

FOVIEA

TRAITEMENT
STATISTIQUE
DES ÉTUDES
CLINIQUES

CHOIX ET TRAITEMENT
DES VARIABLES DANS LES ESSAIS
THÉRAPEUTIQUES

Lors d'un essai thérapeutique, le critère sur lequel vont porter les conclusions et qui va permettre d'apprécier l'efficacité ou toute autre propriété des traitements étudiés doit être choisi avec précision, tant sur le plan clinique que statistique, de manière à ce qu'il apporte le maximum d'informations tout en étant facilement mesurable.

Sur le plan clinique, il doit être si possible direct, sensible, spécifique, reproductible, concordant et stable.

Sur le plan statistique, le choix du type de critère de jugement, qui peut être qualitatif ou quantitatif, va conditionner le mode de calcul et donc l'importance des effectifs à recruter, le mode de présentation des données et le choix des outils statistiques nécessaires à son analyse.

Spécialisé dans le traitement statistique des études cliniques, le Département Biométrie de FOVEA vous propose au travers de ce fascicule quelques notions simples concernant

LE CHOIX ET LE TRAITEMENT DES VARIABLES DANS LES ESSAIS THÉRAPEUTIQUES.

Introduction



A l'origine, la statistique s'appliquait principalement au domaine de la démographie, c'est pourquoi le vocabulaire fait souvent référence à cette discipline en employant les termes population, individu, caractère, ...

Le critère de jugement (*ou **variable** ou **caractère***) est l'un des paramètres qui définit un individu (*ou **unité statistique** ou **unité expérimentale***), lui-même élément d'un ensemble appelé population (*ou **univers statistique***) sur lequel va porter l'analyse statistique.

L'unité statistique peut être un patient, un germe, un membre, ...

Dans une étude clinique, la problématique doit être clairement formulée, la population à étudier précisément définie, les individus concernés ainsi que les caractères qui leur sont attachés bien délimités, notamment le caractère qui constitue le critère de jugement.

Le type d'un caractère est l'élément déterminant de son analyse.

Typologie des variables

On peut classer les variables de différentes façons :

- Une variable pouvant prendre plusieurs formes mutuellement exclusives appelées modalités, classes, valeurs ou réponses, on peut distinguer :
 - **les constantes** comprenant une seule modalité,
 - **les variables bimodales** ou **booléennes** comprenant 2 modalités (*Sexe : Masculin/Féminin, Résultat : Succès-Echec, ...*),
 - **les variables plurimodales** comprenant plus de 2 modalités (*Résultat : Succès/Amélioration/Echec, Intensité d'un effet indésirable : Légère/Modérée/Sévère/Très sévère, ...*).
- les variables simples ou composites (*indice d'évaluation, aire sous la courbe, ...*).
- Les variables :
 - qualitatives (*nominales ou ordinales*),
 - quantitatives (*continues ou discrètes*).

Il n'existe aucune hiérarchie entre les modalités des variables qualitatives nominales (*Sexe : Masculin/Féminin, Yeux : Bleus/Verts/Marrons, ...*).

Par contre, il existe une hiérarchie entre les modalités des variables qualitatives ordinales (*Intensité d'un effet indésirable : Légère/Modérée/Sévère/Très sévère, ...*), mais l'écart entre 2 modalités n'est pas mesurable.

Les variables quantitatives prennent des valeurs qui sont contenues dans un intervalle de variation :

- Les variables quantitatives continues (*critères mesurables*) peuvent prendre toutes les valeurs comprises dans l'intervalle de variation (*Poids, Taille, Taux d'hémoglobine, ...*) ;
- Les variables quantitatives discrètes (*critères dénombrables*) ne peuvent prendre que des valeurs entières comprises dans l'intervalle (*Nombre de comprimés, Nombre d'enfants, ...*).

Traitement de l'information

En passant des variables nominales aux variables continues, la quantité d'informations est de plus en plus riche et les outils statistiques de plus en plus nombreux.

Une variable continue peut être analysée sous forme de classes, donc de la même façon qu'une variable ordinale, mais cela entraîne à la fois une perte de précision dans l'information et une perte de puissance dans les calculs statistiques.

En revanche, une variable ordinale peut être analysée comme une variable continue si elle remplit 3 conditions :

- nombre de cas 100,
- nombre de modalités 5,
- distribution normale.

En ce qui concerne la présentation des données, outre l'utilisation, comme pour les variables nominales et ordinales, de graphiques et de tableaux de fréquences (*absolues, relatives, cumulées*), du mode et de la médiane, les séries de variables quantitatives peuvent être résumées par toute une batterie de paramètres de position et de dispersion.

Sur le plan statistique, seuls les tests du Chi² et de la probabilité exacte de Fisher sont applicables aux variables nominales alors que l'ensemble des tests paramétriques ou non paramétriques peut être appliqué aux variables quantitatives.

Enfin, pour le calcul du nombre de sujets nécessaires, un critère de jugement quantitatif est en général moins consommateur de patients qu'un critère qualitatif (*aux mêmes hypothèses et*).

Typologie des variables



VARIABLES		RELATION ENTRE LES MODALITÉS	TESTS STATISTIQUES	PRÉSENTATION DES DONNÉES
Type	Exemples			
QUALITATIVE	NOMINALE	<p>Sexe</p> <ol style="list-style-type: none"> 1. <i>Masculin</i> 2. <i>Féminin</i> <p>• Relation d'équivalence : = ou</p> <p>La codification sert uniquement à identifier les classes</p>	<ul style="list-style-type: none"> • Chi² • Probabilité exacte de Fisher 	<p>Tableaux</p> <ul style="list-style-type: none"> • Fréquences absolues (<i>effectifs</i>) • Fréquences relatives (<i>pourcentages</i>) • Fréquences relatives cumulées (<i>pourcentages</i>)
	ORDINALE	<p>Intensité d'un effet indésirable</p> <ol style="list-style-type: none"> 1. <i>Légère</i> 2. <i>Modérée</i> 3. <i>Sévère</i> 4. <i>Très sévère</i> <p>• Relation d'équivalence : = ou</p> <p>• Relation d'ordre : < ou ></p> <p>La codification sert à la fois à identifier les classes et à les hiérarchiser entre elles</p>	<ul style="list-style-type: none"> • Test de la médiane • Test de Mann-Whitney-Wilcoxon • Test de Kruskal-Wallis • Test de Siegel-Tukey • Test des carrés des rangs • Test de Kolmogorov-Smirnov • Test de Cramer-von Mises • Test de scores normaux • Test de Jonckheere 	<p>Graphiques</p> <ul style="list-style-type: none"> • Diagramme en bâtons • Polygone des fréquences absolues • Polygone des fréquences relatives • Polygone des fréquences relatives cumulées
QUANTITATIVE	DISCRÈTE	<p>Nombre de comprimés</p> <p>Nombre d'enfants</p> <p>• Relation d'équivalence : = ou</p> <p>• Relation d'ordre : < ou ></p> <p>Ce type de variables permet d'effectuer toutes les opérations arithmétiques</p>	<ul style="list-style-type: none"> • Tests paramétriques (<i>Student, analyse de variance, ...</i>) si distribution normale • Tests non paramétriques (<i>Cf. variables Qualitatives Ordinales</i>) si distribution non normale 	<p>Tableaux</p> <ul style="list-style-type: none"> • Paramètres de position usuels (<i>moyenne arithmétique, médiane, ...</i>) • Paramètres de dispersion usuels (<i>écart-type, étendue, ...</i>) <p>Graphiques</p> <ul style="list-style-type: none"> • Histogramme des fréquences absolues • Polygone des fréquences absolues • Polygone des fréquences relatives cumulées
	CONTINUE	<p>Age</p> <p>Poids</p> <p>Taille</p>		

5

Paramètres caractérisant les distributions des variables quantitatives

PARAMÈTRES DE POSITION		PARAMÈTRES DE DISPERSION	
Paramètre	Définition	Paramètre	Définition
Mode	Valeur la plus fréquente d'une distribution observée (<i>utilisable également pour des variables nominales ou ordinales</i>)	Etendue ou Intervalle de variation ou Amplitude	C'est la différence entre la plus grande et la plus petite des valeurs observées d'une série
Médiane (M)	Valeur centrale d'une série impaire ou moyenne des 2 valeurs centrales d'une série paire (<i>utilisable également pour des variables ordinales</i>)	Intervalle interquartile	C'est la différence entre le 1 ^{er} et le 3 ^{ème} quartile ($Q3-Q1$) (<i>ce paramètre ne tenant pas compte des valeurs extrêmes, mesure la variabilité autour de la médiane</i>)
Moyenne arithmétique (m)	Somme des valeurs divisée par le nombre d'individus (<i>moyenne la plus utilisée</i>)	Ecart moyen arithmétique	C'est la moyenne arithmétique des écarts (<i>en valeur absolue</i>) entre les valeurs observées et leur moyenne arithmétique
Moyenne géométrique (G)	Racine n ^{ème} du produit des n valeurs	Variance	C'est la moyenne arithmétique des carrés des écarts entre les valeurs observées et leur moyenne arithmétique
Moyenne harmonique (H)	Inverse de la moyenne arithmétique des inverses	Ecart-type ou Déviation standard	C'est la racine carrée de la variance
Moyenne quadratique (Q)	Racine carrée de la moyenne des carrés	Coefficient de variation	C'est le rapport entre l'écart-type et la moyenne arithmétique
Quartiles	$H < G < m < Q$ Ce sont les valeurs Q1, Q2 et Q3 de la variable partageant en 4 parties d'effectifs égaux la série des observations classées par ordre non-décroissant Q2 est égal à la médiane M		
Déciles ~ Centiles	Ce sont les valeurs D1, D2, ..., D9 d'une part et C1, C2, ..., C99 d'autre part, qui partagent en 10 (<i>pour les déciles</i>) ou en 100 (<i>pour les centiles</i>) parties d'effectifs égaux la série des observations classées par ordre non-décroissant		

Calcul du nombre de sujets nécessaires



VARIABLES QUALITATIVES	VARIABLES QUANTITATIVES	COMPARAISON DE COURBES DE SURVIE
Etudes classiques de supériorité (2 groupes ~ Situation unilatérale)		
<ul style="list-style-type: none"> $n_{\text{par groupe}} = M / D_s^2$ $M = 2 (z_{1-\alpha} + z_{1-\beta})^2$ <p>(risque de 1^{ère} espèce) est le risque de mettre en évidence une différence qui n'existe pas</p> <p>(risque de 2^{ème} espèce) est le risque de ne pas mettre en évidence une différence qui existe</p> <p>1- (puissance) est la probabilité de mettre en évidence une différence qui existe</p> <p>Les valeurs de $z_{1-\alpha}$ et de $z_{1-\beta}$ peuvent être lues dans la table de répartition de la loi normale centrée réduite</p> <ul style="list-style-type: none"> $D_s = \frac{m_A - m_B}{\sigma}$ $\sigma = m_A - m_B$ <p>m_A et m_B sont les moyennes observées respectivement dans les groupes A et B</p> <p>σ est l'écart-type du critère de jugement</p>	<ul style="list-style-type: none"> $n_{\text{par groupe}} = M / D_s^2$ $M = 2 (z_{1-\alpha} + z_{1-\beta})^2$ $D_s = \sqrt{2} P_A - P_B / \sqrt{P_A Q_A + P_B Q_B}$ <p>P_A et P_B correspondent à la fréquence de réalisation de l'évènement étudié respectivement dans les groupes A et B</p> <p>$Q_A = 1 - P_A$</p> <p>$Q_B = 1 - P_B$</p>	<ul style="list-style-type: none"> $n_{\text{par groupe}} = M / (\ln \frac{B}{A})^2$ $M = 2 (z_{1-\alpha} + z_{1-\beta})^2$ $\ln \frac{B}{A}$ <p>$B > A$</p> <p>A et B correspondent à la médiane de survie respectivement dans les groupes A et B</p>
Etudes d'équivalence		
<ul style="list-style-type: none"> $n_{\text{par groupe}} = M / D_s^2$ $M = 2 (z_{1-\alpha} + z_{1-\beta})^2$ $D_s = \frac{L}{\sigma}$ L est la différence limite fixée pour l'équivalence 	<ul style="list-style-type: none"> $n_{\text{par groupe}} = (1 / L^2) M P_s Q_s$ (Formule de Blackwelder) $M = 2 (z_{1-\alpha} + z_{1-\beta})^2$ L est la différence limite fixée pour l'équivalence <p>P_s correspond à la fréquence de réalisation de l'évènement étudié avec le produit de référence ou produit standard</p> <p>$Q_s = 1 - P_s$</p>	

FOVIA