A MONTE CARLO STUDY OF RANK TESTS FOR REPEATED MEASURES DESIGNS

Michael D. Ernst
Graduate Student

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275-0332

James L. Kepner
Professor of Statistics and Mathematics

Department of Mathematics and Statistics
St. Cloud State University
St. Cloud, Minnesota 56301-4498

Key Words and Phrases: Koch's statistic; rank transformation statistic; Friedman's test; repeated measures; within block correlation.

ABSTRACT

Koch (1969) proposed a rank test to detect fixed treatment effects in complete repeated measures designs where the observations within blocks are assumed to be equally correlated. In deriving the limiting distribution of the rank transformation statistic for this design, Kepner and Robinson (1988) also obtained the limiting distribution of Koch's statistic under substantially less restrictive conditions than those required by Koch. The purpose of this paper is to summarize and disseminate the results of a Monte Carlo study comparing the intermediate sample size performance characteristics of these two statistics, which until now has not been investigated, with two standard competitors, the analysis of variance (ANOVA) F test and Friedman's test.

1. INTRODUCTION

Since the publication of their paper, Kepner and Robinson (1988) have received numerous inquiries asking which statistic, the rank transformation statistic or Koch's statistic, is better in the sense of maintaining a desired significance level while providing good power. The purpose of this paper is to answer this question by comparing the Monte Carlo performance of these two statistics, the ANOVA F test, and Friedman's test. The test statistics and the Monte Carlo study are described in Sections 2 and 3, respectively, representative results are presented in Section 4, and summary remarks appear in Section 5.

2. THE TEST STATISTICS

The linear model is assumed to be

$$X_{ij} = \theta + \beta_i + \tau_j + \varepsilon_{ij}, \tag{1}$$

where $i = 1, \dots, n$ indicates the block, $j = 1, \dots, k$ indicates the treatment, θ is an unknown constant, β_i is the random effect of block i, τ_j is the fixed effect of treatment j subject to the constraint $\sum_{j=1}^{k} \tau_j = 0$, ε_{ij} is the random error term for block i and treatment j and $\{\beta_1, \dots, \beta_n\}$ and $\{\varepsilon_{ij} \mid 1 \le i \le n \text{ and } 1 \le j \le k\}$ are independent sets of mutually independent random variables. The null and research hypotheses are $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$ and H_1 : not all τ_j's equal, respectively. Let R_{ij} be the rank of X_{ij} among all nk observations and define $\bar{R}_{i\cdot} = \frac{1}{k} \sum_{j=1}^{k} R_{ij}$, $\bar{R}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} R_{ij}$, and $\mu = \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} R_{ij} = \frac{(nk+1)}{2}$. Also, let r_{ij} be the rank of X_{ij} among the k observations in block i with $\bar{R}_j = \frac{1}{n} \sum_{i=1}^{n} r_{ij}$. Four competing statistics for detecting treatment effects when this design is appropriate are the ANOVA F statistic,

$$\mathrm{FT} = \frac{(k-1)^{-1}\, n \sum_{j=1}^{k} (\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2}{[(n-1)(k-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k} (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2},$$

where $\bar{X}_{i\cdot} = \frac{1}{k} \sum_{j=1}^{k} X_{ij}$, $\bar{X}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} X_{ij}$, and $\bar{X}_{\cdot\cdot} = \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} X_{ij}$,

Friedman's statistic,

$$\mathrm{FR} = 12n[k(k+1)]^{-1} \sum_{j=1}^{k} \left(\bar{R}_j - \frac{k+1}{2} \right)^2,$$

Koch's statistic,

$$\mathrm{RT}_1 = \frac{(k-1)^{-1}\, n \sum_{j=1}^{k} (\bar{R}_{\cdot j} - \mu)^2}{[n(k-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k} (R_{ij} - \bar{R}_{i\cdot})^2},$$

and the rank transformation statistic,

$$\mathrm{RT}_2 = \frac{(k-1)^{-1}\, n \sum_{j=1}^{k} (\bar{R}_{.j} - \mu)^2}{[(n-1)(k-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k} (R_{ij} - \bar{R}_{i.} - \bar{R}_{.j} + \mu)^2} \,,$$

which is just the ANOVA F statistic with the data replaced by their ranks in the combined sample.

3. DESCRIPTION OF THE MONTE CARLO STUDY

In the Monte Carlo study, values of X_{ij} were generated using model (1), where, without loss of generality, θ was assumed to be zero. The distributions from which the random block and error terms were taken include the uniform, normal, logistic, double exponential, Cauchy, and exponential. Each observation of β_i and ε_{ij} was standardized by subtracting the mean of the distribution from which it was taken and then dividing by its standard deviation. For the Cauchy distribution, 0 and .841345 played the roles of a mean and a standard deviation as suggested by Hogg, Fisher, and Randles (1975); see, for example, Groggel (1987). Nine combinations of n and k were used. For both $k = 3$ and $k = 4$, $n =$ 5, 10, 15 and $k = 5$, $n =$ 10, 15, 20. For each combination of n and k, $\rho = \mathrm{corr}(X_{ij},\ X_{ij'})$, with $j \neq j'$, took on the values 0, .2, .5, and .8 in the cases where this parameter exists. For every k, the five different treatment vectors given in Table 1 were used.

Table 1
Treatment Effect Vectors used in the Monte Carlo Study

	$k = 3$	$k = 4$	$k = 5$
τ_1	$(0, 0, 0)'$	$(0, 0, 0, 0)'$	$(0, 0, 0, 0, 0)'$
τ_2	$(-.3, 0, .3)'$	$(-.25, -.25, .25, .25)'$	$(-.333, -.167, 0, .167, .333)'$
τ_3	$(-.6, 0, .6)'$	$(-.5, -.167, .167, .5)'$	$(-.667, -.333, 0, .333, .667)'$
τ_4	$(-.2, -.1, .3)'$	$(-.15, -.15, -.15, .45)'$	$(-.1, -.1, -.1, -.1, .4)'$
τ_5	$(-.2, -.2, .4)'$	$(-.075, -.075, -.075, .225)'$	$(-.05, -.05, -.05, -.05, .2)'$

For each of the 54 combinations of n, k, and distribution of the block and error terms, 1000 samples were generated using a new random seed for each combination of n and k. In each sample, the block and error terms were multiplied by appropriate constants to obtain the four different values of ρ and to keep $V(X_{ij}) = 1$, thus each component of a $\boldsymbol{\tau}$ is measured in standard deviation units. Model (1) became

$$X_{ij} = a\beta_i + \tau_j + c\varepsilon_{ij}\,,$$

where $\rho = 0$, .2, .5, and .8, as the ordered pair $(a,\ c)$ assumes the values (0, 1), ($\sqrt{.2}$, $\sqrt{.8}$), ($\sqrt{.5}$, $\sqrt{.5}$), ($\sqrt{.8}$, $\sqrt{.2}$), respectively. In the case of the Cauchy distribution, ρ refers to the association induced between two variables in a block by a particular choice of a and c. The test statistics were then calculated for each value of ρ and for each $\boldsymbol{\tau}$. Each statistic was compared to its appropriate $\alpha = .05$ level critical value and the number of rejections counted. In particular, RT_1 and RT_2 were compared to their limiting $(k-1)^{-1}\,\chi^2_{k-1}$ distribution (for details, see Kepner and Robinson (1988)) and to an F–distribution with $k-1$ and $(n-1)(k-1)$ degrees of freedom in the spirit of Iman and Davenport (1980) and others. To obtain the significance level for Friedman's test, the method of randomization was used for $k = 3$ and all n and for $k = 4$ and $n = 5$, while in all other cases Friedman's statistic was compared to a χ^2_{k-1} distribution due to the lack of exact tables.

4. RESULTS

The difference between RT_1 and RT_2 occurs in the denominators of the statistics. For a permutation test, RT_1, with the less complex denominator, would be preferable to RT_2 in the interest of computational speed. But when the two statistics are compared to an F–distribution with $k-1$ and $(n-1)(k-1)$ degrees of freedom and their limiting $(k-1)^{-1}\,\chi^2_{k-1}$ distribution, only RT_2 compared to the F–distribution was able to maintain the .05 significance level. RT_1 compared to the $(k-1)^{-1}\,\chi^2_{k-1}$ distribution was consistently better than when it was compared to the F–distribution while the opposite was true for RT_2; henceforth RT_1 will refer to Koch's statistic compared to the $(k-1)^{-1}\chi^2_{k-1}$ distribution and RT_2 will refer to the rank transformation statistic compared to the F–distribution. The following discussion highlights the power of the statistics with respect to each type of distribution of the random block and error terms.

Normal. As expected, FT was superior in power to RT_1, RT_2 and FR, but RT_2 was only slightly less powerful than FT and remained very competitive for $\rho \leq .5$.

Uniform. FT was more powerful than the other statistics. RT_2 was a close second with RT_1 and FR being the least powerful.

Logistic. When the within block correlation was low, $\rho \leq .2$, RT_2 was the most powerful statistic. For $\rho \geq .5$, FT gained a slight advantage in power over RT_2 while FR was the least powerful since its χ^2 approximation is extremely conservative for small sample sizes (see, for example, Iman and Davenport (1980)).

Double Exponential. RT_2 was the most powerful for $\rho \leq .5$. For $\rho = .8$, all four statistics were very close, with a slight edge in power going to FT, even though FT had difficulty maintaining the significance level, tending to be rather conservative for $k = 4$ and $k = 5$. The χ^2 approximation to FR was again excessively conservative.

Cauchy. As expected, FR, RT_1 and RT_2 performed much better than FT, which was extremely conservative and provided low power. RT_2 was the most powerful only when $\rho = 0$, while FR was the most powerful when $\rho \geq .2$.

Exponential. RT_2 was the most powerful of the four statistics when $\rho \leq .2$. When $k = 3$, RT_1 and RT_2 were more powerful than FR even when $\rho \geq .5$, and when $k = 4$ or 5, they remained very competitive with FR for $\rho \geq .5$. FT was less powerful than the other statistics and had difficulty maintaining the significance level, tending to be very conservative for this skewed distribution.

As an example of typical results, Tables 2A and 2B show the number of rejections out of 1000 for the six distributions and four values of ρ when $k = 3$, $n = 15$ and $\alpha = .05$.

5. CONCLUSION

In this Monte Carlo study, the performance of Koch's statistic (1969), the rank transformation statistic discussed by Kepner and Robinson (1988), the ANOVA F test, and Friedman's test, denoted RT_1, RT_2, FT, and FR, respectively, were compared assuming that the data are obtained from a complete, fixed effects repeated measures design. In addition to considering several symmetric and skewed continuous distributions for both the block and error terms, the within block correlation, $\rho = \text{corr}(X_{ij},\ X_{ij'})$, $j \neq j'$, was controlled and set to 0, .2, .5, and .8 in every case where this parameter exists. The number of treatments considered, k, ranged from 3 to 5 and included both symmetric and skewed treatment effects. The sample sizes, n, were 5, 10, and 15 for $k = 3$ and $k = 4$ and 10, 15, and 20 for $k = 5$.

RT_1 and RT_2 were each compared to an F–distribution with $k - 1$ and $(n-1)(k-1)$ degrees of freedom and a $(k-1)^{-1}\,\chi^2_{k-1}$ distribution. Only RT_2, when compared to the F–distribution, was able to maintain its significance level and compete with FT and FR for power. RT_2 held its significance level very consistently over all cases of n, k, ρ, and distribution of the block and error terms, except for the combination $k = 3$, $n = 5$, (few treatments and very small sample size) and $\rho = 0$, where it tended to be liberal.

FT was the most powerful statistic for the uniform and normal distributions and also for the logistic distribution when $\rho \geq .5$. RT_2 was the most powerful for the logistic distribution when $\rho \leq .2$ and for the double exponential distribution when $\rho \leq .5$. FT was slightly more powerful than the others for the double exponential distribution when $\rho = .8$, but had difficulty maintaining its significance level for $k = 4$ and $k = 5$. FR's poor performance in the logistic and double exponential distributions can be attributed to the conservative nature of its χ^2

approximation when n and k are small. For the Cauchy distribution, RT_2 was the most powerful for $\rho = 0$, while FR was the most powerful for $\rho \geq .2$. RT_2 was the most powerful for the exponential distribution when $\rho \leq .2$ and for $\rho \geq .5$ and $k = 3$. FR was the most powerful for the exponential distribution for $\rho \geq .5$ and $k = 4$ or $k = 5$. These remarks are summarized in Table 3.

Table 3
Summary of Monte Carlo Study Results*

	Uniform	Normal	Logistic	Dbl. Exp.	Cauchy	Exponential	
						$k = 3$	$k = 4, 5$
$\rho = 0$	FT	FT	RT_2	RT_2	RT_2	RT_2	RT_2
$\rho = .2$	FT	FT	RT_2	RT_2	FR	RT_2	RT_2
$\rho = .5$	FT	FT	FT	RT_2	FR	RT_2	FR
$\rho = .8$	FT	FT	FT	FT	FR	RT_2	FR

FT = Normal Theory F–test, RT_2 = rank transformation test, FR = Friedman's test
ρ = correlation in non–Cauchy case between X_{ij} and $X_{ij'}$ $(j \neq j')$
* For example, RT_2 in row 3 of column 4 indicates that the rank transformation statistic was generally the most powerful when $\rho = .5$ and the block and error variables both had double exponential distributions.

BIBLIOGRAPHY

Groggel, D. J. (1987), "A Monte Carlo Study of Rank Tests for Block Designs," *Communications in Statistics, Part B – Simulation and Computation*, 16, 601-620.

Hogg, R. V., Fisher, D. M., & Randles, R. H. (1975), "A Two-Sample Adaptive Distribution-Free Test," *Journal of the American Statistical Association*, 70, 656-661.

Iman, R. L., and Davenport, J. M. (1980), "Approximations of the Critical Region of the Friedman Statistic," *Communications in Statistics, Part A – Theory and Methods*, 9, 571-595.

Kepner, J. L., and Robinson, D. H. (1988), "Nonparametric Methods for Detecting Treatment Effects in Repeated-Measures Designs," *Journal of the American Statistical Association*, 83, 456-461.

Koch, G. G. (1969), "Some Aspects of the Statistical Analysis of 'Split Plot' Experiments in Completely Randomized Layouts," *Journal of the American Statistical Association*, 64, 485-505.

Table 2A

Number of Rejections out of 1000 Replications with $\alpha = .05$, $n = 15$, and $k = 3$

		Normal				Uniform				Logistic			
		FT	RT_1	RT_2	FR	FT	RT_1	RT_2	FR	FT	RT_1	RT_2	FR
	τ_1	48	45	51	59	50	44	54	52	53	44	54	52
	τ_2	246	233	249	193	307	267	284	227	315	325	345	261
$\rho = 0$	τ_3	767	732	745	632	816	751	771	659	822	827	842	739
	τ_4	192	187	199	170	240	205	221	189	242	246	264	198
	τ_5	323	299	318	258	360	318	343	265	381	384	402	320
	τ_1	48	41	48	59	50	45	51	52	53	46	53	52
	τ_2	306	278	297	236	359	307	322	269	387	370	390	316
$\rho = .2$	τ_3	856	822	835	740	896	842	856	740	894	890	896	828
	τ_4	246	212	227	196	279	230	242	217	302	283	300	247
	τ_5	395	355	371	306	437	366	389	318	458	437	458	386
	τ_1	48	38	46	59	50	39	44	52	53	42	45	52
	τ_2	467	412	427	363	534	450	473	396	542	507	523	467
$\rho = .5$	τ_3	970	952	954	903	984	965	967	908	979	973	977	944
	τ_4	362	314	335	288	422	358	379	313	435	408	433	388
	τ_5	600	544	569	477	637	549	569	471	640	613	632	554
	τ_1	48	44	54	59	50	42	49	52	53	39	44	52
	τ_2	856	790	810	740	896	832	850	740	894	846	858	828
$\rho = .8$	τ_3	1000	1000	1000	1000	1000	1000	1000	999	1000	1000	1000	999
	τ_4	767	698	713	629	804	722	738	638	803	748	768	714
	τ_5	954	913	921	875	968	927	932	846	957	934	944	911

Tests: FT = Normal theory F–test, RT_1 = Koch's test, RT_2 = rank transformation test, FR = Friedman's test

ρ = correlation in non–Cauchy cases between X_{ij} and $X_{ij'}$ $(j \neq j')$

$\tau_1 = (0, 0, 0)'$, $\tau_2 = (-.3, 0, .3)'$, $\tau_3 = (-.6, 0, .6)'$, $\tau_4 = (-.2, -.1, .3)'$, $\tau_5 = (-.2, -.2, .4)'$

Table 2B

Number of Rejections out of 1000 Replications with $\alpha = .05$, $n = 15$, and $k = 3$

		Double Exponential				Cauchy				Exponential			
		FT	RT_1	RT_2	FR	FT	RT_1	RT_2	FR	FT	RT_1	RT_2	FR
	τ_1	50	44	54	52	20	44	54	52	44	44	54	52
	τ_2	324	405	419	323	26	96	102	95	345	527	549	439
$\rho = 0$	τ_3	824	881	892	810	38	257	265	210	833	931	934	865
	τ_4	260	309	332	278	22	89	96	184	255	429	454	372
	τ_5	398	482	501	391	24	105	115	106	402	636	662	529
	τ_1	50	44	52	52	20	44	50	52	44	44	50	52
	τ_2	396	431	452	378	27	88	104	107	400	536	560	514
$\rho = .2$	τ_3	893	920	925	869	45	219	235	240	889	954	958	915
	τ_4	309	333	361	327	22	80	88	93	310	428	450	439
	τ_5	474	516	539	467	24	97	107	120	493	649	671	608
	τ_1	50	39	46	52	20	41	46	52	44	41	50	52
	τ_2	562	555	573	550	33	108	120	136	567	652	661	667
$\rho = .5$	τ_3	977	979	981	963	58	250	269	335	966	987	988	978
	τ_4	451	464	487	457	26	88	102	119	467	555	580	589
	τ_5	652	660	675	631	32	113	120	155	677	765	784	776
	τ_1	50	40	44	52	20	38	45	52	44	38	46	52
	τ_2	893	855	867	869	45	144	159	240	889	905	910	915
$\rho = .8$	τ_3	1000	1000	1000	999	149	395	419	617	1000	999	999	999
	τ_4	807	783	804	789	38	117	131	208	831	839	851	876
	τ_5	953	947	955	932	53	162	182	293	953	964	974	965

Tests: FT = Normal theory F–test, RT_1 = Koch's test, RT_2 = rank transformation test, FR = Friedman's test

ρ = correlation in non–Cauchy cases between X_{ij} and $X_{ij'}$ $(j \neq j')$

$\tau_1 = (0, 0, 0)'$, $\tau_2 = (-.3, 0, .3)'$, $\tau_3 = (-.6, 0, .6)'$, $\tau_4 = (-.2, -.1, .3)'$, $\tau_5 = (-.2, -.2, .4)'$