RANK

Notation

Let $y_1 < y_2 < \cdots < y_m$ be m distinct ordered observations for the sample and $C_1, C_2, \ldots, C_m$ be the corresponding sum of caseweights for each value. Define

$$CC_i = \sum_{k=1}^{i} C_k = \text{cumulative sum of caseweights up to } y_i$$

$$W = CC_m = \sum_{k=1}^{m} C_k = \text{total sum of caseweights}$$

Statistics

Rank (R_i)

A rank is assigned to each case based on four different ways of treating ties or caseweights not equal to 1.

For every i, $i = 1, \ldots, m$,

(a) if $C_i \geq 1$

$R_i = CC_{i-1} + 1$ if TIES = LOW

$R_i = CC_i$ if TIES = HIGH

$R_i = CC_{i-1} + (C_i + 1)/2$ if TIES = MEAN

$R_i = i$ if TIES = CONDENSE

(b) if $C_i < 1$

$R_i = CC_{i-1}$	if TIES = LOW
$R_i = CC_i$	if TIES = HIGH
$R_i = CC_{i-1} + C_i/2$	if TIES = MEAN
$R_i = i$	if TIES = CONDENSE

Note: $CC_0 = 0$.

RFRACTION (RF_i)

Fractional rank:

$$RF_i = R_i/W, \; i = 1, \ldots, m$$

PERCENT (P_i)

Fractional rank as a percentage:

$$P_i = \frac{R_i}{W} \times 100, \; i = 1, \ldots, m$$

PROPORTION (F_i): Estimate for Cumulative Proportion

The proportion is calculated for each case based on four different methods of estimating fractional rank:

$F_i = \left(R_i - \frac{3}{8}\right) \Big/ \left(W + \frac{1}{4}\right)$	(BLOM)
$F_i = \left(R_i - \frac{1}{2}\right) \Big/ W$	(RANKIT)
$F_i = \left(R_i - \frac{1}{3}\right) \Big/ \left(W + \frac{1}{3}\right)$	(TUKEY)
$F_i = R_i / (W + 1)$	(Van der Waerden)

Note: F_i will be set to SYSMIS if the calculated value of F_i by the formula is negative.

NORMAL (*a*)

Normal scores that are the *Z*-scores from the standard normal distribution that corresponds to the estimated cumulative proportion *F*. The normal score is defined by

$$a_i = \Psi(F_i), \; i = 1,\ldots,m$$

where Ψ is the inverse cumulative standard normal distribution (PROBIT).

NITLES (*K*)

Assign group membership for the requested number of groups. If *K* groups are requested, the *n*-tile (N_i) for case *i* is defined by

$$N_i = \left\lfloor \frac{R_i K}{W+1} \right\rfloor + 1$$

where $\left\lfloor \frac{R_i K}{W+1} \right\rfloor$ is the greatest integer that is less than or equal to $R_i K / (W+1)$.

SAVAGE (S)

Savage scores based on exponential distribution. The Savage score is calculated by

$$S_i = \begin{cases} \left\{ \left[\left(1 - g_{i_1}\right) l_{i_1+1} + g_{i_2} l_{i_2+1} + \sum_{j=i_1+2}^{i_2} l_j \right] \Big/ C_i \right\} - 1 & i_1 + 2 \le i_2 \\ \left\{ \left[\left(1 - g_{i_1}\right) l_{i_1+1} + g_{i_2} l_{i_2+1} \right] \Big/ C_i \right\} - 1 & i_1 + 1 = i_2 \\ l_{i_1+1} - 1 & i_1 = i_2 \end{cases}$$

where

$$i_1 = \left[CC_{i-1}\right], \quad i_2 = \left[CC_i\right], \quad W^* = \begin{cases} W & \text{if } W \text{ is an integer} \\ \left[W\right]+1 & \text{if } W \text{ is not an integer} \end{cases}$$

$$g_{i_1} = CC_{i-1} - i_1, \quad g_{i_2} = CC_i - i_2$$

and $l_1, \ldots, l_{w^*}$ are defined as the expected values of the order statistics from an exponential distribution; that is

$$l_j = \sum_{K=1}^{j} \frac{1}{W^* - K + 1}$$

References

Blom, G. 1958. *Statistical estimates and transformed beta variables*. New York: John Wiley & Sons, Inc.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. 1983. *Graphical methods for data analysis*. Belmont, Calif.: Wadsworth International Group; Boston: Duxbury Press.

Lehmann, E. L. 1975. *Nonparametrics*: *Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Tukey, J. W. 1962. The future of data analysis. *The Annals of Mathematical Statistics*, 33: 1–67 (Correction: 33: 812)