Lecture 3 | 10.03.202

Statistical inference in a multivariate regression model

Notation overview

- balanced longitudinal profiles $\mathcal{D}_B \equiv \{(\boldsymbol{Y}_i, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in}^\top)^\top;\ i = 1, \dots, N\}$
  - for $n_i = n \in \mathbb{N}$ for all $i = 1, \dots, N$
  - random vectors $(\boldsymbol{Y}_i, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in}^\top)^\top$ are independent with the same length
  - for longitudinal data we do not assume that subject specific measurements are taken at the same time $\Rightarrow$ $\mathcal{D}_B$ generally not a random sample!
  - for multivariate regression model we already assume that the observations in $\mathcal{D}_B$ form a random sample (same error structure) $\Rightarrow$ notation $\mathcal{D}_S$
- population and data model formulation (generic vs. sample model)

$$\boldsymbol{Y} = \boldsymbol{X}^\top \mathbb{B} + \boldsymbol{\varepsilon} \qquad\qquad \mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$$

for generic random vectors $\boldsymbol{Y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^p$ and some matrix with the unknown parameters $\mathbb{B} \in \mathbb{R}^{p \times n}$

The corresponding data (i.e., random sample): $\mathbb{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N)^\top$, $\mathbb{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_N)^\top$, and $\mathbb{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_N)^\top \equiv (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N)^\top, \boldsymbol{u}_i \sim F$

Statistical inference: Likelihood ratio test

- Inference in terms of confidence intervals/regions and hypothesis tests
- General form of the null hypothesis:

$$H_0 : \mathbb{C}_1 \mathbb{B} \mathbb{M}_1 = \mathbb{D}$$

  where $\mathbb{C}_1, \mathbb{M}_1$, and $\mathbb{D}$ are some (suitable) matrices
- The rows of $\mathbb{C}_1$ do inference on the effects of independent variables while the columns of $\mathbb{M}_1$ do inference on particular linear combinations of dependent variables
- In practical applications it is common that $\mathbb{D}$ is a zero matrix (all elements are zeros) and $\mathbb{M}_1 = \mathbb{I}$ (i.e. a unit matrix with ones on the main diagonal) $\hookrightarrow$ alternatively, the model of the form $\mathbb{Y}\mathbb{M}_1 = \mathbb{X}\mathbb{B}\mathbb{M}_1 + \mathbb{U}\mathbb{M}_1$
- Thus, the null hypothesis reduces to

$$H_0 : \mathbb{C}_1 \mathbb{B} = \mathbf{0}$$

  against a general alternative hypothesis of the form $H_A : \mathbb{C}_1 \mathbb{B} \neq \mathbf{0} \in \mathbb{R}^{q \times n}$ (with the rank of the matrix $\mathbb{C}_1$ being equal to $q \in \mathbb{N}$)

Inference: Likelihood ratio test

- consider the null hypothesis of the form $H_0: \ \mathbb{C}_1\mathbb{B} = \mathbb{D}$
  (for some matrix $\mathbb{C} \in \mathbb{R}^{q \times p}$, for some $q \in \shortparallel$, such that $q < p$)
- the model $\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$ can be equivalently expressed as

$$\widetilde{\mathbb{Y}} = \mathbb{Z}\widetilde{\mathbb{B}} + \mathbb{U},$$

  for $\widetilde{\mathbb{Y}} = \mathbb{Y} - \mathbb{X}\mathbb{B}_0$, where $\mathbb{C}_1\mathbb{B}_0 = \mathbb{D}$ (satisfies the null hypothesis), $\mathbb{Z} = \mathbb{X}\mathbb{C}^{-1}$ where $\mathbb{C}^\top = (\mathbb{C}_1^\top, \mathbb{C}_2^\top)$ and $\widetilde{\mathbb{B}} = (\widetilde{\mathbb{B}}_1^\top, \widetilde{\mathbb{B}}_2^\top)^\top = \mathbb{C}(\mathbb{B} - \mathbb{B}_0)$
- the null hypothesis $\mathbb{C}_1\mathbb{B} = \mathbb{D}$ gives that $\widetilde{\mathbb{B}}_1 = \mathbf{0}$ and for the matrix partition $\mathbb{C}^{-1} = (\mathbb{C}^{(1)}, \mathbb{C}^{(2)})$ the projection matrix

$$\mathbb{P}_1 = \mathbb{I} - \mathbb{X}\mathbb{C}^{(2)}(\mathbb{C}^{(2)\top}\mathbb{X}^\top\mathbb{X}\mathbb{C}^{(2)})^{-1}\mathbb{C}^{(2)\top}\mathbb{X}^\top$$

  defines the projection onto the linear subspace orthogonal to the columns of the matrix $\mathbb{X}\mathbb{C}^{(2)}$ (i.e., residuals for the regression onto $\mathbb{C}^{(2)}$ – under the null hypothesis, thus $\widetilde{\mathbb{B}}_1 = \mathbf{0}$)

LRT: Likelihood under the null and alternative

- maximized likelihood under the null hypothesis

$$\ell_0 = |2\pi N^{-1}\widetilde{\mathbb{Y}}^\top \mathbb{P}_1 \widetilde{\mathbb{Y}}|^{-N/2} \cdot \exp\{-\frac{1}{2}Nn\}$$

- maximized likelihood under the alternative hypothesis

$$\ell_1 = |2\pi N^{-1}\widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{P}} \widetilde{\mathbb{Y}}|^{-N/2} \cdot \exp\{-\frac{1}{2}Nn\}$$

- the likelihood ratio test statistic is given as

$$\lambda^{2/N} = |\widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{P}} \widetilde{\mathbb{Y}}| / |\widetilde{\mathbb{Y}}^\top \mathbb{P}_1 \widetilde{\mathbb{Y}}| = |\widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{P}} \widetilde{\mathbb{Y}}| / |\widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{P}} \widetilde{\mathbb{Y}} + \widetilde{\mathbb{Y}}^\top \mathbb{P}_2 \widetilde{\mathbb{Y}}|$$

  and it follows the $\Lambda(n, N-p, q)$ distribution, where $q \in \mathbb{N}$ is the number of rows in $\mathbb{C}_1$ (for $\mathbb{P}_2 = \mathbb{P}_1 - \widetilde{\mathbb{P}}$ – what does it mean geometrically?)

Examples

- Repeated measurements for two groups (two-sample problems):

$$\boldsymbol{Y}_i^{(1)} \sim N_n(\boldsymbol{\mu}_1, \Sigma), \qquad i = 1, \ldots, N_1$$
$$\boldsymbol{Y}_i^{(2)} \sim N_n(\boldsymbol{\mu}_2, \Sigma), \qquad i = 1, \ldots, N_2$$

- Typical testing problems
  - parallel profiles of two groups $H_0 : \mathbb{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$
  - identical profiles for both groups $H_0 : \mathbf{1}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 0$
  - treatment effect $H_0 : \mathbb{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$

- Multiple testing problem
  The statistical test for identical profiles only makes sense if the profiles are parallel; Similarly, if the profiles are parallel, is there any treatment effect at all?

Some useful overview

- statistical tests about some (multivariate) mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ can be often expressed in terms of the null hypothesis $H_0: \ \mathbb{A}\boldsymbol{\mu} = \boldsymbol{a}$ against a general alternative $H_A: \ \mathbb{A}\boldsymbol{\mu} \neq \boldsymbol{a}$ where $\mathbb{A} \in \mathbb{R}^{q \times n}$ and $\boldsymbol{a} \in \mathbb{R}^q$

- for a random sample $\boldsymbol{X}_i \sim N_n(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \dots, N$, with $\Sigma \in \mathbb{R}^{n \times n}$ known, it holds that $\mathbb{A}\overline{\boldsymbol{X}}_N \sim N_n(\mathbb{A}\boldsymbol{\mu}, \frac{1}{N}\mathbb{A}\Sigma\mathbb{A}^\top)$ and the corresponding log-likelihood based test statistic equals $-2\log\lambda = N(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})^\top(\mathbb{A}\Sigma\mathbb{A}^\top)^{-1}(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})$ and it follows (exactly) the χ^2 distribution with $q \in \mathbb{N}$ degrees of freedom

- for $\boldsymbol{X}_i \sim N_n(\boldsymbol{\mu}, \Sigma)$ with $\Sigma \in \mathbb{R}^{n \times n}$ unknown, the log-likelihood test statistic equals to $-2\log\lambda = N\log\left\{1 + (\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})^\top(\mathbb{A}\widehat{\Sigma}_N\mathbb{A}^\top)^{-1}(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})\right\}$ and it follows asymptotically the χ^2 distribution with $q \in \mathbb{N}$ degrees of freedom where $\widehat{\Sigma}_N$ is the sample estimate of the variance-covariance matrix Σ defined as

$$\widehat{\Sigma}_N = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)^\top$$

- alternatively, the estimate $\widehat{\Sigma}_N$ can be also expressed (using the whole data matrix $\mathbb{X} \in \mathbb{R}^{N \times n}$) as $\widehat{\Sigma}_N = \frac{1}{N}\mathbb{X}^\top\mathcal{H}\mathbb{X}$, where $\mathcal{H} \in \mathbb{R}^{N \times N}$ is the so-called centering matrix $\mathcal{H} = \mathbb{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, for the unit matrix $\mathbb{I} \in \mathbb{R}^{N \times N}$ and $\mathbf{1}^\top = (1, \dots, 1) \in \mathbb{R}^N$

Wishart and Hotelling distributions

- under the normality assumption $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \sim N_n(\boldsymbol{\mu}, \Sigma)$ with Σ unknown there also exists an exact test (i.e., a multivariate generalization of the t-test)
- **Wishart distribution** (distribution of random matrices)
  - for $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \sim N_n(\boldsymbol{0}, \Sigma)$ and $\mathbb{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)^\top \in \mathbb{R}^{N \times n}$, it holds that

$$\mathcal{S} = \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i^\top = \mathbb{X}^\top \mathbb{X} \sim W_n(\Sigma, N)$$

    where $\mathcal{S}$ is called the scatter matrix and $W_p(\Sigma, N)$ is the n-dimensional Wishart distribution with N degrees of freedom and the scale matrix Σ (note, that $W_1(1, N)$ is equivalent to the χ^2-distribution with N degrees of freedom)
  - for $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \sim N_n(\boldsymbol{\mu}, \Sigma)$ if can be shown that $\mathbb{X}^\top \mathcal{H} \mathbb{X} \sim W_n(\Sigma, N-1)$
  - if $\mathcal{S} \sim W_n(\Sigma, N)$, $\mathbb{C} \in \mathbb{R}^{q \times n}$ (rank $q \in \mathbb{N}$) $\Longrightarrow \mathbb{C}\mathcal{S}\mathbb{C}^\top \sim W_q(\mathbb{C}\mathcal{S}\mathbb{C}^\top, N)$
- **Hotelling distribution** (generalization of the univariate t-test)
  - for $\boldsymbol{X} \sim N_n(\boldsymbol{0}, \mathbb{I})$ and $\mathcal{S} \sim W_n(\mathbb{I}, N)$ with $\mathbb{I} \in \mathbb{R}^{n \times n}$ being a unit matrix and $\boldsymbol{X} \perp \mathcal{S}$, it holds that

$$N \boldsymbol{X}^\top \mathcal{S}^{-1} \boldsymbol{X} \sim T^2(n, N),$$

    which is the Hotelling's T^2 distribution with parameters $n \in \mathbb{N}$ and $N \in \mathbb{N}$
  - for $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ and $\mathcal{S} \sim W_n(\Sigma, N)$ (Σ with full rank) such that $\boldsymbol{X} \perp \mathcal{S}$, it holds that $\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim N_n(\boldsymbol{0}, \mathbb{I})$ and $\Sigma^{-1/2}\mathcal{S}\Sigma^{-1/2} \sim W_n(\mathbb{I}, N)$ and, therefore, also $N(\boldsymbol{X} - \boldsymbol{\mu})^\top \mathcal{S}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim T^2(n, N)$

Hotelling's T^2 test

- for $\boldsymbol{X}_i \sim N_n(\boldsymbol{\mu}, \Sigma)$ with $\Sigma \in \mathbb{R}^{n\times n}$ unknown, the log-likelihood test statistic equals (as already stated before) to $-2\log\lambda = N\log\left\{1 + (\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})^\top(\mathbb{A}\widehat{\Sigma}_N\mathbb{A}^\top)^{-1}(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})\right\}$ and it follows asymptotically the χ^2 distribution with $q \in \mathbb{N}$ degrees of freedom

- it can be also easily proved that all of the following hold
  - $\overline{\boldsymbol{X}}_N \sim N_n(\boldsymbol{\mu}, \frac{1}{N}\Sigma)$ and also $(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a}) \sim N_q(\mathbb{A}\boldsymbol{\mu} - \boldsymbol{a}, \frac{1}{N}\mathbb{A}\Sigma\mathbb{A}^\top)$
  - under the hull hypothesis it holds $(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a}) \sim N_n(\boldsymbol{0}, \frac{1}{N}\mathbb{A}\Sigma\mathbb{A}^\top)$
  - sample covariance matrix can be obtained as $\widehat{\Sigma}_N = \frac{1}{N}\mathbb{X}^\top\mathcal{H}\mathbb{X}$ and it holds that $N\widehat{\Sigma}_N = \mathbb{X}^\top\mathcal{H}\mathbb{X} \sim W_n(\Sigma, N-1)$ and, moreover, $\overline{\boldsymbol{X}}_N \perp \widehat{\Sigma}_N$
  - thus, for $\mathbb{A}\widehat{\Sigma}_N\mathbb{A}^\top$, it holds that $\mathbb{A}\widehat{\Sigma}_N\mathbb{A}^\top \sim W_q(\Sigma, N-1)$

- bringing now everything together, the test statistic defined as

$$\mathcal{T} = (N-1)(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})^\top(\mathbb{A}\widehat{\Sigma}_N\mathbb{A}^\top)^{-1}(\mathbb{A}\overline{\boldsymbol{X}}_N - \boldsymbol{a})$$

  follows (under the null hypothesis $H_0 : \mathbb{A}\boldsymbol{\mu} = \boldsymbol{a}$) exactly the Hotelling's T^2 distribution with parameters n and $N-1$ (i.e., $\mathcal{T} \sim T^2(n, N-1)$)

- it also holds, that if $\mathcal{T} \sim T^2(n, N)$ then $\frac{N-n+1}{Nn}\mathcal{T} \sim F_{n,N-n+1}$ (where $F_{\cdot,\cdot}$ denotes the F distribution with the corresponding degrees of freedom)

Two sample problems:

Returning now back to the statistical tests for longitudinal profiles, the following test statistics (with the exact Hotelling T^2 distribution under the null hypothesis) are obtained:

- **Parallel profiles**

$$T = \frac{N_1 N_2}{(N_1+N_2)^2}(N_1+N_2-2)\Big[\mathbb{C}(\overline{\boldsymbol{Y}}^{(1)} - \overline{\boldsymbol{Y}}^{(2)})\Big]^\top \Big(\mathbb{C}\widehat{\boldsymbol{\Sigma}}\mathbb{C}^\top\Big)^{-1}\Big[\mathbb{C}(\overline{\boldsymbol{Y}}^{(1)} - \overline{\boldsymbol{Y}}^{(2)})\Big]$$

  where $\widehat{\Sigma} = \frac{1}{N_1+N_2}\Big[\mathbb{Y}_1^\top \mathcal{H}\mathbb{Y}_1 + \mathbb{Y}_2^\top \mathcal{H}\mathbb{Y}_2\Big]$ (i.e., pooled covariance estimate) such that $(N_1+N_2)\widehat{\Sigma} \sim W_n(\Sigma, N_1+N_2-2)$ and $\mathbb{Y}_1 = (Y_{i,j}^{(1)})_{i,j=1}^{N_1,n}$ and $\mathbb{Y}_2 = (Y_{i,j}^{(2)})_{i,j=1}^{N_2,n}$ and the centering matrix $\mathcal{H}$ with proper dimensions
  $\Longrightarrow$ under the null hypothesis, it holds that, $T \sim T^2(n-1, N_1+N_2-2)$

- **Equality of two levels**

$$T = \frac{N_1 N_2}{(N_1+N_2)^2}(N_1+N_2-2)\frac{\Big[\mathbf{1}^\top(\overline{\boldsymbol{Y}}^{(1)} - \overline{\boldsymbol{Y}}^{(2)})\Big]^2}{\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}\mathbf{1}}$$

  and (under the null hypothesis) $T \sim T^2(1, N_1+N_2-2)$

- **Same treatment effect**

$$T = (N_1+N_2-2)(\mathbb{C}\overline{\boldsymbol{Y}})^\top\Big(\mathbb{C}\widehat{\boldsymbol{\Sigma}}\mathbb{C}^\top\Big)^{-1}\mathbb{C}\overline{\boldsymbol{Y}}, \quad \text{for } \overline{\boldsymbol{Y}} = \frac{N_1\overline{\boldsymbol{Y}}^{(1)} + N_2\overline{\boldsymbol{Y}}^{(2)}}{N_1+N_2}$$

  and (under the null hypothesis) $T \sim T^2(n-1, N_1+N_2-2)$

Multivariate model vs. general linear model

- Multivariate regression model $\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$ (wide data format)
  - $n \in \mathbb{N}$ repeated measurements within $N \in \mathbb{N}$ subjects (random sample)
  - repeated measurements taken at the same time-points across subjects
  - time evolution modeled by the set of $\beta_j \in \mathbb{R}^p$ parameters ($j = 1, \dots, n$)
  - the vector of subject's specific covariates $\boldsymbol{X}_i \in \mathbb{R}^p$ fixed over time
  - covariance structure modeled by the matrix Σ, where $\boldsymbol{u}_i \sim N_n(\mathbf{0}, \Sigma)$
  - the data usually form a random sample from the joint distribution $F_{\boldsymbol{Y},\boldsymbol{X}}$

- General linear model for correlated errors $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (long data format)
  - $n \in \mathbb{N}$ repeated measurements within $N \in \mathbb{N}$ subjects (balanced data)
  - the vector of unknown parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ is fixed over time
  - subject's specific covariates $\boldsymbol{X}_{ij} \in \mathbb{R}^p$ may vary with $j \in \{1, \dots, n\}$
  - subjects' independence and within subject's covariance modeled by the variance covariance Σ, where $\varepsilon \sim N(\mathbf{0}, \Sigma)$ (overall dimensionality: Nn)
  - the model can be further generalized for unbalanced data (with $n_i \in \mathbb{N}$)

General linear model with correlated errors

- instead of time-varying β_j and fixed $\boldsymbol{X}_j \in \mathbb{R}^p$ the time evolution can be modeled in terms of the time-varying covariates $\boldsymbol{X}_{ij} \in \mathbb{R}^p$ and fixed $\boldsymbol{\beta} \in \mathbb{R}^p$
  - **Simplification** in terms of the vectors of unknown parameters $\boldsymbol{\beta}_j \in \mathbb{R}^p$ for $j = 1, \ldots, n$ (in the matrix $\mathbb{B} \in \mathbb{R}^{p \times n}$): $\Rightarrow \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_n$
  - **Relaxation** in terms of the subject's specific covariates $\boldsymbol{X}_{ij} \in \mathbb{R}^p$ that are now allowed to change with $j \in \{1, \ldots, n\}$: $\Rightarrow \boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^\top \in \mathbb{R}^p$
- this allows for an alternative formulation of the multivariate (data) model (where $\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$ follows as a special case) in a form

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ \vdots \\ Y_{Nn} \end{pmatrix} = \begin{pmatrix} X_{111} & \ldots & X_{11p} \\ \vdots & \ddots & \vdots \\ X_{1n1} & \ldots & X_{1np} \\ X_{211} & \ldots & X_{21p} \\ \vdots & \ddots & \vdots \\ X_{Nn1} & \ldots & X_{Nnp} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{Nn} \end{pmatrix}$$

- **What are the advantages/disadvantages of both model formulations?**

Matrix formulation

- typical notation (under the multivariate normal distributional assumption) takes the form
$$\boldsymbol{Y} \sim N_{Nn}(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{V}),$$
where $\mathbb{V}$ is a block-diagonal matrix with non-zero blocks of size $n \times n$ (each block $\sigma^2\mathbb{V}_0 \equiv \Sigma$ represents the variance-covariance structure of the repeated measurements within one single subject)

- the variance-covariance matrix $\sigma^2\mathbb{V}$ is estimated by borrowing power across subjects (i.e., replication of $\sigma^2\mathbb{V}_0$ across independent units)

- there can be no specific (parametric) structure assumed for $\mathbb{V}_0$ but it is common to postulate some parametric structure of the matrix $\mathbb{V}_0$

- the estimation of the underlying correlation structure within $\sigma^2\mathbb{V}$ is crucial for a proper statistical inference

Uniform correlation model

- **Assumption:** positive correlation $\rho \in (0,1)$ between any two repeated observations within a given subject
- **Matrix notation:** $\mathbb{V}_0 = (1-\rho)\mathbb{I}_{n\times n} + \rho\mathbf{1}_{n\times n}$
- **Motivation:** the response (random) variable Y_{ij} can be decomposed as

$$Y_{ij} = \mu_{ij} + Z_i + V_{ij},$$

  where $\mu_{ij} = EY_{ij}$ and $Z_i \sim N(0,\nu^2)$ independent of $V_{ij} \sim N(0,\tau^2)$ and it holds that $\rho = \nu^2/(\nu^2+\tau^2)$ and $\sigma^2 = \nu^2 + \tau^2$ (for $\varepsilon_{ij} = Z_i + V_{ij}$)
- **Interpretation:** linear model for the mean of the response with a random intercept (with the variance between subjects $\nu^2 > 0$)

Exponential correlation model

- **Assumption:** covariance between Y_{ij} and Y_{ik} for $i \neq k$ is of the form

$$v_{jk} = \sigma^2 \exp\{-\phi|t_j - t_k|\}$$

  and it decays towards zero as the time separation between repeated observations increases (with the rate of decay given by $\phi > 0$)
- **Matrix notation:** $\mathbb{V}_0 = (v_{jk})_{j,k=1}^{n}$
- **Motivation:** the response (random) variable Y_{ij} can be decomposed as

$$Y_{ij} = \mu_{ij} + W_{ij},$$

  where $W_{ij} = \rho W_{i(j-1)} + Z_{ij}$ for $Z_{ij} \sim N(0, \sigma^2(1 - \rho^2))$ independent (verify, that it holds that $VarY_{ij} = VarW_{ij} = \sigma^2$)
- **Interpretation:** linear model for the mean of the response with with the first order autoregressive correlation structure
- **Generalization:** $Y_{ij} = \mu_{ij} + W_i(t_j)$ for continuous time Gaussian processes $\{X_i(t);\ t \in \mathbb{R}\}$ independent for $i = 1, \dots, N$ and general time points $t_1 < \cdots < t_{ni}$

Towards least squares – two step estimation

- For simplification assume the model $Y_{ij} = a + bX_{ij} + \varepsilon_{ij}$ with no distributional assumption for the error vector $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \dots, \varepsilon_{Nn})^\top$
- Two-stage least squares as a simple estimation method for models where it is needed to distinguish the variance sources (within/between subjects)
- **Stage 1:** longitudinal profiles for each subject $i \in \{1, \dots, N\}$ individually

$$Y_{ij} = A_i + B_i X_{ij} + W_{ij}, \quad j = 1, \dots, n, \quad \text{and } W_{ij} \sim (0, \tau^2), \; i.i.d.$$

  to obtain $\widehat{A}_i = A_i + Z_{ai}$ and $\widehat{B}_i = B_i + Z_{bi}$, for $Z_{ai} \sim (0, v_{ai}^2)$, $Z_{bi} \sim (0, v_{bi}^2)$

- **Stage 2:** OLS analysis of the subject's specific parameter estimates

$$A_i = a + \delta_{ai} \qquad \text{and} \qquad B_i = b + \delta_{bi}$$

  for independent errors $\delta_{ai} \sim (0, \sigma_a^2)$ and $\delta_{bi} \sim (0, \sigma_b^2)$

- **Therefore:** $\widehat{A}_i = a + (\delta_{ai} + Z_{ai})$ and $\widehat{B}_i = b + (\delta_{bi} + Z_{bi})$

Two stage regression model

- both stages can be straightforwardly combined together as

$$\begin{aligned}
Y_{ij} &= A_i + B_1 + W_{ij} \\
&= (a + \delta_i) + (b + \delta_{bi})X_{ij} + W_{ij} \\
&= a + bX_{ij} + \delta_{ai} + \delta_{bi}X_{ij} + W_{ij}\, \underbrace{(\delta_{ai} + \delta_{bi}X_{ij} + W_{ij})}_{\varepsilon_{ij}} \\
&= a + bX_{ij} + \varepsilon_{ij}
\end{aligned}$$

- What is the variance of the of Y_{ij}?
- What is the covariance of two observations Y_{ij} and Y_{ik}, for $j \neq k$?
- What is the covariance of Y_{ij} and Y_{lk}, for $i \neq l$ and $j \neq k$?

Summary

- Two alternative but not equivalent multivariate model formulations

$$\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U} \qquad \text{versus} \qquad \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Estimation of the unknown parameters in $\mathbb{B} \in \mathbb{R}^{p \times n}$ or $\boldsymbol{\beta} \in \mathbb{R}^{p}$
(either in terms of the least squares or the maximum likelihood estimation)

- Decomposition of the overall data variability into two different sources
(the within subject's variability and the between subjects' variability)

- Marginal or hierarchical inference (in terms of the confidence intervals/regions or the statistical tests)

- Two stage estimation approach in the model $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\beta}$
(towards the mixed effect model with fixed and random effects)