CONSEQUENCES OF ASSUMPTION VIOLATIONS REGARDING ONE-WAY ANOVA

Tomáš Marcinko

Abstract

Nearly all classical statistical hypothesis tests are derived under a few fundamental assumptions, which may or may not be met in real world applications, and the classical analysis of variance is no exception. The main aim of this article is to study consequences of the most crucial assumption violations concerning one-way ANOVA tests, mainly their effect on type I errors. The focus will be on the classical F test, as well as on popular procedures of a multiple means comparison. Based on a simulation study the consequences of non-normality and heteroscedasticity will be examined for various sample sizes. The resulting type I errors of classical tests will be compared with those of appropriate nonparametric tests, specifically with the errors of the Kruskal-Wallis test and the multiple means comparison based on pairwise comparisons using Wilcoxon rank sum tests with Bonferroni correction or tests based on bootstrap methodology.

Key words: one-way ANOVA, F test, multiple means comparison, assumption violations, nonparametric alternatives

JEL Code: C12, C14

Introduction

The problem of testing for equality of means in a one-way independent groups design is very frequent in many field of science. Analysis of variance (abbreviated ANOVA) refers to a statistical technique that is used to analyze the differences between means of two or more groups. In other words, the one-way ANOVA, inter alia, tests the null hypothesis that samples in two or more groups are drawn from populations with the same mean values.

In this article we will focus on a very common one-way ANOVA F-test, which is fully justified under the assumptions of independence of observations, normality and homogeneity of group variances (homoscedasticity). However, these assumptions may not be met in real world applications.

The main aim of this article is a simulation study that will examine the consequences of non-normality and heteroscedasticity on a type I error of the ANOVA F-test and two popular multiple means comparison methods that are used to determine which group means differ – the Tukey's HSD (honestly significant difference) test and the Bonferroni correction. The results obtained by these classical tests will then be compared with those of appropriate nonparametric tests, specifically with the Kruskal-Wallis one-way analysis of variance by ranks, where the multiple means comparison will be based on pairwise comparisons using the Wilcoxon rank sum tests with Bonferroni correction or tests based on the bootstrap BC_a (bias-corrected and accelerated) method.

1 Parametric and nonparametric approach to one-way ANOVA

When dealing with fixed-effects one-way ANOVA, the most popular approach of many statisticians, researchers or data analysts is the parametric one, i.e. using the ANOVA F-test, often with some common multiple means comparison methods (e.g. Tukey's HSD) used when the difference between group means is significant.

Let's assume we have a fully randomized experiment with a single factor and we wish to determine, whether the population means in all groups are the same. In order to examine the mere effects of non-normality or heteroscedasticity, we will only consider a balanced design with the number of observation in each group being the same. Let y_{ij}, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$, denote the jth observation in ith group, where I is a number of groups (treatments) and J is a number of observations in each group. Obviously, the total number of observations is $n = IJ$. The means model of one-way analysis of variance can be written as

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

where μ_i, $i = 1, 2, \ldots, I$, are the group means and ε_{ij}, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$, are zero-mean random errors. Hence, the null hypothesis can be written as $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_I$.

The ANOVA F-test assumes that the observations y_{ij} are independent and identically distributed in every group and follow a normal distribution with mean μ_i and variance σ^2, which is the same for all groups. The fundamental procedure of ANOVA is based on comparing the variability between groups with the variability within groups, which can be represented via between-group mean square

$$\mathrm{MS_B} = \sum_{i=1}^{I} J(\bar{y}_i - \bar{y})^2 / (I-1)$$

and within-group mean square

$$\mathrm{MS_W} = \sum_{i=1}^{I}\sum_{j=1}^{J} (y_{ij} - \bar{y}_i)^2 / (n-I)$$

respectively. When the assumptions of independence, normality and homoscedasticity hold then the statistic

$$F = \frac{\mathrm{MS_B}}{\mathrm{MS_W}}$$

follows the F distribution on I – 1 and $n - I$ degrees of freedom. Therefore, we can reject the null hypothesis in favor of the alternative hypothesis (the population means μ_i are not all equal), if the value of the F statistic is greater than a critical value from the F distribution. The technical derivation of the ANOVA F-test can be found in many statistical textbooks (for example Cramér, 1946).

If the null hypothesis is rejected by the ANOVA F-test, the multiple comparison procedures are often used to determine, which group means differ. In a one-way ANOVA involving I group means, we need to make (I – 1) / 2 pairwise comparisons simultaneously. In this article, we will consider two common multiple comparison methods. The first one, the Bonferroni correction (Dunn, 1961), is based on replacing a significance level α with α / I in order to obtain an overall confidence level of simultaneous confidence intervals greater than or equal to 1 – α. Although this method is extremely easy to use, it may also be quite conservative (i.e. having a type I error lower than the significance level). The second multiple comparison method that will be considered is the Tukey's HSD (Tukey, 1953), which is specifically designed for comparing group means in ANOVA setting. It is based on the distribution of a studentized range and simultaneously compares all possible pairs of group means. For further details on these methods see Miller (1981). Over the years several other multiple comparison procedures were developed – a comprehensive review of the most important methods can be found in (Day and Quinn, 1989).

In this article we will compare the results of the ANOVA F-test with the results of the most popular nonparametric alternative – the Kruskal-Wallis one-way analysis of variance by ranks (Kruskal and Wallis, 1952). For multiple means comparison, the Wilcoxon rank-sum tests (Wilcoxon, 1945) with the Bonferroni correction may be used. Another procedure that will be considered is the bootstrap BC_a method (Efron, 1987) with the Bonferroni correction. However, some other methods, such as Neményi test (Neményi, 1963), may be preferable.

2 Simulation study

The following simulation study will focus on the Monte Carlo estimation of a type I error of the parametric and nonparametric ANOVA tests shortly discussed in the previous section under various violations of normality and homoscedasticity.

For the purpose of examining the effects of non-normality on the type I error, data in every group were simulated from the following distributions (hence, the null hypothesis of the ANOVA test holds):

- normal distribution: $X \sim N(100; 100)$
- modified Student's distribution: $X \sim 100 + 5.7735\ t(3)$
- uniform distribution: $X \sim U(82.6795; 117.3205)$
- gamma distribution: $X \sim \Gamma(100; 1)$
- log-normal distribution: $X \sim LN(4.6002; 0.00995)$
- skew normal distribution: $X \sim SN(88.417; 15.303; 3)$
- shifted exponential distribution: $X \sim 90 + Ex(0.1)$
- contaminated normal distribution: $X \sim (1 - \varepsilon)\ N(100; 100) + \varepsilon\ N(100; 10\ 000)$.

Without the loss of generality, all of these distributions (except for a contaminated normal distribution, which has larger variance) were calibrated so that they have the population mean 100 and the variance 100. The first three distributions (normal, modified Student's and uniform) are symmetric around the population mean and the other four distributions (gamma, log-normal, skew normal and shifted exponential) are asymmetric with a gamma distribution being the least skewed ($\gamma_1 = 0.2$) and a shifted exponential distribution being the most skewed ($\gamma_1 = 2$). The skewness of the other two distributions is only moderate ($\gamma_1 = 0.301$ for a log-normal distribution and $\gamma_1 = 0.667$ for a skew normal distribution). Lastly, the contaminated normal distribution is a mixture distribution, where the majority of the population comes from a specified normal distribution, whereas a small proportion of the population ($\varepsilon = 0.05$) comes from a normal distribution with the same mean but much larger variance, i.e. outliers can be drawn from such a population.

For the purpose of examining the mere effects of heteroscedasticity, the group data sets were simulated from a normal distribution with same means but various variances.

After simulating the data, the ANOVA tests were applied and the type I errors were estimated as the proportion of simulated data sets that rejected the null hypothesis, which is actually true. The simulation study consisted of 10 000 simulated data sets so that the Monte Carlo error was sufficiently low. For the bootstrap method 5 000 bootstrap samples were generated. The simulation study was computed in the programming language R version 2.15.3 (R core team, 2013) using additional packages bootstrap, multcomp and sn.

2.1 The effects of non-normality

Tables 1 and 2 provide estimates of type I errors for the ANOVA F-test and the Kruskal-Wallis test, as well as the aforementioned methods of a multiple means comparison, in case of 3 and 5 treatment groups, respectively, under the given violations of normality and for various sample sizes.

It is quite obvious that the ANOVA F-test has a type I error reasonably close to the chosen significance level (for the purpose of this study a significance level 0.05 was used) even for small sample sizes and it only becomes a little bit conservative, if there is substantial skewness (as in case of an exponential distribution) or in case of long-tailed distributions (such as a Student's t distribution on 3 degrees of freedom). And, as expected, with larger sample sizes the type I error gets closer to the nominal significance level. The only exception, when the ANOVA F-test was too conservative even for large sample sizes, is the case of a contaminated normal distribution. This is of course no surprise, as outliers can have a very big influence on the sample mean.

On the other hand, the Kruskal-Wallis test performs really well in respect of a type I error for any underlying distribution including a contaminated normal distribution. Consequently, when we are dealing with contaminated distributions or distributions with possible outliers, the Kruskal-Wallis test should always be preferred. However, we have to keep in mind that in case of a small sample size it can be a bit conservative, as it is in fact an asymptotic test. However, for 3 treatment groups even $J = 10$ seems to be sufficient. Nevertheless, for a larger number of treatment groups a larger sample size might be necessary (e.g. for 5 treatment groups the Kruskal-Wallis test is still a bit conservative even for $J = 20$).

When comparing different methods of multiple means comparison, the Tukey's HSD gives clearly the best results. Only for distributions with substantially long tails or with possible outliers, the Wilcoxon tests with the Bonferroni correction may be less conservative. On the contrary, the bootstrap BC_a method seems to be unusable for multiple comparisons unless a sample size is very large (coverage probabilities of bootstrap intervals are only asymptotically accurate and in this simulation study even for $J = 50$ the BC_a method led to liberal results).

Tab. 1: MC estimation of type I errors for one-way ANOVA (non-normality, $I = 3$)

group size	distribution	ANOVA F-test	KW test	corrected t-tests	Tukey's HSD	corrected Wilcoxon	corrected BCa
$J = 5$	normal	0.0480	0.0428*	0.0402*	0.0491	0.0427*	0.2991*
	t(3)	0.0407*	0.0428*	0.0323*	0.0401*	0.0427*	0.3358*
	uniform	0.0540	0.0428*	0.0452*	0.0532	0.0427*	0.2882*
	gamma	0.0483	0.0428*	0.0408*	0.0487	0.0427*	0.2990*
	log-normal	0.0486	0.0428*	0.0411*	0.0487	0.0427*	0.2991*
	skew normal	0.0471	0.0428*	0.0406*	0.0483	0.0427*	0.3054*
	exponential	0.0390*	0.0428*	0.0324*	0.0382*	0.0427*	0.3519*
	contaminated	0.0372*	0.0430*	0.0306*	0.0382*	0.0423*	0.3671*
$J = 10$	normal	0.0522	0.0493	0.0466	0.0535	0.0416*	0.1461*
	t(3)	0.0445*	0.0493	0.0369*	0.0452*	0.0416*	0.2069*
	uniform	0.0550*	0.0493	0.0473	0.0546*	0.0416*	0.1338*
	gamma	0.0526	0.0493	0.0463	0.0539	0.0416*	0.1474*
	log-normal	0.0528	0.0493	0.0458	0.0539	0.0416*	0.1483*
	skew normal	0.0524	0.0493	0.0448*	0.0532	0.0416*	0.1532*
	exponential	0.0442*	0.0493	0.0364*	0.0439*	0.0416*	0.2123*
	contaminated	0.0343*	0.0484	0.0308*	0.0355*	0.0403*	0.3276*
$J = 20$	normal	0.0520	0.0497	0.0450*	0.0507	0.0435*	0.0860*
	t(3)	0.0469	0.0497	0.0405*	0.0465	0.0435*	0.1471*
	uniform	0.0530	0.0497	0.0452*	0.0529	0.0435*	0.0722*
	gamma	0.0524	0.0497	0.0449*	0.0519	0.0435*	0.0876*
	log-normal	0.0525	0.0497	0.0445*	0.0522	0.0435*	0.0885*
	skew normal	0.0521	0.0497	0.0445*	0.0514	0.0435*	0.0910*
	exponential	0.0467	0.0497	0.0406*	0.0483	0.0435*	0.1319*
	contaminated	0.0317*	0.0487	0.0262*	0.0297*	0.0443*	0.3691*
$J = 50$	normal	0.0511	0.0485	0.0444*	0.0510	0.0441*	0.0606*
	t(3)	0.0479	0.0485	0.0419*	0.0483	0.0441*	0.1061*
	uniform	0.0499	0.0485	0.0448*	0.0513	0.0441*	0.0543
	gamma	0.0507	0.0485	0.0440*	0.0513	0.0441*	0.0581*
	log-normal	0.0510	0.0485	0.0440*	0.0508	0.0441*	0.0586*
	skew normal	0.0503	0.0485	0.0432*	0.0488	0.0441*	0.0609*
	exponential	0.0467	0.0485	0.0401*	0.0471	0.0441*	0.0864*
	contaminated	0.0321*	0.0472	0.0266*	0.0338*	0.0418*	0.3373*

Tab. 2: MC estimation of type I errors for one-way ANOVA (non-normality, I = 5)

group size	distribution	ANOVA F-test	KW test	corrected t-tests	Tukey's HSD	corrected Wilcoxon	corrected BCa
J = 5	normal	0.0477	0.0368*	0.0363*	0.0500	0.0637*	0.8749*
	t(3)	0.0384*	0.0368*	0.0300*	0.0433*	0.0637*	0.9018*
	uniform	0.0522	0.0368*	0.0397*	0.0541	0.0637*	0.8663*
	gamma	0.0478	0.0368*	0.0364*	0.0502	0.0637*	0.8729*
	log-normal	0.0470	0.0368*	0.0368*	0.0499	0.0637*	0.8724*
	skew normal	0.0476	0.0368*	0.0358*	0.0493	0.0637*	0.8771*
	exponential	0.0415*	0.0368*	0.0284*	0.0401*	0.0637*	0.9116*
	contaminated	0.0379*	0.0357*	0.0290*	0.0399*	0.0629*	0.8991*
J = 10	normal	0.0477	0.0398*	0.0370*	0.0504	0.0308*	0.4174*
	t(3)	0.0412*	0.0398*	0.0330*	0.0438*	0.0308*	0.6377*
	uniform	0.0497	0.0398*	0.0385*	0.0518	0.0308*	0.3210*
	gamma	0.0480	0.0398*	0.0372*	0.0500	0.0308*	0.4172*
	log-normal	0.0484	0.0398*	0.0367*	0.0503	0.0308*	0.4231*
	skew normal	0.0477	0.0398*	0.0360*	0.0489	0.0308*	0.4482*
	exponential	0.0426*	0.0398*	0.0309*	0.0420*	0.0308*	0.6655*
	contaminated	0.0327*	0.0388*	0.0267*	0.0344*	0.0301*	0.6319*
J = 20	normal	0.0477	0.0457*	0.0371*	0.0475	0.0376*	0.1270*
	t(3)	0.0417*	0.0457*	0.0329*	0.0435*	0.0376*	0.4089*
	uniform	0.0483	0.0457*	0.0383*	0.0474	0.0376*	0.0982*
	gamma	0.0477	0.0457*	0.0370*	0.0478	0.0376*	0.1287*
	log-normal	0.0480	0.0457*	0.0370*	0.0475	0.0376*	0.1291*
	skew normal	0.0456*	0.0457*	0.0367*	0.0474	0.0376*	0.1415*
	exponential	0.0429*	0.0457*	0.0331*	0.0426*	0.0376*	0.3539*
	contaminated	0.0250*	0.0444*	0.0196*	0.0275*	0.0353*	0.6518*
J = 50	normal	0.0524	0.0503	0.0416*	0.0526	0.0422*	0.0682*
	t(3)	0.0482	0.0503	0.0398*	0.0512	0.0422*	0.2462*
	uniform	0.0523	0.0503	0.0435*	0.0529	0.0422*	0.0573*
	gamma	0.0519	0.0503	0.0419*	0.0535	0.0422*	0.0681*
	log-normal	0.0519	0.0503	0.0415*	0.0536	0.0422*	0.0704*
	skew normal	0.0510	0.0503	0.0410*	0.0517	0.0422*	0.0717*
	exponential	0.0476	0.0503	0.0365*	0.0478	0.0422*	0.1219*
	contaminated	0.0331*	0.0500	0.0250*	0.0330*	0.0426*	0.8313*

2.2 The effects of heteroscedasticity

As it was already mentioned, for the purpose of examining the mere effects of heteroscedasticity the data were simulated from a normal distribution with the mean 100 and the variance 100 k_i, i = 1, 2, …, I, i.e. the group variances may not be the same. The constants

k_i are given in tables 3 and 4, which provide estimates of type I errors for the considered tests under various violations of homoscedasticity and for various sample sizes.

Tab. 3: MC estimation of type I errors for one-way ANOVA (heteroscedastity, I = 3)

group size	k	ANOVA F-test	KW test	corrected t-tests	Tukey's HSD	corrected Wilcoxon	corrected BCa
	(1; 1; 1)	0.0509	0.0446*	0.0415*	0.0521	0.0437*	0.3012*
	(0.9; 0.9; 1.2)	0.0553*	0.0478	0.0446*	0.0536	0.0452*	0.2986*
J = 5	(0.75; 0.75; 1.5)	0.0680*	0.0538	0.0554*	0.0653*	0.0503	0.2941*
	(0.8; 1.0; 1.2)	0.0554*	0.0481	0.0461	0.0553*	0.0448*	0.2966*
	(0.5; 1.0; 1.5)	0.0695*	0.0590*	0.0576*	0.0678*	0.0574*	0.2925*
	(1; 1; 1)	0.0513	0.0470	0.0441*	0.0507	0.0416*	0.1459*
	(0.9; 0.9; 1.2)	0.0548*	0.0487	0.0458	0.0540	0.0411*	0.1465*
J = 10	(0.75; 0.75; 1.5)	0.0649*	0.0546*	0.0559*	0.0630*	0.0448*	0.1403*
	(0.8; 1.0; 1.2)	0.0560*	0.0503	0.0477	0.0545*	0.0412*	0.1448*
	(0.5; 1.0; 1.5)	0.0656*	0.0587*	0.0574*	0.0658*	0.0505	0.1394*
	(1; 1; 1)	0.0503	0.0493	0.0446*	0.0500	0.0427*	0.0899*
	(0.9; 0.9; 1.2)	0.0521	0.0507	0.0468	0.0524	0.0438*	0.0921*
J = 20	(0.75; 0.75; 1.5)	0.0606*	0.0559*	0.0546*	0.0593*	0.0473	0.0878*
	(0.8; 1.0; 1.2)	0.0530	0.0514	0.0467	0.0537	0.0441*	0.0912*
	(0.5; 1.0; 1.5)	0.0614*	0.0613*	0.0540	0.0603*	0.0517	0.0838*
	(1; 1; 1)	0.0490	0.0457*	0.0425*	0.0475	0.0402*	0.0585*
	(0.9; 0.9; 1.2)	0.0505	0.0509	0.0447*	0.0513	0.0427*	0.0589*
J = 50	(0.75; 0.75; 1.5)	0.0595*	0.0569*	0.0513	0.0581*	0.0480	0.0549*
	(0.8; 1.0; 1.2)	0.0511	0.0486	0.0462	0.0520	0.0425*	0.0587*
	(0.5; 1.0; 1.5)	0.0595*	0.0587*	0.0517	0.0584*	0.0534	0.0534

The estimated type I errors imply that the ANOVA F-test is liberal (i.e. having a type I error greater than the significance level) under unequal group variances. Also, the type I error seems to be greater in a situation when all group variances differ in comparison with a situation when only one group variance is different from the others (average group variance being the same) – this effect is more pronounced for a larger number of treatment groups I. Moreover, the Kruskal-Wallis test is shown to be a bit less sensitive to the violation of homoscedasticity and for a larger number of treatment groups the effect of homoscedasticity actually counteracts the effect of inherent conservativeness, which is due to asymptotic nature of this nonparametric test.

On the other hand, the multiple comparison methods (except for the bootstrap BC_a method, which is only usable for very large sample sizes) tend to be also more liberal, when the group variances differ. That is why the inherently more conservative Bonferroni correction of the two-sample t-tests or Wilcoxon sum-rank tests may lead to a type I error that is in fact closer to the nominal significance level when compared to the Tukey's HSD method.

Tab. 4: MC estimation of type I errors for one-way ANOVA (heteroscedastity, I = 5)

group size	k	ANOVA F-test	KW test	corrected t-tests	Tukey's HSD	corrected Wilcoxon	corrected BCa
J = 5	(1; 1; 1)	0.0470	0.0348*	0.0336*	0.0478	0.0585*	0.8793*
	(0.95; ...; 0.95; 1.2)	0.0493	0.0367*	0.0368*	0.0508	0.0590*	0.8667*
	(0.875; ...; 0.875; 1.2)	0.0599*	0.0412*	0.0480	0.0633*	0.0619*	0.8593*
	(0.8; 0.9; 1; 1.1; 1.2)	0.0509	0.0378*	0.0389*	0.0527	0.0611*	0.8593*
	(0.5; 0.75; 1; 1.25; 1.5)	0.0663*	0.0452*	0.0554*	0.0694*	0.0736*	0.8447*
J = 10	(1; 1; 1)	0.0511	0.0453*	0.0377*	0.0495	0.0328*	0.4087*
	(0.95; ...; 0.95; 1.2)	0.0523	0.0439*	0.0386*	0.0510	0.0334*	0.4060*
	(0.875; ...; 0.875; 1.2)	0.0635*	0.0463	0.0511	0.0655*	0.0331*	0.4173*
	(0.8; 0.9; 1; 1.1; 1.2)	0.0528	0.0448*	0.0429*	0.0540	0.0327*	0.4120*
	(0.5; 0.75; 1; 1.25; 1.5)	0.0671*	0.0525	0.0581*	0.0698*	0.0383*	0.3907*
J = 20	(1; 1; 1)	0.0516	0.0489	0.0411*	0.0525	0.0384*	0.1021*
	(0.95; ...; 0.95; 1.2)	0.0556*	0.0511	0.0439*	0.0552*	0.0400*	0.1059*
	(0.875; ...; 0.875; 1.2)	0.0645*	0.0534	0.0548*	0.0644*	0.0404*	0.1046*
	(0.8; 0.9; 1; 1.1; 1.2)	0.0570*	0.0512	0.0443*	0.0572*	0.0414*	0.1077*
	(0.5; 0.75; 1; 1.25; 1.5)	0.0696*	0.0581*	0.0615*	0.0728*	0.0480	0.0984*
J = 50	(1; 1; 1)	0.0475	0.0473	0.0373*	0.0471	0.0373*	0.0693*
	(0.95; ...; 0.95; 1.2)	0.0511	0.0490	0.0399*	0.0510	0.0382*	0.0653*
	(0.875; ...; 0.875; 1.2)	0.0624*	0.0533	0.0541	0.0635*	0.0397*	0.0640*
	(0.8; 0.9; 1; 1.1; 1.2)	0.0521	0.0495	0.0437*	0.0523	0.0389*	0.0607*
	(0.5; 0.75; 1; 1.25; 1.5)	0.0674*	0.0590*	0.0607*	0.0733*	0.0437*	0.0573*

Conclusion

The classical one-way ANOVA F-test is derived under the assumptions that the observations are independent, normally distributed and that the group variances are equal. The aim of this article was to examine the effects of non-normality and heteroscedasticity on type I errors of this test, as well as the Kruskal-Wallis test and tests used in a multiple means comparison.

Based on the results of the simulation study it can be concluded that the non-normality has only a small effect on the type I error of the ANOVA F-test. However, when we are dealing with outliers or a substantially skewed or long-tailed distribution, the Kruskal-Wallis test should be preferred. For a post hoc multiple means comparison the classical Tukey's HSD can be recommended, although it is also conservative for contaminated distributions.

On the other hand, in case of unequal group variances both the ANOVA F-test and the Tukey's HSD may be a bit liberal. Although in this situation the Kruskal-Wallis test gives somewhat better results in respect of a type I error, it cannot be considered robust against heteroscedasticity. Nevertheless, unless we are dealing with a substantial difference between group means (e.g. largest standard deviation is larger than twice the smallest standard deviation) these tests perform reasonably well. However, when the difference between group variances is substantial, tests proposed by Welch (1951) or James (1951) might be preferred.

Acknowledgment

This paper was supported by the Internal Grant Agency of the University of Economics in Prague (project IGA 128/2014 "Consequences of assumption violations of classical statistical methods and the possible use of alternative statistical techniques in economic applications").

References

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Day, R. W., & Quinn, G. P. (1989). Comparisons of Treatments After an Analysis of Variance in Ecology. *Ecological Monographs*, *59* (4), 433-463.

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, *56* (293), 52-64.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, *38* (3-4), 324-329.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47* (260), 583-621.

Miller, R. G. (1981). *Simultaneous Statistical Inference*. New York: Springer.

Neményi, P. B. (1963). *Distribution-free Multiple Comparisons* (doctoral dissertation). Princeton University, Princeton, New Jersey.

R core team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Tukey, J. W. (1953). Some selected quick and easy methods of statistical analysis. *Transactions of the New York Academy of Sciences*, *16* (2), 88-97.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, *38* (3-4), 330-336.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1* (6), 80-83.

Contact

Tomáš Marcinko (xmart14@vse.cz)
University of Economics, Prague
nám. W. Churchilla 4, 130 67 Prague, Czech Republic