

## **PENGUJIAN SIGNIFIKANSI HIPOTESIS NOL DALAM PENELITIAN PSIKOLOGIS**

*Sumadi Suryabrata*

### **PENDAHULUAN**

Walaupun diakui bahwa penerapan pendekatan kualitatif dalam penelitian psikologis cukup terhormat dan layak dilakukan, namun dalam praktek mayoritas penelitian psikologis menggunakan penelitian kuantitatif. Dan yang menerapkan pendekatan kuantitatif itu alat utamanya untuk membuat kesimpulan-kesimpulan ilmiah adalah statistika. Diantara teknik-teknik statistik yang paling banyak digunakan adalah pengujian signifikansi hipotesis nol. Survai yang dilakukan oleh Estes (1997) menunjukkan bahwa dari 11 laporan penelitian yang dimuat di *British Quarterly Journal of Experimental Psychology* edisi Februari 1996 semuanya menggunakan uji signifikansi, dari 17 penelitian yang dimuat di *Journal of Abnormal Psychology* edisi february 1996 semuanya menerapkan uji signifikansi. Greenwald (1993) berusaha mendapatkan informasi dari penulis dan penelaah artikel yang disampaikan kepada *Journal of Personality and Social Psychology* selama periode 3 bulan pada tahun 1973 mengenai apa yang mereka lakukan bila mereka gagal dalam uji signifikansi. Jawaban yang diperoleh adalah 66% melakukan replikasi sebelum menyerahkan naskah, 28% membatalkan tulisannya, dan 6% tetap menyerahkan tulisannya. Dari informasi diatas dapat disimpulkan bahwa pengujian signifikansi hipotesis nol adalah hal yang serius bagi para peneliti dan hal tersebut telah mempunyai dasar yang kuat.

Karena itu adalah hal yang mengejutkan bagi banyak peneliti dibidang psikologi ketika uji signifikansi hipotesis nol itu mendapatkan kritikan tajam (bahkan semacam gugatan) dari sejumlah ahli. Kritikan yang paling tajam adalah yang menyatakan bahwa praktek-praktek uji signifikansi itu justru menghambat kemajuan penelitian psikologis.

Dalam semangat memajukan penelitian psikologis di Indonesia kiranya perlu dicermati isu ini.

## PROSEDUR BAKU

Analisis dan interpretasi data dalam ilmu-ilmu perilaku dan ilmu-ilmu sosial dilakukan dengan menggunakan seperangkat teknik yang boleh dikata sudah baku, yang disebut pengujian signifikansi hipotesis nol (*null hypothesis significance testing*), yang prosedurnya secara garis besar adalah sebagai berikut:

1. Peneliti mulai dengan hipotesis bahwa variabel bebas tertentu mempunyai pengaruh (effect) terhadap variabel tergantung tertentu. Secara umum hipotesis itu dapat dinyatakan sebagai

$$\text{Bukan } (\mu_1 = \mu_2 = \dots \mu_j) \quad (1)$$

Dimana  $\mu_1, \dots, \mu_j$  adalah rata-rata  $j$  distribusi populasi pada variabel tergantung dalam kaitan dengan variabel bebas. Persamaan 1 itu disebut *hipotesis alternatif* atau  $H_1$ .

2. Selanjutnya, peneliti melakukan eksperimen dimana sebanyak  $j$  sampel rambang variabel diperoleh. Eksperimen ini menghasilkan rata-rata sampel  $\bar{X}_1, \dots, \bar{X}_j$  yang merupakan perkiraan bagi rata-rata populasi  $\mu_1, \dots, \mu_j$ . Pada umumnya adalah tidak benar bahwa  $\bar{X}_i = \dots \bar{X}_j$ , artinya akan selalu ada perbedaan diantara berbagai rata-rata sampel. Yang perlu ditentukan adalah apakah perbedaan-perbedaan antara rata-rata yang diobservasi itu hanya disebabkan kesalahan sampling atau karena --- setidaknya-tidaknya --- sebagian berkaitan dengan perbedaan rata-rata pada populasi. Si peneliti harus memberikan alasan bahwa perbedaan-perbedaan itu adalah *nyata*.
3. Sampai disini peneliti menghitung probabilitas (dikenal dengan  $p$ ) mendapatkan perbedaan-perbedaan antara rata-rata  $\bar{X}_1, \dots, \bar{X}_j$  sebesar yang sebenarnya diobservasi jika rata-rata sebanyak  $j$  populasi itu semua sama, atau

$$\mu_1 = \mu_2, \dots, \mu_j \quad (2)$$

Hipotesis bahwa semua rata-rata populasi adalah sama ini disebut *hipotesis nol* atau  $H_0$ .

4. Berdasar atas perbandingan antara  $p$  hitung dengan  $p$  kriteria yang dikenal dengan  $\alpha$  (yang lazim ditetapkan  $\alpha=0,05$ ) si peneliti membuat keputusan binari. Jika  $p$  lebih kecil dari  $\alpha$ , maka peneliti membuat keputusan kuat menolak  $H_0$  dan menerima  $H_1$  (keputusan itu biasanya dinyatakan "efek yang diobservasi secara statistik signifikan"). Apabila  $p$  lebih besar dari  $\alpha$ , maka peneliti membuat "keputusan lemah gagal menolak  $H_0$ ".

Dua tipe kesalahan dapat dilakukan dalam rangka ini. Orang melakukan kesalahan tipe I (*type I error*) bila dia secara tak benar menolak hipotesis nol yang benar. Jika hipotesis nol benar, maka probabilitas kesalahan tipe I, menurut definisinya sama dengan  $\alpha$ . Orang melakukan kesalahan tipe II apabila dia secara tak benar gagal menolak hipotesis nol yang salah/palsu. Jika hipotesis nol salah/palsu, kesalahan Tipe II dilakukan dengan probabilitas  $\beta$ . Pada umumnya kita tidak tahu harga  $\beta$  karena kita tidak mempunyai informasinya, dan tidak ada asumsi mengenai harga-harga  $\mu$ , bila hipotesis nol palsu. *Statistical power* didefinisikan sebagai  $(1-\beta)$ , yang merupakan probabilitas orang secara benar menolak hipotesis nol yang palsu. Karena pada umumnya  $\beta$  tidak diketahui, maka demikian juga *power* itu.

5. Akhirnya, berdasar atas serangkaian keputusan tersebut di atas yaitu, menolak atau gagal menolak serangkaian hipotesis nol peneliti memberi makna pada perangkat data itu, betapapun kompleksnya.

## **RAGAM UJI SIGNIFIKANSI**

Gambaran mengenai prosedur yang disajikan di atas itu adalah penyederhanaan dengan rujukan terutama kepada analisis univariat. Dalam kenyataannya metode uji signifikansi itu cukup kompleks, lebih-lebih pada analisis multivariat. Walaupun para peneliti tentu sudah akrab dengan berbagai ragam uji signifikansi itu, namun kiranya akan berguna untuk menyegarkan ingatan kalau hal tersebut ditelaah balik secara sekilas.

### **A. Uji signifikansi univariat**

#### **1. Univariat, tes satu sampel**

Yang dimaksud dengan tes satu-sampel adalah suatu tes mengenai hipotesis tentang suatu rata-rata populasi, didasarkan pada satu sampel yang ditarik dari populasi itu.

- a. simpangan baku populasi diketahui, tes yang digunakan adalah:

$$z = \frac{\bar{X} - \mu_0}{\delta/\sqrt{N}} \tag{3}$$

dimana  $\bar{X}$  = rata-rata sampel

$\mu_0$  = rata-rata populasi

$\delta$  = simpangan baku populasi

- b. Simpangan baku populasi tak diketahui. Jika simpangan baku populasi tidak diketahui, maka dia diganti dengan perkiraannya, yaitu simpangan baku sampel, yang dihitung dengan rumus.

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$$

Jika distribusi populasinya normal, maka tes statistiknya adalah:

$$t = \frac{\bar{X} - \mu_0}{S\sqrt{N}} \quad (4)$$

dengan  $db = n - 1$

- c. Rancangan berpasangan

Aplikasi tes t satu-sampel yang penting adalah kalau kita punya dua kelompok misalnya kelompok eksperimen dan kontrol, kelompok yang dikenai Metode I dan kelompok yang dikenai Metode II, dst). Daripada menggarap dua perangkat skor dengan rata-rata  $\bar{X}_1$ , dan  $\bar{X}_2$ , disini biasanya orang menggarapnya lewat perbedaan tiap pasangan skor:

$d_i = X_{1i} - X_{2i}$ , yang rata-rata perbedaannya hanyalah satu

Dengan demikian, pertanyaan “Apakah  $X_1$  berbeda secara signifikan dari  $X_2$  ?” diubah menjadi “Apakah  $d$  secara signifikan berbeda dari 0?”

Tes statistiknya adalah

$$t = \frac{\bar{d}}{S_d\sqrt{N}} \quad (5)$$

dengan  $db = N - 1$

## 2. Univariat, tes dua-sampel

Tes dua sampel merujuk kepada situasi-situasi dimana kita mempunyai dua variabel bebas yang diperoleh secara seimbang (seperti dua kelompok perlakuan, kelompok eksperimen dan kelompok kontrol, dan yang sejenisnya) dan kita ingin

menguji signifikansi perbedaan antara rata-rata mereka  $\bar{X}_1$  dan  $\bar{X}_2$  dalam variabel kriterion

- a.  $\delta_1$  dan  $\delta_2$  diketahui.

Tes statistiknya adalah

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{6}$$

Dengan  $db = n_1 + n_2 - 2$

- b.  $\delta_1$  dan  $\delta_2$  tidak diketahui, tetapi diasumsikan keduanya sama.

Dalam hal ini perkiraan terbaik yang tidak bias untuk variansi populasi adalah

$\delta^2 (= \delta_1^2 = \delta_2^2)$  diperoleh melalui

$$S_w^2 = \frac{\Sigma(X_1 - \bar{X}_1)^2 + (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

Tes statistiknya adalah

$$T = \frac{X_1 - \bar{X}_2}{\sqrt{S_w^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \tag{7}$$

Dengan  $db = n_1 + n_2 - 2$

- c. Univariat, tes K-sampel
- d. Situasi yang kita hadapi adalah kelompok sekurang-kurangnya 3 ( $k \geq 3$ ) dan kita ingin menguji hipotesis bahwa semua rata-rata populasinya adalah sama.

$$H_0 = \mu_1 = \mu_2 = \dots = \mu_k$$

Untuk menguji perbedaan ini teknik uji yang cocok adalah nisbah F.

Tes statistiknya adalah

$$F = \frac{MS_b}{MS_w} \tag{8}$$

Dengan  $db = K - 1$  dan  $N - k$

$$MSb = \sum_{k=1}^k n_k (\bar{X}_k - \bar{X}) / (K - 1)$$

$$MSw = \sum_{k=1}^k n_k \left[ \sum_{i=1}^{N_k} (X_{ik} - \bar{X}_k) \right] / (N - K)$$

Dengan keterangan

$X_{ik}$  = skor pada variabel tergantung yang diperoleh subjek ke-I dalam kelompok ke-k.

$X_k$  = rata-rata kelompok ke-k

$X$  = rata-rata total

$N_k$  = besarnya kelompok ke-k

$N$  = jumlah sampel ( $=n_1+n_2+\dots+n_k$ )

Nisbah F hitung ini lalu dibandingkan dengan titik centile ke  $100(1-\alpha)$  dalam diatribusi F dengan db  $K-1$  dan  $N-K$

## B. Uji signifikansi Multivariat

Rationale dasar uji signifikansi, baik univariat maupun multivariat, terletak pada bertanya dan menjawab pertanyaan “Seberapa besar kemungkinan mendapatkan sampel yang menyimpang dari sampel yang diharapkan atas dasar hipotesis nol? Dalam kasus univariat “kemenyimpangan” sampel itu dinilai dari seberapa jauh rata-rata sampel kita itu menyimpang dari rata-rata populasi yang dihipotesiskan. Penyimpangan itu kemudian diterjemahkan kedalam probabilitas – probabilitas kondisional untuk mendapatkan rata-rata sampel menyimpang ini jika hipotesis nol benar. Dalam analisis multivariat rationalenya juga sama, cuma yang dipersoalkan adalah *centroid* sampel. Suatu *centroid* secara geometrik direpresentasikan oleh titik yang koordinatnya sama dengan rata-rata pada sejumlah variabel. Seperti pada analisis univariat, pada analisis multivariat juga ada tes satu sampel, tes dua sampel, dan tes K sampel. Namun dalam tulisan ini hal-hal teknis mengenai itu sengaja tidak disajikan.

## KRITIK TERHADAP UJI SIGNIFIKANSI HIPOTESIS NOL

Bagi peneliti yang menggunakan pendekatan kuantitatif statistika merupakan metode analisis yang *powerful*, karena objektivitasnya tinggi, komunikabilitasnya tinggi, dan daya generalisai tinggi. Karena itu sangat mengejutkan ketika metode

utamanya, yakni uji signifikansi hipotesis nol mendapat kritikan tajam. Sudah pada tahun enam puluhan Cohen (1962) menemukan bukti kelemahan ini. Sedlmeier dan Gigerenzer (1989) menelaah 12 studi empiris yang menghitung taraf kekeliruan pada uji signifikansi statistik dan menemukan bahwa taraf kekeliruan adalah 60%. Lipsey dan Wilson (1993) menelaah sebanyak 302 meta analisis mengenai studi perlakuan yang masing-masing terdiri dari 60 studi; jadi yang mereka telaah meliputi 18.120 studi. Dari 302 meta analisis itu ternyata hanya pada 3 meta analisis hipotesis nolnya benar. Jadi ternyata dalam studi empiris itu hipotesis nol 99% palsu. Karena hipotesis nol pada umumnya palsu, maka tingkat kesalahan yang terjadi adalah bukan 5% kesalahan tipe I, melainkan kesalahan Tipe II yang sangat besar, yang rata-ratanya mencapai 60%. Karena hasil-hasil yang demikian itulah maka Hunter (1997) mengusulkan agar uji signifikansi hipotesis nol dilarang saja.

Abelson (1997) menanggapi kritik-kritik itu (terlebih-lebih usul Hunter) sebagai berlebihan, Apalagi dia menemukan bahwa taraf kesalahan 60% rentangannya sangat luas, ada yang tinggi, ada yang rendah. Karena itu perbaikan metodologi diharapkan dapat mengatasi problem ini.

### **SARAN-SARAN PENGHATI-HATI**

Menghadapi kondisi seperti yang digambarkan di atas sejumlah ahli menyarankan agar para peneliti menerapkan uji signifikansi hipotesis nol itu dengan penuh kearifan. Beberapa diantara kehati-hatian yang disarankan ini disajikan dibawah ini.

a. Loftus (1997) menyarankan dipertimbangkannya enam hal sebagai berikut

(1) ketidakmungkinan hipotesis nol yang khas.

Uji signifikansi hipotesis nol berkuat dengan menguji hipotesis nol yang dalam kenyataannya memang tidak dapat benar. Misalnya, seorang peneliti menyajikan deretan angka pada layar komputer dalam variasi lima macam lamanya, dari 10 milidetik sampai 100 milidetik dan mengukur proporsi deretan yang diingat dengan benar. Dia akan mulai dengan hipotesis:

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \quad (9)$$

Dimana  $\mu$  menunjuk kepada rata-rata masing-masing kondisi. Logika hipotesis nol menuntut bahwa  $\mu$ -  $\mu$  itu tepat bulat, tanpa perbedaan sebarangpun kecilnya. Hipotesis nol mengenai rata-rata yang identik ini sebenarnya tidak realistik. Menurut Loftus yang dipersoalkan sebenarnya bukan ada tidaknya perbedaan, tetapi berapa besarnya perbedaan.

- (2) *Significance versus* pola rata-rata populasi yang mendasarinya.

Pencarian signifikansi statistik sebenarnya hanya mencari bukti bahwa hipotesis nol, yaitu  $\mu_1 = \mu_2 \dots \mu_j$  adalah palsu (tidak benar). Temuan dari upaya ini tidak memberi informasi mengenai pola rata-rata populasi, yang sebenarnya justru perlu dalam penelitian psikologis, guna membuat kesimpulan ilmiah. Hal ini biasanya diatasi dengan *post hoc tests* atau dengan *planned comparisons*. Tetapi kedua teknik ini masing-masing juga mempunyai keterbatasan.

- (3) *Power*

Masalah *power* ini pada umumnya diabaikan. Dalam praktek uji signifikansi hipotesis nol berpusat pada menghindari kesalahan Tipe I ( $\alpha$ ), terutama karena probabilitas kesalahan tipe I ini dapat dihitung. Sebaliknya probabilitas kesalahan Tipe II ( $\beta$ ) dan *power* ( $1-\alpha$ ) pada umumnya tidak dapat dihitung, oleh karena untuk menghitungnya diperlukan hipotesis kuantitatif khusus (misalnya  $\mu_1 = \mu_2 + 10$ ) pada dua kondisi eksperimen), dan hipotesis kuantitatif itu sangat jarang dalam ilmu-ilmu sosial.

- (4) Dikotomi *effect/non-effect* yang artifisial. Pola pikir dikotomis ini membuat orang kurang berhati-hati dalam merumuskan keputusannya. Mestinya orang tidak menyatakan “kita gagal menyimpulkan bahwa hipotesis nol palsu”, tetapi lebih cenderung menyatakan “hipotesis nol benar”. Kebanyakan orang, jika didesak, akan setuju tidak ada perbedaan antara mendapatkan  $p=0,05$  dan  $p=0,051$ . Namun dalam praktek orang lupa bahwa titik potong 0,05 itu adalah riil dan bukan arbitrer. Sehingga dunia penelitian terbagi menjadi ada *real effect* ( $p < 0,05$ ) dan *non-effect* ( $p > 0,05$ ).

- (5) Uji hipotesis nol tidak cukup mampu untuk mencerminkan kompleksitas kenyataan.

Penerapan uji hipotesis nol didasari asumsi-asumsi (a) variabel tergantung dicapai dengan menjumlahkan *effect* numerik (kombinasi linier) *effect* yang ditimbulkan oleh variabel bebas, interaksi antara variabel bebas, dan berbagai “sumber” kesalahan, (b) kesalahan-kesalahan itu berdistribusi normal, (c) distribusi variansi kesalahan-kesalahan itu adalah sama untuk berbagai kondisi.

Jadi sifat teknik analisis itu pada umumnya mendikte teori psikologisnya, sehingga teori psikologis menjadi *generic linear-model theory*.



- (6) Uji signifikansi hipotesis nol hanya memberikan informasi yang tak akurat mengenai validitas hipotesis nol. Ketidacermatan ini telah ditunjukkan oleh ahli-ahli statistika Bayesian. Secara konvensional orang menolak hipotesis nol bila: (a)  $p$  (data yang diobservasi/hipotesis nol)  $< 0.05$ , tetapi menolak hipotesis nol mengimplikasikan bahwa; (b)  $p$  (hipotesis nol/data yang diobservasi adalah kecil). Tanpa tambahan informasi, tidak ada dasar yang logis untuk menyimpulkan validitas Pernyataan (b) berdasarkan temuan pada Pernyataan (a) namun kesimpulan-kesimpulan ini hanya berlaku jika ada kepastian (ada data pendahulu=prior data) bahwa hipotesis nol memang benar. Tanpa kepastian ini sebenarnya besaran 0,05 atau yang lain itu dapat dikatakan adalah ilusi.

## PENUTUP

Uji signifikansi hipotesis nol adalah metode pokok dalam analisis statistik. Statistik sendiri sebagai matematika terapan memberikan alat (*tool*) yang sangat *powerful* bagi peneliti. Namun, betapapun baiknya alat itu, buahnya akan tergantung kepada penggunaannya. Karena itu penggunaan uji signifikansi hipotesis nol secara arif (tidak membabi buta) akan sangat bermanfaat bagi kemajuan psikologi. Beberapa alternatif telah disarankan, seperti (a) penyajian hasil secara grafis dan tidak dalam tabel, (b) penggunaan interval konfidensi, (c) penerapan *post hoc test* dan *planned comparisons*, dan (d) pelaksanaan meta-analisis

## DAFTAR 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.
- Cohen, J. 1962. The statistical power of abnormal-social psychological research, *Journal of Abnormal and Social Psychology*, vol. 65, 145-153.
- Estes, W. K. 1997. Significance testing in psychological Research: Some persisting issues. *Psychological Science*, vol. 8, 18-20.
- Greenwald, A. G. 1993. Consequences of prejudice against the null hypotheses. In G. Keren & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Erlbaum, p. 419-448.
- Harris, J. R. 1997. Significance tests have their place. *Psychological Science*, vol. 8, 8-11.

- Hunter, J. E. 1997. Needed: A ban on the significance test. *Psychological Science*, vol. 8, 3-7.
- Lipsey, M. W. & Wilson, D. B. 1993. The efficacy of psychological, educational, and behavioral treatments: Confirmation from meta-analysis. *American Psychologist*, vol.48, 1181-1209.
- 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.
- Schmidt, F. L. 1996. Statistical significance testing and cumulative knowledge in psychology: Implication for training of researchers. *Psychological Methods*, 1, 115-129.
- Sedlmeier, P., Gigerenzer, G. 1989. Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Shrout, P. E. 1997. Should significance tests be banned ? Introduction to a special section exploring pros and cons. *Psychological Science*, vol. 8. 1-2.
- Tatsuoka, M. M. 1971. *Selected topics in advanced statistics: Significance tests*. Champaign, Il.,:Institute for Personality and Ability Testing.