Exercise 16

Mixed models

In this exercise we studies temperatures measured in cites in the USA. The data gives the normal average January minimum temperature in degrees Fahrenheit with the latitude and longitude of 56 U.S. Cities. For each year from 1931 to 1960, the daily minimum temperatures in January were added together and divided by 31. We want to study whether variation in temperature can be explained by:

Latitude (north-south dimension)
Longitude (west-east dimension)
Elevation (measured at the airport)
Coastal (0=city in inland state, 1=city in coastal state)

(Reference: Peixoto, J.L. (1990) *A property of well-formulated polynomial regression models. American Statistician*, 44, 26-30. Also found in: Hand, D.J., et al. (1994) *A Handbook of Small Data Sets*, London: Chapman & Hall, 208-210. Downloaded from *The Data and Story Library*, see http://lib.stat.cmu.edu/DASL/) The temperature is fake data derived from real data (i.e. data as offered on the DASL website), an therefore still represents some truth.

We study a model with repeated measurements on each subject over time. This is slighly different from the example we saw earlier. In this example the subjects are cities, and the items are years. But the years may show some trend, it will appear that year is both a random and a fixed factor.

Download the table temperature.sav from: http://www.let.rug.nl/~heeringa/statistics/stat03_2013/ and load the table in SPSS.

1. Centre the variables Latitude (mean is 38.970), Longitude (mean is 90.963) and Elevation (974.98).

2. Create side-by-side boxplot diagrams showing by-subject variations (variation over the cities) and by-item variation (variation over the years). What would you expect?

3. Perform the linear mixed model analysis. Perform the following steps.

a. When choosing mixed models, put Cities under Subjects. Click on Continue
b. In the next window, move Temperature to Dependent variable, Coastal to Factor(s) and Year, Latitude, Longitude and Elevation to Covariates.
c. Click on Fixed. Choose the following main effects: Year, Coastal, Elevation, Latitude, Longitude. Choose the following interactions: Coastal * Elevation, Coastal * Latitude, Coastal * Longitude, Elevation * Latitude, Elevation * Longitude. As to the covariates, use the centred variables!!! Click on Continue.
d. Click on Random. Move Year to Model, and City to Combinations. Click on Continue.
e. Click on Estimation. Check Maximum Likelihood (ML). Click on Continue.
f. Click on Statistics. Check Parameter estimates. Click on Continue
g. Click on EM Means. Move Coastal to Display Means for. Check Compaire main effects and choose Bonferroni. Click on Continue.
h. Click on Save. Check Predicted values and Residuals. Click on Continue.
i. Click on OK.

4. Test the assumptions of linearity and homeskedasticity. Make a residual plot and consult it.

5. Test the assumption of multicollinearity. Correlate the covariates (Year, and the centred versions of Latitude, Longitude and Elevation) with each other. Are the correlations smaller than 0.9? Create also a Matrix Scatterplot which shows all pairwise correlations of the four covariates.

6. Check the normality of the residuals. Create a normal quantile plot and perform the Shapiro-Wilk test on the basis of the residuals.

7. Try to find outliers by investigating the residuals. Check whether residuals have an absolute value larger than 3.29, no more than 1% have an absolute value larger than 2.58, no more than 5% have an absolute value larger than 1.96.

a. Standardize the residuals: Analyze, Descriptive Statistics, Descriptives. Move Residuals under Variable(s). Check Save standardized values as variables. Click on OK. A new column contains the standardized residuals.
b. Calculate the absolute standardized values. Use Compute Variable under Tranform in the main menu.
c. Select Transform, Recode into Different Variables. Move the absolute z-scores to Numeric Variable → Output Variable. Click on Old and New Values. Choose Range value trough HIGHEST and enter 3,29. (NB: SPSS requires a comma instead of a point.) Give a new value: 1. Click on Continue. Give a name to the output variable, e.g. Group1. Click on Change, click on OK. Repeat his for 2.58 (Group 2) and 1.96 (Group 3).
d. Go to Analyze, Descriptive Statistics, Frequencies. Move Group1, Group2 and Group 3 under Variable(s). Click on OK.
e. For each group a table is given. The percentage (if it is given) is given in row Valid and column Percent.
f. What do you conclude.

8. Look at the table Tests of Fixed Effects. What do you conclude?

9. Create a graph that shows the interaction between centred Longitude and Coastal. What does the graph tell us?