A drug company sells a skin cream through drugstores in 15 districts. It would like to predict district sales, and collects some data.

District	Sales (gross of jars)	Target population (1000s persons)	Per Capita income (euro)
i	y_i	x_{i1}	x_{i2}
1	162	274	2450
2	120	180	3254
3	223	375	3802
4	131	205	2838
5	67	86	2347
6	169	265	3782
7	81	98	3008
8	192	330	2450
9	116	195	2137
10	55	53	2560
11	252	430	4020
12	232	372	4427
13	144	236	2660
14	103	157	2088
15	212	370	2605

$n = 15$

A 3-d plot suggests a linear relationship

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

where ε_i is an "error term". So we should determine the plane

$$y = b_0 + b_1 x_1 + b_2 x_2$$

where b_0, b_1, b_2 are chosen to minimize the quantity

$$Q = \sum_{i=1}^{n} \left(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2}) \right)^2$$

Here $Q = Q(b_0, b_1, b_2)$, and for a minimum

$$\frac{\partial Q}{\partial b_0} = 0, \quad \frac{\partial Q}{\partial b_1} = 0, \quad \frac{\partial Q}{\partial b_2} = 0 \qquad (*)$$

These normal equations $(*)$ can be expressed in matrix form:

$$B = (X^t X)^{-1} X^t Y \qquad (*)$$

where

$$B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & & \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Equation (*) can be solved to yield

$$b_0 = 3.4526127$$
$$b_1 = 0.4960049$$
$$b_2 = 0.0091490$$

and plane

$$\underline{y} = 3.453 + 0.496 x_1 + 0.009 x_2$$

This can be used to predict the sales y in a new district of size x_1 and income x_2.

General case: $p-1$ independent variables

Given points

$$(y_i, x_{i1}, x_{i2}, \dots, x_{i\,p-1})$$

for $i = 1, 2, \dots, n$ we set

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1\,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2\,p-1} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n\,p-1} \end{pmatrix}$$

and we want to determine a vector

$$B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix}$$

such that $Y - XB$ is "as small as possible".

<u>Aside</u> for $u = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{p-1} \end{pmatrix}$, $v = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_{p-1} \end{pmatrix}$

$\in \mathbb{R}^p$ we define the <u>dot product</u>

$u \cdot v = u^t v = u_0 v_0 + u_1 v_1 + \cdots + u_{p-1} v_{p-1}$.

we define the size of v to be

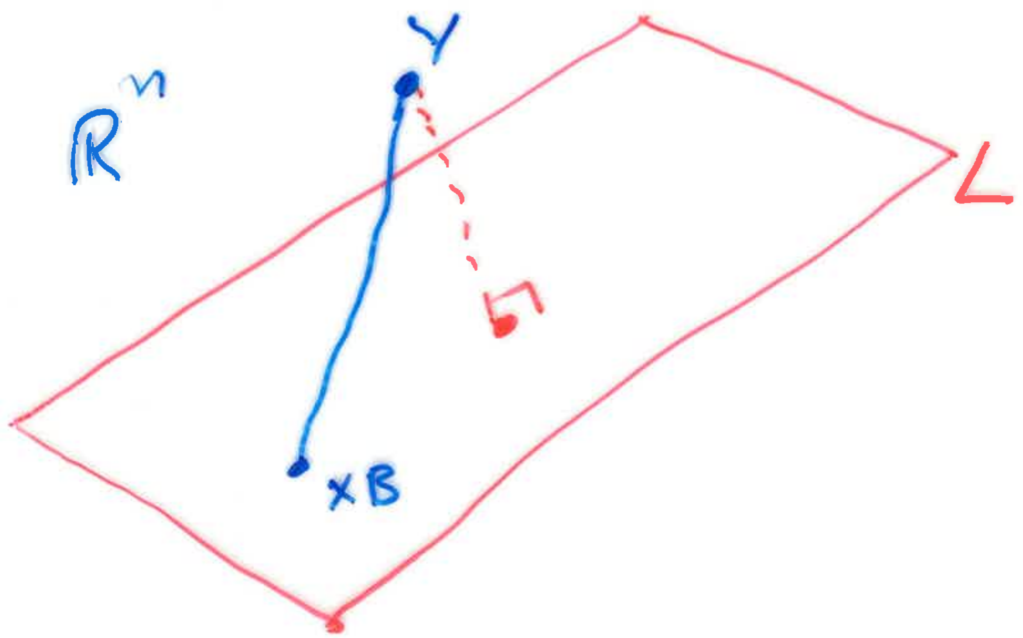
$$\|v\| = \sqrt{v \cdot v}$$

So we want to choose B so that $\|Y - XB\|$ is as small as possible.

Consider the vector space

$$L = \{ Xu : u \in \mathbb{R}^p \} \subseteq \mathbb{R}^n$$

column vector



Now the vector $Y - XB$ is as small as possible if it is perpendicular to the space L.

So we want $Y - XB$ to be perpendicular to every Xu for $u \in \mathbb{R}^p$.

That is, for <u>any</u> $u \in \mathbb{R}^p$ we want

$$(Xu) \cdot (Y - XB) = 0$$

or

$$(u^t X^t)(Y - XB) = 0 .$$

Since this has to be true for all $u \in \mathbb{R}^p$, we need

$$X^t (Y - XB) = 0 .$$

Then

$$X^t Y = X^t X B$$

and

$$\boxed{B = (X^t X)^{-1} X^t Y} \quad (*)$$