Quantitative Analysis in the Social Sciences – Fall semester 2020-2021
Course: Loglinear models and latent class analysis

Deadline to submit assignments: 22 January 2021, 5pm CET
Exams submitted late will not be accepted.

Students should work independently.

Please send an electronic copy of your assignment (and annexes) to the following email address: femke.dekeulenaer@kuleuven.be

Include with your assignment, an annex with a relevant selection of input and output files.

Exercise 1

1. Table 1 presents the unweighted results from a contingency table of geographic region by education level for a sample of Belgian citizens (surveyed in the European Social Survey). The six levels of education range from "Primary education or less" to "University education".
    a. Calculate the local odds ratios for the region by education table.
    b. Interpret the relationship in terms of odds ratios.
2. Due to differential non-response, the distribution of completed interviews for this sample of Belgian citizens does not match that found in the Belgian Labour Force statistics (see table 2 below).
    a. This table can be re-estimated in such a way that all marginal frequencies for the survey data equal those of Belgian Labour Force statistics (Table 2) and the original relationships present in the bivariate survey table (Table 1). This re-estimation can be done via a Deming-Stephan Adjustment. Describe in words what the procedure does, and what the steps are in the adjustment process.
    b. Next, proceed with the re-estimating of the table as described above - i.e. marginal frequencies for the survey data equal those of Belgian Labour Force statistics and the original relationships present in the bivariate survey table are still present in the new table; prove the latter using odds ratios. (Note: the re-estimated table can be calculated by hand or with support of the Lem program.)

Table 1: Region by education – 15-64 year-olds – Belgium (unweighted data from European Social Survey)

	Brussels Capital Region	Flemish Region	Walloon Region	*Total*
Primary or less education	6	62	43	*111*
Lower secondary education	14	169	100	*283*
Higher secondary education	46	327	186	*559*
Higher education, short type	16	174	72	*262*
Higher education, long type	6	44	23	*73*
University education	21	72	42	*135*
Total	*109*	*848*	*466*	*1,423*

Table 2: Region by education – active population in Belgium (based on weighted data from Labour Force Statistics)

	Brussels Capital Region	Flemish Region	Walloon Region	*Total*
Primary or less education	171,796	954,045	596,239	*1,722,080*
Lower secondary education	168,191	1,040,981	604,826	*1,813,998*
Higher secondary education	240,671	1,872,732	991,640	*3,105,043*
Higher education, short type	69,168	726,410	323,215	*1,118,793*
Higher education, long type	60,079	207,366	134,066	*401,511*
University education	168,169	425,786	221,696	*815,651*
Total	*878,074*	*5,227,320*	*2,871,682*	*8,977,076*

For exercise 2 and 3, you can use a **data set of your choice**. If you do not have data from your own research, two data sets have been included on the Toledo page of the course:

- European Social Survey 2008, multi-country data set (take care to select the data from just one country, to avoid analysing a sample that is too large) – this data set can be used for Ex. 2
- US General Social Survey, single-country data set – this data set can be used for Ex. 3

You can also find other interesting datasets via ZACAT - GESIS Online Study Catalogue (https://zacat.gesis.org/webview/) and more recent years for the European Social Survey via https://www.europeansocialsurvey.org/data/

Exercise 2

1. Create a 4 or 5-dimensional **contingency table** in which one of the variables will be treated as the dependent variable. The dependent variable, and at least one of the independent variables, should have at least three categories.
    a. Include the table in your assignment
    b. Describe the variables selected
    c. Formulate an initial set of hypotheses about the expected relationships between the variables (e.g. by referring to earlier studies on the topic)

    Note: take care that there enough observations in your contingency table, but also not too many (e.g. when taking data from a multi-country dataset, such as the European Social Survey or Eurobarometer, start by selecting the data from just one of the countries and focus your analysis on this country).
2. Use Aitkin's Simultaneous Testing Procedure, taking multiple testing into account, to draw conclusions about the different "families of effects" that may contain significant effects that will be important when describing the associations in the table.
3. (Re-)formulate the loglinear model into a multinomial logit model (using dummy coding for both the dependent and independent variables). Explain what is the most important difference between these two models (dependent quantity).
4. Apply a backward selection procedure starting from the logit model with at least all 3-way interactions (also think about how to take into account the outcome of Aitkin's simultaneous testing). Evaluate the acceptable models in terms of L^2 conditional testing and by using the AIC and BIC. Decide which model is the best fitting model.
5. Discuss the model fit of this best fitting model also in terms of the dissimilarity index score, pseudo R^2 and size of the standardized residuals.
6. Your dependent variable has at least three categories (and you have used dummy coding for the dependent variable). Decide whether you will present the results and interpretation in terms of Baseline-Category Logits or Adjacent-Categories Logits. Explain your decision.
7. Present the Baseline-Category or Adjacent-Categories Logits and discuss the parameter estimates, also discuss the associated odds ratios. Note: interpret at least one interaction effect. *You should present your results as if you are writing a basic scientific paper; the interpretation of your results should also include a discussion on how your results fit in with the hypotheses you formulated.*

Exercise 3

1. Create a 4 or 5-dimensional table in which each variable is theoretically a measure of the same underlying latent concept. Each variable is allowed to be dichotomous.
2. Test whether the items actually measure one latent variable by performing an exploratory latent class analysis, paying attention to the number of latent classes. Discuss each step in terms of model fit and draw your conclusion about the final model.
3. Discuss the results of your final latent class model in terms of the size and characteristics of the latent classes identified in your analysis.
4. Formulate at least two hypotheses for a confirmatory latent class analysis. Impose the restrictions for your hypothesis on the model, and draw conclusions.