STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Report S 363

Experimental determination of the power functions of the two-sample rank tests of WILCOXON, VAN DER WAERDEN and TERRY by Monte Carlo techniques for normal parent populations.

by

P. van der Laan and J. Oosterhoff

September 1966

The Mathematical Centre at Amsterdam, founded the 11th of February, 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and the Central Organization for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Introduction

Assume two independent random samples

$$\underline{x}_1^{1)}, \cdots, \underline{x}_m \quad \text{and} \quad \underline{y}_1, \underline{y}_2, \cdots, \underline{y}_n^{1)} \tag{1}$$

are given from populations with continuous cumulative distribution functions $F(x)$ and $G(y)$, respectively. One wishes to test the hypothesis

$$H_0 : F(x) \equiv G(x) \tag{2}$$

against alternative hypotheses of the form

$$H_1 : F(x) \equiv G(x - \delta) \tag{3}$$

for $\delta > 0$ (one-sided test).

Three well-known nonparametric rank tests for this two-sample problem are:

a. the two-sample test of WILCOXON (MANN-WHITNEY),

b. " " " " VAN DER WAERDEN (X - test),

c. " " " " TERRY (FISHER-YATES).

The last two tests are often called "normal scores" tests, since their test statistics are weighted rank sums with weights derived from the normal distribution.

In a former paper [4] the authors discussed a method for determining the power functions of these tests against shift alternatives by Monte Carlo techniques, and for comparing these tests with each other and with parametric tests. In this paper the results of such a Monte Carlo experiment are presented for normal parent distributions with equal but unknown variances (that is, under STUDENT - conditions). The parametric test considered for comparison is the STUDENT t-test for two samples, which is the uniformly most powerful unbiased (UMPU) test for this problem. The sample sizes considered are $m = n = 6$ and $m = n = 10$.

---

1) Random variables will be distinguished from fixed numbers (e.g. from values they assume in an experiment) by underlining their symbols.

In the sequel we shall restrict our attention to the following sizes α of the tests (9 for each sample size)[2]:

$m = n = 6$: α = .004329 , .005 , .007576 , .01 , .020563 , .025 , .046537 , .05 , .089827

$m = n = 10$: α = .004465 , .005 , .009272 , .01 , .025 , .026213 , .05 , .052561 , .095158.

The sizes .005, .01, .025 and .05 were chosen to conform with statistical practice. For the rank tests with these sizes randomization is necessary at the boundary points of the critical regions owing to the discrete character of the test statistics. The five other sizes are exact significance levels of the WILCOXON test statistic; in most cases randomization had to be applied to obtain the same sizes for the normal scores tests.

A shift alternative is characterized by the quantity

$$d = (\varepsilon \underline{x} - \varepsilon \underline{y})/\sigma, \tag{4}$$

where σ is the common standard deviation of both $\underline{x}$ and $\underline{y}$. The shifts considered in our simulation experiment are

$m = n = 6$: $d = 0.2\ (0.2)\ 2.2\ ,\ 2.5$

$m = n = 10$: $d = 0.4\ (0.2)\ 1.4$.

For each shift alternative and each sample size 2000 sample pairs have been drawn from pseudo normal populations using a generator described in [4]. For different shifts and/or sample sizes the estimated powers are thus independent. However, for a given shift they are strongly dependent for different sizes of the tests, since the same samples were used to estimate the powers for different sizes. The choice of 2000 sample pairs was motivated by the following considerations:

---

2) Actually about 30 sizes were considered in the experiment.

(i) the expected width of a two-sided central confidence interval with confidence level 0.98 for the power of a test, based on the Monte Carlo results, is smaller than 0.0525,

(ii) the UMPU test for equality of two powers has very large power against alternatives of interest.

This last statement needs some further elaboration. Let T_1 and T_2 be two tests whose powers are to be tested for equality. Call p_1 the probability of the event: H_0 is rejected by T_1 but not by T_2, p_2 the probability of the event: H_0 is rejected by T_2 but not by T_1, and p_3 the probability of the event: the test results of T_1 and T_2 are identical ($p_1 + p_2 + p_3 = 1$). To test the hypothesis $G_0 : p_1 = p_2$ against $G_1 : : p_1 > p_2$ the UMPU size - α' test is given by the conditional binomial size - α' test for $p_1 = p_2$ against $p_1 > p_2$ given the number of trials with a different test result of T_1 and T_2 (cf. LEHMANN [5], Ch. 4). Consider a fixed alternative $p_1 - p_2 = \Delta > 0$. In [6] it is proved that the (unconditional) power of this test for fixed Δ is a strictly decreasing function of $p = p_1 + p_2$ (provided $\alpha' > 2^{-M}$, where M is the number of trials). Since the test statistics in our experiment are expected to show strong positive correlation, small values of p seem most likely. In [4] approximate values of the power of the UMPU test have been given for some values of p. Exact values of the power for some Δ and p are given in table 1.

TABLE 1

Power of the test for equal powers as a function of p

Δ \ p	.02	.04	.05	.07	.1
.01	.8205	.4661	.3723	.2618	.1803
.02	~1	.9891	.9600	.8605	.6954
.04		~1	~1	~1	.9997

If the normal approximation is used in the conditional binomial test or if the boundary points of the conditional critical regions (where in general randomization is necessary to get the exact size α') are exclu-

ded from the critical regions, the entries in this table will nevertheless be good approximations since M is large.

More details and a brief review of the literature are given in [4].

2. Numerical results

The estimated powers of the three rank tests and the t-test are shown in tables 3 to 20, together with the exact power of the t-test. These powers have also been plotted in the figures 1 to 6.

For each shift and each size the powers of the three rank tests have been tested pairwise for equality with a two-sided equal-tails version of the test discussed in section 1. The two-sided tail probabilities smaller than .1 are shown in table 21. To test the powers of a rank test and the t-test for equality did not seem appropriate since the three one-sided rank tests are unbiased and the t-test is thus uniformly more powerful.

The conjecture, that the probability p of a different test result for a pair of rank tests would be small, was confirmed by the experiment. Comparing the WILCOXON test with one of the normal scores tests, the number of different test results was never larger than 101 (out of 2000 trials); comparing the normal scores tests with each other the number of different test results was always smaller than 43. A confidence upper bound for p with confidence level .999, valid for all sizes, shifts and sample sizes, is then given by (i) .068 if the WILCOXON test is compared with one of the normal scores tests, (ii) .034 if the normal scores tests are compared with each other. From the monotonicity property mentioned in section 1 we conclude, that the entries in the 4^{th} (case (i)) and 2^{nd} (case (ii)) column respectively of table 1 in section 1 may serve as lower confidence bounds with confidence level .999 for the power of the test for equal powers against the alternatives Δ considered. These powers seem large enough to detect any appreciable differences Δ between the powers of the three rank tests.

The critical region of each rank test consists of a set of combinations of x-ranks and y-ranks. As a check of the results the critical

regions were written down explicitly for the sample sizes m = n = 6 and the 9 sizes considered (for m = n = 10 this is hardly possible because of the overwhelming number of combinations of x- and y-ranks in the critical regions). For α = .004329 , .007576 and .020563 all three rank tests have completely identical regions. Obviously the estimated powers should also be equal in these cases. For α = .005 , .01 , .025 , .046537 and .05 the critical regions of the two normal scores tests are also identical, but here randomization is necessary in the boundary points of the critical regions. Since the randomization was independently performed for different tests, small differences between the estimated powers are to be expected. Inspection of table 19 shows that in three cases the differences are even significant at the 10% level of the two-sided test for equality of the powers. In the figures the mean of both powers has been plotted as it is the best available estimate of both powers.

If the results of the Monte Carlo experiment are to be trusted, the experimentally obtained powers of the t-test should be in good agreement with the theoretical powers of the t-test. Large deviations indicate that the normality and/or independence assumption of the generated deviates is not acceptable. For each sample size and each shift a chi square test for goodness of fit to the non-central t-distribution has been applied; the results are given in table 22. Inspection of the tail probabilities suggests that the fit is satisfactory except for the case m = n = 6 and d = 1.4 where the fit is rather bad.

3. Discussion of results

This Monte Carlo experiment was mainly performed to investigate the following problems:

(i) How much power does one loose by using the WILCOXON test instead of the STUDENT t-test under STUDENT-conditions in small samples?

(ii) Are the powers of the normal scores tests under STUDENT-conditions in small samples appreciably larger than the power of the WILCOXON test? Are any differences detectable between the powers of the VAN DER WAERDEN-test and the TERRY-test?

The results of the experiment enable us to give tentative answers to these questions. It is clear from the tables that the powers of the normal scores tests are almost equal; the differences between the estimated powers never exceed .0055 and are never significant at the 1% level.

For m = n = 6 the normal scores tests seem to be only slightly more powerful than the WILCOXON test, if at all. For m = n = 10 the normal scores tests have systematically somewhat larger empirical power than the WILCOXON test, especially for not too small sizes of the tests, but the differences are not very convincing (cf. the relatively small number of small tail probabilities in table 19). The asymptotic PITMAN (local) efficiency of the normal scores tests relative to the t-test for normal shift alternatives is 1 and of the WILCOXON test is $3/\pi \approx$ $\approx .955$, but for small sample sizes the normal scores tests appear to loose much of this advantage.

The estimated loss of power of the rank tests relative to the t-test decreases as the size of the tests increases and is larger for m = n = 6 than for m = n = 10. In table 2 the loss of power (in % of the power of the t-test) is roughly summarized for alternatives for which the power of the t-test lies between .2 and .8 [1].

TABLE 2

Loss of power of the rank tests relative to the t-test

sample sizes	size of the test	Wilcoxon test	normal scores tests
m = n = 6	$\leq .01$	7	$6\frac{1}{2}$
	$> .01$	$4\frac{1}{2}$	4
m = n = 10	$\leq .01$	$4\frac{1}{2}$	4
	$> .01$	4	$2\frac{1}{2}$

1) The estimated powers of the rank tests are compared with the estimated powers of the t-test, not with the exact powers of the t-test.

Since this conclusions are based on a Monte Carlo experiment involving only a moderate number of trials, comparison with other sampling studies seems of interest. Small scale experiments were performed by HEMELRIJK[3] and by DIXON and TEICHROEW[1], involving 50 and 150 trials, respectively. In these studies the power of the WILCOXON test was compared with the power of the t-test; the normal scores tests were not considered. HEMELRIJK finds that for sample sizes $m = n = 10$ and $\alpha = .025$ the empirical power of the WILCOXON test is about 9/10 of the power of the t-test for the alternatives $d = 1.05$ and 1.53. In his study the exact size of the WILCOXON test is not .025 but .02163 since he did not apply randomization in the boundary point of the critical region. Keeping this in mind, his results agree remarkably well with our findings. In this experiments of DIXON and TEICHROEW (who considered two-sided tests for sample sizes $m = n = 5$, 10 and 20) no clear cut differences between the powers of the t-test and the WILCOXON test are observable; for most alternatives the t-test is more powerful, but for other alternatives (e.g. for $d = 1$) the WILCOXON test rejects H_o more often. Recently DOKSUM and THOMPSON[2] estimated the power functions of several distribution-free tests, including the WILCOXON test and (in most cases) both normal scores tests, using 1000, 2000 or 3000 trials. They considered sample sizes $m = n = 5$, 8, 10 and 20, sizes $\alpha = .01$ and .05 and alternatives $d = 0.1$, 0.2, 0.3, 0.5, 1, 1.5 and 2. For $m = n = 5$ the three rank tests are identical (for the sizes considered), for $m = n = 8$ the TERRY-test is slightly more powerful than the VAN DER WAERDEN-test in their experiments, for $m = n = 10$ the situation is reserved (not agreeing with our results). If we compare the estimated power functions of the three rank tests with the exact power of the t-test for $m = n = 10$, the loss of power of the WILCOXON test relative to the t-test is of the same order of magnitude as in our experiment, but the estimated powers of the normal scores tests are often larger than the power of the t-test. The discrepancies between their results and our own may perhaps be explained by the fact that they did not apply randomization and approximated the critical values of the normal scores tests for $m = n = 10$, resulting in different sizes of the tests compared.

Acknowledgement

The authors are much indebted to Prof. J. HEMELRIJK of the Mathematisch Centrum, Amsterdam, for his valuable comments. The computations were performed on a ELECTROLOGICA - X1 computer with an ALGOL system developed by the Mathematisch Centrum. The figures were drawn by a plotter attached to this computer. Mr. E.A. SCHWERDT assisted us with the programming of this plotter.

References

[1] DIXON, W.J. and D. TEICHROEW (1953), Some sampling results on the power of nonparametic tests against normal alternatives. Report U.S. Dept. of Commerce, N.B.S. Abstract Ann. Math. Stat. 25, 175.

[2] DOKSUM, K.A. and R. THOMPSON (1964), Unpublished.

[3] HEMELRIJK, J. (1961), Experimental comparison of Student's and Wilcoxon's two-sample tests. Quant. Methods in Pharmacology, ed. by H. de Jonge, North Holland Publ. Co., 118-134.

[4] LAAN, P. VAN DER and J. OOSTERHOFF (1965), Monte Carlo estimation of the powers of the distribution-free two-sample tests of Wilcoxon, van der Waerden and Terry and comparison of these powers. Statistica Neerlandica 19, 265-275.

[5] LEHMANN, E.L. (1959), Testing statistical hypotheses. J. Wiley & Sons, London.

[6] OOSTERHOFF, J. (1966), A monotonicity property of the power of the test for symmetry in a 2 × 2 table and the sign test, Report S 364 of the Mathematical Centrum, Amsterdam.

TABLE 3

Sample sizes m = n = 6; size α = .004329

shift	power of the test of				
	Student(exact)[1]	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0098	.0120	.0105	.0105	.0105
0.4	.0205	.0155	.0205	.0205	.0205
0.6	.0399	.0400	.0395	.0395	.0395
0.8	.0718	.0795	.0795	.0795	.0795
1.0	.1201	.1215	.1175	.1175	.1175
1.2	.1872	.1865	.1730	.1730	.1730
1.4	.2732	.2675	.2460	.2460	.2460
1.6	.3749	.3700	.3430	.3430	.3430
1.8	.4861	.4705	.4415	.4415	.4415
2.0	.5983	.5895	.5425	.5425	.5425
2.2	.7028	.7125	.6705	.6705	.6705
2.5	.8314	.8245	.7710	.7710	.7710

TABLE 4

Sample sizes m = n = 6; size α = .005

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0112	.0125	.0130	.0120	.0135
0.4	.0233	.0195	.0235	.0220	.0220
0.6	.0449	.0440	.0445	.0420	.0420
0.8	.0799	.0880	.0860	.0830	.0840
1.0	.1322	.1330	.1260	.1250	.1255
1.2	.2039	.2035	.1860	.1915	.1895
1.4	.2943	.2875	.2645	.2650	.2685
1.6	.3997	.3905	.3650	.3655	.3650
1.8	.5129	.5005	.4600	.4595	.4615
2.0	.6248	.6180	.5670	.5640	.5680
2.2	.7271	.7370	.6885	.6900	.6900
2.5	.8496	.8435	.7875	.7880	.7885

1) The exact powers of the Student t-test may be in error by one or two units in the fourth decimal place.

TABLE 5

Sample sizes m = n = 6; size α = .007576

Shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0166	.0185	.0180	.0180	.0180
0.4	.0335	.0295	.0300	.0300	.0300
0.6	.0625	.0575	.0620	.0620	.0620
0.8	.1079	.1235	.1125	.1125	.1125
1.0	.1729	.1710	.1625	.1625	.1625
1.2	.2582	.2625	.2415	.2415	.2415
1.4	.3613	.3530	.3340	.3340	.3340
1.6	.4752	.4660	.4350	.4350	.4350
1.8	.5912	.5790	.5495	.5495	.5495
2.0	.6994	.6940	.6635	.6635	.6635
2.2	.7924	.7955	.7655	.7655	.7655
2.5	.8953	.8945	.8630	.8630	.8630

TABLE 6

Sample sizes m = n = 6; size α = .01

Shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0215	.0205	.0210	.0220	.0210
0.4	.0425	.0380	.0390	.0410	.0400
0.6	.0777	.0740	.0740	.0775	.0765
0.8	.1312	.1420	.1330	.1315	.1295
1.0	.2057	.2040	.1875	.1920	.1895
1.2	.3004	.2990	.2740	.2800	.2780
1.4	.4107	.4040	.3765	.3785	.3755
1.6	.5286	.5160	.4770	.4830	.4800
1.8	.6437	.6375	.5965	.5960	.5985
2.0	.7469	.7415	.7085	.7020	.7030
2.2	.8318	.8360	.7970	.7985	.7930
2.5	.9204	.9215	.8885	.8850	.8825

TABLE 7

Sample sizes m = n = 6 ; size α = .020563

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0418	.0405	.0370	.0370	.0370
0.4	.0779	.0785	.0730	.0730	.0730
0.6	.1338	.1300	.1260	.1260	.1260
0.8	.2123	.2195	.2110	.2110	.2110
1.0	.3123	.3085	.2885	.2885	.2885
1.2	.4284	.4275	.4075	.4075	.4075
1.4	.5511	.5370	.5095	.5095	.5095
1.6	.6687	.6475	.6240	.6240	.6240
1.8	.7715	.7760	.7415	.7415	.7415
2.0	.8534	.8510	.8285	.8285	.8285
2.2	.9128	.9120	.8930	.8930	.8930
2.5	.9656	.9665	.9530	.9530	.9530

TABLE 8

Sample sizes m = n = 6 : size α = .025

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0499	.0480	.0480	.0480	.0470
0.4	.0915	.0875	.0830	.0855	.0860
0.6	.1542	.1535	.1480	.1490	.1500
0.8	.2401	.2470	.2405	.2415	.2415
1.0	.3469	.3390	.3175	.3235	.3235
1.2	.4673	.4645	.4440	.4460	.4475
1.4	.5909	.5760	.5430	.5505	.5490
1.6	.7056	.6835	.6565	.6590	.6595
1.8	.8025	.8025	.7675	.7765	.7760
2.0	.8772	.8765	.8455	.8540	.8525
2.2	.9294	.9280	.9085	.9120	.9125
2.5	.9736	.9735	.9570	.9620	.9615

TABLE 9

Sample sizes m = n = 6 ; size α = .046537

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0874	.0840	.0805	.0785	.0780
0.4	.1503	.1480	.1415	.1450	.1440
0.6	.2377	.2405	.2335	.2350	.2350
0.8	.3474	.3500	.3315	.3370	.3385
1.0	.4717	.4675	.4505	.4555	.4545
1.2	.5988	.6040	.5720	.5735	.5695
1.4	.7158	.6915	.6700	.6735	.6725
1.6	.8133	.7995	.7715	.7745	.7740
1.8	.8868	.8850	.8640	.8680	.8680
2.0	.9369	.9350	.9215	.9205	.9220
2.2	.9677	.9665	.9585	.9590	.9595
2.5	.9899	.9870	.9830	.9820	.9825

TABLE 10

Sample sizes m = n = 6 ; size α = .05

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.0931	.0890	.0880	.0845	.0845
0.4	.1590	.1590	.1490	.1535	.1545
0.6	.2495	.2535	.2425	.2460	.2450
0.8	.3617	.3670	.3415	.3555	.3540
1.0	.4874	.4890	.4660	.4715	.4715
1.2	.6143	.6165	.5860	.5890	.5855
1.4	.7297	.7125	.6820	.6830	.6825
1.6	.8245	.8130	.7815	.7870	.7815
1.8	.8949	.8930	.8705	.8745	.8745
2.0	.9422	.9380	.9255	.9275	.9265
2.2	.9708	.9685	.9605	.9620	.9620
2.5	.9911	.9880	.9835	.9830	.9830

TABLE 11

Sample sizes m = n = 6; size α = .089827

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.2	.1559	.1445	.1535	.1490	.1475
0.4	.2480	.2565	.2460	.2460	.2445
0.6	.3630	.3785	.3475	.3535	.3535
0.8	.4912	.4955	.4765	.4785	.4780
1.0	.6220	.6310	.6065	.6110	.6120
1.2	.7391	.7470	.7240	.7245	.7245
1.4	.8339	.8185	.7945	.7960	.7950
1.6	.9030	.9005	.8775	.8785	.8810
1.8	.9482	.9455	.9335	.9325	.9320
2.0	.9747	.9755	.9675	.9690	.9680
2.2	.9887	.9855	.9805	.9820	.9820
2.5	.997	.9960	.9950	.9950	.9950

TABLE 12

Sample sizes m = n = 10; size α = .004465

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.0358	.0390	.0390	.0385	.0385
0.6	.0813	.0795	.0790	.0785	.0785
0.8	.1607	.1510	.1475	.1475	.1465
1.0	.2783	.2685	.2585	.2595	.2575
1.2	.4268	.4320	.4050	.4115	.4120
1.4	.5867	.5815	.5530	.5585	.5580

TABLE 13

Sample sizes m = n = 10; size α = .005

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.0392	.0420	.0405	.0430	.0405
0.6	.0879	.0865	.0850	.0855	.0855
0.8	.1716	.1630	.1590	.1580	.1585
1.0	.2936	.2795	.2725	.2710	.2720
1.2	.4450	.4495	.4210	.4295	.4300
1.4	.6051	.5995	.5710	.5805	.5815

TABLE 14

Sample sizes m = n = 10; size α = .09272

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.0634	.0640	.0695	.0710	.0695
0.6	.1326	.1320	.1295	.1295	.1285
0.8	.2413	.2395	.2230	.2270	.2305
1.0	.3858	.3815	.3635	.3635	.3595
1.2	.5485	.5500	.5175	.5205	.5240
1.4	.7029	.7035	.6825	.6905	.6915

TABLE 15

Sample sizes m = n = 10; size α = .01

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.0672	.0715	.0725	.0730	.0725
0.6	.1393	.1430	.1385	.1380	.1380
0.8	.2513	.2490	.2305	.2400	.2410
1.0	.3983	.3975	.3780	.3745	.3730
1.2	.5617	.5630	.5300	.5365	.5380
1.4	.7147	.7155	.6960	.7005	.7005

TABLE 16

Sample sizes m = n = 10; size α = .025

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.1330	.1330	.1285	.1295	.1280
0.6	.2452	.2520	.2440	.2445	.2445
0.8	.3948	.3925	.3745	.3810	.3820
1.0	.5622	.5620	.5335	.5460	.5440
1.2	.7186	.7215	.6975	.7040	.7015
1.4	.8414	.8430	.8245	.8270	.8275

TABLE 17

Sample sizes m = n = 10; size α = .026213

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.1375	.1385	.1330	.1315	.1310
0.6	.2520	.2580	.2495	.2520	.2505
0.8	.4032	.4000	.3810	.3870	.3900
1.0	.5709	.5725	.5420	.5570	.5575
1.2	.7262	.7265	.7065	.7115	.7100
1.4	.8470	.8495	.8275	.8335	.8340

TABLE 18

Sample sizes m = n = 10; size α = .05

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.2165	.2185	.2100	.2200	.2180
0.6	.3616	.3600	.3480	.3485	.3500
0.8	.5305	.5315	.5130	.5175	.5135
1.0	.6937	.6920	.6650	.6690	.6680
1.2	.8253	.8315	.7995	.8070	.8095
1.4	.9138	.9145	.8930	.9030	.9040

TABLE 19

Sample sizes m = n = 10; size α = .052561

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.2240	.2285	.2210	.2250	.2265
0.6	.3713	.3710	.3565	.3605	.3600
0.8	.5409	.5430	.5230	.5225	.5205
1.0	.7031	.7000	.6765	.6780	.6805
1.2	.8322	.8405	.8105	.8170	.8170
1.4	.9181	.9195	.8980	.9085	.9095

TABLE 20

Sample sizes m = n = 10; size α = .095158

shift	power of the test of				
	Student(exact)	Student(exper)	Wilcoxon	v.d.Waerden	Terry
0.4	.3310	.3350	.3220	.3240	.3225
0.6	.4998	.4900	.4805	.4835	.4785
0.8	.6687	.6765	.6555	.6565	.6550
1.0	.8084	.8080	.7920	.7955	.7955
1.2	.9045	.9105	.8995	.9015	.9020
1.4	.9594	.9630	.9505	.9580	.9580

TABLE 21

Small tail-probabilities of the two-sided equal-tails test for equal powers of a pair of rank tests

Sample sizes	size α	shift d	Wilcoxon vs. v.d.Waerden	Wilcoxon vs. Terry	Terry vs. v.d.Waerden
m = n = 6	.01	2.2			.0266 [1]
	.025	1.8	.0536	.0604	
		2.0	.0270	.0759	
		2.5	.0309	.0490	
	.046537	0.8	.0433	.0125	
		1.2			.0963 [1]
	.05	.08	.0015	.0041	
		1.6			.0522 [1]
	.089827	0.2		.0652	
		1.6			.0625
m = n = 10	.004329	1.2		.0980	
	.005	0.4			.0625
		1.2	.0396	.0385	
		1.4	.0395	.0460	
	.009272	0.8		.0817	
		1.0			.0762
		1.2			.0923
		1.4	.0888	.0693	
	.01	0.8	.0248	.0170	
		1.2		.0764	
	.025	1.0	.0059	.0334	
	.026213	0.8		.0535	
		1.0	.0009	.0010	
		1.4	.0961		
	.05	0.4	.0060		
		0.8			.0574
		1.2	.0444	.0205	
		1.4	.0055	.0046	
	.052561	0.4		.0762	
		1.4	.0031	.0032	
	.095158	0.6			.0414
		1.4	.0059	.0107	

[1] Both tests have nevertheless exactly the same power; the differences are wholly due to randomization.

TABLE 22

Test of goodness of fit of the experimentally obtained powers of the Student t-test

Sample sizes	shift d	chi square	df	Tail-probability
m = n = 6	0.2	23.52	19	.22
	0.4	17.85	18	.47
	0.6	14.50	18	.70
	0.8	14.91	16	.53
	1.0	13.09	15	.60
	1.2	10.04	14	.76
	1.4	44.65	20	.0012
	1.6	20.47	19	.37
	1.8	15.62	18	.62
	2.0	9.86	16	.87
	2.2	11.01	11	.44
	2.5	11.21	11	.43
m = n = 10	0.4	41.27	32	.13
	0.6	34.13	37	.61
	0.8	20.95	37	.98
	1.0	29.50	36	.77
	1.2	23.04	32	.88
	1.4	12.81	28	.99

Figure 1 sample sizes m=n=6

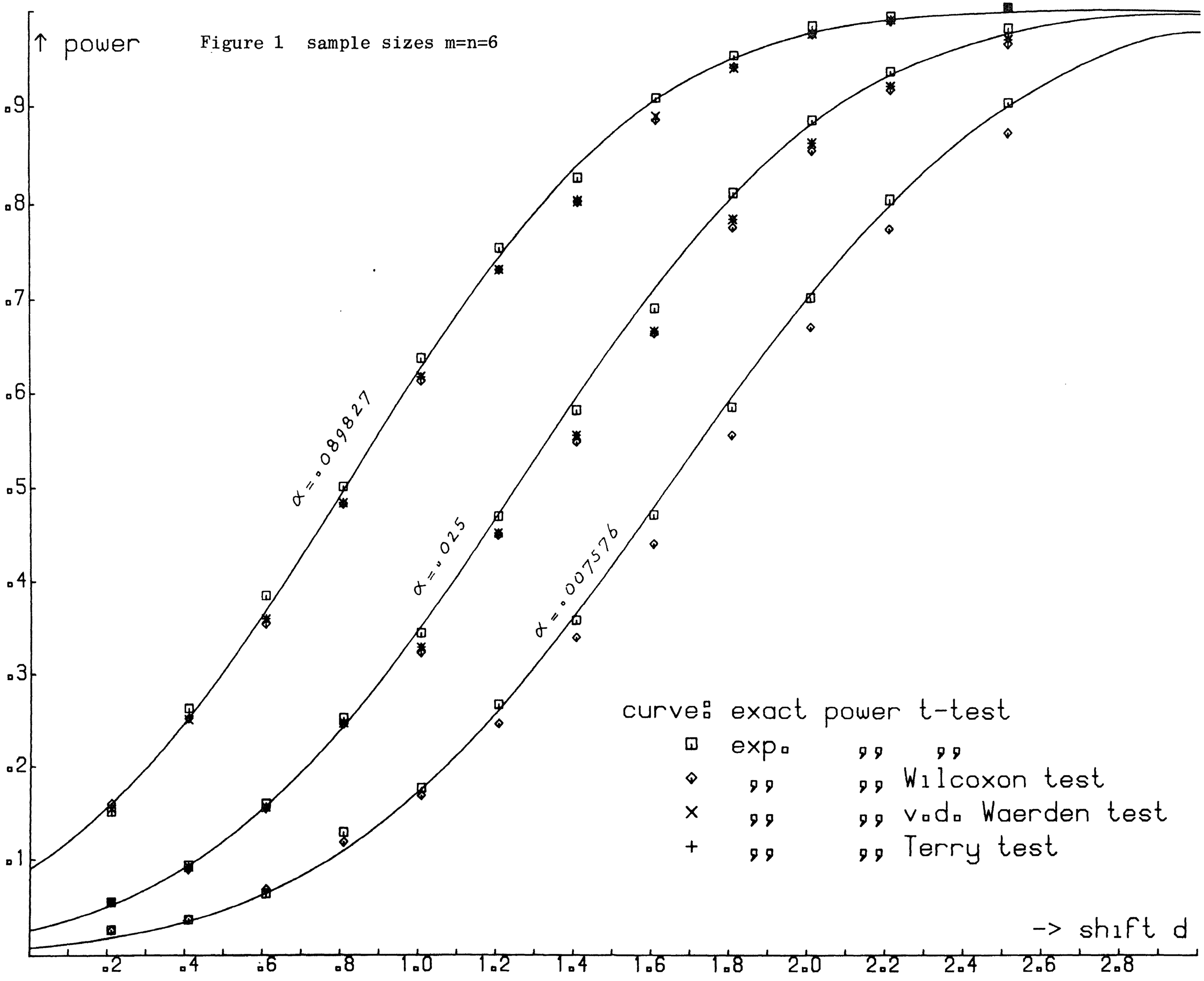

Figure 2 sample sizes m=n=6

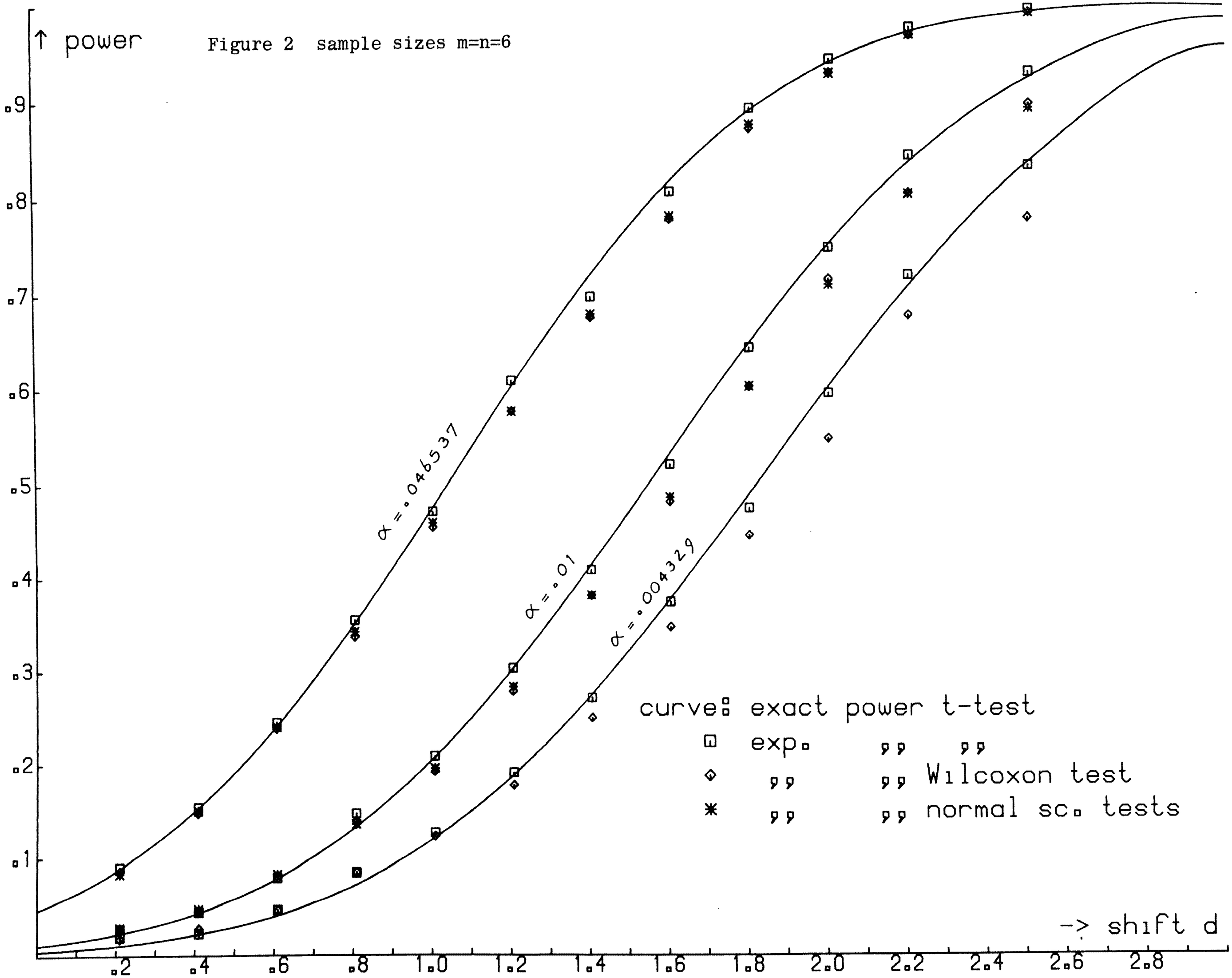

Figure 3 sample sizes m=n=6

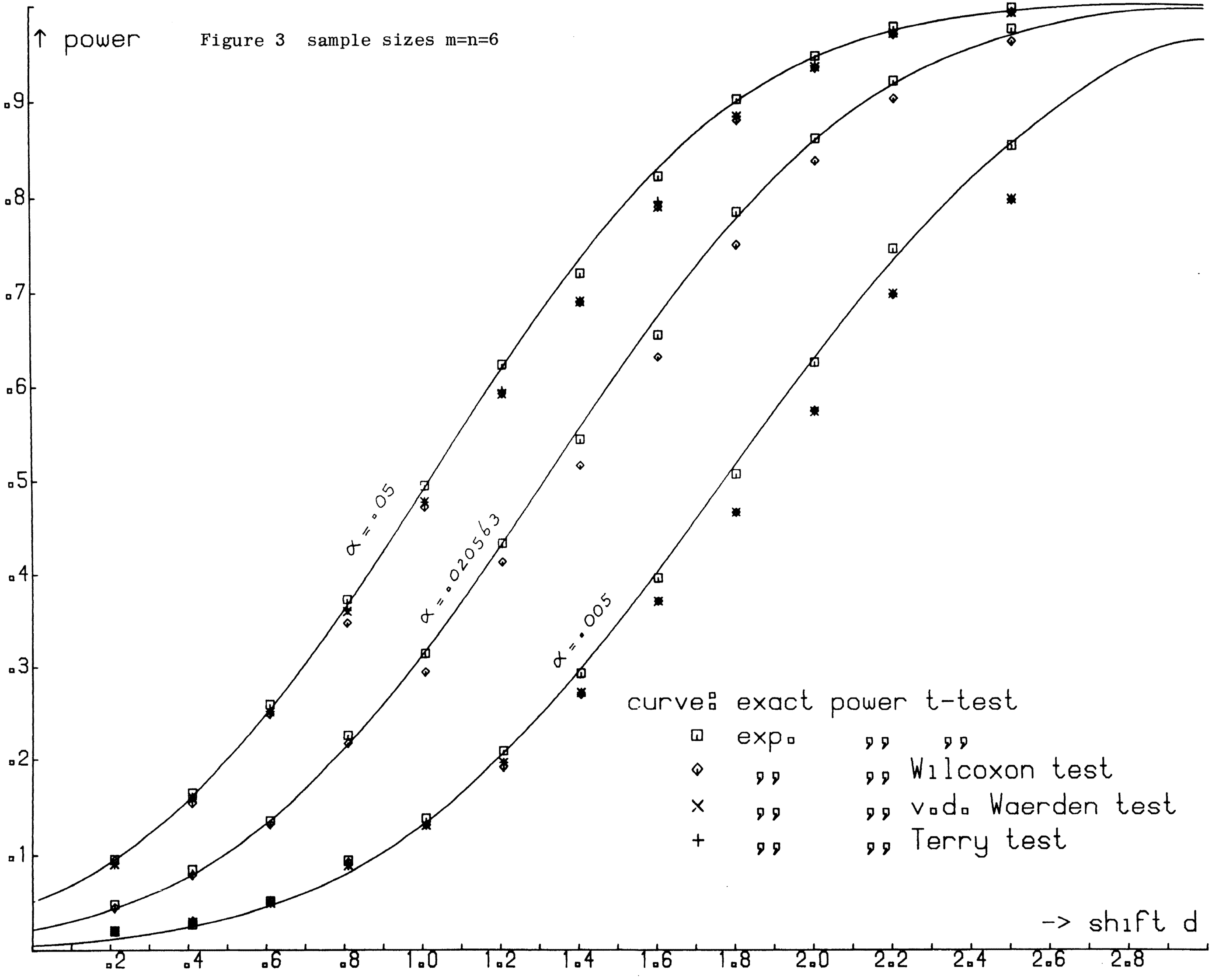

Figure 4 sample sizes m=n=10

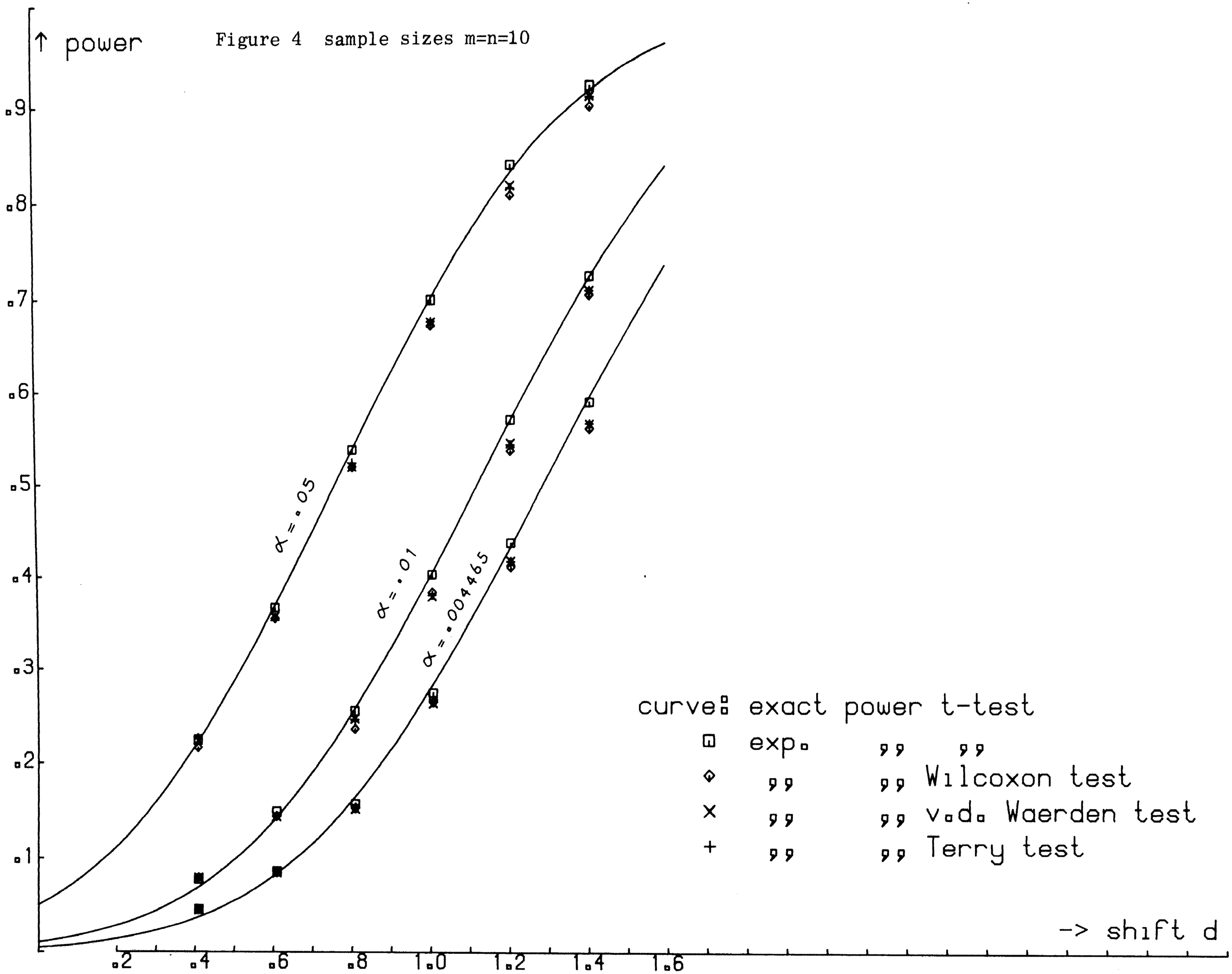

Figure 5 sample sizes m=n=10

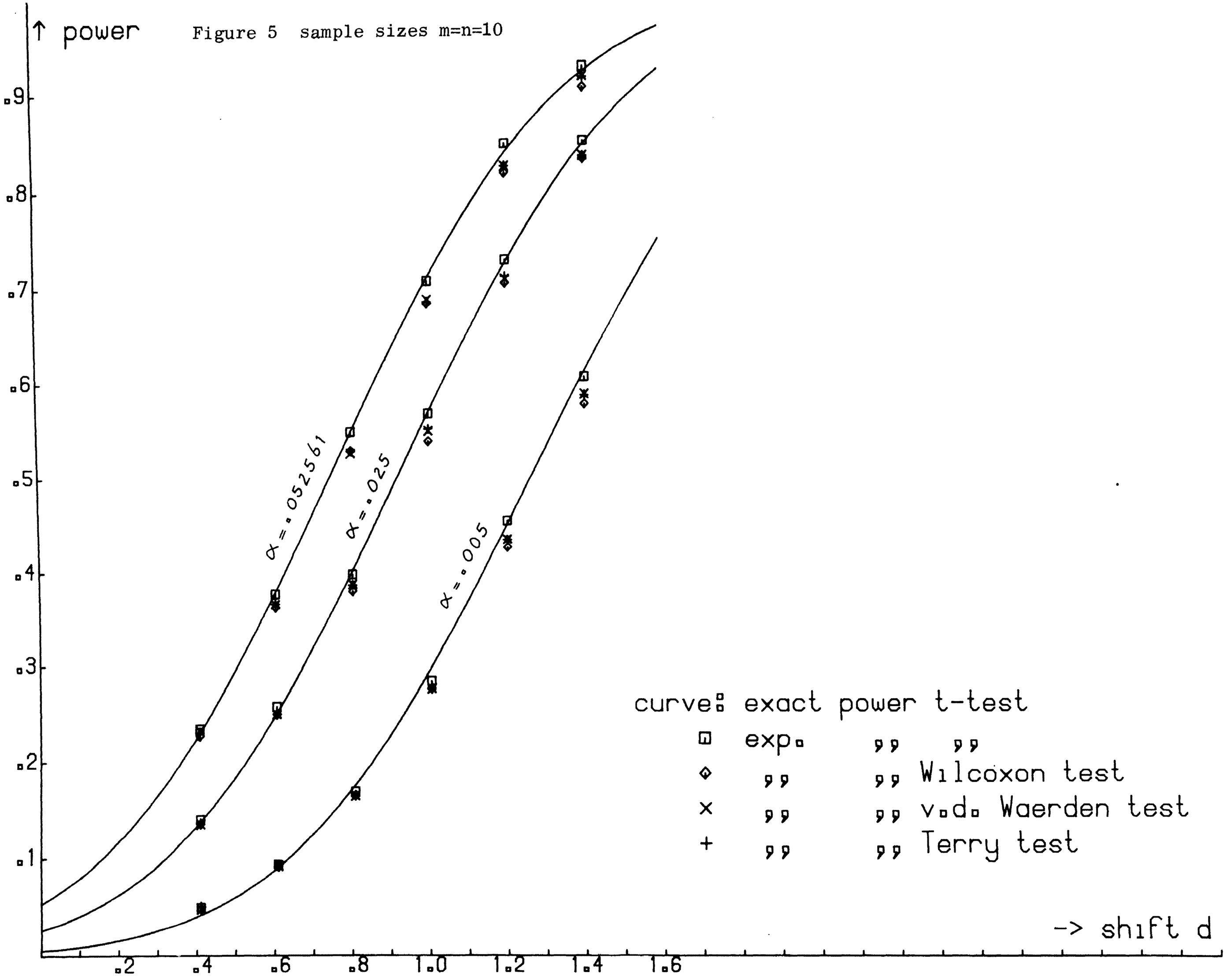

Figure 6 sample sizes m=n=10