

BAB I PENDAHULUAN

1.1 Latar Belakang

Penentuan ukuran sampel merupakan salah satu aspek penting dalam perancangan eksperimen dan survei statistik, terutama dalam studi yang bertujuan untuk mengestimasi proporsi dari populasi (Cochran, 1977). Ukuran sampel yang terlalu kecil dapat menyebabkan estimasi yang tidak akurat, sedangkan ukuran yang terlalu besar bisa menjadi pemborosan sumber daya. Oleh karena itu, diperlukan metode yang efisien dan tepat dalam menentukan ukuran sampel guna memperoleh hasil estimasi yang andal.

Dalam konteks distribusi binomial, salah satu pendekatan yang umum digunakan untuk mengukur ketepatan estimasi proporsi adalah melalui pembentukan selang kepercayaan (*confidence interval*). Selang kepercayaan merupakan rentang nilai yang, dengan tingkat keyakinan tertentu (misalnya 95%), diyakini mengandung nilai parameter populasi yang sebenarnya (Casella & Berger, 2002). Dua metode yang paling sering digunakan untuk membentuk selang kepercayaan proporsi adalah metode Wald dan metode Wilson.

Metode Wald menggunakan pendekatan berdasarkan distribusi normal, di mana selang kepercayaan dibentuk menggunakan estimasi titik dari proporsi ($\hat{p}$) dan simpangan bakunya. Meskipun sederhana dan banyak digunakan karena kemudahannya, metode Wald sering kali tidak memberikan hasil yang akurat, terutama ketika proporsi mendekati 0 atau 1, atau ketika ukuran sampel kecil (Brown, Cai & DasGupta, 2001). Selang kepercayaan yang dihasilkan oleh metode Wald bisa berada di luar batas logis (yaitu kurang dari 0 atau lebih dari 1), serta memiliki tingkat kepercayaan aktual yang lebih rendah dari yang diinginkan.

Sebagai alternatif, metode Wilson yang juga dikenal sebagai score interval diusulkan untuk memperbaiki kelemahan metode Wald. Metode ini mengakomodasi distribusi binomial secara lebih baik dan menghasilkan selang kepercayaan yang lebih simetris dan akurat. Rumus selang kepercayaan Wilson tidak bergantung sepenuhnya pada estimasi titik, melainkan memperhitungkan peran dari nilai *z-score* dan jumlah sampel dalam kedua ujung selang. Metode Wilson terbukti lebih stabil dalam berbagai kondisi, termasuk untuk ukuran sampel kecil dan nilai proporsi ekstrem, serta memiliki tingkat kepercayaan aktual yang lebih dekat dengan nominal (Newcombe, 1998).

Sebagai pendekatan aplikatif dalam penelitian ini, digunakan simulasi untuk merepresentasikan fenomena distribusi binomial secara visual. Dalam simulasi ini, pemain diberikan informasi berupa jumlah total dari sisi yang dilempar secara acak, dan ditugaskan untuk menebak jumlah yang digunakan berdasarkan hasil pelemparan tersebut. Hasilnya akan bergantung pada apakah penjumlahan sisi-sisi dadu sesuai dengan jumlah yang diberikan (misalnya, 6), yang dianggap sebagai kejadian “sukses”. Simulasi ini akan diterapkan skenario nyata dalam menentukan probabilitas sukses



yang kemudian dimodelkan sebagai peubah acak berdistribusi binomial. Dengan memanfaatkan simulasi Monte Carlo, eksperimen permainan dadu ini digunakan untuk membangkitkan data secara acak, yang selanjutnya dianalisis untuk menentukan *expected length* dari selang kepercayaan Wilson.

Dalam mengevaluasi efisiensi dari suatu metode selang kepercayaan, salah satu tolok ukur yang digunakan adalah *expected length* atau panjang harapan dari selang tersebut. *Expected length* menggambarkan seberapa lebar selang kepercayaan yang dibentuk secara rata-rata. Semakin pendek *expected length*, semakin efisien metode tersebut dalam memberikan informasi tanpa mengorbankan tingkat kepercayaan (Zhou & Ling, 2015).

Namun, menghitung *expected length* secara eksak sering kali melibatkan perhitungan integral yang kompleks. Oleh karena itu, digunakan pendekatan aproksimasi orde pertama, yaitu pendekatan matematis yang memanfaatkan ekspansi deret Taylor untuk mendekati nilai *expected length* secara analitik (Mukhopadhyay, 2000). Aproksimasi ini dapat memberikan wawasan yang cukup akurat dengan beban komputasi yang jauh lebih ringan dibandingkan metode eksak.

Dalam penelitian ini, dilakukan perhitungan aproksimasi orde pertama *expected length* dari metode Wilson dalam menentukan ukuran sampel pada distribusi binomial. Tujuan utama dari perbandingan ini adalah untuk mengetahui metode mana yang memberikan efisiensi lebih tinggi dalam membentuk selang kepercayaan yang andal, terutama pada berbagai skenario nilai proporsi dan ukuran sampel. Studi ini penting karena dapat membantu peneliti dan praktisi statistik dalam memilih metode perhitungan ukuran sampel yang lebih tepat guna dan sesuai dengan kondisi data yang dihadapi.

1.2 Batasan Masalah

Batasan masalah dalam penelitian ini adalah

1. Nilai nominal dari setiap selang kepercayaan yang bernilai $1 - \alpha = 95\%$.
2. Nilai banyak dadu yang digunakan (k) adalah 3.
3. Nilai w yang ditetapkan adalah 6.
4. Banyaknya simulasi yang dilakukan adalah 1000 kali.

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, tujuan dari penelitian ini yaitu memperoleh ukuran sampel untuk selang kepercayaan Wilson menggunakan aproksimasi orde pertama dari *Expected Length*.



an

penelitian ini diantaranya:

teoritis, memberikan kontribusi dalam pengembangan bagi penentuan ukuran sampel yang efisien dan akurat serta karakteristik selang kepercayaan Wilson.

2. Memperluas pemahaman tentang keandalan pendekatan aproksimasi dalam penentuan ukuran sampel, yang dapat membantu menghemat sumber daya komputasi dan operasional dalam studi lapangan.

1.5 Teori

1.5.1 Distribusi Binomial

Distribusi binomial digunakan ketika ingin menghitung probabilitas dari jumlah keberhasilan dalam sejumlah uji coba yang dilakukan secara berurutan dan independen. Asumsi-asumsi yang digunakan dalam percobaan binomial adalah (Manurung dkk, 2013):

1. Setiap percobaan memiliki dua kemungkinan hasil (*outcomes*), yakni “sukses” dan “gagal” yang saling independen.
2. Kemungkinan “sukses” ditunjukkan dengan simbol p yang tetap (konstan) dari percobaan ke percobaan berikutnya dan kemungkinan “gagal” ditunjukkan oleh simbol $(1 - p)$.
3. Percobaan-percobaan sebanyak n kali adalah bersifat independen, artinya hasil setiap eksperimen tidak mempengaruhi hasil dari eksperimen yang lain.

Adapun bentuk dari distribusi binomial: (Sugiyarto, 2021):

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \dots, n \quad (1)$$

dengan:

$P(X = x)$ = probabilitas kejadian x terjadi

p = probabilitas “sukses”

n = jumlah total percobaan

x = jumlah sukses dari n kali percobaan

Mean dan variansi dari distribusi binomial secara berturut-turut adalah np dan $np(1 - p)$.

1.5.2 Selang Kepercayaan

Misalkan $X_1, X_2, X_3, \dots, X_n$ merupakan sampel dari peubah acak X dimana X memiliki pdf $f(x; p)$, $\theta \in \Omega$. Diberikan $0 < \alpha < 1$, Misalkan $L = L(X_1, X_2, X_3, \dots, X_n)$ dan $U = U(X_1, X_2, X_3, \dots, X_n)$ merupakan dua statistik. Maka interval (L, U) adalah selang kepercayaan $(1 - \alpha) \cdot 100\%$ untuk θ jika

$$1 - \alpha = P_{\theta}[\theta \in (L, U)] \quad (2)$$

yang mengandung θ adalah $1 - \alpha$ (Hogg, 2019).



1.5.3 Selang Kepercayaan Wilson

Misalkan $X \sim \text{bin}(n, p)$, maka selang kepercayaan Wilson untuk proporsi binomial p memiliki bentuk

$$L(X) = \frac{2X + z_{1-\frac{\alpha}{2}}^2 - z_{1-\frac{\alpha}{2}} \sqrt{z_{1-\frac{\alpha}{2}}^2 + 4X \left(1 - \frac{X}{n}\right)}}{2 \left(n + z_{1-\frac{\alpha}{2}}^2\right)} \quad (3)$$

$$U(X) = \frac{2X + z_{1-\frac{\alpha}{2}}^2 + z_{1-\frac{\alpha}{2}} \sqrt{z_{1-\frac{\alpha}{2}}^2 + 4X \left(1 - \frac{X}{n}\right)}}{2 \left(n + z_{1-\frac{\alpha}{2}}^2\right)} \quad (4)$$

Dengan $z_{1-\frac{\alpha}{2}}$ menyatakan kuantil $1 - \frac{\alpha}{2}$ dari distribusi normal standar (Gonçalves dkk, 2012).

1.5.4 Deret Taylor

Deret Taylor adalah ekspansi fungsi yang cukup halus (terdiferensialkan tak hingga kali) dalam bentuk deret tak hingga yang dibentuk dari turunan-turunan fungsi tersebut pada suatu titik. Definisi umum dari deret Taylor untuk fungsi $g(x)$ di sekitar titik a adalah:

$$g(x) = \sum_{i=0}^{\infty} \frac{g^{(i)}(a)}{i!} (x - a)^i \quad (5)$$

Aproksimasi fungsi oleh deret Taylor disertai suku sisa yang dapat dituliskan sebagai

$$R_k(x) = \frac{g^{(k+1)}(\xi)}{(k+1)!} (x - a)^{k+1} \quad (6)$$

Untuk suatu $\xi \in (a, x)$ (Stewart, 2015).

1.5.5 Penentuan Ukuran Sampel untuk Selang Kepercayaan

Misalkan $X \sim \text{bin}(n, p)$, *Expected Length* dari selang kepercayaan untuk p diberikan oleh



$$p) = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} (U(i) - L(i)) \quad (7)$$

keduanya menyatakan batas atas dan batas bawah dari selang uksek terjadi. Adapun penentuan ukuran sampel berdasarkan

prosedur eksak EL adalah memperoleh nilai n sedemikian sehingga (Gonçalves dkk, 2012)

$$EL(n, p) = \Delta \quad (8)$$

1.5.6 Aproksimasi Orde Pertama

Diberikan suatu peubah acak X , maka ekspansi Taylor dari ekspektasi $g(X)$ di sekitar $E[X]$ diberikan oleh (Benaroya, 2005)

$$E[g(X)] = E \left[g(E[X]) + g'(E[X])(X - E[X]) + \frac{g''(\xi)}{2!}(X - E[X])^2 \right]$$

Untuk suatu $\xi \in (E[X], X)$ sehingga diperoleh

$$E[g(X)] \approx E[g(E[X]) + g'(E[X])(X - E[X])]$$

$$E[g(X)] \approx g(E[X]) + g'(E[X])E[(X - E[X])]$$

karena $E[(X - E[X])] = 0$ maka diperoleh

$$E[g(X)] \approx g(E[X]) \quad (9)$$

Oleh karena itu, aproksimasi orde pertama dari suatu ekspektasi $E[g(X)]$ adalah $g[E(X)]$, yaitu

$$E[g(X)] \cong g[E(X)] \quad (10)$$

Perhatikan bahwa jika $X \sim \text{bin}(n, p)$ dan misalkan $\delta(X) = U(X) - L(X)$, maka

$$EL(n, p) = E[\delta(X)] \cong \delta(E[X]) = U[E(X)] - L[E(X)] = U(np) - L(np)$$

Sehingga penentuan ukuran sampel berdasarkan aproksimasi orde pertama dari EL adalah mencari nilai n sedemikian sehingga

$$U(np) - L(np) = \Delta \quad (11)$$



BAB II METODOLOGI PENELITIAN

2.1 Sumber Data

Data dalam penelitian ini bersumber dari observasi yang berdistribusi $\text{bin}(1, p_{\text{dadu}})$ yang dibangkitkan menggunakan metode Monte Carlo. Data diambil dari sebuah contoh kasus; permainan menebak jumlah dadu yang digunakan. Diberikan sebuah penjumlahan dari sisi-sisi atas k dadu, sebut sebagai w . Dadu-dadu ini akan dilemparkan sebanyak n kali, yang mana akan menghasilkan deretan “sukses” dan “gagal”, yang bergantung pada muncul tidaknya w dari setiap pelemparan. Berdasarkan hasil pelemparan ini akan ditebak berapa dadu yang digunakan pada tiap pelemparan. Tujuan dari permainan ini adalah untuk menemukan ukuran sampel minimum sehingga memperoleh tebakan yang baik untuk setiap putaran permainan.

2.2 Variabel dalam Penelitian

Berdasarkan sumber data yang telah diuraikan, maka dalam penelitian ini akan digunakan jumlah dadu $w = 6$ saat menggunakan $k = 3$ dadu. Atau dengan kata lain, peubah acak dalam penelitian ini adalah peubah acak

$$I \sim \text{bin}(1, P(w = 6 | \text{dadu} = 3))$$

$$I \sim \text{bin}\left(1, \frac{10}{216}\right)$$

2.3 Tahapan Penelitian

Adapun tahapan dalam penelitian ini diantaranya:

1. Menentukan panjang harapan dari selang kepercayaan, yaitu nilai Δ
2. Memperoleh fungsi $n(p)$ dari setiap selang kepercayaan
3. Melakukan spesifikasi p_0 yang digunakan dalam $n(p)$. Berdasarkan informasi prior bahwa $\frac{1}{6^6} \leq p \leq \frac{1}{6}$, pemilihan parameter p_0 yang dapat dilakukan yaitu

$$p_0 = \frac{\frac{1}{6^6} + \frac{1}{6}}{2} = \frac{1 + 6^5}{2 \cdot 6^6}$$

4. Hitung $n(p_0)$ untuk selang kepercayaan Wilson



$$\sim \text{bin}\left(n(p_0), \frac{10}{6^3}\right)$$

hasil

2.4 Diagram Alir

