

直线回归与相关分析

例 1: 如下表数据, 分析 B、T、Z 与 Y 的关系

Y	B	T	Z
47.4	1	1	55.4
59.2	1	2	58.0
10.1	1	3	7.2
62.3	1	1	47.6
59.0	1	2	56.6
27.2	1	3	30.3
54.1	1	1	53.9
40.1	1	2	35.7
29.0	1	3	NA
57.7	2	1	60.1
57.7	2	2	58.5
11.1	2	3	13.4
59.8	2	1	59.2
56.3	2	2	56.0
34.8	2	3	34.8
56.3	2	1	58.1
46.7	2	2	46.3
39.2	2	3	33.1

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。第一行变量如为分类变量, 在其变量名后加 * (如 B*, T*) 表示该变量为分类变量, 否则按自动连续性变量分析。
- 如要做回归分析, 应变量 Y 放第一列 (如无分层变量), 如有分层变量, 分层变量置第一列, 后面紧接着 Y 变量, 其后每列代表一个自变量。
- 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:

两(多)组数据相关与回归分析

标题:

相关与回归分析

输出文件名:

corglm

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:

- 第一行为变量名, 第二行开始为数据。变量名后如加*表示分类变量, 否则按连续性变量分析;
- 如作回归分析, 应变量Y放前面, 后面每列代表一个自变量;
- 如有分层变量, 置第一列。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
1	Y	B	T*	Z	
2	47.4	1	1	55.4	
3	59.2	1	2	58.0	
4	10.1	1	3	7.2	
5	62.3	1	1	47.6	
6	59.0	1	2	56.6	
7	27.2	1	3	30.3	
8	54.1	1	1	53.9	
9	40.1	1	2	35.7	
10	29.0	1	3	NA	
11	57.7	2	1	60.1	
12	57.7	2	2	58.5	
13	11.1	2	3	13.4	
14	59.8	2	1	59.2	
15	56.3	2	2	56.0	
16	34.8	2	3	34.8	

分组变量:

第一列为分组因素

分析内容:

- 相关分析
- 回归分析
- 散点曲线图

开始计算

点击开始分析。

输出结果与解释:

数据

Y	B*	T*	Z
47.4	1	1	55.4
59.2	1	2	58
.....
39.2	2	3	33.1

解释: 首先列出数据, 以核实数据读入是否正确。

相关系数

(spearman)相关系数与 P 值

p\r	Y	B*	T*	Z
Y	NA	0.107417409360943	-0.785136895187755	0.95857668692853
B*	0.671385664719187	NA	0	0.109576738057729
T*	0.000113453278021947	1	NA	-0.779113684407933
Z	1.38578126751554e-09	0.675473541180662	0.000227782276362116	NA

解释：上表以中间对角线分成两部分，右上三角是相关系数，左下是 p 值。表内计算的是 spearman 相关系数（区别于 pearson 相关系数）。对角线是每个变量自身的相关性分析（应该为 1），这里用 NA 略去。

直线回归

单因素分析				多元分析			
自变量	回归系数	标准误	p 值	自变量	回归系数	标准误	p 值
factor(B*) 2	3.4667	8.0215	0.6714	factor(B*) 2	-1.0361	2.7584	0.7137
factor(T*) 2	-3.1	5.123	0.5541	factor(T*) 2	0.2654	3.1039	0.9333
factor(T*) 3	-31.0333	5.123	0	factor(T*) 3	-3.869	6.2922	0.5501
Z	0.963	0.0739	0	Z	0.8704	0.1665	2e-04

模型 R 平方

模型	R 平方	调整的 R 平方
B*	0.0115	-0.0502
T*	0.748	0.7144
Z	0.9189	0.9135
多元模型	0.9249	0.8999

解释：

1. 上表单因素分析模型，即每个自变量与 Y 的单独分析模型，这里有 3 个自变量，即有三个模型的结果。如 T 与 Y 的单因素分析：T 是 3 分类变量，以第一类（T=1）为参照，生成了 2 个哑变量分别表示 T=2 与 T=3，放入模型中，看出 T=2 比 T=1 组，回归系数为-3.1，即 Y 低 3.1，p=0;5541 差异不显著； T=3 比 T=1 组，回归系数-

31.0333, $p=0$ (<0.00001)有非常显著的差异。Z 是连续性变量，回归系数 0.963，表示 Z 每增加 1, Y 增加 0.963, 非常显著的变化。

2. 多因素模型是把 B、T、Z 同时放到一个多元回归模型中，得出 3 个变量的回归系数。看出 B、T 的作用都不显著了。
3. 模型的 R 平方，列出了上述的 3 个单因素模型与 1 个多元模型的 R 平方及调整的 R 平方值。
4. 下面是散点与回归线图输出结果。

