Approximate Tests of Hypotheses in Regression
Models with Grouped Data

Abhaya Indrayan
BRD Medical College
University of Gorakhpur
Gorakhpur, India

and

Jagdish S. Rustagi*
Department of Statistics
The Ohio State University
Columbus, Ohio
Technical Report No. 172

* Supported in part by Contract N00014-78-C-0543 (NR 042-403) by Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Approximate tests of hypotheses in regression models with grouped data.

Techniques for testing hypotheses about parameters in the regression models under the situation of grouped data are provided. A test statistic similar to conventional $F$ statistic is considered. A simulation study performed for a few cases shows that the proposed statistic has an approximate $F$ distribution and is useful in applications.
1. Introduction

In many real situations, observations are available only in certain groups. This process of recording or storing observations in groups directly leads to grouped data. Experimental observations where precise measuring instruments are not available also result in grouped data. Various authors have provided examples of diverse fields where statistical analysis has to depend on grouped data, for reference, see Indrayan and Rustagi (1979). In that paper, the case of approximate maximum likelihood estimates was discussed for regression models. In this paper, we provide techniques for testing hypotheses about parameters in the regression model under the situation of grouped data. We consider a test statistic which is similar to the conventional $F$ statistic for the ungrouped case. A simulation study, performed for a few cases, shows that the proposed statistic has an approximate $F$-distribution.

The approximation here is to the order $O\left(\frac{h}{\sigma^4}\right)$ where $h$ is the length of the interval of recorded observations and $\sigma$ is the standard deviation. In a way, this simulation study confirms the robustness of the $F$-statistic for the regression models in the grouped case and is likely to be very useful in applications.
2. Model and Notation

The notation used here is the same as in Indrayan and Rustagi (1979). It is assumed that there are \( K \) populations, generated by random variables \( y_1, y_2, \ldots, y_K \). Further let \( y_k \) be normally distributed with mean \( \mu_k \) and variance \( \sigma^2 \), where

\[
\mu_k = \beta_1 x_{k1} + \beta_2 x_{k2} + \ldots + \beta_p x_{kp}, \quad k = 1, 2, \ldots, K,
\]

\( \beta_i \)'s are the unknown parameters and \( x \)'s are known constants. It is assumed that there are \( n_k \) independent observations in \( y_k \), denoted as \( y_{uk} \), \( u = 1, 2, \ldots, n_k \), with \( \Sigma n_k = n \). The matrix of observations denoted by

\[
X = (x_{ij})
\]

\( i = 1, 2, \ldots, n \),
\( j = 1, 2, \ldots, p \),

is known.

Suppose the possible values of the random variables \( y_i \) are recorded in intervals \( [C_{i-1}, C_i) \), \( i = \ldots, -2, -1, 0, 1, 2, \ldots \) with \( C_i - C_{i-1} = h \). Let \( N_{ik} \) be the number of observations on \( y_k \) in the interval \( [C_{i-1}, C_i) \) and let this probability be \( \Pi_{ik} \).

Let \( Z \) be a matrix of order \( m \times p \) of known constants and let

\[
\beta' = (\beta_1, \ldots, \beta_p). \quad \text{We are interested in the test of the hypothesis}
\]

\[
H_0: \quad \beta' = \gamma
\]

versus the alternative
$H_1: \mathbf{u} \neq \mathbf{0}$

for some given constant vector $\mathbf{u}$.

In the usual ungrouped case, where $K$ populations are tested for the above hypothesis, the analysis of variance test utilizing the $F$-distribution is generally available in most books, see for example, Rao (1973).

For the grouped data case, the likelihood of the sample $L(\mathbf{u})$ is obtained in terms of multinomials.

Let the maximum likelihood estimates of $\mathbf{u}_{1k}$ under the null hypothesis be denoted by $\mathbf{u}_{1k}(\hat{\omega})$ and under the full model by $\mathbf{u}_{1k}(\hat{\Omega})$. Maximum likelihood estimates were discussed in an earlier paper by the authors (1979). The likelihood ratio given by

$$\lambda = \frac{L(\mathbf{u}_{1k}(\hat{\omega}))}{L(\mathbf{u}_{1k}(\hat{\Omega}))} \quad (2.3)$$

is used to provide tests for the hypotheses (2.2). The asymptotic distribution of $-2 \log \lambda$ follows chi-squared distribution as in the ungrouped case under the approximation ignoring terms of $O\left(\frac{h_i^4}{\sigma^4}\right)$.
3. F-statistic

In the usual ungrouped case, the test of the general linear hypothesis is obtained in terms of an F test. For the sake of completeness, we state below the Fundamental Lemma of Analysis of Variance, Rao (1973).

Lemma 3.1. For the Gauss-Markov model \( (Y, X\beta, \sigma^2 I) \), the test of hypothesis \( H_0 : \beta = \beta_0 \), where \( \beta_0 \) is a given \( m \times p \) matrix of rank \( m \), is given by the statistic

\[
F = \frac{R_1^2 - R_0^2}{\frac{R_0^2}{n - r}}
\]

where \( r = \text{rank of matrix } X \) with

\[
R_0^2 = \min_{C} (Y - X\beta)'(Y - X\beta)
\]

and

\[
R_1^2 = \min_{C} (Y - X\beta)'(Y - X\beta)
\]

The statistic (3.1) has an F-distribution with \( p \) and \( n-r \) degrees of freedom.

The test of hypothesis in the grouped-data case can be similarly obtained in terms of a statistic which is the ratio of sums of squares. Let the midpoint variable \( M \) be defined by the following:

\[
M = m_i \text{ if and only if } Y \in [C_{i-1}, C_i)
\]

\[
i = -2, -1, 0, 1, 2, \ldots
\]

The approximate maximum likelihood estimate of \( \beta \) is given by

\[
\hat{\beta}_0 = (X'X)^{-1}X'M
\]

(3.2)
where $A^-$ denotes the generalized inverse (g-inverse) of the matrix $A$, for reference, see Rao and Mitra (1971) and $\mu_x$ is the vector of mid points resulting from the data. The mid point estimator $\hat{\mu}_0$ leads to a statistic

$$F_0 = \frac{R_{1M}^2 - R_{OM}^2}{R_{OM}^2} \cdot \frac{R_{OM}^2}{n - r}$$

(3.3)

where $R_{OM}^2 = \min_{\beta} (\mu_x - X\beta)'(\mu_x - X\beta)$, and

$$R_{1M}^2 = \min_{C\beta = \alpha} (\mu_x - X\beta)'(\mu_x - X\beta).$$

As usual, we evaluate the sums of squares in the statistic (3.3) by using the following notation.

Suppose $\begin{pmatrix} \tilde{X} \\ X' \end{pmatrix} = \begin{pmatrix} T_a \\ T_c \\ X' \end{pmatrix}$,

then

$$R_{OM}^2 = M'T_aM$$

(3.4)

$$R_{1M}^2 - R_{OM}^2 = (C'M-\mu)'(C'M-\mu)^{-1}$$

(3.5)

To ensure estimability of $C\beta$ we assume that $C = A'X$ for some $A$, with rank of $C = m$.

Without loss of generality, the statistic $F_0$ can be considered for the case of testing the hypothesis that $\beta = 0$. In that case we have

$$R_{OM}^2 = \min_{\beta} (\mu_x - X\beta)'(\mu_x - X\beta)$$

$$= (\mu_x - X\hat{\beta}_0)'(\mu_x - X\hat{\beta}_0)$$

$$= M'M - \hat{\beta}_0X'M.$$
and $R_{1M}^2 = M'M$ so that $R_{1M}^2 - R_{OM}^2 = \mathcal{Q}_0 X'M$. Usually $R_{OM}^2$ is called the Error Sum of Squares (Error SS) and $R_{1M}^2 - R_{OM}^2$ as the Regression Sums of Squares (Regression SS).

To show that the asymptotic distribution of

$$F_0 = \frac{(\text{Regression SS})^2}{m} \div \frac{(\text{Error SS})^2}{n-m}$$  \hspace{1cm} (3.6)

is chi-squared to the order of approximation implied by Sheppard's corrections, Cramér (1974, pp. 359-362) which is assumed to be negligible here, we have the following lemmas.

Lemma 3.2. The asymptotic distribution of

$$\hat{\beta}_0 = (X'X)^{-1} X'M$$  \hspace{1cm} (3.7)

as $n_k \to \infty$ is p-variate normal with mean $\beta_0$ and variance covariance

$$(X'X)^{-1}(\sigma^2 + \frac{h^2}{12})$$

to the order to the order of approximation implied by Sheppard's correction.

Proof: Let $M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_k \end{pmatrix}$

where $M_k$ is $n_k$-vector, $k = 1, 2, \ldots, K$. Suppose

$$(X'X)^{-2} X' = P = (P_1, \ldots, P_k)$$

with $P_i$ being a $p \times n_i$ matrix, $i = 1, 2, \ldots, K$. With the above definitions,

$$\hat{\beta}_0 = \sum_{i=1}^{K} P_i M_i$$  \hspace{1cm} (3.8)

Let the elements of the partitions $P_i$ of the matrix $P$ be denoted by

$P_{i,m}(l)$, $l = 1, 2, \ldots, p$, $m = 1, 2, \ldots, n_i$. Notice that the first $n_1$ columns of $X'$ are identical, the next $n_2$ columns are also identical and so on.

Therefore, the columns of the matrix $P_i$ are all identical for $i = 1, 2, \ldots, K$. 
for each \( i \) reduces to a constant multiple of the sum of \( n_i \) elements of the vector \( \mathbf{M}_i \). Since the components of this sum are independent and identically distributed, by the application of Central Limit Theorem, the vector \( \mathbf{M}_i \) is asymptotically \( p \)-variate normal for each \( i \).

The distribution of \( \hat{\mathbf{b}}_0 \) is consequently also \( p \)-variate normal.

Since \( E(\mathbf{M}_i) = E(\mathbf{X}_i) \) and variance of the components of \( \mathbf{M}_i \) is \( \sigma^2 + \frac{h^2}{12} \) (subject to the approximation implied by Shappard's correction), it follows that

\[
E(\hat{\mathbf{b}}_0) = \mathbf{b}_0 \quad (3.9)
\]

and

\[
\text{Cov}(\hat{\mathbf{b}}_0) = (\mathbf{X}'\mathbf{X})^{-1} (\sigma^2 + \frac{h^2}{12}). \quad (3.10)
\]

The proof of lemma is now complete.

**Lemma 3.3.** To the order of approximation implied by the Sheppard correction, the mean and variance of error sums of squares are given by

\[
E(\text{Error SS}) = (n-p)\sigma_0^2, \quad (3.11)
\]

\[
\text{V}(\text{Error SS}) = 2(n-m)\sigma_0^4 + \frac{h^4}{80\sigma^2} + \frac{h^2}{2\sigma^2} \sigma_0^4 \sum b_{11}^2, \quad (3.12)
\]

where

\[
\sigma_0^2 = \sigma^2 + \frac{h^2}{12}, \quad (3.13)
\]

and

\[
\mathbf{b} = (b_{ij}) = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'. \quad (3.14)
\]
Proof. Applying results of theorem 1 of Seare (1971, p. 55) and (3.10'), we have

\[ E(\text{Error SS}) = n\sigma_0^2 + n\mathbf{x}'\mathbf{X} \mathbf{X}'\mathbf{x} - (n\sigma_0^2 + n\mathbf{x}'\mathbf{X} \mathbf{X}'\mathbf{x}) = (n-m)\sigma_0^2 \]

In a paper by Hsu (1938), it has been shown that a quadratic form

\[ Q = \sum_{i,j} a_{ij} Z_i Z_j \]

with \( E(Q) = 1 \) and \( \text{Var}(Z_i) = \sigma^2 \) for all \( i \), has the following properties:

\[ \text{Var} Q = E(\mu_{41}-3) a_{11}^2 + 2a_{1j}^2 \]

where \( E_{ij} = E(Z_i - E(Z_i))^4 \).

Suppose now,

\[ Q = \frac{\text{Error SS}}{(n-p)\sigma_0^2} \]

so that \( E(Q) = 1 \). Also the matrix \( B \) is symmetric and idempotent with rank \( n-m \) and hence \( \sum_{ij} b_{ij}^2 = n-m \). By assumptions of normality and using Sheppard's corrections, we have

\[ E(M_i - E(M_i))^4 = 3\bar{c}^4 + \frac{h^2}{2}\sigma^2 + \frac{h}{2} \]

for all \( i \). The result (3.12) follows.

From lemma 3.2, we know that the distribution of \((\text{Regression SS})/\sigma_0^2\) as \( n_k \to \infty \), has a non-central chi-squared distribution with \( p \) degrees of freedom with noncentrality parameter

\[ \delta^2 = \frac{1}{2} \left( \mathbf{X}'\mathbf{X} \right) \sigma_0^2 \]

(3.15)

to the order of approximation implied by Sheppard's corrections. Further
since $\sum b_{ii}^2 \leq n-m$, it follows from (3.11) and (3.12) that

\[
\frac{\text{Error SS}}{(n-m)c_0^2} \text{ tends to } 1 \text{ in probability. Hence } mF_0 \text{ as given in (3.3) has a noncentral chi-squared distribution with noncentrality parameter } \delta.
\]

Note that under the null hypothesis, $\delta = 0$, hence the asymptotic distribution of $F_0$ is central chi-squared. Therefore the test can be easily performed.

For small $h$, the distribution of $F_0$ may turn out to be close to the $F$-distribution. Box and Anderson (1955) have developed robust tests for non-normal populations using the following. Assume that the distribution of $(\text{Error SS})/(n-m)c_0^2$ is $\chi^2/\nu$ where degrees of freedom $\nu$ are obtained by a method of moments given by

\[
\nu = n-m - \frac{h^2}{4c_0^2} \sum b_{ii}^2 + O\left(\frac{h^4}{\sigma^4}\right)
\]

(3.16)

For small $h/\sigma$, we have hardly any correction to degrees of freedom and then the test can be performed as an $F$ test. This behavior of the statistic $F_0$ has been studied through simulations and goodness of the approximation is measured in terms of Kolmogoroff-Smirnov statistic in the next section.
4. Simulations

Two models have been considered for simulations.

Model I. 
\[ Y_{uk} \sim N(\alpha + \beta x_k, \sigma^2) \]
\[ u = 1, 2, \ldots, n_k; \quad k = 1, 2, \ldots, K. \]

We consider \( K = 10 \), with \( X_1 = 0, X_2 = 1, \ldots, X_{10} = 10, n_K = 100 \) (same sample sizes for every \( k \)). We consider two cases of \( \alpha, \beta \),

\( \quad \alpha = 10, \quad \beta = 0 \)

\( \quad \alpha = 40 \) and \( \beta = 0. \)

The values of \( \sigma^2 \) are chosen to be

25 or 100.

The size of \( h \) is taken as

0, 2, 3, 4, 5, 10, 15, 20.

The hypothesis considered here is

\[ H_0: \quad \beta = 0 \]

Using the usual IBM Random Number Generator package, samples were generated and then were grouped in intervals of size \( h \). The \( F \) statistic was calculated for the ungrouped data case and \( F_0 \) statistic for the various grouped data cases. The empirical cumulative distribution of the statistics were then computed and were compared with the theoretical \( F \)-distribution using Kolmogorov-Smirnov distribution. Table I and II describe the results of the simulations for
\( n_K = 10 \) and \( n_K = 100 \) respectively. The last column gives the tail probability for significance in both cases. The column with heading \( D \) gives the actual value of Kolmogorov-Smirnov statistic.
Table I

<table>
<thead>
<tr>
<th>h</th>
<th>ν₁</th>
<th>ν₂</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>98</td>
<td>0.6120</td>
<td>0.8481</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>98</td>
<td>0.6311</td>
<td>0.8206</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>98</td>
<td>0.6988</td>
<td>0.8658</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>98</td>
<td>0.4953</td>
<td>0.6669</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>98</td>
<td>0.8531</td>
<td>0.4606</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>98</td>
<td>1.1812</td>
<td>0.1227</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>98</td>
<td>0.8534</td>
<td>0.4602</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>98</td>
<td>1.1241</td>
<td>0.1597</td>
</tr>
</tbody>
</table>

\( \sigma^2 = 25, \alpha = 10 \)

<table>
<thead>
<tr>
<th>h</th>
<th>ν₁</th>
<th>ν₂</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>98</td>
<td>0.9169</td>
<td>0.3699</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>98</td>
<td>0.9339</td>
<td>0.3477</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>98</td>
<td>0.8668</td>
<td>0.4401</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>98</td>
<td>0.6613</td>
<td>0.7743</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>98</td>
<td>0.7018</td>
<td>0.7082</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>98</td>
<td>0.9671</td>
<td>0.3070</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>98</td>
<td>0.8695</td>
<td>0.4362</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>98</td>
<td>1.2566</td>
<td>0.0850</td>
</tr>
</tbody>
</table>

\( \sigma^2 = 25, \alpha = 40 \)

<table>
<thead>
<tr>
<th>h</th>
<th>ν₁</th>
<th>ν₂</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>98</td>
<td>0.6721</td>
<td>0.7570</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>98</td>
<td>0.8685</td>
<td>0.4376</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>98</td>
<td>0.7632</td>
<td>0.6050</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>98</td>
<td>0.5060</td>
<td>0.9600</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>98</td>
<td>0.5713</td>
<td>0.8999</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>98</td>
<td>0.6226</td>
<td>0.8330</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>98</td>
<td>0.7894</td>
<td>0.5614</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>98</td>
<td>0.8204</td>
<td>0.5114</td>
</tr>
</tbody>
</table>

\( \sigma^2 = 100, \alpha = 10 \)

<table>
<thead>
<tr>
<th>h</th>
<th>ν₁</th>
<th>ν₂</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>98</td>
<td>0.7245</td>
<td>0.6701</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>98</td>
<td>0.7987</td>
<td>0.5463</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>98</td>
<td>0.8099</td>
<td>0.5281</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>98</td>
<td>0.8101</td>
<td>0.5278</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>98</td>
<td>0.7078</td>
<td>0.6982</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>98</td>
<td>0.8516</td>
<td>0.4629</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>98</td>
<td>0.9647</td>
<td>0.3098</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>98</td>
<td>0.6869</td>
<td>0.7330</td>
</tr>
</tbody>
</table>

\( \sigma^2 = 100, \alpha = 40 \)
<table>
<thead>
<tr>
<th>$h$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>998</td>
<td>0.8214</td>
<td>0.5097</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>998</td>
<td>0.9069</td>
<td>0.3832</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>998</td>
<td>0.8081</td>
<td>0.5310</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>998</td>
<td>0.7360</td>
<td>0.6507</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>998</td>
<td>0.8411</td>
<td>0.4790</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>998</td>
<td>0.4508</td>
<td>0.9871</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>998</td>
<td>0.9026</td>
<td>0.3891</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>998</td>
<td>0.7345</td>
<td>0.6532</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma^2=25, \alpha=10$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>998</td>
<td>0.5086</td>
<td>0.9562</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>998</td>
<td>0.7267</td>
<td>0.6664</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>998</td>
<td>0.6666</td>
<td>0.7659</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>998</td>
<td>0.5544</td>
<td>0.9184</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>998</td>
<td>0.6424</td>
<td>0.8037</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>998</td>
<td>0.5871</td>
<td>0.8809</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>998</td>
<td>0.7458</td>
<td>0.6342</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>998</td>
<td>1.1031</td>
<td>0.1753</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma^2=25, \alpha=40$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>998</td>
<td>0.7605</td>
<td>0.6096</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>998</td>
<td>0.7578</td>
<td>0.6140</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>998</td>
<td>0.7627</td>
<td>0.6058</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>998</td>
<td>0.9348</td>
<td>0.3466</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>998</td>
<td>0.8685</td>
<td>0.4376</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>998</td>
<td>0.6699</td>
<td>0.7606</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>998</td>
<td>0.6582</td>
<td>0.7792</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>998</td>
<td>0.9366</td>
<td>0.3442</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma^2=100, \alpha=10$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>998</td>
<td>0.7652</td>
<td>0.6017</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>998</td>
<td>0.7107</td>
<td>0.6934</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>998</td>
<td>0.7613</td>
<td>0.6082</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>998</td>
<td>0.7483</td>
<td>0.6300</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>998</td>
<td>0.5413</td>
<td>0.9313</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>998</td>
<td>0.5698</td>
<td>0.9015</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>998</td>
<td>0.8149</td>
<td>0.5201</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>999</td>
<td>0.5546</td>
<td>0.9182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma^2=100, \alpha=40$</td>
<td></td>
</tr>
</tbody>
</table>
Model II. We consider here the regression model

\[ Y_{uk} \sim N(\beta_1 X_{k1} + \beta_2 X_{k2} + \beta_3 X_{k3}, \sigma^2) \]

\[ k = 1, 2, \ldots, 4. \]

Four sets of \( X_{ku} \) are utilized

(i) \( X_{11} = 0 \quad X_{12} = 1 \quad X_{13} = 3 \)
(ii) \( X_{21} = 2 \quad X_{22} = 4 \quad X_{23} = 6 \)
(iii) \( X_{31} = 7 \quad X_{32} = 5 \quad X_{33} = 3 \)
(iv) \( X_{41} = 9 \quad X_{42} = 1 \quad X_{43} = 8 \)

\( h = 0, 10, 50, 100, 200 \)
\( \sigma = 100 \)
\( n_k = 10, 25, 100 \) (same for all \( k \))
\( \beta_1 = \beta_2 = \beta_3 = 0. \)

The hypothesis tested here is

\[ H_0: \beta_2 = \beta_3 = 0 \]

The results are given in Table III and are based on 100 samples.
Table III

<table>
<thead>
<tr>
<th>h</th>
<th>ν₁</th>
<th>ν₂</th>
<th>D</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nₖ = 10 for all k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>37</td>
<td>0.7567</td>
<td>0.6158</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>37</td>
<td>0.6984</td>
<td>0.7138</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>37</td>
<td>0.7600</td>
<td>0.6104</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>37</td>
<td>0.7239</td>
<td>0.6712</td>
</tr>
<tr>
<td>200</td>
<td>2</td>
<td>37</td>
<td>1.2845</td>
<td>0.0738</td>
</tr>
<tr>
<td>nₖ = 25 for all k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>97</td>
<td>0.8039</td>
<td>0.5378</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>97</td>
<td>0.8742</td>
<td>0.4294</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>97</td>
<td>0.8401</td>
<td>0.4805</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>97</td>
<td>0.7487</td>
<td>0.6293</td>
</tr>
<tr>
<td>200</td>
<td>2</td>
<td>97</td>
<td>0.9388</td>
<td>0.3415</td>
</tr>
<tr>
<td>nₖ = 100 for all k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>397</td>
<td>0.7705</td>
<td>0.5927</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>397</td>
<td>0.7468</td>
<td>0.6325</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>397</td>
<td>0.8749</td>
<td>0.4284</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>397</td>
<td>1.1426</td>
<td>0.1469</td>
</tr>
<tr>
<td>200</td>
<td>2</td>
<td>397</td>
<td>1.1312</td>
<td>0.1546</td>
</tr>
</tbody>
</table>
5. Discussion

Comparison of the tail probability of the Kolmogorov-Smirnov statistic for \( h = 0 \) (that is, ungrouped case) with various values of \( h \), shows that on the whole, the empirical cumulative distributions are the same for the cases \( h/\theta \leq 1 \). When \( h/\theta > 1 \), we do not have very good results. In general, one could make the statement that the \( F \)-distribution is fairly a good approximation to the distribution of the statistic \( F_0 \). More extensive simulations may be able to provide further evidence of this correspondence.
References


<table>
<thead>
<tr>
<th>Copies</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistics and Probability Program (Code 436)</td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>Arlington, VA 22217</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Defense Documentation Center</td>
</tr>
<tr>
<td></td>
<td>Cameron Station</td>
</tr>
<tr>
<td></td>
<td>Alexandria, VA 22314</td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>New York Area Office</td>
</tr>
<tr>
<td></td>
<td>715 Broadway - 5th Floor</td>
</tr>
<tr>
<td></td>
<td>New York, New York 10003</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commanding Officer</td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>Branch Office</td>
</tr>
<tr>
<td></td>
<td>Attn: D. A.L. Powell</td>
</tr>
<tr>
<td></td>
<td>Building 114, Section D</td>
</tr>
<tr>
<td></td>
<td>666 Summer Street</td>
</tr>
<tr>
<td></td>
<td>Boston, MA 02210</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commanding Officer</td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>Branch Office</td>
</tr>
<tr>
<td></td>
<td>Attn: Director for Science</td>
</tr>
<tr>
<td></td>
<td>536 South Clark Street</td>
</tr>
<tr>
<td></td>
<td>Chicago, Illinois 60605</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commanding Officer</td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>Branch Office</td>
</tr>
<tr>
<td></td>
<td>Attn: Dr. Richard Lau</td>
</tr>
<tr>
<td></td>
<td>1030 East Green Street</td>
</tr>
<tr>
<td></td>
<td>Pasadena, CA 91107</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>San Francisco Area Office</td>
</tr>
<tr>
<td></td>
<td>One Hallidie Plaza - Suite 601</td>
</tr>
<tr>
<td></td>
<td>San Francisco, CA 94102</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td></td>
<td>Scientific Liaison Group</td>
</tr>
<tr>
<td></td>
<td>Attn: Scientific Director</td>
</tr>
<tr>
<td></td>
<td>American Embassy - Tokyo</td>
</tr>
<tr>
<td></td>
<td>APO San Francisco 96503</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Applied Mathematics Laboratory</td>
</tr>
<tr>
<td></td>
<td>David Taylor Naval Ship Research and Development Center</td>
</tr>
<tr>
<td></td>
<td>Attn: Mr. G. H. Gleissner</td>
</tr>
<tr>
<td></td>
<td>Bethesda, Maryland 20084</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commandant of the Marine Corps</td>
</tr>
<tr>
<td></td>
<td>(Code AX)</td>
</tr>
<tr>
<td></td>
<td>Attn: Dr. A.L. Slafkosky</td>
</tr>
<tr>
<td></td>
<td>Scientific Advisor</td>
</tr>
<tr>
<td></td>
<td>Washington, DC 20380</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Director</td>
</tr>
<tr>
<td></td>
<td>National Security Agency</td>
</tr>
<tr>
<td></td>
<td>Attn: Mr. Stahlly and Dr. Maar (R51)</td>
</tr>
<tr>
<td></td>
<td>Fort Meade, MD 20755</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Navy Library</td>
</tr>
<tr>
<td></td>
<td>National Space Technology Laboratory</td>
</tr>
<tr>
<td></td>
<td>Attn: Navy Librarian</td>
</tr>
<tr>
<td></td>
<td>Bay St. Louis, MS 39522</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Copies</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>U.S. Army Research Office</td>
<td>1</td>
</tr>
<tr>
<td>P.O. Box 12211</td>
<td></td>
</tr>
<tr>
<td>Attn: Dr. J. Chandra</td>
<td></td>
</tr>
<tr>
<td>Research Triangle Park, NC 27706</td>
<td></td>
</tr>
<tr>
<td>Naval Sea Systems Command</td>
<td>1</td>
</tr>
<tr>
<td>(NSEA 03F)</td>
<td></td>
</tr>
<tr>
<td>Attn: Miss B. S. Orleans</td>
<td></td>
</tr>
<tr>
<td>Crystal Plaza #6</td>
<td></td>
</tr>
<tr>
<td>Arlington, VA 20360</td>
<td></td>
</tr>
<tr>
<td>Office of the Director</td>
<td>1</td>
</tr>
<tr>
<td>Bureau of The Census</td>
<td></td>
</tr>
<tr>
<td>Attn: Mr. H. Nisselson</td>
<td></td>
</tr>
<tr>
<td>Federal Building 3</td>
<td></td>
</tr>
<tr>
<td>Washington, DC 20233</td>
<td></td>
</tr>
<tr>
<td>OASD (I&amp;L), Pentagon</td>
<td>1</td>
</tr>
<tr>
<td>Attn: Mr. Charles S. Smith</td>
<td></td>
</tr>
<tr>
<td>Washington, DC 20301</td>
<td></td>
</tr>
<tr>
<td>ARI Field Unit-USAREUR</td>
<td>1</td>
</tr>
<tr>
<td>Attn: Library</td>
<td></td>
</tr>
<tr>
<td>:/o ODCSPER</td>
<td></td>
</tr>
<tr>
<td>HQ USAREUR &amp; 7th Army</td>
<td></td>
</tr>
<tr>
<td>APO New York 09403</td>
<td></td>
</tr>
<tr>
<td>Naval Underwater Systems Center</td>
<td>1</td>
</tr>
<tr>
<td>Attn: Dr. Derrill J. Bordelon</td>
<td></td>
</tr>
<tr>
<td>Code 21</td>
<td></td>
</tr>
<tr>
<td>Newport, Rhode Island 02840</td>
<td></td>
</tr>
<tr>
<td>Library, Code 1424</td>
<td>1</td>
</tr>
<tr>
<td>Naval Postgraduate School</td>
<td></td>
</tr>
<tr>
<td>Monterey, California 93940</td>
<td></td>
</tr>
<tr>
<td>Technical Information Division</td>
<td>1</td>
</tr>
<tr>
<td>Naval Research Laboratory</td>
<td></td>
</tr>
<tr>
<td>Washington, DC 20375</td>
<td></td>
</tr>
</tbody>
</table>