

BAB 2

MENGESTIMASI RELIABILITAS

What good is consistency, if you're consistently wrong? (pepatah Inggris)

A. RELIABILITAS TES ULANG (TEST-RETEST)

Reliabilitas tes ulang adalah menguji keandalan instrumen pengukuran/hasil pengukuran yang didapatkan dari pengukuran secara berulang. Setiap subjek mendapatkan tes yang sama sebanyak dua kali. Estimasi reliabilitas ini dilakukan dengan cara mengkorelasikan hasil pengukuran pertama dan kedua.

1. CONTOH DESAIN PENELITIAN

Ahern (2004) melakukan penelitian untuk mengidentifikasi properti psikometris *Parent-Child Dysfunctional Interaction Scale*. Untuk mendapatkan reliabilitas hasil pengukurannya, ia menggunakan teknik tes ulang yaitu dengan memberikan skala tersebut kepada subjek dengan menggunakan tes ulang kepada 185 orang dengan interval waktu 6 bulan.

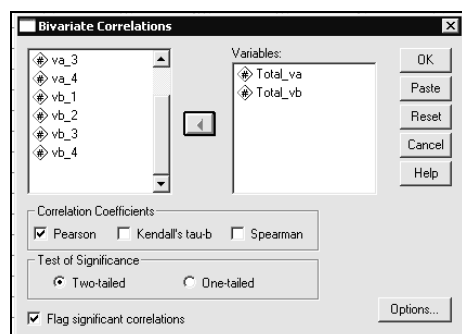
2. LATIHAN

Sebuah tes pemahaman visual diberikan sebanyak dua kali kepada subjek penelitian. Peneliti hendak melihat seberapa jauh reliabilitas alat ukur yang dibuatnya. Kutipan data yang hendak dianalisis dapat dilihat pada tabel berikut ini. Variabel Va_1, Va_2..dst adalah hasil pengukuran pertama. Total_a adalah skor total pengukuran pertama. Vb_1, Vb_2..dst adalah hasil pengukuran kedua. Total_b adalah skor total pengukuran kedua.

No	Va 1	Va 2	Va 3	Va 4	Total a	Vb 1	Vb 2	Vb 3	Vb 4	Total b
1	4	4	4	4	16	4	4	4	4	16
2	5	5	4	2	16	3	3	5	5	16
3	4	4	4	4	16	4	5	4	4	17
4	5	5	5	5	20	4	5	5	4	18

Mari kita menganalisisnya.

- Buka file kerja : “Tes Retest.Sav”
- Klik menu *Analyze* → *Correlation* → *Bivariate*
- Masukkan variabel “total_a” dan “total_b” pada kotak *variable(s)*. Tekan OK!



3. OUTPUT DAN INTERPRETASI

Correlations			
		Total_va	Total_vb
Total_va	Pearson Correlation	1	.919**
	Sig. (2-tailed)		.000
	N	19	19
Total_vb	Pearson Correlation	.919**	1
	Sig. (2-tailed)	.000	
	N	19	19

** . Correlation is significant at the 0.01 level

Hasil analisis menunjukkan bahwa terdapat hubungan yang signifikan antara dua hasil pengukuran pertama dan kedua ($r = 0.919$; $p < 0.001$). Hal ini menunjukkan bahwa hasil pengukuran sangat stabil dari waktu ke waktu (reliabel). Analisis reliabilitas hanya memperhatikan nilai r saja tanpa melihat signifikansi hubungan. Jadi meskipun signifikansi hubungan ini besar ($p < 0.001$), tidak dilaporkan dalam sub bab reliabilitas hasil pengukuran.

4. RUMUS MANUAL

Rumus mencari reliabilitas tes paralel adalah rumus korelasi *product moment* dengan rumus sebagai berikut.

$$r_{xx'} = \frac{S_{XY}}{S_X S_Y}$$

Keterangan

S_{XY} = kovarian antar tes

S_X dan S_Y = deviasi standar tes

Dengan mengoperasikan menu DESCRIPTIVE STATISTICS - DESCRIPTIVE - OPTION kita dapat menemukan nilai deviasi standar masing-masing tes dan dengan mengoperasikan menu CORRELATE - BIVARIATE - OPTION kita dapat menemukan kovarian antar tes.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Total_va	19	7	20	15.05	3.674
Total_vb	19	7	20	15.74	3.970
Valid N (listwise)	19				

Correlations			
		Total_va	Total_vb
Total_va	Pearson Correlation	1	.919**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	242.947	241.263
	Covariance	13.497	13.404
	N	19	19
Total_vb	Pearson Correlation	.919**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	241.263	283.684
	Covariance	13.404	15.760
	N	19	19

** . Correlation is significant at the 0.01 level (2-tailed).

Dari pengoperasian melalui SPSS diketahui kovarian antar tes ($S_{XY} = 13.40$) dan deviasi standar tiap tes ($S_X = 3.674$ dan $S_Y = 3.970$) sehingga nilai reliabilitas yang didapatkan adalah sebesar $r_{xx'} = S_{XY} / (S_X S_Y) = 13.40 / (3.674 * 3.970) = 0.919$

5. PENULISAN LAPORAN PENELITIAN

- Melalui analisis korelasi *product moment* untuk mengestimasi reliabilitas *test-retest* dengan waktu tenggang 2 minggu, didapatkan reliabilitas instrumen pemahaman visual adalah tinggi ($r = 0.919$)
- *In pilot work for the current study subscales of this Parent Report Scale (PRS) manifested relatively strong test-retest reliability estimates (ranging from .71 to .80 over a 4-month interval) (Cole et al. (1996)*

B. RELIABILITAS TES PARALEL

Tes paralel adalah dua buah tes yang disusun dari tujuan ukur yang sama dan blue print yang sama sehingga jumlah item atau taraf kesukarannya untuk tes kognif adalah sama. Secara statistik, tes paralel diasumsikan dengan nilai varian skor tampak dan varian eror yang setara (*parallel*) atau ditambahkan dengan asumsi rerata skor yang setara (*strict parallel*) antar kedua tes.

1. CONTOH DESAIN PENELITIAN

Seorang peneliti hendak melakukan eksperimen mengenai efektivitas penggunaan cerita bergambar untuk meningkatkan penalaran moral remaja dengan menggunakan *pre-test and post-test design*. Untuk mengurangi terjadinya efek pembelajaran ketika subjek dikenakan *post-test* maka peneliti menggunakan dua buah tes yang paralel. Satu diberikan ketika *pre-test* dan satu diberikan ketika *post-test*. Untuk menguji reliabilitasnya peneliti menggunakan teknik estimasi reliabilitas tes paralel.

2. LATIHAN

Sebuah skala hubungan interpersonal berbentuk likert disusun paralel (Form A dan Form B). Masing-masing form terdiri dari 5 aitem. Peneliti hendak melihat koefisien reliabilitas alat ukur yang dibuatnya. Kutipan data yang hendak dianalisis dapat dilihat pada tabel berikut ini. Variabel *a_1*, *a_2*..dst adalah hasil pengukuran pertama. Total_a adalah skor total Skala form A. *b_1*, *b_2*..dst adalah hasil pengukuran kedua. Total_b adalah skor total dari skala form B. Cara mengestimasi reliabilitas tes paralel adalah sama dengan test-retest yaitu dengan cara mengkorelasikan skor total masing-masing form.

No	a_1	a_2	a_3	a_4	a_5	Tot a	b_1	b_2	b_3	b_4	b_5	Tot b
1	4	4	2	2	2	14	2	2	3	4	4	4
2	4	4	2	2	2	14	2	2	2	4	4	4
3	2	4	2	2	2	12	2	4	4	2	2	2
4	2	4	2	2	1	11	2	2	2	3	3	3

Mari kita menganalisisnya.

- Buka file kerja : “Tes Paralel.Sav”
- Klik menu *Analyze* → *Correlation* → *Bivariate*
- Masukkan variabel “total_a” dan “total_b” dalam kotak *variable(s)*. Tekan OK!

3. OUTPUT DAN INTERPRETASI

Correlations		
	Total_a	Total_b
Total_a		
Pearson Correlation	1	,897**
Sig. (2-tailed)		,000
N	19	19
Total_b		
Pearson Correlation	,897**	1
Sig. (2-tailed)	,000	
N	19	19

** . Correlation is significant at the 0.01 level

Hasil analisis melalui korelasi *product moment* menunjukkan bahwa terdapat hubungan yang signifikan antara Form A dan Form B ($r = 0.897$; $p < 0.001$). Hal ini menunjukkan bahwa kedua tes adalah setara (*equivalent*). Sama seperti pada tes-retest analisis reliabilitas hanya memperhatikan nilai r saja tanpa melihat signifikansi hubungan.

4. RUMUS MANUAL

Rumus mencari reliabilitas tes paralel adalah rumus korelasi *product moment* dengan rumus sebagai berikut.

$$r_{xx'} = \frac{S_{XY}}{S_X S_Y}$$

Keterangan

S_{XY} = kovarian antar tes

S_X dan S_Y = deviasi standar tes

Dengan mengoperasikan menu DESCRIPTIVE STATISTICS - DESCRIPTIVE - OPTION kita dapat menemukan nilai deviasi standar masing-masing tes dan dengan mengoperasikan menu CORRELATE - BIVARIATE - OPTION kita dapat menemukan kovarian antar tes.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Total_a	19	6	25	15.58	4.325
Total_b	19	7	25	16.00	4.269
Valid N (listwise)	19				

Correlations

		Total a	Total b
Total_a	Pearson Correlation	1	.897**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	336.632	298.000
	Covariance	18.702	16.556
	N	19	19
Total_b	Pearson Correlation	.897**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	298.000	328.000
	Covariance	16.556	18.222
	N	19	19

** . Correlation is significant at the 0.01 level

Dari pengoperasian melalui SPSS diketahui kovarian antar tes ($S_{XY}=16.556$) dan deviasi standar tiap tes ($S_X=4.325$ dan $S_Y= 4.269$) sehingga nilai reliabilitas yang didapatkan adalah sebesar $r_{xx'} = S_{XY} / (S_X S_Y) = 16.556 / (4.325 * 4.269) = 0.897$

5. PENULISAN LAPORAN PENELITIAN

Melalui analisis korelasi *product moment* didapatkan bahwa reliabilitas skala hubungan interpersonal yang diketahui melalui koefisien ekuivalensi instrumen pemahaman visual adalah tinggi ($r = 0.897$)

D. RELIABILITAS KONSISTENSI INTERNAL

Reliabilitas konsistensi internal adalah reliabilitas yang didapatkan dari pengujian konsistensi antar aitem (antar bagian) menunjukkan konsistensi respon subjek dalam satu alat ukur pada satu kali pengukuran. Koefisien reliabilitas ini dapat langsung diketahui meskipun tes diberikan satu kali saja. Cara yang biasa dipakai adalah melihat hubungan antar belahan pada tes. Dengan membelah tes menjadi beberapa bagian kemudian membandingkan hubungan antar bagian. SPSS menyediakan dua jenis model yaitu model 1 yang tidak melibatkan matriks kovarian dalam analisisnya dan model 2 yang melibatkan matriks kovarian dalam analisisnya. Estimasi reliabilitas dengan teknik konsistensi internal dilakukan dengan melakukan pembelahan item menjadi beberapa belahan. SPSS

menyediakan dua jenis pembelahan, pertama pembelahan 2 bagian (pembelahan panjang setara maupun pembelahan panjang yang tidak setara) misalnya Formula Spearman-Brown serta pembelahan sejumlah itemnya misalnya Formula Alpha.

1.A. CONTOH DESAIN PENELITIAN

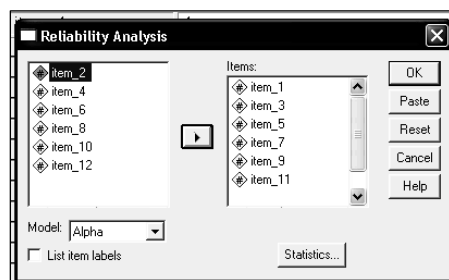
Wijatmiko (2001) hendak menguji alat ukur yang kemandirian yang disusunnya sendiri dengan menggunakan teknik konsistensi internal dengan menggunakan Formula Alpha. Peneliti memberikan alat ukur kepada subjek yang berjumlah 100 orang dan menganalisisnya kemudian mendapatkan bahwa nilai reliabilitas yang didapakkannya memuaskan ($r_{xx}=0.911$)

2.A LATIHAN 1 (KOEFSISIEN ALPHA)

Sebuah skala harga diri berbentuk Likert 4-pilihan (0 sampai 3) yang berjumlah 10 aitem hendak dianalisis reliabilitasnya. Peneliti hendak menguji reliabilitas alat ukur ini dengan menggunakan teknik Alpha Cronbach.

No	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
1	1	1	2	2	1	0	1	1	2	2	2	2
2	0	0	2	1	0	0	2	2	1	2	0	1
3	0	0	2	2	0	0	1	0	2	2	0	0
4	2	2	2	2	1	0	0	2	2	2	1	0

Mari kita menganalisisnya.



- Buka file kerja : “Split Half.Sav”
- Klik menu ANALYZE → SCALE → RELIABILITY ANALYSIS
- Masukkan semua aitem ke dalam kotak ITEMS
- Pilih Model pilih model ALPHA di kotak bagian bawah.
- Tekan menu STATISTICS lalu pilih menu
- Tekan OK!

3.A. OUTPUT DAN INTERPRETASI

Reliability Statistics

Cronbach's Alpha	N of Items
.813	12

Hasil analisis dengan menggunakan teknik *alpha cronbach* didapatkan koefisien reliabilitas alat ukur cukup tinggi yaitu $r_{xx'} = 0.813$. Hal ini menunjukkan bahwa alat ukur tersebut dapat terpercaya.

- Alpha < 0.7 : kurang meyakinkan (inadequate)
- Alpha ≥ 0.7 : baik (*good*)
- Alpha ≥ 0.8 : istimewa (*excellent*) (Nunally, 1978)

4.A. RUMUS MENGHITUNG

Rumus menghitung formula alpha adalah sebagai berikut.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum S_{Yi}^2}{S_{X-tot}} \right)$$

Keterangan

k = jumlah item/belahan

$\sum S_{Yi}^2$ = jumlah varian belahan dalam tes

S_{X-tot} = varian skor total (Cronbach, 1951)

Untuk mencari nilai reliabilitas alpha terlebih dahulu varian kita mencari total varian antar tes ($\sum S_{Yi}^2$). Melalui menu DESCRIPTIVE STATISTICS - DESCRIPTIVE - OPTION kita mendapatkan varian masing-masing item. Varian-varian tersebut kemudian kita jumlahkan.

Descriptive Statistics		
	N	Variance
item_1	10	.767
item_2	10	.844
item_3	10	.489
item_4	10	.500
item_5	10	.544
item_6	10	.889
item_7	10	.544
item_8	10	.544
item_9	10	.622
item_10	10	.844
item_11	10	.900
item_12	10	.711
Valid N (listwise)	10	

Total varian semua item 8.200
($\sum S_{Yi}^2 = 8.200$)

Descriptive Statistics		
	N	Variance
total	10	32.178
Valid N (listwise)	10	

Varian skor total 32.178
($S_{X-tot} = 32.178$)

Dengan mengetahui kedua nilai di atas maka nilai reliabilitas alpha dapat dihitung.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum S_{Yi}^2}{S_{X-tot}} \right) = \frac{12}{11} \times \left(1 - \frac{8.20}{32.178} \right) = 0.813$$

5.A PENULISAN LAPORAN PENELITIAN

- Dengan menggunakan teknik *alpha cronbach* didapatkan bahwa koefisien reliabilitasnya adalah $r_{xx'} = 0.897$
- For the present sample of married Chinese Americans, coefficient alphas for scores on the Support and Conflict scales were moderately high: .87 and .86 for spouse, .88 and .85 for family, and .89 and .78 for friends, respectively (Hwang et al., 2000)

1.B. LATIHAN 2 (KOEFSIEN SPEARMAN-BROWN)

Sebuah skala harga diri berbentuk Likert 4-pilihan (0 sampai 3) yang berjumlah 10 aitem hendak dianalisis reliabilitasnya. Peneliti hendak menguji reliabilitas alat ukur ini dengan menggunakan teknik belah dua (*split half*) dari Spearman-Brown.

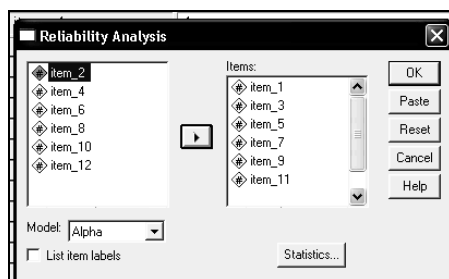
2.B. LATIHAN 2 (KOEFSIEN SPEARMAN-BROWN)

Sebuah skala harga diri berbentuk Likert 4-pilihan (0 sampai 3) yang berjumlah 10 aitem hendak dianalisis reliabilitasnya. Peneliti hendak menguji reliabilitas alat ukur ini dengan menggunakan teknik belah dua (*split half*) dari Spearman-Brown.

No	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
1	1	1	2	2	1	0	1	1	2	2	2	2
2	0	0	2	1	0	0	2	2	1	2	0	1
3	0	0	2	2	0	0	1	0	2	2	0	0
4	2	2	2	2	1	0	0	2	2	2	1	0

Mari kita menganalisisnya.

- Buka file kerja : “Split Half.Sav”
- Sebelum menganalisis, tentukan dulu dengan cara apa kita membagi alat tes mejadi dua bagian. Cara yang kita pilih kali ini adalah ganjil-genap
- Klik menu *Analyze* → *Scale* → *Reliability Analysis*
- Masukkan aitem yang bernomor ganjil (1,3,5,7,9,11) ke dalam kotak *Items*, kemudian disusun dengan memasukkan aitem bernomor genap (2,4,6,8,10,12).
- Pilih Model *Split Half* di kotak bagian bawah. Tekan OK!



2.B. OUTPUT DAN INTERPRETASI

Reliability Statistics			
Cronbach's Alpha	Part 1	Value	.623
		N of Items	6 ^a
	Part 2	Value	.581
		N of Items	6 ^b
Total N of Items			12
Correlation Between Forms			.957
Spearman-Brown Coefficient	Equal Length		.978
	Unequal Length		.978
Guttman Split-Half Coefficient			.978

a. The items are: item_1, item_3, item_5, item_7, item_9, item_11.

b. The items are: item_2, item_4, item_6, item_8, item_10, item_12.

- *Cronbach Alpha*. Menunjukkan reliabilitas tiap bagian yang dihitung dengan menggunakan teknik *Alpha Cronbach*. Terlihat pada output bahwa reliabilitas belahan 1 (ganjil) adalah $r_{xx'} = 0.623$ dan reliabilitas belahan 2 (genap) adalah $r_{xx'} = 0.581$.
- *Correlation Between Forms*. Adalah korelasi antar belahan yang dihitung dengan menggunakan korelasi product momen. Terlihat bahwa korelasi antar bagian tes $r_{xy} = 0.957$ dan nilai reliabilitas yang dihasilkan $r_{xx'} = 0.978$
- *Spearman Brown Coefficient*. Menunjukkan reliabilitas alat ukur yang diestimasi dengan menggunakan teknik belah dua, yaitu $r_{xx'} = 0.978$. Karena jumlah item antara belahan 1 dan 2 adalah sama maka yang kita lihat adalah EQUAL LENGTH

- *Guttman Split Half*. Menunjukkan reliabilitas alat ukur yang diestimasi dengan menggunakan teknik belah dua dengan menggunakan formula dari Guttman yaitu $r_{xx'}=0.978$

5.B. PENULISAN LAPORAN PENELITIAN

Dengan menggunakan teknik belah dua dari Spearman-Brown didapatkan reliabilitas alat ukur harga diri adalah $r_{xx'} = 0.640$

E. KESEPAKATAN ANTAR RATER (KAPPA)

Reliabilitas antar rater ini dipakai menilai konsistensi dua orang rater dalam menilai performansi individu melalui *checklist* yang menghasilkan data nominal. Misalnya untuk mengetahui kesamaan psikolog menilai ada tidaknya simtom depresi pada klien (1=ada, 0=tidak), mengetahui kesamaan antara observer terhadap emosi subjek. Data yang dihasilkan adalah data nominal, misalnya (1=senang, 2=sedih, 3=takut, 4=marah). Semakin banyak kemiripan hasil penilaian antara satu rater dan rater lainnya maka koefisien reliabilitas yang dihasilkan akan tinggi.

1. CONTOH DESAIN STUDI

Schmitz, et.al (2004) hendak menguji peranan informasi di media cetak dalam mempengaruhi persepsi penderita *attention deficit/hyperactivity disorder* (ADHD). Untuk memilih paragraf di dalam artikel yang akan digunakan dalam penelitian, peneliti memberikan 32 artikel media massa kepada 2 orang rater. Tugas rater adalah memilih paragraf mana saja dalam artikel tersebut yang relevan dengan pembahasan ADHD. Dari penelitian tersebut didapatkan nilai konsistensi yang tinggi antar kedua rater yang ditunjukkan dengan nilai kappa(κ)=0.92

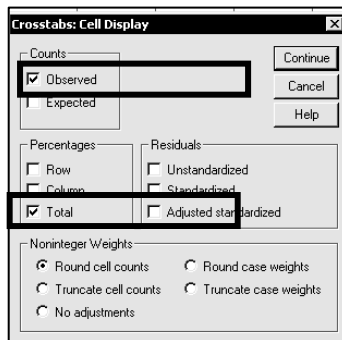
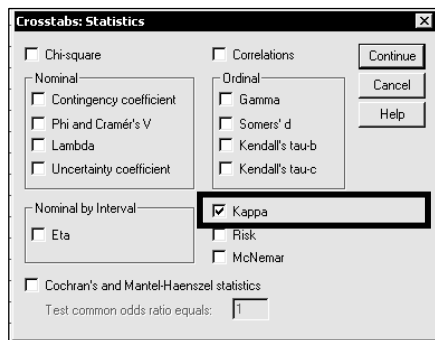
2. LATIHAN

Dua orang rater yang berlatar belakang pendidikan berbeda, yaitu dari psikologi dan non-psikologi diminta untuk menilai tipe kepribadian yang dialami oleh 10 orang subjek. Cara kedua rater menilai skor adalah (1 = tipe A, 2 = Tipe B). Peneliti hendak mencari seberapa jauh kesamaan rater dalam menilai kecemasan yang muncul, apakah latar belakang pendidikan dapat membedakan penilaian terhadap tipe kepribadian. Kutipan data simulasi dilihat pada tabel.

Tabel XX Data Laporan Observasi 2 Orang Rater

No	Rater_1	Rater_2
1	1	1
2	1	2
3	1	1
4	1	2
5	1	1

- Tekan Menu Analyze → Descriptive Statistics → Crosstab
- Masukkan variabel “rater_1” pada *rows* dan “rater_2” pada *column(s)*
- Masuk ke menu *statistics*, lalu pilih menu kappa → tekan *Continue*
- Masuk ke menu *Cells*, lalu pilih menu *Total* di bawah *Percentages* → tekan *Continue*
- Klik OK!



3. OUPUT DAN INTERPRETASI

- Tabel pertama memperlihatkan konsistensi penilaian antar rater. Terlihat bahwa dari 20 (90%) orang yang dinilai, terdapat 18 (60%) orang yang dinilai konsisten. 6 (30%) orang sama-sama dinilai Tipe A dan 12 Orang dinilai sama-sama tipe B. Hanya 2 orang yang dinilai berbeda, rater pertama menilai Tipe A sedangkan rater kedua menilai Tipe B
- Tabel Kedua menunjukkan reliabilitas antar rater yaitu $\kappa=0.783$. *Asymp. Std Error* menunjukkan kesalahan pengukuran terstandar, semakin kecil besarnya koefisien ini, semakin reliabel hasil pengukuran yang dihasilkan.

Rater Pertama * Rater Kedua Crosstabulation

		Rater Kedua		Total
		Tipe A	Tipe B	
Rater Pertama	Tipe A	Count 6	2	8
		% of Total 30,0%	10,0%	40,0%
Tipe B	Count	0	12	12
		% of Total ,0%	60,0%	60,0%
Total		Count 6	14	20
		% of Total 30,0%	70,0%	100,0%

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	,783	,142	3,586	,000
N of Valid Cases		20			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

4. RUMUS MENGHITUNG

Untuk menghitung nilai kappa diperlukan tabel dengan 2x2 yang menunjukkan jumlah ataupun proporsi persetujuan antar rater. Tabel di bawah ini adalah rangkuman dari penilaian antar rater di atas. Terlihat bahwa ada 6 orang yang sama-sama dinilai sebagai tipe A dan ada 12 orang yang dinilai sama-sama tipe B. Sebaliknya ketidaksepakatan terjadi pada 2 orang, yang dinilai berbeda oleh kedua rater.

		Rater 2		Jumlah
		Tipe A	Tipe B	
Rater_1	Tipe A	6	2	8
	Tipe B	0	12	12
Jumlah		6	14	20

		Rater 2		Jumlah
		Tipe A	Tipe B	
Rater_1	Tipe A	A	B	
	Tipe B	C	D	
Jumlah				

Nilai kappa didapatkan melalui transformasi tabel ke persamaan di bawah ini.

$$P_a = \frac{(A + D)}{(A + B + C + D)} = \frac{(6 + 12)}{(6 + 2 + 0 + 12)} = 0.900$$

$$\text{dan } P_c = \frac{(A+B)(A+C)(C+D)(B+D)}{(A+B+C+D)^2} = \frac{(6+2)+(6+0)+(0+12)+(2+12)}{(6+2+0+12)^2} = 0.540$$

$$\kappa = \frac{P_a - P_c}{1 - P_c} = \frac{0.900 - 0.540}{1 - 0.540} = 0.783$$

Fleiss (1981) mengkategorikan tingkat reliabilitas antar rater menjadi tiga kategori, antara lain :

- Kappa < 0.4 : buruk (*bad*)
- Kappa 0.4 - 0.60 : cukup (*fair*)
- Kappa 0.60 - 0.75 : memuaskan (*good*)
- Kappa > 0.75 : istimewa (*excellent*)

5. PENULISAN LAPORAN PENELITIAN

Dalam penulisan laporan ditulis keterangan bahwa koefisien reliabilitas dilaporkan dihitung dengan metode cohenn's kappa dan berapa orang yang menjadi rater. Lebih lengkap lagi jika penulis menyertakan nilai *standard error* pengukuran.

- Dalam penelitian ini Koefisien Cohen's Kappa digunakan untuk menghitung reliabilitas antar 2 rater. Hasil yang didapatkan adalah $\kappa=0.783$ dengan kesalahan standard sebesar 0.142
- *The inter-rater reliability of the CCI by the 2 scorers, as evaluated by the kappa reliability test, was 0.93 with a standard error of 0.10* (Reliabilitas antar rater instrumen CCI oleh dua rater yang dihitung dengan menggunakan uji reliabilitas kappa adalah sebesar 0.93 dengan standar kesalahan sebesar 0.10) (Bernardini, 2004)

Tekadang penulisan melampirkan persentase kesepakatan penilaian antar rater.

- *These two raters achieved 83% agreement. All points of disagreement were then discussed and consensus was reached about how to assign features* (Brownell, 2005).
- *The interrater agreement on first-letter codes for all available choices (N = 216) was 89.8%* (Jepsen, 2003).

F. RELIABILITAS ANTAR RATER (ICC)

Berbeda dengan metode sebelumnya, metode estimasi reliabilitas antar rater ini dipakai jika ada beberapa orang rater menilai individu baik melalui instrumen rating yang menghasilkan data ordinal. Misalnya konsistensi juri menilai performansi peserta kontes karaoke, konsistensi supervisor menilai kualitas hasil kerja anak buahnya.

Pada sub-bab ini reliabilitas antar rater dihitung dengan menggunakan koefisien korelasi antar kelas (*Intraclass Correlation Coefficients*, ICC). ICC menunjukkan perbandingan antara variasi yang diakibatkan atribut yang diukur dengan variasi pengukuran secara keseluruhan.

1. CONTOH DESAIN STUDI

Eisen et al (1998) hendak melakukan mengevaluasi *Brown Assessment of Beliefs Scale* yang mengukur tingkat delusi pasien. Untuk mengetahui reliabilitas alat ukur yang disusunnya, peneliti menggunakan empat orang rater untuk menilai 50 pasien yang memiliki gangguan yang berbeda. Teknik ICC dipakai untuk mengestimasi reliabilitas antar keempat rater.

2. LATIHAN KASUS

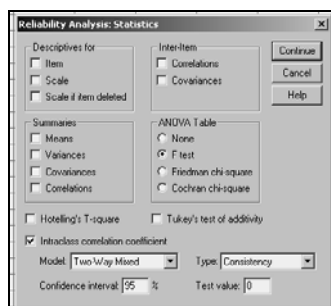
9 orang juri diminta menilai performansi 6 atlet terjun dengan menggunakan *checklist* yang terdiri dari 5 aitem antara yang menggunakan 5 alternatif penyekoran (1 hingga 5). Hasil

penilaian mereka dapat dilihat pada tabel di bawah ini. Reliabilitas dapat diestimasi dengan menggunakan teknik reliabilitas antar rater.

Tabel XX. Hasil observasi oleh rater

	R1	R2	R3	R4	R5	R6	R7	R8	R9
1	5.90	5.90	5.90	5.90	5.80	5.80	5.90	5.90	5.90
2	5.80	5.70	5.80	5.70	5.70	5.70	5.80	5.80	5.80
3	5.50	5.70	5.70	5.80	5.30	5.60	5.50	5.70	5.30
4	5.60	5.80	5.60	5.70	5.40	5.60	5.30	5.60	5.70
5	5.40	5.50	5.40	5.40	5.40	5.40	5.40	5.40	5.50
6	5.50	5.20	5.20	5.30	5.10	5.40	5.20	5.20	5.30

- Pilih Menu ANALYZE → SCALE → RELIABILITY ANALYSIS
- Masukkan kesembilan rater (R1 sampai R9) ke dalam kotak ITEM kemudian pilih STATISTICS
- Klik kotak F-TEST dan INTRACLASS CORRELATION COEFFICIENT
- Pilihlah jenis analisis sesuai dengan default SPSS, yaitu TWO WAY MIXED dan CONSISTENCY dengan CONFIDENCE INTERVAL 95%



3. OUPUT DAN INTERPRETASI

Reliability Statistics

Cronbach's Alpha	N of Items
.974	9

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between People	2.185	5	.437		
Within People				1.919	.084
Between Items	.173	8	.022		
Residual ^a	.450	40	.011		
Total	.622	48	.013		
Total	2.808	53	.053		

Grand Mean = 5.5796

a. Tukey's test for nonadditivity is undefined for dichotomous data.

Intraclass Correlation Coefficient

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig.
Single Measures	.808 ^b	.579	.964	38.883	5.0	40	.000
Average Measures	.974 ^c	.925	.996	38.883	5.0	40	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

- Tabel pertama menunjukkan nilai reliabilitas jika diestimasi dengan menggunakan Koefisien Alpha. Tabel menunjukkan nilai koefisien alpha yang memuaskan ($r_{xx}=0.974$)
- Tabel kedua adalah tabel keluaran analisis melalui ANOVA

- Tabel ketiga menunjukkan keluaran ICC dengan reliabilitas antar rater yang cukup memuaskan ($r_{xx}=0.808$)

4. RUMUS MANUAL

Nilai korelasi (ICC) didapatkan melalui rumus berikut, dengan menggunakan tabel ANOVA yang sudah di atas kita tidak kesulitan untuk mendapatkan nilai korelasi melalui rumus tersebut.

$$r = \frac{MS_{people} - MS_{residual}}{MS_{people} + (df_{people} \times MS_{residual})}$$

$$= \frac{0.473 - 0.011}{0.473 + (8 \times 0.11)} = 0.808$$

Hubungan antara ICC dengan alpha dapat diketahui melalui rumus berikut.

$$\alpha = \frac{k \times r}{1 + (k - 1) \times r}$$

$$= \frac{9 \times 0.808}{1 + (9 - 1) \times 0.808}$$

$$= 0.974$$

5. PENULISAN LAPORAN PENELITIAN

- Dalam penelitian ini Koefisien Cohen's Kappa digunakan untuk menghitung reliabilitas antar rater. Hasil yang didapatkan adalah $\kappa=0.783$ yang menunjukkan konsistensi penilaian antar rater adalah cukup tinggi.
- The ICCs for individual ratings of anxiety (CAS), depression (HDRS), and global impairment (GAS) were .93, .92, and .58 (Lenzenweger dan Loranger, 1989)