Contents

- Example: Intelligence, Lead Pollution and Occupational Status
- Multiple Linear Regression with Two Regressors
- Conditional Regression
- The Properties of the Residual
- Identification of the Regression Coefficients
- Dichotomous Regressors
- Simple Regression and Multiple Linear Regression With Two Regressors
- Linear Quasi-Regression

Example: Intelligence, Lead Pollution and Occupational Status

Does lead pollution of the environment cause a lower intelligence of children?

The study shows a negative correlation (-.14) between the *logarithm of the lead* (X) and the *verbal IQ* (Y) in the sample. If we assume a *linear* regressive dependency of the regressand Y from the regressor X, then we can write the dependency as:

$$E(Y|X) = \alpha_0 + \alpha_1 X$$

where $\alpha_1 < 0$.

Example: Intelligence, Lead Pollution and Occupational Status

Let's consider a third variable, the *occupational status of the parents* of the children, denoted with Z. Now we can ask if there is still a linear regressive dependency between the variables Y and X given a value z of Z. That is: Is there a partial linear regressive dependency of the verbal IQ from the logarithm of the amount of lead with respect to Z. The analysis of the data indicated that Y is with respect to Z linear regressive *independent* of X, i.e.:

$$E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z$$

where

$$\beta_1 = 0.$$

Multiple Linear Regression With Two Regressors: Definition

Definition 9.1

Let X, Y and Z be one-dimensional numerical random variables of a common probability space, and all with a positive and finite variance. The regression $E(Y|X, Z)$ is then called *linear in* (X, Z) if

$$E(Y \mid X, Z) = \beta_0 + \beta_1 X + \beta_2 Z, \quad \beta_0, \beta_1, \beta_2 \in \mathrm{IR}.$$

Multiple Linear Regression With Two Regressors: Figure I 5

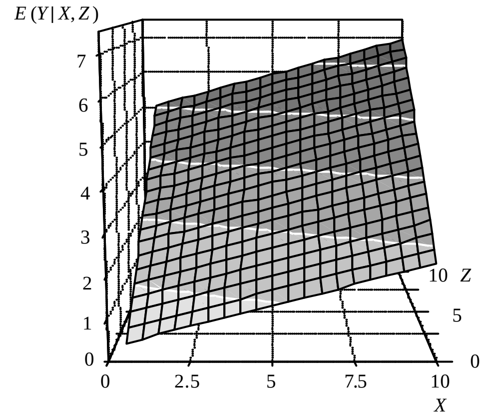

Figure 9.1. Graph of the regression $E(Y|\, X, Z\,) = 0.3 + 0.2 \cdot X + 0.5 \cdot Z$. Every point is (with continuous regressors X and Z) a value $E(Y \mid X = x, Z = z)$ of the regression $E(\,Y| X, Z\,)$.

Conditional Regressions 6

$$E_{Z=z}(Y \mid X) = (\beta_0 + \beta_2\, z) + \beta_1 X,$$

Multiple Linear Regression with Two Regressors: Figure II

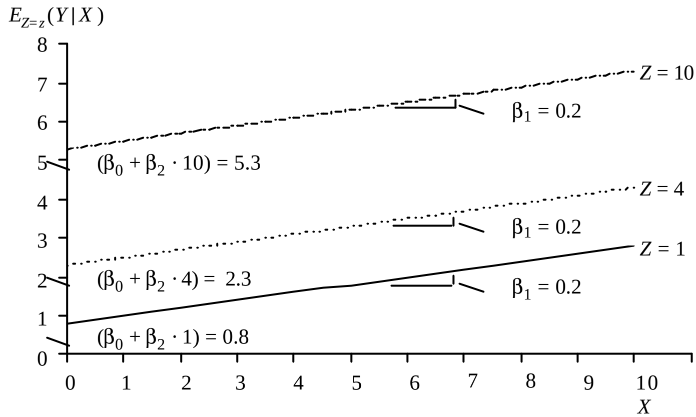

Figure 9.2. The conditional regression lines given partial linear regressiv dependency of the regressand Y of the regressor X with respect to Z. As well as in the preceding figure the regression is based on $E(Y \mid X, Z) = 0.3 + 0.2 \cdot X + 0.5 \cdot Z$.

The Coefficient of Determination and Practical Significance

$$R^2_{Y \mid X, Z} - R^2_{Y \mid Z}$$

$$R^2_{Y \mid X, Z} = \frac{\beta_1^2 Var(X) + \beta_2^2 Var(Z) + 2\beta_1\beta_2 Cov(X, Z)}{Var(Y)}$$

The Properties of the Residual

The residual is defined as:

$$e := Y - E(Y \mid X, Z)$$

All general properties as discussed earlier hold. E.g.

$$E(e \mid X, Z) = 0 \text{ and } E(e) = 0,$$

especially

$$E(e \mid X) = E(e \mid Z) = 0,$$

and

$$Cov(e, X) = Cov(e, Z) = 0.$$

Identification Formulas: General

$$\beta_0 = E(Y) - \beta_1 E(X) - \beta_2 E(Z)$$

$$\beta_1 = \frac{Var(Z)\,Cov(X,Y) - Cov(X,Z)\,Cov(Y,Z)}{Var(X)\,Var(Z) - Cov(X,Z)^2}$$

$$= \frac{Std(Y)}{Std(X)} \cdot \frac{Kor(X,Y) - Kor(X,Z)\,Kor(Y,Z)}{1 - Kor(X,Z)^2}$$

$$\beta_2 = \frac{Var(X)\,Cov(Z,Y) - Cov(X,Z)\,Cov(Y,X)}{Var(X)\,Var(Z) - Cov(X,Z)^2}$$

$$= \frac{Std(Y)}{Std(Z)} \cdot \frac{Kor(Z,Y) - Kor(X,Z)\,Kor(X,Y)}{1 - Kor(X,Z)^2}$$

Identifications: Dichotomous Regressors

If the regressors X and Z are both dichotomous, then the parameters β_0, β_1 und β_2 can be easily computed, because the following equations can be derived:

$$E(Y \mid X = 1, Z = 1) = \beta_0 + \beta_1 + \beta_2$$

$$E(Y \mid X = 1, Z = 0) = \beta_0 + \beta_1$$

$$E(Y \mid X = 0, Z = 1) = \beta_0 + \beta_2$$

$$E(Y \mid X = 0, Z = 0) = \beta_0$$

Solving this linear system of equations leads to

$$\beta_1 = E(Y \mid X = 1, Z = 0) - E(Y \mid X = 0, Z = 0)$$

and

$$\beta_2 = E(Y \mid X = 0, Z = 1) - E(Y \mid X = 0, Z = 0)$$

Simple Regression and Multiple Linear Regression With Two Regressors: I

Theorem 9.1. Given the assumptions of Definition 9.1. the following fact is true: If

$$E(Y \mid X, Z) = \beta_0 + \beta_1 X + \beta_2 Z, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R},$$

and $\beta_2 = 0$ or Z is regressive independent from X, i.e.:

$$E(Z \mid X) = E(Z),$$

then

$$E(Y \mid X) = \alpha_0 + \alpha_1 X,$$

where $\alpha_0 = \beta_0 + \beta_2 E(Z)$ and $\alpha_1 = \beta_1$ and further

$$Cov(X, Z) = 0,$$

$$Var[E(Y|X, Z)] = \beta_1^2\, Var(X) + \beta_2^2\, Var(Z).$$

Simple Regression and Multiple Linear Regression With Two Regressors: II

If the additional condition

$$E(X|\, Z) = E(X),$$

is also met, then

$$Var[E(Y|\, X, Z)] = Var[E(Y \mid X)] + Var[E(Y \mid Z)],$$

$$R^2_{Y|X,Z} = R^2_{Y|X} + R^2_{Y|Z} \ ,$$

$$R^2_{Y|X} = Kor(X,Y)^2$$

$$R^2_{Y|Z} = Kor(Y,Z)^2$$

Simple Regression and Multiple Linear Regression With Two Regressors III

If you can not assume that Z is regressively independent from X but the following equation is true

$E(Z|\, X) = \gamma_0 + \gamma_1 X$,

Then the equation

$$E(Y \mid X) = \alpha_0 + \alpha_1 X,$$

is still true, where $\alpha_0 := \beta_0 + \beta_2\gamma_0$ and $\alpha_1 := \beta_1 + \beta_2\gamma_1$, i.e.

$$E(Y|\, X) = (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)\, X.$$

Multiple Linear Quasi-Regression: I

Definition 9.2. With the assumptions of Definition 9.1. we define the *multiple linear quasi-regression with two regressors,* denoted by $Q(Y \mid X, Z)$, as the linear combination $\beta_0 + \beta_1 X + \beta_2 Z$ of X and Z, which fulfils the following equations:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + n\,,$$

$$E(n) = 0\,,$$

and

$$Cov(n, X) = Cov(n, Z) = 0.$$

Multiple Linear Quasi-Regression: II

Definition 9.3. With the assumptions of Definition 9.1. we can also define $Q(Y \mid X, Z)$, as the linear combination $\beta_0 + \beta_1 X + \beta_2 Z$ of X and Z, which minimizes the following equations (the least square criterion):

$$LS(b_0, b_1, b_2) = E[[Y - (b_0 + b_1 X + b_2 Z)]^2].$$

Those numbers b_0, b_1 and b_2, for which the function $LS(b_0, b_1, b_2)$ is minimal, are denoted as β_0, β_1 und β_2 respectively. The *multiple linear quasi-regression with two regressors* is then defined by:

$$Q(Y \mid X, Z) := \beta_0 + \beta_1 X + \beta_2 Z$$