

多应变量数据分析

例 1：如下表数据，分析两组（GROUP=0、1）INFO、SIMIL、ARITH、PICT 的差异。

GROUP	INFO	SIMIL	ARITH	PICT
0	12	13	16	9
0	13	10	17	9
0	16	18	11	9
0	8	3	7	9
0	6	3	13	9
0	11	8	10	10
0	12	7	9	8
0	8	11	9	3
0	14	12	11	4
0	13	13	13	6
0	13	9	10	9
0	13	10	15	7
0	14	11	12	8
0	15	11	11	10
0	13	10	15	9
0	10	5	8	6
0	10	3	7	7
0	17	13	13	7
0	10	6	10	7
0	10	10	15	8
0	14	7	11	5
0	16	11	12	11
0	10	7	14	6
0	10	10	9	6
0	10	7	10	10
0	15	12	10	6
0	13	15	15	8
0	7	6	5	9
1	9	5	10	8
1	8	8	6	6
1	9	5	9	8
1	10	0	6	2
1	11	9	11	1

1	13	7	12	9
1	11	9	9	8
1	5	3	3	6
1	9	7	8	6
1	7	2	6	4
1	12	9	14	3
1	13	12	11	10

首先在 EXCEL 里把数据按如下格式整理好：

- 第一行为变量名，第二行开始为数据。所有变量应为同一类型，连续型或 0/1 两分类型。
- 如有分层变量，如上表 GROUP，分层变量置第一列。
- 如有缺失值，用 NA 表示（无引号）。

从 EXCEL 里复制粘贴数据(制表符分隔)到数据输入框中，如下图所示：

多应变量分析

标题： 输出文件名： 输出路径(置空表示到我的文档; "."表示到当前工作目录)：

复制粘贴数据到下表：

- 第一行为应变量名，第二行开始为每个对象观察值，每行代表一个观察对象；分组变量置第一列，其后每列代表一个应变量(Y)
- 所有应变量(Y)应为同一类型，连续型或0/1两分类型。缺失数据置空或用NA表示，数据不能含有其它字符或引号。

	A	B	C	D	E	F
1	GROUP	INFO	SIMIL	ARITH	PICT	
2	0	12	13	16	9	
3	0	13	10	17	9	
4	0	16	18	11	9	
5	0	8	3	7	9	
6	0	6	3	13	9	
7	0	11	8	10	10	
8	0	12	7	9	8	
9	0	8	11	9	3	
10	0	14	12	11	4	
11	0	13	13	13	6	
12	0	13	9	10	9	
13	0	13	10	15	7	
14	0	14	11	12	8	
15	0	15	11	11	10	
16	0	13	10	15	9	

- 所有观察值为0/1两分类型 将分组因素按连续性变量分析

点击开始分析。

输出结果与解释：

Data:

GROUP	INFO	SIMIL	ARITH	PICT
0	12	13	16	9
0	13	10	17	9
.....				
1	12	9	14	3
1	13	12	11	10

解释：首先列出数据，以核实数据读入是否正确。

GEE 多元回归分析

每个应变量有一个回归系数模型

	回归系数	标准误	95%区间下限	95%区间上限	P 值
(Intercept)	11.892857	0.523283	10.867223	12.918491	0.000000
factor(x.v1) 1	-2.142857	0.847067	-3.803109	-0.482605	0.011415
factor(x.v2) 1	-2.988095	1.128640	-5.200230	-0.775960	0.008109
factor(x.v3) 1	-2.607143	1.067885	-4.700197	-0.514088	0.014630
factor(x.v4) 1	-1.761905	0.860968	-3.449403	-0.074407	0.040715
factor(v) 2	-2.571429	0.480297	-3.512811	-1.630047	0.000000
factor(v) 3	-0.535714	0.602889	-1.717376	0.645948	0.374229
factor(v) 4	-4.214286	0.611032	-5.411908	-3.016663	0.000000

一个总体回归系数模型

	回归系数	标准误	95%区间下限	95%区间上限	P 值
(Intercept)	11.962500	0.480120	11.021465	12.903535	0.000000
factor(x) 1	-2.375000	0.709684	-3.765981	-0.984019	0.000818
factor(v) 2	-2.825000	0.415388	-3.639161	-2.010839	0.000000
factor(v) 3	-0.675000	0.453650	-1.564154	0.214154	0.136769
factor(v) 4	-4.100000	0.525729	-5.130428	-3.069572	0.000000

两模型比较（对交互作用的 Wald 检验）*

Df	X2	P 值
3	1.26941718254496	0.736407758490737

*检验 GROUP 对不同应变量（INFO SIMIL ARITH PICT）的回归系数是否有显著性差异
适合用一个回归系数来表示 GROUP 的总体效应

解释：上面首先用一个模型（每个应变量有一个回归系数的模型）估计每个应变量的 GROUP 组间差异，得 4 个回归系数，即 factor(x.v1)1、factor(x.v2)1、factor(x.v3)1、factor(x.v4)1，其中 V1、V2、V3、V4 分别表示 4 个应变量（INFO SIMIL ARITH PICT），后面的 1 表示 x=1，即 GROUP =1，GROUP=0 组作为参照。然后把这四个应变量当成 4 次重复测量，用 GEE 估计一个总体的差异（一个总体回归系数模型），得总体回归系数即 factor(x)1 项的回归系数，比较两个模型以判断 4 个应变量的组间差异是否有显著差异，这里得出的 p=0.7364。表示适合用一个回归系数来表示 GROUP 的总体效应。

下面是多变量反应曲线图比较结果及图形输出结果

每单个变量的 t 检验

	N1	均数 1	N2	均数 2	t	自由度	P 值
INFO	28	11.8928571428571	12	9.75	2.29233538918101	38	0.0275148955935611
SIMIL	28	9.32142857142857	12	6.33333333333333	2.40982323607117	38	0.0209096384340342
ARITH	28	11.3571428571429	12	8.75	2.50893337065323	38	0.0164934806784762
PICT	28	7.67857142857143	12	5.91666666666667	2.2775615361973	38	0.0284662107891478

Hotelling T 平方检验(检验均数是否相同)

N1	N2	T 平方	F	自由度 1	自由度 2	P 值
28	12	11.9505339332479	2.75176768199788	4	35	0.0433174128858097

检验平行性

N1	N2	T 平方	F	自由度 1	自由度 2	P 值
28	12	1.17952481097114	0.372481519254044	3	36	0.773328114510886

检验一致性

N1	N2	T 平方	F	自由度 1	自由度 2	P 值
28	12	10.5619798350138	10.5619798350138	1	38	0.00241965706967484

检验反应曲线是否呈水平性

总 SS	组间 SS	自由度 1	组间 MS	组内 SS	自由度 2	组内 MS	F	P 值
1934.4	432.25	3	144.08333	1502.15	156	9.62916667	14.9632	1.31057e-08

