

# Common tests are linear models

See working examples and more details at the accompanying notebook: [https://rpubs.com/lindeloev/tests\\_as\\_linear](https://rpubs.com/lindeloev/tests_as_linear)

		Built-in function in R	Equivalent linear model in R	if N	The model in words	Icon
Simple regression: $\text{lm}(y \sim 1 + x)$	<b>y is independent of x</b> P: One-sample t-test N: Wilcoxon signed-rank	W or L: <code>t.test(y)</code> W or L: <code>wilcox.test(y)</code>	W or L: <code>lm(y ~ 1)</code> W or L: <code>lm(signed_rank(y) ~ 1)</code>	<a href="#">≥14</a>	One number (intercept) predicts y. One number (intercept) predicts the signed rank of y.	
	P: Paired-sample t-test N: Wilcoxon matched pairs	W: <code>t.test(y1, y2, paired=TRUE)</code> W: <code>wilcox.test(y1, y2, paired=TRUE)</code>	W: <code>lm(y2 - y1 ~ 1)</code> W: <code>lm(signed_rank(y2 - y1) ~ 1)</code>	<a href="#">≥14</a>	One intercept predicts the pairwise $y_2 - y_1$ differences. One intercept predicts the pairwise difference in the signed rank of $y_2 - y_1$ .	
	<b>y ~ continuous x</b> P: Pearson correlation N: Spearman correlation	W: <code>cor.test(x, y, method='Pearson')</code> W: <code>cor.test(x, y, method='Spearman')</code>	W: <code>lm(y ~ 1 + x)</code> W: <code>lm(rank(y) ~ 1 + rank(x))</code>	<a href="#">≥10</a>	x multiplied by a number (slope) predicts y. The rank of x multiplied by a number (slope) predicts the rank of y.	
	<b>y ~ discrete x</b> P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	W: <code>t.test(y1, y2, var.equal=TRUE)</code> W: <code>t.test(y1, y2, var.equal=FALSE)</code> W: <code>wilcox.test(y1, y2)</code>	L: <code>lm(y ~ 1 + G2)^A</code> L: <code>gls(y ~ 1 + G2, weights=...^B)^A</code> L: <code>lm(signed_rank(y) ~ 1 + G2)^A</code>	<a href="#">≥11</a>	One intercept per <b>group</b> (i.e., per x) predicts y. One intercept per <b>group</b> (i.e., per x) predicts y (but different variances). One intercept per <b>group</b> (i.e., per x) predicts the signed rank of y.	
Multiple regression: $\text{lm}(y \sim 1 + x_1 + x_2 + \dots)$	P: One-way ANOVA N: Kruskal-Wallis	L: <code>aov(y ~ group)</code> L: <code>kruskal.test(y ~ group)</code>	L: <code>lm(y ~ 1 + G2 + G3 + ... + GN)^A</code> L: <code>lm(rank(y) ~ 1 + G2 + G3 + ... + GN)^A</code>	<a href="#">≥11</a>	One intercept per <b>group</b> (i.e., per x) predicts y. One intercept per <b>group</b> (i.e., per x) predicts the signed rank of y.	
	P: One-way ANCOVA	L: <code>aov(y ~ group + x)</code>	L: <code>lm(y ~ 1 + G2 + G3 + ... + GN + x)^A</code>		One intercept per <b>group</b> plus x multiplied by a number (slope) predicts y. Note: this is discrete AND continuous. All ANCOVAs are ANOVAs with a continuous x.	
	P: Two-way ANOVA	L: <code>aov(y ~ group * sex)</code>	L: <code>lm(y ~ 1 + G2 + G3 + ... + GN + S2 + S3 + ... + SK + G2*S2+G3*S3+...+GN*SK)^A</code>		Interaction: changing <b>sex</b> changes the <b>y ~ group</b> parameters. Note: $G_{2 \text{ to } N}$ is an <b>indicator (0 or 1)</b> for each of N levels of the <b>group</b> variable except for the one modeled by the intercept (1). Similarly for $S_{2 \text{ to } K}$ for <b>sex</b> . Line 1 is main effect of <b>group</b> , line 2 for <b>sex</b> and line 3 is the <b>group X sex</b> interaction. For two levels (e.g. male/female sex), line 2 would just be " $S_2$ " and line 3 would be interactions between just $S_2$ and all $G_{2 \text{ to } N}$ .	[Coming]
	<b>Counts ~ discrete x</b> N: Chi-square test	M: <code>chisq.test(group * sex_table)</code>	<b>Equivalent log-linear R model</b> L: <code>glm(y ~ 1 + G2 + G3 + ... + GN + S2 + S3 + ... + SK + G2*S2+G3*S3+...+GN*SK, family=...)^A</code>		Interaction: changing <b>sex</b> changes the <b>y ~ group</b> parameters. Run glm using the following arguments: <code>glm(model, family=binomial(link='log'))</code> As linear-model, the Chi-square test is $\log(y_i) = \log(N) + \log(\alpha_i) + \log(\beta_j) + \log(\alpha_i\beta_j)$ where $\alpha_i$ and $\beta_j$ are proportions. See more info in <a href="#">the notebook</a> accompanying this table.	Same as Two-way ANOVA
	N: Goodness of fit	L: <code>chisq.test(y)</code>	L: <code>glm(y ~ 1 + G2 + G3 + ... + GN, ...)^A</code>		One intercept per <b>group</b> (i.e., x) predicts y.	1W-ANOVA

List of parametric (P) non-parametric (N) tests and equivalent linear models. The notation  $y \sim 1 + x$  is R shorthand for  $y = 1 \cdot b + a \cdot x$  which most learned in high-school. Models in similar colors are highly similar. Notice how little changes from line to line! For non-parametric models, the linear models are acceptable approximations for sample sizes in the "if N" column (empty = exact). Click links to see more details. Other less accurate approximations exist, e.g., Wilcoxon for sign test and Goodness-of-fit for binomial test. Some R commands require wide format data (W) with multiple values per row while others require long format (L) with one value per row. The signed rank function is `signed_rank = function(x) sign(x) * rank(abs(x))`. The variables  $G_i$  and  $S_i$  are "dummy coded" indicator variables (either 0 or 1) exploiting the fact that when  $\Delta x = 1$  between categories the difference equals the slope. Subscripts (e.g.,  $G_2$  or  $y_1$ ) indicate different columns in data.

<sup>A</sup> See the note to the two-way ANOVA for explanation of the notation.

<sup>B</sup> Same model, but with one variance per group: `gls(value ~ 1 + G2, weights = varIdent(form = ~1|group), method="ML")`.

