

Block 2: Analysing one variable

2.3 Data transformations

2.3.1 Data transformations (detailed contents)

[Draft only: 7 Nov 2011]

Selecting **variables** for analysis, changing the **names** of variables, changing the **values** of variables. Creating **new variables** from existing ones. Selecting **cases** for analysis as a transition from analysing one variable to analysing two or more variables.

These operations are called **data transformations**.

2.3.1.1 Data transformations

[Introductory tutorial, 11 pages]

- 1: Select **cases** for analysis [SELECT IF ~ ~ ~ ~ ]
- 2: Select **variables** for analysis [GET /KEEP ~ ~ ~ ~ ]
- 3: Change the **names** of variables [RENAME VARIABLES ~ ~ ~ ~ ]
- 4: Change or group the **values** of variables [RECODE ~ ~ ~ ~ ]
- 5: Create **new variables** by
  - a: Recoding values [RECODE ~ ~ ~ ~ INTO ~ ~ ~ ~ ]
  - b: Performing calculations based on the values of one or more variables. [COUNT ~ ~ ~ ~ ]  
[COMPUTE ~ ~ ~ ~ ]  
[IF ~ ~ ~ ~ ]  
[DO IF ~ ~ ~ ~ ]  
[ELSE IF ~ ~ ~ ~ ]

2.3.1.2 Exercise to rename variables

[NB: This is for BSA 89: it needs renumbering as a later exercise. There will be a new exercise for BSA 86]

**Exemplar:** British Social Attitudes 1989

**File:** [mybsa86\\_3.sav](#)

**Task 1:** Rename demographic variables

Old name	New name
v1408	marital
v1411	sex
v1412	age

**Task 2:** Create a new variable by recoding **age** into **agegroup** with four categories: 18 -29, 30 -44, 45 -59, 60 and over.

**SPSS commands:** RENAME VARIABLES  
RECODE ~ ~ ~ INTO  
VARIABLE LABELS  
EXECUTE  
VALUE LABELS  
FORMATS  
FREQUENCIES ~ ~ ~ ~  
File > Save As ...



### 2.3.1.3 Conditional frequencies exercise

**Exemplar:** British Social Attitudes 1986

**File:** [bsa86.txt](#) (Raw data file, 23 records per case)

**Research question:** What is the distribution of personal gross earnings? What shape does the distribution have? What is the distribution for women only? What is the distribution for men only?

**Facsimile questions:** Own gross earnings  
Household grid.

**Task 1:** Using **positional** variable names, read in the raw data for sex and gross earnings of respondent from external file **bsa86.txt**. Specify **missing values**, **variable labels** and **value labels** using mixed case text and the **<** and **£** signs.

**Task 2:** Produce a frequency count for each variable.

**Task 3:** Produce separate frequency counts (with barcharts) of gross earnings, first for women only, then for men only.

### 2.3.1.4 Specimen answers for exercise 2.3.1.3 (Conditional frequencies)

[NB: This is the final draft of a very extended and detailed exercise: it's quite long (26 pp) but manageable. It uses data from the 1986 British Social Attitudes survey as an exemplar as preparation for moving from analysing one variable to analysing two or more variables.]

**Task 1:** Using **positional** variable names, read in the raw data for **sex of respondent** and **gross earnings of respondent** from external file **bsa86.txt**.

Specify **missing values**, **variable labels** and **value labels** using mixed case text together with the **£** and **<** (less than) signs.

Save the the syntax file as **mybsa86\_5.sps** data editor as **mybsa86\_5.sav** and in your folder **mybsa86**

**data list**  
**missing values**  
**variable labels**  
**value labels**  
File > Save As ...

**Task 2:** Produce a frequency count for each variable.

**frequencies <varlist> .**

**Task 3:** Produce separate frequency counts (with barcharts) of gross earnings, first for women only, then for men only. Do these one at a time.

**temporary .**  
**select if <logical expression> .**  
**frequencies <varlist>**  
**/barchart .**

The following are in preparation and will be uploaded in the near future.

[Conditional frequencies homework](#)

(repeating exercise 2.3.1.4 on the same variables from the 1989 British Social Attitudes survey.)

[Specimen answer](#) for conditional frequencies homework above

[Intermediate exercises](#) to construct files containing additional variables relevant to variation in own gross earnings. These are based on variables the raw data for which are too complex for beginners to work with and so will use derived variables supplied as a supplement to the original raw data. They will be used in further conditional frequency tables and then, to demonstrate the logic of survey analysis, by progressing to zero order, then first order tables in Block 3.