Journal of Applied Sciences

ISSN 1812-5654
Significant Tests of Coefficient Multiple Regressions by using Permutation Methods

Ali Shadrokh
Department of Statistics, University of Payame Noor, 19395-4697, Islamic Republic of Iran

Abstract: Tests of significance of a single partial regression coefficient in a multiple regression model are often made in situations where the standard assumptions underlying the probability calculation (for example assumption of normally of random error term) do not hold. When the random error term fails to fulfill some of these assumptions, one need resort to some other nonparametric methods to carry out statistical inferences. Permutation methods are a branch of nonparametric methods. This study compared empirical type one error of different permutation strategies that proposed for testing nullity of a partial regression coefficient in a multiple regression model, using simulation and show that the type one error of Freedman and Lane’s strategy is lower to than the other methods.

Key words: Permutation test, exchangeable observation, regression coefficient, empirical type one error, permutative matrix

INTRODUCTION

The first descriptions of permutation tests for linear statistical models, including analysis of variance and regression, can be traced back to the work of Fisher (1935).

Permutation tests were first introduced by Fisher (1935) and Eizharinia et al. (2010). But they were not pleased so much because they needed time consuming calculations. But nowadays these calculations by production of fast and powerful computers that calculating of a p-value is faster than finding an amount of charts for a parametric test. Parametrical tests take place in the area of free- distribution. As we know for doing the question of testing of hypothesis, in addition to hypothesizes like independency of error term, stability of their variance and random sampling, we need to the hypothesis of normal distribution of error term, while for doing permutation tests we don’t need to basic hypothesizes. In fact we for doing a non parametrical test, we use series of simple hypothesizes that leaves the researcher. One of these hypothesizes which in fact is the base for permutation tests is hypothesis of exchange ability of observations which defined like this.

Definition 1: Let’s realize exchange ability of random vector (n) dimensional, \( x = (x_1, x_2, ..., x_n) \) with joint distribution of \( f_{x_1, x_2, ..., x_n} \). \( X \) is called exchangeable if joint density of observations per each permutation vector is suitable. So for each permutation vector which is shown by \( X \) we have:

\[
f_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n) = f_{X'_1, X'_2, ..., X'_n}(x'_1, x'_2, ..., x'_n)
\]

Various permutational strategies have been proposed for testing nullity of a partial regression coefficient in a multiple regression model (AL-Salih et al., 2010; Bughio et al., 2002; Edriss et al., 2008; Kandhiro et al., 2002; Laghari et al., 2003; Rashid et al., 2002; Serhat Odabas et al., 2007; Tariq et al., 2003; Alam, 2004). The proposed permutation methods for such tests have different bases in term of their philosophies and have been proposed in different contexts.

In Anderson and Legendre (1999) point of view just four cases of these strategies are suitable. These four methods are: Manly method (Manly, 1991), Kennedy method (Kennedy and Cade, 1996) Freedman-Lane method (Freedman and Lane, 1983) and Ter braak (Ter-Braak, 1992) but in this article we just compare Kennedy’s method and Freedman and Lane’s method.

KENNEDY’S METHOD

Suppose variable answer Y and variable \( x_1, x_2, ..., x_p \) and assume we have \( n \) observation for regression. Then we have this equation:

\[
3328
\]
where, ε is (sentence wrong) error term and has an
derminate distribution F with by a mean zero variance
σ². We can write the Eq. 1 as the following vector-
matrix:

\[
y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad \epsilon \sim N(0, \sigma^2 I_n) \tag{1}
\]

(2)

\[
Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}, \quad \epsilon \sim N(0, \sigma^2 I_n)
\]

We want to test hypothesis \( H_0 : \beta_p = 0 \) in
canison with \( H_1 : \beta_p \neq 0 \). For doing this test by
permation method we follow Kennedy algorithm like this: we define matrix:

\[
H_i = X_i(X_i'X_i)^{-1}X_i
\]

and then multiple correspondents of Eq. 2 in \( (I, H_i) \) to
solve equation shown here:

\[
\hat{y} = \hat{X}_p \beta_p + \hat{\epsilon}
\]

(3)

Where:

\[
\hat{y} = (I_i - H_i)y, \quad \hat{X}_{p,i} = (I_i - H_i)x_p, \quad \hat{\epsilon}_{i} = (I_i - H_i)\epsilon
\]

**Step 1:** Then by using least error squares, we estimate
\( \beta_p \) like this:

\[
\hat{\beta}_p = \frac{\hat{X}_p'\hat{y}}{\hat{X}_p'\hat{X}_p} = \frac{x_p'(I_i - H_i)y}{x_p'(I_i - H_i)x_p}
\]

After that we estimate \( \beta_p \) amount of statistic test
t = \( \frac{\hat{\beta}_p \sigma}{\hat{\beta}_p} \) calculate and call it referential \( t \).

**Step 2:** We do permutation for \( \hat{y} \) amounts and show
it as \( \hat{y}^* \).

**Step 3:** We make a regression equation on \( \hat{X}_p \), model
of Eq. 3, from found vector. Then we estimate
\( \hat{\beta}_p \) by least sum of square error, like one given
here:

\[
\hat{\beta}_p = \frac{\hat{X}_p'\hat{y}}{\hat{X}_p'\hat{X}_p}
\]

**Step 4:** By repeating second and third step and finding
t's we also find permutation distribution t's, next, by using that we find p-value which is the
relation of amount of permutation statistics that
their absolute value is larger than their referential
E absolute value and at least we admit or reject
zero hypothesis.

**PROBLEM WITH KENNEDY METHOD**

In permutation methods, permuting amounts of \( y \),
causes amounts of \( \epsilon \) permute. Thus a question comes to
mind is that whether this obligatory permutation changes
its parametrical distributions or not? So we first give this
definition.

**Permutative matrix:** Definition 2 permutative matrix is a
square matrix that there are numbers of zero and only a
one each row and the position of number one in each row
is different from other rows. This matrix is shown by
sample P and has a special feature of \( P'P = P P' \).

Now, with respect to the above definition, we analyze
the mentioned problem: We know that:

\[
\epsilon = F(0, \sigma^2 I_n)
\]

so, considering the above definition, we can put the
permuted vector \( \epsilon \) which is known by the notation of \( \epsilon' \)
this way \( \epsilon' \) ptc. As a result:

\[
E(\epsilon') = E(\epsilon) = 0, \quad \text{Var}(\epsilon') = \text{Var}(\epsilon) = \sigma^2 I_n
\]

It is clear that by the permutation of values of \( y \) does
not change the distribution of \( \epsilon \). Now, we study this in
case of the reduced equation of which Kennedy has made
use Eq. 3.

It can easily be proved that:

\[
\epsilon = F(\sigma'(I_n - H_i))
\]

Regarding the definition 2, we can easily prove that:

\[
E(\epsilon') = E(\epsilon) = 0, \quad \text{Var}(\epsilon') = \text{Var}(\epsilon) = \sigma^2 (I_n - H_i)
\]

Which shows the variance of \( \epsilon' \) is dependent on \( H_i \),
thus by any permutation of \( H_i \), \( \sigma^2 \), is multiplied by a new
number that changes the variance. Therefore, by any permutation of \( \hat{y} \), the parameters of distribution of \( \tilde{e} \) changes. So due to the null hypothesis \( H_0: \beta_p = 0 \). The distribution of \( \hat{y} \) also changes by any permutation.

**MODIFIED KENNEDY’S METHOD**

Huh and Jhun (2001) improve this problem, with regard to this point that the matrix \((I, H)\) is the one to a power of its own value and with the rank of \( n' = n-p \), added the following steps to Kennedy’s algorithm:

**Step 1:** First, we work the eigenvectors and rank of the matrix \((I, H)\) out.

**Step 2:** We save the eigenvectors equivalent to the particular of one which equals the rank of in the form of \((I, H)\) the matrix.

**Step 3:** Through the orthogonalizing process of Gram Schumith, we change/tune this matrix to an orthogonal one.

**Step 4:** Divide each column of this matrix by the norm of that column so that an orthogonal unit vector is resulted. We come to realize this new matrix by \( V_i \), the dimensions of which are \((n\times n')\).

**Step 5:** Using the spectrographic analysis and knowing the point that particular measures of matrix \( V_i \) is one, we reduce the matrix \((I, H)\) this way:

\[
(I, H) = V_i V_i' = V_i V_i'
\]

**Step 6:** After reduction of this matrix, we multiply two sides of the Eq. 3 by the matrix \( V_i' \). So that a new equation is resulted in this form:

\[
\tilde{y}_i = \tilde{y}_{n,p} + \tilde{e}_{i,d}, \quad \tilde{e} \sim F(0, \sigma^2 I_n)
\]

Where:

\[
\tilde{y} = V_i \hat{y}, \quad \hat{y} = V_i \tilde{y}, \quad \tilde{e} = V_i' \tilde{e}
\]

Now, we turn to the distribution of the permuted vector of \( \tilde{e} \). To do so, we again refer to the definition 2:

\[
E(\tilde{e}) = E(\tilde{e}P) = 0, \quad \text{Var}(\tilde{e}) = \text{Var}(P) = \sigma^2 I_n
\]

So it is pinned down that \( \tilde{e} = \tilde{e} \) and this gives clue to the fact that this new method keeps the distribution of the error expressions constant and stable even after the permutation.

After the modification improvement of the Eq. 3 and turning it into the form (4), we apply the steps contributed by Kennedy to the new resulted equation.

**SIMULATION**

Anderson and Legendre (1999), having an extent simulation done (ignoring the error in Kennedy’s method), showed that in Friedman-Law method, a Type I error probability is less than that the Kennedy’s methods. Schadorfn and dAubigny (2010), Shadrick (2011) analytically show that the Type I error of Friedman and Lane method is lower than that of Kennedy’s approach.

Here, with the aid of a simulation, we intend to check out whether the claim Anderson and Legendre (1999) and Shadrick (2011) have made remains valid after modifying the error in the method of Kennedy or not? To do so, we calculate the empirical probability of type I error in all the three permuting methods on the basis of a double regression model

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon
\]

considering the following factors:

- The sample size \( \{10, 20, 30\} \)
- The correlation between the two variances \( x_1, x_2 \):
  \( \rho = \{0.1, 0.9\} \)
- The quantity/value of the coefficient of regression which is not tested \( \beta = \{0.5, 1.5\} \)
- The distribution of error expressions: the exponential with the parameter 2

The focused simulation is done employing the software S-plus and the results are drawn as the following charts.

In all four case (showed in four chart) the probability of type one error of Friedman and Lane method is lower than that of Kennedy’s method and modified Kennedy’s method (Huh and Jhun’s method).

The certain considerable point is that, considering Fig. 1a-b, once \( \beta \) equals a small quantity, ignoring the fact that the correlation is high or low and when sample size reaches 30, Friedman-Lane’s method and modified Kennedy’s method which is depicted in the charts as the Huh and Jhun method, both, show a convex quantity.
CONCLUSION

The objective of this article was to select the best test of significance of a single partial regression coefficient in a multiple regression model. Hence the Kennedy’s method, modified Kennedy’s method and Freedman and Lane’s methods compared by simulation. Then, with accordance to the results of the simulation the best method was selected. This led us to the fact that the results Shadrckh’s analytical and results Anderson’s simulation still remain unquestionable after the modification of the problem with Kennedy’s method and it can be asserted that when selecting a method for testing one of the coefficients of multiple regressions, Freedman and Lane’s method is to be the first choice.

REFERENCES


