

SEPUTAR KONTROVERSI UJI SIGNIFIKANSI HIPOTESIS NIHIL

Dicky Hastjarjo

PERDEBATAN TENTANG PERLU-TIDAKNYA UJI SIGNIFIKANSI HIPOTESIS NIHIL

Sejumlah pakar menyatakan bahwa psikologi, ilmu perilaku, dan ilmu sosial mengalami rintangan dan hambatan untuk maju. Salah satu kambing hitamnya adalah kebiasaan berat psikologi dan ilmu sosial dalam menggunakan uji signifikansi hipotesis nihil (Cohen, 1994). Kebiasaan yang sulit dihilangkan ini membuat psikologi seperti serangkaian kuldesak atau lorong buntu metodologis (Loftus, 1996) sehingga psikologi hanya sibuk memutar rodanya tanpa banyak kemajuan (Cohen, 1994), atau maju selambat jalannya gletser (Hunter, 1997),

Semenjak diperkenalkan 75 tahun yang lalu oleh Fisher, banyak kritikan telah dialamatkan kepada pengujian signifikansi hipotesis nihil. Kritikan-kritikan tersebut diantaranya adalah (a) penerapan yang keliru penalaran silogistik deduktif, dan (b) hipotesis nihil selalu salah (Cohen, 1994), serta (a) tingkat kesalahan uji signifikansi bukan sebesar 5% melainkan 60%, dan (b) hipotesis nihil hampir selalu salah (Hunter, 1997). Salah satu alternatif yang disodorkan untuk menggantikan uji signifikansi adalah penggunaan interval kepercayaan (*confidence interval*) (Cohen, 1994; Hunter, 1997). Pada kenyataannya, uji signifikansi hipotesis nihil tetap menjadi metode utama analisis data dalam ilmu-ilmu sosial meskipun telah mendapat kritikan bertubi-tubi (Tyron, 1998). Cohen (1994) menggambarkan uji signifikansi hipotesis nihil bak seorang maharaja tanpa busana yang tanpa rasa malu berlari berputar-putar dalam waktu lama. Situasi ini terjadi oleh karena buku teks menekankan pengujian hipotesis nihil, para dosen mensosialisasikan kepada mahasiswanya penggunaan (*use*) dan penyalahgunaan (*abuse*) uji signifikansi, serta para editor jurnal meminta pengarang naskah untuk melaporkan lebih banyak uji signifikansi (Shrout, 1997).

Menyadari situasi yang demikian ini dalam konvensi *American Psychological Society* (APS) tahun 1996 terdapat simposium yang menyangkut kontroversi uji signifikansi hipotesis nihil. Makalah-makalah, baik yang pro maupun kontra uji signifikansi, diterbitkan dalam rubrik khusus di jurnal *Psychological Science* edisi Januari 1997. Sejumlah pengarang makalah menyarankan untuk melarang sama sekali pengujian signifikansi hipotesis nihil (Hunter, 1997, Shrout, 1997). Saran untuk melarang uji signifikansi hipotesis nihil ini juga telah banyak dikemukakan diantaranya oleh Cohen (1994), Hubbard (1995), dan McGrath (1998). Hubbard (1995) menegaskan bahwa lonceng kematian pengujian signifikansi hanya akan mungkin terjadi jikalau ada perubahan kurikulum di program pascasarjana dengan lebih menekankan penelitian replikasi serta menghilangkan perilaku tergilagila kepada nilai $p < .05$. Judul makalah Hunter (1997) misalnya juga sangat provokatif yakni "Dibutuhkan: satu pelarangan terhadap uji signifikansi". Seorang editor jurnal psikologi pernah menghimbau kepada para pengirim naskah untuk tidak melakukan uji signifikansi (Loftus, 1996, catatan kaki). Seorang editor sebuah jurnal ilmu kesehatan masyarakat bahkan telah meminta calon pengirim naskah hasil penelitian untuk menghilangkan semua acuan mengenai pengujian hipotesis dan signifikansi atau jika tidak mau melakukan hal itu penulis dipersilakan mempublikasikan naskahnya di jurnal lain (Shrout, 1997).

Sejumlah pakar masih mempunyai sikap menyetujui praktek uji signifikansi hipotesis nihil (Abelson, 1997; Baril & Cannon, 1995; Haggen, 1997), atau sekurang-kurangnya tidak setuju terhadap pelarangan uji signifikansi hipotesis nihil (Harris, 1997; Estes, 1997). Barril dan Cannon (1995) tidak setuju terhadap kesimpulan makalah Cohen (1994) diatas dan kedua ahli itu menegaskan bahwa mereka tidak salah menaruh keyakinan pada model hipotesis nihil. Hagen (1997) menyatakan bahwa logika uji signifikansi hipotesis nihil bersifat elegan, luar biasa kreatif, melekat secara dalam pada metode statistika inferensial, serta tidak akan pernah mampu kita cerai meskipun suatu saat kita memutuskan ingin berpisah dengannya. Sebagai penutup makalahnya yang berjudul "Pujian bagi uji signifikansi hipotesis nihil", Hagen (1997) menulis "Ketika kita menggunakan uji signifikansi hipotesis nihil, saya kira kebanyakan dari kita akan terus menggunakannya, janganlah lupa untuk merayakan kecemerlangannya sesekali waktu". Pada saat rubrik khusus mengenai kontroversi pengujian signifikansi hipotesis nihil dimuat di jurnal *Psychological Science* edisi tahun 1997, maka editor jurnal itu mengumumkan bahwa dia tidak akan melarang uji signifikansi statistik selama masa baktinya.

TANGGAPAN TERHADAP PERDEBATAN MENGENAI UJI SIGNIFIKANSI HIPOTESIS NIHIL

Sebagai tanggapan terhadap kontroversi mengenai uji signifikansi, terutama terhadap makalah Cohen di *American Psychologist* yang terbit di tahun 1994, maka *Board of Scientific Affairs (BSA)* dari *American Psychological Association (APA)* pada tahun 1996 membentuk satu komisi yang diberi nama *Task Force on Statistical Inference (TFSI)* (Wilkinson & TFSI, 1999). Salah satu tugas *TFSI* adalah untuk memberi penjelasan mengenai sejumlah isu kontroversial menyangkut aplikasi statistik termasuk uji signifikansi dan alternatif-alternatifnya. *TFSI* diketuai oleh Robert Rosenthal, Robert Abelson, dan Jacob Cohen. Perlu diketahui bahwa sebagaimana dalam uraian diatas maka Robert Abelson adalah ahli psikologi yang bersikap pro uji signifikansi, sedang Jacob Cohen adalah ahli psikologi yang telah berulang kali menggugat keberadaan uji signifikansi hipotesis nihil dalam artikel yang ditulisnya. Sayang bahwa Jacob Cohen tidak sempat melihat perkembangan kerja *TFSI* selanjutnya karena ia meninggal dunia pada awal tahun 1998. *TFSI* juga beranggotakan ahli statistik, pengajar statistik, editor jurnal, pengarang buku statistik, ahli komputer dan sesepuh yang bijaksana. Sementara itu, tokoh-tokoh seperti Lee Cronbach, Paul Meehl, Frederick Mosteller dan John Tukey bertindak sebagai penasihat senior dari *TFSI*.

TFSI merekomendasikan untuk merevisi *APA Publication Manual* edisi tahun 1994 terutama pada Bab Statistik. Rekomendasi tersebut telah disosialisasikan terlebih dahulu kepada publik psikologi dalam bentuk satu artikel sebelum revisi dilakukan pada pedoman publikasi APA itu (Wilkinson, & TFSI, 1999). Rekomendasi yang diajukan mencakup metode (rancangan, pemilihan subjek, dan pengukuran), hasil (komplikasi, dan analisis) dan diskusi (interpretasi, dan kesimpulan).

Rekomendasi yang berkaitan dengan analisis meliputi sejumlah aspek, diantaranya adalah (a) Memilih satu analisis yang cukup secara minimal. Prinsip parsimoni atau dikenal sebagai prinsip *Occam's razor* berlaku bagi pemilihan metode selain bagi pemilihan teori. Jika analisis yang lebih sederhana dapat menjawab permasalahan, maka tidak perlu memakai analisis yang lebih kompleks sekedar untuk memberikan kesan baik bagi pembaca. Ingat gagasan Fisher: "*Begin with an idea. Then pick a method*" (1999, h. 598). Kita dalam melakukan penelitian tidak jarang bersikap terbalik "Menentukan metode dulu, baru cari permasalahan". Misalnya, oleh karena kita baru saja belajar analisis regresi, maka sebelum membuat proposal penelitian kita berkata pada diri sendiri "permasalahan apa yang bisa saya pecahkan dengan analisis regresi", (b) Memilih program komputer. Memilih program *software*

statistik tertentu yang baik itu perlu dilakukan oleh peneliti. Namun ada hal yang lebih penting lagi, yakni peneliti harus memverifikasi hasil analisis statistik, memahami apa maksudnya, dan mengerti bagaimana menghitungnya. *TFSI* juga menambahkan: “*Do not report statistics to a greater precision than is supported by your data simply because they are printed that way by the program*” (1999, h. 598). Tidak jarang juga disekeliling kita ditemukan sejumlah laporan penelitian yang dilampiri berpuluh-puluh halaman hasil perhitungan statististik, meskipun ketika ditanya arti dari angka-angka tadi si pembuat laporan kurang mengerti, (c) Mempertimbangkan asumsi-asumsi. *TFSI* menulis: “*You should take efforts to assure that the underlying assumptions required for the analysis are reasonable given the data*”, (d) Pengujian hipotesis. Melaporkan nilai p aktual akan lebih bagus daripada membuat keputusan dikotomis menolak-menerima hipotesis, dan akan lebih bagus lagi melaporkan interval kepercayaan. Senantiasalah memberikan *effect size estimate* jika melaporkan nilai p. *TFSI* menulis: “*Never use the unfortunate expression “accept the null hypothesis”*”. *Always provide some effect size estimate when reporting a p value*” (1999, h. 399), (e) Menyajikan *Effect Size*. Senantiasalah menyajikan *effect size* untuk hasil-hasil pokok/utama, (f) Menyajikan *Interval Estimates*. Perkiraan interval sebaiknya disajikan untuk setiap *effect size* untuk hasil-hasil utama.

Sebenarnya masih ada beberapa aspek lagi yang direkomendasikan untuk bab analisis pada khususnya atau untuk bab metode pada umumnya, namun untuk kepentingan makalah ini hanya disajikan materi yang relevan saja. Memang rekomendasi *TFSI* tidak sampai pada keputusan melarang praktek uji signifikansi hipotesis nihil, sebab *TFSI* memandang ada sejumlah contoh yang melawan kritikan terhadap uji signifikansi (1999, h. 603; lihat juga Abelson, 1997). Akan tetapi rekomendasi *TFSI* menegaskan bahwa laporan penelitian sebaiknya menyertakan juga interval kepercayaan dan *effect size*. Disamping itu, *TFSI* juga merekomendasikan untuk menyertakan signifikansi teoretis (*theoretical/practical significance*)(1999, h. 603; lihat juga Kirk, 1996 dan makalah Swediati & Bastari dalam edisi ini). Dampak dari rekomendasi ini sudah terlihat, yakni sekitar 13 jurnal telah mengharuskan para pengirim naskah untuk melaporkan *effect size* (Thompson, 2000). Jurnal itu sebagian diantaranya adalah *Contemporary Educational Psychology, Educational and Psychological Measurement, Journal of Applied Psychology, Journal of Consulting and Clinical Psychology, Journal of Experimental Education, Journal of Learning Disabilities, Language Learning, Measurement and Evaluation in Counseling and Development*. Satu pertanyaan yang timbul adalah apakah rekomendasi itu atau nantinya revisi pedoman publikasi APA akan berdampak juga bagi praktek-praktek publikasi penelitian ilmiah di jurnal psikologi di Indonesia atau secara umum pada

praktek-praktek penulisan laporan penelitian ilmiah di bidang psikologi di Indonesia ? Seandainya jawabannya adalah afirmatif tentu saja pengajaran statistika di ruang kuliahpun akan mengalami perubahan juga. “*The winds of change are about us*” tulis Kirk (Kirk, 1996, h. 757).

PUSTAKA

- Abelson, R. P. 1997. On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, vol. 8, 12-15.
- Baril, G. L., & Cannon, J.T. 1995. What is the probability that null hypothesis testing is meaningless? *American Psychologist*, vol. 50, 1098-1099.
- Cohen, J. 1994. The earth is round ($p < .05$). *American Psychologist*, vol. 49, 997-1003.
- Hagen, R. L. 1997. In praise of the null hypothesis statistical test. *American Psychologist*, vol. 52, 15-24.
- Estes, W. K. 1997. Significance testing in psychological Research: Some persisting issues. *Psychological Science*, vol. 8, 18-20.
- Harris, J. R. 1997. Significance tests have their place. *Psychological Science*, vol. 8, 8-11.
- Hubbard, R. 1995. The earth is highly significantly round ($p < .0001$). *American Psychologist*, vol. 50, 1098.
- Hunter, J. E. 1997. Needed: A ban on the significance test. *Psychological Science*, vol. 8, 3-7.
- Kirk, R. E. 1996. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, vol. 56, 746-759.
- Loftus, G. R. 1996. Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, vol. 5, 161-171.
- McGrath, R. E. 1998. Significance testing: Is there something better ? *American Psychologist*, vol. 53, 796-797.
- Shrout, P. E. 1997. Should significance tests be banned ? Introduction to a special section exploring pros and cons. *Psychological Science*, vol. 8, 1-2.
- Swediati, N., & Bastari. 2000. *Keterbatasan Uji Signifikansi Hipotesis Nol*. Makalah disampaikan pada Seminar Kontroversi Penggunaan Prinsip-prinsip Statistik.

Bagian Psikologi Umum dan Eksperimen, Fakultas Psikologi, UGM, Yogyakarta

Thompson, B. 2000. Where the Social Sciences Are Headed: Away from NHST, toward Effect Size & Replicability Evidence. *Http://acs.tamu.edu/bbt 61471*.

Tyron, W.W. 1998. The inscrutable null hypothesis. *American Psychologist, vol.53, 796*.

Wilkinson, L, & Task Force on Statistical Inference. 1999. Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist, vol. 54, 8, 594-604*.