Syntax for Specifying Linear Models in R

Stephen Parry and Erika Mudrak

Suppose `y`, `x`, `x0`, `x1`, `x2`, … are numeric variables and `A`, `B`, `C`, … are factors/categorical variables.

Linear Models (including regression, anova, ancova, etc..)

The following syntax works generally, including in `lm`, `glm`, `aov`, `lme4::lmer`,`lme4::glmer`, `nlme::lme`, `MASS::glmmPQL` , `glmmTMB::glmmTMB` functions.

source: Venebles, Smith and the R Core Team. An Introduction to R: Notes on R: Programming Environment for Data Analysis and Graphics Version 4.4.1 (2024-06-14), Chapter 11
http://cran.r-project.org/doc/manuals/R-intro.pdf

`y ~ x` `y ~ 1 + x`	Both imply the same simple linear regression model of y on x. The first has an implicit intercept term, and the second an explicit one.

`y ~ 0 + x`
`y ~ -1 + x`
`y ~ x - 1`
	Simple linear regression of y on x through the origin (that is, without an intercept term).

`log(y) ~ x1 + x2`	Multiple regression of the transformed variable, log(y), on x1 and x2 (with an implicit intercept term).
`y ~ poly(x,2)` `y ~ 1 + x + I(x^2)`	Polynomial regression of y on x of degree 2. The first form uses orthogonal polynomials, and the second uses explicit powers, as basis.
`y ~ I(x/100)`	Insulate the expression. Inside parentheses all operators have their normal arithmetic meaning, rather than the formula syntax. Here, x is divided by 100 before putting in the model.
`y ~ A`	Analysis of variance model of y, with classes determined by A.
`y ~ A + x`	Analysis of covariance model of y, with classes determined by A, and with covariate x.

`y ~ A*B` `y ~ A + B + A:B`	Two-way analysis of variance model of y on A, B and their interaction
`y ~ A/B` `y ~ A + A:B`	Two-way analysis of variance model of y on B nested in A
`y ~ (A + B + C)^2` `y ~ A*B*C - A:B:C`	Three factor experiment but with a model containing main effects and two factor interactions only. Both formulae specify the same model.
`y ~ (A + B + C + D ...)^n`	All terms in parentheses together with interactions up to order n

`y ~ A * x`
`y ~ A/x`
`y ~ A/(1 + x) - 1`
	Separate simple linear regression models of y on x within the levels of A, with different codings. The last form produces explicit estimates of as many different intercepts and slopes as there are levels in A.

This syntax works in the `aov` function only. Note that aov is designed for balanced designs, and the results can be hard to interpret without balance: beware that missing values in the response(s) will likely lose the balance.

`y ~ A*B + Error(C)`	An experiment with two treatment factors, A and B, and error strata determined by factor C. For example a split plot experiment, with whole plots (and hence also subplots), determined by factor C.

Mixed-Effects Models (models including random effects)

With terms as above including categorical `group` and `block` variables.

The following syntax works to specify the random effects for the `lmer` and `glmer` functions in the `lme4` pacakge, as well as for `glmmTMB::glmmTMB` functions.

From http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#model-specification
Also a good resource: http://lme4.r-forge.r-project.org/book/Ch2.pdf

`(1|group)`	Random group intercept
`(x|group)` `(1+x|group)`	Random slope of x within group with correlated intercept
`(0+x|group)` `(-1+x|group)`	Random slope of x within group: no variation in intercept
`(1|group)+(0+x|group)`	Uncorrelated random intercept and random slope within group
`(1|site/block)` `(1|site)+(1|site:block)`	Intercept varying among sites and among blocks within sites (nested random effects)
`(x|site/block)` `(x|site)+(x|site:block)`	Slope and intercept varying among sites and among blocks within sites
`(1|group1)+(1|group2)`	intercept varying among crossed random effects (e.g. site, year)
`(x1|site)+(x2|block)`	two different effects, varying at different levels
`x*site+(x|site:block)`	fixed effect variation of slope and intercept varying among sites and random variation of slope and intercept among blocks within sites

The following syntax specifies the random effects in the model for the `nlme::lme` (normal distribution only) and `MASS::glmmPQL` (for probability distributions in the exponential family).

`random = ~1 | group`	Random group intercept
`random = ~x | group`	Random slope of x within group with correlated intercept
`random = ~1 | site/block`	Intercept varying among sites and among blocks within sites (nested random effects)
`random = ~x | site/block`	Slope and intercept varying among sites and among blocks within sites