Bagian III Distribusi Sampling

PERISTILAHAN

Untuk mengambil keputusan — terhadap populasi — pengambil keputusan membutuhkan informasi. Informasi yang lengkap diperoleh dari seluruh populasi. Namun, seringkali, untuk mendapatkan informasi yang lengkap — seperti dimaksud di atas — dibutuhkan banyak biaya dan waktu. Oleh karena itu, sampling merupakan cara yang tepat untuk mengestimasi parameter-parameter populasi. Disebut sampling, karena basis yang digunakan untuk mengestimasi parameter-parameter populasi adalah sampel random (acak).

Ada beberapa istilah yang perlu dipahami agar kita mudah memahami konsep-konsep yang berkaitan dengan sampling. Suatu sampel random — disebut juga sampel probabilitas — adalah suatu sampel yang didapat sedemikian sehingga setiap elemen dari suatu populasi memilki suatu kesempatan berdasarkan pengalaman yang pernah terjadi atau pernah diketaui (tidak nol) berada dalam sampel yang terjadi setelah dilakukan sampling. Statistik merupakan besaran yang dipakai untuk menerangkan beberapa sifat karateristik dari suatu sampel, misalnya: rata-rata hitung (*aritmatic mean*), median, dan deviasi standar (simpangan baku) dari suatu sampel. Populasi merupakan kumpulan secara keseluruhan obyek atau orang yang dipelajari (diteliti), darimana suatu sampel diambil. Sampel merupakan kumpulan obyek atau orang yang mewakili populasi. Parameter adalah suatu yang dipakai untuk menerangkan beberapa sifat karakteristik dari suatu populasi. Suatu estimator yang tidak bias adalah suatu besaran statistik yang memiliki nilai harapan yang sama dengan parameter yang diestimasi (populasi). μ lambang rata-rata populasi dan estimatornya $\bar{x}$. π lambang proporsi populasi dan estimatornya $\bar{p}$. σ lambang deviasi standar populasi dan estimatornya s. Untuk nilai s ini, jika faktor koreksinya belum termasuk di dalamnya, estimator yang sesuai adalah

$$s.\sqrt{\frac{n}{n-1}}$$

Catatan, n: besarnya (banyaknya data dalam) sampel.

Jika besarnya sampel (n) lebih kecil dari besarnya populasi (N) dua sampel atau lebih dapat ditarik dari populasi yang sama. Statistik tertentu dapat dihitung untuk setiap sampel yang mungkin ditarik dari populasi itu. Distribusi dari besaran-besaran statistik yang muncul dari sampel-sampel disebut **Distribusi Sampling Statistik.**

Tabel 3.1
Estimator-estimator yang Lazim Digunakan

Parameter Populasi	Estimator
Rata-rata, μ	X
Perbedaan rata-rata dua populasi, $\mu_1 - \mu_2$	$X_1 - X_2$
Proporsi, π	p
Perbedaan proporsi dalam dua populasi, $\pi_1 - \pi_2$	$p_1 - p_2$
Deviasi standard, σ	s

Deviasi standar dari distribusi sampling suatu statistik sering disebut **Standar Error dari Statistik**. Perbedaan antara deviasi standard dan kesalahan standard (*standar error*) adalah bahwa deviasi standard berkaitan dengan nilai-nilai asli, sementara kesalahan standard berkaitan dengan nilai-nilai hasil hitungan.

Perbedaan antara hasil yang diperoleh dari satu sampel (statistik) dan hasil yang telah diperoleh dari populasi (parameter yang bersangkutan) disebut **Sampling Error (kesalahan sampling)**. Suatu kesalahan sampling biasanya terjadi jika peninjauan menyeluruh terhadap populasi tidak dilakukan, tetapi diambil satu sampel untuk menaksir sifat karakteristik populasi itu. Kesalahan sampling diukur dari kesalahan standar statistik dalam bentuk probabilitas di bawah kurva normal. Hasil perhitungan ini menunjukkan ketepatan populasi yang didasarkan atas studi sampel. Makin kecil kesalahan sampling, makin besar ketepatan penaksiran.

METODE-METODE SAMPLING

Berdasarkan jumlah sampel yang diambil, metode sampling dapat diklasifikasikan menjadi 3 macam metode sampling, yaitu: tunggal, ganda, dan multipel. Dengan metode sampling tunggal kita hanya memerlukan satu sampel dari populasi guna pengambilan kesimpulan statistik. Karena hanya diambil satu sampel, besarnya sampel harus cukup memadai untuk menarik kesimpulan statistik. Metode sampling ganda memberikan kemungkinan untuk mengambil sampel kedua guna mengambil kesimpulan jika sampel pertama yang diambil dari populasi yang sama, tidak mampu memberikan keputusan. Kedua sampel kemudian digabung untuk dianalisa hasilnya. Metode sampling multipel prosedurnya sama dengan metode sampling ganda. Perbedaannya adalah bahwa dalam metode ini pengambilan sampel dapat lebih dari dua kali.

Apabila dalam memilih bagian sampel kita mendasarkan pada cara yang dipakai, metode sampling dapat dibedakan menjadi dua, yaitu: sampling kebijaksanaan (*judgement*) dan sampling pemilihan random atau cara acak. Suatu sampel disebut sampel kebijaksanaan, jika bagian-bagian dari sampel dipilih atas kebijaksanaan seseorang (personal). Suatu sampel dikatakan suatu sampel yang diambil secara acak (random), jika pengambilan (pemilihan) bagian dari sampel dilakukan dengan cara sedemikian sehingga setiap bagian dari populasi memiliki kans (kesempatan) yang sama untuk dipilih. Sampel random sering disebut juga sebagai sampel probabilitas, karena masing-masing elemen (bagian) memiliki kans yang sudah diketahui. Cara ini lebih obyektif dan kesalahan sampling dapat diukur. Bentuk umum dari sampling acak adalah sampling acak sederhana, sampling sistematis, sampling berlapis (*stratified)*, dan sampling kelompok (*cluster*).

Sampel random sederhana dipilih dengan cara sedemikian sehingga setiap sampel yang berukuran sama memiliki suatu probabilitas sama untuk terpilih dari populasi. Sebagai contoh dalam suatu populasi dengan tiga unsur (A, B, C), dimungkinkan untuk menarik tiga sampel yang masing-masing memiliki dua unsur (AB, BC, CA). Jika tiap unsur dari tiga sampel itu mempunyai probabilitas sama (1/3) untuk terpilih, sampel yang terambil adalah sampel acak sederhana.

Suatu **sampel sistematis** diperoleh jika unsur-unsur yang dipilih sesuai dengan aturan-aturan tertentu. Cara pemilihan tergantung pada jumlah unsur yang termasuk pada populasi dan besarnya sampel merupakan jumlah unsur dalam populasi, pertama-tama dibagi dengan jumlah unsur yang diinginkan dalam sampel. Hasil baginya akan menunjukkan apakah setiap unsur kesepuluh, keseratus, atau setiap unsur yang kesejuta dari populasi harus dipilih. Bagian pertama dari sampel dipilih secara acak. Jadi sampel sistematis mungkin saja menghasilkan presisi terhadap estimasi populasi yang sama dengan sampel random sederhana apabila unsur pada populasi sesuai dengan aturan random. Sebagai contoh, misalkan bagian pengajaran suatu sekolah memiliki daftar dari 10.000 siswa menurut urutan abjad. Jika diinginkan suatu sampel usia dari 200 siswa, boleh dipilih usia tiap siswa yang kelima puluh dalam daftar (10.000/200 = 50), karena urutan daftar tak berpengaruh apa-apa kepada usia siswa (daftar menurut abjad). Usia yang pertama dapat diambil secara random dalam daftar dari kelompok lima puluh siswa yang pertama. Jika pengambilan pertama usia siswa yang kedelapan belas (18), pilihan kedua adalah nomor siswa ke-68 (= 18 + 50), ketiga 118, dan seterusnya.

Langkah pertama **sampling random berlapis** adalah membagi populasi dalam kelompok-kelompok yang disebut strata, dengan lebih homogen dari populasi. Kemudian bagian-bagian sampel dipilih dari tiap-tiap stratum secara acak ataupun secara sistematis. Taksiran populasi yang didasarkan atas sampel berlapis ini biasanya memiliki presisi lebih besar (atau kesalahan sampling lebih kecil) dibandingkan dengan apabila keseluruhan populasi diambil sebagai sampel dengan cara acak sederhana. Banyaknya unsur yang dipilih dari setiap stratum boleh sebanding dengan jumlah stratum dalam populasi. Sebagai contoh, jika besarnya stratum A adalah 40 persen dari populasi, perbandingan yang sama digunakan untuk memilih bagian sampel dari tiap stratum. Jadi jika besarnya sampel, misalnya n = 200 unsur, 40 persen dari sampel adalah 80 unsur diambil dari stratum A. Jika pemilihannya tidak sebanding, relatif sulit untuk menimbang hasil strata individual.

Suatu **sampel kelompok** diperoleh dengan membagi populasi atas kelompok-kelompok yang cukup untuk pengambilan sampel, sebagian dari kelompok tersebut dipilih secara random atau sistematis. Akhirnya seluruh bagian-bagian itu diambil secara acak atau secara sistematis dari kelompok terpilih untuk memperoleh sampel. Pada cara ini meskipun tidak semua kelompok diambil sebagai sampel namun setiap kelompok benar-benar memiliki kans yang sama untuk terpilih. Sebagai contoh, jika kurang praktis untuk mengambil sampel acak sederhana dari suatu populasi dengan 10.000 keluarga di sebuah kota, kota dapat dibagi menjadi kawasan-kawasan kecil menurut peta kota. Untuk tiap kawasan ditetapkan satu nomor. Sebagian kawasan dipilih secara acak atau secara sistematis dari nomor-nomor yang mewakilinya. Keluarga yang ada dalam kawasan terpilih diwawancarai. Sampel kelompok semacam ini disebut juga **sampel area**.

Distribusi Sampling Rata-rata

Distribusi sampling rata-rata adalah distribusi probabilitas untuk nilai-nilai yang dapat terjadi dari rata-rata sampel yang didasarkan pada sejumlah sampel tertentu. Apabila jumlah sampel n, rata-rata populasi μ, rata-rata sampel — akan bervariasi dari sampel ke sampel yang lain — nilai harapan $E(\bar{x})$ dan deviasi standar distribusi rata-rata $\sigma_{\bar{x}}$ — lebih dikenal dengan istilah **kesalahan standar rata-rata**, adalah

$E(\bar{X}) = \mu$

dan

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Jika sampling dilakukan pada populasi terbatas, saudara harus menggunakan faktor koreksi terhadap kesalahan standar rata-rata. Akan tetapi jika sampel kecil, yaitu kurang dari 5 persen dari populasi, kesalahan standar rata-rata tidak perlu dikoreksi. Formula kesalahan standar rata-rata dengan faktor koreksi adalah

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

Jika deviasi standar populasi tidak diketahui, kesalahan standar rata-rata didekati dengan kesalahan standar rata-rata sampel, dengan formula

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Jika harus menggunakan faktor koreksi, formula kesalahan standar rata-rata menjadi:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

Apabila sampel besar, yaitu $n \geq 30$, kita senantiasa dapat menggunakan distribusi probabilitas normal dengan kesalahan standar rata-rata.

Contoh 1

Seorang analis keuangan mengambil suatu sampel sebesar 10 persen 300 laporan keuangan dan mendapatkan bahwa rata-rata keuntungan Rp148,50 dengan standard deviasi

Rp35,75. Jika rata-rata keuntungan populasi Rp138,00, berapakah probabilitas bahwa keuntungan yang diperoleh — diambil secara random — Rp148,50 atau lebih?

Jawab:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} = \frac{35{,}75}{\sqrt{30}} \sqrt{\frac{300 - 30}{300 - 1}} = 6{,}20$$

$$Z = \frac{X - \mu}{s_{\bar{x}}} = \frac{148{,}50 - 138{,}00}{6{,}20} = +\,1{,}69$$

Dengan demikian,

$P(\bar{X} \geq 148{,}50) = P(Z \geq +\,1{,}69) = 0{,}50 - 0{,}4545 = 0{,}0455$

Contoh 2

Dengan menggunakan ketentuan dan jawaban **Contoh 1** tentukan jumlah yang keuntungan yang diperoleh — diambil secara random — Rp148,50 atau lebih!

Jawab:

$$\begin{aligned} \text{Jumlah} &= N(P\ X \geq 148{,}50)) \\ &= N(P(Z \geq +\,1{,}69)) \\ &= (300)(0{,}0455) \approx 14 \end{aligned}$$

Konfidensi Interval Rata-rata

Metode-metode estimasi interval dalam bagian ini didasarkan pada asumsi bahwa distribusi probabilitas normal dapat digunakan, dengan ketentuan:

a. $n \geq 30$ atau
b. $n < 30$ dengan syarat distribusi populasi normal dan _ diketahui

Suatu konfidensi interval rata-rata adalah suatu estimasi interval yang disusun sesuai dengan rata-rata sampel yang dapat memuat nilai rata-rata dari populasi yang mungkin terjadi. Tingkat atau koefisien konfidensi yang digabungkan dengan suatu konfidensi interval menunjukkan prosentase dari interval tertentu yang akan memuat parameter yang akan diestimasi.

Jika digunakan distribusi probabilitas normal, konfidensi interval untuk rata-rata ditentukan dengan:

$$X \pm Z\,\sigma_{\bar{x}} \quad \text{atau} \quad X \pm Z\,s_{\bar{x}}$$

Koefisien konfidensi yang amat sering digunakan adalah 90 persen, 95 persen, dan 99 persen.

Banyaknya data dalam sampel yang didasarkan pada penggunaan distribusi normal adalah

$$n = \left(\frac{Z\,\sigma}{E}\right)^2$$

Catatan:

Z : nilai standar normal yang digunakan untuk koefisien konfidensi yang ditentukan

σ : deviasi standar populasi

E : faktor kesalahan *plus dan minus* untuk banyaknya sampel.

Distribusi *t-student* digunakan untuk sampel kecil dan distribusi populasi bersifat normal, tetapi σ tidak diketahui. Jika distribusi populasi bersifat normal, distribusi sampling dari rata-rata untuk banyaknya sampel tertentu juga akan berdistribusi normal. Hal ini benar, baik untuk σ diketahui maupun tidak diketahui.

Jika σ tidak diketahui formula $(\bar{x} - \mu)/s_x$ berbeda untuk masing-masing rata-rata sampel. Dengan demikian nilai Z tidak dapat digunakan, karena tidak akan berdistribusi Z lagi. Sebagai penggantinya nilai-nilai distribusi standar normal akan didistribusikan sesuai dengan distribusi t-student yang juga bersifat normal, dengan *degree of freedom* (d.f.) = n - 1 (untuk kasus sampel tunggal). Konfidensi interval untuk menduga rata-rata populasi jika _ tidak diketahui, n < 30 dan distribusi populasi bersifat normal, adalah

$$X \pm t_{df}\, s_{\bar{x}}$$

Contoh 1

Sebuah Perusahaan ingin mengestimasi rata-rata waktu yang dibutuhkan oleh sebuah mesin untuk memproduksi suatu jenis kertas. Diambil secara random 36 rim kertas, waktu rata-rata yang dibutuhkan untuk memproduksi 1 rim adalah 1,5 menit. Jika diasumsikan standard deviasi populasi 0,30 menit, tentukan estimasi interval rata-rata dengan tingkat konfidensi 95 persen.

Jawab:

$X = 1{,}5, \quad \sigma = 0{,}30 \quad n = 36$

Nilai kesalahan standard dari proses sampling:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}30}{\sqrt{36}} = 0{,}05$$

Dengan tingkat konfidensi 95 persen, nilai Z = 1,96. Dengan demikian estimasi interval rata-rata:

$$\bar{X} - Zs_{\bar{x}} < \mu < \bar{X} + Zsx$$

$$1{,}5 - 1{,}96(0{,}05) < \mu < 1{,}5 + 1{,}96(0{,}05)$$

Contoh 2

Diketahui standard deviasi populasi umur bolam 500 jam, tetapi rata-rata umurnya tidak diketahui. Distribusi umur bolam berbentuk distribusi normal. Diambil sampel sebanyak 15 bolam. Ternyata memiliki umur rata-rata 8900 jam. Tentukan untuk mengestimasi rata-rata populasi dengan konfidensi interval a. 95 persen; b. 90 persen!

Jawab:

a. Estimasi rata-rata populasi dengan 95 persen konfidensi interval:

$$\overline{X} \pm Z\sigma_{\bar{x}} = 8900 \pm 1,96 \frac{\sigma}{\sqrt{n}}$$

$$= 8900 \pm 1,69 \frac{500}{\sqrt{15}}$$

$$= 8900 \pm 1,69(129,20)$$

$$= 8647 \text{ sampai dengan } 9153 \text{ jam}$$

b. Estimasi rata-rata populasi dengan 90 persen konfidensi interval:

$$\overline{X} \pm Z\sigma_{\bar{x}} = 8900 \pm 1,65 \frac{\sigma}{\sqrt{n}}$$

$$= 8900 \pm 1,65 \frac{500}{\sqrt{15}}$$

$$= 8900 \pm 1,65(129,20)$$

$$= 8687 \text{ sampai dengan } 9113 \text{ jam}$$

Contoh 3

Dengan menggunakan ketentuan contoh 2, andaikan nilai standard deviasi populasi dan bentuk distribusi umur bolam tidak diketahui, sedangkan besarnya sampel 35 bolam dan standard deviasi sampel 500 jam. Estimasilah rata-rata umur bolam (populasi) dengan konfidensi interval 99 persen!

Jawab:

Kasus ini dapat diselesaikan dengan menggunakan distribusi normal — berdasarkan teorema limit tengah — karena besarnya sampel 35 bolam (lebih besar dari 30).

$$\overline{X} \pm Zs_{\bar{x}} = 8900 \pm 2,58 \frac{\sigma}{\sqrt{n}}$$

$$= 8900 \pm 2,58 \frac{500}{\sqrt{15}}$$

$$= 8900 \pm 2,58(84,46)$$

$$= 8682 \text{ sampai dengan } 9118 \text{ jam}$$

Contoh 4

Seorang analis penelitian pemasaran mengumpulkan secara random 100 dari 400 pelanggan yang menggunakan kupon khusus. Besarnya belanja rata-rata dari 100 pelanggan tersebut adalah Rp24,57 dengan standard deviasi Rp6,60. Dengan konfidensi interval 95 persen, estimasilah:

a. besarnya belanja rata-rata dari 400 pelanggan;
b. jumlah total uang yang dibelanjakan oleh 400 pelanggan!

Jawab:

a. Estimasi besarnya belanja rata-rata dari 400 pelanggan dengan 95 persen konfidensi interval.

$$\bar{X} \pm Zs_{\bar{x}} = 24.57 \pm 1,96 \frac{s}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$= 24.57 \pm 1,96 \frac{6,60}{\sqrt{100}} \sqrt{\frac{400 - 100}{400 - 1}}$$

$$= 24.57 \pm 1,96(0,57)$$

$$= \text{Rp}23,45 \text{ sampai dengan Rp}25,69$$

b. Interval jumlah total uang yang dibelanjakan oleh 400 pelanggan dapat ditentukan sebagai berikut.

$$N(\bar{X} \pm Zs_{\bar{x}}) = 400(\text{Rp}23,45 \text{ sampai dengan Rp } 25,69)$$

$$= \text{Rp}9.380 \text{ sampai dengan Rp}10.276$$

Contoh 5

Dengan menggunakan ketentuan contoh 2, andaikan nilai standard deviasi populasi tidak diketahui, sedangkan besarnya sampel 15 bolam dan standard deviasi sampel 500 jam.

a. Estimasilah rata-rata umur bolam (populasi) dengan konfidensi interval 95 persen!
b. Estimasilah rata-rata umur bolam (populasi) dengan konfidensi interval 90 persen!

Jawab:

Untuk menyelesaikan kasus ini kita gunakan distribusi t karena populasi berdistribusi normal, standard deviasi populasi tidak diketahui, dan jumlah sampel kecil (kurang dari 30).

a. Estimasi rata-rata umur bolam dengan konfidensi interval 95 persen:

$$\bar{X} \pm t_{df} s_{\bar{x}} = 8900 \pm 2,145 \frac{s}{\sqrt{n}}$$

$$= 8900 \pm 2,145 \frac{500}{\sqrt{15}}$$

$$= 8900 \pm 2,145(129,199)$$

$$= 8623 \text{ sampai dengan } 9177 \text{ jam}$$

b. Estimasi rata-rata umur bolam dengan konfidensi interval 90 persen:

$$
\begin{aligned}
\bar{X} \pm t_{df} s_{\bar{x}} &= 8900 \pm 1{,}761 \frac{s}{\sqrt{n}} \\
&= 8900 \pm 1{,}761 \frac{500}{\sqrt{15}} \\
&= 8900 \pm 1{,}761(129{,}199) \\
&= 8672 \text{ sampai dengan } 9128 \text{ jam}
\end{aligned}
$$

Jika sampel kecil dan populasi tidak berdistribusi normal, baik distribusi normal Z maupun distribusi t-student tidak dapat digunakan untuk menentukan konfidensi interval. Untuk itu kita gunakan k - standard. Metode ini dikemukakan oleh **Chebyshev**.

Sebagai penerapan distribusi sampling suatu rata-rata, probabilitas bahwa suatu rata-rata sampel akan tersebar dengan k - standard adalah

$$P(|X - \mu| \leq k\,\sigma_{\bar{x}}) \geq 1 - \frac{1}{k^2}$$

Pertidak-samaan di atas disebut pertidak-samaan Chebyshev. Jika $\sigma_{\bar{x}}$ tidak diketahui, $s_{\bar{x}}$ dapat digunakan, akan tetapi dengan resiko karena adanya fluktuasi.

Interval rata-rata populasi dapat dihitung dengan rumus

$$\bar{X} \pm k\,\sigma_{\bar{x}}$$

atau

$$\bar{X} \pm k\,\sigma_{\bar{x}}$$

Nilai k dapat diperoleh dengan rumus:

koefisien konfidensi = $1 - 1/k^2$

Contoh 6

Rata-rata umur bolam untuk sampel sebanyak 15 bolam adalah 8900 jam dengan standard deviasi 500 jam. Jika distribusi umur bolam tidak dapat diasumsikan berdistribusi normal, tentukan estimasi rata-rata umur populasi bolam pada konfidensi interval 90 persen.

Jawab:

Karena

$$
\begin{aligned}
1 - \frac{1}{k^2} &= 0{,}90 \\
\frac{1}{k^2} &= 1{,}00 - 0{,}90 = 0{,}10
\end{aligned}
$$

$$
\begin{aligned}
0{,}10k^2 &= 1 \\
k^2 &= 10 \\
k &= \sqrt{10} = 3{,}162
\end{aligned}
$$

dapat kita hasilkan

$$X \pm k\frac{s}{\sqrt{n}} = 8900 \pm 3,162\,\frac{500}{\sqrt{15}}$$

$$= 8900 \text{ plusminus } 3,162(129,199)$$
$$= 8492 \text{ sampai dengan } 9308 \text{ jam}$$

KONFIDENSI INTERVAL UNTUK PERBEDAAN DUA RATA-RATA POPULASI

Konfidensi interval untuk tujuan ini disusun seperti susunan untuk mengestimasi rata-rata dengan kesalahan standar perbedaan dua rata-rata. Distribusi normal dapat digunakan, kecuali apabila dua sampel saling berpengaruh atau tidak saling asing. Formula yang digunakan untuk mengestimasi perbedaan dua rata-rata populasi adalah

$$(\overline{X}_1 - \overline{X}_2) \pm Z\sigma_{\bar{x}_1 - \bar{x}_2}$$

atau

$$(\overline{X}_1 - \overline{X}_2) \pm Zs_{\bar{x}_1 - \bar{x}_2}$$

Apabila deviasi standar kedua populasi tidak diketahui kesalahan standar dari perbedaan dua rata-rata adalah

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^{\,2} - {}_{\bar{x}_2}^{\,2}}$$

atau

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^{\,2} - {}_{\bar{x}_2}^{\,2}}$$

Jika jumlah sampel < 30, distribusi t-student digunakan untuk menentukan konfidensi interval dengan $df = n_1 + n_2 - 2$. Dengan demikian konfidensi interval menjadi

$$(\overline{X}_1 - \overline{X}_2) \pm t_{df} s_{\bar{x}_1 - \bar{x}_2}$$

Contoh 1

Seorang analis penelitian pemasaran mengumpulkan secara random 100 dari 400 pelanggan yang menggunakan kupon khusus. Besarnya belanja rata-rata dari 100 pelanggan tersebut adalah Rp24,57 dengan standard deviasi 6,60. Untuk 200 pelanggan dari 900 pelanggan yang tidak menggunakan kupon khusus didapat rata-rata Rp19,60 dengan standard deviasi Rp8,40.

a. Dengan konfidensi interval 95 persen, estimasilah rata-rata belanja dari pelanggan yang tidak menggunakan kupon khusus!
b. Dengan menggunakan konfidensi interval 90 persen, estimasilah perbedaan rata-rata belanja antara pelanggan yang menggunakan kupon khusus dengan pelanggan yang tidak menggunakan kupon khusus!

Jawab:

a. Estimasi rata-rata belanja dari pelanggan yang tidak menggunakan kupon khusus dengan konfidensi interval 95 persen:

$$X \pm Zs_{\bar{x}} = 19{,}60 \pm 1{,}96 \frac{s}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$= 19{,}60 \pm 1{,}96 \frac{8{,}40}{\sqrt{400}} \sqrt{\frac{900 - 200}{900 - 1}}$$

$$= 19{,}60 \pm 1{,}96(0{,}37)$$

$$= \text{Rp}18{,}87 \text{ sampai dengan Rp}20{,}33$$

b. Estimasi perbedaan belanja rata-rata antara pelanggan yang menggunakan kupon khusus dengan pelanggan yang tidak menggunakan kupon khusus:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^{\;2} - {}_{\bar{x}_2}^{\;2}} \quad = \sqrt{(0{,}57)2 + (0{,}37)^2} \quad = 0{,}772$$

$$(\bar{X}_1 - \bar{X}_2) \pm Z\, s_{\bar{x}_1 - \bar{x}_2} = (24{,}57 - 19{,}60) \pm 1{,}65(0{,}67)$$

$$= \text{Rp}3{,}86 \text{ sampai dengan Rp}6{,}08$$

Contoh 2

Sampel random dari 50 penghuni perumahan A memiliki pendapatan rata-rata Rp13.800 dengan standard deviasi Rp2.200. Sampel random dari 50 penghuni perumahan B memiliki pendapatan rata-rata Rp14.600 dengan standard deviasi 2.800. Estimasilah perbedaan pendapatan rata-rata penghuni perumahan A dan perumahan B dengan konfidensi interval 95 persen.

Jawab:

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{2200}{\sqrt{50}} = \text{Rp}311{,}17$$

$$s_{\bar{x}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{2800}{\sqrt{50}} = \text{Rp}396{,}04$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^{\;2} - {}_{\bar{x}_2}^{\;2}}$$

$$= \sqrt{(311{,}17)^2 + (396{,}04)^2} = \text{Rp}503{,}66$$

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\bar{x}_1 - \bar{x}_2} = (13.800 - 14.600) \pm 1{,}96(503{,}66)$$

$$= \text{Rp}1.1787{,}17 \text{ sampai - Rp}187{,}17$$

Contoh 3

Suatu sampel random jenis produk yang dihasilkan suatu perusahaan memiliki: jumlah sampel 12 buah dan berat rata-rata 15,97 ons dengan standard deviasi 0,15 ons. Sedangkan sampel yang lain: jumlah sampel 15 buah dan berat rata-rata 16,14 ons dengan satndard deviasi 0,09 ons. Distribusi berat produk diasumsikan berdistribusi normal. Estimasilah perbedaan rata-rata tersebut dengan konfidensi interval 90 persen!

Jawab:

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{0{,}15}{\sqrt{12}} = 0{,}043 \text{ ons}$$

$$s_{\bar{x}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{0{,}09}{\sqrt{15}} = 0{,}023 \text{ ons}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^{2} + s_{\bar{x}_2}^{2}}$$

$$= \sqrt{(0{,}043)^2 + (0{,}023)^2} = 0{,}049 \text{ ons}$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{df}\, s_{\bar{x}_1 - \bar{x}_2} = (15{,}97 - 16{,}14) \pm 1{,}708(0{,}049)$$

$$= -0{,}25 \text{ ons sampai } 0{,}09 \text{ ons}$$

KONFIDENSI INTERVAL PROPORSI

Hal lain yang perlu diperhatikan — selain yang telah dijelaskan dimuka — adalah jika proporsi populasi tidak diketahui, banyak ahli statistik berpendapat bahwa jumlah sampel sekurang- kurangnya 100.

Jika proporsi sampel $\bar{p}$ dan kesalahan standar proporsi adalah

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

Untuk populasi terbatas dan $n \geq 0{,}05$ N, formula kesalahan standar proporsi menjadi

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \sqrt{\frac{N - n}{N - 1}}$$

Sedangkan konfidensi interval untuk proporsi populasi adalah

$$\bar{p} \pm Z\, s_{\bar{p}}$$

Banyaknya data dalam sampel dengan proporsi populasi π adalah

$$n = \frac{Z^2 \pi (1 - \pi)}{E^2}$$

Catatan:

E: merupakan kesalahan *plus dan minus.*

Jika π tidak pasti, untuk menentukan besarnya sampel digunakan rumus

$$n = \left(\frac{Z}{2\,E}\right)^2$$

Contoh 1

Untuk menguji keberhasilan suatu produk diambil sampel sebanyak 230 unit produk. Ternyata ada 54 unit produk yang dapat dikategorikan berhasil. Estimasilah produk berhasil tersebut dengan konfidensi interval 90 persen!

Jawab:

$$\bar{p} = \frac{54}{230} = 0{,}235$$

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= \sqrt{\frac{(0{,}235)(0{,}765)}{230}} = 0{,}028$$

$$\bar{p} \pm Z\, s_{\bar{p}} = 0{,}235 \pm 1{,}65(0{,}028)$$
$$\approx 0{,}19 \text{ sampai } 0{,}28$$

Contoh 2

Di sebuah kota besar terdapat 800 pompa bensin. Diambil sampel sebanyak 36. Dari jumlah tersebut 20 pompa bensin menawarkan minyak produk luar negeri. Dengan menggunakan konfidensi interval 95 persen estimasilah:

a. proporsi pompa bensin yang menawarkan minyak produk luar negeri;
b. jumlah pompa bensin yang menawarkan minyak produk luar negeri!

Jawab:

a. Proporsi pompa bensin yang menawarkan minyak produksi luar negeri:

$$\bar{p} = \frac{20}{36} \approx 0{,}56$$

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{(0{,}235)(0{,}765)}{230}} = 0{,}083$$

$$\bar{p} \pm Z\, s_{\bar{p}} = 0{,}56 \pm 1{,}96(0{,}083)$$
$$\approx 0{,}40 \text{ sampai } 0{,}72$$

b. Jumlah pompa bensin yang menawarkan minyak produk luar:

$$N(\bar{p} \pm Z\, s_{\bar{p}}) = 800(0{,}40 \text{ sampai dengan } 0{,}72)$$
$$= 320 \text{ sampai dengan } 576$$

KONFIDENSI INTERVAL PERBEDAAN DUA PROPORSI POPULASI

Formulasi konfidensi interval dua proporsi populasi adalah

$$(\bar{p}_1 - \bar{p}_2) \pm Z\, s_{\bar{p}_1 - \bar{p}_2}$$

Dengan

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{s_{\bar{p}_1}^2 - \bar{p}_2^2}$$

Contoh 1

Asumsikan bahwa $p_1 = 0{,}75$ dan $p_2 = 0{,}60$; $n_1 = 50$ dan $n_2 = 100$. Dengan konfidensi interval 90 persen tentukan perbedaan proporsi populasi!

Jawab:

$$s_{\bar{p}_1}^{\,2} = \frac{p_1(1 - p_1)}{n_1} = \frac{(0,75)(0,25)}{50} = 0,00375$$

$$s_{\bar{p}_1}^{\,2} = \frac{p_2(1 - p_2)}{n_2} = \frac{(0,60)(0,40)}{100} = 0,0024$$

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{s_{\bar{p}_1}^{\,2} + {}_{\bar{p}_2}^{\,2}}$$

$$= \sqrt{(0,00375) + (0,0024)} = 0,0784$$

$$(\bar{p}_1 - \bar{p}_2) \pm Z\, s_{\bar{p}_1 - \bar{p}_2} = (0,75 - 0,60) \pm 1,65(0,0784)$$
$$= 0,02064 \text{ sampai dengan } 0,027936$$

Contoh 2

Manajer sebuah kantor akuntan ingin menyelidiki kewajiban pembayaran pajak perusahaan yang ada di kota A dan kota B. Untuk keperluan tersebut masing-masing kota diambil sampel sebanyak 100 laporan keuangan periode tertentu. Dari kota A, 70 perusahaan memenuhi kewajiban pembayaran pajak. Sedangkan dari kota B hanya 50 perusahaan memenuhi kewajiban pembayaran pajak. Estimasilah perbedaan proporsi pemenuhan pembayaran pajak perusahaan yang ada di kota A dan kota B dengan menggunakan konfidensi interval 95 persen!

Jawab:

$$s_{\bar{p}_1}^{\,2} = \frac{p_1(1 - p_1)}{n_1} = \frac{(0,70)(0,30)}{100} = 0,0021$$

$$s_{\bar{p}_1}^{\,2} = \frac{p_2(1 - p_2)}{n_2} = \frac{(0,50)(0,50)}{100} = 0,0025$$

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{s_{\bar{p}_1}^{\,2} + {}_{\bar{p}_2}^{\,2}}$$

$$= \sqrt{(0,0021) + (0,0025)} = 0,068$$

$$(\bar{p}_1 - \bar{p}_2) \pm Z\, s_{\bar{p}_1 - \bar{p}_2} = (0,70 - 0,50) \pm 1,96(0,068)$$
$$= 0,07 \text{ sampai dengan } 0,33$$

SOAL-SOAL

1. Keuntungan rata-rata perusahaan eksportir per bulan apabila dinyatakan dengan US dolar sebesar \$3,400 dengan satndard deviasi \$200. Jika jumlah perusahaan eksportir besar, berapa probabilitas rata-rata keuntungan dari sampel yang besarnya 25 perusahaan:

a. lebih dari \$3,500 b. di antara \$3,350 sampai dengan \$3,450

Jawab: a. 0,0062 b. 0,7888

2. Dengan menggunakan ketentuan soal nomor 1, jika besarnya populasi terbatas, yaitu 100 perusahaan eksportir, tentukan probabilitas rata-rata keuntungan perusahaan eksportir:

   a. lebih dari \$3,500; b. di antara \$ 3,350 sampai dengan \$3,450!

   Jawab: a. 0,0021 b. 0,8502

3. Jumlah bank yang terdapat di suatu kota sebanyak 380 buah. Dari sampel 50 buah bank yang ada di kota tersebut didapat informasi bahwa banyaknya transakasi rata-rata per hari adalah 420,4 dengan standard deviasi 55,7. Buatlah estimasi rata-rata transaksi per hari dari bank-bank yang ada di kota itu dengan konfidensi interval 90 persen!

   Jawab: 408,3 sampai dengan 423,5 transaksi

4. Diambil 16 buah jeruk jenis tertentu sebagai sampel. Berat rata-rata 7,50 ons dengan standard deviasi 1,00 ons. Berat jeruk diasumsikan berdistribusi normal. Estimasilah berat jeruk rata-rata dengan konfidensi interval 80 persen!

   Jawab: 7,16 sampai dengan 7,82 ons

5. Bagian keuangan memiliki staf sebanyak 100 orang. 10 orang ditugas untuk mengerjakan tugas pemeriksaan yang sama. Rata-rata waktu yang mereka perlukan untuk menyelesaikan tugas tersebut 85 menit dengan standard deviasi 15 menit. Andaikan seluruh staf diminta mengerjakan tugas yang sama dengan yang dikerjakan oleh 10 orang tersebut, tentukan batasan waktu yang mereka perlukan untuk menyelesaikan tugas itu dengan konfidensi interval 90 persen!

   Jawab: 70,7 sampai dengan 99,3 menit

6. Dua buah sampel random yang independen $n_1 = n_2 = 30$ yang diambil dari populasi populasi yang memiliki rata-rata sama dan memiliki standard deviasi sama (= 5). Jika nilai rata-rata dari sampel pertama 14 dan dari sampel kedua 16, berapa probabilitas perbedaan rata-rata sampel yang diambil dari kedua populasi lebih besar atau sama dengan 2?

   Jawab: 0,1212

7. Untuk mengetahui efektivitas kerja dua buah perusahaan besar A dan B diambil 30 karyawan dari perusahaan A sebagai sampel dan 40 karyawan dari perusahaan B. Dari sampel-sampel tersebut diperoleh informasi bahwa rata-rata waktu kerja efektip di perusahaan A selama 7,5 jam dengan standard deviasi 1 jam dan di perusahaan B selama 7,05 jam dengan standard deviasi 1,2 jam. Estimasilah perbedaan rata-rata waktu kerja efektip pada kedua perusahaan itu dengan konfidensi interval 90 persen!

   Jawab: 0,02 sampai dengan 0,88 jam

8. Sebuah distributor memesan suatu jenis barang sebanyak 200 unit dari dari perusahaan besar. Diambil 50 unit sebagai sampel untuk diteliti. Ternyata terdapat 5 unit yang tak layak pakai. Estimasilah propors seluruh barang yang tak layak pakai dengan konfidensi interval 95 persen!

   Jawab: 0,03 sampai dengan 0,17

9. Dengan menggunakan ketentuan soal nomor 8 estimasilah jumlah barang yang tak layak pakai dengan konfidensi interval 90 persen!

   Jawab: 8 sampai dengan 32

10. Dua buah agen majalah (sebut A dan B) mengambil suatu majalah masing-masing sebanyak 100 eksemplar setiap penerbitan. Untuk majalah tersebut agen A memiliki sisa rata-rata 12 eksemplar dan agen B memiliki sisa rata-rata 6 eksemplar. Estimasilah persentase perbedaan sisa rata-rata majalah tersebut dengan konfidensi interval 90 persen.

    Jawab: - 0,6 % sampai dengan 12 %

Bagian IV Uji Hipotesis

Hipotesis dapat diartikan sebagai suatu asumsi atau dugaan yang kita tentukan tentang nilai parameter populasi. Keterangan sampel digunakan untuk menguji kenalaran hipotesis.

Setelah sampel yang diambil secara random dikumpulkan kita bandingkan statistik sampel, misalnya rata-rata sampel ($\bar{x}$), dengan parameter dugaan, misalnya rata-rata populasi dugaan (π). Kemudian, kita tentukan menerima atau menolak nilai dugaan seteliti mungkin. Nilai dugaan ditolak hanya jika hasil dari sampel secara nyata tidak mungkin terjadi bahwa hipotesis benar.

Dengan demikian dapat disimpulkan bahwa uji hipotesis dapat digunakan untuk menguji perbedaan nilai parameter, rata-rata, proporsi, maupun kondisi-kondisi tertentu.

Prosedur/langkah-langkah pengujian hipotesis:

Langkah I:

Menentukan hipotesis nol (H_0) dan hipotesis alternatip (H_a). H_0 merupakan hipotesis nilai parameter dugaan yang dibandingkan dengan hasil perhitungan dari sampel. H_0 ditolak hanya jika hasil perhitungan dari sampel tidak mungkin memiliki kebenaran terhadap hipotesis yang ditentukan terjadi. H_a diterima hanya jika H_0 ditolak.

Langkah II:

Menetapkan tingkat signifikansi yang digunakan. Tingkat signifikansi adalah standard statistik yang digunakan untuk menolak H_0. Jika ditentukan tingkat signifikansi 5 persen (α = 0,05), H_0 ditolak hanya jika hasil perhitungan dari sampel sedemikian berbeda dengan nilai dugaan (yang dihipotesakan). Baik hipotesis perbedaan maupun lebih besar akan memiliki kesempatan untuk terjadi 5 persen atau kurang, atau memiliki probabilitas 5 persen atau kurang.

Catatan tentang Kesalahan Tipe I dan Kesalahan Tipe II.

Jika digunakan tingkat signifikansi 5 persen berarti ada probabilitas sebesar 5 persen untuk menolak H_0, ternyata H_0 tersebut benar. Kejadian ini disebut Kesalahan Tipe I. Probabilitas Kesalahan Tipe I selalu sama dengan tingkat signifikansi yang digunakan sebagai dasar untuk menolak H_0; dilambangkan dengan α (alpha), dan α juga melambangkan tingkat signifikansi (dalam *print out* microstat dinyatakan dalam kolom probabilitas).

Misalkan dalam uji dua sisi dengan distribusi Z kita peroleh nilai Z = - 3,03. Ini berarti nilai probabilitas P(0 _ Z < 3,03) = 0,4988 dan nilai _ = 2 (0,5 - 0,4988) = 0,0024. Dengan demikian dapat dikatakan bahwa probabilitas penolakan H_0 padahal H_0 benar adalah 0,24 persen. Dengan kata lain, kemungkinan salah atas keputusan tersebut sebesar 0,24 persen.

Kesalahan Tipe II terjadi jika H_0 diterima padahal H_0 tersebut salah. Probabilitas dari Kesalahan Tipe II dilambangkan dengan β (beta).

Tabel 4.1

Konsekuensi Keputusan Pengujian Hipotesis

Keputusan	**Kemungkinan**	
	H_0 benar	**H_0 salah**
Menerima H_0	penerimaan benar	kesalahan tipe II
Menolak H_0	kesalahan tipe I	penolakan benar

Langkah III:

Memilih uji statistik. Uji statistik akan merupakan salah satu dari statistik sampel (estimator tidak bias dari parameter yang sedang diuji), atau suatu versi yang ditransformasikan dari statistik sampel. Misalnya, menguji suatu nilai hipotesis dari rata-rata populasi, rata-rata dari suatu sampel random yang diambil dari populasi tersebut dapat dipakai sebagai uji statistik. Jika distribusi sampling dari rata-rata merupakan distribusi normal, nilai rata- rata sampel secara khusus ditransformasikan ke suatu nilai Z.

Langkah IV:

Menentukan nilai kritis atau nilai-nilai uji statistik. Setelah memiliki hipotesis nol tertentu, tingkat signifikansi, dan uji statistik yang digunakan, kita tentukan nilai (atau nilai-nilai) uji statistik. Ada kemungkinan terjadi satu atau dua nilai, tergantung pada uji satu sisi atau uji dua sisi. Dalam setiap kasus, nilai kritis mengidentifikasi nilai dari uji statistik untuk menolak hipotesis nol.

Langkah V:

Menghitung nilai hitung dari uji statistik. Misalnya, dalam pengujian nilai rata-rata populasi yang ditentukan, suatu sampel yang diambil secara random kita tentukan, kemudian nilai rata-rata sampel kita hitung. Jika nilai kritis ditentukan dengan nilai Z, nilai rata-rata sampel diubah atau ditransformasikan ke dalam nilai Z.

Langkah VI:

Membuat keputusan. Nilai dari sampel statistik yang diobservasi dibandingkan dengan nilai kritis dari uji statistik. Apabila nilai hitung dari uji statistik berada di daerah penerimaan hipotesis nol kita putuskan menerima hipotesis nol. Dan, jika nilai hitung statistik berada di daerah kritis kita putuskan menolak hipotesis nol. Jika hipotesis nol ditolak, hipotesis alternatip diterima, dan sebaliknya.

RATA-RATA DENGAN NILAI YANG DIHIPOTESAKAN

Bagian ini akan menguraikan tentang pengujian hipotesis antara nilai rata-rata populasi atau sampel dengan nilai rata-rata dugaan (hipotesis). Misalkan, simbol nilai hipotesis dari rata- rata populasi adalah μ_h. Tiga pasangan hipotesis nol tentang rata-rata suatu populasi dengan hipotesis alternatipnya adalah

(a). $H_0: \mu \leq \mu_h$ $H_a: \mu > \mu_h$
(b). $H_0: \mu \geq \mu_h$ $H_a: \mu < \mu_h$
(c). $H_0: \mu = \mu_h$ $H_a: \mu \neq \mu_h$

(a) dan (b) merupakan uji satu sisi. (a) uji sisi kanan, sedangkan (b) uji sisi kiri. (c) uji dua sisi. Pengujian akan ditentukan suatu sampel random sederhana sebanyak n data (kasus) dan rata-rata sampel ($\bar{x}$). Kemudian akan digunakan untuk menghitung nilai uji satistik terhadap nilai rata-rata hipotesis yang ditentukan terlebih dulu. Penerimaan atau penolakan hipotesis nol tergantung pada nilai uji statistik.

Contoh 1

Seorang manager produksi menyatakan bahwa isi sebuah susu kaleng sekurang-kurangnya 32 ons. Ujilah hipotesis dengan tingkat signifikansi (kepercayaan) 1 persen jika pengambilan sampel secara acak 60 kaleng susu diperoleh isi rata-rata 31,98 ons dan standard deviasi sampel 0,10 ons.

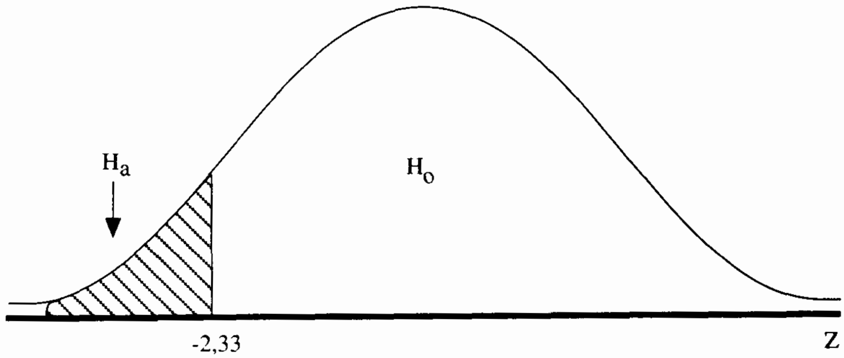
Jawab

1. Hipotesis
   Anggapan bahwa isi rata-rata sekurang-kurangnya 32 ons merupakan hipotesis nol $\mu \geq 32$.

   $H_0: \mu \geq 32$ $H_a: \mu < 32$

2. Nilai Kritis
   Pengujian akan terjadi di sebelah kiri karena H_a menentukan μ kurang dari 32. Besarnya sampel 60 (lebih dari 30), sehingga pengujian statistik menggunakan distribusi Z. Apabila $\mu = 32$, sehingga H_0 benar, kemungkinan untuk menolak kebenaran hipotesis ini (memperhatikan tipe kesalahan I) sebesar $\alpha = 0{,}01$, tingkat signifikansi pengujian. Gambar 1 memperlihatkan daerah penerimaan dan penolakan bagi pengujian yang akan memiliki probabilitas 0,01 dari penolakan H_0 jika $\mu = 32$. Karena nilai $Z_{0,01} = 2{,}33$ kita memiliki aturan pengambilan keputusan.

Gambar 4.1



Menolak H_0 jika nilai uji statistik Z sampel < - 2,33

3. Nilai Hitung

   Nilai uji statistik Z sampel adalah

   $$Z = \frac{\bar{x} - \mu_h}{s_{\bar{x}}}$$

   dengan μ_h = 32. Kita estimasi nilai kesalahan standard rata-rata dengan kesalahan standard sampel dari rata-rata

   $$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{0,1}{\sqrt{60}} = 0,0129$$

   Dengan demikian

   $$Z = \frac{\bar{x} - \mu_h}{s_{\bar{x}}} = \frac{31,98 - 32}{0,0129} = -1,55$$

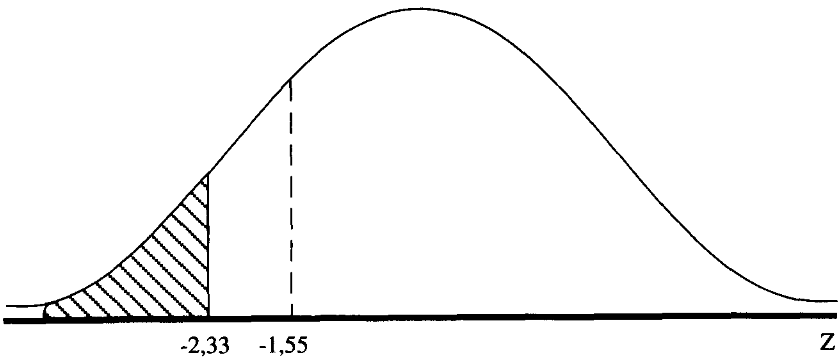
4. Simpulan

   Karena nilai uji statistik Z = - 1,55 tidak kurang dari nilai $Z_{0,01}$ = - 2,33 kita terima H_0. Ini menunjukkan bahwa nilai rata-rata sampel berada di daerah penerimaan H_0. Dengan demikian kita menerima pernyataan manager bahwa isi susu kaleng sekurang-kurangnya 32 ons. Gambar 2 memperlihatkan bahwa H_0 diterima karena nilai uji statistik (nilai sampel) Z berada di daerah penerimaan dari pengujian.

Contoh 2

Setelah diadakan perbaikan, sebuah mesin otomatis dapat memproduksi suatu jenis peralatan yang memiliki diameter 25 milimeter (mm). Sebagian ukuran-ukuran diameter tersebut berdistribusi normal. Rata-rata diameter 10 peralatan yang diambil secara random digunakan untuk menguji apakah mesin tersebut bekerja dengan baik — dapat mengasilkan produk perlatan yang cocok, memiliki diameter 25 mm — atau tidak. (a) Lakukan pengujian hipotesis dengan tingkat signifikansi 5 persen jika rata-rata sampel 25,02 mm dengan standard deviasi sampel 0,024 mm! (b) Apakah maksud dari hasil pengujian?

Gambar 4.2



Jawab

a.1. Hipotesis

Mesin akan beroperasi sesuai dengan keinginan jika rata-rata diameter dari peralatan yang dihasilkan $\mu = 25$ mm. Jika μ kurang dari atau lebih dari 25 mm ($\mu \neq 25$ mm) mesin tidak beroperasi dengan normal. Dengan demikian kita menggunakan pengujian dua sisi.

H_0: $\mu = 25$ mm $\qquad$ H_a: $\mu \neq 25$ mm

2. Nilai Kritis

Karena besarnya sampel hanya 10 (kurang dari 30) kita gunakan distribusi t untuk pengujian statistik. Peluang untuk menolak H_0: $\mu = 25$ apabila benar adalah 0,05 (sama dengan tingkat signifikansi yang digunakan). Kita bagi peluang tersebut ke dalam dua sisi, masing-masing sisi 0,025, seperti terlihat pada Gambar 3, karena kita menggunakan uji dua sisi. Untuk menentukan nilai t pada uji satu sisi dari 0,025, kita perlu mengetahui derajat kebebasan, yaitu

d.f. = n - 1 = 10 - 1 = 9

Besarnya nilai t tabel adalah

$t_{0,025,9} = 2,26$

H_0 akan ditolak jika nilai t sampel kurang dari - 2,26 atau lebih dari 2,26.

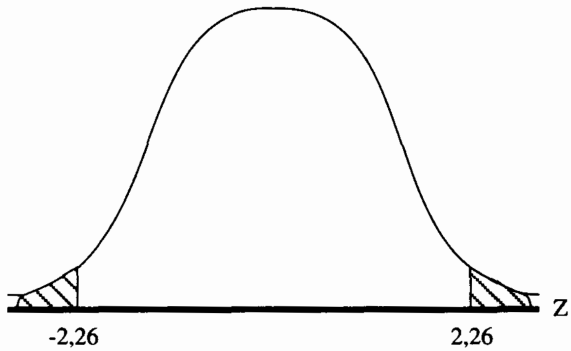
Kita menolak H_0 jika I t sampel I > 2,26

3. Nilai Hitung

Nilai t sampel dapat dihitung dengan menggunakan rumus

$$t = \frac{\bar{x} - \mu_h}{s_{\bar{x}}}$$

Gambar 4.3



Kesalahan standard sampel dari rata-rata dihitung dengan formula

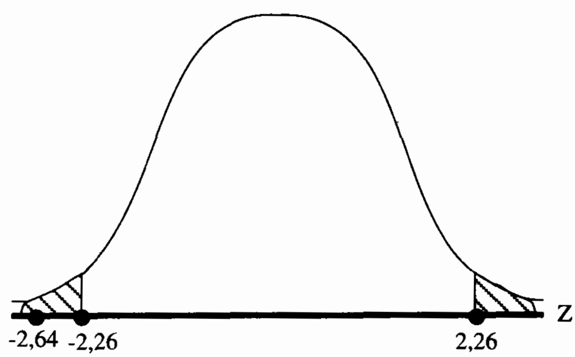
$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{0,024}{\sqrt{10}} = 0,00759$$

Dengan demikian nilai t sampel

$$t = \frac{\bar{x} - \mu_h}{s_{\bar{x}}} = \frac{25,02 - 25}{0,00759} = -2,64$$

4. Simpulan
   Nilai t sampel = 2,64 lebih besar dari 2,26. Oleh karena itu kitak tolak H_0 yang menyatakan bahwa rata-rata diameter produk peralatan 25 mm. Gambar 4.4 memperlihatkan bahwa H_0: μ = 25 ditolak karena nilai t sampel berada dalam daerah penolakan pengujian.

Gambar 4.4



b. Hasil dari pengujian dan kenyataan bahwa rata-rata sampel 25,02 mm, adalah lebih besar dari rata-rata yang diinginkan, 25 mm, berarti bahwa rata-rata populasi lebih dari 25 mm. Dengan demikian mesin harus diperbaiki untuk menghasilkan produk peralatan yang berdiameter lebih kecil.

Perbedaan Rata-rata pada Observasi Berpasangan

Yang dimaksud dengan observasi berpasangan adalah bahwa dua sampel yang digunakan bersifat dependen. Misalkan, kita akan menentukan apakah program latihan (training) akan meningkatkan produktivitas pekerja. Untuk tujuan itu, kita catat produktivitas pekerja sebelum dan sesudah mengikuti program latihan.

Kita gunakan asumsi bahwa perbedaan nilai pasangan (d) mengikuti distribusi normal. Kemudian uji statistiknya adalah nilai t sampel dengan (n - 1) sebagai derajat kebebasan (d.f.).

Simpangan baku dari perbedaan nilai pasangan dapat diperoleh dengan menggunakan rumus:

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

dengan

$$\bar{d} = \frac{\sum d}{n}$$

Kesalahan baku rata-rata didapat dengan rumus:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Dan nilai hitung uji statistik t didapat dengan rumus:

$$t = \frac{\bar{d}}{s_{\bar{d}}}$$

Contoh 3

Produktivitas (jumlah unit yang dihasilkan per hari) dari suatu sampel random 10 pekerja dicatat sebelum dan sesudah mengikuti pelatihan. Berikut ini merupakan pasangan yang didapat; angka pertama pada masing-masing pasangan merupakan output sebelum dan angka kedua ouput sesudah mengikuti pelatihan: (54;60), (56;59), (50;57), (52;56), (55;56), (52;58), (56;62), (53;55), (53;54),(60;64). Ujilah hipotesis dengan tingkat signifikansi 1 persen untuk menentukan apakah rata-rata produktivitas sesudah mengikuti pelatihan lebih besar daripada sebelum mengikuti pelatihan.

Jawab

1. Hipotesis

   Jika rata-rata produktivitas sesudah mengikuti pelatihan lebih besar daripada sebelum mengikuti pelatihan, perbedaan rata-rata populasi D (*sesudah dikurangi sebelum*) lebih besar dari 0. Dengan demikian dapat dilambangkan $D > 0$.

$H_0: D \leq 0 \quad H_a: D > 0$

2. Nilai Kritis

   Pengujian menggunakan satu sisi dengan $\alpha = 0{,}01$. Karena $n = 10$ perbedaan, kita dapatkan derajat kebebasan

   d.f. = n - 1 = 10 - 1 = 9

   Dengan demikian nilai t tabel adalah

   $t_{0{,}01{,}9} = 2{,}82$

   Penolakan H_0 akan terjadi jika nilai t sampel lebih besar 2,82.

3. Nilai Hitung

   Tabel 4.1 untuk mempermudah perhitungan standard deviasi perbedaan sampel.

Tabel 4.1

Pekerja	Produk yang dihasilkan per hari		Perbedaan (d)	$d - \bar{d}$	$(d - \bar{d})^2$
	Sebelum	Sesudah			
1	54	60	6	2	4
2	56	59	3	-1	1
3	50	57	7	3	9
4	52	56	4	0	0
5	55	56	1	-3	9
6	52	58	6	2	4
7	56	62	6	2	4
8	53	55	2	-2	4
9	53	54	1	-3	9
10	60	64	4	0	0
			$\Sigma d = 40$		$\Sigma(d - \bar{d})^2 = 44$
			$\bar{d} = 4$		

Standard deviasi perbedaan sampel dihitung dengan

$$s_d = \sqrt{\frac{\Sigma (d - d)^2}{n - 1}} = \sqrt{\frac{44}{10 - 1}} = 2{,}21$$

Kesalahan standard perbedaan rata-rata sampel diperoleh dengan

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{2,21}{\sqrt{10}} = 0,699$$

Sedangkan nilai t sampel diperoleh dengan

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{4}{0,699} = 5,72$$

4. Simpulan

Nilai t sampel = 5,72, lebih besar daripada nilai $t_{0,01;9} = 2,82$. Dengan demikian kita tolak H_0 yang menyatakan perbedaan produktivitas pekerja per hari sesudah dan sebelum pelatihan lebih kecil atau sama dengan nol. Dengan kata lain kita menerima H_a, yaitu bahwa perbedaan (D) produktivitas pekerja per hari sesudah dengan sebelum pelatihan positip atau lebih besar dari nol.

Contoh 4

Sebuah perusahaan ingin membandingkan kualitas pemakaian dua tipe ban mobil yang berbeda, A dan B. Untuk tujuan tersebut digunakan 5 buah mobil yang menggunakan kedua tipe ban yang diambil secara random. Kelima mobil kemudian dioperasikan (dijalankan) sepanjang jarak tertentu dan masing-masing ban yang digunakan dicatat. Hasil pencatatan terlihat pada Tabel 4.2 Dengan menggunakan data di atas buktikan kesamaan kualitas pemakaian kedua tipe ban dengan tingkat signifikansi 5 persen!

Tabel 4.2

Mobil	Tipe ban	
	A	B
1	10,6	10,2
2	9,8	9,4
3	12,3	11,8
4	9,7	9,1
5	8,8	8,3

Jawab

Perhitungan dengan program *microstat*:

——————— HYPOTHESIS TESTS FOR MEANS ———————

HEADER DATA FOR: C:TIRE LABEL: Kualitas Pemakaian Ban Tipe A dan Tipe B

NUMBER OF CASES: 5 NUMBER OF VARIABLES: 2

DIFFERENCE BETWEEN MEANS: PAIRED OBSERVATIONS

Pengolahan Kualitas Pemakaian Ban Tipe A dan Tipe B
HEADER DATA FOR: C:TIRE LABEL: Kualitas Pemakaian Ban Tipe A dan Tipe B
NUMBER OF CASES: 5 NUMBER OF VARIABLES: 2

```
HYPOTHESIZED DIFF. =  .0000
              MEAN =  .4800
         STD. DEV. =  .0837
        STD. ERROR =  .0374
                 N =  5  (CASES=   1 TO   5)

                 T =  12.8285    (D.F. = 4)   GROUP 1: A
                                              GROUP 2: B
             PROB. =  1.064E-04
```

1. Hipotesis

   Dari hasil hitungan dengan program *microstat* di atas dapat dibuat analisis sebagai berikut. Karena menguji kesamaan berarti hipotesis yang akan kita uji adalah $\_{A} - \_{B} = 0$ atau $D = 0$. Ini menunjukkan pengujian yang akan kita gunakan adalah pengujian dua sisi.

   $H_0: D = 0 \quad H_a: D \neq 0$

   D: perbedaan kualitas pemakaian ban tipe A dan tipe B.

2. Nilai Kritis

   Karena uji dua sisi berarti nilai t tabel adalah

   $t_{0,025,\,4} = 2,776$

3. Nilai hitung t sampel.

   Dari *print out* dapat kita peroleh nilai uji statistik t sampel

   $t = 12,8$

4. Simpulan

   Karena nilai uji statisti t sampel = 12,8 lebih besar dari nilai t tabel = 2,776. Dari hasil tersebut kita menyimpulkan bahwa H_0 ditolak. Dengan kata lain kualitas pemakaian kedua tipe ban tidak sama atau berbeda.

PERBEDAAN RATA-RATA DARI DUA KELOMPOK BERBEDA

Andaikan kita mulai dengan dua populasi yang berdistribusi normal. Populasi I memiliki rata-rata μ_1 dan simpangan baku σ_1. Populasi II memiliki rata-rata σ_2 dan simpangan baku σ_2. Banyaknya sampel untuk masing-masing populasi n_1 dan n_2 dengan rata-rata $\overline{X}_1$ dan $\overline{X}_2$. Dengan demikian selisih rata-rata kedua populasi dan kesalahan baku rata-rata adalah

Jika simpangan baku populasi tidak diketahui, kesalahan baku rata-rata dapat didekati dengan simpangan baku sampel, dengan rumus:

Kita akan menguji hipotesis perbedaan rata-rata dua populasi (μ_1 - μ_2). Tiga pasangan hipotesis yang dapat terjadi adalah

(a). H_0: $\mu_1 - \mu_2 = 0$ $\quad$ H_a: $\mu_1 - \mu_2 \neq 0$
(b). H_0: $\mu_1 - \mu_2 \geq 0$ $\quad$ H_a: $\mu_1 - \mu_2 < 0$
(c). H_0: $\mu_1 - \mu_2 \leq 0$ $\quad$ H_a: $\mu_1 - \mu_2 > 0$

Apabila sampel besar ($n \geq 30$) kita gunakan uji statistik Z. Sedangkan apabila sampel kecil ($n < 30$) kita gunakan uji statistik t.

Contoh 5

Presiden direktur sebuah agen mobil menceriterakan kepada manager pemasaran bahwa ia merasa senang dengan kenaikan jumlah mobil yang terjual setiap hari sepanjang tahun yang lalu, tetapi tidak senang dengan keuntungan bersih per unit mobil yang terjual. Manager pemasaran menyatakan bahwa keuntungan bersih rata-rata per mobil yang terjual tahun ini lebih besar dibandingkan tahun lalu. Untuk membuktikan pernyataan tersebut manager pemasaran mengambil sampel secara random penjualan tahun lalu dan tahun ini masing-masing 35 mobil yang terjual. Manager pemasaran ingin menunjukkan bahwa keuntungan bersih rata-rata tahun ini μ_1 lebih besar daripada keuntungan bersih rata-rata tahun lalu μ_2. Tunjukkan pengujian hipotesis pada tingkat signifikansi 5 persen apabila keuntungan bersih rata-rata sampel tahun ini $\bar{x}_1 = 350$ dan standard deviasi $s_1 = 25$, dan keuntungan rata-rata sampel tahun lalu $\bar{x}_2 = 340$ dan standard deviasi $s_2 = 30$.

Jawab

1. Hipotesis
   Pernyataan manager pemasaran bahwa keuntungan bersih rata-rata tahun ini lebih besar daripada tahun lalu berarti $\mu_1 > \mu_2$ atau $\mu_1 - \mu_2 > 0$.

   H_0: $\mu_1 - \mu_2 \leq 0$ $\quad$ H_a: $\mu_1 - \mu_2 > 0$

2. Nilai Kritis
   Karena H_a menggunakan tanda > kita menggunakan pengujian satu sisi kanan dengan nilai kritis

   $Z_{0,05} = 1,64$

3. H_0 akan ditolak jika nilai hitung Z sampel lebih besar dari 1,64.

$$s_{x_1 - x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{(25)^2}{35} + \frac{(30)^2}{35}}$$

$$= 6,60$$

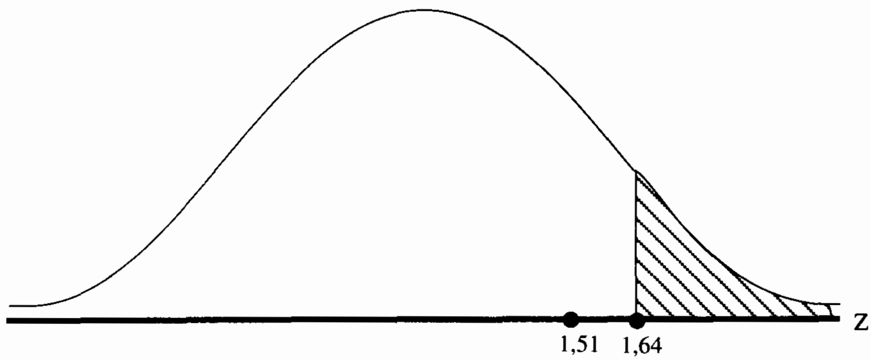
Nilai Z diperoleh dengan formula

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{350 - 340}{6,60} = 1,51$$

4. Simpulan

Nilai hitung Z sampel = 1,51, tidak lebih besar dari 1,64. Dengan demikian Z sampel terletak di daerah penerimaan H_0, berarti kita menerima H_0. Kita dapat mengatakan bahwa pernyataan manager pemasaran bahwa keuntungan bersih rata-rata tahun ini lebih besar daripada tahun lalu adalah salah. Gambar 4.5 memperlihatkan letak nilai hitung Z sampel.

Gambar 4.5



Contoh 6

Sebuah perusahaan ingin menguji dua jenis alat pencampur makanan, jenis I dan jenis II, dengan mengambil sampel secara random, untuk menentukan apakah ada perbedaan berat antara hasil campuran dari alat pencampur jenis I dan jenis II. Informasi dari sampling adalah sebagai berikut:

Jenis I: $n_1 = 12 \bar{x}_1 = 140$; $s_1 = 6$
Jenis II: $n_2 = 15 \bar{x}_2 = 124$; $s_2 = 5$

Tunjukkan pengujian hipotesis dengan tingkat signifikansi 2 persen dan apakah yang dimaksud dengan hasil pengujian?

Jawab

1. Hipotesis

Pengujian yang digunakan untuk menentukan apakah terjadi perbedaan rata-rata adalah pengujian dua sisi.

H_0: $\mu_1 - \mu_2 = 0$ H_a: $\mu_1 - \mu_2 \neq 0$

2. Nilai Kritis

Karena pengujian dua sisi, untuk menentukan nilai t tabel, nilai $\alpha = 0{,}02$ kita bagi dua menjadi $0{,}02/2 = 0{,}01$. Derajat kebebasan

$d.f. = n_1 + n_2 - 2 = 12 + 15 - 2 = 25$

Dengan demikian nilai t sampel adalah

$t_{0{,}01{,}25} = 2{,}49$

Penolakan H_0 hanya terjadi jika | t tabel | > 2,49

3. Nilai Hitung

$$s_{x_1 - x_2} = \sqrt{\frac{(n_1 - 1)_{s_1}^{\ 2} + (n_2 - 1)_{s_2}^{\ 2}}{n_1 + n_2 - 2} + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(12 - 1)(6)^2 + (15 - 1)(5)^2}{12 + 15 - 2} + \left(\frac{1}{12} + \frac{1}{15}\right)}$$

$$= 2{,}116$$

Kemudian nilai t sampel adalah

$$t = \frac{x_1 - x_2}{s_{x_1 - x_2}} = \frac{140 - 124}{2{,}116} = 7{,}56$$

4. Simpulan

Oleh karena nilai t sampel = 7,56 lebih besar dari nilai $t_{0{,}01{,}25} = 2{,}49$, kita tolak H_0. Karena berat hasil campuran dengan alat pencampur jenis I lebih besar daripada dengan alat pencampur jenis II, konsumen akan memilih hasil campuran yang diproduksi dengan alat pencampur jenis I. Sedangkan bagi produsen sebaliknya. Dengan catatan bahwa kondisi yang lain (selain masalah beratnya) sama.

DUA PROPORSI DARI KELOMPOK-KELOMPOK INDEPENDEN

Apabila proporsi sukses dari suatu populasi adalah π, kita dapat mengasumsikan bahwa distribusi proporsi sampel p adalah normal, jika np dan nq (q: proporsi gagal) keduanya lebih dari 5 persen. Pasangan-pasangan hipotesis beda dua proporsi darikelompok-kelompok independen dapat berbentuk:

(a). $H_0: \pi_1 - \pi_2 = 0$ $\quad H_a: \pi_1 - \pi_2 \neq 0$

(b). $H_0: \pi_1 - \pi_2 \leq 0$ $\quad H_a: \pi_1 - \pi_2 > 0$

(c). $H_0: \pi_1 - \pi_2 \geq 0$ $\quad H_a: \pi_1 - \pi_2 < 0$

Andaikan kita mempunyai dua populasi dengan proporsi berturut- turut n_1 dan n_2, sedangkan banyaknya sampel n_1 dan n_2, kesalahan baku beda proporsi dapat diperoleh dengan menggunakan rumus:

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_1 \bar{q}_1}{n_1}}$$

Jika $p_1 = p_2$ rumus di atas dapat diganti menjadi:

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}_c \bar{q}_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Jika p_c tidak diketahui, kesalahan baku beda proporsi dapat dicari dengan rumus:

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}_c \bar{q}_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

dengan:

$$p_c = \frac{\text{jumlah sukses dalam dua sampel}}{n_1 + n_2}$$

Nilai uji statistik dapat dihitung dengan rumus:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sigma_{\bar{p}_1 - \bar{p}_2}}$$

atau

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{s_{\bar{p}_1 - \bar{p}_2}}$$

Contoh 7

Sampel-sampel yang diambil secara random dari 1600 tenaga kerja di daerah I dan 1400 tenaga kerja di daerah II digunakan untuk menentukan apakah proporsi populasi di kedua daerah berbeda. Tunjukkan pengujian hipotesis pada tingkat signifikansi 5 persen apabila jumlah penganggur dari sampel-sampel 120 di daerah I dan 84 di daerah II.

Jawab

1. Hipotesis

   Pernyataan berbeda berarti $\pi_1 \neq \pi_2$ atau $\pi_1 - \pi_2 \neq 0$

   H_0: $\pi_1 - \pi_2 = 0$ H_a: $\pi_1 - \pi_2 \neq 0$

2. Nilai Kritis

   Pengujian dua sisi, nilai probabilitas masing-masing daerah penolakan H_0 adalah 1/2 (0,05) = 0,025, dan nilai

   $Z_{0,025} = 1,96$

   Kita menolak H_0 jika $| Z \text{ sampel} | > 1,96$

3. Nilai Hitung

   Nilai proporsi gabungan adalah

$$\bar{p}_c = \frac{\text{jumlah pengangguran pada kedua sampel}}{n_1 + n_2}$$

$$= \frac{120 + 84}{1600 + 1400}$$

$$= 0{,}068$$

$$\bar{q}_c = 1 - \bar{p}_c = 0{,}9323$$

Kesalahan standard rata-rata diperoleh dengan

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}_c\, \bar{q}_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{(0{,}68)(0{,}932) + \left(\frac{1}{1600} + \frac{1}{1400}\right)}$$

$$= 0{,}00921$$

Nilai-nilai proporsi sampel di daerah I dan daerah II masing-masing

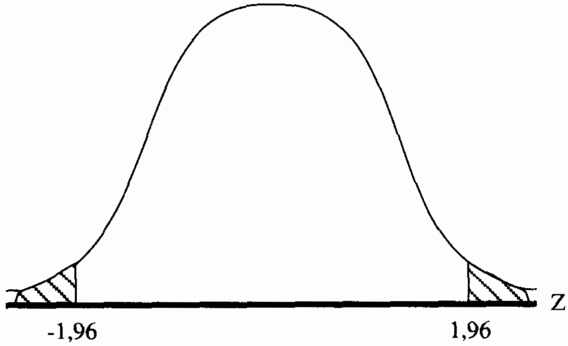
$$\bar{p}_1 = \frac{120}{1600} = 0{,}075$$

$$\bar{p}_2 = \frac{84}{1400} = 0{,}060$$

Nilai Z sampel diperoleh dengan

$$Z = \frac{p_1 - p_2}{s_{p_1 - p_2}} = \frac{0{,}075 - 0{,}060}{0{,}00921} = \frac{0{,}015}{0{,}00921} = 1{,}63$$

Gambar 4.6



4. Simpulan

Karena nilai Z sampel = 1,63 tidak lebih besar dari $Z_{0,025}$ = 1,96, H_0 kita terima. Gambar 4.6 memperlihatkan bahwa letak nilai Z sampel berada di daerah penerimaan H_0. Dengan demikian kita dapat menyimpulkan bahwa proporsi penganggur di daerah I dan daerah II tidak berbeda atau sama.

Contoh 8

Bagian administrasi rumah sakit menganggap penting untuk mengumpulkan dan menghitung data statistik medis yang sangat vital bagi dokter-dokter dan pengambil kebijaksanaan rumah sakit. Catatan sebuah rumah sakit swasta menunjukkan bahwa 52 pria dari 1000 pria dan 23 wanita dari 1000 wanita dirawat karena sakit jantung. Dengan informasi tersebut dapatkah disimpulkan bahwa pasien yang sakit jantung lebih banyak pria dibandingkan dengan wanita?

Jawab

1. Hipotesis

Pernyataan lebih banyak pria berarti $\pi_1 > \pi_2$ atau $\pi_1 - \pi_2 > 0$

$H_0: \pi_1 - \pi_2 \leq 0 \qquad H_a: \pi_1 - \pi_2 > 0$

2. Nilai Kritis

Karena pengujian satu sisi nilai

$Z_{0,05} = 1,645$

Kita menolak H_0 jika Z sampel >1,645.

3. Nilai Hitung

Nilai proporsi gabungan adalah

$$\bar{p}_c = \frac{\text{jumlah penganggur pada kedua sampel}}{n_1 + n_2}$$

$$= \frac{52 + 23}{1000 + 1000}$$

$$= 0,0375$$

$$\bar{q}_c = 1 - \bar{p}_c = 0,9625$$

Kesalahan standard rata-rata diperoleh dengan

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}_c\,\bar{q}_c\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{(0,0375)(0,9625) + \left(\frac{1}{1000} + \frac{1}{1000}\right)}$$

$$= 0,008496$$

Nilai-nilai proporsi sampel pasien pria dan wanita masing-masing

$$\bar{p}_1 = \frac{52}{1000} \; 0{,}052$$

$$\bar{p}_1 = \frac{32}{1000} \; 0{,}032$$

Nilai Z sampel diperoleh dengan

$$Z = \frac{p_1 - p_2}{s_{p1-p2}} = \frac{0{,}052 - 0{,}023}{0{,}008496} = \frac{0{,}029}{0{,}008496} = 3{,}41$$

4. Simpulan
Karena nilai Z sampel = 3,42 lebih besar dari $Z_{0{,}05}$ = 1,645, H_0 kita tolak. Dengan demikian dapat kita simpulkan bahwa pasien pria yang sakit jantung lebih banyak dibanding dengan pasien wanita yang sakit jantung.

PROPORSI DENGAN NILAI YANG DIHIPOTESAKAN

Apabila kita ingin membandingkan proporsi sampel (p) dengan proporsi tertentu (π_0) — dari anggapan populasi —, kesalahan baku proporsi dapat ditentukan dengan menggunakan rumus:

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}$$

Sedangkan nilai uji statistik didapat dengan menggunakan rumus:

$$Z = \frac{\bar{p} - \pi_0}{\sigma_{\bar{p}}}$$

Contoh 9

Dihipotesakan bahwa tidak lebih dari 5 persen bolam yang dihasilkan suatu pabrik tidak memenuhi ketentuan yang disyaratkan. Pengambilan sampel sebanyak n = 100 unit, ditemukan 10 unit tidak memenuhi ketentuan. Ujilah hipotesis pada tingkat signifikansi 5 persen.

Jawab

1. Hipotesis
Pernyataan tidak lebih dari 5 persen berarti proporsi populasi (π_0) $\leq$ 0,05 dan pengujian yang kita gunakan adalah pengujian satu sisi.

H_0: $\pi \leq 0$ $\quad H_a$: $\pi > 0$

2. Nilai Kritis
Dengan uji satu sisi, nilai $Z_{0.05}$ = + 1,65, karena $n\,\pi_0 \geq 5$ dan $n\,(1 - \pi_0) \geq 5$. Kita menolak H_0 jika nilai Z sampel lebih dari 1,65.

3. Nilai Hitung

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}$$

$$= \sqrt{\frac{(0,05)(0,95)}{100}} = \sqrt{\frac{0,0475}{100}} = 0,022$$

$$Z = \frac{\bar{p} - \pi_0}{\sigma_p} = \frac{0,10 - 0,05}{0,022} = \frac{0,05}{0,022} = 1,63$$

4. Simpulan

Dari perhitungan di atas kita perolehnilai Z sampel = 2,27 lebih besar dari nilai $Z_{0,05}$ = 1,65. Oleh karena itu kita menolakH_0. Ini berarti pernyataan bahwatidak lebih dari 5 persen bolam yang dihasilkan suatu pabrik tidak memenuhi ketentuan yang disyaratkan adalahtidak benar.

Contoh 10

Manager personalia suatu perusahaan menyatakan bahwa 50 persen pegawai baru sekurang-kurangnya membutuhkan waktu 1 bulan untuk dapat menyesuaikan diri dengan bidang pekerjaannya. Setelah diadakan penelitian secara random terhadap 30 pegawai baru ternyata 10 orang dapat menyesuaikan diri kurang dari satu bulan. Dapatkah pernyataan manager tersebut ditolak dengan tingkat signifikansi 5 persen.

Jawab

1. Hipotesis

H_0: $\pi \geq 0,5$ $\qquad$ H_a: $\pi < 0$

2. Nilai Kritis

Nilai kritis untuk uji satu sisi, n = 30 dan tingkat signifikansi 5 persen adalah

$Z_{0,05}$ = - 1,65

Kita menolak H_0 jika nilai Z sampel kurang dari - 1,65.

3. Nilai Hitung

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}$$

$$= \sqrt{\frac{(0,5)(0,5)}{30}} = \sqrt{\frac{0,25}{30}} = 0,09$$

$$Z = \frac{\bar{p} - \pi_0}{\sigma_p} = \frac{0,33 - 0,5}{0,09} = \frac{-0,17}{0,09} = 1,88$$

4. Simpulan

   Karena nilai uji statistik Z sampel = - 1,88 kurang dari nilai $Z_{0,05}$ = - 1,65 kita tolak H_0. Ini berarti pernyataan manager personalia bahwa 50 persen pegawai baru sekurang-kurangnya membutuhkan waktu 1 bulan untuk dapat menyesuai-kan diri dengan bidang pekerjaannya adalah salah.

Dua Proporsi Saling Asing dari Satu Kelompok

Dalam satu kelompok sampel — diambil dari satu populasi — dapat terjadi dua proporsi yang saling asing (*mutually exclusive*), tidak mungkin dua kejadian terjadi bersama-sama. Nilai uji statistik dihitung dengan menggunakan rumus:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{1}{n}(\bar{p}_1(1 - \bar{p}_1) + \bar{p}_2(1 - \bar{p}_2) + 2\,\bar{p}_2\,\bar{p}_2)}}$$

Dengan:

$\bar{p}_1$: proporsi sukses kategori I

$\bar{p}_2$: proporsi sukses kategori II

n : besarnya sampel

Contoh 11

Bagian produksi mengatakan bahwa hasil produksi hari ini tidak sama dengan hasil produksi kemarin. Hasil produksi kemarin berdasarkan analisis statistik dapat dikatakan proporsi produk yang baik sama dengan proporsi produk yang tidak baik. Jumlah produk yang dihasilkan hari ini sebanyak 20 unit, 12 unit baik dan 8 unit tidak baik. Dengan data tersebut buktikan kebenaran pernyataan bagian produksi berdasarkan analisis statistik dengan tingkat signifikansi 1 persen.

Jawab:

Dengan menggunkan program *microstat* kita dapatkan hasil:

HYPOTHESIS TEST FOR TWO PROPORTIONS FROM ONE GROUP
(MUTUALLY EXCLUSIVE CATEGORIES)
P1 = .6000   P2 = .4000   SAMPLE SIZE = 20

Z = .913   PROB. = .1807

1. Hipotesis

   H_0: $\pi_1 - \pi_2 = 0$ H_a: $\pi_1 - \pi_2 \neq 0$

2. Nilai Kritis

   Dengan α = 1 persen nilai statistik $|Z_{0,005}|$ = 2,75

3. Nilai Hitung

   Nilai uji statistik Z sampel (dapat dilihat pada hasil hitungan dengan program *microstat* di atas) adalah

   Z = 0,913

4. Simpulan

   Karena nilai uji statistik Z sampel = 0,913 lebih kecil dari $|Z_{0,005}| = 2{,}75$, kita terima H_0. Dengan demikian pernyataan bagian produksi bahwa hasil produksi hari ini tidak sama dengan hasil produksi kemarin kita tolak.

DUA PROPORSI SALING BERINTERAKSI DARI SATU KELOMPOK

Jika dalam suatu sampel ada dua kriteria/kategori dan ada data dalam sampel itu memenuhi kedua kategori tersebut, kita menyebut bahwa dalam sampel tersebut terjadi kondisi saling ganda kategori (*overlapping*). Untuk mendapatkan nilai uji statistik digunakan rumus:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{1}{n}(\bar{p}_1(1-\bar{p}_1) + \bar{p}_2(1-\bar{p}_2) + 2\,\bar{p}_2\,\bar{p}_2 - \bar{p}_{12})}}$$

Dengan:

$\bar{p}_1$: proporsi sukses kategori I

$\bar{p}_2$: proporsi sukses kategori II

$\bar{p}_{12}$: proporsi sukses kategori I dan II secara bersama-sama

n : besarnya sampel

Contoh 12

Misalkan ada 16 perusahaan memproduksi dua jenis barang yang sama. Misalkan untuk memproduksi barang tersebut diperlukan faktor produksi X. Kualitas faktor produksi X ada dua yaitu kualitas I dan kualitas II. 12 perusahaan menggunakan faktor produksi X kualitas I, 4 perusahaan menggunakan faktor produksi X kualitas II, dan 2 perusahaan menggunakan faktor produksi X, baik kualitas I maupun kualitas II. Dengan menggunakan tingkat signifikansi 1 persen, ujilah bahwa proporsi perusahaan yang menggunakan faktor produksi kualitas I sama dengan proporsi perusahaan yang menggunakan faktor produksi kualitas II.

Jawab:

Dengan menggunakan program *microstat* kita dapatkan:

HYPOTHESIS TEST FOR TWO PROPORTIONS FROM ONE GROUP
(OVERLAPPING CATEGORIES)
P1 = .7500 P2 = .2500 SAMPLE SIZE = 16
OVERLAP PROPORTION = .1250
Z = 2.828 PROB. = 2.339E-03

1. Hipotesis

   H_0: $\pi_1 - \pi_2 = 0$ H_a: $\pi_1 - \pi_2 \neq 0$

2. Nilai Kritis

   Dengan α = 1 persen nilai statistik $|Z_{0,005}| = 2,75$

3. Nilai Hitung

   Nilai uji statistik Z sampel (dapat dilihat pada hasil hitungan dengan program *microstat* di atas) adalah

   $Z = 2,828$

4. Simpulan

   Karena nilai uji statistik Z sampel = 2,828 lebih besar dari $|Z_{0,005}| = 2,75$, kita tolak H_0. Dengan demikian pernyataan bahwa proporsi perusahaan yang menggunakan faktor produksi kualitas I sama dengan proporsi perusahaan yang menggunakan faktor produksi kualitas II tidak benar.

CARA LAIN UNTUK MELAPORKAN HASIL PENGUJIAN-PENGUJIAN SECARA STATISTIK: P-VALUE

Probabilitas α dari pembuatan kesalahan tipe I sering disebut sebagai tingkat signifikansi dari pengujian secara statistik. Probabilitas nilai dari pengujian statistik yang diobservasi, atau beberapa, bahkan yang berlawanan dengan hipotesis nol (H_0), mengukur kekuatan dari bukti yang mendukung penolakan. Beberapa peneliti berpendapat bahwa hasil-hasil pengujian statistik pada kondisi signifikan (kita akan menolak) pada tingkat signifikan 5 persen dan tidak pada tingkat signifikan 1 persen. Ini berarti bahwa kita dapat menolak H_0 pada $\alpha = 0,05$ tetapi tidak pada $\alpha = 0,01$.

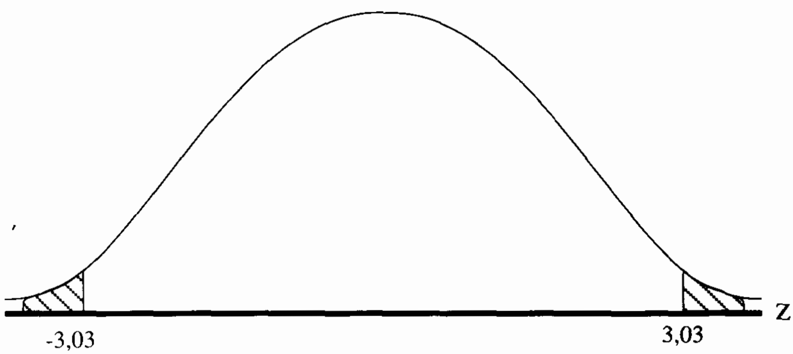
Tingkat signifikan menyatakan probabilitas pengamatan suatu kejadian sampel lebih terhadap yang berlawanan dengan H_0 daripada hasil sampel yang diamati seandainya H_0 benar. Semakin kecil nilai dari probabilitas ini menunjukkan kekuatan dari bukti sampel untuk menolak H_0 semakin besar. Misalnya, suatu pengujian statistik dengan tingkat signifikansi 1 persen lebih memiliki dukungan untuk menolak H_0 daripada pengujian statistik lain dengan tingkat signifikansi 2 persen. Nilai α terkecil yang menjadikan hasil pengujian signifikan secara statistik sering disebut *p-value* terhadap pengujian. Dengan demikian dapat didefinisikan bahwa *p-value* terhadap suatu pengujian hypotesis tertentu merupakan probabilitas untuk memperoleh suatu nilai dari pengujian statistik secara ekstrem atau lebih ekstrem daripada nilai sampel yang terjadi seandainya H_0 benar. *p-value* juga disebut tingkat signifikansi observasi pengujian, apabila *p-value* merupakan nilai α terbesar sehingga kita dapat menolak H_0 dengan menggunakan hasil-hasil sampel yang diobservasi. Beberapa program pengolahan statistik dengan komputer menghitung nilai *p-value* walaupun sudah ada tabel yang dapat digunakan untuk mencari nilai *p-value*.

Contoh 11

Misalkan ditentukan H_0: $\mu = 880$ dan H_a: $\mu \neq 880$. Nilai uji statistik Z sampel = -3,03. Dengan demikian *p-value* untuk pengujian dua sisi adalah probabilitas dari $Z \leq -3{,}03$ atau $Z \geq 3{,}03$, lihat Gambar 4.7; nilai *p-value* merupakan dua kali nilai probabilitas dari nilai uji statistik Z sampel yang diperoleh dari tabel.

Dari tabel nilai Z kita dapatkan bahwa luas daerah di bawah kurva normal di antara Z = 0 dan Z = 3,03 adalah 0,4988. Dengan demikian luas daerah di sebelah kanan Z = 3,03 adalah 0,5 - 0,4988 = 0,0012. Sehingga, karena pengujian dua sisi, nilai _ untuk daerah penolakan dari $Z > 3{,}03$ atau $Z < -3{,}03$ adalah 2(0,0012) = 0,0024. Ini berearti kita dapat menyatakan bahwa nilai *p-value* = 0,0024 atau 0,24 persen. Interpretasi dari nilai *p-value* ini adalah sebagai berikut: Probabilitas bahwa kita akan meneliti suatu nilai Z sebesar 3,03 atau sekecil - 3,03, dengan ketentuan bahwa H_0 benar (misalnya _ = 880) adalah hanya sebesar *p-value* = 0,24 persen.

Gambar 4.7

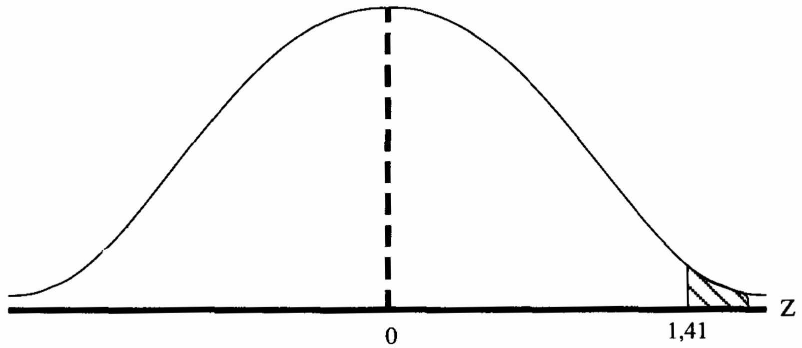


Contoh 12

Misalkan dalam pengujian satu sisi, H_o: $\pi \leq 0{,}10$ H_a: $\pi > 0{,}10$ dan nilai uji statistik Z sampel = 1,41. Dengan demikian nilai *p-value* untuk pengujian ini adalah probabilitas observasi suatu nilai Z yang lebih dari 1,41. Nilai ini merupakan luas daerah di bawah kurva normal di sebelah kanan Z = 1,41. Lihat Gambar 4.8. Dari tabel nilai Z kita dapatkan bahwa luas daerah di bawah kurva normal di antara Z = 0 dan Z = 1,41 adalah 0,4207. Dengan demikian luas daerah di bawah kurva normal di sebelah kanan Z = 1,41, yaitu *p-value*, adalah *p-value* = 0,5 - 0,4207 = 0,0793.

Banyak peneliti mengevaluasi *p-value* dengan empat batasan. Jika *p-value* > 0,05, hasil-hasil penelitian tidak signifikan; jika 0,01 < *p-value* < 0,05, hasil-hasil penelitian dikatakan signifikan; jika 0,001 < *p-value* < 0,01, hasil-hasil penelitian sangat signifikan; jika *p-value* < 0,001, hasil-hasil penelitian dikatakan amat sangat signifikan. Meskipun *p-value* di antara 0,05 dan 0,10 dikatakan tidak signifikan, sering disebut mengarah ke signifikan.

Gambar 4.8



SOAL-SOAL

1. Ujilah hipotesis (nol) bahwa hasil produksi rata-rata perusahaan kimia per hari adalah $\pi = 880$ ton dengan hipotesis (alternatip) lebih besar atau lebih kecil dari 880 ton denggan tingkat signifikansi 5 persen. Besarnya sampel $n = 50$ didapat hasil produksi rata-rata 871 ton dengan standard deviasi 21 ton.

   Jawab: H_0 ditolak.

2. Misalkan sebuah agen paket menyatakan bahwa pengiriman barang per hari 3.000 kg. Untuk menguji pernyataan tersebut diambil sampel sebanyak delapan hari secara random dan diperoleh data sebagai berikut (dalam kg): 3.005, 2.925, 2.935, 2.965, 2.995, 3.005, 2.935, 2.905. Ujilah pernyataan tersebut dengan menggunakan tingkat signifikansi 5 persen.

   Jawab: pernyataan ditolak.

3. Ada banyak metode yang dapat digunakan untuk mengevaluasi arus barang dagangan. Dua metode yang biasa digunakan adalah metode LIFO (*last in, first out*) dan metode FIFO (*first in, first out*), dengan segala kelebihan dan kekurangannya. Untuk keperluan tersebut diambil lima jenis barang yang perhitungan nilai akhir tahun berdasarkan kedua metode tersebut dan dilakukan pada suatu akhir tahun anggaran. Hasil dari perhitungan dapat dilihat pada tabel berikut.

Jenis Barang	Nilai barang (x 1.000)	
	FIFO	LIFO
1	121	117
2	217	198
3	92	105
4	98	86
5	52	49

Dengan tingkat signifikansi 5 persen, ujilah pernyataan bahwa metode LIFO lebih efisien jika dibandingan dengan metode FIFO.

Jawab: Pernyataan ditolak/salah.

4. Sebelum menerapkan program keselamatan kerja dinyatakan bahwa jumlah karyawan yang mengalami kecelakaan lebih banyak dibandingkan dengan sesudah menerapkan program keselamatan kerja. Untuk mengujinya diambil sampel enam perusahaan yang menerapkan program tersebut — dengan jumlah tenaga kerja sama. Gunakan tingkat signifikansi 10 persen.

	Prosedur II					
	1	2	3	4	5	6
Sebelum	38	64	42	70	58	30
Sesudah	31	58	43	65	52	29

Jawab: Pernyataan benar.

5. Sebuah agen pengiriman paket ingin menguji biaya bahan bakar dua jenis kendaraan (jenis I dan jenis II) yang berkapasitas angkut sama. Untuk itu masing-masing jenis diambil sebanyak 100 kendaraan sebagai sampel. Rata-rata kebutuhan dana jenis I Rp6,70 per km dengan varian 0,36 dan jenis II Rp6,54 per km dengan varian 0,40. Dengan tingkat signifikansi 5 persen ujilah bahwa biaya bahan bakar kedua jenis kendaraan sama.

Jawab: biaya bahan bakar kedua jenis kendaraan sama.

6. Sebuah perusahaan ingin menguji dua jenis prosedur kerja suatu proses produksi. Perusahaan tersebut menggunakan 18 orang pekerja baru yang dibagi sama ke dalam dua kelompok kerja untuk mencoba kedua jenis prosedur. Lama waktu yang digunakan oleh masing-masing pekerja dapat dilihat pada tabel di bawah ini.

Prosedur I		Prosedur II	
32	44	35	40
37	35	31	27
35	31	29	32
28	34	25	31
41		34	

Dengan menggunakan tingkat signifikansi apakah dapat disimpulkan bahwa prosedur II lebih baik daripada prosedur I.

Jawab: prosedur II tidak dapat dikatakan lebih baik dari prosedur I.

7. Bagian produksi memeriksa dua jenis mesin (A dan B). Dari hasil produk kedua mesin diambil sampel masing-masing 50 unit. 20 unit produk dari mesin A tidak memenuhi kualitas standard. Sedangkan dari mesin B, 30 unit tidak memenuhi kualitas standard. Apakah perbedaan kedua mesin nyata berdasarkan kualitas produk. Gunakan tingkat signifikansi 5 persen.

   Jawab: ya

8. Misalkan ada pernyataan bahwa satu dari tujuh pria yang hidup di kota besar mengidap penyakit hipertensi. Diambil sampel sebanyak 70 pria yang hidup di berbagai kota besar, ternyata 8 orang mengidap penyakit hipertensi, untuk menguji pernyataan tersebut. Benarkah pernyataan itu. Gunakan tingkat signifikansi 1 persen.

   Jawab: benar

9. Bagian pemasaran perusahaan televisi TAJAM mengharapkan bahwa 40 persen pengunjung Pasar Raya akan memperhatikan demonstrasi produk-produk televisi TAJAM. Setelah diadakan penelitian terhadap 100 pengunjung, ternyata 30 orang saja yang menyatakan memperhatikan demostrasi tersebut. Dapatkah kita menolak anggapan bagian pemasaran itu, dengan tingkat signifikansi (a) 10 persen dan (b) 5 persen.

   Jawab: a. ya b. ya

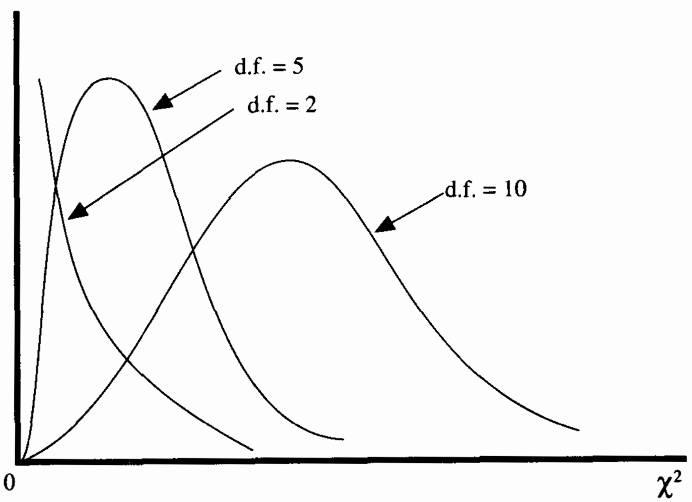
Bagian V Uji Chi Kuadrat

Uji Chi kuadrat merupakan pengujian hipotesis tentang perbandingan antara frekuensi sampel yang benar-benar terjadi (selanjutnya disebut dengan frekuensi observasi, dilambangkan dengan f_0) dengan frekuensi harapan yang didasarkan atas hipotesis tertentu pada setiap kasus atau data (selanjutnya disebut dengan frekuensi harapan, dilambangkan dengan f_e).

DISTRIBUSI CHI-KUADRAT

Ekspresi matematis tentang distribusi chi-kuadrat hanya tergantung pada satu parameter, yaitu derajat kebebasan (d.f.). Ada distribusi chi-kuadrat tertentu untuk masing-masing nilai derajat kebebasan. Misalnya, distribusi Z^2 (kuadrat standard normal) merupakan distribusi chi-kuadrart dengan d.f. = 1. Beberapa contoh distribusi chi-kuadrat yang lain dapat dilihat pada Gambar 5.1. Masing-masing distribusi merupakan distribusi probabilitas, sehingga luas di bawah kurva bernilai 1.

Gambar 5.1



Variabel χ^2 tidak bernilai negatip, sehingga kurva chi-kuadrat tidak mungkin berada di sebelah kiri nilai nol. Gambar 5.1 memperlihatkan bahwa untuk nilai d.f. lebih dari 2 kurva chi-kuadrat condong ke kanan (positip). Untuk nilai d.f. yang sangat besar kurva chi-kuadrat mendekati kurva normal.

Tingkat signifikansi α merupakan daerah di sisi kanan dari distribusi chi-kuadrat. Lambang $\chi^2_{\alpha,d.f.}$ menyatakan nilai χ^2 berarti distribusi chi-kuadrat dengan derajat kebebasan d.f. dan memiliki luas sebesar α pada daerah sisi kanan.

UJI KECOCOKAN

Dalam uji kecocokan *goodness of fit test*, hipotesis nol merupakan suatu ketentuan tentang pola yang diharapkan dari frekuensi-frekuensi dalam barisan kategori-kategori. Pola yang diharapkan harus sesuai dengan asumsi atau anggapan atas kemungkinan kejadian yang sama dan bersifat umum.

Untuk penerimaan hipotesis nol, perbedaan antara frekuensi observasi dengan yang diharapkan harus dapat dilambangkan dengan variabilitas secara sampling pada tingkat signifikansi yang diinginkan. Dengan demikian, uji Chi kuadrat didasarkan pada besarnya perbedaan dari masing-masing kategori dalam distribusi frekuensi. Nilai Chi-kuadrat untuk pengujian perbedaan antara pola frekuensi observasi dan frekuensi harapan adalah

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

dengan:
f_o : frekuensi observasi;
f_e : frekuensi harapan.

Dalam uji kecocokan derajat kebebasan (*degree of freedom, d.f.*) sama dengan jumlah kategori dikurangi jumlah estimator parameter yang didasarkan pada sampel dan dikurang 1. Jika dirumuskan menjadi:

d.f. = k - m - 1

dengan:
k : jumlah kategori data sampel;
m : jumlah nilai-nilai parameter yang diestimasi.

Jika hipotesis nol menyatakan bahwa frekuensi-frekuensi observasi didistribusikan sama dengan frekuensi harapan, tidak ada parameter estimator. Sehingga nilai m = 0.

Contoh 1

Sebuah distributor alat penggilingan padi membagi pasar menjadi 4 wilayah (A, B, C, dan D). Ada informasi bahwa pendistribusian alat penggilingan merata pada setiap wilayah. Untuk membuktikan pernyataan tersebut diambil 40 arsip sebagai sampel. Dari 40 arsip tersebut diperoleh informasi yang tertuang pada Tabel 5.1. Gunakan tingkat signifikansi 5 persen untuk pengujian.

Jawab:

1. Hipotesis
   H_0: distribusi alat penggilingan di keempat wilayah merata (sama)
   H_a: distribusi alat penggilingan di keempat wilayah tidak merata (tidak sama)
2. Nilai Kritis
   d.f. = k - m -1 = 4 - 0 - 1 = 3
   $\chi^2_{0,05,3} = 7,81$
3. Nilai Hitung
   Nilai uji statitik χ^2 adalah

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$= \frac{(6 - 10)^2}{10} + \frac{(12 - 10)^2}{10} + \frac{(14 - 10)^2}{10} + \frac{(8 - 10)^2}{10}$$

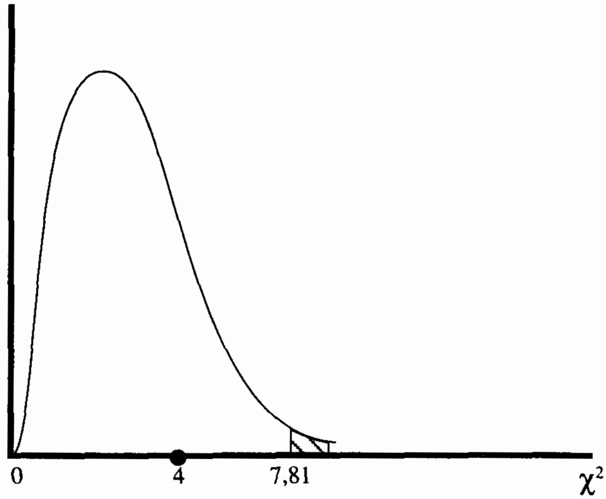
$$= \frac{40}{10} = 4,0$$

4. Simpulan
   Karena nilai statistik χ^2 sampel = 4,0 lebih kecil daripada nilai tabel $\chi^2_{0,05,3} = 7,81$ berarti kita tidak dapat menolak H_0 yang menyatakan bahwa distribusi alat penggilingan di keempat wilayah merata (sama).

Tabel 5.1
Distribusi Alat Penggilingan
di Daerah A, B, C, D

	Wilayah				Total
	A	B	C	D	
Distribusi berdasarkan sampel, f_0	6	12	14	8	40
Distribusi berdasarkan harapan, f_e	10	10	10	10	40

Gambar 5.2



Contoh 2

Misalkan departemen transmigrasi mengharapkan dapat menampung transmigran dari tiga daerah P, Q, dan R masing-masing tiga puluh keluarga (k). Namün kenyataan yang terjadi berasal dari daerah P, Q, dan R berturut-turut 23 k, 36 k, dan 31 k. Dengan data tersebut dapatkah disimpulkan bahwa harapan departemen transmigrasi tercapai. Gunakan $\alpha = 0{,}05$.

Jawab:

1. Hipotesis
   H_0: jumlah transmigran dari ketiga daerah sama (30 k)
   H_a: jumlah transmigran dari ketiga daerah tidak sama, ada yang dianggap tidak sama dengan 30 k
2. Nilai Kritis
   d.f. = k - m -1 = 3 - 0 - 1 = 2
   $\chi^2_{0,05;2} = 5{,}991$
3. Nilai Hitung
   Nilai uji statitik χ^2 adalah

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

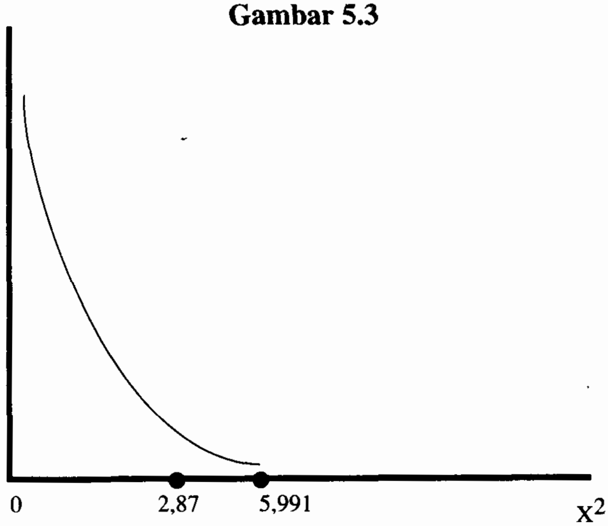
$$= \frac{(23-30)^2}{30} + \frac{(36-30)^2}{30} + \frac{(31-30)^2}{30}$$

$$= \frac{86}{30} = 2{,}87$$

4. Simpulan

Karena nilai statistik χ^2 sampel = 2,87 lebih kecil daripada nilai tabel $\chi^2_{0,05,2}$ = 5,991 berarti kita tidak dapat menolak H_0 yang menyatakan bahwa jumlah transmigran dari ketiga daerah sama (30 k).

Gambar 5.3



Contoh 3

Untuk tujuan-tujuan perencanaan dan pengontrolan suku cadang, perusahaan Chemical Company ingin mengetahui apakah produk suatu cairan kimia berdistribusi normal. Hasil penjualan yang diambil secara random selama 200 hari disajikan pada Tabel 5.3. Rata-rata sampel dan standard deviasi yang dihitung dari 200 sampel jumlah penjualan harian berturut-

Tabel 5.3
Penjualan Selama 200 Hari

Penjualan (dalam ribuan gallon)	Jumlah hari, f
kurang dari 34	0
34,0 dan kurang dari 35,5	13
35,5 dan kurang dari 37,0	20
37,0 dan kurang dari 38,5	35
38,5 dan kurang dari 40,0	43
40,0 dan kurang dari 41,5	51
41,5 dan kurang dari 43,0	27
43,0 dan kurang dari 44,5	10
44,5 dan kurang dari 46,0	1
46,0 atau lebih	0

turut 40.000 gallon dan 2.500 gallon. Nilai-nilai tersebut sebagai parameter estimasi terhadap π dan σ. Pada tingkat signifikansi 5 persen dapatkan disimpulkan bahwa pola penjualan berbentuk distribusi normal?

Jawab:

Untuk menentukan frekuensi harapan kita harus menghitung terlebih dahulu probabilitas. Berikut ini adalah satu contoh perhitungan.

Untuk nilai X = 34

$$Z = \frac{X - \mu}{} = \frac{34 - 40}{2,5} = -2,4$$

Dengan melihat tabel Z kita dapatkan:

P(0 samapi - 2,4) = P(0 sampai 2,4) = 0,4918

Untuk kelas kurang dari 34,0 kita dapatkan nilai probabilitasnya, yaitu:

0,5 - P(0 sampai - 2,4) = 0,5 - 0,4918 = 0,0082

Karena besarnya sampel 200, frekuensi harapannya adalah

f_e = 0,0082(200) = 1,64

Dengan cara yang sama kita dapat menghitung frekuensi-frekuensi harapan kelas-kleas yang lain. Hasil selengkapnya dapat dilihat pada Tabel 6.4.

Jika ditentukan nilai frekuensi harapan terkecil 5, kita dapatkan nilai frekuensi harapan sebanyak 8. Ini terjadi dengan menjadikan satu kelas I dan II, serta menjadikan satu kelas IX dan X. Nilai frekuensi harapan yang baru berturut-turut 7,18 (= 1,64 + 5,54) dan 7,18 (= 5,54 + 1,64).

1. Hipotesis

   H_0: penyebaran data berdistribusi normal
   H_a: penyebaran data tidak berdistribusi normal

2. Nilai kritis

   Dengan adanya parameter estimasi rata-rata dan standard deviasi nilai m = 2. Dengan demikian demikian d.f. = 8 -2-1 = 5. Apabila kita gunakan σ = 0,05 nilai χ^2 tabel adalah
   $\chi^2_{0,05;5}$ = 11,070

3. Nilai hitung

   Nilai statistik χ^2 sampel (*lihat Tabel 5.4*)
   χ^2 = 15,1940

4. Simpulan

Nilai statistik χ^2 sampel = 15,1940 lebih besar daripada nilai $\chi^2_{0,05,5}$ = 11,070. Kita tolak H_0. Ini berarti penyebaran data hasil penjualan harian tidak berdistribusi normal.

Tabel 5.4

Penjualan Selama 200 Hari Dengan Frekuensi Harapan

Penjualan (ribuan gallon)	f_o	Probabilitas (P)	200 P f_e	f_o-f_e	$(f_o-f_e)^2/f_e$
34	0	0,0082	1,64	5,82	4,7176
34,0 dan < 35,5	13	0,0277	5,54		
35,5 dan < 37,0	20	0,0792	15,84	4,16	1,0926
37,0 dan < 38,5	35	0,1592	31,84	3,16	0,3136
38,5 dan < 40,0	43	0,2257	45,14	–2,14	0,1015
40,0 dan < 41,5	51	0,2257	45,14	5,86	0,7607
41,5 dan < 43,0	27	0,1592	31,84	–4,84	0,7357
43,0 dan < 44,5	10	0,0792	15,84	–5,84	2,1531
44,5 dan < 46,0	1	0,0277	5,54	–6,18	5,3193
46,0 atau lebih	0	0,0082	1,64		
				χ^2 =	15,1940

UJI TABEL KONTINGENSI

Tabel kontingensi memuat data yang diperoleh dari sampel random sederhana dan diatur berdasarkan baris dan kolom. Nilai-nilai data tersebut dinamakan frekuensi observasi (f_0).

Dengan uji tabel kontingensi (*contingency table test*) kita dapat menguji apakah dua variabel saling independen. Gagasan ini didasarkan atas anggapan bahwa nilai frekuensi observasi mendekati nilai frekuensi harapan jika kategori-kategori independen. Perbedaan-perbedaan yang besar akan mendukung kita untuk menolak hipotesis independensi.

Apabila banyak baris = r, banyak kolom = k, dan besarnya sampel n, nilai frekuensi harapan baris ke i dan kolom ke j dapat diperoleh dengan rumus:

$$f_{e_{ij}} = \frac{(\Sigma f_{0_i})(E f_{0_j})}{n}$$

dengan derajat kebebasan:

d.f. = (r - 1)(k - 1)

Sedangkan rumus untuk memperoleh nilai χ^2 adalah

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

Contoh 1

Tabel 5.2 menunjukkan pengunjung pada salon TAMPAN pada tanggal 12 Oktober 1993 yang dikategorikan berdasarkan jenis kelamin dan umur. Ujilah hipotesis bahwa jenis kelamin dan umur pengunjung adalah independen dengan tingkat signifikansi $\alpha = 0{,}01$.

Tabel 5.2
Pengunjung Tanggal 12 Oktober 1993

Umur	**Jenis kelamin**		**Total baris**
	Pria	**Wanita**	
Di bawah 30	60	50	110
30 atau lebih	80	10	90
Total kolom	140	60	200

Jawab:

1. Hipotesis
   H_0: jenis kelamin dan umur pengunjung adalah independen
   H_a: jenis kelamin dan umur pengunjung adalah tidak independen
2. Nilai Kritis
   Derajat kebebasan d.f. adalah
   d.f. = (r - 1)(k - 1) = (2 - 1)(2 - 1) = 1
   Nilai uji statistik $\chi^2_{0,01,1} = 6{,}63$
3. Nilai Hitung
   Berikut ini contoh perhitungan nilai frekuensi harapan pada sel jenis kelamin pria dan umur di bawah 30.

$$f_{e_{11}} = \frac{(\Sigma f_{e_1})(\Sigma f_{e_1})}{n} = \frac{(110)(140)}{200} = \frac{15400}{200} = 77$$

   Hasil perhitungan frekuensi harapan selengkapnya dapat dilihat pada Tabel 5.3

Tabel 5.3
Frekuensi Harapan Pengunjung
Tanggal 12 Oktober 1993

Umur	Jenis kelamin		Total baris
	Pria	Wanita	
Di bawah 30 30 atau lebih	77 63	33 27	110 90
Total kolom	140	60	200

Nilai statistik χ^2 sampel adalah

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$= \frac{(60 - 70)^2}{77} + \frac{(50 - 33)^2}{33} + \frac{(80 - 63)^2}{63} + \frac{(10 - 27)^2}{27}$$

$$= 27{,}8$$

4. Simpulan
Dengan tingkat sigifikansi 1 persen H_0 ditolak karena nilai statistik χ^2 sampel = 27,8 lebih besar daripada $\chi^2_{0{,}01{,}1} = 6{,}63$. Ini berarti bahwa jenis kelamin dan umur pengunjung tidak independen adalah benar.

Contoh 2

Dony, pengusaha anggur, telah mengumpulkan opini tentang kualitas buah anggur secara random dari pelanggan-pelanggannya. Pelanggan-pelanggan mencoba anggur-anggur yang dibuat dari buah-buah anggur yang dihasilkan oleh tiga daerah. Kualitas dibedakan menjadi empat kualifikasi pada tingkatan 1 (terbaik) sampai dengan 4. Data sampel diberikan pada Tabel 5.4. Dony ingin mengetahui apakah tingkat kualitas independen atas daerah penghasil buah anggur. Gunakan tingkat signifikansi 5 persen.

Tabel 5.4

Tingkat kualitas	Daerah			Total baris
	I	II	I	
1	15	10	6	31
2	7	13	12	32
3	11	12	8	31
4	3	8	15	26
Total kolom	36	43	41	120

Jawab:

Perhitungan dengan program *microstat*:

——————— CROSSTAB / CHI-SQUARE TESTS ———————

Kualitas Buah Anggur Berdasarkan Daerah Penghasil

OBSERVED VALUES (Cell format: count/ percent:total/ percent:row/ percent:col)

	I	II	III	TOTAL
1	15 12.50 48.39 41.67	10 8.33 32.26 23.26	6 5.00 19.35 14.63	31 25.83
2	7 5.83 21.88 19.44	13 10.83 40.63 30.23	12 10.00 37.50 29.27	32 26.67
3	11 9.17 35.48 30.56	12 10.00 38.71 27.91	8 6.67 25.81 19.51	31 25.83
4	3 2.50 11.54 8.33	8 6.67 30.77 18.60	15 12.50 57.69 36.59	26 21.67
TOTAL	36 30.00	43 35.83	41 34.17	120 100.00

CHI-SQUARE = 14.976 D.F.= 6, PROB. = .0204

1. Hipotesis
   H_0: tingkat kualitas independen atas daerah penghasil
   H_a: tingkat kualitas tidak independen atas daerah penghasil

2. Nilai kritis

   Derajat kebebasan diperoleh dengan rumus:

   d.f. = (r - 1)(k - 1) = (4 - 1)(3 - 1) = 6

   Nilai χ^2 tabel adalah

   $\chi^2_{0,05,6} = 12{,}592$

3. Nilai hitung

   Dari hasil hitungan dengan program *microstat* diperoleh nilai statistik χ^2 sampel:
   $\chi^2 = 14{,}976$

4. Simpulan
   Karena nilai statistik χ^2 sampel = 14,976 lebih besar daripada $\chi^2_{0,05,6} = 12{,}592$ H_0 ditolak. Dengan demikian tingkat kualitas tidak independen atas daerah penghasil.

PENGUJIAN PERBEDAAN DI ANTARA k PROPORSI

Pengujian chi-kuadrat dapat digunakan untuk menguji kesamaan dari dua proporsi atau lebih. Pengujian kesamaan proporsi sama dengan pengujian independensi.

Rumus untuk memperoleh nilai χ^2 adalah

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

a. Uji Proporsi yang Dihipotesakan
   H_0: $\pi = \pi_0$ H_A: $\pi \neq \pi_0$
   π_0: nilai proporsi yang dihipotesakan
   d.f. = k - m -1

b. Uji Beda Dua Proporsi
   H_0: $\pi_1 = \pi_2$ H_A: $\pi_1 \neq \pi_2$
   d.f. = (r -1)(k - 1)

c. Uji Beda k Proporsi
   H_0: $\pi_1 = \pi_2 = \pi_3 = \pi_4$ H_A: hipotesis nol tidak benar.
   d.f. = (r -1)(k - 1)

Contoh 1

Tabel 5.5 berisi tentang data-data pengambilan sampel random dari 200 pekerja. Tabel tersebut memperlihatkan, misalnya, bahwa pekerja yang tidak sekolah (tingkat pendidikan 1) merasa puas terhadap supervisor. Kita akan menguji (pada tingkat $\alpha = 0{,}05$) hipotesis 140

Tabel 5.5

Tingkat kualitas	Tingkat pendidikan			Total baris
	1	2	3	
Puas	12	63	65	140
Tidak puas	8	17	35	60
Total kolom	20	80	100	200

Jawab:
Perhitungan dengan program *microstat*:

CROSSTAB / CHI-SQUARE TESTS
Kepuasan Pekerja Terhadap Supervisor Berdasarkan Pendidikan

OBSERVED VALUES (Cell format: count/ percent:total/ percent:row/ percent:col)

	1	2	3	TOTAL
Puas	12 6.00 8.57 60.00	63 31.50 45.00 78.75	65 32.50 46.43 65.00	140 70.00
Tidak puas	8 4.00 13.33 40.00	17 8.50 28.33 21.25	35 17.50 58.33 35.00	60 30.00
TOTAL	20 10.00	80 40.00	100 50.00	200 100.00

CHI-SQUARE = 5.060, D.F.= 2, PROB. = .0797

1. Hipotesis
   H_0: proporsi sel dalam baris tertentu adalah sama
   H_a: proporsi sel dalam sekurang-kurangnya satu baris tidak sama
2. Nilai kritis
   Derajat kebebasan d.f. = (r - 1)(k - 1) = (2 - 1)(3 - 1) = 2

Dengan $\alpha = 0{,}05$ dan d.f. = 2 nilai χ^2 adalah

$\chi^2_{0,05,2} = 5{,}991$

3. Nilai hitung

   Dari hitungan dengan program *microstat* diperoleh:

   $\chi^2 = 5{,}060$

4. Simpulan

   Karena nilai statistik χ^2 sampel $= 5{,}060$ lebih kecil dari nilai $\chi^2_{0,05,2} = 5{,}991$, berarti H_0 kita terima. Dengan demikian hipotesis proporsi sel dalam baris tertentu adalah sama adalah benar.

SOAL-SOAL

1. Nilai statistika dari 100 mahasiswa terlihat pada Tabel A. Apakah kita dapat mengatakan bahwa distribusi nilai terbagi rata? Ujilah dengan tingkat signifikansi 1 persen!

Tabel A

Nilai				
A	**B**	**C**	**D**	**Total**
30	20	40	10	100

Jawab: Distribusi nilai tidak terbagi rata.

2. Dari data pada Tabel B didapat rata-rata umur peserta kursus 23,3 tahun dengan standard deviasi 3,4 tahun. Dengan tingkat signifikansi 1 persen, apakah frekuensi dari umur tersebar mengikuti pola distribusi probabilitas normal?

Tabel B

Umur	**Jumlah Peserta**
18 dan kurang dari 20	5
20 dan kurang dari 22	18
22 dan kurang dari 24	10
24 dan kurang dari 26	6
26 dan kurang dari 28	5
28 dan kurang dari 30	4
30 dan kurang dari 32	2
Total	50

Jawab: tidak mengikuti pola distribusi probabilitas normal.

3. Dengan menggunakan data sampel random pada Tabel C, ujilah hipotesis bahwa pilihan jenis gula-gula tidak tergantung pada usia. Dengan kata lain antara jenis gula-gula dan usia saling independen! Gunakan $\alpha = 0{,}05$!

Tabel C

Jenis Gula-gula	**Umur**		
	<18	**18-40**	**>40**
Sweet fruit	50	35	35
Tart fruit	19	32	30
Mint	45	38	16

Jawab: pilihan jenis gula-gula tergantung pada usia.

4. Seorang analis keselamatan mengelompokkan stok dari kelompok industri (empat kategori, yaitu I, II, III, IV) menjadi tiga kelompok berdasarkan tingkat keamanannya, yaitu tinggi, rata-rata, dan rendah. Data selengkapnya dapat dilihat pada Tabel D. Ujilah independensi antara tingkat-tingkat keamanan dan klasifikasi industri! Gunakan $\alpha = 0{,}01$.

Tabel D

Tingkat Keselamatan Stok	**Kategori Industri**			
	I	**II**	**III**	**IV**
Tinggi	25	18	29	12
Rata-rata	32	30	42	20
Rendah	13	32	14	33

Jawab: tidak independen.

5. Misalkan suatu departemen menerima laporan yang dibagi ke dalam 2 kelompok (memenuhi syarat dan tidak memenuhi syarat) selama satu tahun tertentu yang dibagi per 4 bulan (I, II, dan III). Data laporan terdapat pada Tabel E. Ujilah proporsi kelompok yang tidak memenuhi syarat dari periode ke periode sama, dengan tingkat signifikansi

Tabel E

Laporan	**Periode**		
	I	**II**	**III**
Memenuhi syarat	427	273	240
Tidak Memenuhi Syarat	23	27	10

Jawab: *proporsi yang tidak memenuhi syarat dari periode ke periode sama.*

6. Sebuah perusahaan yang memiliki empat lokasi produksi ingin mengetahui apakah proporsi tenaga kerja terampil pada masing lokasi sama. Data untuk menyelidiki proporsi tenaga kerja terampil terdapat pada Tabel F. Gunakan tingkat signifikansi 5 persen untuk pengujian!

Tabel F

Tenaga Kerja	Lokasi			
	A	**B**	**C**	**D**
Terampil	89	99	193	82
Tidak Terampil	161	151	148	168

Jawab: *proporsi tenaga terampil pada masing-masing lokasi sama.*