

利用 SPSS 软件分析变量间的相关性

孙逸敏*

(新疆兵团广播电视大学, 新疆 乌鲁木齐 830001)

摘要:通过描述相关分析的研究意义、概念、分类以及利用 SPSS 软件进行变量间相关性分析的举例, 说明了利用 SPSS 软件分析变量间的相关性, 可以使我们了解事物之间的密切程度, 并给事物之间的关系做出定性的分析。

关键词:相关分析; SPSS 软件; 二变量过程; 相关系数; 显著性差异

中图分类号:TP3

文献标识码:A

文章编号:1008—3588(2007)02—0120—04

一、引言

任何事物的存在都不是孤立的, 而是相互联系、相互制约的。为了描述这些事物之间的关系, 人们用变量来表示事物的数量特征, 这些事物之间的关系可以转化为变量之间的关系。通常我们把变量之间的关系归纳为两种类型: 函数关系和统计关系。

函数关系是一一对应的确定性关系即线性关系, 比较容易分析和测量, 可是在现实生活中, 变量之间的关系往往并不那么简单。比如, 职业种类和收入、家庭收入和支出、一个人所受教育程度预期收入、子女身高和父母身高等, 它们之间确定存在某种关系, 但这些关系却无法像函数那样能够用一个确定的函数公式来描述。这样, 一个变量的值不能由另一个变量的值来唯一确定, 这种关系叫做统计关系。事物之间的统计关系不像函数关系那样直接, 但通过大量数据的观察和研究, 可以发现这些变量之间确实存在着某种统计关系, 有的关系强, 有的关系弱, 程度各有差异。那么, 如何测度事物间统计关系的强弱程度呢? 相关分析为我们提供了答案。

二、什么是相关分析

衡量事物之间或称变量之间线性相关程度的强弱, 并用适当的统计指标表示出来, 这个过程就是相关分析。它是研究变量间密切程度的一种常用统计方法。根据变量之间的不同情况, 它分为三类:

1. 线性相关分析。研究两个变量间线性关系

的程度。相关系数是描述这种线性关系程度和方向的统计量, 用 r 来描述。

如果变量 Y 与 X 间是函数关系, 则 $r = 1$ 或 $r = -1$; 如果变量 Y 与 X 间是统计关系, 则 $-1 < r < 1$, 如果 x, y 变化的方向一致, 如身高与体重的关系, 则称为正相关, 则 $r > 0$; 如果 x, y 变化的方向相反, 如吸烟与肺功能的关系, 则称为负相关, 则 $r < 0$; 而 $r = 0$ 表示无线性相关, 一般地, $|r| > 0.95$ 存在显著性相关; $|r| (0.8 \text{ 高度相关}; 0.5 \leq |r| < 0.8 \text{ 中度相关}; 0.3 \leq |r| < 0.5 \text{ 低度相关}; |r| < 0.3 \text{ 关系极弱, 认为不相关})$ 。

相关系数的计算有三种: Pearson、Spearman 和 Kendall。

Pearson 相关系数: 对定距变量的数据进行计算。公式为:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{其中, } r \text{ 为相关系数}$$

数; \bar{x}, \bar{y} 分别是变量 x, y 的均值; x_i, y_i 分别是变量 x, y 的第 i 个观测值。只有正态分布的等间隔测度 (连续) 的变量才使用这种相关分析, 它也是 SPSS 软件系统默认的相关分析方法。

Spearman 和 Kendall 相关系数: 对分类变量的数据或变量值的分布明显非正态或分布不明时, 计算时先对离散数据进行排序或对连续变量的值求秩, 在计算其秩分数间的相关系数。

* [收稿日期] 2007-01-05

[作者简介] 孙逸敏(1978-), 女, 上海市人, 新疆兵团广播电视大学讲师。

2 偏相关分析。当两个变量同时与第三个变量相关时,将第三个变量的影响剔除,只分析另外两个变量之间相关程度的过程。如一个量表中,控制现任职务和现任技术职称两个变量的影响,估计政治面貌与完成科研项目变量之间的相关关系。

3 距离相关分析。是对观测量之间或变量之间相似或不相似程度的一种测量。它用于同一变量内部各个取值间,以考察其相互接近程度;也可用于变量间,以考察预测值对实际值的拟合优度。

通过人工来计算这些复杂的运算比较麻烦,但利用计算机软件来对调查问卷进行相关分析却轻而易举,我们可以使用目前比较常用的 SPSS 软件进行相关性分析。上面介绍的三类相关分析的功能集中于 SPSS 软件中“分析”菜单下的“相关分析”子菜单中,它一般包括以下三个过程。

●二变量(Bivariate)过程此过程用于进行两个(多个)变量间的参数(非参数)相关分析,如果是多个变量,则给出两两相关的分析结果。这是“相关分析”子菜单中最为常用的一个过程,我们对它的使用可能占到相关分析的 95% 以上。下面介绍的利用 SPSS 软件进行变量的相关性分析也是以该过程为主。

●偏相关(Partial)过程如果需要进行相关分析的两个变量其取值均受到其他变量的影响,就可以利用偏相关分析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,这种分析思想和协方差分析非常类似。

●距离(Distances)过程调用此过程可对同一变量内部各观测量之间的数值或各个不同变量间进行距离相关分析。该过程在实际应用中用的比较少。

下面就谈谈二变量过程在调查问卷中不同项目之间相关性程度研究的应用。

三、利用 SPSS 软件进行变量的相关性分析举例

1. 调查方法与工具

(1)被测试对象。对兵团电大开放教育(成人本专科)2002 年至 2004 年毕业的近一千名已毕业生的用人单位进行问卷调查,采用两次分层等距抽样的随机抽样方法,回收“人才培养模式改革和开放教育试点”毕业生追踪调查^④中的调查表(用人单位用)626 份。根据项目内部一致性的分析,各个维度的信度系数均大于 0.8 或接近 0.8,说明各个纬度具有较好的内部一致性,问卷具有良好的信度(问卷总的内部一致性信度系数:Alpha = .9360)。

(2)研究方法。采用调查问卷的方法对某些变

量之间密切关系进行分析。下面就以兵团广播电视大学“人才培养模式改革和开放教育试点”毕业生追踪调查中的调查表(用人单位用)进行问卷调查。调查表主要包括三个部分:毕业生基本情况表、质量评价表和综合评价表。在这里主要讨论毕业生不同背景情况的相关分析,所以只用毕业生基本情况表(如下所示)即可。

- | |
|--|
| a. 现任职务:(局级; 处级; 科级; 普通职员; 其他) |
| b. 技术职称:(高级; 中级; 初级; 其他) |
| c. 政治面貌:(共产党员; 共青团员; 群众; 各民主党派; 其他) |
| d. 工作中的主要角色:(主持或专家; 业务骨干; 一般成员) |
| e. 年终考核成绩:(优秀; 优良; 合格; 不合格) |
| f. 职务职称是否晋升:(是; 否) |
| g. 通过学习获得何种奖励:(省部级; 地市级; 县级; 本单位; 无) |
| h. 发表或出版的研究成果:(专著; 论文; 文艺作品; 译文; 文章报纸; 无) |
| i. 完成的技术革新或科研项目。(省部级; 地市级; 县级; 本单位; 无) |

——毕业生基本情况表——

通过毕业生基本情况表,我们可以进行毕业生不同背景情况的分析,如政治面貌与年终考核成绩的关系、技术职称与发表或出版研究成果的关系等。

(3)相关系数显著性差异的统计意义。由于抽样误差的存在,样本中两个变量间相关系数不为 0,不能说明总体中这两个变量间的相关系数不是 0,因此必须经过检验。检验的零假设是总体中两个变量间的相关系数为 0。SPSS 的相关分析过程给出了该假设成立的概率,公式如下:

$$t = \frac{\sqrt{n-2} \cdot r}{\sqrt{1-r^2}} \quad (r \text{ 是相关系数, } n \text{ 是样本观测量数, } n-2 \text{ 是自由度})^{(9)}$$

数, $n-2$ 是自由度)

当相关系数检验的 t 统计量的显著性概率 $p < 0.05$ 时,说明两个变量间相关性显著,通常在概率值上方用“*”表示;当 $p < 0.01$ 时,说明两个变量间相关性非常显著,通常在概率值上方用“**”表示;当 $p > 0.05$ 时,说明两个变量间没有显著的相关关系,只显示概率值。

2. 数据处理

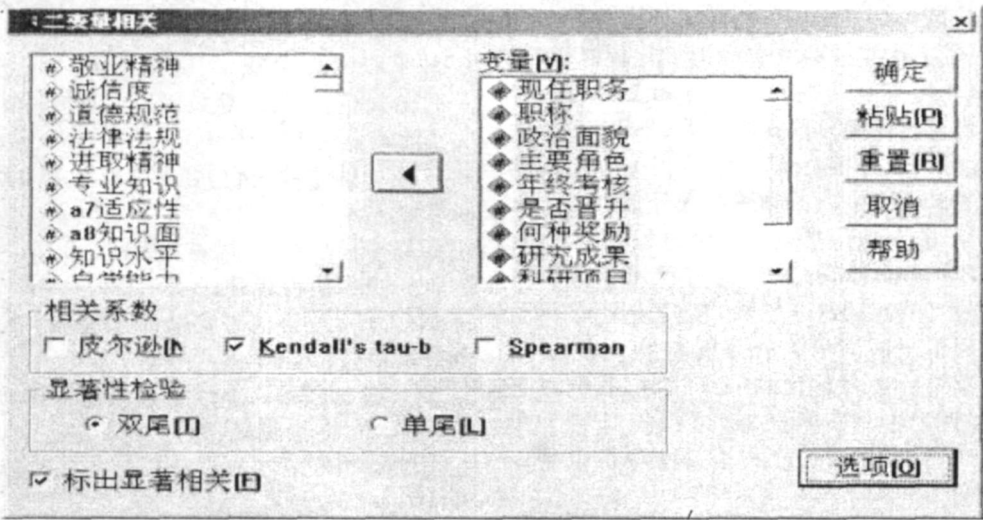
下面我们可以使用 SPSS11.0 对上述问卷进行毕业生不同背景情况的分析和探讨。

(1)对毕业生基本情况表进行变量分析,检验变

量间的关系是连续变量还是等级变量。根据表中的数据可以看出,每道题中的选项并非一个特定的数值,而是一个有序的变量,它的取值大小能够表示观测对象的某种顺序关系,也是基于“质”因素的变量,而不是典型的数据变量。而且,我们对各变量也没有进行正态分布检验,所以我们选择 Kendall's 等级相关的方法来进行变量间的相关分析。

图一:

(2)对各个变量进行相关性分析。在 SPSS 数据窗口中,点击“分析”菜单下的“相关分析”中的“二变量”,在弹出的对话框中将毕业生基本情况表中的九项添加到变量对话框中(如图一),勾上“Kendall's tau- b”相关系数、标出显著性相关两项,进行双尾显著性检验,点击“确定”按钮,系统会自动打开输出窗口。



在系统自动打开的输出窗口中,显示出相关分析表,如下图。

表一

		职务	职称	政治面貌	主要角色	考核	晋升	奖励	成果	项目
现任职务	Correlation Coefficient	1.000	.054	.313(**)	.281(**)	.158(**)	.072	.142(**)	.062	.024
	Sig. (2- tailed)	.	.135	.000	.000	.000	.060	.000	.091	.530
	N	626	621	624	624	619	613	618	616	613

** Correlation is significant at the 0.01 level (2- tailed). 相关系数的显著性概率水平为 0.

01

从上表(表一)中可以看出现任职务与政治面貌、主要角色、年度考核、何种奖励四项的显著性概率值 Sig< 0. 01, 并且有(**)的显著相关的标识,说明现任职务只与政治面貌、主要角色、年度考核、何种奖励有明显的相关性,与其余的职称、研究成果和科研项目没有关系。

表二

		职务	职称	政治面貌	主要角色	年度考核	晋升	奖励	成果	项目
现任职务	Correlation Coefficient	.158(**)	.169(**)	.159(**)	.373(**)	1.000	.214(**)	.258(**)	.145(**)	.129(**)
	Sig. (2- tailed)	.000	.000	.000	.000	.	.000	.000	.000	.001
	N	619	614	618	618	620	611	616	613	607

** Correlation is significant at the 0.01 level (2- tailed). 相关系数的显著性概率水平为 0.01

又如上表(表二)中可以看出, 年终考核与剩下八项都有(*)的标识, 说明它与其余的所有项目都有非常显著的相关性。

同理, 其余项目(如职称、政治面貌、主要角色、晋升、何种奖励、研究成果、科研项目)中的每一项都与所有项目之间都会产生一个对应的相关分析表, 这里就不一一列出了。

最后, 根据相关系数表得出结论。通过所有产生的相关分析表中的数据可以得出以下结论:

(1) 现任职务只与政治面貌、在工作中扮演的角色、年终考核、获得何种奖励显著相关, 与其余等项都不相关。

(2) 现任技术职称除了与现任职务不相关外, 与其余所有项目都显著相关。

(3) 政治面貌除了与研究成果、科研项目不相关外, 与其余所有项目都显著相关。

(4) 在工作中扮演的角色与其余所有项目都显著相关。

(5) 年终考核与其余所有项目都显著相关。

(6) 职务职称是否得到晋升除了与现任职务不相关外, 与其余所有项目都显著相关。

(7) 通过学习, 获得何种奖励与其余所有项目都显著相关。

(8) 发表或出版的研究成果与现任职务、政治面貌两个项目不相关, 与其余所有项目都显著相关。

(9) 完成的技术革新或科研项目与现任职务、政治面貌两个项目不相关, 与其余所有项目都显著相关。

3 结果和探讨

根据相关系数表得出的结论, 我们进一步对毕业生不同背景情况进行深入分析, 可以得出以下三个结论和改进措施。

(1) 调查的学生在发表研究成果和完成技术革

新或科研项目这两个方面与学生目前现任职务和政治面貌没有显著相关, 说明现任职务和政治面貌在科学研究和技术革新方面的激励机制中作用很小, 将来用人单位可以从学生的职称、在工作中的主要角色、年终考核、职务晋升、奖励等方面进行改革, 以此激励学生进行科学研究和技术革新。

(2) 调查的学生在现任技术职称和职务职称是否得到晋升的两个方面, 与现任职务没有相关性关系, 说明在职称的高低与职务的高低没有什么关系, 对学员的工作没有什么影响。

(3) 调查的学生在工作中的角色、年终考核成绩、获得何种奖励这三个方面与学生的所有项目都有很强的相关性, 说明这几项在学生的工作中非常重要, 用人单位需要从这三个方面加大激励措施, 增加学生在工作中的积极性, 提高工作效率, 以增加企业的经济效益。

四、结语

只有使用相关分析了解了事物之间的密切程度, 才能对事物之间的因果关系进行进一步的研究和探索, 我们可以根据相关分析了解事物之间的关系, 再通过回归分析探讨事物之间的因果。利用 SPSS 软件分析变量间的相关性, 使我们在科学研究中减少了复杂繁琐的计算, 方便了我们科学研究, 并帮助我们给事物之间的关系做出定性的分析。

注释:

¹ 杨晓明主编. SPSS 在教育统计中的应用. 北京: 高等教育出版社, 2004, 第 157 页。

^④ 刘香菊等. 兵团广播电视大学“人才培养模式改革和开放教育试点”毕业生追踪调查报告。

(四) 卢纹岱主编. SPSS for Windows 统计分析(第 2 版). 北京: 电子工业出版社, 2002, 第 208 页。

Using SPSS Software to Analyze the Correlation between Variables

Sun Yimin

(Radio and Television University of the XPCC, Urumqi 830001, Xinjiang)

Abstract: This paper makes an explanation of using SPSS software to analyze the correlation between variables by describing the research significance, concept and classification of correlation analysis and by giving the examples of using SPSS software to analyze the correlation between variables. That can make us understand the close links between things and carry out the qualitative analysis of relations between things.

Key words: correlation analysis; SPSS software; bivariate process; correlation coefficient; significant difference