

# Quantitative Analysen medizinischer Prüfungen mittels eines (teil)automatisierten SAS-Programms

Rainer Muche, Brigitte Janz, Beate Einsiedler

Institut für Biometrie

Universität Ulm

Schwabstraße 13

89075 Ulm

rainer.muche@uni-ulm.de

## Zusammenfassung

Seit Änderung der Ärztlichen Approbationsordnung im Jahr 2003 werden vermehrt benotete Prüfungen an den Universitäten durchgeführt, die die Leistungen der Studierenden bescheinigen (sollen). Diese Prüfungen sollen deshalb natürlich fair, objektiv, reliabel und valide sein, um die einzelnen Studierenden gerecht zu bewerten.

Zur Untersuchung dieser Eigenschaften können verschiedene statistische Auswertungen herangezogen werden. Es gibt dazu entsprechende Vorschläge [2], die diese Auswertungen festlegen. Zur Untersuchung der Eigenschaften der eigenen Prüfungen im Pflichtunterricht Biometrie [3] haben wir diese vorgeschlagenen Auswertungen mit SAS umgesetzt.

Auf Anfrage des Studiendekanats wird nun ein (halb-)automatisiertes SAS-Modul zur Auswertung auch verschiedener Prüfungsformate erstellt. Dieses muss ein automatisches Einlesen der Prüfungsdaten (nach Vorgabe der Items und Datenstruktur) und verschiedene Auswahlmöglichkeiten von Auswertungen berücksichtigen.

In dem Beitrag werden die Analysen als Methodik und umgesetzt im Programm vorgestellt.

**Schlüsselworte:** Prüfungen, Reliabilität, Aufgabenschwierigkeit, Trennschärfe

## 1 Einleitung

Seit Änderung der Ärztlichen Approbationsordnung im Jahr 2003 werden vermehrt benotete Prüfungen an den Universitäten durchgeführt, die die Leistungen der Studierenden bescheinigen (sollen). Diese Prüfungen sollen deshalb natürlich fair, objektiv, reliabel und valide sein, um die einzelnen Studierenden gerecht zu bewerten.

Zur Untersuchung dieser Eigenschaften können verschiedene statistische Auswertungen herangezogen werden. Es gibt dazu entsprechende Vorschläge [2], die diese Auswertungen festlegen. Zur Untersuchung der Eigenschaften der eigenen Prüfungen im Pflichtunterricht Biometrie [3] haben wir diese vorgeschlagenen Auswertungen mit SAS programmiert und stellen das Programm Kollegen in der Fakultät zur Verfügung bzw. werten entsprechende Prüfungsdaten aus.

## 2 Datenstruktur der Prüfungsergebnisse

Die Grundlage für die Auswertung von Prüfungen bilden 2 Datensätze, in denen die Prüfungsergebnisse dokumentiert werden.

### Beispieldatei Aufgabenpunkte.xls:

StudentID	Gruppe_sub	Version_sub	Aufg1	Aufg2	Aufg3	Aufg3_mc
73977	A	2	6	5	0	c
74515	A	2	6	8	3	b
40593	A	2	3	6	3	b
83352	A	2	3	8	0	c
53385	A	1	6	8	3	b
22256	B	1	6	8	0	a
48484	B	1	3	8	0	a
19077	B	1	9	8	0	a
81188	B	1	9	8	3	b
74892	B	1	9	8	3	b
49146	B	2	6	8	3	b

**Abbildung 1:** Datenstruktur der Prüfungsergebnisse pro Studierendem (in MS-Excel)

In der ersten Datei werden für jeden Studierenden die Ergebnisse dokumentiert. Folgende Variablen werden benötigt:

Die Datei Aufgabenpunkte.xls enthält die Punkteangaben zu allen Aufgaben eines Studenten in einer Zeile. Dabei erfolgt die Identifikation jedes Studenten mit der fest vorgegebenen Variable „StudentID“. Anzahl und Namen der Aufgaben sind variabel und frei wählbar.

Zusätzlich ist die Angabe von maximal 2 Subanalyseparametern, wie z.B. Gruppe und/oder Version, möglich (Extension\_sub zum Namen des Parameters). Anzahl und Art der Ausprägungen dieser Subanalyseparameter sind ebenfalls variabel und frei wählbar.

Für spezielle Auswertungen der MC-Aufgaben kann nach Wunsch ergänzend zu den Aufgabenpunkten in einer weiteren Spalte die jeweilige MC-Lösung (a-z) jedes Studenten mit gleichem Aufgabennamen und Extension\_mc angegeben werden.

### Beispieldatei Aufgaben.xls:

Variable	Fach	Semester	Maximum	Typ	Anz_Antwort_mc	Ergebnis_mc
Aufg1	Biometrie	SS2008	9	2		
Aufg2	Biometrie	SS2008	8	2		
Aufg3	Biometrie	SS2008	3	1	4	b

**Abbildung 2:** Datenstruktur der Prüfungsaufgaben (in MS-Excel)

Die Angaben zu jeder Aufgabe stehen in der Datei Aufgaben.xls (zeilenorientiert). Dabei müssen die gleiche Anzahl und die gleichen Namen wie in der Datei Aufgabenpunkte.xls vorliegen. Je Aufgabe ist der Typ (FT = Freitext oder MC = Multiple

Choice) und die maximale Punktzahl der Aufgabe anzugeben. Für zusätzliche Auswertungen bei MC-Aufgaben können Anzahl und richtige Lösung (a-z) angegeben werden.

Insgesamt sind so die wesentlichen Informationen für die Prüfungsauswertung zusammengestellt, die in ein SAS-Auswertungsprogramm übernommen werden. Bevor der Programmentwurf erläutert wird, soll zunächst im nächsten Abschnitt die Auswertung angegeben werden.

### 3 Auswertungen / Statistische Kenngrößen

Es gibt einige Vorschläge zur Untersuchung von Prüfungsergebnissen, um deren Eigenschaften insgesamt oder aufgabenspezifisch untersuchen zu können. Um die Prüfungen in der Humanmedizin zu unterstützen und vergleichbar zu machen, wurde in Baden-Württemberg das „Kompetenzzentrum für Prüfungen in der Medizin“ an der Medizinischen Fakultät der Universität Heidelberg eingerichtet.

([http://www.medizinische-fakultaet-hd.uni-heidelberg.de/KompetenzZentrum-fuer-Pruefungen-Komp\\_Med.107007.0.html](http://www.medizinische-fakultaet-hd.uni-heidelberg.de/KompetenzZentrum-fuer-Pruefungen-Komp_Med.107007.0.html)), zuletzt geprüft am 21.06.2009. Ein Vorschlag für Auswertungen wurde von den Mitarbeitern des Kompetenzzentrums als Vorschlag veröffentlicht [2], den wir für unser Programm weitestgehend übernommen haben [1].

Folgende Aspekte werden ausgewertet:

#### Gesamtergebnis:

- Darstellung der Gesamtpunktzahlen  
Tabellarischen Auflistung der Häufigkeiten, ergänzt durch Bestehensgrenzen und Notenschlüssel  
Histogramm

#### Aufgabenschwierigkeitsgrad

- **Aufgabenschwierigkeit** bzw. Itemschwierigkeit:  
bei Aufgaben mit genau einem bzw. keinem Punkt = Mittlere bei dieser Aufgabe erreichte Punktzahl = Relative Häufigkeit der richtigen Antwort

PROC MEANS

- **Normierte Aufgabenschwierigkeit** = P:  
bei Aufgaben mit mehreren Punkten = Relativer (prozentualer) Anteil der maximal erreichbaren (nicht der maximal erreichten) Punktzahl der Aufgabe  
= Mittelwert der erreichten Punkte / Maximum der erreichbaren Punkte dieser Aufgabe

PROC MEANS

### **Trennschärfe**

- **Diskriminationsindex = D:**

Voraussetzung:

Einteilung der Prüflinge in drei Gruppen (gut / mittel / schlecht) anhand der Terzile (unteres und oberes 33 %-Perzentil)

D = Differenz der normierten mittleren Schwierigkeit der Gruppe der "Guten" und der Gruppe der "Schlechten" =  $P_{\uparrow} - P_{\downarrow}$

( $P_{\uparrow}$  = Mittlere erreichte normierte Punktzahl der "Guten")

( $P_{\downarrow}$  = Mittlere erreichte normierte Punktzahl der "Schlechten")

PROC MEANS

- **Trennschärfe über Korrelationsmaße:**

Korrelationskoeffizienten (Pearson-Bravais, Spearman, Kendall)

= Korrelation der in der Aufgabe erreichten Punktzahlen mit der Summe der Punkte in allen Aufgaben

PROC CORR SPEARMAN PEARSON

- **Korrigierte Trennschärfe über Korrelationsmaße:**

Korrelationskoeffizienten (Pearson-Bravais, Spearman, Kendall)

= Korrelation der in der Aufgabe erreichten Punktzahlen

mit der Summe der Punkte in allen anderen Aufgaben als der auszuwertenden

PROC CORR SPEARMAN PEARSON

### **Einzelantwort- oder Distraktorenanalyse bei MC-Aufgaben**

- Häufigkeitsanalyse der Antworten:

Alle Antwortalternativen, ihre absoluten und relativen Häufigkeiten.

(Die relativen Häufigkeiten entsprechen der Aufgabenschwierigkeit.)

- Trennschärfeanalyse von Antworten/Distraktoren:

Erstellung einer Indikator-Variablen für jede Antwortmöglichkeit.

Für jeden Studierenden den Wert 1, wenn er diese Antwort gewählt hat, ansonsten 0.

Berechnung der Trennschärfemaße mit dieser Indikatorvariablen.

### **Reliabilität** (Zuverlässigkeit oder Reproduzierbarkeit)

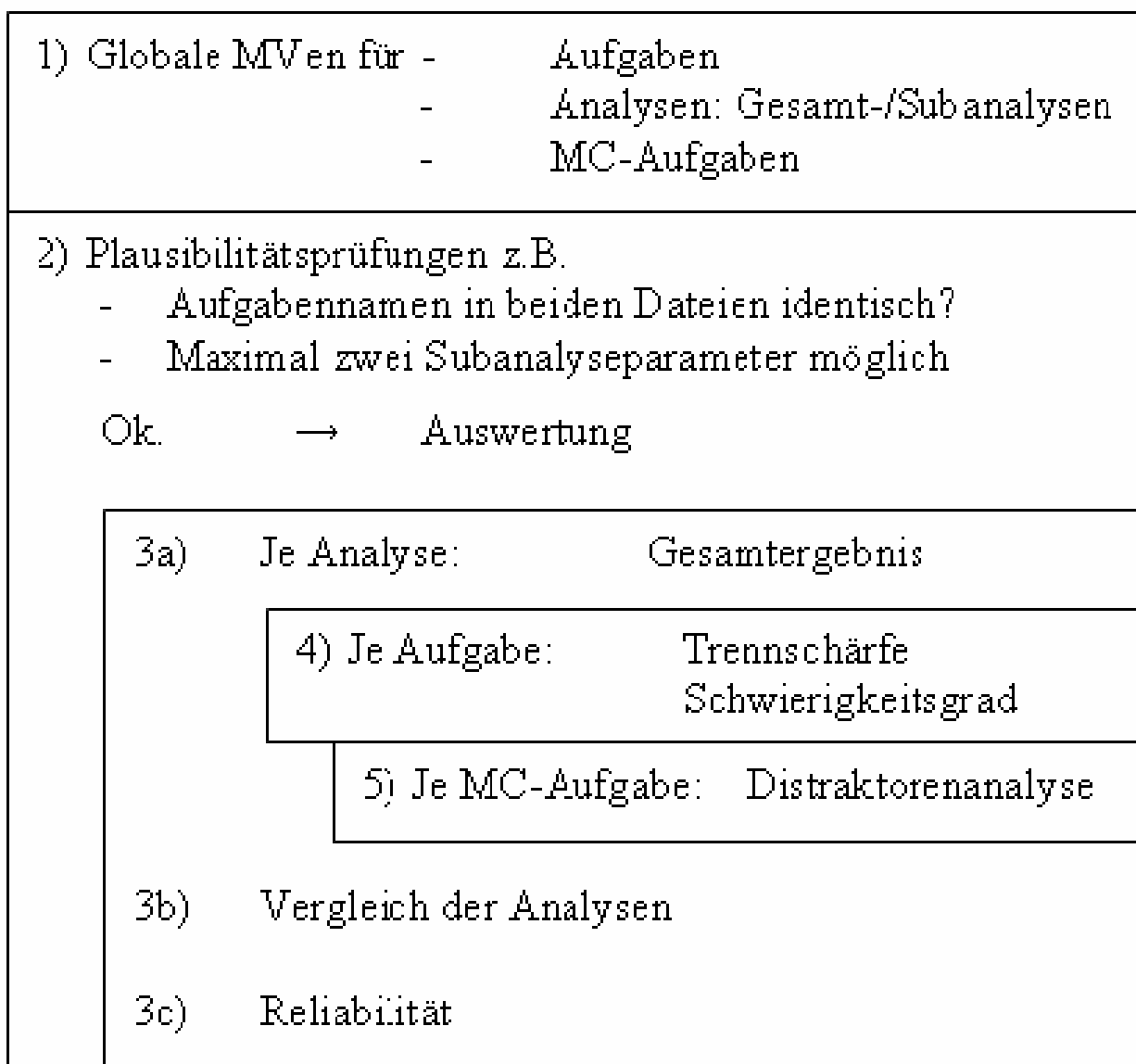
Berechnung mit Cronbachs  $\alpha$ , wobei Cronbachs  $\alpha$  eine untere Grenze für die Reliabilität darstellt.

PROC CORR ALPHA

## 4 Konstruktion des SAS-Auswertungsprogramms

Folgende Struktur wurde zur Umsetzung der verschiedenen Auswertungsansätze gewählt:

- Indirekte Pfade
- Auswertung und Programmsteuerung mittels globaler Makrovariablen (MVen) und Macros



**Abbildung 3:** Programmstruktur

Nach Definition von Makrovariablen (MVen) für die indirekten Pfade (SAS\_EXECFILEPATH) werden im SAS-Programm „Aufgabenpruefung.sas“ Optionen und Formate definiert und alle Makros und Datendateien mit den MVen der Pfade geladen.

Im Hauptprogramm werden dann anhand der Datendateien zu allen Aufgaben und ggf. zu Subanalyseparametern und MC-Aufgaben globale MVEN erzeugt, u. a. für Anzahlen und Variablennamen.

Die Abarbeitung der Auswertungen wird streng hierarchisch mittels Makros durchgeführt. Je Analyse, Aufgabe und MC-Aufgabe werden die jeweiligen Auswertungen im jeweils untergeordneten Makro durchgeführt (siehe Abb. 3). Dabei erfolgt Steuerung und Aufruf der Makros im Wesentlichen durch die globalen MVEN, die ja im gesamten Programm zur Verfügung stehen.

### Beispiel:

Es wird immer eine Gesamtanalyse sowie so viele Subanalysen durchgeführt, wie Ausprägungen des/der Subanalyseparameter vorliegen (zu Aufgabenpunkte.xls, siehe oben).

#### *Biometrie im SS2008* *Übersicht Gesamt-/Subanalysen*

Obs	Analysenr.	Subanalyse ja/nein	(Sub)Analyse- Parameter	Ausprägung	Anzahl Studenten je Ausprägung
1	0	nein	Alle Studenten	Gesamt	11
2	1	ja	GRUPPE_SUB	A	5
3	2	ja	GRUPPE_SUB	B	6
4	3	ja	VERSION_SUB	1	6
5	4	ja	VERSION_SUB	2	5

**Abbildung 4:** Beispiel Auswertungsmöglichkeiten mit Untergruppen

Sind keine Subanalyseparameter in der Datei Aufgabenpunkte.xls angegeben, so wird nur eine Gesamtanalyse durchgeführt.

Ähnlich wird bei den MC-Aufgaben vorgegangen: Die Distraktorenanalyse wird nur für MC-Aufgaben vorgenommen, wenn die jeweiligen MC-Lösungen der Studenten vorliegen und die MC-Angaben zur Aufgabe (Anz\_Antwort\_mc und Ergebnis\_mc) angegeben sind.

Für alle Analysen wird jeweils Gesamtergebnis und Reliabilität bestimmt, sowie für alle Aufgaben Schwierigkeitsgrad und Trennschärfe und ggf. zusätzlich für MC-Aufgaben eine Distraktorenanalyse berechnet. Die Ergebnisse werden in Tabellen und Graphiken dargestellt.

Beispiel einer Übersichtstabelle je Analyse:

**Biometrie im SS2008**  
**Trennschärfe und Diskriminationsindex**  
**Analyse: Alle Studenten (N=11)**

					Trennschärfe (korrigiert*)		Trennschärfe (unkorrigiert)		Diskri- minations- index (korr.*)	Diskri- minations- index (unkorr.)
Aufg.- name	Aufg. typ	Max. Punkte	Mittl. Punkte	Norm. Aufg. schw. P	Spearman r1(*)	Pearson r2(*)	Spearman r1	Pearson r2	D(*)	D
Aufg1	FT	9.0	6.00	0.667	0.3560	0.3242	0.8266	0.8445	0.2000	0.3889
Aufg2	FT	8.0	7.55	0.943	0.3081	0.2542	0.4749	0.5192	0.1250	0.1250
Aufg3	MC	3.0	1.64	0.545	0.2711	0.2577	0.6878	0.6485	0.4643	0.8333

Richtwert für die normierte Aufgabenschwierigkeit P: 0.4 - 0.8

Richtwerte Diskriminationsindex:  $D \geq 0.3$  ist gut,  $0.3 > D \geq 0.2$  ist akzeptabel,  $D < 0.1$  ist schlecht.

Richtwerte Trennschärfe:  $r \geq 0.3$  ist gut,  $0.3 > r \geq 0.2$  ist akzeptabel,  $r < 0.1$  ist schlecht.

(\*) Korrigierte Version mittels der Summe der Punkte aller anderen Aufgaben.

Aufgabentyp: MC = Multiple Choice, FT = Freitext

Auswertung vom 31/03/2009

**Abbildung 5:** Beispiel-Auswertungen

## Fazit / Diskussion / Ausblick

Das Programm wurde kurz vor der KSFE 2009 fertig, so dass bis auf Testläufe an kleinen realen Datensätzen aus dem Institut für Biometrie [1] noch nicht getestet werden konnte. Für die nächste Zeit liegen u.a. Prüfungsergebnisse aus der Dermatologie vor, an denen das Programm validiert und auf einen Routineeinsatz hin überprüft wird.

Mehrere Ausbaumöglichkeiten sind schon identifiziert. So fehlen Auswertungen für MC-Fragen mit Mehrfachantworten. Diese sind bisher nicht aufgenommen worden, da das Kompetenzzentrum für Prüfungen in der Medizin diesen Aufgabentyp nicht empfiehlt. Er wird aber an unserer Fakultät in einigen Prüfungen eingesetzt. Die Ausgabeform, speziell die grafischen Darstellungen sind ebenfalls noch nicht endgültig.

Der Vorteil gegenüber einem Einsatz von kommerziellen Programmen (s. Literatur) ist die Möglichkeit, die Auswertungen selber verändern zu können, um z.B. auf veränderte Anforderungen schnell reagieren zu können.

Der wichtigste Einsatz wird die Auswertung von Prüfungsformen sein, die nicht unbedingt schriftlich mit MC-Aufgaben auskommen können, so den vermehrt eingesetzten

OSCE-Prüfungen (Prüfung von Fertigkeiten) oder unseren Prüfungen im PC-Kurs Biometrie [4].

## Weitere Software und Literatur

Entsprechende, alternative Auswertungssoftware für die oben angegebenen Auswertungen ist kommerziell verfügbar. Dies sind unter anderem

- **KlausurenKlaus:** [www.klausuren-klaus.de](http://www.klausuren-klaus.de) zuletzt besucht am 21.06.2009
- **IMSmedizin:** [www.ims-m.org](http://www.ims-m.org) zuletzt besucht am 21.06.2009

Herr Möltner [2] hat auf der Basis von Excel, R, TeX ebenfalls eine entsprechende Software programmiert (bisher unpubliziert).

## Literatur

- [1] Janz B. (2008): Umsetzung und Validierung / Evaluierung der Prüfungssituation des Lehrprojektes „Biometrie“ im Humanmedizinstudium an der Universität Ulm, Studienarbeit Medizinische Dokumentation, SMD Ulm
- [2] Möltner A., Schellberg D., Jünger J. (2006): Grundlegende quantitative Analysen medizinischer Prüfungen. GMS Z. Med. Ausbildung, 23(3);Doc53
- [3] Muche R., Weirather-Herrlein S., Wildt M., Radlinger K., Seefried K. Jordan J. (2005): Dokumentation des Lehrprojektes „Biometrie“ an der Universität Ulm. Shaker Verlag, Aachen
- [4] Muche R., Kocak S., Jäckel E., Janz B., Einsiedler B. (2009): Automatisierte Unterstützung für Prüfungen in Statistiksoftwarekursen im Humanmedizin-Studium. KSFE 2009, Shaker Verlag, Aachen