GENIOMHE

Multivariate Statistics

by Samuel Ortion

Prof.: Cyril Dalmasso

Fall 2023

Contents

1 Introduction

Definition 1: Long Term Nonprocessor (LTNP)

Patient who will remain a long time in good health condition, even with a large viral load (cf. HIV).

Example 1: Genotype: Qualitative or Quantitative?

$$\text{SNP}: \begin{cases} \text{AA} \\ \text{AB} \\ \text{BB} \end{cases} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix},$$

thus we might consider genotype either as a qualitative variable or quantitative variable.

When the variable are quantitative, we use regression, whereas for qualitative variables, we use an analysis of variance.

2 Linear Model

2.1. Simple Linear Regression

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$\mathbf{Y} = \mathbf{X}\beta + \varepsilon.$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \varepsilon_n \end{pmatrix}$$

Assumptions

(A_1) ε_i are independent;

(A_2) ε_i are identically distributed;

(A_3) ε_i are i.i.d $\sim \mathcal{N}(0, \sigma^2)$ (homoscedasticity).

2.2. Generalized Linear Model

$g(\mathbb{E}(Y)) = X\beta$

with g being

- Logistic regression: $g(v) = \log\left(\frac{v}{1-v}\right)$, for instance for boolean values,
- Poisson regression: $g(v) = \log(v)$, for instance for discrete variables.

2.2.1. Penalized Regression

When the number of variables is large, e.g, when the number of explanatory variable is above the number of observations, if $p >> n$ (p: the number of explanatory variable, n is the number of observations), we cannot estimate the parameters. In order to estimate the parameters, we can use penalties (additional terms).

Lasso regression, Elastic Net, etc.

2.2.2. Statistical Analysis Workflow

Step 1. Graphical representation;

Step 2. ...

$$Y = X\beta + \varepsilon,$$

is noted equivalently as

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}.$$

2.3. Parameter Estimation

2.3.1. Simple Linear Regression

2.3.2. General Case

If $\mathbf{X}^T\mathbf{X}$ is invertible, the OLS estimator is:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{2.1}$$

2.3.3. Ordinary Least Square Algorithm

We want to minimize the distance between $\mathbf{X}\beta$ and $\mathbf{Y}$:

$$\min \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

(See chapter 3).

$$\begin{aligned}
&\Rightarrow \mathbf{X}\beta = proj^{(1,\mathbf{X})}\mathbf{Y} \\
&\Rightarrow \forall v \in w, vy = vproj^w(y) \\
&\Rightarrow \forall i : \\
&\quad \mathbf{X}_i\mathbf{Y} = \mathbf{X}_i X\hat{\beta} \qquad \text{where } \hat{\beta} \text{ is the estimator of } \beta \\
&\Rightarrow \mathbf{X}^T\mathbf{Y} = \mathbf{X}^T\mathbf{X}\hat{\beta} \\
&\Rightarrow (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\hat{\beta} \\
&\Rightarrow \hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}
\end{aligned}$$

This formula comes from the orthogonal projection of $\mathbf{Y}$ on the vector subspace defined by the explanatory variables $\mathbf{X}$

$\mathbf{X}\hat{\beta}$ is the closest point to $\mathbf{Y}$ in the subspace generated by $\mathbf{X}$.

If H is the projection matrix of the subspace generated by $\mathbf{X}$, $X\mathbf{Y}$ is the projection on $\mathbf{Y}$ on this subspace, that corresponds to $\mathbf{X}\hat{\beta}$.

2.4. Sum of squares

$\mathbf{Y} - \mathbf{X}\hat{\beta} \perp \mathbf{X}\hat{\beta} - \mathbf{Y}\,\mathbf{1}$ if $\mathbf{1} \in V$, so

$$\underbrace{\|\mathbf{Y} - \bar{\mathbf{Y}}\,\mathbf{1}\|}_{\text{Total SS}} = \underbrace{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}_{\text{Residual SS}} + \underbrace{\|\mathbf{X}\hat{\beta} - \bar{\mathbf{Y}}\,\mathbf{1}\|^2}_{\text{Explicated SS}}$$

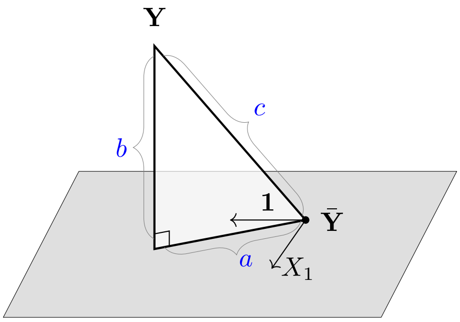

Figure 2.1. Orthogonal projection of $\mathbf{Y}$ on plan generated by the base described by $\mathbf{X}$. a corresponds to $\|\mathbf{X}\hat{\beta} - \bar{\mathbf{Y}}\|^2$ and b corresponds to $\hat{\varepsilon} = \|\mathbf{Y} - \hat{\beta}\mathbf{X}\|^2$ and c corresponds to $\|Y - \bar{Y}\|^2$.

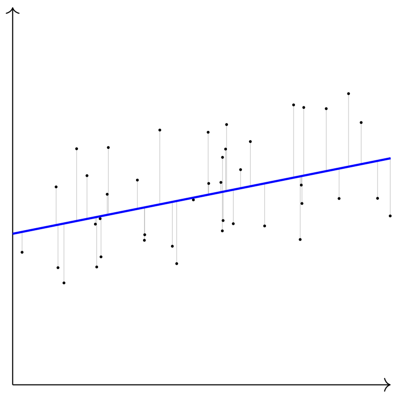

Figure 2.2. Ordinary least squares and regression line with simulated data.

2.5. Coefficient of Determination: R^2

Definition 2: R^2

$$0 \leq R^2 = \frac{\|\mathbf{X}\hat{\beta} - \bar{\mathbf{Y}}\,\mathbf{1}\|^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\,\mathbf{1}\|^2} = 1 - \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\,\mathbf{1}\|^2} \leq 1$$

proportion of variation of $\mathbf{Y}$ explained by the model.

Definition 3: Model dimension

Let $\mathcal{M}$ be a model. The dimension of $\mathcal{M}$ is the dimension of the subspace generated by $\mathbf{X}$, that is the number of parameters in the β vector.
Nb. The dimension of the model is not the number of parameter, as σ^2 is one of the model parameters.

2.6. Gaussian vectors

Definition 4: Normal distribution

Definition 5: Gaussian vector

A random vector $\mathbf{Y} \in \mathbb{R}^n$ is a gaussian vector if every linear combination of its component is ...

Property 1. *$m = \mathbb{E}(Y) = (m_1, \dots, m_n)^T$, where $m_i = \mathbb{E}(Y_i)$*

...

$$\mathbf{Y} \sim \mathcal{N}_n(m, \Sigma)$$

where Σ is the variance-covariance matrix!

$$\Sigma = \mathrm{E}\left[(\mathbf{Y} - m)(\mathbf{Y} - m)^T\right].$$

Remark 1

$$\mathrm{Cov}(Y_i, Y_i) = \mathrm{Var}(Y_i)$$

Definition 6: Covariance

$$\mathrm{Cov}(Y_i, Y_j) = \mathbb{E}\left((Y_i - \mathbb{E}(Y_j))(Y_j - \mathbb{E}(Y_j))\right)$$

When two variable are linked, the covariance is large.
If two variables X, Y are independent, $\mathrm{Cov}(X, Y) = 0$.

Definition 7: Correlation coefficient

$$\mathrm{Cor}(Y_i, Y_j) = \frac{\mathbb{E}\left((Y_i - \mathbb{E}(Y_j))(Y_j - \mathbb{E}(Y_j))\right)}{\sqrt{\mathbb{E}(Y_i - \mathbb{E}(Y_i)) \cdot \mathbb{E}(Y_j - \mathbb{E}(Y_j))}}$$

Covariance is really sensitive to scale of variables. For instance, if we measure distance in millimeters, the covariance would be larger than in the case of a measure expressed in metters. Thus the correlation coefficient, which is a sort of normalized covariance is useful, to be able to compare the values.

Remark 2

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_i) &= \mathbb{E}((Y_i - \mathbb{E}(Y_i))(Y_i - \mathbb{E}(Y_i))) \\
&= \mathbb{E}((Y_i - \mathbb{E}(Y_i))^2) \\
&= \mathrm{Var}(Y_i)
\end{aligned}$$

$$\Sigma = \begin{pmatrix} \mathbb{V}(Y_1) & & & \\ & \ddots & & \\ & \text{Cov}(Y_i, Y_j) & \mathbb{V}(Y_i) & \\ & & & \ddots \\ & & & & \mathbb{V}(Y_n) \end{pmatrix} \tag{2.2}$$

Definition 8: Identity matrix

$$\mathcal{I}_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Theorem 1: Cochran Theorem (Consequence)

Let $\mathbf{Z}$ be a gaussian vector: $\mathbf{Z} \sim \mathcal{N}_n(0_n, I_n)$.

- If V_1, V_n are orthogonal subspaces of $\mathbb{R}^n$ with dimensions n_1, n_2 such that

$$\mathbb{R}^n = V_1 \overset{\perp}{\oplus} V_2.$$

- If Z_1, Z_2 are orthogonal of $\mathbf{Z}$ on V_1 and V_2 i.e. $Z_1 = \Pi_{V_1}(\mathbf{Z}) = \Pi_1 \mathbf{Y}$ and $Z_2 = \Pi_{V_2}(\mathbf{Z}) = \Pi_2 \mathbf{Y}$... (look to the slides)

Definition 9: Chi 2 distribution

If $X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(0, 1)$, then;,

$$X_1^2 + \dots X_n^2 \sim \chi_n^2$$

2.6.1. Estimator's properties

$$\Pi_V = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

$$\hat{m} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

so

$$= \Pi_V \mathbf{Y}$$

According to Cochran theorem, we can deduce that the estimator of the predicted value $\hat{m}$ is independent $\hat{\sigma}^2$

All the sum of squares follows a χ^2 distribution:

...

Property 2.

2.6.2. Estimators consistency

If $q < n$,

- $\hat{\sigma}^2 \underset{n\to\infty}{\overset{\mathbb{P}}{\sigma}}{}^{*2}$.
- If $(\mathbf{X}^T\mathbf{X})^{-1}$...
- ...

We can derive statistical test from these properties.

2.7. Statistical tests

2.7.1. Student t-test

$$\frac{\hat{\theta}-\theta}{\sqrt{\frac{\widehat{\mathbb{V}}(\hat{\theta})}{n}}} \underset{H_0}{\sim} t$$

where

3 Elements of Linear Algebra

Remark 3: vector

Let u a vector, we will use interchangeably the following notations: u and $\vec{u}$

Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

Definition 10: Scalar Product (Dot Product)

$$\begin{aligned} \langle u, v \rangle &= \begin{pmatrix} u_1, \ldots, u_v \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ &= u_1 v_1 + u_2 v_2 + \ldots + u_n v_n \end{aligned}$$

We may use $\langle u, v \rangle$ or $u \cdot v$ notations.

Dot product properties

Commutative $\langle u, v \rangle = \langle v, u \rangle$

Distributive $\langle (u + v), w \rangle = \langle u, w \rangle + \langle v, w \rangle$

$\langle u, v \rangle = \|u\| \times \|v\| \times \cos(\widehat{u, v})$

$\langle a, a \rangle = \|a\|^2$

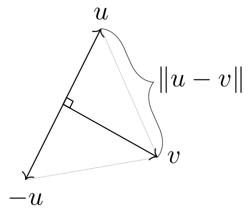

Figure 3.1. Scalar product of two orthogonal vectors.

Definition 11: Norm

Length of the vector.

$$\|u\| = \sqrt{\langle u, v\rangle}$$

$\|u, v\| > 0$

Definition 12: Distance

$$dist(u, v) = \|u - v\|$$

Definition 13: Orthogonality

Remark 4

$$(dist(u, v))^2 = \|u - v\|^2,$$

and

$$\langle v - u, v - u\rangle$$

$$\begin{aligned}\langle v - u, v - u\rangle &= \langle v, v\rangle + \langle u, u\rangle - 2\langle u, v\rangle \\ &= \|v\|^2 + \|u\|^2 \\ &= -2\langle u, v\rangle\end{aligned}$$

$$\begin{aligned}\|u - v\|^2 &= \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle \\ \|u + v\|^2 &= \|u\|^2 + \|v\|^2 + 2\langle u, v\rangle\end{aligned}$$

Proposition 1: Scalar product of orthogonal vectors

$$u \perp v \Leftrightarrow \langle u, v\rangle = 0$$

Indeed. $\|u - v\|^2 = \|u + v\|^2$, as illustrated in Figure 3.1.

$$
\begin{aligned}
&\Leftrightarrow -2\langle u, v\rangle = 2\langle u, v\rangle \\
&\Leftrightarrow 4\langle u, v\rangle = 0 \\
&\Leftrightarrow \langle u, v\rangle = 0
\end{aligned}
$$

□

Theorem 2: Pythagorean theorem

If $u \perp v$, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Definition 14: Orthogonal Projection

Let $y = \begin{pmatrix} y_1 \\ . \\ y_n \end{pmatrix} \in \mathbb{R}^n$ and w a subspace of $\mathbb{R}^n$. $\mathcal{Y}$ can be written as the orthogonal projection of y on w:

$$\mathcal{Y} = proj^w(y) + z,$$

where

$$\begin{cases} z \in w^\perp \\ proj^w(y) \in w \end{cases}$$

There is only one vector $\mathcal{Y}$ that ?

The scalar product between z and (?) is zero.

Property 3. *$proj^w(y)$ is the closest vector to y that belongs to w.*

Definition 15: Matrix

A matrix is an application, that is, a function that transform a thing into another, it is a linear function.

Example 2: Matrix application

Let A be a matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Then,

$$
\begin{aligned}
Ax &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}
\end{aligned}
$$

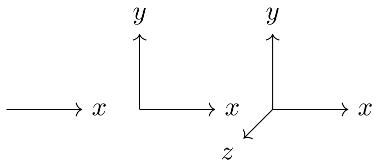

Figure 3.2. Coordinate systems

Example 2 continued

Similarly,

$$\begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 + cx_3 + dx_4 \\ ex_1 + fx_2 + gx_3 + hx_4 \\ ix_1 + jx_2 + kx_3 + lx_4 \end{pmatrix}$$

The number of columns has to be the same as the dimension of the vector to which the matrix is applied.

Definition 16: Tranpose of a Matrix

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$