REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
September, 1996

3. REPORT TYPE AND DATES COVERED
Interim Technical Report, September1996

4. TITLE AND SUBTITLE
Multiple Decision Procedures in Analysis of Variance and Regression Analysis

5. FUNDING NUMBERS
DAAH04-95-01-0165

6. AUTHOR(S)
S. S. Gupta, D. Y. Huang, S. Panchapakesan

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Purdue Research Foundation
Purdue University
West Lafayette, IN 47907

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report #95-44C

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 32922.8 -MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We consider testing of the homogeneity hypothesis in the one-way ANOVA model and testing for the significance of regression in the multiple linear regression model. Unlike in the classical approach, there is no alternative hypothesis to accept when the null hypothesis is rejected. When there is a substantial deviation from the null hypothesis we reject the null hypothesis and also identify the independent variables or the levels that contributed most towards the deviation from the null hypothesis.

19961125 118

14. SUBJECT TERMS
Hypothesis testing, Type I error, Power of Test, Signal-to-Noise ratio

15. NUMBER IF PAGES
11

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

MULTIPLE DECISION PROCEDURES IN ANALYSIS OF VARIANCE AND REGRESSION ANALYSIS

by

S. S. Gupta
Purdue University

D. Y. Huang
Fu Jen Catholic University

and

S. Panchapakesan
Southern Illinois University

Technical Report #95-44C

Department of Statistics
Purdue University

December 1995
Revised June 1996
Revised September 1996

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.

△

MULTIPLE DECISION PROCEDURES IN ANALYSIS OF VARIANCE AND REGRESSION ANALYSIS

by

S. S. Gupta
Purdue University

D. Y. Huang
Fu Jen Catholic University

and

S. Panchapakesan
Southern Illinois University

Abstract

We consider testing of the homogeneity hypothesis in the one-way ANOVA model and testing for the significance of regression in the multiple linear regression model. Unlike in the classical approach, there is no alternative hypothesis to accept when the null hypothesis is rejected. When there is a substantial deviation from the null hypothesis we reject the null hypothesis and also identify the independent variables or the levels that contributed most towards the deviation from the null hypothesis.

Key Words: Hypothesis testing, Type I error, Power of Test, Signal-to-Noise ratio

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.

1. Introduction

Traditional tests of hypotheses are special cases of multiple decision procedures. Usually, they are restricted to two decisions. These testing methods emphasize on a confirmatory approach. The hypotheses are specified *before* the data are collected. However, it is difficult sometimes to specify appropriate hypotheses of interest based on past experience. We take a data analysis approach to the problem of decision making (see Gupta and Huang, 1981).

Usually, an experimenter faces the problem of comparing several categories or populations. The classical approach to this problem is to test homogeneity (null) hypothesis $H_0 : \theta_1 = \ldots = \theta_k$, where $\theta_1, \ldots, \theta_k$ are the unknown values of the parameter θ for these populations. In the case of normal populations with unknown means $\theta_1, \ldots, \theta_k$ and a common unknown variance σ^2, the test can be carried out by means of the F-ratio of the analysis of variance.

In this paper, we formulate above k-sample problem as a multiple decision problem in analysis of variance and regression analysis. Among the early investigators of procedures for such problems are Paulson (1949), Bahadur (1950), Bahadur and Robbins (1950). The formulation of multiple decision procedures in the framework of selection and ranking procedures has been generally accomplished by using either the indifference zone approach or the (random-sized) subset selection approach. The former approach was introduced by Bechhofer (1954). Substantial contribution to the early and subsequent developments in the subset selection theory have been made by Gupta starting from his work in 1956. We will be mainly concerned with multiple decision problems formulated as selection and ranking problems (see Gupta and Panchapakesan (1979)).

In this paper, we discuss inference about the parameters in multiple linear regression model. By a proper reparametrization, ANOVA models can be handled by regression techniques. As Draper and Smith (1981) pointed out, it is useful to appreciate the connection between the two methods of analysis. We consider here only the one-way ANOVA model.

Our interest is not just to test the null hypothesis H_0 against the global alternative. When H_0 is rejected, we want to identify significantly important independent variables and also check the appropriateness (to be explained) of the choice of the variables or the factor levels.

2. Inferences About Regression Parameters

Consider the linear model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{2.1}$$

where $\underline{Y}' = [Y_1, Y_2, \ldots, Y_n]$ is an $n \times 1$ vector of responses, $X = [\underline{1}, \underline{X}_1, \ldots, \underline{X}_{p-1}]$ is an $n \times p (n > p)$ matrix of known constants of rank p, $\underline{\beta}' = [\beta_0, \beta_1, \ldots, \beta_{p-1}]$ is a $1 \times p$ vector of unknown parameters, $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I_n)$, and I_n denotes the $n \times n$ identity matrix. We refer to model (2.1) as the true model of size p. From this model, we obtain $p-1$ so-called reduced models, each of size $p-1$, by dropping one independent variable at a time. Let $X_{(m)}$ denote the "X matrix" of the reduced model obtained by dropping the independent variable X_m, $m = 1, 2 \ldots p-1$. Correspondingly, we have the residual sums of squares for these reduced models denoted by $SS_{p-1,m}$, $m = 1 \ldots, p-1$. Accordingly, the residual sum of squares for the true model is denoted by $SS_{p,1}$. It is known that $SS_{p,1} = \underline{Y}'Q\underline{Y}$ and $SS_{p-1,m} = \underline{Y}'Q_{(m)}\underline{Y}$ where $Q = I_n - X(X'X)X'$ and $Q_{(m)} = I_n - X_{(m)}(X'_{(m)}X_{(m)})^{-1}X'_{(m)}$. Under the true model assumption, it is known that

$$\frac{SS_{p,1}}{\sigma^2} \sim \chi^2_{n-p} \text{ and } \frac{SS_{p-1,m}}{\sigma^2} \sim \chi^2_{n-p,\lambda_{p-1,m}}, m = 1, \ldots, p-1,$$

where $\lambda_{p-1,m} = (X\underline{\beta})'Q_{(m)}(X\underline{\beta})/2\sigma^2, m = 1, 2, \ldots, p-1, \chi^2_\nu$ denotes the (central) chi-square distribution on ν degrees of freedom with noncentrality parameter λ. Thus $E[SS_{p,1}] = (n-p)\sigma^2$ and $E[SS_{p-1,m}] = (n-p+1)\sigma^2 + 2\sigma^2\lambda_{p-1,m}, m = 1, \ldots, p-1$.

We note that $Q_{(m)}, m = 1, 2, \ldots, p-1$, are idempotent and symmetric; thus they are positive semidefinite. Hence $\lambda_{p-1,m}, m = 1, 2, \ldots, p-1$ are nonnegative. Obviously, $\beta_1 = \ldots = \beta_{p-1} = 0$ implies that $\lambda_{p-1,m} = 0$ for $m = 1, \ldots, p-1$. However, the converse is not necessarily true. When $\underline{\beta} \neq \underline{0}, \lambda_{p-1,m}$ can be interpreted as the contribution of X_m in making the regression significant given that the other variables are already in the model. Let

$$\hat{\lambda}_{p-1,m} = \frac{n-p}{2}\left[\frac{SS_{p-1,m}}{SS_{p,1}}\right] - \frac{n-p+1}{2},$$

where

$$\hat{\eta}_m = \frac{SS_{p-1,m} - SS_{p,1}}{SS_{p,1}/(n-p)}$$

has the noncentral F-distribution on 1 and $(n-p)$ degrees of freedom with noncentrality parameter $\lambda_{p-1,m}$, denoted by $F_{1,n-p,\lambda_{p-1,m}}$. We note that the $\hat{\eta}_m$ is the statistic used in the so-called partial F-test for the significance of β_m.

Now, we consider our problem of testing $H_0 : \underline{\beta} = \underline{0}$. The classical F-test for this against the global alternative $H_a : \underline{\beta} \neq \underline{0}$ is designed only to control the probability of type I error. In our formulation, when H_0 is rejected, the decision also includes selecting a subset of the $p-1$ independent variables as significant. When H_0 is false, a correct decision (CD) occurs if H_0 is rejected and the selected subset of the independent variables includes the variable associated with the largest $\lambda_{p-1,m}$. Let $\underline{\lambda} = [\lambda_{p-1,1}, \dots, \lambda_{p-1,p-1}]$ and $\lambda_{p-1,[1]} \leq \dots \leq \lambda_{p-1,[p-1]}$ denote the ordered $\lambda_{p-1,m}$. We require that, for given $0 < \alpha < 1$ and $\frac{1}{p-1} < P^* < 1$,

$$Pr[\text{Reject} H_0 | \underline{\beta} = \underline{0}] \leq \alpha \tag{2.2}$$

and

$$Pr\{CD | \lambda_{p-1,[p-1]} \geq \Delta\} \geq P^* \tag{2.3}$$

where $\Delta > 0$ is specified in advance.

Since $\underline{\beta} = \underline{0}$ implies that $\underline{\lambda} = \underline{0}$, (2.2) is satisfied if

$$Pr[\text{Reject} H_0 | \underline{\lambda} = \underline{0}] \leq \alpha \tag{2.4}$$

and (2.3) is equivalent to

$$\operatorname*{Inf}_{\underline{\lambda}} Pr[CD | \lambda_{p-1,[p-1]} \geq \Delta] = P^*. \tag{2.5}$$

We propose a test of H_0 based on the statistics $SN_m = 10\log\hat{\eta}_m, m = 1, \dots, p-1$, where the log is to base 10. We could just use the $\hat{\eta}_m$, but the transform denoted by SN_m is based on Taguchi's idea of signal-to-noise ratio. Our test procedure is as follows:

Reject H_0 if $SN_i \geq C$ for some i. Include in the selected subset of significant variables all the variables X_i's for which $SN_i \geq C$.

The constant C should satisfy

$$Pr\{SN_i \geq C \text{ for some } i|\underline{\lambda} = \underline{0}\} \leq \alpha \tag{2.6}$$

and

$$\operatorname*{Inf}_{\underline{\lambda}} Pr\{SN_{(p-1)} \geq C|\lambda_{p-1,[p-1]} \geq \Delta\} = P^* \tag{2.7}$$

where SN_{p-1} denotes the SN_i associated with $\lambda_{p-1,[p-1]}$. When $\underline{\lambda} = \underline{0}, SN_i, i = 1, \ldots, p-1$, are correlated each having a central $F_{1,n-p}$ distributions. By using the inequality: $P[\cup A_j] \leq \sum_j P(Aj)$, (2.6) is satisfied if

$$\sum_{m=1}^{p-1} Pr\{SN_m \geq C|\underline{\lambda} = \underline{0}\} = \alpha$$

which gives

$$Pr\{F_{1,n-p} \leq 10^{\frac{C}{10}}\} = 1 - \frac{a}{p-1} = \gamma(say). \tag{2.8}$$

When $\lambda_{p-1,[p-1]} > 0, SN_{(p-1)}$ has the stochastically increasing property in terms of the noncentrality parameter. Thus, (2.7) is satisfied if

$$Pr\{F_{1,n-p;\Delta} \leq 10^{\frac{C}{10}}\} = 1 - P^*. \tag{2.9}$$

We obtain an approximate solution to (2.8) and (2.9) in the form of C as a function of n by using the following lemma of Huang (1996).

Lemma: Let Y have the noncentral F distribution with u, v degrees of freedom, and noncentrality Δ, denoted by $F_{u,v,\Delta}$. Then A

$$Pr\{Y \leq y\} = [1 + e^{-1.794148x}]^{-1} \tag{2.10}$$

where

$$x = \frac{\left(1 - \frac{2}{9v}\right)\left(\frac{uy}{u+\Delta}\right)^{\frac{1}{3}} - \left[1 - \frac{2(u+2\Delta)}{9(u+\Delta)^2}\right]}{\left[\frac{2(u+2\Delta)}{9(u+\Delta)^2} + \frac{2}{9v}\left(\frac{uy}{u+\Delta}\right)^{\frac{2}{3}}\right]^{\frac{1}{2}}}.$$

The maximum absolute error in the approximation in (2.10) is 0.06 for $v \geq 5$. Also $\Delta = 0$ gives the approximation in the central $F_{u,v}$ case.

Now, let $y = 10^{\frac{C}{10}}$, $a_0 = \frac{1}{1.794148} ln(\frac{1}{1-\alpha} - 1), b_0 = \frac{1}{1.794148} ln(\frac{1}{1-P^*} - 1), A = 1 - \frac{2}{9(n-p)}, B = 1 - \frac{2}{9}, E = \left(\frac{1}{1+\Delta}\right)^{\frac{1}{3}}$, and $F = \frac{2(1+2\Delta)}{9(1+\Delta)^2}$. By using the above-stated lemma in equations (2.8) and (2.9), we obtain an approximate solution to (2.8) and (2.9) from the following equations:

$$\frac{Ay^{\frac{1}{3}} - B}{[(1-B) + (1-A)y^{\frac{2}{3}}]^{\frac{1}{2}}} = a_0, \tag{2.11}$$

$$\frac{AEy^{\frac{1}{3}} - (1-F)}{[F + (1-A)E^2 y^{\frac{2}{3}}]^{\frac{1}{2}}} = b_0. \tag{2.12}$$

Squaring both sides of (2.11) and (2.12), and rearranging the terms, we get

$$[A^2 - a_0(1-A)]y^{\frac{2}{3}} - 2ABy^{\frac{1}{3}} + B^2 - a_0(1-B) = 0, \tag{2.13}$$

$$[A^2E^2 - b_0(1-A)E^2]y^{\frac{2}{3}} - 2AE(1-F)y^{\frac{1}{3}} + (1-F)^2 - b_0 F = 0. \tag{2.14}$$

We now eliminate $y^{\frac{2}{3}}$ from (2.13) and (2.14) to obtain

$$\begin{aligned} y^{\frac{1}{3}} &= \frac{[B^2 - a_0^2(1-B)][A^2E^2 - b_0^2E^2(1-A)] - [(1-F)^2 - b_0^2 F][A^2 - a_0^2(1-A)]}{2AB[A^2E^2 - b_0^2E^2(1-A)] - 2AE(1-F)[A^2 - a_0^2(1-A)]} \\ &= G, \text{say}. \end{aligned}$$

Thus $y = G^3$ and $y = 10^{\frac{C}{10}}$, yielding $C = 30 \log G$. (2.15)

Remark: While the data are used to test H_0 and make appropriate decision, the statistics SN_m as signal-to-noise ratio tell also something about the appropriate choice of the independent variables. A negative value of SN_m (or equivalently $\hat{\eta}_m < 1$) shows instability of variance in estimating β_m.

3. Regression Treatment of One-Way ANOVA Model

Consider the ANOVA model

$$Y_{ij} = \mu + \theta_i + \varepsilon_{ij} \quad j = 1, 2, \ldots, J_i; i = 1, 2, \ldots, I, \tag{3.1}$$

where $\varepsilon_{i,j}, j = 1, 2, \ldots, J_i; i = 1, 2, \ldots$ I, are independent and identically distributed normal $N(0, \sigma^2)$ random variables and $\theta_1 + \ldots + \theta_I = 0$. By letting $\beta_i = \mu + \theta_i, i = 1, \ldots, I$, model (3.1) can be written as regression model:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

where $\underline{Y}' = [Y_{11}, \ldots, Y_{1J_1}; \ldots; Y_{I1}, \ldots, Y_{IJ_I}], X = [\underline{X}_1, \ldots, \underline{X}_I], \underline{X}_i' = [0, \ldots, 0; \ldots; \underbrace{1, \ldots 1;}_{i\text{th block}} \ldots; 0, \ldots, 0], i = 1, 2, \ldots, I$, and $\underline{\beta}' = [\beta_1, \ldots, \beta_I]$.

Let $\underline{b} = (b_1, \ldots, b_I)$ where $b_i = \overline{Y}_i = \frac{1}{J_i} \sum_{j=1}^{J_i} Y_i$ is the least squares estimate β_i, $i = 1, \ldots, I$. Since $\theta_i = \beta_i - \frac{1}{I} \sum_{j=1}^{I} \beta_j$, we unbiasedly estimate θ_i by $\hat{\theta}_i = b_i - \frac{1}{I} \sum_{j=1}^{I} b_j$. It is easy to show that $E(\hat{\theta}_i) = \theta_i$, and $\mathrm{Var}(\hat{\theta}_i) = c_i\sigma^2$, where $c_i = (1 - \frac{1}{I})^2 \frac{1}{J_j} + \frac{1}{I^2} \sum_{j \neq i} \frac{1}{J_j}$, $i = 1, 2, \ldots, I$.

Let $\tau_i = \theta_i^2, i = 1, \ldots, I$, and let $\tau_{[1]} \leq \ldots \leq \tau_{[k]}$ denote the ordered τ_i. The θ_i are the treatment effects and $\tau_{[I]}$ is associated with the treatment whose effect is farthest from the average of all treatment effects which is zero. As a treatment effect goes farther from zero, it is said to become more significant. We want to test $H_0: \tau_1 = \ldots = \tau_I = 0$ (which is equivalent to $H_0: \beta_1 = \ldots = \beta_I$) at level α. When H_0 is false, a correct decision occurs if H_0 is rejected and a subset including the treatment associated with $\tau_{[I]}$ is selected. Let $\underline{\tau}' = [\tau_1, \ldots, \tau_I]$. We require that

$$Pr\{\text{ Reject } H_0 | \underline{\tau} = \underline{0}\} = \alpha \tag{3.2}$$

and

$$\text{Inf } Pr\{CD \, | \tau_{[I]} \geq \Delta\sigma^2\} = P^* \tag{3.3}$$

where $\frac{1}{I} < P^* < 1$ and $\Delta > 0$ are specified in advance.

Since $\mathrm{Var}(\hat{\theta}_i) = c_i\sigma^2$, it can be estimated by $c_i \, s^2$, where s^2 is the error mean square (MSE) in the one-way ANOVA. It is known that $\frac{\hat{\theta}_i^2}{c_i \, s^2}$ follows the noncentral $F_{1\,J-I;\lambda_i}$ distribution where $\lambda_i = \frac{\theta_i^2}{c_i\sigma^2}, i = 1, \ldots, I$, and $J = \sum_{i=1}^{I} J_i$. We define our test statistics SN_i by $SN_i = 10 \log \frac{\hat{\theta}_i^2}{c_i \, s^2}, i = 1, \ldots, I$ where the log is to base 10. Our test procedure is as follows:

Reject H_0 if $SN_i \geq C$ for some i. If H_0 is rejected, then include in the selected subset of significant treatments all those treatments for which $SN_i \geq C$.

The constant C should satisfy

$$Pr\{\underset{1\leq i\leq I}{\text{Max}} \; SN_i \geq C | \underline{\tau} = \underline{0}\} = \alpha \tag{3.4}$$

and

$$\underset{\underline{\theta}}{\text{Inf}} \; Pr\{SN_{(I)} \geq C | \tau_{[I]} \geq \Delta\sigma^2\} = P^*, \tag{3.5}$$

where $\lambda_{(I)}$ is the λ_i associated with $\tau_{[I]}$.

When $\underline{\tau} = \underline{0}$, the SN_i are i.i.d. central $F_{1,J-I}$. So, using the arguments employed in Section 2, equation (3.4) is satisfied if

$$Pr\{F_{1,J-I} \leq 10^{\frac{C}{10}}\} = 1 - \frac{\alpha}{I}. \tag{3.6}$$

Now,

$$\begin{aligned} &\underset{\underline{\theta}}{\text{Inf}} \quad \Pr\{SN_{(I)} \geq C | \tau_{[I]} \geq \Delta\sigma^2\} \\ = &\underset{\underline{\theta}}{\text{Inf}} \quad \Pr\{F_{1,J-I;\lambda_{(I)}} \geq C | \tau_{[I]} \geq \Delta\sigma^2\}, \end{aligned}$$

where $\lambda_{(I)}$ is the λ_I associated with $\tau_{(I)}$. Using the stochastically increasing property of the noncentral F in terms of the noncentrality parameter, we can see that equation (3.5) is satisfied if

$$\Pr\{F_{1,J-I;\Delta_1} \leq 10^{\frac{C}{10}}\} = 1 - P^* \tag{3.7}$$

where $\Delta_1 = \frac{\Delta}{\underset{1\leq i\leq I}{\max} c_i}$.

If we now let $\gamma = 1 - \frac{\alpha}{I}, A = 1 - \frac{2}{9(J-I)}$, and define $y, a_0, b_0, B, E,$ and F as in Section 2 with Δ_1 in the place of Δ, then an approximate solution to (3.6) and (3.7) is given by (2.11) and (2.12). Thus the solution is (2.15), namely, $C = 30 \log G$.

Remark: One can use $\frac{\underset{1\leq i\leq I}{Max} \hat{\theta}_i}{s^2}$ as a choice of Δ for a future study. When SN_i is negative, $\frac{\hat{\theta}_i^2}{c_i s^2}$ as an estimate of $[\frac{E(\hat{\theta}_i)}{\sqrt{Var(\hat{\theta}_i)}}]^2$ is less than 1; this usually means that the estimator $\hat{\theta}_i$ of θ_i is unstable.

4. An Example

We illustrate the one-way ANOVA test procedure of Section 3 using the following example of Draper and Smith (1981).

An experiment was conducted using three treatment levels, namely, 0, 100, and 200 mgs of caffeine. Thirty healthy male college students of the same age and with essentially the same physical ability were selected and trained in finger tapping. After the training was completed, ten men were randomly assigned to each treatment level. Neither the men nor the physiologist knew which treatment the men received; only the statistician knew this. Two hours after the treatment was administered, the number of finger taps per minute was recorded for each man.

Let Y_{ij} = number of finger taps per minute of the jth man on the ith treatment, μ = true value for the average number of finger taps in a population of males of which the selected thirty from a random sample, θ_i = the ith treatment effect, that is, the additive effect of the ith treatment over and above (or below) μ, where $\theta_1 + \theta_2 + \theta_3 = 0$, and ε_{ij} = the random effect which is a random deviation from $\mu + \theta_i$ taps per minute for the jth student who received the ith treatment.

With the above definitions, we have the ANOVA model:

$$Y_{ij} = \mu + \theta_i + \varepsilon_{ij}$$

and we assume that the ε_{ij} are iid $N(0, \sigma^2)$.

Now, $\underset{(1\times 10)}{\underline{0}'} = [0, \ldots, 0]$, $\underset{(1\times 10)}{\underline{1}'} = [1, \ldots, 1]$, $\underline{X'}_1 = \underline{1}', \underline{0}', \underline{0}']$, $\underline{X}_2' = [\underline{0}', \underline{1}', \underline{0}']$, $\underline{X}_3' = [\underline{0}', \underline{0}', \underline{1}']$, and $\underline{\beta}' = [\beta_1, \beta_2, \beta_3]$ where $\beta_i = \mu + \theta_i, i = 1, 2, 3$. From the data we have:

$\underset{(1\times 30)}{\underline{Y}'} = [242\ 245\ 244\ 248\ 247\ 248\ 242\ 244\ 246\ 242\ 248\ 246\ 245\ 247\ 248$
$250\ 247\ 246\ 243\ 244\ 246\ 248\ 250\ 252\ 248\ 250\ 246\ 248\ 245\ 250]$,

$\underset{(30\times 3)}{X} = [\underline{X}_1\ \underline{X}_2\ \underline{X}_3], I = 3, J_1 = J_2 = J_3 = 10.$

The regression model is: $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$.

From the output of SAS program, we obtain the following:

$$b_1 = 244.8,\ b_2 = 246.4,\ b_3 = 248.3,\ s^2 = 4.9667.$$

From the results in Section 3, we obtain

$$c_1 = c_2 = c_3 = \frac{1}{15};$$

$$\hat{\theta}_1 = -1.7, \hat{\theta}_2 = -0.1,\ \hat{\theta}_3 = 1.8, \lambda_1, = 8.728, \lambda_2 = 0.03, \lambda_3 = 9.785;$$

$$SN_1 = 9.42384,\ SN_2 = -15.815,\ SN_3 = 9.9203.$$

Suppose we have chosen $\Delta = 0.217$ and $\Delta_1 = 15\Delta = 3.25$ based on the past experience (or a preliminary sample which yielded $\frac{\underset{1 \le i \le 3}{Max} \hat{\theta}_i}{s^2} = 3.25$). Let $\alpha = 0.05$ and $P^* = 0.90$. Then $C = 3.05149$. Since $SN_1 > C$ and $SN_3 > C$, we reject $H_0 :\ \theta_1 = \theta_2 = \theta_3 = 0$ and select treatments 1 and 3 as significant (i.e. sufficiently away from the average effect). On the other hand, SN_2 is not only far less than C, it is negative. We conclude that this treatment level is not stable.

References

Bahadur, R. R. (1950). On a problem in the theory of k populations. *Ann. Math. Statist.* **21**, 362–375.

Bahadur, R. R. and H. Robbins (1950). The problem of the greater mean. *Ann. Math. Statist.* **21**, 469–487. Correction: **22** (1951), 301.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16-39.

Draper, N. and H. Smith (1981). *Applied Regression Analysis, Second Edition.* John Wiley, New York.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. and D. -Y. Huang (1981). *Multiple Decision Theory: Recent Developments.* Lecture Notes in Statistics, Vol. 6, Springer-Verlag, New York.

Gupta, S. S. and S. Panchapakesan (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* John Wiley, New York.

Huang, D. -Y. (1996). Selection procedures in linear models. *J. Statist. Planning and Inf.* **54**, 271-277.

Paulson, E. (1949). A multiple decision procedure for certain problems in analysis of variance. *Ann. Math. Statist.* **20**, 95–98.