Menentukan Distribusi Variabel Random Berdasar Observasi yang Memuat Bentuk Kuadratik

Nurkaromah Dwidayati

Jurusan Matematika FMIPA Universitas Negeri Semarang
noengkd_unnes@yahoo.co.id

Abstrak

Dalam praktek, seringkali dihadapkan pada masalah pengambilan simpulan berdasar informasi atau data yang terbatas (sampel). Suatu pertanyaan yang muncul sejauh manakah simpulan yang diambil dipercayai memiliki kualitas yang baik? Untuk dapat melakukan apresiasi terhadap hasil inferensi tersebut, diperlukan teori tentang distribusi peluang. Berdasar distribusi tersebut, semua pertanyaan tentang populasi dapat dijawab dengan baik. Persoalan yang muncul adalah bagaimana menentukan distribusi variabel random yang merupakan fungsi dari beberapa variabel random yang dapat dinyatakan dalam bentuk kuadrat? Pada pemaparan ini akan dikaji distribusi bentuk kuadratik beserta sifat-sifatnya dan asumsi-asumsi yang harus dipenuhi. Berdasarkan kajian ini ditunjukkan bahwa distribusi variabel random yang dapat dinyatakan dalam bentuk kuadrat adalah chi-kuadrat non sentral. Juga ditunjukkan bahwa variabel random tersebut bersifat independen.

Kata Kunci: variabel random,bentuk kuadratik, distribusi chi-kuadrat non sentral, independensi

PENDAHULUAN

Pada analisis varians, total jumlah kuadrat terpartisi ke dalam komponen jumlah kuadrat, dimana rasionya mengarah pada statistik-F (di bawah kondisi distribusi tertentu) yang cocok untuk uji hipotesis tertentu. Secara umum pada model linear (Searle, 1987) khususnya untuk *unbalanced data* (data mempunyai banyak subklas tak sama), perlu diketahui jumlah kuadrat yang berkaitan dengan proses dalam bentuk kuadratik (*quadratic forms)*.

Bentuk kuadratik dari **y** adalah $y'Ay = \sum_i a_{ii} y_i^2 + \sum_{i \neq j} a_{ij} y_i y_j$ dengan **y** vektor random dan A matriks simetrik dari konstanta (Rencher, 2000).

Contoh:

Misal $y_1, y_2, \ldots, y_n$ sampel acak dari populasi dengan mean μ dan varians σ^2. Total jumlah kuadrat $\sum_{i=1}^{n} y_i^2$ terpartisi kedalam jumlah kuadrat *mean sample* $\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$ dan jumlah kuadrat keduanya adalah:

$$\sum_{i=1}^{n} y_i^2 = \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right) + n\bar{y}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 + n\bar{y}^2$$

Jika **b** matriks px1 maka $b'b = \sum_{i=1}^{p} b_i^2$. Dengan demikian bentuk $\sum_{i=1}^{n} y_i^2$ dapat dinyatakan sebagai bentuk kuadratik:

$$\sum_{i=1}^{n} y_i^2 = y'y = y'Iy \qquad (1)$$

dengan $y' = (y_1, y_2, \ldots, y_n)$.

Didefinisikan vektor $j=(1,1,...,1)'$, maka $\bar{y}$ dapat ditulis sebagai berikut.

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}j'y \text{ ....(ingat bahwa } a'j=j'a=\sum_{i=1}^{n}a_i\text{)}$$

Dengan demikian:

$$n\bar{y}^2=n\left(\tfrac{1}{n}j'y\right)^2=n\left(\tfrac{1}{n}j'y\right)\left(\tfrac{1}{n}j'y\right)=n\left(\frac{1}{n}\right)^2y'jj'y=n\left(\frac{1}{n}\right)^2y'Jy=y'\left(\frac{1}{n}J\right)y$$

Selanjutnya $\sum_{i=1}^{n}(y_i-\bar{y})^2$ dapat ditulis sebagai berikut.

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}\left(y_i^2-2y_i\bar{y}+\bar{y}^2\right)=\sum_{i=1}^{n}y_i^2-2\bar{y}\sum_{i=1}^{n}y_i+n\bar{y}^2=\sum_{i=1}^{n}y_i^2-2\bar{y}n\bar{y}+n\bar{y}^2$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2=y'Iy-y'\left(\frac{1}{n}J\right)y$$

Diperoleh

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=y'\left(I-\frac{1}{n}J\right)y \qquad (2)$$

Dengan demikian bentuk (1) dapat ditulis sebagai berikut.

$$y'Iy=y'\left(I-\frac{1}{n}J\right)y+y'\left(\frac{1}{n}J\right)y \qquad (3)$$

Matriks dari bentuk kuadratik pada (3) mempunyai sifat sebagai berikut.

1. $I=\left(I-\frac{1}{n}J\right)+\frac{1}{n}J$
2. I; $I-\frac{1}{n}J$ dan $\frac{1}{n}J$ idempotent (dalam tulisan ini, matriks idempotent diartikan sebagai matriks simetrik)
3. $\left(I-\frac{1}{n}J\right)\left(\frac{1}{n}J\right)=O$

Bukti ketiga sifat di atas sebagai berikut.

1. Jelas
2. (i) I.I = I maka I *idempotent* (terbukti)

   (ii) $(I-\frac{1}{n}J)^2=(I-\frac{1}{n}J)(I-\frac{1}{n}J)=I-\frac{2}{n}J+\frac{1}{n^2}J^2$

   dengan $J^2=\begin{pmatrix}1&1&\dots&1\\1&1&\dots&1\\\vdots&\vdots&\dots&\vdots\\1&1&\dots&1\end{pmatrix}\begin{pmatrix}1&1&\dots&1\\1&1&\dots&1\\\vdots&\vdots&\dots&\vdots\\1&1&\dots&1\end{pmatrix}=\begin{pmatrix}n&n&\dots&n\\n&n&\dots&n\\\vdots&\vdots&\dots&\vdots\\n&n&\dots&n\end{pmatrix}=nJ$

   Jadi $(I-\frac{1}{n}J)^2=I-\frac{2}{n}J+\frac{1}{n^2}nJ=I-\frac{1}{n}J$.

   Dengan demikian $I-\frac{1}{n}J$ idempotent.(terbukti)

   (iii) $\left(I-\frac{1}{n}J\right)\left(\frac{1}{n}J\right)=\frac{1}{n}J-\frac{1}{n^2}JJ=\frac{1}{n}J-\frac{1}{n^2}nJ=O$ (terbukti)

Dalam praktek, seringkali dihadapkan pada masalah pengambilan simpulan berdasar informasi atau data yang terbatas (sampel). Suatu pertanyaan yang muncul: sejauh manakah simpulan yang diambil dipercayai memiliki kualitas yang baik? Untuk dapat melakukan apresiasi terhadap hasil inferensi tersebut, diperlukan teori tentang distribusi peluang. Berdasar distribusi tersebut, semua pertanyaan tentang populasi dapat dijawab dengan baik. Persoalan yang muncul adalah bagaimana menentukan distribusi variabel random yang merupakan fungsi dari beberapa variabel random yang dapat dinyatakan dalam bentuk kuadra. Pada pemaparan berikut akan dikaji distribusi bentuk kuadratik beserta sifat-sifatnya dan asumsi-asumsi yang harus dipenuhi.

PEMBAHASAN

Mean dan Varians dari Bentuk Kuadrat

Sifat-sifat variabel random dapat dijelaskan melalui ekspektasi matematika, di antaranya adalah dengan memahami mean dan variansnya.

Teorema 1.

Jika y vektor random dengan mean μ dan matriks kovarians Σ dan jika A matriks simetrik dari konstanta maka

$$E[y' Ay] = tr(A\Sigma) + \mu' A\mu \qquad (4)$$

Catatan:

1. Jika $A_{nxn} = (a_{ij})$ maka trace dari matriks A dinotasikan dengan *tr(A)* adalah fungsi skalar yang didefinisikan sebagai jumlah diagonal elemen A maka $tr(A) = \sum_{i=1}^{n} a_{ii}$

   Contoh: Jika $A = \begin{pmatrix} 8 & 4 & 2 \\ 2 & -3 & 6 \\ 3 & 5 & 9 \end{pmatrix}$ maka tr(A) = 8-3+9=14

2. $\Sigma = E[(y-\mu)(y-\mu)'] = E[yy'] - \mu\mu'$ maka

$$E[yy'] = \Sigma + \mu\mu' \qquad (5)$$

Bukti Teorema 1 di atas sebagai berikut.

y'Ay skalar maka **y'Ay =** tr(**y'Ay**).

dengan demikian:

E[**y'Ay]= E[**tr(**y'Ay)] =** E[tr(**Ayy'**)] = tr(E(**Ayy'**)) = tr(**A**E(**yy'**))

E[**y'Ay**] = tr(**A**($\Sigma + \mu\mu'$)) = tr(**A** Σ **+ A** $\mu\mu'$ **)** = tr(**A** Σ **)** **+** tr($\mu' A\mu$)= tr(A Σ) + $\mu' A\mu$

Sebagai catatan, **y'Ay** bukan fungsi linier dari **y,** maka E[**y'Ay**] $\neq E(y')AE(y)$. Berikut ini diberikan contoh sebagai ilustrasi teorema 1 di atas. Misal diambil varians sample:

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} \qquad 6)$$

Pembilang pada persamaan (6) dapat ditulis sebagai berikut.

$\sum_{i=1}^{n}(y_i - \bar{y})^2 = y'\left(I - \frac{1}{n}\right)y$ ....berdasar (2),dengan $y = (y_1, y_2,...y_n)'$

Jika diasumsikan **y** independen, mempunyai distribusi dengan mean μ dan varians σ^2, maka: $E[y] = (\mu, \mu, \ldots, \mu)' = \mu j,\ dengan\ j = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ dan $\operatorname{cov}(y) = \sigma^2 I$

Misal $A = I - \frac{1}{n}J,\quad \Sigma = \sigma^2 I,\ dan\ \mu = \mu j$ maka berdasar (4)

$$E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = E\left[y'\left(I - \frac{1}{n}J\right)y\right]$$

$$= tr\left[\left(I - \frac{1}{n}J\right)(\sigma^2 I)\right] + \mu j'\left(I - \frac{1}{n}J\right)\mu j$$

$$= \sigma^2 tr\left(1 - \frac{1}{n}J\right) + \mu^2\left(j'j - \frac{1}{n}j'jj'j\right)$$

$$E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = \sigma^2\left(tr\ I - tr\left(\frac{1}{n}J\right)\right) + \mu^2\left(n - \frac{1}{n}n^2\right) = \sigma^2(n-1)$$

Dengan demikian:

$$E(s^2) = E\left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}\right] = \frac{1}{n-1}\sigma^2(n-1) = \sigma^2 \tag{7}$$

Pada Teorema 1 di atas, normalitas dari y tidak diasumsikan.

Teorema 2. (Searle, 1971)

Jika y berdistribusi normal p-variat dengan mean μ dan varians Σ, atau ditulis $y \sim N_p(\mu, \Sigma)$ maka

$$\operatorname{var}(y'Ay) = 2tr\left[(A\Sigma)^2\right] + 4\mu' A\Sigma A\mu \tag{8}$$

Bukti:

$\operatorname{var}(y'Ay) = E\left[(y'Ay)^2\right] - (E(y'Ay))^2 = E\left[(y'Ay)^2\right] - (tr(A\Sigma) + \mu'A\mu)^2$

Akan dicari dulu $E\left[(y'Ay)^2\right]$

Tulis: $y'Ay = (y-\mu)'A(y-\mu) + 2\mu'A(y-\mu) + \mu'A\mu$

Sebab:

$$\begin{aligned}
&(y-\mu)'A(y-\mu) + 2\mu'A(y-\mu) + \mu'A\mu \\
&= y'A(y-\mu) - \mu'A(y-\mu) + 2\mu'Ay - 2\mu'A\mu + \mu'A\mu \\
&= y'Ay - y'A\mu - \mu'Ay + \mu'A\mu + 2\mu'Ay - 2\mu'A\mu + \mu'A\mu \\
&= y'Ay - y'A\mu - \mu'Ay + 2\mu'Ay \\
&= y'Ay - y'A\mu + \mu'Ay \\
&= y'Ay - \mu'(y'A)' + \mu'Ay \\
&= y'Ay - \mu'Ay + \mu'Ay \\
&= y'Ay
\end{aligned}$$

Dengan demikian:

$$(y'Ay)^2 = \{(y-\mu)'A(y-\mu) + 2\mu'A(y-\mu) + \mu'A\mu\}^2$$
$$= [(y-\mu)'A(y-\mu)]^2 + 4[\mu'A(y-\mu)]^2 + (\mu'A\mu)^2$$
$$+ 2\mu'A\mu[(y-\mu)'A(y-\mu) + 2\mu'A(y-\mu)] + 4\mu'A(y-\mu)(y-\mu)'A(y-\mu)$$

Misal $z = y - \mu$ maka $E[z] = E(y-\mu) = E[y] - \mu$

Karena $y \sim N_p(\mu, \Sigma)$ maka $E[y] = \mu$. Akibatnya E[z] = 0

$$E\left[(y'Ay)^2\right] = E\left[(z'Az)^2\right] + 4E\left[(\mu'Az)^2\right] + (\mu'A\mu)^2 + 2\mu'A\mu E[z'Az + 2\mu'Az]$$
$$+ 4E[\mu'Azz'Az]$$
$$= E\left[(z'Az)^2\right] + 4E\left[(\mu'Az)^2\right] + (\mu'A\mu)^2 + 2\mu'A\mu + \Sigma trA + 4E[\mu'Azz'Az]$$

$$(zAz)^2 = \sum_i\sum_j\sum_k\sum_l a_{ij}a_{kl}z_iz_jz_kz_l$$

Karena z independen dengan 4 momen pertama maka:

$$E[z_iz_jz_kz_l] = \begin{cases} \mu_4; & (i = j = k = l) \\ \mu_2^2; & (i = j, k = l, i = k, j = l, i = l, j = k) \\ 0; & \text{untuk yang lain} \end{cases}$$

Sehingga $E\left[(y'Ay)^2\right] = \mu_4 \sum_i a_{ii}^2 + \mu_2^2\left(\sum_{i\neq k}\sum a_{ii}a_{kk} + \sum_{i\neq j}\sum a_{ij}^2 + \sum_{i\neq j}\sum a_{ij}a_{ji}\right)$

$$+ (\mu_4 - 3\mu_2^2)a'a + \mu_2^2(tr(A)) + 2trA^2$$

Karena $a_{ij} = a_{ji}$ dan $\sum_i\sum_j a_{ij}^2 = tr\, A^2$ maka $(\mu'Az)^2 = (b'z)^2 = \sum_i\sum_j b_ib_jz_iz_j$

dengan $b' = \mu'A$

dan $(\mu'Az)(z'az) = \sum_i\sum_j\sum_k b_ia_{jk}z_iz_jz_k$ maka:

$$E\left[(\mu'Az)^2\right] = \mu_2\sum_i b_i^2 = \mu_2 b'b = \mu_2\mu'A^2\mu \text{ dan}$$
$$E[(\mu'Az)(z'Az)] = \mu_3\sum_i b_ia_{ii} = \mu_3 b'a = \mu_3\mu'Aa$$

Dengan menggunakan $E[y'Ay] = tr(A\Sigma) + \mu'A\mu$ dan subtitusi ke

$$\text{var}(y'Ay) = E\left[(y'Ay)^2\right] - (E(y'Ay))^2$$

maka diperoleh: $\text{var}(y'Ay) = E\left[(y'Ay)^2\right] - (E(y'Ay))^2$

$$\text{var}(y'Ay) = (\mu_4 - 3\mu_2^2)a'a + 2\mu_2^2 tr\, A^2 + 4\mu_2\mu'A^2\mu + 4\mu_3\mu'Aa$$

Mengingat bahwa $y \sim N_p(\mu, \Sigma)$ maka $\mu_2 = \Sigma;\ \mu_3 = 0;\ dan\ \mu_4 = 3\mu_2^2$

Jadi $\text{var}(y'Ay) = (3\mu_2^2 - 3\mu_2^2)a'a + 2\Sigma^2\, tr\, A^2 + 4\Sigma\mu'A^2\mu = 2tr(A\Sigma)^2 + 4\mu'A\Sigma A\mu$

(terbukti)

Teorema 3. (Bukti lihat Graybill, 1983)

Jika y berdistribusi normal p-variat dengan mean μ dan varians Σ, atau ditulis $y \sim N_p(\mu, \Sigma)$ maka f.p.m dari **y'Ay** adalah :

$$M_{y'Ay}(t) = |I - 2tA\Sigma|^{-1/2} e^{-\frac{1}{2}\mu'\left[I-(I-2tA\Sigma)^{-1}\right]\Sigma^{-1}\mu} \tag{9}$$

Berikut ini akan ditentukan cov(**y,y'Ay**). Dalam hal ini **v**= **y'Ay** suatu variabel random skalar, maka cov(**y,v**) merupakan vektor kolom yang memuat kovarians setiap y_i dan **v**

$$\operatorname{cov}(y,v)=E[(y-E(y))(v-E(v))]=\begin{pmatrix}\sigma_{y_1v}\\ \sigma_{y_2v}\\ \vdots\\ \sigma_{y_pv}\end{pmatrix} \tag{10}$$

Teorema 4. (Searle, 1987)

Jika $y\sim N_p(\mu,\Sigma)$ maka

$$\operatorname{cov}(y,y'Ay)=2\Sigma A\mu \tag{11}$$

Bukti:

Berdasar (10) maka:

$$\operatorname{cov}(y,y'Ay)=E[(y-E(y))(y'Ay-E(y'Ay))]\ =E[(y-\mu)(y'Ay-tr(A\Sigma)-\mu'A\mu)]$$

Kita tulis $y'Ay-\mu'A\mu$ dalam bentuk $y-\mu$, yaitu:

$$y'Ay-\mu'A\mu=(y-\mu)'A(y-\mu)+2(y-\mu)'A\mu\ldots\ldots$$

Jadi diperoleh:

$$\begin{aligned}\operatorname{cov}(y,y'Ay)&=E[(y-\mu)((y-\mu)'A(y-\mu)+2(y-\mu)'A\mu)-tr(A\Sigma)] \qquad (12)\\ &=E[(y-\mu)(y-\mu)'A(y-\mu)]+2E[(y-\mu)(y-\mu)'A\mu]-E[(y-\mu)tr(A\Sigma)]\\ &=0+2\Sigma A\mu-0\end{aligned}$$

Dengan demikian cov (**y,y'Ay**)$=2\Sigma A\mu$ (terbukti)

Corollary 1:

Misal B_{kxp} matriks konstanta maka

$$\operatorname{cov}(By,y'Ay)=2B\Sigma A\mu \tag{13}$$

Teorema 5. (Seber, 1980)

Misal $v=\begin{pmatrix}y\\x\end{pmatrix}$ vektor random terpartisi dengan vektor mean dan matriks kovarians sebagai berikut.

$$\mu=E[v]=E\begin{bmatrix}y\\x\end{bmatrix}=\begin{pmatrix}E(y)\\E(x)\end{pmatrix}=\begin{pmatrix}\mu_y\\ \mu_x\end{pmatrix}\ \text{dan}\ \Sigma=\operatorname{cov}(v)=\operatorname{cov}\begin{pmatrix}y\\x\end{pmatrix}=\begin{pmatrix}\Sigma_{yy}&\Sigma_{yx}\\ \Sigma_{xy}&\Sigma_{xx}\end{pmatrix}$$

Dengan $y_{px1;}$ $x_{qx1;}$ $dan\ \Sigma_{yx}$ berordo pxq.

Misal A matriks konstanta berordo qxp, maka:

$$E[x'Ay]=tr(A\Sigma_{yx})+\mu_x'A\mu_x \tag{14}$$

Bukti: analog bukti Teorema 2.

Contoh:

Untuk mengestimasi kovarians populasi $\sigma_{xy}=E[(x-\mu_x)(y-\mu_y)]$

digunakan kovarians sampel:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \qquad (15)$$

dengan $(x_1, y_1),(x_2, y_2),....,(x_n, y_n)$ sampel random bivariat dari populasi dengan mean μ_x *dan* μ_y, varians σ_x^2 *dan* σ_y^2 dan kovarians σ_{xy}

Persamaan (15) dapat ditulis sebagai berikut.

$$s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n-1}$$

Dalam hal ini:

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}) \\ &= \sum x_i y_i - \bar{y}n\bar{x} - \bar{x}n\bar{y} + n\bar{x}\bar{y} \\ &= \sum x_i y_i - n\bar{x}\bar{y}\end{aligned}$$

Jadi $s_{xy} = \dfrac{x'\left(I - \frac{1}{n}J\right)y}{n-1}$ (16)

dengan $x = (x_1, x_2,..., x_n)'$ dan $y = (y_1, y_2,..., y_n)'$

$(x_i, y_i)\ dan\ (x_j, y_j)$ saling bebas (independen) untuk setiap $i \neq j$, maka vektor $v = \begin{pmatrix} y \\ x \end{pmatrix}$ mempunyai vektor mean dan matriks kovarians sebagai berikut.

$$E\begin{bmatrix} y \\ x \end{bmatrix} = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} = \begin{pmatrix} \mu_y j \\ \mu_x j \end{pmatrix} \text{dan cov}\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} = \begin{pmatrix} \sigma_y^2 I & \sigma_{xy} I \\ \sigma_{xy} I & \sigma_x^2 I \end{pmatrix} \text{dengan } I_{nxn}$$

Dipunyai $A = I - \frac{1}{n}J;\ \Sigma_{yx} = \sigma_{xy}I;\ \mu_x = \mu_x j;\ dan\ \mu_y = \mu_y j$

Dengan demikian:

$$\begin{aligned}E\left[x'\left(I - \frac{1}{n}J\right)y\right] &= tr\left(\left(I - \frac{1}{n}J\right)\sigma_{xy}I\right) + \mu_x j'\left(I - \frac{1}{n}J\right)\mu_y j \\ &= \sigma_{xy} tr\left(I - \frac{1}{n}J\right) + \mu_x\mu_y\left(j'j - \frac{1}{n}j'jj'j\right)\end{aligned}$$

$$E\left[x'\left(I - \frac{1}{n}J\right)y\right] = \sigma_{xy}(n-1)$$

Jadi $E(s_{xy}) = \dfrac{E\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]}{n-1} = \dfrac{(n-1)\sigma_{xy}}{n-1} = \sigma_{xy}$ (17)

Distribusi Chi-kuadrat Non-Sentral

Misal $z_1,z_2,...,z_n$ sampel random berdistribusi normal standar (N(0,1)). Karena z_i sampel random, maka $z_1,z_2,...,z_n$ independen dan setiap z_i berdistribusi N(0,1). Dengan kata lain $z = (z_1,z_2,...,z_n)'$ berdistribusi $N_n(0, I)$, berdasar definisi:

$$\sum_{i=1}^{n} z_i^2 = z'z \sim \chi^2(n) \tag{18}$$

(Amemiya & Anderson, 1990)

Teorema 6.

Jika u~ $\chi^2(n)$ maka (i) E(u) = n (19)

(ii) var(u) = 2n (20)

(iii) $M_u(t) = \dfrac{1}{(1-2t)^{n/2}}$ (21)

Misal $y_1, y_2, ..., y_n$ independen, berdistribusi $N(\mu_i, 1)$ maka $y \sim N_n(\mu, I)$ dengan $\mu = (\mu_1, \mu_2, ..., \mu_n)'$. Pada kasus ini $\sum_{i=1}^{n} y_i^2 = y'y$ tidak berdistribusi chi kuadrat, tetapi $\sum_{i=1}^{n} (y_i - \mu_i)^2 = (y-\mu)'(y-\mu) \sim \chi^2(n)$ karena $(y_i - \mu) \sim N(0,1)$.

Densitas dari v $= \sum_{i=1}^{n} y_i^2 = y'y$ dengan **y** independen berdistribusi $N(\mu_i, 1)$ dinamakan distribusi chi kuadrat non sentral dan dinotasikan dengan $\chi^2(n, \lambda)$ dengan parameter λ didefinisikan sebagai :

$$\lambda = \frac{1}{2}\sum_{i=1}^{n} \mu_i^2 = \frac{1}{2}\mu'\mu \tag{22}$$

Sebagai catatan mean dari v = $\sum_{i=1}^{n} y_i^2 = y'y$ lebih dari mean dari $u = \sum_{i=1}^{n} (y_i - \mu_i)^2$

$$E\left[\sum_{i=1}^{n} (y_i - \mu_i)^2\right] = \sum_{i=1}^{n} E\left[(y_i - \mu_i)^2\right] = \sum_{i=1}^{n} \mathrm{var}(y_i) = \sum_{n=1}^{n} 1 = n \text{ dan}$$

$$E[\sum_{i=1}^{n} y_i^2] = \sum_{i=1}^{n} E\left[y_i^2\right] = \sum_{i=1}^{n} (\sigma_i^2 + \mu_i^2) = \sum_{i=1}^{n} (1 + \mu_i^2) = n + \sum_{i=1}^{n} \mu_i^2 = n + 2\lambda \text{ .berdasar (22)}$$

Teorema 7. (Zhang, 1999)

Jika v~ $\chi^2(n, \lambda)$ maka : (i) E[v]=n+2 λ

(ii) var(v) = 2n +8 λ

(iii) $M_n(t) = \dfrac{1}{(1-2t)^{n/2}} e^{-\lambda[1-1/(1-2t)]}$

Corollary 1.

Jika $\lambda = 0$ berkorespondensi dengan $\mu_i = 0$ untuk setiap i, maka E[v], var(v) dan $M_n(t)$ pada Teorema 7 direduksi ke E[u], var(u) dan $M_n(t)$ untuk khi kuadrat sentral sebagaimana Teorema 6

Teorema 8. (Zhang, 1999)

Jika $v_1, v_2, ..., v_k$ independen berdistribusi $\chi^2(n_i, \lambda_i)$ maka

$$\sum_{i=1}^{k} v_i \sim \chi^2\left(\sum_{i=1}^{k} n_i, \sum_{1}^{k} \lambda_i\right) \tag{23}$$

Corollary 2.

Jika $u_1, u_2, ..., u_k$ independen berdistribusi $\chi^2(n_i)$ maka $\sum_{i=1}^{n} u_i \sim \chi^2\left(\sum_{i=1}^{k} n_i\right)$

Distribusi F dan t Non-Sentral

Distribusi F Non-sentral

Jika $u \sim \chi^2(p)\, dan\, v \sim \chi^2(q)$ serta u dan v independen maka berdasar definisi

$$w = \frac{u/p}{v/q} \sim F(p,q) \tag{24}$$

yang merupakan distribusi F (sentral) dengan derajat kebebasan p dan q.

Mean dari w adalah $E[w] = \dfrac{q}{q-2}$ ...(25)

Misal u berdistribusi chi- kuadrat non sentral, $\chi^2(p,\lambda)$ sedangkan v berdistribusi khi kuadrat sentral, $\chi^2(q)$, dengan u dan v independen maka:

$$z = \frac{u/p}{v/q} \sim F(p,q,\lambda) \tag{26}$$

merupakan distribusi F non sentral dengan parameter non sentral λ, dengan λ seperti parameter nonsentral pada distribusi u (chi kuadrat non sentral). Mean dari z adalah:

$$E[z] = \frac{q}{q-2}\left(1 + \frac{2\lambda}{p}\right) \tag{27}$$

Dalam hal ini E[z]>E[w].

Jika statistik F digunakan untuk uji hipotesis H_0, distribusi ini mempunyai tipikal sentral jika H_0 benar dan nonsentral jika H_0 salah. Dengan demikian distribusi F nonsentral sering digunakan untuk mengevaluasi kuasa dari uji F. Kuasa dari uji adalah peluang untuk menolak H_0 jika diberikan nilai λ. Jika F_α menyatakan titik persentase atas dari distribusi F sentral maka kuasa $P(p,q,\alpha,\lambda)$ dapat didefinisikan sebagai berikut.

$$P(p,q,\alpha,\lambda) = \mathrm{Pr}ob(z > F_\alpha) \tag{28}$$

Di mana z variabel random nonsentral seperti yang didefinisikan pada (26)

Distribusi t Non-sentral

Jika z~N(0,1) dan u $\sim \chi^2(p)$, z dan u independen maka berdasar definisi :

$$t = \frac{z}{\sqrt{u/p}} \sim t(p) \tag{29}$$

yaitu distribusi t (sentral) dengan derajat kebebasan p.

Misal $y \sim N(\mu,1)\, dan\, u \sim \chi^2(p)$, y dan u independen maka

$$t = \frac{y}{\sqrt{u/p}} \sim t(p,\mu) \tag{30}$$

yang merupakan distribusi t nonsentral dengan derajat bebas p dan parameter nonsentral μ. Jika $y \sim N(\mu,\sigma^2)$ maka parameter nonsentral menjadi μ/σ. Dengan demikian $y/\sigma \sim N(\mu/\sigma,1)$.

Distribusi dari Bentuk Kuadratik

Berdasar Teorema 6, jika $y \sim N_n(\mu, I)$ maka $(y-\mu)'(y-\mu) \sim \chi^2(n)$.
Dengan demikian: jika $y \sim N_n(\mu, \Sigma)$ maka

$$(y-\mu)'\Sigma^{-1}(y-\mu) \sim \chi^2(n) \tag{31}$$

Bukti:
Tulis $(y-\mu)'\Sigma^{-1}(y-\mu)$ dalam bentuk sebagai berikut.

$$\begin{aligned}(y-\mu)'\Sigma^{-1}(y-\mu) &= (y-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(y-\mu)\\ &= [\Sigma^{-1/2}(y-\mu)'][\Sigma^{-1/2}(y-\mu)]\\ &= z'z\end{aligned}$$

dengan $z = \Sigma^{-1/2}(y-\mu)$ dan $\Sigma^{-1/2} = (\Sigma^{1/2})$,vektor $z \sim N_n(O, I)$, maka berdasar definisi diperoleh: $z'z \sim \chi^2(n)$.

Teorema 9.
Misal $y \sim N_p(\mu, \Sigma)$, A matriks simetrik dari konstanta berordo pxp dengan rank r, $\lambda = \frac{1}{2}\mu' A\mu$ maka $y'Ay \sim \chi^2(r, \lambda)$ jika dan hanya jika $A\Sigma$ idempoten

Bukti:
Berdasar Teorema 3, f.p.m dari y;Ay adalah

$$M_{y'Ay}(t) = |I - 2tA\Sigma|^{-1/2} e^{-\frac{1}{2}\mu'\left[I-(I-2tA\Sigma)^{-1}\right]\Sigma^{-1\mu}}$$

Nilai eigen dari $I - 2tA\Sigma$ adalah $1 - 2t\lambda_i$; i=1,2,...,p dan λ_i adalah nilai eigen dari $A\Sigma$ .Jadi $|I - 2tA\Sigma| = \prod_{i=1}^{p}(1 - 2t\lambda_i)$.Dengan demikian $(I - 2tA\Sigma)^{-1} = I + \sum_{k=1}^{\infty}(2t)^k(A\Sigma)^k$ berlaku untuk $-1 < 2t\lambda_i < 1$ untuk semua i.

$$M_{y'Ay}(t) = \left(\prod_{i=1}^{p}(I - 2t\lambda_i)^{-1/2}\right)e^{-1/2\mu'\left[-\sum_{k=1}^{\infty}(2t)^k(A\Sigma)^k\right]\Sigma^{-1}\mu}$$

Ambil $A\Sigma$ idempoten dari rank r (rank dari A) maka r dari λ_i sama dengan 1, p-r dari λ_i sama dengan nol dan $(A\Sigma)^k = A\Sigma$. Dengan demikian :

$$\begin{aligned}M_{y'Ay}(t) &= \left(\prod_{i=1}^{r}(I - 2t)^{-1/2}\right)e^{-1/2\mu'\left[\sum_{k=1}^{\infty}(2t)^k\right]A\Sigma\Sigma^{-1}\mu}\\ &= (I-2t)^{-r/2}e^{-1/2\mu'\left[I-(I-2t)^{-1}\right]A\mu}\end{aligned}$$, untuk -1<2t<1 atau -1/2<t<1/2

Jadi $M_{y'Ay}(t) = \dfrac{1}{(I-2t)^{r/2}}e^{-1/2\mu' A\mu[I-I/(I-2t)]}$ yang merupakan f.p.m dari variabel random khi kuadrat nonsentral dengan derajat bebas r = rank(A) dan parameter nonsentral $\lambda = \frac{1}{2}\mu' A\mu$

Corollary 1.
Jika $y \sim N_p(0,1)$ maka $y'Ay \sim \chi^2(r)$ jika dan hanya jika A idempoten dari rank r

Corollary 2.

Jika $y \sim N_p(\mu, \sigma^2 I)$ maka $y'Ay/\sigma^2 \sim \chi^2(r, \mu' A\mu/2\sigma^2)$ jika dan hanya jika A idempoten dari rank r

Independesi dari Bentuk Linier dan Bentuk Kuadrat

Teorema 10:

Misal B_{kxp} matriks konstanta, A_{pxp} matriks simetrik dari konstanta dan $y \sim N_p(\mu, \Sigma)$ maka **By** dan **y'Ay** independen jika dan hanya jika $B\Sigma A = O$

Bukti:

$\Leftarrow$ **A**sumsikan A simetrik dan idempoten, maka $y'Ay = y'A'Ay = (Ay)'Ay$

Ambil $B\Sigma A = O$ maka berdasar (3.43) $B\Sigma A = \mathrm{cov}(By, Ay) = O$

Dengan demikian berdasar teorema 4.4C, maka **By** dan **Ay** independen.

Akibatnya **By** dan fungsi **(Ay)'(Ay)** juga independen.

$\Rightarrow$ Misal **By** dan **y'Ay** independen maka $\mathrm{cov}(By, y'Ay) = 0 \Leftrightarrow 2B\Sigma A\mu = 0$

Karena berlaku untuk μ yang mungkin, maka $B\Sigma A = O$

Dalam hal ini $B\Sigma A = O$ tidak berakibat $A\Sigma B = O$

Corollary 1.

Jika $y \sim N_p(\mu, \sigma^2 I)$ maka **By** dan **y'Ay** independen jika dan hanya jika **BA = O**

Teorema 11.

Jika **A** dan **B** matriks simetrik dari konstanta. Jika $y \sim N_p(\mu, \Sigma)$ maka **y'Ay** dan **y'By** independen jika dan hanya jika $A\Sigma B = O$

Bukti :

$\Leftarrow$ Ambil $A\Sigma B = O$. Asumsikan **A** dan **B** simetrik dan idempoten

$y'Ay = y'A'Ay = (Ay)'Ay$ dan $y'By = y'B'By = (By)'(By)$

Jika $A\Sigma B = O$ maka $A\Sigma B = \mathrm{cov}(Ay, By) = O$

Dengan demikian maka **Ay** dan **By** independen. Akibatnya fungsi

$y'Ay = y'A'Ay = (Ay)'Ay$ dan $y'By = y'B'By = (By)'(By)$ independen

$\Rightarrow$ Misal **A** dan **B** matriks simetrik dari konstanta. Misal $A\Sigma B = O$, maka

$(A\Sigma B)' = B\Sigma A = O$

Corollary 1.

Jika $y \sim N_p(\mu, \sigma^2 I)$ maka **y'Ay** dan **y'By** independen jika dan hanya jika **AB=O** (ekuivalen dengan **BA = O**)

Teorema 12. (bukti lihat Seber, 1980)

Misal $y \sim N_p(\mu, \sigma^2 I)$ dan A_i simetrik dari rank r_i untuk i=1,2,...,k dan

$$y'Ay = \sum_{i=1}^{k} y'A_i y \text{ dengan } A = \sum_{i=1}^{k} A_i \text{ maka:}$$

(i) $y'A_i y/\sigma^2 \sim \chi^2(r_i, \mu' A_i \mu/2\sigma^2)$, i = 1,2,...,k, dan

(ii) $y'A_i y \ dan \ y'A_j y$ independen untuk semua $i \neq j$, dan

(iii) $y'Ay/\sigma^2 \sim \chi^2(r, \mu' A\mu/2\sigma^2)$

jika dan hanya jika memenuhi sebarang dari dua pernyataan berikut.

(a) setiap A_i idempoten

(b) $A_iA_j = O$ untuk semua $i \neq j$

(c) $A = \sum_{i=1}^{k} A_i$ idempoten

atau jika dan hanya jika (c) dan (d) benar, dengan

(d) $r = \sum_{i=1}^{k} r_i$

Corollary 1.

Misal $y \sim N_p(\mu, \sigma^2 I)$ dan A_i simetrik dari rank r_i untuk i=1,2,...,k dan $y'y = \sum_{i=1}^{k} y'A_i y$ maka:

(1) setiap $y'A_i y \sim \chi^2(r_i, \mu' A_i \mu/2\sigma^2)$ dan

(2) y'A$_i$y independen

jika dan hanya jika memenuhi salah satu pernyataan berikut

(a) setiap A_i idempoten

(b) $A_iA_j = O$ untuk semua $i \neq j$

(c) $n = \sum_{i=1}^{k} r_i$

SIMPULAN

Berdasarkan uraian diatas dapat ditarik simpulan sebagai berikut. (1) Jika y berdistribusi normal n-variat dengan mean μ dan varians I maka $(y-\mu)'(y-\mu) \sim \chi^2(n)$. Dalam hal ini $y' = (y_1, y_2, ..., y_n)$. Secara umum jika jika $y \sim N_n(\mu, \Sigma)$ maka $(y-\mu)'\Sigma^{-1}(y-\mu) \sim \chi^2(n)$. (2) Bentuk kuadratik dari **y** adalah **y'Ay** dengan **y** vektor random dan A matriks simetrik dari konstanta. Dalam hal ini $y'Ay = \sum_i a_{ii} y_i^2 + \sum_{i \neq j} a_{ij} y_i y_j$. Jika $y \sim N_p(\mu, \Sigma)$, A matriks simetrik dari konstanta berordo pxp dengan rank r, $\lambda = \frac{1}{2}\mu' A\mu$ maka $y'Ay \sim \chi^2(r, \lambda)$ jika dan hanya jika $A\Sigma$ idempoten. (3) Jika B_{kxp} matriks konstanta, A_{pxp} matriks simetrik dari konstanta dan $y \sim N_p(\mu, \Sigma)$ maka **By** dan **y'Ay** independen jika dan hanya jika $B\Sigma A = O$

DAFTAR PUSTAKA

AmemiyaY.& Anderson T.W.1990. Asymtotic chi-squared test for a large class of factor analysis models. *Annals of Statistics*. **18**, 1453-1463.

Graybill F.A. 1983. *Matrices with Applications in Statistics.* 2nd.ed. Belmont, Calif: Wadsworth.

Rencher A.C. 2000. *Linear Models in Statistics.* New York: John Wiley & Sons.Inc

Searle S.R. 1971. *Linear Models.* New York: John Wiley & Sons.

Searle S.R. 1987. *Linear Models for Unbalanced Data*. New York: John Wiley & Sons.

Seber G.A.F.1980. *The Linear Hypothesis*.2nd.ed. London: Charles Griffin.

Zhang B. 1999. A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*. 86, 535-539.