PERBANDINGAN METODE *CLASSIC* DAN METODE *INVERSE* PADA REGRESI KALIBRASI

SONNY AFRIANSYAH

0303010397

UNIVERSITAS INDONESIA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

DEPARTEMEN MATEMATIKA

DEPOK

2009

PERBANDINGAN METODE *CLASSIC* DAN METODE *INVERSE*

PADA REGRESI KALIBRASI

Skripsi diajukan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Oleh:

SONNY AFRIANSYAH

0303010397

DEPOK

2009

SKRIPSI : PERBANDINGAN METODE *CLASSIC* DAN METODE *INVERSE* PADA REGRESI KALIBRASI

NAMA : SONNY AFRIANSYAH

NPM : 0303010397

SKRIPSI INI TELAH DIPERIKSA DAN DISETUJUI

DEPOK, 26 JUNI 2009

DRA. TITIN SISWANTINING, DEA	DRA. SASKYA MARY, M.Si
PEMBIMBING I	PEMBIMBING II

Tanggal Lulus Ujian Sidang Sarjana : 9 Juli 2009

Penguji I : Dra. Titin Siswantining Hudiyono, DEA

Penguji II : Dra. Rianti Setiadi, M.Si

Penguji III : Helen Burhan, S.Si, M.Si

" Dan bahwasanya seorang manusia tiada memperoleh selain apa yang telah diusahakannya. Dan bahwasanya usahanya itu kelak akan diperlihatkan kepadanya. Kemudian akan diberikan balasan kepadanya dengan balasan yang paling sempurna. Dan bahwasanya Tuhan-mulah kesudahan segala sesuatu… "

(QS. Bintang [53]: 39-42)

…Faidza azamta fa tawakkal 'alallah…

Dedicated to my Lovely Parents :

…Mom and Dad…

…The reason for my existence…

KATA PENGANTAR

Alhamdulillahirabbil'alamin penulis panjatkan kehadirat Allah SWT, karena dengan limpahan rahmat, taufiq, dan hidayah-Nya penulis dapat menyelesaikan penulisan skripsi yang berjudul " **Perbandingan Metode *Classic* dan Metode *Inverse* pada Regresi Kalibrasi** " sesuai dengan waktu yang diharapkan.

Penulisan skripsi ini bertujuan untuk memenuhi salah satu syarat mencapai gelar sarjana dalam kurikulum program S-1 di Departemen Matematika FMIPA Universitas Indonesia Depok.

Penulis menyadari bahwa penulisan skripsi ini tidak akan terselesaikan tanpa bantuan materi, bahan-bahan penulisan, maupun dukungan moril dari semua pihak. Oleh karena itu, pada kesempatan ini perkenankan penulis untuk mengucapkan rasa terima kasih kepada :

- Kedua orang tua penulis yang telah memberikan segalanya yang terbaik, yang tak ternilai harganya kepada penulis. ("I Love U Mom, I Love U Mom, I Love U Mom, and I Love U Dad... Ur the I best ever had...")
- Saudara kandung penulis, Donny JunianSyah Hidayat, yang telah berbagi segalanya dengan penulis. ("Love U Bro..., D Idon segera susul Mas Sonny, selesaikan TA-nya yah...")

- Yang penulis sangat hormati Ibu Dra. Titin Siswantining Hudiyono, DEA selaku pembimbing I dan Ibu Dra. Saskya Mary Soemartojo, M.Si selaku pembimbing II. (”Terima kasih banyak atas bimbingannya Bu, Sonny sangat bersyukur dapat dibimbing oleh Ibu... Ibu sudah seperti orang tua Sonny sendiri... Maafin kekurangan dan kekhilafan Sonny yah Bu... Terima Kasih...”)
- Bapak Gatot F. Hertono, PhD dan Ibu Bevina D. Handari, PhD selaku pembimbing akademik penulis.
- Seluruh dosen Departemen Matematika FMIPA UI, khususnya Almarhum Pak Ponidi, Pak Yudi, Bu Rianti, Bu Netty, dan Mba Helen.
- Seluruh karyawan Departemen Matematika FMIPA UI, khususnya Mba Santi dan Mba Rusmi.
- Rekan-rekan seangkatan penulis, angkatan 2003, khususnya The Last Battalion: Guntoro, Ilham, Diky, Bembi, Gilang, Gunung dan Yessa (”Finally it’s done guys... Congratz...”). Dody, Adri, Arif, dan Tony. (”It’s really hard for me to realize the fact, but I know that U already try to do the best, forgive me and wish U all the best...”)
- Angkatan 97, 2000, 2001, 2002, khususnya Teh Mira Purnamadewi, Bang Ruri, Bang Danu, Mas Rahmat, Mas Adri, Arif, Fuad, dan Iif.

- Angkatan 2004, 2005, 2006, khususnya Harry, Spina, Rani, Miranti, Shinta my lovely sister (”thanks for everything, pardons me for anything...”), Big Best Pals: Aliman, Budi, Rendi, Oza (”jiexpo yuks...”), Gorgeous Rangers: Lee, Rita, Syafirah, Yuri (”Keep united...”)
- Pondok Abadi Society, rumah kedua penulis yang penuh dinamika dan realita hidup, khususnya untuk mahagurudosen H. Erwin Faizal CFP-CFA candidate, Angga Satria Perdana dan Hekal Muin Bajenet. (”Great society with passion and desire...”)
- Bapak Muhammad Syarif Surbakti, SE, Ak., M.Si selaku direktur MUSYS Islamics Financial Consultant yang telah memberikan pengalaman dan pengetahuan selama di Bank Muamalat. (”Jazakallahu Khoiron Katsiro...”)
- Bapak Ir. Parto Kawito, MM, CFP selaku owner Lembaga Pendidikan Sekuritas BINA INSAN dan owner Infovesta Utama yang telah memberikan another path opportunity in careers. (” Terima kasih atas *view*-nya Pak, saya sudah berhasil melewati fund manager representative exam dan broker-dealer representative exam... Now I willing to pass underwriter representative exam... Wish me luck... ”)
- Rekan-rekan di Ikatan Pialang Efek Indonesia, dan rekan-rekan di Asosiasi Analis Efek Indonesia khususnya Mba Viet serta Pak Haryajid Ramelan, CFP, RFC selaku ketua asosiasi. (”Thanks for adding me as a member and for the workshop...”)

- Rekan-rekan di Diamond or Die Unicore, khususnya Pak Yimmy Yoperson, dr. Andhika, Pak Aprizal, Fatahilah, Alvin, Barry, Yugo, dan semua grup penulis. ("Ur core leader will be back someday...")
- Rekan-rekan di Street Bass Club, khususnya Wisnu Pamungkas my fabulous Bass teacher, Bang Franky Sadikin, Pak'de Arya Setiadi dan best buddies di Banana Suite Club, khususnya Sahaja, Rendra-Anggi, Aryo, Rony, Ninie, Desy, dan Putie, serta Team Futsal X_1-9_47.
- Idvan Utama dan Mirza. ("Our single 'jangan kecewakanku' is hits...")
- Sabina Yusuf, Siti Fatimah Az-zahra, Silvi Fransisca, Siti Havitha, Maria Ulfa, Nurina K. Dewi, Dian Nirmala, Wiwie, Ika Euistatikawati, Dezty Nuraeny, Rianti K. Dewi, Bonny R. Gabline, and the last standing Rila Fitria. ("Thank U for the friendship U have offered, rest assured that I'll treasure it in my heart and will always remember that once in my life, I've known someone like U... Love is sweet, love needs sweat, but sweat is not sweet, and sweet is reached with sweat…")
- Semua pihak yang tidak disebutkan penulis. ("Graven in my mind...")

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari sempurna, oleh karena itu penulis juga menyampaikan permohonan maaf apabila terdapat kekurangan. Akhir kata semoga skripsi ini bermanfaat bagi kita semua.

Depok, Juli 2009

Sonny Syah

ABSTRAK

Dalam tugas akhir ini akan dibahas suatu metode pada regresi kalibrasi yang berguna untuk estimasi nilai aktual dan interval kepercayaan nilai aktual dari nilai observasi yang diberikan. Berdasarkan Vocabulary Of International Metrology (VIM), kalibrasi merupakan serangkaian kegiatan yang membentuk hubungan antara nilai yang ditunjukkan oleh instrumen (alat) ukur atau nilai observasi, dengan nilai-nilai yang sudah diketahui atau nilai aktual yang berkaitan dengan besaran yang diukur dalam kondisi tertentu. Metode yang digunakan untuk menaksir nilai aktual adalah metode *classic* dan metode *inverse*, sedangkan metode yang digunakan untuk menaksir interval kepercayaan nilai aktual adalah metode Graybill. Data yang digunakan dalam penelitian ini adalah data kalibrasi alat pengukur suhu yaitu thermocouple yang berasal dari buku Montgomery (2001). Hasil analisis data menunjukkan bahwa hasil taksiran nilai aktual pada metode *inverse* lebih dekat dengan rata-rata nilai aktual bila dibandingkan dengan hasil taksiran nilai aktual yang diperoleh pada metode *classic*.

Kata kunci : *Kalibrasi, Metode Classic, Metode Graybill, Metode Inverse, Regresi Kalibrasi*

x + 74 hlm.; gbr.; tab.; lamp.

Bibliografi: 10 (1967 - 2006)

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Dalam kehidupan sehari-hari, seringkali tertarik untuk mengetahui seberapa akurat hasil pengukuran suatu alat ukur. Seringkali muncul dalam pemikiran, apakah kilogram yang ditunjukkan pada timbangan merepresentasikan berat sebenarnya dari bahan yang ditimbang, atau apakah liter yang ditunjukkan pada pom-meter SPBU merepresentasikan volume bahan bakar sebenarnya dari bahan bakar yang dikeluarkan pom-meter SPBU tersebut. Bagaimanakah cara mengetahuinya, apakah dengan mengukur objek pengukuran secara langsung, atau dengan memastikan alat ukur sesuai dengan standar ukur baik nasional maupun internasional.

Dalam penelitian juga berlaku hal yang sama , misalnya dalam bidang kimia, dimana pengukuran analitik memiliki peranan yang sangat penting. Tujuan dari pengukuran analitik ini adalah untuk menentukan nilai sebenarnya atau nilai aktual dari suatu parameter kuantitas kimia yang terbaca pada suatu alat ukur, contohnya seperti konsentrasi, suhu, pH, dan lain-lain. Apabila hasil pengukuran tidak akurat akan mengakibatkan hasil penelitian yang tidak akurat juga.

Masalah yang muncul pada bahasan tersebut adalah masalah kalibrasi. Berdasarkan Vocabulary Of International Metrology (VIM), kalibrasi merupakan serangkaian kegiatan yang membentuk hubungan antara nilai yang ditunjukan oleh instrumen (alat) ukur, atau sistem pengukuran, atau nilai yang diwakili oleh bahan ukur, dengan nilai-nilai yang sudah diketahui yang berkaitan dari besaran yang diukur dalam kondisi tertentu. Dengan perkataan lain kalibrasi adalah kegiatan untuk menentukan kebenaran konvensional nilai penunjukkan alat ukur dan bahan ukur dengan cara membandingkan terhadap standar ukur yang mampu telusur (traceable) ke standar nasional maupun internasional.

Alat ukur yang digunakan harus terkalibrasi secara rutin yang disertai pernyataan traceable (mampu telusur) untuk tercapainya ketertelusuran atau keakuratan pengukuran sebagai salah satu tolok ukur jaminan mutu suatu hasil penelitian.

Kalibrasi di Indonesia terbagi dua, yaitu :

Kalibrasi Teknis

- Kalibrasi peralatan ukur yang tidak berhubungan langsung dengan dunia perdagangan.
- Dilakukan oleh laboratorium kalibrasi terakreditasi KAN (Komite Akreditasi Nasional) menurut standar ISO/IEC 17025.
- Misalnya kalibrasi thermometer, kalibrasi pH-meter, dan lain-lain.

Kalibrasi Legal (Tera)

- Kalibrasi peralatan ukur untuk keperluan perdagangan.
- Dilakukan oleh direktorat metrologi (BMKG) – departemen perdagangan.
- Misalnya peneraan pom-meter SPBU, peneraan argometer taksi, peneraan timbangan dagang, dan lain-lain.

Tujuan kalibrasi, yaitu :

- Traceable (mampu telusur) ke standar nasional maupun internasional.
- Mendukung sistem mutu yang diterapkan di berbagai industri pada peralatan laboratorium dan produksi yang dimiliki.
- Dapat diketahui seberapa jauh perbedaan atau penyimpangan antara nilai sebenarnya atau nilai aktual dengan nilai observasi yang ditunjukkan oleh alat ukur.

Waktu kalibrasi atau kapan diperlukan kalibrasi :

- Setiap perangkat ukur baru dan secara periodik.
- Ketika suatu perangkat ukur mengalami tumbukan atau getaran yang berpotensi mengubah kalibrasi.
- Ketika hasil observasi alat pengukur dipertanyakan keakuratannya.

Dalam penelitian atau tugas akhir ini, akan dilakukan regresi kalibrasi untuk estimasi nilai sebenarnya atau nilai aktual dari nilai observasi dengan metode *classic* dan metode *inverse* serta estimasi interval kepercayaan berdasarkan metode Graybill.

Data kalibrasi yang digunakan dalam tugas akhir ini adalah kalibrasi pada thermocouple, sehingga kesalahan indikasi atau koreksi dapat ditentukan dan disesuaikan agar menunjukkan temperatur yang sebenarnya dalam derajat celcius pada titik-titik tertentu pada skala thermocouple.

1.2 PERUMUSAN MASALAH

1. Bagaimana estimasi nilai aktual dalam masalah regresi kalibrasi dengan metode *classic* dan metode *inverse*.
2. Estimasi interval kepercayaan nilai aktual pada metode *classic* dan metode *inverse*.
3. Membandingkan metode *classic* dan metode *inverse* berdasarkan hasil estimasi nilai aktual kedua metode tersebut.

1.3 PEMBATASAN MASALAH

1. Regresi kalibrasi yang digunakan dibatasi pada hubungan linear antara nilai aktual dan nilai observasi atau hasil pengukuran suatu alat ukur (Thermocouple).
2. Estimasi nilai aktual dengan metode *classic* dan metode *inverse*.
3. Estimasi interval kepercayan nilai aktual dengan metode Graybill.
4. Membandingkan nilai aktual yang diperoleh melalui metode *classic* dan metode *inverse* berdasarkan kriteria rasio.

1.4 TUJUAN PENELITIAN

1. Mempelajari regresi kalibrasi dengan metode *classic* dan metode *inverse*.
   A. Estimasi nilai aktual dalam masalah kalibrasi.
   B. Estimasi interval kepercayaan nilai aktual.
2. Membandingkan kedua metode estimasi, yaitu metode *classic* dan metode *inverse*.

1.5 SISTEMATIKA PENULISAN

Penulisan pada tugas akhir ini dibagi menjadi lima bab, yaitu :

BAB I Berisikan latar belakang, perumusan masalah, pembatasan masalah, tujuan penulisan, dan sistematika penulisan.

BAB II Membahas landasan teori yang dipakai dalam estimasi nilai aktual dari nilai observasi dan estimasi interval kepercayaan nilai aktual.

BAB III Membahas mengenai regresi kalibrasi. Perumusan regresi kalibrasi, metode estimasi nilai aktual, metode estimasi interval kepercayaan nilai aktual, dan perbandingan.

BAB IV Membahas analisis data.

BAB V Berisikan kesimpulan dan saran untuk tugas akhir ini.

BAB II

LANDASAN TEORI

Pada bab ini akan dijelaskan tentang landasan teori yang mendukung dalam estimasi nilai aktual dengan metode *classic* dan metode *inverse*, estimasi interval kepercayaan nilai aktual, dan perbandingan metode *classic* dengan metode *inverse* pada regresi kalibrasi. Yaitu : peubah acak, ekspektasi, variansi, covariansi, m.g.f. , estimator tidak bias, regresi linear sederhana, koefisien korelasi, koefisien determinasi, outliers, m.l.e., distribusi chi-square, distribusi normal, distribusi t, dan distribusi F.

2.1 PEUBAH ACAK

2.1.1 Definisi

Peubah acak X adalah suatu fungsi yang memetakan setiap elemen dalam ruang sampel, $c \in \mathbb{C}$, pada satu dan hanya satu bilangan riil sedemikian sehingga $X(c) = x$, dengan $x \in \mathbb{R}$. Range dari X adalah bilangan riil $A = \{x : x = X(c), c \in \mathbb{C}\}$, dengan $\mathbb{C}$ adalah ruang sampel dari eksperimen acak.

2.1.2 Peubah acak dan fungsi probabilitas

Misalkan X adalah peubah acak dengan *one-dimensional space* $\mathcal{A}$, yang terdiri dari suatu interval atau gabungan interval. Misalkan f(x) tidak negatif sedemikian sehingga

$$\int_{\mathcal{A}} f(x)dx = 1$$

dimana probabilitas himpunan fungsi P(A), $A \subset \mathcal{A}$, dapat dituliskan dalam f(x) dengan

$$P(A) = \Pr(X \in A) = \int_A f(x)dx$$

maka X disebut dengan peubah acak kontinu, sementara f(x) disebut dengan *probability density function (p.d.f.)* dari X.

2.2 EKSPEKTASI (E)

2.2.1 Definisi

Misalkan X adalah peubah acak dan memiliki p.d.f. f(x) sedemikian sehingga memiliki konvergen absolut.

pada kasus X peubah acak diskrit :

$\sum_x |x| f(x)$, limitnya konvergen menuju nilai tertentu.

maka $E(X) = \sum_x |x| f(x)$.

sedangkan pada kasus X peubah acak kontinu :

$\int_{-\infty}^{\infty} |x| f(x)$, limitnya konvergen menuju nilai tertentu.

maka $E(X) = \int_{-\infty}^{\infty} |x| f(x)$

2.2.2 Sifat

a. E(aX)=aE(X), dengan a adalah sebuah konstanta.

Bukti : a. E(aX)= $\sum_{x} a|x|f(x)$, definisi ekspektasi untuk peubah acak diskrit

$= a\sum_{x} |x| f(x)$, sifat sumasi

$= aE(X)$

b. E(aX)= $\int_{-\infty}^{\infty} a|x|f(x)$, definisi ekspektasi untuk peubah acak kontinu

$= a\int_{-\infty}^{\infty} a|x| f(x)$, sifat integrasi

$= aE(X)$

b. $E(aX + bY) = aE(X) + bE(Y)$

2.3 VARIANSI (σ^2)

2.3.1 Definisi

$$\sigma^2(X) = E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2]$$

karena E adalah operator linear, maka

$$\sigma^2(X) = E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2$$

2.3.2 Sifat

a. $\sigma^2(aX) = a^2\sigma^2(X)$, dengan a adalah sebuah konstanta

Bukti : $\sigma^2(aX) = E([aX - a\mu]^2)$, definisi variansi

$= E([a(X-\mu)]^2)$, sifat bilangan riil

$= E(a^2(X-\mu)^2)$, sifat pangkat

$= a^2E(X-\mu)^2$, sifat ekspektasi (2.2.2 a)

$= a^2\sigma^2(X)$

b. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

c. $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

2.4. COVARIANSI

Covariansi merupakan ukuran kekuatan korelasi antara dua peubah acak atau lebih.

2.4.1 Definisi

Misalkan X dan Y masing-masing adalah peubah acak, dengan $E(X)=\mu_1$ dan $E(Y)=\mu_2$, serta $Var(X)=\sigma_1^2$ dan $Var(Y)=\sigma_2^2$ maka

$$\begin{aligned} Cov(X,Y) &= E\left[(X-\mu_1)(Y-\mu_2)\right] = E(XY-\mu_2X-\mu_1Y+\mu_1\mu_2) \\ &= E(XY)-\mu_2E(X)-\mu_1E(Y)+\mu_1\mu_2 \\ &= E(XY)-\mu_1\mu_2 \end{aligned}$$

2.4.2 Sifat

a. $Cov(X,Y)=0$

artinya peubah acak X dan peubah acak Y tidak berkorelasi.

b. $Cov(X,X)=E(X^2)-\mu_X^2=\sigma^2(X)$

2.5 *MOMENT-GENERATING FUNCTION* (M.G.F.)

2.5.1 Definisi

Misalkan ada bilangan positif h dengan $-h<t<h$ dimana ada ekspektasi matematika $E(e^{tX})$, maka

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \text{, jika X peubah acak kontinu, atau}$$

$$E(e^{tX}) = \sum_{x} e^{tx} f(x) \text{, jika X peubah acak diskrit}$$

Ekspektasi tersebut dinamakan *moment-generating function (m.g.f.)* dari X dan dilambangkan dengan M(t). Jadi M(t)= $E(e^{tX})$.

2.6 ESTIMATOR TIDAK BIAS

2.6.1 Definisi

Suatu statistik dengan ekspektasi matematisnya sama dengan parameter $\hat{\theta}$ yang diestimasinya dikatakan sebagai estimator tidak bias dari parameter θ, yaitu $E(\hat{\theta}) = \theta$.

2.7 REGRESI LINEAR SEDERHANA

Regresi linear adalah salah satu metode analisis statistik untuk mencari hubungan antara dua peubah, dimana kedua peubah tersebut adalah X yang merupakan peubah bebas atau predictor, dan Y yang merupakan peubah terikat atau respons.

Dalam regresi linear sederhana, banyaknya peubah yang digunakan adalah satu peubah X dan satu peubah Y yang diasumsikan memiliki hubungan linier. Model pada regresi linier sederhana adalah :

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Model tersebut dinamakan juga model regresi linier populasi dengan β_0 (intersep) dan β_1 (slope) adalah parameter yang tidak diketahui atau koefisien regresi, sementara ε adalah bagian *random error.*

Terdapat beberapa asumsi dalam regresi linier sederhana tersebut, antara lain adalah :

1. Errornya memiliki nilai harapan atau ekspektasi nol, $E(\varepsilon) = 0$.
2. Errornya memiliki variansi berharga sama atau konstan, $Var(\varepsilon) = \sigma^2$.
3. Errornya saling bebas.
4. Errornya berdistribusi normal.

Dalam menaksir atau mengestimasi koefisien regresi akan digunakan n data berpasangan. Metode yang digunakan adalah metode estimasi *least squares.* Model yang didapat untuk n data berpasangan adalah :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Model tersebut dinamakan juga dengan model regresi linier sampel dengan $i = 1, 2, ..., n$, dan diasumsikan errornya memenuhi asumsi regresi linier sederhana yaitu *normally and independently distributed* dengan $E(\varepsilon) = 0$ dan $Var(\varepsilon) = \sigma^2$ atau dengan perkataan lain $\varepsilon \sim NID(0, \sigma^2)$.

Dengan metode estimasi *least squares* akan diestimasi koefisien regresi sedemikian sehingga jumlah kuadrat antara selisih nilai observasi y_i dengan nilai prediksi $\hat{y}_i$ adalah minimal. Langkahnya adalah sebagai berikut :

Misalkan

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

estimasor *least squares* koefisien regresi, $\hat{\beta}_0$ dan $\hat{\beta}_1$, harus memenuhi persamaan normal berikut :

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \qquad (2.7.1)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0 \qquad (2.7.2)$$

Sederhanakan persamaan (2.7.1) menjadi

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{N} y_i - \frac{\hat{\beta}_1}{n}\sum_{i=1}^{N} x_i$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \tag{2.7.3}$$

dengan

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

kemudian sederhanakan persamaan (2.7.2) menjadi

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \left(\bar{y} - \hat{\beta}_1\bar{x}\right)\sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \left(\bar{y} - \hat{\beta}_1\bar{x}\right) n\bar{x} = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad (2.7.4)$$

dengan

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y} = \sum_{i=1}^{n} (x_i - \overline{x}) y_i = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Dari persamaan (2.7.3) dan (2.7.4), estimator *least squares* koefisien regresi, $\hat{\beta}_0$ dan $\hat{\beta}_1$, adalah $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$ dan $\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$. Dimana $\hat{\beta}_0$ dan $\hat{\beta}_1$, adalah estimator tak bias karena $E(\hat{\beta}_0) = \beta_0$ dan $E(\hat{\beta}_1) = \beta_1$.

Model regresi linier yang tepat atau garis regresi untuk estimasi rata-rata y dari x yang diberikan adalah :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (2.7.5)$$

2.8 KOEFISIEN KORELASI (r)

Koefisien korelasi (r) adalah nilai yang menunjukkan kekuatan dan arah besarnyahubungan linier antara dua peubah dengan $-1 \leq r \leq 1$.

Koefisien korelasi dari sampel didefinisikan sebagai berikut :

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{\frac{1}{2}}}$$

$$= \frac{S_{xy}}{\left[S_{xx}S_{yy}\right]^{\frac{1}{2}}}$$

atau dari persamaan (2.7.4), r juga dapat dinyatakan sebagai :

$$r = \hat{\beta}_1\left(\frac{S_{xx}}{S_{yy}}\right)^{\frac{1}{2}} \qquad (2.8.1)$$

dengan

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

2.9 KOEFISIEN DETERMINASI (R^2)

Dalam regresi linier sederhana, selain koefisien korelasi (r), koefisien determinasi (R^2) juga merupakan suatu nilai yang menunjukkan kekuatan dan arah hubungan linier antara dua peubah acak dengan $0 \leq R^2 \leq 1$. Koefisien determinasi juga merupakan nilai yang menjelaskan kualitas atau seberapa bagus model regresi pada persamaan (2.7.5).

Koefisien determinasi dalam regresi linier sederhana didefinisikan sebagai berikut :

$$r^2 = \hat{\beta}_1^2 \frac{S_{xx}}{S_{yy}} = \hat{\beta}_1 \frac{S_{xy}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = R^2$$

atau

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \tag{2.9.1}$$

2.10 OUTLIERS

Outliers atau pencilan atau data yang menyimpang merupakan nilai observasi yang besar selisihnya dari garis lurus regresi. Karena *outliers* berpengaruh besar pada koefisien regresi dan dapat mengganggu ketepatan metode estimasi *least squares* dalam mengestimasi koefisien regresi, maka *outliers* perlu disisihkan dari kumpulan data.

Oleh karena itu diperlukan suatu metode untuk mendeteksi adanya *outliers*. Beberapa metode yang dapat digunakan dalam mendeteksi *outliers* adalah :

- *Cook's Distance*

  Yaitu nilai atau ukuran yang mengukur jarak *cook's* pada setiap data atau observasi.

  Cook's Distance didefinisikan sebagai :

$$D_i = \frac{r_i^2}{p} \times \frac{h_{ii}}{(1-h_{ii})} \text{ , untuk i = 1, …, n}$$

dengan : h_{ii} = elemen diagonal ke-i dari *hat matrix* $\mathbf{H=X(X'X)^{-1}X'}$ ukuran n x n

r_i = *studentized residuals* = $\dfrac{e_i}{\sqrt{MS_E(1-h_{ii})}} = \dfrac{y_i - \hat{y}_i}{\sqrt{MS_E(1-h_{ii})}}$

p = jumlah parameter dalam model regresi

$$MS_E = \frac{SS_E}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-2}$$

Besarnya D_i sering dibandingkan dengan $F_{\alpha, p, n-p}$. Seringkali digunakan $F_{0,5, p, n-p} \approx 1$, sehingga adanya *outliers* menurut banyak ahli statistika adalah data ke-i yang memiliki *Cook's Distance* > 1.

- *DFfits*

Yaitu nilai atau ukuran yang mengukur perubahan standar deviasi dari nilai yang diprediksi $\hat{y}_i$, bila observasi ke-i disisihkan dari kumpulan data.

DFfits didefiniskan sebagai :

$$DFfits_i = t_i \times \left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}}, \text{ untuk } i = 1, \ldots, n$$

dengan : h_{ii} = elemen diagonal ke-i dari *hat matrix* **H=X(X'X)$^{-1}$X'** ukuran n x n

t_i = *R-student*

$$= \frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}} = \frac{y_i - \hat{y}_i}{\sqrt{\left[\frac{(n-p)MS_E - e_i^2/(1-h_{ii})}{n-p-1}\right](1-h_{ii})}}$$

p = jumlah parameter dalam model regresi

$$MS_E = \frac{SS_E}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-2}$$

Menurut banyak ahli statistika, data ke-i yang memiliki

$\left|DFfits\right| > 2 \cdot \sqrt{p/n}$ dicurigai sebagai *outliers*.

2.11 MAKSIMUM LIKELIHOOD ESTIMATOR (M.L.E.)

Maksimum likelihood estimator adalah suatu metode yang digunakan untuk menaksir suatu parameter yang tidak diketahui dalam suatu fungsi probabilitas.

2.11.1 Definisi

Misalkan sampel acak X_1, X_2, …, X_n dari suatu distribusi yang memiliki bentuk p.d.f. $f(x_i;\theta), \theta \in \Omega$, dengan Ω adalah ruang parameter.

p.d.f bersama dari X_1, X_2, …, X_n adalah $f(x_1;\theta)f(x_2;\theta)...f(x_n;\theta)$.

Fungsi likelihood didefinisikan sebagai

$$L(\theta; x_1, x_2, ..., x_n) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot ... \cdot f(x_n;\theta)$$

Jadi $L(\theta; x_1, x_2, ..., x_n)$ dapat dipandang sebagai fungsi probabilitas bersama dari peubah acak X_1, X_2, …, X_n yang merupakan fungsi dari θ.

Dalam metode maksimum likelihood, akan dicari suatu fungsi nontrivial dari $X_1, X_2, \ldots, X_n$, misalkan $u(x_1, x_2, \ldots, x_n)$, sedemikian sehingga jika θ dalam $L(\theta; x_1, x_2, \ldots, x_n)$ diganti oleh $u(x_1, x_2, \ldots, x_n)$, maka nilai fungsi likelihood akan maksimum. Yaitu $L(u(x_1, x_2, \ldots, x_n); x_1, x_2, \ldots, x_n)$ setidaknya akan sebesar $L(\theta; x_1, x_2, \ldots, x_n)$ untuk setiap $\theta \in \Omega$.

Statistik $u(x_1, x_2, \ldots, x_n)$ disebut dengan *maximum likelihood estimator* (m.l.e.) dari θ dan dilambangkan dengan $\hat{\theta} = u(x_1, x_2, \ldots, x_n)$.

2.12 DISTRIBUSI CHI-SQUARE

Peubah acak kontinu X dikatakan memiliki distribusi *chi-square* dengan derajat bebas r apabila memiliki p.d.f. :

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, && \text{untuk } 0 < x < \infty \\ &= 0 && \text{, lainnya} \end{aligned}$$

dan memiliki m.g.f :

$$M(t) = (1-2t)^{-r/2}, \text{untuk } t < \frac{1}{2}$$

dimana $\mu = r, \sigma^2 = 2r$, dan distribusi *chi-square* merupakan distribusi *gamma* khusus dengan $\alpha = r/2$, dengan r adalah bilangan bulat positif, dan $\beta = 2$.

peubah acak kontinu X dikatakan memiliki distribusi *gamma* apabila memiliki p.d.f. :

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \text{untuk } 0 < x < \infty \\ &= 0 \qquad\qquad\qquad\qquad , \text{lainnya} \end{aligned}$$

dan memiliki m.g.f :

$$M(t) = \frac{1}{(1-\beta t)^{\alpha}}, \text{untuk } t < \frac{1}{\beta}$$

peubah acak kontinu X dikatakan memiliki distribusi *chi-square* dengan derajat bebas r dinyatakan dengan $X \sim \chi^2_{(r)}$.

2.13 DISTRIBUSI NORMAL

Peubah acak kontinu X dikatakan memiliki distribusi *normal* apabila memiliki p.d.f. :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \text{untuk } -\infty < x < \infty$$

dan memiliki m.g.f :

$$M(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$$

peubah acak kontinu X dikatakan memiliki distribusi normal dinyatakan dengan $X \sim N(\mu,\sigma^2)$.

Apabila diketahui bahwa peubah acak kontinu X memiliki distribusi normal standar N(0,1), artinya $\mu = 0$ dan $\sigma^2 = 1$. Sehingga p.d.f. yang dimiliki peubah acak X adalah :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \text{untuk } -\infty < x < \infty$$

dan memiliki m.g.f :

$$M(t) = \exp(\frac{t^2}{2})$$

peubah acak kontinu X dikatakan memiliki distribusi normal standar dinyatakan dengan $X \sim N(0,1)$.

2.14 DISTRIBUSI t

Misalkan peubah acak kontinu W berdistribusi normal standar, $W \sim N(0,1)$, dan peubah acak kontinu V berdistribusi *chi-square* dengan derajat bebas r, $V \sim \chi^2_{(r)}$, dengan W dan V saling bebas.

Definisikan peubah acak kontinu baru yaitu :

$$T = \frac{W}{\sqrt{V/r}}$$

peubah acak kontinu T tersebut dikatakan memiliki distribusi *t* dengan derajat bebas r apabila memiliki p.d.f. :

$$g(t) = \frac{\Gamma\left[(r+1)/2\right]}{\sqrt{\pi r} \cdot \Gamma(r/2)} \cdot \frac{1}{(1+t^2/r)^{(r+1)/2}}, \text{untuk } -\infty < t < \infty$$

2.15 DISTRIBUSI F

Misalkan dua peubah acak kontinu U dan V masing-masing berdistribusi *chi-square* dengan derajat bebas r_1 dan r_2,

$U \sim \chi^2_{(r_1)}$, dan $V \sim \chi^2_{(r_2)}$ dengan U dan V saling bebas.

Definisikan peubah acak kontinu baru yaitu :

$$F = \frac{U/r_1}{V/r_2}$$

peubah acak kontinu F tersebut dikatakan memiliki distribusi *F* dengan derajat bebas r_1 dan r_2 apabila memiliki p.d.f. :

$$\begin{aligned} g(f) &= \frac{\Gamma\left[(r_1+r_2)/2\right](r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \cdot \frac{(w)^{r_1/2-1}}{(1+r_1 w/r_2)^{(r_1+r_2)/2}}, \text{ untuk } 0 < f < \infty \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\, \text{, lainnya} \end{aligned}$$

BAB III

PERBANDINGAN METODE *CLASSIC* DAN METODE *INVERSE* PADA REGRESI KALIBRASI

Dalam bab ini akan dibahas mengenai metode *classic* dan metode *inverse* untuk melakukan estimasi nilai aktual pada regresi kalibrasi, estimasi interval kepercayaan nilai aktual dengan menggunakan metode Graybill, sekaligus perbandingan kedua metode tersebut.

3.1 REGRESI KALIBRASI

Dalam banyak masalah regresi, baik yang melibatkan estimasi atau prediksi, biasanya akan ditentukan berapa nilai y yang bersesuaian bila nilai x diberikan. Namun ada yang terjadi sebaliknya, yaitu akan ditentukan berapa nilai x yang bersesuaian bila nilai y diberikan. Sebagai ilustrasi, misalnya akan dikalibrasi sebuah alat ukur suhu yaitu thermocouple, dimana diasumsikan bahwa suhu observasi atau nilai yang terbaca pada thermocouple memiliki hubungan linear dengan suhu atau nilai aktual yang diukur. Misalkan persamaan regresi linier populasinya sebagai berikut :

$$\text{Suhu observasi} = \beta_0 + \beta_1 \text{ (suhu aktual)} + \varepsilon$$

atau

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (3.1.1)$$

Pada regresi biasa, akan diberikan suhu aktual (x) kemudian ditentukan suhu observasi (y) yang bersesuaian. Sedangkan bila yang terjadi sebaliknya, yaitu bila diberikan suhu observasi (y) kemudian akan ditentukan suhu aktual (x) yang bersesuaian, masalah yang terjadi tersebut dikenal dengan masalah kalibrasi dan sering juga disebut dengan regresi kalibrasi. Estimasi nilai aktual atau suhu aktual pada regresi kalibrasi dapat dilakukan dengan metode *classic* dan metode *inverse*.

3.2 METODE *CLASSIC*

Dengan menggunakan model (3.1.1) dan sejumlah n data suhu berpasangan serta diasumsikan errornya memenuhi asumsi regresi linier sederhana yaitu *normally and independently distributed* dengan $E(\varepsilon)=0$ dan $Var(\varepsilon)=\sigma^2$ atau dengan perkataan lain $\varepsilon \sim NID(0,\sigma^2)$, model tersebut dapat dinyatakan sebagai model regresi linier sampel sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

dengan

$$i = 1, 2, ..., n$$

dimana parameter β_0 dan β_1 akan diestimasi dengan metode estimasi *least squares*. Dari persamaan (2.7.3) dan (2.7.4) didapat $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ dan $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, sehingga didapat garis regresi

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (3.2.1)$$

Persamaan regresi kalibrasi yang bersesuaian dengan (3.2.1) menjadi

$$\hat{x} = \frac{y - \hat{\beta}_0}{\hat{\beta}_1} \qquad (3.2.2)$$

Apabila diberikan suhu observasi (y_0) yang terbaca pada thermocouple, maka estimasi suhu aktual (x_0) berdasarkan metode *classic* adalah

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \qquad (3.2.3)$$

3.3 METODE *INVERSE*

Selain metode *classic,* terdapat metode lain untuk estimasi nilai aktual dari nilai observasi yaitu metode *inverse.*

Dengan menggunakan model (3.1.1) kembali, yaitu

$$y = \beta_0 + \beta_1 x + \varepsilon$$

yang dapat dinyatakan sebagai

$$\beta_1 x = y - \beta_0 - \varepsilon$$

$$x = (-\frac{\beta_0}{\beta_1}) + (\frac{1}{\beta_1})y + (-\frac{1}{\beta_1})\varepsilon$$

$$x = \alpha_0 + \alpha_1 y + \varepsilon' \qquad (3.3.1)$$

dengan

$$\alpha_0 = -\frac{\beta_0}{\beta_1}\ ,\ \ \alpha_1 = \frac{1}{\beta_1}\ ,\ \ \varepsilon' = -\frac{\varepsilon}{\beta_1}$$

Dengan menggunakan sejumlah n data suhu berpasangan serta diasumsikan errornya memenuhi asumsi regresi linier sederhana yaitu *normally and independently distributed* dengan $E(\varepsilon) = 0$ dan $Var(\varepsilon) = \sigma^2$ atau dengan perkataan lain $\varepsilon \sim NID(0, \sigma^2)$, model tersebut dapat dinyatakan sebagai model regresi linier sampel sebagai berikut :

$$x_i = \alpha_0 + \alpha_1 y_i + \varepsilon_i'$$

dengan

$$i = 1, 2, ..., n$$

dimana parameter α_0 dan α_1 akan diestimasi dengan metode estimasi *least squares* sebagai berikut :

Misalkan

$$S(\alpha_0, \alpha_1) = \sum_{i=1}^{n} (x_i - \alpha_0 - \alpha_1 y_i)^2$$

estimasor *least squares* koefisien regresi, $\hat{\alpha}_0$ dan $\hat{\alpha}_1$, harus memenuhi persamaan normal berikut :

$$\frac{\partial S}{\partial \alpha_0} = -2\sum_{i=1}^{n} (x_i - \hat{\alpha}_0 - \hat{\alpha}_1 y_i) = 0 \qquad (3.3.2)$$

$$\frac{\partial S}{\partial \alpha_1} = -2\sum_{i=1}^{n} (x_i - \hat{\alpha}_0 - \hat{\alpha}_1 y_i) y_i = 0 \qquad (3.3.3)$$

Sederhanakan persamaan (3.3.2) menjadi

$$\sum_{i=1}^{n} x_i - n\hat{\alpha}_0 - \hat{\alpha}_1 \sum_{i=1}^{n} y_i = 0$$

$$\hat{\alpha}_0 = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\hat{\alpha}_1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\alpha}_0 = \overline{x} - \hat{\alpha}_1 \overline{y} \qquad (3.3.4)$$

dimana

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Kemudian sederhanakan persamaan (3.3.3) menjadi

$$\sum_{i=1}^{n} x_i y_i - \hat{\alpha}_0 \sum_{i=1}^{n} y_i - \hat{\alpha}_1 \sum_{i=1}^{n} y_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - (\overline{x} - \hat{\alpha}_1 \overline{y}) \sum_{i=1}^{n} y_i - \hat{\alpha}_1 \sum_{i=1}^{n} y_i^2 = 0$$

$$\hat{\alpha}_1 \sum_{i=1}^{n} y_i^2 + (\overline{x} - \hat{\alpha}_1 \overline{y}) n\overline{y} = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\alpha}_1 \left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}$$

$$\hat{\alpha}_1 = \frac{S_{xy}}{S_{yy}} \tag{3.3.5}$$

dengan

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y} = \sum_{i=1}^{n} (x_i - \overline{x}) y_i = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

Garis regresi sekaligus Persamaan regresi kalibrasi yang bersesuaian menjadi

$$\hat{x} = \hat{\alpha}_0 + \hat{\alpha}_1 y \tag{3.3.6}$$

Apabila diberikan suhu observasi (y_0) yang terbaca pada thermocouple, maka estimasi suhu aktual (x_0) berdasarkan metode *inverse* adalah

$$\hat{x}_0 = \hat{\alpha}_0 + \hat{\alpha}_1 y_0 \tag{3.3.7}$$

3.4 KURVA KALIBRASI

Setelah didapatkan persamaan regresi kalibrasi baik pada metode *classic* maupun metode *inverse*, persamaan tersebut digunakan untuk menggambarkan grafik antara nilai observasi dan nilai aktual yang diperoleh. Grafik atau kurva hasil data kalibrasi tersebut dinamakan kurva kalibrasi.

Kurva kalibrasi berguna untuk menentukan secara visual estimasi nilai aktual dari nilai observasi yang diberikan.

3.5 ESTIMASI INTERVAL KEPERCAYAAN NILAI AKTUAL

Setelah estimasi titik nilai aktual dilakukan, selanjutnya dilakukan estimasi interval kepercayaan nilai aktual untuk mengetahui presisi hasil estimasi dengan tingkat kepercayaan α. Metode estimasi interval kepercayaan yang digunakan adalah metode Graybill.

Pada persamaan regresi kalibrasi sebelumnya, baik pada metode *classic* maupun *inverse*, model dasarnya adalah persamaan (3.1.1) sebagai berikut :

$$y = \beta_0 + \beta_1 x + \varepsilon, \qquad \varepsilon \sim NID(0,\sigma^2) \tag{3.5.1}$$

Pada metode Graybill dilakukan dua variasi guna mendapatkan persamaan regresi kalibrasi yang dapat digunakan sebagai estimasi interval kepercayaan. Variasi pertama dilakukan definisi ulang x pada persamaan (3.5.1) sebagai deviasi dari rata-ratanya, yaitu $x - \bar{x}$, sehingga model regresi tersebut dapat dinyatakan sebagai :

$$y = \gamma_0 + \gamma_1 (x - \bar{x}) + \varepsilon, \qquad \varepsilon \sim N(0,\sigma^2) \tag{3.5.2}$$

dimana γ_0 dan γ_1 adalah parameter dengan $\text{cov}(\hat{\gamma}_0, \hat{\gamma}_1) = 0$ (bukti pada lampiran)

Kedua model tersebut, yaitu model pada persamaan (3.5.1) dan (3.5.2), adalah ekivalen bila didefinisikan

$$\gamma_1 = \beta_1$$

dan

$$\beta_0 = \gamma_0 - \gamma_1 \bar{x}$$

Variasi kedua yang dilakukan pada metode Graybill adalah dilakukan observasi tambahan sebanyak $k \geq 1$ pada suhu aktual x_0, sedemikian sehingga sampel yang didapat menjadi n+k data berpasangan yang dinotasikan sebagai berikut :

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), (x_0, y_{n+1}), (x_0, y_{n+2}), \ldots, (x_0, y_{n+k})$$

observasi tambahan sebanyak k yang dinotasikan dengan $y_{n+1}, y_{n+2}, \ldots, y_{n+k}$ tersebut, diasumsikan berasal dari sampel *random* berdistribusi normal dengan mean $\gamma_0 + \gamma_1(x_0 - \bar{x})$ dan variansi σ^2 dengan $x_0, \gamma_0, \gamma_1, \& \sigma^2$ tidak diketahui. Sebelum estimasi interval kepercayaan dapat dilakukan akan diestimasi terlebih dahulu parameter $x_0, \gamma_0, \gamma_1, \& \sigma^2$.

3.5.1 Estimasi γ_0, γ_1, dan x_0

- Akan diestimasi parameter γ_1 sebagai berikut :

Estimasi γ_1 pada model persamaan (3.5.2) sama dengan hasil estimasi parameter β_1 pada model persamaan (3.5.1), yaitu

$$\hat{\gamma}_1 = \hat{\beta}_1$$

dari persamaan (2.7.4) didapat

$$\hat{\gamma}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (3.5.1.1)$$

- Akan diestimasi parameter γ_0 sebagai berikut :

$$\hat{\gamma}_0 = \hat{\beta}_0 + \hat{\gamma}_1 \overline{x}$$

diketahui $\hat{\gamma}_1 = \beta_1$ dan $\hat{\beta}_0 = \overline{y} - \beta_1 \overline{x}$, sehingga

$$\hat{\gamma}_0 = \overline{y} - \beta_1 \overline{x} + \beta_1 \overline{x} = \overline{y} \qquad (3.5.1.2)$$

dengan

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Asumsi normalitas pada ε di persamaan (3.5.2) mengimplikasikan $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ berdistribusi normal juga (bukti pada lampiran).

- Akan diestimasi x_0 berdasarkan n+k data

  $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), (x_0, \overline{y}_0)$

$$\hat{x}_0 = \overline{x} + \frac{\overline{y}_0 - \hat{\gamma}_0}{\hat{\gamma}_1}$$

diketahui $\hat{\gamma}_0 = \overline{y}$, sehingga

$$\hat{x}_0 = \overline{x} + \frac{\overline{y}_0 - \overline{y}}{\hat{\gamma}_1} \qquad (3.5.1.3)$$

dengan

$$\overline{y}_0 = \frac{1}{k}\sum_{i=n+1}^{n+k} y_i$$

3.5.2 Estimasi σ^2

Akan digunakan fungsi *likelihood* untuk mengestimasi variansi σ^2 berdasarkan n+k data. Fungsi *likelihood* tersebut adalah

$$L(\gamma_0, \gamma_1, \sigma^2, x_0 : y_1, x_1, \ldots, y_n, x_n, ; y_{n+1}, y_{n+2}, \ldots, y_{n+k})$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \gamma_0 - \gamma_1(x_i - \overline{x}))^2 \right) \prod_{i=n+1}^{n+k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \gamma_0 - \gamma_1(x_0 - \overline{x}))^2 \right)$$

$$= \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n+k}{2}}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (y_i - \gamma_0 - \gamma_1(x_i - \overline{x}))^2 + \sum_{i=n+1}^{n+k} (y_i - \gamma_0 - \gamma_1(x_0 - \overline{x}))^2 \right) \right)$$

Log likelihoodnya menjadi

$$-\left(\frac{n+k}{2}\right)\log 2\pi - \left(\frac{n+k}{2}\right)\log\sigma^2 - \sum_{i=1}^{n} \frac{(y_i - \gamma_0 - \gamma_1(x_i - \overline{x}))^2}{2\sigma^2} - \sum_{i=n+1}^{n+k} \frac{(y_i - \gamma_0 - \gamma_1(x_0 - \overline{x}))^2}{2\sigma^2}$$

Turunkan log *likelihood* tersebut terhadap σ^2, dan samakan dengan nol, kemudian substitusi γ_0, γ_1, dan x_0 dengan estimatornya, yaitu $\hat{\gamma}_0$, $\hat{\gamma}_1$, dan $\hat{x}_0$, maka akan didapatkan estimator dari σ^2.

turunan parsial terhadap σ^2 adalah

$$\frac{\partial \log L}{\partial \sigma^2}\Big|_{\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0, \hat{\sigma}^2} = -\frac{n+k}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \left(\sum_{i=1}^{n} (y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x}))^2 + \sum_{i=n+1}^{n+k} (y_i - \hat{\gamma}_0 - \hat{\gamma}_1(\hat{x}_0 - \overline{x}))^2 \right)$$

samakan dengan nol

$$-\frac{n+k}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \left(\sum_{i=1}^{n} (y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x}))^2 + \sum_{i=n+1}^{n+k} (y_i - \hat{\gamma}_0 - \hat{\gamma}_1(\hat{x}_0 - \overline{x}))^2 \right) = 0$$

kalikan dengan kedua ruas dengan $\hat{\sigma}^2$ kemudian substitusi $\hat{x}_0 = \bar{x} + \frac{\bar{y}_0 - \bar{y}}{\hat{\gamma}_1}$,

dan $\hat{\gamma}_0 = \bar{y}$ akan menghasilkan

$$-(n+k) + \frac{1}{\hat{\sigma}^2}\left(\sum_{i=1}^{n}(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}(y_i - \bar{y} - \hat{\gamma}_1(\bar{x} + \frac{\bar{y}_0 - \bar{y}}{\hat{\gamma}_1} - \bar{x}))^2\right) = 0$$

sederhanakan menjadi

$$n + k = \frac{1}{\hat{\sigma}^2}\left(\sum_{i=1}^{n}(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}(y_i - \bar{y}_0)^2\right)$$

Dengan demikian *maximum likelihood estimator* untuk σ^2 adalah

$$\hat{\sigma}^2_{ML} = \frac{1}{n+k}\left(\sum_{i=1}^{n}(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}(y_i - \bar{y}_0)^2\right) \qquad (3.5.2.1)$$

Untuk mendapatkan estimator tak bias dari σ^2, selanjutnya akan dihitung ekspektasi dari $\hat{\sigma}^2_{ML}$, tetapi sebelumnya diperlukan perhitungan beberapa covarian sebagai berikut : (bukti pada lampiran)

- $\mathrm{Cov}(y_i, \hat{\gamma}_0) = \frac{\sigma^2}{n}$
- $\mathrm{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = 0$
- $\mathrm{Cov}(y_i, \hat{\gamma}_1) = \frac{\sigma^2(x_i - \bar{x})}{\sum_{k=1}^{n}(x_k - \bar{x})^2}$
- $\mathrm{Cov}(y_i, \bar{y}_0) = \frac{\sigma^2}{k}$

Ekspektasi dari $\hat{\sigma}^2_{ML}$ adalah

$$E(\hat{\sigma}^2_{ML}) = E\left(\frac{1}{n+k}\left(\sum_{i=1}^{n}(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}(y_i - \bar{y}_0)^2\right)\right)$$

berdasarkan operasi linear pada ekspektasi $E(ax+by) = aE(x)+bE(y)$, maka

$$E(\hat{\sigma}^2_{ML}) = \frac{1}{n+k}\left(\sum_{i=1}^{n}E(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}E(y_i - \bar{y}_0)^2\right)$$

berdasarkan definisi variansi yaitu $Var(x) = E(x^2) - (E(x))^2$,

sehingga $E(x^2) = Var(x) + (E(x))^2$, maka

$$E(\hat{\sigma}^2_{ML}) = \frac{1}{n+k}\left(\sum_{i=1}^{n}Var(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x})) + (E(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x})))^2\right)$$

$$+\frac{1}{n+k}\left(\sum_{i=n+1}^{n+k}Var(y_i - \bar{y}_0) + (E(y_i - \bar{y}_0))^2\right)$$

dengan mengunakan sifat penjumlahan pada variansi

yaitu $Var(x+y) = Var(x) + Var(y) + 2Cov(x,y)$, akan didapat

$$E(\hat{\sigma}^2_{ML}) = \frac{1}{n+k}\sum_{i=1}^{n}(Var(y_i) + Var(\hat{\gamma}_0) + (x_i - \bar{x})^2 Var(\hat{\gamma}_1))$$

$$+\frac{1}{n+k}\sum_{i=1}^{n}(-2Cov(y_i,\hat{\gamma}_0) + 2(x_i - \bar{x})Cov(\hat{\gamma}_0,\hat{\gamma}_1) - 2(x_i - \bar{x})Cov(y_i,\hat{\gamma}_1))$$

$$+\frac{1}{n+k}\sum_{i=1}^{n}(E(y_i) - E(\hat{\gamma}_0) - (x_i - \bar{x})E(\hat{\gamma}_1))^2$$

$$+\frac{1}{n+k}\sum_{i=n+1}^{n+k}(Var(y_i) + Var(\bar{y}_0) + (E(y_i) - E(\bar{y}_0))^2)$$

Substitusikan dengan hasil ekspektasi, variansi, dan kovariansi masing-masing, maka akan didapat

$$\frac{1}{n+k}\left(\sum_{i=1}^{n}\left(\sigma^2+\frac{\sigma^2}{n}+\frac{\sigma^2(x_i-\overline{x})^2}{\sum_{j=1}^{n}(x_j-\overline{x})^2}-\frac{2\sigma^2}{n}-\frac{2\sigma^2(x_i-\overline{x})^2}{\sum_{j=1}^{n}(x_j-\overline{x})^2}\right)+\sum_{i=n+1}^{n+k}\left(\sigma^2+\frac{\sigma^2}{k}-\frac{2\sigma^2}{k}\right)\right)$$

sederhanakan menjadi

$$\frac{1}{n+k}\left(\sum_{i=1}^{n}\left(\sigma^2-\frac{\sigma^2}{n}-\frac{\sigma^2(x_i-\overline{x})^2}{\sum_{j=1}^{n}(x_j-\overline{x})^2}\right)+\sum_{i=n+1}^{n+k}\left(\sigma^2-\frac{\sigma^2}{k}\right)\right)=\frac{1}{n+k}\left(n\sigma^2-\sigma^2-\sigma^2+k\sigma^2-\sigma^2\right)$$

didapat

$$E(\hat{\sigma}^2_{ML})=\frac{n+k-3}{n+k}\sigma^2$$

maka

$$\hat{\sigma}^2=\frac{n+k}{n+k-3}E(\hat{\sigma}^2_{ML})$$

substitusikan dengan persamaan (3.5.2.1), didapat

$$\hat{\sigma}^2_{\text{corrected for bias}}=\frac{1}{n+k-3}\left(\sum_{i=1}^{n}(y_i-\hat{\gamma}_0-\hat{\gamma}_1(x_i-\overline{x}))^2+\sum_{i=n+1}^{n+k}(y_i-\overline{y}_0)^2\right) \quad (3.5.2.2)$$

3.5.3 Teorema pendukung

Berdasarkan hasil yang sudah didapat pada persamaan (3.5.1.1), (3.5.1.2), (3.5.1.3), dan (3.5.2.2), berikut adalah teorema yang dibutuhkan dalam menentukan interval kepercayaan nilai aktual x_0 bila diberikan nilai observasi y_0.

Teorema 3.5.3

1. $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ *independent*
2. $\hat{\gamma}_0 \sim N\left(\gamma_{0,} \dfrac{\sigma^2}{n}\right)$ dan $\hat{\gamma}_1 \sim N\left(\hat{\gamma}_{1,} \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$
3. $U = (n+k-3)\dfrac{\hat{\sigma}^2}{\sigma^2}$ berdistribusi χ^2_{n+k-3}
4. $\hat{\sigma}^2$ tidak bergantung dengan $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)$

(Pembuktian teorema pada lampiran)

3.5.4 Interval kepercayaan untuk x_0 bila diberikan y_0

Dengan bantuan hasil-hasil di subbab sebelumnya, yaitu persamaan (3.5.1.1), (3.5.1.2), (3.5.1.3), (3.5.2.2), dan teorema 3.5.3, akan di tentukan interval kepercayaan untuk nilai aktual x_0 bila diberikan nilai observasi y_0 pada regresi kalibrasi. Pada model persamaan (3.5.2), secara intuisi jelas bahwa tidak akan didapatkan interval kepercayaan untuk x_0 bila γ_1 sama dengan nol.

Diketahui bahwa $\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})$ adalah penjumlahan dari bagian yang berdistribusi normal. Oleh karena itu $\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})$ juga berdistribusi normal dengan rata-rata dan variansi sebagai berikut :

- $E(\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x})) = E(\bar{y}_0) - E(\hat{\gamma}_0) - E(\hat{\gamma}_1(x_i - \bar{x}))$

$$= \frac{1}{k}\sum_{i=n+1}^{n+k}\left(\gamma_0 + \gamma_1(x_i - \bar{x})\right) - \gamma_0 - \gamma_1(x_i - \bar{x}) = 0$$

- $Var(\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})) = \sigma^2\left(\frac{1}{k} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$

$$= \sigma^2 A^2 \text{, dimana } A^2 = \left(\frac{1}{k} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

Misalkan $Z = \frac{\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})}{\sqrt{Var\left(\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})\right)}} = \frac{\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})}{\sqrt{\sigma^2 A^2}} \sim N\left(0,1\right)$

Dan dari teorema 3.5.3.(3) didapat :

$$U = (n+k-3)\frac{\hat{\sigma}^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k}(y_i - \bar{y}_0)^2}{\sigma^2} \sim \chi^2_{n+k-3}$$

diketahui bahwa Z dan U saling bebas, karena $\hat{\sigma}^2$ tidak bergantung dengan $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)$. Sehingga akan kita dapatkan $Z \sim N\left(0,1\right)$ dan $U \sim \chi^2_{n+k-3}$ dengan Z dan U saling bebas :

$$T = \frac{Z}{\sqrt{\frac{U}{n+k-3}}} \sim t_{\frac{\alpha}{2};n+k-3}$$

dan

$$P(-t_{\frac{\alpha}{2};n+k-3} \le T \le t_{\frac{\alpha}{2};n+k-3}) = 1-\alpha \qquad (3.5.4.1)$$

Nilai didalam kurung pada persamaan (3.5.4.1) ekivalen dengan

$$T^2 \le t^2_{\frac{\alpha}{2};n+k-3}$$

atau

$$\frac{(\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2}{\hat{\sigma}^2 A^2} \le t^2_{\frac{\alpha}{2};n+k-3}$$

atau

$$(\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2 - \hat{\sigma}^2 A^2 t^2_{\frac{\alpha}{2};n+k-3} \leq 0 \qquad (3.5.4.2)$$

Apabila bagian pertama persamaan (3.5.4.2) dijabarkan dan definisi A^2 disubstitusikan, akan diperoleh pertidaksamaan berikut :

$$\left(\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)(x_0 - \bar{x})^2 - 2\hat{\gamma}_1(\bar{y}_0 - \hat{\gamma}_0)(x_0 - \bar{x}) + \left((\bar{y}_0 - \hat{\gamma}_0)^2 - \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\frac{1}{k} + \frac{1}{n} \right) \right) \leq 0$$

Dapat dilihat bahwa pertidaksamaan tersebut adalah pertidaksamaan kuadratis, yang bisa dituliskan ulang sebagai

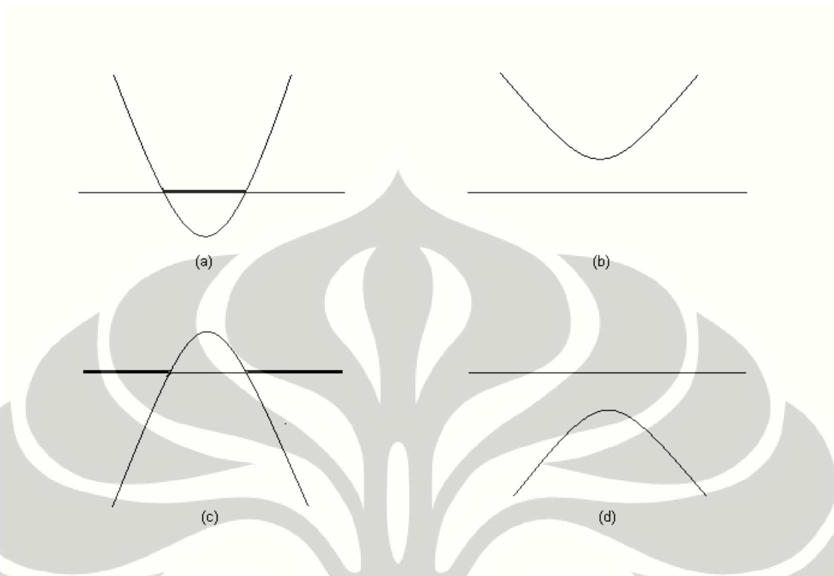
$$q(x_0 - \bar{x}) = a(x_0 - \bar{x})^2 + 2b(x_0 - \bar{x}) + c \leq 0$$

Dimana

$$a = \left(\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right), \; b = -\left(\hat{\gamma}_1(\bar{y}_0 - \hat{\gamma}_0) \right), \; c = \left((\bar{y}_0 - \hat{\gamma}_0)^2 - \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\frac{1}{k} + \frac{1}{n} \right) \right)$$

Jika nilai dari x_o yang memenuhi pertidaksamaan adalah suatu interval, maka nilai tersebut akan membentuk suatu 100(1-α)% interval kepercayaan untuk x_0.

sekarang akan diselidiki pertidaksamaan kuadratis tersebut.



Gambar 3.5.4.1

- Jika diskriminan b^2 - ac dari fungsi kuadratis tersebut adalah negatif, D<0, maka fungsi kuadratis tersebut tidak dapat sama dengan nol. (lihat gambar 3.5.4.1 (b) dan (d)). Dalam kasus ini tidak ada interval kepercayaan yang dihasilkan.
- Jika diskriminan b^2 - ac dari fungsi kuadratis tersebut adalah nol, D=0, maka akan ada titik tunggal z_0, dengan $q(z_0 - \overline{x}) = 0$. Dalam kasus ini, interval kepercayaan untuk x_0 adalah z_0, sehingga interval kepercayaan tidak berarti.

- Jika diskriminan b^2 - ac dari fungsi kuadratis tersebut adalah positif, D>0, maka akan ada dua keadaan yang memungkinkan, (Lihat gambar 3.5.4.1 (a) dan (c)), yaitu :
  - jika $a < 0$, maka nilai dari x_0 dimana $q(x_0 - \bar{x}) \leq 0$ akan membentuk dua interval kepercayaan untuk x_0 yang tak terhingga. (Lihat gambar 3.5.4.1(c)).
  - jika $a > 0$, maka nilai dari x_0 dimana $q(x_0 - \bar{x}) \leq 0$ akan membentuk interval kepercayaan untuk x_0 yang berarti. (Lihat gambar 3.5.4.1(a)).

Dapat disimpulkan bahwa pertidaksamaan $q(x_0 - \bar{x}) \leq 0$ akan menghasilkan interval kepercayaan untuk x_0 yang berarti jika dan hanya jika $a > 0$ dan $b^2 - ac > 0$. Jika $b^2 - ac$ dijabarkan akan didapat :

$$b^2 - ac = \hat{\gamma}_1^2(\bar{y}_0 - \hat{\gamma}_0)^2 - \left(\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)\left((\bar{y}_0 - \hat{\gamma}_0)^2 - \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\frac{1}{k} + \frac{1}{n} \right) \right)$$

$$= \hat{\gamma}_1^2 \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\frac{1}{k} + \frac{1}{n} \right) + \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} (\bar{y}_0 - \hat{\gamma}_0)^2 - \frac{\hat{\sigma}^4 t^4_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \left(\frac{1}{k} + \frac{1}{n} \right)$$

$$= \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\hat{\gamma}_1^2 \left(\frac{1}{k} + \frac{1}{n} \right) + \frac{(\bar{y}_0 - \hat{\gamma}_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \left(\frac{1}{k} + \frac{1}{n} \right) \right)$$

$$= \hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3} \left(\left(\frac{1}{k} + \frac{1}{n} \right) a + \frac{(\bar{y}_0 - \hat{\gamma}_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) \text{ dimana } a = \left(\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

Oleh karena itu, jika $a \geq 0$, maka diskriminan $b^2 - ac \geq 0$. Jadi, pertidaksamaan $q(x_0 - \bar{x}) \leq 0$ akan menghasilkan interval kepercayaan untuk x_0 jika dan hanya jika $a \geq 0$, atau dengan perkataan lain , jika dan hanya jika $\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \geq 0$. Jika $\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2};n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \geq 0$ disederhanakan akan didapat $\frac{\hat{\gamma}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\hat{\sigma}^2} \geq t^2_{\frac{\alpha}{2};n+k-3} = F_{\alpha;1,n+k-3}$, yang merupakan statistik uji dengan tingkat signifikansi α dari $H_0 : \gamma_1 = 0$ dan $H_1 : \gamma_1 \neq 0$.

Berikut adalah prosedur menentukan interval kepercayaan untuk nilai aktual x_0 bila diberikan nilai observasi y_0. Prosedur sebagai berikut :

1. Gunakan persamaan (3.5.1.1), (3.5.1.2), dan (3.5.2.2) untuk menghasilkan estimator nilai dari γ_0, γ_1, dan σ^2.
2. Uji $H_0 : \gamma_1 = 0$ dan $H_1 : \gamma_1 \neq 0$ dengan tingkat signifikansi α.

   Tolak H_0 jika dan hanya jika $\dfrac{\hat{\gamma}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\hat{\sigma}^2} \geq t^2_{\frac{\alpha}{2}:n+k-3} = F_{\alpha:1,n+k-3}$
3. Jika H_0 tidak ditolak, model akan menjadi $y_i = \gamma_0 + \varepsilon_i$, dimana x_0 tidak ada dalam model sehingga tidak ada interval kepercayaan yang dihasilkan.
4. Jika H_0 ditolak, bentuk $100(1-\alpha)\%$ interval kepercayaan untuk nilai aktual x_0 dengan batas kiri

$$\bar{x} + \frac{\hat{\gamma}_1(\bar{y}_0 - \bar{y})}{a} - \frac{t_{\frac{\alpha}{2}:n+k-3}\,\hat{\sigma}}{a} \sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{(\bar{y}_0 - \bar{y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

   dan batas kanan

$$\bar{x} + \frac{\hat{\gamma}_1(\bar{y}_0 - \bar{y})}{a} + \frac{t_{\frac{\alpha}{2}:n+k-3}\,\hat{\sigma}}{a} \sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{(\bar{y}_0 - \bar{y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

   dengan $a = \hat{\gamma}_1^2 - \dfrac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2}:n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

3.6 PERBANDINGAN

Estimasi nilai aktual berdasarkan metode *classic* pada persamaan (3.2.3) dapat dinyatakan sebagai (bukti pada lampiran) :

$$x_{classic} = \bar{x} + (yo - \bar{y})\frac{S_{xx}}{S_{xy}} \qquad (3.6.1)$$

Sedangkan Estimasi nilai aktual berdasarkan metode *inverse* pada persamaan (3.3.7), dimana sudah banyak digunakan karena kesederhanaannya dalam mengestimasi nilai aktual x_0, yaitu dengan langsung mengganti y pada persamaan garis regresi kalibrasi (3.3.6) $\hat{x} = \hat{\alpha}_0 + \hat{\alpha}_1 y$, dengan nilai observasi y_0, dapat dinyatakan sebagai (bukti pada lampiran) :

$$x_{inverse} = \bar{x} + (y_0 - \bar{y})\frac{S_{xy}}{S_{yy}} \qquad (3.6.2)$$

Kedua estimator tersebut akan dibandingkan dengan menggunakan rasio antara $x_{inverse}$ dengan $x_{classic}$, yaitu

$$\frac{x_{inverse}}{x_{classic}}$$

Atau rasio relatif

$$\frac{(x_{inverse} - \bar{x})}{(x_{classic} - \bar{x})}$$

Telah diketahui bersama pada persamaan (2.9.1) bahwa $R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$, dimana R adalah koefisien korelasi. Dari (3.6.1) dan (3.6.2) akan didapat

$$\frac{(x_{inverse} - \overline{x})}{(x_{classic} - \overline{x})} = R^2 \tag{3.6.3}$$

dan

$$\frac{x_{inverse}}{x_{classic}} = 1 + (1 - R^2)(\frac{\overline{x}}{x_{classic}} - 1) \tag{3.6.4}$$

(Bukti (3.6.3) dan (3.6.4) pada lampiran)

Sebagai konsekuensi langsung dari (3.6.3) dan (3.6.4) adalah

- Perbedaan antara estimator *classic* dan estimator *inverse* adalah 0 (kedua metode sama) jika dan hanya jika $R^2 = 1$
- Karena pada kenyataannya $R^2 \leq 1$, estimator *inverse* selalu lebih dekat dengan $\bar{x}$ daripada estimator *classic* karena lebih kecil selisihnya dengan $\bar{x}$

BAB IV

ANALISIS DATA

Sebagai contoh pembahasan dalam bab ini akan dikalibrasi temperatur observasi yang dihasilkan oleh alat ukur suhu yaitu Thermocouple. Data kalibrasi yang diperoleh adalah temperatur aktual x, yang diukur dengan thermometer yang telah tertelusur ke standar internasional. Sedangkan data yang bersesuaian adalah temperatur observasi y, yang diukur dengan thermocouple. Data yang diperoleh berasal dari buku Montgomery (2001), sebagai berikut :

Tabel 4.1 : Temperatur aktual dan temperatur observasi

Observasi (I)	Temperatur Aktual (x_i)	Temperatur Observasi (y_i)
1	100	88.8
2	120	108.7
3	140	129.8
4	160	146.2
5	180	161.6
6	200	179.9
7	220	202.4
8	240	224.5
9	260	245.1
10	280	257.7
11	300	277.0
12	320	298.1
13	340	318.8
14	360	334.6
15	380	355.2
16	400	377.0

Misalkan diperoleh nilai observasi baru yaitu temperatur observasi $y_0 = 200°C$. Akan diestimasi temperatur aktual x_0 beserta interval kepercayaannya yang bersesuaian dengan temperatur observasi y_0 tersebut, dengan metode *classic* maupun dengan metode *inverse*. Kemudian kedua metode akan dibandingkan.

5.1 METODE *CLASSIC*

Untuk estimator *classic* diperoleh model regresi linear sederhana :

$$y = \beta_0 + \beta_1 x + \varepsilon,\ \varepsilon \sim N(0,\sigma^2)\ ,\quad \varepsilon \sim N(0,\sigma^2)$$

dimana parameter β_0 dan β_1 diestimasi dengan metode least squares sebagai $\hat{\beta}_0$ dan $\hat{\beta}_1$.

Tabel 4.1.1 adalah tabel Anova dari model regresi :

Tabel 4.1.1 : Tabel Anova untuk model regresi dari estimator *classic*

NOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	123524.048	1	123524.048	2.099E4	.000[a]
	Residual	82.369	14	5.884		
	Total	123606.418	15			

a. Predictors: (Constant), Temperatur Aktual

b. Dependent Variable: Temperatur Observasi

Dari tabel terlihat bahwa sig=0.000< $\alpha = 0.5$. Artinya model adalah signifikan.

Tabel 4.1.2 adalah tabel summary dari model regresi :

Tabel 4.1.2 : Tabel Summary untuk model regresi dari estimator *classic*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	.999	.999	2.42560

a. Predictors: (Constant), Temperatur Aktual

b. Dependent Variable: Temperatur Observasi

Dari tabel terlihat bahwa $R^2 = 0.999$. Artinya 99.9% variabilitas dalam temperatur observasi dijelaskan oleh model regresi.

Tabel 4.1.3 adalah tabel coefficients dari model regresi :

Tabel 4.1.3 : Tabel Coefficients untuk model regresi dari estimator *classic*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.670	1.753		-3.806	.002
	Temperatur Aktual	.953	.007	1.000	144.896	.000

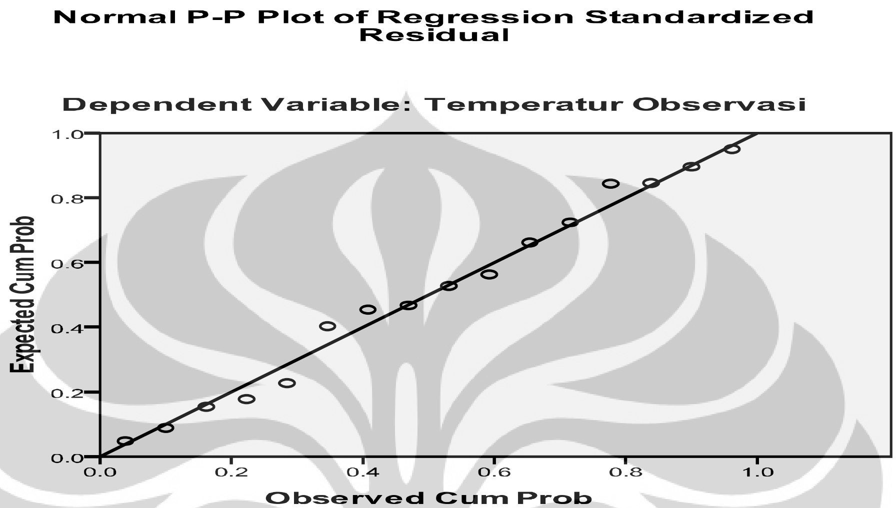
a. Dependent Variable: Temperatur Observasi

Dari tabel terlihat bahwa $\hat{\beta}_0 = -6.670$ dan $\hat{\beta}_1 = 0.953$.

Sehingga diperoleh model regresi :

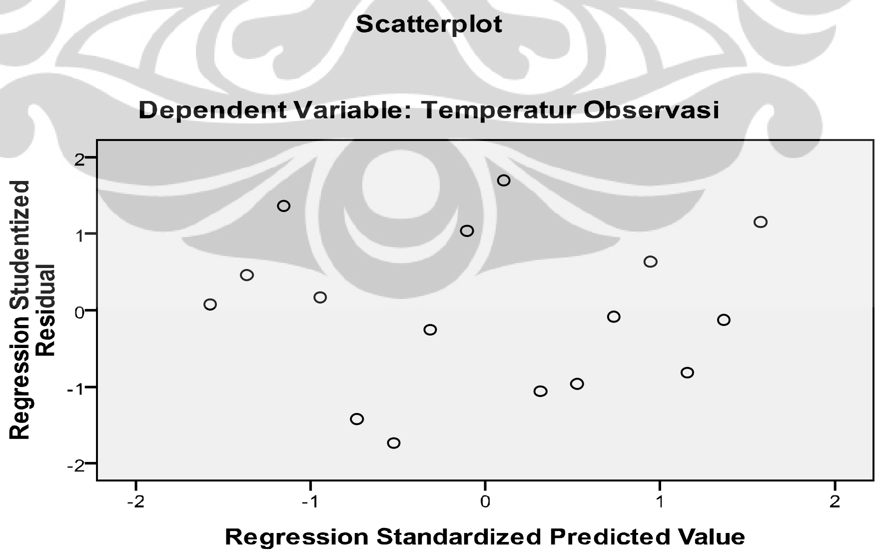
$$\hat{y} = -6.670 + 0.953x$$

Selanjutnya akan diperiksa asumsi normal dari error dengan memeriksa probability plot residual :



Gambar 4.1.1. : Gambar Normal Probability Plot Residuals Metode *Classic*

Kemudian akan diperiksa asumsi Homoskedastis dari hubungan antara nilai yang diprediksi dengan studentized residualnya :



Gambar 4.1.2. : Gambar Scatterplot SRESID vs ZPRED Metode *Classic*

Untuk melihat apakah ada nilai yang mempengaruhi model, akan dilihat melalui Studentized Residuals, Cook's D pada tabel berikut :

Tabel 4.1.4 : Tabel Residuals Statistics model regresi dari estimator *classic*

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Stud. Residual	-1.736	1.696	.008	1.021	16
Cook's Distance	.000	.196	.059	.063	16

a. Dependent Variable: Temperatur Observasi

Pada tabel terlihat bahwa :

- Studentized Residuals : Nilai tertinggi adalah 1.696, yang masih dibawah batas 2.5. Berarti tidak ada outliers.
- Cook's D : Nilai tertinggi adalah 0.196, yang masih dibawah batas 1. Berarti tidak ada outliers.

Secara keseluruhan model memenuhi asumsi regresi linier sehingga layak digunakan untuk prediksi.

Dari persamaan (3.2.3), estimator *classic* untuk x_0 pada $y_0 = 200°C$ adalah :

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{200 - (-6.670)}{0.953} = 216.86°C$$

Sedangkan 95% interval kepercayaan untuk $\hat{x}_0$ adalah

(211.278, 222.419)

4.2 METODE *INVERSE*

Untuk estimator *inverse* diperoleh model regresi linear sederhana :

$$x = \alpha_0 + \alpha_1 y + \varepsilon', \quad \varepsilon' \sim N(0, \sigma^2)$$

dimana parameter α_0 dan α_1 diestimasi dengan least squares sebagai $\hat{\alpha}_0$ dan $\hat{\alpha}_1$.

Tabel 4.2.1 adalah tabel Anova dari model regresi :

Tabel 4.2.1 : Tabel Anova untuk model regresi dari estimator *inverse*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	135909.372	1	135909.372	20994.897	.000[a]
	Residual	90.628	14	6.473		
	Total	136000.000	15			

a. Predictors: (Constant), Temperatur Observasi

b. Dependent Variable: Temperatur Aktual

Dari tabel terlihat bahwa sig=0.000< $\alpha = 0.5$. Artinya model adalah signifikan.

Tabel 4.2.2 adalah tabel summary dari model regresi :

Tabel 4.2.2 : Tabel Summary untuk model regresi dari estimator *inverse*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	.999	.999	2.54430

a. Predictors: (Constant), Temperatur Observasi

b. Dependent Variable: Temperatur Aktual

Dari tabel terlihat bahwa $R^2 = 0.999$. Artinya 99.9% variabilitas dalam temperatur aktual dijelaskan oleh model regresi.

Tabel 4.2.3 adalah tabel coefficients dari model regresi :

Tabel 4.2.3 : Tabel Coefficients untuk model regresi dari estimator *inverse*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.161	1.793		3.994	.001
	Temperatur Observasi	1.049	.007	1.000	144.896	.000

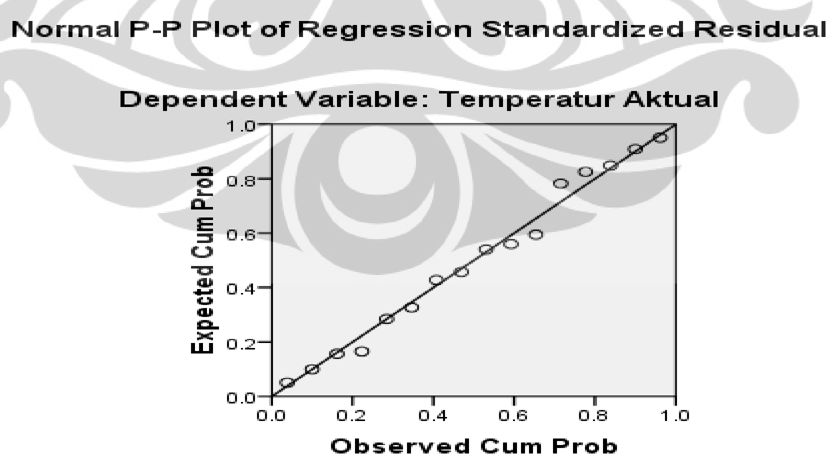
a. Dependent Variable: Temperatur Aktual

Dari tabel terlihat bahwa $\hat{\alpha}_0 = 7.161$ dan $\hat{\alpha}_1 = 1.049$.

Sehingga diperoleh model regresi :

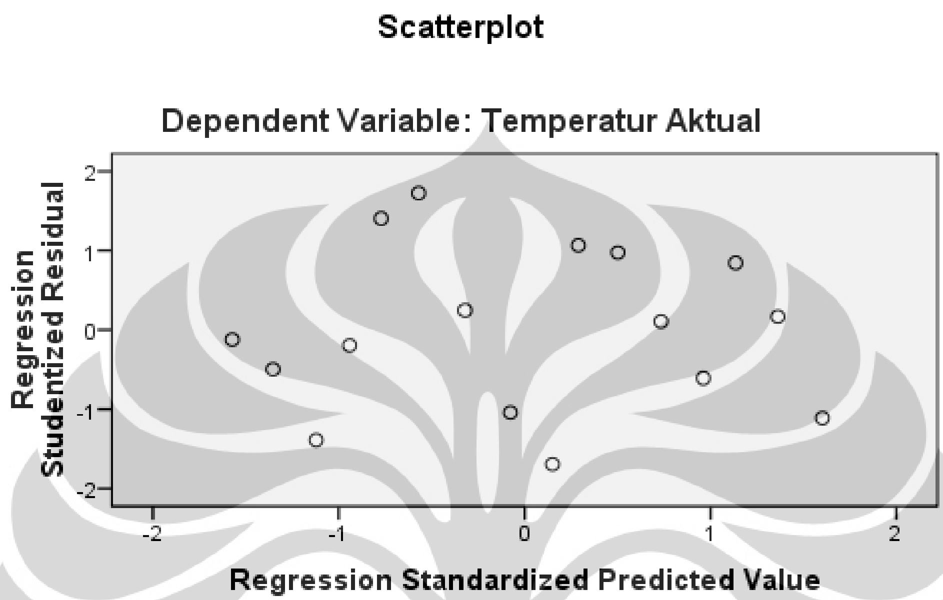
$$\hat{x} = 7.161 + 1.049y$$

Selanjutnya akan diperiksa asumsi normal dari error dengan memeriksa probability plot residual :



Gambar 4.2.1. : Gambar Normal Probability Plot Residuals Metode *Inverse*

Kemudian akan diperiksa asumsi Homoskedastis dari hubungan antara nilai yang diprediksi dengan studentized residualnya :



Gambar 4.2.2. : Gambar Scatterplot SRESID vs ZPRED Metode *Inverse*

Untuk melihat apakah ada nilai yang mempengaruhi model, akan dilihat melalui Studentized Residuals, Cook's D pada tabel berikut :

Tabel 4.2.4 : Tabel Residuals Statistics model regresi dari estimator *inverse*

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Stud. Residual	-1.694	1.724	-.008	1.021	16
Cook's Distance	.001	.189	.059	.062	16

a. Dependent Variable: Temperatur Aktual

Pada tabel terlihat bahwa :

- Studentized Residuals : Nilai tertinggi adalah 1.724, yang masih dibawah batas 2.5. Berarti tidak ada outliers.
- Cook's D : Nilai tertinggi adalah 0.059, yang masih dibawah batas 1. Berarti tidak ada outliers.

Secara keseluruhan, model memenuhi asumsi regresi linier sehingga layak digunakan untuk prediksi.

Dari persamaan (3.3.7), estimator *inverse* untuk x_0 pada $y_0 = 200°C$ adalah :

$$\hat{x}_0 = 7.161 + 1.049 \times 200 = 216,96°C$$

Sedangkan 95% interval kepercayaan untuk nilai aktual $\hat{x}_0$ adalah

(215.428, 218.327)

4.3 PERBANDINGAN

Berdasarkan persamaan (3.6.3), yaitu :

$$\frac{(x\_inverse - \bar{x})}{(x\_classic - \bar{x})} = R^2$$

Dimana $\bar{x} = 250$, diperoleh :

$$\frac{216.96 - 250}{216.86 - 250} = \frac{-33.04}{-33.14} = 0.9969$$

Hasil yang diperoleh sesuai dengan teori, yaitu $R^2 \leq 1$, sehingga estimator *Inverse* akan selalu lebih dekat dengan $\bar{x}$.

Dari interval kepercayaan juga diperoleh bahwa interval kepercayaan nilai aktual pada metode *inverse* lebih sempit daripada interval kepercayaan nilai aktual pada metode *classic.* Artinya interval kepercayaan nilai aktual pada metode *inverse* lebih presisi.

BAB V

KESIMPULAN DAN SARAN

Berdasarkan pembahasan pada bab-bab sebelumnya, diperoleh kesimpulan dan saran yang dituliskan dalam subbab 5.1 dan 5.2 berikut ini.

5.1 KESIMPULAN

Regresi Kalibrasi adalah proses kebalikan dari regresi linear sederhana.

Berdasarkan perbandingan kedua metode, metode *inverse* selalu lebih dekat dengan rata-rata nilai aktual, dan interval kepercayaan nilai aktual yang dihasilkan lebih presisi bila dibandingkan dengan metode *classic.*

5.2 SARAN

Saran yang dapat disampaikan dalam tugas akhir ini adalah diselidiki lebih lanjut mengenai sifat sifat estimator dalam memenuhi sifat statistika yaitu konsisten dan tidak bias.

DAFTAR PUSTAKA

Cheatle, Keith R. 2006. *Fundamentals of Test Measurement Instrumentation.* USA : Instrumentation, Systems, And Automation Society.

Chow, Shein C. 1990. *On the Difference Between the Classical and Inverse Methods of Calibration.* Technometrics.

Graybill, Franklin A. 1976. *Theory and Aplication of the Linear Model.* California: Duxbury Press.

Krutchkoff, R.G. 1967. *Classical and Inverse Methods of Calibration.* Technometrics.

Mendenhall, William. 1996. *A Second Course in Statistics: Regression Analysis*. New Jersey: Prentice-Hall.

Montgomery, Douglas C. 2001. *Introduction To Linear Regression Analysis.* New York: Wiley.

Nicholas, J. V. 2001. *Traceable temperature*. New York: Wiley.

Scroger, M G. 1989. *The Calibration Of Thermocouples.* USA: NIST Publication.

Smith, Harry. 1998. *Applied Regression Analysis.* New York: Wiley.

Tarantola, Albert. 2005. *Inverse Problem Theory*. Philadelphia: Siam.

LAMPIRAN 1

Pembuktian covariansi sub-bab 3.5.2

Bukti :

- $Cov(y_i, \hat{\gamma}_0) = Cov(y_i, \overline{y}) = Cov(y_i, \frac{1}{n}\sum_{j=1}^{n} y_j)$

$$= \frac{1}{n}\sum_{j=1}^{n} Cov(y_i, y_j) = \frac{1}{n}Var(y_i) = \frac{\sigma^2}{n}$$

- $Cov(\hat{\gamma}_0, \hat{\gamma}_1) = Cov(\frac{1}{n}\sum_{i=1}^{n} y_i, \sum_{j=1}^{n} c_j y_j), c_j = \frac{x_j - \overline{x}}{\sum_{k=1}^{n}(x_k - \overline{x})^2}$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}c_j Cov(y_i, y_j) = \sum_{i=1}^{n}\frac{1}{n}c_i Var(y_i) = \frac{\sigma^2}{n}\sum_{i=1}^{n} c_i = 0$$

- $Cov(y_i, \hat{\gamma}_1) = Cov(y_i, \sum_{j=1}^{n} c_j y_j) = \sum_{j=1}^{n} c_j Cov(y_i, y_j)$

$$= c_i Var(y_i) = \sigma^2 c_i = \frac{\sigma^2 (x_i - \overline{x})}{\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

- $Cov(y_i, \overline{y}_0) = Cov(y_i, \frac{1}{k}\sum_{i=n+1}^{n+k} y_i) = \frac{1}{k}Var(y_i) = \frac{\sigma^2}{k}$

LAMPIRAN 2

Pembuktian Teorema 3.5.3

Teorema 3.5.3

5. $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ *independent*

6. $\hat{\gamma}_0 \sim N\left(\gamma_0, \dfrac{\sigma^2}{n}\right)$ dan $\hat{\gamma}_1 \sim N\left(\hat{\gamma}_1, \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$

7. $U = (n+k-3)\dfrac{\hat{\sigma}^2}{\sigma^2}$ berdistribusi χ^2_{n+k-3}

8. $\hat{\sigma}^2$ tidak bergantung dengan $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)$

Bukti :

(1) Akan dibuktikan bahwa $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ saling bebas. Pertama, akan ditunjukkan bahwa $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ berdistribusi normal. Karena keduanya adalah jumlah dari bagian yang berdistribusi normal

$(\hat{\gamma}_0 = \bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$, $\hat{\gamma}_1 = \dfrac{\sum_{i=1}^{n} y_i(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sum_{i=1}^{n} c_i y_i)$, sehingga $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ berdistribusi normal juga.

Kedua, akan ditunjukkan bahwa keduanya tidak berkorelasi :

$$Cov(\hat{\gamma}_0, \hat{\gamma}_1) = Cov(\frac{1}{n}\sum_{i=1}^{n} y_i, \sum_{j=1}^{n} c_j y_j), c_j = \frac{x_j - \overline{x}}{\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}c_j Cov(y_i, y_j) = \sum_{i=1}^{n}\frac{1}{n}c_i Var(y_i) = \frac{\sigma^2}{n}\sum_{i=1}^{n} c_i = 0$$

Karena $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ tidak berkorelasi, maka $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ saling bebas.

(2) Diketahui bahwa $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ berdistribusi normal. Akan dihitung rata-rata dan variansi untuk keduanya.

> rata-rata $\hat{\gamma}_0$ adalah :

$$E(\hat{\gamma}_0) = E(\overline{y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}(\gamma_0 + \gamma_1(x_i - \overline{x})) = \gamma_0 + \frac{\gamma_1}{n}\sum_{i=1}^{n}(x_i - \overline{x})$$

$$= \gamma_0 + \frac{\gamma_1}{n}\left(\sum_{i=1}^{n} x_i - n\overline{x}\right) = \gamma_0$$

> variansi $\hat{\gamma}_0$ adalah :

$$Var(\hat{\gamma}_0) = Var(\overline{y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{\sigma^2}{n}$$

> rata-rata $\hat{\gamma}_1$ adalah :

$$E(\hat{\gamma}_1) = E\left(\frac{\sum_{i=1}^{n} y_i (x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right) = E\left(\sum_{i=1}^{n} c_i y_i \right), c_i = \frac{x_i - \overline{x}}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$= \sum_{i=1}^{n} c_i E(y_i) = \sum_{i=1}^{n} c_i (\gamma_0 + \gamma_1 (x_i - \overline{x})) = \gamma_1 \sum_{i=1}^{n} c_i (x_i - \overline{x}) = \gamma_1$$

> variansi $\hat{\gamma}_1$ adalah :

$$Var(\hat{\gamma}_1) = Var\left(\frac{\sum_{i=1}^{n} y_i (x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right) = Var\left(\sum_{i=1}^{n} c_i y_i \right), c_i = \frac{x_i - \overline{x}}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$= \sum_{i=1}^{n} c_i^2 Var(y_i) = \sum_{i=1}^{n} c_i^2 \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

(3) Misalkan

$S_1^2 = \sum_{i=1}^{n} \left(y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \overline{x}) \right)^2$ dan $S_2^2 = \sum_{i=n+1}^{n+k} (y_i - \overline{y}_0)^2$

sehingga $\hat{\sigma}^2 = \dfrac{1}{n+k-3} \left(S_1^2 + S_2^2 \right)$.

Berdasarkan teorema L.2.A, bahwa $U_1 = \frac{S_1^2}{\sigma^2}$ berdistribusi χ^2_{n-2}

($S_1^2 = (n-2)\hat{\sigma}^2$,$U_1 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$) dan berdasarkan teorema L.2.B.(2),

bahwa $U_2 = \frac{S_2^2}{\sigma^2}$ berdistribusi χ^2_{k-1}. S_1^2 hanya bergantung pada $y_1,...,y_n$ dan S_2^2 hanya bergantung pada $y_{n+1},...,y_{n+k}$. Karena diasumsikan y_i mutually independent , berarti S_1^2 dan S_2^2 saling bebas, sehingga U_1 dan U_2 juga saling bebas. Berdasarkan teorema L.2.C, $U = U_1 + U_2$ berdistribusi χ^2_{n+k-3} dimana

$$U = U_1 + U_2 = \frac{S_1^2}{\sigma^2} + \frac{S_2^2}{\sigma^2} = (n+k-3)\frac{\hat{\sigma}^2}{\sigma^2}$$

(4) Akan ditunjukkan bahwa $\hat{\sigma}^2$ tidak bergantung dengan $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)$. Dari lampiran 1 diketahui bahwa :

- $Cov(y_i, \hat{\gamma}_0) = \frac{\sigma^2}{n}$
- $Cov(y_i, \hat{\gamma}_1) = \frac{\sigma^2 (x_i - \overline{x})}{\sum_{k=1}^{n} (x_k - \overline{x})^2}$

Akan dihitung Covarian berikut :

- $Cov(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x}), \hat{\gamma}_0) = Cov(y_i, \hat{\gamma}_0) - Var(\hat{\gamma}_0) = \frac{\sigma^2}{n} - Var(\overline{y})$

$$= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

- $$Cov(y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x}), \hat{\gamma}_1) = Cov(y_i, \hat{\gamma}_1) - Cov(\hat{\gamma}_1(x_i - \overline{x}), \hat{\gamma}_1)$$

$$= \frac{\sigma^2 (x_i - \overline{x})}{\sum_{k=1}^{n} (x_k - \overline{x})^2} - (x_i - \overline{x}) Var(\hat{\gamma}_1)$$

$$= \frac{\sigma^2 (x_i - \overline{x})}{\sum_{k=1}^{n} (x_k - \overline{x})^2} - \frac{\sigma^2 (x_i - \overline{x})}{\sum_{k=1}^{n} (x_k - \overline{x})^2} = 0$$

Didapat $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$ dan $\hat{\gamma}_0$ tidak berkorelasi, oleh karena itu $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$ dan $\hat{\gamma}_0$ saling bebas. Karena S_1^2 merupakan fungsi dari $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$, S_1^2 dan $\hat{\gamma}_0$ juga saling bebas.

Didapat juga bahwa $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$ dan $\hat{\gamma}_1$ tidak berkorelasi, oleh karena itu $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$ dan $\hat{\gamma}_1$ saling bebas. Karena S_1^2 merupakan fungsi dari $y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \overline{x})$, S_1^2 dan $\hat{\gamma}_1$ juga saling bebas.

Karena S_2^2 hanya bergantung pada k nilai y_i terakhir, yaitu $y_{i+1}, ..., y_{n+k}$, maka S_2^2 serta S_1^2, $\hat{\gamma}_0$, $\hat{\gamma}_1$ dan $\overline{y}_0$ mutually independent, sehingga $\hat{\gamma}_0$, $\hat{\gamma}_1$ dan $\overline{y}_0$ juga bebas dari $S_1^2 + S_2^2$ dan dari $\hat{\sigma}^2$. Sementara $\hat{x}_0$ hanya bergantung pada $\hat{\gamma}_1$ dan $\overline{y}_0$. Dengan demikian didapat $\hat{\sigma}^2$ tidak bergantung dengan $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)$.

Teorema L.2.A

Misalkan $Y = X\gamma + \varepsilon$, ε berdistribusi normal dengan rata-rata 0 dan variansi σ^2, adalah suatu model linear dengan Y suatu peubah acak dengan $n \times 1$ vector observasi. X suatu $n \times p$ ($n > p$) dari suatu nilai yang diketahui dan X memiliki full rank. γ suatu $p \times 1$ vector dari parameter yang tidak diketahui. ε suatu $n \times 1$ dari peubah acak. Maka

$(n-p)\dfrac{\hat{\sigma}^2}{\sigma^2} = U$ berdistribusi χ^2_{n-p}

Bukti :

$$\begin{aligned}
(n-p)\cdot\hat{\sigma}^2 = SS_{Res} &= \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \\
&= \left(Y - \hat{Y}\right)'\left(Y - \hat{Y}\right) \\
&= \left[Y\text{-}X(X'X)^{-1}X'Y\right]'\left[Y\text{-}X(X'X)^{-1}X'Y\right] \\
&= Y'\left[I - X(X'X)^{-1}X'\right]Y
\end{aligned}$$

Akan ditunjukkan bahwa $\left[I - X(X'X)^{-1}X'\right]$ adalah *symmetric idempotent*:

$$\begin{aligned}
\left[I - X(X'X)^{-1}X'\right]\cdot\left[I - X(X'X)^{-1}X'\right] &= I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' \\
&= I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X' \\
&= I - X(X'X)^{-1}X'
\end{aligned}$$

dan

$$\left[I - X(X'X)^{-1}X'\right]' = I' - (X(X'X)^{-1}X')' = I - X(X'X)^{-1}X'$$
$$= I - X(X'X)^{-1}X'$$

Jika $Y \sim N(\mu, \sigma^2 I)$ dan A idempotent dengan rank q, maka :

$$\frac{Y'AY}{\sigma^2} \sim \chi^2_{q,\lambda}$$

dengan $\lambda = \dfrac{\mu' A \mu}{\sigma^2}$

Karena $Y \sim N(\mu, \sigma^2 I)$ dan $\left[I - X(X'X)^{-1}X'\right]$ adalah *idempotent*, maka :

$$\frac{SS_{Res}}{\sigma^2} = \frac{1}{\sigma^2} Y'\left[I - X(X'X)^{-1}X'\right]Y$$

Berdistribusi chi-square, dengan derajat bebasnya adalah rank dari $\left[I - X(X'X)^{-1}X'\right]$, yaitu *trace* dari $\left[I - X(X'X)^{-1}X'\right]$, karena rank dari matriks *idempotent* adalah *trace*-nya.

$$\begin{aligned} \text{trace}(\left[I - X(X'X)^{-1}X'\right]) &= \text{trace}(I_n) - \text{trace}([X(X'X)^{-1}X']) \\ &= n - \text{trace}(X'X(X'X)^{-1}) \\ &= n - \text{trace}(I_p) \\ &= n - p \end{aligned}$$

Diketahui bahwa $E(Y) = \mu = X\gamma$, sehingga

$$\begin{aligned} \lambda = \frac{1}{\sigma^2}\mu'[I - X(X'X)^{-1}X']\mu &= \frac{1}{\sigma^2}\gamma'X'[I - X(X'X)^{-1}X']X\gamma \\ &= \frac{1}{\sigma^2}\gamma'[X'X - X'X(X'X)^{-1}X'X]\gamma \\ &= \frac{1}{\sigma^2}[X'X - X'X]\gamma = 0 \end{aligned}$$

Akhirnya didapat

$$\frac{SS_{Res}}{\sigma^2} = (n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$$

Teorema L.2.B

Misalkan $y_1, y_2 ..., y_n$ adalah sampel acak yang berasal dari distribusi normal dengan rata-rata μ dan variansi σ^2. Maka

1. U_0 berdistribusi χ^2_n, dimana $U_0 = \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2}$
2. U_1 berdistribusi χ^2_{n-1}, dimana $U_1 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\sigma^2}$

Bukti :

(1) Karena y_i berdistribusi normal dengan rata-rata μ dan variansi σ^2,

$\frac{y_i - \mu}{\sigma}$ berdistribusi normal standar. Sehingga $\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2} \sim \chi^2_n$

(2) Misalkan $V_1 = \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2} = V_2 + V_3$

Dimana $S^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)}$

Diketahui $\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ dan $\frac{n(\bar{y} - \mu)^2}{\sigma^2} \sim \chi^2_1$ serta saling independent.

Jadi

$$M_{V_2}(t) = \frac{M_{V_1}(t)}{M_{V_3}(t)} = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}}$$

$$= (1-2t)^{-\frac{n-1}{2}}$$

$M_{V_2}(t)$ adalah Moment Generating Function (MGF) dari χ^2_{n-1}.

maka $\dfrac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \dfrac{(y_i - \overline{y})^2}{\sigma^2} \sim \chi^2_{n-1}$

Teorema L.2.C

Misalkan $X \sim \chi^2_n$ dan $Y \sim \chi^2_p$ serta X dan Y independent, maka $X + Y \sim \chi^2_{n+p}$

Bukti :

$$M_X(t) = (1-2t)^{-\frac{n}{2}}$$

$$M_Y(t) = (1-2t)^{-\frac{p}{2}}$$

$$M_{X+Y}(t) = (1-2t)^{-\frac{n}{2}} \cdot (1-2t)^{-\frac{p}{2}} = (1-2t)^{-\frac{n+p}{2}}$$

$M_{X+Y}(t)$ adalah Moment Generating Function (MGF) dari χ^2_{n+p}

LAMPIRAN 3

Pembuktian persamaan (3.6.1)

Bukti :

Pada estimasi nilai aktual dengan metode *classic*, diketahui :

persamaan regresi kalibrasi (3.2.2), adalah $\hat{x} = \dfrac{y - \hat{\beta}_0}{\hat{\beta}_1}$

dengan $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ dan $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$

maka estimator nilai aktual bila diberikan nilai observasi y_0 (3.2.3), adalah

$$
\begin{aligned}
x_{classic} = \hat{x}_0 &= \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{y_0 - \left(\bar{y} - \hat{\beta}_1\bar{x}\right)}{\hat{\beta}_1} = \frac{\hat{\beta}_1 + \left(y_0 - \bar{y}\right)}{\hat{\beta}_1} \\
&= \bar{x} + \frac{\left(y_0 - \bar{y}\right)}{\hat{\beta}_1} = \bar{x} + \frac{\left(y_0 - \bar{y}\right)}{\left(\dfrac{S_{xy}}{S_{xx}}\right)} \\
&= \bar{x} + \left(y_0 - \bar{y}\right) \times \frac{S_{xx}}{S_{xy}}
\end{aligned}
$$

LAMPIRAN 4

Pembuktian persamaan (3.6.2)

Bukti :

Pada estimasi nilai aktual dengan metode *inverse*, diketahui :

persamaan regresi kalibrasi (3.3.6), adalah $\hat{x} = \hat{\alpha}_0 + \hat{\alpha}_1 y$

dengan $\hat{\alpha}_0 = \overline{x} - \hat{\alpha}_1 \overline{y}$ dan $\hat{\alpha}_1 = \dfrac{S_{xy}}{S_{yy}}$

maka estimator nilai aktual bila diberikan nilai observasi y_0 (3.3.7), adalah

$$
\begin{aligned}
x_{inverse} = \hat{x}_0 = \hat{\alpha}_0 + \hat{\alpha}_1 y_0 &= \left(\overline{x} - \hat{\alpha}_1 \overline{y}\right) + \hat{\alpha}_1 y_0 \\
&= \overline{x} + (y_0 - \overline{y})\hat{\alpha}_1 \\
&= \overline{x} + (y_0 - \overline{y})\frac{S_{xy}}{S_{yy}}
\end{aligned}
$$

LAMPIRAN 5

Pembuktian persamaan (3.6.3)

Bukti :

$$\frac{(x_{inverse} - \overline{x})}{(x_{classic} - \overline{x})}, \text{ substitusikan persamaan (3.6.1) dan (3.6.2)}$$

$$\frac{(x_{inverse} - \overline{x})}{(x_{classic} - \overline{x})} = \frac{\left(\left(\overline{x} + (y_0 - \overline{y})\frac{S_{xy}}{S_{yy}}\right) - \overline{x}\right)}{\left(\left(\overline{x} + (y_0 - \overline{y}) \times \frac{S_{xx}}{S_{xy}}\right) - \overline{x}\right)} = \frac{(y_0 - \overline{y})\frac{S_{xy}}{S_{yy}}}{(y_0 - \overline{y}) \times \frac{S_{xx}}{S_{xy}}} = \frac{S_{xy}^{2}}{S_{xx} S_{yy}} = R^2$$

LAMPIRAN 6

Pembuktian persamaan (3.6.4)

Bukti :

Dari persamaan (3.6.3) diketahui $\frac{(x_{inverse} - \bar{x})}{(x_{classic} - \bar{x})} = R^2$, maka

$$
\begin{aligned}
(x_{inverse} - \bar{x}) &= R^2(x_{classic} - \bar{x}) = x_{classic} \times R^2 - \bar{x} \times R^2 \\
x_{inverse} &= \bar{x} + x_{classic} \times R^2 - \bar{x} \times R^2 = (1-R^2)\bar{x} + x_{classic} \times R^2, \text{bagi kedua ruas dengan } x_{classic} \\
\frac{x_{inverse}}{x_{classic}} &= \frac{(1-R^2)\bar{x}}{x_{classic}} + R^2 = (1-R^2)\left(\frac{\bar{x}}{x_{classic}} - 1\right) + 1 \\
\frac{x_{inverse}}{x_{classic}} &= 1 + (1-R^2)\left(\frac{\bar{x}}{x_{classic}} - 1\right)
\end{aligned}
$$