

DEVELOPMENT OF SOME LOCALLY MOST POWERFUL RANK TESTS FOR CORRELATION

W.J. Conover

Texas Tech University
Lubbock, Texas 79409

ABSTRACT

The Hájek-Sidák (1967, p.71) theorem for locally most powerful rank tests (LMPRT) is extended in this paper to the bivariate case. This enables the locally most powerful rank test for correlation to be developed for continuous random variables under some fairly mild restrictions. Four examples are given to illustrate the ease and practicality of the procedure. The first two examples deal with the bivariate exponential models of Mardia (1970) and Gumbel (1970). The third example uses the bivariate normal distribution, and the fourth example derives the LMPRT for a general correlation model of Morgenstern (1956).

1. INTRODUCTION

There are two primary difficulties in developing tests with good power properties for testing the null hypothesis of independence between two variables X and Y , based on a bivariate random sample (X_i, Y_i) , $i = 1, 2, \dots, n$, against the alternative hypothesis of correlation. One difficulty is in finding a suitable model for the bivariate distribution, and the other is in developing a powerful test for correlation once the model is selected. Some of the results in this paper were previously published in a paper by Shiratata (1974) and by Conover (2001).

The bivariate normal distribution is a convenient model to use for many reasons. The parameter ρ is the linear correlation coefficient, so correlation is convenient to address in this model.

The most powerful test for correlation is well known, and the locally most powerful rank test (LMPRT) uses Fisher-Yates expected normal scores.

But some types of data do not fit the bivariate normal distribution very well. Therefore other classes of bivariate distributions have been developed in an attempt to find something other than the bivariate normal distribution to fit the data, while retaining some of the nice analytical properties found in the bivariate normal distribution. In this paper the general bivariate density function $h(x,y;\theta)$ is considered, with some fairly general restrictions.

One family of bivariate distributions was proposed by Morgenstern (1956). Let $F(x)$ and $G(y)$ be the marginal distribution functions. A bivariate distribution function with those marginals is given by

$$H(x,y;\theta) = F(x)G(y) [1 + \theta\{1 - F(x)\}\{1 - G(y)\}] \quad (1.1)$$

where θ is the parameter that governs the degree of dependence between the random variables. Farlie (1961) found Spearman's rho to be the optimal correlation coefficient for Morgenstern's model (1.1), and studied the efficiency of less than optimal coefficients. Our derivation of the same result is much simpler and with fewer restrictions on the model.

Morgenstern's model (1.1) was generalized by Farlie (1960) to

$$H(x,y;\theta) = F(x)G(y) [1 + \theta A(x)B(y)] \quad (1.2)$$

where $A(x)$ and $B(y)$ are bounded functions such that $A(\infty) = B(\infty) = 0$. The model (1.2), and thus (1.1) also, is a special case of the general model studied in this paper.

Konijn (1956) studied correlation tests for the hypothesis $\theta_2 = \theta_3 = 0$ in the model

$$X = \theta_1 W + \theta_2 Z \qquad Y = \theta_3 W + \theta_4 Z \quad (1.3)$$

where W and Z are independent random variables. Correlation tests for a similar class of alternatives

$$X = (1 - \theta)U + \theta Z \quad Y = (1 - \theta)V + \theta Z \quad (1.4)$$

where U , V and Z are independent random variables, and $\theta \rightarrow 0$, were investigated by Bhuchongkul (1964). Hájek and Sidák (1967, p.75) discuss the nearly identical model

$$X = U + \theta Z \quad Y = V + \theta Z \quad (1.5)$$

These models are more restrictive in their application than the more general model considered in this paper.

In this paper the general alternative distribution $h(x,y;\theta)$ is investigated. A theorem is presented that enables the locally most powerful rank test to be derived under some fairly general conditions. Four examples are given to illustrate the usefulness of this result. While the development stops short of finding the efficiencies of the obtained tests, in some cases the tests are well known, and their efficiencies have already been studied.

2. THE LOCALLY MOST POWERFUL RANK TEST FOR CORRELATION

Let (X,Y) have the joint density function $h(x,y;\theta)$ under H_a and the density $h(x,y;\theta_0) = f(x)g(y)$ under H_0 , the independence hypothesis, where $f(x)$ and $g(y)$ are the marginal density functions of X and Y respectively. In order to derive the locally most powerful rank test for H_0 against H_a , a bivariate rank version of the Neyman-Pearson lemma will be developed, followed by a useful theorem that will enable us to derive such a test.

Neyman-Pearson lemma for bivariate rank tests

Let X_1, \dots, X_n be i.i.d., with density $f(x)$ and ranks $R = (R_1, \dots, R_n)$. Similarly, let Y_1, \dots, Y_n be i.i.d. with density $g(y)$ and ranks $Q = (Q_1, \dots, Q_n)$. The most powerful size α rank test for

$$H_0: \text{the joint density of the } X\text{'s and } Y\text{'s is } \prod_{i=1}^n f(x_i)g(y_i) \quad (2.1)$$

against some simple alternative H_a is given by the critical region defined by the index function

$$\begin{aligned}\Phi(\mathbf{r}, \mathbf{q}) &= 1 \text{ if } P(\mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q} \mid H_a) > k \\ &= a \text{ if } P(\mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q} \mid H_a) = k \\ &= 0 \text{ if } P(\mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q} \mid H_a) < k\end{aligned}\quad (2.2)$$

where k and a are chosen so $E\{\Phi(\mathbf{R}, \mathbf{Q})\} = \alpha$ under H_0 .

Proof: The proof follows from the fact that $P(\mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q})$ is equal for all points (\mathbf{r}, \mathbf{q}) under H_0 , where \mathbf{r} and \mathbf{q} represent permutations of the ranks $1, \dots, n$. Then the critical region with the most power is the region that consists of those points with the greatest probability when H_a is true. Randomization with probability a for points with boundary probabilities k is used only to achieve a significance level exactly equal to α .

Now we are ready to develop a theorem for finding a locally most powerful rank correlation test. Consider independent copies (X_m, Y_m) , $m = 1, \dots, n$ of (X, Y) , with ranks R_m for X_m , and Q_m for Y_m as before. Let the density of (X, Y) be $h(x, y; \theta)$ under H_a , and $h(x, y; \theta_0) = f(x)g(y)$ under H_0 : $\theta = \theta_0$, where $f(x)$ and $g(y)$ are the marginal densities. Also define the scores

$$a(i, j; h) = E \left[\frac{\delta \{h(X_n^{(i)}, Y_n^{(j)}; \theta)\} / \delta \theta \mid_{\theta = \theta_0}}{h(X_n^{(i)}, Y_n^{(j)}; \theta_0)} \mid H_0 \right] \quad (2.3)$$

where $X_n^{(i)}$ and $Y_n^{(j)}$ are the i th and j th order statistics in a random sample of size n from $f(x)$ and $g(y)$ respectively.

Theorem for locally most powerful rank correlation tests

Let J be some open interval around θ_0 . If

1. $h(x, y; \theta) / h(x, y; \theta_0)$ exists for $\theta \in J$,
2. $\delta \{h(x, y; \theta)\} / \delta \theta \mid_{\theta = \theta_0}$ exists for $\theta \in J$,
3. $h(x, y; \theta_0) = \lim_{\theta \rightarrow \theta_0} h(x, y; \theta)$ exists for $\theta \in J$, and
4. $\lim_{\theta \rightarrow \theta_0} \iint \left| \delta \{h(x, y; \theta)\} / \delta \theta \mid_{\theta = \theta_0} \right| dx dy$
 $= \iint \left| \delta \{h(x, y; \theta)\} / \delta \theta \mid_{\theta = \theta_0} \right| dx dy < \infty,$

then the locally ($\delta \rightarrow 0$) most powerful rank test of $H_0: \Theta = \Theta_0$ against $H_a: \Theta = \Theta_0 + \delta$ is given by the test with the critical region

$$\sum_{m=1}^n a(R_m, Q_m; h) > k \quad (2.4)$$

where k is chosen so the test will have an appropriate size α .

Outline of the Proof: The proof resembles the proof on p. 71 of Hájek and Sidák (1967), except the integral is a $2n$ -fold integral over the region defined by both R and Q , and the Neyman-Pearson lemma for bivariate rank tests is invoked where appropriate.

Comment 1. This theorem and its preceding lemma are easily generalized to the p -variate setting, $p > 2$.

Comment 2. A slight variation of this theorem and proof allows us to consider the regression alternative, where the density of (X_m, Y_m) is $h(x, y; c_m \Theta)$, and results in the critical region

$$\sum_{m=1}^n c_m a(R_m, Q_m; h) > k \quad (2.5)$$

which is more in the spirit of the theorem on p.71 of Hájek and Sidák (1967).

Comment 3. This theorem is more general than the bivariate version given on p.75 of Hájek and Sidák (1967), which derived the LMPRT only for the model given by (1.5).

Comment 4. If X and Y are not continuous random variables, then the LMPRT is derived in the same manner described above, but with the joint density $h(x, y; \Theta)$ replaced by the bivariate (or multivariate) Radon-Nikodym derivative of $H(x, y; \Theta)$ with respect to $H(x, y; \Theta_0)$, in the same manner as in Section 6 of Conover (1973) in the univariate case.

Comment 5. The statistic defined by (2.4) is asymptotically normal under some general conditions on the scores. Asymptotic results are discussed in Section 4.

Implementation of the previous theorem to find the scores $a(i, j; h)$ associated with the locally most powerful rank test for

correlation involves the following steps.

1. Find the partial derivative of $h(x,y;\theta)$ with respect to θ and set θ equal to θ_0 .
2. Divide the result in Step 1 by $h(x,y;\theta_0) = f(x)g(y)$.
3. Substitute $X_n^{(i)}$ for x and $Y_n^{(j)}$ for y in the quotient in Step 2, where $X_n^{(i)}$ and $Y_n^{(j)}$ are the i th and j th order statistics in random samples of size n from $f(x)$ and $g(y)$ respectively.
4. Find the expected value of the random variable in Step 3 under H_0 . That is, integrate the product of
 - (a) the result of Step 2,
 - (b) the density function of the i th order statistic from $f(x)$,
 - (c) and the density function of the j th order statistic from $g(y)$,

over the entire range of values of X and Y .

3. FOUR EXAMPLES OF LOCALLY MOST POWERFUL RANK TESTS

These four examples show the ease with which the theorem of Section 2 can be applied to obtain locally most powerful rank tests for correlation. In all four examples the resulting test statistic is known, and the literature citations can be consulted to find tables for small sample sizes, and asymptotic approximations for large sample sizes.

The first two examples involve bivariate distributions, where both marginal distributions are exponential. The model in the first example allows only nonnegative correlations, and may be used when the alternative hypothesis is one of positive correlation. The model in the second example allows only nonpositive correlations, and may be used when the alternative hypothesis is one of negative correlation. In both examples, the locally most powerful rank test uses the top-down correlation coefficient of Iman and Conover (1987). The third example involves the bivariate normal distribution, and the fourth example looks at a very general bivariate distribution.

Example 1. Mardia (1970) presents a bivariate exponential distribution

$$h_1(x, y; \theta) = \frac{1}{1-\theta} e^{-\frac{x+y}{1-\theta}} \sum_{r=0}^{\infty} \left[\frac{\theta xy}{(1-\theta)^2} \right]^r / r! r! ; x > 0, y > 0, 0 < \theta < 1 \quad (3.1)$$

which has exponential marginal densities $\exp(-x)$ and $\exp(-y)$, and which degenerates to the product of those marginal densities $\exp\{-x-y\}$ for $\theta = 0$, representing the case of independence. The correlation coefficient between X and Y is θ .

This model first appeared in Mardia (1962) as a special case of a bivariate gamma distribution that appeared in Kibble (1941). It has been attributed to various authors, such as to Downton (1970) by Hawkes (1972) and others, and to Nagao and Kadota (1971) by Cordova and Rodriguez-Iturbe (1985), Johnson and Kotz (1972), and others. It is widely used as a model for the bivariate exponential distribution. A parametric test of the null hypothesis of independence is apparently unknown. The locally most powerful rank test is derived in the following.

It is easy to show that

$$a(i, j; h_1) = E\{(X_n^{(i)} - 1)(Y_n^{(j)} - 1) \mid H_0\} = (s_n(i) - 1)(s_n(j) - 1) \quad (3.2)$$

where $s_n(i)$ and $s_n(j)$ are the expected values of order statistics from the exponential distribution. That is, step 1 in the previous section involves finding the derivative of $h_1(x, y; \theta)$ with respect to θ , and setting $\theta = 0$. This gives

$$\frac{\delta}{\delta \theta} h_1(x, y; \theta) \Big|_{\theta=0} = e^{-x-y} (x - 1)(y - 1) \quad (3.3)$$

The second step is to divide by $f(x)g(y) = e^{-x-y}$, which gives

$$(x - 1)(y - 1)$$

In the third step the i th and j th order statistics from the exponential distributions $f(x)$ and $g(y)$ replace x and y respectively. Thus the expected values, in step 4, give the LMPRT scores in (3.2).

The scores in (3.2) are given by the formula

$$s_n(i) = \sum_{j=0}^{i-1} \frac{1}{n-j} \quad (3.4)$$

and are sometimes called Savage scores because they were introduced by Savage (1956). Their use in a rank correlation coefficient

$$r_T = \frac{\sum s_n(R_m) s_n(Q_m) - (\sum s_n(i))^2/n}{\sum [s_n(i)]^2 - (\sum s_n(i))^2/n} = \frac{\sum s_n(R_m) s_n(Q_m) - n}{n - s_n(n)} \quad (3.5)$$

was studied by Iman and Conover (1987), and called the top-down correlation coefficient r_T because of its tendency to emphasize the tail values. Exact tables for r_T for $n \leq 14$ are given by Iman (1987). Therefore the locally most powerful rank test of $H_0: \Theta = 0$ against $H_a: \Theta > 0$ in the bivariate exponential distribution given by (3.1) rejects H_0 if and only if $r_T > k$ for a suitably chosen value of k .

Example 2. Gumbel (1960) introduced another bivariate exponential distribution

$$h_2(x, y; \Theta) = \{(1+\Theta x)(1+\Theta y) - \Theta\} e^{-x-y-\Theta xy} \quad x>0, y>0, 0 \leq \Theta \leq 1 \quad (3.6)$$

with non-positive correlation coefficient

$$-1 + \int_0^{\infty} \frac{e^{-y}}{1 + \Theta y} dy \quad (3.7)$$

Note that the correlation coefficient is zero when $\Theta = 0$, and it decreases monotonically as Θ increases. Therefore the LMPRT for correlation is also the LMPRT for Θ . This distribution degenerates to $\exp\{-x-y\}$ under $H_0: \Theta = 0$. This widely known model was studied further by Gumbel (1961) and has been used more recently by Wei (1981) and Barnett (1983). As with the previous model, a parametric test of the null hypothesis of independence is apparently unknown. The locally most powerful rank test is derived in the following.

The optimal scores are again found to be functions of the

Savage scores. Specifically the scores are

$$a(i,j;h_2) = E\{-(X_n^{(i)}-1)(Y_n^{(j)}-1) \mid H_0\} = -(s_n(i)-1)(s_n(j)-1) \quad (3.8)$$

which leads to the locally most powerful rank test that rejects H_0 when $r_T < k$ for some suitably chosen negative number k . Note that the negative value for k is due to the model, which allows only negative correlation in the restricted parameter range for θ .

Example 3. The all-important bivariate normal distribution has density

$$h_3(x,y;\theta) = (2\pi(1-\theta^2))^{-1/2} \exp\{-(x^2+y^2-2\theta xy)/2\} \quad (3.9)$$

and correlation coefficient θ . The scores for the locally most powerful rank test are given by

$$a(i,j;h_3) = E(Z_n^{(i)})E(Z_n^{(j)}) \quad (3.10)$$

where $Z_n^{(i)}$ and $Z_n^{(j)}$ are order statistics from the standard normal distribution. These scores are used in the well-known normal scores statistic first given by Fisher and Yates (1957). This derivation of the locally most powerful rank test for the bivariate normal distribution is much simpler than the previous ones, and uses a more general model than the rather restrictive models (1.3), (1.4) and (1.5).

Example 4. The class of bivariate distributions introduced by Morgenstern (1956) has the bivariate distribution function

$$H(x,y;\theta) = F(x)G(y)[1 + \theta\{1 - F(x)\}\{1 - G(y)\}] \quad (3.11)$$

for any marginal distribution functions $F(x)$ and $G(y)$. This model has been extended by Plackett (1965) and often appears in discussions of bivariate distributions (see for example Mardia, 1970, or Johnson and Kotz, 1972). Due to the unspecified nature of $F(x)$ and $G(y)$ no parametric test is possible. However, rank tests

are possible. In fact the locally most powerful rank test is easily derived, as shown in the following.

When $H(x,y)$ is continuous then the density function is

$$h_4(x,y;\theta) = f(x)g(y)[1 + \theta\{1 - 2F(x)\}\{1 - 2G(y)\}] \quad (3.12)$$

which reduces to the independence case $f(x)g(y)$ when $\theta = 0$. This example shows the full power of the method introduced in this paper for finding the locally most powerful rank test for independence. The scores $a(i,j;h_4)$ in this case reduce to

$$\begin{aligned} a(i,j;h_4) &= E\{(2F(X_n^{(i)}) - 1)(2G(Y_n^{(j)}) - 1)\} \\ &= (2E\{U_n^{(i)}\} - 1)(2E\{U_n^{(j)}\} - 1) \end{aligned} \quad (3.13)$$

where $U_n^{(i)}$ and $U_n^{(j)}$ represent order statistics from the uniform distribution on $(0,1)$. These are the scores used in the Spearman rank correlation coefficient, so Spearman's rho is the locally most powerful rank test for correlation for the entire class of Morgenstern distributions, assuming only that the bivariate distributions are continuous. This result was first obtained by Farlie (1961), but this method of proof is much simpler.

Note that $h_4(x,y;\theta)$ is a density function with marginal densities $f(x)$ and $g(y)$ for all density functions f and g . In particular if f and g are exponential density functions, h_4 is another form of a bivariate exponential distribution. In this case the correlation coefficient is $\theta/4$ (Gumbel, 1960) and it varies only within the narrow domain $[-.25, .25]$. Since the correlation coefficient is a monotonic function of θ , the LMPRT for correlation in this bivariate exponential model uses Spearman's rho, instead of the top-down correlation coefficient of the previous two bivariate exponential models.

4. CONCLUDING REMARKS

Asymptotic normality for the special cases of the test statistic given in the previous section is already known. In general, asymptotic normality results from the following theorem.

Theorem showing asymptotic normality

1. Let $a(i, j; h) = E\{\phi(U_1, V_1) | R_1=i, Q_1=j\}$, where X_m and Y_m are independently distributed according to $F(x)$ and $G(y)$ respectively, $1 \leq m \leq n$, and where $U_m = F^{-1}(X_m)$ and $V_m = G^{-1}(Y_m)$.
2. Assume $0 < \iint [\phi(u, v) - \bar{\phi}]^2 du dv < \infty$, where $\bar{\phi} = \iint \phi(u, v) du dv$ and where integration is over the unit square.
3. Assume H_0 is true. Let $S = \sum a(R_i, Q_j; h)$, where R_i and Q_j are the ranks of X_i (hence U_i) and Y_j (hence V_j) respectively.

Then S is asymptotically (as n gets large) normal with mean

$$\mu = \sum_i \sum_j a(i, j; h) / n$$

and variance given by either

$$\sigma^2 = (n-1) \iint [\phi(u, v) - \bar{\phi}]^2 du dv$$

or

$$\sigma^2 = (n-1)^{-1} \sum_i \sum_j [a(i, j; h) - a(\cdot, j; h) - a(i, \cdot; h) + a(\cdot, \cdot; h)]^2$$

where the dot notation refers to averages over the missing arguments.

Proof. Introduce $T = \sum \phi(U_i, V_i)$ and let $U^{(i)}$ and $V^{(i)}$ be the vectors of order statistics for U and V respectively. Then

$$\begin{aligned} E\{(S - T)^2 | U^{(i)} = u^{(i)} \text{ and } V^{(i)} = v^{(i)}\} \\ = E\{\sum [a(R_i, Q_i; h) - \phi(u^{(R_i)}, v^{(Q_i)})]\}^2 \end{aligned}$$

Let $b(i, j) = a(i, j; h) - \phi(u^{(i)}, v^{(j)})$. Then by Theorem a on p.57 of Hájek and Sidák (1967) the above expression is equal to

$$\begin{aligned} (n-1)^{-1} \sum_i \sum_j [b(i, j) - b(\cdot, j) - b(i, \cdot) + b(\cdot, \cdot)]^2 \\ \leq (n-1)^{-1} \sum_i \sum_j [b(i, j) - b(\cdot, \cdot)]^2 = n^2 (n-1)^{-1} \text{Var}\{b(R_1, Q_1)\} \\ \leq n^2 (n-1)^{-1} E\{a(R_1, Q_1; h) - \phi(u^{(R_1)}, v^{(Q_1)})\}^2. \end{aligned}$$

Therefore, unconditionally,

$$E\{(S - T)^2\} \leq n^2 (n-1)^{-1} E\{a(R_1, Q_1; h) - \phi(U_1, V_1)\}^2$$

$$\text{and } E\left\{\frac{(S - T)^2}{\sigma^2}\right\} = \frac{n^2}{(n-1)^2} \frac{E\{a(R_1, Q_1; h) - \phi(U_1, V_1)\}^2}{\iint [\phi(u, v) - \bar{\phi}]^2 du dv}$$

Because $E\{a(R_1, Q_1; h) - \phi(U_1, V_1)\}^2$ converges to zero for square integrable functions (see Theorem a on page 157 of Hájek and Sidák, 1967) and because $\sigma^2 > 0$, S and T are asymptotically identically distributed. However, T is asymptotically normal by the central limit theorem, which proves the theorem. The alternative form for σ^2 is found on p.57 of Hájek and Sidák (1967).

Hájek and Sidák (1967, p.221) were unable to derive the

asymptotic distribution of correlation statistics under the alternative hypothesis suggested by the model (1.5). We, also, were unable to achieve those results under our more general model. This prevents computing asymptotic relative efficiencies for our model. However, under the model discussed by Farlie (1961), the efficiency of Spearman's rho when Fisher-Yates scores are optimal, or vice-versa, is $(3/\pi)^2 = .912$. Similarly it can be shown that the efficiency of Spearman's rho when the top-down correlation coefficient is optimal, or vice-versa, is $(3/4)^2 = .5625$.

BIBLIOGRAPHY

- Barnett, V. (1983), "Reduced Distance Measures and Transformations in Processing Multivariate Outliers," *Australian Journal of Statistics*, 25(1), 64-75.
- Bhuchongkul, S. (1964), "A Class of Nonparametric Tests for Independence in Bivariate Populations," *The Annals of Mathematical Statistics*, 35, 138-149.
- Conover, W.J. (1973), "Rank Tests for One Sample, Two Samples, and k Samples Without the Assumption of a Continuous Distribution Function," *The Annals of Statistics*, 1(6), 1105-1125
- Conover, W.J. (2001), "Some Locally Most Powerful Tests for Correlation," *Journal of Modern Statistical Methods*, 1.
- Cordova, J.R. and Rodriguez-Iturbe, I. (1985), "On the Probabilistic Structure of Storm Surface Runoff," *Water Resources Research*, 21, 755-763.
- Downton, F. (1970), "Bivariate Exponential Distributions in Reliability Theory," *Royal Statistical Society Journal*, 32, 408-417.
- Farlie, D.J.G. (1960), "The Performance of Some Correlation Coefficients for a General Bivariate Distribution," *Biometrika*, 47, 307-323.
- Farlie, D.J.G. (1961), "The Asymptotic Efficiency of Daniels' Generalized Correlation Coefficients," *Royal Statistical Society Journal*, 23, 128-142.
- Fisher, R.A., and Yates, F. (1957), *Statistical Tables for Biological, Agricultural and Medical Research*, (3rd Ed.), Darien, Conn: Hafner.
- Hájek, J., and Sidák, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.

Hawkes, A.G. (1972), "A Bivariate Exponential Distribution with Applications to Reliability," *Royal Statistical Society Journal*, **34**, 129-133.

Iman, R.L. (1987), "Tables of the Exact Quantiles of the Top-Down Correlation Coefficient for $n = 3(1)14$," *Communications in Statistics, Theory and Methods*, **16(5)**, 1513-1540.

Iman, R.L., and Conover, W.J. (1987), "A Measure of Top-Down Correlation," *Technometrics*, **29(3)**, 351-357.

Johnson, N.L., and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley.

Kibble, W.F. (1941), "A Two-Variate Gamma Type Distribution," *Sankhya*, **5**, 137-150.

Konijn, H.S. (1956), "On the Power of Certain Tests for Independence in Bivariate Populations," *The Annals of Mathematical Statistics*, **27**, 300-323.

Mardia, K.V. (1962), "Multivariate Pareto Distributions," *The Annals of Mathematical Statistics*, **33**, 1008-1015.

Mardia, K.V. (1970), *Families of Bivariate Distributions*, Darien, Conn: Hafner.

Morgenstern, D. (1956), "Einfache Beispiele Zweidimensionaler Verteilungen," *Mitteilungsblatt für Mathematische Statistik*, **8**, 234-235.

Nagao, M., and Kadoya, M. (1971), "Two-Variate Exponential Distribution and its Numerical Table for Engineering Applications," *Bull. Disaster Prev. Res. Inst., Kyoto Univ.*, **20**, 183-215.

Plackett, R.L. (1965), "A Class of Bivariate Distributions," *Journal of the American Statistical Association*, **60**, 516-522.

Savage, I.R. (1956), "Contributions to the Theory of Rank Order Statistics - the Two-sample Case," *The Annals of Mathematical Statistics*, **27**, 590-615.

Shiratata, S. (1974), "Locally Most Powerful Rank Tests for Independence," *Bulletin of Mathematical Statistics*, **16**, 11-21.

Wei, L.J. (1981), "Estimation of Location Difference for Fragmentary Samples," *Biometrika*, **68(2)**, 471-476.

The exponential distribution is widely used for waiting times, and for times to failure.

The bivariate exponential distribution may be an appropriate model for some of the following cases:

1. X = the time interval between the arrival of a customer, and the arrival of the previous customer
Y = the time it takes for the newly-arrived customer to get served
2. X = the time in service of an item (e.g., light bulb) until it fails and requires replacement
Y = the time it takes to replace the item
3. X = the length of service of the first bulb in a two-bulb overhead projector
Y = the length of service of the spare bulb in a two-bulb overhead projector
4. X = the time a telephone is available (not in use) until it rings (assuming no call waiting)
Y = the length of the telephone call until the phone is again available

If the correlation (r) between X and Y is positive, then Mardia's (1970) model may be appropriate. See Example 1. In that case the top-down correlation is the locally most powerful rank test.

If the correlation (r) between X and Y is negative, then Gumbel's (1960) model may be appropriate. See Example 2. In that case the top-down correlation is again the locally most powerful rank test.

The Asymptotic Relative Efficiency of Spearman's rho, when the top-down correlation is the locally most powerful rank test, is 0.5625.

Exact quantiles for the top-down correlation coefficient (from Iman and Conover, 1986)

n	0.001	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
4																			
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			

For an approximation for $n > 14$ use either the standard normal quantile divided by $\text{SQRT}(n-1)$, or more exact approximate tables in Iman and Conover (1987).

Uniform on (0,1)

Uniform on (0,1)	$=LN(1-A_3)$
0.355602	0.466018
0.926145	0.256691
0.925596	0.544969
0.802149	0.675954
0.706076	0.111026
0.926939	0.168096
0.584918	0.110508
0.948088	0.431745
0.163274	0.926084
0.151585	0.80166
0.553606	0.306223
0.541856	0.718528
0.671194	0.742393
0.337565	0.942595

	Pearson's Correlation	Pearson's Correlation	Spearman's Correlation
1			
2	-0.57516		-0.54818

† Spearman's

Correlation **-0.6044**

Top-down
Correlation -0.50597

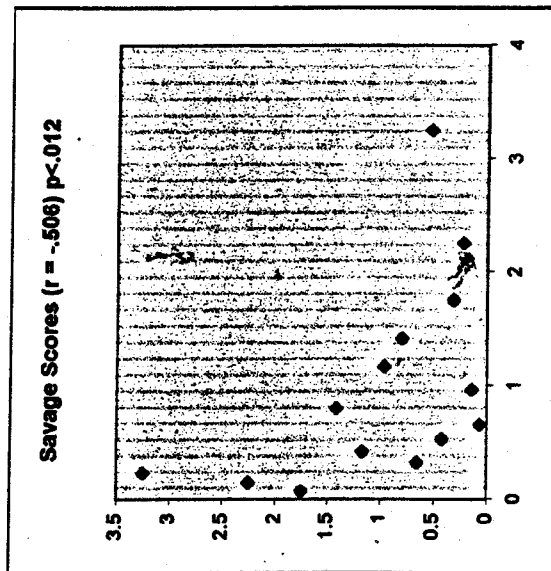
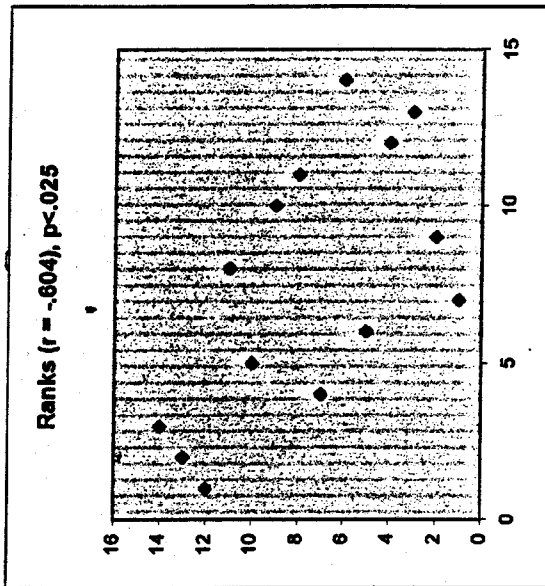
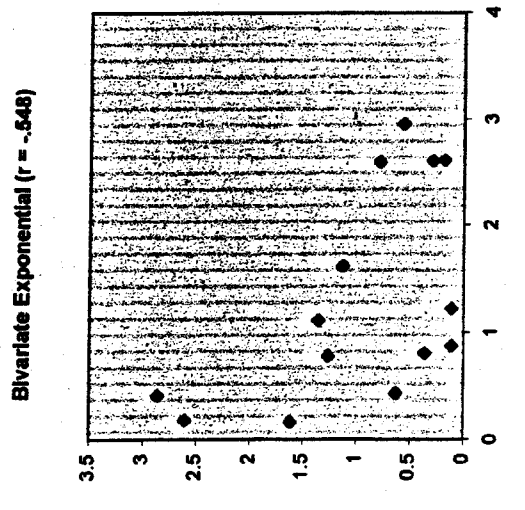
HOW TO FIND SAVAGE SCORES

$$i = 1/(n+1-i) = \text{SUM}(N\$3:N3)$$

1	0.07143	0.07143	0.071429
2	0.07692	0.148352	0.148352
3	0.08333	0.231685	0.231685
4	0.09091	0.322594	0.322594
5	0.1	0.422594	0.422594
6	0.11111	0.533705	0.533705
7	0.125	0.658705	0.658705
8	0.14286	0.801562	0.801562
9	0.16667	0.968229	0.968229
10	0.2	1.168229	1.168229
11	0.25	1.418229	1.418229
12	0.33333	1.751562	1.751562
13	0.5	2.251562	2.251562
14	1	3.251562	3.251562

Savage Scores

0.322594	0.658705
1.751562	0.322594
1.418229	0.801562
1.168229	0.968229
0.968229	0.148352
2.251562	0.231685
0.658705	0.071429
3.251562	0.533705
0.148352	2.251562
0.071429	1.751562
0.533705	0.422594
0.422594	1.168229
0.801562	1.418229
0.231685	3.251562

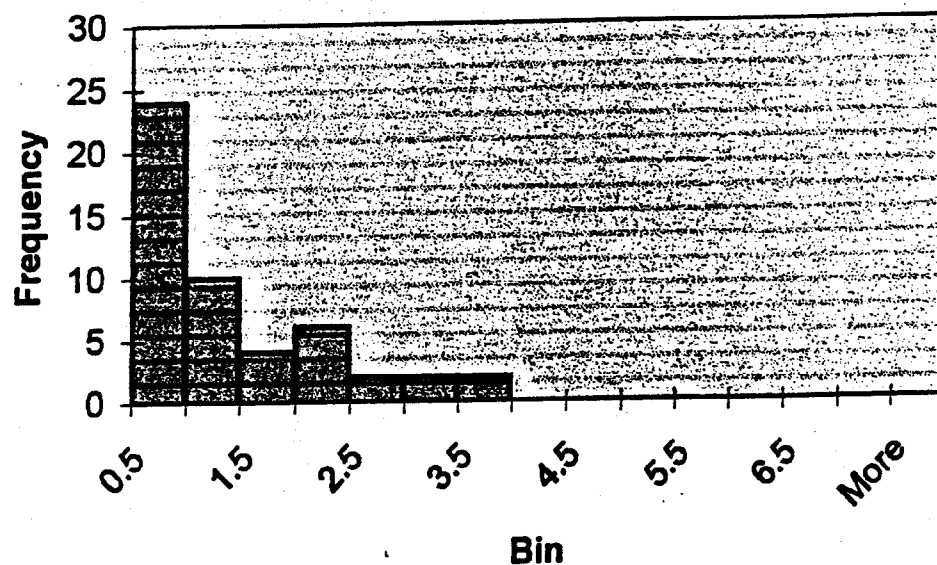


Data Set 1: Positive Correlation

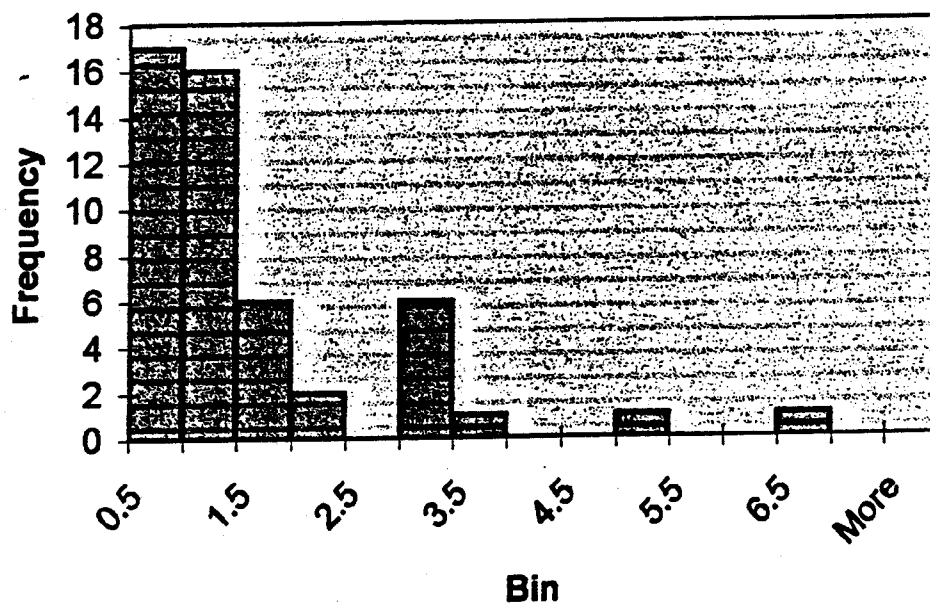
Uniform Distribution		Exponential Distribution		Ranks		scores	scores
0.382	0.355602	0.481267	0.439438	23	16	0.607749	0.380995
0.951689	0.926145	3.0301	2.605654	49	44	3.499205	2.049205
0.756157	0.925596	1.411231	2.598244	38	43	1.395995	1.906348
0.67156	0.802149	1.113401	1.620239	36	40	1.247643	1.570237
0.850429	0.706076	1.899983	1.224435	43	38	1.906348	1.395995
0.65685	0.926939	1.069587	2.616456	35	45	1.180976	2.215872
0.601886	0.584918	0.921017	0.879279	34	29	1.118476	0.853847
0.545518	0.948088	0.788598	2.958206	30	47	0.901466	2.665872
0.259865	0.163274	0.300923	0.178259	11	8	0.245662	0.172463
0.014222	0.151585	0.014324	0.164386	1	7	0.02	0.149207
0.265145	0.553606	0.308082	0.806553	14	27	0.324646	0.764914
0.943815	0.541856	2.879113	0.780572	48	25	2.999205	0.683247
0.350017	0.671194	0.430809	1.112286	19	34	0.47196	1.118476
0.414594	0.337565	0.53545	0.411833	26	15	0.723247	0.352424
0.202582	0.489242	0.226376	0.67186	9	22	0.196272	0.572034
0.125889	0.602466	0.134548	0.922475	6	31	0.126479	0.951466
0.539445	0.225105	0.775323	0.255027	29	12	0.853847	0.271303
0.332774	0.229682	0.404626	0.260952	17	13	0.410407	0.297619
0.930357	0.927061	2.66437	2.618129	47	46	2.665872	2.415872
0.894345	0.998169	2.247575	6.302833	46	50	2.415872	4.499205
0.218421	0.430342	0.246439	0.562718	10	20	0.220662	0.504218
0.526933	0.920164	0.748517	2.527775	28	42	0.808392	1.781348
0.87994	0.150853	2.119765	0.163523	45	6	2.215872	0.126479
0.856441	0.598682	1.941009	0.913	44	30	2.049205	0.901466
0.167089	0.685781	0.182828	1.157666	8	36	0.172463	1.247643
0.362743	0.813227	0.450582	1.67286	21	41	0.537552	1.670237
0.378277	0.689261	0.475261	1.1688	22	37	0.572034	1.319072
0.81106	0.052797	1.666325	0.054242	41	3	1.670237	0.061241
0.0983	0.992523	0.103474	4.895919	4	49	0.082518	3.499205
0.262673	0.215155	0.304724	0.24227	13	11	0.297619	0.245662
0.753624	0.436689	1.400897	0.573924	37	21	1.319072	0.537552
0.260292	0.190527	0.3015	0.211372	12	9	0.271303	0.196272
0.325571	0.387371	0.39389	0.489996	16	17	0.380995	0.410407
0.280221	0.549089	0.328811	0.796485	15	26	0.352424	0.723247
0.56563	0.399518	0.833859	0.510022	32	18	1.004097	0.44071
0.046297	0.050508	0.047403	0.051828	3	2	0.061241	0.040408
0.556139	0.001495	0.812243	0.001497	31	1	0.951466	0.02
0.798975	0.194739	1.604324	0.216588	39	10	1.479328	0.220662
0.960997	0.06711	3.244126	0.069468	50	4	4.499205	0.082518
0.112796	0.532334	0.119681	0.760002	5	24	0.104257	0.644786
0.808893	0.744102	1.654922	1.362978	40	39	1.570237	1.479328
0.357006	0.953429	0.441619	3.066772	20	48	0.504218	2.999205
0.154027	0.408094	0.167268	0.524407	7	19	0.149207	0.47196
0.39198	0.67391	0.497547	1.120581	24	35	0.644786	1.180976
0.457289	0.610858	0.611179	0.943812	27	32	0.764914	1.004097
0.596088	0.1377	0.906557	0.148151	33	5	1.059653	0.104257
0.031892	0.491134	0.032411	0.675571	2	23	0.040408	0.607749
0.816156	0.553758	1.69367	0.806895	42	28	1.781348	0.808392
0.397412	0.617573	0.506522	0.961216	25	33	0.683247	1.059653
0.335673	0.286447	0.408981	0.337498	18	14	0.44071	0.324646

Column 1 Column 2		Column 1 Column 2		Column 1 Column 2		Column 1 Column 2	
Column 1	1	Column 1	1	Column 1	1	Column 1	1
Column 2	0.166783	1 Column 2	0.207295	1 Column 2	0.18213685	1 Column 2	0.175186

Histogram



Histogram



Data Set 2: Positive Correlation

Uniform		Bivariate Exponential		Ranks		Savage Scores	
0.466018	0.303903	0.627392	0.362267	24	11	0.644786	0.245662
0.256691	0.679647	0.296644	1.138332	16	31	0.380995	0.951466
0.544969	0.674978	0.78739	1.123862	28	30	0.808392	0.901466
0.675954	0.948515	1.126871	2.96647	32	48	1.004097	2.999205
0.111026	0.386334	0.117688	0.488304	9	15	0.196272	0.352424
0.168096	0.005158	0.184038	0.005171	10	1	0.220662	0.02
0.110508	0.558916	0.117104	0.81852	8	27	0.172463	0.764914
0.431745	0.967467	0.565186	3.425509	21	49	0.537552	3.499205
0.926084	0.381664	2.604828	0.480724	44	14	2.049205	0.324646
0.80166	0.695242	1.617774	1.188238	40	32	1.570237	1.004097
0.306223	0.227607	0.365604	0.258262	19	7	0.47196	0.149207
0.718528	0.76693	1.267722	1.456417	37	36	1.319072	1.247643
0.742393	0.795495	1.356321	1.587165	38	38	1.395995	1.395995
0.942595	0.15302	2.857618	0.166078	45	4	2.215872	0.082518
0.511582	0.245125	0.716583	0.281203	25	9	0.683247	0.196272
0.980438	0.806055	3.934148	1.64018	47	40	2.665872	1.570237
0.103946	0.796197	0.109755	1.590603	7	39	0.149207	1.479328
0.693106	0.884823	1.181252	2.161286	34	45	1.118476	2.215872
0.254158	0.758171	0.293242	1.419526	15	35	0.352424	1.180976
0.554918	0.832606	0.809497	1.787405	29	41	0.853847	1.670237
0.294473	0.460952	0.34881	0.61795	18	18	0.44071	0.44071
0.531144	0.027345	0.75746	0.027725	26	2	0.723247	0.040408
0.04944	0.054598	0.050704	0.056145	3	3	0.061241	0.061241
0.043733	0.222968	0.044718	0.252274	2	6	0.040408	0.126479
0.989959	0.64687	4.601119	1.04092	49	29	3.499205	0.853847
0.993774	0.50029	5.079057	0.693727	50	21	4.499205	0.537552
0.86578	0.747276	2.008272	1.375458	43	34	1.906348	1.118476
0.308939	0.946837	0.369527	2.934388	20	47	0.504218	2.665872
0.535844	0.486953	0.767535	0.667388	27	20	0.764914	0.504218
0.985809	0.78341	4.25514	1.52975	48	37	2.999205	1.319072
0.213904	0.582293	0.240677	0.872975	13	28	0.297619	0.808392
0.094119	0.157109	0.098847	0.170918	5	5	0.104257	0.104257
0.849574	0.51149	1.894286	0.716396	41	22	1.670237	0.572034
0.713584	0.318979	1.250309	0.384163	36	12	1.247643	0.271303
0.862239	0.87289	1.982238	2.062706	42	44	1.781348	2.049205
0.270363	0.40492	0.315209	0.519059	17	17	0.410407	0.410407
0.947844	0.39375	2.953514	0.500463	46	16	2.415872	0.380995
0.243294	0.539262	0.27878	0.774926	14	24	0.324646	0.644786
0.176305	0.557878	0.193955	0.81617	11	26	0.245662	0.723247
0.70745	0.479141	1.229118	0.652275	35	19	1.180976	0.47196
0.098849	0.321421	0.104083	0.387754	6	13	0.126479	0.297619
0.684805	0.518601	1.154563	0.731059	33	23	1.059653	0.607749
0.083865	0.88525	0.087591	2.165003	4	46	0.082518	2.415872
0.769921	0.871456	1.469332	2.051484	39	43	1.479328	1.906348
0.599841	0.992553	0.915894	4.900009	30	50	0.901466	4.499205
9.16E-05	0.250984	9.16E-05	0.288995	1	10	0.02	0.220662
0.664602	0.720847	1.092436	1.275996	31	33	0.951466	1.059653
0.438154	0.866176	0.576528	2.011232	23	42	0.607749	1.781348
0.179174	0.236579	0.197444	0.269946	12	8	0.271303	0.172463
0.436964	0.552263	0.574412	0.803549	22	25	0.572034	0.683247

Column 1	Column 2	Column 1	Column 2	Column 1	Column 2	Column 1
Column 1	1	Column 1	1	Column 1	1	Column 1
Column 2	0.30382	1 Column 2	0.058951	1 Column 2	0.281633	1 Column 2
						0.049993

Data Set 3: Negative Correlation

Uniform Distribution		Exponential Distribution		Ranks		scores	
0.382	0.466018	0.481267	0.627392474	23	24	0.607749	0.644786
0.951689	0.256691	3.0301	0.296643686	49	16	3.499205	0.380995
0.756157	0.544969	1.411231	0.787389783	38	28	1.395995	0.808392
0.67156	0.675954	1.113401	1.126871237	36	32	1.247643	1.004097
0.850429	0.111026	1.899983	0.11768767	43	9	1.906348	0.196272
0.65685	0.168096	1.069587	0.18403817	35	10	1.180976	0.220662
0.601886	0.110508	0.921017	0.117104229	34	8	1.118476	0.172463
0.545518	0.431745	0.788598	0.56518564	30	21	0.901466	0.537552
0.259865	0.926084	0.300923	2.604828266	11	44	0.245662	2.049205
0.014222	0.80166	0.014324	1.617773592	1	40	0.02	1.570237
0.265145	0.306223	0.308082	0.365604298	14	19	0.324646	0.47196
0.943815	0.718528	2.879113	1.267721547	48	37	2.999205	1.319072
0.350017	0.742393	0.430809	1.356321126	19	38	0.47196	1.395995
0.414594	0.942595	0.53545	2.857618361	26	45	0.723247	2.215872
0.202582	0.511582	0.226376	0.71658322	9	25	0.196272	0.683247
0.125889	0.980438	0.134548	3.934147733	6	47	0.126479	2.665872
0.539445	0.103946	0.775323	0.109754648	29	7	0.853847	0.149207
0.332774	0.693106	0.404626	1.18125244	17	34	0.410407	1.118476
0.930357	0.254158	2.66437	0.293241694	47	15	2.665872	0.352424
0.894345	0.554918	2.247575	0.809496874	46	29	2.415872	0.853847
0.218421	0.294473	0.246439	0.348810377	10	18	0.220662	0.44071
0.526933	0.531144	0.748517	0.75745989	28	26	0.808392	0.723247
0.87994	0.04944	2.119765	0.050703979	45	3	2.215872	0.061241
0.856441	0.043733	1.941009	0.044718141	44	2	2.049205	0.040408
0.167089	0.989959	0.182828	4.60111944	8	49	0.172463	3.499205
0.362743	0.993774	0.450582	5.079057197	21	50	0.537552	4.499205
0.378277	0.86578	0.475261	2.008272019	22	43	0.572034	1.906348
0.81106	0.308939	1.666325	0.369526995	41	20	1.670237	0.504218
0.0983	0.535844	0.103474	0.767534553	4	27	0.082518	0.764914
0.262673	0.985809	0.304724	4.255139785	13	48	0.297619	2.999205
0.753624	0.213904	1.400897	0.240676653	37	13	1.319072	0.297619
0.260292	0.094119	0.3015	0.09884742	12	5	0.271303	0.104257
0.325571	0.849574	0.39389	1.894285784	16	41	0.380995	1.670237
0.280221	0.713584	0.328811	1.250309241	15	36	0.352424	1.247643
0.56563	0.862239	0.833859	1.982238233	32	42	1.004097	1.781348
0.046297	0.270363	0.047403	0.31520878	3	17	0.061241	0.410407
0.556139	0.947844	0.812243	2.953513507	31	46	0.951466	2.415872
0.798975	0.243294	1.604324	0.278779891	39	14	1.479328	0.324646
0.960997	0.176305	3.244126	0.193955484	50	11	4.499205	0.245662
0.112796	0.70745	0.119681	1.229118211	5	35	0.104257	1.180976
0.808893	0.098849	1.654922	0.104082946	40	6	1.570237	0.126479
0.357006	0.684805	0.441619	1.154563258	20	33	0.504218	1.059653
0.154027	0.083865	0.167268	0.087591397	7	4	0.149207	0.082518
0.39198	0.769921	0.497547	1.469332364	24	39	0.644786	1.479328
0.457289	0.599841	0.611179	0.91589407	27	30	0.764914	0.901466
0.596088	9.16E-05	0.906557	9.15597E-05	33	1	1.059653	0.02
0.031892	0.664602	0.032411	1.092436143	2	31	0.040408	0.951466
0.816156	0.438154	1.69367	0.576527916	42	23	1.781348	0.607749
0.397412	0.179174	0.506522	0.197444335	25	12	0.683247	0.271303
0.335673	0.436964	0.408981	0.574411743	18	22	0.44071	0.572034

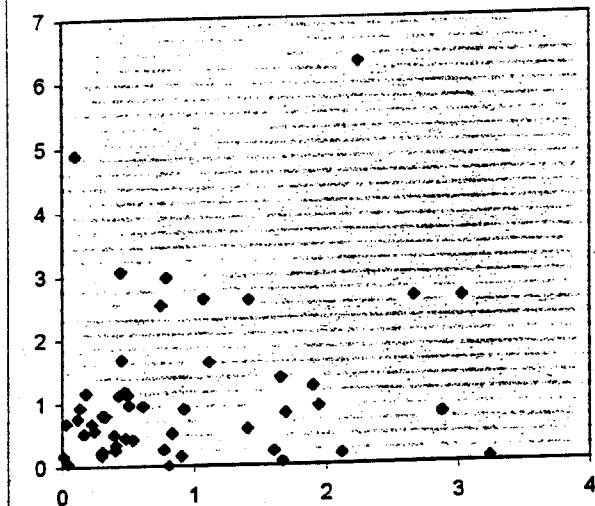
Column 1	Column 2	Column 1	Column 2	Column 1	Column 2	Column 1
Column 1	1	Column 1	1	Column 1	1	Column 1
Column 2	-0.439328	1 Column 2	-0.36762093	1 Column 2	-0.40389	1 Column 2
						-0.341062

Data Set 4: Negative Correlation

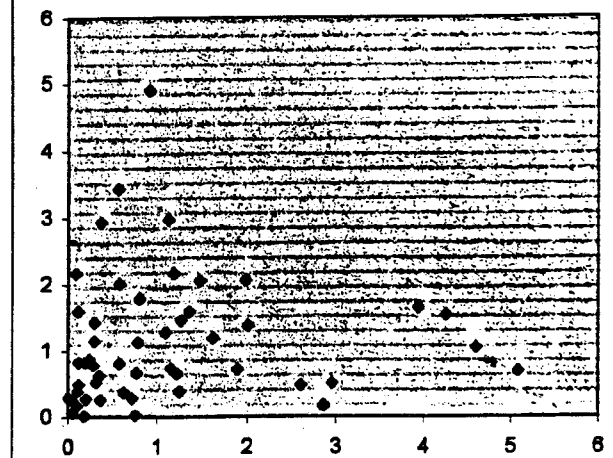
Uniform		Bivariate Exponential		Ranks		Savage Scores	
0.910306	0.303903	2.411353	0.362267	47	11	2.665872	0.245662
0.100314	0.679647	0.10571	1.138332	6	31	0.126479	0.951466
0.903867	0.674978	2.342019	1.123862	46	30	2.415872	0.901466
0.789026	0.948515	1.556018	2.96647	44	48	2.049205	2.999205
0.40144	0.386334	0.513229	0.488304	24	15	0.644786	0.352424
0.451765	0.005158	0.601052	0.005171	27	1	0.764914	0.02
0.497024	0.558916	0.687214	0.81852	29	27	0.853847	0.764914
0.250771	0.967467	0.28871	3.425509	13	49	0.297619	3.499205
0.808313	0.381664	1.651893	0.480724	45	14	2.215872	0.324646
0.284555	0.695242	0.33485	1.188238	15	32	0.352424	1.004097
0.958678	0.227607	3.186359	0.258262	50	7	4.499205	0.149207
0.223121	0.76693	0.25247	1.456417	12	36	0.271303	1.247643
0.576434	0.795495	0.859045	1.587165	30	38	0.901466	1.395995
0.695029	0.15302	1.187537	0.166078	41	4	1.670237	0.082518
0.405713	0.245125	0.520393	0.281203	26	9	0.723247	0.196272
0.270211	0.806055	0.315	1.64018	14	40	0.324646	1.570237
0.62215	0.796197	0.973259	1.590603	34	39	1.118476	1.479328
0.382153	0.884823	0.481514	2.161286	21	45	0.537552	2.215872
0.947508	0.758171	2.947098	1.419526	49	35	3.499205	1.180976
0.033418	0.832606	0.033989	1.787405	1	41	0.02	1.670237
0.656911	0.460952	1.069765	0.61795	38	18	1.395995	0.44071
0.62804	0.027345	0.98897	0.027725	36	2	1.247643	0.040408
0.072329	0.054598	0.075078	0.056145	5	3	0.104257	0.061241
0.774468	0.222968	1.489294	0.252274	43	6	1.906348	0.126479
0.636311	0.64687	1.011456	1.04092	37	29	1.319072	0.853847
0.403394	0.50029	0.516498	0.693727	25	21	0.683247	0.537552
0.159276	0.747276	0.173492	1.375458	11	34	0.245662	1.118476
0.043214	0.946837	0.044176	2.934388	3	47	0.061241	2.665872
0.334574	0.486953	0.407328	0.667388	17	20	0.410407	0.504218
0.59801	0.78341	0.911329	1.52975	32	37	1.004097	1.319072
0.114872	0.582293	0.122023	0.872975	8	28	0.172463	0.808392
0.470717	0.157109	0.636233	0.170918	28	5	0.808392	0.104257
0.043519	0.51149	0.044495	0.716396	4	22	0.082518	0.572034
0.376202	0.318979	0.471928	0.384163	20	12	0.504218	0.271303
0.134861	0.87289	0.144865	2.062706	9	44	0.196272	2.049205
0.687185	0.40492	1.162144	0.519059	40	17	1.570237	0.410407
0.396954	0.39375	0.505762	0.500463	23	16	0.607749	0.380995
0.657308	0.539262	1.070922	0.774926	39	24	1.479328	0.644786
0.590869	0.557878	0.89372	0.81617	31	26	0.951466	0.723247
0.364879	0.479141	0.45394	0.652275	18	19	0.44071	0.47196
0.319742	0.321421	0.385284	0.387754	16	13	0.380995	0.297619
0.721122	0.518601	1.27698	0.731059	42	23	1.781348	0.607749
0.042085	0.88525	0.042996	2.165003	2	46	0.040408	2.415872
0.386273	0.871456	0.488205	2.051484	22	43	0.572034	1.906348
0.102145	0.992553	0.107747	4.900009	7	50	0.149207	4.499205
0.938353	0.250984	2.786324	0.288995	48	10	2.999205	0.220662
0.372143	0.720847	0.465442	1.275996	19	33	0.47196	1.059653
0.136784	0.866176	0.14709	2.011232	10	42	0.220662	1.781348
0.599048	0.236579	0.913913	0.269946	33	8	1.059653	0.172463
0.626179	0.552263	0.983978	0.803549	35	25	1.180976	0.683247

Column 1	Column 2	Column 1	Column 2	Column 1	Column 2	Column 1
Column 1	1	Column 1	1	Column 1	1	Column 1
Column 2	-0.36194	1 Column 2	-0.29665	1 Column 2	-0.4206	1 Column 2
						-0.31448

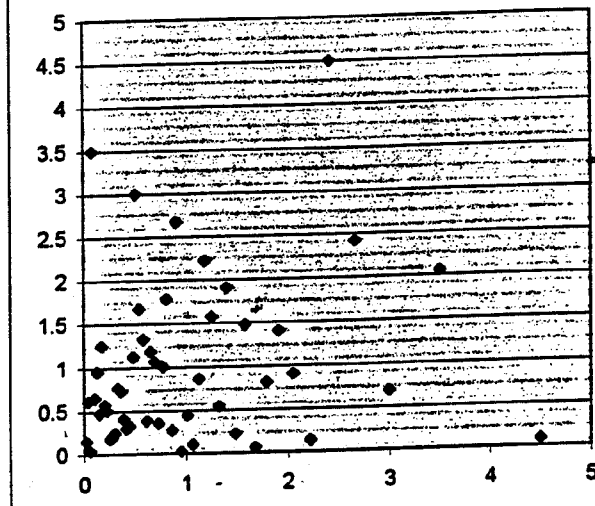
Bivariate Exponential ($r = .207$)



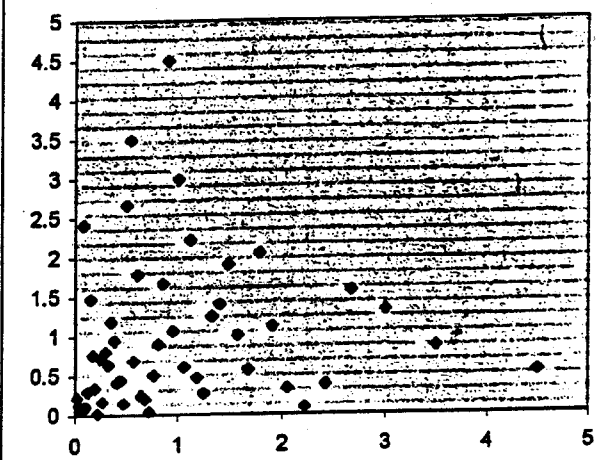
Bivariate Exponential ($r = .059$)



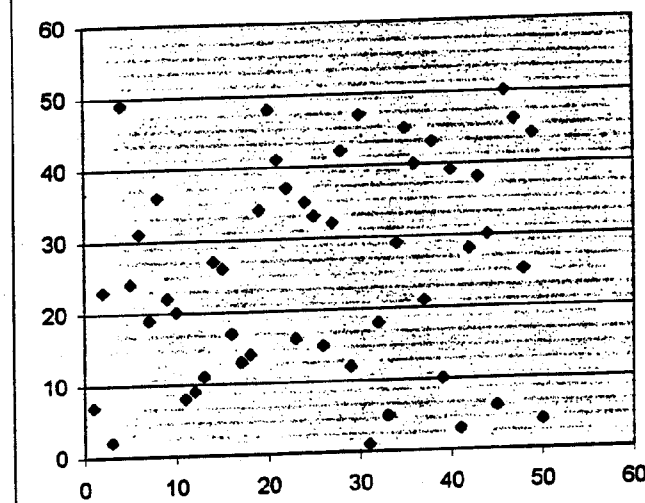
Savage Scores ($r = .175$)



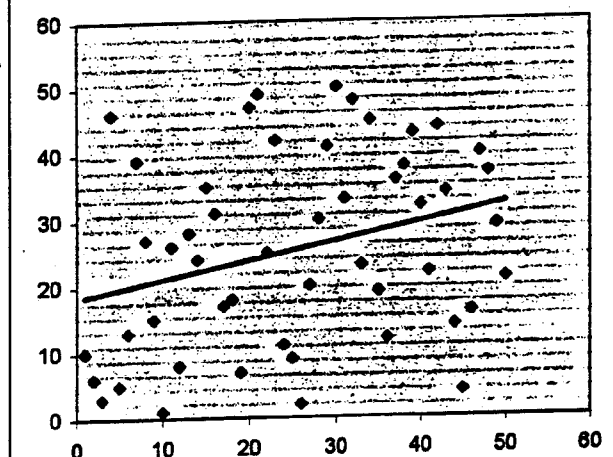
Savage Scores ($r = .050$)



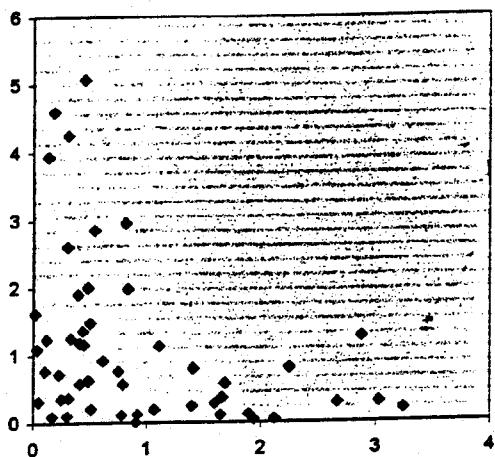
Ranks ($r = .182$)



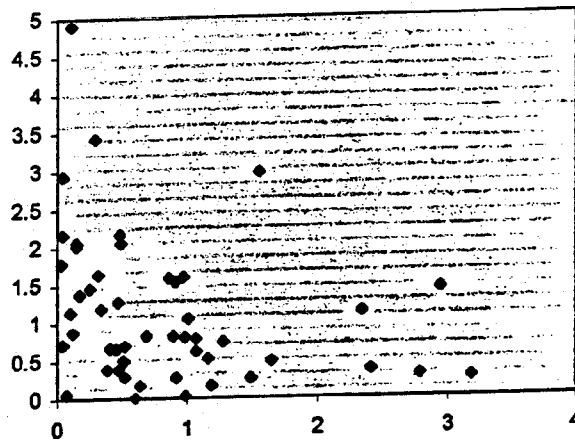
Ranks ($r = .282$)



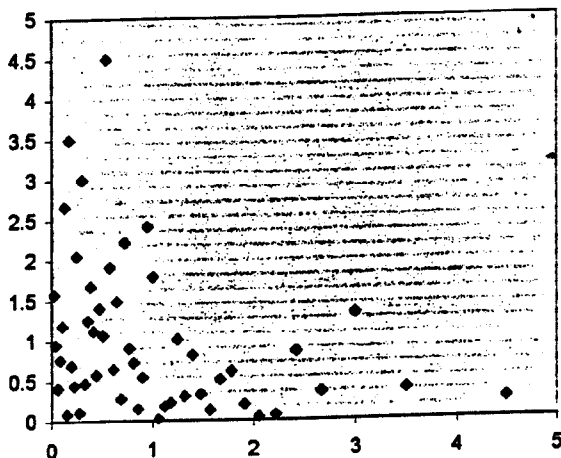
Bivariate Exponential ($r = -.368$)



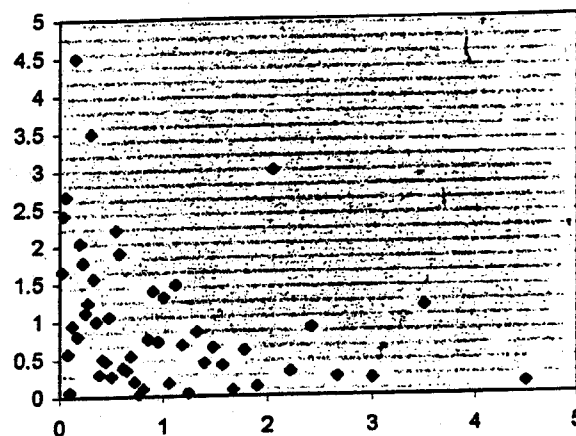
Bivariate Exponential ($r = -.297$)



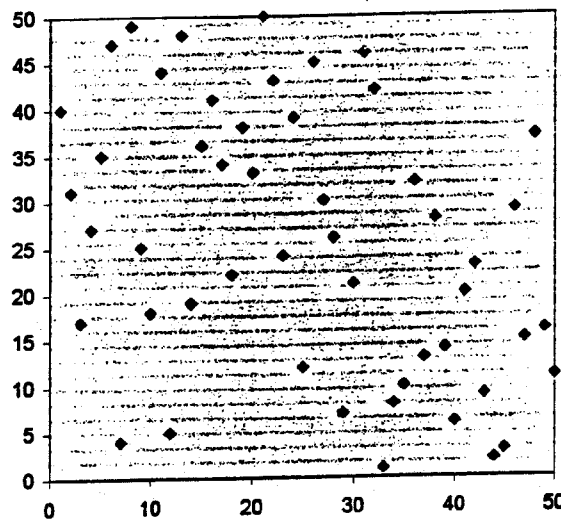
Savage Scores ($r = -.341$)



Savage Scores ($r = -.314$)



Ranks ($r = -.404$)



Ranks ($r = -.421$)

