

Uji Keberartian Koefisien *Raw Agreement*

MEGA ANISA RACHIM, TETI SOFIA YANTI, LISNUR WACHIDAH

Jurusan Statistika Universitas Islam Bandung

ABSTRAK

Dalam kehidupan sehari-hari seringkali terjadi dua penilai atau dua alat ukur dihadapkan pada suatu permasalahan untuk menyelesaikan, menentukan, mengukur dan mendiagnosis satu atau beberapa objek penilaian atau pengukuran ke dalam satu kategori dari beberapa kategori yang mungkin. Dari dua penilai atau dua alat ukur dapat diketahui tingkat kesesuaian antara dua penilai atau dua alat ukur tersebut. Dalam penulisan ini akan dikemukakan koefisien kesesuaian *raw agreement* beserta pengujiannya menggunakan dua metode yaitu menggunakan statistik Stouffer's *Z* dan Binomial eksak. *Raw Agreement* dapat mengukur kesesuaian antara dua penilai atau dua alat ukur dengan dua atau lebih kategori yang dinilai atau diukur. Sebagai aplikasi penulis menggunakan data dua alat diagnosis Appendisititis atau radang usus buntu pada pasien penderita nyeri perut kanan bawah di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung pada tahun 2008, penelitian Dr. Dono Pranoto, 2008. Setelah diuji, ternyata dua metode tersebut menghasilkan kesimpulan yang berbeda. Berdasarkan pengujian Stouffer's *Z* didapat hasil yaitu tidak ada kesesuaian yang signifikan antara dua alat diagnosis Appendisititis dan berdasarkan pengujian Binomial eksak untuk keberartian koefisien *ra* ada kesesuaian antara dua alat diagnosis Appendisititis.

Kata kunci: Koefisien raw agreement, statistik Stouffer's Z dan Binomial eksak

1. PENDAHULUAN

Dalam kehidupan sehari-hari seringkali terjadi dua penilai atau dua alat ukur dihadapkan pada suatu permasalahan untuk menyelesaikan, menentukan, mengukur dan mendiagnosis satu atau beberapa objek penilaian atau pengukuran ke dalam satu kategori dari beberapa kategori yang mungkin. Dua penilai atau dua alat ukur dapat diketahui tingkat kesesuaian antara dua penilai atau dua alat ukur tersebut. Pada kasus-kasus khusus, misalkan dua alat diagnosis dihadapkan pada masalah untuk mendiagnosis atau mengklasifikasikan beberapa objek penilaian (pasien) ke dalam satu kategori dari beberapa kategori yang mungkin. Berdasarkan gambaran diatas, sangat wajar apabila ingin diketahui ukuran kesesuaian antara dua penilai tersebut dalam mengklasifikasikan kategori.

Cohen (dalam Yan Firwan, 1999), mengemukakan suatu ukuran kesepakatan antara dua penilai dalam mengklasifikasikan beberapa subjek ke dalam satu dari dua kategori yang mungkin. Ukuran tersebut dikenal dengan statistik Cohen Kappa, yang mengasumsikan bahwa skala pengukurannya nominal. Ukuran statistik Cohen Kappa sudah biasa digunakan di bidang kesehatan. Nilai statistik Kappa yang mendekati satu menunjukkan kesepakatan yang sangat baik diantara kedua penilai sedangkan yang mendekati nol menunjukkan kesepakatan lemah diantara kedua penilai. Dalam pengukuran Cohen Kappa dapat menggunakan tabel kontingensi 2×2 berpasangan yang terdiri dari dua kategori misalkan sebuah jawaban ya dan tidak, sakit dan tidak sakit, dan sebagainya.

Dalam penulisan ini, akan membahas tentang alternatif lain dari pengukuran koefisien kesesuaian beserta pengujian hipotesisnya yang diperkenalkan oleh Alexander, Michael dan Patrick (2007) yaitu koefisien kesesuaian atau yang lebih dikenal *raw agreement (ra)*. Ukuran kesesuaian ini digunakan untuk mengukur dua penilai dengan kategori penilaian terdiri dari dua atau lebih kategori yang dinilai. Kelebihan *raw agreement* dibandingkan Cohen Kappa adalah dapat mengukur tingkat kesesuaian dengan dua atau lebih kategori yang dinilai, seluruh skala pengukuran dapat digunakan dan juga dapat dilakukan uji keberartian *raw agreement (ra)*. Tujuannya untuk mengetahui keberartian tingkat kesesuaian antara dua penilai atau dua alat ukur dengan dua atau lebih kategori penilaian atau pengukuran, maka dilakukan dua pengujian keberartian koefisien *ra* yaitu pengujian koefisien *ra* menggunakan statistik Stouffer's *Z* dan pengujian keberartian koefisien *ra* menggunakan Binomial eksak. Kedua pengujian tersebut dapat diaplikasikan salah satunya untuk mengetahui kesesuaian

dua alat diagnosis Appendisitis atau radang usus buntu pada pasien penderita nyeri perut kanan bawah di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung pada tahun 2008, penelitian Dr. Dono Pranoto, 2008.

2. TINJAUAN PUSTAKA

2.1 Koefisien *Raw Agreement*

Koefisien *raw agreement* (ra) merupakan pengukuran proporsi kesesuaian antara dua penilai dengan dua atau lebih kategori yang dinilai. Untuk melakukan analisis koefisien *raw agreement* (ra) ada beberapa asumsi yang harus dipenuhi yaitu:

- (1) Seluruh skala pengukuran dapat digunakan
- (2) Kategori yang dapat digunakan dua atau lebih kategori
- (3) Data berpasangan
- (4) Kedua penilai memiliki kualifikasi (kemampuan) yang sama

Tabel analisis yang digunakan berupa tabel *raw agreement* $I \times I$ untuk data berpasangan, sebagai berikut:

Tabel 1
Tabel *Raw Agreement* (ra) $I \times I$

		Penilai 1 Kategori			
		1	2	...	I
Penilai 2 Kategori	1	n_{11}	n_{12}	...	$n_{1.}$
	2	n_{21}	n_{22}	...	$n_{2.}$
	⋮
	I	$n_{.1}$	$n_{.2}$...	n

Dari data yang ada pada tabel tersebut dapat dilakukan penghitungan koefisien ra , adapun rumus yang digunakan sebagai berikut:

$$ra = \frac{1}{n} \sum_{i=1}^I n_{ii} \quad (1)$$

dengan:

n_{ii} = frekuensi pada sel i, i

n = banyaknya pengamatan

I = banyaknya kategori penilaian

Batasan nilai *raw agreement* (ra) adalah $0 \leq ra \leq 1$, andaikan terdapat kesesuaian yang sempurna, proporsi kesesuaian akan sama dengan satu sehingga $ra = 1$. Sebaliknya, apabila sama sekali tidak terdapat kesesuaian atau tampak semata-mata karena kebetulan, proporsi kesesuaian akan sama dengan nol sehingga $ra = 0$.

2.2 Pengujian Keberartian Koefisien ra Menggunakan Statistik Stouffer's Z

Distribusi Uniform merupakan distribusi peluang diskrit yang paling sederhana dimana setiap kejadian memiliki peluang yang sama, maka distribusi Uniform mempunyai persamaan sebagai berikut :

$$P(X = x) = \frac{1}{k} ; x = x_1, x_2, \dots, x_k.$$

Dengan menggunakan distribusi Uniform, didapat masing-masing sel dari tabel *raw agreement*

memiliki probabilitas yang sama yaitu $\frac{1}{I}$, tetapi yang diperhatikan adalah kedua penilai atau

dua alat ukur yang mempunyai kesesuaian yang sama (jumlah sel diagonal utama dari tabel *raw agreement*), jadi probabilitas untuk diagonal utama tabel *raw agreement* adalah

$I\left(\frac{1}{I^2}\right) = \frac{1}{I}$ dan masing-masing sel pada diagonal utama pada tabel *raw agreement* memiliki probabilitas $\frac{1}{I}$.

Sehingga perumusan hipotesisnya sebagai berikut :

$H_0 : ra = \frac{1}{I}$; Tidak ada kesesuaian antara penilai atau alat ukur 1 dan penilai atau alat ukur 2

$H_1 : ra > \frac{1}{I}$; Ada kesesuaian antara penilai atau alat ukur 1 dan penilai atau alat ukur 2

Statistik uji yang digunakan yaitu statistik Stouffer's Z yang mengikuti distribusi Normal Baku dengan rumus sebagai berikut :

$$Z = \frac{1}{\sqrt{I}} \sum_i z_{ii} ; \quad i = 1, 2, \dots, I \tag{2}$$

dengan :

$$z_{ii} = \frac{n_{ii} - \frac{n}{I^2}}{\sqrt{\frac{n}{I^2}}} \tag{3}$$

n_{ii} = frekuensi pada sel i, i

n = banyaknya pengamatan

I = banyaknya kategori penilaian

Dengan kriteria uji, menggunakan daftar distribusi normal baku, H_0 ditolak jika nilai $Z \geq Z_{0,5-\alpha}$, dalam hal lainnya H_0 diterima atau H_0 ditolak jika $p\text{-value} \leq \alpha$.

$P\text{-value}$ adalah nilai peluang untuk menolak H_0 yang seharusnya diterima, dengan rumus sebagai berikut :

$$P(z \geq Z) = 0,5 - \alpha \tag{4}$$

Peluang yang biasa digunakan $p \leq 0,05$, atau $p \leq 0,01$.

2.3 Pengujian Keberartian Koefisien ra Menggunakan Binomial Eksak

Jika jumlah pengamatan yang diteliti kecil ($n \leq 30$). Distribusi binomial dinyatakan dengan $b(x; n, p)$. Suatu percobaan bernoulli yang dapat menghasilkan "sesuai" dengan probabilitas p dan "tidak sesuai" dengan probabilitas $q = 1 - p$ maka distribusi probabilitas variabel acak binomial X dengan jumlah sesuai dalam n percobaan independen adalah

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, x=0, 1, 2, \dots, n$$

Sehingga peluang kesesuaian satu sisi untuk jumlah total pengamatan sebagai berikut:

$$P = \sum_{j=\sum n_{ii}}^n \binom{n}{j} p^j q^{n-j} \tag{5}$$

dengan :

n = banyaknya pengamatan

j = frekuensi pada sel ii

p = probabilitas yang menyatakan "sesuai" yang ditaksir dengan $\frac{1}{I}$

$q = 1 - p$

Prosedur untuk pengujian hipotesis akan diuraikan sebagai berikut :

Perumusan hipotesis yang digunakan sama seperti hipotesis untuk pengujian statistik Stouffer's Z.

Statistik Uji yang digunakan yaitu :

- Jika jumlah pengamatan yang diteliti kecil ($n \leq 30$), maka p -value yang digunakan yaitu pada Persamaan (6).
- Ketika jumlah pengamatan besar ($n > 30$), maka dilakukan pengujian keberartian koefisien ra menggunakan pengujian Binomial dengan pendekatan distribusi Normal Baku.

Statistik uji yang digunakan mengikuti distribusi Normal Baku dengan rumus sebagai berikut :

$$Z_{bin} = \frac{\sum_i n_{ii} - np}{\sqrt{npq}} \quad (6)$$

dengan :

n_{ii} = frekuensi pada sel ii

n = banyaknya pengamatan

p = probabilitas yang menyatakan "sesuai" yang ditaksir dengan $\frac{1}{I}$

$q = 1 - p$

Dengan kriteria uji, yaitu :

- H_0 ditolak jika p -value $\leq \alpha$. Dalam hal lainnya, H_0 diterima.
- Dengan menggunakan daftar distribusi normal baku, H_0 ditolak jika nilai $Z \geq Z_{0,5-\alpha}$, dalam hal lainnya H_0 diterima atau H_0 ditolak jika p -value $\leq \alpha$.

3. APLIKASI

Appendisitis akut merupakan suatu keadaan akut abdomen yang paling sering ditemukan memerlukan tindakan bedah sesegera mungkin. Dari keseluruhan kasus akut abdomen yang memerlukan tindakan bedah, salah satunya adalah pada appendisitis akut. Diagnosis pada pasien dengan suspek appendisitis akut dilakukan berdasarkan pemeriksaan fisik diagnostik, laboratorium, dan beberapa pemeriksaan penunjang seperti foto polos abdomen, USG, dan CT scan. Terdapat penelitian mengenai pendekatan diagnosis appendisitis akut untuk lebih akurat. Penelitian tersebut dilakukan mulai dari gejala klinik, pemeriksaan laboratorium yang khas, sampai dengan pemeriksaan penunjang. Beberapa peneliti kemudian memberikan skoring terhadap gejala klinik dan pemeriksaan laboratorium yang ditemukan. Berdasarkan hal ini maka berkembanglah berbagai macam skor untuk menilai dan membantu klinisi dalam mendiagnosis appendisitis akut. Terdapat dua alat diagnosis appendisitis atau radang usus buntu pada penderita sakit nyeri perut kanan bawah, yaitu sistem skoring Alvarado dan sistem skoring Tzanakis, dimana kedua sistem skoring tersebut mempunyai sensitifitas diatas 90% (dalam Dr. Dono Pranoto,2007).

Tabel 2

Tabel *Raw Agreement* 2 x 2 Berpasangan Dua Alat Ukur Tentang Radang Usus Buntu Pada 64 Penderita Sakit Nyeri Perut Kanan Bawah di RSHS Bandung Tahun 2008

		Alat Ukur 1 (Alvarado)		Jumlah
		Akut	Tidak Akut	
Alat Ukur 2 (Tzanakis)	Akut	41	7	48
	Tidak Akut	16	0	16
Jumlah		57	7	64

Sumber: *Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung Tahun 2008*

Objek pengamatannya adalah penderita yang mengeluh nyeri perut kanan bawah yang dirawat dibagian bedah unit gawat darurat, dan penderita yang dikonsulkan dari bagian lain dengan keluhan yang serupa di Rumah Sakit Hasan Sadikin Bandung tahun 2008, dalam penelitian Dr. Dono Pranoto,2008. Banyaknya pasien yang diamati sebanyak 64 pasien penderita yang

mengeluh nyeri perut kanan bawah pada kelompok umur < 21 Tahun, 21 - 40 Tahun, dan > 40 Tahun. Pada pengamatan yang dilakukan ke – 64 pasien penderita yang mengeluh nyeri perut kanan bawah diperiksa dengan menggunakan dua alat diagnosis appendisititis atau radang usus buntu yaitu sistem skoring Alvarado dan sistem skoring Tzanakis. Agar dapat diketahui kesesuaian antara dua alat diagnosis Appendisititis. Data yang diperoleh disajikan dalam tabel *raw agreement 2x2* dengan dua kategori penilaian, seperti pada Tabel 2.

Berdasarkan penghitungan koefisien *raw agreement*, dan pengujian keberartian dengan menggunakan statistik Stouffer's Z dan pengujian Binomial, dengan menggunakan persamaan (1), (2), (3), (4) dan (6) maka didapat hasil pada Tabel 3 sebagai berikut:

Tabel 3
Hasil Penghitungan dan Pengujian *Raw Agreement (ra)*

Koefisien <i>raw agreement (ra)</i>	0,6406	
	Metode Pengujian	
	Statistik Stouffer's Z	Binomial Eksak (n > 30)
Statistik Uji	1,5909	2,25
<i>P-value</i>	0,0559	0,0122
Kriteria Uji ($\alpha = 5\%$)	Hipotesis nol diterima	Hipotesis nol ditolak
Kriteria Uji ($\alpha = 10\%$)	Hipotesis nol ditolak	Hipotesis nol ditolak

4. KESIMPULAN

Adapun kesimpulan yang dapat dikemukakan dalam penulisan ini adalah:

1. Berdasarkan data pasien mengenai nyeri perut kanan bawah yang dirawat di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung, penghitungan koefisien *raw agreement (ra)* diperoleh koefisien *raw agreement* sebesar 0,6406 yang artinya sistem skoring Alvarado dan sistem skoring Tzanakis dalam mendiagnosis radang usus buntu terhadap pasien penderita nyeri perut kanan bawah memiliki kesesuaian sebesar 64,06%.
2. Berdasarkan pengujian keberartian koefisien *raw agreement (ra)* dengan menggunakan statistik Stouffer's Z, didapat hasil yaitu dengan menggunakan $\alpha = 5\%$ dikatakan bahwa tidak ada kesesuaian yang signifikan antara sistem skoring Alvarado dan sistem skoring Tzanakis dalam mendiagnosis Radang Usus Buntu terhadap pasien penderita nyeri perut kanan bawah yang dirawat di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung. Dalam pengertian lain, sistem skoring Alvarado dan sistem skoring Tzanakis tidak dapat saling menggantikan atau dapat digunakan keduanya.
3. Berdasarkan pengujian keberartian koefisien *raw agreement (ra)* dengan menggunakan Binomial dengan jumlah pengamatan lebih dari 30, didapat hasil yaitu dengan menggunakan $\alpha = 5\%$ dapat dikatakan bahwa ada kesesuaian yang signifikan antara sistem skoring Alvarado dan sistem skoring Tzanakis dalam mendiagnosis radang usus buntu terhadap pasien penderita nyeri perut kanan bawah yang dirawat di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung. Dalam pengertian lain, skoring Alvarado dan sistem skoring Tzanakis dapat digunakan salah satu karena kedua alat tersebut memiliki kesesuaian.
4. Dalam data pengamatan diagnosis radang usus buntu terhadap pasien penderita nyeri perut kanan bawah yang dirawat di Bagian Bedah Unit Gawat Darurat di Rumah Sakit Hasan Sadikin Bandung dengan menggunakan sistem skoring Alvarado dan sistem skoring Tzanakis, memiliki hasil pengujian yang berbeda. Dimana pengujian Binomial eksak dengan menggunakan $\alpha = 5\%$, memberikan hasil *p-value* lebih kecil dari pengujian Stouffer's Z yaitu 0,0122. Untuk itu dalam kasus ini jika menggunakan $\alpha = 5\%$ lebih baik menggunakan pengujian Binomial eksak.

5. Jika menggunakan $\alpha = 10\%$, kedua metode pengujian sistem skoring Alvarado dan sistem skoring Tzanakis tidak memberikan hasil yang berbeda. Oleh karena itu, pasien dapat memilih pemeriksaan dengan menggunakan sistem skoring Alvarado atau pemeriksaan dengan menggunakan sistem skoring Tzanakis tetapi pemeriksaan dengan menggunakan sistem skoring Alvarado lebih terjangkau biayanya karena menggunakan alat laboratorium dibandingkan dengan sistem skoring Tzanakis menggunakan alat ultrasound yang lebih tinggi biayanya.

DAFTAR PUSTAKA

- [1]. Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20: 37 – 46.
- [2]. Firwan, Yan. (1999). *Koefisien Kesepakatan dan Model Trinomial Untuk Mengukur Kesepakatan Dua Orang Penilai*, Skripsi. Jurusan Statistika, FMIPA, Unisba.
- [3]. M. Parker, Philip. (2009). “*Specialty Definition: P-value*”. Diakses tanggal 17 Juli 2009 pada pukul 14.20 WIB, dari <http://www.websters-online-ictionary.org/P-/P-value.html>
- [4]. Pranoto, Dono. (2007). *Perbandingan Sistem Skoring Berdasarkan Tzanakis Dengan Sistem Skoring Berdasarkan Alvarado Untuk Menilai Keakuratan Dalam Mendiagnosis Appendicitis Akut Di Rumah Sakit Dr. Hasan Sadikin Bandung*, Usulan Penelitian. Program Pendidikan Dokter Spesialis I, Fakultas Kedokteran, Unpad.
- [5]. *Sistem skor Alvarado*. (2009). Diakses tanggal 15 Juli 2009 pada pukul 07.51 WIB, dari <http://www.bedahugm.net/Bedah-Digesti/Appendicitis-akut.html>
- [6]. *Sistem skoring dalam radang usus buntu (Appendicitis)*. 2009. Diakses tanggal 15 Juli 2009 pada pukul 07.44 WIB, dari <http://www.abdopain.com/diagnosis-of-appendicitis.html>
- [7]. Sudjana. (1996). *Metode Statistika*. Penerbit Tarsito. Bandung.
- [8]. Von Eye, A., Schauerhuber, M., and Mair, P. (2007). “*Significance Tests for the Measure of Raw Agreement*”.