

The Geometry of Linear Models

STATS 762 – Lecture Notes 1

Arden Miller

Department of Statistics, University of Auckland

Linear Models and Geometry

There is a rich geometry associated with the statistical linear model. Understanding this geometry can provide insight in much of the analysis associated with regression analysis.

- ▶ The idea is to write the regression model as a vector equation and explore the implications of this equation using a basic understanding of vector spaces.
- ▶ Need to review aspects of vectors and vector spaces.

The Basics of Vectors

For our purposes, a vector is a “ n -tuple” of real numbers which we denote

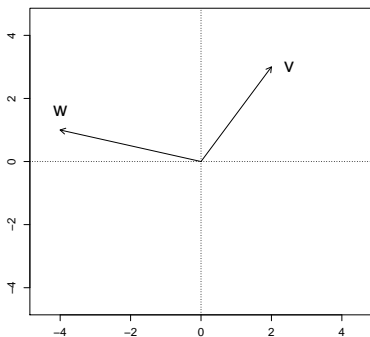
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

boldface will be used to indicate vectors (and matrices)

- ▶ We will think of a vector as the coordinates of a point in n -dimensional space – often we will use a directed line segment that extends from the origin to these coordinates to help us visualise concepts.

Example: 2-component Vectors

Two-component vectors can be displayed as directed line segments on a standard scatterplot.



$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$$

Vector Addition

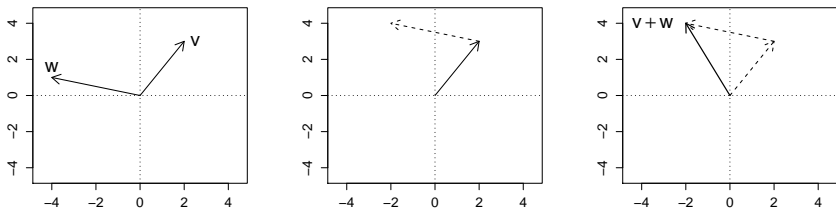
The sum of two vectors is obtained by adding their corresponding entries:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$$

For example: $\mathbf{v} + \mathbf{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -4 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$

Visualising Vector Addition

Visually, we translate the starting point of one the vectors to the endpoint point of the other.



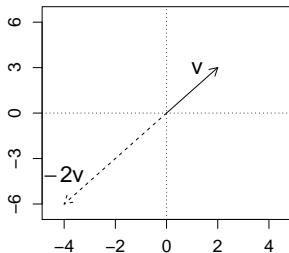
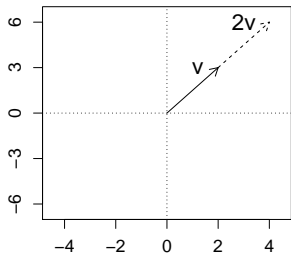
The sum is the vector from the origin to the new endpoint.

Scalar Multiplication of Vectors

To multiply a vector by a constant, simply multiply each entry by that constant:

$$k \times \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} k \times v_1 \\ k \times v_2 \\ \vdots \\ k \times v_n \end{bmatrix}$$

Visualising Scalar Multiplication



If we multiply a vector by a constant, the resulting vector has the same direction (or the opposite direction if the constant is negative) as the original vector but its length has been multiplied by the constant.

Some Basic Vector Algebra

For vectors \mathbf{u} , \mathbf{v} and \mathbf{w} and scalars k_1 and k_2 :

1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
3. $k_1(\mathbf{v} + \mathbf{w}) = k_1\mathbf{v} + k_1\mathbf{w}$
4. $(k_1 + k_2)\mathbf{v} = k_1\mathbf{v} + k_2\mathbf{v}$

- ▶ Pretty much any algebraic property that applies to the addition and multiplication of real numbers will apply to vectors as well.

The Linear (Regression) Model

The linear model can be written as an equation which relates the value of a response variable Y to the values of one or more explanatory variables:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \epsilon$$

- ▶ All of the β 's are fixed constants but are unknown
- ▶ ϵ is a random variable that is assumed to have a $N(0, \sigma^2)$ distribution.
- ▶ As a result Y is a random variable with mean $\mu = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ and variance σ^2 .

The Data

Suppose we have n observed values for the response y_1 through y_n . For observation i , denote the values of the explanatory variables as x_{i1} through x_{ik} and arrange the data in a table:

| Obs. | Resp. | X_1 | X_2 | ... | X_k |
|----------|----------|----------|----------|-----|----------|
| 1 | y_1 | x_{11} | x_{12} | ... | x_{1k} |
| 2 | y_2 | x_{21} | x_{22} | ... | x_{2k} |
| \vdots | \vdots | \vdots | \vdots | | \vdots |
| n | y_n | x_{n1} | x_{n2} | ... | x_{nk} |

A Set of Equations

For each observation y_i , we can write:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i$$

Stacking all of these equations gives:

$$\begin{array}{rcllclcl} y_1 & = & \beta_0 & + & \beta_1 x_{11} & + & \beta_2 x_{12} & + & \dots & + & \beta_k x_{1k} & + & \epsilon_1 \\ y_2 & = & \beta_0 & + & \beta_1 x_{21} & + & \beta_2 x_{22} & + & \dots & + & \beta_k x_{2k} & + & \epsilon_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_n & = & \beta_0 & + & \beta_1 x_{n1} & + & \beta_2 x_{n2} & + & \dots & + & \beta_k x_{nk} & + & \epsilon_n \end{array}$$

The Linear Model as a Vector Equation

The previous set of equations can be rewritten as a vector equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + \dots + \beta_k \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Using boldface to represent vectors, this becomes:

$$\mathbf{y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + \boldsymbol{\epsilon}$$

Catheter Length Example

For 12 young patients, catheters were fed from a principal vein into the heart. The catheter length was measured as was the height and weight of the patients. Is it possible to predict the necessary catheter length based on height and weight?

Reference: S Weisberg, Applied Linear Regression, Wiley, 1980, page 218.

Catheter Length Data

| Patient | Height (in.) | Weight (lbs.) | Catheter (cm) |
|---------|--------------|---------------|---------------|
| 1 | 42.8 | 40.0 | 37 |
| 2 | 63.5 | 93.5 | 50 |
| 3 | 37.5 | 35.5 | 34 |
| 4 | 39.5 | 30.0 | 36 |
| 5 | 45.5 | 52.0 | 43 |
| 6 | 38.5 | 17.0 | 28 |
| 7 | 43.0 | 38.5 | 37 |
| 8 | 22.5 | 8.5 | 20 |
| 9 | 37.0 | 33.0 | 34 |
| 10 | 23.5 | 9.5 | 30 |
| 11 | 33.0 | 21.0 | 38 |
| 12 | 58.0 | 79.0 | 47 |

Catheter Regression Model

We can explore using a regression model that relates the necessary catheter length to the height and weight of the patient:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

- ▶ Y is catheter length.
- ▶ X_1 is patient height.
- ▶ X_2 is patient weight.
- ▶ ϵ represents patient-to-patient variability.

The Vector Equation for the Catheter Data

$$\begin{bmatrix} 37 \\ 50 \\ 34 \\ 36 \\ 43 \\ 28 \\ 37 \\ 20 \\ 34 \\ 30 \\ 38 \\ 47 \\ y \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} 42.8 \\ 63.5 \\ 37.5 \\ 39.5 \\ 45.5 \\ 38.5 \\ 43.0 \\ 22.5 \\ 37.0 \\ 23.5 \\ 33.0 \\ 58.0 \\ x_1 \end{bmatrix} + \beta_2 \begin{bmatrix} 40.0 \\ 93.5 \\ 35.5 \\ 30.0 \\ 52.0 \\ 17.0 \\ 38.5 \\ 8.5 \\ 33.0 \\ 9.5 \\ 21.0 \\ 79.0 \\ x_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon \end{bmatrix}$$

Fixed Vectors and Random Vectors

The linear model contains two types of vectors:

1. *Fixed vectors* are vectors of constants – these are vectors that you would study in a Maths course.
 - ▶ $\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_k$ are all fixed vectors.
2. *Random vectors* are vectors of random variables. Thus a random vector has a distribution.
 - ▶ ϵ is a random vector.

Some Stuff about Random Vectors

A random vector \mathbf{V} that contains random variables V_1, \dots, V_p can be thought of as a vector that has a density function or as a collection of random variables.

- ▶ The distribution for \mathbf{V} is determined by the joint distribution of V_1, \dots, V_p .
- ▶ The expected value of \mathbf{V} represents its “average location” and is a fixed vector given by:

$$E(\mathbf{V}) = E \begin{bmatrix} V_1 \\ \vdots \\ V_p \end{bmatrix} = \begin{bmatrix} E(V_1) \\ \vdots \\ E(V_p) \end{bmatrix}$$

- ▶ We will use the notation $\mu_{\mathbf{V}}$ to represent $E(\mathbf{V})$.

More Stuff about Random Vectors

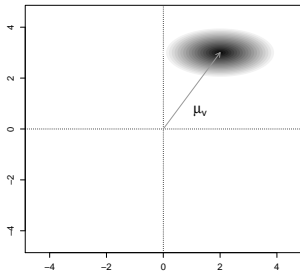
To summarise how \mathbf{V} varies about $\mu_{\mathbf{V}}$, both the variability of the elements and how they vary relative to each other must be considered (the variances of individual elements and the covariances between pairs of elements).

- ▶ It is convenient, to put these variances and covariances into a matrix which we will call $\Sigma_{\mathbf{V}}$ or $\text{Cov}(\mathbf{V})$.

$$\Sigma_{\mathbf{V}} = \begin{bmatrix} \text{var}(V_1) & \text{cov}(V_1, V_2) & \cdots & \text{cov}(V_1, V_p) \\ \text{cov}(V_2, V_1) & \text{var}(V_2) & \cdots & \text{cov}(V_2, V_p) \\ \vdots & \vdots & & \vdots \\ \text{cov}(V_p, V_1) & \text{cov}(V_p, V_2) & \cdots & \text{var}(V_p) \end{bmatrix}$$

The Density of a Random Vector

Conceptually, it is useful to think the density function for a random vector as a cloud in R^n that indicates the plausible end points for the random vector: the vector is more likely to end in a region where the cloud is dense than one where it is not dense.



Working with Random Vectors

If we add a fixed vector \mathbf{C} to a random vector \mathbf{V} , the resulting vector $\mathbf{U} = \mathbf{V} + \mathbf{C}$ is a random vector with:

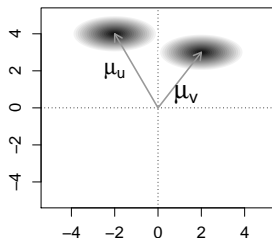
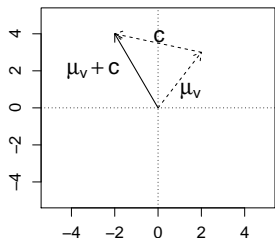
$$\mu_{\mathbf{U}} = \mu_{\mathbf{V}} + \mathbf{C} \quad \text{and} \quad \Sigma_{\mathbf{U}} = \Sigma_{\mathbf{V}}$$

► E.g. If $\mu_{\mathbf{V}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$

then $\mu_{\mathbf{U}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -4 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$

Working with Random Vectors

The mean of the vector has been shifted but how the vector varies about its mean stays the same.



The Distribution of the Errors

The linear model assumes that the errors are independent, $N(0, \sigma^2)$ observations.

- ▶ The joint distribution of the ϵ_i 's is multivariate Normal with $E(\epsilon_i) = 0$, $\text{var}(\epsilon_i) = \sigma^2$ and $\text{cov}(\epsilon_i, \epsilon_j) = 0$ for all i and $j \neq i$.
- ▶ Thus the joint density function is:

$$f(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\epsilon_1^2 + \epsilon_2^2 + \dots + \epsilon_n^2)/2\sigma^2}$$

The Distribution of ϵ

On slide 11 we defined the random vector ϵ :

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad \text{where the } \epsilon_i \text{'s are independent } N(0, \sigma^2) \text{ random variables.}$$

The random vector ϵ is designated as being $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$:

$$\mu_{\epsilon} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \Sigma_{\epsilon} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

The Density “Cloud” for ϵ

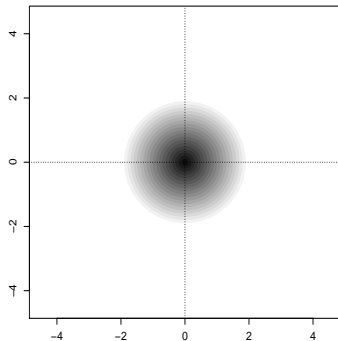
For ϵ , the density can be written as:

$$f(\epsilon) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\|\epsilon\|^2/2\sigma^2} \quad \text{where} \quad \|\epsilon\|^2 = \epsilon^t \epsilon = \epsilon_1^2 + \dots + \epsilon_n^2$$

- ▶ Since $\|\epsilon\|$ is the length of ϵ , the density function depends on the length of ϵ but not on its direction.
- ▶ The density decreases as $\|\epsilon\|^2$ increases.
- ▶ Conceptually, this density is a n -dimensional fuzzy ball centered at the origin.

The Density “Cloud” for a $N(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ Vector

A two dimensional $N(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ random vector would have a density cloud like this:



Is the Response Vector Fixed or Random?

It depends:

- ▶ When we are talking about the properties of the linear model, then the response vector is a random vector which we will denote as \mathbf{Y} .
- ▶ However, when we are talking about a particular data set, then the response vector contains the observed values of the response which we will denote by \mathbf{y} . Technically, \mathbf{y} is a fixed vector which represents a particular realisation of the random vector \mathbf{Y} .

The Distribution of \mathbf{Y}

The linear model represents \mathbf{Y} as the sum of a fixed vector and a random vector:

$$\mathbf{Y} = \underbrace{\beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k}_{\text{fixed vector}} + \underbrace{\boldsymbol{\epsilon}}_{\text{random vector}}$$

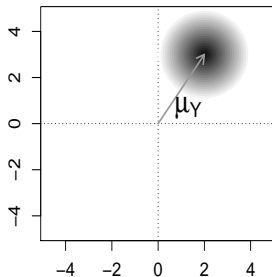
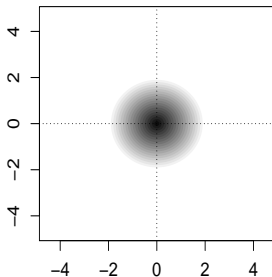
- ▶ \mathbf{Y} is a random vector with:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{Y}} &= \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + \boldsymbol{\mu}_{\boldsymbol{\epsilon}} \\ &= \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k\end{aligned}$$

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2 \mathbf{I}_n$$

The Density Cloud of \mathbf{Y}

\mathbf{Y} has the same density as ϵ except that it is centered around $\mu_{\mathbf{Y}}$ rather than the origin.



The Mean

The linear model restricts the possibilities for μ_Y to vectors that can be formed by taking linear combinations of the vectors $\mathbf{1}$, \mathbf{x}_1 , \dots \mathbf{x}_k :

$$\mu_Y = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots \beta_k \mathbf{x}_k$$

- ▶ In words, the mean vector must be a linear combination of $\mathbf{1}$, \mathbf{x}_1 , \dots \mathbf{x}_k .

Vector Spaces

For our purposes, we only need to consider vectors which contain real numbers and the usual definitions of vector addition and scalar multiplication. In this case, a *vector space* is any collection of vectors that is closed under addition and scalar multiplication.

- ▶ This means that if we take two vectors \mathbf{u} and \mathbf{v} from a vector space then any linear combination $k_1\mathbf{u} + k_2\mathbf{v}$ must also be in that vector space.
- ▶ As a result, the zero vector must be in all vector spaces.

Definition of R^n

Let R^n be the set of all n -component vectors where each component is a real number.

- ▶ R^n is a vector space under the usual definitions of vector addition and scalar multiplication.
 - ▶ The real numbers are closed under addition and multiplication.

Subspaces of R^n

We will need to consider the different subspaces of R^n .

- ▶ Any subset of the vectors in R^n which is itself a vector space is called a subspace of R^n .
- ▶ Since we use the same definitions for addition and scalar multiplication as before, all we really need to check is that the subset of vectors is closed under addition and scalar multiplication.

The Basis of a Vector Space

The usual way to define a subspace is by identifying a set of vectors that form a *basis*.

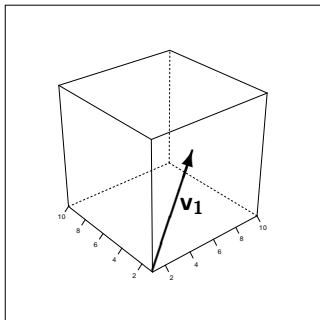
- ▶ Suppose we take any finite collection of vectors from R^n and consider the set of vectors produced by taking all possible linear combinations of these vectors. Our method of generating this subset guarantees that it will be closed under addition and scalar multiplication and thus be a subspace of R^n .
- ▶ Further, suppose that none of the vectors in original collection can be generated as a linear combination of the other vectors – i.e. none of these is redundant. Then this collection of vectors is called a basis for the vector space they generate.

R^3 as an Example

R^3 is a useful example since we can think of it as representing the space around us.

Consider the vector space generated by a single vector \mathbf{v}_1 in R^3 : the subspace consists of \mathbf{v}_1 and all scalar multiples of \mathbf{v}_1 .

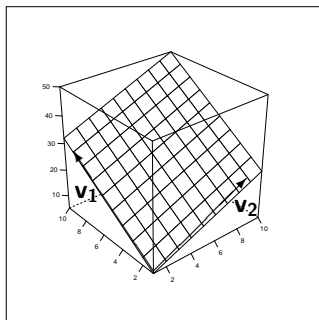
- ▶ This subspace can be thought of as a infinite line in R^3



R^3 as an Example

Now consider the vector space generated by two vectors, \mathbf{v}_1 and \mathbf{v}_2 , in R^3 .

- ▶ Provided that $\mathbf{v}_1 \neq k \times \mathbf{v}_2$ (i.e. they are not co-linear) then the subspace generated by \mathbf{v}_1 and \mathbf{v}_2 is a plane.



The Subspaces of R^3

The subspaces of R^3 can be categorised as follows

1. The origin itself.
2. Any line through the origin.
3. Any plane through the origin.
4. R^3 itself.

Items ~~3~~¹ and 4 on this list are technically subspaces of R^3 but are not of much practical interest – they are referred to as the “improper subspaces.”

Dimensions of Subspaces

Notice that our categories are based on the dimensions of the subspaces.

- ▶ The origin is considered 0-dimensional.
- ▶ Lines are 1-dimensional as they can be defined by a single vector.
- ▶ Planes are 2-dimensional as 2 (non-colinear) vectors are needed to define a plane.
- ▶ R^3 is 3-dimensional.

A Basis of a Subspace

Suppose that for a subspace S we have vectors $v_1 \dots v_k$ such that every vector in S can be expressed as a linear combination of $v_1 \dots v_k$. Then $v_1 \dots v_k$ is said to **span** S .

Vectors $v_1 \dots v_k$ are said to be **linearly independent** if it is not possible to express any one of them as a linear combination of the others.

A set of vectors $v_1 \dots v_k$ is a **basis** for a subspace S if

- (i) $v_1 \dots v_k$ span S
- (ii) $v_1 \dots v_k$ are linearly independent.

The Dimension of a Subspace

For any subspace S , there are an infinite number of bases. However, each of these will consist of *exactly the same number of vectors*. The number of vectors in a basis for S is called the **dimension** of S .

- ▶ E.g. for a line in R^3 , a basis will consist of one vector that falls on that line – lines are 1-dimensional.
- ▶ For any plane in R^3 , any set of 2 linearly independent (non-colinear) vectors that fall on that plane are a basis – planes are 2-dimensional.
- ▶ Any set of 3 linearly independent vectors in R^3 will be a basis for R^3 itself.

Extending to R^n

The subspaces of R^n can be categorised by their dimension:

- ▶ The origin itself.
- ▶ Any line through the origin (1-dimensional).
- ▶ Any plane through the origin (2-dimensional)
- ▶ Any 3-dimensional hyperplane through the origin.
- ▶ \vdots
- ▶ Any $(n - 1)$ -dimensional hyperplane through the origin.
- ▶ R^n itself.

Back to the Regression Model

For the regression model:

$$\mathbf{Y} = \boldsymbol{\mu}_Y + \epsilon \quad \text{where} \quad \boldsymbol{\mu}_Y = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k$$

The relationship between Y and the explanatory variables is modeled through the fixed vector $\boldsymbol{\mu}_Y$.

- ▶ Since $\boldsymbol{\mu}_Y$ is a linear combination of the vectors $\mathbf{1}$, \mathbf{x}_1 , \dots , \mathbf{x}_k , it must be an element of the vector space spanned by $\mathbf{1}$, \mathbf{x}_1 , \dots , \mathbf{x}_k – we will call this the *model space*.
- ▶ Assuming that $\mathbf{1}$, \mathbf{x}_1 , \dots , \mathbf{x}_k are linearly independent, the model space is a subspace of R^n of dimension $k + 1$.

Matrix Form of the Regression Model

The regression model as a vector equation:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ \vdots \\ x_{n1} \end{bmatrix} + \dots + \beta_k \begin{bmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

We can write this more compactly by combining the explanatory variable vectors into a matrix:

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

Matrix Form for Catheter Data

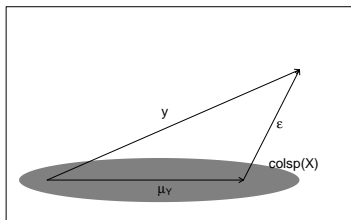
$$\underbrace{\begin{bmatrix} 37 \\ 50 \\ 34 \\ 36 \\ 43 \\ 28 \\ 37 \\ 20 \\ 34 \\ 30 \\ 38 \\ 47 \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} 1 & 42.8 & 40.0 \\ 1 & 63.5 & 93.5 \\ 1 & 37.5 & 35.5 \\ 1 & 39.5 & 30.0 \\ 1 & 45.5 & 52.0 \\ 1 & 38.5 & 17.0 \\ 1 & 43.0 & 38.5 \\ 1 & 22.5 & 8.5 \\ 1 & 37.0 & 33.0 \\ 1 & 23.5 & 9.5 \\ 1 & 33.0 & 21.0 \\ 1 & 58.0 & 79.0 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

Geometric Representation

Thus we have:

$$\mathbf{Y} = \boldsymbol{\mu}_Y + \boldsymbol{\epsilon} \quad \text{where} \quad \boldsymbol{\mu}_Y = \mathbf{X}\boldsymbol{\beta}$$

Notice that we have defined the model space as the subspace of R^n spanned by the columns of \mathbf{X} – another name for this subspace is the *column space* of \mathbf{X} denoted as $\text{colsp}(\mathbf{X})$.



Model Fitting

We can divide the model fitting procedure into two steps:

Step 1: Find an estimate of μ_Y based on the distribution of Y – we will call this estimate $\hat{\mu}_Y$.

Step 2: Use $\hat{\mu}_Y$ to estimate the unknown parameters for the regression model $(\beta_0, \beta_1, \dots, \beta_k)$.

Step 1: Finding $\hat{\mu}_Y$

The regression model restricts μ_Y to the subspace of R^n spanned by the explanatory vectors (we called this the model space). Since the distribution of Y is centered around μ_Y , it makes sense to define $\hat{\mu}_Y$ as the point in the model space that is closest to Y .

- ▶ To find this point, we take the orthogonal projection of Y onto the model space.

Orthogonal Projection Matrices

To find the orthogonal projection of the observed response vector \mathbf{y} onto $\text{colsp}(\mathbf{X})$, we can pre-multiply \mathbf{y} by a projection matrix \mathbf{H} given by:

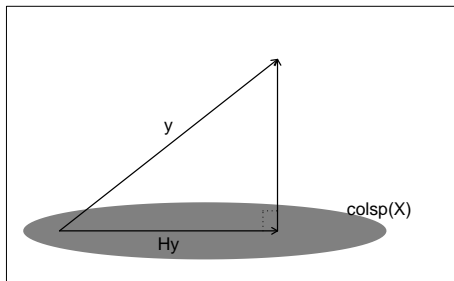
$$\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$$

- ▶ This method of producing a projection matrix works for any matrix \mathbf{X} which has linearly independent columns.

Orthogonal Projection of \mathbf{Y}

Projecting \mathbf{y} onto $\text{colsp}(\mathbf{X})$ gives our estimated mean vector for \mathbf{Y} (i.e the fitted values):

$$\hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = \mathbf{H}\mathbf{y}$$



Catheter Data Analysis using R

In R, we can create the \mathbf{X} matrix and the \mathbf{y} vector for the catheter data as follows:

```
> x1<-c(42.8,63.5,37.5,39.5,45.5,38.5,  
+       43.0,22.5,37.0,23.5,33.0,58.0)  
> x2<-c(40.0,93.5,35.5,30.0,52.0,17.0,  
+       38.5, 8.5,33.0, 9.5,21.0,79.0)  
> X<-cbind(1,x1,x2)  
> y<-matrix(c(37,50,34,36,43,28,37,20,34,30,38,47),12,1)
```

↑ ↑
rows cols

Fitted Values for the Catheter Example

Then we can project \mathbf{y} on to the colsp(\mathbf{X}) to get $\hat{\mu}_Y$ as follows:

```
> H<-X%*%solve(t(X)%*%X)%*%t(X)
> H%*%y
```

```
      [,1]
[1,] 37.03954
[2,] 51.62559
[3,] 35.06266
[4,] 34.43313
[5,] 39.90170
[6,] 31.73815
[7,] 36.79505
[8,] 26.74188
[9,] 34.47955
[10,] 27.14373
[11,] 31.34342
[12,] 47.69560
```

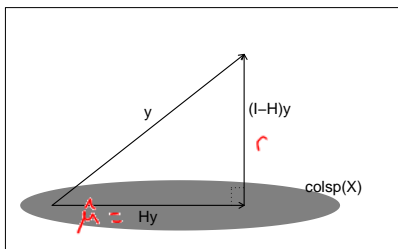
$\hat{\mu}$

$t()$ - transpose
%*% m. mult.
solve(.) inverse.

The Residual Vector

The vector of residuals is defined as:

$$\begin{aligned}\mathbf{r} &= \mathbf{y} - \hat{\boldsymbol{\mu}}_Y \\ &= \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y}\end{aligned}$$



$\mathbf{I} \equiv$ identity matrix

$$\mathbf{I}\mathbf{y} = \mathbf{y}$$

$$\mathbf{I} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$



Least Squares

The orthogonal projection of \mathbf{y} minimises the distance between \mathbf{y} and $\hat{\boldsymbol{\mu}}$. From the previous picture it is clear that this distance between is equal to the length of the residual vector \mathbf{r} which we denote as $\|\mathbf{r}\|$. Recalling some linear algebra:

$$\|\mathbf{r}\| = \sqrt{\mathbf{r}^t \mathbf{r}} = \sqrt{r_1^2 + r_2^2 + \dots + r_n^2}$$

- ▶ Thus choosing $\hat{\boldsymbol{\mu}}$ to minimise $\|\mathbf{r}\|$ is the same as minimising the sum of the squared residuals (least squares).

Parameter Estimates

Since the column vectors of \mathbf{X} are linearly independent, they form a basis for $\text{colsp}(\mathbf{X})$. Thus there is a unique linear combination of the columns of \mathbf{X} that produce $\hat{\boldsymbol{\mu}}$. Putting the coefficients for this relation in a vector $\hat{\boldsymbol{\beta}}$, we get $\hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

$$\begin{aligned} \text{Combining:} \quad & \hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \\ \text{with:} \quad & \hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \\ \text{gives:} \quad & \hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \end{aligned}$$

Parameter Estimates for Catheter Data

To get $\hat{\beta}$ for our catheter data:

```
> solve(t(X)%*%X)%*%t(X)%*%y
      [,1]
20.3757645
x1  0.2107473
x2  0.1910949
```