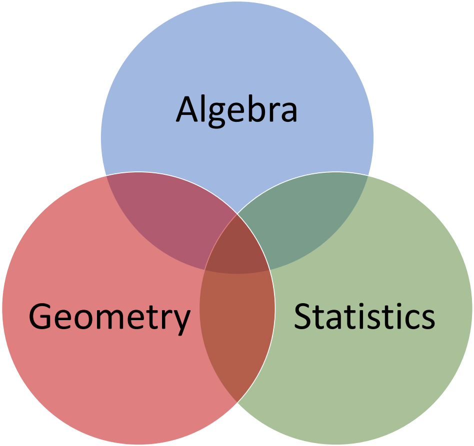
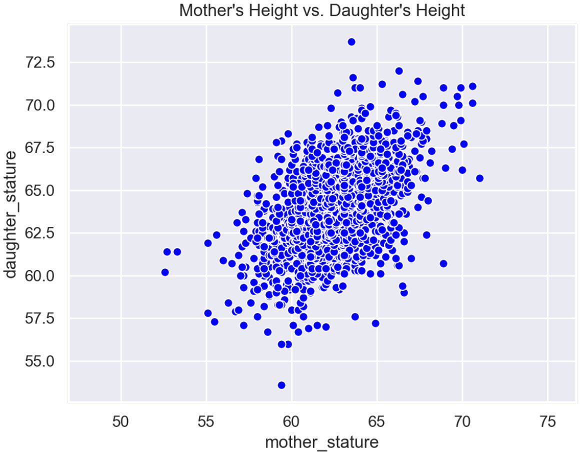
Linear Regression

Foundations of Data Analysis

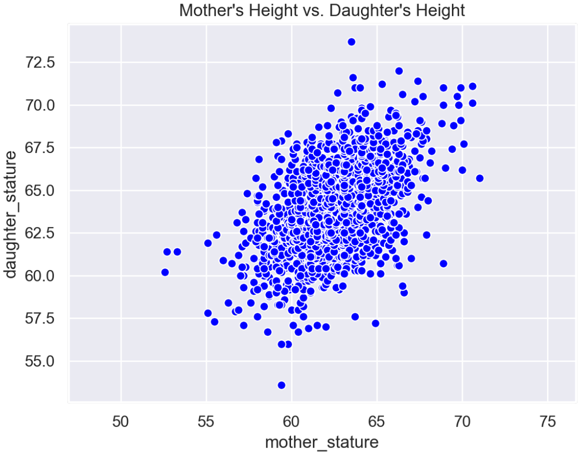
February 20, 2020


Algebra
Geometry
Statistics

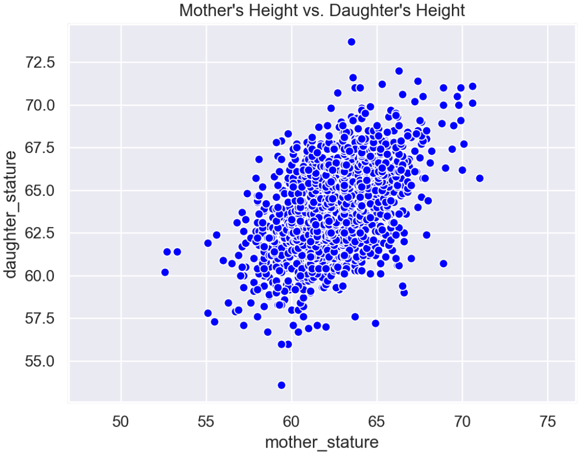
Is there a relationship between the heights of mothers and their daughters?



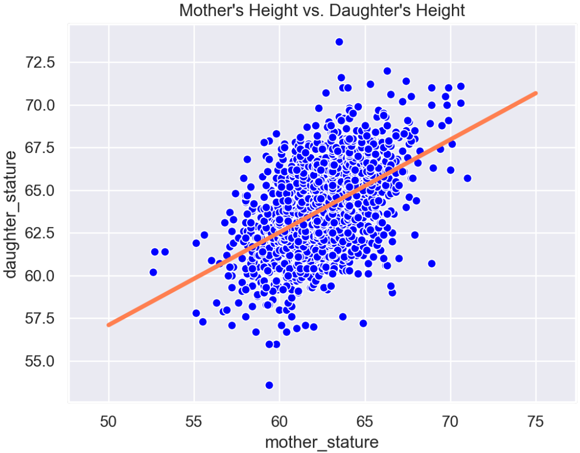
If you know a mother's height, can you predict her daughter's height with any accuracy?



Linear regression is a tool for answering these types of questions.



It models the relationship as a straight line.



Regression Setup

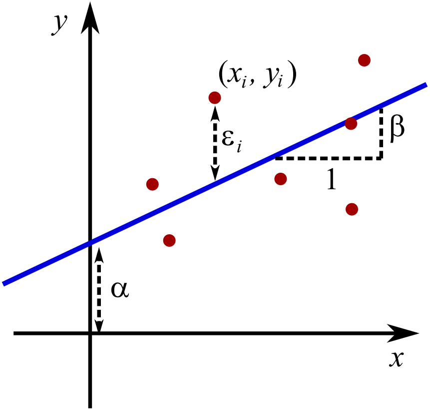
When we are given real-valued data in pairs:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathbb{R}^2$$

Example:
x_i is the height of the ith mother
y_i is the height of the ith mother's daughter

Linear Regression

Model the data as a line:



$$y_i = \alpha + \beta x_i + \epsilon_i$$

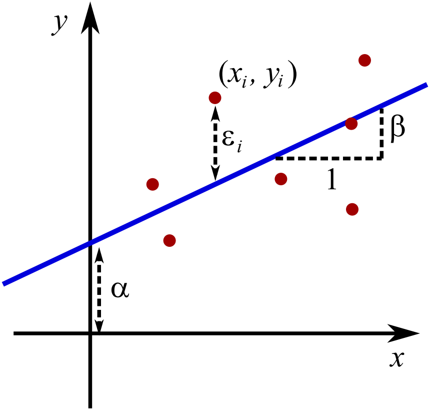
α : intercept

β : slope

ϵ_i : error

Geometry: Least Squares

We want to fit a line as close to the data as possible, which means we want to **minimize the errors**, ϵ_i.



$$y_i = \alpha + \beta x_i + \epsilon_i$$

α : intercept

β : slope

ϵ_i : error

Geometry: Least Squares

Taking the line equation: $y_i = \alpha + \beta x_i + \epsilon_i$

Geometry: Least Squares

Taking the line equation: $y_i = \alpha + \beta x_i + \epsilon_i$

Rearrange to get: $\epsilon_i = y_i - \alpha - \beta x_i$

Geometry: Least Squares

Taking the line equation: $y_i = \alpha + \beta x_i + \epsilon_i$

Rearrange to get: $\epsilon_i = y_i - \alpha - \beta x_i$

We want to minimize the **sum-of-squared errors (SSE)**:

$$\mathrm{SSE}(\alpha, \beta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Least Squares: Step 1

Center the data by removing the mean:

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i = x_i - \bar{x}$$

Least Squares: Step 1

Center the data by removing the mean:

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i = x_i - \bar{x}$$

Note: $\sum_{i=1}^{n} \tilde{y}_i = 0$ and $\sum_{i=1}^{n} \tilde{x}_i = 0$

Least Squares: Step 1

Center the data by removing the mean:

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i = x_i - \bar{x}$$

Note: $\sum_{i=1}^{n} \tilde{y}_i = 0$ and $\sum_{i=1}^{n} \tilde{x}_i = 0$

We'll first get a solution: $\tilde{y} = \alpha + \beta\tilde{x}$, then shift it back to the original (uncentered) data at the end

Least Squares: Step 2

Take derivative of $\mathrm{SSE}(\alpha, \beta)$ wrt α and set to zero:

$$0 = \frac{\partial}{\partial \alpha} \mathrm{SSE}(\alpha, \beta)$$

Least Squares: Step 2

Take derivative of $\mathrm{SSE}(\alpha, \beta)$ wrt α and set to zero:

$$0 = \frac{\partial}{\partial \alpha} \mathrm{SSE}(\alpha, \beta) = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i)^2$$

Least Squares: Step 2

Take derivative of $\mathrm{SSE}(\alpha, \beta)$ wrt α and set to zero:

$$\begin{aligned} 0 = \frac{\partial}{\partial \alpha} \mathrm{SSE}(\alpha, \beta) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i)^2 \\ &= -2 \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i) \end{aligned}$$

Least Squares: Step 2

Take derivative of $\mathrm{SSE}(\alpha, \beta)$ wrt α and set to zero:

$$
\begin{aligned}
0 = \frac{\partial}{\partial \alpha} \mathrm{SSE}(\alpha, \beta) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i)^2 \\
&= -2 \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i) \\
&= -2 \sum_{i=1}^{n} \tilde{y}_i + 2n\alpha + 2\beta \sum_{i=1}^{n} \tilde{x}_i
\end{aligned}
$$

Least Squares: Step 2

Take derivative of $\mathrm{SSE}(\alpha, \beta)$ wrt α and set to zero:

$$
\begin{aligned}
0 = \frac{\partial}{\partial \alpha} \mathrm{SSE}(\alpha, \beta) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i)^2 \\
&= -2 \sum_{i=1}^{n} (\tilde{y}_i - \alpha - \beta \tilde{x}_i) \\
&= -2 \sum_{i=1}^{n} \tilde{y}_i + 2n\alpha + 2\beta \sum_{i=1}^{n} \tilde{x}_i
\end{aligned}
$$

Using $\sum \tilde{y}_i = \sum \tilde{x}_i = 0$, we get

$$\hat{\alpha} = 0$$

Least Squares: Step 3

With $\alpha = 0$, we are left with

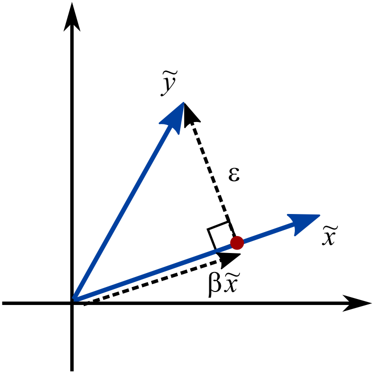
$$\tilde{y}_i = \beta \tilde{x}_i + \epsilon_i$$

Or, in vector notation:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \beta \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Least Squares: Step 3

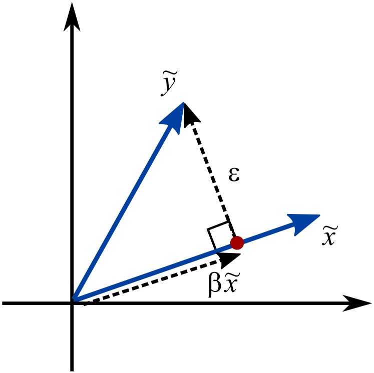
$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \beta \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$



Minimizing $\mathrm{SSE}(\alpha, \beta) = \sum \epsilon_i^2 = \|\epsilon\|^2$ is projection!

Least Squares: Step 3

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \beta \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$



Minimizing $\mathrm{SSE}(\alpha, \beta) = \sum \epsilon_i^2 = \|\epsilon\|^2$ is projection!

Solution is $\hat{\beta} = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\|^2}$

Shifting Back to Uncentered Data

So far, we have:

$$\tilde{y}_i = \hat{\beta}\tilde{x}_i + \epsilon_i$$

Shifting Back to Uncentered Data

So far, we have:

$$\tilde{y}_i = \hat{\beta}\tilde{x}_i + \epsilon_i$$

Expanding out $\tilde{x}_i$ and $\tilde{y}_i$ gives

$$(y_i - \bar{y}) = \hat{\beta}(x_i - \bar{x}) + \epsilon_i$$

Shifting Back to Uncentered Data

So far, we have:

$$\tilde{y}_i = \hat{\beta}\tilde{x}_i + \epsilon_i$$

Expanding out $\tilde{x}_i$ and $\tilde{y}_i$ gives

$$(y_i - \bar{y}) = \hat{\beta}(x_i - \bar{x}) + \epsilon_i$$

Rearranging gives

$$y_i = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta}x_i + \epsilon_i$$

Shifting Back to Uncentered Data

So far, we have:

$$\tilde{y}_i = \hat{\beta}\tilde{x}_i + \epsilon_i$$

Expanding out $\tilde{x}_i$ and $\tilde{y}_i$ gives

$$(y_i - \bar{y}) = \hat{\beta}(x_i - \bar{x}) + \epsilon_i$$

Rearranging gives

$$y_i = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta}x_i + \epsilon_i$$

So, for the uncentered data, $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

Probability: Maximum Likelihood

So far, we have only used geometry, but if our data is random, shouldn't we be talking about probability?

Probability: Maximum Likelihood

So far, we have only used geometry, but if our data is random, shouldn't we be talking about probability?

To make linear regression probabilistic, we model the errors as Gaussian:

$$\epsilon_i \sim N(0, \sigma^2)$$

Probability: Maximum Likelihood

So far, we have only used geometry, but if our data is random, shouldn't we be talking about probability?

To make linear regression probabilistic, we model the errors as Gaussian:

$$\epsilon_i \sim N(0, \sigma^2)$$

The likelihood is

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right)$$

Probability: Maximum Likelihood

The log-likelihood is then

$$\log L(\alpha, \beta) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2 + \text{const.}$$

Probability: Maximum Likelihood

The log-likelihood is then

$$\log L(\alpha, \beta) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\epsilon_i^2 + \text{const.}$$

Maximizing this is equaivalent to minimizing SSE!

$$\max \log L = \min \sum \epsilon_i^2 = \min \text{SSE}$$