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 disiplin 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. Kendatipun 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, dalam Bagian 1.4, kita akan membicarakan skala-skala pengukuran. Bagian 1.5 mengantar kita ke konsep-konsep dasar statistika *nonparametrik*, yang tidak lain merupakan pokok pembahasan buku ini. Dalam Bagian 1.6 kita diajak meninjau ke depan, secara sekilas, ke Bab

2 sampai dengan Bab 10, sedangkan Bagian 1.7 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 melakukan pengamatan tentang mahasiswa-mahasiswa di suatu perguruan tinggi tertentu saja. Kedua peneliti itu sama-sama menganggap bahwa bagi mereka populasi adalah kumpulan mahasiswa yang akan menjadi obyek penyelidikan mereka.

Dalam beberapa konteks tertentu, 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 atau 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 mahasiwa 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 Inggris bisa dipandang 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 tak terhingga ini.

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 kita harus meneliti setiap anggota populasi termaksud untuk mengumpulkan informasi yang akan mendasari penarikan 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 tabel-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 desa, 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)* ρ. 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 seorang 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 riil/*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.

1.2 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 saling 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 hampir 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:

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

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

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

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

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

Guna menguji hipotesis itu, peneliti memilih *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

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

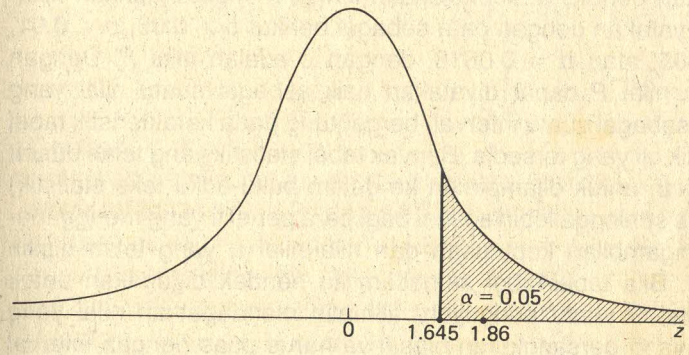
Di sini $\bar{x}_1$ dan $\bar{x}_2$ adalah nilai-nilai rata-rata sampel yang dihitung dari sampel-sampel berukuran n_1 dan n_2, yang berturut-turut ditarik dari populasi 1 dan populasi 2, D_0 adalah selisih hipotesis *(hypothesized difference)* antara nilai-nilai rata-rata populasi, dan s_p^2 diperoleh 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 atau 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 1.1 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

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

Andaikan pula bahwa taraf nyata $\alpha = 0.05$, bahwa statistik uji memiliki distribusi normal standar, dan bahwa nilai z hasil perhitungan adalah 1.86. Apabila kita mengacu ke Tabel A.2, kita melihat bahwa nilai kritis statistik uji untuk $\alpha = 0.05$ dengan uji satu sisi adalah 1.645. Gambar 1.1 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 level, 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 diyatakan 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 1.1

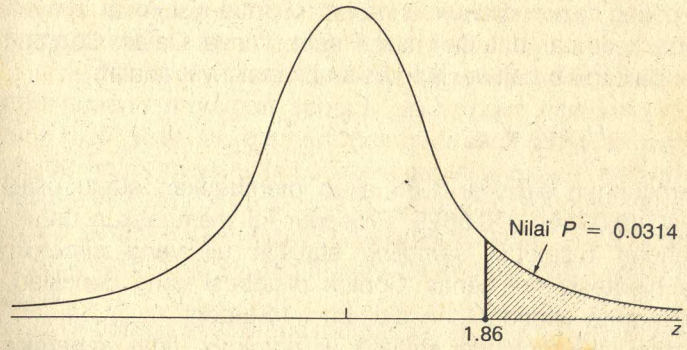
Andaikan hipotesis nol dan hipotesis tandingannya adalah

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

Selanjutnya andaikan pula bahwa statistik uji yang tepat adalah z dan bahwa berdasarkan perhitungan nilai $z = 1.86$. Kita mengacu ke Tabel A.2 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 1.2.

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

Gambar 1.2 Harga statistik uji z hasil perhitungan (1.86) dan harga *P*-nya untuk suatu uji hipotesis satu-sisi

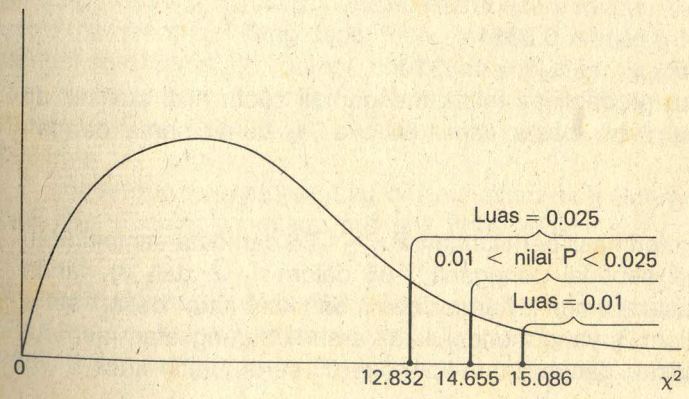


Contoh 1.2

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, 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 $X^2_{0.995} = 16.750$, kita bisa membuat pernyataan begini: nilai $P < 0.005$. Contoh ini dijelaskan dalam Gambar 1.3.

Gambar 1.3 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* satu-sisinya. Dalam Contoh 1.1, misalnya, andaikanlah bahwa hipotesis-hipotesisnya adalah

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

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 standar dan 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

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

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 $1 - \beta$, dengan β 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 β 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 matematik 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.

1.3 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 $(1 - \alpha)$% percaya bahwa interval tunggal yang kita buat mengandung parameter yang sedang kita duga. Dalam kedua interpretasi itu, koefisien kepercayaan yang digunakan adalah 100$(1 - \alpha)$%.

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

$$P(L_0 < \theta < U_0) = 1 - \alpha \qquad \textbf{(1.1)}$$

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

$$C(L < \theta < U) = 1 - \alpha \qquad \textbf{(1.2)}$$

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. Coba perhatikan hipotesis-hipotesis

$$H_0 : \theta = \theta_0, \qquad H_1 : \theta \neq \theta_0$$

dengan taraf nyata α. Nilai-nilai parameter θ yang mungkin, yang terdapat dalam 100 $(1 - \alpha)$% dari interval kepercayaan $L < \theta < U$ adalah nilai-nilai yang berkesesuaian dengan hipotesis nol. Nilai-nilai θ 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 $(1 - \alpha)$% dari interval kepercayaan tidak mengandung nilai parameter θ_0 yang dihipotesiskan, kita menolak H_0 pada taraf nyata α. Jika θ_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).

1.4 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). Jadi kita boleh mengklasifikasikan (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 ini sering 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-peristiwa 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 titik 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 katakan 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).

1.5 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 ringkas, 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 nonparametrik 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), Quenoille (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.

1.6 CAKUPAN BUKU INI

Tekanan dalam buku ini adalah pada penerapan metode-metode statistik nonparametrik. Di mana saja tersedia, contoh-contoh dan latihan-latihan di sini 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.

Berikut ini kami menyajikan ringkasan pokok-pokok pembahasan dalam bab-bab mendatang.

Bab 2 *Prosedur-prosedur yang menggunakan data dari sebuah sampel tunggal* Dalam bab ini, kita membahas prosedur-prosedur untuk menduga dan menguji hipotesis tentang parameter-parameter lokasi bila kita berminat terhadap karakteristik-karakteristik suatu populasi tunggal. Di sini kita juga membahas sebuah uji keacakan dan sebuah uji kecenderungan *(test for trend)*.

Bab 3 *Prosedur-prosedur yang menggunakan data dari dua sampel bebas* Bila data yang tersedia terdiri atas sampel-sampel bebas yang masing-masing berasal dari dua populasi yang berbeda, prosedur-prosedur dalam bab ini sangat tepat. Di sini kita membahas prosedur-

prosedur untuk menguji kesamaan di antara parameter-parameter lokasi dengan kecanggihan yang merentang dari uji kilat Tukey *(Tukey's quick test)* hingga uji Mann-Whitney, yang dipandang sebagai uji-uji nonparametrik yang cukup berdaya guna untuk data yang berasal dari dua sampel. Kita juga membahas teknik-teknik untuk menduga beda antara dua parameter populasi, serta uji-uji untuk memeriksa kesamaan di antara dua parameter penyebaran *(dispersion parameter)*. Sebagai tambahan, kami mengetengahkan sebuah uji *run,* sebuah uji untuk memeriksa reaksi-reaksi ekstrem, dan uji eksak Fisher.

Bab 4 *Prosedur-prosedur yang menggunakan data dari dua sampel berhubungan* Prosedur-prosedur yang disajikan di sini dapat diterapkan bila data terdiri atas dua sampel yang berhubungan secara khusus. Data hasil pengamatan mungkin berupa pengukuran-pengukuran terhadap subjek-subjek yang sama, yakni sebelum dan sesudah suatu perlakuan tertentu atau berupa pengukuran-pengukuran terhadap subjek-subjek yang berbeda, yang telah dipadankan menurut suatu atau beberapa kriteria. Kita juga membahas prosedur-prosedur untuk menduga interval serta sebuah uji untuk digunakan bila data yang dihadapi berupa frekuensi-frekuensi.

Bab 5 *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 6 *Prosedur-prosedur yang menggunakan data dari tiga sampel bebas atau lebih* Bab ini membahas prosedur-prosedur yang merupakan bentuk-bentuk analog nonparametrik atas analisis varians satu-arah parametrik. Bab ini juga membahas sebuah prosedur pembandingan-

berganda serta sebuah prosedur yang digunakan bila *urutan besaran* parameter-parameter lokasinya ditetapkan dalam hipotesis tandingan.

Bab 7 *Prosedur-prosedur yang menggunakan data dari tiga sampel berhubungan atau lebih* Bab ini meliputi uji-uji yang digunakan bila seseorang menghendaki suatu uji nonparametrik yang analog dengan analisis varians untuk suatu rancangan blok teracak *(randomized block design)*. Bab ini juga membahas kasus blok-blok yang tidak lengkap serta sebuah uji yang digunakan bila hipotesis tandingannya berurut.

Bab 8 *Uji-uji keselarasan (goodness-of-fit tests)* Dalam bab ini, kita membahas uji-uji keselarasan yang paling sering kita gunakan—yakni, uji kai-kuadrat dan uji Kolmogorov-Smirnov. Kita juga membahas sebuah prosedur pembuatan sekumpulan keyakinan untuk fungsi distribusi suatu populasi.

Bab 9 *Korelasi peringkat dan ukuran-ukuran asosiasi yang lain* Bab ini membahas uji-uji asosiasi yang paling sering digunakan.

Bab 10 *Analisis regresi linier sederhana* Bab ini mengetengahkan beberapa prosedur nonparametrik yang dapat digunakan dengan model regresi linier sederhana. Kita membahas uji-uji dan prosedur-prosedur pembentukan interval kepercayaan guna menentukan koefisien kemiringan *(slope coefficient)* dan parameter intersepsi. Ke dalam cakupan pembahasan kami juga telah menyertakan sebuah uji untuk memeriksa kesejajaran antara dua garis regresi serta prosedur pembentukan interval kepercayaan untuk mendapatkan beda antara dua parameter kemiringan.

1.7 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. Akhirnya, untuk setiap prosedur kami juga menyediakan latihan-latihan. Latihan-latihan ini diadakan dengan dua tujuan: Pertama, untuk menjelaskan penggunaan-penggunaan suatu uji secara tepat, dan kedua, sebagai sarana untuk menentukan apakah Anda telah menguasai teknik-teknik perhitungan yang diberikan, cara menyusun hipotesis, dan cara menggunakan kaidah pengambilan keputusan yang sesuai.

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.

Daftar acuan

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.
8 Kempthorne, Oscar, and Leroy Folks, *Probability, Statistics, and Data Analysis*, Ames, Iowa: Iowa State University Press, 1971.
9 Lindgren, Bernard W., *Statistical Theory*, third edition, New York: Macmillan, 1976.
10 Bahn, Anita K., "*P* and the Null Hypothesis," *Ann. Intern. Med.*, 76 (1972), 674.
11 Bakan, David, *On Method*, San Francisco: Jossey-Bass, 1967.
12 Brewer, James K., Letter to the Editor, *Amer. Statist.*, 29 (1975), 171.
13 Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, New York: Academic Press, 1969.
14 Duggan, T. J., and C. W. Dean, "Common Misinterpretation of Significance Levels in Sociology Journals," *Amer. Sociol.*, 3 (1968), 45–46.
15 Gold, David, "Statistical Tests and Substantive Significance," *Amer. Sociol.*, 4 (1969), 42–46.
16 Kish, Leslie, "Some Statistical Problems in Research Design," *Amer. Sociol. Rev.*, 24 (1959), 328–338.
17 McGinnis, R., "Randomization and Inference in Sociological Research," *Amer. Sociol. Rev.*, 23 (1958), 408–414.
18 Pitman, E. J. G., *Lecture Notes on Nonparametric Statistical Inference*. Columbia University, spring 1948, cited in Capon, Jack, "Asymptotic Efficiency of Certain Locally Most Powerful Rank Tests," *Ann. Math. Statist.*, 32 (1961), 88–100.
19 Noether, G. E., "On a Theorem of Pitman," *Ann. Math. Statist.*, 26 (1955), 64–68.
20 Stuart, A., "The Asymptotic Relative Efficiency of Tests and the Derivatives of Their Power Functions," *Skandinavisk Aktuarietidskrift*, 37 (1954), 163–169.
21 Stuart, A., "The Measurement of Estimation and Test Efficiency," *Bull. Int. Statist. Inst.*, Part III, 36 (1956), 79–86.
22 Lehmann, E. L., *Testing Statistical Hypotheses*, New York: Wiley, 1959.
23 Fraser, D. A. S., *Nonparametric Methods in Statistics*, New York: Wiley, 1957.
24 Blyth, C. R., "Note on Relative Efficiency of Tests," *Ann. Math. Statist.*, 29 (1958), 898–903.
25 Smith, K., "Distribution-Free Statistical Methods and the Concept of Power Efficiency," in L. Festinger and D. Katz (eds.), *Research Methods*

in the Behavioral Sciences, New York: Dryden, 1953, pp. 536–577.
26 Wilson, Warren R., and Howard Miller, "A Note on the Inconclusiveness of Accepting the Null Hypothesis," *Psychol. Rev.*, 71 (1964), 238–242.
27 Edwards, Ward, "Tactical Note on the Relation between Scientific and Statistical Hypotheses," *Psychol. Bull.*, 63 (1965), 400–402.
28 Feinberg, William E., "Teaching the Type I and Type II Errors: The Judicial Process," *Amer. Statist.*, 25 (June 1971), 30–32.
29 Rodger, R. S., "Type I Errors and Their Decision Basis," *Br. J. Math. Statist. Psychol.*, 20 (1967), 51–62.
30 Rodger, R. S., "Type II Errors and Their Decision Basis," *Br. J. Math. Statist. Psychol.*, 20 (1967), 187–204.
31 Edgington, Eugene S., "Statistical Inference and Nonrandom Samples," *Psychol. Bull.*, 66 (1966), 485–487.
32 Ungerleider, Harry E., and Courtland C. Smith, "Use and Abuse of Statistics," *Geriatrics*, 22 (February 1967), 112–120.
33 Brewer, James K., and Ruth Dailey Knowles, "Some Statistical Considerations in Nursing Research," *Nursing Res.* 23 (1974), 68–70.
34 Ahrens, S. J., "Statistical Tests of Significance: Truth, Paradox, or Folly?" *Res. Quart. Am. Assoc. Health, Phys. Educ. Rec.*, 42 (1971), 436–440.
35 Barnard, G. A., "The Meaning of a Significance Level," *Biometrika*, 34 (1947), 179–182.
36 Chandler, Robert E., "The Statistical Concepts of Confidence and Significance," *Psychol. Bull.*, 54 (1957), 429–430.
37 Labovitz, Sanford, "Criteria for Selecting a Significance Level: A Note on the Sacredness of 0.05," *Amer. Sociol.*, 3 (1968), 220–222.
38 Lykken, David T., "Statistical Significance in Psychological Research," *Psychol. Bull.*, 70 (1968), 151–159.
39 Krause, Merton S., "Insignificant Differences and Null Explanations," *J. Gen. Psychol.*, 86 (1972), 217–220.
40 Morrison, D. E., and R. E. Henkel, *The Significance Test Controversy—A Reader*, Chicago: Aldine, 1970.
41 O'Brien, Thomas C., and Bernard J. Shapiro, "Statistical Significance—What?" *Math. Teacher*, 61 (1968), 673–676.
42 Rozeboom, W. W., "The Fallacy of the Null Hypothesis Significance Test," *Psychol. Bull.*, 57 (1960), 416–428.
43 Selvin, H. C., "A Critique of Tests of Significance in Survey Research," *Amer. Sociol. Rev.*, 22 (1957), 519–527.
44 Skipper, James K., Jr., Anthony L. Guenther, and Gilbert Nass, "The Sacredness of .05: A Note Concerning the Uses of Statistical Levels of Significance in Social Science," *Amer. Sociol.*, 2 (1967), 16–18.

45 Stone, M., "Role of Significance Testing—Some Data With a Message," *Biometrika*, 56 (1969), 485–493.
46 Winch, R. F., and D. T. Campbell, "Proof? No. Evidence? Yes. The Significance of Tests of Significance," *Amer. Sociol.*, 4 (May 1969), 140–143.
47 Zeisel, H., "The Significance of Insignificant Differences," *Public Opinion Quart.*, 19 (1955), 319–321.
48 Editorial. "Significance of Significant," *N. Eng. J. Med.*, 278 (1968), 1232.
49 Bross, Irwin D. J., "The Role of the Statistician: Scientist or Shoe Clerk," *Amer. Statist.*, 28 (1974), 126–127.
50 Godambe, V. P., and D. A. Sprott, *Foundations of Statistical Inference, A Symposium*, Minneapolis: Winston Press, 1972.
51 Kiefer, J., "Statistical Inference," *Math. Spectrum*, 3 (Fall 1970), 1–11.
52 Lurie, William, "The Impertinent Questioner: The Scientist's Guide to the Statistician's Mind," *Amer. Scientist*, 46 (1958), 57–61.
53 Natrella, Mary G., "The Relation between Confidence Intervals and Tests of Significance," *Amer. Statist.*, 14 (February 1960), 20–22, 38.
54 Kirk, Roger E. (ed.), *Statistical Issues: A Reader for the Behavioral Sciences*, Monterey: Brooks Cole, 1972.
55 Stevens, S. S., "On the Theory of Scales of Measurement," *Science*, 103 (1946), 677–680.
56 Stevens, S. S., "Mathematics, Measurement and Psychophysics," in S. S. Stevens (ed.), *Handbook of Experimental Psychology*, New York: Wiley, 1951.
57 Stevens, S. S., "Measurement Statistics, and the Schemapiric View," *Science*, 161 (1968), 849–856.
58 Anderson, Norman H., "Scales and Statistics: Parametric and Nonparametric," *Psychol. Bull.*, 58 (1961), 305–316.
59 Boneau, C. Alan, "A Note on Measurement Scales and Statistical Tests," *Amer. Psychol.*, 16 (1961), 260–261.
60 Gaito, John, "Scale Classification and Statistics," *Psychol. Rev.*, 67 (1960), 277–278.
61 Lord, F. M., "On the Statistical Treatment of Football Numbers," *Amer. Psychol.*, 8 (1953), 750–751.
62 Baker, Bela O., Curtis D. Hardyck, and Lewis F. Petrinovich, "Weak Measurement vs. Strong Statistics: An Empirical Critique of S. S. Stevens' Proscriptions on Statistics," *Educ. Psychol. Measurement*, 26 (1966), 291–309.
63 Campbell, N. R., "Symposium: Measurement and Its Importance for Philosophy," *Proc. Aristotelian Soc.*, Suppl., 17 (1938), London: Harrison and Sons.
64 Gardner, Paul Leslie, "Scales and Statistics," *Rev. Educ. Res.*, 45 (Winter 1975), 43–57.
65 Kendall, M. G., and R. M. Sundrum, "Distribution-Free Methods and Order Properties," *Rev. Int. Statist. Inst.* 21 (1953), 124–134.
66 Savage, I. R., *Bibliography of Nonparametric Statistics*, Cambridge, Mass.: Harvard University Press, 1962.
67 Bradley, James V., *Distribution-Free Statistical Tests*, Englewood Cliffs, N. J.: Prentice-Hall, 1968.
68 Conover, W. J., *Practical Nonparametric Statistics*, New York: Wiley, 1971.
69 Gibbons, Jean Dickinson, *Nonparametric Methods for Quantitative Analysis*, New York: Holt, Rinehart, and Winston, 1976.
70 Hollander, Myles, and Douglas A. Wolfe, *Nonparametric Statistical Methods*, New York: Wiley, 1973.
71 Lehmann, E. L., *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day, 1975.
72 Maxwell, A. E., *Analyzing Quantitative Data*, New York: Wiley, 1961.
73 Mosteller, F., and R. E. K. Rourke, *Sturdy Statistics*, Reading, Mass.: Addison–Wesley, 1973.
74 Noether, G., *Introduction to Statistics: A Fresh Approach*, Boston: Houghton Mifflin, 1971.
75 Pierce, Albert, *Fundamentals of Nonparametric Statistics*, Belmont, California: Dickenson, 1970.
76 Quenouille, M. H., *Rapid Statistical Calculations. A Collection of Distribution-Free and Easy Methods of Estimation and Testing*, second edition, London: Griffin, 1972.
77 Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956.
78 Senders, V. L., *Measurement and Statistics*, New York: Oxford University Press, 1958.
79 Tate, Merle W., and Richard C. Clelland, *Nonparametric and Shortcut Statistics in the Social, Biological, and Medical Sciences*, Danville, Ill.: Interstate, 1957.
80 Wilcoxon, Frank, and Roberta A. Wilcox, *Some Rapid Approximate Statistical Procedures*, revised, Pearl River, N. Y.: Lederle Laboratories, 1964.
81 David, H. A., *Order Statistics*, New York: Wiley, 1970.
82 Edgington, Eugene S., *Statistical Inference: The Distribution-Free Approach*, New York: McGraw-Hill, 1969.

83 Fraser, D. A. S., *Nonparametric Methods in Statistics*, New York: Wiley, 1957.
84 Gibbons, Jean Dickenson, *Nonparametric Statistical Inference*, New York: McGraw-Hill, 1971.
85 Hájek, Jeroslav, *A Course in Nonparametric Statistics*, San Francisco: Holden-Day, 1969.
86 Hájek, J., and Šidák, Z., *Theory of Rank Tests*, New York: Academic Press, 1967.
87 Kraft, Charles H., and Constance van Eeden, *A Nonparametric Introduction to Statistics*, New York: Macmillan, 1968.
88 Noether, Gottfried E., *Elements of Nonparametric Statistics*, New York: Wiley, 1967.
89 Puri, Mandan Lal (ed.), *Nonparametric Techniques in Statistical Inference*, Cambridge: Cambridge University Press, 1970.
90 Sarhan, S. E., and B. G. Greenberg (ed.), *Contributions to Order Statistics*, New York: Wiley, 1962.
91 Walsh, John E., *Handbook of Nonparametric Statistics*, Princeton, N. J.; D. van Nostrand, 1962.
92 Walsh, John E., *Handbook of Nonparametric Statistics*. II, Princeton, N. J.; D. van Nostrand, 1965.
93 Walsh, John E., *Handbook of Nonparametric Statistics*. III, Princeton, N. J.: D. van Nostrand, 1968.
94 Kendall, M. G., *Rank Correlation Methods*, fourth edition, New York: Hafner, 1970.
95 Lancaster, H. O., *The Chi-Squared Distribution*, New York: Wiley, 1969.
96 Ruist, E., "Comparison of Tests for Nonparametric Hypotheses," *Ark. Matematik* 3 (1955), 133–163.
97 Hettmansperger, Thomas P., and Joseph W. McKean, "A Graphical Representation for Nonparametric Inference," *Amer. Statist.*, 28 (1974), 100–102.
98 Buchanan, William, "Nominal and Ordinal Bivariate Statistics: The Practitioner's View," *Amer. J. Polit. Sci.* 18 (1974), 625–646.
99 Noether, Gottfried E., "The Nonparametric Approach in Elementary Statistics," *Math. Teacher*, 67 (1974), 123–126.
100 Scheffé, H., "Statistical Inference in the Nonparametric Case," *Ann. Math. Statist.*, 14 (1943), 305–332.