AUTOKORELASI DALAM REGRESI LINIER SEDERHANA

Akmal Ahadi Risza[1], Sigit Nugroho[2] dan Fachri Faisal[2]

[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRAK

Asumsi klasik dalam regresi linier sederhana mengasumsikan bahwa error (galat) tidak saling berkorelasi. Asumsi ini seringkali dilanggar ketika menggunakan data deret waktu (*time series*). Error yang berkorelasi menurut urutan waktu dikatakan saling berautokorelasi atau berkorelasi serial. Tujuan dari penelitian ini adalah untuk memberikan penjelasan mengenai pengertian autokorelasi, apa akibat yang ditimbulkan autokorelasi, bagaimana mendeteksi keberadaan autokorelasi, dan langkah-langkah remedial bagi permasalahan autokorelasi. Pada skripsi ini, permasalahan autokorelasi dapat dihilangkan dengan mencari model regresi yang lebih tepat atau memasukkan variabel yang "hilang" ke dalam model regresi. Langkah lain untuk mengatasi permasalahan autokorelasi adalah dengan menggunakan transformasi terhadap variabel. Prosedur Cochrane-Orcutt, Hidreth-Lu, Prais-Wisten, dan Perbedaan Pertama (*First Differences*) merupakan beberapa prosedur yang bergantung pada langkah ini.

Kata Kunci *: autokorelasi, regresi linier, transformasi terhadap variabel*

I. PENDAHULUAN

Regresi merupakan suatu model yang dapat digunakan untuk menggambarkan hubungan antara satu atau lebih peubah bebas (independen) dengan satu peubah takbebas (dependen) (Weisberg, 1985). Istilah regresi pertama kali diperkenalkan oleh seorang antropolog Inggris yang bernama Sir Francis Galton pada tahun 1855. Selain untuk menggambarkan hubungan antara peubah bebas dengan peubah takbebas, regresi juga dapat digunakan untuk kegiatan pengestimasian nilai peubah takbebas jika nilai peubah bebasnya diketahui.

Salah satu jenis regresi yang sering digunakan untuk menggambarkan hubungan antara satu peubah bebas dengan satu peubah takbebas dalam bentuk persamaan linier disebut regresi linier sederhana.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \ i = 1,2,...,n$$

dengan :

Y = peubah takbebas (dependen).

X = peubah bebas (independen).

ε_i = galat (error).

Bila persamaan di atas menggunakan lebih dari satu peubah bebas, maka model regresi linier tersebut disebut regresi linier berganda.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i \qquad ; i = 1,2,...,n$$

Parameter $\beta_0, \beta_1, ..., \beta_k$ pada persamaan regresi linier di atas diduga dengan $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$. Penduga parameter-parameter tersebut dapat diperoleh dengan menggunakan metode kuadrat terkecil dengan cara meminimumkan jumlah kuadrat galat.

Penggunaan metode kuadrat terkecil dalam regresi linier harus memenuhi beberapa asumsi agar diperoleh penduga yang baik. Asumsi-asumsi yang harus dipenuhi tersebut sering disebut dengan asumsi klasik. Asumsi klasik tersebut mengharuskan hubungan antara X dengan Y bersifat linier, tidak ada korelasi antara error dengan variabel independen, tidak ada korelasi diantara error pengamatan, tidak ada multikolinieritas sempurna, dan error mengikuti distribusi Normal dengan rata-rata nol dan varian yang konstan (*Homokedastisitas*) (Pindyck & Rubinfield, 1991).

Penggunaan regresi sebagai suatu teknik analisa sering kali berhadapan dengan data deret waktu (*time series*). Salah satu pelanggaran terhadap asumsi klasik yang sering terjadi ialah terjadinya korelasi diantara error menurut urutan waktu atau sering disebut dengan autokorelasi.

Permasalahan autokorelasi merupakan salah satu bentuk pelanggaran asumsi klasik yang dapat menimbulkan permasalahan yang cukup serius. Permasalahan autokorelasi akan mempengaruhi sifat-sifat yang dimiliki penduga metode kuadrat terkecil sehingga memerlukan jalan keluar yang tepat untuk mengatasi permasalahan tersebut.

Dari latar belakang permasalahan di atas, muncul ketertarikan untuk membahas lebih mendalam permasalahan autokorelasi terutama dalam regresi linier sederhana, sehingga dapat menjelaskan secara lebih luas apa yang dimaksud dengan autokorelasi, bagaimana cara untuk melihat/mendeteksi keberadaan autokorelasi, apa akibat yang ditimbulkan autokorelasi, dan bagaimana cara/jalan keluar untuk mengatasi permasalahan autokorelasi.

Tujuan dari penelitian ini adalah memberikan penjelasan mengenai pengertian autokorelasi, bagaimana cara mendeteksi keberadaan autokorelasi, apa akibat yang ditimbulkan autokorelasi, dan bagaimana cara mengatasi persoalan autokorelasi, terutama dalam regresi linier sederhana

II. TINJAUAN PUSTAKA

2.1 Autokorelasi

Menurut Gujarati (1999) autokorelasi merupakan korelasi diantara anggota seri observasi yang disusun menurut urutan waktu (dalam data *time series*) atau menurut urutan tempat/ruang (dalam data *cross section*). Autokorelasi atau juga sering disebut korelasi serial merupakan suatu bentuk pelanggaran terhadap asumsi klasik yang lebih sering/dominan terjadi ketika regresi linier sebagai sebagai suatu teknik analisa menggunakan data deret waktu (*time series*), walaupun autokorelasi juga dapat terjadi dalam data *cross section*. Autokorelasi dapat disebabkan beberapa faktor, diantaranya karena manipulasi data, tidak memasukkan peubah (variabel) yang berpengaruh, atau karena kesalahan model.

Permasalahan autokorelasi akan mengakibatkan $Cov(\varepsilon_t, \varepsilon_{t-s}) \neq 0$, sehingga terjadi korelasi antara error (galat) dari suatu waktu pengamatan dengan error dari waktu pengamatan lain. Ketika terjadi autokorelasi namun asumsi-asumsi klasik lainnya terpenuhi, maka autokorelasi akan menghasikan matriks varian-kovarian error seperti berikut :

$$V^* = E[\varepsilon\varepsilon'] = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma^2 \end{bmatrix} \tag{2.1}$$

Elemen-elemen selain diagonal utama pada matriks varian-kovarian di atas seharusnya bernilai 0, yang menandakan tidak adanya korelasi diantara error.

2.2 Autoregresif Order Pertama (*First-order Autoregressive*)

Salah satu bentuk autokorelasi sering dijumpai ialah dalam bentuk hubungan sebagai berikut :

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad (2.2)$$

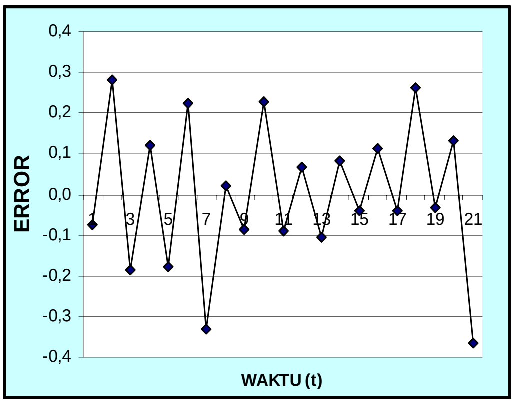
dengan :

ρ = Koefisien autokorelasi ($-1 < \rho < 1$).

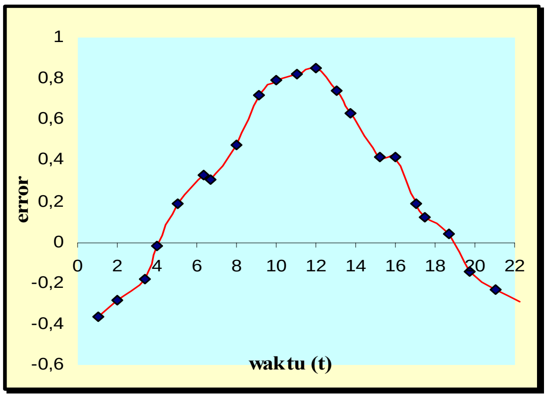
u_t = Error, dengan $u_t \sim N(0,\sigma^2)$ dan u_t saling bebas.

Persamaan (2.2) disebut dengan model autoregresif order pertama (*first-order autoregressive*), dan merupakan model yang paling populer untuk memodelkan autokorelasi atau korelasi serial (Thomas, 1996).

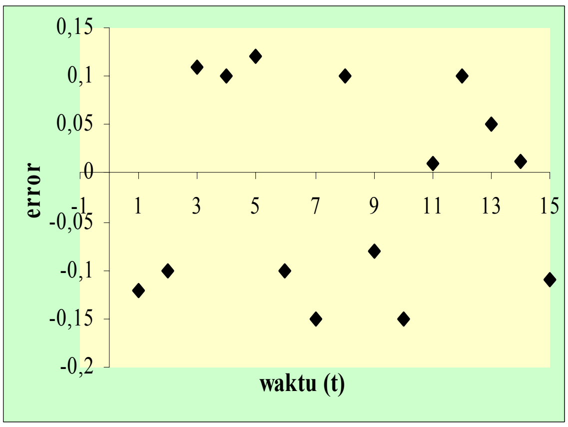
Berdasarkan model autoregresif order pertama, nilai ρ akan menentukan arah dan besarnya hubungan antara ε_{t-1} dengan ε_t. Nilai $\rho > 0$ akan mengakibatkan nilai error dari satu waktu pengamatan ke pengamatan berikutnya cenderung memiliki tanda yang sama, sehingga ketika ε_{t-1} bernilai negatif maka ε_t juga akan cenderung memiliki nilai yang negatif, dan begitu juga sebaliknya. Pola error seperti ini menandakan terjadinya autokorelasi positif. Sedangkan jika nilai $\rho < 0$, maka error akan cenderung untuk berganti-ganti tanda dari satu waktu pengamatan ke waktu pengamatan berikutnya, sehingga ketika ε_{t-1} bernilai negatif, nilai ε_t akan cenderung memiliki nilai positif, dan begitu juga sebaliknya. Pola korelasi error seperti ini sering disebut dengan autokorelasi negatif. Wujud autokorelasi positif dan negatif dapat dilihat pada gambar di bawah ini :



Gambar 2.1. Autokorelasi Negatif



Gambar 2.2. Autokorelasi Positif



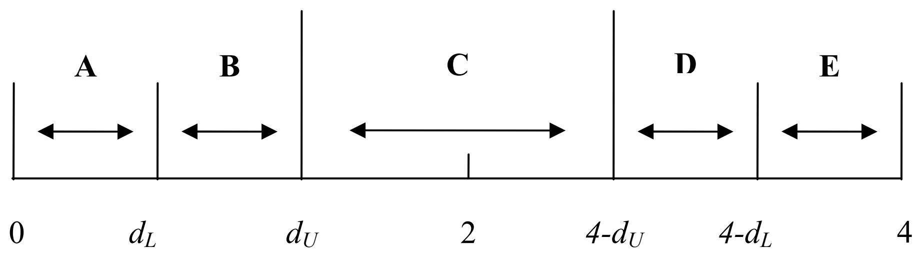
Gambar 2.3. Error yang Acak (tidak berkorelasi)

2.3 Pendeteksian Autokorelasi

Pendeteksian keberadaan autokorelasi dapat dilakukan dengan beberapa cara, diantaranya adalah dengan menggunakan metode grafik atau menggunakan uji statistik.. Metode grafik merupakan metode pendeteksi keberadaan autokorelasi yang paling sederhana yang bekerja dengan cara menggambarkan tebaran galat (error) yang diperoleh dari model regresi menurut urutan waktu/pengamatan. Bila tebaran galat (error) terhadap waktu membentuk pola tertentu, seperti pada Gambar 2.1.a dan Gambar 2.2.b, maka ada kecurigaan terjadi autokorelasi (positif atau negatif). Salah satu uji statistik yang cukup terkenal dan hampir selalu dilampirkan pada setiap keluaran paket komputer untuk mendeteksi autokorelasi ialah uji Durbin-Watson. Statistik Durbin-Watson didefinisikan sebagai berikut :

$$d = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2} \tag{2.3}$$

Nilai-nilai dari statistik Durbin-Watson mengenai kriteria ada atau tidaknya autokorelasi dapat digambarkan seperti berikut :



Gambar 2.4. Daerah Keputusan Statistik Durbin-Watson

Keterangan :
A = Daerah penolakan H_0 (ada korelasi positif).
B = D = Daerah tanpa keputusan *(inconclusive)*.
C = Daerah penerimaan H_0 (tidak ada autokorelasi).
E = Daerah penolakan H_0 (ada korelasi negatif).

III. AKIBAT DAN REMEDIAL AUTOKORELASI

Permasalahan autokorelasi khususnya autoregresif order pertama tidak mempengaruhi sifat ketakbiasan (*unbiasedness*) dan kelinieran (*linearity*). Namun menurut Gujarati (1999), penggunaan metode kuadrat terkecil tanpa memperhitungkan keberadaan autoregresif order pertama akan mengakibatkan varian $\hat{\beta}_0, \hat{\beta}_1$ yang dihasilkan menjadi bias dan cenderung *underestimated*. Pendapat senada juga diungkapkan oleh Thomas (1996) yang menyatakan bahwa autoregresif order pertama akan menyebabkan varian $\hat{\beta}_0, \hat{\beta}_1$ yang dihasilkan oleh metode kuadrat terkecil pada kondisi normal akan menjadi bias dan cenderung *underestimated* ketika terjadi autokorelasi positif $(\rho > 0)$. Oleh karena itu, permasalahan autokorelasi khususnya autoregresif order pertama akan mengakibatkan $\hat{\beta}_0, \hat{\beta}_1$ tidak lagi mempunyai varian yang minimum (terkecil). Selain itu, autokorelasi juga mengakibatkan s^2 menjadi bias dan cenderung *underestimated* terutama ketika terjadi autokorelasi positif (Thomas, 1996).

Bias dari varian $\hat{\beta}_0, \hat{\beta}_1$ dan varian error (galat) akan berdampak terhadap standar error $\hat{\beta}_0, \hat{\beta}_1$, selang (interval) kepercayaan $\hat{\beta}_0, \hat{\beta}_1$, pengujian parameter $\hat{\beta}_0, \hat{\beta}_1$, dan nilai koefisien determinasi (R^2) yang dihasilkan metode kuadrat terkecil pada kondisi normal. Perhitungan standar error yang bergantung dari nilai varian akan menyebabkan standar error juga menjadi bias dan cenderung *underestimated* ketika terjadi autokorelasi positif. Bias dari standar error akan berdampak terhadap pembentukan selang (interval) kepercayaan $\hat{\beta}_0, \hat{\beta}_1$, yang mengakibatkan selang kepercayaan akan cenderung lebih kecil dibandingkan dengan yang sesungguhnya sehingga mengurangi keakuratannya. Pengujian apakah ada pengaruh linier X terhadap Y dengan menggunakan uji-t juga akan memberikan hasil yang tidak tepat dan mengakibatkan hasil pengujian menjadi tidak valid. Nilai varian error yang bias dan cenderung *underestimated* akan mengakibatkan nilai koefisien determinasi (R^2) menjadi *overestimated*.

Langkah awal yang dapat dilakukan untuk mengatasi permasalahan autokorelasi adalah dengan menyelidiki penyebab terjadinya autokorelasi. Kadangkala dengan mengetahui penyebab terjadinya autokorelasi, maka permasalahan autokorelasi dapat dihilangkan. Jika langkah tersebut tidak membuahkan hasil maka dapat dilakukan transformasi terhadap peubah bebas dan peubah takbebas.

$$Y_t' = \beta_0' X_t' + \beta_1' X_t' + u_t \tag{3.1}$$

dengan :

$$Y_t' = Y_t - \rho Y_{t-1}$$

$$X_t' = X_t - \rho X_{t-1}$$

$$\beta_1' = \beta_1$$

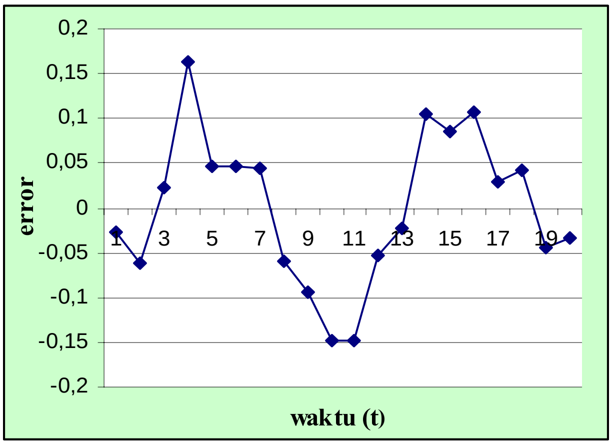
$$\beta_0' = \beta_0 (1 - \rho)$$

$$u_t = \varepsilon_t - \rho \varepsilon_{t-1}$$

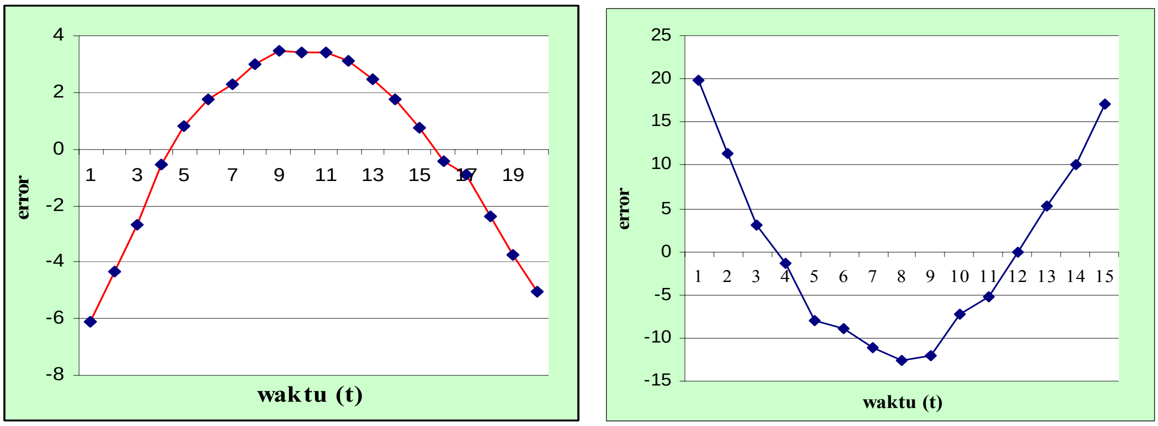
Penggunaan metode transformasi ini memerlukan informasi mengenai nilai ρ. Oleh karena itu, dapat digunakan beberapa prosedur pemilihan nilai ρ yang berdasarkan cara dan kriteria tertentu. Beberapa diantaranya adalah Prosedur Cochrane-Orcutt (1949), Hildreth-Lu (1960), Prais-Wisten (1954), dan Perbedaan Pertama (*First Differences Procedure*).

IV. STUDI KASUS

Plot error terhadap waktu merupakan salah satu rangkaian pemeriksaan error (sisa) yang bertujuan untuk melihat kecocokan suatu model regresi. Pola-pola tebaran error terhadap waktu diharapkan tidak membentuk pola-pola tertentu sehingga asumsi keindependenan galat terpenuhi. Namun pola tebaran error terhadap waktu seringkali membentuk pola-pola yang sistematis. Beberapa diantaranya adalah seperti berikut :

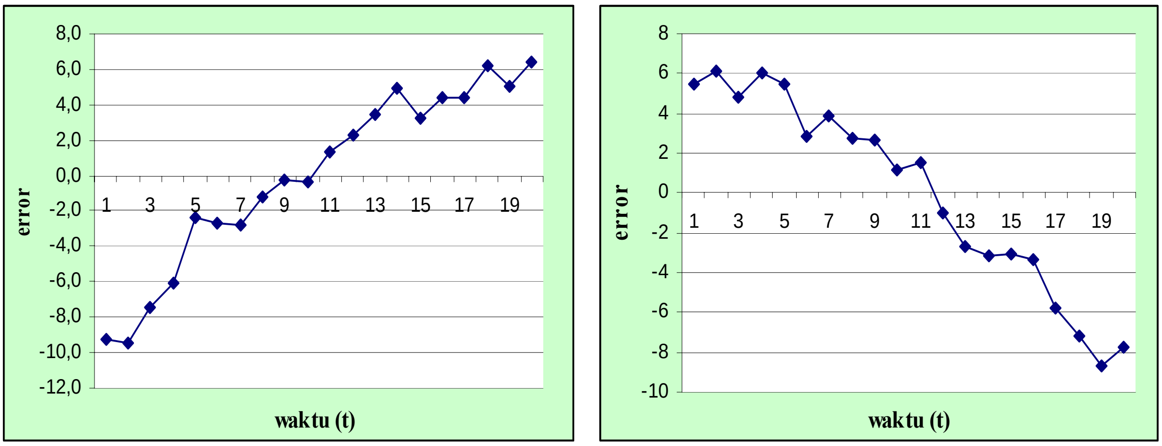


Gambar 4.1. Pola Error Seperti Siklik



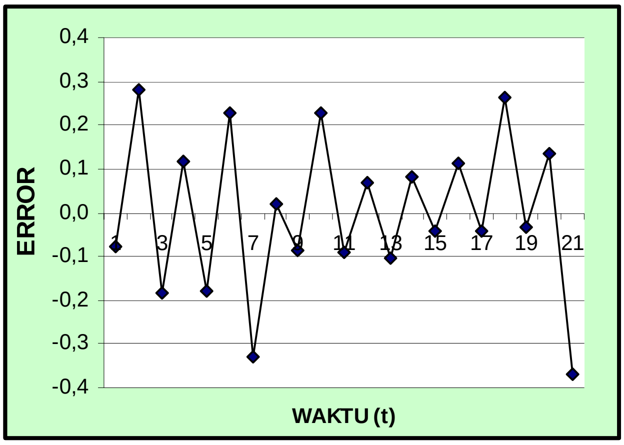
(a) **(b)**

Gambar 4.2. Pola Error Seperti Kurva



(a) **(b)**

Gambar 4.3. Pola Error Seperti Garis Lurus



Gambar 4.4. Pola Error Berganti-ganti Tanda

4.1 Langkah Remedial

Pengujian keberadaan autokorelasi dengan menggunakan uji Durbin-Watson terhadap pola-pola error pada Gambar 4.1, 4.2, dan 4.3 menunjukkan terjadinya autokorelasi positif, sedangkan terhadap Gambar 4.4 menunjukkan terjadinya autokorelasi negatif. Hasil pengujian ini bermanfaat dalam pemilihan prosedur yang tepat untuk mengatasi permasalahan autokorelasi.

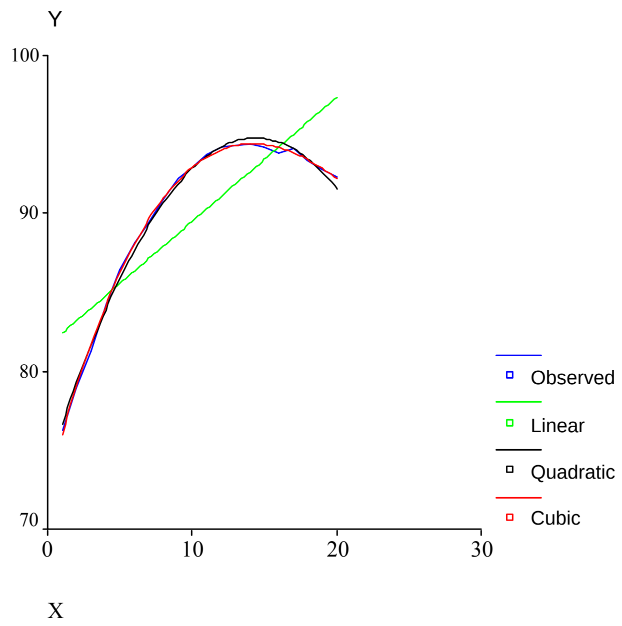
Penyebab terjadinya autokorelasi seringkali karena tidak memasukkan variabel yang berpengaruh ke dalam model regresi atau salah dalam menentukan model regresi yang digunakan. Gambar 4.2.a dan 4.2.b memperlihatkan error yang berbentuk kurva, yang menandakan bahwa model regresi yang digunakan belum tepat. Model regresi yang seharusnya digunakan ialah model non-linier. Gambar 4.3.a dan 4.3.b memperlihatkan pola error seperti garis lurus sehingga error membesar/mengecil seiring waktu. Pola error seperti ini menunjukkan bahwa adanya pengaruh waktu atau variabel lain yang belum dimasukkan ke dalam model.

Langkah remedial dengan menggunakan transformasi terhadap variabel dependen maupun independen dapat dilakukan apabila usaha memasukkan variabel yang berpengaruh atau mengganti model regresi yang digunakan tidak membuahkan hasil atau tidak dapat dilakukan. Langkah remedial untuk contoh-contoh kasus di atas adalah sebagai berikut:

4.1.1 Mengganti Model atau Memasukkan Variabel yang Berpengaruh

4.1.1.1 Pola Error Seperti Kurva

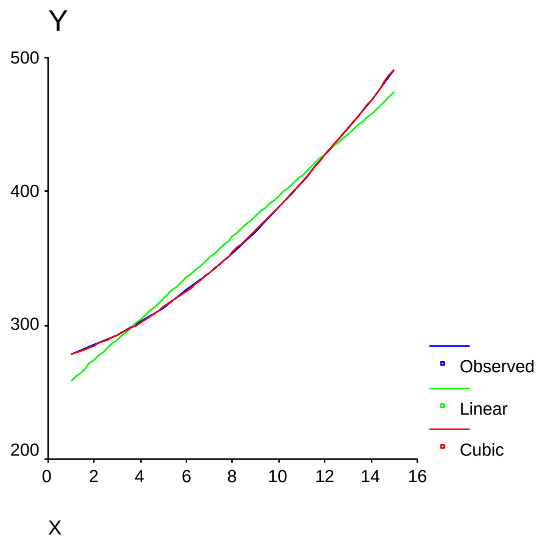
Pola error pada Gambar 4.2.a memperlihatkan terbentuknya pola error yang berbentuk kurva. Hal ini mengindikasikan bahwa model regresi linier sederhana bukanlah model yang tepat untuk digunakan. Kecocokan antara data dan beberapa model regresi dapat dilihat pada gambar berikut:



Gambar 4.5. Kecocokan Data Gambar 4.2.a dengan Beberapa Model

Gambar di atas memperlihatkan bahwa model regresi yang tepat ialah kubik. Penggunaan regresi linier sederhana mengakibatkan error membentuk pola seperti kurva. Penggantian model regresi yang digunakan menghasilkan nilai d = 1,92. Nilai batas atas (d_U) dan batas bawah (d_L) untuk $n = 20$, $k = 3$, dan taraf keberartian 1% adalah 1,41 dan 0,77. Pengujian dua arah memberikan kesimpulan bahwa H_0 tidak ditolak untuk taraf keberartian 2% karena nilai $d > d_U$ dan $4 - d = 2{,}08 > d_U$.

Pola error pada Gambar 4.2.b juga memperlihatkan pola error yang mirip dengan Gambar 4.2.a yang menandakan bahwa model yang digunakan juga belum tepat. Kecocokan antara data dan beberapa model regresi dapat dilihat pada gambar berikut:



Gambar 4.6. Kecocokan Data Gambar 4.2.b dengan Beberapa Model

Gambar di atas juga menunjukkan bahwa model regresi linier sederhana bukanlah model yang tepat dan model yang lebih tepat ialah kubik. Penggunaan regresi linier sederhana akan mengakibatkan error membentuk kurva seperti Gambar 4.2.b.

Penggantian model regresi yang digunakan menghasilkan nilai d = 2,38. Nilai batas atas (d_U) dan batas bawah (d_L) untuk n = 15, k = 3, dan taraf keberartian 1% adalah 1,46 dan 0,59. Pengujian dua arah juga memberikan kesimpulan bahwa H_0 tidak ditolak untuk taraf keberartian 2% karena nilai $d > d_U$ dan $4 - d = 1{,}62 > d_U$.

4.1.1.2 Pola Error Seperti Garis Lurus

Pola error pada Gambar 4.3.a memperlihatkan terbentuknya pola linier positif, sehingga error membesar seiring waktu. Pola seperti ini terjadi karena adanya pengaruh waktu yang belum dimasukkan ke dalam model.

Penambahan variabel waktu ke dalam model menghasilkan nilai d = 1,64. Nilai batas atas (d_U) dan batas bawah (d_L) untuk n = 20, k = 2, dan taraf keberartian 1% adalah 1,27 dan 0,86. Pengujian dua arah memberikan kesimpulan bahwa H_0 tidak ditolak untuk taraf keberartian 2 % karena nilai $d > d_U$ dan $4 - d = 2{,}36 > d_U$.

Sedangkan pola error pada Gambar 4.3.b memperlihatkan terbentuknya pola linier negatif, sehingga error mengecil seiring waktu. Pola error seperti ini juga menunjukkan adanya pengaruh variabel waktu yang belum dimasukkan ke dalam model.

Penambahan variabel waktu ke dalam model menghasilkan nilai d = 1,41. Nilai batas atas (d_U) dan batas bawah (d_L) untuk n = 20, k = 2, dan taraf keberartian 1% adalah 1,27 dan 0,86. Nilai $d > d_U$ dan $4 - d = 2{,}59 > d_U$ mengakibatkan H_0 tidak ditolak untuk taraf keberartian 2 %.

Permasalahan autokorelasi seperti Gambar 4.2 dan 4.3 ternyata dapat diatasi tanpa perlu melakukan transformasi terhadap variabel dependen maupun independen. Permasalahan autokorelasi seperti Gambar 4.2 ternyata dapat dihilangkan dengan mencari model regresi yang lebih tepat. Sedangkan pola error pada Gambar 4.3 dapat diatasi dengan memasukkan variabel yang berpengaruh ke dalam model.

4.1.2 Transformasi Variabel

Bila usaha mencari model regresi yang lebih tepat atau memasukkan variabel yang berpengaruh ke dalam model tidak berhasil, maka langkah lain yang dapat ditempuh adalah dengan melakukan transformasi terhadap variabel. Penggunaan prosedur-prosedur yang berdasarkan transformasi terhadap variabel untuk contoh kasus Gambar 4.1 dan 4.4 adalah seperti berikut :

4.1.2.1 Pola Error Seperti Siklik

Pengujian keberadaan autokorelasi terhadap pola error pada Gambar 4.1 menunjukkan terjadinya autokorelasi positif ($\rho > 0$). Korelasi antara ε_t dan ε_{t-1} dapat dilihat pada tabel berikut :

Tabel 4.1. Koefisien Korelasi Pola Error Seperti Siklik

		ε_t	ε_{t-1}
ε_t	Pearson Correlation	1	,630(**)
	Sig. (2-tailed)	.	,004
	N	19	19
ε_{t-1}	Pearson Correlation	,630(**)	1
	Sig. (2-tailed)	,004	.
	N	19	19

** Correlation is significant at the 0.01 level (2-tailed).

Tabel koefisien korelasi di atas memperlihatkan adanya korelasi positif yang cukup tinggi dan signifikan antara ε_t dan ε_{t-1}. Prosedur-prosedur yang dapat digunakan untuk mengatasi permasalahan autokorelasi positif diantaranya adalah prosedur Cochrane-Orcutt, Hildreth-Lu, Prais-Wisten, dan Perbedaan Pertama (*First Differences*).

Penggunaan prosedur Cochrane-Orcutt, Prais-Wisten, Hildreth-Lu, dan Perbedaan Pertama (*First Differences*) terhadap pola error seperti siklik (Gambar 4.1) menghasilkan output sebagai berikut :

Tabel 4.2. Penerapan Beberapa Prosedur Remedial Terhadap Pola Error Seperti Siklik

Prosedur	$\hat{\beta}_0$	$\hat{\beta}_1$	$se(\hat{\beta}_0)$	$se(\hat{\beta}_1)$	MSE (s^2)
Metode Kuadrat Terkecil	-1,455	0,176	0,214	0,0010	0,0070
Cochrane-Orcutt	-0,487	0,170	0,722	0,0045	0,0047
Hildreth-Lu	1,775	0,160	1,450	0,0070	0,0040
Prais-Wisten	-1,268	0,175	0,365	0,0024	0,0046
First Differences	-0,231	0,168	-	0,0050	0,0048

4.1.2.2 Pola Error Berganti-ganti Tanda

Pengujian keberadaan autokorelasi terhadap pola error pada Gambar 4.4 menunjukkan terjadinya autokorelasi negatif ($\rho < 0$). Korelasi antara ε_t dan ε_{t-1} dapat dilihat pada tabel berikut :

Tabel 4.3. Koefisien Korelasi Pola Error Berganti-ganti Tanda

		ε_t	ε_{t-1}
ε_t	Pearson Correlation	1	-,674(**)
	Sig. (2-tailed)	.	,001
	N	20	20
ε_{t-1}	Pearson Correlation	-,674(**)	1
	Sig. (2-tailed)	,001	.
	N	20	20

** Correlation is significant at the 0.01 level (2-tailed).

Tabel di atas memperlihatkan korelasi negatif yang cukup besar dan signifikan antara ε_t dan ε_{t-1} sehingga memerlukan langkah remedial yang tepat. Prosedur-prosedur yang dapat digunakan unutk mengatasi permasalahan autokorelasi negatif diantaranya adalah Prosedur Cochrane-Orcutt, Hildreth-Lu, dan Prais-Wisten.

Penggunaan prosedur Cochrane-Orcutt, Prais-Wisten, dan Hildreth-Lu terhadap pola error berganti-ganti tanda (Gambar 4.4) menghasilkan output sebagai berikut :

Tabel 4.4. Penerapan Beberapa Prosedur Remedial Terhadap Pola Errror Berganti-ganti Tanda

Prosedur	$\hat{\beta}_0$	$\hat{\beta}_1$	$se(\hat{\beta}_0)$	$se(\hat{\beta}_1)$	MSE (s^2)
Metode Kuadrat Terkecil	2,696	0,449	0,715	0,024	0,034
Cochrane-Orcutt	2,306	0,462	0,556	0,018	0,020
Hildreth-Lu	2,322	0,462	0,503	0,017	0,019
Prais-Wisten	2,385	0,460	0,533	0,018	0,020

Tabel 4.2 dan 4.4 memperlihatkan hasil-hasil penggunan prosedur-prosedur yang berdasarkan transformasi variabel. Prosedur-prosedur tersebut menghasilkan nilai $\hat{\beta}_1$ yang hampir sama untuk kedua contoh kasus. Nilai $\hat{\beta}_0$ yang dihasilkan juga hampir sama untuk kasus pola error yang berganti-ganti tanda, walaupun pada kasus pola error berbentuk siklik cukup berbeda. Tabel 4.2 juga menunjukkan $se(\hat{\beta}_1)$ yang dihasilkan metode kuadrat terkecil adalah yang terkecil jika dibandingkan dengan beberapa prodsedur tersebut. Hal ini membuktikan bahwa permasalahan autokorelasi positif mengakibatkan $se(\hat{\beta}_1)$ *underestimated*.

Menurut Neter *et al*. (1990), jika hasil yang diperoleh dari prosedur-prosedur remedial tidak jauh berbeda dengan metode kuadrat terkecil, maka hasil yang diperoleh dari metode kuadrat terkecil masih dapat dipergunakan. Namun jika sebaliknya, lebih baik digunakan hasil yang diperoleh dari prosedur-prosedur tersebut.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Autokorelasi atau juga sering disebut korelasi serial merupakan korelasi diantara error pengamatan menurut urutan waktu yang lebih sering/dominan terjadi pada data deret waktu (*time series*). Autokorelasi mengakibatkan penduga metode kuadrat terkecil tidak lagi mempunyai varian minimum, namun tetap merupakan penduga yang linier dan takbias. Pendeteteksian keberadaan autokorelasi dapat dilakukan dengan mengunakan metode grafik atau uji statistik. Uji statistik yang dapat digunakan diantaranya adalah uji Durbin-Watson atau Lagrange-Multiplier Test. Permasalahan autokorelasi untuk contoh kasus pola error seperti kurva (Gambar 4.2) dapat diatasi dengan menggantikan model yang digunakan. Sedangkan untuk pola error seperti garis lurus dapat dihilangkan dengan memasukkan variabel waktu ke dalam model. Penggunaan prosedur-prosedur remedial untuk contoh kasus pola error seperti kurva (Gambar 4.1) dan pola error berganti-ganti tanda (Gambar 4.4) menghasilkan output yang tidak jauh berbeda. Hasil tersebut juga tidak jauh berbeda dengan yang dihasilkan metode kuadrat terkecil.

5.2 Saran

Fokus kajian skripsi ini menitikberatkan permasalahan autoregresif order pertama yang terjadi pada model regresi linier sederhana. Penelitian lebih lanjut dapat membahas permasalahan autokorelasi dengan ruang lingkup pembahasan yang lebih luas dan dengan contoh-contoh kasus yang lebih beragam sehingga dapat menambah pengetahuan tentang permasalahan autokorelasi. Skripsi ini mudah-mudahan dapat menambah wawasan di bidang regresi, khususnya mengenai permasalahan autokorelasi.

DAFTAR PUSTAKA

Draper, N.R. dan H. Smith. 1992. *Analisis Regresi Terapan.* edisi kedua. PT Gramedia Pustaka Utama. Jakarta.

Gujarati, D. 1999. *Ekonometrika Dasar*. edisi keenam. Penerbit Erlangga. Jakarta.

Gujarati, D. 1999. *Essentials of Econometrics*. Mc Graw-Hill Companies. Singapore.

Maddala, G.S. 1989. *Introduction to Econometrics*. Mac Millan Publishing Company. Singapore.

Neter, J. *et al*. 1990. *Applied Linear Statistical Models*. 3rd ed. Richard D. Irwin Inc. Singapore.
Pindyck, R.S. and D.L. Rubinfield. 1991. *Econometrics Model & Economic Forecast*. 3rd ed. Mc Graw-Hill International Edition. Singapore.
Sembiring, R.K. 1995. *Analisis Regresi*. Penerbit ITB. Bandung.
Stewart, J. 1991. *Econometrics*. Philip Allan. England.
Studenmund, A.H. 2001. *Using Econometrics a Practical Guide*. 4th ed. Adisson Wesley Longman Inc. USA.
Supranto, J. 1984. *Ekonometrika*. Jilid 2. Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia. Jakarta.
Thomas, R.L. 1996. *Modern Econometrics An Introduction*. Addison-Wesley. England.
Weisberg, S. 1985. *Applied Linear Regression*. 2nd ed. John-Willey and Sons Inc. USA.
Wonnacott, R.J. and T.H. Wonnacott. 1981. *Regression: a Second Course in Statistics*. John Wiley and Sons, Inc. New York.

ANALISIS KONJOIN PREFERENSI MAHASISWA DALAM PEMILIHAN FLASH DISK

(Study Kasus Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Bengkulu Tahun 2004 – 2007)

Novi Susanti[1], Sigit Nugroho[2] dan Jose Rizal[2]

[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

Abstrak

Flash disk merupakan media penyimpan data yang dapat dihubungkan dengan komputer. Atribut-atribut yang terdapat pada *flash disk* dapat mempengaruhi pilihan konsumen dalam membeli *flash disk*. Mahasiswa merupakan salah satu komunitas konsumen *flash disk*, sehingga perlu diketahui preferensi mahasiswa dalam memilih *flash disk*. Penelitian ini bertujuan untuk mengetahui perkembangan *flash disk* di kalangan mahasiswa sehingga dapat membantu distributor dalam memasarkan *flash disk*. Penelitian ini menggunakan *cluster random sampling*, dimana sampel diambil dari Mahasiswa Matematika FMIPA Universitas Bengkulu yang memiliki *flash disk*. Preferensi ditentukan menggunakan analisis konjoin. Dengan metode *choice based conjoint*, atribut yang diamati adalah faktor merek, harga, kapasitas penyimpanan, fitur tambahan dan ukuran. Menggunakan bantuan SPSS versi 16.0, didapat rata-rata nilai *utility* setiap level dan rata-rata nilai *importance* dari semua atribut yang diamati. Berdasarkan nilai *Importance,* dapat diketahui bahwa pertimbangan pertama mahasiswa dalam memilih *flash disk* adalah Kapasitas penyimpanannya dengan nilai 24,29%. Kombinasi atribut yang paling disenangi mahasiswa adalah kombinasi *flash disk* dengan merek kingstone, harga murah, memiliki kapasitas 1G, dengan fitur tambahan MP3, dan mempunyai ukuran yang kecil sehingga mudah dibawa kemana-mana.

Kata kunci : *Analisis Konjoin, Choice based, importance*

Pendahuluan

Pada riset pemasaran banyak ditemukan bagaimana cara mendesain suatu produk yang banyak diminati oleh konsumen, salah satunya adalah produk *flash disk*. Sebagaimana lazimnya sebuah produk, terdapat beberapa atribut yang mempengaruhi konsumen untuk membeli *flash disk* yaitu kapasitas, ukuran, harga, fitur tambahan, dan bahan. Pengukuran dan analisis dalam penelitian pemasaran untuk memilih suatu produk biasanya dilakukan dengan menggunakan analisis konjoin.

Analisis konjoin (*Conjoint Analysis)* merupakan suatu metode analisis dalam analisis multivariat, analisis ini mulai diperkenalkan pada tahun 1970-an (Cattink and Wittink, 1992). Analisis ini digunakan untuk membantu mendapatkan atau komposisi atribut-atribut suatu produk baik baru maupun lama yang paling banyak disukai konsumen. Hasil utama konjoin adalah suatu bentuk (desain) produk barang atau jasa atau objek tertentu yang diinginkan oleh sebagian besar responden (Santoso, 2002). Menurut Hair *et al*., (1998), dalam prosesnya analisis konjoin akan memberikan ukuran kuantitatif terhadap tingkat kegunaan (*utility)* dan kepentingan relatif (*relatif importance*) suatu atribut dari produk

Terdapat beberapa ketentuan dalam memilih metode yang akan digunakan dalam analisis konjoin(Hair *et al*., 1998), yaitu :

- Choice-Based Conjoint (CBC). Digunakan apabila jumlah atribut ≤ 6
- Traditional Conjoint, digunakan apabila jumlah atribut < 10
- Adaptive Conjoint Analysis (ACA), digunakan apabila jumlah atribut ≥ 10

Merancang Kombinasi Atribut

Metode yang digunakan untuk merancang kombinasi taraf dari atribut pada penelitian ini adalah dengan menggunakan *Choice-Based Conjoint (CBC).* Metode ini mulai popular pada awal tahun 1990-an, dan saat ini banyak digunakan serta mendapat perhatian yang sangat besar oleh kalangan peneliti dan praktisi pemasaran..

Menurut Hair *et al.*, (1998) keunggulan utama *Choice-Based* dibandingkan metode lain adalah prosedur pengumpulan datanya secara langsung mencerminkan prilaku pasar. Metode *Choice-Based Conjoint,* tidak tepat digunakan dalam penelitian dengan jumlah atribut yang banyak. Green dan Srinivasan (1990), menyatakan bahwa 6-10 atribut adalah jumlah maksimum atribut yang dapat menggunakan konsep *full-profile* dalam analisis konjoin.

Metode Analisis

Konjoin termasuk dalam *Multivariate Dependence Method***,** yaitu hubungan antara variabelnya dependensi. Hubungan dependensi yaitu jika variabel-variabel yang diteliti secara teoritis dapat dipisahkan kedalam variabel-variabel respon dan variabel penjelas (Santoso, 2002) dengan model sebagai berikut :

$$Y = X_1 + X_2 + X_3 + \ldots + X_k$$

(metrik atau *non*metrik) (*non*metrik)

Variabel independen (X) adalah faktor dan berupa data *non*metrik. Termasuk disini adalah bagian dari faktor (level). Sedangkan variabel dependen (Y) adalah pendapat keseluruhan (*overall preference*) dari seorang responden terhadap sekian faktor dan level pada sebuah produk.

Secara umum model dasar analisis konjoin untuk pilihan responden (r_i) untuk setiap faktor dan level ditulis dalam bentuk :

$$r_i = \beta_0 + \sum_{j=1}^{p} u_{jk_{ji}} \tag{1}$$

Keterangan:

r_i	:	Kegunaan atau Utility total
β_0	:	Intersep model respon
$u_{jk_{ji}}$	:	Sumbangan *the part worth* atau utility dari factor ke-*j* level ke-k_{ji} .
p	:	Jumlah atribut.

fungsi kegunaan (*utility function*) yang nantinya akan diduga sangat dipengaruhi oleh model yang akan digunakan dibawah ini :

1. Faktor diskret $$\hat{u}_{jk} = \begin{cases} \hat{\alpha}_{jk} & untuk,\ k = 1,...,m_j - 1 \\ -\sum_{j=1}^{m_j-1} \hat{\alpha}_{jk}, & untuk \quad k = m_j \end{cases} \tag{2}$$
2. Faktor linier $$\hat{u}_{jk} = \hat{\beta}_j x_k \tag{3}$$
3. Faktor ideal-point $$\hat{u}_{jk} = \hat{\gamma}_{j1} z_{jk} + \hat{\gamma}_{j2} z^2{}_{jk} \tag{4}$$

Saat ini terdapat beberapa metode atau prosedur yang dapat digunakan untuk menyelesaikan model dari analisis konjoin, salah satunya adalah metode regresi dengan variabel *dummy*.

Untuk atribut ke- j dengan k_j level, variabel *dummy*nya adalah :

Tabel.1 Variabel *Dummy* Atribut ke- j dan Level k_j

Level	x_1	x_2	$\dots$	X_{k_j-1}
1	1	0	$\dots$	0
2	0	1	$\dots$	0
3	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$k_j - 1$	0	0	$\dots$	1
k_j	0	0	$\dots$	0

Langkah yang paling penting dalam analisis konjoin adalah mengestimasi kegunaan (*utility function*) atau tingkat kepentingan relatif individu (*individual level part worth*).

Untuk mendapatkan nilai-nilai $u_{jk_{ji}}$ tersebut, langkah yang harus dilakukan adalah mengestimasi model dasar analisis konjoin dengan persamaan regresi linier ganda dengan variabel *dummy*. Maka persamaan regresinya adalah :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \qquad (5)$$

Untuk menaksir parameter pada persamaan (5) maka akan digunakan metode kuadrat terkecil.

Pentingnya suatu atribut, misalnya RANGE_i dinyatakan dalam kisaran *Part Worth* melintasi level dari atribut, yaitu :

$$\text{RANGE}_i = \{\max(u_{jk_{ji}}) - \min(u_{jk_{ji}})\}, \text{ untuk setiap i} \qquad (6)$$

Selanjutnya, pentingnya suatu atribut digunakan untuk menyakinkan kepentingan relatif dengan atribut lainnya. Kepentingan relatif disimbolkan dengan IMP yang ditentukan melalui formula berikut :

$$IMP_i = \frac{RANGE}{\sum_{i=1}^{p} RANGE_i} \times 100\% \qquad (7)$$

Setelah didapatkan nilai-nilai $u_{jk_{ji}}$, maka kisaran *part worth* RANGE_i dan timbangan kepentingan relatif IMP_i akan diperoleh. Kisaran *part worth* dan timbangan kepentingan relatif ini memberikan dasar untuk menginterpretasikan hasil. Angka IMP_i yang terbesar menunjukkan preferensi terbesar terhadap level-level pada atribut tertentu.

Uji Realibilitas dan Validitas

Tahapan selanjutnya yang perlu dilakukan dalam analisis konjoin adalah menilai keandalan dan kesahihan. Jika perosedur konjoin menggunakan regresi dengan variabel dummy, maka ketepatan/kecocokan dari estimasi model digunakan nilai koefisien determinasi berganda R^2 (Supranto, 2004). Koefisien determinasi (R^2) adalah persentase keragaman variabel bebas yang dapat dijelaskan oleh model persamaan regresi. Nilai (R^2) persamaan regresi yang makin mendekati 100% menunjukkan bahwa makin banyak keragaman variabel bebas yang dapat dijelaskan dari persamaan regresi tersebut.

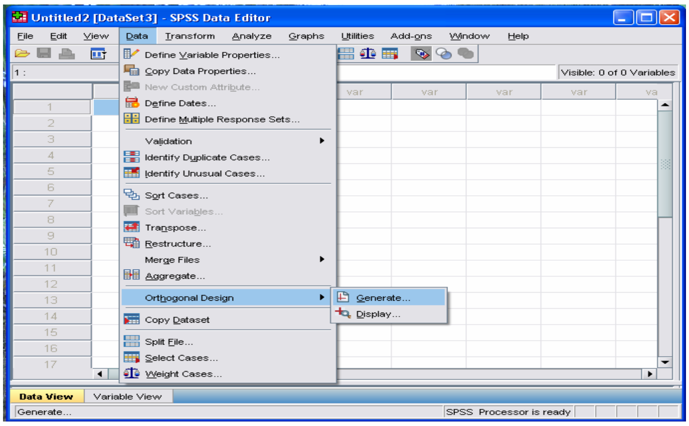
ANALISIS DATA

Analisis konjoin untuk mengetahui preferensi mahasiswa Matematika FMIPA yang menggunakan *flash disk* dilakukan dengan menggunakan Program SPSS Versi 16.0. Proses analisis konjoin dilakukan melalui tiga langkah yaitu:

1. Merancang Kartu Stimuli

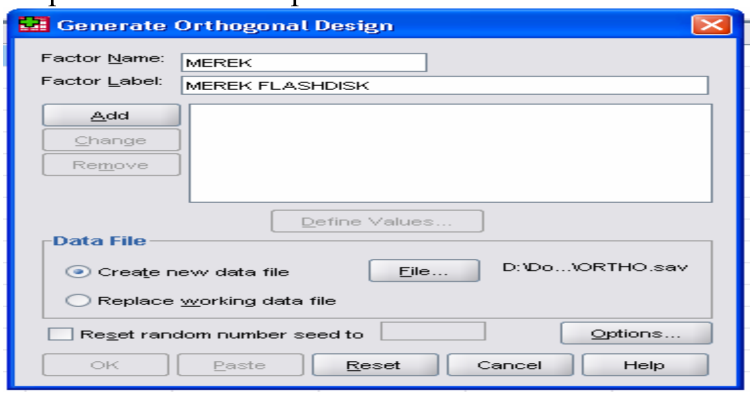
Buka Program SPSS, untuk membuat stimuli dengan orthogonal design

►Dari Menu, buka **Data** kemudian pilih **Orthogonal Design**, lalu klik **Generate...**



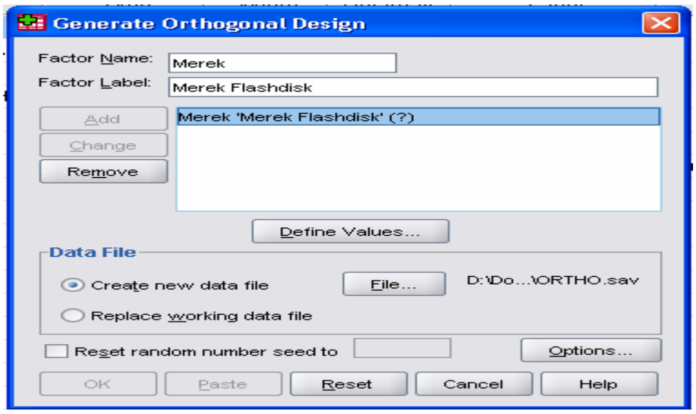
Gambar 1. Tampilan SPSS Data Editor

- Setelah itu akan tampil kotak dialog "**Generate Orthogonal Design**",
- Langkah selanjutnya setelah tampilan muncul kemudian mendeskripsikan variabel-variabel yang akan digunakan dalam analisis. Ketik variabel MEREK pada Factor Name, kemudian Merek *Flash disk* pada **Factor Label**, untuk lebih jelasnya dapat dilihat dari tampilan berikut ini :



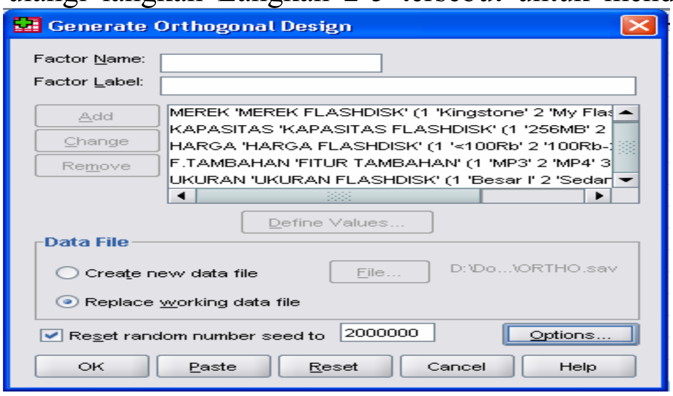
Gambar 3. Pemberian Nama Atribut Pada Kotak Dialog Generate Orthogonal Design

- Selanjutnya klik **Add**, kemudian akan tampil MEREK 'Merek Flashdisk'(?) klik item ini, dapat dilihat dari tampilan berikut ini :



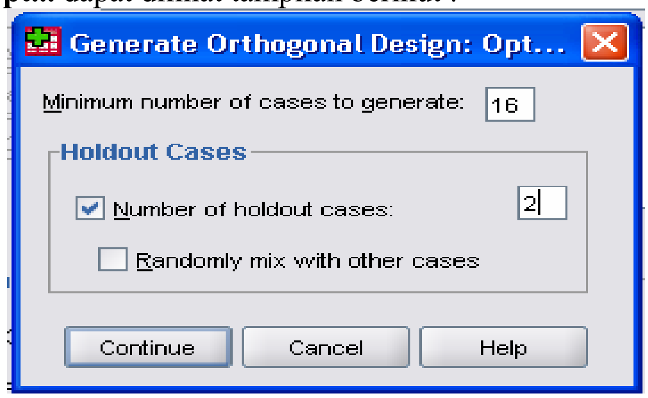
Gambar 4. Tampilan Nama Atribut Pada Kotak Dialog Generate Orthogonal Design

- Selanjutnya klik **Define Values....** Maka akan tampil kotak dialog **Generate Design Define Values**. Masukkan masing-masing nilai pada kotak **Value** dan nama pada kotak **Label** untuk setiap taraf dari faktor. Untuk faktor merek, value dan label. Kemudian klik **Continue**.
- Kemudian ulangi langkah Langkah 2-5 tersebut untuk mendefenisikan faktor



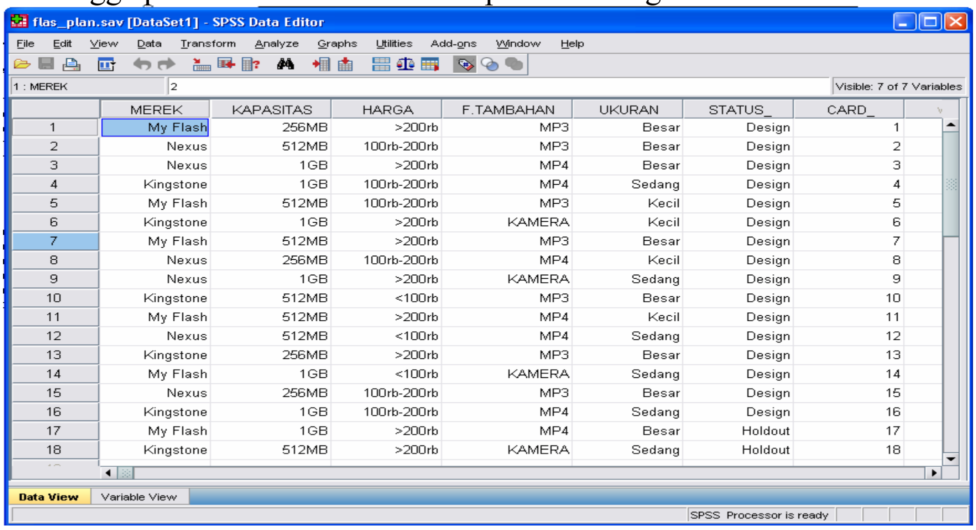
Gambar 5. Atribut Yang Sudah di Definisikan Pada Kotak Dialog

- Pilih **Replace working data file** pada kotak pilihan **Data File**
- Pilih **Reset random number seed to** dan masukkan angka **2000000.**
- Klik menu **Options**, sehingga akan tampil kotak dialog **Generate Orthogonal Design: Opt...** dapat dilihat tampilan berikut :



Gambar 6. kotak Dialog Generate Orthogonal Design: Opt...

- Langkah selanjutnya karena pada penelitian ini diinginkan adanya 18 stimuli, masukkan angka 18 pada **Minimum number of cases to generate**. Pada kotak **Holdout Cases**, pilih **Number of houldout cases** dan masukkan angka 2.
- Lalu klik **Continue**. Pada kotak dialog **Generate Orthogonal Design**, klik **OK** Sehingga pada Data Editor akan tampil data sebagai berikut :

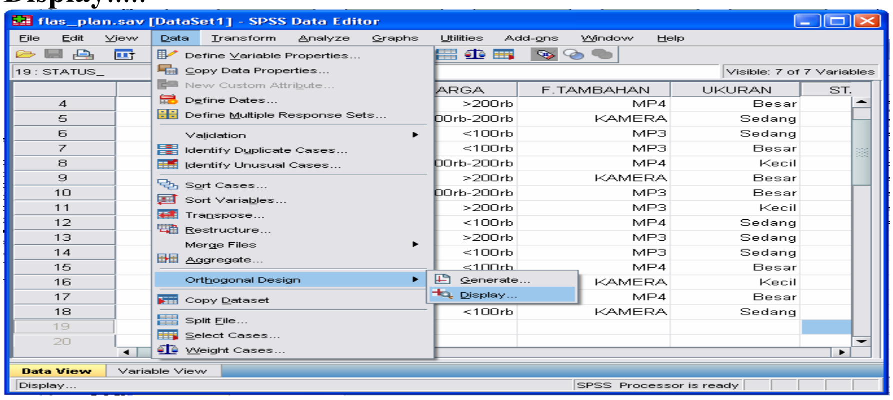


	MEREK	KAPASITAS	HARGA	F.TAMBAHAN	UKURAN	STATUS_	CARD_
1	My Flash	256MB	>200rb	MP3	Besar	Design	1
2	Nexus	512MB	100rb-200rb	MP3	Besar	Design	2
3	Nexus	1GB	>200rb	MP4	Besar	Design	3
4	Kingstone	1GB	100rb-200rb	MP4	Sedang	Design	4
5	My Flash	512MB	100rb-200rb	MP3	Kecil	Design	5
6	Kingstone	1GB	>200rb	KAMERA	Kecil	Design	6
7	My Flash	512MB	>200rb	MP3	Besar	Design	7
8	Nexus	256MB	100rb-200rb	MP4	Kecil	Design	8
9	Nexus	1GB	>200rb	KAMERA	Sedang	Design	9
10	Kingstone	512MB	<100rb	MP3	Besar	Design	10
11	My Flash	512MB	>200rb	MP4	Kecil	Design	11
12	Nexus	512MB	<100rb	MP4	Sedang	Design	12
13	Kingstone	256MB	>200rb	MP3	Besar	Design	13
14	My Flash	1GB	<100rb	KAMERA	Sedang	Design	14
15	Nexus	256MB	100rb-200rb	MP3	Besar	Design	15
16	Kingstone	1GB	100rb-200rb	MP4	Sedang	Design	16
17	My Flash	1GB	>200rb	MP4	Besar	Holdout	17
18	Kingstone	512MB	>200rb	KAMERA	Sedang	Holdout	18

Gambar 7. Rancangan Stimuli Pada Data Editor

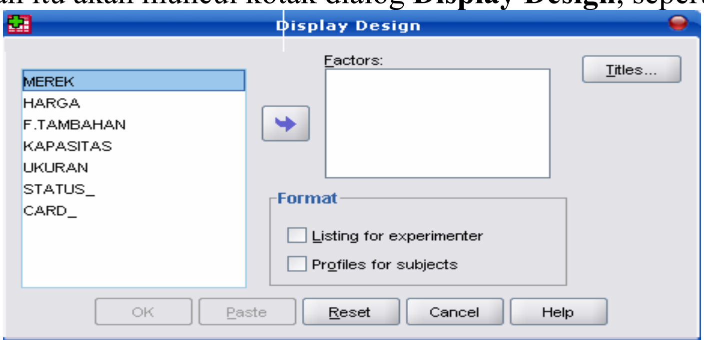
- Pilih menu **File**-> **Save** dan simpan Data Editor dengan nama file **flas_plan.sav**
- Kemudian langkah selanjutnya untuk display orthogonal design :

  ► Dari menu buka **Data** kemudian pilih **Orthogonal Design**, lalu klik **Display**.....



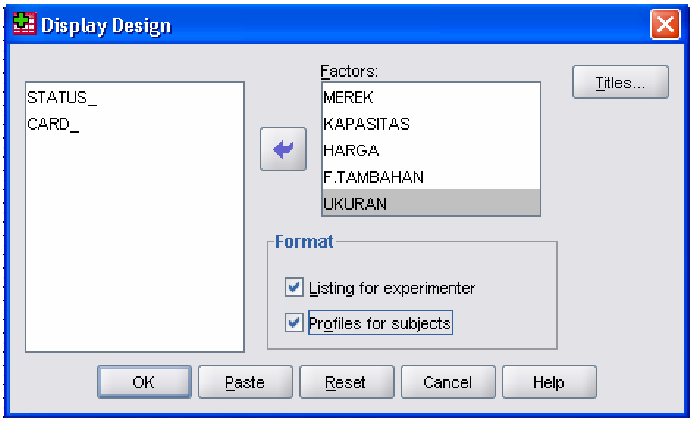
Gambar 8. Tampilan SPSS Data Editor Untuk Display

- Setelah itu akan muncul kotak dialog **Display Design**, seperti berikut :



Gambar 9. Kotak Dialog Display Design

- Masukkan variabel **Merek**, **Kapasitas**, **Harga**, **F.Tambahan** dan **Ukuran** ke dalam kotak **Factors**. Klik item variabel, kemudian klik tanda . Pada kotak **Format**, pilih **Listing for experimenter** untuk menampilkan seluruh stimuli ke dalam satu tabel dan **Profiles for subjects** untuk menampilkan setiap stimuli ke dalam tabel-tabel terpisah. Sehingga kotak dialog menjadi seperti berikut :

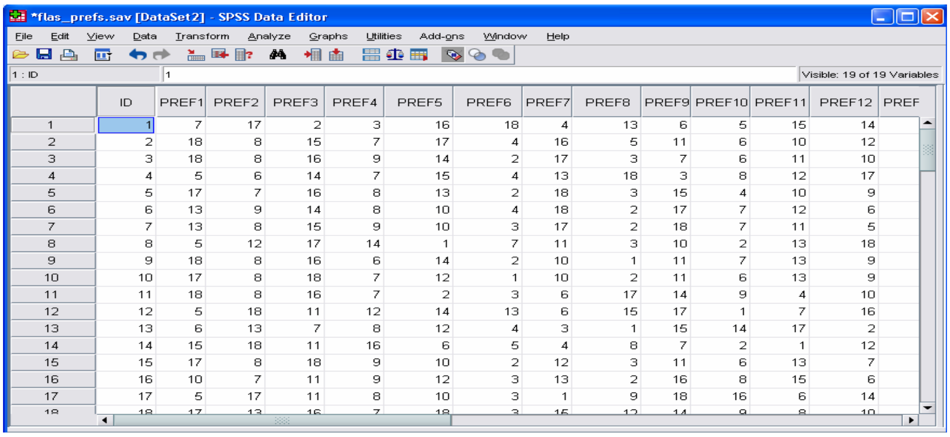


Gambar 10. Kotak Format Pada Kotak Dialog Display Design

- Klik **OK** untuk menjalankan analisis.

2. Membuat Data Preferensi

Setelah dilakukan pengumpulan data, maka data preferensi dimasukkan ke satu Data Editor dengan nama **flas_prefs.sav**. tampilan data pada program SPSS adalah sebagai berikut :



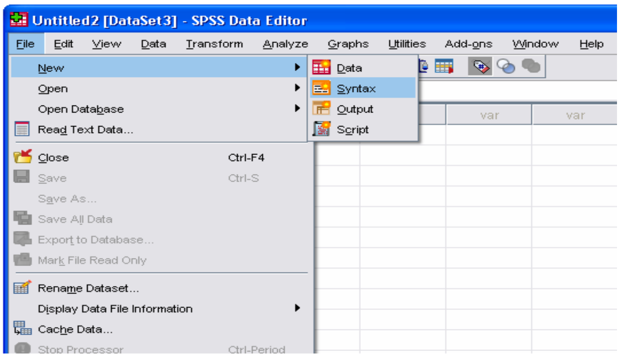
	ID	PREF1	PREF2	PREF3	PREF4	PREF5	PREF6	PREF7	PREF8	PREF9	PREF10	PREF11	PREF12
1	1	7	17	2	3	16	18	4	13	6	5	15	14
2	2	18	8	15	7	17	4	16	5	11	6	10	12
3	3	18	8	16	9	14	2	17	3	7	6	11	10
4	4	5	6	14	7	15	4	13	18	3	8	12	17
5	5	17	7	16	8	13	2	18	3	15	4	10	9
6	6	13	9	14	8	10	4	18	2	17	7	12	6
7	7	13	8	15	9	10	3	17	2	18	7	11	5
8	8	5	12	17	14	1	7	11	3	10	2	13	18
9	9	18	8	16	6	14	2	10	1	11	7	13	9
10	10	17	8	18	7	12	1	10	2	11	6	13	9
11	11	18	8	16	7	2	3	6	17	14	9	4	10
12	12	5	18	11	12	14	13	6	15	17	1	7	16
13	13	6	13	7	8	12	4	3	1	15	14	17	2
14	14	15	18	11	16	6	5	4	8	7	2	1	12
15	15	17	8	18	9	10	2	12	3	11	6	13	7
16	16	10	7	11	9	12	3	13	2	16	8	15	6
17	17	5	17	11	8	10	3	1	9	18	16	6	14

Gambar 11. Data Editor Untuk Data Preferensi Responden

3. Membuat Syntax Konjoin

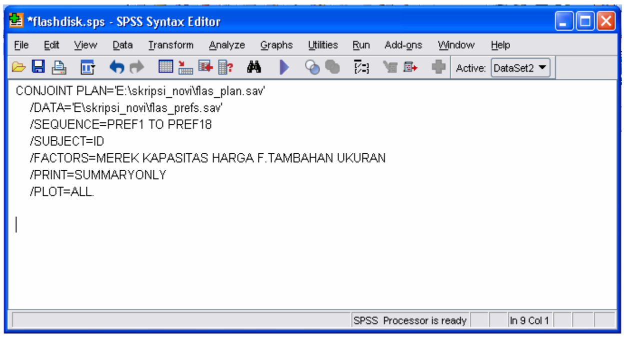
. Langkah-langkah penulisan dan eksekusi syntax dilakukan dengan cara sebagai berikut :

- Pada Data Editor yang kosong, klik **File**. Kemudian pada menu **New**, pilih **Syntax**.



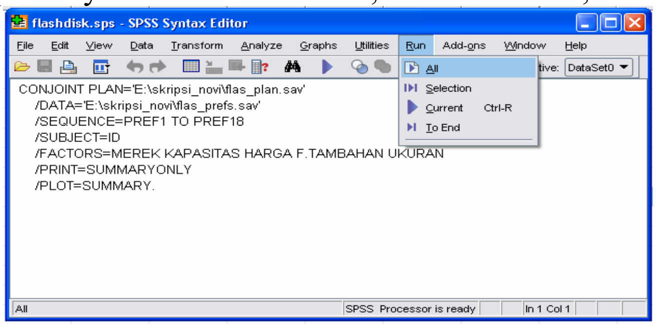
Gambar 12. Menampilkan Syntax Editor Pada SPSS 16.0

- Pada layar Syntax Editor masukkan kalimat perintah untuk melakukan analisis konjoin,:



Gambar 13. CONJOINT Command Pada Syntax Editor

- Dari tampilan Syntax Editor SPSS 16 , klik menu **Run**, kemudian klik **All**.



Gambar 14. Mengeksekusi CONJOINT Command Pada Syntax Editor

Hasil dan Interpretasi Analisis Konjoin

1. Nilai *Utility* (Kegunaan)

Nilai (*utility*) adalah nilai yang menyatakan utilitas masing-masing level dalam faktor. Apabila dalam grafik *utility* adalah positif, maka berarti responden tersebut menyukai level tersebut, dan apabila negatif berarti responden tidak menyukai level tersebut. Nilai *utility* secara umum dapat dilihat pada tabel berikut :

Tabel 2. Rata-rata Nilai *Utility* (Kegunaan) Pada Atribut *Flash Disk*.

Utility

		Utility Estimate	Std. Error
MEREK	Kingstone	.268	.403
	My Flash	.006	.278
	Nexus	-.274	.328
KAPASITAS	256MB	-.154	.429
	512MB	-.168	.279
	1GB	.322	.462
HARGA	<100rb	.011	.359
	100rb-200rb	.031	.406
	>200rb	-.043	.291
F.TAMBAHAN	MP3	.022	.249
	MP4	.002	.292
	KAMERA	-.024	.310
UKURAN	Besar	.064	.253
	Sedang	-.184	.302
	Kecil	.120	.289
(Constant)		8.482	.216

Masing–masing nilai *utility* adalah variabel x_{ij} atribut ke-i level ke-j dengan nilai konstanta $\beta_0 = 8.482$, maka model analisis konjoin untuk preferensi mahasiswa Matematika FMIPA dalam pemilihan *flash disk* adalah :

$$r_i = 0.268\, x_{11} + 0.006\, x_{12} - 0.274\, x_{13} - 0.154\, x_{21} - 0.168\, x_{22} + 0.322\, x_{23} + 0{,}011\, x_{31} + 0.031\, x_{32} - 0.043\, x_{33} + 0.022\, x_{41} + 0.002\, x_{42} - 0.024\, x_{43} + 0.064\, x_{51} - 0.168\, x_{52} + 0.12\, x_{53} + 8.482$$

2. Nilai *Importance* (Kepentingan)

Dari analisis diperoleh nilai *importance* (kepentingan), yang mana nilai tersebut merupakan gabungan pendapat responden terhadap faktor yang dimaksud. Nilai *importance* digunakan untuk mengetahui faktor mana yang dianggap terpenting oleh responden dalam memilih *flash disk*. Nilai yang tertinggi dianggap faktor yang terpenting dalam memilih suatu produk. Hasil analisis konjoin untuk nilai *importance* secara umum dapat dilihat pada tabel dibawah ini :

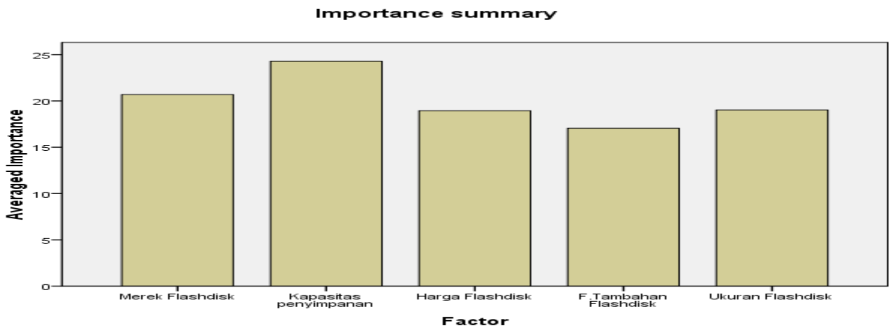
Tabel 3. Rata-rata Nilai *Importance* (Kepentingan) Atribut *Flash Disk*

Importance Values

MEREK	20.690
KAPASITAS	24.289
HARGA	18.944
F.TAMBAHAN	17.046
UKURAN	19.031

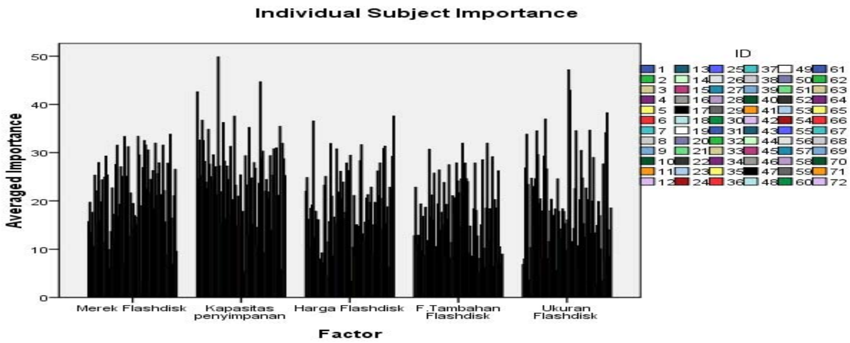
Averaged Importance Score

Nilai *importance* di atas dapat disajikan pada gambar berikut :



Gambar 15. Nilai *Importance* Atribut *Flash disk*

Sedangkan grafik nilai *importance* untuk 72 responden gambar berikut:



Gambar 16. Distribusi Penilaian Setiap Responden Terhadap Faktor

Berdasarkan tabel korelasi dapat diketahui hubungan (korelasi) antara data responden dengan data sebenarnya yang bertujuan untuk mengukur ketepatan/kecocokan estimasi model. Output untuk nilai korelasi adalah sebagai berikut:

Tabel 4. Nilai Correlation Responden Terhadap Atribut *Flash Disk*

Correlations[a]

	Value	Sig.
Pearson's R	.512	.021
Kendall's tau	.310	.048
Kendall's tau for Holdouts	1.000	.

a. Correlations between observed and estimated preferences

Pada tabel korelasi angka signifikan untuk uji Pearson's R dan Kendall's tau dibawah **0,05** maka kedua uji tersebut berada pada taraf signifikan, maka Ho ditolak. Hal ini berarti memang ada korelasi yang nyata antara hasil konjoin dengan pendapat responden tersebut. Dengan demikian bahwa pendapat 72 responden tersebut bisa diterima untuk menggambarkan keinginan populasi pembeli *flash disk*.

Kesimpulan

Berdasarkan hasil analisis konjoin rata-rata responden menganggap kapasitas lebih penting dibanding fitur tambahan dan atribut lainnya. Diketahui responden lebih senang *flash disk* dengan merek kingstone, harga murah, memiliki kapasitas 1G, dengan fitur tambahan MP3, dan mempunyai ukuran yang kecil sehingga mudah dibawah kemana-mana.

Daftar Pustaka

Cattin, P., and D. R. Wittink. 1982. Commercial use of conjoint analysis: A survey. *Journal of Marketing*, 46:3, 44–53. http://www. Sawtoothsoftware. com.

Febriyani, C. 2001. *Pengembangan Produk Dengan Analisis Konjoin*. Jakarta

Hair, JF. *et al*. 1998. *Multivariate Analysis Fifth Edition*, New Jersey: Prentice-Hall International.

Johnson, W. and Wichern, D. 1998 . *Applied Multivariat Statistical Analysis.* Prentice Hall, Englewood Cliffs, New Jassey.

Wittink, D. *et al*. 1992. *The Number Of Levels Effect In Conjoint*. www. Sawtoothsoftware. com

PERAMALAN CURAH HUJAN DI KOTA BENGKULU DENGAN METODE PEMULUSAN EKSPONENSIAL

RA Mirdaliafianti[1], Sigit Nugroho[2] dan Baki Swita[2]
[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu
[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRAK

Pengembangan sektor pariwisata, sektor pertanian, dan sektor penerbangan dipengaruhi oleh iklim dan cuaca diantaranya adalah curah hujan. Tujuan penelitian ini adalah untuk mendapatkan metode pemulusan eksponensial yang terbaik mengenai ramalan curah hujan di kota Bengkulu.

Penelitian ini menggunakan data curah hujan yang didapat dari Badan Meteorologi dan Geofisika sebagai data sekunder pada tahun 1985 sampai dengan Januari 2007.

Berdasarkan analisis metode pemulusan eksponensial diperoleh kesimpulan bahwa metode pemulusan eksponensial tunggal dan metode pemulusan eksponensial ganda metode dua parameter dari Holt lebih baik.

Kata kunci : *peramalan, pemulusan eksponensial, Holt*

PENDAHULUAN

Latar Belakang

Propinsi Bengkulu mempunyai banyak sumber daya alam yang dapat dikembangkan, antara lain sektor pertambangan, pariwisata, industri, dan pertanian. Sedangkan di kota Bengkulu, sebagai ibukota propinsi, potensi sumber daya alam yang dapat dikembangkan adalah sektor pariwisata, sektor pertanian, dan sektor penerbangan.

Pengembangan sektor-sektor tersebut dipengaruhi oleh iklim dan cuaca di kota Bengkulu diantaranya adalah curah hujan. Sehingga data curah hujan dapat dijadikan acuan oleh para pengambil kebijakan, perencana, dan pelaksana lapangan. Dengan mempertimbangkan keadaan curah hujan yang dihadapi, sehingga peramalan curah hujan sangat diperlukan.

Salah satu teknik peramalan dalam statistika adalah metode deret waktu (*Time Series*). Metode deret waktu terdiri dari beberapa bentuk, salah satunya adalah metode pemulusan (*smoothing*). Metode pemulusan ini dapat dilakukan dengan dua pendekatan yakni metode perataan (*average*) dan metode pemulusan eksponensial (*exponential smoothing*). Pemulusan eksponensial adalah teknik yang dapat memberikan ketepatan dalam ramalan jangka pendek (Arsyad, 2001). Karena curah hujan merupakan faktor penting dalam kehidupan dan data pengamatan banyaknya curah hujan dapat disajikan dalam metode pemulusan eksponensial, maka metode pemulusan ini dapat digunakan untuk meramalkan curah hujan.

Metode pemulusan eksponensial terdiri dari: (1) pemulusan eksponensial tunggal, (2) pemulusan eksponensial tunggal: pendekatan adaptif, (3) pemulusan eksponensial ganda: metode linier satu parameter dari Brown, (4) metode pemulusan eksponensial ganda: metode dua parameter dari Holt (Makridakis, 1999). Tidak ada satupun referensi yang mengatakan bahwa metode yang paling baik untuk peramalan adalah mengikuti metode tertentu, akan tetapi tergantung data yang ada. Berdasarkan alasan tersebut penulis akan membandingkan hasil peramalan metode pemulusan eksponensial yaitu pemulusan eksponensial tunggal dengan pemulusan eksponensial tunggal pendekatan adaptif dan pemulusan eksponensial ganda metode linier satu parameter dari Brown dengan metode pemulusan eksponensial ganda metode dua parameter dari Holt.

Tujuan

Tujuan penelitian ini adalah :

1. Membandingkan hasil peramalan curah hujan di kota Bengkulu antara metode pemulusan eksponensial tunggal dan metode pemulusan eksponensial tunggal pendekatan adaptif.
2. Membandingkan hasil peramalan curah hujan di kota Bengkulu antara metode pemulusan eksponensial ganda metode linier satu parameter dari Brown dan metode pemulusan eksponensial ganda metode dua parameter dari Holt.

TINJAUAN PUSTAKA

Konsep Curah Hujan

Menurut Lakitan (2002), salah satu unsur iklim yang sering dipakai oleh para pakar untuk mengklasifikasikan iklim adalah curah hujan. Pada kenyataannya, klasifikasi iklim di Indonesia seluruhnya memang dikembangkan dengan menggunakan curah hujan sebagai kriteria utamanya. Hal ini dilakukan karena keragaman (variasi) curah hujan untuk wilayah Indonesia sangat nyata.
Curah hujan adalah banyaknya hujan yang menutupi permukaan tanah jika air hujan itu tidak meresap, tidak mengalir, dan tidak menguap, yang dinyatakan dalam satuan milimeter (mm)(Mustofa, 2002).

Uji Data Deret Waktu

Uji Stasioneritas

Tujuan dari uji stasioneritas adalah untuk mendapatkan nilai rata-rata stabil dan *random error* sama dengan nol, sehingga model yang diperoleh memiliki kemampuan yang tepat dalam melakukan prediksi.

Dalam melakukan penelitian ini, uji stasioner yang digunakan adalah metode grafik. Metode grafik digunakan untuk mendeteksi apakah suatu series data stasioner atau tidak secara visual dapat dilihat plot/grafik data observasi terhadap waktu. Apabila data stasioner maka grafiknya akan mempunyai kecendrungan konstan di sekitar nilai rata-ratanya dengan amplitudo yang relatif tetap atau tidak terlihat adanya *trend* naik atau turun.

Metode Pemulusan Eksponensial

Metode pemulusan eksponensial adalah suatu prosedur yang mengulang perhitungan secara terus menerus dengan menggunakan data terbaru. Metode ini didasarkan pada perhitungan rata-rata data-data masa lalu secara eksponensial (Lincolin, 1995 *dalam* Anonim 1, 2007).

Metode Pemulusan Eksponensial Tunggal

Metode pemulusan eksponensial tunggal dikembangkan dengan sebuah persamaan awal, yaitu:

$$F_{t+1} = \alpha X_t + (1-\alpha) F_t$$

Metode Pemulusan Eksponensial Tunggal: Pendekatan Adaptif

Persamaan awal untuk peramalan dengan metode pemulusan eksponensial tunggal dengan tingkat respon adaptif serupa dengan persamaan diatas kecuali bahwa nilai α diganti dengan α_t.

di mana :

$$F_{t+1} = \alpha_t X_t + (1-\alpha_t) F_t$$

$$\alpha_{t+1} = \pm \left| \frac{E_t}{M_t} \right|$$

$$E_t = \beta e_t + (1-\beta) E_{t-1}$$

$$M_t = \beta |e_t| + (1-\beta) M_{t-1}$$

$$e_t = X_t - F_t$$

Keterangan :

F_{t+1}	:	Nilai ramalan pada periode ke-$t+1$
X_t	:	Data aktual pada periode ke-t
E_t	:	Nilai kesalahan yang dihaluskan
M_t	:	Nilai kesalahan absolut yang dihaluskan
e_t	:	Nilai kesalahan

Dengan α dan β merupakan parameter antara 0 dan 1.

Metode Pemulusan eksponensial Ganda: Metode Linier Satu Parameter dari Brown

Menurut Assauri (1984) *dalam* anonim 1 (2007), dasar pemikiran dari metode eksponensial tunggal maupun ganda adalah bahwa nilai pemulusan akan terdapat pada waktu sebelum data sebenarnya apabila pada data tersebut terdapat komponen trend. Oleh karena itu untuk nilai-nilai pemulusan tunggal perlu ditambahkan nilai pemulusan ganda guna menyesuaikan trend. Adapun formula yang digunakan adalah sebagai berikut:

Pemulusan tunggal : $S'_t = \alpha X_t + (1-\alpha) S'_{t-1}$

Pemulusan ganda : $S''_t = \alpha S'_t + (1-\alpha) S''_{t-1}$

Pemulusan trend : $a_t = S'_t + (S'_t - S''_t) = 2S'_t - S''_t$

$$b_t = \frac{\alpha}{1-\alpha}(S'_t - S''_t)$$

Peramalan : $F_{t+m} = a_t + b_t m$

Keterangan :

S'_t : Nilai pemulusan eksponensial Tunggal
S''_t : Nilai pemulusan eksponensial ganda
b_t : Pemulusan trend
F_{t+m} : Nilai ramalan pada periode m
X_t : Nilai X aktual pada periode *ke-* t
α : Konstanta dengan nilai $0 < \alpha < 1$

Metode Pemulusan Eksponensial Ganda: Metode Dua-Parameter dari Holt

Metode pemulusan ganda lain yang dapat digunakan untuk menangani trend yang linier adalah metode dua parameter dari Holt. Pada metode Holt nilai trend tidak dimuluskan dengan pemulusan ganda secara langsung, tetapi proses pemulusan trend dilakukan dengan menggunakan parameter yang berbeda dengan parameter yang digunakan pada pemulusan data asli. Pada metode Brown hanya terdapat satu parameter saja dan estimasi nilai trend masih sangat sensitif sekali terhadap fluktasi random. Metode Holt memberikan banyak kefleksibelan dalam menyeleksi komponen trend. Metode Holt secara matematis ditulis pada tiga persamaan berikut:

Pemulusan total : $S_t = \alpha X_t + (1-\alpha)(S_{t-1} + b_{t-1})$

Pemulusan trend : $b_t = \gamma + (S_t - S_{t-1}) + (1-\gamma) b_{t-1}$

Peramalan : $F_{t+m} = S_t + b_t m$

Keterangan :

S_t	:	Nilai pemulusan eksponensial
m	:	Periode yang diramalkan
b_t	:	Pemulusan trend
F_{t+m}	:	Nilai ramalan pada periode m
X_t	:	Nilai X aktual pada periode *ke*- t
α	:	Konstanta pemulusan dengan nilai $0 < \alpha < 1$
γ	:	Konstanta pemulusan untuk pemulusan trend $0 < \gamma < 1$

Ukuran Ketepatan Ramalan

Jika X_i merupakan data aktual untuk periode i dan F_i merupakan ramalan atau nilai kecocokkan untuk periode yang sama, maka kesalahan didefinisikan sebagai:

$$e_i = X_i - F_i$$

Jika terdapat nilai pengamatan dan ramalan untuk n periode waktu, maka akan terdapat n buah kesalahan. Bagi pemakai ramalan, ketepatan dalam peramalan adalah hal yang paling penting. Untuk mengevaluasi kesalahan peramalan dapat menggunakan *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *Mean Absolute Error* (MAE). Kriteria ukuran ketepatan ramalan dapat didefinisikan sebagai berikut:

Nilai Tengah Kesalahan Absolut (*Mean Absolute Error*)

$$MAE = \sqrt{\frac{\sum_{i=1}^{n} |X_i - F_i|}{n}}$$

Nilai tengah kesalahan kuadrat (*Root Mean Squared Error*)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - F_i)^2}{n}}$$

Nilai Tengah Kesalahan Persentase Absolut (*Mean Absolute Percentage Error*)

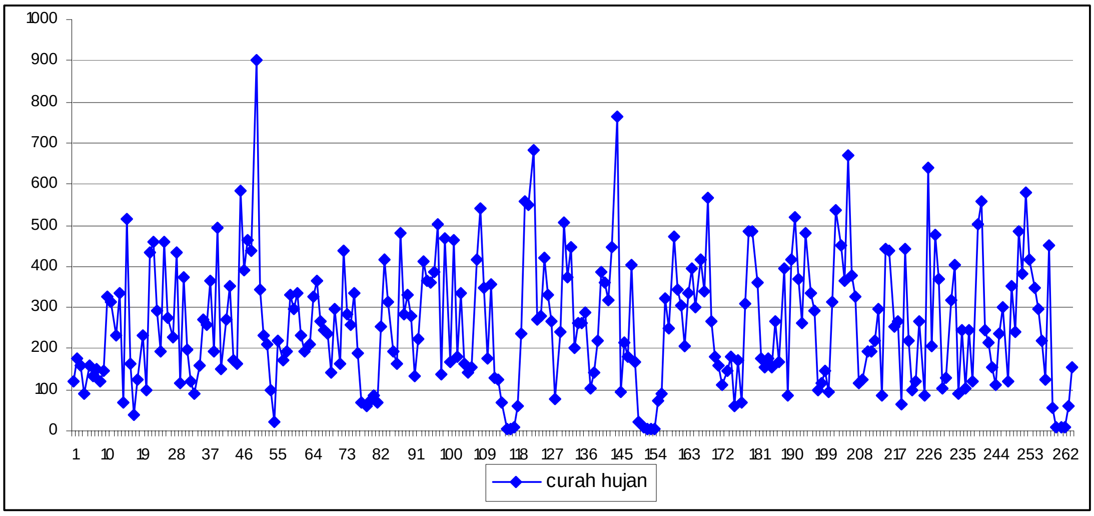
$$MAPE = \sqrt{\frac{100\sum_{i=1}^{n}\frac{|X_i - F_i|}{X_i}}{n}}$$

Model terbaik adalah model yang memiliki nilai statistik paling minimum, yang berarti penyimpangan nilai peramalan dengan nilai aktualnya paling kecil.

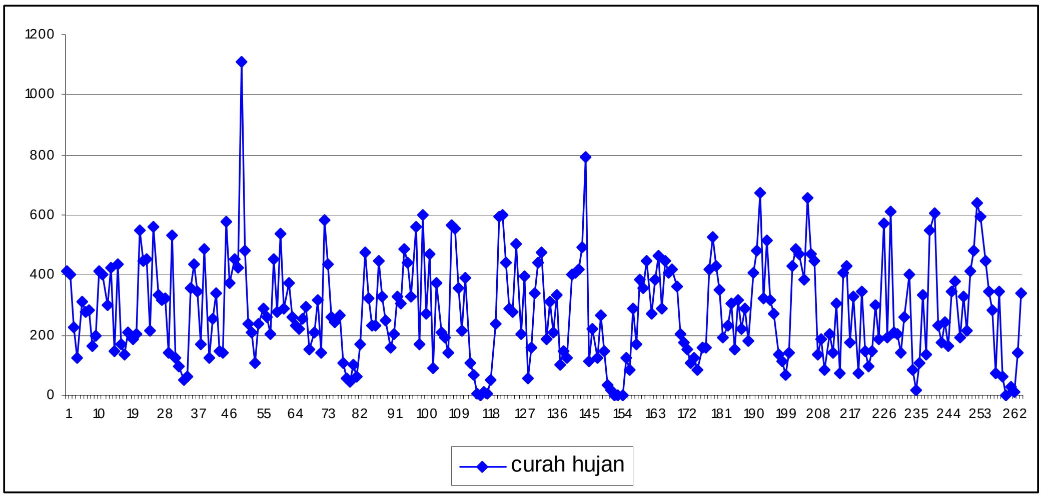
PEMBAHASAN

Uji Stasioneritas

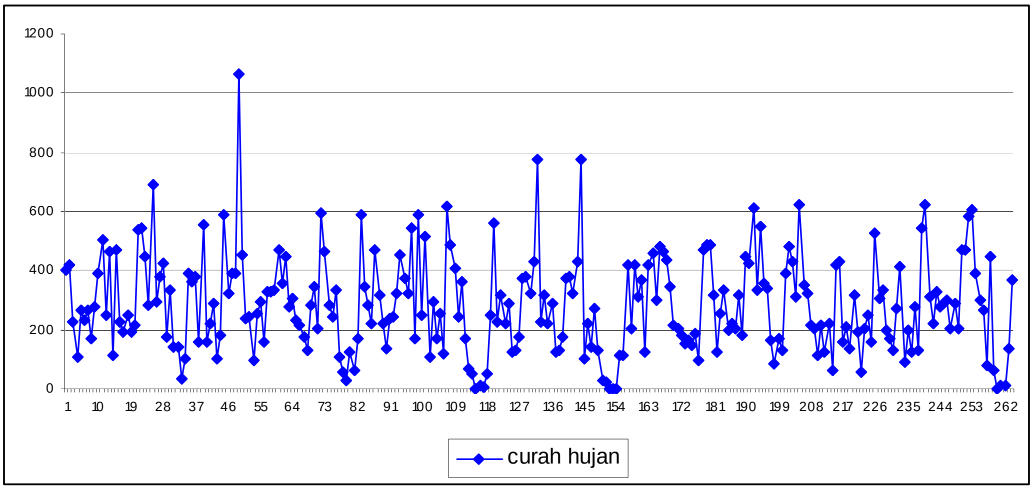
Data banyaknya curah hujan di kota Bengkulu yang diukur bulanan dari bulan Januari 1985 sampai dengan Januari 2007 diperoleh dari Badan Meteorologi dan Geofisika (BMG) Pulau Baai Bengkulu, dituangkan dalam Lampiran 2. Proses eksekusi data secara keseluruhan dilakukan dengan menggunakan *software microsoft excel*. Dengan *microsoft excel* dapat dihasilkan grafik untuk data orisinil yang dituangkan dalam Gambar 1, Gambar 2,
dan Gambar 3 sebagai gambaran kestasioneran data.



Gambar 1. Grafik Series Data Curah Hujan Pos Padang Harapan



Gambar 2. Grafik Series Data Curah Hujan Pos Pulau Baai



Gambar 3. Grafik Series Data Curah Hujan Pos Padang Kemiling

Dari Gambar 1, Gambar 2, dan Gambar 3 dapat dilihat bahwa data stasioner karena grafiknya mempunyai kecenderungan konstan disekitar nilai rata-ratanya dengan amplitudo yang relatif tetap atau tidak terlihat adanya trend naik atau turun. Menurut Makridakis (1999), jika datanya stasioner maka metode pemulusan eksponensial tunggal merupakan pendekatan yang cukup baik.

Metode Pemulusan Eksponensial Tunggal

Tujuan dari analisis pemulusan eksponensial tunggal pada penelitian ini adalah untuk mendapatkan nilai ramalan ke depan. Rumus nilai ramalan metode pemulusan eksponensial tunggal adalah $F_{t+1} = \alpha X_t + (1-\alpha)F_t$, sehingga peramalan yang dilakukan hanyalah untuk meramalkan satu periode saja yaitu bulan Januari 2007. Untuk meramalkan curah hujan di kota Bengkulu Januari tahun 2007 digunakan konstanta pemulusan α = 0.1, α = 0.2, α = 0.3. Kemudian akan dilihat seberapa besar tingkat ketepatan ramalan pemulusan eksponensial dengan menggunakan ukuran ketepatan ramalan RMSE, MAE, dan MAPE yang diperoleh dengan menggunakan rumus (19), (20), (21) yang hasilnya dapat dilihat pada Tabel 1 berikut.

Tabel 1. Ukuran Ketepatan Ramalan Januari 1985 - Desember 2006

POS	RMSE	MAE	MAPE
Padang Harapan			
A = 0.1	163.6965139	11.54416362	15.47486539
A = 0.2	167.4606852	11.68015323	14.78661861
A = 0.3	169.6940131	11.71368058	14.00066043
Pulau Baai			
A = 0.1	179.2551609	12.19724402	25.55191786
A = 0.2	183.4238002	12.26980429	23.74839084
A = 0.3	185.5282077	12.26436386	21.80976394
Padang Kemiling			
A = 0.1	172.0590373	11.80870495	25.20924829
A = 0.2	175.9714562	11.96119417	23.42054099
A = 0.3	178.2595901	12.01548787	21.57876117

Dari Tabel 1 diatas dapat disimpulkan bahwa dari ketiga pos pengamatan nilai RMSE dan MAE yang minimum diperoleh dengan menggunakan konstanta pemulusan α = 0.1. Sedangkan nilai MAPE yang minimum diperoleh pada konstanta pemulusan α = 0.3. bandingan Data dengan Forecasting Metode Pemulusan

Metode Pemulusan Eksponensial Tunggal Pendekatan Adaptif

Nilai untuk periode mendatang (*t+1*) didapat dari jumlah nilai peramalan saat ini (saat *t*) dengan hasil

perkalian antara parameter pemulusan (α_t) yang dapat berubah secara otomatis bilamana terdapat perubahan dalam pola data dasar, yang akan memberikan nilai minimum MSE.

Dapat dilihat dari Tabel 2 nilai MAPE dan MAE yang minimum diperoleh jika digunakan konstanta pemulusan α = 0.1 & β = 0.3. Sedangkan nilai RMSE yang minimum diperoleh jika digunakan konstanta pemulusan α = 0.1 & β = 0.1.

Berdasarkan data periode Januari 1985 sampai dengan Desember 2006, maka dapat diperoleh nilai ramalan pada periode Januari 2007, dan besar selisih nilai ramalan dengan data aslinya seperti yang terdapat pada Tabel 5.

Tabel 5. Nilai Ramalan Periode Januari 2007

POS	Pemulusan Eksponensial Tunggal		
	Xt = Data	Ft = Ramalan	Xt - Ft
Padang Harapan			
α = 0.1	234	194.4634265	39.53657349
α = 0.2	234	124.3684659	109.6315341
α = 0.3	234	79.07554178	154.9244582
Pulau Baai			
α = 0.1	570	219.4701700	350.5298300
α = 0.2	570	149.1911574	420.8088426
α = 0.3	570	107.5610198	462.4389802
Padang Kemiling			
α = 0.1	510	217.9691988	292.0308012
α = 0.2	510	146.8148389	363.1851611
α = 0.3	510	104.3838509	405.6161491
	Pemulusan Eksponensial Tunggal Adaptif		
Padang Harapan	Xt = Data	Ft = Ramalan	Xt – Ft
α = 0.1 & β = 0.1	234	94.2686293	139.7313707
α = 0.1 & β = 0.2	234	65.4985490	168.5014510
α = 0.1 & β = 0.3	234	52.9430967	181.0569033
Pulau Baai			
α = 0.1 & β = 0.1	570	142.1659316	427.8340684
α = 0.1 & β = 0.2	570	97.87935647	472.1206435
α = 0.1 & β = 0.3	570	193.4759504	376.5240496
Padang Kemiling			
α = 0.1 & β = 0.1	510	137.0536827	372.9463173
α = 0.1 & β = 0.2	510	94.71873296	415.2812670
α = 0.1 & β = 0.3	510	204.3912557	305.6087443

Tabel 5 menggambarkan bahwa metode pemulusan eksponensial tunggal dengan α = 0.1 dapat meramalkan jumlah curah hujan periode Januari 2007 sebesar 194.4634265 pada pos pengamatan Padang Harapan yang paling mendekati data asli jika dibandingkan dengan metode pemulusan eksponensial tunggal dengan pendekatan adaptif dan dengan selisih sebesar 39.53657349. Begitu pula pada pos pengamatan Pulau Baai dan Padang Kemiling, hasil ramalan yang didapat mendekati data aslinya adalah dengan menggunakan metode pemulusan eksponensial tunggal dan α = 0.1.

Peramalan dengan Metode Pemulusan Eksponensial Ganda Metode Linier Satu Parameter dari Brown

Pada metode ini ramalan curah hujan di kota Bengkulu masing-masing pos pengamatan diperoleh dari perbedaan antara nilai pemulusan tunggal dan ganda serta ditambahkan kepada nilai pemulusan tunggal dan disesuaikan untuk trend. Sehingga di dapat persamaan (14). Kemudian dari hasil ramalan akan dilihat seberapa besar tingkat ketepatan ramalan pada pada masing-masing α dengan menggunakan ukuran RMSE, MAE, dan MAPE yang hasilnya dapat dilihat pada Tabel 6 berikut.

Tabel 6. Ukuran Ketepatan Ramalan Maret 1985 – Desember 2006

Padang Harapan	RMSE	MAE	MAPE
α = 0.1	161.5038	11.44607	12.57762
α = 0.2	165.1949	11.49913	10.32270
α = 0.3	169.3897	11.66571	9.565448
Pulau Baai			
α = 0.1	175.6324	11.97095	19.13587
α = 0.2	178.2802	11.90575	12.91208
α = 0.3	181.5079	11.95555	11.31437
Padang Kemiling			
α = 0.1	168.5725	11.57574	18.50583
α = 0.2	171.9496	11.54482	12.40580
α = 0.3	175.9880	11.64163	11.11222

Dapat disimpulkan dari Tabel 6 diatas bahwa nilai RMSE dan MAE yang minimum diperoleh jika digunakan konstanta pemulusan $\alpha = 0{,}1$ pada masing-masing pos pengamatan, sedangkan nilai MAPE yang minimum diperoleh jika digunakan konstanta pemulusan $\alpha = 0{,}3$.

Berdasarkan data periode Januari 1985 sampai dengan Januari 2006, maka dapat diperoleh nilai ramalan pada periode Januari 2007 dengan rumus $F_{t+m} = a_t + b_t m$ dengan t pada periode Januari 2006, m =1 untuk Februari 2006, m = 2 untuk Maret 2006 seterusnya sampai m = 12 untuk Januari 2007.

Tabel 7. Ramalan Curah Hujan di Kota Bengkulu Februari 2006 sampai dengan Januari 2007 Pos Pengamatan Padang Harapan

Bulan	Xt	Ft		
		α = 0.1	A = 0.2	α = 0.3
Feb-06	349	380.6394	462.739	499.7397
Mar-06	294	385.3256	447.8414	448.6957
Apr-06	221	377.1083	406.2976	363.9412
Mei-06	125	352.2465	332.413	240.1677
Jun-06	452	305.9157	218.3628	67.72809
Jul-06	55	344.9442	324.854	307.1074
Agt-06	7	277.2481	154.5565	8.824696
Sep-06	7	202.6418	-8.03028	-225.467
Okt-06	7	135.2595	-128.136	-348.671
Nov-06	61	74.90178	-212.919	-397.556
Des-06	155	37.48718	-225.841	-318.467
Jan-07	234	33.88721	-148.746	-99.3556

Jadi Tabel 7 dan Lampiran 9 menjelaskan bahwa nilai ramalan yang mendekati data asli untuk Februari, Maret, dan November dengan menggunakan konstanta pemulusan α = 0.1. Dan untuk bulan April, Mei, Juli, dan Agustus nilai ramalan yang mendekati data asli dengan menggunakan konstanta pemulusan $\alpha = 0.3$ karena pada periode itu data cenderung menurun. Pada Juni, Oktober, November, Desember, dan Januari nilai ramalan yang mendekati data asli dengan konstanta pemulusan $\alpha = 0.1$ karena data mengalami kenaikan. Untuk September dengan konstanta $\alpha = 0.2$ karena pada periode itu data mangalami kenaikan.

Berdasarkan Tabel 8 dibawah ini dan Lampiran 10 dapat disimpulkan nilai ramalan yang mendekati data asli untuk Februari, Juni, Desember, dan Januari diperoleh dengan menggunakan $\alpha = 0.1$ karena data merngalami kenaikan. Dan September dengan konstanta $\alpha = 0.2$ data mengalami kenaikan. Sedangkan untuk bulan Maret, April, Oktober, dan November nilai ramalan yang mendekati data asli diperoleh dengan menggunakan konstanta $\alpha = 0.1$. Dan untuk bulan Mei, Juli, dan Agustus dengan konstanta pemulusan $\alpha = 0.3$ karena data cenderung menurun.

Tabel 8. Ramalan Curah Hujan di Kota Bengkulu Februari 2006 sampai dengan Januari 2007 Pos Pengamatan Pulau Baai

Bulan	Xt	Ft		
		α = 0.1	α = 0.2	α = 0.3
Feb-06	449	447.8183	560.3857	589.4023
Mar-06	345	464.8573	560.9892	464.2345
Apr-06	283	457.1527	495.2597	339.6642
Mei-06	76	435.0377	424.4387	73.89757
Jun-06	346	364.3647	257.0410	163.3029
Jul-06	65	366.6063	273.5970	-65.3064
Agt-06	1	295.3993	119.5547	-269.631
Sep-06	29	213.2987	-39.3060	-340.769
Okt-06	14	147.0072	-138.126	-379.490
Nov-06	144	83.32860	-217.934	-193.779
Des-06	340	65.77857	-171.853	228.1784
Jan-07	570	111.5959	26.65726	839.9888

Tabel 9. Ramalan Curah Hujan di Kota Bengkulu Februari 2006 sampai dengan Januari 2007 Pos Pengamatan Padang Kemiling

Bulan	Xt	Ft		
		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
Feb-06	390	446.205	546.7595	617.9961
Mar-06	301	450.6965	523.5602	541.6252
Apr-06	265	428.4383	449.5623	409.3153
Mei-06	77	4061662	381.7531	297.0082
Jun-06	447	341.7527	227.3308	57.40406
Jul-06	65	370.797	310.7799	260.5806
Agt-06	1	301.805	148.9901	-3.0318
Sep-06	9	222.0354	-16.4797	-232.364
Okt-06	10	151.8296	-134.421	-345.285
Nov-06	136	88.97066	-217.479	-388.377
Des-06	368	69.46367	-176.938	-210.528
Jan-07	510	122.289	46.26084	261.4966

Pada Tabel 9 dan Lampiran 11 menggambarkan nilai ramalan untuk Februari, Maret, dan Oktober mendekati data asli dengan menggunakan konstanta $\alpha = 0.1$ karena pada periode tersebut data cenderung menurun. Pada bulan April, Mei, dan Juli nilai ramalan mendekati data asli dengan menggunakan konstanta pemulusan $\alpha = 0.3$ karena data cenderung turun. Untuk Juni, November, dan Desember nilai ramalan mendekati data pada konstanta pemulusan $\alpha = 0.1$ karena data mengalami kenaikan. Untuk Oktober nilai ramalan mendekati data dengan konstanta pemulusan $\alpha = 0.1$ cenderung naik. Untuk Agustus dan Januari nilai ramalan mendekati data dengan konstanta pemulusan $\alpha = 0.3$ dan untuk bulan September nilai ramalan mendekati data pada pemulusan eksponensial $\alpha = 0.2$ karena data mengalami kenaikan.

Peramalan dengan Metode Pemulusan Eksponensial Ganda Metode Dua Parameter Holt

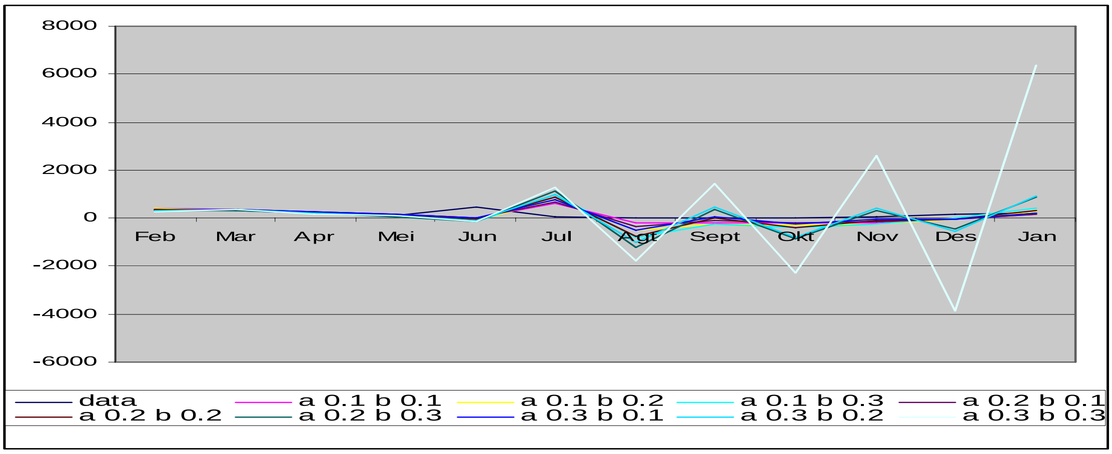
Pada peramalan curah hujan di kota Bengkulu dengan menggunakan metode pemulusan eksponensial ganda metode dua parameter Holt dapat memuluskan nilai trend dengan parameter yang berbeda dari parameter yang digunakan pada deret asli.

Pertama-tama nilai pengamatan tahun 1985 digunakan untuk memperoleh nilai *intercept* dan *slope trend* linier. Sehingga nilai ramalan untuk Januari 1986 sampai dengan Januari 2006 dapat diperoleh dengan menggunakan rumus (15), (16), (17), dan konstanta pemulusan digunakan untuk ramalan periode ke depan $\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.3$, $\beta = 0.1$, $\beta = 0.2$, $\beta = 0.3$. Perbandingan sembilan ramalan dapat dilihat dari ukuran RMSE, MAE, dan MAPE pada Tabel 10 di bawah ini.

Pada Tabel 10 dibawah ini menjelaskan dari ketiga pos pengamatan nilai RMSE, MAE, dan MAPE yang minimum diperoleh jika digunakan konstanta pemulusan $\alpha = 0.1, \beta = 0.1$.

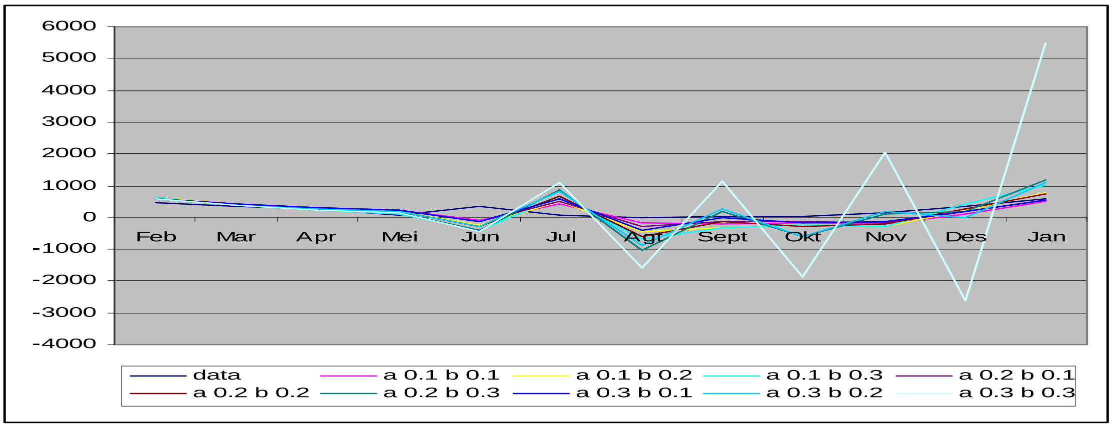
Seperti dengan pemulusan eksponensial ganda metode linier satu parameter dari Brown, pemilihan α dan β yang optimal akan menghasilkan nilai ramalan yang mendekati data yang sebenarnya. Pada kasus ini nilai α = 0.1, β = 0.1 memberikan ramalan yang lebih baik dibandingkan dengan α dan β yang lain. Keutamaan metode ini adalah mampu menyesuaikan S_t secara langsung untuk trend sebelumnya, yaitu b_{t-1} dengan menambahkan nilai pemulusan terakhir, yaitu S_{t-1}. Hal ini membantu untuk menghilangkan kelambatan dan menempatkan S_t ke dasar perkiraan nilai data saat ini. Selain itu metode ini dapat memuluskan trend.

Berdasarkan data Januari 1985 sampai dengan Januari 2006, nilai ramalan curah hujan di kota Bengkulu bulan Februari 2006 sampai dengan Januari 2007 dapat diperoleh menggunakan rumus $F_{t+m} = S_t + b_t m$ dengan t pada periode Januari 2006, t = 253, dan m = 1 untuk Februari 2006, m = 2 untuk Maret 2006 dan seterusnya m = 12 untuk Januari 2007.



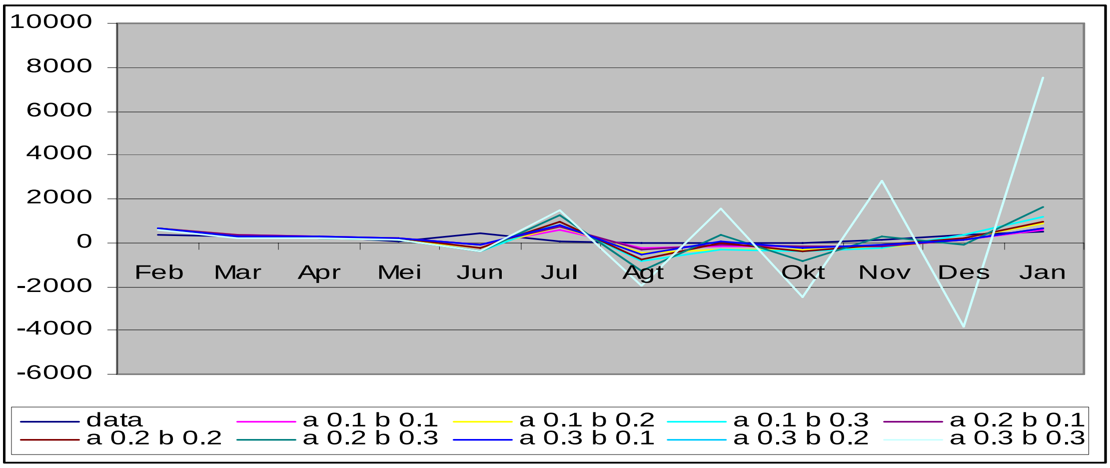
Gambar 7. Ramalan Pemulusan Eksponensial Ganda Dua Parameter Holt Februari 2006 sampai dengan Januari 2007

Berdasarkan Gambar 7 dan Lampiran 12 dapat diketahui bahwa nilai ramalan untuk bulan Februari dan September nilai ramalan yang mendekati data asli diperoleh dengan menggunakan $\alpha = 0.2$ dan $\beta = 0.2$ karena pada periode tersebut data cenderung menurun. Sedangkan Maret dan Juli dengan konstanta pemulusan masing-masing $\alpha = 0.1$ dan $\beta = 0.3$, dan $\alpha = 0.1$ dan $\beta = 0.1$ karena pada periode tersebut data cenderung menurun. Untuk bulan April dan Mei nilai ramalan yang mendekati data sebenarnya dengan menggunakan konstanta $\alpha = 0.3$ dan $\beta = 0.2$. Untuk bulan Juni dan Agustus dengan konstanta pemulusan $\alpha = 0.1$ dan $\beta = 0.1$. Untuk bulan November dan Januari dengan konstanta pemulusan $\alpha = 0.3$ dan $\beta = 0.1$, serta untuk bulan Oktober dan Desember dengan konstanta pemulusan $\alpha = 0.2$ dan $\beta = 0.1$, dan $\alpha = 0.1$ dan $\beta = 0.3$ karena data mengalami kenaikan.



Gambar 8. Ramalan Pemulusan Eksponensial Ganda Dua Parameter Holt Februari 2006 sampai dengan Januari 2007 Pos Pulau Baai

Gambar 8 dan Lampiran 13 menjelaskan nilai ramalan untuk bulan Februari dan Maret nilai ramalan yang mendekati data asli diperoleh dengan menggunakan $\alpha = 0.3$ dan $\beta = 0.3$ karena pada periode tersebut data cenderung menurun. Dan untuk Mei dan Juli nilai ramalan yang mendekati data sebenarnya diperoleh dengan menggunakan konstanta pemulusan $\alpha = 0.1$ dan $\beta = 0.3$ dan $\alpha = 0.1$ dan $\beta = 0.1$ karena data cenderung menurun. Sedangkan April, Oktober, November, dan Desember nilai ramalan yang mendekati data sebenarnya dengan menggunakan konstanta pemulusan masing-masing $\alpha = 0.1$ dan $\beta = 0.2$, $\alpha = 0.2$ dan $\beta = 0.1$, $\alpha = 0.2$ dan $\beta = 0.3$, dan $\alpha = 0.2$ dan $\beta = 0.2$. Untuk Juni dan Agustus dengan konstanta pemulusan $\alpha = 0.1$ dan $\beta = 0.1$. Untuk bulan September dan Januari dengan konstanta pemulusan $\alpha = 0.3$ dan $\beta = 0.1$ karena data mengalami kenaikan.



Gambar 9. Ramalan Pemulusan Eksponensial Ganda Dua Parameter Holt Februari 2006 sampai dengan Januari 2007 Pos Padang Kemiling

Dari Gambar 9 dan Lampiran 14 dapat disimpulkan nilai ramalan untuk bulan Juni, Agustus, Oktober, dan Desember nilai ramalan yang mendekati data asli diperoleh dengan menggunakan masing-masing $\alpha = 0.1$ dan $\beta = 0.1$, $\alpha = 0.1$ dan $\beta = 0.1$, $\alpha = 0.2$ dan $\beta = 0.1$, dan $\alpha = 0.1$ dan $\beta = 0.3$ karena pada periode tersebut data mengalami kenaikan. Sedangkan untuk April, Juli dan Januari nilai ramalan yang mendekati data sebenarnya diperoleh dengan menggunakan konstanta pemulusan $\alpha = 0.1$ dan $\beta = 0.1$. Dan untuk Februari, Maret, Mei, September, dan November dengan konstanta pemulusan masing-masing $\alpha = 0.3$ dan $\beta = 0.3$, $\alpha = 0.2$ dan $\beta = 0.2$, $\alpha = 0.1$ dan $\beta = 0.3$, $\alpha = 0.3$ dan $\beta = 0.1$, dan $\alpha = 0.1$ dan $\beta = 0.3$ karena data cenderung turun.

Setelah nilai ramalan diperoleh dan dibandingkan dengan nilai sebenarnya pada bulan Februari 2006 sampai dengan Januari 2007 dapat dikatakan bahwa dari keempat metode pemulusan eksponensial tidak ada nilai ramalan yang persis sama dengan data asli. Tetapi nilai ramalan tersebut tidak dapat disalahkan karena baik buruknya hasil ramalan tergantung pada bagaimana data histori diperoleh, karena jika data asli diperoleh dengan cara sembarangan tentu saja nilai ramalan tidak akan akurat dan sebaliknya akan memberikan hasil yang akurat jika data histori berasal dari pengamatan akurat.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan uraian pada pembahasan sebelumnya dapat disimpulkan bahwa:

a. Peramalan curah hujan di kota Bengkulu pada tiga titik pengamatan, yaitu pos pengamatan Padang Harapan, Pulau Baai, dan Padang Kemiling dengan metode pemulusan eksponensial tunggal lebih baik daripada metode pemulusan eksponensial tunggal pendekatan adaptif dengan menggunakan α =0.1 hasil ramalan mendekati data aslinya.
b. Peramalan curah hujan dengan metode pemulusan eksponensial ganda dua parameter dari Holt lebih baik daripada pemulusan eksponensial ganda metode linier satu parameter dari Brown.

Saran

Dari pembahasan dan analisis serta kesimpulan yang diperoleh, saran-saran yang dapat disampaikan adalah sebagai berikut:

a. Sebaiknya untuk data historis yang digunakan harusnya diperoleh dari pengamatan yang akurat sehingga akan dihasilkan nilai ramalan dengan tingkat ketepatan yang lebih baik.
b. Sebaiknya dapat dilakukan penelitian lanjut terhadap kasus ini dengan menggunakan metode-metode pemulusan eksponensial yang lain.

DAFTAR PUSTAKA

Anonim 1. 2007. *Metode Peramalan Bisnis dan Upaya Memperoleh Akurasi yang Lebih Baik*. http:// dickyrahardi. blogspot. Com/ index. html. 17 Maret 2007; 15:15.

Anonim 2. 2007. *Klasifikasi Iklim*. http:// mbojo. wordpress. Com/ . 1 Juni 2007; 09:35.

Anonim 3. 2007. *Rencana Strategis Tahunan*. http:// www.bdg.lapan.go.id/index.php?nama=reinstra&opt=detail&id=37. 1 Juni 2007; 09:35.

Arsyad, L. 1994. *Peramalan Bisnis, Edisi Pertama,* Yogyakarta: BPFE.

Clark, J. R. and Schkade. 1983. *Statistical Analysis for Administrative Decisions*. USA: Ohio South-Westren Publishing CO.

Haymans, A. 1990. *Teknik Peramalan Bisnis dan Ekonomi.* Jakarta: Rineka Cipta.

Kuncoro, M. 2004. *Metode Kuantitatif Teori dan Aplikasi untuk Bisnis dan Ekonomi*. Yogyakarta: UPPAMPYKPN.

Lakitan, B. 2002. *Dasar-Dasar Klimatologi*, Jakarta: PT. Raja Grafindo Persada.

Makridakis, S. S. C. Wheelwright, and V. E. McGee. 1999. *Metode dan Aplikasi Peramalan*. Jakarta: Erlangga.
Maryati, M. C. 2001. *Statistik Ekonomi dan Bisnis Plus*. Yogyakarta: UPPAMPYKPN
Ma'mur, M. T. 1994. *Ilmu Pengetahuan Bumi dan Antariksa*. Jakarta: Balai Pustaka.
Mustofa, A. 2000. *Kamus Lingkungan*. Jakarta: PT. Rineka Cipta.
Sugiarto, dan Harijono. 2000. *Peramalan Bisnis*. Jakarta: PT. Gramedia Pustaka Utama.
Sutrisno, H. 2004. *Statistik (Jilid 3) Edisi II*. Yogyakarta: Andi Offset.

REGRESI LINIER DENGAN VARIABEL BONEKA UNTUK MELIHAT BEBERAPA FAKTOR YANG DI DUGA BERPENGARUH TERHADAP INDEKS PRESTASI SEMESTER PERTAMA

Titin Sumarnii[1], Sigit Nugroho[2] dan Jose Rizal[2]

[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRAK

Dalam Analisis Regresi Linier, variabel tak bebas dapat bersifat kuanttitatif dan kualitatif (kategori). Untuk mengakomodasi adanya variabel kategori dalam regresi linier dapat digunakan variabel boneka (dummy variable). Indeks Prestasi (IP) mahasiswa semester pertama, dapat dipengaruhi oleh beberapa faktor diantaranya nilai hasil Ujian Nasional (UN), asal sekolah, jenis kelamin, pendidikan orang tua, dan pekerjaan orang tua. Untuk melihat beberapa faktor yang diduga berpengaruh terhadap IP Mahasiswa semester pertama ini dapat digunakan variabel boneka. Hasil penelitian menunjukkan bahwa hanya variabel asal sekolah yang memberikan pengaruh signifikan terhadap IP mahasiswa semester pertama.

Kata Kunci : Regresi Linier, Peubah Boneka

1. Pendahuluan

Dalam analisis regresi biasanya variabel-variabel yang digunakan merupakan variabel kuantitatif, baik itu variabel bebas maupun tak bebas. Namun seringkali terjadi bukan hanya variabel-variabel bebas kuantitatif saja yang mempengaruhi variabel tak bebas, ada juga variabel-variabel kualitatif (kategori) yang juga ikut mempengaruhi, seperti jenis kelamin, tingkat pendidikan, dan lain sebagainya (Gaspert, 1991). Untuk mengakomodasi adanya variabel bebas yang bersifat kualitatif tersebut dalam model regresi dapat dilakukan dengan menggunakan variabel boneka (variabel dummy). Santoso (2005) mengatakan bahwa variabel boneka adalah variabel yang digunakan untuk membuat kategori data yang bersifat kualitatif. Variabel boneka ini biasanya menggunakan nilai “0” dan “1”. Kedua nilai ini tidak menunjukkan bilangan (numerik), tetapi hanya sebagai identifikasi kelas atau kategorinya saja. Misalnya, 1 jika masuk kategori tertentu dan 0 jika masuk kategori lainnya.

Keberhasilan studi mahasiswa secara akademis (prestasi belajar) dapat dilihat dari nilai indeks prestasinya. Besarnya Indeks Prestasi (IP) setiap mahasiswa tentunya akan berbeda-beda karena setiap orang memiliki kemampuan dan latar belakang akademik yang berbeda-beda. Tujuan penelitian ini adalah untuk membentuk suatu model regresi yang variabel bebasnya bersifat kuantitatif dan kualitatif dan melihat beberapa faktor yang diduga berpengaruh terhadap indeks prestasi seorang mahasiswa.

2. Regresi Linier Ganda

Analisis regresi ganda digunakan untuk melihat hubungan antara dua variabel bebas atau lebih terhadap satu variabel tak bebas, atau untuk membuktikan ada atau tidaknya hubungan fungsional antara dua buah variabel bebas atau lebih dengan sebuah variabel tak bebas. Hubungan fungsional dalam regresi ini diharapkan berlaku untuk populasi berdasarkan data sampel yang diambil secara acak dari populasi yang bersangkutan. Hubungan fungsional tersebut dituliskan dalam bentuk persamaan matematika (disebut persamaan regresi) yang akan bergantung pada parameter-parameter (Sudjana, 1996).

Model regresi yang mengandung *k* variabel bebas (*X*) dan 1 variabel tak bebas (*Y*) dapat dituliskan sebagai

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon \tag{1}$$

Data pengamatan dari persamaan (2.1) dapat dinyatakan dalam bentuk pasangan data $(y_i, x_{i1}, x_{i2}, \ldots, x_{ik})$ dengan i = 1, 2, 3, ..., n. Pasangan data ini dapat disusun dalam notasi matriks sebagai berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{ik} \\ 1 & x_{21} & x_{22} & \cdots & x_{ik} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{11} & x_{12} & \cdots & x_{ik} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

Persamaan diatas dapat ditulis dalam bentuk sederhana, yaitu $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$.

Keterangan :

$\mathbf{Y}$: merupakan vektor variabel tak bebas $n \times 1$

$\mathbf{X}$: menyatakan matriks variabel bebas ukuran $n \times (k+1)$

β : vektor parameter ukuran $(k+1) \times 1$

ε : vektor galat ukuran $n \times 1$

Selanjutnya taksirannya dapat ditulis $\widehat{\mathbf{Y}} = \mathbf{Xb}$, dengan $\mathbf{b} = \widehat{\beta}$, vektor taksiran dari β (Sembiring, 2003).

2.1. Pendugaan Parameter dengan Metode Kuadrat Terkecil

Untuk menduga parameter pada persamaan regresi, dapat digunakan metode kuadrat terkecil. Aunuddin (1989) mengatakan bahwa pendugaan parameter regresi dengan metode ini dapat dilakukan dengan meminimumkan jumlah kuadrat simpangan y_i terhadap $E(y_i)$.

Uraian tentang pendugaan parameter ini dijelaskan oleh Sembiring (2003). Vektor $\mathbf{b} = (b_0, b_1, \ldots., b_k)'$ sebagai taksiran dari vektor parameter $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ dicari dengan meminimumkan jumlah kuadrat galat

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik}\right)^2 \tag{3}$$

Kemudian *J* diturunkan secara parsial terhadap $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ dan samakan dengan nol, yakni :

$$\frac{\partial J}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik}) = 0$$

$$\frac{\partial J}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - ... - \beta_k x_{ik})x_{i1} = 0$$
$$\frac{\partial J}{\partial \beta_2} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - ... - \beta_k x_{ik})x_{i2} = 0$$
$$\vdots$$
$$\frac{\partial J}{\partial \beta_k} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - ... - \beta_k x_{ik})x_{ik} = 0$$

Setelah disusun kembali dan semua parameter diganti dengan penaksirnya, sistem persamaan ini ditulis sebagai

$$\begin{aligned}
&\sum_{i=1}^{n} y_i = nb_0 + b_1\sum_{i=1}^{n} x_{i1} + b_2 x_{i2} + ... + b_k x_{ik} \\
&\sum_{i=1}^{n} y_i x_{i1} = b_0\sum_{i=1}^{n} x_{i1} + b_1\sum_{i=1}^{n} x_{i1}^2 + b_2\sum_{i=1}^{n} x_{i2} x_{i1} + ... + b_k\sum_{i=1}^{n} x_{ik} x_{i1} \\
&\sum_{i=1}^{n} y_i x_{i2} = b_0\sum_{i=1}^{n} x_{i2} + b_1\sum_{i=1}^{n} x_{i1} x_{i2} + b_2\sum_{i=1}^{n} x_{i2}^2 + ... + b_k\sum_{i=1}^{n} x_{ik} x_{i2} \\
&\vdots \\
&\sum_{i=1}^{n} y_i x_{ik} = b_0\sum_{i=1}^{n} x_{ik} + b_1\sum_{i=1}^{n} x_{i1} x_{ik} + b_2\sum_{i=1}^{n} x_{i2} x_{ik} + ... + b_k\sum_{i=1}^{n} x_{ik}^2
\end{aligned} \tag{4}$$

Persamaan (4) disebut persamaan normal, dapat ditulis dalam bentuk matriks sebagai berikut :

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y} \tag{5}$$

Bila matriks **X'X** tidak singular artinya matriks tersebut memiliki invers, maka persamaan normal (5) mempunyai penyelesaian yang tunggal, yaitu :

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \tag{6}$$

2.2. Pengujian Koefisien Regresi Secara Keseluruhan

Pengujian secara keseluruhan dimaksudkan untuk melihat hubungan linier antara variabel tak bebas Y_i i=1,2,…,n dan variabel bebas $X_1, X_2, ..., X_k$ secara bersama-sama, dengan n adalah banyaknya pengamatan. Pendekatan hipotesisnya adalah :

$$\mathrm{H_0} : \beta_1 = \beta_2 = ... = \beta_k = 0$$
$$\mathrm{H_a} : \text{minimal ada satu } \beta_j \neq 0$$

Penolakan $\mathrm{H_0} : \beta_j = 0$ menyatakan bahwa paling sedikit satu variabel bebas $X_1, X_2, ..., X_k$ memberikan sumbangan yang nyata pada model tersebut (Montgomery and Hines, 1990).

Jumlah Kuadrat Total merupakan penjumlahan dari Jumlah Kuadrat Regresi (JKR) dan Jumlah Kuadrat Sisa (JKS). Untuk analisis regresi linier sederhana penjelasan rumusnya adalah sebagai berikut (Sembiring, 2003) :

$$(\mathbf{Y} - \overline{\mathbf{Y}})'(\mathbf{Y} - \overline{\mathbf{Y}}) = (\hat{\mathbf{Y}} - \overline{\mathbf{Y}})'(\hat{\mathbf{Y}} - \overline{\mathbf{Y}}) + (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) \tag{7}$$

JKT = JKR + JKS

Selanjutnya melakukan pengujian dengan Uji F Anova dengan membuat tabel Anova sebagai berikut.

Tabel 1. Anova dalam Analisis Regresi Linier Berganda

Sumber	**Jumlah Kuadrat (JK)**	**Derajat Bebas (db)**	**Kuadrat Tengah (KT)**	**F_{hitung}**
Regresi	JKR	k	KTR= $\frac{JKR}{k}$	$\frac{\text{KTR}}{\text{KTS}}$
Sisa/residu	JKS	$n-k-1$	KTS= $\frac{JKS}{n-k-1}$	
Total	JKT	$n-1$		

Rumus-rumus Jumlah Kuadrat dapat dijabarkan sebagai berikut.

$$\begin{aligned}\text{JKS} &= \sum_{i=1}^{n}(\text{y}_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n}\varepsilon_i^2 \\ &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\end{aligned}$$

Substitusikan $\boldsymbol{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y}\text{ - }\mathbf{Xb}$ sehingga dihasilkan :

$$\begin{aligned}\text{JKS} &= (\mathbf{Y}\text{ - }\mathbf{Xb})'(\mathbf{Y}\text{ - }\mathbf{Xb}) \\ &= \mathbf{Y'Y}\text{ - }\mathbf{b'X'Y}\text{ - }\mathbf{Y'Xb} + \mathbf{b'X'Xb}, \quad \text{dimana } \mathbf{b'X'Y} = \mathbf{Y'Xb} \qquad (8) \\ &= \mathbf{Y'Y}\text{ - }\mathbf{2b'X'Y} + \mathbf{b'X'Xb}\end{aligned}$$

Dari persamaan (5) diketahui bahwa $(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$, sehingga persamaan di atas menjadi

$$\begin{aligned}\text{JKS} &= \mathbf{Y'Y} - \mathbf{2b'X'Y} + \mathbf{b'X'Y} \\ &= \mathbf{Y'Y} - \mathbf{b'X'Y}\end{aligned} \qquad (9)$$

Sementara,

$$\text{JKT} = \sum_{i=1}^{\text{n}}(\mathbf{Y_i} - \overline{\mathbf{Y}})^2 = \mathbf{Y'Y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \qquad (10)$$

Oleh karena itu, persamaan (8) dapat ditulis kembali menjadi

$$\text{JKS} = \mathbf{Y'Y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left(\mathbf{b'X'Y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right) \text{ atau}$$

$$\text{JKS = JKT - JKR}$$

Sehingga,

$$\text{JKR} = \mathbf{b'X'Y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{11}$$

Prosedur pengujian untuk $\text{H}_\text{a} : \beta_j \neq 0$ adalah :

$$\text{F}_{\text{hitung}} = \frac{\frac{\text{JKR}}{\text{k}}}{\frac{\text{JKR}}{\text{n}-\text{k}-1}} = \frac{\text{KTR}}{\text{KTS}} \tag{12}$$

H_0 ditolak jika $\text{F}_{\text{hitung}} > \text{F}_{(\alpha;\text{k};\text{n}-\text{k}-1)}$ (Montgomery and Hines, 1990).

Dengan α adalah tingkat signifikansi dan $1-\alpha$ adalah tingkat kepercayaan. Apabila H_0 ditolak artinya terdapat pengaruh yang nyata dari variabel bebas terhadap variabel takbebas. Untuk mengetahui variabel bebas mana yang memberikan kontribusi yang besar dalam mempengaruhi variabel tak bebas, maka dilakukan pengujian koefisien regresi secara individual.

2.3 Pengujian Koefisien Regresi Secara Individual

Uji ini dilakukan untuk mengetahui apakah variabel bebas secara parsial mempunyai pengaruh yang signifikan terhadap variabel tak bebas. Pengujian secara individual ini dilakukan dengan menggunakan *uji* t.

Hipotesis untuk pengujian nyata beberapa koefisien regresi secara individu β_j, adalah (Montgomery and Hines, 1990).:

$\text{H}_0 : \beta_j = 0$ artinya, tidak ada pengaruh variabel bebas $ke-j$ terhadap variabel tak bebas.

$\text{H}_\text{a} : \beta_j \neq 0$ artinya, ada pengaruh variabel bebas $ke-j$ terhadap variabel tak bebas.

Jika $\text{H}_0 : \beta_j = 0$ tidak ditolak, maka ini menunjukkan bahwa x_1 dapat dihilangkan dari model tersebut. Uji statistik untuk pengujian ini adalah

$$\text{t}_{\text{hitung}} = \frac{\widehat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}; j = 1,2,3,\ldots,k \tag{13}$$

dimana C_{jj} adalah elemen diagonal $\mathbf{(X'X)}^{-1}$ yang berhubungan dengan β_j dan $\hat{\sigma}^2$ adalah penduga varian kesalahan/sisa atau Kuadrat Tengah Sisa. Hipotesis nol $\text{H}_0 : \beta_j = 0$ ditolak jika $|\text{t}_{\text{hitung}}| > \text{t}_{\alpha/2,\text{n}-\text{k}-1}$. Apabila H_0 ditolak berarti variabel x_j berpengaruh nyata dan dapat digunakan sebagai penduga.

3. Regresi dengan Variabel Bebas Kualitatif

Salah satu cara yang dapat digunakan untuk membangun model regresi yang variabel bebasnya mengandung variabel kualitatif adalah dengan menggunakan variabel boneka. Dalam analisis regresi, variabel boneka ini digunakan nomor kode 1 untuk pengamatan yang masuk kategori 1 dan kode 0 untuk pengamatan yang masuk kategori lainnya (Djarwanto, 2001). Bisa juga sebaliknya, 0 untuk kategori 1 dan 0 untuk kategori lainnya. Jadi, variabel boneka sifatnya biner, nilainya 0 atau 1 tergantung pada apakah pengamatan berasal dari populasi dengan sifat tertentu atau bukan (Sembiring, 2003).

Untuk variabel bebas kualitatif yang memiliki *k* kategori maka bisa dibangun (*k*-1) variabel boneka (Draper and Smith, 1992). Misalnya ingin memperkirakan nilai variabel *Y* yang dipengaruhi oleh satu variabel bebas kualitatif yang mempunyai dua kategori, misalnya kategori 1 dan kategori 2, dan satu variabel bebas kuantitatif (x_2). Sehingga dapat dibuat satu buah peubah boneka, yang disimbolkan dengan D_1. Simbol D_1 ini hanya untuk membedakan mana yang peubah boneka dan mana yang merupakan variabel kuantitatif.

Cara pengerjaan untuk masalah ini ada dua cara. Pertama, sampel yang berasal dari kategori 1 dan kategori 2 dianalisis secara terpisah. Jika ukuran sampelnya besar, maka cara ini tidak menimbulkan masalah, hanya saja pekerjaan harus dilakukan dua kali untuk sampel dari kategori 1 dan kategori 2. Bila ukuran sampel tidak cukup besar, maka cara ini mempunyai kelemahan besar. Cara kedua ialah dengan memasukkan ke dalam model regresi apa yang disebut variabel dummy (variabel boneka).
Modelnya adalah (Sembiring, 2003):

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i \quad (14)$$

D_{i1} adalah variabel bebas kualitatif yang bernilai
0 jika individu ke-i masuk kategori 1
1 jika individu ke-i masuk kategori 2
X_{i2} adalah variabel bebas kuantitatif

X_{i3},..., X_{ik} adalah variabel lainnya yang diduga ikut menentukan nilai y. Jadi, variabel boneka bersifat biner, nilainya 0 dan 1, tergantung pada apakah pengamatan berasal dari populasi dengan sifat tertentu atau bukan, dalam hal ini adalah kategori 1 atau kategori 2. Perbandingan nilai pengamatan dari kategori 1 dan kategori 2 kemudian dapat dilakukan setelah menghitung $\hat{y}$ untuk kategori 1 (D_1=0) dan untuk kategori 2 (D_1=1).

Regresi untuk pengamatan yang berasal dari kategori 1 adalah :

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1(0) + \beta_2 X_2 + ... + \beta_k X_k \\ &= \beta_0 + \beta_2 X_2 + ... + \beta_k X_k \end{aligned}$$

Sementara untuk pengamatan yang berasal dari kategori 2, regresinya adalah :

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1(1) + \beta_2 X_2 + ... + \beta_k X_k \\ &= (\beta_0 + \beta_1) + \beta_2 X_2 + ... + \beta_k X_k \end{aligned}$$

4. Metodologi

Penelitian ini adalah penelitian terapan yang mengambil sampel Mahasiswa Jurusan Matematika FMIPA UNIB angkatan 2004/2005, 2005/2006, dan 2006/2007 yang berasal dari SMA Negeri dan Swasta dalam Kota Bengkulu. Data yang digunakan dalam penelitian ini merupakan data skunder yang diperoleh dari arsip Jurusan Matematika FMIPA UNIB.

Dalam penelitian ini, yang menjadi variabel tak bebas (Y) adalah IP mahasiswa semester pertama. Sementara variabel bebasnya (X) adalah :

- nilai UN bidang studi Matematika (X_1).
- nilai UN bidang studi Bahasa Indonesia (X_2).
- nilai UN bidang studi Bahasa Inggris (X_3).
- asal sekolah (variabel boneka), dikategorikan menjadi :
  - Kategori 1 : SMAN 2, SMAN 5, dan SMA Carolus.
  - Kategori 2 : SMAN 1, SMAN 3, SMAN 4, SMAN 6, SMAN 7, dan SMAN 8.
  - Kategori 3 : SMA Swasta, STM, dan SMEA
- jenis kelamin : laki-laki dan perempuan
- pendidikan orang tua (ayah dan ibu), dikategorikan menjadi :
  Pendidikan Dasar : SD/SMP
  Pendidikan Menengah : SLTA
  Pendidikan Tinggi : Perguruan Tinggi
- pekerjaan orang tua (ayah dan ibu), dikategorikan menjadi :
  - Pegawai pemerintah
  - Pegawai swasta
  - Wiraswasta
  - Lain-lain

5. Hasil Penelitian

Dari hasil analisis per variabel, diperoleh hasil bahwa hanya variabel asal sekolah yang memberikan pengaruh yang signifikan terhadap IP. Hal ini terbukti dari hasil analisis menunjukkan bahwa nilai P-Value untuk uji F secara keseluruhan bagi D_4 dan D_5 adalah 0,004 < $\alpha = 0{,}05$. Hasil uji Anova disajikan dalam Tabel 1 berikut.

Tabel 1. Tabel Anova Analisis dengan Variabel Asal Sekolah

Model	**Jumlah Kuadrat**	**Derajat Bebas**	**Kuadrat Tengah**	**F hitung**	**P-Value**
Regresi	1,798	2	0,99	6,536	0,004
Sisa	5,089	37	0,138		
Total	6,888	39			

Kemudian hasil uji secara individualnya, dihasilkan P-Value untuk konstanta, koefisien D_4 dan koefisien D_5 berturut-turut adalah 0,000, 0,001, dan 0,002. Semua nilai ini lebih kecil dari $\alpha = 0{,}05$, yang berarti bahwa koefisien regresi signifikan. Hal ini berarti bahwa variabel asal sekolah berpegaruh secara signifikan terhadap IP. Hasil uji ini lebih jelasnya dapat dilihat pada Tabel 10. berikut.

Tabel 10. Hasil Uji Koefisien Regresi Secara Individual dengan Uji t Untuk Variabel Asal Sekolah

Model	**Koefisien**	**T Hitung**	**P-Value**
Konstanta	1,895	7,226	0,000
D_4	1,006	3,615	0,001
D_5	0,988	3,282	0,002

Persamaan regresinya adalah :

$$Y = 1{,}895 + 1{,}006\ D_4 + 0{,}988\ D_5.$$

Misalkan ada seorang mahasiswa yang berasal dari SMA kategori I (SMAN 2, SMAN 5, dan SMA Carolus) maka nilai IP yang diprediksi adalah

$$Y = 1{,}895 + 1{,}006 . (1) + 0{,}988 . (0)$$
$$Y = 2{,}901$$

Jika seorang mahasiswa berasal dari SMA Kategori 2 (SMA Negeri selain SMA 2 dan SMA 5) maka IP yang dapat diprediksi adalah

$$Y = 1{,}895 + 1{,}006 . (0) + 0{,}988 . (1)$$
$$Y = 2{,}883$$

Jika seorang mahasiswa berasal dari SMA Kategori 3 (SMA Swasta selain SMA Carolus) maka IP yang dapat diprediksi adalah

$$Y = 1{,}895 + 1{,}006 . (0) + 0{,}988 . (0)$$
$$Y = 1{,}895.$$

Jadi, dari ketiga kategori asal sekolah, sekolah yang dapat memberikan prediksi IP terbaik adalah SMA kategori 1 yaitu SMAN 2, SMAN 5 dan SMA Carolus.

Selanjutnya, analisis dengan hanya menyertakan variabel asal sekolah ini memberikan nilai koefisien determinasi R^2 Adjusted = 0,261. Yang berarti bahwa hanya 26,1% variasi IP dapat dijelaskan oleh variasi asal sekolah mahasiswa tersebut. Artinya, sebanyak 73,9% variasi IP tersebut dijelaskan oleh faktor lain. Angka koefisien determinasi ini, meskipun variabel yang masuk ke dalam model hanya dua buah variabel boneka, tetapi memberikan angka koefisien determinasi yang lebih besar dari pada analisis-analisis sebelumnya. Hal ini disebabkan karena, variabel asal sekolah memberikan pengaruh yang signifikan terhadap IP, sementara pada analisis sebelumnya, variabel bebas yang masuk ke dalam model tidak memberikan pengaruh terhadap IP secara bersama-sama.

Oleh karena, uji F menyatakan bahwa terdapat pengaruh yang signifikan antara variabel asal sekolah dengan variabel IP dan melalui Uji t diperoleh hasil bahwa koefisien regresi signifikan, maka model yang terbentuk tersebut telah dapat digunakan. Namun, mengingat besarnya koefisien determinasi hanya 26,1% maka model tersebut belum begitu baik untuk digunakan sebagai prediksi.

6. Kesimpulan

Dari hasil penelitian yang diperoleh, dapat disimpulkan bahwa tidak terdapat pengaruh yang signifikan antara variabel kuantitatif Nilai UN Matematika, Bahasa Indonesia, dan Bahasa Inggris terhadap IP Mahasiswa Matematika FMIPA UNIB semester pertama, dan variabel kualitatif Jenis Kelamin, Pendidikan Orang Tua, dan Pekerjaan Orang Tua terhadap IP Mahasiswa Matematika FMIPA UNIB semester pertama. Sementara variabel asal sekolah (D4 dan D5) memberikan pengaruh yang signifikan terhadap (D4 dan D5) terhadap IP Mahasiswa Matematika FMIPA UNIB semester pertama. Adapun persamaan regresi yang terbentuk adalah :

$$Y = 1{,}895 + 1{,}006\ D_4 + 0{,}988\ D_5.$$

Dari model tersebut, dapat diprediksi bahwa mahasiswa yang dapat memperoleh IP terbaik adalah mahasiswa yang berasal dari SMA 2, SMA 5, atau SMA Carolus (SMA Kategori 1) yaitu sebesar 2,901. Namun, koefisien determinasi yang diperoleh hanya 0,261 yang berarti bahwa sebesar 26,1% variasi IP dijelaskan oleh Asal sekolah mahasiswa. Selebihnya, sebesar 73,9% variasi IP dijelaskan oleh faktor lain, seperti lingkungan tempat tinggal, minat belajar, dan lain-lain.

DAFTAR PUSTAKA

Aunuddin, 1989. *Bahan Pengajaran Analisis Data*. Depdikbud. IPB, Bogor.

Draper, N. and Smith, H., 1992. *Analisis Regresi Terapan (Terjemahan)*. Gramedia Pustaka Utama, Jakarta.

Gaspert V, 1991. *Ekonometrika Terapan I*. Penerbit Tarsito Bandung, Bandung.

Montgomery, D.C and Hines, W.W. 1990. *Probabilita dan Statistik dalam Ilmu ekayasa dan Manajemen*. UI-Press, Jakarta.

Santoso S., 2005. *Menggunakan SPSS untuk Statistik Parametrik*. Elek Media Komputindo, Jakarta

Sembiring, R.K., 2003. *Analisis Regresi*. Penerbit ITB, Bandung.

Sudjana, 1996. *Metoda Statistika*. Tarsito, Bandung.

KAJIAN DUA SAMPEL INDEPENDEN DENGAN UJI MEDIAN, MANN-WHITNEY-WILCOXON, DAN KOLMOGOROV-SMIRNOV

Wahyuni Saputri[1], Sigit Nugroho[2], dan Fachri Faisal[2]
[1]Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu
[2]Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

Abstrak

Statistik nonparametrik memiliki metode untuk menguji hipotesis pada dua sampel independen. Metode pengujian yang digunakan adalah uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov. Penulisan ini bertujuan untuk mengkaji penggunaan uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov dalam menganalisis hipotesis. Metode penulisan yang digunakan adalah studi literatur. Data yang digunakan adalah data yang diambil dari buku literatur. Data terdiri dari dua tipe yaitu data untuk sampel kecil dan sampel besar. Hasil penelitian menunjukkan bahwa ketiga metode meiliki hasil pengujian hipotesis yang sama, walaupun prosedur analisis ketiganya berbeda. Hal ini menunjukkan bahwa besar sampel tidak mempengaruhi pengambilan keputusan dari ketiga metode.

Kata kunci : *Statistik Nonparametrik, Dua Sampel Independen, Uji Median, Mann Whitney-Wilcoxon, Kolmogorov-Smirnov*.

Pendahuluan

Dua sampel independen merupakan sampel sampel yang saling independen. Independen yang dimaksudkan bukan hanya sampel-sampel yang akan dianalisis tetapi juga unsur-unsur penyusun sampel. Sehingga dimungkinkan keindependenan ini dilihat pada sampel yang diambil dari populasi yang sama.

Analisis yang dapat dilakukan pada dua sampel independen adalah menguji apakah dua sampel mewakili populasi yang berbeda dalam hal parameter lokasi seperti nilai tengah. Kedua, apakah dua sampel independen berasal dari distribusi yang sama.

Analisis untuk dua sampel independen pada statistik parametrik dapat menggunakan uji t atau uji F. Prosdedur uji-uji ini dapat diterapakan pada data yang berdistribusi normal.Asumsi data berdistribusi normal tidak selamanya dapat dipenuhi. Sehingga diperlukan metode alternatif yang memiliki asumsi bahwa data tidak harus berdistribusi normal. Metode ini adalah statistik nonparametrik.

Metode yang dibahas untuk dua sampel independen pada statistik nonparametrik antara lain uji Median, Uji Mann-Whitney-Wilcoxon, dan uji Kolmogorov-Smirnov. Metode lain yang dapat digunakan pada kasus ini adalah uji Eksak Fisher, uji Kilat Tukey, uji Khai-Kuadrat, uji Peringkat Moses, uji permutasi untuk dua sampel independen, dan lain-lain. Pada skala pengukuran data minimal nominal dapat digunakan uji Eksak Fisher dan uji Khai-Kuadrat dua sampel independen. Skala data pengukuran minimal ordinal menggunakan uji Median, Uji Mann-Whitney-Wilcoxon dan uji Kolmogorov-Smirnov.

Namun ketiga metode di atas memiliki asumsi dan kriteria tersendiri dalam analisis data. Sehingga pengunaannya juga disesuaikan dengan kriteria yang diisyaratkan. Hal ini membuat peneliti ingin mengetahui bagaimana penerapan uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov pada kasus yang sama.

Tujuan dari penulisan ini adalah untuk mempelajari prosedur uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov. Selain itu, ketiga uji tersebut akan dibandingkan untuk melihat persamaan atau perbedaan hasil yang mungkin terjadi pada suatu kasus yang sama.

Pengantar Teori Statistik Nonparametrik

Statistik inferensia mengenal dua prosedur analisis data, yaitu untuk statistik parametrik dan nonparametrik. Pemilihan prosedur pengujian hipotesis dilakukan berdasarakan data yang didapat. Data yang dapat dianalisis dengan statistik parametrik merupakan data yang memenuhi asumsi seperti, berdistribusi normal, skala pengukuran data minimal interval. Jika data yang diperoleh tidak memenuhi asumsi tersebut, maka analisis data dapat dilakukan dengan menggunakan metode lain yaitu statistik nonparametrik. Jika data yang tidak memenuhi asumsi statistik parametrik dipaksakan untuk menggunakan prosedur tersebut, maka hasil yang didapat dapat tidak valid.

Statistik nonparametrik memiliki asumsi yang lebih sederhana dari statistik parametrik. Hal ini dapat dilihat pada skala pengukuran data yang dapat analisis, skala pengukuran datanya dapat ordinal, bahkan nominal. Pada statistik nonparametrik, distribusi data tidak harus normal atau berdistribusi bebas, dalam hal pengujian hipotesis, statistik nonparametrik hanya membutuhkan spesifikasi secara garis besar dari distribusinya.

Ada beberapa kriteria umum yang harus dipenuhi jika ingin menggolongkan suatu prosedur atau metode dikatakan nonparametrik adalah

1. Metode ini digunakan untuk skala pengukuran data minimal nominal.
2. Metode ini digunakan untuk skala pengukuran data minimal ordinal.
3. Metode ini juga dapat digunakan pada skala pengukuran data interval atau rasio, dimana fungsi distribusi dari variabel acaknya tidak diketahui.

Kriteria di atas merupakan salah satu solusi untuk menetapkan analisis data yang digunakan pada statistik nonparametrik.

Sebagian besar uji-uji nonparametrik menggunakan uji peringkat dalam menganalisa data. Dalam hal ini, data asli yang didapat dari observasi diubah menjadi skor dalam rank (peringkat) dan dalam beberapa hal menggunakan bentuk tanda (sign). Dengan kata lain, sebagian besar prosedur uji nonparametrik tidak mengolah data asli dalam hal menganalisi data.

Dua Sampel Independen

Saat analisis data, terkadang ditemukan sampel-sampel independen atau tidak saling berhubungan. Sampel dikatakan independen jika memenuhi kriteria berikut : pertama unsur-unsur pada sampel pertama sama sekali tidak tergantung pada unsur-unsur sampel yang diambil berikutnya. Keduan, masing-masing unsur juga saling bebas dari setiap unsur lain dalam sampel itu. Dengan demikian, asal populasi tidak begitu mempengaruhi keindependenan sampel karena dapat saja sampel-sampel independen ini diambil dari populasi yang sama.

Penyelesaian atau penganalisaan untuk kasus seperti ini tentunya berbeda dengan sampel yang berhubungan. Pada statistik parametrik dapat dibuat analisa pada jenis sampel ini. Metode yang biasa digunakan adalah uji t untuk menguji apakah dua rata-rata populasi sama besar. Uji F untuk menguji apakah dua varians populasi sama.

Jika memiliki dua sampel yang independen, ada beberapa hal yang dapat dianalisis. Hal tersebut adalah, untuk melihat apakah dua sampel independen ini berasal dari populasi yang sama. Kedua, apakah dua sampel mewakili populasi yang berbeda dalam hal parameter lokasi seperti median. Ketiga apakah dua sampel berasal dari populasi yang memiliki sembarang perbedaan, dan apakah dua sampel independen memiliki distribusi populasi yang sama. Sehingga secara umum dapat dikatakan bahwa tujuan prosedur ini adalah untuk menduga beda atau selisih parameter tertentu pada kedua populasi dan juga untuk menguji hipotesis tentang kedua populasi.

Uji Median

Uji median merupakan suatu prosedur untuk menguji apakah dua sampel independen berbeda mediannya. Maksudnya, uji median ini memberikan kita informasi tentang mungkin atau tidaknya dua sampel independen telah ditarik dari populasi yang memiliki median yang sama. Kedua sampel acak yang diambil dapat memiliki besar sampel yang berbeda.

Asumsi-asumsi yang dibutuhkan (Conover, 1980)

- 以 Masing-masing sampel adalah sampel acak.
- 以 Masing-masing sampel saling independen ; $X_1, X_2, \ldots, X_n$ dan $Y_1, Y_2, \ldots, Y_m$.
- 以 Skala pengukuran data yang digunakan minimal ordinal.
- 以 Variable yang diamati kontinu.

Hipotesis yang diuji

H_0: Dua populasi memiliki median sama.

$$M_1 = M_2$$

H_1: Dua populasi memiliki median berbeda.

$$M_1 \neq M_2 \qquad M_1 < M_2 \qquad M_1 > M_2$$

Langkah prosedur uji Median

- Menentukan Median Gabungan (Grand Median) dari kedua sampel.
- Menyusun data pada table kontingensi 2 x 2 berikut,

Tabel Kontingensi Uji Median

	Sampel		Jumlah
	1	2	
Di Atas Median Gabungan	O_{11}	O_{12}	n_1
Di Bawah atau sama denganMedian Gabungan	O_{21}	O_{22}	n_2
Total	C_1	C_2	N

Keterangan :

O_{11} = Banyaknya hasil pengamatan dari sampel 1 yang nilainya lebih besar dari Median gabungan.

O_{12} = Banyaknya hasil pengamatan dari sampel 2 yang nilainya lebih besar dari Median gabungan.

O_{21} = Banyaknya hasil pengamatan dari sampel 1 yang nilainya lebih kecil dari Median gabungan.
O_{22} = Banyaknya hasil pengamatan dari sampel 2 yang nilainya lebih kecil dari Median gabungan.
C_1 = Jumlah pengamatan pada sampel 1
C_2 = Jumlah pengamatan pada sampel 2
n_1 = $O_{11} + O_{12}$
n_2 = $O_{21} + O_{22}$

Statistik uji untuk r *x* c tabel kontingensi,

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

dengan r = 2 dan c = 2 menjadi,

$$T = \frac{\left(O_{11} - E_{11}\right)^2}{E_{11}} + \frac{\left(O_{12} - E_{12}\right)^2}{E_{12}} + \frac{\left(O_{21} - E_{21}\right)^2}{E_{21}} + \frac{\left(O_{22} - E_{22}\right)^2}{E_{22}}$$

$$= \frac{O_{11}^2}{E_{11}} + \frac{O_{12}^2}{E_{12}} + \frac{O_{21}^2}{E_{21}} + \frac{O_{22}^2}{E_{22}} + E_{11} + E_{12} + E_{21} + E_{22} - \left(2O_{11} + 2O_{12} + 2O_{21} + 2O_{22}\right)$$

$$= \frac{O_{11}^2}{E_{11}} + \frac{O_{12}^2}{E_{12}} + \frac{O_{21}^2}{E_{21}} + \frac{O_{22}^2}{E_{22}} + E_{11} + E_{12} + E_{21} + E_{22} - 2\left(O_{11} + O_{12} + O_{21} + O_{22}\right)$$

diketahui bahwa $O_{11} + O_{12} + O_{21} + O_{22} =$ N,

$$T = \frac{O_{11}^2}{E_{11}} + \frac{O_{12}^2}{E_{12}} + \frac{O_{21}^2}{E_{21}} + \frac{O_{22}^2}{E_{22}} + E_{11} + E_{12} + E_{21} + E_{22} - 2N$$

subtitusikan $E_{ij} = \frac{n_i C_j}{N}$

$$T = \frac{N\,O_{11}^2}{n_1 C_1} + \frac{N\,O_{12}^2}{n_1 C_2} + \frac{N\,O_{21}^2}{n_2 C_1} + \frac{N\,O_{22}^2}{n_2 C_2} + \frac{n_1 C_1}{N} + \frac{n_1 C_2}{N} + \frac{n_2 C_1}{N} + \frac{n_2 C_2}{N} - 2N$$

$$= \frac{N\,O_{11}^2}{n_1 C_1} + \frac{N\,O_{12}^2}{n_1 C_2} + \frac{N\,O_{21}^2}{n_2 C_1} + \frac{N\,O_{22}^2}{n_2 C_2} + \frac{\left(n_1 + n_2\right)\left(C_1 + C_2\right)}{N} - 2N$$

karena $n_1 + n_2 =$ N dan $C_1+C_2 =$ $O_{11} + O_{12} + O_{21} + O_{22}$ =N,

$$= \frac{N\,O_{11}^2}{n_1C_1} + \frac{N\,O_{12}^2}{n_1C_2} + \frac{N\,O_{21}^2}{n_2C_1} + \frac{N\,O_{22}^2}{n_2C_2} - N$$

$$= \frac{N\,O_{11}^2 n_2C_2}{n_1n_2C_1C_2} + \frac{N\,O_{12}^2 n_2C_1}{n_1n_2C_1C_2} + \frac{N\,O_{21}^2 n_1C_2}{n_1n_2C_1C_2} + \frac{N\,O_{22}^2 n_1C_1}{n_1n_2C_1C_2} - \frac{N\,n_1n_2C_1C_2}{n_1n_2C_1C_2}$$

$$= \frac{N}{n_1n_2C_1C_2}\left(O_{11}^2n_2C_2 + O_{12}^2n_2C_1 + O_{21}^2n_1C_2 + O_{22}^2n_1C_1 - n_1n_2C_1C_2\right)$$

karena $n_1 = O_{11} + O_{12}$, $n_2 = O_{21} + O_{22}$, $C_1 = O_{11} + O_{21}$, $C_2 = O_{12} + O_{22}$ maka,

$$T = \frac{N}{n_1n_2C_1C_2}\left(O_{11}(C_1 - O_{21})n_2C_2 + O_{12}(C_2 - O_{22})n_2C_1 + O_{21}^2n_1C_2 + O_{22}^2n_1C_1 - n_1n_2C_1C_2\right)$$

$$= \frac{N}{n_1n_2C_1C_2}\left(O_{11}n_2C_1C_2 - O_{11}O_{21}n_2C_2 + O_{12}n_2C_1C_2 - O_{12}O_{22}n_2C_1 + O_{21}^2n_1C_2 + O_{22}^2n_1C_1 - [(O_{11} + O_{12})n_2C_1C_2]\right)$$

$$= \frac{N}{n_1n_2C_1C_2}\left(O_{11}n_2C_1C_2 - O_{11}O_{21}n_2C_2 + O_{12}n_2C_1C_2 - O_{12}O_{22}n_2C_1 + O_{21}^2n_1C_2 + O_{22}^2n_1C_1 - O_{11}n_2C_1C_2 - O_{12}n_2C_1C_2\right)$$

$$= \frac{N}{n_1n_2C_1C_2}\left(O_{21}^2n_1C_2 + O_{22}^2n_1C - O_{11}O_{21}n_2C_2 - O_{12}O_{22}n_2C_{11}\right)$$

$$= \frac{N}{n_1n_2C_1C_2}\left(O_{11}^2O_{22}^2 + O_{12}^2O_{21}^2 - 2O_{11}O_{12}O_{21}O_{22}\right)$$

sehingga didapat,

$$T = \frac{N}{n_1n_2C_1C_2}\left(O_{11}O_{22} - O_{12}O_{21}\right)^2 \qquad (3.1)$$

Kaidah Pengambilan Keputusan

Nilai T_{hitung} dibandingkan dengan T yang terdapat di Tabel Kai-Kuadrat dengan derajat bebas 1. Jika nilai $T_{hitung} > T_{tabel}$ maka H_0 ditolak dan H_1 dapat diterima.

Uji Mann-Whitney-Wilcoxon

Uji Wilcoxon-Mann-Whitney digunakan untuk menguji hipotesis nol tentang kesamaan parameter lokasi. Selain itu, uji ini juga dapat diterapkan untuk menguji hipotesis tentang sembarang perbedaan dalam populasi antara kedua sampel.

Asumsi-asumsi yang digunakan untuk metode ini (Conover, 1980) adalah

- Masing-masing sampel adalah sampel acak
- Masing-masing sampel independen satu sama lain $X_1, X_2, \ldots, X_m$ dan $Y_1, Y_2, \ldots, Y_n$.

- Skala pengukuran yang digunakan sekurang-kurangnya ordinal.
- Variabel yang diamati kontinu.
- Apabila terdapat perbedaan antara fungsi distribusi populasi, maka perbedaan yang dimaksud adalah perbedaan distribusi lokasi.

Hipotesis

Misalkan E*(X)* dan *E(Y)* masing-masing merupakan nilai harapan untuk populasi 1 dan 2, yang mewakili *X* dan *Y*.

H_0 : $E(X) = E(Y)$

H_1 : $E(X) \neq E(Y)$

Uji Mann-Whitney-Wilcoxon merupakan jumlah dari banyaknya $Y < X$ pada susunan terurut dari dua sampel independen

$$X_1, X_2, X_3, \ldots, X_m \qquad \text{dan} \qquad Y_1, Y_2, Y_3, \ldots, Y_n$$

di dalam suatu barisan terurut $m + n = N$ yang terurut dari nilai yang terkecil sampai yang terbesar. Diasumsikan bahwa dua sampel berdistribusi kontinu.

$$D_{ij} = \begin{cases} 1 & jika\ Y_j < X_i \qquad untuk\ semua\ i = 1,2,\ldots,m \\ 0 & jika\ Y_j > X_i \qquad\qquad\qquad j = 1,2,\ldots,n \end{cases}$$

Keterangan :

Y_j = Observasi yang berasal dari kelompok 2

X_i = Observasi yang berasal dari kelompok 1

Statistik uji Mann-Whitney-Wilcoxon

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

persamaan di atas dapat disederhanakan menjadi.

$$\begin{aligned} U' &= mn - \left[R_m - \frac{m}{2}(m+1) \right] \\ &= mn + \frac{m}{2}(m+1) - R_m \end{aligned} \tag{1}$$

dan

$$\begin{aligned} U' &= mn - \left[R_n - \frac{n}{2}(n+1) \right] \\ &= mn + \frac{n}{2}(n+1) - R_n \end{aligned} \tag{2}$$

Nilai U' yang didapat dari persamaan di atas akan diperbandingkan satu sama lain. Pengambilan keputusan, nilai U' yang digunakan adalah nilai U' yang paling kecil.

Kaidah Pengambilan Keputusan

Ada beberapa kriteria yang digunakan pada pengambilan keputusan dalam uji Mann-Whitney-Wilcoxon. Kriteria itu adalah,

- Sampel yang sangat kecil.

  Jika nilai m atau n yang lebih kecil atau sama 8 maka perhitungan rumus dapat dilakukan. Nilai U'' yang didapat dari perhitungan akan dicari nilai *p* pada Tabel

Nilai U dengan menggunakan *m* dan *n*. Jika nilai $p_{hitung} > \alpha$ yang sudah ditetapkan, maka terima H_0.

- Sampel $8 < n \leq 20$

  Sampel yang berukuran sedang yang dimasukkan dalam rumus (1) dan (2) tidak dapat lagi menggunakan Tabel Nilai U, tabel yang digunakan adalah Tabel Nilai untuk n > 8. Tabel Nilai untuk n > 8 memberikan nilai kritis U untuk uji satu arah dengan taraf nyata pengujian : 0.001; 0.01; 0.025; dan 0.05. Sedangkan untuk uji dua arah, taraf nyata pengujiannya adalah 0.002; 0.02; 0.05; dan 0.1. Bila yang diperoleh dari perhitungan lebih kecil atau sama dengan U dari Tabel Nilai untuk n > 8, maka H_0 ditolak.

- Sampel Besar $m \text{ atau } n > 20$

  Ukuran sampel yang besar tidak memungkinkan lagi penggunaan Tabel Nilai untuk n > 8 dan Tabel Nilai U. Ukuran sampel yang besar membuat distribusi sampling U secara cepat mendekati distribusi normal dengan

  Mean = $\mu_U = \dfrac{mn}{2}$

  Dengan standar deviasi = $\sigma_U = \sqrt{\dfrac{mn\,(m+n+1)}{12}}$

  Sehingga bila *m atau n* > 20 maka penentuan signifikansi suatu harga U observasi dengan,

$$z = \frac{U' - \mu_U}{\sigma_U} = \frac{U' - \dfrac{mn}{2}}{\sqrt{\dfrac{(m)(n)(m+n+1)}{12}}} \tag{3}$$

Pada kasus sampel besar kedua nilai U' dapat digunakan dalam pengambilan keputusan. Jika nilai z dari persamaan (3) lebih besar dari z nilai kritis pada Tabel Normal maka H_0 ditolak.

Uji Kolmogorov-Smirnov

Uji Kolmogorov-Smirnov merupakan suatu uji yang dipakai untuk menguji hipotesis nol yang menyatakan bahwa dua sampel bebas berasal dari populasi-populasi yang identik dalam hal lokasi dan distribusi

Asumsi-asumsi yang harus dipenuhi (Daniels, 1989) adalah

- ♦ Masing-masing sampel independen satu sama lain $X_1, X_2, \ldots, X_n$ dan $Y_1, Y_2, \ldots, Y_m$.
- ♦ Skala pengukuran yang digunakan sekurang-kurangnya ordinal.

Hipotesis

a. Hipotesis Dua arah

Misalkan *F(x)* dan *G(x)* masing-masing merupakan fungsi distribusi untuk populasi 1 dan 2, yang mewakili *X* dan *Y.*

$H_0 : F(x) = G(x)$ untuk semua *x*
$H_1 : F(x) \quad G(x)$ untuk paling sedikit satu nilai x

b. Hipotesis Satu arah

$H_0 : F(x) \leq G(x)$ untuk semua *x*
$H_1 : F(x) > G(x)$ untuk paling sedikit satu nilai x atau sebaliknya.

$Sn_1(X)$ adalah nilai pengamatan kumulatif salah satu sampel yang besarnya adalah $Sn_1(X) = k/n_1$. Dimana k adalah jumlah yang sama atau lebih kecil daripada X, sedangkan n_1 merupakan jumlah observasi di sampel 1. $Sn_2(X)$ merupakan nilai pengamatan kumulatif dari sampel lain, yang besarnya $Sn_2(X) = k/n_2$, dimana n_2 adalah jumlah observasi pada sampel 2. Sehingga uji Kolmogorov-Smirnov untuk dua sampel adalah

Uji satu arah

$$D = maksimum\left[Sn_1(X) - Sn_2(X)\right] \tag{4}$$

Uji dua arah

$$D = maksimum\left|Sn_1(X) - Sn_2(X)\right| \tag{5}$$

Kaidah Pengambilan Keputusan

Ada beberapa kriteria yang digunakan pada pengambilan keputusan dalam uji Kolmogorov-Smirnov. Kriteria itu adalah,

- Sampel kecil (n_1 dan $n_2 < 40$)
  - Sampel $n_1 = n_2$
    Apabila sampel yang diambil berukuran kecil, n_1 dan $n_2 \leq 40$, dimana n_1=n_2. Tabel Nilai D untuk Sampel sama dapat digunakan untuk menguji H_0. Pada tabel ini terdapat nilai D_{tabel}. Jika $D_{hitung} > D_{tabel}$ maka H_0 ditolak. Selain itu terdapat juga nilai N *(N = n_1= n_2)*. Perlu juga diperhatikan uji satu arah atau uji dua arah yang dipergunakan untuk mentukan signifikansinya.
  - Sampel $n_1 < n_2$
    Sedangkan untuk $n_1 < n_2$ maka tabel yang digunakan adalah Tabel Nilai D untuk Sampel Tidak Sama. Jika $D_{hitung} > D_{tabel}$ maka H_0 ditolak, pada nilai signifikansi tertentu.
- Sampel Besar (n_1 dan n_2 >40)
  - Sampel $n_1 = n_2$

    Jumlah sampel yang besar *(N = n_1= n_2 > 40)* menyebabkan terjadi perubahan rumus yang digunakan untuk mencari D_{tabel}. Tabel yang digunakan tetap L Tabel Nilai D untuk Sampel sama, namun nilai D didapat dengan menggunakan rumus yang terdapat di Tabel Nilai D untuk Sampel sama, sesuai dengan α yang ditentukan. Jika $D_{hitung} > D_{tabel}$ maka H_0 ditolak.
  - Sampel $n_1 \neq n_2$

Apabila n_1 dan n_2 keduanya lebih besar dari 40, maka tabel yang dipergunakan adalah Tabel Nilai D untuk Sampel Tidak Sama. Pada Tabel Nilai D untuk Sampel Tidak Sama tidak diperlukan nilai N. Penggunakan Tabel Nilai D untuk Sampel Tidak Sama adalah menentukan nilai D pengamatan berdasarkan rumus (3.10.a) atau (3.10.b). Lalu bandingkan nilai pengamatan ini dengan nilai dari Tabel Nilai D untuk Sampel Tidak Sama. Tolak H_0 pada taraf yang ditentukan menurut kejadian. Pada Tabel Nilai D untuk Sampel Tidak Sama terdapat berbagai nilai kritis yang diberikan 0.1, 0.05, 0.025, 0.01, 0.005, 0.001. Nilai kritis ini ditentukan terlebih dahulu. Misalkan nilai kritis yang ditentukan adalah 0.05, maka pada Tabel Nilai D untuk Sampel Tidak Sama dicari nilai kritis 0.05 lalu lihat rumus yang sepadan untuk menetukan nilai D pengamatan. Nilai D yang telah dihitung dengan Tabel Nilai D untuk Sampel Tidak Sama dibandingkan dengan nilai kritis 0.05 yang sudah ditetapkan. H_0 ditolak jika nilai $D_{hitung} > D_{tabel}$.

Studi Kasus

Suatu penelitian dilakukan pada beberapa bank. Penelitian ini bertujuan untuk mengetahui tentang ada tidaknya perbedaan kualitas manajemen antara Bank-Bank yang dianggap favorit oleh masyarakat dan Bank yang tidak favorit. Penelitian ini menggunakan sampel 12 Bank yang dianggap tidak favorit (Kelompak A) dan 15 Bank yang dianggap favorit (Kelompok B). Selanjutnya kedua kelompok Bank tersebut diukur kualitas manajemennya dengan menggunakan instrumen khusus yang terdiri dari butir-butir pertanyaan. Skor penilaian yang diberikan tertinggi 40 dan terendah 0 (α = 0,05). (Ritongga, 1987)

Data .

Kelompok A	
16	15
18	10
10	12
12	15
16	16
14	11

Kelompok B		
19	27	25
19	23	27
21	27	23
25	19	19
26	19	29

Rumusan Masalah.

Adakah perbedaan kualitas manajemen yang signifikan antara Bank yang favorit dan tidak favorit.

Hipotesis

H_0 : Tidak terdapat perbedaan kualitas manajemen yang signifikan antara bank yang favorit dan tidak favorit.

H_1 : Terdapat perbedaan kualitas manajemen yang signifikan antara bank yang favorit dan tidak favorit.

Penyelesaian kasus di atas akan menggunakan Uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov.

Analisis Kasus I

Kasus yang telah disebutkan di atas akan dianalisis dengan tiga metode, yaitu Uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov. Hasil dari ketiga metode tersebut akan dilihat persamaan atau perbedaan dalam pengujian hipotesis.

Analisis dengan Uji Median

Untuk menganalisis kasus dua sampel independen tentang uji kualitas manajemen bank di atas, dapat dilakukan dengan Uji Median. Langkah-langkah yang dilakukan adalah,

Langkah 1. Dua sampel yang telah diambil, gabungkan menjadi satu kelompok dan dicari median gabungannya.

Kelompok Gabungan :

$$\underbrace{10\ 10\ 11\ 12\ 12\ 14\ 15\ 15\ 16\ 16\ 16\ 18\ 19}\ \mathbf{19}\ \underbrace{19\ 19\ 19\ 21\ 23\ 23\ 25\ 25\ 26\ 27\ 27\ 27\ 29}$$

Median dari kelompok gabungan adalah 19.

Langkah 2. Kelompokkan data dari kedua kelompok kedalam tabel kontingensi 2 x 2 berikut ini,

Tabel Kontingensi Uji Median

Kategori	Sampel		Jumlah
	A	B	
Sampel di atas 19	0	11	11
Sampel di Bawah atau sama dengan 19	11	5	16
Total	11	16	27

Langkah 3. Gunakan uji statistik untuk median (rumus 3.2)

$$T = \frac{27\ \left[(11\ x\ 11) - (0\ x\ 5)\right]^2}{(11\ x\ 16)\cdot(11\ x\ 16)}$$

$$= 12.76$$

Langkah 4. Dari perhitungan didapat T_{hitung} adalah 12,76. Berdasarkan Lampiran 1 untuk dk=1 dan $\alpha = 0{,}05$; nilai T_{tabel} adalah 3,814. Karena $T_{hitung} > T_{tabel}$ maka berdasarkan kaidah pengambilan keputusan uji median, maka H_0 tidak dapat diterima atau dengan kata lain H_1diterima.

Kesimpulan. Terdapat perbedaan yang signifikan antara kualitas manajemen antara Bank yang favorit dan tidak favorit. Bank favorit memiliki kualitas manajemen yang lebih baik dibandingkan dengan Bank tidak favorit.

Analisis dengan Uji Mann-Whitney-Wilcoxon

Setelah melakukan analisis dengan menggunakan uji Median, sekarang analisis dilakukan dengan menggunakan uji Mann-Whitney-Wilcoxon. Langkah-langkah yang dilakukan dalam uji Mann-Whitney-Wilcoxon adalah,

Langkah 1. Data dari kedua kelompok digabungkan membentuk kelompok baru. Lalu beri peringkat. Jika terdapat nilai yang sama, maka peringkat yang diberikan adalah peringkat rata-rata.

Tabel Peringkat Kelompok A dan B

Kelompok A (Nilai Kualitas)	**Peringkat Gabungan**	**Kelompok B (Nilai Kualitas)**	**Peringkat Gabungan**
16	10	19	15
18	12	19	15
10	1.5	21	18
12	4.5	25	21.5
16	10	26	23
14	6	27	25
15	7.5	23	19.5
10	1.5	27	25
12	4.5	19	15
15	7.5	19	15
16	10	25	21.5
11	3	27	25
		23	19.5
		19	15
		29	27
	$R_m = 78$		$R_n = 300$

Langkah 2. Masukkan Nilai R_n dan R_m pada rumus (3.7.a) dan (3.7.b) dengan $n = 15$ dan $m = 12$.

$$U' = 15 \cdot 12 \; - \left[300 - \frac{15}{2}(15+1) \right]$$
$$= 180 + (15 \cdot 8) - 300$$
$$= 0$$

$$U' = 15 \cdot 12 - \left[78 - \frac{12}{2}(12+1) \right]$$
$$= 180 + 6 \cdot 13 - 78$$
$$= 180$$

Langkah 3. Dari kedua nilai U' yang didapat, dipilih nilai U' yang paling kecil untuk dibandingkan dengan $U_{tabel.}$ Jadi nilai U'_{hitung} yang dipilih adalah 0. sedangkan nilai U_{tabel} untuk $\alpha = 0.05$ untuk uji dua sisi dan m = 12; n = 15 adalah 49. Karena $U'_{hitung} < U_{tabel}$ maka H_0 tidak dapat diterima.

Kesimpulan. Dari hasil menunjukkan bahwa H_0 ditolak sehingga dapat disimpulkan bahwa terdapat perbedaan kualitas manajemen antara Bank yang favorit dan tidak favorit.

Analisis dengan Uji Kolmogorov-Smirnov

Analisis terakhir yang digunakan untuk menganalisis kasus 1 adalah uji Kolmogorov-Smirnov. Langkah-langkah yang dilakukan adalah,

Langkah 1. Data kelompok A dan B dibuat dalam bentuk tabel frekuensi kumulatif.
R = nilai tertinggi-nilai terendah
= 29 – 10 = 19
K = 1 + 3.322 log N dimana N = m+n
= 1 + 3.322 (log 27)
= 5,75 dibulatkan menjadi 5

Panjang kelas $I = \frac{R}{K} = \frac{19}{5} = 3{,}8$ sehingga panjang kelas 4

Tabel Nilai Kumulatif

	Kualitas manajemen				
	10-13	14-17	18-21	22-25	26-29
Kelompok A	5	11	12	0	0
Kelompok B	0	0	6	10	15

Tabel $Sn_1(X)$ - $Sn_2(X)$

	Kualitas Manajemen				
	10-13	14-17	18-21	22-25	26-29
$Sn_1(X)$	5/12	11/12	12/12	12/12	12/12
$Sn_2(X)$	0	0	6/15	10/15	15/15
$Sn_1(X) - Sn_2(X)$	25/60	55/60	36/60	20/60	0

Langkah 2. Dicari nilai D dengan menggunakan rumus (3.10.b) karena uji yang dupakai adalah uji dua arah dengan α =0.05

$$D = maksimum\left|Sn_1(X) - Sn_2(X)\right|$$

$$D = {}^{55}\!/_{60}$$

dengan demikian, D = 0, 92 .

Langkah 3. Dari Lampiran 6 didapat nilai D dengan $n_1 = 12$, $n_2 = 15$ dan α =0.05 (Uji dua arah) adalah ½ atau 0,5.

Kesimpulan. Karena nilai $D_{hitung} > D_{tabel}$ maka H_0 tidak dapat diterima.

Sehingga dapat disimpulkan bahwa terdapat perbedaan kualitas manajemen antara Bank yang favorit dan tidak favorit.

<u>Kesimpulan Kasus</u>

Kasus I telah dianalisis dengan menggunakan ketiga metode seperti di atas. Hasil hari ketiganya menunjukkan bahwa, uji Median menyimpulkan bahwa H_0 ditolak. Begitu juga dengan dua uji lainnya, uji Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov, menunjukkan hasil pengujian hipotesis sama yaitu penolakkan terhadap H_0. Dengan demikian dapat dikatakan bahwa ketiga metode analisis (Uji Median, Mann-Whitney-Wilcoxon, dan Kolmogorov-Smirnov) memiliki hasil atau kesimpulan yang sama yaitu menolak H_0 pada kasus dengan sampel berukuran sedang.

Kesimpulan

Analisis dua sampel independen nonparametrik dapat menggunakan uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov. Uji Median, Mann-Whitney dan

Kolmogorov-Smirnov memiliki hasil pengujian hipotesis yang sama pada sampel sedang.
Besarnya sampel tidak mempengaruhi pengambilan hipotesis yang dihasilkan oleh ketiga metode analisis dua sampel independen (Uji Median, Mann-Whitney-Wilcoxon dan Kolmogorov-Smirnov), walaupun prosedur uji dari ketiga metode berbeda.

Daftar Pustaka

[1] Chase, C.I. 1987. *Elementary Statistical Procedures*. Third Edition. McGraw-Hill. New York

[2] Conover, W.J. 1980. *Practical Nonpametric Statistics.* Second Edition. Jhon Wiley and Sons. New York.

[3] Daniels, W.W. 1989. *Statistika Nonparametric Terapan.* Edisi Terjemahan Gramedia. Jakarta.

[4] Elifson, K.W, and R. Runyon. 1990. *Fundamental of Social Statistics*. Second Edition. McGraw-Hill International Edition. Singapore.

[5] Gibbon, J.D. 1985. *Nonparametric Statistical Inference*. Marcel Dekker. New York, NY.

[6] Kanji, G.K. 1999. *100 Statistical Test*. New Edition. Sage Publication. London.

[7] Randles, R.H, and D.A. Wolfe. 1979. *Introduction to The Theory of Nonparametric Statsitics*. Jhon Wiley and Sons. New York.

[8] Ritonga, A. 1987. *Statistik Terapan Untuk Penelitian*. Universitas Indonesia. Jakarta.

[9] Siegel, S. 1997. *Nonparametric Statistics for Behavioral Sciences.* McGraw-Hill. New York.

[10] Sprent, P. 1991. *Metode Statistik Nonparametrik Terapan.* Edisi Terjemahan Universitas Indonesia. Jakarta.

[11] Walpole, R.E dan R. H. Myers. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*. Edisi keempat. ITB. Bandung

[12] Wampold, B.E, and C.J. Drew. 1990. *Theory and Application of Statistics.* McGraw-Hill International Edition. Singapore.

KAJIAN UJI MANN-WHITNEY DAN UJI PERINGKAT BERTANDA WILCOXON

Oleh
Yelvarina[1], Sigit Nugroho[2] dan Baki Swita[2]
[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu
[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRAK

Tulisan ini mengkaji statistik nonparametrik khususnya uji Mann-Whitney dan uji peringkat bertanda Wilcoxon serta membandingkan kedua uji ini dengan uji t. Metode yang digunakan adalah simulasi data dengan menggunakan paket program Microsoft Excel. Hasil analisis menunjukkan bahwa untuk data yang diketahui bentuk distribusinya yaitu berdistribusi normal, uji parametrik dengan menggunakan uji t memberikan hasil yang lebih baik daripada uji nonparametrik dengan menggunakan uji Mann-Whitney dan uji peringkat bertanda Wilcoxon. Sedangkan untuk data yang bukan berdistribusi normal (Kai-kuadrat, Gamma, dan Seragam), uji nonparametrik dengan menggunakan uji Mann-Whitney dan uji peringkat bertanda Wilcoxon memberikan hasil yang lebih baik daripada uji parametrik dengan uji t.

Kata kunci : *Statistik Parametrik, Statistik Nonparametrik, uji Mann-Whitney, uji Peringkat Bertanda Wilcoxon.*

Pendahuluan

Salah satu tujuan pokok dari statistika adalah melakukan inferensia statistika, yaitu penarikan kesimpulan tentang parameter populasi berdasarkan data sampel. Untuk melakukan inferensia statistika diperlukan pengetahuan tentang distribusi sampling atau distribusi penarikan sampel.

Distribusi sampling dari banyak uji statistik yang diterapkan pada pengujian hipotesis didasarkan pada asumsi mengenai populasinya dari mana sampel diambil. Misalnya pada uji-t untuk menguji perbedaan dua nilai tengah populasi, pengujian bertumpu pada asumsi mengenai bentuk distribusi populasinya (yaitu, berdistribusi normal) dan pada asumsi nilai parameter ragam populasi adalah sama. Pengujian-pengujian asumsi yang didasarkan pada asumsi tentang distribusi populasi atau ukuran parameter populasi dinamakan uji parametrik dan uji statistiknya dinamakan statistika parametrik.

Metode statistik nonparametrik, seperti statistik uji Mann-Whitney dan uji peringkat bertanda Wilxocon, menggunakan sejumlah asumsi yang kuat. Misalnya, tidak memperhatikan bentuk distribusi populasi dan variabel yang diamati adalah variabel acak kontinu. Jika asumsi tersebut dipenuhi, uji Mann-Whitney maupun uji peringkat bertanda Wilcoxon mempunyai kuasa yang lebih besar dari pada uji parametrik yang kegunaannya sama. Uji Mann-Whitney dipakai apabila peneliti tidak mengetahui karakteristik kelompok item yang menjadi sumber sampelnya. Metode ini dapat diterapkan terhadap data yang diukur dengan skala ordinal. Sedangkan pada uji peringkat bertanda Wilcoxon digunakan jika besaran maupun arah perbedaan relevan untuk menentukan apakah terdapat perbedaan yang sesungguhnya antara pasangan data yang diambil dari satu sampel atau dua sampel yang saling terkait

Statistik Parametrik dan Nonparametrik

Suatu tes statistik parametrik adalah suatu tes yang modelnya menetapkan adanya syarat-syarat tertentu tentang parameter populasi yang merupakan sumber sampel penelitian. Syarat-syarat itu biasanya tidak diuji dan dianggap sudah dipenuhi. Tes-tes parametrik juga menuntut bahwa skor-skor yang dianalisis merupakan hasil suatu pengukuran yang sedikitnya berkekuatan sebagai skala interval.

Uji t (t-*test*)

Uji t merupakan uji statistik parametrik yang digunakan untuk menguji hipotesis tentang kesamaan dari rataan populasi dengan varian yang tidak diketahui. Uji t ini digunakan untuk data yang bertipe numerik misalnya volume air, lama hidup bohlam yang diasumsikan memiliki sebaran normal. Uji ini menghasilkan apa yang disebut statsitik uji t dengan basis perhitungan adalah selisih antara rata-rata yang didapat dari data dengan rata-rata yang dihipotesiskan, dan dibandingkan dengan nilai t-tabel dengan derajat bebas tertentu.

Uji t Independent

1. Uji Hipotesis Tentang Mean Populasi
2. Uji Hipotesis Beda Dua Mean Populasi – Dua Sampel Independent.

Harga uji statistik dari sampel-sampel kecil yang diasumsikan bahwa kedua populasi berdistribusi normal dan diasumsikan varian kedua populasi sama ($\sigma_1^2 = \sigma_2^2$) tapi tidak diketahui adalah:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right]\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$$

keterangan :

$\bar{x}_1 = rataan\ sampel\ 1$

$\bar{x}_2 = rataan\ sampel\ 2$

$n_1 = besar\ sampel\ 1$

$n_2 = besar\ sampel\ 2$

$S_1^2 = \text{var}ian\ sampel\ 1$

$S_2^2 = \text{var}ian\ sampel\ 2$

dengan derajat kebebasan (df) :

$df = n_1 + n_2 - 2$

Apabila diasumsikan kedua populasi berdistribusi normal dan diasumsikan varian kedua populasi tidak sama ($\sigma_1^2 \neq \sigma_2^2$) dan tidak diketahui, maka harga uji statistik t dinyatakan dalam rumus:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

keterangan :

$\bar{x}_1 = rataan\ sampel\ 1$

$\bar{x}_2 = rataan\ sampel\ 2$

$n_1 = besar\ sampel\ 1$

$n_2 = besar\ sampel\ 2$

$S_1^2 = \mathrm{var}ian\ sampel\ 1$

$S_2^2 = \mathrm{var}ian\ sampel\ 2$

Uji t Berpasangan

Misalkan peubah X_A dan X_B diamati secara berpasangan artinya dalam setiap pengukuran yang diukur adalah pasangan [A,B] karena pengamatannya secara berpasangan maka dalam setiap pengamatan X_A dan X_B tidak lagi bebas sesamanya walaupun bebas antara pasangan yang satu dengan pasangan yang lainnya.
Harga uji statistiknya dihitung dengan rumus:

$$t = \frac{\bar{d}}{S_d / n}$$

keterangan :

$\bar{d}$ = Mean dari harga-harga d (perbedaan harga-harga yang berpasangan)

S_d = Deviasi standar dari harga-harga d.

n = Banyaknya pasangan

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} d_i^{\ 2} - n\bar{d}^2}{n-1}}$$

Kekuatan Efisiensi Suatu Uji Hipotesis

Sebuah kriterium untuk mengevaluasi unjuk kerja *(performance)* suatu uji adalah efisiensi. Patokan yang paling sering digunakan untuk mengukur efisiensi suatu uji nonparametrik adalah *efisiensi relatif asimptotiknya (Asymptotic Relative Efficiency / ARE)*. Efisiensi yang tinggi adalah karakteristik yang harus dimiliki oleh suatu uji.

Pada umumnya kekuatan suatu tes statistik meningkat dengan meningkatnya ukuran sampel (N). Jadi makin besar sampel, makin besar kuasa uji karena makin besar sampel, kemungkinan berbuat kesalahan tipe II menjadi kecil, yaitu kesalahan yang disebabkan karena kita menerima H_0 padahal H_0 itu salah

Uji Mann-Whitney

Bila angka sama terjadi, berikan kepada masing-masing kedua observasi itu rata-rata rangking yang akan dimiliki seandainya angka sama itu tidak terjadi. Akibat dari rangking-rangking yang sama adalah mengubah variabilitas himpunan rangking itu. Dengan demikian, koreksi untuk angka sama ditetapkan pada deviasi standar distribusi sampling statistik Mann-Whitney sebagai berikut :

$$\sigma_U = \sqrt{\left(\frac{n_1 n_2}{N(N-1)}\right)\left(\frac{N^3 - N}{12} - \Sigma T\right)}$$

keterangan :

$N = n_1 + n_2$

$T = \frac{t^3 - t}{12}$ (di mana t banyak observasi yang berangka sama untuk suatu rangking tertentu.)

ΣT = harga-harga T semua kelompok yang memiliki observasi-observasi berangka sama.

Untuk menghitung nilai statistik uji hasil pengamatan, kedua sampel digabungkan dan memeringkatkan semua hasil pengamatan dalam sampel tersebut dari yang paling kecil hingga yang paling besar. Hasil-hasil pengamatan dengan nilai-nilai yang sama diberi peringkat yang sama dengan rata-rata dari posisi-posisi peringkat yang semestinya andaikata kasus angka sama tidak terjadi. Kemudian peringkat-peringkat hasil pengamatan dijumlahkan dari masing-masing populasi 1 dan populasi 2. selanjutnya ditentukan statistik uji untuk masing-masing populasi sebagai berikut :

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ (dari populasi 1)}$$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \text{ (dari populasi 2)}$$

keterangan :

R_1 = Jumlah peringkat hasil-hasil pengamatan yang merupakan sampel dari populasi 1.

R_2 = Jumlah peringkat hasil-hasil pengamatan yang merupakan sampel dari populasi 2.

n_1 = Jumlah pengamatan pada sampel pertama.

n_2 = Jumlah pengamatan pada sampel kedua.

Uji Peringkat Bertanda Wilcoxon

Uji peringkat bertanda Wilcoxon digunakan jika besaran maupun arah perbedaan relevan untuk menentukan apakah terdapat perbedaan yang sesungguhnya antara data yang satu dengan data yang lainnya. Uji peringkat bertanda Wilcoxon tidak hanya memanfaatkan informasi tentang arah tetapi juga besarnya perbedaan pasangan nilai itu.

Simulasi Dua Sampel Saling Bebas

Untuk tiap tipe sebaran data (Normal, Kai-kuadrat, Gamma, dan Seragam) dilakukan sebanyak 1000 simulasi untuk membandingkan prosedur Mann-Whitney dan uji t independen. Tiap simulasi dibangkitkan data X berukuran n_1 = 12 dan data Y berukuran n_2 = 15.

Prosedur uji Mann-Whitney adalah sebagai berikut :

1. Pembangkitan data independen X dan Y dengan ukuran data yang tidak sama.
2. Melakukan pemeringkatan gabungan dua kelompok data dari 1 sampai dengan $N = n_1 + n_2$.
3. Menentukan harga U (Statistik Mann-Whitney).
4. Pengambilan keputusan dengan menolak atau menerima H_0.
   Jika harga statistik U mempunyai kemungkinan yang sama besar atau lebih kecil dari $W_{\alpha/2}$, maka H_0 ditolak dan sebaliknya. Nilai $W_{\alpha/2}$ dapat dilihat pada tabel A yang terdapat pada Lampiran 1.

Tabel 1. Perbandingan Hasil Analisis Statistik Parametrik (uji t independen) dan Statistik Nonparametrik (uji Mann-Whitney).

Distribusi	n_1	n_2	Perbedaan kesimpulan
Normal dengan σ^2 sama dan μ berbeda dimana $\sigma_1^2 = 10$ $\sigma_2^2 = 10$ $\mu_1 = 200$ $\mu_2 = 300$	12	15	Tidak ada perbedaan
Normal dengan σ^2 dan μ berbeda dekat (sama) dimana $\sigma_1^2 = 10$ $\sigma_2^2 = 10$ $\mu_1 = 245$ $\mu_2 = 250$	12	15	34 perbedaan
Normal dengan σ^2 berbeda dan μ sama dimana $\sigma_1^2 = 10$ $\sigma_2^2 = 15$ $\mu_1 = 200$ $\mu_2 = 200$	12	15	11 perbedaan
Normal dengan σ^2 dan μ berbeda dimana $\sigma_1^2 = 10$ $\sigma_2^2 = 15$ $\mu_1 = 100$ $\mu_2 = 300$	12	15	Tidak ada perbedaan
Kai-kuadrat (0,30)	12	15	28 perbedaan
Gamma (0,100)	12	15	54 perbedaan
Seragam (0,100)	12	15	3 perbedaan

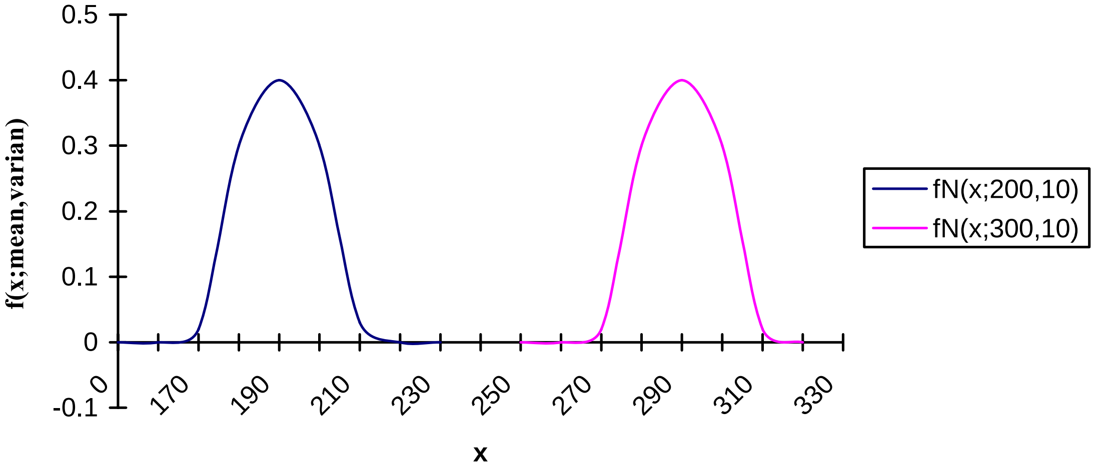
Pada tabel di atas dapat dilihat bahwa untuk data berdistribusi normal dengan asumsi varian sama dan mean kedua populasi berbeda tidak terdapat perbedaan kesimpulan yang mana kesimpulan yang diperoleh uji Mann-Whitney mengikuti kesimpulan yang diperoleh uji t. Dengan kata lain apabila uji t menolak H_0 maka kesimpulan dari uji Mann-Whitney juga menolak H_0. Hal ini menunjukkan bahwa uji Mann-Whitney dapat juga dipakai untuk data berdistribusi normal dengan asumsi varian sama dan mean berbeda (Lampiran 4). Sedangkan untuk data yang berdistribusi normal dengan asumsi varian dan mean kedua populasi sama terdapat 34 perbedaan pada kesimpulannya. Hal ini menunjukkan bahwa uji t lebih baik dari uji Mann-Whitney pada kondisi data berdistribusi normal dengan asumsi varian dan mean populasi sama (Lampiran 5).

Untuk data berdistribusi normal dengan varian populasi yang berbeda dan mean populasi sama hanya 11 kesimpulan dari uji Mann-Whitney yang tidak mengikuti kesimpulan yang diperoleh uji t. Ini berarti uji t tetap lebih baik dari pada uji Mann-

Whitney yaitu sesuai dengan jenis datanya yaitu berdistribusi normal (Lampiran 6). Pada data berdistribusi normal dengan varian yang berbeda dan mean populasi berbeda dimana tidak ada perbedaan sama sekali dalam pengambilan kesimpulan antara kesimpulan yang diperoleh uji t dengan kesimpulan yang diperoleh uji Mann-Whitney. Artinya, uji Mann-Whitney dapat dipakai pada kondisi data berdistribusi normal dengan varian yang berbeda dan mean berbeda (Lampiran 7).

Dari uraian di atas menunjukkan bahwa ada pengaruh dalam pemilihan parameter-parameter yang digunakan, dimana semakin jauh beda varian dan mean populasi 1 dengan varian dan mean populasi 2 maka semakin sedikit perbedaan yang terjadi antara kedua uji tersebut.

Pengaruh mean (μ) dan Varian (σ^2) dapat diilustrasikan sebagai berikut :



Gambar 2. Pengaruh Mean dan Varian Pada Distribusi Normal

Statistik nonparametrik dengan menggunakan uji Mann-Whitney untuk data yang berdistribusi Kai-kuadrat terdapat 28 perbedaan kesimpulan dengan uji t dimana diantara 1000 hasil simulasi data tersebut hanya sedikit saja kesimpulan dari uji t yang tidak mengikuti kesimpulan yang diperoleh uji Mann-Whitney. Hal ini menunjukkan bahwa uji Mann-Whitney lebih baik daripada uji t sesuai dengan jenis data yaitu data bukan berasal dari distribusi normal atau data berdistribusi Kai-kuadrat (Lampiran 8).

Hasil analisis untuk data simulasi berdistribusi Gamma dengan menggunakan uji Mann-Whitney terdapat 54 perbedaan kesimpulan dengan kesimpulan yang diperoleh uji t. Hal ini berarti uji t tidak terlalu baik digunakan untuk data besdistribusi Gamma (Lampiran 9).

Uji Mann-Whitney dan uji t independen memberikan hasil yang tidak terlalu jauh berbeda untuk data simulasi yang berdistribusi seragam, yaitu antara kesimpulan yang diperoleh uji Mann-Whitney dengan uji t hanya terdapat 3 perbedaan pengambilan kesimpulan sehingga uji t tersebut dapat dipakai untuk manganalisis data simulasi berdistribusi seragam (Lampiran 10).

Simulasi Dua Sampel Saling Berpasangan

Untuk masing-masing tipe sebaran data (Normal dan Seragam) dilakukan sebanyak 500 simulasi untuk membandingkan prosedur uji peringkat bertanda

Wilcoxon dan uji t berpasangan. Tiap simulasi dibangkitkan data X_1 dan X_2 yang berukuran sampel sebanyak n = 12.

Prosedur uji peringkat bertanda Wilcoxon adalah sebagai berikut :

1. Pembangkitan data berpasangan X_1 dan X_2 dengan ukuran sampel yang sama.
2. Memberikan rangking atau peringkat untuk harga-harga d dengan memperhatikan tandanya.
3. Menetapkan harga T (yaitu jumlah yang lebih kecil dari kedua kelompok rangking yang memiliki tanda yang sama).
4. Menetapkan nilai N, yaitu banyak total kedua harga d yang memiliki tanda.
5. Melakukan pengambilan kesimpulan dengan menolak H_0 atau menerima H_0 dengan kriteria pengujian sebagai berikut:
   H_0 ditolak jika harga T observasi < nilai T tabel dan H_0 diterima jika harga T observasi > nilai T tabel.

Tabel 2. Perbandingan Hasil Analisis Statistik Parametrik (uji t berpasangan) dan Statistik nonparametrik (uji peringkat bertanda Wilcoxon).

Distribusi	n	Perbedaan kesimpulan
Normal dengan σ^2 sama dan μ berbeda dekat dimana $\sigma_1^2 = 9$ $\sigma_2^2 = 9$ $\mu_1 = 100$ $\mu_2 = 103$	12	16 perbedaan
Normal dengan σ^2 sama dan μ berbeda jauh dimana $\sigma_1^2 = 9$ $\sigma_2^2 = 9$ $\mu_1 = 100$ $\mu_2 = 115$	12	10 perbedaan
Normal dengan σ^2 berbeda dan μ berbeda dekat dimana $\sigma_1^2 = 9$ $\sigma_2^2 = 25$ $\mu_1 = 100$ $\mu_2 = 103$	12	7 perbedaan
Seragam (0, 100)	12	6 perbedaan
Seragam (0, 100) dan (100, 200)	12	Tidak ada perbedaan
Seragam (0, 75) dan (25, 100)	12	41 perbedaan

Berdasarkan tabel 2 terlihat bahwa dari data yang berdistribusi normal dengan asumsi varian sama dan mean populasi berbeda dekat diperoleh 16 perbedaan kesimpulan antara uji peringkat bertanda Wilcoxon dan uji t berpasangan. Hal ini menunjukkan bahwa uji t berpasangan lebih baik dari uji peringkat bertanda Wilcoxon untuk kondisi data berdistribusi normal dengan varian populasi sama dan mean berbeda dekat (Lampiran 11).

Kesimpulan yang diperoleh uji peringkat bertanda Wilcoxon untuk data berdistribusi normal dengan varian populasi sama dan mean populasi berbeda jauh hanya 10 kesimpulan yang tidak mengikuti uji t berpasangan. Ini berarti antara uji t berpasangan dan uji peringkat bertanda Wilcoxon tetap lebih baik uji t berpasangan (Lampiran 12).

Pada data simulasi yang berdistribusi normal dengan asumsi varian berbeda dan mean populasi dekat, kesimpulan antara uji t dengan uji peringkat bertanda Wilcoxon hanya terdapat 7 perbedaan. Hal ini menunjukkan bahwa uji peringkat bertanda

Wilcoxon hampir seefisien uji t berpasangan untuk yang berasumsi varian populasi berbeda dan mean populasi berbeda dekat (Lampiran 13).

Uji statistik nonparametrik dengan menggunakan uji peringkat bertanda Wilcoxon untuk data yang berdistribusi Seragam (0,100) terdapat 6 perbedaan kesimpulan dengan kesimpulan uji statistik parametrik dengan menggunakan uji t berpasangan. Hal ini berarti diantara 500 hasil simulasi data hanya 6 kesimpulan dari uji t yang tidak mengikuti kesimpulan dari uji peringkat bertanda Wilcoxon.

Uji peringkat bertanda Wilcoxon dan uji t berpasangan memberikan hasil yang tidak ada perbedaan sama sekali untuk data yang berdistribusi Seragam (0,100) dan (100,200) sehingga kedua uji tersebut dapat dipakai untuk menganalisis data simulasi berdistribusi Seragam (0,100) dan (100,200) karena kedua nilai parameter dari populasi 1 dan populasi 2 tidak saling berkaitan (Lampiran 15).

Dari simulasi untuk data yang berdistribusi Seragam (0,75) dan (25,100) dengan menggunakan uji peringkat bertanda Wilcoxon diperoleh hasil sebanyak 41 perbedaan kesimpulan dengan kesimpulan uji t berpasangan. Jadi pada kondisi data yang berdistribusi Seragam (0,75) dan (25,100) sudah terlihat jelas bahwa uji peringkat bertanda Wilcoxon jauh lebih baik dari uji t berpasangan (Lampiran 16).

Statistik parametrik dengan menggunakan uji t dapat dilakukan jika beberapa persyaratan terpenuhi, diantaranya adalah populasi yang dianalisis harus berdistribusi normal. Sehingga pada data yang berdistribusi normal, uji t lebih baik dari pada uji statistik nonparametrik dengan menggunakan uji Mann-Whitney atau uji peringkat bertanda Wilcoxon.

Statistik nonparametrik dengan menggunakan uji Mann-Whitney atau uji peringkat bertanda Wilcoxon tidak mengasumsikan bentuk distribusi data sehingga pada data yang bukan berdistribusi normal, uji nonparametrik lebih baik daripada uji parametrik karena statistik nonparametrik tidak harus memakai suatu parameter tertentu, seperti keharusan adanya Mean, Standar Deviasi, Varian, dan lainnya. Seandainya pada uji nonparametrik digunakan pada data berdistribusi normal, bisa saja mengakibatkan terjadi pelanggaran asumsi. Pelanggaran asumsi dapat berakibat fatal terhadap kesimpulan yang diambil.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil simulasi data dan pembahasan sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Untuk data yang diketahui bentuk distribusinya yaitu berdistribusi normal, uji parametrik dengan menggunakan uji t memberikan hasil yang lebih baik daripada uji nonparametrik dengan menggunakan uji Mann-Whitney dan uji peringkat bertanda Wilcoxon.
2. Untuk data yang bukan berdistribusi normal (Kai-kuadrat, Gamma, dan Seragam), uji nonparametrik dengan menggunakan uji Mann-Whitney dan uji peringkat bertanda Wilcoxon memberikan hasil yang lebih baik daripada uji parametrik dengan uji t.

Saran

Penelitian ini dibatasi pada uji statistik nonparametrik, yaitu uji Mann-Whitney dan uji peringkat bertanda Wilcoxon dan aplikasinya untuk beberapa nilai parameter.

Untuk penelitian selanjutnya disarankan mengkaji mengenai uji Mann-Whitney dan uji peringkat bertanda Wilcoxon untuk nilai parameter-parameter yang lain.

DAFTAR PUSTAKA

Anonim. 2007. *Teknik dan Metode Analisis Data.* http://www.euronet.nl/users/warnar/demostatistiek/stat/mwutwo.htm

Bungin, B. 2005. *Metodologi Penelitian Kuantitatif : Komunikasi, Ekonomi dan Kebijakan Publik Serta Ilmu-Ilmu Sosial Lainnya*. Jakarta : PT Gramedia Pustaka Utama.

Daniel, W. 1989. *Statistik Nonparametrik Terapan*. Jakarta : PT Gramedia.

Djarwanto, P.S. 2001. *Mengenal Beberapa Uji Statistik Dalam Penelitian*. Surakarta : Universitas Sebelas Maret.

Harinaldi. 2005. *Prinsip-Prinsip Statistik Untuk Teknik dan Sains*. Jakarta : Erlangga.

Murti, B. 1996. *Penerapan Metode Statistik Nonparametrik Dalam Ilmu-Ilmu Kesehatan.* Jakarta : PT Gramedia Pustaka Utama.

Nasoetion, A.H. dan A.R. Rambe. 1984. *Teori Statistika untuk Ilmu-ilmu Kuantiatif.* Jakarta : Bhratara Karya Aksara.

Runyon, R.P. and A. Haber. 1989. *Fundamental Of Behavioral Statistics*, (Singapore: Mc GRAW – HILL International Edition, 1989).

Saleh, S. 1996. *Statistik Nonparametrik*. Yogyakarta: Universitas Gadjah Mada.

Sanders, D.H. 1990. *Statistics : A Fresh Approach*, (Singapore: Mc GRAW – HILL International Edition).

Siegel, S. 1997. *Statistik Nonparametrik Untuk Ilmu-Ilmu Sosial*. Jakarta : PT Gramedia.

Spigel, M. 2004. *Statistik.* Terjemahan. Jakarta : Penerbit Erlangga.

Supranto, J. 1989. *Statistik : Teori dan Aplikasi*, Jakarta: PT Gelora Aksara Pratama.