Lecture 7 | 09.04.2024

Statistical inference

in a normal linear model

Overview

- In general, the random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ i = 1, \dots, n\}$ from some joint distribution $F_{(Y,\boldsymbol{X})}$ (a generic random vector $(Y, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{p+1}$)
- the underlying structure (i.e., linear model) is assumed to hold

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

- the model can be equivalently also 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 above specifies the (conditional) mean structure $(\mathbb{X}\beta)$ and the (conditional) variance-covariance structure $(\sigma^2\mathbb{I})$ of the random vector $\boldsymbol{Y}$ given the random matrix $\mathbb{X}$

Overview

- In general, the random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ i = 1, \ldots, n\}$ from some joint distribution $F_{(Y,\boldsymbol{X})}$ (a generic random vector $(Y, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{p+1}$)
- the underlying structure (i.e., linear model) is assumed to hold

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

- the model can be equivalently also 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 above specifies the (conditional) mean structure ($\mathbb{X}\beta$) and the (conditional) variance-covariance structure ($\sigma^2\mathbb{I}$) of the random vector $\boldsymbol{Y}$ given the random matrix $\mathbb{X}$
- the joint distribution function $F_{Y,\boldsymbol{X}}(y, \boldsymbol{x})$ can be factorized as

$$F_{Y,\boldsymbol{X}}(y, \boldsymbol{x}) = F_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) \cdot F_{\boldsymbol{X}}(\boldsymbol{x})$$

and $F_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ is assumed to be a conditional normal distribution

Typical linear model assumptions

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

Typical linear model assumptions

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

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

- $\varepsilon|\mathbb{X} \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$
- $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$
- 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

- the normal model $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbb{I})$ is assumed to hold
- the unknown parameters to be estimated are $\boldsymbol{\beta} \in \mathbb{R}^p$, and $\sigma^2 > 0$
- statistical inference (confidence intervals, statistical tests) could be, however, also performed with respect to the parameters β and σ^2 but, it can be also of some interest to do inference about some linear combination(s) of $\boldsymbol{\beta}$
- from the practical point of view, we are interested in the parameter vector $\boldsymbol{\beta}$ itself but also some (reasonable) linear combinations $\boldsymbol{l}^\top\boldsymbol{\beta}$, for some (fixed) vector $\boldsymbol{l} \in \mathbb{R}^p$

Parameter estimation in a normal model

There are basically two standard techniques for the parameter estimation under the linear model formulation:

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

Parameter estimation in a normal model

There are basically two standard techniques for the parameter estimation under the linear model formulation:

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

In both situations the estimates 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 $p \in \mathbb{N}$
- $\widehat{\sigma^2} = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$, where $\widehat{Y}_i = Y_i - \boldsymbol{X}_i^\top\widehat{\beta}$

Parameter estimation in a normal model

There are basically two standard techniques for the parameter estimation under the linear model formulation:

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

In both situations the estimates 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 $p \in \mathbb{N}$
- $\widehat{\sigma^2} = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$, where $\widehat{Y}_i = Y_i - \boldsymbol{X}_i^\top\widehat{\beta}$

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., mean, variance, distribution, etc.)

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 in the model
- therefore, it is also of some interest to perform statistical inference about some linear combination of the unknow vector of paramters—inference about some different parametrization of the mean structure
- 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 $\boldsymbol{\theta} = \mathbb{L}\boldsymbol{\beta} = (\boldsymbol{l}_1^\top \boldsymbol{\beta}, \dots, \boldsymbol{l}_m^\top \boldsymbol{\beta})^\top = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$ be some linear combinations of the original β vector
- instead of performing the inference about $\boldsymbol{\beta} \in \mathbb{R}^p$ we can be interested in performing the statistical inference about $\boldsymbol{\beta} \in \mathbb{R}^m$

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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 $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$ and $\widehat{\beta} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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 $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\beta, \sigma^2\mathbb{I})$ and $\widehat{\beta} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$

Then the following holds:

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

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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

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 $\boldsymbol{\beta} \in \mathbb{R}^p$

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
- $SSe/\sigma^2 \sim \chi^2_{n-p}$ 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 $\boldsymbol{Y}|\mathbb{X} \sim N_n(\mathbb{X}\boldsymbol{\beta}, \sigma^2\mathbb{I})$ and $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ is the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$

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
- $SSe/\sigma^2 \sim \chi^2_{n-p}$ 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 $\mathbb{V} = \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top = (v_{ij})_{i,j=1}^m$

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 $\beta \in \mathbb{R}^p$

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
- $SSe/\sigma^2 \sim \chi^2_{n-p}$ 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 $\mathbb{V} = \mathbb{L}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{L}^\top = (v_{ij})_{i,j=1}^m$
- $\frac{1}{m}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\Big(MSe \cdot \mathbb{V}\Big)^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim F_{m,n-p}$, provided that $rank(\mathbb{L}) = m \leq p$

Inference in a normal linear model

- **Inference about some β_j**
  - 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 $H_0 : \beta_j = \beta_j^{(0)}$

Inference in a normal linear model

- **Inference about some β_j**
  - 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 $H_0 : \beta_j = \beta_j^{(0)}$

- **Simultaneous confidence region for β**
  - $\mathcal{S}(\alpha) = \{\beta \in \mathbb{R}^p;\ \frac{1}{p}(\beta - \widehat{\beta})^\top (MSe^{-1}\mathbb{X}^\top\mathbb{X})(\beta - \widehat{\beta}) < F_{p,n-p}(1-\alpha)\}$, which is an elipsoid with the center $\widehat{\beta}$, 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 $H_0 : \beta = \beta^{(0)}$

Model based predictions

- Model utilization for
  - characterization of the conditional distribution of Y given $\boldsymbol{X}$
  - explaining the effect of some covariate X_j on the variable Y
  - prediction of a new observation Y_{new} given the observed value of $\boldsymbol{X}_{new}$

Model based predictions

- Model utilization for
  - characterization of the conditional distribution of Y given $\boldsymbol{X}$
  - explaining the effect of some covariate X_j on the variable Y
  - prediction of a new observation Y_{new} given the observed value of $\boldsymbol{X}_{new}$

- straightforward prediction in terms of the estimated conditional expectation $\widehat{\mu}_{new} = \boldsymbol{X}_{new}^{\top}\widehat{\beta}$
- however, can we do better (e.g., accounting for the variability in Y_{new})?

Model based predictions

- Model utilization for
  - characterization of the conditional distribution of Y given $\boldsymbol{X}$
  - explaining the effect of some covariate X_j on the variable Y
  - prediction of a new observation Y_{new} given the observed value of $\boldsymbol{X}_{new}$

- straightforward prediction in terms of the estimated conditional expectation $\widehat{\mu}_{new} = \boldsymbol{X}_{new}^{\top}\widehat{\beta}$
- however, can we do better (e.g., accounting for the variability in Y_{new})?
- distributional assumption

$$Y_{new}|\boldsymbol{X}_{new} \sim N(\boldsymbol{X}_{new}^{\top}\beta, \sigma^2)$$

where $(Y_{new}, \boldsymbol{X}_{new})$ is independent of $\{(Y_i, \boldsymbol{X}_i);\ i = 1, \ldots, n\}$

Theoretical background of the prediction

- **Formally**

$$Y_{new} = \boldsymbol{X}_{new}^{\top}\beta + \varepsilon_{new}, \quad \text{for } \varepsilon_{new} \sim N(0, \sigma^2)$$

- **Theoretical property**

$$P[Y_{new} \in (\mathbb{X}_{new}^{\top}\beta \pm u_{1-\alpha/2}\sigma)] = 1 - \alpha$$

- **Theoretical property**

$$P\left[Y_{new} \in (\mathbb{X}_{new}^{\top}\widehat{\beta} \pm t_{1-\alpha/2}(n-p)\sqrt{1 + \boldsymbol{X}_{new}^{\top}(\mathbb{X}^{\top}\mathbb{X})^{-1}\boldsymbol{X}_{new}})\right] = 1 - \alpha$$

Summary

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