Behavior of nonparametric tests in longitudinal designs

Stefanie Hayoz
Institute of Mathematical Statistics and Actuarial Science.
University of Bern, Switzerland

Keywords: Nonparametric procedures. Longitudinal data. Repeated measurements. Simulation study.
AMS: 62G10, 62G35

Abstract

Nonparametric procedures need less model assumptions than the analogous parametric ones. Moreover the nonparametric procedures are more robust against deviations from the model assumptions. It is well known that the Friedman test is used to analyze the treatment effect in the randomized complete block design. The Friedman test could also be used in a longitudinal design to analyze the time effect. Alternatives are an ANOVA type test and the parametric ANOVA test for repeated measures. The three tests are compared with respect to the power and the control of the nominal level by means of a simulation study. This study incorporates different covariance structures that are used to model the dependence among the repeated measurements and the results are based on 40'000 replications for each situation considering small sample sizes.

1. Introduction

In this paper three tests to analyze the treatment effect in the randomized complete block design (see e.g. [3]) with k treatments and n blocks are compared. The blocks could be individuals and the treatments could be time points, then we would have the longitudinal LD-F1 design of Brunner and Langer (see e.g. [1]). The observations are denoted by $X_{ij}, i \leq n$ and $j \leq k$ and their marginal distributions by F_{ij}. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ denote the vector of all treatment effects, $\boldsymbol{I}_k$ the $k \times k$ identity matrix, $\boldsymbol{J}_k$ the $k \times k$ matrix of 1's, and $\boldsymbol{P}_k = \boldsymbol{I}_k - \frac{1}{k}\boldsymbol{J}_k$. Then the null hypothesis that all treatment effects are the same can be formulated as $H_0^\mu : \boldsymbol{P}_k\boldsymbol{\mu} = \mathbf{0}$. Under the assumption that the marginal distributions do not differ from block to block, i.e. $F_{1j} = \ldots = F_{nj} = F_j, \forall j \leq k$, the null hypothesis that the marginal distributions are the same for all treatments can be formulated as $H_0^F : \boldsymbol{P}_k\boldsymbol{F} = \mathbf{0}$, where $\boldsymbol{F} = (F_1, \ldots, F_k)'$. It is known that the relation between the two null hypotheses is $H_0^F : \boldsymbol{P}_k\boldsymbol{F} = \mathbf{0} \Rightarrow H_0^\mu : \boldsymbol{P}_k\boldsymbol{\mu} = \mathbf{0}$.

1.1. Friedman test

In order to use the Friedman test the marginal distributions must come from a continuous distribution, otherwise a tie correction has to be made.

Furthermore the observations do not have to be independent but interchangeable, thus the covariance matrix of the data must have a compound symmetry structure. The marginal distributions may differ from block to block. The observations are ranked in each block separately. The rank of X_{ij} is denoted by R_{ij}. Let $\bar{R}_{.j} = \frac{1}{n}\sum_{i=1}^{n} R_{ij}$. The distribution of the Friedman test statistic

$$Q = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(\bar{R}_{.j} - \frac{(k+1)}{2} \right)^2$$

under the null hypothesis $H_0^\mu : \boldsymbol{P}_k \boldsymbol{\mu} = \boldsymbol{0}$ is

$$P_{H_0^\mu}(Q = c) = \frac{\#\{\text{Arrangements with } Q = c\}}{(k!)^n} \ .$$

For small n and k tables of this distribution exist. For $n \to \infty$ the distribution of Q can be approximated by a χ^2 distribution with $k-1$ degrees of freedom. For more detailed information about the Friedman test and tables of the distribution see e.g. [3].

1.2. ANOVA type test

In order to use the ANOVA type test the marginal distributions must be equal in each block, i.e. $F_{1j} = \ldots = F_{nj} = F_j, \forall j \le k$. Furthermore there must exist some variability in the observations, i.e. not all observations can be equal. All $N = nk$ observations are ranked together (not blockwise). To describe differences in the distributions, the relative treatment effects $p_j = \int H dF_j$ can be used, where $H = \frac{1}{k}\sum_{j=1}^{k} F_j$. A consistent estimate of $\boldsymbol{p} = (p_1, \ldots, p_j)'$ is $\hat{\boldsymbol{p}} = (\bar{R}_{.1} - \frac{1}{2}, \ldots, \bar{R}_{.k} - \frac{1}{2})'/N$. To test the hypothesis $H_0^F : \boldsymbol{P}_k \boldsymbol{F} = \boldsymbol{0}$, the ANOVA type statistic

$$A = \frac{n}{tr(\boldsymbol{P}_k \hat{\boldsymbol{V}}_n)} \hat{\boldsymbol{p}}' \boldsymbol{P}_k \hat{\boldsymbol{p}} = \frac{n}{N^2 tr(\boldsymbol{P}_k \hat{\boldsymbol{V}}_n)} \sum_{j=1}^{k} (\bar{R}_{.j} - \bar{R}_{..})^2$$

can be used, where

$$\hat{\boldsymbol{V}}_n = \frac{1}{N^2(n-1)} \sum_{i=1}^{n} (\boldsymbol{R}_i - \bar{\boldsymbol{R}}_.)(\boldsymbol{R}_i - \bar{\boldsymbol{R}}_.)'$$

is an estimator for the covariance matrix $\boldsymbol{V}_n$, with $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{ik})'$ and $\bar{\boldsymbol{R}}_. = (\bar{R}_{.1}, \ldots, \bar{R}_{.k})'$. If $tr(\boldsymbol{P}_k \hat{\boldsymbol{V}}_n) > t_0 > 0$ (i.e. variability in the observations) the distribution of A can be approximated under $H_0 : \boldsymbol{P}_k \boldsymbol{F} = \boldsymbol{0}$ with a $F(\hat{f}, \infty)$ distribution, where

$$\hat{f} = \frac{(tr(\boldsymbol{P}_k \hat{\boldsymbol{V}}_n))^2}{tr(\boldsymbol{P}_k \hat{\boldsymbol{V}}_n \boldsymbol{P}_k \hat{\boldsymbol{V}}_n)} \ .$$

For more detailed information about the relative treatment effects and the ANOVA type test see e.g. [1].

1.3. ANOVA test for repeated measures

A parametric test for this situation is the ANOVA test for repeated measures. In order to use this test the covariance matrix of the data must have a compound symmetry structure and the data must come from a normal distribution. The mean squares MS_T and MS_R are defined as follows:

$$MS_T = \frac{SS_T}{k-1} = \frac{1}{k-1}\sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{X}_{.j} - \bar{X}_{..})^2 = \frac{n}{k-1}\sum_{j=1}^{k}(\bar{X}_{.j} - \bar{X}_{..})^2$$

and

$$MS_R = \frac{SS_R}{(n-1)(k-1)} = \frac{1}{(n-1)(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 \ ,$$

where

$$\bar{X}_{i.} = \frac{1}{k}\sum_{i=1}^{n} X_{ij} \ , \quad \bar{X}_{.j} = \tfrac{1}{n}\textstyle\sum_{j=1}^{k} X_{ij} \quad \text{and } \bar{X}_{..} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} X_{ij} \ .$$

Under $H_0^{\mu} : \boldsymbol{P}_k \ \boldsymbol{\mu} = \boldsymbol{0}$ the ANOVA test statistic

$$F = \frac{MS_T}{MS_R} = \frac{SS_T(n-1)}{SS_R}$$

has a $F(n-1,(n-1)(k-1))$ distribution. For detailed information see e.g. [2].

2. Simulation study

To learn more about the behavior of the three mentioned tests in different situations, i.e. how they react to violations of the assumptions and how good the approximations of the distributions hold for small sample sizes, they are compared with respect to the power and the control of the nominal level by means of a simulation study.

2.1. Simulation of the nominal level

The randomized complete block design with $n = 10, 20, 30$ blocks and $k = 4, 6, 8$ treatments is considered. The dependence among the observations of a block is modelled by CS, AR(1) and MA(1) for correlation $\rho = 0.8, 0.5$ and 0.3 (for MA(1) only $\rho = 0.5$ and 0.3). The data are simulated with an underlying normal distribution, i.e. $X_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$ such that $\mu_j = 0, \forall j \leq k$

and $\sigma^2 = 1$. For each situation $n_{sim} = 40'000$ replications are made. The simulation error is ± 0.00109 for $\alpha = 0.05$. In figure 1 the simulation results for $\rho = 0.5$ and $\alpha = 0.05$ are displayed.

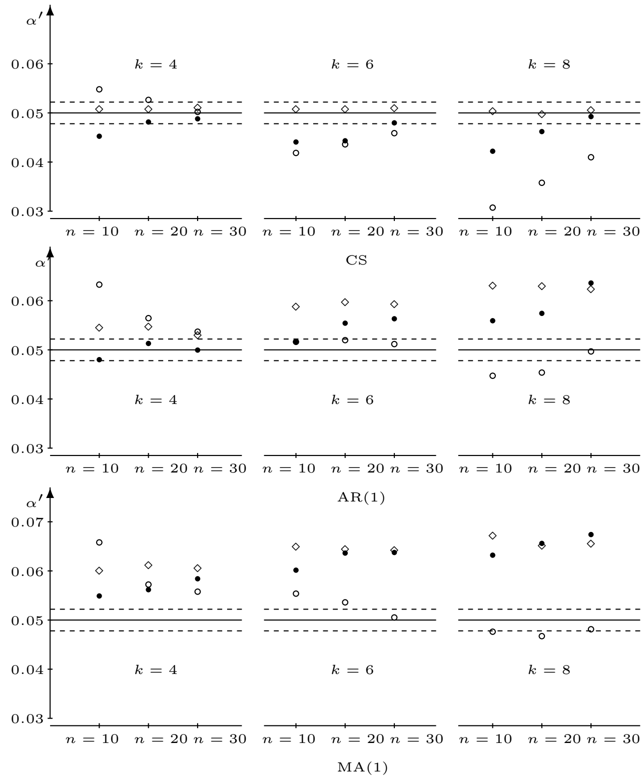

Figure 1: Simulated probabilties for the type I error α' for the Friedman test (•), the ANOVA type test (◦) and the ANOVA test for repeated measures (⋄) for $\alpha = 0.05$ modelled with the three covariance structures CS, AR(1) and MA(1) with correlation $\rho = 0.5$ and an underlying normal distribution.

Because the distribution of the ANOVA type test is determined only approximately, it holds the nominal level better with increasing sample size n. With increasing number of groups k the ANOVA type test becomes more conservative because the estimator $\hat{f}$ of f is only approximately unbiased. For AR(1) and MA(1) with large correlations ρ and number of treatments k the Friedman test and the ANOVA test are liberal, for small ρ and k they control the nominal level better because a violation of the compound symmetry assumption is less severe for small ρ and k. For CS the ANOVA test holds the nominal level very good, the Friedman test only if the sample size n is large enough because the distribution is determined only approximately.

2.2. Correction of the p-values

To ensure a correct comparison of the power, the p-values should be corrected such that the true level is equivalent to the nominal level. Therefore the p-values are multiplicated with a correction factor that is the ratio of the nominal level α and the α-quantile of the p-values that were obtained by the simulations under H_0.

2.3. Simulation of the power

The randomized complete block design with $n = 10$ blocks and $k = 4$ treatments is considered. The dependence among the observations of a block is modelled by CS, AR(1) and MA(1) for correlation $\rho = 0.5$. The data are simulated with an underlying normal distribution, i.e. $X_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$ such that $\sigma^2 = 1$. Simulations are performed for two different patterns of alternatives, namely a linear trend of the form $\mu_j = \delta \cdot \frac{j}{4}, \forall j \leq 4$ and an alternative with a change point of the form $\mu_1 = \mu_2 = 0$, $\mu_3 = \mu_4 = \delta$. For each situation $n_{sim} = 40'000$ replications were made. The simulation error is $\pm\sqrt{p(1-p)/40'000}$. In figure 2 the simulation results for the first kind of alternative are displayed.

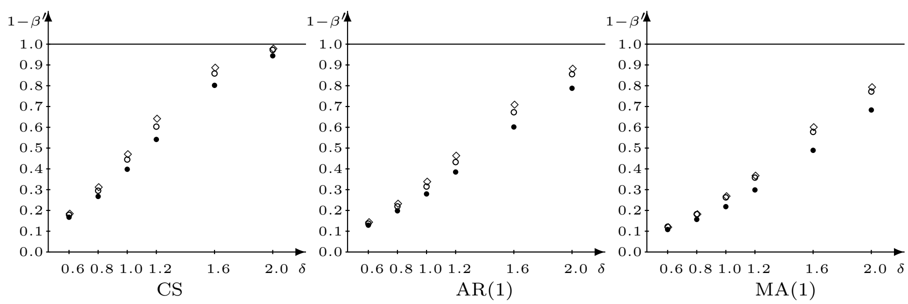

Figure 2: Power of the Friedman Test (•), the ANOVA type test (◦) and the ANOVA test for repeated measures (⋄) for an alternative of the type $\mu_j = \delta \cdot \frac{j}{4}$, $j \leq 4$ for $\alpha = 0.05$ with $n = 10$ blocks and $k = 4$ groups modelled with the three covariance structures CS, AR(1) and MA(1) with correlation $\rho = 0.5$ and an underlying normal distribution.

For all tests and covariance structures the larger the correlation is, the faster the power increases with increasing δ. In all cases the power of the Friedman test is smaller and increases more slowly with increasing δ than the power of the ANOVA type test. The power of the ANOVA type test is in all cases smaller and increases more slowly with increasing δ than the power of the ANOVA test. For small δ the differences are not relevant. For CS the power increases faster with increasing δ than for AR(1) and for AR(1) faster than for MA(1). For the alternatives of the structure $\mu_j = \delta \cdot \frac{j}{4}$ the power increases for all tests faster than for alternatives of the structure $\mu_1 = \mu_2 = 0$, $\mu_3 = \mu_4 = \delta$.

However the differences are rather small.

Simulations with other underlying distributions are in progress.

Acknowledgements: I would like to thank Prof. Dr. Jürg Hüsler for his support.

3. Bibliography

[1] Brunner, E., Domhof, S. and Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments.* Wiley, New York.

[2] Davis, C. S. (1952). *Statistical methods for the analysis of repeated measurements.* Springer, New York.

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