h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Prof. Dr. Andreas Thümmel

Hochschule Darmstadt – University of Applied Sciences
Fachbereich Mathematik und Naturwissenschaften

Schöfferstr. 3
Geb. C10, Raum 9.31
D-64295 Darmstadt

Tel.: +49 (6151) 163-7951
Email: thuemmel@h-da.de
Web: http://www.thuemmel.eu

**Analyseschritte für GLM und Multilineare Regression**

Stand: 03.10.2017

**Präambel**

Die Analyse eines linearen Modells ist nicht einfach und erfordert ein gezieltes Vorgehen, um die gefundenen Effekte zu plausibilisieren, das Modell zu stabilisieren und final zu bewerten.

Die hier erstellte Anleitung stellt den Analyseprozeß in kurzen Worten stichwortartig dar. Es wird empfohlen, diesen Prozeß in Ihrem Booklet (zu den Vorlesungen QM, QM2, TAO, ggf. Statistik 1/2), der Abschlußarbeit (BA/MA) oder Projekten entsprechend zu verfolgen – jedoch nicht absolut. Mit Überraschungen ist insb. bei den realitätsnahen Fallbeispielen oder bei Echtdaten natürlich stets zu rechnen.

Bei Unklarheiten bitte melden (per Email, bitte beachten Sie unbedingt die Kommunikationsregeln, siehe Homepage).

**Analyseprozess**

1. Analyse ohne Reduktion von Variablen / Faktoren und ohne Wechselwirkungen

Ziel: erstmal die Haupteffekte „sehen“. Diese sollten sich in einem reduzierten Modell ebenfalls in ähnlicher Signifikanz / Bedeutung wiederfinden. Wenn nicht, dann wirft das ggf. Fragen auf bzgl. der dann notwendig im Modell enthaltenen Wechselwirkungen. Insb. die Beitrags%-Werte sind dann zu plausibilisieren.

   a. Sind R^2, R^2(prog.) genügend hoch ?
   b. Sind die VIF-Werte < 5 ?
   c. Term „Fehlende Anpassung“ nicht signifikant ?
   d. Sind Haupteffekte plausibel (z.B. bekannt) ?
   e. Verhältnis der Beitrag% zueinander und zum Fehler-Beitrag

2. Hinzunahme möglichst vieler WW-Terme und Terme höherer Ordnung (Freiheitsgrade ggf. = 0? Achtung: sehr instabiles Modell, Reduktion unbedingt erforderlich) und Aktivieren der Schrittweisen Reduktion (Rückwärtseliminierung).

   Vorbereiten der Reduktion nach dem Prinzip der Sparsamkeit so, dass das Modell stabil wird. Wechselwirkungen vom Grad 3 und mehr sind ggf. trotz Signifikanz zu eliminieren (unter aller gebotener Vorsicht), da diese recht selten erwartet werden.

   a. Sind Wechselwirkungen signifikant ? Sind diese plausibel ?
   b. Sind nichtlineare Terme signifikant ? Sind diese plausibel ?
   c. Ergibt sich bzgl. der Haupteffekte das gleiche oder ein sehr ähnliches Modell ?

3. Eliminieren von signifikanten, aber unbedeutenden Termen

Wesentlich für die Kontrolle eines Modells ist der Beitrags%-Satz. Die Sensitivität der betrachteten Zielgröße wird auch bei Nichtsignifikanz der unabhängigen Variablen sehr hoch sein. Ggf. diesen Fall in Einzeldiagrammen (Haupteffekte- und Wechselwirkungsdiagrammen) deutlich.

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Prof. Dr. Andreas Thümmel

Hochschule Darmstadt – University of Applied Sciences
Fachbereich Mathematik und Naturwissenschaften

Schöfferstr. 3
Geb. C10, Raum 9.31
D-64295 Darmstadt

Tel.: +49 (6151) 163-7951
Email: thuemmel@h-da.de
Web: http://www.thuemmel.eu

   a. Sind die Beitrags%-Werte für signifikante Terme max. im Bereich von 50% des Fehler-Beitrag%-Wertes, dann nehmen Sie diesen Term ebenfalls aus dem Modell

4. VIF-Werte eliminieren

Abhängigkeiten in den Unabhängigen kann zu undeutlichen Modellen führen. Die manuelle Reduktion von abhängigen Termen führt eigentlich zu einer Reduktion des Anpassungsgrades. Dieses ist bei mehreren so auffälligen Variablen zu überprüfen durch kombinatorische Modellreduktion.

   a. Ggf. sind mehrere Eliminierungen notwendig; das Modell sollte final keine VIF-Werte > 5 mehr enthalten und die im Modell verbliebenen Terme sollten aus praktischer Sicht kausal sein
   b. Eliminierung kombinatorisch

5. Aus dem so nach dem Prinzip der Sparsamkeit minimalen Modell signifikante Faktoren / Variablen rausnehmen

Das bis zu diesem Punkt entstandene Modell nun testen, ob auch tatsächlich alle Faktoren / Variablen signifikant und mit hohem Beitrags%-Wert in das Modell eingehen. Zudem am aktuellen Modell die Stichprobengröße als ausreichend für ein stabiles Modell überprüfen.

   a. Ist R^2 dann dennoch weiterhin genügen groß ?
   b. Ist die Differenz zwischen R^2 und R^2(prog) genügend klein (Faustregel: <= 10%) ? Wenn nicht, dann ist die Stichprobengröße zu klein → weitere Stichprobenentnahme, möglichst Balanciert (überprüfen durch auszählen) und orthogonal (überprüfen durch Berechnung der Korrelationskoeffizienten zwischen den Prädiktoren). Im Falle DoE: Berechnung der Stichprobeneinstellungen der Prädiktoren durch D-Optimalität

**Schlußbemerkung**

Ziel dieser Betrachtungen ist es schlußendlich, ein Modell zu bekommen, daß

- Genügend hohe Anpassungsgrade R^2 und (vor allem !) R^2(prog) besitzt (Stabilität)
- Keine VIF-Werte > 5 enthält (Optimierung liefert eindeutige Einstellungen)
  Alle gefundenen signifikanten Terme plausibel sind anhand der folgenden weiteren Kriterien (zu überprüfen anhand „Wissen“ bzw. „Erfahrung“ bzgl. des betrachteten Prozesses, d.h. Expertenwissen erforderlich für eine qualitative Bewertung):

   a. Vorzeichen
   b. Verhältnis der Terme zueinander
   c. Größenordnung der Terme (Hebelwirkungen)
   d. physikalisch-technische Wirkung ist plausibel bzw. erwartet (bekannt)