1. Introduction to SPSS programming

Karl B Christensen
http://publicifsv.sund.ku.dk/~kach/SPSS

Basic concept of SPSS programming

SPSS syntax is like a recipe - a series of instructions to be executed in a specified sequence.

- Write a SPSS syntax[1]
- Let SPSS interpret your program and do some statistical calculations
- SPSS responds by giving results in some format

Need to know rules for the SPSS language.

[1] or point and click and then click Paste

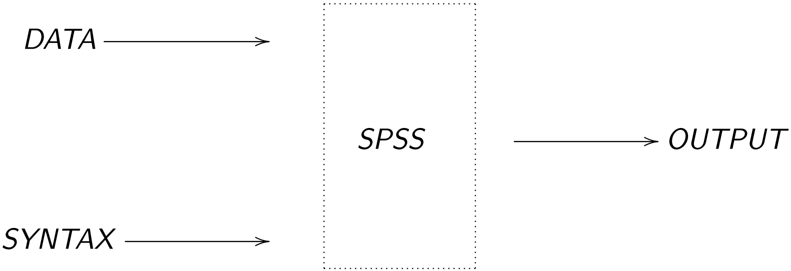
Typical use of SPSS for statistical analysis

1. You have data in some format (SPSS, Excel, SAS, text ...)
2. Get the data into SPSS
3. Look at the data using SPSS
4. Transform or select part of the data: make data ready for statistical analysis
5. Choose appropriate SPSS procedure. Generate syntax.
6. Get your results out of SPSS and SPSS syntax
7. Make sure that SPSS did what you asked for[2]
8. Interpret SPSS output

   Save data and SPSS syntax. Then you can reproduce your results later on and the steps you have taken are documented.

[2] have you, e.g., asked SPSS to compute the logarithm of a negative number?

The SPSS system



DATA

SYNTAX

SPSS

OUTPUT

- Point-and-click is convenient, but dangerous.
- Save syntax
- Collect the fragments into coherent `.sps` files that can be run from scratch.
- Test your programs in a freshly started SPSS session.

Advantages and disadvantages of SPSS

+ Reads Excel
+ You can see your data while you work
+ Cover many statistical methods
+ Easy to use

- Unflexible in advanced programming
- Hard to make good-looking graphics

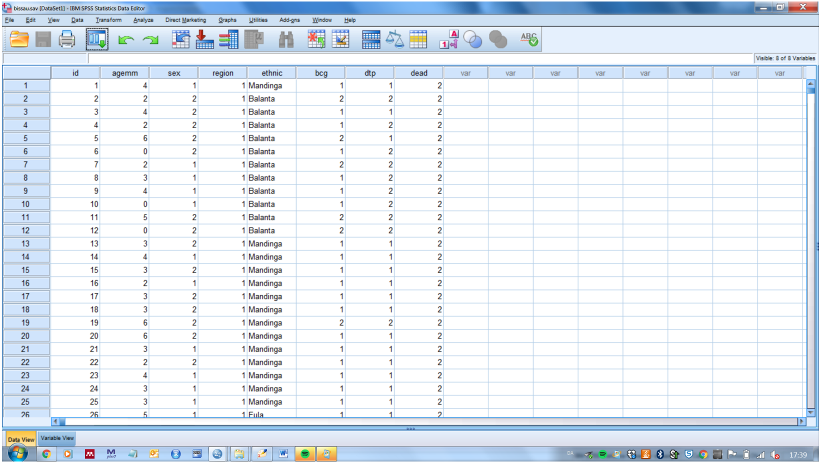
- Does not help you do reproducible research

SPSS windows

Framework for program development and data handling

- **Data** You can see your SPSS data
- **Syntax** Write you syntax
- **Output** Results

SPSS Data window

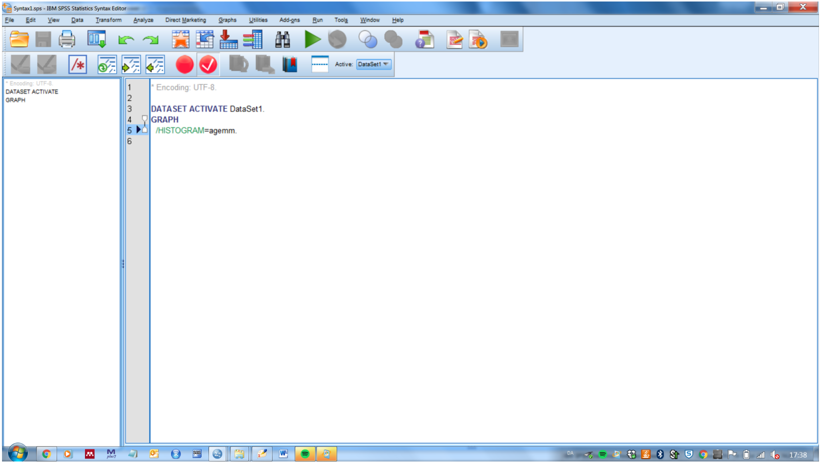


bissau.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 8 of 8 Variables

	id	agemm	sex	region	ethnic	bcg	dtp	dead	var	var	var	var	var	var	var	var
1	1	4	1	1	Mandinga	1	1	2								
2	2	2	2	1	Balanta	2	2	2								
3	3	4	2	1	Balanta	1	1	2								
4	4	2	2	1	Balanta	1	2	2								
5	5	6	2	1	Balanta	2	1	2								
6	6	0	2	1	Balanta	1	2	2								
7	7	2	1	1	Balanta	2	2	2								
8	8	3	1	1	Balanta	1	2	2								
9	9	4	1	1	Balanta	1	2	2								
10	10	0	1	1	Balanta	1	2	2								
11	11	5	2	1	Balanta	2	2	2								
12	12	0	2	1	Balanta	2	2	2								
13	13	3	2	1	Mandinga	1	1	2								
14	14	4	1	1	Mandinga	1	1	2								
15	15	3	2	1	Mandinga	1	1	2								
16	16	2	1	1	Mandinga	1	1	2								
17	17	3	2	1	Mandinga	1	1	2								
18	18	3	2	1	Mandinga	1	1	2								
19	19	6	2	1	Mandinga	2	2	2								
20	20	6	2	1	Mandinga	1	1	2								
21	21	3	1	1	Mandinga	1	1	2								
22	22	2	2	1	Mandinga	1	1	2								
23	23	4	1	1	Mandinga	1	1	2								
24	24	3	1	1	Mandinga	1	1	2								
25	25	3	1	1	Mandinga	1	1	2								
26	26	5	1	1	Fula	1	1	2								

SPSS Syntax window

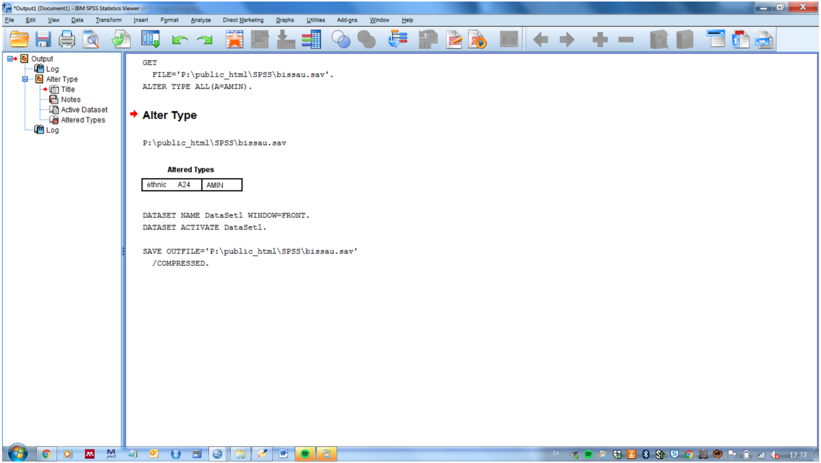


```
* Encoding: UTF-8.

DATASET ACTIVATE DataSet1.
GRAPH
  /HISTOGRAM=agemm.

```

SPSS Output window



*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Output
Log
Alter Type
Title
Notes
Active Dataset
Altered Types
Log
17:37

```
GET
  FILE='P:\public_html\SPSS\bissau.sav'.
ALTER TYPE ALL(A=AMIN).
```

Alter Type

P:\public_html\SPSS\bissau.sav

Altered Types

ethnic	A24	AMIN

```
DATASET NAME DataSet1 WINDOW=FRONT.
DATASET ACTIVATE DataSet1.

SAVE OUTFILE='P:\public_html\SPSS\bissau.sav'
  /COMPRESSED.
```

Output window

- Has a panel on the left that is a 'table of contents' of all your results.
- Free text edit is available in the output window

Data sets

Observations × variables :

sex	age	weight	name
1	8	25	John
2	5	17	Anna
2	13	48	Maria
⋮	⋮	⋮	⋮

- Variable names: sex age weight
- Variable types: Character or Numeric
  - 'Scale'
  - 'Ordinal'
  - 'Nominal'

The Juul data set

Serum IGF-I (Insulin-like Growth Factor) reference data set

Age	N	Source
0-5	44	Circumcision, hernia operation
5-20	833	4 schools in the Copenhagen area
20+	153	Hospital staff

Anders Juul et al., Dep. GR, Rigshosp.

AGE	age
MENARCHE	1st menstrual period occurred (1/2, 2 for yes)
SEXNR	1 for boys, 2 for girls
SIGF1	Serum IGF-I
TANNER	Puberty stage (1-5)
TESTVOL	Testicular volume
WEIGHT	weight

http://publicifsv.sund.ku.dk/~kach/SPSS/juul2.sav

The Guinea-Bissau data set

Data set called bissau.sav from rural Guinea-Bissau, West-Africa: 5273 children visited when being less than 7 months of age and followed for approximately six months. Registration of vaccination status, weight, etc at visit and deaths registered during follow-up.

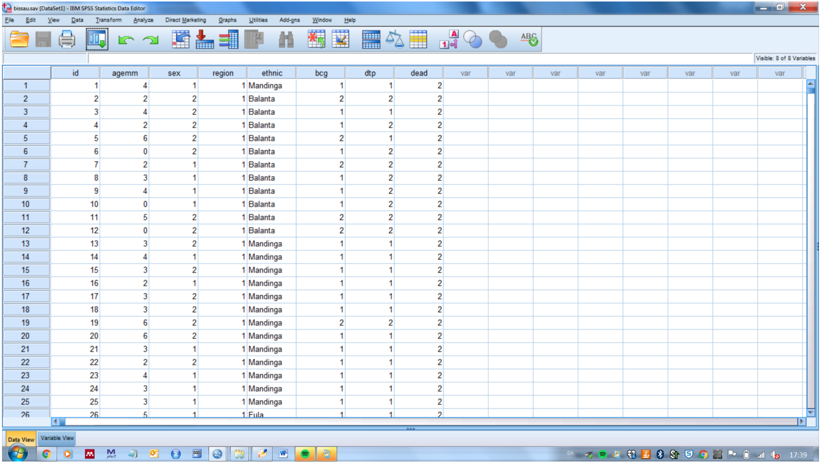
Data in SPSS

Two possibilites

- 'Data View'
- 'Variable View'

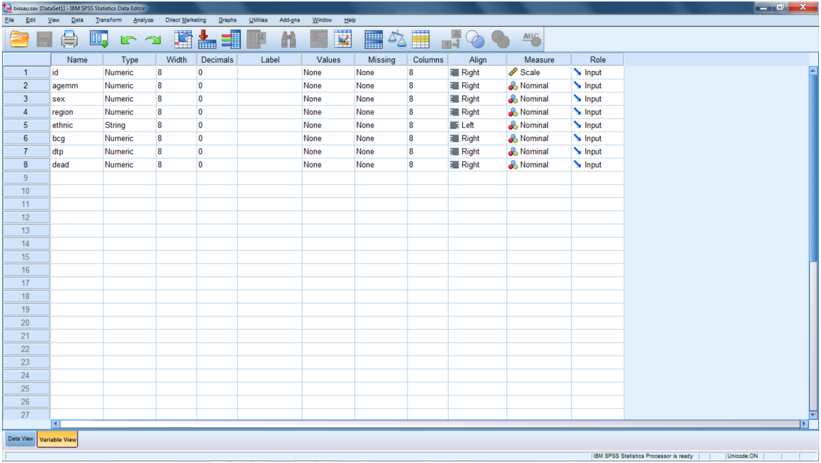
in 'Variable View' we can change the type between 'Scale', 'Ordinal' and 'Nominal' (in the column 'Measure').

'Data view'



	id	agemm	sex	region	ethnic	bcg	dtp	dead	var	var	var	var	var	var	var	var
1	1	4	1	1	Mandinga	1	1	2								
2	2	2	2	1	Balanta	2	2	2								
3	3	4	2	1	Balanta	1	1	2								
4	4	2	2	1	Balanta	1	2	2								
5	5	6	2	1	Balanta	2	1	2								
6	6	0	2	1	Balanta	1	2	2								
7	7	2	1	1	Balanta	2	2	2								
8	8	3	1	1	Balanta	1	2	2								
9	9	4	1	1	Balanta	1	2	2								
10	10	0	1	1	Balanta	1	2	2								
11	11	5	2	1	Balanta	2	2	2								
12	12	0	2	1	Balanta	2	2	2								
13	13	3	2	1	Mandinga	1	1	2								
14	14	4	1	1	Mandinga	1	1	2								
15	15	3	2	1	Mandinga	1	1	2								
16	16	2	1	1	Mandinga	1	1	2								
17	17	3	2	1	Mandinga	1	1	2								
18	18	3	2	1	Mandinga	1	1	2								
19	19	6	2	1	Mandinga	2	2	2								
20	20	6	2	1	Mandinga	1	1	2								
21	21	3	1	1	Mandinga	1	1	2								
22	22	2	2	1	Mandinga	1	1	2								
23	23	4	1	1	Mandinga	1	1	2								
24	24	3	1	1	Mandinga	1	1	2								
25	25	3	1	1	Mandinga	1	1	2								
26	26	5	1	1	Fula	1	1	2								

Variable view



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	0		None	None	8	Right	Scale	Input
2	agemm	Numeric	8	0		None	None	8	Right	Nominal	Input
3	sex	Numeric	8	0		None	None	8	Right	Nominal	Input
4	region	Numeric	8	0		None	None	8	Right	Nominal	Input
5	ethnic	String	8	0		None	None	8	Left	Nominal	Input
6	bcg	Numeric	8	0		None	None	8	Right	Nominal	Input
7	dtp	Numeric	8	0		None	None	8	Right	Nominal	Input
8	dead	Numeric	8	0		None	None	8	Right	Nominal	Input
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											

Syntax

Write syntax and ”submit” it to SPSS. Standard free text editing. Two ways

- Write syntax
- Point-and-click and click 'Paste'. This generates syntax.

Open `bissau.sav`. Type

```
DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=agemm
  /ORDER=ANALYSIS.
```

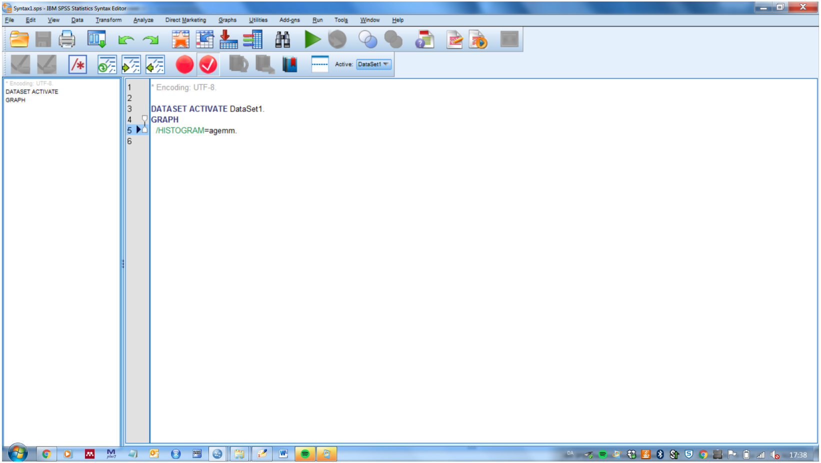
or

http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif1.gif.

Note: click on 'Paste'. Not on 'OK'.

Submit syntax and get output in the output window

Go to syntax window. Click on the green triangle.



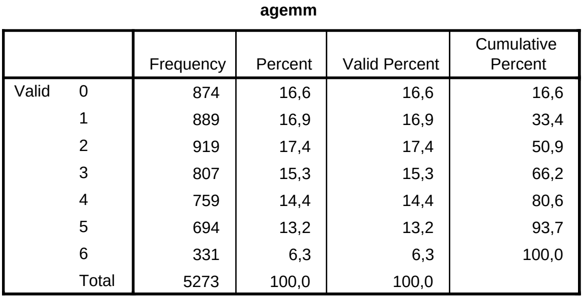
Active: DataSet1

```
* Encoding: UTF-8.

DATASET ACTIVATE DataSet1.
GRAPH
 /HISTOGRAM=agemm.
```

Submit syntax and get output in the output window

Output



agemm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	874	16,6	16,6	16,6
	1	889	16,9	16,9	33,4
	2	919	17,4	17,4	50,9
	3	807	15,3	15,3	66,2
	4	759	14,4	14,4	80,6
	5	694	13,2	13,2	93,7
	6	331	6,3	6,3	100,0
	Total	5273	100,0	100,0	

Exercise 1

Find the data set `bissau.sav` on the homepage, download it and open it in SPSS.

1. How many variables, how many observations ?
2. Are the numeric variables correctly classified into 'Scale', 'Ordinal' and 'Nominal' ?
3. Tabulate the variables `dead, bcg, dtp`

Data manipulation: recode

Recode age: point-and-click

http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif2.gif

click 'Paste':

```
DATASET ACTIVATE DataSet1.
RECODE agemm (0 thru 1=1) (2 thru 4=2) (5 thru 6=3) INTO newage.
VARIABLE LABELS  newage 'recoded'.
EXECUTE.
```

Syntax

Write syntax and "submit" it to SPSS. Standard free text editing.

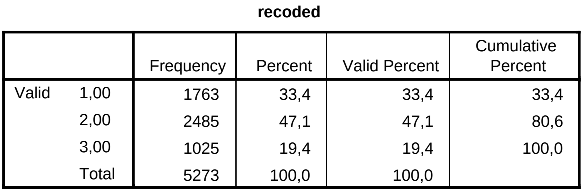
Two ways

- Write syntax
- Point-and-click and click 'Paste'. This generates syntax.

Open `bissau.sav`. Type

```
DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=newage
  /ORDER=ANALYSIS.
```

output



recoded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	1763	33,4	33,4	33,4
	2,00	2485	47,1	47,1	80,6
	3,00	1025	19,4	19,4	100,0
	Total	5273	100,0	100,0	

SPSS is a programming language

A SPSS program is a “recipe”: A series of instructions to be executed in a specified sequence.

- SPSS is not a spreadsheet. Output is output, and does not change automatically if data are changed
- Some rules and conventions are necessary for SPSS to be able to interpret its instructions

```
DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=agemm
  /ORDER=ANALYSIS.

DATASET ACTIVATE DataSet1.
RECODE agemm (0 thru 1=1) (2 thru 4=2) (5 thru 6=3) INTO newage.
VARIABLE LABELS  newage ’recoded’.
EXECUTE.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=newage
  /ORDER=ANALYSIS.
```

Syntax

- Roughly speaking, SPSS syntax consist of two kinds of steps
  - steps that define data sets by reading raw data, computing transformed variables, selecting cases, etc.
  - steps that contain standard procedures that operate *on* data sets.
- Normal arrangement of a SPSS program is to put data steps at the beginning, but they can occur intermixed

Comments are very helpful when you reread an old SPSS program. Two ways of making comments in SPSS programs:

```
* This is a comment and will continue to be a comment until the terminating period.
/* This is a comment and will continue to be a comment until the terminating asterisk-sl
```

Syntax with comments

```
/* recode the age variable */
DATASET ACTIVATE DataSet1.
RECODE agemm (0 thru 1=1) (2 thru 4=2) (5 thru 6=3) INTO newage.
VARIABLE LABELS  newage ’recoded’.
EXECUTE.
/* tabulate transformed variable */
DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=newage
  /ORDER=ANALYSIS.
```

Data manipulation: compute variable

In the data set juul2.sav we compute BMI

http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif3.gif

```
* DataSet2 is the juul data set.

DATASET ACTIVATE DataSet2.
COMPUTE BMI=weight/(height/100) ** 2.
EXECUTE.
```

Descriptive statistics

Histograms

http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif4.gif

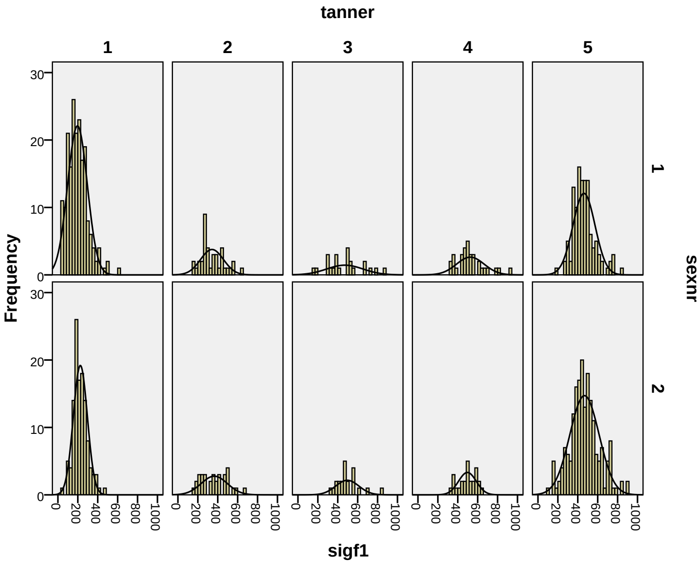
```
GRAPH
  /HISTOGRAM(NORMAL)=sigf1
  /PANEL COLVAR=tanner COLOP=CROSS ROWVAR=sexnr ROWOP=CROSS.
```

Boxplots

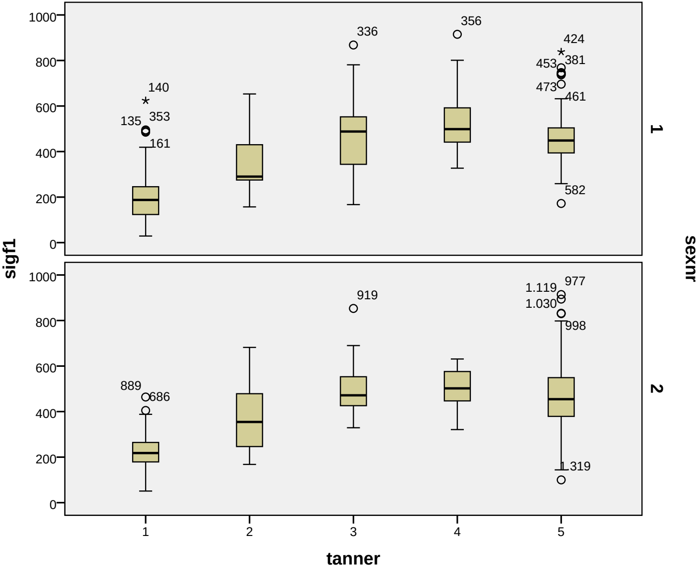
http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif5.gif

```
EXAMINE VARIABLES=sigf1 BY tanner
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /PANEL ROWVAR=sexnr ROWOP=CROSS.
```

Descriptive statistics



Descriptive statistics



Descriptive statistics

Compute means, medians and more:

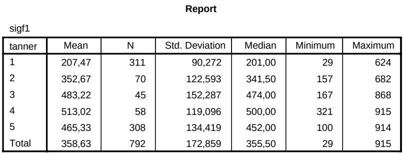
http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif6.gif

```
MEANS TABLES=sigf1 BY tanner
  /CELLS=MEAN COUNT STDDEV MEDIAN MIN MAX.
```

http://publicifsv.sund.ku.dk/~kach/SPSS/F1_gif7.gif

```
FREQUENCIES VARIABLES=sigf1
  /NTILES=4
  /STATISTICS=STDDEV MEAN MEDIAN
  /ORDER=ANALYSIS.
```

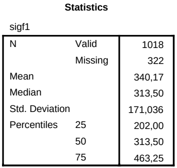
Descriptive statistics



Report

sigf1

tanner	Mean	N	Std. Deviation	Median	Minimum	Maximum
1	207,47	311	90,272	201,00	29	624
2	352,67	70	122,593	341,50	157	682
3	483,22	45	152,287	474,00	167	868
4	513,02	58	119,096	500,00	321	915
5	465,33	308	134,419	452,00	100	914
Total	358,63	792	172,859	355,50	29	915



Statistics

sigf1

N	Valid	1018
	Missing	322
Mean		340,17
Median		313,50
Std. Deviation		171,036
Percentiles	25	202,00
	50	313,50
	75	463,25

Exercise 2

Look at the `juul2.sav` data set

1. Make a new variable `log(SIGF1)`
2. Compare the distribution of this variable across genders and across Tanner groups.
3. Is the normal distribution a suitable description of the distribution ?