II. LANDASAN TEORI

2.1. Data Kategori

Wallpole (1995), mendefinisikan data kategori sebagai data yang diklasifikasikan menurut kriteria tertentu. Data kategori disebut juga data nonmetrik atau data yang bukan merupakan hasil pengukuran. Data kategori merupakan data kualitatif sehingga untuk dapat dianalisis dengan menggunakan rumus matematika/statistika perlu diberi kode (*coding*) berupa angka. Analisis matematika/statistika terhadap data kategori dilakukan berdasarkan hasil membilang (*counting*) pada setiap kategori/pasangan kategori.

Klasifikasi data kategori adalah :

1. Kategori Nominal.
2. Kategori Ordinal.

2.2. Tabel Kontingensi

Untuk memudahkan tampilan dan pembacaan data kategori, tabel kontingensi adalah metode yang tepat. Agresti (1990) mendefinisikan tabel kontingensi ini sebagai teknik penyusunan data yang cukup sederhana untuk melihat hubungan antara beberapa variabel dalam satu tabel.

Watson (1993) menerangkan bahwa jika data dari suatu variabel acak yang diambil dari suatu populasi diklasifikasikan dalam dua variabel kategori atau kriteria, maka salah satu kelas dapat direpresentasikan sebagai baris dalam tabel dan kelas yang lain direpresentasikan sebagai kolom. Secara umum tabel dengan *r* baris dan *c* kolom dikenal dengan tabel kontingensi (*contingency table*) atau tabulasi silang (*cross tabulation*).

Baris	Kolom					Total Baris
	1	2	3	…	c	
1	Y_{11}	Y_{12}	Y_{13}	…	Y_{1c}	R_1
2	Y_{21}	Y_{22}	Y_{23}	…	$Y2c$	R_2
3	Y_{31}	Y_{32}	Y_{33}	…	Y_{3c}	R_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
r	Y_{r1}	Y_{r2}	Y_{r3}	…	Y_{rc}	R_r
Total Kolom	C_1	C_2	C_3	...	C_c	n

Keterangan :

Y_{ij} : frekuensi observasi (f_0) yaitu nilai observasi sampel acak pada baris ke-*i* dan kolom ke- *j*

n : ukuran sampel $= \sum_{i=1}^{r}\sum_{j=1}^{c} Y_{ij} = \sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j$

Jika f_0 adalah frekuensi observasi, f_0 dapat diduga oleh frekuensi harapan (*expected frequency*) f_e, frekuensi harapan dari observasi baris ke-*i* dan kolom ke-j dirumuskan dengan :

$$fe_{ij} = \frac{(total\ baris)(total\ kolom)}{ukuran\ sampel} = \frac{R_i C_j}{n}$$

(Watson, *et.al.*, 1993).

2.3. Distribusi Multinomial

Sebuah populasi dengan suatu variabel kategori yang terdiri dari *k* kelas (*k* adalah konstanta intejer, $k \geq 2$) disebut dengan populasi multinomial (*multinomial* *population*). Jika proporsi dari elemen-elemen yang termasuk dalam setiap kelas tidak dipengaruhi oleh pemilihan sampel maka model yang tepat adalah distribusi multinomial (*multinomial distribution*). Distribusi multinomial adalah distribusi bersama untuk suatu peubah acak $Y_1, Y_2, \cdots, Y_k$, yang beranggotakan suatu sampel acak berukuran *n* yang termasuk dalam tiap *k* kelas populasi multinomial, Y_j adalah anggota kelas ke-*j*. Parameter dalam suatu distribusi multinomial terdiri dari *n* yang bersifat tetap dan $p_1, p_2, \ldots, p_k$ adalah proporsi kelas ke-*i*. Dengan asumsi $\sum_{i=1}^{k} Y_i = n$ dan $\sum_{i=1}^{k} p_i = 1$.

Fungsi Distribusi Peluang (*Probability Distribution Function*) untuk distribusi multinomial $Y_1, Y_2, \cdots, Y_k$ anggota *k* kelas adalah :

$$P(Y_1, Y_2, \cdots, Y_k) = n! \frac{p_1^{Y_1} p_2^{Y_2} \cdots p_k^{Y_k}}{Y_1! Y_2! \ldots Y_k!} = n! \prod_{i=1}^{k} \frac{p_i^{Y_i}}{Y_i!}$$

(Watson,*et.al*, 1993).

Dengan mengetahui sebaran dari suatu variabel kategori, maka pendugaan terhadap nilai frekuensi observasi dalam suatu tabel kontingensi dapat dilakukan dengan menggunakan *Maximum Likelihood Estimation*.

2.4. Metode Pendugaan Maksimum (*Maximum Likelihood Estimation)*

Salah satu metode pendugaan parameter dari suatu fungsi distribusi adalah metode pendugaan maksimum (*maximum likelihood estimation*). Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari suatu distribusi $f(x_i, \theta)$ dimana θ adalah parameter, $\theta \in \Omega$. Fungsi kepekatan peluang bersama dari $X_1, X_2, \ldots, X_n$ adalah

$f(x_1,\theta)f(x_2,\theta)f\cdots(x_n,\theta)$. Fungsi kepekatan peluang bersama ini dilihat sebagai fungsi θ, yang selanjutnya disebut dengan fungsi *likelihood* (*L*) sampel acak dan ditulis :

$$L(\theta;x_1,x_2,\ldots,x_n)=f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta)$$

Melalui fungsi *likelihood* ini, dapat ditentukan suatu fungsi nontrivial dari $x_1,x_2,\ldots,x_n$ yaitu $u(x_1,x_2,\ldots,x_n)$, sehingga θ dapat digantikan oleh $u(x_1,x_2,\ldots,x_n)$ yang mengakibatkan fungsi *likelihood* (*L(*θ)) akan maksimum.

Fungsi (*L(*θ)) akan dapat maksimum dengan menentukan derivatif pertama dari logaritma fungsi (*L(*θ)) terhadap θ yang sama dengan nol.

$$\frac{\partial \ln L(\theta;x_1,x_2,\ldots,x_n)}{\partial\theta}=0$$

Selanjutnya dengan melakukan penurunan matematis maka akan didapat

$$\hat{\theta}=u(x_1,x_2,\ldots,x_n).$$

$\hat{\theta}$ disebut dengan penduga maksimum likelihood (Hoog dan Craig, 1995).

2.5. Model Log-Linear

Model log –linear merupakan salah satu bentuk khusus dari model linear umum. Analisis log-linear adalah bentuk pemodelan dari suatu tabel kontingensi dua arah yang menganalisis variabel kategori berdasarkan logaritma frekuensi sel. Log-linear dapat digunakan untuk menganalisis hubungan atau asosiasi antar variabel. Log-linear tidak hanya dapat digunakan untuk menganalisis hubungan antara dua variabel kategori, melainkan juga dapat digunakan untuk menganalisis tabel kontingensi yang terdiri dari tiga atau lebih variabel (*multi-way contingency tables*) (Angela, 2009).

Strategi dasar dalam pemodelan log-linear adalah membentuk model berdasarkan frekuensi pengamatan dalam tabel silang dari suatu variabel kategori. Model yang dihasilkan akan merepresentasikan frekuensi harapan yang mungkin berbeda atau menyerupai frekuensi pengamatan.

Bentuk umum model log-linear :

$$\log E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$$

Keterangan :

E(Y_{ijk}) = frekuensi harapan dalam setiap sel

μ = *Intercept* atau konstanta atau rata-rata umum

α_i = parameter pengaruh tingkat ke-*i* dari faktor α

β_j = parameter pengaruh tingkat ke-*j* dari faktor β

γ_k = parameter pengaruh tingkat ke-*k* dari faktor γ

$\alpha\beta_{ij}$ = parameter pengaruh interaksi tingkat ke-*i* dan ke-*j* dari faktor α dan faktor β

$\alpha\gamma_{ik}$ = parameter pengaruh interaksi tingkat ke-*i* dan ke-*k* dari faktor α dan faktor γ

$\beta\gamma_{jk}$ = parameter pengaruh interaksi tingkat ke-*j* dan ke-*k* dari faktor β dan faktor γ

$\alpha\beta\gamma_{ijk}$ = parameter pengaruh interaksi tingkat ke-*i*, ke-*j* dan ke-*k* dari faktor α, β dan faktor γ

Dimana penduga parameternya adalah :

$$\mu = \frac{\sum_i \sum_j \sum_k \log E(Y_{ijk})}{IJK}$$

$$\alpha_i = \frac{\sum_j \sum_k \log E(Y_{ijk})}{JK} - \mu$$

$$\beta_j = \frac{\sum_i \sum_k \log E(Y_{ijk})}{IK} - \mu$$

$$\gamma_k = \frac{\sum_i \sum_j \log E(Y_{ijk})}{IJ} - \mu$$

$$\alpha\beta_{ij} = \frac{\sum_k \log E(Y_{ijk})}{K} - \frac{\sum_j \sum_k \log E(Y_{ijk})}{JK} - \frac{\sum_i \sum_k \log E(Y_{ijk})}{IK} - \mu$$

$$\alpha\gamma_{ik} = \frac{\sum_j \log E(Y_{ijk})}{J} - \frac{\sum_j \sum_k \log E(Y_{ijk})}{JK} - \frac{\sum_i \sum_j \log E(Y_{ijk})}{IJ} - \mu$$

$$\beta\gamma_{jk} = \frac{\sum_i \log E(Y_{ijk})}{I} - \frac{\sum_i \sum_k \log E(Y_{ijk})}{IK} - \frac{\sum_i \sum_j \log E(Y_{ijk})}{IJ} - \mu$$

$$\alpha\beta\gamma_{ijk} = \log \hat{Y}_{ijk} - \frac{\sum_k \log E(Y_{ijk})}{K} - \frac{\sum_j \log E(Y_{ijk})}{J} - \frac{\sum_i \log E(Y_{ijk})}{I} - \frac{\sum_j \sum_k \log E(Y_{ijk})}{JK} - \frac{\sum_i \sum_k \log E(Y_{ijk})}{IK} - \frac{\sum_i \sum_j \log E(Y_{ijk})}{IJ} - \mu$$

Jumlah dari parameter untuk semua index adalah nol, yaitu :

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \sum_i \alpha\beta_{ij} = \sum_j \alpha\gamma_{ij} = \cdots = \sum_k \alpha\beta\gamma_{ijk} = 0$$

(Agresti, 1992).

Model umum diatas disebut juga Model Jenuh (*Saturated Model*) karena model tersebut memuat kemungkinan pengaruh setiap faktor beserta interaksinya.

Selain itu, ada pula model independen sebagai partisi dari model jenuh log-linear yaitu model log-linear yang hanya melibatkan parameter pengaruh faktor utama dan tidak melibatkan parameter interaksi antar faktor. Model ini digunakan jika tidak ada interaksi antar variabel. Bentuk umumnya sebagai berikut :

$$\log E(Y_{ijkl}) = \mu + \alpha_i + \beta_j + \gamma_k$$

Selain model independen dan model jenuh, dikenal pula model hirarki log-linear, Model tersebut adalah partisi dari model jenuh. Model hirarki menghilangkan parameter-parameter faktor yang tidak signifikan berpengaruh dan hanya melibatkan faktor yang signifikan berpengaruh. Contohnya :

$$\log E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\gamma_{ik}$$

Pembentukan model statistika adalah mencari model sesederhana mungkin yang dapat mencocokkan data tanpa harus melibatkan parameter faktor yang komplek dan berlebihan (Agus, 2005). Pada model log-linear jenuh (*saturated model*)

$$\log E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$$

melibatkan semua parameter faktor yang terlibat beserta interaksinya secara lengkap. Kelebihan parameter ini menyebabkan model jenuh selalu mencocokkan data dengan baik, tetapi tujuan pemodelan yang diinginkan adalah mencari model sederhana yang dapat mencocokkan data dengan tepat tanpa harus melibatkan parameter faktor yang berlebihan. Oleh karena itu diperlukan pengujian parameter (Nala, 2003).

Sebagai indikasi apakah suatu parameter faktor atau parameter interaksi faktor itu signifikan berpengaruh terhadap model, diperlukan suatu statistik yang dikenal dengan Rasio Kesamaan Chi-kuadrat (*Likelihood Ratio Chi-square*) dilambangkan dengan G^2. Statistika G^2 berdistribusi Chi-kuadrat (Agung, 2004).

Rasio Kesamaan Chi-kuadrat (G^2) dapat dihitung secara manual berdasarkan rumus :

$$G^2 = 2\sum_i \sum_j \sum_k \left[fo_{ijk} \ln\left(\frac{fo_{ijk}}{fe_{ijk}} \right) \right]$$

(Agung, 2004).

Dengan hipotesis

Ho : tidak ada pengaruh faktor interaksi

H_1 : ada pengaruh faktor interaksi

Pengambilan keputusan berdasarkan perbandingan nilai G^2 terhadap nilai χ^2 *tabel* atau berdasarkan nilai *p-value*. Jika nilai statistik G^2 lebih besar dari nilai χ^2 *tabel* atau *p-value* $< \alpha$ (taraf nyata) maka disimpulkan bahwa interaksi tersebut signifikan berpengaruh dan perlu dimasukkan dalam model (Agung, 2004).

2.6. Pengujian Chi Square

Pengujian mengenai keindependenan variabel kategori atau kriteria klasifikasi baris dan kolom pada tabel kontingensi dihitung berdasarkan frekuensi harapan dari tiap cell (f_e) dan menggunakan pengujian statistik χ^2.

Uji ini digunakan untuk menguji apakah ada hubungan antara dua variabel kategori (data kualitatif). Pada uji ini digunakan tabel kontingensi dengan banyaknya baris *r* dan banyaknya kolom *c* (tabel kontingensi *r x c*).

Pengujian hipotesis yang dilakukan adalah:

Ho : tidak ada hubungan antara baris dan kolom

H_1 : ada hubungan antara baris dan kolom

Statistik Ujinya adalah: $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(fo_{ij} - fe_{ij}\right)^2}{fe_{ij}}$

keterangan:

fo_{ij} : frekuensi observasi pada baris ke-*i* dan kolom ke-*j*

fe_{ij} : frekuensi harapan pada baris ke-*i* dan kolom ke-*j*,

n : ukuran sampel

dengan derajat bebas $v = (c-1)(r-1)$

Pengambilan keputusan didasarkan pada hal dibawah ini :

a. Berdasarkan perbandingan χ^2 hitung dan tabel

Jika χ^2 hitung < χ^2 tabel, maka Ho tidak ditolak

Jika χ^2 hitung > χ^2 tabel, maka Ho ditolak

b. Berdasarkan probabilitas (*p-value*)

Jika probabilitas > α maka H_0 tidak ditolak

Jika probabilitas < α maka H_0 ditolak

(Sanders & Smidt, 2000)

2.7. Uji z

Dua hal penting dalam statistika inferensia adalah pendugaan paramater dan pengujian hipotesis statistika. Pengujian hipotesis dilakukan untuk menjawab suatu pertanyaan hipotesis yang merupakan suatu dugaan sementara.

Salah satu cara pengujian hipotesis statistik yang umum digunakan adalah dengan menggunakan Pengujian nilai z (*z-test*). Uji z didasarkan pada pendekatan nilai pengamatan terhadap nilai z. Suatu pengamatan θ dari suatu populasi yang mempunyai nilai tengah μ_θ dan simpangan baku σ_θ, mempunyai nilai *z* yang didefinisikan sebagai :

$$z = \frac{\theta - \mu_\theta}{\sigma_\theta}$$

Dengan perumusan hipotesis :

$$Ho : \theta = \theta_0$$
$$H_1 : \theta > \theta_0 \quad atau \quad \theta < \theta_0$$

atau dikenal dengan uji satu arah.

$$Ho : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

atau dikenal dengan uji dua arah.

Dengan taraf nyata sebesar α maka pengambilan keputusan tolak Ho dilakukan jika $z > z_{\alpha}$ atau $z < -z_{\alpha}$ untuk pengujian satu arah. Sedangkan tolak Ho jika $|z| < z_{\alpha/2}$ untuk pengujian dua arah. α didefinisikan sebagai nilai peluang tolak Ho padahal Ho benar, oleh karena itu nilai α dibuat sekecil mungkin. Nilai α yang umum digunakan adalah 0,01 atau 0,05 (Sanders & Smidt, 2000).