KETERBATASAN UJI SIGNIFIKANSI HIPOTESIS NOL

Nonny Swediati dan Bastari

KRITIKAN TERHADAP UJI SIGNIFIKANSI HIPOTESIS NOL

Paper ini ditulis berdasarkan telaah dan adaptasi dari 2 paper yang masing-masing ditulis oleh Cohen (1994) dan Kirk (1996). Kedua pakar statistika ini telah mencoba untuk membahas permasalahan di atas sekaligus memberikan pemecahannya. Namun demikian, tidak semua pokok-pokok pikiran mereka akan dituangkan dalam tulisan ini. Hanya bagian-bagian tertentu saja yang dianggap penting yang akan dibicarakan, sebagai bahan renungan bagi kita, peneliti dan pengajar, khususnya dalam bidang pendidikan, psikologi, dan ilmu-ilmu sosial.

Menurut Kirk (1996) terdapat tiga kritik atau keterbatasan yang paling banyak disebut-sebut dalam uji signifikansi terhadap hipotesis nol, yaitu:

1. Prosedur ini tidak menjawab hal-hal yang ingin diketahui oleh para peneliti. Jelasnya, ada perbedaan penekanan pertanyaan antara *scientific inference* dengan uji signifikansi terhadap hipotesis nol. Dalam *scientific inference* yang ingin diketahui adalah berapakah probabilitas kebenaran hipotesis nol (H_o) dari satu set data (D) yang diperoleh, $p(H_o|D)$. Sebaliknya, uji signifikansi hipotesis nol menekankan pada pertanyaan berapakah probabilitas untuk memperoleh data tersebut jika hipotesis nolnya benar, $p(D|H_o)$. Sayangnya, hal ini tidak berarti bahwa jika kita memperoleh data dengan $p(D|H_o)$ rendah maka $p(H_o|D)$ juga akan rendah. Pernyataan yang kurang benar juga adalah jika *nilai* ***p*** yang relatif kecil dikaitkan dengan uji statistik, misalnya kurang dari 0,05, maka hipotesis nolnya mungkin salah. Alasan yang kurang benar itu dilanjutkan lagi dengan banyaknya peneliti yang percaya bahwa (a) *nilai p* merupakan probabilitas bahwa hipotesis nol benar, dan (b) pelengkap dari *nilai p* adalah probabilitas bahwa hasil yang signifikan itu akan dijumpai pada replikasi yang dilakukan.
2. Uji signifikansi hipotesis nol merupakan kegiatan yang tidak ada gunanya. John Tukey (1991) menyatakan bahwa " Pengaruh dari perlakuan A dan B selalu

berbeda dalam beberapa angka desimal untuk setiap A dan B." Oleh karena itu, pertanyaan 'Apakah kedua perlakuan mempunyai pengaruh yang berbeda?' adalah pertanyaan yang sama sekali naif. Karena hipotesis nol selalu *salah*, maka keputusan untuk menolak pernyataan itu hanyalah menunjukkan bahwa suatu design penelitian mempunyai *power* untuk mendeteksi keadaan yang sebenarnya, yang pengaruhnya belum dapat dipastikan apakah besar atau bermanfaat. Menurut Cohen (1994) hal yang ritual dan sering dikerjakan dalam uji signifikansi hipotesis nol menyebabkan para peneliti lebih mengontrol *error* tipe I yang tidak muncul karena semua hipotesis nol adalah *salah*, sementara itu akan muncul *error* tipe II yang akan melebihi batas penerimaan, bahkan sampai sebesar 0,50 atau 0,80.

3. Dengan menggunakan taraf signifikansi tertentu dalam uji hipotesis, seorang peneliti menjadi terbelenggu dalam membuat keputusan, yaitu mereduksi suatu kontinum ke dalam dua kategori *reject* atau *do not reject* Ho. Keadaan tersebut dapat membuat peneliti kesituasi yang *anomali* (dua peneliti dengan kasus yang sama tetapi memperoleh kesimpulan yang berbeda). Satu peneliti memperoleh nilai $p < 0,06$ dan mengambil kesimpulan *not to reject* Ho. Peneliti yang lain menggunakan sampel yang agak lebih banyak dan mendapatkan nilai $p < 0,05$ dengan kesimpulan *to reject* Ho. Mana yang paling Anda sukai?

   Jawaban Rosnow and Rosenthal dalam Kirk (1994): *Surely, God loves the .06 nearly as much as the 0.05"*. Masih dalam kaitannya dengan keputusan yang berdasarkan dikotomi tersebut, beberapa peneliti salah menginterpretasikan bahwa kegagalan dalam menolak hipotesis nol adalah sebagai bukti untuk menerimanya.

INFORMASI TAMBAHAN YANG DIPERLUKAN DALAM PROSEDUR UJI SIGNIFIKANSI HIPOTESIS NOL

Dari pembahasan sebelumnya, nampak bahwa hasil uji signifikansi terhadap Ho mempunyai beberapa kelemahan, dan kurang informatif. Menurut Kirk (1994) banyak peneliti yang menekankan pentingnya untuk menolak hipotesis nol, dan 'besar kecilnya' dari nilai *p*. Seharusnya, para peneliti lebih menekankan pada data, yaitu apakah data mendukung hipotesis ilmiah (*Scientific hypothesis*). Sebagai contoh Menurut Fisher (1925) dalam Kirk (1994) pada waktu menggunakan uji signifikansi dengan anava perlu diberi informasi pelengkap, yaitu informasi tentang rasio korelasi atau eta, yang isinya menunjukkan kekuatan hubungan antara variabel independen dan dependen.

Ada beberapa cara untuk melengkapi informasi hasil uji hipotesis, yaitu *confidence interval, measures of effect magnitude*, analisa data secara eksploratori, dan metode grafik. Selanjutnya, dalam paper ini yang dibahas hanya dua pertama. Menurut Kirk (1994), ada dua cara yang banyak dikenal untuk melengkapi informasi dalam uji hipotesis adalah menggunakan *confidence interval* dan pengukuran terhadap *effect magnitude* dari satu/lebih dari semua pengukuran.

1. Confidence Interval

Apa yang dapat dipelajari oleh seorang peneliti jika gagal untuk menolak Ho? Sebenarnya hasil yang disampaikan tidak hanya masalah menyatakan arah perbedaan perlakuan/kondisi yang diberikan (Hasil dari uji hipotesis). Karena pilihan keputusan hanya dua, *rejected* dan *retained* dan informasi lanjut yang tersedia tidak ada, maka yang akan dipelajari seorang peneliti dari hasil uji hipotesis menjadi terbatas. Informasi lain yang dibutuhkan misalnya berapa besar perbedaan A dan B, dan berapa besar *error* pada waktu melakukan estimasi. Jadi informasi bahwa A lebih besar dari B saja kurang membantu dalam menyampaikan hasil. Informasi tambahan yang diperlukan dapat diperoleh dari *confidence interval* (CI) yang berisikan semua informasi/nilai yang tersedia pada uji signifikansi, lebih dari itu pada CI juga memuat range nilai termasuk di dalam range itu adalah 'perbedaan yang sebenarnya'. Proses untuk memperoleh CI disebut estimasi interval

Kelebihan-kelebihan CI dibandingkan dengan hasil uji hipotesis terutama bila dikaitkan dengan *inference mean*.

a. Hasil dari estimasi interval adalah pernyataan tentang satu parameter atau beberapa parameter yang kita teliti. Pada uji hipotesis, pernyataannya tentang *derived score*, yaitu *z* atau *t*, atau tentang probabilitas, *p*. Keduanya digunakan untuk mengambil kesimpulan tentang satu/lebih parameter.

b. Estimasi interval memperlihatkan langsung pengaruh dari variasi sampel yang dilakukan secara random, terutama jumlah sampel. Ingat, jumlah sampel yang kecil menghasilkan CI yang lebih lebar, sedangkan jumlah sampel besar menghasilkan CI yang lebih pendek. Lebar-pendeknya suatu interval memberikan informasi langsung tentang ketepatan estimasi untuk tujuan kita. Dalam uji hipotesis, taraf signifikansi dipengaruhi oleh dua hal yang sulit dijelaskan tanpa adanya penelitian lanjut, yaitu (a) perbedaan antara hal yang dihipotesiskan dan bagaimanakah kebenarannya, dan (b) ikut sertanya sejumlah variasi sampling. Sebagai contoh, Nilai *t* yang besar muncul dalam penelitian di mana perbedaan

meannya kecil tetapi sampelnya besar, atau perbedaan meannya besar tetapi sampelnya kecil.

c. Dalam uji hipotesis kita mudah bingung/terlena pada pernyataan "*statistically significant difference* daripada hal yang penting lainnya. Dengan adanya CI masalah tersebut dapat dieliminasi. Suatu contoh. Skor-skor IQ dari suatu sampel memiliki $\overline{X} = 103$, $s_x = 20$, dan *n* = 1600. Jika kita melakukan uji hipotesis yang menyatakan bahwa $\mu_x = 100$, dan kita memperoleh *t* kira-kira +6 dengan p< .000001. Hebat! Namun demikian perhitungan dengan interval estimate tentang μ_x menunjukkan bahwa $C(102 \le \mu_x \le 104) = .95$ akan membawa kita kepada realitas/keadaan yang sesungguhnya. Kembali lagi ke masalah, pentingnya hasil penelitian tergantung kepada keputusan peneliti.

d. Hasil dari uji hiptesa adalah menyatakan bahwa suatu kondisi tertentu yang dinyatakan dalam bentuk hipotesis nol dapat dipertahankan (Ho dipertahankan) atau hipotesis nol ditolak (Ho ditolak). Sedangkan pada Interval estimate lebih menekankan pada nilai dalam suatu range yang dapat menjelaskan parameter yang kita teliti.

Meskipun *confidence interval* ini mempunyai banyak kelebihan dibandingkan dengan hasil uji hipotesis, tetapi dalam penelitian-penelitian pendidikan, psikologi maupun ilmu-ilmu sosial lain masih jarang sekali digunakan. Kenyataan yang sering kita lihat adalah strategi mengambil keputusan dengan cara 'menolak atau mempertahankan Ho'. Hasil dari uji hipotesis tersebut yang tidak banyak memberikan informasi tentang hal-hal yang kita ingin ketahui, dan peneliti sudah puas dengan *nilai p* saja tanpa memeriksa kembali datanya.

Sejauh ini *confidence interval* yang kita bahas di atas masih terbatas pada sampel yang kita peroleh/teliti. Dengan kemajuan ilmu dalam statistik dan tersedianya komputer dengan kapasitas pengolahan data yang besar dan cepat, banyak peneliti melakukan simulasi data yang hasilnya akan memperkaya informasi yang kita perlukan, terutama dalam kaitannya dengan mengestimasi parameter yang kita teliti, dan juga mengestimasi *confidence interval*. Dengan simulasi ini memungkinkan kita untuk melakukan resampling tanpa harus berasumsi bahwa CI untuk parameter-parameter yang kita teliti mengikuti distribusi normal. Konsep dasar dalam melakukan resampling bahwa resampling tidak menggunakan asumsi distribusi probabilitas, tetapi langsung menghitung distribusi emperik dari estimasi parameter-parameternya. Dengan membuat sampel yang berkali-kali berdasarkan sampel aslinya, kemudian kita tinggal mengestimasi nilai parameter dari masing-masing sampel. Setelah semua

diestimasi, kita dapat membuat histogram dari nilai-nilai parameter tersebut, dan bahkan dapat menghitung CI untuk estimasi parameternya dari distribusi yang sebenarnya. Sebagai informasi, dua model *resampling* yang banyak digunakan adalah *Jacknife* dan *bootstrap*. Teori-teori yang mendasari *resampling* dan pembahasannya dapat dipelajari di buku-buku statistik baru.

2. Pengukuran terhadap *Effect magnitude*.

Untuk melengkapi informasi yang diperoleh dari uji hipotesis, Kirk (1996) banyak membahas cara-cara pengukuran terhadap *effect magnitude*. Menurut Kirk pengukuran terhadap *effect magnitude* dikategorikan dalam salah satu dari 3 kategori berikut (a) *measures of strength of association* ; (b) *measures of effect size* (biasanya, perbedaan mean yang telah distandardisasi), dan (c) *other measures*. Dari 3 kategori *effect magnitude* tersebut yang banyak mendapat perhatian adalah (a) *strength of association* dan (b) *effect size*. Namun demikian, (c) pengukuran-pengukuran lain mulai mendapat perhatian para peneliti. Menurut Kirk (1994) pengukuran terhadap *effect magnitude* ini belum banyak dilakukan digunakan oleh peneliti-peneliti bidang pendidikan maupun psikologi.

KESIMPULAN

Banyak hasil penelitian-penelitian di Indonesia di bidang ilmu-ilmu sosial termasuk di dalamnya pendidikan dan psikologi yang menggunakan uji signifikansi hipotesis. Seperti yang telah dibahas sebelumnya bahwa uji hipotesis tersebut mempunyai kelemahan-kelemahan dan perlu dilengkapi dengan prosedur lanjut untuk melengkapi informasi yang dihasilkan dari uji hipotesis, sehingga hasil penelitian tersebut lebih bermakna. Prosedur-prosedur lanjut tersebut telah banyak dibahas dan banyak digunakan di negara-negara barat. Di samping itu, dalam paper ini juga dibicarakan sekilas tentang penelitian dengan *resampling* melalui simulasi yang merupakan alternatif menarik untuk melakukan penelitian.

Kita sebagai peneliti perlu mulai menyesuaikan diri dan mengikuti perkembangan statistik yang ada, tidak hanya memfokuskan kepada uji signifikansi hipotesis nol saja, tetapi kita harus mulai dengan signifikansi praktis tentang data yang kita teliti.

KEPUSTAKAAN

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, December 1994, p. 997-999.*

Hair, J.F, Anderson, R.E., Tatham, R.L., and Black, W.C. (1998). *Multivariate data analysis*. Fifth edition. Prentice-Hall International, Inc.

Kirk, R.E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, Vol. 56. No. 5; October, p. 746-759.*