

ANOVA & GLM(Univariate) & Linear Regression

	One-Way ANOVA	GLM → Univariate	Linear Regression
统计本质	线性模型	线性模型	线性模型
自变量	离散型 (1 个)	离散型 (≥ 1 个)	连续型
虚拟变量	隐式编码	隐式编码	---
F 检验	组间效应	主效应/简单效应/交互作用	回归模型
结果是否等价	---	One-Way ANOVA	依赖虚拟变量编码 ¹

DataSet: 方差分析-单因素方差分析-cholesterol.sav.

¹:Dummy Variable Encoding

	D1(1time?)	D2(2times?)	D3(4times?)	D4(drugD?)
1time	1	0	0	0
2times	0	1	0	0
4times	0	0	1	0
drugD	0	0	0	1
drugE	0	0	0	0

Note: drugE is reference group / baseline category;

$$\text{Model: } Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \varepsilon$$

β_0 : Mean of reference group, β_j : the mean difference between group j and the reference group.

Summary:

1. ANOVA = Regression with Dummy Variables, 一般线性模型可自动完成虚拟变量编码;
2. 单因素方差分析比较的是不同组的均值; 一般线性模型把这种比较转换成一个线性模型; 而线性回归中, 分组变量必须通过虚拟变量编码转化为数值形式。只要虚拟变量编码正确, 这三种方法在模型层面是完全等价的。