STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 181–194, https://doi.org/10.59139/stattrans-2025-033
Received – 05.02.2025; accepted – 24.04.2025

Multivariate two-sample permutation test with directional alternative for categorical data

Stefano Bonnini[1], **Michela Borghesi**[2]

Abstract

This paper presents a distribution-free test, based on the permutation approach, on treatment effects with a multivariate categorical response variable. The motivating example is a typical case-control biomedical study, performed to investigate the effect of the treatment called "assisted motor activity" (AMA) on the health of comorbid patients affected by "low back pain" (LBP), "hypertension" and "diabetes". Specifically, the goal was to test whether the AMA determines an improvement in the functionality and the perceived health status of patients. Two independent samples (treated and control group) were compared according to 13 different binary or ordinal outcomes. The null hypothesis of the test consists in the equality in the distribution of the multivariate responses of the two groups, whereas under the alternative hypothesis, the health status of the treated patients is better. The approach proposed in this work is based on the Combined Permutation Test (CPT) method, which is suitable for analyzing multivariate categorical data in the presence of confounding factors. A stratification of the groups and intra-stratum permutation univariate two-sample tests are conducted to avoid the potential confounding effects. P-values from the partial tests are combined using the CPT approach to create a suitable test statistic for the overall problem.

Key words: nonparametric statistics, permutation test, multivariate statistics, categorical data.

1. Introduction

This study involves applying a distribution-free test using a permutation approach to address a multivariate biostatistical problem. The main goal of this work is to test the effect of "assisted motor activity" (AMA) on the health of patients affected by "low back

[1] Department of Economics and Management, University of Ferrara, Ferrara, Italy. E-mail: stefano.bonnini@unife.it. ORCID: https://orcid.org/0000-0002-7972-3046.
[2] Department of Economics and Management, University of Ferrara, Ferrara, Italy. E-mail: michela.borghesi@unife.it. ORCID: https://orcid.org/0000-0003-1872-5766.

pain" (LBP), "hypertension" and "diabetes". AMA is a therapy that focuses on targeted physical exercises to regain mobility restrictions resulting from various causes.

In the literature, some studies were conducted on the effect of motor activity on patients affected by low back pain, especially on proving that treatments based on specific physical exercises can improve the health of patients (Ferreira et al., 2007; Gordon and Bloxham, 2016). This type of analysis almost always consists of a randomized controlled trial and parametric methods are used which often require large sample sizes to obtain reliable results (Macedo et al., 2012; Macedo et al., 2014).

This study is a case-control experiment that compares two independent samples. The samples consist of a treated group of 27 patients (group 1) and a control group of 20 patients (group 0). At time t_0 (prior to treatment), there is no significant difference in health status between the two groups. A comparison is made at time t_1, following the treatment for group 1. Health status is assessed using 13 different binary or ordinal outcomes. The null hypothesis states that the distribution of multivariate responses is equal for groups 1 and 0, while the alternative hypothesis suggests that the treated patients have a better health status. This alternative hypothesis is directional, indicating a multivariate stochastic dominance issue for ordinal variables. Since being over 60 years old and having LBP could be a risk factor, a confounding factor s is established, where $s = 1$ if the patient is over 60 years old and has LBP, and $s = 0$ otherwise.

The methodology outlined in this study is grounded in the Combined Permutation Test (CPT), which is effective for handling multivariate categorical data and addressing the issue of confounding factors (Pesarin and Salmaso, 2010; Bonnini et al., 2014). The combined permutation testing approach has been successfully applied in a wide range of contexts (Alibrandi et al., 2022). Specifically, it has seen extensive use in empirical research (Toma et al., 2017), involving both numerical and categorical variables (Bonnini et al., 2023), large-scale data scenarios (Simon and Tibshirani, 2012), and regression analyses (Bonnini and Borghesi, 2022). Furthermore, this method has proven effective in testing both directional and non-monotonic hypotheses (Bonnini et al., 2024a; Stute et al., 1998), in the analysis of count data (Bonnini et al., 2024b), and in numerous other applications. For more detailed information on the methodology of the nonparametric combination, refer to Bonnini et al. (2024).

To mitigate confounding effects by comparing comparable patients regarding the confounder, we implement stratification of the groups and conduct intra-stratum permutation univariate two-sample tests. Considering there are 13 components in the multivariate response and 2 strata, the total number of partial tests amounts to 26. Utilizing the CPT methodology to combine the p-values from these partial tests yields a test statistic that is appropriate for the overall analysis. From the application point of view, the aim is to evaluate if AMA leads to enhancements in the functionality and perceived health condition of patients with comorbidities, using a significance level $\alpha = 0.05$.

Section 2 is dedicated to the description of the statistical problem and Section 3 includes a description of the proposed methodological solution. In Section 4 a simulation study is carried out, Section 5 is dedicated to the case study and the results are reported and commented in the conclusions (Section 6).

2. Statistical problem

Let $X_{1,sv}$ and $X_{0,sv}$ represent the v-th outcome or, equivalently, the v-th component of the multivariate response, in the stratum s for the treated and the control group respectively, with $s = 0,1$ and $v = 1, \dots, 13$. The partial problem related to the v-th outcome and the stratum s consists of testing

$$H_{0,sv}: X_{1,sv} =^d X_{0,sv} \tag{1}$$

versus the alternative hypothesis

$$H_{1,sv}: X_{1,sv} >^d X_{0,sv}, \tag{2}$$

where $=^d$ denotes equality in distribution and $>^d$ indicates stochastic dominance. Such hypotheses may be written as

$$H_{0,sv}: F_{1,sv}(x) = F_{0,sv}(x)\ \forall x \tag{3}$$

and

$$H_{1,sv}: F_{1,sv}(x) \leq F_{0,sv}(x)\ \forall x \text{ and } \exists x | F_{1,sv}(x) < F_{0,sv}(x)\,, \tag{4}$$

where $F_{j,sv}(x)$ represent the cumulative distribution function of $X_{j,sv}$, with $j = 0,1$.

Under the null hypothesis, for the v-th outcome, both the intra-stratum partial null hypothesis $H_{0,1v}$ and $H_{0,0v}$ are true, thus $H_{0,v}: H_{0,1v} \cap H_{0,0v}$. Similarly, $H_{1,v}: H_{1,1v} \cup H_{1,0v}$, with the same notation. Hence, the overall null and alternative hypothesis of the problem can be denoted by

$$\begin{cases} H_0: \bigcap_{v=1}^{13} H_{0,v} \\ H_1: \bigcup_{v=1}^{13} H_{1,v} \end{cases} \tag{5}$$

For problems of stochastic dominance for ordered categorical variables, the literature offers quite a long list of exact and approximate solutions (see Agresti and Klingenber, 2005; Han et al., 2004; Hirotsu, 1986; Loughin and Scherer, 1998; Loughin, 2004; Lumely, 1996; Nettleton and Banerjee, 2000). For the univariate case, most of the methodological solutions are based on the restricted maximum likelihood ratio test (see Cohen et al., 2000; Silvapulle and Sen, 2005; Wang, 1996). Asymptotic null distributions of the test statistic are mixtures of chi-squared, implying that the mixture's weights depend on the unknown population distribution. In general, this is a complex problem with no easy solution especially among the parametric methods and in particular the likelihood approach.

Nonparametric solutions are proposed by Pesarin (1994), Brunner and Munzel (2000), Pesarin (2001), Troendle (2002), Pesarin and Salmaso (2006), Agresti (1999), and Arboretti and Bonnini (2009). The presented problem can be solved by performing $13 \times 2 = 26$ partial permutation tests (Anderson-Darling type test statistics) and combining the p-values first with respect to the strata (within each variable) and then with respect to the variables (Pesarin, 2010; Pesarin and Salmaso, 2010; Bonnini et al., 2014).

3. Methodological solution

The absolute frequency of the number of statistical units on which such a category is observed (i.e. the j-th ordered category) within the stratum s for the v-th variable in the treated and control group is $f_{1j,sv}$ and $f_{0j,sv}$ respectively. For that reason, the cumulative frequencies in the two groups can be denoted by $F_{1j,sv} = \sum_{r=1}^{j} f_{1r,sv}$ and $F_{0j,sv} = \sum_{r=1}^{j} f_{0r,sv}$ respectively. For the partial test concerning the null and the alternative hypothesis, the Anderson-Darling type test statistic is used

$$T_{sv} = \sum_{j=1}^{k_v-1} \left(F_{0j,sv} - F_{1j,sv}\right)\left[F_{\cdot j,sv}\left(n_s - F_{\cdot j,sv}\right)\right]^{-0.5}, \tag{6}$$

where k_v is the number of ordered categories of the v-th variable, $F_{\cdot j,sv} = F_{0j,sv} + F_{1j,sv}$ and $n_s = F_{\cdot k_v,sv}$. In order to address the testing concerning the v-th variable, specifically testing $H_{0,v}$ against $H_{1,v}$, a first-level combination of the significance level functions of the partial tests of the two strata may be applied. If $L_{sv}(t_{sv}) = P[T_{sv} \geq t_{sv}|\boldsymbol{X}]$ is the significance level function for the s-th stratum and the v-th variable given the observed dataset $\boldsymbol{X}$, for any $t_{sv} \in \mathbb{R}$, according to the permutation distribution, a suitable combined test statistic for the v-th variable is

$$T'_v(t_{1v}, t_{0v}) = \max\left[\left(1 - L_{1v}(t_{1v})\right)\left(1 - L_{0v}(t_{0v})\right)\right], \tag{7}$$

for any couple of values $(t_{1v}, t_{0v}) \in \mathbb{R}^2$, where $L_{sv}(t_{sv}) = P[T_{sv} \geq t_{sv}|H_0, \boldsymbol{X}]$ and $\boldsymbol{X}$ is the observed dataset.

Similarly, to solve the general multivariate problem, a second-level combination concerning the variables may be carried out. Let $L'_v(t'_v) = P[T'_v \geq t'_v|\boldsymbol{X}]$ denote the significance level function of T'_v for any $t'_v \in \mathbb{R}$. The second-level combined test statistic is the following:

$$T''_v(t'_1, \dots, t'_{13}) = \max\left[\left(1 - L'_1(t'_1)\right), \dots, \left(1 - L'_{13}(t'_{13})\right)\right]. \tag{8}$$

In the end, H_0 is rejected if the p-value of the combined test is less than or equal to the significance level $\alpha = 0.05$, specifically if $L''(t''_{obs}) \leq \alpha$, where $L''(t'') = P[T'' \geq t''|\boldsymbol{X}]$ with $t'' \in \mathbb{R}$.

Probabilities and p-values are calculated based on the null permutation distributions derived from permuting the rows of $\boldsymbol{X}$, as the condition of exchangeability holds under the null hypothesis.

Generally, in the case of a multivariate two-sample permutation test, the general CPT procedure in the presence of $q \geq 2$ response variables concern testing $H_0: \bigcap_{j=1}^{q} H_0^{(j)}$ against the alternative hypothesis $H_1: \bigcup_{j=1}^{q} H_1^{(j)}$. The decision rule consists in the rejection of H_0 for large values of T, as follows:

$$\hat{\lambda}_{j,0} = \hat{P}\left(T_j \geq T_{j,obs}|H_0^{(j)}\right) = \hat{L}_j\left(T_{j,obs}\right) \tag{9}$$

$$\hat{\lambda}_{j,b}^{*} = \hat{P}\left(T_j \geq T^{*}_{j,b}|H_0^{(j)}\right) = \hat{L}_j\left(T^{*}_{j,b}\right). \tag{10}$$

An alternative resampling strategy (based on Glivenko-Cantelli theorem) consists in carrying out B (e.g. 10000) independent random samplings from the permutation space (CMC method).

Finally, the combined p-value can be computed in the following way:

$$T^{*}_{comb,b} = \psi\left(\lambda^{*}_{1,b}, \dots, \lambda^{*}_{q,b}; w_1, \dots, w_q\right)$$

with $\psi: (0,1)^{2q} \rightarrow \mathbb{R}$ a suitable combining function, that must satisfy the following conditions:

1. ψ is monotonic decreasing with respect to each argument: $\psi\left(\dots, \lambda_{j,b}, \dots\right) > \psi\left(\dots, \lambda'_{j,b}, \dots\right)$ if $\lambda_{j,b} < \lambda'_{j,b}$,
2. upper and lower bound: $\psi\left(\dots, \lambda_{j,b}, \dots\right) \rightarrow \bar{\psi}$ if $\lambda_{j,b} \rightarrow 0$ and $\psi\left(\dots, \lambda_{j,b}, \dots\right) \rightarrow \underline{\psi}$ if $\lambda_{j,b} \rightarrow 1$,
3. limited acceptance region: $\forall \alpha \in (0,1)\ t_{comb,\alpha} < \bar{\psi} < \infty$.

In the literature, one of the most commonly used combining functions is the combination function of Tippett: $\psi_T = \max_j\left(1 - \lambda_j\right)$.

4. Simulation study

In order to study the power behavior of the proposed methodological approach, consisting of combined permutation tests, a Monte Carlo simulation study is carried out. It mainly requires exchangeability under H_0 and it is distribution-free. It is suitable for both two-sided and one-sided alternatives.

The application of the test, the data random generations, and the comparative performance assessments were carried out through original *R* scripts created by the authors. For the evaluation of the methods, according to the null permutation distribution of the test statistics, 1000 random permutations were carried out. For each setting, 1000 datasets (replicates) were randomly generated.

Sample data were randomly simulated from normal q-variate distributions with null mean vector and variance-covariance matrix Σ and then transformed into q-variate categorical data. Formally, we have:

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

where ρ represents the Pearson correlation and the covariance between each couple of components of the multivariate normal random variable.

Let z_{ti} denote the observation related to the t-th component observed on the i-th statistical unit, with $t = 1, \ldots, q$ and $i = 1, \ldots, n$. The rule of the transformation into categorical variables is as follows:

$$x_{ti} = \begin{cases} \text{category 1 if } z_{ti} \leq -1.5 \\ \text{category 2 if } -1.5 < z_{ti} \leq 0 \\ \text{category 3 if } 0 < z_{ti} \leq 1.5 \\ \text{category 4 if } z_{ti} > 1.5 \end{cases}$$

Therefore, the setting parameters in the simulations are the following:

- n: sample size of the control and the treated group,
- q: number of response variables,
- δ: mean value of each underlying normal random variable for the treated group, i.e. for each component of the q-variate response (for the control group the mean is zero),
- ρ: Pearson correlation index that represents the dependence between each couple of the k components of the multivariate response.

The nominal significance level chosen for the simulations, under all the settings considered in the study, is $\alpha = 0.05$.

Looking at Table 1, we can say that the proposed solution, under the null hypothesis, respect the nominal α level. This means that the test is well approximated.

Table 1: Rejection rates of the proposed solution under H_0, with $\delta = 0$, $\rho = 0.30$ and $\alpha = 0.05$

n	q	Rejection rates
10	5	0.004
100		0.004
10	10	0.008
100		0.002

Source: authors' analysis.

Furthermore, in Figure 1, the consistency of the proposed methodology is shown. The greater the sample size n, the greater the power of the proposed nonparametric procedure. The power tends to 1 when n diverges.

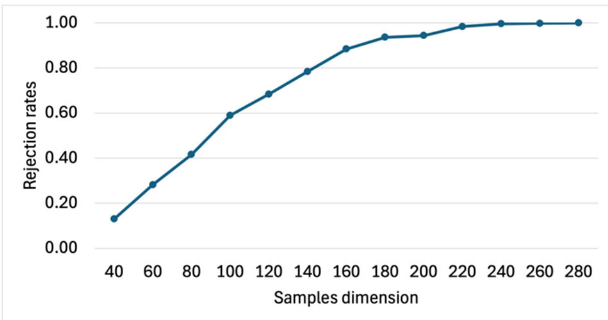

Figure 1: Rejection rates of the proposed test under H_1 as a function of n, with $q = 5$, $\delta = 0.05$, $\rho = 0.30$ and $\alpha = 0.05$

Source: authors' analysis.

The Combined Permutation Tests are exact, unbiased and consistent. CPTs are suitable when the problem can be broken down into several sub-problems (multi-aspect, multivariate, multistrata, multiple comparisons, …) with possible different types of alternatives. Finally, they are suitable for complex alternatives (stochastic ordering, umbrella alternatives, multivariate one-sided test, …), for multivariate numeric, categorical, and mixed variables.

5. Case study

The case study concerns a test on the effect of “assisted motor activity” (AMA) on the health of patients affected by “low back pain” (LBP), “hypertension” and “diabetes”.

The health status is measured according to 13 different binary or ordinal outcomes. The components of the ordinal multivariate response variable are listed below:

- X_1 – Perceived Health: Self-assessment on the general state of health (1-bad, 2-neutral, 3-good),
- X_2 – Moderate Activity: Self-assessment on the ability to perform moderate physical activity (1-no, 2-partial, 3-yes),
- X_3 – Stair Climbing: Self-assessment on the difficulty in stair climbing (1-yes, 2-partial, 3-no),
- X_4 – Physical Performance: Physical performance lower than expected in the last month (1-yes, 2-no),
- X_5 – Activity Limitations: Need to limit some types of activity in the last month (1-yes, 2-no),
- X_6 – Emotional State: Physical performance lower than expected due to emotional state in the last month (1-yes, 2-no),
- X_7 – Mind Concentration: Decrease of mind concentration in the last month due to emotional state (1-yes, 2-no),
- X_8 – Pain: Difficulty in daily activities due to pain in the last month (1-very much, 2-somewhat, 3-undecided, 4-not much, 5-not at all),
- X_9 – Calm and Serenity: Frequency of calm and serenity in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always),
- X_{10} – Full of Energy: Frequency of feeling full of energy in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always),
- X_{11} – Discouraged and Sad: Frequency of feeling discouraged and sad in the last month (1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never),
- X_{12} – Social Activities: Frequency of negative effects of health and emotional state on social activities in the last month (1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never),
- X_{13} – Stress Level: Self-assessment of the level of stress (1-very high, 2-high, 3-average, 4-moderate, 5-low, 6-very low).

In this application, the null hypothesis of no effect of AMA on the health of treated patients (t.p.) is tested against the alternative hypothesis of positive effect of AMA on the health of treated patients. In formula we have the following:

$$H_0\colon [health_1 =^d health_0] \cap [activ_1 =^d activ_0] \cap \ldots \cap [stress_1 =^d stress_0]$$
$$\equiv H_0\colon \bigcap_{v=1}^{13} H_{0,v}$$

vs

$$H_1: [health_1 >^d health_0] \cup [activ_1 >^d activ_0] \cup \dots \cup [stress_1 >^d stress_0]$$
$$\equiv H_0: \bigcup_{v=1}^{13} H_{1,v}.$$

Looking at Table 1, the joint frequency distribution of treatment and confounding factors are reported. It is evident that, considering the stratified methodology, we deal with small sample sizes. Hence, the need to use a non-parametric approach becomes crucial.

Table 2: Joint frequency distribution of treatment and confounding factor

Stratum	**Group**		**Total**
	0	**1**	
Over 60 with LBP (s=1)	10	7	**17**
Oher patients (s=0)	10	20	**30**
Total	**20**	**27**	**47**

Source: authors' analysis.

Each partial test can be broken down into two sub-partial tests, i.e. one for each stratum. Let $s = 1$ and $s = 0$ denote the stratum of over 60 with LBP and other patients respectively:

$$H_0: \bigcap_{v=1}^{13} [H_{0,v0} \cap H_{0,v1}] \equiv H_0: \bigcap_{v=1}^{13} \bigcap_{s=0}^{1} H_{0,vs}$$

vs

$$H_0: \bigcup_{v=1}^{13} [H_{1,v0} \cup H_{1,v1}] \equiv H_0: \bigcup_{v=1}^{13} \bigcup_{s=0}^{1} H_{1,vs}.$$

Summarizing the characteristics of the problem, we are in the presence of a multivariate test with 13 marginal variables, categorical data, a directional alternative hypothesis (stochastic dominance), a stratified test to avoid confounding effects and small sample sizes.

The application of the described method leads an overall p-value of 0.019. Thus, we have empirical evidence in favor of the hypothesis of significant effect of AMA on the health of patients (i.e. significant effect of the treatment).

To attribute the overall significance to some specific outcomes (second level partial tests) we need to adjust the partial p-values to control the Family Wise Error Rate (FWER), i.e. the probability of wrong rejection of one or more partial null hypotheses.

According to the Bonferroni-Holm procedure the adjusted p-values are computed as follows:

$$\tilde{p}_{(k)} = \max_{i \leq k} \left\{ \min\left[1, (13 + 1 - i) p_{(i)}\right] \right\}, \tag{11}$$

where $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(q)}$ are the sorted p-values of the partial tests.

Table 2: Adjusted p-values of the partial tests (significance in bold)

Test	Variable	Adjusted p-value
T_1	Perceived Health	0.361
T_2	Moderate Activity	0.357
T_3	Stair Climbing	0.232
T_4	Physical Performance	0.444
T_5	Activity Limitations	0.444
T_6	Emotional State	0.444
T_7	Mind Concentration	0.304
T_8	Pain	0.444
T_9	Calm and Serenity	0.444
T_{10}	Full of Energy	0.444
T_{11}	Discouraged and Sad	**0.019**
T_{12}	Social Activities	0.444
T_{13}	Stress Level	0.258

Source: authors' analysis.

According to Table 2, the positive effect of the treatment on the health of patients can only be attributed to a decrease in sadness and discouragement.

6. Conclusions

This work aims to test the null hypothesis of no effect of AMA on the health of treated patients versus the alternative hypothesis of a positive effect of AMA on the health of treated patients. The state of health is a multivariate categorical variable and for such a problem of stochastic dominance, the combined permutation test method is a valid solution. The proposed approach represents a valid solution for the described testing problem, whose complexity is due to the multivariate nature of the response, the categorical data, the directional alternative hypothesis, the presence of confounders and the small sample sizes. The application to the case study provides an overall p-value of the CPT equal to 0.019, which provides empirical evidence in favor of the hypothesis of the significant effect of AMA on the health of patients. Furthermore, looking at the partial p-values, the positive effect of the treatment on the health of patients can only be attributed to a decrease in sadness and discouragement. The good power behavior and in particular the consistency of the proposed test are proved in the simulation study.

The proposed method offers several promising directions for future research. First, it could be extended to accommodate longitudinal multivariate categorical data, allowing for the analysis of repeated measures or follow-up studies while preserving the nonparametric nature of the test. Moreover, integration with hierarchical or mixed-effects frameworks would enable its application in multi-center studies or clustered data settings. Another valuable advancement would be the development of software packages to facilitate wider adoption of the method in applied research. Robust handling of missing data through imputation strategies combined with permutation logic also represents a key area for enhancement. Furthermore, the extension of the method to comparisons involving more than two groups could broaden its applicability in complex experimental designs. In terms of application, the approach may prove particularly useful in the evaluation of public policy interventions using categorical indicators in socio-economic studies. Additional relevant fields include educational research (e.g. assessing student outcomes through ordinal rating scales), marketing and consumer behavior (e.g. preference or satisfaction surveys), and environmental epidemiology, where categorical health outcomes are compared across groups exposed to different environmental conditions. These areas all stand to benefit from a flexible, robust, and distribution-free testing procedure such as the one proposed in this study.

References

Alibrandi, A.; Giacalone, M.; Zirilli, A., (2022). Psychological stress in nurses assisting Amyotrophic Lateral Sclerosis patients: A statistical analysis based on Non-Parametric Combination test. *Mediterr. J. Clin. Psychol.*, 10.

Agresti, A., (1999). Modelling ordered categorical data: recent advances and future challenges. *Statistics in medicine*, 18(17-18), pp. 2191–2207.

Agresti, A., Klingenberg, G. B., (2005). Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *J. R. Stat. Soc.*, Ser. C 4, pp. 691–706

Arboretti Giancristofaro, R., and Bonnini, S., (2009). Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Statistical Methods and Applications*, 18, pp. 221–236.

Bonnini, S., Borghesi, M., (2022). Relationship between Mental Health and Socio-Economic, Demographic and Environmental Factors in the COVID-19 Lockdown Period—A Multivariate Regression Analysis. *Mathematics*, 10, pp. 3237.

Bonnini, S., Borghesi, M., and Giacalone, M., (2023). Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with application to Circular Economy. *Ann. Oper. Res.*, 342, pp. 1655–1670.

Bonnini, S., Borghesi, M., and Giacalone, M., (2024a). Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments. *Appl. Stoch. Models Bus. Ind.*, 40, pp. 389–407.

Bonnini, S., Borghesi, M., and Giacalone, M., (2024b). Semi-parametric approach for modeling overdispersed count data with application to industry 4.0. *Socio-Econ. Plan. Sci.*, 95, p. 101976.

Bonnini, S., Corain, L., Marozzi, M., and Salmaso, L., (2014). *Nonparametric Hypothesis Testing. Rank and Permutation Methods with Applications in R.* Wiley, Chichester.

Bonnini, S., Assegie, G. M., and Trzcinska, K., (2024). Review about the Permutation Approach in Hypothesis Testing. *Mathematics*, 12, pp. 2617–1.

Brunner, E., Munzel, U., (2000). The nonparametric Behrens–Fisher problem: asymptotic theory and small-sample approximation. *Biom. J.*, 42, pp. 17–25

Cohen, A., Kemperman, J. H. B., Madigan, D., and Sakrowitz, H. B., (2000). Effective directed tests for models with ordered categorical data. *Aust. N. Z. J. Stat.*, 45, pp. 285–300

Ferreira, M. L., Ferreira, P. H., Latimer, J., Herbert, R. D., Hodges, P. W., Jennings, M. D., ... and Refshauge, K. M., (2007). Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomized trial. *Pain*, 131(1-2), pp. 31–37.

Gordon, R., Bloxham, S., (2016, April). A systematic review of the effects of exercise and physical activity on non-specific chronic low back pain. In *Healthcare* (Vol. 4, No. 2, p. 22). MDPI.

Han, K. E., Catalano, P. J., Senchaudhuri, P., and Metha, C., (2004). Exact analysis of dose-response for multiple correlated binary outcomes. *Biometrics,* 60, pp. 216–224

Hirotsu, C., (1986). Cumulative chi-squared statistic as a tool for testing goodness-of-fit. *Biometrika,* 73, pp. 165–173

Loughin, T. M., (2004). A systematic comparison of methods for combining p-values from independent tests. *Comput. Stat. Data Anal.*, 47, pp. 467–485

Loughin, T. M., Scherer, P. N., (1998). Testing for association in contingency tables with multiple column responses. *Biometrics*, 54, pp. 630–637

Lumely, T., (1996). Generalized estimating equations for ordinal data: a note on working correlation structures. *Biometrics*, 52, pp. 354–361

Macedo, L. G., Latimer, J., Maher, C. G., Hodges, P. W., McAuley, J. H., Nicholas, M. K., ... and Stafford, R., (2012). Effect of motor control exercises versus graded activity in patients with chronic nonspecific low back pain: a randomized controlled trial. *Physical therapy*, 92(3), pp. 363–377.

Macedo, L. G., Maher, C. G., Hancock, M. J., Kamper, S. J., McAuley, J. H., Stanton, T. R., ... and Hodges, P. W., (2014). Predicting response to motor control exercises and graded activity for patients with low back pain: preplanned secondary analysis of a randomized controlled trial. *Physical therapy*, 94(11), pp. 1543–1554.

Marcus, R., Eric, P., and Gabriel, K. R., (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3), pp. 655–660

Nettleton, D., Banerjee, T., (2000). Testing the equality of distributions of random vectors with categorical components. *Comput. Stat. Data Anal.*, 37, pp. 195–208

Pesarin, F., (1994). Goodness-of-fit testing for ordered discrete distributions by resampling techniques. *Metron,* 52, pp. 57–71.

Pesarin, F., (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley: Chichester.

Pesarin, F, Salmaso, L., (2006). Permutation tests for univariate and multivariate ordered categorical data. *Austrian Journal of Statistics,* 35, pp. 315–324.

Pesarin F., Salmaso L., (2010). *Permutation Tests for Complex Data. Theory, Applications, and Software*. Wiley, Chichester.

Silvapulle, M. J., Sen, P. K., (2005). *Constrained Statistical Inference, Inequality, Order, and Shape Restrictions*. Wiley, New York

Simon, N., Tibshirani, R., (2012). A permutation approach to testing interactions in many dimensions. *arXiv*, arXiv, 1206.6519.

Stute, W., Thies, S., and Zhu, L., (1998). Model checks for regression: An innovation process approach. *Ann. Statist.*, 26, pp. 1916–1934.

Toma, P., Miglietta, P. P., Zurlini, G., Valente, D., and Petrosillo, I., (2017). A non-parametric bootstrap-data envelopment analysis approach for environmental policy planning and management of agricultural efficiency in EU countries. *Ecol. Indic*., 83, pp. 132–143.

Troendle, J. F., (2002). A likelihood ratio test for the nonparametric Behrens–Fisher problem. *Biom. J.,* 44(7), pp. 813–824

Wang, Y., (1996). A likelihood ratio test against stochastic ordering in several populations. *J. Am. Stat. Assoc*., 91, pp. 1676–1683