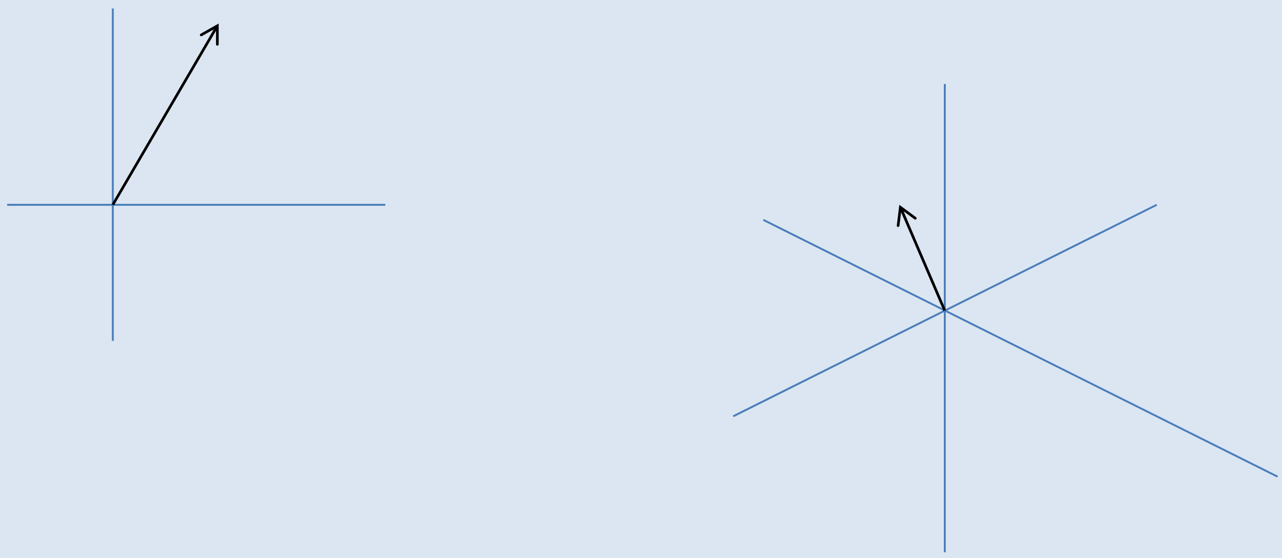
Vectors

The Australian Curriculum mentions vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$



Why not vectors in $\mathbb{R}^4, \mathbb{R}^5, \ldots, \mathbb{R}^n$?

Length

Vector $\boldsymbol{a} \in \mathbb{R}^n$ has length or magnitude $|\boldsymbol{a}| = \sqrt{{a_1}^2 + {a_2}^2 + \cdots + {a_n}^2}$

where the numbers a_i are the coordinates with respect to an orthogonal basis.

In terms of the usual scalar product, this is $|\boldsymbol{a}| = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$

Orthogonality

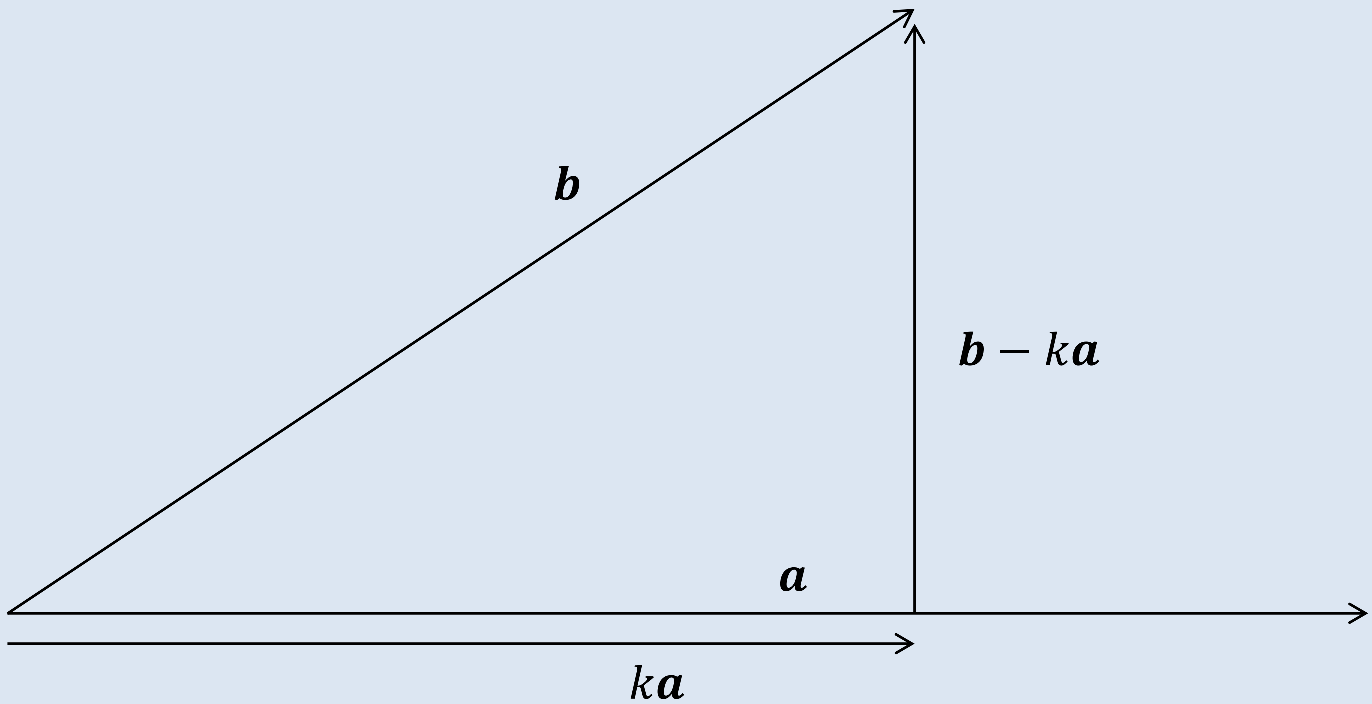
Vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ are said to be orthogonal if $\boldsymbol{a} \cdot \boldsymbol{b} = 0$.

With distance and orthogonality defined in this way, using the scalar product, we can think of the orthogonal projection $P_{\boldsymbol{a}}(\boldsymbol{b})$ of $\boldsymbol{b}$ onto $\boldsymbol{a}$ as the best approximation to $\boldsymbol{b}$ as a scalar multiple of $\boldsymbol{a}$.

Projection

With distance and orthogonality defined using the scalar product, we can think of the orthogonal projection $P_{\boldsymbol{a}}(\boldsymbol{b})$ of $\boldsymbol{b}$ onto $\boldsymbol{a}$ as the scalar multiple of $\boldsymbol{a}$ that represents the best approximation to $\boldsymbol{b}$.

The orthogonal projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ minimises the distance $|\boldsymbol{b} - k\boldsymbol{a}|$.



$|\boldsymbol{b} - k\boldsymbol{a}|$ is minimised when $k\boldsymbol{a}$ and $\boldsymbol{b} - k\boldsymbol{a}$ are orthogonal.

That is, $|\boldsymbol{b} - k\boldsymbol{a}|$ is minimised when $k\boldsymbol{a} = P_{\boldsymbol{a}}(\boldsymbol{b})$

Fishy experiment

A sample of 160 fish were given 5 treatments, each in 4 levels. The response measured was the amount of *atarinine* developed in the brains of the fish over a 20-month period.

The treatments are called *Finglish*, *Mathematics*, *Drama*, *Business* and *Sport*, and in the experiment they were administered in amounts 0, 1, 2 and 3 units to particular subjects according to a randomised design.

Model

We think some linear combination of the amounts of each treatment will be a predictor of the observed results.

$$A = u_1 F + u_2 M + u_3 D + u_4 B + u_5 S + \epsilon$$

We expect that the experimental results will enable us to discover the values of the coefficients u_i.

F(english)	*M(aths)*	*D(rama)*	*B(usiness)*	*S(port)*	*Atarinine*
3	2	0	0	0	1.33
3	3	1	0	0	1.01
0	2	1	3	0	0.89
2	3	2	3	1	1.21
1	2	1	0	1	1.11
1	2	0	1	0	0.28
3	1	1	2	0	0.49
3	0	3	3	2	1.29
3	0	3	0	0	0.77
0	0	0	2	0	0.86
3	0	1	1	2	0.45
3	2	2	2	2	1.25
1	1	0	0	1	1.02
0	0	2	2	0	0.61
1	2	2	3	3	0.55
1	1	1	0	1	0.83

$$\begin{bmatrix} A \\ \vdots \end{bmatrix} = \begin{bmatrix} F & M & D & B & S \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} + \boldsymbol{\epsilon}$$

$$Y = X\boldsymbol{u} + \boldsymbol{\epsilon}$$

We can allow the coefficient vector $\boldsymbol{u}$ to vary until we find the orthogonal projection of the observed vector Y onto the column space of the matrix X.

We require

$$(Y - X\boldsymbol{u}) \cdot X\boldsymbol{u} = 0$$

That is,

$$(X\boldsymbol{u})^T (Y - X\boldsymbol{u}) = 0$$

$$\boldsymbol{u}^T X^T Y - \boldsymbol{u}^T X^T X\boldsymbol{u} = 0$$

$$\boldsymbol{u}^T (X^T Y - X^T X\boldsymbol{u}) = 0$$

Now, the vector $\boldsymbol{u}$ of coefficients is variable. So, we must have

$$X^T Y - X^T X\boldsymbol{u} = \mathbf{0}$$

since the zero vector is the only one orthogonal to every vector.

But then,

$$\boldsymbol{u} = (X^T X)^{-1} X^T Y$$

This gives the 'least-squares' best fitting approximation $X\boldsymbol{u}$ to the observed vector Y.

The matrix X has 160 rows and 5 columns. So, $X^T X$ is a 5×5 matrix.

According to Excel it is:

531	345	336	339	327
345	526	327	335	341
336	327	555	361	325
339	335	361	540	323
327	341	325	323	496

It’s inverse is:

0.003927	-0.00096	-0.00118	-0.00045	-0.00104
-0.00096	0.004101	-0.00086	-0.00095	-0.001
-0.00118	-0.00086	0.004307	-0.00133	-2.4E-05
-0.00045	-0.00095	-0.00133	0.003895	-0.00107
-0.00104	-0.001	-2.4E-05	-0.00107	0.004091

$$X^T Y = \begin{matrix} 200.21 \\ 180.44 \\ 202.32 \\ 219.46 \\ 186.24 \end{matrix}$$

$$\text{So, } \boldsymbol{u} = \begin{matrix} 0.122575 \\ 0.144808 \\ 0.109237 \\ 0.081076 \\ 0.072415 \end{matrix}$$

Model:

$$A = 0.12F + 0.14M + 0.11D + 0.08B + 0.07S$$

According to the model, each of the treatments had a positive effect on the expected development of atarinine, but with different intensities.

Interestingly, the model suggests that *mathematics* was the most important contributor to the levels of atarinine, closely followed by *finglish* and *drama*.

Some of the error terms, however, were very large in this experiment – up to 400% in a few subjects.

Hence, the model needs refinement as the underlying theory does not fully explain the experimental results.

The point

Some students will want more and your explanations will be clearer if you have gone further yourself .

Thank you