Sains Malaysiana 49(4)(2020): 885-898
http://dx.doi.org/10.17576/jsm-2020-4904-18

Pengujian Hipotesis Berbilang: Perbandingan Simulasi Monte Carlo Berdasarkan Ralat Jenis I

(Comparative Multiple Hypothesis Testing: Comparison of Monte Carlo Simulation Based on Type-1 Error)

NORA MUDA* & NOR SYAFAWATI JANI

ABSTRAK

Pengujian hipotesis berbilang merupakan pengujian yang melibatkan ujian serentak lebih daripada satu hipotesis dan digunakan untuk mengawal kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR) dengan meminimumkan Ralat Jenis I. Kajian ini bertujuan untuk membuat perbandingan ujian pengujian hipotesis berbilang bagi ujian-t iaitu pengujian antara dua kumpulan sampel melalui perbandingan antara ujian Bonferroni*, ujian* Holm*, ujian* Hochberg*, ujian* Hommel*, ujian* Benjamini-Hochberg *dan ujian* Benjamini-Yekutieli *dengan mengikut keadaan yang tertentu iaitu nilai α, bilangan ujian, m dan jenis taburan yang berbeza. Perbandingan pengujian hipotesis berbilang berdasarkan kebarangkalian Ralat Jenis I bagi kes kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR) dijalankan berdasarkan simulasi Monte Carlo. Didapati, bagi kumpulan min yang sama iaitu {0,0} bagi kesemua keadaan, Ralat Jenis I bernilai sifar. Hal ini kerana kesemua ujian gagal menolak hipotesis nol dan terbukti menyatakan kesemua hipotesis nol adalah benar. Selain itu, aras keertian 0.01 tidak sesuai digunakan bagi kesemua keadaan kerana aras keertian ini dikatakan sangat jitu. Bagi kumpulan min yang berbeza iaitu {0,1}, ujian* Benjamini-Yekutieli *sesuai digunakan bagi mengawal kadar penemuan palsu (FDR) kerana dapat meminimumkan Ralat Jenis I dengan baik berbanding dengan ujian lain. Manakala bagi kadar ralat berkumpulan (FWER), ujian* Hommel *sesuai digunakan berbanding dengan ujian lain. Hal ini kerana ujian ini dapat mengawal dengan baik dan meminimumkan Ralat Jenis I.*

Kata kunci: Kadar penemuan palsu; kadar ralat berkumpulan; ujian Benjamini-Hochberg*; ujian* Bonferroni*; ujian* Holm

ABSTRACT

Multiple hypothesis testing is a test that involves more than one hypothesis test which run simultaneously and is used to control group error rate (FWER) and false discovery rate (FDR) by minimizing Type I Error. This study aims to compare multiple hypothesis testing tests for t-test; test between two group samples by comparing between Bonferroni test, Holm test, Hochberg test, Hommel test, Benjamini-Hochberg test, and Benjamini-Yekutieli test according to specific conditions namely α value, number of tests, m and different types of distribution. Comparison of multiple hypothesis testing based on probability of Type I error for group error rate (FWER) and false discovery rate (FDR) was performed based on Monte Carlo simulation. It is found that for the group with that same mean {0,0} in all cases, the Type I error is zero. This is because all tests failed to reject the null hypothesis and proved that all null hypotheses were true. Also, the significance level of 0.01 is not appropriate for all situations because it is said to be very accurate. For different mean groups of {0,1}, the Benjamini-Yekutieli test is best used to control the false discovery rate (FDR) as it minimizes Type I error better than other tests. For group error rates (FWER), the Hommel test is applicable compared to other tests. This is because this test can control and minimize Type I Errors.

Keywords: Benjamini-Hochberg test; Bonferroni test; false discovery rate (FDR); group error rate (FWER); Holm test

PENGENALAN

Pengujian hipotesis statistik merupakan perbandingan parameter antara cerapan dengan teori. Pengujian hipotesis perlu dilakukan dengan menyatakan hipotesis nol dan hipotesis alternatif terlebih dahulu. Seterusnya, menghitung statistik ujian, menetapkan rantau penolakan dan membuat keputusan menolak hipotesis nol atau gagal menolak hipotesis nol. Pengujian hipotesis terbahagi kepada dua iaitu pengujian hipotesis tunggal dan pengujian hipotesis berbilang. Bagi pengujian hipotesis tunggal, kebiasaannya kita menguji hipotesis nol dan hipotesis alternatif dengan menggunakan suatu statistik ujian. Penolakan hipotesis nol dan menerima hipotesis alternatif apabila statistik ujian berada dalam rantau penolakan dengan mengikut keadaan yang telah ditetapkan. Terdapat beberapa kemungkinan untuk membuat salah satu daripada dua jenis ralat iaitu Ralat Jenis I dan Ralat Jenis II. Ralat Jenis I atau dikenali sebagai positif palsu terjadi apabila membuat keputusan untuk menolak hipotesis nol walhal hipotesis nol benar. Manakala, Ralat Jenis II atau dikenali sebagai negatif palsu terjadi apabila tidak menolak atau gagal menolak hipotesis nol walhal hipotesis alternatif adalah benar.

Pengujian hipotesis berbilang pula merujuk kepada pengujian yang serentak lebih daripada satu hipotesis serta mengubah suai nilai-*p* individu setiap ujian yang dijalankan dengan mengekalkan kadar ralat keseluruhan. Apabila menjalankan pengujian hipotesis berbilang, jika mengikut peraturan rantau penolakan yang sama untuk ujian yang berasingan atau tidak bersandar, akan menyebabkan kebarangkalian melakukan sekurang-kurangnya satu Ralat Jenis I adalah lebih tinggi berbanding dengan aras nominal yang digunakan untuk setiap ujian terutamanya apabila bilangan ujian, *m* yang digunakan adalah besar. Bagi *m* ujian yang tidak bersandar, jika α merupakan aras penolakan untuk setiap nilai-*p*, maka kebarangkalian ralat jenis satu menjadi seperti $(1 - \alpha)_m$. Hal ini kerana $0 < \alpha < 1$, mengikut $(1 - \alpha)_m < (1 - \alpha)$ dan kebarangkalian tidak membuat Ralat Jenis I dalam ujian $m > 1$ adalah kecil berbanding dengan menggunakan ujian hipotesis tunggal. Oleh itu, kebarangkalian membuat sekurang-kurangnya satu Ralat Jenis I dalam ujian *m* adalah lebih tinggi berbanding kes bagi satu ujian. Sebagai contoh, nilai-$p < 0.05$ digunakan sebagai rantau penolakan untuk 100 ujian dan kebarangkalian membuat sekurang-kurangnya satu Ralat Jenis I ialah 0.99.

Terdapat banyak set data yang mengandungi bilangan pemboleh ubah yang banyak dan diperoleh daripada uji kaji yang telah dijalankan dalam pelbagai bidang seperti biologi, bioperubatan, epidemiologi, genom dan analisis imej. Set data ini meningkat seiringan dengan teknologi dan penyelidikan yang semakin maju. Oleh itu, set data yang banyak ini adalah salah satu sebab kebanyakan penyelidik menggunakan pengujian hipotesis berbilang dan masih digunakan sehingga kini. Godfrey (1985), Pocock et al. (1987) dan Smith et al. (1987) membuat kajian mengenai perbandingan kajian antara sampel laporan daripada jurnal-jurnal perubatan dan mendapati penyelidik tidak melihat kepada data yang berbilang menyebabkan keputusan yang diperoleh mempunyai perbezaan dalam mengawal keadaan. Sebagai contoh dalam kajian DNA yang melakukan 10,000 ujian hipotesis yang berasingan iaitu menggunakan pengujian hipotesis tunggal menyebabkan kekangan masa dan kos dalam menganalisis kerana ujian yang dijalankan banyak. Oleh itu, pengujian hipotesis berbilang diperlukan dalam bidang-bidang tersebut untuk mendapatkan keputusan yang tepat berbanding menggunakan pengujian hipotesis tunggal.

Oleh itu, untuk menangani masalah ini, pengujian hipotesis berbilang seperti ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* telah diperkenalkan bagi membuat ujian individu lebih konservatif bagi meminimumkan bilangan Ralat Jenis I dengan mengekalkan kadar ralat untuk keseluruhan ujian. Pemilihan ujian-ujian ini adalah berdasarkan perkembangan semasa dan terkini hasil penyelidikan berdasarkan situasi tertentu. Sebagai contoh, Holm (1979) telah membuat kajian mengenai prosedur menolak ujian berbilang secara berurutan dengan mudah. Beliau telah menyatakan ujian *Bonferroni* adalah prosedur yang mudah dengan menolak hipotesis nol apabila nilai-*p* lebih kecil daripada $\frac{\alpha}{m}$ dengan α ialah aras keertian dan *m* ialah bilangan ujian hipotesis. Manakala Simes (1986) pula mengubah suai kaedah *Bonferroni* berdasarkan penyusunan nilai-p daripada ujian individu yang merupakan kurang konservatif berbanding dengan ujian klasik *Bonferroni* tetapi masih mudah untuk digunakan. Hommel (1988) pula menggunakan prinsip Marcus et al. (1976) dengan mengawal aras keertian, *α*. Hochberg (1988) pula memperkenalkan prosedur yang mudah bagi ujian berbilang berdasarkan nilai-*p* individu. Prosedur ini lebih tepat berbanding ujian Holm (1979) dalam prosedur berurutan dalam menolak hipotesis. Dengan membuat perbandingan, kajian ini mendapati ujian *Bonferroni* lebih mudah digunakan berbanding dengan ujian lain.

Berbeza dengan Wright (1992) yang menyatakan keputusan daripada ujian serentak bagi nilai-*p* terubah suai, jika nilai-*p* terubah suai bagi hipotesis individu lebih kecil berbanding dengan aras keertian, α yang telah ditetapkan, maka hipotesis perlu ditolak dengan kadar ralat tidak melebih daripada *α*. Benjamini dan Hochberg (1995) telah mengemaskini ujian mereka dengan mengawal kadar penemuan palsu (FDR) dalam pengujian hipotesis berbilang. Mereka memperkenalkan penghampiran yang lain dalam masalah pengujian hipotesis berbilang dan dikenali sebagai mengawal jangkaan perkadaran penyalahan dalam menolak hipotesis iaitu kadar penemuan palsu (FDR). Kaedah sedia ada diperbaiki lagi oleh Aickin dan Gensler (1996) dengan membuat perbandingan antara ujian *Bonferroni* dan ujian *Holm* yang memperkenalkan ujian menyusun nilai-*p* secara menurun. Kedua-dua ujian ini dibandingkan dengan menggunakan data kanser payu dara. Benjamini dan Yekutieli (2001) memperkenalkan ujian pengujian hipotesis berbilang bagi mengawal kadar penemuan palsu (FDR) bagi kes bersandar. Ujian ini lebih berkuasa dan dapat mengawal kadar penemuan palsu (FDR) dengan baik walaupun prosedur bagi ujian ini agak rumit.

Kajian mengenai DNA adalah kajian yang baru dalam bidang bioteknologi dalam membolehkan pemantauan tahap ungkapan sel-sel yang terdapat dalam beribu gen secara serentak. Aplikasi pengujian hipotesis berbilang digunakan dalam uji kaji mikro tatasusunan ini oleh Dudoit et al. (2003). Kaedah baru diperkenalkan lagi oleh Abdi (2010) yang mengkaji mengenai kaedah Holm-Sidak dan Holm-Bonferroni berurutan. Perkembangan kaedah kajian diteruskan lagi oleh Sinclair et al. (2013) yang mengkaji mengenai pelarasan aras *α* bagi pengujian berbilang untuk pemboleh ubah yang bersandar. Apabila lebih daripada satu ujian tidak bersandar digunakan, kemungkinan dalam mengenal pasti kesan signifikan meningkat berdasarkan bilangan analisis yang dijalankan.

Kajian simulasi Monte Carlo dengan mengambil kira nilai Ralat Jenis I dan kuasa ujian serta membandingkan hasil ujian statistik seperti ujian-*t* dan ujian ANOVA (Analisis Varians) dengan kriteria Bradley telah dijalankan oleh beberapa penyelidik seperti Abdul Rahman dan Lai (2014) yang meneruskan kajian yang dilakukan oleh Nor Aishah et al. (2012) yang memperkenalkan kaedah pseudo-median dalam ujian Wilcoxon bagi membuat perbandingan antara beberapa kumpulan tak bersandar yang mempunyai saiz sampel yang kecil dan tidak memenuhi andaian kenormalan. Abdul Rahman dan Lai (2014) telah menjalankan pelbagai ujian tak-berparameter dan mendapati ujian Mann-Whitney paling kerap dipilih. Mereka telah memperkenalkan ujian pengubahsuaian Mann-Whitney (RMW) dan mendapati ujian RMW dapat mengawal ralat jenis I dengan baik sungguhpun kuasanya rendah. Walau bagaimanapun, ujian ini dijalankan bagi membuat perbandingan terhadap dua kumpulan tak bersandar.

Dalam kajian ini, dua sampel digunakan bagi membuat perbandingan kehomogenan min dengan menggunakan ujian-t dengan varians tergembleng. Perbezaan bagi ujian ini adalah apabila ujian ini perlu diulang bagi kumpulan yang berbeza dengan keadaan yang berbeza. Satu kaedah yang dapat menjimatkan masa menjalankan ujian adalah melalui pengujian hipotesis berbilang kerana ia dijalankan secara serentak. Sampel yang digunakan dalam ujian berbilang ini dijana daripada taburan Normal, taburan Khi-Kuasa Dua dan taburan Lognormal. Seterusnya, nilai-p yang diperoleh daripada ujian tersebut digunakan dalam ujian pengujian hipotesis berbilang iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* serta dibandingkan dengan mengikut beberapa keadaan yang telah ditetapkan. Oleh itu, dapat ditentukan ujian yang terbaik dengan mengikut keadaan yang ditetapkan.

DATA DAN SKOP KAJIAN

Kajian ini dijalankan bagi membuat perbandingan dan menentukan pengujian hipotesis berbilang bagi dua sampel dengan mengikut beberapa keadaan yang tertentu dengan memperoleh kebarangkalian Ralat Jenis I bagi kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR). Beberapa keadaan digunakan di dalam kajian ini iaitu aras keertian, α yang berbeza, bilangan ujian, m, yang berbeza dan jenis taburan yang berbeza.

Kajian ini dikendalikan dengan bilangan ujian, m yang berbeza iaitu 5, 10, 20, 25, 30 dan 50. Selain itu, kajian ini menggunakan saiz sampel yang sama iaitu 10 sampel dan varians yang sama iaitu 1 daripada tiga taburan yang berbeza iaitu taburan normal, taburan khi-kuasa dua dan taburan log-normal. Manakala, terdapat dua nilai min iaitu 0 dan 1. Bagi taburan khi-kuasa dua menggunakan 3 darjah kebebasan kerana mewakili taburan pencongan sederhana. Bagi tujuan membandingkan kadar Ralat Jenis I bagi kes kadar ralat berkumpulan (FWER) dan kadar ralat penemuan palsu (FDR), uji kaji ini dijalankan mengikut bilangan ujian yang ditentukan bagi setiap taburan.

KAEDAH KAJIAN

Kajian simulasi Monte Carlo digunakan dalam menghitung kadar Ralat Jenis I bagi kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR). Simulasi Monte Carlo mengandungi ujian simulasi yang kerap dijalankan dengan menggunakan data rawak dan menghasilkan keputusan yang mempunyai bentuk yang berbeza. Penggunaan simulasi Monte Carlo dapat mengelakkan kekeliruan yang terdapat dalam membuat statistik ujian. Algoritma bagi pengiraan kadar Ralat Jenis I adalah seperti berikut:

Pertama, menjana sampel rawak daripada taburan normal, taburan khi-kuasa dua dan taburan log-normal. Andaikan hipotesis nol adalah benar iaitu H_0: $\mu_X = \mu_Y$. Kedua, menghitung nilai statistik ujian. Ketiga, menghitung nilai-p dan Keempat, ulang langkah (1) hingga langkah (3) mengikut bilangan ujian.

KAEDAH MENENTUKAN KESAMAAN MIN ANTARA DUA SAMPEL

Selepas data dijana, ujian bagi kesamaan min antara dua sampel perlu dijalankan. Katakanlah $X = x_1, x_2, ..., x_n$ dan $Y = y_1, y_2, ..., y_n$, yang $X \sim N\,(0,1)$ dan $Y \sim N\,(0,1)$ adalah populasi yang tidak bersandar dan saiz sampel adalah n. Bagi menguji kesamaan min antara dua sampel, hipotesis nol dan hipotesis alternatif dinyatakan terlebih dahulu seperti berikut :

$H_0 : \mu_X = \mu_Y$, kedua-dua populasi min adalah sama

$H_1 : \mu_X \neq \mu_Y$, kedua-dua populasi min adalah tidak sama

Terdapat beberapa kaedah menguji kesamaan min iaitu ujian-t varians tergembeleng, ujian-t *Welch* dan ujian *Alexander Govern*. Kajian ini menggunakan ujian-t varians tergembeleng bagi dua sampel. Ujian-t digunakan apabila andaian dipenuhi iaitu sampel adalah taburan normal dan saiz sampel kecil, $n < 30$ serta varians bagi kedua-dua sampel adalah sama. Nilai varians sepunya σ^2 dianggar menggunakan varians tergembeleng, $S_p^{\,2}$ yang merujuk kepada min berpemberat bagi kedua-dua varians sampel $S_1^{\,2}$ dan $S_2^{\,2}$. Varians tergembeleng bagi ujian ini boleh dirumuskan iaitu:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Statistik ujian bagi ujian-t adalah seperti berikut:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} \quad \frac{1}{n_2}\right)}}$$

Dengan n_1 dan n_2 adalah saiz sampel bagi sampel pertama dan sampel kedua, $\overline{X}_1$ dan $\overline{X}_2$ adalah min bagi saiz sampel μ_1 dan μ_2 ialah min populasi bagi populasi pertama dan populasi kedua. Darjah kebebasan bagi ujian ini adalah $df = n_1 + n_2 - 2$. Nilai-p bagi ujian ini dikira dengan menggunakan persamaan $p = Kb\left(t > T_{\frac{\alpha}{2}, n_1+n_2-2}\right)$. Apabila nilai-*p* lebih besar daripada 0.05, maka hipotesis nol gagal ditolak. Oleh itu, min antara dua sampel adalah sama. Manakala, apabila nilai-*p* lebih kecil daripada 0.05, hipotesis nol ditolak, maka min antara dua sampel adalah tidak sama.

KAEDAH PENGUJIAN HIPOTESIS BERBILANG

Selepas ujian-*t* dijalankan dan menghasilkan nilai-*p,* nilai-*p* tersebut digunakan dalam pengujian hipotesis berbilang. Pengujian hipotesis berbilang merujuk kepada pengujian yang melibatkan ujian serentak lebih daripada satu hipotesis dan bilangan ujian, *m* yang banyak. Pengujian ini membuat perbandingan nilai-*p* dengan nilai aras keyakinan α yang telah dilaraskan untuk mengawal kadar Ralat Jenis I bagi kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR). Bagi pengujian hipotesis berbilang, hipotesis nol dan hipotesis alternatif merujuk kepada hipotesis bagi kesamaan min bagi dua sampel dan dinyatakan terlebih dahulu seperti berikut:

$$H_0 : \{H_{01}, H_{02}, ..., H_{0m}\}$$

$$H_1 : \{H_{11}, H_{12}, ..., H_{1m}\}$$

UJIAN *BONFERRONI*

Menurut Perneger (1998), ujian *Bonferroni* merupakan salah satu daripada pelarasan asas digunakan untuk kadar Ralat Jenis I serta mengawal kadar ralat berkumpulan (FWER). Ujian Bonferroni melaraskan aras *α* bagi statistik ujian yang dianggap sebagai signifikan berdasarkan jumlah bilangan ujian yang dijalankan dan mengandaikan statistik ujian tersebut adalah tidak bersandar. Apabila ujian yang melibatkan hipotesis nol H_1, i = 1, 2, ..., bagi mengekalkan kadar Ralat Jenis I yang dilaraskan dengan aras *α* secara serentak dan setiap nilai-p akan dibandingkan dengan $\frac{\alpha}{m}$. Ujian *Bonferroni* mengandaikan setiap *m* hipotesis adalah benar dan kadar Ralat Jenis I akan berlaku jika nilai-$p \le \frac{\alpha}{m}$ bagi salah satu daripada hipotesis yang benar. Ujian *Bonferroni* merupakan ujian yang mudah untuk digunakan dan andaian bagi setiap ujian adalah bersandar dipenuhi. Algoritma bagi ujian *Bonferroni* adalah seperti berikut: Pertama, bilangan ujian, *m* dijalankan. Kedua, menghitung nilai-p dari setiap ujian dan Ketiga, menolak hipotesis dan dikatakan signifikan jika nilai-$p \le \frac{\alpha}{m}$.

UJIAN *HOLM*

Ujian *Holm* merupakan kaedah yang digunakan dalam kes yang sama dengan ujian *Bonferroni* tetapi ujian ini lebih baik digunakan dan dirumuskan oleh Holm (1979). Ujian *Holm* adalah kaedah yang berurutan daripada ujian *Bonferroni* yang sentiasa menolak hipotesis nol lebih banyak daripada ujian *Bonferroni* dan mengawal aras *α* dengan baik serta mengandaikan setiap ujian adalah benar. Algoritma bagi ujian *Holm* adalah seperti berikut:

Pertama, nilai-*p* dari *m* hipotesis yang dijalankan. Kedua, menyusun nilai-p_j , $p_j = p_{1,} p_{2,} ..., p_{m,}$ secara menaik. Ketiga, membuat perbandingan nilai-p_j dengan $\frac{\alpha}{m-j+1}$, j = 1, 2, 3,…, *m*, *m* ialah bilangan ujian dan nilai-*p* yang dikatakan signifikan apabila nilai-P_j kurang daripada $\frac{\alpha}{m-j+1}$ dan Keempat, bermula daripada nilai-*p* yang kecil dan digunakan untuk membuat perbandingan daripada *j*=1 hingga *m* ujian sehingga terjumpa nilai-*p* yang tidak signifikan yang pertama iaitu gagal menolak hipotesis nol.

UJIAN *HOCHBERG*

Menurut Hochberg (1988), ujian *Hochberg* adalah sama dengan ujian *Holm*. Perbezaan antara dua ujian ini adalah ujian *Hochberg* menyusun nilai-*p* secara menurun dan lebih berkuasa berbanding dengan ujian *Holm*. Walaupun ujian ini amat rumit dalam pengiraan tetapi ujian *Hochberg* dikatakan lebih baik kerana dapat mengawal kadar ralat berkumpulan (FWER) dengan lebih baik berbanding ujian *Bonferroni* dan ujian *Holm*. Algoritma bagi ujian *Hochberg* adalah seperti berikut:

Pertama, menyusun nilai-p_j, $p_j = p_{1,} p_{2,} ..., p_m$, secara menurun. Kedua, membuat perbandingan nilai-p_j dengan $\frac{\alpha}{m-j+1}$, j = 1, 2, 3, …, *m*, *m* ialah bilangan ujian dan Ketiga, Langkah 2 diteruskan dan diberhentikan jika terdapat nilai-*p* yang pertama kurang daripada $\frac{\alpha}{m-j+1}$ dan nilai-*p* yang seterusnya dikatakan signifikan.

UJIAN *HOMMEL*

Menurut Hommel (1988), ujian Hommel adalah ujian yang lebih baik dan sedikit rumit dalam pengiraannya tetapi ujian ini adalah sebagai alternatif dan kurang digunakan. Ujian Hommel akan menolak hipotesis nol lebih banyak berbanding ujian Hochberg dan ujian ini mudah dengan menggunakan teknologi pengaturcaraan. Algoritma bagi ujian ini adalah seperti berikut:

Pertama, menolak kesemua nilai-$p \le \frac{\alpha}{j}$. Kedua, $J = max_{i=1,2,...,m}\{p_{m-i+k} > \frac{k\alpha}{i}$ for $k = 1, 2, \ldots, i\}$ dan Ketiga, jika tidak terdapat nilai yang maksimum, kesemua hipotesis nol akan ditolak ataupun menolak $\left\{H_{0i} : p_i \le \frac{\alpha}{k}\right\}$.

UJIAN *BENJAMINI-HOCHBERG*

Benjamini dan Hochberg (1995) merupakan orang pertama yang memperkenalkan kaedah yang mengawal kadar penemuan palsu (FDR) dan ujian ini digunakan sehingga kini. Ujian *Benjamini-Hochberg* mengandaikan nilai-*p* adalah tidak bersandar atau positif bersandar antara ujian serta mengawal FDR pada aras $\pi_0\alpha$, yang π_0 ialah perkadaran bagi hipotesis benar. Berikut adalah algoritma bagi ujian *Benjamini-Hochberg*:

Pertama, nilai-*p*, p_j disusun secara menaik. Kedua, mencari ujian yang pangkat tinggi, *j* bagi $p_j \leq \frac{j\alpha}{lm}$ dan ujian yang ke-1, 2,.., *j* dianggap sebagai signifikan.

UJIAN *BENJAMINI-YEKUTIELI*

Ujian *Benjamini-Yekutieli* diperkenalkan oleh Yoav Benjamini dan Daniel Yekutieli pada tahun 2001 kerana menyatakan ujian *Benjamini-Hochberg* mengawal dengan baik dan sesuai untuk digunakan dalam FDR dan dalam pengujian hipotesis berbilang. Oleh itu, Benjamini dan Yekutieli memperkenalkan ujian ini sebagai salah satu kaedah untuk mengawal FDR apabila ujian yang dijalankan adalah bersandar dengan hipotesis yang benar. Algoritma bagi ujian *Benjamini-Yekutieli* adalah seperti berikut:

Pertama, menyusun nilai-*p*, $p_{(1)}$, $p_{(2)}$, ..., $p_{(m)}$ secara menaik. Kedua, mencari ujian yang pangkat tinggi, *j* bagi $p_j \leq \frac{j\alpha}{lm}$,dengan $l = \sum_{k=1}^{m} \frac{1}{k}$ dan Ketiga, ujian yang ke-1, 2,…, *j* dianggap sebagai signifikan.

KADAR RALAT JENIS I

Definisi bagi kebarangkalian Ralat Jenis I ialah α = *Kb*(Tolak H_o | Walhal ia benar). Keadaan ini berlaku apabila sesuatu penyataan yang diperoleh benar sedangkan ia palsu dan kadar Ralat Jenis I dianggap sebagai ralat yang tepat. Dalam pengujian hipotesis berbilang, kebanyakan pengkaji daripada bidang perubatan dan biologi menggunakan kadar Ralat Jenis I sebagai keputusan pentaabiran statistik dan mengawal kadar Ralat Jenis I. Apabila melakukan pengujian hipotesis berbilang dan jika mengikut peraturan yang sama untuk menolak ujian yang tidak bersandar kebarangkalian untuk membuat Ralat Jenis I adalah tinggi daripada aras nominal yang digunakan untuk setiap ujian terutamanya bilangan ujian, *m* yang banyak. Oleh itu, kajian ini menggunakan beberapa pendekatan untuk mengawal kadar Ralat Jenis I iaitu kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR). Selain itu, kajian ini menggunakan aras keertian α=0.05, 0.01 dan 0.1 untuk membuat perbandingan antara statistik ujian. Jika sesuatu ujian mendapat keputusan kadar Ralat Jenis I yang menghampiri nilai α dianggap sangat baik dan dapat mengawal FWER dan FDR. Bradley (1978) mengemukakan alasan bahawa prosedur statistik dianggap sebagai teguh apabila nilai α dengan syarat kebarangkalian ralat jenis satu berada dalam selang 0.5α < $\hat{\alpha} < 1.5\alpha$ dikenali sebagai kriteria liberal Bradley.

KADAR RALAT BERKUMPULAN (FWER)

Kadar ralat berkumpulan (FWER) bermaksud kebarangkalian melakukan sekurang-kurangnya satu kadar Ralat Jenis I iaitu FWER = *Kb*(V≥1). Selain itu, kadar Ralat Jenis I berkumpulan (FWER) digunakan dalam pelbagai bidang seperti biologi dan masih digunakan sehingga kini. Pada mulanya, kaedah pengujian hipotesis berbilang hanya menumpukan dalam mengawal kadar ralat berkumpulan (FWER) berbanding kadar ralat yang lain. Terdapat beberapa kaedah pengujian hipotesis berbilang yang mengawal FWER iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg* dan ujian *Hommel*. Ujian-ujian tersebut menolak hipotesis nol sedikit dan gagal menolak hipotesis nol apabila mempunyai data yang banyak. Kebanyakan pengkaji mengkritik FWER kerana sangat ketat dan konservatif. Oleh itu, penyelidik banyak menggunakan FDR dan kaedah FDR untuk mengawal kadar Ralat Jenis I. Rumus bagi kadar ralat berkumpulan (FWER) adalah seperti berikut:

$$\text{FWER} = Kb(\text{V} \geq 1)$$

$$\text{FWER} = \frac{V}{m_0}$$

dengan V ialah bilangan kadar ralat jenis satu atau positif palsu dan m_0 ialah bilangan hipotesis benar.

KADAR PENEMUAN PALSU (FDR)

Kebanyakan pendekatan moden yang berkaitan dengan pengujian hipotesis berbilang fokus dalam mengawal kadar penemuan palsu (FDR). Kadar penemuan palsu kurang konservatif berbanding dengan FWER dan kebanyakan kaedah FDR masih dalam pengubahsuaian. Definisi kadar penemuan palsu (FDR) ialah perkadaran kadar Ralat Jenis I antara hipotesis yang ditolak, FDR = $E\left(\frac{V}{R}\right)$. Apabila kesemua hipotesis nol benar, FDR akan setara dengan FWER. Walau bagaimanapun, bilangan hipotesis nol yang benar adalah kurang daripada bilangan hipotesis. Apabila $m_0 < m$, FDR lebih kecil atau sama dengan FWER. Oleh itu, ujian bagi FWER boleh digunakan untuk mengawal FDR. Terdapat beberapa ujian yang mengawal FDR iaitu ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli*. Rumus bagi kadar penemuan palsu (FDR) adalah seperti berikut:

$$FDR = E\left(\frac{V}{R}\right)$$

dengan *V* ialah bilangan kadar Ralat Jenis I atau positif palsu dan *R* ialah bilangan hipotesis yang ditolak.

Hasil dan Perbincangan

Kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR) digunakan dalam kajian ini untuk mengawal kadar Ralat Jenis I. Ralat Jenis I diukur dengan menggunakan kriteria liberal Bradley dengan syarat kebarangkalian Ralat Jenis I mesti berada dalam selang $0.5\alpha < \hat{\alpha} < 1.5\alpha$ dan ujian tersebut dianggap sebagai teguh dan bagus digunakan dengan memenuhi kesemua keadaan. Kajian ini menggunakan tiga aras keertian yang berbeza iaitu 0.05, 0.01 dan 0.1. Pada aras keertian 0.05, Ralat Jenis I mesti berada dalam selang $0.025 \leq \hat{\alpha} \leq 0.075$ bagi aras keertian 0.01 pula berada dalam selang $0.005 \leq \hat{\alpha} \leq 0.015$ dan aras keertian 0.1 berada dalam selang $0.5\alpha \leq \hat{\alpha} \leq 0.15$.

Kebarangkalian Ralat Jenis I bagi Kadar Penemuan Palsu (FDR)

Keputusan simulasi bagi kebarangkalian Ralat Jenis I bagi saiz sampel sama iaitu {10,10}, varians sama iaitu {1,1} dan min sama iaitu {0,0} bagi aras keertian 0.05, 0.01 dan 0.1 serta bilangan ujian yang berbeza adalah kosong. Hal ini kerana kesemua ujian hipotesis berbilang iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* gagal menolak hipotesis nol dengan nilai-*p* lebih besar dengan aras keertian. Oleh itu, terbukti yang menyatakan kesemua hipotesis nol adalah benar kerana min bagi sampel pertama dan kedua adalah sama.

Keputusan simulasi bagi saiz sampel sama iaitu {10,10} dengan menggunakan varians yang sama {1,1} dan min yang tidak sama {0,1} untuk ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* bagi taburan Normal, taburan Khi-Kuasa Dua dan taburan Lognormal bagi aras keertian 0.05 dalam Jadual 1.

JADUAL 1. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.05

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
	5	*Bonferroni*	0.40	0.00	0.20
		Holm	0.80	0.00	0.20
		Hochberg	0.80	0.00	0.20
		Hommel	0.80	0.00	0.20
		Benjamini-Hochberg	0.80	0.40	0.20
		Benjamini-Yekutieli	0.60	0.00	0.20
	10	*Bonferroni*	0.30	0.00	0.10
		Holm	0.30	0.00	0.10
		Hochberg	0.30	0.00	0.10
		Hommel	0.30	0.00	0.10
		Benjamini-Hochberg	0.70	0.00	0.20
		Benjamini-Yekutieli	0.30	0.00	0.10
	20	*Bonferroni*	0.15	**0.05**	**0.05**
		Holm	0.15	**0.05**	**0.05**
		Hochberg	0.15	**0.05**	**0.05**
		Hommel	0.15	**0.05**	**0.05**
		Benjamini-Hochberg	0.50	**0.05**	**0.05**
		Benjamini-Yekutieli	**0.05**	0.00	**0.05**
0.05	25	*Bonferroni*	**0.04**	**0.04**	0.12
		Holm	**0.04**	**0.04**	0.12
		Hochberg	**0.04**	**0.04**	0.12
		Hommel	0.08	**0.04**	0.12
		Benjamini-Hochberg	0.44	**0.04**	0.20
		Benjamini-Yekutieli	**0.04**	0.00	**0.04**
	30	*Bonferroni*	**0.03**	**0.03**	**0.06**
		Holm	**0.07**	**0.03**	**0.06**
		Hochberg	**0.07**	**0.03**	**0.06**
		Hommel	**0.07**	**0.03**	**0.06**
		Benjamini-Hochberg	0.43	**0.03**	0.20
		Benjamini-Yekutieli	**0.03**	0.00	**0.03**

50	*Bonferroni*	**0.06**	0.00	**0.06**
	Holm	**0.06**	0.00	**0.06**
	Hochberg	**0.06**	0.00	**0.06**
	Hommel	0.08	0.00	**0.06**
	Benjamini-Hochberg	0.30	0.00	0.18
	Benjamini-Yekutieli	**0.06**	0.00	**0.04**

Nota: Nilai yang ditanda hitam menunjukkan Ralat Jenis I dalam selang $0.025 \leq \hat{\alpha} \leq 0.075$

Jadual 1 menunjukkan keputusan bagi kebarangkalian Ralat Jenis I bagi saiz sampel dengan kumpulan varians yang sama, min yang berbeza dan bilangan ujian yang berbeza bagi taburan yang berbeza iaitu taburan normal, taburan khi-kuasa dua dan taburan log-normal bagi aras keertian 0.05. Bagi taburan normal, beberapa ujian berada dalam selang kriteria liberal Bradley apabila bilangan ujian yang digunakan tinggi iaitu 20 ujian dan ke atas. Ujian *Benjamini-Yekutieli* dapat mengawal dengan baik apabila bilangan ujian 20, 25, 30 dan 50 kerana berada dalam selang kriteria liberal Bradley berbanding ujian-ujian yang lain. Manakala bagi ujian *Bonferroni*, ujian *Holm* dan ujian *Hochberg* masing-masing teguh dengan kriteria liberal Bradley pada bilangan ujian 25, 30 dan 50. Di samping itu, ujian *Hommel* teguh apabila bilangan ujian yang digunakan besar iaitu pada bilangan ujian 30. Bagi ujian *Benjamini-Hochberg* tidak sesuai digunakan untuk taburan normal kerana nilai kebarangkalian bagi ujian ini sangat tinggi dan tidak dapat mengawal dengan baik.

Bagi taburan khi-kuasa dua, ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel* dan ujian *Benjamini-Hochberg* menunjukkan berada dalam selang kriteria liberal Bradley bagi bilangan ujian 20, 25 dan 30 sahaja dan dapat mengawal Ralat Jenis I dengan baik. Walau bagaimanapun, apabila bilangan ujian yang kecil dan sangat besar iaitu 5, 10 dan 50 ujian-ujian ini tidak berada dalam selang. Manakala, bagi ujian *Benjamini-Yekutieli* tidak sesuai digunakan bagi taburan ini kerana tidak mendapat mengawal kadar penemuan palsu (FDR) dan kebarangkalian Ralat Jenis I tidak berada dalam selang kriteria liberal Bradley untuk kesemua bilangan ujian yang digunakan.

Manakala, bagi taburan log-normal, pada bilangan ujian 20, kesemua ujian berada dalam selang kriteria liberal Bradley. Ujian *Benjamini-Yekutieli* dapat mengawal kadar penemuan palsu (FDR) dengan baik kerana apabila bilangan ujian 20, 25, 30 dan 50 menunjukkan kebarangkalian Ralat Jenis I berada dalam selang kriteria liberal Bradley. Bagi ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg* dan ujian *Hommel* masing-masing berada dalam selang kriteria liberal Bradley apabila bilangan ujian 30 dan 50. Manakala bagi ujian *Benjamini-Hochberg,* hanya dapat mengawal kadar penemuan palsu (FDR) apabila kebarangkalian Ralat Jenis I bagi bilangan ujian 20 dan bagi bilangan ujian yang lain tidak berada dalam selang kriteria liberal Bradley.

Keputusan simulasi bagi saiz sampel sama iaitu {10,10} dengan menggunakan varians yang sama {1,1} dan min yang tidak sama {0,1} untuk ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* bagi taburan normal, taburan khi-kuasa dua dan taburan log-normal bagi aras keertian 0.01 ditunjukkan seperti di Jadual 2.

JADUAL 2. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.01

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
	5	*Bonferroni*	0.20	0.00	0.20
		Holm	0.20	0.00	0.20
		Hochberg	0.20	0.00	0.20
		Hommel	0.20	0.00	0.20
		Benjamini-Hochberg	0.20	0.00	0.20
		Benjamini-Yekutieli	0.20	0.00	0.20
	10	*Bonferroni*	0.10	0.00	0.10
		Holm	0.10	0.00	0.10
		Hochberg	0.10	0.00	0.10
		Hommel	0.10	0.00	0.10
		Benjamini-Hochberg	0.10	0.00	0.10
0.01		*Benjamini-Yekutieli*	0.10	0.00	0.10

20	*Bonferroni*	0.05	0.00	0.05
	Holm	0.05	0.00	0.05
	Hochberg	0.05	0.00	0.05
	Hommel	0.05	0.00	0.05
	Benjamini-Hochberg	0.05	0.00	0.05
	Benjamini-Yekutieli	0.05	0.00	0.05
25	*Bonferroni*	0.04	0.00	0.04
	Holm	0.04	0.00	0.04
	Hochberg	0.04	0.00	0.04
	Hommel	0.04	0.00	0.04
	Benjamini-Hochberg	0.04	0.00	0.04
	Benjamini-Yekutieli	0.04	0.00	0.04
30	*Bonferroni*	0.03	0.00	0.03
	Holm	0.03	0.00	0.03
	Hochberg	0.03	0.00	0.03
	Hommel	0.03	0.00	0.03
	Benjamini-Hochberg	0.03	0.00	0.03
	Benjamini-Yekutieli	0.03	0.00	0.03
50	*Bonferroni*	0.40	0.00	0.00
	Holm	0.40	0.00	0.00
	Hochberg	0.40	0.00	0.00
	Hommel	0.40	0.00	0.00
	Benjamini-Hochberg	0.06	0.00	0.04
	Benjamini-Yekutieli	0.40	0.00	0.00

Nota: Nilai yang ditanda hitam menunjukkan Ralat Jenis I dalam selang $0.005 \leq \hat{\alpha} \leq 0.015$

Jadual 2 menunjukkan keputusan bagi kebarangkalian Ralat Jenis I bagi kes kadar penemuan palsu (FDR) dengan menggunakan saiz sampel yang sama dengan kumpulan varians yang sama, kumpulan min yang berbeza dan bilangan ujian yang berbeza bagi taburan Normal, taburan Khi-Kuasa Dua dan taburan Lognormal bagi aras keertian 0.01. Kesemua ujian iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* tidak berada dalam selang kriteria liberal Bradley bagi kesemua taburan dan bilangan ujian. Oleh itu, sebanyak 99% keyakinan bahawa hipotesis nol gagal ditolak dan dapat dibuat kesimpulan min bagi dua sampel adalah sama.

Keputusan simulasi bagi saiz sampel sama iaitu {10,10} dengan menggunakan varians yang sama {1,1} dan min yang tidak sama {0,1} untuk ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* bagi taburan Normal, taburan Khi-Kuasa Dua dan taburan LogNormal bagi aras keertian 0.1 ditunjukkan dalam Jadual 3.

JADUAL 3. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.1

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
	5	*Bonferroni*	0.60	0.40	0.20
		Holm	0.80	0.40	0.20
		Hochberg	0.80	0.40	0.20
		Hommel	0.80	0.40	0.20
		Benjamini-Hochberg	0.80	0.40	0.80
		Benjamini-Yekutieli	0.80	0.40	0.20
	10	*Bonferroni*	0.40	0.00	0.20
		Holm	0.50	0.00	0.20
		Hochberg	0.50	0.00	0.20
		Hommel	0.60	0.00	0.20
		Benjamini-Hochberg	0.70	0.30	0.60
		Benjamini-Yekutieli	0.60	0.00	**0.10**

	20	*Bonferroni*	0.20	**0.05**	**0.05**
		Holm	0.20	**0.05**	**0.05**
		Hochberg	0.20	**0.05**	**0.05**
		Hommel	0.20	**0.05**	**0.05**
		Benjamini-Hochberg	0.20	0.30	**0.15**
		Benjamini-Yekutieli	0.60	0.00	**0.05**
0.1	25	*Bonferroni*	0.16	0.04	**0.12**
		Holm	0.16	0.04	**0.12**
		Hochberg	0.16	0.04	**0.12**
		Hommel	0.20	0.04	**0.12**
		Benjamini-Hochberg	0.56	0.04	0.20
		Benjamini-Yekutieli	0.16	0.00	**0.12**
	30	*Bonferroni*	**0.13**	0.03	**0.10**
		Holm	0.17	0.03	**0.10**
		Hochberg	0.17	0.03	**0.10**
		Hommel	0.17	0.03	**0.10**
		Benjamini-Hochberg	0.60	0.03	0.20
		Benjamini-Yekutieli	0.17	0.00	**0.10**
	50	*Bonferroni*	**0.12**	0.02	**0.10**
		Holm	**0.14**	0.02	**0.10**
		Hochberg	**0.14**	0.02	**0.10**
		Hommel	0.16	0.02	**0.10**
		Benjamini-Hochberg	0.56	0.02	0.22
		Benjamini-Yekutieli	0.18	0.00	**0.10**

Nota: Nilai yang ditanda hitam menunjukkan Ralat Jenis I dalam selang $0.005 \leq \hat{\alpha} \leq 0.015$

Berdasarkan Jadual 3, didapati beberapa ujian dapat mengawal kadar penemuan palsu (FDR) apabila bilangan ujian yang besar iaitu 30 dan 50 bagi sampel yang mengikuti taburan Normal. Ujian *Bonferroni* didapati lebih baik berbanding ujian-ujian lain dalam mengawal kadar penemuan palsu (FDR) pada bilangan ujian 30 dan 50. Manakala, kebarangkalian Ralat Jenis I bagi ujian *Holm* dan ujian *Hochberg* berada dalam selang kriteria liberal Bradley pada bilangan ujian 50 sahaja. Bagi ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli*, kebarangkalian Ralat Jenis I tidak berada dalam selang kriteria liberal Bradley. Oleh itu, ujian-ujian tersebut tidak sesuai digunakan bagi sampel yang mengikuti taburan Normal pada aras keertian 0.1 kerana tidak dapat mengawal kadar penemuan palsu (FDR) dengan baik.

Bagi taburan khi-kuasa dua, ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg* dan ujian *Hommel* dapat mengawal kadar penemuan palsu (FDR) pada bilangan ujian 20 sahaja. Didapati, ujian *Benjamini-Hochberg* tidak dapat mengawal FDR bagi kesemua bilangan ujian yang digunakan. Manakala ujian *Benjamini-Yekutieli* tidak sesuai digunakan apabila menggunakan data simulasi daripada taburan Khi-Kuasa Dua. Hal ini kerana terdapat nilai kosong pada keputusan kebarangkalian Ralat Jenis I dan tidak berada dalam selang kriteria liberal Bradley serta tidak dapat mengawal kadar penemuan palsu (FDR).

Manakala bagi taburan Lognormal pula, ujian *Benjamini-Yekutieli* dapat mengawal kadar penemuan palsu (FDR) dengan baik pada bilangan ujian yang kecil dan besar iaitu 10, 20, 25, 30 dan 50 ujian. Manakala bagi ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg* dan ujian *Hommel*, ujian-ujian ini dapat mengawal FDR pada bilangan ujian 20, 25, 30 dan 50. Hanya pada bilangan ujian 20 sahaja ujian *Benjamini-Hochberg* dapat mengawal FDR tetapi ujian ini tidak sesuai digunakan apabila sampel mengikuti taburan Lognormal.

KEBARANGKALIAN RALAT JENIS I BAGI KADAR RALAT BERKUMPULAN (FWER)

Keputusan simulasi bagi kadar ralat berkumpulan (FWER) bagi kebarangkalian Ralat Jenis I bagi saiz sampel sama iaitu {10,10}, varians sama iaitu {1,1} dan min sama iaitu {0,0} bagi aras keertian 0.05, 0.01 dan 0.1 serta bilangan ujian iaitu 5, 10, 20, 25, 30 dan 50 adalah kosong. Hal ini kerana kesemua ujian hipotesis berbilang iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* gagal menolak hipotesis nol dengan nilai-p lebih besar dengan aras keertian 0.05, 0.01 dan 0.1. Oleh itu, terbukti kesemua hipotesis nol adalah benar kerana min bagi sampel pertama dan kedua adalah sama.

Jadual 4 menunjukkan keputusan simulasi ujian-t dengan pengujian hipotesis berbilang bagi ujian

Bonferroni, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* mengikut keadaan berikut: saiz sampel sama {10,10} dengan menggunakan varians sama {1,1} dan min berbeza {0,1} untuk taburan Normal, taburan Khi-Kuasa Dua dan taburan LogNormal pada aras keertian 0.05.

JADUAL 4. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.05

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
	5	*Bonferroni*	0.20	0	0.10
		Holm	0.40	0	0.10
		Hochberg	0.40	0	0.10
		Hommel	0.40	0	0.10
		Benjamini-Hochberg	0.40	0.20	0.10
		Benjamini-Yekutieli	0.30	0	0.10
	10	*Bonferroni*	0.15	0	**0.05**
		Holm	0.15	0	**0.05**
		Hochberg	0.15	0	**0.05**
		Hommel	0.15	0	**0.05**
		Benjamini-Hochberg	0.35	0	0.10
		Benjamini-Yekutieli	0.15	0	**0.05**
	20	*Bonferroni*	**0.075**	**0.025**	**0.025**
		Holm	**0.075**	**0.025**	**0.025**
		Hochberg	**0.075**	**0.025**	**0.025**
		Hommel	**0.075**	**0.025**	**0.025**
		Benjamini-Hochberg	0.25	**0.025**	**0.025**
		Benjamini-Yekutieli	**0.025**	0	**0.025**
0.05	25	*Bonferroni*	0.02	0.02	**0.06**
		Holm	0.02	0.02	**0.06**
		Hochberg	0.02	0.02	**0.06**
		Hommel	**0.04**	0.02	**0.06**
		Benjamini-Hochberg	0.22	0.02	0.10
		Benjamini-Yekutieli	0.02	0	0.02
	30	*Bonferroni*	**0.02**	0.017	**0.03**
		Holm	**0.03**	0.017	**0.03**
		Hochberg	**0.03**	0.017	**0.03**
		Hommel	**0.03**	0.017	**0.03**
		Benjamini-Hochberg	0.22	0.017	0.10
		Benjamini-Yekutieli	0.02	0	0.017
	50	*Bonferroni*	**0.03**	0	**0.03**
		Holm	**0.03**	0	**0.03**
		Hochberg	**0.03**	0	**0.03**
		Hommel	**0.04**	0	**0.03**
		Benjamini-Hochberg	0.15	0	0.09
		Benjamini-Yekutieli	**0.03**	0	0.02

Nota: Nilai yang ditanda hitam menunjukkan Ralat Jenis I dalam selang $0.025 \leq \hat{\alpha} \leq 0.075$

Berdasarkan Jadual 4, didapati ujian *Hommel* dapat mengawal kadar ralat berkumpulan (FWER) dengan baik bagi sampel yang mengikuti taburan Normal pada bilangan ujian 20, 25, 30 dan 50 berbanding dengan ujian-ujian lain. Manakala ujian *Bonferroni* dapat mengawal FWER pada semua taburan pada bilangan ujian 20 sahaja dan hanya taburan Normal dan Lognormal sahaja pada bilangan ujian 50. Begitu juga ujian *Benjamini-Yekutieli* yang hanya dapat mengawal FWER pada bilangan ujian 20 untuk taburan Normal dan Lognormal dan hanya taburan

Normal sahaja bagi bilangan ujian 50. Ujian *Holm* dan ujian *Hochberg* pula hanya dapat mengawal kadar ralat berkumpulan (FWER) bagi taburan Lognormal sahaja apabila kebarangkalian Ralat Jenis I berada dalam kriteria liberal Bradley pada bilangan ujian 20, 30 dan 50.

Hanya pada bilangan ujian 20 sahaja kesemua ujian (kecuali ujian *Benjamini-Yekutieli*) dapat mengawal kadar ralat berkumpulan (FWER) bagi taburan khi-kuasa dua. Ujian-ujian lain gagal menolak hipotesis nol. Manakala, bagi taburan Lognormal, hampir kesemua ujian bagi pengujian hipotesis berbilang dapat mengawal kadar FWER pada semua bilangan ujian kecuali pada bilangan ujian yang kecil iaitu 5.

Keadaan ujian pada situasi saiz sampel sama iaitu {10,10} dengan menggunakan varians yang sama {1,1} dan kumpulan min yang tidak sama {0,1} untuk semua ujian dan tiga taburan dipilih ditunjukkan dalam Jadual 5.

JADUAL 5. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.01

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
	5	*Bonferroni*	0.10	0.00	0.10
		Holm	0.10	0.00	0.10
		Hochberg	0.10	0.00	0.10
		Hommel	0.10	0.00	0.10
		Benjamini-Hochberg	0.10	0.00	0.10
		Benjamini-Yekutieli	0.10	0.00	0.10
	10	*Bonferroni*	0.05	0	0.05
		Holm	0.05	0	0.05
		Hochberg	0.05	0	0.05
		Hommel	0.05	0	0.05
		Benjamini-Hochberg	0.05	0	0.05
		Benjamini-Yekutieli	0.05	0	0.05
	20	*Bonferroni*	0.025	0	0.025
		Holm	0.025	0	0.025
		Hochberg	0.025	0	0.025
		Hommel	0.025	0	0.025
		Benjamini-Hochberg	0.025	0	0.025
		Benjamini-Yekutieli	0.025	0	0
0.01	25	*Bonferroni*	0.02	0	0.02
		Holm	0.02	0	0.02
		Hochberg	0.02	0	0.02
		Hommel	0.02	0	0.02
		Benjamini-Hochberg	0.02	0	0.02
		Benjamini-Yekutieli	0.02	0	0
	30	*Bonferroni*	0.02	0	0.017
		Holm	0.02	0	0.017
		Hochberg	0.02	0	0.017
		Hommel	0.02	0	0.017
		Benjamini-Hochberg	0.02	0	0.017
		Benjamini-Yekutieli	0.02	0	0
	50	*Bonferroni*	0.02	0	0
		Holm	0.02	0	0
		Hochberg	0.02	0	0
		Hommel	0.02	0	0
		Benjamini-Hochberg	0.02	0	0.02
		Benjamini-Yekutieli	0.02	0	0

Didapati, kesemua ujian tidak dapat mengawal kadar ralat berkumpulan (FWER) dengan baik pada aras keertian 0.01 bagi semua taburan dan semua bilangan ujian. Pada aras keertian 0.10 pula, saiz sampel {10,10} dengan varians sama {1,1} dan kumpulan min berbeza {0,1} yang mengikuti taburan Khi-Kuasa Dua sahaja yang dapat mengawal FWER pada bilangan ujian yang kecil iaitu 5 ujian bagi semua ujian pengujian hipotesis berbilang seperti ditunjukkan pada Jadual 6.

JADUAL 6. Kebarangkalian bagi ralat jenis I bagi min berbeza pada aras keertiaan 0.1

Alfa	Bilangan ujian	Ujian	Taburan normal	Taburan Khi-Kuasa Dua	Taburan Lognormal
0.10	5	*Bonferroni*	0.30	**0.10**	**0.10**
		Holm	0.40	**0.10**	**0.10**
		Hochberg	0.40	**0.10**	**0.10**
		Hommel	0.40	**0.10**	**0.10**
		Benjamini-Hochberg	0.40	**0.10**	0.40
		Benjamini-Yekutieli	0.40	**0.10**	**0.10**
	10	*Bonferroni*	0.20	0	**0.10**
		Holm	0.25	0	**0.10**
		Hochberg	0.25	0	**0.10**
		Hommel	0.30	0	**0.10**
		Benjamini-Hochberg	0.35	**0.15**	0.30
		Benjamini-Yekutieli	0.30	0	**0.05**
	20	*Bonferroni*	**0.10**	0.025	0.025
		Holm	**0.10**	0.025	0.025
		Hochberg	**0.10**	0.025	0.025
		Hommel	**0.10**	0.025	0.025
		Benjamini-Hochberg	0.30	***0.15***	0.075
		Benjamini-Yekutieli	**0.10**	0	0.025
	25	*Bonferroni*	**0.08**	0.02	**0.06**
		Holm	**0.08**	0.02	**0.06**
		Hochberg	**0.08**	0.02	**0.06**
		Hommel	**0.10**	0.02	**0.06**
		Benjamini-Hochberg	0.28	0.02	**0.10**
		Benjamini-Yekutieli	**0.08**	0	**0.06**
	30	*Bonferroni*	**0.07**	0.017	**0.05**
		Holm	**0.08**	0.017	**0.05**
		Hochberg	**0.08**	0.017	**0.05**
		Hommel	**0.08**	0.017	**0.05**
		Benjamini-Hochberg	0.30	0.017	**0.10**
		Benjamini-Yekutieli	**0.08**	0	**0.05**
	50	*Bonferroni*	**0.06**	0.01	**0.05**
		Holm	**0.07**	0.01	**0.05**
		Hochberg	**0.07**	0.01	**0.05**
		Hommel	**0.08**	0.01	**0.05**
		Benjamini-Hochberg	0.28	0.01	**0.11**
		Benjamini-Yekutieli	**0.09**	0	**0.05**

Nota: Nilai yang ditanda hitam menunjukkan Ralat Jenis I dalam selang $0.05 \le \hat{\alpha} \le 0.15$

Bagi taburan Lognormal, semua ujian pada semua bilangan ujian (kecuali 20) dapat mengawal FWER pada aras 90% tetapi berbeza dengan taburan Khi-Kuasa Dua yang hanya signifikan pada bilangan ujian yang kecil iaitu 5 pada kesemua ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli.* Hanya ujian *Benjamini-Hochberg* yang signifikan pada bilangan ujian 10 dan 20. Sebaliknya pula bagi taburan Normal yang didapati semua ujian adalah signifikan iaitu dapat mengawal kadar ralat berkumpulan (FWER) pada bilangan ujian yang besar iaitu melebihi 20 ujian kecuali bagi ujian *Benjamini-Hochberg*. Ini bermakna ujian *Benjamini-Hochberg* hanya signifikan bagi taburan Khi-Kuasa Dua dan pada bilangan ujian yang kecil sahaja.

KESIMPULAN

Sesuatu ujian dikatakan terbaik apabila kebarangkalian Ralat Jenis I bagi kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR) berada dalam selang kriteria liberal Bradley iaitu $0.5\alpha < \hat{\alpha} < 1.5\alpha$ atau menghampiri aras keertian α. Bagi data tertabur secara Normal, Khi-Kuasa Dua dan Lognormal dengan kumpulan min yang sama iaitu {0,0} bagi aras keertian 0.05, 0.01 dan 0.1 serta bilangan ujian yang berbeza bagi kedua-dua kes iaitu kadar ralat berkumpulan (FWER) dan kadar penemuan palsu (FDR), didapati ia mempunyai kebarangkalian Ralat Jenis I sifar. Ini memberi kesimpulan bahawa kesemua ujian hipotesis berbilang iaitu ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* gagal menolak hipotesis nol dan tidak signifikan. Tiada perbezaan min antara kedua-dua kumpulan dan dapat dibuat kesimpulan bahawa kedua-dua kumpulan adalah daripada populasi taburan yang sama. Manakala keputusan ujian-*t* bagi berbilang pengujian dengan andaian min bagi kedua-dua sampel adalah berbeza didapati tidak signifikan pada 99% keyakinan pada semua keadaan ujian. Ini kerana aras keertian 0.01 tidak sesuai digunakan bagi keadaan-keadaan ini kerana aras keertian ini dikatakan sangat jitu iaitu 99% berbanding aras keertian yang lain.

Apabila menjalankan pengujian hipotesis berbilang bagi ujian-*t* yang kedua-dua sampel mempunyai nilai min yang berbeza, didapati ujian *Benjamini-Yekutieli* adalah ujian yang terbaik untuk digunakan bagi mengawal kadar penemuan palsu (FDR) dengan data tertabur secara Normal dan Lognormal pada semua keadaan ujian, tetapi pada aras keertian 90%, ujian *Benjamini-Yekutieli* dikatakan terbaik hanya apabila data tertabur secara Lognormal. Walau bagaimanapun, ujian *Benjamini-Yekutieli* tidak sesuai digunakan apabila data tertabur secara Khi-Kuasa Dua pada aras keyakinan 90% dan 95%. Manakala, ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel* dan ujian *Benjamini-Hochberg* adalah ujian terbaik yang boleh digunakan bagi menguji perbezaan min antara dua sampel bagi bilangan ujian 20, 25 dan 30 dengan keyakinan 95% dan 90% apabila sampel mengikuti taburan Khi-Kuasa Dua. Manakala ujian *Bonferroni* dapat mengawal kadar penemuan palsu (FDR) pada 90% keyakinan sahaja apabila data tertabur secara Normal dengan bilangan ujian 30 dan 50 berbanding dengan ujian lain.

Apabila sampel yang digunakan mempunyai nilai min yang berbeza dan bagi mengawal kadar ralat berkumpulan (FWER) bagi data tertabur secara Normal, Khi-Kuasa Dua dan Lognormal, ujian *Hommel* adalah ujian yang terbaik digunakan apabila bilangan ujian yang digunakan kecil dan besar bagi taburan Normal iaitu 20, 25, 30 dan 50 ujian, taburan Khi-Kuasa Dua dengan 20 ujian dan taburan Lognormal dengan 10, 20, 25, 30, 50 ujian kerana ia sentiasa konsisten pada aras keertian 0.05 dan 0.1. Oleh itu, ujian *Hommel* dapat dilihat sangat mudah digunakan apabila bilangan ujian yang digunakan kecil dan besar. Manakala ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* juga sesuai digunakan bagi keadaan-keadan tersebut tetapi ujian *Benjamini-Yekutieli* tidak sesuai digunakan apabila data tertabur secara taburan Khi-Kuasa Dua. Kesimpulannya, ujian yang terbaik bagi mengawal kadar penemuan palsu (FDR) ialah ujian *Benjamini-Yekutieli*. Manakala ujian *Hommel* dapat mengawal kadar ralat berkumpulan (FWER). Hal ini kerana ujian *Benjamini-Yekutieli* dan ujian *Hommel* dikatakan konsisten apabila kebarangkalian Ralat Jenis I sentiasa berada dalam selang kriteria liberal Bradley. Keputusan yang diperoleh daripada kajian ini selari dengan keputusan-keputusan daripada kajian lepas. Hal ini kerana menurut kajian lepas iaitu Benjamini dan Yekutieli (2001) menyatakan ujian *Benjamini-Yekutieli* mengawal kadar penemuan palsu (FDR). Menurut Wright (1992), ujian Hommel baik digunakan bagi mengawal kadar ralat berkumpulan (FWER). Maka dengan perbandingan pengujian hipotesis berbilang antara ujian *Bonferroni*, ujian *Holm*, ujian *Hochberg*, ujian *Hommel*, ujian *Benjamini-Hochberg* dan ujian *Benjamini-Yekutieli* dengan mengikut keadaan yang tertentu iaitu nilai α, bilangan ujian. m dan jenis taburan dapat dijalankan dengan menentukan ujian yang terbaik bagi kesemua keadaan. Ini dapat menjimatkan masa dan kos ketika menjalankan pengujian hipotesis berbilang bagi kumpulan sampel yang banyak terutama dalam bidang biologi atau perubatan.

PENGHARGAAN

Penghargaan kepada Universiti Kebangsaan Malaysia untuk geran penyelidikan GUP-2016-052.

RUJUKAN

Abdi, H. 2010. Holm's Sequential Bonferroni Procedure. Dlm. *Encyclopedia of Research Design*, disunting oleh Salkind, N. Thousand Oaks, CA: Sage.

Abdul Rahman Othman. & Lai Choo Heng. 2014. Sensitivity analysis of the refinement to the Mann-Whitney Test. *Sains Malaysiana* 43(7): 1095-1100.

Aickin, M. & Gensler, H. 1996. Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *America Journal of Public Health* 86(5): 726-728.

Benjamini, Y. & Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29: 1165-1188.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300.

Bradley, J.V. 1978. Robustness. *British Journal of Mathematics and Statistical Psychology* 31: 144-151.

Dudoit, S., Shaffer, J.P. & Boldrick, J.C. 2003. Multiple hypothesis testing in microarray experiments. *Statistical Science* 18(1): 71-103.

Godfrey, K. 1985. Comparing the means of several groups. *New England Journal of Medicine* 311: 1450-1456.

Hochberg, Y. 1988. A Sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75(4): 800-802.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandanian Journal of Statistics* 6: 65-70.
Hommel, G. 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 75(2): 383-386.
Marcus, R., Eric, P. & Gabriel, K.R. 1976. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63(3): 655-660.
Nor Aishah Ahad, Abdul Rahman Othman. & Sharipah Soaad Syed Yahaya. 2012. Performance of two-samples pseudo-median procedure. *Sains Malaysiana* 41(9): 1149-1154.
Pocock, S.J., Hughes, M.D. & Lee, R.J. 1987. Statistical problems in the reporting of clinical trials. *New England Journal of Medicine* 317: 426-432.
Perneger, T.V. 1998. What's wrong with Bonferroni adjustments. *BMJ* 316(7139): 1236-1238.
Simes, R.J. 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73(3): 751-754.
Sinclair, J.K., Taylor, P.J. & Hobbs, S.J. 2013. Alpha level adjustments for multiple dependent variable analyses and their applicability: A review. *World Academic Press, World Academic Union* 7(1): 017-020.
Smith, D.G., Clemens, J., Crede, W., Harvey, M. & Gracely, E.J. 1987. Impact of multiple comparisons in randomized clinical trials. *The American Journal of Medicine* 83(3): 545-550.
Wright, S.P. 1992. Adjusted P-values for simultaneous inference. *Biometrics* 48: 1005-1013.

Jabatan Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

*Pengarang untuk surat-menyurat; email: noramuda@ukm.edu.my

Diserahkan: 16 Julai 2019
Diterima: 27 Disember 2019