ON THE CORRELATION BETWEEN RANKING AND NUMERIC VARIABLES

by

(nstance L. Wood and Foster B. Cady
University of Kentucky and Cornell University

BU-721-M* October 1980

Abstract

Considered is the dependence between a ranking variable and a continuous numeric variable for experiments where the experimental unit is a group of secondary units and the ranking is done within the experimental unit. If the continuous variable is assumed to be normal, a monotone function of the square of the Pearson product moment correlation is computed for each experimental unit and then averaged over all experimental units. This statistic is shown to have an F-distribution. However, if the continuous variable may not be assumed to be normal, Spearman's Rank Correlation is used. The randomization and large sample distributions for this latter case are given.

---

* In the Biometric Unit Mimeo Series, Cornell University, Ithaca, New York.

1. Introduction

Often the dependence between a ranking variable and a continuous numeric variable is of interest. Numerous counting and ranking procedures exist for testing independence based on a single bivariate sample of observations. Perhaps the best known is Spearman's rank order correlation (see Gibbons, 1976). Here we consider experiments where the experimental unit consists of a group of secondary units. Within each experimental unit, the secondary units are ranked according to some characteristic and also are measured with respect to a numeric variable. Of interest is the independence of the two characteristics.

Cunningham (1980) describes an experiment consisting of nine pens, each containing four chickens. Each pen is kept under observation and the chickens are assigned rank according to the observed "pecking order" within the pen, say R. Also recorded is the weight of each chicken, say Y. The hypothesis of interest is the independence of weight and pecking order.

In general, for b experimental units composed of n secondary units, we observed (R_{ij}, Y_{ij}), $i = 1, \cdots, b$, $j = 1, \cdots, n$ and we wish to test the independence of $\underset{\sim}{R}_i = (R_{i1}, \cdots, R_{in})'$ and $\underset{\sim}{Y}_i = (Y_{i1}, \cdots, Y_{in})'$. Our approach is essentially to form the square of an appropriate correlation coefficient and then to average these statistics over blocks. The choice of correlation coefficient is dependent upon the assumptions one might be willing to make about $\underset{\sim}{Y}_1, \cdots, \underset{\sim}{Y}_b$.

If $\underset{\sim}{Y}_i$ is assumed to be multivariate normal with mean $\mu \underset{\sim}{1}_{(n \times 1)}$ and covariance matrix $\sigma^2[(1+\rho)\underset{\sim}{I}_{n \times n} - \rho \underset{\sim}{J}_{n \times n}]$, for $\sigma^2 > 0$, $0 \leq \rho < 1$ and $\underset{\sim}{J}$ an $(n \times n)$ matrix of ones, then the original data (R_{ij}, Y_{ij}) are used to form the product moment correlation within each experimental unit. Due to the assumption of homogeneous errors within the experimental unit, we replace the sums of squares among the $\underset{\sim}{Y}_i$'s with a pooled unbiased estimator of σ^2. However, if the above normality assumption is untenable, $\underset{\sim}{Y}_i$ is replaced by the vector of within-experimental unit ranks for that characteristic, say $(S_{i1}, \cdots, S_{in})'$. This results in a Spearman's Rank Order Correlation being computed within each experimental unit. The former situation is considered in the next section, with the latter considered in Section 3.

This approach may be extended to experimental units which do not have constant size. The analysis which is presented in the next section is valid without modification. The randomization distributions in the third section may be computed as described, but are not given here.

2. Product Moment Correlation

Under the assumption of normality of the measured variables, we consider the test statistic

$$r_p^2 = \sum_{i=1}^{k} \left\{ \left[\sum_{j=1}^{n} Y_{ij}(R_{ij} - \bar{R}_{i.}) \right]^2 \Big/ s_p^2 \left[\sum_{i=1}^{n} (R_{ij} - \bar{R}_{i.})^2 \right] \right\} \tag{2.1}$$

where

$$s_p^2 = \left(\sum_{i=1}^{b} (n-1)s_i^2 - \left[\sum_{j=1}^{n} Y_{ij}(R_{ij} - \bar{R}_{i.}) \right]^2 \Big/ \sum_{i=1}^{n} (R_{ij} - \bar{R}_{i.})^2 \right) \Big/ \sum_{i=1}^{b} (n-2)$$

and

$$s_i^2 = (n-1)^{-1} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2 , \qquad i = 1, \cdots, b \quad .$$

Since

$$\bar{R}_{i.} = (n+1)/2 \quad \text{and} \quad \sum_{i=1}^{n} (R_{ij} - \bar{R}_{i.})^2 = n(n-1)(n+1)/12 \quad ,$$

the above statistic reduces to

$$r_p^2 = \frac{12}{bn(n+1)(n-1)} \sum_{i=1}^{b} \left[\sum_{j=1}^{n} Y_{ij}(R_{ij} - n+1/2) \right]^2 \Big/ s_p^2 \quad .$$

In a test of the hypothesis of independence between $\underset{\sim}{R}_i$ and $\underset{\sim}{Y}_i$, $i = 1, \cdots, b$, large values of r_p^2 will lead to rejection. However, we need to establish that the null hypothesis distribution of r_p^2 is given by $F_{b,b(n-2)}$, Snedecor's F with b and b(n-2) degrees of freedom.

Under H_0, $\underset{\sim}{Y}_1, \cdots, \underset{\sim}{Y}_b$ are independent of $\underset{\sim}{R}_1, \cdots, \underset{\sim}{R}_b$. Therefore, for any x, $-\infty < x < \infty$,

$$P_0\left[\sum_{j=1}^{n} Y_{ij}(R_{ij} - \bar{R}_{i.}) \Big/ \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})^2) \le x\right]$$

$$= E\left[P_0\left(\sum_{j=1}^{n} Y_{ij}(R_{ij} - \bar{R}_{i.}) \Big/ \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})^2 \le x \,|\, \underset{\sim}{R}_i\right)\right]$$

$$= E\left[P_0\left(\sum_{i=1}^{n} Y_{ij}\lambda_{ij} \le x \,|\, \underset{\sim}{R}_i\right)\right] \quad ,$$

where

$$\lambda_{ij} = (R_{ij} - \bar{R}_{i.}) \Big/ \left(\sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})^2\right)^{\frac{1}{2}} \quad ,$$

$$\sum_{j=1}^{n} \lambda_{ij} = 0 \qquad \text{and} \qquad \sum_{j=1}^{n} \lambda_{ij}^2 = 1 \quad .$$

Therefore

$$P_0\left[\sum_{j=1}^{n} Y_{ij}\lambda_{ij} \le x \,|\, \underset{\sim}{R}_i\right] = \Phi(x/\sigma) \quad ,$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Since the conditional distribution is functionally independent of $\underset{\sim}{R}_i$, under the null hypothesis

$$\sum_{j=1}^{n} Y_{ij}(R_{ij} - \bar{R}_{i.}) \Big/ \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})^2$$

is normal with mean zero and variance σ^2 .

Similarly, conditional on $\underset{\sim}{R}_i$

$$\sum_{j=1}^{n} (Y_{ij}(R_{ij} - \bar{R}_{i.}) \Big/ \sum_{j=1}^{n} (R_{ij} - \bar{R}_{i.})^2$$

is independent of s_p^2 . Since neither conditional distribution depends upon $\underset{\sim}{R}_i$, under H_0, the two statistics are independent.

Combining these results, we see that the numerator of r_p^2 is distributed as $\sigma^2\chi_b^2$ and independent of s_p^2 . This gives the desired result.

3. Rank Order Correlation

Next we consider an analogous distribution-free procedure where the Y_{ij} are replaced by their within-experimental unit ranks, S_{ij} . The test statistic then becomes

$$R_p^2 = \frac{144}{bn^2(n+1)^2(n-1)^2} \sum_{i=1}^{b} \Big[\sum_{j=1}^{n} S_{ij}(R_{ij} - \bar{R}_{ij}) \Big]^2 \quad .$$

Note that s_p^2 has been replaced by $n(n+1)(n-1)/12 = \Sigma_j(S_{ij} - \bar{S}_{i.})^2$. Again large values of R_p^2 lead to rejection of the hypothesis of independence.

Under the null hypothesis of independence of $\underset{\sim}{Y}_i$ and $\underset{\sim}{R}_i$, R_p^2 is distribution-free. In fact, since Y_{ij} are assumed to be equally correlated and in fact exchangeable, all possible rank configurations are equally likely. Through actual enumeration of all possible values of R_p^2 and the corresponding frequencies, the randomization distributions of R_p^2 for $n = 3, 4, 5$ have been tabulated. These are given in the Appendix.

Next we consider the large sample properties (in both b and n) of R_p^2 . By methods similar to that given in Section 2, it can be shown that under the hypothesis of independence, for $n \geq 4$,

$$bVar_0(R_p^2) = \left[\frac{2}{n(n-1)} + \frac{3(2n-3)}{n(n-1)(n-2)(n-3)}\right]$$

$$-6\left[\frac{144}{n^2(n-1)^2(n+1)^2(n-2)(n-3)}\right]\left[\sum_{j=1}^{n}(R_{1j}-\bar{R}_{1.})^4\right]$$

$$+\left[\frac{144}{n^2(n-1)^2(n+1)^2}\right]\left[\frac{n(n+1)}{(n-1)(n-2)(n-3)}\right]\left[\sum_{j=1}^{n}(R_{1j}-\bar{R}_{1.})^4\right]^2$$

$$= \frac{2}{n(n-1)} + o[n^{-1}(n-1)^{-1}] \quad .$$

Further, for fixed b as $n \to \infty$, from the well known fact that Spearman's Rank Correlation has an asymptotic normal distribution,

$$b\sqrt{n(n-1)}\ R_p^2 \text{ is asymptotically a } \chi_b^2 \text{ d.f.} \quad .$$

On the other hand, for fixed n as $b \to \infty$, from the independence of the experimental units, $\sqrt{b}[\sqrt{n(n-1)}\ R_p^2 - 1]/\sqrt{Var_0(R_p^2)}$ is asymptotically standard normal.

These two results give asymptotic approximations to the distribution of R_p^2; the choice of approximation depending upon the relative sizes of b and n . Comparisons of the exact and asymptotic critical points are given in Table 1.

4. Example

Cunningham (1980) gives the data on "pecking order" and weights of chickens caged four to a pen (see Table 2).

Table 1

Exact and Asymptotic Critical Points (p = .05)

b	2	3	4	5	6	7	8	9
				n = 3				
Exact	1.00 (.111)	1.00 (.037)	1.00 (.012)	.85 (.0453)	.875 (.018)	.786 (.045)	.813 (.020)	.75 (.0424)
χ_b^2	–	–	.969	.904	.857	.821	.791	.768
Normal (1)	–	.957	.883	.833	.796	.767	.744	.725
Normal (2)	.75	.704	.677	.658	.644	.634	.625	.618
				n = 4				
Exact	.82 (.049)	.6667 (.046)	.63 (.036)	.576 (.049)	.56 (.0451)	.543 (.044)	.525 (.048)	.511 (.050)
χ_b^2	.865	.752	.685	.639	.606	.580	.560	.543
Normal (1)	.766	.676	.624	.589	.563	.543	.526	.512
Normal (2)	.690	.625	.586	.559	.539	.524	.512	.502
				n = 3				
Exact	.625 (.042)	.527 (.049)	.4875 (.049)	.4580 (.049)				
χ_b^2	.6697	.582	.5305	.4950				
Normal (1)	.5914	.5239	.4837	.4562				
Normal (2)	.552	.497	.464	.441				

(1) Normal approximation using asymptotic mean and variance.

(2) Normal approximation using exact mean and variance.

Table 2

Relationship of Position in Pecking Order Within Pen

to Body Weight

Pecking Order /Pen	Weight (grams) 1	2	3	4	5	6	7	8	9
1	1880	1880	1300	1600	1380	1800	1000	1680	1800
2	1920	1520	1700	1830	1520	1780	1740	1460	1840
3	1600	1980	1500	1530	1520	1360	1520	1760	1720
4	1830	1680	1880	1820	1380	2000	2000	1500	1000
ΣY_{ij}	1808	1765	1595	1693	1450	1735	1565	1675	1590
$\Sigma(Y_{ij} - \bar{Y})^2$	61,766	126,700	188,300	73,476	19,600	217,100	541,100	69,100	471,600
$\Sigma Y_{ij}(R_{ij} - \bar{R})$	-235	-70	770	175	-140	90	1390	330	1260

Therefore,

$$\sum_{i=1}^{b}\left[Y_{ij}(R_{ij} - \bar{R})\right]^2 = 1,587,600 \quad ,$$

$$s_p^2 = [1,768,742 - (1,587,600)/5]/18 = 80,623.44$$

and from (2.1)

$$r_p^2 = (45)^{-1}(1,587,600/80,623.44) = 0.438 \quad ,$$

which is not significant when compared with the critical points of $F_{9,18}$.

References

1) Cunningham, D. L. (1980). Social aspects of feeding behavior in chickens. Unpublished manuscript, Dept. of Poultry Sci., Cornell Univ., Ithaca, NY.

2) Gibbons, J. D. (1976). Nonparametric Methods for Quantitative Analysis. Holt, Rinehart and Winston, New York.

APPENDIX: Randomization Distributions of R_p^2

Given below are the randomization distributions of R_p^2 for $n = 3, 4, 5$. In particular, $p = P[R_p^2 > x | n = k]$ is displayed. For some cases only the upper 25% of the distribution is given. Those distributions fitting the asymptotic approximations also are omitted (see Table 1).

$n = 3$

$b = 2$

x	p
0.2500	1.0000
0.6250	0.5556
1.0000	0.1111

$b = 3$

x	p
0.2500	1.0000
0.5000	0.7037
0.7500	0.2593
1.0000	0.0370

$b = 4$

x	p
0.2500	1.0000
0.4375	0.8025
0.6250	0.4074
0.8125	0.1111
1.0000	0.0123

$b = 5$

x	p
0.2500	1.0000
0.4000	0.8683
0.5500	0.5391
0.7000	0.2099
0.8500	0.0453
1.0000	0.0041

$b = 6$

x	p
0.2500	1.0000
0.3750	0.9122
0.5000	0.6488
0.6250	0.3196
0.7500	0.1001
0.8750	0.0178
1.0000	0.0014
1.3229	0.1250

$b = 7$

x	p
0.2500	1.0000
0.3571	0.9415
0.4643	0.7366
0.5714	0.4294
0.6786	0.1733
0.7857	0.0453
0.8929	0.0069
1.0000	0.0005

$b = 8$

x	p
0.2500	1.0000
0.3438	0.9610
0.4375	0.8049
0.5313	0.5318
0.6250	0.2587
0.7188	0.0879
0.8125	0.0197
0.9063	0.0026
1.0000	0.0002

$b = 9$

x	p
0.2500	1.0000
0.3333	0.9740
0.4167	0.8569
0.5000	0.6228
0.5833	0.3497
0.6667	0.1448
0.7500	0.0424
0.8333	0.0083
0.9167	0.0010
1.0000	0.0001

n = 4

b = 2

x	p
0.0	1.0000
0.0200	0.9931
0.0400	0.9653
0.0800	0.9375
0.1000	0.8819
0.1600	0.7708
0.1800	0.6597
0.2000	0.6458
0.2600	0.6181
0.3200	0.5625
0.3400	0.5208
0.3600	0.4375
0.4000	0.4306
0.5000	0.2639
0.5000	0.2222
0.5200	0.2083
0.5800	0.1806
0.6400	0.1250
0.6800	0.0625
0.8200	0.0486
1.0000	0.0069

b = 3

x	p
0.0	1.0000
0.0133	0.9994
0.0267	0.9959
0.0400	0.9890
0.0533	0.9844
0.0667	0.9774
0.0800	0.9497
0.1067	0.9219
0.1200	0.8941
0.1333	0.8368
0.1467	0.8299
0.1600	0.8229
0.1733	0.7859
0.1867	0.7720
0.2133	0.7442
0.2267	0.7390
0.2400	0.6904
0.2533	0.6678
0.2667	0.6644
0.2800	0.6227
0.2933	0.5394
0.3200	0.5324
0.3333	0.4491
0.3333	0.4387
0.3467	0.4369
0.3600	0.4091
0.3867	0.4016
0.4000	0.3461
0.4267	0.3183
0.4400	0.3027
0.4533	0.2436
0.4667	0.2350
0.4800	0.2280
0.5067	0.1655
0.5467	0.1516
0.5600	0.1256
0.5733	0.1047
0.6000	0.1030
0.6400	0.0613
0.6667	0.0457
0.6667	0.0353
0.6800	0.0336
0.7200	0.0301
0.7600	0.0231
0.7867	0.0075
0.8800	0.0058
1.0000	0.0006

b = 4

x	p	x	p
0.0	1.0000	0.3700	0.4104
0.0100	0.9999	0.3800	0.3564
0.0200	0.99[illegible]	0.3900	0.3448
0.0300	0.9984	0.4000	0.3356
0.0400	0.9969	0.4100	0.2939
0.0500	0.9953	0.4200	0.2869
0.0600	0.9907	0.4300	0.2603
0.0700	0.9814	0.4400	0.2522
0.0800	0.9753	0.4500	0.2511
0.0900	0.9706	0.4600	0.2163
0.1000	0.9519	0.4700	0.1886
0.1100	0.9322	0.4800	0.1862
0.1200	0.9299	0.4900	0.1810
0.1300	0.9160	0.5000	0.1428
0.1400	0.8890	0.5000	0.1368
0.1500	0.8798	0.5100	0.1365
0.1600	0.8705	0.5200	0.1284
0.1700	0.8576	0.5400	0.1062
0.1800	0.8449	0.5500	0.0900
0.1900	0.8191	0.5700	0.0854
0.2000	0.8133	0.5800	0.0749
0.2100	0.8052	0.5900	0.0599
0.2200	0.7651	0.6000	0.0576
0.2300	0.7350	0.6100	0.0564
0.2400	0.7304	0.6300	0.0356
0.2500	0.7026	0.6400	0.0333
0.2500	0.6453	0.6600	0.0294
0.2600	0.6451	0.6700	0.0224
0.2700	0.6324	0.6800	0.0190
0.2800	0.6229	0.7000	0.0187
0.2900	0.5840	0.7300	0.0117
0.3000	0.5678	0.7500	0.0065
0.3100	0.5307	0.7600	0.0046
0.3200	0.5207	0.7900	0.0042
0.3300	0.5181	0.8200	0.0034
0.3400	0.4706	0.8400	0.0008
0.3500	0.4394	0.9100	0.0006
0.3600	0.4336	1.0000	0.0000

b = 6

x	p
0.4320	0.2328
0.4400	0.2222
0.4480	0.2087
0.4560	0.1891
0.4640	0.1859
0.4720	0.1637
0.4800	0.1543
0.4880	0.1523
0.4960	0.1345
0.5040	0.1146
0.5120	0.1107
0.5200	0.1071
0.5280	0.0865
0.5360	0.0829
0.5440	0.0752
0.5520	0.0669
0.5600	0.0666
0.5680	0.0551
0.5760	0.0493
0.5840	0.0488
0.5920	0.0450
0.6000	0.0349
0.6000	0.0331
0.6080	0.0330
0.6160	0.0314
0.6320	0.0225
0.6400	0.0193
0.6560	0.0177
0.6640	0.0144
0.6720	0.0116
0.6800	0.0112
0.6880	0.0110
0.7040	0.0067
0.7120	0.0064
0.7280	0.0047
0.7360	0.0034
0.7440	0.0029
0.7600	0.0029
0.7840	0.0019
0.8000	0.0008
0.8000	0.0006
0.8080	0.0006
0.8320	0.0005
0.8560	0.0004
0.8720	0.0001
0.9280	0.0001
1.0000	0.0000

	b = 2				b = 3			
	x	p	x	p	x	p	x	p
n = 5	0.0	1.0000	0.3400	0.2942	0.3500	0.2472	0.5433	0.0427
	0.0050	0.9975	0.3600	0.2842	0.3533	0.2436	0.5467	0.0405
	0.0100	0.9808	0.3650	0.2706	0.3567	0.2371	0.5500	0.0394
	0.0200	0.9531	0.3700	0.2539	0.3600	0.2255	0.5533	0.0379
	0.0250	0.9431	0.4000	0.2339	0.3633	0.2234	0.5600	0.0319
	0.0400	0.9097	0.4050	0.2272	0.3667	0.2167	0.5667	0.0314
	0.0450	0.8997	0.4100	0.2203	0.3767	0.2029	0.5700	0.0294
	0.0500	0.8831	0.4250	0.1983	0.3800	0.1989	0.5733	0.0272
	0.0650	0.8275	0.4450	0.1617	0.3833	0.1904	0.5767	0.0265
	0.0800	0.7942	0.4500	0.1517	0.3867	0.1837	0.5800	0.0256
	0.0850	0.7875	0.4850	0.1294	0.3900	0.1810	0.5900	0.0246
	0.0900	0.7653	0.4900	0.1206	0.3933	0.1746	0.5933	0.0239
	0.1000	0.7375	0.5000	0.1106	0.4000	0.1654	0.5967	0.0230
	0.1250	0.7242	0.5050	0.0972	0.4033	0.1648	0.6000	0.0210
	0.1300	0.6919	0.5200	0.0917	0.4067	0.1560	0.6033	0.0207
	0.1450	0.6586	0.5300	0.0883	0.4100	0.1483	0.6067	0.0180
	0.1600	0.6386	0.5450	0.0750	0.4167	0.1453	0.6167	0.0169
	0.1700	0.6342	0.5650	0.0694	0.4200	0.1402	0.6233	0.0151
	0.1800	0.6008	0.5800	0.0594	0.4267	0.1308	0.6300	0.0137
	0.1850	0.5892	0.5850	0.0572	0.4300	0.1304	0.6333	0.0132
	0.2000	0.5503	0.6250	0.0417	0.4333	0.1262	0.6400	0.0121
	0.2050	0.5269	0.6400	0.0383	0.4367	0.1242	0.6467	0.0120
	0.2250	0.5136	0.6500	0.0358	0.4400	0.1155	0.6567	0.0100
	0.2450	0.4747	0.6800	0.0225	0.4433	0.1145	0.6600	0.0096
	0.2500	0.4647	0.7250	0.0186	0.4467	0.1114	0.6667	0.0075
	0.2600	0.4214	0.7450	0.0119	0.4533	0.1022	0.6700	0.0069
	0.2650	0.4058	0.8100	0.0086	0.4567	0.0996	0.6800	0.0067
	0.2900	0.3858	0.8200	0.0042	0.4600	0.0964	0.6867	0.0056
	0.3050	0.3525	0.9050	0.0025	0.4633	0.0934	0.6967	0.0060
	0.3200	0.3292	1.0000	0.0003	0.4667	0.0868	0.7033	0.0053
	0.3250	0.3242	0.4330	0.0670	0.4700	0.0856	0.7100	0.0040
					0.4730	0.0849	0.7200	0.0035
					0.4800	0.0803	0.7233	0.0035
					0.4833	0.0798	0.7500	0.0027
					0.4867	0.0768	0.7533	0.0026
					0.4900	0.0708	0.7600	0.0019
					0.4967	0.0698	0.7667	0.0018
					0.5000	0.0638	0.7867	0.0011
					0.5067	0.0616	0.8100	0.0010
					0.5100	0.0609	0.8167	0.0007
					0.5133	0.0564	0.8300	0.0004
					0.5167	0.0531	0.8733	0.0003
					0.5267	0.0491	0.8800	0.0001
					0.5367	0.0474	0.9367	0.0001
					0.5400	0.0449	1.0000	0.0000