Modul Kuliah ke 12
Statistik

Membandingkan Multi Proporsi dan Tes Independen Serta Goodness of Fit Distribusi Normal

Kompetensi:

Setelah membaca modul kuliah ini, diharapkan mahasiswa mampu:

a. Memahami cara membandingkan multi proporsi, Tes Independen dan good fit
b. Mencari penyelesaian pada perbandingan multi proporsi, Tes Independen dan good fit
c. Menarik kesimpulan pada hipotesis multi proporsi, Tes Independen dan good fit

I. Menguji Kesetaraan Proporsi diatas Tiga Populasi

Tes statistik digunakan dalam melakukan uji hipotesis yang didasarkan pada distribusi chi-square. Tes ini bersifat kategorikal dan dapat digunakan untuk pengujian hipotesis dalam :

a. Menguji kesetaraan proporsi populasi pada tiga populasi atau lebih
b. Menguji independensi dua kategori dari variabel
c. Menguji apakah distribusi probabilitas untuk suatu populasi mengikuti distribusi probabilitas.

Rumus untuk mencari frekuensi expected dengan asumsi H_0 adalah benar seperti berikut:

$$e_{ij} = \frac{(Row\ i\ Total)(Column\ j\ total)}{Total\ Sample\ Size}$$

Dimana:

e_{ij} = Harapan (expected) yang inginkan

Tes statistik Chi-square adalah sebagai berikut:

$$X^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Dimana:

f_{ij} = frekuensi observasi untuk sell pada bari i dan kolom j

e_{ij} = expected frekuensi untuk sell pada bari i dan kolom j

(k – 1) adalah derajat ketinggian (df)

Langkah tes chi-square untuk kesetaraan proporsi populasi untuk kategori $k \geq 3$ populasi adalah sebagai berikut:

1. Buatlah hipotesis nol dan hipotesis alternatif sebagai berikut:

   H_0: $p_1 = p_2 = \ldots = p_k$

   H_a: Tidak semua proporsi populasi adalah sama

2. Pilihlah sample random dari masing-masing populasi yang disajikan dalam tabel 2 baris dan k kolom sebagai f_{ij}.
3. Asumsi H_0 adalah benar, maka hitunglah frekuensi expected e_{ij} dari data frekuensi.
4. Hitunglah tes statistiknya:

$$X^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

5. Menguji hipotesis aturan ditolak
   - Pendekatan p-value : Tolak H_0 jika p-value $\leq \alpha$
   - Pendekatan kritikal : Tolak H0 jika $X^2 \geq X^2_\alpha$

   Dimana:

   Data berdistribusi chi-square dengan df = k – 1 dan α adalah level signifikasinya.

Contoh 1

Dalam suatu penelitian tertentu ingin membandingkan loyalitas pelanggan untuk tiga merk mobil untuk membeli kembali merk tersebut yaitu : Chevrolet impala, ford fusion, dan honda accord.

Pemilik mobil dari masing-masing tiga mer mobil membentuk tiga populasi untuk dipelajari dan kondisi data proporsi dari tiga populasi yang menarik adalah sebagai berikut:

Membeli kembali	**Chevrolet impala**	**ford fusion**	**honda accord**	**Total**
Yes	69	120	123	312
No	56	80	52	188
Total	125	200	175	500

Dari kasus tersebut ujilah kesetaraan proporsi populasi pada ketiga merk tersebut dengan level signifikasi $\alpha = 0{,}05$.

Jawaban

a. Hipotesis pada kasus di atas adalah sebagai berikut:
   H_0: $p_1 = p_2 = p_3$
   H_a: Tidak semua proporsi populasi adalah sama

b. Data untuk f_{ij} diambil pada data history pada soal seperti berikut:
f_{ij} =

Membeli kembali	Chevrolet impala	ford fusion	honda accord	Total
Yes	69 (f_{11})	120 (f_{12})	123 (f_{13})	312
No	56 (f_{21})	80 (f_{22})	52 (f_{23})	188
Total	125	200	175	500

c. Untuk mencari e_{ij} dapat dihitung seperti berikut:

$$e_{11} = \frac{312}{500}(125) = 78$$
$$e_{12} = \frac{312}{500}(200) = 124{,}8$$
$$e_{13} = \frac{312}{500}(175) = 109{,}2$$
$$e_{21} = \frac{188}{500}(125) = 47$$
$$e_{22} = \frac{188}{500}(200) = 75{,}2$$
$$e_{23} = \frac{188}{500}(175) = 65{,}8$$

Tabel e_{ij} dapat dilihat seperti berikut:

Membeli kembali	Chevrolet impala	ford fusion	honda accord	Total
Yes	78 (e_{11})	124,8 (e_{12})	109,2 (e_{13})	312
No	47 (e_{21})	75,2 (e_{22})	65,8 (e_{23})	188
Total	125	200	175	500

d. Hasil menghitung tes statistik rumus di atas adalah sebagai berikut:

$$X^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Membeli kembali	Merk	Frekuensi Observasi (f_{ij})	Frekuensi Expected (e_{ij})	Difference ($F_{ij} - e_{ij}$)	Kuadrat Difference $(F_{ij} - e_{ij})^2$	$(F_{ij} - e_{ij})^2/e_{ij}$
Yes	Impala	69	78	-9	81	1,04
Yes	Fusion	120	124,8	-4,8	23,04	0,18
Yes	Accord	123	109,2	13,8	190,44	1,74
No	Impala	56	47	9	81	1,72
No	Fusion	80	75,2	4,8	23,04	0,31
No	Accord	52	65,8	-13,8	190,44	2,89
	Total	500	500			$X^2 = 7{,}89$

e. Menguji hipotesis dengan pendekatan p-value dan kritikal.

1) Pendekatan p-value :
Hasil X^2 = 7,89 dengan df = k-1 = 3 -1 = 2
Dengan menggunakan tabel chi-square, maka :
Nilai p-value = antara 0,025 sampai 0,01
Hasil pada tabel chi-square dapat dilihat pada tabel di bawah ini:

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**	**.005**
χ^2 Value (2 *df*)	4.605	5.991	7.378	9.210	10.597

$\chi^2 = 7.89$

Uji p-value dengan α = 0,05

Tolak H_0 jika p-value ≤ α

Apakah p-value ≤ α

Apakah 0,025 --- 0,01 ≤ 0,05

Pernyataan tersebut merupakan pernyataan yang benar, maka kesimpulan uji p-value adalah H_0 ditolak.

2) Pendekatan kritikal :

α = 0,05, dengan df = 2, maka nilai dari X_{α}^{2} dapat dilihat pada tabel chi-square berikut:

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**	**.005**
χ^2 Value (2 *df*)	4.605	$\boxed{5.991}$	7.378	9.210	10.597

Dari tabel di atas dapat kita peroleh bahwa nilai X^2 = 5,991
Uji kritikalnya sebagai berikut:
Tolak H_0 jika $X^2 \geq X_{\alpha}^{2}$
Apakah $X^2 \geq X_{\alpha}^{2}$
Apakah 7,89 ≥ 5,991

Pernyataan tersebut merupakan pernyataan yang benar, maka kesimpulan uji kritikal adalah H_0 ditolak.

Kesimpulan dari uji p-value dan uji ktritikal menghasilkan kesimpulan yang sama dan konsisten dengan hasilnya adalah H_0 ditolak. Artinya Tidak semua proporsi populasi memiliki kesetaraan/ kesamaan proporsinya dan ada proporsi dari populasi yang berbeda.

II. Tes Independen

Penggunaan yang penting dari uji chi-square melibatkan penggunaan data sampel untuk menguji independensi dua variabel kategori. untuk tes ini kita mengambil satu sampel dari suatu populasi

yang dicatat pengamatan untuk dua variabel kategori. Kita hitung jumlah respons untuk setiap kombinasi kategori pada variabel 1 dan kategori untuk variabel 2. Hipotesis nol untuk tes ini adalah apakah dua variabel kategori independen. Dengan demikian, tes ini disebut sebagai tes independensi.

Langkah tes chi-square untuk menguji indenpendensi kategori variabel adalah sebagai berikut:

1. Buatlah hipotesis nol dan hipotesis alternatif sebagai berikut:

   H_0: Dua kategori variabel adalah independen

   H_a: Dua kategori variabel adalah tidak independen

2. Pilihlah sample random dari masing-masing populasi yang disajikan dalam tabel 2 baris dan kolom sebagai f_{ij}.
3. Asumsi H_0 adalah benar, maka hitunglah frekuensi expected e_{ij} dari data frekuensi.
4. Hitunglah tes statistiknya:

$$X^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

5. Menguji hipotesis aturan ditolak
   - Pendekatan p-value : Tolak H_0 jika p-value $\leq \alpha$
   - Pendekatan kritikal : Tolak H0 jika $X^2 \geq X_\alpha^2$

   Dimana:

Data berdistribusi chi-square dengan df = ((baris – 1) (kolom-1)) dan α merupakan level dari signifikasinya.

Contoh 2

Sebuah industri minuman akan melakukan penelitian preperensi tiga yaitu ringan, sedang dan berat. Data yang diambil sebanyak 200 responden untuk menilai tiga preperensi tersebut dengan kategori Pria dan wanita. Pihak perusahaan ingin mengetahui independensi kategori dari pria terhadap wanita dan sebaliknya. Hipotesis pada awal penelitian adalah :

H_0: Preperensi minuman independen pada gender

H_a: Preperensi minuman tidak independen pada gender

Data history yang didapat hasil observasi yang telah dilakukan kepada sejumlah pelanggan minuman adalah sebagai berikut:

Preperensi	Male	Female	Total
Ringan	51	39	90
Sedang	56	21	77
Berat	25	8	33
	132	68	200

Dari kasus tersebut ujilah independensi kategori gender dengan level signifikasi α = 0,05.

Jawaban

a. Hipotesis pada kasus di atas adalah sebagai berikut:

H_0: Preperensi minuman independen pada gender
Ha: Preperensi minuman tidak independen pada gender

b. Data untuk fij diambil pada data history pada soal seperti berikut:
f_{ij} =

Preperensi	ford fusion	honda accord	Total
Ringan	51 (f_{11})	39 (f_{12})	90
Sedang	56 (f_{21})	21 (f_{22})	77
Berat	25 (f_{31})	8 (f_{32})	33
Total	132	68	200

f. Untuk mencari e_{ij} dapat dihitung seperti berikut:

$$e_{11} = \frac{90}{200}(132) = 59{,}4$$
$$e_{12} = \frac{90}{200}(68) = 30{,}6$$
$$e_{21} = \frac{77}{200}(132) = 50{,}82$$
$$e_{22} = \frac{77}{200}(68) = 26{,}18$$
$$e_{31} = \frac{33}{200}(132) = 21{,}78$$
$$e_{32} = \frac{33}{200}(68) = 11{,}22$$

Tabel e_{ij} dapat dilihat seperti berikut:

Preperensi	ford fusion	honda accord	Total
Ringan	59,4 (e_{11})	30,6 (e_{12})	90
Sedang	50,82 (e_{21})	26,18 (e_{22})	77
Berat	21,78 (e_{31})	11,22 (e_{32})	33
Total	132	68	200

g. Hasil menghitung tes statistik rumus di atas adalah sebagai berikut:

$$X^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Merk	**Frekuensi Observasi (f_{ij})**	**Frekuensi Expected (e_{ij})**	**Difference ($f_{ij} - e_{ij}$)**	**Kuadrat Difference ($f_{ij} - e_{ij})^2$**	**$(f_{ij} - e_{ij})^2/e_{ij}$**
Ringan	51	59,4	-8,4	70,56	1,19
Ringan	39	30,6	8,4	70,56	2,31
Sedang	56	50,82	5,18	26,83	0,53
Sedang	21	26,18	-5,18	26,83	1,02
Berat	25	21,78	3,22	10,37	0,48
Berat	8	11,22	-3,22	10,37	0,92
Total	200	200			X^2 = 6,45

h. Menguji hipotesis dengan pendekatan p-value dan kritikal.

1) Pendekatan p-value :
Hasil X^2 = 6,45 dengan df = (b – 1)(k – 1) = (3 -1)(2-1) = (2)(1)= 2
Dengan menggunakan tabel chi-square, maka :
Nilai p-value = antara 0,05 sampai 0,025
Hasil pada tabel chi-square dapat dilihat pada tabel di bawah ini:

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**	**.005**
χ^2 Value (2 *df*)	4.605	5.991	7.378	9.210	10.597

$\chi^2 = 6.45$

Uji p-value dengan α = 0,05

Tolak H_0 jika p-value ≤ α

Apakah p-value ≤ α

Apakah 0,05 --- 0,025 ≤ 0,05

Pernyataan tersebut merupakan pernyataan yang benar, maka kesimpulan uji p-value adalah H_0 ditolak.

2) Pendekatan kritikal :

α = 0,05, dengan df = 2, maka nilai dari X^2_α dapat dilihat pada tabel chi-square berikut:

Area in Upper Tail	.10	.05	.025	.01	.005
χ^2 Value (2 *df*)	4.605	$\boxed{5.991}$	7.378	9.210	10.597

Dari tabel di atas dapat kita peroleh bahwa nilai X_α^2 = 5,991
Uji kritikalnya sebagai berikut:
Tolak H_0 jika $X^2 \geq X_\alpha^2$
Apakah $X^2 \geq X_\alpha^2$
Apakah $6{,}45 \geq 5{,}991$

Pernyataan tersebut merupakan pernyataan yang benar, maka kesimpulan uji kritikal adalah H_0 ditolak.

Kesimpulan dari uji p-value dan uji ktritikal menghasilkan kesimpulan yang sama dan konsisten dengan hasilnya adalah H_0 ditolak. Artinya Uji kategori preperensi terhadap gender tidak independen.

III. Tes Goodness of Fit Probabilitas Distribusi Normal

Goodness of fit test untuk distribusi probabilitas normal didasarkan pada penggunaan distribusi chi-square. frekuensi yang diamati untuk beberapa kategori sampel data dibandingkan dengan frekuensi yang diharapkan dengan syarat populasi memiliki distribusi probabilitas normal.

Langkah pengujian Goodness of Fit distribusi probabilitas normar adalah sebagi berikut:

1. Buatlah hipotesis nol dan hipotesis alternatif sebagai berikut:

   H_0: Populasi memiliki distribusi probabilitas normal
   H_a: Populasi tidak memiliki distribusi probabilitas normal

2. Pilihlah sample random dari masing-masing populasi
   a. Hitunglah sample mean dan standar deviasi sample
   b. Tentukan k interval dari nilai yang menjadi frekuensi minimal 5 interval
   c. Catat frekuensi observasi disajikan dalam tabel sebagai f_i.
3. Hitunglah frekuensi expected ei dari data frekuensi pada masing-masing interval. Kalikan sample size dengan probabilitas dari variabel random normal pada interval.
4. Hitunglah tes statistiknya:

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

d. Menguji hipotesis aturan ditolak
   - Pendekatan p-value : Tolak H_0 jika p-value $\leq \alpha$
   - Pendekatan kritikal : Tolak H_0 jika $X^2 \geq X_\alpha^2$

   Dimana:

Data berdistribusi chi-square dengan df = k - p – 1, p = jumlah parameter dan α signifikasinya.

Contoh 3

Sebuah perusahaan menerima pelamar baru untuk menjadi karyawan diperusahaannya. Untuk mengetahui populasi dari nilai tes aptitude pelamar yang masuk apakah berdistribusi normal, maka dilakukan pengumpulan data nilai tes aptitude pada 50 secara random seperti berikut ini;

71	66	61	65	54	93	60	86	70	70
73	73	55	63	56	62	76	54	82	79
76	68	53	58	85	80	56	61	61	64
65	62	90	69	76	79	77	54	64	74
65	65	61	56	63	80	56	71	79	84

Dari data tersebut diperoleh mean sample dan standar deviasi sample seperti berikut:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3421}{50} = 68{,}42$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{5310{,}0369}{49}} = 10{,}41$$

Hipotesis pada kasus tersebut adalah:

H_0: Populasi dari nilai tes memiliki distribusi probabilitas normal dengan mean 68,42 dan standar deviasi 10,41.

H_a: Populasi dari nilai tes tidak memiliki distribusi probabilitas normal dengan mean 68,42 dan standar deviasi 10,41.

Hitunglah tes statistiknya dan apa kesimpulan uji hipotesisnya dengan $\alpha = 0{,}1$!

Jawaban

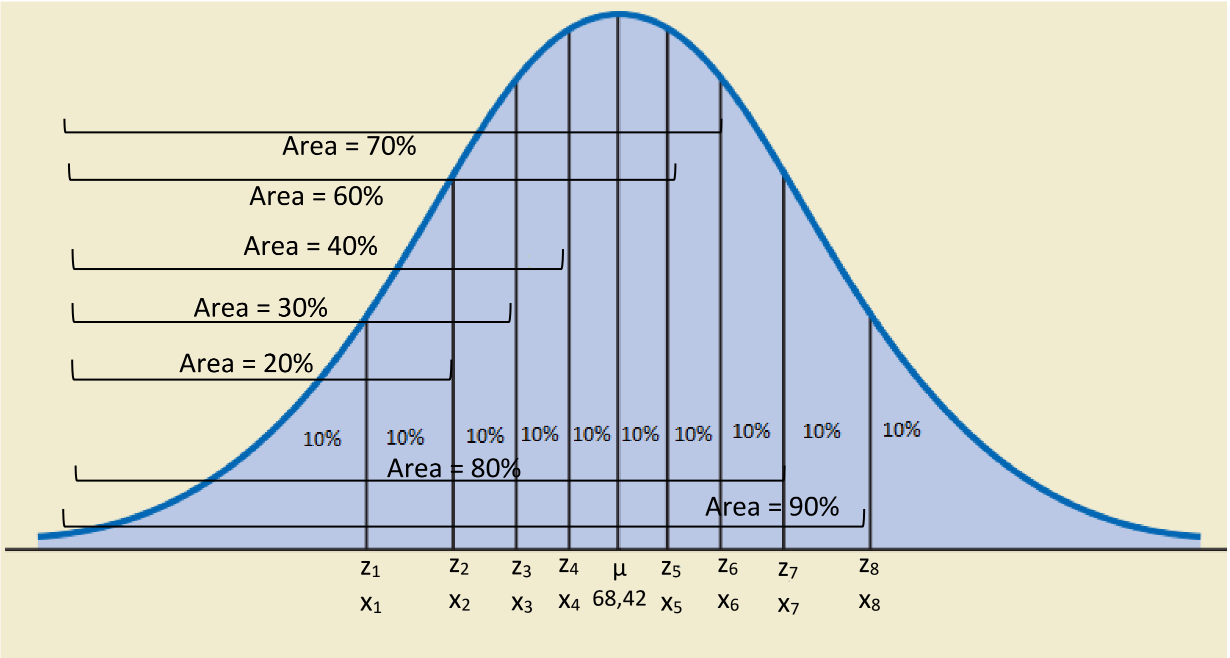
Menggunakan langkah analisis pengujian hipotesis, kita akan menjawab dari kasus di atas seperti berikut ini:

1. Hipotesis nol dan hipotesis alternatif sebagai berikut

   H_0: Populasi dari nilai tes memiliki distribusi probabilitas normal dengan mean 68,42 dan standar deviasi 10,41.

   H_a: Populasi dari nilai tes tidak memiliki distribusi probabilitas normal dengan mean 68,42 dan standar deviasi 10,41.
2. Dari sample yang ada kita memperoleh:
   a. Mean sample = 68,42 dan standar deviasi = 10,41
   b. Jumlah data yang diperoleh sebanyak 50 secara distribusi normal, maka kita dapat membagi gambar area distribusi normal menjadi 10 bagian dengan masing-maisng luas area adalah 1/10 = 0,1. Dari masing-masing area pada gambar distribusi normal, kita dapat

mencari nilai z score pada batas masing-masing bagian. Kemudia dari nilai z score ini tiap batas bagian pada gambar distribusi normal tersebut kita dapat mencari nilai dari x_i pada masing-masing batas seperti berikut:



Gambar 4. Pembagian area distribusi normal

Dari garbar di atas kita dapat mencari Z score dengan menggunakan tabel distribusi normal dan menggunakan rumus z score kita mendapat nilai dari x_i seperti berikut:

$$z = \frac{xi - \mu}{\sigma} \rightarrow$$

$$x_i = z\sigma + \mu$$

Tabel 3. Hasil perhitungan x_i

Luas area	Desimal	Zscore	Xi
10%	0,10	z_1 = -1,285	x_1 = -1,285 (10,41)+ 68,42 = 55,10
20%	0,20	z_2 = -0,845	x_2 = -0,845 (10,41)+ 68,42 = 59,68
30%	0,30	z_3 = -0,525	x_3 = -0,525 (10,41)+ 68,42 = 63,01
40%	0,40	z_4 = -0,255	x_4 = -0,255 (10,41)+ 68,42 = 65,82
50%	0,50	z_5 = 0,0	x_5 = 0,00 (10,41)+ 68,42 = 68,42
60%	0,60	z_6 = 0,255	x_6 = 0,255 (10,41)+ 68,42 = 71,02
70%	0,70	z_7 = 0,525	x_7 = 0,525 (10,41)+ 68,42 = 73,83
80%	0,80	z_8 = 0,845	x_8 = 0,845 (10,41)+ 68,42 = 77,16
90%	0,90	z_9 = 1,285	x_9 = 1,285 (10,41)+ 68,42 = 81,74

c. Masukkan data dari sample ke dalam interval masing-masing sesuai dengan nilai batasnya.

Interval nilai tes	**Observed Frekuensi (f_i)**	**Expected Frekuensi (e_i)**
Dibawah 55,10	5	5
55,10 – 59,68	5	5
59,68 – 63,01	9	5
63,01 – 65,82	6	5
65,82 – 68,42	2	5
68,42 – 71,02	5	5
71,02 – 73,83	2	5
73,83 – 77,16	5	5
77,16 – 81,74	5	5
Diatas 81,74	6	5
Total	50	50

d. Hasil tes statistik rumus di atas adalah sebagai berikut:

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

Merk	**Frekuensi Observasi (f_i)**	**Frekuensi Expected (e_i)**	**Difference ($f_i - e_i$)**	**Kuadrat Difference ($f_i - e_i$)^2**	**($f_i - e_i$)^2/e_i**
Dibawah 55,10	5	5	0	0	0
55,10 – 59,68	5	5	0	0	0
59,68 – 63,01	9	5	4	16	3,2
63,01 – 65,82	6	5	1	1	0,2
65,82 – 68,42	2	5	-3	9	1,8
68,42 – 71,02	5	5	0	0	0
71,02 – 73,83	2	5	-3	9	1,8
73,83 – 77,16	5	5	0	0	0
77,16 – 81,74	5	5	0	0	0
Diatas 81,74	6	5	1	1	0,2
Total	50	50			$X^2 = 7,2$

a. Menguji hipotesis dengan pendekatan p-value dan kritikal.

3) Pendekatan p-value :
   Hasil $X^2 = 7,2$ dengan df = (k - p - 1) = (10 - 2 - 1) = 7
   Dengan menggunakan tabel chi-square, maka :
   Nilai p-value = antara 0,9 sampai 0,1
   Hasil pada tabel chi-square dapat dilihat pada tabel di bawah ini:

Degrees of Freedom	Area in Upper Tail						
	.995	.99	.975	.95	.90	.10	.05
1	.000	.000	.001	.004	.016	2.706	3.841
2	.010	.020	.051	.103	.211	4.605	5.991
3	.072	.115	.216	.352	.584	6.251	7.815
4	.207	.297	.484	.711	1.064	7.779	9.488
5	.412	.554	.831	1.145	1.610	9.236	11.070
6	.676	.872	1.237	1.635	2.204	10.645	12.592
7	.989	1.239	1.690	2.167	2.833	12.017	14.067

Uji p-value dengan α = 0,10

Tolak H_0 jika p-value ≤ α

Apakah p-value ≤ α

Apakah 0,90 --- 0,10 ≤ 0,10

Pernyataan tersebut merupakan pernyataan yang salah karena kedua nilai p-value tidak lebih kecil dari α, maka kesimpulan uji p-value adalah H_0 diterima.

4) Pendekatan kritikal :

α = 0,10, dengan df = 7, maka nilai dari X_α^2 dapat dilihat pada tabel chi-square berikut:

Degrees of Freedom	Area in Upper Tail						
	.995	.99	.975	.95	.90	.10	.05
1	.000	.000	.001	.004	.016	2.706	3.841
2	.010	.020	.051	.103	.211	4.605	5.991
3	.072	.115	.216	.352	.584	6.251	7.815
4	.207	.297	.484	.711	1.064	7.779	9.488
5	.412	.554	.831	1.145	1.610	9.236	11.070
6	.676	.872	1.237	1.635	2.204	10.645	12.592
7	.989	1.239	1.690	2.167	2.833	$\boxed{12.017}$	14.067

Dari tabel di atas dapat kita peroleh bahwa nilai X_α^2 = 12,017
Uji kritikalnya sebagai berikut:
Tolak H_0 jika $X^2 \geq X_\alpha^2$
Apakah $X^2 \geq X_\alpha^2$
Apakah 7,2 ≥ 12,017

Pernyataan tersebut merupakan pernyataan yang salah, maka kesimpulan uji kritikal adalah H_0 diterima.

Kesimpulan dari uji p-value dan uji ktritikal menghasilkan kesimpulan yang sama dan konsisten dengan hasilnya adalah H_0 diterima. Artinya Uji goodness of Fit tidak terdistribusi secara normal.

Daftar Pustaka

Statistics for Business and Economics (13e)" Anderson, Sweeney, Williams, Camm, Cochran 2017, Cengage Learning