Prana Ugiana Gio
Rezzy Eko Caraka
Dian Utami Sutiksno
Ansari Saleh Ahmar

$$\chi^2 = \sum_{j=1}^k \frac{(f_{b_j} - f_{e_j})^2}{f_{e_j}}$$

Analisis Hubungan Antara Variabel Kategori dengan STATCAL SPSS & Minitab

Analisis Hubungan Antara Variabel Kategori dengan STATCAL SPSS & Minitab

Prana Ugiana Gio (**CEO STATCAL**), lahir di Tanjung Pinang, 1 Oktober 1989. Menyelesaikan Program Studi S1 Matematika di Universitas Sumatera Utara (USU) 2012 dan S2 Matematika USU 2014. Saat ini sedang melanjutkan pendidikan S3 Matematika (2017-sekarang) di USU. Prana Ugiana Gio mendirikan konsultan olah data medan, yang mana telah membimbing mahasiswa S1, S2, dan S3 dalam bidang pengolahan data lebih dari 1300 klien selama 3 tahun.

Rezzy Eko Caraka (**CFO STATCAL**) lahir pada 27 Januari 1994 di Tanjung Balai Karimun Provinsi Kepulauan Riau. Menyelesaikan Pendidikan S1 Statistika Universitas Diponegoro dengan masa studi 3 tahun 5 bulan pada tahun 2015 dan S2 *by research School of Mathematical Sciences Faculty of Science and Technology The National University of Malaysia* dengan bidang riset *Statistics and data science*. Pada Agustus 2017 Mendapatkan penghargaan dari Malaysia *Digital Economy Corporation (MDEC)* pada kegiatan *Big Data Analysis in Medicine* dan *Best Talent JALUMA@4.0 Ministry of Higher Education Malaysia*.

Dian Utami Sutiksno (**Promotion and Brand STATCAL**), menempuh pendidikan S3 (Doktor Ilmu Manajemen) Universitas Padjadjaran Bandung, kompetensi *management, strategic management, marketing* dan *strategic marketing*. Dosen tetap PNS di Politeknik Negeri Ambon sejak tahun 2001, selain sebagai aktifis sosial di beberapa LSM, juga aktif dalam organisasi profesi Dosen Indonesia.

Ansari Saleh Ahmar (**Copyright STATCAL**), lahir tanggal 16 April 1988 di Kabupaten Takalar Propinsi Sulawesi Selatan. Gelar Sarjana Matematika (Sarjana Sains) diperoleh Tahun 2009 tepatnya tanggal 14 November 2009. Pada Tahun 2010 berkesempatan melanjutkan studinya di Program Pascasarjana S2 Matematika (M.Sc.) Universitas Gadjah Mada, Yogyakarta tahun 2012. Saat ini, menjadi Dosen Program Studi Statistika FMIPA Universitas Negeri Makassar dan juga merupakan Presiden dari Komunitas Kolaborasi Publikasi Indonesia (KO2PI).

Analisis Hubungan antara Variabel Kategori dengan STATCAL, SPSS dan Minitab

Prana Ugiana Gio
Rezzy Eko Caraka
Dian Utami Sutiksno
Ansari Saleh Ahmar

KATA PENGANTAR

Alhamdulillah, puji syukur atas kehadiran Allah SWT, karena atas izin-Nya, penulis dapat terus mempertahankan semangat untuk menulis, dan akhirnya dapat menyelesaikan tulisan ini. Hadirnya tulisan ini, tidak semata-mata atas usaha penulis sendiri, melainkan atas izin-Nya. Sungguh suatu kebahagiaan bagi penulis bisa berbagi sebagian kecil ilmu pengetahuan milik-Nya melalui tulisan yang berjudul “**Analisis Hubungan Antara Variabel Kategori dengan STATCAL, SPSS dan Minitab**”.

Ucapan terima kasih penulis sampaikan kepada semua pihak yang telah membantu dalam rangka penyelesaian tulisan ini. Penulis menyadari bahwa tulisan ini tentunya masih perlu perbaikan, sehingga penulis mengharapkan kritik dan saran yang membangun dari para pembaca agar tulisan ini dapat menjadi lebih baik. Kritik dan saran dapat ditujukan ke alamat email gioprana89@gmail.com.

Medan, 3 Mei 2018

Prana Ugiana Gio

Robert Kurniawan

Dina Nazriani

DAFTAR ISI

- Bagian 1: Contoh Data Variabel Kategori [1]
- Bagian 2: Data Nominal dan Ordinal [3]
- Bagian 3: Perbedaan Data Nominal dan Ordinal [3]
- Bagian 4: Berbagai Penjelasan Mengenai Hubungan Antara Variabel Kategori [4]
- Bagian 5: Tabel Kontingensi (Contingency Table) [5]
- Bagian 6: Frekuensi Pengamatan (Observed Frequency) dan Frekuensi Harapan (Expected Frequency) [6]
- Bagian 7: Uji Chi-Square Pearson dan Fisher's Exact [8]
- Bagian 8: Benarkah Uji Chi-Square Pearson Hanya Untuk 2 x 2? [10]
- Bagian 9: Perhitungan Manual Uji Chi-Square Pearson [11]
- Bagian 10: Penyelesaian dengan STATCAL [14]
- Bagian 11: Penyelesaian dengan SPSS [21]
- Bagian 12: Penyelesaian dengan Minitab [25]
- Bagian 13: Apakah Uji Chi-Square Pearson Dapat Digunakan untuk 1x2 atau 2x1 atau 3x1 atau 1x3 atau 1x4 atau 4x1??? [28]
- Bagian 14: Contoh Kasus dengan STATCAL, SPSS dan Minitab [31]
- Bagian 15: Mengukur Keeratan Hubungan Variabel Kategori dengan Korelasi Cramer's V [35]

Bagian 1: Contoh Data Variabel Kategori

Berikut diberikan contoh data variabel kategori.

Tabel 1.1 Contoh Data Variabel Kategori

Responden	Jenis Kelamin	Hobi	Tingkat Pendidikan	Tingkat Kebahagiaan
1	Laki-Laki	Olahraga	S1	Kurang Bahagia
2	Laki-Laki	Olahraga	S1	Kurang Bahagia
3	Laki-Laki	Olahraga	S1	Kurang Bahagia
4	Laki-Laki	Memasak	S1	Kurang Bahagia
5	Laki-Laki	Olahraga	S1	Kurang Bahagia
6	Perempuan	Memasak	S1	Kurang Bahagia
7	Perempuan	Membaca	S1	Kurang Bahagia
8	Perempuan	Membaca	S1	Kurang Bahagia
9	Perempuan	Olahraga	S1	Kurang Bahagia
10	Perempuan	Membaca	S1	Kurang Bahagia
11	Laki-Laki	Memasak	S1	Kurang Bahagia
12	Laki-Laki	Olahraga	S1	Bahagia
13	Laki-Laki	Membaca	S1	Bahagia
14	Laki-Laki	Memasak	S1	Bahagia
15	Laki-Laki	Olahraga	S1	Bahagia
16	Laki-Laki	Olahraga	S1	Kurang Bahagia
17	Laki-Laki	Olahraga	S1	Sangat Bahagia
18	Laki-Laki	Olahraga	S1	Sangat Bahagia
19	Laki-Laki	Memasak	S1	Sangat Bahagia
20	Laki-Laki	Olahraga	S1	Sangat Bahagia
21	Perempuan	Membaca	S2	Bahagia
22	Perempuan	Membaca	S2	Bahagia
23	Perempuan	Memasak	S2	Bahagia
24	Perempuan	Membaca	S2	Bahagia
25	Perempuan	Membaca	S2	Bahagia
26	Laki-Laki	Memasak	S2	Bahagia
27	Laki-Laki	Olahraga	S2	Bahagia
28	Laki-Laki	Membaca	S2	Bahagia
29	Laki-Laki	Memasak	S2	Bahagia
30	Laki-Laki	Olahraga	S2	Bahagia
31	Laki-Laki	Olahraga	S2	Bahagia
32	Laki-Laki	Olahraga	S2	Kurang Bahagia
33	Laki-Laki	Olahraga	S2	Kurang Bahagia
34	Laki-Laki	Memasak	S2	Kurang Bahagia
35	Laki-Laki	Olahraga	S2	Kurang Bahagia
36	Perempuan	Memasak	S2	Sangat Bahagia
37	Perempuan	Memasak	S2	Bahagia
38	Perempuan	Memasak	S2	Sangat Bahagia

39	Perempuan	Membaca	S2	Sangat Bahagia
40	Perempuan	Membaca	S2	Sangat Bahagia
41	Laki-Laki	Memasak	S3	Sangat Bahagia
42	Laki-Laki	Olahraga	S3	Sangat Bahagia
43	Laki-Laki	Membaca	S3	Sangat Bahagia
44	Laki-Laki	Memasak	S3	Sangat Bahagia
45	Laki-Laki	Olahraga	S3	Sangat Bahagia
46	Laki-Laki	Olahraga	S3	Sangat Bahagia
47	Laki-Laki	Olahraga	S3	Sangat Bahagia
48	Laki-Laki	Olahraga	S3	Sangat Bahagia
49	Laki-Laki	Memasak	S3	Sangat Bahagia
50	Laki-Laki	Olahraga	S3	Sangat Bahagia
51	Perempuan	Memasak	S3	Sangat Bahagia
52	Perempuan	Membaca	S3	Kurang Bahagia
53	Perempuan	Membaca	S3	Kurang Bahagia
54	Perempuan	Olahraga	S3	Kurang Bahagia
55	Perempuan	Membaca	S3	Kurang Bahagia
56	Laki-Laki	Memasak	S3	Sangat Bahagia
57	Laki-Laki	Olahraga	S3	Bahagia
58	Laki-Laki	Membaca	S3	Bahagia
59	Laki-Laki	Memasak	S3	Bahagia
60	Laki-Laki	Olahraga	S3	Bahagia

Tabel 1.1 merupakan contoh data variabel kategori. Diketahui:

- ⇒ Variabel jenis kelamin merupakan variabel kategori, yakni terdiri dari 2 kategori, laki-laki dan perempuan.
- ⇒ Variabel hobi merupakan variabel kategori, yakni terdiri dari 3 kategori, memasak, olahraga dan membaca.
- ⇒ Tingkat pendidikan merupakan variabel kategori, yakni terdiri dari 3 kategori, S1, S2, dan S3.
- ⇒ Tingkat kebahagiaan merupakan variabel kategori, yakni terdiri dari 3 kategori, kurang bahagia, bahagia dan sangat bahagia.

Bagian 2: Data Nominal dan Ordinal

Berdasarkan data pada Tabel 1.1, data pada jenis kelamin dan hobi merupakan data **nominal**, sementara data pada tingkat pendidikan dan tingkat kebahagiaan merupakan data **ordinal**.

Bagian 3: Perbedaan Data Nominal dan Ordinal

Perbedaan antara data nominal dan data ordinal adalah pada data nominal tidak memperhatikan urutan atau tingkatan. Maksudnya laki-laki tidak lebih tinggi dibandingkan perempuan atau sebaliknya perempuan tidak lebih tinggi dibandingkan laki-laki. Begitu juga memasak tidak lebih tinggi olahraga atau sebaliknya olahraga tidak lebih tinggi dibandingkan memasak.

Namun pada data ordinal memperhatikan urutan atau tingkatan. Misalkan pada variabel tingkat pendidikan, jenjang pendidikan S2 lebih tinggi dibandingkan jenjang pendidikan S1, jenjang pendidikan S3 lebih tinggi dibandingkan jenjang pendidikan S2. Namun perhatikan bahwa

Berapakah $S1 + S2$??

Berapakah

Berapakah $S3 - S1$???

Data tersebut adalah data kategori, di mana data kategori tidak bisa dilakukan operasi matematika. Seringkali kita memberi label untuk data kategori, misalkan angka 1 untuk label S1, angka 2 untuk label S2 dan angka 3 untuk label S3. Selanjutnya angka-angka tersebut digunakan untuk operasi matematika, misalkan

$$1 + 2 = 3 \text{ (} S1 + S2 = S3 \text{)}$$

$$1 - 2 = -1 \text{ (} S1 - S2 = \text{?????)}$$

$$3 \times 2 = 6 \text{ (} S3 \times S2 = \text{?????)}$$

Perhatikan bahwa angka-angka tersebut (1, 2 dan 3) pada dasarnya hanya berupa pengkodean saja (bukan angka sejati) sehingga tidak tepat untuk digunakan dalam operasi matematika.

Bagian 4: Berbagai Penjelasan Mengenai Hubungan Antara Variabel Kategori

Berikut disajikan kutipan penjelasan dari beberapa buku terkait hubungan antara variabel kategori.

⇒ Andy Field (2009:688) dalam bukunya yang berjudul “*Discovering Statistics Using SPSS, 3rd Edition*” menyatakan sebagai berikut.

“If we want to see whether **there’s a relationship between two categorical variables** (i.e. does the amount of cats that line dance relate to the type of training used?) **we can use the Pearson’s chi-square test** (Fisher, 1922; Pearson, 1900). This is an extremely elegant statistic based on the simple idea of comparing the **frequencies you observe in certain categories to the frequencies you might expect** to get in those categories by chance. All the way back in Chapters 2, 7 and 10 we saw that if we fit a model to any set of data we can evaluate that model using a very simple equation (or some variant of it).”

⇒ Prem S. Mann (2013:535) dalam bukunya yang berjudul “*Introductory Statistics, 8th Edition*” menyatakan sebagai berikut.

“In a **test of independence for a contingency table**, we test the **null hypothesis that the two attributes (characteristics) of the elements of a given population are not related** (that is, they are independent) against the alternative hypothesis that the two characteristics are related (that is, they are dependent). For example, we may want to test if the affiliation of people with the Democratic and Republican parties is independent of their income levels. We perform such a test by using the chi-square distribution. As another example, **we may want to test if there is an association between being a man or a woman and having a preference for watching sports or soap operas on television.**”

⇒ Alan Agresti dan Barbara Finlay (2009:221) dalam bukunya yang berjudul “*Statistical Methods for the Social Sciences, 4th Edition*” menyatakan sebagai berikut.

“Recall that we say there is **an association between two variables** if the distribution of the response variable changes in some way as the value of the explanatory variable changes. **In comparing groups, an association exists if the population means or population proportions differ between the groups.**

This chapter presents **methods for detecting and describing associations between two categorical variables**. The methods of this chapter help us answer a question such as, "**Is there an association between happiness and whether one is religious?**"

Bagian 5: Tabel Kontingensi (*Contingency Table*)

Tabel 5.1 diberikan contoh tabel kontingensi 2×3 .

Tabel 5.1 Contoh Tabel Kontingensi					
		Hobi			Total
		Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	12	24	4	40
	Perempuan	6	2	12	20
Total		18	26	16	60

Perhatikan bahwa:

- ⇒ Perhatikan bahwa " 2×3 " maksudnya adalah tabel kontingensi tersebut terdiri dari 2 baris dan 3 kolom.
- ⇒ Nama lain dari tabel kontingensi adalah ***cross-tabulation*** (tabulasi silang).
- ⇒ Tabel kontingensi dapat berukuran 2×3 , 2×4 , 3×4 , dan seterusnya.
- ⇒ Angka pertama menyatakan **jumlah baris**, sedangkan angka kedua menyatakan **jumlah kolom**. Jadi, tabel kontingensi 3×4 memiliki 3 baris dan 4 kolom.
- ⇒ Tabel kontingensi menyajikan informasi frekuensi untuk setiap kombinasi kategori.
- ⇒ Berdasarkan Tabel 5.1, diketahui dari 40 laki-laki, 12 di antaranya hobi memasak, 24 olahraga, dan 4 membaca. Sementara dari 20 perempuan, 6 di antaranya hobi memasak, 2 olahraga dan 12 membaca.
- ⇒ Perhatikan bahwa pada Tabel 5.1 terdiri dari 6 **cell** atau kotak ($2 \times 3 = 6$) atau terdiri dari 6 kombinasi kategori, yakni $\{(\text{laki-laki, memasak}), (\text{laki-laki, olahraga}), (\text{laki-laki, membaca}), (\text{perempuan, memasak}), (\text{perempuan, olahraga}), (\text{perempuan, membaca})\}$.

Prem S. Mann (2013:534) dalam bukunya yang berjudul “*Introductory Statistics, 8th Edition*” menyatakan sebagai berikut.

“11.3.1 A Contingency Table Often we may have information on **more than one variable for each element**. Such information can be summarized and presented using a **two-way classification table**, which is also called **a contingency table or cross-tabulation**. Suppose a university has a total of 20,758 students enrolled. By classifying these students based on gender and whether these students are full-time or part-time, we can prepare Table 11.5, which provides an example of a contingency table. Table 11.5 has two rows (one for males and the second for females) and two columns (one for full-time and the second for part-time students). Hence, it is also called a 2 x 2 (read as “two by two”) contingency table.

A contingency table can be of any size. For example, it can be 2x3, 3x2, 3x3, or 4x2. Note that in these notations, **the first digit refers to the number of rows in the table**, and the **second digit refers to the number of columns**. For example, a 3x2 table will contain three rows and two columns. In general, an RxC table contains R rows and C columns.”

Alan Agresti dan Barbara Finlay (2009:221) dalam bukunya yang berjudul “*Statistical Methods for the Social Sciences, 4th Edition*” menyatakan sebagai berikut.

“8.1 Contingency Tables Data for the analysis of categorical variables are displayed in contingency tables. This type of table displays the number of subjects observed at all combinations of possible outcomes for the two variables.”

Bagian 6: Frekuensi Pengamatan (Observed Frequency) dan Frekuensi Harapan (Expected Frequency)

Apa itu frekuensi pengamatan (*observed frequency*)? Apa itu frekuensi harapan (*expected frequency*)?

- ⇒ Berdasarkan Tabel 5.1, diketahui jumlah laki-laki yang hobi memasak sebanyak 12.
Nah, 12 itu merupakan frekuensi pengamatan (*observed frequency*).
- ⇒ Berdasarkan Tabel 5.1, diketahui jumlah laki-laki yang hobi olahraga sebanyak 24.
Nah, 24 itu merupakan frekuensi pengamatan (*observed frequency*).
- ⇒ Berdasarkan Tabel 5.1, diketahui jumlah laki-laki yang hobi membaca sebanyak 4.
Nah, 4 itu merupakan frekuensi pengamatan (*observed frequency*).
- ⇒ Berdasarkan Tabel 5.1, diketahui jumlah perempuan yang hobi memasak sebanyak 6.
Nah, 6 itu merupakan frekuensi pengamatan (*observed frequency*).
- ⇒ Berdasarkan Tabel 5.1, diketahui jumlah perempuan yang hobi olahraga sebanyak 2.
Nah, 2 itu merupakan frekuensi pengamatan (*observed frequency*).

⇒ Berdasarkan Tabel 5.1, diketahui jumlah perempuan yang hobi membaca sebanyak 12.

Nah, 12 itu merupakan frekuensi pengamatan (*observed frequency*).

Untuk frekuensi harapan (*expected frequency*) perlu dihitung. Misalkan akan dihitung frekuensi harapan pada kotak baris laki-laki dan kolom memasak.

$$\begin{aligned} \text{Frekuensi Harapan}_{(\text{laki-laki; memasak})} &= \frac{(\text{jumlah laki-laki}) \times (\text{jumlah memasak})}{\text{Total}} \\ &= \frac{(40) \times (18)}{60} \\ &= 12 \end{aligned}$$

Tabel 6.1 Frekuensi Harapan pada Kotak Baris Laki-Laki dan Kolom Memasak

		Hobi			Total
		Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	12 (12)	24	4	40
	Perempuan	6	2	12	20
Total		18	26	16	60

Berdasarkan perhitungan sebelumnya, diperoleh frekuensi harapan (*expected frequency*) pada kotak baris laki-laki dan kolom memasak, yakni 12 (frekuensi harapan diapit buka-tutup kurung). Misalkan lagi akan dihitung frekuensi harapan (*expected frequency*) pada kotak baris perempuan dan kolom olahraga.

$$\begin{aligned} \text{Frekuensi Harapan}_{(\text{perempuan; olahraga})} &= \frac{(\text{jumlah perempuan}) \times (\text{jumlah olahraga})}{\text{Total}} \\ &= \frac{(20) \times (26)}{60} \\ &= 8,6667 \\ &\cong 8,7 \end{aligned}$$

Tabel 6.2 Frekuensi Harapan pada Kotak Baris Perempuan dan Kolom Olahraga

		Hobi			Total
		Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	12 (12)	24	4	40
	Perempuan	6	2 (8,6667)	12	20
Total		18	26	16	60

Berdasarkan perhitungan sebelumnya, diperoleh frekuensi harapan (*expected frequency*) pada kotak baris perempuan dan kolom olahraga, yakni 8,7 (frekuensi harapan diapit buka-tutup kurung). Dalam SPSS, dapat ditampilkan nilai-nilai dari frekuensi harapan untuk setiap kotak. Perhatikan Tabel 6.3.

Tabel 6.3 Frekuensi Pengamatan (*Observed Frequency*) dan Frekuensi Harapan (*Expected Frequency*) (Output SPSS)
Jenis Kelamin * Hobi Crosstabulation

			Hobi			Total
			Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	Count	12	24	4	40
		Expected Count	12.0	17.3	10.7	40.0
		% within Jenis Kelamin	30.0%	60.0%	10.0%	100.0%
	Perempuan	Count	6	2	12	20
		Expected Count	6.0	8.7	5.3	20.0
		% within Jenis Kelamin	30.0%	10.0%	60.0%	100.0%
Total	Count		18	26	16	60
	Expected Count		18.0	26.0	16.0	60.0
	% within Jenis Kelamin		30.0%	43.3%	26.7%	100.0%

Dalam SPSS, istilah bahasa inggris untuk frekuensi harapan adalah *expected count*.

Bagian 7: Uji Chi-Square Pearson dan Fisher's Exact

Uji chi-square Pearson dan uji *Fisher's exact* dapat digunakan untuk menguji signifikansi hubungan dua variabel kategori. Namun pada uji chi-square Pearson memiliki kelemahan. Apa kelemahannya? Kutipan pernyataan dari beberapa buku berikut menyatakan kelemahan uji chi-square Pearson. Alan Agresti dan Barbara Finlay (2009:227) dalam bukunya yang berjudul "*Statistical Methods for the Social Sciences, 4th Edition*" menyatakan sebagai berikut.

"Sample Size Requirements *The chi-squared test, like one and two-sample z tests for proportions, is a large sample test. The chi-squared distribution is the sampling distribution of the χ^2 test statistic only if the sample size is large. A rough guideline for this requirement is that the expected frequency f_e should exceed 5 in each cell. Otherwise, the chi-squared distribution may poorly approximate the actual distribution of the χ^2 statistic"*

Prem S. Mann (2013:537) dalam bukunya yang berjudul “*Introductory Statistics, 8th Edition*” menyatakan sebagai berikut.

“To apply a chi-square test of independence, the sample size should be large enough so that the expected frequency for each cell is at least 5. If the expected frequency for a cell is not at least 5, we either increase the sample size or combine some categories. Examples 11–6 and 11–7 describe the procedure to make tests of independence using the chi-square distribution.”

Andy Field (2009:690) dalam bukunya yang berjudul “*Discovering Statistics Using SPSS, 3rd Edition*” menyatakan sebagai berikut.

“There is one problem with the chi-square test, which is that the sampling distribution of the test statistic has an approximate chi-square distribution. The larger the sample is, the better this approximation becomes and in large samples the approximation is good enough to not worry about the fact that it is an approximation. However, in small samples, the approximation is not good enough, making significance tests of the chi-square distribution inaccurate. This is why you often read that to use the chi-square test the expected frequencies in each cell must be greater than 5 (see section 18.4). When the expected frequencies are greater than 5, the sampling distribution is probably close enough to a perfect chi-square distribution for us not to worry. However, when the expected frequencies are too low, it probably means that the sample size is too small and that the sampling distribution of the test statistic is too deviant from a chi-square distribution to be of any use.”

Berdasarkan uraian di atas, dapat disimpulkan bahwa syarat pada penggunaan uji chi-square Pearson, ketika terdapat *cell* atau kotak, dengan nilai frekuensi harapan (*expected frequency*) kurang dari 5, maka distribusi sampling statistik chi-square Pearson akan buruk dalam hal mendekati distribusi chi-square. Sehingga akan memperoleh hasil yang tidak akurat. Tabel 7.1 disajikan hasil berdasarkan SPSS untuk uji chi-square Pearson dan *Fisher’s exact*.

Tabel 7.1 Hasil Uji Chi-Square Pearson dan Fisher’s Exact (Output SPSS)

Chi-Square Tests						
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	9.090 ^a	2	.011	.010		
Likelihood Ratio	9.313	2	.009	.012		
Fisher's Exact Test	8.658			.010		
Linear-by-Linear Association	1.799 ^b	1	.180	.211	.130	.068
N of Valid Cases	60					

a. 3 cells (50.0%) have expected count less than 5. The minimum expected count is 2.60.

b. The standardized statistic is 1.341.

Berdasarkan Tabel 7.1:

- ⇒ Terdapat pernyataan “**3 cells (50.0%) have expected count less than 5**”. Pernyataan tersebut memberitahu kita bahwa ada 3 *cell* atau kotak dengan nilai frekuensi harapan (*expected frequency*) < 5 .
- ⇒ Meskipun terdapat *cell* dengan nilai frekuensi harapan < 5 , **uji chi-square Pearson tetap dapat digunakan, dengan memperhatikan kolom *Exact Sig. (2-sided)*** (bukan kolom *Asymp. Sig. 2-sided*). Hasil pada kolom *Exact Sig. (2-sided)* merupakan hasil eksak (sebenarnya).
- ⇒ Hasil-hasil pada kolom *Exact Sig. (2-sided)* dapat digunakan tanpa memperdulikan syarat frekuensi harapan harus > 5 , baik itu uji chi-square Pearson ataupun *Fisher's exact*. Mengapa demikian? Hal dikarenakan hasil-hasil pada kolom *Exact Sig. (2-sided)* menggunakan metode perhitungan pendekatan eksak, yang mana terpenuhinya atau tidak dari persyaratan frekuensi harapan > 5 , tetap memberikan hasil yang akurat (eksak / sebenarnya).

Alan Agresti dan Barbara Finlay (2009:228) dalam bukunya yang berjudul “*Statistical Methods for the Social Sciences, 4th Edition*” menyatakan sebagai berikut.

“For 2x2 contingency tables, a small-sample test of independence is Fisher's exact test, discussed in Section 7.5. This test extends to tables of arbitrary size $r \times c$ but requires specialized software, such as SAS (the EXACT option in PROC FREQ) or SPSS (the Exact module).

If you have such software, you can use the exact test for any sample size. You don't need to use the chi-squared approximation. For Table 8.4, the exact test also gives $P\text{-value} = 0.0003$.”

Bagian 8: Benarkah Uji Chi-Square Pearson Hanya Untuk 2×2 ?

Tidak benar bahwa uji chi-square Pearson hanya diperuntukkan 2×2 . Uji chi-square Pearson dapat digunakan untuk ukuran selain dari 2×2 , misalkan 2×3 , 3×2 , 2×5 , 6×6 , dan seterusnya.

Prem S. Mann (2013:534) dalam bukunya yang berjudul “*Introductory Statistics, 8th Edition*” menyatakan sebagai berikut.

“A contingency table can be of any size. For example, it can be 2x3, 3x2, 3x3, or 4x2. Note that in these notations, **the first digit refers to the number of rows in the table**, and the **second digit refers to the number of columns**. For example, a 3x2 table will contain three rows and two columns. In general, an RxC table contains R rows and C columns.”

Bagian 9: Perhitungan Manual Uji Chi-Square Pearson

Data pada Tabel 5.1 disajikan kembali pada Tabel 9.1.

Tabel 9.1 Contoh Tabel Kontingensi

		Hobi			Total
		Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	12	24	4	40
	Perempuan	6	2	12	20
Total		18	26	16	60

Berdasarkan data pada Tabel 9.1:

- ⇒ Terdapat kecenderungan laki-laki suka olahraga, namun tidak suka membaca.
- ⇒ Terdapat kecenderungan perempuan suka membaca, namun tidak suka olahraga.

Akan digunakan uji chi-square Pearson untuk menguji apakah terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%. Apakah jenis kelamin mempengaruhi seseorang dalam hal pemilihan hobi.

Tahap pertama adalah menghitung nilai frekuensi harapan (*expected frequency*) untuk setiap kotak atau *cell*.

$$\begin{aligned}
 f_e(\text{laki} - \text{laki}; \text{memasak}) &= \frac{(\text{jumlah laki} - \text{laki}) \times (\text{jumlah memasak})}{\text{Total}} \\
 &= \frac{(40) \times (18)}{60} \\
 &= 12
 \end{aligned}$$

$$\begin{aligned}
 f_e(\text{laki} - \text{laki}; \text{olahraga}) &= \frac{(\text{jumlah laki} - \text{laki}) \times (\text{jumlah olahraga})}{\text{Total}} \\
 &= \frac{(40) \times (26)}{60} \\
 &= 17,333333
 \end{aligned}$$

$$\begin{aligned}
 f_e(\text{laki} - \text{laki}; \text{membaca}) &= \frac{(\text{jumlah laki} - \text{laki}) \times (\text{jumlah membaca})}{\text{Total}} \\
 &= \frac{(40) \times (16)}{60} \\
 &= 10,666667
 \end{aligned}$$

$$\begin{aligned}
 f_e(\text{perempuan}; \text{memasak}) &= \frac{(\text{jumlah perempuan}) \times (\text{jumlah memasak})}{\text{Total}} \\
 &= \frac{(20) \times (18)}{60} \\
 &= 6
 \end{aligned}$$

$$\begin{aligned}
 f_e(\text{perempuan}; \text{olahraga}) &= \frac{(\text{jumlah perempuan}) \times (\text{jumlah olahraga})}{\text{Total}} \\
 &= \frac{(20) \times (26)}{60} \\
 &= 8,666667
 \end{aligned}$$

$$\begin{aligned}
 f_e(\text{perempuan}; \text{membaca}) &= \frac{(\text{jumlah perempuan}) \times (\text{jumlah membaca})}{\text{Total}} \\
 &= \frac{(20) \times (16)}{60} \\
 &= 5,333333
 \end{aligned}$$

Berdasarkan perhitungan frekuensi harapan di atas, disajikan Tabel 9.2. Tabel 9.2 menyajikan frekuensi pengamatan (*observed frequency*) dan frekuensi harapan (*expected frequency*).

Tabel 9.2 Frekuensi Pengamatan dan Frekuensi Harapan

		Hobi			Total
Jenis Kelamin	Laki-Laki	Memasak	Olahraga	Membaca	
	Perempuan				
	Laki-Laki	12 (12)	24 (17,3333)	4 (10,6667)	40
	Perempuan	6 (6)	2 (8,6667)	12 (5,3333)	20
Total		18	26	16	60

Berdasarkan Tabel 9.2, tidak ada satu *cell* pun dengan nilai frekuensi harapan (*expected frequency*) < 5.

Tahap kedua adalah menghitung nilai statistik chi-square Pearson. Nilai statistik chi-square Pearson dinotasikan dengan lambang χ^2 , dihitung dengan rumus sebagai berikut.

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

Berikut perhitungan nilai statistik chi-square Pearson.

$$\begin{aligned} \chi^2 &= \sum \frac{(f_o - f_e)^2}{f_e} \\ &= \frac{(12 - 12)^2}{12} + \frac{(24 - 17,3333)^2}{17,3333} + \frac{(4 - 10,6667)^2}{10,6667} + \\ &\quad \frac{(6 - 6)^2}{6} + \frac{(2 - 8,6667)^2}{8,6667} + \frac{(12 - 5,3333)^2}{5,3333} \\ \chi^2 &= 20,192 \end{aligned}$$

Berdasarkan hasil perhitungan, diperoleh nilai statistik chi-square Pearson 20,192. Dengan menggunakan SPSS diperoleh hasil yang sama (Tabel 9.3).

Tabel 9.3 Hasil Perhitungan Statistik Chi-Square Pearson dengan SPSS
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	20.192 ^a	2	.000	.000		
Likelihood Ratio	21.371	2	.000	.000		
Fisher's Exact Test	20.211			.000		
Linear-by-Linear Association	5.796 ^b	1	.016	.018	.012	.008
N of Valid Cases	60					

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.33.

b. The standardized statistic is 2.407.

Tahap ketiga adalah menghitung nilai kritis chi-square. Untuk menentukan nilai kritis chi-square, terlebih dahulu menghitung derajat bebas (*degree of freedom*). Derajat bebas dihitung dengan rumus sebagai berikut.

$$df = (r - 1)(c - 1)$$

Perhatikan bahwa r dan c masing-masing menyatakan jumlah baris dan kolom dalam tabel kontingensi. Untuk tabel kontingensi dengan ukuran 2×3 , $r = 2$ dan $c = 3$, maka diperoleh $df = (2 - 1)(3 - 1) = 2$. Nilai kritis chi-square dengan derajat bebas 2 dan tingkat signifikansi 5% adalah 5,99. Nilai kritis chi-square dapat dihitung dengan *Microsoft Excel*, seperti pada Gambar 9.1.

D	E	F	G
	Derajat Bebas	Tingkat Signifikansi	Nilai Kritis Chi-Square
	2	0.05	5.991464547

Gambar 9.1 Menghitung Nilai Kritis Chi-Square dengan Microsoft Excel

Tahap terakhir adalah pengambilan keputusan terhadap hipotesis. Diketahui nilai statistik chi-square adalah $20,192 >$ nilai kritis chi-square 5,99, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%.

Bagian 10: Penyelesaian dengan STATCAL

Contoh kasus pada Bagian 9 akan diselesaikan dengan STATCAL. Data aslinya tersaji pada Tabel 1.1. Untuk menginput data di STATCAL, data perlu disajikan seperti pada Tabel 10.1.

Tabel 10.1 Data Jenis Kelamin dan Hobi dari 40 Responden

Responden	Jenis Kelamin	Hobi
1	1	2
2	1	2
3	1	2
4	1	1
5	1	2

6	2	1
7	2	3
8	2	3
9	2	2
10	2	3
11	1	1
12	1	2
13	1	3
14	1	1
15	1	2
16	1	2
17	1	2
18	1	2
19	1	1
20	1	2
21	2	3
22	2	3
23	2	1
24	2	3
25	2	3
26	1	1
27	1	2
28	1	3
29	1	1
30	1	2
31	1	2
32	1	2
33	1	2
34	1	1
35	1	2
36	2	1
37	2	1
38	2	1
39	2	3
40	2	3
41	1	1
42	1	2
43	1	3
44	1	1
45	1	2
46	1	2
47	1	2
48	1	2
49	1	1
50	1	2
51	2	1

52	2	3
53	2	3
54	2	2
55	2	3
56	1	1
57	1	2
58	1	3
59	1	1
60	1	2

Berdasarkan data pada Tabel 10.1:

- ⇒ Untuk variabel jenis kelamin, angka 1 menyatakan laki-laki, sementara angka 2 menyatakan perempuan.
- ⇒ Untuk variabel hobi, angka 1 menyatakan memasak, angka 2 menyatakan olahraga dan angka 3 menyatakan membaca.

Data pada Tabel 10.1 diinput dalam STATCAL seperti pada Gambar 10.1.

Categorical Data

Name of Variable

Jenis Kelamin
 Hobi

Label

	V1	V2	V3
1	Laki-Laki	Memasak	
2	Perempuan	Olahraga	
3		Membaca	
4			
5			

Set Number of Column

1 28

Set Number of Row

1 70

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	1	2							
2	1	2							
3	1	2							
4	1	1							
5	1	2							
6	2	1							
7	2	3							
8	2	3							
9	2	2							
10	2	3							
11	1	1							
12	1	2							
13	1	3							
14	1	1							
15	1	2							

Gambar 10.1 Input Data Kategori dalam STATCAL

Selanjutnya pilih **Statistics => Association between Categorical Variables** (Gambar 10.2).

Gambar 10.2 Association between Categorical Variables

Perhatikan Gambar 10.3. Pindahkan variabel jenis kelamin ke kotak sebelah kanan (kotak atas) dan juga pindahkan variabel hobi ke kotak sebelah kanan (kotak bawah). Hasilnya dapat dilihat pada bagian **Result**.

Gambar 10.3 Association between Categorical Variables

Pada bagian **Result**, tersaji output grafik batang frekuensi, seperti pada Gambar 10.4 sampai dengan Gambar 10.7.

Gambar 10.4 Grafik Batang Frekuensi berdasarkan Jenis Kelamin

Berdasarkan grafik batang frekuensi pada Gambar 10.4, diketahui dari 60 responden, terdapat 40 (66,67%) responden dengan jenis kelamin laki-laki, sementara 20 (33,33%) responden dengan jenis kelamin perempuan.

Gambar 10.5 Grafik Batang Frekuensi berdasarkan Hobi

Berdasarkan grafik batang frekuensi pada Gambar 10.5, diketahui dari 60 responden, terdapat 18 (30%) responden dengan hobi memasak, 26 (43,33%) responden dengan hobi olahraga dan 16 (26,67%) responden dengan hobi membaca.

Gambar 10.6 Grafik Batang Frekuensi berdasarkan Jenis Kelamin dan Hobi

Gambar 10.7 Grafik Batang Frekuensi berdasarkan Jenis Kelamin dan Hobi

Berdasarkan grafik batang frekuensi pada Gambar 10.6 dan Gambar 10.7, diketahui:

- ⇒ Dari 18 responden yang hobi memasak, 12 (67%) di antaranya laki-laki, sementara 6 (33%) perempuan.
- ⇒ Dari 26 responden yang hobi olahraga, 24 (92%) di antaranya laki-laki, sementara 2 (8%) perempuan.
- ⇒ Dari 16 responden yang hobi membaca, 4 (25%) di antaranya laki-laki, sementara 12 (75%) perempuan.

Pada bagian **Result** juga terdapat *output* dari hasil uji chi-square Pearson dan *Fisher's exact*, seperti pada Gambar 10.8 dan Gambar 10.9.

Pearson's Chi-Squared Test

Show entries Search:

Statistic of Chi-Squared	Degree of Freedom	P-Value	Chi-Squared Critical Value (5%)
20.1923	2	0	5.9915

Gambar 10.8 Hasil Uji Chi-Square Pearson

Berdasarkan hasil uji uji chi-square Pearson pada Gambar 10.8, diketahui:

- ⇒ Nilai statistik chi-square Pearson 20,1923.
- ⇒ Nilai kritis chi-square dengan derajat bebas 2 dan tingkat signifikansi 5% adalah 5,9915.
- ⇒ Nilai statistik chi-square Pearson 20,1923 > nilai kritis chi-square 5,9915, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%.
- ⇒ Cara lain pengambilan keputusan, diketahui nilai *P-Value* 0,000 < tingkat signifikansi 0,05, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%.

Fisher's Exact Test for Count Data

Show entries Search:

	p_value
Fisher's Exact Test	0

Gambar 10.9 Hasil Uji Fisher's Exact

⇒ Selain uji chi-square Pearson, dapat juga digunakan uji *Fisher's exact*. Diketahui nilai *P-Value* $0,000 < \text{tingkat signifikansi } 0,05$, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%.

Bagian 11: Penyelesaian dengan SPSS

Contoh kasus pada Bagian 9 akan diselesaikan dengan SPSS. Input data pada Tabel 10.1 dalam SPSS seperti pada Gambar 11.1.

Jenis_Kelamin	Hobi	var
1.00	2.00	
1.00	2.00	
1.00	2.00	
1.00	1.00	
1.00	2.00	
2.00	1.00	
2.00	3.00	
2.00	3.00	
2.00	2.00	
2.00	3.00	
1.00	1.00	
1.00	2.00	
1.00	3.00	
1.00	1.00	
1.00	2.00	
1.00	2.00	
1.00	2.00	
1.00	1.00	
1.00	2.00	
2.00	3.00	

	Jenis_Kelamin	Hobi
1	Laki-Laki	Olahraga
2	Laki-Laki	Olahraga
3	Laki-Laki	Olahraga
4	Laki-Laki	Memasak
5	Laki-Laki	Olahraga
6	Perempuan	Memasak
7	Perempuan	Membaca
8	Perempuan	Membaca
9	Perempuan	Olahraga
10	Perempuan	Membaca
11	Laki-Laki	Memasak
12	Laki-Laki	Olahraga
13	Laki-Laki	Membaca
14	Laki-Laki	Memasak
15	Laki-Laki	Olahraga
16	Laki-Laki	Olahraga
17	Laki-Laki	Olahraga
18	Laki-Laki	Olahraga
19	Laki-Laki	Memasak
20	Laki-Laki	Olahraga
21	Perempuan	Membaca
22	Perempuan	Membaca
23	Perempuan	Memasak
24	Perempuan	Membaca
25	Perempuan	Membaca

Gambar 11.1 Input Data Kategori dalam SPSS

Selanjutnya pilih **Analyze => Descriptive Statistics => Crosstabs** (Gambar 11.2), sehingga muncul tampilan kotak **Crosstabs** (Gambar 11.3). Pindahkan variabel jenis kelamin ke kotak **Row(s)**, sementara pindahkan variabel hobi ke kotak **Column(s):** (Gambar 11.3). Selanjutnya pilih **Exact**, sehingga muncul tampilan kotak **Exact Tests** (Gambar 11.4). Pada Gambar 11.4, pilih **Exact**. Kemudian pilih **Continue**. Maka tampilan kembali seperti pada Gambar 11.3.

Gambar 11.2

Gambar 11.3

Gambar 11.4

Selanjutnya pilih **Statistics**, sehingga muncul tampilan **Crosstabs: Statistics** (Gambar 11.5). Pada Gambar 11.5, pilih **Chi-Square**. Kemudian pilih **Continue**, sehingga tampilan kembali seperti pada Gambar 11.3.

Gambar 11.5

Selanjutnya pilih **Cells**, sehingga muncul kotak **Crosstabs: Cell Display** (Gambar 11.6). Pilih **Counts: Observed** dan **Percentages: Row**. Selanjutnya pilih **Continue** dan **OK**.

Gambar 11.6

Tabel 11.1 Tabel Kontingensi Jenis Kelamin dan Hobi
Jenis Kelamin * Hobi Crosstabulation

			Hobi			Total
			Memasak	Olahraga	Membaca	
Jenis Kelamin	Laki-Laki	Count	12	24	4	40
		% within Jenis Kelamin	30.0%	60.0%	10.0%	100.0%
	Perempuan	Count	6	2	12	20
		% within Jenis Kelamin	30.0%	10.0%	60.0%	100.0%
Total		Count	18	26	16	60
		% within Jenis Kelamin	30.0%	43.3%	26.7%	100.0%

Berdasarkan Tabel 11.1, diketahui:

- ⇒ Dari 40 responden laki-laki, 12 (30%) di antaranya hobi memasak, 24 (60%) olahraga dan 4 (10%) membaca.
- ⇒ Dari 20 responden perempuan, 6 (30%) di antaranya hobi memasak, 2 (10%) olahraga dan 12 (60%) membaca.

Tabel 11.2 Hasil Uji Chi-Square Pearson dan Fisher's Exact
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	20.192 ^a	2	.000	.000		
Likelihood Ratio	21.371	2	.000	.000		
Fisher's Exact Test	20.211			.000		
Linear-by-Linear Association	5.796 ^b	1	.016	.018	.012	.008
N of Valid Cases	60					

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.33.

b. The standardized statistic is 2.407.

Berdasarkan hasil pada Tabel 11.2:

- ⇒ Pertama perhatikan pernyataan “**0 cells (.0%) have expected count less than 5**” yang menunjukkan tidak ada satu *cell* pun yang memiliki nilai frekuensi harapan < 5 .
- ⇒ Karena nilai frekuensi harapan untuk semua *cell* ≥ 5 , maka perhatikan nilai *P-Value* pada baris **Pearson Chi-Square** dan kolom **Asymp. Sig. (2-sided)**, yakni bernilai $0,000 < \text{tingkat signifikansi } 5\%$, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%. Diketahui juga nilai statistik chi-

square Pearson adalah $20,192 >$ nilai kritis chi-square yang telah dihitung sebelumnya, yakni $5,9915$, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5% .

⇒ Selain uji chi-square Pearson, dapat juga digunakan uji *Fisher's exact*. Diketahui nilai *P-Value* (baris *Fisher's Exact Test*, kolom *Exact. Sig. (2-sided)*) $0,000 <$ tingkat signifikansi $0,05$, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5% .

Bagian 12: Penyelesaian dengan Minitab

Contoh kasus pada Bagian 9 akan diselesaikan dengan Minitab. Input data pada Tabel 1.1 dalam Minitab seperti pada Gambar 12.1. Selanjutnya pilih **Stat** => **Tables** => **Cross Tabulation and Chi-Square** (Gambar 12.2), sehingga muncul kotak **Cross Tabulation and Chi-Square** (Gambar 12.3).

	C1-T	C2-T	C3
	Jenis Kelamin	Hobi	
1	Laki-Laki	Olahraga	
2	Laki-Laki	Olahraga	
3	Laki-Laki	Olahraga	
4	Laki-Laki	Memasak	
5	Laki-Laki	Olahraga	
6	Perempuan	Memasak	
7	Perempuan	Membaca	
8	Perempuan	Membaca	
9	Perempuan	Olahraga	
10	Perempuan	Membaca	
11	Laki-Laki	Memasak	
12	Laki-Laki	Olahraga	
13	Laki-Laki	Membaca	
14	Laki-Laki	Memasak	
15	Laki-Laki	Olahraga	
16	Laki-Laki	Olahraga	
17	Laki-Laki	Olahraga	
18	Laki-Laki	Olahraga	

Gambar 12.1 Input Data Kategori dalam Minitab

Gambar 12.2

Gambar 12.3

Pada Gambar 12.3, masukkan variabel jenis kelamin pada kotak **For rows**, masukkan variabel hobi pada kotak **For columns**. Pada bagian **Display**, pilih **Counts** dan **Row percents**. Selanjutnya pilih **Chi-Square**, sehingga muncul tampilan **Cross Tabulation – Chi-Square** (Gambar 12.4). Pada bagian **Display**, pilih **Chi-Square analysis**. Kemudian pilih **OK** dan **OK**. Hasilnya diperlihatkan pada Gambar 12.5.

Gambar 12.4

Rows: Jenis Kelamin		Columns: Hobi			
	Memasak	Membaca	Olahraga	All	
Laki-Laki	12	4	24	40	
	30.00	10.00	60.00	100.00	
Perempuan	6	12	2	20	
	30.00	60.00	10.00	100.00	
All	18	16	26	60	
	30.00	26.67	43.33	100.00	
Cell Contents:		Count			
		% of Row			
Pearson Chi-Square = 20.192, DF = 2, P-Value = 0.000					
Likelihood Ratio Chi-Square = 21.371, DF = 2, P-Value = 0.000					

Gambar 12.5

Gambar 12.5 merupakan hasil dari Minitab. Diketahui:

- ⇒ Dari 40 responden laki-laki, 12 (30%) di antaranya hobi memasak, 24 (60%) olahraga dan 4 (10%) membaca.
- ⇒ Dari 20 responden perempuan, 6 (30%) di antaranya hobi memasak, 2 (10%) olahraga dan 12 (60%) membaca.
- ⇒ Diketahui nilai statistik chi-square Pearson adalah 20,192.
- ⇒ Nilai *P-Value* 0,000 < tingkat signifikansi 0,05, maka terdapat hubungan yang signifikan antara jenis kelamin dan hobi pada tingkat signifikansi 5%.

Bagian 13: Apakah Uji Chi-Square Pearson Dapat Digunakan untuk 1×2 atau 2×1 atau 3×1 atau 1×3 atau 1×4 atau 4×1 ???

Misalkan diberikan data seperti pada Tabel 13.1.

Tabel 13.1

Jenis Kelamin	Hobi
Laki-Laki	Masak
Laki-Laki	Masak
Laki-Laki	Olahraga
Laki-Laki	Membaca
Laki-Laki	Membaca
Laki-Laki	Membaca

Data pada Tabel 13.1 disajikan dalam SPSS seperti pada Gambar 13.1.

	Jenis_Kelamin	Hobi
1	Laki-Laki	Memasak
2	Laki-Laki	Memasak
3	Laki-Laki	Olahraga
4	Laki-Laki	Membaca
5	Laki-Laki	Membaca
6	Laki-Laki	Membaca

Gambar 13.1

Berdasarkan Gambar 13.1, diketahui:

- ⇒ Data pada variabel jenis kelamin hanya terdiri dari 1 kategori, yakni laki-laki (tidak ada perempuan).
- ⇒ Data pada variabel hobi terdiri dari 3 kategori, yakni memasak, olahraga dan membaca.

Selanjutnya akan diuji apakah terdapat hubungan yang signifikan antara jenis kelamin dan hobi. Pengujian menggunakan metode statistika chi-square Pearson. Gambar 13.2 dan Gambar 13.3 merupakan langkah-langkah dalam SPSS untuk uji chi-square Pearson.

Gambar 13.2

Gambar 13.3

Warnings

No measures of association are computed for the crosstabulation of Jenis Kelamin * Hobi. At least one variable in each 2-way table upon which measures of association are computed is a constant.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Jenis Kelamin * Hobi	6	100.0%	0	.0%	6	100.0%

Jenis Kelamin * Hobi Crosstabulation

			Hobi			Total
			Memasak	Olahraga	Membaca	
Jenis Kelamin Laki-Laki	Count		2	1	3	6
	Expected Count		2.0	1.0	3.0	6.0
Total	Count		2	1	3	6
	Expected Count		2.0	1.0	3.0	6.0

Chi-Square Tests

	Value
Pearson Chi-Square	. ^a
N of Valid Cases	6

a. No statistics are computed because Jenis Kelamin is a constant.

Gambar 13.4

Gambar 13.4 merupakan hasilnya. Perhatikan bahwa **nilai statistik chi-square Pearson tidak muncul**. Mengapa demikian? Perhatikan pernyataan **“No statistics are computed because jenis kelamin is a constant”**. Hal ini karena data pada variabel jenis kelamin hanya memiliki 1 kategori, yakni laki-laki. Paling tidak, jumlah kategori untuk suatu variabel kategori minimal 2.

Seandainya data pada variabel jenis kelamin diperbaiki seperti pada Gambar 13.5, maka nilai statistik chi-square muncul, seperti pada Gambar 13.6. Berdasarkan data pada Gambar 13.5, jumlah kategori pada variabel jenis kelamin sebanyak 2, yakni laki-laki dan perempuan. Nilai statistik chi-square Pearson juga muncul, yakni 1,200.

	Jenis_Kelamin	Hobi
1	Laki-Laki	Memasak
2	Laki-Laki	Memasak
3	Laki-Laki	Olahraga
4	Perempuan	Membaca
5	Laki-Laki	Membaca
6	Laki-Laki	Membaca
7		

Gambar 13.5 Data Jenis Kelamin Memiliki 2 Kategori (Laki-Laki dan Perempuan)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Jenis Kelamin * Hobi	6	100.0%	0	.0%	6	100.0%

Jenis Kelamin * Hobi Crosstabulation

			Hobi			Total	
			Memasak	Olahraga	Membaca		
Jenis Kelamin	Laki-Laki	Count	2	1	2	5	
		Expected Count	1.7	.8	2.5	5.0	
	Perempuan	Count	0	0	1	1	
		Expected Count	.3	.2	.5	1.0	
Total		Count	2	1	3	6	
		Expected Count	2.0	1.0	3.0	6.0	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.200 ^a	2	.549
Likelihood Ratio	1.588	2	.452
Linear-by-Linear Association	.862	1	.353
N of Valid Cases	6		

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .17.

Gambar 13.6

Bagian 14: Contoh Kasus dengan STATCAL, SPSS dan Minitab

Misalkan diberikan data seperti pada Tabel 14.1.

Tabel 14.1 Contoh Data

Jenis Kelamin	Hobi
Laki-Laki	Olahraga
Laki-Laki	Olahraga
Laki-Laki	Olahraga
Laki-Laki	Olahraga
Laki-Laki	Membaca
Laki-Laki	Membaca
Perempuan	Memasak
Perempuan	Memasak
Perempuan	Memasak
Perempuan	Olahraga

Data pada Tabel 14.1 diinput dalam STATCAL, SPSS dan Minitab seperti pada Gambar 14.1, Gambar 14.2 dan Gambar 14.3.

Categorical Data

Name of Variable

Jenis Kelamin
Hobi

Label

	V1	V2	V3	
1	Laki-Laki	Olahraga		
2	Perempuan	Membaca		
3		Memasak		
4				

Set Number of Column

1 20

Set Number of Row

1 50

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	1	1							
2	1	1							
3	1	1							
4	1	1							
5	1	2							
6	1	2							
7	2	3							
8	2	3							
9	2	3							
10	2	1							
11									
12									
13									
14									

Gambar 14.1 Input Data Kategori dalam STATCAL

1 : Jenis_Kelamin 1.0

	Jenis_Kelamin	Hobi
1	Laki-Laki	Olahraga
2	Laki-Laki	Olahraga
3	Laki-Laki	Olahraga
4	Laki-Laki	Olahraga
5	Laki-Laki	Membaca
6	Laki-Laki	Membaca
7	Perempuan	Memasak
8	Perempuan	Memasak
9	Perempuan	Memasak
10	Perempuan	Olahraga

Gambar 14.2 Input Data Kategori dalam SPSS

Minitab - Untitled - [Worksheet 1 ***]

	C1-T	C2-T
	Jenis Kelamin	Hobi
1	Laki-Laki	Olahraga
2	Laki-Laki	Olahraga
3	Laki-Laki	Olahraga
4	Laki-Laki	Olahraga
5	Laki-Laki	Membaca
6	Laki-Laki	Membaca
7	Perempuan	Memasak
8	Perempuan	Memasak
9	Perempuan	Memasak
10	Perempuan	Olahraga

Gambar 14.3 Input Data Kategori dalam Minitab

Uji chi-square Pearson dan *Fisher's exact* akan digunakan untuk menguji apakah terdapat hubungan yang signifikan antara jenis kelamin dan hobi, pada tingkat signifikansi 5%. Gambar 14.4 merupakan hasil dari SPSS. Berdasarkan hasil SPSS pada Gambar 14.4 diketahui:

- ⇒ Dari 6 responden laki-laki, 4 responden di antaranya hobi olahraga dan 2 responden hobi membaca.
- ⇒ Dari 4 responden perempuan, 1 responden hobi olahraga dan 3 responden hobi memasak.

- ⇒ Berdasarkan hasil **Chi-Square Tests**, terdapat pernyataan “**6 cells (100.0%) have expected count less than 5**”. Maksudnya, terdapat 6 *cell* dengan nilai frekuensi harapan kurang dari 5.
- ⇒ Uji chi-square Pearson tetap dapat digunakan dengan melihat nilai pada kolom **Exact Sig. (2-sided)** (bukan kolom **Asymp. Sig (2-sided)**).
- ⇒ Diketahui nilai *P-Value* pada kolom **Exact Sig. (2-sided)** dan baris **Pearson Chi-Square** adalah $0,071 > \text{tingkat signifikansi } 0,05$, maka tidak terdapat hubungan yang signifikan antara jenis kelamin dan hobi.
- ⇒ Selain uji chi-square Pearson, dapat juga menggunakan uji *Fisher’s exact*. Diketahui nilai *P-Value* pada kolom **Exact Sig. (2-sided)** dan baris **Fisher’s Exact Test** adalah $0,071 > \text{tingkat signifikansi } 0,05$, maka tidak terdapat hubungan yang signifikan antara jenis kelamin dan hobi.

Jenis Kelamin * Hobi Crosstabulation

			Hobi			Total
			Olahraga	Membaca	Memasak	
Jenis Kelamin	Laki-Laki	Count	4	2	0	6
		Expected Count	3.0	1.2	1.8	6.0
		% within Jenis Kelamin	66.7%	33.3%	.0%	100.0%
	Perempuan	Count	1	0	3	4
		Expected Count	2.0	.8	1.2	4.0
		% within Jenis Kelamin	25.0%	.0%	75.0%	100.0%
	Total	Count	5	2	3	10
		Expected Count	5.0	2.0	3.0	10.0
		% within Jenis Kelamin	50.0%	20.0%	30.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	6.667 ^a	2	.036	.071		
Likelihood Ratio	8.456	2	.015	.071		
Fisher's Exact Test	5.555			.071		
Linear-by-Linear Association	3.868 ^b	1	.049	.071	.048	.038
N of Valid Cases	10					

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .80.

b. The standardized statistic is 1.967.

Gambar 14.4 Hasil berdasarkan SPSS

Gambar 14.5 merupakan hasil berdasarkan STATCAL. Dikarenakan terdapat *cell* dengan nilai frekuensi harapan < 5 , maka perhatikan hasil untuk uji *Fisher's exact*. Diketahui nilai *P-Value* adalah $0,0714 > \text{tingkat signifikansi } 0,05$, maka tidak terdapat hubungan yang signifikan antara jenis kelamin dan hobi.

Gambar 14.5 Hasil berdasarkan STATCAL

Tabulated statistics: Jenis Kelamin, Hobi

* NOTE * Fisher's exact test available only for 2 x 2 tables.

Rows: Jenis Kelamin Columns: Hobi

	Memasak	Membaca	Olahraga	All
Laki-Laki	0 1.800	2 1.200	4 3.000	6 6.000
Perempuan	3 1.200	0 0.800	1 2.000	4 4.000
All	3 3.000	2 2.000	5 5.000	10 10.000

Cell Contents: Count
Expected count

Pearson Chi-Square = 6.667, DF = 2
Likelihood Ratio Chi-Square = 8.456, DF = 2

* WARNING * 1 cells with expected counts less than 1
* WARNING * Chi-Square approximation probably invalid

* NOTE * 6 cells with expected counts less than 5

Gambar 14.6 Hasil berdasarkan Minitab

Gambar 14.6 merupakan hasil berdasarkan Minitab. Terdapat peringatan “NOTE 6 cells with expected counts less than 5”.

Bagian 15: Mengukur Keeratan Hubungan Variabel Kategori dengan Korelasi Cramer's V

Pada pembahasan sebelumnya, uji chi-square Pearson dan uji *Fisher's exact* digunakan untuk menguji signifikansi suatu hubungan variabel kategori. Namun tidak dapat mengukur seberapa erat hubungan atau korelasi kedua variabel kategori tersebut. Terdapat beberapa ukuran untuk mengukur keeratan hubungan antara dua variabel kategori, yakni korelasi *Cramer's V*.

Andy Field (2009:698) dalam bukunya yang berjudul "*Discovering Statistics Using SPSS, 3rd Edition*" menyatakan sebagai berikut.

"Phi: This statistic is accurate for 2×2 contingency tables. However, for tables with greater than two dimensions the value of phi may not lie between 0 and 1 because the chi-square value can exceed the sample size. Therefore, Pearson suggested the use of the **coefficient of contingency**.

Contingency Coefficient: This coefficient ensures a value between 0 and 1 but, unfortunately, it seldom reaches its upper limit of 1 and for this reason Cramer devised Cramer's V.

Cramer's V: When both variables have only two categories, phi and Cramer's V are identical. However, when variables have more than two categories Cramer's statistic can attain its maximum of one – unlike the other two – and so it is the most useful.

Gambar 15.1 diperlihatkan menu **Phi and Cramer's V** dalam SPSS. Sementara Gambar 15.2 diperlihatkan menu **Cramer's V-square statistic** dalam Minitab.

Gambar 15.1 Cramer's V dalam SPSS

Gambar 15.2 Cramer's V dalam Minitab

Gambar 15.3 disajikan nilai *Cramer's V* berdasarkan SPSS untuk contoh kasus pada Bagian 14.

Symmetric Measures				
		Value	Approx. Sig.	Exact Sig.
Nominal by Nominal	Phi	.816	.036	.071
	Cramer's V	.816	.036	.071
N of Valid Cases		10		

Gambar 15.3 Cramer's V dalam SPSS

Diketahui keeratan hubungan antara jenis kelamin dan hobi berdasarkan ukuran *Cramer's V* adalah 0,816. Diketahui nilai korelasi $0,816 > 0,5$, maka jenis kelamin dan hobi berkorelasi kuat.

Andy Field (2009:170) dalam bukunya yang berjudul "*Discovering Statistics Using SPSS, 3rd Edition*" menyatakan sebagai berikut.

"We also saw in section 2.6.4 that because the correlation coefficient is a standardized measure of an observed effect, it is a commonly used measure of the size of an effect and that values of ± 1 represent a small effect, ± 3 is a medium effect and ± 5 is a large effect (although I re-emphasize my caveat that these canned effect sizes are no substitute for interpreting the effect size within the context of the research literature)."

Berdasarkan uraian di atas:

⇒ Nilai korelasi $\pm 0,1$ menyatakan *small effect* (pengaruh lemah).

⇒ Nilai korelasi $\pm 0,3$ menyatakan *medium effect* (pengaruh sedang).

⇒ Nilai korelasi $\pm 0,5$ menyatakan *large effect* (pengaruh besar).

Gambar 15.4 disajikan nilai *Cramer's V-square* berdasarkan Minitab untuk contoh kasus pada Bagian 14. Diketahui nilai *Cramer's V-square* adalah 0,66667, maka nilai *Cramer's V* adalah $\sqrt{0,66667} = 0,816$. Gambar 15.5 merupakan nilai *Cramer's V* berdasarkan STATCAL.

Tabulated statistics: Jenis Kelamin, Hobi

Rows: Jenis Kelamin Columns: Hobi

	Memasak	Membaca	Olahraga	All
Laki-Laki	0 1.800	2 1.200	4 3.000	6 6.000
Perempuan	3 1.200	0 0.800	1 2.000	4 4.000
All	3 3.000	2 2.000	5 5.000	10 10.000

Cell Contents: Count
Expected count

Pearson Chi-Square = 6.667, DF = 2
Likelihood Ratio Chi-Square = 8.456, DF = 2

* WARNING * 1 cells with expected counts less than 1
* WARNING * Chi-Square approximation probably invalid

* NOTE * 6 cells with expected counts less than 5

Cramer's V-square 0.66667

Gambar 15.4 Cramer's V dalam Minitab

Gambar 15.5 Cramer's V dalam STATCAL

Referensi Bacaan

- [1] John Maindonald dan W. John Braun, 2010, *Data Analysis and Graphics Using R, An Example-Based Approach 3rd Edition*, Cambridge University Press.
- [2] Gareth James, Daniela Witten, Trevor Hastie dan Robert Tibshirani, 2014, *An Introduction to Statistical Learning with Applications in R*, Springer.
- [3] Peter Dalgaard, 2008, *Introductory Statistics with R, 2nd Edition*, Springer.
- [4] Michael J. Crawley, 2015, *Statistics, An Introduction Using R, 2nd Edition*, John Wiley and Sons, Ltd.
- [5] Sanders & Smidth, 2000, *Statistics, A First Course, 6th Edition*, McGraw-Hill.
- [6] Douglas C. Montgomery dan George C. Runger, 2014, *Applied Statistics and Probability for Engineers, 6th Edition*, John Wiley & Sons.
- [7] Alan Agresti dan Barbara Finlay, 2009, *Statistical Methods for the Social Sciences, 4th Edition*, Prentice Hall.
- [8] Andy Field, 2009, *Discovering Statistics Using SPSS, 3rd Edition*, Sage.
- [9] Damodar N. Gujarati, 2003, *Basic Econometrics, 4th Edition*, McGraw-Hill.
- [10] Prem S. Mann, 2013, *Introductory Statistics, 8th Edition*, John Wiley and Sons.
- [11] Murray R. Spiegel dan Larry J. Stephens, 2008, *Statistics 4th Edition*, McGraw-Hill Companies.
- [12] Jim Albert, 2009, *Bayesian Computation with R, 2nd Edition*, Springer.
- [13] Nick T. Thomopoulos, 2013, *Essentials of Monte Carlo Simulation, Statistical Methods for Building Simulation Models*, Springer.
- [14] Sugiyono, 2015, *Metode Penelitian Manajemen*, Penerbit ALFABETA.
- [15] Yvonne Augustine dan Robert Kristaung, 2013, *Metodologi Penelitian Bisnis dan Akuntansi*, Dian Rakyat.
- [16] W. J. Conover, 1999, *Practical Nonparametric Statistics 3rd Edition*, John Wiley and Sons.
- [17] Paul H. Kvam dan Brani Vidakovic, 2007, *Nonparametric Statistics with Applications to Science and Engineering*, John Wiley and Sons.