, with 20 students per class[1]. Create two random variables based on normal distribution, with mean = 2 and standard deviation = 0.5 for GPA and mean = 10 (i.e., hours per day) and std = 3 for Webtime. Table 1 describes the data:

Table 1. students.sav

Student ID (id)	Class ID (j)	GPA	Webtime
1	1	…	…
…	…	…	…
200	10	…	…

2. Aggregate students.sav to classes.sav, based on j, with two new variables GPA.mean (the average GPA of the class) and GPA.sd (the standard deviation of the class). Table 2 describe the data:

Table 2. classes.sav

Class ID (j)	GPA.mean	GPA.sd
1	…	…
…	…	…
200	…	…

3. Merge students.sav and classes.sav (based on which variable?) and save the results in students-classes.sav as described in Table 3:

Table 3. students-classes.sav

Student ID (id)	Class ID (j)	GPA	Webtime	GPA.mean	GPA.sd
1	1	…	…	…	…
…	…	…	…	…	…
200	10	…	…	…	…

4. Create a random sample of 200 websites, assuming each being visited by a student of the same ID. Create a variable (content) based on normal distribution with mean = 0 and standard deviation = 1. Table 4 describes the data:

Table 4. websites.sav

Website ID (id)	content
1	…
…	…
200	…

5. Merge websites.sav with students-classes.sav, create WebtimeXContent by multiplying the two relevant variables, and save the results to students-classes-websites.sav, as described in Table 5:

Table 3. students-classes.sav

Student ID (id)	Class ID (j)	GPA	Webtime	GPA.mean	GPA.sd	Content	WebtimeXContent
1	1	…	…	…	…	…	
…	…	…	…	…	…	…	
200	10	…	…	…	…	…	

6. Run a regression analysis with GPA being the dependent variable (individual-level), GPA.mean (class-level), GPA.sd (class-level), and WebtimeXContent (individual-level) being the independent variables. Try to explain what each of the three independent variables really measures here, regardless of the significance of their effects.

[1] See the revised PPT of last time on Web Mining Lab website to learn how to create a hierarchically nested sample.