

## Logistische Regression mit Messwiederholungen: Anwendung von PROC GENMOD in SAS

Birgit Hay  
Institut für Biometrie,  
Universität Ulm  
Schwabstraße 13  
89070 Ulm  
Birgit.Hay@uni-ulm.de

Silvia Sander  
Institut für Biometrie,  
Universität Ulm  
Schwabstraße 13  
89070 Ulm  
Silvia.Sander@uni-ulm.de

Manfred Weiß  
Klinik für Anästhesiologie,  
Universitätsklinik Ulm  
Steinhövelstraße 9  
89075 Ulm  
Manfred.Weiss@uniklinik-  
ulm.de

Martina Kron  
Institut für Biometrie,  
Universität Ulm  
Schwabstraße 13  
89070 Ulm  
Martina.Kron@uni-ulm.de

### Zusammenfassung

Bei schwerstkranken Patienten der Anästhesiologischen Intensivstation kommt es oft zu extremen Blutzuckerentgleisungen.

Wie häufig kommt es zu Entgleisungen und welche Faktoren sind Risikofaktoren? Anhand eines Datensatzes von 804 Patienten mit 7409 Verlaufsbeobachtungen sollen Risikofaktoren identifiziert werden. Als mögliche Risikofaktoren gibt es sowohl einmalig erhobene Parameter (z.B. Alter, Geschlecht) als auch täglich untersuchte Parameter (z.B. Therapien). Die Zielgrößen Hypoglykämie ( $<80\text{mg/dl}$  vs. normal) und Hyperglykämie ( $\geq 150\text{mg/dl}$  vs. normal) werden anhand der Blutzuckerwerte täglich ermittelt.

Die Identifikation wichtiger Risikofaktoren erfolgt für jede Zielgröße durch multiple logistische Regressionsmodelle mit Messwiederholungen in SAS 9.2. Rohe und adjustierte Odds Ratios mit Konfidenzintervallen können für Verlaufsdaten in SAS 9.2 mit der Prozedur GENMOD ausgewertet werden. Jedoch ist die Rückwärtselimination von Einflussgrößen in PROC GENMOD nicht implementiert und muss schrittweise „von Hand“ durchgeführt werden.

**Schlüsselwörter:** multiple logistische Regression, Messwiederholungen, PROC GENMOD

## 1 Motivation

Bei schwerstkranken Patienten der Anästhesiologischen Intensivstation kommt es oft zu extremen Blutzuckerentgleisungen. Da diese Entgleisungen die Organfunktionen und somit auch die Mortalität beeinflussen, erhalten die Patienten prophylaktisch eine inten-

sivierte Insulintherapie (IIT). Trotz dieser IIT lässt sich der Blutzuckerwert häufig nicht im Normbereich zwischen 80 mg/dl und 150 mg/dl halten.

Fragen:

Wie häufig gibt es Hypoglykämien ( $BZ < 80\text{mg/dl}$ )?

→ welche Risikofaktoren gibt es für Hypoglykämie?

Wie häufig gibt es Hyperglykämien ( $BZ \geq 150\text{mg/dl}$ )?

→ welche Risikofaktoren gibt es für Hyperglykämie?

Wie lässt sich dies in SAS lösen?

## 2 Durchführung und Ergebnisse

Die Daten erwachsener Patienten der Jahre 2007 und 2008, die mindestens drei Tage auf der Intensivstation lagen, werden analysiert.

→ Es liegen 7409 Verlaufsbeobachtungen von 804 Patienten vor.

Zur Ermittlung der beiden **Zielgrößen** werden die täglich mehrfach gemessenen Blutzuckerwerte (in mg/dl) jeden Tag sowohl in **Hypoglykämie** (ja vs. normal) als auch in **Hyperglykämie** (ja vs. normal) dichotomisiert.

Als **Einflussgrößen** sollen Variablen betrachtet werden, die sowohl **einmalig** (z.B. Alter (J), Geschlecht, Aufnahmeparameter (z.B. OP-Art)) als auch **täglich** (z.B. IIT, Therapie (z.B. Steroid), Status (z.B. Sepsis)) erhoben werden.

Die Identifikation wichtiger Risikofaktoren erfolgt für beide Zielgrößen durch logistische Regressionsmodelle mit Messwiederholungen [1]. Dabei wird der Einfluss einmalig und täglich erhobener Parameter auf die Zielgrößen mit der Prozedur GENMOD [2] berechnet.

### 2.1 Rohe Odds Ratios

Zuerst werden für jede Einflussvariable rohe Odds Ratios und 95%-Konfidenzintervalle berechnet. Um diese Odds Ratios mit den dazugehörigen Konfidenzintervallen zu erhalten, ist die Angabe des ESTIMATE Statements notwendig. Das REPEATED Statement ist erforderlich, da für jeden Patienten die Zielgröße und teilweise die Einflussgrößen täglich erhoben werden.

Programmbeispiel für eine Einflussgröße (Zielgröße Hypoglykämie)

```
TITLE 'Hypoglykämie (BZ<80 vs normal) ' ;
PROC GENMOD ;
  CLASS patnr
        iit (PARAM=ref REF='nein') ;
  MODEL hypo = iit
        / DIST=BIN LINK=LOGIT ;
  REPEATED SUBJECT=patnr / TYPE=CS ;
  ESTIMATE 'IIT' iit 1 / EXP ;
RUN ;
```

## SAS-Output Beispiel für eine Einflussgröße (Zielgröße Hypoglykämie)

Hypoglykämie (BZ&lt;80 vs normal)

Analysis Of GEE Parameter Estimates  
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept	-2.1035	0.0992	-2.2979	-1.9090	-21.20	<.0001
IIT ja	0.8717	0.1145	0.6474	1.0960	7.62	<.0001

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits	L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits	Chi-Square	Pr > ChiSq
IIT	0.7051	0.6564 0.7495	0.8717	0.1145	0.05	0.6474 1.0960	58.00	<.0001
Exp(IIT)			<u>2.3910</u>	0.2737	0.05	<u>1.9105</u> <u>2.9922</u>		

## 2.2 Rückwärtselimination

Anschließend werden alle potentiellen Einflussgrößen in einem Gesamtmodell simultan betrachtet, um dann mittels Rückwärtselimination schrittweise wichtige Parameter ( $p \leq 0.05$ ) zu identifizieren. Um den globalen p-Wert für die vierstufige Einflussgröße OP-Art zu erhalten, ist das CONTRAST Statement erforderlich.

### Programmbeispiel für Gesamtmodell (Zielgröße Hypoglykämie)

```
TITLE 'Hypoglykämie (BZ<80 vs normal)          GESAMTMODELL' ;
PROC GENMOD ;
  CLASS patnr
        iit      (PARAM=ref REF='nein')
        opart    (PARAM=ref REF='neuroch')
        steroid  (PARAM=ref REF='nein')
        adren    (PARAM=ref REF='nein')
        noradr   (PARAM=ref REF='nein')
        geschl   (PARAM=ref REF='m')
        sepsis   (PARAM=ref REF='nein') ;
  MODEL hypo = iit opart steroid adren noradr geschl sepsis alter
            / DIST=BIN LINK=LOGIT ;
  REPEATED SUBJECT=patnr / TYPE=CS ;
  CONTRAST 'OP-Art' opart 1 0 0 ,
            opart 0 1 0 ,
            opart 0 0 1 ;
RUN ;
```

SAS-Output Beispiel für Gesamtmodell (Zielgröße Hypoglykämie)

Hypoglykämie (BZ<80 vs normal) GESAMTMODELL

Analysis Of GEE Parameter Estimates  
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept		-3.2733	0.2747	-3.8117	-2.7348	-11.91	<.0001
iit	ja	0.7249	0.1225	0.4848	0.9651	5.92	<.0001
opart	Ba	-0.4324	0.1739	-0.7732	-0.0915	-2.49	0.0129
opart	Ge	-0.1929	0.1792	-0.5442	0.1583	-1.08	0.2817
opart	Tr	-0.2166	0.1861	-0.5812	0.1481	-1.16	0.2444
steroid	ja	0.2813	0.1350	0.0167	0.5459	2.08	0.0372
adren	ja	0.3309	0.2014	-0.0638	0.7257	1.64	0.1004
noradr	ja	0.3530	0.1345	0.0892	0.6167	2.62	0.0087
geschl	w	0.2233	0.1345	-0.0404	0.4869	1.66	0.0970
sepsis	ja	-0.0748	0.1138	-0.2979	0.1482	-0.66	<u>0.5109</u>
alter		0.0198	0.0038	0.0124	0.0272	5.23	<.0001

Contrast Results for GEE Analysis

Contrast	DF	Chi-Square	Pr > ChiSq	Type
OP-Art	3	6.16	0.1039	Score

Im ersten Schritt wird aus dem Ergebnis des Gesamtmodells die Variable mit dem größten p-Wert eliminiert und ein neues Modell (im Bsp. ohne ‚sepsis‘) gerechnet. Dieses Vorgehen wird sooft wiederholt, bis nur noch Variablen mit p-Werten  $\leq 0.05$  im Modell sind.

### 2.3 Endmodell

Sind alle Einflussgrößen mit p-Werten  $> 0.05$  eliminiert, ergibt sich das Endmodell. Wichtige Prädiktoren für das Auftreten einer **Hypoglykämie** sind die Merkmale Alter bei Aufnahme sowie die Insulin-, Noradrenalin- und Steroidgabe am Untersuchungstag.

#### Programmbeispiel für Endmodell (Zielgröße Hypoglykämie)

```
TITLE 'Hypoglykämie (BZ<80 vs normal)          ENDMODELL' ;
PROC GENMOD ;
  CLASS patnr
        iit      (PARAM=ref ref='nein')
        steroid (PARAM=ref ref='nein')
        noradr   (PARAM=ref ref='nein') ;
  MODEL hypo = iit steroid noradr alter
              / DIST=BIN LINK=LOGIT ;
  REPEATED SUBJECT=patnr / TYPE=CS ;
  ESTIMATE 'IIT'      iit 1 / EXP ;
  ESTIMATE 'Steroid' steroid 1 / EXP ;
  ESTIMATE 'Noradr.' noradr 1 / EXP ;
  ESTIMATE 'Alter'   alter 1 / EXP ;
RUN ;
```

## SAS-Output Beispiel für Endmodell (Zielgröße Hypoglykämie)

Hypoglykämie (BZ&lt;80 vs normal) ENDMODELL

Analysis Of GEE Parameter Estimates  
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept	-3.3540	0.2487	-3.8414	-2.8667	-13.49	<.0001
IIT ja	0.7410	0.1176	0.5105	0.9714	6.30	<.0001
Steroid ja	0.2615	0.1315	0.0038	0.5192	1.99	0.0467
Noradr ja	0.3645	0.1015	0.1655	0.5635	3.59	0.0003
Alter	0.0188	0.0036	0.0117	0.0259	5.19	<.0001

## Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits	L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits	Chi-Square	Pr > ChiSq
IIT	0.6772	0.6249 0.7254	0.7410	0.1176	0.05	0.5105 0.9714	39.71	<.0001
Exp(IIT)			<u>2.0980</u>	0.2467	0.05	<u>1.6661</u> <u>2.6417</u>		
Steroid	0.5650	0.5010 0.6270	0.2615	0.1315	0.05	0.0038 0.5192	3.96	0.0467
Exp(Steroid)			<u>1.2989</u>	0.1708	0.05	<u>1.0038</u> <u>1.6807</u>		
Noradr.	0.5901	0.5413 0.6373	0.3645	0.1015	0.05	0.1655 0.5635	12.89	0.0003
Exp(Noradr.)			<u>1.4398</u>	0.1462	0.05	<u>1.1800</u> <u>1.7568</u>		
Alter	0.5047	0.5029 0.5065	0.0188	0.0036	0.05	0.0117 0.0259	26.90	<.0001
Exp(Alter)			<u>1.0190</u>	0.0037	0.05	<u>1.0118</u> <u>1.0262</u>		

Für die zweite Zielgröße **Hyperglykämie** sind ebenso die Merkmale Alter bei Aufnahme sowie die Insulin-, Noradrenalin- und Steroidgabe wichtig, außerdem sind die Operationsart und das Vorliegen einer Sepsis am Untersuchungstag von Bedeutung.

### 3 Fazit

Rohe und adjustierte Odds Ratios mit Konfidenzintervallen können für Verlaufsdaten in SAS 9.2 mit der Prozedur GENMOD ausgewertet werden. Jedoch ist die Rückwärtselimination in PROC GENMOD nicht implementiert und muss schrittweise „von Hand“ durchgeführt werden.

### Literatur

- [1] Hosmer DW, Lemeshow S: Applied logistic regression. New York, Wiley, 2000
- [2] SAS Institute Inc., Cary, NC, USA, Online-Dokumentation