Survey Analysis Workshop

Block 3: Analysing two variables (and sometimes three)

Section 3.2: Three (or more) variables

Sub-section 3.2.1 Elaboration

© Copyright 2019 John F Hall

[New tutorial 12 May 2019: Draft only]

3.2.1.6 Earnings differences 2009: Extracting and saving selected variables

Data source:British Social Attitudes Survey, 20091(UKDS SN 6695)

Previous session: <u>3.2.1.5 Earnings differences 2009: Download and check file</u>

Page Contents:

- 2 <u>Research questions</u>
- 2 Elaboration model
- 3 Variables to be extracted

Intermediate stage 1:

- 6 <u>Making a copy of the file</u>
- 9 Locating the selected variables

Intermediate stage 2:

- 10 Extracting the subset of variables
- 12 Checking the data file
- 14 Discarding cases with no earnings
- 16 <u>Tidy up the data dictionary</u>
- 17 Save the reduced file
- 19 Checking the variable properties
- 26 Changing the display of variable attributes

¹ National Centre for Social Research. (2011). *British Social Attitudes Survey, 2009*. [data collection]. UK Data Service. SN: 6695, <u>http://doi.org/10.5255/UKDA-SN-6695-1</u>

Previous research questions:

1: Is there a difference between the earnings (from paid work) of men and women? See sessions: 2.3.1.6.2: Specimen answer for tasks 3 and 4 3.1.4.1 Income differences work-through

2: What other variables might account for differences in earnings?

See sessions: <u>3.1.4.2 Income differences - Build working file</u> <u>3.1.4.3 Income differences for test variables</u> 3.1.4.4 Income differences - Choose test variables and cutting points

3: What effect do they have by themselves? See session: <u>3.1.4.5 Income differences for derived test variables</u>

Current research question:

What happens to any differences in earnings between men and women when controlling for these other variables?

Elaboration² model

 $X \rightarrow Y \cdot T$ (the effect of X on Y controlling for T) where:

Y = Dependent variable X = Independent variable T = Test variable(s)

Y (Dependent)	X (Independent)	Tn (Test or control)		
Gross earnings from paid work	Sex	T1 T2 T3 T4 T5 T6 T7 T8	Working full time or part time Employee or self employed Economic sector Socio-economic grade of work Years of full-time education Qualifications Age Geographical region	

This session demonstrates how to:

- a) extract the variables needed to test the above elaboration model
- b) save them in a much smaller file.

² (See <u>Elaboration</u> (extract from Jim Ring's <u>Statistical Notes</u> specially written for this course)

Variables to be extracted

Dependent variable

[REarn] [if working] R's own gross or total earnings, before income tax+national insurance

Valid values range from 1 to 20 denoting grouped earnings per calendar month. Value -1 "Item not applicable" is declared as missing. Values 97 "Refused information", 98 "Don't know" and 99 "Refused" are not declared as missing.

[REarnQ] [if working] Quartile groups pf R's gross earnings

Derived variable in which **[REarn]** has been regrouped into four categories: this helps to keep contingency tables small and manageable. Values **7** "Refused information" and **8** "Don't know" are **not declared as missing**.

Independent variable [Rsex] "Sex of respondent"

[Rsex] "SEX OF respondent? :Q356" [sic]

[Rsex] is coded **1** "Male" **2** "Female" and has no missing values. Users may prefer to rename it as **[sex]** or **[gender]** according to their preferences.

Weighting

Some analyses may also require the weighting factor [Wtfactor] "Final BSA weights"

Test variables:

(Demographics)

[Rage] Age of respondent last birthday

There are two existing groupings for age, but it may be preferable to create new groupings

[Ragecat] Age of respondent (grouped into 7 categories" [RAgeCat2] Age of respondent (grouped into 6 categories

(Work-related)

If the respondent is working, several work-related variables are available:

[Remploye]	" Is R an employee or self-employed currently? "
[ROcsect2]	" Resp SEG <grouped> [pre-SOC2000]"</grouped>
[RNSocCI]	" Respondent : social class [pre-SOC2000] "

There is no single variable for working part-time (Under 30 hours a week) and working full- time (30 or more hours a week). There are two separate derived variables, one for employees and another for the self-employed:

[EjbHrCal] "Hours R works per week, including overtime [employee]. **[SJbHrCal]** "Hours R works per week, including overtime [self-employed].

These variables³ are mutually exclusive, so will have to be combined into a single variable.

³ [EjbHrCal] and [SJbHrCal] are spelled [ejbhrcai] and [sjbhrcai]

(Education)

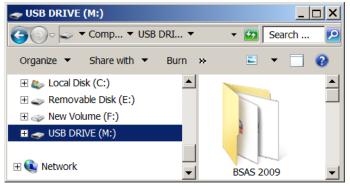
[Tea]	Age of R when completed continuous full-time education
[HEdqual2]	Highest educational qualification obtained (postgrad separate)

(Geography)

[GOR2]	Government Office Region (2003 version)
[Country]	Country of interview (England, Scotland or Wales)

In session <u>3.2.1.5 Earnings differences 2009: Download and check file</u> we created a new folder BSAS 2009 on USB Drive **M**:

Fig. 1:



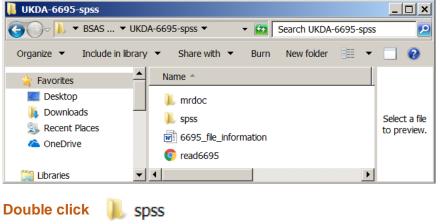
.. copied zip file **6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4** (downloaded from UKDS) from **Downloads** to folder **BSAS 2009** and extracted the contents.

Fig. 2:

Name 🔺	Date modified	Туре
👢 UKDA-6695-spss	24/07/2018 00:50	File Folder
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4.zip	09/10/2018 09:47	7-Zip.zip



Fig. 3:



UKDA-6695-spss

Fig. 4:

🔋 spss				
	9 ▼ UKDA-6695-spss ▼ spss ▼	🝷 🔯 Search	n spss	<u> </u>
Organize 🔻 Include in	library ▼ Share with ▼ Burn New folder	U File Shi	redder 🔠 🔻	
🔺 Favorites 🔺	Name 🔶	Date modified	Туре	
Desktop	🗼 spss12	08/02/2011 12:09	File Folder	Select a file to preview.
 Recent Places OneDrive 	<		Þ	preview.

Double click 🛛 👢 spss12

Fig. 5:		
📕 spss12		<u>_ 🗆 ×</u>
Gor ↓ ▼ spss ▼ spss12	✓ ☑ Search spss12	2
Organize 🔻 Include in librar	y ▼ Share with ▼ Burn New folder 🔠 ▼	
Favorites	Name ^	
E Desktop	bsa09	Colorto fla
🐌 Downloads	-	Select a file to preview.
😓 Recent Places		
🛆 OneDrive		

Double click ight bsa09

File is the active file.

Fig. 6:

*bsa09.	*bsa09.sav [DataSet1] - IBM SPSS Statistics Data Editor										
jle Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help											
😑 🗄 🖨 💷 🗠 🤉 🕌 🏭 📴 🛍 📟 🖾 📟 🐴 🚟 📲 🐼 🧠 🧠											
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Serial	Numeric	6	0	Serial Number:Q1	None	None	10	≡ Right	🛷 Scale	ゝ Input 🥤
2	SPoint	Numeric	3	0	Sample point:Q310	None	None	7	≡ Right	🛷 Scale	ゝ Input
3	StratID	Numeric	4	0	Stratification ID:Q311	None	None	9	疆 Right	🚓 Nominal	S Input
4	PopBand	Numeric	1	0	Population Density < Quartiles of GB> :Q316	{1, 0-2.789	None	9	疆 Right	🚓 Nominal	S Input
5	GOR2	Numeric	2	0	Government office region 2003 version:Q320	{1, North E	None	5	疆 Right	🚓 Nominal	💊 Input 🔤
Data View	Data View Variable View										
							IBM SPS	SS Statistics F	Processor is ready	Unicode:O	N

The file contains 847 variables, far too many for this exercise.

The SPSS icon in the task bar will show ⁴

Fig. 7:

- *Syntax1 IBM SPSS Statistics Syntax Editor
- *Output1 [Document1] IBM SPSS Statistics Viewer
- *bsa09.sav [DataSet1] IBM SPSS Statistics Data Editor

Intermediate stage 1: Making a copy of the file

Never work on an original file! Make a copy of the file

If your task bar does not show:

Fig. 8:

*Syntax1 - IBM SPSS Statistics Syntax Editor

... you must open a new Syntax Editor with:

File >> New >> Syntax

Fig. 9:

🍓 *bsa09.sav [DataSet1] - IBM SPSS Statistics Data Editor								
<u>File</u> dit <u>V</u> iew <u>D</u> ata	Transform Analyze	Direct <u>M</u> arketing	g <u>G</u> raphs <u>U</u> tilitie:	s E <u>x</u> tensions	Window Help	p		
<u>N</u> ew Open	<u>۲</u>	Data		- A			ABG	
Import Data	•	Syntax	Label	Values	Missing	Columns	Align	Measure
	Ctrl+F4	Cutput	Serial Number	None	None	10	■ Right	🖋 Scale 🥤
	Ctrl+S	S <u>c</u> ript ▶	Sample point:	None	None	7	■ Right	🛷 Scale
S <u>a</u> ve As	04110	0	Stratification I	None	None	9	≡ Riaht	🐣 Nominal 💌
🖶 Save A <u>l</u> l Data								
Expor <u>t</u>	•			IBM CDCC	Statistics Proce	seor is roady	Unicode:C	
hark File Read Only				IDIVI SPS3	Statistics Proce	ssor is ready	Unicode.C	

Fig. 10: Blank Syntax Editor



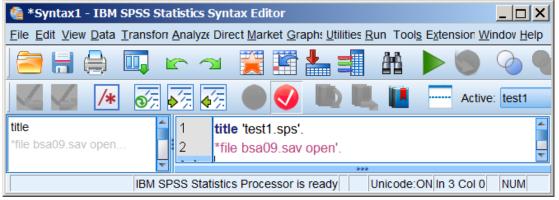
⁴ The author has set SPSS to open a new **Syntax Editor** at startup.

In the Syntax Editor write:

title 'test1.sps'. *file bsa09.sav open'.

[Good practice is to give each *.sps file a **title**] [Comment (ignored by SPSS) to remind you what you are doing]

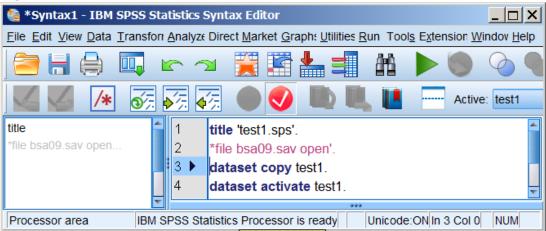
Fig. 11:



Now write:

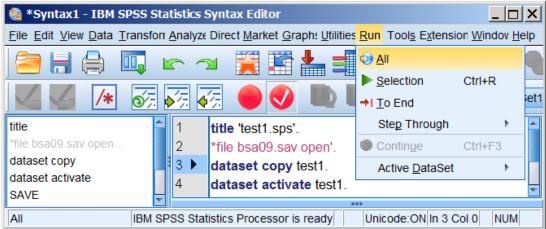
dataset copy test1. dataset activate test1. [Creates a copy of the fil] [Opens it as ***Untitled2 [test1]** which becomes the active file]

Fig. 12:



Either: Run >> All

Fig. 13:



Or: Highlight all 4 lines:

Fig. 14:

🍓 *Syntax1 - IBM SPSS Sta	atistics Syntax Editor							
<u>File</u> <u>E</u> dit <u>V</u> iew <u>D</u> ata <u>T</u> ransform	File Edit View Data Transforr Analyze Direct Market Graphe Utilities Run Tools Extension Windov Help							
) 🚍 🔚 🖨 🛄 II	r 🛥 🧮 🖺 📥 💷 👫 🕨 🕥 🌒							
	🟹 🎻 🔴 🍼 ា 🖳 🛄 🔤 Active: DataSet1							
title	1 title 'test1.sps'.							
*file bsa09.sav open	2 *file bsa09.sav open'.							
dataset copy	3 dataset copy test1.							
dataset activate	4 dataset activate test1.							
•								
IBM SP	SS Statistics Processor is ready Unicode:ON In 4 Col 23 NUM							

Press the green arrow: by to make a copy of the file and open it.

The copy opens as **Untitled2 [test1]** and becomes the active file.

Fig. 15: New Data Editor

🍓 *Untitled2 [test1] - IBM SPSS Statistics Data Editor									
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze Dire	ect <u>M</u> arketing	<u>G</u> raphs <u>U</u> tilitie				
🗁 🗄		5 2	iii 🕌 i	ч п					
	Name	Туре	Width	Decimals	Label				
1	Serial	Numeric	6	0	Serial Number				
2	SPoint	Numeric	3	0	Sample point:				

The task bar displays:

Fig. 16: SPSS icon in task bar

- 😫 *Syntax1 IBM SPSS Statistics Syntax Editor
- Tailor *Output1 [Document1] IBM SPSS Statistics Viewer
- *bsa09.sav [DataSet1] IBM SPSS Statistics Data Editor
- *Untitled2 [test1] IBM SPSS Statistics Data Editor

Without making any changes, save the file as test1.sav.

In the Syntax Editor add a line:

SAVE OUTFILE = 'M:\BSAS 2009\test1.sav'.

Fig. 16:				
🍓 *Syntax1 - IBM SPSS	Stat	stics Syntax Editor		
<u>File</u> <u>E</u> dit <u>V</u> iew <u>D</u> ata <u>T</u> ranst	fori <u>/</u>	nalyz∈ Direct <u>M</u> arket <u>G</u> rapl	h: <u>U</u> tilities <u>R</u> un Tool	l <u>s</u> E <u>x</u> tension <u>W</u> indov <u>H</u> elp
) 🚔 🖶 🖨 🗔		- 🗠 🧮 🖺		
🖌 🦉 🖊 🐼		ka 😽 🔴 🌖	II) II, II	Active: test1
title		1 title 'test1.sps'.		
*file bsa09.sav open		2 *file bsa09.sav op	pen'.	
dataset copy		3 dataset copy tes	st1.	
dataset activate		4 dataset activate		
SAVE		5	10011	
		6 SAVE OUTFILE		9\test1.sav' .
Processor area	SP:	S Statistics Processor is re	eady Unicode:	ON In 6 Col 42 NUM

Make sure the cursor is on the same line: press the green arrow:

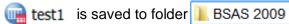


Fig. 17:

BSAS 2009								
G → L ▼ Computer ▼ USB DRIVE (M:) ▼ BSAS 2009 ▼ Search BSAS 2009 Search BSAS 2009								
Organize 🔻 Share with 🔻 Bu	Organize ▼ Share with ▼ Burn New folder							
		Name 🔺	Date modified	Туре	Size 🔺			
Computer Local Disk (C:)		🐌 UKDA-6695-spss	24/07/2018 00:50	File Folder		Select a file		
Wew Volume (F:)		6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB	to preview.		
🥪 USB DRIVE (M:)		🕞 test1	12/05/2019 11:14	SPSS Statis	3,686 KB			

UKDA-6695-spss	24/07/2018 00:50	File Folder	
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB
🕞 test1	12/05/2019 11:14	SPSS Statis	3,686 KB

Locating the selected variables

The variables needed are:

Dependent:	[REarn] [REarnQ]
Independent:	[Rsex]
Test: Age: Work: Education: Geography: Other:	[Rage] [Ragecat] [RAgeCat2] [Remploye] [EjbHrCal] [SJbHrCal] [rocsect2] [rnsoccl] [Tea] [hedqual2] [GOR2] [country] [wtfactor]

How do we extract them and save them to a smaller file?

Double click on und test1

Fia. 18:

_	10										
te 🔓	test1.sav [dataSet2] - IBM SPSS Statistics Data Editor										
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	Analyze Dire	ect <u>M</u> arketing	Graphs Utilities	s E <u>x</u> tensions	Window <u>H</u> elp	p		
			<u></u>		м	#1 💹 🔛		🛄 📕		ABG	
		Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1		Serial	Numeric	6	0	Serial Number	None	None	10	■ Right	🖋 Scale 🛛 🖆
2	2	SPoint	Numeric	3	0	Sample point:	None	None	7	≡ Right	🖋 Scale
3	1	StratID	Numeric	4	0	Stratification I	None	None	9	≡ Right	🗞 Nominal 🚽
		4							;		۱.
Data	View	Variable View									
							IBM SPSS	Statistics Proce	ssor is ready	Unicode:C	N N

Intermediate stage 2: Extracting the subset of variables

To extract only the 17 variables needed for this exercise and save them in a new file test2.sav

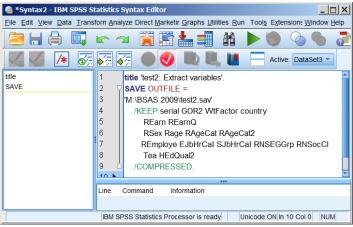
Open a new Syntax Editor

Fig. 19:	
🍓 *Syntax2 - IBM SPSS Statistics Syntax Editor	
<u>File Edit View Data Transfori Analyze Direct Market Graph: Utilities Run Tools Extension W</u>	ndov <u>H</u> elp
) 😑 🖶 🖨 📖 🖛 🛥 🧮 🖀 📥 💷 👫 🕨 🍥	
🛛 🜌 🌽 😹 🏹 🌑 🔍 🖿 Active	: DataSet3
IBM SPSS Statistics Processor is ready Unicode:ON In 1 Col 0	NUM

. . and write the following:

```
title 'test2: Extract variables'.
SAVE OUTFILE =
'M:\BSAS 2009\test2.sav'
  /KEEP serial GOR2 WtFactor country
    REarn REarnQ
    RSex Rage RAgeCat RAgeCat2
    REmploye EjbHrCal SJbHrCal RNSEGGrp RNSocCl
    Tea HEdQual2
  /COMPRESSED.
```

Fig. 20:



Make sure the cursor is somewhere inside the block of syntax and press the green arrow



File 🙀 test2 will be saved to folder 📗 BSAS 2009

Fig. 21:								
🔋 BSAS 2009								
😋 🕞 🗸 🔻 Computer 🕶								
Organize 🔻 Share with 🔻	Burn	New folder		U File	e Shredder	• 🗌 😧		
		Name *	Date modified	Туре	Size	4		
Computer Local Disk (C:)		🗼 UKDA-6695-spss	24/07/2018 00:50	File Folder				
New Volume (F:)		6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB	Select a file		
🧼 USB DRIVE (M:)		in test1	12/05/2019 11:14	SPSS Statis	3,686 KB —	to preview.		
🔃 Network		est2	12/05/2019 11:14	SPSS Statis	120 KB	•		
- Hetholk	-				▶			

Fig. 22:

👢 UKDA-6695-spss	24/07/2018 00:50	File Folder	
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB
🕞 test1	12/05/2019 11:14	SPSS Statis	3,686 KB
🕞 test2	12/05/2019 11:14	SPSS Statis	120 KB

Using **/KEEP** has reduced the file size from 3,656 kb to 120 kb.

Double click on united test2

File **test2** is now the active file.

Fig. 23:

	t 2.s av Edit	v [DataSet6] - View Data			ditor ect Marketing	Graphs Utilitie:	s Extensions	Window Hel	n			
										ABG		
		Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1		Serial	Numeric	6	0	Serial Number	None	None	10	Right	🛷 Scale	ゝ Input
2		GOR2	Numeric	2	0	Government of	{1, North E	None	5	■ Right	\delta Nominal	ゝ Input
3		WtFactor	Numeric	10	4	final BSA weig	None	None	12	■ Right	🗞 Nominal	ゝ Input
4		Country	Numeric	1	0	Country of inte	{1, England	None	9	■ Right	🗞 Nominal	💊 Input
5		REarn	Numeric	2	0	R's own gross	{-1, Skip, n	None	7	■ Right	🗞 Nominal	ゝ Input
6		REarnQ	Numeric	2	0	respondent ea	{-1, Skip,n	None	8	■ Right	🗞 Nominal	💊 Input
7		RSex	Numeric	1	0	SEX OF respo	{1, Male}	None	5	■ Right	💑 Nominal	ゝ Input
8		RAge	Numeric	2	0	What was R's	{97, 97+}	None	6	■ Right	🛷 Scale	💊 Input
9		RAgeCat	Numeric	2	0	Age of respon	{1, 18-24}	None	9	■ Right	💑 Nominal	ゝ Input
10		RAgeCat2	Numeric	2	0	Age of respon	{1, 18-24}	None	10	■ Right	ቆ Nominal	💊 Input
11		REmploye	Numeric	2	0	Is R an emplo	{-3, Skip, n	-1	9	■ Right	💑 Nominal	ゝ Input
12		EJbHrCal	Numeric	2	0	Hours R works	{-1, Not em	LO1	10	■ Right	🚓 Nominal	ゝ Input
13		SJbHrCal	Numeric	2	0	Hours R works	{-1, Not sel	LO1	10	■ Right	💑 Nominal	ゝ Input
14		RNSEGGrp	Numeric	4	0	Resp:SEG <gr< td=""><td>{-1, Never</td><td>LO1</td><td>10</td><td>■ Right</td><td>🖋 Scale</td><td>ゝ Input</td></gr<>	{-1, Never	LO1	10	■ Right	🖋 Scale	ゝ Input
15		RNSocCl	Numeric	2	0	Respondent :	{-1, Never	LO1	8	■ Right	🗞 Nominal	ゝ Input
16	·	Теа	Numeric	2	0	R how old whe	{1, 15 or u	None	5	■ Right	🗞 Nominal	ゝ Input
17		HEdQual2	Numeric	2	0	Highest educat	{1, Postgra	None	10	■ Right	🗞 Nominal	ゝ Input
18												
Data \	_	1 Variable View										
								IBM SP	SS Statistics	Processor is ready	/ Unicode:O	N

Using /KEEP has reduced the number of variables from 847 to 17

Checking the data file

Open a new Syntax Editor

Fig. 24:

1	Syntax	3 - IB	M SPSS	Statistics S	yntax Edi	tor							
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct Marketing	<u>G</u> raphs	<u>U</u> tilities	<u>R</u> un	Tool <u>s</u>	Extensions	<u>W</u> indow	<u>H</u> elp
				5 2	1	🛐 🏪 🗐	A A) @		a	2	
		/>	k 0	a 🗸 🎸		🥑 🐚 🛢		Ac	tive: Da	taSet3	-		
				1 ► 2 2									

						IBM SPSS St	atistics Pr	rocessor is	ready	Un	icode:ON In	1 Col 0	MUM

Write in:

title 'test3: Check data file'. display dictionary. display labels.

Fig. 25: 🔮 *Syntax3 - IBM SPSS Statistics Syntax Editor <u>File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Run Tools Extensions Window Help</u> R ÅÅ. G 5 $\overline{\mathbf{n}}$ a -**₫**, **↓**, **↓**, Active: DataSet3 - \checkmark 1 title 'test3: Check data file'. 2 display dictionary. Ŧ 3 🕨 display labels. IBM SPSS Statistics Processor is ready Unicode:ON In 3 Col 15 NUM

Run >> <u>A</u>ll

Fig. 26:						
<u>R</u> un	Run Tools Extensions Window					
😔 <u>A</u> ll						
▶ Selection Ctrl+R						

Table 1: Data dictionary for selected variables (condensed)

Variable	Line	Label	Level	Format	Missing
Serial	1	Serial Number:Q1	Scale	F6	
GOR2	2	Government office region 2003 version:Q320	Nominal	F2	
WtFactor	3		Nominal	F10.4	
Country	4	Country of interview England, Scotland or Wales? :Q333	Nominal	F1	
REarn	5	R's own gross or total earnings, before income tax+national	Nominal	F2	
		insurance?:Q1376			
REarnQ	6	respondent earnings quartiles (dv):Q1377	Nominal	F2	
RSex	7		Nominal	F1	
RAge	8		Scale	F2	
RAgeCat	9	Age of respondent(grouped)<7 category> dv:Q446	Nominal	F2	
RAgeCat2	10	Age of respondent(grouped)<6 category> dv:Q447	Nominal	F2	
REmploye	11	Is R an employee or self-employed currently? dv:Q987	Nominal	F2	-1
EJbHrCal	12	······································	Nominal	F2	Lo thru -1
SJbHrCal	13	Hours R works per week, including overtime [self-employed].	Nominal	F2	Lo thru -1
		DV:Q1010			
RNSEGGrp	14	Resp:SEG <grouped>[pre-SOC2000]best est.Q</grouped>	Scale	F4	Lo thru -1
RNSocCl	15	Respondent : social class[pre-SOC2000]best estimate dv:Q1022	Nominal	F2	Lo thru -1
Теа	16	R how old when completed continuous fulltime	Nominal	F2	
		education?[compressed]dv:Q1196			
HEdQual2	17	Highest educational qual obtained (postgrad separate) - dv:Q1262	Nominal	F2	

[NB: The above table has been heavily edited to delete columns Role, Width, Align and to combine Print Format and Write Format in a single column Format.]

Table 2: Variable labels

		Variable Labels
Variable	Position	Label
Serial	1	Serial Number:Q1
GOR2	2	Government office region 2003 version:Q320
WtFactor	3	final BSA weights
Country	4	Country of interview England, Scotland or Wales? :Q333
REarn	5	R's own gross or total earnings, before income tax+national
		insurance?:Q1376
REarnQ	6	respondent earnings quartiles (dv):Q1377
RSex	7	SEX OF respondent? :Q356
RAge	8	What was R's age last birthday? :Q357
RAgeCat	9	Age of respondent(grouped)<7 category> dv:Q446
RAgeCat2	10	Age of respondent(grouped)<6 category> dv:Q447
REmploye	11	Is R an employee or self-employed currently? dv:Q987
EJbHrCal		
SJbHrCal	13	Hours R works per week, including overtime [self-employed]. DV:Q1010
RNSEGGrp	14	Resp:SEG <grouped>[pre-SOC2000]best est.Q*</grouped>
RNSocCl	15	Respondent : social class[pre-SOC2000]best estimate dv:Q1022
Теа	16	R how old when completed continuous fulltime
		education?[compressed]dv:Q1196
HEdQual2	17	Highest educational qual obtained (postgrad separate) - dv:Q1262

Variables in the working file

The **question numbers** at the end of each label are useful for checking against the original questionnaire, but are effectively superfluous and could be deleted: so could **[compressed]** and **dv:** (derived variable).

* There is no question number in the variable label for [RNSEGGrp]

Across all waves of the survey, the same question or item always has the same variable name, but the question numbers will be different in each year and consequently be incompatible between years. Similarly, variable types, formats, widths, labels and value labels also vary between years. When combining data for different years, all variable properties will need to be standardised.

Further file modifications

To show screenshots of the **Syntax Editor** on the following pages would be repetitive and tedious. From this point only the **commands** and **output** are shown.

Discarding cases with no earnings

We need to discard cases who have no earnings from paid work.

frequencies rearnq.

Table 3: Respondent earnings quartiles

REarnQ respondent earnings quartiles (dv):Q1377

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	-1 Skip,not in paid work	1558	45.5	45.5	45.5
	1 less than 11999	395	11.5	11.5	57.1
	2 12000- 19999	414	12.1	12.1	69.2
	3 20000- 31999	467	13.7	13.7	82.8
	4 32000 or more	413	12.1	12.1	94.9
	7 Refused information	147	4.3	4.3	99.2
	8 Don't know	27	0.8	0.8	100.0
	Total	3421	100.0	100.0	

[NB: The £ sign was not available in SPSS 12]

Values -1 " Skip,not in paid work" 7 " Refused information" and 8 " Don't know" need to be treated as missing.

missing values REarn (-1 97 thru 99) REarnQ (-1, 7, 8) **frequencies** rearnq.

KEaning Respondent earnings quarties								
		Frequency	Percent	Valid Percent	Cumulative Percent			
Valid	1 less than 11999	395	11.5	23.4	23.4			
	2 12000- 19999	414	12.1	24.5	47.9			
	3 20000- 31999	467	13.7	27.6	75.5			
	4 32000 or more	413	12.1	24.5	100.0			
	Total	1689	49.4	100.0				
Missing	-1 Skip,not in paid work	1558	45.5					
	7 Refused information	147	4.3					
	8 Don't know	27	0.8					
	Total	1732	50.6					
Total		3421	100.0					

 Table 4: Respondent earnings quartiles

 REarnQ Respondent earnings quartiles

To select only those cases with non-missing values for [REarnQ]:

select if (not (missing (rearnq))). **frequencies** rearnq.

Table 5: Respondent earnings quartiles (non-earners excluded)

	Keanila								
			_		Cumulative				
		Frequency	Percent	Valid Percent	Percent				
Valid	less than 11999	395	23.4	23.4	23.4				
	12000- 19999	414	24.5	24.5	47.9				
	20000- 31999	467	27.6	27.6	75.5				
	32000 or more	413	24.5	24.5	100.0				
	Total	1689	100.0	100.0					

The number of cases in the file has been reduced from 3421 to 1689.

There is no variable denoting the year of survey:

compute year = 2009. **variable labels** year 'Year of survey'.

Variable [year] is appended to the file:

Fig. 27:

	Name	Туре	Width	Decimals	Label
16	Теа	Numeric	2	0	R how old whe
17	HEdQual2	Numeric	2	0	Highest educat
18	year	Numeric	8	0	Year of survey
	-				

Tidy up the data dictionary

The data dictionary now needs editing to add or correct measurement levels, missing values and variable and value labels.

Levels of measurement:

variable level

RSEx GOR2 country (nominal) REarn REarnQ RAgeCat RAgeCat2 REmploye EJbHrCal SJbHrCal RNSEGGrp to HEdQual2 (ordinal) year WtFactor (scale).

Formats

formats

year (f4.0) serial (f10.0) REarn GOR2 (f3.0) REarnQ RAgeCat RAgeCat2 REmploye EJbHrCal SJbHrCal RNSEGGrp RNSocCl Tea HEdQual2 country (f2.0) wtfactor (f6.4).

Variable labels

Variable labels can be edited manually in the Data Editor, but it's much safer to use syntax:

variable labels

inable labe	
year	'Year of survey'
Serial	'Serial Number'
GOR2	' Government office region: 2003 version '
WtFactor	'Final BSA weights '
Country	' Country of interview: England, Scotland or Wales '
REarn	" R's own gross earnings, before Income Tax and National Insurance "
REarnQ	'Respondent earnings quartiles '
RSex	'Sex of respondent '
RAge	"What was R's age last birthday? "
RAgeCat	' Age of respondent (grouped: 7 categories) '
RAgeCat2	Age of respondent (grouped: 6 categories)
REmploye	' Is R an employee or self-employed currently? '
EjbHrCal	'Hours R works per week, including overtime [employee].'
SJbHrCal	'Hours R works per week, including overtime [self-employed].'
RNSEGGrp	o' Resp:SEG <grouped> [pre-SOC2000] '</grouped>
RNSocCl	' Respondent : social class [pre-SOC2000] '
Теа	'R how old when completed continuous full time education '
HEdQual2	' Highest educational qualification obtained '.

[NB: Double quotes are needed for [REarn] and [Rage] because the labels contain single quotes].

Missing values

No missing values are declared for [REarn] and [REarnQ]

Fig. 28:

618	REarn	Numeric	2	0	R's own gross	{-98, Don't	None	•
619	REarnQ	Numeric	2	0	respondent ea	{-1, Skip,n	None	1

These two variables have already been checked (see frequency tables on page 13 above and on pages 13 to 16 of <u>3.2.1.5 Earnings differences 2009: Download and check file</u>)

Variable **[REmploye]** has two negative values declared as missing: **-3** " Skip, not now in paid work" and **-1** " Skip, never had a job".

It also has two other values which we need to treat as missing: 8 "Don't know" and 9 "Refusal".

SPSS only allows 3 missing values eg (7, 8, 9) two of which can be lower and upper limits of a value range eg (-1, 97 thru 99). Where there are more than 3 values to be treated as missing, one method is to recode all positive missing values to negative and specify missing values as (lo thru -1). If you do this you must use the ADD VALUE LABELS command, otherwise the existing labels will be deleted.

recode REmploye (8 = -8)(9 = -9). **add value labels** REmploye -8 " Don't know" -9 " Refusal".

missing values

REarn (-1 97 thru 99) REarnQ (-1, 7, 8) RAgeCat (8) RAgeCat2 (9) REmploye (-9 thru -1) EJbHrCal SJbHrCal (-1, 5 thru 9) RNSEGGrp RNSocCl (-1, 7, 8) Tea (6 thru 99) HEdQual2 (7, 9).

Variable [year] needs to be at the beginning of the file.

To make files easier to navigate, experienced researchers often re-order the variables:

[Dependent >> Independent >> Test]

. . and arrange the test variables in blocks defined by their source topic, in this case:

Age >> Work >> Education >> Geography

Save the reduced file

SAVE OUTFILE = 'M:\BSAS 2009\test4.sav' /KEEP year serial Rearngrp REarn REarnQ RSex Rage RAgeCat RAgeCat2 REmploye EjbHrCal SJbHrCal RNSEGGrp RNSocCl Tea HEdQual2 GOR2 Country WtFactor /COMPRESSED.

File **[III]** test3 will be saved to folder

BSAS 2009 on USB Drive M:



BSAS 2009						_ 🗆 >
🌀 🕞 🗏 🔻 Computer 🔻	👻 🚺 Se	arch BSAS 2009				
Organize 👻 Share with	Burn	New folder		U File	e Shredder 🔠 👻	
		Name *	Date modified	Туре	Size	
🦓 Homegroup		🐌 UKDA-6695-spss	24/07/2018 00:5	50 File Folder		
💐 Computer		6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:4	2 ZIP File	4,229 KB	Select a fi
bocal Disk (C:)		🕞 test1	12/05/2019 11:1	4 SPSS Statis	3,686 KB	to preview
New Volume (F:)		🕞 test2	12/05/2019 11:1	4 SPSS Statis	120 KB	
🥪 USB DRIVE (M:)	•	test3 Type: SPSS Statistics Size: 119 KB	Data Document 17:2	27 SPSS Statis	63 KB	

Fig. 29:

UKDA-6695-spss	24/07/2018 00:50	File Folder	
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB
uest1	12/05/2019 11:14	SPSS Statis	3,686 KB
uest2	12/05/2019 11:14	SPSS Statis	120 KB
🕞 test3	12/05/2019 17:27	SPSS Statis	63 KB

🔚 test3.sav [DataSet:

Checking the variable properties

Double click on in test3

which becomes the active file

Fig. 30: 🔩 test3.sav [DataSet11] - IBM SPSS Statistics Data Editor <u>_ D ×</u> <u>E</u>dit <u>V</u>iew <u>D</u>ata Direct Marketing <u>U</u>tilities Extensions Window Help Transform Analyze File Graphs 1 × K 2 h <u>
</u> 9 446 Name Width Decimals Label Values Missing Columns Measure Role Туре Align 1 vear Numeric 4 0 Year of survey None None 10 ■ Right 🖋 Scale ゝ Input 2 Serial Number 🖋 Scale Serial Numeric 10 0 None None 10 Right ゝ Input 3 REarn 2 0 97 - 99 - 1 7 ≡ Right Ordinal Numeric R's own gross... {-1, Skip, n.. 🔪 Input 4 REarnQ Numeric 2 0 Respondent e... {-1, Skip,n. -1, 7, 8 8 Right 📶 Ordinal ゝ Input 5 RSex Numeric 2 0 Sex of respon... {1, Male} None 5 ≣ Right 🚴 Nominal 🔪 Input 6 RAge Numeric 2 0 What was R's... {97, 97+}. None 6 ■ Right 🖋 Scale ゝ Input 7 RAgeCat Numeric 2 0 Age of respon... {1, 18-24} ... 8 9 ■ Right 🚽 Ordinal ゝ Input 2 RAgeCat2 Numeric 0 Age of respon... {1, 18-24}... 9 10 ■ Right J Ordinal 🔪 Input 8 9 REmploye Numeric 2 0 Is R an emplo... {-9, Refusa... -9 - -1 9 ■ Right 📶 Ordinal ゝ Input 10 EJbHrCal 2 0 Hours R works... {-1, Not em... 5 - 9, -1 10 ■ Right 📕 Ordinal Numeric > Input 11 SJbHrCal Numeric 2 0 Hours R works... {-1, Not sel... 5 - 9, -1 10 疆 Right Ordinal ゝ Input RNSEGGrp 2 Resp:SEG <g... {-1, Never ... -1, 7, 8 ■ Right 📕 Ordinal 12 Numeric 0 10 ゝ Input RNSocCl 2 0 Right 13 Respondent : ... {-1, Never ... 8 Ordinal Numeric -178 > Input 6 - 99 14 Теа Numeric 2 0 R how old wh... {1, 15 or u... 5 ■ Right 📕 Ordinal ゝ Input 15 HEdQual2 Numeric 2 0 Highest educ... {1, Postgra... 7, 9 10 理 Right 📶 Ordinal 🔪 Input 16 GOR2 Numeric 2 0 Government o... {1, North E... None 5 ■ Right 🚴 Nominal ゝ Input 17 2 0 Country of int... {1, England... 9 ≡ Right \delta Nominal Country Numeric None ゝ Input 12 18 6 Final BSA wei... None Scale ゝ Input WtFactor Numeric 4 None ■ Right 19 Data View Variable View IBM SPSS Statistics Processor is ready Unicode:ON

Correct measurement levels have now been assigned.

The variable labels are also clearer:

Fig. 31:

a test3	3.sav [DataSet10] - IBM SPSS S	tatistics Data	a Editor		_ 🗆 >		
ile <u>E</u> o	dit <u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze D	irect Marketing	<u>G</u> raphs <u>U</u> tilities E <u>x</u> tensions <u>W</u> indow <u>H</u> elp			
					11 📰 📰 🚍 🐴 📰 📑 🐼 畅 🦇			
	Name	Туре	Width	Decimals	Label			
1	year	Numeric	4	0	Year of survey			
2	Serial	Numeric	10	0	Serial Number			
3	REarn	Numeric	2	0	R's own gross earnings, before Income Tax and National Insurance			
4	REarnQ	Numeric	2	0	Respondent earnings quartiles			
5	RSex	Numeric	2	0	Sex of respondent			
6	RAge	Numeric	2	0	What was R's age last birthday?			
7	RAgeCat	Numeric	2	0	Age of respondent (grouped: 7 categories)			
8	RAgeCat2	Numeric	2	0	Age of respondent (grouped: 6 categories)			
9	REmploye	Numeric	2	0	Is R an employee or self-employed currently?			
10	EJbHrCal	Numeric	2	0	Hours R works per week, including overtime [employee].			
11	SJbHrCal	Numeric	2	0	Hours R works per week, including overtime [self-employed].			
12	RNSEGGrp	Numeric	2	0	Resp:SEG <grouped> [pre-SOC2000]</grouped>			
13	RNSocCl	Numeric	2	0	Respondent : social class [pre-SOC2000]			
14	Tea	Numeric	2	0	R how old when completed continuous full time education			
15	HEdQual2	Numeric	2	0	Highest educational qualification obtained			
16	GOR2	Numeric	2	0	Government office region: 2003 version			
17	Country	Numeric	2	0	Country of interview: England, Scotland or Wales			
18	WtFactor	Numeric	6	4	Final BSA weights			
10	4							
Data Vie	ew Variable Viev	v						
Data Vi	variable viev	v						
IBM SPSS Statistics Processor is ready Unicode:ON								

display labels.

Table 6: \	Variable	labels
------------	----------	--------

Variable Labels

Variable	Position	Label
year	1	Year of survey
Serial	2	Serial Number
REarn	3	R's own gross earnings (before Income Tax and National Insurance)
REarnQ	4	Respondent earnings quartiles
RSex	5	Sex of respondent
RAge	6	What was R's age last birthday?
RAgeCat	7	Age of respondent (grouped: 7 categories)
RAgeCat2	8	Age of respondent (grouped: 6 categories)
REmploye	9	Is R an employee or self-employed currently?
RNSEGGrp	10	Resp:SEG <grouped> [pre-SOC2000]</grouped>
RNSocCl	11	Respondent : social class [pre-SOC2000]
Tea	12	R how old when completed continuous full time education
HEdQual2	13	Highest educational qualification obtained
GOR2	14	Government office region: 2003 version
Country	15	Country of interview: England, Scotland or Wales
WtFactor	16	Final BSA weights

Variables in the working file

Check variable properties

Check 1: Dependent variable

For the purposes of elaboration our dependent variable will be [REarnQ]

frequencies rearnq.

Та	h	ما	7	-
	D	e		•

REarnQ

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	less than 11999	395	23.4	23.4	23.4
	12000- 19999	414	24.5	24.5	47.9
	20000- 31999	467	27.6	27.6	75.5
	32000 or more	413	24.5	24.5	100.0
	Total	1689	100.0	100.0	

[NB: The £ sign was not available in SPSS 12]

Check 2: Independent variable

frequencies rsex.

Table 8:

	RSex								
					Cumulative				
		Frequency	Percent	Valid Percent	Percent				
Valid	Male	831	49.2	49.2	49.2				
	Female	858	50.8	50.8	100.0				
	Total	1689	100.0	100.0					

Check 3: Earnings of men and women

crosstabs rsex by rearnq /cells count row.

Table 9:

RSex * REarnQ Crosstabulation

				REar	nQ		Total
			less than	12000-	20000-	32000 or	
			11999	19999	31999	more	
RSex	Male	Count	102	186	247	296	831
		% within RSex	12.3%	22.4%	29.7%	35.6%	100.0%
	Female	Count	293	228	220	117	858
		% within RSex	34.1%	26.6%	25.6%	13.6%	100.0%
Total		Count	395	414	467	413	1689
		% within RSex	23.4%	24.5%	27.6%	24.5%	100.0%

Check 4: Test variables

Person related

frequencies RageCat RAgeCat2.

Table	Table 10: RAgeCat Age of respondent (grouped: 7 categories)								
		Frequency	Percent	Valid Percent	Cumulative Percent				
Valid	1 18-24	109	6.5	6.5	6.5				
	2 25-34	357	21.1	21.1	27.6				
	3 35-44	489	29.0	29.0	56.5				
	4 45-54	425	25.2	25.2	81.7				
	5 55-59	174	10.3	10.3	92.0				
	6 60-64	94	5.6	5.6	97.6				
	7 65+	41	2.4	2.4	100.0				
	Total	1689	100.0	100.0					

Table11:

RAgeCat2 Age of respondent (grouped: 6categories)

	Ingeea			l'eupeur ceate	jeniee/
					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	1 18-24	109	6.5	6.5	6.5
	2 25-34	357	21.1	21.1	27.6
	3 35-44	489	29.0	29.0	56.5
	4 45-54	425	25.2	25.2	81.7
	5 55-64	268	15.9	15.9	97.6
	6 65-97	41	2.4	2.4	100.0
	Total	1689	100.0	100.0	

We shall need to group **Rage** into fewer categories, but using different cutting points.

Work related

frequencies REmploye EJbHrCai SJbHrCai RNSEGGrp RNSocCI.

Table 12:

REmploye Is R an employee or self-employed currently?									
		Frequency	Percent	Valid Percent	Cumulative Percent				
Valid	1 Emp	1467	86.9	86.9	86.9				
	2 SEmp	222	13.1	13.1	100.0				
	Total	1689	100.0	100.0					

Table 13:

EJbHrCal Hours R works per week, including overtime [employee].

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	76	4.5	5.2	5.2
	2 16-23 hours a week	163	9.7	11.2	16.4
	3 24-29 hours a week	84	5.0	5.8	22.1
	4 30 or more hours a week	1137	67.3	77.9	100.0
	Total	1460	86.4	100.0	
Missing	-1 Not employee	222	13.1		
-	5 Varies too much to say	5	0.3		
	8 Don't know	2	0.1		
	Total	229	13.6		
Total		1689	100.0		

Table 14:

SJbHrCal Hours R works per week, including overtime [self-employed].

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 10-15 hours a week	13	0.8	6.0	6.0
	2 16-23 hours a week	36	2.1	16.5	22.5
	3 24-29 hours a week	14	0.8	6.4	28.9
	4 30 or more hours a week	155	9.2	71.1	100.0
	Total	218	12.9	100.0	
Missing	 1 Not self-employed 	1467	86.9		
-	5 Varies too much to say	3	0.2		
	9 Refusal	1	0.1		
	Total	1471	87.1		
Total		1689	100.0		

Table 15:

RNSEGGrp Resp SEG <grouped> [pre-SOC2000]

		[p.		Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	1 Professional/employers/managers	359	21.3	21.4	21.4
	2 Intermediate non-manual	501	29.7	29.9	51.3
	3 Junior non-manual	227	13.4	13.5	64.8
	4 Supervisor/skilled manual	308	18.2	18.4	83.1
	5 Semi-skilled/personal services	239	14.2	14.2	97.4
	6 Unskilled manual	44	2.6	2.6	100.0
	Total	1678	99.3	100.0	
Missing	7 Armed forces	5	0.3		
	8 Inadequately described/not stated	6	0.4		
	Total	11	0.7		
Total		1689	100.0		

[Values 7 and 8 will be treated as missing. Only 11 cases, so hardly any effect on statistical analysis]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 I (SC=1)	108	6.4	6.4	6.4
	2 II (SC=2)	657	38.9	39.2	45.6
	3 III (non-manual) (SC=3)	343	20.3	20.4	66.0
	4 III (manual) (SC=4)	312	18.5	18.6	84.6
	5 IV (SC=5)	204	12.1	12.2	96.8
	6 V (SC=6)	54	3.2	3.2	100.0
	Total	1678	99.3	100.0	
Missing	7 Armed forces	5	0.3		
_	8 Insufficient information	6	0.4		
	Total	11	0.7		
Total		1689	100.0		

RNSocCL Respondent : social class [pre-SOC2000]

[Values 7 and 8 will be treated as missing. Only 11 cases, so hardly any effect on statistical analysis]

Education related

frequencies Tea HEdQual2.

Table 16:

Table 16:

Tea R how old when completed continuous full time education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 15 or under	908	26.5	27.2	27.2
	2 16	965	28.2	28.9	56.1
	3 17	305	8.9	9.1	65.2
	4 18	423	12.4	12.7	77.9
	5 19 or over	740	21.6	22.1	100.0
	Total	3341	97.7	100.0	
Missing	6 Still at school	12	0.4		
_	7 Still at college or university	57	1.7		
	97 Other answer (WRITE IN)	5	0.1		
	98 Don't know	3	0.1		
	99 Refusal	3	0.1		
	Total	80	2.3		
Total		3421	100.0		

[Treating values 6 and 7 as missing makes sample more homogenous and [Tea] smoothly ordinal.]

Table 17:

HEdQual2 Highest educational qualification obtained

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Postgraduate degree	174	5.1	5.6	5.6
	2 First degree	449	13.1	14.5	20.1
	3 Higher educ below degree	357	10.4	11.5	31.7
	4 A level or equiv	521	15.2	16.8	48.5
	5 O level or equiv	603	17.6	19.5	68.0
	6 CSE or equiv	244	7.1	7.9	75.9
	8 No qualification	746	21.8	24.1	100.0
	Total	3094	90.4	100.0	
Missing	7 Foreign or other	34	1.0		
_	9 DK/Refusal/NA	293	8.6		
	Total	327	9.6		
Total		3421	100.0		

[Treating values 7 and 9 as missing makes [hedqual2] smoothly ordinal.]

Geographic

frequencies GOR2 Country.

Table 18:

I able	18:				
	GOR2 Governm	nent office re	gion 2003 y	version:Q320	
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 North East	100	5.9	5.9	5.9
	2 North West	188	11.1	11.1	17.1
	3 Yorkshire and Humberside	136	8.1	8.1	25.1
	4 East Midlands	119	7.0	7.0	32.1
	5 West Midlands	164	9.7	9.7	41.9
	6 SW	156	9.2	9.2	51.1
	7 Eastern	183	10.8	10.8	61.9
	8 Inner London	80	4.7	4.7	66.7
	9 Outer London	115	6.8	6.8	73.5
	10 South East	228	13.5	13.5	87.0
	11 Wales	75	4.4	4.4	91.4
	12 Scotland	145	8.6	8.6	100.0
	Total	1689	100.0	100.0	

Table 19:

Country Country of interview England, Scotland or Wales? :Q333

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 England	1469	87.0	87.0	87.0
	2 Scotland	145	8.6	8.6	95.6
	3 Wales	75	4.4	4.4	100.0
	Total	1689	100.0	100.0	

Fig. 33:

BSAS 2009									
Image: Search BSAS 2009 ✓ Image: Search BSAS 2009									
Organize 👻 Share with 🗨	Burn	New folder		U File	Shredder 🔠 🦷	- 🗆 📀			
•		Name ^	Date modified	Туре	Size				
🦓 Homegroup		📙 UKDA-6695-spss	24/07/2018 00:50	File Folder					
💐 Computer		6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB				
tocal Disk (C:)		iest1	12/05/2019 11:14	SPSS Statis	3,686 KB	Select a fi to previev			
🥪 New Volume (F:)		iest2	12/05/2019 11:14	SPSS Statis	120 KB				
🥪 USB DRIVE (M:)		🖷 test3	12/05/2019 17:27	SPSS Statis	63 KB				
🔃 Network	•	u test4	12/05/2019 18:24	SPSS Statis	65 KB				

Fig. 34:

UKDA-6695-spss	24/07/2018 00:50	File Folder	
6695spss_fa3ff1f37a5f7dd3c4ff6d62b3923ac4	29/04/2019 20:42	ZIP File	4,229 KB
📻 test1	12/05/2019 11:14	SPSS Statis	3,686 KB
🙀 test2	12/05/2019 11:14	SPSS Statis	120 KB
📻 test3	12/05/2019 17:27	SPSS Statis	63 KB
🙀 test4	12/05/2019 18:24	SPSS Statis	65 KB



Double click on

File test4 is now the active file.

Fig. 35:

e <u>E</u> dit	View Data	Transform A	Analyze Dir	ect <u>M</u> arketing	Graphs Utilitie	s Extensions	Window Help		AB6		
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	year	Numeric	4	0	Year of survey	None	None	10	≡ Right	🛷 Scale	ゝ Input
2	Serial	Numeric	10	0	Serial Number	None	None	10	■ Right	🛷 Scale	ゝ Input
3	rearngrp	Numeric	2	0	Quartile earnin	{1, Q1}	97 - 99, -1	10	≡ Right	📲 Ordinal	ゝ Input
4	REarn	Numeric	2	0	R's own gross	{-1, Skip, n	97 - 99, -1	7	≡ Right	📲 Ordinal	ゝ Input
5	REarnQ	Numeric	2	0	Respondent e	{-1, Skip,n	-1, 7, 8	8	≡ Right	📲 Ordinal	ゝ Input
6	RSex	Numeric	2	0	Sex of respon	{1, Male}	None	5	≡ Right	🗞 Nominal	ゝ Input
7	RAge	Numeric	2	0	What was R's	{97, 97+}	None	6	≡ Right	🛷 Scale	ゝ Input
8	RAgeCat	Numeric	2	0	Age of respon	{1, 18-24}	8	9	≡ Right	📲 Ordinal	ゝ Input
9	RAgeCat2	Numeric	2	0	Age of respon	{1, 18-24}	9	10	≡ Right	📲 Ordinal	ゝ Input
10	REmploye	Numeric	2	0	Is R an emplo	{-9, Refusa	-91	9	≡ Right	📲 Ordinal	ゝ Input
11	EJbHrCal	Numeric	2	0	Hours R works	{-1, Not em	5 - 9, -1	10	≡ Right	📲 Ordinal	ゝ Input
12	SJbHrCal	Numeric	2	0	Hours R works	{-1, Not sel	5 - 9, -1	10	≡ Right	📲 Ordinal	ゝ Input
13	RNSEGGrp	Numeric	2	0	Resp:SEG <g< td=""><td>{-1, Never</td><td>-1, 7, 8</td><td>10</td><td>≡ Right</td><td>📲 Ordinal</td><td>ゝ Input</td></g<>	{-1, Never	-1, 7, 8	10	≡ Right	📲 Ordinal	ゝ Input
14	RNSocCl	Numeric	2	0	Respondent :	{-1, Never	-1, 7, 8	8	■ Right	📲 Ordinal	ゝ Input
15	Теа	Numeric	2	0	R how old wh	{1, 15 or u	6 - 99	5	≡ Right	📲 Ordinal	ゝ Input
16	HEdQual2	Numeric	2	0	Highest educ	{1, Postgra	7, 9	10	≡ Right	🚮 Ordinal	💊 Input
17	GOR2	Numeric	2	0	Government o	{1, North E	None	5	≡ Right	🗞 Nominal	💊 Input
18	Country	Numeric	2	0	Country of int	{1, England	None	9	≡ Right	\delta Nominal	💊 Input
19	WtFactor	Numeric	6	4	Final BSA wei	None	None	12	■ Right	🛷 Scale	ゝ Input
20	4										
	Variable View										

Changing the display of variable attributes

The author's preferred display of variable attributes is:

Name Missing Decimals Measure Label Values

View >> Customize Variable View

Fig. 35:

a test4.	sav [DataSet1]	3] - IBM SPSS	Statistics Data	Editor							- 0	12
Ble Ed	tt View Dat	a <u>t</u> ransform	Analyze De	ett Mattelin	g Graphs Use	es Eglensions	Wedaw He	dp.			20.00	22
21	V Statu Took				H =		1 🛲 📲	00	*6			
		Editor		Necimals	Label	Values	Missing	Columns	Align	Measure	Role	
1				-	Year of survey	None	None	10	Right	# Scale	> Input	
2	Eonts				Serial Number	None	None	10	Right Right	/ Scale	> Input	
3		✓ Grd Lines			R's own gross.	(-1, Skip, n.,	97 . 99 . 1	7	Right .	Ordinal	> Input	
4	H H Yana				Respondent e	{-1, Skip,n	-1, 7, 8	8	all Right	d Ordinal	> Input	
5	and the second second	Angepalied Data		_	Sex of respon	{1, Male}-	None	5	Right	& Normal	> Input	
6	B' Qusta	imze Variabie V	ice.		What was R's	(97, 97+)	None	6	Right 8	/ Scale	> Input	
7	1 Di Ogla		Ctrl+T		Age of respon.	(1, 18-24)	8	9	Right	Ordinal	> Input	
8	RAgeCat2	Numeric	2	0	Age of respon.	(1, 18.24)	9	10	Right	Ordinal	> Input	
9	REmploye	Numeric	2	0	Is R an emplo	(-9. Refusa	-91	9	Right	Ordinal	> Input	

Fig. 36a:

Fig. 36a:		Fig. 36b:	
🍓 Customize Variable View	×	🍓 Customize Variable View	X
Customize Variable View Reorder and select attributes to display in the Variat Show Name Name Yes Yes Width Decimals Label Values Nissing Columns Align Role		Customize Variable View Reorder and select attributes to display in the Varia Name Name Name Align Role Role Customize Customize Role Customize Role Columns Align Role Columns Columns Align Role Columns Col	
Restore <u>D</u> efaults		Restore <u>D</u> efaults	
OK Cancel Help		OK Cancel Help	

Use the blue arrows to move variables up or down the priority list. Uncheck to hide attributes you don't really need. Press OK

*test4	.sav [DataSet1]	- IBM SPSS Statist	ics Data Editor				_ [
ile <u>E</u> dit	t <u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u> naly	ze Direct <u>M</u> arketir	ig <u>G</u> raphs <u>L</u>	<u>J</u> tilities E <u>x</u> tens	ions <u>W</u> indo	ow <u>H</u> elp	
		<u>r</u> a 📱			📴 🚍 4			6
	Name	Measure	Label	Values	Missing	Decimals	Туре	
1	year	🛷 Scale	Year of survey	None	None	0	Numeric	
2	Serial	🛷 Scale	Serial Number	None	None	0	Numeric	
3	rearngrp	🚽 Ordinal	Quartile earnin	{1, Q1}	97 - 99, -1	0	Numeric	
4	REarn	🚽 Ordinal	R's own gross	{-1, Skip, n	97 - 99, -1	0	Numeric	
5	REarnQ	🚽 Ordinal	Respondent e	{-1, Skip,n	-1, 7, 8	0	Numeric	
6	RSex	\delta Nominal	Sex of respon	{1, Male}	None	0	Numeric	
7	RAge	🛷 Scale	What was R's	{97, 97+}	None	0	Numeric	
8	RAgeCat	🚽 Ordinal	Age of respon	{1, 18-24}	8	0	Numeric	
9	RAgeCat2	🚽 Ordinal	Age of respon	{1, 18-24}	9	0	Numeric	
10	REmploye	🚽 Ordinal	Is R an emplo	{-9, Refusa	-91	0	Numeric	
11	EJbHrCal	🚽 Ordinal	Hours R works	{-1, Not em	5 - 9, -1	0	Numeric	
12	SJbHrCal	🚽 Ordinal	Hours R works	{-1, Not sel	-1, 5, 9	0	Numeric	
13	RNSEGGrp	🚽 Ordinal	Resp:SEG <g< td=""><td>{-1, Never</td><td>-1, 7, 8</td><td>0</td><td>Numeric</td><td></td></g<>	{-1, Never	-1, 7, 8	0	Numeric	
14	RNSocCl	🛃 Ordinal	Respondent :	{-1, Never	-1, 7, 8	0	Numeric	
15	Теа	🛃 Ordinal	R how old wh	{1, 15 or u	6 - 99	0	Numeric	
16	HEdQual2	🛃 Ordinal	Highest educ	{1, Postgra	7, 9	0	Numeric	
17	GOR2	\delta Nominal	Government o	{1, North E	None	0	Numeric	
18	Country	\delta Nominal	Country of int	{1, England	None	0	Numeric	
19	WtFactor	🖋 Scale	Final BSA wei	None	None	4	Numeric	
20								
	4							
Data Viev	Variable View							
			IBM SPS	SS Statistics Proc	essor is readv	Unicod	le:ON	

----27

Since the variables are all Numeric, we don't really need Type either, but this format is much easier to navigate.

Ctrl S to save the file.

End of session: 3.2.1.6 Earnings differences 2009: Extracting and saving selected variables

Back to: 3.2.1.5 Earnings differences 2009: Download and check file

Back to: 3.2 Three (or more) variables

Next session: 3.2.1.7: Earnings differences 2009: Elaboration