

## SAS-Makro zur Analyse von disproportionalen Substichproben am Beispiel der ISAAC Phase Two Studie

Andrea Kleiner  
 Institut für Epidemiologie  
 Helmholtzstraße 22  
 89081 Ulm  
 andrea.kleiner@uni-ulm.de

Gisela Büchele  
 Institut für Epidemiologie  
 Helmholtzstraße 22  
 89081 Ulm  
 gisela.buechele@uni-ulm.de

Gudrun Weinmayr  
 Institut für Epidemiologie  
 Helmholtzstraße 22  
 89081 Ulm  
 gudrun.weinmayr@uni-ulm.de

Stephan Weiland  
 Institut für Epidemiologie  
 Helmholtzstraße 22  
 89081 Ulm  
 stephan.weiland@uni-ulm.de

### Zusammenfassung

**Einführung:** Bei großen epidemiologischen Studien, wie der Phase Zwei der ISAAC Studie (International Study of Asthma and Allergies in Childhood), bietet es sich an bei der Erhebung von aufwendigen Daten (Untersuchungsdaten, Blutabnahme) disproportional Substichproben zu verwenden. So haben, in der ISAAC Phase Zwei Studie, einige Studienzentren bei den Haut-Prick-Tests oder der Untersuchung zu Neurodermitis auf diese Substichprobenziehung zurückgegriffen. Aufgrund der Disproportionalität kann eine konventionelle logistische Regression, für die zentrumsspezifische Berechnung der Odds Ratios, zu verzerrten Ergebnissen führen. Empfohlen wird die Anwendung eines gewichteten logistischen Regressions-Modell.

**Methode:** Eine gewichtete logistische Regression war bis SAS Version 9 mit der SUDAAN Software (SAS aufrufbare Version) möglich. Ab SAS Version 9.1 wird hierfür die Prozedur Surveylogistic verwendet. Da sich einzelne Substichproben (wie zum Beispiel die des Haut-Prick-Tests und die der Untersuchung auf Neurodermitis) innerhalb der Zentren unterscheiden können, ist ein Programm notwendig, welches für jedes neue Modell die Gewichte modell- und zentrumsspezifisch berechnen kann. Das hier vorgestellte Makro erfüllt diese Vorgaben und berechnet einfache Modelle mit einer unabhängigen Variablen genauso wie Modelle mit mehreren Adjustierungsvariablen.

**Ergebnisse und Diskussion:** Die Prozedur Surveylogistic und die Gewichtung gewährleisten unverzerrte Ergebnisse. Das Makro ist sehr benutzerfreundlich programmiert, so erhält der Anwender lediglich durch das Angeben der Modellvariablen und der erwünschten Zentren ein Ergebnis, das immer mit den aktuellen modell- und zentrumsspezifischen Gewichten berechnet wird. Odds Ratios und Beta-Schätzer werden in eine Excel-Datei ausgegeben, was eine einfache Weiterverarbeitung, z.B. Erstellung einer Präsentationsgraphik, ermöglicht.

**Schlüsselwörter:** SAS Makro, PROC SURVEYLOGISTIC, gewichtete logistische Regression

## 1 Einführung

Einige Studienzentren der ISAAC Phase Zwei Studie (International Study of Asthma and Allergies in Childhood) verwendeten für die Datenerfassung von aufwendigen Untersuchungen (z.B.: Haut-Prick-Test, Neurodermitis-Diagnose) keine vollständigen Stichproben sondern disproportional gezogene Substichproben. Laut Studienprotokoll bestehen diese aus 100 Teilnehmer mit und 100 Teilnehmer ohne keuchende Atmung. Die tatsächlichen Zahlen weichen jedoch, je nach Teilnahme, von diesen Vorgaben ab. So sind die einzelnen Substichproben von Zentrum zu Zentrum, und von Modul zu Modul verschieden.

### 1.1 Disproportionale Substichproben

Beispiel:

In einer Studienpopulation mit 1677 Teilnehmern befinden sich 142 Teilnehmer mit keuchender Atmung und 1535 ohne keuchende Atmung. Das entspricht einem Verhältnis von 8% zu 92%. Entnimmt man hieraus, für die Substichprobe, jeweils 100 Teilnehmer mit und 100 Teilnehmer ohne keuchende Atmung, zieht man somit aus der Gesamtstichprobe 70% mit keuchender Atmung ( $100/142$ ) und 7% ohne keuchende Atmung ( $100/1535$ ).

Werden bei der Odds Ratio Berechnung, mit einer konventionellen logistischen Regression, Variablen aus solchen Substichproben zur Bildung des Modells verwendet, kann dies zu einer Verzerrung der Ergebnisse führen. Das gilt immer dann, wenn die Variable „keuchende Atmung“ nicht als abhängige Variable im Modell verwendet wird. Um eine solche Verzerrung zu vermeiden ist es notwendig die Schätzer der Substichproben, auf die Anzahl der Kinder mit und ohne keuchende Atmung in der vollen Stichprobe, hochzugewichten.

### 1.2 Umgang mit disproportionalen Substichproben in der ISAAC Studie

Um Verzerrungen der Ergebnisse zu vermeiden, werden die Odds Ratios in der ISAAC Phase Zwei Studie mit einer gewichteten logistischen Regression berechnet. Bis SAS Version 9 ist eine solche Regression mit der SUDAAN Software (SAS aufrufbare Version) und der Prozedur RLogist berechnet worden. Ab SAS Version 9.1 wird hierfür die Prozedur Surveylogistic verwendet. Da sich die Zusammensetzung der Substichproben von Zentrum zu Zentrum und von Modul zu Modul unterscheidet, ist es besser die Prozedur in ein Makro einzubinden. Diese Programmierweise ermöglicht eine relativ einfache Umsetzung des Zieles für jedes Modell die Gewichte modell- und zentrumsspezifisch, automatisch bei jedem Aufruf, neu zu berechnen.

## 2 Methode

### 2.1 Grundstruktur des Makros

Der Makroaufruf ist für den Anwender relativ einfach gehalten. Nach dem Einlesen des Datensatzes, in dem bereits wichtige Informationen für die Gewichts-Berechnung dauerhaft gespeichert sind, müssen beim eigentlichen Aufruf lediglich die Modellvariablen angegeben werden und die Zentren, für die das Modell berechnet werden soll. Im Makro selbst werden einige Hilfsvariablen gebildet, die für die zentrums- und modell-spezifische Gewichts-bildung relevant sind. Außerdem laufen einige DATA Schritte ab, um eine Excel-Tabelle mit den Ergebnissen zu erhalten. Des Weiteren werden zwei weitere Makros aufgerufen, einmal für die Ermittlung der Kennzahlen für die Gewichte, und einmal zur Berechnung der Odds Ratios mit der Prozedur Surveylogistic. Diese laufen für jedes Zentrum separat ab.

### 2.2 Datensatzstruktur, Aufruf des Makros, wichtige Hilfsvariablen

Vor dem eigentlichen Makro-Aufruf werden, anhand der %INCLUDE Anweisung, alle benötigten Makros angelegt und der Datensatz temporär eingelesen.

Im Datensatz beginnen die Variablen eines Moduls jeweils mit denselben zwei Buchstaben, somit kann erkannt werden, welche Variable zu welchem Modul gehört. Beispiele hierfür:

- WH: Fragen zu Asthma, pfeifende Atemgeräusche und Atemnot
- SP: Variablen für die Haut-Prick-Test Untersuchung
- ED: Variablen für die Untersuchung auf Neurodermitis
- DC: Variablen zu den demographischen Kenndaten

Jedes Modul besitzt eine Modul-Info-Variable, woraus man erkennen kann in welchen Zentren, in diesem Modul, eine volle Stichprobe vorliegt und bei welchen eine disproportionale Substichprobe:

- SP\_S=0: Zentrum besitzt im Modul Haut-Prick-Test eine volle Stichprobe.
- SP\_S=1: Zentrum besitzt im Modul Haut-Prick-Test eine disproportionale Substichprobe.

Eine weitere wichtige Information ist die Anzahl der Teilnehmer, pro Zentrum, mit und ohne keuchende Atmung in der vollen Stichprobe:

- WHyf: Anzahl Teilnehmer mit keuchender Atmung in der vollen Stichprobe
- WHnf: Anzahl Teilnehmer ohne keuchende Atmung in der vollen Stichprobe
- WHynf: Die Summe von WHyf und WHnf, also Gesamtzahl der Teilnehmer pro Zentrum

Während des Einlesens des Datensatzes wird die Stratumvariable gebildet, die für Teilnehmer mit keuchender Atmung den Wert „1“ annimmt und für Teilnehmer ohne keuchende Atmung den Wert „2“:

```
strata=wh02x;
if strata=0 then strata=2;
```

Im Makro-Aufruf werden die Modellvariablen angegeben und die Zentren, für die das Modell berechnet werden soll:

```
%LRS1
(ed01x
, sp09allx
, E:\temp\
, 11, 12, 13, 14, 15, 16, 17, 23, 24,
25, 26, 27, 28, 29, 30, 32, 33, 34, 35,
36, 37, 38, 39, 40, 42, 44, 45, 46, 48, 49
, i=30, model=1);
```

Parameter-Definition:

```
%macro LRS1
(DV, IV, path, centre1, centre2, centre3, centre4, centre5, centre6, centre7,
centre8, centre9, centre10, centre11, centre12, centre13, centre14,
centre15, centre16, centre17, centre18, centre19, centre20, centre21,
centre22, centre23, centre24, centre25, centre26, centre27, centre28,
centre29, centre30, i=1, model=1);
```

In der ersten Zeile befindet sich die Angabe der abhängigen Variablen (ED01x: Neurodermitis ja/nein); die zweiten Zeile enthält Angaben zur unabhängigen Variable (SP09allx: positiver Haut-Prick-Test ja/nein). Eventuelle Adjustierungsvariablen (z.B. Alter, Geschlecht) werden, durch Leerzeichen getrennt, hinten angehängt. Mit der Pfadangabe erfolgt die Definition des Zielordners für die Exceldatei, in der die Ergebnisse gespeichert werden. Danach werden die Nummern der Zentren aufgezählt, für die die Odds Ratios berechnet werden sollen. In der letzten Zeile sind wichtige Informationen für den Makorablauf angegeben: Anzahl der Zentren, für die das Modell berechnet werden soll ( $i=30$ ) und Anzahl der unabhängigen Variablen, die sich im Modell befinden ( $model=1$ ).

Zu Beginn des LRS1 Makros werden einige Hilfsvariablen gebildet:

```
%let n=%eval(&model+1);
%let var&n = &DV;
%let mm&n = %qsubstr(&&dv, 1, 2);

%local j;
%local k;
%do j=1 %to %eval(&model+1);
    %let var&j = %scan(&iv, &j);

    %do k=1 %to &model;
        %let mm&k=%qsubstr(&&var&k, 1, 2);
    %end;
```

```

%let temp=%quote(&var&&j);
%if &j=1 %then %let freq=&temp;
%if &j > 1 and &j <= &model %then %let
freq=&freq*&temp;

%if &j=1 %then %let hilf1=&temp%str(=.M);
%if &j > 1 and &j <= &model %then %let
hilf1=&hilf1 %str(OR) &temp%str(=.M);

%let info=%quote(&mm&&j%str(_S));

%if &j=1 %then %let keep=&info;
%if &j > 1 and &j <= &model %then %let keep=&keep &info;
%if &j=1 %then %let summe=&info;
%if &j > 1 and &j <= &model %then %let summe=&summe ,&info;

%end;

```

Für unser Beispiel-Modell ergeben sich folgende Hilfsvariablen:

- Var2: ED01x
- Mm2: ED
- Var1: SP09allx
- Mm1: SP
- Freq: sp09allx
- Hilf1: 'sp09allx=.M'
- Keep: sp\_s
- Summe: sp\_s

Adjustiert man das Modell, zum Beispiel nach Geschlecht (DC01), sehen die Hilfsvariablen wie folgt aus:

- Var3:ED01x
- Mm3: ED
- Var1: sp09allx
- Var2: dc01
- Mm1: SP
- Mm2: DC
- Freq: sp09allx\*dc01
- Hilf1: ',sp09allx=.M or DC01=.M'
- Keep sp\_s dc\_s
- Summe: sp\_s, dc\_s

## 2.3 Berechnung der Gewichte

Wichtige Kennzahlen für die Berechnung der Gewichte sind die Anzahl der Teilnehmer mit keuchender Atmung und ohne keuchende Atmung innerhalb des Modells. Diese werden für jedes Zentrum, mittels des Makros %freq, einzeln ermittelt:

```

Proc Sort data=wtwork;by strata;run;
%Do z=1 %TO &i ;
%LET cc=&centre&&z;

```

```
%put ----- &cc -----;  
%freq(&cc);  
%END;Title;
```

Für  $z=1$  wird der Variable `cc` der Wert für das erste Zentrum zugeordnet, d.h. das Makro berechnet die Kennzahlen für das Zentrum 11.

Im Makro `%freq` wird zuerst die Hilfsvariable `freqvar` gebildet:

```
%let freqvar=&&freq %str(*) &DV;
```

Für unser Beispiel-Modell erhält die Variable folgenden Wert: `sp09allx*ED01x`. Dieser findet Verwendung in der nachfolgenden Proc Freq Anweisung:

```
Proc Freq data=wtwork;Ods output List=freq;  
tables strata*&freqvar /list ;  
Title "Frequency List for Centre: &centre";  
where centre=&centre;run;quit;  
ods output close;
```

Die Anzahl der Teilnehmer mit keuchender Atmung (`why_model`) und ohne keuchende Atmung (`whn_model`), sowie die Gesamtzahl der Teilnehmer im Modell (`cumall`), kann mit folgenden Anweisungen ermittelt werden:

```
Data freq2;set freq end=lastobs;  
keep centre strata cumfrequency why;;  
where strata=1;  
IF lastobs then why_model=cumfrequency;  
run;
```

```
Data freq3;set freq end=lastobs;  
keep centre strata cumfrequency cumall;  
where strata=2;  
IF lastobs then cumall=cumfrequency;  
run;
```

```
Data freq&centre;  
merge freq2 freq3;  
by centre;  
keep centre why_model whn_model cumall dropout;  
whn_model=cumall-why_model;  
IF why_model=. or whn_model=. then delete;  
run;
```

Am Ende des Makros werden die Modellinformationen der einzelnen Zentren im Datensatz `modelfreq` zusammengefasst. Die eigentliche Gewichtsrechnung erfolgt im Makro `%LRS1`:

```
IF strata=1 then wtmin=(1/(why_model/whyf));
```

```
IF strata=2 then wtmin=(1/(whn_model/whnf));
IF strata=.M then wtmin=.;
```

Nun wird die Summe der Modul-Informationsvariablen berechnet. Falls diese, für ein Zentrum, den Wert 0 annimmt, muss nicht hochgewichtet werden, da das Modell bereits aus einer vollen Stichprobe besteht. Das Gewicht wtmin erhält dann den Wert 1 und die Stratumvariable wird ebenfalls auf 1 gesetzt:

```
sample=sum(%unquote (&summe) , %unquote (&&mm&n%str (_S) ));

IF sample eq 0 THEN DO;
    wtmin=1;strata=1;
END;
```

## 2.4 Die Prozedur Surveylogistic

Zur Berechnung der Odds Ratios wird die Prozedur Surveylogistic, im Makro %rlogmin, ebenfalls für jedes Zentrum einzeln aufgerufen:

```
%Do j=1 %TO &i ;
    %LET cc=&centre&&j;
    %put ----- &cc -----;
    %rlogmin(&cc);
%END;
```

Für jedes Zentrum wird ein Datensatz gebildet, auf den die Prozedur Surveylogistic zugeht. Ausgeschlossen werden Teilnehmer, die fehlende Werte in der Stratumvariablen haben.

Im Makro %rlogmin erfolgt der Aufruf der Prozedur Surveylogistic.

```
ods output parameterestimates=betas oddsratios=risk;

proc surveylogistic data=data&centre;
title "OR subsample for centre: &centre";
    model &DV (desc)= &IV ;
    strata strata;
    weight wtmin;
run;
ods output close;
```

Die für jedes Zentrum erstellten Datensätze „betas“, mit den Beta-Schätzern, werden im Datensatz „exportest“ zusammengefasst. Ebenso werden die Datensätze „risk“, sie enthalten die Odds Ratios der einzelnen Zentren, zum Datensatz „export“ zusammengefasst. Nach dem Zusammenspielen von „export“ und „exportest“ mit einem weiteren Datensatz, dieser beinhaltet Zentrumsinformationen, wird der so entstehende Datensatz, mittels Proc export nach Excel exportiert.

### 3 Ergebnisse und Diskussion

#### 3.1 Ergebnisse des Programms

Nachfolgend werden einige Ergebnisse des Makros dargestellt. Im SAS Output-Fenster (Abb. 1) werden die Ergebnisse der Freq Prozedur des Makros %freq ausgegeben, die zur Ermittlung der Kennzahlen für die Gewichte im Modell dient. Der SAS Datensatz modelfreq (Abb. 2) enthält, für jedes Zentrum, Anzahl der Teilnehmer mit und ohne keuchende Atmung im Modell.

Frequency List for Centre: 46						
The FREQ Procedure						
strata	sp09allx	ED01x	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	0	28	16.47	28	16.47
1	0	1	5	2.94	33	19.41
1	1	0	16	9.41	49	28.82
1	1	1	4	2.35	53	31.18
2	0	0	67	39.41	120	70.59
2	0	1	12	7.06	132	77.65
2	1	0	34	20.00	166	97.65
2	1	1	4	2.35	170	100.00

Frequency Missing = 848

**Abbildung 1:** Ergebnis der Prozedur freq, für die Ermittlung der Teilnehmer mit und ohne keuchende Atmung im Modell, für ein Zentrum

	centre	why_model	cumall	whn_model
1	11	111	992	881
2	12	153	1039	886
3	13	86	610	524
4	14	84	968	884
5	15	38	881	843
6	16	54	628	574

**Abbildung 2:** Ausschnitt aus dem Datensatz modelfreq mit den Kennzahlen für die Gewichts Berechnung für jedes Zentrum.

Nachfolgend Ergebnisse der Prozedur Surveylogistic (Abb. 3) und Ausschnitte der Datensätze „export“ und „exportest“ (Abb. 4).



Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	wald Chi-Square	Pr > ChiSq
Intercept	1	-1.7200	0.2911	34.9171	<.0001
sp09allx	1	-0.3179	0.5351	0.3530	0.5524

  

Odds Ratio Estimates			
Effect	Point Estimate	95% wald Confidence Limits	
sp09allx	0.728	0.255 2.077	

Abbildung 3: Ausschnitt des Ergebnisses der Prozedur Surveylogistic für ein Zentrum

Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio	centre	dependent	independent	method
2.440	0.540	11.000	11	ed01x	sp09allx	SURVEYLogistic
1.990	0.810	4.920	12	ed01x	sp09allx	SURVEYLogistic
4.380	1.640	11.700	13	ed01x	sp09allx	SURVEYLogistic
2.340	1.110	4.950	14	ed01x	sp09allx	SURVEYLogistic
0.840	0.250	2.870	15	ed01x	sp09allx	SURVEYLogistic
2.220	0.960	5.160	16	ed01x	sp09allx	SURVEYLogistic

  

	sp09allxBETA	Standard Error	centre
1	0.8925	0.7679	11
2	0.6898	0.4611	12
3	1.4761	0.5016	13
4	0.8500	0.3823	14
5	-0.1727	0.6269	15
6	0.7995	0.4294	16

Abbildung 4: Ausschnitt aus den Datensätzen export und exportest

Durch das Verbinden beider oben abgebildeter Dateien und durch einige Formatierungen erhält man einen Ergebnis-Datensatz, der nach Excel exportiert wird (Abb. 5).

	A	B	C	D	E	F	G	H	I	J
1	country	name	centre	model	method	sp09allx_OR	sp09allx_LOWOR	sp09allx_UPOR	sp09allxBETA	sp09allxSEBETA
2	Spain	Cartagena	11	ed01x = sp09allx	SURVEYLogistic	2.44	0.54	11	0.892499675	0.767929956
3	Spain	Almeria	12	ed01x = sp09allx	SURVEYLogistic	1.99	0.81	4.92	0.689786155	0.461131691
4	Spain	Madrid	13	ed01x = sp09allx	SURVEYLogistic	4.38	1.64	11.7	1.476141203	0.501603621
5	Spain	Valencia	14	ed01x = sp09allx	SURVEYLogistic	2.34	1.11	4.95	0.85003221	0.382292138
6	Albania	Tirana	15	ed01x = sp09allx	SURVEYLogistic	0.84	0.25	2.87	-0.17269587	0.626866681
7	Estonia	Tallinn	16	ed01x = sp09allx	SURVEYLogistic	2.22	0.96	5.16	0.799494375	0.429449533

Abbildung 5: Ausschnitt der exportierten Excel-Datei mit Zentrumsinformation, Modellinformation, den Odds Ratios und den Beta-Schätzern

## 3.2 Diskussion

Eine Verzerrung der Ergebnisse, bei der Berechnung von Odds Ratios in disproportional gezogenen Substichproben, kann durch die Verwendung der Prozedur Surveylogistic vermieden werden. Ein Beispiel für solche Ergebnisse ist die Odds ratio zwischen Neurodermitis Diagnose und Haut-Prick-Test im Zentrum 46. Dort wurde für beide Module eine disproportionale Substichprobe gewählt. Die Ergebnisse aus der logistischen Regression: 0,92 (0,37-2,28) und der gewichteten logistischen Regression: 0,73 (0,26-2,08) unterscheiden sich.

Die Einbindung der Prozedur Surveylogistic in ein Makro ermöglicht eine automatische, zentrumsspezifische Berechnung des Gewichts gezielt für jedes Modell. So benötigt der Anwender keine detaillierten Informationen über eventuell vorhandene Substichproben im zu berechnenden Modell, und eine manuelle Berechnung der Gewichte entfällt ebenfalls. Auch sind für die Anwendung des Makros keine spezifischen Makro-Kenntnisse notwendig. Alle relevanten Ergebnisse werden in einer Excel-Tabelle ausgegeben und können auch in anderen Softwareanwendungen einfach weiterverwendet werden (z.B.: Erstellung von Präsentationsgrafiken). Analog zu der Berechnung der Odds Ratios ist eine Berechnung gewichteter Prävalenzen durch die Prozedur Surveymeans möglich, ebenso der Einsatz von linearen Regressionsmodellen mit der Prozedur Surveyreg.

## Literatur

- [1] Weiland SK, Björkstén B, Brunekreef B, Cookson WOC, von Mutius E, Strachan DP and the international Study of Asthma and allergies in Childhood Phase II Study Group. Phase II of the International Study of Asthma and allergies in Childhood (ISAAC II): rationale and methods. Eur Respir J 2004; 24: 1-7
- [2] Richardson D, Rzehak P, Klenk J, Weiland SK. Analysis of Case-Control Data for Outcomes Other Than the One that Defined the Index Case Series. Epidemiology (in press)

## Anhang A: Vollständiger Programmcode des Makros:

Programm: Macro\_SurveyLog\_Call\_Min.sas; Makroaufruf und Einlesen des Datensatzes:

```
option mprint symbolgen mtrace spool;

filename min "G:\Daten1\I2CDC\Andrea\Makrosurveylogistic";
%INCLUDE min(WTWORK.sas); *Anlegen des temporären Datensatzes wtwork;
%INCLUDE min(Macro_SurveyLog_I_Min.sas); *Anlegen von %LRS1;
%INCLUDE min(Macro_Freq_Min.sas); *Anlegen von %freq;
%INCLUDE min(Macro_SurveyLog_II_Min.sas); *Anlegen von %Rlogmin;
```

%LRS1

```
(ed01x
,sp09allx
,E:\temp\
,11,12,13,14,15,16,17,23,24,
25,26,27,28,29,30,32,33,34,35,
36,37,38,39,40,42,44,45,46,48,49
,i=30,model=1);
```

Programm: WTWORk.sas; zum Einlesen des temporären Datensatzes

```
libname total 'G:\Daten1\I2CDC\SASdaten\I2CDC_total';
libname in 'G:\Daten1\I2CDC\SASdaten';
```

```
Data wtwork;
set total.i2cdc_total_05;
    strata=wh02x;* Def. of Srata Variable 1,2 Kodiert;
    if strata=0 then strata=2;
where partpup=1 and 8<=age<=12;
run;
```

Programm: Macro\_SurveyLog\_I\_Min.sas (Makro: LRS1); Bildung von Hilfsvariablen und dem Export-Datensatz:

```
%macro LRS1(DV,IV,path,centre1,centre2,centre3,centre4,centre5,
centre6,centre7,centre8,centre9,centre10,centre11,centre12,centre13,
centre14,centre15,centre16,centre17,centre18,centre19,centre20,
centre21,centre22,centre23,centre24,centre25,centre26,centre27,
centre28,centre29,centre30,i=1,model=1);
```

```
%let count=0;
%put &count;
```

```
/*die abhängige Variable*/
%let n=%eval(&model+1);
%let var&n = &DV;
%let mm&n = %qsubstr(&&dv,1,2);
```

```
/*Die unabhängige(n) Variable(n)*/
%local j;
%local k;
%do j=1 %to %eval(&model+1);
    %let var&j = %scan(&iv,&j);*Übergabe der jew. Variablennamen;
    %do k=1 %to &model;
        %let mm&k=%qsubstr(&&var&k,1,2);*die ersten 2
Buchstaben der unabhängigen Variablen werden ermittelt;
    %end;
```

```
/*Hilfsvariable für Proc Freq Statement*/
%let temp=%quote(&var&&j);
%if &j=1 %then %let freq=&temp;
%if &j > 1 and &j <= &model %then %let freq=&freq*&temp;
```

```

%if &j=1 %then %let hilf1=&temp%str(=.M);
%if &j > 1 and &j <= &model %then %let hilf1=&hilf1
%str(OR) &temp%str(=.M);

/*Hilfsvariable*/
%let info=%quote(&mm&&j%str(_S));

/*Hilfsvariable für das Keep statement*/
%if &j=1 %then %let keep=&info;
%if &j > 1 and &j <= &model %then %let keep=&keep &info;
%if &j=1 %then %let summe=&info;
%if &j > 1 and &j <= &model %then %let summe=&summe , &info;
%end;

/*****AUFRUF MACRO FREQ*****/
Proc Sort data=wtwork;by strata;run;
%Do z=1 %TO &i ;
    %LET cc=&centre&&z;%put ----- &cc -----;
    %freq(&cc);
%END;Title;
/*****/

Data zwischen;set wtwor;keep centre strata whyf whnf;run;
Proc Sort;by centre;run;
Data wtmin; merge modelfreq zwischen;by centre; run;

Data wtmin;
set wtmin;
IF strata=1 then wtmin=(1/(why_model/whyf)) ;
IF strata=2 then wtmin=(1/(whn_model/whnf)) ;
IF strata=.M then wtmin=.;
run;

Proc Sort data=wtwork;by centre;run;

Data wtwor1;merge wtwor wtmin;by centre;run;

Data wtwor2;
set wtwor1;
sample=sum(%unquote(&summe),%unquote(&&mm&n%str(_S)));

IF sample=. THEN DO;
    wtmin=.;strata=.;
END;
IF sample eq 0 THEN DO;
    wtmin=1;strata=1;
END;
run;

/*****AUFRUF MACRO RLOGMIN*****/
Data wtwor3;
set wtwor2(keep=&DV &IV wt: strata: centre partpup age);

```

```

where strata in (1,2);
run;
Proc Sort;by centre strata;run;
  %let count=0;
  %put &count;
  %Do j=1 %TO &i ;
  %LET cc=&centre&&j; %put ----- &cc -----;
  %rlogmin(&cc);
  %END;
/*****
/*****Excel Export*****/
Proc Sort data=export;by centre;run;
proc sort data=exportest;by centre;run;
Proc Sort data=in.centre;by centre;run;

data export;merge export exportest in.centre; by centre; run;

Data export;
length country $20.
      Name $20.
      centre 8.
      model $100.
      Method $20.;
set export;
model="&dv = &iv";
drop independent dependent ccode;
where &var1._OR ne .;
run;

Proc Print data=export;run;

%let model=&dv._&iv;

PROC EXPORT DATA=export
      OUTFILE= "&path&model"
      DBMS=EXCEL2000 REPLACE; RUN;
%mend;

```

Programm: Macro\_Freq\_Min.sas (Makro %freq); Berechnung der Kennzahlen für die Gewichte

```

%macro freq(centre);

%let count=%eval(&count+1);
%put &count;

%let freqvar=&&freq %str(*) &DV;

*Anlegen leerer Freq Datensätze;
Data freq;strata=.;centre=.;why_model=.;whn_model=.;cumall=.;run;
Data freq2;Data freq3;run;

```

```
Proc Freq data=wtwork;Ods output List=freq;
tables strata*&freqvar /list ;
Title "Frequency List for Centre: &centre"; where
centre=&centre;run;quit;
ods output close;
```

```
Data freq2;set freq end=lastobs;
keep centre strata cumfrequency why;;
centre=&centre;
where strata=1;
IF lastobs then why_model=cumfrequency;run;
```

```
Data freq3;set freq end=lastobs;
keep centre strata cumfrequency cumall;
centre=&centre;
where strata=2;
IF lastobs then cumall=cumfrequency;run;
```

```
Data freq&centre;
merge freq2 freq3;by centre;
keep centre why_model whn_model cumall;
whn_model=cumall-why_model;
IF why_model=. or whn_model=. then delete;
run;
```

```
%IF &count=1 %THEN %DO ;
    data modelfreq;
    set freq&centre;
    run;
%END;
%ELSE %DO ;
    data modelfreq;
    set modelfreq freq&centre;
    run;
%end;
```

```
proc print noobs;run;
proc sort;by centre;run;
%mend;
```

Programm: Macro\_RLogist\_II\_Min.sas (Makro %rlogmin); Berechnung der gewichteten Odds Ratios

```
%macro rlogmin(centre);

%let string="";
%let betastr="";

*Neue leere Risk und Betas Datensätze anlegen;
Data risk;modelrhs=.;Data betas;modelrhs=.;run;
```

```

%let count=%eval(&count+1);
%put &count;

/*****Proc RLOGIST*****/
data data&centre; set wtwork3; where centre=&centre; run;

ods output parameterestimates=betas oddsratios=risk;
proc surveylogistic data=data&centre;
title "OR subsample for centre: &centre";
model &DV (desc)= &IV ;
strata strata;
weight wtmin;
run;
ods output close;

/*****Odds Ratios*****/
%DO k=1 %TO &model;
data risk; set risk; n=_N_; run;

Data risk&k; set risk;
where n=%eval(&k);
keep OddsRatioEst lowercl uppercl;
OddsRatioEst=round(OddsRatioEst, .01);
lowercl=round(lowercl, .01);
uppercl=round(uppercl, .01);
Rename oddsratioest=&&var&k.._OR lowercl=&&var&k.._LOWOR
      uppercl=&&var&k.._UPOR;
%let string=&string risk&k;
run;
%END;

Data rlogist;
merge &string;
centre=&centre;
dependent="&DV";
independent="&IV";
method="SURVEYLogistic";
run;

%IF &count=1 %THEN %DO ;
      data export;set rlogist;run;
%END;
%ELSE %DO ;
      data export;set export rlogist;      run;
%end;

/****ESTIMATES*****/
%DO k=1 %TO &model;
data betas; set betas; n=_N_; run;

Data beta&k;

```

```
set betas;  
where n=%eval(&k+1);  
keep estimate stderr ;  
rename estimate=&&var&k..BETA stderr=&&var&k..SEBETA;  
%let betastr=&betastr beta&k;  
run;  
%END;  
  
Data est; merge &betastr; centre=&centre; run;  
  
%IF &count=1 %THEN %DO ;  
    data exportest; set est; run;  
%END;  
%ELSE %DO ;  
    data exportest;set exportest est;run;  
%end;  
%mend;
```