# SOME DISTRIBUTION THEORY RESULTS FOR A REGRESSION MODEL

PRAKASH C. JOSHI

(Received Sept. 13, 1972)


## Summary

The joint distribution of two weighted residuals for a normal theory regression model is derived and some of its properties are studied. A useful bound depending on the residual variances for the correlation coefficient between any two residuals is obtained. An application of this bound in the detection of a single outlier is also considered.


*Some key words*: Linear regression; Residuals; Correlations; Normal distribution; Outliers.

## 1. Introduction

Consider a normal theory linear regression model described by

$$\boldsymbol{y} \text{ is distributed as } N(\boldsymbol{X}'\boldsymbol{\beta}, \sigma^2\boldsymbol{I}) ,$$

where $\boldsymbol{y}$ is an $(n\times 1)$ observation vector, $\boldsymbol{X}$ is a known $(m\times n)$ matrix of rank $r$ $(r\leqq m<n)$, $\boldsymbol{\beta}$ is an $(m\times 1)$ parameter vector and $\boldsymbol{I}$ is the identity matrix of order $n$. For this model, the residual vector, $\boldsymbol{e}$, and the residual sum of squares, $S^2$, are given by

$$\boldsymbol{e}=\boldsymbol{\Lambda y} \qquad \text{and} \qquad S^2=\boldsymbol{y}'\boldsymbol{\Lambda y}$$

respectively, where $\boldsymbol{\Lambda}=\boldsymbol{I}-\boldsymbol{X}'(\boldsymbol{XX}')^-\boldsymbol{X}=((\lambda_{ij}))$ is an idempotent matrix of rank $n-r$. Here $\boldsymbol{A}^-$ denotes a generalized inverse of $\boldsymbol{A}$ satisfying $\boldsymbol{AA}^-\boldsymbol{A}=\boldsymbol{A}$ (Rao [7], p. 24).

Now $\boldsymbol{e}$ has a multivariate normal distribution with mean vector $\boldsymbol{0}$ and covariance matrix $\sigma^2\boldsymbol{\Lambda}$. Further $S^2/\sigma^2$ has a $\chi^2$ distribution with $n-r$ degrees of freedom. Let $s_\nu$ be a root mean square estimator of $\sigma$ based on $\nu$ degrees of freedom and independent of $\boldsymbol{y}$. For $i=1,\cdots,n$, let

$$(1) \qquad z_i=e_i/\sqrt{\lambda_{ii}} \qquad \text{and} \qquad w_i=z_i/S_p ,$$

where $S_p^2=S^2+\nu s_\nu^2$ is the pooled sum of squares based on $p=n-r+\nu$

degrees of freedom. In this paper, the joint distribution of $w_i$ and $w_j$ is derived and some of its properties are studied. These results are also mentioned in Joshi [4].

## 2. Marginal and joint distribution of weighted residuals

Throughout this paper we will assume that $|\rho_{ij}|<1$ for all $i\neq j$, where $\rho_{ij}=\lambda_{ij}/(\lambda_{ii}\lambda_{jj})^{1/2}$ is the correlation coefficient between the residuals $e_i$ and $e_j$. It will be shown in Section 4 that a sufficient condition for its validity is that $\lambda_{ii}+\lambda_{jj}>1$ $(i\neq j)$.

Without any loss of generality we take $\sigma=1$ and consider the distribution of $\boldsymbol{w}'=(w_1, w_2)$. Let $\rho_{12}=\rho$ and $\boldsymbol{P}$ be the correlation matrix of $\boldsymbol{z}'=(z_1, z_2)$. By our assumption, $\boldsymbol{P}$ is positive definite and $p$ is greater than 1. For $p=2$, $\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w}\equiv 1$, so that $\boldsymbol{w}'$ has a singular distribution. We therefore assume that $p\geqq 3$.

Let $Q=S_p^2-\boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z}$. Now using the independence of $z_1$, $(z_2-\rho z_1)/(1-\rho^2)^{1/2}$ and $Q$ we see that the joint distribution of $\boldsymbol{z}$ and $Q$ is

$$g(\boldsymbol{z}, Q)=CQ^{(p-4)/2}\exp\{-1/2(Q+\boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z})\}\ ,$$

where $C$ is a generic constant. Make a transformation

$$Q=Q\ ,\qquad w_i=z_i/(Q+\boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z})^{1/2}\ ,\quad i=1, 2\ .$$

The jacobian of the transformation, after some simplifications, is $Q/(1-\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w})^2$ and the joint distribution of $\boldsymbol{w}$ and $Q$ is

$$g(\boldsymbol{w}, Q)=CQ^{(p-2)/2}(1-\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w})^{-2}\exp[-Q/\{2(1-\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w})\}]\ .$$

Integrating out $Q$ from 0 to $\infty$, we get

$$g(\boldsymbol{w})=C(1-\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w})^{(p-4)/2}\ , \tag{2}$$

where the region of positive density is the interior of the ellipse $\boldsymbol{w}'\boldsymbol{P}^{-1}\boldsymbol{w}=1$, i.e.,

$$w_1^2-2\rho w_1w_2+w_2^2=1-\rho^2$$

and the constant $C$ is given by

$$C=(p-2)/2\pi\sqrt{(1-\rho^2)}\ .$$

Integrating out $w_1$ or $w_2$ we see that the marginal distribution of $w_i$ is

$$f(w_i)=\frac{1}{B\{1/2, 1/2(p-1)\}}(1-w_i^2)^{(p-3)/2}\ ,\qquad -1\leqq w_i\leqq 1\ .$$

This is the same for all $i$ and can be derived by other methods also (Srikantan [8]) and is valid for $p \geqq 2$.

## 3. Related results

McFadden [5] has studied the bivariate distribution given at (2), viz.,

$$g(x, y)=\frac{p-2}{2\pi\sqrt{(1-\rho^2)}}\left(1-\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right)^{(p-4)/2} \tag{3}$$

inside the ellipse

$$x^2-2\rho xy+y^2=1-\rho^2 . \tag{4}$$

He has derived a diagonal expansion for $g(x, y)$ and has also obtained the characteristic function of $x$ and $y$. Here, we consider some additional results concerning this distribution.

A method for evaluating the bivariate probability $\Pr(x \geqq h,\ y \geqq k) = L(h, k, \rho, p)$ for $\rho=-1/(n-1)$ by fitting an increasing number of planes to the density surface is given by Quesenberry and David [6]. We will now describe an alternative method of finding $L(h, k, \rho, p)$ by expressing it as a single integral and then using numerical integration.

It is clear that we only need to consider the case when both $h$ and $k$ are non-negative. Let 0 be the origin and $A$ be the point $(h, k)$ which

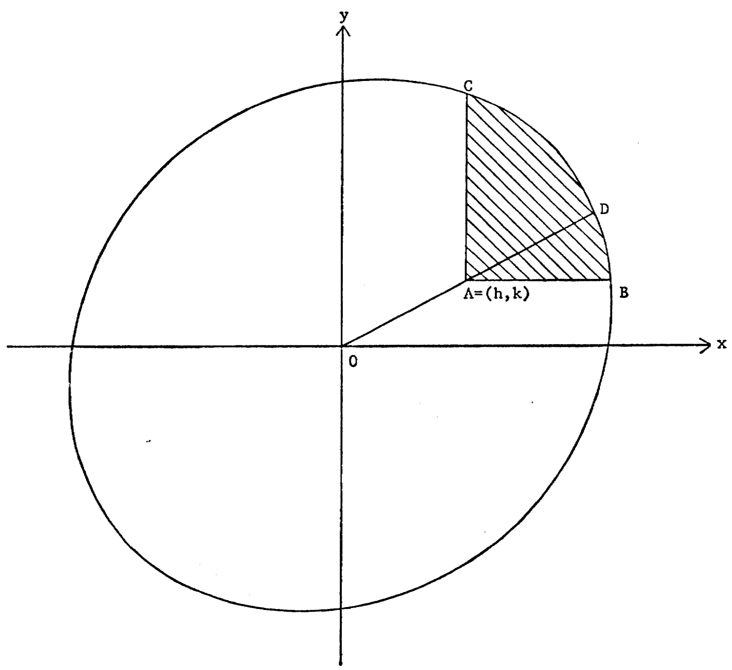


Fig. 1 The region of integration: $x \geqq h$, $y \geqq k$ and $x^2-2\rho xy+y^2 \leqq 1-\rho^2$.

lies inside the ellipse (4) in the $(x, y)$-plane (Fig. 1). The region of integration is the shaded area $ABCA$, where $B=(\rho k+(1-\rho^2)^{1/2}(1-k^2)^{1/2}, k)$ and $C=(h, \rho h+(1-\rho^2)^{1/2}(1-h^2)^{1/2})$.

Let $D$ be the point of intersection of the extended line $0A$ and the ellipse in the first quadrant and let $M(h, k, \rho, p)=\Pr((x, y)\in ABDA)$. Then due to symmetry

$$L(h, k, \rho, p)=M(h, k, \rho, p)+M(k, h, \rho, p) . \tag{5}$$

To find an expression for $M(h, k, \rho, p)$, we assume that $k>0$ and make a transformation $r\cos\theta=(x-\rho y)/\sqrt{(1-\rho^2)}$ and $r\sin\theta=y$. Then

$$\begin{aligned} M(h, k, \rho, p)&=\int_{\arctan(k/\sqrt{(1-k^2)})}^{\arctan(k\sqrt{(1-\rho^2)}/(h-\rho k))}\int_{k\operatorname{cosec}\theta}^{1}\frac{p-2}{2\pi}r(1-r^2)^{(p-4)/2}dr d\theta \\ &=\frac{1}{2\pi}\int_{\arctan(k/\sqrt{(1-k^2)})}^{\arctan(k\sqrt{(1-\rho^2)}/(h-\rho k))}(1-k^2\operatorname{cosec}^2\theta)^{(p-2)/2}d\theta . \end{aligned}$$

On putting $\tan\theta=k(1-u^2)^{1/2}/(h-uk)$, this can be rewritten as

$$M(h, k, \rho, p)=\frac{1}{2\pi}\int\frac{k(k-uh)}{(h^2+k^2-2uhk)\sqrt{(1-u^2)}}\left(1-\frac{h^2+k^2-2uhk}{1-u^2}\right)^{(p-2)/2}du ,$$

where the range of integration is

$$hk-(1-h^2)^{1/2}(1-k^2)^{1/2}\leqq u\leqq\rho . \tag{6}$$

Note that this holds for $k=0$ as well, because $M(h, 0, \rho, p)=0$. The expression for $M(k, h, \rho, p)$ is similar, with $h$ and $k$ interchanged. Substituting this in equation (5) we get

$$L(h, k, \rho, p)=\frac{1}{2\pi}\int\frac{1}{\sqrt{(1-u^2)}}\left(1-\frac{h^2+k^2-2uhk}{1-u^2}\right)^{(p-2)/2}du ,$$

where the range of integration is given at (6). Putting $u=\cos w$, we finally have

$$\begin{aligned} L(h, k, \rho, p)=\frac{1}{2\pi}&\int_{\arccos\rho}^{\arccos(hk-(1-h^2)^{1/2}(1-k^2)^{1/2})} \\ &\cdot\left(1-\frac{h^2+k^2-2hk\cos w}{\sin^2 w}\right)^{(p-2)/2}dw . \end{aligned}$$

This gives the desired expression for non-negative values of $h$ and $k$. For other values of $h$, $k$, the bivariate probabilities can be evaluated by using the symmetry properties of $g(x, y)$.

Doornbos ([3], p. 21) has also considered this distribution in a different context. He has shown that $\Pr(x\leqq h,\ y\leqq k)\leqq\Pr(x\leqq h)\Pr(y\leqq k)$, where $h$ and $k$ are of the same sign and $\rho=-1/(n-1)$. Following an identical approach, this can be extended for other values of $\rho\leqq 0$. An application of this inequality for the detection of outliers is given in Joshi [4]. It should be noted that this result is not true for $\rho>0$, since

$$\Pr(x\leqq 0,\ y\leqq 0)=\frac{1}{4}+\frac{1}{2\pi}\arcsin\rho\ .$$

## 4. Correlations between residuals

Clearly, the correlation matrix $\boldsymbol{R}=((\rho_{ij}))$ depends on the design matrix $\boldsymbol{X}$. Except in some cases, which in general occur in the analysis of variance problems, it is difficult to write down the matrix $\boldsymbol{R}$ theoretically. In some statistical problems it is necessary to compute $\boldsymbol{R}$ and know about the magnitude of correlation coefficients; for example, in the detection of outliers we want $|\rho_{ij}|$ to be less than 1 (Anscombe [1], Joshi [4]). Here, we show that

$$\rho_{ij}^2\leqq(1-\lambda_{ii})(1-\lambda_{jj})/(\lambda_{ii}\lambda_{jj})\ . \tag{7}$$

PROOF. Since $\boldsymbol{A}=\boldsymbol{I}-\boldsymbol{\Lambda}=((a_{ij}))$ is a covariance matrix, hence $a_{ij}^2\leqq a_{ii}a_{jj}$. But $a_{ii}=1-\lambda_{ii}$ and $a_{ij}=-\lambda_{ij}$. Consequently $\lambda_{ij}^2\leqq(1-\lambda_{ii})(1-\lambda_{jj})$ and the result follows.

Equation (7) shows that if $\lambda_{ii}+\lambda_{jj}>1$, then $|\rho_{ij}|<1$. Note that $\boldsymbol{\Lambda}$ is an idempotent matrix of rank $n-r$ and therefore $\operatorname{trace}(\boldsymbol{\Lambda})=\operatorname{rank}(\boldsymbol{\Lambda})$, that is, $\sum_{k=1}^{n}\lambda_{kk}=n-r$. Consequently if $n$ is large compared to $r$, then the condition $\lambda_{ii}+\lambda_{jj}>1$ is likely to hold.

For a particular regression model, some of the $\rho_{ij}$'s could be close to zero and then the universal bound given at (7) may not be sharp. However, for moderate values of $|\rho_{ij}|$, it is expected to be reasonably good. This is illustrated in the following examples.

1. *Regression on one variable* ($m=1$). For the simple case $\mathrm{E}(y_i)=\beta$ $(i=1,\cdots,n)$, we have $\lambda_{ii}=(n-1)/n$ $(i=1,\cdots,n)$ and $\lambda_{ij}=-1/n$ $(i\neq j)$. Consequently $\rho_{ij}=-1/(n-1)$ and equation (7) yields $|\rho_{ij}|\leqq 1/(n-1)$.

2. *Simple linear regression* ($m=2$). Let

$$\boldsymbol{X}=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_n\end{bmatrix}$$

be a matrix of rank 2. Without any loss of generality we assume that $\sum x_i=0$. Then

$$\lambda_{ii}=\frac{n-1}{n}-\frac{x_i^2}{\sum x_i^2}\qquad(i=1,\cdots,n)\,,$$

$$\lambda_{ij}=-\left(\frac{1}{n}+\frac{x_i x_j}{\sum x_i^2}\right) \qquad (i\neq j)\,.$$

With the help of these values, the bound for $|\rho_{ij}|$ can be compared with the exact values. If $x_i$'s are equally spaced, then $(n-4)/n<\lambda_{ii}\leqq (n-1)/n$ (Anscombe [1]) and hence we must have $|\rho_{ij}|<4/(n-4)$ for all $i\neq j$. Of course, for individual $\rho_{ij}$'s the bound in (7) may be considerably better than this crude bound. Thus, for example, for $x_i=i-(n+1)/2$ $(i=1,\cdots,n)$, we have

$$\lambda_{11}=\frac{(n-1)(n-2)}{n(n+1)}\,, \qquad \lambda_{22}=\frac{(n-2)(n^2-2n+13)}{n(n^2-1)}\,, \qquad \lambda_{12}=\frac{-4(n-2)}{n(n+1)}\,.$$

This gives $\rho_{12}=-4/(n^2-2n+13)^{1/2}$ and the upper bound for $|\rho_{12}|$ is given by

$$U=\left\{\rho_{12}^2+\frac{12(n+1)}{(n-1)(n-2)^2(n^2-2n+13)}\right\}^{1/2}.$$

Even for $n$ as small as 5, we have $\rho_{12}=-0.756$ and $U=0.802$.

3. *Quadratic regression* $(m=3)$. Let

$$\boldsymbol{X}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \end{bmatrix}$$

be of rank 3. In an unpublished report, Chew [2] has considered this example to show that the correlations can be large even for moderate values of $n$. For $n=21$ and $x_i=i-11$ $(i=1,\cdots,21)$ his values for $\lambda_{11}$, $\lambda_{22}$ and $\rho_{12}$ are 0.6437, 0.7596 and $-0.4143$ respectively. Equation (7) now gives $|\rho_{12}|\leqq 0.4186$, which is quite close to the true value.

As an application of (7), we consider the problem of detection of a single outlier in linear regression. Now a test statistic for an outlier in either direction is (Joshi [4])

$$B=\max_i |w_i|$$

where $w_i$ is given at (1). It is clear that $B$ will have a maximum for a single $i$, say $i=i_0$. A large value of $B$ then indicates that $y_{i0}$ is an outlier. Let $b_\alpha$ be the upper $\alpha\%$ point of $B$ and $b_\alpha^*$ be an upper limit for $b_\alpha$ obtained by solving the first Bonferroni inequality

$$n\,\Pr(|w_1|>b_\alpha^*)=\alpha\,,$$

that is,

$$I_{1-b_\alpha^{*2}}\left(\frac{p-1}{2}, \frac{1}{2}\right)=\frac{\alpha}{n} ,$$

where

$$I_x(p, q)=\frac{1}{B(p, q)} \int_0^x u^{p-1}(1-u)^{q-1}du .$$

It is known (Srikantan [8], Joshi [4]) that $b_\alpha^*$ coincides with $b_\alpha$ provided that

$$b_\alpha^*>\left\{\frac{1}{2}(1+|\rho_{ij}|)\right\}^{1/2} \qquad (i\neq j) ,$$

where $\rho_{ij}$'s are as in Section 2. Srikantan [8] has used this result for obtaining the true percentage points $b_\alpha^2$ of statistic $B^2$ for some special regression models. Here, we consider the regression models for which all the residuals have equal variance, viz. $(n-r)/n$. In this case

$$B=\left(\frac{n}{n-r}\right)^{1/2} \frac{\max|e_i|}{S_p} ,$$

so that the observation with the largest absolute residual is a possible outlier. Further, equation (7) reduces to

$$|\rho_{ij}|\leqq r/(n-r) \tag{8}$$

and hence $b_\alpha^*$ coincides with $b_\alpha$ provided that

$$b_\alpha^* >\left\{\frac{n}{2(n-r)}\right\}^{1/2} . \tag{9}$$

The 5 and 1% values of $b_\alpha^*$ for $\nu=0$, $r=1(1)4$ and $n\leqq 20$ are tabulated in Table 1. For $r=1$ and 2, $b_\alpha^{*2}$ has been also tabulated by Srikantan [8]. Wilks [9] has tabulated $1-b_\alpha^{*2}$ for $r=1$ in a slightly different context. It is clear that equations (8) and (9) are useful only for $n>2r$. The lower and upper values of $n$ for which the inequality at (9) holds are given in Table 2. Note that (9) does not hold for any $n$ in the case $r=4$ and $\alpha=0.05$. For other cases, although (9) does not hold for some small values of $n$, yet we feal that $b_\alpha^*=b_\alpha$ for these values also.

Table 1 Approximate percentage points $b_\alpha^*$ of statistic $B$ obtained by solving $I_{1-b_\alpha^{*2}}\left(\frac{n-r-1}{2}, \frac{1}{2}\right)=\frac{\alpha}{n}$

| $n/r$ | $\alpha=0.05$ | | | | $\alpha=0.01$ | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 3 | 0.9997 | | | | 1.0000 | | | |
| 4 | 0.9875 | 0.9998 | | | 0.9975 | 1.0000 | | |
| 5 | 0.9587 | 0.9900 | 0.9999 | | 0.9859 | 0.9980 | 1.0000 | |
| 6 | 0.9245 | 0.9635 | 0.9917 | 0.9999 | 0.9665 | 0.9875 | 0.9983 | 1.0000 |
| 7 | 0.8907 | 0.9302 | 0.9671 | 0.9929 | 0.9433 | 0.9690 | 0.9888 | 0.9986 |
| 8 | 0.8593 | 0.8965 | 0.9347 | 0.9699 | 0.9190 | 0.9462 | 0.9710 | 0.9897 |
| 9 | 0.8306 | 0.8649 | 0.9014 | 0.9385 | 0.8951 | 0.9222 | 0.9487 | 0.9727 |
| 10 | 0.8046 | 0.8359 | 0.8697 | 0.9056 | 0.8721 | 0.8983 | 0.9249 | 0.9509 |
| 11 | 0.7810 | 0.8095 | 0.8405 | 0.8740 | 0.8504 | 0.8752 | 0.9011 | 0.9273 |
| 12 | 0.7594 | 0.7855 | 0.8139 | 0.8446 | 0.8300 | 0.8534 | 0.8780 | 0.9036 |
| 13 | 0.7397 | 0.7636 | 0.7896 | 0.8178 | 0.8109 | 0.8329 | 0.8562 | 0.8806 |
| 14 | 0.7217 | 0.7436 | 0.7674 | 0.7933 | 0.7931 | 0.8137 | 0.8355 | 0.8586 |
| 15 | 0.7050 | 0.7252 | 0.7472 | 0.7709 | 0.7763 | 0.7956 | 0.8162 | 0.8379 |
| 16 | 0.6895 | 0.7083 | 0.7285 | 0.7504 | 0.7606 | 0.7787 | 0.7980 | 0.8185 |
| 17 | 0.6751 | 0.6926 | 0.7114 | 0.7316 | 0.7458 | 0.7628 | 0.7810 | 0.8002 |
| 18 | 0.6618 | 0.6780 | 0.6955 | 0.7142 | 0.7319 | 0.7479 | 0.7650 | 0.7831 |
| 19 | 0.6492 | 0.6644 | 0.6807 | 0.6982 | 0.7187 | 0.7339 | 0.7500 | 0.7670 |
| 20 | 0.6375 | 0.6517 | 0.6670 | 0.6832 | 0.7063 | 0.7207 | 0.7358 | 0.7519 |

Table 2 Lower and upper values of $n$ for which $b_\alpha^*$ is the true percentage point

| $r$ | $\alpha=0.05$ | | $\alpha=0.01$ | |
|---|---|---|---|---|
| | lower | upper | lower | upper |
| 1 | 3 | 13 | 3 | 18 |
| 2 | 5 | 12 | 5 | 17 |
| 3 | 7 | 11 | 7 | 17 |
| 4 | | | 9 | 16 |

The author wishes to thank Professor H. A. David in the preparation of this paper. Part of this work was done while the author was at the University of North Carolina at Chapel Hill and was supported by the U.S. Army Research Office, Durham, Grant No. DA-ARO(D)-31-124-G-746.

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

## REFERENCES

[1] Anscombe, F. J. (1960). Rejection of outliers, *Technometrics*, **2**, 123-147.
[2] Chew, V. (1964). Tests for the rejection of outlying observations, *RCA Systems Analysis Tech. Memo.*, No. 64-7.
[3] Doornbos, R. (1966). *Slippage Tests*, Mathematisch Centrum, Amsterdam.
[4] Joshi, P. C. (1972). Some slippage tests of mean for a single outlier in linear regression, *Biometrika*, **59**, 109-120.
[5] McFadden, J. A. (1966). A diagonal expansion in Gegenbauer polynomials for a class of second-order probability densities, *SIAM J. Appl. Math.*, **14**, 1433-1436.
[6] Quesenberry, C. P. and David, H. A. (1961). Some tests for outliers, *Biometrika*, **48**, 379-390.
[7] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, Wiley, New York.
[8] Srikantan, K. S. (1961). Testing for the single outlier in a regression model, *Sankhyā*, A, **23**, 251-260.
[9] Wilks, S. S. (1963). Multivariate statistical outliers, *Sankhyā*, A, **25**, 407-426.