Kajian Hubungan Koefisien Korelasi Pearson (ρ), Spearman-*Rho* (r), Kendall-*Tau* (τ), Gamma (G) , dan Somers (d_{yx})

Resi Vusvitasari[1], Sigit Nugroho[2], dan Syahrul Akbar[2]
[1]Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu
[2]Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

Abstrak

Penelitian ini bertujuan mengkaji tentang hubungan koefisien korelasi Pearson (r), Spearman-*rho* (ρ), Kendall-*tau* (τ), Gamma (G) dan Somers (d_{yx}) serta mempelajari penggunaan dari masing-masing koefisien korelasi untuk skala ordinal. Metode yang digunakan dalam penulisan ini adalah studi literatur dan data yang digunakan adalah data simulasi yang dibuat menggunakan program komputer Microsoft EXCEL. Data simulasi terdiri dari dua, yaitu data tidak normal (seragam) dan data normal yang dibangkitkan dari data seragam. Hasil penelitian menunjukkan bahwa untuk data seragam, koefisien korelasi yang diberikan oleh koefisien korelasi Spearman-*rho* (ρ) dan Kendall-*tau* (τ) lebih besar dibandingkan dengan Koefisien korelasi Pearson (r). Dan untuk data normal, koefisien korelasi yang diberikan oleh koefisien korelasi Pearson (r) lebih besar dibandingkan koefisien korelasi Spearman-*rho* (ρ) dan Kendall-*tau* (τ). Ini membuktikan bahwa koefisien korelasi Pearson sesuai digunakan untuk data yang berdistribusi normal, sedangkan koefisien korelasi Spearman-*rho* (ρ) dan Kendall-*tau* (τ) digunakan untuk data yang tidak normal. Penelitian juga menunjukan adanya hubungan yang linier antara koefisien korelasi Gamma dan Somers.

Kata Kunci : Korelasi, Ukuran Asosiasi , Tabel Kontingensi, Statistika Nonparametrik.

Pendahuluan

Kekuatan dan sifat ketergantungan antar variabel merupakan masalah sentral yang ingin diketahui pada suatu penelitian. Kadang peneliti ingin mengetahui apakah terdapat hubungan antara dua variabel dan seberapa kuat hubungan kedua variabel tersebut. Uji statistika yang mengukur keeratan hubungan antara dua variabel ini disebut analisis korelasi (*correlation*). Ukuran untuk menentukan kuatnya atau derajat keeratan hubungan antar dua variabel dinamakan koefisien korelasi (*the correlation coefficient*).

Dalam statistika parametrik, koefisien korelasi antara dua variabel (*bivariat*e) yang biasa digunakan adalah koefisien korelasi momen hasil kali Pearson, yang dinotasikan dengan r. Dimana skala data pengamatan serendah-rendahnya adalah interval atau rasio. Jika data pengamatan adalah berupa skala ordinal, dalam hal ini untuk uji korelasi statistika nonparametrik, maka ada beberapa koefisien korelasi yang dapat digunakan, yaitu koefisien korelasi peringkat Spearman-*rho* (ρ), Kendall-*tau* (τ), Gamma (G), dan Somers (d_{yx}). Dari keempat koefisien korelasi ini banyak peneliti yang mungkin ingin tahu kapan harus menggunakan koefisien korelasi peringkat Spearman-*rho* (ρ), Kendall-*tau* (τ), Gamma (G), dan Somers (d_{yx}) dalam suatu penelitian. Untuk mengetahui penggunaan dari masing-

masing koefisien korelasi untuk data berskala ordinal di atas, maka perlu dipelajari tentang sifat-sifat dari keempat koefisien korelasi ini. Sehingga nantinya para peneliti dapat mengetahui dan benar-benar tepat dalam memilih koefisien korelasi mana yang akan digunakan dalam penelitian.

Tujuan dari penulisan skripsi ini adalah untuk mengkaji tentang hubungan koefisien korelasi momen hasil kali Pearson (r), Spearman-*rho* (ρ), Kendall-*tau* (τ), Gamma (G) dan Somers (d_{yx}) serta mempelajari penggunaan dari koefisien korelasi peringkat Spearman-*rho* (ρ), Kendall-*tau* (τ), koefisien korelasi Gamma (G), dan Somers (d_{yx}) dalam suatu penelitian. Kemudian akan diberikan contoh kasus yang terkait dengan pembahasan.

Pengantar Teori Statistika Nonparametrik

Peneliti sering kali mengalami kesulitan untuk memperoleh data kontinu pada penelitian yang mengikuti distribusi normal. Hal ini salah satunya karena jumlah data sampel yang didapat tidak cukup banyak sehingga tidak memenuhi distribusi normal. Selain itu, banyak pengukuran data dilakukan secara kualitatif dan data dalam penelitian yang diperoleh sering berupa kategori yang hanya dapat dihitung frekuensinya atau berupa data yang hanya dapat dibedakan berdasarkan tingkatan atau rankingnya.

Menghadapi kasus data kategorikal (nominal) atau data ordinal seperti itu, jelas peneliti tidak mungkin mempergunakan metode statistika parametrik. Karena apabila asumsi-asumsi tidak dapat terpenuhi, akan menghasilkan suatu kesimpulan yang tidak valid. Kesulitan-kesulitan dalam data tetap harus diatasi supaya analisis data bisa dilakukan dan menghasilkan suatu kesimpulan yang valid. Sebagai gantinya diciptakan oleh pakar metode statistika alternatif yang sesuai yaitu metode statistika nonparametrik sebagai pelengkap statistika parametrik.

Metode statistika nonparametrik merupakan suatu metode analisis data tanpa memperhatikan bentuk distribusinya sehingga statistika ini sering juga disebut metode bebas sebaran (*distribution free methods*), karena model uji statistikanya tidak menetapkan syarat-syarat tertentu tentang bentuk distribusi parameter populasinya. Artinya bahwa metode statistika nonparametrik ini tidak menetapkan syarat bahwa observasi-observasinya harus ditarik dari populasi yang berdistribusi normal dan tidak menetapkan syarat homoskedastisitas (*homoscedasticity*).

Selain tidak menetapkan syarat mengenai distribusi populasinya, statistika nonparametrik juga tidak menetapkan syarat-syarat mengenai parameter-parameter populasi yang merupakan induk sampel penelitiannya.

Suatu metode statistika dapat dikatakan nonparametrik apabila memenuhi paling sedikit satu kriteria dibawah ini :

1. Metode ini digunakan untuk data pengamatan dengan skala nominal
2. Metode ini digunakan untuk data pengamatan dengan skala ordinal
3. Metode ini digunakan untuk data pengamatan dengan skala interval atau rasio, dimana distribusi populasinya tidak diketahui.

Pemilihan macam uji statistika nonparametrik mana yang paling sesuai didasarkan pada beberapa kriteria. Pertama didasarkan pada skala pengukuran variabel penelitiannya, baik itu skala nominal, ordinal, atau skala interval/rasio. Pada dasarnya uji yang sesuai bagi variabel dengan skala nominal atau ordinal adalah uji statistika nonparametrik, namun terdapat juga uji statistika nonparametrik yang berlaku pada variabel yang berskala interval. Kedua, pemilihan uji statistika nonparametrik didasarkan pada banyaknya sampel penelitian, apakah berupa sampel tunggal, dua sampel berpasangan atau dua sampel independen, atau sampelnya lebih dari dua buah yang berpasangan atau yang independen. Ketiga, pemilihan uji nonparametrik didasarkan pada jenis penelitiannya, apakah berupa uji kesesuaian (*goodness-of-fit*), uji banding, uji independensi, atau apakah berupa uji keterikatan (korelasi) antara dua kumpulan atribut atau dua kumpulan skor. Contoh uji statistika nonparametrik diantaranya adalah uji binomial, uji median, uji tanda, uji *Kolmogorov-Smirnov*, uji korelasi peringkat, uji *Wilcoxon-Mann-Whitney*, uji *Friedman* dan lain sebagainya.

Analisis Korelasi Statistika Nonparametrik

Analisis korelasi merupakan uji statistika yang mengukur keeratan hubungan antara dua variabel. Keeratan hubungan antara dua variabel dapat diukur kekuatannya. Indeks yang mengukur keeratan hubungan dua variabel disebut koefisien korelasi. Nilai koefisien korelasi paling (*r*) dapat dinyatakan sebagai berikut :

$$-1 \leq r \leq 1 \qquad [1]$$

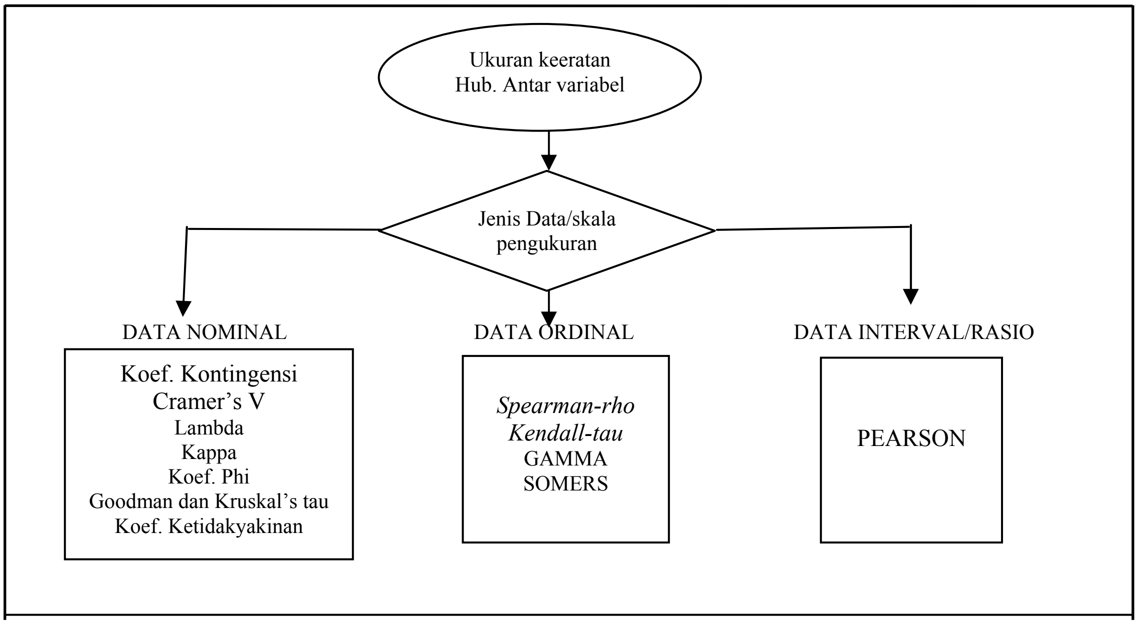
r = 1, hubungan X dan Y sempurna dan positif (mendekati 1, yaitu hubungan sangat kuat dan positif).
r = -1, hubungan X dan Y sempurna dan negatif (mendekati -1, yaitu hubungan sangat kuat dan negatif).
r = 0, hubungan X dan Y lemah sekali atau tidak ada hubungan.

Sama halnya dengan statistika parametrik, analisis korelasi pada statistika nonparametrik juga mempelajari apakah ada hubungan antar dua variabel. Hanya pada korelasi nonparametrik, data atau variabel yang akan diuji dan diukur korelasinya adalah data nominal atau ordinal. Sebagai contoh, apakah motivasi seseorang mempengaruhi kepuasan bekerja orang tersebut. Di sini variabel motivasi ataupun kepuasan kerja adalah data ordinal, karena tidak mungkin motivasi dan kepuasan diukur seperti pengukuran tinggi badan atau berat badan yang secara riil dapat dilihat.

Dalam statistika parametrik, Koefisien korelasi yang dikenal luas dan paling sering digunakan adalah koefisien korelasi momen hasil kali Pearson yang dinotasikan dengan *r*, dimana rumus *r* adalah sebagai berikut:

$$r = \frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\left[\left(\sum\left(X-\bar{X}\right)^2\right)\left(\sum\left(Y-\bar{Y}\right)^2\right)\right]^{\frac{1}{2}}} \qquad [2]$$

dengan X dan Y adalah variabel-variabel yang diamati dan banyaknya sampel pengamatan. Perhitungan dalam teknik korelasi ini mensyaratkan bahwa populasi asal sampel mempunyai dua varian (bivariat) dan berdistribusi normal. Selain itu teknik korelasi ini dalam aplikasinya digunakan untuk mengukur korelasi data dengan skala pengukuran interval atau rasio. Sedangkan dalam statistika nonparametrik, untuk kasus pengukuran analisis korelasi antara dua kumpulan skor dapat digolongkan berdasarkan pada skala pengukuran variabel penelitiannya (skala nominal, ordinal, atau interval-rasio). Adapun penggolongan uji korelasi itu adalah sebagai berikut:



Gambar 1. *Pembagian Korelasi*

Dari bagan di atas, dapat dilihat bahwa koefisien korelasi nonparametrik untuk jenis data dengan skala pengukuran ordinal, terdapat empat macam koefisien korelasi yang dapat digunakan, yaitu koefisien korelasi peringkat *Spearman-rho* yang dinotasikan dengan ρ, koefisien korelasi *Kendall-tau* yang dinotasikan dengan τ, koefisien korelasi Somers yang dinotasikan dengan d_{yx}, dan koefisien korelasi Gamma yang dinotasikan dengan G. Keempat koefisien korelasi ini didasarkan pada ranking. Hanya saja antara koefisien korelasi peringkat *Spearman-rho* dengan koefisien korelasi *Kendall-tau*, Gamma, dan Somers ada sedikit perbedaan. Dimana untuk koefisien korelasi peringkat *Spearman-rho* memperhitungkan besarnya perbedaan rank pasangan nilai pengamatan (X_i, Y_i), sedangkan untuk koefisien korelasi *Kendall-tau*, koefisien korelasi Somers, dan koefisien korelasi Gamma hanya memperhitungkan kekuatan asosiasi berdasarkan arah pasangan nilai pengamatan (X_i, Y_i) dan tidak memperhitungkan besarnya perbedaan pasangan nilai pengamatan (X_i, Y_i) seperti pada koefisien korelasi peringkat *Spearman-rho* atau dengan kata lain koefisien korelasi *Kendall-tau*, koefisien korelasi Somers, dan koefisien korelasi Gamma didasarkan pada konsep pasangan *concordant* (*C*) dan *discordant* (*D*).

Untuk mengetahui keeratan hubungan antara dua variabel, tidak hanya dilihat dari besarnya nilai koefisien korelasi antara dua variabel yang diberikan. Akan tetapi perlu juga dilakukan uji signifikansi, dalam hal ini pengujian hipotesis dari kedua variabel tersebut.

Hipotesis-hipotesis

Uji dua arah

H_0 : X_i dan Y_i saling bebas
H_1 : X_i dan Y_i dependen (hubungan positif atau negatif) .

Uji satu arah (Positif)
H_0 : X_i dan Y_i saling bebas
H_1 : X_i dan Y_i dependen (hubungan positif) .

Uji satu arah (Negatif)
H_0 : X_i dan Y_i saling bebas
H_1 : X_i dan Y_i dependen (hubungan negatif) .

Koefisien Korelasi Nonparametrik Untuk Skala Ordinal

Koefisien korelasi merupakan ukuran yang menyatakan keeratan hubungan antara dua variabel. Koefisien korelasi bivariat yang paling lama dan banyak digunakan adalah korelasi yang dikembangkan oleh *Karl Pearson*. Perhitungan dalam korelasi ini didasarkan pada data sebenarnya (variabel asli). Secara statistik, koefisien korelasi momen hasil kali Pearson atau sering disingkat dengan koefisien korelasi Pearson yang dinotasikan dengan r dirumuskan sebagai berikut:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)\right]^{\frac{1}{2}}} \qquad [3]$$

Dalam aplikasinya koefisien korelasi ini digunakan untuk mengukur keeratan hubungan di antara hasil-hasil pengamatan dari populasi yang mempunyai dua varian (bivariat). Perhitungan dalam tekhnik korelasi ini mensyaratkan bahwa populasi asal sampel mempunyai dua varian dan berdistribusi normal. Selain itu teknik korelasi ini dalam aplikasinya digunakan untuk mengukur korelasi data interval atau rasio.

1. **Koefisien korelasi peringkat Spearman - *rho* (ρ)**

Ukuran korelasi nonparametrik yang analog dengan koefisien korelasi Pearson (r) adalah koefisien korelasi yang dikembangkan oleh *Charles Spearman* (1908) yaitu koefisien korelasi peringkat Spearman. Statistik ini kadang disebut dengan Spearman- *rho*, dan dinotasikan dengan ρ. Jika pada koefisien korelasi Pearson (r) digunakan untuk mengetahui korelasi data kuantitatif (skala interval dan rasio), maka pada koefisien korelasi peringkat Spearman-*rho* digunakan untuk pengukuran korelasi pada statistik nonparametrik (skala ordinal). Ini merupakan ukuran korelasi yang menuntut kedua variabel diukur sekurang-kurangnya dalam skala ordinal sehingga obyek-obyek penelitiannya dapat diranking dalam dua rangkaian berurut.

Misal data terdiri dari sampel acak bivariat berukuran n, yaitu $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$. Misalkan $R(X_i)$ adalah rank dari X_i dibandingkan dengan nilai X lainnya, untuk $i = 1,2,\ldots,n$. $R(X_i) = 1$ jika X_i adalah nilai X terkecil dari $X_1, X_2,\ldots, X_n$, $R(X_i) = 2$ jika X_i adalah nilai X terkecil kedua, dan seterusnya dengan rank n ditandai sebagai nilai X_i terbesar. Begitu juga untuk $R(Y_i)$. Jika di antara nilai X_i atau di antara nilai Y_i terdapat angka sama, maka masing-masing nilai yang sama diberi peringkat rata-rata dari posisi-posisi yang seharusnya.

Rumus koefisien korelasi peringkat Spearman- *rho* merupakan turunan rumus koefisien korelasi Pearson, yaitu

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2\right]^{\frac{1}{2}}} \qquad [4]$$

dimana untuk koefisien korelasi peringkat Spearman-*rho* (ρ), variabel asli diganti dengan rank-ranknya, maka X_i diganti dengan $R(X_i)$ dan Y_i diganti dengan $R(Y_i)$. Sehingga rumus koefisien korelasi peringkat Spearman-*rho* (ρ) adalah

$$\begin{aligned}\rho &= \frac{\sum_{i=1}^{n} \left[R(X_i) - \overline{R(X)}\right]\left[R(Y_i) - \overline{R(Y)}\right]}{\left[\sum_{i=1}^{n} \left(R(X_i) - \overline{R(X)}\right)^2 \sum_{i=1}^{n} \left(R(Y_i) - \overline{R(Y)}\right)^2\right]^{\frac{1}{2}}} \\ &= \frac{\sum_{i=1}^{n} \left[R(X_i) - \frac{n+1}{2}\right]\left[R(Y_i) - \frac{n+1}{2}\right]}{\left(\frac{n(n^2-1)}{12} \cdot \frac{n(n^2-1)}{12}\right)^{\frac{1}{2}}} \\ &= \frac{\sum_{i=1}^{n} \left[R(X_i) - \frac{n+1}{2}\right]\left[R(Y_i) - \frac{n+1}{2}\right]}{\frac{n(n^2-1)}{12}}\end{aligned} \qquad [5]$$

untuk mempermudah perhitungan, maka persamaan diatas dapat disederhanakan sebagai berikut

$$\rho = 1 - \frac{6\sum_{i=1}^{n} \left[R(X_i) - R(Y_i)\right]^2}{n(n^2-1)} = 1 - \frac{6T}{n(n^2-1)} \qquad [6]$$

dimana $\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \left[R(X_i) - R(Y_i)\right]^2$, yaitu jumlah kuadrat dari selisih-selisih antara rank-rank X_i *dan* Y_i untuk masing-masing pengamatan.

Langkah-langkah untuk menghitung koefisien korelasi Spearman-*rho* (ρ) adalah sebagai berikut :

- Berilah peringkat untuk masing-masing pengamatan X mulai dari 1 hingga n, juga untuk pengamatan Y beri peringkat mulai dari 1 sampai n.
- Tentukan harga $\sum_{i=1}^{n} d_i^2$, yaitu jumlah kuadrat dari selisih-selisih antara rank-rank X_i *dan* Y_i untuk masing-masing pengamatan.
- Gunakan persamaan [6] untuk menghitung ρ.

2. Koefisien korelasi Kendall - *Tau* (τ)

Koefisien korelasi yang kedua yang biasa digunakan untuk mengukur kekuatan korelasi untuk data penelitian dengan skala pengukuran ordinal adalah koefisien korelasi yang dikenalkan oleh *M.G. Kendall* (1938) yaitu koefisien korelasi Kendall-*tau* yang dinotasikan dengan τ. Koefisien korelasi ini memiliki sifat yang sama dengan koefisien korelasi peringkat Spearman-*rho*, tetapi berbeda dasar logikanya. Jika untuk koefisien korelasi peringkat Sperman-*rho* didasarkan pada peringkat (*rank*), dimana baik variabel X dan variabel Y masing-masing kita ranking. Sedangkan untuk koefisien korelasi Kendall-

tau salah satu variabelnya yang diberi peringkat (diurutkan), yaitu variabel X saja atau variabel Y saja dalam hal ini biasanya adalah variabel X. Sedangkan variabel Y akan dilihat apakah nilai variabel Y itu searah (konkordan) atau berlawanan arah (diskordan) dengan variabel X yang sudah diurutkan.

Jika ada data bivariat $(X_i, Y_i), i = 1,2,...,n$ dimana X dan Y sekurang-kurangnya berskala ordinal. Maka untuk setiap pasangan nilai observasi $(X_i, Y_i)\ dan\ (X_j, Y_j)\ untuk\ i \neq j$ dapat didefinisikan pasangan nilai sebagai berikut :

i. Pasangan $(X_i, Y_i)\ dan\ (X_j, Y_j)$ konkordan, jika $(X_i - X_j)(Y_i - Y_j) > 0$
   artinya adalah jika $X_i > X_j\ maka\ Y_i > Y_j$ atau jika $X_i < X_j\ maka\ Y_i < Y_j$ sehingga $(X - X)\ dan\ (Y - Y)$ memiliki tanda yang sama, yaitu sama-sama positif atau sama-sama negatif dengan hasil kali yang selalu positif.

ii. Pasangan $(X_i, Y_i)\ dan\ (X_j, Y_j)$ diskordan, jika $(X_i - X_j)(Y_i - Y_j) < 0$
   artinya adalah jika $X_i > X_j\ maka\ Y_i < Y_j$ atau jika $X_i < X_j\ maka\ Y_i > Y_j$ sehingga $(X - X)\ dan\ (Y - Y)$ memiliki tanda yang berlawanan dengan hasil kali yang selalu negatif.

Secara keseluruhan, untuk n pengamatan ada sebanyak $\binom{n}{2} = \frac{n(n-1)}{2}$ pasangan yang mungkin. Jika ada sebanyak C pasangan yang searah (konkordan) dan D pasangan yang berlawanan arah (diskordan), maka Kendall-*tau* dapat dihitung sebagai berikut:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)} \qquad [7]$$

Langkah-langkah untuk menghitung koefisien korelasi Kendall-*tau* (τ) adalah sebagai berikut :

- Susunlah pasangan-pasangan (X_i,Y_i) dalam sebuah kolom menurut besarnya nilai-nilai pengamatan X, dari nilai pengamatan X yang paling kecil. Disini dapat dikatakan bahwa nilai-nilai X berada dalam urutan yang wajar (*natural order*).
- Perbandingkan setiap nilai pengamatan Y satu demi satu dengan setiap nilai Y yang ada di sebelah bawahnya. Jika nilai Y yang di bawah lebih besar dari Y yang di atasnya, maka arah nilai pengamatannya sama (konkordan). Dan jika nilai Y yang di bawah lebih kecil dari Y yang di atasnya, maka arah nilai pengamatannya berlawanan (diskordan).
- Tetapkan C sebagai banyaknya pasangan konkordan dan D banyaknya pasangan diskordan.
- gunakan persamaan [7] untuk menghitung τ.

Koefisien korelasi Gamma (*G*)

Sebelumnya sudah dibahas dua koefisien korelasi untuk dua variabel dengan skala pengukuran ordinal, yaitu Spearman-*rho* dan Kendall-*tau*. Akan tetapi, jika data pasangan pengamatan banyak mengandung angka sama atau ada situasi dimana data pengamatan ditampilkan dalam bentuk tabel kontingensi, maka penggunaan koefisien korelasi Spearman-*rho* dan Kendall-*tau* akan kurang efektif. Dengan demikian untuk data pasangan pengamatan yang keduanya bertipe ordinal dan ditampilkan dalam bentuk tabel kontingensi, koefisien korelasi yang dapat digunakan adalah koefisien korelasi Gamma (*G*) dan koefisien korelasi Somers (d_{yx}).

Koefisien korelasi yang ketiga yang dapat digunakan untuk mengukur korelasi untuk data penelitian dengan skala pengukuran ordinal adalah koefisien korelasi Gamma, yang dinotasikan dengan *G*. Koefisien korelasi ini dikenalkan oleh *Goodman* dan *Kruskal* (1954). Koefisien korelasi ini memiliki dasar logika yang sama dengan koefisien korelasi Kendall-*tau*, yaitu didasarkan pada banyaknya pasangan konkordan (*C*) dan pasangan diskordan (*D*).

Misalkan ada dua pengamatan bivariat *X* dan *Y*, dimana keduanya merupakan variabel terurut. Pengamatan X_i terdiri dari $X_1, X_2, \ldots, X_k$, $i = 1,2,\ldots,k$ dimana $X_1 < X_2 < \cdots < X_k$. Begitu juga dengan pengamatan Y_j terdiri dari $Y_1, Y_2, \ldots, Y_r$, $j = 1,2,\ldots,r$ dimana $Y_1 < Y_2 < \cdots < Y_r$.

Untuk menghitung statistik *G* dari dua pasangan pengamatan untuk data ordinal, $X_1, X_2, \ldots, X_k$ dan $Y_1, Y_2, \ldots, Y_r$ yang disusun dalam tabel kontingensi seperti dibawah ini,

Tabel 1. Tabel Kontingensi Data Kategorik Peringkat.

	X_1	X_2	…	X_k	Total
Y_1	n_{11}	n_{12}	…	n_{1k}	R_1
Y_2	n_{21}	n_{22}	…	n_{2k}	R_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
Y_r	n_{r1}	n_{r2}	…	n_{rk}	R_r
Total	C_1	C_2	…	C_k	N

maka statistik *G* didefinisikan sebagai berikut,

$$G = \frac{C-D}{C+D} \qquad [8]$$

dimana $C = \sum_{i.j} n_{ij} N_{ij}^{+}$ dengan $i = 1,2,\ldots,r-1$ dan $j = 1,2,\ldots,k-1$ [9]

$D = \sum_{i.j} n_{ij} N_{ij}^{-}$ dengan $i = 1,2,\ldots,r-1$ dan $j = 1,2,\ldots,k$ [10]

N_{ij}^{+} *dan* N_{ij}^{-} didefinisikan sebagai berikut:

$$N_{ij}^{+} = \sum_{p=i+1}^{j} \sum_{q=j+1}^{k} n_{pq} \quad \text{dan} \quad N_{ij}^{-} = \sum_{p=i+1}^{j} \sum_{q=1}^{k-1} n_{pq}$$

Langkah-langkah menghitung koefisien korelasi Gamma (*G*) adalah sebagai berikut:

- Hitung banyaknya pasangan konkordan dan diskordan dari tabel kontingensi yang diberikan. Dimana untuk menghitung banyaknya pasangan konkordan dapat digunakan persamaan [9] dan untuk menghitung banyaknya pasangan diskordan gunakan persamaan [10].
- Setelah banyaknya pasangan konkondan (*C*) dan diskordan (*D*) sudah diketahui, substitusikan *C* dan *D* ke persamaan [8].

3. **Koefisien korelasi Somers** (d_{yx})

Koefisien korelasi yang dapat digunakan untuk mengukur kekuatan korelasi untuk data penelitian dimana kedua variabel berskala ordinal dan data ditampilkan dalam bentuk tabel kontingensi selain koefisien korelasi Gamma (*G*) adalah koefisien korelasi Somers, yang dinotasikan dengan (d_{yx}). Koefisien korelasi ini dikenalkan oleh *Somers* (1962). Koefisien korelasi ini juga memiliki dasar logika yang sama dengan koefisien korelasi *Kendall-tau* dan Gamma, yaitu didasarkan pada banyaknya pasangan konkordan (*C*) dan pasangan diskordan (*D*).

Untuk menghitung statistik (d_{yx}) dari dua buah pengamatan terurut *X* dan *Y*, yaitu $X_1, X_2, \ldots, X_k$ dan $Y_1, Y_2, \ldots, Y_r$. Diberikan dalam tabel kontingensi dibawah ini,

Tabel 1. Tabel Kontingensi Data Kategorik Peringkat.

	X_1	X_2	…	X_k	Total
Y_1	n_{11}	n_{12}	…	n_{1k}	R_1
Y_2	n_{21}	n_{22}	…	n_{2k}	R_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
Y_r	n_{r1}	n_{r2}	…	n_{rk}	R_r
Total	C_1	C_2	…	C_k	N

Sehingga statistik (d_{yx}) didefinisikan sebagai berikut :

$$d_{yx} = \frac{2(C-D)}{N^2 - \sum_{i=1}^{k} C_i^2} \qquad [11]$$

N adalah banyaknya pengamatan dan C_i merupakan frekuensi marginal dari nilai pengamatan *X*. Statistik (d_{yx}) menyatakan selisih proporsi pasangan konkordan dan diskordan diantara pasangan dengan nilai pasangan pengamatan yang berangka sama untuk variabel *X*.

Langkah-langkah menghitung koefisien korelasi Somers (d_{yx}) adalah sebagai berikut:

- Dengan cara yang sama seperti koefisien korelasi Gamma (*G*), hitung banyaknya pasangan konkordan dan diskordan dari tabel kontingensi yang diberikan menggunakan persamaan [9] dan [10].
- Selanjutnya hitung jumlah kuadrat dari banyaknya frekuensi dalam tiap baris di setiap kolomnya.

- Kemudian substitusikan ke persamaan [11].

Hubungan antara Gamma (G) dan Somers (d_{yx})

Dari informasi yang diberikan, diketahui bahwa rumus koefisien korelasi Gamma (*G*) adalah sebagai berikut : $G = \frac{C-D}{C+D}$

dan koefisien korelasi Somers (d_{yx}) adalah sebagai berikut : $d_{yx} = \frac{2(C-D)}{N^2 - \sum_{i=1}^{k} C_i^2}$

sehingga hubungan antara Gamma (*G*) dan Somers (d_{yx}) adalah

$$\frac{G}{d_{yx}} = \frac{\frac{C-D}{C+D}}{\frac{2(C-D)}{N^2 - \sum_{i=1}^{k} C_i^2}} = \frac{N^2 - \sum_{i=1}^{k} C_i^2}{2(C+D)}$$

dengan demikian

$$G = \left(\frac{N^2 - \sum_{i=1}^{k} C_i^2}{2(C+D)} \right) d_{yx} \qquad [12]$$

atau

$$d_{yx} = \left(\frac{2(C+D)}{N^2 - \sum_{i=1}^{k} C_i^2} \right) G \qquad [13]$$

untuk data yang sama, dalam perhitungan *G* dan (d_{yx}) nilai $\frac{2(C+D)}{N^2 - \sum_{i=1}^{k} C_i^2}$ konstan.

Teladan Penerapan

Sebagai ilustrasi, data yang digunakan untuk perhitungan koefisien korelasi nonparametrik untuk skala ordinal adalah data simulasi. Simulasi data ini merupakan dua buah data berpasangan (*X,Y*). Data simulasi terdiri dari dua jenis, yaitu data tidak normal (seragam) dan data normal. Dimana simulasi data dibuat menggunakan program komputer Microsoft EXCEL.

1. Perhitungan Koefisien Korelasi Pearson, Spearman -*rho*,dan *Kendall-tau*

Perhitungan koefisien korelasi Pearson, koefisien korelasi Spearman-*rho*, dan koefisien korelasi Kendall-*tau* (τ), data simulasi yang digunakan itu sama. Data sebanyak 100 sampel, dimana setiap sampel terdiri dari 12 pengamatan. Data simulasi di sini ada dua macam, yang pertama data simulasi dengan sebaran seragam. Kemudian data tadi dibangkitkan sehingga datanya berdistribusi normal. Salahsatu contoh sampelnya adalah sebagai berikut:

Tabel 2. Simulasi Data Seragam

	Simulation Data	
	X	Y
1	48	1
2	55	2
3	48	3
4	7	4
5	17	5
6	84	6
7	87	7
8	22	8
9	45	9
10	56	10
11	74	11
12	34	12

Tabel 3. Simulasi Data Normal

	Simulation Data	
	X	Y
1	-10	11
2	-2	-6
3	4	-17
4	3	-16
5	3	0
6	-4	7
7	-2	-2
8	-13	9
9	2	-2
10	-4	-3
11	0	8
12	-5	2

Setelah simulasi data untuk masing-masing sampel dibuat, maka langkah selanjutnya adalah menghitung masing-masing koefisien korelasi Pearson, Spearman-*rho*, dan Kendall-*tau* (τ) untuk masing-masing sampel baik itu data seragam maupun data normal. Dengan demikian dapat diperoleh secara keseluruhan nilai koefisien korelasi Pearson, Spearman-*rho*, dan Kendall-*tau* untuk 100 sampel baik untuk data seragam maupun data normal yang ditampilkan pada tabel D dan tabel E (di lampiran).

Tabel D menunjukkan nilai masing-masing koefisien korelasi Pearson, Spearman-*rho*, dan Kendall-*tau* dari 100 sampel untuk data seragam. Untuk data seragam, diharapkan bahwa nilai koefisien korelasi Spearman-*rho* dan Kendall-*tau* lebih baik (lebih besar) dibandingkan dengan koefisien korelasi Pearson. Dari Tabel D dapat dilihat bahwa dari 100 sampel kebanyakan nilai-nilai koefisien korelasi yang diberikan oleh koefisien korelasi Spearman-*rho* dan Kendall-*tau* cenderung lebih besar dibandingkan dengan koefisien korelasi Pearson. Ini sesuai dengan yang diharapkan, karena data pengamatan merupakan sampel acak dengan distribusi seragam dan koefisien korelasi yang baik digunakan untuk menghitung data semacam itu adalah koefisien korelasi Spearman-*rho* dan Kendall-*tau* dimana data pengamatannya tidak berdistribusi normal. Sedangkan koefisien korelasi Pearson digunakan untuk menghitung koefisien korelasi dimana data pengamatannya berdistribusi normal, sehingga untuk data seragam nilai koefisien korelasi Pearson yang diberikan lebih kecil.

Tabel E menunjukkan nilai masing-masing koefisien korelasi Pearson, Spearman-*rho*, dan Kendall-*tau* dari 100 sampel untuk data normal. Dan untuk data normal, diharapkan bahwa nilai koefisien korelasi Pearson lebih baik (lebih besar) dibandingkan dengan koefisien korelasi Spearman-*rho* dan Kendall-*tau*. Dari Tabel E untuk ketiga nilai koefisien korelasi, dapat dilihat bahwa dari 100 sampel kebanyakan nilai-nilai koefisien korelasi Pearson lebih besar dibandingkan dengan koefisien korelasi Spearman-*rho* dan Kendall-*tau*. Ini sesuai dengan yang diharapkan, karena koefisien korelasi Pearson memang baik digunakan untuk data pengamatan yang berdistribusi normal. Sedangkan koefisien korelasi Spearman-*rho* dan Kendall-*tau* digunakan untuk data yang tidak berdistribusi normal sehingga nilai koefsien yang diberikan lebih kecil dibandingkan dengan koefisien korelasi Pearson.

Perhitungan Koefisien Korelasi Gamma (*G*) dan Somers (d_{yx})

Untuk perhitungan koefisien korelasi Gamma (*G*) dan Somers (d_{yx}), data simulasi yang digunakan juga sama. Data sebanyak 100 sampel, dimana setiap sampel terdiri dari 1000 pengamatan. Data simulasi di sini juga dua macam yaitu data seragam dan data normal. Karena koefisien korelasi Gamma (*G*) dan Somers (d_{yx}) merupakan suatu ukuran asosiasi dimana data pengamatannya berupa data kategori peringkat dan ditampilkan dalam bentuk tabel kontingensi sehingga data dibangkitkan ke bentuk tabel kontingensi. Disini masing-masing pengamatan *X* dan *Y* dibagi menjadi beberapa kelas, dalam hal ini dibagi ke dalam beberapa kategori. Berikut contoh simulasi data untuk salah satu sampel dengan 1000 pengamatan untuk data simulasi dari data seragam dan data normal.

Tabel 4. Data Simulasi Kategorik Seragam

	X	Y			
1	0.84352	0.46344	4	2	42
2	0.45049	0.51671	2	2	22
3	0.66206	0.33482	3	1	31
4	0.65512	0.17671	3	0	30
5	0.55677	0.29441	2	1	21
6	0.03931	0.25021	0	1	1
7	0.55919	0.99774	2	4	24
8	0.54403	0.13505	2	0	20
9	0.17538	0.63256	0	3	3
10	0.20846	0.51695	1	2	12
. . .					
1000	0.68083	0.61992	3	3	33

Tabel 5. Data Simulasi Kategorik Normal

	X	Y			
1	17.45115	14.72188	1	1	11
2	31.07570	26.42992	3	2	32
3	32.10446	21.43789	3	2	32
4	35.45599	42.42763	3	4	34
5	38.26679	34.41158	3	3	33
6	29.08129	23.76656	2	2	22
7	29.15713	40.98426	2	4	24
8	42.04156	8.83054	4	0	40
9	33.67225	38.77020	3	3	33
10	27.38279	41.45445	2	4	24
. . .					
1000	22.02840	44.96036	2	4	24

Dari tabel 4.9 dan 4.10 di atas untuk tiga kolom terakhir merupakan alat bantu untuk membangkitkan data dari data pengamatan *X* dan *Y* yang diberikan ke bentuk data kategorik, dalam hal ini dibuat ke dalam bentuk tabel kontingensi. Setelah melalui beberapa proses, maka data pengamatan *X* dan *Y* di atas menghasilkan tabel kontingensi sebagai berikut:

Tabel 6.Tabel Kontingensi Data Seragam

	0	1	2	3	4	
0	51	39	41	48	38	217
1	31	28	34	35	34	162
2	38	45	52	53	41	229
3	48	41	31	40	38	198
4	43	43	37	28	43	194
	211	196	195	204	194	1000

Tabel 7.Tabel Kontingensi Data Normal

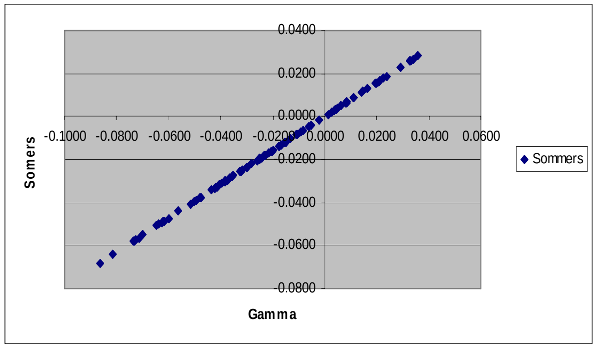
	0	1	2	3	4	
0	0	1	10	6	3	20
1	2	18	53	52	21	146
2	6	47	115	110	48	326
3	11	45	125	115	41	337
4	2	15	47	50	13	127
	21	126	350	333	126	956

Setelah tabel kontingensi dibuat, dapat dihitung nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}).

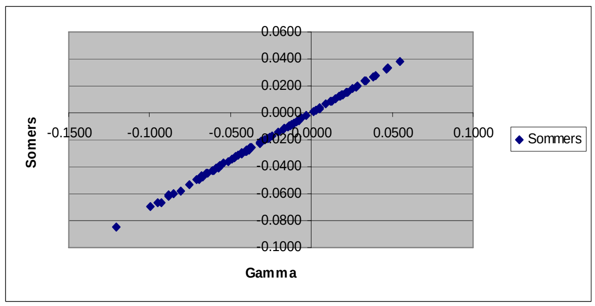
Tabel F dan tabel G (pada lampiran) menunjukan nilai-nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}) untuk data seragam dan data normal. Baik untuk data normal ataupun data seragam, kebanyakan nilai koefisien korelasi Gamma (G) yang diberikan cenderung lebih besar dibandingkan dengan koefisien korelasi Somers (d_{yx}). Ini disebabkan koefisien korelasi Gamma tidak memperhatikan banyaknya data kembar, sedangkan untuk koefisien korelasi Somers (d_{yx}) banyaknya data kembar untuk pengamatan X diperhatikan.

Baik untuk data seragam maupun data normal dari tabel nilai-nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}) untuk 100 sampel, nilai *slope* (kemiringan garis) yang diberikan berturut-turut sebesar 0.7882 dan 0.7021. Nilai *slope* menunjukkan dua hal, yaitu arah hubungan dan besarnya perubahan pada nilai-nilai koefisien korelasi Gamma (G) yang terjadi sehubungan dengan perubahan pada nilai-nilai koefisien korelasi Somers (d_{yx}).

Arah hubungan dapat dilihat dari tanda aljabar (positif atau negatif) pada nilai *slope*. Karena nilai *slope* yang diberikan oleh Tabel F dan tabel G bernilai positif, ini menyatakan bahwa arah hubungan antara nilai-nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}) adalah positif baik untuk data seragam ataupun data normal. Dimana hubungan yang positif menunjukkan bahwa kenaikan nilai koefisien korelasi Gamma (G) diikuti oleh kenaikan pada nilai koefisien korelasi Somers (d_{yx}) dan sebaliknya penurunan nilai koefisien korelasi Gamma (G) diikuti oleh penurunan pada nilai koefisien korelasi Somers (d_{yx}) . Untuk melihat pola hubungan antara nilai-nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}) baik untuk data seragam maupun data normal dapat dilihat pada gambar grafik di bawah berikut ini:



Gambar 2. *Hubungan linier antara Gamma dan Somers untuk data seragam*



Gambar 3. *Hubungan linier antara Gamma dan Somers untuk data normal*

Pola hubungan antara nilai-nilai koefisien korelasi Gamma (G) dan Somers (d_{yx}) baik untuk data seragam maupun data normal yang terlihat pada gambar 2 dan gambar 3 di atas adalah hubungan yang bersifat *linier* karena dapat dihampiri oleh sebuah garis lurus. Sehingga hubungan antara nilai-nilai koefisien korelasi Gamma (G) dan nilai-nilai koefisien korelasi Somers (d_{yx}) adalah *linier*.

Kesimpulan

Koefisien korelasi yang dapat digunakan untuk skala data ordinal adalah koefisien korelasi Spearman-*rho* (ρ), Kendall-*tau* (τ), Gamma (G), dan Somers (d_{yx}). Untuk data yang tidak normal (data seragam), nilai koefisien korelasi yang diberikan oleh koefisien korelasi Spearman- *rho* dan Kendall- *tau* lebih besar dibandingkan dengan koefisien korelasi Pearson. Sedangkan untuk data normal nilai koefisien korelasi Pearson lebih besar dibandingkan dengan koefisien korelasi Spearman- *rho* dan Kendall- *tau*. Hal ini menunjukkan bahwa :

- Koefisien korelasi Pearson (r) baik digunakan jika data pengamatan berdistribusi normal dan skala data serendah-rendahnya adalah interval atau rasio.
- Koefisien korelasi Spearman-*rho* (ρ) dan Kendall-*tau* (τ) baik digunakan untuk pasangan pengamatan yang tidak berdistribusi normal.

Koefisien korelasi Gamma (G) dan Somers (d_{yx}) digunakan untuk pasangan pengamatan dengan skala data ordinal dalam bentuk kategorik peringkat (data ditampilkan dalam bentuk tabel kontingensi) dan Koefisien korelasi Gamma (G) dan Somers (d_{yx}) menunjukkan hubungan yang linier.

DAFTAR PUSTAKA

[1] Agresti, 1984. *Analysis of Ordinal Categorical Data.* John Wiley and Sons. New York.

[2] Anonim. 2003. *Introduction to Exact Nonparametric Inference*. http://www.cytel.com/Products/StatXact/Intro_Nonparametric_Inference.pdf

[3] Anonim. 2000. *Correlation*. http://www.blackwellpublishing.com/content/BPL_Images/Content_store/Sample_chapter/9781405127806/Petrie%20sample%20Ch26.pdf

[4] Aryee, M. 2002. *Measures of Association*. http://academic.shu.edu/eop/worksheets/exac2126/PRE--Measures%20of%20Association--1203.doc

[5] Azizi.2005. *Analisis Berstatistik Lanjutan*. http://www.geocities.com/kheru2006/vii.htm

[6] Conover, W.J. 1971. *Practical Nonparametric Statistics*. Wiley International Edition. John Wiley and Sons. New York, NY.

[7] Daniel, W. 1989. *Statistika Nonparametrik Terapan*. Penerbit PT. Gramedia. Jakarta.

[8] Djarwanto, 1997. *Statistika Nonparametrik*. BPFE - Yogyakarta. Yogyakarta.

[9] Elifson, K.W, and R. Runyon. 1990. *Fundamental of Social Statistics*. Second Edition. McGraw-Hill International Edition. Singapore.

[10] Gibbon, J.D. 1985. *Nonparametric Statistical Inference*. Marcel Dekker. New York, NY.

[11] Loether, H. and D.G. McTavish.1988. *Descriptive and Inferential Statistics*: *An Introduction*, *Third Edition*. Allyn and Bacon. Needham Heights, MA.

[12] Lohninger, H. 2006. *Ordinal Association*. http://www.statisticssolutions.com/ordinal-association.htm

[13] SAS Institute, 1999. *Measures of Association*. http://v8doc.sas.com/sashtml/stat/chap28/sect20.htm

[14] Scheaffer R.L. 1999. *Categorical Data Analysis*. http://courses.ncssm.edu/math/Stat_Inst/PDFS/Categorical%20Data%20Analysis.pdf

[15] Siegel, S., and J. Castellan, Jr. 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill International Edition. Singapore.