

Uji Normalitas

Bilson Simamora, 29 Juni 2018

Prinsip normalitas berlaku untuk *error* (Rawlings et al., 1998; Osborne dan Water, 2002; Willlaim et al., 2013). *Error* yang disebut juga *residual* adalah perbedaan antara nilai hasil observasi dan nilai prediksi yang diperoleh melalui model regresi sebenarnya (*true regression model*) yang berlaku untuk populasi secara keseluruhan. *Residual* adalah perbedaan antara nilai observasi dengan nilai prediksi yang diperoleh dengan menggunakan model regresi estimasi. Untuk setiap kombinasi nilai prediktor, diasumsikan distribusi residual adalah normal.

Williams et al. (2013) menambahkan bahwa apabila *error* berdistribusi normal, kita dapat mengambil kesimpulan tentang populasi walaupun ukuran sampel kecil. Pelanggaran terhadap asumsi ini, menurut mereka, mengutip White dan MacDonald (1980), dapat menurunkan efisiensi estimator. Apabila *error* berdistribusi normal, maka *ordinary least square* (OLS) adalah teknik yang paling efisien dari semua estimator yang tidak bias. Menurut mereka, apabila *error* tidak berdistribusi normal, nilai koefisien t dan F mungkin tidak mengikuti distribusi t dan F. Mereka juga mengatakan bahwa prinsip normalitas variabel tidak bisa diterapkan apabila estimator menggunakan skala dikotomi (misalnya: pria-wanita, hujan-tidak hujan, terang-gelap).

Bagaimana kalau tidak normal?

1. Tambah jumlah sampel karena semakin besar ukuran sampel, data semakin mengarah ke distribusi normal.
2. Deteksi Keberadaan *outliers* yang dapat menyebabkan data berdistribusi tidak normal.
3. Lakukan transformasi regresi linier menjadi regresi *polynomial*. Ada dua persamaan yang umumnya digunakan, yaitu persamaan pangkat dua dan pangkat tiga tergantung pada pola distribusi datanya. Persamaan pangkat dua digunakan untuk distribusi yang bersifat parabolik, sedangkan persamaan pangkat tiga digunakan untuk distribusi data yang bersifat sinusoidal.
4. Lakukan transformasi regresi linier menjadi regresi logaritmik. Ada empat pilihan model, yaitu *log-liner*, *linier-log* dan *log-log*. Prinsip-prinsip operasi logaritma berlaku dalam pendekatan ini.

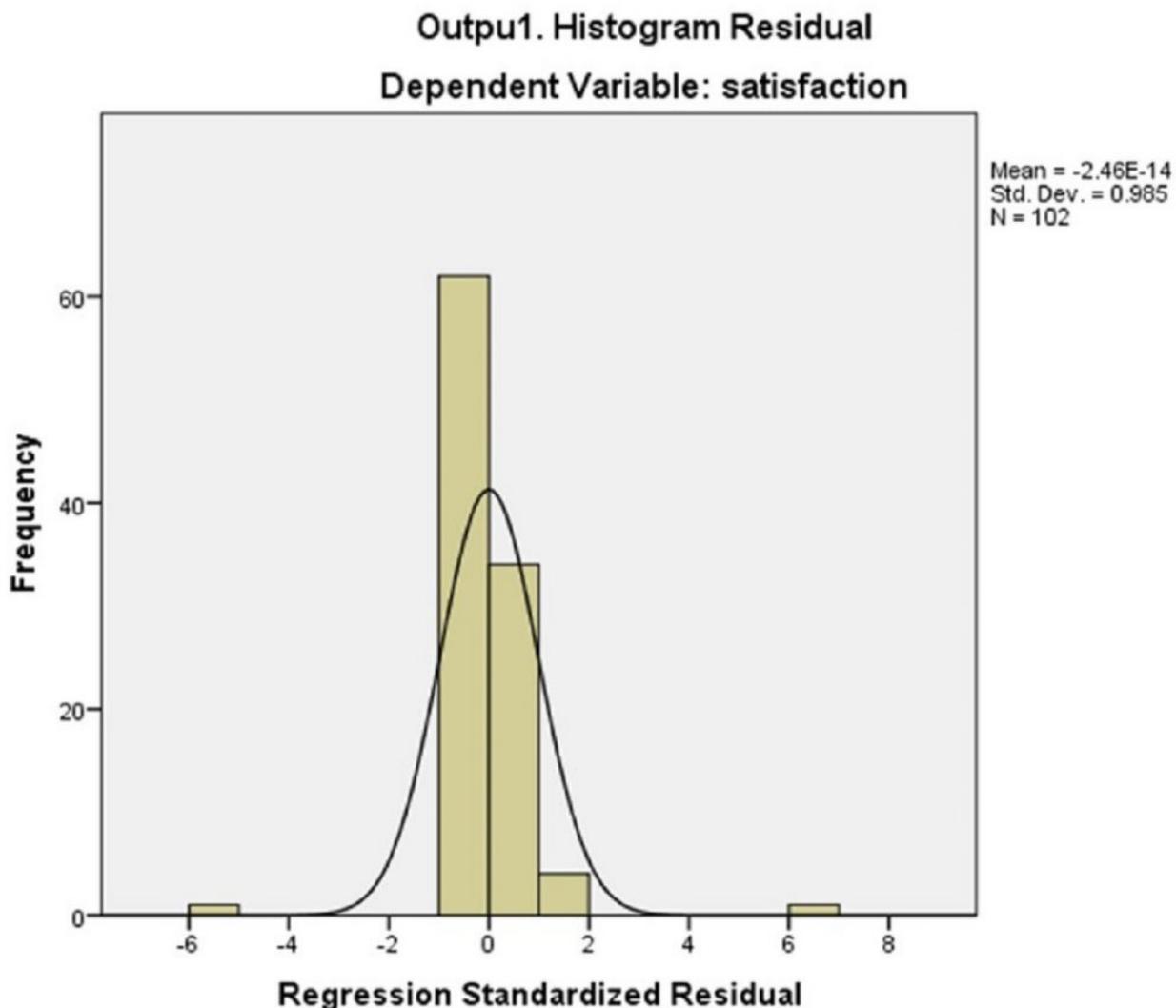
Transformasi regresi linier menjadi *polynomial regression* dan *logarithmic regression* dapat ditemukan pada buku-buku statistika. Pembaca bisa pula membaca buku "Teknik Regresi untuk Riset Manajemen dan Bisnis" karangan penulis.

Sebagai contoh, buka file 'contoh regresi.sav ([klik link ini](#)). Untuk membuka file ini tentu komputer saudara harus punya program SPSS.

Pertama-tama, lakukan regresi dengan langkah-langkah pada SPSS:

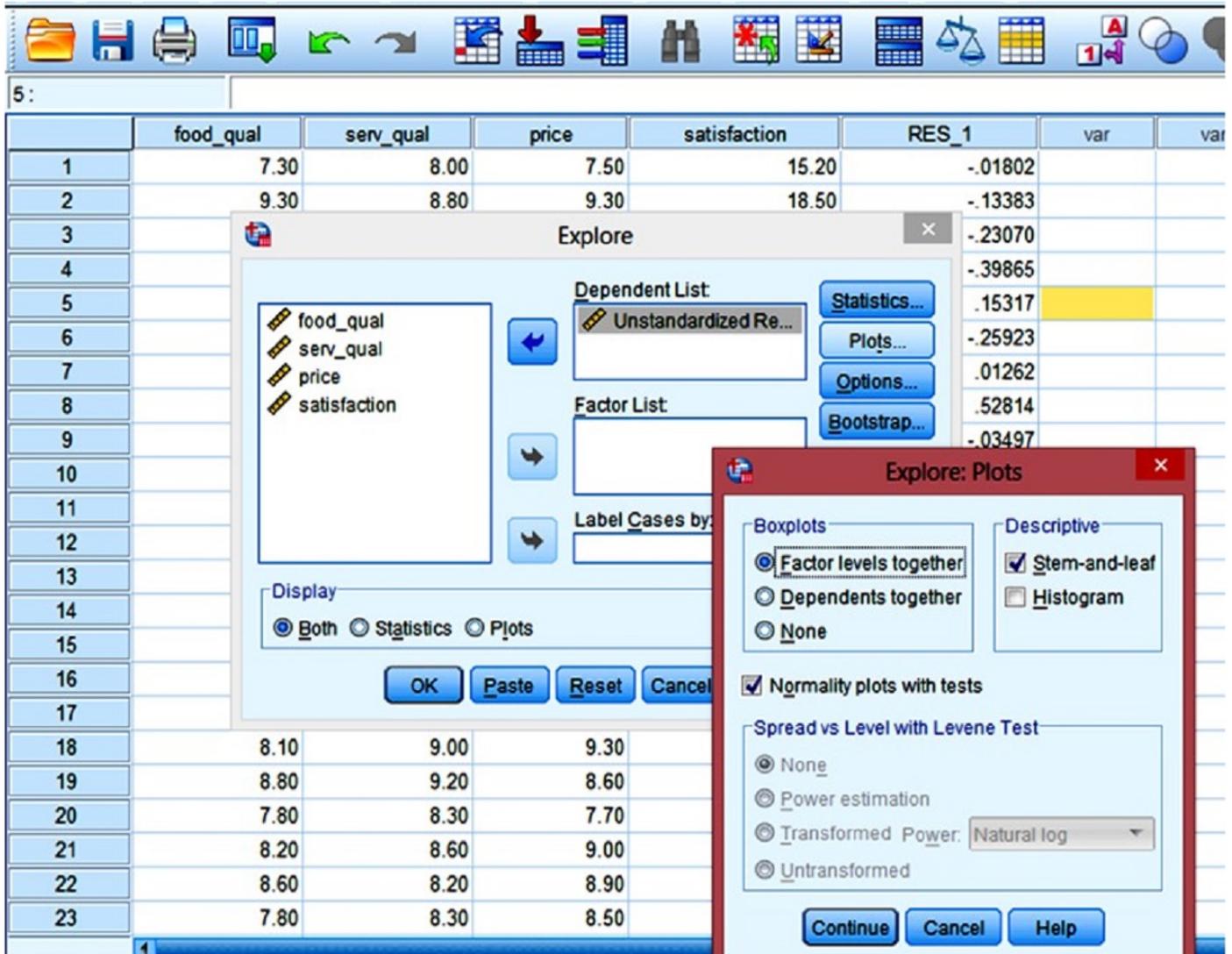
Analysis>Regression>Linier. Kemudian, pada kotak dialog SPSS, isikan variabel dependen dan independen. Jangan lupa meng-klik tombol 'save' agar SPSS memberikan data *unstandardized residuals*. Klik tombol 'Plot' lalu tandai *Normal probability plot, Histogram*. Pada sel X masukkan *ZPRES dan *ZRESID masukkan pada sel Y, kemudian *Continue*. Lalu klik OK pada kotak yang pertama muncul tadi.

Karena perhatian kita adalah menguji normalitas, maka pusat perhatian kita yang pertama distribusi residual, seperti ditunjukkan pada *Output 1*. Kita dapat melihat bahwa distribusi residual yang ditampilkan dalam bentuk histogram tidak mengikuti pola kurva normal. Namun, untuk memastikan apakah residual berdistribusi normal atau tidak, kita dapat menggunakan uji Kolmogorov-Smirnov dan Shapiro-Wilk.



Perlu diketahui dari proses sebelumnya SPSS telah memberikan data residual yang disimpan sebagai variabel RES_1. Nah, sekarang kita menguji apakah variabel ini berdistribusi normal atau tidak. Langkah-langkahnya adalah: *Analyze>Descriptive Statistics>Explore*. Pada kotak yang muncul, masukkan *Unstandardized Residuals* (RES_1) ke sel *Dependent List*. Kemudian, klik tombol 'Plot' dan tandai

Normality plots with test. Klik tombol Statistic, kemudian tandai pilihan 'Outliers'. Kemudian, klik OK agar SPSS melakukan proses.



Hipotesis Ho yang diuji adalah: 'Data residual berdistribusi normal'. Dengan nilai Kolmogorov-Smirnov (nilai KS=0.231, Sig.=0.000) dan uji Shapiro-Wilk (nilai SW=0.551, Sig.=0.000), kita tidak memiliki cukup bukti untuk menerima Ho. Dengan kata lain, terdapat cukup bukti untuk menolak Ho dan menyatakan bahwa data residual berdistribusi tidak normal. Dengan tidak diterimanya Ho berarti asumsi homoskedastisitas belum terpenuhi. Bagaimana selanjutnya?

Output 2. Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.231	102	.000	.551	102	.000

a. Lilliefors Significance Correction

Apa yang kita lakukan sekarang? Menambah jumlah responden dapat dilakukan apabila proses penelitian masih memungkinkannya. Kalau tidak, maka langkah kedua dapat kita lakukan, yaitu mendeteksi *outliers*. Pendeteksian outliers dapat memanfaatkan hasil analisis SPSS, seperti ditampilkan pada *Output 3*.

Pada *Output 3* terlihat bahwa *outliers* adalah kasus yang nilai *residual*-nya ≤ -3.27 (satu responden, yaitu no. 52) dan ≥ 0.83 (lima responden, yaitu no. 23, 48, 51, 75 dan 100). Selanjutnya, keenam responden tersebut kita keluarkan dari data. Konsekuensinya, ukuran sampel berubah dari 102 menjadi 96. *Tips*: Dalam mengeluarkan responden atau case dalam SPSS, mulailah dari nomor paling besar.

Kita bisa mendeteksi secara manual. Namun, pada saat menandai tombol "*Outliers*" pada proses di atas, sebenarnya kita sudah memerintah SPSS agar mengidentifikasi "*Outliers*".

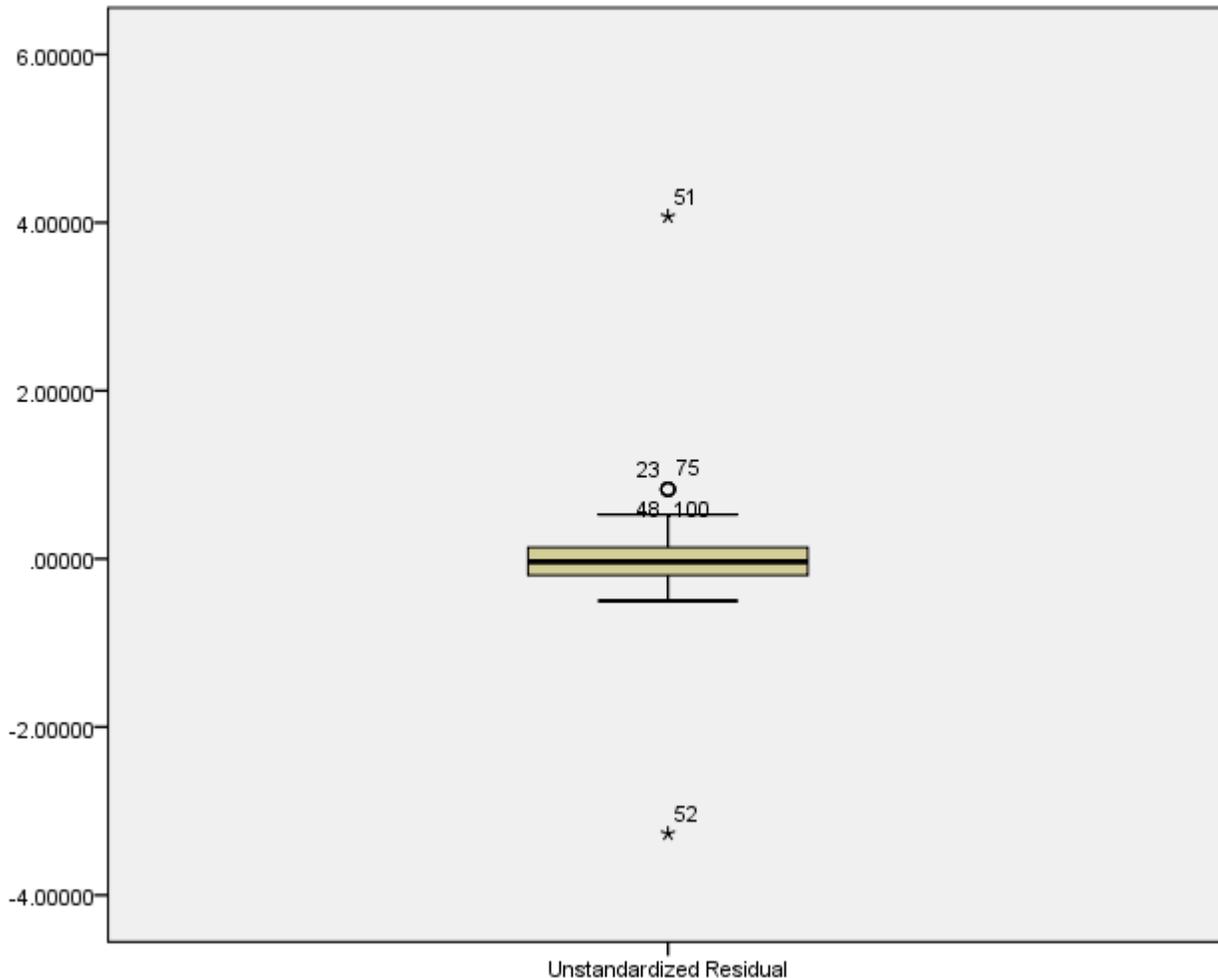
□ **Output 3.** Unstandardized Residual Stem-and-Leaf Plot

Frequency	Stem	&	Leaf
1.00	Extremes		(= -3.27)
1.00	-4	.	9
8.00	-3	.	00666999
14.00	-2	.	00033344445555
22.00	-1	.	1111111333335588889999
17.00	-0	.	11111111113335557
9.00	0	.	111133338
14.00	1	.	11113333455557
3.00	2	.	111
3.00	3	.	000
.00	4	.	
5.00	5	.	12222
5.00	Extremes		(> $=.83$)

Stem width:	.10000
Each leaf:	1 case(s)

Output 4. *Outliers*

Observed Value



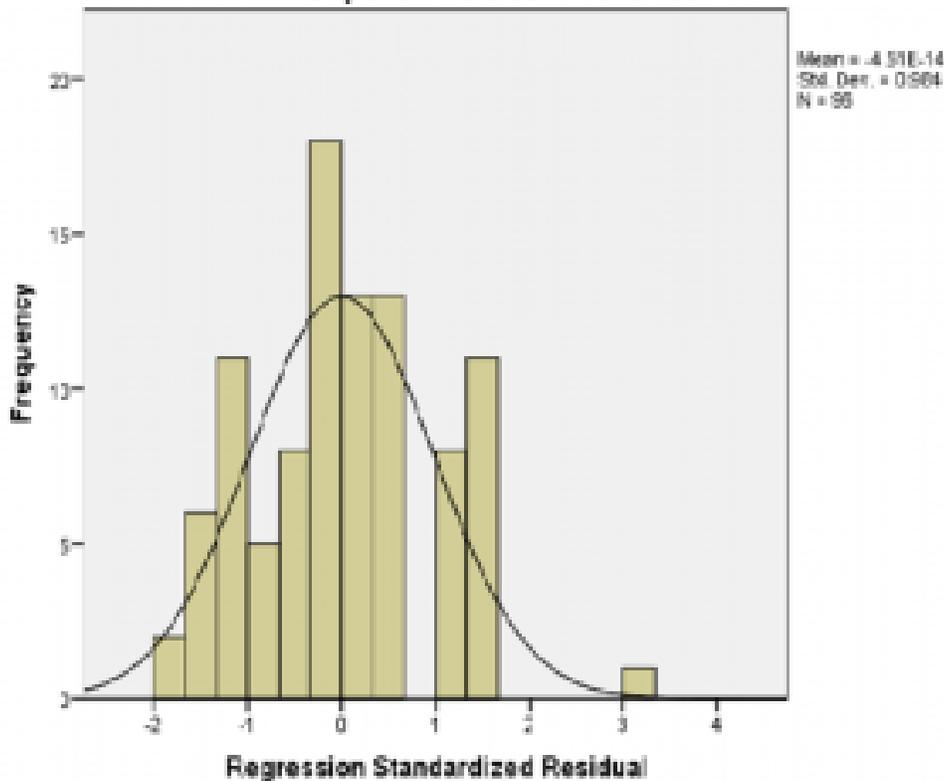
Setelah *outliers* dikeluarkan, lakukan kembali regresi dengan langkah-langkah pada SPSS: *Analysis > Regression > Linier*. Kemudian, pada kotak yang muncul isikan variabel dependen dan independen. Jangan lupa meng-klik tombol 'save' agar SPSS memberikan data *unstandardized residuals*. Klik tombol 'Plot' lalu tandai *Normal probability plot, Histogram*. Pada sel X masukkan *ZPRES dan pada sel Y masukkan *ZRESID, kemudian *Continue*. Klik tombol *Statistics* dan pastikan pilihan *Estimates, Model fit* dan *Collinearity diagnostics* telah ditandai. Kemudian klik OK.

Langsung saja kita memeriksa distribusi residual yang dinyatakan dalam bentuk histogram (*Output 4*). Terlihat bahwa distribusi residual tidak mengikuti sepenuhnya kurva distribusi normal. Namun, untuk memastikan apakah residual berdistribusi normal atau tidak, kita kembali menggunakan uji Kolmogorov-Smirnov dan Shapiro-Wilk. Prosedur sebelumnya kita gunakan lagi. Langkah-langkahnya adalah: *Analyze > Descriptive Statistics > Explore*. Pada kotak yang muncul, masukkan *Unstandardized Residuals (RES_1)* ke sel *Dependent List*. Kemudian, klik tombol 'Plot' dan tandai *Normality plots with test*. Kemudian, klik OK untuk melakukan proses.

Output 4

Histogram

Dependent Variable: satisfaction



Pada *Output 5* terlihat bahwa nilai Kolmogorov-Smirnov=0.72 dengan Sig.=0.20 sedangkan nilai Shapiro-Wilk=0.975 dengan nilai Sig.=0.060. Dengan demikian, pada batas $\alpha=0.05$, tidak cukup bukti untuk menolak 'Ho: Data residual berdistribusi normal'. Dengan demikian, data sudah memenuhi syarat normalitas.

Output 5. Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.072	96	.200*	.975	96	.060

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Kalau uji Kolmogorov-Smirnov (KS) dan Shapiro-Wilk (SW) memiliki nilai sig.>0.05 tidak ada masalah terkait normalitas residual. Pertanyaannya, bagaimana kalau salah satu 'nilai sig.'-nya <0.05? Menurut Razali dan Wah (2011), apabila jumlah sampel 30 atau lebih kecil, uji KS dan SW memiliki kekuatan yang relatif sama. Pada rentang jumlah sampel antara 30 sampai 2000,

uji SW lebih *powerful* dibanding uji KS. Sedangkan pada sampel besar (di atas 2000) kedua uji memiliki kekuatan sama. Dengan demikian, apabila jumlah sampel antara 30 sampai 2000, maka uji SW lebih meyakinkan.

Daftar Referensi (maaf belum lengkap)

Razali, N.M. & Wah, Y.B. (2011). Power comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling test. *Jornal of Statistical Modelings and Analytics*, Vol 2 (1), 21-33. Retrieved from https://www.researchgate.net/publication/267205556_Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirnov_Lilliefors_and_Anderson-Darling_Tests, 29 Juni 2018.