FREQUENCIES

If the absolute value of any observation is greater than 10^{13}, no calculations are done. For sorting of the observations, see Appendix 6. For information on percentiles for grouped data, see Appendix 8.

Notation

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

X_k	Value of the variable for case *k*
w_k	Weight for case *k*
NV	Number of distinct values the variable assumes
N	Number of cases
W	Sum of weights of the cases

Basic Statistics

The values are sorted into ascending order and the following calculated:

Sum of Weights of Cases Having Each Value of *X*

$$f_j = \sum_{i=1}^{N} w_i k_i \qquad j = 1, 2, \ldots, NV$$

where

$$k_i = \begin{cases} 1 & \text{if } X_i = X_j \\ 0 & \text{otherwise} \end{cases}$$

where X_j is the *j*th largest distinct value of *X*.

Relative Frequency (Percentage) for each Value of *X*

$$Rf_j = \left(\frac{f_j}{W'} \right) \times 100$$

where

$W' = \sum_{i=1}^{NV} f_i$ (sum over all categories including those declared as missing values)

Adjusted Frequency (Percentage)

$$Af_j = \left(\frac{f_j}{W} \right) \times 100$$

where

$W = \sum_{i=1}^{NV} f_i k_i$ (sum over nonmissing categories)

and

$$k_i = \begin{cases} 0 & \text{if } X_i \text{ has been declared missing} \\ 1 & \text{otherwise} \end{cases}$$

For all X_j declared missing, an adjusted frequency is not printed.

Cumulative Frequency (Percentage)

$$Cf_j = \sum_{i=1}^{j} f_i$$

Minimum

$$\min_k X_k$$

Maximum

$$\max_k X_k$$

Mode

Value of X_j which has the largest observed frequency. If several are tied, the smallest value is selected.

Range

Maximum – Minimum

The *p* th percentile

Find the first score interval (*x2*) containing more than *tp* cases.

$$p\text{th percentile} = \begin{cases} x_2 & \text{if } tp - cp_1 \geq 100/W \\ \left\{1 - \left[(W+1)\,p/100 - cc_1\right]\right\} x_1 & \text{if } tp - cp_1 < 100/W \\ \quad + \left[(W+1)\,p/100 - cc_1\right] x_2 & \end{cases}$$

where

$tp = (W+1)\,p/W$
$cp_1 < tp < cp_2$
x_1 and x_2 are the values corresponding to cp_1 and cp_2, respectively
cc_1 is the cumulative frequency up to x_1
cp_1 is the cumulative percent up to x_1

Mean

$$\overline{X} = \frac{\sum_{j=1}^{NV} f_j X_j}{W}$$

Moments about the mean are calculated as:

$$M_j = \sum_{i=1}^{NV} f_i \left(X_i - \overline{X} \right)^j \qquad j = 2, 3, 4$$

Variance

$$S^2 = \frac{M_2}{(W-1)}$$

Standard Deviation

$$S = \sqrt{S^2}$$

Standard Error of the Mean

$$SEM = \frac{S}{\sqrt{W}}$$

Skewness (computed if $W \geq 3$ and $S^2 > 0$) (Bliss, 1967, p. 144)

$$g_1 = \frac{WM_3}{(W-1)(W-2)S^3} \qquad se(g_1) = \sqrt{\frac{6W(W-1)}{(W-2)(W+1)(W+3)}}$$

Kurtosis (computed if $W \geq 4$ and $S^2 > 0$)

$$g_2 = \frac{W(W+1)M_4 - 3(W-1)M_2^2}{(W-1)(W-2)(W-3)S^4} \qquad se(g_2) = \sqrt{\frac{4\left(W^2-1\right)se(g_1)^2}{(W-3)(W+5)}}$$

References

Blalock, H. M. 1972. *Social statistics*. New York: McGraw-Hill.

Bliss, C. I. 1967. *Statistics in biology*, Volume 1. New York: McGraw-Hill.