Statistika, Vol. 22 No. 1, 1 – 11
Mei 2022

Kode R dan Selang Kepercayaan Korelasi Berdasarkan Empirical Likelihood serta Implementasinya pada Korelasi PDRB dengan Jumlah Kasus Covid-19 di Indonesia

SULIADI

Program Studi Statistika Fakultas MIPA Universitas Islam Bandung, Indonesia
e-mail: suliadi@gmail.com; suliadi@unisba.ac.id

ABSTRAK

Korelasi merupakan suatu ukuran untuk melihat kekuatan hubungan linier di antara dua variabel. Ada beberapa metode yang biasa digunakan untuk mengukur korelasi, di antaranya adalah korelasi pearson, peringkat spearman dan kendall tau. Metode yang biasa digunakan untuk mengukur korelasi untuk variabel bertipe numerik adalah korelasi pearson. Metode ini mensyaratkan bahwa kedua variabel tersebut berdistribusi normal bivariat. Oleh karena itu inferensia korelasi pearson hasilnya akan valid jika asumsi tersebut terpenuhi. Dalam praktik, sering kali kenormalan data tersebut tidak dapat dipenuhi. Satu pendekatan telah diajukan dalam pembuatan selang kepercayaan berdasarkan *empirical likelihood.* Metode ini adalah metode bebas distribusi yang artinya tidak ada asumsi bahwa data harus berdistribusi tertentu. Dalam artikel ini kami membahas penyusunan selang kepercayaan korelasi pearson berdasarkan metode *empirical likelihood* dan juga menyediakan kode perintah R Language untuk pembuatan selang kepercayaan tersebut. Kami menerapkan metode tersebut pada kasus hubungan antara PDRB dan jumlah kasus Covid-19 berdasarkan data provinsi di Indonesia Tahun 2020. Kami mendapatkan adanya hubungan yang sangat kuat antara PDRB dan jumlah kasus Covid-19 di Indonesia dengan korelasi sebesar 0.939 dan dengan metode tersebut diperoleh batas bawah selang kepercayaan 99% adalah 0.872 dan batas atasnya adalah 0.962.

Kata Kunci: *Empirical likelihood*, Koefisien korelasi pearson, Kode R, PDRB, Covid-19

ABSTRACT

Correlation is a measure to see the strength of the linear relationship between two variables. There are several methods commonly used to measure correlation, including Pearson's correlation, Spearman's rank and Kendall tau. The method commonly used to measure correlations for numeric type variables is Pearson correlation. This method requires that the two variables have a bivariate normal distribution. Therefore, the Pearson correlation inferential results will be valid if these assumptions are met. In practice, often the normality of the data cannot be met. One approach has been proposed in constructing confidence intervals based on empirical likelihood. This method is a distribution-free method, which means that there is no assumption that the data must have a certain distribution. In this article, we discuss the construction of the Pearson correlation confidence interval based on the empirical likelihood method and also provide the R Language command code for constructing the confidence interval. We apply this method to the case of the relationship between GRDP and the number of Covid-19 cases based on provincial data in Indonesia in 2020. We found a very strong relationship between GRDP and the number of Covid-19 cases in Indonesia with a correlation of 0.939 and with this method the lower limit of the 99% confidence interval was 0.872 and the upper limit was 0.962.

Keywords: Empirical likelihood, Pearson correlation coefficient, R Code, GRDP, Covid-19.

1. PENDAHULUAN

Korelasi adalah suatu ukuran untuk melihat kekuatan hubungan linier di antara dua variabel. Ukuran ini banyak sekali digunakan pada berbagai bidang seperti sosial, ekonomi, biologi, maupun di bidang ilmu lainya. Ada beberapa metode yang dapat digunakan untuk mengukur besarnya korelasi baik yang bersifat parametrik maupun nonparametrik, seperti korelasi pearson, peringkat spearman dan kendall tau.

Korelasi pearson merupakan metode yang paling sering digunakan sebagai ukuran korelasi untuk data bertipe numerik. Metode ini diperkenalkan oleh Karl Pearson pada tahun 1895. Metode ini mempersayaratkan bahwa data berasal dari distribusi nomal bivariat. Jika asumsi ini terpenuhi, maka korelasi pearson merupakan penduga korelasi yang paling baik, serta memberikan selang kepercayaan dan juga pengujian hipotesis yang valid. Metode yang biasa digunakan untuk membuat selang kepercayaan adalah modifikasi transformasi-z dari Fisher yang diajukan oleh Hotelling (1953). Beberapa penelitian telah dilakukan untuk memperbaiki kinerja dari modifikasi transformasi-z dari Fisher seperti yang dilakukan oleh Weerakkody & Givaruangsawat (1995), Sun & Wong (2007) dan Tian & Wilding (2008). Metode yang diajukan pada penelitian tersebut di atas didasarkan pada distribusi normal bivariat. Oleh karena itu metode tersebut tidak lagi valid jika asumsi kenormalan data tidak terpenuhi.

Hu et al. (2018) mengajukan metoe untuk membangun selang kepercayaan untuk koefisien korelasi berdasarkan *empirical likelihood*. Metode ini bersifat bebas distribusi, sehingga dapat diterapkan untuk sembarang distribusi. Hu et al. (2018) melaporkan bahwa metode ini kinerjanya setara dengan metode modifikasi transformasi-z dari Fisher untuk data berdistribusi normal bivariat dalam hal cakupan peluangnya. Untuk data nonnormal, metode ini lebih unggul dibandingkan metode lain.

Artikel ini membahas bagaimana menyusun selang kepercayaan untuk koefisien korelasi berdasarkan *empirical likelihood* yang diusulkan oleh Hu et al. (2018), serta implementasinya dalam bentuk kode program dalam R. Metode tersebut diterapkan pada kasus hubungan antara PDRB dan jumlah kasus Covid-19 di Indonesia untuk data Tahun 2020.

2. METODE PENELITIAN

Misalkan (X,Y) adalah vektor acak bivariat dengan nilai tengah μ dan ragam Σ di mana

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}, \qquad \text{... (1)}$$

Koefisien korelasi didefinisikan oleh

$$\rho = \frac{E(X-\mu_x)(Y-\mu_y)}{\sqrt{E(X-\mu_x)^2}\sqrt{E(Y-\mu_y)^2}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y},$$

di mana $-1.0 \le \rho \le 1.0$. Penduga koefisien korelasi pearson adalah

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}. \qquad \text{... (2)}$$

Fisher (1915) dan Hotelling (1953) mengajukan distribusi sampling untuk r, akan tetapi bentuknya sangat komplek. Beberapa penelitian juga telah dilakukan untuk mendapatkan distribusi sampling yang lebih sederhana seperti yang dilakukan oleh Sun & Wong (2007) dan Nie et al. (2011). Meskipun demikian, pendekatan tranformasi yang digunakan oleh Hotteling adalah yang paling sederhana dan paling banyak digunakan. Jika (X,Y) adalah vektor acak dari bivariat normal, maka $(1/2)\ln[(1+r)/(1-r)]$ secara aproksimai akan berdistribusi normal dengan nilai tengah $(1/2)\ln[(1+\rho)/(1-\rho)]$ dan ragam $1/(n-3)$. Oleh karena itu

$$z = \frac{\sqrt{n-3}}{2}\left[\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho}{1-\rho}\right)\right] = \frac{\sqrt{n-3}}{2}\ln\left[\frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}\right]$$

secara aproksimasi, z akan berdistribusi normal (0,1). Berdasarkan distribusi z tersebut maka dapat disusun selang kepercayaan untuk ρ.

Distribusi sampling dari (1/2)ln[(1+r)/(1-r)] di atas disusun berdasarkan asumsi bahwa (X,Y) berdistribusi normal bivariat. Oleh karena itu selang tersebut valid jika asumsi kenormalan data terpenuhi. Dalam praktik, sangat sering asumsi tersebut tidak dapat dipenuhi. Hu et al. (2018) mengajukan suatu pendekatan lain untuk menyusun selang kepercayaan tanpa memerlukan asumsi distribusi data. Dengan kata lain metode ini bersifat bebas distribusi. Metode yang diajukan oleh Hu, et al ini didasarkan pada metode *empirical likelihood* (Owen, 1988; Owen, 2001).

Empirical likelihood merupakan metode untuk menyusun selang kepercayaan berdasarkan distribusi empiris data tanpa mengasumsikan distribusi tertentu. Kelebihan dari *empirical likelihood* adalah bahwa meskipun data tidak diasumsikan dari distribusi tertentu (*free distribution*), akan tetapi selang kepercayaan dibentuk dengan menggunakan pendekatan fungsi likelihood, dalam hal ini menggunakan pendekatan *likelihood ratio test*.

Misalkan $w_1, w_2, \ldots, w_n$ adalah sampel acak berukuran n, di mana w_i bisa berupa vektor atau skalar dari suatu distribusi dengan parameter θ. Misalkan juga $p_1, p_2, \ldots, p_n$ adalah satu set bobot peluang untuk pengamatan ke-1, 2, ..., n, di mana $0 < p_i < 1.0$ dan $\Sigma p_i = 1.0$. *Empirical likelihood* untuk θ didefinisikan sebagai (Owen,1988; Owen, 1991; Owen, 2001; Chen & Van Keilegom, 2009)

$$L_n(\theta) = \max \prod_{i=1}^{n} p_i. \qquad \ldots (3)$$

Jika tidak ada konstrain apa-apa, maka $L(\theta)$ akan maksimum jika $\hat{p}_1 = \hat{p}_2 = \cdots = \hat{p}_n = 1/n$. Pada kondisi ini

$$L(p_1, p_2, \ldots, p_n) \le L(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n).$$

Dengan kata lain bahwa solusi yang memaksimumkan $L(\theta)$ pada (3) akan diperoleh jika setiap sampel diberi bobot yang sama yaitu 1/n, yang tidak lain adalah peluang dari distribusi uniform diskrit.

Misalkan $g(w_i,\theta)$ adalah suatu fungsi dari sampel dan parameter, yang memenuhi $g(w_i,\theta)=0$. Sebagai contoh jika $\theta = E(W_i) = \mu$, maka $g(w_i,\mu) = w_i-\mu$. Parameter θ diperoleh dengan memaksimumkan fungsi $L(\theta)$ dengan kendala

$$\sum_{i=1}^{n} p_i = 1.0 \quad \text{dan} \sum_{i=1}^{n} p_i g(w_i, \theta) = 0.$$

Untuk pengujian hipotesis H_0: $\theta = \theta_0$ maka kendala di atas menjadi

$$\sum_{i=1}^{n} p_i = 1.0 \quad \text{dan} \sum_{i=1}^{n} p_i g(w_i, \theta_0) = 0. \qquad \ldots (4)$$

Ratio likelihood untuk hipotesis di atas adalah

$$R(\theta_0) = \frac{L(\theta_0)}{L_n(\theta)} = \frac{\prod p_i}{1/n} = \prod n p_i \qquad \ldots (5)$$

atau dalam bentuk log $\log R = \sum_{i=1}^{n} \log(np_i)$. Bobot peluang di bawah H_0, yaitu $p_1, p_2, \ldots, p_n$ dapat diperoleh dari

$$G(\theta_0) = \max_{p_1, p_2, \ldots, p_n} \sum_{i=1}^{n} \log(np_i) \left| p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(w_i, \theta_0) = 0 \right. . \qquad \ldots (6)$$

Dengan pengganda lagrange, maka fungsi tujuan di atas dapat dituliskan sebagai

$$G = \sum_{i=1}^{n} \log(np_i) + \gamma\left(\sum_{i=1}^{n} p_i - 1\right) - n\lambda \sum_{i=1}^{n} p_i g(w_i, \theta_0).$$

Solusi dari $p_1, p_2, \ldots, p_n$ diperoleh melalui

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\lambda g(w_i, \theta) + \gamma = 0; \quad \frac{\partial G}{\partial n\lambda} = \sum p_i g(w_i, \theta) = 0; \quad \frac{\partial G}{\partial \gamma} = \left(\sum p_i\right) - 1 = 0. \qquad \ldots (7)$$

Karena $\Sigma p_i[\partial G/\partial p_i]=0$, maka akan diperoleh (Owen, 2001) $\gamma = -n$ dan

$$p_i = 1/[n(1 + \lambda g(w_i, \theta))]. \qquad \ldots (8)$$

Berdasarkan hasil tersebut, maka λ dapat diperoleh sebagai solusi dari

$$\sum_{i=1}^{n} \frac{g(w_i,\theta)}{n+n\lambda g(w_i,\theta)} = 0 \,. \qquad \ldots (9)$$

Pada kondisi H_0, maka λ harus dipilih sehingga memenuhi

$$\sum_{i=1}^{n} \frac{g(w_i,\theta_0)}{n+n\lambda g(w_i,\theta_0)} = 0 \,. \qquad \ldots (10)$$

Jika kondisi tersebut terpenuhi, maka akan diperoleh $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$ yang juga memenuhi H_0 dan konstrain pada (4).

Selang kepercayaan untuk parameter θ berdasarkan empirical likelihood pertama kali diajukan oleh Owen (1988) berdasarkan teorema yang dibangunnya, bahwa untuk n → ∞ maka -2log R(θ_0), untuk R(θ_0) pada persamaan (5) akan konvergen dalam distribusi ke χ_p^2. Berdasarkan hal tersebut, maka pengujian hipotesis dan pendugaan selang kepercayaan dapat dilakukan berdasarkan distribusi chi-square.

3. HASIL DAN PEMBAHASAN

SK untuk Koefisien Korelasi Berdasarkan Empirical Likelihood

Hu et al. (2018) telah mengajukan metode untuk membangun selang kepercayaan berdasarkan *empirical likelihood*. Metode tersebut memanfaatkan *influence function* untuk korelasi dalam penyusunan selang kepercayaan. Metode ini mempunyai kinerja yang bagus untuk data normal dan kinerjanya setara dengan metode transformasi dari Hotelling (1953) dalam hal *coverage probability* dan taraf signifikansi. Berikut ini penurunan metode tersebut yang disampaikan oleh Hu et al. (2018).

Misalkan $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ adalah sampel acak dari distribusi bivariat (X,Y) dengan nilai tengah dan matriks kovarians diberikan oleh (1). Berdasarkan hal tersebut, maka

$$E\left(\frac{X-\mu_x}{\sigma_x}\cdot\frac{Y-\mu_y}{\sigma_y}-\rho\right)=0.$$

Oleh karena itu log empirical likelihood untuk ρ diberikan oleh

$$L(\rho)=\prod_{i=1}^{n} p_i;\;\; p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(g(w_i)-\rho)=0. \quad \ldots (11)$$

di mana $g(w_i)=\frac{x_i-\mu_x}{\sigma_x}\cdot\frac{y_i-\mu_y}{\sigma_y}$. Sebagaimana pada persamaan (5), ratio likelihood diberikan oleh

$$R(\rho)=\prod np_i$$

Karena $\mu_x, \mu_y, \sigma_x, \sigma_y$ biasanya tidak diketahui, maka dapat diganti oleh penduganya yaitu $\bar{x}, \bar{y}, S_x, S_y$, di mana S_x dan S_y, masing-masing adalah simpangan baku sampel untuk *X* dan *Y*. Sebagaimana pada persamaan (5) dan (6) maka log *ratio likelihood* untuk (11) diberikan

$$G(\theta_0)=\max_{p_1,p_2,\ldots,p_n} \sum_{i=1}^{n}\log(np_i) \bigg| p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(g(w_i)-\rho)=0,$$

sehingga berdasarkan hasil pada (7)-(10) maka penduga bagi p_i diperoleh melalui

$$p_i = 1/[n\{1+\lambda[g(w_i)-\rho]\}] \qquad \ldots (12)$$

di mana λ diperoleh sebagai solusi dari

$$\sum_{i=1}^{n} \frac{g(w_i)-\rho}{n+n\lambda[g(w_i)-\rho]}=0.$$

Hu et al. (2018) menyatakan bahwa diperlukan perhitungan yang agak komplek dari A agar supaya dari [-2A log R(ρ)] secara aproksimasi berdistribusi chi-square. Mereka menyarankan menggunakan *influence function* untuk ρ yang didefinisikan sebagai pengganti g(w_i)-ρ, yaitu

$$g_I(w_i\rho) = \left(\frac{x_i-\mu_x}{\sigma_x}.\frac{y_i-\mu_y}{\sigma_y}\right) - (1/2)\rho\left[\left\{\left(\frac{x_i-\mu_x}{\sigma_x}\right)^2 - 1\right\} + \left\{\left(\frac{y_i-\mu_y}{\sigma_y}\right)^2 - 1\right\}\right]. \quad \text{... (13)}$$

Nilai $\mu_x, \mu_y, \sigma_x, \sigma_y$ dapat diganti oleh penduganya yaitu $\bar{x}, \bar{y}, S_x, S_y$ seperti sebelumnya, sedangkan ρ pada (13) diambil seperti pada H_0. Empirical likelihood untuk rho berdasarkan *influence function* di atas diberikan oleh

$$L_I(\rho) = \sup_{p_1, p_2, \ldots p_n}\left\{\prod_{i=1}^n p_i;\, p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_I(w_i, \rho) = 0\right\}$$

Dengan cara yang sama sebagaimana pada (7), maka solusi bagi p_i diperoleh seperti pada (12), yaitu

$$p_i = 1/[n\{1+\lambda g_I(w_i, \rho)\}] \quad \text{... (14)}$$

sedangkan λ diperoleh sebagai solusi dari

$$\sum_{i=1}^n \frac{g_I(w_i, \rho)}{n + n\lambda g_I(w_i, \rho)} = 0. \quad \text{... (15)}$$

Hu et al. (2018) menunjukkan bahwa

$$l_I(\rho) = -2 \log R(\rho) \xrightarrow{d} \chi^2_{(1)}.$$

Oleh karena itu selang kepercayaan bagi ρ diperoleh melalui

$$\{\rho : l_I(\rho) \leq \chi^2_{(1-\alpha;1)}\}. \quad \text{... (16)}$$

Implementasi pada R

Implementasi metode di atas memerlukan alur pemikiran yang jelas agar dapat diimplementasi ke dalam suatu kode perintah atau program komputer. Rasionalisasi kode perintah di R adalah sebagai berikut.

Untuk mencari batas bawah dan batas atas selang, maka harus dicari suatu ρ* yang memenuhi persamaan **Error! Reference source not found.**. Kami menggunakan metode bisection dengan mencoba ke arah kanan dan kiri nilai yang memenuhi persamaan tersebut. Beberapa hal yang perlu diperhatikan adalah bahwa -1.0 ≤ ρ ≤ 1.0 dan nilai l_I (ρ) minimum pada ρ = r, di mana r adalah penduga bagi koefisien korelasi pearson pada persamaan (2). Oleh karena itu pencarian ρ* dilakukan ke arah kiri menuju -1.0 untuk mendapatkan batas bawah selang dan ke arah kanan menuju 1.0 untuk mendapatkan batas atas selang. Untuk mendapatkan solusi bagi λ kami menggunakan package nleqslv dari Hasselman (2018). Algoritma untuk mendapatkan selang kepercayaan untuk ρ berdasarkan *empirical likelihood* adalah sebagai berikut.

A. Penentuan batas bawah
   1) Tentukan δ=0.1dan r_0 = r dari persamaan (2)
   2) r_0 = r_0 - δ dan cari λ, solusi bagi (15) dan hitung p_i menggunakan (14).
   3) Hitung likelihood ratio $l_I(r_0) = -2 \log R(r_0)$.
   4) Jika $l_I(r_0) < \chi^2_{(1-\alpha;1)}$, kembali A.2.
   5) Jika ketelitian sudah cukup, *stop*. Selainnya set: r_0 = r_0 + δ; δ = 0.1*δ dan kembali ke A.2.

B. Penentuan batas atas
   1) Tentukan δ=0.1dan r_0 = r dari persamaan (2)
   2) r_0 = r_0 + δ dan cari λ, solusi bagi (15) dan hitung p_i menggunakan (14).

3) Hitung *likelihood ratio* $l_l(r_0) = -2 \log R(r_0)$.

4) Jika $l_l(r_0) < \chi^2_{(1-\alpha;1)}$, kembali B.2.

5) Jika ketelitian sudah cukup, *stop*. Selainnya set: $r_0 = r_0 - \delta$; $\delta = 0.1*\delta$ dan kembali ke B.2.

Kode R dari algoritma di atas dapat dilihat pada lampiran. Perintah untuk mendapatkan selang kepercayaan adalah

bisec.CI(xx=x,yy=y,alpha1=0.95)

di mana `xx` dan `yy` adalah vektor data dan `alpha1` adalah (1-α) untuk taraf signifikansi α untuk membuat selang kepercayaan dengan lebar (1-α)x100%. Output dari fungsi ini bertipe list yang berisi

`Korelasi`	: koefisien korelasi pearson
`alpha1`	: nilai (1-α)
`Lower`	: batas bawah selang kepercayaan
`Upper`	: batas atas selang kepercayaan
`LL.Upper`	: nilai -2log R(ρ) untuk rho = Lower
`LL.Lower`	: nilai -2log R(ρ) untuk rho = Upper

Berikut ini contoh berdasarkan data dari distribusi normal bivariat dengan korelasi ρ = -0.3 dan 0.9 dan n = 100. Sebagai catatan bahwa $\chi^2_{(0.95;1)} = 3.841459$.

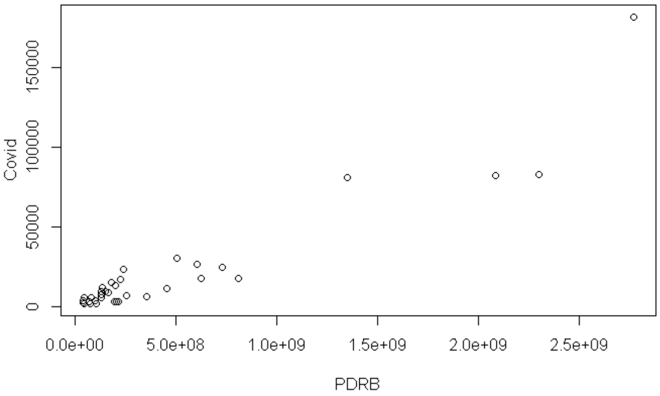
```
> library(mvtnorm)
> ## rho=-0.3
> d<-rmvnorm(100,sigma=matrix(c(1,-0.3,-0.3,1),ncol=2,nrow=2))
> x=d[,1]
> y=d[,2]
>
> bisec.CI(xx=x,yy=y)
$Korelasi
[1] -0.3419467
$alpha1
[1] 0.95
$Lower
[1] -0.5146661
$Upper
[1] -0.1420663
$LL.Upper
[1] 3.841459
$LL.Lower
[1] 3.841459
```

Nilai dugaan korelasi pearsonnya adalah -0.3419467 dengan batas bawah dan batas atas selang kepercayaan 95% berdasarkan metode *empirical likelihood* masing-masing adalah -0.5146661 dan -0.1420663.

```
> ## rho=0.9
> d<-rmvnorm(100,sigma=matrix(c(1,0.9,0.9,1),ncol=2,nrow=2))
> x=d[,1]
> y=d[,2]
>
> bisec.CI(xx=x,yy=y)
$Korelasi
[1] 0.9052214
$alpha1
[1] 0.95
$Lower
[1] 0.8622063
$Upper
[1] 0.9358708
$LL.Upper
[1] 3.841458
$LL.Lower
```

```
[1] 3.841459
```

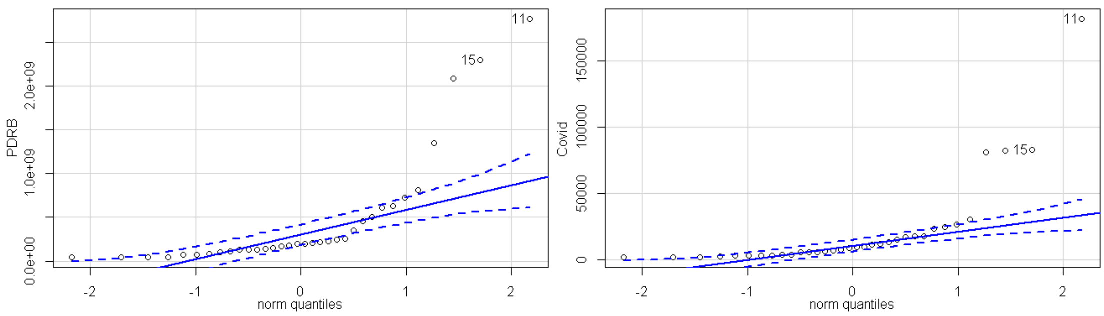
Nilai dugaan korelasi pearsonnya adalah 0.9052214 dengan batas bawah dan batas atas selang kepercayaan 95% berdasarkan metode *empirical likelihood* masing-masing adalah 0.8622063 dan 0.9358708.



Gambar 1 Pola hubungan antara PDRB dan jumlah terkonfirmasi positif Covid-19 dari 34 provinsi di Indonesia.

Korelasi antara PDRB dan Kasus Positif Covid-19

Covid-19 adalah suatu istilah yang merujuk pada Coronavirus Disease 2019, yaitu suatu penyakit yang disebabkan oleh virus SARS-Cov-2. Penyakit ini telah menjadi pandemi yang mematikan di seluruh dunia. Virus ini mulai terdeteksi di Indonesia pada bulan Maret 2020 dan semenjak itu kasus positif terinfeksi Covid-19 terus meningkat. Pemerintah telah mengeluarkan berbagai aturan untuk meredam laju penularan virus tersebut seperti Keputusan Presiden No. 12 Tahun 2020 tentang Penetapan Bencana Nonalam Penyebaran Coronavirus Disease 2019 (COVID-19) Sebagai Bencana Nasional dan juga Peraturan Meteri Kesehatan No.9 Tahun 2020 tentang Pedoman Pembatasan Sosial berskala Besar dalam Rangka Percepatan Penanganan Coronavirus Disease 2019. Inti dari kebijakan tersebut adalah mengurangi kegiatan yang dilakukan secara bersama-sama, di mana hal tersebut akan berimbas kepada kegiatan ekonomi. Salah satu tolok ukur dari aktivitas ekonomi adalah produk domestik regional bruto (PDRB). PDRB merupakan jumlah transaksi yang terjadi di masyarakat berupa nilai tambah yang dihasilkan dari seluruh kegiatan ekonomi. Berdasarkan hal tersebut, maka penting untuk mengetahui bagaimana hubungan antara kasus terkonfirmasi positif Covid-19 dan PDRB sebagia tolok ukur kegiatan ekonomi. Data yang digunakan adalah banyaknya kasus positif Covid-19 pada tahun 2020 per provinsi (https://covid19.go.id) dan data PDRB per provinsi berdasarkan harga berlaku 2020 (BPS, 2021).



Gambar 2 Plot kuantil-kuantil normal untuk PDRB (kiri) dan jumlah terkonfirmasi positif Covid-19 dari 34 provinsi di Indonesia

Pola hubungan antara jumlah kasus terkonfirmasi Covid-19 dan PDRB ditunjukkan oleh Gambar 1. Dari gambar tersebut tampak sekali ada hubungan linier antara kedua variabel. Gambar 2 menyajikan plot kuantil-kuantil (qq-plot). Dari plot tersebut tampak bahwa kedua variabel tidak berdistribusi normal. Untuk memastikan kenormalan data, dilakukan uji menggunakan metode shapiro wilks. Untuk pengujian bivariat normal, digunakan generalisasi dari shapiro-wilks test untuk normal multivariat yang ada pada package `mvShapiroTest` (Gonzalez-Estrada & Villasenor-Alva, 2013). Hasil pengujian disajikan pada Tabel 1. Dari tabel diperoleh bahwa masing-masing variabel PDRB dan jumlah terkonfirmasi positif Covid-19 tidak berdistribusi normal. Demikian juga kedua variabel tersebut juga tidak berdistribusi normal bivariat.

Tabel 1 Hasil pengujian kenormalan data dengan H_0: data berdistribusi normal

Variabel	**P-value**	**Kesimpulan**
Covid-19	4.235e-08	Tidak berdistribusi normal
PDRB	4.877e-09	Tidak berdistribusi normal
Covid-19 dan PDB	1.005e-09	Tidak berdistribusi normal bivariat

Karena asumsi kenormalan data tidak terpenuhi, maka diterapkan metode *empirical likelihood* dan kode perintah R di atas untuk menduga selang kepercayaan koefisien korelasi antara PDRB dan jumlah terkonfirmasi positif Covid-19 pada provinsi-provinsi di Indonesia dengan tingkat kepercayaan 99%. Berikut ini perintah dan hasilnya:

```
> bisec.CI(xx=PDRB,yy=Covid,alpha1=0.99)
$Korelasi
[1] 0.9388992
$alpha1
[1] 0.99
$Lower
[1] 0.8717677
$Upper
[1] 0.9615555
$LL.Upper
[1] 6.634894
$LL.Lower
[1] 6.634897
```

Dari *output* di atas diperoleh nilai dugaan koefisien korelasi pearson antara PDRB dan jumlah terkonfirmasi positif Covid-19 pada provinsi-provinsi di Indonesa adalah 0.939 dengan batas bawah selang kepercayaan 99% adalah 0.872 dan batas atasnya adalah 0.962. Dari hasil ini tampak adanya hubungan yang sangat kuat antara PDRB sebagai representasi aktivitas ekonomi dan jumlah terinfeksi Covid-19.

4. SIMPULAN DAN SARAN

Metode yang diajukan oleh Hu et al. (2018) sangat fleksibel diterapkan dalam praktik keseharian karena tidak memerlukan asumsi distribusi tertentu. Metode ini valid diterapkan untuk data-data bertipe numerik. Selain itu dengan kode program R yang disertakan, hal itu akan mempermudah penerapannya dalam praktik.

Dari analisis terhadap data PDRB dan data jumlah positif Covid-19 disimpulkan adanya hubungan yang sangat kuat antara PDRB dan jumlah kasus Covid-19 di Indonesia. Berdasarkan hasil tersebut maka kebijakan pemerintah dengan menerapkan pembatasan sosial berskala besar (PSBB) dan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) untuk menurunkan tingkat penularan sudah tepat. Dengan menerapkan kebijakan tersebut diharapkan akan menurunkan jumlah kasus positif terinfeksi Covid-19, meskipun hal tersebut juga akan menurunkan akitivitas ekonomi yang pada akhirnya akan menurunkan PDRB.

DAFTAR PUSTAKA

BPS. (2021). Produk Domestik Regional Bruto Provinsi-Provinsi di Indonesia Menurut Pengeluaran 2016-2020. Badan Pusat Statistik, Indoesia.

Chen, S.X., I. Van Keilegom. (2009). A review on empirical likelihood methods for Regression. TEST Vol. 18, p415–447.

Fisher, R.A. (1915). Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. Biometrika, Vl. 10, 507–521.

Gonzalez-Estrada, E. and J.A. Villasenor-Alva. (2013). mvShapiroTest: Generalized Shapiro-Wilk test for multivariate normality. R package version 1.0.
https://CRAN.R-project.org/package=mvShapiroTest.

Hasselman, B. (2018). nleqslv: Solve Systems of Nonlinear Equations. R package version 3.3.2.
https://CRAN.R-project.org/package=nleqslv.

Hotelling, H. (1953). New Light on the Correlation Coefficient and its Transform. Journal of the Royal Statistical Society, Series B, Vol. 15,p193–232.

Hu, H., A. Jung & G. Qin. (2018). Interval Estimation for the Correlation Coefficient. The American Statistician, DOI: 10.1080/00031305.2018.1437077.

Nie, L., Y. Chen, and H. Chu. (2011). Asymptotic Variance of Maximum Likelihood Estimator for the Correlation Coefficient from a BVN Distribution withOne Variable Subject to Censoring. Journal of Statistical Planning and Inference, Vol. 141, p392–401.

Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. Biometrika, Vol. 75, p237-249.

Owen, A.B. (1991). Empirical Likelihood for Linear Models. The Annals of Statistics. Vol. 19, p1725-1747.

Owen, A. B. (2001). Empirical likelihood. Chapman & Hall/CRC. New York, USA.

Sun, Y. and A.C.M. Wong. (2007). Interval Estimation for the Normal Correlation Coefficient. Statistic & Probability Letters, Vol. 77, p1652–1661.

Tian, L. and G.E. Wilding. (2008). Confidence Interval Estimation of a Common Correlation Coefficient. Computational Statistics and Data Analysis, Vol. 52, p4872–4877.

Weerakkody, G.J., and S. Givaruangsawat. (1995). Estimating the Correlation Coefficient in the Presence of Correlated Observations from a Bivariate Normal Population. Communication in Statistics—Theory and Methods, Vol. 24, p1705–1719.

LAMPIRAN

Kode program R untuk pembuatan selang kepercayaan berdasarkan empirical likelihood.

```
##mencari batas atas dan batas bawah##
library(nleqslv)

flambda<-function(x,rho=rho,vi=vi){
  n<-length(vi)
  temp=sum(vi/(1+x*vi))
  temp<-temp/n
  return(temp)
} ## end function

cari.lambda<-function(rho=rho,vi=vi){
  n<-length(vi)
  min<-((1/n)-1)/min(vi)
  max<-((1/n)-1)/max(vi)
  ave<-(min+max)/2
```

```
  l0<-(nleqslv(0,flambda,rho=rho,vi=vi))$x
  lmin<-(nleqslv(min,flambda,rho=rho,vi=vi))$x
  lmax<-(nleqslv(max,flambda,rho=rho,vi=vi))$x
  lave<-(nleqslv(ave,flambda,rho=rho,vi=vi))$x
  pi0<-(1/n)*(1/(1+l0*vi))
  pimin<-(1/n)*(1/(1+lmin*vi))
  pimax<-(1/n)*(1/(1+lmax*vi))
  piave<-(1/n)*(1/(1+lave*vi))
  if ((min(pi0)>=0)&(max(pi0)<=1)){
    lambda=l0
    pi=pi0  }
  else
    if ((min(pimin)>=0)&(max(pimin)<=1)){
      lambda=lmin
      pi=pimin  }
  else
    if ((min(pimax)>=0)&(max(pimax)<=1)){
      lambda=lmax
      pi=pimax  }
  else
    if ((min(piave)>=0)&(max(piave)<=1)){
      lambda=lave
      pi=piave  }
  else{
    lambda<-NA
    pi<-rep(NA,n)
  }
  return(lambda)
} ## end function

cari.pi<-function(lambda,rho=rho,vi=vi){
  n<-length(vi)
  pi<-(1/n)*(1/(1+lambda*vi))
  return(pi)
}## end function

#prob<-cari.pi(lambda,rho=rho,vi=vi)
cari.2ll<-function(rho,vi){
  lambda<-cari.lambda(rho=rho,vi)
  prob<-cari.pi(lambda,rho=rho,vi)
  n<-length(vi)
  ll<-2*sum(log(1+lambda*vi))
  return(ll)
} ## end function

bisec.CI<-function(xx,yy,alpha1=0.95){
  chisquare=qchisq(alpha1,1)
  zx=scale(xx)
  zy=scale(yy)
  min<--1
  max<-1
  rata<-cor(xx,yy)
  delta<-10e-10

  ##Cari batas atas
  a<-rata
  b<-(rata+max)/2
  c<-max
```

```
  Delta=1
  i<-1
  while (Delta>delta & i<100){
    binit<-b
    rho=b
    vi=(zx*zy-rho)-(1/2)*rho*((zx^2-1)+(zy^2-1))
    llb<-cari.2ll(rho=b,vi=vi)
    data<-c(i,a,b,c,llb)
    i<-i+1
    if (llb<chisquare){
      a<-b
      b<-(a+c)/2
    } else
    {c=b
    b<-(a+c)/2
    }
    Delta<-abs(binit-b)/abs(b)
  }
  upper<-b
  llupper<-llb

  ##Cari batas bawah###
  a<-min
  b<-(rata+min)/2
  c<-rata
  Delta=1
  i=1
  while (Delta>delta & i<100){
    binit<-b
    rho=b
    vi=(zx*zy-rho)-(1/2)*rho*((zx^2-1)+(zy^2-1))
    llb<-cari.2ll(rho=b,vi=vi)
    data<-c(i,a,b,c,llb)
    i<-i+1
    if (llb<chisquare){
      c<-b
      b<-(a+c)/2
    } else
    {a=b
    b<-(a+c)/2
    }
    Delta<-abs(binit-b)/abs(b)
  }
  lower<-b
  lllower<-llb

return(list(Korelasi=cor(xx,yy),alpha1=alpha1,Lower=lower,Upper=upper,LL.Upper=llupper,L
L.Lower=lllower))
} ## end function
```