Resampling-based simultaneous confidence intervals to compare scale using deviances

By: Scott J. Richter & Melinda H. McCann

Richter, S.J. & McCann, M.H. (2023). Resampling-based simultaneous confidence intervals to compare scale using deviances. *Communications in Statistics: Simulation and Computation*, 52(7), 3146-3155. (DOI: https://doi.org//10.1080/03610918.2021.1928194).

Abstract:
Resampling-based confidence intervals based on medians are examined for pairwise comparison of scale. Methods that have been found in the literature to be effective for comparing scale for two groups are extended to the case of all pairwise comparisons, using a Tukey-type adjustment to ensure simultaneous confidence. Coverage rate and interval width estimates were computed using simulated data. A permutation-based method was found to have the best coverage properties.

Keywords: Deviance | Permutation test | Scale parameter

Article:

1. Introduction

Pairwise comparison of scale parameters may be of interest in many areas, including industrial quality control, agricultural production and experimental education (Marozzi 2011). However, it is well known that the parametric F-test for comparing variances of two treatments, as well as parametric tests for more than two treatments (e.g., tests due to Bartlett (1937), Cochran (1937), and Hartley (1950)) are generally not robust to nonnormality (see Sharma and Kibria 2013). Consequently, more robust tests of scale parameters are of interest.

There are many comparative studies of tests for comparing scale differences in the literature. Levene (1960) proposed using the ANOVA F-test on the absolute deviations from the mean. Brown and Forsythe (1974) proposed instead using absolute deviations from the median, which they referred to as the *W50* test. While no uniformly best test for scale has been demonstrated in the literature, the *W50* test has been recommended as showing good overall performance with respect to power and robustness to nonnormality in several comparative studies, including Keselman, Gaines, and Clinch (1979), Conover, Johnson, and Johnson (1981), and Balakrishnan and Ma (1990). O'Brien (1979) proposed a modification of Levene's test (OB50) which has been recommended over the *W50* test for lighter-tailed distributions (O'Brien 1979, Olejnik and Algina 1988). Marozzi (2011) considered the *W50* test, as well as resampling-based versions, and found that the resampling-based versions tended to be more robust and have higher power. They recommended the permutation *W50* test as a computationally simple robust test, although a slightly modified version of *W50* (*L50*, Pan (1999) was found to often have slightly higher power. Marozzi (2012a) subsequently proposed a combined version of *OB50* and *L50* that, although it was not always most powerful, showed good overall performance. Marozzi (2012b) also showed that a modification to the adaptive test of Hall and Padmanabhan (1997) improved its performance. Richter and McCann (2017) found that the permutation *RMD* test due to Higgins (2004), also based on absolute deviations from the median, was generally superior to *W50*, especially for heavier-tailed distributions.

In this paper, we extend the resampling-based versions of the *RMD* test to the problem of simultaneous pairwise confidence intervals. The method of Richter and McCann (2007) is used to control the simultaneous confidence level, and estimated coverage rates and interval length are compared.

2. Simultaneous confidence intervals using permutation tests

Consider a one-way layout with t treatments and n_i observations per treatment. We assume a location-scale model, $y_{ij}=\mu_i+\sigma_i e_{ij}, i=1,\ldots,t, j=1,\ldots,n_i$, where μ_i and σ_i are the location and scale parameters, respectively, of treatment i, and εi_j are independent and identically distributed with median 0. It is desired to construct simultaneous confidence intervals for $\frac{\sigma i}{\sigma i'}$ for all $i\neq i'$.

2.1 *Individual confidence intervals* based on inverting the permutation distribution

Consider first the case where k=2. We will use $log(\frac{\widehat{\sigma 1}}{\hat{\sigma}2})$ as the test statistic to carry out a permutation test. A confidence interval for the pairwise scale ratio, $\frac{\sigma 1}{\sigma 2}$, can be constructed as follows. Find two constants, d_l and d_u that, when subtracted from the observed log-ratio, lead to the smallest p-value of the permutation test greater than or equal to $\alpha/2$. Then the $100(1-\alpha)$ % confidence interval for $log(\frac{\sigma 1}{\sigma 2})$ is $d_l \leq log\ (\frac{\sigma 1}{\sigma 2}) \leq du$. That this is indeed a $100(1-\alpha)$ % confidence interval for $log(\frac{\sigma 1}{\sigma 2})$ is verified by Good (2000, p.210).

The values d_l and d_u can be found by first finding the percentiles of the permutation distribution, and then determining dl and du as $dl=log(\frac{\hat{\sigma}1}{\hat{\sigma}2})-p_{1}-\alpha/2$ and $du=log(\frac{\hat{\sigma}1}{\hat{\sigma}2})+p_{1}-\alpha/2$, where $log(\frac{\hat{\sigma}1}{\hat{\sigma}2})$ is the observed log-ratio. Then exp (d_l) and exp (d_u) are the endpoints for the $100(1-\alpha)$% confidence interval for $\frac{\sigma 1}{\sigma 2}$.

2.2. *Simultaneous confidence intervals when* $\boldsymbol{k}$ >2

Richter and McCann (2017) proposed a restricted permutation method to provide simultaneous confidence for pairwise comparison of location parameters. This can be extended to the present case of computing intervals for scale parameters as follows.
First, construct the reference distribution, consisting of the maximum of the pairwise log-ratios, $max_{1\leq i<i'\leq k}\ [log(\frac{\hat{\sigma}i}{\hat{\sigma}i'})]$, across all pairs, for each permutation. Next, the $1-\alpha/2$ percentile of the reference distribution, $p(max)_{1-\alpha/2}$, is determined. Then for each pairwise difference, the constants $d(max)_l$ and $d(max)_u$ are found

as $d(max)_l = \log(\frac{\hat{\sigma}i}{\hat{\sigma}i'}) - p(max)1-\alpha/2$ and $d(max)_u = \log(\frac{\hat{\sigma}i}{\hat{\sigma}i'}) + p(max)_{1-\alpha/2}$, where $log(\frac{\hat{\sigma}i}{\hat{\sigma}i'})$ is the observed log-ratio for comparing groups i and i'.

Richter and McCann (2007) proved that for the case of testing all pairwise comparisons, using the distribution of the absolute maximum median difference across all pairs, where randomization is performed within each pair, provides strong control of the familywise error rate (FWER). Since the confidence interval procedure is found by inverting the test procedure, the same percentiles used for the testing procedure will be used to construct the confidence intervals. Thus, since the acceptance region of the testing procedure is used to construct the confidence intervals, the simultaneous confidence level must also be controlled. This follows from Good (2000; Theorem 4.1, p.210) when we make the appropriate generalization to a set of tests with their corresponding confidence intervals, and consider simultaneous instead of individual coverage. Specifically, let $A(\boldsymbol{\Delta})$, $\boldsymbol{\Delta}=\{\Delta ij, i,j=1,\ldots,k, i<j\}$ be the set of all values in the sample space where the simultaneous testing procedure of Richter and McCann (2007), detailed in Sec. 2.1, would not be rejected if the values Δij were simultaneously tested. Now let $\boldsymbol{S}(\boldsymbol{X})$ be the set of simultaneous confidence intervals described above that result from a specific $\boldsymbol{X}$ in the sample space. Note that a vector $\boldsymbol{\Delta}$ of Δij values will be included in $\boldsymbol{S}(\boldsymbol{X})$ if and only if this $\boldsymbol{\Delta}$ would not be rejected by the simultaneous procedure described above. Consequently, $P(\boldsymbol{\Delta}\in \boldsymbol{S}(\boldsymbol{X}))=P(\boldsymbol{X}\in A(\boldsymbol{\Delta}))\geq 1-\alpha$, and thus, simultaneous coverage at the $1-\alpha$ level is guaranteed.

3. Simultaneous bootstrap confidence intervals

Consider a one-way layout with t treatments and ni observations per treatment. We assume a location-scale model, $\boldsymbol{y_{ij}=\mu_i+\sigma_i e_{ij}, i=1,\ldots,t, j=1,\ldots,n_i}$, where μ_i and σ_i are the location and scale parameters, respectively, of treatment i, and ε_{ij} are independent and identically distributed with median 0. It is desired to construct simultaneous confidence intervals for $\frac{\sigma i}{\sigma i'}$ for all $i\neq i'$. The estimator for σi will be based on absolute deviances, or the absolute deviations $dev_{ij}=|y_{ij}-median_i|$, where $median_i$ is the sample median for the i^{th} treatment. We define the estimator as the mean of the absolute deviances: $\boldsymbol{\hat{\sigma}i=\frac{1}{ni}\sum^{ni}{}_{j=1} dev_{ij}}$.

3.1 Individual confidence intervals

Intervals for $\frac{\sigma 1}{\sigma 2}$ can be constructed directly using bootstrap sampling, where independent bootstrap samples of size ni and ni' are selected from the combined observed samples at each iteration. The percentile method uses as endpoints the quantiles $b_{\alpha/2}$ and $b_1-\alpha/2$ of the bootstrap distribution of $\frac{\widehat{\sigma i}}{\widehat{\sigma i'}}$. Alternatively, a bootstrap interval estimator is proposed based on the pivotal quantity, $FB_{i,i'}=\frac{\widehat{\sigma i}/\widehat{\sigma i'}}{\widehat{\sigma i}/\widehat{\sigma i'}}$. The $1-\alpha$ level interval estimator is then $(\frac{\widehat{\sigma i}/\widehat{\sigma i'}}{FB1-\alpha/2}$, $\frac{\widehat{\sigma i}/\widehat{\sigma i'}}{FB\alpha/2})$, where $FB_1-\alpha/2$ is the $1-\alpha/2$ quantile of the bootstrap distribution.

3.2. *Simultaneous confidence intervals when* $\boldsymbol{k}$ >2

For the percentile method outlined in Sec. 3.1, a Bonferroni correction may be used to achieve simultaneous confidence.

For the estimator based on the pivotal quantity $FB_{i,i'} = \frac{\widehat{\sigma_i} / \widehat{\sigma_{i'}}}{\sigma_i / \sigma_{i'}}$ described in Sec. 3.1, we again use the method of Richter and McCann (2017) from Section 2.2. Independent bootstrap samples are selected from the k observed samples at each iteration, and the estimated scale ratio, $\frac{\widehat{\sigma_i}}{\widehat{\sigma_{i'}}}$, calculated for all $i \neq i'$. Finally the maximum, $max_{\, i \neq i'}(\frac{\widehat{\sigma_i}}{\widehat{\sigma_{i'}}})$is calculated, and these values are used to construct the bootstrap distribution of $max_{i \neq i'}$ $[FB_{i,i'}]$. Let $maxFB_{\alpha/2}$ be the α/2 quantile of the bootstrap distribution. Then the 1−α level confidence interval for $\frac{\sigma_i}{\sigma_{i'}}$ is given by $(\frac{\widehat{\sigma_i} / \widehat{\sigma_{i'}}}{maxFB_{1-\alpha/2}}, \frac{\widehat{\sigma_i} / \widehat{\sigma_{i'}}}{maxFB_{\alpha/2}})$. The set of $k(k-1)/2$ intervals will have approximate simultaneous confidence level 1−α.

4. Simulation

4.1. Simulation details

A simulation was conducted to compare estimated power and FWER for the procedures described in Sec. 2, controlling FWER using both the technique of Richter & McCann as well as a Bonferroni correction.

The methods compared were:

- FP_{MAX}: The procedure based on the permutation distribution of the absolute maximum median difference across all pairs (Sec. 2.2)
- FP_{BON}: The procedure based on the permutation distribution of the absolute median difference using a Bonferroni correction (Sec. 3.2)
- FB_{MAX}: Simultaneous bootstrap intervals using the maximum *F*-pivot statistic across all pairs (Sec. 4.2)
- FB_{BON}: Simultaneous bootstrap intervals using the F-pivot statistic and a Bonferroni correction
- B_{BON}: Simultaneous bootstrap intervals using the sample scale ratio statistic and a Bonferroni correction
- F_{BON}: Simultaneous parametric F-ratio intervals using a Bonferroni correction

Several combinations of sample sizes were chosen to investigate large, small, equal and unequal sizes. Settings at both five and nine groups were examined. Several different distributions, chosen to represent both symmetric and skewed distributions with lighter and heavier tails, were used to generate data.

1. Normal. The normal distribution has pdf $f(y|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\,[-\frac{y^2}{2\sigma^2}]$, where σ is the scale parameter. The normal distribution has skewness 0 and kurtosis excess of 0.

2. Laplace. The pdf is $f(y|\sigma) = \frac{1}{2\sigma} \exp\,[-\frac{|y|}{2\sigma}]$, where again σ is the scale parameter. The Laplace distribution is symmetric (skewness 0), but will have heavier tails than a normal distribution (kurtosis excess 3) with the same value of the scale parameter σ, as the standard deviation is $\sqrt{2\sigma}$.

3. Nonstandardized t (μ=0, ν=2). The pdf is $f(x|\nu,\sigma)=)\frac{\Gamma(\frac{\nu+1}{2}}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma}}(1+\frac{1}{\nu}(\frac{x}{\sigma})^2)^{-\frac{\nu+1}{2}}$, where σ is the scale parameter. This distribution is also symmetric, but will have heavier tails than the Laplace distribution.

4. Exponential. The pdf is $f(y|\sigma)=\frac{1}{\sigma}\exp(-\frac{y}{\sigma})$, where σ is the scale parameter. The exponential distribution is skewed (skewness coefficient 2) and has a lighter right tail than the lognormal distribution (kurtosis excess 6).

5. Lognormal (μ=0, σ^2 =0.49). The pdf is $f(y|\sigma)=\frac{1}{\sigma\sqrt{2\pi y}}e^{\frac{-(lny)2}{2\sigma 2}}$. This distribution is more skewed (skewness 2.9) and has heavier tails than the exponential distribution (kurtosis excess 18).

Marozzi (2016) suggested that at least 253 random permutations are necessary with 1000 random data sets if the goal of the simulation is to estimate the power of a test and only a "rough" estimate of the permutation p-value is required, while Keller-McNulty and Higgins (1987) recommended a random sample of at least 1600 permutations to estimate the exact p-value for a permutation test. Since precise estimation of the permutation distribution was considered important, a slightly conservative 2,000 random permutations were utilized.

4.2. Simulation Results

Representative simulation results are presented in this section. Tables 1–8 show estimated coverage of 95% simultaneous intervals, and Tables 9–12 show estimated mean interval lengths. Additional results are available in the Supplementary Material.

Table 1. Simultaneous coverage, σ_i = (4,2,1,1,1); n_i = 10.

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.977	0.947	0.950	0.950	0.925
FP_{BON}	0.948	0.915	0.931	0.851	0.802
FB_{MAX}	0.985	0.930	0.943	0.951	0.906
FB_{BON}	0.987	0.931	0.942	0.918	0.810
B_{BON}	0.987	0.927	0.953	0.955	0.920
F_{BON}	0.973	0.752	0.636	0.601	0.287

Table 2. Simultaneous coverage, 9 groups, σ_i = (4,2,1,1,1,1,1,1,1); n_i = 10.

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.979	0.970	0.952	0.960	0.929
FP_{BON}	0.919	0.907	0.906	0.815	0.740
FB_{MAX}	0.985	0.934	0.961	0.955	0.917
FB_{BON}	0.977	0.940	0.937	0.870	0.704
B_{BON}	0.980	0.923	0.947	0.937	0.906

F_{BON}	0.969	0.506	0.540	0.497	0.112

Table 3. Simultaneous coverage, σ_i = (4,2,1,1,1); n_i = 100.

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.986	0.983	0.956	0.973	0.952
FP_{BON}	0.978	0.973	0.949	0.958	0.950
FB_{MAX}	0.953	0.940	0.947	0.949	0.941
FB_{BON}	0.960	0.942	0.922	0.910	0.910
B_{BON}	0.964	0.942	0.913	0.926	0.947
F_{BON}	0.958	0.622	0.081	0.230	0.652

Table 4. Simultaneous coverage, 9 groups, σ_i = (4,2,1,1,1,1,1,1,1,1); n_i= 100.

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.988	0.952	0.956	0.973	0.945
FP_{BON}	0.965	0.958	0.968	0.968	0.919
FB_{MAX}	0.960	0.941	0.947	0.949	0.939
FB_{BON}	0.957	0.897	0.922	0.910	0.861
B_{BON}	0.945	0.808	0.913	0.926	0.908
F_{BON}	0.965	0.637	0.081	0.953	0.005

Table 5. Simultaneous coverage, σ_i = (4,2,1,1,1); n_i = (10,20,20,30,30).

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.975	0.969	0.949	0.946	0.936
FP_{BON}	0.969	0.958	0.939	0.899	0.850
FB_{MAX}	0.972	0.959	0.919	0.920	0.857
FB_{BON}	0.958	0.945	0.904	0.875	0.793
B_{BON}	0.911	0.872	0.855	0.815	0.790
F_{BON}	0.974	0.751	0.541	0.548	0.154

4.2.1. Simultaneous coverage

FP_{MAX} always showed coverage at least at the nominal level and was the only method to do so. All methods generally provided adequate coverage with normally distributed data, but all methods besides FP_{MAX} showed less than nominal coverage in some cases for nonnormal data, especially with small to moderate sample sizes and with more groups. In particular, FP_{MAX} was the only method to maintain adequate coverage with unequal, small to moderate sample sizes (Tables 5,6), with the exception of the lognormal distribution with 9 groups (Table 6), although it still had the best coverage, and coverage became adequate as sample sizes increased. FB_{MAX} showed adequate coverage, except with lognormal data where the coverage tended to be below nominal level. FB_{BON} and B_{BON} generally had inadequate coverage for nonormal distributions. F_{BON} had unacceptable coverage for all nonnormal cases.

The coverage advantage of FP_{MAX} over FP_{BON} and FB_{MAX} over FB_{BON} increased as the number of groups, and thus the number of pairwise comparisons increased. This may seem counterintuitive, since for theoretical continuous sampling distributions typically used, the interval width for the Bonferroni-adjusted intervals can increase without bound as the number of comparisons increases. However, this is not the case for permutation and bootstrap distributions, and thus while the number of comparisons increases there is essentially a ceiling for the interval width, and thus it can happen that the coverage decreases.

Table 6. Simultaneous coverage, 9 groups, σ_i= (4,2,1,1,1,1,1,1,1); n_i= (10,10,10,20,20,20,30,30,30).

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.978	0.952	0.945	0.948	0.901
FP_{BON}	0.948	0.928	0.908	0.872	0.821
FB_{MAX}	0.973	0.941	0.947	0.933	0.874
FB_{BON}	0.934	0.897	0.881	0.842	0.700
B_{BON}	0.870	0.808	0.806	0.772	0.733
F_{BON}	0.947	0.637	0.340	0.395	0.057

Table 7. Simultaneous coverage, σ_i= (4,2,1,1,1); n_i= (120,120,90,90,60).

Distribution Statistic	Normal	Laplace	Nonst-t	Exponential	Lognormal
FP_{MAX}	0.985	0.988	0.966	0.973	0.954
FP_{BON}	0.977	0.976	0.969	0.968	0.939
FB_{MAX}	0.956	0.956	0.933	0.922	0.928
FB_{BON}	0.948	0.961	0.942	0.929	0.904
B_{BON}	0.948	0.952	0.937	0.929	0.922
F_{BON}	0.961	0.638	0.275	0.428	0.062

Table 8. Simultaneous coverage, 9 groups, σ_i= (4,2,1,1,1,1,1,1,1); n_i= (60,60,60,90,90,90,120,120,120)

Distribution Statistic	*Normal*	*Laplace*	*Nonst-t*	*Exponential*	*Lognormal*
FP_{MAX}	0.992	0.974	0.969	0.973	0.942
FB_{MAX}	0.950	0.947	0.950	0.927	0.926
FB_{BON}	0.946	0.927	0.934	0.884	0.852
B_{BON}	0.938	0.914	0.939	0.897	0.885
F_{BON}	0.967	0.504	0.093	0.203	0.009

Table 9. Minimum, median and maximum estimated mean interval lengths σ_i= (4,2,1,1,1); n_i = 10.

Statistic	FP_{MAX}	FP_{BON}	FB_{MAX}	FB_{BON}	B_{BON}	F_{BON}
			Normal			
Minimum	2.90	2.03	2.90	3.03	3.30	2.76
Median	6.74	7.68	7.20	8.21	8.36	5.87

Maximum	22.73	35.40	19.84	20.74	23.35	15.73
			Laplace			
Minimum	2.26	2.34	3.00	3.22	3.41	2.68
Median	9.21	10.47	9.29	10.38	10.62	6.30
Maximum	41.62	60.23	39.05	47.93	51.40	17.00
			Nonstandardized-t			
Minimum	2.62	2.88	3.08	3.37	3.91	2.74
Median	8.77	9.80	9.30	10.06	10.41	6.13
Maximum	132.35	177.09	94.66	116.70	60.64	45.82
			Exponential			
Minimum	2.77	2.63	3.91	4.03	4.15	2.19
Median	8.77	9.80	9.30	10.06	10.41	6.13
Maximum	132.35	177.09	94.66	116.70	60.64	45.82
			Lognormal			
Minimum	2.00	2.04	3.60	3.57	4.74	1.75
Median	14.43	14.45	16.00	17.21	16.34	6.78
Maximum	768.08	957.26	437.50	644.00	116.85	88.66

Table 10. Minimum, median and maximum estimated mean interval lengths, 9 groups, σ_i= (4,2,1,1,1,1,1,1,1); n_i = 10.

Statistic	FP_{MAX}	FP_{BON}	FB_{MAX}	FB_{BON}	B_{BON}	F_{BON}
			Normal			
Minimum	2.92	2.99	4.27	4.57	4.71	3.75
Median	6.54	7.04	7.33	9.17	9.37	5.78
Maximum	15.65	19.66	16.28	44.29	44.75	10.03
			Laplace			
Minimum	3.77	3.77	4.54	5.39	5.73	3.33
Median	8.99	9.58	9.85	11.55	12.03	6.17
Maximum	27.11	31.32	24.46	35.36	33.00	11.71
			Nonstandardized-t			
Minimum	3.29	3.44	4.32	4.83	4.97	3.40
Median	8.95	9.26	9.91	11.89	12.01	6.20
Maximum	264.24	314.13	184.72	324.53	143.06	46.26
			Exponential			
Minimum	3.66	3.27	5.07	5.42	5.86	3.12
Median	10.65	10.22	12.33	14.91	14.73	6.24
Maximum	62.45	65.23	35.35	136.29	84.13	15.57
			Lognormal			
Minimum	4.91	4.16	6.65	6.84	7.51	2.89
Median	16.46	14.70	19.72	22.08	19.96	7.11
Maximum	315.75	352.23	273.00	545.25	193.62	35.18

Table 11. Minimum, median and maximum estimated mean interval lengths, σ_i= (4,2,1,1,1); n_i= (10,20,20,30,30).

Statistic	FP_{MAX}	FP_{BON}	FB_{MAX}	FB_{BON}	B_{BON}	F_{BON}
			Normal			

Minimum	2.38	2.23	2.68	2.59	2.59	2.02
Median	5.46	6.36	5.27	5.89	5.30	4.49
Maximum	13.37	16.26	11.98	22.13	13.99	7.93
			Laplace			
Minimum	2.54	2.28	2.53	2.36	2.30	1.66
Median	6.68	7.73	6.28	7.00	6.36	4.41
Maximum	27.80	28.47	36.11	36.44	22.20	11.19
			Nonstandardized-t			
Minimum	2.40	2.22	2.81	2.50	2.32	1.67
Median	6.79	7.68	6.49	7.08	6.60	4.49
Maximum	80.81	88.52	67.61	92.03	25.83	20.98
			Exponential			
Minimum	2.82	2.26	2.31	2.54	2.75	1.52
Median	7.75	9.00	7.12	7.94	7.41	4.43
Maximum	34.4	63.04	37.31	64.32	28.26	12.39
			Lognormal			
Minimum	2.95	2.40	2.77	2.61	3.07	1.37
Median	10.44	11.52	9.77	10.40	10.08	4.66
Maximum	212.59	311.72	174.50	265.27	67.90	33.16

4.2.2. Interval width

As might be expected from the coverage results, FP_{MAX} tended to produce wider intervals, although the median interval width was usually not substantially larger. In particular, the width of FP_{MAX} was noticeably larger for the case of unequal small to moderate sample sizes, which was the case where all of the other methods had inadequate coverage.

Table 12. Minimum, median and maximum estimated mean interval lengths, 9 groups, σ_i= (4,2,1,1,1,1,1,1,1); n_i= (10,10,10,20,20,20,30,30,30).

Statistic	FP_{MAX}	FP_{BON}	FB_{MAX}	FB_{BON}	B_{BON}	F_{BON}
			Normal			
Minimum	2.56	2.72	3.27	3.45	2.80	2.64
Median	4.84	5.55	5.10	6.17	4.81	4.20
Maximum	11.12	12.89	10.44	63.90	16.57	6.19
			Laplace			
Minimum	2.33	2.54	3.29	3.65	3.02	2.10
Median	6.22	7.36	6.58	7.62	6.47	4.27
Maximum	21.74	24.00	21.12	41.31	20.83	8.43
			Nonstandardized-t			
Minimum	2.98	3.00	3.66	3.70	3.05	2.37
Median	6.18	6.90	6.71	7.59	6.09	4.28
Maximum	26.69	39.39	38.80	44.43	29.72	13.87
			Exponential			
Minimum	2.99	3.27	4.54	3.77	3.54	2.20
Median	6.18	6.90	6.71	7.59	6.09	4.28
Maximum	32.38	36.99	30.05	103.97	37.85	12.10
			Lognormal			

Minimum	3.19	3.49	3.97	4.43	3.67	2.17
Median	9.95	11.05	11.70	12.99	9.64	4.58
Maximum	106.53	92.06	256.38	208.55	43.75	19.27

5. Example

Stejskal, Aulicky, and Pekar (2009) was interested in variance differences of microcapsule size for five different formulations of pesticide. The authors were concerned about nonnormal distributions (See Figure 1) and thus used a permutation version of Hartley's *F*-max test as an omnibus test of equal variances. Pairwise differences between groups based on sample variances were discussed but not analyzed statistically. We reanalyze these data to produce simultaneous intervals for all pairwise comparisons of the formulations using the permutation-based procedure (FP_{MAX}). We found statistical evidence that formulations 1, 2 and 4 have higher variability than formulations 3 and 5. Formulation 1 was estimated to have scale parameter at least 30–31% higher than Formulations 3 and 5, while Formulation 2 at least 14–15% higher, and Formulation 4 at least about 8–10% higher, with simultaneous 95% confidence (See Table 13).

Figure 1. Distributions of microcapsule size for five different formulations of pesticide

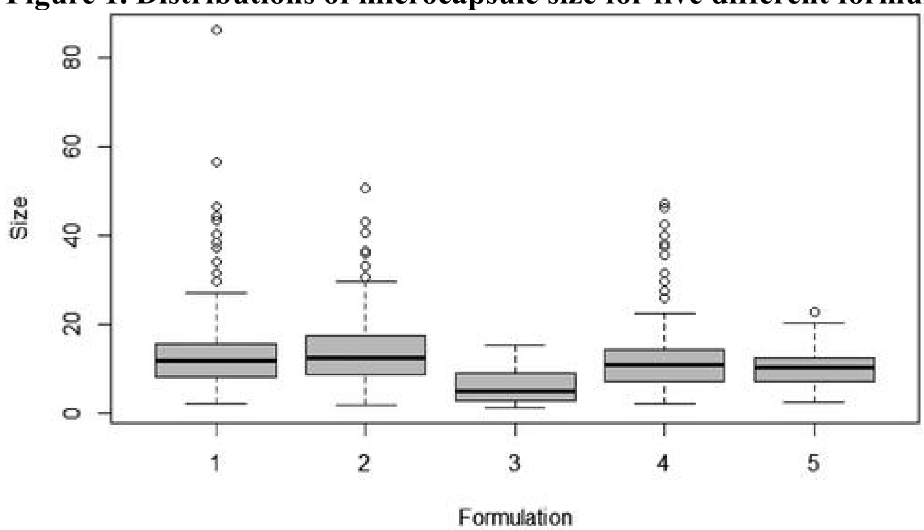

Table 13. Observed mean deviance ratios and simultaneous 95% confidence intervals for scale ratios using FP_{MAX}.

		Ratio (95% confidence interval)			
Formulation	Mean deviance	Formulation 2	Formulation 3	Formulation 4	Formulation 5
1	6.95	1.14 (0.70,1.86)	2.14 (1.31, 3.50)	1.20 (0.74, 1.96)	2.12 (1.30, 3.46)
2	6.08		1.88 (1.15, 3.06)	1.05 (0.64, 1.71)	1.85 (1.14, 3.02)
3	3.24			0.56 (0.34, 0.91)	0.99 (0.61,1.61)
4	5.78				1.76 (1.08, 2.88)
5	3.28				

6. Discussion

Resampling-based procedures for estimating scale parameters provide an alternative to other parametric and semiparametric procedures, most of which rely on approximate standard errors and large sample approximations. The permutation-based procedure presented here provides a distribution-free procedure for estimating pairwise differences between scale parameters, controlling the familywise simultaneous confidence level. It provides a viable alternative to parametric procedures that perform poorly under nonnormality.

Since the coverage of FP_{MAX} was sometimes conservative, especially with fewer groups and smaller sample sizes, ways to decrease the interval width will be a subject of future research. Since the permutation distribution can be highly discrete in these situations, an idea similar to the mid-*p*-value used for permutation tests will be investigated.

References

Balakrishnan, N., and C. W. Ma. 1990. A comparative study of various tests for the equality of two population variances. *Journal of Statistical Computation and Simulation* 35 (1–2):41–89. https://doi.org/10.1080/00949659008811234.

Bartlett, 1937. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London* 160:268–82.

Brown, M. B., and A. B. Forsythe. 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association* 69 (346):364–7. doi:10.1080/01621459.1974.10482955.

Cochran, W. G. 1937. Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society* 4:102–18.

Conover, W. J., M. E. Johnson, and M. M. Johnson. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23 (4):351–61. doi:10.1080/00401706.1981.10487680.
Good, P. 2000. *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York:Springer-Verlag.

Hall, P., and A. R. Padmanabhan. 1997. Adaptive inference for the two-sample scale problem. *Technometrics* 39(4):412–22. doi:10.1080/00401706.1997.10485160.

Hartley, H. O. 1950. The use of range in analysis of variance. *Biometrika* 37:271–80. doi:10.2307/2332380.

Higgins, J. J. 2004. *Introduction to modern nonparametric statistics*. Pacific Grove, CA: Duxbury.

Keller-McNulty, S., and J. J. Higgins. 1987. Effect of tail weight and outliers on power and Type-I error of Robust permutation tests for location. *Communications in Statistics - Simulation and Computation* 16 (1):17–36. doi:10.1080/03610918708812575.

Keselman, H. J., P. A. Gaines, and J. J. Clinch. 1979. Tests for homogeneity of variance. *Communications in Statistics - Simulation and Computation* 8 (2):113–9. doi:10.1080/03610917908812108.

Levene, H. 1960. Robust tests for equality of variances. In *Contributions to Probability and Statistics*, ed. I. Olkinand H. Hotelling, 278–92. Palo Alto, CA: Stanford University Press.

Marozzi, M. 2011. Levene type tests for the ratio of two scales. *Journal of Statistical Computation and Simulation* 81 (7):815–26. https://doi.org/10.1080/00949650903499321.

Marozzi, M. 2012a. A distribution free test for the equality of scales. *Communications in Statistics - Simulation and Computation* 41 (6):878–89. doi:10.1080/03610918.2012.625335.

Marozzi, M. 2012b. A modified hall-Padmanabhan Test for the homogeneity of scales. *Communications in Statistics - Theory and Methods* 41 (16–17):3068–78. doi:10.1080/03610926.2011.606483.

Marozzi, M. 2016. Multivariate tests based on interpoint distances with application to magnetic resonance imaging. *Statistical Methods in Medical Research* 25 (6):2593–610. doi:10.1177/0962280214529104.
O'Brien, R. G. 1979. A general ANOVA method for robust tests of additive models for variances. *Journal of the American Statistical Association* 74:877–80.

Olejnik, S. F., and J. Algina. 1988. Tests of variance equality when distributions differ in form and location. *Educational and Psychological Measurement* 48 (2):317–29. doi:10.1177/0013164488482005.

Pan, G. 1999. On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation* 63 (1):59–71. doi:10.1080/00949659908811949.

Richter, S. J., and M. H. McCann. 2007. Multiple comparisons using medians and permutation tests. *Journal of Modern Applied Statistical Methods* 6 (2):399–412. doi:10.22237/jmasm/1193889900.

Richter, S. J., and M. H. McCann. 2017. Permutation tests of scale using deviances. *Communications in Statistics -Simulation and Computation* 46 (7):5553–65. doi:10.1080/03610918.2016.1165844.

Sharma, D., and B. M. Kibria. 2013. On some test statistics for testing homogeneity of variances: A comparative study. *Journal of Statistical Computation and Simulation* 83 (10):1944–63. doi:10.1080/00949655.2012.675336.

Stejskal, V., R. Aulicky, and S. Pekar. 2009. Brief exposure of Blattella germanica (Blattodea) to insecticides formulated in various microcapsule sizes and applied on porous and non-porous surfaces. *Pest Management Science* 65 (1):93–8. doi:10.1002/ps.1651.