

Uji Chi-Square Data Kategorik

PEMBELAJARAN DARING KOLABORASI

**UNIVERSITAS MUHAMMADIYAH PALANGKA RAYA-
STMIK PALANGKA RAYA**



Uji Chi Square

1

Uji chi-square merupakan uji statistik non parametrik, dimana tidak mensyaratkan datanya harus berdistribusi normal.

2

Data yang digunakan merupakan data kategorik (nominal atau ordinal).

3

Uji chi square dilakukan untuk sampel tunggal dan sampel independent.

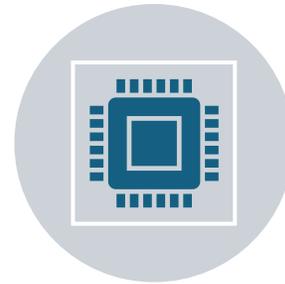
1. Uji Chi-Square (χ^2) Sampel Tunggal (*Goodness of Fit Test*)

- **Tujuan:** menentukan apakah distribusi frekuensi yang diamati dari satu variable kategorik sesuai dengan distribusi frekuensi yang diharapkan (berdasarkan teori atau hipotesis tertentu).
- **Contoh dibidang informatika:** menguji apakah distribusi system operasi yang digunakan oleh Mahasiswa informatika mengikuti distribusi system operasi yang digunakan secara global.
- **Contoh dibidang farmasi:** menguji apakah kandungan zat aktif dalam tablet yang diproduksi dalam satu batch seragam dan sesuai dengan spesifikasi yang ditetapkan.

Asumsi Uji Chi Square Sampel Tunggal



data berupa data kategorik
(nominal atau ordinal)



data dalam sel harus
berupa frekuensi atau
hitungan (bukan
persentase)



Kategori bersifat saling
eksklusif (satu observasi
hanya masuk ke dalam
satu kategori)



Frekuensi yang diharapkan
untuk setiap kategori
minimal 5 (aturan praktis).



Hipotesis

- **Hipotesis nol (H_0):** distribusi yang diamati **sesuai** dengan distribusi frekuensi yang ditetapkan
 - **Hipotesis alternatif (H_1):** distribusi yang diamati **tidak sesuai** dengan distribusi frekuensi yang ditetapkan
-

Langkah-Langkah Uji Chi-Square untuk Sampel Tunggal

- 1) **Tentukan hipotesis nol dan alternatif**
- 2) **Tentukan frekuensi yang diamati (O).** Ini adalah data yang dikumpulkan.
- 3) **Tentukan frekuensi yang diharapkan (E).** Ini dihitung berdasarkan teori atau hipotesis yang ingin diuji. Biasanya, jika mengharapkan proporsi yang sama di semua kategori, maka $E = (\text{Total Frekuensi}) / (\text{Jumlah Kategori})$. Jika ada proporsi spesifik yang diharapkan, kalikan proporsi tersebut dengan total frekuensi.
- 4) **Hitung nilai chi-square** dengan rumus:

$$\chi^2 = \sum [(O - E)^2 / E]$$

dimana O adalah frekuensi yang diamati dan E adalah frekuensi yang diharapkan.

Langkah-Langkah Uji Chi-Square Sampel Tunggal

5) **Tentukan derajat kebebasan (dk)**, dengan rumus:

$$dk = \text{jumlah kategori} - 1$$

6) **Tentukan nilai p (*p-value*)**. Gunakan table distribusi Chi-Square atau kalkulator statistik dengan dk yang telah dihitung dan nilai χ^2 yang diperoleh.

7) **Tafsirkan hasil.**

Jika nilai $p \leq$ tingkat signifikansi (α), maka H_0 ditolak. Berarti ada perbedaan signifikan antara frekuensi yang diamati dan yang diharapkan.

Jika nilai $p >$ tingkat signifikansi (α), maka H_0 diterima. Berarti tidak ada perbedaan signifikan antara frekuensi yang diamati dan yang diharapkan.

Contoh 1.

Sebuah rumah sakit ingin mengevaluasi efektivitas sistem resep elektronik (e-prescription) baru yang diterapkan. Salah satu indikator efektivitas adalah kecepatan dokter dalam menulis resep. Rumah sakit tersebut memiliki data historis yang menunjukkan bahwa, dengan sistem manual sebelumnya, distribusi waktu penulisan resep (dalam menit) adalah sebagai berikut:

- **< 2 menit:** 20%
- **2-5 menit:** 50%
- **> 5 menit:** 30%

Setelah implementasi sistem e-prescription, diambil sampel acak sebanyak 150 resep yang ditulis menggunakan sistem baru. Hasilnya seperti pada Tabel berikut.

Dengan tingkat signifikansi 0,05, ujilah apakah distribusi waktu penulisan resep dengan sistem e-prescription berbeda secara signifikan dibandingkan dengan sistem manual sebelumnya.

Waktu penulisan resep (menit)	Jumlah resep (frekuensi yg diamati)
< 2 menit	45
2 – 5 menit	70
> 5 menit	35



Penyelesaian.

1) Hipotesis:

H0 : Distribusi waktu penulisan resep dengan system e-prescription sama dengan distribusi penulisan resep dengan system manual

H1 : Distribusi waktu penulisan resep dengan system e-prescription berbeda dengan distribusi penulisan resep dengan system manual.

2) Frekuensi yang diamati (O) : sudah diberikan dalam table soal.

Penyelesaian.

3) Frekuensi yang diharapkan (E) :

Waktu penulisan resep (menit)	Persentase (Sistem Manual)	Frekuensi yang diharapkan (E)
< 2 menit	20% = 0,20	$0,20 * 150 = 30$
2 – 5 menit	50% = 0,50	$0,50 * 150 = 75$
> 5 menit	30% = 0,30	$0,30 * 150 = 45$

Penyelesaian.

4) Hitung statistik chi-square χ^2 :

Waktu penulisan resep (menit)	O	E	$O - E$	$(O - E)^2$	$(O - E)^2 / E$
< 2 menit	45	30	15	225	7,5
2 – 5 menit	70	75	-5	25	0,33
> 5 menit	35	45	-10	100	0,22
Total	150	150			10,05

$$\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right] = 7,5 + 0,33 + 2,22 = 10,05$$

Penyelesaian.

5) Hitung derajat kebebasan (dk) :

$$dk = \text{jumlah kategori} - 1 = 3 - 1 = 2$$

6) Hitung Nilai p, menggunakan table distribusi chi-square atau kalkulator statistik dengan $dk = 2$ dan χ^2 , diperoleh nilai p sekitar 0,0066.

7) Tafsirkan hasil:

Karena nilai $p = 0,0066 < \text{Tingkat signifikansi } (\alpha) = 0,05$, maka H_0 ditolak, berarti menerima H_1 . Dengan demikian dapat disimpulkan bahwa dengan Tingkat signifikansi 0,05 terdapat cukup bukti untuk menyatakan bahwa distribusi waktu penulisan resep dengan system e-prescription berbeda secara signifikan dibandingkan dengan system manual sebelumnya. Terlihat bahwa dengan system e-prescription, proporsi resep yang ditulis dalam waktu >5 menit lebih rendah dibandingkan dengan system manual. Hal ini menunjukkan bahwa system e-prescription berpotensi meningkatkan efisiensi dalam penulisan resep.

2. Uji Chi-Square Sampel Independent

Tujuan: menentukan apakah ada hubungan yang signifikan antara dua variable kategorik. Uji ini membandingkan frekuensi yang diamati dari data dengan frekuensi yang diharapkan jika kedua variable tersebut tidak berhubungan.

Contoh:

- menguji apakah ada hubungan antara tingkat pendidikan apoteker dengan tingkat kepuasan pasien terhadap pelayanan informasi obat
 - menguji apakah ada hubungan antara penggunaan aplikasi kesehatan dengan tingkat keberhasilan terapi.
-

Asumsi-asumsi

1

Kedua variable harus kategorik (nominal atau ordinal)

2

Data dalam sel harus berupa frekuensi atau hitungan data (bukan persentase)

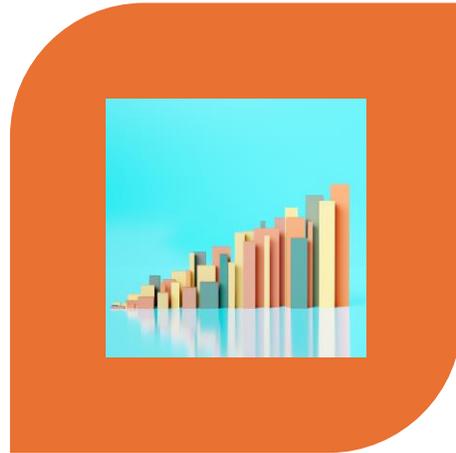
3

Tingkat variable bersifat saling eksklusif

4

Tidak ada sel dalam table kontingensi yang memiliki hitungan yang diharapkan kurang dari 5

Hipotesis



Hipotesis Nol (H0) :

Tidak Ada Hubungan Antara
Kedua Variable Kategorik



Hipotesis Alternatif (H1):

Ada Hubungan Antara
Kedua Variable Kategorik

Langkah-Langkah Sampel Independent dengan Uji Chi-Square

- 1) Tentukan formulasi hipotesis (H0 dan H1)
- 2) Buat table kontingensi, yaitu table yang menunjukkan frekuensi yang diamati untuk setiap kombinasi kategori dari kedua variable.
- 3) Hitung frekuensi yang diharapkan, yaitu frekuensi yang akan diharapkan untuk dilihat dalam setiap sel table jika hipotesis nol benar. Frekuensi yang diharapkan (E) ini dapat dihitung dengan rumus berikut:

$$E = (\text{Baris Total} * \text{Kolom Total}) / \text{Total Keseluruhan}$$

Langkah-Langkah Uji Chi-Square

4) Hitung statistic Chi-Square, gunakan rumus:

$$\chi^2 = \sum [(O - E)^2 / E]$$

dimana O adalah frekuensi yang diamati dan E adalah frekuensi yang diharapkan.

5) Tentukan derajat kebebasan dengan rumus:

$$df = (\text{jumlah baris} - 1) * (\text{jumlah kolom} - 1)$$

- 6) Cari nilai p (p-value). Nilai p adalah probabilitas untuk mendapatkan statistic Chi-Square seekstrem atau lebih ekstrem daripada yang diamati jika hipotesis nol benar. Nilai p diperoleh dari table distribusi Chi Square atau kalkulator statistic.
- 7) Tafsirkan hasilnya. Jika nilai p kurang dari Tingkat signifikansi ($p - value < \alpha$) maka H0 ditolak. Ini berarti terdapat bukti yang mendukung bahwa ada hubungan antara kedua variable, dan sebaliknya.

Contoh 2.

Sebuah Perusahaan melakukan survei terhadap 200 orang yang dipilih secara acak, dengan pertanyaan apakah mereka lebih suka menonton komedi atau drama di televisi. Hasil survei diringkas dalam table kontingensi berikut:

Lakukan Uji Chi Square untuk menentukan apakah ada hubungan antara jenis kelamin dan preferensi menonton menggunakan taraf signifikansi 5%!

	Komedi	Drama	Total
Laki-laki	50	30	80
Perempuan	80	40	120
Total	130	70	200

Penyelesaian.

1) Tentukan formulasi hipotesis (H0 dan H1)

H0 : Tidak ada hubungan antara jenis kelamin dan preferensi menonton

H1 : Ada hubungan antara jenis kelamin dan preferensi menonton

2) Tabel Kontingensi

	Komedi	Drama	Total
Laki-laki	50	30	80
Perempuan	80	40	120
Total	130	70	200

3) Frekuensi yang diharapkan:

$$(E = (\text{Baris Total} * \text{Kolom Total}) / \text{Total Keseluruhan})$$

$$E (\text{laki-laki, Komedi}) = (80 * 130) / 200 = 52$$

$$E (\text{laki-laki, Drama}) = (80 * 70) / 200 = 28$$

$$E (\text{Perempuan, Komedi}) = (120 * 130) / 200 = 78$$

$$E (\text{Perempuan, Drama}) = (120 * 70) / 200 = 42$$

Table kontingensi dengan frekuensi yang diharapkan:

	Komedi		Drama		Total
	O	E	O	E	
L	50	52	30	28	80
P	80	78	40	42	120
Total	130	-	70	-	200

4) Menghitung statistic Chi-Square:

$$\chi^2 = \sum[(O - E)^2 / E],$$

$$\begin{aligned} \chi^2 &= \frac{(50 - 52)^2}{52} + \frac{(30 - 28)^2}{28} + \frac{(80 - 78)^2}{78} + \frac{(40 - 42)^2}{42} \\ &= 0,077 + 0,143 + 0,051 + 0,095 = 0,366, \end{aligned}$$

5) Menghitung derajat kebebasan:

$$df = (\text{jumlah baris} - 1) * (\text{jumlah kolom} - 1)$$

$$Df = (2 - 1) * (2 - 1) = 1$$

6) Menghitung nilai p:

Nilai p yang sesuai untuk $\chi^2 = 0,366$ dengan $df=1$, diperoleh berdasarkan table distribusi chi - square yaitu 0,545.

7) Kesimpulan

Karena nilai $p = 0,545 >$ Tingkat signifikansi $= 0,05$, maka H_0 diterima. Ini berarti bahwa tidak ada cukup bukti untuk menyimpulkan bahwa ada hubungan antara jenis kelamin dan preferensi menonton di televisi.

Catatan:

Perhitungan manual uji chi square bisa sangat rumit dan memakan waktu terutama dengan table kontingensi yang besar. Oleh karena itu, pengujian dapat dilakukan dengan bantuan software statistic.

Soal Latihan.

- 1) Sebuah perusahaan farmasi mengklaim bahwa obat baru mereka memiliki efek samping yang sama pada semua kelompok usia. Untuk menguji klaim tersebut, dilakukan survei terhadap 200 pasien yang mengonsumsi obat, dan diperoleh data sebagai berikut:

Kelompok usia	Efek samping	Tidak ada efek samping	Total
< 30 tahun	20	80	100
30 – 59 tahun	30	70	100
\geq 60 tahun	40	60	100
Total	90	210	300

Ujilah hipotesis untuk mengetahui apakah proporsi efek samping sama pada semua kelompok usia, gunakan Tingkat signifikansi 5%.

Soal Latihan.

2) Sebuah perusahaan ingin mengetahui apakah ada hubungan antara jenis kelamin pengguna dan preferensi system operasi (Windows, macOS, Linux). Data survey diperoleh sebagai berikut:

Jenis kelamin	Windows	macOS	Linux	Total
Laki-laki	50	30	20	100
Perempuan	40	40	20	100
Total	90	70	40	200

Ujilah hipotesis untuk mengetahui apakah ada hubungan antara jenis kelamin dan preferensi system operasi, gunakan tingkat signifikansi 5%.