STIMULUS (C1, J2, L1, L5, O)
(11 / 10 / 2020, © Monfort, Dicostat2005, 2005-2020)

Dans certains **domaines de connaissance** (biologie, psychologie) on peut appliquer à des **unités statistiques** (bactéries, individus) un **stimulus** ou même plusieurs stimuli :

(a) biologie : caractéristiques physico-chimiques d'un milieu ambiant pour une bactérie ou un virus. Ainsi, on peut placer une charge virale (grippe, rage, COVID19, etc) dans des milieux combinant eg température, hygrométrie, Ph, luminosité (UV) ou oxygène ;

(b) psychologie : épreuve de type I. PAVLOV pour un animal, test d'aptitude pour une personne physique (eg QI), etc.

Chaque stimulus est appliqué à différents « **niveaux** » (de façon continue ou discontinue, suivant sa nature), appelés « **doses** ». Plusieurs stimuli de types différents peuvent ainsi combiner des doses de type différent, chacune étant appelée **traitement** (cf **plan d'expérience**).

Chaque **réponse** aux stimuli peut être

(a) parfois de type continu, la réponse étant graduellement variable en fonction des traitements appliqués à chaque unité. Ainsi, la « gravité » d'une pathologie peut varier en fonction de l'accumulation d'une charge virale dans l'organisme infecté ;

(b) parfois de type discontinu. Dans ce cas, on assimile souvent cette réponse à une **variable qualitative** généralement dichotomique, appelée **variable quantale**, ou variable par « quanta ». La proportion de réponses, dans un sens ou dans l'autre, constitue, notamment, un **paramètre d'intérêt** usuel. Aussi, l'emploi d'un **modèle qualitatif,** tel que le **modèle Logit**, le **modèle Probit**, ou le **modèle Normit**, est-il courant dans ce type de **situation statistique** : l'objectif de ces modèles est d'expliciter (estimer) la probabilité (proportion théorique) de réponse en fonction de divers **facteurs**.

La structure générale de ceux-ci se décrit comme suit.

(i) On considère un **espace probabilisé** $(\Omega, \mathcal{T}, P)$, une **variable endogène** $\eta : \Omega \mapsto \{0, 1\}$ (réponse), un **vecteur aléatoire** $\xi = (\xi_1 ,..., \xi_K)' : \Omega \mapsto \mathbf{R}^K$ (représentant les stimuli) et une **fonction de répartition** F. On suppose alors que la **loi** P^η (conditionnelle en ξ) de la variable η est définie par :

(1) $\quad P^\eta (y) = [F (\xi' b)]^{y} [1 - F (\xi' b)]^{1 - y}, \qquad \forall y \in N_1$,

où $b \in \mathbf{R}^K$ représente un **paramètre** à estimer : les paramètres b_k ($k \in N_K$*) traduisent l'effet des facteurs « exogènes ».

Ainsi, lorsque F correspond, à une **fonction Logit** (resp **fonction Normit**, resp **fonction Probit**), on définit un **modèle Logit** (resp un modèle Normit, un modèle Probit).

(ii) L'estimation s'effectue généralement à l'aide de la **méthode du maximum de vraisemblance**, la **vraisemblance** du modèle s'écrivant (dans le cas où l'on admet l'indépendance entre les **unités expérimentales**, ie entre leurs « réactions » individuelles) :

(2) $L_N (y, b) = \Pi_{n=1}^{N} P^{\eta} (y_n) = \Pi_{n=1}^{N} \{F (X_n b)\}^{y(n)} . \{1 - F (X_n b)\}^{1 - y(n)}$,

où $y = (y_1 ,..., y_N)'$ est un ensemble d'observations de η et $X_n = (x_{n1} ,..., x_{nK})$ $(n = 1..., N)$ une suite d'observations de ξ , disposée en ligne, et constituant une **matrice** $X \in M_{NK}$ (**R**).

L'**équation de vraisemblance** est alors de la forme (non linéaire) suivante :

(3) $X' y = X' F_N (b)$,

où $F_N (b) = (F (X_1 b) ,..., F (X_N b))'$.