

Pendekatan Bayesian pada Uji Korelasi dan Penerapannya pada Pengujian Hubungan antara Penyebaran Covid 19 dengan Kepadatan Penduduk di Jawa Barat

Siska Ghania Putri*, Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*siskaghaniaput@gmail.com, suliadi@gmail.com

Abstract. Correlation is an analysis to measure the closeness of the linear relationship between two variables. The most widely used correlation test is the Pearson correlation with the t-student distribution approach. Another alternative to the correlation test is to use the Bayesian method, this method uses test statistics on the ordinary frequentist t test and the level of the test is equivalent to the t test, so it can be easily calculated using an ordinary calculator or spreadsheet in excel and can measure evidence that can support the hypothesis. Alternative. The test uses the Restricted Most Powerful Bayesian Test (RMPBT) approach. This approach can draw conclusions from both frequentist and bayesian points of view, because the RMPBT approach involves frequentist methods in the process.. West Java is one of the largest contributors to the confirmed cases of Covid 19 and is the province with the 2nd highest population density in Indonesia. The purpose of this study is to carry out a Bayesian approach to the Correlation Test and its application to testing the relationship between the spread of Covid 19 and population density in West Java. The data used is secondary data in 2020. The results of the Pearson correlation of 0.620 and Bayes Factor 44.67 the strength of evidence to support H_1 is strong.

Keywords: Correlation Coefficient, Bayesian, RMPBT.

Abstrak. Korelasi adalah analisis untuk mengukur keeratan hubungan linear antar dua variabel. Uji Korelasi yang paling banyak digunakan adalah korelasi Pearson dengan pendekatan distribusi *t-student*. Alternatif lain untuk uji korelasi adalah dengan menggunakan metode Bayesian, metode ini menggunakan statistic uji pada uji t *frequentist* biasa dan taraf ujinya setara dengan uji t, sehingga dapat dengan mudah dihitung dengan menggunakan kalkulator biasa atau spreadsheet di excel serta dapat mengukur bukti yang dapat mendukung hipotesis alternatif. Pengujian menggunakan pendekatan *Restricted Most Powerfull Bayesian Test* (RMPBT). Pendekatan ini dapat menarik kesimpulan dari sudut pandang *frequentist* maupun bayesian, karena pendekatan RMPBT melibatkan metode *frequentist* dalam prosesnya. Jawa Barat menjadi salah satu penyumbang terbanyak kasus terkonfirmasi Covid 19 dan menjadi Provinsi dengan Kepadatan Penduduk tertinggi ke 2 di Indonesia. Tujuan dari penelitian ini adalah melakukan pendekatan Bayesian pada Uji Korelasi dan penerapannya pada pengujian hubungan antara penyebaran Covid 19 dengan kepadatan Penduduk Di Jawa Barat. Data yang digunakan adalah data sekunder tahun 2020. Didapatkan hasil korelasi Pearson 0,620 dan *Bayes Factor* 44,67 kekuatan bukti untuk mendukung H_1 adalah kuat.

Kata Kunci: Koefisien Korelasi, Bayesian, RMPBT.

1. Pendahuluan

Korelasi adalah analisis untuk mengukur keeratan hubungan linear antar dua variabel. Uji Korelasi yang paling banyak digunakan adalah korelasi Pearson dengan pendekatan distribusi *t-student*. Dalam pengujian korelasi, para peneliti menggunakan metode klasik (*frequentist*) dengan menggunakan statistik uji atau *p-value* untuk pengujian hipotesis, karena perhitungan lebih mudah dan analisisnya tersedia di hampir semua software statistika. Namun, metode *frequentist* ini memiliki kelemahan salah satunya diantaranya adalah memberikan sedikit informasi terkait tingkat kebenaran hipotesis alternatif ketika menolak hipotesis nol (Wang dkk, 2019).

Alternatif lain adalah menggunakan metode Bayesian (Jeffreys, 1961; Kass&Raftery, 1995), metode ini memungkinkan peneliti untuk mengukur kekuatan bukti yang mendukung hipotesis alternatif dengan menghitung *bayes factor*. Metode Bayesian merupakan metode yang digunakan untuk menggabungkan informasi dari data sampel dengan informasi sebelumnya (prior). Dalam pendekatan Bayesian, parameter model tidak bersifat *fixed*, tetapi merupakan sebuah peubah acak. Dalam Bayesian, informasi awal dari parameter disebut dengan prior, karena bersifat acak maka parameter tersebut mempunyai distribusi yang disebut dengan distribusi prior. *Prior believe* (kepercayaan awal) dapat diperbarui menggunakan teorema Bayesian ketika data amatan diperoleh (William, 2007)

Goddard (2016) mengajukan metode untuk model linear dengan membatasi kelas hipotesis alternatif yang mungkin ke *g-prior* yaitu *Restricted Most Powerful Bayesian Test* (RMPBT). *G-prior* lalu Wang, dkk (2019) mengembangkan pengujian korelasi berbasis Bayesian berdasarkan RMPBT. Kelebihan metode ini adalah dapat mengontrol galat tipe I, menggunakan statistik uji pada uji t *frequentist* biasa dan taraf ujinya setara dengan uji t, sehingga dapat dengan mudah dihitung menggunakan kalkulator biasa atau spreadsheet di excel.

Covid 19 adalah penyakit yang disebabkan oleh penyebaran virus dengan golongan *coronavirus* yaitu *severe acute respiratory syndrome coronaviru-2* (SARs-CoV-2). Covid 19 menyerang sistem pernapasan dan kekebalan tubuh manusia dengan gejala yang muncul setelah masa inkubasi (WHO, 2020). Salah satu faktor penyebaran Covid 19 adalah kepadatan penduduk. Kepadatan penduduk memiliki peran penting dalam penyebaran Covid 19 di Indonesia, hal ini mengacu pada kenyataan bahwa daerah perkotaan dengan kepadatan penduduk yang lebih tinggi dibandingkan daerah pedesaan akan menyebabkan penularan penyakit lebih cepat dengan rantai penyebaran yang lebih kompleks (Hardianto, 2020). Berdasarkan penelitian yang dilakukan oleh Arif dan Sengupta (2020), terdapat 3 negara dengan kepadatan penduduk yang tinggi di India Selatan dengan penyebaran kasus Covid-19 yang tinggi. Hal ini, memperkuat dugaan adanya hubungan antara Covid-19 dengan kepadatan penduduk.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini adalah “Bagaimana Pendekatan Bayesian pada Uji Korelasi dan Penerapannya pada Pengujian Hubungan antara Penyebaran Covid-19 dengan Kepadatan Penduduk di Jawa Barat?”. Selanjutnya tujuan dalam penelitian ini adalah untuk mengetahui pendekatan Bayesian pada uji korelasi dan penerapannya pada pengujian hubungan antara penyebaran Covid 19 dengan kepadatan penduduk di Jawa Barat.

2. Metodologi

Panjang Data dan Sumber Data

Data dari penelitian ini adalah hasil survey sekunder instansional yaitu pengumpulan data dari instansi-instansi berupa studi literature yang ada serta melakukan kajian terdahulu. Data diperoleh dari instansi pemerintah yaitu Pusat Koordinasi dan Informasi Covid-19 Provinsi Jawa Barat (Pikobar, 2020) dan Badan Pusat Statistik Provinsi Jawa Barat. Komponen dari penelitian ini yaitu 27 kota/kabupaten di Jawa Barat sebagai berikut:

Tabel 1. Data Terkonfirmasi Covid 19 dan Kepadatan Penduduk di Jawa Barat

Wilayah	Terkonfirmasi covid19	Kepadatan Penduduk
Kab. Bogor	4304	2246
Kab. Sukabumi	1905	596
Kab. Cianjur	209	590
Kab. Bandung	3952	2167
Kab. Garut	2672	858
Kab. Tasikmalaya	325	688
Kab. Ciamis	593	849
Kab. Kuningan	1598	979
Kab. Cirebon	2214	2244
Kab. Majalengka	1271	1005
Kab. Sumedang	976	760
Kab. Indramayu	788	852
Kab. Subang	503	851
Kab. Purwakarta	1643	1177
Kab. Karawang	5692	1435
Kab. Bekasi	8338	3183
Kab. Bandung Barat	1438	1313
Kab. Pangandaran	135	398
Kota Bogor	3581	9510
Kota Sukabumi	936	6854
Kota Bandung	6319	14970
Kota Cirebon	582	8627
Kota Bekasi	14007	14886
Kota Depok	12289	12403
Kota Cimahi	2753	15798
Kota Tasikmalaya	1977	3869
Kota Banjar	310	1615

Sumber: Pikobar, 2020 & BPS Jabar, 2020

Variabel respon (Y) dalam penelitian ini adalah jumlah terkonfirmasi Covid 19 dan variabel predictor (X) pada penelitian ini adalah Kepadatan Penduduk di Jawa Barat.

Prosedur Penelitian

Teknik pengolahan data menggunakan Minitab dan Excel. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Deskripsi data jumlah terkonfirmasi Covid 19 dan kepadatan penduduk.
2. Pendekatan Bayesian pada uji Korelasi tingkat kabupaten/kota Provinsi Jawa Barat 2020.

Pendekatan Bayesian pada uji korelasi dilakukan dengan langkah-langkah sebagai berikut:

1. Mendeskripsikan data
2. Menghitung korelasi Pearson
3. Uji Bayesian untuk korelasi dengan hipotesis uji:

$$H_0 : \rho_{xy} = 0$$

$$H_1 : \rho_{xy} \neq 0$$

4. Menghitung *Bayes factor*
5. Menghitung nilai *threshold*
6. Kesimpulan

3. Pembahasan dan Diskusi

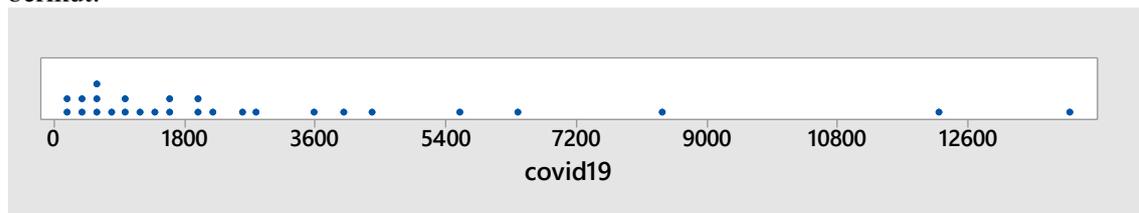
Deskripsi Data

Mendeskripsikan data bertujuan untuk mendapatkan informasi dari variabel yang akan diteliti tanpa bermaksud mengambil kesimpulan untuk umum tetapi hanya menjelaskan variabel tersebut dengan menggunakan analisis deskriptif. Penelitian ini dimaksudkan untuk mengetahui hubungan antara Covid 19 dengan kepadatan penduduk. Dengan deskripsi data adalah sebagai berikut:

Tabel 2. Deskriptif Variabel

Variabel	N	Mean	StDev	Minimum	Maximum
y	27	3011	687	135	1400
x	27	4101	971	398	1575

Bedasarkan tabel diatas, rata-rata jumlah terkonfirmasi Covid 19 (Y) sebanyak 3011 kasus, dengan sebaran data sebesar 687 dan jumlah terendah sebanyak 135 kasus serta jumlah kasus tertinggi sebanyak 1400 kasus. Sementara rata-rata jumlah kepadatan penduduk (X) sebanyak 4101 jiwa/km dengan sebaran data sebesar 971, dan jumlah terendah sebesar 398 dan tertinggi sebesar 1575. Dengan menggunakan dotplot sebagai representasi distribusi data yang terdiri dari kumpulan titik data pada sebuah skala sederhana. Dotplot dapat dilihat sebagai berikut:



$$= 0.620$$

Sebagai tambahan informasi terkait ada/tidaknya hubungan antara Covid 19 dengan kepadatan penduduk, maka ditentukan kekuatan hubungan/korelasi sebagai berikut:

Korelasi atau hubungan yang sempurna: $\rho = -1$ atau $\rho = 1$,

Sangat kuat: $-1 < \rho < -0.85$ atau $0.85 < \rho < 1$,

Cukup kuat: $-0.85 < \rho < -0.65$ atau $0.65 < \rho < 0.85$,

Sedang: $-0.65 < \rho < -0.50$ atau $0.50 < \rho < 0.65$,

Rendah: $-0.50 < \rho < -0.35$ atau $0.35 < \rho < 0.50$,

Sangat Rendah: $-0.35 < \rho < 0$ atau $0 < \rho < 0.35$

(Suliadi, 2015)

Berdasarkan acuan kekuatan hubungan korelasi diatas, $r_{xy} = 0.620$ menunjukkan bahwa penyebaran Covid-19 dan Kepadatan Penduduk memiliki kekuatan hubungan sedang dan arah hubungannya adalah positif, jadi jika Kepadatan Penduduk naik maka penyebaran Covid-19 cenderung naik begitu juga sebaliknya.

Uji Korelasi dengan metode Bayesian

Disini akan dibahas bagaimana uji Bayesian untuk penyebaran Covid-19 Kepadatan Penduduk dengan menggunakan *Restricted Most Powerful Bayesian Test* (RMPBT). Pendekatan ini adalah mekanisme sederhana untuk mengubah statistik uji klasik dan *p-value* menjadi *Bayes Factor*. Teknik ini, memilih distribusi prior dibawah H_1 yang memaksimalkan kuasa uji Bayesian, hasil adalah *Bayes Factor* dengan *threshold* (titik potong) yang tepat dan dapat mengontrol peluang *error* Tipe I sebagaimana pada uji *frequentist* (Wang, dkk. 2019). Dalam metode ini, uji yang dilakukan adalah Bayes tetapi setara dengan *frequentist* dengan taraf uji α dengan hipotesis ujinya adalah:

$H_0 : \rho_{xy} = 0$; tidak ada hubungan antara penyebaran Covid-19 dengan kepadatan penduduk di Jawa Barat.

$H_1 : \rho_{xy} \neq 0$; ada hubungan antara penyebaran Covid-19 dengan kepadatan penduduk di Jawa Barat.

Bayes Factor dan *threshold* diperoleh:

$$t = r_{xy} \sqrt{\frac{(n-2)}{(1-r_{xy}^2)}} = 3.9510$$

$$t_{1-\frac{\alpha}{2},(n-2)} = 2.0595$$

$$BF_{10} = t_{1-\frac{\alpha}{2},(n-2)}^{n-2} \left[\frac{t_{1-\frac{\alpha}{2},(n-2)}^2 + \frac{t^2}{(n-2)}}{1 + \frac{t^2}{(n-2)}} \right]^{-\frac{(n-1)}{2}}$$

$$BF_{10} = 44.66604$$

Nilai *threshold* nya δ adalah sebagai berikut

$$\delta = \left[\frac{t_{1-\frac{\alpha}{2},(n-2)}^2 + (n-2)}{t_{1-\frac{\alpha}{2},(n-2)}^2 / (n-1)} \right]^{\frac{(n-1)}{2}}$$

$$\delta = 2.23671$$

Dengan nilai $BF_{10} = 44,66604$ dan $\delta = 2,23671$ sehingga $BF_{10} > \delta$. Dapat disimpulkan tolak H_0 dan terima H_1 , uji tersebut setara dengan taraf uji 5%, agar peneliti yang biasa menggunakan metode *frequentist* dapat mudah memahami. Dengan ditolaknya H_0 dapat disimpulkan bahwa adanya hubungan antara penyebaran Covid-19 dengan kepadatan penduduk. Dengan kategori penolakan H_0 sebagai berikut:

Tabel 3. Kekuatan bukti menolak H_0

BF_{10}	Strength of Evidence
1 to 3	Kecil
3 to 20	Sedang
20 to 150	Kuat
>150	Sangat Kuat

Sumber: Held, L. & Bove, D. S. (2014)

Berdasarkan tabel 1, dengan $BF_{10} = 44.66604$ disimpulkan bahwa kekuatan bukti untuk menolak H_0 dan menerima H_1 adalah kuat. Hal ini juga menunjukkan bahwa 44.67 kali lebih mungkin untuk menolak H_0 daripada menerima H_0 . Oleh karena itu, dapat ditarik kesimpulan bahwa ada hubungan antara penyebaran Covid-19 dengan kepadatan penduduk.

Berdasarkan teorema bayes, peluang posterior yang mendukung H_j didefinisikan sebagai:

$$P(H_j|Y) = \frac{\pi_j m_j(Y)}{\pi_0 m_0(Y) + \pi_1 m_1(Y)} = \left[1 + \frac{\pi_0}{\pi_1} \frac{1}{BF_{10}} \right]^{-1}$$

Dimana π_j adalah peluang prior dari H_j sehingga ($\pi_1 + \pi_0 = 1$) untuk masing-masing $j = 1,0$. Jika tidak ada pengetahuan tentang prior, biasanya ditetapkan peluang prior yang sama untuk kedua hipotesis yaitu ($\pi_1 = \pi_0 = \frac{1}{2}$) yang disebut dengan asumsi *equipoise*. Maka dari itu, probabilitas posterior yang mendukung H_1 adalah 0.9781.

4. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

Berdasarkan pendekatan Bayesian pada uji Korelasi, dengan korelasi Pearson sebesar $r_{xy} = 0.620$ dan uji $t = 3.9510$ dengan $t_{1-\frac{\alpha}{2};(n-2)} = 2.0595$, didapatkan bahwa $BF_{10} = 44,66604$ dan $\delta = 2,23671$ maka H_0 ditolak karena $BF_{10} > \delta$. Artinya, terdapat hubungan antara penyebaran Covid-19 dengan kepadatan penduduk di Jawa Barat. Dengan probabilitas prior yang mendukung H_1 sebesar 0,9781. Dengan BF_{10} sebesar 44,66604 dikategorikan memiliki bukti untuk menolak H_0 dan menerima H_1 adalah kuat. Artinya, dengan menggunakan metode Bayesian dapat mengukur bukti yang mendukung H_1 .

Acknowledge

berterima kasih kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga dapat terselsaikannya studi ini. Kedua orang tua dan suami tercinta karena telah mendukung sepenuhnya dan mendoakan penulis sehingga penulis dapat menyelesaikan studi ini. Kepada Bapak Suliadi, S.Si., M.Si., Ph.D., yang telah meluangkan waktu untuk membimbing penulis hingga studi ini selesai. Serta ibu/bapak Dosen Program Studi Statistika Universitas Islam Bandung yang telah membekali penulis ilmu sehingga penulis dapat menyelesaikan studi ini. Teman-teman yang telah memberi dukungan dan berjuang bersama hingga penulis dapat menyelesaikan studi ini hingga akhir. Pihak-pihak terkait yang telah mendukung hingga studi ini selesai.

Daftar Pustaka

- [1] Arif, M., & Sengupta, S. 2020. *Nexus between population density dan COVID19 pandemic in the south Indian State. :A geo-statistical approach*, 1-22.
- [2] BPS. 2020. *Data Kependudukan*. Jawa Barat.
- [3] Goddard, S. D. 2016. *Restricted Most Powerful Bayesian Tests for Linear Models*. Scandinavian Journal Statistics, 1162-1177.
- [4] Hardianto, J. 2020. *Korelasi Kepadatan Penduduk dan Penyebaran Covid-19*. Rujak Centre for Urban Studied.
- [5] Jeffreys, H. 1961. *Theory of Probability*. Oxford: UK Oxford University Press.
- [6] Kass, R. E., & Raftery, A. E. 1995. *Bayes Factor*. Journal of the American Statistical Association, 90, 377-39.
- [7] Pikobar. 2020. *Data Penyebaran Covid-19*. Retrieved from pikobar.jabarprov.go.id. Diakses pada 25 Februari 2021.
- [8] Suliadi. 2015. *Analisis Regresi*. Diktat Kuliah Program Studi Statistika.
- [9] Wang, M., Chen, F., Lu, T., Dong, J. 2019. *Bayesian T Test for Correlation and Partial Correlation*. Journal of Aplied Statistic.
- [10] WHO. 2020. *Coronavirus*. Retrieved from www.who.int/health-topics/coronavirus. Diakses pada 8 Maret 2021.
- [11] William, M. B, (2007). *Introduction to Bayesian Statistics, 2nd ed*. New Jersey: John Wiley.
- [12] Irawadi Jody Alwin, Sunendiari Siti. (2021). *Penerapan dan Perbandingan Tiga Metode Analisis Pohon Keputusan pada Klasifikasi Penderita Kanker Payudara*. Jurnal Riset Statistika, 1(1), 19-27.