

SPSS——相关分析



K房K事

得之我幸，失之非命！

18 人赞同了该文章

相关分析简介

- 相关分析的指标体系：对于任何类型的变量，都可用相应的指标进行相关关系的考察，测量相关程度的相关系数有很多。
- (1)、连续变量的相关指标：

一般使用积差相关系数，即**Pearson相关系数**表示。其数值介于-1~1之间。当两个变量间的相关性达到最大，散点呈一条直线是取值为±1，正负号表示相关的方向。如果两变量无关，取值为0。

积差相关系数严格讲只适用于两变量呈线性关系，其有一定适用条件。当数据不满足适用条件可考虑使用**Spearman等级相关系数**来解决。

- (2)、有序变量的相关指标：

往往称为“一致性”。指行变量等级高的列变量等级也高，行变量等级低的列变量等级也低。两个指标的含义：当按两个变量的取值列出交叉表后，**P** 代表两倍的一致对子数，**Q** 代表两倍不一致的对子数。一致对子数就是指行变量等级高的列变量等级也高。统计量：

Gamma统计量：描述有序分类数据联系强度的度量，介于-1~1之间。 $\gamma = \frac{P - Q}{P + Q}$

Kendall's Tau_b: $\tau = \frac{P - Q}{\frac{n(n-1)}{2}}$ ，在±1之间。在“相关”-“双变量”中也有此值。

Kendall's Tau_c: 在Kendall's Tau_b校正

Somers"d: 略

- (3)、名义变量的相关指标：（由卡方衍生）

列联/相依系数 (contingency coefficient)：值0~1，越大表明两变量间相关性越强。

Phi和Cramer's V: phi的指标的绝对值越大，相关性越强。V的取值0~1之间

λ 系数：反映自变量对因变量的预测效果。

▲ 赞同 18 ▼

● 2 条评论

- SPSS中相应的功能

(1)、交叉表：统计量的子对话框



(1) **相关性 (R)** 复选框：适用于两个连续性变量的分析，计算行列变量的**Pearson**相关系数和**Spearman**相关系数。

(2) **“按区间标定”**：包含一个变量为数值变量，另一个变量为分类变量时度量两者关联度的指标。Eta的平方表示“由组间差异所解释的因变量的方差的比例”。即 $\frac{S_{\text{组间}}}{S_{\text{总}}}$ ，一共给出两个Eta

值：行变量为因变量和列变量为因变量的情况。

(3) **“有序”**复选框：反映分类变量一致性的指标，只能在两个变量均属于**“有序分类”**时使用。

(4) **“名义”**复选框：反映分类变量相关性的指标，在**有序和无序**分类是均可使用。但两变量为有序时，采用**“有序”**复选框。

(5) **“Kappa”**：计算内部一致性系数。

(6) **“风险”**：计算OR值（比数比）和RR值（相对危险度）

(2) **“相关”子菜单**:(针对连续性变量的相关分析常用)

双变量 (bivariate) 过程：进行**“两个/多个变量间的参数/非参数相关分析”**。多个变量给出两两相关分析的结果。

偏相关 (Partial) 过程：对其他变量进行控制，输出控制其他变量影响后的相关系数。

距离(Distance)过程：多用于因子分析、聚类分析和多维尺度分



简单相关分析

1. 方法原理：

- (1) 直线相关：两变量呈线性关系；
- (2) 曲线相关：存在相关趋势，但并非线性；
- (3) 正相关和负相关：两变量增加趋势是否一致；
- (4) 完全相关：相关程度亲密无间的程度。

注：当变量为有序变量或者名义变量时，一般不考虑直线、曲线相关问题。

2. 系数计算：

$$r = \frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}} = \frac{\sum_{i=1}^n \frac{(X-\bar{X})(Y-\bar{Y})}{n-1}}{\sqrt{\sum_{i=1}^n (X-\bar{X})^2/n-1} \sqrt{\sum_{i=1}^n (Y-\bar{Y})^2/n-1}}$$

r 取值-1~1, $r>0$, 正相关; $r<0$, 负相关; $|r|$ 接近1, 相关性好, 接近0, 相关性差。

3. 相关系数的检验方法：

相关系数计算后需对其进行检验, $H_0: \rho = 0$, 两变量间无直线相关关系。主要是T检验。SPSS会给出相关系数值和最终的P值。

4. 积差相关系数的适用条件：

首先：考虑两变量是否可能存在相关关系。

- (1) 积差相关系数只适用线性相关的情形，对曲线相关等并不适用
- (2) 样本极端值对积差相关系数的影响较大。要慎重，剔除或者变换
- (3) 积差相关系数要求相应变量呈双变量正态分布。（较宽松）

其次：做散点图

- (1) 看出是否存在相关趋势；
- (2) 呈现线性趋势还是曲线趋势，是否可直接使用线性相关的积差相关系数；
- (3) 是否有明显极端点。

案例：考察总信心指数值和年龄的相关性。

首先：变量均为连续变量，考虑使用两个连续性变量的相关性指标。

其次：绘制散点图（略）

➔ 非参数检验

假设检验汇总

	原假设	检验	显著性	决策者
1	S3. 年龄 的分布为正态分布，平均值为 36.358，标准偏差为 12.86。	单样本 Kolmogorov-Smirnov 检验	.000 ¹	拒绝原假设。
2	总指数 的分布为正态分布，平均值为 95.894，标准偏差为 21.00。	单样本 Kolmogorov-Smirnov 检验	.000 ¹	拒绝原假设。

显示渐进显著性。 显著性水平为 .05。

¹已校正 Lilliefors

正态性检验

两则均不为正态，但由于条件3可放宽。因而采用Pearson和Spearman双重检验。

步骤：“分析”——“相关”——“双变量相关性”



结果：



相关性

		S3. 年龄	总指数
S3. 年龄	Pearson 相关性	1	-.219**
	显著性（双尾）		.000
	N	1147	1147
总指数	Pearson 相关性	-.219**	1
	显著性（双尾）	.000	
	N	1147	1147

** . 在置信度（双测）为 0.01 时，相关性是显著的。

pearson相关性结果

非参数相关性

相关性

		S3. 年龄	总指数
斯皮尔曼等级相关系数	S3. 年龄 相关系数	1.000	-.213**
	显著性（双尾）	.	.000
	N	1147	1147
总指数	相关系数	-.213**	1.000
	显著性（双尾）	.000	.
	N	1147	1147

** . 相关性在 0.01 级别显著（双尾）。

Spearman相关性结果

可以看出：给出相关系数、P值、样本数。相关pearson系数为-0.219，Spearman相关系数为-0.213.且P值<0.05.所以：两变量间的负相关是有统计学意义的，随着年龄的增加，总指数呈现减少的趋势。

注：**Spearman等级相关系数**，又称秩相关系数，是利用两变量的秩次大小进行线性相关分析的，对原始变量的分布不做要求，属于非参数统计方法。

Kendall等级相关系数：在“交叉表”——“统计量”中。用于反映分类变量的相关性的指标，适用于两个变量均为有序分类的情况。

注意：秩相关系数和等级相关系数（均丢失变量某些信息）的绝对值都小于积差相关系数。

偏相关分析

▲ 赞同 18 ▼

● 2 条评论

分析变量间的关系是，在计算积差相关系数，秩相关系数（spearman系数）、kendall等级相关系数是都没考虑第三方的影响。偏相关分析指在相关的基础上考虑两因素以外的各种影响因素，重新来考察两因素的关联程度。

偏相关系数计算公式：
$$r_{12(3)} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \times \sqrt{1 - r_{23}^2}}$$

r_{ij} 为变量i和j的相关系数。

2. 案例：控制家庭收入的影响后考察年龄对总指数的作用

首先：查看三个变量之间的相关性。



添加3个变量



首发于
SPSS 数据分析

相关性				
		S3. 年龄	总指数	Qs9
S3. 年龄	Pearson 相关性	1	-.219**	-.138**
	显著性 (双尾)		.000	.000
	N	1147	1147	992
总指数	Pearson 相关性	-.219**	1	.084**
	显著性 (双尾)	.000		.008
	N	1147	1147	992
Qs9	Pearson 相关性	-.138**	.084**	1
	显著性 (双尾)	.000	.008	
	N	992	992	992

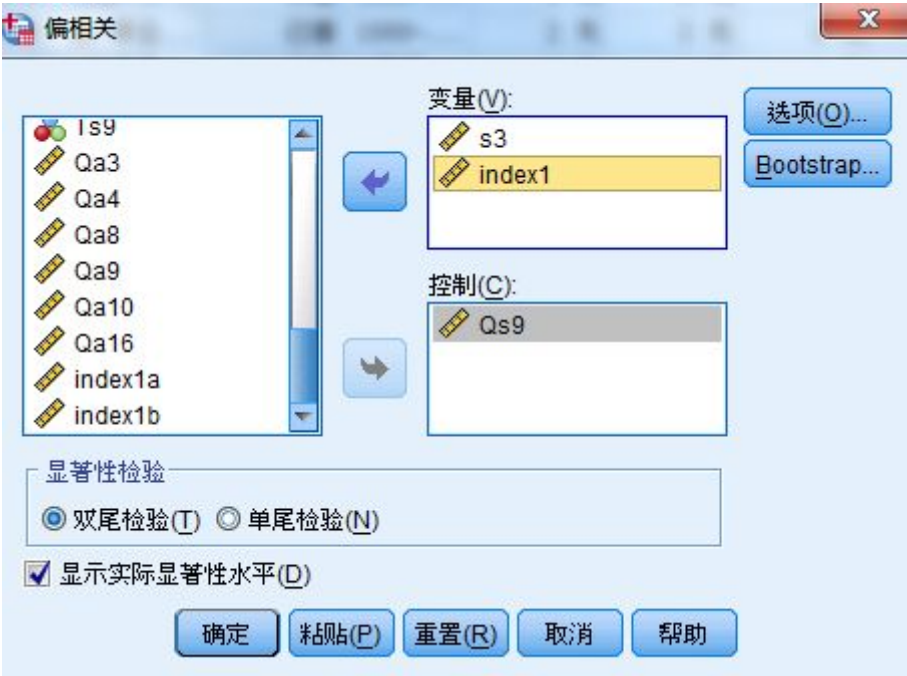
** . 在置信度 (双测) 为 0.01 时, 相关性是显著的。

两两之间的相关性

结果显示：两两相关性都具有统计学意义。

其次：采用偏相关进行分析。

“分析”——“相关”——“偏相关”



偏相关输入

结果：



首发于
SPSS 数据分析

相关性

控制变量			S3. 年龄	总指数
Qs9	S3. 年龄	相关性	1.000	-.216
		显著性（双侧）	.	.000
		df	0	989
总指数	相关性	相关性	-.216	1.000
		显著性（双侧）	.000	.
		df	989	0

偏相关分析的结果

控制变量Qs9后，年龄和总指数之间偏相关系数为-0.216，且具有统计学意义。在控制家庭收入后，年龄和总指数之间仍存在负相关性。

Distance过程

不是很懂这块，待续。。。就是没有续集的意思

发布于 2017-11-20

SPSS 数据分析

文章被以下专栏收录



SPSS 数据分析

进入专栏

推荐阅读



知乎



首发于
SPSS 数据分析

步骤1：确认待分析的原始变量之间是否存在较强的相关关系。可采用计算“相关系数矩阵”“...

K房K事 发表于SPSS ...



SPSSAU教程03
指标解读

2 条评论

切换为时间排序

写下你的评论...



夏天的胖子

1 年前



赞



SPSSAU

17 天前

偏相关分析挺好的 但好像一般用的不多

赞

赞同 18



2 条评论