

1.2 Definitions of statistics

Given a set of values $\{y_1, y_2, \dots, y_n\}$ for any individual variable Y , the following summary statistics can be calculated by clicking on **Tools > Summary Stats...** in PRIMER 8:

- **Average:** $\bar{y} = \sum_{i=1}^n y_i / n$, the average (or mean)
- **Median:** m , the median value
- **Sum:** $\sum_{i=1}^n y_i$, the sum of the values
- **Minimum:** $\min(y_i)$, the minimum value
- **Maximum:** $\max(y_i)$, the maximum value
- **Quantiles:** q_α , the value corresponding to a given (α -)quantile in the empirical distribution of values. Quantiles must be chosen by the end-user and must be in the range (0, 1).
- **Range:** the range; i.e., $(\max(y_i) - \min(y_i))$, the difference between the maximum and minimum values
- **IQR:** the inter-quartile range; i.e., $(q_{0.75} - q_{0.25})$, the difference between the upper and lower quartile.
- **Standard deviation:** s , the standard deviation; i.e., the square root of the variance.
- **Variance:** $s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1)$, an unbiased estimate of the variance.
- **Sample size:** n , the number of values
- **Standard error:** $\sqrt{s^2/n}$, the standard error of the mean
- **Symmetry:** α -symmetry statistic, with α chosen by the end-user (default $\alpha = 0.05$). For symmetric data, the median (m) is equidistant from the α -quantile and the $(1-\alpha)$ -quantile. The α -symmetry statistic is defined as $(m - q_\alpha) / (q_{1-\alpha} - q_\alpha)$ for a given quantile (α). A value close to 0.5 indicates symmetry, a value < 0.5 indicates right-skewness, and a value > 0.5 indicates left-skewness.
- **Skewness:** k_3 , the skewness coefficient; i.e., $k_3 = \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{(n-1)(n-2) \cdot s^3}$. A value close to zero indicate symmetry. A positive value indicates right-skewness; a negative value indicates left-skewness. See [Sheskin \(2011\)](#).
- **Kurtosis:** k_4 , the kurtosis coefficient; i.e., $k_4 = \frac{\left[\sum_{i=1}^n (y_i - \bar{y})^4 / (n-1) \right] - 3 \left[\sum_{i=1}^n (y_i - \bar{y})^2 / (n-2) \right]^2}{(n-2)(n-3) \cdot s^4}$. A value close to zero indicates a mesokurtic distribution. A positive value indicates a leptokurtic distribution (pointy, with broad tails). A negative value indicates a platykurtic distribution (flat-topped, with short tails). See [Sheskin \(2011\)](#).
- **Number of zeros:** the number of zeros.
- **Singletons:** the number of ones (useful for count data).
- **Doubletons:** the number of twos (useful for count data).
- **Number of nonzeros (frequency):** the number of non-zero values; e.g., if the variable consisted of counts of an organism, this would be the frequency of

occurrences of that organism across the set of values (samples).

- **Smallest number above threshold:** $\text{min}_{y \geq y_t}$ the smallest value in the set that occurs above a specified threshold value (y_t), chosen by the end-user. For example, to obtain the smallest non-zero value in a set of non-negative values, specify $y_t=0$. Here is another example: suppose a variable consists of lead (Pb) concentrations measured from sediment. It may be useful to identify the smallest concentration value recorded above the detection limit of the instrument. Knowing the smallest non-zero (or detected) value can be handy for choosing an appropriate constant (c) to add for a transformation such as $\log(y+c)$, when the variable contains zero values.
 - **Largest number below threshold:** $\text{max}_{y \leq y_t}$ the largest value in the set that occurs below a specified threshold value (y_t), chosen by the end-user. This option has similar uses to the previous one, but for non-positive data.
-

Revision #121

Created 14 November 2024 23:35:15 by Marti

Updated 24 September 2025 21:02:20 by Marti