PENDUGAAN JUMLAH LIMFOSIT *ACTUAL* BERDASARKAN NILAI *OUTPUT* ALAT *HEMATOLOGY ANALYZER* MENGGUNAKAN REGRESI INVERS

Helma, Husni

Jurusan Matematika FMIPA UNP, Padang
Akademi Kebidanan Alifah Padang
helma667@yahoo.co.id

ABSTRACT

The development of science and technology led to the discovery of the new tools that can simplify the work. To facilitate counting the number of lymphocytes in the blood sample can be used hematology analyzer tool. Because the factors of production equipment can sometimes count the number of lymphocytes (output) is lower or higher than that calculated manually (actual. If the output value issued by the hematology analyzer tool is not the same as the actual value, it may cause an error in the interpretation of the circumstances relating to the value of the data. This problem can be solved by the inverse regression. Thus, the problem in this research is " What form of estimating the value of the actual number of lymphocytes in the blood sample based on the value of output hematology analyzer using inverse regression?" The conclusion of the research, the value of the actual number of lymphocytes can be estimated both classical as well as inverse

Keywords: *Hematology Analyzer, Actual Value, Output Value, Inverse Regression*

PENDAHULUAN

Perkembangan ilmu dan teknologi menyebabkan adanya temuan terhadap alat-alat baru yang dapat mempermudah pekerjaan. Contohnya alat pengukur tekanan darah, baik digital maupun non digital, alat-alat laboratorium, alat penghitung jumlah sel darah otomatis, dan lain-lain. Karena faktor produksi peralatan tersebut kadang-kadang menghasilkan keluaran (*output*) lebih rendah atau lebih tinggi dari yang seharusnya (*actual*/nyata). Untuk itu, sebelum peralatan tersebut digunakan harus dilakukan telaah terhadap *output* yang dihasilkan.

Apabila nilai *output* yang dikeluarkan oleh sebuah alat pengukuran tidak sama dengan nilai *actual* (tidak akurat), tentulah dapat menyebabkan suatu kekeliruan dalam penafsiran keadaan yang berhubungan dengan nilai pengukuran. Jika hal ini tidak diketahui dan si pengukur harus meng ambil keputusan, maka dapat mengakibat kan keputusan yang dihasilkan tidak akan sesuai dengan permasalahan yang akan diselesaikan.

Salah satu contoh dari masalah tersebut adalah pada alat penghitung jumlah sel darah otomatis (*hemotology analyzer*). Pada pengukuran terhadap suatu sampel darah, jumlah limfosit yang dikeluarkan oleh alat (nilai *output*) kadang-kadang lebih tinggi dari jumlah limfosit apabila dihitung secara manual (nilai *actual*). Jika nilai *output* yang dihasilkan lebih tinggi dari nilai *actual*-nya maka dokter/perawat yang menggunakan alat tersebut akan membuat keputusan yang tidak sesuai dengan keadaan pasien. Misalnya, berdasarkan pengukuran dengan alat tersebut diputuskan jumlah limfosit pasien berada pada kategori normal, padahal sebenarnya jumlah limfosit pasien tersebut berada pada kategori rendah. Tentulah penyakit pasien tidak teratasi sesuai dengan masalahnya.

Untuk mengatasi permasalahan, diperlukan suatu cara agar alat tetap dapat digunakan tetapi keputusan dari nilai *output* yang dihasilkan dapat diinterpretasikan sesuai dengan nilai *actual*. Cara ini berkait an dengan adanya suatu transformasi informasi yang menghubungkan kedua nilai tersebut.

Fenomena yang ada pada persoalan adalah nilai *actual* mempengaruhi nilai *output*, bukanlah nilai *output* mempengaruhi nilai *actual*. Hal ini dapat dikatakan bahwa terjadi hubungan sebab akibat (hubungan kausal) antara kedua keadaan. Jika hubungan tersebut bersifat linear maka menurut Montgomery (2006: 12) hubungan dapat ditelaah salah satunya dengan menggunakan analisis regresi linear sederhana

Pada analisis regresi linear seder hana, hubungan antar variabel dikaitkan dengan suatu persamaan yang disebut dengan model regresi linear sederhana. Model tersebut (Montgomery, 2006: 12) adalah $y = \beta_0 + \beta_1 x + \varepsilon$. Karena nilai *actual* mempengaruhi nilai *output*, maka dalam masalah ini nilai *output* merupakan variabel respon y dan nilai *actual* merupakan variabel regressor x.

Nilai parameter β_0 dan β_1 pada model dapat diduga berdasarkan data sampel. Jika b_0 merupakan penduga β_0 dan b_1 merupakan penduga β_1, maka diperoleh model regresi linear dugaan, yaitu $\hat{y} = b_0 + b_1 x$. Model regresi linear dugaan menunjukkan bahwa nilai *output* dapat diduga berdasarkan nilai *actual*.

Berdasarkan model regresi linear dugaan, dapat diketahui nilai *output* jika nilai *actual* diketahui. Karena salah satu kegunaan dari regresi adalah untuk prediksi, maka dapat dibentuk suatu interval terhadap nilai yang diduga. Interval dari nilai yang diduga tersebut dinamakan dengan *prediction interval* (Montgomery, 2006: 33). Apabila yang akan diduga adalah nilai *output* maka akan diperoleh *prediction interval* untuk nilai *output* yang berkaitan dengan suatu nilai *actual* diketahui. Begitu pula halnya, apabila yang akan diduga adalah nilai *actual* maka akan diperoleh *prediction interval* untuk nilai *actual* yang berkaitan dengan suatu nilai *output* diketahui

Pada permasalahan di atas, yang akan diduga adalah nilai *actual* berdasarkan nilai *output*. Akibatnya, model regresi tidak dapat digunakan secara langsung. Menurut Montgomery (2006: 488), hal ini dapat diselesaikan dengan pendekatan kebalikan dari regresi yang disebut dengan *inverse estimation*. *Inverse estimation* merupakan invers dari regresi linear sederhana.

Secara umum, terdapat dua metode dalam menyelesaikan invers dari regresi linear sederhana, yaitu metode klasik dan metode invers (Ryan, 2007: 257). Metode klasik memandang secara terbalik dari analisis regresi linear, sedangkan metode invers menukarkan peranan variabel regressor menjadi variabel respon dan variabel respon menjadi variabel regressor dari analisis regresi linear sederhana (Parker, 2010).

Pada metode klasik, untuk menduga suatu nilai *actual* x_o yang berkaitan dengan nilai *output* y_o digunakan model regresi linear dugaan $\hat{y}_o = b_0 + b_1 x_o$. Akibatnya, diperoleh bentuk pendugaan untuk variabel regressor x_o , yaitu

$$\hat{x}_o = \frac{y_o - b_0}{b_1}$$

Sedangkan pada metode invers, untuk menduga suatu nilai *actual* x_o yang berkaitan dengan nilai *output* y_o digunakan model dugaan $\hat{x}_o = a_0 + a_1 y_o$.

Model $\hat{x} = a_0 + a_1 y$ merupakan model dugaan dari bentuk model regresi linear sederhana $x = \alpha_0 + \alpha_1 y + \varepsilon'$. Dalam hal ini, a_0 merupakan penduga α_0 dan a_1 merupakan penduga α_1 .

Berdasarkan kedua metode kebalikan dari regresi linear sederhana, maka hal ini dapat digunakan untuk menduga nilai *actual* yang berkaitan dengan nilai *output* yang dihasilkan oleh suatu peralatan. Sehingga, keakuratan *output* suatu peralatan bukan lagi menjadi suatu permasalahan, karena informasi dari peralatan tersebut dapat ditransformasi sesuai dengan keadaan nilai *actual*-nya.

Metode klasik dan metode invers mungkin dapat memberikan dugaan yang berbeda terhadap nilai x_o yang tidak diketahui. Begitu pula untuk *prediction interval* nilai x_o, terdapat perbedaan (Parker, 2010).

Untuk itu, perlu suatu telaah terhadap hasil yang diberikan oleh alat ukur. Sehingga, masalah pada penelitian ini adalah "Apa bentuk pendugaan nilai *actual* jumlah limfosit dalam suatu sampel darah berdasarkan nilai *output hematology analy zer* dengan menggunakan regresi invers?"

METODE PENELITIAN

Pada penelitian ini, pendekatan yang digunakan adalah analisis teori analisis regresi balikan (*inverse*) dan diterapkan pada penafsiran jumlah limfosit yang dihasilkan alat *hemotology analyzer*. Data berasal dari 50 sampel darah. Sampel darah diambil berdasarkan *accidental sampling*. Alat *hemotology analyzer* yang digunakan adalah Coulter STKS

Adapun langkah-langkah yang dilaku kan untuk menjawab permasalahan yang dikemukakan adalah sebagai berikut. Pertama, menelaah tentang masalah yang terjadi pada alat *hemotology analyzer*, yaitu dalam menentukan sifat data pengamatan pada alat (nilai *output*) dan sifat data pengamatan secara manual (nilai *actual*). Kedua, menelaah asumsi-asumsi apakah yang dipenuhi oleh data yang mengacu kepada analisis regresi linear sederhana. Ketiga, menentukan model regresi linear yang cocok untuk data tersebut. Keempat, menentukan parameter-parameter pada model. Sehingga, dapat ditentukan nilai *actual* dan *prediction interval* dengan menggunakan metode klasik Kelima, menentukan jumlah limfosit nilai *actual* dan *prediction interval* dengan meng gunakan metode invers.

HASIL DAN PEMBAHASAN

Misal terdapat n sampel hasil pengukuran dari suatu alat. Nilai yang dikeluarkan oleh alat disebut dengan nilai *output* dari alat dan nilai yang sebenarnya disebut dengan nilai *actual*.

Variabel Respon adalah Nilai *Output* Dari Alat

Hubungan nilai *actual* dan nilai *output* dapat ditunjukkan dengan menggunakan model regresi linear sederhana

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

Jika dilihat dari nilai variabel respon y dan nilai regressor x maka kedua nilai variabel ini terlihat bervariasi. Untuk nilai variabel respon y yang sama dapat berkaitan dengan nilai variabel regressor x yang berbeda, begitu pula sebaliknya. Akibatnya, apabila diketahui suatu nilai *output* (nilai variabel respon y) maka tentulah yang diinginkan suatu nilai *actual* (nilai variabel regressor x). Untuk itu, pada kajian ini dilakukan pemusatan terhadap variabel regressor x .

a. Pemusatan Terhadap Variabel Regressor X

Pemusatan terhadap variabel regressor x dilakukan dengan mengguna kan $(x - \bar{x})$. Pemusatan ini merupakan langkah-langkah dalam Metode Graybill (Thonnard, 2006). Jika dilakukan pemu satan tersebut maka diperoleh

$$y_i = \gamma_0 + \gamma_1 (x_i - \bar{x}) + \varepsilon_i$$

Maka $\varepsilon_i = y_i - \gamma_0 - \gamma_1 (x_i - \bar{x})$

Untuk itu, akan ditentukan nilai $\hat{\gamma}_0$ dan $\hat{\gamma}_1$ sebagai pendugaan parameter γ_0 dan γ_1 dengan menggunakan metode kuadrat terkecil.

Jumlah kuadrat galat adalah

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [\, y_i - \gamma_0 - \gamma_1 (x_i - \bar{x}) \,]^2$$

Estimator kuadrat terkecil dari γ_0 dan γ_1, yaitu $\hat{\gamma}_0$ dan $\hat{\gamma}_1$, haruslah memenuhi

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

Karena $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ maka $\hat{\gamma}_0 = \bar{y}$

dan $\hat{\gamma}_1 = \dfrac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

Dengan demikian, model prediksinya adalah

$$\hat{y}_i = \bar{y} + \hat{\gamma}_1 (x_i - \bar{x})$$

dimana $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ dan $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

b. Pendugaan Untuk Nilai *Actual*

Pada suatu pengukuran dengan alat, seringkali nilai *actual* tidak diketahui, tetapi nilai *output* dapat diketahui dari keluaran alat. Misalkan terdapat sampel berukuran $n + k$. Maka pendugaan untuk x_0 adalah

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

dengan $\hat{\gamma}_1 \neq 0$ dan $\bar{y}_0 = \frac{1}{k}\sum_{i=n+1}^{n+k} y_i$

Pada suatu pengukuran, untuk satu nilai observasi baru y_0 akan sama dengan pendugaan nilai variabel x_0 pada metode klasik, sehingga $\bar{y}_0 = y_0$. Akibatnya, $\hat{\gamma}_1 = \hat{\beta}_1$. Dengan demikian,

$$\hat{\beta}_1 (\hat{x}_0 - \bar{x}) = y_0 - \bar{y}$$

Karena $\hat{\beta}_0 = \bar{y} - \hat{\gamma}_1 \bar{x}$, maka

$$\hat{\beta}_1 \hat{x}_0 = y_0 - \hat{\beta}_0$$

Sehingga, diperoleh pendugaan untuk nilai *actual* adalah

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \bar{x} + \frac{\bar{y}_0 - \hat{\gamma}_0}{\hat{\gamma}_1}$$

Karena nilai *actual* bervariasi maka akan ditentukan interval untuk nilai *actual* tersebut. Untuk menentukan interval nilai *actual* diperlukan hal yang berikut ini.

c. Pendugaan Variansi (σ^2)

Misalkan terdapat sampel berukuran n + k. Maka untuk n sampel, diperoleh sisaan atau residualnya adalah

$$e_i = y_i - \hat{y}_i = y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (\hat{x}_0 - \bar{x})$$

Untuk k sampel berikutnya, yaitu nilai observasi y berukuran $k \geq 1$ pada x_0 yang tidak diketahui, jumlah kuadrat sisaannya adalah

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

Maka diperoleh

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

Untuk satu nilai observasi baru y_0 akan ditentukan prediksi nilai *actual* x_0. Karena nilai *actual* bervariasi maka dapat dipandang hanya untuk satu nilai observasi baru y_0, yaitu $k = 1$. Sehingga,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n+1} e_i^2}{n-2}$$

Karena hanya terdapat satu data nilai observasi baru y_0 , maka $\bar{y}_0 = y_{n+1}$. Akibatnya,

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

Parameter-parameter yang telah diduga di atas dengan penduga $\hat{\gamma}_0$, $\hat{\gamma}_1$, dan $\hat{\sigma}^2$ merupakan penduga tak bias linear terbaik (*Best Linear Unbiased Estimator / BLUE*).

d. Prediction *Interval* (Selang Prediksi) Untuk Nilai *Actual*

Berdasarkan yang telah dikemukakan di atas, yaitu untuk suatu nilai observasi baru y_0 akan ditentukan prediksi nilai *actual* x_0. Dalam hal ini, $y_i = \bar{y}_0$. Akibatnya,

$$\text{Var}\,[\,\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1 (x_0 - \bar{x})\,] =$$

$$\sigma^2 + \frac{\sigma^2}{n} + (x_0 - \bar{x})^2 \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} =$$

$$\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]$$

Perhatikan bahwa $P(-t_{\alpha/2,(n-2)} \leq T \leq t_{\alpha/2,(n-2)}) = 1 - \alpha$. Dengan timgkat kepercayaan $(1 - \alpha)$ 100% , maka

$$\left[\frac{\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1 (x_0 - \bar{x})}{\hat{\sigma}\, A} \right]^2 \leq t^2_{\alpha/2,(n-2)}$$

$$[\,\bar{y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1 (x_0 - \bar{x})\,]^2 -$$

$$\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right] t^2_{\alpha/2,(n-2)} \leq 0$$

dengan $A^2 = 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

Akibatnya,

$$\left[\hat{\gamma}_1^{\,2} - \frac{\hat{\sigma}^2 t^2_{\alpha/2,(n-2)}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right] (x_i - \bar{x})^2 -$$

$$2\,\hat{\gamma}_1\,(\bar{y}_0 - \hat{\gamma}_0)\,(x_0 - \bar{x}) +$$

$$\left[(\bar{y}_0 - \hat{\gamma}_0)^2 - \hat{\sigma}^2 t^2_{\alpha/2,(n-2)} \left(1 + \frac{1}{n} \right) \right] \leq 0$$

Sehingga, diperoleh prediksi nilai *actual* x_0 untuk selang kepercayaan $(1 - \alpha)$ 100% sebagai berikut.

$$(x_0 - \bar{x}) = \frac{\hat{\gamma}_1 (\bar{y}_0 - \hat{\gamma}_0)}{a} \pm \frac{\hat{\sigma} t_{\alpha/2,(n-2)} \sqrt{\left\{ a \left(1 + \frac{1}{n} \right) + \frac{(\bar{y}_0 - \hat{\gamma}_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right\}}}{a}$$

dengan $a = \hat{\gamma}_1^{\,2} - \frac{\hat{\sigma}^2 t^2_{\alpha/2,(n-2)}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$,

Variabel Respon adalah Nilai *Actual*

Nilai dugaan variabel regresor dapat diperoleh dengan menukarkan peranan variabel regresor menjadi variabel respon, begitu pula sebaliknya. Apabila meng gunakan model regresi linier x atas y, haruslah dipenuhi asumsi bahwa y merupakan suatu nilai tertentu dan bukan variabel acak, dan juga x haruslah dipenuhi asumsi bahwa x merupakan variabel acak.

Model regresi linear sederhana x atas y adalah $x = \alpha_1 + \alpha_2\, y + \varepsilon'$. Jika diberikan nilai untuk variabel x, misalkan x_o (tidak diketahui) dan dapat diamati nilai y yang berkorespondensi dengan x_o , katakanlah y_o. Dugaan untuk x_o adalah

$$\hat{x}_o = a_0 + a_1 y_o$$

dimana $a_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

Sama halnya dengan regresi linear seder hana y atas x , penduga parameter a_0 dan a_1 diperoleh dengan menggunakan metode kuadrat terkecil seperti yang telah dijelas kan pada kajian teori.

Pendugaan interval untuk nilai sebenarnya x_o dengan tingkat kepercayaan $(1-\alpha)$ 100% dapat pula dihitung dengan menggunakan

$$\hat{x}_o - t_{\alpha/2,n-2} \sqrt{MS_{Res} \left(1 + \frac{1}{n} + \frac{(y_o - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \right)}$$

$$\leq x_o \leq \hat{x}_o +$$

$$t_{\alpha/2,n-2}\sqrt{MS_{Res}\left(1+\frac{1}{n}+\frac{(y_o-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)}$$

Data berikut ini merupakan data dari sampel darah yang berasal dari 50 orang pasien di rumah sakit. Pada sampel darah tersebut dihitung jumlah sel-sel darah per-100 leukosit. Sel-sel darah jenis leukosit yang dihitung pada penelitian ini adalah limfosit. Banyaknya sel darah limfosit dalam leukosit berdasarkan nilai rujukan sel darah adalah 17% sampai 40%.

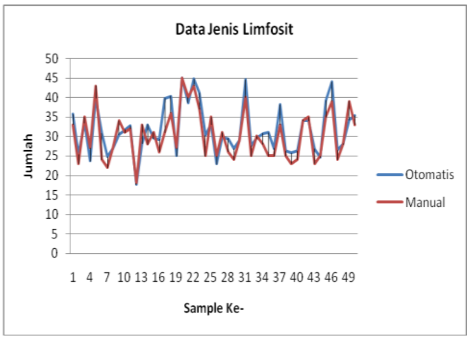
Sel-sel darah limfosit biasanya di hitung secara manual. Karena adanya alat *hemotology analyzer* Coulter STKS untuk menghitung sel-sel darah, maka sel-sel darah dihitung secara otomatis. Data sel darah limfosit yang diukur/dihitung secara manual dan otomatis dari 50 sampel darah dapat dilihat pada Tabel 1.

Tabel 1. Jumlah Limfosit dalam Sampel Darah yang Dihitung Secara Otomatis (O) dan Manual (M)

No.	O	M	No.	O	M
1	35.7	33	14	33	28
2	25.4	23	15	29.5	31
3	33.5	35	16	29.2	26
4	23.7	27	17	39.8	31
5	39.8	43	18	40.3	36
6	30.8	24	19	25	27
7	24.8	22	20	45	45
8	27	27	21	38.6	40
9	30.5	34	22	44.8	43
10	31.7	31	23	41.2	37
11	32.7	32	24	30.3	25
12	17.7	18	25	33.1	35
13	28.1	33	26	22.9	25

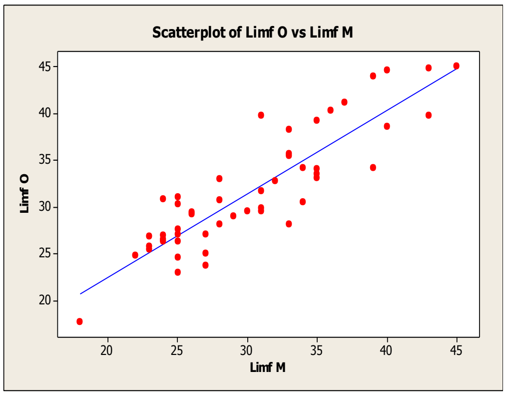
No.	O	M	No.	O	M
27	29.8	31	39	25.8	23
28	29.4	26	40	26.3	24
29	26.9	24	41	34.1	34
30	29	29	42	34	35
31	44.6	40	43	26.8	23
32	27.6	25	44	24.6	25
33	29.5	30	45	39.2	35
34	30.7	28	46	44	39
35	31	25	47	26.5	24
36	27	25	48	28.1	28
37	38.2	33	49	34.2	39
38	26.3	25	50	35.4	33

Data sel darah limfosit dibuat gambar pencaran data (plot data) agar dapat dilihat kecenderungan data. Adapun plot data untuk kedua jenis sel darah dapat dilihat pada Gambar 1



Gambar 1. Plot Data Limfosit

Pada plot data terlihat bahwa pengukuran secara otomatis memberikan hasil yang lebih tinggi daripada pengukuran secara manual. Data sel darah limfosit digambarkan hubungannya (*scatterplot*) antara data yang berasal dari pengukuran secara manual dan secara otomatis. Hal ini bertujuan untuk melihat pola hubungan yang terjadi antara pengukuran secara manual dan secara otomatis. Adapun *scatterplot* dapat dilihat pada Gambar 2.



Gambar 2. *Scatterplot* Data Limfosit

Pada *scatterplot* data jenis limfosit terlihat bahwa terdapat hubungan linear antara data hasil pengukuran secara otomatis dengan data hasil pengukuran secara manual. Lebih lanjut lagi, terlihat adanya suatu penambahan konstanta pada hubungan tersebut. Hal ini ditunjukkan oleh garis regresi linear yang tidak mendekati titik O.

Berdasarkan hal di atas, yaitu terdapatnya hubungan linear antara pengukuran secara otomatis dengan pengukuran secara manual, maka regresi linear sederhana dapat digunakan untuk analisis hubungan selanjutnya. Adapun hasil regresi linear sederhana pada kedua data tersebut dapat dilihat pada Tabel 2

Tabel 2. Hasil Analisis Regresi Linear Data Sel Darah Limfosit

No.	Sel Darah Limfosit
1.	Persamaan Regresi Linear Limf O = 4.64 + 0.893 Limf M
2.	Penduga Parameter $\hat{\beta}_0$ = 4.64 , nilai P = 0.034 $\hat{\beta}_1$ = 0.893 , nilai P = 0.000
3.	R-Sq = 77.8%
4.	Model regresi linear signifikan Nilai F = 168.14 , nilai P = 0.000
5.	Durbin-Watson statistic = 1.73511

Pada Tabel 2, dapat dilihat bahwa nilai F cukup besar dan nilai P dari masing-masing jenis sel darah adalah 0.000. Hal ini berarti untuk menunjukkan hubungan kedua pengukuran tersebut, model regresi linear dapat digunakan (signifikan) pada masing-masing jenis sel darah.

Model regresi pada jenis sel darah limfosit adalah Limf O = 4.64 + 0.893 Limf M. Dalam hal ini, $\hat{\beta}_0$ = 4.64 dengan nilai P = 0.034 dan $\hat{\beta}_1$ = 0.893 dengan nilai P = 0.000. Hal ini berarti, parameter β_0 tidak dapat dikatakan memiliki pengaruh pada model tersebut, yaitu tidak memiliki pengaruh terhadap penghitungan secara otomatis. Untuk parameter β_1, dapat dikatakan bahwa terdapat hubungan linear antara data hasil pengukuran secara manual dengan data hasil pengukuran secara otomatis dengan taraf kesalahan 1%. Model ini 77.8% dapat menggambarkan tentang hubungan pengukuran tersebut.

Nilai $\hat{\beta}_1$ pada jenis sel darah limfosit adalah 0.893 $\approx$ 0.90. Dengan demikian, alat penghitung sel darah otomatis tersebut mempunyai kekonsistenan dalam perban dingan nilai pengamatan.

Sebelum menggunakan model regresi yang telah diperoleh, maka semua asumsi dasar harus diuji terlebih dahulu. Apabila semua asumsi dasar telah dipenuhi oleh model, maka model dapat digunakan. Berikut ini akan diuji semua asumsi dasar yang harus dipenuhi oleh model regresi linear.

a. Hubungan Antara Respon Y dengan Regressor X Adalah Linear

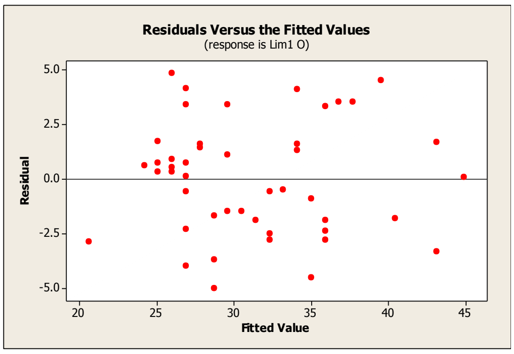
Seperti yang telah dikemukakan sebelumnya, bahwa terdapat hubungan linear antara data hasil pengukuran secara manual dengan data hasil pengukuran secara otomatis.

b. Rataan dari Galat adalah Nol**.**

Galat ε_i diduga dengan mengguna kan sisaan (*residual*) $e_i = y_i - \hat{y}_i$. Rataan sisa adalah nol. Hal ini berarti, asumsi rataan dari galat adalah nol sudah terpenuhi.

c. Variansi σ^2 dari Galat adalah Konstan

Apabila dibuat plot antara sisaan e_i dengan nilai dugaan (*fitted value*) $\hat{y}_i$ pengukuran secara otomatis untuk masing-masing jenis sel darah maka diperoleh Gambar 3.



Gambar 3. Plot Residuals Versus TheFitted Values

Karena sisaan termuat pada *a horizontal band* dan tersebar secara acak (tidak berpola), maka dapat dikatakan bahwa variansi σ^2 dari galat adalah konstan. Dengan demikian, tidak ada keraguan terhadap kecacatan model.

d. Galat Tidak Berkorelasi

Uji asumsi terhadap galat tidak berkorelasi dapat dilakukan dengan melihat nilai Durbin-Watson statistic. Pada jenis sel darah limfosit diperoleh nilai Durbin-Watson statistic-nya adalah 1.73511. Karena nilainya dekat ke 2, maka dapat disimpulkan bahwa galat tidak berkorelasi.

e. Galat Berdistribusi Normal

Untuk menguji kenormalan galat digunakan Uji Anderson-Darling. Uji tersebut dilakukan terhadap sisaan. Karena nilai P yang diperoleh adalah 0.704, maka galat berdistribusi normal.

Secara keseluruhan, semua asumsi dasar telah dipenuhi oleh model regresi linear yang telah dibentuk. Untuk itu, model tersebut dapat digunakan untuk analisa selanjutnya.

Berdasarkan persamaan

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \bar{x} + \frac{\bar{y}_0 - \hat{\gamma}_0}{\hat{\gamma}_1}$$

dapat diduga nilai *actual* atau nilai pengukuran secara manual. Berikut ini akan diberikan contoh untuk menentukan nilai pengukuran secara manual jumlah limfosit apabila diketahui nilai pengukuran secara otomatis pada suatu sampel darah dengan menggunakan persamaan tersebut.

Misalkan, nilai suatu pengukuran jumlah limfosit secara otomatis adalah y_0 = 35. Maka pendugaan nilai pengukuran jumlah limfosit secara manualnya (nilai *actual*) adalah

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = 33.998$$

Karena $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 30.28$

$$\hat{\gamma}_1 = \hat{\beta}_1 = 0.893$$

$$\hat{\gamma}_0 = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 31.662$$

$$\bar{y}_0 = y_0 = 35$$

maka dengan menggunakan $\hat{x}_0 = \bar{x} + \frac{\bar{y}_0 - \hat{\gamma}_0}{\hat{\gamma}_1}$ diperoleh $\hat{x}_0 = 34.018$

Karena penghitungan jumlah limfosit secara manual menghasilkan bilangan bulat maka diperoleh hasil, yaitu $\hat{x}_0 = 34$.

Hasil dugaan untuk penghitungan jumlah limfosit secara manual terlihat lebih kecil dari penghitungan jumlah limfosit secara otomatis. Jika diperhatikan Gambar 1, maka hal ini tidaklah salah. Plot data jenis limfosit memperlihatkan bahwa pengukuran secara otomatis memberikan hasil yang cenderung lebih tinggi daripada pengukuran secara manual.

Selanjutnya, diduga variansi

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

9.1 Pada tingkat kepercayaan 95% , maka selang pendugaan jumlah limfosit *actual* yang bersesuaian dengan jumlah limfosit yang dihitung secara otomatis y_0 = 35 adalah sebagai berikut ini.

Perhatikan $p = \hat{\gamma}_1^{\,2} - \frac{\hat{\sigma}^2 t^2_{\alpha/2,(n-2)}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$$p = 0.7793$$

Sehingga,

$$27.315 \leq x_0 \leq 40.895$$

Jadi, jika jumlah limfosit yang dihitung secara otomatis adalah y_0 = 35, maka nilai dugaan untuk jumlah limfosit yang dihitung secara manual adalah $\hat{x}_0$ = 34. Pada tingkat kepercayaan 95%, maka limfosit yang dihitung secara manual tersebut berjumlah antara 27.315 sampai 40.895.

Apabila masalah regresi tersebut dipandang secara terbalik, yaitu untuk mendapatkan *inverse estimator* dari x_0 , harus dibentuk terlebih dahulu model regresi linear sederhana dari nilai pengu kuran secara manual (x) terhadap nilai pengukuran secara otomatis (y). Model regresi linear sederhananya adalah

Limf M = 2.68 + 0.872 Limf O

Adapun hasil analisis lainnya dapat dilihat pada Tabel 3.

Tabel 3. Hasil Analisis Data Limf M = 2.68 + 0.872 Limf O

No.	Statistik
1.	Penduga Parameter $\hat{\beta}_0$ = 2.68 , nilai P = 0.222 $\hat{\beta}_1$ = 0.872 , nilai P = 0.000
2.	R-Sq = 77.8%
3.	Model regresi linear signifikan Nilai F = 168.14 , nilai P = 0.000
4.	Durbin-Watson statistic = 1.80526

Berdasarkan Tabel 3, dapat dikatakan bahwa model regresi linear tersebut dapat digunakan (signifikan) untuk menunjukkan hubungan kedua pengukuran, karena nilai F cukup besar dengan nilai P = 0.000. Dalam hal ini, $\hat{\beta}_0$ = 2.68 dengan nilai P = 0.222 dan $\hat{\beta}_1$ = 0.872 dengan nilai P = 0.000, yang berarti parameter β_0 tidak memiliki pengaruh terhadap penghitungan secara manual dan terdapat hubungan linear antara data hasil pengukuran secara otomatis dengan data hasil pengukuran secara manual. Model ini 77.8% dapat meng gambarkan tentang hubungan pengukuran tersebut.

Sebelum menggunakan model reg resi yang telah diperoleh, maka semua asumsi dasar harus diuji terlebih dahulu. Apabila semua asumsi dasar telah dipenuhi, maka model dapat digunakan. Berikut ini akan diuji semua asumsi dasar yang harus dipenuhi oleh model regresi linear Limf M = 2.68 + 0.872 Limf O.

Hasil uji semua asumsi dasar yang dilakukan seperti langkah di atas, maka semua asumsi dasar dipenuhi oleh model regresi linear yang telah dibentuk. Untuk itu, model tersebut dapat digunakan untuk analisa selanjutnya.

Berdasarkan persamaan regresi linear yang telah diperoleh, yaitu

Limf M = 2.68 + 0.872 Limf O

dapat diduga nilai pengukuran secara manual untuk nilai pengukuran jumlah limfosit secara otomatis y_0 = 35. Adapun nilainya adalah

$$\hat{x}_0 = 2.68 + (0.872)(35) = 33.2$$

Karena penghitungan jumlah limfosit secara manual menghasilkan bilangan bulat maka diperoleh hasil $\hat{x}_0$ = 33.

Variansi dugaan untuk model regresi linear ini adalah $\hat{\sigma}^2$ = 8.9. Sehingga, pada tingkat kepercayaan 95%, selang dugaan jumlah limfosit yang dihitung secara manual yang bersesuaian dengan jumlah limfosit yang dihitung secara otomatis y_0 = 35 adalah sebagai berikut ini.

Perhatikan

$$t_{\alpha/2,n-2}\sqrt{MS_{Res}\left(1+\frac{1}{n}+\frac{(y_o-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)}$$

$$= (1.96)\sqrt{(8.9)\left(\frac{51}{50}+\frac{3.338^2}{50824.16}\right)}$$

$$= (1.96)\sqrt{(8.9)(1.020)}$$

$$= 5.906$$

Maka

$$27.294 \leq x_0 \leq 39.106$$

Jadi, jika jumlah limfosit yang dihitung secara otomatis adalah y_0 = 35, maka nilai dugaan untuk jumlah limfosit yang dihitung secara manual adalah $\hat{x}_0$ = 33. Pada tingkat kepercayaan 95%, maka limfosit yang dihitung secara manual tersebut berjumlah antara 27.294 sampai 39.106.

Apabila dibandingkan kedua hasil yang didapatkan dari metode di atas maka terlihat bahwa hasilnya sedikit berbeda. Keputusan akhir tergantung pada diagnosa dokter dan gejala penyakit yang dirasakan oleh pasien.

PENUTUP

Berdasarkan hasil penelitian diperoleh kesimpulan bahwa bentuk pendugaan nilai *actual* jumlah limfosit dalam suatu sampel darah berdasarkan nilai *output hematology analyzer* dengan menggunakan regresi invers adalah jika variabel respon adalah nilai *output* dari alat, maka

Limf O = 4.64 + 0.893 Limf M

Jika jumlah limfosit yang dihitung secara otomatis adalah $y_0 = 35$, maka nilai dugaan untuk jumlah limfosit yang dihitung secara manual adalah $\hat{x}_0 = 34$. Pada tingkat kepercayaan 95%, maka limfosit yang dihitung secara manual tersebut berjumlah antara 27.315 sampai 40.895

Jika variabel respon adalah nilai *actual*, maka

Limf M = 2.68 + 0.872 Limf O

Jika jumlah limfosit yang dihitung secara otomatis adalah $y_0 = 35$, maka nilai dugaan untuk jumlah limfosit yang dihitung secara manual adalah $\hat{x}_0 = 33$. Pada tingkat kepercayaan 95%, maka limfosit yang dihitung secara manual tersebut berjumlah antara 27.294 sampai 39.106

DAFTAR PUSTAKA

Chang, G.A., Kerns, G.J., Lee, Dj., dan Stanek, G.L. 2009. **Jurnal Volume Statistika Pendidikan 17**. Youngs town State University. www.amstat.org/publications/jse/v17n2/datasets.chang.html

Graybill, Franklin A., and Iyer, Hariharan K. 1962. **Regression Analysis Concepts and Applications**. Vol. 33. Jhon W. Turkey.

Krutchkoff, R. G. 1967. **Classical and Inverse Regression Methods of Calibration.** Technometrics, Vol. 9, No. 3, pp. 425-439. Published by: American Statistical Association and American Society for QualityStable. www.jstor.org/stable/1266511 .

Montgomery, D.C., Peck, E.A., dan Vining, G.G. 2006. **Introduction to Linear Regression Analysis**. New York: Jhon Wiley and Sons, Inc.

Ryan, T.P. 2007. **Modern Engineering Statistics.** New York: Jhon Wiley and Sons, Inc.

Sarwoko. 2007. **Statistik Inferensi, untuk Ekonomi dan Bisnis**. Yogyakarta: Andi.

Seber, Goerge A. F., and Lee, Alan J. 2003. **Linear Regression Analysis Second Edition**. Wiley Series In Probability and Statistics.

Sembiring. 1995. **Analisis Regresi**. Bandung: ITB Bandung.

Thonnard, M. 2006. **Convidence Intervals in Inverse Regression**. Technische Universiteit Endhoven, Department of Mathematics and Computer Science.

Qudratullah, M.F. 2013. **Analisis Regresi Terapan. Teori, Contoh Kasus, dan Aplikasi dengan SPSS.** Yogyakarta: Andi.