

Introduktion til SAS.

Per Kragh Andersen

1

SAS

- Display manager (programmering)
 - Vinduer:
 - * program editor (med syntaks-check)
 - * log
 - * output
 - reproducerbart (program teksten kan gemmes og **eksakt** den samme analyse kan reproduceres)
 - enkelt at dokumentere.

2

- SAS analyst
 - menu/skema-orienteret overbygning
 - skriver og kører programmer for en
 - ingen udenadslære, ingen syntaksfejl
 - let at importere filer med andre formater

Men:

- man får ikke alt med
- det er tungt at bruge i længden
- reproducerbarheden mistes.

3

Eksempel: Blodtryk og fedme

OBESE: vgt/idealvgt, BP: systolisk blodtryk

SEX	OBESE	BP
male	1.31	130
male	1.31	148	female	1.25	98
male	1.19	146	female	1.24	110
male	1.11	122	female	1.27	118
male	1.34	140	female	1.57	116
male	1.17	146	female	1.30	118
male	1.56	132	female	1.32	138
male	1.18	110	female	1.41	142
male	1.04	124	female	1.21	124
male	1.03	150	female	1.20	120
..	female	2.20	136
..	female	1.64	136
..	female	1.73	208

4

Data ligger i tekst filen `t:\bp.txt`, som indeholder følgende variable

- **SEX**: karaktervariabel
- **OBESE**: fedmegrad, dvs. vægt/ideelvægt
- **BP**: systolisk blodtryk

Dvs.:

- **3 variable**
- **102 observationer**

5

Indlæsning og udskrivning i SAS:

```
data bp;
  infile 't:\bp.txt' firstobs=2;
  input sex $ obese bp;
run;

proc print data=bp;
  var sex obese bp;
run;
```

Her laves et midlertidigt datasæt `bp`, som kun eksisterer indenfor det program, der er tale om.

Derefter skrives det ud, dvs. der kommer en liste i `output`-vinduet.

6

Direkte programmering

- Indlæsning og datamanipulationer:

Datastep:

```
data a;
  < evt. indlsning >
  < data manipulation >
run;
```

- Procedurekald, `proc xxx`

```
proc XXX data = a ;
  < procedure specifikationer >
run;
```

Datamanipulationer **kun** mellem `"data a;"` og `"run;"`.

7

Eksempel:

```
data bp;
  infile 't:\bp.txt' firstobs=2 ;
  input sex obese bp ;
  bp150 = (bp>=150);
run ;

proc print data=bp;
  var sex obese bp bp150;
run ;

proc freq data=bp;
  table sex * bp150 ;
run;
```

8

Indtastning af program i Editor vinduet:

- Piletaster, `backspace`, `delete`, `Home` og `End` virker som de plejer.
- Kvajetast i kommandobjælken, eller `Ctrl-Z`.
- Gyldige SAS-ord bliver blå når de er korrekt stavet.

Når programmet er kørt (`Run`→`Submit`, den lille mand der løber eller `F8`) kommer der resultater i:

- **Log-vindue:**
Her kan man se, hvordan kørslen er gået
 - hvor mange observationer, man har
 - hvor mange variable, der er
 - om der var nogen fejl
 - hvilke sider, der er skrevet af hvilke procedurer
- **Output-vindue:** selve resultaterne (hvis der er nogen).

- **Graph-vindue** (måske)
 - Her gemmes evt. højopløselige plots i rækkefølge

Der skiftes mellem vinduerne ved at klikke på `Windows` i kommandobjælken.

Ændringer i programmet:

Når man kører programmet, bliver **hele** teksten i `Editor`-vinduet kørt. Hvis man kun ønsker at køre **en del** af programmet, skal man blot først markere den del, man vil køre.

Bemærk:

- SAS-kørslerne (`Log`- `Output`- og `Graph`-vindue) kumulerer, dvs. alt bliver gemt fortløbende.
 - Slet en gang imellem, vælg `Edit`→`Clear`, eller `Ctrl-E`.
 - **Lad være** med at printe ud!
- Husk at gemme selve programmet ind imellem.

Det er god tone at specificere input-datasættet i alle procedurekald:

```
data bp;
  infile 't:\bp.txt' firstobs=2 ;
  input sex obese bp ;
run ;

proc univariate data = bp ;
  var obese bp ;
run ;
```

SAS-Datasæt

Et SAS-datasæt er en fil hvor data ligger i et særligt format.

Man skelner mellem at lave

- **Permanente** SAS-datasæt:
data sasuser.bp;
her, men måske noget andet hjemme
- **Temporære** (midlertidige) SAS-datasæt:
data bp;
forsvinder, når I lukker ned for SAS

13

Ganske tilsvarende, når man skal anvende sådanne, men lige først skal modificere dem:

- **Permanente** SAS-datasæt:
data ny ;
set sasuser.bp ;
...
run ;
- **Temporære** (midlertidige) SAS-datasæt:
data ny ;
set bp ;
...
run ;

14

Brug af proc freq.

```
data bp;  
  infile 't:\bp.txt' firstobs=2;  
  input sex $ obese bp;  
run;  
  
data bp; set bp;  
  bp150 = (bp>150);  
  fed=(obese>1.3);  
run;  
  
proc freq data=bp;  
  table sex*bp150/nocol nopercnt relrisk chisq;  
run;
```

15

Brug af proc freq.

TABLE OF SEX BY BP150

SEX	BP150		
	Frequency		
Row Pct	0	1	Total
-----+-----+-----+			
female	54	4	58
	93.10	6.90	
-----+-----+-----+			
male	42	2	44
	95.45	4.55	
-----+-----+-----+			
Total	96	6	102

16

STATISTICS FOR TABLE OF SEX BY BP150

Statistic	DF	Value	Prob
Chi-Square	1	0.250	0.617
Likelihood Ratio Chi-Square	1	0.256	0.613
Continuity Adj. Chi-Square	1	0.006	0.940
Mantel-Haenszel Chi-Square	1	0.247	0.619
Fisher's Exact Test (Left)			0.478
(Right)			0.820
(2-Tail)			0.697
Phi Coefficient		-0.049	
Contingency Coefficient		0.049	
Cramer's V		-0.049	

17

etc. etc.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95%	
		Confidence Bounds	
Case-Control	0.643	0.112	3.680
Cohort (Col1 Risk)	0.975	0.887	1.073
Cohort (Col2 Risk)	1.517	0.291	7.912

Sample Size = 102

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

18

Brug af proc freq: Mantel-Haenszel analyse.

```

data bp;
  infile 't:\bp.txt' firstobs=2;
  input sex $ obese bp;
run;

data bp; set bp;
  bp150 = (bp>150);
  fed=(obese>1.3);
run;

proc freq data=bp;
  table fed*sex*bp150/nocol nopercent relrisk chisq cmh;
run;

```

19

The FREQ Procedure

Table 1 of sex by bp150
Controlling for fed=0

sex	bp150		Total
Frequency	0	1	
Row Pct			
female	21	1	22
	95.45	4.55	
male	33	1	34
	97.06	2.94	
Total	54	2	56

19-1

Statistics for Table 1 of sex by bp150
Controlling for fed=0

Statistic	DF	Value	Prob
Chi-Square	1	0.0998	0.7520

etc. etc.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits
Case-Control (Odds Ratio)	0.6364	0.0377 10.7328
Cohort (Col1 Risk)	0.9835	0.8825 1.0960
Cohort (Col2 Risk)	1.5455	0.1019 23.4472

19-2

Table 2 of sex by bp150
Controlling for fed=1

sex	bp150		Total
female	33	3	36
male	9	1	10
Total	42	4	46

19-3

Statistics for Table 2 of sex by bp150
Controlling for fed=1

Statistic	DF	Value	Prob
Chi-Square	1	0.0274	0.8686

etc. etc.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits
Case-Control (Odds Ratio)	1.2222	0.1131 13.2082
Cohort (Col1 Risk)	1.0185	0.8102 1.2805
Cohort (Col2 Risk)	0.8333	0.0969 7.1684

19-4

Summary Statistics for sex by bp150
Controlling for fed

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0064	0.9364
2	Row Mean Scores Differ	1	0.0064	0.9364
3	General Association	1	0.0064	0.9364

19-5

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value
Case-Control (Odds Ratio)	Mantel-Haenszel Logit	0.9287 0.9322
Cohort (Col1 Risk)	Mantel-Haenszel Logit	0.9958 0.9898
Cohort (Col2 Risk)	Mantel-Haenszel Logit	1.0713 1.0571

19-6

Brug af proc freq: Mantel-Haenszel analyse.

Af og til har man data som på forhånd er tabellerede, dvs. man ikke har data for de enkelte individer, men kun en tabel som f.eks:

	Age ≥ 60		Age < 60	
	MI yes	MI no	MI yes	MI no
SBP ≥ 140	9	115	20	596
SBP < 140	6	73	21	1171

Hvis man gerne vil tabellere disse tal og bruge `proc freq` til at regne på dem ville det være yderst upraktisk at skulle danne et datasæt med i dette tilfælde 2011 observationer. Derfor tillader `proc freq` at man lader hver observation i datasættet repræsentere flere personer. Antallet af personer for en given kombination af variablene skal være i en variabel som man nævner i et `weight`-statement:

Type of Study	Method	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel Logit	0.1463 0.1510	5.8937 5.7549
Cohort (Col1 Risk)	Mantel-Haenszel Logit	0.8945 0.8975	1.1086 1.0916
Cohort (Col2 Risk)	Mantel-Haenszel Logit	0.2017 0.1955	5.6903 5.7147

Breslow-Day Test for
Homogeneity of the Odds Ratios

Chi-Square	0.1209
DF	1
Pr > ChiSq	0.7280

19-7

Brug af proc freq: Mantel-Haenszel analyse.

```
data KSTab55;
input age $ sbp $ mi $ antal;
cards;
ge60 ge140 ja 9
ge60 ge140 nej 115
ge60 lt140 ja 6
ge60 lt140 nej 73
lt60 ge140 ja 20
lt60 ge140 nej 596
lt60 lt140 ja 21
lt60 lt140 nej 1171
;
run;
```

```

proc freq;
tables sbp*age*mi/nocol nopercnt relrisk cmh;
weight antal;
run;

/* Lg mrke til weight-kommandoen. Kommandoen 'cmh'
   giver Mantel-Haenszel analysen.
*/

```

TABLE 1 OF AGE BY MI
CONTROLLING FOR SBP=ge140

AGE	MI		
Frequency			
Row Pct	ja	nej	Total
ge60	9	115	124
	7.26	92.74	
lt60	20	596	616
	3.25	96.75	
Total	29	711	740

etc. etc.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Bounds	
Case-Control	2.332	1.036	5.251
Cohort (Col1 Risk)	2.235	1.043	4.793
Cohort (Col2 Risk)	0.959	0.911	1.009

Sample Size = 740

TABLE 2 OF AGE BY MI
CONTROLLING FOR SBP=lt140

AGE	MI		
Frequency			
Row Pct	ja	nej	Total
ge60	6	73	79
	7.59	92.41	
lt60	21	1171	1192
	1.76	98.24	
Total	27	1244	1271

... etc.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95%	
		Confidence Bounds	
Case-Control	4.583	1.795	11.705
Cohort (Col1 Risk)	4.311	1.791	10.375
Cohort (Col2 Risk)	0.941	0.883	1.002

Sample Size = 1271

26

SUMMARY STATISTICS FOR AGE BY MI
CONTROLLING FOR SBP

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	13.179	0.001
2	Row Mean Scores Differ	1	13.179	0.001
3	General Association	1	13.179	0.001

27

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95%	
			Confidence Bounds	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.961	1.648	5.322
	Logit	3.115	1.686	5.754
Cohort (Col1 Risk)	Mantel-Haenszel	2.817	1.611	4.928
	Logit	2.965	1.667	5.273
Cohort (Col2 Risk)	Mantel-Haenszel	0.951	0.926	0.977
	Logit	0.951	0.914	0.990

The confidence bounds for the M-H estimates are test-based.

Breslow-Day Test for Homogeneity of the Odds Ratios

Chi-Square = 1.180 DF = 1 Prob = 0.277

Total Sample Size = 2011

28

Brug af proc genmod: Logistisk regression.

```

data bp;
  infile 't:\bp.txt' firstobs=2;
  input sex $ obese bp;
run;

data bp; set bp;
  bp150 = (bp>150);
run;

proc genmod data=bp descending;
  class sex;
  model bp150=sex/dist=bin type3;
  estimate 'm vs. f' sex -1 1 / exp;
run;

```

29

The GENMOD Procedure

Model Information

Data Set WORK.BP
 Distribution Binomial
 Link Function Logit
 Dependent Variable bp150
 Observations Used 102

Class Level Information

Class	Levels	Values
sex	2	female male

30

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	100	45.3826	0.4538
Scaled Deviance	100	45.3826	0.4538
Pearson Chi-Square	100	102.0000	1.0200
Scaled Pearson X2	100	102.0000	1.0200
Log Likelihood		-22.6913	

Algorithm converged.

32

Response Profile

Ordered Value	bp150	Total Frequency
1	1	6
2	0	96

PROC GENMOD is modeling the probability that bp150='1'.

Parameter Information

Parameter	Effect	sex
Prm1	Intercept	
Prm2	sex	female
Prm3	sex	male

31

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square
Intercept	1	-3.0445	0.7237	-4.4630 -1.6260	17.1
sex female	1	0.4418	0.8901	-1.3028 2.1865	0.1
sex male	0	0.0000	0.0000	0.0000 0.0000	.
Scale	0	1.0000	0.0000	1.0000 1.0000	

Parameter Pr > ChiSq

Intercept	<.0001
sex female	0.6196
sex male	.

33

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
sex	1	0.26	0.6130

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits	Chi-Square
m vs. f	-0.4418	0.8901	0.05	-2.1865 1.3028	0.2
Exp(m vs. f)	0.6429	0.5722	0.05	0.1123 3.6795	

Label	Pr > ChiSq
m vs. f	0.6196
Exp(m vs. f)	