KAJIAN BEBERAPA UJI KENORMALAN

Siti Karomah[1], Sigit Nugroho[2] dan Fachri Faisal[2]
[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu
[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRACT

This research is aimed to learn about Normal Distribution and the normality tests. The statistical tests that was used in the normality tests are Lilliefors, Shapiro-Wilk, Skewness, Kurtosis, D'Agostino-Pearson, and Jarque Bera. Literature study is used in this experiment using data simulation generated from Normal, Chi Square, Exsponential, and Uniform distribution. The result shows that not all of normality tests is good and suitable for all distribution data. In general, Shapiro-Wilk test is good and suitable for testing normality for both Normal and non-Normal distribution, while D'Agostino-Pearson test is only recommended for non-Normal generated data.

Keywords: *Normal Distribution, Statistical tests, Normality tests*

1. PENDAHULUAN

Penggunaan metode statistika dalam membantu pemecahan masalah di berbagai bidang seperti ekonomi, bisnis, pertanian, teknik, psikologi, kedokteran, pendidikan, dan bidang-bidang spesifik lainnya terasa semakin dibutuhkan. Dalam hal ini metode statistik merupakan alat analisis dari data kuantitatif atau data kualitataif yang bertujuan untuk pengambilan keputusan dan peramalan (Djarwanto, 2001).

Sebelum kesimpulan tersebut dibuat, data yang telah terkumpul itu terlebih dahulu dipelajari, dianalisis atau diolah. Pembuatan suatu kesimpulan harus dilakukan dengan baik, cermat, teliti, hati-hati, serta mengikuti cara-cara dan teori yang benar. Dengan demikian kesimpulan yang dibuat dari penelitian dan pengamatan tersebut dapat dipertanggungjawabkan.

Misalnya, terdapat suatu sampel berukuran n dengan nilai pengamatan $y_1, y_2, ..., y_n$, kemudian akan timbul pertanyaan apakah populasi asal sampel itu dapat dianggap menyebar normal atau tidak (Nasoetion & Barizi,1985). Oleh karena itu, untuk mengetahui apakah suatu sampel berdistribusi normal atau tidak, maka perlu dilakukan uji kenormalan.

Uji kenormalan adalah suatu uji yang dilakukan untuk mengetahui apakah suatu data memiliki Distribusi Normal. Data yang memilki Distribusi Normal merupakan salah satu syarat dilakukannya uji parametrik. Sedangkan data yang tidak mempunyai Distribusi Normal dapat dianalisis dengan menggunakan uji nonparametrik (Patria, 2008).

Dalam uji statistik seperti uji χ^2, T, dan F suatu data diasumsikan berdistribusi normal. Secara kasat mata, kenormalan suatu data dapat ditaksir dengan menggunakan metode grafik. Tetapi metode grafik tidak dapat digunakan jika perbedaan antara Distribusi Normal dan distribusi sampel adalah signifikan. Uji-uji yang dapat digunakan untuk menaksir kenormalan adalah Kai-Kuadrat, Anderson Darling, Ryan Joner, Kolmogrov-Smirnov, Shapiro-Wilk, Jarque Bera, Lilliefors, D'Agostino-Pearson, Skewness, dan Kurtosis.

Berdasarkan uraian di atas penulis tertarik untuk mempelajari dan mengkaji prosedur uji yang cocok dan layak digunakan untuk menguji kenormalan suatu sampel data. Untuk lebih memahami konsep-konsep di atas akan digunakan data simulasi yang diterapkan untuk menguji kenormalan dengan uji Lilliefors, Shapiro-Wilk, Skewness, Kurtosis, D'Agostino-Pearson, dan Jarque Bera.

2. LANDASAN TEORI

2.1 Distribusi Normal

Distribusi Normal merupakan distribusi teoritis dari variabel acak yang kontinu. Sebagian besar variabel acak kontinu yang diaplikasikan di berbagai bidang, pada umumnya memiliki distribusi yang dapat didekati dengan Distribusi Normal. Karena distribusinya kontinu, maka cara menghitung peluangnya dilakukan dengan menentukan luas dibawah kurvanya (Dajan, 1996).

Distribusi Normal dikenalkan oleh Abraham De Moivre pada tahun 1733. Pada saat itu distribusi ini digunakan untuk menghampiri peluang yang berhubungan dengan variabel acak binomial ketika parameter n besar. Hasil ini kemudian dikembangkan oleh Laplace dan yang lainnya, dan tercakup dalam teori peluang yang dikenal dengan teorema limit pusat (Ross, 1987).

2.2 Karakteristik Distribusi Normal

Distribusi Normal berbentuk sebuah lonceng *(bell-shape)*, yang juga sering disebut *bell-shape distribution*. Sebagai model teoritis, Distribusi Normal memiliki empat karateristik yang bersifat kumulatif, yaitu:

- Unimodal
  Sifat unimodal (uni = satu dan Modal = modos) mengandung pengertian bahwa setiap Distribusi Normal hanya memiliki satu modus.
- Simetris
  Suatu distribusi disebut simetris jika setengah bagian dari suatu distribusi sama dan sebangun (identik) dengan setengah bagian lainnya
- Identik
  Dari dua karateristik di atas (*unimodal dan simetris*), ketiga ukuran gejala pusat (modus, median, dan rata-rata) distribusi sama besar (identik). Dengan kata lain, pada Distribusi Normal, modus = median = rata-rata.
- Asimtotik
  Oleh karena nilai terkecil dan nilai terbesar pada suatu distribusi data kontinu bersifat tak hingga, maka tidak ada suatu daerah pun di bawah kurva normal yang memiliki frekuensi (probabilitas) sama dengan nol. Berdasarkan asumsi seperti itu, maka kurva Distribusi Normal tidak akan pernah menyentuh absisnya.

Model distribusi lain mungkin ada yang memiliki salah satu atau beberapa dari keempat karateristik itu, akan tetapi hanya Distribusi Normal yang memiliki semuanya.

2.3 Persamaan Kurva Normal

Fungsi kepekatan peluang (*probability density function*) variabel acak normal dengan rataan μ dan variansi σ^2 dapat dinyatakan sebagai (Dixon & Massey, 1991):

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

untuk $-\infty < x < \infty$; $-\infty < \mu < \infty$; $\sigma > 0$

Dari turunan pertama dan kedua fungsi kepekatan peluang variabel acak normal dapat diperoleh lima sifat kurva normal, yaitu :

1. Modus terdapat pada $x = \mu$
2. Kurva setangkup terhadap sumbu tegak yang melalui rataan μ
3. Kurva memiliki titik belok pada $x = \mu \pm \sigma$, cekung dari bawah bila $\mu - \sigma < x < \mu + \sigma$, dan cekung dari atas untuk nilai x lainnya.
4. Kedua ujung kurva normal mendekati asimtot sumbu datar bila nilai x bergerak menjahui μ baik ke kiri maupun ke kanan
5. Seluruh luas di bawah kurva dan di atas sumbu datar sama dengan 1

Distribusi Normal ini mempunyai rataan μ dan varian σ^2, dengan fungsi pembangkit momen yang dirumuskan sebagai berikut (Ross, 1987):

$$M_x(t)=e^{\left(\mu t+\frac{\sigma^2 t^2}{2}\right)} \tag{2}$$

2.4 Distribusi Normal Standar

Untuk mengatasi kesulitan dalam menghitung integral fungsi kepekatan peluang normal, maka dibuat tabel luas kurva normal sehingga memudahkan penggunaanya. Akan tetapi tidak akan mungkin membuat tabel yang berlainan untuk setiap nilai μ dan σ. Karena seluruh pengamatan setiap variabel acak normal X dapat ditransformasi menjadi himpunan pengamatan baru satu variabel acak normal Z dengan rataan 0 dan variansi 1, maka permasalahan tersebut dapat diatasi. Hal ini dapat dilakukan dengan trasformasi:

$$Z=\frac{X-\mu}{\sigma} \tag{3}$$

Jika X mendapat suatu nilai x, maka nilai Z padanannya diberikan oleh persamaan (3). Jadi, untuk nilai X antara $x = x_1$ dan $x = x_2$ maka variabel acak Z akan mempunyai nilai antara $z_1 = \frac{x_1 - \mu}{\sigma}$ dan $z_2 = \frac{x_2 - \mu}{\sigma}$, sehingga probabilitasnya dapat dinyatakan dengan:

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) \tag{4}$$

dan fungsi kepekatan peluang normal standar dirumuskan sebagai berikut:

$$\phi=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \tag{5}$$

Dengan adanya transformasi ini, kebutuhan akan banyaknya tabel luas kurva normal yang dibutuhkan telah diperkecil menjadi satu, yaitu tabel Distribusi Normal Standar (Walpole & Myers, 1995).

2.5 Distribusi Normal Kumulatif

Secara umum, fungsi distribusi kumulatif dari Distribusi Normal yang kontinu dengan rataan μ dan variansi σ^2 dapat dirumuskan sebagai :

$$F(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \qquad (6)$$

Secara matematis, fungsi distribusi pada persamaan (6) merupakan fungsi X yang tidak pernah menurun, akan mendekati nol jika $X \to -\infty$ dan mendekati satu jika $X \to \infty$. Sedangkan fungsi distribusi kumulatif normal standar dirumuskan sebagai:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{z} e^{-\frac{1}{2}u^2} du \qquad (7)$$

2.6 Distribusi yang Diturunkan dari Distribusi Normal

2.6.1 Distribusi Kai Kuadrat (χ^2)

Jika $Z_1, Z_2, ..., Z_n$ adalah variabel acak Distribusi Normal Standar yang saling bebas, maka :

$$X = Z_1^2 + Z_2^2 + ... + Z_n^2 \qquad (8)$$

dikatakan mempunyai Distribusi Kai Kuadrat dengan n derajat bebas, yang dinotasikan dengan $X \sim \chi_n^2$. Fungsi kepekatan peluang Distribusi Kai Kuadrat dirumuskan oleh:

$$f(x) = \frac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{1}{2}}\Gamma\left(\frac{n}{2}\right)} \qquad (9)$$

Sedangkan fungsi pembangkit momen dari variabel acak Kai Kuadrat dengan n derajat bebas ditunjukkan dengan persamaan berikut :

$$E\left[e^{tX}\right] = (1-2t)^{\frac{-n}{2}} \qquad (10)$$

Selain sebagai fungsi pembangkit momen dari variabel acak Kai-Kuadrat, persamaan (10) juga dikenal sebagai fungsi pembangkit momen dari variabel acak gamma dengan parameter $\left(\frac{n}{2}, \frac{1}{2}\right)$. Oleh karena keunikan fungsi pembangkit momen tersebut, maka kedua distribusi ini adalah identik.

2.6.1 Distribusi t

Jika Z dan V adalah variabel acak yang saling bebas dengan Z berdistribusi Normal Standar dan V berdistribusi Kai Kuadrat dengan derajat bebas n, maka :

$$T_n = \frac{Z}{\sqrt{\frac{V}{n}}} \tag{11}$$

dikatakan memiliki Distribusi-t dengan derajat bebas n. Fungsi kepekatan peluang dari T_n adalah turunan dari fungsi distribusi kumulatif $F(x)$ yang dirumuskan oleh:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} \tag{12}$$

dengan Γ adalah fungsi gamma.
Distribusi-t diterbitkan pada tahun 1908 dalam suatu makalah oleh W.S. Gosset. Dalam menurunkan distribusi ini, Gosset menganggap sampel berasal dari populasi normal. Karena Distribusi-t setangkup dengan rataan nol, maka $t_{1-\alpha} = -t_\alpha$.

2.6.3 Distribusi F

Distribusi F didefiniskan sebagai nisbah dua peubah acak Kai Kuadrat yang saling bebas dan masing-masing dibagi dengan derajat kebebasannya. Misalkan, χ_n^2 dan χ_m^2 adalah variabel acak Kai-Kuadrat yang saling bebas dengan n dan m berturut-turut adalah derajat kebebasannya. Maka variabel acak $F_{n,m}$ didefinisikan oleh $F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$. $F_{n,m}$ dikatakan memiliki distribusi F dengan n dan m derajat bebas.

Fungsi kepekatan peluang dari $F_{n,m}$ adalah turunan dari fungsi distribusi kumulatif $F(x)$ yang diberikan oleh:

$$f(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)(n)^{\frac{n}{2}}(m)^{\frac{m}{2}}\,x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)(m+nx)^{(n+m)/2}} \tag{13}$$

Derajat kebebasan yang berkaitan dengan variabel acak Kai-Kuadrat pada pembilang F selalu ditulis lebih dahulu, kemudian diikuti oleh derajat kebebasan yang berkaitan dengan variabel acak Kai-Kuadrat pada penyebut. Jadi, kurva Distribusi F tidak hanya tergantung pada kedua parameter n dan m tetapi juga tergantung pada urutan keduanya ditulis.

3. UJI-UJI KENORMALAN

Banyak uji statistik yang memerlukan data berdistribusi Normal. Untuk itu, pemeriksaan terhadap kenormalan data adalah kriteria dalam proses analisis data. Berikut ini akan dibahas tentang beberpa uji yang dapat digunakan untuk memeriksa kenormalan suatu data.

3.1 Uji Lilliefors (L)

Uji Lilliefors didesain untuk menaksir Distribusi Normal ketika satu dan/atau kedua dari parameter populasi (μ dan σ^2) tidak diketahui. Kedua parameter yang digunakan untuk uji ini diestimasi dari sampel data. Prosedur uji Lilliefors ini sama dengan prosedur untuk uji Kolmogorov-Smirnov, kecuali pada parameter yang digunakan. Pada uji Lilliefors parameter yang digunakan adalah nilai dari rata-rata sampel $\bar{X}$ dan variansi s^2.

Suatu data $x_1,\ldots,x_n$ yang diambil dari sampel acak berukuran n dari sebuah fungsi distribusi yang tidak diketahui dan dinotasikan dengan $F(x)$. Apabila

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{14}$$

sebagai penduga bagi μ, dan

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{15}$$

sebagai penduga bagi σ, serta:

$$Z_i = \frac{X_i - \bar{X}}{s} \quad untuk\ i = 1, 2, \ldots, n \tag{16}$$

Uji statistik Lilliefors dihitung dari nilai Z_i sebagai pengganti dari sampel acak aslinya. Uji ini dirumuskan oleh (Conover, 1971):

$$L = \sup_x \left| F^*(x) - S(x) \right| \tag{17}$$

Uji Lilliefors ini digunakan untuk menguji hipotesis:
H_0: Sampel acak mempunyai Distribusi Normal dengan rataan dan varian tertentu
H_1: Fungsi distribusi dari nilai-nilai X_i tidak Normal

H_0 akan ditolak pada taraf nyata α jika nilai dari uji statistik Lilliefors (L) lebih besar dari nilai $w_{1-\alpha}$ dan H_0 tidak dapat ditolak jika nilai dari L lebih kecil dari nilai $w_{1-\alpha}$.

3.2 Uji Shapiro-Wilk (SW)

Uji kenormalan ini dikembangkan oleh Samuel Shapiro dan Martin Wilk pada tahun 1965. Pada saat ini, uji Shapiro-Wilk menjadi uji kenormalan yang lebih disukai kerena memiliki kekuatan uji yang lebih baik dibandingkan uji-uji alternatif dari bermacam-macam range. Uji ini tergantung pada korelasi antara data yang diberikan dan kecocokan angka normalnya.

Tidak seperti beberapa uji-uji yang lain, uji ini tidak perlu menentukan nilai dari rataan dan varian terlebih dahulu, serta valid untuk menguji kenormalan dengan ukuran sampel antara 3 sampai 50. Nilai dari uji statistik Shapiro-Wilk ini adalah positif, yaitu lebih kecil atau sama dengan satu (antara 0 dan 1) (Peng *et al*, 2008).

Uji statistik Shapiro-Wilk (SW) dirumuskan sebagai berikut (Marques de Sa, 2007):

$$SW = \frac{\left[\sum_{i=1}^{k} a_i \left(x_{(n-i+1)} - x_{(i)}\right)\right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{18}$$

dimana $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ merupakan pengamatan tertata yang bebas dari skala dan tittik pusat. Jarak dari kesimetrian nilai data sekitar nilai tengah diukur dengan:

$$\left(x_{(n-i+1)} - x_{(i)}\right) \quad ; \quad untuk\ i = 1, 2, ..., k$$

dimana $k = \frac{(n+1)}{2}$ untuk n ganjil dan $k = \frac{(n)}{2}$ untuk n yang lainnya.

Dengan menggunakan uji statistik Shapiro-Wilk ini akan diuji hipotesis:
H_0: Data ditarik dari populasi yang berdistribusi Normal
H_1: Data ditarik dari populasi yang berdistribusi tidak Normal

Hipotesis nol ini akan ditolak pada taraf nyata α jika nilai dari uji satatistik SW kurang dari nilai yang terdapat pada Tabel quantil statistik uji Shapiro-Wilk. H_0 tidak dapat ditolak jika nilai dari SW lebih besar dari nilai Tabel persentil uji Shapiro-Wilk.

3.3 Uji Skewness dan Kurtosis

Selain rata-rata dan varian, ada dua pengukuran lain yang dapat memberikan informasi diskriptif tentang suatu distribusi. Dua pengukuran ini adalah Skewness dan Kurtosis yang masing-masing menunjukkan momen ketiga dan momen keempat dari suatu distribusi. Hays & Winkler (1971) *dalam* Shenkin (2003) mengatakan bahwa peristilahan momen digunakan untuk menunjukkan nilai harapan dari perbedaan kekuatan variabel acak. Persamaan umum untuk suatu momen dirumuskan:

$$m_i = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^i}{n} \tag{19}$$

Skewness

Skewness (kemiringan) adalah pengukuran yang menggambarkan derajat untuk distribusi yang simetris. Pengukuran yang paling tepat untuk skewness adalah dengan menggunakan nilai dari momen ketiga di sekitar rata-rata (m_3) (Shenkin, 2003):

$$m_3 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n} \tag{20}$$

Kemudian persamaan (20) digunakan untuk menghitung koefisien uji statistik Skewness yang dilambangkan dengan g_1 dan dirumuskan:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{3}{2}}} \tag{21}$$

D'Agostino *et al*. (1990) *dalam* Sheskin (2003) mengubah nilai dari g_1 kedalam statistik $\sqrt{b_1}$. Ini adalah estimasi dari parameter populasi yang menunjuk pada $\sqrt{\beta_1}$ yang juga digunakan untuk menggambarkan Skewness dan dirumuskan oleh:

$$\sqrt{b_1} = \frac{(n-2)g_1}{\sqrt{n(n-1)}} \tag{22}$$

Jika distribusinya Normal maka nilai $Skewness \sim N\left(0, \frac{6}{n}\right)$.

Interpretasi dari nilai $\sqrt{b_1}$:

- Jika $\sqrt{b_1}$ = 0, maka distribusi adalah simetris
- Jika $\sqrt{b_1}$ > 0, maka distribusi akan miring ke kanan
- Jika $\sqrt{b_1}$ < 0, maka distribusi akan miring ke kiri

Dengan uji statistik Skewness ini akan diuji hipotesis bahwa:

H_0 : Distribusi populasi yang menggambarkan sampel adalah simetris ($\sqrt{b_1} = 0$)

H_1 : Distribusi populasi yang menggambarkan sampel tidak simetris ($\sqrt{b_1} \neq 0$)

H_0 akan ditolak pada taraf nyata α jika nilai dari uji satatistik Skewness yang telah distandarlisasi lebih besar dari nilai Distribusi t dan H_0 tidak dapat ditolak jika nilai dari Skewness lebih kecil dari nilai Distribusi t.

Kurtosis

Menurut D'Agostino et al (1990) *dalam* Shenkin (2003), kata kurtosis berarti lengkungan. Secara umum kurtosis didefinisikan sebagai pengukuran yang menggambarkan derajat kelancipan suatu distribusi. Pengukuran kurtosis digunakan untuk menentukan apakah data berasal dari distribusi populasi Normal. Kurtosis menggambarkan momen keempat disekitar rata-rata (m_4):

$$m_4 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n} \tag{23}$$

Kemudian persamaan (23) digunakan untuk menghitung koefisien statistik uji Kurtosis yang dilambangkan dengan g_2 dan dirumuskan:

$$g_2 = \frac{m_4}{m_2^2} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^2} \tag{24}$$

Pengukuran Kurtosis dengan menggunakan momen keempat ini selalu menghasilkan bilangan positif. Oleh karena itu, pada tahun 1988 Moors mengembangkan suatu bilangan dari pengukuran alternatif untuk Kurtosis, yaitu dengan menghitung k_4 yang dirumuskan oleh:

$$k_4 = \left[\frac{\left(\left[\sum_{i=1}^{n}(X_i - \bar{X})^4 n(n+1)\right] \Big/ (n-1)\right) - 3\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}{(n-2)(n-3)}\right] \tag{25}$$

Walaupun pengukuran k_4 ini tidak menggambarkan keadaan yang sesungguhnya dari momen keempat sekitar rata-rata. Tetapi, dengan penggunaan k_4 dapat diasumsikan juga suatu bilangan negatif. Metode ini merupakan metode yang baik untuk mengestimasi Kurtosis. Dengan pengukuran alternatif pada persamaan ini, koefisien uji statistik Kurtosis (g_2) dapat juga dihitung dengan:

$$g_2 = \frac{k_4}{s^2} \tag{26}$$

Interpretasi dari nilai g_2:

- Jika $g_2 = 0$, maka distribusi adalah mesokurtik
- Jika $g_2 > 0$, maka distribusi akan menjadi leptokurtik
- Jika $g_2 < 0$, maka distribusi akan menjadi platykurtik

Zar (1999) *dalam* Shenkin (2003) mengatakan bahwa parameter populasi β_2 digunakan untuk menggambarkan Kurtosis. Statistik yang digunakan untuk mengestimasi nilai β_2 adalah b_2 yang dirumuskan oleh:

$$b_2 = \frac{(n-2)(n-3)g_2}{(n+1)(n-1)} + \frac{3(n-1)}{(n+1)} \tag{27}$$

Jika distribusinya Normal maka nilai $Kurtosis \sim N\left(3, \frac{24}{n}\right)$.

Interpretasi dari nilai b_2:

- Jika $b_2 = \frac{3(n-1)}{(n+1)}$, maka distribusinya adalah mesokurtik
- Jika $b_2 > \frac{3(n-1)}{(n+1)}$, maka distribusinya akan menjadi leptokurtik
- Jika $b_2 < \frac{3(n-1)}{(n+1)}$, maka distribusinya akan menjadi platykurtik

Jika ukuran sampel meningkat maka nilai dari b_2 akan mendekati 3.

Uji statistik Kurtosis ini akan digunakan untuk menguji hipotesis:
H_0 : Distribusi populasi yang menggambarkan sampel adalah mesokurtik ($\beta_2 = 3$)
H_1 : Distribusi populasi yang menggambarkan sampel bukan mesokurtik ($\beta_2 \neq 3$)

Jika nilai dari uji satatistik Kurtosis yang telah distandarlisasi lebih besar dari nilai Distribusi t pada taraf nyata α , maka H_0 akan ditolak dan sebaliknya.

3.4 Uji D'Agostino-Pearson (DAP)

Uji D'Agostino-Pearson dikembangkan oleh D'Agostino dan Pearson pada tahun 1973. Uji D'Agostino-Pearson menganalisis data untuk menentukan Skewness (untuk mengukur kesimetrian dari distribusi) dan Kurtosis (untuk mengukur bentuk dari distribusi). Kemudian menghitung seberapa jauh masing-masing dari nilai ini berbeda dari nilai harapan dengan Distribusi Normal dan menghitung nilai p dari penjumlahan kuadrat dari ketidakcocokan ini. Uji ini adalah kombinasi antara uji D'Agostino Skewness dan Ascombe-Glynn Kurtosis (Oztuna *et al*, 2006). Statistik uji D'Agostino-Pearson dirumuskan sebagai berikut:

$$DAP = Z_{g_1}^2 + Z_{g_2}^2 \tag{28}$$

dimana:

$$Z_{g_1} = E \ln\left(F + \sqrt{F^2 + 1}\right) \tag{29}$$

$$F = \frac{A}{\sqrt{\frac{2}{C-1}}} \tag{30}$$

$$E = \frac{1}{\sqrt{\ln D}} \tag{31}$$

$$D = \sqrt{C} \tag{32}$$

$$C = \sqrt{2(B-1)} - 1 \tag{33}$$

$$B = \frac{3\left(n^2 + 27n - 70\right)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} \tag{34}$$

$$A = g_1\sqrt{\frac{(n+1)(n+3)}{6(n-2)}} \tag{35}$$

Dan

$$Z_{g_2} = \frac{1 - \frac{2}{9K} - \sqrt[3]{L}}{\sqrt{\frac{2}{9K}}} \tag{36}$$

$$L = \frac{1 - \frac{2}{K}}{1 + H\sqrt{\frac{2}{K-4}}} \tag{37}$$

$$K = 6 + \frac{8}{J}\left[\frac{2}{J} + \sqrt{1 + \frac{4}{J^2}}\right] \tag{38}$$

$$J = \frac{6\left(n^2 - 5n + 2\right)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}} \tag{39}$$

$$H = \frac{(n-2)(n-3)\left|g_2\right|}{(n+1)(n-1)\sqrt{G}} \tag{40}$$

$$G = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \tag{41}$$

Dengan menggunakan uji statistik D'Agostino-Pearson ini akan diuji hipotesis:

H_0: Sampel berasal dari populasi yang berdistribusi Normal
H_1: Sampel tidak berasal dari populasi yang berdistribusi Normal

Tolak H_0 pada taraf nyata α jika nilai dari uji satatistik D'Agostino-Pearson (DAP) lebih besar dari nilai χ^2 dengan derajat bebas 2 dan H_0 tidak dapat ditolak jika nilai dari DAP lebih kecil dari nilai χ^2.

3.5 Uji Jarque Bera (JB)

Uji ini diusulkan oleh Jarque dan Bera pada tahun 1980 dan hanya digunakan untuk menguji kenormalan. Uji Jarque Bera diketahui mempunyai kekuatan uji yang sangat baik dan mudah dalam perhitungannya serta biasanya digunakan dalam bidang ekonometrik.
Uji ini menghitung koefisien dari Skewness dan Kurtosi dari variabel acak. Jika distribusinya Normal maka Skewness akan bernilai 0 dan Kurtosis bernilai 3. Tetapi, jika sampelnya tidak berdistribusi Normal maka nilai Skewness akan jauh dari nol dan nilai Kurtosis akan jauh dari 3.
uji statistik Jarque Bera yang dirumuskan sebagai berikut (Gujarati, 1999):

$$JB = n\left(\frac{(g_1)^2}{6} + \frac{(g_2 - 3)^2}{24}\right) \tag{42}$$

Uji statistik JB dihampiri oleh distribusi Kai Kuadrat dengan derajat bebas 2.

Dengan uji Jarque Bera ini akan diuji hipotesis:

H_0: Sampel berasal dari populasi yang berdistribusi Normal
H_1: Sampel tidak berasal dari populasi yang berdistribusi Normal

Tolak H_0 pada taraf nyata α jika nilai dari uji satatistik Jarque Bera (JB) lebih besar dari nilai χ^2 dengan derajat bebas 2 dan H_0 tidak dapat ditolak jika nilai dari JB lebih kecil dari nilai χ^2.

4. TELADAN PENERAPAN

Sebagai ilustrasi, data yang digunakan untuk menguji kenormalan adalah data simulasi. Simulasi ini terdiri dari data berdistribusi Normal dan Distribusi tidak Normal (distribusi Kai kuadrat, Eksponensial, dan Seragam). Data simulasi ini dibuat dan diolah menggunakan program komputer Microsoft Excel.

4.1 Hasil dan Pembahasan

Suatu uji statistik dikatakan baik jika mempunyai kemungkinan yang kecil untuk menolak H_0 apabila H_0 itu benar dan mempunyai kemungkinan yang besar untuk menolak H_0, kalau H_0 itu salah. Dalam simulasi ini diharapkan bahwa untuk sampel data yang berdistribusi Normal, hipotesis nol yang menyatakan sampel data berasal dari populasi yang berdistribusi Normal diterima, sedangkan untuk sampel yang berasal dari distribusi tidak Normal (distribusi Kai kuadrat, Eksponensial, dan Seragam) diharapkan hipotesis nol ditolak. Tabel di bawah ini merupakan persentase penolakan dari simulasi yang dilakukan untuk setiap uji kenormalan.

Tabel 1. Persentase Penolakan untuk Tiap Tipe Distribusi

Uji Statistik	**Persentase Penolakan**			
	Normal	**Kai Kuadrat**	**Eksponensial**	**Seragam**
Lilliefors	2.39	7.15	81.82	15.36
Shapiro-Wilk	4.48	38.76	100	86.10
Skewness	3.23	36.68	97.20	0
Kurtosis	2.29	13.06	58.74	7.17
D'Agostino-Pearson	79.79	91.29	100	99.89
Jarque Bera	3.12	28.70	93.29	0

Untuk sampel data yang berdistribusi Normal diharapkan bahwa persentase penolakan yang diberikan oleh keenam uji kenormalan ini harus lebih kecil dari persentase penerimaannya. Lima dari enam uji ini telah sesuai dengan yang diharapkan, karena persentase penolakannya lebih kecil dari persentase penerimaannya. Hal ini sesuai dengan kenyataan bahwa sampel yang digunakan berasal dari Distribusi Normal. Tetapi, ada satu uji yaitu D'Agostino-Pearson yang tidak sesuai dengan yang diharapkan karena persentase penolakannya lebih besar dari persentase penerimaannya. Hal ini disebabkan oleh perhitungan uji ini menggunakan gabungan dari hampiran distribusi sampel untuk Skewness dan Kurtosis. Jika semakin jauh perbedaan antara nilai Skewness dan Kurtosis dari simulasinya dengan nilai Skewness dan Kurtosis untuk Distribusi Normal, maka kemungkinan penolakan hipotesis nol oleh uji D'Agostino-Pearson ini semakin besar.

Sedangkan untuk sampel data yang berdistribusi tidak Normal diharapkan bahwa persentase penolakan harus lebih besar dari pada persentase penerimaan. Dari Tabel 7 di atas dapat dilihat bahwa untuk sampel data yang berdistribusi Eksponensial, persentase penolakan dari semua uji ini lebih besar dari persentase penerimaannya. Ini sesuai dengan kenyataan bahwa sampel data yang digunakan berasal dari populasi berdistribusi tidak normal .

Tetapi, dari sampel data yang berdistribusi tidak Normal lainnya (Kai Kuadrat dan Seragam) tidak semua uji memenuhi kriteria ini. Hal ini disebabkan oleh parameter yang digunakan dalam simulasi ini merupakan bilangan bulat acak yang berbeda-beda. Untuk Distribusi Kai Kuadrat disebabkan oleh db, jika db semakin besar maka kemungkinan uji-uji statistik ini untuk menolak hipotesis bahwa sampel data berdistribusi Normal semakin kecil. Sedangkan untuk Distribusi Seragam disebabkan oleh parameter a dan b, jika semakin kecil nilai dari parameter a dan b yang digunakan, maka kemungkinan uji-uji kenormalan ini untuk menolak hipotesis nol semakin kecil.

5. Kesimpulan

- Distribusi Normal merupakan distribusi yang paling penting karena distribusi ini memberikan hampiran yang baik terhadap distribusi teoritis lainnya dan banyak uji statistik yang mengasumsikan datanya mengikuti Distribusi Normal dalam proses analisisnya..
- Secara keseluruhan uji kenormalan yang baik untuk sampel data berdistribusi Normal maupun tidak Normal adalah uji Shapiro-Wilk karena uji ini merupakan pengamatan tertata yang bebas dari skala dan titik pusat serta tergantung pada korelasi antara data yang diberikan dan kecocokan angka normalnya. Sedangkan uji statistik D'Agostino-Pearson dianjurkan untuk data yang dibangkitkan dari Distribusi tidak Normal.

DAFTAR PUSTAKA

Conover, W.J. 1971. *Practical Nonparametric Statistics*. John Wiley and Sons. New York

Dajan, A. 1986. *Pengantar Metode Statistik*. PT Pustaka LP3ES. Jakarta

Dixon, W.J. & F.J. Massey. 1991. *Pengantar Analisis Statistik*. Gadjah Mada University Press. Yogyakarta

Djarwanto. 2001. *Mengenal Beberapa Uji Statistik dalam Penelitian*. Liberty. Yogyakarta

Furqon. 2004. *Statistika Terapan untuk Penelitian*. Alpabeta. Bandung

Gujarati, D. 1999. *Essentials of Econometrics*. McGraw Hill. Boston

Hogg, R.V and E.A. Tanis. 1989. *Probability and Statistical Inference*. MACMILLAN Publishing Company. New York

Keskin, S. 2006. *Comparison of Several Univariate Normality Test Regarding Type I Error Rate and Power of the T[illegible] Simulation Based Small Samples*. http://www.insinet.net/jasr/2006/296-300.pdf

Mandes, M and A. Pala. 2003. *Type I Error Rate and Power of Three Normality Tests*. http://www.ansijournals.com/itj/2003/135-139.pdf

Marques de Sa, J.P. 2007. *Applied Statistics Using SPSS, STATISTICAL, MATLAB, and R*. Springer Berlin Heidelberg. New York

Nasoetion, A.H. dan Barizi. 1985. *Metode Statistika untuk Penarikan Kesimpulan*. PT Gramedia. Jakarta

Oztuna, D. *et al*. 2006. *Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power Under Different Distributions*. http://journals.tubitak.gov.tr/medical/issues/sag.06-36-3/sag-363-7-0510-10.pdf

Patria, B. 2008. *Uji Kenormalan.* http://inparametric.com/bhinablog/download/uji_normalitas.pdf

Peng, G. *et al*. 2004. *Testing Normality of Data Using SAS.* http://www.lexjansen.com/pharmasug/2004/posters/po04.pdf

Ross, S.M. 1987. *Introduction to Probability and Statistics for Engineers and Scientists*. John Wiley & Sons. New York

Seber, G.A.F. 1984. *Multivariate Observation*. John Wiley and Sons. New York

Sheskin, D.J. 2003. *Parametric and Nonparametric Statistical Procedure*. Chapman and Hall (CRC). Boca Raton

Walpole, R.E. and R.H. Myers. 1995. *Ilmu Peluang dan Statistik untuk Insinyur dan Ilmuwan*. ITB. Bandung