

# 常见的统计检验是线性模型

英文版最后更新于 2019 年 6 月 28 日，中文版最后更新于 2019 年 8 月 31 日。也可查阅 [Python 版](#)！

从配套的笔记查看可运行例子和更多细节：  
英文版：<https://lindeloef.github.io/tests-as-linear>，  
中文版：<https://cosx.org>

	常见名称	R 内置函数	R 中的等价线性模型	精确近似	线性模型的文字解释	图标
单变量回归: $\text{lm}(y \sim 1 + x)$	<b>y 独立于 x</b> P: 单样本 t 检验 N: Wilcoxon 符号秩检验	<code>t.test(y)</code> <code>wilcox.test(y)</code>	$\text{lm}(y \sim 1)$ $\text{lm}(\text{signed\_rank}(y) \sim 1)$	✓ <a href="#">N &gt; 14</a>	从一个数字（截距，比如说平均值）来预测 y。 -（相同，但是它预测 y 的符号秩。）	
	P: 配对样本 t 检验 N: Wilcoxon 配对组检验	<code>t.test(y1, y2, paired=TRUE)</code> <code>wilcox.test(y1, y2, paired=TRUE)</code>	$\text{lm}(y2 - y1 \sim 1)$ $\text{lm}(\text{signed\_rank}(y2 - y1) \sim 1)$	✓ <a href="#">N &gt; 14</a>	从一个截距来预测配对 y2-y1。 -（相同，但是它预测 y2-y1 的符号秩。）	
	<b>y ~ 连续变量 x</b> P: Pearson 相关系数 N: Spearman 相关系数	<code>cor.test(x, y, method='Pearson')</code> <code>cor.test(x, y, method='Spearman')</code>	$\text{lm}(y \sim 1 + x)$ $\text{lm}(\text{rank}(y) \sim 1 + \text{rank}(x))$	✓ N > 10	从一个截距加上 x 乘以一个数值（斜率）来预测 y。 -（相同，但是使用了 x 和 y 的秩）	
	<b>y ~ 离散变量 x</b> P: 双样本 t 检验 N: Mann-Whitney U 检验	<code>t.test(y1, y2, var.equal=TRUE)</code> <code>t.test(y1, y2, var.equal=FALSE)</code> <code>wilcox.test(y1, y2)</code>	$\text{lm}(y \sim 1 + G2)^A$  $\text{glm}(y \sim 1 + G2, \text{weights}=\dots^B)$ $\text{lm}(\text{signed\_rank}(y) \sim 1 + G2)^A$	✓ ✓ N > 11	从第 1 组的截距（如果是第 2 组的话加多一个相差值）来预测 y。 -（相同，但是每一组用不同的方差，而不是用着共同的方差。） -（相同，不过这个预测了 y 的符号秩。）	
多变量回归: $\text{lm}(y \sim 1 + x2 + \dots)$	P: 单因素方差分析 (one-way ANOVA) N: Kruskal-Wallis 检验	<code>aov(y ~ group)</code> <code>kruskal.test(y ~ group)</code>	$\text{lm}(y \sim 1 + G2 + G3 + \dots + Gn)^A$ $\text{lm}(\text{rank}(y) \sim 1 + G2 + G3 + \dots + Gn)^A$	✓ N > 11	从第 1 组的截距（如果不是第 1 组的话加多一个相差值）来预测 y。 -（相同，不过这个预测了 y 的秩。）	
	P: 单因素协变量分析 (one-way ANCOVA)	<code>aov(y ~ group + x)</code>	$\text{lm}(y \sim 1 + G2 + G3 + \dots + Gn + x)^A$	✓	-（相同，不过加上了 x 的斜率。） 注意：这里是离散和连续的混合情况。单因素协变量分析是单因素方差分析加上一个连续的 x。	
	P: 双因素方差分析 (two-way ANOVA)	<code>aov(y ~ group * sex)</code>	$\text{lm}(y \sim 1 + G2 + G3 + \dots + Gn + S2 + S3 + \dots + Sk + G2 * S2 + G3 * S3 + \dots + Gn * Sk)$	✓	交互项: sex 的变化会导致 y ~ group 参数的变化。 注意: G2 ... Gn 是 group 变量的每一个非截距的可能取值的示性 (0 或 1) 变量。 同理地, S2 ... Sk 是 sex 变量的示性变量。 第一条线 (Gi) 是 group 变量的主要效应, 第二条线 (Si) 是 sex 变量的主要效应, 第三条线是 group * sex 的交互效应。 对于两种可能取值范围 (如男女性别) 情况, 第 2 条线则是 S2, 而第 3 条线是 S2 乘以每个 Gi。	(待绘制)
	计数 ~ 离散 x N: 卡方检验	<code>chisq.test(groupXsex_table)</code>	等价的 <del>对数</del> 线性模型 $\text{glm}(y \sim 1 + G2 + G3 + \dots + Gn + S2 + S3 + \dots + Sk + G2 * S2 + G3 * S3 + \dots + Gn * Sk, \text{family}=\dots)^A$	✓	交互项: (和双因素方差分析一致) 注意: 使用以下参数运行 glm 函数: <code>glm(model, family=poisson())</code> 。 对于线性模型, 卡方检验是 $\log(y_{ij}) = \log(N) + \log(\alpha_i) + \log(\beta_j) + \log(\alpha_i \beta_j)$ , 其中 $\alpha_i$ 和 $\beta_j$ 是比率。建议查阅配套笔记获得更多信息。	和双因素方差分析一致
	N: 拟合优度检验	<code>chisq.test(y)</code>	$\text{glm}(y \sim 1 + G2 + G3 + \dots + Gn, \text{family}=\dots)^A$	✓	(和单因素方差分析一致, 建议查阅卡方检验笔记。)	和单因素方差分析一致

常见的参数 (P, parametric) 和非参 (N, non-parametric) 检验，以及等价的线性模型。记号  $y \sim 1 + x$  是 R 对于大部分我们在学校学习的  $y = 1 \cdot b + a \cdot x$  的快捷表达方式。相似颜色的模型本身也非常相似，真的，你们可以看看它们的一些颜色其实是相同的！就非参模型而言，对于样本量不小的情况，线性模型是足够好的近似了（见“精确近似”一列，点击链接查看对应模拟）。其余没那么精确的近似也是存在的，比如说 Wilcoxon 检验和符号检验，以及拟合优度检验和二项检验。符号秩函数的定义是 `signed_rank = function(x) sign(x) * rank(abs(x))`。变量 Gi 和 Si 是示性变量（取值只能是 0 或 1），揭示出当类别之间  $\Delta x = 1$  的时候，差值等于斜率。下标（如 G2 或 y1）表示数据中的不同列。lm 对于所有非连续变量模型都需要长格式数据。可以在 <https://lindeloef.github.io/tests-as-linear>（英文版）或 <https://cosx.org>（中文版）里找到以上内容的进一步解释和可运行例子。

<sup>A</sup>查看双因素方差分析笔记获取记号的解释。  
<sup>B</sup>相同模型，但是每个组有自己的方差：`glm(value ~ 1 + G2, weights = varIdent(form = ~1|group), method="ML")`。



原作者：Jonas Kristoffer Lindeløf  
<https://lindeloef.net>