| Variable | Example | Type of Regression | R function / R function for mixed models |
|---|---|---|---|
| Continuous | Quality of Life, linear scales | linear | `lm()` |
| | | | `lmer(), glmmTMB()` |
| Binary | Success yes/no | binary logistic | `glm(family=binomial)` |
| | | | `glmer(*), glmmTMB(*)` |
| Trials (or proportion of *counts*) | 20 successes out of 30 trials | logistic[1] | `glm(cbind(trial,success), family=binomial)` |
| | | | `glmer(*), glmmTMB(*)` |
| Count data | Number of usages, counts of events | Poisson | `glm(family=poisson)` |
| | | | `glmer(*), glmmTMB(*)` |
| Count data, with excess zeros or overdispersion | Number of usages, counts of events (with higher variance than mean of response) | negative binomial | `glm.nb()` |
| | | | `glmer.nb(), glmmTMB(family=nbinom)` |
| Count data with very many zeros (inflation) | see count data, but response is modelled as mixture of Bernoulli & Poisson (two sources of zeros) | zero-inflated | `zeroinfl()` |
| | | | `glmmTMB(ziformula, family=poisson)` |
| Count data, with very many zeros (inflation) and overdispersion | Number of usages, counts of events (with higher variance than mean of response) | zero-inflated negative binomial | `zeroinfl(dist="negbin")` |
| | | | `glmmTMB(ziformula, family=nbinom)` |
| Count data, zero-truncated | see count data, but only for positive counts (hurdle component models zero-counts) | hurdle (Poisson) | `hurdle()` |
| | | | `glmmTMB(family=truncated_poisson)` |
| Count data, zero-truncated and overdispersion | see "Count data, zero-truncated", but with higher variance than mean of response | hurdle (neg. binomial) | `vglm(family=posnegbinomial)` |
| | | | `glmmTMB(family=truncated_nbinom)` |
| Proportion / Ratio (without zero and one) | Percentages, proportions of *continuous* data | Beta[1] | `betareg()` |
| | | | `glmmTMB(family=beta)` |
| Proportion / Ratio (including zero and one) | Percentages, proportions of *continuous* data | Beta-Binomial, zero-inflated Beta, ordered Beta[2] | `BBreg(), betabin(), ordbetareg(), vglm(family=betabinomial)` |
| | | | `glmmTMB(ziformula, family=beta_family/ betabinomial/ ordbeta), ordbetareg()` |
| Ordinal | Likert scale, worse/ok/better | ordinal, proportional odds | `polr(), clm()` |
| | | | `clmm(), mixor(), MCMCglmm()` |

| Variable | Example | Type of Regression | R function / R function for mixed models |
|---|---|---|---|
| Cumulative, multinomial | No natural order of categories, like red/green/blue | cumulative link, multinomial | `multinom()`, `clm()`,`bracl()`, `brmultinom()` |
| | | | `clmm()`, `mixor()`, `MCMCglmm()` |
| Continuous, right-skewed | Financial data, reaction times | Gamma | `glm(family=Gamma)` |
| | | | `glmer(*)`, `glmmTMB(*)` |
| (Semi-)Continuous, (right) skewed, probably spike at zero (zero-inflation) | Financial data, probably exponential dispersion of variance | Tweedie | `glm(family=tweedie)`, `cpglm()` |
| | | | `cpglmm()`,`glmmTMB(family=tweedie)` |
| (Semi-)Continuous, skewed, zero-inflation | Normal distribution, negative values censored and stacked on zero | Tobit | `censReg()`, `tobit()` |
| | | | `semLme()` |
| Continuous, but truncated or outliers | | truncated | `censReg()`, `tobit()`, `vglm(family=tobit)` |
| Continuous, but exponential growth | wildlife populations, financial investments | log-transformed, non-linear | `glm(family=Gaussian("log"), nls()` |
| | | | `glmmTMB(family=Gaussian("log"), nlmer()` |
| Proportion / Ratio with > 2 categories | Biomass partitioning in plants (ratio of leaf, stem and root mass) | Dirichlet | `DirichReg()` |
| Time-to-Event | Survival-analysis, time until event/death occurs | Cox (proportional hazards) | `coxph()` |
| | | | `coxme()` |

* Indicates same family-option for mixed models as for their non-multilevel counterparts.

[1] Note that ratios or proportions from *count data*, like `cbind(successes, failures)`, are modelled as logistic regression with `glm(cbind(successes, failures), family=binomial())`, while ratios from *continuous data* (where the response ranges from 0 to 1) are modelled using beta-regression.

[2] Usually, zero-inflated models are used when 0 or 1 come from a separate process or category. However, when the 0/1 values are most consistent with censoring rather than with a separate category/process, the *ordered beta regression* is probably a better choice (i.e., 0 mean "below detection", not "something qualitatively different happened") (Source: https://twitter.com/bolkerb/status/1577755600808775680)

Last update: 2023/04/25