Suplemen Responsi **ANALISIS DATA KATEGORIK (STK351)** **Departemen Statistika – FMIPA IPB**			**Pertemuan** **5**
Pokok Bahasan	**Sub Pokok Bahasan**	**Referensi**	**Waktu**
Uji Khi-Kuadrat	▪ Uji Kebebasan ▪ Uji Kehomogenen ▪ Uji Kebaikan Suai untuk Beberapa Sebaran Diskret	Applied Nonparametric Statistic Daniel (1990)	Jumat 12 Okt 2012 15.30 – 16.30

Uji Kebebasan Khi-Kuadrat

Salah satu pertanyaan paling sering diajukan dalam suatu penelitian adalah "apakah dua peubah saling berhubungan?". Misalnya, seorang konsultan ingin mengetahui apakah jenis kelamin berhubungan dengan tingkat kesibukan aktivitas seseorang. Atau mungkin juga seorang dokter ingin mengatahui apakah ada kaitan antara jenis kelamin dengan perilaku merokok. Apabila tidak saling berhubungan, kedua peubah tersebut dikatakan saling bebas *(independent)*. Jadi, dua peubah dikatakan saling bebas apabila distribusi satu peubah di antaranya tidak dipengaruhi oleh distribusi peubah lainnya. Apabila dua peubah saling bebas, nilai pada satu peubah tidak dapat digunakan untuk menghitung ataupun memperkirakan nilai pada peubah lainnya, demikian sebaliknya. Uji kebabasan yang biasa digunakan adalah uji Khi-Kuadrat. Dalam bahasa Inggris, uji ini disebut sebagai *Chi-Square Test for Independence*.

Asumsi

a. Data terdiri dari sebuah contoh acak berukuran *n* dari beberapa populasi
b. Pengamatan-pengamatan pada contoh dapat dikategori-silangkan berdasarkan dua kriteria, sehingga setiap pengamatan akan berada pada satu – dan hanya satu – kategori dari setiap kriteria. Kriteria yang dimaksud adalah peubah-peubah yang diamati.
c. Peubah yang diamati bersifat kategorik, atau dapat juga berupa peubah kuantitatif yang pengukurannya dapat dinyatakan dalam kategori numerik.

Hipotesis

H_0 : Dua peubah saling bebas
H_1 : Dua peubah *tidak* saling bebas

Statistik Uji

Misalkan ada satu tabel kontingensi yang manampilkan frekuensi atau banyaknya pengamatan. Dari frekuensi teramati dan frekuensi harapan, dapat dihitung selisih atau jarak antar keduanya. Ketika hipotesis awal benar, dapat dihitung statistik :

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

Statistik ini akan mendekati sebaran χ^2 dengan derajat bebas (r-1) (c-1); dalam hal ini r adalah banyaknya baris *(row)* dan c adalah banyaknya kolom *(column)* pada tabel kontingensi.

Sedangkan frekuensi harapan sendiri dapat dihitung dengan rumus :

$$E_{ij} = \frac{n_{i.} n_{.j}}{n}$$

Kaidah Keputusan

Hipotesis awal bahwa dua peubah saling bebas ditolak pada taraf nyata α apabila nilai hitung dari statistik X^2 lebih besar dari nilai $\chi^2_{1-\alpha}$ dengan derajat bebas (r-1) (c-1). (Tabel Khi-Kuadrat, A.11).

<u>Tabel Kontingensi 2 x 2</u> Untuk tabel kontingensi dengan ukuran r=c=2 seperti ditunjukkan pada *layout* dibawah ini :

Kriteria 1	Kriteria 2		
	1	2	Total
1	a	b	a + b
2	c	d	c + d
Total	a + c	b + d	$n = a + b + c + d$

Statistik X^2 secara sederhana dapat diperoleh dengan rumus :

$$X^2 = \frac{n(ad - bc)^2}{(a+c)(b+d)(c+d(a+b)}$$ yang menyebar $\chi^2_{1-\alpha}$ dengan derajat bebas 1.

Contoh :

Seorang konsultan ingin mengetahui hubungan antara jenis kelamin (wanita, pria) dengan tingkat kesibukan (tinggi, menengah, rendah) seseorang. Data frekuensi ditampilkan dalam tabel kontingensi di bawah ini. Ujilah apakah jenis kelamin dan tingkat kesibukan memiliki hubungan yang signifikan?

Jenis Kelamin	Tingkat kesibukan			
	Tinggi	Sedang	Rendah	Total
Wanita	5	26	4	35
Pria	16	35	5	56
Total	21	61	9	91

Sumber data : Minitab 15 Sample Data : Exh_tabl.MTW

<u>Hipotesis</u> : H_0 : Jenis kelamin dan tingkat kesibukan saling bebas
H_1 : Jenis kelamin dan tingkat kesibukan *tidak* saling bebas

<u>Statistik Uji</u> : Statistik uji dapat diperoleh melalui prosedur :

Frek. Teramati *(O)* Frek. Harapan *(E)*	Jenis Kelamin	Tingkat kesibukan			
		Tinggi	Sedang	Rendah	Total
	Wanita	5	26	4	35
		8.08	23.46	3.46	
	Pria	16	35	5	56
		12.92	37.54	5.54	
	Total	21	61	9	91

Menghitung frekuensi harapan, misalnya untuk *cell* jenis kelamin=wanita dan tingkat kesibukan=tinggi, $E_{11} = n_{1.}n_{.1}/n$ = (35)(21)/91 = 8.08

Sehingga, dapat dihitung nilai statistik uji :

$$X^2 = \frac{(5-8.08)^2}{8.08} + \frac{(26-23.46)^2}{23.46} + \cdots + \frac{(5-5.54)^2}{5.54} = 2.49$$

Dengan derajat bebas (2-1)(3-1) = 2

Keputusan : Berdasarkan tabel A.11, diperoleh $\chi^2_{(1-0.05)(db=2)} = 5.991$. Karena $\chi^2_{1-0.05} > X^2$ maka hipotesis nol tidak ditolak dan simpulkan bahwa jenis kelamin dan tingkat kesibukan saling bebas (alfa 5%). Dalam hal ini 0.1 < *p-value* < 0.95.

Uji Kehomogenan Khi-Kuadrat

Uji kehomogenan khi-kuadrat *(Chi-Square Test of Homogeneity)* digunakan untuk menguji hipotesis nol bahwa dua (atau lebih) populasi homogen. Uji kehomogenan khi-kuadrat sangat mirip dengan uji kebebasan khi-kuadrat, baik dari segi prosedur maupun rumus-rumus yang digunakan. *Perbedaan keduanya* terletak pada dua hal penting, yaitu *prosedur penarikan contoh dan pemikiran yang melandasi perhitungan frekuensi harapan*. Selain asumsi yang digunakan pada uji kebebasan, tentu saja uji kehomogenan membutuhkan asumsi tambahan yaitu *contoh yang diuji harus saling bebas.*

Contoh :

Di bawah ini adalah tabel jumlah penduduk berdasarkan jenjang pendidikan dan jenis kelamin. Jika diasumsikan saling bebas, ujilah apakah keenam jenjang pendidikan homogen menurut jenis kelamin :

Pendidikan	Jenis Kelamin		
	Wanita	Pria	Total
Bachelor	118	321	439
College	281	523	804
Doctorate	6	23	29
HS-Grad	286	551	837
Master	40	106	146
No HS-Grad	96	235	331
Total	827	1759	2586

Sumber data : http://archive.ics.uci.edu/ml/datasets/Adult.

Hipotesis : H_0 : Keenam populasi yang direpresentasikan oleh kelompok jenjang pendidikan adalah homogen menurut jenis kelamin

H_1 : Keenam populasi tidak homogen

Statistik Uji :

Frek. Teramati *(O)* Frek. Harapan *(E)*	Pendidikan	Jenis Kelamin Wanita	Pria	Total
	Bachelor	118	321	439
		104.4	298.6	
	College	281	523	804
		257.1	546.9	
	Doctorate	6	23	29
		9.3	19.7	
	HS-Grad	286	551	837
		267.7	569.3	
	Master	40	106	146
		46.7	99.3	
	No HS-Grad	96	235	331
		105.9	225.1	
	Total	827	1759	2586

Sehingga, dapat dihitung nilai statistik uji :

$$X^2 = \frac{(108-104.4)^2}{104.4} + \frac{(281-257.1)^2}{257.1} + \cdots + \frac{(235-225.1)^2}{225.1} = 13.06$$

Dengan derajat bebas (6-1)(2-1) = 5

Keputusan : Berdasarkan tabel A.11, diperoleh $\chi^2_{(1-0.05)(db=5)} = 11.070$. Karena $X^2 > \chi^2_{1-0.05}$ maka hipotesis nol ditolak dan simpulkan bahwa keenam populasi yang direpresentasikan oleh kelompok jenjang pendidikan tidak homogen menurut jenis kelamin (taraf nyata 5%). Dalam hal ini 0.01 < *p-value* < 0.025.

Uji Kebaikan Suai Khi-Kuadrat

Uji kebaikan suai khi-kuadrat *(Chi-Square Goodness of Fit Test)* sangat mirip dengan uji khi-kuadrat untuk kebebasan dan kehomogenan. Statistik uji untuk ketiga pengujian ini dihasilkan dari perbandingan antara frekuensi teramati dengan frekuensi harapannya. Uji kebaikan suai khi-kuadrat biasanya digunakan untuk memeriksa apakah contoh acak berasal dari populasi yang mengikuti sebaran tertentu.

Asumsi

a. Data terdiri dari *n* contoh acak yang saling bebas
b. Skala pengukuran *boleh* nominal
c. Pengamatan dapat dikategorikan ke dalam *r* kategori. Jumlah pengamatan yang berada pada kategori tertentu disebut sebagai *frekuensi teramati* dari kategori tersebut.

Hipotesis

H_0 : Contoh berasal dari populasi yang menyebar menurut sebaran tertentu
H_1 : Contoh bukan berasal dari populasi yang menyebar menurut sebaran tertentu

Statistik Uji

Statistik uji kebaikan suai khi-kuadrat adalah :

$$X^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}$$

yang mendekati sebaran χ^2 dengan derajat bebas (r-1), dalam hal ini r adalah banyaknya kategori. Sedangkan frekuensi harapan dapat dihitung dengan rumus : $E_i = np_i$. Dalam hal ini p adalah peluang pengamatan ada pada kategori ke-i, i=1, 2, ..., r.

Contoh (Sebaran Seragam) :

Di bawah ini adalah data frekuensi mahasiswa berdasarkan perolehan nilai yang diambil secara acak dari sebuah kelas.

Nilai	A	B	C	D	E
Frekuensi	12	15	13	11	9

Apakah nilai mahasiswa di kelas tersebut seragam?

<u>Hipotesis</u> : H_0 : Nilai mahasiswa di kelas tersebut seragam (atau, nilai mahasiswa mengikuti sebaran seragam)
H_1 : Nilai mahasiswa tidak seragam

<u>Statistik Uji</u> : Sebaran seragam mempunyai peluang p_i = 1/r untuk semua i, i = 1, 2 ..., r. Sehingga, kita akan memperoleh p = 1/5 = 0.2 untuk kelima kategori (nilai mahasiswa). Dengan demikian, untuk n=60 frekuensi harapan bagi setiap kategori adalah E = (0.2)(60) = 12

Nilai	Frekuensi (O)	Frekuensi harapan (E)	$(O – E)^2$ / E
A	12	12	0
B	15	12	0.75
C	13	12	0.083
D	11	12	0.083
E	9	12	0.75
Total	60	60	1.66

Diperoleh $X^2 = 1.66$ dengan derajat bebas 5 – 1 =4.

<u>Keputusan</u> : Berdasarkan tabel A.11, diketahui $\chi^2_{(1-0.05)(db=4)} = 9.488$. Karena $X^2 < \chi^2_{1-0.05}$ maka hipotesis nol tidak ditolak dan simpulkan bahwa nilai mahasiswa disuatu kelas tersebut (mendekati) seragam. Dalam hal ini 0.1 < *p-value* < 0.95.

Contoh (Sebaran Binomial) :

Misalkan ada 1000 keranjang yang masing-masing berisi 10 buah jeruk. Beberapa di antara jeruk-jeruk tersebut ada yang membusuk. Diperoleh data sebagai berikut :

Banyaknya jeruk yang membusuk dalam satu keranjang	0	1	2	3	4	5	6
Banyaknya keranjang	334	369	191	63	22	12	9

(Sumber : http://www.math.unb.ca/~rolf)

Apakah banyaknya jeruk yang membusuk dalam keranjang mengikuti sebaran Binom(10,*p*) ?

Hipotesis : H_0 : Banyaknya jeruk yang membusuk dalam keranjang menyebar binomial
H_1 : Banyaknya jeruk yang membusuk dalam keranjang tidak menyebar binomial

Statistik Uji : Peluang jeruk membusuk, *p,* tidak diketahui. Sehingga, kita hanya dapat menduganya dengan proporsi jeruk yang membusuk sebagai berikut :

$$\widehat{p} = \frac{\#\text{ jeruk yg membusuk}}{\#\text{ jeruk}} = \frac{(0)(334)+(1)(369)+\cdots+(6)(9)}{(1000)(10)} = 0.11$$

Dari tabel binomial (A.1), peluang binomial(n=10,p=0.11) untuk r=0, 1, ..., 6 berturut-turut adalah 0.3118, 0.3854, 0.2143, 0.0706, 0.0153, 0.0023, 0.0002.

r	*O*	*P* (*r* | binom(10,0.11)	*E*	$(O-E)^2/E$
0	334	0.3118	311.8	1.58
1	369	0.3854	385.4	0.70
2	191	0.2143	214.3	2.53
3	63	0.0706	70.6	0.82
4	22	0.0153	15.3	2.93
5 atau 6	21	0.0025	2.5	136.90
Total	1000			145.46

Catatan : Karena peluang dan frekuensi harapan untuk *r=6* sangat kecil, maka digabung dengan *r*=5

Diperoleh $X^2 = 145.46$ dengan derajat bebas 6 – 1 – 1 = 4.

Ketika menghitung nilai dugaan bagi parameter binomial *p,* kita kehilangan 1 derajat bebas. Sehingga, derajat bebas bagi statistik uji ini menjadi 6 – 1 – 1 = 4, bukan 6 – 1 = 5.

Keputusan : Berdasarkan tabel A.11, diketahui $\chi^2_{(1-0.05)(db=4)} = 9.488$. Karena $X^2 > \chi^2_{1-0.05}$ maka hipotesis nol ditolak dan simpulkan bahwa banyaknya jeruk yang membusuk dalam keranjang tidak menyebar binomial (taraf nyata 5%). Dalam hal ini *p-value* < 0.005.

Contoh (Sebaran Poisson) :

Di bawah ini adalah data dari banyaknya kedatangan nasabah per menit di sebuah bank yang diamati selama periode 200 menit. Ujilah apakah data ini menyebar poisson?

Kedatangan	0	1	2	3	4	5	6	7	8
Frekuensi	14	31	47	41	29	21	10	5	2

(Sumber : http://courses.wcupa.edu/rbove/Berenson)

<u>Hipotesis</u> : H_0 : Banyaknya kedatangan per menit menyebar Poisson
H_1 : Banyaknya kedatangan per menit tidak menyebar Poisson

<u>Statistik Uji</u> : Parameter sebaran Poisson, λ (mean), tidak diketahui. Namun, itu dapat diduga dengan rata-rata kedatangan per menit sebagai berikut :

$$\hat{\lambda} = \bar{X} = \frac{\#\ \text{kedatangan}}{\text{periode waktu}} = \frac{(0)(14)+(1)(31)+\cdots+(8)(2)}{200} = 2.90$$

Nilai $\hat{\lambda}$ selanjutnya dapat digunakan untuk menduga peluang poisson untuk $\hat{\lambda} = 2.90$ dan r = 0, 1, ..., 8, melalui tabel Poisson atau dengan rumus :

$$P(r) = \frac{\hat{\lambda}^r \exp(-\hat{\lambda})}{r!}.$$

Misalnya, untuk r=1 diperoleh $P(r=8) = \frac{2.9^8 \exp(-2.9)}{8!} = \frac{275.25}{40320} = 0.0068$

r	*O*	*P* (*r* | Poisson(λ=2.9)	*E*	$(O-E)^2/E$
0	14	0.0550	11.00	0.8182
1	31	0.1596	31.91	0.0265
2	47	0.2314	46.27	0.0112
3	41	0.2237	44.73	0.3126
4	29	0.1622	32.43	0.3648
5	21	0.0940	18.81	0.2574
6	10	0.0455	9.09	0.0890
7	5	0.0188	3.77	0.4089
8	2	0.0068	1.37	0.3012
Total	200			2.5899

Diperoleh $X^2 = 2.59$ dengan derajat bebas 9 – 1 – 1 = 7.

<u>Keputusan</u> : Berdasarkan tabel A.11, diketahui $\chi^2_{(1-0.05)(db=7)} = 14.067$. Karena $X^2 < \chi^2_{1-0.05}$ maka hipotesis nol tidak ditolak dan simpulkan bahwa banyaknya kedatangan pelanggan per menit di sebuah bank mengikuti sebaran Poisson (taraf nyata 5%). Dalam hal ini 0.1 < *p-value* < 0.95.

<u>Tugas</u> : Buku Daniel (1990) hal. 190 latihan 5.6, dan hal. 202 latihan 5.20, dan hal. 317 latihan 8.5 (ketiga soal menggunakan taraf nyata 5%)