



Modul 7

FPM 226-Methodologi Penelitian Fisioterapi II

Materi 7

Multiple Comparison Test

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

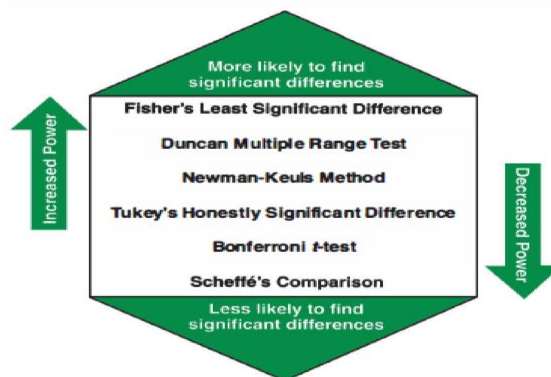
2019

Pendahuluan

Ketika analisis varians menghasilkan F-ratio yang signifikan, peneliti dibenarkan menolak *null hypothesis* dan menyimpulkan bahwa tidak semua k populasi sama; Namun, hasil ini memberitahu kita apa-apa tentang yang berarti secara signifikan berbeda dari yang lain berarti. Dalam materi ini, kita menjelaskan prosedur statistik yang paling umum digunakan untuk memutuskan yang berarti berbeda secara signifikan. Prosedur ini disebut *multiple comparison test*. Setiap tes melibatkan urutan ranking *mean*. Perbedaan pairwise antara *mean* diuji terhadap nilai kritis untuk menentukan apakah perbedaan cukup besar untuk menjadi signifikan.

Perbedaan utama antara berbagai tes terletak pada tingkat yang ditawarkan terhadap *type I* dan *type II error*. Terkait kesalahan *type I*, membutuhkan mean yang jauh berbeda untuk menetapkan signifikansi. Sebagian besar *multiple comparison* diklasifikasikan sebagai *post hoc* karena diputuskan setelah analisis varians telah selesai. Ini dianggap perbandingan yang tidak direncanakan yang didasarkan pada eksplorasi hasil. Oleh karena itu, tes ini paling berguna ketika hipotesis alternatif umum telah diusulkan.

Tiga jenis tes yang paling sering dilaporkan yaitu Tukey's *honestly significant difference method*, the *Newman-Keuls test*, dan the *Scheffe comparison*. Yang lain tapi kurang sering digunakan yaitu *Duncan's multiple range test* dan *Fisher's least significant difference*. Tes ini umumnya dianggap terlalu liberal, mengakibatkan terlalu besar risiko kesalahan *type I error* (gambar 1).



Gambar 1: *Multiple Comparison Test Berdasarkan Power*

Multiple comparison tests lainnya diklasifikasikan sebagai *a priori*, atau perbandingan yang direncanakan, karena direncanakan sebelum pengumpulan data berdasarkan alasan penelitian. Secara teknis, perbandingan ini sesuai bahkan ketika *F-test* tidak signifikan, karena direncanakan sebelum data dikumpulkan.

Tingkat *Type I Error*

Kita mendiskusikan ketidaktepatan menggunakan beberapa *t-test* ketika lebih dari dua perbandingan dibuat dalam satu kumpulan data. Masalah ini didasarkan pada penetapan α . Pada $\alpha = .05$, kita membatasi diri untuk 5% kesempatan bahwa kita akan mengalami peristiwa acak menemukan perbedaan yang signifikan ketika tidak ada. Kita mengambil risiko 5% bahwa kita akan salah jika kita mengatakan bahwa *mean* kelompok berbeda untuk setiap perbandingan tunggal.

Kita harus membedakan rasio kesalahan dari situasi di mana ditetapkan pada .05 untuk setiap beberapa perbandingan dalam satu eksperimen. Meskipun benar bahwa $\alpha = .05$ untuk setiap perbandingan individu, potensi kumulatif kesalahan untuk himpunan perbandingan sebenarnya lebih besar daripada .05. Probabilitas kumulatif ini disebut *familywise error rate* (α FW) dan mewakili probabilitas membuat setidaknya satu *type I error* dalam satu set dari perbandingan statistik.

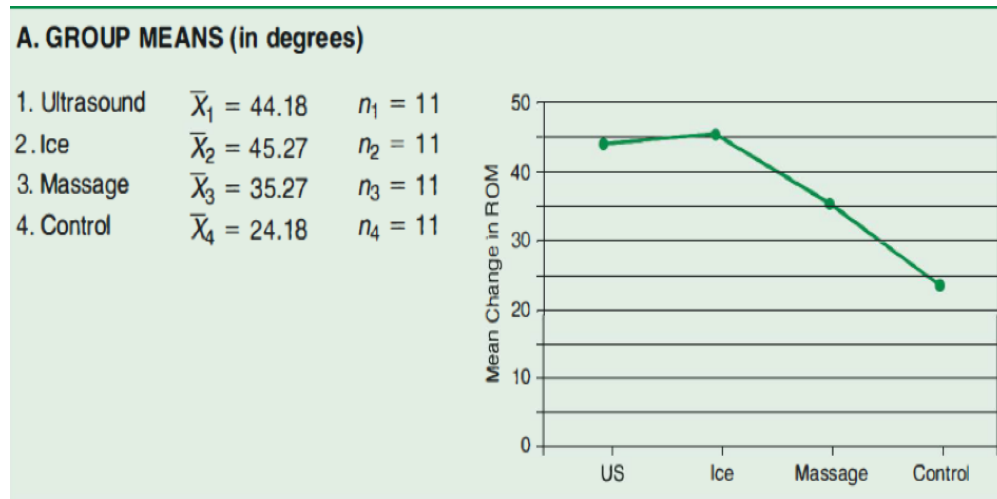
One-Tailed Versus Two-Tailed Tests

Perbandingan *post hoc* biasanya diuji menggunakan *two-tailed probabilities*. Jika kontras tertentu tidak ditentukan sebelumnya, hal ini mengikuti arah perbedaan yang tidak dapat diprediksi. *One-tailed tests* dapat dilakukan untuk perbandingan yang direncanakan jika bukti yang mendukung hipotesis terarah cukup kuat, umumnya statistik lebih aman untuk melakukan *two-tailed tests*. Kontras yang melibatkan *two-tailed tests* akan selalu didasarkan pada perbedaan absolut antar *mean*. *One-tailed tests* harus menghasilkan rasio statistik yang mendukung *directional hypothesis*.

Rasio Statistik pada *Multiple Comparison Test*

Untuk menggambarkan konsep *multiple comparison tests*, di bawah ini adalah contoh yang membandingkan efek *ultrasound (US)*, *ice* dan *friction massage* terhadap

nyeri pada 44 pasien dengan *elbow tendonitis*. *Mean* pada empat kelompok, yang ditunjukkan pada tabel 1A, mewakili perubahan nyeri pada tiga kelompok intervensi dan kelompok kontrol. Terdapat sebelas subjek dalam setiap kelompok. *Mean plot* menunjukkan distribusi setiap kelompok.



Tabel 1A. Data *Group Means*

Null hypothesis pada studi ini menunjukkan tidak ada perbedaan di antara ke empat kelompok

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Hasil analisis varians pada data ini, yang ditunjukkan pada tabel 1B menunjukkan signifikan ($p < .001$), dan sekarang kita akan melihat perbedaan di antara *mean*. Proses pengujian perbedaan antara beberapa *mean* cukup konsisten untuk semua prosedur *multiple comparison*. Dalam setiap tes, *mean* diatur dalam urutan ukuran, dan perbedaan antara pasangan *mean* yang diperoleh, seperti yang ditunjukkan pada tabel 1 C. Tabel ini menunjukkan perbedaan absolut antara semua pasangan *mean*, menggunakan format triangular.

B. ANALYSIS OF VARIANCE					
	Sum of Squares	df	Mean Square	F	Sig.
Between groups	3158.09	3	1052.70	11.89	.000
Within groups (Error)	3541.64	40	88.54		

C. TABLE OF MEAN DIFFERENCES				
	4	3	1	2
Means	Control 24.18	Massage 35.27	Ultrasound 44.18	Ice 45.27
4	24.18	—	11.09	20.00
3	35.27	—	8.91	10.00
1	44.18	—	—	1.09

Tabel 1B dan 1C. Hasil Analisis dan *Mean Differences*

Dengan $k = 4$, akan ada total $4(4-1)/2 = 6$ perbandingan. Entri dalam tabel adalah *pairwise mean differences*. Nilai tidak dimasukkan di bawah diagonal untuk menghindari pengulangan. Setiap *pairwise mean differences* diuji terhadap *minimum significant difference (MSD)*. Jika perbedaan absolut antara sepasang *mean* sama dengan atau lebih besar dari MSD, maka dianggap signifikan.

$$|\bar{X}_1 - \bar{X}_2| \geq \text{Minimum significant difference}$$

Jika perbedaan pairwise lebih kecil dari perbedaan MSD, berarti tidak berbeda secara signifikan satu sama lain. Perhitungan MSD berdasarkan *error mean square (MSe)*, yang berasal dari analisis varians, dan nilai kritis yang diambil dari tabel statistik. MSe mencerminkan tingkat varians dalam kelompok (antara subyek). Secara logis, semakin besar varians dalam kelompok, semakin kecil kemungkinan kita akan melihat perbedaan yang signifikan antara sarana. Nilai penting untuk MSD digunakan secara berbeda, tergantung pada jumlah berarti dibandingkan dan jenis tingkat kesalahan digunakan.

Tukey's Honestly Significant Difference (HSD)

Tukey mengembangkan salah satu prosedur *multiple comparison* sederhana, yang disebut *honestly significant difference (HSD)*. Prosedur Tukey menetapkan *familywise error rate*, sehingga mengidentifikasi probabilitas bahwa satu atau lebih dari *pairwise* akan dinyatakan signifikan secara keliru. Oleh karena itu, tes ini menawarkan proteksi terhadap *type I error*. *Tukey's HSD test* dihitung dengan menggunakan

studentized range statistic (simbol q). Nilai-nilai kritis q ditemukan dalam tabel A. 6 berikut:

TABLE A.6 CRITICAL VALUES OF THE STUDENTIZED RANGE STATISTIC, q , FOR TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) AND NEWMAN-KEULS (NK) COMPARISONS

df_e	α	$r = \text{number of means (HSD) or size of comparison interval (NK)}$										
		2	3	4	5	6	7	8	9	10	11	12
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.48
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03	8.18
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65	7.78
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.49
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.28

Tabel 2.1 Critical Value *Studentized Range Statistic*

TABLE A.6 CRITICAL VALUES OF THE STUDENTIZED RANGE STATISTIC, *q*, FOR TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) AND NEWMAN-KEULS (NK) COMPARISONS

24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69	5.76
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37	5.44
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29

Tabel 2.2 Critical Value *Studentized Range Statistic* (Lanjutan)

Statistik *q* dipengaruhi oleh jumlah keseluruhan *mean* yang dibandingkan. Di bagian atas tabel A. 6, angka atau "kisaran" berarti dibandingkan diberikan simbol *r*. Secara logis, karena jumlah sampel *mean* meningkat, ukuran perbedaan antara *mean* terbesar dan terkecil juga akan meningkat, bahkan ketika H_0 benar. Statistik *q* menyediakan mekanisme untuk menyesuaikan nilai kritis untuk memperhitungkan efek dari jumlah *mean* yang lebih besar.

Minimum Significant Difference

Perbedaan MSD untuk prosedur HSD Tukey sebagai berikut:

$$MSD = q \sqrt{\frac{MS_e}{n}}$$

dimana MS_e adalah *mean square error* dan *n* adalah jumlah subjek dalam setiap kelompok (dengan asumsi ukuran sampel yang sama *), dan *q* diambil dari lampiran tabel A. 6, untuk tingkat yang diinginkan dari α , d F_e , dan jumlah berarti, *r*. Untuk contoh

yang kita gunakan, $q = 3,79$ untuk $\alpha = .05$, $r = 4$, dan $df_e = 40$. Dengan demikian, nilai MSD adalah sebagai berikut:

$$MSD = 3.79 \sqrt{\frac{88.54}{11}} = 10.75$$

MSD dibandingkan dengan setiap *pairwise mean difference* dalam tabel 3.. Perbedaan absolut yang sama atau lebih besar dari nilai ini adalah signifikan. Sebagai contoh, perbedaan antara terbesar dan terkecil berarti (X2-X4) adalah 1,09. Nilai ini melebihi MSD dan, oleh karena itu, signifikan. Menurut hasil ini, tiga kelompok eksperimental berbeda dari kelompok kontrol, tetapi kelompok treatment tidak berbeda satu sama lain.

Analisis komputer data ini disajikan dalam hal *subset* homogen. Dalam *output* ini, setiap *subset* (tercantum dalam kolom yang sama) merepresentasikan bahwa *mean* tidak berbeda secara signifikan. Dan yang tercantum dalam kolom terpisah secara signifikan berbeda dari satu sama lain.

Newman-Keuls (NK) Test

Newman-Keuls (NK) test (kadang-kadang disebut *Student -Newman-Keuls test*) mirip dengan metode Tukey, kecuali menggunakan tingkat kesalahan perbandingan. Oleh karena itu, menetapkan tingkat *type I error* untuk setiap kontras *pairwise*, bukan untuk seluruh rangkaian perbandingan. Secara keseluruhan, ketika jumlah perbandingan meningkat, kemungkinan melakukan *type I error* lebih besar menggunakan prosedur ini dibanding menggunakan *Tukey's test*.

TABLE 21.2 SIGNIFICANT DIFFERENCES (*) FOR TUKEY'S HSD TEST ($\alpha = .05$)

A. MINIMUM SIGNIFICANT DIFFERENCE: 10.75

B. TABLE OF MEAN DIFFERENCES

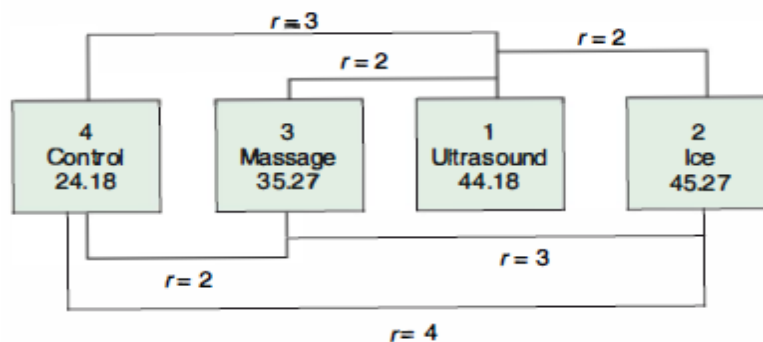
		4	3	1	2
	Means	Control 24.18	Message 35.27	Ultrasound 44.18	Ice 45.27
4	24.18	—	11.09*	20.00*	21.09*
3	35.27		—	8.91	10.00
1	44.18			—	1.09

C. OUTPUT: Homogeneous Subsets of Means

GROUP		N	Subset for alpha = .05	
			1	2
Tukey HSD	Control	11	24.18	
	Message	11		35.27
	US	11		44.18
	Ice	11		45.27

Tabel 3. MSD, Mean Difference dan Output Tukey's Test

Metode Newman-Keuls juga didasarkan pada kisaran *studentized q*; Namun, nilai *q* digunakan secara berbeda untuk setiap kontras, tergantung pada jumlah cara yang berdekatan, *r*, dalam interval perbandingan secara berurutan. Untuk mengilustrasikan bagaimana ini diterapkan, pertimbangkan empat contoh berarti untuk studi tendonitis, peringkat dalam urutan ukuran secara meningkat adalah : (4) kontrol, (3) *massage*, (1) *US*, (2) *ice*. Jika kita membandingkan dua cara yang lebih kecil, perbandingan interval untuk *control-massage* mencakup dua cara yang berdekatan. Oleh karena itu, $r = 2$ untuk perbandingan. Jika kita membandingkan cara terbesar dan terkecil, interval untuk *control-ice* berisi empat cara yang berdekatan (4-3-1-2), dan $r = 4$. Demikian pula, jika kita membandingkan mean *massage* dan *ice*, interval perbandingan berisi tiga cara yang berdekatan (3-1-2), jadi $r = 3$.



Gambar 2. Perbandingan Interval Grup Mean Berdasarkan Urutan

Karena itu, perbandingan interval merupakan dengan mengurutkan *mean* untuk perbandingan tertentu. Berbeda dengan pendekatan Tukey yang menggunakan satu perbedaan penting untuk semua perbandingan, *Newman-Keuls test* akan menggunakan perbedaan penting yang lebih besar ketika *r* meningkat. Ini menyesuaikan pada fakta bahwa perbedaan yang lebih besar diharapkan dengan berbagai cara yang lebih besar, bahkan ketika H_0 adalah benar.

Minimum Significant Difference

Perbedaan MSD *Newman-Keuls test* adalah sebagai berikut:

$$MSD = q_{(r)} \sqrt{\frac{MS_e}{n}}$$

dimana nilai $q(r)$ diperoleh dari tabel A. 6 untuk setiap interval perbandingan. Untuk contoh yang kita gunakan, q untuk $\alpha = .05$ dan $d Fe = 40$ untuk interval perbandingan $r = 2, 3,$ dan 4 :

$$q_{(r=2)} = 2.86$$

$$q_{(r=3)} = 3.44$$

$$q_{(r=4)} = 3.79$$

Dengan $MSe = 88,54$ dan $n = 11$, kita menemukan MSD:

$$MSD(r = 2) = 2.86 \sqrt{\frac{88.54}{11}} = 8.11$$

$$MSD(r = 3) = 3.44 \sqrt{\frac{88.54}{11}} = 9.76$$

$$MSD(r = 4) = 3.79 \sqrt{\frac{88.54}{11}} = 10.75$$

MSD ini dibandingkan dengan perbedaan mean sesuai pada tabel 4. Perbedaan signifikan ditandai dengan tanda bintang. Sebagai contoh, perbedaan *mean* untuk *control* dan *ice* adalah 21,09, yang melebihi perbedaan kritis 10,75 untuk $r = 4$. Oleh karena itu, dua cara ini secara signifikan berbeda. Perbedaan mean antara *massage* dan *ice* adalah 10,00, yang melebihi perbedaan kritis 9,76 untuk $r = 3$. Kedua cara ini juga secara signifikan berbeda satu sama lain. Perbedaan antara *US* dan *ice* adalah 1,09, yang tidak melebihi perbedaan kritis 8,11 untuk $r = 2$. Ini berarti tidak berbeda secara signifikan. Dari enam perbandingan, lima signifikan. Tes ini menunjukkan bahwa

tiga kelompok eksperimental berbeda dari kelompok kontrol, *US* dan *ice* berbeda dengan *massage*.

TABLE 21.3 SIGNIFICANT DIFFERENCES (*) FOR THE NEWMAN-KEULS TEST ($\alpha = .05$)

A. MINIMUM SIGNIFICANT DIFFERENCES:
 Comparison Intervals: $r = 2$ MSD = 8.11
 $r = 3$ MSD = 9.76
 $r = 4$ MSD = 10.75

B. TABLE OF MEAN DIFFERENCES

	4	3	1	2
	Control 24.18	Massage 35.27	Ultrasound 44.18	Ice 45.27
3 24.18	—	11.09* ($r = 2$)	20.00* ($r = 3$)	21.09* ($r = 4$)
2 35.27		—	8.91* ($r = 2$)	10.00* ($r = 3$)
1 44.18			—	1.09 ($r = 2$)

C. OUTPUT: HOMOGENEOUS SUBSETS OF MEANS

		N	Subset for alpha = .05		
GROUP			1	2	3
Newman-Keuls	Control	11	24.18		
	Massage	11		35.27	
	US	11			44.18
	Ice	11			45.27

Tabel 4. MSD, Mean Difference dan Output Newman-Keuls Test

Hasil ini juga ditampilkan pada tabel 4 untuk *subset mean*. Karena *mean* untuk kelompok kontrol tercantum dalam kolom sendiri, yang berbeda dengan *mean* lain. Hal yang sama berlaku untuk *mean* untuk kelompok *massage*. Mean *US* dan *ice* tercantum dalam kolom yang sama, menunjukkan bahwa mereka tidak berbeda satu sama lain. Perlu dicatat bahwa MSD *Newman-Keuls test* dengan $r = 4$ sama dengan *Tukey test* (dalam hal ini 10,75). Prosedur *Tukey* menggunakan satu MSD untuk semua perbandingan, sedangkan *Newman-Keuls test* menyesuaikan perbedaan minimum untuk interval perbandingan yang lebih kecil.

Oleh karena itu, perbedaan minimum akan lebih rendah menggunakan metode Newman-Keuls. Akibatnya, *Newman-Keuls test* dapat mengakibatkan perbedaan yang lebih signifikan dan lebih kuat dari dua perbandingan. Namun, karena prosedur Newman-Keuls tidak mengontrol *familywise error rate*, itu akan menghasilkan lebih banyak *type I error* daripada metode Tukey dalam jangka panjang.

Scheffe Comparison

Scheffe comparison adalah *multiple comparison* yang paling fleksibel dan paling ketat dari post hoc. Hal ini didasarkan pada *F-distribution*. Ini adalah tes konservatif karena mengadopsi *familywise error rate* yang berlaku untuk semua. Hal ini memberikan perlindungan yang kuat terhadap *type I error*, tetapi juga membuat prosedur jauh lebih kuat daripada tes lain yang telah dijelaskan. Scheffe telah merekomendasikan bahwa tingkat yang kurang ketat signifikansi digunakan, seperti = .10, untuk menghindari *type II error*.

Minimum Significant Difference

MSD untuk Scheffe comparison adalah sebagai berikut:

$$\text{MSD} = \sqrt{(k - 1)F} \sqrt{\frac{2MS_e}{n}}$$

di mana k adalah jumlah *mean* dalam serangkaian perbandingan, dan F adalah nilai kritis untuk dfb dan Untuk contoh yang digunakan, k = 4 dan F = 2,84 untuk 3 dan 40 derajat kebebasan pada $\alpha = .05$. Karena itu MSD adalah:

$$\text{MSD} = \sqrt{(4 - 1)(2.84)} \sqrt{\frac{2(88.54)}{11}} = 11.71$$

Semua perbedaan antar *mean* berarti harus memenuhi atau melampaui nilai ini agar signifikan. Oleh karena itu, seperti yang dilambangkan dengan tanda bintang di tabel sebelumnya analisis ini menghasilkan dua perbandingan yang signifikan (lebih sedikit daripada metode Newman-Keuls atau Tukey), menunjukkan kekuatan rendah

yang terkait dengan *Scheffe Comparison*. Menurut tes ini, kelompok kontrol dan *massage* tidak secara signifikan berbeda satu sama lain, dimana mereka dianggap berbeda secara signifikan dengan tes lain.

TABLE 21.4 SIGNIFICANT DIFFERENCES (*) FOR THE SCHEFFÉ COMPARISON ($\alpha = .05$)

A. MINIMUM SIGNIFICANT DIFFERENCE: 11.71

B. TABLE OF MEAN DIFFERENCES

		4	3	1	2
	Means	Control 24.18	Massage 35.27	Ultrasound 44.18	Ice 45.27
4	24.18	—	11.09	20.00*	21.09*
3	35.27		—	8.91	0.00
1	44.18			—	1.09

C. OUTPUT: Homogeneous Subsets of Means

		N	Subset for alpha = .05	
	GROUP		1	2
Scheffé	Control	11	24.18	
	Massage	11	35.27	35.27
	US	11		44.18
	Ice	11		45.27

Tabel 5. MSD, Mean Difference dan Output Scheffe Comparison Test

Bonferroni t-Test

Para peneliti sering kali menetapkan kontras spesifik kepentingan tertentu sebelum pengumpulan data. Kontras ini biasanya berkaitan dengan ekspektasi teoritis data. Ketika perbandingan direncanakan di muka dan ketika mereka relatif terbatas jumlahnya, *a priori test* dapat digunakan. Alasan untuk aplikasi yang valid dari perbandingan yang direncanakan harus dibentuk sebelum data dikumpulkan, sehingga pilihan hipotesis tertentu tidak dapat dipengaruhi oleh data. Karena peneliti tidak selalu

tertarik pada semua kontras yang mungkin, itu sebenarnya tidak perlu untuk menguji keseluruhan null hypothesis dengan analisis varians. Terlepas dari apakah ANOVA menunjukkan signifikan F-rasio, perbandingan perencanaan dapat dibuat.

Perbandingan Bonferroni (juga disebut *Dunn's multiple comparison procedure*) adalah perbandingan yang direncanakan, menggunakan familywise error rate yang merupakan jumlah tingkat signifikansi perbandingan. Pada dasarnya, prosedur membagi secara merata di antara set kontras direncanakan, sehingga setiap kontras diuji. Oleh karena itu, jika seorang peneliti menginginkan probabilitas keseluruhan .05 untuk satu set dari empat kontras, masing-masing perbandingan akan harus mencapai signifikansi di .05 atau $p = .013$. Proses penyesuaian α , yang disebut penyetelan Bonferroni (atau koreksi), digunakan sebagai perlindungan terhadap *type I error*. *Bonferroni test* didasarkan pada *student's t-distribution*, dengan penyesuaian yang dilakukan untuk jumlah kontras yang dilakukan dalam satu set data.

Minimum Significant Difference

MSD pada *Bonferroni test* dapat dihitung dengan menggunakan rumus berikut:

$$MSD = t(B) \sqrt{\frac{2MS_e}{n}}$$

Untuk α_{FW} , DFE, dan c , di mana c adalah jumlah total perbandingan dalam percobaan. Melanjutkan dengan contoh yang telah kita gunakan, untuk enam perbandingan yang dilakukan di $\alpha_{FW} = .05$, dengan DFE = 40, kita menemukan $t(B) = 2,77$. Karenanya:

$$MSD = 2.77 \sqrt{\frac{2(88.54)}{11}} = 11.11$$

Semua perbedaan *pairwise* dibandingkan dengan MSD yang signifikan, seperti yang ditunjukkan pada tabel 21,5. Dalam kasus ini, tiga dari enam perbandingan yang signifikan. Menurut hasil ini, tiga kelompok intervensi berbeda dari kontrol.

TABLE 21.5 SIGNIFICANT DIFFERENCES (*) FOR THE BONFERRONI t-TEST ($\alpha = .05$)

A. MINIMUM SIGNIFICANT DIFFERENCE: 11.11

B. TABLE OF MEAN DIFFERENCES

		4	2	1	2
Means		Control 24.18	Massage 35.27	Ultrasound 44.18	Ice 45.27
4	24.18	—	11.09*	20.00*	21.09*
3	35.27		—	8.91	10.00
1	44.18			—	1.09

C. OUTPUT: Homogeneous Subsets of Means

			Subset for alpha = .05	
	GROUP	N	1	2
Bonferroni	Control	11	24.18	
	Massage	11		35.27
	US	11		44.18
	Ice	11		45.27

Tabel 5. MSD, Mean Difference dan Output Bonferroni t-Test

Referensi

1. Keppel G. Design and Analysis: A Researcher's Handbook (4th ed.). Englewood Cliffs, NJ: Prentice Hall, 2004.
2. Green SB, Salkind NJ, Akey TM. Using SPSS for Windows and Macintosh: Analyzing and Understanding Data (4th ed.). Upper Saddle River, NJ: Prentice Hall, 2004.
3. Dennis Howitt, Duncan Cramer. Introduction to SPSS Statistics in Psychology For Version 19 and earlier Fifth edition. Pearson Education Limited Edinburgh Gate Harlow Essex CM20 2JE England 2011.