The Maximum Likelihood Estimate of the Fraction Defective Under Curtailed Multiple Sampling Plans

D. K. Shah and A. G. Phatak

Department of Statistics
Baroda University
Baroda, India

In this paper we have obtained the maximum likelihood estimate of the fraction defective under a fully-curtailed double sampling plan. The asymptotic variance of the maximum likelihood estimate is also obtained. The results are generalized to multiple sampling plan under full-curtailment. A simple device for executing a curtailed multiple sampling plan and the calculation of the ASN, the probability of acceptance therefrom is discussed.

Key Words
Acceptance Sampling Plan by Attributes
ASN
Asymptotic Variance of the MLE
Curtailed Multiple Sampling Plans
Double Sampling Plan
Easy execution of Sampling Plans
Estimation of Fraction Defective
Maximum Likelihood Estimate of the Fraction Defective
Multiple Sampling Plan

1. Introduction

In the multiple sampling plan (MSP) by attributes a sequence of k samples of size n, (i = 1, 2, . . . , k) is taken from a lot of size N. The design of the sampling plan specifies 2k numbers a and r, (i = 1, 2, . . . , k). If the accumulated number of defectives, d, is equal to or less than a, , the lot is accepted. If d, is equal to or greater than r, , the lot is rejected. If d, falls between a, and r, , the decision of acceptance or rejection of the lot is deferred until the next sample of size n,+1 is inspected. The constants a, and r, , known as acceptance and rejection numbers, are predetermined numbers satisfying the conditions

(i) 0 ≤ a1 ≤ a2 ≤ · · · ≤ a2k-1 ≤ a2k ,
(ii) r1 ≤ r2 ≤ · · · ≤ rk ,
(iii) a, + 1 < r, , i = 1, 2, . . . , k - 1,
(iv) a, + 1 = r, ,
(v) a, ≤ ∑i=1 k ni ,
(vi) r, ≤ ∑i=1 k ni .

The condition (iv) ensures that not more than k samples are required to inspect.

If the inspection has no other purpose than to determine which inspection lots to accept and which to reject, it would be obvious to stop the inspection as soon as the decision of acceptance or rejection is reached. This leads to curtailment in the inspection. Two forms of curtailed inspection can be distinguished. The plan which considers curtailment in the inspection arising due to observing enough defectives to reject a lot is termed here as a semi-curtailed plan, following the terminology of the Statistical Research Group [7]. Similarly the plan which considers curtailment in the inspection arising due to observing either enough defectives to reject a lot or enough nondefectives to accept a lot is called a fully-curtailed plan.

In this paper we have defined a fully-curtailed MSP. A double sampling plan (DSP) is a particular case of a MSP for k = 2. We have obtained the maximum likelihood estimate (mle) of the fraction defective and the asymptotic variance of the mle for a fully-curtailed DSP. Generalization of these results for a fully-curtailed MSP is also discussed. Furthermore, we have obtained an expression for the ASN for a fully-curtailed DSP. A simple device for executing MSP and the calculation of the ASN and the probability of acceptance therefrom are explained with a numerical illustration.

2. Statement of a Fully-Curtailed MSP

We give below the statement of a fully-curtailed MSP.

The concern here is with an attributes acceptance plan in which individual units randomly selected from a lot of size N are inspected one at a time till one of the following 2k events occurs: (a,) r, defectives are observed and the number of units inspected is greater than ∑i=1 k ni and less than or equal to ∑i=1 k ni , (b,) g, nondefectives are observed and the number of units inspected is greater than ∑i=1 k ni.
and less than or equal to \( \sum_{i=1}^{k} n_i \) for \( i = 1, 2, \cdots, k \) where \( n_0 \) is assigned a value zero.

The decision rule is then to reject the lot if one of the \( k \) events of the set \( \{ \alpha_i \} \) occurs and to accept the lot if one of the \( k \) events of the set \( \{ \beta_i \} \) occurs.

The relation between the constants given in the above statement and those of the corresponding uncurtailed MSP is

\[
n_i = n_i, \quad r_i = r_i, \quad g_i = \sum_{j=1}^{i} n_j - a_i.
\]

Statement of a fully-curtailed DSP can be easily obtained from the above by considering \( k = 2 \).
The events \( \alpha_1, \alpha_2, \beta_1 \), \( \beta_2 \) of this case are designated as \( E_i \) \((i = 1, 2, 3, 4)\) for convenience in the following sections.

3. Estimation of the Fraction Defective in Case of a Fully-Curtailed DSP

Let the process average proportion of defectives be \( p \) and for sufficiently large lots it can be considered as the probability of selecting a defective in a single trial. Furthermore, let the probability \( p \) remain constant from trial to trial and the trials be stochastically independent.

Let \( Y \) denote the number of units inspected when the inspection is stopped due to the occurrence of the event \( E_i \) \((i = 1, 2, 3, 4)\). Let \( A_i \) \((i = 1, 2, 3, 4)\) be the set of possible values attained by \( Y \). Then

\[
A_1 = \{ r_1, r_1 + 1, \cdots, n_1 \},
A_2 = \{ r_2 - r_1 + n_1 + 1, \cdots, r_2, r_2 + n_1 + 2, \cdots, n_1 + n_2 \},
A_3 = \{ g_1, g_1 + 1, \cdots, n_1 \},
A_4 = \{ g_2 - g_1 + n_1 + 1, g_2 - g_1 + n_1 + 2, \cdots, n_1 + n_2 \}.
\]

Lastly define a random variable \( T \) as follows

\[
T = i \text{ if } E_i \text{ occurs, } i = 1, 2, 3, 4.
\]

Then the joint probability function of the random variables \( Y \) and \( T \) can be expressed as

\[
P(Y = y, T = i) = f_i(y; p) \quad y \in A_i, \quad i = 1, 2, 3, 4 \quad \text{(3.1)}
\]

where

\[
f_i(y; p) = \begin{cases} \left(\frac{y - 1}{y_i - 1}\right) p^{y_i} q^{y - y_i}, & y \in A_i, \quad i = 1, 2, 3, 4 \quad \text{(3.2)} \\ 0 & \text{elsewhere} \end{cases}
\]

and

\[
f_i(y; p) = \sum_{u=1}^{n_i} \left(\frac{n_i}{y_i - u}\right) \left(\frac{b_2 - u}{b_2 - b_1}\right) p^{y_i} q^{y - y_i}, \quad \text{(3.3)}
\]

where \( f_i(y; p) \) is the probability function of observing \( y \) out of \( n_i \) trials with a probability of success \( p \).
NOTE

\[ \frac{\text{Total number of defectives observed}}{\text{Total number of units inspected}}. \quad (3.10) \]

4. THE ASYMPTOTIC VARIANCE OF THE MLE

Differentiating the logarithm of the likelihood function, \( L \), given by (3.8) twice it is found that

\[ -\frac{\partial^2 \log L}{\partial p^2} = \frac{q - p}{p^2 q^2}, \]

where

\[ q = \frac{\text{Total number of defectives observed}}{\text{Total number of units inspected}} \]

Noting that

(i) \( E[\text{Total number of defectives observed}] = p E[\text{Total number of units inspected}] = pm \) (ASN)

Where the expression for ASN is given in the next section, and (ii) the asymptotic variance of the maximum likelihood estimate of \( p \) is given by

\[ V(\hat{p}) = -\frac{1}{E\left(\frac{\partial^2 \log L}{\partial p^2}\right)} \]

one has

\[ V(\hat{p}) = \left[ \frac{q - p}{pq^2} pm(ASN) + \frac{m}{q} (ASN) \right]^{-1} \]

\[ = -\frac{pq}{m(ASN)} \quad (4.1) \]

Furthermore, it has been verified that the maximum likelihood estimate of the fraction defective and its asymptotic variance under fully-curtailed MSP have the same features as those of (3.10) and (4.1) respectively. These features were also observed in fully-curtailed and semi-curtailed single sampling plans by Phatak and Bhatt [5].

5. THE ASN FOR A FULLY-CURTAILED DSP

The average sample number, ASN, is merely the average number of units inspected. Thus

\[ \text{ASN} = E(Y). \]

Then it follows from (3.6) that

\[ E(Y) = \sum_{y} y \sum_{r=1}^{k} j_{r}(y; p) \]

\[ = \sum_{y_{x_{1}}} y_{x_{1}}(y; p) + \sum_{y_{x_{2}}} y_{x_{2}}(y; p) + \sum_{y_{x_{3}}} y_{x_{3}}(y; p) + \sum_{y_{x_{4}}} y_{x_{4}}(y; p) \]

(5.1)

Each of the four summation terms of the right hand side of the above expression can be expressed as given below:

\[ \sum_{y_{x_{1}}} y_{x_{1}}(y; p) = \frac{r_{1}}{p} [1 - B(p, n_{1} + 1, r_{1})] \quad (5.2) \]

\[ \sum_{y_{x_{2}}} y_{x_{2}}(y; p) = \sum_{t=1}^{k} \left( \frac{n_{2}}{q_{1} - t} \right) p^{n_{2} - r_{1} - 1} q^{r_{1} - 1} \]

\[ \cdot \left[ \frac{b_{2} + 1 - t}{p} [1 - B(p, n_{2} + 1, b_{2} + 1 - t)] \right] + n_{1} \cdot B(p, n_{2}, b_{2} - t) \]

(5.3)

\[ \sum_{y_{x_{3}}} y_{x_{3}}(y; p) = \frac{q_{1}}{q} B(p, n_{1} + 1, n_{1} - q_{1}) \]

(5.4)

\[ \sum_{y_{x_{4}}} y_{x_{4}}(y; p) = \sum_{t=1}^{k} \left( \frac{n_{4}}{q_{1} - t} \right) p^{n_{4} - r_{1} - 1} q^{r_{1} - 1} \]

\[ \cdot \left[ \frac{q_{2} - q_{1} + t}{q} [1 - B(p, n_{2} + 1, b_{2} - t)] \right] \]

(5.5)

\[ \text{where} \]

\[ B(p, n, k) = \sum_{x=0}^{n} \binom{n}{x} p^{x} (1 - p)^{n-x} \]

These expressions have been obtained using the identity

\[ \sum_{x=0}^{n} \binom{n}{x} p^{x} (1 - p)^{n-x} = 1 - B(p, r + k, k - 1) \]

whose proof can be had in [3, 4].

The ASN which is the sum of the above four terms given by (5.2) through (5.5) can be expressed as

\[ \text{ASN} = \frac{r_{1}}{p} [1 - B(p, n_{1} + 1, r_{1})] \]

\[ + \frac{q_{1}}{q} B(p, n_{1} + 1, n_{1} - q_{1}) \]

\[ + \sum_{t=1}^{k} \left( \frac{n_{2}}{q_{1} - t} \right) p^{n_{2} - r_{1} - 1} q^{r_{1} - 1} \]

\[ \cdot \left[ \frac{b_{2} + 1 - t}{p} [1 - B(p, n_{2} + 1, b_{2} + 1 - t)] \right] \]

\[ + n_{1} \cdot B(p, n_{2}, b_{2} - t) \]

(5.6)

One can calculate the ASN using (5.6) with the help of the usual binomial tables such as the tables of the Cumulative Binomial Probability Distribution [6]. However, for large values of \( n \), the binomial tables including [6] do not give the cumulative probability at an unit interval for \( n \). In that case the following recurrence relation

\[ B(p, n + r, k) = \sum_{x=0}^{r} \binom{r}{x} p^{x} (1 - p)^{r-x} B(p, n + 1, k - x) \]

(5.7)

may be found useful.
6. EASY EXECUTION OF A FULLY-CURTAILED MSP WITH A NUMERICAL ILLUSTRATION

The complexity in executing the plans increases when one wants to execute a fully-curtailed MSP for \( k \geq 3 \). For easy execution of any MSP-curtailed or uncurtailed a device is introduced. This device is to use a grid or the so called storage board [1]. This device is also useful in easily determining the ASN and the probability of acceptance [2]. This device consists of the usual graph (grid) with rectangular coordinates whose ordinates denote defectives and abscissa denote the number of units inspected. A point with coordinates \((y, d)\) represents the occurrence of \( d \) defectives in the inspection of \( y \) units. For a given sampling plan firstly one plots the boundary points of acceptance and rejection regions in accordance with the statement of a given plan. Furthermore, one finds the number of paths for reaching to every boundary point from the origin \((0, 0)\). Then the ASN of a given plan is

\[
ASN = \sum_p y_p, \quad (6.1)
\]

where

(i) \( p_y \) = probability of accessing to a boundary point \((y, d)\)

\[
= \frac{\text{number of paths reaching to a boundary point } (y, d)}{p^y q^d}. \quad (6.2)
\]

(ii) \( y \) = the abscissa for a boundary point \((y, d)\),

(iii) \( \sum_p \) denotes the summation over all the boundary points.

Then during the execution of the plan, as the inspection of units of a lot progresses, starting from the origin one plots a point one unit horizontally to the right when a nondefective occurs and one unit horizontally to the right and one unit vertically up if a defective occurs. The procedure of plotting the points on the graph (and thereby the process of inspection) is continued till the path joining the points plotted reaches one of the boundary points of rejection and acceptance regions.

We now illustrate the results with reference to the following fully-curtailed triple sampling plan

\[
N = 150
\]

\[
n_1 = 4, n_2 = 5, n_3 = 6, r_1 = 3, r_2 = 5, r_3 = 0, \quad g_1 = 3, g_2 = 6, g_3 = 10.
\]

A little perspective will reflect that the boundary points \((y, d)\) of the acceptance region of the above plan are \((3, 0), (4, 1), (8, 2), (9, 3), (14, 4), (15, 5)\) and of the rejection region are \((3, 3), (4, 3), (7, 5), (8, 5), (9, 5), (11, 6), (12, 6), (13, 6), (14, 6), (15, 6)\).

The number of different paths accessible from \((0, 0)\) to the respective boundary points of the acceptance region are

\[
1, 3, 6, 24, 60, 300
\]

and to those of the rejection region are

\[
1, 3, 6, 18, 36, 60, 120, 180, 240, 300.
\]

Then using the relation \((6.1)\) we find that the ASN of this plan is the sum of the following six terms

\[
3(p^3 q^0) + 4(3p^5 q), \quad 7(6p^5 q^3) + 8(18p^5 q^4) + 9(36p^6 q^5),
\]

\[
11(60p^6 q^5) + 12(120p^6 q^6) + 13(180p^6 q^7) + 14(240p^6 q^8) + 15(300p^6 q^9),
\]

\[
3(p^3 q^0) + 4(3p^5 q), \quad 8(6p^5 q^5) + 9(24p^5 q^6),
\]

\[
14(60p^6 q^9) + 15(300p^6 q^{10}). \quad (6.3)
\]

The probability of rejection of a lot is the sum of the quantities involved in brackets in first three terms of the above display whereas the probability of acceptance of a lot is the sum of the quantities involved in brackets in the last three terms of the same display.

Using the model sampling method the above plan was administered on 25 lots each with fraction defective equal to 0.2. The following is the summary of the inspection data:

(i) Only one lot is rejected by finding 3 defectives. These 3 defectives are observed during the inspection of 4 units.

(ii) 11 lots are accepted by finding 3 nondefectives where the number of units inspected in each case is 3. Whereas 7 lots are accepted by finding 3 nondefectives where the number of units inspected in each case is 4.

(iii) 2 lots are accepted by finding 6 nondefectives where the number of units inspected in each case is 8 whereas 2 lots are accepted by finding 6 nondefectives where the number of units inspected in each case is 9.

(iv) 2 lots are accepted by finding 10 nondefectives where the number of units inspected in each case is 15.

It is realized from the above summary that the total number of defectives observed is 30 and the total number of units inspected is 190. Hence the maximum likelihood estimate of the fraction defective is \(0.23\). Substituting the hypo-
theoretical value of $p$, namely 0.2 in (6.3) one finds that the ASN = 4.3396. The asymptotic variance = $(0.2)(0.8)/(25 \times 4.3396) = 0.00147$. In practice one may use $\hat{p}$ to calculate the estimate of the asymptotic variance when one does not know the true value of $p$.

References


Approximating the Distribution of the Sample $R^2$ in Best Subset Regressions

George Diehr and Donald R. Hoflin

Graduate School of Business Administration
University of Washington
Seattle, Washington

This note presents research on the problem of determining the distribution of the usual sample $R^2$ statistic in multiple regression studies where the variables to be included in the regression equation are the subset of $k$ variables, from a set of $m$ variables, which maximize the sample $R^2$ value or satisfy some similar criterion. A Monte-Carlo approach was used to estimate certain percentile points of the distribution of $R^2$ under the null hypothesis of independence between the dependent variable and the $m$ independent variables. A function has been developed which appears to provide a good approximation to percentile points of the $R^2$ distribution.

KEY WORDS
Subset Regression
$R^2$ Distribution

1. INTRODUCTION
There is considerable interest in the problem of determining the best $k$ of $m$ variable regression (the best subset problem) for a set of $n$ observations on $m$ independent variables and one dependent variable. Both heuristic algorithms, such as stepwise regression (Draper and Smith, [3]; Effroymson, [4]; Mantel, [9]), and more recently exact techniques (Garside, [6]; Beale et al, [2]; Furnival, [5]; Hocking and Leslie, [7]; Schatzoff et al, [10]) have been programmed for this problem of determining the subset of $k$ variables which maximize the sample multiple correlation coefficient, $R^2$. A major problem with such data analytic techniques is significance level testing of the resulting sample $R^2$. The typical $F$ test used to evaluate $R^2$ is clearly biased except when the entire set of $k = m$ variables is included in the regression equation.

Unfortunately, the distribution theory for the sample $R^2$ statistic in best subset regressions is hopelessly complex. Thus, we have utilized a Monte-Carlo approach to determine an approximation formula for its distribution. We have studied its distribution in the particular case of a population of $m+1$ independent normally distributed variates. This restriction to independence among the independent variables (i.e., diagonal population covariance matrix) means that the use of our results for significance level testing of $R^2$ are conservatively biased when the covariances among the independent variables are non-zero. (This contrasts the opposite bias in the usual $F$ statistic.)

2. RESEARCH DESIGN
The distribution of the sample $R^2$ in best subset regression on $n$ observations was estimated by a Monte-Carlo approach, detailed below. A function was then devised to fit the Monte-Carlo results. The function parameters can be determined from two known points of the $R^2$ distribution. Sample tables have been provided using this function for several values of $k, m,$ and $n$.

The Monte-Carlo experiment was performed in a direct fashion: $n$ observations were selected from an $m+1$ dimensional multivariate normal population with zero mean and identity covariance matrix. For the sample, the "best" $k$ variable regression was determined by exhaustive enumeration of the $(m \choose k)$ regressions for values of $k$ from 1 to $m$. For each value of $k$, the resulting maximum $R^2$ was determined and saved. Then, the procedure was repeated 100 times. Ordering the 100 sample $R^2$ values for particular values of $k, m,$ and $n$ provided estimates, $R^2(k, m, n, \alpha)$ of the upper $\alpha$ percentile points of the sample $R^2$ distribution, $R^2(k, m, n, \alpha)$. The values of $k, m,$ and $n$ used were $m = 2(1)8, k = 1(1)m,$ and $n = 10, 25, 50, 100, 200$.

Figure 1 graphically presents some of the Monte-Carlo results along with "hypothesized" and known values of $R^2$, (see discussion below). This figure includes a plot for each of the five values of sample size, $n = 10, 25, 50, 100, 200$, for the case of eight available independent variables ($m = 8$). The graphs indicate the estimated 90th percentile

Received Aug. 1972; revised June 1973.
3. APPROXIMATION TO THE BEST SUBSET

$R^2$ DISTRIBUTION

The $R^2$ distribution has the parameters $k$, $m$, and $n$; i.e., $R^2(k, m, n, \alpha)$ where $\alpha$ is the probability of a larger $R^2$ value. When $k = m$ the distribution is related to the known distribution of the usual $F$ statistic with $n$ and $n - m - 1$ degrees of freedom by:

$$R^2(k, m, n, \alpha) = mF(m, n - m - 1, \alpha)/(n - m - 1 + mF(m, n - m - 1, \alpha))$$

(3.1)

The second known value of the $R^2$ distribution is obtained at $k = 1$. Since we assume that the observations are drawn from a multivariate normal population with diagonal covariance matrix, if we test each variable at level $\alpha^*$ the probability that one or more variables will be significant is given by:

$$\alpha = 1 - (1 - \alpha^*)^m$$

Since we want the overall test to be at level $\alpha$ the best variable should be tested at level $\alpha^*$, given by:

$$\alpha^* = 1 - (1 - \alpha)^{1/m}$$

(3.2)

The test statistic used for the best single variable is given by $F(1, n - 2, \alpha^*)$. Using transformation (3.1) above we obtain the proper critical value for testing a sample $R^2$ at $k = 1$.

Given these two known points in the distribution of $R^2$, we visually examined some of the Monte-Carlo results to determine if a function could be found which, when fit through the known points, would also fit the intermediate Monte-Carlo estimates of $R^2(k, m, n, \alpha)$ for fixed values of $m$, $n$, and $\alpha$, for $k = 2(1)m - 1$. The approximation, $\hat{R}^2(k, m, n, \alpha)$ is given by:

$$\hat{R}^2(k, m, n, \alpha) = w(1 - v^*)$$

(3.3)

with $w$ and $v$ determined from the known values $R^2(1, m, n, \alpha)$ and $R^2(m, m, n, \alpha)$. Note that (3.3) has the desired property that $\hat{R}^2(0, m, n, \alpha) = 0$.

Graphs such as appear in Figure 1 suggest that the function (3.3) provides a good fit to the true distribution. The fit was tested by comparing the percent deviations of the Monte-Carlo results from the known points $R^2(1, m, n, \alpha)$ and $R^2(m, m, n, \alpha)$ to the deviations from the estimated points, $\hat{R}^2(k, m, n, \alpha)$ for $k = 2(1)m - 1$. Both $t$ tests and non-parametric tests of these deviations provided test statistics insignificant at the 50% level.

The deviations at the known points also provide indication of any bias in the average over the
100 Monte-Carlo repetitions. Since we are using the Monte-Carlo results only indirectly to suggest and test our function (3.3), their average error in estimating each $R^2$ value (on the order of 8% of the $R^2$ value) is not particularly critical. What is critical, is that there be little or no bias in estimating the true $R^2$ value else our choice of function (3.3) might exhibit the same bias. A test of average percent deviation of 70 Monte-Carlo estimates from true $R^2$ values (at $k = 1$ and $k = m$) produced a t-statistic of $-0.02$ supporting our conjecture that the Monte-Carlo results have not introduced bias in our selection of (3.3) or our testing of it as an approximation to $R^2$ at intermediate ($k = 2(1)m - 1$) points.

4. USE OF APPROXIMATION (3.3) AND THE SAMPLE TABLES

Tables of the percentile points of $R^2$ have been provided for $m = 5$ and $m = 10$ at several values of $k$ and several sample sizes, as computed from (3.3). For example, in a regression using the best $k = 5$ of $m = 10$ independent variables and sample size 31, Table 2 indicates the following percentile points: $R^2_{.10} = .468$, $R^2_{.05} = .513$, and $R^2_{.01} = .609$.

As an example of the use of formula (3.3) we tested the sample $R^2$ on an example from Draper and Smith [3, pp. 195-203]. Their example illustrates a stepwise regression for a sample of 17 observations on 5 independent variables. Their selection for a $k = 2$ variable regression produces $R^2 = .574$ and we desire to test this value for significance at the 0.05 and 0.01 levels. The two known values of the distribution at the 0.05 level are found from:

1. $R^2(5, 5, 17, 5\%)$
   \[ = 5F(5, 11, 5\%)/(11 + 5F(5, 11, 5\%)) \]
   \[ = 5(3.20)/(11 + 5(3.20)) \]
   \[ = 0.593 \]

2. From (3.2), $\alpha^* = 1 - (1 - .05)^{1/5}$
   \[ = 1.03% \approx 1.0% \]
   \[ R^2(1, 5, 17, 5\%)/(15 + F(1, 15, 1\%)) \]
   \[ = 0.305 \]

Next, (3.3) must be solved iteratively for the parameters $w$ and $v$. A good starting value is a value slightly larger than $R^2(5, 5, 17, 5\%)$, say 0.6. Solving for $v$ at $k = 1$, gives:

\[ v = 1 - R^2(1, 5, 17, 5\%)/w = 1 - .365/.6 = 0.39 \]

Then, solving for $w$ at $k = m$, gives:

\[ w = R^2(5, 5, 17, 5\%)/(1 - v^2) = .539/.991 = 0.603 \]

Two additional iterations produce insignificant changes in $v$ and $w$, and final values: $v = .395$, $w = .605$.

Evaluation of (3.3) for $k = 2$ gives the desired estimate of the critical value:

\[ R^2(2, 5, 17, 5\%) = .605(1 - .395^2) = 0.511 \]

Similar manipulations for a significance level of one percent yields: $R^2(2, 5, 17, 1\%) = .363$. Thus, the sample $R^2$ value of 0.574 is significant at the five percent level but not at the one percent level. The usual $F$ test produces a critical $R^2$ value of 0.482 at the 1% level. Thus, the use of the standard $F$ test could lead to a rejection of the null hypothesis at a level quite different from that produced by our approach.

5. CONCLUSION

The research presented here provides one approach to testing the significance level of sample $R^2$ values in best subset regression problems. Using the tables presented here or standard tables of the $F$ distribution and formula (3.3) the empirical researcher can

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{k} & \textbf{a} & \textbf{10} & \textbf{20} & \textbf{50} & \textbf{100} \\
\hline
1 & .10 & .326 & .203 & .095 & .050 \\
& .05 & .356 & .249 & .119 & .082 \\
& .01 & .506 & .334 & .163 & .088 \\
\hline
2 & .10 & .484 & .290 & .136 & .042 \\
& .05 & .535 & .343 & .163 & .087 \\
& .01 & .666 & .448 & .222 & .120 \\
\hline
3 & .10 & .523 & .328 & .154 & .081 \\
& .05 & .593 & .382 & .183 & .097 \\
& .01 & .717 & .448 & .244 & .132 \\
\hline
\end{tabular}
\caption{Percentile Points of $R^2$ Distribution for $m = 5$.}
\end{table}

\textit{Technometrics}, Vol. 16, No. 2, May 1974
**Table 2—Percentile Points of $R^2$ Distribution for $m = 10$.**

<table>
<thead>
<tr>
<th>n - m - 1</th>
<th>k</th>
<th>α</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.10</td>
<td>.298</td>
<td>.298</td>
<td>.298</td>
<td>.298</td>
<td>.298</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.442</td>
<td>.442</td>
<td>.442</td>
<td>.442</td>
<td>.442</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
</tr>
<tr>
<td>2</td>
<td>.10</td>
<td>.469</td>
<td>.469</td>
<td>.469</td>
<td>.469</td>
<td>.469</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.531</td>
<td>.531</td>
<td>.531</td>
<td>.531</td>
<td>.531</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
</tr>
<tr>
<td>3</td>
<td>.10</td>
<td>.568</td>
<td>.568</td>
<td>.568</td>
<td>.568</td>
<td>.568</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
<td>.632</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.745</td>
<td>.745</td>
<td>.745</td>
<td>.745</td>
<td>.745</td>
</tr>
<tr>
<td>5</td>
<td>.10</td>
<td>.657</td>
<td>.657</td>
<td>.657</td>
<td>.657</td>
<td>.657</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.716</td>
<td>.716</td>
<td>.716</td>
<td>.716</td>
<td>.716</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.811</td>
<td>.811</td>
<td>.811</td>
<td>.811</td>
<td>.811</td>
</tr>
</tbody>
</table>

guard against the use of regression equations which overfit the data. Furthermore, we feel that a more valid indication of the "significance level" of a best subset or stepwise regression $R^2$ is given by our approach rather than by the standard approach. While biased on the conservative side, the use of approximation (3.3) is, to our knowledge, the only test of overall $R^2$ value which takes into account the number of independent variables which have been searched for the best subset.

The problem of stopping rules in variable selection and the choice of value for $k$ remain to be solved although we feel that the use of the tables presented here may aid the researcher in the selection of $k$. For example, in comparing subsets at different values of $k$, a large drop in significance level of $R^2$ as $k$ increases would encourage use of the smaller subset. Other problems which require more investigation include the estimation of "slippage" or "shrinkage"—the difference in predictability on the sample and on the population [Lachenbruch, 8] and the development of a test statistic for individual regression parameters in best subset problems.

Our appreciation to the referees for many helpful suggestions.

**References**


Some Results on Non stationary First Order Autoregression

Edward J. Wegman
The University of North Carolina
Chapel Hill, North Carolina

Consider the first order autoregressive model, \( X_t = a(t)X_{t-1} + \epsilon_t, \ t = 1, 2, \ldots \). The mean and covariance structure is computed under the hypotheses that \( X_t \) is not second order stationary and that \( \epsilon_t \) is an uncorrelated sequence of random variables with \( E\epsilon_t = 0 \) and \( \text{var} (\epsilon_t) = \sigma_t^2 \), not constant. Two applications are mentioned.

Consider the first order autoregressive time series model

\[
X_t = a(t)X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots . \tag{1}
\]

The usual assumptions for this model are:
1. \( a(t) = a, \) constant.
2. \( X_t \) is second order stationary, i.e. \( E(X_t) = 0 \) and \( \text{cov} (X_t, X_{t-s}) = \gamma_s \), a function not depending on \( t \).
3. \( E(\epsilon_t) = 0, \) \( \text{var} (\epsilon_t) = \sigma_t^2 \), a constant and that \( \epsilon_t \) is uncorrelated with \( \epsilon_{t'}, \ t \neq t' \) and \( \epsilon_t \) is uncorrelated with \( X_{t'}, \ t' < t \).

Under these assumptions, one may derive (see [1, pp. 57-58]).

\[
\gamma_s = \frac{a^s \sigma_t^2}{1 - a^2},
\]

relating the covariance structure of the process to the elementary constants. The constants, \( a \) and \( \sigma_t^2 \), may be estimated by least squares procedures.

The variance-covariance structure may be easily derived without assumptions 1 and 2. Moreover, under less severe restraints on \( a(t) \), the function, \( a(t) \), may be estimated with least squares techniques. Let us also assume \( \text{var} (X_t) = \sigma_t^2 \), not necessarily constant.

Suppose \( X_0 = x_0 \) is the known initial condition. We write \( EX_t \) to mean the conditional expectation of \( X_t \) given \( X_0 = x_0 \). (All expectations will be conditioned on \( X_0 = x_0 \)). From (1), it is easy to see

\[
EX_t = a(t)EX_{t-1}, \quad t = 1, 2, \ldots .
\]

Using the formula recursively, we obtain

\[
EX_t = \left( \prod_{j=1}^{t} a(j) \right)x_0, \quad t = 1, 2, \ldots . \tag{2}
\]

Now multiplying both sides of (1) by \( X_{t-s} \), we have

\[
X_tX_{t-s} = a(t)X_{t-1}X_{t-s} + \epsilon_tX_{t-s}, \quad s > 0,
\]

so that by taking expectations, we have

\[
E(X_tX_{t-s}) = a(t)E(X_{t-1}X_{t-s}).
\]

Subtracting the product of the appropriate means,

\[
\text{cov} (X_t, X_{t-s}) = a(t) \text{cov} (X_{t-1}, X_{t-s}). \tag{3}
\]

Applying (3) recursively,

\[
\text{cov} (X_t, X_{t-s}) = a(t)a(t-1) \cdots a(t-s+1) \cdot \text{var} (X_{t-s}). \tag{4}
\]

Thus the system of covariances depends on the variances. Applying (1) recursively leads to

\[
X_t = \left( \prod_{j=1}^{t} a(j) \right)x_0 + \sum_{j=1}^{t-1} a(t) \cdots a(t-j+1)a(t-j+1) \epsilon_{t-j} + \epsilon_t.
\]

Subtracting (2) from this yields

\[
(X_t - EX_t) = \left( \prod_{j=1}^{t} a(j) \right)(X_0 - EX_0) + \sum_{j=1}^{t-1} a(t) \cdots a(t-j+1)a(t-j+1) \epsilon_{t-j} + \epsilon_t.
\]

Since \( X_0 = x_0 = EX_0 \)

\[
(X_t - EX_t) = \sum_{j=1}^{t-1} a(t) \cdots a(t-j+1)a(t-j+1) \epsilon_{t-j} + \epsilon_t.
\]

Squaring, taking expected values and recalling \( \epsilon_t \) is uncorrelated with \( \epsilon_{t'}, \ t \neq t' \),

Received June 1973; revised Aug. 1973.
\[ \text{var} (X_i) = \sum_{j=1}^{t-1} [a(t) \cdots a(t-j+1)]^2 \sigma_{t-i}^2 + \sigma_i^2. \]  

(5)

Combining (4) and (5) yields the complete covariance structure, the only assumptions being that \( E(\varepsilon_t) = 0, \text{var} (\varepsilon_t) = \sigma_t^2, \varepsilon_t \) is uncorrelated with \( \varepsilon_t \) for \( t \neq t' \) and with \( X_t, t' < t \).

Adding the assumptions that \( \sigma_t^2 = \sigma^2, \) a constant and that \( a(t) = a, \) a constant allows (5) to be rewritten as

\[ \text{var} (X_i) = \sigma^2 \sum_{i=0}^{t-1} a^i. \]

Using the formula for a geometric sum,

\[ \text{var} (X_i) = \frac{\sigma^2 (1 - a^{t+1})}{1 - a^2}. \]

Combining this equation with (4) yields

\[ \text{cov} (X_i, X_{i-1}) = \frac{a^2 \sigma^2 (1 - a^{t+1})}{1 - a^2} \]

(6)

which differs by a factor of \( (1 - a^2) \) from the usual stationary covariance.

In (1), it is of considerable interest to estimate \( a(t) \) Suppose that we may parameterize \( a(t) \) (either exactly or as an approximation) as an \( l \)-th degree polynomial. Then we may use standard least squares procedures to determine the coefficients. In (1) let us assume \( \varepsilon_t \) has constant variance and assume we have \( n \) observations \( x_0, \cdots, x_n \). Then we wish to minimize

\[ \sum_{i=1}^{n} \left( x_i - \left( \sum_{j=0}^{i} a_{i-j}x_{i-j} \right) \right)^2. \]

Taking partial derivatives with respect to \( a_{i-j} \) and equating to zero, we arrive at the normal equations

\[ \sum_{i=1}^{n} \varepsilon_i x_{i-j} = \sum_{j=0}^{i} a_{i-j} \left( \sum_{k=0}^{n-1} \varepsilon_k x_{i-k-1} \right), \quad k = 0, \cdots, l. \]

Solving these equations simultaneously gives an estimate, \( \hat{a}_{i-j} \) of \( a_{i-j} \), \( j = 0, \cdots, l \), and hence an estimate of \( a(t) \). Note in the particular case \( a(t) = a, \) a constant, then the estimate of \( a \) is \( \hat{a}_0 \) which is

\[ \left[ \sum_{i=1}^{n} x_i x_{i-1} \right] \left[ \sum_{i=1}^{n} x_{i-1} \right]^{-1}. \]

Note also that it is possible that \( a(t) \) may be more conveniently parameterized as some other orthogonal series, notably a trigonometric series. Similar least squares techniques are appropriate here.

Example 1: Consider a capacitor, with capacitance \( C \) in farads, discharging its initial charge \( Q \) in coulombs across a resistance \( R \) in ohms. At time \( t \) in seconds, it is well-known that the remaining charge, \( q(t) \), is given by

\[ q(t) = Qe^{-t/RC}. \]

From a purely empirical point of view, however, it is reasonable to expect that the charge at time \( t \) is a percentage of the charge at the previous time \( t - 1 \).

It is thus reasonable to expect that the relationship, \( q(t) = a(t)q(t - 1) \), holds. It is, however, unlikely with real measurements that we will find \( q(t) - a(t)q(t - 1) = 0 \) but rather some non-zero error term, say \( \varepsilon_t \). Thus we have

\[ q(t) = a(t)q(t - 1) + \varepsilon_t, \quad t = 1, 2, \cdots \]

the nonstationary autoregressive model. Here, we can identify \( a(t) \) with \( \exp (-1/RC) \). Thus, we can use nonstationary autoregression to estimate the physical constants, in this case \( 1/RC \), the so-called time constant of the circuit. Other examples of situations in which nonstationary autoregression is appropriate are radioactive decay, population growth (or decline) and decay of learning curves.

Example 2: Notice that knowing the mean and variance of the process \( X_t \) allows us to project the trajectory of the curve as well as allowing us to give confidence intervals for this trajectory. This would, for example in connection with radioactive decay provide some measure of accuracy for \( C_{14} \) dating.

Suppose that \( X_t \) is the proportion of \( C_{14} \) content of an object \( t \) time units in the past. The nature of radioactive decay suggests that

\[ X_t = aX_{t-1} + \varepsilon_t, \quad t = 1, 2, \cdots \]

with \( \text{var} (\varepsilon_t) = \sigma^2 \), where, of course, \( a \) is some number greater than or equal to 1. Based on objects whose antiquity is known, \( a \) and \( \sigma^2 \) may be estimated. Thus the trajectory (backwards through time) of \( X_t \) may be projected and using the normal assumption, a confidence interval \( (l, u_t) \) for \( EX_t \) may be formed. The usual estimate of age is given by that \( t \) for which \( EX_t = S \) where \( S \) is standard proportion of \( C_{14} \) content for a currently living organism. Clearly an older limit on the age is given by the \( t \) for which \( l = S \) and a younger limit by the \( t \) for which \( u_t = S \).

Reference

Approximating the Negative Binomial via the Positive Binomial

M. R. Binns
Statistical Research Service
Canada Department of Agriculture
Ottawa, Ontario

Bartko (1966) compared several approximations to the negative binomial, but did not include among them the equivalent of the ordinary normal approximation to the positive binomial. A table of maximum errors for this approximation is presented.

Key Words
Negative Binomial
Approximation

Bartko (1966) displayed tables of maximum errors for various approximations to the negative binomial distribution, but did not include the approximation which is derived directly from the normal approximation to the positive binomial distribution: the probability of less than \( c \) successes in \( n \) positive binomial tries with parameter \( p \), is equal to the probability of more than \( n \) negative binomial tries to obtain \( c \) successes with the same parameter \( p \). Using this well-known fact (Bartko, 1967), we may derive from the positive binomial normal approximation:

\[
\Pr(\text{less than } c \text{ successes} \mid n, p) = \Phi \left[ \frac{c - np - 0.5}{\sqrt{np(1 - p)}} \right]
\]

a normal approximation to the negative binomial:

\[
\Pr(\text{more than } n \text{ tries} \mid c, p) = \Phi \left[ \frac{c - np - 0.5}{\sqrt{np(1 - p)}} \right].
\]

Therefore the approximation to the distribution function of the number \( N(\geq c) \) of tries is

\[
\Pr(\text{number of tries} \leq n \mid c, p) = \Phi \left[ \frac{np - c + 0.5}{\sqrt{np(1 - p)}} \right].
\]

In the usual notation for the negative binomial distribution \( c \) is replaced by \( k \) and \( N \) is replaced by \( k + x \). Therefore \( x \geq 0 \) is distributed according to the negative binomial distribution with parameters \( p \) and \( k \):

\[
P_x = \Pr(x \mid p, k) = \frac{\Gamma(k + x)}{\Gamma(x + 1)\Gamma(k)} p^x(1 - p)^k.
\]

Let

\[NB(x, k, p) = \sum_{i=x}^{\infty} P_i\]

and

\[ANB(x, k, p) = \Phi \left[ \frac{xp - kq + 0.5}{\sqrt{(x + k)pq}} \right],\]

where

\[q = 1 - p.\]

Bartko tabulated the maximum error \( M(k, p) \) defined by

\[M(k, p) = \max_{x} [NB(x, k, p) - ANB(x, k, p)]\]

where \( NB(x, k, p) \) is his \( i \)-th approximation to \( NB(x, k, p) \). Values of \( M \) with \( ANB(x, k, p) \) in place of \( NB(x, k, p) \) are displayed in Table 1. Comparison with Bartko’s tables show that \( ANB \) is better than the straightforward normal approximation, and for \( p < 0.7 \) it is better than the Poisson Gram Charlier series. It is, however, inferior everywhere to the Cram Paulson approximation and it is also inferior, at least for small \( x \), to Guenther’s (1972) approximation.

No credit is claimed for originality in these calculations; it is, however, useful to extend Bartko’s list of approximations, especially since the relation between the negative binomial and positive binomial

Received May 1973; revised Aug. 1973.
Table 1—Maximum errors of the approximation.

<table>
<thead>
<tr>
<th>Values of p</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.050</td>
<td>0.020</td>
<td>0.011</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.027</td>
<td>0.018</td>
<td>0.011</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>0.20</td>
<td>0.022</td>
<td>0.015</td>
<td>0.009</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td>0.30</td>
<td>0.019</td>
<td>0.010</td>
<td>0.006</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>0.40</td>
<td>0.008</td>
<td>0.006</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>0.50</td>
<td>0.009</td>
<td>0.006</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>0.60</td>
<td>0.011</td>
<td>0.007</td>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>0.70</td>
<td>0.012</td>
<td>0.009</td>
<td>0.010</td>
<td>0.007</td>
<td>0.005</td>
</tr>
<tr>
<td>0.80</td>
<td>0.010</td>
<td>0.009</td>
<td>0.011</td>
<td>0.013</td>
<td>0.009</td>
</tr>
<tr>
<td>0.90</td>
<td>0.030</td>
<td>0.010</td>
<td>0.026</td>
<td>0.024</td>
<td>0.012</td>
</tr>
<tr>
<td>0.95</td>
<td>0.078</td>
<td>0.099</td>
<td>0.053</td>
<td>0.040</td>
<td>0.027</td>
</tr>
</tbody>
</table>

is well known and the normal approximation to the positive binomial is probably used more often than any other.

References to more complicated approximations to the positive (and hence to the negative) binomial distribution may be found in Molenaar (1973).

References