czechitas

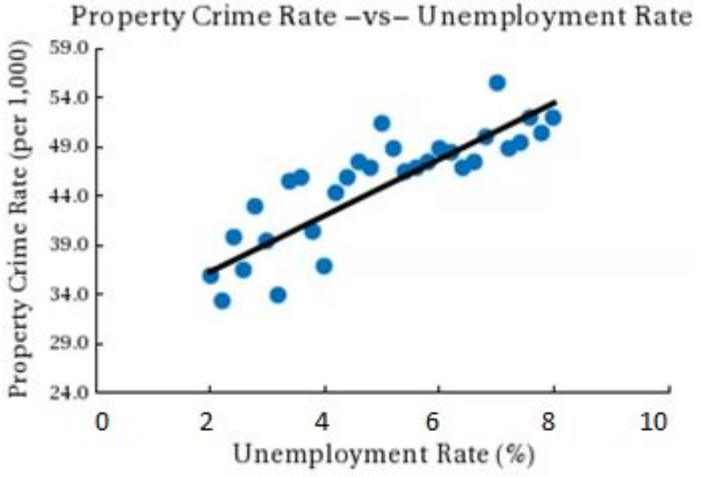
L4: Hypotheses Testing with Linear Regression - Summary

Michal Hakala, Pavel Fišer

9th of May, 2023

Mean Linear Regression

- Dependent/explained variable y_i
- Independent/explanatory variable(s) x_i
- Regression error ε_i



Scalar Notation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ \dots + \beta_k x_{ki} + \varepsilon_i$$

Matrix Notation

$$y_i = x_i' \beta + \varepsilon_i \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad x_i = \begin{pmatrix} 1 \\ x_{1i} \\ \vdots \\ x_{ki} \end{pmatrix}$$

czechitas

Test of Independence

$$crime\ rate = \beta_0 + \beta_1\, unemployment\ rate + \varepsilon$$

Null hypothesis H_0: There is no relationship between unemployment rate and crime rate.

$$H_0: \beta_1 = 0$$

We reject the null hypothesis, if our calculated β_1 *is "far enough from zero".*

czechitas

Test of Independence

Null hypothesis H_0: There is <u>no relationship</u> between unemployment rate and crime rate.

$$H_0: \beta_1 = 0$$

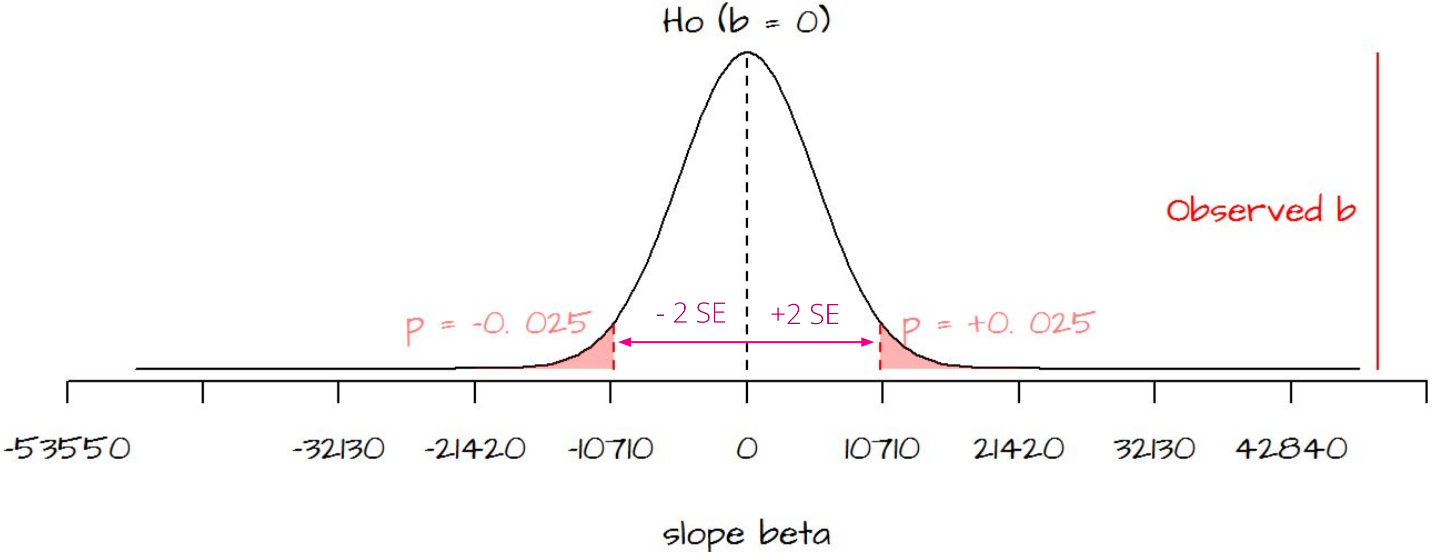
Distribution of Beta Estimates

- $\widehat{\beta_1}$ – estimate,
- β_1 – true beta (not observed)
- Test = answering question how unlikely it is to obtain $\widehat{\beta_1}$ when $\beta_1 = 0$

$$\widehat{\beta_1} \sim N(\beta_0, Standard\ Error^2)$$

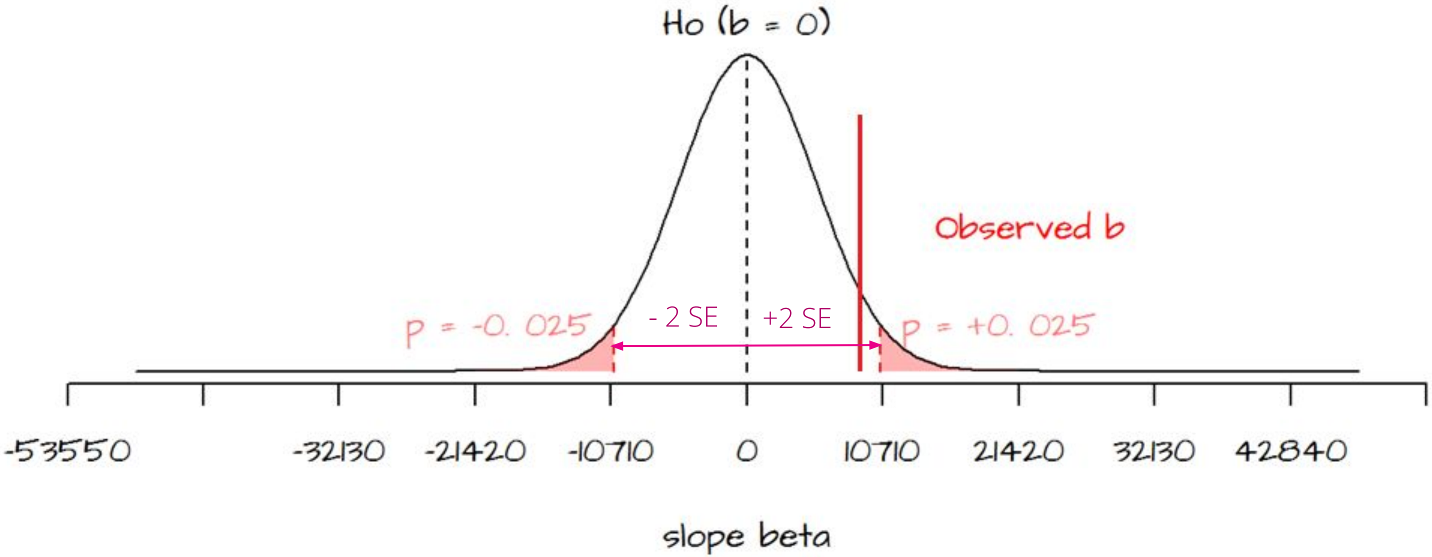
czechitas

Hypotheses Testing with OLS: Is Beta “far enough from zero”?



Here, we reject H_0 with 99,99% confidence

Hypotheses Testing with OLS: Is Beta "far enough from zero"?



And so on..

Interpretation of Results

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3526.6672	374.290	9.422	0.000	2788.515	4264.820
facebook	0.1885	0.009	21.893	0.000	0.172	0.206
newspaper	-0.0010	0.006	-0.177	0.860	-0.013	0.011
youtube	0.0458	0.001	32.809	0.000	0.043	0.049

- Coef – beta estimate
- Std err – standard error
- t – t-ratio

$$t = \frac{\hat{\beta}_j}{Standard\ Error\ (\hat{\beta}_j)} \sim N(0,1)$$

- P>|t| – p-value

“Probability of obtaining $\hat{\beta}_j$ or worse value with respect to $\mathbb{H}_0$”

4. No Omitted Variable

Causal Impact

"Impact of X on Y while everything else remains the same."

Wage Example

Consider two models

$$wage = \alpha_0 + \alpha_1 education + e$$

$$wage = \beta_0 + \beta_1 education + \beta_2 ability + u$$

- Problem: education and ability is correlated $\Rightarrow$
  α_1 captures impact of education and partially impact of ability
- "Solution": we have to add ability to the model to control it ("keep it the same")
- Omitted Variable Bias: $\alpha_1 - \beta_1$

czechitas

Assumptions Summary

We need to check that **model assumptions** hold.

1. Linear relationship → biased betas
2. No multicollinearity → high variance of beta estimates
3. Random sample → biased betas
4. No omitted variable → biased betas

---

5. Homoskedasticity → invalid inference (use robust errors)
6. Normality → invalid inference (large sample desired)

Values of betas

Validity of inference

czechitas