

## **Erfahrungen beim Portieren eines Makro-Pakets von SAS 9.1 auf SAS 9.4**

Sandra Müller  
Universität Ulm  
Schwabstr. 13  
89075 Ulm

sandra.mueller@uni-ulm.de

Rainer Muche  
Universität Ulm  
Schwabstr. 13  
89075 Ulm

rainer.muche@uni-ulm.de

### **Zusammenfassung**

Das SAS-Makropaket „Prognosemodellierung auf Basis der logistischen Regression“ ist 2004 entstanden und u.a. 2005 auf der KSFE in Berlin vorgestellt worden [1,2]. Damit wird eine Vorgehensweise zur Prognosemodellierung auf Basis der logistischen Regression vorgestellt, deren Umsetzung in der Praxis durch neu entwickelte SAS-Makros bzw. den sinnvollen Einbau bereits vorhandener SAS-Makros unterstützt wird. Die Modellierung und Überprüfung der Prognosegüte erfolgt dabei im Wesentlichen in drei Schritten: Modellentwicklung, Bestimmung der Prognosegüte und Modellvalidierung.

Das Makropaket (14 Hauptmakros, 23 Hilfsmakros) wurde in der SAS-Version 8.2 programmiert und 2006 auf Kompatibilität mit der damals neuen SAS-Version 9.1 geprüft und überarbeitet [3]. Jetzt soll das Makropaket im Rahmen einer großen Beobachtungsstudie in der Neurologie wieder eingesetzt werden, um ein Prognosemodell zu generieren. Dafür ist eine Überprüfung der Makros in der SAS-Version 9.4 notwendig.

Dabei sind einige Änderungen in SAS aufgefallen, die die Lauffähigkeit der Makros unter der neuen Version beeinträchtigen. Unter anderem sind Neuerungen, die die Nutzung der SAS-Software unter der neuen Version erleichtern, schwierig für den Ablauf „alter“ Makros. Die Abwärtskompatibilität ist nicht vollständig gegeben. So ist z.B. die Einführung von ODS-Output (z.B. als HTML) und automatischer Grafikausgabe z.B. in den Prozeduren PROC UNIVARIATE und PROC LOGISTIC für die Lauffähigkeit alter Makros problematisch. Außerdem wird in den neuen Versionen genauer gerechnet. So ist z.B. die Option BINWIDTH in PROC LOGISTIC für einige Änderungen der Ausgabe der Ergebnisse zwischen den beiden Versionen 9.1 und 9.4 verantwortlich.

In dem Beitrag werden die notwendigen Änderungen, die für eine Wiederherstellung der Lauffähigkeit notwendig waren, erläutert und Korrekturmöglichkeiten aufgezeigt. Die Makros sowie die Dokumentation [4] sind auf der Homepage des Instituts [5] verfügbar.

**Schlüsselwörter:** Makro, SAS-Version, Logistische Regression, Binwidth, ODS

## 1 Einleitung

Das SAS-Makropaket „Prognosemodellierung auf Basis der logistischen Regression“ ist 2004 entstanden und u.a. 2005 auf der KSFE in Berlin vorgestellt worden [1,2]. Damit wird eine Vorgehensweise zur Prognosemodellierung auf Basis der logistischen Regression vorgestellt, deren Umsetzung in der Praxis durch neu entwickelte SAS-Makros bzw. den sinnvollen Einbau bereits vorhandener SAS-Makros unterstützt wird. Die Modellierung und Überprüfung der Prognosegüte erfolgt dabei im Wesentlichen in drei Schritten: Modellentwicklung, Bestimmung der Prognosegüte und Modellvalidierung.

Das Makropaket (14 Hauptmakros, 23 Hilfsmakros) wurde in der SAS-Version 8.2 programmiert und 2006 auf Kompatibilität mit der damals neuen SAS-Version 9.1 geprüft und überarbeitet [3]. Jetzt soll das Makropaket im Rahmen einer großen Beobachtungsstudie in der Neurologie wieder eingesetzt werden, um ein Prognosemodell zu generieren. Dafür ist eine Überprüfung der Makros in der SAS-Version 9.4 notwendig.

## 2 Das SAS-Makropaket

**„Prognose ist eine Vorhersage über den zukünftigen Verlauf einer Krankheit nach ihrem Beginn.“** Nach dieser Definition können Prognosen in der Medizin Therapieentscheidungen direkt oder indirekt mitbestimmen und sollten daher so zuverlässig wie möglich erstellt werden. Neben der subjektiven ärztlichen Einschätzung zum zukünftigen Krankheitsverlauf können Prognosen auch auf Grundlage entwickelter, mathematischer Modelle gegeben werden. Dabei handelt es sich oft um verallgemeinerte lineare Regressionsmodelle, wie z.B. das multiple logistische Regressionsmodell, das im Fall dichotomer Zielgrößen, wie sie häufig im klinischen Alltag beobachtet werden, zur Anwendung kommt.

Im Folgenden wird *eine* Vorgehensweise zur Prognosemodellierung auf Basis der logistischen Regression vorgestellt, deren Umsetzung in der Praxis durch neu entwickelte SAS-Makros bzw. den sinnvollen Einbau bereits vorhandener SAS-Makros unterstützt wird. Die Modellierung und Überprüfung der Prognosegüte erfolgt dabei im Wesentlichen in drei Schritten: (1) Modellentwicklung, (2) Bestimmung der Prognosegüte und (3) Modellvalidierung [1,2].

Im Folgenden werden die Makros (s. Abbildung 1) kurz beschrieben:

### **Modellentwicklung**

#### **PM\_DESCRIPTION.MAC.SAS**

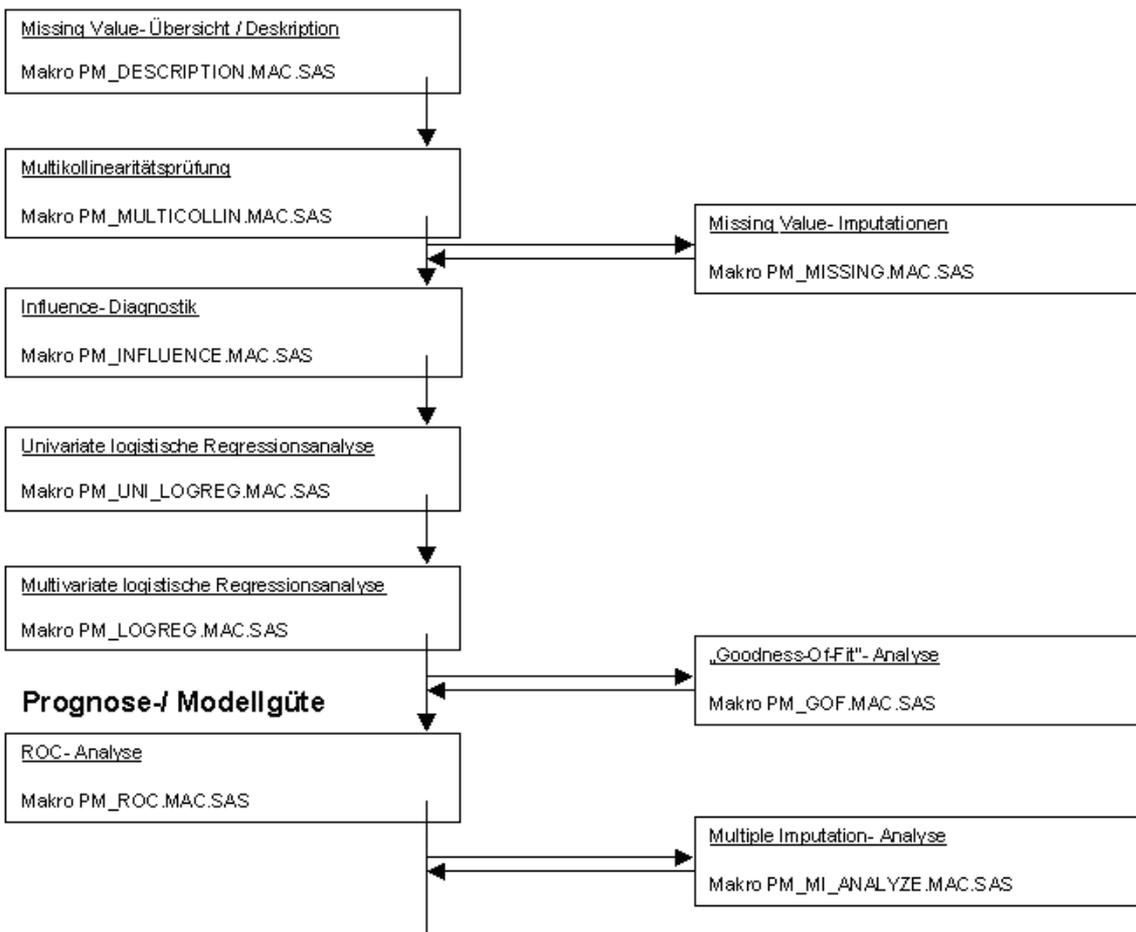
Deskriptive Auswertung der Modellvariablen. Complete Case-Analyse sowie Auszählung der fehlenden Werte pro Beobachtung möglich.

#### **PM\_MULTICOLLIN.MAC.SAS**

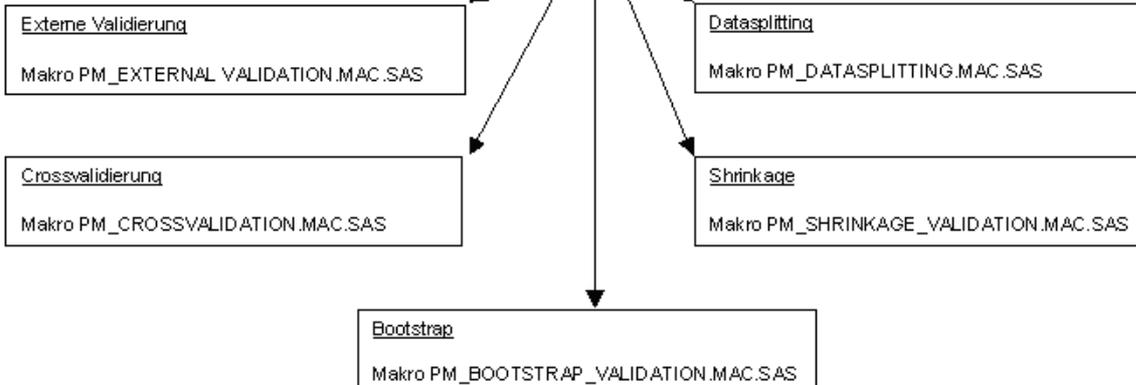
Untersuchung der Multikollinearität durch paarweise Korrelationen, Varianzinflationsfaktoren und Condition-Indizes (Belsley 1991). Auswertung mit linearer Regression, gewichtet mit geschätzter Wahrscheinlichkeit aus PROC LOGISTIC (Allison 1999).

## Schematischer Ablauf der Prognosemodellierung auf Basis der logistischen Regression

### Modellentwicklung



### Modellvalidierung



**Abbildung 1:** Schematischer Ablauf der Prognosemodellierung auf Basis der logistischen Regression

### **PM\_MULTICOLLIN.MAC.SAS**

Untersuchung der Multikollinearität durch paarweise Korrelationen, Varianzinflationsfaktoren und Condition-Indizes (Belsley 1991). Auswertung mit linearer Regression, gewichtet mit geschätzter Wahrscheinlichkeit aus PROC LOGISTIC (Allison 1999).

### **PM\_MISSING.MAC.SAS / PM\_MI\_ANALYZE.MAC.SAS**

Complete-Case-Analyse ist Standard. Möglichkeiten zur Ersetzung fehlender Werte: Single Imputation (PROC STDIZE), Multiple Imputation (PROC MI). Zusammenfassen der Prognosegüten über Makro PM\_MI\_ANALYZE.MAC.SAS

### **PM\_INFLUENCE.MAC.SAS**

Identifikation einflussreicher Beobachtungen für die Modellierung. Residuenanalyse (Veränderung der Pearson-Statistik nach Entfernen einer Beobachtung).

### **PM\_UNI\_LOGREG.MAC.SAS**

Univariater Einfluss der Variablen im Modell. Complete Case-Analyse möglich. Zur Untersuchung der Modellierung stetiger Variablen univariat „Fractional Polynomials“ (Royston, Altman 1994) installiert.

### **PM\_LOGREG.MAC.SAS**

Multiple logistische Regression. Korrekturmöglichkeiten für „Data Separation“ (FL-Makro, Heinze 2002). Probleme in PROC LOGISTIC umgangen. (TEST Statement mit CLASS, Ausgabe Datensätze für ROC-Analyse).

### **PM\_GOF.MAC.SAS**

Überprüfung der Modellanpassung. Neben Parametern aus PROC LOGISTIC spezielle Tests für Sparseness (wenige Beobachtungen pro Merkmalskombination) integriert (Makros von Kuss 2002, Pulkstenis/Robinson 2002).

## **Bestimmung der Prognosegüte / ROC-Analyse**

### **PM\_ROC.MAC.SAS**

Es werden Cutpoint-abhängige Gütemaße (Sensitivität, Spezifität, prädiktive Werte) sowie globale Gütemaße (AUC, Somer's D, Emax, Brier Score) ausgegeben. ROC-kurven wahlweise mit punktwisen und simultanen Konfidenzbändern (Hilgers 1991).

## **Modellvalidierung**

Zur Abschätzung des Bias der Überschätzung der Prognosegüte werden fünf verschiedene Methoden angeboten:

### **PM\_EXTERNAL\_VALIDATION.MAC.SAS**

Bei der externen Validierung wird die Prognosegüte des Modells anhand eines zweiten, unabhängigen Datensatzes bestimmt.

### **PM\_DATASPLITTING.MAC.SAS**

Durch Data-Splitting wird der Datensatz geteilt. Ein Teil wird zur Modellentwicklung, der zweite zur Validierung (s. externe Validierung) genutzt.

### **PM\_CROSSVALIDATION.MAC.SAS**

Ziehen ohne Zurücklegen: Datensatz in K Teile geteilt; anschließend an K-1 Teilen das Modell entwickelt und am K-ten Teil validiert. Wiederholung für alle K.

Methoden: K-fold Crossvalidation, adjusted Crossvalidation (Davison, Hinkley 1997), Jackknife-Crossvalidation.

**PM\_BOOTSTRAP\_VALIDATION.MAC.SAS**

Ziehen mit Zurücklegen: Neue Datensätze erzeugen, daran Modell validieren.

Methoden: Simple -, Enhanced Bootstrap (Efron, Tibshirani 1994), Mean Model.

**PM\_SHRINKAGE\_VALIDATION.MAC.SAS**

Korrektur der Regressionskoeffizienten durch Shrinkage-Faktoren. Methoden: heuristischer und globaler Shrinkage (van Houwelingen 1990), parameterbezogener Shrinkage (Sauerbrei 1999).

Das komplette Paket besteht aus 102 Makro-Dateien, davon enthalten 14 Dateien die eigentlichen Makroaufrufe (Hauptmakros). In 65 Dateien sind Fehlermeldungen gespeichert, die verbleibenden 23 Makros sind Hilfsmakros, die von anderen Makros aufgerufen werden. Die Hilfsmakros kommen teilweise von externen Stellen, anderen Universitäten etc. Der Standard-Aufruf der Makros ist in Abbildung 2 zu sehen.

```

%MACRO macroname ( data      =,
                   resp_var  =,
                   event     =1,
                   xvar      =,
                   cvar      =,
                   ref       =,
                   miss      =0,
                   ...
                   weitere spezifische Parameter,
                   ...
                   macro_path =,
                   );

%MEND macroname;

```

Abbildung 2: Standard-Aufruf für die Makros

### 3 Überprüfung der Makros

Zum Testen der Makros wurde derselbe Datensatz aus der rehabilitationsmedizinischen Forschung benutzt, an dem auch die Makros in [1] entwickelt wurden. Beim ersten Erproben wurde die Situation dargestellt, dass das Makro mit allen dort beschriebenen Einstellungen der Parameter gestartet wurde. Verglichen wurde der Output des Makros mit dem Output, der in [1] abgedruckt ist. Außerdem wurden die Meldungen im LOG-Fenster nach Warnungen und Fehler untersucht und unter Umständen erzeugten Dateien auf ihre Korrektheit überprüft.

Im Jahr 2006 wurde das Makro-Paket, das in der SAS Version 8 programmiert wurde, auf Kompatibilität mit der damals neuen SAS-Version 9.1 geprüft und überarbeitet. Es wurden damals einige Notwendigkeiten für Änderungen gefunden, die in [3] dokumentiert sind. Die jetzige Überprüfung wird anhand der Makros für die Version 9.1 in der SAS-Version 9.4 durchgeführt und auf Auffälligkeiten untersucht. Mögliche Änderungen werden im Kapitel 4 beschrieben, die den „Standard“-Output wiederherstellen.

## 4 Auffälligkeiten / notwendige Änderungen

### 4.1 Grafiken

In den SAS Makros werden Grafikoptionen wie Speicherort und Dateiformat durch GOPTIONS gesteuert. In SAS Version 9.4 werden diese Angaben jedoch vollständig ignoriert, wenn ODS GRAPHICS eingeschaltet ist. Dieser Konflikt kann jedoch durch die Einfügung der Anweisung „ODS GRAPHICS OFF;“ vor der Grafikausgabe gelöst werden.

Zum Beispiel wird im Makro PM\_DESCRIPTION.MAC.SAS die Variablen deskriptiv untersucht und unter Anderem für stetige Variablen ein Histogramm erstellt.

Fehlerbeschreibung:

Das Makro gibt Histogramme aller stetigen Variablen aus. Durch Makrovariablen kann das Format und der Speicherort gewählt werden. Diese Angaben werden jedoch ignoriert und alle Graphiken werden als PNG im Arbeitsordner gespeichert.

Fehlerbehebung:

Das Format und der Speicherort der Histogramme werden in GOPTIONS eingestellt. Diese Angaben und andere Formatierungen der Graphik scheinen allerdings ignoriert zu werden. Das Problem wird gelöst, in dem die ODS Graphikausgabe kurzzeitig unterbrochen wird. Dazu wird in Zeile 1760 vor den GOPTIONS das Statement

```
ODS GRAPHICS OFF;
```

eingefügt. Nach der Ausgabe des Histogramms über die PROC UNIVARIATE wird diese Einstellungen durch Einfügen von

```
ODS GRAPHICS;
```

in Zeile 1771 wieder aufgehoben.

Zudem neu seit SAS Version 9.1 ist die Ausgabe von Plots in PROC LOGISTIC. Diese Funktion wird in den SAS Makros jedoch absichtlich unterdrückt, um den bisherigen Aufbau der Makros beizubehalten. So gibt es für die Erstellung von ROC-Kurven auch weiterhin ein eigenes Makro, in dem aus den gespeicherten Daten der Kurve eine Grafik erzeugt und gespeichert wird, um z.B. Konfidenzintervallbänder angeben zu können, die in der Standardausgabe nicht enthalten sind.

Fehlerbeschreibung Graphiken:

Das Makro erzeugt eine Graphik der ROC-Kurve des finalen Modells im Arbeitsordner, obwohl dies laut der Makro-Beschreibung nicht vorgesehen ist.

Fehlerbehebung:

Bei der Programmierung der Makros war es mit der damaligen SAS Version noch nicht möglich mit Hilfe der PROC LOGISTIC eine ROC-Kurve zu erstellen. Deshalb wurde dies anders gelöst und in ein eigenes Makro, PM\_ROC.MAC.SAS, ausgelagert. Durch die Änderungen in SAS Version 9.4 wird dieses separate Makro obsolet und könnte in das Makro PM\_LOGREG.MAC.SAS integriert werden. Dies wäre allerdings mit einem nicht geringen Programmieraufwand und anschließenden Validierungstests verbunden. Deshalb soll die bisherige Aufteilung der Makros erhalten bleiben und die Ausgabe der ROC-Kurve in PM\_LOGREG.MAC.SAS unterdrückt werden. Dies geschieht in den Zeilen 399 und 892 im Hilfsmakro LOGREG2.MAC.SAS durch die Option PLOTS=NONE im Aufruf von PROC LOGISTIC.

**4.2 BINWIDTH-Option in PROC LOGISTIC**

Bei der Ausgabe von PROC LOGISTIC gab es geringe numerische Abweichungen. Als Ursache stellte sich die geänderte Standardbelegung der Option BINWIDTH im MODEL-Statement der Prozedur heraus. Dieser Parameter gibt die Intervalllänge an, in der die berechneten Wahrscheinlichkeiten bei der Bestimmung der konkordanten, diskordanten und gebundenen Paare zusammengefasst werden. Je größer der angegebene Wert ist, desto mehr gebundene Paare gibt es.

Fehlerbeschreibung numerische Abweichungen:

Im Abschnitt "Association of Predicted Probabilities and Observed Responses" der Ausgabe gibt es numerische Abweichungen (siehe Abbildung 3). Bei den Prozentsätzen der konkordanten, diskordanten und gebundenen Paare sind diese jeweils geringer als ein Prozentpunkt. Bei den Assoziationsmaßen Somers' D und Gamma liegen die Abweichungen im Bereich von  $10^{-3}$ .

Fehlerbehebung:

Zunächst wurde eine genauere Zahlendarstellung durch die Umstellung von SAS 32bit auf 64bit als Grund der numerischen Abweichungen vermutet. Dies konnte durch einen Test mit SAS Version 9.3 32bit und einem anschließendem Vergleich der Ergebnisse jedoch ausgeschlossen werden.

Eine solche numerische Abweichung wurde auch in einem Forenbeitrag der SAS Support Community [6] diskutiert. Als Grund wurde hierbei eine geänderte Standardbelegung der Option BINWIDTH im MODEL-Statement der Prozedur PROC LOGISTIC angegeben. Dieser Parameter gibt die Intervalllänge an, in der die berechneten Wahrscheinlichkeiten bei der Bestimmung der konkordanten, diskordanten und gebundenen Paare zusammengefasst werden. Je größer der angegebene Wert ist, desto mehr gebundene Paare gibt es. Bisher verwendete SAS aus Performanzgründen eine BINWIDTH von 0.002. Ab SAS Version 9.4 wird hingegen bei einem binären Zielereignis und weniger als 5.000.000 Beobachtungen standardmäßig ohne Binning, also mit einem Wert von 0, gerechnet. Dies ist laut einer SAS Usage Note insbesondere der Fall, wenn ein ROC Statement angegeben ist, die Daten der ROC-Kurve gespeichert werden oder eine ROC-Kurve ausgegeben wird. In diesen Fällen werden alle Angaben der Option

BINWIDTH ignoriert und es wird kein Binning verwendet. Die genannten Ausnahmefälle treten im Makro PM\_LOGREG.MAC.SAS und im Hilfsmakro LOGREG2.MAC.SAS jedoch nicht auf.

Die Berechnung ohne Binning ist genauer als mit und sollte damit auch weiterhin erhalten bleiben. Gleichzeitig soll aber auch ein Vergleich der Ausgabe mit der aus SAS Version 9.1 unter den gleichen Voraussetzungen möglich sein. Dies betrifft auch alle anderen Makros, die das Makro PM\_LOGREG.MAC.SAS und dessen Ergebnisse direkt verwenden. Daher wird in den Hauptmakros:

PM\_LOGREG.MAC.SAS, PM\_ROC.MAC.SAS,  
 PM\_CROSSVALIDATION.MAC.SAS,  
 PM\_BOOTSTRAP\_VALIDATION.MAC.SAS,  
 PM\_EXTERNAL\_VALIDATION.MAC.SAS,  
 PM\_SHRINKAGE\_VALIDATION.MAC.SAS

und den Hilfsmakros

LOGREG2.MAC.SAS, BOOT\_MODEL\_VALIDATION.MAC.SAS

BOOT\_MEAN\_VALIDATION.MAC.SAS

eine neue Makrovariable BINWIDTH eingefügt. Zudem wird die neue Fehlermeldung error66.sas erstellt, die bei einer Eingabe der BINWIDTH kleiner 0 oder größer gleich 1 ausgegeben wird.

| Ausgabe in Version 9.1 (Binwidth = 0.002)                     |       |           |       | Ausgabe in Version 9.4 (Binwidth = 0)                         |       |           |       |
|---|-------|-----------|-------|---|-------|-----------|-------|
| Association of Predicted Probabilities and Observed Responses |       |           |       | Association of Predicted Probabilities and Observed Responses |       |           |       |
| Percent Concordant  | 87.8  | Somers' D | 0.757 | Percent Concordant  | 87.9  | Somers' D | 0.758 |
| Percent Discordant  | 12.0  | Gamma     | 0.759 | Percent Discordant  | 12.1  | Gamma     | 0.758 |
| Percent Tied  | 0.2   | Tau-a     | 0.194 | Percent Tied  | 0.0   | Tau-a     | 0.194 |
| Pairs   | 64521 | c         | 0.879 | Pairs   | 64521 | c         | 0.879 |

**Abbildung 3:** Vergleich der Ausgaben in PROC LOGISTIC mit dem alten und neuen Default für BINWIDTH in Version 9.1 und 9.4

### 4.3 Hosmer-Lemeshow-Test für Goodness-of-Fit in PROC LOGISTIC

Für die Anpassungsgüte (Goodness of Fit) wird der Zusammenhang zwischen den beobachteten Werten  $y$  und den durch das Modell geschätzten Werten  $\hat{y}$  untersucht. Auch hier kommt wieder die PROC LOGISTIC zum Einsatz. In dieser Prozedur kann der Hosmer-Lemeshow-Test zur Überprüfung der Anpassungsgüte angefordert werden.

#### Fehlerbeschreibung:

Beim Hosmer and Lemeshow Goodness-of-Fit Test gibt es Unterschiede in der Klasseinteilung und damit numerische Abweichungen bei  $\chi^2$  und dem dazugehörigen p-Wert.

Fehlerbehebung:

Die 710 Beobachtungen wurden bisher auf zehn Klassen mit jeweils 69 bis 73 Beobachtungen aufgeteilt. In SAS Version 9.4 ist die Aufteilung allerdings gleichmäßig mit 71 Beobachtungen pro Klasse. Laut dem SAS User's Guide [7] werden bei der Klasseneinteilung Gruppen von Beobachtungen mit der gleichen geschätzten Wahrscheinlichkeit nicht getrennt, was zu unterschiedlichen Klassengrößen führen kann. Es besteht daher die Vermutung, dass auch hier eine unterschiedliche BINWIDTH die Ursache der Abweichungen ist. Leider wird jede Angabe eines Wertes für die BINWIDTH ignoriert, wenn die Option LACKFIT zur Berechnung des Hosmer and Lemeshow Goodness-of-Fit Tests angegeben wird. Die Vermutung lässt sich daher nicht bestätigen.

Da die Berechnung mit einer gleichmäßigen Klasseneinteilung für die Analyse genauer ist, bleibt das Makro PM\_GOF.MAC.SAS unverändert.

**4.4 Stichprobenziehung für Bootstrap-Validierung**

Das Prinzip des Ziehens mit Zurücklegen kommt bei der Bootstrap-Validierung zum Einsatz. Die SAS Makros verwendeten zur Erzeugen von Bootstrap-Samples bisher einen DATA-Step mit SELECT-Statement. Hierbei stellte sich jedoch heraus, dass die Seed-Option nicht vollständig funktioniert. Ob dies durch eine fehlerhafte Implementierung oder den Versionswechsel von SAS Version 9.1 zu Version 9.4 verursacht wurde, ist nicht gewiss. Um das Problem zu lösen wird stattdessen nun die Prozedur PROC SURVEYSELECT verwendet.

Fehlerbeschreibung:

Um bei mehrmaliger Durchführung der Bootstrap-Validierung die gleichen Ergebnisse zu erhalten, kann beim Aufruf des Makros eine positive Zahl als Seed übergeben werden. Dadurch erhält man allerdings auch immer die gleichen Bootstrap-Samples und dadurch nur einen Satz Bootstrap-Schätzer. Werden für die Validierung beispielsweise 200 Samples verlangt, so erhält man 200 identische Samples. Eine sinnvolle Schätzung der Parameter ist damit nicht möglich.

Fehlerbehebung:

Bisher wird in einer Schleife jeweils ein neues Sample mit dem angegebenen Seed erzeugt und ausgewertet. Mit Hilfe der PROC SURVEYSELECT außerhalb dieser Schleife werden nun stattdessen alle Samples auf einmal erzeugt und dann in Teildatensätze aufgespalten.

Der zugehörige Code in den Hilfsmakros BOOT\_MEAN\_VALIDATION und BOOT\_MODEL\_VALIDATION jeweils ab Zeile 214 sieht wie folgt aus (siehe Abbildung 4):

```
/* Gesamten Bootstrap-Datensatz erzeugen */
PROC SURVEYSELECT DATA=&boot OUT=boot_all
    seed=&random method=urs samprate=1
    outhits rep=&anzahl_samples NOPRINT;
RUN;

%DO km=1 %TO &anzahl_samples;
    DATA boot&km (DROP=Replicate);
    SET boot_all (WHERE=(Replicate=&km));
    RUN;
/* Auswertung des aktuellen Bootstrap-
Datensatzes */
%END;
```

Abbildung 4: SAS-Code für die Erzeugung der Bootstrap-Samples

## 5 Zusammenfassung

Das Makropaket wurde in SAS Version 8 geschrieben und getestet. Im Jahr 2005 wurde es zudem in SAS Version 9.1 getestet und gegebenenfalls korrigiert. Die jetzige Aufgabe war nun, dieses auf SAS Version 9.4 zu testen und eventuelle Fehler zu korrigieren. Einige Fehler waren sofort gefunden, andere wiederum waren etwas kniffliger zu entdecken und zu beheben. Das Testen der Makros bezog sich hier „nur“ auf die Hauptmakros, die vom Benutzer selbst aufgerufen werden.

Die Hilfsmakros, auf die diese Makros zugreifen, kommen teilweise von externen Stellen, Autoren, anderen Universitäten etc. Sie sind zum Teil noch in SAS Version 6 geschrieben und stellen daher die Frage, ob die Kompatibilität zu SAS Version 9.4 gewährleistet werden kann. In unseren Tests sind hier allerdings keine Versionskonflikte aufgefallen.

Insgesamt sind die Unstimmigkeiten der Ergebnisse aus Version 9.1 und 9.4 nachvollziehbar, wenn man sie kennt. Einführung von ODS und nachfolgenden Default-Einstellungen (z.B. HTML-Output) sowie der Einbau von Grafikausgaben in Auswertungsprozeduren machen die Nutzung von SAS in der neuesten Version um einiges leichter. Standardisierte Makros sind dann aber nicht unbedingt mehr lauffähig oder laufen anders, als vorab definiert. Eine 100%ige Rückwärtskompatibilität ist nicht gegeben, wäre aber schön. So ist doch einiges an Aufwand zu betreiben, um die Lauffähigkeit in der neuen Version wieder herzustellen, dass die Standardabläufe und Standardergebnisse zu erzielen sind.

Eine genauere Dokumentation der Überprüfung [4] kann von den Autoren angefordert werden und ist neben dem SAS-Makropaket von der Instituts-Internetseite [5] herunterladbar.

**Literatur**

- [1] R. Muche, C. Ring, C. Ziegler: Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression. Shaker Verlag, Aachen, 2005
- [2] R. Muche, C. Ring, C. Ziegler: Ein SAS-Makro Paket für die Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression. In: E. Rödel, R.-H. Bödeker (Hrsg.): SAS-Verbindung von Theorie und Praxis. Proceedings der 9. Konferenz der SAS-Anwender in Forschung und Entwicklung 2005 in Berlin. Shaker Verlag, Aachen (2005), S. 313-321
- [3] R. Muche, F. Ruthenberg: SAS V9 kompatibel zu V8? Ein Fallbeispiel aus PROC LOGISTIC. In: K. Kaiser, R.-H. Bödeker (Hrsg.): Statistik und Datenanalyse mit SAS. Proceedings der 10. Konferenz der SAS-Anwender in Forschung und Entwicklung 2006 in Hamburg. Shaker Verlag, Aachen (2006), S. 167-170
- [4] S. Müller: SAS-Makros zur Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression – Getestet in SAS Version 9.4. Interner Institutsbericht, Institut für Epidemiologie und Med. Biometrie, Universität Ulm (2016)
- [5] R. Muche: Internetseite für Download der Makros und Begleitdokumente:  
<http://www.uni-ulm.de/med/epidemiologie-biometrie/forschung/methodik/sas-makros-fuer-prognosemodellierung.html>
- [6] SAS Knowledge Base: Usage Note 45767: Computing the statistics in "Association of Predicted Probabilities and Observed Responses" table  
<http://support.sas.com/kb/45/767.html>
- [7] SAS/STAT(R) 14.1 User's Guide: The LOGISTIC Procedure - The Hosmer-Lemeshow Goodness-of-Fit Test  
[http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug\\_logistic\\_details32.htm](http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_logistic_details32.htm)