
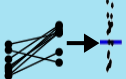
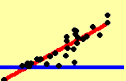
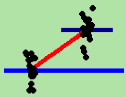
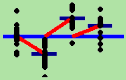
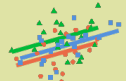


# Common statistical tests are linear models

Last updated: 28 June, 2019. Also check out the [Python version!](#)

See worked examples and more details at the accompanying notebook: <https://lindeloef.github.io/tests-as-linear>

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear 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	t.test(y) wilcox.test(y)	$\text{lm}(y \sim 1)$ $\text{lm}(\text{signed\_rank}(y) \sim 1)$	✓ <a href="#">for N &gt; 14</a>	One number (intercept, i.e., the mean) predicts <b>y</b> . - (Same, but it predicts the <i>signed rank</i> of <b>y</b> .)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y1, y2, paired=TRUE) wilcox.test(y1, y2, paired=TRUE)	$\text{lm}(y_2 - y_1 \sim 1)$ $\text{lm}(\text{signed\_rank}(y_2 - y_1) \sim 1)$	✓ <a href="#">for N &gt; 14</a>	One intercept predicts the pairwise <b>y<sub>2</sub>-y<sub>1</sub></b> differences. - (Same, but it predicts the <i>signed rank</i> of <b>y<sub>2</sub>-y<sub>1</sub></b> .)	
	<b>y ~ continuous x</b> P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	$\text{lm}(y \sim 1 + x)$ $\text{lm}(\text{rank}(y) \sim 1 + \text{rank}(x))$	✓ <a href="#">for N &gt; 10</a>	One intercept plus <b>x</b> multiplied by a number (slope) predicts <b>y</b> . - (Same, but with <i>ranked x</i> and <b>y</b> )	
	<b>y ~ discrete x</b> P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y1, y2, var.equal=TRUE) t.test(y1, y2, var.equal=FALSE) wilcox.test(y1, y2)	$\text{lm}(y \sim 1 + G_2)^A$ $\text{gls}(y \sim 1 + G_2, \text{weights}=\dots^B)^A$ $\text{lm}(\text{signed\_rank}(y) \sim 1 + G_2)^A$	✓ ✓ <a href="#">for N &gt; 11</a>	An intercept for <b>group 1</b> (plus a difference if <b>group 2</b> ) predicts <b>y</b> . - (Same, but with one variance <i>per group</i> instead of one common.) - (Same, but it predicts the <i>signed rank</i> of <b>y</b> .)	
Multiple regression: $\text{lm}(y \sim 1 + x_1 + x_2 + \dots)$	P: One-way ANOVA N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N)^A$ $\text{lm}(\text{rank}(y) \sim 1 + G_2 + G_3 + \dots + G_N)^A$	✓ <a href="#">for N &gt; 11</a>	An intercept for <b>group 1</b> (plus a difference if group $\neq 1$ ) predicts <b>y</b> . - (Same, but it predicts the <i>rank</i> of <b>y</b> .)	
	P: One-way ANCOVA	aov(y ~ group + x)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + x)^A$	✓	- (Same, but plus a slope on <b>x</b> .) <i>Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.</i>	
	P: Two-way ANOVA	aov(y ~ group * sex)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_K + G_2*S_2 + G_3*S_3 + \dots + G_N*S_K)$	✓	Interaction term: changing <b>sex</b> changes the <b>y ~ group</b> parameters. <i>Note: G<sub>2</sub> to G<sub>N</sub> is an indicator (0 or 1) for each non-intercept levels of the group variable. Similarly for S<sub>2</sub> to S<sub>K</sub> for sex. The first line (with G<sub>i</sub>) is main effect of group, the second (with S<sub>j</sub>) for sex and the third is the group * sex interaction. For two levels (e.g. male/female), line 2 would just be "S<sub>2</sub>" and line 3 would be S<sub>2</sub> multiplied with each G<sub>i</sub>.</i>	[Coming]
	<b>Counts ~ discrete x</b> N: Chi-square test	chisq.test(groupXsex_table)	<b>Equivalent log-linear model</b> $\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_K + G_2*S_2 + G_3*S_3 + \dots + G_N*S_K, \text{family}=\dots)^A$	✓	Interaction term: (Same as Two-way ANOVA.) <i>Note: Run glm using the following arguments: glm(model, family=poisson()) As linear-model, the Chi-square test is <math>\log(y_i) = \log(N) + \log(\alpha_i) + \log(\beta_j) + \log(\alpha_i\beta_j)</math> where <math>\alpha_i</math> and <math>\beta_j</math> are proportions. See more info in <a href="#">the accompanying notebook</a>.</i>	Same as Two-way ANOVA
	N: Goodness of fit	chisq.test(y)	$\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N, \text{family}=\dots)^A$	✓	(Same as One-way ANOVA and see Chi-Square note.)	1W-ANOVA

List of common 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 of us learned in school. Models in similar colors are highly similar, but really, notice how similar they *all* are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. 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. `lm` requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at <https://lindeloef.github.io/tests-as-linear>.

<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")`.

