Anadolu Tarım Bilim. Derg., 2009,24(2):93-97
Anadolu J. Agric. Sci., 2009,24(2):93-97

Araştırma
Research

USE AND COMPARISON OF PERMUTATION TESTS IN LINEAR MODELS

Hasan ÖNDER [1*] Zeynel CEBECİ[2]

[1] Ondokuz Mayıs University, Faculty of Agriculture, Dep. of Animal Science, Biometry & Genetics Unit, 55139 Samsun-Turkey
[2] Çukurova University, Faculty of Agriculture, Dep. of Animal Science, Biometry & Genetics Unit, 01330 Adana-Turkey

[*]e-mail: hasanonder@gmail.com

Received Date:19.01.2009 Accepted Date: 20.04.2009

ABSTRACT: *F* and *t*-test are generally used to test significance of hypothesis and/or model parameters. Although parametric tests are considerably effective, they can be ineffective when the assumptions needed by model are not provided, which is a usual situation for many data sets. In this case, permutation test not affected by the assumptions can be applied as a non-parametric method. In this study, permutation tests such as permutation of raw data, permutation of residuals under full model and permutation of residuals under restricted model are compared for multiple linear regression, completely randomized designs, randomized block design and Latin square design in terms of the Type I error rates, and performance of each tests are studied via animal science data. Results from this study indicate that permutation tests yields more reliable results than parametric tests in terms of Type I error rate, and permutation tests are recommended in order to reduce Type I errors.
Keywords: Linear models, Resampling methods, Permutation tests

PERMÜTASYON TESTLERİNİN DOĞRUSAL MODELLERDE UYGULANMASI VE KARŞILAŞTIRILMASI

ÖZET: Genellikle hipotezin ve/veya model parametrelerinin testi için *F* ve *t*-testleri kullanılır. Parametrik testler parametrik olmayan karşıtlarına göre daha etkili olsa da, pek çok veri seti için gerekli olan model varsayımlarının sağlanamadığı durumlarda, etkilerini yitirmektedirler. Bu durumda, varsayımlardan etkilenmeyen Permütasyon testleri parametrik olmayan bir yöntem olarak uygulanabilmektedir. Bu çalışmada, ham verinin permütasyonu, kalıntıların tam permütasyonu, kalıntıların kısmi permütasyonu yöntemleri, çoklu doğrusal regresyon, tesadüf parselleri, tesadüf blokları ve Latin kare deneme desenleri için I. tip hata olasılıkları bakımından karşılaştırılmıştır. Yöntemlerin karşılaştırılmasında hayvancılık verileri kullanılmıştır. Sonuç olarak, I Tip hata olasılığı bakımından Permütasyon testlerinin parametrik yöntemlere göre daha güvenilir sonuçlar ürettiği ve daha yüksek I. Tip hatadan kaçınmak için önerilebilecekleri görülmüştür.
Anahtar sözcükler: Doğrusal modeller, Yeniden örnekleme yöntemleri, Permütasyon testleri

1. INTRODUCTION

The first description of permutation tests one of the resampling methods for linear statistical models can be traced back to the works of Fisher (1935) and Pitman (1937) in the first half of the 20th Century. Since permutation tests are computationally intensive; their employment in data analyses did not receive much attention until the widespread use of powerful computers (Anderson & Robinson, 2001).

Because of its independency from the distribution, permutation tests are successful in many cases where parametric tests are not. The assumptions of permutation tests are exchangeability and relabelability of data. If the null hypothesis is established correctly, exchangeability and relabelability are obtained. If the null hypothesis is correctly established, there will be no effect on the result even when the observations between two groups are exchanged.

The aim of this study is to compare permutation of raw data, permutation of residuals under full model and permutation of residuals under restricted model on linear models such as multiple regression, completely randomized design, randomized block design and Latin Square design.

2. MATERIAL AND METHOD

All data which were used in this study was originated from previous experiments on small ruminant (Darcan, 2004) and animal nutrition (Serbester et al, 2005) carried at Çukurova University. The data used for multiple regression analysis was taken from a study on sheep research with sample size of 8. Pulse number (PN) was selected as response variable and explanatory variables were selected as rectal heat (RH) and respiration number (RN).

For randomized block design, species (sheep, goat; *n*=16) factor on rectal heat was examined and four different sampling times in a day (6–7 am, 12–13 pm, 18–19 pm and 0–1 am) were used to block factor. For completely randomized design, effect of sampling times was removed from same data set without changing the data and only species variable used as a factor.

The data used for Latin square analysis was taken from a study on animal nutrition with sample size of 25. In this model, animal, group and ration effects were used as row, column and treatment, respectively. To analyze the data NPMANOVA and DISTLM statistical software was utilized (Anderson, 2000; Anderson, 2003). To examine whether data has

Summary of the PhD Thesis

normal distribution or not, Anderson-Darling normality test was performed by use of MINITAB statistical software.

Permutation of raw data (PRD) permutes the raw observations. The essential requirement for this to work is that the distribution of the observations must be similar to the distribution of errors under the null hypothesis. This method may not be true if there is an outlier in data set.

Permutation of residuals under the reduced model (PRR) can be generally referred to model-based permutation. In this model residuals of the linear model are used as the permutable units of the test.

Permutation of residuals under the full model (PRF) permutes residuals uses the residuals from the full regression model as the permutable units of the test. The rationale for the method is that it uses the estimates of $\beta_{1.2}$ as part of the test, but also uses the original estimate of $\beta_{2.1}$ as part of the permutational procedure.

Normality of the data set and/or existence of outliers were used as a criteria when the compare models.

2.1. Multiple Regressions

When X_1 is constant and n combinations do exists (missing values and duplications are unimportant), the possible combinations can be shown as (X_{1j}, Y_j), j = 1,2,...,n. Similarly if X_2 is constant, possible combinations can be shown as (X_{2j}, Y_j), j = 2,3,...,n. Hence there are n - 1 combinations in this case, in turn, there are n-2 and n-3 combinations for X_3 and X_4 respectively. Finally the number of all possible combinations between X and Y is $n!$. In a multiple linear regression F value for statistical significance is calculated as indicated in equation given as (Kleinbaum et al, 1998);

$$F = \frac{MS_{reg}}{MS_{error}}.$$

The number of F values is $n!$ which will be handled by changing the order of Y. Let F^*, $*$ = 1,2,...,$n!$ and F_j^* is j^{th} element of F^*. Then F^*= (F_1^*, F_2^*,..., F_k^*). When it is assumed that F_j^* is jth element of set of F values under the null hypothesis, the experimental distribution of F_j which is jth element of F value estimated by OLS (Ordinary Least Squares) can be given as follows:

$$P(F_j < F_j^1) = \frac{\text{Number of the } F^* \text{ less than } F_j^1}{n!}$$

$$P(F_j = F_j^1) = \frac{\text{Number of the } F^* \text{ equal to } F_j^1}{n!}$$

$$P(F_j > F_j^1) = \frac{\text{Number of the } F^* \text{ greather than } F_j^1}{n!}$$

For all j = 1,2,...,k, the computing method given above can be applied and calculated for each probability. The significance test of regression equation can be applied as determining the position of F_j^1. If either $P(F_j < F_j^1)$ or $P(F_j > F_j^1)$ is small enough null hypothesis, H_0:$\beta_j = 0$ is rejected by two-tailed test.

Only PRD and PRF methods were used on multiple regression analysis because of software limitation.

2.2. Variance Analysis

e_{ij} term in the mathematical model which is called as error term can be divided into two pieces such as technical error which is a random variable equal to the difference of the conceptual observed response y_{ij}, measurement error, and treatment error. By use of permutation test technical error tends to be zero. Thus, only treatment error can remain in the model. For this reason, permutation tests can yield more reliable results (Good, 2000).

To calculate a P value, the F value obtained from the original data is compared with the distribution of F^* values obtained by permutation test. The empirical frequency distribution of F^* is entirely exposed because the number of possible relabeling data is finite. Type I error rate for the null hypothesis is calculated as dividing the number of F^* equals to or greater than F by total number of F.

$$P = \frac{(\text{number of. } F^* \geq F)}{n!}$$

This P value provides an exact test for the null hypothesis, when there are no differences among groups.

2.2.1. Completely randomized design

The possible number of relabeled data sets for one-way ANOVA which has t groups with n replicates can be calculated by the equation of $(tn)!/[t!(n!)^t]$ (Anderson, 2001). Here, it is essential to state that original data is a member of permutation set. Only PRD method was used on analysis of completely randomized design because of software limitation.

2.2.2. Randomized Block Design

The possible number of relabeled data sets for randomized block design which has b blocks and t treatments can be calculated by the equation of $(t!)^b$ (Mielke and Berry, 2001; Anderson, 2001). PRD, PRR and PRF methods were used on analysis of randomized block design.

2.2.3. Latin Square Design

The possible number of relabeled data sets for Latin square design can be calculated by $t!\,(t-1)!$ (Scheffé, 1959).

To put into practice the permutation test for Latin square design exactly it would be simplest to base it on the sum of squares for numbers (i.e., levels of factor *C*), namely on $\sum_k T_k^2$ where T_k is total of the *t* observations where factor *C* is at level *t*. the number of different squares in a transformation set is $t!(t-1)!$ times the number of Standard squares in the set. Since the statistic is invariant under the $t!$ permutation of the levels of *C*, the number of different values the statistic takes on with equal probability for a given transformation set is $(t-1)!$ times the number of standard squares in the set. PRD, PRR and PRF methods were used on analysis of randomized block design.

3. RESULTS AND DISCUSSION

3.1. Multiple Regressions

For multiple regression analysis, conventional analysis OLS (Ordinary Least Squares), available permutation methods such as (PRD) permutation of raw data, (PRF) permutation of residuals under the full model were performed and Type I error rates were observed respectively, 0.009, 0.0047 and 0.0054. These results shows that permutation tests produce smaller Type I error rates than OLS, but there are no considerable differences in Type I error rates among any of methods. Result of Anderson-Darling test shows that data has not normal distribution, so one of the assumption of OLS was not confirmed. In this situation, OLS method is having a tendency to yield errant results, and it can be indicated that Type I error rates obtained from PRD and PRF are more reliable than OLS results.

Type I error rate obtained from permutation of raw data is smaller than permutation of residuals (full model), but this difference is not significant. Anderson and Legendre (1999) declared that there are no difference affect the making decision between PRD and PRF by means of Type I error rates, and they suggested PRD method. Anderson (2001) suggested PRD method for small sample sizes. It can be preferred because it is distribution free and needs less computer time. Tanizaki (2001) also proposed permutation methods for significance tests for regression models when the data has not normal distribution. Obtained findings for this study also support the study results of Anderson and Legendre (1999).

3.2. Completely Randomized Design

For completely randomized design, ANOVA and PRD which is suitable for completely randomized design are performed and Type I error rates were observed respectively, 0.005 and 0.0068. According to the results, permutation tests produce higher Type I error rates than ANOVA, but there are no differences affect the making decision in Type I error rates between these methods. Result of Anderson-Darling test shows that data has not normal distribution. When the methods examined for this design, there is no significant difference between these methods to change the decision about the null hypothesis for specie factor. Önder (2007), Routledge (1997) and Anderson (2001) notify that permutation tests produce more reliable results than ANOVA because permutation tests is distribution free and permutation tests equalize the technical error to zero. For these reasons, PRD method can be preferred for completely randomized design.

3.3. Randomized Block Design

For randomized block design, conventional analysis ANOVA, available permutation methods such as (PRD) permutation of raw data, (PRF) permutation of residuals under the full model and (PRR) permutation of residuals under the reduced model were performed and Type I error rates were shown in Table 1.

Table 1. Type I error rates for randomized block design by use of ANOVA, PRD, PRF, and PRR.

	ANOVA	PRD	PRR	PRF
Species	0.002	0.0033	0.0033	0.0002
Time	0.024	0.0261	0.0045	0.0028
Species x Time	0.423	0.4238	0.4212	0.4590

When the results of randomized block design examined with assistance of Table 1, it is clear that PRD and PRR permutation methods produced higher Type I error rates than ANOVA but PRF method produced smaller Type I error rate than ANOVA for species factor. PRD method produced higher Type I error rates than ANOVA but PRF and PRR methods produced smaller Type I error rate than ANOVA for time factor which is used as block. PRD and PRF methods produced higher Type I error rates than ANOVA but PRR method produced smaller Type I error rate than ANOVA for species x time interaction term.

Anderson (2001) notify that PRR method should be used when the size of sample is higher than 10 and researcher recommended the use of PRD method when the size of sample is smaller than 10 because of unreliability of residuals. Routledge (1997) and Anderson (2001) notify that permutation tests produce more reliable results than ANOVA because permutation tests is distribution free and permutation tests equalize the technical error to zero as they notified for completely randomized design. Anderson (2001) notify that PRF method may lead to incorrect results in interaction analysis for randomized block design. In this study size of sample was higher than 10 and this situation PRR method can be preferred because data did not satisfy the assumptions of ANOVA.

3.4. Latin Square Design

For Latin Square design, conventional analysis ANOVA, available permutation methods such as (PRD) permutation of raw data, (PRF) permutation of residuals under the full model and (PRR) permutation of residuals under the reduced model were performed and Type I error rates were shown in Table 2.

Table 2. Type I error rates for Latin Square design by use of ANOVA, PRD, PRR, and PRF.

	ANOVA	PRD	PRR	PRF
Row	0.203	0.4348	0.4308	0.4308
Column	0.052	0.1587	0.1602	0.1602
Treatment	0.030	0.2640	0.0313	0.0313

When the PRD, PRR and PRF methods which is performable on Latin Square design compared with ANOVA, as seen in Table 2, all of three permutation methods produced higher Type I error rates than ANOVA for row, column and treatment factors. It is essential to state that discrepancy between PRD and ANOVA methods change the decision about the null hypothesis. While the ANOVA method rejects the null hypothesis, PRD method accepts the null hypothesis. The other permutation methods such as PRR and PRF reject the null hypothesis at significance level of 5%.

Routledge (1997) informed that permutation tests produce more reliable results from the point of Type I error rate for Latin Square design by reason of the fact that permutation methods can eliminate technical error. Selection of permutation method depends on size of sample for variance analysis as Anderson (2001) notified.

Taking into account of information that permutation tests should be used only if number of treatment is equal or higher than 5 on Latin Square design, it is understood that number of observation for Latin Square design contain at least 25 observation units. Interpretation of this information suggests the use of permutation of residuals under the reduced model (PRR) for Latin Square design because permutation tests never applied on Latin Square design which has smaller than 25 observations.

4. CONCLUSION

In this study, permutation methods such as (PRD) permutation of raw data, (PRF) permutation of residuals under the full model and (PRR) permutation of residuals under the reduced model were compared with conventional methods. It is determined that permutation tests are more reliable on multiple linear regression, completely randomized design, randomized block design, and Latin Square design.

In multiple linear regression models, if the data has not normal distribution and/or there is high correlation between explanatory variables, PRF can be preferred in the presence of outlier/s.

For completely randomized design, PRD method can be preferred. When the size of sample is higher than 10, PRR method can be recommended. When the size of sample is smaller than 10, PRD method can be proposed for randomized block design.

For Latin Square design, comparison of permutation tests has not mentioned in previous studies according to obtained references. Result of this study showed that PRR method should be used for Latin Square design.

It is the most important factor to recommend the use of permutation tests that it equalize the technical error, one of the components of error term, to zero and only treatment error remained in the error term. It is well known that data taken from biological studies generally do not satisfy the assumption of the parametric methods. If data does not receive assumptions or structure of the data is not known, permutation tests can be performed to obtain more reliable results. Otherwise, the statistical decision may lead to misinterpretation of the results because of making Type I error for the hypothesis. Permutation tests and parametric methods yield similar results when the data fulfill the necessary assumptions for the parametric method. In this case use of parametric methods can be preferred in terms of computer time and simple calculation effort.

For the future studies, it is understood that comparison of resampling methods such as permutation, Bootstrap and Jackknife with one another and/or parametric methods such as F and t tests is necessary.

5. ACKNOWLEDGEMENTS

This study is summary of PhD thesis and supported by Academic Research Project Unit of Çukurova University with project number of FBE.2003D-2.

6. REFERENCES

Anderson, M.J., Legendre, P. 1999. An Empirical Comparison of Permutation Methods for Tests of Partial Regression Coefficients in a Linear Model. J. Statist. Comput. Simul. 62, 271-303.

Anderson, M.J. 2000. NPMANOVA: A FORTRAN Computer Program for Non-Parametric Multivariate Analysis of Variance (For Any Two Factor ANOVA Design) Using Permutation Tests. Department of Statistics, University of Auckland.

Anderson, M.J. 2001. Permutation Tests for Univariate and Multivariate Analysis of Variance and Regression. Can. J. Fish. Aquat. Sci. 58, 626-639.

Anderson, M.J., Robinson, J. 2001. Permutation Tests for Linear Models. Aust. N. Z. Stat. 43(1), 75 – 88.

Anderson, M.J. 2003. DISTLM: A FORTRAN Computer Program to Calculate a Distance-Based Multivariate Analysis for a Linear Model. Department of Statistics, University of Auckland.

Darcan, N. 2004. Data obtained from small ruminant research. (Personal communication).

Fisher, R.A. 1935. The Design of Experiments. Oliver and Body, Edinburgh.

Good, P. 2000. Permutation Tests, A Practical Guide to Resampling Methods for Testing Hypotheses. Springer-Werlag New York Inc.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A. 1998. Applied Regression Analysis and Other Multivariable Methods. Brooks/Cole Publishing Company.

Mielke, P.W., Berry, K.J. 2001. Permutation Methods: A Distance Function Approach. Springer-Werlag New York Inc.

Önder, H. 2007**.** Using Permutation Tests to Reduce Type I and II Errors for Small Ruminant Research. J. Appl. Anim. Res. 32 (1), 69 – 72.

Pitman, E.J.G. 1937. Significance Tests Which May be Applied to Samples from Any Population. Royal Staistical Society Supplement. Part I. 4, 119-130.

Routledge, R.D. 1997. P-Values from Permutation and F-Tests. Computational Statistics & Data Analysis. 24, 379 – 386.

Scheffé, H. 1959. The Analysis of Variance, John Willey & Sons, Inc. New York.

Serbester, U., Görgülü, M., Kutlu, H.R., Yurtseven, S, Arieli, A., Kowalski, Z.M. 2005. The Effects of Sprinkler+Fan, Fish Meal or Dietary Fat on Milk Yield and Milk Composition of Dairy Cows in Mid Lactation During Summer. Journal of Animal and Feed Sciences 14, 639 – 653.

Tanizaki, H. 2001. On Small Sample Properties Permutation Tests: An Independence Test Between Two Samples and A Significance Test For Regression Models. (http://ht.econ.kobe-u.ac.jp/~tanizaki/cv/working/permute.pdf Last visit: 12/03/2007).