A Measure of Concordance When There are Many Traits

Jai Won Choi, National Center for Health Statistics
Balgobin Nandram, Worcester Polytechnic Institute/National Center for Health Statistics
Jai Won Choi, Room 915, NCHS, 6525 Belcrest Road, Hyattsville, MD 20782

Key Words: Mixed Variable; Correlation; Nominal Value; Nonparametric; Prior Weight

Abstract: We consider two populations, each of which contains at least one object. On each object there are several traits measured. We obtain a simple nonparametric measure of concordance to assess the degree of positive relationship between the two populations, and in this measure the different traits can be weighted. We consider two cases, one in which each population has one object and the other case in which each population includes more than one object. Using the one to one corresponding common variables, our formula gives a similarity measure which ranges between 0 and 1 with increasing similarity from 0 to 1. We perform a simulation to assess the sensitivity of the measure to the changes in the weights. When the respective variables are widely spread out, the score obtained from the transformed values appears to be the better measurement.

1. Introduction

Statisticians routinely calculate the correlation between two groups when the measurements are similar, both continuous and discrete variables (e.g., weight and sex). Pearson (1896) presented mathematical formula of correlation estimation and many papers have been published since then on the same topic (Falk and Well 1997, Zhenng and Matis 1994, Rodgers and Nicewander 1988, Nelson 1998). When the variables are mixed and there is no common distribution, the calculation of correlation between two groups is not possible. Others use the ranks of the variables instead when the variables do not have common distribution (Steel and Torrie 1960, Goodman and Kruskal 1954 1963 1972, Kendall 1949b). But the ranking can be done for the same type of variables; therefore, the mixed variables can not be ranked.

No method is available to obtain relation between two groups with mixed variables. We present a simple method to calculate similarity between two groups when each group includes mixed variables and when one variable in that group matches to another variable in the second group. Our method can be used only when such one to one matching is possible for the variables in the groups and when the each variable can be properly quantified to a positive number.

This method is useful to many areas. For example, one may compare two persons when they have common mixed variables such as weight, height, age, sex, race, education, and income. Similarly we may compare two companies of different sizes and products, two countries of different population and culture, and two hurricanes of different forces and directions. It may also be applicable to form strata in sample surveys, combining similar subunits. When there are more traits, the measure of similarity is more reliable. The proper quantification of common variables is very important factor for the measurement.

Let $X_1, \ldots, X_n$ be the mixed variables of size n from one group. Two variables, X_i and X_j for $i \neq j$, in the same group are entirely different types. Some of the variables may be correlated as discussed in the conclusion. Similarly $Y_1, \ldots, Y_n$ are the mixed variables of size n in the other group. Again any two variables, Y_i and Y_j for $i \neq j$, are entirely different. Let X_i from the first group and Y_i from the second group represent the i^{th} common variables from the two groups. Our measure is nonparametric, and really does not require existence of any moments. However, if inference is required about the concordance measure, then one would require the mean and variance of the mixed variables in each group. Then, we assume that the two common variables, X_i and Y_i have common expected value and variance.

This concept is extended to two populations with more than one object. In this case we construct the possible pairs of objects between the two groups, and calculate the measure of concordance for each pair. The sum of all the concordance from these pairs is divided by the number of the pairs to measure the similarity between the two groups.

We discuss a simple method to calculate the concordance between two groups with mixed variables. In Section 2, we describe the measure of concordance when each group includes only one object. In the second part, we extend the comparison of two groups when the each group includes more than one object. Finally, we discuss how to transform common variables to obtain better measures when the common variables are widely spread out. In Section 3, we include some comments and possible extension.

2. Measure of Concordance

First, we construct a measure of concordance to compare two groups, each including one object (e.g., two persons). Second we construct a measure of concordance between two groups, each including more than one object. Last section discuss how to transform the variables to have better measures when these variables are wide spread. In each section, we include examples to illustrate the measure of concordance.

2.1 Comparison of Two Groups with One Object

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be the mixed variables of the object in group X and group Y, respectively in the population. The corresponding sample variables are $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ from each group. Here x_i and y_i are common variables for $i = 1, \ldots, n$. For instance, in comparing two persons, x_1 and y_1 are races for the two persons, x_2 and y_2 are education, x_3 and y_3 are sex, x_4 and y_4 are weights and so on. These variables could also be ranks, percents, counts, measurements, nominal variables, and may take any other types that are quantifiable with positive numbers.

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the n pairs of common variables from the two groups with common expectation and variance. We assign a weight to each pair. Define the weight W_i for $i = 1, \ldots, n$ with the constraint $\sum_{i=1}^{n} W_i = 1$. The individual weights could be different according to its importance among the n traits in the overall picture. For the equal weight, $W_i = 1/n$ for each term.

When proper weights are assigned to each term, we measure the concordance as the weighted average of ratios of common part of X_i and Y_i divided by the square root of X_i and Y_i. The common part of two variables may be considered as some form of "correlation" between them. When $x_i > 0$ and $y_i > 0$ for all i and one to one correspondence is possible for the common variables between two groups, we define the measure of concordance, denoted by CN(x,y),

$$CN(x, y) = \sum_{i=1}^{n} W_i \frac{min(x_i, y_i)}{\sqrt{(x_i y_i)}}. \qquad (1)$$

When all terms are considered similar in importance, we may assign the same weights (i.e.,$W_i = 1/n$) to all terms. When the minimum value is x_i (i.e., $min(x_i, y_i) = x_i$) for all i, the term under the summation sign in the equation (1) is simplified to $CN(x, y) = \sum_i W_i(\sqrt{x_i/y_i})$. Similarly, when the minium value is y_i (*i.e.*, $min(x_i, y_i) = y_i$) for all i, it reduces to $CN(x, y) = \sum_i W_i(\sqrt{y_i/x_i})$. Thus, CN(x,y) is a weighted average of terms in (0,1]. If the two groups are the same (i.e., $x_i = y_i$ for all i), $CN(x, y) = 1$ and $CN(x, y) \to 0$ when the difference between x_i and y_i becomes large for all i. We note that the CN measure is bounded by 0 and 1 (i.e.,$0 < CN(x, y) \leq 1$).

The CN measure (1) has other properties. It is invariant to scale, it is not invariant to location. $CN(ax, ay)$ =CN(x,y) for any scalar $a > 0$, but $CN(x + a, y + a) \neq CN(x, y)$ for any locator $a > 0$. Let $y = ax$ for a scalar a. Then $CN(x, y) = 1/\sqrt{a}$ for $a > 1$ and $CN(x, y) = \sqrt{a}$ for $0 < a < 1$.

Since there is only one observation from each object and two objects are different, the variance of NC(x,y) estimator is not available. There may be correlation between some of $x_1, \ldots, x_n$ and similarly for $y_1, \ldots, y_n$. If the group includes more than one object as in Section 2.2, we can obtain the variance as seen in Section 2.2.

In practice, the value x (or y) in denominator may be zero and in this case CN measure is undefined. To avoid such cases, the x (or y) can be rescaled to make it different from zero, adding a small number to each x and y. It may also happen to have negative values, and we may adjust them similarly adding a constant number to x and y.

Example 1

With the seven mixed variables (i.e., height, weight, age, sex, race, income, education), we can measure the concordance of John and Paul. The respective common variables for John and Paul are (6′, 5′), (120 lbs, 110 lbs), (25 yrs, 30 yrs), (*male*, male), (white, black), (50 k/yr, 40 k/yr) and (college graduate, high school graduate). Here the mean and variance of the mixed variables for each person is meaningless, and it is not sensible to calculate the Pearson's correlation coefficient. To quantify sex, 1 for male and 2 for female; similarly, 1 for white and 2 for black; 4 for college graduate and 2 for high school graduate. These indicators are quantified by (6, 5), (120, 110), (25, 30), (1, 1), (1, 2), (5, 4), (4, 2). When equal weights are assigned to these 7 variables (i.e., 1/7), CN measure of concordance between the John and Paul is

$$CN = \frac{1}{7}(\sqrt{5/6} + \sqrt{110/120} + ... + \sqrt{2/4}) = 0.87. \quad (2)$$

However, suppose that we want to emphasize the race component in their relationship, and one half of the weight is assigned to race and the remaining half is evenly divided to the six variables. Then, the CN measure of similarity is 0.80, which is reduced from 0.87 with the equal weight case.

In addition, our measure of concordance can be interpreted as the correlation between two sets of variables if they are expected to have a large positive correlation. When two groups are highly correlated with Pearson's correlation coefficient greater than 0.80 and when each group includes the measurements of same type, the CN measure is almost the same as Pearson's correlation coefficient and CN rank correlation is same as Spearman's rank correlation.

Example 2

We compare the CN measure of the data to two methods often used to measure correlation. Pearson's correlation and Spearman's rank correlation. We obtain the ranks for the numbers of eggs laid and similarly for the number of ovulated follicles, and paired them for 14 hans. Then we calculate the rank correlation by CN methods and Spearman's correlation.

In studying the use of ovulated follicles in determining eggs laid by ring-necked pheasant (Kabat et al., 1948), the eggs laid and ovulated follicles of 14 hans are: (Eggs, Follicles) = (39, 37), (29, 34), (49, 52), (28,26), (31, 32), (25, 25), (49, 55), (57, 65), (51, 44), (21, 25), (42, 45), (38, 26), (34, 29), (47, 30).

Table 1: Comparison of CN measures with other correlation coefficients

Methods	correlation
Pearson	0.92
CN	0.92
CN rank	0.87
Spearman Rank	0.87

Table 1 shows that Pearson's correlation coefficient and the CN measure are the same. We also present Spearman's rank correlation and the CN measure for the ranked data; these are the same. This shows that the CN measure is a sensible measure of the relationship between two commensurate variables which are expected to have a moderate to high correlation.

Example 3: A Small Scale Simulation

We use the data in Example 1 to study the sensitivity of CN(x,y) to changes in the weights W_i. We generate weights from the Dirichlet distribution with mean 1/7 for the seven cells and we allow τ to vary when $\tau/7$ are the parameters of Dirichlet distribution. Small τ reflects very heterogeneous weights and large τ very homogeneous weights. We allow τ to vary: $\tau =$ 5, 10, 25, 50, 75, 100, 200, 400. At each of these eight design points, we generate 100 sets of weights and computed CN(x,y).

In Table 2, we present a five number summary of CN(x,y) at each value of τ. As τ varies from 5 to 400, the average of CN(x,y) decreases smoothly from a range of (0.74, 0.98) to (0.86, 0.89). The values of CN(x,y) fluctuate about 0.87, the value corresponding to equal weight case. Thus, if the weights are varying widely, the variation in CN(x,y) is moderate, and if the weights are very similar (1/7), there is virtually no variation of CN(x,y) values.

Table 2: Five number summary of the CN measure by τ

τ	Min	Q_1	Q_2	Q_3	Max
5	.74	.84	.87	.90	.98
10	.78	.84	.87	.90	.94
25	.81	.86	.87	.88	.93
50	.83	.86	.87	.88	.91
75	.84	.86	.87	.88	.90
100	.84	.86	.87	.88	.90
200	.85	.87	.87	.88	.90
400	.86	.87	.87	.87	.89

2.2 Comparison of Two Groups with More Than One Object

When each group includes more than one object and each object is characterized with mixed variables as discussed, the work in Section 1 can be generalized to accommodate this complex case.

Suppose that group A includes objects $a_1, \ldots, a_m$ and group B includes objects $b_1, \ldots, b_m$. There would be m^2 pairs of (a_i, b_i) for $i = 1, \ldots, m^2$. The object a_i of the group A is characterized by a string of n mixed variables, $x_{i1}, \ldots, x_{in}$, and the object b_i of the group B is characterized by a string of $y_{i1}, \ldots, y_{in}$ mixed variables. The common variables,

x_{ij} and y_{ij}, for the j-th pair, $j = 1, \dots, n$, from the i-th pair, $i = 1, \dots, m^2$, are identified first. When proper weights are assigned to each pair for the i-th group, the CN measure of equation (1) can be extended to measure the similarity between the two groups A and B.

$$CN(A, B) = \frac{1}{m^2} \sum_{i=1,m^2} \sum_{j=1,n} W_j \frac{min(x_{ij}, y_{ij})}{\sqrt{(x_{ij} y_{ij})}} \quad (3)$$

The n weights are same for all i, then we can reduce the weights W_{ij} to W_j with constraint $\sum_{j=1,n} W_j = 1$ where we may use equal weight $W_j = 1/n$ for all j . The measurement CN(A,B) of similarity is also bounded by 0 and 1 as we have discussed in Section 2.1.

Example 4.

Two groups, A and B, of students are compared where group A has 3 students from the freshmen class and group B has 2 students from the senior class. We want to measure the similarity between these two groups according to the 6 variables. Group A includes three students identified with characteristics (19, M, W, 1*st*, 4.0, English), (18, F, W, 1*st*, 4.0, Sociology), and (19, M, W, 1*st*, 4.0, History) for their age, sex, race, years in college, grade point average, and major field. Similarly, group B includes two students with characteristics of (21, F, B, 4*th*, 2.0, math) and (22, M, B, 4*th*, 2.0, physics), respectively for their age, sex, race, years in college, grade point average, and major field.

Quantifying is a problem when the variable is nonnumeric. Sex may be indicated by 1 for male and 2 for female, and similarly race by 1 for white and 2 for black. It is not easy to quantify the major field, and there may be many ways we to do so. Liberal arts and sciences may be differentiated by two digit number, 10 for liberal arts and 20 for science, and within each field the major may be designated by one digit numbers 1 for English, 2 for Sociology, and 3 for History. According to this rule, we may quantify English major with 11, Sociology major with 12, and History major with 13. Similarly, we can use 21 for Mathematics and 22 for Physics. With these numeric indicators, group A has (19, 1, 1, 1, 4.0, 11), (18, 2, 1, 1, 4.0, 12), and (19, 1, 1, 1, 4.0, 13) for the 3 freshmen students, and Group B has (21, 2, 2, 4, 2.0, 21), and (22, 1, 2, 4, 2.0, 22) for the 2 senior students.

There are 6 possible pairs between these two groups. Giving equal weights to each pair, we obtain 0.74 for the CN measure of the two groups. Thus, there is at least moderate concordance between these two groups.

2.3 Variance of Common Variables

When the variables are all commensurate, it is possible to measure uncertainty about estimates of the CN measure; otherwise this is not sensible. For example, we can assume the existence of second moments.

Using Taylor's expansion, and assume mean $E(x_{ij}) = \mu_{xi}$, $E(y_{ij}) = \mu_{yi}$, variance $Var(x_{ij}) = \sigma_{xi}^2$, $Var(y_i) = \sigma_{yi}^2$ and covariance $Cov(x_{ij}, y_{ij}) = \sigma_{ij}$, the variance of $W_j\sqrt{x_{ij}/y_{ij}}$ of the jth term in the ith group (similarly for $W_j((\sqrt{y_{ij}/x_{ij}}))$, is given by

$$Var(W_j\sqrt{x_{ij}/y_{ij}}) = \frac{W_j^2}{4\mu_{yi}^2}[(\mu_{yi}/\mu_{xi})\sigma_{xi}^2$$

$$+(\mu_{xi}/\mu_{yi})\sigma_{yi}^2 + 2\sigma_{ij}] \ + \ O(n^{-1}).$$

The variances and covariance may be estimated by usual method of moment. When $E(x_{ij}) = E(y_{ij}) = \mu_i$, variance $Var(x_{ij}) = Var(y_{ij}) = \sigma_i^2$ and covariance $\sigma_{ij} = 0$, the variance is reduced to $\frac{W_j^2\sigma_i^2}{2\mu_i^2}$. When the n terms are independent and the minimum value is x_{ij} (i.e., y_{ij}) for all i, overall variance is the sum of these variances:

$$Var(CN(x,y)) = \Sigma_{ij}\frac{W_j^2}{2\mu_i^2}\sigma_i^2 + O(n^{-1}). \quad (4)$$

When Var(x_{ij}) and Var(y_{ij}) are different and $Cov(x_{ij}, y_{ij}) = \sigma_{ij} \neq 0$, above variance has to take these into account, and keep the separate terms of two variances and covariance. When the minimum value is not decided consistently in the n terms, we use $Var(\sqrt{x_{ij}/y_{ij}})$ when the $ijth$ term includes $min(x_{ij}, y_{ij}) = x_{ij}$ and the $Var(\sqrt{y_{ij}/x_{ij}})$ when the $ijth$ term includes $\min(x_{ij}, y_{ij}) = y_{ij}$.

2.4 Transformation of Common Variables

The quantity $\sqrt{x/y}$ may often overestimate or underestimate the real values when the ranges of all common values like x and y are spread widely or narrowly, while these specific x and y are within a narrow range or wider range. To correct this problem, first we find the minimum and maximum values of common variables. We subtract the common minimum from x and y variables and divide this resulting difference by the difference between common maximum and minimum values. Then, we apply the CN measure to these transformed variables.

Suppose that we know the maximum value B and minimum value A, and that 95 percent of all common objects are greater than this minimum value. We adjust all variables as $(y-A)/(B-A)$ and $(x-A)/(B-A)$, and a term of CN measure is $\sqrt{\frac{y-A}{x-A}}$ instead of $\sqrt{y/x}$ for $x > y$. The adjusted ratio is a more representative value for the relationship of x and y. We need to know the minimum value A for this adjustment.

Example 5.

In Example 1, we have the seven pairs of common variables: $(6,5)$ for height, $(120,110)$ for weight, $(25,30)$ for age, $(1,1)$ for sex, $(1,2)$ for race, $(5,4)$ for income, $(4,2)$ for education. The first pair is the heights of two persons. The major problem is the find the minimum height of people. For now suppose that the minimum height is 4 feet. Then the transformed heights are $[(5-4)/(6-4) = 1/2$ and $\sqrt{1/2} = 0.71$. Similarly setting the minimum weights of people to 90 lbs, the weight $(120,110)$ becomes $(110-90)/(120-90) = 2/3$ and $\sqrt{2/3} = 0.82$ instead of $\sqrt{110/120} = 0.96$. The minimum of age is set to 1, then adjusted term for ages $(30,25)$ is $\sqrt{25-1/30-1} = 0.91$. Male is 1 and female is 2 and $\sqrt{1-1/2-1} = 0$. But, when both sex are male or female, we define that $\sqrt{0/0} = 1$ for the same sex, and $\sqrt{0/1} = 0$ for different sex, and similarly for race. The minimum for income are set to 1 and the incomes $(4,5)$ provides a CN term of $\sqrt{4-1/5-1} = 0.87$. For the minimum of education is set to 1, the common education variables $(2,4)$ gives $\sqrt{2-1/4-1} = 0.58$.

Using these transformed variables, CN(John, Paul) is:

$CN(John, Paul) = \frac{1}{7}(0.71 + 0.82 + 0.91$
$+ 1 + 0 + 0.87 + 0.58) = 0.70$

Previous value for $CN(John, Paul)$ was:

$CN(John, Paul) = \frac{1}{7}(0.91 + 0.92 + 0.91$
$+ 1 + 0.71 + 0.89 + 0.71) = 0.87.$

The traits that incur most changes are 0.71 reduced to 0.00 for race, the 0.71 reduced to 0.58 for education, 0.91 reduced to 0.71 for height, 0.92 reduced to 0.82 for weight. These reduced values appear more reasonable than the original scores. Hence the new CN measure may be a better estimate.

Example 6.

In Example 4, the age, sex, race, years in college, grade point average, and major field are the characteristics of students. We may set the minimum and maximum of these values at $(1,90)$ for age, $(1,2)$ for sex, $(1,2)$ for race, $(1,4)$ for the years of education, $(1,4)$ for grade point average, and $(10,29)$ for major field.

Group A: (19, 1, 1, 1, 4.0, 11), (18, 2, 1, 1, 4.0, 12), and (19, 1, 1, 1, 4.0, 13) for three freshmen students, and Group B: (21, 2, 2, 4, 2.0, 21), and (22, 1, 2, 4, 2.0, 22) for two senior students. By transforming these values, we have Group A: (18, 0, 0, 0, 3, 1), (17, 1, 0, 0, 3, 2), and (18, 0, 0, 0, 3, 1) for the three students, and Group B: (20, 1, 1, 3, 1, 12), and (21, 0, 1, 3, 1, 12) for the two students.

The CN values of six pairs are: ($\sqrt{18/20}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{1/3}$, $\sqrt{1/11}$),
($\sqrt{17/20}$, $\sqrt{1/1}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{1/3}$, $\sqrt{2/11}$),
($\sqrt{18/20}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{1/3}$, $\sqrt{3/11}$),
($\sqrt{18/21}$, $\sqrt{0/0}$, $\sqrt{0/1}$, $\sqrt{0/3}$, $\sqrt{1/3}$), $\sqrt{1/12}$),
($\sqrt{17/21}$, $\sqrt{0/1}$, $\sqrt{0/1}$, $\sqrt{0/3}$, $\sqrt{1/3}$, $\sqrt{2/12}$),
($\sqrt{18/21}$, $\sqrt{0/0}$, $\sqrt{0/1}$, $\sqrt{0/3}$, $\sqrt{1/3}$, $\sqrt{3/12}$).

The quantity $\sqrt{0/0}$ is set to 1 to indicate that the two numbers indicate the same. The average sum of these is CN(A,B) $= (1/6)(1/7)[1.83 + 2.93 + 2.05 + 2.80 + 1.89 + 3.01] = 0.35$. The original score of 0.743 is reduced considerably. This happens mainly because the CN terms of opposite sex or race is zero.

3. Conclusion

We are not aware of any method to calculate the similarity between two objects with mixed variables. When there are no common variables (or distributions) as seen in the examples, we can not use the usual correlation methods. Those mixed variables may be nominal values, measurements, counts, and qualities that indicate an object's characteristics. When the variables for each object are mixed, we present a way to calculate the concordance between two objects. We extend this concept between two groups to include more than one object.

When the variables are similar in each object and Pearson's correlation $r > 0.8$, we found that the CN measure is approximately the same as the Pearson's correlation coefficient for continuous variables and that the CN rank measure is about same as the Spearman's rank correlation when the ranks are used for CN measure. The transformed variables may provide similar scores as those of Pearson's correlation coefficient even if the Pearson's correlation $r < 0.8$.

One problem is the correlated variables. The few variables of an object may be correlated. For example, the income and education or height and weight of a person might be correlated. If such correlation exist, an extra adjustment is needed.

An interesting problem is how to compare parametrically two populations with mixed multivariate characteristics. For example, a random sample of multivariate observations is obtained from one population, and an independent random sample from another population and these samples are to be used to compare the two populations. There are both continuous and discrete variables for each multivariate observation, and all the variables are jointly correlated. A simple example is when each individual in the sample has their sex and income measured. Then, one can take a Bernoulli random variable for sex and a lognormal distribution for income, and these two are correlated. We believe that this is a challenging problem.

References

Falk, R. and Well, A.D. (1997), "Many Faces of the Correlation Coefficient," J. of Statistics Education, v.5, N.3, 1-15.

Galton, F. (1888), "Co-relations and Their Measurement, Chiefly From Anthropometric Data," Proceedings of the Royal Society of London, 45, 135-145.

Goodman, L. A. and Kruskal, W. H. (1954), "Measures of Association for Cross-classifications," J. Amer. Statis. Assoc. 49, 732-764.

Goodman, L. A. and Kruskal, W. H. (1963), "Measures of Association for Cross-classifications,III" J. Amer. Statis. Assoc. 58, 310-364.

Goodman, L. A. and Kruskal, W. H. (1972), "Measures of Association for Cross-classifications,IV" J. Amer. Statis. Assoc. 67, 415-421.

Kabat, Buss, I. O. and Meyer R. K. (1948) "The use of ovulated follicles in determining eggs laid by ring-necked pheasant," J. Wildlife Management, 12: 399-416.

Kendall, M. G. (1949b), "Rank and Product moment correlation," Biometrika, 36, 177.

Nelson, R. B. (1998), "Correlation, Regression lines, and Moments of Inertia," The American Statistician, Vol. 52, 343-345.

Pearson, K. (1896), "Mathematical Contributions to the Theory of Evolution: III. Regression, Heredity, and Panmixia," Philosophical Transactions of the Royal Society of London, Ser. A, 187, 253-318.

Rodgers, J. L. and Nicewander, W. A. (1988), "Thirteen Ways to Look at the Correlation Coefficient," The American Statistician, Vol. 42, 59-66.

Steel, R. and Torrie, J. (1960), Principles and Procedures of Statistics, McGraw-Hill Book Company, Inc. New York, pp409.

Stigler, S. (1989), "Francis Galton's Account of the Invention of Correlation," Statistical Science, 4, 73-86.

Zheng, Q. and Matis, J. H. (1994), "Correlation Coefficient Revisited," The American Statistician", Vol. 48, No.3.