ABSTRAK

Pengendalian terhadap mutu menjadi sangat penting karena menyangkut kepuasan konsumen sebagai pelanggan. Dalam statistik dikenal pengendalian mutu secara statistik atau apa yang disebut *Statistical Process Control* (SPC). Untuk memulai proses tersebut data harus diasumsikan berdistribusi normal. Asumsi inilah yang harus diuji kebenarannya agar produk yang dihasilkan nanti benar-benar sesuai dengan kriteria yang diharapkan.

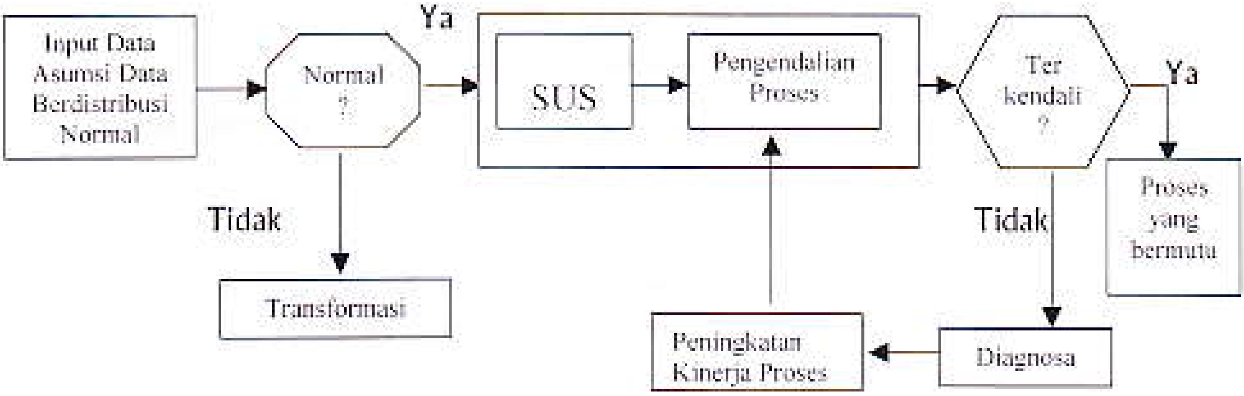
Pengujian kenormalan data disini adalah data univariat. Untuk menguji data tersebut dibuat suatu statistik penguji, yaitu berupa tabel dari nilai R^2 hasil simulasi.

Kata Kunci : *Statistical Process Control*, Statistik Penguji, Data Univariat.

PEMBUATAN STATISTIK PENGUJI KENORMALAN DATA UNIVARIAT

1. Pendahuluan

Mutu adalah suatu hal yang memegang peranan penting dalam sebuah industri. Industri-industri akan selalu berusaha untuk meningkatkan mutu produk. Cara yang dapat dilakukan adalah dengan pengendalian proses secara statistik. Diagram pengendalian proses secara statistik ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Diagram Alir Pengendalian Proses
SUS = Start-Up Stage

2. Masalah Penelitian

Yang menjadi masalah dalam penelitian ini adalah bagaimana membuat statistik penguji untuk menguji kenormalan data univariat yang berdasarkan statistik terurut dari distribusi beta disertai dengan data aplikasi.

3. Tinjauan Pustaka

Distribusi normal univariat (atau normal saja) didefinisikan sebagai berikut.

Definisi 3.1. Variabel random X dikatakan berdistribusi normal dengan mean μ dan variansi σ^2, jika fungsi kepadatan peluangnya adalah

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

dengan $-\infty < x < \infty$, $\sigma^2 > 0$.

Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari distribusi normal $N(\mu, \sigma^2)$. Maka (lihat Hogg and Craig) diperoleh :

$$1.\ \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$2.\ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},\ \text{dengan}\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Selanjutnya akan dikemukakan sifat lain dari distribusi normal yang akan digunakan untuk membentuk statistik penguji.

Definisi 3.2 Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari populasi yang memiliki fungsi distribusi yang kontinu dengan fungsi kepadatan peluang f(x), a < x < b. Misalkan Y_1 yang paling kecil dari $X_i, i = 1,2,\ldots,n$. Y_2 adalah yang terkecil berikutnya dan yang terbesar Y_n dari $X_i, i = 1,2,\ldots,n$. Statistik $Y_1, Y_2, \ldots Y_n$ disebut statistik terurut dari sampel random itu.

Pada teorema berikut dikemukakan fungsi kepadatan peluang gabungan dari $Y_1, Y_2, \ldots Y_n$.

Teorema 3.1.a Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari populasi yang memiliki fkp f(x) yang kontinu. Fungsi kepadatan peluang gabungan dari $Y_1, Y_2, \ldots Y_n$ adalah:

$$g(y_1, y_2, \ldots, y_n) = n!\, f(y_1) f(y_2) \ldots f(y_n), \qquad a < y_1 < y_2 < \ldots < y_n < b$$

$$= 0 \qquad \text{yang lainnya}$$

Selanjutnya dikemukakan pula fungsi kepadatan peluang marginal dari Y_k.

Teorema 3.1.b. Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari populasi yang memiliki fkp f(x) dengan fungsi distribusinya F. Jika $Y_1 \le Y_2 \le \ldots \le Y_n$ merupakan statistik terurut, maka fungsi kepadatan peluang marginal dari Y_k adalah :

$$g_k(y_k) = \int_a^{y_k} \ldots \int_a^{y_2} \int_{y_k}^{b} \ldots \int_{y_{n-1}}^{b} n!\, f(y_1) f(y_2) \ldots f(y_n)\, dy_n \ldots dy_{k+1}\, dy_1 \ldots dy_{k-1}$$

$$= \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)^{k-1}(1 - F(y_k))^{n-k} f(y_k)\right],\ a < y_k < b$$

4. Metode Penelitian

Pada penelitian ini digunakan metode kajian pustaka, di mana peneliti membahas lebih dalam literatur-literatur yang berhubungan dengan masalah penelitian yakni teori mengenai bagaimana membuat statistik penguji kenormalan khusus untuk data univariat yang berdasarkan pada statistik terurut dari distribusi beta.

5. Hasil Penelitian dan Pembahasan.

5.1. Pembentukan Statistik Penguji.

Pada bagian ini akan dijelaskan tentang pembentukan statistik penguji yang dapat digunakan untuk menguji kenormalan.
Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari X. Berdasarkan sampel ini akan diuji hipotesis :

H_0 : X berdistribusi Normal.

H_1 : X tidak berdistribusi Normal.

Untuk itu, perhatikan statistik berikut (lihat Small (1978)).

$$r_i^2 = \frac{\left(X_i - \overline{X}\right)^2}{s^2}, i = 1,2,\ldots,n$$

Pada kedua teorema berikut (lihat Gnanadesikan dan Kettenring (1972)) dikemukakan sifat-sifat dari r_i^2 yang diperlukan untuk pembentukan statistik penguji.

Teorema 5.1.a. Di bawah $H_0, r_i^2 = \dfrac{\left(X - \overline{X}\right)^2}{s^2} \sim \dfrac{(n-1)^2}{n} Beta\left(\dfrac{1}{2}, \dfrac{n-2}{2}\right), i = 1,2,\ldots$

Bukti (lihat Djauhari (1998)):

Kita tahu bahwa $\dfrac{(n-1)s^2}{\sigma^2} = \sum_{k=1}^{n} \dfrac{\left(X_k - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$

Akan ditentukan terlebih dahulu distribusi dari $\left(X_i - \overline{X}\right)$

$$\begin{aligned}
\left(X_i - \overline{X}\right) &= X_i - \frac{\sum_{k=1}^{n} X_k}{n} \\
&= \left(X_i - \frac{1}{n}X_i\right) - \frac{1}{n}\sum_{k \neq i} X_k \\
&= \left(1 - \frac{1}{n}\right)X_i - \frac{1}{n}\sum_{k \neq i} X_k
\end{aligned}$$

$$E\left(X_i - \overline{X}\right) = E(X_i) - E\left(\overline{X}\right) = \mu - \mu = 0$$

$$\begin{aligned}
Var\left(X_i - \overline{X}\right) &= Var\left[\left(1 - \frac{1}{n}\right)X_i - \frac{1}{n}\sum_{k \neq i} X_k\right] \\
&= Var\left[\left(1 - \frac{1}{n}\right)X_i\right] + Var\left[\frac{1}{n}\sum_{k \neq i} X_k\right]
\end{aligned}$$

$$= \left(1-\frac{1}{n}\right)^2 Var(X_i) + \frac{1}{n^2} Var\left(\sum_{k\neq i} X_k\right)$$

$$= \left(1-\frac{1}{n}\right)^2 \sigma^2 + \frac{n-1}{n^2}\sigma^2$$

$$= \left[\frac{(n-1)^2}{n^2} + \frac{(n-1)}{n^2}\right]\sigma^2$$

$$= \frac{n-1}{n^2}(n-1+1)\sigma^2 = \frac{n-1}{n}\sigma^2$$

Diperoleh $\left(X_i - \overline{X}\right) \sim N\left(0, \frac{n-1}{n}\sigma^2\right)$. Akibatnya, $\frac{n}{n-1}\frac{\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{(1)}$

Kemudian, perhatikan statisik berikut :

$$\frac{\left(X_i - \overline{X}\right)^2}{s^2} = (n-1)\frac{\frac{\left(X_i - \overline{X}\right)^2}{\sigma^2}}{\sum_{k=1}^{n}\frac{\left(X_k - \overline{X}\right)^2}{\sigma^2}}$$

atau,

$$\frac{\left(X_i - \overline{X}\right)^2}{s^2} = \frac{(n-1)^2}{n}\frac{\frac{n}{n-1}\frac{\left(X_i - \overline{X}\right)^2}{\sigma^2}}{\sum_{k=1}^{n}\frac{\left(X_k - \overline{X}\right)^2}{\sigma^2}}$$

Akan tetapi, $\sum_{k=1}^{n}\frac{\left(X_k - \overline{X}\right)^2}{\sigma^2} = \frac{n}{n-1}\frac{\left(X - \overline{X}\right)^2}{\sigma^2} + \sum_{k=1}^{n} A_k \frac{\left(X_k - \overline{X}\right)^2}{\sigma^2}$

dengan $A_k = \begin{cases} -\frac{1}{n-1} & \text{untuk } k = i \\ 1 & \text{untuk } k \neq i \end{cases}$

Oleh karena itu kita peroleh :

$$\frac{\left(X_i - \overline{X}\right)^2}{s^2} = \frac{(n-1)^2}{n}\frac{\frac{n}{n-1}\frac{\left(X_i - \overline{X}\right)^2}{\sigma^2}}{\frac{n}{n-1}\frac{\left(X - \overline{X}\right)}{\sigma^2} + \sum_{k=1}^{n} A_k \frac{\left(X_k - \overline{X}\right)^2}{\sigma^2}}$$

atau

$$\frac{n}{(n-1)^2}\frac{(X_i-\overline{X})^2}{s^2}=\frac{\frac{n}{n-1}\frac{(X_i-\overline{X})^2}{\sigma^2}}{\frac{n}{n-1}\frac{(X-\overline{X})}{\sigma^2}+\sum_{k=1}^{n}A_k\frac{(X_k-\overline{X})^2}{\sigma^2}}=\frac{A}{A+B}$$

di mana $A\sim\chi^2_{(1)}$ dan $B\sim\chi^2_{(n-2)}$ dan A dan B saling bebas. Jika A dan (A+B) dibagi oleh 2, maka $A\sim Gamma\left(\frac{1}{2},1\right)$ dan $B\sim Gamma\left(\frac{n-2}{2},1\right)$. Ini berarti bahwa :

$$\frac{n}{(n-1)^2}\frac{(X_i-\overline{X})^2}{s^2}\sim Beta\left(\frac{1}{2},\frac{n-2}{2}\right)$$

atau

$$\frac{(X_i-\overline{X})^2}{s^2}\sim\frac{(n-1)^2}{n}Beta\left(\frac{1}{2},\frac{n-2}{2}\right)$$

(terbukti)

Akibat.

$$U_i=\frac{n}{(n-1)^2}r_i^2\sim Beta\left(\frac{1}{2},\frac{n-2}{2}\right), i=1,2,\ldots,n.$$

Jika $U_{(1)}<U_{(2)}<\ldots<U_{(n)}$ adalah statistik terurut dari $U_1,U_2,\ldots,U_n$, Small (1978) dan Hahn Shapiro (1967) menunjukkan teorema berikut.

Teorema 5.1.b. Jika H_0 benar dan $U_i^*=E(U_{(i)})$, maka $U_{(i)}$ adalah fungsi linier dari U_i^*

Sekarang akan dicari nilai U_i^* yang merupakan nilai ekspektasi dari statistik terurut $U_{(i)}$. Jika F adalah fungsi distribusi dari U_i, ekspektasi dari statistik $U_{(i)}$ adalah sebagai berikut :

$$E(u_{(i)})=\frac{n!}{(i-1)!(n-i)!}\int_0^1 u_i[F(u_i)]^{i-1}[1-F(u_i)]^{n-i}dF(u_i),\qquad i=1,2,\ldots,n$$

Ekspektasi di atas sulit untuk ditentukan nilainya secara eksak. Oleh karena itu digunakan aproksimasi (Hahn dan Shapiro (1967), hal. 293) sebagai berikut :

$U_i^*=E(u_{(i)})=F^{-1}\left(\frac{i-\alpha}{n-\alpha-\beta+1}\right),\ i=1,2,\ldots,n$ dengan $\alpha=-\frac{1}{2}$ dan $\beta=\frac{1}{2}\left(\frac{n-4}{n-2}\right)$

Dengan demikian U_i^* adalah jawab dari persamaan

$$\frac{i-\alpha}{n-\alpha-\beta+1} = \int_0^{U_i^*} f(u_i)du_i, \quad i=1,2,\ldots,n \qquad \ldots\ldots\ldots\ldots(*)$$

di mana f(ui) adalah fungsi kepadatan peluang dari Ui , yakni

$$f(u_i) = \frac{1}{Beta\left(\frac{1}{2},\frac{n-2}{2}\right)} u_i^{\frac{1}{2}-1}(1-u_i)^{\frac{n-2}{2}-1}, \quad i=1,2,\ldots,n$$

Akan tetapi

$$Beta\left(\frac{1}{2},\frac{n-2}{2}\right) = \int_0^1 u_i^{\frac{1}{2}-1}(1-u_i)^{\frac{n-2}{2}-1}du$$

Jadi persamaan (*) diatas dapat dituliskan sebagai berikut :

$$\frac{i-\alpha}{n-\alpha-\beta+1} = \frac{\int_0^{U_i^*} u^{\frac{1}{2}-1}(1-u_i)^{\frac{n-2}{2}-1}du}{\int_0^1 u_i^{\frac{1}{2}-1}(1-u_i)^{\frac{n-2}{2}-1}du}, \qquad i=1,2,\ldots,n$$

Hasil perhitungan diatas memberikan nilai-nilai U_i^*, untuk i=1,2,..,n. Dengan menggunakan program Maple V diperoleh tabel U_i^* berikut, untuk n=5,6,...,10.

Tabel 5.1.a Nilai U_i^*

i \ n	5	6	7	8	9	10
1	0.03500	0.01927	0.01175	0.07769	0.00531	0.00383
2	0.09945	0.05485	0.03334	0.02178	0.01500	0.01079
3	0.20226	0.11176	0.06769	0.04398	0.03017	0.02160
4	0.35450	0.19600	0.11769	0.07586	0.05169	0.03678
5	0.58570	0.32071	0.18929	0.12024	0.08107	0.05722
6		0.52470	0.29542	0.18260	0.12090	0.08432
7			0.47799	0.27534	0.17626	0.12045
8				0.44050	0.25883	0.17032
9					0.40969	0.24475
10						0.38370

Menurut Teorema 2, dibawah H_0 benar, $U_{(i)}$ merupakan fungsi linier dari U_i^* dengan regresi linier sederhana kita punya persamaan $U_{(i)}$ = a + bU_i^*. Dari persamaan regresi linier ini didapat nilai koefisien determinasi dari $U_{(i)}$ dan U_i^* atau R^2. Uji regresi linieritas antara $U_{(i)}$ dan U_i^* adalah dengan nilai R^2−nya. Jadi statistik uji untuk menguji kenormalan adalah nilai R^2.

Permasalah yang muncul adalah seberapa besar nilai R^2 sehingga kita tidak menolak H_0. Untuk itu kita perlu tahu distribusi statistik uji sehingga untuk suatu nilai daerah kritis (α) tertentu dapat diambil suatu keputusan apakah H_0 diterima atau ditolak. Tetapi untuk menentukan distribusi dari R^2 tidak mudah, untuk itu dilakukan simulasi data normal. Simulasi data normal ini dengan menggunakan program MINITAB.

Simulasi data yang akan dilakukan adalah untuk data N(0,1).

Contoh simulasi untuk n = 5, subgrup = 30.

1. Bangkitkan 5 buah data acak berdistribusi N(0,1) sebanyak 30 kali.

$$
\begin{array}{llll}
X_{11} & X_{12} & \cdots\cdots & X_{15} \\
X_{21} & X_{22} & \cdots\cdots & X_{25} \\
\cdots\cdots\cdots & & & \\
X_{301} & X_{302} & \cdots\cdots & X_{305}
\end{array}
$$

2. Transformasikan X_{k1} X_{k2} ...... X_{k5} menjadi $r_{k1}, r_{k2}, \ldots, r_{k5}$.

$$r_{ki}^2 = X_{ki}^2 - \frac{4^2}{5} Beta\left(\frac{1}{2}, \frac{3}{2}\right)$$

Kemudian hitung $U_{ki} = \frac{5}{4^2} r_{ki}^2$, $k = 1,2,\ldots,30$ dan $i = 1,2,\ldots,5$

3. Urutkan $U_{k1}, U_{k2}, \ldots, U_{k5}$, diperoleh $U_{(k1)} < U_{(k2)} < \ldots < U_{(k5)}$; $k = 1,2,\ldots,30$

4. Regresikan $U_{(ki)}$ terhadap U_{ki}^2, i = 1,2,...,5, untuk setiap k = 1,2,...,30.

5. Untuk masing-masing subgrup, regresikan antara $U_{(ki)}$ dengan U_{ki}^2, diperoleh 30 buah nilai R^2, dan terdapat nilai R^2 terkecil.

5.2. Pengujian Kenormalan.

Misalkan $X_1, X_2, \ldots, X_n$ sampel random dari X. Berdasarkan sampel ini akan diuji hipotesis :

H_0 : X berdistribusi Normal.

H_1 : X tidak berdistribusi Normal.

Langkah-langkah pengujian :

Langkah 1 : Hitung r_i^2 , untuk i = 1,2,...,n.

Langkah 2 : Hitung nilai U_i

$$U_i = \frac{n}{(n-1)^2} * r_i^2, \quad i = 1,2,\ldots,n$$

Langkah 3 : Regresikan $U_{(i)}$ terhadap U_i^* , hitung nilai R^2.

Ketiga langkah tersebut untuk tiap-tiap n(n=5,6,...,10) diulangi dengan banyak subgrup=20,21,...,30. Hasil simulasi tersebut memberikan nilai R^2 yang disajikan pada Tabel 5.2 berikut.

Tabel 5.2 Nilai R^2 Simulasi

Subgrup \ n	5	6	7	8	9	10
20	0.8377	0.7225	0.7032	0.8757	0.5819	0.8569
21	0.7937	0.7949	0.7359	0.7868	0.8032	0.7086
22	0.6401	0.7956	0.7789	0.8160	0.7892	0.8312
23	0.7824	0.7709	0.8265	0.7761	0.7855	0.8159
24	0.7985	0.8193	0.7700	0.8007	0.8142	0.7554
25	0.8419	0.6531	0.8271	0.7604	0.8273	0.7763
26	0.7731	0.8046	0.7703	0.7442	0.7855	0.8001
27	0.6918	0.7541	0.6128	0.9897	0.7743	0.7743
28	0.7400	0.7388	0.7460	0.7574	0.7566	0.7071
29	0.7823	0.8104	0.7798	0.7439	0.7639	0.8404
30	0.6185	0.8065	0.7715	0.8301	0.7618	0.7139

Nilai R^2 simulasi inilah yang dijadikan sebagai statistik penguji asumsi kenormalan data univariat. Jika R^2 hitung > R^2 simulasi dapat disimpulkan bahwa H_0 cendrung untuk diterima atau dengan kata lain asumsi kenormalan cendrung dipenuhi.

6. Kesimpulan

- Statistik uji yang dapat digunakan dalam pengujian kenormalan data univariat adalah nilai koefisien determinasi atau koefisien penentu (R^2).
- Jika nilai R^2 hasil simulasi lebih kecil dari R^2 hitung, maka dapat disimpulkan H_0 cendrung untuk diterima atau dengan kata lain asumsi kenormalan cendrung dipenuhi.

7. Ucapan Terima Kasih

Atas terlaksananya penelitian ini maka tidak lupa peneliti mengucapkan terima kasih kepada :

1. Lembaga Penelitian Universitas Andalas yang telah mendanai penelitian ini.
2. Suamiku tercinta Aidinil Zetra, yang telah banyak membantu dalam penyelesaian penelitian ini.
3. Pihak-pihak lain yang tidak dapat penulis tuliskan satu persatu yang telah berkenan membantu sejak awak hingga selesainya artikel ini.

8. Daftar Pustaka

1. Blom, G. (1958). Statistical Estimates and Transformed Beta-variabel. New York.
2. Djauhari, M.A (1998). A Unifying Concept of X Chart and X-Bar Chart When Subgrup Sizes are Equal. Proceeding ITB.
3. Feigenbaum, A.V. (1951). Total Quality Control. McGraw-Hill.
4. Granadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observation. Wiley & Son, New York.
5. Gnanadesikan, R. & Kettenring, J.R. (1972). Robust Estimate, Residual, and Outlier Detection With Multiresponse Data. Biometrics 28, 81-124.
6. Hahn, G.G & Shapiro, S. (1967). Statistical Model in Enginering. Wiley & Son, New York.
7. Hoog, R. V. dan Craig, A.T. (1978). Introduction to Mathematical Statistics. MacMillan Publishing Co, New York, NY.
8. Montgomery, D.C. (1991). Introduction to Statistical Quality Control. John Wiley & Son, New York, NY.
9. Seber G.A.F. (1984). Multivariate Observation. John Wiley & Son, New York, NY.
10. Sembiring, R.K. (1995). Analisa Regresi. Penerbit ITB, Bandung.
11. Small, N.J.H. (1978). Plotting Squred Radii. Biometrika, Vol. 65, hal. 657-658.