

UJI NORMALITAS DATA MENGGUNAKAN METODE EMPIRICAL DISTRIBUTION FUNCTION DENGAN MEMANFAATKAN MATLAB DAN MINITAB 19

Rifqi Arief Permana¹, Diana Ikasari²

¹Program Studi Sarjana Matematika Universitas Diponegoro
Tembalang, Jawa Tengah

²Sistem Informasi Universitas Gunadarma
Depok, Jawa Barat

rifqi.apermana@gmail.com, ikasari.diana@gmail.com

ABSTRAK

Beberapa variasi metode dapat digunakan dalam melakukan uji normalitas data untuk menentukan apakah data memiliki distribusi normal atau tidak. Metode tersebut dapat menghasilkan keputusan yang berbeda sehingga dapat membuat ambigu para praktisi dalam melakukan uji statistik. Diperlukan adanya suatu metode uji normalitas yang dapat menghasilkan keputusan yang konsisten dan akurat. Penelitian ini bertujuan untuk melakukan pengujian normalitas data dengan menggunakan beberapa metode seperti: Anderson-Darling test, Kolmogorov-Smirnov test, Ryan-Joiner test. Penelitian ini menggunakan data sekunder yang diambil dari Badan Pusat Statistik Provinsi Jawa Barat dan Banyumas berupa data jumlah SMA Negeri di Provinsi Jawa Barat pada tahun 2021 dan data SD Negeri di Kabupaten Banyumas pada tahun 2021. Dari hasil uji yang dilakukan menggunakan software MATLAB dan Minitab19, diperoleh bahwa data jumlah SMA Negeri di Jawa Barat pada tahun 2021 berdistribusi tidak normal karena hasil dari uji normalitas menunjukkan $p\text{-value} < \alpha$, dengan tingkat signifikansi $\alpha = 0,005$, sedangkan data jumlah SD Negeri di Kabupaten Banyumas pada tahun 2021 berdistribusi normal karena hasil uji normalitasnya menunjukkan $p\text{-value} > \alpha$.

Kata Kunci: Uji Normalitas, Matlab, Minitab 19, p-value

ABSTRACT

Some variations of methods can be used to test the normality of the data to determine whether the data has a normal distribution or not. These methods can produce different decisions so that it can make practitioners ambiguous in carrying out statistical tests. It is necessary to have a normality test method that can produce consistent and accurate decisions. The aim of this study is to test the normality of the data by using several methods such as: Anderson-Darling test, Kolmogorov-Smirnov test, Ryan-Joiner test. This study uses secondary data taken from the Central Statistics Agency of West Java Province in the form of data on the number of Public Senior High Schools in West Java in 2021 and data on Public Elementary School in Banyumas Regency in 2021. From the results of tests carried out using MATLAB and Minitab19 software, it was found that data on the number of Public Senior High Schools in West Java in 2021 was not normally distributed because the results of the normality test showed $p\text{-value} < \alpha$, with a significance level of $\alpha = 0,005$, while the data on the number of Public Elementary Schools in Banyumas Regency in 2021 has a normal distribution because the results of the normality test show $p\text{-value} > \alpha$.

Key Word: Normality test, Matlab, Minitab 19, p-value.

PENDAHULUAN

Statistika terdiri dari dua jenis yaitu statistik deskriptif dan statistik inferensial. Keduanya tidak dapat dipisahkan satu dengan yang lainnya. Statistika deskriptif dapat didefinisikan sebagai kumpulan kegiatan yang berhubungan pengumpulan data, pengolahan, dan penyajian data dalam bentuk yang baik (Arisena, 2018). Statistika inferensial didefinisikan sebagai alat bantu mengolah data, menganalisis data, menarik kesimpulan

dan membuat keputusan. Bentuk contoh dari statistik inferensial adalah statistik parametrik dan statistik non parametrik. Penggunaan uji statistik parametrik dan uji non parametrik berdasarkan pada distribusi data yang digunakan sebagai salah satu bentuk asumsi dasar. Jika data memiliki distribusi normal maka statistik parametrik dapat digunakan, namun apabila data memiliki jika distribusi tidak normal maka yang digunakan adalah statistik non parametrik. Data populasi akan

memiliki nilai distribusi normal jika rata-rata nilainya sama dengan modenyanya serta memiliki nilai sama dengan mediannya dan sebagian nilai atau skor berpusat atau terkumpul di posisi tengah (Malik, 2013).

Statistika berupaya untuk melakukan pemeliharaan agar data yang digunakan memiliki hasil yang berada pada nilai rata-rata atau memiliki hasil yang disebut dengan istilah kewajaran. Dalam melakukan uji kewajaran tersebut, perlu dilakukan suatu pengujian normalitas (Matondang, 2012). Pengujian normalitas data dapat dilakukan dengan berbagai metode seperti: Anderson-Darling test, Kolmogorov-Smirnov test, Pearson Chi-Square test, Cramer-von Mises test, Shapiro-Wilk test, Fisher's cumulate test (Wahjudi, 2007). Matondang (2012) mengemukakan bahwa ada dua entuk pengujian normalitas yang digunakan untuk pendekatan statistik parametrik diantaranya adalah uji Lilliefors dan uji ChiKuadrat. Öner (2017) juga memperkenalkan bentuk metode uji normalitas yang dapat digunakan yaitu dengan metode Kolmogorov-Smirnov, Kesimetrisan dan Kurtosis (Uji Skewness-Kurtosis). Metode-metode yang telah disebutkan di atas memiliki kelebihan masing-masing, penelitian yang dilakukan oleh Dahlan (2009) menyebutkan bahwa uji Kolmogorov-Smirnov lebih tepat untuk sampel yang lebih dari 50 sedangkan Matondang (2012) menyebutkan bahwa Lilliefors biasanya digunakan untuk rentang data yang tidak melebihi 50. Uji Shapiro-Wilk yang pada umumnya penggunaannya terbatas untuk sampel yang kurang dari 50 agar menghasilkan keputusan yang akurat. Uji Skewness-Kurtosis yang dapat mengambil keputusan suatu uji normalitas jika digunakan pada data dengan nilai rata-rata lebih kecil dari standar deviasi (Sintia, 2022).

Matlab (*Matrix Laboratory*) merupakan suatu lingkungan komputasi numerikal dan bahasa pemrograman generasi keempat. Matlab dapat digunakan untuk persoalan - persoalan matriks dan aljabar linear lainnya yang dapat diselesaikan dengan singkat dan tepat. Analisis numerik merupakan bagian dari ilmu statistika yang sering digunakan untuk membuat keputusan di berbagai bidang ilmu

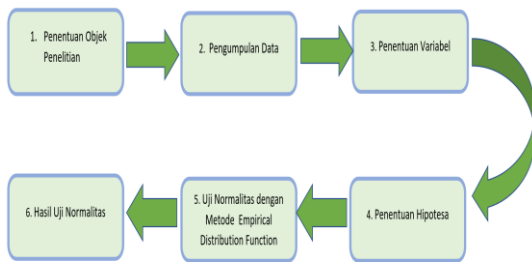
Teknik, seperti arsitektur, teknik sipil, dan bahkan teknik industri. Beberapa manfaat dan fungsi dari matlab diantaranya adalah melakukan pengolahan data menjadi lebih mudah. Data yang diolah ini adalah data riset, dimana MATLAB merupakan suatu program yang dapat digunakan dalam memlakukan validasi hasil riset dengan berbagai metode, hal lain yang membuat matlab banyak digunakan. Penggunaan matlab dalam penelitian juga dapat memvisualisasikan hasil data riset dengan jelas. Penggunaan matlab juga dapat membantu dalam melakukan simulasi yaitu dengan membuat suatu bentuk pemodelan atau algoritma untuk menyelesaikan masalah serta melakukan pengujian keberhasilan model atau algoritma tersebut dengan mensimulasikan hasil akhir (Rahmalia, 2021).

Berdasarkan uraian diatas maka penulis akan melakukan uji normalitas data dengan menggunakan beberapa metode seperti: Anderson-Darling test, Kolmogorov-Smirnov test, Pearson Chi-Square test, Cramer-von Mises test, Shapiro-Wilk test, Fisher's cumulate test. Hal tersebut dilakukan untuk mengetahui dan membandingkan hasil uji data sampling yang memiliki nilai normalitas atau tidak memiliki nilai normalitas. Penelitian ini menggunakan data sekunder yang diambil dari Badan Pusat Statistik Provinsi Jawa Barat berupa data jumlah SMA Negeri di Provinsi Jawa Barat pada tahun 2021 dan data data SD Negeri di Kabupaten Banyumas pada tahun 2021 yang diambil dari Badan Pusat Statistik Kabupaten Banyumas sebagai data uji normalitas.

METODE PENELITIAN

Penelitian ini merupakan penelitian non reaktif (*non-reactive research*) dengan menggunakan data yang sudah terkumpul sebelumnya (data sekunder). Adapun tahapan penelitian yang dilakukan adalah menentukan objek penelitian, yang dilanjutkan dengan pengumpulan data kemudian menentukan variabel yang akan digunakan dalam melakukan uji normalitas. Berdasarkan variabel yang telah ditetapkan maka dilanjutkan dengan penentuan hipotesa untuk selanjutnya dilakukan uji normalitas dengan metode Empirical Distribution Function yang

pada tahap akhir akan menghasilkan nilai penentuan hasil uji normalitas yang telah dilakukan. Adapun tahapan penelitian terlihat pada gambar 1.



Gambar 1. Tahapan Penelitian

1. Penentuan Objek Penelitian

Objek penelitian adalah uji normalitas data sampling terhadap populasi SMA Negeri yang berada di wilayah propinsi Jawa Barat serta populasi seluruh SD Negeri di wilayah kabupaten Banyumas

2. Pengumpulan Data

Pengumpulan data yang dilakukan pada penelitian ini menggunakan data sekunder yang diambil dari Badan Pusat Statistik propinsi Jawa Barat berupa data jumlah SMA Negeri yang tersebar di 28 (dua puluhdelapan) wilayah Jawa Barat dan Badan Pusat Statistik Kabupaten Banyumas berupa data jumlah Sekolah Dasar yang tersebar di 27 (Dua puluh tujuh) wilayah Banyumas. Tabel data yang digunakan terlihat pada tabel 1 dan tabel 2.

Tabel 1. Data Sampling Jumlah SMA Negeri Jawa Barat Tahun 2021

No	Wilayah	Jumlah Sekolah
1.	Jawa Barat	512
2.	Bogor	45
3.	Sukabumi	27
4.	Cianjur	18
5.	Bandung	19
6.	Garut	32
7.	Tasikmalaya	18
8.	Ciamis	17
9.	Kuningan	19
10.	Cirebon	21
11.	Majalengka	16
12.	Sumedang	15
13.	Indramayu	23
14.	Subang	21
15.	Purwakarta	17
16.	Karawang	30
17.	Bekasi	44
18.	Bandung Barat	17
19.	Pangandaran	5
20.	Kota Bogor	10
21.	Kota Sukabumi	5
22.	Kota Bandung	27
23.	Kota Cirebon	9
24.	Kota Bekasi	22
25.	Kota Depok	16
26.	Kota Cimahi	6
27.	Kota Tasikmalaya	10
28.	Kota Banjar	3

Tabel 2. Data Sampling jumlah Sekolah Dasar di wilayah Banyumas tahun 2021 (BPS Banyumas)

No	Wilayah	Jumlah Sekolah
1.	Lumbir	35
2.	Wangon	45
3.	Jatilawang	36
4.	Rawalo	22
5.	Kebasen	29
6.	Kemranjen	33
7.	Sumpiuh	30
8.	Tambak	27
9.	Somagede	19
10.	Kalibagor	21
11.	Banyumas	31
12.	Patikraja	28
13.	Purwojati	20
14.	Ajibarang	33
15.	Gumelar	32
16.	Pekuncen	34
17.	Cilongok	43
18.	Karanglewas	22
19.	Kedungbanteng	26
20.	Baturraden	24
21.	Sumbang	37
22.	Kembaran	37
23.	Sokaraja	30
24.	Purwokerto Selatan	23
25.	Purwokerto Barat	18
26.	Purwokerto Timur	20
27.	Purwokerto Utara	21

3. Penentuan Variabel

Hal penting yang harus diperhatikan selama melakukan uji normalitas adalah melihat hipotesis uji normalitas dari hasil keputusan yang diperoleh. Apabila variabel data berasal

dari distribusi normal, maka kita bisa menggunakan pengujian statistika parametrik. Namun, apabila variabel data berasal dari distribusi tidak normal, maka kita bisa melakukan penanggulangan pada variabel data yang tidak normal tersebut. Apabila tidak bisa dilakukan penanggulangan, maka kita hanya bisa menggunakan pengujian statistika nonparametrik agar hasil yang diperoleh sesuai yang diharapkan.

Misalkan y_1, y_2, \dots, y_n adalah nilai terurut dari nilai sampel x_1, x_2, \dots, x_n .

Misalkan i menyatakan frekuensi dari y_k dalam order ke- k .

Fungsi distribusi empiris didefinisikan sebagai:

$$F_n(y) = \begin{cases} 0, & y < y_1 \\ i/n, & y_k \leq y < y_{k+1}, k = 1, 2, \dots, n-1 \\ 1, & y \geq y_n \end{cases} \quad (1)$$

Kolmogorov-Smirnov test

Statistik uji Kolmogorov-Smirnov dihitung menggunakan statistik D_n yang didefinisikan sebagai:

$$D_n = \sup_x |F_n(x) - F_0(x)| \quad (2)$$

Statistik uji D_n tidak lain adalah nilai terbesar dari jarak vertikal antara $F_n(x)$ dan $F_0(x)$, yaitu:

$$D_n = \max\{|D_n^-, |D_n^+|\} \quad (3)$$

dengan $D_n^- = F(x_{k-1}) - F_0(x_k)$ dan $D_n^+ = F(x_k) - F_0(x_k)$

Statistik KSz dihitung sebagai berikut:

$$KSz = \sqrt{n}D_n \quad (4)$$

4. Penentuan Hipotesa

Hipotesis untuk pengujian kecocokan (*fitness*) jika data sampel atau variabel mengikuti distribusi normal (yaitu pengujian kecocokan) adalah sebagai berikut:

$$H_0: x \in N(\mu, \sigma^2) \quad (5)$$

$$H_1: x \notin N(\mu, \sigma^2) \quad (6)$$

H_0 diterima jika p -value $> \alpha$, dengan α adalah tingkat signifikansi. Nilai α yang sering digunakan adalah $\alpha = 0.05 = 5\%$.

H_0 ditolak (H_1 diterima) jika p -value $< \alpha$.

5. Uji Normalitas Menggunakan Metode Empirical Distribution Function

Uji coba dalam menggunakan Metode Empirical Distribution dapat dilakukan dengan berbagai test, dalam penelitian ini dilakukan uji normalitas dengan menghitung p -value untuk uji Kolmogorov-Smirnov dengan algoritma terlihat pada gambar 2.

```

Step 1: Enter sample x vector
Step 2: y_i = sort(x)
Step 3: i = 1, 2, ..., n
Step 4: F_n(y) = i / n
Step 5: mean(y) = Σ y_i / n
Step 6: σ(y) = √[Σ (y_i - mean(y))^2 / (n-1)]
Step 7: z_i = (y_i - mean(y)) / σ(y)
Step 8: F_0(y) = u_i = φ(z_i)
Step 9: D_plus = abs(F_0(y) - i / n)
Step 10: D_minus = abs(F_0(y) - (i - 1) / n)
Step 11: D_n = max(D_plus, D_minus)
Step 12: KSz = √n D_n
    
```

Gambar 2. Algoritma Uji Normalitas

6. Hasil Uji Normalitas

Uji normalitas yang dilakukan pada penelitian ini menghasilkan data pada tabel yang digunakan dengan nilai berdistribusi normal atau berdistribusi tidak normal melalui indicator nilai p -value.

HASIL DAN PEMBAHASAN

Berdasarkan uji normalitas yang dilakukan terhadap data sampling yang digunakan pada tabel 1 menggunakan software matlab, menggunakan sepuluh uji normalitas, yaitu; Kolmogorov-Smirnov Limiting Form, Kolmogorov-Smirnov Stephens Method, Kolmogorov-Smirnov Marsaglia Method, Kolmogorov-Smirnov Lilliefors Test, Anderson-Darling Test, Cramer-Von Mises Test, Shapiro-Wilk Test, Shapiro-Francia Test, Jarque-Bera Test, dan D'Agostino and Pearson Test. Hasil uji dapat dilihat pada Gambar 3.

Test Name	Test Statistic	p-value	Normality(1:Normal,0:Not Normal)
KS Limiting Form	2.2672	0.0001	0
KS Stephens Modification	2.3318	0.0100	0
KS Marsaglia Method	2.2672	0.0000	0
KS Lilliefors Modification	0.4285	0.0000	0
Anderson-Darling Test	8.2523	0.0000	0
Cramer-Von Mises Test	1.6496	0.0000	0
Shapiro-Wilk Test	0.2758	0.0000	0
Shapiro-Francia Test	0.2492	0.0000	0
Jarque-Bera Test	698.3531	0.0000	0
D'Agostino & Pearson Test	67.1491	0.0000	0

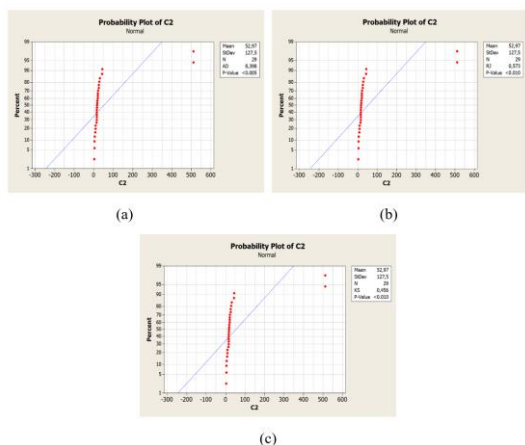
```

ans =
    2.2672    0.0001    0
    2.3318    0.0100    0
    2.2672    0.0000    0
    0.4285    0.0000    0
    8.2523    0.0000    0
    1.6496    0.0000    0
    0.2758    0.0000    0
    0.2492    0.0000    0
    698.3531    0    0
    67.1491    0.0000    0
    
```

Gambar 3. Hasil Uji Normalitas Data Pada Tabel 1 menggunakan software MATLAB

Berdasarkan hasil uji normalitas pada Gambar 3, dapat dilihat bahwa semua uji normalitas menghasilkan $p\text{-value} < \alpha$, dengan tignkat signifikansi $\alpha = 0,005$. Maka dapat disimpulkan bahwa data pada Tabel 1 berdistribusi tidak normal.

Selanjutnya dilakukan uji normalitas menggunakan software Minitab19. Digunakan tiga uji normalitas yaitu; Anderson-Darling, Ryan-Joiner, dan Kolmogorov-Smirnov. Hasil uji dapat dilihat pada Gambar 4.



Gambar 4. Hasil Uji Normalitas Data Pada Tabel 1 menggunakan software Minitab19 menggunakan uji (a) Anderson-Darling, (b) Ryan-Joiner, (c) Kolmogorov-Smirnov

Gambar 4 menampilkan diagram pencar data dan $p\text{-value}$ yang dihasilkan oleh masing-masing uji normalitas. Uji normalitas Anderson-Darling menghasilkan $p\text{-value} < 0,005$, uji normalitas Ryan-Joiner menghasilkan $p\text{-value} < 0,010$, dan uji normalitas Kolmogorov-Smirnov

menghasilkan $p\text{-value} < 0,010$. Dari hasil tersebut, dapat disimpulkan ketiga uji normalitas tersebut menunjukkan bahwa data pada Tabel 1 berdistribusi tidak normal.

Pengujian data sampling pada tabel 2 dilakukan uji normalitas data pada Tabel 2 menggunakan software MATLAB. Hasil uji normalitas dapat dilihat pada Gambar 5.

Test Name	Test Statistic	p-value	Normality(1:Normal,0:Not Normal)
KS Limiting Form	0.5760	0.8944	1
KS Stephens Modification	0.5930	0.1500	1
KS Marsaglia Method	0.5760	0.8585	1
KS Lilliefors Modification	0.1108	0.2000	1
Anderson-Darling Test	0.3742	0.4161	1
Cramer-Von Mises Test	0.0449	0.5920	1
Shapiro-Wilk Test	0.9529	0.2515	1
Shapiro-Francia Test	0.9595	0.3064	1
Jarque-Bera Test	1.2415	0.5375	1
D'Agostino & Pearson Test	1.3932	0.4983	1

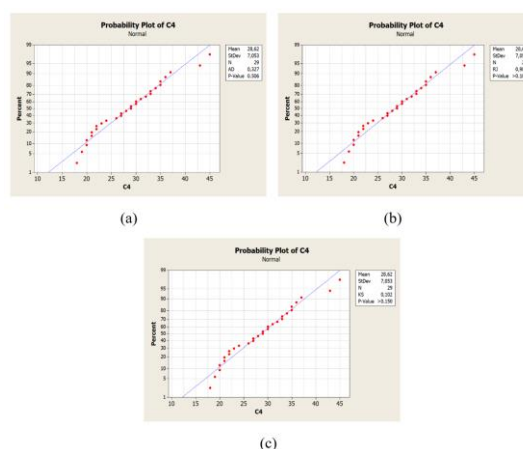
```

ans =
    0.5760    0.8944    1.0000
    0.5930    0.1500    1.0000
    0.5760    0.8585    1.0000
    0.1108    0.2000    1.0000
    0.3742    0.4161    1.0000
    0.0449    0.5920    1.0000
    0.9529    0.2515    1.0000
    0.9595    0.3064    1.0000
    1.2415    0.5375    1.0000
    1.3932    0.4983    1.0000
    
```

Gambar 5. Hasil Uji Normalitas Data Pada Tabel 2 menggunakan software MATLAB

Berdasarkan hasil uji normalitas pada Gambar 5, dapat dilihat bahwa semua uji normalitas menghasilkan $p\text{-value} > \alpha$, dengan tignkat signifikansi $\alpha = 0,005$. Maka dapat disimpulkan bahwa data pada Tabel 2 berdistribusi normal.

Selanjutnya dilakukan uji normalitas menggunakan software Minitab19. Digunakan tiga uji normalitas yaitu; Anderson-Darling, Ryan-Joiner, dan Kolmogorov-Smirnov. Hasil uji dapat dilihat pada Gambar 6.



Gambar 6. Hasil Uji Normalitas Data Pada Tabel 2 menggunakan software Minitab19 menggunakan uji

(a) Anderson-Darling, (b) Ryan-Joiner, (c)
Kolmogorov-Smirnov

Gambar 6 menampilkan diagram pencar data dan p -value yang dihasilkan oleh masing-masing uji normalitas. Uji normalitas Anderson-Darling menghasilkan p -value = 0,506, uji normalitas Ryan-Joiner menghasilkan p -value > 0,100, dan uji normalitas Kolmogorov-Smirnov menghasilkan p -value > 0,150. Dari hasil tersebut, dapat disimpulkan ketiga uji normalitas tersebut menunjukkan bahwa data pada Tabel 2 berdistribusi normal.

SIMPULAN DAN SARAN

Berdasarkan uji normalitas menggunakan *software* MATLAB dan Minitab19 yang telah dilakukan untuk mengetahui dan membandingkan hasil uji data sampling yang memiliki nilai normalitas atau tidak memiliki nilai normalitas, diperoleh bahwa data sampling jumlah SMA Negeri di Jawa Barat pada tahun 2021 berdistribusi tidak normal karena hasil dari uji normalitas menunjukkan p -value < α , dengan tingkat signifikansi $\alpha = 0,005$, sedangkan data sampling untuk jumlah SD Negeri di wilayah Banyumas pada tahun 2021 berdistribusi normal karena hasil uji normalitasnya menunjukkan p -value > α . Penelitian ini dapat dilanjutkan dengan menggunakan tools dan metode berbeda untuk mengetahui keakuratan hasil sebagai bentuk perbandingan.

UCAPAN TERIMAKASIH

Terima kasih Universitas Gunadarma dan Universitas Diponegoro atas kesempatan yang diberikan, serta seluruh pihak yang telah membantu terselesaikannya penulisan ini.

DAFTAR PUSTAKA

- Arisena, G. M. K. (2018). Buku Ajar Pengantar Statistika. 1–46.
- Badan Pusat Statistik Provinsi Jawa Barat. (n.d.). Retrieved November 9, 2022, from <https://jabar.bps.go.id/indicator/28/159/1/jumlah-sekolah-menengah-atas.html>
- BPS Kabupaten Banyumas. (n.d.). Retrieved November 9, 2022, from <https://banyumaskab.bps.go.id/statistictable/2022/04/08/433/jumlah->

sekolah-sekolah-dasar-sd-di-bawah-kementerian-pendidikan-dan-kebudayaan-menurut-kecamatan-di-kabupaten-banyumas-2020-2021-dan-2021-2022.html

- Dahlan, M.S. (2009). Statistik untuk Kedokteran dan Kesehatan, Edisi 4 (Deskriptif, Bivariat dan Multivariat, dilengkapi Aplikasi dengan Menggunakan SPSS). Jakarta: Salemba Medika.
- Malik, A. (2013). Pengantar Statistika Pendidikan. A Psicanalise Dos Contos de Fadas. Tradução Arlene Caetano, 466.
- Matondang, Z. (2011). Pengujian Normalitas Data. Taburlaasa PPS UNIMED, 1–8.
- Öner, M., & Kocakoç, I. D. (2017). JMASM 49: A compilation of some popular goodness of fit tests for normal distribution: Their algorithms and MATLAB codes (MATLAB). *Journal of Modern Applied Statistical Methods*, 16(2), 547–575. <https://doi.org/10.22237/jmasm/1509496200>
- Rahmalia, N. (2021). MATLAB: Definisi, Kegunaan, dan Sistemnya - Glints Blog. <https://glints.com/id/lowongan/matlab-adalah/#.Y2udCnZBy3B>
- Sintia, I., Pasarella, M. D., & Nohe, D. A. (2022). Perbandingan Tingkat Konsistensi Uji Distribusi Normalitas pada Kasus Tingkat Pengangguran di Jawa. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya Terbitan II*, 322–333.
- Wahjudi, D. (2007). Power Dari Uji Kenormalan Data.