Statistika dan Probabilitas
Minggu ke 14: **Uji Hipotesis**

Haryadi
haryadi_ump@yahoo.co.id

Universitas Muhammadiyah Palangka Raya
December 16, 2023

Peneliti sering menyatakan suatu sifat populasi berdasarkan informasi sampel. Suatu pernyataan atau klaim tentang parameter populasi dinamakan **hipotesis statistik**. Karena didasarkan pada data sampel, maka hipotesis tersebut bisa benar atau bisa tidak benar.

Definisi 1. Uji hipotesis *adalah suatu aturan dimana setelah data sampel diperoleh maka akan menuntun kepada diterima atau ditolaknya suatu hipotesis.*

Suatu hipotesis yang akan diuji dinamakan **hipotesis nol**, dituliskan H_0. Hipotesis yang berbeda dengan hipotesis nol dinamakan **hipotesis alternatif**, dituliskan H_1.

Suatu hipotesis yang jika benar ternyata menggambarkan secara lengkap distribusi populasi, dinamakan **hipotesis sederhana**. Jika tidak demikian maka dinamakan **hipotesis komposit**. Sebagai contoh perhatikan dua hipotesis berikut.

(a) Rata-rata berat badan mahasiswa UM Palangkaraya adalah $\mu = 60$

(b) Rata-rata berat badan mahasiswa UM Palangkaraya adalah $\mu \leq 60$

Hipotesis (a) merupakan hipotesis sederhana, sedangkan hipotesis (b) merupakan hipotesis komposit. Perhatikan bahwa jika hipotesis (a) benar, maka nilai parameter populasi secara tegas dinyatakan dengan $\mu = 60$; sebaliknya pada hipoteis (b) jika benar maka nilai parameter populasi μ tidak dapat tergambar secara tegas.

Contoh 1. Kita klaim bahwa rata-rata berat badan mahasiswa UM Palangkaraya adalah 60 kg. Untuk membuktikan benar atau tidaknya klaim tersebut, maka perlu diuji. Dalam hal ini hipotesis nolnya adalah "rata-rata berat badan mahasiswa UM Palangkaraya adalah 60 kg". Hipotesis alternatifnya misalnya adalah "rata-rata berat badan mahasiswa UM Palangkaraya tidak sama dengan 60 kg". Hipotesis alternatif lainnya misalnya adalah "rata-rata berat badan mahasiswa UM Palangkaraya adalah kurang dari 60 kg".

Perhatikan kembali contoh 1. Jika hipotesis nolnya adalah "rata-rata berat badan mahasiswa UM Palangkaraya adalah 60 kg" dan hipotsis alternatifnya adalah "rata-rata berat badan mahasiswa UM Palangkaraya tidak sama dengan 60 kg", maka didituliskan

$$H_0 : \quad \mu = 60$$

melawan hipotesis

$$H_1 : \quad \mu \neq 60.$$

Karena H_0 dan H_1 merupakan pernyataan yang komplementer, maka mudah dipahami bahwa jika kita **menerima** H_0 tentu kita **tidak menerima** H_1 atau **menolak** H_1. Demikian pula jika kita menolak H_0 berarti kita menerima H_1. Dari sini diperlukan batasan yang jelas bagaimana suatu hipotesis diterima atau ditolak.

Untuk menguji hipotests H_0 diperlukan data sampel. Misalkan diambil sampel berukuran n yang nilai-nilai observasinya $X_1, X_2, \cdots, X_n$. Perhatikan bahwa $(X_1, X_2, \cdots, X_n)$ merupakan titik pada ruang berdimensi n. Nilai-nilai tersebut akan digunakan sebagai dasar diterimanya atau ditolaknya H_0. Selanjutnya dibentuk daerah berdimeni n yang dinamakan **daerah kritis**, dituliskan dengan C. Ini berarti uji statistik yang ditentukan oleh daerah kritis adalah

$$\text{terima } H_0 \quad \text{jika} \quad (X_1, X_2, \cdots, X_n) \notin C,$$

dan

$$\text{tolak } H_0 \quad \text{jika} \quad (X_1, X_2, \cdots, X_n) \in C.$$

Karena data yang digunakan untuk menerima atau menolak suatu hipotesis adalah data sampel, maka tidak dapat dipastikan apakah hipotesisi tersebut benar atau salah. Dalam pengambilan kesimpulan, kalaupun ada kesalahan tentu kita berharap kesalahan tersebut sekecil mungkin. **Menolak** suatu **pernyataan yang benar** tentu merupakan suatu kesalahan. Demikian pula, **menerima** suatu **pernyataan yang salah** tentu merupakan suatu kesalahan.

Definisi 2. *Kesalahan jenis I adalah ditolaknya H_0 padahal H_0 benar. Kesalahan jenis II adalah diterimanya H_0 padahal H_0 salah.*

Definisi 3. Tingkat signifikansi *suatu uji hipotesis, dituliskan α, adalah peluang terjadinya kesalahan jenis I.*

Dengan demikian, jika C adalah daerah kritis, maka

$$\alpha = P(H_0 \text{ ditolak padahal } H_0 \text{ benar }).$$

1 Uji tentang mean populasi normal

1.1 Uji hipotesis dengan σ^2 diketahui

Diketahui $X_1, X_2, \cdots, X_n$ adalah sampel random dari populasi berdistribusi normal dengan mean μ dan varian σ^2 yang diketahui. Misalkan akan diuji hipotesis

$$H_0 : \quad \mu = \mu_0$$

melawan hipotesis

$$H_1 : \quad \mu \neq \mu_0$$

dimana μ_0 suatu konstanta.

Karena $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ adalah estimator untuk μ, maka cukup beralasan bahwa H_0 diterima jika $\overline{X}$ tidak berbeda jauh dengan μ_0. Oleh karena itu daerah kritis hipotesis ini adalah

$$C = \{(X_1, X_2, \cdots, X_n) : |\overline{X} - \mu_0| > c\} \tag{1}$$

dengan c suatu konstanta.

Jika akan diuji hipotesis pada tingkat signifikansi α, maka nilai c pada (1) harus dicari sehingga kesalahan jenis pertamanya adalah α, yaitu dicari c sehingga

$$P(|\overline{X} - \mu_0| > c) = \alpha \tag{2}$$

Dengan membagi kedua ruas ketaksamaan $|\overline{X} - \mu_0| > c$ dengan $\sigma/\sqrt{n}$ diperoleh

$$\left| \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \right| > \frac{c\sqrt{n}}{\sigma}$$

Karena $\overline{X}$ berdistribusi normal dengan mean μ dan varian σ^2/n, maka

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

berdistribusi normal standar. Oleh karena itu persamaan (2) dapat dituliskan menjadi

$$P\left(|Z| > \frac{c\sqrt{n}}{\sigma}\right) = \alpha \tag{3}$$

Karena $|Z| > \frac{c\sqrt{n}}{\sigma}$ ekivalen dengan $-Z < \frac{c\sqrt{n}}{\sigma}$ atau $\frac{c\sqrt{n}}{\sigma} < Z$, maka

$$P\left(|Z| > \frac{c\sqrt{n}}{\sigma}\right) = P\left(-Z < \frac{c\sqrt{n}}{\sigma}\right) + P\left(\frac{c\sqrt{n}}{\sigma} < Z\right) = 2P\left(\frac{c\sqrt{n}}{\sigma} < Z\right).$$

Oleh karena itu

$$P\left(\frac{c\sqrt{n}}{\sigma} < Z\right) = \frac{\alpha}{2}$$

Karena Z berdistribusi normal standar, maka

$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2},$$

atau

$$c = \frac{\sigma z_{\alpha/2}}{\sqrt{n}}.$$

Dengan demikian H_0 ditolak jika $|\overline{X} - \mu_0| > \frac{\sigma z_{\alpha/2}}{\sqrt{n}}$ atau

$$\frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| > z_{\alpha/2}.$$

Uraian di atas dirangkum dengan aturan berikut.

Aturan 1. Untuk menguji hipotesis

$$H_0 : \mu = \mu_0$$

melawan hipotesis

$$H_1 : \mu \neq \mu_0$$

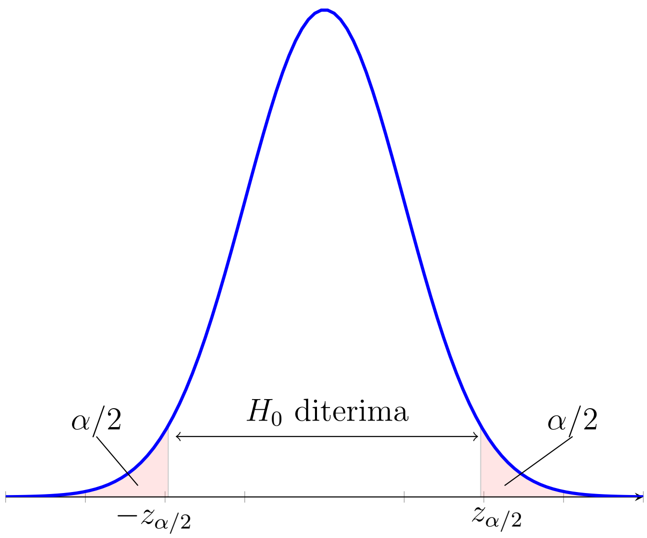
digunakan aturan

$$\begin{aligned} \textit{terima} \quad H_0 \quad \textit{jika} \quad z &= \frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| \leq z_{\alpha/2} \\ \textit{tolak} \quad H_0 \quad \textit{jika} \quad z &= \frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| > z_{\alpha/2}. \end{aligned}$$

dengan $z_{\alpha/2}$ diperoleh dari tabel normal standar sehingga

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Daerah kritis penerimaan hipotesis disajikan pada Gambar 1.



Gambar 1: Daerah kritis

Contoh 2. Akan diuji suatu pernyataan bahwa rata-rata jumlah anak per KK di Palangkaraya adalah 3. Diambil sampel random berukuran 100 dan diperoleh rata-rata sampel $\overline{x} = 2.84$. Jika diketahui $\sigma = 0.8$ akan diuji hipotesis tersebut pada tingkat signifikansi $\alpha = 0.05$. Dalam hal ini hipotesis yang akan diuji

$$H_0 : \mu = 3$$

melawan hipotesis

$$H_1 : \mu \neq 3.$$

Penyelesaian. Berdasarkan tabel normal standar, $z_{\alpha/2} = z_{0.025} = 1.96$. Diperoleh

$$z = \frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| = \frac{\sqrt{100}}{0.8}\left|2.84 - 3\right| = 2.0.$$

Karena $z > z_{0.025}$ ini berarti H_0 ditolak dan disimpulkan bahwa rata-rata banyaknya anak per KK di Palangkaraya tidak sama dengan 3.

Dapat terjadi $H_0 : \mu \leq \mu_0$ dan hipotesis alternatifnya adalah $H_1 : \mu > \mu_0$. Untuk menguji hipotesis demikian kita gunakan aturan berikut.

Aturan 2. Untuk menguji hipotesis:

$$H_0 : \mu \leq \mu_0$$

melawan hipotesis

$$H_1 : \mu > \mu_0$$

digunakan aturan

$$\begin{aligned} &terima \quad H_0 \quad jika \quad z = \frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu_0\right) \leq z_\alpha \\ &tolak \quad H_0 \quad jika \quad z = \frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu_0\right) > z_\alpha \end{aligned}$$

Demikian pula dapat terjadi $H_0 : \mu \geq \mu_0$ dan hipotesis alternatifnya adalah $H_1 : \mu < \mu_0$.

Aturan 3. Untuk menguji hipotesis:

$$H_0 : \mu \geq \mu_0$$

melawan hipotesis

$$H_1 : \mu < \mu_0$$

digunakan aturan

$$\begin{aligned} &terima \quad H_0 \quad jika \quad z = \frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu_0\right) \geq -z_\alpha \\ &tolak \quad H_0 \quad jika \quad z = \frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu_0\right) < -z_\alpha \end{aligned}$$

Uji hipotesis pada Aturan 2 dan 3 dinamakan uji hipotesis satu sisi.

Contoh 3. Semua rokok yang beredar di pasaran mengandung nikotin paling sedikit 1.6 mg per batang rokok. Suatu perusahaan rokok mengklaim dengan suatu metode tertentu dapat menurunkan kadar nikotin kurang dari 1.6 mg per batang rokok. Untuk menguji klaim tersebut, sampel berukuran 20 dari perusahaan tersebut dianalisis. Diketahui devisi standar nikotin pada rokok adalah 0.8 mg. Jika rata-rata nikotin ke 20 rokok pada sampel tersebut adalah 1.54, apakah klaim perusahaan tersebut dapat diterima pada tingkat signifikansi 5 persen?

Penyelesaian. Akan diuji hipotesis

$$H_0 : \mu \geq 1.6 \quad \text{melawan} \quad H_1 : \mu < 1.6.$$

Berdasarkan tabel normal diperoleh $z_{-\alpha} = z_{-0.05} = -1.65$. Berdasarkan data di atas diperoleh

$$z = \frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu_0\right) = \frac{\sqrt{20}}{0.8}(1.54 - 1.6) = -0.336.$$

Karena $z \geq z_{-0.05}$ maka H_0 diterima, yang berarti klaim perusahaan tersebut tidak benar pada tingkat signifikansi 5 persen.

1.2 Uji hipotesis dengan σ^2 tidak diketahui

Dalam situasi yang lebih umum, mean populasi μ dan varian populasi σ^2 biasanya tidak diketahui.

Aturan 4. Untuk menguji hipotesis

$$H_0 : \mu = \mu_0$$

melawan hipotesis

$$H_1 : \mu \neq \mu_0$$

digunakan aturan

$$\begin{aligned} &\textit{terima} \quad H_0 \quad \textit{jika} \quad t = \frac{\sqrt{n}}{S}\left|\overline{X} - \mu_0\right| \leq t_{\alpha/2,n-1} \\ &\textit{tolak} \quad H_0 \quad \textit{jika} \quad t = \frac{\sqrt{n}}{S}\left|\overline{X} - \mu_0\right| > t_{\alpha/2,n-1} \end{aligned}$$

dimana S^2 adalah varian sampel dan n ukuran sampel.

Contoh 4. Pemerintah mengklaim bahwa kebutuhan air bersih rata-rata rumah tangga adalah 350 galon per hari. Untuk membuktikan klaim tersebut, suatu studi terhadap 20 rumah tangga dilaksanakan dan diperoleh data berikut.

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Berdasarkan data sampel tersebut, apakah klaim pemerintah dapat diterima pada tingkat signifikansi 10 persen?

Penyelesaian. Akan diuji hipotesis

$$H_0 : \mu = 350 \quad \text{melawan} \quad H_1 : \mu \neq 350.$$

Berdasarkan data tersebut diperoleh

$$\overline{X} = 353.8 \quad \text{dan} \quad S = 21.8478.$$

Dengan demikian

$$t = \frac{\sqrt{n}}{S}\left|\overline{X} - \mu_0\right| = \frac{\sqrt{20}}{21.8478}|353.8 - 350| = 0.7778$$

Berdasarkan tabel t diperoleh $t_{\alpha/2,n-1} = t_{0.05,19} = 1.729$. Karena $t < t_{0.05,19}$ berarti H_0 diterima, yang berarti bahwa klaim pemerintah dapat diterima dengan tingkat signifikansi 10 persen.

Uji hipotesis satu sisi untuk varian populasi σ tidak diketahui diberikan dalam aturan-aturan berikut.

Aturan 5. Uji hipotesis satu sisi:

$$H_0 : \mu \leq \mu_0$$

melawan hipotesis

$$H_1 : \mu > \mu_0$$

digunakan aturan

$$\begin{aligned} terima \quad H_0 \quad jika \quad t = \frac{\sqrt{n}}{S}\left(\overline{X} - \mu_0\right) \leq t_{\alpha,n-1} \\ tolak \quad H_0 \quad jika \quad t = \frac{\sqrt{n}}{S}\left(\overline{X} - \mu_0\right) > t_{\alpha,n-1} \end{aligned}$$

dimana S^2 adalah varian sampel dan n ukuran sampel.

Aturan 6.

$$H_0 : \mu \geq \mu_0$$

melawan hipotesis

$$H_1 : \mu < \mu_0$$

digunakan aturan

$$\begin{aligned} terima \quad H_0 \quad jika \quad t = \frac{\sqrt{n}}{S}\left(\overline{X} - \mu_0\right) \geq -t_{\alpha,n-1} \\ tolak \quad H_0 \quad jika \quad t = \frac{\sqrt{n}}{S}\left(\overline{X} - \mu_0\right) < -t_{\alpha,n-1} \end{aligned}$$

dimana S^2 adalah varian sampel dan n ukuran sampel.

Contoh 5. Hasil pengukuran *delay* pada suatu jaringan (dalam *mili second*) adalah sebagai berikut:
148, 165, 200, 145, 130, 197, 188, 170, 155, 143, 146, 156, 120, 120, 124.

Menurut stadard TIPHON, ketegori "sangat baik" untuk parameter delay

adalah jika delay kurang dari 150 ms. Berdasarkan data hasil pengamatan di atas, ujilah apakah kategory delay pada jaringan tersebut "sangat bagus"?

Penyelesaian. Berdasarkan data di atas, mean dan standard deviasi sampelnya adalah $\bar{x} = 153.8$ dan $s = 26.16$. Dengan $\mu_0 = 150$ dan $n = 15$, diperoleh

$$t = \frac{\sqrt{15}}{26.16}(153.8 - 150) = 0.563$$

Selanjutnnya dari tabel t dengan $\alpha = 0.05$ diperoleh $t_{0.05,14} = 1.761$. Karena $t < t_{0.05,14}$ maka disimpulkan bahwa H_0 diterima yang berarti benar bahwa delaynya termasuk kategori sangat baik.

2 Uji kesamaan mean dua populasi

Jika peneliti ingin mengetahui efek suatu jenis obat dalam menyembuhkan suatu penyakit, maka setidaknya dia mencoba obat tersebut ke suatu sampel dan membandingkan hasilnya dengan sampel lain yang tidak diberi obat, yakni kita memiliki dua sampel yang mendapat perlakuan berbeda.

Untuk mengetahui apakah suatu metode belajar tertentu memberikan hasil yang berbeda dengan metode belajar yang ada, tentu peneliti harus mencobakan metode tertentu tersebut dan membandingkan hasilnya dengan metode yang ada.

Dari kedua contoh di atas, yang akan diselidiki dapat berupa kesamaan atau perbedaan mean kedua populasi.

2.1 Varian populasi diketahui

Misalkan $X_1, X_2, \cdots, X_n$ dan $Y_1, Y_2, \cdots, Y_m$ adalah dua sampel independen dari dua populasi yang berdistribusi normal, yang masing-masing memiliki mean μ_x dan μ_y yang tidak diketahui, dan varian σ_x^2 dan σ_y^2 yang diketahui. Andaikan akan diuji hipotesis

$$H_0 : \mu_x = \mu_y$$

melawan hipotesis

$$H_1 : \mu_x \neq \mu_y$$

Karena $\overline{X}$ adalah estimator μ_x dan $\overline{Y}$ adalah estimator μ_y, maka $\overline{X} - \overline{Y}$ dapat digunakan untuk mengestimasi $\mu_x - \mu_y$. Karena H_0 dapat dituliskan $H_0 : \mu_x - \mu_y = 0$, maka cukup beralasan untuk menerima H_0 bilamana $\overline{X} - \overline{Y}$ tidak berbeda jauh dengan 0. Selanjutnya, karena $\overline{X} - \overline{Y}$ berdistribusi normal dengan mean $\mu_x - \mu_y$ dan varian $\sigma_x^2/n + \sigma_y^2/m$, maka

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

berdistribusi normal standar. Ini berakibat jika H_0 benar, yakni $\mu_x - \mu_y = 0$, maka

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

berdistribusi normal standar. Dengan demikian belaku

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \leq z_{\alpha/s}\right) = 1 - \alpha.$$

Aturan 7. Untuk menguji

$$H_0 : \mu_x = \mu_y$$

melawan hipotesis

$$H_1 : \mu_x \neq \mu_y$$

dengan tingkat signifikansi α digunakan aturan

$$\begin{aligned} terima \quad H_0 \quad jika \quad & z = \frac{\left|\overline{X} - \overline{Y}\right|}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \leq z_{\alpha/2} \\ tolak \quad H_0 \quad jika \quad & z = \frac{\left|\overline{X} - \overline{Y}\right|}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} > z_{\alpha/2}. \end{aligned}$$

2.2 Varian populasi tidak diketahui

Misalkan $X_1, X_2, \cdots, X_n$ dan $Y_1, Y_2, \cdots, Y_m$ adalah dua sampel independen dari dua populasi yang berdistribusi normal, yang masing-masing memiliki mean μ_x dan μ_y yang tidak diketahui, dan varian σ_x^2 dan σ_y^2 juga tidak diketahui. Lebih lanjut dimisalkan $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Andaikan akan diuji hipotesis

$$H_0 : \mu_x = \mu_y$$

melawan hipotesis

$$H_1 : \mu_x \neq \mu_y$$

Hipotesis H_0 akan diterima jika $\overline{X} - \overline{Y}$ tidak berbeda jauh dengan 0. Varian masing-masing sampel adalah

$$S_x^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

$$S_y^2 = \frac{\sum_{i=1}^{m}(Y_i - \overline{Y})^2}{m-1}.$$

Dapat ditunjukan bahwa

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

berdistibusi t dengan dejarat bebas $n+m-2$, dengan

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}.$$

Oleh karen itu jika H_0 benar, yakni jika $\mu_x - \mu_y = 0$, maka statistik

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

berdistribusi t dengan dejarat bebas $n+m-2$.

Aturan 8 (σ_x^2 dan σ_y^2 sama dan tidak diketahui)**.** Untuk menguji

$$H_0 : \mu_x = \mu_y$$

melawan hipotesis

$$H_1 : \mu_x \neq \mu_y$$

dengan tingkat signifikansi α digunakan aturan

$$\begin{aligned}
\textit{terima} \quad H_0 \qquad \textit{jika} \qquad |t| &= \frac{|\overline{X} - \overline{Y}|}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \geq t_{\alpha/2, n+m-2} \\
\textit{tolak} \quad H_0 \qquad \textit{jika} \qquad |t| &= \frac{|\overline{X} - \overline{Y}|}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} > t_{\alpha/2, n+m-2}
\end{aligned}$$

Contoh 6. Dua puluh lima pria sebuah institusi dipilih secara random dan diamati tekanan darah sistoliknya. Dari 25 pria tersebut tercatat 11 perokok dan 14 bukan perokok, dan hasil pengamatan tekanan tersebut adalah sebagai berikut:

Perokok	Bukan perokok
124	120
134	130
136	122
125	128
133	129
127	118
135	122
131	116
133	127
125	135
118	120
	122
	120
	115
	123

Apakah ada perbedaan rata-rata tekanan darah sistolik antara perokok dan bukan perokok pada tingkat signifinasi 5 persen? Bagaimana jika digunakan tingkat signifikansi 1 persen?

Penyelesaian. Dalam soal ini hipotesis yang akan diuji adalah

$$H_0 : \mu_x = \mu_y$$

melawan hipotesis

$$H_1 : \mu_x \neq \mu_y.$$

Karena kedua sampel diambil dari sebuah institusi, maka varian populasi dianggap sama. Kita gunakan indeks x menyatakan prokok dan indeks y bukan perokok. Berdasarkan data tersebut diperoleh

$$n_x = 11 \quad \overline{X}_x = 129.18 \quad S_x^2 = 32.76, \quad n_y = 14 \quad \overline{X}_y = 123.36 \quad S_y^2 = 32.86,$$

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (m_y - 1)S_y^2}{n_x + m_y - 2} = \frac{10 \cdot 32.76 + 13 \cdot 32.86}{23} = 32.82$$

$$|t|\frac{|\overline{X}-\overline{Y}|}{\sqrt{S_p^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}=\frac{|129-123|}{\sqrt{32.82\left(\frac{1}{11}+\frac{1}{14}\right)}}=2.52$$

Karena $t_{0.025,23} = 2.069$ (yang berarti $t_{0.025,23} < |t|$), maka H_0 ditolak, yang berarti bahwa rata-rata tekanan sistolik antara perokok dan bukan perokok berbeda pada tingkat signifikansi 5 persen.

Karena $t_{0.005,23} = 2.807$ (yang berarti $t_{0.005,23} > |t|$), maka H_0 diterima, dengan kata lain rata-rata tekanan sistolik antara perokok dengan bukan perokok tidak berbeda pada tingkat signifikansi 1 persen.

2.3 Varian populasi tidak diketahui dan tidak sama

Dimisalkan varian populasi σ_x^2 dan σ_y^2 tidak diketahui dan tidak sama. Karena S_x^2 adalah estimator untuk σ_x^2 dan S_y^2 estimator untuk σ_y^2 maka uji hipotesis

$$H_0 : \mu_x = \mu_y \quad \text{melawan} \quad H_1 : \mu_x \neq \mu_y$$

bisa didasarkan pada statistik

$$t=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}$$

Jika n dan m cukup besar, maka t akan mendekati distribusi normal standar.

Aturan 9 (σ_x^2 dan σ_y^2 tidak sama dan tidak diketahui)**.** Untuk n dan m besar

$$\textit{terima} \quad H_0 \quad \textit{jika} \quad |t|=\left|\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}\right|\leq z_{\alpha/2}$$

$$\textit{tolak} \quad H_0 \quad \textit{jika} \quad |t|=\left|\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}\right|> z_{\alpha/2}$$

3 Uji t berpasangan

Misalkan peneliti ingin mengetahui apakah ada perbedaan hasil pengukuran berat antara neraca O-haus dengan neraca pegas. Untuk tujuan tersebut diambil sampel n objek kemudian ditimbang beratnya dengan kedua alat. Oleh karena itu didapatkan n pasang pengamatan yang dituliskan

$$(X_i, Y_i), \quad i = 1, 2, \cdots, n$$

dengan X_i dan Y_i berturut-turut menyatakan hasil pengukuran dengan neraca O-haus dan neraca pegas objek ke i.

Suatu cara untuk menguji hipotesis adanya perbedaan hasil pengukuran berat tersebut adalah dengan menghitung selisih kedua hasil pengukuran. Misalkan W_i menyatakan selisih antara pengukuran dengan neraca o-haus dan pengukuran dengan neraca pegas. Jika tidak ada perbedaan hasil pengukuran, maka W_i akan memiliki mean 0. Oleh karena itu kita dapat menuliskan uji hipotesisnya sebagai berikut.

Aturan 10. Uji t berpasangan (t-paired test). Untuk menguji hipotesis

$$\begin{aligned} &H_0 : \mu_w = 0 \quad \text{melawan} \quad H_1 : \mu_w \neq 0 \\ &\text{terima } H_0 \text{ jika } -t_{\alpha/2,\,n-1} < \sqrt{n} \cdot \tfrac{\overline{W}}{S_w} < t_{\alpha/2,\,n-1} \\ &\text{tolah } H_0 \text{ untuk yang lainnya} \end{aligned}$$

Contoh 7. Ingin diketahui apakah ada perbedaan hasil pengukuran berat benda antara necara o-haus dan neraca pegas. Untuk itu diambil 8 sampel benda dan timbang dengan kedua neraca. Misalkan hasil pengukurannya (dalam newton) adalah sebagai berikut.

No. Objek	1	2	3	4	5	6	7	8
Neraca Ohaus	3.2	4.5	5.3	9.6	4.6	8.0	7.3	2.2
Necara pegas	3.1	4.6	5.1	9.5	4.7	8.1	7.1	2.0

Penyelesaian. Misalkan w_i menyatakan selisih hasil pengukuran neraca ohaus dan neraca pegas. Beradasarkan data di atas diperoleh

No. Objek	1	2	3	4	5	6	7	8
w_i	0.1	-0.1	0.2	0.1	-0.1	-0.1	0.2	0.2

Dapat dihitung bahwa mean selisih kedua hasil pengukuran adalah $\bar{w} = 0.0625$ dan deviasi standarnya adalah $s_w = 0.1408$. Selanjutnya, karena $n = 8$, maka

$$\sqrt{n}\frac{\overline{w}}{s_w} = \sqrt{8}\frac{0.0625}{0.1408} = 1.2555$$

dan $t_{\alpha/2,\,n-1} = t_{0.025,\,7} = 2.3646$. Karena

$$-t_{\alpha/2,\,n-1} < \sqrt{n}\frac{\overline{W}}{S_w} < t_{\alpha/2,\,n-1}$$

maka H_0 diterima, yang berarti bahwa tidak ada perbedaan hasil pengukuran kedua jenis alat pada tingkat signifikansi 5 persen.

4 Uji hipotesi tentang varian populasi normal

Dalam aplikasi peneliti sering ingin mengetahui apakah informasi yang tertera pada kemasan suatu produk cukup dapat dipercaya. Sebagai contoh, apakah suatu jenis pupuk memiliki kadar seperti tertera pada labelnya, dan apakah daya tumbuh benih sesuai dengan yang tertera di labelnya. Dalam hal demikian, peneliti dapat menguji apakah varian kadar tersebut tidak berbeda dengan standar varian yang diijinkan.

Misalkan $X_1, X_2, \cdots, X_n$ adalah sampel random dari populasi normal dengan mean μ dan varian σ^2 yang keduanya tidak diketahui. Akan diuji hipotesis

$$H_0 : \sigma^2 = \sigma_0^2$$

melawan hipotesis alternatif

$$H_1 : \sigma^2 \neq \sigma_0^2$$

dengan σ_0^2 suatu nilai tertentu.

Perlu diingat kembali bahwa $(n-1)S^2/\sigma^2$ berdistribusi chi-square dengan derajat bebas $n-1$.

Aturan 11. Untuk menguji hipotesis

$$H_0 : \sigma^2 = \sigma_0^2$$

melawan hipotesis alternatif

$$H_0 : \sigma^2 \neq \sigma_0^2$$

digunakan aturan

terima H_0 jika $\quad \chi^2_{1-\alpha/2,n-1} \leq \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha/2,n-1}$
tolak H_0 untuk lainnya

Contoh 8. Ingin diketahui apakah varian kadar paracetamol suatu jenis obat tidak melebihi batas yang diijinkan pihak berwenang. Misalkan kadar yang ijinkan adalah memiliki varian 40 mg. Suatu sampel 12 obat diukur kadar paracetamolnya dan hasilnya adalah sebagai berikut (dalam mg).

512 532 502 510 508 502 505 511 510 507 509 512

Apakah varian paracematmol dapat diterima pada tingkat signifikansi 1 persen?

Penyelesaian. Diketahui $\sigma_0^2 = 40$, $n = 12$ dan $\alpha = 0.01$. Berdasarkan data dapat dihitung $s^2 = 60$, $\alpha/2 = 0.005$ sehingga $1 - \alpha/2 = 0.995$. Berdasarkan tabel chi-square diperoleh $\chi^2_{0.995,11} = 2.603$ dan $\chi^2_{0.005,11} = 26.757$.

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{(12-1)(60)}{40} = 16.5$$

Karena $\chi^2_{0.95,11} \leq \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{0.05,11}$, maka H_0 diterima. Ini berarti pada tingkat signifikansi 1 persen, varian paracematol jenis obat tersebut tidak melebihi batas yang diijinkan pihak berwenang.

5 Uji hipotesis kesamaan varian dua populasi normal

Jika peneliti memiliki dua populasi, maka mungkin ia tertarik untuk mengetahui apakah variabilitas kedua populasi sama atau berbeda. Sebagai contoh, apakah variabilitas hasil panen padi lokal sama dengan variabilitas hasil panen padi unggul nasional.

Diketahui $X_1, X_2, \cdots, X_n$ dan $Y_1, Y_2, \cdots, Y_m$ dua sampel independen dari dua populasi normal yang masing-masing memiliki parameter μ_x, σ_x^2 dan μ_y, σ_y^2. Akan diuji

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{melawan} \quad H_1 : \sigma_x^2 \neq \sigma_y^2.$$

Varian masing-masing sampel adalah

$$S_x^2 = \frac{1}{n-1} \sum_i^n (X_i - \bar{X})^2$$

$$S_y^2 = \frac{1}{m-1} \sum_i^m (Y_i - \bar{Y})^2.$$

Ingat kembali bahwa $(n-1)S_x^2/\sigma_x^2$ dan $(m-1)S_y^2/\sigma_y^2$ adalah variabel random indepnden dengan derajat bebas masing-masing $n-1$ dan $m-1$. Oleh karena itu

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$$

berdistribusi F dengan derajat bebas pembilang $n-1$ dan derajat bebas penyebut $m-1$.

Aturan 12. Untuk menguji hipotesis

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{melawan} \quad H_1 : \sigma_x^2 \neq \sigma_y^2$$

pada tingkat signifikansi α digunakan aturan

$$\begin{array}{lll} \text{terima} & H_0 & \text{jika} \quad F_{1-\alpha/2,\, n-1, m-1} < S_x^2/S_y^2 < F_{\alpha/2, n-1, m-1} \\ \text{tolak} & H_0 & \text{untuk yang lain} \end{array}$$

Contoh 9. Ingin diketahui apakah ada perbedaan variabilitas hasil panen padi lokal dengan padi unggul nasional pada tingkat signifikansi 5 persen. Suatu sampel padi lokal berukuran 10 dan sampel padi unggul nasional berukuan 13 diambil. Hasil pengamatan (dalam ton per hektar) adalah sebagai berikut.

Padi lokal	1.5	2.3	2.5	1.9	3.0	2.5	1.7	1.8	2.0	2.3
Padi unggul nasional	3.9	4.0	4.2	4.1	4.0	3.8	3.9	4.3	4.5	4.4
	3.8	3.9	4.0							

Penyelesaian. Berdasarkan data di atas diperoleh

$$S_x^2 = 0.205 \quad S_y^2 = 0.0509$$

$$S_x^2/S_y^2 = 0.205/0.0509 = 4.0277$$

Dapat dicari di dalam Tabel f atau dengan aplikasi Excell bahwa $F_{0.975,9,12} = 0.26$ dan $F_{0.025,9,12} = 4.44$, yang berarti H_0 ditolak. Dengan demikian disimpulkan terdapat perbedaan variabilitas produktivitas kedua jenis padi.

6 Uji Goodness of Fit

Misalkan satu mata uang logam seimbang dilontarkan 1000 kali. Frekuensi harapan terjadinya sisi angka tentu $1000 \cdot \frac{1}{2} = 500$ kali. Namun demikian frekuensi terjadinya sisi angka hasil observasi bisa berbeda dengan frekuensi harapan tersebut. Untuk menguji apakah hasil percobaan ini sesuai dengan hasil teoritis, bisa digunakan satu uji statistik.

Goodness of fit adalah suatu uji statistik untuk menentukan apakah suatu populasi memiliki distribusi tertentu.

Misalkan peneliti memiliki k persitiwa $E_1, E_2, \cdots, E_k$ yang masing-masing dapat terjadi dengan peluang $p_1, p_2, \cdots, p_k$. Jika diambil sampel random berukuran n dari populasi ini, nilai observasi untuk peristiwa $E_1, E_2, \cdots, E_k$ dapat dinyatakan sebagai variabel random $X_1, X_2, \cdots, X_k$. Frekuensi harapan peristiwa-peristiwa tersebut masing-masing adalah $np_1, np_2, \cdots, np_k$. Dalam bentuk tabel dapat dinyatakan sebagai berikut.

Peristiwa	E_1	E_2	$\cdots$	E_k
Frekuensi observasi	x_1	x_2	$\cdots$	x_k
Frekuensi harapan	np_1	np_2	$\cdots$	np_k

Aturan 13. Untuk menguji apakah nilai observasi menyimpang terhadap nilai harapan digunakan aturan sebagai berikut.

$$\chi^2 = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} + \cdots + \frac{(X_k - np_k)^2}{np_k} \tag{4}$$

$$\begin{array}{llll} \text{terima} & H_0 & \text{jika} & \chi^2 < \chi^2_{\alpha,k-1} \\ \text{tolak} & H_0 & \text{jika} & \chi^2 \geq \chi^2_{\alpha,k-1} \end{array}$$

Contoh 10. Satu dadu dilontarkan 120 kali. Hasil observasi dan nilai harapan setiap sisi dinyatakan dalam tabl berikut.

Sisi	1	2	3	4	5	6
Nilai observasi	22	23	18	19	21	17
Nilai harapan	20	20	20	20	20	20

Jika dadu tersebut seimbang, tentu peneliti berharap setiap sisi memiliki peluang 1/6. Dengan demikian dalam 120 lontaran, setiap sisi diharapkan terjadi $\frac{1}{6}120 = 20$ kali. Namun berdasarkan hasil observasi, ternyata tidak semua sisi terjadi 20 kali. Apakah ini berarti dadu tersebut tidak seimbang? Untuk mengetahuinya kita gunakan uji chi-square.

$$\begin{aligned} \chi^2 &= \tfrac{(22-20)^2}{20} + \tfrac{(23-20)^2}{20} + \tfrac{(18-20)^2}{20} + \tfrac{(19-20)^2}{20} + \tfrac{(21-20)^2}{20} + \tfrac{(17-20)^2}{20} \\ &= 1.4. \end{aligned}$$

Berdasarkan tabel, $\chi^2_{0.05,5} = 11.070$. Karena nilai yang dihitung $\chi^2 < \chi^2_{0.05,5}$ maka hipotesis nol diterima, dengan kata lain tidak cukup bukti untuk mengatakan bahwa dadu tidak seimbang.

7 Uji Independen

Dalam fenomena sehari-hari, seseorang sering ingin mengetahui ada tidaknya hubungan antara dua variabel. Sebagai contoh, apakah tingkat pendapatan masyarakat independen terhadap keputusan untuk menolak atau menerima kenaikan BBM. Misalkan ada 3 tingkat pendapatan: miskin, sedang dan kaya. Dalam hal ini populasi memiliki dua karakteristik, yaitu yang setuju dan yang tidak setuju dengan kenaikan BBM. Kedua karakteristik masyarakat terbagi dalam 3 golongan, yaitu miskin, sedang dan kaya.

Misalkan kita memiliki suatu populasi yang dapat diklasifikasikan menjadi dua karakteristik, namakan X dan Y. Misalkan ada r nilai yang mungkin untuk kelas X dan ada s nilai yang mungkin untuk klas Y. Peluang kelas X terjadi nilai i dan kelas Y terjadi nilai j ditulis P_{ij}. Jadi

$$P_{ij} = P(X = i, Y = j)$$

Setiap anggota populasi diasumsikan independen. Peluang sebarang anggota populasi memiliki sifat X dengan nilai i adalah

$$p_i = \sum_{j=1}^{s} P_{ij}$$

Peluang sebarang anggota populasi memiliki sifat Y dengan nilai j adalah

$$q_j = \sum_{i=1}^{r} P_{ij}$$

Misalkan kita ingin menguji hipotsis bahwa anggota populasi dengan sifat X dan Y independen, yaitu

$$H_0 : p_{ij} = p_i q_j$$

melawan hipotesis

$$H_1 : p_{ij} \neq p_i q_j$$

Karena nilai p_i dan q_i tidak dinyatakan secara spesifik di dalam hipotesis, maka kita harus menduga nilai tersebut. Banyaknya anggota populasi dengan karakteristik X dan berpendapat i adalah

$$N_i = \sum_{j=1}^{r} N_{ij},$$

dan estimator untuk p_i adalah

$$\hat{p}_i = \frac{N_i}{n}, \quad i = 1, 2, \cdots, r.$$

Banyaknya anggota populasi dengan karakeristik Y dan memilih j adalah

$$M_j = \sum_{i=1}^{r} N_{ij},$$

dan estimator untuk q_i adalah

$$\hat{q}_i = \frac{M_j}{n}, \quad j = 1, 2, \cdots, s.$$

Jika H_0 benar, maka

$$E(N_{ij}) = np_iq_j$$

dan

$$T = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{(N_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

berdistribusi chi-square dengan derajat bebas $(r-1)(s-1)$. Oleh karena itu diperoleh aturan berikut.

Aturan 14. Untuk menguji hipotesis

$$H_0 : p_{ij} = p_iq_j$$

melawan hipotesis

$$H_1 : p_{ij} \neq p_iq_j$$

digunakan aturan

$$T = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{(N_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

$$\begin{array}{llll} \text{terima} & H_0 & \text{jika} & T \geq \chi^2_{\alpha,(r-1)(s-1)} \\ \text{tolak} & H_0 & \text{untuk lainnya} & \end{array}$$

Contoh 11. Untuk mengetahui apakah keputusan menolak atau menerima kenaikan BBM independen terhadap tingkat pendapatan masyarakat, diambil sampel random berukuran 200 orang. Hasil observasi diberikan dalam tabel berikut (dinamakan tabel contingency).

Tingkat pendapatan	Miskin	Sedang	Kaya	Jumlah
Setuju	30	32	32	94
Tidak Setuju	44	35	27	106
Jumlah	74	67	59	200

Dari tabel tersebut diperoleh

$$\hat{p}_1 = \tfrac{94}{200} = 0.47$$
$$\hat{p}_2 = \tfrac{106}{200} = 0.53$$
$$\hat{q}_1 = \tfrac{74}{200} = 0.37$$
$$\hat{q}_2 = \tfrac{67}{200} = 0.335$$
$$\hat{q}_3 = \tfrac{59}{200} = 0.295$$
$$n\hat{p}_1\hat{q}_1 = (200)(0.47)(0.37) = 34.78$$
$$n\hat{p}_1\hat{q}_2 = (200)(0.47)(0.335) = 31.49$$
$$n\hat{p}_1\hat{q}_3 = (200)(0.47)(0.295) = 27.73$$
$$n\hat{p}_2\hat{q}_1 = (200)(0.53)(0.37) = 39.22$$
$$n\hat{p}_2\hat{q}_2 = (200)(0.53)(0.335) = 35.51$$
$$n\hat{p}_2\hat{q}_3 = (200)(0.53)(0.295) = 31.27$$

$$\begin{aligned} T &= \tfrac{(30-34.78)^2}{34.78} + \tfrac{(32-31.49)^2}{31.49} + \tfrac{(32-27.73)^2}{27.73} \\ &\quad + \tfrac{(44-39.22)^2}{39.22} + \tfrac{(35-35.51)^2}{35.51} + \tfrac{(27-31.27)^2}{31.27} \\ &= 2.5415. \end{aligned}$$

Karena $(r-1)(s-1) = 2$ dan $\chi^2_{0.05,2} = 5.911$, maka H_0 terima, yakni keputusan untuk menerima atau menolak kenaikan BBM adalah independen.