KUASA HIPOTESIS DALAM PARADIGMA NON PARAMETRIK

oleh :

Anung Prasetyo Nugroho, SE.,M.MA.
Swaidatul Masluhiya AF., S.Si., M.Ked.Trop.

Penerbit : UNITRI Press, Anggota IKAPI
Jalan Telagawarna, Tlogomas,Malang
Telp. (0341) 565500 Fax (0341) 565522

BAB 1
Pendahuluan dan tinjauan

Pokok pembahasan statistika pada hakikatnya mencakup kegiatan-kegiatan, gagasan-gagasan, serta hasil-hasil yang sangat beraneka ragam. Mereka yang berkecimpung di bidang statistika biasanya memaklumi kenyataan bahwa displin ini terbagi atas dua golongan besar, yakni: *statistika deskriptif* dan *statistika inferensial*. Statistika deskriptif berkaitan dengan kegiatan pencatatan dan peringkasan hasil-hasil pengamatan terhadap kejadian-kejadian atau karakteristik-karakteristik manusia, tempat dan sebagainya, secara kuantitatif. Catatan-catatan mengenai jumlah kelahiran, kematian, dan perkawinan per tahun disebut statistik. Demikian pula deskripsi mengenai usia, tingkat pendidikan, serta komposisi etnik penduduk yang tinggal di suatu daerah. Adapun inferensi statistik, atau statistika inferensial, adalah statistika yang menyangkut kegiatan penarikan kesimpulan dari fakta-fakta seperti tersebut di atas serta pengambilan keputusan berdasarkan fakta-fakta itu.

Buku ini terutama dimaksudkan untuk mempelajari statistika inferensial dengan pengandaian bahwa sebagian besar pembaca sekurang-kurangnya pernah mengikuti mata kuliah pengantar di bidang ini. Kedatipun demikian, tinjauan ulang secara sekilas tentang beberapa konsep yang penting tentu tidak akan merupakan beban tambahan bahkan mungkin akan sangat membantu. Sebab itu tiga bagian pertama bab ini disediakan untuk sekali lagi membahas dasar-dasar statistika secara umum. Kemudian, pada bagian Bab 4, kita akan membicarakan skala-skala pengukuran. Bagian Bab 5 mengantar kita ke konsep-konsep dasar statistika nonpara*metrik*, yang tidak lain merupakan pokok pembahasan buku ini. Dalam Bagian Bab 6 kita diajak meninjau ke depan, secara sekilas, dari Bab 1 sampai dengan Bab 11, sedangkan Bagian Bab 7 menjelaskan format yang akan kita gunakan untuk menyajikan teknik-teknik statistik dalam bab-bab selanjutnya.

1.1 BEBERAPA TERMINOLOGI PENTING

Dalam bagian ini kita mendefinisikan beberapa istilah yang akan digunakan dalam bab-bab mendatang. Istilah-istilah ini merupakan bagian dari perbendaharaan kata para ahli statistika. Adapun istilah-istilah yang lain nanti akan kita definisikan begitu kita jumpai.

Populasi Kita akan menggunakan istilah *populasi* bila yang kita maksudkan adalah suatu kumpulan, entah kumpulan orang, tempat, atau benda. Sedangkan mengenai kumpulan mana yang membentuk populasi dalam suatu pembicaraan, ini bergantung pada minat dan kepentingan masing-masing peneliti. Peneliti yang satu mungkin ingin membuat pengamatan tentang mahasiswa-mahasiswa di semua perguruan tinggi di Amerika Serikat. Peneliti yang lain mungkin ingin membuat pengamatan tentang mahasiswa-mahasiswa di suatu perguruan tinggi tententu saja. Kedua peneliti itu sama-sama menganggap bahwa bagi mereka populasi adalah kumpulan mahasiswa yang akan menjadi obyek penyelidikan mereka.

Dalam beberapa konteks tersebut, kita juga akan menggunakan istilah populasi untuk kumpulan *pengukuran* atau *collection of measurements* (kadang-kadang disebut pengamatan) yang dilakukan terhadap sekumpulan orang, tempat, atau benda. Sebagai contoh, kalau kita berminat menyelidiki usia semua mahasiswa di suatu perguruan tinggi, yang kita pandang sebagai populasi adalah kumpulan usia tadi. Lebih khusus lagi, kita mungkin mendefinisikan populasi sebagai *kumpulan orang, tempat, atau benda yang paling besar* (termasuk pengukuran-pengukuran) yang kita *minati*. Populasi bisa terhingga, bisa pula tak terhingga.

Yang berikut ini adalah beberapa contoh *populasi* terhingga : segenap mahasiswa yang saat ini terdaftar di suatu perguruan tinggi;segenap penghuni di suatu daerah tertentu dalam kurun waktu yang tertentu;semua barang dengan jenis tertentu yang dihasilkan suatu pabrik pada suatu hari.

Sedangkan yang berikut ini adalah beberapa contoh *populasi* tak terhingga : segenap mahasiswa yang pernah terdaftar (dahulu, sekarang, atau di masa mendatang) di suatu perguruan tinggi; semua orang yang pernah, sedang, serta akan menghuni suatu daerah tertentu; semua barang dengan jenis tertentu yang pernah dan akan diproduksi oleh suatu pabrik.

Populasi bisa nyata *(real)*, bisa pula *hipotetik (hypothetical)*. Contoh populasi nyata adalah segenap mahasiswa yang sekarang terdaftar di suatu universitas. Sedangkan contoh populasi hipotetik adalah sebagai berikut. Andaikan Anda sedang merancang sebuah eksperimen untuk mengevaluasi kesangkilan *(effectiveness)* tiga jenis obat penenang, dan merencanakan akan memilih sejumlah kelinci percobaan secara acak *(random)* yang masing-masing akan diberi salah satu dari ketiga jenis obat tadi. Kita bisa menganggap masing-masing dari ketiga kelompok kelinci percobaan ini sebagai sampel dari suatu populasi yang beranggotakan sejumlah besar kelinci percobaan yang dapat diberi obat tersebut. Populasi semacam ini tidak ada belum ada; namun demikian *hipotetik* atau *potensial* (masih mungkin ada).

Sampel Sampel adalah bagian dari suatu populasi. Andaikanlah bahwa suatu populasi tertentu terdiri atas semua mahasiswa yang ada di suatu perguruan tinggi. *Bagian*, atau *subhimpunan (subset)*, dari semua mahasiswa yang terdaftar di jurusan statistika akan membentuk suatu kumpulan yang disebut sampel. Kita dapat mengenali sampel-sampel lain dalam populasi yang sama. Sebagai contoh, para mahasiswa yang mengambil kuliah bahasa ing gris bisa dipandang sebagai sebagai sebuah sampel, demikian pula para mahasiswa yang telah menikah, atau para mahasiswa yang memiliki tanda izin parkir di kampus bagi kendaraan mereka.

Kita, tentu saja, boleh mengambil sampel dari populasi yang tak terhingga, sebagaimana halnya dari populasi yang terhingga. Seorang ahli ilmu sosial, misalnya, mungkin tertarik pada beberapa karakteristik orang-orang dewasa yang pernah, sedang, dan akan tinggal di Greenwich Village, New York. Ilmuwan itu pasti akan mengakui bahwa populasi ini tak terhingga. Sampel yang diambil dari sejumlah orang dewasa yang kini tinggal di Greenwich Village akan merupakan sampel dari populasi yang akan tak terhingga.

Sampel acak Inferensi statistik antara lain mencakup tahapan penarikan kesimpulan mengenai suatu populasi atas dasar informasi yang terkandung dalam sebuah sampel. Apabila populasi yang dihadapi cukup besar atau bahkan tak terhingga, tentu tidak praktis atau tidak mungkin jika harus kita harus meneliti setiap anggota populasi termaksud untuk mengumpulkan informasi yang akan mendasari penarika kesimpulan tentang populasi itu secara keseluruhan. Bila kita menggunakan inferensi statistik untuk menarik kesimpulan tentang populasi berdasarkan informasi yang berasal dari sampel-sampel, sampel yang dipilih secara asal-asalan tentu saja tidak cukup. Kesahihan *(validity)* hasil-hasil yang diperoleh melalui metode inferensi statistik bergantung pada asumsi bahwa sampel bertipe khusus, yang disebut sampel acak *(random sample),* telah digunakan dalam proses itu.

Guna mendapatkan sampel acak berukuran *n*, kita memilihnya dengan cara sedemikian rupa sehingga probabilitas atau peluangnya untuk terpilih telah diketahui atau ditentukan sejak awal. Tipe sampel acak yang paling sederhana adalah *sampel acak sederhana (simple random sample).* Sementara jenis sampel acak yang lain misalnya adalah sampel acak berlapis *(stratified random sample)* dan *sampel gerombol gugus (cluster sample).*

Sampel yang dipilih karena mudah diperoleh *(samples of convenience)* Kalau Anda baru saja usai mempelajari statistika dan benak Anda masih penuh dengan keyakinan bahwa pemilihan sampel secara acak dalam arti yang semurni-murninyalah yang merupakan dasar bagi prosedur-prosedur inferensial, Anda mungkin terkejut bila melihat sampel-sampel yang digunakan pada kebanyakan riset yang dilaporkan dalam berbagai kepustakaan ilmiah. Alih-alih sampel acak yang dipilih dengan bantuan table-tabel bilangan teracak, sebagaimana dijelaskan dalam buku-buku teks statistika elementer, Anda akan menemukan sampel-sampel seperti "pasien-pasien yang berobat selama tiga bulan pertama pada suatu tahun," atau "semua murid kelas satu di Sekolah Dasar X," atau "para sukarelawan yang berbadan sehat." Sampel-sampel semacam itu dipergunakan orang semata-mata karena telah tersedia dan mudah diperoleh. Lalu, bagaimanakah kita bisa merasionalkan penggunaan sampel-sampel tersebut untuk membuat inferensi (kesimpulan)?

Dunn (1, hlm. 12) dan Remington serta Schork (2, hlm. 93) menyarankan agar kita meneliti sifat dasar populasi yang sampel-sampel sedemikiannya dapat dianggap acak. Sebagai contoh, apabila suatu sampel terdiri atas semua murid kelas satu di salah sebuah sekolah dasar, di pinggiran kota, mungkin saja sampel itu dianggap sebagai sampel acak dari populasi berupa semua murid kelas satu yang belajar di sekolah-sekolah sejenis di kawasan serupa. Armitage (3, hlm. 99, 100) mengemukakan rasionalisasi yang sedikit berbeda atas penggunaan *samples of convenience* ini. Colton (4, hlm. 4 – 7) menyoroti masalah yang sama ketika membahas perbedaan antara *populasi target* (populasi yang akan dibuatkan kesimpulannya) dan populasi yang disampelkan (populasi yang anggotanya betul-betul dijadikan sampel).

Statistik *Statistik* (catatan: bukan 'statistika') didefinisikan sebagai ukuran deskriptif (semacam informasi ringkas) yang dihitung dari data sampel. Statistik-statistik yang tidak asing bagi Anda yang telah mengenal statistika adalah rata-rata sampel *(sample mean)* $\bar{x}$, varians sampel *(sample variance)* s^2, dan koefisien korelasi sampel *(sample correlation coefficient)* r.

Parameter *Parameter* didefinisikan sebagai ukuran yang digunakan untuk menggambarkan suatu populasi. Contoh-contohnya antara lain adalah rata-rata populasi *(population mean)* μ, varians populasi *(population variance)* σ^2, dan koefisien korelasi populasi *(population correlation coefficient)* p. Parameter biasanya tidak diketahui; dan dengan statistiklah harga-harga parameter itu kita taksir atau kita estimasi. Sebagai contoh, kita boleh menggunakan rata-rata sampel $\bar{x}$ untuk menaksir rata-rata populasi μ yang tidak diketahui dari populasi yang sampelnya kita ambil.

Dalam analisis statistik nonparametrik, parameter yang cukup menarik perhatian adalah *median* populasi. Parameter ini, dalam analisis nonparametrik, sering menggantikan rata-rata populasi sebagai ukuran untuk lokasi atau tendensi sentral yang lebih disukai.

Variabel acak Kita biasanya mengandaikan bahwa data numerik yang akan kita analisis secara statistik adalah hasil-hasil prosedur penyampelan acak *(random sampling*) atau eksperimen acak. Himpunan hasil-hasil sedemikian disebut *variabel acak (random variable).* Dalam proses penyampelan atau eksperimen, kita mengamati satu, atau lebih dari satu, harga variabel acak itu. Sebagai contoh, waktu yang diperlukan oleh subyek dewasa untuk bereaksi terhadap suatu rangsangan *(stimulus)* adalah variabel acak. Apabila kita memberikan rangsangan tadi kepada seorang dewasa yang dipilih secara acak dan mendapatkan waktu reaksi 0.15 detik, maka 0.15 adalah harga variabel acak ini.

Variabel kontinyu *(continuous variable)* Suatu variabel acak disebut *kontinyu* bila harga-harga yang dapat diasumsikannya terdiri atas semua bilangan sejati (bilangan rill/*real numbers*) di suatu interval. Waktu reaksi terhadap suatu rangsangan adalah salah satu contoh variabel kontinyu.

Variabel diskret *(discrete variable)* Bila suatu variabel acak hanya dapat mengasumsikan harga-harga dengan jumlah yang terhingga *(finite)* atau tak terhingga namun masih dapat dihitung *(countable)*, maka variabel acak ini disebut variabel diskret. Dengan kata lain, banyaknya harga-harga itu bisa terhingga, bisa pula tak terhingga, namun dapat dihitung. Jumlah pasien yang berobat ke sebuah rumah sakit dalam suatu kurun waktu tertentu adalah salah satu contoh variabel acak diskret.

BAB 2
PENGUJIAN HIPOTESIS

Dalam buku ini kita akan berhubungan dengan dua jenis inferensi statistik: yaitu *pengujian hipotesis (hypothesis testing)* dan *estimasi interval* atau penaksiran/pendugaan interval *(interval estimation).* Kita akan membahas pengujian hipotesis dalam bagian ini, dan estimasi interval dalam bagian berikutnya.

Kita boleh mendefinisikan hipotesis sebagai pernyataan mengenai satu atau beberapa populasi. Secara umum kita juga boleh membedakan hipotesis atas: *hipotesis riset* dan *hipotesis statistik*. Hipotesis riset adalah hipotesis yang dirumuskan oleh seorang peneliti ahli (*sample surveyor* atau *experimenter*) yang biasanya bukan seorang ahli statistika. Oleh sebab itu hipotesis riset sering merupakan hasil firasat atau kecurigaan yang didasarkan atas pengamatan secara cermat serta lama oleh si peneliti ahli yang bersangkutan.

Sebagai contoh, seorang guru mungkin menaruh suatu kecurigaan, berdasarkan pengalaman mengajarnya selama bertahun-tahun, bahwa kondisi-kondisi fisik tertentu di ruangan kelas menghambat proses belajar. Seorang dokter yang mengamati bahwa beberapa pasiennya menjadi sesak napas setelah meminum suatu obat tertentu mungkin menaruh kecurigaan bahwa obat tersebut memiliki dampak sampingan yang merugikan, sekurang-kurangnya bagi beberapa pasien. Kecurigaan sedemikian mengantar ke pembentukan hipotesis riset seperti “Murid-murid kelas tiga memperoleh nilai tinggi pada ulangan-ulangan matematika bila temperatur ruangan selama jam-jam pelajaran tidak lebih dari 20°C,” dan “Sesak napas setelah pemberian obat A lebih sering terjadi pada pasien-pasien dengan tekanan darah tinggi ketimbang pada yang bertekanan darah normal atau rendah.”

Kita mengenal dua hipotesis statistik, yakni: *hipotesis nol (null hypothesis*; yang kita beri notasi H_0) dan *hipotesis tandingan (alternative hypothesis*; yang kita beri notasi H_1).

Hipotesis nol adalah hipotesis yang kita uji. Prosedur pengujiannya, yang berlandaskan informasi berasal dari data sampel yang tepat, menghasilkan salah satu dari dua keputusan statistik sebagai berikut: (1) keputusan untuk menolak hipotesis nol (karena dianggap salah) atau (2) keputusan untuk *tidak* menolak hipotesis nol karena sampel tidak memiliki bukti yang cukup untuk membenarkan penolakannya. Bila kita menolak hipotesis nol, berarti kita menerima bahwa hipotesis tandingannya benar. Kita dapat berbuat begitu karena hipotesis nol dan hipotesis tandingan dinyatakan sedemikian rupa sehingga keduanya sering berdiri sendiri dan melengkapi. Biasanya-namun tidak selalu-hipotesis tandingan sama dengan hipotesis riset. Jadi dalam banyak hal, hipotesis tandingan adalah pernyataan tentang sesuatu yang hamper kita yakini kebenarannya.

Apabila kita menolak hipotesis nol, kita menerima hipotesis tandingan dengan keyakinan yang lebih besar dibanding bila kita "menerima" hipotesis nol karena kita tidak mampu menolaknya. Bukti yang menunjang kebenaran suatu hipotesis tidaklah semeyakinkan bukti yang menjatuhkan hipotesis itu.

Uji hipotesis bisa *dua sisi (two-sided/two-tailed/nondirectional;*tanpa arah/dwi arah), bisa pula *satu sisi (one-sided/one-tailed/directional*;searah/eka arah). Yang berikut ini adalah contoh pernyataan hipotesis nol dan hipotesis tandingannya bila parameter-parameter yang ingin kita ketahui adalah rata-rata populasi μ_1 untuk populasi 1, dan rata-rata populasi μ_2 untuk populasi 2, dengan pengujian yang bersifat dua sisi:

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

Di sini hipotesis nol menyatakan bahwa rata-rata kedua populasi itu sama, sedangkan hipotesis tandingannya menyatakan bahwa rata-rata keduanya tidak sama. Dalam hal ini seorang peneliti bisa bertanya, "Dapatkah saya menyimpulkan bahwa kedua populasi itu memiliki rata-rata yang berbeda?" Peneliti itu mungkin merasa bahwa pertanyaannya akan lebih berarti bila berbunyi sebagai berikut, “Dapatkah saya menyimpulkan bahwa populasi 1 memiliki rata-rata yang lebih besar ketimbang populasi 2?” Dalam hal ini, si peneliti melakukan suatu uji satu sisi, dan dengan demikian hipotesis nol serta hipotesis tandingannya adalah

$$H_0 : \mu_1 \leq \mu_2, \quad H_1 : \mu_1 > \mu_2$$

Peneliti boleh pula mengajukan pertanyaan yang mengarah ke uji satu sisi atau uji eka arah sedemikian rupa sehingga hipotesis-hipotesis statistiknya adalah

$$H_0 : \mu_1 \geq \mu_2, \quad H_1 : \mu_1 < \mu_2$$

Guna menguji hipotesis itu, peneliti memiliki *statistik uji (test statistic)* yang paling tepat dan menetapkan distribusinya bila H_0 benar. Sebagai contoh, bila hipotesis itu bersangkut-paut dengan selisih antara dua rata-rata populasi, statistik uji yang digunakan biasanya adalah

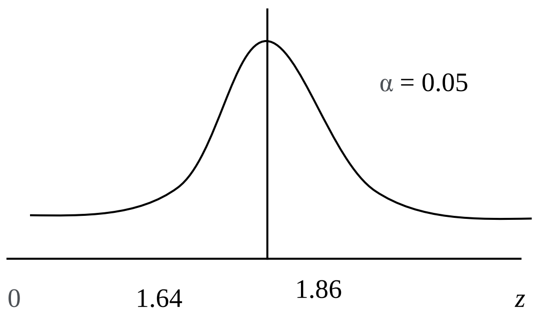
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{(s_p^2/n_1) + (s_p^2/n_2)}}$$

Di sini $\bar{x}_1$ dan $\bar{x}_2$ adalah nilai-nilai rata-rata sampel yang dihitung dari sampel-sampel berukuran n_1 dan n_2, yang berturut-turut ditarik dari populasi 1 dan populasi 2, D_0 adalah selisih hipotesis *(hypothesized difference)* antara nilai-nilai rata-rata populasi, dan s_p^2 diperoleh melalui penggabungan *(pooling)* kedua varians sampel. Apabila asumsi-asumsi yang telah ditentukan terpenuhi dan H_0 benar, maka t memiliki distribusi t-Student dengan derajat bebas *(degree of freedom)* $n_1 + n_2 - 2$.

Dari data sampel yang teramati, kita dapat menghitung harga statistik ujinya dan bertanya kepada diri sendiri, "Apakah nilai ini luar biasa ekstrem (entah sangat besar atau sangat kecil) jika H_0 benar?" Dengan kata lain, kita ingin tahu apakah besar nilai statistik uji hasil perhitungan cukup ekstrem sehingga hipotesis nolnya pantas kita tolak. Sebelum memeriksa data sampel, banyak investigator yang merumuskan kaidah pengambilan keputusan terlebih dahulu. Kaidah ini mengatakan, sesuai dengan hukum sebab-akibat, bahwa mereka akan menolak H_0 bila probabilitas untuk mendapatkan suatu harga statistik uji yang besarnya tertentu atau lebih ekstrem-bila H_0 benar-sama dengan atau kurang dari suatu bilangan kecil α. Kebanyakan penulis buku statistika dasar menyatakan bahwa α adalah *taraf nyata (level of significance).* Ada pula penulis yang menyebut α *ukuran uji (size of the test)*; sebagai contoh, Mood, Graybill, dan Boes (5) menggunakan istilah ini. Bila orang menggunakan pendekatan dengan kaidah pengambilan keputusan, mereka biasanya memilih α sebesar 0.05 atai 0.01, atau kadang-kadang 0.10.

Nilai kritis (critical value) suatu statistik uji adalah nilai yang begitu ekstrem sehingga probabilitas untuk mendapatkan nilai tersebut atau yang lebih ekstrem, bila H_0 benar, sama dengan α. Dengan demikian , kita boleh menyatakan kaidah pengambilan keputusan *(decision rule)* menurut nilai-nilai kritis. Dalam suatu uji satu sisi atau uji eka arah, umpamanya, kaidah pengambilan keputusan menyuruh kita menolak H_0 jika nilai statistik uji hasil perhitungan lebih ekstrem (entah lebih besar atau lebih kecil, bergantung pada arah hipotesis tandingannya) daripada nilai kritisnya. Dalam uji dua sisi atau uji dwi arah kita menghadapi dua nilai kritis. H_0 kita tolak bila nilai statistik uji hasil perhitungan lebih besar daripada nilai kritis yang besar atau lebih kecil daripada nilai kritis yang kecil.

Gambar 2.1 Harga kritis statistik uji *z* (1.645) dan harga hasil perhitungannya (1.86) untuk suatu uji hipotesis satu-sisi



Penggunaan kaidah pengambilan keputusan untuk pengujian hipotesis dapat kita jelaskan secara grafis. Andaikanlah bahwa hipotesis-hipotesis yang ingin kita uji adalah

$$H_0 : \mu_1 \leq \mu_2, \quad H_1 : \mu_1 > \mu_2$$

Andaikan pula bahwa taraf nyata α = 0.05, bahwa statistik uji memiliki distribusi normal standar, dan bahwa nilai z hasil perhitungan adalah 1.86. Apabila kita mengacu ke Tabel A.2 (baca: Tabel Statistik), kita melihat bahwa nilai kritis statistik uji untuk α = 0.05 dengan uji satu sisi adalah 1.645. Gambar 2.1 memperlihatkan distribusi z, nilai kritisnya, serta nilai hasil perhitungannya. Karena 1.86 lebih besar dari 1.645, maka H_0 kita tolak.

Nilai-nilai *P* Cara lain untuk memutuskan apakah data sampel menyatakan kecurigaan terhadap hipotesis nol adalah dengan menentukan peluang mengamati suatu nilai statistik uji, bila H_0 benar, yang setidak-tidaknya sama ekstrem dengan nilai yang sungguh-sungguh merupakan hasil pengamatan. Peluang atau probabilitas ini dikenal dengan berbagai sebutan, antara lain: *critical value, descriptive level of significance, prob value,* dan *associated probability.* Di sini kita akan menggunakan istilah *nilai P* untuk probabilitas, sesuai dengan teladan yang diberikan oleh Gibbons dan Pratt (6) dalam artikel mereka mengenai interpretasi dan metodologi nilai-nilai *P*. Hodges dan Lehmann (7, hlm. 317) mengajak kita menganggap nilai *P* (yang mereka sebut peluang nyata atau *significance probability*) "sebagai suatu ukuran, dalam bentuk bilangan yang bersahaja, atas kadar keheranan yang dialami oleh seseorang yang mempercayai hipotesis nol pada suatu eksperimen."

Kebanyakan penulis di kepustakaan ilmiah mengetengahkan nilai-nilai *P* yang dinyatakan dengan cara sebagai berikut $p > 0.05$, $p < 0.01$, $0.01 < p < 0.05$, atau $p = 0.0618$, dengan p adalah nilai *P*. Dengan demikian suatu nilai *P* dapat dinyatakan baik sebagai suatu nilai yang eksak maupun sebagai suatu interval, bergantung pada karakteristik tabel distribusi statistik uji yang tersedia. Banyak tabel statistik yang telah dibuat (atau diikhtisarkan untuk dilampirkan ke dalam buku-buku teks statistik) sedemikian rupa sehingga lebih sesuai bagi para peneliti yang menggunakan kaidah pengambilan keputusan dan nilai-nilai α yang telah dipilih terlebih dahulu. Bila tabel-tabel semacam itu hendak digunakan untuk menentukan nilai *P* suatu pengujian, alih-alih mendapatkan nilai yang eksak, peneliti yang bersangkutan biasanya harus puas dengan interval yang diperolehnya. Dalam contoh berikut, kita dapat menemukan nilai *P* yang eksak.

Contoh 2.1

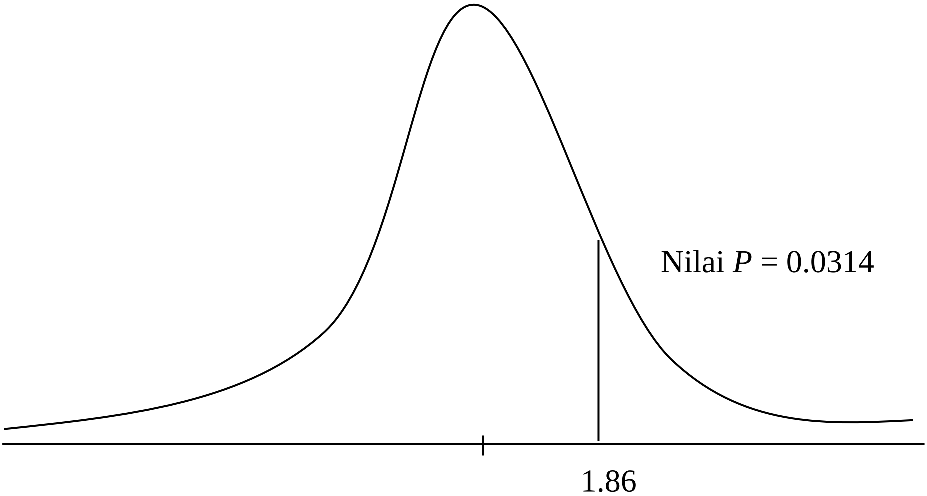
Andaikan hipotesis nol dan hipotesis tandingannya adalah

$$H_0 : \mu_1 \leq \mu_2, \quad H_1 : \mu_1 > \mu_2$$

Selanjutnya andaikan pula bahwa statistik uji yang tepat adalah z dan bahwa berdasarkan perhitungan nilai z = 1.86. Kita mengacu ke Tabel A.2 (baca: Tabel Statistik) dan melihat bahwa luas daerah di sebelah kanan z = 1.86 adalah 0.5 – 0.4686 = 0.0314. Jadi, peluang atau probabilitas untuk memperoleh suatu nilai z yang sebesar atau lebih besar daripada 1.86, bila H_0 benar, adalah 0.0314. Karena itu nilai P di sini adalah 0.0314. Keadaan ini diterangkan dalam Gambar 2.2.

Sekarang perhatikan sebuah contoh yang mengharuskan kita menyajikan nilai P sebagai suatu interval karena kemampuan tabel yang tersedia terbatas.

Gambar 2.2 Harga statistik uji *z* hasil perhitungan (1.86) dan harga *P*-nya untuk suatu uji hipotesis satu-sisi

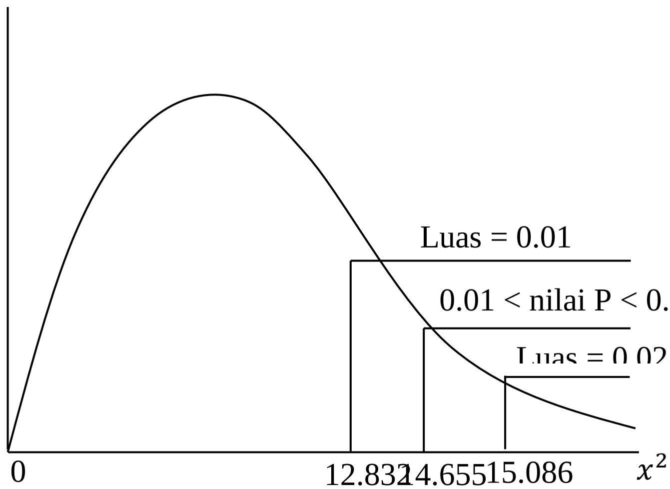


Contoh 2.2

Andaikan bahwa, dalam eksperimen kita, statistik ujinya mengikuti suatu distribusi kai-kuadrat dengan derajat bebas lima, dan bahwa nilai statistik uji hasil perhitungan adalah 14.665. Untuk nilai-nilai statistik uji yang cukup besar, kita akan menolak H_0. Apabila kita menengok ke Tabel A.12 (baca: Tabel Statistik), di baris derajat bebas lima, kita menjumpai bahwa 14.665 terletak di antara $X^2_{0.975}$ = 12.832 dan $X^2_{0.99}$ = 15.086. Dengan demikian nilai *P* hanya dapat kita nyatakan sebagai 0.01 < nilai *P* < 0.025.

Umpama data sampel menghasilkan nilai statistik uji sebesar 17.335. Karena 17.335 lebih besar daripada $X^2_{0.975}$ = 16.750, kita bisa membuat pernyataan begini: nilai P < 0.005. Contoh ini dijelaskan dalam Gambar 2.3.

Gambar 2.3 Harga statistik uji kai-kuadrat hasil perhitungan (14.655) dan harga *P*-nya untuk suatu uji hipotesis



Menentukan nilai-nilai *P* dalam uji-uji dua-sisi atau dwi arah bisa menimbulkan masalah. Sebagaimana yang ditunjukkan oleh Gibbons dan Pratt (6), yang paling lazim dilakukan dalam uji dua-sisi, nilai *P*-nya dinyatakan sebagai kelipatan dua dari nilai P satusisinya. Dalam Contoh 2.1, misalnya, andaikanlah bahwa hipotesis-hipotesisnya adalah

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

Nilai z hasil perhitungan sebesar 1.86 akan memberikan suatu nilai P yang besarnya 2(0.0314) = 0.0628. Prosedur ini memuaskan untuk kasus-kasus dengan distribusi sampling statistik uji yang simetrik (setangkup) bila hipotesis nol benar. Contoh distribusi yang demikian adalah distribusi normal standard an distribusi t-Student.

Bagaimanapun, jika distribusi statistik uji untuk H_0 tidak simetrik, penggandaan nilai P satu-sisi guna mendapatkan nilai P dua-sisi bisa menghasilkan suatu nilai P yang lebih besar dari 1, atau hal-hal lain yang mustahil. Gibbons dan Pratt (6), yang membahas beberapa pilihan prosedur dalam kasus dua-sisi, menyukai cara menyatakan nilai P untuk masing-masing arah atau sisi berikut pernyataan tentang arah keberangkatan untuk hipotesis nolnya. Misalkan kerangka ini kita ikuti, dengan hipotesis-hipotesis yang dihadapi adalah

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

Dan statistik uji yang paling sesuai adalah z. Jika harga z hasil perhitungan adalah 1.86, kita boleh meringkaskan hasilnya berupa suatu nilai P dalam salah satu dari beberapa cara berikut (cara-cara yang lain juga masih ada).

1 $P\ (z \geq 1.86 \mid H_0) = 0.0314$
2 $P\ (z \geq 1.86 \mid \mu_1 = \mu_2) = 0.0314$
3 Peluang atau probabilitas untuk mengamati suatu nilai statistik uji *sebesar* atau *lebih besar dari* 1.86 bila H_0 benar sama dengan 0.0314.

Jika kita memperoleh hasil perhitungan z = -1.86 dari data sampel (dan menggunakannya sebagai pengganti 1.86 dalam 1, 2, dan 3), tanda ketidaksamaan dalam 1 dan 2 harus dibalik, dan kata-kata “besar” serta “lebih besar” dalam 3, yang menunjukkan arah keberangkatan menurut H_0, harus digantikan dengan kata-kata “kecil” serta “lebih kecil”.

Untuk nilai hasil perhitungan +1.86 dalam contoh ini, kita dapat menyatakan bahwa “nilai P = 0.0314 menghendaki nilai μ_1 yang lebih besar.” Untuk z = -1.86, kita dapat menyatakan bahwa “nilai P = 0.0314 menghendaki nilai μ_2 yang lebih besar.”

Suatu nilai P memberi kita informasi yang lebih baik ketimbang pernyataan-pernyataan seperti “perbedaan (*difference*) itu nyata pada taraf 0.05” atau “H_0 dapat ditolak pada taraf 0.01.” Ini merupakan alasan utama untuk menerima penggunaan nilai-nilai P dalam pengungkapan hasil-hasil riset. Bila menerima suatu nilai P, Anda dapat memilih taraf nyata Anda sendiri, atau taraf pada saat Anda ingin melepaskan keyakinan bahwa H_0 benar dan mulai menerima keyakinan bahwa H_1 benar.

Beberapa pengarang membedakan *pengujian hipotesis* dari *pengujian kebermaknaan (significance testing)*. Mereka menggunakan istilah "pengujian hipotesis" bila mereka menerapkan kaidah pengambilan keputusan berdasarkan suatu nilai α yang dipilih sejak awal, dan "pengujian kebermaknaan" bila mereka menyatakan nilai P. Pembahasan lebih lanjut tentang perbedaan ini dapat dijumpai pada Kempthorne dan Folks (8) dan Lindgren (9).

Karena dalam statistika yang Anda kenal sebelum ini, untuk pengujian hipotesis, Anda mungkin telah terbiasa dengan pendekatan kaidah keputusan (dengan α yang telah dipilih terlebih dahulu) ketimbang pendekatan menggunakan perhitungan nilai-nilai P, kami menggunakan kedua pendekatan tersebut dalam contoh-contoh dan latihan-latihan pada bagian awal buku ini. Setelah itu kita akan meninggalkan penggunaan metode yang pertama guna membiasakan diri dengan pendekatan menggunakan nilai P. Dengan mengikuti kedua pendekatan itu, kami berharap bahwa Anda akan memperoleh pemahaman yang lebih baik tentang nilai-nilai P (yang mungkin merupakan konsep baru bagi Anda) serta pemahaman yang lebih baik tentang hubungan antara kedua pendekatan tersebut. Dalam contoh-contoh dan latihan-latihan yang hanya menghitung nilai-nilai *P*, Anda boleh mempraktekkan pendekatan menggunakan taraf nyata Anda sendiri untuk mengevaluasi hasil-hasilnya.

Beberapa tanggapan yang dikemukakan oleh Bahn (10) mengenai nilai-nilai *P* mungkin cukup menarik bagi Anda.

Kebermaknaan statistik versus kebermaknaan praktis

Di samping kebermaknaan statistik (*statistical significance*), konsep penting lain yang juga timbul ketika kita mencoba mengevaluasi hasil-hasil riset adalah *kebermaknaan praktis* atau (*kebermaknaan substantif*. Sayang sekali, istilah kebermaknaan statistik acap kali digunakan untuk pengertian yang maksudnya adalah kebermaknaan praktis. Apabila analisis secara statistika terhadap temuan-temuan riset menyingkapkan hasil-hasil yang bermakna secara statistika, Anda tidak boleh serta-merta menganggap bahwa penemuan itu harus memiliki suatu kebermaknaan praktis. Andaikan kita berminat mengetahui apakah dua buah nilai rata-rata populasi sama besar. Sampel-sampel yang cukup besar akan menampakkan suatu perbedaan, betapapun kecilnya, namun demikian mungkin hanya perbedaan yang agak besar yang ada manfaatnya. Demikian pula, sampel-sampel kecil mungkin tidak mampu mengungkapkan perbedaan-perbedaan (atau hubungan-hubungan) antara kedua populasi yang bermakna secara praktis.

Karena pengertian kebermaknaan statistik dan kebermaknaan praktis berbeda, maka Anda sebaiknya berhati-hati dalam menggunakan istilah-istilah tersebut. Anda harus menggunakan istilah “nyata” (*significant*), bila yang Anda maksudkan adalah kebermaknaan statistik, yaitu untuk mengacu ke hasil-hasil sampel. Jadi, sebagai contoh, Anda boleh mengatakan, “nilai-nilai rata-rata sampel berbeda nyata,” bila Anda memaksudkan bahwa beda yang teramati menghasilkan suatu nilai *P* yang cukup kecil sehingga menyebabkan Anda menolak hipotesis nol yang menyatakan tidak adanya perbedaan antara nilai-nilai rata-rata populasi. Anda harus menghindari ungkapan-ungkapan seperti “nilai-nilai rata-rata populasi berbeda nyata” sebab, jika yang Anda maksudkan adalah kebermaknaan statistik, terminologi yang digunakan tidak tepat; sedangkan jika yang Anda maksudkan adalah kebermaknaan praktis, Anda sebaiknya menggunakan istilah lain, selain “nyata” (*significant*) agar tidak membingungkan. Tentu saja, hindari pula ungkapan-ungkapan seperti ”dalam hipotesis nol dinyatakan bahwa nilai-nilai rata-rata kedua populasi berbeda nyata,” sebab uji-uji hipotesis secara statistika tidak dapat menentukan hal-hal yang menyangkut kebermaknaan praktis. Hanya orang yang berpengetahuan sangat mendalam di bidang yang diteliti itulah yang berhak memutuskannya (suatu kriterium yang ahli statistik pun sering tidak mampu memenuhinya).

Kerancuan di sekitar pengertian kebermaknaan statistik dan kebermaknaan praktis telah dibahas oleh Bakan (11), Brewer (12), Cohen (13), Duggan dan Dean (14), Gold (15), Kish (16), dan McGinnis (17).

Kuasa suatu uji hipotesis

Kuasa (*power*) suatu uji hipotesis adalah peluang atau probabilitas untuk menolak hipotesis nol bila hipotesis itu salah. Kuasa boleh juga didefinisikan sebagai 1 – *ß*, dengan *ß* adalah peluang untuk menerima suatu hipotesis nol yang salah. Anda mungkin ingat bahwa menerima suatu hipotesis nol yang salah disebut kesalahan tipe II (*type II error*), dan bahwa menolak suatu hipotesis nol yang benar adalah kesalahan tipe I (*type I error*). Peluang terjadinya kesalahan tipe I biasanya dinyatakan dengan α. Pada umumnya, kita menghendaki suatu uji yang tinggi kuasanya. Untuk kebanyakan uji yang kita bahas dalam bab-bab selanjutnya, kita akan mengomentari apa yang kita ketahui tentang kuasa ujinya masing-masing.

Efisiensi suatu uji hipotesis

Sebuah kriterium lain untuk mengevaluasi unjuk kerja (*performance*) suatu uji adalah *efisiensi*. Patokan yang paling sering digunakan untuk mengukur efisiensi suatu uji nonparametrik adalah *efisiensi relatif asimptotik*-nya (*asymptotic relative efficiency/ARE*). Karena konsep efisiensi relatif asimptotik dipopulerkan oleh Pitman (18), maka efisiensi ini sering disebut *efisiensi Pitman*. Meskipun pembahasan konsep ARE secara mendalam membutuhkan penguasaan matematika tingkat yang lebih lanjut daripada yang dibutuhkan dalam buku ini, kita dapat mengatakan bahwa efisiensi yang tinggi adalah karakteristik yang harus dimiliki oleh suatu uji.

Dalam berbagai situasi praktis; ARE suatu uji merupakan aproksimasi yang baik terhadap *efisiensi relatif*-nya. Efisiensi relatif suatu uji A terhadap uji B (untuk H_0, H_1 , *α,* dan *ß* yang sama) adalah perbandingan n_B/n_A, dengan n_A adalah ukuran sampel uji A dan n_B adalah ukuran sampel uji B. Jika n_A lebih kecil daripada n_B, efisiensi uji A relatif terhadap uji B lebih besar dari 1, dan kita mengatakan bahwa uji A lebih efisien dibanding uji B. Kita lebih menyukai uji yang membutuhkan ukuran sampel lebih kecil apabila kondisi-kondisinya sama, karena makin kecil ukuran sampel-

sampel yang digunakan, umumnya makin kecil pula biaya, waktu serta kebutuhan lain yang diperlukan.

Dalam bab-bab mendatang, kita akan mengulas efisiensi pada hampir semua uji yang kita bahas. Jika Anda berminat mempelajari rincian-rincian matematika tentang ARE, Anda dapat mengacu ke artikel-artikel yang ditulis oleh Noether (19) dan Stuart (20, 21) serta buku-buku karangan Lehmann (22) dan Fraser (23). Untuk definisi-definisi lain tentang efisiensi relatif, lihat Blyth (24). Artikel yang ditulis oleh Smith (25) pun agaknya cukup menarik.

Kepustakaan mengenai berbagai segi yang berkaitan dengan inferensi statistik cukup banyak. Yang berikut ini mungkin berharga atau menarik bagi Anda. Wilson dan Miller (26), misalnya, membahas "mengapa keputusan menerima hipotesis nol tidak meyakinkan (*inconclusive*)." Edwards (27) mengkaji hubungan antara hipotesis ilmiah dan hipotesis statistik. Feinberg (28) dan Rodger (29, 30) menulis tentang kesalahan-kesalahan tipe I dan tipe II. Edgington (31) membahas sampel-sampel tidak acak (*nonrandom*) dalam inferensi statistik, dan Ungerleider serta Smith (32) mengulas penggunaan statistika secara salah.

Artikel oleh Brewer dan Knowles (33) berisi pembahasan sederhana mengenai kuasa (*power*). Sejumlah ulasan tambahan tentang kebermaknaan statistik antara lain terdapat dalam artikel-artikel oleh Ahrens (34), Barnard (35), Chandler (36), Labovitz (37), Lykken (38), Krause (39), Morrison dan Henkel (40), O'Brien dan Shapiro (41), Rozeboom (42), Selvin (43), Skipper (44), Stone (45), Winch dan Campbell (46), Zeisel (47), serta pada suatu editorial dalam *New England Journal of Medicine* (48). Bross (49), Godambe dan Sprott (50), Kiefer (51), dan Lurie (52) terutama membahas pandangan-pandangan umum yang berkaitan dengan inferensi statistik.

BAB 3
PENDUGAAN (ESTIMASI)

Dalam banyak hal, para peneliti mungkin ingin membuat keputusan yang berkaitan dengan nilai numerik suatu parameter populasi sebagai ganti (atau selain) keingintahuan tentang apakah mereka dapat menolak hipotesis nol yang menyatakan bahwa nilai parameter itu sama dengan beberapa nilai tertentu. Guna mendapatkan keputusan tentang besar (*magnitude*) nilai-nilai parameter populasi berdasarkan data sampel, kita menggunakan proses yang disebut pendugaan (*estimation*).

Kita mengenal dua macam pendugaan, yakni: *pendugaan titik (point estimation*) dan *pendugaan interval (interval estimation).* Dalam pendugaan titik, kita menghitung suatu nilai tunggal-yang disebut *dugaan (estimate)*-dari data sampel dan mengajukannya sebagai calon untuk parameter yang ingin kita duga. Dalam kebanyakan kasus, orang lebih menyukai *dugaan interval (interval estimate)* dan menganggapnya lebih bermanfaat. Suatu dugaan interval antara lain terdiri atas dua nilai yang mungkin merupakan parameter yang sedang diduga-yaitu sebuah nilai bawah dan sebuah nilai atas. Kedua nilai ini membatasi suatu interval (selang) yang memungkinkan kita mengekspresikan suatu derajat kepercayaan *(degree of confidence)* bahwa interval itu memuat parameter yang kita duga. Oleh sebab itu, dugaan interval sering disebut *interval kepercayaan* atau interval keyakinan *(confidence interval)*.

Kita mengekspresikan derajat kepercayaan kita terhadap suatu interval kepercayaan melalui penggunaan *koefisien kepercayaan (confidence coefficient),* baik yang berupa suatu bilangan antara 0 dan 1 maupun suatu persentase. Jika kita menggunakan koefisien kepercayaan 0.95 atau 95%, misalnya, itu mengandung arti bahwa kita 95% percaya bahwa interval yang kita maksudkan mengandung parameter yang kita duga.

Kita membuat interval-interval sedemikian rupa sehingga kita boleh menginterpretasikan interval-interval itu dalam dua cara. Salah satu di antaranya adalah *interpretasi probabilistik (probabilistic interpretation)* yang berlandaskan kenyataan bahwa dalam penyampelan yang berulang-ulang, 100 $(1 - \alpha)$% dari interval-interval yang dibuat dengan cara serupa (dan dengan ukuran sampel yang sama) berisi parameter yang sedang diduga. Interpretasi ini berlaku untuk semua interval kepercayaan yang kita buat. Dalam praktek, kita hanya membuat sebuah interval tunggal, dan terhadap interval tunggal inilah kita menerapkan interpretasi yang lain, yaitu *interpretasi praktis (practical interpretation)*. Dalam mengekspresikan interpretasi praktis, kita mengatakan bahwa kita 100 (1 – α)% percaya bahwa interval tunggal yang kita buat mengandung parameter yang sedang kita duga. Dalam kasus interpretasi itu, koefisien kepercayaan yang digunakan adalah 100 (1 – α)%.

Kita boleh mengekspresikan suatu interval kepercayaan untuk parameter ø secara probabilistik sebagai berikut

$$P(L_0 < ø < U_0) = 1 - \alpha$$

(3.1)

Dengan L_0 dan U_0 adalah variabel-variabel acak yang memenuhi pernyataan probabilitas itu. Segera setelah kita menetapkan nilai-nilai L_0 dan U_0, katakanlah L dan U, Ekspresi 1.1 bukan lagi merupakan suatu variabel untung-untungan *(change variable)*, melainkan suatu interval yang pasti. Inilah interval tunggal yang biasa dibuat dalam penerapan sesungguhnya. Kita dapat mengekspresikan interval tunggal tersebut secara ringkas sebagai berikut

$$C(L < ø < U) = 1 - \alpha$$

(3.2)

Dengan C adalah kependekan dari *confidence* (kepercayaan) dan menunjukkan bahwa ekspresi ini lebih merupakan suatu pernyataan kepercayaan ketimbang suatu pernyataan probabilitas.

Sebagaimana yang mungkin Anda perkirakan, pendugaan interval dan pengujian hipotesis saling berkaitan. Contoh perhatikan hipotesis-hipotesis

$$H_0 : \o = \o_0, \quad H_1 : \o \neq \o_0$$

dengan taraf nyata α. Nilai-nilai parameter $\o$ yang mungkin, yang terdapat dalam 100 (1 – α)% dari interval kepercayaan $L < \o < U$ adalah nilai-nilai yang berkesesuaian dengan hipotesis nol. Nilai-nilai $\o$ yang mungkin, yang berada di luar interval itu tidak berkesesuaian dengan hipotesis nol. Dengan demikian kita boleh menguji H_0 menggunakan interval kepercayaan. Jika 100 (1 – α)% dari interval kepercayaan tidak mengandung nilai parameter $\o_0$ yang dihipotesiskan, kita menolak H_0 pada taraf nyata α. Jika $\o_0$ terdapat dalam interval yang bersangkutan, kita tidak menolak H_0 (pada taraf nyata α). Natrella (53) membahas hubungan antara interval-interval kepercayaan dan uji-uji nyata. Beberapa artikel yang dikutip dalam bagian ini, serta dalam bagian-bagian yang lain, terdapat dalam bab tentang pengujian hipotesis dan interval-interval kepercayaan dalam sebuah buku kumpulan makalah oleh Kirk (54).

BAB 4
SKALA-SKALA PENGUKURAN

Seorang peneliti yang menganalisis data numerik senantiasa berkepentingan dengan sifat dasar skala yang digunakan untuk pengukuran-pengukuran. Kebanyakan peneliti mengikuti pandangan-pandangan tentang pengukuran dan skala-skala pengukuran yang dikemukakan oleh Stevens (55, 56, 57). Stevens (55) mendefinisikan pengukuran sebagai pemberian angka-angka terhadap benda-benda atau peristiwa-peristiwa menurut kaidah-kaidah tertentu, dan menunjukkan bahwa kaidah-kaidah yang berbeda menghendaki skala-skala serta pengukuran-pengukuran yang berbeda pula. Stevens mendefinisikan empat macam skala pengukuran, yaitu: *nominal*, *ordinal*, *interval*, dan *ratio*.

Skala nominal Skala ini merupakan skala yang paling lemah di antara keempat skala pengukuran. Sesuai dengan nama atau sebutannya, skala nominal membedakan benda atau peristiwa yang satu dengan yang lainnya berdasarkan nama (predikat). Jika kita boleh mengklarifikasikan (menyebut) barang-barang yang dihasilkan pada suatu proses ban berjalan dengan predikat cacat atau tidak cacat. Bayi yang baru lahir bisa laki-laki, bisa pula perempuan. Pasien-pasien di rumah sakit jiwa bisa dibeda-bedakan atas penderita *schizophrenic, manic-depressive, psychoneurotic,* dan sebagainya.

Tidak jarang kita menggunakan nomor-nomor yang dipilih sekehendak hati, sebagai pengganti nama-nama atau sebutan-sebutan, untuk membedakan benda-benda atau peristiwa-peristiwa berdasarkan beberapa karakteristik tertentu. Sebagai contoh, kita boleh menggunakan nomor 1 untuk menyebut kelompok barang-barang yang cacat dari suatu proses produksi dan 0 untuk kelompok barang-barang yang tidak cacat. Skala nominal biasanya juga digunakan bila kita berminat terhadap jumlah benda atau peristiwa yang termasuk ke dalam masing-masing kategori nominal. Sebagai contoh, kita mungkin ingin mengetahui komposisi jumlah pasien di sebuah rumah sakit jiwa yang menurut diagnosis menderita *schizophrenic, manic-depressive, psychoneurotic,* dan sebagainya. Data semacam inisering disebut data hitung (count data) atau data frekuensi).

Skala ordinal Skala pengukuran berikutnya yang lebih presisi atau lebih canggih adalah *skala ordinal*. Kita membeda-bedakan benda atau peristiwa yang satu dengan yang lain yang diukur dengan skala ordinal berdasarkan jumlah relatif beberapa karakteristik tertentu yang dimiliki oleh masing-masing benda atau peristiwa.

Pengukuran ordinal memungkinkan segala sesuatu disusun menurut peringkatnya masing-masing. Tenaga penjualan, misalnya, bisa diperingkat dari yang “paling buruk” sampai yang “paling baik” berdasarkan kepribadian mereka. Para peserta kontes kecantikan dapat diperingkat dari yang paling kurang cantik sampai yang paling cantik. Penyakit dapat diperingkat dari yang paling ringan sampai yang paling berbahaya. Kalau kita bermaksud memeringkat *n* buah benda berdasarkan suatu ciri tertentu, kita boleh menetapkan nomor 1 untuk benda yang ciri tertentunya paling kurang, nomor 2 untuk benda yang ciri tertentunya kedua paling kurang, dan seterusnya hingga nomor *n*, untuk benda dengan kadar ciri tertentu yang paling tinggi. Sebagai contoh, para peserta lomba lari dapat diberi peringkat 1, 2, 3, …, berdasarkan urut-urutan waktu yang mereka perlukan untuk mencapai garis finis. Data semacam ini sering disebut *data peringkat (rank data)*.

Beda-beda antara peringkat yang satu dan yang berikutnya tidak perlu sama. Sebagai contoh, tiga orang mahasiswa yang mengikuti suatu ujian boleh diberi peringkat kesatu, kedua, dan ketiga berdasarkan urut-urutan waktu yang diperlukan untuk menyelesaikan ujian tersebut. Bagaimanapun, ini tidak berarti bahwa selisih waktu antara nomor 1 dan nomor 2 sama dengan antara nomor 2 dan nomor 3. Mahasiswa yang pertama, misalnya, mungkin menyelesaikan ujiannya lima menit lebih cepat dibanding mahasiswa kedua, sementara yang belakangan ini menyelesaikannya delapan menit lebih cepat dibanding mahasiswa ketiga. Bagaimanapun, kalau yang tersedia bagi kita untuk analisis hanya peringkat, kita tidak dapat mengetahui besar perbedaan antara pengukuran-pengukuran yang telah diperingkat itu.

Skala interval Apabila benda-benda atau peristiwa-pristiwa yang kita selidiki dapat dibeda-bedakan antara yang satu dan lainnya kemudian diurutkan, dan bilamana perbedaan-perbedaan antara peringkat yang satu dan lainnya mempunyai arti (yakni, bila satuan pengukurannya tetap), di sini *skala interval* dapat diterapkan. Skala interval yang benar memiliki sebuah titik nol, tetapi titk nol ini bisa dipilih secara sembarang. Contoh pengukuran interval yang tidak asing bagi kita adalah pengukuran temperatur dalam derajat Fahrenheit atau derajat Celsius (*centigrade*). Namun demikian, perlu dimaklumi bahwa titik nol baik pada termometer Fahrenheit maupun pada termometer Celsius *tidak* menunjukkan tidak adanya temperatur, yakni sesuatu yang kita ukur dengan alat tersebut.

Sebagai contoh, andaikanlah bahwa empat buah benda A, B, C, dan D secara berturut-turut kita beri nilai *(score)* 20, 30, 60, dan 70, melalui pengukuran menggunakan skala interval. Karena yang digunakan adalah skala interval, maka kita bisa mengatakan bahwa beda antara 20 dan 30 sama dengan beda antara 60 dan 70. Dengan demikian, jarak yang sama antara anggota-anggota masing-masing pasangan nilai itu menunjukkan beda yang sama dalam hal kadar ciri atau sifat yang kita ukur. Bagaimanapun, skala interval tidak menjadikan kita berhak berbicara tentang *ratio* antara dua buah nilai. Dalam contoh kita, kita tidak dapat mengatakan bahwa nilai 60 untuk C dan nilai 30 untuk B mengandung arti bahwa ciri atau sifat yang dimiliki oleh C dua kali lipat yang dimiliki oleh B.

Skala *ratio* Apabila pengukuran-pengukuran yang kita lakukan memiliki sifat-sifat yang terdapat pada ketiga skala yang pertama serta sifat tambahan bahwa *ratio* antara masing-masing pengukuran mempunyai arti, maka skala pengukuran di sini disebut *skala ratio*. Pengukuran-pengukuran dengan skala *ratio* yang tidak asing bagi kita misalnya adalah pengukuran tinggi dan pengukuran berat. Kita dapat mengatakan bahwa seseorang yang beratnya 90 kg memiliki kelebihan berat 30 kg dibanding yang beratnya 60 kg (sebagaimana yang dapat kita katakana bila menggunakan skala interval).

Dengan skala *ratio*, kita juga dapat mengatakan bahwa orang yang beratnya 90 kg dua kali lebih berat daripada orang yang beratnya 45 kg. Skala *ratio* merupakan aras pengukuran yang paling tinggi.

Stevens berpendapat bahwa prosedur-prosedur statistik yang sesuai untuk penggunaan dengan data empirik ditentukan oleh skala pengukuran yang dipakai ketika pengamatan. Banyak ahli statistika dan peneliti yang mengikuti pandangan ini bila mereka mengerjakan analisis-analisis statistik. Bagaimanapun, ada pula yang tidak sependapat dengan Stevens; khususnya mereka tidak setuju terhadap anggapan Stevens bahwa aras pengukuran menentukan sifat dasar operasi-operasi statistik yang diperbolehkan. Di antara mereka yang menentang itu adalah Anderson (58), Boneau (59), Gaito (60), dan Lord (61). Artikel-artikel lain yang cukup menarik ditulis oleh Baker dkk. (62), Campbell (63), serta Gardner (64).

BAB 5
STATISTIKA NONPARAMETRIK

Kuliah atau kursus statistika pengantar umumnya menekankan pembahasan pada *prosedur-prosedur statistik parametrik*. Anda tentu ingat bahwa prosedur-prosedur ini antara lain mencakup uji-uji yang berlandaskan distribusi *t*-Student, analisis varians, analisis korelasi, dan analisis regresi. Salah satu karakteristik prosedur-prosedur ini adalah bahwa kelayakan penggunaannya untuk maksud-maksud inferensi (penyimpulan) bergantung pada asumsi-asumsi tertentu. Prosedur-prosedur inferensial dalam analisis varians, misalnya, mengasumsikan bahwa sampel-sampel telah ditarik dari populasi-populasi berdistribusi normal dengan varians-varians yang sama.

Karena populasi-populasi yang kita kaji tidak selalu memenuhi asumsi-asumsi yang mendasari uji-uji parametrik, kita kerap kali membutuhkan prosedur-prosedur inferensial dengan kesahihan *(validity)* yang tidak bergantung pada asumsi-asumsi yang kaku. Dalam banyak hal, prosedur-prosedur statistik nonparametrik memenuhi kebutuhan ini karena tetap sahih meski hanya berlandaskan asumsi-asumsi yang sangat umum. Sebagaimana yang akan kita bahas secara lebih mendalam nanti, prosedur-prosedur nonparametrik juga memenuhi kebutuhan-kebutuhan lain para peneliti.

Berdasarkan kesepakatan, dua tipe utama prosedur statistik yang dianggap nonparametrik adalah: (1) prosedur-prosedur nonparametrik murni dan (2) prosedur-prosedur bebas-distribusi *(distribution-free procedures).* Secara singkat, dapat dikatakan bahwa prosedur-prosedur nonparametrik tidak berkepentingan dengan parameter-parameter populasi. Sebagai contoh, dalam buku ini kita akan membicarakan uji-uji keselarasan (uji kompatibilitas; *goodness of fit test*) dan uji keacakan (*test for randomness*) di mana kita berkepentingan dengan beberapa karakteristik yang bukan nilai parameter populasi. Kesahihan prosedur bebas-distribusi tidak bergantung pada bentuk fungsi populasi yang sampelnya telah kita ambil. Terutama di kalangan para penulis Amerika, kedua jenis prosedur itu telah biasa dianggap prosedur nonparametrik. Perbedaan antara istilah-istilah nonparametrik dan bebas-distribusi dibahas oleh Kendall dan Sundrum (65).

Keunggulan dan kelebihan statistika nonparametrik

Berikut ini dikemukakan beberapa keuntungan yang dapat diperoleh melalui penggunaan prosedur-prosedur statistik nonparametrik.

1. Karena kebanyakan prosedur nonparametrik memerlukan asumsi dalam jumlah yang minimum, maka kemungkinan untuk digunakan secara salah pun kecil.
2. Untuk beberapa prosedur nonparametrik, perhitungan-perhitungan dapat dilaksanakan dengan cepat dan mudah, terutama bila terpaksa dikerjakan secara manual. Jadi penggunaan prosedur-prosedur ini menghemat waktu yang diperlukan untuk perhitungan. Ini bisa dijadikan bahan pertimbangan yang penting bila hasil pengkajian harus segera tersaji atau bila mesin hitung berkemampuan tinggi tidak tersedia.
3. Para peneliti dengan dasar matematika serta statistika yang kurang biasanya menemukan bahwa konsep-konsep dan metode-metode prosedur nonparametrik mudah dipahami.
4. Prosedur-prosedur nonparametrik boleh diterapkan bila data telah diukur menggunakan skala pengukuran yang lemah, sebagaimana bila hanya data hitung atau data peringkat yang tersedia untuk analisis.

Kekurangan dan kelemahan statistika nonparametrik

Bagaimanapun, prosedur-prosedur nonparanetrik bukannya tanpa kelemahan. Berikut ini adalah beberapa kelemahan yang cukup menonjol.

1. Karena perhitungan-perhitungan yang dibutuhkan untuk kebanyakan prosedur nonparametrik cepat dan sederhana, prosedur-prosedur ini kadang-kadang digunakan untuk kasus-kasus yang lebih tepat bila ditangani dengan prosedur-prosedur parametrik. Cara seperti ini sering menyebabkan pemborosan informasi.
2. Kendatipun prosedur nonparametrik terkenal karena prinsip perhitungannya yang sederhana, pekerjaan hitung-menghitung *(arithmetic)*-nya sendiri acap kali membutuhkan banyak tenaga serta menjemukan.

Kapan prosedur nonparametrik digunakan?

Di bawah ini dikemukakan beberapa situasi yang tepat bila ditangani dengan prosedur nonparametrik.

1. Bila hipotesis yang harus diuji tidak melibatkan suatu parameter propulasi.

2. Bila data telah diukur dengan skala yang lebih lemah dibanding yang dipersyaratkan oleh prosedur parametrik yang semestinya digunakan. Sebagai contoh, data mungkin terdiri atas data hitung atau data peringkat, sehingga menghalangi penerapan prosedur parametrik yang semestinya lebih tepat.
3. Bila asumsi-asumsi yang diperlukan agar penggunaan suatu prosedur parametrik menjadi sahih tidak terpenuhi. Dalam banyak hal, rancangan suatu proyek riset mungkin menganjurkan penggunaan prosedur parametrik tertentu. Bagaimanapun, pemeriksaan data mungkin mengungkapkan bahwa salah satu atau beberapa asumsi yang mendasari pengujian betul-betul tidak bisa dipatuhi. Dalam hal begitu, prosedur nonparametrik acap kali merupakan pengganti satu-satunya.
4. Bila hasil-hasil riset harus segera disajikan dan perhitungan-perhitungan terpaksa dikerjakan secara manual.

Kepustakaan mengenai statistika nonparametrik cukup luas. Dalam tahun 1962, bibliografi yang disusun oleh Savage (66) berisi sekitar 3000 judul. Bibliografi yang paling baru dengan sendirinya memuat jumlah judul yang berlipat ganda. Buku-buku tentang metode-

metode nonparametrik yang tidak membutuhkan latar belakang matematika yang kuat antara lain disusun oleh Bradley (67), Conover (68), Gibbons (69), Hollander dan Wolfe (70), Lehmann (71), Maxwell (72), Mosteller dan Rourke (73), Noether (74), Pierce (75), Quenoiller (76), Siegal (77), Senders (78), Tate dan Clelland (79), serta Wilcoxon dan Wilcox (80). Adapun buku-buku yang mempersyaratkan penguasaan matematika yang lebih tinggi adalah yang disusun oleh David (81), Edgington (82), Fraser (83), Gibbons (84), Hájek (85), Hájek dan Šidák (86), Kraft dan van Eeden (87), Noether (88), Puri (89), Sarhan dan Greenberg (90), serta Walsh (91, 92, 93). Kami menganjurkan buku karya Kendall tentang korelasi peringkat (94) bila Anda ingin mendalami pokok permasalahan ini. Jika Anda berminat mempelajari *kai*-kuadrat, buku oleh Lancaster (95) tentu menarik bagi Anda.

Artikel yang ditulis oleh Ruist (96) berisi suatu tinjauan historis yang baik tentang statistika nonparametrik. Hettmansperger dan McKean (97) menyajikan suatu grafik yang berguna untuk menggambarkan hubungan antara statistik-statistik uji nonparametrik dan prosedur-prosedur pendugaan yang berkaitan dengan statistik-statistik tersebut. Artikel-artikel oleh Buchanan (98), Noether (99), dan Scheffé juga menarik.

BAB 6
CAKUPAN BUKU INI

Tekanan dalam buku ini adalah pada penerapan metode-metode statistik nonparametrik. Di mana saja tersedia, contoh-contoh menggunakan data asli, yang dikutip terutama dari hasil-hasil riset yang dipublikasikan dalam berbagai terbitan berkala ilmiah. Kami berharap bahwa penggunaan situasi yang nyata serta data yang asli akan membuat buku ini lebih menarik bagi Anda. Kami telah mengambil persoalan-persoalan yang berasal dari sumber-sumber yang sedapat mungkin sangat beragam guna menunjukkan betapa luas jangkauan penerapan teknik-teknik yang diketengahkan. Teknik-teknik statistik yang dibahas di sini pun sangat beragam. Teknik-teknik itu terutama adalah yang paling memiliki peluang untuk terbukti bermanfaat bagi para peneliti dan paling mungkin muncul dalam kepustakaan riset. Selain pengujian hipotesis, pembahasan dalam buku teks ini juga meliputi pendugaan interval.

Uji-uji kai-kuadrat untuk memeriksa independensi ketidak terkaitan dan homogenitas Di sini kita membahas uji kai-kuadrat, yang mungkin merupakan prosedur nonparametrik yang paling luas penggunaannya. Dalam hal ini ada dua situasi yang kita liput. Dalam situasi yang pertama, data berasal dari suatu sampel tunggal yang terdiri atas subjek-subjek yang diklasifikasikan secara silang berdasarkan dua kriteria, dengan tujuan menentukan apakah kita harus menyimpulkan bahwa kedua kriteria klasifikasi itu bertalian. Dalam situasi yang kedua, kita terlebih dahulu menetapkan dua populasi atau lebih, kemudian dari masing-masing populasi itu kita menarik sebuah sampel. Tujuan kita dalam hal ini adalah menentukan apakah kita harus menyimpulkan bahwa populasi-populasi itu tidak homogen berkenaan dengan beberapa karakteristik tertentu yang dimiliki.

BAB 7
FORMAT DAN ORGANISASI

Dalam menyajikan prosedur-prosedur statistik ini, kami telah mengadopsi suatu format yang sengaja dirancang untuk memudahkan Anda. Setiap prosedur pengujian hipotesis dalam buku ini dipecah menjadi empat bagian, yakni: (1) asumsi-asumsi, (2) hipotesis-hipotesis, (3) statistik uji, dan (4) kaidah pengambilan keputusan.

Dengan demikian, untuk suatu uji tertentu, dengan cepat Anda dapat menentukan asumsi-asumsi yang mendasari uji itu, hipotesis-hipotesis yang sesuai, cara menghitung statistik uji, serta cara menentukan apakah suatu hipotesis nol harus ditolak. Setiap kali, langkah pertama yang kita lakukan adalah membicarakan suatu topik secara umum. Setelah itu kita menggunakan sebuah contoh untuk menjelaskan penerapannya.

Andaikata diperlukan, untuk suatu uji yang diketengahkan, kita mungkin saja membahas kasus angka sama *(ties)*, aproksimasi bila sampelnya besar, dan kuasa serta efisiensi uji yang bersangkutan. Untuk masing-masing prosedur, kami mencantumkan acuan-acuan yang mungkin ingin Anda baca bila Anda berminat mempelajari suatu prosedur secara lebih mendalam atau ingin lebih yakin tentang suatu topik yang berkaitan.

Dalam bab-bab mendatang, kami membagi acuan-acuan pada daftar kepustakaan atas dua jenis: yang pertama adalah acuan-acuan yang bertalian dengan bagian pembahasan utama, dan tergolong ke dalam kepustakaan statistik; yang kedua adalah acuan-acuan yang bertalian dengan contoh-contoh dan latihan-latihan, yang tergolong ke dalam kepustakaan riset. Nomor setiap acuan yang disebutkan dalam pokok pembahasan sengaja ditambahi huruf T di depannya, sedangkan nomor-nomor acuan yang disebutkan pada contoh-contoh dan latihan-latihan ditambahi huruf E. Demikian pula, di depan setiap nomor tabel yang disajikan di bagian Apendiks kami menambahkan huruf A.

BAB 8
Uji-uji kai-kuadrat untuk memeriksa ketidaktergantungan dan homogenitas

Di antara prosedur-prosedur statistik dengan penerapan yang sangat luas terdapat uji-uji kai-kuadrat *(chi-square tests)* untuk memeriksa ketidaktergantungan dan homogenitas. Uji-uji ini dilandaskan pada sebuah teknik yang diperkenalkan dalam tahun 1900 oleh Karl Pearson (T – 1), yang telah digelari "bapak ilmu statistika" (T – 2). Pearson waktu itu mempermasalahkan keselarasan *(goodness of fit)* antara data yang teramati dan kurva-kurva frekuensi teoritisnya. Dalam makalahnya yang penting (T – 1), ia telah merunut uji kai-kuadrat untuk memeriksa keselarasan.

Seperti yang akan Anda lihat, uji-uji kai-kuadrat untuk memeriksa ketidaktergantungan dan homogenitas, pada hakikatnya, adalah uji-uji keselarasan *(goodness of fit tests).* Pada keduanya, prosedur uji selalu meliputi pembandingan frekuensi-frekuensi yang teramati dengan frekuensi-frekuensi yang diharapkan bila hipotesis nol yang ditetapkan benar. Jadi kita mengukur keselarasan antara frekuensi-frekuensi yang teramati dan yang diharapkan. Apabila hasil pengukuran menunjukkan bahwa keselarasan tersebut cukup buruk, maka hipotesis nol kita tolak. Baik-buruknya keselarasan antara frekuensi-frekuensi yang teramati dan yang diharapkan ditentukan dengan cara memperbandingkan ukuran keselarasan hasil perhitungan terhadap suatu harga yang sesuai pada suatu distribusi yang dikenal sebagai *distribusi kai-kuadrat (chi-square distribution).*

Banyak ahli statistika yang beranggapan bahwa uji-uji kai-kuadrat untuk memeriksa ketidaktergantungan dan homogenitas bukan prosedur nonparametrik. Dalam kasus dua sampel, uji-uji ini tidak lain merupakan cara alternatif untuk menguji hipotesis nol yang menyatakan bahwa proporsi-proporsi dua populasi sama besar. Dalam kasus sampel banyak, uji-uji kai-kuadrat memungkinkan kita membuat inferensi tentang parameter-parameter distribusi multinomial. Bagaimanapun, pembahasan uji-uji kai-kuadrat dalam buku-buku statistika nonparametrik sudah lazim, dan kita tidak akan menyimpang dari tradisi itu dengan menghilangkan topik ini dari pembicaraan kita.

Dalam bagian mendatang, secara singkat kita akan mengulas beberapa sifat matematika distribusi kai-kuadrat, dan dalam dua bagian selanjutnya kita membahas uji-uji ketidaktergantungan dan homogenitas. Bagian terakhir disediakan untuk menyajikan serangkaian teknik khusus yang memungkinkan para peneliti memanfaatkan uji-uji kai-kuadrat dengan sebaik-baiknya.

8.1 SIFAT-SIFAT MATEMATIK DISTRIBUSI KAI-KUADRAT

Sebagaimana yang mungkin telah Anda ketahui dari kuliah pengantar statistika, Anda dapat mentransformasikan distribusi normal mana pun ke distribusi normal standar (yang memiliki rata-rata 0 dan deviasi standar 1) menggunakan rumus

$$Z_i = \frac{X_i - \mu}{\sigma}$$

(8.1)

dengan Z_i adalah sebuah nilai dari distribusi normal standar, X_i sebuah nilai dari distribusi normal yang akan ditransformasikan, dan μ serta σ berturut-turut adalah rata-rata dan deviasi standar dari distribusi ini.

Kalau kita mengkuadratkan varian-varian *(variants/variates)* Z_i dalam Persamaan 8.1, yaitu kalau kita menghitung

$$Z_i^2 = \left(\frac{X_i - \mu}{\sigma}\right)^2$$

(8.2)

maka Z^2 mengikuti suatu distribusi kai-kuadrat. Jadi, kalau kita secara acak dan secara bebas memilih nilai-nilai beberapa variabel acak X yang terdistribusi normal, menghitung nilai-nilai normal standarnya, dan mengkuadratkan nilai-nilai standar tersebut, maka kita memiliki variabel acak yang berdistribusi kai-kuadrat.

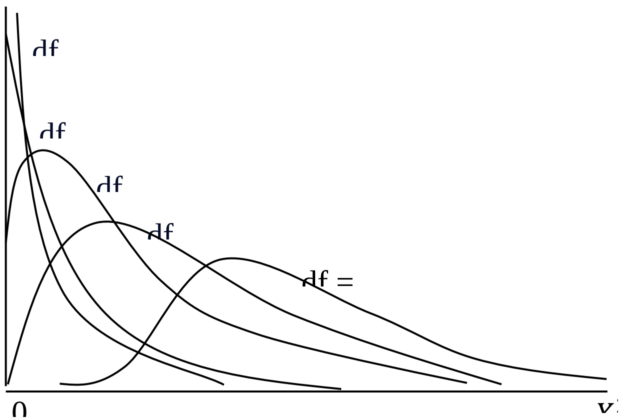
Kalau kita secara acak dan secara bebas memilih sebuah sampel yang terdiri atas dua nilai variabel acak *X* yang terdistribusi normal, mentransformasikannya ke nilai-nilai normal standar dengan menghitung $Z_i = (X_i - \mu)/ \sigma$ dan $Z_2 = (X_2 - \mu)/ \sigma$, mengkuadratkan nilai-nilai *Z* yang dihasilkan itu, dan menghitung jumlah nilai-nilai kuadrat tersebut, maka kita mendapatkan variabel $Z_1^2 + Z_2^2$, yang juga memiliki distribusi kai-kuadrat. Pada umumnya, kalau kita mengikuti prosedur ini untuk sampel berukuran *n*, kita akan memperoleh variabel

$$Z_1^2 + Z_2^2 + Z_3^2 + . . . + Z_n^2$$

Variabel ini juga akan memiliki distribusi kai-kuadrat.

Distribusi kai-kuadrat yang satu berbeda dengan yang lain sesuai dengan besarnya derajat bebas masing-masing variabelnya. Istilah "derajat bebas" di sini mengacu ke banyaknya varian normal standar bebas yang kita kuadratkan dan jumlahkan guna mendapatkan variabel yang memiliki distribusi kai-kuadrat. Variabel kai-kuadrat dituliskan dengan notasi X^2, simbol yang pertama kali digunakan oleh Pearson (T – 3). Sebuah *subscript*, yang menyatakan derajat bebas, boleh pula diimbuhkan untuk membedakan distribusi yang satu dengan yang lain. Jadi X_1^2, X_2^2, dan X_n^2 berturut-turut menerangkan variabel-variabel yang memiliki distribusi kai-kuadrat dengan derajat-derajat bebas 1, 2, dan *n*. Distribusi-distribusi kai-kuadrat untuk beberapa derajat bebas tampak dalam Gambar 8.1. Sesuai dengan teorema limit sentral, bila *n* semakin besar, maka distribusi kai-kuadrat mendekati distribusi normal.

Gambar 8.1 Distribusi kai-kuadrat untuk beberapa derajat bebas



Tabel A.12 (baca: Tabel Statistik) menyediakan nilai-nilai kai-kuadrat untuk berbagai nilai derajat bebas. Subskrip pada X^2 yang dicantumkan di bagian atas setiap kolom menunjukkan proporsi daerah di bawah kurva kai-kuadrat yang terletak di sebelah kiri nilai yang dimuat dalam tabel. Sebagai contoh, $X^2_{0.95}$ menunjukkan bahwa 95% dari daerah di bawah kurva kai-kuadrat untuk derajat bebas 1 terletak di sebelah kiri 3.841.

Nilai rata-rata suatu distribusi kai-kuadrat sama dengan derajat bebasnya, dan variansnya sama dengan dua kali derajat bebasnya. Sebagai contoh, rata-rata distribusi kai-kuadrat dengan derajat bebas 10 adalah 10 dan variansnya adalah 20.

Pembahasan yang lebih mendalam mengenai sifat-sifat matematik distribusi kai-kuadrat bisa dijumpai di kebanyakan buku teks tentang statistika matematik. Buku yang secara khusus membahas distribusi kai-kuadrat ditulis oleh Lancaster (T – 4).

Dalam bagian-bagian mendatang kita akan menjumpai bahwa, menurut beberapa hipotesis nol, ukuran (statistik) untuk keselarasan antara data yang teramati dan data yang diharapkan kurang-lebih memiliki distribusi kai-kuadrat bila hipotesis nol itu benar. Karena itu, dalam setiap uji, distribusi kai-kuadrat dengan derajat-derajat bebas yang sesuai dijadikan sebagai pembanding standar bagi nilai statistik uji hasil perhitungan sehingga kita dapat memutuskan apakah harus menolak hipotesis nol itu atau tidak.

BAB 9
UJI KAI-KUADRAT UNTUK MEMERIKSA KETIDAKTERGANTUNGAN

Sebuah pertanyaan dalam riset yang sering muncul adalah apakah dua variabel yang diminati saling berkaitan. Sebagai contoh, seorang ahli sosiologi mungkin ingin tahu apakah aras pendidikan formal berkaitan dengan pendapatan. Seorang pejabat lembaga perlindungan konsumen mungkin berminat untuk mengetahui apakah nilai perabotan rumah tangga berkaitan dengan mutunya. Atau seorang ahli gizi mungkin ingin tahu apakah perolehan gizi para pelajar bertalian dengan unjuk kerja akademik mereka.

Apabila antara dua variabel tidak ada pertalian, maka kita mengatakan bahwa keduanya bebas (tidak saling mempengaruhi). Dua variabel disebut independen bila distribusi yang satu sama sekali tidak bergantung pada (tidak dipengaruhi oleh) distribusi yang lain. Apabila dua variabel tidak berkaitan (yakni, bila keduanya saling bebas), maka meskipun kita mengetahui nilai salah satu variabel untuk suatu subjek, ini tidak akan membantu kita dalam menentukan nilai variabel yang lain untuk subjek yang sama. Sebagai contoh, apabila harga dan mutu suatu perabot rumah tangga saling tidak mempengaruhi, seseorang yang mengetahui harga perabot tersebut tidak menjadi lebih ahli dalam meramalkan mutunya ketimbang orang yang tidak mengetahui harga barang tadi. Di pihak lain, apabila dua variabel berhubungan, pengetahuan tentang yang satu akan bermanfaat dalam meramalkan nilai-nilai yang diminati pada yang lain.

Uji kai-kuadrat untuk memeriksa ketidaktergantungan inilah yang akan kita gunakan untuk memutuskan apakah dua variabel dalam suatu populasi saling bebas.

Asumsi-asumsi

A. Data terdiri atas sebuah sampel acak sederhana berukuran dari suatu populasi yang diminati.
B. Hasil-hasil pengamatan dalam sampel boleh diklasifikasikan secara silang *(cross-classified)* menurut dua kriteria, sehingga masing-masing hasil pengamatan memenuhi salah satu dan hanya satu aras dari masing-masing kriterium. Kriteria di sini adalah variabel-variabel yang diminati dalam situasi yang dihadapi.

Tabel 9.1 Tabel kontingensi untuk uji ketidakterkaitan kai-kuadrat

Kriterium klasifikasi pertama	**Kriterium klasifikasi kedua**						
	Tingkat						
Tingkat	**1**	**2**	**…**	***j***	**…**	**c**	**Jumlah**
1	n_{11}	n_{12}	…	n_{1j}	…	n_{1c}	n_{1}
2	n_{21}	n_{22}	…	n_{2j}	…	n_{2c}	n_{2}
.							.
.							.
.							.
i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ic}	n_{i}
.							.
.							.
.							.
r	n_{r1}	n_{r2}	…	n_{rj}	…	n_{rc}	$n_{r.}$
Jumlah	$n_{.1}$	$n_{.2}$	…	$n_{.j}$	…	$n_{.c}$	n

Data boleh disajikan dalam tabel kontingensi seperti Tabel 9.1. Di sini, banyaknya subjek teramati, n_{ij}, yang termasuk salah satu aras dari masing-masing kriterium ditempatkan dalam sel yang terbentuk oleh perpotongan antara baris ke-*i* dan kolom ke-*j*. Isi sel semacam ini disebut *frekuensi sel teramati (obsevered cell frequency),* dan biasa dituliskan dengan notasi O_{ij}, jadi $O_{ij} = n_{ij}$. Frekuensi sel yang teramati O_{ij} menunjukkan banyaknya subyek dalam sampel yang selain memenuhi aras ke-*i* dari kriterium klasifikasi yang pertama juga memenuhi aras ke-*j* dari kriterium kedua.

Hipotesis-hipotesis

H_0 = Kedua kriteria klasifikasi bebas

H_1 = Kedua kriteria klasifikasi tidak bebas

Statistik uji

Kita menghitung statistik uji berdasarkan asumsi bahwa H_0 benar, yaitu bahwa kedua kriteria klasifikasi bebas. Sebagaimana yang telah kita tekankan, uji ketidaktergantungan kai-kuadrat memperbandingkan hasil-hasil yang teramati dengan hasil-hasil yang diharapkan bila H_0 benar. Tegasnya, uji ini meliputi pembandingan frekuensi-frekuensi sel yang teramati terhadap frekuensi-frekuensi sel yang diharapkan bila H_0 benar.

Guna mendapatkan frekuensi-frekuensi sel yang diharapkan, kita menggunakan hukum probabilitas elementer sebagai berikut: jika dua peristiwa saling independen, peluang untuk terjadi secara bersamaan sama dengan hasil perkalian besarnya peluang masing-

masing peristiwa. Bila H_0 benar-yaitu, bila kedua kriteria klasifikasi bebas-peluang untuk memasukkan suatu subjek ke dalam hitungan untuk sel *ij* sama dengan peluang untuk memasukkannya ke dalam hitungan untuk baris ke-*i* kali peluang untuk memasukkannya ke dalam hitungan untuk kolom ke-*j*. Dari data sampel, peluang-peluang yang belakangan ini berturut-turut dapat kita duga dengan n_i/n dan n_j/n. Selanjutnya kita dapat menuliskan dugaan peluang masuknya suatu subjek ke dalam hitungan untuk sel *ij* sebagai berikut:

$$P \text{ (subjek dihitung untuk sel } ij) = \left(\frac{n_{i.}}{n}\right) \left(\frac{n_{.j}}{n}\right)$$
(9.3)

Untuk mendapatkan E_{ij}, frekuensi yang diharapkan untuk sel *ij*, kita mengalikan dugaan peluang di atas dengan total ukuran sampel. Jadi, frekuensi yang diharapkan untuk sel pada tabel kontingensi seperti Tabel 9.1 adalah

$$E_{ij} = \text{n} \left(\frac{n_{i.}}{n}\right) \left(\frac{n_{.j}}{n}\right)$$
(9.4)

Karena *n* yang menjadi pembilang dalam Persamaan 9.4 saling meniadakan dengan salah satu *n* yang menjadi penyebut, maka persamaan ini dapat disederhanakan menjadi

$$E_{ij} = n_{i.}n_{.j}/\text{n}$$

(9.5)

Bentuk persamaan yang terakhir ini menunjukkan bahwa kita dengan mudah dapat menghitung frekuensi sel yang diharapkan dengan mengalihkan total-total baris dan kolom yang sesuai dan membagi hasilnya dengan total ukuran sampel.

Bilamana kita telah mempunyai frekuensi-frekuensi sel yang teramati dan frekuensi-frekuensi sel yang diharapkan, maka yang kita minati adalah besarnya beda-beda antara keduanya. Tegasnya, kita ingin tahu apakah beda-beda itu cukup kecil sehingga bisa dianggap kebetulan *(sampling variability)* bila H_0 benar, atau apakah beda-beda itu sedemikian besar sehingga penjelasan lain (misalnya bahwa H_0 salah) diperlukan. Dari frekuensi-frekuensi yang diharapkan dan yang teramati itu kita boleh menghitung suatu statistik uji yang mencerminkan besarnya beda-beda antara keduanya. Apabila H_0 benar, statistik uji ini kurang-lebih memiliki distribusi X^2 dengan derajat bebas (r – 1)(c – 1), di mana r adalah banyaknya baris dan c banyaknya kolom dalam tabel kontingensi. Statistik uji ini adalah

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

(9.6)

Apabila beda-beda antara frekuensi-frekuensi yang teramati dan diharapkan besar, maka X^2 juga besar; sedangkan apabila beda-beda tersebut kecil, maka X^2 pun kecil.

Kaidah pengambilan keputusan

Kita boleh menolak hipotesis nol yang menyatakan ketergantungan pada taraf nyata α jika nilai statistik uji X^2 hasil perhitungan lebih besar daripada nilai X^2_{1-} α dalam tabel untuk derajat bebas (r – 1)(c – 1).

Contoh 9.1

Data dalam Tabel 9.2, yang dilaporkan oleh Monteiro (E – 1), adalah data tentang 2765 penduduk Pulau Rhode yang diklasifikasikan menurut penghasilan dan selang waktu sejak kali terakhir berkonsultasi dengan dokter. Kita ingin tahu apakah data ini menyediakan bukti yang memadai untuk menunjukkan adanya kaitan antara besar penghasilan dan panjang selang waktu sejak konsultasi terakhir dengan dokter. Dengan kata lain, kita ingin tahu apakah kita dapat menyimpulkan bahwa kedua variabel di atas tidak bebas.

Tabel 9.2 Saat konsultasi terakhir dengan dokter, menurut klasifikasi pendapatan. Rhode Island. 1971.

Konsultasi terakhir dengan dokter

Pendapatan	**Dalam 6 bulan terakhir**	**Dalam 7 bulan hingga satu tahun yang lalu**	**Lebih dari setahun yang lalu**	**Jumlah**
Less than $3000	186	38	35	259
$3000-$4999	227	54	45	326
$5000-$6900	219	78	78	375
$7000-$9999	355	112	140	607
$10,000 plus	653	285	259	1197
Jumlah	1640	567	557	2746

Sumber: Lois A. Monteiro, "Expense Is No Object…: Income and Physician Visits Reconsidered," J. Health Soc. Beh., 14 (1973), 99-115

Hipotesis-hipotesis

H_o: Penghasilan dan selang waktu sejak konsultasi terakhir dengan dokter saling bebas

H_1: Kedua variabel tidak bebas

Statistik uji

Dengan persamaan 9.5 kita mendapati bahwa frekuensi sel yang diharapkan untuk sel 11 adalah

E_{11} = (259) (1640)/2764 = 153.68

Perhitungan-perhitungan serupa menghasilkan frekuensi-frekuensi yang diharapkan untuk sel-sel yang lain, yaitu yang tampak dalam tanda kurung pada Tabel 9.3.

Tabel 9.3 Frekuensi teramati dan frekuensi harapan untuk Contoh 9.1 (Frekuensi harapan adalah yang dalam kurung)

Konsultasi terakhir dengan dokter

Pendapatan	**Dalam 6 bulan terakhir**	**Dalam 7 bulan hingga satu tahun yang lalu**	**Lebih dari setahun yang lalu**	**Jumlah**

Less than $3000	186 (153.68)	38 (53.13)	35 (52.19)	259
$3000-$4999	227 (193.43)	54 (66.87)	45 (65.70)	326
$5000-$6999	219 (222.50)	78 (76.93)	78 (75.57)	375
$7000-$9999	355 (360.16)	112 (124.52)	140 (122.32)	607
$10.000 plus	653 (710.23)	285 (245.55)	259 (241.22)	1197
Jumlah	1640	567	557	2746

Nilai statistik uji hasil perhitungan menggunakan Persamaan 9.6 adalah

$$X^2 = \frac{(186-153.68)^2}{153.68} + \frac{(227-193.43)^2}{193.43} + \dots + \frac{(259-241.22)^2}{241.22} = 47.90$$

Derajat bebas di sini adalah (5 – 1)(3 – 1) = 8.

Keputusan

Karena 47.90 lebih besar daripada $X^2_{0.995} = 21.955$, kita dapat menolak H_0 pada taraf nyata 0.005. Jadi, kita berkesimpulan bahwa penghasilan dan selang waktu sejak konsultasi terakhir dengan dokter tidak bebas. Nilai P dalam contoh ini kurang dari 0.005.

Frekuensi-frekuensi harapan yang kecil Statistik

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

Kurang-lebih memiliki distribusi X^2 bila H_0 benar, tetapi ini hanya berlaku jika frekuensi-frekuensi harapan E_{ij} besar. Mengenai sampai berapa besar E_{ij} itu sehingga penerapan uji kai-kuadrat sahih hingga kini belum disepakati secara bulat oleh para ahli statistika. Beberapa penulis telah menganjurkan nilai minimum 10 untuk E_{ij}. Bagaimanapun, untuk tabel kontingensi dengan derajat bebas lebih dari satu, Cochran (T – 5, T – 6) menyatakan bahwa frekuensi sel minimum sekecil 1 diperbolehkan bila 20% atau lebih sedikit dari sel-sel memiliki frekuensi-frekuensi harapan kurang dari 5. Apabila X^2 memiliki derajat bebas kurang dari 30 dan frekuensi harapan minimum 2 atau lebih, Cochran (T – 6) menyatakan bahwa penggunaan tabel-tabel X^2 yang biasa umumnya memadai.

Baris-baris yang bersebelahan dan/atau kolom-kolom yang bersebelahan dalam suatu tabel kontingensi boleh digabungkan guna mendapatkan frekuensi-frekuensi sel harapan yang minimum. Masalah yang menyangkut frekuensi-frekuensi harapan yang kecil juga telah dikaji oleh Katti dan Sastry (T – 7), Nass (T – 8), Roscoe dan

Byars (T – 9), Tate dan Hyer (T – 10, T – 11), serta Yarnold (T – 12).

Tabel kontingensi 2 x **2** Manakala masing-masing dari dua kriteria klasifikasi terpenuhi pada dua aras, tabel kontingensi yang dihasilkan terdiri atas dua baris dan dua kolom yang membentuk empat buah sel. Tabel semacam ini disebut tabel kontingensi 2 x 2 dan umumnya tampak seperti Tabel 9.4. Bila kita menerapkan kaidah (r – 1)(c – 1) kita menjumpai bahwa derajat bebas di sini adalah satu. Guna menghitung X^2 dari data pada tabel 2 x 2 kita dapat menggunakan rumus pintas sebagai berikut.

$$X^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(c+d)(a+b)}$$

(9.7)

Prosedur untuk menganalisis data dari suatu tabel 2 x 2 dijelaskan dalam contoh di bawah.

Tabel 9.4 Tabel kontingensi 2 x 2

	Kriterium klasifikasi kedua		
Kriterium klasifikasi pertama	**1**	**2**	**Jumlah**
1	a	b	$a + b$
2	c	d	$c + d$
Jumlah	$a + c$	$b + d$	n

Contoh 9.2

Abse dkk. (E – 2) melaporkan data dalam Tabel 9.5 tentang kebiasaan merokok di malam hari dan penyakit kanker paru-paru. Data ini diperoleh dari 56 subjek. Kita ingin tahu apakah kita boleh menyimpulkan dari data ini bahwa merokok di malam hari berkaitan dengan kanker paru-paru.

Tabel 9.5 Kebiasaan merokok malam dan kanker paru-paru pada 56 subjek

	Merokok malam		
Kanker paru-paru	**Ya**	**Tidak**	**Jumlah**
Ya	20	16	36
Tidak	6	14	20
Jumlah	26	30	56

Sumber: D. Wilfred Abse, Marilyn M. Wilkins, Gordon Kirshner, Don. L. Weston, Robert S. Brown, and W. D. Buxton, "Self-Frustation, Nighttime Smoking and Lunc Cancer," Psychosom. Med., 34 (1972), 395 – 404

Hipotesis-hipotesis

H_0 = Merokok malam dan kanker paru-paru tidak saling berhubungan

H_1 = Kedua variabel berhubungan (tidak independen)

Statistik uji

Dengan data dari Tabel 9.5 kita menggunakan persamaan 9.7 untuk menghitung

$$X^2 = \frac{56\,[(20)(14)-(16)(6)]^2}{(26)(30)(20)(36)} = 3.376$$

Keputusan

Karena X^2 = 3.376 lebih kecil dari 3.841, nilai kai-kuadrat dalam tabel untuk derajat bebas satu dan α = 0.05, kita tidak dapat menolak H_0 pada taraf nyata 0.05. Oleh sebab itu kita menyimpulkan bahwa pada taraf nyata 0.05 kedua variabel-kebiasaan merokok malam dan kanker paru-paru-mungkin tidak berhubungan (0.10 > nilai P > 0.05).

Masalah kecilnya frekuensi yang diharapkan mungkin pula dihadapi dalam situasi-situasi yang menyangkut tabel kontingensi 2 x 2. Dalam kasus-kasus begini pun, orang acap kali mengikuti anjuran Cochran (T – 5, T – 6). Cochran menganjurkan agar uji kai-kuadrat tidak digunakan bila n < 20. Untuk 20 < n < 40, ia juga menganjurkan agar uji kai-kuadrat tidak digunakan bila frekuensi harapan terkecil kurang dari 5. Untuk n > 40, tak ada frekuensi harapan dalam tabel kontingensi 2 x 2 yang harus kurang dari 1.

Koreksi Yates Pada tahun 1934, Yates (T – 13) mengusulkan suatu prosedur koreksi yang dipakai bila X^2 dihitung menggunakan data dari tabel kontingensi 2 x 2. Prosedur yang dikenal sebagai *koreksi Yates untuk kontinyuitas* ini tidak lain merupakan pengurangan 0.5 dari nilai mutlak ad – bc dalam pembilang di Persamaan 9.7. Maksud prosedur ini adalah untuk "mengoreksi" distribusi X^2 kontinyu yang digunakan dalam aproksimasi terhadap distribusi X^2 diskret. Bila kita menggunakan koreksi tersebut, Persamaan 9.7 menjadi

$$X_c^2 = \frac{n(\backslash ad - bc \backslash - 0.5n)^2}{(a+c)(b+d)(c+d)(a+b)}$$

(9.8)

Kendatipun koreksi Yates dahulu dipakai secara luas, sejumlah kritik telah ditimpakan ke prosedur ini, antara lain dari Conover (T – 14), Grizzle (T – 15), Lancaster (T – 16), Pearson (T – 17), dan Plackett (T – 18). Sebagai akibatnya, belakangan ini timbul kecenderungan untuk tidak menggunakan lagi koreksi tersebut. Untuk mengetahui ulasan-ulasan lain tentang koreksi Yates, baca artikel-artikel yang ditulis oleh Liddell (T – 19) serta Mantel dan Greenshouse (T – 20).

Kalau kita menerapkan koreksi Yates pada analisis data di Contoh 9.2 yang tampak dalam Tabel 9.5, kita memperoleh

$$X_c^2 = \frac{56\ [|(20)(14)-(16)(6)|-0.5\ (56)]^2}{(26)(30)(20)(36)} = 2.427$$

Perhatikan bahwa pada aras kebermaknaan taraf nyata 0.05, kita sampai ke kesimpulan yang sama, baik bila kita menggunakan atau tidak menggunakan koreksi Yates (harga P > 0.10).

BAB 10
UJI KAI-KUADRAT UNTUK MEMERIKSA HOMOGENITAS

Dalam statistika elementer, kita telah mempelajari penggunaan distribusi normal untuk mendekati distribusi binomial pada pengujian hipotesis nol yang menyatakan bahwa dua proporsi populasi p_1 dan p_2 sama besar. Hipotesis nol ini biasanya ditulis sebagai H_0: p_1= p_2. Sebagai alternatif, hipotesis nol tadi bisa dinyatakan dengan H_0: kedua populasi homogen (dalam hal proporsi subjek-subjek dengan karakteristik yang diminati). Pada lazimnya, kita menarik sebuah sampel dari masing-masing populasi yang diselidiki, lalu mengklasifikasikan subjek-subjeknya menurut apakah subjek-subjek tersebut memiliki karakteristik yang diminati. Hasil-hasilnya boleh diperagakan dalam tabel kontingensi 2 x 2 seperti tabel 10.1. Dengan jelas kita dapat melihat keserupaan antara Tabel 10.1 dan tabel kontingensi 2 x 2 yang terdahulu (Tabel 9.4). Dan sudah barang tentu kita dapat menghitung statistik X^2 dari data pada Tabel 10.1 menggunakan Persamaan 9.7.

Tabel 10.1

Tabel kontingensi 2 x 2

	Adanya Karakteristik yang diminati		
Sampel	**Ya**	**Tidak**	**Jumlah**

1	*a*	*b*	*a* + *b*
2	*c*	*d*	*c* + *d*
Jumlah	*a* + *c*	*b* + *d*	*n*

Kendati, baik pada pengujian homogenitas maupun pengujian ketidaktergantungan kita menghitung X^2 menggunakan rumus yang sama, uji-uji ini berbeda dalam dua hal penting: prosedur penyampelan dan pola pemikiran yang melandasi perhitungan frekuensi-frekuensi yang diharapkan. Dalam pembicaraan mendatang kita akan membahas perbedaan-perbedaan ini secara lebih rinci.

Kita menguji hipotesis nol yang menyatakan bahwa kedua populasi asal kedua sampel homogen dengan membandingkan nilai X^2 hasil perhitungan dengan nilai X^2 dalam tabel untuk derajat bebas satu. Uji seperti ini disebut *uji homogenitas kai-kuadrat* atau *uji kai-kuadrat untuk memeriksa homogenitas.* Prosedur pengujian yang dijelaskan di atas untuk tabel kontingensi 2 x 2 hanyalah suatu kasus khusus. Apabila tabel kontingensi terdiri atas baris dan kolom, kita boleh meringkaskan prosedur pengujian kita sebagai berikut.

Asumsi-asumsi

A. Sampel-sampel yang diamati bebas.
B. Sampel-sampel yang diamati acak.

C. Masing-masing subjek dalam populasi boleh diklasifikasikan ke dalam salah satu dari dua kategori yang paling eksklusif, berdasarkan apakah subjek tersebut memiliki atau tidak memiliki karakteristik yang diminati.

Hipotesis-hipotesis

H_0 : Populasi-populasi asal sampel homogen

H_1 : Populasi-populasi asal sampel tidak homogen.

Statistik uji

Statistik uji kita adalah

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

(10.2)

Untuk tabel kontingensi 2 x 2, kita boleh menghitung menggunakan Persamaan 9.7.

Kaidah pengambilan keputusan

Kita menolak H_0 jika nilai X^2 hasil perhitungan lebih besar dari atau sama dengan nilai X^2 dalam tabel untuk derajat bebas (r – 1)(c – 1).

Melalui contoh berikut, prosedur untuk kasus tabel kontingensi 2 x 2 akan menjadi jelas bagi kita.

Contoh 10.1.

Richardson dkk. (E – 14) melaporkan ada dan tidaknya *respiratory distress syndrome* (RDS) dalam dua kelompok bayi. Kelompok 1 terdiri atas 42 bayi yang selaput janinnya pecah tidak sampai 24 jam sebelum kelahiran, sementara kelompok 2 terdiri atas 22 bayi yang selaputnya pecah lebih dari 24 jam sebelum kelahiran. Data tersebut tampak dalam Tabel 5.18 dan boleh kita gunakan untuk menguji hipotesis nol yang menyatakan bahwa kedua populasi homogen. Misalnya α = 0.05.

Tabel 10.2 Ada dan tidaknya *respiratory distress syndrome* (RDS) dalam dua kelompok bayi

	RDS		
Kelompok	**Ya**	**Tidak**	**Jumlah**
1	27	15	42
2	7	15	22

Jumlah	34	30	64

Hipotesis-hipotesis

H_0 = Kedua populasi yang diwakili oleh kedua kelompok bayi dalam penelitian tersebut homogen dalam hal adanya RDS

H_1 = Kedua populasi tidak homogen

Statistik uji

Dengan Persamaan 9.7 kita mendapatkan

$$X^2 = \frac{64\,[(27)(15)-(15)(7)]^2}{(34)(30)(22)(42)} = 6.112$$

Keputusan

Karena 6.112 > 3.841, tolaklah H_0 dan simpulkan bahwa kedua populasi tidak homogen (0.025 > nilai *P* > 0.01).

Uraian dan ulasan mengenai frekuensi harapan yang kecil dan koreksi Yates dalam hubungan dengan tabel 2 x 2 yang berlaku untuk uji ketidaktergantungan juga berlaku untuk uji homogenitas pada kasus 2 x 2.

Tabel kontingensi *r* x 2 Kita dapat memperluas cakupan uji homogenitas kai-kuadrat untuk kasus dengan tiga populasi atau lebih yang subjek-subjeknya dapat diklasifikasikan ke dalam salah satu dari dua kategori yang saling eksklusif. Data untuk uji semacam ini biasanya diperoleh dengan menarik sampel-sampel acak independen dari ke-*r* populasi, dan semuanya dapat disajikan dalam suatu tabel kontingensi r x 2, dengan ke-*r* barisnya menyatakan ke-*r* populasi dan kedua kolomnya menyatakan kedua klasifikasi karakteristik yang diminati. Uji homogenitas dalam situasi begini akan dijelaskan dalam contoh berikut.

Contoh 10.2

Mims dkk. (E – 15) melaporkan karakteristik subjek-subjek yang menghadiri suatu seminar lima hari tentang seksualitas manusia seperti tampak dalam Tabel 10.3. Dalam uji homogenitas berikut, kita mengandaikan bahwa keempat kelompok subjek tersebut merupakan sampel-sampel acak sederhana yang bebas dari keempat populasi yang tercantum di ujung-ujung baris.

Tabel 10.3 Status perkawinan empat kelompok subjek yang menghadiri suatu kursus tentang seksualitas manusia

Kelompok	*Single*	Kawin atau cerai	Jumlah
Mhs. Kedokteran	50	20	70
Mhs. Perawat Kesehatan	12	25	37
Mahasiswa bidang lain	6	8	14
Pemimpin kelompok	1	21	22
Jumlah	69	74	143

*Sumbe*r: Fern Mims, Rosalee Yeaworth, and Stephen Horstein, “Effectiveness of an Interdisciplinary Course in Human Sexuality,” Nursing Res., 23 (1974), 248 – 253. Copyright 1974, The American Journal of Nursing Company

Hipotesis-hipotesis

H_0 : Keempat populasi yang diselidiki homogen dalam hal status perkawinan

H_1 : Keempat populasi tidak homogen dalam hal status perkawinan

Statistik uji

Setelah dihitung, frekuensi-frekuensi sel yang diharapkan tampak dalam Tabel 5.20.

Tabel 10.4 Frekuensi-frekuensi sel harapan untuk contoh 10.2

Kelompok	*Single*	Kawin atau cerai	Jumlah
Mhs. Kedokteran	33.776	36.224	70
Mhs. Perawat kesehatan	17.853	19.147	37
Mahasiswa bidang lain	6.755	7.245	14
Pemimpin kelompok	10.615	11.385	22
Jumlah	69	74	143

Dengan frekuensi-frekuensi harapan dan frekuensi-frekuensi teramati pada Tabel 5.19, kita dapat menghitung

$$X^2 = \frac{(50-33.776)^2}{33.776} + \frac{(20-36.224)^2}{36.224} + \ldots + \frac{(21-11.385)^2}{11.385} = 35.761$$

Keputusan

Karena 35.761 > X_3^2 = 12.838, kita dapat menolak H_0 pada taraf nyata 0.005. Dengan demikian, berdasarkan data yang dilaporkan, kita menyimpulkan bahwa keempat populasi tersebut tidak homogen dalam hal status perkawinan. Nilai *P* di sini kurang dari 0.005.

Tabel kontingensi r x **c** Acap kali, subjek-subjek dari suatu populasi dapat dimasukkan ke dalam tiga buah kategori atau lebih yang saling eksklusif untuk karakteristik tertentu. Apabila subjek dalam sampel-sampel dari dua populasi diklasifikasikan ke dalam salah satu dari tiga kategori atau lebih, hasilnya dapat diperagakan dalam suatu tabel kontingensi 2 x *c*. Dalam kasus yang lebih umum, kita sering harus mengklasifikasikan subjek-subjek dari tiga populasi atau lebih ke dalam tiga kategori atau lebih yang saling eksklusif. Dalam kondisi begini, data boleh diperagakan melalui tabel kontingensi *r* x *c* seperti Tabel 10.5, dengan ke-*r* barisnya yang menyatakan ke-*r* populasi dan ke-*c* kolomnya yang menyatakan ke-*c* kategori klasifikasi. Kita boleh melakukan uji homogenitas kai-kuadrat terhadap data dalam tabel kontingensi tersebut. Dengan sendirinya uji ini mirip dengan uji ketidaktergantungan untuk tabel kontingensi *r* x *c* yang dibahas di bagian terdahulu.

Tabel 10.5 Tabel kontingensi *r* x *c* untuk uji homogenitas kai-kuadrat

Kategori klasifikasi

Populasi	**1**	**2**	**…**	***j***	**…**	***c***	**Jumlah**
1	n_{11}	n_{12}	…	n_{1j}	…	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2j}	…	n_{2c}	$n_{2.}$
.							.
.							.
.							.
i	n_{i1}	n_{i2}		n_{ij}		n_{ic}	$n_{i.}$
.							.
.							.
.							.
r	n_{r1}	n_{r2}	…	n_{rj}	…	n_{rc}	$n_{r.}$
Jumlah	$n_{.1}$	$n_{.2}$	…	$n_{.j}$	…	$n_{.c}$	n

Salah satu hal penting yang membuat uji homogenitas kai-kuadrat berbeda dari uji ketidaktergantungan kai-kuadrat adalah pola pemikiran yang mendasari perhitungan frekuensi-frekuensi harapannya. Dalam uji ketidaktergantungan yang diterangkan di bagian terdahulu, kita menghitung frekuensi-frekuensi sel harapan untuk tabel kontingensi dengan pengandaian bahwa kedua kriteria klasifikasi yang digunakan bebas. Selanjutnya, sesuai dengan hukum probabilitas, peluang masuknya subjek ke gabungan dua taraf dari masing-masing kriterium sama dengan hasil kali kedua peluang masuknya subjek ke salah satu taraf.

Dalam uji homogenitas, kita menghitung frekuensi-frekuensi sel harapan dengan pengandaian bahwa populasi-populasi yang terwakilkan dalam tabel kontingensi homogen sehubungan dengan variabel yang diminati. Apabila ini benar, maka kita boleh memperlakukan beberapa sampel sebagai suatu sampel tunggal yang besar sejauh masih berkaitan dengan kriterium atau variabel klasifikasi. Akibatnya, untuk populasi mana pun yang diminati dalam penelitian, dugaan terbaik atas proporsi subjek-subjek yang masuk ke dalam salah satu kategori yang tersedia dapat kita peroleh dengan membagi total dari semua sampel yang masuk ke kategori tersebut dengan total semua subjek dalam penelitian.

Sebagai contoh, dalam Tabel 10.5, dugaan terbaik atas proporsi ke-$n_1.$ subjek dalam sampel dari populasi 1 yang masuk ke kategori 1 dari kriterium klasifikasi sama dengan $n_{.1}$/n, yakni total besar dalam kategori 1 yang dibagi dengan total banyaknya subjek dalam penelitian. Kemudian, untuk mendapatkan banyaknya subjek yang diharapkan dari ke-$n_1.$ subjek yang masuk ke kategori 1, kita mengalikan proporsi yang diharapkan dalam kategori itu dengan n_1 sehingga

$$E_{11} =\left(\frac{n_{.1}}{n}\right)n_{1.}$$

Jadi, perhatikan bahwa kita bisa mendapatkan frekuensi sel harapan yang mana pun dengan cara membagi hasil perkalian antara jumlah-jumlah tepi yang bersangkutan dengan total jumlah *(grand total).* Ingatlah bahwa ini sama dengan prosedur yang digunakan untuk menghitung frekuensi-frekuensi sel harapan dalam uji ketidaktergantungan. Dengan demikian kita melihat bahwa meskipun uji ketidaktergantungan dan uji homogenitas dalam hal komputasi identik, pola pemikiran yang mendasari perhitungan frekuensi-frekuensi yang diharapkan berbeda.

Kedua uji itu juga berbeda dengan cara yang biasa ditempuh untuk pengumpulan data. Untuk uji ketidaktergantungan, lazimnya investigator membentuk suatu sampel tunggal dari sejumlah subjek yang berasal dari sebuah populasi, dan kemudian mengklasifikasikan subjek-subjek tersebut secara silang menurut dua kriteria yang diminati. Dalam pada itu, untuk uji homogenitas, investigator biasanya menetapkan dahulu dua populasi atau lebih yang diminati sebelum mengumpulkan data, dan menarik sebuah sampel bebas dari masing-masing populasi yang telah ditetapkan. Sesudah pengumpulan data, investigator menempatkan subjek-subjek dari masing-masing sampel, secara terpisah, ke dalam salah satu dari dua kategori klasifikasi atau lebih. Dalam kedua kasus di atas, kita boleh meringkaskan hasil-hasilnya dalam tabel kontingensi. Yang berikut ini adalah contoh penerapan uji kai-kuadrat untuk memeriksa homogenitas.

Tabel 10.6 Reaksi-reaksi terhadap pertanyaan: “Adakah masalah pencemaran udara di sekitar Anda?”

Daerah tempat tinggal	Tidak	Ya	Ragu-ragu	Tidak tahu	Jumlah
Rawmarsh	5	31	2	2	40
Treeton	10	21	4	5	40
Wath	11	20	7	2	40
Jumlah	26	72	13	9	120

Sumber: Geoffrey Wall, "Public Response to Air Pollution in South Yorkshire, England," Environ. and Beh., 5 (June 1973), 239; dicetak ulang dengan izin dari Sage Publications

Contoh 10.3

Ketika menyelidiki taraf kesadaran dan keprihatinan masyarakat terhadap pencemaran udara, Wall (E – 16) mewawancarai sampel-sampel dari tiga daerah yang berlainan di Inggris, yang masing-masing terdiri atas 40 penduduk. Tabel 5.22 menampilkan tanggapan-tanggapan mereka atas pertanyaan: "Adakah masalah pencemaran udara di sekitar tempat tinggal Anda?" Untuk ini kita akan menggunakan prosedur pengujian hipotesis yang baru saja kita pelajari. Misalkan $\alpha = 0.05$.

Hipotesis-hipotesis

H_0 : Ketiga populasi masyarakat di atas homogen dalam hal pengetahuan tentang ada-tidaknya masalah pencemaran udara

H_1 : Ketiga populasi tersebut tidak homogen

Statistik uji

Sesudah dihitung frekuensi-frekuensi sel harapan tampak seperti dalam Tabel 10.7. Dengan frekuensi-frekuensi harapan ini serta frekuensi-frekuensi teramati dari Tabel 10.7, kita menghitung

$$X^2 = \frac{(5-8.6667)^2}{8.6667} + \frac{(31-24)^2}{24} + \ldots + \frac{(2-3)^2}{3} = 10.391$$

Tabel 10.7 Frekuensi-frekuensi sel harapan untuk Contoh 10.3

Daerah tempat tinggal	**Tidak**	**Ya**	**Ragu-ragu**	**Tidak tahu**	**Jumlah**
Rawmarsh	8.6667	24	4.3333	3	40
Treeton	8.6667	24	4.3333	3	40
Wath	8.6667	24	4.3333	3	40
Jumlah	26.0001	72	13.0000	9	120

Keputusan

Karena 10.391 < X_6^2 = 12.592, maka kita tidak dapat menolak H_0, dan kita berkesimpulan bahwa populasi-populasi tersebut mungkin homogen dalam hal pengetahuan tentang ada-tidaknya masalah pencemaran udara (nilai P > 0.10).

Metode-metode untuk menangani masalah kecilnya frekuensi harapan dalam kasus tabel kontingensi r x c yang diberikan di bagian terdahulu berlaku pula untuk kasus uji homogenitas.

BAB 11
PUSPARAGAM

Dalam bagian ini kami ingin memperkenalkan beberapa topik yang menarik bagi para pengguna uji-uji kai-kuadrat untuk memeriksa ketidaktergantungan serta homogenitas, karena penguasaan teknik-teknik yang akan dibahas di sini akan sangat meningkatkan potensi analitik mereka. Kendatipun kami tidak dapat menyajikan pembahasan yang mendalam mengenai topik-topik ini, di akhir bab ini Anda akan menjumpai daftar acuan yang mudah-mudahan dapat membantu Anda dalam mempelajari masing-masing topik secara lebih terinci.

Penyekatan kai-kuadrat Manakala suatu uji ketidaktergantungan atau uji homogenitas kai-kuadrat dilakukan terhadap data dari tabel kontingensi 2 x 2, penafsiran terhadap hasil-hasilnya tidak berbelit-belit dan tidak bermakna ganda. Dalam uji ketidaktergantungan, yang kita simpulkan adalah entah bahwa kedua variabel yang dikaji mungkin bebas atau bahwa keduanya berhubungan. Dalam uji homogenitas, yang kita simpulkan adalah entah bahwa kedua populasi asal sampel mungkin homogen atau bahwa keduanya tidak homogen, bergantung pada apakah kita menolak hipotesis nol atau tidak. Apabila kita melakukan uji ketidaktergantungan dan uji homogenitas kai-kuadrat terhadap data dari tabel kontingensi dengan derajat bebas lebih dari satu, penafsiran hasil-hasilnya relatif kurang tegas. Kalau kita menolak hipotesis nol yang menyatakan ketidaktergantungan, misalnya, kita tidak tahu apakah kurangnya ketidaktergantungan berlaku untuk seluruh tabel atau hanya di antara kategori-kategori tertentu dari kedua variabel yang diminati. Kalau kita tidak menolak hipotesis tentang ketidaktergantungan, kita tidak tahu apakah ketidaktergantungan berlaku secara menyeluruh untuk semua kategori atau hanya di antara kategori-kategori tertentu, sebab ketidaktergantungan di antara beberapa kategori dapat menutupi atau membaurkan ketergantungan di antara yang lain. Alhasil, nilai yang dihitung untuk keseluruhan tidak nyata.

Kita dapat menanggulangi kesulitan di atas menggunakan suatu teknik yang dikenal sebagai *teknik penyekatan kai-kuadrat* atau *partitioning of chi-square*. Pada dasarnya, prosedur ini merupakan pemecahan tabel kontingensi r x c yang besar menjadi tabel-tabel yang lebih kecil. Selanjutnya kita menganalisis tabel-tabel yang lebih kecil ini secara terpisah guna menyelidiki apakah tabel besar kita berisi daerah-daerah dengan hipotesis tentang ketidaktergantungan atau hipotesis tentang homogenitas yang dapat kita tolak.

Karya kepeloporan di bidang ini antara lain dilakukan oleh Irwin (T – 21). Kastenbaum (T – 22), Kimball (T –

23), dan Lancaster (T – 24, T – 25). Metode-metode mereka telah dibahas dan dijelaskan oleh Castellan (T – 26) dan Maxwell (T – 27). Makalah yang lebih baru mengenai masalah ini telah ditulis oleh Bresnahan dan Shapiro (T – 28), dan Shaffer (T – 29). Brunden (T – 30), membahas sebuah metode penyekatan untuk suatu tabel 2 x *C* yang memungkinkan pembandingan suatu kelompok *placebo* dengan semua kelompok terapi lain melalui analisis terhadap tabel-tabel 2 x 2 yang dipilih secara tepat.

Tabel-tabel kontingensi multidimensi Sejauh ini, kita selalu berurusan dengan analisis terhadap data yang dapat diperagakan dalam tabel kontingensi dua-arah *(two-way contingency table).* Tetapi tidak mustahil ada eksperimen atau survei yang datanya dapat ditampilkan secara bermakna dalam suatu tabel kontingensi dengan dimensi lebih tinggi. Hoyt dkk. (T – 31), misalnya, menganalisis sebuah tabel kontingensi empat-arah yang mengklasifikasikan 13 968 lulusan sekolah lanjutan atas berdasarkan (a) posisi dalam salah satu dari ketiga kelompok peringkat di antara para lulusan, (b) status mereka pada suatu waktu tertentu setelah lulus (umpama: kuliah di perguruan tinggi, belajar di lembaga pendidikan kejuruan, bekerja purna waktu, dan lain-lain), (c) jenis kelamin, dan (d) status pekerjaan orangtua. Judul-judul tepi *(marginal headings)* untuk tabel mereka tampak dalam Tabel 5.34. Tabel itu sendiri terdiri atas 2 x 7 x 4 x 3 = 168 buah sel untuk memuat frekuensi-frekuensi yang berkaitan dengan subkelompok-subkelompok yang telah ditetapkan bagi ke-13 968 subjek.

Tabel 11.1 Judul-judul tepi pada tabel kontingensi yang mengklasifikasikan 13 968 lulusan SMA

		Peringkat di SMA													
		Sepertigaan bawah					**Sepertigaan tengah**					**Sepertigaan atas**			
Status selepas SMA:		**1**	**2**	**3**	**4**		**1**	**2**	**3**	**4**		**1**	**2**	**3**	**4**
Jenis kelamin (1)	1														
Status	2														
Pekerjaan ayah	3														
	4														
	5														
	6														
	7														
Jenis kelamin (2)	1														
Status	2														
Pekerjaan ayah	3														
	4														
	5														
	6														
	7														

Lewis (T – 32) menulis sebuah ulasan umum tentang metode-metode untuk menganalisis tabel-tabel multidimensi, berikut sejumlah prosedur yang menurut pendapatnya memiliki perhitungan paling sederhana. Artikel-artikel lain yang menarik adalah yang ditulis oleh Lancaster (T – 33), Plackett (T – 34), Shaffer (T – 35), Smith (T – 36), Sutcliffe (T – 37), dan Goodman (T – 38, T – 39), (T – 40, T – 41). Daftar rujukan yang lain bisa ditemui dalam Lancaster (T – 4).

Penggabungan tabel-tabel kontingensi Data tabel kontingensi yang berkaitan dengan riset yang sama mungkin kadang-kadang tersedia. Dalam hal ini, yang patut dipertanyakan adalah apakah dan bagaimana sejumlah bukti yang terpisah-pisah dapat digabungkan untuk menguji hipotesis tentang ketidaktergantungan atau homogenitas di antara baris-baris dan kolom-kolom.

Armitage (T – 42) dan Cochran (T – 6) membahas beberapa metode untuk menggabungkan tabel-tabel kontingensi yang sebelumnya pernah digunakan. Cochran (T – 6), yang melihat sejumlah kekurangan pada beberapa dari prosedur itu, menyarankan sebuah metode yang menggunakan pembobotan dalam proses penggabungan (pooling). Dalam artikelnya (T – 43), Radhakrishna menguraikan metode Cochran secara panjang lebar. Catatan yang diberikan oleh Nelson (T – 44) agaknya juga menarik.

Pengujian kecenderungan dalam tabel-tabel kontingensi Apabila ke-*c* kategori klasifikasi pada suatu tabel kontingensi 2 x *c* bisa diurutkan secara wajar, seperti bila kategori-kategori tersebut adalah kelompok-kelompok umur, komponen kai-kuadratnya yang memiliki kecenderungan linier dapat diisolasi dan diuji kebermaknaannya. Armitage (T – 42, T – 45) dan Cochran (T – 6) membahas prosedur ini secara rinci. Smith (T – 46) menguraikan prosedur ini menggunakan data dari sebuah studi tentang selera konsumen. Melalui uji kai-kuadrat yang menyeluruh, hipotesis nol yang menyatakan bahwa selera terhadap suatu produk tertentu dan usia konsumen bebas tidak dapat ditolak pada taraf nyata 0.10. Ternyata, dengan menerapkan uji untuk memeriksa kecenderungan ini, ia dapat menyimpulkan, pada taraf nyata 0.01, bahwa proporsi konsumen yang cenderung memilih produk di atas bertambah sesuai dengan bertambahnya usia. Steel dan Torrie (T – 47) juga menyajikan contoh angka untuk menerangkan teknik ini. Pemilihan cara pengelompokan yang tepat untuk pengujian kecenderungan dalam konteks ini dibahas oleh Connor (T – 48).

Uji-uji statistik simultan menggunakan data kategorik Dalam survei pengumpulan sampel yang lazim, tanggapan-tanggapan terhadap sejumlah pertanyaan yang beragam biasanya dirangkum dalam beberapa tabel kontingensi dua-arah. Apabila kita melakukan uji kai-kuadrat terhadap masing-masing tabel dan memperbandingkan nilai-nilai X^2 hasil perhitungan dengan nilai-nilai dari tabel distribusi kai-kuadrat seperti biasa guna mengetahui taraf nyata masalah penafsiran akan timbul karena uji-uji yang beragam ini umumnya tidak bebas – semua uji ini melibatkan subjek-subjek yang sama. Jensen dkk. (T – 49) telah memperkenalkan sebuah prosedur untuk melaksanakan beberapa uji kai-kuadrat secara simultan kendatipun di antara statistik-statistik ujinya terhadap korelasi. Penulis-penulis ini juga menyediakan tabel-tabel khusus yang bisa memudahkan penerapan prosedur yang disarankan itu.

Penentuan ukuran sampel Ketika suatu teknik statistik hendak digunakan, ukuran sampelnya merupakan bahan pertimbangan yang penting. Ini pula yang berlaku bila analisis kai-kuadrat akan digunakan. Holt dkk. (T – 50) memberi kita sebuah rumus untuk menduga ukuran sampel yang dibutuhkan untuk suatu tabel kontingensi 2 x 2. Dalam makalahnya yang lain, Holt (T – 51) menguraikan perunutan sebuah rumus penduga ukuran sampel untuk suatu tabel *r* x *c*.

Kuasa Sejumlah makalah yang membahas kuasa uji-uji kai-kuadrat antara lain ditulis oleh Diamond (T – 52), Harkness dan Katz (T – 53), Meng dan Chapman (T – 54), serta Mitra (T – 55). Kuasa uji-uji kai-kuadrat untuk memeriksa kecenderungan linier dalam proporsi-proporsi telah dibahas oleh Chapman (T – 56).

Penggunaan analisis kai-kuadrat secara salah Karena analisis kai-kuadrat merupakan salah satu teknik statistik yang digunakan secara luas, tidak mengherankan bila teknik ini juga digunakan secara salah. Lewis dan Burke (T – 57, T – 58) membahas beberapa contoh yang menurut mereka merupakan penyalahgunaan teknik analisis kai-kuadrat. Makalah mereka ini telah mengundang komentar-komentar tambahan oleh Peters (T – 59), Pastore (T – 60), Edwards (T – 61), dan Burke (T – 62). Kumpulan makalah yang lengket tentang ini terdapat dalam sebuah buku yang disunting oleh Steger (T – 63). Analisis kai-kuadrat yang diduga keras telah digunakan secara salah juga diulas oleh Rich dkk. (T – 64) dan Wright (T – 65).

Komentar-komentar tambahan Dalam pembahasan oleh Feinstein dan Ranshaw (T – 66) Anda mungkin tertarik pada sebuah prosedur untuk melaksanakan perhitungan uji kai-kuadrat 2 x 2 secara cepat. Smith (T – 67) memperkenalkan sebuah nomogram untuk menghitung statistik kai-kuadrat yang mungkin bermanfaat dalam penggunaan di lapangan atau bila hasil-hasil analisis harus segera diketahui. Russell (T – 68) telah menulis sebuah program komputer yang mampu menyusun frekuensi-frekuensi dan membentuk tabel-tabel kontingensi dari daftar pertanyaan dan data lain yang tersimpan pada memori komputer. Hill dan Pike (T – 69) menyajikan sebuah algoritma untuk menentukan peluang bahwa X^2 untuk derajat bebas tertentu melebihi nilai yang telah ditetapkan. Sebuah algoritma lain yang diajukan oleh Goldstein (T – 70) dimaksudkan untuk mengevaluasi kuantil pada suatu tingkat probabilitas tertentu untuk distribusi kai-kuadrat dengan derajat bebas yang ditetapkan. Masalah optimalisasi penentuan klasifikasi untuk tabel-tabel kontingensi telah dibahas oleh Hamdan (T – 71). Chernoff (T – 72) menulis tentang derajat bebas untuk kai-kuadrat.

DAFTAR PUSTAKA

1 Dunn, Olive J., *Basic Statistics: A Primer for the Biomedical Sciences,* New York : Wiley, 1964.

2 Remington, Richard D., and M. Anthony Schork, *Statistics with Applications to the Biological and Health Sciences,* Englewood Cliffs, N.J. : Prentice-Hall, 1970.

3 Armitage, P., *Statistical Methods in Medical Research,* Oxford and Edinburgh : Blackwell Scientific Publications, 1971.

4 Colton, Theodore, *Statistics in Medicine*, Boston : Little, Brown, 1974.

5 Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes, *Introduction to the Theory of Statistics*, third edition, New York : McGraw-Hill, 1974.

6 Gibbons, Jean D., and John W. Pratt, "*P*-Values : Interpretation and Methodology," *Amer. Statist*., 29 (1975), 20-25.

7 Hodges, J. L., Jr., and E. L. Lehmann, *Basic Concepts of Probability and Statistics*. second edition, San Francisco : Holden-Day, 1970.

T – 1 Pearson, Karl, "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *The London, Edinburgh and Dublin Philosoph. Mag. J. Sci.* (5th Series), 50 (1900), 157 – 175. Reprinted in Karl Pearson's Early Statistical Papers, London : Cambridge University Press, 1948.

T – 2 Wilks, Samuel S., "Karl Pearson : Founder of the Science of Statistics," *Scientific Monthly*, 53 (1941), 249 – 253.

T – 3 Pearson, Karl, "Mathematical Contributions to the Theory of Evolution. III, Regression, Heredity, and Panmixia," *Philosoph. Trans. Roy. Soc. A*, 187 (1896), 253 – 318.

T – 4 Lancaster, H, O., *The Chi-Squared Distribution*, New York : Wiley, 1969.

T – 5 Cochran, W. G., "The X^2 Test of Goodness of Fit," *Ann. Math. Statist.*, 23 (1952), 315 – 345.

T – 6 Cochran, W. G., "Some Methods for Strengthening the Common X^2 Tests, *Biometrics*, 10 (1954), 417 – 451.

T – 7 Katti, S. K., and A. N. Sastry, “Biological Examples of Small Expected Frequencies and the Chi-Square Test,” *Biometrics,* 21 (1965), 49 – 54.