MULTIVARIATE STATISTICS AND HILBERT SPACES

Luigi Beghi

Multivariate Statistics' main purpose is to define and subsequently statistically validate models of mathematical relationships among a finite set of "*measurable attributes*" (*variables*) $\{X_1, \ldots, X_n\}$ characterizing a certain domain of investigation.
Within this general frame we take into consideration the "*best fitting* " problem, where the *measurable attributes* are subdivided into a subset of "*independent or explanatory variables*" $\{X_1, \ldots, X_p\}$ and another subset $\{Y_1, \ldots, Y_q\}$ of "*dependent variables*"; a *mathematical model* of functional dependence of the Y's variables on the X's is introduced, together with an *optimality criterion* allowing for the determination of the numerical values of the parameters present in the model on the base of available experimental data. A distinct sample of experimental data will allow for the statistical validation of the model.
Let us now consider the case of "*linear least squares best fitting*", with a single dependent variable Y and a set $\{X_1, \ldots, X_p\}$ of dependent variables.

Given a sample of experimental data $\{(x_{i1}, \ldots, x_{ip}, y_i)\}$, $i = 1, \ldots, N$ and the mathematical model of *linear dependence* of Y on the X's :

$$Y = \alpha_1 X_1 + \ldots + \alpha_p X_p + \gamma ,$$

(where $\alpha_1, \ldots, \alpha_p, \gamma$ are the model's *parameters*) , the *optimality criterion* for the determination of the numerical values of parameters is the *minimization* of the *total sum of squares of residuals*

$$F(\alpha_1, \ldots, \alpha_p, \gamma) = \Sigma_i [y_i - y'_i]^2$$

where the $\{y'_i\}$ are the estimated values of the dependent variable Y, obtained according to the relation :

$$y'_i = \alpha_1 x_{i1} + \ldots + \alpha_p x_{ip} + \gamma , I = 1, \ldots, N$$

A possible solution of this problem, in the absence of mathematical *constraints* on the parameters is the "*analytical*" one, is obtained through the *vector differential equation*

$$grad\, F = 0 ,$$

equivalent to the system of linear differential equations

$$\partial F / \partial \alpha_1 = 0 , \ldots , \partial F / \partial \alpha_p = 0 , \partial F / \partial \gamma = 0$$

leading to the solution of the system of linear equations :

$$\mathrm{var}(X_1)\,\alpha_1 + \mathrm{covar}(X_1, X_2)\,\alpha_2 + \ldots + \mathrm{covar}(X_1, X_p)\,\alpha_p = \mathrm{covar}(X_1, Y)$$

$$. \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; .$$

$$\mathrm{covar}(X_p, X_1)\,\alpha_1 + \mathrm{covar}(X_p, X_2)\,\alpha_2 + \ldots + \mathrm{var}(X_p)\,\alpha_p = \mathrm{covar}(X_p, Y)$$

together with the relation :

$$\gamma = E(Y) - \alpha_1 E(X_1) - \ldots - \alpha_p E(X_p) \quad ,$$

where $E(Y)$, $E(X_1)$, . . . , $E(X_p)$ are the *expected values* (i.e.the *arithmetic means*) and $\mathrm{var}(X_1)$, . . . , $\mathrm{var}(X_p)$, $\mathrm{covar}(X_1, X_2)$, . . . , $\mathrm{covar}(X_p, Y)$ are the *variances* and the *covariances* of the variables considered above.

An equivalent approach is offered by the introduction of a *Hilbert spaces* on these variables and the subsequent use of the so-called *orthogonality principle*, as it will be shown in what follows.

Given a *Probability Space* and a generic set of random variables $V_1, \ldots, V_k$ with finite expected values $E(V_1), \ldots, E(V_k)$ and a *non singular variance-covariance matrix* B (i.e. $\det B \neq 0$) :

$$B = \begin{pmatrix} \mathrm{var}(V_1) & \mathrm{covar}(V_1, V_2) & \ldots & \mathrm{covar}(V_1, V_k) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{covar}(V_k, V_1) & \mathrm{covar}(V_k, V_2) & \ldots & \mathrm{var}(V_k) \end{pmatrix}$$

and finally assuming, for the sake of simplicity, $E(V_1) = \ldots = E(V_k) = 0$, we define the *norms* of these r.v.'s :

$$\| V_1 \| = \mathrm{var}(V_1)^{1/2} , \ldots , \| V_k \| = \mathrm{var}(V_k)^{1/2}$$

and their « *scalar products* » :

$$< V_r , V_s > = \mathrm{covar}(V_r, V_s) , \; r, s = 1 , \ldots , k$$

in such a way that

$$\| V_r \| = < V_r , V_r >^{1/2} \; .$$

The *scalar product* that we just defined satisfies the following properties:

1) $< x , y > = < y , x >$,
2) $< x , y + z > = < x , y > + < x , z >$
3) $< \lambda x , y > = | \lambda | < x , y >$ **(where λ is an arbitrary real number, called *scalar*)**
4) $< x , x > \geq 0$, $< x , x > = 0 \Rightarrow x = 0$.

We say that x , y are *orthogonal* if $< x , y > = 0$ while their *norms* are positive.

The set of random variables $\{V_1 , \ldots , V_k\}$ with the *scalar product* defined above constitutes a "*Hilbert Space*" *of finite dimension* k .

Given a sub-set $\{X_1 , \ldots , X_p\}$ ($p \leq k$) such that the determinant of their variance-covariance matrix $\neq 0$, we define as *linear variety* $L[X_1 , \ldots , X_p]$ the set of all their possible *linear combinations*

$$\alpha_1 X_1 + \ldots + \alpha_p X_p$$

where $\alpha_1 , \ldots , \alpha_p$ are arbitrary real numbers.

Considering now a generic random variable . Y of the same Hilbert Space , we define as its *orthogonal projection* Y' on $L[X_1 , \ldots , X_p]$ the random variable Y' of $L[X_1 , \ldots , X_p]$, i.e. $Y' = \alpha_1 X_1 + \ldots + \alpha_p X_p$, such that it satisfies the following orthogonality conditions:

$$< Y - Y' , X_1 > = 0 \ , \ldots , \quad < Y - Y' , X_p > = 0$$

This relations are equivalent to the system of equations:

$$< X_1 , X_1 > \alpha_1 + < X_1 , X_2 > \alpha_2 + \ldots + < X_1 , X_p > \alpha_p = < X_1 , Y >$$

. . . .

$$< X_p , X_1 > \alpha_1 + < X_p , X_2 > \alpha_2 + \ldots + < X_p , X_p > \alpha_p = < X_p , Y >$$

Orthogonality principle **: the orthogonal projection Y' on $L[X_1 , \ldots , X_p]$ exists and is unique and satisfies the minimality condition**

$$\| Y - (\alpha^*_1 X_1 + \ldots + \alpha^*_p X_p) \| = \min$$

The values of $(\alpha^*_1 , \ldots , \alpha^*_p)$ satisfy the system of linear equations : written above.

This system of linear equation is the same written before in terms of variances and covariances !

TO BE NOTICED : we can easily extend the concept of Hilbert space to random variables with expected values $\neq 0$: it will be sufficient in this case to associate to them the r. v. $X'_1 = X_1 - E(X_1)$, . . . , $X'_p = X_p - E(X_p)$, $Y' = Y - E(Y)$ and thereafter apply for these new random variables the procedure illustrated before !