quantile {stats} R Documentation

Sample Quantiles

Description

The generic function `quantile` produces sample quantiles corresponding to the given probabilities. The smallest observation corresponds to a probability of 0 and the largest to a probability of 1.

Usage

```
quantile(x, ...)

## Default S3 method:
quantile(x, probs = seq(0, 1, 0.25), na.rm = FALSE,
         names = TRUE, type = 7, ...)
```

Arguments

`x`
: numeric vector whose sample quantiles are wanted, or an object of a class for which a method has been defined (see also ‘details’). `NA` and `NaN` values are not allowed in numeric vectors unless `na.rm` is `TRUE`.

`probs`
: numeric vector of probabilities with values in *[0,1]*. (Values up to `2e-14` outside that range are accepted and moved to the nearby endpoint.)

`na.rm`
: logical; if true, any `NA` and `NaN`'s are removed from `x` before the quantiles are computed.

`names`
: logical; if true, the result has a `names` attribute. Set to `FALSE` for speedup with many `probs`.

`type`
: an integer between 1 and 9 selecting one of the nine quantile algorithms detailed below to be used.

`...`
: further arguments passed to or from other methods.

Details

A vector of length `length(probs)` is returned; if `names = TRUE`, it has a `names` attribute.

`NA` and `NaN` values in `probs` are propagated to the result.

The default method works with classed objects sufficiently like numeric vectors that `sort` and (not needed by types 1 and 3) addition of elements and multiplication by a number work correctly. Note that as this is in a namespace, the copy of `sort` in **base** will be used, not some S4 generic of that name. Also note that that is no check on the ‘correctly’, and so e.g. `quantile` can be applied to complex vectors which (apart from ties) will be ordered on their real parts.

There is a method for the date-time classes (see "`POSIXt`"). Types 1 and 3 can be used for class "`Date`" and for ordered factors.

Types

`quantile` returns estimates of underlying distribution quantiles based on one or two order statistics from the supplied elements in `x` at probabilities in `probs`. One of the nine quantile algorithms discussed in Hyndman and Fan (1996), selected by `type`, is employed.

All sample quantiles are defined as weighted averages of consecutive order statistics. Sample quantiles of type *i* are defined by:

$$Q[i](p) = (1 - \gamma)\, x[j] + \gamma\, x[j+1],$$

where $1 \le i \le 9$, $(j-m)/n \le p < (j-m+1)/n$, $x[j]$ is the *j*th order statistic, *n* is the sample size, the value of γ is a function of $j = floor(np + m)$ and $g = np + m - j$, and *m* is a constant determined by the sample quantile type.

Discontinuous sample quantile types 1, 2, and 3

For types 1, 2 and 3, $Q[i](p)$ is a discontinuous function of *p*, with $m = 0$ when $i = 1$ and $i = 2$, and $m = -1/2$ when $i = 3$.

Type 1

Inverse of empirical distribution function. $\gamma = 0$ if $g = 0$, and 1 otherwise.

Type 2

Similar to type 1 but with averaging at discontinuities. $\gamma = 0.5$ if $g = 0$, and 1 otherwise.

Type 3

SAS definition: nearest even order statistic. $\gamma = 0$ if $g = 0$ and *j* is even, and 1 otherwise.

Continuous sample quantile types 4 through 9

For types 4 through 9, $Q[i](p)$ is a continuous function of *p*, with *gamma = g* and *m* given below. The sample quantiles can be obtained equivalently by linear interpolation between the points $(p[k],x[k])$ where $x[k]$ is the *k*th order statistic. Specific expressions for $p[k]$ are given below.

Type 4

$m = 0$. $p[k] = k / n$. That is, linear interpolation of the empirical cdf.

Type 5

$m = 1/2$. $p[k] = (k - 0.5) / n$. That is a piecewise linear function where the knots are the values midway through the steps of the empirical cdf. This is popular amongst hydrologists.

Type 6

$m = p$. $p[k] = k / (n + 1)$. Thus $p[k] = E[F(x[k])]$. This is used by Minitab and by SPSS.

Type 7

$m = 1-p$. $p[k] = (k - 1) / (n - 1)$. In this case, $p[k] = mode[F(x[k])]$. This is used by S.

Type 8

$m = (p+1)/3$. $p[k] = (k - 1/3) / (n + 1/3)$. Then $p[k] =\sim median[F(x[k])]$. The resulting quantile estimates are approximately median-unbiased regardless of the distribution of `x`.

Type 9

$m = p/4 + 3/8$. $p[k] = (k - 3/8) / (n + 1/4)$. The resulting quantile estimates are approximately unbiased for the expected order statistics if `x` is normally distributed.

Further details are provided in Hyndman and Fan (1996) who recommended type 8. The default method is type 7, as used by S and by **R** < 2.0.0.

Author(s)

of the version used in **R** >= 2.0.0, Ivan Frohne and Rob J Hyndman.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

Hyndman, R. J. and Fan, Y. (1996) Sample quantiles in statistical packages, *American Statistician* **50**, 361–365.

See Also

`ecdf` for empirical distributions of which `quantile` is an inverse; `boxplot.stats` and `fivenum` for computing other versions of quartiles, etc.

Examples

```
quantile(x <- rnorm(1001)) # Extremes & Quartiles by default
quantile(x,  probs = c(0.1, 0.5, 1, 2, 5, 10, 50, NA)/100)

### Compare different types
p <- c(0.1, 0.5, 1, 2, 5, 10, 50)/100
res <- matrix(as.numeric(NA), 9, 7)
for(type in 1:9) res[type, ] <- y <- quantile(x,  p, type = type)
dimnames(res) <- list(1:9, names(y))
round(res, 3)
```

[Package *stats* version 3.3.0 Index]