

斯皮尔曼相关系数的假设检验

分为小样本和大样本两种情况：
小样本情况，即 $n \leq 30$ 时，直接查临界值表即可。

1.小样本

*样本相关系数 r 必须大于等于表中的临界值，才能得出显著的结论。

<i>n</i>	单尾检验的显著水平			
	.05	.025	.01	.005
	双尾检验的显著水平			
	.10	.05	.02	.01
4	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.622	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467
35	0.283	0.335	0.394	0.433
40	0.264	0.313	0.368	0.405
45	0.248	0.294	0.347	0.382
50	0.235	0.279	0.329	0.363
60	0.214	0.255	0.300	0.331
70	0.190	0.235	0.278	0.307
80	0.185	0.220	0.260	0.287
90	0.174	0.207	0.245	0.271
100	0.165	0.197	0.233	0.257

Zar, J. H. (1972). Significance testing of the Spearman rank correlation coefficient. *Journal of the American Statistical Association*, 67, 578. Re- printed with permission from the *Journal of the American Statistical Association*. Copyright © 1972 by the American Statistical Association. All rights reserved.

2.大样本

大样本情况下, 统计量 $r_s \sqrt{n-1} \sim N(0, 1)$

$$H_0: r_s = 0, H_1: r_s \neq 0$$

我们计算检验值 $r_s \sqrt{n-1}$, 并求出对应的 p 值与 0.05 相比即可。

R =

1.0000	0.0301	-0.2430	-0.1990	0.0624	0.1099
0.0301	1.0000	0.1305	0.0898	0.0216	-0.0488
-0.2430	0.1305	1.0000	0.2626	0.0219	-0.0801
-0.1990	0.0898	0.2626	1.0000	-0.0910	-0.0029
0.0624	0.0216	0.0219	-0.0910	1.0000	-0.0399
0.1099	-0.0488	-0.0801	-0.0029	-0.0399	1.0000



1 x 6 double
591x6 double
Test of rho=0

检验值 $z^* = 0.0301 \sqrt{591-1} = 0.731126$

```
>> disp((1-normcdf(0.7311))*2)  
0.4647
```

p 值大于 0.05, 因此我们无法拒绝原假设。
(和 0 没有显著差异)

两个相关系数的比较

斯皮尔曼相关系数和皮尔逊相关系数选择:

1. 连续数据, 正态分布, 线性关系, 用 **pearson** 相关系数是最恰当, 当然用 **spearman** 相关系数也可以, 就是效率没有 **pearson** 相关系数高。
2. 上述任一条件不满足, 就用 **spearman** 相关系数, 不能用 **pearson** 相关系数。
3. 两个定序数据之间也用 **spearman** 相关系数, 不能用 **pearson** 相关系数。

定序数据是指仅仅反映观测对象等级、顺序关系的数据, 是由定序尺度计量形成的, 表现为类别, 可以进行排序, 属于品质数据。

例如: 优、良、差;

我们可以用 1 表示优差、2 表示良、3 表示优, 但请注意, 用 2 除以 1 得出的 2 并不代表任何含义。定序数据最重要的意义代表了一组数据中的某种逻辑顺序。

