mathematics

MDPI

Article

A Non-Parametric Test for a Two-Way Analysis of Variance

Stefano Bonnini [1], Michela Borghesi [1], Gianfranco Piscopo [2,*], Massimiliano Giacalone [3]

[1] Department of Economics and Management, University of Ferrara, Via Voltapaletto 11, 44121 Ferrara, Italy; stefano.bonnini@unife.it (S.B.); michela.borghesi@unife.it (M.B.)
[2] Department of Mathematics and Applications "Renato Caccioppoli", University of Naples "Federico II", Via Cintia, Monte S. Angelo, 80126 Napoli, Italy
[3] Capua (CE) - Department of Economy, University of Campania "Luigi Vanvitelli", Corso Gran Priorato di Malta, 81043 Capua CE, Italy; massimiliano.giacalone@unicampania.it
* Correspondence: gianfranco.piscopo@unina.it

Abstract: The methodology carried out in this work is based on non-parametric inference. The problem is framed as a regression analysis, and the solution is derived using the permutation approach. The proposed test does not rely on the assumption that the distribution of the response follows a specific family of probability laws, unlike other parametric approaches. This makes the test powerful, particularly when the typical assumptions of parametric approaches, such as the normality of data, are not satisfied and parametric tests are not reliable. Furthermore, this method is more flexible and robust with respect to parametric tests. A permutation test on the goodness-of-fit of a multiple regression model is applied. Hence, proposed solution consists of the application of permutation tests on the significance of the single coefficients and then a combined permutation test (CPT) to solve the overall goodness-of-fit testing problem. Furthermore, a Monte Carlo simulation study was performed to evaluate the power of the previously mentioned permutation approach, comparing it with the conventional parametric *F*-test for ANOVA and the bootstrap combined test, both commonly discussed in the literature on this statistical problem. Finally, the proposed non-parametric test was applied to real-world data to investigate the impact of age and smoking habits on medical insurance costs in the USA. The findings suggest that smoking and being at least 50 years old significantly contribute to increased medical insurance costs.

Keywords: non-parametric inference; permutation tests; ANOVA

MSC: 62J10; 62G10; 62J05

Academic Editor: MinJae Lee

Received: 30 January 2025
Revised: 19 March 2025
Accepted: 27 March 2025
Published: 29 March 2025

Citation: Bonnini, S.; Borghesi, M.; Piscopo, G.; Giacalone, M. A Non-Parametric Test for a Two-Way Analysis of Variance. *Mathematics* **2025**, *13*, 1131. https://doi.org/10.3390/math13071131

1. Introduction

The paper deals with a two-way analysis of variance, in the presence of two two-level factors and a numeric response. The motivating example of this work concerns the US population's medical insurance cost, and the effect on it of age and being a smoker. Recently, the actuarial modelling of insurance claims has emerged as a significant research topic in the health insurance industry and is primarily used for establishing appropriate premiums [1]. This is crucial for enticing and keeping insured individuals, as well as for effectively managing current plan members. Nevertheless, because of the numerous elements influencing medical insurance expenses and the complexities involved, there are challenges in precisely developing a predictive model for it. Elements such as demographic data, health conditions, access to healthcare based on location, lifestyle habits, and characteristics of providers can significantly influence the anticipated expenses of medical

insurance. Additional important aspects such as the extent of coverage, the kind of plan, the deductible amount, and the customer's age at the time of enrollment also significantly influence the possible expenses associated with medical insurance [2].

In some countries, such as the United States, it is nowadays essential for the population to have health insurance [3,4]. One of the largest yearly expenses individuals face is their health insurance coverage. Health insurance represents a third of the GDP, and medical care is necessary for everyone at different levels. Shifts in medical practices, trends in pharmaceuticals, and political influences are just some of the numerous factors that create yearly variations in healthcare expenses [5]. Some research indicates that insurance companies ought to adjust their policies based on the smoking habits, BMI, and age of their clients in order to offer insurance plans that cater to the public's needs [6].

Grasping the factors that impact an individual's health insurance premium is essential for insurance companies to set appropriate prices. When making informed decisions, the premium should always be the primary focus for users. Several empirical works in the literature investigate the main factors that significantly correlate with health insurance costs. Charges are positively related to smoking and age. The importance of health insurance is growing as individuals' lifestyles and health issues evolve. Since a medical condition can affect anyone unexpectedly and can have significant psychological and financial repercussions, it is challenging to determine when these issues will arise. Given this context, some authors [7] aimed to predict health insurance costs based on the following factors: age, sex, geographical location, smoking status, body mass index (BMI), and number of children [8,9]. Age, BMI, chronic diseases, and family history of cancer have been shown to be the most influential factors determining the premium price [10,11].

One of the goals of the paper is that the empirical findings of this study will assist policymakers, insurers, and prospective medical insurance purchasers in making informed choices about selecting policies that fit their individual requirements. To study the effect of age and smoking on insurance costs, we use sample data. The problem under study can be traced back to a two-way ANOVA layout. We are interested in testing the significance of both the main effect of each factor and the interaction between the two. Some solutions have been carried out in the methodological literature for the ANOVA problem. Some researchers have examined the two-way ANOVA model with varying cell frequencies without assuming equal error variances. By employing a generalized approach to calculate p-values, the traditional F-tests on the interaction and main effects are expanded to accommodate heteroscedasticity. The generalized F-tests introduced in the aforementioned study can be applied in significance testing or in fixed level testing according to the Neyman–Pearson framework. Simulation studies reveal that, although the method shows enhanced power in the presence of heteroscedasticity, the size of the test does not surpass the predetermined significance level [12]. Another work states that determining the a priori power for univariate repeated measures (RM) ANOVA designs with two or more factors is currently difficult due to the lack of accurate methods for estimating the error variances used in power calculations. Hence, Monte Carlo simulation procedures have been used to estimate power under different experimental conditions [13]. The main problem remains that this method provides only an estimate and not an adequate procedure for the statistical problem. In another study by Toothaker and Newman (1994) [14], the ANOVA F-test and several non-parametric competitors for two-way design were compared in terms of empirical α and power. Through a simulation study, they confirmed that the ANOVA F-test suffers from conservative α and power for the mixed normal distribution.

As said, in this work, the goal is to investigate the main and interaction effects of crucial factors, such as being a smoker (yes or no) and age (less than 50 years old and greater than or equal to 50 years old) in an experimental framework where the response

variable is the charges of medical insurance costs. The proposed methodological approach is based on a non-parametric inference. The problem is framed within the context of a regression analysis, and the resolution relies on a permutation method. The suggested test, in contrast to parametric methods, does not necessitate the assumption that the response distribution adheres to a particular set of probability distributions. This type of test is highly effective, particularly when the usual assumptions of parametric methods (like data normality) are unmet and the validity of parametric tests is compromised [15]. Moreover, the proposed approach demonstrates greater flexibility and robustness compared to parametric tests, and it can be viewed in opposition to stepwise regression [16]. Actually, it is based on a permutation test to evaluate the goodness-of-fit of a multiple regression model. This solution is based on a multiple testing approach, that jointly assesses the significance of all single regression coefficients, including both main effects and interaction effects. The method consists of combining the p-values of the partial tests on the regression coefficients. The approach of combined permutation tests has been effectively implemented across various contexts [17,18]. In particular, it has been widely applied in empirical studies [19], with numeric variables but also categorical data [20], for big data problems [21], in regression analysis [22,23], to test directional and non-monotonic hypotheses [24,25], with count data [24] and in many other problems.

Permutation goodness-of-fit assessments, using partial sums or cumulative sums of residuals, have been suggested for linear regression models [26]. In order to evaluate the influence of covariates, the authors of [27] introduced the concept of combining non-parametric permutation tests. The notion of treating the test for the validity of a multivariate linear model as a simultaneous test was put forward by [28] within the context of rotation tests. Moreover, in the literature, several recent articles discuss approaches for assessing goodness-of-fit through the permutation approach in linear regression models, as well as the use of non-parametric permutation tests to evaluate the influence of covariates. For instance, [29] proposes a non-parametric test to assess the effect of covariates on the cure rate in mixture cure models, employing a bootstrap method to approximate the null distribution of the test statistic. On the other hand, a previous study generalizes the metric-based permutation test for the equality of covariance operators to multiple samples of functional data, utilizing a non-parametric combination methodology to merge pairwise comparisons into a global test [30].

The rest of the paper is organized as follows. Section 2 presents the statistical problem. The methodological proposed solution is explained in Section 3. Section 4 contains the Monte Carlo simulation study. The case study is described in Sections 5 and 6 includes the overall conclusions.

2. Statistical Problem

Let us consider a model with binary variables as predictors, to represent the main effects and the interaction effects of two factors with two levels. This situation was first considered by [31] because it is typical of experimental designs where subjects are first divided into homogeneous subgroups (blocks) and then randomly assigned to various treatment levels. Under the null hypothesis, all the treatment effects are null. On the other hand, in the alternative hypothesis, at least one effect is not equal to zero.

Let x_{kji} denote the i-th observation of the response variable related to the combination of levels (k, j), that is the value observed on the i-th unit (i-th replication) when factor 1 is at level k and factor 2 is at level j ($k = 1, \ldots, K$; $j = 1, \ldots, C$). Data are assumed to be realizations of random variables X_{kji} which behave according to the following linear model:

$$X_{kji} = \mu + \theta_k + \gamma_j + (\theta\gamma)_{kj} + \varepsilon_{kji}, \quad (1)$$

where θ_k and γ_j are the main effects of factor 1 at level k and factor 2 at level j, respectively, $(\theta\gamma)_{kj}$ is the interaction effect related to factor 1 at level k and factor 2 at level j and ε_{kji} are exchangeable errors, with zero mean and unknown continuous distribution [15]. In this situation, without loss of generality, for ease of interpretation, it is also possible to consider the usual constraints:

$$\sum_{k=1}^{K} \theta_k = 0, \ \sum_{j=1}^{C} \gamma_j = 0, \ \sum_{k=1}^{K} (\theta\gamma)_{kj} = 0 \ \forall j, \ \sum_{j=1}^{C} (\theta\gamma)_{kj} = 0 \ \forall k. \tag{2}$$

Typically, three potential testing problems could be examined independently:

1. $H_{0\theta} : [\theta_k = 0 \text{ for every } k]$, against $H_{1\theta} : [\theta_k \neq 0 \text{ for some } k]$ (significance of main effect of factor 1);
2. $H_{0\gamma} : \left[\gamma_j = 0 \text{ for every } j\right]$, against $H_{1\gamma} : \left[\gamma_j \neq 0 \text{ for some } j\right]$ (significance of main effect of factor 2);
3. $H_{0\theta\gamma} : \left[(\theta\gamma)_{kj} = 0 \text{ for every } (k,j)\right]$, against $H_{1\theta\gamma} : \left[(\theta\gamma)_{kj} \neq 0 \text{ for some } (k,j)\right]$ (significance of interaction effect).

Under the null hypothesis related to one of the three problems, exchangeability is applicable only among certain data blocks (determined by the $C \times K$ combinations of levels), and thus only synchronized permutations are permitted [32]. For instance, if we want to compare two treatments of factor 1, corresponding to levels k_1 and k_2 of the factor, where $1 \leq k_1 < k_2 \leq K$, we can shuffle data between blocks having the same level of factor 2, one featuring factor 1 at level k_1 and the other at level k_2. In other words, by representing the data block with (k, j) as having factor 1 at level k and factor 2 at level j, we can interchange data between (k_1, j) and (k_2, j) for each j, ensuring that the permutations between the C pairs of blocks remain synchronized. These synchronized permutations are a valid alternative because it is among the permutation methods used for the two-way (M)ANOVA. However, this method only works when the aim is to test only one factor and consider the others as confounders. This is different from our proposal because we want to verify the significance of the estimates of all the single coefficients jointly considered.

The data can be alternatively represented using a different notation. Let $y_1, y_2, \ldots, y_n$ be the observed data of the response, where n is the total number of observations and y_i is the i-th observation of the response, with $i = 1, \ldots, n$. If we consider the case of two factors at two levels, i.e., $K = C = 2$, and assume y_i to be a realization of the random variable Y_i, an alternative linear model in the framework of the linear regression analysis is the following:

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_{12} D_{i1} D_{i2} + \varepsilon_i, \tag{3}$$

where D_{iv} is a dummy variable that takes the value 0 if the i-th observation is associated to the v-th factor at level 1, and takes the value 1 if the i-th observation is related to the v-th factor at level 2, with $i = 1, \ldots, n$ and $v = 1, 2$. In the regression model, $\varepsilon_1, \ldots, \varepsilon_n$ are exchangeable random errors and $\beta_0, \beta_1, \beta_2, \beta_{12}$ are unknown parameters, with β_0 as the expected value of the response corresponding to the baseline (both the factors at level 1), β_1 and β_2 as the main effects of factor 1 and factor 2, respectively, and β_{12} as the interaction effect.

3. Permutation Solution

Regarding the regression coefficients of the reparameterized model (3), we are interested in the following system of hypotheses:

$$\begin{cases} H_0: & \beta_1 = \beta_2 = \beta_{12} = 0 \\ H_1: & \overline{H_0} \end{cases}. \tag{4}$$

Given that we aim to determine which coefficients have significant estimates when the null hypothesis is rejected, our approach involves conducting multiple tests to assess the significance of the single regression coefficients. A solution to the aforementioned statistical problem concerns the use of a combined permutation test (CPT) [33]. The overall testing problem can be considered a multiple test, composed by the partial tests on the significance of the regression coefficients' estimates [34].

The system of hypotheses of the partial test concerning the main effect of the vth factor is

$$\begin{cases} H_0^v : \beta_v = 0 \\ H_1^v : \beta_v \neq 0 \end{cases}$$

with $v = 1, 2$. Similarly, the system of hypotheses of the partial test concerning the interaction effect of the two factors is

$$\begin{cases} H_0^{12} : \beta_{12} = 0 \\ H_1^{12} : \beta_{12} \neq 0. \end{cases}$$

Therefore, the null and alternative hypotheses of the overall problem can be represented as follows:

$$\begin{cases} H_0: & H_0^1 \bigcap H_0^2 \bigcap H_0^{12} \\ H_1: & H_1^1 \bigcup H_1^2 \bigcup H_1^{12}, \end{cases} \tag{5}$$

where the intersection symbols mean that all the partial null hypotheses are true under H_0, and the union symbols indicate that, under H_1, at least one partial alternative hypothesis is true.

The proposed solution consists of the application of permutation tests on the significance of the single coefficients and then a CPT to solve the overall testing problem by combining the inferential results of the partial tests. Reasonable test statistics for the partial tests on the single coefficients are the absolute values of suitable estimators of the parameters. Formally, an appropriate partial test statistic for the main effect of the vth factor is

$$T_v = |\hat{\beta}_v|$$

with $v = 1, 2$. Therefore, a suitable partial test statistic for the interaction effect is

$$T_{12} = |\hat{\beta}_{12}|.$$

The null distribution of the test statistics is obtained by permuting the rows of the design matrix, keeping the vector of observed values of the response fixed, and recalculating for each permutation the estimates of the coefficients and, consequently, the values of the test statistics. This non-parametric method is very useful for solving complex problems and the main advantage with respect to parametric methods is that the multivariate distribution of the trivariate test statistic does not need to be known, and in particular, the dependence structure between variables does not need to be explicitly modelled or specified [35].

Permutation tests are distribution-free, hence they are flexible and robust with respect to the departure from normality [32].

The core concept of a combined permutation test is to identify an appropriate statistic for each partial test and to aggregate the permutation p-values from these tests to address the overall problem [33]. In our case, the absolute values of the least squares estimators of the regression coefficients serve as suitable and effective test statistics for the partial tests. The procedure follows these steps:

1. Computation of the vector of observed values of the test statistics $T_0 = |b_1|, |b_2|, |b_{12}| = T(X)$;
2. B independent random permutations of the rows of the X matrix: $X_1^*, X_2^*, \ldots, X_B^*$;
3. Computation of the values of the test statistic vector for the B dataset permutations T_b^* and the corresponding vector of p-values λ_b^* with $b = 1, 2, \ldots, B$;
4. Computation of the value of the combined test statistic for each permutation and for the observed dataset through the combination of the partial p-values with a suitable function ψ: $T_{\text{comb}} = \psi(\lambda_b^*)$;
5. Computation of the p-value of the combined test according to the null permutation distribution.

The dependence of the partial tests is implicitly taken into account through the permutation of the rows of the design matrix. Since each partial null hypothesis is rejected for large values of the test statistic, without loss of generality, we assume the same rejection rule for the combined test statistic of the overall problem. Therefore, a suitable combined test statistic ψ is:

$$T_{\text{comb}} = -2\sum_j \ln(\lambda_j) = -2 \cdot [\ln(\lambda_1) + \ln(\lambda_2) + \ln(\lambda_{12})], \tag{6}$$

where λ_1, λ_2 and λ_{12} are the p-values of the partial tests. Equation (6) represents Fisher's combination function which, among the various combination functions available in the literature, exhibits a good power behavior regardless of the proportion of true partial alternative hypothesis [36,37]. Since such a permutation ANOVA for the linear regression models is defined as a multiple tests, in the case of a rejection of the null hypothesis in favor of the alternative and, in order to attribute the overall significance to specific partial tests (i.e., to specific coefficient estimates), the control of the family-wise error (FWE) is necessary [38,39]. Family-wise error rate (FWE) refers to the probability of making at least one type I error (false positive) when performing multiple statistical tests on the same dataset [40]. It is a key concept in multiple comparisons and hypothesis testing, particularly when testing multiple hypotheses simultaneously. When conducting multiple statistical tests, the likelihood of incorrectly rejecting at least one true null hypothesis (type I error) increases. In other words, to avoid the inflation of the type I error of the overall test, we must adjust the partial p-values. A suitable method is the one based on the Bonferroni–Holm rule [41]. This correction procedure works as follows. Given a set of m partial p-values:

1. all p-values are sorted in order of smallest to largest;
2. if the smallest p-value is greater than or equal to α/m, the procedure is stopped and no p-value must be considered significant. Otherwise, go to step three;
3. the smallest p-value is considered significant, then the second smallest p-value is compared to $\alpha/(m-1)$. If the second smallest p-value is greater than or equal to $\alpha/(m-1)$, the procedure is stopped and no other p-value is considered significant. Otherwise, go to step four;
4. repeat steps 2 and 3 on the remaining p-values until you find the first non-significance.

The method's application to the real data of the case study, along with the simulation analysis, was conducted using original R scripts developed by the authors. Actually, for the regression model, the R function *lm()* was carried out. The graphs were created with the corresponding R functions *plot()*, *hist()* and *boxplot()*.

4. Monte Carlo Simulation Study

A Monte Carlo simulation analysis was conducted to assess the power behavior of the non-parametric test discussed in the paper (CPT) in comparison to the traditional parametric F-test for the ANOVA and to the bootstrap combined test (TBC), also present in the literature concerning this statistical problem. The previously mentioned competitor, TBC, was implemented by taking inspiration from the technique reported in [42,43]. In practice, this method, unlike CPT, generates the B simulated datasets by resampling the original data with replacement. Once the bootstrapped datasets are generated, the technique for combining the partial p-values, related to the significance of the individual regression coefficients, remains the same as in the CPT procedure.

Let n and q denote the sample size and the number of binary factors of the model, respectively (in our specific case $q = 2$). The binary data were simulated by randomly generating values from normal distributions and then transforming such values into binary data. Specifically, $n \times q$ matrix $\mathbf{Z}$ was simulated by randomly generating n observations from a q-variate normal distribution with null mean vector and variance–covariance matrix $\mathbf{\Sigma}$, i.e., $\mathbf{Z} \sim \mathcal{N}_q(\mathbf{0_q}, \mathbf{\Sigma})$. We assumed that the variance of each component of the underlying bivariate normal distribution was constant and equal to σ_z^2 and the correlation between each couple of components ρ_z. Consequently, $\mathbf{\Sigma} = \sigma_z^2[\rho_z \mathbf{J_q} + (1 - \rho_z)\mathbf{I_q}]$, where $\mathbf{J_q}$ denotes the $q \times q$ all-ones matrix and $\mathbf{I_q}$ is the $q \times q$ identity matrix. In our simulations, we set $\sigma_z = 1$ and $\rho_z = 0.30$. According to the extensive notation, we have:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_z^2 & \sigma_z^2\rho_z & \dots & \sigma_z^2\rho_z \\ \sigma_z^2\rho_z & \sigma_z^2 & \dots & \sigma_z^2\rho_z \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_z^2\rho_z & \sigma_z^2\rho_z & \dots & \sigma_z^2 \end{pmatrix}.$$

Then, the model design matrix $\mathbf{X}$ was obtained by transforming the elements of $\mathbf{Z}$ into binary data and adding the column of ones and the products between factors to take into account the constant of the model and the interaction, respectively. As said, we focused on the case $q = 2$. Let θ_1 and θ_2 denote the probability of success (or population proportion of ones) of the underlying Bernoulli random variables corresponding to the first and the second factors, respectively. The binary simulated value x_{iv} corresponding to the ith observation of the vth factor was computed as follows:

$$x_{iv} = I_{(-\infty, \Phi^{-1}(\theta_v))}(z_{iv}/\sigma_z),$$

where $I_A(\cdot)$ is the indicator function of the set A, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and $\Phi^{-1}(\cdot)$ its inverse, that is the Gaussian quantile function, with $v = 1, 2$. In the simulations, we considered the realistic values $\theta_1 = 0.30$ and $\theta_2 = 0.20$. Hence, the simulated matrix of predictors is the following:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} \end{pmatrix}.$$

In order to simulate the values of the numeric response, we assumed three different distributions for the model random errors, mainly to evaluate the effect of the distribution skewness on the inferential results. The n i.i.d. errors, simulated from probability distributions with different skewness, were rescaled to respect a distribution variance of $\sigma_\varepsilon = 12$, very similar to that of the case study, and centred on the mean value 0. Let ε_i and $\tilde{\varepsilon}_i$ denote the error randomly generated and the transformed one (rescaled and centered), respectively. The three cases are:

1. $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, i.e., errors generated from a normal distribution, such that $skewness = 0$, then $\tilde{\varepsilon}_i = \varepsilon_i$;
2. $\varepsilon_i \sim \chi^2_{13}$, i.e., errors generated from a chi-square distribution with 13 degrees of freedom, such that $skewness = \sqrt{8/13} = 0.78$ (moderate asymmetry), then $\tilde{\varepsilon}_i = (\varepsilon_i - 13)\sigma_\varepsilon/\sqrt{26}$;
3. $\varepsilon_i \sim \chi^2_2$, i.e., errors generated from a chi-square distribution with 2 degrees of freedom, such that $skewness = \sqrt{8/2} = 2$ (high asymmetry), then $\tilde{\varepsilon}_i = (\varepsilon_i - 2)\sigma_\varepsilon/2$.

The values of the dependent variable were obtained by adding the simulated deterministic part of the model and the transformed simulated errors, according to the regression model reported in Equation (3). Hence, the vector of n simulated values of the response is computed as follows:

$$\tilde{y} = \mathbf{X}\beta + \tilde{\varepsilon},$$

where $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n)'$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})'$, and $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)'$.

For every simulation setting, the total number of datasets created for power estimation was 1000, and the count of random permutations to establish the null distribution of the test statistics was $B = 1000$. The approximation would certainly improve with B = 10,000, which is the choice we made in the case study. Nevertheless, using 1000 permutations in the conditional Monte Carlo method is adequate to achieve a satisfactory level of approximation, ensuring reliable outcomes from the simulations [15,33] and computational efficiency. To address the computational limit of the permutation approach and guarantee both robustness and feasibility of the procedure, either Monte Carlo approximations or parallel computing techniques could be also considered [44,45]. In the simulations, a comparative performance analysis of the CPT method and the classic F-test of the parametric approach was carried out.

The setting parameters which vary in the simulations are the following:

- n is the sample size,
- β_0, β_1, β_2 and β_{12} are the coefficients of the regression equation.
- $skewness = \sqrt{8/df}$, where df represents the degrees of freedom of the chi-square distribution of errors ($\to \infty$ in case of normality)

Firstly, simulations were carried out under the null hypothesis (H_0), with $\beta_0 = 13$ and $\beta_1 = \beta_2 = \beta_{12} = 0$. Figure 1 reports the rejection rates of the compared tests as a function of the sample size n. The rejection rate refers to the proportion of times H_0 is rejected in a simulation. Indeed, under H_0, these values should remain below the predefined significance level α to prove that the tests are well approximated. It is evident that all three tests are well approximated because the rejection rates are all under or very close to the significance level $\alpha = 0.05$. Even though the competitor TBC is often above or close to the reference value α.

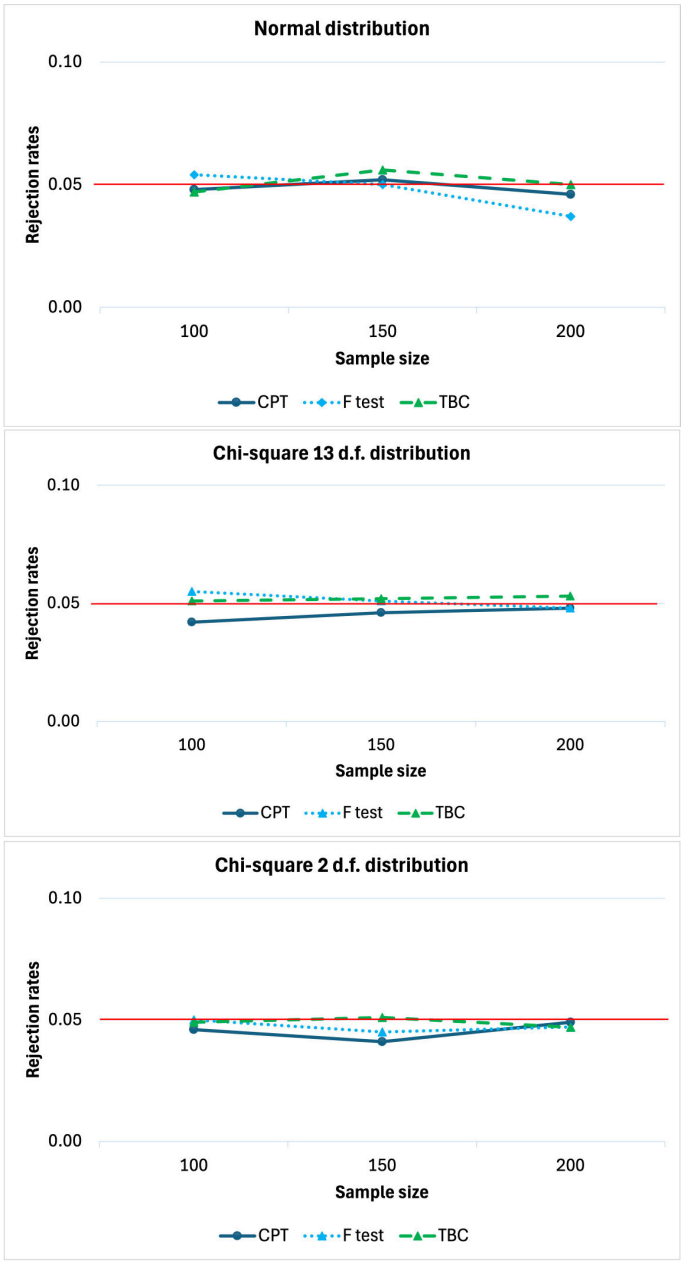

Figure 1. Rejection rates under H_0, with $\rho_z = 0.3$, $\sigma_z = 1$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\beta_0 = 13$, $\beta_1 = \beta_2 = \beta_{12} = 0$ and $\alpha = 0.05$.

We studied the power behavior under H_1 in the scenario with positive main effects of the two factors and null interaction effect, similarly to the empirical results of the case study. The power of a statistical test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. This probability is typically represented by a curve that illustrates the test's power as a function of sample size, number of partial tests, or other parameters of interest. The power under such a scenario, as a function of the sample size, is reported in Figure 2 to evaluate the consistency of the test. In general, all the three tests are consistent. The power of the proposed method is slightly greater than that of the parametric F-test and also the non-parametric method based on a bootstrap approach (TBC). Furthermore, when n diverges, the proposal tends to infinity faster. We carried out the simulations under the same scenarios except $\beta_{12} = 10$ (to consider the case of positive interaction effect as well). The results confirm that the CPT is more powerful than the

parametric F-test and than the TBC approach. Additionally both of them are consistent (see Appendix A).

Furthermore, in Figure 3, the effect of skewness can be assessed when $n = 100$. The rejection rates of the three tests are represented as a function of the asymmetry of the error term distribution. We remind that the skewness is null when the errors are simulated under normality, and corresponds to $\sqrt{\frac{8}{df}}$ when the assumed error distribution is chi-square with df degrees of freedom. We considered four scenarios, by simulating under normality and under χ^2_{13}, χ^2_8, and χ^2_2, corresponding to the skewness values 0, 0.78, 1 and 2, respectively. Again, the CPT method seems to be more powerful regardless of the departure of the distribution skewness from the symmetric case typical of normality.

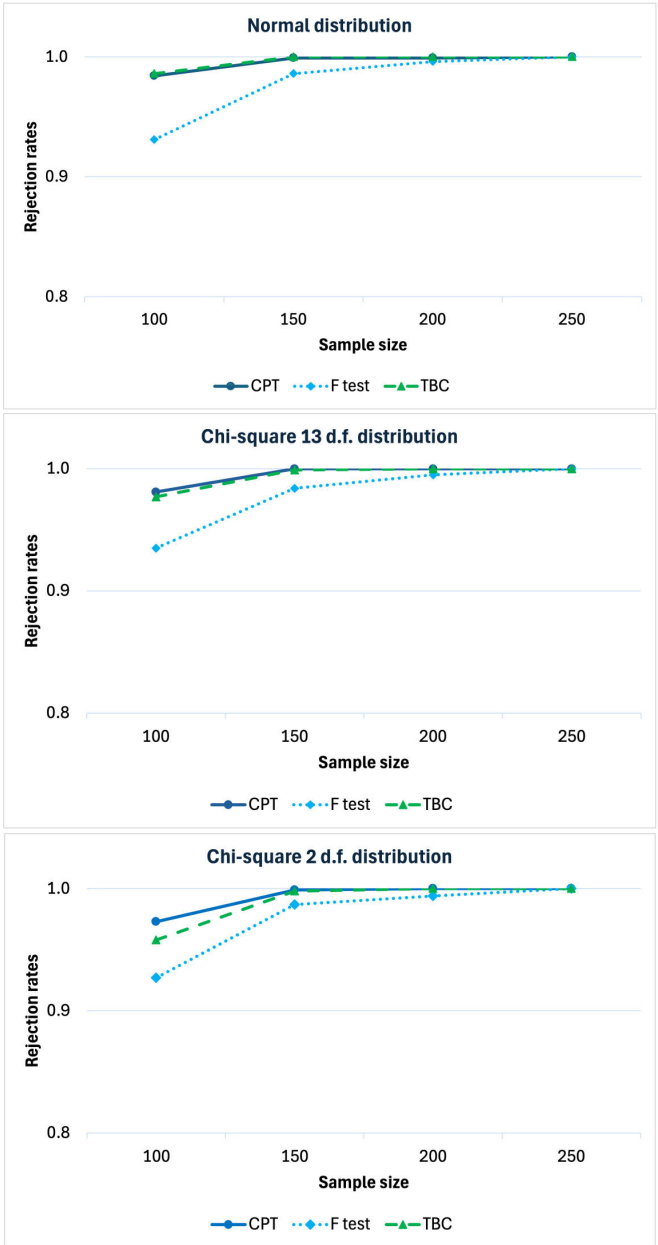

Figure 2. Rejection rates under H_1, with $\rho_z = 0.3$, $\sigma_z = 1$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\beta_0 = 6$, $\beta_1 = 7$, $\beta_2 = 23$, $\beta_{12} = 0$ and $\alpha = 0.05$.

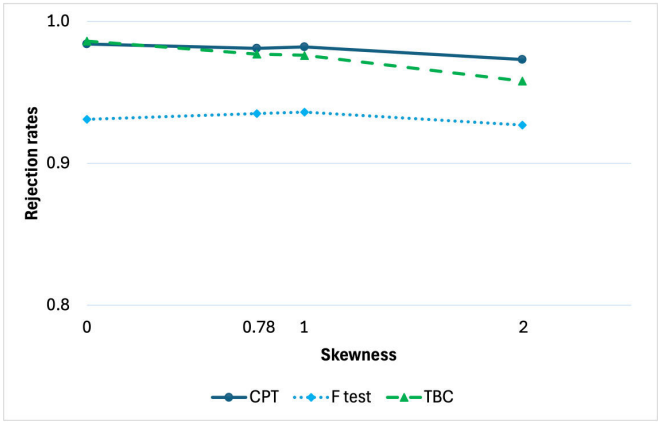

Figure 3. Rejection rates under H_1 as a function of the asymmetry of the error term, with $n = 100$, $\rho_z = 0.3$, $\sigma_z = 1$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\beta_0 = 6$, $\beta_1 = 7$, $\beta_2 = 23$, $\beta_{12} = 0$ and $\alpha = 0.05$.

5. Application to Real Data

As said, the application example concerns the study of the effect of age and smoking habit on the medical insurance cost in the USA. Data relate to a small random sample of the American population selected for a survey on the topic. We observed from [46] that attributes such as being a smoker, BMI, and age are the most important factors in determining medical insurance costs. However, since BMI in our dataset exhibits limited variability, resulting in a generally homogeneous patient population in this regard, we decided not to include it among the factors and instead considered only smoking status and age. The data source is Kaggle (https://www.kaggle.com/datasets/mirichoi0218/insurance, accessed on 20 December 2024).

The regression model under consideration for such a case study is the following:

$$Y_i = \beta_0 + \beta_1 \cdot smoker_i + \beta_2 \cdot age_i + \beta_{12} \cdot interaction_i + \varepsilon_i \tag{7}$$

where y_i is the charge of medical insurance costs for the i-th person, smoker$_i$ is the dummy variable that takes the value 1 if the i-th person is a smoker and 0 otherwise, age$_i$ is a dummy variable that takes the value 1 if the i-th person is at least 50 years old and 0 otherwise, and interaction$_i$ is the product between smoker$_i$ and age$_i$, hence it takes the value 1 if both the dummy variables take the value 1, in other words, if the level of both factors changes with respect to the baseline (from no-smoker under 50 y.o. to smoker over 50 y.o.).

In the literature, the concept that the allocation of health care expenses is significantly influenced by age is widespread, a trend that becomes increasingly important as the baby boom generation grows older. After the first year, health-care costs are at their lowest for children, gradually rise throughout adulthood, and surge dramatically after the age of 50 [47]. It was proved that the annual costs for older adults are roughly four to five times higher than those for individuals in their early teenage years [48]. For this reason, we took 50 years as a threshold to define *age* as a dummy variable [49]. The considered sample was obtained with a random selection from the American adult population, composed of 1338 people. This random selection implies that the exchangeability of statistical units under the null hypothesis holds.

Figure 4 shows the scatter diagram, the normal Q–Q plot, the histogram and the density plot of the residuals. According to such plots, the assumptions of uncorrelated, homoskedastic, zero-mean errors seem to be plausible, but the distributions of the errors could be asymmetric. In particular, by examining the histogram and the density plot of the residuals, it can be observed that the distribution appears asymmetric, suggesting that the data are skewed to the left. This violates one of the typical assumptions of classic

linear regression according to which errors follow a normal distribution. On the other hand, the same conclusion can be drawn from the normal Q–Q plot of residuals. In fact, the dots deviate from the straight line, indicating that the data may not follow the assumed theoretical distribution and therefore violate the normality assumption of the residuals. As a result, the assumption of normality for the model errors appears to be unmet and the use of the CPT test is preferable [50]. Indeed, as previously mentioned, the CPT method allows for the relaxation of the normality assumption, and there is no requirement to assume a specific family of probability distributions for the model error terms.

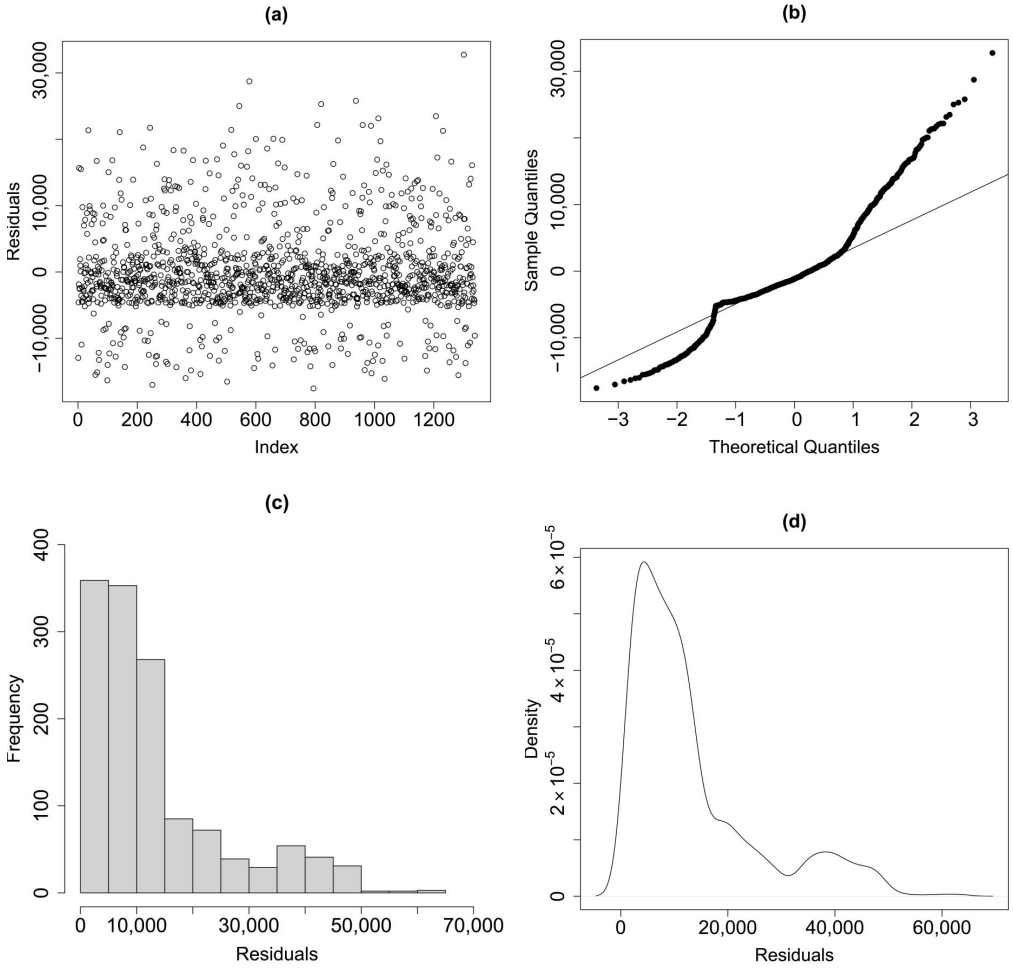

Figure 4. (**a**) Scatter plot of the residuals. (**b**) Normal Q−Q plot of residuals. (**c**) Histogram of the residuals. (**d**) Density plot of the residuals.

The boxplots of the response versus age and of the response versus being a smoker are reported in Figure 5. Such plots confirm the previous statements concerning the distribution asymmetry.

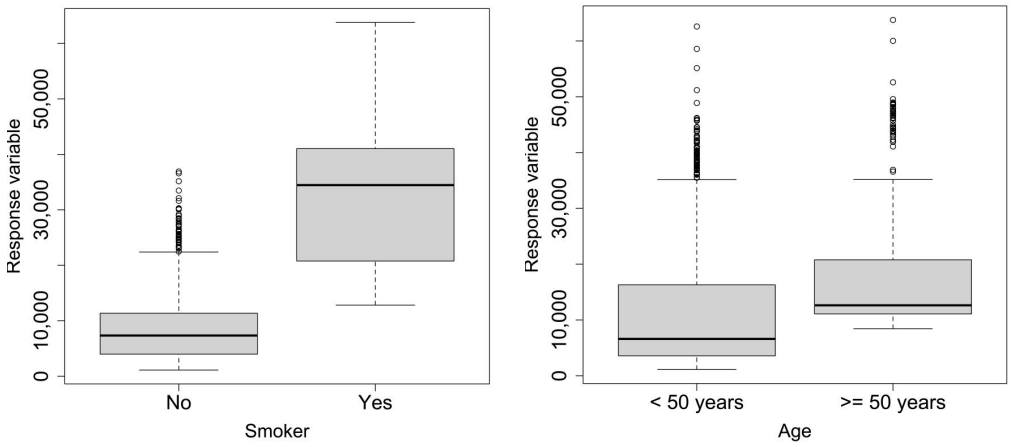

Figure 5. Boxplot of charge of medical insurance costs versus age and versus being a smoker.

In Figure 6, the main effect plots and the interaction plot are represented. From a descriptive point of view, a variation in the means in both the main effect plots is shown, which is much more evident in the case of being a smoker. Conversely, in the interaction plot, the two segments tends to be parallel so there may not be an interaction between the two factors.

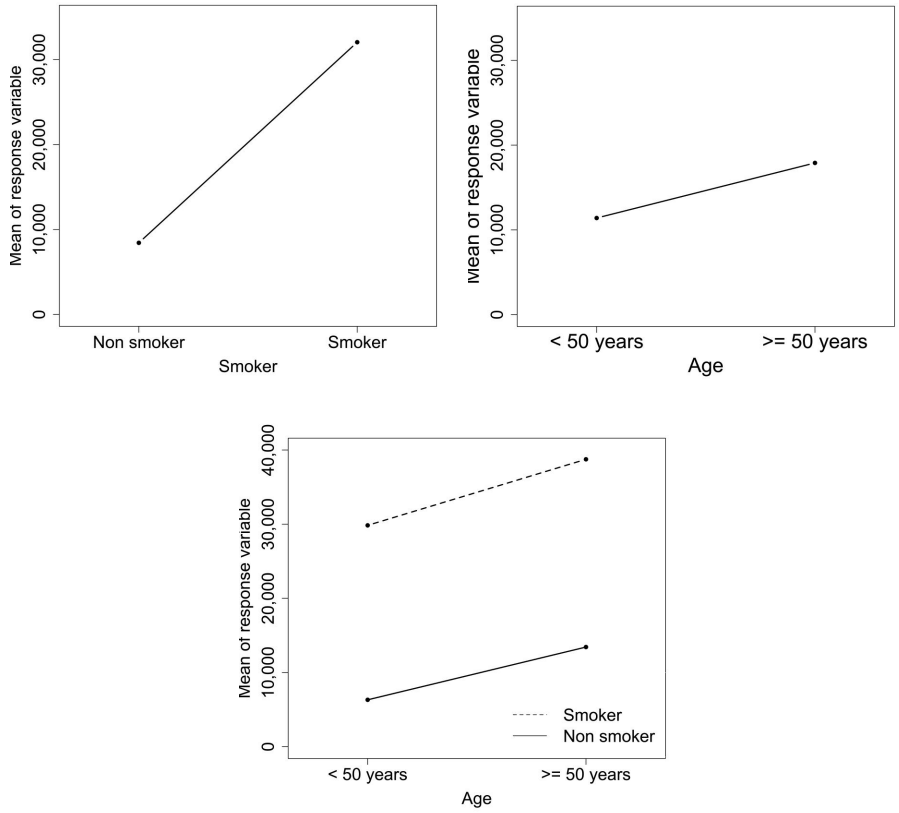

Figure 6. Main effect plots and interaction plot.

The goal of the analysis is to test the system of hypotheses defined in Section 3. To this aim, the non-parametric test presented was applied to such data. The significance level was set to 0.05. The overall p-value of the CPT is equal to 0.0002. This is less than the significance level $\alpha = 0.05$ which leads to the rejection of the null hypothesis of no effect. The partial p-values related to the tests on the significance of the coefficient estimates

adjusted with the Bonferroni–Holm correction are shown in Table 1. According to such adjusted p-values, the significance of the overall test can be attributed to the regression coefficients related to being a smoker and age but not to the interaction. Since the estimates of the regression coefficients are positive in both cases, we can state that being a smoker and being at least 50 years old positively influence medical insurance costs.

Table 1. Estimation of the coefficients, partial p-values on the significance of the coefficients estimates adjusted with the Bonferroni–Holm method (significance in bold).

	Estimate	**Adjusted p-Values**
Intercept	6313.9	
Smoker	23,525.3	**0.0003**
Age	7117.0	**0.0003**
Interaction	1792.1	0.3464

6. Concluding Remarks

This work is inspired by the study of the effects of age and smoking status on the insurance cost for the population in the USA. The main scientific contribution of the work consists of the application of an innovative non-parametric method, based on the CPT approach, to jointly test the main effects of the two binary factors and their interaction effect. The proposed method involves using permutation tests to assess the significance of individual coefficients, and combining the p-values of such partial tests to address the two-way ANOVA problem. The good power performance of the CPT compared to the F-test was shown in the Monte Carlo simulation study. The test was proved to be powerful, unbiased and consistent, regardless of the skewness of the error distribution.

The application of the proposed method to the interesting case study of medical insurance costs led to empirical evidence in favor of the hypothesis that age and being a smoker, each has a positive effect, whilst there is not an interaction effect. The findings of this research will practically support policymakers, insurance providers, and potential medical insurance buyers in making conscious decisions about choosing policies that meet their specific needs. A future extension of this work could explore the method's performance on datasets with more complex dependence structures or alternative parametric models.

Possible future developments related to this work may involve the implementation of parallel computing techniques. These techniques serve as valuable tools for accelerating a wide range of algorithms, including Monte Carlo simulations. Indeed, the use of parallel computing can significantly enhance the efficiency of Monte Carlo simulations. Furthermore, the proposed method could be expanded in future work to cases of ANOVA with higher-order interactions or multi-level factors. For instance, in a three-way ANOVA (i.e., with three factors), it would be necessary to examine three main effects, three two-way interactions, and one three-way interaction.

Author Contributions: The authors equally contributed to the paper. Conceptualization, S.B., M.B., G.P. and M.G.; methodology, S.B., M.B., G.P. and M.G.; software, S.B., M.B., G.P. and M.G.; validation, S.B., M.B., G.P. and M.G.; formal analysis, S.B., M.B., G.P. and M.G.; investigation, S.B., M.B., G.P. and M.G.; resources, S.B., M.B., G.P. and M.G.; data curation, S.B., M.B., G.P. and M.G.; writing—original draft preparation, S.B., M.B., G.P. and M.G.; writing—review and editing, S.B., M.B., G.P. and M.G.; visualization, S.B., M.B., G.P. and M.G.; supervision, S.B., M.B., G.P. and M.G.; project administration, S.B., M.B., G.P. and M.G.; funding acquisition, S.B., M.B., G.P. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available upon request in the case of publication of the paper.

Acknowledgments: The authors wish to thank the anonymous reviewers and the assistant editor for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

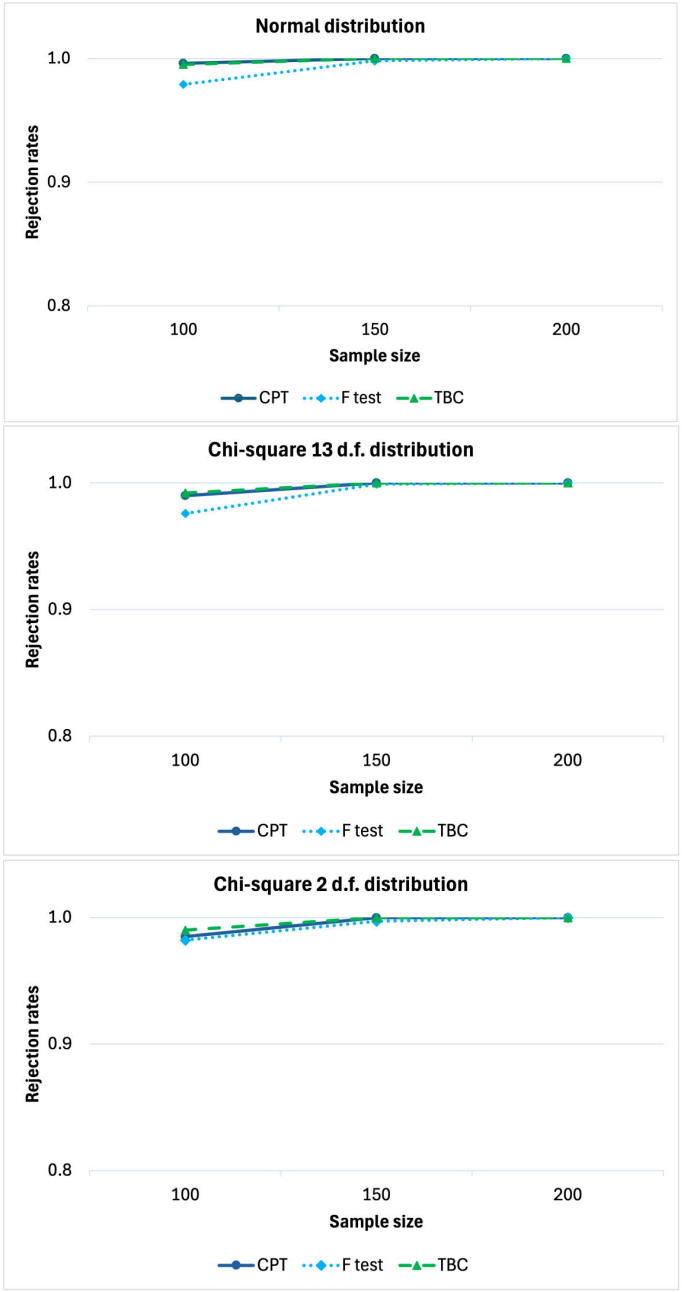

Figure A1. Rejection rates under H_1, with $\rho_z = 0.3$, $\sigma_z = 1$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\beta_0 = 6$, $\beta_1 = 7$, $\beta_2 = 23$, $\beta_{12} = 10$ and $\alpha = 0.05$.

References

1. Duncan, I.; Loginov, M.; Ludkovski, M. Testing alternative regression frameworks for predictive modeling of health care costs. *N. Am. Actuar. J.* **2016**, *20*, 65–87.

2. Orji, U.E.; Ugwuishiwu, C.H.; Nguemaleu, J.C.; Ugwuanyi, P.N. Machine learning models for predicting bank loan eligibility. In Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, Abuja, Nigeria, 15–17 May 2022; pp. 1–5.
3. Riedel, L.M. Health insurance in the United States. *AANA J.* **2009**, *77*, 439–444.
4. Koibichuk, V.V.; Khan, B.; Drozd, S.A. *The USA Medical Insurance as a Stimulating Factor to Increase Labour Efficiency*; Sumy State University: Sumy, Ukraine, 2023.
5. Patra, G.K.; Kuraku, C.; Konkimalla, S.; Boddapati, V.N.; Sarisa, M.; Reddy, M.S. An Analysis and Prediction of Health Insurance Costs Using Machine Learning-Based Regressor Techniques. *J. Data Anal. Inf. Proc.* **2024**, *12*, 581–596.
6. Alzoubi, H. M.; Sahawneh, N.; AlHamad, A. Q.; Malik, U.; Majid, A.; and Atta, A. Analysis Of Cost Prediction In Medical Insurance Using Modern Regression Models. In Proceedings of the 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 6–7 October 2022; pp. 1–10.
7. Shyamala Devi, M.; Swathi, P.; Purushotham Reddy, M.; Deepak Varma, V.; Praveen Kumar Reddy, A.; Vivekanandan, S.; Moorthy, P. Linear and ensembling regression based health cost insurance prediction using machine learning. In *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics*; Springer: Singapore, 2021; Volume 2, pp. 495–503.
8. Kaushik, K.; Bhardwaj, A.; Dwivedi, A.D.; Singh, R. Machine learning-based regression framework to predict health insurance premiums. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7898.
9. Rohilla, V.; Chakraborty, S.; Kumar, R. Deep learning based feature extraction and a bidirectional hybrid optimized model for location based advertising. *Multimed. Tools Appl.* **2022**, *81*, 16067-16095.
10. Gu, D.; Sung, H. Y.; Calfee, C. S.; Wang, Y.; Yao, T.; Max, W. Smoking-Attributable Health Care Expenditures for US Adults With Chronic Lower Respiratory Disease. *JAMA Netw. Open* **2024**, *7*, e2413869.
11. Orji, U.; Ukwandu, E. Machine learning for an explainable cost prediction of medical insurance. *Mach. Learn. Appl.* **2024**, *15*, 100516.
12. Ananda, M.M.; Weerahandi, S. Two-way ANOVA with unequal cell frequencies and unequal variances. *Stat. Sin.*,**1997**, *7*, 631–646.
13. Potvin, P.J.; Schutz, R.W. Statistical power for the two-factor repeated measures ANOVA. *Behav. Res. Methods Instrum. Comput.* **2000**, *32*, 347–356.
14. Toothaker, L.E.; Newman, D. Nonparametric competitors to the two-way ANOVA. *J. Educ. Stat.* **1994**, 19, 237–273.
15. Pesarin, F. *Multivariate Permutation Tests with Applications in Biostatistics*; Wiley: Chichester, YK, 2001.
16. Harrar, S.W.; Bathke, A.C. A non-parametric version of the Bartlett-Nanda-Pillai multivariate test. Asymptotics, approximations, and applications. *Am. J. Math. Manag. Sci.* **2008**, *28*, 309–335.
17. Yu, Y.; Zeng, L.; Wu, M.; Li, C.; Qiu, Y.; Liu, J.; Yang, F.; Xia, P. Exploring amyotrophic lateral sclerosis patients' experiences of psychological distress during the disease course in China: A qualitative study. *BMJ Open* **2024**, *14*, e082398.
18. Alibrandi, A.; Giacalone, M.; Zirilli, A. Psychological stress in nurses assisting Amyotrophic Lateral Sclerosis patients: A statistical analysis based on Non-Parametric Combination test. *Mediterr. J. Clin. Psychol.* **2022**, *10*(2).
19. Toma, P.; Miglietta, P.P.; Zurlini, G.; Valente, D.; Petrosillo, I. A non-parametric bootstrap-data envelopment analysis approach for environmental policy planning and management of agricultural efficiency in EU countries. *Ecol. Indic.* **2017**, *83*, 132–143.
20. Bonnini, S.; Borghesi, M.; Giacalone, M. Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with application to Circular Economy. *Ann. Oper. Res.* **2023**, 342, 1655–1670.
21. Simon, N.; Tibshirani, R. A permutation approach to testing interactions in many dimensions. *arXiv*, **2012**, arXiv:1206.6519.
22. Bonnini, S.; Borghesi, M. Relationship between Mental Health and Socio-Economic, Demographic and Environmental Factors in the COVID- 19 Lockdown Period-A Multivariate Regression Analysis, *Mathematics* **2022** *10*, 3237.
23. Das, P. Linear regression model: Goodness of fit and testing of hypothesis. In *Econometrics in Theory and Practice: Analysis of Cross Section, Time Series and Panel Data with Stata*; Springer Singapore: Singapore, 2019; Volume 15, pp. 75–108.
24. Bonnini, S.; Borghesi, M.; Giacalone, M. Semi-parametric approach for modeling overdispersed count data with application to industry 4.0. *Socio-Econ. Plan. Sci.* **2024**, *95*, 101976.
25. Bonnini, S.; Borghesi, M.; Giacalone, M. Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments. *Appl. Stoch. Models Bus. Ind.* **2024b**, *40*, 389–407.
26. Stute, W.; Thies, S.; Zhu, L. Model checks for regression: An innovation process approach. *Ann. Statist.* **1998**, *26*, 1916–34.
27. Basso, D.; Finos, L. Exact multivariate permutation tests for fixed effects in mixed models. *Commun. Stat. Theory* **2012**, *41*, 2991–3001.
28. Solari, A.; Finos, L.; Goeman, J.J. Rotation-based multiple testing in the multivariate linear model. *Biometrics* **2014**, *70*, 954–61.
29. López-Cheda, A.; Jácome, M.A.; Van Keilegom, I.; Cao, R. Nonparametric covariate hypothesis tests for the cure rate in mixture cure models. *Stat. Med.* **2020**, *39*, 2291–2307.
30. Cabassi, A.; Pigoli, D.; Secchi, P.; Carter, P. A. Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology. *Electron. J. Statist.* **2017**, *11*, 3815–3840.

31. Hollander, M.; Wolfe, D.A. *Nonparametric Statistical Methods*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 1999.
32. Pesarin F.; Salmaso L. *Permutation tests for complex data: Applications and software*, Wiley series in probability and statistics, 2010.
33. Bonnini, S.; Assegie, G.M.; Trzcinska, K. Review about the Permutation Approach in Hypothesis Testing. *Mathematics* **2024** *12*, 2617.
34. Gagnon-Bartsch, J.; Shem-Tov, Y. The classification permutation test: A flexible approach to testing for covariate imbalance in observational studies. *Ann. Appl. Stat.* **2019**, *13*, 1464–1483. https://doi.org/10.1214/19-AOAS1241.
35. Bonnini, S. Testing for heterogeneity with categorical data: Permutation solution vs. bootstrap method. *Commun. Stat. Theory Methods* **2014**, *43*, 906–917.
36. Li, Q.; Hu, J.; Ding, J.; Zheng, G. Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. *Biostatistics* **2014**, *15*, 284–295.
37. Yu, X.; Li, D.; Xue, L. Fisher's combined probability test for high-dimensional covariance matrices. *J. Am. Stat. Assoc.* **2024**, *119*, 511–524.
38. Westfall, P.H.; Young, S.S. On adjusting p-values for Multiplicity. *Biometrics* **1992**, *49*, 941–945.
39. Westfall, P.H.; Young, S.S. P-value adjustments for multiple tests in multivariate binomial models. *J. Am. Stat. Assoc.* **1989**, *84*, 780–786.
40. Tukey, J. W., The problem of multiple comparisons, **1953**, Princeton University. .
41. Watson, S.I.; Akinyemi, J.O.; Hemming, K. Permutation-based multiple testing corrections for P P-values and confidence intervals for cluster randomized trials. *Stat. Med.* **2023**, *42*, 3786–3803.
42. López-Cheda, A.; Cao, R.; Jácome, M.A.; Van Keilegom, I. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comput. Stat. Data Anal.* **2017**, *105*, 144–165.
43. Albert, M.; Bouret, Y.; Fromont, M.; Reynaud-Bouret, P. Bootstrap and permutation tests of independence for point processes. *Ann. Statist.* **2015**, *43*, 2537–2564.
44. Fang, J.; Huang, C.; Tang, T.; Wang, Z. Parallel programming models for heterogeneous many-cores: A comprehensive survey. *Ccf Trans. High Perform. Comput.* **2020**, *2*, 382–400.
45. Wolfe, M. Parallelizing compilers. *Acm Comput. Surv. (Csur)* **1996**, *28*, 261–262.
46. ul, Hassan, C.A.; Iqbal, J.; Hussain, S.; AlSalman, H.; Mosleh, M.A.; Sajid, Ullah, S. A computational intelligence approach for predicting medical insurance cost. *Math. Probl. Eng.* **2021**, *2021*, 1162553.
47. Meerding, W.J.; Bonneux, L.; Polder, J.J.; Koopmanschap, M.A.; van der Maas, P.J. Demographic and Epidemiological Determinants of Healthcare Costs in Netherlands: Cost of Illness Study. *Br. Med. J.* **1998**, *317*, 111–115.
48. Bradford, D.F.; ; Max, D.A. *Implicit Budget Deficits: The Case of a Mandated Shift to Community-Rated Health Insurance*; NBER working paper no. 5514; National Bureau of Economic Research: Cambridge, MA, USA, 1996.
49. Alemayehu, B.; Warner, K.E. The lifetime distribution of health care costs. *Health Serv. Res.* **2004**, *39*, 627–642.
50. Westfall P.H. Kurtosis as peakedness, 1905–2014. "r.i.p.". *Am. Stat.* **2014**, *68*, 191–195.