

Changes in Variance and the Detection of Trends

Markus Neuhäuser 

Department of Mathematics, Informatics, and Technology, RheinAhrCampus, Koblenz University of Applied Sciences, Joseph-Rovan-Allee 2, 53424 Remagen, Germany; neuhaeuser@hs-koblenz.de

Abstract

Background: Tests for a trend in location are appropriate when there is an ordered alternative such as, for example, when it is assumed that the effect does not decrease with increasing doses of a drug or fertilizer. Classical trend tests for normally distributed data as well as the nonparametric Jonckheere trend test can have inflated type I error rates when variances differ between groups. Here, different approaches suggested to handle heterogeneous variances are investigated in combination with the Williams trend test. **Methods:** A simulation study was performed to compare the Jonckheere trend test with competing tests. The different tests were investigated for normal and non-normal data and also applied to a data set on sizes of walnuts opened by birds in various stages of a winter. **Results:** With one exception, all investigated trend tests can have an inflated type I error rate when variances differ. Only a nonparametric multiple contrast test based on relative effects showed an acceptable type I error rate in all scenarios considered in the simulation. **Conclusions:** The Williams trend test in combination with the nonparametric multiple contrast test based on relative effects can be suggested for routine use. With this procedure, an increase in variance cannot cause a significant result in the test for trend.

Keywords: test for trend; heteroscedasticity; nonparametric tests; Jonckheere test; Williams test

1. Introduction

Many statistical methods such as the one-way analysis of variance (ANOVA) assume homoscedasticity, that is, homogeneity of variances. Thus, although the variances of the different groups are unknown it is assumed that they are equal. However, in applications variances often differ [1]. For example, the placebo group in clinical trials might have a smaller variance [2]. Ogenstad [3] (p. 497) wrote that “the assumption of homoscedasticity ... is usually made for simplicity and mathematical ease rather than anything else.”

In this paper, I consider trend tests which are appropriate when more than two groups are compared and the alternative is constrained because it can be assumed that an effect does not decrease with increasing doses. For instance, this might be sensible when increasing doses of a drug or fertilizer are investigated. The dose levels are treated as a discrete variable. Then, an ANOVA approach is possible and selecting a specific model is not required [1].

Several trend tests were proposed in the literature, such as the tests introduced by Bartholomew [4], Williams [5], and Jonckheere [6]. Shorack [7] extended Bartholomew’s test to other designs such as the two-way layout, including a rank analog for the nonparametric case. Approaches using permutation combination-based tests were also developed [8]; for further details and extensions to the multivariate case, the reader is referred to Basso et al. [9].



Academic Editor: Wei Zhu

Received: 18 May 2026

Revised: 18 June 2026

Accepted: 22 June 2026

Published: 24 June 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

When variances differ, trends may arise as a consequence of a change in variance. Gould [10] (p. 33) wrote: “apparent trends can be generated as by-products, or side consequences, of expansions and contractions in the amount of variation.” Gould [10,11] presented two examples: lineages increase in body size over evolutionary time (according to Cope’s rule) as a consequence of increased species diversity; and in baseball the disappearance of the 0.4 hitting was caused by a decrease in standard deviations, whereas the mean batting average was constant over 100 years.

Empirical studies often show increases in variance. For example, Bracht et al. [12] investigated patients with hepatitis C in five different fibrosis stages. The biomarker MFAP4 was measured (U/mL), and its mean increased from 9.23 in the first stage F0 to 24.62 in the fifth stage F4; the corresponding standard deviations increased from 6.92 to 15.83.

In another study, the number of erythrocytes in female rats treated with six different doses of sodium dichromate dihydrate was reported [13]. Mean $\pm$ standard deviations were 8.30 ± 0.18 for the zero dose, 8.60 ± 0.15 for the lowest non-zero dose, and 9.62 ± 0.29 for the highest dose. In both examples, a distinct increase in variability was observed.

In the case of heteroscedasticity, classical trend tests for normally distributed data such as Bartholomew’s test and multiple contrast tests, but also the nonparametric Jonckheere trend test, can have inflated type I error rates [14,15]. This anti-conservativeness of trend tests might be a statistical explanation for the above-mentioned phenomenon that trends might be generated as by-products of changes in variance.

When there is some association between means and variances it might be sensible to consider the coefficient of variation (CV). The CV is defined as the ratio of the standard deviation divided by the mean, or the absolute value of the mean in the presence of negative data. However, it is usually required that all observations of the considered sample are nonnegative and the mean is positive [16]. In some applications it is meaningful to consider the CV. For example, in crop yield data the CV often decreases with increasing means according to a power-law relationship, and therefore the CV is used when testing for crop yield stability [17].

However, here tests for a trend in location are investigated. In this case, analogous to the Behrens-Fisher problem, it is tested whether the location parameters of different groups differ and show a trend, irrespective of a possible difference in variances. Unequal variances could even occur in the absence of a treatment effect. Consequently, in this case, it is useful to test for differences in location parameters while adjusting for a possible heteroscedasticity, as in the Welch t test.

In the following sections different approaches suggested to handle heterogeneous variances shall be investigated in combination with trend tests. The competing tests are compared in a simulation study and are based on a real data set with heterogeneous variances.

2. Materials and Methods

Here, tests for a trend in location are investigated. The outcome variable in the k groups is assumed to be continuous. Observations within groups are assumed to be independent and identically distributed, and mutual independence between samples is assumed. The location parameter in group i is denoted by ϑ_i , $i = 1, \dots, k$. The null hypothesis is $H_0: \vartheta_1 = \dots = \vartheta_k$, whereas the alternative is constrained and one-sided $H_1: \vartheta_1 \leq \dots \leq \vartheta_k$ with $\vartheta_1 < \vartheta_k$. Of course, the alternative could also be formulated as a decreasing trend. However, two-sided trend tests should not be considered because there is no sensible interpretation [18].

The standard nonparametric trend test is the Jonckheere test [1,6], which also is the NTP (National Toxicology Program) standard statistical test for evaluating a dose–response relationship in organ weight data [19]. The test statistic T_J is the sum of Mann-Whitney

scores U_{ij} : $T_J = \sum_{i=1}^{k-1} \sum_{j=i+1}^k U_{ij}$. The test can be carried out as an asymptotic test based on the asymptotic normality of T_J [18] or as a permutation test [1]. However, the Jonckheere test is designed for homoscedastic data, so that an alternative method is needed for the common case of unequal variances. Here, it is proposed to combine the Williams trend test [5,13] with a variance estimation that is robust under heteroscedasticity.

The Williams test can be formulated as a multiple contrast test [13]. In general, a contrast is a vector (whose coefficients add up to zero) that assigns weights to the different groups. This enables varying treatment comparisons that can be combined in a multiple contrast test. In the case of a balanced design and $k = 3$ groups, the Williams test compares groups 1 and 3, and the two pooled groups 2 and 3 are compared with the first group. In the case of $k = 4$ the groups 1 and 4 are compared, the two pooled groups 3 and 4 are compared with the first group, and the pooled groups 2–4 are compared with the first group. The null hypothesis H_0 is rejected in favor of a trend when at least one contrast is significant according to the adjusted p -values.

Since the Williams test is a multiple contrast test, it can readily be combined with approaches developed for multiple contrasts. Here, I combine the following three approaches with the Williams test. Herberich et al. [20] proposed a heteroscedastic consistent sandwich covariance estimation function. This function is available in the R package sandwich, version 3.1-1 [21]. Another procedure was introduced by Hasler and Hothorn [22] using an approximate multivariate t -distribution, available in the R package SimComp, version 3.6 [23].

As a further alternative approach, a nonparametric multiple contrast test (MCTP) based on relative effects [24] is investigated. This method, abbreviated as MCTP, uses rank-based multiple contrast tests; it does not require normality, takes the correlation between different contrasts into account, and provides simultaneous confidence intervals which are compatible with the test decision. This MCTP is available in the R package nparcomp (version 3.0) and can be combined with Williams-type trend contrasts [25]. Here, the asymptotic approximation method called “mult.t” is used in order to apply a multivariate t -distribution with a Satterthwaite approximation [25]. For this MCTP, heterogeneous variances are allowed under the null hypothesis. The hypotheses are formulated in terms of relative effects [18]. When comparing two samples, the relative effect can be defined as $P(X < Y) + 0.5 P(X = Y)$ where X is from group 1 and Y from group 2. If the relative effect is 0.5, there is no tendency that one of the samples takes smaller or greater values than the other sample, without any assumptions about the shape of the distributions [1]. The formal definition of relative effects for more than two groups is given by Brunner et al. [18] (p. 61).

A Monte Carlo simulation study was performed using R (version 4.5.2); 10,000 simulation runs were generated for each configuration. The permutation test with the Jonckheere statistic was carried out based on 1000 permutations. Please note that, according to Bonini et al. [26], 1000 permutations are sufficient. Designs with three and four groups, different distributions, balanced and unbalanced sample sizes as well as homogeneous and heterogeneous variances were investigated. R code for the simulation is provided as Supplementary Materials.

3. Results

3.1. Simulation Study

Simulation results for normally distributed data are given in Table 1. The simulated type I error rates confirm that the Jonckheere test cannot control the significance level when variances differ. In some situations with $\alpha = 5\%$ the actual type I error rate increases up to approx. 9%, even for balanced sample sizes. In some other situations, the Jonckheere test is conservative with an actual type I error rate smaller than 3%. These results hold for both

the asymptotic and the permutation Jonckheere test. In contrast, the various Williams tests have an acceptable actual type I error rate much closer to α .

However, the situation is different when data are not normally distributed, see Tables 2 and 3. Then, the Williams test combined with the variance estimation approaches from Herberich [20] as well as Hasler and Hothorn [22] can have unacceptably high type I error rates (>0.1 in some cases), whereas the actual type I error rate of the combination Williams test and MCTP is acceptable in all investigated scenarios. To be precise, the simulated type I error rate ranges from 0.046 to 0.057 (within all scenarios investigated, see Tables 1–3).

Table 1. Simulated type I error rates of different trend tests for normally distributed data with an equal mean of 0 in all groups and homogeneous as well as heterogeneous variances ($\alpha = 0.05$).

Standard Deviations σ	Jonckheere Asymptotic	Jonckheere Permutation	----- Herberich -----	Williams Test Hasler	----- MCTP -----
<i>k</i> = 3, sample sizes: 10, 10, 10					
1, 1, 1	0.050	0.048	0.045	0.050	0.052
1, 1, 2	0.071	0.070	0.048	0.052	0.054
1, 1, 3	0.080	0.079	0.048	0.051	0.054
1, 1, 4	0.087	0.085	0.050	0.051	0.055
4, 4, 1	0.038	0.037	0.050	0.050	0.049
<i>k</i> = 3, sample sizes: 12, 12, 6					
1, 1, 1	0.049	0.052	0.050	0.054	0.052
1, 1, 2	0.072	0.730	0.054	0.055	0.052
1, 1, 3	0.082	0.085	0.059	0.053	0.049
1, 1, 4	0.089	0.091	0.059	0.052	0.046
4, 4, 1	0.030	0.029	0.051	0.054	0.050
<i>k</i> = 3, sample sizes: 6, 12, 12					
1, 1, 1	0.049	0.050	0.050	0.050	0.056
1, 1, 2	0.066	0.068	0.047	0.052	0.056
1, 1, 3	0.074	0.076	0.048	0.051	0.057
1, 1, 4	0.078	0.079	0.047	0.051	0.057
4, 4, 1	0.047	0.048	0.056	0.047	0.047
<i>k</i> = 4, sample sizes: 10, 10, 10, 10					
1, 1, 1, 1	0.050	0.050	0.048	0.052	0.055
1, 1, 1, 4	0.085	0.084	0.048	0.049	0.049
1, 1, 4, 4	0.062	0.062	0.051	0.052	0.051
1, 4, 4, 4	0.031	0.030	0.048	0.050	0.048
4, 4, 4, 1	0.032	0.033	0.049	0.049	0.048
<i>k</i> = 4, sample sizes: 13, 13, 7, 7					
1, 1, 1, 1	0.047	0.047	0.047	0.052	0.050
1, 1, 1, 4	0.083	0.084	0.056	0.051	0.047
1, 1, 4, 4	0.077	0.078	0.053	0.049	0.048
1, 4, 4, 4	0.037	0.039	0.055	0.051	0.048
4, 4, 4, 1	0.026	0.028	0.041	0.044	0.046
<i>k</i> = 4, sample sizes: 7, 7, 13, 13					
1, 1, 1, 1	0.045	0.046	0.050	0.049	0.052
1, 1, 1, 4	0.083	0.084	0.051	0.053	0.053
1, 1, 4, 4	0.043	0.044	0.050	0.053	0.053
1, 4, 4, 4	0.028	0.028	0.046	0.050	0.046
4, 4, 4, 1	0.039	0.039	0.055	0.048	0.051

Hence, the Jonckheere test as well as the Williams test together with the approaches of Herberich or Hasler should not be applied in the common situation that differences in variability can exist between groups. In this case, only the Williams test with MCTP is a suitable option.

For normally distributed data with homogeneous variances and balanced sample sizes, all investigated tests are appropriate. In these scenarios, there are some power differences between the tests. The power comparison depends on the trend pattern, but the differences between the tests are small (see Table 4). Hence, one can suggest the Williams trend test in combination with MCTP for routine use.

The proposed approach, that is, the Williams test in combination with MCTP, is a nonparametric procedure with the additional advantage that it can also be applied to non-continuous data [24,25]. Therefore, the actual type I error rate is also simulated for data distributed according to a Poisson distribution. Again, the simulated actual type I error rates are close to α and acceptable, as displayed in Table 5.

Table 2. Simulated type I error rates of different trend tests for non-normally distributed data and $k = 3$ groups ($\alpha = 0.05$).

	Jonckheere Asymptotic	Jonckheere Permutation	----- Herberich	Williams Test Hasler	----- MCTP
<i>t</i> distributions with <i>df</i> degrees of freedom					
sample sizes: 10, 10, 10					
<i>df</i> = 3	0.049	0.048	0.040	0.045	0.052
<i>df</i> = 5	0.046	0.046	0.039	0.043	0.045
sample sizes: 12, 12, 6					
<i>df</i> = 3	0.045	0.048	0.039	0.044	0.051
<i>df</i> = 5	0.048	0.049	0.045	0.048	0.053
sample sizes: 6, 12, 12					
<i>df</i> = 3	0.047	0.049	0.041	0.041	0.050
<i>df</i> = 5	0.048	0.048	0.048	0.047	0.056
exponential distributions with rate = 3					
sample sizes: 10, 10, 10					
	0.048	0.047	0.051	0.058	0.052
sample sizes: 12, 12, 6					
	0.047	0.048	0.037	0.041	0.050
sample sizes: 6, 12, 12					
	0.046	0.047	0.086	0.095	0.054
chi ² distributions with <i>df</i> degrees of freedom					
sample sizes: 10, 10, 10					
<i>df</i> = 1	0.048	0.047	0.054	0.060	0.052
<i>df</i> = 3	0.048	0.048	0.053	0.060	0.051
sample sizes: 12, 12, 6					
<i>df</i> = 1	0.048	0.048	0.038	0.041	0.052
<i>df</i> = 3	0.043	0.044	0.036	0.041	0.049
sample sizes: 6, 12, 12					
<i>df</i> = 1	0.047	0.049	0.106	0.114	0.054
<i>df</i> = 3	0.044	0.046	0.081	0.089	0.053

Table 3. Simulated type I error rates of different trend tests for non-normally distributed data and $k = 4$ groups ($\alpha = 0.05$).

	Jonckheere Asymptotic	Jonckheere Permutation	----- Herberich	Williams Test Hasler	----- MCTP
<i>t</i> distributions with <i>df</i> degrees of freedom					
sample sizes: 10, 10, 10, 10					
<i>df</i> = 3	0.050	0.049	0.042	0.044	0.050
<i>df</i> = 5	0.053	0.054	0.048	0.052	0.054
sample sizes: 13, 13, 7, 7					
<i>df</i> = 3	0.049	0.049	0.041	0.045	0.052
<i>df</i> = 5	0.048	0.050	0.043	0.047	0.048
sample sizes: 7, 7, 13, 13					
<i>df</i> = 3	0.045	0.046	0.041	0.041	0.048
<i>df</i> = 5	0.047	0.048	0.046	0.046	0.052
exponential distributions with rate = 3					
sample sizes: 10, 10, 10, 10					
	0.053	0.052	0.065	0.071	0.055
sample sizes: 13, 13, 7, 7					
	0.054	0.055	0.048	0.052	0.053
sample sizes: 7, 7, 13, 13					
	0.048	0.049	0.093	0.099	0.053
chi ² distributions with <i>df</i> degrees of freedom					
sample sizes: 10, 10, 10, 10					
<i>df</i> = 1	0.052	0.051	0.065	0.073	0.051
<i>df</i> = 3	0.050	0.049	0.061	0.068	0.051
sample sizes: 13, 13, 7, 7					
<i>df</i> = 1	0.046	0.049	0.045	0.051	0.051
<i>df</i> = 3	0.047	0.049	0.043	0.048	0.049
sample sizes: 7, 7, 13, 13					
<i>df</i> = 1	0.048	0.049	0.109	0.120	0.056
<i>df</i> = 3	0.049	0.050	0.082	0.089	0.054

Table 4. Simulated power of different trend tests for normally distributed data with an equal variance of 1 in all groups and balanced sample sizes of 10 per group ($\alpha = 0.05$).

Means μ	Jonckheere Asymptotic	Jonckheere Permutation	----- Herberich	Williams Test Hasler	----- MCTP
<i>k</i> = 3					
0, 0, 1.2	0.79	0.78	0.76	0.77	0.77
0, 0.6, 1.2	0.81	0.81	0.79	0.80	0.79
0, 1.2, 1.2	0.79	0.78	0.88	0.89	0.88
<i>k</i> = 4					
0, 0, 0, 1.2	0.73	0.73	0.73	0.74	0.73
0, 0.4, 0.8, 1.2	0.86	0.85	0.80	0.81	0.80
0, 0, 1.2, 1.2	0.94	0.93	0.86	0.87	0.86
0, 1.2, 1.2, 1.2	0.74	0.73	0.91	0.91	0.90

Table 5. Simulated type I error rates of Williams test with MCTP for Poisson distributed data with different means λ ($\alpha = 0.05$).

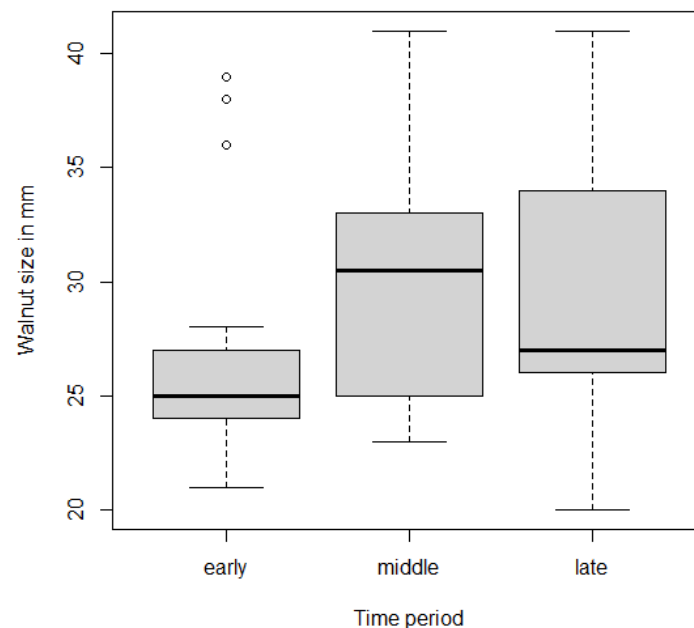
	$\lambda = 3$	$\lambda = 5$
$k = 3$, sample sizes: 10, 10, 10	0.050	0.053
$k = 3$, sample sizes: 12, 12, 6	0.049	0.052
$k = 3$, sample sizes: 6, 12, 12	0.055	0.055
$k = 4$, sample sizes: 10, 10, 10, 10	0.050	0.050
$k = 4$, sample sizes: 13, 13, 7, 7	0.051	0.053
$k = 4$, sample sizes: 7, 7, 13, 13	0.051	0.052

3.2. Example Data: Sizes of Walnuts Opened by Carrion Crows

In late summer and autumn, carrion crows (*Corvus c. corone*) hide walnuts for the winter. Larger nuts are harder to open and are not eaten initially. However, as the winter progresses, even larger nuts are eaten [27,28]. Here, I analyze the size of the opened walnuts (in mm), observed by Josef H. Reichholf during the season 2003–2004 on the premises of the Bavarian State Collection of Zoology in Munich, Germany. Dates are categorized into three categories: an early period before 13 September, a middle period from 13 September to 31 January, and a late period starting from 1 February. There are 67 observations in total, 22 in each of the first two periods and 23 in the late period.

A Tukey boxplot of the data is displayed in Figure 1. Differences in variance are obvious; the standard deviation increases from 4.78 in the early period to 5.15 in the middle period and 5.73 in the late period. However, there is a change in means as well, from 26.59 to 29.95 and 29.35.

The Jonckheere test applied to this data set gives a p -value of 0.0220 (based on 100,000 permutations). The Williams tests with the variance estimation according to Herberich [20] as well as Hasler and Hothorn [22] lead to even smaller p -values of 0.0191 and 0.0177, respectively. However, this significance at the 5% significance level could be caused by an inflated type I error rate due to the heterogeneous variances.

**Figure 1.** Boxplot for the example data on sizes of walnuts opened by carrion crows.

Therefore, the Williams test in combination with MCTP is also applied; it gives a p -value of 0.0095. Thus, we can conclude that there is a trend in walnut sizes irrespective of the increase in variance. The related estimates for the relative effects (and the respective

simultaneous 95–confidence intervals) are 0.38 (0.30 to 0.46) for the first period, 0.57 (0.48 to 0.65) for the second period, and 0.56 (0.48 to 0.63) for the third period. Hence, the first period seems to be different from the other periods, with a corresponding trend pattern such as $\vartheta_1 < \vartheta_2 = \vartheta_3$. This is consistent with the observation that larger nuts are rarely eaten in the early stage of the winter.

4. Discussion and Conclusions

Heterogeneous variances are common in applications. As a consequence, tests for a trend in location can have inflated type I error rates, and trends may appear as by-products of heteroscedasticity. Here, it is shown that the problem of inflated type I error rates persists even when approaches proposed for heterogeneous variances are applied. However, the simulation shows that one approach is acceptable: the combination of the Williams trend test with MCTP, a nonparametric multiple contrast test based on relative effects [24,25]. The latter combination is readily available in R with the following code:

```
library(nparcomp)
mctp(X ~ Y, data = d, type = "Williams", alternative = "greater", asy.method = "mult.t")
```

where X is the outcome variable, Y the grouping variable, and d denotes the data frame. As mentioned above, the asymptotic approximation method called “mult.t” is used in order to apply a multivariate t-distribution with a Satterthwaite approximation.

This procedure can be proposed for routine use when, as usual in applications, the distribution of the underlying data is unknown and the variances might differ between the groups. It also provides estimates for relative effects and corresponding simultaneous confidence intervals as illustrated using example data on sizes of walnuts opened by carrion crows in various periods of winter. Moreover, it can be applied in case of non-continuous data.

Finally, it should be noted that only the one-way layout is investigated in this study. However, a similar solution can be implemented for the case of dependent samples. The nonparametric multiple contrast test MCTP is also available for a repeated measures design, using the function `mctp.rm`. This design will be investigated in future research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/stats9040067/s1>, R code for the analysis of the example data and the simulation study.

Funding: This research received no external funding.

Data Availability Statement: R code for the analysis of the example data and the simulation study is available as Supplementary Materials. The R code also includes the raw data of the example.

Acknowledgments: The author expresses his sincere thanks to Josef H. Reichholf (retired from the Bavarian State Collection of Zoology) for providing the example data set of his observations.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	analysis of variance
CV	coefficient of variation
df	degrees of freedom
MCTP	multiple contrast test procedure

References

1. Neuhäuser, M. *Nonparametric Statistical Tests: A Computational Approach*; CRC Press: Boca Raton, FL, USA, 2012.
2. Bender, R.; Lange, S.; Ziegler, A. Wichtige Signifikanztests. *Dtsch. Med. Wochenschr.* **2007**, *132*, e24–e25. [[CrossRef](#)] [[PubMed](#)]
3. Ogenstad, S. The use of generalized tests in medical research. *J. Biopharm. Stat.* **1998**, *8*, 497–508. [[CrossRef](#)] [[PubMed](#)]
4. Bartholomew, D.J. A test of homogeneity for ordered alternatives. *Biometrika* **1959**, *46*, 36–48. [[CrossRef](#)]
5. Williams, D.A. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics* **1971**, *27*, 103–117. [[CrossRef](#)] [[PubMed](#)]
6. Jonckheere, A.R. A distribution-free k-sample test against ordered alternatives. *Biometrika* **1954**, *41*, 133–145. [[CrossRef](#)]
7. Shorack, G.R. Testing against ordered alternatives in model I analysis of variance; normal theory and nonparametric. *Ann. Math. Stat.* **1967**, *38*, 1740–1752. [[CrossRef](#)]
8. Bonnini, S.; Prodi, N.; Salmaso, L.; Visentin, C. Permutation approach for stochastic ordering. *Commun. Stat. Theory Methods* **2014**, *43*, 2227–2235. [[CrossRef](#)]
9. Basso, D.; Pesarin, F.; Salmaso, L.; Solari, A. *Permutation Tests for Stochastic Ordering and ANOVA*; Springer: New York, NY, USA, 2009.
10. Gould, S.J. *Full House: The Spread of Excellence from Plato to Darwin*; Harmony Books: New York, NY, USA, 1996.
11. Gould, S.J. Trends as changes in variance: A new slant on progress and directionality in evolution. *J. Paleontol.* **1988**, *62*, 319–329. [[CrossRef](#)]
12. Bracht, T.; Mölleken, C.; Ahrens, M.; Poschmann, G.; Schlosser, A.; Eisenacher, M.; Stühler, K.; Meyer, H.E.; Schmiegel, W.H.; Holmskov, U.; et al. Evaluation of the biomarker candidate MFAP4 for non-invasive assessment of hepatic fibrosis in hepatitis C patients. *J. Transl. Med.* **2016**, *14*, 201. [[CrossRef](#)] [[PubMed](#)]
13. Hothorn, L.A. *Statistics in Toxicology Using R*; CRC Press: Boca Raton, FL, USA, 2016.
14. Neuhäuser, M.; Hothorn, L.A. Parametric location-scale and scale tend tests based on Levene's transformation. *Comput. Stat. Data Anal.* **2000**, *33*, 189–200. [[CrossRef](#)]
15. Neuhäuser, M. One-sided two-sample and trend tests based on a modified Baumgartner-Weiss-Schindler statistic. *J. Nonparametric Stat.* **2001**, *13*, 729–739. [[CrossRef](#)]
16. Stepniak, C. Coefficient of variation. In *International Encyclopedia of Statistical Science*, 2nd ed.; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2025; pp. 487–488.
17. Döhning, T.F.; Reckling, M. Detecting global trends of cereal yield stability by adjusting the coefficient of variation. *Eur. J. Agron.* **2018**, *99*, 30–36. [[CrossRef](#)]
18. Brunner, E.; Bathke, A.C.; Konietzschke, F. *Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs*; Springer: Cham, Switzerland, 2018.
19. Jeffers, A.; Konrad, K.; Larson, G.; Allen-Moyer, K.; Cunny, H.; Shockley, K. Simulation methodologies to determine statistical power in laboratory animal research studies. *Lab. Anim.* **2024**, *58*, 486–492. [[CrossRef](#)] [[PubMed](#)]
20. Herberich, E.; Sikorski, J.; Hothorn, T. A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE* **2010**, *5*, e9788. [[CrossRef](#)] [[PubMed](#)]
21. Zeileis, A. Object-oriented computation of sandwich estimators. *J. Stat. Softw.* **2006**, *16*, 1–16. [[CrossRef](#)]
22. Hasler, M.; Hothorn, L.A. A Dunnett-type procedure for multiple endpoints. *Int. J. Biostat.* **2011**, *7*, 3. [[CrossRef](#)]
23. Hasler, M.; Kluss, C. *SimComp: Simultaneous Comparisons for Multiple Endpoints*, R Package, version 3.6; R Foundation: Vienna, Austria, 2025.
24. Konietzschke, F.; Hothorn, L.A.; Brunner, E. Rank-based multiple test procedures and simultaneous confidence intervals. *Electron. J. Stat.* **2012**, *6*, 738–759. [[CrossRef](#)]
25. Konietzschke, F.; Placzek, M.; Schaarschmidt, S.; Hothorn, L.A. nparcomp: An R software package for nonparametric multiple comparisons and simultaneous confidence intervals. *J. Stat. Softw.* **2015**, *61*, 1–17.
26. Bonnini, S.; Assegie, G.M.; Trzcinska, K. Review about the permutation approach in hypothesis testing. *Mathematics* **2024**, *12*, 2617. [[CrossRef](#)]
27. Reichholf, J.H. Winternutzung versteckter Walnüsse durch Rabenkrähen *Corvus c. corone*. *Ornithol. Mitteilungen* **2004**, *56*, 257–262.
28. Reichholf, J.H. *Rabenschwarze Intelligenz*; F.A. Herbig: Munich, Germany, 2009.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.