Statistics functions in R and spreadsheets

Bert Van Vreckem

Contents

This cheat sheet gives an overview of the most common statistics functions for spreadsheets in English and Dutch (LibreOffice Calc, Excel), and for the R programming language.

1 Univariate statistics

x denotes a cell range (spreadsheet) or list/array/table (R).

Function	R	Spreadsheet (EN)	Spreadsheet (NL)
Mean, average	`mean(x)`	`=AVERAGE(x)`	`=GEMIDDELDE(x)`
Population variance	–	`=VAR.P(x)`	`=VAR.P(x)`
Population standard deviation	–	`=STDEV.P(x)`	`=STDEV.P(x)`
Sample variance	`var(x)`	`=VAR(x)`, `=VAR.S(x)`	`=VAR(x)`, `=VAR.S(x)`
Sample standard deviation	`sd(x)`	`=STDEV(x)`, `=STDEV.S(x)`	`=STDEV(x)`, `=STDEV.S(x)`
Median	`median(x)`	`=MEDIAN(x)`	`=MEDIAAN(x)`
Minimum	`min(x)`	`=MIN(x)`	`=MIN(x)`
Maximum	`max(x)`	`=MAX(x)`	`=MAX(x)`
Quartile	–	`=QUARTILE(x, type)`†	`=KWARTIEL(x, type)`†
Percentile	`quantile(x, alphas)`‡	`=PERCENTILE(x, alpha)`‡	`=PERCENTIEL(x, alpha)`‡

† `type`: 0 = min, 1 = 25% (1st quartile) , 2 = 50% (median), 3 = 75% (3rd quartile), 1 = max

‡ `alpha` is a number in [0, 1] denoting the percentile rank (0 = minimum, .5 = median, 1 = max). In R, you can specify an array of the desired percentiles, e.g. `quantile(x, c(0, .33, .67, 1))`.

2 Bivariate statistics

- x denotes the cell range (spreadsheet) or list/array/table (R) containing values of the *independent variable*.
- y denotes the cell range (spreadsheet) or list/array/table (R) containing values of the *dependent variable*.

Function	R	Spreadsheet (EN)	Spreadsheet (NL)
Pearson's correlation coefficient (R)	`cor(x, y)`	`=PEARSON(y, x)`	`=PEARSON(y, x)`
Determination coefficient (R^2)		`=RSQ(y, x)`	`=R.KWADRAAT(y, x)`
Covariance	`cov(x, y)`	`=COVAR(x, y)`	`COVARIANTIE.S(x, y)`

3 Probability density of the normal distribution

- X is a normally distributed stochastic variable with mean `m` and standard deviation `s`, or `X ~ Nor(m, s)`. `x` is a number drawn from X.
  - `P(X < x)` is the probability that a number is drawn from X smaller than `x` (left tail probability)
- Z is the standard normal distribution, or `Z ~ Nor(0, 1)`. `z` is a number drawn from Z.
  - `P(Z < z)` is the probability that a number is drawn from Z smaller than `z` (left tail probability)

Function	R	Spreadsheet (EN)	Spreadsheet (NL)
z-transformation	`z <- (x - m)/s`	`=STANDARDIZE(x, m, s)`	`=NORMALISEREN(x, m, s)`
`P(Z < z)`	`pnorm(z)`	`=NORMSDIST(z)`	`=STAND.NORM.VERD(z)`
`P(X < x)`	`pnorm(x, m, s)`	`=NORMDIST(x, m, s)`	`=NORM.VERD(x, m, s)`
`z` so `P(Z < z) = p`	`qnorm(p)`	`=NORM.S.INV(p)`	`=NORM.S.INV(p)`
`x` so `P(X < x) = p`	`qnorm(p, m, s)`	`=NORMINV(p, m, s)`	`=NORM.INV.N(p, m, s)`

4 Resources

- Van Der Elst, J. (2012). *Statistiek met Excel.* Derde druk. Uitgeverij De Boeck.