

**PENERAPAN REGRESI LINIER MULTIVARIAT
PADA DISTRIBUSI UJIAN NASIONAL 2014
(Pada Studi Kasus Nilai Ujian Nasional 2014 SMP Negeri 1 Sayung)**

Vica Nurani¹, Sudarno², Rita Rahmawati³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

vica.nurani@yahoo.com,

ABSTRACT

National Exam is a measurement and assessment activities accession of national competency standards on specific subjects as well as a requirement that a student continue to pursue higher education. If we want to know the relationship between national exam score and semester score using multivariate linear regression analysis. Multivariate linear regression is the linear regression model with more than one response variables Y correlated and one or more predictor variables X . In the multivariate linear regression analysis, model selection is the important thing. This is because the selection of the best models in the multivariate linear regression analysis depends on the number of predictor variables involved in the model. The purpose of this study was to determine the best model in the multivariate linear regression analysis using the criteria of Mean Square Error (MSE). The result showed using MSE criterion obtained the best model with the smallest MSE value for 17424540. The best model obtained consists of six predictor variables and four response variables. The effect from response to predictor is 74,512%.

Keywords : *National Exam, Multivariate Linear Regression, MSE Criterion, Best Model.*

1. PENDAHULUAN

Ujian Nasional adalah kegiatan pengukuran dan penilaian pencapaian standard kompetensi lulusan secara nasional pada mata pelajaran tertentu dan Ujian Sekolah adalah kegiatan pengukuran dan penilaian kompetensi peserta didik yang dilakukan oleh sekolah penyelenggara program pendidikan kesetaraan untuk semua mata pelajaran. Pada jenjang SMP mata pelajaran yang diujikan dalam Ujian Nasional meliputi Mata Pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan Ilmu Pengetahuan Alam (IPA). Dalam penulisan Tugas Akhir ini, penulis ingin mengetahui apakah nilai Semester V dan VI pada mata pelajaran yang diujikan pada Ujian Nasional tersebut berhubungan dengan empat Nilai Akhir mata pelajaran tersebut.

Metode analisis yang cocok digunakan dalam penulisan ini adalah analisis regresi linier multivariat. Analisis regresi adalah metode statistika yang digunakan untuk mengetahui hubungan antara peubah respon (y) dan peubah prediktor (x). Analisis regresi linier multivariat adalah model regresi linier dengan lebih dari satu variabel respon (y) yang saling berkorelasi dan satu atau lebih variabel prediktor (x) (Johnson dan Wichern, 2007).

Pada regresi linier, besarnya variansi variabel respon yang dapat dijelaskan oleh variabel prediktor, tergantung pada banyaknya variabel yang terlibat di dalam model. Pemilihan variabel prediktor dalam analisis regresi dilakukan untuk menyeleksi variabel yang tidak signifikan dan tetap mempertahankan variabel yang signifikan secara statistik terhadap model. Dalam penelitian ini ingin diketahui model terbaik pada analisis regresi linier multivariat dengan menggunakan kriteria *Mean Square Error* (MSE).

2. TINJAUAN PUSTAKA

Analisis regresi linier multivariat adalah model regresi linier dengan lebih dari satu variabel respon (Y) yang saling berkorelasi dan satu atau lebih variabel prediktor (X) (Johnson dan Wichern, 2007). Misalkan terdapat variabel respon berjumlah p yaitu y_1, y_2, \dots, y_p dan variabel prediktor berjumlah q yaitu x_1, x_2, \dots, x_q maka model regresi linier multivariat p respon adalah sebagai berikut:

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{q1}x_q + \varepsilon_1 \\ y_2 &= \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{q2}x_q + \varepsilon_2 \\ &\vdots \\ y_p &= \beta_{0p} + \beta_{1p}x_1 + \beta_{2p}x_2 + \dots + \beta_{qp}x_q + \varepsilon_p \end{aligned} \quad (1)$$

2.1. Uji Korelasi antar Variabel Respon

Korelasi sering diukur untuk mengetahui keeratan hubungan antara masing-masing variabel. Uji korelasi antar variabel dapat digunakan untuk menguji variabel random apakah bersifat saling bebas atau tidak. Uji yang digunakan adalah uji *Bartlett of Sphericity* dengan menggunakan matriks korelasi sampel. Berikut uji *Bartlett of Sphericity*:

$$\begin{aligned} H_0 : \mathbf{R} &= \mathbf{I} && \text{(Antar variabel respon saling bebas)} \\ H_1 : \mathbf{R} &\neq \mathbf{I} && \text{(Antar variabel respon tidak saling bebas)} \end{aligned}$$

Statistik uji:

$$\chi^2_{hitung} = - \left\{ n - 1 - \frac{2p - 5}{6} \right\} \ln |\mathbf{R}|$$

Tolak H_0 jika $\chi^2_{hitung} \geq \chi^2_{\alpha; \frac{1}{2}p(p-1)}$, artinya variabel respon tidak saling bebas. (Basilevsky, 1994)

2.2. Estimasi Kuadrat Terkecil Model Regresi Multivariat

Salah satu tujuan dari analisis regresi adalah mengembangkan persamaan yang akan memprediksi respon untuk diberikan nilai-nilai variabel prediktor. Untuk mendapatkan model yang cocok dengan persamaan (1), maka harus ditentukan nilai untuk koefisien regresi (\mathbf{B}) dan varian error (σ^2) yang konsisten dengan data. Estimasi kuadrat terkecil untuk $\hat{\mathbf{B}}$ ditulis dengan persamaan:

$$\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2)$$

2.3. Uji Signifikansi Parameter

Pengujian hipotesis ini merupakan pengujian untuk seluruh koefisien regresi β_{jk} dan \mathbf{B}_1 terhadap \mathbf{Y} . Hipotesis yang digunakan adalah sebagai berikut:

$$\begin{aligned} H_0 : \mathbf{B}_1 &= \mathbf{0} \\ H_1 : \mathbf{B}_1 &\neq \mathbf{0} \end{aligned}$$

dimana \mathbf{B}_1 mencakup seluruh baris dari matriks \mathbf{B} kecuali baris pertama.

Statistik uji yang digunakan adalah Wilk's lamda:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \frac{|\mathbf{Y}^T \mathbf{Y} - \hat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y}|}{|\mathbf{Y}^T \mathbf{Y} - n \bar{\mathbf{y}} \bar{\mathbf{y}}^T|}$$

Tolak H_0 jika $\Lambda \leq \Lambda_{\alpha, p, q, (n-q-1)}$. Nilai $\Lambda_{\alpha, p, q, (n-q-1)}$ merupakan nilai kritis dari tabel Wilk's Lamda.

(Rencher, 2002)

2.4. Kriteria Pemilihan Model dengan Mean Square Error

Dalam regresi linier multivariat, beberapa variabel prediktor mungkin akan lebih berpotensi mempengaruhi variabel respon daripada variabel prediktor yang lain. Maka harus dilakukan penyederhanaan dalam model dengan menyeleksi semua model yang mungkin. Salah satu kriteria yang dapat digunakan adalah *Mean Square Error*.

$$MSE_p = S_p = \frac{E_p}{n-p} \quad (3)$$

dengan

$$E_p = \mathbf{Y}^T \mathbf{Y} - \widehat{\mathbf{B}}_p^T \mathbf{X}_p^T \mathbf{Y}$$

untuk mengubah persamaan tersebut ke dalam bentuk skalar, dapat digunakan $tr(\mathbf{S}_p)$ atau $|\mathbf{S}_p|$. Kriteria pemilihan subset adalah dengan memilih subset yang memiliki nilai minimum dari $tr(\mathbf{S}_p)$ atau $|\mathbf{S}_p|$.

(Rencher, 2002)

2.5. Uji Subset X

Uji ini dilakukan untuk mengetahui subset \mathbf{X} signifikan terhadap model atau tidak. Untuk menjelaskan hipotesis ini, \mathbf{B} dipartisi ke bentuk:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_r \\ \mathbf{B}_d \end{bmatrix}$$

Hipotesis yang digunakan ditulis sebagai berikut:

$$H_0 : \mathbf{B}_d = \mathbf{0}$$

$$H_1 : \mathbf{B}_d \neq \mathbf{0}$$

Statistik uji yang digunakan adalah:

$$\Lambda = \frac{\frac{|\mathbf{Y}^T \mathbf{Y} - \widehat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y}|}{|\mathbf{Y}^T \mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}^T|}}{\frac{|\mathbf{Y}^T \mathbf{Y} - \widehat{\mathbf{B}}_r^T \mathbf{X}_r^T \mathbf{Y}|}{|\mathbf{Y}^T \mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}^T|}} = \frac{|\mathbf{Y}^T \mathbf{Y} - \widehat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y}|}{|\mathbf{Y}^T \mathbf{Y} - \widehat{\mathbf{B}}_r^T \mathbf{X}_r^T \mathbf{Y}|} = \frac{\Lambda_f}{\Lambda_r}$$

Kriteria penolakan, H_0 ditolak jika $\Lambda \leq \Lambda_{\alpha,p,h,(n-q-1)}$ di mana h adalah banyaknya variabel prediktor yang direduksi dalam model tereduksi. Nilai $\Lambda_{\alpha,p,h,(n-q-1)}$ adalah nilai kritis dari tabel Wilk's Lamda.

(Rencher, 2002)

2.6. Hubungan antara Variabel Respon dan Prediktor

Pada regresi linier multivariat, ukuran yang digunakan dalam mengukur hubungan antara variabel respon dan variabel prediktor adalah dengan menggunakan rasio korelasi Fisher yang disarankan oleh Wilk.

$$\eta_\Lambda^2 = 1 - \Lambda$$

Nilai η_Λ^2 terletak di antara 0 dan 1, artinya semakin mendekati satu maka semakin erat hubungan antara variabel respon dengan prediktor.

(Rencher, 2002)

2.7. Asumsi Regresi Linier Multivariat

2.7.1. Uji Normal Multivariat

Data yang diambil berasal dari populasi normal jika residualnya memenuhi asumsi normal multivariat yang dinyatakan dalam $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Ada dua cara yang dapat dilakukan untuk memeriksa asumsi normal multivariat, yang pertama adalah

dengan membuat plot *Chi Square* (untuk $p \geq 2$). Hipotesis yang digunakan untuk menguji residual berdistribusi normal yaitu:

H_0 : Residual data berdistribusi normal multivariat

H_1 : Residual data tidak berdistribusi normal multivariat

Cara yang kedua adalah dengan melihat banyaknya nilai d_i^2 yang kurang dari nilai kuantil *Chi Square*.

2.7.2. Uji Kesamaan Matriks Varian Kovarian

Salah satu asumsi yang harus dipenuhi dalam melakukan pemodelan regresi multivariat adalah residual memiliki matriks varian-kovarian yang homogen. Untuk menguji syarat ini dapat dipergunakan statistik uji Box's M. Hipotesis yang digunakan adalah sebagai berikut:

H_0 : $\Sigma_1 = \Sigma_2 = \dots = \Sigma_p$ (Matriks varian kovarian residual homogen)

H_1 : Minimal ada satu $\Sigma_j \neq \Sigma_k$, untuk $j \neq k$ (Matriks varian kovarian residual tidak homogen)

Statistik uji:

$$C = (1 - u)M$$

$$= (1 - u) \left\{ \left[\sum_l (n_l - 1) \right] \ln |S_{pooled}| - \sum_l [(n_l - 1) \ln |S_l|] \right\}$$

H_0 ditolak jika $C \geq \chi^2_{\alpha; (p(p+1)(g-1))/2}$, yang artinya bahwa matriks varian kovarian tidak homogen.

(Johnson and Wichern, 2007)

2.7.3. Uji Independensi Residual

Residual $\epsilon_1, \epsilon_2, \dots, \epsilon_p$ dikatakan bersifat saling bebas (independen) jika matriks korelasi antar residual membentuk matriks identitas. Untuk menguji kebebasan antar residual ini dilakukan uji *Bartlett Sphericity*. Hipotesis yang digunakan adalah:

H_0 : $\mathbf{P} = \mathbf{I}$ (Residual bersifat saling bebas)

H_1 : $\mathbf{P} \neq \mathbf{I}$ (Residual bersifat tidak saling bebas)

Statistik uji:

$$\chi^2_{hitung} = - \left\{ n - 1 - \frac{2p - 5}{6} \right\} \ln |\mathbf{P}|$$

Tolak H_0 jika $\chi^2_{hitung} \geq \chi^2_{\alpha; \frac{1}{2}p(p-1)}$, artinya variabel respon tidak saling bebas.

(Basilevsky, 1994)

3. METODE PENELITIAN

3.1. Data

Data yang digunakan adalah data nilai Ujian Nasional (UN) 2014 dan data nilai kelas IX tahun ajaran 2013 / 2014. Data tersebut merupakan data sekunder yang diperoleh dari Daftar Kolektif Hasil Ujian Nasional dan Buku Leger Nilai SMP Negeri 1 Sayung.

3.2. Variabel Penelitian

Penggunaan variabel pada tugas akhir ini terdiri atas empat variabel respon (\mathbf{Y}) yaitu Y_1 = Nilai Akhir Bahasa Indonesia, Y_2 = Nilai Akhir Bahasa Inggris, Y_3 = Nilai Akhir Matematika dan Y_4 = Nilai Akhir IPA serta 8 variabel prediktor (\mathbf{X}) yaitu X_1 = Nilai Bahasa Indonesia Semester 1, X_2 = Nilai Bahasa Indonesia Semester 2, X_3 = Nilai Bahasa

Inggris Semester 1, X_4 = Nilai Bahasa Inggris Semester 2, X_5 = Nilai Matematika Semester 1, X_6 = Nilai Matematika Semester 2, X_7 = Nilai IPA Semester 1, dan X_8 = Nilai IPA Semester 2.

3.3. Metode Analisis

Tahapan analisis data adalah Menguji Korelasi Antar Variabel Respon, Uji Normal Multivariat, Membuat Model Lengkap Regresi Linier Multivariat, Melakukan Pemilihan Model Terbaik dengan Memilih Variabel Prediktor dengan kriteria MSE, Membuat Model Regresi Linier Multivariat, Uji Asumsi Regresi Linier Multivariat,

4. HASIL DAN PEMBAHASAN

4.1. Uji Korelasi antar Variabel Respon

Perhitungan korelasi antar variabel respon ini menggunakan *software* SPSS 18.0. Secara ringkas korelasi untuk masing-masing variabel respon disajikan pada tabel berikut:

Tabel 1. Korelasi antar Variabel Respon

Variabel Respon	Y_1	Y_2	Y_3	Y_4
Y_1	1	0,363	0,278	0,143
Y_2	0,363	1	0,402	0,222
Y_3	0,278	0,402	1	0,304
Y_4	0,143	0,222	0,304	1

Uji *Bartlett of Sphericity* dapat juga digunakan untuk mengetahui hubungan antar variabel respon secara keseluruhan dengan hipotesis:

H_0 : Antar variabel respon saling bebas

H_1 : Antar variabel respon tidak saling bebas

Hasil yang diperoleh adalah $\chi^2_{hitung} = 43,923$ lebih besar dari $\chi^2_{0,05;6} = 12,02$ sehingga dapat disimpulkan bahwa antar variabel respon tidak saling bebas atau dependen, maka data dapat digunakan pada analisis regresi linier multivariat.

4.2. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan menggunakan kriteria MSE yaitu dengan meregresikan seluruh variabel prediktor (\mathbf{X}) ke seluruh variabel respon (\mathbf{Y}) sehingga didapat 255 kombinasi.

Tabel 2. Nilai MSE untuk Setiap Model Regresi

No.	Prediktor	MSE	No.	Prediktor	MSE
1	X_1	25549340	129	$X_2X_3X_4X_6$	23883474
2	X_2	27746450	130	$X_2X_3X_4X_7$	20295175
3	X_3	33975586	131	$X_2X_3X_4X_8$	24747400
4	X_4	38053001	132	$X_2X_3X_5X_6$	23900317
5	X_5	36762806	133	$X_2X_3X_5X_7$	21388135
6	X_6	34608080	134	$X_2X_3X_5X_8$	24673130
7	X_7	27900100		:	
8	X_8	40580804		:	
9	X_1X_2	22440557	229	$X_1X_2X_4X_5X_6X_7$	17424540
10	X_1X_3	24347233		:	
11	X_1X_4	24289409	255	$X_1X_2X_3X_4X_5X_6X_7X_8$	17930081

Berdasarkan Tabel 2 didapat nilai MSE yang terkecil ada pada model dengan variabel prediktor X_1, X_2, X_4, X_5, X_6 dan X_7 . Maka dapat dikatakan bahwa model regresi linier multivariat terbaik adalah model dengan variabel prediktor X_1, X_2, X_4, X_5, X_6 dan X_7 . Sehingga model regresi linier multivariat terbaiknya adalah:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \hat{Y}_4 \end{bmatrix} = \begin{bmatrix} -36,903 & 0,495 & 0,391 & -0,108 & -0,242 & 0,124 & 0,756 \\ -5,131 & 0,478 & 0,148 & 0,689 & 0,055 & 0,039 & -0,459 \\ 16,255 & 1,085 & 0,184 & 0,583 & 0,068 & -0,168 & 1,064 \\ 54,284 & -0,186 & 0,314 & 0,112 & 0,219 & -0,702 & 0,501 \end{bmatrix} \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix}$$

4.3. Uji Kecocokan Model Terbaik

Setelah ditemukan model terbaik, dilakukan pengujian signifikansi dari model untuk mengetahui apakah model tersebut cocok atau tidak diterapkan pada data. Uji yang dilakukan adalah uji Wilk's Lamda dengan hipotesis:

$$H_0 : \mathbf{B}_1 = \mathbf{0} \text{ (Parameter tidak signifikan secara serentak terhadap model)}$$

$$H_1 : \mathbf{B}_1 \neq \mathbf{0} \text{ (Parameter signifikan secara serentak terhadap model)}$$

Hasil yang diperoleh adalah $\Lambda = 0,257848$ kurang dari $\Lambda_{0,05;4;6;257} = 0,61055$ sehingga dapat disimpulkan bahwa parameter signifikan secara serentak terhadap model.

4.4. Uji Subset X

Setelah mendapat model tereduksi (model terbaik) kemudian dilakukan pengujian variabel prediktor atau subset X apakah variabel tersebut signifikan terhadap model. Uji yang digunakan adalah uji Wilk's Lamda dengan hipotesis:

$$H_0 : \mathbf{B}_d = \mathbf{0} \text{ (Parameter tidak signifikan terhadap model)}$$

$$H_1 : \mathbf{B}_d \neq \mathbf{0} \text{ (Parameter signifikan terhadap model)}$$

Hasil yang diperoleh adalah:

Tabel 3. Nilai Wilk's Lamda untuk Setiap Variabel Prediktor

Variabel Prediktor	Nilai Wilk's Lamda	Nilai p-value
X_1	0,838	0,003
X_2	0,897	0,043
X_4	0,913	0,083
X_5	0,951	0,331
X_6	0,913	0,080
X_7	0,789	0,000

Pada Tabel 4 terlihat bahwa hanya variabel X_1, X_2 dan X_7 yang signifikan, artinya ketiga variabel tersebut berpengaruh terhadap model. Sedangkan variabel $X_4, X_5,$ dan X_6 tidak signifikan, artinya ketiga variabel tersebut tidak berpengaruh terhadap model.

4.5. Hubungan antara Variabel Respon dan Prediktor

Nilai hubungan antara variabel respon dan variabel prediktor dapat dicari dengan menggunakan nilai Wilk's Lamda dari model tereduksi. Didapat nilai sebesar 0,74215 artinya variabel respon mempengaruhi variabel prediktor sebesar 74,215% dan sisanya dipengaruhi oleh faktor lain.

4.6. Uji Asumsi Residual

Asumsi-asumsi yang diperlukan dalam Regresi Linier Multivariat adalah:

4.6.1. Residual Berdistribusi Normal Multivariat

Hipotesis distribusi normal multivariat adalah:

H_0 : Residual data berdistribusi normal multivariat

H_1 : Residual data tidak berdistribusi normal multivariat

Hasil yang diperoleh adalah $p\text{-value} = 0,5634$ lebih dari $\alpha = 0,05$ sehingga dapat disimpulkan bahwa residual berdistribusi normal multivariat.

4.6.2. Uji Homogenitas Residual

Berdasarkan hasil perhitungan dengan menggunakan *software* SPSS 21, hipotesis untuk uji asumsi independensi residual adalah:

H_0 : $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4$ (Matriks varian kovarian residual homogen)

H_1 : Minimal ada satu $\Sigma_j \neq \Sigma_k$, untuk $j \neq k$ (Matriks varian kovarian residual tidak homogen)

Hasil yang didapat adalah $C = 29,940$ kurang dari $\chi^2_{0,05;30} = 43,77$ sehingga dapat disimpulkan bahwa matriks varian kovarian homogen.

4.6.3. Uji Independensi Residual

Berdasarkan hasil perhitungan dengan menggunakan *software* SPSS 21, hipotesis untuk uji asumsi independensi residual adalah:

H_0 : $P = I$ (Residual bersifat saling bebas)

H_1 : $P \neq I$ (Residual bersifat tidak saling bebas)

Hasil yang diperoleh adalah $\chi^2_{hitung} = 10,425$ lebih besar dari $\chi^2_{0,05;6} = 12,59$ sehingga didapat kesimpulan bahwa residual bersifat tidak saling bebas.

5. KESIMPULAN

Pemilihan model terbaik dilakukan menggunakan kriteria *Mean Square Error*. Model dengan nilai MSE terkecil ditetapkan sebagai model terbaik. Hasil yang didapat adalah model dengan 6 variabel prediktor yaitu Nilai Bahasa Indonesia Semester 1, Nilai Bahasa Indonesia Semester 2, Nilai Bahasa Inggris Semester 2, Nilai Matematika Semester 1, Nilai Matematika Semester 2, Nilai IPA Semester 1. Model yang terbentuk adalah:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \hat{Y}_4 \end{bmatrix} = \begin{bmatrix} -36,903 & 0,495 & 0,391 & -0,108 & -0,242 & 0,124 & 0,756 \\ -5,131 & 0,478 & 0,148 & 0,689 & 0,055 & 0,039 & -0,459 \\ 16,255 & 1,085 & 0,184 & 0,583 & 0,068 & -0,168 & 1,064 \\ 54,284 & -0,186 & 0,314 & 0,112 & 0,219 & -0,702 & 0,501 \end{bmatrix} \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix}$$

Pengaruh variabel respon terhadap variabel prediktor adalah sebesar 50,43% dan sisanya dipengaruhi oleh faktor lain.

DAFTAR PUSTAKA

Basilevsky, A. 1994. *Statistical Factor Analysis and Related Methods, Theory and Application*. John Wiley & Sons Inc. New York.

Johnson, R.A and Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*, Sixth Edition. Prentice Hall International. New Jersey.

Rencher, A. C. 2002. *Methods of Multivariate Analysis*, Second Edition. John Wiley & Sons Inc. New York.