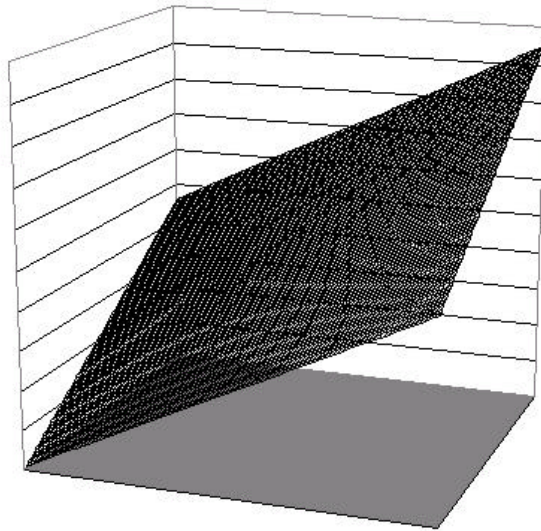


Referat: Multiple Regression in der Anwendung mit SPSS  
Seminar 1847: *Evaluation pädagogisch-psychologischer Maßnahmen*  
Leitung: Frau P.D. Elisabeth Kals  
Referenten: Andreas Nack (Matr.Nr. 47 04 78)  
Jan Seifert (Matr.Nr. 43 79 43)



# Multiple Regression in der Anwendung mit SPSS



## Inhalt

<b>1</b>	<b>Vorbemerkungen</b>	<b>2</b>
<b>2</b>	<b>Hintergründe zur Erinnerung</b>	<b>2</b>
2.1	Bedeutung der einzelnen Parameter	2
2.2	Signifikanztests	2
<b>3</b>	<b>Praktische Umsetzung: Voraussetzungen &amp; Fehlerquellen</b>	<b>3</b>
3.1	<b>Vor der Datenerhebung</b>	<b>3</b>
3.1.1	Störvariablen und Scheinkorrelationen	3
3.1.2	Rückwärts gerichtete Kausalität	3
3.1.3	Meßgenauigkeit	4
3.1.4	Intervallskalierung	5
3.1.5	Stichprobengröße	5
3.1.6	Auswahl der Variablen	5
3.2	<b>Vor der Regression</b>	<b>6</b>
3.2.1	Falsche Polung von Skalenwerten	6
3.2.2	Missing Values beim Aggregieren von Skalenwerten	6
3.3	<b>Nach der Regression</b>	<b>7</b>
3.3.1	Ausreißer	7
3.3.2	Linearität und Normalverteilung der Zusammenhänge	8
3.3.3	Statistische Verteilung der Residuen	9
3.3.4	Varianzhomogenität (Homoskedastizität) und Unabhängigkeit der Residuen	10
3.3.5	Signifikanz	10
3.3.6	Effekt	11
3.3.7	Power	12
3.4	<b>Komplexere Probleme</b>	<b>12</b>
3.4.1	Suppression	12
3.4.2	Multi-Kollinearität	12
3.4.3	Moderator-Effekte	13
3.4.4	Seriell abhängige Daten	14
3.5	<b>Modellselektion: Bestimmung der besten Regressionsgleichung</b>	<b>15</b>
3.5.1	Theoriegeleitet	15
3.5.2	Blockweise	15
3.5.3	Vorwärts oder Rückwärts	16
3.5.4	Die Kombination von Vorwärts und Rückwärts: Schrittweise	16
3.5.5	Diskussion	16
<b>4</b>	<b>Literatur</b>	<b>17</b>

## 1 Vorbemerkungen

Wir stützen uns in dieser Hausarbeit auf SPSS™ Version 8.0. Alle Menüpunkte, Buttonbezeichner, Schalter und Dialogboxen sind zur besseren Lesbarkeit **fett** markiert.

Wir verwenden in dieser Hausarbeit im wesentlichen Schaubilder und Ergebnisse, die sich bei multiplen Regressionen im Umgang mit der Datei Autos.sav (Cars.sav) ergeben haben. Diese Datei ist eines der mitgelieferten SPSS-Beispiele. Für unsere Regression wollten wir bestimmen, durch welche Variablen eines Autos der Benzinverbrauch vorhergesagt werden kann. (Wenn man unserem Beispiel Glauben schenken mag, dann wären das lediglich das Gewicht und das Baujahr).

Außerdem wollen wir anmerken, daß Grundkenntnisse in der Methodenlehre ganz allgemein, und speziell auch die Grundprinzipien der multiplen Regression zum Verständnis notwendig sind.

## 2 Hintergründe zur Erinnerung

### 2.1 Bedeutung der einzelnen Parameter

I. Das *Regressionsmodell* oder auch *Regressionsgleichung*:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon$$

Wenn alle  $x_i$  und  $y_i$  z-standardisiert werden, sieht die Formel entsprechend anders aus:

$$Z(y) = \beta_0 + \beta_1 Z(x_1) + \beta_2 Z(x_2) + \dots + \beta_n Z(x_n) + Z(\varepsilon)$$

- II. Die Konstante  $b_0 / \beta_0$  ist die *Regressionskonstante* und bezeichnet (in Anlehnung an die bivariate Regression) den Y-Achsenabschnitt. Die übrigen Werte  $b_i / \beta_i$  sind die *Regressionskoeffizienten*. Sie geben an, mit welcher Gewichtung ein Prädiktor auf die abhängige Variable  $y$  wirkt. Die standardisierten Regressionskoeffizienten  $\beta_i$  entsprechen darüber hinaus den Semipartial-Korrelationskoeffizienten.
- III.  $R$  ist die multiple Korrelation.
- IV.  $R^2$  ist die (prozentuale) aufgeklärte Varianz, die in  $Y$  vorhergesagt werden kann, wenn man  $Y$  durch eine lineare Regression mittels  $x_1, \dots, x_m$  beschreibt.

### 2.2 Signifikanztests

Der Signifikanztest über das gesamte Regressionsmodell erfolgt über die F-Verteilung. Zusätzlich müssen die einzelnen Regressionskoeffizienten auf Signifikanz geprüft werden, was über t-Tests erfolgt.

Generell gibt es zu den Signifikanztest bei SPSS (und vergleichbaren Statistikprogrammen) zu sagen, daß sie eher optimistische Ergebnisse liefern, d.h. sie benennen einen Effekt schnell als signifikant und riskieren dabei einen Fehler (EFRON & TIBSHIRANI 1986).

### **3 Praktische Umsetzung: Voraussetzungen & Fehlerquellen**

#### **3.1 Vor der Datenerhebung**

##### **3.1.1 Störvariablen und Scheinkorrelationen**

Prinzipiell könnte jede Variable (außer dem Kriterium und den Prädiktoren) eine Störvariable sein. Randomisierung entfällt bei der multiplen Regression leider meist als Möglichkeit, Einflüssen der Störvariablen entgegenzuwirken. So müssen also die Störvariablen auch erhoben werden, um sie kontrollieren zu können. Natürlich kann man nicht unendlich viele Variablen erheben. Daher ist es notwendig, sich vorher zu überlegen, von welchen Variablen es Sinn macht, sie als potentielle Störvariable zu erfassen.

Danach gilt es zu klären, ob eine Variable tatsächlich eine Störvariable ist. Variablen, die nur mit Prädiktoren korrelieren, sind unproblematisch und müssen im Regressionsmodell nicht berücksichtigt werden. Bei Variablen, die mit dem Kriterium zusammenhängen jedoch nicht mit den Prädiktoren, sollte man sich die Frage stellen, ob man diese nicht als Prädiktor mit ins Modell aufnehmen sollte. Hierbei spielen sicher inhaltliche und theoretische Überlegungen eine Rolle, aber auch die stetige Absicht eines Versuchsleiters, Fehlervarianz zu minimieren.

Korreliert eine Variable aber sowohl mit dem Kriterium als auch mit mindestens einem Prädiktor, könnten die Schätzungen der Regressionskoeffizienten verzerrt werden, wenn man diese Variable nicht in das Regressionsmodell aufnimmt. Denn es könnte sein, daß der Zusammenhang zwischen Kriterium und Prädiktor nur eine Scheinkorrelation ist, die durch die Störvariable vermittelt wird.

##### **3.1.2 Rückwärts gerichtete Kausalität**

Auch kausale Effekte des Kriteriums auf die Prädiktoren könnten durch Randomisierung natürlich vermieden werden, aber gerade die multiple Regression wird meist unter Bedingungen eingesetzt, in denen Randomisierung eben nicht möglich ist. Manchmal läßt sich Kausalität noch aus den logischen Zusammenhängen ableiten. Daß beispielsweise

Körpergröße auf die Selbstsicherheit eines Menschen wirken könnte, erscheint jedem plausibel; der umgekehrte Fall wäre dagegen einfach nur Unsinn.

Leider ist dem Problem rückwärts gerichteter Kausalität mit der multiplen Regression in keinsten Weise beizukommen. Eine Alternative wären hier eventuell die Pfadanalyse oder die Strukturgleichungsmodelle. Die einzige Möglichkeit, die man aber bei der multiplen Regression hat, ist es, sich Zusammenhänge in beiderlei Richtung zu veranschaulichen und zu verstehen, um sich danach zu fragen, wie sinnvoll oder wahrscheinlich der rückwärts gerichtete Fall wäre. Zum Beispiel könnte einer wissen wollen, ob ein städtisches Wohnumfeld Toleranz fördert; tatsächlich könnte es aber auch sein, daß tolerante Menschen sich bevorzugt in Städten niederlassen.

Die kausale Wirkung des Kriteriums auf die Prädiktoren kann für das Ergebnis der multiplen Regression von nicht unerheblicher Bedeutung sein. Man muß verzerrte Regressionskoeffizienten in Kauf nehmen, wenn das Kriterium auf *mehrere* Prädiktoren wirkt.

Abbildung 1 zeigt ein idealisiertes Beispiel. Dort hat das Kriterium K kausale Wirkung auf die Prädiktoren P1 ( $r = 0,683$ ) und P2 ( $r = 0,845$ ). Aus diesem Grund besteht zwischen den Prädiktoren eine Scheinkorrelation ( $r = 0,553$ ). Diese *Scheinkorrelation* wird von der Regressionsanalyse entdeckt, mißverstanden und mittels Partialisierung eliminiert. Das Regressionsmodell schließlich sieht wie folgt aus:  $K = 0,673 \cdot P2 + 0,311 \cdot P1 (+ P0)$ . Verglichen mit den realen Gegebenheiten aus Abbildung 1 muß man zugeben, daß dieses Modell die Wirklichkeit nicht sonderlich gut widerspiegelt.

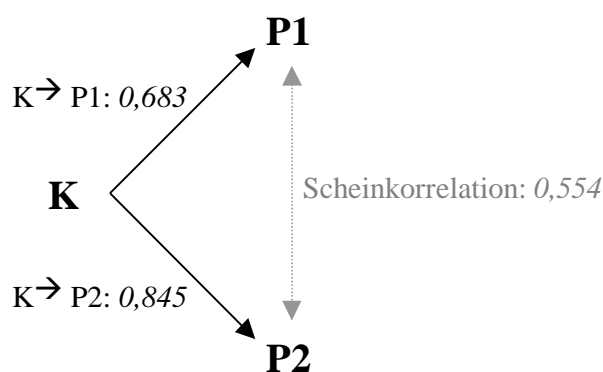


Abbildung 1

### 3.1.3 Meßgenauigkeit

Je ungenauer die Variablen gemessen werden – je kleiner also deren Reliabilität – desto schlechter können die Koeffizienten  $R$ ,  $R^2$  und die  $\beta$ -Gewichte geschätzt werden. Die

zusätzlich entstehende Fehlervarianz belastet darüber hinaus die Hypothesen. Darum muß der Psychometrie (z.B. der Skalenkonstruktion) viel Aufmerksamkeit und Mühe gewidmet werden.

### 3.1.4 Intervallskalierung

Verletzung des Intervallskalenniveaus: Die Voraussetzung des Intervallskalenniveaus ist für die multiple Regression so gut wie immer notwendig. Darauf muß insbesondere bei der Skalenkonstruktion geachtet werden, da hier die Gefahr besonders groß ist, ins Ordinalskalenniveau abzugleiten, wenn die Abstufungen der Ratings keine gleich großen Abstände abbilden.

Ausnahme bilden zweistufige Nominalvariablen (z.B. Geschlecht), die als Prädiktoren auch in die Regressionsgleichung eingehen dürfen. In diesem Fall beschreibt das  $\beta$ -Gewicht einer solchen Variablen den Unterschied zwischen den beiden Faktorstufen im Kriterium. Die Stufen der Nominalvariablen werden in neuen Variablen (einer sog. Dummy- oder Indikatorvariablen) durch Zahlen ersetzt. Näheres hierzu in BORTZ (1993, Kap. 14).

### 3.1.5 Stichprobengröße

Die Stichprobengröße aufgrund von Power- und Signifikanzgründen läßt sich am leichtesten mit G-Power (ERDFELDER, FAUL & BUCHNER 1996) ermitteln. Allerdings ist diese Vorgehensweise alleine nicht ausreichend, da zur Schätzung stabiler Parameter mindestens (und *aller*mindestens) 10 Probanden pro Prädiktor (GEDIGA & KUHNT 1999) von Nöten sind.

### 3.1.6 Auswahl der Variablen

Insbesondere wenn man eine schrittweise Regression (siehe 3.5) rechnet, spielt die Anzahl der Variablen eine große Rolle. Da bei der schrittweisen Regression jeweils der „signifikanteste“ Zusammenhang ausgewählt wird, um die Regressionsgleichung aufzubauen, spielt man dem Zufall in die Hände. Es muß - allein aufgrund von Zufallsschwankungen - einige Prädiktoren geben, die signifikant mit der abhängigen Variablen zusammenhängen. Je mehr Variablen man benutzt, desto größer wird das Problem.

Ein weiteres Problem bildet die Stabilität der Schätzungen. In der Literatur wird vorgeschlagen, daß man mindestens 10mal (besser 20mal) soviel Fälle haben soll wie Prädiktoren (GEDIGA & KUHNT 1999). Es macht also auch deswegen keinen Sinn, alle möglichen Variablen zu erheben. Es sollte eine subtile Auswahl getroffen werden. Nicht zu viele Prädiktoren, um die genannten Probleme zu umgehen; aber auch nicht zu wenige, um die gewünschten Zusammenhänge zu entschlüsseln.

Beide Probleme lassen sich eigentlich relativ leicht angehen: Man führt die Regression (ob schrittweise oder nicht) an einem Teil (z.B. der ersten Hälfte) des Datensatzes durch und benutzt die Ergebnisse der Regression aus dem „Lerndatensatz“ in einem „Testdatensatz“ (also z.B. der zweiten Hälfte). Ein Plot von vorhergesagten Y-Werten (auf der Basis der Regressionskoeffizienten aus dem Lerndatensatz) gegen die beobachteten Y-Werte im Testdatensatz zeigt, wie gut die Regression tatsächlich ist.

Diese sogenannte *Kreuzvalidierungstechnik* ist einfach durchzuführen, wird aber leider bisher von SPSS nicht automatisiert angeboten. Ein wichtiger Index ist Varianz des Kreuzvalidierung-Residuums, d.h. der Differenz von vorhergesagten und beobachteten Werten. Das Verhältnis der Varianz des Kreuzvalidierung-Residuums und der Varianz der abhängigen Variablen ergibt ein Fehlermaß, das optimalerweise nahe Null sein sollte (aber durchaus größer sein kann als 1!).

## 3.2 Vor der Regression

### 3.2.1 Falsche Polung von Skalenwerten

In jeder vernünftigen Skala sind sowohl positiv formulierte Items enthalten, als auch negative. Um die Werte vernünftig aggregieren zu können, müssen alle Items auf dieselbe Polung gebracht werden. Dabei hilft der Compute-Befehl oder **Transformieren→Berechnen (Transform→Compute)**. Unter **Zielvariable (Target Variable)** wird der Name des Items angegeben. Unter **Numerischer Ausdruck (Numeric Expression)** wird der Befehl wie folgt angegeben: **[Stufenzahl der Variablen]-[Item]+1**. Falls also ein Item mit der Bezeichnung **berge2** umgepolt werden soll, das auf 6 Stufen gemessen wurde, lautet das so: **6-berge2+1**.

### 3.2.2 Missing Values beim Aggregieren von Skalenwerten

Wenn mehrere Items einer Fragebogenskala vor der Regression zu dem anschließenden Prädiktor zusammengefaßt werden sollen, sollte darauf geachtet werden, daß genügend Items zum Addieren zur Verfügung stehen. Wenn mehrere Items von einem Probanden nicht beantwortet wurden, dann macht es keinen Sinn diesen Skalenwert zu aggregieren und ihn in die Regression aufzunehmen.

In SPSS wird die neue Skalenvariable über **Transformieren→Berechnen (Transform→Compute)** errechnet. Unter **Zielvariable (Target Variable)** wird der Name der neuen Variable eingegeben. Unter **Numerischer Ausdruck (Numeric Expression)** wird der Befehl wie folgt angegeben: **MEAN.3(VAR1,VAR2,VAR3,VAR4,VAR5)**, wobei **.3**

nach **MEAN** die Anzahl der minimal erforderlichen Werte angibt, also werden in diesem Beispiel maximal zwei Missing Werte geduldet.

### 3.3 Nach der Regression

#### 3.3.1 Ausreißer

Die Kleinste-Quadrate-Schätzung der Regressionsgewichte ist besonders empfindlich gegenüber Ausreißern in den Prädiktoren (siehe Abbildung 2).

Bei Erhebungen per Fragebogen ist das meist kein Problem, da der Antwortbereich fix vorgegeben ist (z.B. *1 = stimme voll zu* bis *6 = lehne voll ab*). Bei freiem Antwortformat besteht hier allerdings ein enormes Risiko.

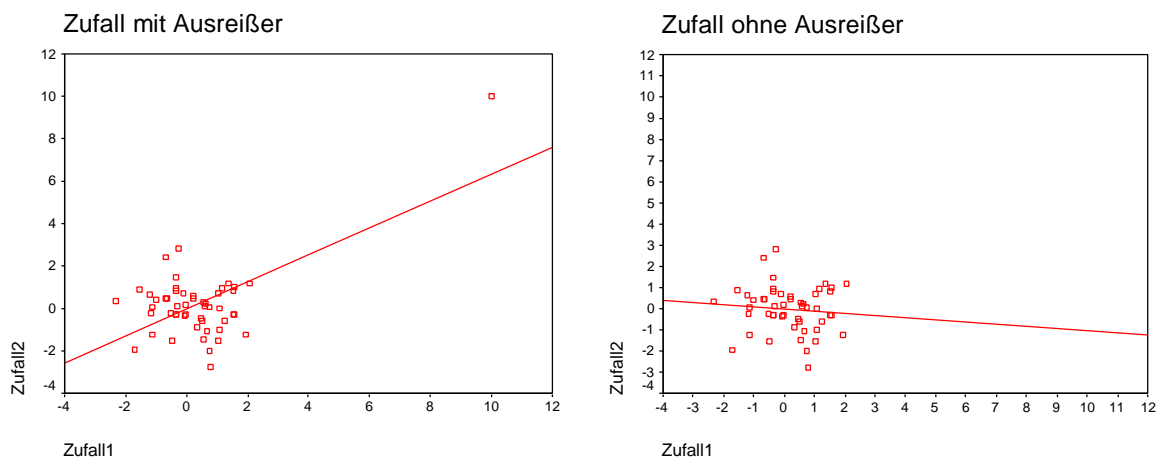


Abbildung 2 – Kein Zusammenhang zwischen Zufall1 und Zufall2 - mit und ohne Ausreißer

Mit einer grafischen Ausgabe von Residuum und vorhergesagten Werten läßt sich die Regression auf Ausreißer testen. Dazu öffnet man den Sub-Dialog **Diagramme (Plots)**. Auf der X-Achse wird der standardisierte vorhergesagte Wert benutzt (SPSS-Variablen **\*ZPRED**). Durch die Standardisierung kann man erkennen, welche vorhergesagten Werte extrem vom Mittelwert abweichen, was auf Ausreißern in den X-Werten hinweist. Auf der Y-Achse das studentisierte (=standardisierte) ausgeschlossene Residuum für jeden Fall berechnet (SPSS-Variablen **\*SDRESID**). Hierbei bedeutet „ausgeschlossen“, daß für jeden Fall eine Regression berechnet wird, wobei dieser Fall nicht in Berechnung eingeht. Der Abstand der Vorhersage aus den anderen Werten auf den beobachteten Wert dieses Falles wird dann standardisiert. Im Diagramm bilden dann wieder Werte, die betragsmäßig größer als 3,5 sind (auf X und Y), Hinweise auf mögliche Ausreißer (GEDIGA & KUHN 1999).

Weitere Checks auf ausreißende Beobachtungen bilden die Cook-Distanzen der einzelnen Beobachtungen. Cooks Distanzen sind ein Maß dafür, wie stark die Werte eines



Probanden das Ergebnis der Regression beeinflussen. Im Sub-Dialog **Speichern (Save)** die Option **nach Cook (Cook's)** berechnet eine neue Spalte mit Werten im Datenblatt. Dort müssen nur noch Werte größer *Eins* gesucht werden (STEVENS 1996). Diese sind meistens bedenklich. In aller Regel erbringt der Plot der studentisierten ausgeschlossenen Residuen gegen die standardisierten vorhergesagten Werte die gleiche Information wie die Inspektion von Cook-Distanzen.

Neben diesen beiden Verfahren bietet SPSS noch eine ganze Reihe von Alternativen. STEVENS (1996) geht ausführlicher auf die verschiedenen Möglichkeiten ein.

Die ‚Therapie‘ bei Ausreißern besteht meist darin, sie zu eliminieren. Allerdings muß man damit vorsichtig sein. Es ist ganz sicher *nicht* wissenschaftlich korrekt, einfach alle diejenigen Probanden/Fälle aus den Daten zu entfernen, die mir eine Hypothese kaputt machen. Zudem können solch unausgewogenen Verteilungen der Datenpunkte auch auf ganz andere Probleme hindeuten (z.B. einen Stichprobenfehler). Wenn allerdings Datensätze gelöscht werden, dann kann das nicht geschehen, ohne daß eine technische Panne nachgewiesen wäre.

### 3.3.2 Linearität und Normalverteilung der Zusammenhänge

Das lineare Modell der Regression ist nicht in der Lage nicht-lineare Zusammenhänge zu erkennen, also beispielsweise quadratische oder exponentielle (Abbildung 3 zeigt ein bivariates Beispiel).

Die Linearitätsannahme muß für jeden Zusammenhang zwischen dem Kriterium und den einzelnen Prädiktor überprüft werden. Auch wenn man viele Prädiktoren hat, ist das mit SPSS sehr einfach. Im Sub-Dialog **Diagramme (Plots)** muß lediglich die Option **Alle partiellen Diagramme erzeugen (Produce all partial plots)** angewählt werden. Als Ergebnis liefert SPSS ebenso viele Scatterdiagramme (**Partielles Regressionsdiagramm von...**), wie Prädiktoren vorhanden sind. Die Punktwolken in diesen Diagrammen müssen einigermaßen einen linearen Zusammenhang erkennen lassen. Abbildung 3 dagegen zeigt eine deutliche Verletzung der Linearitätsannahme. Derart deutlich sind diese Verletzungen in der Realität allerdings recht selten.

Eine weitere Voraussetzung läßt sich ebenfalls mit Hilfe der **Partiellen Regressionsdiagramme** überprüfen. Die jeweiligen Korrelationen zwischen Kriterium und Prädiktor müssen bivariat normalverteilt sein.

Im Scatterdiagramm prüft man das an der Form der Punktwolke. Läßt sich die Wolke ungefähr als Ellipse ausmachen, dann kann man diese Annahme als erfüllt betrachten.

Mit dieser Voraussetzung kann man allerdings einigermaßen nachlässig umgehen, weil Verteilung der Variablen nicht die Schätzung der  $\beta$ -Koeffizienten oder der Teststatistik beeinträchtigt. Sie ist allerdings notwendig, um gefundene Zusammenhänge linear interpretieren zu können. Insgesamt aber ist die multiple Regression recht stabil, was die Verletzung dieser Annahme angeht.

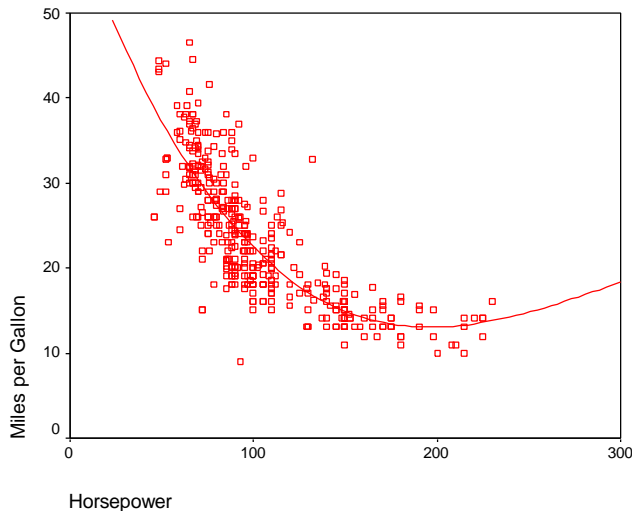


Abbildung 3 – non-linearer Zusammenhang zwischen Motorleistung und Benzinverbrauch; mit polynomialer Kurvenanpassung

### 3.3.3 Statistische Verteilung der Residuen

Die mathematische Prozedur der multiplen Regression setzt für den Mittelwert aller Residuen voraus, er sei gleich Null.

Wenn im Sub-Dialog **Diagramme (Plots)** mindestens ein **Streudiagramm (Scatter)** mit X und Y definiert wird, dann gibt SPSS eine Tabelle **Residuenstatistik (Residuals Statistic)** aus. In dieser Tabelle läßt sich der Mittelwert der Residuen ablesen (**nicht standardisierte Residuen**), er sollte möglichst nahe bei Null liegen.

Ob die Residuen auch normalverteilt sind, läßt sich mittels eines weiteren Schaubildes überprüfen. Dieses läßt sich parallel zur Regression miterstellen über den Sub-Dialog **Diagramme**→**Diagramme der standardisierten Residuen (Plots**→**Standardized residual plots)**. Die beiden Optionen **Histogramm (Histogramm)** und **Normalverteilungsdiagramm (Normal probability plot)** resultieren in jeweils einem Schaubild (siehe Abbildung 4).

Im **P-P-Diagramm (P-P-Plot)** müssen die Punkte möglichst nahe an der 1. Winkelhalbierenden liegen, damit man die Normalverteilungsannahme bezüglich der Residuen als gegeben betrachten kann.

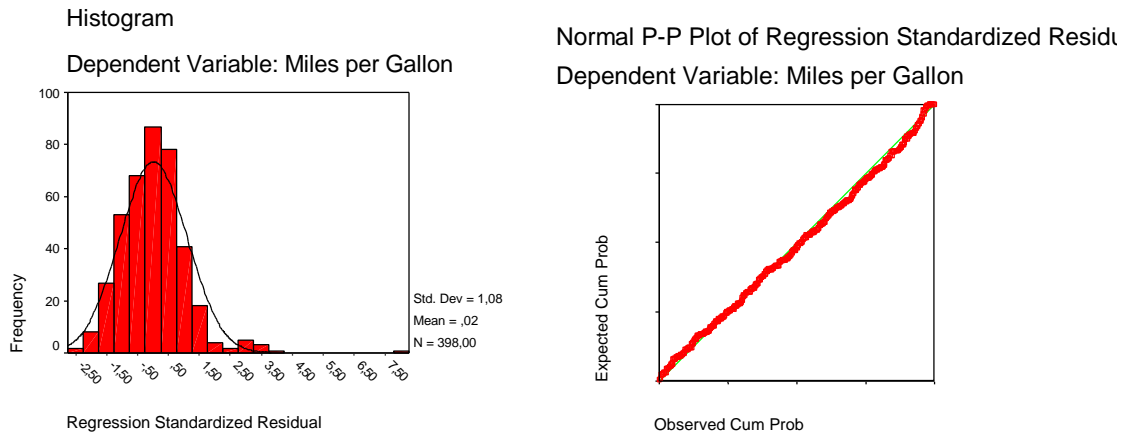


Abbildung 4

### 3.3.4 Varianzhomogenität (Homoskedastizität) und Unabhängigkeit der Residuen

Zum sinnvollen Einsatz der multiplen Regression besteht zum einen die Forderung, daß die Residuen nicht in Abhängigkeit zu den vorhergesagten Werten streuen. Anders formuliert: die Varianz der Residuen muß unabhängig sein von dem Kriterium.

Zur Überprüfung nutzen wir dasselbe Schaubild wie zur Überprüfung von Ausreißern (siehe 3.3):  $\hat{Y}$  - standardisierte vorhergesagte Werte,  $X$  - studentisierte ausgeschlossene Residuen. Abbildung 5 zeigt idealisierte Darstellungen, die auf Verletzungen der beiden Voraussetzungen hindeuten.

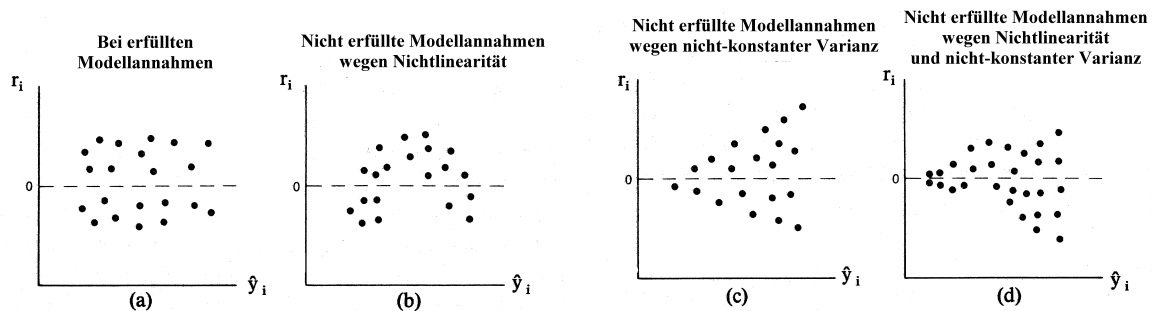


Abbildung 5 - (aus STEVENS 1996)

Im Falle heterogener Varianzen gibt es eine Möglichkeit, die Daten rechnerisch aufzubereiten und das Problem damit zu umgehen. Jede Beobachtung wird gewichtet „mit dem Kehrwert der Wurzel der Fehlervarianz“. Wer allerdings genau wissen will, wie diese sogenannte *Methode der gewichteten Kleinst-Quadrat-Schätzung* funktioniert, sei auf einschlägige Darstellungen verwiesen (U.R.T. 1999 S. 23ff. oder RYAN 1997, S.60ff.).

### 3.3.5 Signifikanz

Der F-Wert mit Signifikanzniveau läßt sich in der Tabelle ANOVA ablesen. Damit sind aber noch nicht die einzelnen Koeffizienten auf ihre Signifikanz geprüft. Zwar liefert SPSS in

der Tabelle **Koeffizienten (Coefficients)** eine Teststatistik, aber bei mehreren Einzeltests macht sich die Kumulierung des  $\alpha$ -Fehlers bemerkbar. Das ist hier nicht berücksichtigt und muß etwas umständlich über eine MANOVA ermöglicht werden. Dieser Spezialfall der MANOVA kann in SPSS 8.0 nicht mehr über Dialog erreicht werden. Daher muß man wie folgt vorgehen: **Datei**→**Neu**→**Syntax (File**→**New**→**Syntax)**. Daraufhin öffnet SPSS ein neues Fenster „**SPSS Syntax-Editor**“. Dort müssen folgende Zeilen eingegeben werden, k (in der zweiten Zeile) wird durch das jeweilige Kriterium ersetzt und var1 bis var6 wird ersetzt durch die Prädiktoren:

```
MANOVA
  k WITH var1 var2 var3 var4 var5 var6
/PRINT PARAM(ESTIM)
/CINTERVAL JOINT(.95) UNIVARIATE(BONFER)
/METHOD=UNIQUE
/ERROR WITHIN+RESIDUAL
/DESIGN .
```

Jetzt nur noch **Ausführen**→**Alles (Run**→**All)** und die MANOVA wird ausgeführt.

### 3.3.6 Effekt

Vor allem, wenn die Probandenzahl über das hinausgeht, was zur Identifikation eines bestimmten Effektes nötig gewesen wäre (zur Schätzung stabiler Parameter siehe 3.1.5), ist die Post-Hoc Analyse des Effektes wichtig.

Die Effektgröße bei der bivariaten und der multiple Regression ist an die aufgeklärte Varianz  $R^2$  (=  $r^2$  in der bivariaten linearen Regression) gekoppelt. Nach COHEN (1988) unterscheidet man 3 Klassen von Effektstärken:

Tabelle 1

Größe des Effekts	klein	mittel	groß
$R^2$	0.0196	0.13	0.26
$f^2$	0.02	0.15	0.35

Natürlich sind diese Angaben von COHEN nicht absolut zu verstehen, sondern machen nur Sinn im Rahmen einer Kosten-Nutzen-Analyse (sensu E. Naumann).

Obwohl die Größe „aufgeklärte Varianz“ in Varianzanalyse und multipler Regression identisch ist, sind die Werte für die Effekte unterschiedlich. Dies liegt an der Methode, wie COHEN die Effektgrößen bestimmte: COHEN identifizierte diese Größen aus publizierten Ergebnissen und die sind natürlich stark unterschiedlich, da bei Regressionsanalysen und Korrelationsberechnungen meist größere Stichproben zum Einsatz kommen als in Experimenten, für die oft die Varianzanalyse zu Testzwecken benutzt wird.

### **3.3.7 Power**

Es ist nach Abschnitt 3.3.6 bei nicht-signifikanten Effekten ebenfalls unerlässlich, post-hoc die Power zu überprüfen.

## **3.4 Komplexere Probleme**

### **3.4.1 Suppression**

Mit der multiplen Regression werden zwei Probleme gleichzeitig angegangen: Zum einen wird die abhängige Variable linear mit den unabhängigen Variablen verknüpft, zum anderen werden aber auch alle unabhängigen Variablen linear miteinander verrechnet. Bei dieser Verrechnung werden die für die Erklärung der unabhängigen Variablen nicht-interessante Zusammenhänge zwischen den Prädiktoren eliminiert. Dieser Effekt nennt sich Suppressions-Effekt und ist zunächst durchaus gewollt. Allerdings: Wenn dieser Effekt groß wird, kann die Interpretation der Ergebnisse schwierig werden!

Ein krasser Anstieg von  $\beta$ -Gewichten in der schrittweisen Regression deuten auf Suppression hin. Die  $\beta$ -Gewichte der geschätzten Regressionsgeraden sind sehr vorsichtig zu interpretieren. Die Ursache der Suppression ist zunächst am besten theoretisch zu ergründen.

### **3.4.2 Multi-Kollinearität**

Die extreme Form einer Suppression, da in dieser Situation die Prädiktoren (fast perfekt) linear zusammenhängen und somit (fast) keine Varianz für die Regression auf die abhängige Variable übrig bleibt. Ergebnisse in der Nähe der Multi-Kollinearität sind wertlos, da die Prädiktoren (fast) alle Varianz untereinander unterdrücken und die Vorhersage für Y nur auf einem geringen Varianzanteil innerhalb der Prädiktoren beruhen.

Der Folgen dieses Phänomens gibt es drei und diese sind nicht allzu erwünscht. Also drei gute Gründe, um Multi-Kollinearität zu vermeiden:

1. Multi-Kollinearität verstärkt die Varianzen der Regressionskoeffizienten ( $\beta$ -Gewichte). Das bedeutet, die Schätzung der wahren Koeffizienten wird ungenau. Das kann soweit gehen, daß ein Gewicht als negativ ausgewiesen wird, obwohl es realiter ein positives Vorzeichen trägt.
2. Multi-Kollinearität erschwert es ungemein, die tatsächliche Bedeutung eines Prädiktors abzuschätzen, denn dieser hängt im Modell sehr stark von den übrigen Prädiktoren ab.
3. Multi-Kollinearität führt zu erheblichen Problemen bei der Suche nach der besten oder „wahren“ Regressionsgleichung. Tatsächlich gilt, je höher Multi-Kollinearität, desto mehr

Modelle lassen sich aufstellen, zwischen denen statistisch keine Unterscheidung mehr möglich ist.

Das Erkennen der Multi-Kollinearität ist einfach: Statistikpakete melden diesen Zustand auf Anforderung (Sub-Dialog **Statistik (Statistics)**, Option **Kollinearitätsdiagnose (Collinearity diagnostics)**).

Bei schrittweiser Regression ist ein sehr großer Sprung in den  $\beta$ -Gewichten ein Indiz für Multi-Kollinearität. Solche Sprünge lassen sich in der Tabelle **Koeffizienten (Coefficients)** anhand der Variance Inflation Factors (VIF) ermitteln. Nach MYERS (1990) kann folgende Daumenregel zur Hilfe genommen werden. VIFs größer als 10 sollten den Experimentator zum Nachdenken verleiten. Er sollte sich dann über Lösungsstrategien Gedanken machen.

Der Umgang mit Multi-Kollinearität kann zu einem kniffligen Problem werden. Anscheinend gibt es zu viele Variablen, die alle ‚das Gleiche‘ vorhersagen. In übersichtlichen Situationen können daher einzelne Variablen aus der Regression ausgeschlossen werden (allerdings nicht ohne Begründung). Oder einige Prädiktoren werden zu einer neuen Variable aggregiert. Aber auch hier gilt, daß diese neue Variable auch noch irgendein Sinn machen muß.

Eine weitere Möglichkeit ist es, die Variablenzahl durch eine Faktorenanalyse zu reduzieren. Wenn 30 verschiedene Prädiktoren erhoben werden, die zudem untereinander erheblich korrelieren, dann ist es fraglich, ob auch tatsächlich 30 gänzlich unterschiedliche Konstrukte erfaßt werden. Die Faktorenanalyse ermittelt die übergeordneten (orthogonalen) Faktoren, die anschließend als Prädiktoren in die multiple Regression eingehen.

### **3.4.3 Moderator-Effekte**

Ein sehr unangenehmes Problem in der multiplen Regression sind Moderator-Effekte, die allgemein als Interaktionseffekte zwischen den Prädiktoren zu beschreiben sind und somit Nicht-Linearitäten höherer Ordnung bilden. Moderatorvariablen verändern somit den linearen Zusammenhang zwischen abhängiger Variable und den Prädiktoren.

Manchmal werden mit Moderatorvariablen fälschlicherweise auch solche Variablen bezeichnet, bei deren Aufnahme in die Regressionsgleichung der lineare Zusammenhang zwischen den Prädiktoren und Kriterium erheblich höher wird: Dieser Effekt ist aber durch einen Suppression zu erklären! Deshalb wäre es angemessener derartige Variablen als Suppressorvariablen zu bezeichnen.

Moderator-Effekte sind (im Gegensatz zu Suppressions-Effekten) in der Regressionsgleichung schwer zu identifizieren. Desweiteren gibt es auch keine einfache Kur

für Moderator-Effekte im linearen Modell. Ein Moderator-Effekt ist eine Nicht-Linearität. Immerhin lassen sich vier Möglichkeiten nennen, wie man mit diesem Problem umgehen kann:

1. Regressionsanalysen auf Substichproben durchführen, die durch Fallunterscheidungen auf einer Moderatorvariablen gebildet werden. Sind signifikante  $\beta$ -Gewichte in den Substichproben substantiell unterschiedlich, dann liegt höchstwahrscheinlich ein Moderator-Effekt vor. Die Ergebnisse kann man dann nicht in einer Regressionsgleichung zusammenfassen, sondern es müssen dann eben mehrere Regressionsgleichungen für die unterschiedlichen Fälle benutzt werden.
2. Weglassen einer Variablen, die die Moderation (mit-)verursacht und ersetzen durch eine andere, theoretisch ähnliche Variable, die besser zu linearisieren ist.
3. Effekt der Moderation abschätzen. Wenn der Effekte nicht sehr hoch ist, kann man mit der linearen Regression (mitsamt der Moderatorvariablen) als Näherung erster Ordnung leben.
4. Modellieren der Nicht-Linearität durch Aufnahme von Polynomen in die Regressionsgleichung. Die normale multiple Regression kann benutzt werden, wenn die Zuverlässigkeit der Prädiktoren hoch ist. Dann darf eine weitere „Variable“  $x_{12}=x_1*x_2$  in die Gleichung mit aufgenommen werden.

#### 3.4.4 Seriell abhängige Daten

Die Inferenzstatistik in der üblichen Regression setzt voraus, daß die Modellresiduen unabhängig sind. Anderenfalls sind (unter anderem) die Signifikanzbeurteilungen zu den einzelnen Regressionskoeffizienten verfälscht. Stammt jede Beobachtung von  $y$  (dem Kriterium) von einem anderen Probanden, dann kann die Unabhängigkeit der Residuen als gesichert gelten. Gehen aber Daten von mehreren Meßzeitpunkten desselben Probanden in die Regression ein, ist diese Annahme gefährdet. Dann können die Kriteriumswerte derselben  $V_{pn}$  gemeinsamen (modellfremden) Einflüssen unterliegen, so daß ihre Residualvariablen verbotenerweise korrelieren. Eben dieses Risiko besteht beispielsweise, wenn aus den Evaluationsdaten eines Trainings ein Regressionsmodell ermittelt werden soll.

Zur Prüfung dieser Bedingung kann der Durbin-Watson-Koeffizient (**Durbin-Watson** im Sub-Dialog **Statistik (Durbin-Watson in Statistics)**) eingesetzt werden. Dieser Koeffizient erscheint in der Tabelle **Modellzusammenfassung (Model Summary)**. Bei welchen Koeffizienten der Experimentator beruhigt aufatmen kann, läßt sich leider nicht pauschal sagen. Dazu ist das Trend-Handbuch von SPSS nötig, dort ist die notwendige Tabelle abgedruckt mit jeweils zwei Werte  $dU$  und  $dL$ . Wenn der Durbin-Watson-

Koeffizient  $DW = dU$  ist, dann kann die Nullhypothese beruhigt angenommen werden, das heißt die Residuen sind *nicht* seriell abhängig. Wenn  $DW$  zwischen  $dU$  und  $dL$  liegt, ist die Entscheidung nicht zu beantworten, ob Abhängigkeit vorliegt.

Was aber zu tun bleibt, wenn dieser Test ungünstig ausfällt, kann in dieser Hausarbeit leider nicht beantwortet werden. Es sei auf die Umdrucke des Rechenzentrums (U.R.T 1997, 1998) oder der Fachliteratur verwiesen.

### **3.5 Modellselektion: Bestimmung der besten Regressionsgleichung**

Modellselektion ist ein schwieriges Unterfangen. Meistens bieten verschiedene Suchstrategien auch verschiedene Lösungen an, die rein statistisch betrachtet absolut gleichwertig sind. Dabei bleibt zusätzlich die Frage offen, welche Modelle womöglich noch existieren, aber von den vorliegenden Strategien nicht entdeckt werden.

#### **3.5.1 Theoriegeleitet**

Theoriegeleitetes Vorgehen ist an sich nur möglich, wenn ich bereits ein Modell/eine Regressionsgleichung habe, die ich an der vorliegenden Stichprobe validieren möchte. All die übrigen Situationen zeichnen sich eben dadurch aus, daß ich das gewünschte Modell nicht kenne. Trotzdem ist es unumgänglich, sich durch theoretische Überlegungen und Annahmen an ein vernünftiges Modell heranzupirschen (wie noch zu zeigen sein wird).

#### **3.5.2 Blockweise**

Einmalige Benutzung der Regression mit  $m$  Prädiktoren (SPSS-Methode **Einschluß (Enter)**). Man gibt diese Prädiktoren fest vor. Nur in diesem Fall sind die in SPSS angegebenen Signifikanztest richtig (GEDIGA & KUHN 1999)! Die Gefahr ist hoch, zu viele Prädiktoren zu benutzen und irrelevante Variablen in die Gleichung aufzunehmen.

Schrittweise multiple Regression vorwärts mit fester Reihenfolge (SPSS-Methode **Einschluß (Enter)** mit vorgegebenen Blöcken). Ähnlich der vorab vorgestellten Methode, wobei die Methode auf mehrere Sätze von Prädiktoren angewandt wird. Man sortiert die Prädiktoren nach Wichtigkeit und bricht ab, wenn  $k < m$  Variablen das Kriterium ausreichend beschreiben. Probleme: (a) Solch eine feste Reihenfolge der Wichtigkeit kann oft nicht angegeben werden. (b) Wann ist eine Variable ausreichend beschrieben? Statistische Tests, die in diesem Zusammenhang oft durchgeführt werden, sind wenig hilfreich und oft genug nicht einmal richtig!



Auf dieselbe Weise, wie eine feste Anzahl an Prädiktoren auf einmal ins Modell aufgenommen werden können, lassen sich diese auch wieder entfernen (SPSS-Methode **Ausschluß (Remove)**).

### **3.5.3 Vorwärts oder Rückwärts**

Schrittweise multiple Regression vorwärts mit vom Computer zu bestimmender Reihenfolge (SPSS-Methode **Vorwärts (Forward)**). Wie vorher, die Reihenfolge wird jedoch durch den jeweils „signifikantesten“ Prädiktor festgelegt. Diese (meistgewählte) Strategie hat das Problem, daß ein zu Anfang „signifikanter“ Prädiktor in Laufe der Aufnahme weiterer Prädiktoren insignifikant werden kann, da die anderen Prädiktoren diesen linear aufklären. Dann muß das ganze Verfahren noch einmal ohne diesen Prädiktor wiederholt werden.

Schrittweise Regression rückwärts (SPSS-Methode **Rückwärts (Backward)**). Statt sukzessive signifikante Prädiktoren in die Regressionsgleichung aufzunehmen, wird zunächst die lineare Regression mit einer Vielzahl von Prädiktoren berechnet und dann schrittweise der am wenigsten signifikante Prädiktor entfernt. Im Kontrast zur Vorwärts-Methode kann es hier vorkommen, daß ein eliminiertes Prädiktor zu einem späteren Zeitpunkt wieder wertvolle Beiträge zur Varianzaufklärung liefern könnte. Während also die Vorwärts-Methode riskiert, unnütze Variablen in die Gleichung aufzunehmen, werden hier womöglich wichtige Variablen übersehen.

### **3.5.4 Die Kombination von Vorwärts und Rückwärts: Schrittweise**

Zuerst wird hier (SPSS-Methode **Schrittweise (Stepwise)**) analog zur Vorwärts-Methode vorgegangen. Neu ist jedoch, daß nach Aufnahme eines Prädiktors die bereits vorhandenen Koeffizienten erneut auf Signifikanz überprüft werden. Wegen der (Multi-Kollinearitäten) kann es vorkommen, daß eine vorher als bedeutsam eingestufte Variable nun wieder eliminiert werden muß.

Diese Methode erscheint auf den ersten Blick die plausibelste und beste Strategie zur Modellauswahl zu sein. Simulationsrechnungen und Anwendungsbeispiele widerlegen jedoch diese Vermutung (HOCKING 1976). Auch sie ist nicht perfekt.

### **3.5.5 Diskussion**

Die Fragestellung nach der besten Methode läßt sich nur dann eindeutig lösen, wenn die potentiellen Einflußgrößen im ‚Pool‘ nahezu unkorreliert sind. In diesem Fall liefern alle vorgestellten Methoden dasselbe Ergebnis. Bei hohen Multi-Kollinearitäten ist die Frage nach dem besten Modell unbestimmt, da (zumindest oft) ein Teil der Prädiktoren gegeneinander

ausgetauscht werden kann. Welche dieser Prädiktoren letztlich in das Modell aufgenommen werden läßt sich *nicht statistisch* entscheiden.

Für die eine routinemäßige Anwendung herrschen über die beste Strategie geteilte Meinungen. MAGER (1982) favorisiert die Rückwärts-Methode; BORTZ (1993, S. 427) plädiert für die Vorwärts-Methode, nachdem die Variablen theoretisch in Blöcke vorstrukturiert worden sind.

## **4 Literatur**

- BORTZ (1993) „*Statistik für Sozialwissenschaftler*“; Berlin: Springer
- EFRON B. & TIBSHIRANI R. (1986) „*Bootstrap methods for standard errors, confidence intervals and other measures of of statistical accuracy*“; *Statistical Science* 1, S.54-77
- ERDFELDER E., FAUL F., & BUCHNER A. (1996) „[GPOWER](#): A general power analysis program. *Behavior Research Methods*“, *Instruments, & Computers*, 28, 1-11
- GEDIGA G. & KUHN T. (1999) „*Praktische Methodenlehre*“; <http://luce.psych.uni-osnabrueck.de/ggediga/www/pm98/>
- HOCKING R.R. (1976) „*The analysis and selection of variables in linear regression*“; *Biometrics* 32, S.1-49
- MAGER H. (1982) „*Moderne Regressionsanalyse*“; Salle & Sauerländer
- MYERS R. (1990) „*Classical and modern regression with applications*“; 2te Auflage, Boston, MA: Duxbury Press
- RYAN T.S. (1997) „*Modern Regression Methods*“; New York: Wiley
- STEVENS, J. (1996) „*Applied multivariate statistics for the social sciences*“; Mahwah: Lawrence Erlbaum
- U.R.T. (1997) „*Zeitreihenanalyse mit SPSS-Trends*“; Universitäts-Rechenzentrum Trier
- U.R.T. (1998) „*Einsatz von SPSS-Trends bei der Regressionsanalyse mit seriell abhängigen Daten*“; Universitäts-Rechenzentrum Trier
- U.R.T. (1999) „*Lineare Regressionsanalyse mit SPSS*“; Universitäts-Rechenzentrum Trier