

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.

PENGUJIAN HIPOTESIS NIHIL: UJI SIGNIFIKANSI DAN INTERVAL KEPERCAYAAN

Djemari Mardapi

PENDAHULUAN

1. Sikap orang terhadap statistik bermacam-macam, ada yang senang, kagum, ada yang sinis, curiga, dan tidak senang. Bahkan ada yang mengatakan bahwa orang mudah berbohong dengan statistik. Seseorang yang sebagian anggota badannya di almari pendingin dan sebagian lainnya di alat pemanas, disimpulkan keadaan orang tersebut normal. Tentunya kesimpulan ini hanya melihat harga rerata saja tanpa memperhatikan besarnya *range* dan simpangan baku.
2. Sebagian orang menghindari statistik karena bias filsafat, kekakuan statistik, atau karena salah konsep terhadap disiplin statistik (Glass & Hopkins, 1984). Mereka cenderung bekerja berdasarkan tradisi, intuisi, kekuasaan atau rasa kewajaran. Namun lambat laun, orang memerlukan adanya cara yang sistematis dan objektif yang dapat digunakan pada penelitian empiris, yaitu statistik.
3. Statistik adalah ilmu menjawab pertanyaan berdasarkan data empiris (Olson, 1987). Definisi ini menunjukkan bahwa statistik adalah alat atau teknik untuk menyajikan dan mengolah data yang diperoleh melalui pengukuran sehingga menjadi informasi yang berguna bagi pembuat kebijakan. Sesuai fungsinya sebagai alat, statistik tidak dapat mengubah data yang tidak baik menjadi baik.
4. Data yang diolah dengan statistik harus mengandung kesalahan yang kecil. Untuk memperoleh data yang memiliki kesalahan yang kecil, alat ukur yang digunakan harus sah dan handal, dan cara penggunaan alat ukur juga harus baku. Data dengan banyak kesalahan tidak akan dapat diubah menjadi baik dengan teknik statistik apapun. Sebagai alat, kualitas keluarannya tergantung

8. pada kualitas masukannya. Oleh karena itu, statistik membutuhkan data yang sah dan handal.
5. Banyak teknik statistik yang dapat digunakan untuk mengolah data. Dalam menentukan teknik yang tepat, jenis data yang akan diolah harus diketahui terlebih dahulu. Berdasarkan peringkat pengukuran, data dapat dikategorikan menjadi empat, yaitu data nominal, data ordinal, data interval, dan data rasio. Setiap jenis data memiliki teknik statistik tertentu sesuai dengan pertanyaan penelitian akan hipotesis penelitian. Data yang diolah dengan statistik kemudian diartikan hasilnya dengan menggunakan bahasa yang mudah dipahami oleh semua orang, khususnya para pembuat kebijakan.
6. Ada dua aspek utama ilmu statistik, yaitu statistik deskriptif dan statistik inferensial. Statistik deskriptif berhubungan dengan metode merangkum data tanpa ada niat untuk membuat kesimpulan di luar data tersebut. Kesimpulan hanya berlaku pada data yang diteliti. Statistik inferensial berhubungan dengan metode membuat generalisasi terhadap populasi berdasarkan data pada sampel. Dalam hal ini, statistik digunakan untuk menaksir besarnya parameter, yaitu deskripsi numerik tentang populasi. Deskripsi numerik tentang sampel disebut dengan statistik.
7. Pada prinsipnya statistik inferensial ingin menaksir besarnya harga parameter, yaitu besaran pada populasi. Ada dua cara yang dapat digunakan untuk menaksir besarnya parameter, yaitu taksiran titik dan taksiran interval. Taksiran titik hasilnya hanya satu harga, dan digunakan untuk melakukan uji signifikansi. Taksiran interval hasilnya berupa batas atas dan bawah suatu parameter, dan digunakan untuk menguji hipotesis dengan menggunakan interval kepercayaan. Jadi taksiran titik dan taksiran interval dapat digunakan untuk menguji hipotesis.
8. Statistik inferensial selalu berhubungan dengan hipotesis nihil dan hipotesis alternatif. Setelah dilakukan taksiran parameter selanjutnya dilakukan pengujian hipotesis atau uji signifikansi yaitu untuk menentukan apakah perbedaan antara besarnya statistik dari sampel dengan besarnya parameter pada populasi benar nyata atau disebabkan karena kesalahan acak. Bila perbedaan ini bukan karena kesalahan acak, maka disimpulkan secara statistik perbedaannya signifikan.
9. Pengujian hipotesis nihil dengan uji signifikansi dan interval kepercayaan, masing-masing memiliki keunggulan dan kelemahan. Oleh karena itu, makalah ini akan membahas pengujian hipotesis nihil dengan cara taksiran titik dan interval kepercayaan. Walaupun ada perbedaan antara dua cara tersebut, namun

10. keduanya membutuhkan informasi bentuk distribusi pencuplikan untuk menentukan statistik yang tepat.
10. Penentuan teknik statistik yang tepat juga ditentukan sejauhmana hasilnya tetap tegar bila asumsi yang mendasarinya tidak dipenuhi. Berdasarkan asumsi distribusi pencuplikan dan berdasarkan peringkat pengukuran, ada dua teknik statistik yang dapat digunakan, yaitu parametrik dan nonparametrik. Statistik parametrik menggunakan asumsi bentuk distribusi peluang, sedang statistik nonparametrik cenderung tidak menggunakan asumsi bentuk distribusi peluang, namun daya statistik non parametrik cenderung lebih rendah dibanding statistik parametrik.

TAKSIRAN PARAMETER

1. Hasil taksiran harus memenuhi dua kriteria, yaitu akurat dan teliti. Akurat artinya bebas dari kesalahan sistematis dan teliti artinya secara relatif bebas dari kesalahan acak (Olson, 1987). Untuk itulah setiap penelitian harus melaporkan kualitas datanya dengan menunjukkan bukti kesahihan dan kehandalan instrumen atau alat ukur.
2. Ada dua cara untuk menaksir besarnya harga parameter, yaitu dengan taksiran titik dan taksiran interval. Taksiran titik adalah harga numerik tunggal yang diperoleh berdasarkan data sampel dan digunakan sebagai indikasi harga parameter. Taksiran titik diikuti dengan informasi tentang galat baku taksiran. Taksiran interval adalah dua bilangan numerik yang ditentukan berdasarkan data sampel dan diikuti dengan pernyataan besarnya suatu interval kepercayaan bahwa suatu interval mencakup besarnya parameter populasi (Olson, 1987). Masing-masing cara ini memiliki keunggulan dan keterbatasan. Besarnya taksiran ini selanjutnya digunakan untuk pengujian hipotesis nihil.
3. Ada tiga sifat penaksir parameter yang harus dielaah, yaitu tidak bias, konsistensi, dan efisiensi (Glass & Hopkins, 1984). Penaksir parameter dikatakan tidak bias bila rerata distribusi pencuplikan sama dengan besarnya parameter pada populasi. Ada beberapa penaksir yang tidak bias, di antaranya adalah: statistik rerata \bar{x} sebagai penaksir μ yang tidak bias; statistik s^2 dengan pembagi $(n-1)$ adalah penaksir σ^2 yang tidak bias, tetapi bila pembagiannya n menjadi penaksir yang bias. Penaksir yang bias di antaranya adalah: statistik r untuk menaksir besarnya ρ ; statistik s^2 untuk menaksir besarnya σ^2 dengan pembagi n .

4. Penaksir dikatakan konsisten walaupun hasil taksirannya bias, apabila hasil taksirannya semakin mendekati harga parameter bila ukuran cuplikan semakin banyak. Besarnya statistik s akan mendekati besarnya σ apabila ukuran cuplikannya banyak. Sifat penaksir yang ketiga adalah efisiensi relatif, yaitu ketepatan hasil taksiran dari suatu penaksir. Hal ini dilihat dari variasi statistik dari cuplikan ke cuplikan, yaitu derajat galat pencuplikan. Semakin kecil galat ini akan semakin baik statistik yang digunakan. Menurut Glass dan Hopkins (1984), efisiensi relatif lebih penting dari pada bias dan konsistensi.

TAKSIRAN TITIK

1. Taksiran titik adalah harga numerik tunggal yang diperoleh berdasarkan data cuplikan digunakan sebagai indikasi harga parameter. Taksiran titik pada statistik selalu diikuti dengan galat baku taksiran, dihitung berdasarkan bentuk distribusi pencuplikan. Semakin kecil galat baku ini akan semakin teliti hasil estimasi, sebaliknya semakin besar galat baku taksiran akan semakin kurang teliti hasil estimasi.
2. Taksiran titik dapat digunakan untuk menguji hipotesis nihil dan sering disebut dengan uji signifikansi. Uji signifikansi selalu disertai dengan informasi peringkat signifikansi atau kesalahan tipe I, dan kesalahan tipe II. Pengujian hipotesis dengan pendekatan Neyman-Pearson menggunakan nilai alpha, yaitu kesalahan tipe I untuk menentukan daerah penolakan hipotesis nihil. Menurut Huberty (1987), pada penelitian penajagan bisa digunakan alpha 0,10 atau lebih besar, pada penelitian dengan ubahan yang terjadi pada masa lalu disarankan menggunakan alpha 0,01, sedang untuk pengujian ganda dengan statistik (*multiple tests*) disarankan menggunakan alpha 0,10 sampai 0,20. Nilai alpha ini tetap besarnya dan ditentukan sejak awal, saat merancang suatu penelitian.
3. Pengujian hipotesis dengan pendekatan Fisher tidak menggunakan nilai *alpha* tetapi menggunakan nilai p , yaitu peluang kesalahan menolak hipotesis nihil bila hipotesis ini benar. Nilai p adalah ubahan acak dan sering digunakan pada laporan-laporan penelitian yang ditulis di jurnal dan selalu muncul sebagai hasil *print-out* paket program komputer untuk analisis data seperti SPSS. Besarnya p untuk menyatakan signifikan atau tidak diserahkan pada pembuat kebijakan.
4. Informasi lain yang perlu diketahui dalam pengujian hipotesis nihil dengan uji signifikansi adalah daya statistik, yaitu besarnya peluang menolak hipotesis nihil

yang salah. Setiap pengujian ditungjukan agar daya statistiknya tinggi, namun daya statistik ini ditentukan oleh nilai *alpha* dan besarnya ukuran cuplikan.

5. Huberty (1987) mengajukan pendekatan Hybrid berdasarkan pembahasan pada pendekatan Neyman-Pearson dan Fisher untuk menguji hipotesis nihil. Pada pendekatan ini, saat merancang penelitian ditetapkan besarnya alpha, beta, dan kuadrat eta. Alpha adalah kesalahan tipe-I, beta kesalahan tipe-II, sedang kuadrat eta adalah indeks variasi yang dapat dijelaskan. Huberty menganjurkan rasio beta dan alpha sekitar 0,30. Besarnya kuadrat eta yang dianjurkan adalah 0,10 dan hasil yang diperoleh dari perhitungan digunakan untuk membuat penafsiran hasil analisis data.

6. Uji signifikansi menggunakan informasi taksiran titik untuk menguji hipotesis nihil. Semua pengujian hipotesis nihil dapat menggunakan uji signifikansi, sehingga uji signifikansi banyak digunakan pada pengujian hipotesis dengan teknik statistik mulai yang sederhana seperti uji beda rerata pada univariat sampai uji beda pada multivariat. Pengujian hipotesis nihil pada analisis regresi, analisis jalur, analisis faktor, dan persamaan model struktural menggunakan uji signifikansi.

INTERVAL KEPERCAYAAN

Konsep interval kepercayaan adalah sebagai berikut:

1. Andalkan μ adalah harga parameter populasi yang ingin ditaksir besarnya berdasarkan data pada sampel. Pada pendekatan ini harus ditentukan batas atas dan batas bawah interval. Misalnya, batas bawahnya 80 dan batas atasnya adalah 100. Apakah harga μ berada pada batas interval tersebut, tentu jawabannya kita tidak tahu. Apabila taksiran ini dilakukan berulang-ulang dengan memilih cuplikan berulang-ulang dengan ukuran dan prosedur yang sama akan diperoleh interval yang berbeda-beda, yaitu batas atas dan batas bawah yang berbeda.
2. Pertanyaan yang timbul adalah apakah interval yang diperoleh akan mencakup harga parameter tersebut? Tentu jawabannya tidak tahu, karena bisa mencakup harga parameter dan bisa tidak. Setiap pengambilan cuplikan akan menghasilkan batas interval yang berbeda, ada yang mencakup harga parameter dan ada yang tidak. Apabila dipilih cuplikan secara acak berulang-ulang sebanyak 100 kali, dan ada 95 buah interval yang mencakup harga parameter, maka dikatakan interval kepercayaan 95 persen harga parameter berada pada batas interval (Blommers & Forsyth 1977). Atau dengan kata lain dapat dikatakan bahwa

3. Lebar interval menentukan kecermatan estimasi, dan besarnya interval keyakinan menentukan besarnya keyakinan kita tentang letak harga μ . Besarnya interval ditentukan oleh bentuk distribusi pencuplikan. Apabila bentuk distribusinya simetris maka letak batas interval atas dan bawah juga simetris. Berdasarkan pada teorema limit pusat, walaupun distribusi populasi induk tidak simetris, namun bila ukuran cuplikan yang dipilih banyak, maka distribusi pencuplikan harga rerata akan cenderung simetris (Walpole & Myers, 1989).
4. Interval kepercayaan dapat digunakan untuk menguji hipotesis nihil, karena pengujian hipotesis nihil pada prinsipnya adalah menaksir besarnya parameter. Glass dan Hopkins (1984) memberi contoh penggunaan interval kepercayaan untuk menaksir besarnya parameter pada distribusi normal, distribusi yang persegi, dan distribusi yang tidak simetris. Pada interval kepercayaan 68 % hasil taksiran ketiga distribusi tersebut untuk ukuran cuplikan yang lebih besar dari 25 tidak jauh berbeda hasilnya. Oleh karena itu, walau bentuk distribusinya tidak simetris, konsep interval kepercayaan dapat digunakan untuk menaksir besarnya parameter.
5. Banyak teknik statistika yang menggunakan interval kepercayaan untuk menguji hipotesis nihil. Menurut Lehmann (1986), interval kepercayaan adalah famili pengujian hipotesis. Di antaranya adalah uji univariate, uji-t, anava, dan anakova. Berikut ini beberapa contoh penggunaan interval kepercayaan untuk menguji hipotesis nihil:
 - Uji univariate dengan distribusi normal atau distribusi-t.
 - 95 % interval kepercayaan:
 - .95 IK = $\bar{X} \pm 1,96 \sigma_{\bar{X}}$
 - Uji beda rerata pada Varians
 - $(1-\alpha)IK = \bar{X}_j \pm \sqrt{(1-\alpha)F_{1-\alpha/2, v_w} (MS_w / n_j)}$
 - $= \bar{X}_j \pm (1-\alpha)t_{v_w} S_{Xj}$
 - Korelasi dengan Z-Transformasi Fisher

- Banyak digunakan pada pengujian hipotesis tentang beda rerata, karena distribusi pencuplikannya simetris, sedang untuk pengujian hipotesis dengan statistik multivariat sulit untuk digunakan, karena penghitungannya menjadi kompleks.

KOMPARASI PENGUJIAN HIPOTESIS NIHIL

1. Strategi pengujian hipotesis melibatkan suatu keputusan berkaitan dengan hipotesis statistik H_0 , yaitu keputusan apakah H_0 benar atau tidak. Hipotesis statistik merupakan pernyataan numerik tentang populasi. Pengujian hipotesis statistik berdasarkan data empirik yang diperoleh melalui pengukuran baik melalui tes maupun nontes. Pengujian hipotesis nihil dengan uji signifikansi dan interval kepercayaan membutuhkan informasi tentang bentuk distribusi pencuplikannya.
2. Pengujian hipotesis dengan uji signifikansi menggunakan taksiran titik disertai dengan peringkat signifikansi untuk menolak atau menerima hipotesis nihil. Penggagas pengujian hipotesis adalah Neyman dan Pearson, sedang penggagas pengujian signifikansi adalah Fisher (Huberty, 1987). Neyman-Pearson menggunakan nilai-alpha untuk menentukan daerah penolakan hipotesis nihil. Fisher menggunakan besarnya peluang kesalahan p untuk menolak hipotesis nihil.
3. Selanjutnya berdasarkan pada kelemahan pada uji signifikansi dan uji hipotesis, Huberty mengajukan pendekatan Hybrid untuk menguji hipotesis yaitu dengan menambahkan kriteria beta, yaitu kesalahan tipe II, dan eta kuadrat.
4. Interval kepercayaan dapat digunakan untuk menguji hipotesis nihil, terutama untuk hipotesis dengan statistik univariate. Pengujian hipotesis nihil dengan interval kepercayaan menggunakan uji dua sisi atau hipotesis alternatif dua sisi. Misalnya, interval kepercayaan 95 %, berarti sisi kiri 2,5 % dan sisi kanan 2,5%. Analisis data dengan statistik multivariat belum ada yang menggunakan interval kepercayaan untuk menguji hipotesis nihil (Hair, Anderson, cs, Tatham, Black, 1998; Stevens, 1986). Penggunaan interval kepercayaan seperti pada uji signifikansi memerlukan informasi bentuk distribusi pencuplikan (*sampling distribution*) dan simpang bakunya.
5. Uji signifikansi dan interval kepercayaan membutuhkan informasi tentang bentuk distribusi pencuplikannya. Pendekatan yang pertama menggunakan peringkat kesalahan atau peringkat signifikansi, sedang pendekatan yang kedua

menggunakan interval kepercayaan. Sering dibuat analogi antara peringkat kesalahan dan interval kepercayaan. Peringkat kesalahan 5 % analogi dengan interval kepercayaan 95 %. Uji signifikansi dapat menguji hipotesis alternatif satu sisi dua sisi, namun interval kepercayaan cenderung menguji hipotesis dua sisi. Dalam hal ini antara hipotesis dua sisi dan satu sisi juga ada yang masih mempermasalahakan.

6. Cohen (Pillemer, 1991) salah satu pakar statistik yang menyatakan bahwa pengujian hipotesis pada bidang sosial terlalu berlebihan. Sebagai alternatif ia menganjurkan menggunakan ukuran efek – beda rerata antar kelompok, korelasi, besaran apa saja pada fenomena yang tepat untuk suatu kontak penelitian. Cohen juga menganjurkan penggunaan ukuran efek disertai interval kepercayaan untuk menjelaskan batas nilai suatu parameter dibandingkan menggunakan uji hipotesis dan nilai peluang.

7. Keunggulan uji signifikansi adalah kemudahan dalam menentukan menolak atau menerima hipotesis nihil, baik menggunakan alpha maupun yang menggunakan besarnya peluang statistik seperti yang dibahas oleh Huberty. Apabila tingkat kesalahan tipe I lebih kecil dari besarnya *alpha* yang ditetapkan atau apabila besarnya statistik lebih besar dari harga kritik, maka hipotesis nihil ditolak atau dinyatakan data tidak mendukung hipotesis nihil. Kelemahan pendekatan ini adalah ketergantungan pada ukuran cuplikan, karena semakin besar ukuran cuplikan akan semakin besar peluang menolak hipotesis nihil.

8. Contoh perbandingan antara penggunaan uji signifikansi dan interval kepercayaan:

Misalkan: $H_0: \mu = 100$

Statistik diperoleh rerata = 101 dengan $n = 2000$. Pada $\alpha = 0,01$ dan dengan $\sigma = 15$, hipotesis nihil akan ditolak. Namun perbedaan 101 dan 100 sebesar 1 secara praktis tidak berarti, walau secara statistik signifikan. Oleh karena itu, penentuan kebijakan tidak saja berdasarkan pada signifikansi statistik, tetapi juga dilihat dari signifikansi secara praktis. Dengan menggunakan interval kepercayaan diperoleh batas interval sebesar 100,12 sampai 101,86, sehingga disimpulkan bahwa harga parameter tercakup pada interval kepercayaan 99 %, sehingga hipotesis nihil tidak ditolak. Jadi uji signifikansi dipengaruhi oleh ukuran cuplikan, sedang interval kepercayaan tidak dipengaruhi oleh ukuran cuplikan.

9. Menurut Box (Glass & Hopkins, 1984) uji signifikansi pada umumnya terlalu berlebihan, dan dalam banyak kasus pernyataan signifikan akan lebih baik bila dilengkapi dengan interval kepercayaan untuk menyatakan batas interval yang mencakup besarnya parameter.

10. Untuk pengujian hipotesis yang sederhana, uji signifikansi dan interval kepercayaan dapat digunakan. Penggunaan interval kepercayaan tampak lebih baik dibanding hanya dengan uji signifikansi karena pengaruh dari ukuran cuplikan dan pentingnya signifikansi secara praktis.

11. Analisis data untuk pengujian suatu model kausal, seperti analisis regresi, analisis jalur, model struktural sulit bila menggunakan interval kepercayaan, sehingga cenderung menggunakan taksiran titik. Oleh karenanya, penggunaan uji signifikansi disertai dengan signifikansi secara praktis lebih mudah digunakan untuk pengujian model kausal.

12. Hasil uji signifikansi dan interval kepercayaan ditentukan oleh kesahihan dan kehandalan data yang dianalisis. Kesalahan hasil analisis data tidak saja ditentukan oleh uji signifikansi atau interval kepercayaan tetapi juga ditentukan oleh kesalahan pengukuran. Oleh karena itu, setiap penelitian harus mencantumkan indeks kehandalan dan kesahihan data.

KESIMPULAN

1. Pengujian hipotesis nihil univariat yang sederhana dengan pendekatan interval kepercayaan tampak memberikan informasi lebih baik dibanding dengan uji signifikansi. Namun apabila signifikansi secara statistik disertai dengan signifikansi secara praktis, uji signifikansi masih tetap bisa digunakan. Bahkan dianjurkan menggunakan keduanya, yaitu uji signifikansi disertai informasi interval kepercayaan untuk mengetahui letak parameter.

2. Pengujian suatu model lebih mudah menggunakan uji signifikansi disertai tafsiran signifikansi secara praktis, karena kompleksitas penghitungan. Model hubungan ubahan yang menggunakan model persamaan struktural dan model pengukuran sekaligus meninjau kesalahan pengukuran cenderung menggunakan uji signifikansi (Pedhazur, 1982). Demikian pula analisis data dengan statistik multivariat cenderung menggunakan uji signifikansi.

DAFTAR PUSTAKA

- Blommers, P. J., & Forsyth, R. A. (1977). *Elementary statistical methods in psychology and education*. Boston: Houghton Mifflin Company.
- Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Hair, E. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. London: Prentice-Hall International.
- Hays, W. K. (1981). *Statistics*. New York: Holt, Rinehart and Winston.
- Huberty, C. J. (1987). On statistical testing. *Educational researcher*, 16, 8, 4-9.
- Lehmann, E. L. (1986). *Testing statistical hypothesis*. New York: John-Wiley & Sons.
- Olson, C. L. (1987). *Statistics: Making sense of data*. Boston: Allyn and Bacon, Inc.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart and Winston.
- Pillemer, D. B. (Des. 1991). One-versus two-tailed hypothesis tests in contemporary educational research. *Educational researchers*, 20, 9, 13-17.
- Stevens, James. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Walpole, R. E., & Myers, R. H. (1989). *Probability statistics for engineers and scientist*. Penerjemah: Sembiring, R. K. *Ilmu Peluang dan statistika untuk insinyur dan ilmuwan*. Bandung: Penebiti ITB.

KETERBATASAN UJI SIGNIFIKANSI HIPOTESIS NOL

Nonny Swehati 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_0) dari satu set data (D) yang diperoleh, $P(H_0|D)$. Sebaliknya, uji signifikansi hipotesis nol menekankan pada pertanyaan berapakah probabilitas untuk memperoleh data tersebut jika hipotesis nolnya benar, $P(D|H_0)$. Sayangnya, hal ini tidak berarti bahwa jika kita memperoleh data dengan $P(D|H_0)$ rendah maka $P(H_0|D)$ 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