Linear Algebra Ib:

Overdetermined Systems

and

Least-Squares Approximation

Overdetermined System

Say B is $m \times n$ with $m > n$, and

g is $m \times 1$:

can we find a z (n vector) such that

$$\underset{m\times n}{B}\;\underset{n\times 1}{z} \overset{?}{=} \underset{m\times 1}{g} \quad ?$$

Consider $m = 3$, $n = 2$:

$$\underset{(3)}{m} > \underset{(2)}{n}$$

$$\underset{3\times 2}{\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix}} \underset{2\times 1}{\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}} \overset{?}{=} \underset{3\times 1}{\begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}}$$

equate corresponding elements of vector Bz, and vector g

or

$$\left.\begin{aligned} B_{11} z_1 + B_{12} z_2 &= g_1 \\ B_{21} z_1 + B_{22} z_2 &= g_2 \\ B_{31} z_1 + B_{32} z_2 &= g_3 \end{aligned}\right\}$$

3 equations in 2 unknowns

(a) row perspective (example)

case I

$$\underset{B}{\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 2 & -3 \end{pmatrix}} \underset{z}{\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}} \overset{?}{=} \underset{g_{I}}{\begin{pmatrix} \frac{5}{2} \\ 2 \\ -2 \end{pmatrix}}$$

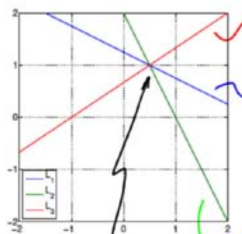
$$\begin{aligned} 1z_1 + 2z_2 &= \tfrac{5}{2} \\ 2z_1 + 1z_2 &= 2 \\ 2z_1 - 3z_2 &= -2 \end{aligned}$$

case II

$$\underset{B}{\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 2 & -3 \end{pmatrix}} \underset{z}{\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}} \overset{?}{=} \underset{g_{II}}{\begin{pmatrix} 0 \\ 2 \\ -4 \end{pmatrix}}$$

$$\begin{aligned} 1z_1 + 2z_2 &= 0 \\ 2z_1 + 1z_2 &= 2 \\ 2z_1 - 3z_2 &= -4 \end{aligned}$$

case I



$2x_1 - 3x_2 = -2$
(solutions of third equation)

$1x_1 + 2x_2 = \frac{5}{2}$
(solutions of first equation)

$2x_1 + 1x_2 = 2$
(solutions of second equation)

x_2

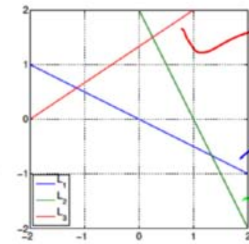
x_1

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

Solution of all three equations
$\Rightarrow$ solution of $Bz = g$
(but "unstable")

Note: any 2 equations suffice; 3rd equation is redundant, but not inconsistent.

case II



solutions of third equation

solutions of first equation

solutions of second equation

x_2

x_1

there is no point z which satisfies all three equations
$\Rightarrow$ no solution to $Bz = g$

Note: third equation is inconsistent with other two equations.

(b) column perspective (example)

case I

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3/2 \\ 0 \end{pmatrix}$$

$$B \quad z \quad \stackrel{?}{=} \quad g_I$$

or

$$\underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{col 1 of } B} z_1 + \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{\text{col 2 of } B} z_2 = \underbrace{\begin{pmatrix} 1 \\ 3/2 \\ 0 \end{pmatrix}}_{g_I}$$

case II

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3/2 \\ 2 \end{pmatrix}$$

$$B \quad z \quad \stackrel{?}{=} \quad g_{II}$$

or

$$\underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{col 1 of } B} z_1 + \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{\text{col 2 of } B} z_2 = \underbrace{\begin{pmatrix} 1 \\ 3/2 \\ 2 \end{pmatrix}}_{g_{II}}$$

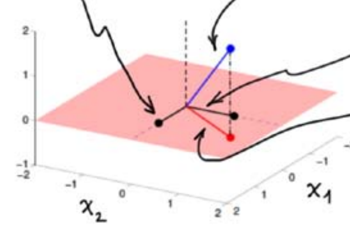
col 1 of B

g_{II}

col 2 of B

g_I

x_3



x_2

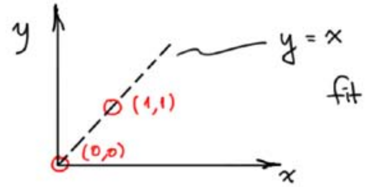
x_1

g_I can be expressed as linear combination (z) of columns of B
$\Downarrow$
$Bz = g_I$ has a solution
(but "unstable")

g_{II} can not be expressed as linear combination (z) of columns of B
$\Downarrow$
$Bz = g_{II}$ has no solution

(c) "Fitting" Perspective (Line to Data)

2 data points



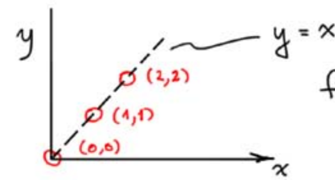
fit line to data
↳ $\beta_0^{true} + \beta_1^{true} x$

$$\beta_0^{true} + \beta_1^{true} \cdot 0 = 0 \qquad \text{first point on line}$$
$$\beta_0^{true} + \beta_1^{true} \cdot 1 = 1 \qquad \text{second point on line}$$

$\hookrightarrow \beta_0^{true} = 0, \; \beta_1^{true} = 1$

3 data points:

case I



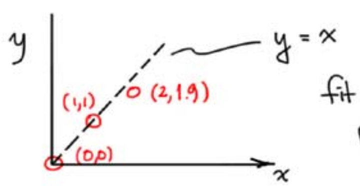
fit line to data
$\beta_0^{true} + \beta_1^{true} x$

$$\beta_0^{true} + \beta_1^{true} \cdot 0 = 0 \qquad \text{first point on line}$$
$$\beta_0^{true} + \beta_1^{true} \cdot 1 = 1 \qquad \text{second point on line}$$
$$\beta_0^{true} + \beta_1^{true} \cdot 2 = 2 \qquad \text{third point on line}$$

$$\left.\underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}}_{\substack{X \\ \text{"B"}}} \underbrace{\begin{pmatrix} \beta_0^{true} \\ \beta_1^{true} \end{pmatrix}}_{\substack{\beta^{true} \\ \text{"z"}}} = \underbrace{\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}}_{\substack{Y \\ \text{"g"}}} \right\} \qquad \beta_0^{true} = 0, \; \beta_1^{true} = 1$$

3 data points:

Case II



fit line to data
$\beta_0^{true} + \beta_1^{true} x$

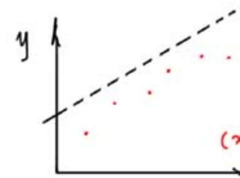
$$\beta_0^{true} + \beta_1^{true} \cdot 0 = 0 \qquad \text{first point on line ?}$$
$$\beta_0^{true} + \beta_1^{true} \cdot 1 = 1 \qquad \text{second point on line ?}$$
$$\beta_0^{true} + \beta_1^{true} \cdot 2 = 1.9 \qquad \text{third point on line ?}$$

$$\left.\underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_0^{true} \\ \beta_1^{true} \end{pmatrix}}_{\beta^{true}} \overset{?}{=} \underbrace{\begin{pmatrix} 0 \\ 1 \\ 1.9 \end{pmatrix}}_{Y} \right\}$$

NO SOLUTION
but somehow "close"

m data points

least squares



candidate $\underbrace{\beta_0 + \beta_1 x}_{Y_{model}(x;\beta)}$

$(x_i, Y_i), \; 1 \le i \le m$

Assume $Y_i = Y_{model}(x_i; \beta^{true})$ + "noise" for some β^{true}.

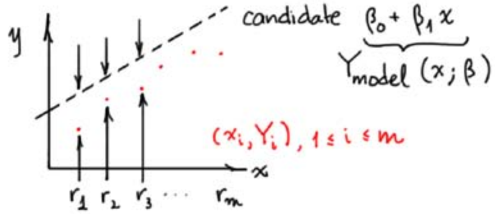
But
$$\begin{cases} \beta_0^{true} + \beta_1^{true} x_1 = Y_1 \\ \beta_0^{true} + \beta_1^{true} x_2 = Y_2 \\ \quad\vdots \\ \beta_0^{true} + \beta_1^{true} x_m = Y_m \end{cases} \quad \text{or} \quad \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_0^{true} \\ \beta_1^{true} \end{pmatrix}}_{\beta^{true}} = \underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}}_{Y}$$

has no solution. So...

How do we find an estimate $\hat{\beta}$ for β^{true}?

m data points

least squares



Define residuals

$$r_1(\beta) \equiv Y_1 - Y_{\text{model}}(x_1;\beta) = Y_1 - (\beta_0 + \beta_1 x_1)$$

$$r_2(\beta) \equiv Y_2 - Y_{\text{model}}(x_2;\beta) = Y_2 - (\beta_0 + \beta_1 x_2)$$

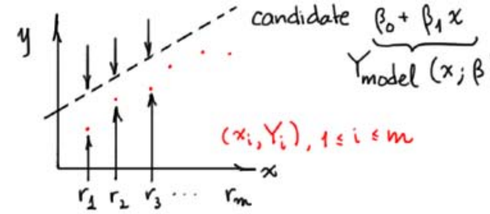
$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$r_m(\beta) = Y_m - Y_{\text{model}}(x_m;\beta) = Y_m - (\beta_0 + \beta_1 x_m)$$

and choose $\hat\beta$ to minimize (over all β) $\underbrace{\sum_{i=1}^{m} r_i^2}_{0 \text{ iff all points lie on a line}}$.

m data points

least squares



Note

$$\begin{array}{ccccccc} & m\times 1 & & m\times 1 & & m\times 2 & 2\times 1 \\ r(\beta) \equiv & \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} & \equiv & \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} & - & \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}}_{X} & \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\beta} \end{array}$$

$$\underbrace{\qquad X \quad \beta \qquad}_{X\beta}$$

so

$$r_i(\beta) = Y_i - (X\beta)_i, \quad 1 \le i \le m.$$

measurement at x_i model prediction at x_i

A General Linear Model (to which to fit data)

Let $(x_{(1)}, \ldots, x_{(p)})$ be our independent variables (p in total).

Let y be our dependent variable. predict in terms of

Let $h_j(x)$, $1 \le j \le n-1$, be prescribed functions.

Let β_j, $0 \le j \le n-1$, be (unknown) coefficients.

Then define
$$Y_{\text{model}}(x;\beta) = \beta_0 + \sum_{j=1}^{n-1} \beta_j h_j(x),$$
$$= \sum_{j=0}^{n-1} \beta_j h_j(x)$$

(for $h_0(x) = 1$).

We postulate that for some β, β^{true},

$$y = Y_{\text{model}}(x, \beta^{\text{true}})$$
$$= \sum_{i=0}^{n-1} \beta_j^{\text{true}} h_j(x).$$

Platonic/ideal

(or "noise-free" measurements, or $\mathbb{E}$(measurements), ...)

Note the model $Y_{\text{model}}(x;\beta)$ is

linear in β $\qquad \beta_2^2,\ e^{\beta_3}$ ✗

but not (necessarily)

linear in x $\qquad h_1(x) = 1/x_{(1)},\ e^{x_{(2)}}$ ✓.

matrix form

Given (x_i, Y_i), $1 \le i \le m$, (a vector of independent variables) $\quad Y_i = Y_{model}(x_i; \beta^{true}) + \text{"noise"}$

$$\underset{m\times 1}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}; \quad \underset{m\times n}{X} = \begin{pmatrix} 1 & h_1(x_1) & h_2(x_1) & \cdots & h_{n-1}(x_1) \\ 1 & h_1(x_2) & h_2(x_2) & & h_{n-1}(x_2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & h_1(x_m) & h_2(x_m) & \cdots & h_{n-2}(x_m) \end{pmatrix}.$$

$$X_{ij} = h_{j-1}(x_i) \quad 1 \le i \le m, \quad 1 \le j \le n$$

(Assume columns of X are independent.)

Note

$$r_1 = Y_1 - Y_{model}(x_1; \beta)$$

$$= Y_1 - \left(\beta_0 + \sum_{j=1}^{n-1} \beta_j h_j(x_1)\right) = Y_1 - \underbrace{(X\beta)_1}_{m\times 1} \text{ first component}$$

$$r_2 = Y_2 - (X\beta)_2$$

$$\vdots$$

or

$$\boxed{\underset{m\times 1}{r} \equiv \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \underset{m\times 1}{Y} - \overset{m\times 1}{\underset{m\times n}{X}\,\underset{n\times 1}{\beta}}.}$$

example: Simple$_0$

$p = 0$, $x \equiv$ null $\qquad y$

$n = 1$, $h_0(\cdot) \equiv 1$, $\beta \equiv \beta_0$

$Y_{model}(\cdot\,; \beta) \equiv \beta_0$

FIT DATA to a CONSTANT

$$\underset{m\times 1}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} \qquad \underset{\substack{m\times n \\ (m\times 1)}}{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

example: Simple$_1$

$p = 1$, $x_{(1)} \equiv x$ (say) $\qquad y$

$n = 2$, $h_0(x) \equiv 1$, $h_1(x) \equiv x$

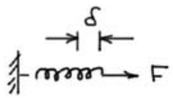
$\beta = (\beta_0 \;\; \beta_1)^T$

$Y_{model}(x; \beta) \equiv \beta_0 + \beta_1 x$

FIT DATA to a LINE

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}$$

Note: $\frac{dY_{model}}{dx} = \beta_1$; differentiation of noisy data example: spring



$$p = 1,\ x_{(1)} = x \equiv \delta/\delta_{max}$$

$$y \equiv F$$

$$n = 3,\ h_0 = 1,\quad h_1(x) \equiv x,\quad h_2(x) = x^2 \ ↯$$

$$h_1(\delta/\delta_{max}) \equiv \delta/\delta_{max} \qquad h_2(\delta/\delta_{max}) = (\delta/\delta_{max})^2$$

$$\beta = (\beta_0\ \beta_1\ \beta_2)^T$$

F_1 $\qquad (\delta/\delta_{max})_1 \qquad (\delta/\delta_{max})_1^2 \qquad (\delta/\delta_{max})_2$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{pmatrix}$$

example: Nusselt number (heat transfer)

$$Nu = \alpha (Re)^{\gamma} (Pr)^{\delta} \qquad \text{not linear}$$

but

$$\log(Nu) = \log\alpha + \gamma \log(Re) + \delta \log(Pr) \qquad \text{linear}$$

so choose

$$p = 2,\ x_{(1)} \equiv \log(Re),\ x_{(2)} = \log(Pr) \qquad \text{or} \quad Re \quad Pr$$

$$y = \log(Nu)$$

$$n = 3,\ h_0 = 1,\ h_1(x) = x_{(1)},\ h_2(x) = x_{(2)} \qquad \log \quad \log$$

$$\log(Re) \qquad \log(Pr)$$

$$\beta = (\beta_0\ \beta_1\ \beta_2)^T$$

$$(\beta_0 \equiv \log\alpha,\ \beta_1 \equiv \gamma,\ \beta_2 \equiv \delta)$$

and hence

$$Y = \begin{pmatrix} \log Nu_1 \\ \log Nu_2 \\ \vdots \\ \log Nu_m \end{pmatrix} \qquad X = \begin{pmatrix} 1 & \log Re_1 & \log Pr_1 \\ 1 & \log Re_2 & \log Pr_2 \\ \vdots & \vdots & \vdots \\ 1 & \log Re_m & \log Pr_m \end{pmatrix}$$

such that (say)

$$r_1 \equiv Y_1 - (X\beta)_1 = \log Nu_1 - \beta_0 - \beta_1 \log Re_1 - \beta_2 \log Pr_1$$

$$r_2 \equiv \ldots$$

$$\vdots$$

or $r \equiv Y - X\beta$.

General Least-Squares Formulation

Review: data + model yields

$$Y, \quad X\beta.$$

In general,

$$X\beta^{true} = Y$$

has no solution.

Look for estimate of β^{true}, $\hat\beta$, such that

$$\hat\beta \text{ minimizes } \sum_{i=1}^{m} r_i^2(\beta)$$

$$r_i(\beta) \equiv Y_i - (X\beta)_i,\quad 1 \le i \le m.$$

matrix form

Note

$$r(\beta) \equiv Y - X\beta \qquad m\times 1$$

and

$$\sum_{i=1}^{m} r_i^2(\beta) = \underbrace{r^T(\beta)\,r(\beta)}_{\|r(\beta)\|^2} = (r_1\ r_2\ \cdots\ r_m)\begin{pmatrix} r_1\\ r_2\\ \vdots\\ r_m\end{pmatrix}.$$

Define

$$J(\beta) \equiv r^T(\beta)\,r(\beta)\,; \qquad \text{note } J \text{ is a scalar}$$

then

$\hat{\beta}$ minimizes $J(\beta)$.

So

$$J(\beta) = r^T(\beta)\,r(\beta) = (Y - X\beta)^T(Y - X\beta)$$

$$= (Y^T - (X\beta)^T)(Y - X\beta)$$

$$= (Y^T - \beta^T X^T)(Y - X\beta)$$

$$= Y^T(Y - X\beta) - \beta^T X^T(Y - X\beta)$$

$$= \underbrace{Y^T Y}_{1\times m\ m\times 1} - \underbrace{Y^T X\beta}_{1\times m\ m\times n\ n\times 1} - \underbrace{\beta^T X^T Y}_{1\times n\ n\times m\ m\times 1} + \underbrace{\beta^T X^T X\beta}_{1\times n\ n\times m\ m\times n\ n\times 1}$$

pause

note $Y^T X\beta = (Y^T X\beta)^T$ scalar
$= \beta^T X^T Y$

$$= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

It can be shown that the unique† $\hat{\beta}$ which

minimizes $J(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$

is the solution to the "normal" equation

$$\underbrace{\underbrace{X^T X}_{n\times m\ m\times n}}_{n\times n}\ \underbrace{\hat{\beta}}_{n\times 1} = \underbrace{\underbrace{X^T Y}_{n\times m\ m\times 1}}_{n\times 1}.$$

Hence in practice,

solve normal equation for $\hat{\beta}$, then

$J(\hat{\beta}) < J(\beta)$ for any $\beta \neq \hat{\beta}$: BEST FIT.

† iff X has independent columns

Denote

$$\hat{Y} = X\hat{\beta}$$

such that

$$\hat{Y}_i = (X\hat{\beta})_i = Y_{\text{model}}(x_i\,;\hat{\beta})$$

$=$ model prediction for best-fit $\hat{\beta}$.

Then

$$J(\hat{\beta}) = (Y - X\hat{\beta})^T(Y - X\hat{\beta})$$

$$= (Y - \hat{Y})^T(Y - \hat{Y})$$

$$= \|Y - \hat{Y}\|^2 < \|Y - X\beta\|^2 \text{ for any } \beta \neq \hat{\beta}.$$

example : Simple_0 $\quad p = 0,\ x = \text{null}$

$n = 1,\ h_0 = 1$

$\beta = \beta_0$ FIT to a CONSTANT

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} \qquad X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

First, from scratch:

$$r(\beta_0) = Y - X\beta_0 \qquad (r_i = Y_i - \beta_0)$$

$$J(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta$$

$$\left\{ \begin{aligned} Y^T Y &= \sum_{i=1}^m Y_i^2 = c_0 \\ X^T Y &= (1\ 1\ \ldots\ 1)\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \sum_{i=1}^m Y_i = m\bar{Y} \quad \text{scalar} \\ X^T X &= (1\ 1\ \ldots\ 1)\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = m \quad \text{scalar} \end{aligned} \right.$$

so

$$\left. \begin{aligned} & J(\beta) = c_0 - 2\beta_0 m \bar{Y} + \beta_0^2 m \\ & \frac{dJ}{d\beta_0}(\beta_0) = -2m\bar{Y} + 2\beta_0 m; \quad \frac{dJ}{d\beta_0}(\hat{\beta}_0) = 0 \Rightarrow \hat{\beta}_0 = \bar{Y} \\ & \frac{d^2J}{d\beta_0^2}(\beta_0) = 2m > 0 \quad (\Rightarrow \hat{\beta}_0 \text{ a minimizer}) \end{aligned} \right\}$$

Second, from the general formula,

$$\left. \begin{aligned} X^T X \hat{\beta} &= X^T Y \\ m\,\hat{\beta}_0 &= m\bar{Y} \\ \hat{\beta}_0 &= \bar{Y}\,. \end{aligned} \right\}$$