Universität Ulm | 89069 Ulm | Germany

Fakultät für Mathematik und Wirtschaftswissenschaften

Dipl. WiWi. Sabrina Böck
Institut für Wirtschaftspolitik

Ludwig-Erhard-Stiftungsprofessur

Wintersemester 2008/2009

# Tutorium zur Makroökonomik

# Das lineare Regressionsmodell

## 1 Beispiel

Hier wird die lineare Einfachregression anhand eines Beispiels dargestellt.

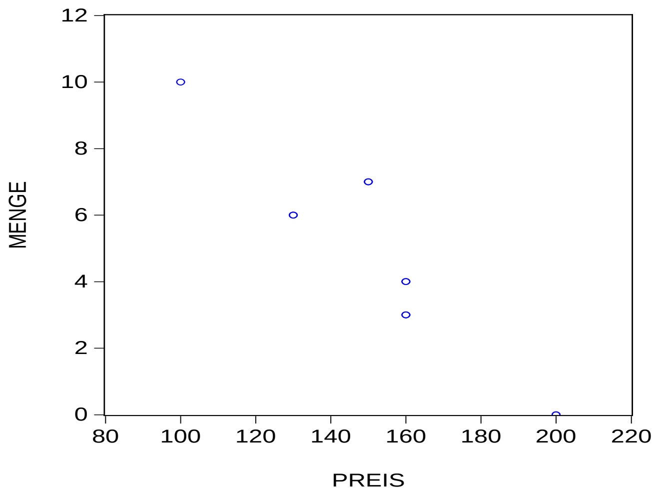
Ein renommierter Juvelier möchte eine neue Zielgruppe ansprechen und Schmuck speziell für Studenten auf den Markt bringen. Für die Festlegung des Abgabepreises soll zunächst eine Preis-Absatz-Funktion ermittelt werden. Dazu wurde in n = 6 Geschäften ein Testverkauf durchgeführt. Man erhielt sechs Wertepaare mit dem Ladenpreis x (in Euro) eines Schmuckstücks und der verkauften Menge y an Schmuckstücken:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Laden | $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
| Preis eines Schmuckstücks | $x_i$ | 200 | 160 | 150 | 160 | 130 | 100 |
| Verkaufte Menge | $y_i$ | 0 | 3 | 7 | 4 | 6 | 10 |

Helmholtzstr. 20, Raum E 02
Tel. 0731 50-24265, Fax-24262

http://www.mathematik.uni-ulm.de/wipo
Sabrina.Böck@uni-ulm.de

Als Streudiagramm von Preis und abgesetzter Menge an Schmuckstücken ergibt sich folgende Grafik.




**Berechnung der Regressionsgeraden**

Man geht von folgendem statistischen Modell aus:

Man betrachtet zwei Variablen y und x, die vermutlich ungefähr in einem linearen Zusammenhang stehen:

$Y \approx \alpha + \beta x.$

Auf die Vermutung des linearen Zusammenhangs kommt man, wenn man das obige Streudiagramm betrachtet, dort erkennt man, dass die eingetragenen Punkte nahezu auf einer Linie liegen.

Im Weiteren sind $x$ als unabhängige und $Y$ als abhängige Variable definiert. Es existieren von $x$ und $y$ je n Beobachtungen $x_i$ und $y_i$, wobei $i$ von 1 bis $n$ geht. Der funktionale Zusammenhang $Y = f(x)$ zwischen $x$ und $Y$ kann nicht exakt festgestellt werden, da $\alpha + \beta x$ von einer *Störgröße* $\epsilon$ überlagert wird. Diese Störgröße ist als Zufallsvariable (der Grundgesamtheit) konzipiert, die nichterfassbare Einflüsse (menschliches Verhalten oder Messungenauigkeiten oder ähnliches) darstellt.

$Y \approx \alpha + \beta x + \epsilon$ oder genauer $y_i = \alpha + \beta x_i + \epsilon_i$.

Da $\alpha$ und $\beta$ nicht bekannt sind, kann $y$ nicht in die Komponenten $\alpha + \beta x$ und $\epsilon$ zerlegt werden. Des Weiteren soll eine Schätzung für die Parameter $\alpha$ und $\beta$ durch $\hat{\alpha}$ und $\hat{\beta}$ gefunden werden, damit ergibt sich

$y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{\epsilon}_i$

mit dem Residuum $\hat{\epsilon}_i$ der Stichprobe. Das Residuum gibt die Differenz zwischen der Regressionsgerade $\hat{\alpha} + \hat{\beta} x_i$ und den Messwerten $y_i$ an. Des Weiteren bezeichnet man mit $\hat{y}_i$ den Schätzwert (fitted value) für $y_i$ und es gilt

$\hat{y}_i = \hat{\alpha} + \hat{\beta x}_i$

und somit kann man das Residuum schreiben als $\hat{\epsilon}_i = y_i - \hat{y}_i$.

Eine Möglichkeit, die Gerade zu schätzen ist die Methode der kleinsten Quadrate (OLS, Ordinary Least Squares). Man legt eine Gerade so durch den Punkteschwarm, dass die Quadratsumme der Residuen, also der senkrechten Abweichungen $\hat{\epsilon}_i$ der Punkte von dieser Ausgleichsgeraden minimiert wird.

In der folgenden Grafik sieht man nun die Beobachtungspunkte und die geschätzte Regressionsgerade.