

Modul

CHI-SQUARE: GOODNESS OF FIT TEST

Pengantar

Chi Square adalah salah satu alat analisis yang paling sering digunakan pada statistik, dengan tujuan untuk Uji Homogenitas, Uji Independensi dan Uji Goodness of Fit Test. Pada modul ini akan dibahas kegunaan Chi Square untuk Uji Goodness of Fit Test, sedangkan Uji Independensi dibahas pada buku BUKU LATIHAN SPSS: PARAMETRIK.

Uji Goodness of Fit Test pada prinsipnya bertujuan untuk mengetahui apakah sebuah distribusi data dari sampel mengikuti sebuah distribusi teoritis tertentu atau tidak. Sebagai contoh, jika sebuah dadu dilempar, maka kemungkinan mendapat angka '5' adalah $1/6$, juga kemungkinan untuk angka yang lain. Inilah yang disebut distribusi teoritis sebuah dadu, karena terdiri dari 6 mata dadu yang mempunyai kemungkinan seimbang untuk muncul dalam sekali pelemparan. Seandainya dilakukan pelemparan 120 kali, seharusnya tiap mata dadu *secara teoritis* akan muncul masing-masing $1/6 \times 120 = 20$ kali (angka 1 muncul 20 kali, angka 2 muncul 20 kali dan seterusnya). Namun tentu kenyataan tidaklah persis sama, bisa saja angka 1 muncul hanya 10 kali, tapi angka 3 muncul 24 kali dan kemungkinan lain. Untuk mengetahui apakah kenyataan tersebut *masih bisa dianggap selaras (fit)* dengan distribusi teoritis, akan digunakan uji Goodness of Fit.

Dengan demikian, Goodness of Fit Test akan membandingkan dua distribusi data, yakni yang teoritis (frekuensi harapan) dan yang sesuai kenyataan (frekuensi observasi). Uji ini hampir sama dengan uji Binomial, hanya jika pada binomial hanya ada *dua kemungkinan jawaban*, pada uji Goodness of Fit ada *lebih dari dua kemungkinan*.

Dua Kasus berikut menjelaskan penggunaan Chi Square pada Goodness of Fit Test

KASUS 1:

Selama ini Manajer Pemasaran sabun mandi HARUM menganggap bahwa konsumen *sama-sama* menyukai tiga warna sabun mandi yang diproduksi, yaitu putih, hijau dan kuning. Untuk mengetahui apakah pendapat Manajer tersebut benar, kepada dua belas orang responden ditanya warna sabun mandi yang paling disukainya.

Berikut adalah data kuesioner tersebut:

(lihat file **CHI-SQUARE**):

Responden	Warna Pilihan
RENNY	putih
RONNY	hijau
SUGIARTO	kuning
DEDDY	kuning
SINTHA	hijau
SUSSY	kuning
LILY	kuning
LINNA	kuning
LANNY	putih
RUBEN	putih
RICKY	hijau
ROSSY	hijau

Catatan data:

- Data untuk variabel NAMA (Nama Responden) adalah berjenis STRING dan bukan Numerik, karena nama seseorang mengandung karakter (huruf). Dengan demikian dalam proses perhitungan SPSS, variabel ini tidak akan ditampilkan.
- Data untuk variabel WARNA adalah dalam bentuk kategori, yang dalam input SPSS digunakan kode 1, 2 dan 3, dengan mengisi bagian VALUES pada VARIABLE VIEW.

Adapun kode dan warna:

KODE	WARNA
1	Putih
2	Hijau
3	Kuning

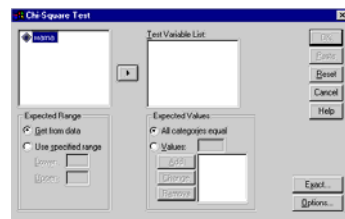
- Arti data diatas: konsumen bernama Renny senang memakai sabun warna putih, sedangkan Ronny mempunyai favorit warna sabun hijau. Demikian seterusnya untuk data yang lain.

Langkah:

Karena delapan data diatas berjenis nominal (kategorisasi, dengan kode 1 untuk putih dan sebagainya), maka digunakan uji non parametrik.

- Buka file **chi square**
- Dari menu **Analyze**, pilih sub menu **Nonparametric Test**, lalu pilihan **chi square...**

Tampak di layar:



Pengisian:

- TEST VARIABLE LIST atau nama variabel yang akan diuji. Sesuai kasus, masukkan variabel **warna**.
- EXPECTED RANGE. Disini karena data sudah diinput, maka pilihan tetap pada GET FROM DATA.
- EXPECTED VALUES. Jika dilihat pada kasus, Manajer menganggap kesenangan terhadap warna adalah sama, yang berarti semua warna *seharusnya* dipilih secara merata (sepertiga konsumen

memilih warna putih, sepertiga memilih warna hijau dan sepertiga memilih warna kuning). Dengan demikian, tetap pada pilihan ALL CATEGORIES EQUAL.

Abaikan bagian lain dan tekan OK.

Output:

(lihat file **CHI SQUARE**)

NPar Tests

Chi-Square Test

Frequencies

WARNA

	Observed N	Expected N	Residual
putih	3	4,0	-1,0
hijau	4	4,0	,0
kuning	5	4,0	1,0
Total	12		

Test Statistics	
	WARNA
Chi-Square ^a	,500
df	2
Asymp. Sig.	,779

^a. 3 cells (100,0%) have expected frequencies less than 5. The minimum expected cell frequency is 4,0.

Analisis:

1. Dari tabel WARNA terlihat bahwa putih dipilih oleh 3 responden, warna hijau 4 responden dan warna kuning 5 responden. Ini adalah frekuensi yang didapat dari input data (observed).
2. Dari tabel WARNA untuk kolom EXPECTED, karena distribusi seharusnya merata, maka masing-masing warna *seharusnya* diminati oleh masing-masing 4 responden (dari 12 responden dibagi 3 warna).
3. kolom RESIDUAL dari tabel WARNA adalah selisih antara kolom Observed dan kolom Expected. Seperti untuk warna putih, angka residual adalah 3 – 4 atau -1. Demikian untuk dua angka residual yang lain.

4. Proses pengambilan keputusan:

a. HIPOTESIS:

Ho: $p_1=p_2=p_3=1/3$

Hi: $p_1 \neq p_2 \neq p_3 \neq 1/3$

NB: Ho menyatakan seharusnya distribusi sampel responden mengikuti distribusi teoritis, yaitu semua warna dipilih merata.

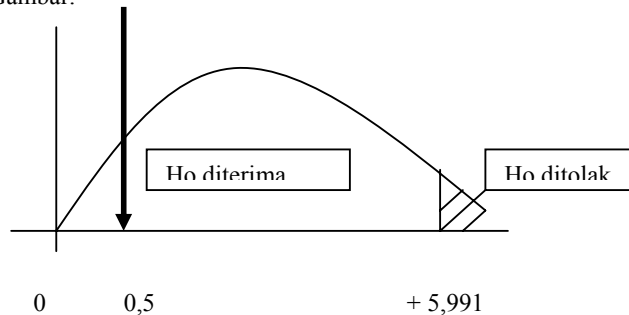
b. Dasar pengambilan keputusan:

- Dengan membandingkan Chi-Square Hitung dengan Chi-Square tabel:
 - Jika Chi-Square Hitung < Chi-Square tabel, Ho diterima
 - Jika Chi-Square Hitung > Chi-Square tabel, Ho ditolak
- Dengan melihat angka probabilitas, dengan ketentuan:
 - Probabilitas >0,05 maka Ho diterima
 - Probabilitas <0,05 maka Ho ditolak

c. Keputusan:

- Perbandingan Chi-Square Hitung dengan Chi-Square tabel:
 - Dari tabel Output TEST STATISTICS terlihat angka Chi Square hitung adalah 0,500
 - Mencari Chi-Square Tabel:
Derajat kebebasan (df)= jumlah baris-1= 3-1 = 2
Tingkat kepercayaan=95%, berarti tingkat signifikansi 100%-95%=5%.
Pada $\chi^2_{(0,05;2)}$ didapat angka 5,991

Gambar:



Karena Chi-Square hitung (χ^2) < Chi-Square tabel, maka H_0 diterima.

- Berdasar angka Probabilitas:

Karena angka pada kolom EXACT. SIG adalah 0,779 yang adalah > 0,05, maka H_0 diterima.

Kedua cara menghasilkan keputusan yang sama, yaitu H_0 diterima. Hal ini berarti konsumen menyukai ketiga warna sabun mandi secara proporsional, dalam arti tidak ada warna yang lebih disukai dari lainnya. Perhatikan walaupun dalam sampel warna putih paling sedikit peminatnya, dan warna kuning paling banyak yang menyukai, namun setelah diuji dengan Chi-Square, ternyata distribusi sampel tersebut *masih sesuai* dengan distribusi teoritis (yang seharusnya), yaitu warna disukai secara merata.

Penggunaan EXACT TEST

Jika dilihat tabel output diatas, khususnya pada TEST STATISTICS, terlihat keterangan bahwa ada 3 sel yang mempunyai nilai EXPECTED dibawah angka 5 (yaitu hanya bernilai 4). Selain itu, jumlah frekuensi total tidak mencapai 20 (yaitu hanya 12). Hal ini menyebabkan keputusan yang diambil bisa diragukan valid tidaknya. Untuk itu, SPSS menyediakan option EXACT untuk menguji hasil chi-square dengan kondisi jumlah data dibawah 20 dan ada sel yang berangka dibawah 5.

Prosedur penggunaan Exact:

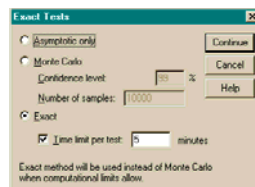
- Buka file **chi square**
- Dari menu **Analyze**, pilih sub menu **Nonparametric Test**, lalu pilihan **chi square...**

Tampak di layar kotak dialog CHI-SQUARE TEST dengan langkah pengisian persis seperti sebelumnya:

- TEST VARIABLE LIST, masukkan variabel **warna**.
- EXPECTED VALUES, tetap pada pilihan ALL CATEGORIES EQUAL.

Sekarang buka kotak **EXACT...**

Tampak di layar:



aktifkan pilihan **Exact** dan juga aktifkan bagian **Time limit per test** , dan kemudian ketik **5**.

NB: ketik 5 berarti SPSS diberi waktu maksimal 5 menit untuk mencari ketepatan uji chi-square.

- Abaikan bagian lain dan tekan OK.

OUPUT:

(lihat file **chi square exact.spo**)

NPar Tests

Chi-Square Test

Frequencies

WARNA			
	Observed N	Expected N	Residual
1.00 putih	3	4.0	-1.0
2.00 hijau	4	4.0	.0
3.00 kuning	5	4.0	1.0
Total	12		

Test Statistics	
	WARNA
Chi-Square ^a	.500
df	2
Asymp. Sig.	.779
Exact Sig.	.935
Point Probability	.313

a. 3 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 4.0.

Analisis:

Perhatikan tampilan tambahan yakni Exact Sig. Dan Point Probability. Terlihat angka signifikansi dari proses penambahan ketepatan hasil uji sebelumnya (Exact) yang ternyata juga diatas 0,05 (angka Exact 0,935). Dengan demikian, kesimpulan yang diambil sama saja seperti sebelumnya (asymptotic significant) yang bernilai 0,779 dan diatas 0,05 pula. Pada banyak kasus, angka hasil uji Exact dengan tanpa uji Exact (hanya asymp. Sig) menghasilkan kesimpulan yang sama, walaupun pada beberapa kasus, angka dan kesimpulan dari uji Exact bisa berbeda. Jika terjadi perbedaan kesimpulan, maka hasil uji Exact yang dijadikan pedoman keputusan akhir.

KASUS 2:

Berkaitan dengan kasus 1, Manajer Pemasaran menduga bahwa untuk daerah pemasaran Jawa Tengah, kesukaan konsumen akan warna sabun mandi justru tidak merata. Dia beranggapan konsumen yang suka warna putih dan hijau masing-masing hanya 25%, sedangkan 50% konsumen menyukai warna kuning.

Untuk membuktikan dugaannya, Manajer Pemasaran tersebut kembali meminta pendapat 30 responden yang berdomisili di Jawa Tengah, dengan data sebagai berikut:

(lihat file **CHI-SQUARE2**):

Responden	Warna Pilihan	Responden	Warna Pilihan
RENNY	putih	HANNY	kuning
RITA	putih	ANNY	kuning
SLAMET	putih	SUDIMAN	putih

ALAN	putih	GUNAWAN	putih
BUDIMAN	hijau	ANITA	putih
GUNADI	hijau	CICILIA	putih
GIMAN	kuning	FERRY	putih
LINNA	putih	SILVI	hijau
LANNY	putih	AGUS	hijau
RUBEN	hijau	NANA	kuning
RICKY	hijau	SUGIANTO	Kuning
ROSSY	hijau	DEWI	Hijau
YUNI	putih	ROSALIND	Hijau
CYNTHIA	putih	SUSANA	Putih
ROSALINDA	putih	ROBERT	Putih

Keterangan:

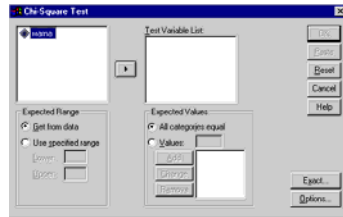
kode dan warna sabun mandi:

KODE	WARNA
1	Putih
2	Hijau
3	Kuning

Langkah:

- Buka file **chi square2**
- Dari menu **Analyze**, pilih sub menu **Nonparametric Test**, lalu pilihan **chi square...**

Tampak di layar:



Pengisian:

- TEST VARIABLE. Masukkan variabel **warna**.
- EXPECTED RANGE. Pilihan tetap pada GET FROM DATA.
- EXPECTED VALUES. Jika dilihat pada kasus, Manajer menganggap kesenangan terhadap warna adalah TIDAK SAMA. Untuk itu dilakukan pengisian distribusi teoritis dengan langkah:

- o Pilih (klik mouse) pada VALUES. Terlihat kotak sebelah kanan terbuka dan berwarna putih, yang berarti *harus diisi*.
- o Karena kode warna 1 adalah putih, maka warna ini yang pertama diisi. Sesuai kasus yang menyatakan warna putih diduga disukai hanya 25% konsumen, ketik **25**

NB: tidak perlu ditambah dengan tanda %, karena SPSS otomatis menganggap input dalam bentuk persentase.

Kemudian tekan tombol **Add** untuk memasukkan nilai tersebut ke kotak dibawahnya.

- o Kode warna berikut (2) adalah hijau. Sesuai kasus yang menyatakan warna hijau diduga disukai hanya 25% konsumen, ketik **25**

Kemudian tekan tombol **Add** untuk memasukkan nilai tersebut ke kotak dibawahnya.

- o Kode warna terakhir (3) adalah kuning. Sesuai kasus yang menyatakan warna kuning diduga disukai 50% konsumen, ketik **50**

Kemudian tekan tombol **Add** untuk memasukkan nilai tersebut ke kotak dibawahnya.

- Abaikan bagian lain dan tekan OK.

Output:

(lihat file **CHI SQUARE2.spo**)

NPar Tests

Chi-Square Test

Frequencies

WARNA			
	Observed N	Expected N	Residual
putih	16	7.5	8.5
hijau	9	7.5	1.5
kuning	5	15.0	-10.0
Total	30		

Test Statistics	
	WARNA
Chi-Square ^a	16.600
df	2
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 7.5.

Analisis:

1. Dari tabel WARNA terlihat bahwa putih dipilih oleh 16 responden, warna hijau 9 responden dan warna kuning 5 responden. Ini adalah frekuensi yang didapat dari input data (observed).
2. Dari tabel WARNA untuk kolom EXPECTED, karena distribusi tidak merata, maka distribusi teoritis masing-masing warna adalah:
 - Putih = $25\% \times 30 = 7,5$
 - Hijau = $25\% \times 30 = 7,5$
 - Kuning = $50\% \times 30 = 15$
3. Kolom RESIDUAL dari tabel WARNA adalah selisih antara kolom Observed dan kolom Expected. Seperti untuk warna putih, angka residual adalah $16 - 7,5$ atau $+ 8,5$. Demikian untuk dua angka residual yang lain.
4. Proses pengambilan keputusan:
 - a. HIPOTESIS:

Ho: Distribusi data sesuai dengan distribusi teoritis

Hi: Distribusi data tidak sesuai dengan distribusi teoritis

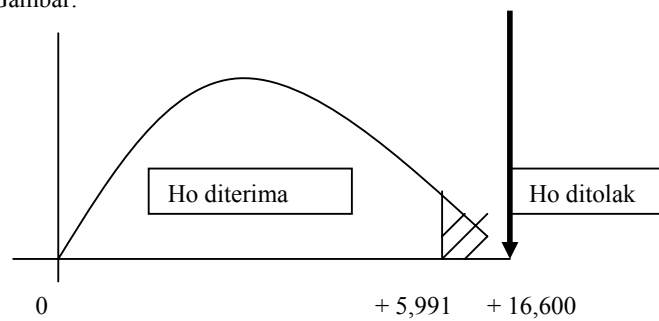
b. Dasar pengambilan keputusan:

- Dengan membandingkan Chi-Square Hitung dengan Chi-Square tabel.
- Dengan melihat angka probabilitas

c. Keputusan:

- Perbandingan Chi-Square Hitung dengan Chi-Square tabel:
 - Dari tabel Ouput TEST STATISTICS terlihat angka Chi Square hitung adalah 16,600
 - Mencari Chi-Square Tabel:
Derajat kebebasan (df)= 2 dan Tingkat signifikansi 5%
Pada $\chi^2_{(0,05;2)}$ didapat angka 5,991

Gambar:



Karena Chi-Square hitung (χ^2) > Chi-Square tabel, maka Ho ditolak.

- Berdasar angka Probabilitas:

Karena angka pada kolom EXACT. SIG adalah 0,00 yang adalah jauh lebih kecil dari 0,05, maka Ho ditolak.

Kedua cara menghasilkan keputusan yang sama, yaitu H_0 ditolak. Hal ini berarti dugaan Manajer Pemasaran diatas adalah tidak benar, atau distribusi sampel yang didapat ternyata tidak sesuai (fit) dengan distribusi teoritis (yang dikemukakan Manajer Pemasaran sebelumnya).

Sehingga, jika Manajer Pemasaran sebelumnya menduga sabun mandi warna putih hanya 25% peminatnya, ternyata dari survei konsumen malah diminati hampir separo konsumen. Sebaliknya warna kuning ternyata malah kurang diminati konsumen.

NB: disini tidak perlu dilakukan uji tambahan Exact, karena tidak ada sel expected yang bernilai dibawah 5 (lihat keterangan di tabel output test statistics).

LATIHAN

BUKA FILE **deskriptif.sav**

1. Lakukan GOODNESS OF FIT TEST pada variabel SIKAP, dengan asumsi bahwa ketiga jenis Sikap mempunyai proporsi yang sama.
2. Lakukan GOODNESS OF FIT TEST pada variabel USIA, dengan asumsi bahwa ketiga jenis Sikap mempunyai proporsi yang sama.