

斯皮尔曼等级相关系数

维基百科，自由的百科全书

在统计学中,以查尔斯·斯皮尔曼命名的**斯皮尔曼等级相关系数**,经常用希腊字母 ρ (rho) 或者 r_s 表示。它是衡量两个变量的依赖性的非参数指标。它利用单调方程评价两个统计变量的相关性。如果数据中没有重复值,并且当两个变量完全单调相关时,斯皮尔曼相关系数则为+1或-1。

目录

定义和计算

相关度量

解释

示例

显著性的确定

基于斯皮尔曼相关系数的一致性分析

参见

引文

外部链接

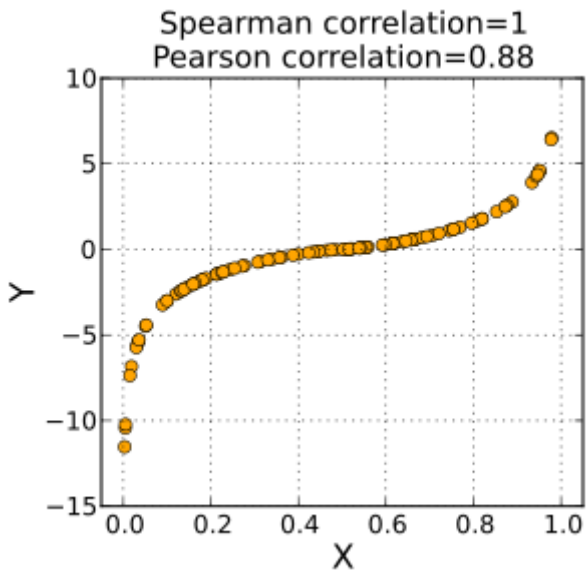
定义和计算

斯皮尔曼相关系数被定义成 等级变量之间的皮尔逊相关系数。^[1] 对于样本容量为 n 的样本, n 个 原始数据 $\mathbf{X}_i, \mathbf{Y}_i$ 被转换成等级数据 $\mathbf{x}_i, \mathbf{y}_i$, 相关系数 ρ 为

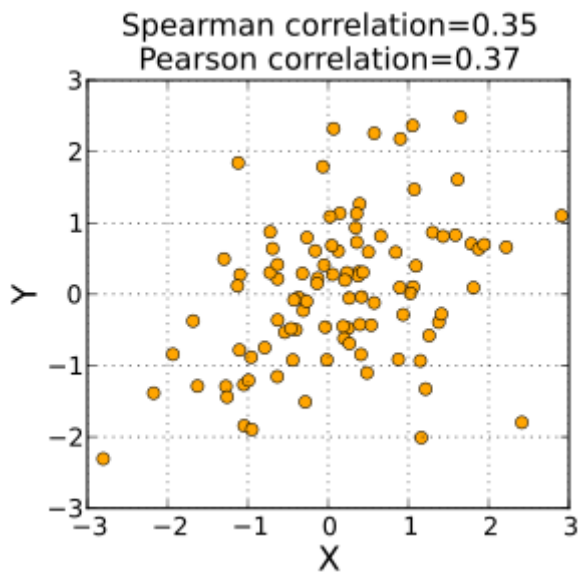
$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

原始数据依据其在总体数据中平均的降序位置，被分配了一个相应的等级。如下表所示：

变量 X_i	降序位置	等级 x_i
0.8	5	5
1.2	4	$\frac{4+3}{2} = 3.5$
1.2	3	$\frac{4+3}{2} = 3.5$
2.3	2	2
18	1	1



斯皮尔曼等级相关系数为1表明两个被比较的变量是相关的,即使它们之间的关系并非线性的。相较而言,它并未给出完整的皮尔逊相关系数。



当数据大致分布并没有明显的离群点，皮尔逊相关系数的值和斯皮尔曼相关系数的值是相似的。

实际应用中，变量间的连结是无关紧要的，于是可以通过简单的步骤计算 ρ 。^{[1][2]} 被观测的两个变量的等级的差值 $d_i = x_i - y_i$ ，则 ρ 为

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

相关度量

度量一对观测数据的统计依赖性还有其他的几种度量指标: 在相关性和依赖性中有谈及。其中最常用的是皮尔逊积矩相关系数。

斯皮尔曼相关也可称为 "级别相关";^[3] 也就是说，被观测数据的 "等级" 被替换成 "级别"。在连续的分布中，被观测数据的级别，通常总是小于等级的一半。然而，在这个案例中，级别和等级相关系数是一致的。更一般的，被观测数据的 "级别" 与估计的总体样本的比值小于给定的值，即被观测值的一半。也就是说，它是相应的等级系数的一种可能的解决方案。虽然不常用，"级别相关" 还是仍然有被使用。^[4]

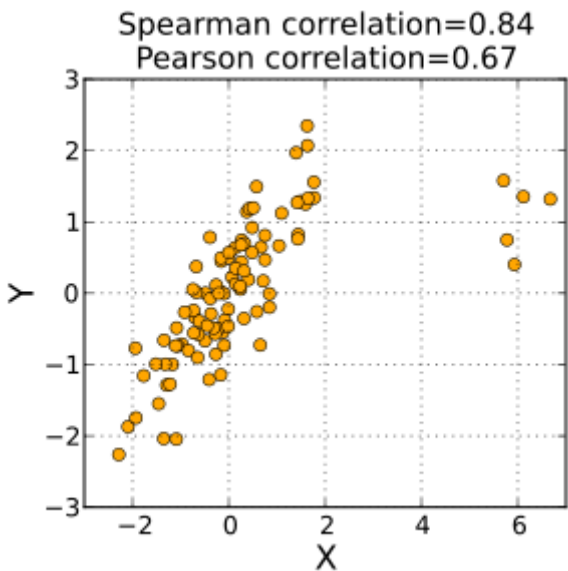
解释

斯皮尔曼相关系数表明 X (独立变量) 和 Y (依赖变量) 的相关方向。如果当 X 增加时， Y 趋向于增加，斯皮尔曼相关系数则为正。如果当 X 增加时， Y 趋向于减少，斯皮尔曼相关系数则为负。斯皮尔曼相关系数为零表明当 X 增加时 Y 没有任何趋向性。当 X 和 Y 越来越接近完全的单调相关时，斯皮尔曼相关系数会在绝对值上增加。当 X 和 Y 完全单调相关时，斯皮尔曼相关系数的绝对值为 1。完全的单调递增关系意味着任意两对数据 X_i, Y_i 和 X_j, Y_j ，有 $X_i - X_j$ 和 $Y_i - Y_j$ 总是同号。完全的单调递减关系意味着任意两对数据 X_i, Y_i 和 X_j, Y_j ，有 $X_i - X_j$ 和 $Y_i - Y_j$ 总是异号。

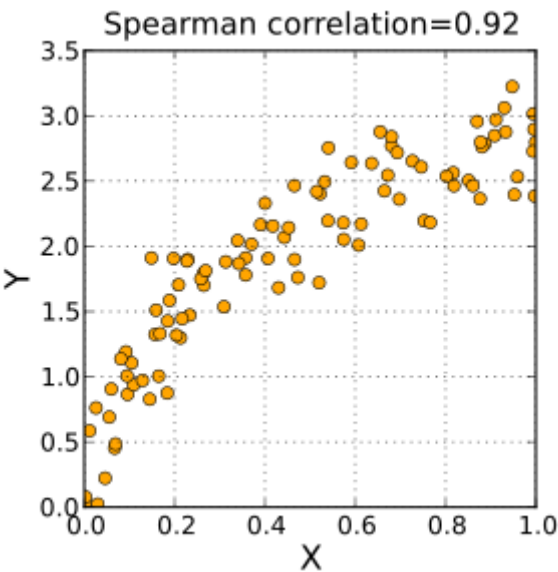
斯皮尔曼相关系数经常被称作 "非参数" 的。这里有两层含义。首先，当 X 和 Y 的关系是由任意 单调函数 描述的，则它们是完全皮尔逊相关的。与此相应的，皮尔逊相关系数只能给出由线性方程描述的 X 和 Y 的相关性。其次，斯皮尔曼不需要先验知识 (也就是说, 知道其参数) 便可以准确获取 X 和 Y 的采样概率分布。

示例

在此例中，我们要使用下表所给出的原始数据计算一个人的智商和其每周花在电视上的小时数的相关性。

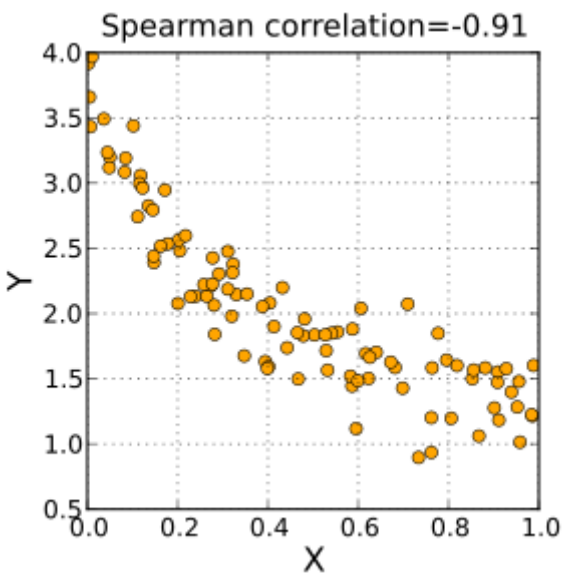


对样本中的显著离群点，斯皮尔曼相关系数比皮尔逊相关系数不敏感。



正的斯皮尔曼相关系数反应两个变量X 和 Y 单调递增的趋势。

智商, X_i	每周花在电视上的小时数 Y_i
106	7
86	0
100	27
101	50
99	28
103	29
97	20
113	12
112	6
110	17



首先，我们必须根据以下步骤计算出 d_i^2 ，如下表所示。

负的斯皮尔曼相关系数反应两个变量X 和 Y 单调递减的趋势。

- 排列第一列数据(X_i)。创建新列 x_i 并赋以等级值 1,2,3,... n 。
- 然后，排列第二列数据(Y_i)。创建第四列 y_i 并相似地赋以等级值1,2,3,... n 。
- 创建第五列 d_i 保存两个等级列的差值(x_i 和 y_i)。
- 创建最后一列 d_i^2 保存 d_i 的平方。

智商, X_i	每周花在电视上的小时数 Y_i	等级 x_i	等级 y_i	d_i	d_i^2
86	0	1	1	0	0
97	20	2	6	-4	16
99	28	3	8	-5	25
100	27	4	7	-3	9
101	50	5	10	-5	25
103	29	6	9	-3	9
106	7	7	3	4	16
110	17	8	5	3	9
112	6	9	2	7	49
113	12	10	4	6	36

根据 d_i^2 计算 $\sum d_i^2 = 194$ 。样本容量 n 为 10。将这些值带入方程

$$\rho = 1 - \frac{6 \times 194}{10(10^2 - 1)}$$

得 $\rho = -0.175757575...$

P-value = 0.6864058 (使用 t分布)

这个值很小表明上述两个变量的关系很小。原始数据不能用于此方程中，相应的，应使用皮尔逊相关系数计算等级。

显著性的确定

一种确定被观测数据的 ρ 值是否显著不为零(r 总是有 $-1 \leq r \leq 1$)的方法是计算它是否大于 r 的概率，作为原假设，并使用分层排列测试进行检验。这种方法的优点之处在于它考虑了样本中的数据个数和在使用样本计算等级相关系数的风险。

另外的一种方法是使用皮尔逊积矩中使用到的费雪变换。也就是， ρ 的置信区间和零检验可以通过费雪变换获得

$$F(r) = \frac{1}{2} \ln \frac{1+r}{1-r} = \operatorname{arctanh}(r).$$

如果 $F(r)$ 是 r 的Fisher变换，则

$$z = \sqrt{\frac{n-3}{1.06}} F(r)$$

是 r 的z-值，其中， r 在统计依赖($\rho = 0$).^{[5][6]}的零假设下 近似服从标准 正态分布。

显著性为

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

其在零假设下近似服从自由度为 $n-2$ 的t分布。^[7] A justification for this result relies on a permutation argument.^[8]

一般地，斯皮尔曼相关系数在有三个或更多条件的情况下是有用的。并且，它预测观测数据有一个特定的顺序。例如，在同一任务中，一系列的个体会被尝试多次，并预测在多次尝试过程中，性能会得到提升。在这种情况下，对条件间趋势的显著性检验由E. B. Page^[9] 发展了,并通常称为给定序列下的Page趋势测验。

基于斯皮尔曼相关系数的一致性分析

经典的一致性分析 是一种统计方法，它给两个标称变量赋给一个分数。通过这种方法，两个变量间的皮尔逊相关系数被最大化了。

有一种被称为级别相关分析的等价方法，它最大化了斯皮尔曼相关系数或肯德尔相关系数^[10]

参见

- Kendall tau rank correlation coefficient
- Rank correlation
- Chebyshev's sum inequality rearrangement inequality(These two articles may shed light on the mathematical properties of Spearman's ρ .)
- Pearson product-moment correlation coefficient, a similar correlation method that measures the "linear" relationships between the raw numbers rather than between their ranks.

引文

- Myers, Jerome L.; Well, Arnold D., Research Design and Statistical Analysis 2nd, Lawrence Erlbaum: 508, 2003, ISBN 0-8058-4037-0
- Maritz. J.S. (1981)*Distribution-Free Statistical Methods* Chapman & Hall. ISBN 0-412-15940-6 (page 217)
- Yule, G.U and Kendall, M.G. (1950), "An Introduction to the Theory of Statistics", 14th Edition (5th Impression 1968). Charles Griffin & Co. page 268
- Piantadosi, J.; Howlett, P; Boland, J. (2007) "Matching the grade correlation coefficient using a copula with maximum disorder", *Journal of Industrial and Management Optimization* 3 (2), 305–312
- Choi, S.C. (1977) Test of equality of dependent correlations. *Biometrika*, 64 (3), pp. 645–647
- Fieller, E.C.; Hartley, H.O.; Pearson, E.S. (1957)*Tests for rank correlation coefficients*. *J Biometrika* 44, pp. 470–481
- Press, Vetterling, Teukolsky, and Flannery (1992) *Numerical Recipes in C: The Art of Scientific Computing, 2nd Edition*, page 640

8. Kendall, M.G., Stuart, A. (1973)*The Advanced Theory of Statistics, Volume 2: Inference and Relationship* Griffin. ISBN 0-85264-215-6(Sections 31.19, 31.21)
 9. Page, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*. 1963,**58** (301): 216–230.[doi:10.2307/2282965](#)
 10. Kowalczyk, T; Pleszczyńska E. , Ruland F (eds.). *Grade Models and Methods for Data Analysis with Applications for the Analysis of Data Populations. Studies in Fuzziness and Soft Computing* vol. 151. Berlin Heidelberg New York: Springer Verlag. 2004. ISBN 978-3-540-21120-4
- G.W. Corder, D.I. Foreman, "Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach", Wiley (2009)
 - C. Spearman, "The proof and measurement of association between two things" *Amer. Psychol.*, 15 (1904) pp. 72–101
 - M.G. Kendall, "Rank correlation methods", Griffin (1962)
 - M. Hollander, D.A. Wolfe, "Nonparametric statistical methods", Wiley (1973)
 - J. C. Caruso, N. Clif, "Empirical Size, Coverage, and Power of Confidence Intervals for Spearman's Rho", *Ed. and Psy. Meas.*, 57 (1997) pp. 637–654

外部链接

- ["Understanding Correlation vs. Copulas in Excel"](#)by Eric Torkia, Technology Partnerz 2011
 - [Table of critical values of \$\rho\$ for significance with small samples](#)
 - [A calculator that shows the working out for Spearman's correlation](#)
 - [Spearman's rank online calculator](#)
 - [Chapter 3 part 1](#) shows the formula to be used when there are ties
 - [Spearman's rank correlation](#) Simple notes for students with an example of usage by biologists and a spreadsheet for [Microsoft Excel](#) for calculating it (a part of materials for a *Research Methods in Biology* course).
-

取自“<https://zh.wikipedia.org/w/index.php?title=斯皮尔曼等级相关系数&oldid=39538246>”

本页面最后修订于2016年3月28日 (星期一) 16:33。

本站的全部文字在[知识共享 署名-相同方式共享 3.0协议](#)之条款下提供，附加条款亦可能应用。（请参阅[使用条款](#)）
Wikipedia®和维基百科标志是[维基媒体基金会](#)的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是在美国佛罗里达州登记的501(c)(3)[免税](#)、非营利、慈善机构。