BU-894-M

Tests for Equality of Variances with Paired Data

CHARLES E. McCULLOCH*

The normal theory test for equality of variances with paired data is shown to be nonrobust to violation of the assumption of normality. Nonparametric tests are shown to provide a much safer alternative with little loss of efficiency.

Key words: robustness; Pearson's product-moment correlation coefficient; Spearman's rank correlation coefficient; jackknifing.

*Charles E. McCulloch is Assistant Professor, Biometrics Unit, Cornell University, Ithaca, NY 14853. This paper is BU-894-M in the Biometrics Unit Series, Cornell University, Ithaca, NY 14853.

1. INTRODUCTION AND NOTATION

The F-test for equality of variances in two independent normal samples is well-known to be nonrobust to the assumption of normality. For example, see Conover *et al.* (1981). Pitman (1939) proposed a test for paired, normally distributed data based on the correlation between the sums and differences within the pairs. Ekbohm (1981) conjectured that Pitman's test would also be nonrobust, though calculations by Bansal and Srivastava (1977) had not supported this conjecture for the two-sided test. They concluded that, "In each case the sum of the two-tail contents is not very different from the normal theory value. Hence, on the whole, the two-sided test is very little affected by nonnormality as compared to one-tailed tests." Bell, Rothstein and Li (1982) conducted a simulation which showed the size of Pitman's test could be larger than nominal for nonnormal distributions. They recommended the use of a method proposed by Rothstein *et al.* (1981), which jackknifes the log of the ratio of the sample variances. In Section 2 below, some calculations are done which shed light on the sensitivity of Pitman's test. A simple alternative test is proposed. In Section 3 are reported the results of a simulation study which is more extensive than the one described in Bell, Rothstein and Li (1982).

We now establish the basic notation. Let (X_1,Y_1), (X_2,Y_2), $\cdots$, (X_n,Y_n) denote i.i.d. pairs of observations and let $D_i = X_i-Y_i$ $(1\leq i\leq n)$ and $S_i = X_i+Y_i$ $(1\leq i\leq n)$. Pitman (1939) noted that $\mathrm{Cov}(S_i,D_i) = \sigma_X^2-\sigma_Y^2$ and thus a test of $H_0:\sigma_X^2=\sigma_Y^2$ is equivalent to $H_0:\rho_{DS}=0$, where ρ_{DS} denotes the population correlation between D_i and S_i. When (X_i,Y_i) are bivariate normal, (D_i,S_i) are also bivariate normal and therefore a test of $H_0:\rho_{DS}=0$ can be made referring r_{DS}, the Pearson product-moment correlation coefficient between D_i and S_i to the usual tables for significance. Explicitly, the two-sided test of $H_o: \sigma_X^2 = \sigma_Y^2$ versus $H_A: \sigma_X^2 \neq \sigma_Y^2$ is given by: reject H_0 if and only if $|r_{DS}| \geq r_\alpha$. This test is recommended, for example, in Snedecor and Cochran (1980, p. 190); a table of r_α can also be found there (Table A11(i)). Finally, we will also need the following notation:

$$\mu_{r,s} \equiv E\left[(X-\mu_X)^r (Y-\mu_Y)^s\right]$$

and

$$m_{r,s} \equiv \frac{1}{n} \sum_{i=1}^{n} (X_i-\bar{X})^r (Y_i-\bar{Y})^s \quad .$$

2. PERFORMANCE OF PITMAN'S STATISTIC AND AN ALTERNATIVE

The behavior of Pitman's test statistics under $H_0:\rho_{DS}=0$ and hence the size of the test can be elucidated by calculation of $\mathrm{Var}(n^{\frac{1}{2}}r_{DS})$. Kowalski (1972) states that, "It is the variance of r which is most vulnerable to nonnormality ····." Also, for normally or approximately normally distributed data when $\rho_{DS}=0$, the distribution of $n^{\frac{1}{2}}r_{DS}$ is well approximated by a standard normal distribution. Thus, calculation of $\mathrm{Var}(n^{\frac{1}{2}}r_{DS})$ should give valuable information.

Before deriving the variance and the normal approximation, it is useful to note the following algebraic identity

Lemma

$$r_{DS} = \frac{m_{2,0}-m_{0,2}}{\sqrt{(m_{0,2}+m_{2,0})^2-4m_{1,1}^2}} = \frac{S_X^2-S_Y^2}{\sqrt{(S_X^2+S_Y^2)^2-4S_{XY}^2}} \quad .$$

To derive the asymptotic distribution of r_{DS} we use the known fact that $(m_{2,0},m_{1,1},m_{0,2})$ is asymptotically multivariate normal (Cramér, 1946, p. 366) and hence a multivariate delta method can be applied. In the case of $\rho_{DS}=0$, the asymptotic distribution takes a simple form.

Proposition. Let r_{DS} be the Pearson product-moment correlation of $D_i=X_i-Y_i$ and $S_i=X_i+Y_i$, where (X_i,Y_i) are i.i.d. such that $\mu_{4,0}$, $\mu_{0,4}$ and $\mu_{2,2}$ are finite and nonzero. Assume $\rho_{DS}=0$ $(\sigma_X^2=\sigma_Y^2\equiv\sigma^2)$, then

$$n^{\frac{1}{2}}r_{DS} \sim AN\left(0,\frac{1}{4\sigma^4}\left[\frac{\mu_{4,0}+\mu_{0,4}-2\mu_{2,2}}{1-\rho_{XY}^2}\right]\right) . \qquad (2.1)$$

Proof. We follow the development of Serfling (1980, pp. 122-124). Regarding r_{DS} as a function of its numerator and denominator we see that, under $\rho_{DS}=0$, it can be approximated by the (nonzero) differential $(S_X^2-S_Y^2)/E\left[\left((S_X^2+S_Y^2)^2-4S_{XY}^2\right)^{\frac{1}{2}}\right]$. Since the numerator is asymptotically normal, the asymptotic distribution of $n^{\frac{1}{2}}r_{DS}$ will also be normal. Since $\rho_{DS}=0$, the mean will be zero and the variance can be calculated as

$$\mathrm{Var}_{\infty}(n^{\frac{1}{2}}r_{DS}) = \mathrm{Var}_{\infty}\left(n^{\frac{1}{2}}(S_X^2-S_Y^2)\right) / E_{\infty}^2\left[\left((S_X^2+S_Y^2)^2-4S_Y^2\right)^{\frac{1}{2}}\right] ,$$

where $\mathrm{Var}_{\infty}(\cdot)$ denotes asymptotic variance and $E_{\infty}[\cdot]$ denotes asymptotic mean. To the requisite orders of n,

$$\begin{aligned}\mathrm{Var}(S_X^2-S_Y^2) &= \mathrm{Var}(S_X^2) + \mathrm{Var}(S_Y^2) - 2\mathrm{Cov}(S_X^2,S_Y^2)\\ &= \frac{1}{n}\left[\mu_{4,0} - \frac{n-3}{n-1}\sigma^4\right] + \frac{1}{n}\left[\mu_{0,4} - \frac{n-3}{n-1}\sigma^4\right]\\ &\quad - \frac{2}{n}[\mu_{2,2}-2\sigma^4] + o\left(\frac{1}{n}\right)\\ &= \frac{1}{n}[\mu_{4,0}+\mu_{0,4}-2\mu_{2,2}] + o\left(\frac{1}{n}\right)\end{aligned}$$

and

$$E\left[\left((S_X^2+S_Y^2)^2-4S_{XY}^2\right)^{\frac{1}{2}}\right] = \left((\sigma^2+\sigma^2)^2 - 4\mu_{1,1}^2\right)^{\frac{1}{2}} + o(1) \quad .$$

Thus,

$$\begin{aligned}\mathrm{Var}_{\infty}(n^{\frac{1}{2}}r_{DS}) &= \frac{\mu_{4,0}+\mu_{0,4}-2\mu_{2,2}}{4(\sigma^4-\mu_{1,1}^2)}\\ &= \frac{1}{4\sigma^4}\left[\frac{\mu_{4,0}+\mu_{0,4}-2\mu_{2,2}}{1-\rho_{XY}^2}\right] \quad .\end{aligned}$$

Two Corollaries are immediate.

Corollary: If, in addition to the assumptions of the Proposition, (X_i,Y_i) are bivariate normal, then (2.1) simplifies to

$$\mathrm{Var}(n^{\frac{1}{2}}r_{DS}) = 1 \quad .$$

Corollary: If, in addition to the assumptions of the Proposition, X_i and Y_i are uncorrelated with $\mu_{4,0}=\mu_{0,4}$, then (2.1) simplifies to

$$\mathrm{Var}(n^{\frac{1}{2}}r_{DS}) = \frac{1}{2}(\kappa+2) \quad ,$$

where κ = kurtosis of X_i and Y_i.

These corollaries shed much light on the behavior of r_{DS} as a test statistic. Under the conditions of the second Corollary we see that the variance is larger (smaller) than the normal distribution case whenever the kurtosis is greater (less) than zero. Thus, for sample sizes large enough for the normal approximation to be valid, the size of the test will be larger (smaller) than nominal according as the kurtosis is greater (less) than zero. For an extreme case like the exponential distribution with $\kappa=6$, the variance will quadruple (nominal size .05 would be exceeded by a factor of 6). Such effects will not dissipate as the sample size increases.

In the more general case, the size of the test will be larger (smaller) than nominal according as $\mu_4/\sigma^4-\mu_{2,2}/\sigma^4$ is greater (less) than $2(1-\rho_{XY}^2)$. This is a multivariate version of heavy tailedness.

An easy way to simulate correlated pairs of variables is to generate W_1 and W_2 which are independent with mean zero and variance one and then set

$$\begin{aligned} X &= \mu_X + \sigma_X W_1 \\ Y &= \mu_Y + \sigma_Y(\rho W_1+(1-\rho^2)\ W_2) \quad . \end{aligned} \tag{2.2}$$

In investigating the correlation between X and Y, the means and variances $(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2)$ are irrelevant and it is easiest to take $\mu_X=\mu_Y=0$ and $\sigma_X=\sigma_Y=1$. In that case, the general formula for $\mathrm{Var}(n^{\frac{1}{2}}r_{DS})$ also simplifies as shown below.

<u>Proposition</u>: Assume that the covariance structure between X and Y is the same as that of W_1 and $\rho W_1 + (1-\rho^2)^{\frac{1}{2}}W_2$, where W_1 and W_2 are independent with means zero, variance one and common fourth moment $\mu_4 = E[W_1^4] = \kappa_W+3$. Then

$$\mathrm{Var}(n^{\frac{1}{2}} r_{DS}) = \frac{1}{2}\,[\rho^2(3-\mu_4)+(\mu_4-1)]$$
$$= \frac{1}{2}\,[2+\kappa_W(1-\rho^2)] \quad . \tag{2.3}$$

Proof: Straightforward computations show that, under (2.2),

$$E[X] = E[Y] = 0$$
$$E[X^2] = E[Y^2] = 1$$
$$E[XY] = \rho$$
$$E[X^4] = \mu_4$$
$$E[Y^4] = \rho^4\mu_4 + 6\rho^2(1-\rho^2) + (1-\rho^2)^2\mu_4$$
$$E[X^2Y^2] = \rho^2\mu_4 + (1-\rho^2) \quad .$$

Plugging these in to (2.1) yields (2.3).

Remarks: Depending on whether the kurtosis is greater than or less than zero, the variance has its minimum or maximum at $\rho=0$. Figure 1 illustrates the two cases $\kappa_W=2$ and $\kappa_W=-1.2$ (the kurtosis for a uniform distribution). Thus Pitman's test can be expected to have maximum type I error rate at $\rho=\pm1$ for distributions with kurtosis less than zero and at $\rho=0$ for distributions with kurtosis greater than zero. It also implies that kurtosis greater than zero will lead to a liberal test, while kurtosis less than zero will lead to a conservative test.

In their simulation study, Bell, Rothstein and Li (1982) found that the empirical sizes of Pitman's test were consistently too large. The results on $\mathrm{Var}(n^{\frac{1}{2}} r_{DS})$ indicate why this is so since all of the distributions they used were heavy-tailed. However, the results also indicate that no simple change, such as an increase of the critical values, will improve the approximation to the null distribution. Thus, the suggestion by Bell, Rothstein and Li (1982) to try to improve that approximation is probably not easily achieved.

Bell, Rothstein and Li (1982) also evaluated a nonparametric test proposed by Rothstein *et al.* (1978). That test calculates the n pseudo-values $L_{-1}, L_{-2}, \cdots, L_{-n}$ (removing one observation at a time) for log S_X^2/S_Y^2 and rejects H_0: $\sigma_X^2=\sigma_Y^2$ versus H_A: $\sigma_X^2\neq\sigma_Y^2$ if $|n^{\frac{1}{2}}\bar{L}/S_L|$ exceeds a t-distribution quantile with n-1 degrees of freedom. As an alternative to the above suggestion and to Pitman's test, we propose the use of Spearman's rank correlation coefficient on the D_i and S_i, denoted by $\tilde{r}_{DS}$.

The use of $\tilde{r}_{DS}$ has several advantages over the jackknife procedure. It is easily computed, it has an extensively tabulated distribution in the null hypothesis case of independence and it has known asymptotic relative efficiency to Pitman's test under a variety of distributions. In the next section we evaluate these tests via a simulation study.

3. A SIMULATION STUDY

A simulation study was performed to compare the size and power of five tests:

PEARS: Pitman's test which uses Pearson's product-moment correlation coefficient as a test statistic. The test is: reject H_0 iff $|r_{DS}| \geq r_\alpha$, where r_α denotes the usual normal theory critical values.

SPEAR: The proposed test based on Spearman's rank correlation coefficient. The test is: reject H_0 iff $|\tilde{r}_{DS}| \geq \tilde{r}_\alpha$, where $\tilde{r}_\alpha$ denotes the usual critical values (given in Snedecor and Cochran, 1980, Table A11(ii)).

JRATIO: Jackknife tests based on, respectively, log S_X^2/S_Y^2, r_{DS}, and Fisher's z-transform of r_{DS}. In each case, calculate

JFISHER: pseudo-values L_{-1}, L_{-2}, $L_{-3}, \cdots, L_{-n}$. Next form $\left|n^{\frac{1}{2}}\bar{L}/S_L\right|$ and reject H_0 if it is bigger than the critical value. In each case, preliminary simulations were run to see if the normal or t-distribution percentage points provided a better approximation to the critical values. The t-distribution was used for JRATIO and JPEARS. Normal percentage points were used for JFISHER. These tests were studied in Bell, Rothstein and Li (1982) and Rothstein *et al.* (1981).

Details of the simulation techniques are given in the appendix.

Figures 2, 3 and 4 show the size of the tests for normally distributed data (nominal α is .10), sample sizes 10, 27 and 52 and a range of values of ρ_{DS}. Only JPEARS failed to control the size to an acceptable level. It was therefore eliminated from further study. JRATIO appeared to be slightly liberal for values of ρ close to -1 or 1. The sizes of PEARS and SPEAR, a check on the simulation, were within sampling error of the nominal sizes in all cases (due to discreteness, the sizes of SPEAR for $n=10$ are .096, .048 and .01 for the .10, .05 and .01 tests).

Next, the sizes of the tests were investigated under nonnormal distributions. Using the device described by (2.2), X and Y were generated from W_i which had exponential (EXPO), uniform(0,1)(UNIF), normal(0,1) contaminated with 10% normal(0,9) (N/N) and 5% and 1% standard Cauchy contaminating a normal(0,1) (N/C5 and N/C1).

Figures 5, 6 and 7 show the sizes of the tests at a nominal α of .10 for these nonnormal distributions. Only SPEAR came close to controlling the type I error rate. Performance of the other tests was especially poor for the exponential, N(0,1)/5% C(0,1) and the N(0,1)/10% N(0,9) distributions.

Finally, Figure 8 shows the power of the tests for normal distributions. Of course, PEARS has the highest power, but the nonparametric tests are competitive. In view of the poor control of type I error by PEARS for nonnormal distributions, this seems like a small price to pay to achieve close to the proper size.

4. CONCLUSIONS

Theoretical results predicting the sensitivity of Pitman's test to the kurtosis of the underlying distribution were upheld by a simulation. Pitman's test was found to greatly exceed the nominal size for distributions with high kurtosis. A test based on Spearman's rank correlation coefficient was much better at controlling the type I error rate at close to nominal. It performed better than a nonparametric jackknife procedure proposed by Rothstein *et al.* (1981). The test based on Spearman's correlation also has advantages in terms of ease of computation and tabulation of its null distribution.

APPENDIX

All simulations were run on an IBM PC-XT. Uniform pseudo-random numbers were generated via the method of Wichmann and Hill (1982). Normal pseudo-random numbers were generated via an acceptance-rejection version of the Box-Muller algorithm. All other distributions were generated using inverse c.d.f.'s.

Each simulation utilized 3,600 replications. This number was chosen to estimate the size of the tests at a nominal $\alpha = .10$ to within $\pm .01$ $\left(2\sqrt{\frac{.9(.1)}{3600}} = .01\right)$. The different tests were compared using common random numbers. More simulations were run than are reported here. The following additional simulations were performed:

N = 27 ρ = 0,.9 all nonnormal distributions

N = 27 ρ = 0,.9 variance ratio = 1, 1.5, 2, 3, 4, normal

N = 52 ρ = .9 UNIF, N/C5, EXPO

Various nonnormal power studies.

Also, information for a nominal alpha of .05 and .01 was recorded in all the simulations. The details are available from the author.

REFERENCES

BANSAL, A. K., and SRIVASTAVA, R. B. L., (1977), "Robustness of the Test for Equality of Variances," *Journal of the Indian Statistical Association,* 15, 81-90.

BELL, W. D., ROTHSTEIN, S. M., and LI, W., (1982), "Empirical Study of Some Nonparametric Tests for Dispersion of Correlated Data," *Journal of Statistical Computation and Simulation,* 15, 9-15.

CONOVER, W. J., JOHNSON, M. E., and JOHNSON, M. M., (1981), "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Shelf Bidding Data," *Technometrics,* 23, 351-361.

CRAMÉR, H., (1945), *Mathematical Methods in Statistics,* Princeton, NJ: Princeton University Press.

EKBOHM, G., (1981), "A Test for Equality of Variances in the Paired Case with Incomplete Data," *Biometrical Journal,* 23, 261-265.

KOWALSKI, C. J., (1972). "On the Effects of Non-normality on the Distribution of the Sample Product-Moment Correlation Coefficient," *Applied Statistic,* 21, 1-12.

PITMAN, E. J. G., (1939), "A Note on Normal Correlation," *Biometrika,* 31, 9-12.

ROTHSTEIN, S. M., BELL, W. D., PATRICK, J. A., and MILLER, H., (1981), "A Jackknife Test of Homogeneity of Variance with Paired Replicates Data," *Psychometrika,* 46, 35-40.

SNEDECOR, G. W., and COCHRAN, W. G., (1980), *Statistical Methods, 7th* ed., Ames, IA: Iowa State University Press.

SERFLING, R. J., (1980), *Approximation Theorems in Mathematical Statistics,* New York: Wiley.

WICHMANN, B. A., and HILL, I. D., (1982), Algorithm AS183, "An Efficient and Portable Pseudo-Random Number Generator," *Applied Statistics,* 31, 188-190.

Titles of Figures

Figure 1: Asymptotic variance of r

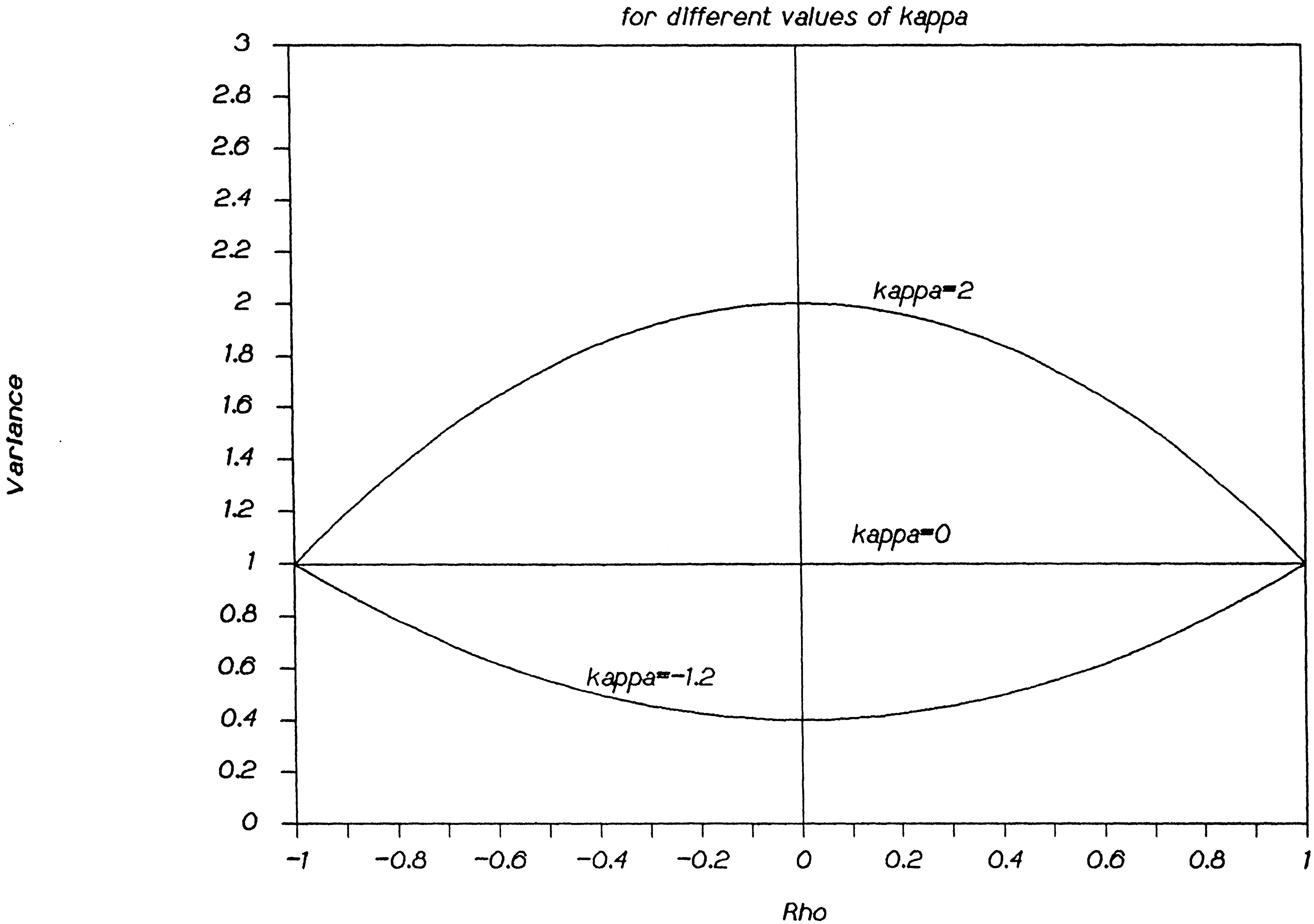

FIGURE 2: Empiric sizes at alpha=0.1

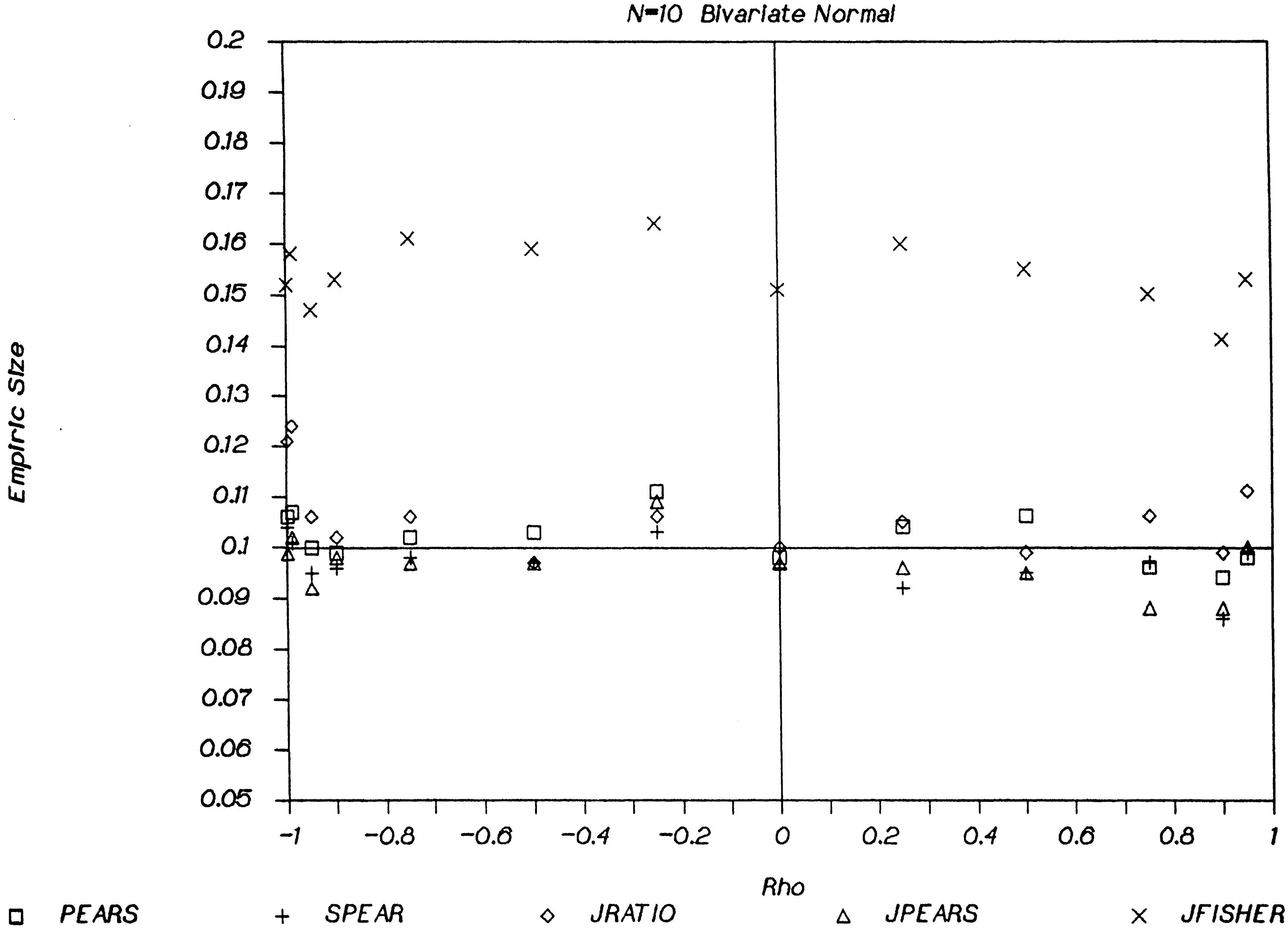

FIGURE 3: Empiric sizes at alpha=0.1

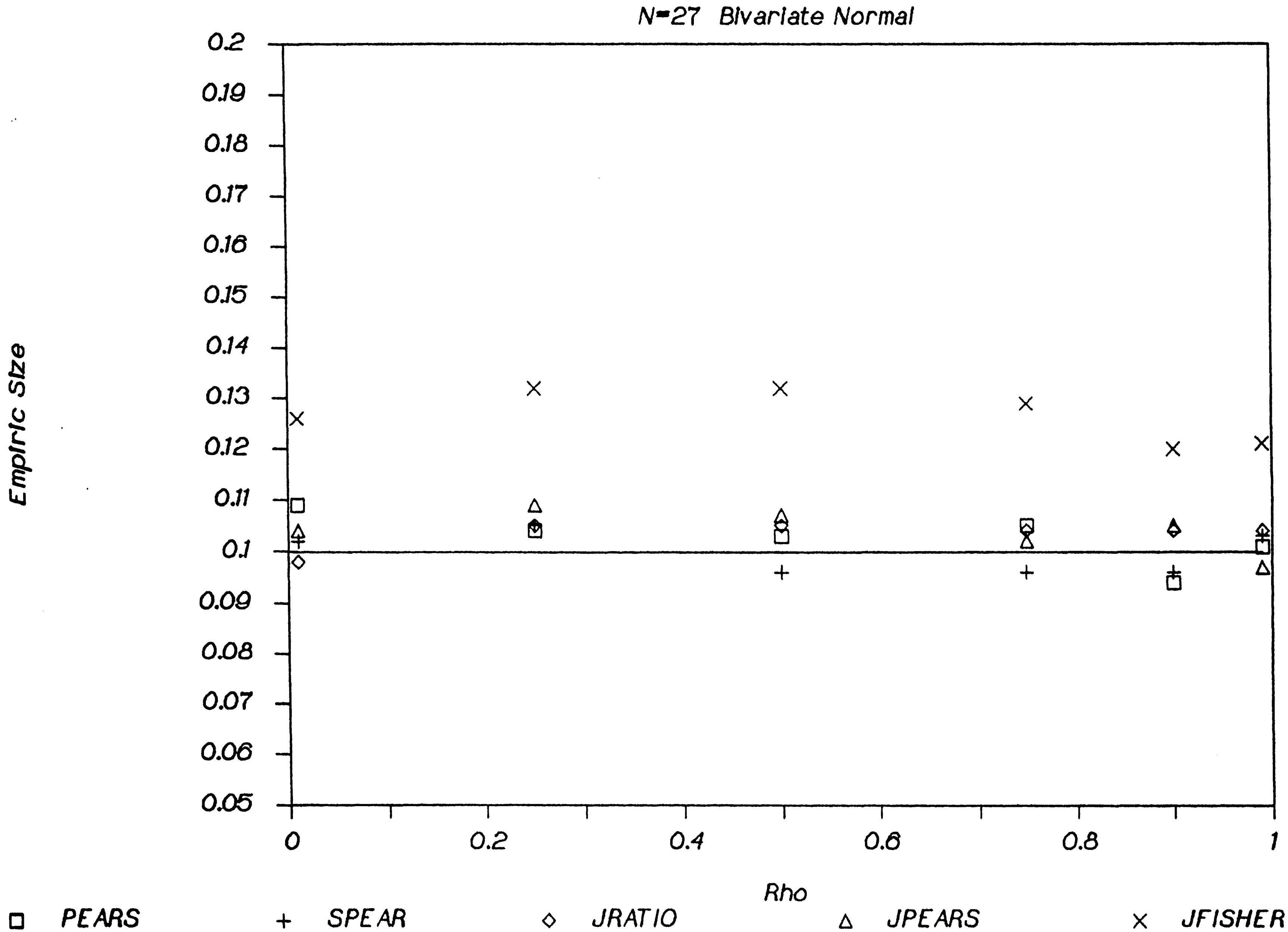

FIGURE 4: Empiric sizes at alpha=0.1

N=52 Bivariate Normal

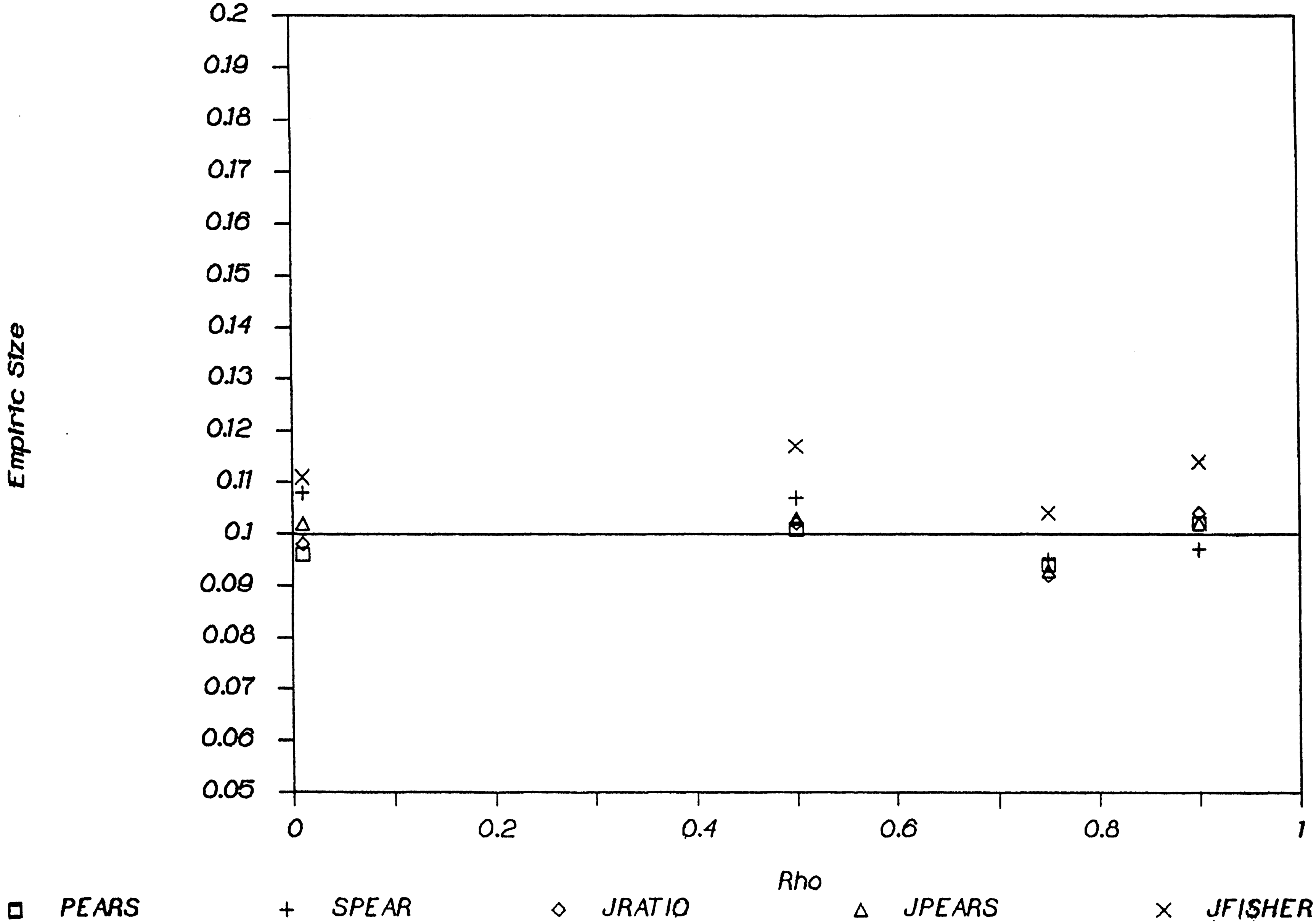

FIGURE 5: Empiric sizes at alpha=0.1

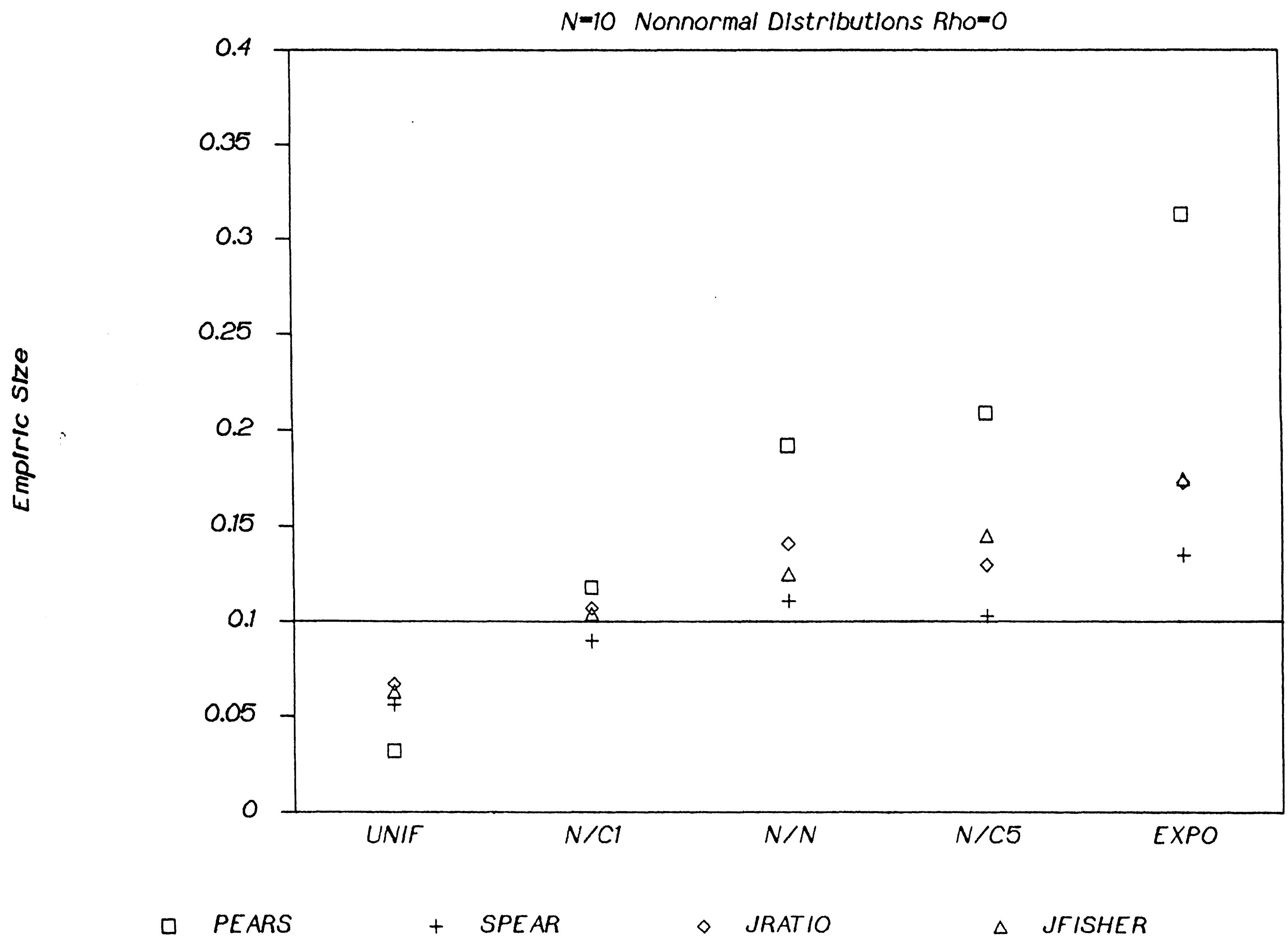

FIGURE 6: Empiric sizes at alpha=0.1

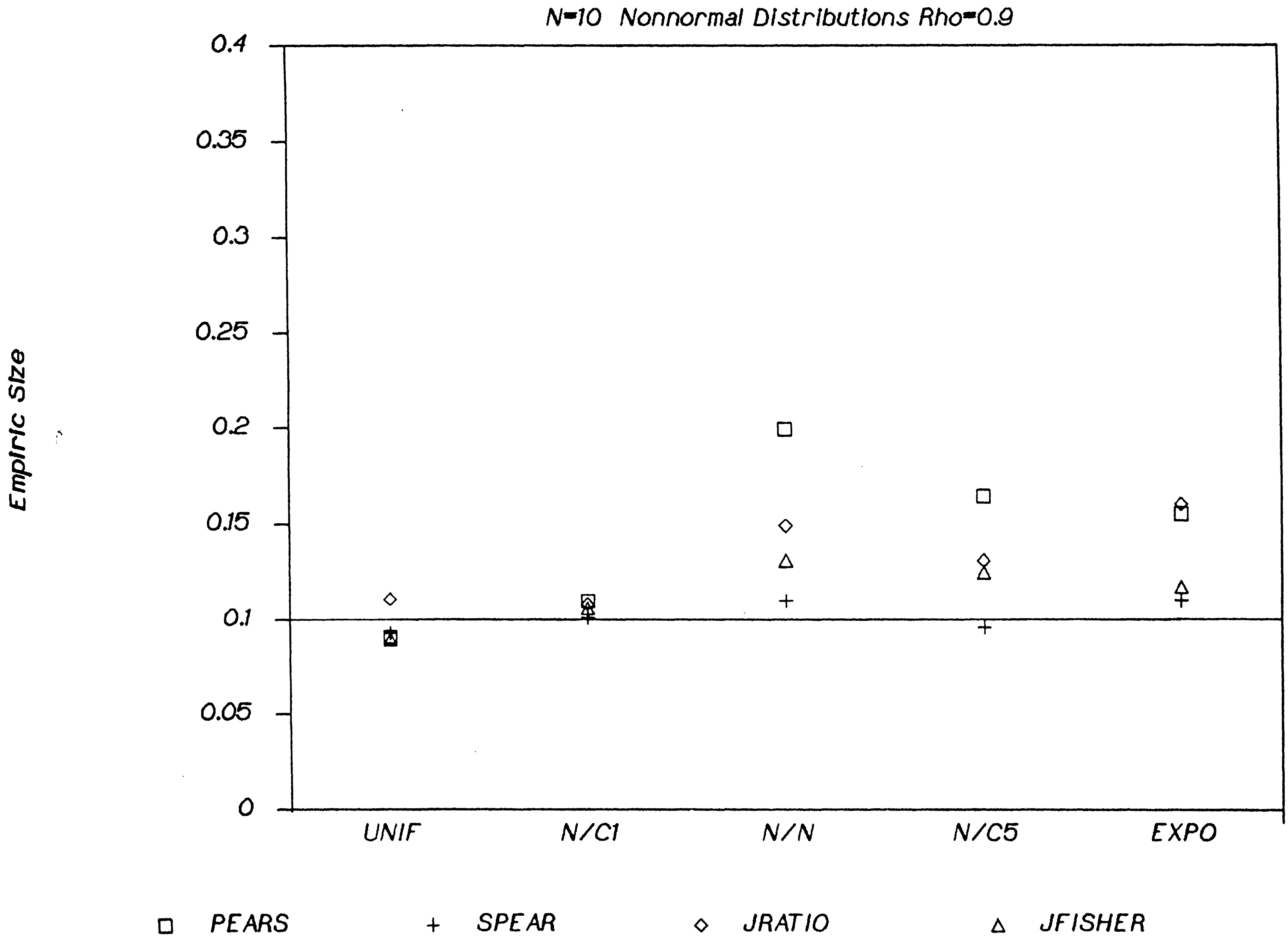

FIGURE 7: Empiric sizes at alpha=0.1

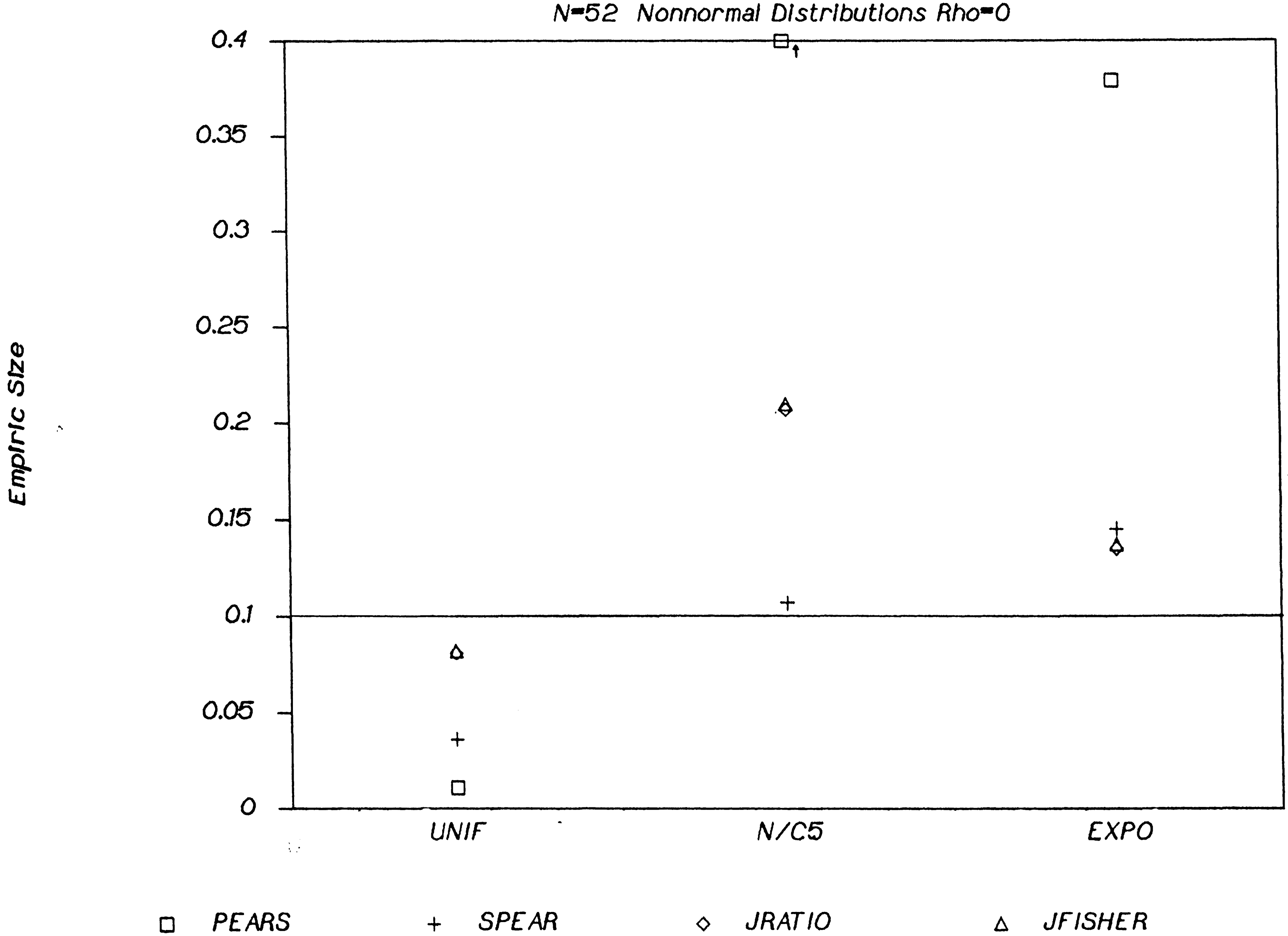

FIGURE 8: Empiric power at alpha=0.1