

# Pendekatan Metode Bayesian pada Uji Korelasi *Kendall Tau* pada Pengujian Hubungan antara Luas Lahan Padi dan Produktivitas Padi di Indonesia

Riyanto Wahyunardi\*, Suliadi

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

\*belutse7@gmail.com, suliadi@gmail.com

**Abstract.** Correlation is used to measure the closeness of the relationship between two variables (bivariate). Pearson's Product Moment is a measure of correlation in parametric statistics, while Rank Spearman, and Kendall Tau are included in nonparametric statistical correlation measures. The Kendall Tau correlation has an advantages over the Pearson correlation, which is stronger against outliers and violations of normality. A new method to test the Kendall Tau correlation was proposed by Van Doorn, et al (2018), namely the Bayes approach with parametric yoking where the yoking parametric is the same as the prior distribution in the parametric case. The advantage is that the prior distribution is independent of the sample size. In this study, we will apply a new method proposed by Van Doorn et al (2018) to examine the relationship between paddy land area and paddy productivity in Indonesia. The results can be used to measure the extent to which paddy land area can support the effectiveness and efficiency of rice productivity in Indonesia. The results showed that by using bayesian method in correlation test Kendall Tau ( $\tau$ ) obtained the value of  $BF_{01} = 0.0366$  and  $BF_{10} = 27.334$ . Where  $H_0$  rejected because of  $BF_{01} < 1$ . This means that there is a relationship between Paddy Land Area and Paddy Productivity. This  $H_0$  rejection ( $H_1$  acceptance) is supported by strong evidence because the  $BF_{10}$  value is at intervals of  $10 < BF_{10} < 30$ .

**Keywords:** Paddy Land Area and Paddy Productivity, Kendall Tau Correlation, Bayesian Method.

**Abstrak.** Korelasi digunakan untuk mengukur keeratan hubungan antara dua variabel (*bivariate*). *Product Moment Pearson* merupakan ukuran korelasi dalam statistik parametrik, sedangkan *Rank Spearman*, dan *Kendall Tau* termasuk ke dalam ukuran korelasi statistik nonparametrik. Korelasi *Kendall Tau* memiliki kelebihan dibandingkan korelasi *Pearson* yaitu lebih kuat terhadap pencilan dan pelanggaran normalitas. Metode baru untuk menguji korelasi *Kendall Tau* dikemukakan oleh Van Doorn, dkk (2018) yaitu pendekatan Bayes dengan parametrik yoking dimana parametrik yoking ini sama saja seperti distribusi *prior* pada kasus parametrik. Kelebihannya distribusi *prior* menjadi tidak tergantung pada ukuran sampel. Dalam penelitian ini kami akan menerapkan metode baru yang diusulkan oleh Van Doorn, dkk (2018) untuk menguji hubungan antara luas lahan padi dan produktivitas padi di Indonesia. Hasilnya dapat digunakan untuk mengukur sejauh mana luas lahan padi (X) dapat mendukung efektifitas dan efisiensi produktivitas padi (Y) di Indonesia. Hasil penelitian menunjukkan bahwa dengan menggunakan metode Bayesian pada uji korelasi *Kendall Tau* ( $\tau$ ) didapatkan nilai  $BF_{01} = 0,0366$  dan  $BF_{10} = 27,334$ . Dimana  $H_0$  di tolak karena  $BF_{01} < 1$ . Artinya ada hubungan antara Luas Lahan Padi dengan Produktivitas Padi. Penolakan  $H_0$  (penerimaan  $H_1$ ) ini didukung oleh bukti yang kuat karena nilai  $BF_{10}$  berada pada interval  $10 < BF_{10} < 30$ .

**Kata Kunci:** Luas Lahan Padi dan Produktivitas Padi, Korelasi *Kendall Tau*, Metode Bayesian.

## 1. Pendahuluan

Metode statistika yang mempelajari tentang korelasi terdapat pada statistik parametrik dan statistik nonparametrik. Dalam statistik parametrik, ukuran korelasi yang digunakan adalah koefisien korelasi *Product Moment Pearson*. Korelasi yang termasuk dalam kategori statistik nonparametrik ini diantaranya *Rank Spearman*, *Kendall Tau*, *Poin Biserial*, *Phi*, *Gamma*, dan *Somers*.

Korelasi *Product Moment Pearson* merupakan korelasi yang digunakan untuk data berskala interval atau rasio, serta cocok digunakan untuk data yang tidak memiliki pencilan. Namun apabila datanya memiliki pencilan atau adanya pelanggaran asumsi, disarankan analisis korelasi menggunakan statistik nonparametrik, misalnya menggunakan korelasi *Kendall Tau* yaitu lebih kuat terhadap pencilan dan pelanggaran normalitas (Kendall dan Gibbons 1990).

Prinsip perhitungan koefisien korelasi *Kendall Tau* dengan cara mengurutkan salah satu variabel biasanya variabel  $X$ . Sedangkan variabel  $Y$  akan dilihat apakah nilai variabel  $Y$  itu searah (konkordan) atau berlawanan arah (diskordan) dengan variabel  $X$  yang sudah diurutkan.

Menguji signifikansi koefisien korelasi *Kendall Tau* jika  $n \leq 10$  dapat dicari dengan menggunakan tabel koefisien korelasi *Kendall Tau*. Apabila nilai  $\tau$  hitung lebih besar dari  $\tau$  tabel maka pengujian signifikan (tolak  $H_0$ ), sebaliknya jika  $\tau$  hitung lebih besar dari  $\tau$  tabel maka pengujian tidak signifikan (terima  $H_0$ ). Namun jika  $n > 10$ , distribusi yang digunakan adalah pendekatan normal (Siegel, 1994:273)

Pengujian koefisien korelasi *Kendall Tau* menggunakan pendekatan metode bayes sebelumnya pernah dilakukan pada penelitian Yuan dan Johnson (2008). Metode Bayes ini memiliki kelebihan yaitu tidak hanya sekedar menolak atau menerima hipotesis  $H_0$ , namun apabila  $H_0$  ditolak dan  $H_1$  diterima, akan terdapat kategorisasi seberapa kuat  $H_0$  tersebut ditolak. Untuk mengatasi kekurangan-kekurangan dari metode sebelumnya, Van Doorn, dkk (2018) mengajukan metode baru yaitu pendekatan bayes dengan menggunakan parametrik yoking. Hal ini bertujuan agar hipotesis satu tidak tergantung kepada ukuran sampel. Metode bayes ini mendefinisikan kembali  $H_1$  sehingga tidak bergantung pada ukuran sampel.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut:

1. “Bagaimana prosedur pengujian hipotesis analisis korelasi *Kendall Tau* menggunakan metode Bayes dengan parametrik yoking”
2. “Bagaimana hasil penerapan pendugaan metode Bayes pada korelasi *Kendall Tau* terhadap kekuatan hubungan antara luas lahan padi dan produktivitas padi di 34 provinsi pada tahun 2020”

Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Mengetahui prosedur pengujian hipotesis analisis korelasi *Kendall Tau* menggunakan metode Bayes dengan parametrik yoking.
2. Untuk mengetahui hasil penerapan pendugaan metode Bayes pada *Kendall Tau* terhadap kekuatan hubungan antara luas lahan padi dan produktivitas padi di 34 provinsi pada tahun 2020.

## 2. Metodologi

### Bahan

Data dan sumber data yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh dari Badan Pusat Statistik Indonesia (BPS) tahun 2020. Data ini mempunyai skala pengukuran interval. Variabel pada data ini yang akan digunakan yaitu luas lahan padi dan produktivitas padi tahun 2020 di 34 Provinsi Indonesia yaitu dari mulai Provinsi Aceh sampai Papua dengan total keseluruhan ada 34 Provinsi (BPS, 2020).

### Variabel Penelitian

Variabel – variabel yang digunakan dalam penelitian dapat dilihat pada Tabel 1 di bawah ini:

**Tabel 1.** Variabel Penelitian

Variabel	Keterangan
Y	Produktivitas Padi
X	Luas Lahan Padi

**Metode Analisis Penelitian**

1. Deskripsi data Luas Lahan Padi dan Produktivitas Padi di 34 provinsi tahun 2020 meliputi uji normalitas data.
2. Melakukan Pendekatan Bayes pada Uji korelasi Kendall Tau dan penerapannya pada pengujian se-berapa erat hubungan antara Luas Lahan Padi dengan Produktivitas Padi di 34 provinsi tahun 2020.

**3. Pembahasan dan Diskusi****Deskripsi Data**

Deskripsi data merupakan tahap awal yang dilakukan untuk mengetahui gambaran umum dari data yang digunakan pada penelitian. Berikut adalah data Luas Lahan Padi dan Produktivitas Padi dari 34 Provinsi yang ada di Indonesia.

**Tabel 2.** Data Luas Lahan Padi dan Produktivitas Padi

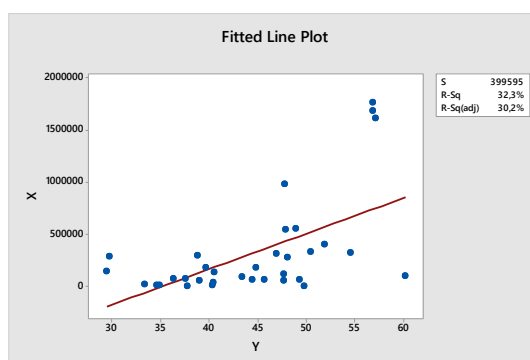
No.	Provinsi	Luas Lahan padi (ha)	Produktivitas padi (ku/ha)
1.	Aceh	320.752,85	54,62
2.	Sumatera Utara	400.300,96	51,87
3.	Sumatera Barat	309.365,07	46,90
...	...	...	...
...	...	...	...
...	...	...	...
...	...	...	...
31.	Maluku	29.643,85	40,42
32.	Maluku Utara	10.608,22	40,33
33.	Papua Barat	7.420,16	34,90
34.	Papua	52.712,71	39,00

Sumber : BPS, 2020

**Tabel 3.** Statistik Deskriptif Luas Lahan Padi dan Produktivitas Padi

Variabel	Total	N	Mean	SE. Mean	St. Dev	Min.	Q1	Median	Q3	Max.
X	34	34	317.259	82.011	478.205	360	52.174	122.468	321.927	1.761.882
Y	34	34	44,52	1,37	7,96	29,48	38,59	45,31	49,48	60,20

Berdasarkan hasil Tabel 3 diperoleh nilai *mean* dari variabel X (Luas Lahan Padi) sebesar 317.259 ha dan nilai *mean* dari variabel Y (Produktivitas Padi) sebesar 44,52 ku/ha dengan nilai minimum dari variabel X sebesar 360 ha dan sebesar 29,48 ku/ha dari variabel Y, nilai maksimum dari variabel X sebesar 1.761.882 ha dan sebesar 60,20 ku/ha dari variabel Y. Pola hubungan variabel X dengan Y sebagai berikut :

**Gambar 1.** Hubungan Luas Lahan Padi dengan Produktivitas

Dari hasil plot Gambar 1 menunjukkan adanya kecenderungan yang tidak terlalu kuat hubungan antara variabel X dengan variabel Y. Selain itu tampak ada tiga titik yang jauh dari pola umum.

#### **Pendekatan Bayesien pada uji Korelasi *Kendall Tau* dan Penerapannya pada pengujian seberapa erat hubungan antara Luas Lahan Padi dengan Produktivitas Padi di Indonesia.**

Akan dicari terlebih dahulu koefisien korelasi *Kendall Tau* lalu selanjutnya pengujian hipotesis dengan pendekatan Bayesien. Sebelumnya akan dicari nilai *S* dengan rumus sebagai berikut:

$$\hat{t} = \frac{S}{\frac{n(n-1)}{2}} \quad (1)$$

Dari rumus (1) kita mencari *S* terlebih dahulu yaitu dengan mengurutkan data, lalu dicari *concordant* dan *discordant*. Banyaknya *Y* yang *concordant* dinyatakan sebagai *P* dan *discordant* sebagai *Q* sehingga nanti untuk mencari *S* dapat menggunakan persamaan  $S = P - Q$ . Sebelumnya data penelitian diurutkan terlebih dahulu dari data terkecil ke terbesar sebagai berikut:

**Tabel 4.** Data pengurutan Luas Lahan Padi dan Produktivitas Padi

Provinsi	Variabel X	Variabel Y	R(X)	R(Y)	Concordant	Discordant
Kep. Riau	359,96	37,8	1	8	26	7
DKI Jakarta	934,24	49,85	2	27	7	25

Papua Barat	7.420,16	34,9	3	5	27	4
Maluku Utara	10.608,22	40,33	4	12	21	9
Kalimantan Utara	11.605,46	34,66	5	4	26	3
Kep. Bangka Belitung	20.451,36	33,39	6	3	26	2
Maluku	29.643,85	40,42	7	13	20	7
Gorontalo	50.556,73	47,68	8	21	12	14
Papua	52.712,71	39	9	10	20	5
....	....	....	....	....	....	....
Aceh	320.752,85	54,62	26	30	3	5
Banten	325.450,19	50,5	27	28	4	3
Sumatera Utara	400.300,96	51,87	28	29	3	3
Lampung	544.061,31	47,88	29	23	4	1
Sumatera Selatan	551.242,08	48,92	30	25	3	2
Sulawesi Selatan	978.192,54	47,83	31	22	3	0
Jawa Barat	1.613.828,78	57,13	32	33	0	2
Jawa Tengah	1.684.746,24	56,9	33	32	0	1
Jawa Timur	1.761.881,97	56,88	34	31	0	0
JUMLAH					$P = 387$	$Q = 174$

Dari tabel 4. didapat *concordant* dan juga *discordant*. Dari tabel diatas nilai *concordant* pertama diperoleh dari R(Y) dimana pada baris pertama yang bernilai > 8 ada sebanyak 26, lalu baris kedua yang bernilai > 27 ada sebanyak 7, dan seterusnya sampai bawah. Sedangkan *discordant* diperoleh dari nilai R(Y) dimana pada baris pertama yang bernilai < 8 ada sebanyak 7, lalu baris kedua yang bernilai < 27 ada sebanyak 25, dan seterusnya sampai bawah. Dimana P adalah jumlah *concordant* sebesar 387 dan Q adalah jumlah *discordant* sebesar 174. Sehingga nilai dari koefesien korelasi Kendall Tau adalah :

$$\hat{\tau} = 0,3797$$

Selanjutnya akan dilakukan pengujian apakah korelasinya signifikan atau tidak secara *frekuentist*.

Hipotesis

$H_0$  : Variabel X dan Y saling bebas

$H_1$  :  $\tau > 0$ , terdapat hubungan positif antara Luas Lahan Padi dan Produktivitas Padi

Taraf  $\alpha = 5\%$ , dengan ukuran sampel  $n = 34$  dan statistik uji  $\hat{\tau} = 0,3797$

Dilihat dari tabel harga-harga kritis statistik uji Kendall Tau, karena uji pihak kanan maka kriteria ujinya yaitu :

Tolak  $H_0$  jika  $\hat{\tau} > \tau^*$

$\tau^*$  diperoleh dari table Kendall Tau, karena  $n = 34$  dengan  $\alpha = 5\%$ , diperoleh  $\tau^* = 0,201$  sedangkan  $\hat{\tau} = 0,3797$ . Karena  $0,3797 > 0,201$  maka  $H_0$  ditolak sehingga data yang kita yakin sebesar 95% terdapat hubungan positif antara Luas Lahan Padi dan Produktivitas Padi.

#### Uji Bayesien untuk Korelasi Kendall Tau

Hipotesis uji yang digunakan dengan parametrik yoking adalah sebagai berikut:

$H_0 : \theta = 0$  ; tidak ada hubungan antara Luas Lahan Padi dengan Produktivitas Padi di Indonesia.

$H_1 : \theta \neq 0$  ; ada hubungan antara Luas Lahan Padi dengan Produktivitas Padi di Indonesia.

Selanjutnya setelah menentukan hipotesis, maka akan menghitung nilai  $T^*$  yang nantinya akan digunakan untuk pengujian Bayes Factor sebagai berikut:

$$T^* = \frac{\sum_{1 \leq i < j \leq n} Q((x_i, y_i), (x_j, y_j))}{\sqrt{n(n-1)(2n+5)/18}} \quad (2)$$

$$\text{dimana } Q((x_i, y_i), (x_j, y_j)) = \begin{cases} -1 & \text{jika } (x_i - x_j)(y_i - y_j) < 0 \\ +1 & \text{jika } (x_i - x_j)(y_i - y_j) > 0 \end{cases}$$

**Tabel 5.** Indikator  $Q$  Luas Lahan Padi dan Produktivitas Padi

$I$	$J$	$(x_i - x_j)$	$(y_i - y_j)$	$Q$
1	2	320752,85 - 400300,96	54,62 - 51,87	-1
1	3	320752,85 - 309365,07	54,62 - 46,9	+1
....	....	....	....	....
2	3	400300,96 - 309365,07	51,87 - 46,9	+1
....	....	....	....	....
30	33	64656,69 - 7420,16	49,36 - 34,9	+1
....	....	....	....	....
33	34	7420,16 - 52712,71	34,9 - 39	+1
JUMLAH				$Q = 213$

Dari tabel 5 menyatakan hasil untuk indikator  $Q$  yang akan digunakan pada rumus  $T^*$ . Sehingga nilai  $T^* = 3,2713$

Selanjutnya akan dicari nilai dari  $BF$  (Bayes Factor) dengan rumus sebagai berikut:

$$BF_{01} = \frac{\exp(-\frac{T^{*2}}{2})}{\int_{-1}^1 \exp\left(-\frac{(r - \frac{3}{2}r\sqrt{n})^2}{2}\right) \left(\pi \frac{2^{-2\alpha}}{B(\alpha, \alpha)} \times \cos\left(\frac{\pi r}{2}\right)^{2\alpha-1}\right) dr} \quad (3)$$

Dengan didalam rumus tersebut terdapat prior dan porterior rumus prior bagi Kendall Tau sebagai berikut:

$$p(\tau) = \pi \frac{2^{-2\alpha}}{B(\alpha, \alpha)} \times \cos\left(\frac{\pi \tau}{2}\right), \quad \tau \in (-1, 1) \quad (4)$$

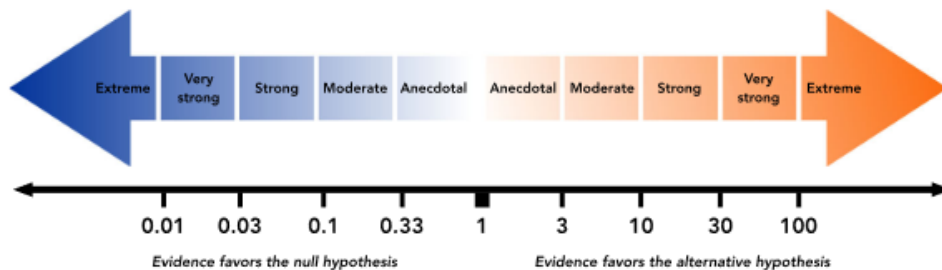
Dan rumus *posterior* sebagai berikut:

$$p(T^*|\theta_0 + \Delta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(T^* - \frac{3}{2}\Delta\sqrt{n})^2}{2}\right) \quad (5)$$

Dari rumus (3) didapat nilai *BF* (Bayes Factor) sebagai berikut:

$$BF_{01} = 0,036584 \approx 0,0366$$

untuk Nilai  $BF_{01}$  diatas diperoleh sebesar 0,0366. Karena Gambar 2 menyajikan kriteria untuk  $BF_{10}$  dimana hubungan dengan  $BF_{01} = \frac{1}{BF_{10}}$  maka  $BF_{10} = \frac{1}{0,0366} = 27,33422 \approx 27,334$ .



**Gambar 2.** Diagram Klasifikasi Penafsiran  $BF_{10}$  (Faktor Bayes) (Quintana dan Williams, 2018).

#### 4. Kesimpulan

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

1. Berdasarkan hasil pengujian didapat hipotesis uji korelasi *Kendall Tau* ( $\tau$ ) menggunakan metode Bayesian dengan parametrik yoking adalah lebih mendukung untuk menerima  $H_1$  dan apabila  $n$  semakin banyak tidak akan merubah atau membingungkan untuk menerima  $H_0$  atau  $H_1$ .
2. Berdasarkan Metode Bayesian pada uji korelasi *Kendall Tau* ( $\tau$ ) didapatkan bahwa  $BF_{01} = 0,0366$  dan  $BF_{10} = 27,334$ . Maka  $H_0$  di tolak karena  $BF_{01} < 1$ . Artinya ada hubungan antara Luas Lahan Padi dengan Produktivitas Padi. Penolakan  $H_0$  (penerimaan  $H_1$ ) ini didukung oleh bukti yang kuat karena  $10 < BF_{10} < 30$ .

#### Acknowledge

Terimakasih kepada seluruh pihak yang membantu memberi saran maupun masukan sampai terlaksananya penelitian ini.

#### Daftar Pustaka

- [1] Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E. (2018). Bayesian Inference for Kendall's Rank Correlation Coefficient: Online Appendix. 1.
- [2] <https://www.bps.go.id/indicator/53/1498/1/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html> [internet, diakses pada tanggal 1 juni 2021].
- [3] Kendall, M., and Gibbons, J. D. (1990), Rank Correlation Methods, New York: Oxford University Press.
- [4] Quintana, D. S., & Williams, D. R. (2018). Bayesian Alternative for Common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry*.
- [5] Siegel, Sidney. (1994). Statistik Nonparametrik Untuk Ilmu-ilmu Sosial. PT. Gramedia Pustaka Utama: Jakarta, 273.
- [6] Shofwani Sheila Ghazia, Kudus Abdul. (2021). Penentuan Kriteria Pengunjung dalam Pemilihan Green Hotel di Kota Bandung Menggunakan Metode Discrete Choice Experiment dengan Desain Choice Sets Kombinatorial. *Jurnal Riset Statistika*, 1(1), 1-9.