

A Comparison of Correlation Coefficients via a Three-Step Bootstrap Approach

Tahani A. Maturi (Corresponding author)

Department of Mathematical Sciences, University of Durham
Durham, DH1 3LE, England
E-mail: tahanistat@hotmail.com

Anga Elsayigh

Ain Shams University, Faculty of Commerce, Abbassia 11566, Cairo, Egypt
E-mail: angasoft@hotmail.com

Abstract

In this paper we compare ten correlation coefficients using a three-step bootstrap approach (TSB). A three-step bootstrap is applied to determine the optimal repetitions, B , to estimate the standard error of the statistic with certain degree of accuracy. The coefficients in question are Pearson product moment (r), Spearman's rho (ρ), Kendall's tau (τ), Spearman's Footrule (F_r), Symmetric Footrule (C), the Greatest deviation (R_g), the Top - Down (r_T), Weighted Kendall's tau (τ_w), Blest (ν), and Symmetric Blest's coefficient (ν^*). We consider a standard error criterion for our comparisons. However, since the rank correlation coefficients suffer from the tied problem that results from the bootstrap technique, we use existing modified formulae for some rank correlation coefficients, otherwise, the randomization tied-treatment is applied.

Keywords: A three-step bootstrap, Correlation coefficients, Pearson product moment, Spearman's rho, Kendall's tau, Spearman's Footrule, Symmetric Footrule, Greatest deviation, Top-Down, Weighted Kendall's tau, Blest, Symmetric Blest's coefficient

1. Introduction

One may be interested in the relationship between two factors or two variables and would wish to represent this relationship by a number or even using a statistical technique to make an inference. This number is called a correlation coefficient. The most common and well known correlation coefficient is the Pearson moment product coefficient. Some people may use this coefficient immediately ignoring the bivariate normality assumption of the data. Others may use a nonparametric rank correlation coefficient such as Spearman rho or Kendall tau for the same purpose. However the Pearson coefficient examines different aspects compared to Spearman and Kendall. Pearson coefficient considers the linearity of the relationship whereas Spearman and Kendall study the monotonicity of this relationship.

In some circumstances, we may have data with some outliers, in which case using the Greatest deviation correlation coefficient would be more suitable due to its robustness property against outliers. However, in other situations, assigning more emphasis on the top of the observations is required. For this purpose, several nonparametric coefficients are suggested, namely: the Top-Down, Weighted Kendall's tau, Blest, and Symmetric Blest's coefficient.

Usually the variance of the correlation coefficients is derived under the assumption that the null hypothesis that there is no correlation is true. Otherwise, it is intractable to calculate the variance without this assumption. Surely one would prefer using the correlation coefficient which produces small variation, and, consequently less standard error.

Borkowf (2000) presented a new nonparametric method for estimating the variance of Spearman's rho by calculating ρ from a two-way contingency table with categories defined by the bivariate ranks. His method is a computer method depending on the data at hand like the bootstrap and jackknife methods. There are claims that his method is more accurate than the bootstrap and jackknife methods. However, it is more complicated and the differences in accuracy are small in comparison with other methods.

In this paper, we will apply the bootstrap method. Yet, the common issue is how many replications, B , should run to get the most accurate results required. One such approach is introduced by Andrews and Buchinsky (2000; 2001; 2002) which is called a three-step approach. We will use this approach to determine the optimal number of replications for different degrees of accuracy. In addition we will use it for comparison, precisely, to estimate the standard error of the estimators, in our case the correlation coefficients, without imposing the null hypothesis. However, by using the bootstrap technique, the rank correlation coefficients suffer from tied problem. Therefore, we will use the existing adjusted formula for Spearman and Kendall coefficients (Hollander & Wolfe, 1999); otherwise, the randomization tied-treatment is applied (Gibbons, 1971).

Overviews of the correlation coefficients that are used in comparisons are given in Section 2. In Section 3, we will introduce some bootstrap notation and motivations. Then, in Section 4, we will summarize the three-step approach for estimating the standard error. The application of this approach on the correlation coefficients is introduced via the example in Section 5. Finally, conclusions are reported in the last section.

2. Overview of the correlation coefficients

Our data set consists of n observations, where (x_i, y_i) , $i = 1, 2, \dots, n$ is a random sample from a continuous bivariate population, and let p_i and q_i be the ranks of x_i and y_i , respectively. Here, we have ten correlation coefficients; Pearson moment product coefficient as a parametric correlation coefficient (Hollander & Wolfe, 1999; Gibbons, 1971), five nonparametric correlation coefficients: Spearman's rho, Kendall's tau (Hollander & Wolfe, 1999; Gibbons, 1971; Kendall, 1975), Spearman's Footrule or simply Footrule (Diaconis & Graham, 1977; Franklin, 1988; Salama & Quade, 2002; Ury & Kleinecke, 1979), Symmetric Footrule (Salama et al., 2001), the Greatest deviation (Gideon & Hollister, 1987), and four weighted nonparametric correlation coefficients: the Top-Down (Iman & Conover, 1987), Weighted Kendall's tau (Shieh, 1998), Blest (Blest, 2000), and Symmetric Blest's coefficient (Genest & Plante, 2003).

In Table 1 we summarize these coefficients which are involved in our comparisons. In the second and third columns their formulae are stated, yet the fourth column contains the adjusted formulae for Spearman and Kendall coefficients for when ties occur. For others coefficients where no adjusted formula is found in current literature, we will use the randomization tool (Gibbons, 1971) to deal with tied problem. The justification behind using the randomization method is that it behaves randomly as the bootstrap does. However, the most important property is that the randomization methods do not affect the null distribution of the rank correlation coefficient, so we do not need to adapt the null distribution for these coefficients.

3. Bootstrap Motivation and Notation

The bootstrap technique was first introduced by Efron in 1979 (Efron & Tibshirani, 1994). It is a computer intensive method of statistical analysis that uses resampling to calculate standard errors, confidence intervals and significance tests. There are various applications of the bootstrap techniques in life sciences such as medical, social, and business science. They are applicable as a parametric or semi-parametric or nonparametric technique. In this paper we will consider the nonparametric bootstrap since the correlation coefficients (except Pearson) in question are nonparametric rank correlation coefficients.

Actually, the most common problem in bootstrap literature is choosing the optimal number of repetitions, B . By choosing different, small values it can result in different answers. By choosing extremely large values it gives a more accurate result, however more costly. Andrews and Buchinsky (2000; 2001; 2002) introduced a three-step approach to determine the number of repetitions B with pre-fixed degrees of accuracy which is applied for many different bootstrap problems, such as estimating the standard error, confidence interval, and p-value for classical statistical techniques.

The aim is to achieve the desired level of accuracy by choosing B . Accuracy is measured by the percentage deviation of the bootstrap standard error estimate, confidence interval length, test's critical value, test's p-value, or bias-corrected estimate based on B bootstrap simulations from the corresponding ideal bootstrap quantities for which $B = \infty$. A bound of the relevant percentage deviation, pdB , is specified such as the actual percentage deviation is less than this bound with a specified probability, $(1 - \delta)$ tends to one. That is, for given (pdB, δ) , the optimal number of repetitions, B^* , satisfies

$$P^* \left(100 \frac{|\hat{\lambda}_{B^*} - \hat{\lambda}_{\infty}|}{\hat{\lambda}_{\infty}} \leq pdB \right) = 1 - \delta, \quad (1)$$

where λ is a quantity of interest (the standard error in our case), $\hat{\lambda}_{\infty}$ is an "ideal" bootstrap estimate, and $\hat{\lambda}_B$ is the bootstrap approximation of $\hat{\lambda}_{\infty}$ based on B bootstrap repetitions. Also, P^* here represent the probability with respect to the randomness in the bootstrap samples.

In the rest of this section we will present the notation which we will use in the bootstrap framework. Our observed data is a sample of size n : $\mathbf{X} = (X_1, \dots, X_n)'$, where $X_i = (x_i, y_i)$; $i = 1, 2, \dots, n$. Let $\mathbf{X}^* = (X_1^*, \dots, X_n^*)'$ be a bootstrap sample of size n based on the original sample \mathbf{X} . When the original sample \mathbf{X} is comprised of independent and identically distributed (*iid*) or independent and nonidentically distributed (*inid*) random variables, the bootstrap sample \mathbf{X}^* often is an *iid* sample of size n drawn from some distribution \hat{F} . In this paper, \hat{F} is the empirical distribution due to that we use the nonparametric bootstrap.

Let $\hat{\theta} = \hat{\theta}(\mathbf{X})$ be an estimator of a parameter θ_0 based on the sample \mathbf{X} . Let $\hat{\theta}^* = \hat{\theta}(\mathbf{X}^*)$ denote the bootstrap estimator. Let $\{\mathbf{X}_b^* : b = 1, \dots, B\}$ denote B *iid* bootstrap samples, each with the same distribution as \mathbf{X} . Let $\hat{\theta}_b^* = \hat{\theta}(\mathbf{X}_b^*)$, for $b = 1, \dots, B$, denote the corresponding B bootstrap estimators.

4. A three-step method for choosing the number of bootstrap repetitions B for standard error estimates

One of the application is to bootstrap standard error estimates for a scalar estimator $\hat{\theta}$, in our case $\hat{\theta}$ is a correlation coefficient. The quantities $(se, \hat{se}_{\infty}, \hat{se}_B)$ in this case are the standard error, the "ideal" bootstrap standard error estimator, and

the bootstrap standard error estimator based on B bootstrap repetitions, respectively, where E^* represent the expectation with respect to the randomness in the bootstrap samples. These quantities are given as

$$se = [E[\hat{\theta}(X) - E(\hat{\theta}(X))]^2]^{1/2}, \tag{2}$$

$$\hat{se}_\infty = [E^*[\hat{\theta}(X^*) - E^*(\hat{\theta}(X^*))]^2]^{1/2}, \tag{3}$$

$$\hat{se}_B = \left[\frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \frac{1}{B} \sum_{c=1}^B \hat{\theta}_c^* \right)^2 \right]^{1/2}. \tag{4}$$

The three-step method, introduced by Andrews and Buchinsky (2000), depends on the estimation of the coefficient of excess kurtosis, γ_2 , of the bootstrap distribution of the parameter estimator, where

$$\gamma_2 = \frac{E^*[\hat{\theta}(X^*) - E^*(\hat{\theta}(X^*))]^4}{\hat{se}_\infty^4} - 3.$$

By using the three-step method we aim to choose a value of B to achieve a desired accuracy of \hat{se}_B for estimating \hat{se}_∞ using pre-specified values of (pdb, δ) . The method involves the following steps:

Step(1): Suppose $\gamma_2 = 0$, and a preliminary value of B , denoted B_0 where

$$B_0 = \text{int} \left[\frac{5000 z_{1-\delta/2}^2}{(pdb)^2} \right]$$

and $\text{int}(a)$ be the smallest integer greater than or equal to a , z_α represents the α quantile of the standard normal distribution.

Step(2): Simulate B_0 bootstrap, and compute

$$\hat{\gamma}_{2B_0} = \frac{\frac{1}{B_0-1} \sum_{b=1}^{B_0} (\hat{\theta}_b^* - \hat{\mu}_{B_0})^4}{(\hat{se}_{B_0})^4} - 3, \quad \hat{\mu}_{B_0} = \frac{1}{B_0} \sum_{b=1}^{B_0} \hat{\theta}_b^*.$$

Simulate R , say $R = 500$, bootstrap samples from $(\hat{\theta}_1^*, \dots, \hat{\theta}_{B_0}^*)$, which can be written as $\Theta_r^{**} = (\theta_{1r}^{**}, \dots, \theta_{B_0r}^{**})$, $r = 1, \dots, R$. Then compute

$$\hat{\gamma}_{2B_0R} = 2\hat{\gamma}_{2B_0} - \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_2(\Theta_r^{**})$$

where

$$\hat{\gamma}_2(\Theta_r^{**}) = \frac{\frac{1}{B_0-1} \sum_{b=1}^{B_0} (\hat{\theta}_{rb}^{**} - \frac{1}{B_0} \sum_{b=1}^{B_0} \hat{\theta}_{rb}^{**})^4}{\left[\frac{1}{B_0-1} \sum_{b=1}^{B_0} (\hat{\theta}_{rb}^{**} - \frac{1}{B_0} \sum_{c=1}^{B_0} \hat{\theta}_{rc}^{**})^2 \right]^2} - 3.$$

Step(3): Let $B^* = \max(B_0, B_1)$, where

$$B_1 = \text{int} \left[\frac{2500 z_{1-\delta/2}^2 (2 + \hat{\gamma}_{2B_0R})}{(pdb)^2} \right]$$

such that, if $\hat{\gamma}_{2B_0R} \leq 0$ then the desired number of bootstrap repetitions is $B^* = B_0$, and if $\hat{\gamma}_{2B_0R} > 0$, compute $B_1 - B_0$ additional bootstrap estimate $\{\hat{\theta}_b^* : b = B_0 + 1, \dots, B_1\}$ before computing \hat{se}_{B^*} .

5. Example

Let us consider the data from a study designed to ascertain the relative importance of the various factors contributing to tuna quality and to find objective methods for determining quality parameters and consumer preference (Hollander & Wolfe, 1999). Table 2 gives values of the Hunter L measure of lightness (X), along with panel scores (Y) for nine lots of canned tuna (also see Figure 1).

The preliminary and the final values of the number of the replications for estimating the standard errors of our correlation coefficients are given in Tables 3 and 4. We will categorize our comparison in three blocks; a parametric correlation coefficient (Pearson), nonparametric rank correlation coefficients (Spearman, Kendall, Spearman's Footrule, Symmetric Footrule and the Greatest deviation coefficient). Finally, weighted nonparametric rank correlation coefficients (Top-Down, Weighted Kendall, Blest's coefficient and Symmetric Blest's coefficient).

For example, let us focus on the situation when $pdb = 5\%$ and $\delta = 0.01$, as in Table 5. Here the preliminary value of B_0 (for all correlation coefficients) is 1327 which is quite small as a starting point for any simulation, then the adjusted number of replications, B_1 , is given in the third column in Table 5, which varies from one correlation coefficient to another. For example, for the Pearson correlation coefficient, this number, B_1 is more than double the preliminary value (increased by 177%). However, for Footrule B_1 is just increased by 2.4% (from 1327 to 1527). For Kendall and Weighted Kendall, the number of replications, B_1 , increased almost the same (between 38.9% and 39.41%). Among the nonparametric correlation coefficients, Spearman has the highest increase in the number of replications, increased by 72.57%.

According to the differences between the observed (calculated value from the data) and the bootstrap simulation estimate, we can see that for Kendall correlation coefficient, the difference between the actual value of the statistic (observed value) and the bootstrap estimate is small. However, the largest difference between the observed and the estimated values is for Weighted Kendall since we choose $m = 5$ which means we ignore about a half of the data but this difference gap will reduce and is equal to that of Kendall at $m = n = 9$. Also, the differences are quite similar for Pearson, Spearman, Footrule, and Greatest, which are all between 0.0312 and 0.0520.

With respect to the standard error estimate, we found that the Pearson correlation coefficient has the smallest standard error, however, the Weighted Kendall has the largest value of standard error. In fact it is twice as much as the standard error of Pearson correlation coefficient. It is for the same reason that Weighted Kendall ignores some of the data (in our case 4 pairs of observations are ignored). In block 2, we found that the Greatest correlation coefficient has the smallest standard error which is expected since our data has some outliers (see Figure 1). However, Spearman and Footrule have large standard errors within this block.

In block 3, when our interest is focused on the initial (top) data, we noticed that the Symmetric Blest's coefficient has a smaller standard error than its asymmetric version and is also smaller than the other weighted correlation coefficients. However, as we mentioned above, Weighted Kendall has the largest standard error. Generally, the standard errors for correlation coefficients in block 3 are larger than those in block 1 and 2.

6. Conclusion

To conclude, one should use the Pearson correlation coefficient if the data meets the normality assumption, otherwise, the Greatest deviation performs well especially when the data has outliers. However, when we want emphasis on the initial (top) data, the Symmetric Blest's coefficient has lowest standard error amongst other weighted correlation coefficients.

References

- Andrews D. W. K. and Buchinsky M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, 68(1), 23-51.
- Andrews D. W. K. and Buchinsky M. (2001). Evaluation of a three-step method for choosing the number of bootstrap repetitions. *Journal of Econometrics*, 103, 345-386.
- Andrews D. W. K. and Buchinsky M. (2002). On the number of bootstrap repetitions for Bca confidence intervals. *Econometric Theory*, 18(4), 962-984.
- Blest D. C. (2000). Rank correlation - An alternative measure. *Australian & New Zealand Journal of Statistics*, 42(1), 101-111.
- Borkowf C. B. (2000). A new nonparametric method for variance estimation and confidence interval construction for Spearman's rank correlation. *Computational Statistics & Data Analysis*, 34, 219-241.
- Diaconis P. and Graham R. L. (1977). Spearman's Footrule as a measure of disarray. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39, 262-268.
- Efron B. and Tibshirani R. J. (1994). *An introduction to the bootstrap*. Chapman & Hall.
- Franklin L. A. (1988). Exact tables of Spearman's Footrule for $N=11(1)18$ with estimate of convergence and errors for the normal approximation. *Statistics & Probability Letters*, 6, 399-406.
- Genest C. and Plante J. (2003). On Blest's measure of rank correlation. *The Canadian Journal of Statistics*, 31(1), 1-18.
- Gibbons J. D. (1971). *Nonparametric statistical inference*. McGRAW-HILL KOGAKUHA, LTD.
- Gideon R. A. and Hollister R. A. (1987). A rank correlation coefficient resistant to outliers. *Journal of the American Statistical Association: Theory and Methods*, 82(398), 656-666.
- Hollander M. and Wolfe D. A. (1999). *Nonparametric statistical methods*, 2nd edition. John Wiley & Sons, INC.
- Iman R. L. and Conover W. J. (1987). A measure of Top - Down correlation. *Technometrics*, 29(3), 351-357.
- Kendall M. (1975). *Rank correlation methods*, 4th edition. Charles Griffin & Company LTD.

Salama I. A. and Quade D. (2002). Computing the distribution of Spearman's Footrule in $O(n^4)$ time. *Journal of Statistical Computation and Simulation*, 72(11), 895-898.

Salama I. A., Quade D. and Abdelfattah E. H. (2001). The symmetric Footrule. *Communications in Statistics - Theory and Methods*, 30(6), 1099-1109.

Shieh G. S. (1998). A weighted Kendall's tau statistic. *Statistics & Probability Letters*, 39, 17-24.

Ury H. K. and Kleinecke D. C. (1979). Tables of the distribution of Spearman's Footrule. *Applied Statistics*, 28(3), 271-275.

Table 1. Overview of some correlation coefficients

Corr.Coeff.	Formula	Definition	Modified formula for tied ranks
Pearson	$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$		
Spearman's rho	$\rho = 1 - \frac{6D_s^2}{n(n^2-1)}$	$D_s^2 = \sum_{i=1}^n (p_i - q_i)^2$	$\rho = \frac{n(n^2-1) - 6D_s^2 - \frac{1}{2} \left[\sum_{i=1}^g t_i(t_i^2-1) + \sum_{j=1}^h u_j(u_j^2-1) \right]}{\sqrt{[n(n^2-1) - \sum_{i=1}^g t_i(t_i^2-1)][n(n^2-1) - \sum_{j=1}^h u_j(u_j^2-1)]}}$ where g denotes the number of tied p groups, t_i is the size of tied p group i , h is the number of tied q groups, and u_j is the size of tied q group j (Hollander & Wolfe, 1999).
Kendall's Tau	$\tau = \frac{2K}{n(n-1)}$	$K = \sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)$ where $\text{sgn}(a) = 1$ if $a > 0$, $\text{sgn}(a) = 0$ if $a = 0$ and $\text{sgn}(a) = -1$ if $a < 0$.	$\tau = \frac{2K}{\sqrt{[n(n-1) - \sum_{i=1}^g t_i(t_i-1)][n(n-1) - \sum_{j=1}^h u_j(u_j-1)]}}$ where g denotes the number of tied X groups, t_i is the size of tied X group i , h is the number of tied Y groups, and u_j is the size of the tied Y group j (Hollander & Wolfe, 1999).
Spearman's Footrule	$F = \begin{cases} 1 - \frac{4D}{n^2} & \text{if } n \text{ is even} \\ 1 - \frac{4D}{n^2-1} & \text{if } n \text{ is odd} \end{cases}$	$D(p, q) = \sum_{i=1}^n q_i - p_i $	
Symmetric Footrule	$C = \begin{cases} \frac{4S}{n^2} & \text{if } n \text{ is even} \\ \frac{4S}{n^2-1} & \text{if } n \text{ is odd} \end{cases}$	$S(p, q) = \frac{D(p, q^*) - D(p, q)}{2}$ $D(p, q^*) = \sum_{i=1}^n p_i + q_i - (n+1) $	
Greatest deviation	$R_g = \frac{d(\epsilon o q) - d(q)}{\lceil \frac{n}{2} \rceil}$ where $\lceil \cdot \rceil$ is the greatest integer	$d_i(q) = \sum_{j=1}^i I(q_j > i)$ $d(q) = \max\{d_i(q)\}$ $d_i(\epsilon o q) = \sum_{j=1}^i I(q_j < n+1-i)$ $d(\epsilon o q) = \max\{d_i(\epsilon o q)\}$	
Top - Down	$r_T = \frac{\sum_{i=1}^n S p_i S q_i - n}{n - S_1}$	$S_i = \sum_{j=i}^n \frac{1}{j}$ (Savage scores)	
Weighted Kendalls	$\tau_w = \frac{2Q}{m(m-1)}$ where m is the number of interesting rankings	$Q = \sum_{1 \leq j < i \leq m} \text{sgn}(q_i - q_j)$ and $p = \frac{m}{n}, m = \lceil (n+1)p \rceil$, where $\lceil \cdot \rceil$ is the greatest integer	
Blests Correlation	$v = \frac{2w+1}{n-1} - \frac{12w}{n(n+1)^2(n-1)}$	$w = \sum_{i=1}^n (n-i+1)^2 q_i$	
Symmetric Blest	$v^* = \frac{2w+1}{n-1} - \frac{12(w_1+w_2)}{n(n+1)^2(n-1)}$	$w_1 = \frac{1}{2} \sum_{i=1}^n (n+1-i)^2 q_i$ $w_2 = \frac{1}{2} \sum_{i=1}^n (n+1-q_i)^2 i$	

Table 2. Tuna Lightness and Quality data set

Lot	Hunter L value (X)	Panel scores (Y)
1	44.4	2.6
2	45.9	3.1
3	41.9	2.5
4	53.3	5.0
5	44.7	3.6
6	44.1	4.0
7	50.7	5.2
8	45.2	2.8
9	60.1	3.8

Table 3. TSB results of estimating the standard errors of the correlation coefficients

Pearson	Observed (r) = 0.5712							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	\hat{r}	se_B	B_0	B_1	\hat{r}	se_B
20	49	23	0.6621	0.2157	83	65	0.6365	0.2257
15	86	68	0.6368	0.2238	148	977	0.6165	0.2185
10	193	1180	0.6190	0.2193	332	1294	0.6188	0.2174
5	769	1992	0.6160	0.2180	1327	3669	0.6232	0.2187
1	19208	39025	0.6203	0.2186	33175	67616	0.6201	0.2193
Spearman	Observed (ρ) = 0.6000							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	$\hat{\rho}$	se_B	B_0	B_1	$\hat{\rho}$	se_B
20	49	56	0.5907	0.2727	83	83	0.5537	0.2867
15	86	88	0.5398	0.3254	148	346	0.5285	0.3008
10	193	420	0.5371	0.2965	332	570	0.5421	0.2895
5	769	1207	0.5471	0.3026	1327	2290	0.5548	0.2938
1	19208	29889	0.5569	0.2951	33175	51594	0.5555	0.2965
Kendall	Observed (τ) = 0.4444							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	$\hat{\tau}$	se_B	B_0	B_1	$\hat{\tau}$	se_B
20	49	40	0.4659	0.2448	83	74	0.4473	0.2574
15	86	75	0.4474	0.2564	148	336	0.4198	0.2726
10	193	402	0.4235	0.2674	332	534	0.4310	0.2638
5	769	1029	0.4345	0.2708	1327	1839	0.4401	0.2673
1	19208	22832	0.4449	0.2667	33175	39455	0.4445	0.2661
Footrule	Observed (F_t) = 0.2000							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	\hat{F}_t	se_B	B_0	B_1	\hat{F}_t	se_B
20	49	43	0.1816	0.2970	83	65	0.1663	0.3209
15	86	63	0.1581	0.3047	148	157	0.1166	0.3196
10	193	190	0.1202	0.3265	332	340	0.1253	0.3109
5	769	777	0.1344	0.3109	1327	1359	0.1527	0.3071
1	19208	20843	0.1496	0.3115	33175	35778	0.1495	0.3131
Symmetric Footrule	Observed (C) = 0.4500							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	\hat{C}	se_B	B_0	B_1	\hat{C}	se_B
20	49	49	0.4643	0.2481	83	83	0.4578	0.2662
15	86	82	0.4570	0.2521	148	281	0.4461	0.2615
10	193	321	0.4327	0.2571	332	507	0.4250	0.2726
5	769	1083	0.4344	0.2611	1327	1772	0.4391	0.2577
1	19208	26162	0.4363	0.2642	33175	44740	0.4363	0.2651

Table 4. TSB results of estimating the standard errors of the correlation coefficients, contd.

Greatest	Observed (R_g) = 0.2000							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	\hat{R}_g	se_B	B_0	B_1	\hat{R}_g	se_B
20	49	51	0.2706	0.1952	83	104	0.2596	0.2444
15	86	108	0.2630	0.2440	148	204	0.2235	0.2404
10	193	259	0.2247	0.2434	332	390	0.2369	0.2345
5	769	790	0.2311	0.2348	1327	1444	0.2312	0.2351
1	19208	22685	0.2311	0.2297	33175	37942	0.2301	0.2296
Top-Down	Observed (r_T) = 0.7392							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	\hat{r}_T	se_B	B_0	B_1	\hat{r}_T	se_B
20	49	38	0.6375	0.3793	83	61	0.6290	0.3615
15	86	61	0.6247	0.3615	148	185	0.6169	0.3767
10	193	260	0.6020	0.3770	332	367	0.5986	0.3714
5	769	986	0.6281	0.3607	1327	1861	0.6329	0.3544
1	19208	25664	0.6341	0.3537	33175	44175	0.6333	0.3548
Weighted Kendall	Observed (τ_w) = 0.8000, $m = 5$							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	$\hat{\tau}_w$	se_B	B_0	B_1	$\hat{\tau}_w$	se_B
20	49	44	0.4859	0.3721	83	65	0.4762	0.3702
15	86	65	0.4720	0.3784	148	175	0.4155	0.4304
10	193	217	0.4132	0.4181	332	327	0.4160	0.4069
5	769	968	0.4355	0.4209	1327	1850	0.4329	0.4150
1	19208	26252	0.4394	0.4200	33175	45077	0.4402	0.4189
Blest	Observed (ν) = 0.6000							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	$\hat{\nu}$	se_B	B_0	B_1	$\hat{\nu}$	se_B
20	49	40	0.6012	0.2655	83	74	0.5808	0.2818
15	86	71	0.5825	0.2776	148	344	0.5722	0.3009
10	193	344	0.5492	0.3151	332	648	0.5503	0.3153
5	769	1247	0.5400	0.3171	1327	1994	0.5313	0.3169
1	19208	27959	0.5391	0.3154	33175	49867	0.5381	0.3150
Symmetric Blest	Observed (ν^*) = 0.6250							
	$\delta = .05$				$\delta = .01$			
<i>pdb</i> (%)	B_0	B_1	$\hat{\nu}^*$	se_B	B_0	B_1	$\hat{\nu}^*$	se_B
20	49	37	0.6159	0.2492	83	68	0.5968	0.2645
15	86	66	0.5981	0.2614	148	342	0.5855	0.2869
10	193	347	0.5685	0.2981	332	667	0.5667	0.2980
5	769	1246	0.5593	0.2987	1327	1975	0.5514	0.2973
1	19208	28237	0.5586	0.2977	33175	50212	0.5575	0.2976

Table 5. TSB results when *pdb*=5% and $\delta=0.01$

Corr.Coeff.	Observed (O)	B_1	$100 \frac{B_1 - B_0}{B_0}$	Estimate (E)	$ O - E $	se_B
Pearson	0.5712	3669	176.49	0.6232	0.0520	0.2187
Spearman	0.6000	2290	72.57	0.5548	0.0452	0.2938
Kendall	0.4444	1839	38.58	0.4401	0.0043	0.2673
Footrule	0.2000	1359	2.41	0.1527	0.0473	0.3071
Symmetric footrule	0.4500	1772	33.53	0.4391	0.0109	0.2577
Greatest	0.2000	1444	8.82	0.2312	0.0312	0.2351
Top-down	0.7392	1861	40.24	0.6329	0.1063	0.3544
Weighted Kendall	0.8000	1850	39.41	0.4329	0.3671	0.4150
Blest	0.6000	1994	50.26	0.5313	0.0687	0.3169
Symmetric Blest	0.6250	1975	48.83	0.5514	0.0736	0.2973

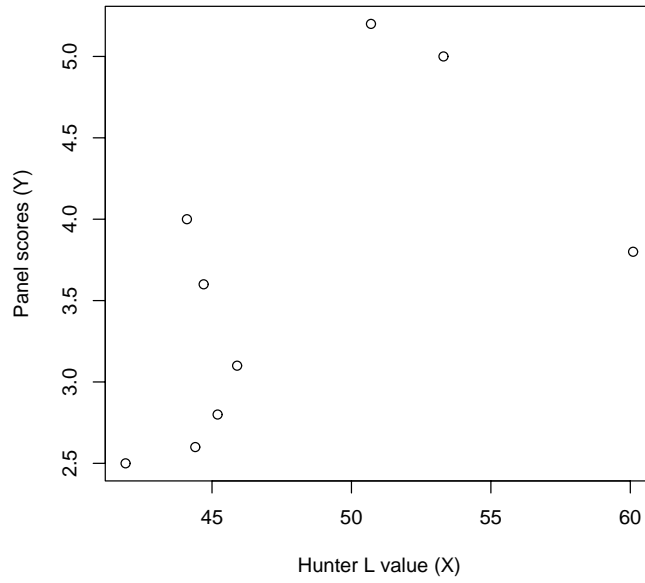


Figure 1. A simple scatter plot for Tuna Lightness and Quality data