

Appendix 8: Grouped Percentiles

Two summary functions, GMEDIAN and GPTILE are used in procedures such as Frequencies and Graph, to calculate the percentiles for the data which are grouped by specifying a value for each grouping. It is assumed that the actual data values give represent midpoints of the grouped intervals.

Notation

The following notation is used throughout this appendix unless otherwise stated:

| | |
|---------------------|---|
| $x_1 < \dots < x_k$ | Distinct observed values with frequencies (caseweights) c_1, \dots, c_k |
| k | Number of distinct observed data points |
| p | percentile/100 (a number between 0 and 1) |
| cc_l | Cumulative frequency up to and including x_l |

$$cc_l = \sum_{i=1}^{l-1} c_i + 0.5 * c_l \quad l = 1, \dots, n$$

Finding Percentiles

To find the $100p$ th grouped percentile, first find i such that $cc_{i-1} \leq wp < cc_i$, where w is the total sum of caseweights which is equal to $\sum_{j=1}^k c_j$. Then the grouped percentile is

$$(1 - R)x_{i-1} + Rx_i$$

2 Appendix 8

where

$$R = \frac{wp - cc_{i-1}}{cc_i - cc_{i-1}}$$

Note the following:

- If $wp < cc_1$, the grouped percentile is system missing and a warning message “Since the lower bound of the first interval is unknown, some percentiles are undefined” is produced.
- If $wp > cc_k$, the grouped percentile is system missing and a warning message “Since the upper bound of the last interval is unknown, some percentiles are undefined” is produced.
- If $wp = cc_k$, the grouped percentile is equal to x_k .