



PANORAMA

JURNAL EKONOMI, PENDIDIKAN & PARIWISATA

NUSANTARA

MENINGKATKAN KUALITAS DAN KREATIVITAS

EDISI Ke-6

Januari - Juni 2009

EFEKTIFITAS PENGELOLAAN DIKLAT PARIWISATA

(Studi Peningkatan Efektifitas, Kinerja, Mutu dan Kepuasan Pelanggan pada Pusat Pelatihan dan Sertifikasi Kepariwisataan (PPSK) Dinas Pariwisata Provinsi DKI Jakarta)

Oleh : *Witarsa Tambunan*
(Dosen STEIN, Jakarta) 1 - 20

PROGRESS REPORT PROGRAM PEMBERDAYAAN MASYARAKAT KELURAHAN (PPMK) TAHUN 2002 - 2007 KELURAHAN DUREN SAWIT, KOTA ADMINISTRASI JAKARTA TIMUR

(Suatu Laporan Penelitian : 187 Rukun Tetangga (RT) Sekelurahan Duren Sawit)

Oleh : *Imam Subarkah*
(Dosen STEIN, Jakarta) 21 - 30

HUBUNGAN MOTIVASI DAN KOMPETENSI DENGAN PRESTASI KERJA KARYAWAN PERUSAHAAN KONTRAKTOR

Oleh : *Mardiaman*
(Dosen Universitas Mpu Tantular, Jakarta) 31 - 38

POSISI TRANSPORTASI DALAM PARIWISATA

Oleh : *Nani Tambunan*
(Dosen Universitas Mpu Tantular, Jakarta) 39 - 48

ANALISIS SISTEM PENGENDALIAN YANG EFEKTIF

Oleh : *Darmadi A. Karim*
(Dosen STEIN, Jakarta) 49 - 55

IDENTIFIKASI DAN PENGARUH KEBERADAAN DATA PENCILAN (OUTLIER)

(Studi Kasus Jumlah Kunjungan Wisman dan Pengunjung Asing ke Indonesia Melalui Pintu Masuk Makasar Antara Bulan Januari 2007 s.d. Juli 2008)

Oleh : *Salman Paludi*
(Dosen STEIN, Jakarta) 56 - 62

Diterbitkan oleh

SEKOLAH TINGGI ILMU EKONOMI PARIWISATA INTERNASIONAL

JL. RAYA KALIMALANG NO. 2A JAKARTA TIMUR, 13620

Tlp. (021) 8629472 fax. 8629471

www.panoramanusantara.wordpress.com



SUSUNAN REDAKSI

Pembina

*P. Silitonga, MBA (Ketua Yayasan LBPP)
Beryamin Sopaheluwakan, SE (Ketua STEIN)*

Pemimpin Umum/Penanggung Jawab

Bonifasius MH. Natunggolan, S.Si

Mitra Bestari

Parulian Sihotang, Ph.D

Pemimpin Redaksi

Salman Paludi, S.Si

Anggota Redaksi

Prof. Dr. H. Siti Ibrahim Boechari, M.Sc

Edy Karyadi, S.Si

Dra. Yanhi Setyawan

Drs. Jamotari Harefa, M.M

Ir. Mardiaman, M.T

H. Darmadi A. Karim, SH, MM

Meylani Tuti, S.S

H. Supardi, SE

Pelaksana Teknis

Soerjanto

Setyo Widianto

Rosmaya Silitonga

Estu Pujianto

Sekretariat Redaksi

*Kampus Sekolah Tinggi Ilmu Ekonomi Pariwisata
Internasional (STEIN)*

*Jl. Raya Kalimantan No. 2A Jakarta Timur 13620,
telp. 8629472 fax. 8629471*

www.panoramanusantara.wordpress.com

PENGANTAR REDAKSI

Pembaca Panorama Nusantara yang budiman,

Puji syukur kepada Tuhan Yang Maha Esa kami ucapkan karena berkat rahmat dan karunia-Nya-lah kami berhasil menerbitkan edisi keenam majalah ilmiah Panorama Nusantara ini. Panorama Nusantara adalah majalah ilmiah Sekolah Tinggi Ilmu Ekonomi Pariwisata Internasional (STEIN) yang diharapkan dapat menjadi wadah/tempat bagi para dosen untuk meningkatkan kualitas dan kreativitas khususnya dalam bidang penelitian ilmiah.

Pada edisi keenam atau edisi Januari – Juni 2009 kali ini menampilkan tulisan-tulisan ilmiah dari rekan-rekan dosen yang diawali oleh tulisan *Witarsa Tambunan*, disusul oleh tulisan *Imam Subarkah, Mardiaman, Nani Tambunan, Darmadi A. Karim* dan diakhiri oleh tulisan *Salman Paludi*.

Akhir kata, melalui forum redaksi ini kami menunggu "sumbangan" tulisan-tulisan ilmiah selanjutnya dari para pembaca dan pecinta Panorama Nusantara.

Redaksi

Majalah Ilmiah Panorama Nusantara terbit 2 (dua) kali dalam setahun, yaitu edisi Januari – Juni dan edisi Juli – Desember. Majalah Panorama Nusantara terbit pertama kali pada bulan November 2006 untuk edisi Juli – Desember 2006.

IDENTIFIKASI DAN PENGARUH KEBERADAAN DATA PENCILAN (OUTLIER) (Studi Kasus Jumlah Kunjungan Wisman dan Pengunjung Asing ke Indonesia Melalui Pintu Masuk Makasar Antara Bulan Januari 2007 s.d. Juli 2008)

Oleh : Salman Paludi
Dosen STEIN, Jakarta

Abstract

An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs. Outliers, being the most extreme observations, may include the sample maximum or sample minimum, or both, depending on whether they are extremely high or low. However, the sample maximum and minimum are not always outliers because they may not be unusually far from other observations.

Deletion of outlier data is a controversial practice frowned on by many scientists and science instructors; while mathematical criteria provide an objective and quantitative method for data rejection, they do not make the practice more scientifically or methodologically sound, especially in small sets or where a normal distribution cannot be assumed. Rejection of outliers is more acceptable in areas of practice where the underlying model of the process being measured and the usual distribution of measurement error are confidently known. An outlier resulting from an instrument reading error may be excluded but it is desirable that the reading is at least verified.

Key words ; outlier, data, boxplot

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan ilmu pengetahuan dan teknologi pada saat ini semakin maju. Hal ini dilatarbelakangi dengan adanya penelitian dan pengembangan yang dilakukan oleh para peneliti dan ilmuwan. Dalam melakukan penelitian, sipeneliti biasanya akan berhubungan dengan data dan angka-angka dari system yang sedang ditelitinya. Dalam penelitian, peneliti biasanya menggunakan model, suatu hubungan fungsional antara peubah. Dengan model itu kita berusaha memahami, menerangkan, mengendalikan dan kemudian memprediksikan kelakuan system yang kita teliti.

Secara umum, model merupakan penyerdahanaan dan abstraksi dari keadaan alam yang sesungguhnya. Keadaan alam yang ingin diteliti biasanya amat rumit dan kemampuan kita menelitinya secara keseluruhan amat terbatas, karena itu kita perlu menyederhanakannya sesuai dengan akal kita menghadapinya. Dari pengalaman dimasa lalu atau dari dugaan mengenai hubungan antara peubah dalam system yang diteliti, dirumuskan perkiraan kelakuan system tersebut dalam berbagai situasi. Si peneliti mengaharapkan bahwa model tersebut merupakan teori tentang cara kerja system yang ia teliti. Rumusan hubungan tersebut, yang selanjutnya dinyatakan dalam bentuk hipotesis, seterusnya diuji

berdasarkan data statistic yang dikumpulkan kemudian. Model yang dibicarakan disini ialah berbentuk fungsi yang pembentukanya biasanya dengan regresi. Model yang baik menangkap kecenderungan umum yang terdapat pada data. Ada dua hal yang menyebabkan kecocokan model dengan data sampel masih jelek, pertama modelnya belum tepat dan kedua, adanya pencilan yang tidak mengikuti pola umum data.

1.2. Tujuan

Adapun tujuan dari penulisan makalah ini adalah untuk :

- Menjelaskan definisi pencilan.
- Menjelaskan metode-metode yang dapat dipergunakan dalam mengidentifikasi keberadaan pencilan.

1.3. Manfaat

Manfaat yang dapat diperoleh dari penyusunan makalah ini adalah kita dapat memahami tindakan yang perlu dilakukan dalam mengidentifikasi serta menanggulangi keberadaan pencilan dalam data yang akan dianalisis.

1.4. Batasan Masalah

Dalam tulisan ini hanya menjelaskan dan identifikasi apa itu pencilan, namun tidak menjabarkan dampaknya terhadap kecocokan model regresi dan penyebab keberadaan data pencilan pada data kunjungan wisman ke Indonesia melaui pintu Makasar antara bulan Januari 2007 s.d. bulan Juli 2008.

2. TINJAUAN PUSTAKA

2.1. Pencilan

2.1.1. Definisi Pencilan dan Analisis Residual

Umumnya pengamatan yang dicurigai sebagai outlier, influential observations, dan high leverage dikategorikan ke dalam pelanggaran asumsi. Maka lebih tepat jika digunakan analisis residual. Berikut ini adalah beberapa definisi outlier :

1. Ferguson (1961), Outlier adalah suatu data yang menyimpang dari sekumpulan data yang lain.
2. Barnett (1981), Outlier adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data.
3. R.K. Sembiring (1950), Outlier adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi.
4. Weissberg (1985), Jika terdapat masalah yang berkaitan dengan outlier , maka diperlukan alat diagnosis yang dapat mengidentifikasi masalah outlier , salah satunya dengan menyisihkan outlier dari kelompok data kemudian menganalisis data tanpa outlier.

Metoda yang digunakan dalam hubungannya dengan outlier (pencilan), influential observations (pengamatan berpengaruh), dan high leverage (pengaruh tinggi) adalah analisis residual. Residual banyak memegang peranan penting dalam pengujian untuk model regresi karena residual itu sendiri merupakan sisa pada suatu pengamatan .

Residual atau sisaan di definisikan sebagai berikut : $e = y_i - \hat{y}$, $i = 1, 2, \dots, n$. Kendati namanya mungkin meyesatkan, residual bukan berarti sampah yang tidak berguna. Residual kaya akan berbagai informasi dan karena itu merupakan bagian yang penting dalam setiap analisa data. Informasi dari data semula tidak terserap oleh model akan menjadi residual. Jika semua pola yang ada pada data telah masuk kedalam model maka residual akan berbentuk acak. Tetapi jika model yang digunakan tidak mampu mengambil semua pola yang ada pada data maka residual akan mempunyai kecenderungan tertentu. Dalam hal itu, model belumlah baik betul, dalam arti masih dapat disempurnakan. Ketidakcocokan model dengan data dilihat dengan mengamati residual. Residual secara kasar, member keterangan tentang data yang tidak mengikuti pola umum model yang digunakan, ditandai dengan residualnya yang relative besar. Residual yang relative besar dapat merupakan petunjuk bahwa modelnya belum cocok ataupun pengamatannya barangkali merupakan pencilan. Membuang

data pencilan mungkin tindakan yang keliru, data tersebut mungkin berasal dari bibit unggul.

2.1.2. Dampak Pencilan

Keberadaan data pencilan akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut :

- Residual yang besar dari model yang terbentuk atau $E[e] \neq 0$
- Varians pada data tersebut menjadi lebih besar
- Taksiran interval memiliki rentang yang lebar

2.1.3. Identifikasi Pencilan

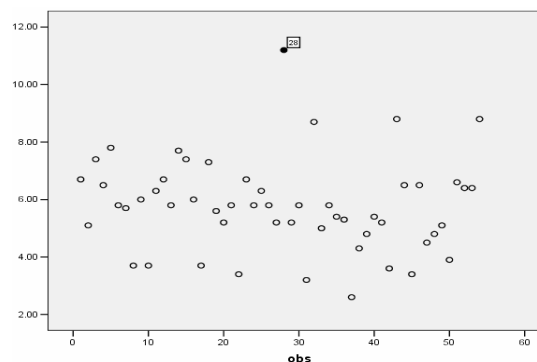
Suatu data mungkin letaknya terpencil, tapi bila pengaruhnya terhadap koefisien kecil maka kita tidak perlu member perhatian besar padanya. Makin besar ukuran sampel n makin kecil pengaruh suatu titik data. Pengaruh suatu data mungkin merupakan pencilan bila menggunakan suatu model, tapi tidak bila model lain yang digunakan.

Dalam statistik ruang, data pencilan harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data pencilan tersebut perlu dihilangkan atau tidak. Terdapat beberapa metode untuk menentukan batasan pencilan dalam sebuah analisis, yaitu :

a. Metode Grafis

Untuk melihat apakah terdapat pencilan pada data, dapat dilakukan dengan memplot antara data dengan observasi ke- i ($i = 1, 2, 3, \dots, n$) seperti gambar berikut :

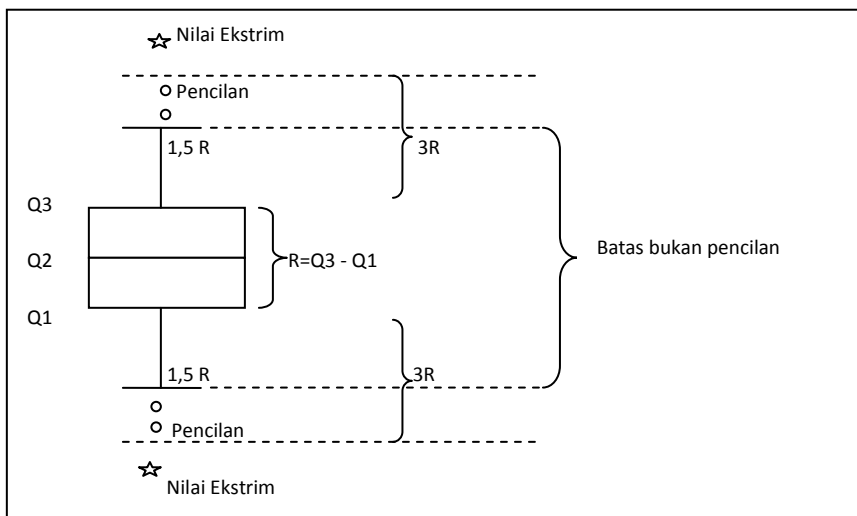
Gambar 1. Contoh *scatter-plot* dari data dengan observasi ke- i



Dari contoh di atas terdapat salah satu data, yakni observasi ke-28 yang mengindikasikan merupakan pencilan. Selain melalui *scatter-plot* di atas, jika sudah didapatkan model regresi maka dapat dilakukan dengan cara memplot antara residual (e) dengan nilai prediksi y (\hat{y}). Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya pencilan. Kelemahan

dari metode ini adalah keputusan bahwa suatu data merupakan pencilan sangat bergantung pada *judgement* peneliti, karena hanya mengandalkan visualisasi grafis, untuk itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut. Dalam rangka meminimumkan kesalahan teknis, maka pendeteksian dilakukan melalui perhitungan statistis yang akan dijelaskan pada bagian selanjutnya.

b. Boxplot



Metode ini merupakan yang paling umum yakni dengan mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q3 - Q1$. Data-data pencilan dapat ditentukan yaitu nilai yang kurang dari $1.5 \cdot IQR$ terhadap kuartil 1 dan nilai yang lebih dari $1.5 \cdot IQR$ terhadap kuartil 3. Gambar 2. Skema identifikasi pencilan menggunakan IQR atau *boxplot*

c. Leverage Values, DfFITS, Cook's Distance, dan DfBETA(s)

Sebelum menjelaskan ketentuan untuk metode di atas, terlebih dahulu didefinisikan arti dari masing-masing metode :

- *Leverage Values*; menampilkan nilai *leverage* (pengaruh) terpusat.
- *DfFITS* atau *Standardized DfFIT*; menampilkan nilai perubahan dalam harga yang diprediksi bilamana *case* tertentu dikeluarkan, yang sudah distandarkan.
- *Cook's Distance*; menampilkan nilai jarak Cook
- *DfBETA(s)*; menampilkan nilai perubahan koefisien regresi sebagai hasil perubahan yang disebabkan oleh pengeluaran *case* tertentu. Digunakan untuk mendeteksi pencilan pada variabel bebas.

Adapun ketentuan yang berlaku dalam pengambilan keputusan adanya pencilan atau tidak adalah sebagai berikut :

Gambar 3. Kriteria pengambilan keputusan adanya pencilan atau tidak

$$\text{Jika } \left\{ \begin{array}{l} \text{Leverage Values} > (2p - 1)/n \\ \text{DfFITS} > 2 \cdot \sqrt{p/n} \\ \text{Cook's Distance} > F(0.5; p, n-p) \\ \text{DfBETA}(s) > 2/\sqrt{n} \end{array} \right\} \begin{array}{l} \text{Pencilan} \\ \text{(Outlier)} \end{array}$$

Ket. : n = Jumlah observasi (sampel); p = Jumlah parameter

2.1.4. Tindakan Terhadap Pencilan

Tujuan analisis data pencilan bukan hanya mencari poal umum data tapi juga mencari data yang mana saja yang tidak mengikuti pola umum. Bila ternyata hasil identifikasi menunjukkan adanya pencilan, maka yang dapat dilakukan adalah membuang/menghilangkan data pengamatan tersebut, jika tidak memberikan pengaruh setelah dilakukan pengujian. Karena bagaimanapun juga keberadaan data pencilan mengganggu proses analisis.

Untuk menanggulangi pencilan pada data, yaitu dengan mengeluarkan atau membuang observasi ke-i pada data yang

diduga merupakan pencilan. Kemudian dilakukan pengujian kembali untuk mendeteksi terdapat atau tidaknya pencilan pada data sampai tidak terdapat lagi pencilan pada data tersebut.

Meskipun pencilan identik dengan data yang tidak bagus, akan tetapi ia merupakan bagian terpenting dari data, karena menyimpan informasi tertentu. Membuang data pencilan mungkin tindakan yang amat keliru, data tersebut mungkin berasal dari bibit unggul. Bayangkan bila kita menganalisa data produksi susu sapi ada sapi yang menghasilkan susu sapi yang jauh melebihi sapi-sapi lainnya. Data ini merupakan pencilan dan bila diabaikan atau membuang informasi seperti ini berarti membuang bibit unggul. Contoh lainnya ialah jika kita menangani data penyebaran emas dalam tambang, maka kita lebih tertarik melihat data yang tidak mengikuti pola umum. Emas tersebar dalam bongkah. Bongkah yang besar lebih jarang ditemui, jadi lebih merupakan pencilan. Dengan demikian praktik membuang pengamatan dari himpunan data regresi seharusnya tidak dilakukan tanpa alasan yang jelas, karena data itu mungkin pencilan.

2. STUDI KASUS DAN PEMBAHASAN

Dalam penelitian ini, data yang digunakan bersumber pada data jumlah wisatawan mancanegara dan pengunjung asing yang masuk melalui pintu Makasar tahun 2007 yang disajikan dalam tabel berikut ini :

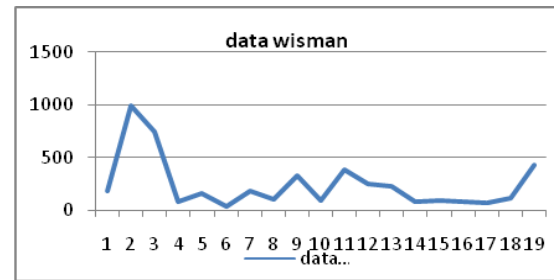
Tabel 1 : Jumlah wisatawan yang masuk

No	Bulan	Jumlah Wisman & Pengunjung asing
1	Januari '07	187
2	Februari	989
3	Maret	741
4	April	78
5	Mei	154
6	Juni	36
7	Juli	178
8	Agustus	99
9	September	323
10	Oktober	88
11	Nopember	383
12	Desember	246
13	Januari '08	228
14	Februari	79
15	Maret	94
16	April	78
17	Mei	67
18	Juni	112
19	Juli	430

Sumber : Ditjen Imigrasi, BPS dan Angka Pura I & II

Diolah kembali oleh Pusat Pengelolaan Data dan Sistem Jaringan
www.budpar.go.id

Berdasarkan data di atas dapat dibuat grafiknya sebagai berikut :

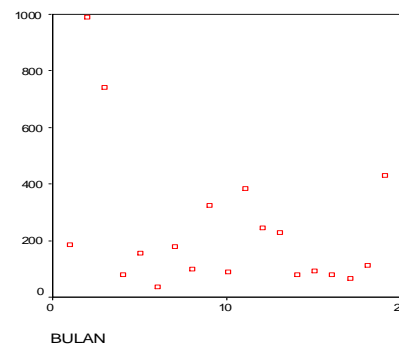


Berdasarkan data di atas terlihat bahwa data bulan Februari dan Maret 2007 jauh di atas data-data yang lain pada umumnya. Hal ini diperkuat dengan bentuk grafik garisnya yang mencolok pada data bulan Februari dan Maret 2007. Bila dibandingkan dengan data bulan Februari dan Maret 2008 terlihat data bulan Februari dan Maret 2007 memiliki perbedaan yang sangat signifikan, sehingga bukan merupakan data musiman.

2.1. Identifikasi data pencilan

Untuk mendeteksi apakah data bulan Februari dan Maret merupakan data pencilan akan dilakukan uji-uji sebagai berikut :

a. Metode Grafis



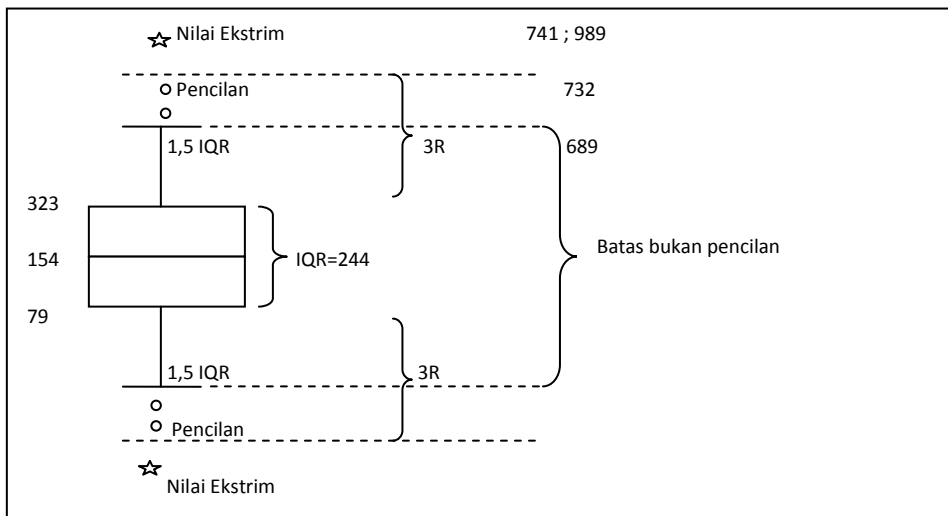
Berdasarkan identifikasi sebaran data (scatter) di atas terlihat data no.2 dan 3 (Februari dan Maret 2007) letaknya jauh bila dibandingkan dengan data-data lainnya sehingga dapat dicurigai sebagai data pencilan.

b. Boxplot

Statistics		
WISMAN		
N	Valid	19
	Missing	0
Percentiles	25	79,00
	50	154,00
	75	323,00

Descriptives				Statistic	Std. Error
WISMAN	1	Mean		241,58	57,233
		95% Confidence Interval for Mean	Lower Bound	121,34	
			Upper Bound	361,82	
		5% Trimmed Mean		211,48	
		Median		154,00	
		Variance		62236,702	
		Std. Deviation		249,473	
		Minimum		36	
		Maximum		989	
		Range		953	
		Interquartile Range		244,00	
		Skewness		2,048	,524
		Kurtosis		4,094	1,014

Berdasarkan data hitungan statistic di atas terlihat nilai quartile ($Q1=79$, $Q2=154$ dan $Q3=323$) sedangkan nilai Interquartile range ($IQR=244$). Jadi nilai $1,5 IQR=366$. Sehingga dapat dibuat ilustrasinya :

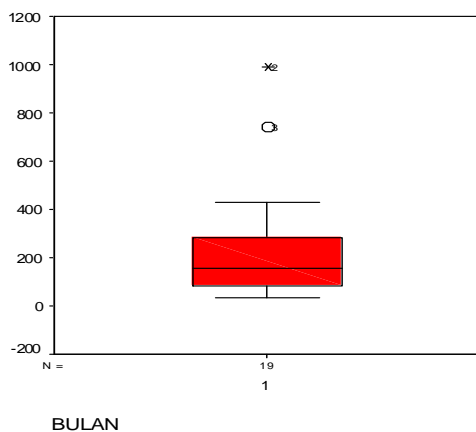


Jadi berdasarkan ilustrasi di atas terlihat bahwa batas atas pencilan adalah data yang nilainya antara 689 sampai 732, di atas nilai 732 merupakan nilai ekstrim. Jika dilihat data bulan Februari 2007 = 989 dan Maret 2007 = 741 maka dapat dikatakan bahwa data tersebut bukan hanya sebatas pencilan melainkan sudah termasuk data ekstrim.

Berdasarkan box plot di atas terlihat data no. 2 dan 3 (Februari dan Maret 2007) berada di atas batas.

Identifikasi berdasarkan penghitungan SPSS disajikan berikut ini :

Box plot



WISMAN Stem-and-Leaf Plot for BULAN= 1

```

Frequency  Stem & Leaf
      8,00   0 . 36777899
      4,00   1 . 1578
      2,00   2 . 24
      2,00   3 . 28
      1,00   4 . 3
      2,00 Extremes (>=741)
Stem width:   100
Each leaf:    1 case(s)

```

Berdasarkan hitungan dengan SPSS terlihat bahwa ada 2 data ekstrim, yaitu data yang lebih besar atau sama dengan 741.

2.2. Pengaruh keberadaan data pencilan

Jika data pencilan dihilangkan maka data yang hilang akan diganti dengan data hasil penghitungan *missing value* dengan metode *linear trend at point*.

No	Bulan	wisman	Tanpa pencilan	Trend wisman
1	Januari '07	187	187	187
2	Februari	989	-	140
3	Maret	741	-	143
4	April	78	78	78
5	Mei	154	154	154
6	Juni	36	36	36
7	Juli	178	178	178
8	Agustus	99	99	99
9	September	323	323	323
10	Oktober	88	88	88
11	Nopember	383	383	383
12	Desember	246	246	246
13	Januari '08	228	228	228
14	Februari	79	79	79
15	Maret	94	94	94
16	April	78	78	78
17	Mei	67	67	67
18	Juni	112	112	112
19	Juli	430	430	430

Sumber : data diolah 2008

Hasil perbandingan penghitungan statistic deskriptif :

Descriptives					
BULAN				Statistic	Std. Error
WISMAN	1	Mean		241,58	57,233
		95% Confidence Interval for Mean	Lower Bound	121,34	
			Upper Bound	361,82	
		5% Trimmed Mean		211,48	
		Median		154,00	
		Variance		62236,702	
		Std. Deviation		249,473	
		Minimum		36	
		Maximum		989	
		Range		953	
		Interquartile Range		244,00	
		Skewness		2,048	,524
		Kurtosis		4,094	1,014
TREND(WISMAN)	1	Mean		165,44	25,479
		95% Confidence Interval for Mean	Lower Bound	111,91	
			Upper Bound	218,97	
		5% Trimmed Mean		157,94	
		Median		140,12	
		Variance		12334,306	
		Std. Deviation		111,060	
		Minimum		36	
		Maximum		430	
		Range		394	
		Interquartile Range		149,00	
		Skewness		1,228	,524
		Kurtosis		,738	1,014

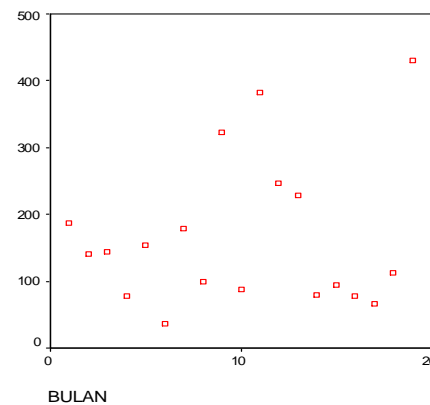
Berdasarkan tabel hitungan di atas terlihat bahwa ada perbedaan yang cukup signifikan antara besarnya nilai variansi data wisman (ada pencilan) = 62236,702 dan data trend wisman (pencilan dihilangkan) = 12334,306. Karena besarnya perbedaan nilai variansinya maka dapat mempengaruhi uji kenormalan data. Hal ini dapat terlihat dalam tabel uji normalitas data dengan metode Kolmogorov-Smirnov berikut ini :

Tests of Normality

	BULAN	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
	WISMAN 1	,230	19	,009	,738	19	,000
	TREND(WISMAN) 1	,173	19	,139	,864	19	,012

a. Lilliefors Significance Correction

Dari tabel di atas terlihat bahwa data wisman nilai signifikasinya 0,009 dan data trend(wisman) nilai signifikasinya 0,139. Karena data dikatakan normal jika nilai signifikasinya lebih besar dari 0,05 maka data wisman, yang memiliki data pencilan tidak memenuhi anggapan kenormalan, atau dapat dikatakan bahwa data wisman tidak normal. Namun demikian data yang memiliki data pencilan belum tentu pasti tidak normal, karena data trend wisman sebenarnya masih menyimpan data pencilan yang lain. Hal ini terlihat pada hasil penghitungan berikut ini :



TREND(WISMAN) Stem-and-Leaf Plot for BULAN= 1

Frequency Stem & Leaf

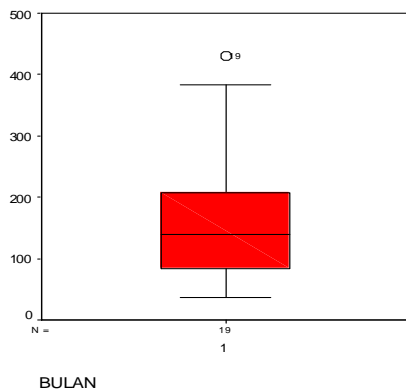
```

1,00 0 . 3
7,00 0 . 6777899
3,00 1 . 144
3,00 1 . 578
2,00 2 . 24
,00 2 .
1,00 3 . 2
1,00 3 . 8
1,00 Extremes (>=430)

```

Stem width: 100

Each leaf: 1 case(s)



Dari penghitungan dan gambar box plot di atas terlihat ada 1 data yang di atas batas pencilan, yaitu data no.19 atau Juli 2008.

KESIMPULAN DAN SARAN

Kesimpulan :

Berdasarkan analisa-analisa di atas dapat diambil kesimpulan sebagai berikut :

1. Pencilan (*outlier*) adalah suatu data yang jauh berbeda dibandingkan terhadap keseluruhan data, untuk mengidentifikasinya dapat menggunakan metode boxplot.
2. Keberadaan data pencilan mempengaruhi nilai variansi data sehingga dapat mempengaruhi uji kenormalan data.
3. Data kunjungan wisman ke Indonesia melalui pintu Makasar pada bulan Januari 2007 s.d. Juli 2008 setidaknya memiliki 2 (dua) data pencilan, yaitu data bulan Februari 2007 dan Maret 2007.

Saran :

1. Banyak informasi yang didapat dari data pencilan, untuk itu jangan langsung mengilangkan data pencilan dari data walaupun data tersebut sangat mengganggu.
2. Hendaknya ada penelitian lanjutan mengenai data pencilan/ ekstrim dalam data kunjungan wisman ke Indonesia melalui pintu Makasar antara bulan Januari 2007 s.d. Juli 2008.
3. Saran untuk penelitian selanjutnya adalah supaya peneliti lebih banyak membaca dan mendapatkan bahan tentang pencilan, supaya mendapatkan lebih banyak cara untuk mendeteksi dan menanggulangi pencilan pada data.

DAFTAR PUSTAKA

- Atkinson A.C (1981). Two Graphical Display for Outlying and Influential Observation in Regression. *Technometrics* Division, MathSoft, Seattle, WA. Rousseeuw, P.J. (1991), "Diagnostic Plot for Regression Outlier and Leverage Point," *Statistical Software Newsletter*, 127-129.
- Gujarati (1988). *Basic Econometrics* 2nd. Ed., Mc Graw-Hill Book. Co. New York
- Hawkins ,D.M.,1994.The Feasible Solution Algorithm for Least TrimmedSquares
- MathSoft (1999a), *S-Plus 2000 User's Guide*, Chapter 9. Data Analysis Products
- Outliers and Leverage Points," *Journal of the American Statistical Association*, 85, 633-651.
- Sembiring RK, 1995, *Analisa Regresi*, ITB, Bandung
- Soemartini, 2005. *Bahan Kuliah Pencilan (Outlier)*, Jurusan Statistika , UNPAD-Bandung .
- www.budpar.go.id
- www.wikipedia.org