

基礎統計分析皆是線性模型

請參考原作者 [github 主頁](https://github.com/lindeloev)，取得更多範例與詳細說明：

<https://lindeloev.github.io/tests-as-linear>

	分析方法	R 內建函式程式碼	線性模型表示式	可用條件	線型模型說明	圖例
簡單迴歸 $\text{lm}(y \sim 1 + x)$	y 獨立於 x P: 單一樣本 t 檢定 N: Wilcoxon 符號檢定	<code>t.test(y)</code> <code>wilcox.test(y)</code>	$\text{lm}(y \sim 1)$ $\text{lm}(\text{signed_rank}(y) \sim 1)$	✓ for N > 14	以單一參數 (截距 · 如平均值) 預測 y - (同上 · y 的預測值是名義或序列數值)	
	P: Paired-samplet 檢定 N: Wilcoxon 匹配檢定	<code>t.test(y1, y2, paired=TRUE)</code> <code>wilcox.test(y1, y2, paired=TRUE)</code>	$\text{lm}(y_2 - y_1 \sim 1)$ $\text{lm}(\text{signed_rank}(y_2 - y_1) \sim 1)$	✓ for N > 14	以單一截距預測 y₂-y₁ - (同上 · y₂-y₁ 的預測值是名義或序列數值)	
	y ~ x (x 是連續變項) P: 皮爾森相關 N: 斯皮爾曼相關	<code>cor.test(x, y, method='Pearson')</code> <code>cor.test(x, y, method='Spearman')</code>	$\text{lm}(y \sim 1 + x)$ $\text{lm}(\text{rank}(y) \sim 1 + \text{rank}(x))$	✓ for N > 10	以單一截距加上 x 的加權 (乘以斜率) 預測 y . - (同上 · x 與 y 都是名義或序列數值)	
	y ~ x (x 是類別變項) P: Two-samplet 檢定 P: Welch 氏 t 檢定 N: Mann-Whitney U	<code>t.test(y1, y2, var.equal=TRUE)</code> <code>t.test(y1, y2, var.equal=FALSE)</code> <code>wilcox.test(y1, y2)</code>	$\text{lm}(y \sim 1 + G_2)^A$ $\text{gls}(y \sim 1 + G_2, \text{weights}=\dots^B)^A$ $\text{lm}(\text{signed_rank}(y) \sim 1 + G_2)^A$	✓ ✓ for N > 11	以單一截距代表 第 1 組 (含與 第 2 組 的差異) 預測 y - (同上 · 以各組變異數取代單一變異數) - (同上 · y 的預測值是名義或序列數值)	
多元迴歸 $\text{lm}(y \sim 1 + x_1 + x_2 + \dots)$	P: 單因子變異數分析 N: Kruskal-Wallis 檢定	<code>aov(y ~ group)</code> <code>kruskal.test(y ~ group)</code>	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N)^A$ $\text{lm}(\text{rank}(y) \sim 1 + G_2 + G_3 + \dots + G_N)^A$	✓ for N > 11	以單一截距代表 第 1 組 (含組間差異不等於 1) 預測 y - (同上 · y 的預測值是名義或序列數值)	
	P: 單因子共變數分析	<code>aov(y ~ group + x)</code>	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + x)^A$	✓	- (同上 · 增加 x 的斜率) 註: 共變數分析比起變異數分析 · 多出一個連續變項 x	
	P: 二因子變異數分析	<code>aov(y ~ group * sex)</code>	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_K + G_2 * S_2 + G_3 * S_3 + \dots + G_N * S_K)$	✓	G_N*S_K : 組間差異 y ~ group 在性別(sex)之間的差異 註: 除了截距(1)的其他項目 · G_{2 to N} · S_{2 to K} 表示各組別及性別的 虛擬變項 (數值為 0 或 1) · 模型表示式第一行集合組間主要效果 G_i · 第二行集合性別的主要效果 S_i · 第三行集合兩者的所有交互作用項目 group * sex · 各階層相互對應 · 如第二行的 S₂ · 對應第三行的 S₂ 與各組別 G_i 之交交互作用項目。	[製作中]
	y ~ x (y 是計數; x 是類別變項) N: 卡方檢定	<code>chisq.test(groupXsex_table)</code>	等價的對數線性模型 $\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_K + G_2 * S_2 + G_3 * S_3 + \dots + G_N * S_K, \text{family}=\dots)^A$	✓	G_N*S_K : (同二因子變異數分析) 註: <code>glm</code> 函式參數格式 <code>glm(model, family=poisson())</code> 卡方檢定的完整線性模型表示式是 $\log(y_i) = \log(N) + \log(\alpha_i) + \log(\beta_i) + \log(\alpha_i \beta_i)$ α_i 與 β_i 都是比例 · 詳見本網頁說明。	同二因子變異數分析
	N: 適合度檢定	<code>chisq.test(y)</code>	$\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N, \text{family}=\dots)^A$	✓	(同單因子共變數分析 · 其他資訊請參考卡方檢定內註)	同單因子變異數分析

基礎母數方法(P)與無母數方法(N)以及相等功能的線性模型。R 程式碼 $y \sim 1 + x$ 等同許多統計課程教授的迴歸式 $y = 1 \cdot b + a \cdot x$ 。同色欄位內的模型有許多相似處，但請勿必了解其中細節！留意無母數方法，線性模型僅是合理逼近一定樣本數以上的狀況 (請參考“可用條件”一欄 · 點擊超連結可見模擬結果)。某些方法並不完全符合線性迴歸 · 例如 Wilcoxon 符號檢定與二項式檢定的適合度考驗。部分 R 程式碼裡有作者自建函式 `signed_rank = function(x) sign(x) * rank(abs(x))`。變項符號 **G_i** 與 **S_i** 都是「**虛擬變數**」(dummy variables)。數值為 0 或 1 · 數值間的差異表示自變數內的類別差異等於斜率。下標記號 (例如 **G₂** 或 **y₁**) 表示資料裡的各欄位。R 函式 `lm` 必須匯入長表格(long-format)資料 · 才能處理有非連續變項的模型。詳細說明與範例請參考原作者網頁 <https://lindeloev.github.io/tests-as-linear>

^A 請參考二因子 ANOVA 的線型模型說明。

^B 與此功能相同的程式碼 · 但加權係數根據各組變異數 `gls(value ~ 1 + G2, weights = varIdent(form = ~1|group), method="ML")`

