Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)
Volume 03 , No.1 (2014), hal 1- 6.

# UKURAN SAMPEL DAN DISTRIBUSI SAMPLING DARI BEBERAPA VARIABEL RANDOM KONTINU

**Muhammad Nurudin, Muhlasah Novitasari Mara, Dadan Kusnandar**

## INTISARI

*Ukuran sampel mempunyai peranan yang penting dalam sebuah penelitian. Ukuran sampel yang diambil tidak hanya mempengaruhi hasil penelitian, tetapi juga menentukan banyaknya biaya dan waktu yang dibutuhkan dalam melakukan penelitian tersebut. Teorema limit pusat menjelaskan bahwa jika sampel berukuran cukup besar, maka distribusi samplingnya akan mendekati Distribusi Normal apapun bentuk awal distribusinya. Penelitian ini dilakukan untuk mengkaji hubungan antara ukuran sampel dan distribusi samplingnya. Berbagai ukuran sampel dari beberapa variabel random kontinu dibangkitkan melalui proses simulasi. Distribusi sampling untuk rata-rata sampel yang dihasilkan melalui proses simulasi tersebut kemudian diuji kenormalannya dengan uji Kolmogorov-Smirnov. Hasil simulasi menunjukkan bahwa ukuran sampel yang dibutuhkan agar distribusi samplingnya mendekati Normal sangat bergantung pada bentuk distribusi datanya. Namun demikian terdapat kecenderungan bahwa setiap distribusi sampling akan mendekati Normal dengan semakin besarnya ukuran sampel.*

***Kata Kunci:*** *Gamma, Uniform, Weibull, Teorema Limit Pusat, Kolmogorov-Smirnov.*

## PENDAHULUAN

Distribusi Normal merupakan salah satu bentuk distribusi yang penting dalam statistika. Banyak metode dan teknik analisis memerlukan asumsi kenormalan sebagai asumsi dasarnya. Berbagai fenomena dalam kehidupan sehari-hari dapat direpresentasikan mengikuti pola Distribusi Normal. Salah satu teorema yang penting berkaitan dengan Distribusi Normal adalah teorema limit pusat. Teorema limit pusat menyatakan bahwa jika dari suatu populasi diambil sampel berukuran cukup besar, maka distribusi sampling dari rata-rata sampel akan mendekati Distribusi Normal apapun bentuk awal distribusinya. Konvensi yang sering digunakan berkaitan dengan penerapan teorema limit pusat ada tiga yaitu; (i) pada umumnya distribusi sampling dari rata-rata sampel akan mendekati Distribusi Normal jika ukuran sampelnya lebih besar dari 30; (ii) jika distribusi populasi asalnya simetris, maka distribusi sampling dari rata-rata sampel akan mendekati Distribusi Normal pada saat ukuran sampelnya lebih besar dari 15; (iii) jika populasi asalnya berdistribusi Normal, maka distribusi sampling dari rata-rata sampel juga akan berdistribusi Normal berapapun ukuran sampelnya [1]. Kriteria umum dalam menggunakan teorema limit pusat banyak yang mendasarkan pada ukuran sampel sama dengan 30, tanpa melihat terlebih dahulu bagaimana bentuk awal dari distribusinya. Hanya dengan kriteria umum seperti ini mungkin tidak cocok untuk berbagai bentuk distribusi peluang, sebab banyak jenis distribusi peluang lain yang memiliki karakteristik yang berbeda dengan Distribusi Normal. Uji Kolmogorov-Smirnov dapat digunakan untuk menguji suatu distribusi peluang apakah distribusi sampling untuk rata-ratanya mendekati Distribusi Normal atau tidak.

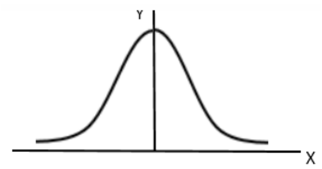
Suatu penelitian yang baik menggunakan sampel sebanyak 10% sampai 30% dari populasi agar diperoleh sampel dengan tingkat kepercayaan yang sesuai [2]. Sebuah studi simulasi *Monte Carlo* oleh Smith dan Wells menunjukkan bahwa ukuran sampel 15 sudah sesuai dengan teorema limit pusat untuk Distribusi Normal, tetapi untuk Distribusi Bimodal pada saat ukuran sampel 30 baru sesuai dengan teorema limit pusat. Sedangkan untuk Distribusi Uniform, ukuran sampel 30 masih belum cocok dijadikan sebagai ukuran sampel untuk merepresentasikan teorema limit pusat [3]. Penelitian mengenai ukuran sampel juga dilakukan menggunakan data berdistribusi Gamma [4]. Uji yang dilakukan dalam penelitian tersebut adalah uji *Shapiro-Wilk*. Pengujian normalitas Distribusi Gamma diuji pada berbagai ukuran sampel dan berbagai parameter. Hasil dari penelitian menunjukkan bahwa teorema limit pusat berlaku pada data tersebut, hanya saja bentuk distribusi peluang ukuran sampelnya bervariasi. Penelitian tentang penentuan ukuran sampel juga dilakukan dari Distribusi Weibull dengan menggunakan uji *Shapiro-Wilk*. Uji ini dilakukan dengan membandingkan nilai parameter distribusi berupa kelipatan 5 dari 1 sampai 100 dengan berbagai ukuran sampel. Diperoleh kesimpulan bahwa hasil penelitian cukup representatif untuk dijadikan dasar bahwa ukuran sampel 30 merupakan nilai yang tepat dan sesuai dengan teorema limit pusat [5].

Berdasarkan teorema limit pusat, dijelaskan bahwa ukuran sampel 30 sudah dianggap Normal apapun bentuk awal distribusinya, padahal distribusi lain ada yang sangat berbeda karakteristiknya dengan Distribusi Normal, meskipun ada beberapa distribusi yang mendekati karakteristik Distribusi Normal. Tujuan penelitian ini adalah untuk mengkaji hubungan antara ukuran sampel dengan bentuk distribusi samplingnya pada Distribusi Uniform, Distribusi Normal, Distribusi Gamma, dan Distribusi Weibull.

## DISTRIBUSI NORMAL, UNIFORM, GAMMA DAN WEIBULL

Distribusi peluang adalah himpunan semua nilai yang mungkin untuk suatu variabel beserta dengan nilai peluangnya masing-masing. Berdasarkan bentuk variabelnya, distribusi peluang dibedakan menjadi dua jenis, yaitu distribusi peluang variabel random kontinu dan distribusi peluang variabel random diskrit. Distribusi peluang variabel random kontinu diantaranya adalah Distribusi Normal, Distribusi Gamma, dan Distribusi Weibull, sedangkan Distribusi Uniform dapat berupa distribusi peluang variabel random kontinu maupun distribusi peluang variabel random diskrit [6].

Distribusi Normal merupakan salah satu distribusi peluang yang penting dan sering digunakan, karena bentuk kurva normal menyerupai lonceng yang menunjukkan keseimbangan antara luas rata-rata ke kanan dan ke kiri masing-masing mendekati 50 %. Distribusi Normal memiliki satu modus sehingga hanya terdapat satu titik puncak (Gambar 1).




Gambar 1. Kurva Distribusi Normal

Bentuk kurva Distribusi Normal dipengaruhi oleh dua parameter yaitu simpangan baku ($\sigma$) dan rata-rata ($\mu$). Nilai dari $\sigma$ menentukan bentangan dari kurva sedangkan nilai dari $\mu$ menentukan pusat simetrisnya. Fungsi kepadatan peluang variabel random Normal $X$, dengan rata-rata $\mu$ dan varians $\sigma^2$, adalah sebagai berikut [7]:

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

Distribusi Uniform kontinu disebut juga sebagai Distribusi Seragam kontinu, distribusi ini memiliki sebaran peluang yang sama pada seluruh interval [$\alpha,\beta$], dengan fungsi kepadatan peluang sebagai berikut [6]:

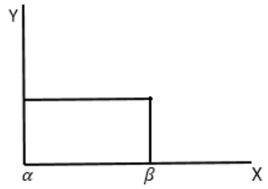
$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{, jika } x \in [\alpha, \beta] \\ 0 & \text{, untuk } x \text{ yang lain} \end{cases}$$

Distribusi Uniform kontinu mempunyai rata-rata dan varians sebagai berikut:

$$E(X) = \mu = \frac{\beta+\alpha}{2}$$

$$Var(X) = \sigma^2 = \frac{(\beta-\alpha)^2}{12}$$

Kurva Distribusi Uniform dapat dilihat pada Gambar 2.




Gambar 2. Kurva Distribusi Uniform

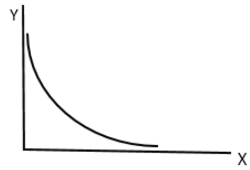
Distribusi Gamma merupakan bagian dari Distribusi peluang kontinu yang memiliki dua parameter yaitu $\alpha$ dan $\beta$, dimana $\alpha$ merupakan *shape parameter*, sedangkan $\beta$ merupakan *scale parameter*. Distribusi Gamma didefinisikan dari fungsi Gamma yang dirumuskan sebagai berikut:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \text{ untuk } \alpha > 0$$

Variabel random kontinu $X$ berdistribusi Gamma, dengan parameter $\alpha$ dan $\beta$, bila fungsi kepadatan peluangnya berbentuk sebagai berikut:

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{, dengan } x > 0 \\ 0 & \text{, untuk } x \text{ yang lain} \end{cases}$$

Kurva Distribusi Gamma ditunjukkan pada Gambar 3.




Gambar 3. Kurva Distribusi Gamma

Rata-rata dan varians Distribusi Gamma adalah sebagai berikut:

$$E(X) = \mu = \alpha\beta$$

$$Var(X) = \sigma^2 = \alpha\beta^2$$

Distribusi Weibull pertama kali diperkenalkan oleh Waloddi Weibull seorang fisikawan dari Swedia pada tahun 1939. Fungsi kepadatan peluang dari Distribusi Weibull dapat diketahui apabila *shape parameter* $\alpha$ dan *scale parameter* $\beta$ diketahui. Variabel random kontinu $X$ berdistribusi Weibull, dengan parameter $\alpha$ dan $\beta$ jika fungsi kepadatan peluangnya berbentuk sebagai berikut:

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} & \text{, dengan } x > 0 \\ 0 & \text{, untuk } x \text{ yang lain} \end{cases}$$

Rata-rata dan varians dari Distribusi Weibull adalah sebagai berikut:

$$E(X) = \mu = \alpha^{-1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$Var(X) = \sigma^2 = \alpha^{-2/\beta} \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 \right\}$$

Gambar 4 berikut menunjukkan Kurva dari Distribusi Weibull.


Gambar 4. Kurva Distribusi Weibull

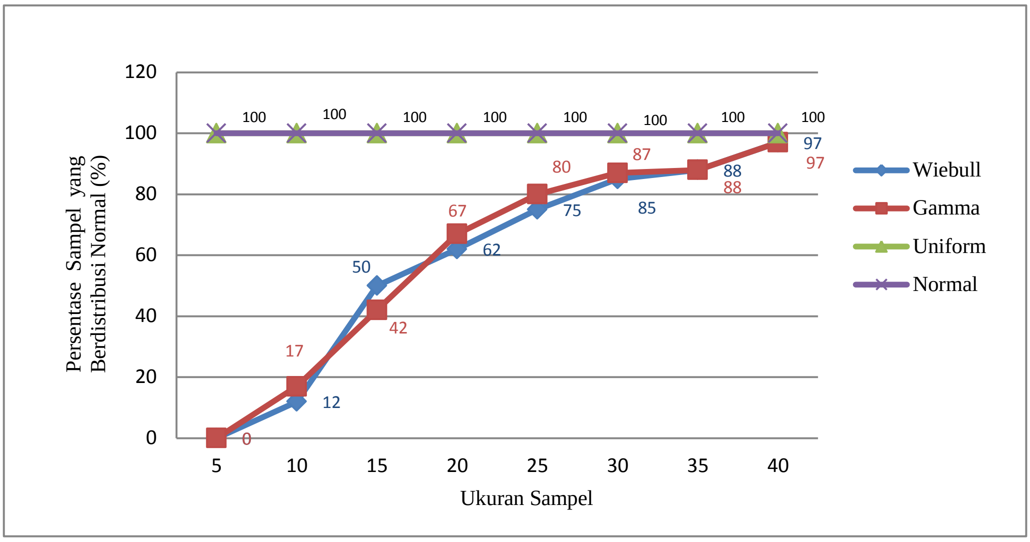
## UJI KOLMOGOROV-SMIRNOV

Uji Normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk mengetahui kenormalan data. Salah satu alat yang dapat digunakan untuk mengetahui normalitas data adalah uji Kolmogorov-Smirnov. Keunggulan uji kolmogorov-Smirnov dibandingkan dengan uji normalitas lainnya adalah uji ini dapat digunakan untuk data yang sangat kecil tanpa harus menggabungkan data yang akan diuji terlebih dahulu, sehingga hasil yang diperoleh lebih akurat. Secara sistematis uji Kolmogorov-Smirnov diawali dengan mengurutkan data yang sudah dirata-ratakan dari data yang terkecil sampai yang terbesar. Selanjutnya dilakukan perhitungan $Z_i = \frac{\bar{x}_i - \bar{\bar{x}}}{S}$ dimana $i = 1,2,3,\ldots,N$, dengan $\bar{x}_i$ merupakan rata-rata dari sampel ke-*i*, $\bar{\bar{x}}$ adalah rata-rata dari seluruh rata-rata sampel yang dibangkitkan dan $S$ merupakan simpangan baku dari rata-rata sampel. Nilai $Z_i$ selanjutnya dibandingkan dengan nilai pada tabel normal baku. Nilai-nilai yang diperoleh merupakan fungsi distribusi kumulatif observasi dari data yang dilambangkan dengan $S_N\left(\bar{x}\right)$. Rumus untuk Kolmogorov-Smirnov adalah $D = maksimum\left|S_N\left(\bar{x}\right) - F_0\left(\bar{x}\right)\right|$, dimana $F_0\left(\bar{x}\right)$ merupakan distribusi frekuensi kumulatif teoritis. Uji hipotesis dalam penelitian ini adalah hipotesis $H_0$ bahwa data berdistribusi Normal lawan $H_1$ data tidak berdistribusi Normal. Taraf nyata yang digunakan adalah 0,05 yang dapat diketahui nilainya dengan pendekatan rumus $\frac{1{,}36}{\sqrt{N}}$ [8]. Pada Uji Kolmogorov-Smirnov yang akan dibandingkan adalah $D_{Hitung}$ dengan $D_N$, dimana $D_{Hitung}$ merupakan nilai dari hasil uji Kolmogorov-Smirnov dan $D_N$ merupakan nilai pada tabel Kolmogorov-Smirnov. Terima $H_0$ jika $D_{Hitung} \leq D_N$ artinya distribusi sampling untuk rata-rata tersebut mendekati Distribusi Normal, dan tolak $H_0$ jika $D_{Hitung} > D_N$ artinya distribusi sampling untuk rata-rata tersebut tidak mendekati Distribusi Normal. Setelah dilakukan pengujian menggunakan uji Kolmogorov-Smirnov, selanjutnya pengambilan kesimpulan apakah data mendekati Distribusi Normal atau tidak mendekati Distribusi Normal.

## HASIL SIMULASI

Proses awal simulasi dilakukan dengan membangkitkan data berupa variabel random yang terdiri dari berbagai ukuran sampel dengan parameter masing-masing distribusi sampling dari beberapa variabel random kontinu. Parameter yang digunakan adalah sebagai berikut: untuk Distribusi Uniform, parameter yang digunakan yaitu nilai minimum = 0 dan nilai maksimum = 1; untuk Distribusi Normal parameter yang digunakan yaitu rata-rata = 0 dan varians = 1; sedangkan untuk Distribusi Gamma dan Distribusi Weibull parameter yang digunakan adalah $\alpha = 1$ dan $\beta = 1$. Pembangkitan data dilakukan sebanyak 1000 kali pengulangan dengan ukuran sampel *n* = 5, 10, 15, 20, 25, 30, 35 dan 40, sehingga diperoleh 1000 data dengan ukuran sampel yang sama untuk masing-masing ukuran sampel. Masing-masing data tersebut dicari rata-rata sampelnya. Kemudian nilai rata-rata diurutkan terlebih dahulu dari yang terkecil hingga terbesar sebelum dilakukan pengujian normalitasnya menggunakan Uji Kolmogorov-Smirnov. Langkah ini dilakukan sebanyak 100 kali pengulangan untuk setiap ukuran sampel. Langkah terakhir dihitung berapa banyak yang Normal dan yang tidak Normal.

Simulasi dan pengujian normalitas pada penelitian ini dilakukan dengan program R versi R2.13.2. Hasil pengujian normalitas terhadap distribusi sampling dari beberapa variabel random kontinu dengan uji Kolmogorov-Smirnov disajikan menggunakan grafik yang ditunjukkan pada gambar 5.




Gambar 5. Hasil Analisis Distribusi sampling untuk Rata-rata dari Empat Distribusi

Hasil simulasi menunjukkan bahwa ukuran sampel tidak berpengaruh terhadap kenormalan distribusi sampling dari data yang berdistribusi Normal dan Distribusi Uniform (Gambar 5). Distribusi sampling yang dibangkitkan dari kedua distribusi tersebut semuanya berdistribusi Normal berapapun ukuran sampelnya. Hasil Distribusi Normal sesuai dengan konvensi yang ada yaitu jika distribusi sampling untuk rata-rata sampel dari sebuah populasi berdistribusi Normal, maka distribusi samplingnya juga akan beristribusi Normal [1].

Untuk Distribusi Gamma terdapat kecenderungan peningkatan persentase distribusi sampling yang berdistribusi Normal dengan meningkatnya ukuran sampel. Pada Distribusi Gamma untuk ukuran sampel 5, menunjukkan tidak ada satu pun distribusi sampling untuk rata-ratanya yang berdistribusi Normal, sedangkan pada ukuran sampel 10, jumlah sampel yang berdistribusi Normal sebanyak 17%. Hasil simulasi menunjukkan pada saat ukuran sampel 30, jumlah sampel yang berdistribusi Normal meningkat menjadi sebanyak 87%, Sedangkan pada saat ukuran sampel sama dengan 40, jumlah distribusi sampling untuk rata-rata yang berdistribusi Normal sebanyak 97%. Pada ukuran sampel 40 diperoleh hampir seluruh ulangan mendekati Distribusi Normal.

Kecenderungan yang sama juga terjadi pada data yang dibangkitkan dari Distribusi Weibull. Pada Distribusi Weibull untuk ukuran sampel 5 diperoleh hasil sama dengan kondisi pada Distribusi Gamma yaitu tidak ada satupun sampel yang berdistribusi Normal dari 100 ulangan yang dilakukan. Pada saat ukuran sampel sama dengan 10, persentase sampel yang berdistribusi Normal sebanyak 12%. Untuk ukuran sampel 30 diperoleh jumlah sampel yang berdistribusi Normal sebanyak 85%, Kondisi ini hanya berbeda 2% ulangan lebih sedikit dari Distribusi Gamma. Distribusi sampling untuk rata-rata yang ukuran sampelnya 40 diperoleh jumlah data yang berdistribusi Normal sebanyak 97%. Hal ini sama dengan banyaknya distribusi sampling untuk rata-rata yang berdistribusi Normal pada Distribusi Gamma dengan ukuran sampel yang sama.

Secara keseluruhan, bahwa dengan semakin besarnya ukuran sampel maka distribusi sampling untuk rata-rata akan mendekati Distribusi Normal apapun bentuk awal distribusinya. Hal ini dapat dilihat dari hasil penelitian yang telah dilakukan. Meskipun terdapat perbedaan hasil distribusi sampling untuk rata-rata yang berdistribusi Normal antara Distribusi Uniform, Distribusi Normal, Distribusi Gamma dan Distribusi Weibull, namun terdapat kecenderungan yang sama yaitu setiap distribusi sampling dari beberapa variabel random kontinu tersebut akan mendekati Normal dengan semakin besarnya ukuran sampel.

## PENUTUP

Berdasarkan hasil analisis dan pembahasan dalam penelitian ini, dapat ditarik kesimpulan bahwa rata-rata sampel yang diambil dari populasi yang berdistribusi Normal dan Berdistribusi Uniform menghasilkan distribusi sampel yang mengikuti kaidah Distribusi Normal. Hal ini terjadi untuk semua ukuran sampel. Sedangkan kenormalan dari distribusi sampling bagi rata-rata sampel pada Distribusi Gamma dan distribusi Weibull terjadi ketika ukuran sampelnya cukup besar, yaitu lebih besar dari 30. Secara keseluruhan, semakin besar ukuran sampel maka distribusi sampling bagi rata-rata sampel akan mendekati Distribusi Normal dan teorema limit pusat berlaku pada keempat distribusi.

**DAFTAR PUSTAKA**

[1]. Kusnandar, D. *Metode Statistik dan Aplikasinya dengan MINITAB dan Excel*. Madyan Press, Yogyakarta. 2004.

[2]. Israel, Glenn D. 2013. *Determining Sample Size,* Program Evaluation and Organizational Development, IFAS Ektension, University of Florida. PEOD-6. June.

[3]. Smith, Z, R., Wells,C S. Central Limit Theorem and Sample Size, *Paper presented at the annual meeting of the Northeastern Educational Research Association.* University of Massachusetts Amherst, New York. 2006.

[4]. Chang, H. J, Wu, C.H., Ho, J. F., Chen, P, Y. On sample Size in Using central limit TheoremI for Gamma Distribution. *Information and Management Sciences*,**19**:153-174. 2008.

[5]. Wu, C, H., Chang, H, J., Huang, K, C. Determination of Sample Size in Using Central Limit Theorem for Weibull Distribution. *Information and Management Sciences*, **17**:31-46. 2006.

[6]. Harinaldi. *Statistik untuk Teknik dan Sains*. Erlangga, Jakarta. 2005.

[7]. Walpole dan Myers. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan.* ITB, Bandung. 1995.

[8]. Siegel, S. *Statistik Non Parametrik untuk Ilmu-ilmu Sosial.* Gramedia, Jakarta. 1992.

MUHAMMAD NURUDIN : Jurusan Matematika, FMIPA UNTAN Pontianak, nurudin_7777@yahoo.com

MUHLASAH NOVITASARI MARA : Jurusan Matematika, FMIPA UNTAN Pontianak, noveemara@gmail.com

DADAN KUSNANDAR : Jurusan Matematika, FMIPA UNTAN Pontianak, dkusnand@yahoo.com