

## PENENTUAN UKURAN SAMPEL UNTUK SURVEY PILKADA MENGGUNAKAN PENDEKATAN BAYES

NENENG SUNENSIH<sup>1</sup>, ENNY SUPARTINI<sup>2</sup>,  
TITI PURWANDARI<sup>3</sup>

1Departemen Statistika FMIPA UNPAD, sunengsih@unpad.ac.id

2Departemen Statistika FMIPA UNPAD, arthinii@yahoo.com

3Departemen Statistika FMIPA UNPAD, titipurwandari@yahoo.com

**Abstrak.** Penentuan ukuran sampel untuk sampling multinomial dapat dibangun melalui konsep bayesian, dimana metodenya didasarkan pada Daerah Toleransi  $R$  yang memuat vektor parameter yang diteliti dengan peluang tertentu dan bergantung pada rata – rata seluruh sampel yang mungkin. Dalam metode ini terdapat dua tipe Daerah Toleransi, yaitu elipsoidal dengan titik pusat pada rata – rata distribusi posterior bersyarat matriks sampel  $X$  dan daerah hiperkubik yang berisikan seluruh interval yang simetris terhadap rata – rata posterior masing-masing parameter. Konsep tersebut pada penelitian ini akan diaplikasikan pada penentuan ukuran sampel yang akan digunakan untuk survey pilkada, dimana sebagai ukuran informasi priornya adalah perolehan suara partai pendukung calon kada pada pemilihan umum dua periode sebelumnya.

*Kata kunci:* teorema bayes, tolerance region, informasi prior.

### 1. Pendahuluan

Pemilihan Kepala Daerah atau Pilkada merupakan bentuk kegiatan berdemokrasi yang sudah hampir menjadi pesta rakyat, sehingga hasil perolehan suara dari para calon Kepala Daerah diharapkan segera diketahui. Untuk memenuhi keinginan ini dapat diperoleh melalui perhitungan cepat secara sampling, dimana sudah barang tentu pertanyaan awal yang akan muncul adalah berapa besar ukuran sampel yang harus dipilih. Pada tahun 2018 akan dilaksanakan pilkada serentak di 17 provinsi 39 kota dan 115 kabupaten, tetapi meskipun Kabupaten Karawang tidak akan melaksanakan pilkada pada tahun depan, teori yang dibahas dalam penelitian ini akan diaplikasikan pada penentuan ukuran sampel untuk perhitungan cepat pada pilkada di Karawang yang akan melaksanakan pemilihan Bupati pada tahun 2020, dalam keperluan ini yang menjadi unit analisisnya adalah berupa pemilih. Pada umumnya calon kepala daerah terdiri atas lebih dari satu pasang calon, sehingga dalam penentuan ukuran sampelnya digunakan sampling multinomial. Penentuan ukuran sampel yang menggunakan konsep presisi dari penaksir terhadap parameternya ( $\theta$ ), telah dirumuskan oleh Cochran dengan konsep klasik melalui pendekatan distribusi normal pada distribusi binomial untuk tingkat presisi 100% sebagai berikut :

$$n \geq \chi_{(1),\alpha}^2 \frac{\theta(1-\theta)}{d^2}$$

Formula di atas hanya berlaku untuk penentuan ukuran sampel dari populasi yang memiliki dua kategori atau kelas dimana dalam perhitungannya memerlukan nilai parameter  $\theta$ , nilai parameter ini bisa diganti oleh nilai taksirannya berdasarkan informasi awal. Untuk keperluan yang lebih luas, yaitu ketika populasinya terdiri atas lebih dari dua kelas, maka perlu dicari formulanya. Pada penelitian ini akan dibahas penentuan ukuran sampel untuk sampling multinomial dan digunakan melalui pendekatan Bayes, dimana konsep dari presisi akan tetap digunakan sebagai dasar perhitungannya.

## 2. Metode

Penelitian ini dilakukan untuk menentukan ukuran sampel apabila sampelnya berdistribusi multinomial seperti yang ditulis oleh Adcock, yang diaplikasikan untuk survey Pilkada di Karawang yang memiliki calon lebih dari satu melalui pendekatan Bayes. Adcock & Smith mendefinisikan bahwa, jika  $\theta$  adalah vektor dari parameter populasi yang diperhatikan,  $\mathbf{X}$  adalah vektor data sampel maka Daerah Toleransi  $\mathbf{R}(\mathbf{X})$  dibangun berdasarkan persamaan berikut :

$$P(\theta \in \mathbf{R}(\mathbf{X}) | \mathbf{X}) = 1 - \alpha$$

Melalui analisis pre-posterior di bawah ini :

$$\int P(\theta \in \mathbf{R}(\mathbf{X}) | \mathbf{X}) p(\mathbf{X}) d\mathbf{X} = 1 - \alpha \quad (1)$$

memungkinkan untuk memilih  $\mathbf{R}(\mathbf{X})$  sebagai dasar dalam penentuan rumusan ukuran sampel  $n$ . Selanjutnya fungsi likelihood untuk sampling multinomial dengan  $k$  kelas dalam notasi standar dituliskan sebagai

$$\rho(\mathbf{n} | \theta) = K \prod_i \theta_i^{n_i} \quad (2)$$

Dalam hal ini  $K$  adalah konstanta dan  $n_i$  adalah ukuran sampel dari kelas ke- $i$ , densitas prior ini diperlukan karena dalam Bayesian parameter dianggap sebagai variabel acak, sehingga adanya informasi prior sangat diperlukan dalam pembentukan distribusi posterior, dimana dalam hal ini diasumsikan bahwa distribusi prior dari parameternya mengikuti distribusi Dirichlet yang dinyatakan dengan  $D(\mathbf{v})$  berikut:

$$\rho(\theta) = K' \prod_i \theta_i^{v_i-1} = D(\mathbf{v}) \quad (3)$$

Pemilihan distribusi Dirichlet dianggap merupakan *conjugate prior* yang wajar (prior yang terkait dengan pola model likelihood datanya), sehingga fungsi densitas posterior dari  $\theta$  dengan syarat  $\mathbf{n}$  dapat dituliskan sebagai :

$$p(\theta | \mathbf{n}) = D(\mathbf{v} + \mathbf{n})$$

(4)

Rata – rata dari parameter individu  $\theta_i$  adalah :

$$E(\theta_i) = v_i / v = e_i \quad \text{dengan} \quad v = \sum v_i$$

Varians dari parameter individu  $\theta_i$  adalah :

$$V(\theta_i) = \frac{e_i(1-e_i)}{(v+1)}$$

Besaran  $e_i$  umumnya selalu tersedia, sedangkan  $v$  merupakan ukuran sederhana dari *prior knowledge* dimana nilai  $v$  terbesar merupakan informasi prior terbaik. Penyelesaian Persamaan (1) untuk Persamaan (2) dan (3) tidaklah mudah, sehingga selanjutnya diasumsikan distribusi posterior untuk Persamaan (4) adalah pendekatan pada distribusi singular normal berikut :

$$\theta | \mathbf{n} \square N(\phi, \mathbf{V} / (v+n+1))$$

Penentuan ukuran sampel ditentukan berdasarkan dua tipe Daerah Toleransi, yaitu elipsoidal dan hiperkubik.

a. Elipsoidal

Ukuran sampel dibentuk berdasarkan persamaan :

$$R = \left\{ \theta : (\theta - \phi)^T \mathbf{V}^{-1} (\theta - \phi) \leq d^2 \right\}$$

Akan menghasilkan Daerah Toleransi  $\sum (\theta - \phi)^2 / \sigma_i \leq d^2$

Dan formula ukuran sampel :

$$n + v + 1 \geq \frac{\chi^2_{(k-1), \alpha}}{d^2}$$

b. Hiperkubik

Daerah Toleransi diperoleh dari interseksi interval individu :

$$R_i = \left\{ \theta_i : |\theta_i - \phi_i| \leq (\sigma_i(1-\phi_i))^{1/2} d \right\}$$

Presisinya proporsional pada *standard error* posterior dari parameter yang diteliti, rumusan yang mudah didapat pada penerapan pertidaksamaan Bonferroni sehingga diperoleh formula ukuran sampel :

$$n + \nu + 1 \geq \frac{\chi_{(1),\alpha/k}^2}{d^2} \quad (6)$$

Selain kedua persamaan diatas, dapat digunakan juga persamaan konservatif berikut :

$$n + \nu + 1 \geq \frac{\chi_{(1),\alpha/k}^2(0.25)}{d^2} \quad (7)$$

Hasil dari persamaan 5, 6 dan 7 akan berbeda karena memperlihatkan perbedaan dari Daerah Toleransi yang di pilih.

Dari ketiga persamaan untuk menentukan ukuran sampel mengandung  $\nu$  yang bisa ditaksir melalui informasi Prior berdasarkan data masa lalu. Dalam hal ini informasi Prior tersebut kaitannya dengan Pilkada dihitung berdasarkan perolehan suara partai pada dua Pemilu Legislatif tahun 2009 dan 2014 di Kabupaten Karawang. Nilai  $\nu$  didapat dari solusi persamaan berikut :

$$E(Q) = \frac{2 \left( 1 - \sum_i e_i^2 \right)}{(\nu + 1)}$$

Dalam hal ini  $Q = (w_1 - w_2)^T (w_1 - w_2)$  sedangkan  $w_1$  dan  $w_2$  adalah penaksir dari  $\theta$  dan  $e_i$  ditaksir oleh rata-rata dari  $w$ .

### 3. Hasil dan Pembahasan

Penelitian ini mengambil asumsi terdapat empat calon Bupati Kabupaten Karawang dimana :

1. Calon nomor urut 1 Independen
2. Calon nomor urut 2 didukung oleh PDIP
3. Calon nomor urut 3 didukung oleh DEMOKRAT
4. Calon nomor urut 4 didukung oleh GOLKAR

Calon Independen diasumsikan didukung oleh pemilih diluar partai PDIP, DEMOKRAT dan GOLKAR. Berdasarkan perolehan suara partai di Kabupaten Karawang pada PILEG 2009 dan 2014 yang bersumber dari website KPUD Kabupaten Karawang yang selengkapnya disajikan dalam Tabel 1 berikut :

**Tabel 1. Perolehan Suara di Kabupaten Karawang**

No.	PARTAI	PILEG'09	PILKADA'15	PILEG'14
1	Independen	40.97%	21.60%	45.95%
2	PDIP	18.45%	18.97%	24.50%
3	DEMOKRAT	23.63%	51.06%	8.69%
4	GOLKAR	16.94%	7.76%	20.87%

Selanjutnya untuk menentukan ukuran sampel yang akan digunakan pada survey PILKADA di Kabupaten Karawang menggunakan Persamaan 5, 6, dan 7 diambil  $\alpha$  sebesar 5% dan Presisi  $100d\% = 5$  diperoleh hasil sebagai berikut :

**Tabel 2. Ukuran Sampel Berdasarkan Daerah Toeransi**

<b>Daerah Toleransi</b>	<b><i>n</i></b>
Elipsoida	3081
Hiperkubik	2609
Konservatif	652

Tabel 2 diatas menunjukkan bahwa ketiga pendekatan tersebut memiliki hasil yang berbeda sesuai dengan konsekuensi dari daerah toleransi yang dipilih. Apabila akan digunakan angka tersebut satuannya adalah pemilih, karena pada PILKADA 2015 di Kabupaten Karawang terdapat jumlah Pemilih tetap sebanyak 1.563.281 pada 2628 TPS.

#### **4. Kesimpulan**

Besarnya ukuran sampel dari sampel yang berdistribusi Multiomial melalui metode penekatan Bayes dapat ditentukan melalui Daerah Toleransi Elipsoida dan Hiperkubik juga dengan presisi yang konstan atau Konservatif menghasilkan besaran yang berbeda karena Daerah Toleransi Hiperkubik tidak memperhatikan banyaknya kelas.

#### **Referensi**

- [1] Adock, C.J. (1987). A Bayesian approach to calculating sample size for multinomial sampling. *The Statistician*, 36(2/3), 155-159.
- [2] Adock, C.J. (1993). An Improved Bayesian Procedure for Calculating Sample Size in Multinomial Sampling. *The Statistician*, 42(2), 91-95.
- [3] Cochran, William G. (1977). Sampling Techniques. *John Wiley & Sons Inc.*
- [4] <http://kpu-karawang.go.id/> yang diakses pada 2 Agustus 2017