

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

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	<code>multinom()</code> , <code>clm()</code> , <code>brac1()</code> , <code>brmultinom()</code>
			<code>clmm()</code> , <code>mixor()</code> , <code>MCMCglmm()</code>
Continuous, right-skewed	Financial data, reaction times	Gamma	<code>glm(family=Gamma)</code>
			<code>glmer(*)</code> , <code>glmmTMB(*)</code>
(Semi-)Continuous, (right) skewed, probably spike at zero (zero-inflation)	Financial data, probably exponential dispersion of variance	Tweedie	<code>glm(family=tweedie)</code> , <code>cpglm()</code>
			<code>cpglm()</code> , <code>glmmTMB(family=tweedie)</code>
(Semi-)Continuous, skewed, zero-inflation	Normal distribution, negative values censored and stacked on zero	Tobit	<code>censReg()</code> , <code>tobit()</code>
			<code>semLme()</code>
Continuous, but truncated or outliers		truncated	<code>censReg()</code> , <code>tobit()</code> , <code>vglm(family=tobit)</code>
Continuous, but exponential growth	wildlife populations, financial investments	log-transformed, non-linear	<code>glm(family=Gaussian("log"), nls())</code>
			<code>glmmTMB(family=Gaussian("log"), nlmer())</code>
Proportion / Ratio with > 2 categories	Biomass partitioning in plants (ratio of leaf, stem and root mass)	Dirichlet	<code>DirichReg()</code>
Time-to-Event	Survival-analysis, time until event/death occurs	Cox (proportional hazards)	<code>coxph()</code>
			<code>coxme()</code>

* 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: 2022/11/18