Analisis Matematis Model Proyeksi Penduduk Aritmatik, Geometrik, Eksponensial, dan Logistik

Sebuah Catatan Kuliah Pemodelan Lingkungan

Abdur Rohman, Ph.D

2024-06-24

Daftar Isi

Ada banyak model matematika yang digunakan para ilmuwan untuk keperluan proyeksi penduduk. Dalam catatan ini akan saya bahas empat model matematika yang sederhana yaitu model aritmatik, model geometrik, model eksponensial, dan model logistik. Model-model yang akan saya bahas ini saya sebut sederhana karena hanya mempertimbangkan hubungan antara dua variabel saja yaitu **waktu** dan **jumlah penduduk** pada waktu itu. Tidak dipertimbangkan faktor-faktor penyumbang jumlah penduduk yaitu kelahiran, kematian, dan migrasi pada setiap waktu. Juga tidak mempertimbangkan dinamika variabel-variabel itu berdasarkan usia, jenis kelamin, dan faktor demografis lainnya. Selain itu, model-model yang dibahas ini tidak mempertimbangkan adanya korelasi serial pada data penduduk. Jadi dalam model-model ini data penduduk pada satu waktu dengan waktu yang lain dianggap tidak berhubungan. Ini jelas asumsi yang kurang realistis.

Meski demikian, empat model ini tetap layak dibahas karena dapat membantu kita menghasilkan prediksi yang cukup tinggi akurasinya secara cepat dengan data yang terbatas. Selain itu, empat model ini saya bahas karena banyak digunakan dalam Teknik Lingkungan, di antaranya dalam perencanaan Sistem Penyediaan Air Minum. Dalam Peraturan Menteri Pekerjaan Umum nomor 18/PRT/M/2007 tentang Penyelenggaraan Pengembangan Sistem Penyediaan Air Minum pada Lampiran I dijelaskan empat model untuk memperkirakan proyeksi jumlah penduduk, yaitu model **arithmatik, geometrik, least square**, dan **trend logistic**.

Bila dilihat rumus matematika dari keempat model ini, dapat dipastikan bahwa rumus arithmatik dan least square dalam Permen PU 18/2007 ini pada dasarnya sama, yaitu sama-sama model linear. Perbedaannya adalah metode **arithmatik** itu menghitung nilai gradien dari dua titik data saja, sedangkan **least square** menghitung nilai gradien dari semua data. Jadi empat model di Permen PU 18/2007 itu dapat dikelompokkan menjadi tiga model saja, yaitu model linear, geometrik, dan logistik. Hanya saja, dalam rumus model trend logistic di Permen PU 18/2007 itu terdapat kesalahan tanda pada penyebutnya. Mestinya tandanya + tapi tertulis -.

Dalam BPS(2010) disebutkan tiga model penghihtungan proyeksi penduduk yaitu aritmatik, geometrik, dan eksponensial. Dari kedua referensi ini kita dapatkan empat model yaitu aritmatik (linear), geometrik, eksponensial, dan logistik.

Dalam bagian berikut ini akan dijelaskan asumsi dari tiap model dan penjelasn rumus matematika yang berkaitan dengan asumsi tersebut.

1 Asumsi pada Model Proyeksi Aritmatik

Model aritmatik memiliki asumsi bahwa **perubahan jumlah penduduk dalam satuan jiwa** dari tahun ke tahun bersifat tetap. Misalnya: Penduduk suatu wilayah diasumsikan bertambah sebanyak 100 jiwa tiap tahun secara konstan, tidak berubah pada tahun berapapun.

Misalkan P_t menunjukkan jumlah penduduk pada tahun t, P_0 menunjukkan jumlah penduduk awal, dan diasumsikan jumlah penduduk itu bertambah sebanyak N jiwa tiap tahun. Maka:

$$
\begin{aligned}
&P_1 = P_0 + N \\
&P_2 = P_1 + N = P_0 + N + N = P_0 + 2N \\
&P_3 = P_2 + N = P_0 + 3N
\end{aligned}
$$

Dari simulasi sederhana ini didapat:

$$P_t = P_0 + N \cdot t$$

Secara umum, ini identik dengan rumus persamaan linear:

$$y_x = c + mx$$

Kita dapat menggunakan metode least square untuk menemukan m dan c terbaik dari data penduduk BPS yang kita analisis. Penerapan metode least square itu dapat dilihat di lampiran. Hasil yang didapat sama dengan hasil pada rumus least square pada Permen PU nomor 18/2007 tersebut.

2 Asumsi Pada Model Proyeksi Penduduk Geometrik

Asumsi pada model geometrik adalah bahwa perubahan jumlah penduduk berlangsung dengan **persentase tetap** dari waktu ke waktu. Misalnya: Suatu kota jumlah penduduknya tiap tahun bertambah 1% dari jumlah penduduknya di tahun sebelumnya.

Misalkan P_t menunjukkan jumlah penduduk pada tahun t, P_0 menunjukkan jumlah penduduk awal, dan diasumsikan jumlah penduduk itu bertambah sebanyak r persen tiap tahun (dalam contoh di atas r = 5%). Maka:

$$\begin{aligned}
P_1 &= P_0 + r \cdot P_0 \\
&= P_0(1+r) \\
P_2 &= P_1 + r \cdot P_1 \\
&= P_1(1+r) = P_0(1+r)(1+r) = P_0(1+r)^2 \\
P_3 &= P_2 + r \cdot P_2 \\
&= P_2(1+r) \\
&= P_0(1+r)^2(1+r) \\
&= P_0(1+r)^3
\end{aligned}$$

Dari simulasi sederhana ini didapat:

$$P_t = P_0(1+r)^t$$

3 Asumsi Pada Model Proyeksi Penduduk Eksponensial dan Logistik

Misalkan P_t menunjukkan jumlah penduduk pada waktu t. Maka dP_t menunjukkan **perubahan penduduk** dan dt menunjukkan **perubahan waktu**. Perbandingan antara perubahan penduduk dan perubahan waktu $\frac{dP_t}{dt}$ menunjukkan **perubahan penduduk tiap waktu**.

Dalam model **eksponensial** perubahan penduduk tiap waktu ini diasumsikan terjadi dengan **persentase tetap** (Vandermeer, 2010). Misalnya persentase itu kita sebut $r_{eksponensial}$. Nilai persentase ini dalam model eksponensial tetap tanpa dipengaruhi jumlah penduduk.

Untuk mendapatkan nilai perubahan penduduk tiap waktu dalam jiwa tiap waktu pada waktu tertentu, persentase ini dikalikan jumlah penduduk pada waktu tertentu itu. Secara matematis *dalam model eksponensial*:

$$\frac{dP_t}{dt}|_{eksponensial} = r_{eksponensial} \cdot P_t = k \cdot P_t \tag{1}$$

dengan k adalah sebuah konstanta yang nilainya dapat diperoleh dari data penduduk dan tahun dengan menggunakan berbagai algoritma, di antaranya Gauss-Newton dan Leverberg-Marquardt.

Bila kita menggunakan **R** (R Core Team, 2023), algoritma Gauss-Newton dapat kita jalankan dengan fungsi `nls()`. Algoritma Leverberg-Marquardt dapat kita jalankan dengan fungsi `nlsLM` dari paket `minpack.lm` (Elzhof et.al., 2023).

Model logistik menggunakan asumsi bahwa perubahan penduduk sepanjang waktu terjadi dengan **persentase yang berubah secara linear seiring dengan perubahan penduduk** (Vandermeer, 2010). Artinya, persentase perubahan penduduk terus berkurang atau terus bertambah mengikuti berkurang atau bertambahnya penduduk.

4 Asumsi Pada Model Proyeksi Penduduk Logistik

Dalam model logistik nilai persentase perubahan penduduk (sebut saja $r_logistik$) diasumsikan nilainya terus berubah secara berbanding terbalik dengan jumlah penduduk.

Secara matematis (Vandermeer, 2010) :

$$\frac{dP_t}{dt}|_{logistik} = r_{logistik} \cdot P_t = (a - b \cdot P_t) \cdot P_t \tag{2}$$

dengan a dan b adalah suatu konstanta yang dapat diperoleh dari data.

Dengan asumsi ini, apabila jumlah penduduk (P_t) terus bertambah, maka nilai persentase perubahannya akan terus berkurang. Persentase pertumbuhan mendekati nol apabila a nilainya mendekati $b \cdot P_t$.

5 Rumus Jumlah Penduduk dalam Model Eksponensial dan Logistik

Kita dapat menggunakan cara *pemisahan variabel* untuk mendapatkan jumlah penduduk di setiap waktu, P_t.

5.1 Rumus Model Eksponensial

Persamaan Perhitungan 1 dapat diuraikan sebagai berikut.

$$\begin{aligned}
\frac{dP_t}{dt} &= k \cdot P_t \\
\frac{dP_t}{P_t} &= k \cdot dt \\
\int \frac{dP_t}{P_t} &= \int k \cdot dt \\
\ln P_t + c_1 &= k \cdot t + c_2 \\
\ln P_t &= k \cdot t + c_3 && \left(dengan\ c_3 = c_2 - c_1 \right) \\
P_t &= e^{k \cdot t + c_3} && \left(Ingat\ bila\ \ln y = x\ maka\ y = e^x \right) \\
P_t &= e^{k \cdot t} \cdot e^{c_3}
\end{aligned}$$

Saat $t = 0$, $P_t = e^0 \cdot e^{c_3} = e^{c_3}$. Ini berarti bahwa e^{c_3} nilainya sama dengan *penduduk pada tahun pertama* dalam data penduduk yang dianalisis. Misalkan kita simbolkan P_0. Maka:

$$P_t = P_0 \cdot e^{k \cdot t}$$

5.2 Rumus Model Logistik

Persamaamn Perhitungan 2 dapat diuraikan menjadi sebagai berikut.

$$\begin{aligned}
\frac{dP_t}{dt} &= [a - (b \cdot P_t)] \cdot P_t \\
\frac{dP_t}{(a - b \cdot P_t) \cdot P_t} &= dt \\
dP_t \cdot \frac{1}{(a - b \cdot P_t) \cdot P_t} &= dt
\end{aligned}$$

Bentuk di ruas kiri perlu dipecah sehingga menjadi lebih sederhana supaya mudah diintegralkan. Bentuk sederhananya:

$$\frac{1}{(a - b \cdot P_t) \cdot P_t} = \frac{k_1}{a - (b \cdot P_t)} + \frac{k_2}{P_t}$$

dengan k_1 dan k_2 adalah konstanta-konstanta yang harus kita cari nilainya dalam a dan b.

Untuk memahami perubahan bentuk ini, mari kita lihat contoh berikut.

$$\frac{2}{3} + \frac{4}{5} = \frac{2 \cdot 5 + 4 \cdot 3}{3 \cdot 5} = \frac{22}{3 \cdot 5}$$

Ini berarti

$$\frac{22}{3 \cdot 5} = \frac{k_1}{3} + \frac{k_2}{5}$$

Kita dapat menemukan k_1 dan k_2 yang sesuai yaitu dengan $5 \cdot k_1 + 3 \cdot k_2 = 22$. Dalam contoh ini didapat salah satu solusinya $k_1 = 2$ dan $k_2 = 4$.

$$\begin{aligned}
\frac{1}{[a - (b \cdot P_t)] \cdot P_t} &= \frac{k_1}{a - (b \cdot P_t)} + \frac{k_2}{P_t} \\
\frac{1}{[a - (b \cdot P_t)] \cdot P_t} &= \frac{k_1 \cdot P_t + (k_2 \cdot a) - (k_2 \cdot b \cdot P_t)}{[a - (b \cdot P_t)] \cdot P_t} \\
\frac{1}{[a - (b \cdot P_t)] \cdot P_t} &= \frac{[k_1 - (k_2 \cdot b)] \cdot P_t + (k_2 \cdot a)}{[a - (b \cdot P_t)] \cdot P_t} \\
1 &= [k_1 - (k_2 \cdot b)] \cdot P_t + k_2 \cdot a \\
0 \cdot P_t + 1 &= [k_1 - (k_2 \cdot b)] \cdot P_t + k_2 \cdot a
\end{aligned}$$

Dari sini kita dapatkan:

$$\begin{aligned}
k_2 \cdot a &= 1 \\
k_2 &= \frac{1}{a}
\end{aligned}$$

dan

$$\begin{aligned}
k_1 - (k_2 \cdot b) &= 0 \\
k_1 &= k_2 \cdot b \\
k_1 &= \frac{b}{a}
\end{aligned}$$

Dengan demikian, bentuk sederhananya adalah:

$$\begin{aligned}
\frac{1}{(a - b \cdot P_t) \cdot P_t} &= \frac{k_1}{a - (b \cdot P_t)} + \frac{k_2}{P_t} \\
&= \frac{b}{a} \cdot \frac{1}{a - (b \cdot P_t)} + \frac{1}{a} \cdot \frac{1}{P_t}
\end{aligned}$$

Bentuk ini lalu kita masukkan ke persamaan awal, lalu kita integralkan:

$$
\begin{aligned}
dP_t \cdot \frac{1}{(a - b \cdot P_t) \cdot P_t} &= dt \\
\frac{1}{a} \cdot \frac{1}{a - (b \cdot P_t)} \cdot dP_t + \frac{b}{a} \cdot \frac{1}{P_t} \cdot dP_t &= dt \\
\int \frac{b}{a} \cdot \frac{1}{a - (b \cdot P_t)} \cdot dP_t + \int \frac{1}{a} \cdot \frac{1}{P_t} \cdot dP_t &= \int dt \\
\frac{b}{a} \int \frac{1}{a - (b \cdot P_t)} \cdot dP_t + \frac{1}{a} \int \frac{1}{P_t} \cdot dP_t &= \int dt
\end{aligned}
$$

Bentuk integral di suku pertama ruas kiri bisa diuraikan dengan substitusi.

$$
\begin{aligned}
u &= a - (b \cdot P_t) \\
\frac{du}{dP_t} &= -b \\
dP_t &= -\frac{1}{b} \cdot du \\
\int \frac{1}{a - (b \cdot P_t)} \cdot dP_t &= \int \frac{1}{u} \cdot \left(-\frac{1}{b} \right) \cdot du \\
&= -\frac{1}{b} \int \frac{1}{u} \cdot du \\
&= -\frac{1}{b} \ln u + c_1 \\
&= -\frac{1}{b} \ln(a - b \cdot P_t) + c_1 \qquad \left(\textit{dengan syarat } a - b \cdot P_t > 0 \right)
\end{aligned}
$$

Dengan hasil ini kita bisa melanjutkan proses integral:

$$
\begin{aligned}
\frac{b}{a}\int \frac{1}{a-(b\cdot P_t)}\cdot dP_t + \frac{1}{a}\int \frac{1}{P_t}\cdot dP_t &= \int dt \\
\frac{b}{a}\left(-\frac{1}{b}\ln(a-(b\cdot P_t))+c_1\right)+\frac{1}{a}\ln P_t + c_2 &= t + c_3 \\
-\frac{1}{a}\ln(a-b\cdot P_t)+\frac{1}{a}\ln P_t &= t + c_4 && \left(\textit{dengan } c_4 = c_3 - c_2 - \frac{b}{a}c_1\right) \\
\frac{1}{a}\big[\ln P_t - \ln(a-(b\cdot P_t))\big] &= t + c_4 \\
\ln P_t - \ln(a-(b\cdot P_t)) &= at + c_5 && \left(\textit{dengan } c_5 = a\cdot c_4\right) \\
\ln\left(\frac{P_t}{a-(b\cdot P_t)}\right) &= at + c_5 && \left(\textit{Ingat:}\ \ln x - \ln y = \ln\left(\frac{x}{y}\right)\right) \\
\frac{P_t}{a-(b\cdot P_t)} &= e^{at+c_5} \\
P_t &= (a\cdot e^{at+c_5}) - (b\cdot P_t\cdot e^{at+c_5}) \\
P_t\left(1 + b\cdot P_t\cdot e^{at+c_5}\right) &= a\cdot e^{at+c_5} \\
P_t &= \frac{a\cdot e^{at+c_5}}{1+b\cdot e^{at+c_5}}
\end{aligned}
$$

Persamaan ini dapat disederhanakan dengan mengalikan pembilang dan penyebutnya masing-masing dengan $b\cdot e^{at+c_5}$ sehingga menjadi:

$$
\begin{aligned}
P_t &= \frac{a/b}{(1/b)\cdot e^{-at-c_5}+1} \\
P_t &= \frac{a/b}{(e^{-c_5}/b)\cdot e^{-at}+1}
\end{aligned}
$$

Karena masing-masing dari a, b, dan c_5 adalah konstanta, kita bisa tulis dengan lebih sederhana menjadi

$$
P_t = \frac{K}{1+f\cdot e^{g\cdot t}}
$$

dengan K, f, dan g adalah kontanta-konstanta yang nilainya dapat diperoleh dari data penduduk dan tahun menggunakan berbagai algoritma, di antaranya Gauss-Newton dan Levenberg Marquardt.

Dalam rumus logistik ini K menunjukkan nilai batas jumlah penduduk yang tidak akan dilampaui sampai kapanpun. Ini bisa dipahami dari limit berikut.

$$\lim_{t\to\infty} \frac{K}{1+f\cdot e^{g\cdot t}} = \frac{K}{1+0} = K \qquad (untuk \;\; g \leq 0)$$
$$\lim_{t\to\infty} \frac{K}{1+f\cdot e^{g\cdot t}} = \frac{K}{1+\infty} = 0 \qquad (untuk \;\; g > 0)$$

Nilai f dapat dipahami dari perbandingan antara K dan penduduk di tahun awal P_0 (yaitu penduduk saat $t = 0$).

$$P_0 = \frac{K}{1+f\cdot e^0}$$
$$1+f = \frac{K}{P_0}$$
$$f = \frac{K}{P_0} - 1$$

Nilai g menunjukkan nilai persentase di awal pertumbuhan penduduk, yang kemudian terus berkurang bila penduduk terus bertambah. Ini bisa dilihat dari $r_{logistik} = a - by$ dan dalam analisis di atas didapat $g = a$. Nilai g negatif bila jumlah penduduk terus bertambah sepanjang waktu. Nilai g positif bila jumlah penduduk terus berkurang sepanjang waktu.

6 Referensi

Kementerian Pekerjaan Umum (2007). Peraturan Menteri Pekerjaan Umum nomor 18/PRT/M/2007 tentang Penyelenggaraan Pengembangan Sistem Penyediaan Air Minum. https://jdih.pu.go.id/internal/assets/assets/produk/PermenPUPR/2007/06/permen18-2007.pdf https://peraturanpedia.id/peraturan-menteri-pekerjaan-umum-dan-perumahan-rakyat-nomor-18-prt-m-2007/

BPS (Badan Pusat Statistik) (2010). Pedoman Penghitungan Proyeksi Penduduk dan Angkatan Kerja https://media.neliti.com/media/publications/50042-ID-pedoman-penghitungan-proyeksi-penduduk-dan-angkatan-kerja.pdf

Vandermeer, J. (2010) How Populations Grow: The Exponential and Logistic Equations. Nature Education Knowledge 3(10):15 https://www.nature.com/scitable/knowledge/library/how-populations-grow-the-exponential-and-logistic-13240157/

R Core Team (2023). R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Elzhov, T.V., Mullen, K.M., Spiess, AN, Bolker, B. (2023). minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. https://CRAN.R-project.org/package=minpack.lm

https://www.pnas.org/doi/pdf/10.1073/pnas.6.6.275

Rumus Metode Numerik Ordinary Least Square pada Regresi Linear

Abdur Rohman, Ph.D

2023-10-18

Contoh Data

Metode OLS tanpa Matriks

$$
\begin{aligned}
E &= \sum_{k=1}^{n} \epsilon_k^2 = \sum_{k=1}^{n} (y_{prediksi_k} - y_{data_k})^2 \\
&= \sum_{k=1}^{n} (mx_k + c - y_k)^2 \\
&= (mx_1 + c - y_1)^2 + (mx_2 + c - y_2)^2 + ... + (mx_n + c - y_n)^2 \\
&= (mx_1 + c)^2 - 2(mx_1 + c)y_1 + y_1^2 + \\
&\quad (mx_2 + c)^2 - 2(mx_2 + c)y_2 + y_2^2 + ... \\
&\quad + (mx_n + c)^2 - 2(mx_n + c)y_n + y_n^2 \\
&= m^2x_1^2 + 2mx_1c + c^2 - 2mx_1y_1 - 2cy_1 + y_1^2 \\
&\quad + m^2x_2^2 + 2mx_2c + c^2 - 2mx_2y_2 - 2cy_2 + y_2^2 + ... \\
&\quad + m^2x_n^2 + 2mx_nc + c^2 - 2mx_ny_n - 2cy_n + y_n^2 \\
&= m^2(x_1^2 + x_2^2 + ... + x_n^2) + 2mc(x_1 + x_2 + ... + x_n) + c^2n \\
&\quad - 2m(x_1y_1 + x_2y_2 + ... + x_ny_n) - 2c(y_1 + y_2 + ... + y_n) \\
&\quad + y_1^2 + y_2^2 + ... + y_n^2 \\
&= m^2 \sum_{k=1}^{n} x_k^2 + 2mc \sum_{k=1}^{n} x_k + c^2n - 2m \sum_{k=1}^{n} x_ky_k \\
&\quad - 2c \sum_{k=1}^{n} y_k + \sum_{k=1}^{n} y_k^2 \\
&= m^2 \cdot S_1 + 2mc \cdot S_2 + c^2n - 2m \cdot S_3 - 2c \cdot S_4 + S_5
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial c} &= 0 \\
0 + 2mS_2 + 2cn - 0 - 2S_4 + 0 &= 0 \\
c &= \frac{S_4}{n} - m\frac{S_2}{n} \\
c &= \frac{\sum_{k=1}^{n} y_k}{n} - m\frac{\sum_{k=1}^{n} x_k}{n} \\
c &= y_{rata-rata} - m \cdot x_{rata-rata}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial m} &= 0 \\
2mS_1 + 2cS_2 + 0 - 2S_3 - 0 + 0 &= 0 \\
m &= \frac{S_3}{S_1} - c\frac{S_2}{S_1} \\
m &= \frac{S_3}{S_1} - \left(\frac{S_4}{n} - m\frac{S_2}{n}\right)\left(\frac{S_2}{S_1}\right) \\
m &= \frac{nS_3}{nS_1} - \frac{S_4S_2}{nS_1} + m\frac{S_2^2}{nS_1} \\
m(nS_1 - S_2^2) &= nS_3 - S_4S_2 \\
m &= \frac{nS_3 - S_4S_2}{nS_1 - S_2^2} \\
m &= \frac{n \cdot \sum_{k=1}^{n} x_k y_k - \left(\sum_{k=1}^{n} y_k \cdot \sum_{k=1}^{n} x_k\right)}{n \cdot \sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2}
\end{aligned}
$$

Metode OLS dengan Matriks

$$
\begin{aligned}
r &= Y - X \cdot \beta \\
r^T &= (Y - X \cdot \beta)^T = Y^T - X^T \cdot \beta \\
r^T \cdot r &= (Y^T - X^T \cdot \beta)(Y - X \cdot \beta) \\
&= Y^T \cdot Y - Y^T \cdot X \cdot \beta - X^T \cdot \beta \cdot Y + X^T \cdot \beta \cdot X \cdot \beta
\end{aligned}
$$

Supaya nilai $r^T \cdot r$ sekecil mungkin, harus dipenuhi syarat turunan fungsi, yaitu turunan pertamanya terhadap β bernilai 0.

$$\begin{aligned}
\frac{\partial(r^T \cdot r)}{\partial\beta} &= 0 \\
\frac{\partial(Y^T \cdot Y)}{\partial\beta} - \frac{\partial(Y^T \cdot X \cdot \beta)}{\partial\beta} - \frac{\partial(X^T \cdot \beta \cdot Y)}{\partial\beta} + \frac{\partial(X^T \cdot \beta \cdot X \cdot \beta)}{\partial\beta} &= 0 \\
0 - Y^T \cdot X - X^T \cdot Y + X^T \cdot X \cdot 2\beta &= 0 \\
X^T \cdot X \cdot 2\beta &= Y^T \cdot X + X^T \cdot Y
\end{aligned}$$

Ruas kanan dapat disederhanakan menjadi $2X^T \cdot Y$ atau $2Y^T \cdot X$ karena $Y^T \cdot X = X^T \cdot Y$. Karena ini perkalian matriks, ruas kiri dapat dipindah ke kanan dengan cara mengalikan matriks selain β dengan inversnya dari kiri.

Maka:

$$\begin{aligned}
X^T \cdot X \cdot 2\beta &= 2X^T \cdot Y \\
X^T \cdot X \cdot \beta &= X^T \cdot Y \\
\left(X^T \cdot X\right)^{-1} \cdot (X^T \cdot X) \cdot \beta &= \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y \\
I \cdot \beta &= \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y \\
\beta &= \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y
\end{aligned}$$