A is idempotent -> $A^2 = A$
SVD
V = Eigenvector
$L = diag(l1, l2,\dots,ln)$

-> $VLV' = (V*L*V') * (V*L*V')$

$(V*L*V') * (V*L*V') = V * L * (V' V) * L * V'$
$(V' V)$ = Identity matrix

$V * L * (V' V) * L * V'$
=> $V * L * L * V'$
=> $V * L^2 * V'$

$V * L^2 * V' = V * L * V'$ l(remove V and V’ by multiplying V and V’ bc $V*V'$ = Identity matrix)

$L^2 = L$
$L^2 - L = 0$
=> $L(L - \text{Identitymatrix})$

Applied Data Analysis (ADA)

Erhard Cramer, Maria Kateri

Institute of Statistics
Pontdriesch 14-16

erhard.cramer@rwth-aachen.de, maria.kateri@rwth-aachen.de
www.isw.rwth-aachen.de

A preliminary note

- **Please read carefully the slides on the course concept uploaded in RWTHmoodle!**
- The lecture is split into two parts:
  - Part I: **Linear Models** (Cramer)
    Lectures from April 13 to May 19
    Tutorials: April 23, May 7, 21
  - Break: May 24–28 (Pentecost week)
  - Part II: **Generalized Linear Models** (Kateri)
    Lectures from June 1 to July 23
    Tutorials: June 18, July 2, 16
- Parts I & II will be held as distance teaching course.

Just to let you know who is talking to you on linear models...

Prof. Dr. Erhard Cramer

- Institute of Statistics
- Pontdriesch 14-16
- Office: Room no. 313, third floor
- erhard.cramer@rwth-aachen.de

In the first part, we consider...

a class of statistical models, so called **linear models**, generated by the equation

$$\mathbf{Y} = \mathrm{B}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with

- $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ vector of observations,
- B design matrix,
- $\boldsymbol{\beta}$ parameter vector
- $\boldsymbol{\varepsilon}$ (random) error term (not observable).

Part I: Linear Models

Chapter I.1

Preliminaries

Notation, Linear Algebra & Probability

Topics

- **To be discussed...**
  - properties of vectors & matrices, rank, trace, singular value decomposition, etc.
  - Moore-Penrose general inverse
  - Image, kernel, orthogonal projectors
  - Random vectors, expectations, covariance matrix
  - selected probability distributions on the real line connected to the normal distribution
  - non-central χ^2- and F-distribution

Part I: Linear Models

Chapter I.1

Preliminaries

Linear Algebra

Notation & basic definitions

I.1.1 Notation (vectors and matrices)

- $\mathbb{R}^p$: p-dimensional Euclidean space
- $\mathbb{R}^{p\times q}$: set of all $(p \times q)$-matrices
- vectors are written in bold italics: $\boldsymbol{x} = (x_i)_{1\leqslant i\leqslant p} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$
- random vectors are written in capital bold italics: $\boldsymbol{X} = (X_i)_{1\leqslant i\leqslant p} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$
- matrices are written in capitals: $A = (a_{ij})_{1\leqslant i\leqslant p, 1\leqslant j\leqslant q} = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ldots & \vdots \\ a_{p1} & \cdots & a_{pq} \end{pmatrix}$.

Notation & basic definitions

I.1.2 Notation (special vectors and matrices)

- $A = \text{diag}(a_1, \ldots, a_p)$: diagonal matrix with diagonal elements $a_1, \ldots, a_p$
- $\mathbb{1}_p \in \mathbb{R}^p$: vector of ones, $0 \in \mathbb{R}^p$ zero vector
- $e_{1,p}, \ldots, e_{p,p}$: standard basis of $\mathbb{R}^p$
- $I_p = \text{diag}(1, \ldots, 1)$: p-dimensional identity matrix
- $\mathbb{1}_{p\times p} = \mathbb{1}_p \mathbb{1}_p'$: matrix of ones
- $E_p = I_p - \frac{1}{p}\mathbb{1}_{p\times p}$: ortho-projection matrix

- $\|x\| = \sqrt{\sum_{i=1}^p x_i^2}$ denotes the (Euclidean) norm of a vector $x \in \mathbb{R}^p$.
- $\text{rank}(A)$ denotes the rank of a matrix A.
- $\det(A)$ denotes the determinant of a squared matrix A.
- $\text{trace}(A)$ denotes the trace of a squared matrix $A \in \mathbb{R}^{p\times p}$, i.e., $\text{trace}(A) = \sum_{i=1}^p a_{ii}$

- The transpose of a matrix A is denoted by A'.
- The inverse of a matrix $A \in \mathbb{R}^{p\times p}$ is denoted by A^{-1} (provided it exists), i.e., $AA^{-1} = A^{-1}A = I_p$.

Notation & basic definitions

I.1.3 Definition

- A matrix $A \in \mathbb{R}^{p \times p}$ is called symmetric if $A = A'$.
- A matrix $A \in \mathbb{R}^{p \times p}$ is called an orthogonal matrix if $AA' = A'A = I_p$.
- A matrix $A \in \mathbb{R}^{p \times p}$ is called positive (non-negative) definite if $A = A'$ and

$$\boldsymbol{x}' A \boldsymbol{x} > (\geqslant) 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^p \setminus \{\boldsymbol{0}\}.$$

  For short, we write $A > 0$ or $A \geqslant 0$, respectively.
- $\mathbb{R}^{p \times p}_{>0}$: set of all positive definite $(p \times p)$-matrices
- $\mathbb{R}^{p \times p}_{\geqslant 0}$: set of all non-negative definite $(p \times p)$-matrices

Some linear algebra

I.1.4 Lemma

Let $A, C \in \mathbb{R}^{p\times p}$ with $\det(AC) \neq 0$ and $B \in \mathbb{R}^{k\times p}$, $1 \leqslant k \leqslant p$. Then:

1. $(C')^{-1} = (C^{-1})'$
2. $(AC)' = C'A'$
3. $(AC)^{-1} = C^{-1}A^{-1}$
4. $\text{rank}(B') = \text{rank}(B)$
5. $\text{rank}(BC) = \text{rank}(B)$
6. For $D \in \mathbb{R}^{p\times k}$, we have $\text{rank}(BD) = \text{rank}(DB)$.
7. $\text{rank}(BB') = \text{rank}(B'B) = \text{rank}(B)$
8. $\det(AC) = \det(CA) = \det(A) \cdot \det(C)$ for all $A, C \in \mathbb{R}^{p\times p}$
9. $\text{trace}(A + C) = \text{trace}(A) + \det(C)$
10. $\text{trace}(BD) = \text{trace}(DB)$ for all $D \in \mathbb{R}^{p\times k}$

Singular Value Decomposition (SVD)

I.1.5 Theorem

Let $\Sigma \in \mathbb{R}^{p\times p}_{\geqslant 0}$. Then, the **singular value decomposition** (eigen decomposition) of Σ is given by

$$\Sigma = V\Lambda V',$$ V' = V^(-1)

where $\lambda_1 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ denote the eigenvalues and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ the corresponding (orthonormal) eigenvectors of Σ. Further, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ and $V = [\boldsymbol{v}_1 \mid \cdots \mid \boldsymbol{v}_p]$ with $V'V = VV' = I_p$.

Furthermore, with the definitions $\Lambda^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})$ and $\Sigma^{1/2} = V\Lambda^{1/2}V'$, we have

- $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ and $(\Sigma^{1/2})' = \Sigma^{1/2}$.
- $\Sigma^{1/2}$ is non-negative definite.

  ☞ $\Sigma^{1/2}$ is called the **root of Σ**.
- Notice that, for a regular matrix Σ, $(\Sigma^{1/2})^{-1} = \Sigma^{-1/2}$ where $\Sigma^{-1/2} = V\Lambda^{-1/2}V'$ and $\Lambda^{-1/2} = \operatorname{diag}(\sqrt{\lambda_1^{-1}}, \ldots, \sqrt{\lambda_p^{-1}})$.

Moore-Penrose (general) inverse of a matrix

I.1.6 Theorem

The Moore-Penrose (general) inverse of a matrix A is denoted by A^+, i.e., A^+ is the unique matrix satisfying the four equations:

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad (AA^+)' = AA^+, \quad (A^+A)' = A^+A.$$

It has the following properties:

- $(A^+)^+ = A$
- $(A^+)' = (A')^+$
- $A^+ = A^+(A^+)'A'$
- $A = AA'(A^+)'$
- $A' = A'AA^+$
- $A' = A^+AA'$
- $(AA')^+ = (A')^+A^+$
- $A^+ = (A'A)^+A'$
- $A^+ = A'(AA')^+$

- If $A \in \mathbb{R}^{p \times q}$ exhibits the SVD $A = U\Lambda V'$, then A^+ has the SVD $A^+ = V\Lambda^+U'$ where Λ^+ is the Moore-Penrose inverse of the matrix Λ.
- If $A \in \mathbb{R}^{p \times p}$ is a regular matrix then $A^+ = A^{-1}$.

Image, Kernel, Orthogonal Projectors

I.1.7 Definition

- For a matrix $A \in \mathbb{R}^{p \times q}$, let
  - $\text{Ker}(A) = \{\mathbf{x} \in \mathbb{R}^q \mid A\mathbf{x} = \mathbf{0}\}$ be the kernel (null space) of A.
  - $\text{Im}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^q\}$ be the image of A.
- For a linear subspace $\mathscr{A} \subseteq \mathbb{R}^p$, $\mathscr{A}^\perp = \{\mathbf{y} \in \mathbb{R}^p \mid \mathbf{x}'\mathbf{y} = 0 \text{ for all } \mathbf{x} \in \mathscr{A}\}$ denotes the corresponding orthogonal space.
- Let $\mathscr{A}, \mathscr{B} \subseteq \mathbb{R}^p$ be linear subspaces with $\mathscr{A} \cap \mathscr{B} = \{\mathbf{0}\}$. Then, $\mathscr{A} \oplus \mathscr{B} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathscr{A}, \mathbf{y} \in \mathscr{B}\}$ is called the direct sum of $\mathscr{A}, \mathscr{B}$.

Notice that $\text{Ker}(A) \subseteq \mathbb{R}^q$ and $\text{Im}(A) \subseteq \mathbb{R}^p$ are linear subspaces.

I.1.8 Definition

A matrix $Q \in \mathbb{R}^{p \times p}$ is called

- idempotent if $Q^2 = Q$
- orthogonal projector on a linear subspace $\mathscr{A} \subseteq \mathbb{R}^p$ if
  - for any $\mathbf{x} \in \mathscr{A}$: $Q\mathbf{x} = \mathbf{x}$
  - for any $\mathbf{y} \in \mathscr{A}^\perp$: $Q\mathbf{y} = \mathbf{0}$

Properties of Orthogonal Projectors and Moore-Penrose inverse

I.1.9 Lemma

- Orthogonal projectors on a linear subspace $\{0\} \neq \mathscr{A} \subseteq \mathbb{R}^p$ are unique.
- $Q \in \mathbb{R}^{p\times p}$ is an orthogonal projector (on $\mathrm{Im}(Q)$) iff $Q^2 = Q$ and $Q' = Q$.

I.1.10 Theorem

Let $A \in \mathbb{R}^{p\times q}$ with Moore-Penrose inverse A^+ and define $P_1 = I_q - A^+A$, $P_2 = I_p - AA^+$. Then:

- P_1 and P_2 are orthogonal projectors, respectively, that is, $P_i^2 = P_i$, $P_i' = P_i$, $i = 1, 2$.
- $Q = AA^+ = A(A'A)^+A'$ is the (unique) orthogonal projector on $\mathrm{Im}(A)$.
- $A^+A = A'(AA')^+A$ is the (unique) orthogonal projector on $\mathrm{Im}(A')$.
- $\mathrm{Ker}(A) = \mathrm{Im}(P_1)$, $\mathrm{Im}(A) = \mathrm{Ker}(P_2)$.
- $\mathrm{Ker}(A^+) = \mathrm{Im}(P_2)$, $\mathrm{Im}(A^+) = \mathrm{Ker}(P_1)$
- $\mathrm{Im}(A) = \mathrm{Ker}(A^+)^\perp$, $\mathrm{Im}(A^+) = \mathrm{Ker}(A)^\perp$,
- $\mathrm{Ker}(A) \oplus \mathrm{Im}(A^+) = \mathbb{R}^q$, $\mathrm{Ker}(A^+) \oplus \mathrm{Im}(A) = \mathbb{R}^p$

Part I: Linear Models

Chapter I.1

Preliminaries

Probability

Expectations of random vectors and random matrices

I.1.11 Definition (expectation of random vectors and random matrices)

1. The **expectation of a random vector** $\mathbf{X} = (X_1, \ldots, X_p)'$ is defined by the vector of means, that is,

$$\mathrm{E}\mathbf{X} = \begin{pmatrix} \mathrm{E}X_1 \\ \vdots \\ \mathrm{E}X_p \end{pmatrix};$$

subsequently, we use the notation $\boldsymbol{\mu} = \mathrm{E}\mathbf{X}$;

2. The **expectation of a random matrix** $\mathscr{X} = (X_{ij})_{1 \leqslant i \leqslant p, 1 \leqslant j \leqslant q}$ is defined by the matrix of means, that is,

$$\mathrm{E}\mathscr{X} = \begin{pmatrix} \mathrm{E}X_{11} & \cdots & \mathrm{E}X_{1q} \\ \vdots & \ddots & \vdots \\ \mathrm{E}X_{p1} & \cdots & \mathrm{E}X_{pq} \end{pmatrix}.$$

In the following, all expectations are supposed to exist.

I.1.12 Lemma

1. Let $\mathbf{X} = (X_1, \ldots, X_p)'$ be a p-dimensional random vector and $A \in \mathbb{R}^{k \times p}, \mathbf{b} \in \mathbb{R}^k$. Then:

$$\mathsf{E}(A\mathbf{X} + \mathbf{b}) = A\mathsf{E}(\mathbf{X}) + \mathbf{b}\,.$$

2. Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be p-dimensional random vectors and $A_1, \ldots, A_n \in \mathbb{R}^{k \times p}$. Then:

$$\mathsf{E}\Big(\sum_{j=1}^{n} A_j \mathbf{Z}_j\Big) = \sum_{j=1}^{n} A_j \mathsf{E}(\mathbf{Z}_j) \in \mathbb{R}^k\,.$$

I.1.13 Definition (variance-covariance matrix)

Let $\mathbf{X} = (X_1, \ldots, X_p)'$, $\mathbf{Y} = (Y_1, \ldots, Y_q)'$ be random vectors. Then, the **covariance matrix** of $\mathbf{X}$ and $\mathbf{Y}$ is defined by

$$\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} \operatorname{Cov}(X_1, Y_1) & \cdots & \operatorname{Cov}(X_1, Y_q) \\ \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_p, Y_1) & \cdots & \operatorname{Cov}(X_p, Y_q) \end{pmatrix}.$$

The **variance-covariance matrix** of $\mathbf{X}$ is defined by $\Sigma = \operatorname{Cov}(\mathbf{X}) = \operatorname{Cov}(\mathbf{X}, \mathbf{X})$.

I.1.14 Remark

Defining the random matrix $\mathscr{C}_{X,Y} = \big(\mathbf{X} - E(\mathbf{X})\big)\big(\mathbf{Y} - E(\mathbf{Y})\big)'$, we get

1. $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = E\mathscr{C}_{X,Y}$
2. $\operatorname{Cov}(\mathbf{X}) = \operatorname{Cov}(\mathbf{X}, \mathbf{X}) = E\mathscr{C}_{X,X}$
3. $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}\mathbf{Y}') - E\mathbf{X} \cdot E\mathbf{Y}'$
4. $\operatorname{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}') - E\mathbf{X} \cdot E\mathbf{X}'$

Covariance matrices are always non-negative definite, that is, $\operatorname{Cov}(\mathbf{X}) \geqslant 0$.

I.1.15 Notation (block matrix)

A matrix $A \in \mathbb{R}^{(p+q)\times(k+r)}$ can be written as a **block matrix**

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{with } A_{11} \in \mathbb{R}^{p\times k}, A_{12} \in \mathbb{R}^{p\times r}, A_{21} \in \mathbb{R}^{q\times k}, A_{22} \in \mathbb{R}^{q\times r}.$$

I.1.16 Lemma

With the notation from Definition I.1.13, we get for $A \in \mathbb{R}^{k\times p}, B \in \mathbb{R}^{r\times q}, \mathbf{b} \in \mathbb{R}^{k}, \mathbf{c} \in \mathbb{R}^{r}$:

1. $\mathrm{Cov}\,(A\mathbf{X} + \mathbf{b}, B\mathbf{Y} + \mathbf{c}) = A\mathrm{Cov}\,(\mathbf{X}, \mathbf{Y})B'$,
2. $\mathrm{Cov}\,(A\mathbf{X} + \mathbf{b}) = A\mathrm{Cov}\,(\mathbf{X})A'$,
3. $\mathrm{Cov}\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{bmatrix} \mathrm{Cov}\,(\mathbf{X}) & \mathrm{Cov}\,(\mathbf{X}, \mathbf{Y}) \\ \mathrm{Cov}\,(\mathbf{Y}, \mathbf{X}) & \mathrm{Cov}\,(\mathbf{Y}) \end{bmatrix}$,
4. $\mathrm{Cov}\,(\mathbf{X}, \mathbf{Y}) = \mathrm{Cov}\,(\mathbf{Y}, \mathbf{X})'$.

Using Lemma I.1.16, we can write with $\Sigma_{XY} = \mathrm{Cov}\,(\mathbf{X}, \mathbf{Y})$:

$$\Sigma_{\binom{X}{Y}} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \overset{(*)}{=} \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma'_{XY} & \Sigma_{YY} \end{bmatrix}.$$

Probability distributions on $\mathbb{R}$

I.1.17 Remark (density functions of distributions on $\mathbb{R}$)

- Normal distribution $N(\mu, \sigma^2)$:

$$f(x) = \varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$$

- χ^2-distribution $\chi^2(p)$ with $p \in \mathbb{N}$ degrees of freedom:

$$f(x) = \frac{1}{2^{p/2}\Gamma(p/2)}\, x^{p/2-1}\, e^{-x/2}, \quad x > 0$$

- t-distribution $t(p)$ with $p \in \mathbb{N}$ degrees of freedom:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{p\pi}\,\Gamma\left(\frac{p}{2}\right)} \left(1 + \frac{x^2}{p}\right)^{-(p+1)/2}, \quad x \in \mathbb{R}$$

- F-distribution $F(p, q)$ with $p \in \mathbb{N}$ numerator and $q \in \mathbb{N}$ denominator degrees of freedom:

$$f(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{n/2} \frac{x^{p/2-1}}{\left(1+\frac{p}{q}\,x\right)^{\frac{p+q}{2}}}, \quad x > 0$$

Connections of probability distributions

I.1.18 Notation

The notation $X_1, \ldots, X_k \overset{\text{iid}}{\sim} P$ means that the random variables $X_1, \ldots, X_k$ are independent and identically distributed (iid) with $X_1 \sim P$.

The same notation is used for samples of random vectors.

I.1.19 Proposition

1. Let $X \sim N(0,1)$ and $\mu \in \mathbb{R}, \sigma > 0$. Then, $\mu + \sigma X \sim N(\mu, \sigma^2)$.
2. Let $X_1, \ldots, X_p \overset{\text{iid}}{\sim} N(0,1)$. Then, $\sum_{i=1}^{p} X_i^2 \sim \chi^2(p)$.
3. Let $X \sim \chi^2(p)$ and $Z \sim \chi^2(q)$ be independent random variables. Then, $X + Z \sim \chi^2(p+q)$.
4. Let $X \sim N(0,1)$ and $Z \sim \chi^2(p)$ be independent random variables. Then, $\frac{X}{\sqrt{\frac{1}{p}Z}} \sim t(p)$.
5. Let $X \sim \chi^2(p)$ and $Z \sim \chi^2(q)$ be independent random variables. Then, $\frac{\frac{1}{p}X}{\frac{1}{q}Z} \sim F(p,q)$.

Non-central χ^2- and F-distribution

I.1.20 Remark

- Given independent random variables $X_1, \ldots, X_p$ with $X_i \sim N(\mu_i, 1)$, $\mu_i \in \mathbb{R}$, $1 \leqslant i \leqslant p$, the distribution of

$$\sum_{i=1}^{p} X_i^2$$

is called non-central χ^2-distribution $\chi^2(p, \delta)$ with $p \in \mathbb{N}$ degrees of freedom and non-centrality parameter $\delta = \frac{1}{2} \sum_{i=1}^{p} \mu_i^2 \geqslant 0$.

Clearly, $\chi^2(p) = \chi^2(p, 0)$.

- Let $X \sim \chi^2(p, \delta)$ and $Z \sim \chi^2(q)$ be independent random variables. Then,

$$\frac{\frac{1}{p} X}{\frac{1}{q} Z} \sim F(p, q, \delta),$$

that is, the ratio has a non-central F-distribution $F(p, q, \delta)$ with $p \in \mathbb{N}$ numerator and $q \in \mathbb{N}$ denominator degrees of freedom and non-centrality parameter $\delta \geqslant 0$.

Clearly, $F(p, q) = F(p, q, 0)$.