

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika merupakan ilmu yang berperan dalam kehidupan nyata (Diez dkk., 2015). Statistika berperan dalam menganalisis data dan menarik kesimpulan mengenai karakteristik suatu populasi berdasarkan informasi yang diperoleh dari sampel. Data sampel yang telah dikumpulkan kemudian dianalisis untuk memperoleh informasi yang dapat digunakan dalam menjawab pertanyaan penelitian serta mendukung pengambilan keputusan (Diez dkk., 2019). Secara umum, statistika dibedakan menjadi dua cabang utama, yaitu statistika deskriptif dan statistika inferensi. Statistika deskriptif digunakan untuk menyajikan, merangkum, dan menggambarkan karakteristik data melalui tabel, grafik, maupun ukuran pemusatan dan penyebaran data. Di sisi lain, statistika inferensi biasanya dibentuk melalui dua pendekatan utama, yaitu estimasi parameter dan pengujian hipotesis.

Parameter merupakan nilai numerik yang menggambarkan karakteristik suatu populasi, seperti rata-rata, median, atau ukuran penyebaran data, yang pada umumnya tidak diketahui nilainya secara langsung dan perlu diduga berdasarkan data sampel. Estimasi diterapkan untuk menduga nilai parameter populasi yang tidak diketahui berlandaskan data sampel, sedangkan uji hipotesis diaplikasikan untuk menilai kebenaran dugaan terhadap parameter. Dengan demikian, statistika inferensi menjadi salah satu solusi untuk menghasilkan ukuran statistik yang mampu merepresentasikan karakteristik data secara tepat, terkhusus pada data pengamatan yang umumnya menunjukkan variasi antar observasi (Casella & Berger, 2002).

Salah satu karakteristik yang kerap kali dikaji dalam statistika inferensi adalah ukuran pemusatan yang merupakan salah satu parameter yang penting untuk merepresentasikan karakteristik data. Ukuran pemusatan menggambarkan nilai yang menjadi acuan utama dari sekumpulan data, sehingga kerap kali digunakan sebagai dasar dalam menarik kesimpulan mengenai populasi berdasarkan data sampel. Beberapa ukuran pemusatan yang umum digunakan antara lain *mean*, median, dan modus. Di antara ketiganya, *mean* merupakan ukuran pemusatan yang

kerap kali digunakan karena melibatkan seluruh nilai data dalam perhitungannya sehingga *mean* menjadi ukuran yang informatif dan mudah diinterpretasikan ketika data relatif homogen. Namun demikian, karena memperhitungkan seluruh nilai data secara langsung, *mean* menjadi sensitif terhadap keberadaan pencilan atau nilai ekstrem sehingga nilainya dapat bergeser cukup jauh dan estimasi yang dihasilkan bisa menjadi bias serta kurang merepresentasikan data yang sebenarnya (Bulanov, 2021).

Berbeda halnya dengan median, yang nilainya ditentukan berdasarkan posisi data setelah diurutkan, bukan berdasarkan besaran nilainya secara langsung, sehingga median bersifat lebih *robust* atau tidak mudah terpengaruh oleh pencilan maupun perubahan ekstrem pada data. Oleh karena itu, median kerap kali digunakan karena dianggap lebih mampu merepresentasikan pusat data dibanding *mean*, terutama pada data yang mengandung pencilan, karena median tidak terlalu dipengaruhi oleh nilai ekstrem (Rousseelet & Wilcox, 2018). Kondisi inilah yang mendasari perlunya ukuran pemusatan alternatif yang lebih *robust*, salah satunya median, yang dipandang lebih stabil dalam menggambarkan karakteristik data ketika terdapat pencilan atau penyebaran data yang tidak simetris. Ukuran pemusatan yang *robust* merupakan ukuran yang relatif tidak sensitif terhadap keberadaan pencilan maupun perubahan ekstrem pada data, sehingga mampu memberikan hasil estimasi yang lebih stabil. Kondisi ini menjadi penting karena dalam penerapannya data yang diperoleh kerap kali tidak berdistribusi normal atau tidak simetris, banyak variabel kontinu yang tidak mengikuti distribusi normal dan cenderung mengandung nilai ekstrem. Oleh karena itu, diperlukan metode inferensi yang sesuai untuk median dalam berbagai kondisi distribusi data (Sainani, 2012).

Walaupun median memiliki keunggulan dalam merepresentasikan pusat data, proses inferensi pada median tidak sesederhana pada ukuran pemusatan lain seperti *mean*. Secara matematis *mean* lebih mudah dianalisis karena tidak bergantung pada pengurutan data, sedangkan median berkaitan dengan data yang diurutkan sehingga distribusi samplingnya menjadi lebih sulit ditentukan dan proses inferensinya lebih kompleks (Iseh, 2023). Selain itu, metode inferensi untuk rata-rata telah dikembangkan lebih komprehensif dibandingkan dengan median, sehingga metode untuk median kurang terdukung. Di lain sisi, median juga

menunjukkan sifat yang tidak sederhana, apabila ukuran sampel kecil kerap kali terdapat bias (Rousseelet & Wilcox, 2018). Akibatnya, inferensi pada median memerlukan pendekatan khusus agar dapat dihasilkan kesimpulan yang valid.

Kompleksitas inferensi median tampak jelas terutama pada tahapan pengujian hipotesis. Pengujian hipotesis kerap kali lebih difokuskan pada parameter rata-rata karena prosedurnya lebih sederhana dan memiliki landasan teoritis yang kuat. Hal ini tidak terlepas dari sifat distribusi sampling *mean* yang relatif mudah didekati, terutama melalui teorema limit pusat, sehingga prosedur pengujian hipotesis terhadap *mean* dapat dirumuskan secara eksplisit dan telah banyak tersedia dalam berbagai literatur statistika. Sementara itu, pengujian hipotesis terhadap median, khususnya pada distribusi eksponensial, tidak dapat dilakukan secara langsung dan memerlukan pendekatan yang lebih spesifik. Hal ini disebabkan oleh kompleksitas distribusi sampling dari estimator median yang tidak selalu memiliki bentuk yang sederhana, terutama pada ukuran sampel terbatas. Oleh karena itu, diperlukan pengembangan prosedur uji hipotesis yang sesuai agar inferensi terhadap median dapat dilakukan secara lebih akurat dan dapat dipertanggungjawabkan secara statistik.

Salah satu cara yang dapat diterapkan adalah dengan menetapkan tipe distribusi spesifik pada data, sehingga analisis mengenai median bisa dilakukan dengan cara yang lebih sistematis. Dalam hal ini, distribusi probabilitas memiliki peranan krusial sebagai landasan dalam mengembangkan metode inferensi. Dalam statistika inferensi, distribusi probabilitas digunakan untuk memodelkan pola data serta menjadi dasar dalam proses analisis statistik. Dalam konteks analisis median, pemilihan distribusi tertentu menjadi penting agar karakteristik data dapat direpresentasikan secara akurat. Salah satu distribusi probabilitas yang memiliki pola atau bentuk tidak simetris adalah distribusi eksponensial.

Distribusi eksponensial sering digunakan untuk memodelkan waktu tunggu hingga terjadinya suatu kejadian, seperti waktu antar kedatangan dalam sistem antrian, waktu kegagalan suatu komponen, maupun waktu respons dalam sistem layanan (Ross, 2014). Distribusi eksponensial memiliki sifat *memoryless*, yaitu peluang terjadinya suatu kejadian di masa mendatang tidak bergantung pada waktu yang telah berlalu, yang menjadi karakteristik unik dibandingkan distribusi lainnya

(Balakrishnan & Basu, 1996). Kesederhanaan bentuk matematisnya juga menjadikan distribusi ini kerap kali digunakan dalam analisis keandalan (*reliability analysis*) dan data *lifetime* (Montgomery & Runger, 2014). Oleh karena itu, distribusi eksponensial ini menjadi relevan dalam mengkaji parameter, khususnya median, secara lebih terstruktur.

Distribusi eksponensial merupakan salah satu distribusi yang dapat digunakan sebagai pendekatan untuk menganalisis karakteristik median secara lebih spesifik. Median dinyatakan sebagai nilai yang membagi distribusi menjadi dua bagian dengan sama atau setara dengan persentil ke-50 (Ross, 2014). Median dinyatakan sebagai nilai m yang memenuhi peluang kumulatif $P(X \leq m) = 0,5$ (Casella & Berger, 2002). Dalam distribusi eksponensial, median memiliki bentuk khusus dan dapat diinterpretasikan sebagai waktu terjadinya suatu kejadian pada 50% populasi. Namun, dalam penerapannya nilai dari suatu median itu tidak diketahui secara langsung, sehingga perlu dilakukan proses inferensi untuk menentukan berdasarkan data sampel. Bentuk khusus tersebut menunjukkan bahwa median pada distribusi eksponensial dapat dinyatakan sebagai fungsi eksplisit dari parameter distribusinya, sehingga estimasi median dapat dilakukan secara tidak langsung melalui estimasi parameter tersebut, alih-alih melalui pendekatan non-parametrik berbasis statistik tataan (*order statistics*) yang distribusi samplingnya jauh lebih kompleks sebagaimana telah diuraikan sebelumnya. Pendekatan parametrik semacam ini memungkinkan diperolehnya estimator median dengan sifat-sifat statistik yang lebih terdefinisi, sehingga proses estimasi maupun pengujian hipotesis terhadap median dapat dilakukan secara lebih sistematis.

Untuk dapat menduga nilai median tersebut, baik melalui proses estimasi maupun pengujian hipotesis, dibutuhkan pendekatan metodologis yang sesuai dengan karakteristik distribusi eksponensial. Pendekatan yang kerap kali digunakan adalah metode *maximum likelihood estimation* (MLE), yang digunakan untuk estimasi parameter distribusi berdasarkan data sampel. MLE kerap kali digunakan karena memiliki sifat yang baik, seperti konsisten, efisien, serta optimal secara asimtotik dalam berbagai kondisi (Poudyal, 2021). Meskipun demikian, MLE bukan satu-satunya metode dalam proses estimasi, salah satu metode lain adalah metode momen, di mana metode momen ini adalah metode tertua dalam

menemukan titik estimator (Casella & Berger, 2002). Namun, metode momen memiliki keterbatasan karena estimator yang dihasilkan tidak selalu efisien dan dalam beberapa kasus memiliki variansi yang lebih besar dibandingkan estimator yang diperoleh melalui metode *Maximum likelihood estimation* (MLE), sehingga masih memerlukan pengembangan lebih lanjut (Casella & Berger, 2002). Dengan keunggulan tersebut, MLE menjadi pilihan yang kuat karena kemampuan dalam menghasilkan estimator yang lebih baik dalam banyak penelitian.

Berbagai penelitian telah dilakukan untuk mengkaji estimasi median pada distribusi eksponensial dengan pendekatan yang beragam. Shawiesh (2010) mengembangkan metode estimasi median menggunakan pendekatan *Maximum Likelihood Estimation* (MLE) dan menunjukkan bahwa median dapat diperlakukan sebagai parameter inferensi yang penting. Namun, penelitian tersebut masih terbatas pada aspek estimasi tanpa mengembangkan prosedur pengujian hipotesis yang secara eksplisit ditujukan untuk median. Sementara itu, Al-Athari (2024) mengembangkan interval kepercayaan median yang bersifat *robust* terhadap *outlier*, namun masih belum menyediakan prosedur pengujian hipotesis yang memadai untuk median dalam distribusi eksponensial. Meskipun demikian, sebagian besar kajian terhadap median sebagai ukuran pemusatan alternatif masih terbatas, terutama dalam pengembangan prosedur inferensi yang mencakup estimasi dan pengujian hipotesis secara terintegrasi. Di sisi lain, Mahdavi (2013) mengembangkan metode pengujian hipotesis yang *robust* untuk mengatasi sensitivitas *mean* terhadap *outlier*, namun tetap berfokus pada *mean* sebagai parameter utama.

Dengan demikian, diperlukan pengembangan teknik inferensi untuk median dalam distribusi eksponensial, yang mencakup estimasi dan pengujian hipotesis secara terintegrasi, sebagaimana belum dilakukan secara eksplisit pada penelitian-penelitian sebelumnya. Ketiadaan prosedur pengujian hipotesis yang dirancang khusus untuk median ini bukan sekadar celah teoritis, melainkan berimplikasi praktis: tanpa prosedur tersebut, peneliti kerap kali terpaksa menggunakan uji hipotesis berbasis *mean* meskipun median justru dipilih karena sifat *robust*-nya terhadap pencilan, sehingga tujuan awal penggunaan median menjadi tidak tercapai secara optimal. Hal ini menjadi krusial terutama pada data *lifetime* atau *reliability*

yang umumnya berukuran sampel terbatas dan tidak simetris, sebagaimana telah diuraikan sebelumnya, sehingga validitas kesimpulan statistik mengenai median sulit dipertanggungjawabkan secara teoritis apabila hanya mengandalkan prosedur uji yang dirancang untuk *mean*. Berbeda dengan Shawiesh (2010) yang hanya berfokus pada estimasi median, Al-Athari (2024) yang mengembangkan interval kepercayaan tanpa disertai prosedur uji hipotesis, maupun Mahdavi (2013) yang menitikberatkan pada *mean* sebagai parameter utama, penelitian ini mengembangkan prosedur estimasi sekaligus pengujian hipotesis median secara khusus pada distribusi eksponensial. Oleh karena itu, penelitian ini bertujuan untuk mengestimasi median pada distribusi eksponensial serta menguji hipotesis terhadap median populasi berdasarkan data sampel, sekaligus mengilustrasikan metode estimasi yang digunakan dan kesesuaiannya terhadap karakteristik distribusi eksponensial. Dengan demikian, penelitian ini diharapkan dapat melengkapi kekosongan tersebut dan berkontribusi dalam menghasilkan metode inferensi median yang lebih akurat.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, penelitian ini membahas masalah seperti berikut:

1. Bagaimana mengestimasi median pada distribusi eksponensial?
2. Bagaimana mengkonstruksi interval konfidensi median pada distribusi eksponensial?
3. Bagaimana menguji hipotesis median pada distribusi eksponensial?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Pendugaan parameter median distribusi eksponensial dilakukan menggunakan metode *Maximum likelihood estimation* (MLE).
2. Distribusi sampling dari estimator median diturunkan berdasarkan sifat-sifat estimator *Maximum Likelihood*.
3. Pengujian hipotesis terhadap median dilakukan dengan pendekatan parametrik yang didasarkan pada asumsi distribusi eksponensial.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan pendekatan statistika inferensi untuk mengestimasi nilai median pada distribusi eksponensial serta melakukan uji hipotesis terhadap median populasi berdasarkan data sampel. Penelitian ini juga bertujuan untuk mengilustrasikan metode estimasi median dan kesesuaiannya terhadap karakteristik distribusi eksponensial melalui pendekatan analitik maupun simulasi.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah menambah pemahaman mengenai penerapan metode estimasi dan pengujian median pada distribusi eksponensial sebagai salah satu bentuk ukuran pemusatan data pada distribusi kontinu. Pendekatan ini diharapkan dapat menjadi alternatif yang lebih representatif dibandingkan penggunaan *mean*, terutama dalam kasus data yang memiliki kemencengan (*skewness*) tinggi, serta memberikan kontribusi terhadap pengembangan analisis statistik pada bidang keandalan dan waktu antar-kejadian (*interarrival time*).

