

Chi-Square Test (χ^2)

1. Saat melakukan pelemparan koin, kita mengharapkan 30 kali kepala dan 30 kali ekor dalam sebuah eksperimen sebanyak 60 kali. Realitanya, hasilnya jarang tepat sempurna → perlu uji Chi-Square
2. Uji Chi-Square (χ^2) adalah uji statistik non-parametrik yang digunakan untuk:
 - a. Menguji perbedaan secara signifikan antara frekuensi yang diamati dengan frekuensi yang diharapkan
 - b. Menguji kebebasan antar faktor dari data dalam tabel kontingensi
 - c. Menguji kedekatan data sampel dengan suatu fungsi distribusi, seperti binomial, poisson, atau normal
3. Uji Chi-Square (χ^2) digunakan ketika:
 - Ketika ada sedikit atau tidak ada asumsi tentang parameter atau distribusi dari populasi terkait
 - Digunakan pada data kategorikal atau data kualitatif dengan menggunakan tabel kontingensi
 - Digunakan untuk mengevaluasi sampel dan proporsi yang tidak berpasangan atau bersifat independent satu sama lain.
 - Jumlah sampel (N) harus cukup besar
4. Distribusi Chi-Square:
 - Uji chi-square mengukur sejauh mana frekuensi yang diamati berbeda dari frekuensi yang diharapkan, dan hasilnya dibandingkan dengan distribusi chi-square untuk mengevaluasi signifikansi.
 - Total area yang berada di bawah area kurva adalah 1
 - Kurva selalu menceng kanan
 - Semakin besar jumlah derajat kebebasan, kurva distribusi chi-square semakin mendekati distribusi normal
 - Rata-rata dari distribusi chi-square adalah derajat kebebasan, sedangkan variansnya adalah 2 kali dari derajat kebebasan.
 - Ketika nilai derajat kebebasan adalah 3 atau lebih, puncak dari distribusi chi-square sama dengan df-2. Nilai ini merupakan cara untuk menentukan nilai modus pada distribusi chi-square.
5. Ada dua macam uji Chi-Square (χ^2):
 - a. Uji kecocokan model → mengetahui apakah suatu data terdistribusi sesuai dengan distribusi atau pola tertentu (goodness of fit test)
 - b. Uji independensi → mengetahui apakah dua variabel kategorik saling bebas atau tidak saling bebas

6. Syarat uji independensi

- Tidak ada cell dengan nilai frekuensi observasi (o_i) sebesar 0 (nol)
- Apabila bentuk tabel kontingensi 2×2 , maka tidak boleh ada 1 cell yang memiliki frekuensi harapan/expected (e_i) kurang dari 5 → Uji Fisher Exact Probability
- Apabila bentuk tabel lebih dari 2×2 , maka jumlah cell dengan frekuensi harapan yang kurang dari 5 tidak boleh lebih dari 20% dari keseluruhan data yang dimiliki.

Uji kecocokan model atau goodness of fit

- Ilustrasi: pelemparan sebuah dadu memiliki hipotesis adalah dadu seimbang, yang setara dengan menguji hipotesis bahwa distribusi outcome (setiap mata dadu) adalah berdistribusi seragam diskrit (discrete uniform distribution)

$$f(x) = \frac{1}{6}, x = 1, 2, \dots, 6$$

- Misalkan dilakukan pelemparan sebuah dadu sebanyak 120 kali dan setiap hasil dicatat. Secara teoritis, jika dadu seimbang, setiap mata dadu akan muncul sebanyak 20 kali. Hasil pelemparan ditunjukkan pada tabel berikut:

Frekuensi “teramati” dan “yang diharapkan” dari 120 kali pelemparan sebuah dadu

Face	1	2	3	4	5	6
Observed	20	22	17	18	19	24
Expected	20	20	20	20	20	20

- Statistik uji untuk uji kecocokan model:

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

χ^2 adalah nilai variabel acak yang diperkirakan distribusi samplingnya sangat dekat dengan distribusi khi-kuadrat dengan derajat bebas $\nu = k - 1$

o_i adalah frekuensi yang diamati (observed)

e_i adalah frekuensi yang diharapkan

- Langkah pengujian hipotesis:

Tujuan: mengetahui apakah dadu seimbang

- Hipotesis

H_0 : data hasil pelemparan dadu berdistribusi uniform

H_1 : data hasil pelemparan dadu tidak berdistribusi uniform

- Taraf signifikansi

$$\alpha = 5\% = 0.05$$

c. Statistik uji

$$\begin{aligned}\chi^2 &= \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \\ &= \frac{(20 - 20)^2}{20} + \frac{(22 - 20)^2}{20} + \frac{(17 - 20)^2}{20} + \frac{(18 - 20)^2}{20} + \frac{(19 - 20)^2}{20} \\ &\quad + \frac{(24 - 20)^2}{20} \\ &= 1.7\end{aligned}$$

d. Keputusan:

$$v = k - 1 = 6 - 1 = 5$$

$$\chi_{hitung}^2 = 1.7$$

$$\chi_{0.05;5}^2 = 11.070$$

Karena $\chi_{hitung}^2 < \chi_{0.05;5}^2$, maka H_0 tidak ditolak (diterima)

e. Kesimpulan:

Dengan tingkat signifikansi 5%, dapat disimpulkan bahwa data hasil pelemparan dadu berdistribusi uniform (dadu seimbang)

Uji independensi (Test for Independence)

1. Prosedur uji chi-square yang dibahas pada bagian uji kecocokan model dapat juga digunakan untuk menguji hipotesis independensi dua variabel kategorik
2. Contoh: ingin diketahui apakah pendapat penduduk di suatu daerah tentang reformasi pajak baru tidak bergantung pada tingkat pendapatan mereka. Sampel acak terdiri dari 1000 pemilih terdaftar dari suatu negara yang diklasifikasikan apakah mereka termasuk dalam kelompok berpenghasilan rendah, menengah, atau tinggi; dan apakah mereka mendukung reformasi pajak atau tidak. Frekuensi yang diamati disajikan pada tabel berikut yang dikenal sebagai tabel kontingensi 2 X 3

	Income Level			Total
Tax Reform	Low	Medium	High	
For	182	213	203	598
Against	154	138	110	402
Total	336	351	313	1000

3. Langkah pengujian hipotesis
Tujuan: ingin mengetahui apakah pendapat mengenai Tax Reform dipengaruhi oleh tingkat penghasilan

a. Hipotesis

H_0 : Pendapat mengenai Tax Reform dan tingkat penghasilan tidak saling mempengaruhi ($o_{ij} = e_{ij}$)

H_1 : Pendapat mengenai Tax Reform dan tingkat penghasilan saling mempengaruhi ($o_{ij} \neq e_{ij}$)

b. Taraf signifikansi

$$\alpha = 5\% = 0.05$$

c. Statistik uji:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

	Income Level			Total
Tax Reform	Low	Medium	High	
For	182	213	203	598
Against	154	138	110	402
Total	336	351	313	1000

Menghitung nilai ekspektasi (e_i):

L = sampel terpilih pada low-income level $\rightarrow P(L) = \frac{336}{1000}$

M = sampel terpilih pada medium-income level $\rightarrow P(M) = \frac{351}{1000}$

H = sampel terpilih pada high-income level $\rightarrow P(F) = \frac{598}{1000}$

F = sampel terpilih setuju pada tax reform $\rightarrow P(A) = \frac{402}{1000}$

A = sampel terpilih tidak setuju (against) pada tax reform $\rightarrow P(H) = \frac{313}{1000}$

$$P(L \cap F) = P(L)P(F) = \frac{336}{1000} \cdot \frac{598}{1000}$$

$$P(L \cap A) = P(L)P(A) = \frac{336}{1000} \cdot \frac{402}{1000}$$

$$P(M \cap F) = P(M)P(F) = \frac{351}{1000} \cdot \frac{598}{1000}$$

$$P(M \cap A) = P(M)P(A) = \frac{351}{1000} \cdot \frac{402}{1000}$$

$$P(H \cap F) = P(H)P(F) = \frac{313}{1000} \cdot \frac{598}{1000}$$

$$P(H \cap A) = P(H)P(A) = \frac{313}{1000} \cdot \frac{402}{1000}$$

Contoh menghitung e_i untuk pemilih dengan low income dan setuju dengan

$$\text{tax reform: } \frac{336}{1000} \cdot \frac{598}{1000} \cdot 1000 = \frac{(336)(598)}{1000} = 200.9$$

Sehingga rumusnya menjadi: $e_{ij} = \frac{(\text{column total}) \times (\text{row total})}{\text{total}}$

	Income Level			Total
Tax Reform	Low	Medium	High	
For	182 (200.9)	213 (209.9)	203 (187.2)	598
Against	154 (135.1)	138 (141.1)	110 (125.8)	402
Total	336	351	313	1000

$$\begin{aligned} \chi^2 &= \sum_{i=1}^r \sum_{j=1}^k \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \\ &= \frac{(182 - 200.9)^2}{200.9} + \frac{(213 - 209.9)^2}{209.9} + \frac{(203 - 187.2)^2}{187.2} + \frac{(154 - 135.1)^2}{135.1} \\ &\quad + \frac{(138 - 141.1)^2}{141.1} + \frac{(110 - 125.8)^2}{125.8} \\ &= 7.85 \end{aligned}$$

d. Keputusan:

$$v = (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$$

$$\chi_{hitung}^2 = 7.85$$

$$\chi_{0.05;2}^2 = 5.991$$

Karena $\chi_{hitung}^2 > \chi_{0.05;2}^2$, maka H_0 ditolak

f. Kesimpulan:

Dengan tingkat signifikansi 5%, dapat disimpulkan bahwa pendapat mengenai tax reform dan tingkat penghasilan saling mempengaruhi (ada dependensi antara tax reform dan tingkat penghasilan)