

Pengaruh Distribusi Data Terhadap Hasil Uji Korelasi Studi Pada Uji Pearson Product Moment, Rank Spearman, dan Rank Kendall Tau

Ratna Yuniarti¹, Hartiani², Harizahayu³

¹Institut Teknologi Sosial dan Kesehatan Muhammadiyah Selong, ratna.matika@gmail.com

²Institut Teknologi Sosial dan Kesehatan Muhammadiyah Selong, hartiani78@gmail.com

³Politeknik Negeri Medan, harizahayu@polmed.ac.id

Abstract, Correlation analysis is used if you want to know whether there is a correlation between two phenomena. To determine the type of correlation to be used, researchers need to consider whether or not the assumption of normality and data characteristics are met. The purpose of this study was to compare the results of the analysis using several correlation tests with both parametric and non-parametric approaches. The method used is to provide data simulation with three types of data characteristics, namely normal, skewed and data containing outliers. The test used is the correlation in the parametric approach with the Pearson Product Moment Test, while for the non-parametric approach is the Spearman Rank and Kendall Tau Rank tests. Furthermore, a case study is given. The results show that in correlation testing without considering data distribution and data characteristics. Can produce inaccurate conclusions.

Keywords: *Data Distribution, Pearson Product Moment Correlation Test, Spearman Rank, Kendall Tau Rank*

Abstrak. Analisis korelasi digunakan jika ingin mengetahui ada tidaknya korelasi antara dua fenomena. Untuk menentukan jenis korelasi yang akan digunakan peneliti perlu mempertimbangkan terpenuhi atau tidak asumsi normalitas dan karakteristik data. Tujuan penelitian ini adalah untuk membandingkan hasil analisis dengan menggunakan beberapa uji korelasi baik dengan pendekatan parametric dan non parametric. Metode yang digunakan dengan memberikan simulasi data dengan tiga jenis karakteristik data yaitu normal, skewed dan data yang terdapat outlier. Adapun uji yang digunakan adalah korelasi pada pendekatan parametric dengan Uji Pearson Product Moment, sedangkan untuk pendekatan nonparametric adalah uji Rank Spearman dan Rank Kendall Tau. Selanjutnya diberikan studi kasus. Hasil menunjukkan bahwa dalam pengujian korelasi tanpa memperhatikan distribusi data dan karakteristik data. Dapat menghasilkan kesimpulan yang tidak akurat.

Kata Kunci: *Distribusi Data, Uji Korelasi Pearson Product Moment, Rank Spearman, Rank Kendall Tau*

1 Pendahuluan

Analisis hubungan antar variabel merupakan bagian penting dalam penelitian kuantitatif, terutama ketika peneliti ingin mengetahui ada tidaknya asosiasi atau korelasi di antara dua fenomena [1]. Salah satu pendekatan yang paling umum digunakan dalam analisis hubungan adalah uji korelasi Pearson, yang mengukur kekuatan dan arah hubungan linear antar dua variabel numerik

[2]. Koefisien korelasi Pearson digunakan untuk menilai kekuatan hubungan linear antara dua variable pada pendekatan statistic parametrik [3]. Metode ini diaplikasikan ketika variabel-variabel yang dianalisis diukur dalam skala interval atau rasio. Oleh karena itu, pengujian signifikansi terhadap koefisien korelasi Pearson mengharuskan data sampel berdistribusi normal, karena statistik uji yang digunakan mengikuti distribusi t.

Jika asumsi normalitas tidak terpenuhi atau skala pengukuran tidak termasuk dalam interval maupun rasio, maka pendekatan alternatif yang tepat adalah menggunakan metode korelasi dari statistika nonparametrik, seperti Spearman atau Kendall. Sejumlah studi telah menunjukkan pentingnya pemilihan jenis korelasi berdasarkan distribusi data. Schober menegaskan bahwa penggunaan koefisien korelasi Pearson pada data yang tidak memenuhi asumsi normalitas dapat menghasilkan estimasi hubungan yang bias [4].

Statistik nonparametrik digunakan dalam analisis data ketika asumsi distribusi normal tidak terpenuhi[5]. Dalam konteks ini, jenis skala pengukuran yang umum digunakan adalah skala ordinal dan nominal. Skala ordinal memberikan informasi mengenai urutan atau peringkat, namun tidak memberikan keterangan mengenai jarak antar nilai. Untuk menganalisis data dalam skala ordinal, dua koefisien korelasi yang sering digunakan adalah Rank Spearman dan Rank Kendall. Kedua metode ini memungkinkan pengukuran asosiasi antara dua variabel yang dinyatakan dalam bentuk peringkat [6]. Koefisien korelasi Rank Spearman dan Rank Kendall memiliki keunggulan dalam fleksibilitas penggunaannya, karena tidak mensyaratkan asumsi normalitas data. Hal ini menjadikan keduanya sangat sesuai untuk data ordinal, terutama ketika terdapat skor yang identik. Namun demikian, keberadaan banyak skor yang sama dapat menurunkan efektivitas pengukuran asosiasi yang dihasilkan oleh kedua metode tersebut [7]. Oleh karena itu, pemilihan metode korelasi yang tepat perlu mempertimbangkan karakteristik data serta tujuan analisis yang ingin [8].

Permasalahan muncul ketika banyak peneliti, khususnya mahasiswa, secara langsung menggunakan uji Pearson tanpa terlebih dahulu melakukan eksplorasi terhadap karakteristik data yang dimiliki. Praktik ini berisiko menghasilkan kesimpulan yang tidak akurat. Perbedaan hasil uji Pearson dan Spearman sering kali tidak diperhatikan, padahal keduanya dapat menghasilkan interpretasi korelasi yang berbeda terhadap data yang sama apabila asumsi normalitas tidak diperhatikan[9].

Penelitian yang berkaitan dengan pelayanan public juga membutuhkan uji korelasi misalnya pada pelayanan kesehatan tentang hubungan kualitas pelayanan posyandu dengan kepuasan ibu hamil, hubungan antara kelengkapan dokumen kependudukan dengan penetapan sebagai penerima bantuan social dan lain-lain. Sejauh ini, penelitian yang membahas secara komparatif penggunaan uji korelasi parametrik dan non-parametrik terhadap masih terbatas, terutama dalam konteks data pendidikan di tingkat perguruan tinggi. Padahal, pemahaman terhadap penggunaan metode korelasi yang tepat sangat penting dalam memastikan validitas hasil penelitian [10]

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk membandingkan hasil uji korelasi Pearson, Spearman dan kendall, guna memberikan gambaran yang lebih komprehensif mengenai kelebihan dan keterbatasan masing-masing metode dalam konteks distribusi berbeda.

2 Metode

Penelitian ini merupakan literatur review pada uji korelasi pada pendekatan parametric dan non parametric. Kemudian diberikan contoh simulasi data untuk membandingkan hasil uji pada masing-masing karakteristik data. Selain itu diberikan simulasi data yang kontekstual penerapan uji korelasi untuk mengetahui hubungan antar variabel.

Adapun langkah-langkah dalam penelitian ini adalah sebagai berikut:

- 1) Studi literatur tentang beberapa uji korelasi pada pendekatan parametrik dan non parametric. Adapun uji korelasi dengan pendekatan parametrik pada penelitian ini adalah Uji Pearson Product Moment. Teknik korelasi yang digunakan untuk mencari hubungan variabel bebas (X) dengan variabel terikat (Y) dan data berbentuk interval dan rasio yaitu dengan Uji Pearson Product Moment [12]. Adapun rumus yang dikemukakan adalah

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{\{n.\sum X^2 - (\sum X)^2\}\{n.\sum Y^2 - (\sum Y)^2\}}} \quad (1)$$

Sementara untuk pendekatan nonparametrik adalah korelasi Spearman Rank, jenis data yang dikorelasikan adalah data ordinal atau berjenjang atau rangking, serta data dari kedua variable tidak harus membentuk distribusi normal[13]. Adapun rumus Spearman Rank adalah

$$\rho = 1 - \frac{6 \sum b_i^2}{n(n^2 - 1)} \quad (2)$$

Dimana:

ρ : koefisien korelasi spearman Rank.

Korelasi Kendal Tau (τ) digunakan untuk mencari hubungan dan menguji hipotesis antara dua variabel atau lebih, bila datanya berbentuk ordinal atau rangking[13].

$$\tau = \frac{\sum A - \sum B}{N(N-1)/2} \quad (3)$$

Dimana :

τ = koefisien korelasi Kendal Tau yang besarnya ($-1 < \tau < 1$)

A= Jumlah rangking atas

B= Jumlah rangking bawah

N= Jumlah anggota sampel

- 2) Kemudian diberikan simulasi data dengan karakteristik data diasumsikan berdistribusi normal, data terdapat outlier, dan data skewed. Dilanjutkan dengan melakukan uji korelasi dengan masing-masing uji yaitu uji Pearson Product Moment, uji spearman Rank, uji Kendal Tau.
- 3) Membandingkan hasil uji dan melakukan interpretasi
- 4) Diberikan simulasi data kontekstual pada penerapan uji Pearson Product Moment, uji spearman Rank, uji Kendal Tau. Membandingkan hasil uji dan melakukan interpretasi

3 Hasil dan Pembahasan

Pada penelitian ini akan diberikan simulasi untuk tiga jenis data dan membanding hasilnya. Data berdistribusi normal merupakan salah satu syarat untuk melakukan uji korelasi dengan pendekatan parametric menggunakan rumus dengan Uji Pearson Product Moment, sedangkan untuk uji Korelasi Spearman Rank dan Korelasi Kendal Tau (τ) data dari kedua variable tidak harus membentuk distribusi normal.

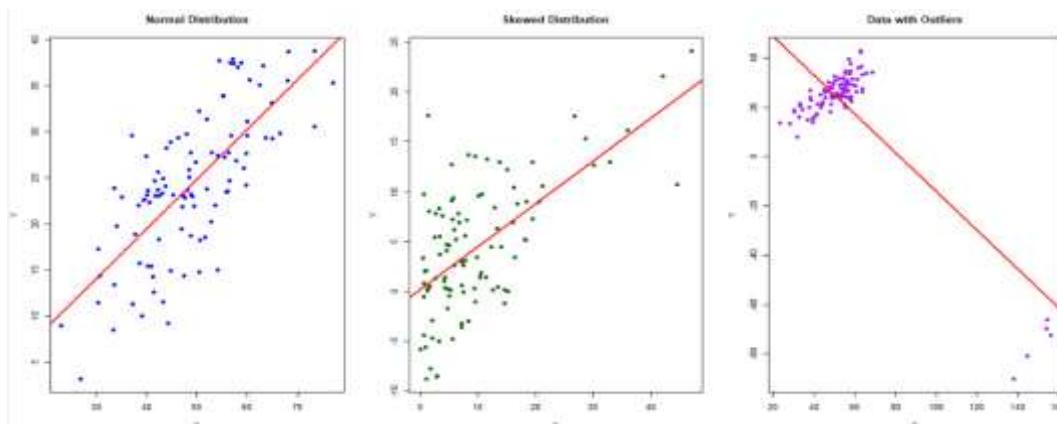
Berikut diberikan simulasi data di R dengan hasil sebagai berikut:

- Pertama kita mengaktifkan package library(tidyverse), library(ggpubr), library(writexl).
- Dibangkitkan 100 data untuk X dan Y berdistribusi normal dengan mean X = 50 dan standar deviasi X = 10, sedangkan untuk data Y dengan mean = 0 dan standar deviasi = 5.

```
x_normal <- rnorm(n, mean = 50, sd = 10)
y_normal <- 2 * x_normal + rnorm(n, 0, 10)
```
- Dibangkitkan data Skewed distribution (exponential)

```
x_skewed <- rexp(n, rate = 0.5)
y_skewed <- 3 * x_skewed + rnorm(n, 0, 1)
```
- Dibangkitkan data yang mengandung outlier

```
x_outlier <- rnorm(n, 50, 10)
x_outlier <- x_outlier[seq(1, n, 10)] + 50
y_outlier <- 2 * x_outlier + rnorm(n, 0, 10)
```
- Visualisasi scatterplot berdasarkan tiga jenis data yang dibangkitkan sebelumnya sebagai berikut

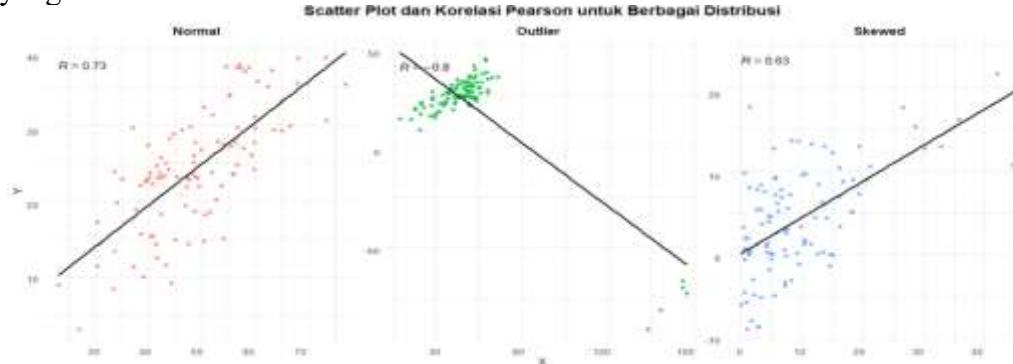


Gambar 1. Scatterplot Data

Gambar 1 di atas adalah visualisasi scatterplot berwarna biru untuk data berdistribusi normal, scatterplot berwarna hijau untuk data berdistribusi skewed, dan, scatterplot berwarna ungu dengan data mengandung outlier.

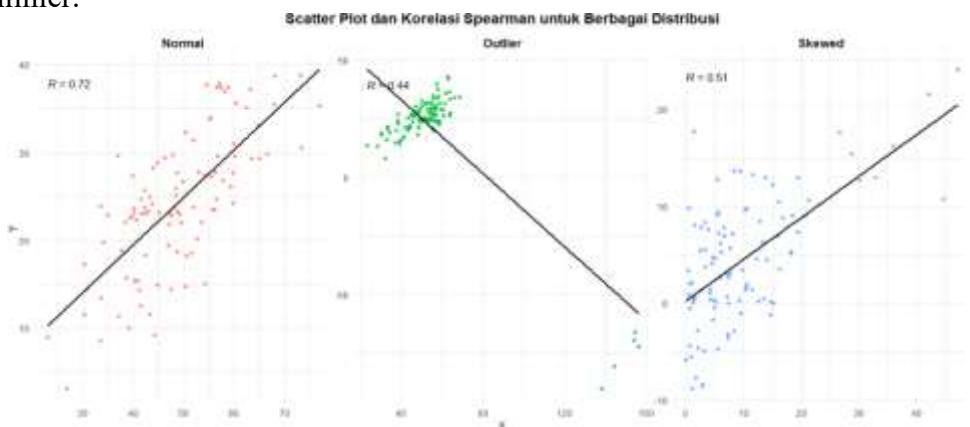
Perbandingan hasil analisis menggunakan tiga jenis korelasi untuk beberapa jenis data, pada Gambar 2 (scatterplot berwarna orange) dibawah ini menunjukkan bahwa pada data yang memenuhi asumsi normal didapatkan bahwa Uji Pearson Product Moment dengan koefisien korelasi $r = 0,73$ adalah hasil terbaik. Karena data yang linier dan normal cocok untuk dianalisis dengan Uji

Pearson Product Moment. Sementara pada jenis data Skewed (scatterplot berwarna biru) hasil koefisien korelasi $r = 0,63$ masih kuat. Sedangkan untuk jenis data yang mengandung outlier (scatterplot berwarna hijau) menghasilkan koefisien korelasi $r = -0,8$ artinya bahwa uji korelasi Pearson Product Moment tidak kuat terhadap outlier. Maka tidak disarankan menggunakan uji ini pada data yang terdeteksi outlier.



Gambar 2. Koefisien Korelasi Uji Pearson

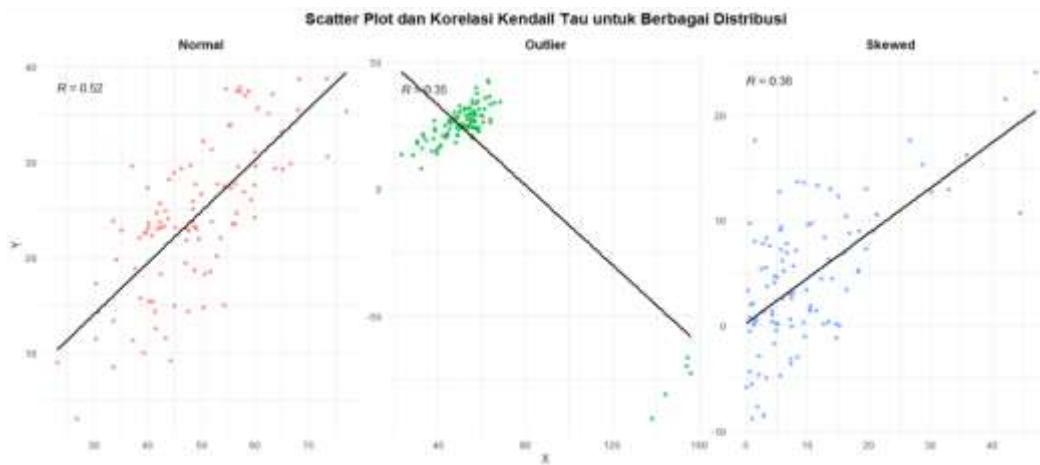
Sementara untuk Gambar 3 menunjukkan hasil analisis korelasi menggunakan Uji Rank Spearman menunjukkan hasil bahwa pada data berdistribusi normal (scatterplot berwarna orange) didapat bahwa koefisien korelasi $r = 0,72$ adalah hasil korelasi yang kuat. Uji ini baik untuk mengukur hubungan yang monotonik (nonlinier) dan tidak akan seefektif uji Pearson Product Moment pada data linier. Namun jika dibandingkan pada data yang mengandung outlier (scatterplot berwarna biru) dan berdistribusi skewed (scatterplot berwarna hijau) hasil korelasi dengan uji Rank Spearman menunjukkan hasil yang kuat. Meskipun data tidak berdistribusi normal. Sehingga uji ini disarankan untuk data yang tidak memenuhi normalitas dan data yang nonlinier.



Gambar 3. Koefisien Korelasi Uji Rank Spearman

Terakhir pada Gambar 4 adalah analisis menggunakan uji Rank Kendall Tau, dimana pada data yang berdistribusi normal (scatterplot berwarna orange) tetap dapat menghasilkan analisis yang baik. Begitu juga pada data Skewed (scatterplot berwarna biru) dan data yang mengandung outlier (scatterplot berwarna hijau), uji ini tetap kuat. Sehingga uji ini juga disarankan pada data yang tidak memenuhi

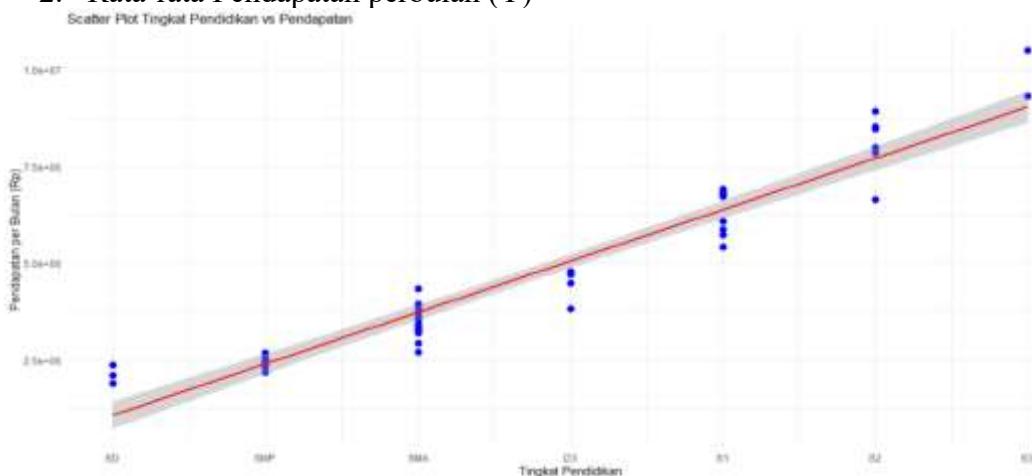
asumsi normalitas dan data dengan ukuran kecil < 30 . Nilai korelasi lebih kecil dibanding Spearman (dalam skala), sehingga bisa terlihat "kurang kuat".



Gambar 4. Koefisien Korelasi Uji Rank Kendall Tau

Pada studi kasus ini dibangkitkan 50 data simulasi tentang hubungan tingkat pendidikan dengan rata-rata pendapatan perbulan. tulisan akan diberikan studi kasus menggunakan data dua variable yaitu:

1. Tingkat Pendidikan (X),
2. Rata-rata Pendapatan perbulan (Y)



Gambar 5. Scatterplot Data

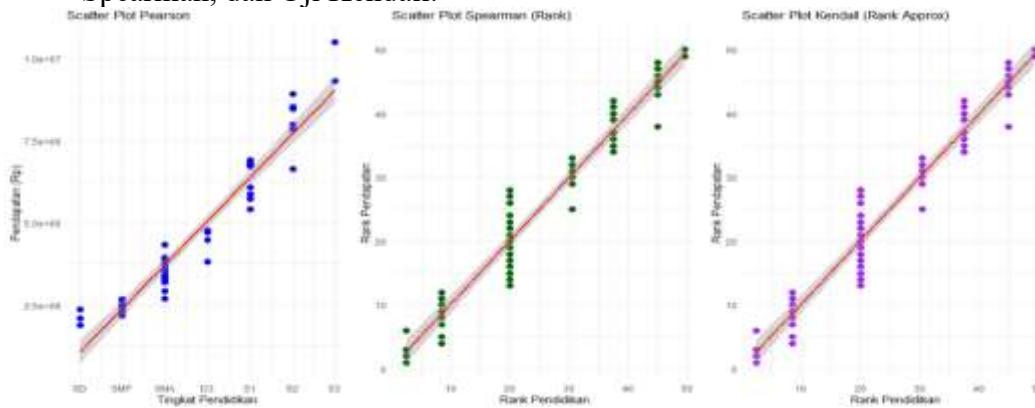
Gambar 5 di atas menunjukkan visualisasi data tingkat pendidikan dan rata-rata pendapatan perbulan. Pada kasus ini jenis data di variabel X adalah data ordinal karena tingkat pendidikan SD, SMP, SMA, S1, S2, S3. Sedangkan variable Y adalah data rasio.

Untuk mengetahui data memenuhi asumsi normalitas data diuji normalitasnya dengan Kolmogorov smirnov. Adapun hasil uji menunjukkan bahwa data berdistribusi normal.

Asymptotic one-sample Kolmogorov-Smirnov test

```
data: data_simulasi_50$Pendidikan_Num
D = 0.20483, p-value = 0.03012
alternative hypothesis: two-sided
```

3. Kemudian dilakukan uji korelasi dengan Uji Pearson Product Moment, Uji Spearman, dan Uji Kendall.



Gambar 6. Koefisien Korelasi Uji Pearson, Rank Spearman, Rank Kendall Tau

Pada Gambar 6 di atas menunjukkan perbandingan scatterplot hasil uji korelasi dengan Uji Pearson, Uji Rank Spearman, dan Uji Rank Kendall Tau. Adapun perbandingan koefisien korelasi pada Tabel 1 dibawah ini dengan P-Value masing-masing uji menunjukkan hubungan yang sangat kuat pada semua jenis uji yang digunakan. Namun, pemilihan uji terbaik disini tidak dilihat dari besarnya koefisien korelasi tetapi syarat karakteristik data pada masing-masing uji harus terpenuhi.

Misalnya Pada tabel 1 Uji pearson menunjukkan hasil korelasi 0.9605500 sangat kuat. Namun data studi kasus tersebut berjenis ordinal dan rasio tidak sesuai dengan

Tabel 1. Perbandingan Koefisien korelasi

Metode Uji	Koefisien	P value
Pearson	0.9605500	2.500418e-28
Spearman	0.9726656	4.294623e-32
Kendall Tau	0.8968504	9.611790e-18

Sumber: data diolah

Pada pembahasan sebelumnya telah dijabarkan bahwa untuk pemilihan uji perlu memperhatikan jenis data. Untuk data simulasi ini berjenis ordinal dan rasio sehingga disarankan untuk menggunakan uji spearman dan kendall tau. Sedangkan Uji Pearson tidak disarankan untuk data ordinal. Jenis data ini berpotensi memengaruhi hasil uji korelasi, terutama jika metode yang digunakan tidak sesuai dengan karakteristik distribusi data tersebut. Sayangnya, banyak peneliti dan praktisi statistik yang menggunakan metode korelasi secara default tanpa memperhatikan jenis data dan bentuk distribusi data yang dianalisis.

4 Kesimpulan

Berdasarkan hasil penelitian tersebut ada beberapa hal yang dapat disimpulkan dalam pemilihan Uji Korelasi dalam menganalisis hubungan :

1. Penggunaan pendekatan parametrik dengan uji Pearson Product Momen yaitu pada data yang memenuhi asumsi normalitas dan jenis data pada skala interval dan rasio.
2. Penggunaan pendekatan nonparametric dengan uji Rank Spearman dan Rank Kendall Tau bebas asumsi normalitas dan jenis data pada skala ordinal dan nominal
3. Penggunaan uji yang tidak sesuai dengan karakteristik data dapat menghasilkan kesimpulan yang tidak akurat.

5 Daftar Pustaka

- [1] L. C. Guglielmetti, F. Faber-Castell, L. Fink, and R. N. Vuille-dit-Bille, “Statistics decrypted—a comprehensive review and smartphone-assisted five-step approach for good statistical practice,” *Langenbecks Arch. Surg.*, vol. 407, no. 2, pp. 529–540, Mar. 2022, doi: 10.1007/s00423-021-02360-0.
- [2] R. E. Walpole, *Introduction to Statistics*. New York: Mc Millan, 1962.
- [3] M. I. Hasan, *statistik 2*, 2nd ed. JAKARTA: PT BU,MI AKSARA, 2003.
- [4] P. Schober, C. Boer, and L. A. Schwarte, “Correlation Coefficients: Appropriate Use and Interpretation,” *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.
- [5] H.-Y. Kim, “Statistical notes for clinical researchers: Nonparametric statistical methods: 2. Nonparametric methods for comparing three or more groups and repeated measures,” *Restor. Dent. Endod.*, vol. 39, no. 4, p. 329, 2014, doi: 10.5395/rde.2014.39.4.329.
- [6] G. Norman, “Likert scales, levels of measurement and the ‘laws’ of statistics,” *Adv. Health Sci. Educ.*, vol. 15, no. 5, pp. 625–632, Dec. 2010, doi: 10.1007/s10459-010-9222-y.
- [7] N. Minois *et al.*, “Using Poisson–gamma model to evaluate the duration of recruitment process when historical trials are available,” *Stat. Med.*, vol. 36, no. 23, pp. 3605–3620, Oct. 2017, doi: 10.1002/sim.7365.
- [8] J. Hauke and T. Kossowski, “Comparison of Values of Pearson’s and Spearman’s Correlation Coefficients on the Same Sets of Data,” *QUAGEO*, vol. 30, no. 2, pp. 87–93, Jun. 2011, doi: 10.2478/v10117-011-0021-1.
- [9] M. M. Mukaka, “Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research”.
- [10] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2. ed., Reprint. New York, NY: Psychology Press, 2009.
- [11] R. Ridwan, *DasarDasar Statistika*. in 15. Bandung: ALFABETA, 2018.
- [12] S. Sugiyono, *Statistika untuk Penelitian*. Bandung: ALFABETA, 2016.
- [13] A. Supangat, *Statistika dalam Kajian Deskriptif, Inferensi, dan Nonparametrik*. JAKARTA: Kencana Prenada Media Group, 2010.