KELAYAKAN FORMULA DAN DISTRIBUSI KOEFISIEN-KOEFISIEN REGRESI UNTUK PENELITIAN BIDANG BIOLOGI

Suhardi Djojoatmodjo

Fakultas Biologi Universitas Kristen Duta Wacana - Yogyakarta

Abstrak

Sudah diperlihatkan bagaimana χ^2 dapat digunakan dengan tepat untuk menguji kebaikan dan kesesuaian terhadap frekuensi-frekuensi. Dan model yang digunakan untuk penyelidikan kebaikan-kesesuaian adalah dengan garis-garis regresi. Kenyataannya bahwa : Ini adalah suatu permasalahan yang lebih rumit. Dalam penelitian ini diperlihatkan bahwa distribusi χ^2 hanya mensuplai suatu estimasi, dan dalam analisis variansi diambil distribusi *z* atau *F*. Sebelum aplikasi, pada umumnya dikenali lebih dahulu tentang distribusi *z* yang ternyata terus menguat secara tidak terduga. Selain daripada itu dalam penelitian ini juga dikaji temuan-temuan para peneliti pendahulu yang tertarik terhadap hubungan-hubungan dan penggunaan-penggunaannya berkaitan dengan distribusi tersebut. Temuan-temuan tersebut juga diperlakukan dalam penelitian ini sebagai distribusi χ^2 yang dimodifikasi. Diperlihatkan pula dalam penelitian ini bahwa : Metode tersebut diperluas ke regresi non-linier, dan interpretasi yang lebih akurat pada "rasio korelasi". Pada bagian lain sebagai perluasan dari bagian yang pertama dimunculkan data regresi. Diperlihatkan bahwa signifikansi koefisien-koefisien pada formulasi regresi linier atau non-linier, sederhana atau multiganda, juga yang bersifat bias diperlakukan secara tepat dengan "student's" *t-test*.

Kata kunci : uji kebaikan dan kesesuaian, distribusi χ^2, distribusi *z* atau *F,* regresi non-linier, multiganda, bias dan *t-test*

PENDAHULUAN

Dua tipe kuantitas yang mencirikan statistika modern, yakni *probable error* dan uji *kelayakan*. Uji kelayakan Pearson mudah diterapkan pada masalah distribusi frekuensi sebagai esensi justifikasi metode *a posterior* reduksi himpunan data. Model dalam penelitian ini adalah perluasan uji Slutsky dan Pearson yang berlaku terhadap kesesuaian formula regresi, rasio korelasi. Diperlihatkan oleh Fisher, 1922 : Kelayakan uji-t- χ^2 dapat diterapkan secara akurat hanya jika penilaian diberikan untuk jumlah konstanta yang sesuai dalam merekonstruksi populasi teoritis. Koreksi perlu diberikan untuk semua kasus, tetapi yang terpenting khususnya dalam tabel kontingensi, dan fakta kesalahan penggunaan uji kelayakan adalah memperbesar *P* yang diperoleh. Oleh karena itu perlu perluasan penyelidikan uji kelayakan terhadap garis regresi. Kesalahan yang muncul biasanya dikarenakan mengabaikan jumlah konstanta yang sesuai; disamping hal-hal yang muncul dalam penyelidikan sebelumnya. Kesalahan konsep juga terjadi dalam distribusi rasio korelasi η. Dalam hal ini juga disajikan penyelesaian permasalahan tentang distribusi koefisien-koefisien regresi umtuk sampel-sampel kecil. Hal terpenting yaitu kurva distribusi χ^2 kini tidak lagi bertipe III sebagai basis Elderton, melainkan tipe IV.

TINJAUAN PUSTAKA

Aplikasi Tabel Elderton. Misalkan untuk *x, y* sebarang; n_p : jumlah observasi untuk $x = x_p$, dan n_{pqi} untuk $y = y_q$; $\bar{y}_p$: mean nilai *y* terobservasi untuk *x* tertentu, sehingga $n_p \bar{y}_p = S_p(n_{pq} y_q)$. Misalkan *n* : grup sampel random dari populasi; ε : konstanta; *y* : variabel bebas terhadap mean m_p dengan deviasi standar σ_p. Sampel n_p memenuhi mean $\bar{y}_p$ independent terhadap jumlah array, dan m_p : mean $\bar{y}_p$ dari sampel random. Simak Fakta Pearson [1, hal.240]. Perbedaan prinsip ini terletak pada kesederhanaan hasil yang diperoleh senantiasa tidak mengeliminir kuantitas-kuantitas sampel yaitu formula distribusi statistika yang diselidiki, namun hanya terfokus pada kuantitas parameter populasi yang tidak diketahui. Kemudian, jika array populasinya normal untuk array-array ukuran tertentu serta n_p besar maka deviasi standar $\bar{y}_p$ = $\frac{\sigma_p}{\sqrt{n_p}}$ didistribusikan secara normal. Pearson menjelaskan bahwa nilai $\bar{y}_p$ untuk array-array ukuran berbeda tidak berdistribusi normal, tetapi leptokurtic bahkan untuk array-array yang cukup besar.. Lebih mudahnya jika array populasinya normal maka kuantitas $Z_p = \sqrt{n_p}(\bar{y}_p - m_p)$ terdistribusi secara normal sekitar nol, dengan deviasi standar σ_p, independent terhadap ukuran array.

Jika array populasi adalah variabel maka $\sigma_p = \sigma$ konstan, dan kuantitas $S(Z_p^2) = S\left\{ n_p(\bar{y}_p - m_p)^2 \right\}$ untuk array lainnya, yaitu jumlah kuadrat *a* kuantitas variabel independent, normal dan sama, dan konsekuensinya jika ditulis $\chi^2 \sigma^2 = S(Z_p^2)$ maka χ^2 didistribusikan sebagai ukuran kelayakan biasa. Dalam menerapkan tabel Elderton harus memberikan *n'* satu lebihnya dari jumlah derajad bebas, perhatikan [2]. Jika nilai m_p : *a* priori maka diambil *n' = a + 1*, agar formula regresi sesuai terhadap persamaan linear y_p maka direduksi jumlah derajad bebas dengan jumlah konstanta yang sesuai. Jadi, jika m_p : fungsi linear *x*, dan garis lurus bersesuaian maka *n' = a – 1*, dan nilai χ^2 sebagai uji apakah m_p berada dalam realita yang memenuhi syarat fungsi linier *x* atau tidak. Demikian pula, jika polinomial kubik dalam *x* bersesuaian maka *n' = a - 3*

TUJUAN DAN MANFAAT PENELITIAN

Tujuan dari penelitian ini adalah mencoba untuk memodifikasi distribusi χ^2. Manfaat dari penelitian ini adalah (1) dapat diperluas ke regresi non-linier, dan interpretasi yang lebih akurat pada "rasio korelasi". (2) dapat dilihat signifikansi koefisien-koefisien regresi linier atau non-linier, sederhana atau multiganda, juga yang bersifat bias diperlakukan secara tepat dengan "student's" *t-test*.

METODE PENELITIAN

4.1. σ penentu keeksakan distribusi χ^2. Jika σ ditentukan dari data maka distribusi χ^2 eksak, sehingga kesalahan yang muncul dari sumber ini mempengaruhi χ^2, dan σ dapat diestimasi dari data keseluruhan, yang dikenal dengan *akurasi* urutan lebih tinggi daripada kuantitas-kuantitas yang mensuport χ^2, untuk data yang tidak begitu banyak perlu mengestimasi simpangan. Dari [3] momen kedua sampel normal n_p didistribusikan sedemikian rupa sehingga frekuensi yang termuat dalam kisaran ds_p^2 adalah proporsional dengan $\sigma^{-(n_p - 1)}(s_p^2)^{-\frac{n_p s_p^2}{2\sigma^2}} d(s_p^2)$; kans di mana semua nilai s_p^2 terobservasi termuat dalam kisaran yang ditetapkan merupakan produk *a*-kuantitas, untuk semua distribusi in-dependent. Konsekuensinya, nilai σ-optimum atau nilai probabilitas kesalahan terkecil didapat dari perkalian maksimum $\sigma .s_p^2$n. Dengan mengambil logaritma dan diferensial-nya diperoleh : $\frac{\partial L}{\partial \sigma} = \frac{S(n_p^2 p^2)}{\sigma^3} - \frac{S(n_p - 1)}{\sigma}$. Jadi nilai σ^2 - optimum adalah s^2 di mana $(n\text{-}a)s^2 = S(n_p s_p^2)$. Jadi metode ini mengestimasi σ menjadi $\chi^2 = \frac{S(z_x^2)}{s^2}$; Lantas distribusi statistiknya ditentukan dari distribusi s^2 yang memiliki jenis sama dengan distribusi terkait. Untuk s^2 : jumlah kuadrat N-kuantitas variabel independent berlaku $S(n_p s_p^2) = S(y - \bar{y}_p)^2$. Dan $S_p(y) = n_p \bar{y}_p)^2$, untuk a terrestriksi linear. Oleh karena itu, jika distribusi s^2 sebagai fungsi elemen frekuensi ∂f dipandang dari sudut element varian maka berlaku $\partial f \varpi \propto t^{\frac{1}{2}(N...a-2)} e^{-\frac{t}{2\sigma^2}} \partial \tau;$ di mana t : $s^2(N - a)$. Analog jika τ : $\chi^2 s^2$ maka berlaku $\partial f \propto \tau^{\frac{1}{2}(a-p-3)} e^{-\frac{\tau}{2\sigma^2}} \partial \tau;$ dengan $p+1$ kostanta. Dan dua distribusi ini independent, yang satu tergantung pada deviasi mean sampel normal, dan lainnya pada meannya. Jadi χ^2 berdistribusi $(N - \alpha)\frac{\tau}{t}$, sehingga dengan men-distribusikan $\tau = \frac{\chi^2 t}{N - a}$ dalam $t^{\frac{N-a-2}{2}} \tau^{\frac{a-p-3}{2}} e^{-\frac{t}{2\sigma^2}(t+\tau)} \partial t \partial \tau$ dan mengabaikan konstanta diperoleh : $(\chi^2)^{\frac{a-p-3}{2}} t^{\frac{N-p-3}{2}} e^{-\frac{t}{2\sigma^2}(1+\frac{\chi^2}{N-a})} \partial t\, \partial \chi^2$, dengan mengambil integral terhadap t dari 0 ke ∞ diperoleh : $\chi^{2\frac{a-p-3}{2}}(1 + \frac{\chi^2}{N - a})^{-\frac{N-p-1}{2}} d\chi^2$, oleh karena itu variansi s^2 meng-ubah bentuk kurva distribusi eksak χ^2 dari Tipe III ke Tipe VI. Perubahan ini sangat nyata untuk N besar, yaitu $(1 + \frac{\chi^2}{N - a})^{-\frac{N-p-1}{2}} \xrightarrow{n >>} e^{-\frac{1}{2}\chi^2}$.

4.2. Sifat Estimasi Kurva Tipe VI. Diketahui : Kurva dengan persamaan

$$\partial f = \frac{(N - a)^{-\frac{a-p-2}{2}} . \frac{N - p - 3}{2}!}{\frac{N - a - 2}{2}! \frac{a - p - 3}{2}!} \chi^{\frac{a-p-3}{2}} (1 + \frac{\chi}{N - a})^{-\frac{N-p-1}{2}} \partial \chi$$

Terhadap tipe III : $\partial f = \frac{2^{-\frac{a-p-1}{2}}}{\frac{a - p - 3}{2}!} . \chi^{\frac{a-p-3}{2}} e^{-\frac{1}{2}\chi} \partial \chi$. Jika χ^2 << maka bentuk dua kurva sangat serupa, yang terakhir merupakan distribusi χ^2, sesuai tabel Elderton, dengan $n' = a - p$. Rasio ordinat-ordinat ujung kurva diperoleh dengan memperluas multiplier konstan kurva yang pertama dalam N^{-1}-kuantitas yaitu $1 + \frac{(n'-1)(n'-3)}{4N}$. Oleh karena itu, untuk nilai P tinggi maka dengan tabel Elderton *1-p* dikoreksi dengan mengalikan dengan faktor ini. Sehingga observasi posisi mean dan mode mendekati pusat kurva :

Tipe	Mean	Mode.
III	$a - p - 1$	$a - p$ -3
VI	$(a - p - 1)\frac{n - a}{N - a - 2}$	$(a - p - 1)\frac{n - a}{N - a + 2}$

Oleh karena itu, mean naik dan mode turun dengan proporsi yang sama. Untuk x yang lebih tinggi maka kurva tidak begitu serupa, dan untuk inilah terutama nilai P dibutuhkan, sehingga diperoleh koreksi P untuk sejumlah faktor dalam N^{-1}. Rasio ordinat-ordinatnya adalah:

$$1+\frac{1}{4N}\left\{ x^2-2(n'-1)x+(n'-1)(n'-3) \right\},\ \text{di mana}\ 2^{\frac{n'-1}{2}}.\frac{n'-3}{2}!\,P_{a'}(x)=\int_{\infty}^{\infty} x^{\frac{n'-3}{2}} e^{-\frac{1}{2}x}\,dx,$$

memiliki koreksi $\frac{1}{4N}\left\{(n'-1)(n'+1)P_{n'+4}-2(n'-1)^2P_{n'+2}+(n'-1)(n'-3)P_{n'}\right\}$

$$=\frac{n'-1}{4N}\left\{(n'+1)P_{n'+4}-2(n'-1)P_{n'+2}+(n'-3)P_{n'}\right\},$$

dengan asumsi tabel-tabel kurva Tipe VI memenuhi syarat.

4.3. Rasio Korelasi. Jika formula regresi Y : fungsi x, dan berlaku hubungan $NR^2{s_y}^2=S\left\{n_p(Y_p-\bar{y})^2\right\};\ N{s_y}^2=S(y-\bar{y})^2$ di mana $\bar{y}$: mean semua y terobservasi; maka $S\left\{n_p(\bar{y}_p-Y_p)^2\right\}$ minimum , asalkan Y sesuai dengan data sehingga untuk variasi-variasi proporsional $Y-\bar{y}$, berlaku $N(1-R^2){s_y}^2=S\left\{n_{pq}(y-Y_p)^2\right\}$; formula paralel rasio korelasi : $N(1-\eta^2){s_y}^2=SS\left\{n_{pq}(y-Y_p)^2\right\}=(N-a)s^2;$ dengan pengurangan didapat :

$$N(1-\eta^2){s_y}^2=S\left\{n_p(\bar{y}_p-Y_p)^2\right\}=\chi^2 s=\chi^2\frac{N}{N-a}(1-\eta^2){\sigma_y}^2;$$

atau $\chi^2=(N-a)\frac{\eta^2-R^2}{1-\eta^2}$; untuk uji signifikansi η^2-R^2 disubstitusikan ke tabel Elderton dengan $n'=a-p$, di mana $p+1$: jumlah konstanta yang sesuai dengan garis regresi. Dengan demikian, untuk formula regresi linear $\chi^2=(N-a)\frac{\eta^2-R^2}{1-\eta^2}$; dan $n'=a-1$ diperoleh korelasi untuk Tipe VI yang sesuai dengan sebelumnya. Bentuk eksak distribusi η itu sendiri sulit diperoleh, namun dalam praktiknya η digunakan untuk menguji validitas formula regresi linear atau lainnya. Untuk tujuan ini dibutuhkan variabel kuantitas $(\eta^2-R^2)(1-\eta^2)$ dan bukan distribusi η, dan ekspresi-ekspresi di atas diwakili kurva Tipe III, dan probabilitas kesesuaian yang lebih besar terjadi secara kebetulan diperoleh dari tabel Elderton.

4.4. Komparasi Formula. Metode Slutsky (4, hal. 83) memperkarakan homoscedastic, analog dengan capaian proses di atas metode ini menggunakan 4 deviasi: (1) Meng-estimasi σ^2 dengan mean deviasi standar array, bukan kuadrat-kuadratnya, (2) Membagi totalnya dengan N bukan $N-a$. (3) Menggunakan tabel Elderton dengan $n'=a+1$ dan $n'=a-p$. (4) Digunakan distribusi Tipe II sebagai distribusi eksak.

4.4.1 Keampuhan Pearson. [1, hal. 294-51) jika permukaan homocedastic maka diambil $s^2(1-\eta^2)$, jika regresi linear maka $1-\eta^2$ diganti $1-r^2$. Secara umum Pearson lebih tertarik menggunakan $1-r^2$ dalam memecahkan data Slutsky.

4.4.2 Keuntungan Estimasi. Keuntungan praktis yang diperoleh adalah : Keeratan kesesuaian kurva Tipe III yang sesuai dalam lingkungan median; sedangkan korelasi untuk P hanya merujuk pada bentuk χ^2 statistik itu sendiri. Jika array tidak terdefinisi berkenaan dengan distribusi y, maka diambil $\frac{1}{N-1}S(y-\bar{y})^2$ sebagai estimasi terbaik variansi keseluruhan observasi; artinya estimasi mean terhadap variansi setiap array ada-lah $\frac{1}{N-1}SS\{n_{pq}(y-\bar{y}_p)^2\}$. Jadi $(1-\eta^2)S(y-\bar{y})^2 = SS\{n_{pq}(y-\bar{y}_p)^2\}$ menunjukkan mean $(1-\eta^2) = \frac{N-a}{N-1}$, dan $\eta^2 = \frac{a-1}{N-1}$; Pearson membahas distribusi η dalam [5], yaitu sekalipun array tidak terdefinisi sama sekali, namun perlu $\bar{\eta}$ positif, karena untuk uji η yang berbeda secara signifikan dari nol bukan hanya perlu mengetahui kesalahan standar η, namun juga nilai mean yang bervariasi. Standar kesalahan η untuk array yang tidak terdiferensialkan sudah dievaluasi sebelumnya [6] pada $1/\sqrt{N}$, yaitu $\eta^2 = \frac{a-1}{N}$; disimpulkan mean $\eta = \sqrt{\frac{a-1}{N}}$. Dalam kasus $p = 0$, $R = 0$, garis regresi yang sesuai adalah $Y = \bar{y}$. Maka distribusi dari $(N-a)\frac{\eta^2}{1-\eta^2}$ adalah:

$$df = \frac{(N-a)^{-\frac{a-1}{2}} \cdot \frac{N-3}{2}!}{\frac{N-a-2}{2}!\,\frac{a-3}{2}!} x^{\frac{a-2}{2}} \left(1+\frac{x}{N-a}\right)^{-\frac{N-1}{2}} dx$$

; dalam kurva Tipe VI, untuk N yang besar distribusi η tidak mengacu ke arah normalitas sebagaimana diperkirakan Pearson, namun distribusi η^2 cenderung ke kurva Tipe III, yaitu : $\bar{\eta} = \frac{\frac{a-2}{2}!}{\frac{a-3}{2}!} \cdot \frac{\frac{N-3}{2}!}{\frac{N-2}{2}!} \propto \frac{\frac{a-2}{2}!}{\frac{a-3}{3}!}\sqrt{\frac{2}{N}}\left(1+\frac{3}{4N}\right)$; untuk nilai mean η dan η^2 ini sesuai dengan nilai sebelumnya. Jadi mean η^2 bersesuain dengan mean Pearson, hanya saja nilai akurat untuk mean dan standar deviasi berbeda. Pengaruh komparasi garis-garis ini, hanya terjadi jika jumlah arraynya besar, sehingga distribusi η jauh dari normal, dan signifikansi η terobservasi diuji dengan akurasi χ^2. Jika jumlah array besar didapat : $\sigma^2{}_n = \bar{\eta}^2 - \bar{\eta}^2 = \frac{1}{2N}\left(1-\frac{a}{N}\right)$ sebagai estimasi pertama, di mana faktor kedua diabaikan.

4.4.3 Perbedaan Metode. Dalam penelitian ini dilibatkan sejumlah konstanta dalam penyesuaian formula regresi, dijelaskan dalam [2] yang berbeda dengan metode Slutsky dan Pearson. Jika $\chi^2 = S\left\{\frac{n_p - \bar{n}_p}{\bar{n}_p}\right\}$; di mana $\bar{n}_p$: jumlah observasi terestimasi; n_p : jumlah terobservasi dalam sel, maka nilai *n'* dalam tabel Elderton bukan merupakan jumlah total sel, melainkan jumlah lebihnya dari jumlah nilai $n_p - \bar{n}_p$; yang dapat dispesifikasikan secara independent, yakni jika nilai $\bar{n}_p$ direkonstruksi dari data sampel maka (*n'-1*) adalah jumlah derajad bebas tersisa setelah rekonstruksi. Analog untuk garis regresi $\chi^2 = \frac{1}{s^2}S\left\{n_p(\bar{y}_p - Y_p)^2\right\}$ dan jika *a* : jumlah array, *n'-1 = a*, maka U_p ditetapkan melalui sampel secara independen. Jika Y_p : nilai formula regresi yang sesuai dengan sampel maka jumlah nilai $\bar{y}_p - Y_p$ yang dapat dispesifikasikan secara independen direduksi dengan jumlah konstanta yang sesuai. Misalnya, jika polynomial kubik yang bersesuaian maka jumlah derajad bebas adalah $a - 4$ sehingga $n' = a - 3$.

4.5. Distribusi Koefisien Regresi. Yang lazim dilakukan adalah mensimulasi jenis data lantas diuji kesesuaian garis-garis regresi. Dikatakan Pearson [1, hal.258] : Keterbatasan metode terhadap data merupakan salah satu defisiensi paling serius dalam metode statistik selama ini. Bahwa permasalahan yang bersifat obyektif sudah jelas dari kepercayaan di mana setelah dilihat hasilnya kesesuaian-kesesuaian yang sangat buruk akan ditolak segera, untuk jelasnya simak (9, Fisher, 1921). Untuk menelaah formula regresi linear sederhana $Y = a + b(x - \bar{x})$ dengan *a, b* : koefisien-koefisien yang ditentu-kan oleh persamaan $a = \bar{y}$ dan $b = \dfrac{S(y(x-\bar{x}))}{S(x-\bar{x})^2}$; dalam hal ini *a*, *b* : fungsi-fungsi orthogonal terhadap *x* terobservasi, dengan variasi sampling saling independen. "Student" [7] memperlihatkan adanya kemungkinan melacak kesalahan *a*; karena jika nilai x tertentu dan standar deviasi *y* adalah σ maka *a* berdistribusi normal, sehingga ${\sigma_a}^2 = \dfrac{\sigma^2}{n}$. Jadi jika α adalah populasi *a*, dan $\tau = \dfrac{\alpha - a}{\sigma}\sqrt{n}$, maka τ berdistribusi normal sekitar nol dengan standar deviasi σ. Jika σ^2 tidak diketahui maka estimasi terbaik dari sampel adalah $\chi^2 = (n-2)\dfrac{s^2}{\sigma^2}$, $\partial f = \dfrac{1}{\frac{n-4}{2}!}\left(\dfrac{\chi^2}{2}\right)^{\frac{n-4}{2}} e^{-\frac{1}{2}\chi^2}\partial\left(\dfrac{\chi^2}{2}\right)$. Distribusi 2 kuantititas *s* dan *a* independent; dari "Student" diperoleh distribusi kuantitas yang ditentukan dari sampel, yakni :

$$Z = \frac{\tau}{x} = \frac{(a-a)\sqrt{n}}{\sqrt{S(y-Y)^2}}\text{; Untuk } \partial f = \frac{1}{\frac{n-4}{2}!}\left(\frac{\chi^2}{2}\right)^{\frac{n-4}{2}} e^{-\frac{1}{2}\chi^2} d\left(\frac{\chi^2}{2}\right).\frac{1}{\sqrt{2\pi}} e^{-\frac{\chi^2\tau^2}{2}}.\chi.\partial z$$

$$= \frac{1}{\sqrt{\pi}}.\frac{1}{\frac{n-4}{2}!}.\left(\frac{\chi^2}{2}\right)^{\frac{n-3}{2}} e^{-\frac{1}{2}(1+z^2)} d\left(\frac{\chi^2}{2}\right).\partial z$$

dengan pengintegralan terhadap χ^2 dari $0 \to \infty$, diperoleh : $\dfrac{1}{\sqrt{\pi}}.\dfrac{\frac{n-3}{2}!}{\frac{n-4}{2}!}.\dfrac{dz}{(1+z^2)^{\frac{n-1}{2}}}$,

sedangkan kurva Tipe VII diperoleh melalui "Student", dengan *n* tereduksi, dan penyesuaian garis regresi derajad pertama. Demikian juga untuk *b*, jika $Z = \dfrac{(b-\beta)\sqrt{S(x-\bar{x})^2}}{\sqrt{S(y-Y)^2}}$ diperoleh ${\sigma_b}^2 = \dfrac{\sigma^2}{S(x-\bar{x})^2}$ yaitu sesuai distribusi sebelumnya, dengan β : nilai populasi koefisien regresi.

Argument diatas juga berlaku untuk garis-garis regresi yang melibatkan sejumlah koefisien. Jika persamaan regresi : $Y = a + b\chi_1 + c\chi_2 + \cdots + k\chi_p$, di mana $\chi_1, \chi_2, \cdots, \chi_p$: fungsi orthogonal terhadap *x* ter-observasi maka $S(\chi_a\chi_b) = 0$; dalam kasus terpenting jika χ_p : polynomial derajad *p* dalam *x*, ortogonal terhadap polynomials derajad lebih rendah maka menurut [9] diperoleh $k = \dfrac{S(y\chi_p)}{S(\chi_p)}$; dan jika $s^2 = \dfrac{1}{n-p-1}S(\eta - Y)^2$ maka $\sigma_k^2 = \dfrac{\sigma^2}{S(\chi_p)}$, distribusi *s* diberikan oleh $\partial f = \dfrac{1}{\frac{n-p-2}{2}!}\left(\dfrac{\partial^2}{2}\right)^{\frac{n-p-2}{2}} e^{-\frac{1}{2}\chi^2}\partial\left(\dfrac{\chi^2}{2}\right).e^{-\frac{1}{2}\chi^2}.\,d\left(\dfrac{1}{2}\chi^2\right)$; di mana $X^2 = (n-p-1)\dfrac{s^2}{\sigma^2}$.

Konsekuensinya distribusi $Z = \dfrac{(K-k)\sqrt{S(X_p)^2}}{\sqrt{S(y-Y)^2}}$ adalah kurva Tipe VII :

$\partial f = \dfrac{1}{\sqrt{\pi}}.\dfrac{\frac{n-p-2}{2}!}{\frac{n-p-3}{2}!}.\dfrac{\partial z}{\left(1+z^2\right)^{\frac{n-p}{2}}}$ dan dalam hal ini, jika $p+1$ konstanta sudah sesuai, maka semua koefisien regresi lainnya akan didistribusikan dengan cara serupa, hanya dengan mensubstitusikan fungsi x sesuai untuk χ_p.. "Student" [8] menyajikan Tabel integral distribusi probabilitas Tipe VII untuk nilai $n-p$ dari 0 ke 30. Tabel-tabel ini berada dalam bentuk yang sesuai untuk uji signifikansi koefisien regresi terobservasi. Untuk sampel-sampel yang lebih besar kurva akan cukup normal dengan variansi $z = \dfrac{1}{n-p-3}$. Kegunaan kurva "Student" untuk distribusi kesalahan dalam mean sampel. Selanjutnya dengan menetapkan $z = \dfrac{\bar{x}-\bar{x}'}{\sqrt{S(x-\bar{x})^2 + S(x'-\bar{x}')}}.\sqrt{\dfrac{nn'}{n+n'}}$ maka distribusi z ditentulan oleh

$$\partial f = \frac{1}{\sqrt{\pi}}.\frac{\frac{n+n'-3}{2}!}{\frac{n+n'-4}{2}!}.\frac{\partial z}{\left(1+z^2\right)^{\frac{n+n'-1}{2}}}$$

HASIL DAN PEMBAHASAN

Jika nilai x yang sama diobservasi dalam setiap kasus maka metode kompirasi dapat diterapkan secara langsung terhadap koefisien-koefisien regresi. Khusus untuk variabel tunggal dapat memilih faktor-faktor persamaan regresi, dengan beberapa faktor yang tidak berkorelasi. Sistem regresi yang digunakan adalah : $Y = b_1x_1 + b_2x_2 + ... + b_px_p$; di mana $x_1, x_2, \ldots, x_p$ nilai terobservasi; p : variabel independent dengan korelasi mutual tertentu. Akurasi koefisien-koefisien regresi hanya dipengaruhi oleh korelasi-korelasi yang muncul dalam sampel, sehingga jika dikonstruksi determinan

$$\Delta = \begin{bmatrix} S(x_1^2) & S(x_1x_2) & \cdots & S(x_1x_p) \\ S(x_1x_2) & S(x_2^2) & \cdots & S(x_2x_p) \\ \vdots & \vdots & \vdots & \vdots \\ S(x_1x_p) & S(x_2x_p) & \cdots & S(x_p^2) \end{bmatrix}$$

dari nilai sampel maka $\sigma_{b1}{}^2 = \dfrac{\sigma^2\Delta_n}{\Delta}$, di mana Δ_{11} : minor $S(x_1{}^2)$. Konsekuensinya, jika $z = \dfrac{(b_1-\beta_1)\sqrt{\Delta}}{\sqrt{S(y-Y)^2}\sqrt{\Delta_{11}}}$ maka $\partial f = \dfrac{1}{\sqrt{\pi}}.\dfrac{\frac{n-p-2}{2}!}{\frac{n-p-3}{2}!}.\dfrac{\partial z}{\left(1+z^2\right)^{\frac{n-p}{2}}}$ artinya z terdistribusi ke distribusi Tipe III

KESIMPULAN

1) Dalam proses uji kesesuaian harus diperhatikan jumlah derajad bebas yang diabsorbsi terhadap garis regresi yang berkaitan. 2) Distribusi Tipe III pada tabel Elderton tidak eksak untuk uji garis-garis regresi, namun tabel tersebut bisa digunakan sebagai basis estimasi yang bermanfaat. 3) Distribusi eksak χ^2 diberikan oleh kurva Tipe VI, di mana untuk sampel besar mendekati distribusi Tipe III. 4) Untuk array yang tak terdiferensialkan, distribusi η^2 diberikan oleh kurva Tipe I; untuk sampel besar kurva ini mendekati distribusi Tipe III. (5) Distribusi sampel random memuat koefisien regresi diperlakukan dengan metode "Student" untuk distribusi mean sampel normal, dan sebagaimana dalam kasus tersebut mengarah ke kurva distribusi Tipe VII, di mana untuk sampel-sampel besar dengan cepat mendekati normalitas. Bahwa sejumlah koefisien regresi mungkin dikalkulasi dengan aman dari sampel ukuran moderat.

DAFTAR PUSTAKA

[1] K. Pearson, 1916, *On The Application Of Goodness Of Fit Tables To Ujit Regresion Curves and Theritical Curves Used To Describe Observational or Experimental Data,* Biom, 239-61

[2] R.A Fisher, 1922, *On The significance of X^2 From Contingency Tables And On The Calculation Of P, J.R.S.S.,* LXXXV, pp. 87-94

[3] R.A Fisher, 1915, *Frequency Distributin Of The Values Of The Correlation Coefficient In Samples From an Indefinitely Large Population,* LXXVII, 78-84

[4] E. Slutsky, 1913, *On The Criterion Of Goodness Of Fit Of The Regression Lines and The Best Method Of Fitting Them To The Data,* J.R.S.S, LXXVII, 78-84

[5] K. Pearson, 1911, *On A Correction To Be Made To The Correlation Ratio,* Biom, VIII, 254-6

[6] K. Pearson, 1905, *On The General Theory Of Skew Correlation And Non-Linear Regression,* Drapers Company Research Menoirs: Dulau and Co

[7] Student, 1908, *The Probable Error Of a Mean,* Biom, VI, pp. 1-25

[8] Student, 1917, *Tables For Estimating The Probability That The Mean Of A Unique Sample Of Observations Lies Between - ∞ And Any Given Distance Of The Mean Of The Population From Which The Sample Is Drawn,* Biom, XI, 414-17

[9] R. A Fisher, 1921, *An Examination Of The Yield Of Dressed Grain From Broadbalk, Journal Of Agricultural Science*, XI, 107-35