# Verkehrsökonometrie für Bachelor-Studierende

## Sommersemester 2024, Lösungsvorschläge zu Übung Nr. 13

### Lösungsvorschlag zu Aufgabe 13.1: ÖPNV-Nutzungshäufigkeit – Bivariate lineare Regression

Das bivariate lineare Modell lautet

$$\hat{y}(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \tag{1}$$

Die exogenen Variablen sind $x_1$ (Fahrpreis pro Fahrt in €) und $x_2$ (effektive Geschwindigkeit in km/h). Die endogene Variable $y$ (bzw. der Schätzer $\hat{y}$ für die lineare Regressionsfunktion) beschreibt die Nutzungshäufigkeit des ÖPNV (Fahrten pro Jahr). Es liegen Realisierungen aus $n = 5$ Städten vor.

(a) Die Minimierung des mittleren quadratischen Fehlers (*"least square", LQ*) des bivariaten Modells führt auf das lineare Gleichungssystem

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} s_{1y} \\ s_{2y} \end{pmatrix} \tag{2}$$

mit der allgemeinen Lösung

$$\hat{\beta}_1 = \frac{s_{22}s_{1y} - s_{12}s_{2y}}{\det S}, \tag{3}$$

$$\hat{\beta}_2 = \frac{s_{11}s_{2y} - s_{21}s_{1y}}{\det S}. \tag{4}$$

Die Matrixelemente der Varianz-Kovarianz-Matrix der exogenen Variablen lauten (ggf. den Verschiebungssatz nutzen, wie in der ersten Zeile gezeigt):

$$s_{11} = \frac{1}{n}\sum_{i=1}^{5}(x_{1i} - \bar{x}_1)^2 = \frac{1}{n}\sum_{i=1}^{5}\left(x_{1i}^2 - \bar{x}_1^2\right) = \underline{\underline{0.416}}, \tag{5}$$

$$s_{22} = \frac{1}{n}\sum_{i=1}^{5}(x_{2i} - \bar{x}_2)^2 = \underline{\underline{14.0}}, \tag{6}$$

$$s_{12} = s_{21} = \frac{1}{n}\sum_{i=1}^{5}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) = \underline{\underline{1.60}}. \tag{7}$$

Die eingehenden Mittelwerte sind

$$\bar{x}_1 = \frac{1}{n}\sum_i x_{1i} = \underline{\underline{2.40\,€}} \tag{8}$$

$$\bar{x}_2 = \frac{1}{n}\sum_i x_{2i} = \underline{\underline{24\,\mathrm{km/h}}}\,. \tag{9}$$

Das Matrixelement $s_{21}$ erhält man durch Vertauschen der beiden Faktoren, so dass $s_{12} = s_{21}$ ist.

Der Kovarianzvektor der abhängigen Variable $y$ mit den unabhängigen Variablen $x_1$ und $x_2$ lautet

$$s_{1y} = \frac{1}{n}\sum_{i=1}^{5}(x_{1i} - \bar{x}_1)(y_i - \bar{y}) = \underline{\underline{-13.2}}\,, \tag{10}$$

$$s_{2y} = \frac{1}{n}\sum_{i=1}^{5}(x_{2i} - \bar{x}_2)(y_i - \bar{y}) = \underline{\underline{0.00}}\,, \tag{11}$$

mit $\bar{x}_1 = 2.40$ und $\bar{x}_2 = 24$ von oben und

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n=5} y_i = \underline{\underline{200\,\mathrm{Fahrten/Jahr}}}\,. \tag{12}$$

(b) Die Anstiegsparameter $\beta_1$ und $\beta_2$ ergeben sich aus der Lösung (3) und (4) des linearen Gleichungssystems (2).

Die Determinante der $2 \times 2$-Matrix ist

$$\det S = s_{11}s_{22} - s_{12}s_{21} = 3.264\,. \tag{13}$$

Damit ergeben sich die linearen Anstiegsparameter zu

$$\hat{\beta}_1 = \frac{14 \cdot (-13.2) - 1.6 \cdot 0}{3.264} = \underline{\underline{-56.62}}\,, \tag{14}$$

$$\hat{\beta}_2 = \frac{0.416 \cdot 0 - 1.6 \cdot (-13.2)}{3.264} = \underline{\underline{6.471}}\,. \tag{15}$$

Ein Anstieg um 1 € im Preis führt also zu $\approx 56.6$ Fahrten pro Jahr *weniger*. Eine Geschwindigkeitserhöhung führt hingegen zu $\approx 6.5$ Fahrten *mehr*.

Eine Preiserhöhung um 1 € würde also im Mittel zu einer relativen Verringerung der Jahresfahrten um den Faktor $\frac{\bar{y}-56.6}{\bar{y}} = 0.72$ führen. Dies entspricht einem Rückgang um $1 - 0.72 = 0.28$ bzw. 28%.

Um den Verlust an Fahrten aus 1 € Preiserhöhung auszugleichen, müßte die Geschwindigkeit um

$$-\frac{\hat{\beta}_1}{\hat{\beta}_2} = 8.75\,\mathrm{km/h} \tag{16}$$

zunehmen.

Die Elastizitäten $\epsilon_1$ und $\epsilon_2$ sind definiert als

$$\epsilon_{1/2} = \frac{\text{relative Änderung von } y}{\text{relative Änderung von } x_1 \text{ bzw. } x_2}$$

im Grenzfall sehr kleiner Änderungen. Sie werden am Punkt $(\bar{x}_1, \bar{x}_2, \bar{y})$ ausgewertet, so dass es sich um "Punktelastizitäten" handelt:

$$\epsilon_1 = \frac{\bar{x}_1}{\bar{y}} \frac{\partial \hat{y}}{\partial x_1}, \qquad \epsilon_2 = \frac{\bar{x}_2}{\bar{y}} \frac{\partial \hat{y}}{\partial x_2}. \tag{17}$$

Diese Elastizitäten sind per Definition einheitenlos, was, in Anbetracht der komplexen Einheiten der nichtskalierten Sensitivitäten, z.B.

$$[\beta_1] = \frac{\text{Fahrten}^2}{\text{Euro Jahr}}$$

sehr vorteilhaft und vergleichsweise anschaulich ist. Die Ableitungen des *linearen* Modells (1) sind die Anstiegsparameter $\beta_i$, so dass sich ergibt:

$$\epsilon_1 = \frac{\bar{x}_1}{\bar{y}} \hat{\beta}_1 = \frac{2.40}{200} \cdot (-56.6) = \underline{\underline{-0.68}}, \tag{18}$$

$$\epsilon_2 = \frac{\bar{x}_2}{\bar{y}} \hat{\beta}_2 = \frac{24.0}{200} \cdot 6.47 = \underline{\underline{0.78}}. \tag{19}$$

Da in beiden Fällen $|\epsilon| < 1$ ist, ändert sich $y$ relativ weniger stark als die $x_i$ (auch: "unterproportional" oder "y ist elastisch").

Näherungsweise ergibt eine Erhöhung der Ticketkosten um 1% eine Abnahme der Fahrten um 0.68%, wohingegen eine 1%ige Erhöhung der mittleren Geschwindigkeit zu 0.78% mehr Fahrten führt.

Wenn man die Geschwindigkeit in eine *Zeit* umrechnet, kann man auf einen "Zeit-Geld-Umrechnungsfaktor" schließen. Aus $\hat{\beta}_1 dx_1 + \hat{\beta}_2 dx_2 = 0$ ergibt sich

$$\frac{dx_1}{dx_2} = -\frac{\hat{\beta}_2}{\hat{\beta}_1} = 0.114 \, \frac{€}{\text{km/h}}. \tag{20}$$

Angenommen wird nun, dass eine Erhöhung der Geschwindigkeit um 1 km/h einer Zeitersparnis von 1 min entspricht. Damit bekommt man den "Zeit-ist-Geld-Faktor"

$$0.114 \frac{€}{\text{min}} = \underline{\underline{6.86 \, \frac{€}{\text{h}}}}. \tag{21}$$

(c) Die univariaten Modelle lauten

$$\hat{v}(x_1) = \hat{\beta}_0^* + \hat{\beta}_1^* x_1, \tag{22}$$

$$\hat{w}(x_2) = \hat{\beta}_0' + \hat{\beta}_2' x_2. \tag{23}$$

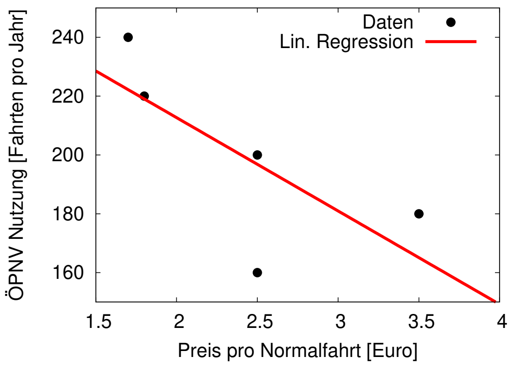
Die Anstiegsparameter ergeben sich aus (vgl. Statistik II)

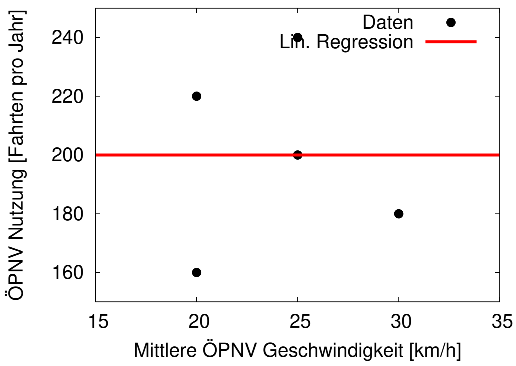
$$\hat{\beta}_1^* = \frac{s_{1y}}{s_{11}} \quad \text{bzw.} \quad \hat{\beta}_2' = \frac{s_{2y}}{s_{22}}, \tag{24}$$

mit den Zahlenwerten

$$\hat{\beta}_1^* = \frac{-13.2}{0.416} = \underline{\underline{-31.7}}, \tag{25}$$

$$\hat{\beta}_2' = \frac{0.0}{14.0} = \underline{\underline{0.0}}. \tag{26}$$




Die Anstiegs- bzw. Abfallrate für den Preis ist $\approx$ nur halb so groß wie im bivariaten Modell. Weiterhin zeigt das univariate Modell quasi keinerlei Abhängigkeit von der mittleren Geschwindigkeit. Konsequenterweise sind nun die Preis- und Zeitelastizitäten

$$\epsilon_1^* = -0.35, \quad \tilde{\epsilon}_2 = 0$$

statt $\epsilon_1 = -0.68$ und $\epsilon_2 = 0.78$.

Offensichtlich entspricht im Modell (23) die durch $\hat{\beta}_2' = 0$ beschriebene fehlende statistische Abhängigkeit der Nutzungshäufigkeit von der Geschwindigkeit nicht den tatsächlichen Begebenheiten. Dies ist durch die zwei folgenden Tatsachen zu erklären:

- Ein weiterer relevanter Einflussfaktor (hier der Fahrpreis) wird nicht berücksichtigt.
- Es gibt eine lineare Abhängigkeit der berücksichtigten und der nichtberücksichtigten Variablen, welche durch die positive Kovarianz $s_{12}$ bzw. den positiven Korrelationskoeffizienten
$$\rho_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = 0.66 \tag{27}$$
beschrieben wird.

Beide Tatsachen sind notwendig: Hätten die unabhängigen Variablen keine Korrelation, enthielte man durch Modell (23) zwar keine Information über die Fahrpreisabhängigkeit, die Geschwindigkeitsabhängigkeit würde aber korrekt wiedergegeben. Wäre andererseits der Fahrpreis irrelevant, würde auch eine Korrelation nicht stören. Hier jedoch kompensieren die tendenziell bei besseren ÖV-Systemen teureren Fahrpreise gerade die höhere Attraktivität und man erhält eine falsche Aussage!

Diese Fehlerquelle gilt allgemein, da bessere Produkte im Allgemeinen teurer sind. Es ist deshalb wichtig, möglichst alle relevanten Einflussfaktoren (unabhängige Variablen) zu erfassen.

Ansonsten ist das Problem „fehlspezifiziert". Allgemein liegt eine Fehlspezifikation vor, wenn:

(1) eine wesentliche erklärende Variable nicht berücksichtigt wurde (wie hier),

(2) eine 100%ig mit anderen Variablen oder Linearkombinationen davon korrelierte erklärende Variable mitgeführt wird oder

(3) eine irrelevante erklärende Variable mitgeführt wird.

Der letztere Fall ist vergleichsweise harmlos. Der entsprechende Anstiegsparameter ist dann nicht signifikant von null verschieden. Der zweite Fall führt zu viel größeren Streubreiten, aber zu keiner systematischen Verfälschung des Mittelwertes. Fall (1) hingegen kann genau dies bewirken, wie diese Aufgabe zeigte. Er ist deshalb die gefährlichste Art der Fehlspezifikation.

## Zur Information bzw. Ausblick Master-VL "Methoden Ökonometrie"

Fehlergrenzen des korrekt spezifizierten Modells mit zwei exogenen Variablen