

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



SKRIPSI

Oleh :

VICA NURANI

24010211130033

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2015

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

Oleh :

VICA NURANI

24010211130033

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Statistika pada Jurusan Statistika

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2015

HALAMAN PENGESAHAN I

Judul : Penerapan Regresi Linier Multivariat pada Distribusi Ujian Nasional
2014 (Studi Kasus Nilai Ujian Nasional 2014 SMP Negeri 1 Sayung)

Nama : Vica Nurani

NIM : 24010211130033

Telah diujikan pada sidang Tugas Akhir tanggal 25 Juni 2015 dan dinyatakan
lulus pada tanggal 30 Juni 2015.

Semarang, 30 Juni 2015

Mengetahui,

a.n. Ketua Jurusan Statistika
Sekretaris Jurusan Statistika
FSM UNDIP

Panitia Penguji Ujian Tugas Akhir
Ketua,

Drs. Agus Rusgiyono, M.Si
NIP. 196408131990011001

Dra. Hj. Dwi Ispriyanti, M.Si
NIP. 195709141986032001

HALAMAN PENGESAHAN II

Judul : Penerapan Regresi Linier Multivariat pada Distribusi Ujian Nasional
2014 (Studi Kasus Nilai Ujian Nasional 2014 SMP Negeri 1 Sayung)

Nama : Vica Nurani

NIM : 24010211130033

Telah diujikan pada sidang Tugas Akhir tanggal 25 Juni 2015.

Semarang, 30 Juni 2015

Pembimbing I

Pembimbing II

Drs. Sudarno, M.Si

NIP. 196407091992011001

Rita Rahmawati, M.Si

NIP. 198009102005012002

KATA PENGANTAR

Puji syukur kepada Allah SWT yang telah memberikan rahmat, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan penulisan tugas akhir dengan judul **“Penerapan Regresi Linier Multivariat pada Distribusi Ujian Nasional 2014 (Studi Kasus Nilai Ujian Nasional 2014 SMP Negeri 1 Sayung)”**.

Penulis menyadari tanpa bantuan dari berbagai pihak, laporan ini tidak akan dapat diselesaikan. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Ibu Dra. Hj. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Drs. Sudarno, M.Si selaku dosen pembimbing I.
3. Ibu Rita Rahmawati, M.Si selaku dosen pembimbing II.
4. Bapak/Ibu dosen Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
5. Semua pihak yang telah membantu dalam penulisan laporan ini, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan. Sehingga saran dan kritik dari segala pihak yang bersifat membangun sangat penulis harapkan demi kesempurnaan penulisan selanjutnya.

Semarang, Juni 2015

Penulis

ABSTRAK

Ujian Nasional untuk tingkat SMP ke bawah adalah kegiatan pengukuran dan penilaian pencapaian standar kompetensi lulusan secara nasional pada mata pelajaran tertentu serta merupakan syarat agar seorang siswa dapat melanjutkan ke jenjang pendidikan yang lebih tinggi. Ingin diketahui hubungan antara nilai ujian nasional dengan nilai semester dengan menggunakan analisis regresi linier multivariat. Regresi linier multivariat adalah model regresi linier dengan lebih dari satu variabel respon Y yang saling berkorelasi dan satu atau lebih variabel prediktor X. Pada analisis regresi linier multivariat pemilihan model terbaik merupakan hal yang penting. Hal ini dikarenakan pemilihan model terbaik pada analisis regresi linier multivariat bergantung pada banyaknya variabel prediktor yang terlibat dalam model. Tujuan penelitian ini adalah mengetahui model terbaik pada analisis regresi linier multivariat dengan menggunakan kriteria *Mean Square Error* (MSE). Hasil analisis menunjukkan dengan menggunakan kriteria MSE didapat model terbaik dengan nilai MSE terkecil sebesar 17424250. Model terbaik yang didapat terdiri dari enam variabel prediktor dan empat variabel respon. Pengaruh variabel respon terhadap variabel prediktor adalah sebesar 74,215%.

Kata Kunci : Ujian Nasional, Regresi Linier Multivariat, Kriteria MSE, Model Terbaik.

ABSTRACT

National Exam for Junior High School and under 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 17424250. 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.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN I	ii
HALAMAN PENGESAHAN II	iii
KATA PENGANTAR	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR SIMBOL.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
 BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Permasalahan	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penulisan	5
 BAB II TINJAUAN PUSTAKA	
2.1 Analisis Regresi Linier Multivariat	6
2.2 Uji Korelasi antar Variabel Respon	8
2.3 Distribusi Normal Multivariat	9
2.4 Estimasi Kuadrat Terkecil Model Regresi Multivariat	10

2.4.1 Uji Signifikansi Parameter	13
2.4.2 Kriteria Pemilihan Model dengan <i>Mean Square Error</i>	14
2.4.3 Uji Subset X	15
2.5 Pengaruh Variabel Respon Terhadap Variabel Prediktor	16
2.6 Asumsi Regresi Linier Multivariat.....	17
2.6.1 Uji Normal Multivariat	17
2.6.2 Uji Kesamaan Matriks Varian Kovarian	18
2.6.3 Uji Independensi Residual	19

BAB III METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data	20
3.2 Variabel Penelitian	20
3.3 Metode Analisis	21

BAB IV HASIL DAN PEMBAHASAN

4.1 Uji Korelasi antar Variabel Respon	24
4.2 Model Awal Regresi Linier Multivariat	26
4.3 Uji Kecocokan Model Awal Regresi Linier Multivariat	27
4.4 Pemilihan Model Terbaik.....	29
4.5 Uji Kecocokan Model Terbaik	34
4.6 Uji Subset X	36
4.7 Pengaruh Variabel Respon Terhadap Variabel Prediktor	37
4.8 Uji Asumsi Residual	37
4.8.1 Residual Berdistribusi Normal Multivariat.....	38
4.8.2 Uji Homogenitas Residual	41
4.8.3 Uji Independensi Residual	41

BAB V KESIMPULAN	43
DAFTAR PUSTAKA	45
LAMPIRAN	48

DAFTAR SIMBOL

Y	: Matriks variabel respon
X	: Matriks variabel prediktor
Y_1	: Nilai Akhir Bahasa Indonesia
Y_2	: Nilai Akhir Bahasa Inggris
Y_3	: Nilai Akhir Matematika
Y_4	: Nilai Akhir IPA
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
X_8	: Nilai IPA Semester 2
n	: Banyak pengamatan
p	: Banyak variabel respon
q	: Banyak variabel prediktor
β	: Koefisien regresi
B	: Matriks koefisien regresi
ε	: Matriks residual
μ	: Vektor rata-rata
Σ	: Matriks varian kovarian

- $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: Distribusi normal multivariat
- E : Matriks jumlah kuadrat error
- MSE : *Mean Square Error*
- h : Banyak variabel prediktor pada model tereduksi
- Λ : *Lamda*, Koefisien Wilk's Lamda
- η^2 : *Eta Square*, Rasio korelasi Fisher
- d_i^2 : Jarak Mahalanobis

DAFTAR TABEL

	Halaman
Tabel 1 Korelasi antar Variabel Respon	24
Tabel 2 Nilai MSE untuk Setiap Model Regresi	30
Tabel 3 Nilai Wilk's Lamda untuk Setiap Variabel Prediktor	36
Tabel 4 Keputusan dan Kesimpulan Uji Wilk's Lamda	37
Tabel 5 Nilai Jarak Mahalanobis dan Kuantil <i>Chi-Square</i>	38

DAFTAR GAMBAR

	Halaman
Gambar 1 Alur Penyelesaian Masalah	23
Gambar 2 Plot <i>Chi-Square</i> untuk Uji Asumsi Normal Multivariat	39

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data yang Digunakan	47
Lampiran 2 Korelasi antar Variabel Respon.....	51
Lampiran 3 Nilai Wilk's Lamda Model Lengkap.....	52
Lampiran 4 Nilai MSE Seluruh Model	53
Lampiran 5 Nilai Wilk's Lamda Model Terbaik	58
Lampiran 6 Nilai Wilk's Lamda Subset X	59
Lampiran 7 Uji Normal Multivariat	60
Lampiran 8 Uji Box's M	61
Lampiran 9 Uji <i>Bartlett</i>	62

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan adalah pembelajaran pengetahuan, keterampilan, dan kebiasaan sekelompok orang yang diturunkan dari satu generasi ke generasi berikutnya melalui pengajaran, pelatihan atau penelitian. Pendidikan sering terjadi di bawah bimbingan orang lain, tetapi juga memungkinkan secara otodidak. Pendidikan di Indonesia adalah seluruh pendidikan yang diselenggarakan di Indonesia, baik itu secara terstruktur maupun tidak terstruktur. Secara terstruktur, pendidikan di Indonesia menjadi tanggung jawab Kementerian Pendidikan dan Kebudayaan Republik Indonesia (Kemdikbud), dahulu bernama Departemen Pendidikan Nasional Republik Indonesia (Depdiknas). Pendidikan umumnya dibagi menjadi tahap seperti pra sekolah, sekolah dasar, sekolah menengah dan perguruan tinggi. Di Indonesia, semua penduduk wajib mengikuti program wajib belajar pendidikan dasar selama sembilan tahun, enam tahun di sekolah dasar atau madrasah ibtidaiyah dan tiga tahun di sekolah menengah pertama atau madrasah tsanawiyah. Saat ini, pendidikan di Indonesia diatur melalui Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.

Sekolah Menengah Pertama atau yang biasa disingkat dengan SMP merupakan jenjang pendidikan dasar pada pendidikan formal di Indonesia setelah lulus sekolah dasar (atau sederajat). Sekolah Menengah Pertama ditempuh dalam waktu 3 tahun, mulai dari kelas 7 sampai kelas 9. Saat ini Sekolah Menengah Pertama menjadi program Wajar 9 Tahun (SD, SMP). Agar seorang siswa dapat

melanjutkan ke jenjang pendidikan selanjutnya, siswa tersebut harus menyelesaikan seluruh program pembelajaran, memperoleh nilai minimal baik pada penilaian akhir untuk seluruh mata pelajaran, lulus Ujian Sekolah dan Ujian Nasional.

Menurut Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 97 Tahun 2013 yang dimaksud dengan Ujian Sekolah adalah kegiatan pengukuran dan penilaian kompetensi peserta didik yang dilakukan oleh sekolah penyelenggara program pendidikan kesetaraan untuk semua mata pelajaran dan Ujian Nasional adalah kegiatan pengukuran dan penilaian pencapaian standar kompetensi lulusan secara nasional pada mata pelajaran tertentu. Nilai Akhir adalah nilai gabungan antara nilai Ujian Sekolah dengan nilai Ujian Nasional. Ujian Nasional dan Ujian Sekolah dilaksanakan oleh siswa yang menduduki tahun terakhir pada suatu jenjang pendidikan. Biasanya siswa yang telah memasuki tahun terakhir akan belajar lebih giat dan lebih keras dari tahun sebelumnya, mengingat Indonesia mempunyai nilai minimal tersendiri untuk seorang siswa agar bisa dinyatakan lulus.

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. Pada penulisan ini variabel respon yang digunakan adalah empat Nilai Akhir mata pelajaran Ujian Nasional yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, dan Ilmu Pengetahuan Alam. Penulis menduga bahwa empat

nilai tersebut saling berkorelasi satu sama lain. Oleh karena itu metode analisis yang cocok digunakan dalam penulisan ini adalah analisis regresi linier multivariat.

Menurut Sembiring (2003), analisis regresi adalah metode statistika yang digunakan untuk mengetahui hubungan antara peubah respon (y) dan peubah prediktor (x). Dalam analisis regresi dapat ditemukan model tertentu, suatu hubungan fungsional antar peubah, dengan model tersebut dapat diprediksi kelakuan sistem yang diteliti. Terdapat dua analisis regresi yaitu analisis regresi linier dan non linier. Untuk melihat hubungan antara variabel respon dan sejumlah variabel prediktor secara simultan dapat digunakan analisis regresi linier. Rencher (2002) menyatakan analisis regresi linier dapat dibagi dalam tiga kasus menurut jumlah variabelnya. Pertama, regresi linier sederhana yaitu yang melibatkan satu variabel x dan satu variabel y . Kedua, regresi linier berganda yaitu regresi yang melibatkan beberapa variabel x dan satu variabel y . Regresi linier yang ketiga adalah regresi linier multivariat.

Analisis regresi linier sederhana dan berganda terbatas pada variabel respon yang hanya berjumlah satu. Jika terdapat kasus dengan variabel respon yang lebih dari satu maka dapat digunakan analisis regresi linier multivariat. 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. Terdapat beberapa metode pemilihan variabel prediktor salah satunya adalah *All Possible Subset* (R^2 , MSE, C_p). Dalam penelitian ini ingin diketahui model terbaik pada analisis regresi linier multivariat dengan menggunakan kriteria *Mean Square Error* (MSE).

1.2 Permasalahan

Berdasarkan latar belakang yang telah diuraikan, terdapat beberapa permasalahan yang dapat diangkat antara lain bagaimana mendapat estimasi parameter agar dapat diketahui model lengkap dari data. Selain itu bagaimana melakukan pemilihan model terbaik dengan kriteria *Mean Square Error* (MSE). Jika telah mendapat model terbaik ingin diketahui juga apakah terdapat hubungan antara variabel respon dengan prediktor, lalu menguji asumsi regresi linier multivariat pada residual agar didapat model terbaik dengan asumsi yang terpenuhi.

1.3 Batasan Masalah

Penulisan tugas akhir ini dibatasi pada pemilihan 4 variabel respon yaitu 4 mata pelajaran yang diujikan pada Ujian Nasional serta pemilihan 8 variabel prediktor yaitu nilai Semester V dan VI pada mata pelajaran yang diujikan pada Ujian Nasional. Data di ambil secara acak dengan ukuran sampel 100.

1.4 Tujuan Penulisan

Tujuan dari penulisan tugas akhir ini adalah:

1. Mendapatkan model lengkap regresi linier multivariat dari data.
2. Mendapatkan model terbaik berdasarkan kriteria *Mean Square Error* (MSE).
3. Mendapatkan besarnya pengaruh variabel respon terhadap variabel prediktor.
4. Menguji asumsi regresi linier multivariat.