

Lecture 03: Classical Inferential Statistics I: Basics and Confidence Intervals

3.1 Expectation and Covariance Matrix
of the Ordinary Least Squares (OLS)
Estimator

3.2 Confidence Intervals

$$\frac{2 \sigma_{\epsilon}}{\sqrt{n}}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\end{aligned}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \underline{\underline{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\epsilon}}}\end{aligned}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \underline{\underline{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\epsilon}}}\end{aligned}$$

Expectation value

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\boldsymbol{\epsilon}) = \boldsymbol{\beta}$$

3.1. Ordinary Least Squares (OLS) Estimator: Expectation and Covariance

- ▶ Only stochasticity: residual errors ϵ according to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$
- ▶ The OLS estimator is linear in \mathbf{y} :

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \epsilon) \\ &= \underline{\underline{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon}}\end{aligned}$$

Expectation value

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\epsilon) = \boldsymbol{\beta}$$

The OLS estimator of parameter-linear models is **un-biased** under the mild condition $E(\epsilon) = \mathbf{0}$ for all the data points

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\mathbf{V}_{\hat{\beta}} \stackrel{\text{def}}{=} E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right)$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\mathbf{V}_{\hat{\beta}} \stackrel{\text{def}}{=} E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right)$$

[insert $\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon \rightarrow$]

$$= E \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon ((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon)' \right)$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\mathbf{V}_{\hat{\beta}} \stackrel{\text{def}}{=} E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right)$$

$$\text{[insert } \hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \rightarrow] = E \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \right)' \right)$$

$$\text{[transpose and inversion rules } \rightarrow] = E \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \right)$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\begin{aligned}
 \mathbf{V}_{\hat{\beta}} &\stackrel{\text{def}}{=} E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) \\
 \text{[insert } \hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)'\right) \\
 \text{[transpose and inversion rules } \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
 \text{[} E(\cdot) \text{ acts only on } \epsilon \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
 \end{aligned}$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\begin{aligned}
 \mathbf{V}_{\hat{\beta}} &\stackrel{\text{def}}{=} E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) \\
 \text{[insert } \hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)'\right) \\
 \text{[transpose and inversion rules } \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
 \text{[} E(\cdot) \text{ acts only on } \epsilon \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 \text{[Gauß-Markow } \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{\epsilon}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
 \end{aligned}$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\begin{aligned}
 \mathbf{V}_{\hat{\beta}} &\stackrel{\text{def}}{=} E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) \\
 \text{[insert } \hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)'\right) \\
 \text{[transpose and inversion rules } \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
 \text{[} E(\cdot) \text{ acts only on } \epsilon \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 \text{[Gauß-Markow } \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{\epsilon}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 \text{[def inverse matrix } \rightarrow] &= \sigma_{\epsilon}^2(\mathbf{X}'\mathbf{X})^{-1}
 \end{aligned}$$

OLS estimator: variances and covariances

- ▶ Gauß-Markow conditions $\rightarrow \epsilon \sim \text{i.i.d}N(0, \sigma^2) \rightarrow \hat{\beta}$ is normal distributed
- ▶ In this case, the complete error characteristics are specified by the expectation value and the **variance-covariance matrix** $\mathbf{V}_{\hat{\beta}}$

$$\begin{aligned}
 \mathbf{V}_{\hat{\beta}} &\stackrel{\text{def}}{=} E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) \\
 \text{[insert } \hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)'\right) \\
 \text{[transpose and inversion rules } \rightarrow] &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
 \text{[} E(\cdot) \text{ acts only on } \epsilon \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 \text{[Gauß-Markow } \rightarrow] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{\epsilon}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 \text{[def inverse matrix } \rightarrow] &= \sigma_{\epsilon}^2(\mathbf{X}'\mathbf{X})^{-1}
 \end{aligned}$$

The variance-covariance matrix depends only on the values of the exogenous factors!

Results

- ▶ Ordinary least squares (OLS) estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ **Variance-Covariance matrix** of the estimation errors (provided the errors are i.i.d.) can be written in terms of the **Hesse matrix H** of the objective function SSE:

$$\begin{aligned} \mathbf{V}_{\hat{\beta}} &= E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2 \mathbf{H}^{-1}, \\ H_{jk} &= \left. \frac{\partial^2 S}{\partial \beta_j \partial \beta_k} \right|_{\beta = \hat{\beta}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- ▶ Variances of estimation errors: $V(\hat{\beta}_j) = V_{jj}$
- ▶ Correlation of estimation errors: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- ▶ Distribution of the normalized estimation errors: $\frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$

Results

- ▶ Ordinary least squares (OLS) estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ **Variance-Covariance matrix** of the estimation errors (provided the errors are i.i.d.) can be written in terms of the **Hesse matrix \mathbf{H}** of the objective function SSE:

$$\begin{aligned} \mathbf{V}_{\hat{\beta}} &= E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2 \mathbf{H}^{-1}, \\ H_{jk} &= \left. \frac{\partial^2 S}{\partial \beta_j \partial \beta_k} \right|_{\beta = \hat{\beta}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- ▶ Variances of estimation errors: $V(\hat{\beta}_j) = V_{jj}$
- ▶ Correlation of estimation errors: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- ▶ Distribution of the normalized estimation errors: $\frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$

Results

- ▶ Ordinary least squares (OLS) estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ **Variance-Covariance matrix** of the estimation errors (provided the errors are i.i.d.) can be written in terms of the **Hesse matrix \mathbf{H}** of the objective function SSE:

$$\begin{aligned} \mathbf{V}_{\hat{\beta}} &= E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2 \mathbf{H}^{-1}, \\ H_{jk} &= \left. \frac{\partial^2 S}{\partial \beta_j \partial \beta_k} \right|_{\beta = \hat{\beta}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- ▶ Variances of estimation errors: $V(\hat{\beta}_j) = V_{jj}$
- ▶ Correlation of estimation errors: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- ▶ Distribution of the normalized estimation errors: $\frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$

Results

- ▶ Ordinary least squares (OLS) estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ **Variance-Covariance matrix** of the estimation errors (provided the errors are i.i.d.) can be written in terms of the **Hesse matrix \mathbf{H}** of the objective function SSE:

$$\begin{aligned} \mathbf{V}_{\hat{\beta}} &= E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2 \mathbf{H}^{-1}, \\ H_{jk} &= \left. \frac{\partial^2 S}{\partial \beta_j \partial \beta_k} \right|_{\beta = \hat{\beta}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- ▶ Variances of estimation errors: $V(\hat{\beta}_j) = V_{jj}$
- ▶ Correlation of estimation errors: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- ▶ Distribution of the normalized estimation errors: $\frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$

Results

- ▶ Ordinary least squares (OLS) estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ **Variance-Covariance matrix** of the estimation errors (provided the errors are i.i.d.) can be written in terms of the **Hesse matrix H** of the objective function SSE:

$$\begin{aligned} \mathbf{V}_{\hat{\beta}} &= E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2 \mathbf{H}^{-1}, \\ H_{jk} &= \left. \frac{\partial^2 S}{\partial \beta_j \partial \beta_k} \right|_{\beta = \hat{\beta}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- ▶ Variances of estimation errors: $V(\hat{\beta}_j) = V_{jj}$
- ▶ Correlation of estimation errors: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- ▶ Distribution of the normalized estimation errors: $\frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$

Estimation of the residual variance

The above cannot be applied directly since the residual variance σ^2 is unknown and must be estimated by the minimum SSE $S(\hat{\beta})$:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_i (y_i - \hat{y}(x_i))^2 = \frac{S(\hat{\beta})}{n-p}$$

This is a generalisation of the variance estimator of a random variable Y with no exogenous variables (model $Y = \beta_0 + \epsilon$, $\epsilon \sim i.i.d.$, OLS estimator $\hat{y} = \hat{\beta}_0 = \bar{y}$):

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_i (y_i - \bar{y})^2$$

Argue qualitatively why an estimator with n instead of $n-1$ in the denominator would be biased

if we knew the expectation $E(Y)$, i.e., the true value $\mu = \beta_0$, we could use the unbiased estimator $\hat{\sigma}^2 = 1/n \sum_i (y_i - \mu)^2$. Otherwise, we need to replace the unknown μ by its estimator \bar{y} . If, by chance, the sample gives $\bar{y} > \mu$, then also the individual data y_i tend to be above μ , so $\sum_i (y_i - \bar{y})^2 < \sum_i (y_i - \mu)^2$. The same is true if \bar{y} happens to be $< \mu$. So, $\sum_i (y_i - \bar{y})^2$ is systematically smaller than $\sum_i (y_i - \mu)^2$ which must be compensated for by a smaller denominator.

Info: In [Derivation for experts](#), the unbiased estimators for the simple case ($p = 1, \hat{\beta}_0 = \bar{y}$) and the general case $p > 1$ are formally derived.

Results if the variance needs to be estimated

- ▶ Estimated variance-covariance matrix:

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}} = 2\hat{\sigma}^2 \mathbf{H}^{-1} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- ▶ *Normalized* parameter estimation errors (see [derivation for experts](#)):

$$\begin{aligned} \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \frac{\sigma}{\hat{\sigma}} = Z \frac{\sigma}{\hat{\sigma}} \\ &= Z \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}} = \frac{Z}{\sqrt{\frac{\sum_{i=1}^{n-p} Z_i^2}{n-p}}} \end{aligned}$$

- ▶ A random variable with a standard Gaussian in the numerator and the normalized square root of a $\chi^2(n-p)$ distributed random variable in the denominator defines a **student-t distribution** with $n-p$ degrees of freedom:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} \sim T(n-p)$$

Results if the variance needs to be estimated

- ▶ Estimated variance-covariance matrix:

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}} = 2\hat{\sigma}^2 \mathbf{H}^{-1} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- ▶ *Normalized* parameter estimation errors (see [derivation for experts](#)):

$$\begin{aligned} \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \frac{\sigma}{\hat{\sigma}} = Z \frac{\sigma}{\hat{\sigma}} \\ &= Z \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}} = \frac{Z}{\sqrt{\frac{\sum_{i=1}^{n-p} Z_i^2}{n-p}}} \end{aligned}$$

- ▶ A random variable with a standard Gaussian in the numerator and the normalized square root of a $\chi^2(n-p)$ distributed random variable in the denominator defines a **student-t distribution** with $n-p$ degrees of freedom:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} \sim T(n-p)$$

Results if the variance needs to be estimated

- ▶ Estimated variance-covariance matrix:

$$\hat{\mathbf{V}}_{\hat{\beta}} = 2\hat{\sigma}^2 \mathbf{H}^{-1} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- ▶ *Normalized* parameter estimation errors (see [derivation for experts](#)):

$$\begin{aligned} \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \frac{\sigma}{\hat{\sigma}} = Z \frac{\sigma}{\hat{\sigma}} \\ &= Z \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}} = \frac{Z}{\sqrt{\frac{\sum_{i=1}^{n-p} Z_i^2}{n-p}}} \end{aligned}$$

- ▶ A random variable with a standard Gaussian in the numerator and the normalized square root of a $\chi^2(n-p)$ distributed random variable in the denominator defines a **student-t distribution** with $n-p$ degrees of freedom:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} \sim T(n-p)$$

Multivariate distribution function of $\hat{\beta}$

The distribution of the errors $\Delta\hat{\beta} = \hat{\beta} - \beta$ obeys a multivariate normal distribution:

$$f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{1}{2} \Delta\hat{\beta}' \mathbf{V}^{-1} \Delta\hat{\beta} \right] = \exp \left[-\frac{\Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}}{2\sigma_{\epsilon}^2} \right].$$

Relation to the maximum-likelihood-method (→ Lecture 07:)

Expand the SSE $S(\beta)$ around $\hat{\beta}$ to second order:

$$S(\beta) - S(\hat{\beta}) \approx \frac{1}{2} \Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta} = \Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}$$

$$\Rightarrow f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{S(\beta) - S(\hat{\beta})}{2\sigma_{\epsilon}^2} \right]$$

and with the estimated residual variance $\hat{\sigma}_{\epsilon}^2 = S(\hat{\beta})/(n-p)$

$$\hat{f}_{\hat{\beta}}(\beta) \propto \exp \left[-\frac{(n-p)}{2} \left(\frac{S(\beta)}{S(\hat{\beta})} - 1 \right) \right]$$

Multivariate distribution function of $\hat{\beta}$

The distribution of the errors $\Delta\hat{\beta} = \hat{\beta} - \beta$ obeys a multivariate normal distribution:

$$f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{1}{2} \Delta\hat{\beta}' \mathbf{V}^{-1} \Delta\hat{\beta} \right] = \exp \left[-\frac{\Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}}{2\sigma_{\epsilon}^2} \right].$$

Relation to the maximum-likelihood-method (→ Lecture 07:)

Expand the SSE $S(\beta)$ around $\hat{\beta}$ to second order:

$$S(\beta) - S(\hat{\beta}) \approx \frac{1}{2} \Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta} = \Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}$$

$$\Rightarrow f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{S(\beta) - S(\hat{\beta})}{2\sigma_{\epsilon}^2} \right]$$

and with the estimated residual variance $\hat{\sigma}_{\epsilon}^2 = S(\hat{\beta})/(n-p)$

$$\hat{f}_{\hat{\beta}}(\beta) \propto \exp \left[-\frac{(n-p)}{2} \left(\frac{S(\beta)}{S(\hat{\beta})} - 1 \right) \right]$$

Multivariate distribution function of $\hat{\beta}$

The distribution of the errors $\Delta\hat{\beta} = \hat{\beta} - \beta$ obeys a multivariate normal distribution:

$$f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{1}{2} \Delta\hat{\beta}' \mathbf{V}^{-1} \Delta\hat{\beta} \right] = \exp \left[-\frac{\Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}}{2\sigma_{\epsilon}^2} \right].$$

Relation to the maximum-likelihood-method (→ Lecture 07:)

Expand the SSE $S(\beta)$ around $\hat{\beta}$ to second order:

$$S(\beta) - S(\hat{\beta}) \approx \frac{1}{2} \Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta} = \Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}$$

$$\Rightarrow f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{S(\beta) - S(\hat{\beta})}{2\sigma_{\epsilon}^2} \right]$$

and with the estimated residual variance $\hat{\sigma}_{\epsilon}^2 = S(\hat{\beta})/(n-p)$

$$\hat{f}_{\hat{\beta}}(\beta) \propto \exp \left[-\frac{(n-p)}{2} \left(\frac{S(\beta)}{S(\hat{\beta})} - 1 \right) \right]$$

Multivariate distribution function of $\hat{\beta}$

The distribution of the errors $\Delta\hat{\beta} = \hat{\beta} - \beta$ obeys a multivariate normal distribution:

$$f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{1}{2} \Delta\hat{\beta}' \mathbf{V}^{-1} \Delta\hat{\beta} \right] = \exp \left[-\frac{\Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}}{2\sigma_{\epsilon}^2} \right].$$

Relation to the maximum-likelihood-method (→ Lecture 07:)

Expand the SSE $S(\beta)$ around $\hat{\beta}$ to second order:

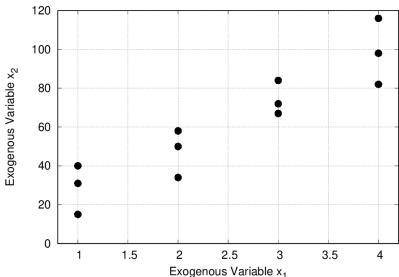
$$S(\beta) - S(\hat{\beta}) \approx \frac{1}{2} \Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta} = \Delta\hat{\beta}' \mathbf{X}'\mathbf{X} \Delta\hat{\beta}$$

$$\Rightarrow f_{\hat{\beta}}(\Delta\hat{\beta}) \propto \exp \left[-\frac{S(\beta) - S(\hat{\beta})}{2\sigma_{\epsilon}^2} \right]$$

and with the estimated residual variance $\hat{\sigma}_{\epsilon}^2 = S(\hat{\beta})/(n-p)$

$$\hat{f}_{\hat{\beta}}(\beta) \propto \exp \left[-\frac{(n-p)}{2} \left(\frac{S(\beta)}{S(\hat{\beta})} - 1 \right) \right]$$

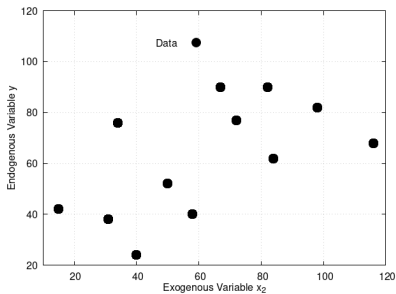
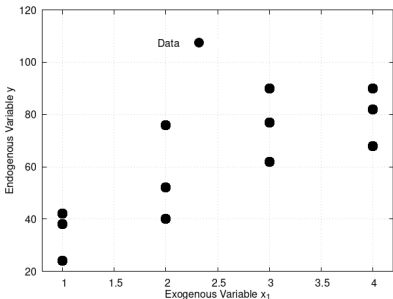
Example of correlated errors: modeling the demand for hotel rooms



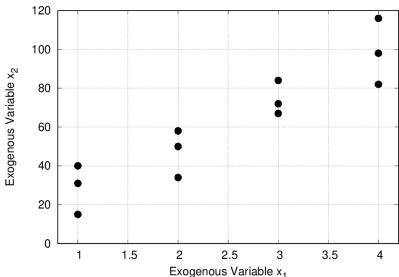
- ▶ The example of Lecture 02:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- ▶ Exogenous factors: $x_0 = 1$, x_1 : proxy for quality [# stars]; x_2 : price [€/night].
- ▶ Endogenous variable: booking rate [%]
- ▶ The demand is positively correlated with both the quality and the price



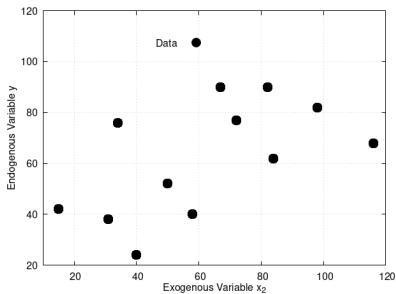
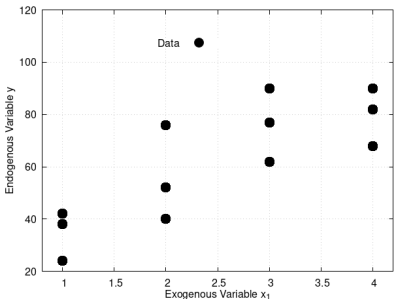
Example of correlated errors: modeling the demand for hotel rooms



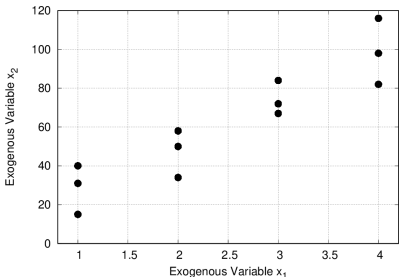
- ▶ The example of Lecture 02:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- ▶ Exogenous factors: $x_0 = 1$, x_1 : proxy for quality [# stars]; x_2 : price [€/night].
- ▶ Endogenous variable: booking rate [%]
- ▶ The demand is positively correlated with both the quality and the price



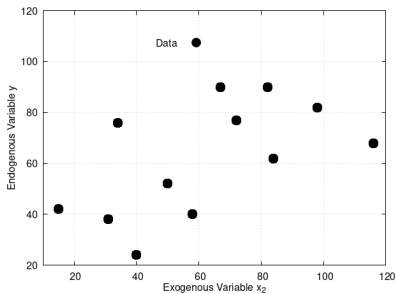
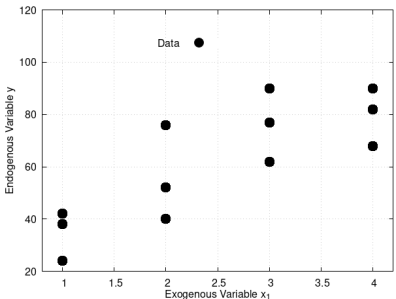
Example of correlated errors: modeling the demand for hotel rooms



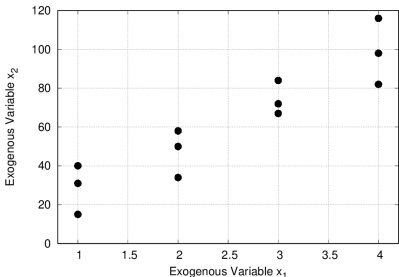
- ▶ The example of Lecture 02:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- ▶ Exogenous factors: $x_0 = 1$, x_1 : proxy for quality [# stars]; x_2 : price [€/night].
- ▶ Endogenous variable: booking rate [%]
- ▶ The demand is positively correlated with both the quality and the price



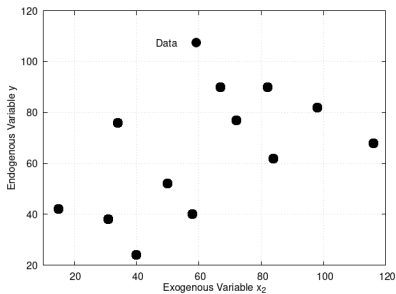
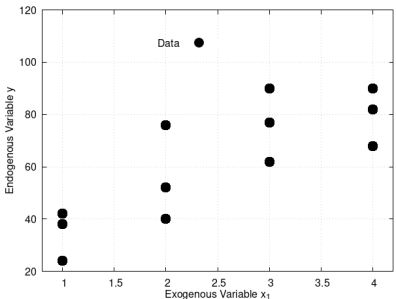
Example of correlated errors: modeling the demand for hotel rooms



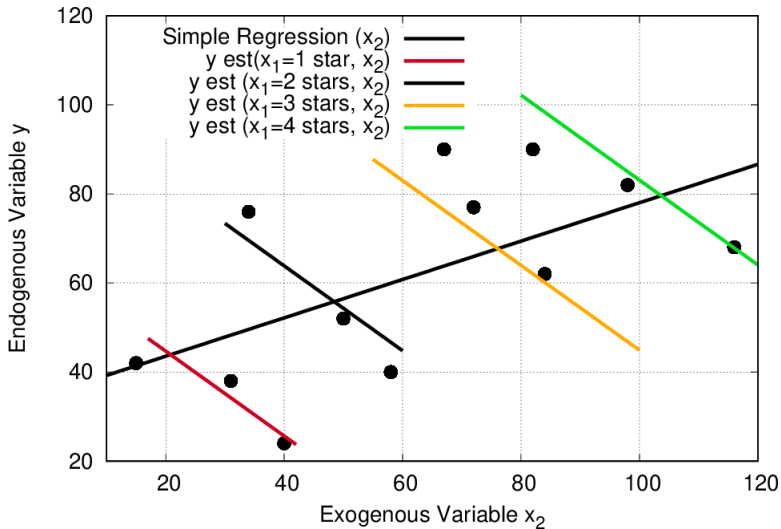
- ▶ The example of Lecture 02:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- ▶ Exogenous factors: $x_0 = 1$, x_1 : proxy for quality [# stars]; x_2 : price [€/night].
- ▶ Endogenous variable: booking rate [%]
- ▶ The demand is positively correlated with both the quality and the price

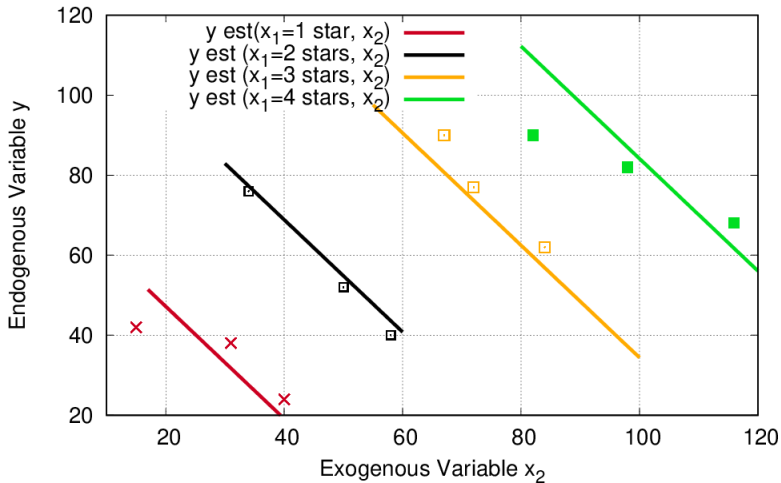


Residual errors for fitted parameters



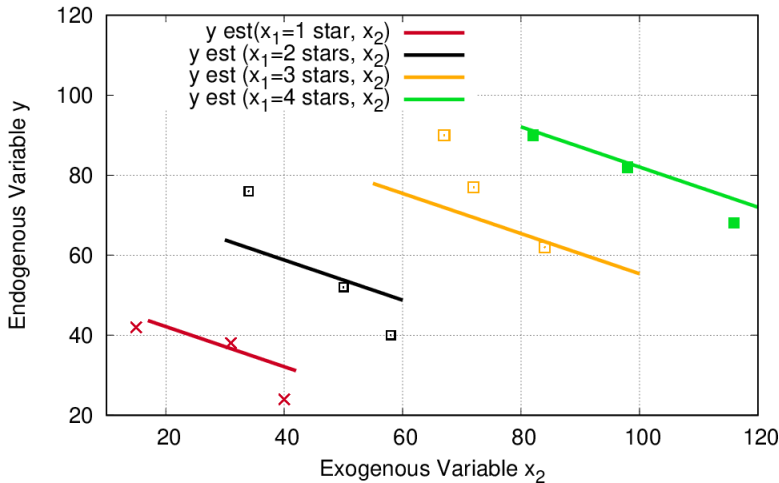
Effect of mis-fit parameters I: small effect if β_1 and β_2 have opposite misfits

β_1 and β_2 um $\Delta\beta_1$ bzw. $-\Delta\beta_2$ verschoben



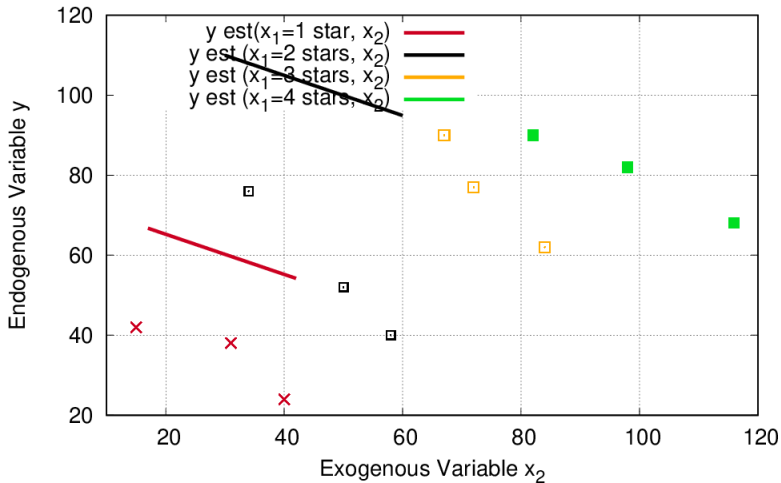
Effect of mis-fit parameters II: small effect if β_1 and β_2 have opposite misfits

β_1 and β_2 um $-\Delta\beta_1$ bzw. $+\Delta\beta_2$ verschoben



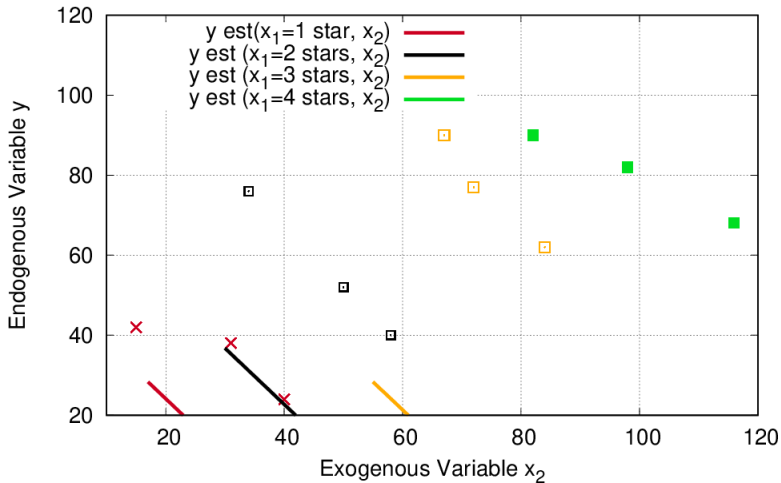
Effect of mis-fit parameters III: large effect if β_1 and β_2 have both positive misfits

β_1 and β_2 um $\Delta\beta_1$ bzw. $\Delta\beta_2$ verschoben

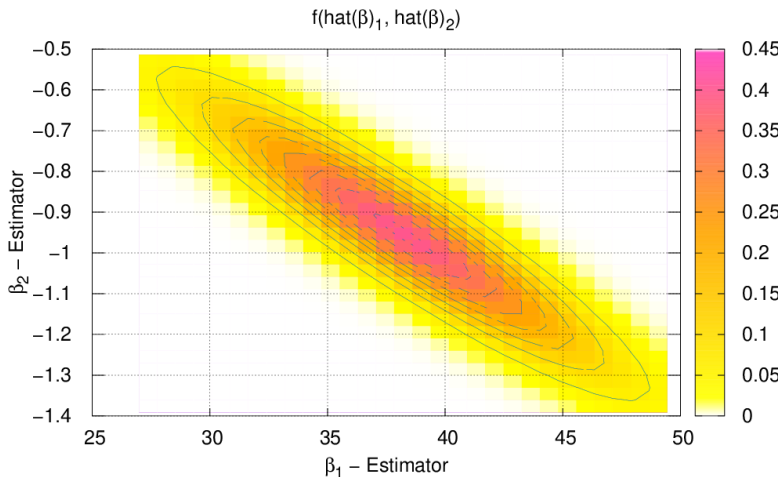


Effect of mis-fit parameters IV: large effect if β_1 and β_2 have both negative misfits

β_1 and β_2 um $-\Delta\beta_1$ bzw. $-\Delta\beta_2$ verschoben



All this results in a negative correlation
between the estimation errors for β_1 and β_2



Special case 1: No exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- ▶ System matrix: $\mathbf{X} = (1, 1, \dots, 1)'$
- ▶ OLS estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\mathbf{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$$

- ▶ Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- ▶ Distribution of the estimator (if $\epsilon \sim i.i.d.N(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n-1)$$

Special case 1: No exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- ▶ System matrix: $\mathbf{X} = (1, 1, \dots, 1)'$
- ▶ OLS estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\mathbf{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$$

- ▶ Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- ▶ Distribution of the estimator (if $\epsilon \sim i.i.d.N(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n-1)$$

Special case 1: No exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- ▶ System matrix: $\mathbf{X} = (1, 1, \dots, 1)'$
- ▶ OLS estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\mathbf{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$$

- ▶ Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- ▶ Distribution of the estimator (if $\epsilon \sim i.i.d.N(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n-1)$$

Special case 1: No exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- ▶ System matrix: $\mathbf{X} = (1, 1, \dots, 1)'$
- ▶ OLS estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\mathbf{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$$

- ▶ Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- ▶ Distribution of the estimator (if $\epsilon \sim i.i.dN(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$
$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n-1)$$

Special case 1: No exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- ▶ System matrix: $\mathbf{X} = (1, 1, \dots, 1)'$
- ▶ OLS estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\mathbf{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$$

- ▶ Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- ▶ Distribution of the estimator (if $\epsilon \sim i.i.d.N(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$
$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n - 1)$$

Special case 2: Simple linear regression

- ▶ Model (with $x_1 = x$): $y = \beta_0 + \beta_1 x + \epsilon$
- ▶ System matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

- ▶ OLS estimator (with $s_x^2 = 1/n(\sum x_i^2 - n\bar{x})$):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$$

$$\hat{\beta}_1 = \begin{pmatrix} -\frac{\bar{x}}{ns_x^2} & \frac{1}{ns_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}} = \frac{s_{xy}}{s_x^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Special case 2: Simple linear regression

- ▶ Model (with $x_1 = x$): $y = \beta_0 + \beta_1 x + \epsilon$
- ▶ System matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

- ▶ OLS estimator (with $s_x^2 = 1/n(\sum x_i^2 - n\bar{x})$):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$$

$$\hat{\beta}_1 = \begin{pmatrix} -\frac{\bar{x}}{ns_x^2}, \frac{1}{ns_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}} = \frac{s_{xy}}{s_x^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Special case 2: Simple linear regression

- ▶ Model (with $x_1 = x$): $y = \beta_0 + \beta_1 x + \epsilon$
- ▶ System matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

- ▶ OLS estimator (with $s_x^2 = 1/n(\sum x_i^2 - n\bar{x})$):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$$

$$\hat{\beta}_1 = \begin{pmatrix} -\frac{\bar{x}}{ns_x^2}, \frac{1}{ns_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}} = \frac{s_{xy}}{s_x^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Simple linear regression (ctnd)

- ▶ Variance-covariance matrix (assuming w/o loss of generality $\bar{x} = 0$):

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{ns_x^2} \end{pmatrix}$$

- ▶ Variance of the estimator $\hat{y}(x)$ (x is deterministic):

$$V(\hat{y}(x)) = V(\hat{\beta}_0 + \hat{\beta}_1 x) = V_{00} + x^2 V_{11} + 2x V_{01} = \frac{\sigma^2}{n} \left(1 + \frac{x^2}{s_x^2} \right)$$

- ▶ Distribution of the estimator for $y(x)$:

$$\hat{y}(x) \sim N(y(x), V(\hat{y}(x)))$$

If σ^2 has to be estimated by $\hat{\sigma}^2$, the normalized estimators for β_0 , β_1 and $y(x)$ are $\sim T(n-2)$.

Simple linear regression (ctnd)

- ▶ Variance-covariance matrix (assuming w/o loss of generality $\bar{x} = 0$):

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{ns_x^2} \end{pmatrix}$$

- ▶ Variance of the estimator $\hat{y}(x)$ (x is deterministic):

$$V(\hat{y}(x)) = V(\hat{\beta}_0 + \hat{\beta}_1 x) = V_{00} + x^2 V_{11} + 2x V_{01} = \frac{\sigma^2}{n} \left(1 + \frac{x^2}{s_x^2} \right)$$

- ▶ Distribution of the estimator for $y(x)$:

$$\hat{y}(x) \sim N(y(x), V(\hat{y}(x)))$$

If σ^2 has to be estimated by $\hat{\sigma}^2$, the normalized estimators for β_0 , β_1 and $y(x)$ are $\sim T(n-2)$.

Simple linear regression (ctnd)

- ▶ Variance-covariance matrix (assuming w/o loss of generality $\bar{x} = 0$):

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{ns_x^2} \end{pmatrix}$$

- ▶ Variance of the estimator $\hat{y}(x)$ (x is deterministic):

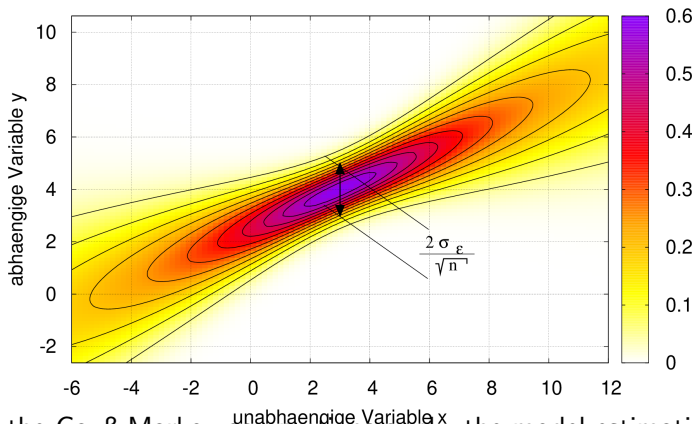
$$V(\hat{y}(x)) = V(\hat{\beta}_0 + \hat{\beta}_1 x) = V_{00} + x^2 V_{11} + 2x V_{01} = \frac{\sigma^2}{n} \left(1 + \frac{x^2}{s_x^2} \right)$$

- ▶ Distribution of the estimator for $y(x)$:

$$\hat{y}(x) \sim N(y(x), V(\hat{y}(x)))$$

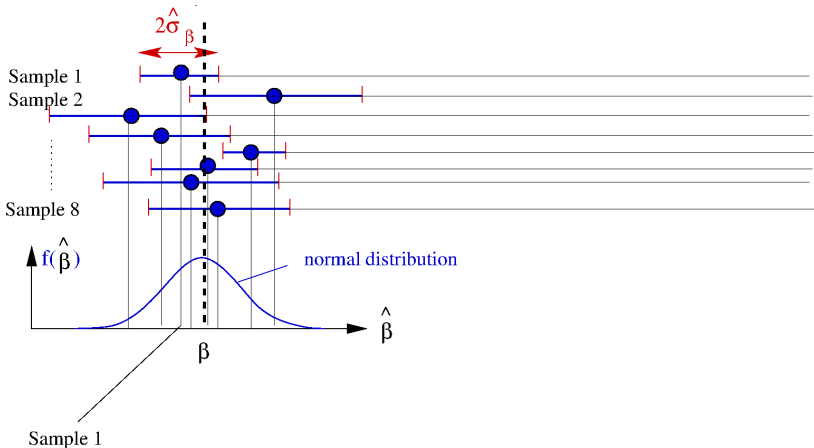
If σ^2 has to be estimated by $\hat{\sigma}^2$, the normalized estimators for β_0 , β_1 and $y(x)$ are $\sim T(n-2)$.

Probability density for $\hat{y}(x)$ for simple linear regression

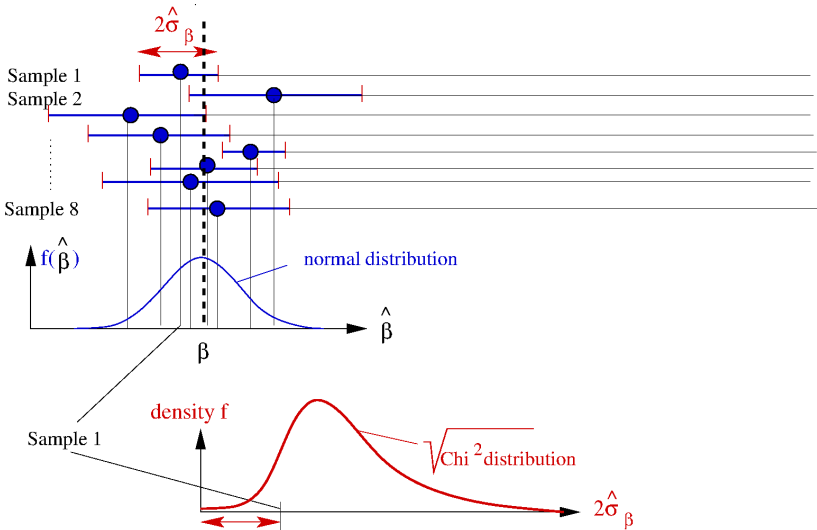


- ▶ If the Gauß-Markov assumptions apply, the model estimation errors $\hat{y}(x) - y(x)$ are Gaussian distributed
- ▶ The expectation and variance depends on x ; the standard error is hyperbola-shaped.

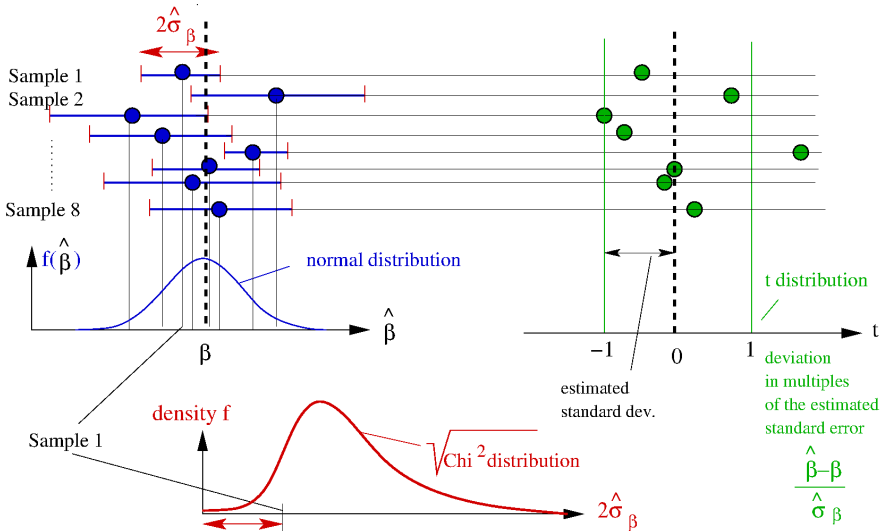
3.2. Confidence Intervals: where the Student-t distribution comes from



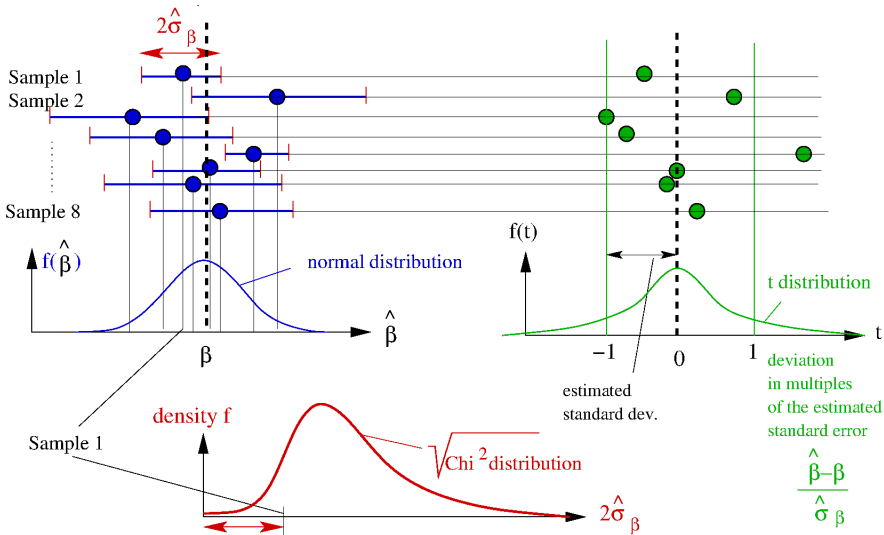
3.2. Confidence Intervals: where the Student-t distribution comes from



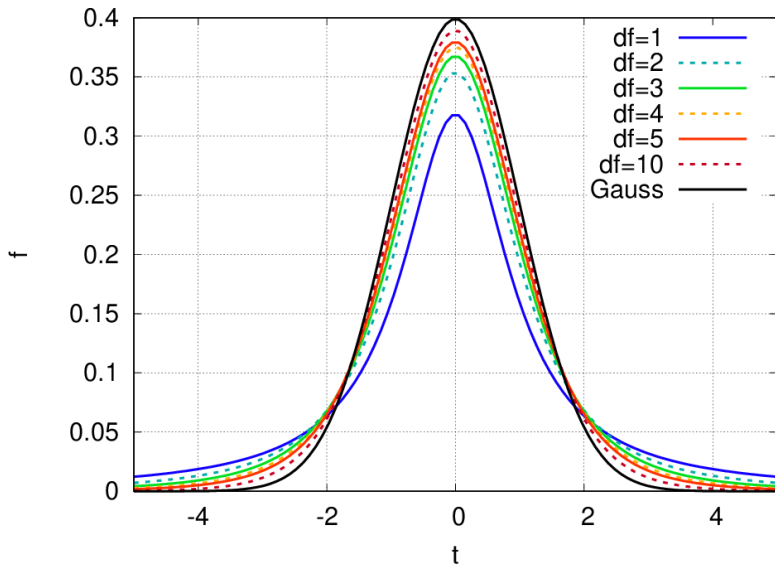
3.2. Confidence Intervals: where the Student-t distribution comes from



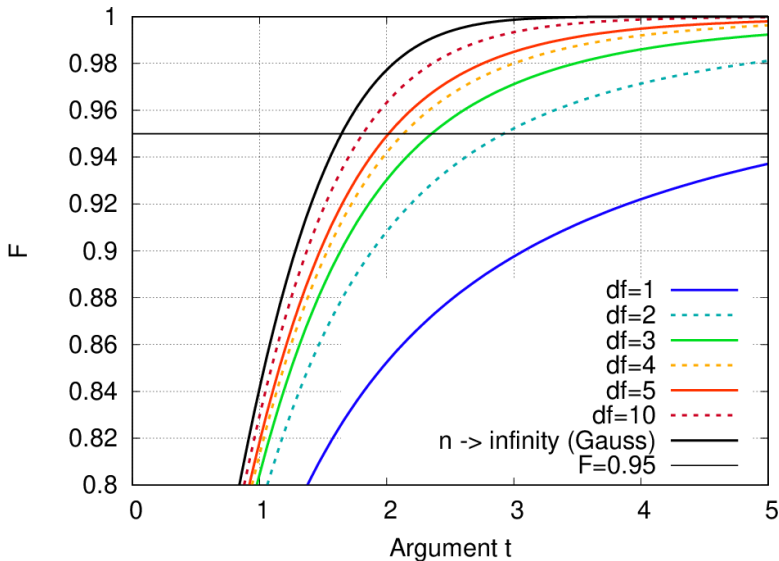
3.2. Confidence Intervals: where the Student-t distribution comes from



Densities of standard normal vs. Student-t distribution



Distributions of standard normal vs. Student-t-distribution



Calculation of the confidence intervals (CI)

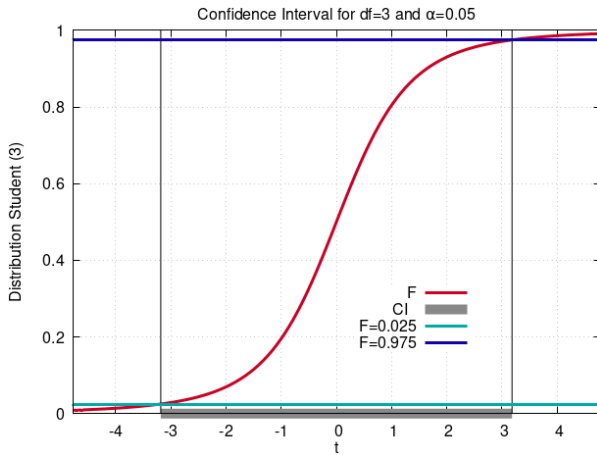
$$CI_{\beta_j}^{(\alpha)} : \beta_j \in \left[\hat{\beta}_j - \Delta\hat{\beta}_j, \hat{\beta}_j + \Delta\hat{\beta}_j \right], \quad \Delta\hat{\beta}_j = t_{1-\alpha/2}^{(n-p)} \hat{\sigma} \hat{\beta}_j.$$

- ▶ $t_{1-\alpha/2}$: Quantile (inverse of) the distribution function
- ▶ CI "uncertainty principle": Higher sensitivity implies higher α error.

Calculation of the confidence intervals (CI)

$$CI_{\beta_j}^{(\alpha)} : \beta_j \in \left[\hat{\beta}_j - \Delta \hat{\beta}_j, \hat{\beta}_j + \Delta \hat{\beta}_j \right], \quad \Delta \hat{\beta}_j = t_{1-\alpha/2}^{(n-p)} \hat{\sigma} \hat{\beta}_j.$$

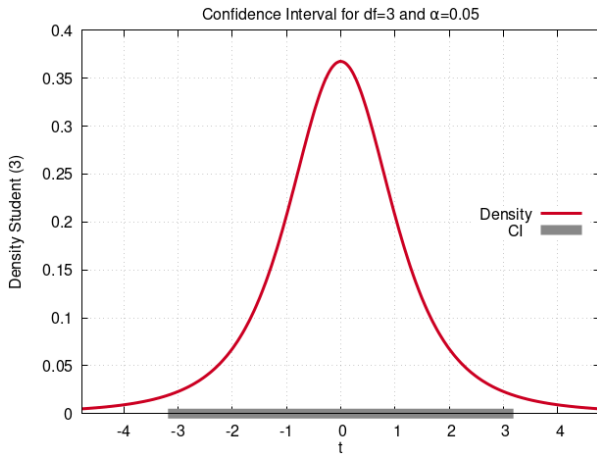
- ▶ $t_{1-\alpha/2}$: Quantile (inverse of) the distribution function
- ▶ CI "uncertainty principle": Higher sensitivity implies higher α error.



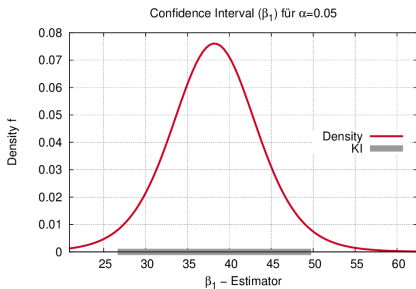
Calculation of the confidence intervals (CI)

$$CI_{\beta_j}^{(\alpha)} : \beta_j \in \left[\hat{\beta}_j - \Delta \hat{\beta}_j, \hat{\beta}_j + \Delta \hat{\beta}_j \right], \quad \Delta \hat{\beta}_j = t_{1-\alpha/2}^{(n-p)} \hat{\sigma} \hat{\beta}_j.$$

- ▶ $t_{1-\alpha/2}$: Quantile (inverse of) the distribution function
- ▶ CI “uncertainty principle”: Higher sensitivity implies higher α error.



Hotel example: CI for the appraisal for “stars” β_1 (full model)



$$\text{Model: } y(\mathbf{x}) = \sum_j \beta_j x_j + \epsilon$$

Factors:

$$x_0 = 1, x_1: \# \text{stars}, x_2: \text{price}$$

Confidence interval (CI):

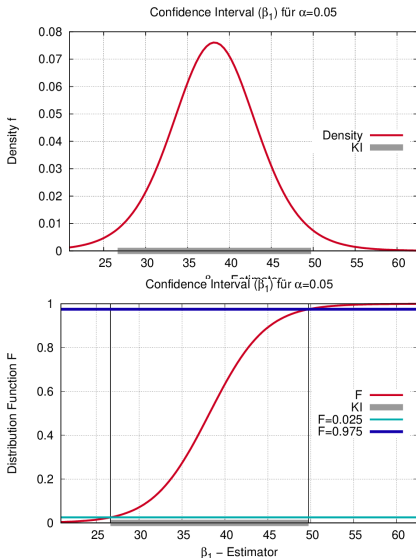
$$\beta_1 \in \left[\hat{\beta}_1 - \Delta \hat{\beta}_1^{(\alpha)}, \hat{\beta}_1 + \Delta \hat{\beta}_1^{(\alpha)} \right]$$

$$\Delta \hat{\beta}_1^{(\alpha)} = t_{1-\alpha/2}^{(n-3)} \sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}(\hat{\beta}_1) = \hat{\sigma}_\epsilon^2 \left[(\mathbf{X}'\mathbf{X})^{-1} \right]_{11}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-3} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Hotel example: CI for the appraisal for "stars" β_1 (full model)



$$\text{Model: } y(\mathbf{x}) = \sum_j \beta_j x_j + \epsilon$$

Factors:

$$x_0 = 1, x_1: \# \text{stars}, x_2: \text{price}$$

Confidence interval (CI):

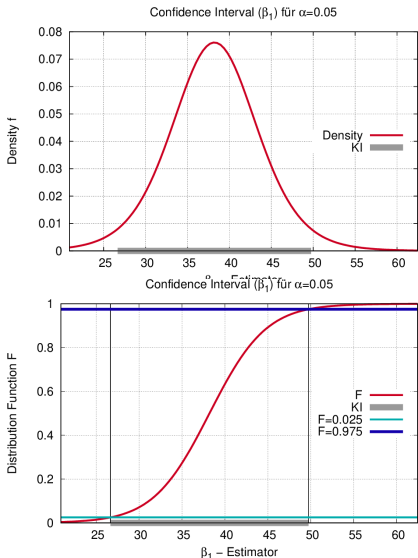
$$\beta_1 \in \left[\hat{\beta}_1 - \Delta \hat{\beta}_1^{(\alpha)}, \hat{\beta}_1 + \Delta \hat{\beta}_1^{(\alpha)} \right]$$

$$\Delta \hat{\beta}_1^{(\alpha)} = t_{1-\alpha/2}^{(n-3)} \sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}(\hat{\beta}_1) = \hat{\sigma}_\epsilon^2 \left[(\mathbf{X}'\mathbf{X})^{-1} \right]_{11}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-3} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Hotel example: CI for the appraisal for "stars" β_1 (full model)



$$\text{Model: } y(\mathbf{x}) = \sum_j \beta_j x_j + \epsilon$$

Factors:

$$x_0 = 1, x_1: \# \text{stars}, x_2: \text{price}$$

Confidence interval (CI):

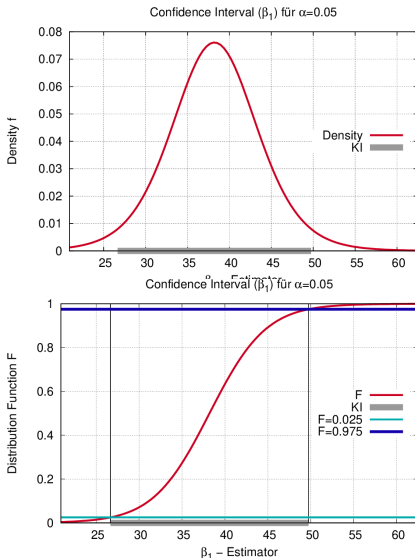
$$\beta_1 \in \left[\hat{\beta}_1 - \Delta \hat{\beta}_1^{(\alpha)}, \hat{\beta}_1 + \Delta \hat{\beta}_1^{(\alpha)} \right]$$

$$\Delta \hat{\beta}_1^{(\alpha)} = t_{1-\alpha/2}^{(n-3)} \sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}(\hat{\beta}_1) = \hat{\sigma}_\epsilon^2 \left[(\mathbf{X}'\mathbf{X})^{-1} \right]_{11}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-3} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Hotel example: CI for the appraisal for "stars" β_1 (full model)



$$\text{Model: } y(\mathbf{x}) = \sum_j \beta_j x_j + \epsilon$$

Factors:

$$x_0 = 1, x_1: \text{\#stars}, x_2: \text{price}$$

Confidence interval (CI):

$$\beta_1 \in \left[\hat{\beta}_1 - \Delta \hat{\beta}_1^{(\alpha)}, \hat{\beta}_1 + \Delta \hat{\beta}_1^{(\alpha)} \right]$$

$$\Delta \hat{\beta}_1^{(\alpha)} = t_{1-\alpha/2}^{(n-3)} \sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}(\hat{\beta}_1) = \hat{\sigma}_\epsilon^2 \left[(\mathbf{X}'\mathbf{X})^{-1} \right]_{11}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-3} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$