Article

On Comparing and Classifying Several Independent Linear and Non-Linear Regression Models with Symmetric Errors

Pan Ji-jun [1], Mohammad Reza Mahmoudi [2,*], Dumitru Baleanu [3] and Mohsen Maleki [4]

[1] College of Mathematics, Dianxi Science and Technology, Normal University, Lincang 677000, China; elinhd2013@gmail.com

[2] Department of Statistics, Faculty of Science, Fasa University, Fasa 74616 86131, Iran

[3] Department of Mathematics, Faculty of Art and Sciences, Cankaya University Balgat, Ankara 06530, Turkey; dumitru@cankaya.edu.tr

[4] Department of Statistics, Faculty of Science, Shiraz University, Shiraz 71946 85115, Iran; m.maleki.stat@gmail.com

* Correspondence: mahmoudi.m.r@fasau.ac.ir

Abstract: In many real world problems, science fields such as biology, computer science, data mining, electrical and mechanical engineering, and signal processing, researchers aim to compare and classify several regression models. In this paper, a computational approach, based on the non-parametric methods, is used to investigate the similarities, and to classify several linear and non-linear regression models with symmetric errors. The ability of each given approach is then evaluated using simulated and real world practical datasets.

---

1. Introduction

In many situations, we aim to study the effects of variables $X_1, \ldots, X_k$ on variable Y. Simple and multiple regressions are data analysis techniques to model these effects. The authors of the references [1-2] applied simple and multiple linear regression models in different science fields, such as agriculture, biology, material, mechanical engineering, and signal processing. In many real world problems, scientists want to compare the relationship between the dependent variable and independent variables in several separate datasets.

The comparison of the correlation between the variables X and Y in two separate datasets, different techniques was provided by [3-5]. The comparison of the correlation between the variables X and Y in a dataset, and the correlation between the two variables X and W in another dataset, resulted in different methods developed by [6-10]. The correlation between the variables X and Y in a dataset, and the correlation between two variables W and Z in another dataset, were compared by different methods in [9, 11-12]. The comparison and classification of two, and more simple linear regression models, have been considered in [13-16]. The comparison of two regression models has been reported in [14-22].

In the present research, we aim to compare and classify several linear and non-linear regression models that fitted on several independent datasets. The non-parametric methods are used to construct an approach to investigate the similarity and to classify the linear and non-linear regression models. A given approach is then evaluated using simulation and real world studies. The introduced approach is powerful and applicable in its ability to compare any linear or non-linear regression models.

2. Models Comparing and Classification

Assume $(X_{1j}, \dots, X_{kj}, Y_j)$, $j = 1, \dots, n_i$, is a sample dataset of size n_i, from $(X_1, \dots, X_k, Y)$. The equations of *m* linear or non-linear regression models can be written by:

$$Y_{ij} = f_i(X_{1j}, \dots, X_{kj}) + \varepsilon_{ij}, j = 1, \dots, n_i, \quad i = 1, \dots, m, \tag{1}$$

such that for $i = 1, \dots, m$, $\varepsilon_{ij}, j = 1, \dots, n_i$, are zero-mean symmetric random variables with unknown and equal variance σ_i^2.

By considering Equation (1), consequently, the conditional expectation of Y based on $f_i(X_1, \dots, X_k)$, that we show it by $\theta_i(X_1, \dots, X_k)$, is given by:

$$\theta_i(X_1, \dots, X_k) = E(Y|f_i(X_1, \dots, X_k)) = f_i(X_1, \dots, X_k). \tag{2}$$

In real-word problems the aim is to test the hypothesis $H_0 : \theta_1(X_1, \dots, X_k) = \theta_2(X_1, \dots, X_k) = \cdots = \theta_m(X_1, \dots, X_k)$. Under the rejection of H_0, we conclude that at least two models of the *m* regression models are not statistically similar, and if H_0 is accepted then it can be concluded that the *m* regression models are statistically equal.

The regression equations can be represented by:

$$\boldsymbol{Y_i} = \boldsymbol{f_i}(\boldsymbol{X_1}, \dots, \boldsymbol{X_k}) + \boldsymbol{\varepsilon_i}, \quad i = 1, \dots, m, \tag{3}$$

such that $\boldsymbol{Y_i} = (y_1, \dots, y_{n_i})^T, i = 1, \dots, m$, are the values for the dependent variable Y, $\boldsymbol{X_1} = (x_{11}, \dots, x_{1n_i})^T, \dots, \boldsymbol{X_k} = (x_{k1}, \dots, x_{kn_i})^T, i = 1, \dots, m$ are the values for the independent variables $(X_1, \dots, X_k)$, $\boldsymbol{f_i}(\boldsymbol{X_1}, \dots, \boldsymbol{X_k}) = (f_i(x_{11}, \dots, x_{k1}), \dots, f_i(x_{1n_i}, \dots, x_{kn_i}))^T$, and $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T, i = 1, \dots, m$, are zero-mean random variables with unknown and equal variance σ_i^2.

First, all *m* regression models are estimated by

$$\widehat{\boldsymbol{Y_i}} = \widehat{\boldsymbol{f}}_i(\boldsymbol{X_1}, \dots, \boldsymbol{X_k}), \quad i = 1, \dots, m, \tag{4}$$

for all $n = distinct\ points\ (n_1 \cup n_2 \cup \dots \cup n_m)$ values of $(\boldsymbol{X_1}, \dots, \boldsymbol{X_k})$, where $\widehat{\boldsymbol{Y_i}} = (\hat{y}_{i1}, \dots, \hat{y}_{in})^T, i = 1, \dots, m$, are the estimated values for dependent variable Y, based on i[th] regression model. Since $\varepsilon_i, i = 1, \dots, m$, are zero-mean symmetric random variables, consequently, $\hat{y}_{i1}, \dots, \hat{y}_{in}$, $i = 1, \dots, m$, are unbiased estimators for $\theta_i(X_1, \dots, X_k), i = 1, \dots, m$, respectively. In other words, $\hat{y}_{i1}, \dots, \hat{y}_{in}, i = 1, \dots, m$, are random variables with mean $\theta_i(X_1, \dots, X_k), i = 1, \dots, m$.

Remark 1: $n = distinct\ points\ (n_1 \cup n_2 \cup \dots \cup n_m)$ means that the repeated points are assumed once.

Now, to compare the fitted regression models, the Friedman test [23-26] will be applied on *n* couples $(\hat{y}_{11}, \dots, \hat{y}_{m1}), \dots, (\hat{y}_{1n}, \dots, \hat{y}_{m1})$.

The Friedman test that is a non-parametric alternative to the repeated measures is used to compare related datasets (datasets that are repeated on the same subjects). This test is commonly applied when dataset do not follow the parametric conditions, such as normality assumption.

2.1. Classification

In previous discussion, if H_0 is false, then we conclude that the mechanism of one model or mechanisms of some models are significantly different from the other models. However, to determine which models are significantly different from each other, the sign test or Wilcoxon test are applied in order to compare each of the regression model pairs.

3. Simulation Study

This section assesses the ability of the introduced approach simulation datasets. First, the different datasets from different regression models are produced. Then, we compute the values of the Estimated Type I error probability $(\hat{\alpha})$ and the Estimated Power $(\hat{\pi})$ of the introduced approach. For comparison, the Wilcoxon and Friedman tests are applied. The simulations are accomplished after 1000 runs and using the *R 3.5.3* software (*R Development Core Team*, 2018) on a PC (Processor: Intel(R) CoreTM(2) Duo CPU T7100 @ 1.80GHz 1.80GHz, RAM: 2.00GB, System Type: 32-bit).

Example 1. Assume the simple linear regression model:

$$Y = \beta X + \varepsilon, \tag{5}$$

such that ε and X are independent.

Example 2. Let

$$Y = \beta_0 + \beta_1 X + \beta_2 X \varepsilon, \tag{6}$$

such that ε and X are independent.

Example 3. Assume:

$$Y = 1 + \beta X + \varepsilon, \tag{7}$$

such that ε and X are independent.

Example 4. Assume the multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + 2\beta_2 X_2 + \varepsilon, \tag{8}$$

such that ε, X_1 and X_2 are independent.

Example 5. For the first dataset, assume the simple nonlinear regression model:

$$Y = e^X + \varepsilon, \tag{9}$$

such that ε and X are independent.

For the second and the third datasets let $Y = \{e^X + \varepsilon, 1 + \beta X + \varepsilon\}$, and $Y = \{e^X + \varepsilon, 1 + \beta X + \varepsilon, 2X + \varepsilon\}$, respectively.

Figures 1 and 2 shows the density plots of the some parts of the response variable Y. As it can be seen in these figures, the density plots are symmetric, but not necessarily normal (Figure 2).
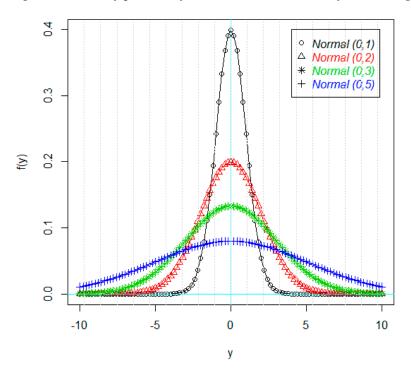

Figure 1. The density plots of the some parts of the response variable Y (Black or Pie: Normal (0,1); Red or Triangle: Normal (0,2); Green or Star: Normal (0,3); Blue or Plus: Normal (0,5)).
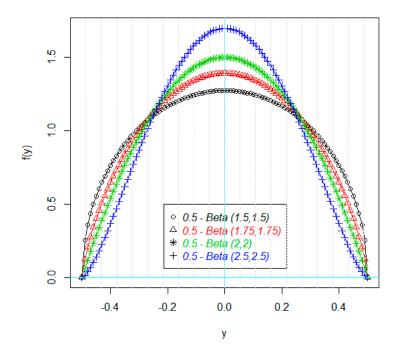
Figure 2. The density plots of the some parts of the response variable Y (Black or Pie: 0.5 – Beta (1.5,1.5); Red or Triangle: 0.5 – Beta (1.75,1.75); Green or Star: 0.5 – Beta (2,2); Blue or Plus: 0.5 – Beta (2.5,2.5)).

The values of $\hat{\alpha}$ (first four rows) and $\hat{\pi}$ (other rows) for Examples 1 to 5 are summarized in Tables 1 to 5, respectively. As Tables 1-5 indicate the values of $\hat{\alpha}$ are very close to size test ($\alpha = 0.05$), and consequently the introduced approach can be controlled the type I error. Also the values of $\hat{\pi}$ show that the given technique can distinguished between the null and alternative hypotheses.

Table 1. The values of $\hat{\alpha}$ and $\hat{\pi}$ for Example 1.

| | | β | | (n_1, n_2, n_3) | | | |
ε	X	Second	Third	$(10, 10, 10)$	$(20, 40, 60)$	$(50, 75, 100)$	$(75, 100, 150)$
Uniform (−2,2)	*Normal* (0,0.25)	1	1	0.053	0.051	0.051	0.049
Uniform (−2,2)	*Exponential* (5)	1	1	0.052	0.052	0.051	0.048
Normal (0,0.5)	*Normal* (0,0.25)	1	1	0.053	0.052	0.050	0.049
Normal (0,0.5)	*Exponential* (5)	1	1	0.053	0.052	0.050	0.049
Uniform (−2,2)	*Normal* (0,0.25)	1	2	0.738	0.882	0.934	0.958
Uniform (−2,2)	*Exponential* (5)	1	2	0.753	0.801	0.950	0.981
Normal (0,0.5)	*Normal* (0,0.25)	1	2	0.754	0.854	0.945	0.972
Normal (0,0.5)	*Exponential* (5)	1	2	0.749	0.889	0.913	0.970
Uniform (−2,2)	*Normal* (0,0.25)	1	3	0.703	0.825	0.941	0.993
Uniform (−2,2)	*Exponential* (5)	1	3	0.710	0.859	0.917	0.975
Normal (0,0.5)	*Normal* (0,0.25)	1	3	0.728	0.824	0.910	0.984
Normal (0,0.5)	*Exponential* (5)	1	3	0.707	0.864	0.934	0.953
Uniform (−2,2)	*Normal* (0,0.25)	2	1	0.768	0.828	0.913	0.978
Uniform (−2,2)	*Exponential* (5)	2	1	0.703	0.824	0.928	0.951
Normal (0,0.5)	*Normal* (0,0.25)	2	1	0.794	0.846	0.930	0.968
Normal (0,0.5)	*Exponential* (5)	2	1	0.794	0.800	0.903	0.955
Uniform (−2,2)	*Normal* (0,0.25)	2	2	0.745	0.813	0.946	0.971
Uniform (−2,2)	*Exponential* (5)	2	2	0.718	0.858	0.937	0.981
Normal (0,0.5)	*Normal* (0,0.25)	2	2	0.784	0.866	0.901	0.953

Normal (0,0.5)	*Exponential* (5)	2	2	0.726	0.821	0.944	0.999
Uniform (−2,2)	*Normal* (0,0.25)	2	3	0.795	0.849	0.924	0.982
Uniform (−2,2)	*Exponential* (5)	2	3	0.755	0.856	0.928	0.961
Normal (0,0.5)	*Normal* (0,0.25)	2	3	0.763	0.845	0.936	0.988
Normal (0,0.5)	*Exponential* (5)	2	3	0.710	0.865	0.914	0.975

Table 2. The values of $\hat{\alpha}$ and $\hat{\pi}$ for Example 2.

		$(\beta_0, \beta_1, \beta_2)$		(n_1, n_2, n_3)			
ε	X	Second	Third	$(10, 10, 10)$	$(20, 40, 60)$	$(50, 75, 100)$	$(75,100,150)$
Uniform (−1,1)	*Normal* (0,1)	(2,1,2)	(2,1,2)	0.052	0.052	0.051	0.049
Uniform (−1,1)	*Exponential* (1)	(2,1,2)	(2,1,2)	0.053	0.051	0.050	0.049
Normal (0,2)	*Normal* (0,1)	(2,1,2)	(2,1,2)	0.053	0.052	0.051	0.049
Normal (0,2)	*Exponential* (1)	(2,1,2)	(2,1,2)	0.052	0.052	0.051	0.049
Uniform (−1,1)	*Normal* (0,1)	(2,1,2)	(0,2,1)	0.770	0.843	0.903	0.981
Uniform (−1,1)	*Exponential* (1)	(2,1,2)	(0,2,1)	0.743	0.817	0.909	0.979
Normal (0,2)	*Normal* (0,1)	(2,1,2)	(0,2,1)	0.771	0.842	0.918	0.992
Normal (0,2)	*Exponential* (1)	(2,1,2)	(0,2,1)	0.791	0.855	0.934	0.967
Uniform (−1,1)	*Normal* (0,1)	(2,1,2)	(3,2,1)	0.737	0.891	0.941	0.997
Uniform (−1,1)	*Exponential* (1)	(2,1,2)	(3,2,1)	0.798	0.860	0.932	0.988
Normal (0,2)	*Normal* (0,1)	(2,1,2)	(3,2,1)	0.740	0.849	0.947	0.993
Normal (0,2)	*Exponential* (1)	(2,1,2)	(3,2,1)	0.712	0.827	0.916	0.997
Uniform (−1,1)	*Normal* (0,1)	(0,2,1)	(2,1,2)	0.782	0.837	0.932	0.966
Uniform (−1,1)	*Exponential* (1)	(0,2,1)	(2,1,2)	0.780	0.830	0.936	0.960
Normal (0,2)	*Normal* (0,1)	(0,2,1)	(2,1,2)	0.720	0.857	0.945	0.998
Normal (0,2)	*Exponential* (1)	(0,2,1)	(2,1,2)	0.767	0.897	0.902	0.958
Uniform (−1,1)	*Normal* (0,1)	(0,2,1)	(0,2,1)	0.790	0.809	0.921	0.992
Uniform (−1,1)	*Exponential* (1)	(0,2,1)	(0,2,1)	0.741	0.814	0.935	0.992
Normal (0,2)	*Normal* (0,1)	(0,2,1)	(0,2,1)	0.710	0.844	0.945	0.981
Normal (0,2)	*Exponential* (1)	(0,2,1)	(0,2,1)	0.760	0.871	0.906	0.972
Uniform (−1,1)	*Normal* (0,1)	(0,2,1)	(3,2,1)	0.776	0.807	0.919	0.969
Uniform (−1,1)	*Exponential* (1)	(0,2,1)	(3,2,1)	0.701	0.875	0.928	0.963
Normal (0,2)	*Normal* (0,1)	(0,2,1)	(3,2,1)	0.780	0.803	0.936	0.987
Normal (0,2)	*Exponential* (1)	(0,2,1)	(3,2,1)	0.720	0.886	0.923	0.960

Table 3. The values of $\hat{\alpha}$ and $\hat{\pi}$ for Example 3.

		β		(n_1, n_2, n_3)			
ε	X	Second	Third	$(10, 10, 10)$	$(20, 40, 60)$	$(50, 75, 100)$	$(75,100,150)$
Uniform (−1,1)	Geometric (0.4)	1	1	0.053	0.051	0.050	0.049
Uniform (−1,1)	Binomial (2, 0.7)	1	1	0.053	0.051	0.051	0.050
Normal (0,0.5)	Geometric (0.4)	1	1	0.053	0.051	0.051	0.050
Normal (0,0.5)	Binomial (2, 0.7)	1	1	0.053	0.051	0.051	0.050
Uniform (−1,1)	Geometric (0.4)	1	2	0.724	0.846	0.924	0.996
Uniform (−1,1)	Binomial (2, 0.7)	1	2	0.734	0.813	0.942	0.952
Normal (0,0.5)	Geometric (0.4)	1	2	0.737	0.818	0.914	0.959
Normal (0,0.5)	Binomial (2, 0.7)	1	2	0.764	0.819	0.949	0.998
Uniform (−1,1)	Geometric (0.4)	1	5	0.797	0.808	0.904	0.959
Uniform (−1,1)	Binomial (2, 0.7)	1	5	0.760	0.869	0.919	0.978
Normal (0,0.5)	Geometric (0.4)	1	5	0.793	0.843	0.917	0.988
Normal (0,0.5)	Binomial (2, 0.7)	1	5	0.765	0.876	0.910	0.983
Uniform (−1,1)	Geometric (0.4)	2	1	0.742	0.868	0.934	0.954
Uniform (−1,1)	Binomial (2, 0.7)	2	1	0.730	0.810	0.925	0.966

Normal (0,0.5)	Geometric (0.4)	2	1	0.725	0.867	0.911	0.981
Normal (0,0.5)	Binomial (2, 0.7)	2	1	0.769	0.868	0.930	0.996
Uniform (−1,1)	Geometric (0.4)	2	2	0.763	0.816	0.905	0.982
Uniform (−1,1)	Binomial (2, 0.7)	2	2	0.706	0.895	0.935	0.951
Normal (0,0.5)	Geometric (0.4)	2	2	0.723	0.866	0.909	0.981
Normal (0,0.5)	Binomial (2, 0.7)	2	2	0.765	0.857	0.903	0.974
Uniform (−1,1)	Geometric (0.4)	2	5	0.710	0.867	0.910	0.950
Uniform (−1,1)	Binomial (2, 0.7)	2	5	0.764	0.837	0.904	0.981
Normal (0,0.5)	Geometric (0.4)	2	5	0.778	0.891	0.933	0.987
Normal (0,0.5)	Binomial (2, 0.7)	2	5	0.726	0.819	0.946	0.967

Table 4. The values of $\hat{\alpha}$ and $\hat{\pi}$ for Example 4.

X_1	X_2	$(\beta_0, \beta_1, \beta_2)$		(n_1, n_2, n_3)			
		Second	Third	$(\mathbf{10, 10, 10})$	$(\mathbf{20, 40, 60})$	$(\mathbf{50, 75, 100})$	(75,100,150)
Uniform (0,2)	Exponential (5)	(2,1,2)	(2,1,2)	0.052	0.052	0.050	0.049
Uniform (0,2)	Geometric (0.3)	(2,1,2)	(2,1,2)	0.053	0.052	0.050	0.049
Binomial (3, 0.5)	Exponential (5)	(2,1,2)	(2,1,2)	0.052	0.052	0.051	0.049
Binomial (3, 0.5)	Geometric (0.3)	(2,1,2)	(2,1,2)	0.052	0.051	0.050	0.048
Uniform (0,2)	Exponential (5)	(2,1,2)	(0,2,1)	0.734	0.893	0.923	0.961
Uniform (0,2)	Geometric (0.3)	(2,1,2)	(0,2,1)	0.787	0.887	0.947	0.964
Binomial (3, 0.5)	Exponential (5)	(2,1,2)	(0,2,1)	0.766	0.813	0.943	0.973
Binomial (3, 0.5)	Geometric (0.3)	(2,1,2)	(0,2,1)	0.762	0.897	0.909	0.993
Uniform (0,2)	Exponential (5)	(2,1,2)	(3,2,1)	0.706	0.866	0.936	0.966
Uniform (0,2)	Geometric (0.3)	(2,1,2)	(3,2,1)	0.746	0.882	0.946	0.960
Binomial (3, 0.5)	Exponential (5)	(2,1,2)	(3,2,1)	0.716	0.875	0.948	0.975
Binomial (3, 0.5)	Geometric (0.3)	(2,1,2)	(3,2,1)	0.757	0.811	0.939	0.950
Uniform (0,2)	Exponential (5)	(0,2,1)	(2,1,2)	0.792	0.866	0.936	0.985
Uniform (0,2)	Geometric (0.3)	(0,2,1)	(2,1,2)	0.768	0.824	0.902	0.995
Binomial (3, 0.5)	Exponential (5)	(0,2,1)	(2,1,2)	0.773	0.841	0.933	0.983
Binomial (3, 0.5)	Geometric (0.3)	(0,2,1)	(2,1,2)	0.795	0.801	0.940	0.992
Uniform (0,2)	Exponential (5)	(0,2,1)	(0,2,1)	0.790	0.891	0.912	0.953
Uniform (0,2)	Geometric (0.3)	(0,2,1)	(0,2,1)	0.784	0.855	0.924	0.951
Binomial (3, 0.5)	Exponential (5)	(0,2,1)	(0,2,1)	0.739	0.842	0.908	0.961
Binomial (3, 0.5)	Geometric (0.3)	(0,2,1)	(0,2,1)	0.749	0.880	0.905	0.963
Uniform (0,2)	Exponential (5)	(0,2,1)	(3,2,1)	0.745	0.854	0.918	0.956
Uniform (0,2)	Geometric (0.3)	(0,2,1)	(3,2,1)	0.739	0.825	0.946	0.955
Binomial (3, 0.5)	Exponential (5)	(0,2,1)	(3,2,1)	0.743	0.883	0.926	0.960
Binomial (3, 0.5)	Geometric (0.3)	(0,2,1)	(3,2,1)	0.734	0.840	0.918	0.976

Table 5. The values of $\hat{\alpha}$ and $\hat{\pi}$ for Example 5.

ε	X	Y		(n_1, n_2, n_3)			
		Second	Third	$(\mathbf{10, 10, 10})$	$(\mathbf{20, 40, 60})$	$(\mathbf{50, 75, 100})$	(75,100,150)
Uniform (−2,2)	Normal (0,0.5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.052	0.051	0.051	0.048
Uniform (−2,2)	Poisson (5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.053	0.051	0.051	0.049
Normal (0,0.25)	Normal (0,0.5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.052	0.051	0.051	0.049
Normal (0,0.25)	Poisson (5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.052	0.051	0.050	0.048
Uniform (−2,2)	Normal (0,0.5)	$e^X + \varepsilon$	$1 + \beta X + \varepsilon$	0.787	0.895	0.901	0.965
Uniform (−2,2)	Poisson (5)	$e^X + \varepsilon$	$1 + \beta X + \varepsilon$	0.787	0.829	0.930	0.974
Normal (0,0.25)	Normal (0,0.5)	$e^X + \varepsilon$	$1 + \beta X + \varepsilon$	0.725	0.848	0.912	0.991
Normal (0,0.25)	Poisson (5)	$e^X + \varepsilon$	$1 + \beta X + \varepsilon$	0.759	0.898	0.944	0.984
Uniform (−2,2)	Normal (0,0.5)	$e^X + \varepsilon$	$2X + \varepsilon$	0.734	0.891	0.949	0.962
Uniform (−2,2)	Poisson (5)	$e^X + \varepsilon$	$2X + \varepsilon$	0.788	0.811	0.921	0.981

Normal (0,0.25)	*Normal* (0,0.5)	$e^X + \varepsilon$	$2X + \varepsilon$	0.759	0.877	0.941	0.965
Normal (0,0.25)	*Poisson* (5)	$e^X + \varepsilon$	$2X + \varepsilon$	0.704	0.868	0.948	0.989
Uniform (−2,2)	*Normal* (0,0.5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.798	0.845	0.908	0.956
Uniform (−2,2)	*Poisson* (5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.753	0.809	0.927	0.989
Normal (0,0.25)	*Normal* (0,0.5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.731	0.865	0.910	0.990
Normal (0,0.25)	*Poisson* (5)	$e^X + \varepsilon$	$e^X + \varepsilon$	0.731	0.820	0.906	0.962
Uniform (−2,2)	*Normal* (0,0.5)	$1 + \beta X + \varepsilon$	$1 + \beta X + \varepsilon$	0.723	0.897	0.934	0.960
Uniform (−2,2)	*Poisson* (5)	$1 + \beta X + \varepsilon$	$1 + \beta X + \varepsilon$	0.799	0.807	0.949	0.982
Normal (0,0.25)	*Normal* (0,0.5)	$1 + \beta X + \varepsilon$	$1 + \beta X + \varepsilon$	0.713	0.877	0.916	0.952
Normal (0,0.25)	*Poisson* (5)	$1 + \beta X + \varepsilon$	$1 + \beta X + \varepsilon$	0.743	0.872	0.925	0.965
Uniform (−2,2)	*Normal* (0,0.5)	$1 + \beta X + \varepsilon$	$2X + \varepsilon$	0.725	0.892	0.901	0.996
Uniform (−2,2)	*Poisson* (5)	$1 + \beta X + \varepsilon$	$2X + \varepsilon$	0.795	0.886	0.944	0.959
Normal (0,0.25)	*Normal* (0,0.5)	$1 + \beta X + \varepsilon$	$2X + \varepsilon$	0.707	0.821	0.925	0.972
Normal (0,0.25)	*Poisson* (5)	$1 + \beta X + \varepsilon$	$2X + \varepsilon$	0.798	0.825	0.924	0.974

4. Real data

In this section, a practical real data is considered to study the power of the introduced approach in real world problems. Drought is a damaging natural phenomenon. To prevent this phenomenon, the hydrologists model and predict the drought datasets in a standard time period. In this research, the average monthly rainy days (1966 – 2010) at three Iranian synoptic stations (Fasa, Sarvestan, and Shiraz) was considered and modeled.

To model and forecast the average monthly rainy days, different polynomial regression models of orders 1 to 3 (linear, quadratic and cubic) and exponential model were fitted to datasets. The formulas of the considered models are as following:

Linear model: $Y = \beta_0 + \beta_1 X + \varepsilon$ Quadratic model: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$. (11)

$\quad\quad$ ➢ $\quad$ Cubic model: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$. $\quad\quad\quad\quad$ (12)

$\quad\quad\quad$ ➢ $\quad$ Exponential model: $Y = \beta_0 + \beta_1 e^{\beta_2 X} + \varepsilon$. $\quad\quad\quad$ (13)

The numerical computations are done using the R 3.5.3 software (Library 'nlstools', lm() function for linear regression and nls() function for nonlinear regression) and Minitab 18 software.

The results of fitted regression models are summarized in Table 6. It can be observed that, for all of the stations, respectively, the polynomial regression of order 3 (cubic), and the exponential models, had the most R-square (R^2) and the least root mean square error (*RMSE)* between all fitted models.

Table 6. Indices to evaluate the fitted regression models.

Model	Station	R Square	RMSE
Linear	Fasa	0.624	1.693
	Sarvestan	0.638	1.516
	Shiraz	0.689	1.501
Quadratic	Fasa	0.734	1.350
	Sarvestan	0.743	1.285
	Shiraz	0.767	1.265
Cubic	Fasa	0.895	0.910
	Sarvestan	0.899	0.855
	Shiraz	0.976	0.529
Exponential	Fasa	0.767	0.978
	Sarvestan	0.778	0.926
	Shiraz	0.876	0.713

Now, we use the proposed approach to compare and classify these stations, for each model. The result of Friedman test is shown in Table 7. This table indicated that the fitted cubic and exponential models are significantly different in these stations (p<0.05). Also, there is no significant difference between the fitted linear and quadratic models in these stations (p >0.05).

Table 7. Friedman test to compare the stations.

Model	p
Linear	0.123
Quadratic	0.224
Cubic	<0.001
Exponential	<0.001

As Table 8 indicates, we can classify the stations in two clusters, for cubic and exponential models. First cluster: Fasa and Sarvestan, and second cluster: Shiraz.

Table 8. Wilcoxon test to compare and classify the stations.

Model		Stations	P
Cubic	Pair 1	Shiraz - Fasa	0.011
	Pair 2	Shiraz - Sarvestan	0.003
	Pair 3	Fasa - Sarvestan	0.144
Exponential	Pair 1	Shiraz - Fasa	0.019
	Pair 2	Shiraz - Sarvestan	<0.001
	Pair 3	Fasa - Sarvestan	0.112

5. Conclusion

In many real world problems, researchers wish to compare and classify the regression models in several datasets. In this paper, the non-parametric methods were used to construct an approach to investigate the similarity of some linear and non-linear regression models with symmetric errors. Particular approaches were evaluated using simulation and practical datasets. A simulation study indicated that the introduced approach controlled the Type I error. Also the proposed technique distinguished well between null and alternative hypotheses. The introduced approach also had many advantages. First, it was powerful. Second, it was not too computational. Third, it could be applied to compare any linear or non-linear regression models. Fourth, this method did not need the normality of errors and could be applied for all models with symmetric errors.

References

1. Wan, J.; Zhang, D.; Xu, W.; Guo, Q. Parameter Estimation of Multi Frequency Hopping Signals Based on Space-Time-Frequency Distribution. *Symmetry* **2019**, *11*, 648.
2. Sajid, M.; Shafique, T.; Riaz, I.; Imran, M.; Jabbar Aziz Baig, M.; Baig, S.; Manzoor, S. Facial asymmetry-based anthropometric differences between gender and ethnicity. *Symmetry* **2018**, *10*, 232.
3. Mahmouudi, M.R.; Maleki, M.; Pak, A. Testing the Difference between Two Independent Time Series Models. *Iran. J. Sci. Technol. Trans. A Sci.* **2017**, *41*, 665–669.
4. Fisher, R.A. On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample. *Metron* **1921**, *1*, 3–32.
5. Mahmoudi, M.R.; Mahmoodi, M. Inferrence on the Ratio of Correlations of Two Independent Populations. *J. Math. Ext.* **2014**, *7*, 71–82.
6. Howell, D.C. *Statistical Methods for Psychology*, 6th ed.; Thomson Wadsworth: Stamford, CT, USA, 2007.
7. Hotelling, H. The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters. *Ann. Math. Stat.* **1940**, *11*, 271–283.

8.   Williams, E.G. The Comparison of Regression Variables. *J. R. Stat. Soc. Ser. B* **1959**, *21*, 396–399.

9.   Steiger, J.H. Tests for Comparing Elements of a Correlation Matrix. *Psychol. Bull.* **1980**, *87*, 245–251.

10. Meng, X.; Rosenthal, R.; Rubin, D.B. Comparing Correlated Correlation Coefficients. *Psychol. Bull.* **1992**, *111*, 172–175.

11. Peter, C.C.; Van Voorhis, W.R. *Statistical Procedures and Their Mathematical Bases*; McGraw-Hill: New York, NY, USA, 1940.

12. Raghunathan, T.E.; Rosenthal, R.; Rubin, D.B. Comparing Correlated but Nonoverlapping Correlations. *Psychol. Methods* **1996**, *1*, 178–183.

13. Liu, W.; Jamshidian, M.; Zhang, Y. Multiple Comparison of Several Linear Regression Lines. *J. R. Stat. Soc. Ser. B* **2004**, *99*, 395–403.

14. Liu, W.; Hayter, A.J.; Wynn, H.P. Operability Region Equivalence: Simultaneous Confidence Bands for the Equivalence of Two Regression Models Over Restricted Regions. *Biom. J.* **2007**, *49*, 144–150.

15. Liu, W.; Jamshidian, M.; Zhang, Y.; Bertz, F.; Han, X. Pooling Batches in Drug Stability Study by Using Constant-width Simultaneous Confidence Bands. *Stat. Med.* **2007**, *26*, 2759–2771.

16. Liu, W.; Jamshidian, M.; Zhang, Y.; Bertz, F.; Han, X. Some New Methods for the Comparison of Two Linear Regression Models. *J. Stat. Plan. Inference* **2007**, *137*, 57–67.

17. Hayter, A.J.; Liu, W.; Wynn, H.P. Easy-to-Construct Confidence Bands for Comparing Two Simple Linear Regression Lines. *J. Stat. Plan. Inference* **2007**, *137*, 1213–1225.

18. Jamshidian, M.; Liu, W.; Bretz, F. Simultaneous Confidence Bands for all Contrasts of Three or More Simple Linear Regression Models over an Interval. *Comput. Stat Data Anal.* **2010**, *54*, 1475–1483.

19. Marques, F.J.; Coelho, C.A.; Rodrigues, P.C. Testing the equality of several linear regression models. *Comput. Stat.* **2016**, *32*, 1453–1480.

20. Mahmoudi, M.R.; Mahmoudi, M.; Nahavandi, E. Testing the Difference between Two Independent Regression Models. *Commun. Stat. Theory Methods* **2016**, *45*, 6284–6289.

21. Mahmoudi, M.R. On Comparing Two Dependent Linear and Nonlinear Regression Models. *J. Test. Eval.* **2018**, *47*, 449–458.

22. Mahmoudi, M.R.; Maleki, M.; Pak, A. Testing the Equality of Two Independent Regression Models. *Commun. Stat. Theory Methods* **2018**, *47*, 2919–2926.

23. Conover, W.J. *Practical Nonparametric Statistics*, 3rd ed.; John Wiley: Hoboken, NJ, USA, 1980.

24. Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. R. Stat. Soc. Ser. B* **1937**, *32*, 675–701.

25. Friedman, M. A Correction: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. R. Stat. Soc. Ser. B* **1939**, *34*, 109.

26. Friedman, M. A Comparison of Alternative Tests of Significance for the Problem of *m* Rankings. *Ann. Math. Stat.* **1940**, *11*, 86–92.