Journal of Modern Applied Statistical Methods

Volume 1 | Issue 1 Article 23

5-1-2002

Asymptotic And Exact Tests In 2 x C Ordered Categorical Contingency Tables With *StatXact* 2.0 - 4.0

Margaret Posch
Wayne State University

Journal Of Modern Applied Statistical Methods
Winter 2002, Vol. 1, No. 1, 167-175

1538 - 9472/02/$30.00

Asymptotic And Exact Tests In 2 × C Ordered Categorical Contingency Tables With *StatXact* 2.0 - 4.0

Margaret Posch
Office of the Dean
College of Education
Wayne State University

The purpose of this study was to compare the statistical power of a variety of exact tests in the 2 × C ordered categorical contingency table using *StatXact* software. The Wilcoxon Rank Sum, Expected Normal Scores, Savage Scores (or its Log Rank equivalent), and Permutation tests were studied. Results indicated that the procedures were nearly the same in terms of comparative statistical power.

Keywords: Ordered-categorical contingency tables, 2 x C, Statistical power, Nonparametric tests

Introduction

The purpose of this study was to compare the statistical power of a variety of exact tests in the 2 × C ordered categorical contingency table using *StatXact* software. The Wilcoxon Rank Sum, Expected Normal Scores, Savage Scores (or its Log Rank equivalent), and Permutation tests were studied. Results indicated that the procedures were nearly the same in terms of comparative statistical power.

The development and wide-spread availability of personal computers with increased power in the early 1980s to the present have provided user-friendly statistical packages which make it possible for the applied researcher to easily carry out computation-intensive statistical procedures with high-speed and accuracy. Randomization and permutation tests, examples of computer-intensive procedures, yield exact p values instead of asymptotic p values. Mehta and Patel (1995) underscored the importance of using exact p values with an example, where the asymptotic Pearson Chi Squared (χ^2) test for row and column interaction in a 3 x 9 contingency table produced an observed test statistic of $\chi^2 = 22.29$. The p value associated with this obtained value for $\nu(16)$ is 0.1342. However, the p value associated with the exact distribution of χ^2 for the tail area to the right of 22.29 is 0.0013, which indicates there was a significant row and column interaction in the contingency table. This clearly demonstrates the power superiority of using the exact p value.

There are other reasons to support the preference of randomization and permutation tests over asymptotically-derived procedures in applied small samples research. Consider again the asymptotic χ^2 test. It is well known that for sparse contingency tables, "the usual chi-squared asymptotic distribution... is not likely to yield accurate p-values" (Mehta & Patel, 1995, p. 577). Moreover, the nature of small samples research lends itself to one or more (if not most) cells with expected values less than five. The statistical literature is replete with warnings about conducting the asymptotic χ^2 test under these conditions. Some authors protest if even a single cell has an expected frequency of less than five, though others permit up to twenty percent of the cells with low expectancies.

The analysis of continency data, which results in the commonly applied asymptotic χ^2 test, is frequent in behavioral and educational research. Of particular interest in this article is the ordinal categorical layout, which is comprised of two categorical groups with ordinal level outcomes. For example, in a study on the research experiences of doctoral students and publication rates after graduation, Troup-Leasure, Eichelberger, and Zigmond (1992) analyzed a 2x3 layout of apprenticeship (yes or no) and rate of publication per year (zero, less than one, one or more). Grissom (1994) examined a 2x3 layout of two types of marital therapies for which three levels of outcome (divorce, no change, improve) occurred. Frequently, in applied ordinal categorical studies such as these, the asymptotic χ^2 or asymptotic Wilcoxon test is calculated, or the results are analyzed solely with descriptive statistics such as percentages or correlations.

The Current Study

Given the propensity of the 2 x *c* ordinal categorical layout in applied research, the purpose of this study was to compare the statistical properties of computer-intensive nonparametric tests. The competing tests were the asymptotic χ^2, which was included for comparison purposes, and the computer-intensive exact versions of the following four tests: Wilcoxon Rank Sum (WRS), Expected Normal Scores (ENS), Savage Scores (SS) or its Log Rank

Margaret Posch, is Assistant Professor-Research, Office of the Dean, College of Education at Wayne State University. Her expertise is in research methods and statistical analysis in school transitional programs. She has served two terms as President of the Detroit Chapter of the American Statistical Association. Contact her at m.posch@wayne.edu.

equivalent, and Permutation (P).

Methodology

Considerable small samples research has been conducted on the properties of various statistics using pseudo-random number generators to model real data sets. However, Micceri (1989) pointed out the "need for careful data scrutiny prior to analysis" (p.161), and therefore, real data sets were obtained for the current study using the sampling plan discussed below.

Sampling Plan

Twenty-nine education and psychology journals were canvassed, primarily from the four year period of 1992 - 1996. A list of journals, the number of articles in which 2 x *c* data sets were found, and the number of data sets, are compiled in Table 1. The competing tests (asymptotic χ^2, and the exact WRS, ENS, SS, and P) were then applied to the data sets obtained from these studies.

The software used was *StatXact* 2.0-4.0 (Gajjar, Mehta, Patel, & Senchaudhuri, 1992) for exact nonparametric inference. (Note that version 5.0 has subsequently been released.) Results below permit a direct comparison of these statistics on real data commonly found in social and behavioral science research. The hypotheses being tested in all cases is the test for interaction. In some situations, this is a test of independence with regard to the outcome variable. In other cases, the data sets are simply cross tabulations of demographic variables.

Results

The survey yielded 149 ordinal categorical data sets of a 2 x *c* design collected from 73 articles published in refereed journals. Four distinct formats were encountered in the data: 1) the simple 2 x *c* design; 2) a 4 x *c* design, which consisted of a table with two distinct sets of two categories. For example, Table 2 shows four categories of psychological problems according to grade levels (used ordinally). For this data set, two 2 x *c* tables were arbitrarily constructed to include Emotional and Conduct as the categories of the first table and ADD and LD as the categories in the second 2 x *c*. 3) Nested designs were handled

Table 1. Education and Psychology Journals from 1992 - 1996 Canvassed.

Psychology (# of Articles, Data Sets)	Education (# of Articles, Data Sets)
American Journal of Community Psychology (1, 1)	*American Educational Research Journal* (2, 8)
American Journal of Family Therapy (0, 0)	*Education* (1, 2)
American Journal of Psychology (0, 0)	*Educational Researcher* (1, 1)
Basic and Applied Social Psychology (0, 0)	*Harvard Educational Review* (1, 1)
Developmental Psychology (0, 0)	*Journal for the Education of the Gifted* (1, 2)
Educational Psychologist (0, 0)	*Journal for Research in Mathematics Education* (3, 3)
Journal of Applied Psychology (1, 4)	*Journal of Negro Education* (1, 8)
Journal of Clinical Child Psychology (1, 2)	*Journal of Special Education* (1, 1)
Journal of Consulting and Clinical Psychology (1, 5)	*Journal of Teacher Education* (0,0)
Journal of Psychology: Interdisciplinary and Applied (3, 4)	*Teacher Education and Special Education* (0, 0)
Perceptual and Motor Skills (14, 19)	
Psychological Bulletin (0, 0)	
Psychological Reports (14, 29)	
Psychological Science (2, 5)	
Psychology and Aging (4, 7)	
Psychology in the Schools (2, 2)	
Professional Psychology: Research and Practice (2, 4)	
Reading Psychology (2, 2)	
School Psychology International (1,5)	

as multiple data sets. For example, Cherian (1992) reported parental education on an ordinal scale of low, middle and high for boys and girls from families with varying life status of parent(s) (Table 3). This example yielded two other data sets, by gender and by life status, both of which engaged the same ordinal levels of parent education, or *c*. 4) Other nested designs encouraged the summing of subdivisions, as noted in Table 4. The table portrayed several subcategories of men and women who were placed in ordinal levels of young, middle-aged or senior age groups. The subdivisions of each gender category were summed to develop a single data set.

Sample sizes for the selected studies ranged from ten to 19,256, and contained from three to eight ordinal levels (*c*) of outcome data. Ninety-seven (65.1%) of the studies had three ordinal levels, 32 (21.5%) had 4 levels, 15 (10.1%) had 5 levels, one (.7%) had six levels, one (.7%) had seven levels, and three (2%) studies had eight

Table 2. Teacher's Perceptions Of Psychological Problems In Students (Morrow, 1995).

Category	K-3	4-6	7-8	9-12
Emotional	187	110	23	56
Conduct	190	116	76	55
ADD	190	118	23	54
LD	184	113	23	54

Table 3. Frequencies For Low-, Middle-, And High-Scoring Subjects On Parental Education From Families With One Or Both Parents Dead And With Both Parents Alive (Cherian, 1992).

One/Both Parents Dead	Low	Middle	High
Boys	9	89	13
Girls	39	147	11
Both Parents Alive			
Boys	21	121	55
Girls	58	279	178

Table 4. Number Of Note Pairs Matched By Age Group, Sex, And Occupational Level (Black, 1993).

Group	Young (18-29)	Middle-aged (30-64)	Senior (65+)
Men			
Blue Collar	4	10	3
White Collar	3	18	4
Professional	0	7	3
Student	3	0	0
Homemaker	0	0	0
Women			
Blue Collar	0	0	0
White Collar	3	3	1
Professional	0	2	1
Student	2	0	0
Homemaker	1	5	4

ordinal levels.

The p values compiled in Table 5 for each nonparametric exact test, in addition to the asymptotic p values for the χ^2 test, include an asterisk if the value was significant at $\alpha = 0.0500$. *StatXact Turbo* software carried out the decimal to four places; thus, p values of 0.05 with values higher than zero in the third or fourth decimal place were not reported as significant.

Error messages sometimes occurred with larger data sets: "The problem (was) too large for the test option". When this occurred, the test was rerun with the "Monte Carlo (MO) option". P values for the Monte Carlo option result from employing 20,000 iterations (99% confidence level), and are indicated in parentheses in Table 5. *StatXact* reported Monte Carlo results either as a confidence interval, or, for example as with data set 41, as 0.0053 ± 0.0013. (Note that these lower and upper bounds of the interval are not included in Table 5.)

Table 5. P Values For Tests Computed With *StatXact*.

ID	WRS	SS	ENS	P	χ^2	ID	WRS	SS	ENS	P	χ^2
001	.4225	.4901	.4572	.4415	.9336	039	.0000*	.0000*	.0000*	.0000*	.0000*
002	.3467	.2980	.3167	.3467	.8424	040	.5000	.4675	.4838	.6169	.6005
003	.4073	.4526	.4073	.4397	.9208	041	(.0053)*	(.0091)*	(.0067)*	.0056*	.1342
004	.5000	.4675	.4838	.6169	.6005	042	(.0163)*	(.0231)*	(.0171)*	.0161*	.2012
005	.0261*	.0613	.0260*	.0392*	.1146	043	.1801	.4194	.1775	.3230	.0807
006	.2812	.3907	.3158	.2511	.0056*	044	.0263*	.0235*	.0263*	.0330*	.1311
007	.0122*	.0122*	.0122*	.0203*	.0207*	045	.1903	.0985	.2053	.2570	.0941
008	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0000*	046	.0117*	.0014*	.0328*	.0574	.0002*
009	.0042*	.0012*	.0026*	.0026*	.0044*	047	.0000*	.0000*	.0000*	.0000*	.0000*
010	.4658	.4768	.4844	.5071	.9747	048	.0000*	.0000*	.0000*	.0000*	.0000*
011	.4570	.4390	.4522	.4535	.9975	049	.0289*	.0289*	.0289*	.0294*	.1025
012	.5831	.2441	.5831	.5831	.0123*	050	(0.0)*	(0.0)*	(0.0)*	.0000*	.0002*
013	.0024*	.0015*	.0024*	.0024*	.0085*	051	.4254	(.1731)	.4773	.4683	.0161*
014	.0201*	.0805	.0201*	.0302*	.0360*	052	.0029*	.0290*	(.0041)*	.0040*	.0238*
015	.0001*	.0001*	.0000*	.0001*	.0002*	053	.0000*	(.0001)*	.0000*	.0000*	.0000*
016	.0000*	.0001*	.0000*	.0000*	.0001*	054	.0014*	(.0001)*	.0020*	.0038*	.0023*
017	.0016*	.0018*	.0019*	.0021*	.0130*	055	.0938	.0812	.0914	.1210	.3271
018	.0160*	.0511	.0195*	.0216*	.0953	056	.0000*	.0000*	.0000*	.0000*	.0001*
019	.0004*	.0014*	.0004*	.0004*	.0017*	057	.5000	.4619	.5000	.5568	.9004
020	.2755	.4146	.3834	.4857	.0756	058	.2007	.1464	.2007	.2052	.4380
021	.2968	.2184	.2662	.2902	.6945	059	.0000*	.0000*	.0000*	.0000*	.0000*
022	.1475	.2185	.1667	.1864	.5167	060	.1580	.2309	.1580	.1584	.1049
023	.4488	.2234	.4909	.4916	.1238	061	.3574	.4266	.3510	.3510	.2900
024	.0001*	.0012*	.0002*	.0005*	.0003*	062	.2486	.1256	.2070	.2175	.5384
025	.0000*	.0000*	.0000*	.0000*	.0000*	063	.0400*	.0257*	.0559	.0492*	.1791
026	.0135*	.3396	.0501	.0825	.0000*	064	.3997	.4201	.3997	.4228	.9574
027	.0000*	.0005*	.0000*	.0000*	.0000*	065	(.4912)	(.4985)	(.4918)	.5017	.9999
028	.0324*	.0178*	.0324*	.0330*	.0889	066	.4592	.4771	.4585	.4648	.9228
029	.3467	.4884	.3467	.3738	.3865	067	(0.0)*	(0.0)*	(0.0)*	.0000*	.0000*
030	.0525	.0332*	.0525	.0535	.1581	068	.0000*	.0000*	.0000*	.0000*	.0363*
031	.4414	.4803	.4414	.4649	.9315	069	.3836	.1582	.4673	.4675	.0363*
032	.0000*	.0000*	.0000*	.0000*	.0000*	070	.3836	.1582	.4673	.4675	.0363*
033	.4959	.4901	.4971	.5169	.9996	071	.3836	.1582	.4673	.4675	.0363*
034	.5769	.3288	.4462	.5769	.4117	072	.3836	.1582	.4673	.4675	.0363*
035	.2464	.2715	.2587	.2791	.8731	073	.1562	.1876	.1597	.1703	.7475
036	.3798	.3543	.3744	.4082	.9813	074	.1575	.1779	.1514	.1718	.7512
037	.4361	.4448	.4407	.4727	.9968	075	.0810	.0726	.0806	.0858	.3301

Table 5. *P* Values For Tests Computed With StatXact.

ID	WRS	SS	ENS	P	χ^2	ID	WRS	SS	ENS	P	χ^2
077	.0269*	.0333*	.0276*	.0332*	.1470	114	.1272	.0892	.1584	.1626	.3264
078	.0913	.1182	.0913	.1291	.2751	115	.2536	.2814	.2751	.2982	.7118
079	.3556	.3333	.3556	.3556	.5988	116	.5257	.5041	.5046	.5257	1.0000
080	.0007*	.0043*	.0010*	.0011*	.0070*	117	.0000*	.0000*	.0000*	.0000*	.0000*
081	.0000*	.0000*	.0000*	.0000*	.0000*	118	.0051*	.0047*	.0051*	.0059*	.0215*
082	.0454*	.0265*	.0570	.0715	.0828	119	.0023*	.0123*	.0030*	.0077*	.0039*
083	.4433	.4234	.4569	.4904	.9181	120	.0000*	.0000*	.0000*	.0000*	.0000*
084	.0219*	.0037*	.0265*	.0431*	.0014*	121	.0034*	.0034*	.0034*	.0034*	.0006*
085	.0475*	.0622	.0562	.0583	.3469	122	.0010*	.0013*	.0010*	.0010*	.0051*
086	.0000*	.0000*	.0000*	.0000*	.0000*	123	.0526	.0868	.0518	.0530	.1404
087	.0748	.1062	.0748	.0777	.2591	124	.4368	.4680	.4369	.4680	.9644
088	.0136*	.0152*	.0136*	.0164*	.0824	125	.0000*	.0000*	.0000*	.0000*	.0000*
089	.2759	.1823	.2759	.2799	.4065	126	.5581	.4730	.5581	.5581	.8956
090	.0136*	.0152*	.0136*	.0164*	.0824	127	.2151	.2697	.2151	.2195	.3871
091	.5569	.5223	.5223	.5569	1.0000	128	.0000*	.0000*	.0000*	.0000*	.0000*
092	.5569	.5223	.5223	.5569	1.0000	129	.0269*	.0138*	.0251*	.0260*	.0779
093	.1829	.1939	.1829	.2010	.6344	130	.4672	.2885	.4697	.4695	.5258
094	.3696	.3762	.3696	.3911	.9396	131	.3995	.4954	.3941	.4255	.5917
095	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0000*	132	.3685	.3681	.3667	.4036	.9015
096	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0002*	133	.0742	.0131*	.0845	.0690	.0308*
097	(0.0)*	(0.0)*	(0.0)*	.0000*	.0000*	134	.0000*	.0000*	.0000*	.0000*	.0000*
098	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0001*	135	.3160	.3571	.3160	.3430	.8505
099	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0000*	136	.1665	.2038	.1419	.1419	.0000*
100	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0001*	137	.4699	.4397	.4706	.4949	.9.810
101	.0000*	(0.0)*	.0000*	.0000*	.0000*	138	.2250	.3601	.2786	.2913	.4336
102	(0.0)*	(0.0)*	(0.0)*	(0.0)*	.0000*	139	.2150	.0809	.1794	.1794	.2050
103	.4691	.4926	.4691	.4838	.9073	140	.3748	.4329	.3748	.4286	.9040
104	.3824	.2651	.3417	.3167	.7407	141	.3605	.2346	.3444	.3527	.5897
105	.4852	.4943	.4878	.4949	.9999	142	.0383*	.0136*	.0208*	.0382*	.0689
106	.2804	.2551	.2754	.2776	.7979	143	.4740	.3986	.4302	.5000	.8977
107	.2868	.2580	.2812	.2832	.7988	144	.0767	.0367*	.0534	.0785	.1733
108	.3600	.1486	.3265	.3159	.0681	145	.0000*	.0000*	.0000*	.0000*	.0002*
109	.2849	.1507	.2637	.2577	.2359	146	.0000*	.0000*	.0000*	.0000*	.0000*
110	.2980	.1602	.2760	.2705	.2455	147	.3776	.4209	.3776	.4160	.9146
111	.3280	.3801	.3326	.3564	.8453	148	.0016*	.0061*	.0016*	.0028*	.0092*
112	.3993	.4312	.3993	.4459	.9365	149	.0257*	.0302*	.0257*	.0275*	.0178*
113	.3565	.3815	.3565	.3712	.7920						

The prevailing question is whether one exact test is more powerful than another. Table 6 carries the question of power one step further, displaying the frequency and the percent of significant results for all levels of ordinal outcomes. The results indicates comparable results for all procedures.

Conclusion

Clearly, the χ^2 should be avoided in analyzing ordered categorical contingency tables, for both theoretical reasons and in terms of power to detect row and column interactions. On the basis of this study, the exact versions of the Wilcoxon, Expected Normals, Savage, and Permutation tests appear to be comparable in terms of statistical power.

Table 6. Frequency (%) of Significant Results of Exact Tests for Each Level of c, $\alpha = 0.0500$.

	Number Of Ordinal Levels, (N Of Data Sets)					
Test	3 (N=97)	4 (N=32)	5 (N=15)	6 (N=1)	7 (N=1)	8 (N=3)
WRS	41 (.25)	8 (.29)	7 (.27)	0 (0)	0 (0)	0 (0)
SS	40 (.25)	7 (.27)	5 (.19)	0 (0)	0 (0)	0 (0)
ENS	39 (.25)	6 (.22)	6 (.23)	0 (0)	0 (0)	0 (0)
P	39 (.25)	6 (.22)	8 (.31)	1 (1.00)	0 (0)	0 (0)

References

Basu, A. P. (1968). On a generalized savage statistic with applications to life testing. *Annals of Mathematical Statistics, 39,* 1591-1604.

Bhattacharyya, G. K., & Johnson, R. A. (1977). *Statistical concepts and methods.* NY: Wiley.

Black, S. T. (1993). Comparing genuine and simulated suicide notes: A new perspective. *Journal of Consulting and Clinical Psychology, 61,* 699-702.

Blair, R. C. (1980). A comparison of the power of the two independent means t test to that of the Wilcoxon's rank-sum test for samples of various populations. Unpublished doctoral dissertation, University of South Florida, Tampa, FL.

Bradley, J. V. (1968). *Distribution-free statistical tests.* Englewood Cliffs, NJ: Prentice-Hall.

Chernoff, H., & Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist, 29,* 972-994.

Cherian, V. I. (1992). Relation of parental education and life status to academic achievement by Xhosa children. *Psychological Reports,* 70, 947-956.

Fisher, R. A., & Yates, F. (1938). *Statistical tables for biological, agricultural and medical research.* Edinburgh: Oliver and Boyd.

Gajjar, Y., Mehta, C. R., Patel, N. & Senchaudhuri, P. (1992). *StatXact-Turbo statistical software for exact nonparametric inference user manual.* MA: Cytel Software.

Grissom, R. J. (1994). Statistical analysis of ordinal categorical status after therapies. *Journal of Consulting and Clinical Psychology,* 62, 281-284.

Haber, A. (1990). Comments on The test of homogeneity for 2x2 contingency tables: A review of and some personal opinions on the controversy by G. Camilli. *Psychological Bulletin,* 108, 146-149.

Hajek, J. (1969). *A course in nonparametric statistics.* San Francisco, CA: Holden-Day.

Harter, H. L. (1969). *Order statistics and their use in testing and estimation. Vol. 2.* U. S. Government Printing Office, Washington, DC.

Hodges, J. L., Jr., & Lehmann, E. L. (1960). Comparison of the Normal Scores and Wilcoxon tests. Proceedings of the *Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1,* 307-317. Berkeley: University of California Press.

Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.) New York: Holt, Rinehart, & Winston.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics, 18,* 50-60.

Marascuilo, L. & Serlin, R. (1988). *Statistical methods for the social and behavioral sciences.* NY: W. H. Freeman.

May, R. B., & Hunter, M. A. (1990). Some advantages of permutation tests. In B. Zumbo (Chair), *Alternatives to classical statistical procedures.* Symposium conducted at the annual meeting of the Canadian Psychological Association, Ottawa. (Abstract in *Canadian Psychology,* 31, 238).

Mehta, C. R. & Patel, N. (1992). *StatXact Turbo 2.0: Statistical software for exact nonparametric inference user manual.* MA: Cytel Software.

Mehta, C. R. & Patel, N. (1995). *StatXact 3 for Windows: Statistical software for exact nonparametric inference user manual.* MA: Cytel Software.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Morrow, V. L. (1995). Teachers evaluate psychological problems and personal counseling needs of students. *Education, 116,* 130-136.

Norween, E. W. (1989). *Computer intensive methods for testing hypotheses: An introduction.* NY: Wiley.

Owen, D.B. (1962). *Handbook of statistical tables*. Reading, MA: Addison-Wesley.

Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric tests*. NY: John Wiley.

Savage, R. (1956). Contributions to the theory of rank order statistics: The two-sample case. *Annals of Mathematical Statistics, 27,* 590-615.

Solorzano, D. G. (1992). An exploratory analysis of the effects of race, class, and gender on student and parent mobility aspirations. *Journal of Negro Education, 61*, 30-44.

Troup-Leasure, K. , Eichelberger, R. T., & Zigmond, N. (1992). Research experiences of doctoral students and publication after graduation. *Teacher Education and Special Education, 15*, 183-193.

Van der Waerden, B. L. (1952). Order Tests for the two-sample problem and their power. *Indagationes Math, 14*, 453-458.

Van der Waerden, B. L. (1953). Errata. 80.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1*, 80-82.

Appendix

Studies Where Data Sets Were Located (Ordered According To ID in Table 5)

[1]Babad, E. (1995). Can accurate knowledge reduce wishful thinking in voters' predictions of election outcomes? *The Journal of Psychology, 129*, 285-300.

[2,3] Weiserbs, B., & Gottlieb, J. (1995). The perception of risk over time as a factor influencing attitudes toward children with physical disabilities. *The Journal of Psychology, 129*, 689-699.

[4]Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62*, 543-549.

[5]Speer, D. C. (1994). Can treatment research inform decision makers? Nonexperimental method issues and examples among older outpatients. *Journal of Consulting and Clinical Psychology*, 62, 560-568.

[6]Dancer, L. S., Anderson, A. J., & Derlin, R. L. (1994). Use of log-linear models for assessing differential item functioning in a measure of psychological functioning. *Journal of Consulting and Clinical Psychology, 62*, 710-717.

[7]Gauthier, J., Côté, G., & French, D. (1994). The role of home practice in the thermal biofeedback treatment of migraine headache. *Journal of Consulting and Clinical Psychology, 62*, 180-184.

[8]Storfer, M. D. (1995). Problems in left-right discrimination in a high-IQ population. *Perceptual and Motor Skills*, 81, 491-497.

[9]Zaslavsky, O. & Peled, I. (1996). Inhibiting factors in generating examples by mathematics teachers and student teachers: The case of binary operation. *Journal for Research in Mathematics Education, 27*, 67-78.

[10,11]Morrow, V. L. (1995). Teachers evaluate psychological problems and personal counseling needs of students. *Education, 116*, 130-136.

[12,13]Halvari, H. & Gjesme, T. (1995). Trait and state anxiety before and after competitive performance. *Perceptual and Motor Skills, 81*, 1059-1074.

[14]Forns-Santacana, M. & Martorell-Balanzó, B. (1996). Relationships of personality factors with clinical dimensions and school achievement. *Perceptual and Motor Skills, 82*, 243-253.

[15]Toro, P. A. & McDonnell, D. M. (1992). Beliefs, attitudes, and knowledge about homelessness: A survey of the general public. *American Journal of Community Psychology, 20*, 53-78.

[16]Cherian, V. I. (1994). Self-reports of corporal punishment of Xhosa children from broken and intact families and their academic achievement. *Psychological Reports, 74*, 867-874.

[17]Cherian, V. I. (1994). Relationship between parental aspiration and academic achievement of Xhosa children from broken and intact families. *Psychological Reports, 74*, 835-840.

[18]Hamm, N. H. (1994). Outcomes of the Minnesota smoking prevention program. *Psychological Reports, 75*, 880-882.

[19]Ines, T. M. & Sacco, W. P. (1992). Factors related to correspondence between teacher ratings of elementary student depression and student self-ratings. *Journal of Consulting and Clinical Psychology, 60*, 140-142.

[20-31]Norris, F. H. (1992). Epidemiology of trauma: Frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60*, 409-418.

[32]Eysenck, H. J. (1992). The effects of psychotherapy: An evaluation. *Journal of Consulting and Clinical Psychology, 60*, 659-663.

[33]Sanders, M. R., Shepherd, R. W., Cleghorn, G., & Woolford, H. (1994). The treatment of recurrent abdominal pain in children: A controlled comparison of cognitive-behavioral family intervention and standard pediatric care. *Journal of Consulting and Clinical Psychology, 62*, 306-314.

[34]Black, St. T. (1993). Comparing genuine and simulated suicide notes: A new perspective. *Journal of Consulting and Clinical Psychology, 61*, 699-702.

[35-37]Jacobson, N. S., Fruzzetti, A. E., Dobson, K., Whisman, M., & Hops, H. (1993). Couple therapy as a treatment for depression: II. The effects of relationship quality and therapy on depressive relapse. *Journal of Consulting and Clinical Psychology 61*, 516-519.

[38,39]Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

[40]Uehara, E. S., Smukler, M., & Newman, F. L. (1994). Linking resource use to consumer level of need: Field test of the Level of Need-care Assessment (LONCA) method. *Journal of Consulting and Clinical Psychology, 62*, 695-709.

[41-43]Kingma, J. (1994). Age and gender distributions of pedestrian accidents across the life-span. *Perceptual and Motor Skills*, 79, 1680-1682.

[44,45]Finkel, D., McGue, M. & Pedersen, N. (1995). Genetic influences on memory performance in adulthood: Comparison of Minnesota and Swedish twin data. *Psychology and Aging, 10*, 437-446.

[46,47]Bell, P. F., McKenna, J. P., & Digman, Jr., R. H. (1995). Should psychologists obtain prescribing privileges? A survey of family physicians. *Professional Psychology: Research and Practice, 26*, 371-376.

[48]Simonson, S. D. (1995). A historical view of content area reading instruction. *Reading Psychology: An International Quarterly, 16*, 99-147.

[49]Kragler, S. (1995). The transition from oral to silent reading. *Reading Psychology: An International Quarterly, 16*, 395-408.

[50-54]Zimba, R. F. (1995). Secondary school students' risks that may promote HIV infection and the spread of AIDS. *School Psychology International, 16*, 67-78.

[55-60]Fukunishi, I. (1995). Psychological acceptance and alexithymia in spinal cord injury patients. *Psychological Reports, 76*, 475-481.

[61]Isoaho, R., Keistinen, T., Laippala, P. & Kivelä, S. (1995). Chronic obstructive pulmonary disease and symptoms related to depression in elderly persons. *Psychological Reports, 76*, 287-297.

[62,63]Anstey, K., Stankov, L., & Lord, S. (1993). Primary aging, secondary aging, and intelligence. *Psychology and Aging, 8*, 562-570.

[64]Schulz, R., Musa, D., Staszewski, J., & Siegler, R. (1994). The relationship between age and major league baseball performance: Implications for development. *Psychology and Aging, 9*, 274-286.

[65,66]Spiro, III, A., Schnurr, P. P., & Aldwin, C. M. (1994). Combat-related posttraumatic stress disorder symptoms in Older Men. *Psychology and Aging, 9*, 17-26.

[67]Goldberg, C. J. & Botvin, G. J. (1993). Assertiveness in Hispanic adolescents: Relationship to alcohol use and abuse. *Psychological Reports, 73*, 227-238.

[68]Trankina, M. L. (1993). Gender differences in attitudes toward science. *Psychological Reports, 73*, 123-130.

[69-72]Cherian, V. I. (1993). Gender, socioeconomic status, and mathematics achievement by Xhosa children. *Psychological Reports, 73*, 771-778.

[73,74]Erdman, R. A. M., Kooijman, M. Passchier, J., & Stronks, D. L. (1993). The Dutch version of the Nottingham Health Profile: Investigations of psychometric aspects. *Psychological Reports, 72*, 1027-1035.

[75-77]Nagy, S. & Nagy, M. C. (1992). Longitudinal examination of teachers' burnout in a school district. *Psychological Reports, 71*, 523-531.

[78,79]Beer, J. & Beer, J. (1992). Depression, self-esteem, suicide ideation, and GPAs of high school students at risk. *Psychological Reports, 71*, 899-902.

[80,81]Cherian, V. I. (1992). Relation of parental education and life status to academic achievement by Xhosa children. *Psychological Reports, 71*, 947-956.

[82-84]Matthew, M. D., Franz, R. S., & Weaver, C. N. (1992). Perceived financial status of male and female college professors. *Psychological Reports, 70*, 199-202.

[85]Kalliopuka, M. (1992). Social desirability related to children's age, sex, and willingness to help. *Psychological Reports, 70*, 479-482.

[86]May, C. P., Hasher, L. & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science, 4*, 326-330.

[87-90]Costa, M. M. & Gatz, M. (1992). Determination of authorship credit in published dissertations. *Psychological Science, 3*, 354 - 357.

[91-94]Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology, 79*, 993-997.

[95-102]Solorzano, D. G. (1992). An exploratory analysis of the effects of race, class, and gender on student and parent mobility aspirations. *Journal of Negro Education, 61*, 30-44.

[103]Kowner, R. & Ogawa, T. (1993). The contrast effect of physical attractiveness in Japan. *The Journal of Psychology, 127*, 51-64.

[104,105]Adelman, H. S., Barker, L. A., & Nelson, P. (1993). A study of a school-based clinic: Who uses it and who doesn't? *Journal of Clinical Child Psychology, 22*, 52-59.

[106-111]Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behavior: A multiyear, multischool study. *American Educational Research Journal, 30*, 179-215.

[112,113]Newman, R. S. & Schwager, M. T. (1995). Students' help seeking during problem solving: Effects of grade, goal, and prior achievement. *American Educational Research Journal, 32*, 352-376.

[114]Powell, F. C. & Wanzenried, J. W. (1993). Perceptions of Bush, Clinton, and Perot in relation to frequency of presidential debate viewing. *Perceptual and Motor Skills, 77*, 35-41.

[115]Hicks, R. A. & Bautista, J. (1993). Snoring and nightmares. *Perceptual and Motor Skills, 77*, 433-434.

[116]Kunz, J. (1993). Ice cream preference: Gender differences in taste and quality. *Perceptual and Motor Skills, 77,* 1097-1098.

[117]Hicks, R. A., Conti, P. A., & Nellis, T. (1992). Arousability and stress-related physical symptoms: A validation study of Coren's Arousal Predisposition Scale. *Perceptual and Motor Skills, 74,* 659-662.

[118-122]Troup-Leasure, K. Eichelberger, R. T., & Zigmond, N. (1992). Research experiences of doctoral students and publication after graduation. *Teacher Education and Special Education, 15,* 183-193.

[123]Lidz, C. S. (1992). The extent of incorporation of dynamic assessment into cognitive assessment courses: A national survey of school psychology trainers. *The Journal of Special Education, 26,* 325-331.

[124]English, L. D. (1993). Children's strategies for solving two- and three-dimensional combinatorial problems. *Journal for Research in Mathematics Education, 24,* 255-273.

[125]Ferrini-Mundy, J. & Gaudard, M. (1992). Secondary school calculus: Preparation or pitfall in the study of college calculus? *Journal for Research in Mathematics Education, 23,* 56-71.

[126,127]Ford, D. Y. (1993). Support for the achievement ideology and determinants of underachievement as perceived by gifted, above-average, and average Black students. *Journal for the Education of the Gifted, 16,* 280-298.

[128]Orfield, G. (1992). Money, equity, and college access. *Harvard Educational Review, 52,* 337-372.

[129]Pickett, W. & Burrill, D. F. (1994). The use of quantitative evidence in research: A comparative study of two literatures. *Educational Researcher, 23,* 18-21.

[130,131]Wrobel, N. H. (1993). Effect of patient age and gender on clinical decisions. *Professional Psychology: Research and Practice, 24,* 206-212.

[132]Harris, J. (1992). Teacher-completed child behavior checklist ratings as a function of classroom-based interventions: A pilot study. *Psychology in the Schools, 29,* 42-52.

[133]Santos de Barona, M. (1992). Infant and preschool evaluation services: Implications for school districts and personnel. *Psychology in the Schools, 29,* 237-247.

[134]Carrillo-de-la-Peña & Luengo, M. A. (1994). Time estimation and juvenile delinquency. *Perceptual and Motor Skills, 79,* 1559-1565.

[135]Eddowes, E. A., Aldridge, J., & Culpepper, S. (1994). Primary teachers' classroom practices and their perceptions of children's attention problems. *Perceptual and Motor Skills, 79,* 787-790.

[136]Bortoli, L. & Robazza, C. (1994). The Motor Activity Anxiety test. *Perceptual and Motor Skills, 79,* 299-305.

[137,138]Gaulin, C. A. & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79,* 55-64.

[139-144]Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402-408.

[145,146]Tanaka, S. (1994). Effects of a moderately difficult task on preschool children's concentration and their subsequent choices of task. *Perceptual and Motor Skills, 78,* 699-700.

[147]Nkaya,H. N., Huteau, M., & Bonnet, J. (1994). Retest effect on cognitive performance on the Raven-38 Matrices in France and in the Congo. *Perceptual and Motor Skills, 78,* 503-510.

[148,149]Grissom, R. J. (1994). Statistical analysis of ordinal categorical status after therapies. *Journal of Consulting and Clinical Psychology, 62,* 281-284.