

Kort intro til SAS

Forår 2017

Janne Petersen
Judith L Jacobsen
Lene Theil Skovgaard

1

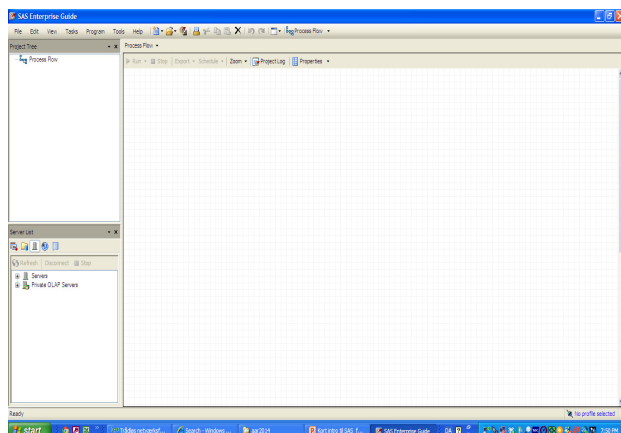
Hvorfor SAS

- Kan "alt"
- Alle ph.d. studerende har gratis adgang
 - Fra universitetet eller hospitalerne
- Kode --- hjælp fra Enterprise Guide

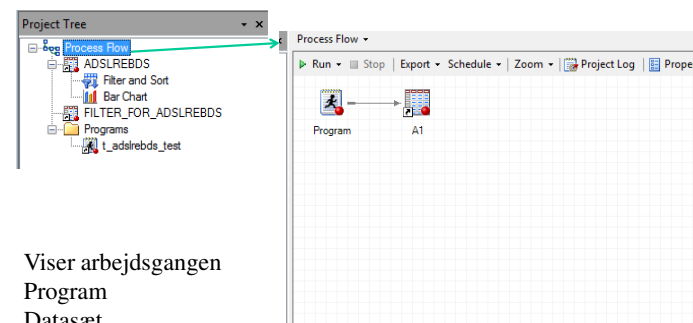
Alternative programmer SPSS, R og STATA

2

Enterprise Guide

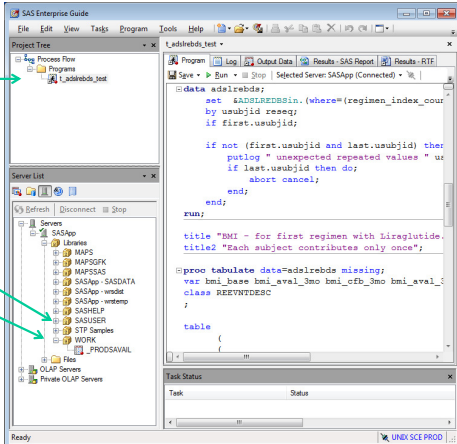


Arbejdspladsen



Viser arbejdsgangen
Program
Datasæt
Handler

Folder struktur etc.



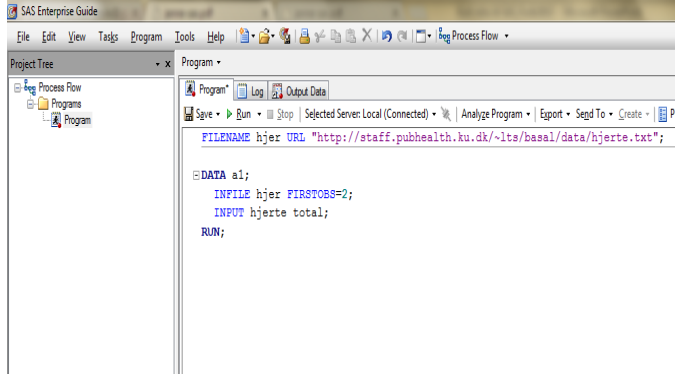
Process flow
Programmer

Filer og mapper opsætning

Datasæt kan gemmes
permanent i SASUSER

Og midlertidige i Work
mappe

Faneblade – Program

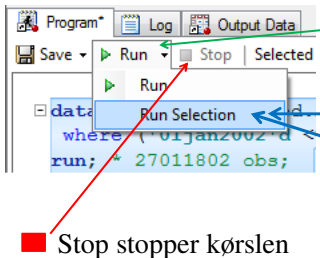


```
FILENAME hjer URL "http://staff.pubhealth.ku.dk/~lts/basal/data/hjerte.txt";

DATA a1;
  INFILE hjer FIRSTOBS=2;
  INPUT hjer total;
RUN;
```

Her skriver vi koden med hvad SAS skal gøre

Run Program

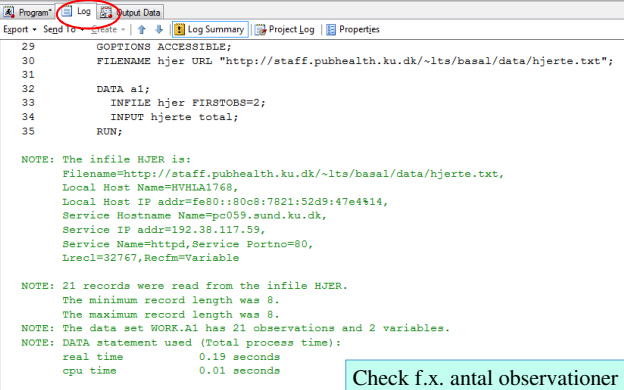


Kører hele programmet

Drop down menu ELLER
F8 kører den markerede
del af et program

Stop stopper kørslen
Den er rød under kørsel
Ellers grå

Faneblad – Log



```
29 GOPTIONS ACCESSIBLE;
30 FILENAME hjer URL "http://staff.pubhealth.ku.dk/~lts/basal/data/hjerte.txt";
31
32 DATA a1;
33   INFILE hjer FIRSTOBS=2;
34   INPUT hjer total;
35   RUN;
```

NOTE: The infile HJER is:
Filename=http://staff.pubhealth.ku.dk/~lts/basal/data/hjerte.txt,
Local Host Name=HVHIA1768,
Local Host IP addr=fe80::80c8:7821:52d9:47e4%14,
Service Hostname Name=pc059.sund.ku.dk,
Service IP addr=192.38.117.59,
Service Name=httpd,Service Portno=80,
Lrecl=32767,Recfm=Variable

NOTE: 21 records were read from the infile HJER.
The minimum record length was 8.
The maximum record length was 8.

NOTE: The data set WORK.A1 has 21 observations and 2 variables.

NOTE: DATA statement used (Total process time):
real time 0.19 seconds
cpu time 0.01 seconds

Check f.x. antal observationer

Loggen er jeres ven

Note (grøn) Informationer der ikke er fejl, men oplysninger om hvad SAS har gjort, f.eks. Antal observationer i et datasæt

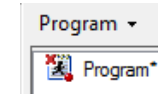
Warning (blålig) Viser fejl som SAS selv har rettet---tjek at det er gjort rigtigt

Error (rød) Syntaks fejl, som sas ikke har kunnet rette, programmet er ikke gennemført

9

Fanen sladrer om loggen

- Et rødt x fortæller at der er errors i program kørslen



- Ligeledes vil en gul trekant vises i øverste venstre hjørne, hvis der er en warning



10

Faneblad – Output

	usubjid	sex	brthdm	rfstdm	read_code	diabdiagdm	diabtype	dt	res
1	1000020	M	1946-01-01	1996-06-26	C10F.00	2004-02-20	Type 2	2	
2	1000036	M	1945-01-01	1991-02-28	C109.00	2002-05-22	Type 2	2	
3	10006142	M	1951-01-01	2000-09-08	C10F.00	1996-06-20	Type 2	2	
4	10007513	F	1943-01-01	1999-12-13	C10F.00	1997-07-16	Type 2	2	
5	10008467	M	1964-01-01	1999-09-16	C10F.00	2003-08-12	Type 2	2	
6	10012490	M	1948-01-01				Type 2	2	
7	1001480	M	1948-01-01				Type 2	2	
8	10019643	F	1954-01-01				Type 2	2	
9	10022119	M	1956-01-01				Type 2	2	
10	100289	M	1957-01-01				Type 2	2	
11	10032294	M	1940-01-01				Type 2	2	
12	10034636	M	1950-01-01				Type 2	2	
13	10048476	F	1961-01-01				Type 2	2	
14	10064165	F	1954-01-01				Type 2	2	
15	10064174	F	1946-01-01				Type 2	2	
16	10067318	M	1939-01-01	2000-11-17	C10F.00	1993-09-28	Type 2	2	
17	1007018	M	1955-01-01	1995-12-18	C10F.00	2003-10-17	Type 2	2	
18	1007040	F	1948-01-01	2004-02-08	C10F.00	2006-04-20	Type 2	2	

Variable

Numeriske variable er altid tal.

Eks. Alder med værdier 45, 37 ...

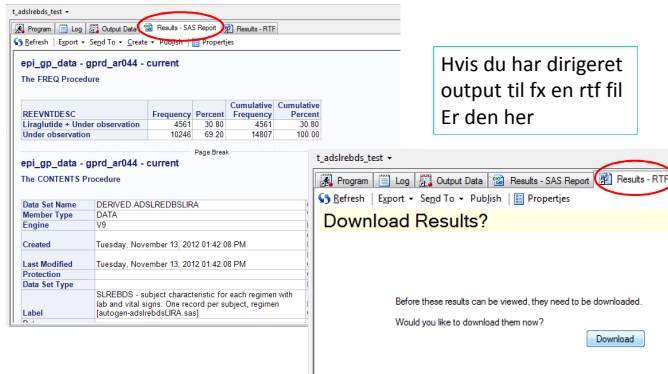
Man kan kun beregne gennemsnit mm. på numeriske variable

Tekst variable er primært bogstaver men kan også være tal.

Eks. Month med værdier "jan", "feb" ...

12

Faneblad – Results



Hvis du har dirigeret output til fx en rtf fil
Er den her

Frequency	Percent	Cumulative Frequency	Cumulative Percent
4561	30.80	4561	30.80
10246	69.20	14807	100.00

Download Results?

Before these results can be viewed, they need to be downloaded.
Would you like to download them now?

Download

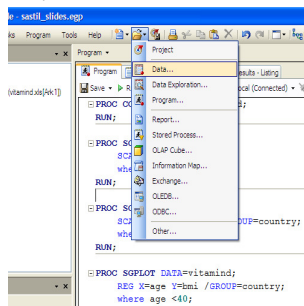
Gem – faner eller projekt

- Shift Ctrl S (eller:  Save) gemmer programmet (eller loggen, eller...)
- Ctrl S gemmer projektet
 - Arbejdsfladen,
 - de (temporære) dataset,
 - output, etc. – alt gemmes

Indlæs data---Metode 1

1. Download excel datasættet fra Lene Theil Skovgaards hjemmeside

2.



Indlæser sas datasæt
(.sas7bdat), excel filer mm.

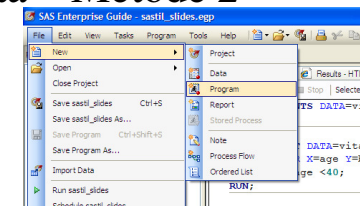
SAS kode, der indlæser
data genereres automatisk

15

Indlæs data---Metode 2

Åben et nyt program
-og skriv

**Denne metode virker
bedst på dette kursus
hvis net adgang!!!**



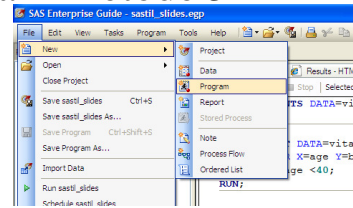
DATA a1;
INFILE "http://staff.pubhealth.ku.dk/~lts/basal/data/fil1.txt" **URL FIRSTOBS=2;**
INPUT hjerte total;
RUN;

Navne på alle de variable, der er i
datasættet, i rigtig rækkefølge

16

Indlæs data---Metode 3

Åben et nyt program
-og skriv



```
DATA a1;  
  INFILE "C:\kurser\basalstat\hjerne.txt" FIRSTOBS=2;  
  INPUT hjerte total;  
RUN;
```

Navne på alle de variable,
der er i datasættet, i rigtig
rækkefølge

Denne metode virker bedst
på dette kursus hvis
ingen net adgang, og
data downloadet

17

SAS programmering

SAS-kode foregår i to forskellige typer trin

DATA (data trin):

1. Indlæser et datasæt
2. Laver nye variable, sletter variable mm.
3. Generer nyt datasæt med de nye rettelser

PROC (procedure trin):

1. Producerer informationer om data, f.eks. tabeller, grafer og analyser

18

Data trin

```
DATA SASUSER.ny;  
  SET WORK.gammel;  
RUN;
```

Indlæser datasættet "gammel" fra work og
gemmer datasættet ny i sasuser-de er helt
ens!

Husk ; efter hver linie

- **WORK** er midlertidigt
- **SASUSER** er permanent

19

Noter---skriv mange!

Noter omringes af /* og */ og læses ikke af SAS
(genvej til note - **CTRL ***)

Eksempel:

```
DATA ny;  
  SET WORK.gammel;/*work er ikke nødvendigt*/  
RUN;
```

20

Definer ny variabel

```
DATA ny;  
  SET gammel;  
  BMI=vagt/(hoejde**2);  
RUN;
```

Datasættet "gammel" skal indeholde variable vagt og hoejde, det nye datasæt "ny" indeholder også BMI

21

"Pas på missing

```
DATA ny;  
  SET gammel;  
  BMI=vagt/(hoejde**2);  
  IF .z<BMI<24 THEN BMI2gr=1;  
  IF BMI>=24 THEN BMI2gr=2;  
RUN;
```

Ny variable BMI2gr inddeler BMI i 2 grupper alt efter om BMI er over eller under 24. ".z<" sikrer at missing bliver ved at være missing₂₂

Operatorer

		Forklaring
=	EQ	Lig med
^=	NE	Forskellig fra
<	LT	Mindre end
>	GT	Større end
<=	LE	Mindre end eller lig
>=	GE	Større end eller lig
	IN	Tilhører
&	AND	Og
	OR	Eller
^	NOT	Negation

23

Regnefunktioner

	Forklaring
*	Multiplikation
/	Division
+	Plus
-	Minus
**	Potensopløftning $a**b = a^b$
Exp(variable)	Eksponentialfunktion
Log(variable)	Naturlig logaritme
Log2(variable)	2-tals logaritme
Log10(variable)	10-tals logaritme
ABS(variable)	Numerisk værdi
ROUND(variable)	Afrundet værdi

24

Procedure trin

SAS indeholder en lang række procedurer, der er udviklet på forskellige tidspunkter og derfor ikke altid er konsistente i notationen

Ligesom datatrin starter med **DATA** starter procedure trin med **PROC**

25

Procedure struktur

```
PROC procnavn DATA=datasætnavn;  
    ordre vedrørende udførelsen  
RUN;
```

Datasætnavn angiver hvilket datasæt proceduren skal arbejde på

26

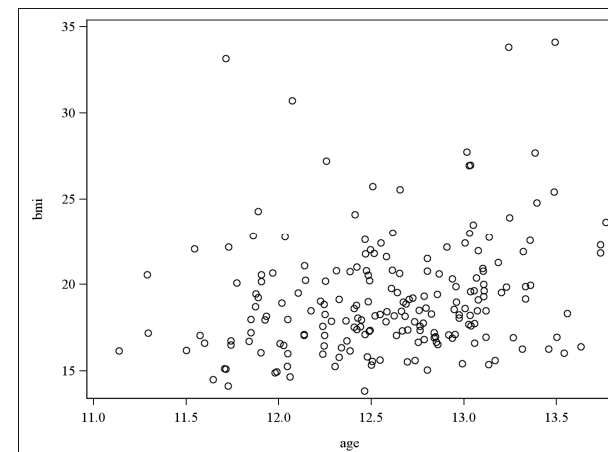
Scatter plot

```
PROC SGPlot DATA=vitamind;  
    SCATTER X=age Y=bmi;  
RUN;
```

Plotter bmi mod age.

Husk at overveje hvad der skal være på X-aksen og på Y-aksen. Det, der er på Y-aksen ses som funktion af det på X-aksen.

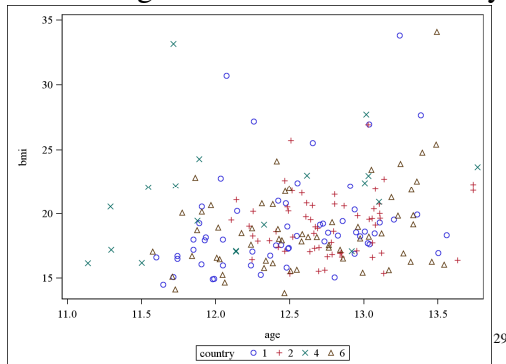
27



28

Scatter plot med grupper

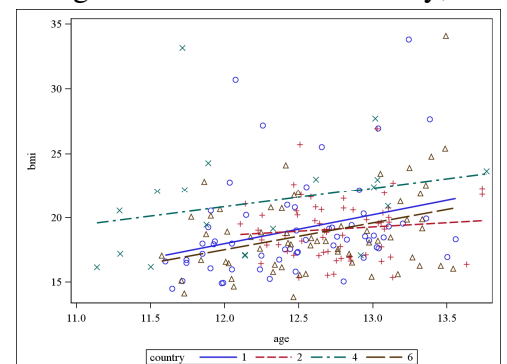
```
PROC SGPLOT DATA=vitamind;  
  SCATTER X=age Y=bmi/GROUP=country;  
RUN;
```



29

Scatter plot med linier

```
PROC SGPLOT DATA=vitamind;  
  REG X=age Y=bmi/GROUP=country;  
RUN;
```



30

SAS hjælp

- Google is your friend
- Enterprise Guide editoren
- F1 hjælp
- <http://support.sas.com/documentation/onlinedoc/base/>

SAS online kurser

- <http://www.ats.ucla.edu/stat/seminars/>
Specielt dem med film er ok

31

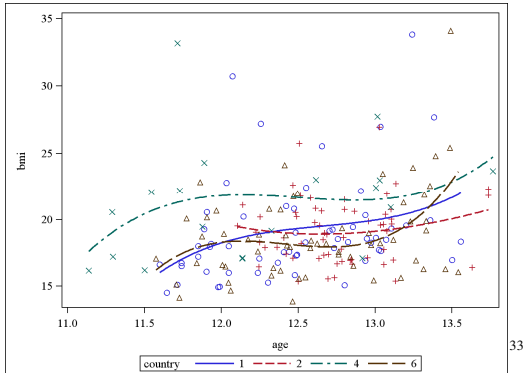
RESTEN

Resten af disse slides er nyttige informationer og ting I sikkert vil få brug for men vil ikke blive gennemgået!

32

Scatter plot med kurver

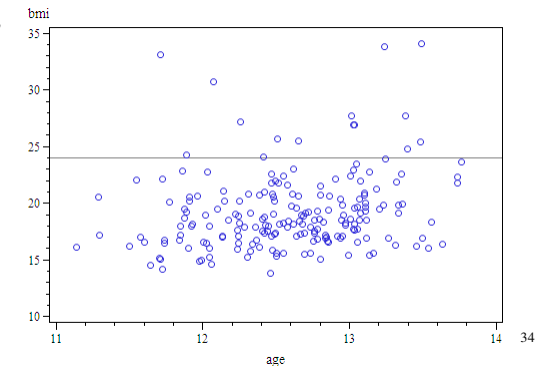
```
PROC SGPLOT DATA=vitamind;
  REG X=age Y=bmi/GROUP=country DEGREE=3;
  RUN;
```



33

Scatter plot referencelinie

```
PROC GGPLOT DATA=vitamind;
  PLOT bmi*age/VREF=24;
  RUN;
```

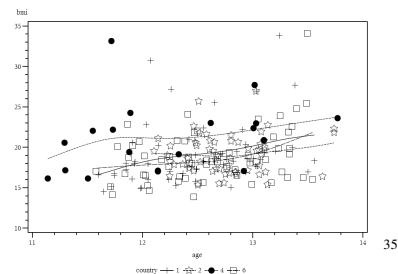


34

Scatter plot med mere kontrol

```
PROC GGPLOT DATA=vitamind;
  PLOT bmi*age=country;
  SYMBOL1 v=plus C=Black I=SM70S;
  SYMBOL2 v=star C=Black I=SM70S;
  SYMBOL3 v=dot C=Black I=SM70S;
  SYMBOL4 v=square C=Black I=SM70S;
  RUN;
```

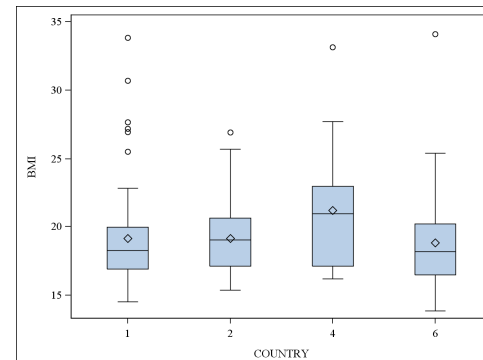
Der er mange forskellige muligheder vha. symbol statements, Enterprise Guide kan hjælpe med muligheder.



35

Box plot

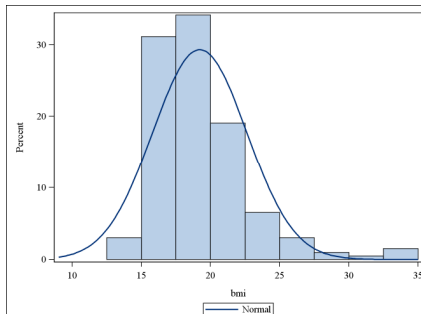
```
PROC SGPLOT DATA=vitamind;
  VBOX bmi/CATEGORY=country;
  RUN;
```



36

Histogram

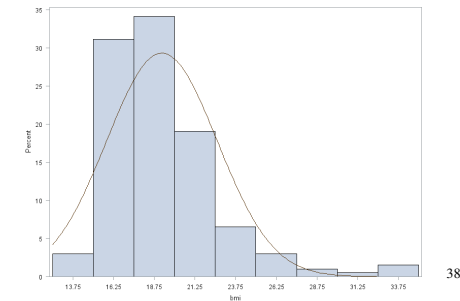
```
PROC SGPLOT DATA=vitamind;
  HISTOGRAM bmi;
  DENSITY bmi;
RUN;
```



37

Histogram – alternativ kodning

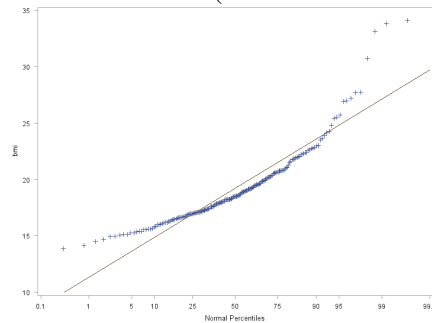
```
PROC UNIVARIATE DATA=VITAMIND;
  VAR bmi;
  HISTOGRAM bmi/NORMAL;
RUN;
```



38

Probability Plot

```
PROC UNIVARIATE DATA=VITAMIND;
  VAR bmi;
  PROBLOT bmi/NORMAL(MU=est SIGMA=est);
RUN;
```



39

Beregning af fraktiler

```
PROC UNIVARIATE DATA=VITAMIND;
  VAR bmi;
  OUTPUT OUT=p_res pctlpre=P pctlpts=2.5,97.5;
RUN;
```

Her beregnes 2,5 og 97,5 fraktilerne, og resultatet ligger i datasættet "p_res"

40

Deskriptiv statistik

```
PROC MEANS DATA=vitamind N NMISS MEAN
MEDIAN P25 P75 MAXDEC=2;
```

```
VAR bmi age;
```

```
RUN;
```

Variable	N	Miss	Mean	Median	25th Pctl	75th Pctl
bmi	412	0	23.60	23.03	18.56	27.64
age	412	0	43.21	69.47	12.64	71.84

Udover **N NMISS MEAN MEDIAN P25 P75 MAXDEC** kan man skrive et mange andre options f.eks. **Stddev** for spredning.

41

Tabel

```
PROC FREQ DATA=VITAMIND;
```

```
TABLE country;
```

```
RUN;
```

country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	112	27.18	112	27.18
2	114	27.67	226	54.85
4	60	14.56	286	69.42
6	126	30.58	412	100.00

42

Tovejs Tabel

```
PROC FREQ DATA=VITAMIND;
```

```
TABLE country*category/NOPERCENT NOCOL;
```

```
RUN;
```

43

Table of country by category				
country	category			
Frequency				
Row Pct	1	2		Total
1	59	53		112
	52.68	47.32		
2	60	54		114
	52.63	47.37		
4	19	41		60
	31.67	68.33		
6	61	65		126
	48.41	51.59		
Total	199	213		412 ¹⁴

Resultat formater

- Klik på **Tools**
- Vælg **Options**
- Klik på **Resultater**
- Vælg de former I gerne vil have jeres resultater på f.eks. .RTF

Alternativ, behold default og eksporter output efterfølgende fra outputvinduet

45

Manglende værdier

Tekstvariable Manglende værdier registreres ved blankt felt

Numerisk variable Manglende værdi registreres ved . eller .x

Pas på . tæller som minus uendelig. Så skriv

IF .z<var <8 **THEN**... i stedet for

IF var <8 **THEN**...

46

Genvejstaster

Lister over genveje, vælg:

- Help
- SAS Enterprise Guide Help
- Keyboard Shoutcuts

	Forklaring
F8	Kører programmet
Ctrl i	Formaterer koden pænt
F1	Hjælp
Ctrl Shift H	Søger online for markeret ord
Ctrl *	/**/ Til udkommentering

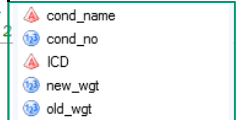
47

Udvalgt genvej

CTRL + Shift + V

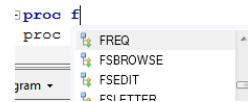
Åbner auto færdiggørelses vindue med en liste af data set kolonne navne.

```
data hes_charlson;  
merge hes_icd_hypo(drop=poi  
by icd;  
if ja1  
run; * 2
```



Variable som er det dataset du arbejder på

Hjælp ved hvert skridt



- Ved Proc
Forskellige procedurer foreslås ved hvert bogstav, indsnævres listen

```
proc sort data=  
data hes_charls  
merge hes_icd_  
by icd ;  
if ja1 and ja2  
run; * 232561 o  
/* data not_hes  
merge hes_icd_  
by icd;  
if not ja1 and  
run; * 134 obs ;
```

- Ved data
De tilgængelige dataset foreslås

CTRL + D
Åbner auto færdiggørelses
vindue med en liste af data set,
som blev lavet ved et DATA step.

Syntax Hjælp

```
proc sort data=hes_icd_hypo; by icd; run;
```

Keyword: [PROC](#)

Context: [GLOBAL STATEMENT] PROC statement

Syntax: PROC procedure-name <options>;

Begins a PROC step. The PROC step consists of a group of SAS statements that call and execute a procedure, usually with a SAS data set as input.

Search: [The Product Documentation](#), [Samples & SAS Notes](#), [Papers](#)

Hover (svæv) musen over **proc**

- En definition åbenbares

Procedure Syntax

```
proc sort data=  
data  
merg  
by i  
if j  
run;
```

Keyword: [SORT](#)

Context: [PROCEDURE DEFINITION] PROC SORT

Syntax: PROC SORT <collating-sequence-option> <other option(s)>;
BY <DESCENDING> variable-1 <...><DESCENDING> variable-n;

The SORT procedure orders SAS data set observations by the values of one or more character or numeric variables. The SORT procedure either replaces the original data set or creates a new data set. PROC SORT produces only an output data set.

Search: [The Product Documentation](#), [Samples & SAS Notes](#), [Papers](#)

Hover (svæv) musen over **sort**

- En definition og syntax hjælp åbenbares

Split skærm

Option 1: **Split skærm for at se forskellige sektioner af samme fane**

- Højre klik på den fane du ser på (der er også et ikon på øverste menu linje)
- Vælg Split – Stacked, Side by Side eller Both



Option 2: **Split skærm for at se forskellige faner**

- Vælg View fra Hoved Menuen
- Klik 'Workspace layout'
- Vælg Stacked eller side by side.
- Hver 'skærm del' har en drop down i venstre top hjørne hvor du kan bestemme hvad du vil se (e.g. Log OG Program).

Split skærm til at se forskellige sektioner af samme fane

Smart hvis man har et langt program Og vil se top og bund samtidigt.

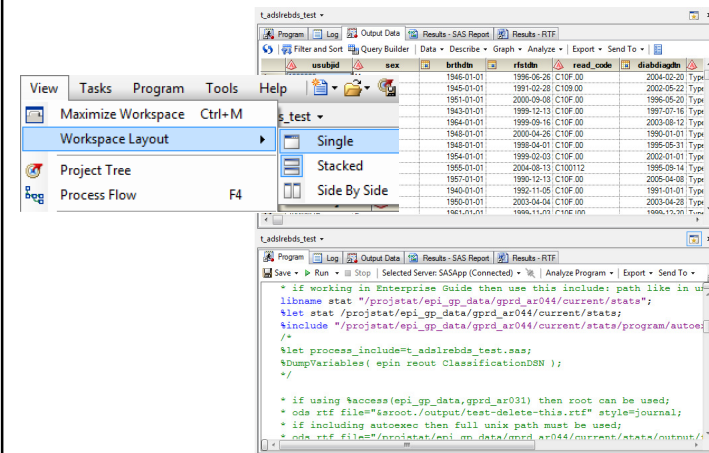
```

* If working in Enterprise Guide then use this include path like in u
libname stat "/projstat/epi_gp_data/gprd_ar044/current/ata";
%let stat /projstat/epi_gp_data/gprd_ar044/current/ata;
%include "/projstat/epi_gp_data/gprd_ar044/current/ata/program/autocoe
/*
%let process_include=t_adslrebbs_test.sas;
%DumpVariables( epin reout ClassificationDSN );
*/

proc tabulate data=adslrebbs missing;
var Hb1c_base Hb1c_aval_3mo Hb1c_cfb_3mo Hb1c_aval_3mo_d;
class REEVNTDESC
;
table
(
  Hb1c_base="Hb1c before"
  Hb1c_aval_3mo="Hb1c 3mo"
  Hb1c_cfb_3mo="Change"
)
* (n=F#F6.0 (mean std)*F#F5.1 (min max)*F#F5.1)
Hb1c_aval_3mo_d="Day"* (min max)*F#F5.0
)
;
REEVNTDESC
all / rspace=40 box="Some text"
;

```

Split skærm til forskellige faner



Resultater i word

```

ODS RTF FILE="c:\Janne\res_anal.rtf";
PROC FREQ DATA=VITAMIND;
  TABLE country*category/NOPERCENT NOCOL;
RUN;
ODS RTF CLOSE;

```

Alt hvad der står mellem ODS RTF FILE og ODS RTD CLOSE; kommer over i word dokumentet "c:\Janne\res_anal.rtf"

Resultater i word – lidt pænere

```

ODS RTF FILE="c:\Janne\res_anal.rtf" BODYTITLE
STARTPAGE=NO STYLE=Journal;
ODS NOPROCTITLE;
PROC FREQ DATA=VITAMIND;
  TABLE country*category/NOPERCENT NOCOL;
RUN;
ODS RTF CLOSE;

```