Lecture 6 | 07.04.2026

Statistical inference
in a normal linear model

Overview

- **Population model** $Y = \boldsymbol{X}^\top \beta + \varepsilon$ and the corresponding random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ i = 1, \dots, n\}$ drawn from the joint distribution $F_{(Y,\boldsymbol{X})}$ of some generic random vector $(Y, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{p+1}$ (where $\varepsilon \sim N(0, \sigma^2)$)

- The underlying structure (i.e., the model) is also assumed to hold for

$$Y_i = \boldsymbol{X}_i^\top \beta + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \text{ where } \varepsilon_i \sim N(0, \sigma^2)$$

- The model can be also equivalently expressed in a matrix notation as

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{\varepsilon}, \qquad \text{where } \boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \mathbb{I})$$

Overview

- **Population model** $Y = \boldsymbol{X}^\top \beta + \varepsilon$ and the corresponding random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ i = 1, \dots, n\}$ drawn from the joint distribution $F_{(Y,\boldsymbol{X})}$ of some generic random vector $(Y, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{p+1}$ (where $\varepsilon \sim N(0, \sigma^2)$)

- The underlying structure (i.e., the model) is also assumed to hold for

$$Y_i = \boldsymbol{X}_i^\top \beta + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \text{ where } \varepsilon_i \sim N(0, \sigma^2)$$

- The model can be also equivalently expressed in a matrix notation as

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{\varepsilon}, \qquad \text{where } \boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \mathbb{I})$$

- The model formulations used above specify the following:
  - the (conditional) mean structure of $\boldsymbol{Y}$ given $\mathbb{X}$ (i.e, $E[Y|\mathbb{X}] = \mathbb{X}\beta$)
  - the (conditional) variance-covariance structure of $\boldsymbol{Y}$ (i.e., $Var\,\boldsymbol{Y} = \sigma^2 \mathbb{I}$)
  - independence of Y_i and Y_j for any $i \neq j$ (zero correlation + normality)

Overview

- **Population model** $Y = \boldsymbol{X}^\top \beta + \varepsilon$ and the corresponding random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ i = 1, \dots, n\}$ drawn from the joint distribution $F_{(Y,\boldsymbol{X})}$ of some generic random vector $(Y, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{p+1}$ (where $\varepsilon \sim N(0, \sigma^2)$)

- The underlying structure (i.e., the model) is also assumed to hold for

$$Y_i = \boldsymbol{X}_i^\top \beta + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \text{ where } \varepsilon_i \sim N(0, \sigma^2)$$

- The model can be also equivalently expressed in a matrix notation as

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{\varepsilon}, \qquad \text{where } \boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \mathbb{I})$$

- The model formulations used above specify the following:
  - the (conditional) mean structure of $\boldsymbol{Y}$ given $\mathbb{X}$ (i.e, $E[Y|\mathbb{X}] = \mathbb{X}\beta$)
  - the (conditional) variance-covariance structure of $\boldsymbol{Y}$ (i.e., $Var\,\boldsymbol{Y} = \sigma^2 \mathbb{I}$)
  - independence of Y_i and Y_j for any $i \neq j$ (zero correlation + normality)

- Moreover, the joint distribution function $F_{(Y,\boldsymbol{X})}(y, \boldsymbol{x})$ can be factorized as

$$F_{(Y,\boldsymbol{X})}(y, \boldsymbol{x}) = F_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) \cdot F_{\boldsymbol{X}}(\boldsymbol{x}), \quad \forall (y, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^p$$

  where $F_{Y|\boldsymbol{X}}$ is the normal distribution and $F_{\boldsymbol{X}}$ does not depend on β, σ^2

Typical linear regression model assumptions

- **Ordinary linear regression model**
  - random sample $(Y_i, \boldsymbol{X}_i^\top)^\top$, $i = 1, \dots, n$ from the joint distribution $F_{(Y, \boldsymbol{X})}$
  - mean specification $E[\boldsymbol{Y}|\mathbb{X}] = \mathbb{X}\beta$, respectively $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$
  - variance specification $Var(\boldsymbol{Y}|\mathbb{X}) = \sigma^2\mathbb{I}$, resp. $Var(\varepsilon) = \sigma^2\mathbb{I}$

- **Normal linear regression model**
  - random sample $(Y_i, \boldsymbol{X}_i^\top)^\top$, $i = 1, \dots, n$ from the joint distribution $F_{(Y, \boldsymbol{X})}$
  - such that $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$

Typical linear regression model assumptions

- **Ordinary linear regression model**
  - random sample $(Y_i, \boldsymbol{X}_i^\top)^\top$, $i = 1, \dots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$
  - mean specification $E[\boldsymbol{Y}|\mathbb{X}] = \mathbb{X}\beta$, respectively $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$
  - variance specification $Var(\boldsymbol{Y}|\mathbb{X}) = \sigma^2\mathbb{I}$, resp. $Var(\varepsilon) = \sigma^2\mathbb{I}$

- **Normal linear regression model**
  - random sample $(Y_i, \boldsymbol{X}_i^\top)^\top$, $i = 1, \dots, n$ from the joint distribution $F_{(Y,\boldsymbol{X})}$
  - such that $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$

The formulation of the normal linear model above also implies the following:

- $\varepsilon|\mathbb{X} \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$
- $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$
- thus, the error terms $\varepsilon_1, \dots, \varepsilon_n$ form a random sample from a univariate normal distribution with the zero mean and the variance $\sigma^2 > 0$

Parameter estimation in the normal model

Recall, that there are basically two standard techniques for the parameter estimation under the normal linear regression model formulation:

- **Least Squares**
- **Maximum Likelihood**

Parameter estimation in the normal model

Recall, that there are basically two standard techniques for the parameter estimation under the normal linear regression model formulation:

- **Least Squares**
- **Maximum Likelihood**

In both situations the estimates for $\beta \in \mathbb{R}^p$ are given by the formulae

- $\widehat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X} \boldsymbol{Y}$, where $\mathbb{X}^\top \mathbb{X}$ is of a full rank, i.e. $p \in \mathbb{N}$

Under the ML estimation, the estimate for $\sigma^2 > 0$ can be also obtained

- $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$, where $\widehat{Y}_i = Y_i - \boldsymbol{X}_i^\top \widehat{\beta}$ are the fitted values

(the residual sum of squares (RSS) divided by the number of subjects $n \in \mathbb{N}$)

Parameter estimation in the normal model

Recall, that there are basically two standard techniques for the parameter estimation under the normal linear regression model formulation:

- **Least Squares**
- **Maximum Likelihood**

In both situations the estimates for $\beta \in \mathbb{R}^p$ are given by the formulae

- $\widehat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X} \boldsymbol{Y}$, where $\mathbb{X}^\top \mathbb{X}$ is of a full rank, i.e. $p \in \mathbb{N}$

Under the ML estimation, the estimate for $\sigma^2 > 0$ can be also obtained

- $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$, where $\widehat{Y}_i = Y_i - \boldsymbol{X}_i^\top \widehat{\beta}$ are the fitted values

(the residual sum of squares (RSS) divided by the number of subjects $n \in \mathbb{N}$)

Both estimates—quantities $\widehat{\beta}$ and $\widehat{\sigma^2}$—are random quantities (random vector and random variable) and, therefore, it is reasonable to investigate their statistical properties (e.g., the mean, variance, distribution, etc.)

(note, that the ML estimate $\widehat{\sigma^2}$ is biased and, instead, an unbiased estimate $s^2 = RSS/(n-p)$ is often used as a more appropriate alternative – also reffered to as the Mean Squared Error, MSe)

Model utilization for a prediction of some new Y

- One of the principal roles of the regression model is use the information in $\boldsymbol{X} \in \mathbb{R}^p$ (typically easily accessible) to learn something relevant (e.g. the conditional mean) about the variable Y (which is typically not observed in a straightforward way)—typically applied for $(Y_{new}, \boldsymbol{x}_{new}^\top)^\top$ independent from the original sample where $\boldsymbol{x}_{new} \in \mathbb{R}^p$ is known and $Y_{new} \in \mathbb{R}$ is unknown

Model utilization for a prediction of some new Y

- One of the principal roles of the regression model is use the information in $\boldsymbol{X} \in \mathbb{R}^p$ (typically easily accessible) to learn something relevant (e.g. the conditional mean) about the variable Y (which is typically not observed in a straightforward way)—typically applied for $(Y_{new}, \boldsymbol{x}_{new}^\top)^\top$ independent from the original sample where $\boldsymbol{x}_{new} \in \mathbb{R}^p$ is known and $Y_{new} \in \mathbb{R}$ is unknown

- For the parameter estimate $\widehat{\beta} \in \mathbb{R}^p$ in a normal linear model it holds that
$$\widehat{\beta} \sim N_p(\beta, \sigma^2(\mathbb{X}^\top \mathbb{X})^{-1})$$

Model utilization for a prediction of some new Y

- One of the principal roles of the regression model is use the information in $\boldsymbol{X} \in \mathbb{R}^p$ (typically easily accessible) to learn something relevant (e.g. the conditional mean) about the variable Y (which is typically not observed in a straightforward way)—typically applied for $(Y_{new}, \boldsymbol{x}_{new}^\top)^\top$ independent from the original sample where $\boldsymbol{x}_{new} \in \mathbb{R}^p$ is known and $Y_{new} \in \mathbb{R}$ is unknown

- For the parameter estimate $\widehat{\beta} \in \mathbb{R}^p$ in a normal linear model it holds that
$$\widehat{\beta} \sim N_p(\beta, \sigma^2(\mathbb{X}^\top\mathbb{X})^{-1})$$

- For the new observation from the same model $Y_{new} = \boldsymbol{x}_{new}^\top\beta + \varepsilon_{new}$ it holds
$$Y_{new} \sim N(\boldsymbol{x}_{new}^\top\beta, \sigma^2)$$

Model utilization for a prediction of some new Y

- One of the principal roles of the regression model is use the information in $\boldsymbol{X} \in \mathbb{R}^p$ (typically easily accessible) to learn something relevant (e.g. the conditional mean) about the variable Y (which is typically not observed in a straightforward way)—typically applied for $(Y_{new}, \boldsymbol{x}_{new}^\top)^\top$ independent from the original sample where $\boldsymbol{x}_{new} \in \mathbb{R}^p$ is known and $Y_{new} \in \mathbb{R}$ is unknown

- For the parameter estimate $\widehat{\beta} \in \mathbb{R}^p$ in a normal linear model it holds that
$$\widehat{\beta} \sim N_p(\beta, \sigma^2(\mathbb{X}^\top\mathbb{X})^{-1})$$

- For the new observation from the same model $Y_{new} = \boldsymbol{x}_{new}^\top\beta + \varepsilon_{new}$ it holds
$$Y_{new} \sim N(\boldsymbol{x}_{new}^\top\beta, \sigma^2)$$

- Thus, given the **independence of Y_{new} and $(Y_i, \boldsymbol{X}_i)$** for $i = 1, \dots, n$ it holds
$$(Y_{new} - \widehat{\mu}_{new}) \equiv (Y_{new} - \boldsymbol{x}_{new}^\top\widehat{\beta}) \sim N(\boldsymbol{x}_{new}^\top\beta, \sigma^2(1 + \boldsymbol{x}_{new}^\top(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{x}_{new}))$$

Model utilization for a prediction of some new Y

- One of the principal roles of the regression model is use the information in $\boldsymbol{X} \in \mathbb{R}^p$ (typically easily accessible) to learn something relevant (e.g. the conditional mean) about the variable Y (which is typically not observed in a straightforward way)—typically applied for $(Y_{new}, \boldsymbol{x}_{new}^\top)^\top$ independent from the original sample where $\boldsymbol{x}_{new} \in \mathbb{R}^p$ is known and $Y_{new} \in \mathbb{R}$ is unknown

- For the parameter estimate $\widehat{\beta} \in \mathbb{R}^p$ in a normal linear model it holds that
$$\widehat{\beta} \sim N_p(\beta, \sigma^2(\mathbb{X}^\top\mathbb{X})^{-1})$$

- For the new observation from the same model $Y_{new} = \boldsymbol{x}_{new}^\top\beta + \varepsilon_{new}$ it holds
$$Y_{new} \sim N(\boldsymbol{x}_{new}^\top\beta, \sigma^2)$$

- Thus, given the **independence of Y_{new} and $(Y_i, \boldsymbol{X}_i)$** for $i = 1, \dots, n$ it holds
$$(Y_{new} - \widehat{\mu}_{new}) \equiv (Y_{new} - \boldsymbol{x}_{new}^\top\widehat{\beta}) \sim N(\boldsymbol{x}_{new}^\top\beta, \sigma^2(1 + \boldsymbol{x}_{new}^\top(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{x}_{new}))$$

- Thus, the corresponding **prediction interval for Y_{new}** and some $\alpha \in (0,1)$ is
$$P\left[Y_{new} \in \left(\boldsymbol{x}_{new}^\top\widehat{\beta} \pm t_{1-\alpha/2}(n-p)\sqrt{MSe(1 + \boldsymbol{x}_{new}^\top(\mathbb{X}^\top\mathbb{X})^{-1}\boldsymbol{x}_{new})}\right)\right] = 1 - \alpha$$
$\hookrightarrow$ where MSe is used as a surrogate for the unknown $\sigma^2 > 0$

Model utilization for an inference about β

- The normal model $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, where $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$ is assumed to hold
- The unknown parameters to be estimated are $\beta \in \mathbb{R}^p$, and $\sigma^2 > 0$

Model utilization for an inference about β

- The normal model $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, where $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$ is assumed to hold
- The unknown parameters to be estimated are $\beta \in \mathbb{R}^p$, and $\sigma^2 > 0$

- **Statistical inference** (involves confidence intervals and statistical tests) can be also performed with respect to the parameters β and σ^2
(or it can be of some interest to do inference about some linear combination(s) of β)
- **From the practical point of view**, we are interested in the parameter vector $\beta \in \mathbb{R}^p$ itself but also some (reasonable) linear combinations of the form $\boldsymbol{l}^\top\beta$, for some (fixed) vector $\boldsymbol{l} \in \mathbb{R}^p$ (or $\mathbb{L}\beta$ for a matrix $\mathbb{L} \in \mathbb{R}^{m\times p}$)

Model utilization for an inference about β

- The normal model $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, where $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$ is assumed to hold
- The unknown parameters to be estimated are $\beta \in \mathbb{R}^p$, and $\sigma^2 > 0$

- **Statistical inference** (involves confidence intervals and statistical tests) can be also performed with respect to the parameters β and σ^2
(or it can be of some interest to do inference about some linear combination(s) of β)
- **From the practical point of view**, we are interested in the parameter vector $\beta \in \mathbb{R}^p$ itself but also some (reasonable) linear combinations of the form $\boldsymbol{l}^\top\beta$, for some (fixed) vector $\boldsymbol{l} \in \mathbb{R}^p$ (or $\mathbb{L}\beta$ for a matrix $\mathbb{L} \in \mathbb{R}^{m\times p}$)
  - the process of building the final model – an inference on some elements of the unknown vector $\beta \in \mathbb{R}^p$ can help to decide which covariates—the elements of $\boldsymbol{X} \in \mathbb{R}^p$—should be included in the model
(e.g., statistical tests about a model and its submodel)
  - interpretation of the final model – the inference allows to use the estimated parameters in $\widehat{\beta} \in \mathbb{R}^p$ and to make statistically valid conclusions about the unknown (underlying) population
(with respect to the population, interest is typically not given to β itself)

Model utilization for an inference about β

- The normal model $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, where $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$ is assumed to hold
- The unknown parameters to be estimated are $\beta \in \mathbb{R}^p$, and $\sigma^2 > 0$

- **Statistical inference** (involves confidence intervals and statistical tests) can be also performed with respect to the parameters β and σ^2
  (or it can be of some interest to do inference about some linear combination(s) of β)
- **From the practical point of view**, we are interested in the parameter vector $\beta \in \mathbb{R}^p$ itself but also some (reasonable) linear combinations of the form $\boldsymbol{l}^\top\beta$, for some (fixed) vector $\boldsymbol{l} \in \mathbb{R}^p$ (or $\mathbb{L}\beta$ for a matrix $\mathbb{L} \in \mathbb{R}^{m\times p}$)
  - the process of building the final model – an inference on some elements of the unknown vector $\beta \in \mathbb{R}^p$ can help to decide which covariates—the elements of $\boldsymbol{X} \in \mathbb{R}^p$—should be included in the model
    (e.g., statistical tests about a model and its submodel)
  - interpretation of the final model – the inference allows to use the estimated parameters in $\widehat{\beta} \in \mathbb{R}^p$ and to make statistically valid conclusions about the unknown (underlying) population
    (with respect to the population, interest is typically not given to β itself)

Note: Formally, one can be also interested in some more complex transformations of the unknown vector of parameters but linear contrasts assumed above preserve simplicity and explicit formulas...

Linear combinations of the model parameters

- The unknown vector of parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ is used to model the conditional mean structure $E[\boldsymbol{Y}|\mathbb{X}]$ but specific interpretation (meaning) of the elements of $\boldsymbol{\beta}$ depends on the parametrization that is used

- Therefore, it is also of some interest to be able to perform statistical inference about some appropriate linear combinations of the unknown parameters $\boldsymbol{\beta} \in \mathbb{R}^p$

- Let $\mathbb{L} \in \mathbb{R}^{m \times p}$ be a matrix with nonzero rows $\boldsymbol{l}_1^\top, \dots, \boldsymbol{l}_m^\top$ and let $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta} = (\boldsymbol{l}_1^\top \boldsymbol{\beta}, \dots, \boldsymbol{l}_m^\top \boldsymbol{\beta})^\top = (\theta_1, \dots, \theta_m)^\top \in \mathbb{R}^m$ be some linear combination of the original parameter $\boldsymbol{\beta} \in \mathbb{R}^p$ vector

- Thus, instead of performing the statistical inference about $\boldsymbol{\beta} \in \mathbb{R}^p$ the statistical inference is focusing on $\boldsymbol{\theta} \in \mathbb{R}^m$ instead (or $\mathbb{L}\boldsymbol{\beta}$ respectively)

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$ Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m \times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\boldsymbol{0}, \sigma^2\mathbb{M})$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\mathbf{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\mathbf{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$
- Random vectors $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally (given $\mathbb{X}$) independent

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$ and $\widehat{\beta} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\beta \in \mathbb{R}^p$ Moreover, $\boldsymbol{\theta} = \mathbb{L}\beta$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\boldsymbol{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$
- Random vectors $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally (given $\mathbb{X}$) independent
- Random vector $\widehat{\boldsymbol{\theta}}$ and SSe are conditionally (given $\mathbb{X}$) independent

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\boldsymbol{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$
- Random vectors $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally (given $\mathbb{X}$) independent
- Random vector $\widehat{\boldsymbol{\theta}}$ and SSe are conditionally (given $\mathbb{X}$) independent
- $MSe(n-p)/\sigma^2 = SSe/\sigma^2 \sim \chi^2_{n-p}$ and $\|\widehat{\boldsymbol{Y}} - \mathbb{X}\boldsymbol{\beta}\|^2/\sigma^2 \sim \chi^2_p$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$ and $\widehat{\beta} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\beta \in \mathbb{R}^p$
Moreover, $\boldsymbol{\theta} = \mathbb{L}\beta$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\beta}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\boldsymbol{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$
- Random vectors $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally (given $\mathbb{X}$) independent
- Random vector $\widehat{\boldsymbol{\theta}}$ and SSe are conditionally (given $\mathbb{X}$) independent
- $MSe(n-p)/\sigma^2 = SSe/\sigma^2 \sim \chi^2_{n-p}$ and $\|\widehat{\boldsymbol{Y}} - \mathbb{X}\beta\|^2/\sigma^2 \sim \chi^2_p$
- $T_j = \frac{\widehat{\theta}_j - \theta_j}{\sqrt{MSe\cdot v_{jj}}} \sim t_{n-p}$, where $\mathbb{V} = \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top = (v_{ij})_{i,j=1}^m$

Statistical properties of $\widehat{\beta}$ and $\widehat{\theta}$

Recall, that we are working with the **normal linear model** of the form $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$ Moreover, $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m\times p}$, such that $rank(\mathbb{L}) = m$

Then the following holds:

- $\widehat{\boldsymbol{\theta}} = \mathbb{L}\widehat{\boldsymbol{\beta}}$ is the (BLUE) estimate for $\boldsymbol{\theta} \in \mathbb{R}^m$
- $\widehat{\boldsymbol{Y}}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{H})$
- $\boldsymbol{U}|\mathbb{X} \sim N_n(\boldsymbol{0}, \sigma^2\mathbb{M})$
- $\widehat{\boldsymbol{\theta}} \sim N_m(\boldsymbol{\theta}, \sigma^2\mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top)$
- Random vectors $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{U}$ are conditionally (given $\mathbb{X}$) independent
- Random vector $\widehat{\boldsymbol{\theta}}$ and SSe are conditionally (given $\mathbb{X}$) independent
- $MSe(n-p)/\sigma^2 = SSe/\sigma^2 \sim \chi^2_{n-p}$ and $\|\widehat{\boldsymbol{Y}} - \mathbb{X}\boldsymbol{\beta}\|^2/\sigma^2 \sim \chi^2_p$
- $T_j = \frac{\widehat{\theta}_j - \theta_j}{\sqrt{MSe \cdot v_{jj}}} \sim t_{n-p}$, where $\mathbb{V} = \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top = (v_{ij})_{i,j=1}^m$
- $\frac{1}{m}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \Big(MSe \cdot \mathbb{V}\Big)^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim F_{m,n-p}$

Inference in a normal linear model

- **Inference about some $\beta_j \in \mathbb{R}$** (one element in $\boldsymbol{\beta} \in \mathbb{R}^p$)
  - confidence interval $\widehat{\beta}_j \pm t_{n-p}(1-\alpha/2)\sqrt{MSe \cdot v_{jj}}$, where $Var\widehat{\beta}_j = \sigma^2 v_{jj}$
  - statistical tests of the null hypothesis $H_0 : \beta_j = \beta_j^{(0)}$ against some H_A

Inference in a normal linear model

- **Inference about some $\beta_j \in \mathbb{R}$** (one element in $\boldsymbol{\beta} \in \mathbb{R}^p$)
  - confidence interval $\widehat{\beta}_j \pm t_{n-p}(1-\alpha/2)\sqrt{MSe \cdot v_{jj}}$, where $Var\widehat{\beta}_j = \sigma^2 v_{jj}$
  - statistical tests of the null hypothesis $H_0 : \beta_j = \beta_j^{(0)}$ against some H_A

- **Simultaneous confidence region for $\boldsymbol{\beta}$**
  - $\mathcal{S}(\boldsymbol{\alpha}) = \{\boldsymbol{\beta} \in \mathbb{R}^p;\ \frac{1}{p}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top (MSe^{-1}\mathbb{X}^\top\mathbb{X})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) < F_{p,n-p}(1-\alpha)\}$, which is an elipsoid with the center $\widehat{\boldsymbol{\beta}}$, the shape matrix $MSe \cdot (\mathbb{X}^\top\mathbb{X})^{-1}$ and the diameter $\sqrt{kF_{p,n-p}(1-\alpha)}$
  - statistical test of the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ against some H_A

Summary

- **Simple inference in the normal linear model**
  - confidence intervals and statistical tests for elements of $\boldsymbol{\beta} \in \mathbb{R}^p$
  - confidence intervals for some linear combination $\boldsymbol{l}^\top \boldsymbol{\beta}$, where $\boldsymbol{l} \in \mathbb{R}^p$
- **Simultaneous inference for vector parameters**
  - confidence regions and statistical tests for the whole vector $\boldsymbol{\beta} \in \mathbb{R}^p$
  - confidence regions for some linear combinations $\mathbb{L}\boldsymbol{\beta}$, where $\mathbb{L} \in \mathbb{R}^{m \times p}$
- **Prediction in the normal linear model**
  - point prediction for a new value of Y given the observed $\boldsymbol{X} = \boldsymbol{x} \quad (\boldsymbol{x}_{new})$
  - interval prediction for a new value of Y given the observed $\boldsymbol{X} = \boldsymbol{x} \quad (\boldsymbol{x}_{new})$