Permutation tests for coefficients of variation in general one-way ANOVA models

Markus Pauly[1] and Łukasz Smaga[2]

[1]Faculty of Statistics
TU Dortmund University
Dortmund, Germany

[2]Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Poznań, Poland

Coefficient of variation

- The coefficient of variation (CV)
$$c = \frac{\sigma}{\mu}$$
is a unitless dispersion measure for data given on a ratio scale.
- It has many applications as, for example:
  - guide of the performance and repeatability of measurements in clinical trials (Feltz and Miller, 1996),
  - a reliability tool in engineering control charts (Castagliola et al., 2013),
  - a measure of risk in empirical finance and psychology (Ferri and Jones, 1979; Weber et al., 2004),
  - supply chain management in distributional and procurement logistic (Wanke and Zinn, 2004),
  - quantifying variability in genetics (Wright, 1952).

Statistical inference for CV

- Many of confidence intervals and tests for CVs are based on the estimator

$$\widehat{c} = \frac{s}{\bar{x}}$$

  of the CV or unbiased modifications thereof.
- As the asymptotic distribution of $\widehat{c}$ depends on potentially unknown model parameters as the curtosis, many methods are derived for parametric models.
- Assuming normality, testing for equality of CVs have been proposed by, e.g., Feltz and Miller (1996) - the current gold standard, Forkman (2009), Krishnamoorthy and Lee (2014), whereas Aerts and Haesbroeck (2017) investigate multivariate CVs under elliptic symmetry.

- If the model is correctly specified, most of these methods perform fairly well. Otherwise, however, the procedures may not be reliable in general.
- To this end, we consider a statistic of Wald-type in a general model and equip it with a permutation technique to assure good finite sample behaviour.
- The resulting permutation test is finitely exact if data is exchangeable and is asymptotically correct in general.

- We consider a general k-sample model ($k \geq 2$) given by independent random variables

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad 1 \leq i \leq k, \ 1 \leq j \leq n_i,$$

where $\epsilon_{i1}, \ldots, \epsilon_{in_i}$ are independent and identically distributed with

$$\mathbb{E}(\epsilon_{i1}) = 0, \ \mathbb{E}(\epsilon_{i1}^2) = \sigma_i^2 > 0, \ \sup_{1 \leq i \leq k} \mathbb{E}(\epsilon_{i1}^4) < \infty.$$

- X_{ij} describes the j-th observation in group i, n_i the i-th sample size and $N = n_1 + \cdots + n_k$ denotes the total sample size.

Hypotheses

- We define the CV of the i-th group as

$$c_i = \frac{\sigma_i}{\mu_i}$$

(additionally assuming $\mu_i \neq 0$) and set

$$\beta_i = \frac{\mu_i}{\sigma_i}$$

for the corresponding standardized mean.

- The null hypothesis of equal CVs:

$$H_0 : c_1 = \cdots = c_k$$

(assuming $\mu_i \neq 0$, $i = 1, \ldots, k$) or equivalently of equal standardized means

$$H_0 : \beta_1 = \cdots = \beta_k.$$

- The natural plug-in estimators:

$$\widehat{c}_i = \frac{\widehat{\sigma}_i}{\bar{X}_{i\cdot}} \text{ and } \widehat{\beta}_i = \frac{\bar{X}_{i\cdot}}{\widehat{\sigma}_i},$$

  where $\bar{X}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $\widehat{\sigma}_i^2 = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2$ are the sample mean and variance in group i, $i = 1, \ldots, k$.
- The estimators $\widehat{c}_i$ and $\widehat{\beta}_i$ are consistent and asymptotically normal (as $\min(n_1, \ldots, n_k) \to \infty$) under the given assumptions.

Lemma

Set $\theta_i := (\mu_i, \sigma_i^2)^\top \in \mathbb{R} \times (0, \infty)$. Then we have for each $i = 1, \ldots, k$:
(a) for the CV estimator $\widehat{c}_i = \widehat{\sigma}_i / \bar{X}_{i\cdot}$ assuming that $\mu_i \neq 0$:

$$\sqrt{n_i}\,(\widehat{c}_i - c_i) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} h_{\theta_i}(X_{ij}) + o_P(1),$$

where

$$h_{\theta_i}(X_{i1}) = \frac{(X_{i1} - \mu_i)^2 - \sigma_i^2}{2\mu_i\sigma_i} - \frac{\sigma_i(X_{i1} - \mu_i)}{\mu_i^2}.$$

Moreover, $\mathbb{E}(h_{\theta_i}(X_{i1})) = 0$ and

$$\mathrm{Var}(h_{\theta_i}(X_{i1})) = \frac{\sigma_i^4}{\mu_i^4} - \frac{\mathbb{E}(X_{i1}^3) - 3\mu_i\sigma_i^2 - \mu_i^3}{\mu_i^3} + \frac{\mathbb{E}(X_{i1}^4) - 4\mu_i\mathbb{E}(X_{i1}^3) + 6\mu_i^2\sigma_i^2 + 3\mu_i^4 - \sigma_i^4}{4\mu_i^2\sigma_i^2}.$$

Lemma

(b) for the standardized means estimator $\widehat{\beta}_i = \bar{X}_{i\cdot}/\widehat{\sigma}_i$ *we have*

$$\sqrt{n_i}\left(\widehat{\beta}_i - \beta_i\right) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} h_{\theta_i,inv}(X_{ij}) + o_P(1),$$

where $h_{\theta_i,inv}(X_{i1}) = -(\mu_i/\sigma_i)^2 h_{\theta_i}(X_{i1})$ *for* $\mu_i \neq 0$ *and* $h_{\theta_i,inv}(X_{i1}) = X_{i1}/\sigma_i$ *for* $\mu_i = 0$. *Furthermore,* $\mathbb{E}(h_{\theta_i,inv}(X_{i1})) = 0$ *and*

$$\mathrm{Var}(h_{\theta_i,inv}(X_{i1})) = (\mu_i/\sigma_i)^4 \mathrm{Var}(h_{\theta_i}(X_{i1}))$$

for $\mu_i \neq 0$ *and* $\mathrm{Var}(h_{\theta_i,inv}(X_{i1})) = 1$ *for* $\mu_i = 0$.

Lemma

Under the assumptions of the above Lemma and presuming $\mu_i \neq 0$ we set

$$p(\mu_i, \sigma_i) = \frac{\mu_i \sqrt{\frac{\sigma_i^2}{\mu_i^2} + 1} - \sigma_i}{2\mu_i \sqrt{\frac{\sigma_i^2}{\mu_i^2} + 1}} \in (0, 1).$$

Then we have $\mathrm{Var}(h_{\theta_i}(X_{i1})) = 0$ (as well as $\mathrm{Var}(h_{\theta_i,inv}(X_{i1})) = 0$) if and only if X_{i1} has a specific two-point distribution given by

$$X_{i1} = \begin{cases} \mu_i + \frac{\sigma_i^2}{\mu_i} + \sigma_i \sqrt{\frac{\sigma_i^2}{\mu_i^2} + 1}, & \textit{with probability } p(\mu_i, \sigma_i), \\ \mu_i + \frac{\sigma_i^2}{\mu_i} - \sigma_i \sqrt{\frac{\sigma_i^2}{\mu_i^2} + 1}, & \textit{with probability } 1 - p(\mu_i, \sigma_i). \end{cases} \tag{1}$$

- Assuming $n_i/N \to p_i > 0$ for $i = 1, \dots, k$, we have

$$\sqrt{N}(\widehat{c}_i - c_i) \xrightarrow{d} N\left(0, \frac{\mathrm{Var}(h_{\theta_i}(X_{i1}))}{p_i}\right)$$

  and

$$\sqrt{N}(\widehat{\beta}_i - \beta_i) \xrightarrow{d} N\left(0, \frac{\mathrm{Var}(h_{\theta_i,inv}(X_{i1}))}{p_i}\right)$$

  for each $i = 1, \dots, k$.
- To estimate $\mathrm{Var}(h_{\theta_i}(X_{i1}))$ and $\mathrm{Var}(h_{\theta_i,inv}(X_{i1}))$, we use the empirical sample means $\bar{X}_{i\cdot}$, variances $\hat{\sigma}_i^2$, and third and fourth moments $n_i^{-1}\sum_{j=1}^{n_i} X_{ij}^3$ and $n_i^{-1}\sum_{j=1}^{n_i} X_{ij}^4$, respectively.
- Denote the resulting estimators of $\mathrm{Var}(h_{\theta_i}(X_{i1}))$ and $\mathrm{Var}(h_{\theta_i,inv}(X_{i1}))$ by S_i^2 and $S_{i,inv}^2$ respectively.

Test statistics

- Following, e.g., Feltz and Miller (1996) and Chung and Romano (2013), we propose to use statistics of James-type (James, 1951):

$$Z_N = \sum_{i=1}^{k} \frac{n_i}{S_i^2} \left[\widehat{c}_i - \frac{\sum_{i=1}^{k} n_i \widehat{c}_i / S_i^2}{\sum_{i=1}^{k} n_i / S_i^2} \right]^2,$$

$$Z_{N,inv} = \sum_{i=1}^{k} \frac{n_i}{S_{inv,i}^2} \left[\widehat{\beta}_i - \frac{\sum_{i=1}^{k} n_i \widehat{\beta}_i / S_{inv,i}^2}{\sum_{i=1}^{k} n_i / S_{inv,i}^2} \right]^2.$$

- Under the above assumptions:

$$Z_N|_{H_0} \xrightarrow{d} \chi^2_{k-1}$$

and

$$Z_{N,inv}|_{H_0} \xrightarrow{d} \chi^2_{k-1}.$$

- Asymptotic χ^2-tests:

$$\varphi_N = \mathbf{1}\{Z_N > \chi^2_{k-1,1-\alpha}\},$$

and

$$\varphi_{N,inv} = \mathbf{1}\{Z_{N,inv} > \chi^2_{k-1,1-\alpha}\}$$

for H_0, where $\chi^2_{l,\alpha}$ denotes the α-quantile of the χ^2_l-distribution.

- Keeping the pooled data

$$(Y_1, \ldots, Y_N) = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}, \ldots, X_{k1}, \ldots, X_{kn_k})$$

fixed, a permutation π is uniformly chosen from the symmetric group $\mathcal{S}_N$ and the test statistic, say Z_N, is recalculated with the permuted sample

$$(Y_{\pi(1)}, \ldots, Y_{\pi(N)}).$$

- The permutation Z_N-test is given by

$$\varphi_N^\pi = \mathbf{1}\{Z_N > c_{1-\alpha}^\pi\},$$

where $c_{1-\alpha}^\pi$ denotes the (conditional) $(1-\alpha)$-quantile of the distribution function of $Z_N(Y_{\pi(1)}, \ldots, Y_{\pi(N)})$ given by

$$t \mapsto \hat{R}_{Z_N}(t) = \frac{1}{N!} \sum_{\pi \in \mathcal{S}_N} \mathbf{1}(Z_N(Y_{\pi(1)}, \ldots, Y_{\pi(N)}) \le t).$$

- By construction, this test is exact if $(Y_1, \ldots, Y_N)$ are exchangeable.

- Chung and Romano (2013)

Theorem

Under the above assumptions and some additional ones, the permutation distributions of Z_N *and* $Z_{N,inv}$ *mimic the asymptotic null distribution, that is we have convergence in probability under* H_0 *as* $\min_i n_i \to \infty$

$$\hat{R}_{Z_N}(t) = \frac{1}{N!} \sum_{\pi \in \mathcal{S}_N} \mathbf{1}(Z_N(Y_{\pi(1)}, \ldots, Y_{\pi(N)}) \leq t) \xrightarrow{P} \chi^2_{k-1}(t)$$

$$\hat{R}_{Z_{N,inv}}(t) = \frac{1}{N!} \sum_{\pi \in \mathcal{S}_N} \mathbf{1}(Z_{N,inv}(Y_{\pi(1)}, \ldots, Y_{\pi(N)}) \leq t) \xrightarrow{P} \chi^2_{k-1}(t).$$

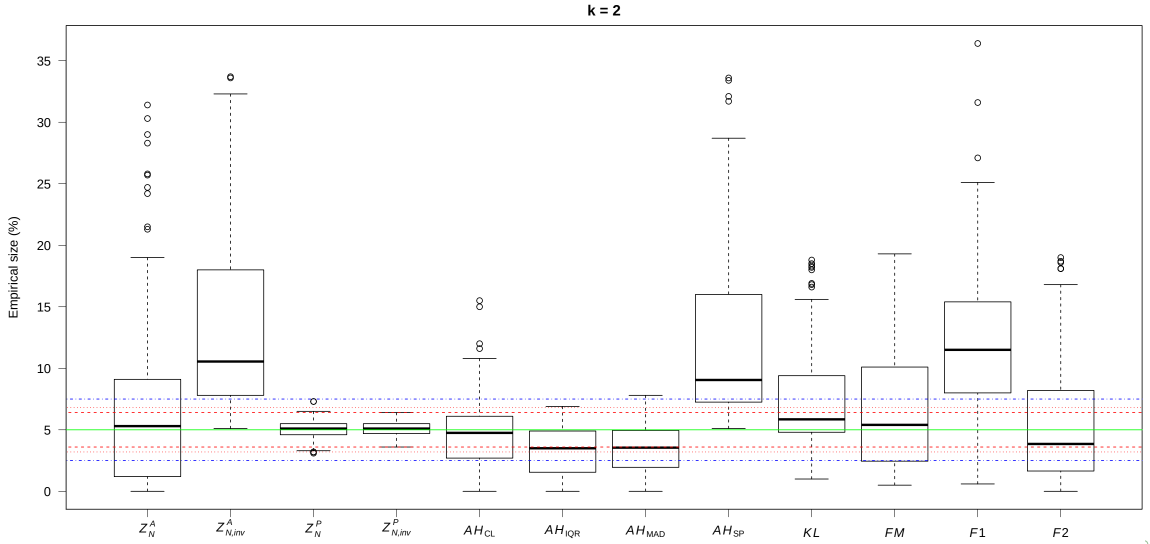
Moreover, the probabilities that the Z_N*- and* $Z_{N,inv}$*-permutation tests reject* H_0 *tend to* α*.*

- comparison of tests in terms of size control and power
- $k = 2$
  - variance-ratio F tests Wright (1952), Bader and Lehman (1965) (the $F1$-test) and by Forkman (2009) (the $F2$-test)
- $k \geq 2$
  - an approximate FM-test (Feltz and Miller, 1996) - the current gold standard
  - a modified signed-likelihood ratio test (Krishnamoorthy and Lee, 2014)
  - Aerts and Haesbroeck (2017) - procedures based on different estimations of the CVs as a classical, an interquartile range, a median absolute deviation or a semi-parametric estimation (the $\mathrm{AH}_{\mathrm{CL}}$, $\mathrm{AH}_{\mathrm{IQR}}$, $\mathrm{AH}_{\mathrm{MAD}}$, $\mathrm{AH}_{\mathrm{SP}}$ tests)

- Aerts and Haesbroeck (2017)
- $k = 2$ or $k = 3$
- equal and unequal distributions across the k groups:
  1. power exponential distribution with parameter $\beta = 2$,
  2. normal distribution,
  3. power exponential distribution with parameter $\beta = 0.5$,
  4. Student t_5-distribution with five degrees of freedom,
  5. normal distribution in the first group (resp. first and second group for $k = 3$) and power exponential distribution with parameter $\beta = 0.5$ in the second (resp. third for $k = 3$) group.

- $\mu_i = 1/c_i$, $\sigma_i^2 = 1$ in group i, where c_i is the i-th CV, $i = 1, \dots, k$
- under null hypothesis: $c_i \in \{0.05, 0.1, 0.5, 1, 1.5, 2\}$
- under alternative hypothesis:
  - for $k = 2$, $(c_1, c_2) \in \{(0.07, 0.1), (0.13, 0.1), (0.5, 1), (1.5, 1)\}$
  - for $k = 3$, $(c_1, c_2, c_3) \in \{(0.07, 0.1, 0.1), (0.13, 0.1, 0.1), (0.5, 1, 1), (1.5, 1, 1)\}$
- $(n_1, n_2) \in \{(4, 7), (10, 15), (25, 30), (50, 50)\}$ for $k = 2$
  $(n_1, n_2, n_3) \in \{(20, 30, 25), (50, 50, 50)\}$ for $k = 3$
- 1000 simulation replications
- 1000 Monte Carlo runs
- $\alpha = 5\%$
- When the observations are assumed to be generated from distribution F (e.g., normal or power exponential), the $\mathrm{AH}_{\mathrm{CL}}$, $\mathrm{AH}_{\mathrm{IQR}}$ and $\mathrm{AH}_{\mathrm{MAD}}$ tests were performed assuming this distribution, i.e., the asymptotic variances of and consistency factors for the estimators used in these tests were computed for F to ensure their asymptotic correctness.

Simulation results



Simulation results

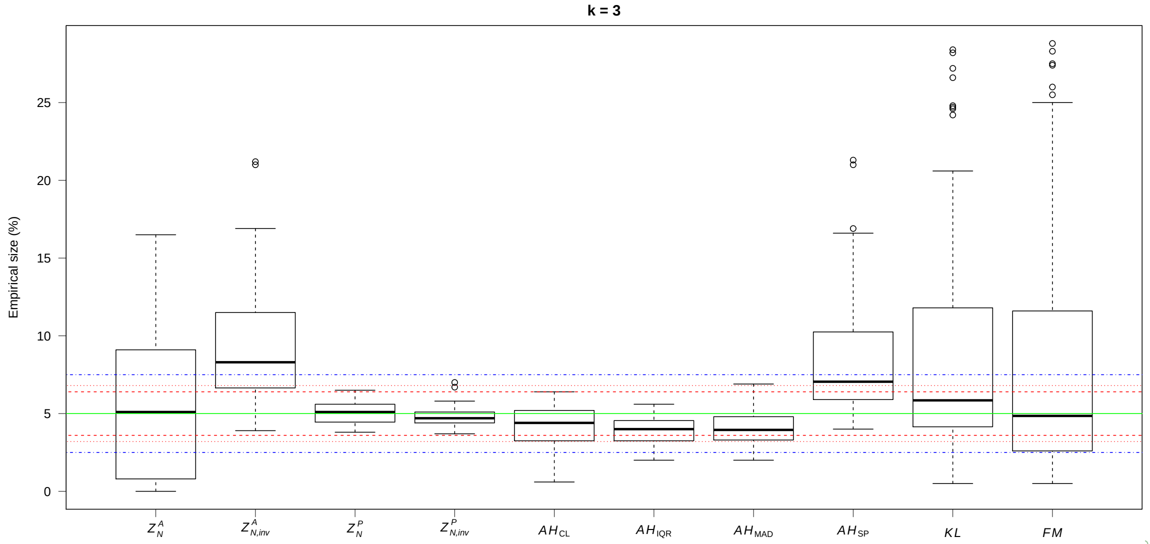
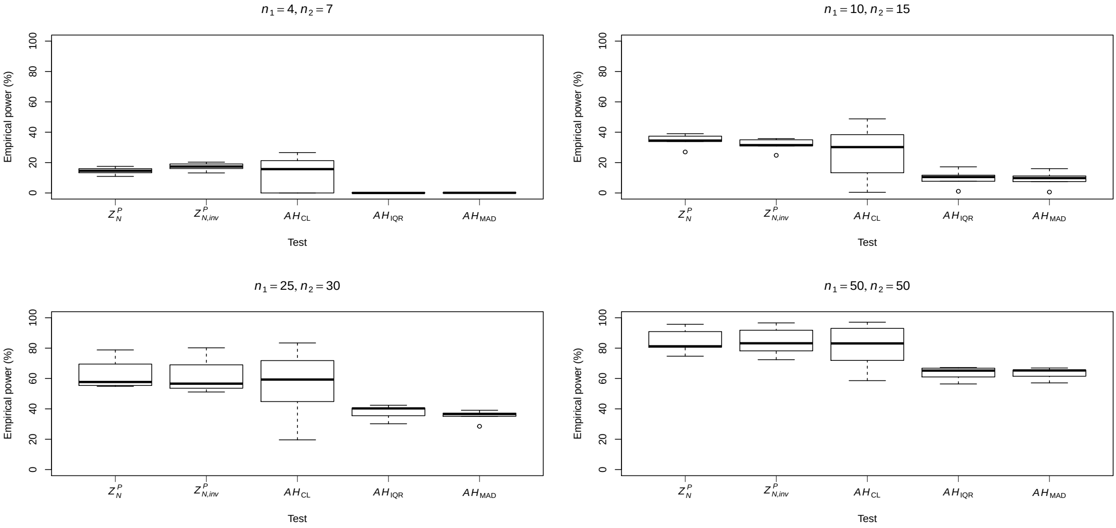


Table: (Bradley's liberal criterion of robustness) Proportions (as percentages) of the empirical sizes of the tests obtained from all cases considered smaller than 2.5% or greater than 7.5% or both.

	Z_N^A	$Z_{N,inv}^A$	Z_N^P	$Z_{N,inv}^P$	AH_{CL}	AH_{IQR}	AH_{MAD}	AH_{SP}	KL	FM	$F1$	$F2$
$k = 2$												
< 2.5	37.5	0.00	0.00	0.00	23.33	36.67	30.00	0.0	5.83	25.00	6.67	39.17
> 7.5	32.5	78.33	0.00	0.00	10.83	0.00	0.83	72.5	33.33	34.17	77.50	28.33
< 2.5 or > 7.5	70.0	78.33	0.00	0.00	34.17	36.67	30.83	72.5	39.17	59.17	84.17	67.50
$k = 3$												
< 2.5	35.00	0.00	0.00	0.00	15.00	3.33	3.33	0.00	10.00	23.33		
> 7.5	35.00	65.00	0.00	0.00	0.00	0.00	0.00	43.33	40.00	36.67		
< 2.5 or > 7.5	70.00	65.00	0.00	0.00	15.00	3.33	3.33	43.33	50.00	60.00		



$n_1 = 4, n_2 = 7$
$n_1 = 10, n_2 = 15$
$n_1 = 25, n_2 = 30$
$n_1 = 50, n_2 = 50$
Empirical power (%)
0
20
40
60
80
100
Z_N^P
$Z_{N,inv}^P$
AH_{CL}
AH_{IQR}
AH_{MAD}
Test

Real data example 1

- biochemical analysis for weight disorders (Smith, Gnanadesikan and Hughes, 1962)
- The data set consists of different biometricals (e.g., pH, pigment creatinine, concentration of choline) and characteristics (e.g., volume) of urine specimens of young men classified into four groups according to their degree or type of weight disorder:
  - Group 1 - severe underweight ($n_1 = 12$)
  - Group 2 - slight underweight ($n_2 = 14$)
  - Group 3 - slight obesity ($n_3 = 11$)
  - Group 4 - severe obesity ($n_4 = 8$)
- It was already used for illustrational purposes in multivariate analysis of variances and covariances (Smith, Gnanadesikan and Hughes, 1962; Morrison, 1990) as well as for classification tasks (Górecki and Łuczak, 2013).
- We consider the CVs of concentration of choline and test the preceding null hypothesis separately in the first two, three and all four groups.

Real data example 1

- $\widehat{c}_1 = 96.96\%$, $\widehat{c}_2 = 52.22\%$, $\widehat{c}_3 = 63.72\%$, $\widehat{c}_4 = 89.54\%$

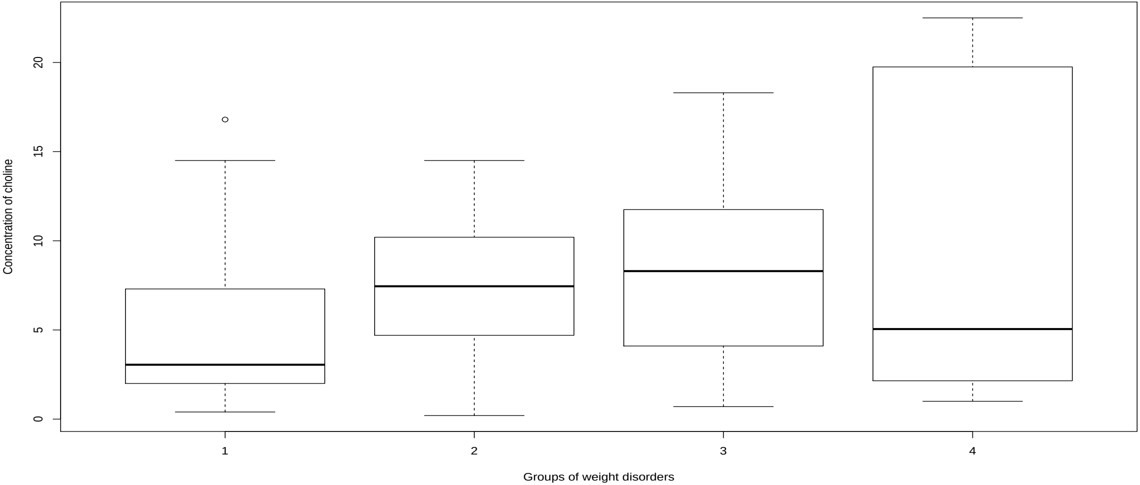


Table: P-values of the tests for comparison of the concentration of choline separately in the first two, three and all four groups of weight disorders.

Groups	Z_N^A	$Z_{N,inv}^A$	Z_N^P	$Z_{N,inv}^P$	AH_{CL}	AH_{IQR}	AH_{MAD}	AH_{SP}	KL	FM	$F1$	$F2$
1-2	0.0133	0.0293	0.0144	0.0320	0.1241	0.1749	0.5331	0.0953	0.1437	0.1434	0.9094	0.1574
1-3	0.0459	0.0496	0.0622	0.0564	0.2861	0.3594	0.7423	0.2405	0.3319	0.2883		
1-4	0.0644	0.1043	0.1157	0.1390	0.4113	0.2543	0.7501	0.3108	0.4750	0.4606		

- quality assurance study for medical laboratories (Fung and Tsang, 1998)
- In that study, the quality of laboratory technology for haematology and serology in Hong Kong was investigated by means of CVs.
- For illustrative purposes, we focus on the hemoglobin (Hb) measurements of abnormal samples and compare them for the two years 1995 ($n_1 = 65$) and 1996 ($n_2 = 73$).

Real data example 2

- $\widehat{c}_1 = 1.75\%$ and $\widehat{c}_2 = 2.63\%$ for all data and $\widehat{c}_2 = 1.71\%$ after removing the outlier

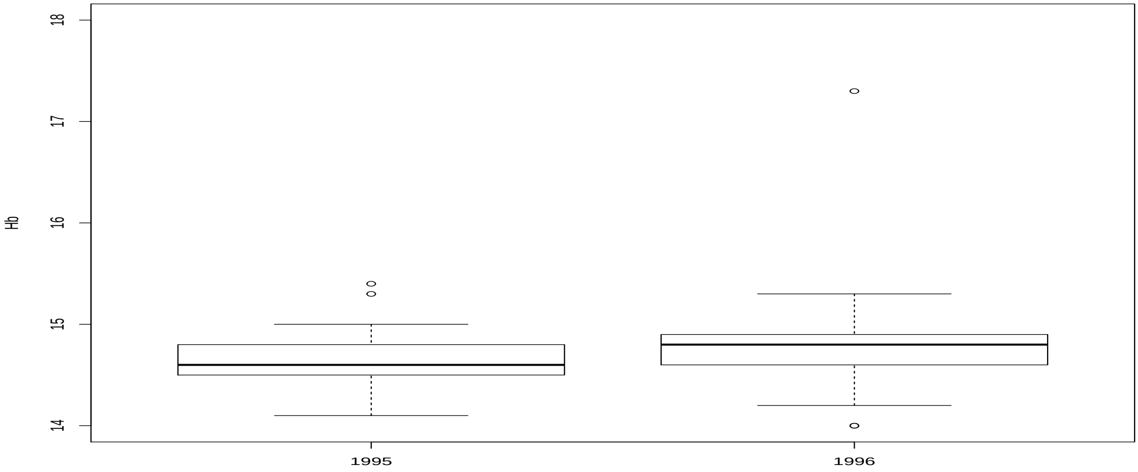


Table: P-values of the tests for comparison of the measurements of Hb in the abnormal sample in 1995 and in 1996.

	Z_N^A	$Z_{N,inv}^A$	Z_N^P	$Z_{N,inv}^P$	AH_{CL}	AH_{IQR}	AH_{MAD}	AH_{SP}	KL	FM	$F1$	$F2$
all data	0.2521	0.1143	0.6060	0.6387	0.0012	0.9455	0.9455	0.1168	0.0009	0.0011	0.0029	0.0011
without outlier	0.8577	0.8579	0.8616	0.8620	0.8381	0.9456	0.1355	0.8575	0.8457	0.8344	0.8949	0.8318

- We studied the problem of testing equality coefficients of variation or standardized means.
- We proposed two statistics of James-type (plain and inverse) and shown the asymptotic χ^2-tests are asymptotically correct regardless of the underlying distribution of the data, which is in general not the case for existing methods.
- However, large sample sizes are needed to obtain satisfactory type-*I*-error control for the asymptotic χ^2-tests. To this end, we additionally proposed permutation versions of both methods. They are both finitely exact under exchangeable situations and also proven to be asymptotically valid in general.

- Moreover, in an extensive simulation study and two real data examples, the permutation tests showed the overall best finite sample properties among all of the 12 investigated procedures.
- It is, of course, tempting to investigate whether the proposed permutation approach extends to the multivariate setting (Albert and Zhang, 2010; Aerts, Haesbroeck and Ruwet, 2015; Aerts and Haesbroeck, 2017) or whether they can also be combined with robust estimators.

1. Aerts, S., Haesbroeck, G. and Ruwet, C. (2015). Multivariate coefficients of variation: Comparison and influence functions. Journal of Multivariate Analysis 142, 183–198.
2. Aerts, S., Haesbroeck, G. (2017). Robust asymptotic tests for the equality of multivariate coefficients of variation. Test 26, 163–187.
3. Albert, A. and Zhang, L. (2010). A novel definition of the multivariate coefficient of variation. Biometrical Journal 52, 667–675.
4. Bader, R.S., Lehman, W.H. (1965). Phenotypic and genotypic variation in odontometric traits of the house mouse. American Midlands Naturalist 74, 28–38.
5. Castagliola P., Achouri A., Taleb H., Celano G., Psarakis S. (2013). Monitoring the coefficient of variation using control charts with run rules. Qual Technol Quant Manag 10, 75–94.
6. Chung, E.Y., Romano, J.P. (2013). Exact and asymptotically robust permutation tests. The Annals of Statistics 41, 484–507.
7. Feltz, C.J., Miller, G.E. (1996). An asymptotic test for the equality of coefficients of variation from k population. Statistics in Medicine 15, 647–658.

8 Ferri, M. G., Jones, W. H. (1979). Determinants of financial structure: A new methodological approach. The Journal of Finance 34(3), 631–644.

9 Forkman, J. (2009). Estimator and tests for common coefficients of variation in normal distributions. Communications in Statistics - Theory and Methods 38, 233–251.

10 Fung, W. K., Tsang, T. S. (1998). A simulation study comparing tests for the equality of coefficients of variation. Statistics in Medicine 17, 2003–2014.

11 Górecki, T., Łuczak, M. (2013). Linear discriminant analysis with a generalization of the Moore-Penrose pseudoinverse. International Journal of Applied Mathematics and Computer Science 23(2), 463–471.

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

13 Krishnamoorthy, K., Lee, M. (2014). Improved tests for the equality of normal coefficients of variation. Computational Statistics 29, 215–232.

14. Morrison, D. (1990). Multivariate Statistical Methods. McGraw-Hill Series in Probability and Statistics, McGraw-Hill, New York, NY.
15. Pauly M., Smaga Ł. (2020). Asymptotic permutation tests for coefficients of variation and standardized means in general one-way ANOVA models. Statistical Methods in Medical Research DOI: 10.1177/0962280220909959
16. Smith, H., Gnanadesikan, R., Hughes, J.B. (1962). Multivariate analysis of variance (MANOVA). Biometrics 18(1), 22–41.
17. Wanke, P.F. Zinn, W. (2004). Strategic logistics decision making. International Journal of Physical Distribution & Logistics Management 34(6), 466–478.
18. Weber, E. U., Shafir, S., Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. Psychological Review 111(2), 430–445.
19. Wright, S. (1952). The genetics of quantitative variability. In: Reeve, E.C.R. and C. Waddington (eds.). Quantitative Inheritance. pp. 5–41. H.M.S.O., London.

Thank you for your attention!