

FACTEUR EXPÉRIMENTAL (L4)

(03 / 04 / 2018)

On appelle **facteur expérimental** une **variable** susceptible d'influer sur les **unités expérimentales** utilisées (**matériel expérimental**), donc sur les résultats de l'expérience. En pratique, un **facteur** est un ensemble de conditions de même nature pouvant influencer le phénomène illustré par l'expérience et, notamment, la **variable endogène** qui mesure les résultats de l'expérience (cf **dispositif expérimental**, **plan d'expérience**).

Ainsi, on peut considérer des facteurs « température », « pression », « hygrométrie », « luminosité » (physique : météorologie) ; du facteur « dosage » (biologie) ; des facteurs « exposition », « pluviométrie », « type de sol », « variété végétale » (écologie : agronomie) ; du facteur « **stimulus** » (psychologie), etc.

(i) Lors d'une expérience, on choisit généralement, dans l'ensemble, plus ou moins vaste, des facteurs possibles, un sous-ensemble fini $F = \{F_1, \dots, F_H\}$, constitué de H facteurs F_h . Chacun d'eux peut être assimilé à une **variable statistique** : soit une variable certaine (ou mathématique), soit une **variable aléatoire**.

Une telle variable peut être :

(a) une **variable quantitative**, ie une variable numérique : eg une variable continue (température, pression, degré hygrométrique, etc) ou une variable discontinue (stimulus fractionné par doses) ;

(b) une **variable qualitative** : luminosité, couleur, action (ou inaction) d'un certain type, etc.

Les différentes valeurs, numériques ou non, d'un facteur F_h sont appelées ses **niveaux**, ou parfois ses **modalités** : niveaux de température, de pression, d'humidité, de dosage (ie **doses**), etc.

En général, les niveaux d'un facteur F_h sont en nombre fini : ainsi F_h comportera l_h niveaux $F_{h,i(h)}$ ($i_h = 1, \dots, l_h$), en notant aussi $i(h)$ pour désigner i_h , et ces niveaux sont souvent simplement notés $i_h \in \{1, \dots, l_h\}$. On peut alors noter $\mathcal{F}_h = \{F_{h1}, \dots, F_{hi(h)}\}$ et $\mathcal{I}_h = \{1, \dots, l_h\}$.

(ii) On appelle aussi **facteur** F_h une application $f_h : A \mapsto F_h$ (ou $A \mapsto \mathcal{F}_h$).

Un **facteur équilibré** est alors tq $\text{card } f_h^{-1}(F_{hi(h)})$ ou $\text{card } f_h^{-1}(i_h)$ ne dépend pas de h , $\forall i_h \in \mathcal{I}_h$.

Si $f_h : A \mapsto F_h$ et $f_k : A \mapsto F_k$ sont deux facteurs expérimentaux, on dit que f_h est un **facteur emboîté** dans f_k ssi il existe une application $\psi : F_h \mapsto F_k$ tq $\psi \circ f_h = f_k$. On dit alors que le niveau $F_{hi(h)}$ est emboîté dans le niveau $F_{ki(k)}$ ssi l'on a $\psi(F_{hi(h)}) = F_{ki(k)}$.

(iii) On distingue généralement entre (cf schéma ci-dessous) :

typologie simplifiée des caractéristiques factorielles

	observable	inobservable
connu	directement contrôlable	indirectement contrôlable
inconnu	identifiable	non contrôlable

(a) **facteurs connus** (en général **observables**) et **facteurs inconnus** (observables ou **inobservables**). Il arrive, en effet, qu'on ignore si certaines variables ont une influence réelle sur le **phénomène** étudié (ces variables sont donc négligées à tort). Inversement, on peut croire (à tort) que certaines variables ont une action sur le phénomène, alors que ce n'est pas le cas (artéfact, causes causées et non causantes, etc). L'un des buts de l'analyse statistique est de déterminer leur statut « réel » ;

(b) **facteurs observables** et **facteurs inobservables**. Les premiers sont généralement ceux que le dispositif d'observation (cf **production statistique**), qui concerne un matériel expérimental donné, tente de mesurer (cf **système d'observation**). Les seconds font généralement partie des variables que l'**homme de l'art** ou le **statisticien** soupçonnent influencer sur les résultats sans pouvoir les observer (concepts non « mesurables »). Certains facteurs peuvent même ne pas être identifiables (ou « repérables »).

Il est cependant parfois possible de « construire » des facteurs inobservables : eg facteurs cachés et **vecteurs propres** associés, en **analyse des données**.

(c) **facteurs directement contrôlables** et **facteurs indirectement contrôlables**.

Seuls les facteurs connus et observables peuvent être intégrés à une expérience. Les autres interviendront sur les conditions de réalisation de l'expérience en tant que **facteurs « parasites »**, ou **facteurs « importuns »**, généralement considérés comme des **variables aléatoires** : ceci explique, notamment, la mise en oeuvre de divers **modèles statistiques** (**modèle linéaire**, **modèle d'analyse de la variance**, **modèle d'analyse de la covariance**, analyse de la **surface de réponse**) en association avec la plupart des plans d'expérience.

Un facteur connu (observable ou non) n'est pas toujours contrôlable : eg les grandeurs définissant un « climat » en météorologie. Un facteur inconnu (observable ou non) n'est naturellement jamais contrôlable.

(d) **facteurs entièrement contrôlables**, **facteurs partiellement contrôlables** et **facteurs non contrôlables**.

(iv) Pour éviter de fausser une expérience à cause de l'action possible de facteurs non contrôlables, le **principe de randomisation** est souvent mis en oeuvre (cf **plan d'expérience, randomisation**).

(v) On appelle **combinaison de facteurs**, ou parfois **conjugaison de facteurs**, ou encore, le plus souvent, simplement **traitement**, la mise en oeuvre simultanée (ou conjointe), en cours d'expérience, d'un **niveau** de chaque facteur.

Un **traitement** est donc repérable par un H-uple $(F_{1\ i(1)}, \dots, F_{H\ i(H)})$, ie, en simplifiant, par $(i_1, \dots, i_H) = I$. On note alors $\mathcal{T} = \prod_{h=1}^H \mathcal{T}_h$ l'ensemble des traitements possibles ($\text{Card } \mathcal{T} = \prod_h I_h = K$). Parmi ceux-ci, seuls certains seront effectivement appliqués : on note $\mathcal{T}_e \subset \mathcal{T}$ l'ensemble de ces derniers, et l'on pose $K_e = \text{Card } \mathcal{T}_e$.

Par ailleurs, on appelle **traitement partiel** un sous-indice du multi-indice I .

Lorsque $H = 1$, les notations se simplifient. On note F le facteur considéré, $N_i^* = \{1, \dots, I\}$ l'ensemble de ses niveaux i (encore noté I , lorsqu'aucune ambiguïté n'en résulte) et N_i le nombre total des unités expérimentales soumises au niveau $i \in I$, l'une quelconque d'entre elles étant repérée par l'indice $n_i \in \{1, \dots, N_i\}$. Dans ce cas, la notion de traitement se confond avec celle de niveau du facteur F .

Lorsque $H = 2$, on note F et G les deux facteurs, N_i^* et N_j^* (ou simplement I et J) l'ensemble de leurs niveaux respectifs et N_{ij} le nombre total d'unités soumises au traitement (i, j) , chacune d'entre elles étant repérée par l'indice $n_{ij} \in \{1, \dots, N_{ij}\}$.

(vi) L'intérêt de la notion de traitement vient de ce qu'elle permet la prise en compte d'éventuelles **interactions entre facteurs** (cf **interaction, interaction factorielle, répétition, traitement**).

Enfin, lorsque aucune **structure** particulière n'est mentionnée, l'ensemble de tous les traitements possibles (resp effectifs) est aussi parfois noté $\mathcal{K}$ (resp $\mathcal{K}_e$) au lieu de $\mathcal{T}$ (resp $\mathcal{T}_e$) et les traitements $I = (i_1, \dots, i_H)$ sont alors notés $k \in \{1, \dots, K\}$ (resp $k_e \in \{1, \dots, K_e\}$), ce qui les suppose classés dans un certain ordre. Dans les notations précédentes, on peut parfois utiliser le même symbole K (resp K_e) pour représenter l'ensemble des traitements, son cardinal ou l'indice du dernier traitement (selon l'ordre précédent).