

Nama : Stefanus Edward Azarya Manoppo
NIM : 00000121955

STRUCTURED ASSIGNMENT 8

CE319 Probability and

Statistic Academic Year

2025/2026

Instruction:

- Perform these tasks using R Programming Language. **Attach your code, result, and give your analysis.**
 - Make a report consist of syntax, result and analysis in 1 PDF file with file name **NIM(WithoutZero)_Name_Task3**
1. Use chi-square goodness of fit and test of independence to solve the following problems.
 - a. Use a chi-square goodness-of-fit test to determine whether the observed frequencies are distributed the same as the expected frequencies. Use $\alpha = 0.05$

Category	f_o	f_e
1	53	68
2	37	42
3	32	33
4	28	22
5	18	10
6	15	8

Syntax

```
1 # --- SOAL A: Goodness of Fit Test ---
2
3 # Memasukkan data
4 observed <- c(53, 37, 32, 28, 18, 15) # f0
5 expected <- c(68, 42, 33, 22, 10, 8) # fe
6
7 # Menghitung probabilitas untuk setiap kategori berdasarkan expected values
8 # Karena chisq.test di R membutuhkan argumen 'p' sebagai probabilitas (proporsi), bukan nilai mentah.
9 probs <- expected / sum(expected)
10
11 # Melakukan Chi-square Goodness-of-Fit Test
12 hasil_a <- chisq.test(x = observed, p = probs)
13
14 # Menampilkan hasil
15 print(hasil_a)
```

Output

```
> # --- SOAL A: Goodness of Fit Test ---
>
> # Memasukkan data
> observed <- c(53, 37, 32, 28, 18, 15) # f0
> expected <- c(68, 42, 33, 22, 10, 8) # fe
>
> # Menghitung probabilitas untuk setiap kategori berdasarkan expected values
> # Karena chisq.test di R membutuhkan argumen 'p' sebagai probabilitas (proporsi), bukan nilai mentah.
> probs <- expected / sum(expected)
>
> # Melakukan Chi-square Goodness-of-Fit Test
> hasil_a <- chisq.test(x = observed, p = probs)
>
> # Menampilkan hasil
> print(hasil_a)

      Chi-squared test for given probabilities

data:  observed
X-squared = 18.096, df = 5, p-value = 0.002829
```

Analisis Langkah demi Langkah berdasarkan Output R

1. Definisi Hipotesis

Sebelum melihat angka, ditetapkan dulu apa yang diuji:

- **H0 (Hipotesis Nol):** Distribusi frekuensi yang diamati (f_o) SAMA atau sesuai dengan frekuensi yang diharapkan (f_e). Artinya, perbedaan angka hanyalah kebetulan.
- **H1 (Hipotesis Alternatif):** Distribusi frekuensi yang diamati BERBEDA secara signifikan dengan frekuensi yang diharapkan.

2. Analisis P-Value (Kunci Utama)

- Nilai P-value: 0.002829
- Tingkat Signifikansi (Alpha): 0.05

Logika Keputusan:

- Jika P-value < 0.05 , maka Tolak H_0 .
- Jika P-value > 0.05 , maka Terima H_0 .

Hasil:

Karena $0.002829 < 0.05$, maka keputusan statistik adalah Tolak H_0 .

3. Analisis X-squared (Nilai Chi-Square)

- Nilai X-squared: 18.096
- Degrees of Freedom (df): 5

Penjelasan:

Angka 18.096 merepresentasikan total besarnya "selisih" atau penyimpangan antara data asli (observasi) dengan data harapan (ekspektasi).

- Semakin BESAR nilai X-squared, semakin besar ketidakcocokan data.
- Semakin KECIL nilai X-squared (mendekati 0), semakin pas/cocok datanya.

Nilai 18.096 tergolong cukup besar, yang mengonfirmasi bahwa selisih antara apa yang diobservasi dan apa yang diharapkan itu nyata.

4. Kesimpulan Akhir

Berdasarkan hasil uji statistik tersebut:

"Terdapat perbedaan yang signifikan antara frekuensi yang diobservasi dengan frekuensi yang diharapkan."

Dalam konteks soal:

Data kategori 1 sampai 6 TIDAK terdistribusi sesuai dengan ekspektasi teoretisnya. Ada kategori tertentu yang nilainya jauh lebih tinggi atau jauh lebih rendah dari yang seharusnya diprediksi oleh model.

- b. A group people are interviewed to determine whether the type of music most listened to by people in their region is independent of the age. The interview result is given below. Determine whether music preference is independent of age. Use $\alpha = 0.05$

		Type of Music			
		Rock	RnB	Pop	Classic
Age	15-25	140	32	5	18
	26-35	134	41	52	8
	36-45	154	27	8	13

Syntax

```
1 # --- SOAL B: Test of Independence ---
2
3 # Memasukkan data ke dalam matriks (Contingency Table)
4 data_musik <- matrix(c(140, 32, 5, 18,
5                        134, 41, 52, 8,
6                        154, 27, 8, 13),
7                      nrow = 3,
8                      byrow = TRUE)
9
10 # Memberikan nama baris dan kolom agar lebih rapi (opsional)
11 rownames(data_musik) <- c("15-25", "26-35", "36-45")
12 colnames(data_musik) <- c("Rock", "RnB", "Pop", "Classic")
13
14 # Melakukan Chi-square Test of Independence
15 hasil_b <- chisq.test(data_musik)
16
17 # Menampilkan hasil
18 print(data_musik) # Cek tabel
19 print(hasil_b)
```

Output

```
> # --- SOAL B: Test of Independence ---
>
> # Memasukkan data ke dalam matriks (Contingency Table)
> data_musik <- matrix(c(140, 32, 5, 18,
+                        134, 41, 52, 8,
+                        154, 27, 8, 13),
+                      nrow = 3,
+                      byrow = TRUE)
>
> # Memberikan nama baris dan kolom agar lebih rapi (opsional)
> rownames(data_musik) <- c("15-25", "26-35", "36-45")
> colnames(data_musik) <- c("Rock", "RnB", "Pop", "Classic")
>
> # Melakukan Chi-square Test of Independence
> hasil_b <- chisq.test(data_musik)
>
> # Menampilkan hasil
> print(data_musik) # Cek tabel
      Rock RnB Pop Classic
15-25  140  32   5     18
26-35  134  41  52      8
36-45  154  27   8     13
> print(hasil_b)

      Pearson's Chi-squared test

data:  data_musik
X-squared = 64.92, df = 6, p-value = 4.48e-12
```

Analisis Mendalam Output R

1. Definisi Hipotesis

Dimulai dengan menetapkan apa yang diuji:

- H0 (Hipotesis Nol): Tipe musik yang disukai INDEPENDEN (tidak bergantung) terhadap usia. Artinya, pola kesukaan musik sama saja di semua kelompok umur.
- H1 (Hipotesis Alternatif): Tipe musik yang disukai DEPENDEN (bergantung) pada usia. Artinya, ada hubungan antara umur seseorang dengan jenis musik yang mereka pilih.

2. Analisis Hasil Statistik (Dari Gambar)

A. P-value (4.48e-12)

- Nilai: 4.48e-12 adalah penulisan ilmiah.
- Artinya: 0.00000000000448 (sangat mendekati nol).
- Bandingkan dengan Alpha (0.05): Nilai ini jauh lebih kecil daripada 0.05.
- Keputusan: Karena P-value < 0.05, Tolak H0 secara mutlak.

B. X-squared (64.92)

- Nilai Chi-square hitung adalah 64.92.
- Angka ini sangat besar untuk degree of freedom (df) 6.
- Artinya: Terdapat penyimpangan (selisih) yang sangat masif antara data yang diobservasi (kenyataan survei) dengan data yang diharapkan (jika diasumsikan tidak ada hubungan).

C. Degrees of Freedom (df = 6)

- Dihitung dari: $(\text{Baris} - 1) \times (\text{Kolom} - 1)$
- $(3 - 1) \times (4 - 1) = 2 \times 3 = 6$. (Sudah benar sesuai tabel 3x4).

3. Kesimpulan Akhir

Berdasarkan data output tersebut:

- Secara Statistik: Menolak Hipotesis Nol (H_0). Bukti statistik sangat kuat.
- Secara Konteks Masalah: Preferensi musik TIDAK INDEPENDEN terhadap usia.
- Artinya: Umur seseorang sangat mempengaruhi jenis musik yang mereka sukai. Kelompok umur yang berbeda memiliki selera musik yang berbeda secara signifikan.

4. Insight Tambahan

Jika melihat sekilas tabel data di gambar R:

- Perhatikan kolom "Pop".
- Usia 15-25: hanya 5 orang.
- Usia 36-45: hanya 8 orang.
- Tapi Usia 26-35: meledak menjadi 52 orang.

Lonjakan drastis pada kategori "Pop" di usia 26-35 inilah penyumbang terbesar tingginya nilai X-squared (64.92), yang membuktikan bahwa selera musik memang bergantung pada kelompok usianya.