RÉGRESSION (D2, J)
(21 / 10 / 2021, © Monfort, Dicostat2005, 2005-2021)

Dans chaque **domaine de connaissance**, l'**observation** d'un **phénomène** quelconque consiste à décrire diverses parties, ou constituants de ce phénomène, que le **statisticien** considère comme des **unités statistiques** supposées pertinentes pr à ce phénomène. Par suite, cette observation porte sur diverses variables attachées à ces unités constituantes. On observe ainsi, sur chacune d'elles, une liste de « **descripteurs** », conceptualisés dans la notion de **variable**. La **famille** constituée par ces variables est alors supposée « gouvernée » par une **loi de probabilité** que le statisticien cherche à caractériser.

Le concept probabiliste de **régression** est un exemple fondamental de **relation fonctionnelle** restreinte à des **variables numériques**. Il s'agit donc d'une **caractéristique légale** associée à la **famille des lois** sous-jacentes au phénomène.

Les va considérées peuvent donc a priori être scalaires ou vectorielles, ou encore être des **variables discrètes** ou des **variables continues**.

La **théorie de la régression** conduit à distinguer, au sein des variables prises en compte :

(a) certaines, dites **variables endogènes**, ou **variables expliquées**, ou **effets** ;

(b) d'autres, appelées **variables exogènes**, **variables explicatives**, ou **causes** (cf aussi **causalité**).

Par ailleurs, chacune de ces deux sous-familles de **va** peut comporter :

(a) des « variables univariées », ou « variables simples », ou encore « variables uniques » : ce sont des descripteurs de type « scalaire » ou « à une dimension » (eg taille, poids, pression, vitesse, nombre d'enfants, etc) ;

(b) des « variables multivariées », ou « variables complexes », ou encore « variables multiples » : eg « liste » ou « ensemble » de variables simples », ou encore « variable vectorielle » (eg couple « taille x poids », triplet « longitude, latitude, altitude » de localisation, etc) (cf **variable multidimensionnelle**, **loi multidimensionnelle**, **vecteur aléatoire**).

Autrement dit, la notion de régression fait référence (plus ou moins implicite) au schéma général suivant.

types de régressions

variable		exogène	
	type	simple	multiple
endogène	simple	numérique	numérique
	multiple	numérique	numérique

Une régression constitue une façon d'exprimer les endogènes, alors appelées « régressandes », en fonction des exogènes, appelées « régresseurs ». Ce concept probabiliste, exprimé dans l'**espace des variables**, correspond à la notion de **loi scientifique**, tq l'**homme de l'art** peut la concevoir (cf **loi**).

Il s'agit d'une importante notion mathématique (ie probabiliste) qui se traduit, en **Statistique** (F. GALTON), sous la forme de **« liaison moyenne »** ou, plus généralement, de **« liaison centrale »**, entre les deux groupes de variables précédents. Cette liaison est construite par **conditionnement** entre ces variables.

Selon le **contexte**, ce concept comporte des dénominations spécifiques : cf notamment **modèle de régression multiple**, **régressions multiples**, **régression multidimensionnelle**, **modèle d'interdépendance**.

(i) Soit $(\Omega, \mathcal{T})$ un **ensemble fondamental**, $(\mathcal{Z}, \mathcal{D}) = (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{C})$ un **espace d'observation** numérique produit et $\zeta = (\xi, \eta) : \Omega \mapsto \mathcal{X} \times \mathcal{Y}$ un **couple aléatoire**. On privilégie la seconde variable η, appelée **variable endogène**, à valeurs dans $\mathcal{Y}$. La première variable ξ, à valeurs dans $\mathcal{X}$, est appelée **variable exogène**. On admet généralement que la **loi de probabilité** $P^{(\xi, \eta)}$ de (ξ, η) parcourt une famille $\mathcal{P}^{(\xi, \eta)}$ de **lp**, et que l'une d'elles est la « vraie » **loi** qui gouverne le phénomène.

La loi $P^{(\xi, \eta)}$ est généralement complexe. Pour étudier η, deux attitudes sont possibles :

(a) on peut s'intéresser seulement à sa loi P^{η}, ie à la **loi marginale** (ou « **loi propre** ») de η , mais on perd l'information qui solidarise η avec ξ ;

(b) dans la mesure où l'on cherche à préciser le « comportement » de η pr à ξ, on doit supposer l'existence d'une **dépendance** de η pr à ξ ; le concept de **loi conditionnelle** de η relativement à ξ est le concept approprié pour formaliser cette dépendance. Cette loi est notée $\mathcal{L}(\eta / \xi)$ ou $P^{(\eta / \xi)}$. A la famille $\mathcal{P}^{(\xi, \eta)}$ correspond alors la famille $\mathcal{P}^{(\eta / \xi)}$ des lois conditionnelles $P^{(\eta / \xi)}$ précédentes.

Chacune des lois conditionnelles se calcule à l'aide du **théorème des probabilités composées**. Ainsi, avec des variables numériques et un **modèle dominé** par une **mesure positive** (ie possédant des lois à densités), le calcul « formel » $P(A \cap B) = P(A) \cdot P(B / A)$ conduit à exprimer la **densité conditionnelle** selon : $g(y / x) = h(x, y) / f(x)$, où $h(x, y)$ désigne la **densité** (jointe) du **couple aléatoire** (ξ, η) et $f(x)$ la **densité marginale** en x.

(ii) Dans le cadre de la régression, l'**inférence statistique** ne porte pas sur les **lois conjointes** $P^{(\xi, \eta)}$, ni même sur les lois conditionnelles $P^{(\eta / \xi)}$, mais simplement sur une **caractéristique** de ces dernières. Cette caractéristique, notée $C(\eta / \xi)$, de $P^{(\eta / \xi)}$ est une **caractéristique conditionnelle**, puisqu'elle dépend de ξ : $C(\eta / \xi)$ prend ses valeurs dans $\mathcal{Y}$. Elle est définie par le « **mode opératoire** » spécifique retenu (calcul d'**espérance**, de **mode**, de **quantile**, etc). C'est donc l'image de $\mathcal{P}^{(\eta / \xi)}$ dans $\mathcal{Y}$ définie par une **application caractéristique** (conditionnelle) $c : \mathcal{P}^{(\eta / \xi)} \mapsto \mathcal{Y}$.

La notion (probabiliste) de **fonction de régression** se déduit alors directement de $C(\eta / \xi)$. Elle est définie comme l'**application** :

(0) $\quad x \in \mathcal{X} \mapsto r = \rho(x) = C(\eta / \xi = x) \in \mathcal{Y}$ (p.s.),

dans laquelle $r \in \mathcal{Y}$ est la valeur (générique) de la caractéristique associée à une valeur $x \in \mathcal{X}$.

La fonction ρ est un **concept probabiliste** qui résulte de la loi $P^{(\eta / \xi)}$ à laquelle on applique le mode opératoire (application caractéristique conditionnelle) définissant ρ. Quand $P^{(\eta / \xi)}$ varie dans $\mathcal{P}^{(\eta / \xi)}$, ρ varie dans la famille $\mathcal{R}$ des fonctions de régression associée à la famille $\mathcal{P}^{(\eta / \xi)}$.

L'écriture générale précédente est souvent qualifiée de « non paramétrée », ou « non paramétrique ». On parle alors de « **régression non paramétrée** » ou de « **régression non paramétrique** ».

(iii) Si la famille initiale $\mathcal{P}^{(\xi, \eta)}$ est de type connu (eg **famille exponentielle** : **lois gaussiennes**, etc), les fonctions ρ associée aux $C(\eta / \xi = x)$ résultent directement, par calcul, des lois $P^{(\xi, \eta)}$: elles n'ont donc pas besoin d'être spécifiées « séparément ».

(iv) En général, $\mathcal{P}^{(\xi, \eta)}$ n'est pas donnée (ni connue) a priori. On doit donc définir une **spécification** particulière pour ρ, eg admettre la véracité d'une hypothèse particulière, eg :

(a) ρ est de type semi-paramétrique (symbole s), avec $\rho_s(x) = \rho(x, \theta)$, où ρ n'est pas connue mais dépend d'un **paramètre d'intérêt** interprétable $\theta \in \Theta$ (cf **modèle semi-paramétrique**) ;

(b) ρ est de type paramétrique (symbole p), avec $\rho_p(x) = \rho(x, \theta)$, où ρ est connue (eg linéaire, logarithmique ou quadratique) et dépend aussi d'un paramètre $\theta \in \Theta$ (cf **modèle paramétrique**).

Ces spécifications (**hypothèses**) doivent faire l'objet de **tests** pour être validées. Notamment, l'importance d'une **erreur de spécification** relative à ρ peut s'apprécier à partir d'une « distance » δ (ρ_s , ρ) ou δ (ρ_p , ρ) entre la « vraie » fonction de régression ρ (parfois notée ρ^*) et celle (ρ_s ou ρ_p) retenue pour l'analyse.

Dans l'appréciation de cette erreur, la composition de la « liste » des variables endogènes η ou celle de la liste des variables exogènes ξ peu(ven)t aussi intervenir : ce sont les formes analytiques retenues pour ρ_s ou ρ_p qui déterminent ces listes (inclusion, ou exclusion, de variables).

En effet, le statisticien n'a pas toujours connaissance de la totalité des variables à prendre en compte ;

(a) certaines variables, relevant du champ de connaissance auquel le phénomène appartient, peuvent, pour diverses raisons :

$(a)_1$ être négligées car relevant d'autres phénomènes de ce champ ;

$(a)_2$ être inobservables (cf **inobservabilité**) ;

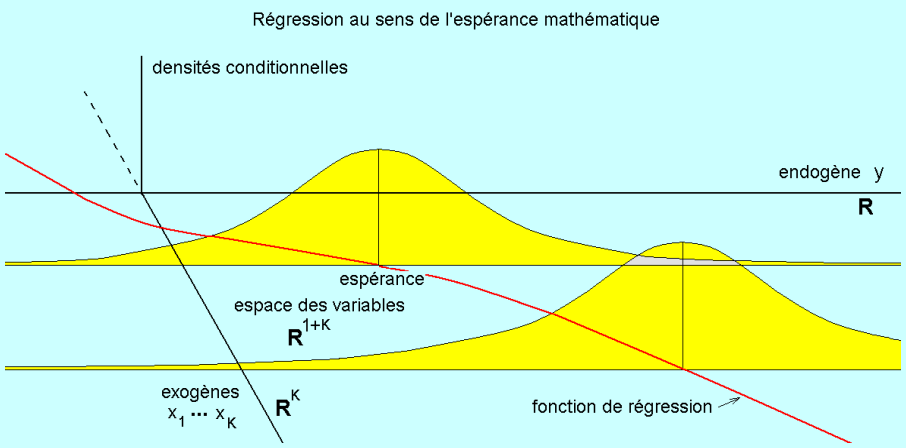
(b) certaines variables peuvent aussi décrire un phénomène relevant d'un autre champ de connaissance.

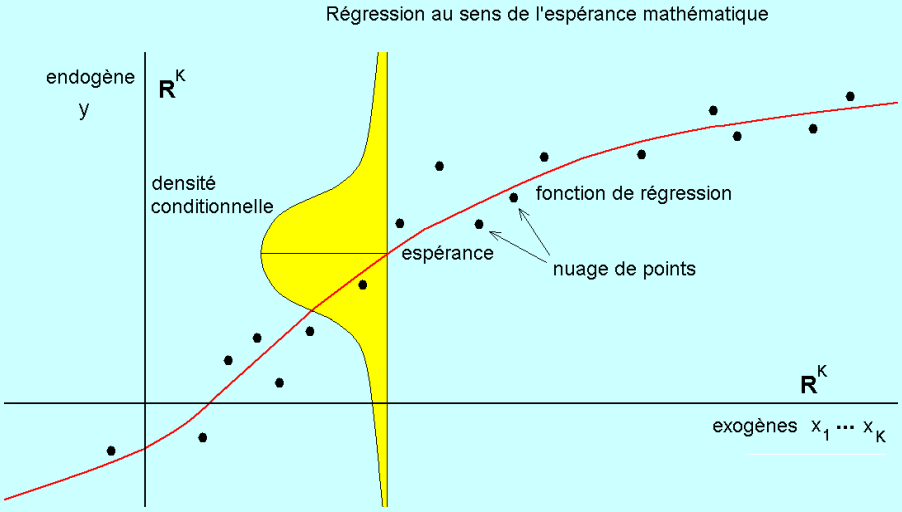
(v) La caractéristique considérée pour définir une régression est le plus souvent, par nature, une caractéristique de **centralité** (cf aussi **paramètre de position**) :

(a) l'**espérance mathématique** est très largement utilisée, en raison de ses propriétés analytiques commodes (**opérateur linéaire**, affinité avec les notions de **forme quadratique** et de **dispersion**, **théorème de la limite centrale**) ;

Ainsi, lorsque la notion d'**espérance conditionnelle** a un sens, elle se concrétise par les concepts usuels de **courbe de régression**, **surface de régression** ou, plus généralement, **variété de régression**. L'équation générale correspondante s'écrit, dans l'**espace des variables**, et sous forme paramétrée (cf schémas ci-dessous pour l'espérance) :

(1) $\quad x \in \mathcal{X} \mapsto r = \rho(x) = E(\eta / \xi = x)$ (p.s.).





La **forme explicite** (1) est très courante.

Il existe aussi une **forme implicite**, dérivée de la notion de **fonction d'interdépendance** (cf **interdépendance**, **modèle d'interdépendance**) ;

(b) d'autres caractéristiques de centralité (**médiane**, **mode**) peuvent aussi être mises en oeuvre. Elles ont conduit aux notions de :

$(b)_1$ **régression médiane**, cas particulier de **régression quantilaire** (cf aussi **quantile**) ;

$(b)_2$ **régression modale**.

En effet, l'espérance peut posséder un sens clair lorsqu'elle est associée à une **loi symétrique** et dotée de moments d'ordres 1 ou 2. Mais lorsque la loi sous-jacente au phénomène considéré est asymétrique ou sans moment, ces dernières caractéristiques sont souvent plus appropriées (cf **forme légale**) ;

(c) cependant, la caractéristique considérée ne se réduit pas nécessairement à une caractéristique de centralité. Elle peut, alternativement, être une caractéristique de **dispersion** (cf **paramètre d'échelle**), ou encore une **caractéristique de forme**, résumant donc la **forme** des lois (cf **forme légale**) ;

(d) enfin, cette caractéristique peut ne pas être « ponctuelle » (scalaire ou vectorielle), mais « ensembliste » : « plages » de valeurs décrivant la centralité (eg **zone** centrale), la dispersion, la forme.

(vi) La fonction de régression $x \in \mathcal{X} \mapsto r = \rho(x) = C(\eta / \xi = x) \in \mathcal{Y}$ est à la base d'une « **décomposition tautologique** » triviale (mais fondamentale) de type b = a + (b - a) (cas d'un **groupe algébrique** additif), ou encore b = a . (b / a) (cas d'un groupe multiplicatif). Dans le cas additif, la va η se décompose alors selon (on suppose les opérations sensées) :

(2) $\eta = C(\eta / \xi = x) + \{\eta - C(\eta / \xi = x)\} = \rho(x) + \varepsilon$,

où $\varepsilon = \eta - C(\eta / \xi = x)$ joue le rôle de **perturbation aléatoire**, souvent appelée **« erreur sur l'équation »**. Cette perturbation, à caractère synthétique, peut s'interpréter comme un écart dû à la non prises en compte de certaines variables pertinentes (cf (iv) supra).

Dans ce cas, et pour que l'approche conserve son homogénéité, la va ε doit posséder des propriétés stochastiques compatibles avec le mode opératoire (application caractéristique c) qui définit $C(. / \xi)$, ie :

(a) la caractéristique de centralité $C(\varepsilon / \xi)$ se substitue à l'espérance, et doit alors vérifier $C(\varepsilon / \xi) = 0$;

(b) la **variabilité** de ε doit être définie à l'aide d'une caractéristique de **dispersion** appropriée (ie qui dépend de la nature de C) et notée eg $D(\varepsilon / \xi)$: on peut alors définir comme **indicateur de dispersion** cette même caractéristique appliquée à l'écart, ie $C\{\eta - C(\eta / \xi = x) / \xi = x\}$.

(vii) Le cadre précédent est un cadre « théorique », défini dans l'espace des variables. Il permet à l'**homme de l'art** de mettre en forme une théorie (cf **loi**) : spécification des listes de variables, de la forme de la liaison, etc.

Il en est ainsi dans toutes les sciences : une « loi » scientifique est alors une liaison de forme spécifique entre diverses variables décrivant un phénomène particulier. Le concept probabiliste de régression intervient donc de façon centrale. Ainsi :

(a) physique : lois de R. BOYLE - E. MARIOTTE, loi de I. NEWTON, lois de la relativité restreinte ;

(b) biologie : lois de la génétique, lois des réseaux neuronaux ;

(c) écologie : loi de K. SPRENGEL - J. de LIEBIG, loi de V.E. SHELFORD, lois de prédation ou de compétition ;

(d) psychologie : lois décrivant l'action de divers facteurs sur l'intelligence, la mémoire, l'affectif ;

(e) sociologie : lois d'association, de différenciation, d'opposition, etc. Plus particulièrement (sociologie : économie), et dans un seul but pédagogique, un modèle keynésien (très élémentaire) peut s'écrire :

$C = c \,.\, R + b + \varepsilon_C$ (fonction de consommation, multiplicateur),

(2)' $I = k \,.\, (R - R_{-1}) + \varepsilon_I$ (fonction d'investissement, accélérateur),

$R = C + I$ (identité comptable),

où R désigne le revenu national (R_{-1} étant son retardé d'une période), C la consommation, I l'investissement. Les paramètres d'intérêt sont $c \in]0, 1[$ (propension marginale à consommer), $b \geq 0$ (consommation « incompressible ») et $k \geq 0$ (rapport entre le capital marginal et la production marginale, en anglais ICOR), ε_C et ε_I jouant le rôle de variables d'écart (perturbations aléatoires).

La fonction de consommation est une « loi » traduisant, sous forme affine, la façon dont le consommateur utilise son revenu, ainsi que l'effet sur l'économie que ce dernier peut entraîner. La fonction d'investissement est une loi traduisant ici, sous forme très élémentaire (ie linéaire), la façon dont les variations de production sont investies, ainsi que les fluctuations d'investissement que cela peut entraîner.

(viii) On peut supposer l'existence d'un **changement de nature** des lois selon le **domaine de connaissance**. Dans l'approche constructive des cinq domaines fondamentaux, les lois d'un « niveau » donné doivent, en principe, se déduire de celles du niveau précédent.

De même, on peut observer que la « nature » même de certaines lois peut différer, non seulement entre domaines, mais aussi à l'intérieur des domaines. Ainsi (économie), il existe une propriété de « no bridge » entre les lois macroéconomiques et les lois microéconomiques : l'agrégation ne conserve invariantes la forme des lois microéconomique que dans des cas très restreints (relations affines).

(ix) Pour procéder à l'**inférence statistique**, on utilise le « produit » d'un **système d'observation** (système descriptif, système de mesure :

(a) **données** (ou **observations**) de conception non statistique : données « spontanées », données « opportunistes », etc ;

(b) ou données statistiquement élaborées (**expériences**, **sondages**).

Ce **système statistique** (**dispositif expérimental**, etc) permet ainsi de disposer d'un ensemble d'observations Y de η et X de ξ : ces observations se réfèrent, en général, aux mêmes **unités statistiques** (données en **coupes instantanées**) ou aux mêmes instants d'observation (cas des **processus** ou des **séries temporelles**), ou encore aux mêmes **zones** de l'espace. C'est grâce à ces observations que la « théorie » résumée dans la définition d'une régression peut être « appréciée » et validée.

De plus, l'existence de « liaisons » éventuelles entre les observations elles-mêmes permet d'enrichir et d'étendre la notion de régression, mais aussi de la rendre plus complexe (cf **complexité**) : ceci conduit à définir, dans l'**espace des observations**, les concepts de **modèle de régression** ou de **modèle d'interdépendance**. Chacun d'eux constitue donc un **modèle statistique** particulier, dont le « paramètre » (de type fonctionnel) est la fonction de régression ou la fonction d'**interdépendance**.

En pratique, l'étude des données ne se borne pas à formuler un modèle tq (1) ou (2). Elle spécifie d'emblée un modèle reliant les observations (X, Y) des variables (ξ, η) concernées. Elle incorpore, en outre, un maximum d'informations : structure des observations, particularités stochastiques (inter-relations, variabilités, corrélations, etc).

Cependant, ce modèle doit toujours pouvoir se particulariser (ou se « marginaliser ») sous la forme théorique (1) ou (2) qui l'« engendre » : seule cette forme possède une interprétation concrète et spontanée. L'**homme de l'art** qui est conduit à formaliser de tels modèles n'est pas toujours un **statisticien** : ainsi (économie), l'économiste s'intéresse aux équations tq (2)' plutôt qu'à celles résultant de leur « **écriture statistique** ».

Cette écriture statistique (dans l'**espace des observations**) est la suivante. Dans le cas le plus élémentaire, les observations du couple (ξ, η) constituent un N-**échantillon iid** (X, Y), et le **modèle statistique** considéré s'écrit $(\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{C}, \mathcal{P}^{(\xi, \eta)})^{\otimes N}$, ou encore sous la forme $\{(\mathcal{X} \times \mathcal{Y})^N, (\mathcal{B} \otimes \mathcal{C})^{\otimes N}, (\mathcal{P}^{(\xi, \eta)})^{\otimes N}\}$. La définition précédente de la régression se transpose alors à ce contexte : l'équation de régression initiale est ici représentée sous la « **forme observée** » $Y = \Phi(X) + U$, où Φ est l'application multivariée associée à cette « réécriture ».

Ainsi, lorsque ξ est un **vecteur aléatoire** réel, η une **vars**, $y = (y_N, ..., y_N)'$ le vecteur aléatoire constitué des observations y_n de η et $X \in M_{NK}(\mathbf{R})$ la **matrice d'observation** aléatoire dont les lignes X_n sont les observations correspondantes de $\xi = (\xi_1, ..., \xi_K)$ (liste des **variables exogènes**), la **« régression observée »** se réécrit, $\forall\, n \in N_N^*$:

(3) $\quad y_n = C(\eta / \xi = X_n) + u_n$, $\quad$ avec $C(u_n / \xi = X_n) = 0$,

ie :

(3)' $\quad y_n = \rho(X_n) + u_n$, $\quad$ avec $C(u_n / \xi = X_n) = 0$,

soit, sous forme vectorielle :

(3)" $y = \Phi(X) + U$, avec $\Gamma(U / X) = 0$,

où l'application Γ résulte de l' « empilement » des $C(u_n / \xi = X_n)$ $(n = 1, ..., N)$.

Selon la forme de la loi conditionnelle de η / ξ et la nature de la caractéristique conditionnelle retenue, les observations vérifiant (3) ne seront pas, en général, réparties de façon approximativement égale pr à la variété définie par C. Ceci est cependant vérifié dans le cas d'une loi symétrique et de l'espérance conditionnelle (cf **forme légale**).

(x) Par suite, l'**estimation** d'une régression peut s'effectuer à l'aide de méthodes générales, qui peuvent dépendre la nature de la fonction de régression ρ :

(a) **méthodes non paramétrique** : **méthode du noyau**, **méthode des fonctions orthogonales** ou **méthode des polynômes orthogonaux**, **méthode des fonctions splines** ;

(b) méthodes paramétriques (de type usuel) : **méthode du maximum de vraisemblance**, **méthode des moments**, méthode à distance minimale (**estimateur à distance minimale** ou **estimateur à distance minimum**).

(xi) Selon la nature (individuelle, temporelle, spatiale, etc) des observations (X, Y), ou encore selon leur structuration ou leur disposition (cf **échantillon**), de nombreuses formes de modèles de régression ou d'interdépendance on été définies (cf **classification des modèles**), qui prennent en compte de façon variée et adaptée les « liaisons » entre variables ou observations déjà indiquées : corrélations inter-individuelles, corrélations inter-temporelles, corrélations inter-zones, etc.

Ainsi, l'activité statistique conduit à compléter la notion de **« liaison centrale »** entre variables par (1) à l'aide de **liaisons centrales « statistiques »** (ou « stochastiques ») : ces dernières sont l'objet d'hypothèses diverses, appelées **« hypothèses stochastiques »** du modèle. Ces hypothèse ne font que préciser certaines propriétés relatives à la famille initiale $\mathscr{P}^{(\xi,\eta)}$.

(xii) Pour illustrer ce point avec l'espérance conditionnelle, on suppose que la fonction ρ de (1) s'écrit sous forme paramétrique tq :

(4) $y = f(x, b)$,

où $x = (x_1, ..., x_K) \in \mathbf{R}^K$, $b \in \mathbf{R}^Q$ et $y \in \mathbf{R}$. Cette forme signifie que y s'exprime en fonction de x selon :

(5) $y = E(\eta / \xi = x) + (y - E(\eta / \xi = x)) = f(x, b) + \varepsilon$,

ie que :

(6) $\eta = f(\xi, b) + \varepsilon.$

La va (inobservable) ε, appelée **perturbation de l'équation**, est supposée exercer un **effet** « moyen » nul sur la variable η, ie :

(7) $E(\varepsilon / \xi) = 0.$

Les observations (supposées iid) sont indicées par $n \in N_N^*$ et l'on note y le vecteur des observations de η et X la matrice des observations de ξ. L'équation (6) peut s'écrire, pour chaque observation y_n de η et $X_n = (x_{n1}, ..., x_{nK})$ (n-ième ligne de X) de ξ :

(8) $y_n = f(X_n, b) + u_n, \qquad \forall n \in N_N^*,$

où u_n est une « **copie** » de ε. Sous forme matricielle, les équations (8) s'écrivent :

(9) $y = F(b) + u,$

où H : $\mathbf{R}^Q \mapsto \mathbf{R}^N$ est une fonction connue, qui dépend de X, $b \in \mathbf{R}^Q$ et $u = (u_1, ..., u_N)$ est un vecteur aléatoire dont les coordonnées sont de moyenne nulle, ie $E\, u_n = 0$, $\forall n \in N_N^*$ (ie $E(u / X) = 0$).

(xiii) La construction précédente peut être davantage spécifiée si les observations, indicées par $n \in N_N^*$, peuvent être supposées sans corrélation (ou indépendantes) entre elles : en effet, une **expérience aléatoire** peut parfois être conçue en sorte que cette hypothèse soit vérifiée (au moins approximativement ou asymptotiquement). Dans d'autres circonstances, une telle hypothèse est plausible ou simplement commode (mais elle devra, en principe, être préalablement testée). Elle s'exprime selon :

(10) $V(u / X) = V(y / X) = \sigma^2 I_N,$

où $\sigma^2 = V\, u_n$ est la **variance** commune des perturbations u_n (resp des endogènes y_n), ce qui suppose donc une hypothèse simplificatrice (à tester éventuellement, elle aussi) d'**homogénéité** au second ordre (ie en variance) des perturbations (resp des endogènes). Dans l'exemple, on a été conduit à préciser (ie à « spécifier ») le modèle initial (6) compte tenu d'une hypothèse stochastique (7) sur la perturbation (quand à son influence moyenne) et d'une seconde hypothèse stochastique (10) sur les copies de cette perturbation (quand à leur influence en termes de **variabilité** sur le vecteur d'intérêt y).

(xiv) La distinction terminologique entre **« régression »** et **« modèle de régression »** (resp entre **« interdépendance »** et **« modèle d'interdépendance »**) n'est pas absolue, mais de simple commodité :

(a) on réserve plutôt :

(a)$_1$ les termes **équation de « régression »** (ou simplement « régression ») et **équation d'« interdépendance »** (ou simplement « interdépendance ») à des équations définies, dans l'**espace des variables**, à partir de l'espérance conditionnelle, et dont toutes les variables sont numériques (de type « continu »). Ces notions correspondent à l'un des concepts de **loi scientifique** ;

(a)$_2$ les expressions « **modèle de régression** » ou « **modèle d'interdépendance** » à des équations définies dans l'**espace des observations** et dont les toutes les variables sont numériques (et « continues »). Ces notions préparent l'**inférence statistique** puisqu'elles incorporent les équations précédentes dans une **représentation statistique** ;

(b) on confond souvent la procédure ou la méthode (d'estimation) utilisée avec le modèle de régression ou d'interdépendance auxquel elle s'applique. Ainsi, on parle de :

(b)$_1$ de « **régression pas à pas** » pour désigner une procédure itérative (ou adaptative) d'estimation et de test fondée sur une régression, etc ;

(b)$_2$ de « **régression orthogonale** » pour désigner une méthode d'estimation des paramètres d'un modèle de régression ;

(b)$_3$ de **régression modale** pour désigner une régression basée sur la notion de **mode**, de **régression quantilaire** pour désigner une régression basée sur une notion de **quantile** (cf aussi **fonction quantile**), etc.

(xv) Diverses extensions ou applications du modèle de régression (ou du modèle d'interdépendance) ont été définies :

(a) ainsi, les concepts usuels de régression ont été étendus à d'autres situations :

(a)$_1$ types de variables différents : va discrètes, va qualitatives ;

(a)$_2$ caractéristiques conditionnelles différentes (cf supra).

Ces variables ou caractéristiques définissent des modèles « voisins » dans leur esprit, dénommés avec des expressions terminologiques spécifiques : cf eg **modèle d'analyse de la variance**, **modèle d'analyse de la covariance**, **modèle à variable dépendante qualitative**, **modèle qualitatif**, **modèle quantitatif**, **modèle mixte** ;

(b) d'autres extensions prennent en compte eg :

(b)$_1$ l'existence d'observations en **temps** continu (cf **modèle de processus**) ;

(b)$_2$ l'existence d'une **censure** possible sur les observations ;

(b)$_3$ l'existence d'observations atypiques (**aberrations**) dans le cas où la famille $\mathscr{P}^{(\xi, \eta)}$, ou $\mathscr{P}^{(X, Y)}$, concerne des **lois à queue épaisse**, ou des **mélanges de lois**, où qui ne sont pas les « bonnes lois » (**erreur de spécification** sur la liste (ξ, η) des variables) (cf aussi **robustesse**) ;

(b)$_4$ l'existence de contraintes diverses portant sur les observations ou sur les paramètres d'intérêt (cf **contrainte sur les paramètres**, **contrainte sur les variables**).
&^{22}