FES 524: Natural Resources Data Analysis

Reading 6.2: Correlation structures

Contents

1 Correlation matrices

This reading delves a little deeper into *correlation structures.* Like with statistical models, the mathematical representation of correlations can be an efficient way to describe and understand correlation structures. You will only be asked to describe any correlations on assignments in words, not with mathematical notation, but the goal of introducing correlations as symbols is to help in your overall understanding of correlations.

A correlation matrix is one way to describe the within-subject correlations among errors. A correlation matrix is a square matrix, meaning there are as many rows as columns. Each element of the matrix stores the pairwise correlation between the errors of the repeated measures. For example, the value in the first row and third column stores the correlation between ϵ_{i1} and ϵ_{i3}, the errors for the first and third repeated measurements of subject i. Correlations do not depend on order, so the element in the first row and third column will be equal to the element of the third row and first column. We call a matrix like this *symmetric*, since the values are mirrored over the diagonal. We will denote the correlation matrix for the within-subject errors as $\mathbf{\Omega}_i$ and the covariance matrix $\mathbf{\Sigma}_i$, and individual elements as $\mathbf{\Omega}_{i,jk}$ for the element in the jth row and k^{th} column of $\mathbf{\Omega}_i$. When we eventually discuss the full correlation and covariance matrix of all the errors collectively, we will drop the subscript i.

1.1 No correlation

In the absence of correlation between errors of repeated measures of subject i, we have

$$\text{corr}(\boldsymbol{\epsilon}_i) = \mathbf{\Omega}_i = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}_m$$

where $\mathbf{I}_m$ stands for the *identity matrix* with m rows and m columns, where m is the number of repeated measurements for the subject. Without correlation, the correlation matrix for the errors within a subject is the identity matrix since all random variables have a correlation of 1 with themselves, but zero correlation among different errors.

1.2 A matrix with correlation

With correlation among errors within a subject, we might have something like

$$\text{corr}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Omega}_i = \begin{bmatrix} 1 & 0.2 & -0.1 \\ 0.2 & 1 & 0.5 \\ -0.1 & 0.5 & 1 \end{bmatrix}$$

for a case when $m = 3$. Again, we have 1's along the diagonal since a random variable is always perfectly correlated with itself, but some non-zero correlations in the *off-diagonals*. This example also makes it easier to see the symmetry of a correlation matrix. There is a positive correlation between ϵ_{i1} and ϵ_{i2} equal to 0.2, which we can see in the elements $\boldsymbol{\Omega}_{i,12}$ and $\boldsymbol{\Omega}_{i,21}$.

1.3 Symbolic representation

It is common to use ρ_{jk} to denote the correlation between variables j and k (the errors in this case). Thus, a symbolic representation you will see is

$$\text{corr}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Omega}_i = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{12} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1m} & \rho_{2m} & \cdots & 1 \end{bmatrix}.$$

Notice that we don't switch the order of the subscripts since the correlation in row j and column k has to be equal to the correlation in row k and column j. This makes the symmetry clearer when we switch to a symbolic representation of the matrix. Similarly, if all errors within a subject have the same correlation, we would make this clear by using the notation

$$\text{corr}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Omega}_i = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$$

Here, we dropped the subscripts to make it clear that all the off-diagonal elements are the same.

2 The covariance matrix