

Block 1: From questionnaire to SPSS saved file

1.4.3: Missing values - a note

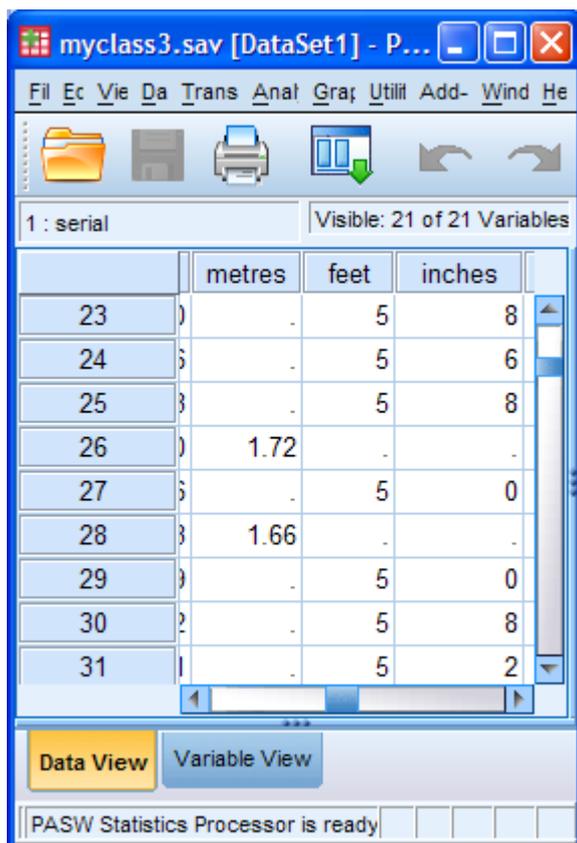
[23 November 2010]

All surveys have questions for which the responses will be missing for some reason or other (eg 'Don't know', 'Not Answered', 'Not Applicable').

We now need to complete the first version of the saved file for the data from our pre-course questionnaire by specifying, for each variable, any values to be treated as missing and which therefore need to be omitted from most statistical calculations. In practice, for most surveys, it's preferable to declare missing values before adding variable and value labels (which take a lot longer to prepare) so that you can do some preliminary analyses without waiting for the labels.

In fact there aren't any missing values coded as such in this data set, but there were one or two items left blank by some students, especially by the ones who didn't know their height in metres! If these data had been read in as alphabetic (strings) the blanks would have been treated as valid values, but because we read the data in as numeric, the blanks were treated as missing by SPSS when it found them and automatically given the system missing value **sysmis**.

These appear as dots (periods) ' .' in the data file.



Extract from Data Editor for myclass3.sav showing dots to replace blanks

## MISSING VALUES

The **MISSING VALUES** command is used to specify those values of each variable which are to be treated as 'missing' (typically Don't know, Refused, etc.). These values are known as **user-missing** since the user defines them as such. Up to **three** such values can be defined.

**General format:**

```
MISSING VALUES    <variable list>    ( <value list> )
                   / <variable list>    ( <value list> )
                   / ~ ~ ~ ~ .
```

where <value list> can be:

- (i) up to three numeric or alphanumeric values separated by commas or blank spaces
- (ii) for numeric values only, a **range** of the form:

<value> **THRU** <value>

**LOWEST THRU** <value>

<value> **THRU HIGHEST**

Because you are allowed to specify three values and because the keyword **THRU** only uses two of these, you can specify a third value in addition to the two used with **THRU**.

Example:

```
missing values    sex          (0 , 9)
                  /age         (0 , 97 thru 99)
                  /occupation  ('Unemployed')
                  /q1 to q10   (0)
```

Variables which have the same missing values can be declared together, eg:

```
missing values    v8 to v12 v14 (8,9) .
```

In addition to user-defined missing values, SPSS has another **system-missing** value which it assigns to alphabetic characters and embedded blanks found in numeric fields when reading raw data. (SPSS issues a warning message when this happens). This assigned value is known inside SPSS as **SYSMIS** and is displayed as a dot '.' on output. Thus a variable in SPSS can in effect have up to four missing values specified, two of which may be the lower and upper limits of a range of values.

When reading data in numeric format, SPSS automatically sets blanks to the system missing value **sysmis**. If you want to set blanks in numeric fields to a value other than **sysmis** you can use the special **SET** command eg:

```
SET    BLANKS = -1 .
```

**End of Block 1:**

**From questionnaire to SPSS saved file**

**Next tutorials are in Block 2:**

**Analysing one variable**

[\[Back to Block 1 menu\]](#)

or

[\[Forward to Block 2 menu\]](#)