UNIVERSITÄT SIEGEN

Fachbereich 5
Lehrstuhl für Marketing
Univ.-Professor Dr. Hanna Schramm-Klein

Dipl.-Kfm. Sascha Steinmann
Universität Siegen
Lehrstuhl für Marketing
steinmann@marketing.uni-siegen.de

Sommersemester 2010

Marktforschung

Übungsaufgaben zu Regressionsanalyse

Aufgabe 1:

Worin besteht der Unterschied zwischen der einfachen und der multiplen Regressionsanalyse?

Aufgabe 2:

Erläutern Sie die Schritte zur Ermittlung der Regressionsgeraden anhand des konkreten Beispiels aus der Vorlesung. Verdeutlichen Sie das Grundprinzip der Regression anhand einer Grafik für den bivariaten Fall.

Aufgabe 3:

Welche Kriterien haben Sie in der Vorlesung für die Beurteilung einer Regressionsfunktion kennengelernt? Welche inhaltlichen Aussagen lassen sich aus dem Bestimmtheitsmaß R^2, den standardisierten und den nicht-standardisierten Regressionskoeffizienten ableiten?

Aufgabe 4:

Bei K=12 Familienvätern wurden die jährlichen Ausgaben (in Euro) für ein bestimmtes Konsumgut erhoben. Der Wert derselben Größe wurde auch bei ihren Söhnen erfasst. Die erfassten Daten sind in der nachstehenden Tabelle zusammengefasst.

	K	1	2	3	4	5	6	7	8	9	10	11	12
Väter	x_k	60	80	70	90	65	100	120	80	95	65	85	100
Söhne	y_k	65	90	65	80	60	110	110	60	90	70	100	85

a) Zeichnen und interpretieren Sie das zugehörige Streudiagramm.

b) Führen Sie eine Kleinst-Quadrate-Schätzung durch und schätzen Sie die Ausgaben eines Sohnes, dessen Vater x_0=110 Euro für das betrachtete Gut ausgegeben hat. Bitte achten Sie auf einen strukturierten und nachvollziehbaren Lösungsweg.

c) Berechnen Sie das zugehörige Bestimmtheitsmaß und interpretieren Sie Ihr Ergebnis. Bitte achten Sie auf einen strukturierten und nachvollziehbaren Lösungsweg.

Bearbeitungshinweise zu Aufgabe 4:

$$b_0 = \frac{\sum x_k^2 \sum y_k - \sum x_k \sum x_k y_k}{K \sum x_k^2 - (\sum x_k)^2}$$

$$b_1 = \frac{K \sum x_k y_k - \sum x_k \sum y_k}{K \sum x_k^2 - (\sum x_k)^2}$$

$$R^2 = \frac{\sum_{k=1}^{K} (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^{K} (y_k - \bar{y})^2}$$

Aufgabe 5:

Welche Voraussetzungen müssen für die (multivariate) Regressionsanalyse erfüllt sein?

Musterlösung (Skizze):

Aufgabe 1:

Der Unterschied liegt in der in der Anzahl der berücksichtigten unabhängigen Variablen.

Diese werden häufig auch als Prädiktoren bezeichnet:

- einfache Regression = eine unabhängige Variable (ein Prädiktor)
- multiple/mulitvariate Regression = mindestens zwei unabhängige Variablen (mindestens zwei Prädiktoren)

Aufgabe 2:

Die drei wesentlichen Schritte der Regressionsanalyse:

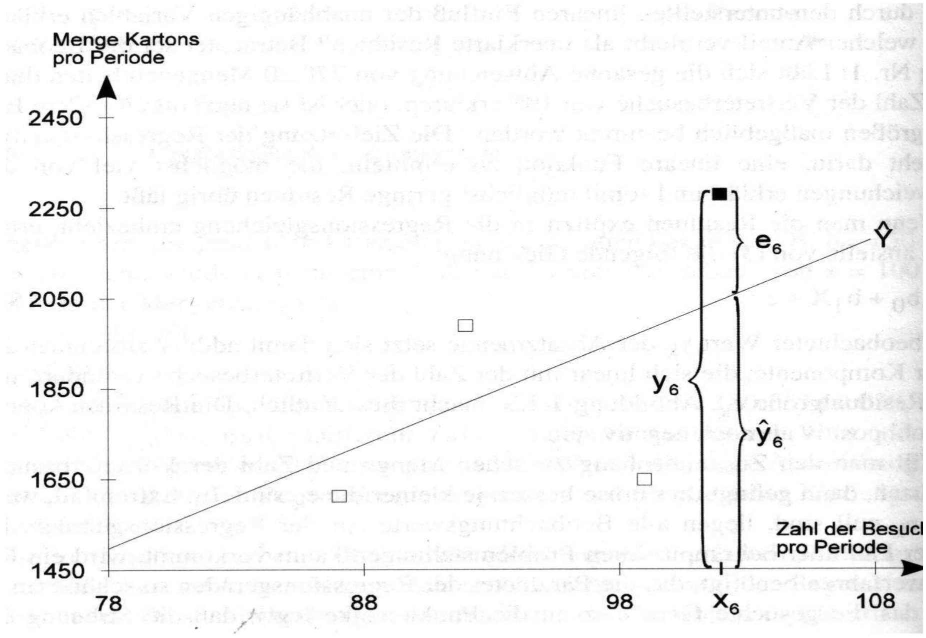
(1) Formulierung des Modells

(2) Schätzung der Regressionsfunktion

(3) Prüfung der Regressionsfunktion

Erläuterung der einzelnen Schritte siehe Vorlesungsfolien!!

Grundprinzip der Regressionsanalyse im bivariaten Fall (lineare Einfachregression):



Der Anteil aller Abweichungen der Beobachtungswerte von ihrem gemeinsamen Mittelwert (gekennzeichnet in der Grafik als Bereich y_6) lässt sich zum Teil durch einen linearen Einfluss der unabhängigen Variablen erklären (verdeutlicht durch die Linie; erklärter Anteil ist durch den Bereich $\hat{y}_6$ gekennzeichnet). Ein bestimmter Anteil verbleibt jedoch als unerklärte Residuen (e_6 in der Abbildung). Gesucht wird eine Funktion der Form $y = b_0 + b_1x_1$, für die gilt: Quadrat der verbleibenden Residuen ist minimal ($\Sigma e^2 \rightarrow$ min.). Die einzelnen Funktionsparameter (im bivariaten Fall: b_0 und b_1) werden hierfür mittels Kleinst-Quadrate-Schätzung (KQS) ermittelt.

Aufgabe 3:

die Kriterien sind ja in der Frage schon aufgeführt:

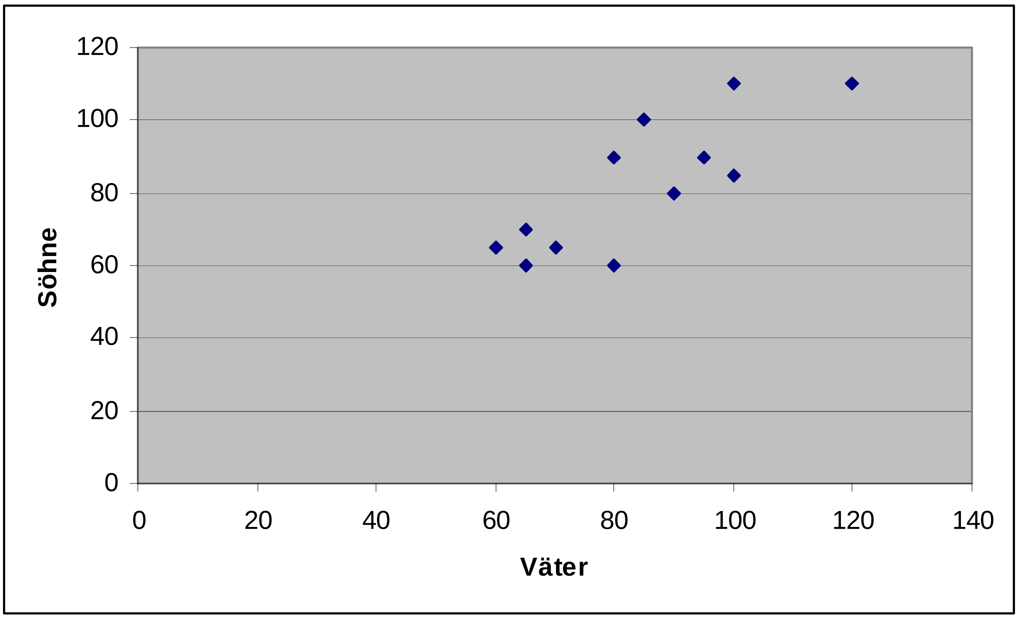
<u>Bestimmtheitsmaß R^2 =</u> sagt aus, welcher Anteil der Varianz/Schwankungen des Kriteriums (AV) durch die Prädiktoren (UV) erklärt werden kann; kann Werte zwischen 0 (keine Erklärung) und 1 (vollständige Erklärung) annehmen.

<u>nicht standardisierter Regressionskoeffizient =</u> ist Bestandteil der Regressionsfunktion der Form $y = b_0 + b_1x_1 + \ldots + b_jx_j$; besagt in unserem Beispiel aus der Vorlesung, dass eine marginale Änderung der Variable „Vertreterbesuche“ (also ein Besuch mehr) zu einer Änderung von ca. 11 in der Variable „Menge“ führt. Also: ein Vertreterbesuch mehr → ca. 11 Kartons mehr verkauft. Ähnliches gilt für die Variable „Ausgaben“. Daraus lässt sich aber nicht ableiten, dass bspw. die Vertreterbesuche einen stärkeren Einfluss auf die AV ausüben, als die Ausgaben.

<u>standardisierter Regressionskoeffizient =</u> drückt den relativen Anteil der Schwankungen der AV aus, der durch diesen Prädiktor erklärt wird. An den standardisierten Regressionskoeffizienten aus dem Vorlesungsbeispiel lässt sich schlussfolgern, dass die Ausgaben einen größeren Einfluss auf die verkaufte Menge nehmen, als die Vertreterbesuche.

Aufgabe 4:

a) Streudiagramm



Augenscheinlich liegt eine positive Beziehung zwischen den Ausgaben der Väter und den Ausgaben der Söhne vor, d.h. mit steigenden Ausgaben der Väter, gehen auch höhere Ausgaben der Söhne für das Konsumgut einher.

b) Hier werden lediglich die für die Berechnung der Funktionsparameter notwendigen Zwischenschritte sowie die berechneten Funktionsparameter und zu ermittelnde Wert der Regressionsfunktion bei Ausgaben von x_0 = 110 Euro angegeben. Die einzelnen Schritte für die Berechnung führen Sie zur Übung bitte selber durch. Für einen nachvollziehbaren Lösungsweg können Sie sich zum Beispiel an den in den Vorlesungsfolien angeführten Arbeitstabellen orientieren.

$\sum x_k^2 = 88500$

$\sum x_k = 1010$

$\sum y_k = 985$

$\sum x_k y_k = 85850$

$b_0 = 11{,}074$

$b_1 = 0{,}844$

$\hat{y}_k = 11{,}074 + 0{,}844 x_k$

$y_0 = 103{,}914$

d) $R^2 = 0{,}6681$; zur Interpretation des Bestimmtheitsmaßes vgl. die Vorlesungsfolien.

Aufgabe 5: wichtige Voraussetzungen für die Regressionsanalyse (mit wenigen Ergänzungen zu den Vorlesungsfolien**)**

- **Metrisches Messniveau** sowohl für die abhängigen als auch für die unabhängigen Variablen.
- **Lineare Beziehung** zwischen Prädiktoren und Kriterium (vgl. Backhaus et al. 2006 S. 81)
- **Additive Verknüpfung** der unabhängigen Variablen
- **Stichprobe** ≥ **Anzahl** der Parameter (b_0, $b_1 \ldots b_j$) Faustregel: min. 3 - 5x so groß wie Anzahl der Parameter
- Die **Residuen sind normalverteilt** ($e_k \sim N(0,\sigma 2)$)
- Unkorrelierte Residuen (**keine Autokorrelation**)
- keine (perfekte) **Multikollinearität**