Universität Duisburg-Essen
Lehrstuhl für Ökonometrie
Dr. Yannick Hoga
MSc. Martin Arnold

Wintersemester 2022/23

Übungsblatt 4 — Einfaches Regressionsmodell II

Zu diesem Übungsblatt empfehlen wir neben der Lektüre von Kapitel 4 des Lehrbuches *Introduction to Econometrics* von *Stock & Watson* eine Aufarbeitung mithilfe der Kapitel 4 in unserem Online-Companion *Introduction to Econometrics with R*.

Aufgabe 3 — Anwendungsbeispiel: Lohnregression

Eine Regression des Wochenlohns (WL, in Euro) auf das Alter (in Jahren), welcher eine Zufallsstichprobe von vollzeitbeschäftigten Hochschulabsolventen im Alter von 25 bis 65 Jahren zugrunde liegt, lieferte folgendes Ergebnis:

$$\widehat{WL} = 696.7 + 9.6 \times Alter, \quad R^2 = 0.023, \quad SER = 624.1$$

(a) Erläutern Sie die Beudeutung der Werte 696.7 und 9.6.

Lösung:

9.6: Der Wochenlohn steigt um geschätzte 9.6 € pro Jahr an Lebensalter.

696.7: Der Y-Achsenabschnitt der Regressionsgerade ist hier der geschätzte Wochenlohn für einen 0-Jährigen).

(b) Der Standardfehler der Regression (standard error of the regression, SER) beträgt 624.1. Welche Einheit besitzt dieser? (Euro? Jahre? Keine Einheit?)

Lösung:

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} \widehat{u}_i^2}$$

Der SER hat folglich stets dieselbe Einheit wie die abhängige Variable (hier €).

(c) Der Anpassungskoeffizient R^2 für die Regression beträgt 0.023. Welche Einheit besitzt dieser?

Lösung:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} = 0.023,$$

d.h. 2.3% der beobachteten Variation in den Wochenlöhnen wird druch das Modell erklärt. R^2 ist stets ein *Anteil* und hat folglich keine Einheit.

(d) Welcher geschätzte Wochenlohn ergibt sich für einen 25 bzw. für eine/n 45 Jahre alten Vollzeitbeschäftigte/n aus der Regression?

Lösung:

25 Jahre: $696.7 + 9.6 \cdot 25 = 936.7$€

45 Jahre: $696.7 + 9.6 \cdot 45 = 1128.7$€

(e) Wird die Regression eine zuverlässige Schätzung für den Wochenlohn von 99-jährigen Vollzeitbeschäftigten liefern?

Lösung:

Nein, denn der Datensatz beruht auf Erhebungen von 25-65 jährigen vollzeitarbeitenden Hochschulabsolventinnen, d.h. unsere Stichprobe bildet diesen Teil der Population nicht ab.

(f) Sind die Fehler normalverteilt? Benutzen Sie Ihr Wissen über die Einkommensverteilung!

Lösung:

Vermutlich nicht. Einkommsverteilungen sind rechtsschief und besitzten ein höhere Wölbung als Normalverteilungen. Häufig wird für Einkommensdaten eine logarithmische Normalverteilung angenommen.

(g) Das Durchschnittsalter der Vollzeitbeschäftigten in der Stichprobe beträgt 41.6 Jahre. Was ist der Durschnittswert für den Wochenlohn?

Lösung:

Wir nutzen die Schwerpunkteigenschaft der KQ-Schätzung und erhalten $\overline{WL} = 1096.06$. Es ist

$$\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{X} = 696.7 + 9.6 \cdot 41.6 = 1096.06.$$

(h) Wie könnte das Modell unter Umständen noch verbessert werden?

Lösung:

- WL logarithmieren, d.h. $\log(WL) = \beta_0 + \beta_1 \times \mathit{Alter} + u$ schätzen.
- Bspw. $\sqrt{\mathit{Alter}}$ statt *Alter* als Regressor nutzen:

  Der Zusammenhang zwischen Alter und Wochenlohn wird vermutlich nicht-linear sein. Abnehmender Übungseffekt: Lohnzuwachs nimmt mit Lebensalter ab.

R-Hausaufgabe

Zeigen Sie, wie man die gefitteten Werte des in Teilaufgabe 1 (e) geschätzten quadratischen Modells

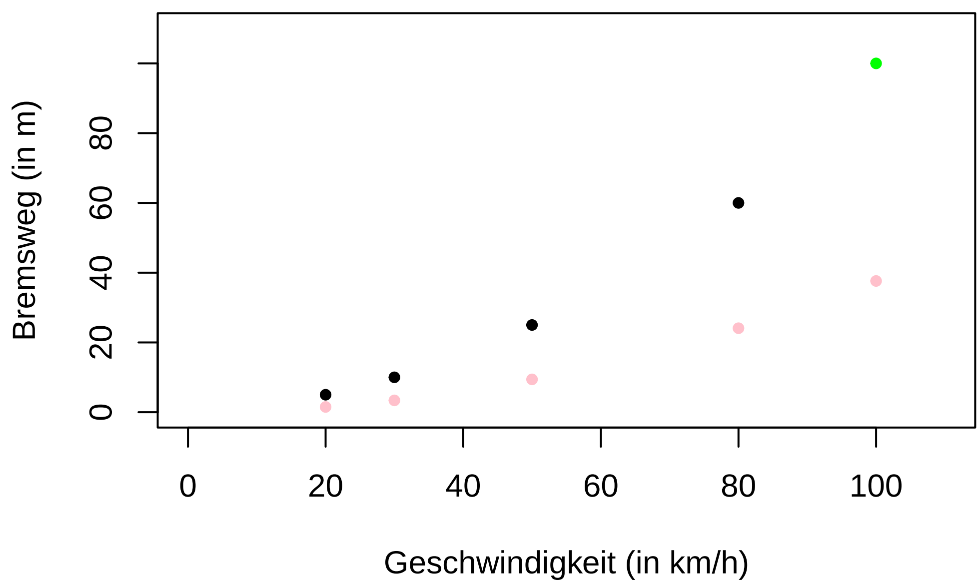
$$\widehat{\text{Bremsweg}}_i = 0.009836 \cdot \text{Geschwindigkeit}_i^2$$

in den Plot aus (d) einzeichnen kann.

Lösung:

```r
# Vektoren definieren
Geschwindigkeit <- c(20, 30, 50, 80, 100)
Bremsweg <- c(5, 10, 25, 60, 10)

# Modell schätzen
mod4 <- lm(Bremsweg ~ I(Geschwindigkeit^2) - 1)

# Gefittete Werte des Modells
```

```r
Bremsweg_fitted <- fitted(mod4)

# Beobachtungen ohne Ausreißer plotten (von (d) kopiert)
plot(Geschwindigkeit[-5], Bremsweg[-5], pch = 20,
         ylim = c(0, 110),
         xlim = c(0, 110),
         ylab = "Bremsweg (in m)",
         xlab = "Geschwindigkeit (in km/h)")

# Korrigierte Beobachtung einzeichnen
points(100, 100,
       col = "Green",
       pch = 20)

# Geschätzte Bremswege einzeichnen
points(Geschwindigkeit, Bremsweg_fitted, col = "pink", pch = 20)
```




```r
#### mit lines() Vorhersagen auf Basis des korrigierten Datensatzes einzeichnen ####

# Datensatz korrigieren
Geschwindigkeit <- c(20, 30, 50, 80, 100)
Bremsweg <- c(5, 10, 25, 60, 100)

# Beobachtungen plotten
plot(Geschwindigkeit, Bremsweg, pch = 20,
         ylim = c(0, 110),
         xlim = c(0, 110),
         ylab = "Bremsweg (in m)",
         xlab = "Geschwindigkeit (in km/h)")

# Modell schätzen
mod_korrigiert <- lm(Bremsweg ~ I(Geschwindigkeit^2) - 1)
```

```r
# Geschwindigkeiten 0,1,2,...,100 km/h
G <- seq(0, 100, 1)

# Vorhersagen des Bremswegs für Vektor G
B <- predict(mod_korrigiert, new = data.frame("Geschwindigkeit" = G))

# Linie, die Datenpunkte verbindet, einzeichnen
lines(x = G, y = B, lwd = 1)
```