University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

5. Regression Models with Dummy Explanatory Variables

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2025

Basic definitions

DUMMY VARIABLES

The following names occur for such a variable:

- dummy
- binary
- dihotomous
- indicator
- artificial

Observed unit has the analysed property → Value 1

Observed unit does not have the analysed property → Value 0

Basic definitions

WARNINGS

1. **A categorical variable (attribute) with *m* possible values requires (*m* – 1) dummy variables.**

   "Dummy trap" – Perfect multicollinearity!

2. **Value of the categorical variable that is assigned value 0 for the dummy variable is called the *base* (benchmark, control, reference) *value*.**

3. **Switching the assigned values (0 to 1 and vice versa) for dummy variables does not affect the (absolute) values of regression coefficients.**

Analysis of variance models

ANALYSIS OF VARIANCE (ANOVA) MODELS

A Regression model with one attribute with two possible values

$$y_i = \beta_1 + \beta_2 D_i + u_i$$

$$E(y_i | D_i = 0) = \beta_1$$

$$E(y_i | D_i = 1) = \beta_1 + \beta_2$$

(arithmetic mean)

? $$H_0 : \beta_2 = 0 \ ; \ H_1 : \beta_2 \neq 0$$

i $$|t(b_2)| > t_c$$

Analysis of variance models

B **Regression model with one attribute with several (three) possible values**

$$y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + u_i$$

$$E(y_i | D_{1i} = 0, D_{2i} = 0) = \beta_1$$

$$E(y_i | D_{1i} = 1, D_{2i} = 0) = \beta_1 + \beta_2$$

$$E(y_i | D_{1i} = 0, D_{2i} = 1) = \beta_1 + \beta_3$$

$H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0$ $\qquad$ $H_0 : \beta_3 = 0 \; ; \; H_1 : \beta_3 \neq 0$

$|t(b_2)| > t_c$ $\qquad$ $|t(b_3)| > t_c$

Analysis of variance models

C **Regression model with two attributes with two values each**

$$y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + u_i$$

$$E(y_i | D_{1i} = 0, D_{2i} = 0) = \beta_1$$
$$E(y_i | D_{1i} = 1, D_{2i} = 0) = \beta_1 + \beta_2$$
$$E(y_i | D_{1i} = 0, D_{2i} = 1) = \beta_1 + \beta_3$$
$$E(y_i | D_{1i} = 1, D_{2i} = 1) = \beta_1 + \beta_2 + \beta_3$$

? $H_0 : \beta_2 = 0 \;;\; H_1 : \beta_2 \neq 0$ $\quad$ $H_0 : \beta_3 = 0 \;;\; H_1 : \beta_3 \neq 0$

i $|t(b_2)| > t_c$ $\quad$ $|t(b_3)| > t_c$

Analysis of covariance models

ANALYSIS OF COVARIANCE (ANCOVA) MODELS

D Regression model with one numerical explanatory variable and one attribute with two possible values

$$y_i = \beta_1 + \beta_2 D_i + \beta_3 x_{3i} + u_i$$

$$E(y_i | D_i = 0, x_{3i}) = \beta_1 + \beta_3 x_{3i}$$

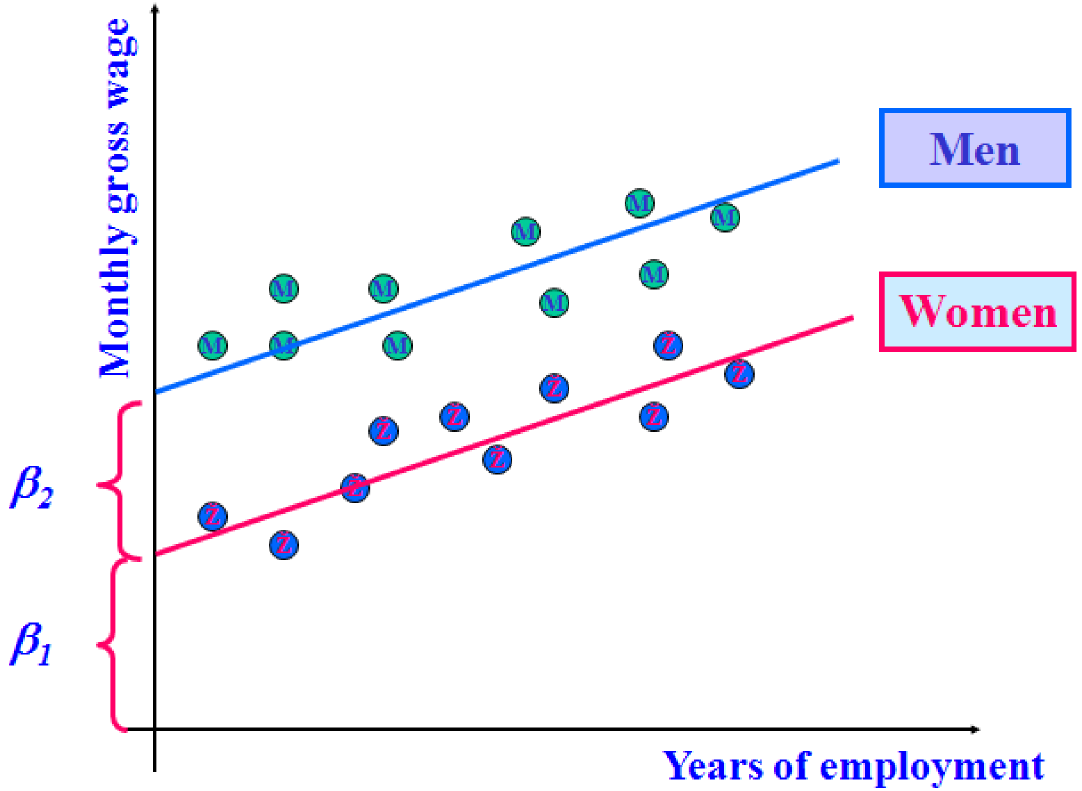
$$E(y_i | D_i = 1, x_{3i}) = (\beta_1 + \beta_2) + \beta_3 x_{3i}$$

(conditional expected value)

$$H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0$$

$$|t(b_2)| > t_c$$

Analysis of covariance models



Regression coefficient β_2 is also called the *differential intercept coefficient.*

Analysis of covariance models

E **Regression model with one numerical explanatory variable and one attribute with several (three) possible values**

$$y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 x_{4i} + u_i$$

$$E(y_i | D_{1i} = 0, D_{2i} = 0, x_{4i}) = \beta_1 + \beta_4 x_{4i}$$

$$E(y_i | D_{1i} = 1, D_{2i} = 0, x_{4i}) = (\beta_1 + \beta_2) + \beta_4 x_{4i}$$

$$E(y_i | D_{1i} = 0, D_{2i} = 1, x_{4i}) = (\beta_1 + \beta_3) + \beta_4 x_{4i}$$

? $H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0$ $\qquad$ $H_0 : \beta_3 = 0 \; ; \; H_1 : \beta_3 \neq 0$

i $|t(b_2)| > t_c$ $\qquad$ $|t(b_3)| > t_c$

Analysis of covariance models

F Regression model with one numerical explanatory variable and two attributes with two values each

$$y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 x_{4i} + u_i$$

$$E(y_i | D_{1i} = 0, D_{2i} = 0, x_{4i}) = \beta_1 + \beta_4 x_{4i}$$

$$E(y_i | D_{1i} = 1, D_{2i} = 0, x_{4i}) = (\beta_1 + \beta_2) + \beta_4 x_{4i}$$

$$E(y_i | D_{1i} = 0, D_{2i} = 1, x_{4i}) = (\beta_1 + \beta_3) + \beta_4 x_{4i}$$

$$E(y_i | D_{1i} = 1, D_{2i} = 1, x_{4i}) = (\beta_1 + \beta_2 + \beta_3) + \beta_4 x_{4i}$$

? $H_0: \beta_2 = 0 \; ; \; H_1: \beta_2 \neq 0$ $\quad$ $H_0: \beta_3 = 0 \; ; \; H_1: \beta_3 \neq 0$

i $|t(b_2)| > t_c$ $\quad$ $|t(b_3)| > t_c$

Comparison of two regression models

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

a) Chow test

$$F = \frac{(RSS - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)}$$

b) Application of dummy variables

Multiplicative or interaction dummy variable

$$y_i = \beta_1 + \beta_2 D_i + \beta_3 x_i + \beta_4 (D_i x_i) + u_i$$

$$E(y_i | D_i = 0, x_i) = \beta_1 + \beta_3 x_i$$

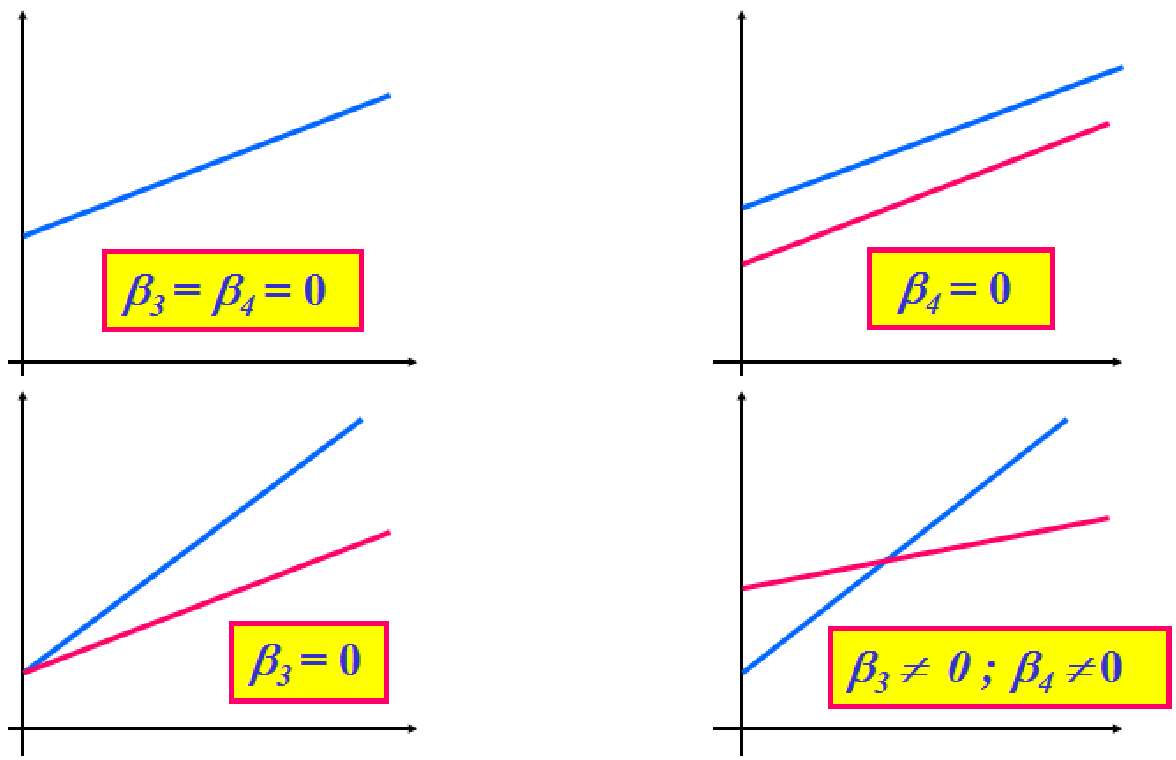
$$E(y_i | D_i = 1, x_i) = (\beta_1 + \beta_2) + (\beta_3 + \beta_4) x_i$$

$$H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0 \qquad H_0 : \beta_4 = 0 \; ; \; H_1 : \beta_4 \neq 0$$

Comparison of two regression models

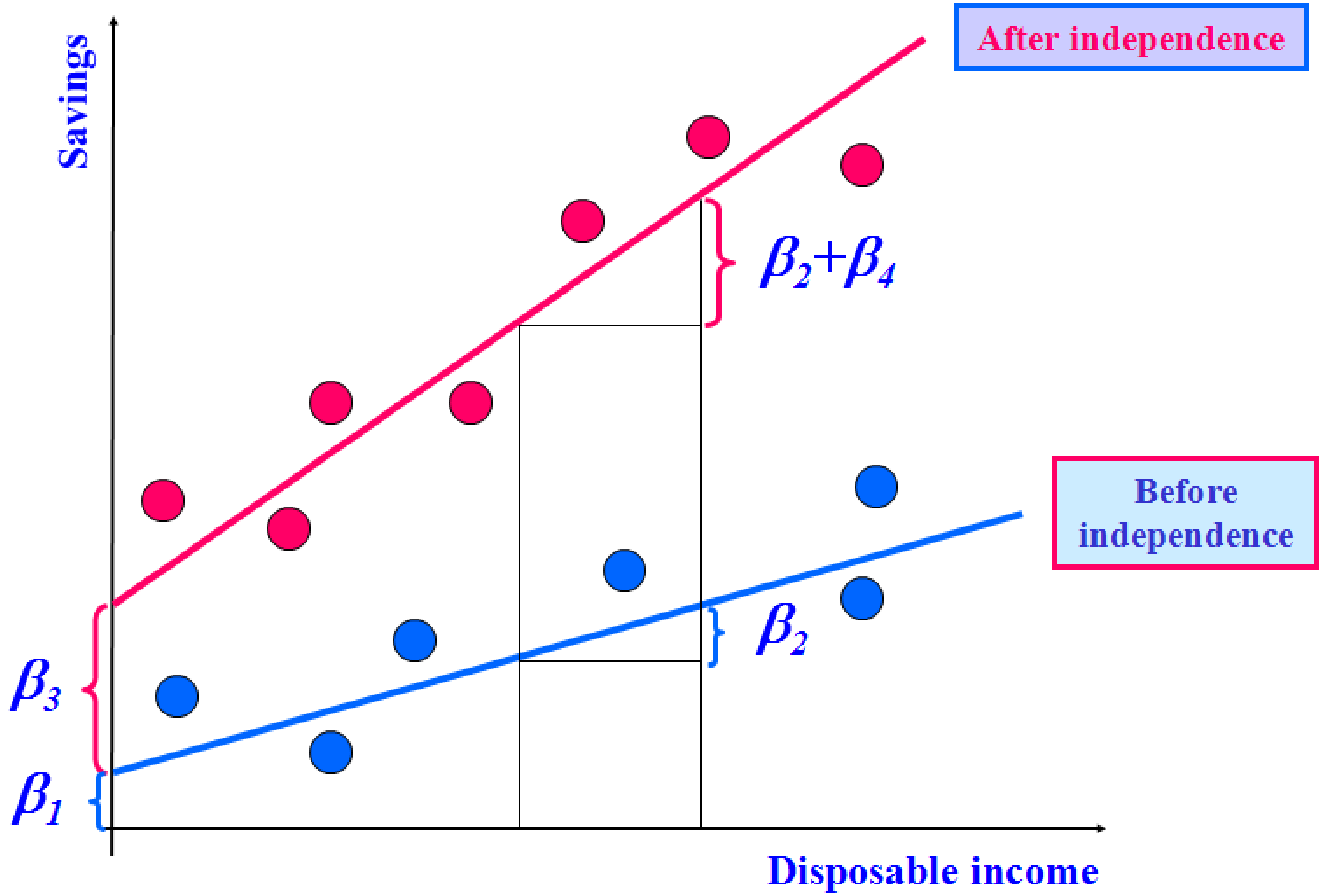
Application of one dummy variable offers four possibilities (models)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 D_t + \beta_4 (x_{2t} D_t) + u_t$$



EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION OF MBAS AMBA ACCREDITED

Comparison of two regression models



Comparison of two regression models

Advantages of dummy variables compared to statistical tests (Chow test)

1. **Only one regression model is estimated**

2. **Differences in one or more regression coefficients allowed**

   β_3 – differential intercept coefficient
   β_4 – differential slope coefficient

3. **No degrees of freedom are lost when analysing statistical significance**

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

5. Regression Models with Dummy Explanatory Variables

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2025