

Selang Kepercayaan Koefisien Korelasi Berdasarkan Empirical Likelihood dan Penerapannya pada Data Rata-Rata Lama Sekolah dan Penduduk Miskin Kota/Kabupaten di Indonesia

Anisa Nur Fitriani*

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

*anisanurf40@gmail.com

Abstract. The correlation coefficient is an analysis that studies the relationship between two variables. Hu, et al (2018) introduce the Empirical Likelihood method as an alternative that can be used to build a confidence interval for the correlation coefficient if the distribution nonnormal or unknown. In this paper, we apply the empirical probability method from Hu, et al (2018) in constructing a confidence interval for the Pearson correlation coefficient of Average Length of Schooling and Poor Population in Cities/Regencies in Indonesia in 2020. The results of the confidence interval for the average data years of schooling and poor population of Cities/Regencies in Indonesia in 2020 is $[-0.6231173; -0.4599864]$. That means there is a significant linear relationship between the Average Length of Schooling and the City/Regency Poor Population Rate in Indonesia in 2020.

Keywords: Correlation Coefficient, Confidence Interval, Z-Transformation Method, Empirical Likelihood Method.

Abstrak. Koefisien korelasi adalah suatu analisis yang mempelajari tentang hubungan antara dua variabel. Untuk membangun selang kepercayaan pada koefisien korelasi, yang biasa digunakan adalah Metode Z-Transformation dimana data harus berdistribusi normal, Hu, dkk (2018) memperkenalkan metode Empirical Likelihood sebagai salah satu alternatif yang dapat digunakan untuk membangun selang kepercayaan pada koefisien korelasi jika data tidak berdistribusi normal. Dalam makalah ini akan menerapkan metode empirical likelihood dari Hu, dkk (2018) dalam membangun selang kepercayaan untuk koefisien korelasi Pearson pada data Rata-rata Lama Sekolah dan Penduduk Miskin di Kota/Kabupaten di Indonesia tahun 2020. Hasil selang kepercayaan untuk data rata-rata lama sekolah dan penduduk miskin Kota/Kabupaten di Indonesia tahun 2020 adalah $[-0.6231173 ; -0.4599864]$. Dalam pengujian hipotesisnya disimpulkan bahwa ada hubungan linier yang signifikan antara Rata-rata Lama Sekolah dengan Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia tahun 2020.

Kata Kunci: Koefisien Korelasi, Selang Kepercayaan, Metode Z-Transformation, Metode Empirical Likelihood.

A. Pendahuluan

Koefisien korelasi adalah suatu analisis yang mempelajari tentang hubungan antara dua variabel. Dalam analisis, koefisien korelasi menentukan seberapa kuat hubungan antara dua variabel yang berada dalam interval $-1 \leq r \leq +1$. Koefisien korelasi memiliki sifat parametrik dan nonparametrik. Korelasi yang biasa digunakan adalah Korelasi Pearson, Korelasi Rank Spearman, dan Korelasi Tau Kendall. Fisher mengusulkan metode berbasis Z-Transformation untuk pengujian Korelasi Pearson dan digunakan untuk membuat selang kepercayaan koefisien korelasi. Hu, dkk (2018) mengusulkan metode Empirical Likelihood untuk membangun selang kepercayaan koefisien korelasi. Untuk metode empirical likelihood didasarkan pada data tanpa mengasumsikan data tersebut berasal dari distribusi tertentu. Meskipun metode empirical likelihood termasuk kategori nonparametrik, tetapi pendugaan selang kepercayaan dan pengujian hipotesis didasarkan atas metode maksimum likelihood.

Empirical Likelihood adalah metode nonparametrik yang sangat baik untuk membangun selang kepercayaan (Owen, 2001). Metode empirical likelihood juga tidak membutuhkan perhitungan statistik pivot dan tidak memerlukan informasi distribusi data. Untuk metode empirical likelihood didasarkan pada data tanpa mengasumsikan data tersebut berasal dari distribusi tertentu. Meskipun metode empirical likelihood termasuk kategori nonparametrik, tetapi pendugaan selang kepercayaan dan pengujian hipotesis didasarkan atas metode maksimum likelihood. Berdasarkan pendahuluan, dapat dilihat masalah yang dapat diambil adalah bagaimana menerapkan metode Empirical Likelihood untuk membangun selang kepercayaan untuk koefisien korelasi antara Rata-rata Lama Sekolah dan Penduduk miskin Kota/Kabupaten di Indonesia tahun 2020. Data tersebut merupakan data sukender yang diperoleh dari Badan Pusat Statistika (BPS).

B. Metodologi

Bahan

Data dan sumber data yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh dari Badan Pusat Statistik Indonesia (BPS) tahun 2020. Variabel pada data ini yang akan digunakan yaitu Rata-rata Lama Sekolah (X) dan Tingkat Penduduk Miskin (Y) dengan total keseluruhan adalah 548 Kota/Kabupaten di Indonesia.

Variabel Penelitian

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

Tabel 1. Variabel Penelitian

Variabel	Definisi
Rata-rata Lama Sekolah (X)	Tahun Rata-rata Lama Sekolah
Penduduk Miskin (Y)	Persentase Penduduk Miskin

Metode Analisis Penelitian

- Membuat deskripsi data Rata-rata Lama Sekolah dan penduduk miskin di Indonesia.
- Mencari nilai Batas Bawah (BB) dan Batas Atas (BA) dengan bantuan program Rstudio
Dalam mencari nilai BB dan BA selang berpatokan pada beberapa hal:
 - $-1 < \rho < 1$, sehingga $BB > -1$ dan $BA < +1$
 - Misalkan r adalah nilai dugaan korelasi, maka $BB < r < BA$, sehingga untuk mencari BB, pencarian bisa dilakukan diantara -1 dan r sebab $-1 < BB < r$. Demikian juga untuk mencari BA, pencarian dilakukan diantara r dan $+1$ sebab $r < BA < +1$.
 - BB dan BA adalah suatu nilai yang menyebabkan $\ell_I(BB) \leq \chi^2_{(1,0.95)} = 3.84$ dan $\ell_I(BA) \geq \chi^2_{(1,0.95)} = 3.84$ (persamaan 2.52)
 - Log Likelihood (ρ) akan maximum pada saat $\rho = r$, sehingga $\ell_I(\rho) =$

$-2 \log(\rho)$ akan minimum saat $\rho = r$

Berdasarkan hal tersebut, algoritma untuk mencari BB dan BA selang adalah sebagai berikut:

Hitung nilai dugaan koefisien korelasi pearson r menggunakan persamaan (2.1)

Mencari Batas Bawah (BB)

- (i) Tentukan nilai awal $a = -1; c = r$
- (ii) $BB = \frac{a+c}{2}$
- (iii) Dengan $\rho = BB$, hitung $\hat{V}_1(W_i, \rho)$
- (iv) Cari λ yang memenuhi persamaan (2.49) dan metode *newton Raphson* pada dengan menggunakan persamaan (2.50)
- (v) Hitung $-2\log$ untuk $\rho = BB$ yaitu $\ell_1(BB)$ menggunakan persamaan (2.51)
- (vi) Jika $\ell_1(BB) < 3.84$, set $c = BB$; jika $\ell_1(BB) > 3.84$, set $a = BB$
- (vii) Jika BB konvergen maka stop, selainnya kembali ke (ii)

Mencari BA

- (i) Tentukan nilai awal $a = r; c = +1$
- (ii) $BA = \frac{a+c}{2}$
- (iii) Dengan $\rho = BA$, hitung $\hat{V}_1(W_i, \rho)$
- (iv) Cari λ yang memenuhi persamaan (2.49) dan metode *newton Raphson* pada dengan menggunakan persamaan (2.50)
- (v) Hitung $-2\log$ untuk $\rho = BA$ yaitu $\ell_1(BA)$ menggunakan persamaan (2.51)
- (vi) Jika $\ell_1(BA) < 3.84$, set $c = BA$; jika $\ell_1(BA) > 3.84$, set $a = BA$
- (vii) Jika BA konvergen maka stop, selainnya kembali ke (ii)

3. Membuat Analisis dan Kesimpulan.

C. Hasil dan Pembahasan

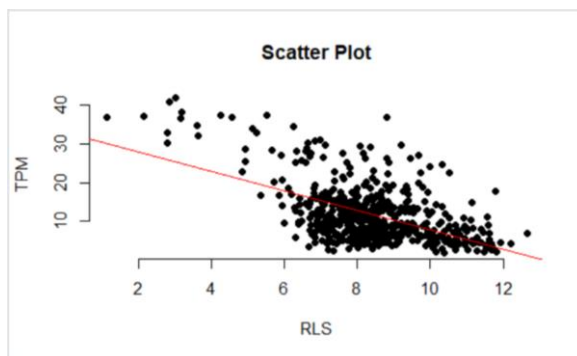
Deskripsi data

Pada penjelasan deskripsi data ini menggunakan analisis deskriptif untuk variabel X (Persentase Rata-rata Lama Sekolah) dan variabel Y (Persentase Tingkat Kemiskinan). Pada penelitian ini bertujuan mengetahui hubungan antara variabel X dan variabel Y. Dengan deskripsi data sebagai berikut:

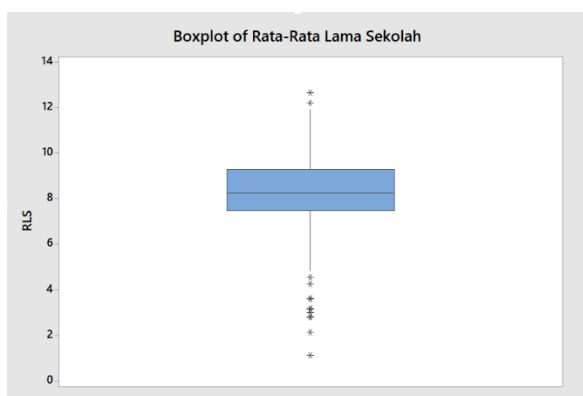
Tabel 2. Variabel Penelitian

Variabel	Rata-rata	Minimum	Maximum	StDev	N
X	8.361387	1.13	15.92	1.599267	548
Y	11.85956	2.02	41.76	7.370789	548

Dari hasil perhitungan diatas, rata-rata(mean) pada rata-rata lama sekolah di Kota/Kabupaten Indonesia (X) sebesar 8.36 tahun, dengan sebaran data sebesar 1.60 tahun, tingkat terendah pada Nduga Provinsi Papua sebesar 1.13 tahun, tingkat tertinggi pada Kota Makassar Provinsi Sulawesi Selatan sebesar 15.92 tahun. Sementara rata-rata pada tingkat kemiskinan di Kota/Kabupaten Indonesia (Y) sebesar 11.86%, dengan sebaran data sebesar 7.37%, tingkat terendah pada Kabupaten Badung, Provinsi Bali sebesar 2.02%, dengan tingkat tertinggi pada Kabupaten Deiyai, Provinsi Papua sebesar 41.76%.. Setelah mengetahui deskripsi variabel, data akan dideskripsikan dengan scatter plot Gambar 1



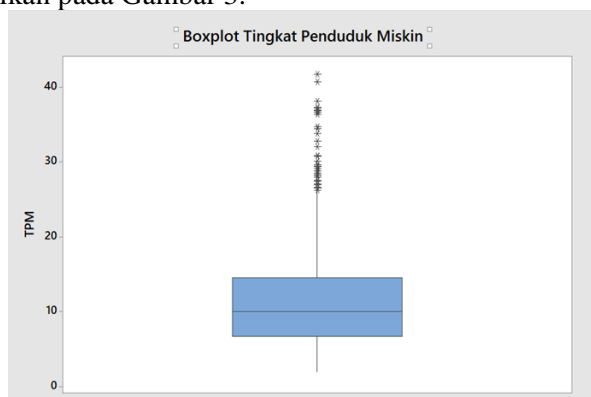
Gambar 1. Scatterplot



Gambar 2. Boxplot data Rata-rata Lama Sekolah di Kota/Kabupaten di Indonesia

Gambar 1 menunjukkan pola hubungan negatif, yaitu kecenderungan berlawanan arah dimana rata-rata lama sekolah naik, maka tingkat penduduk miskin Kota/Kabupaten di Indonesia turun, sebaliknya jika rata-rata lama sekolah turun, maka tingkat penduduk miskin Kota/Kabupaten di Indonesia naik. Gambar 2 menyajikan boxplot untuk Rata-rata Lama Sekolah di Kota/Kabupaten di Indonesia tahun 2020.

Dari gambar 2 diperoleh nilai Q1 sebesar 7.84, dengan nilai tengah 8.26 dan nilai Q3 sebesar 9.38. Boxplot diatas menunjukkan adanya outlier atau pencilan yang berada diatas garis dan dibawah garis. Data pencilan yang berada dibawah garis adalah berasal dari Provinsi Papua yaitu Nduga, Puncak, Pegunungan Bintang, Deiyai, Mamberamo Tengah, Puncak Jaya, Yahukimo, Paniai dengan nilai rata-rata lama sekolah yang rendah. Sedangkan data pencilan yang berada diatas garis adalah Kota Banda Aceh, Provinsi Aceh dengan nilai rata-rata lama sekolah yang tinggi. Boxplot untuk data tingkat penduduk miskin Kota/Kabupaten di Indonesia pada tahun 2020 disajikan pada Gambar 3.



Gambar 3. Boxplot data Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia

Dari gambar 3 diperoleh nilai Q1 pada variabel tingkat penduduk miskin sebesar 6.78, dengan nilai tengah atau median sebesar 10.05 dan nilai Q3 sebesar 14.57. Boxlot diatas menunjukkan banyak pencilan atau outlier yang berada di atas garis. Data outlier tersebut berasal dari beberapa Kota dari Provinsi di Indonesia dimana tidak semua disebutkan yaitu Provinsi Maluku (Kepulaun Aru, Maluku Barat Daya), Provinsi Nusa Tenggara Timur (Manggarai Timur, Timor Tengah Selatan, Sumba Barat, Sumba Timur), Provinsi Papua (Deiyai, Puncak, Pegunungan Bintang, Waropen), Provinsi Papua Barat (Manokwari Selatan, Maybarat, Wondama).

Korelasi Pearson

Korelasi Pearson adalah korelasi yang digunakan untuk mengetahui hubungan antara dua variabel. Korelasi pearson biasa disebut dengan korelasi product-moment. Korelasi pearson biasanya digunakan pada data skala interval dan rasio. Misal $(X_i, Y_i), i = 1, 2, \dots, n$ adalah nilai pengamatan (X, Y) , maka korelasi pearson dinyatakan sebagai berikut (Suliadi, 2015):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \dots (2.1)$$

Pada rumus (2.1) ini akan dicari nilai r dibantu dengan *Software R Studio* dengan menggunakan package (*cor.test*). Dari hasil output yang diperoleh pada lampiran dapat dilihat bahwa nilai r adalah -0.544819 . Atau dengan menggunakan rumus dari persamaan

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{-3512.9716}{\sqrt{1399.0379} \sqrt{29717.7038}} = -0.544819037$$

Kekuatan hubungan atau korelasi, menunjukkan bahwa Rata-rata Lama Sekolah dan Tingkat Kemiskinan di Kota/Kabupaten Indonesia memiliki kekuatan hubungan yang sedang dan arah hubungan yang negatif, artinya jika Rata-rata Lama Sekolah naik maka Tingkat Kemiskinan cenderung turun, dan sebaliknya.

Mencari Selang Kepercayaan Koefisien Korelasi berdasarkan Empirical Likelihood

Mencari nilai λ dengan metode Newtown Raphson yang digunakan untuk mencari selang kepercayaan, namun dalam mencari nilai λ sangat kompleks, sehingga proses mencari selang kepercayaan untuk nilai λ menggunakan syntax yang sudah tersedia pada software R Studio. Pada pencarian nilai selang kepercayaan empirical likelihood dimana nilai $\ell_1(\rho)$ akan konvergen ke $\chi^2_1(1 - \alpha)$ dengan $\alpha = 0.05$ dimana nilai $\chi^2_{1,0.95}$ diperoleh 3.84 (Owen, 2001).

Dari hasil perhitungan selang kepercayaan dengan metode *empirical likelihood* berbasis *influence function* untuk ρ (IFEL) pada output dengan bantuan package (*nleqslv, emplik, psychometric*) dan menggunakan metode bisection yaitu mencoba nilai $\rho = r$ ke arah kiri dan ke kanan, maka diperoleh nilai $\ell_1(\rho)$ dengan BA dan BB: $[3.84 ; 3.84]$ yang artinya tingkat kepercayaan sesuai dengan nilai $\chi^2_{1,0.95} = 3.84$, dengan selang kepercayaannya diperoleh: $[-0.6231173 ; -0.4599864]$ Artinya secara statistik dengan tingkat kepercayaan 95% hubungan rata-rata lama sekolah dan tingkat penduduk miskin Kota/Kabupaten di Indonesia pada tahun 2020 adalah antara -0.6231173 hingga -0.4599864 .

Uji Hipotesis:

$H_0 : \rho = 0$; Tidak ada hubungan linier yang signifikan antara Rata-rata Lama Sekolah dengan Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia tahun 2020.

$H_1 : \rho \neq 0$; Ada hubungan linier yang signifikan Rata-rata Lama Sekolah dengan Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia tahun 2020.

maka kesimpulannya adalah menolak H_0 , sebab selang kepercayaan yang diperoleh tidak mencakup nol, artinya ada hubungan linier yang signifikan antara Rata-rata Lama Sekolah dengan Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia tahun 2020.

D. Kesimpulan

Berdasarkan hasil pembahasan dan diskusi untuk selang kepercayaan koefisien korelasi

berdasarkan *Empirical Likelihood* berbasis *influence function* untuk ρ (IFEL) dan penerapannya pada data rata-rata lama sekolah dan penduduk miskin Kota/Kabupaten di Indonesia sehingga diperoleh selang kepercayaan dengan tingkat kepercayaan sebesar 95% antara -0.6231173 hingga -0.4599864 dengan nilai korelasi (r) sebesar -0.544819 yang artinya hubungan antara rata-rata lama sekolah dan tingkat penduduk miskin Kota/Kabupaten di Indonesia pada tahun 2020 berada diantara -0.6231173 hingga -0.4599864 . Dengan hasil uji hipotesisnya ada hubungan linier yang signifikan antara Rata-rata Lama Sekolah dengan Tingkat Penduduk Miskin Kota/Kabupaten di Indonesia tahun 2020.

Acknowledge

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

Daftar Pustaka

- [1] Hu, X. Jung, A. & Qin, G. (2018). Interval Estimation for the Correlation Coefficient, Department of Mathematics and Statistics.
- [2] Owen, A. B. (2001). Empirical likelihood. New York: Chapman & Hall/CRC.
- [3] Suliadi. (2015). *Analisis Regresi*. Diktat Kuliah Program Studi Statistika Unisba