

METODE ESD (*EXTREME STUDENTIZED DEVIATION*) UNTUK PENGUJIAN *MULTIPLE OUTLIERS*

Oleh:

Masta Hutajulu

Pendidikan Matematika, STKIP Siliwangi Bandung
mastahutajulu@yahoo.com

ABSTRAK

Analisis data pada suatu proses produksi merupakan hal yang esensial untuk dilakukan. Data yang valid menjadikan data tersebut baik. Salah satu ciri data yang baik adalah terhindar dari pencilan (*outliers*). Untuk mendeteksi pencilan (*outlier*) yang terdapat pada suatu data diperlukan suatu metode, yang antara lain melalui metode ESD. Metode ESD dapat mendeteksi pencilan tunggal (*single outliers*) ataupun pencilan ganda (*multiple outliers*). Konsep utama dari metode ini, adalah mengukur jarak setiap data terhadap mean sampelnya.

Kata Kunci : Metode ESD, pencilan

ABSTRACT

Data analyse of a production process is an essential part. The accuracy of the data analyse depends on the validity of the data. One indicator of valid data is evitable from outliers. It requires certain method to examine the outliers from data to decide whether the data is valid or not. ESD method is the recommended method, which will be studied thoroughly in this final report. ESD method is a method that is able to detect single outlier as well as multiple outliers. The main concept of the method is measuring the difference of each data to the mean of its sample.

Keywords: Outlier, ESD Method

I. PENDAHULUAN

Berbagai masalah dalam proses sebuah produksi dihadapi produsen. Analisis data pada suatu proses produksi menjadi penting dilakukan supaya pemaksimalan hasil produksi yang terkait dengan pendapatan dapat diperoleh. Produk dikatakan memberi hasil maksimum jika hasil produksi bernilai jual (baik) atau sebaliknya.

Jika selanjutnya baik buruknya kualitas barang dipermasalahkan, maka dibutuhkan analisis proses produksi. Produk yang dihasilkan terjamin dan laku dipasaran akan berdampak pada peningkatan kualitas proses produksi. Disamping itu dapat saja muncul pertanyaan lain misalnya bagaimanacara yang efektif untuk meningkatkan kualitas proses yang berdampak pada pengurangan kerugian yang dialami perusahaan.

Hal-hal di atas membuat perlu ditempuh berbagai proses untuk mendapatkan barang baik dalam fisik maupun jasa dengan kualitas yang baik. Faktor dominan disini adalah melakukan kontrol terhadap proses produksi secara berkala dan analisis data yang keat didasarkan pada data yang valid, yang terbebas dari data yang menyimpang seperti pencilan (*outliers*) agar tingkat kesalahan dalam pengambilan keputusan dapat dikurangi. Analisis data pencilan cukup menarik untuk dikaji dan penggunaan metode isini adalah metode ESD.

Metode ESD merupakan metode untuk pengujian atau pendeteksian *outliers*. Konsep utama dari metode ini adalah mengukur jarak tiap data untuk membangun statistik penguji ESD. Distribusi ESD dihasilkan dengan cara menurunkan Distribusi Beta.

Identifikasi Masalah

Agar mendapatkan mutu data yang baik maka data harus dianalisis dengan prosedur yang tepat. Data yang baik haruslah merupakan data yang valid. Salah satu ciri bahwa dikatakan baik adalah terbebas dari data yang menyimpang yaitu data yang tidak terdapat pencilan (*outliers*).

Peminimalan kesalahan dalam pengambilan keputusan inilah yang menjadi salah satu indikator “baik-tidaknya” suatu data. Untuk mendeteksi suatu pencilan maka diperlukan statistik penguji (*statistical test*). Pada sekelompok data mungkin saja terjadi pencilan tunggal (*single outliers*) ataupun pencilan ganda (*multiple outliers*) merupakan data yang menyimpang sifatnya dari data lainnya. Dan penyimpangan tersebut berakibat yang tidak baik bagi data keseluruhan. Pengujian terhadap pencilan tunggal (*single outliers*) maupun pencilan ganda (*multiple outliers*) dapat dilakukan jika kita memiliki statistik penguji.

Disini akan diperkenalkan statistik penguji ESD, untuk mendeteksi keberadaan sejumlah pencilan (*outliers*) pada data dengan asumsi bahwa data berdistribusi Normal. Dan metode ESD merupakan metode yang relative mudah digunakan karena dapat dipakai pada *software* Excel yang pada umumnya banyak ditemui.

Tujuan Penelitian

Tujuan dari penelitian ini adalah memahami secara lebih baik metode ESD dan penggunaan *software* Excel dan SPSS dalam analisis data pencilan (*outliers*).

Batasan Masalah

Batasan pembahasan dalam penulisan laporan ini adalah bagaimanakah penentuan titik kritis berdasarkan Distribusi statistik penguji ESD untuk kasus pencilan ganda (*Multiple outliers*) dengan penerapan algoritma perancangan *software* Excel dan

SPPS untuk mencari titik kritis. Titik kritis tersebut adalah : $c = \sqrt{G^{-1}(\sqrt[m]{1-\alpha})}$.

Dalam hal ini G berdistribusi Beta, m adalah banyaknya data, α tingkat kepercayaan (*level of significant*). Karena G adalah fungsi berdistribusi Beta maka tidak mudah mencari solusi invers G untuk c sehingga algoritma perancangan *software* komputer sangat diperlukan untuk mencari G^{-1} (G invers).

II. ALGORITMA PENENTUAN TITIK KRITIS

Proses penentuan titik kritis untuk kasus penelitian tunggal (*single outliers*) dan kasus pencilan ganda (*multiple outliers*) akan dibahas pada bab berikut ini.

1. Kasus Penelitian Tunggal (*single outliers*)

Pada ulasan sebelumnya dijelaskan bahwa formula titik kritis yang digunakan adalah:

$$c = \sqrt{G^{-1}(\sqrt[m]{1-\alpha})}.$$

Dengan demikian untuk mencari nilai-nilai c , dapat digunakan perangkat lunak excel dengan langkah-langkah sebagai berikut:

- 1) Masuk pada *microsoft* Excel
- 2) Pada kolom A dalam hal ini A, tuliskan nilai-nilai m yang dikehendaki, yaitu ukuran sampel
- 3) Letakkan kursor di kolom B, hitung nilai akar pangkat m dari $(1-\alpha)$ dengan menggunakan perintah $1.^{(1/A_1)} = (1-\alpha)^{(1/A_1)}$. Ini dilakukan untuk menghitung satu nilai m saja. Untuk nilai m selanjutnya kita tinggal membuat copy perintahnya saja dengan cara memposisikan kursor mouse di layar komputer pada ujung kanan bawah sehingga tandanya berubah menjadi sebuah tanda (+), kita drag ke bawah sampai dengan nilai m yang terakhir yang kita kehendaki.
- 4) Setelah diperoleh nilai-nilai dari no.2 dan no.3, letakkan kursor pada kolom c, klik insert, function, beta invers. Akan muncul kotak untuk menghitung nilai dari beta invers tersebut.
- 5) Isilah baris pertama (pada kotak komunikasi tersebut) dengan nilai-nilai pada kolom B. Untuk mempercepat proses tulisan B, (apabila nilai di kolom B mulai pada baris ke-1, baris ke-2 isi dengan 0,5 dan baris ke-3 dengan $(A_1 - 2)/2$, baris ke-4 isi dengan nol dan baris ke-5 isi dengan 1, tekan enter, maka nilai c^2 akan tampak pada kolom c pada baris pertama (c_1). Untuk menghitung nilai c^2 yang lainnya cukup kita men-drag seperti pada no.3.
- 6) Selanjutnya pada kolom D dihitung titik c yang dicari dengan menggunakan perintah $'=sqrt(c_1)'$ pada baris pertama (D_1). Untuk menghitung nilai titik yang lainnya, cukup kita men-drag seperti no.3 atau no.5.

2. Kasus pencilan Ganda (*Multiple outliers*)

Memperhatikan bahasan pada kasus *single outliers* maka pada kasus *multiple outliers* dimisalkan ada k buah data yang dicurigai sebagai *outliers*. Langkah pengujian yang dilakukan adalah :

$$1) w_{(1)} \text{ dibandingkan dengan titik } c_1 = \sqrt{G_1^{-1}({}^m\sqrt{1-\alpha})}.$$

$$2) w_{(2)} \text{ dibandingkan dengan titik } c_2 = \sqrt{G_2^{-1}({}^m\sqrt{1-\alpha})}.$$

.....

$$n). w_{(k)} \text{ dibandingkan dengan titik } c_k = \sqrt{G_k^{-1}({}^m\sqrt{1-\alpha})}.$$

Cat: G_i adalah fungsi distribusi dari Beta ($1/2, (m-i-1)/2$)

III. STUDY KASUS DAN APLIKASI METODE ESD DENGAN PROGRAM KOMPUTER

Penggunaan metode ESD untuk pengujian pencilan Ganda (*multiple outlier*) akan dibahas berikut ini. Data yang digunakan ada 2 (dua) macam, yaitu data hipotesis (Rosner, 1983) dan data riil yang penulis peroleh dari Jurusan Farmasi ITB (Suryadi, 1994).

Sebelum data tersebut analisis, terlebih dahulu diuji distribusi dari setiap data tersebut. Pengujian konormalan dilakukan dengan uji *Kolmogorov-Smirnov*. Uji Normalitas dari setiap data dapat dilihat pada lampiran.

Contoh penerapan I

Data berikut ini merupakan data hipotesis yang penulis gunakan agar lebih mudah untuk memahami dan mengerti bagaimana penerapan metode ESD yang telah dijelaskan pada bab-bab sebelumnya untuk data yang ada.

Diberikan data sebagai berikut :

Tabel 1 : Data hipotesis

Obs ke-1	Nilai
1	2,1
2	2,6
3	2,4
4	2,5
5	2,3

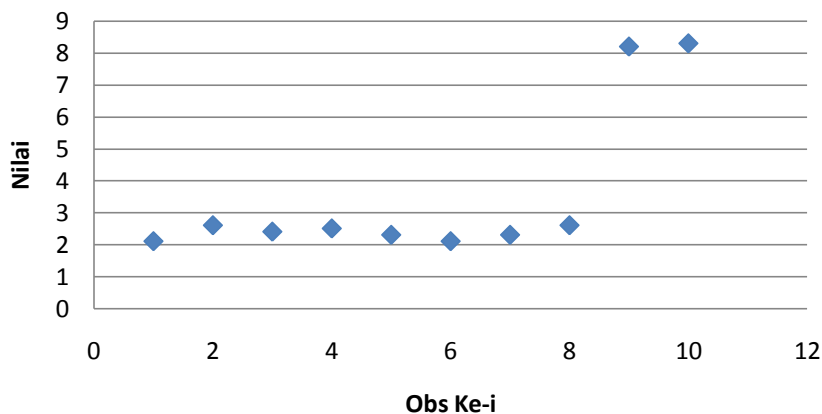
Obs ke-1	Nilai
6	2,1
7	2,3
8	2,6
9	8,2
10	8,3

$\bar{x} = 3,54$ (mean sampel / rata-rata sampel)

$s = 2,48873$ (simpangan baku sampel)

Terhadap data diatas akan dilakukan pengujian data ekstrim yang terkandung didalamnya apakah merupakan *outliers* atau bukan. Secara sepintas dari data diatas dapat saja diduga bahwa data ke-9 dan ke-10 yaitu 8,2 dan 8,3 adalah *outliers*.

Untuk lebih jelas dapat dilihat dalam grafik dibawah ini. Terlihat bahwa 2 data terakhir yaitu data ke-9 dan data ke-10 terletak jauh dari data lainnya.



Grafik 1: Plot nilai terhadap nomor observasi

Kita akan menguji hipotesis tersebut, kita tentukan :

H_0 : 2 Data ekstrim adalah bukan *outliers*

H_1 : 2 Data ekstrim adalah *outliers*

Kemudian dilakukan uji satu demi satu H_0 dengan mengetes *outliers* sebagai berikut:

1. Lakukan perhitungan tiap tahap dengan sampel yang telah berkurang satu setelah data yang memiliki statistik V_i terbesar *dihapus*

Tabel 2 : Tahapan Pengerjaan pada Data hipotesis

Obs ke-i	x_i	Tahap Pertama		Tahap kedua		Tahap ketiga	
		$ x_i - \bar{x} /s$	v_i	$ x_i - \bar{x} /s$	v_i	$ x_i - \bar{x} /s$	v_i
1	2,1	0,5786	0,2033	0,4661	0,1748	1,3154	0,5315
2	2,6	0,3777	0,1327	0,2103	0,0789	1,1902	0,4809
3	2,4	0,4581	0,1609	0,3126	0,1172	0,1879	0,0759
4	2,5	0,4179	0,1468	0,2615	0,0980	0,6890	0,2784
5	2,3	0,4982	0,3638	0,3638	0,1364	0,3132	0,1265
6	2,1	0,3777	0,2033	0,4661	0,1749	1,3154	0,5315
7	2,3	0,4982	0,1751	0,3638	0,1364	0,3132	0,1265
8	2,6	0,3777	0,1327	0,2103	0,0789	1,1902	0,4809
9	8,2	1,8724	0,6579	2,6545	0,9954		
10	8,3	1,9126	0,6720				
Maks v_i			0,6720		0,9954		0,4809
W_{tabel}			0,80347		0,82947		0,8582
\bar{x}		3,54		3,01111		2,36250	
S		2,48873		1,95477		0,19955	

Dengan $v_i = \sqrt{U_i}$ dan $U_i = \frac{m(x_i - \bar{x})^2}{(m-1)s^2}$

2. Perhatikan data pada Tabel 2

- Tahap pertama, Pengujian untuk data yang memiliki jarak terjauh dari mean sampel yang dalam hal ini statistik pengujian adalah $W_1 = V_{(10)}$. Dari perhitungan diperoleh $W_{tabel} = 0,80347$ untuk $m = 10$ dan $\alpha = 5\%$ maka $W_1 < W_{tabel}$.
- Tahap kedua, setelah data yang memiliki jarak jauh dari mean sampel dari tahap pertama kita *hapus*, kita lakukan pengujian dengan mencari lagi data yang memiliki jarak terbesar terhadap mean sampel statistik pengujiannya ialah $W_2 = V_{(9)}$. Dari perhitungan diperoleh $W_{tabel} = 0,829469$ untuk $m = 9$ dan $\alpha = 5\%$ maka $W_2 > W_{tabel}$.

- c. Tahap ketiga, setelah data yang memiliki jarak terjauh dari mean sampel dari tahap kedua kita *hapus*. Kita lakukan pengujian lagi dengan mencari data yang memiliki jarak terbesar terhadap mean sampel dan diperoleh statistik pengujian adalah $W_3 = V_{(8)}$. Dari perhitungan diperoleh $W_{tabel} = 0,858202$ untuk $m = 8$ dan $\alpha = 5\%$ maka $W_3 < W_{tabel}$.
3. Hasil perhitungan menunjukkan bahwa $W_1 < W_{tabel}$, $W_2 > W_{tabel}$, $W_3 < W_{tabel}$ dengan indeks $i=2$ adalah indeks terbesar sehingga $W_i > W_{tabel}$. Artinya data yang memiliki jarak jauh dengan mean sampel pada tahap ke-k dan sebelumnya merupakan *outliers*. Dalam kasus ini disimpulkan bahwa 8,3 dan 8,2 adalah *outliers*.

IV. KESIMPULAN DAN SARAN

1. Kesimpulan

Ada beberapa kesimpulan yang dapat diambil :

- a. Adanya pencilan (*outliers*) sangat mengganggu terhadap analisis statistik. Apabila hal ini dibiarkan maka dapat mengakibatkan kerugian yang tidak sedikit terhadap masalah kualitas barang produksi. Hal ini disebabkan informasi yang kita terima dari data proses produksi yang kita analisa. Melenceng jauh dari yang sebenarnya sehingga dapat mengakibatkan pengambilan keputusan yang salah. Oleh sebab itu data-data yang analisa sebaiknya terbebas dari pencilan (*outliers*).
- b. Pencilan tunggal (*Single outlier*) ataupun pencilan ganda (*multiple outliers*) pada sebuah data merupakan hal yang esensial untuk dideteksi. Salah satu metode untuk mendeteksi pencilan (*outlier*) adalah metode ESD (*Extreme Studentized Deviation*).
- c. Metode ESD merupakan sebuah metode yang mampu mendeteksi pencilan tunggal (*Single Outlier*) ataupun pencilan ganda (*Multiple outliers*).
- d. Pengujian terhadap beberapa *outliers* sangat diperlukan. Dan metoda pengujiannya bias secara satu persatuan tetapi dengan prosedur yang telah dijelaskan pada bab sebelumnya.

2. Saran

Mengingat pentingnya mengetahui kehadiran *outliers* dalam data, maka disarankan:

- a. Pada metode ESD pendeteksian *outliers* dilakukan dengan memeriksa satu persatu yang dianggap sebagai *outliers*. Metode tersebut cenderung lama. Oleh karena itu diperlukan metode lain sebagai pembandingan yang dapat mendeteksi *outliers* sekaligus.
- b. Program yang disajikan oleh penulis juga masing dapat disempurnakan lagi dengan menggunakan *software* (bahasa Pemrograman) yang lainnya sehingga mempermudah penentuan titik kritis tersebut.

DAFTAR PUSTAKA

- Anggraeni, N. (2001). *Distribusi Eksak Statistik Penguji ESD untuk Pendeteksian Single Outliers pada Sampel Normal Univariate*. Tugas Akhir Institut Teknologi Bandung (ITB).
- Djauhari, M.A. (1998). *A Unifying Concept of X-chart and X-barchart when subgroup sizes are Equal*. Departement of Mathematics Institut Teknologi Bandung, Proceedings ITB 30(1)
- Herryanto, N. (1995). *Satistika Matematika*. Universitas Pendidikan Indonesia Bandung.
- Hogg, V & Craig, T. (1995). Fifth Edition. *Introduction to Mathematical Statistics*. New Jersey: Prentice Hall.
- Rosner, B. (1983). *Selected Outlier Identification Test-Generalized ESD many Outliers Proceure*. Technometrics, Vol. 25 No. 2.
- Suryadi, H. (1994). *Pemisahan dan Penetapan Kadar bberapa Komponen Obat Influenza dalam Sediaan Tablet dengan Metoda Kromatografi Cair Kinerja Tinggi*. Tesis Jurusan farmasi Institut Teknologi Bandung.