

Chapitre I: Les Tests Statistiques

Une des applications les plus fréquentes de la statistique est de tester la liaison ou la relation entre deux variables.

Test de l'Hypothèse Nulle :

La formulation de l'hypothèse nulle est la première étape du test d'hypothèse. L'hypothèse nulle, notée H_0 , est une hypothèse d'absence de différence. De manière générale, c'est une hypothèse d'indépendance.

Exemple 1

A partir d'un groupe de 100 malades, on a constitué un groupe de 54 malades traités par un traitement A et un groupe de 46 malades auxquels on a administré un traitement B. les résultats sont consignés au tableau ci-dessous

	Malades traités	Succès	
		Effectif	%
Traitement A	54	43	79,6
Traitement B	46	31	67,4

Il s'agit donc de savoir si les sujets soumis au traitement A sont caractérisés par la même fréquence relative de succès (ou d'échecs) que les sujets soumis au traitement B.

Donc dans notre exemple, H_0 pourrait s'énoncer :

- le traitement et le résultat produit sont indépendants ;
- il n'existe pas de relation entre le traitement et le résultat produit ;
- l'utilité du traitement A n'est pas différente de celle du traitement B ;
- il n'existe pas de différence entre les deux traitements.

Sous forme abrégée, si $p(A)$ et $p(B)$ désignent les fréquences relatives de succès observées pour les deux traitements, on peut écrire H_0 sous la forme :

$$H_0 : p(A) = p(B)$$

L'hypothèse nulle peut être vérifiée et est, soit rejetée, soit gardée. Cette décision est basée sur les tests statistiques qui sont appelés aussi tests de signification.

Le test statistique donne la probabilité p que le hasard puisse expliquer les résultats.

- Si la probabilité p est inférieure ou égale à un certain seuil, appelé seuil de signification, on rejette H_0 et on dit que la différence est significative.
- Si p est supérieur au seuil, on ne rejette pas H_0 et on dit que la différence n'est pas significative.

Le seuil de signification est habituellement fixé à 5 %. Il s'agit d'une convention très largement adoptée.

En tout cas, le rôle du hasard ne peut pas être éliminé totalement. Il faut bien comprendre que le jugement est fondé sur une probabilité et n'offre pas une sécurité absolue.

- Si on rejette H_0 mais cette décision risque d'être malvenue : C'est le risque de première espèce.
- Il existe un deuxième risque qui survient lorsque H_0 est retenue de façon injustifiée. C'est le risque de deuxième espèce.

En conséquence que l'on rejette ou que l'on garde une hypothèse nulle, on prend le risque de commettre un type d'erreur.

1. Test Khi Deux (X^2)

Le X^2 teste la liaison entre deux variables qualitatives. De façon générale le test du X^2 s'applique à tous les tableaux de contingence quelque soit le nombre de modalités de deux variables qualitatives.

Le principe du test du X^2 consiste à calculer pour chacune des cases du tableau de contingence l'écart quadratique :

$$\frac{(o - c)^2}{c}$$

Et à en faire la somme : $X^2 = \sum \left[\frac{(o - c)^2}{C} \right]$

Dans l'expression de l'écart quadratique,

→ o: est l'effectif observé de la case ;

→ c : est l'effectif calculé (appelé aussi effectif théorique) correspondant à la case et qui devrait être observé si H_0 était vraie.

À côté du calcul du X^2 , il faut déterminer le nombre de degrés de liberté (dl en abrégé). Celui-ci, dans un tableau de contingence à L lignes et C colonnes (c modalités pour l'autre variable), est tout simplement :

$$dl = (l - 1) \cdot (c - 1).$$

Exemples

Reprenons les données du tableau de l'exemple 1 ci-dessus relatives à la comparaison de deux traitements A et B.

Disposons les données comme dans le tableau de travail 2 où les effectifs observés dans chacune des quatre cases sont facilement déduits par différence.

L'effectif calculé correspondant à chaque effectif observé est entre parenthèses.

	succès	échec	Total
Traitement A	43(40,0)	11(14,0)	54
Traitement B	31(34,0)	15(12,0)	46
Total	74	26	100

Illustrons la façon dont l'effectif théorique correspondant à l'effectif observé 43 a été calculé.

Il s'agit donc de l'effectif des sujets qui ont reçu le traitement A et qui sont des succès.

Selon H_0 , les groupes A et B ont la même distribution en matière de résultats obtenus. Le nombre de succès au traitement A devrait être : $(74/100) \times 54 = 40,0$

Appliqué à notre exemple, le calcul du khi-deux donne :

$$X^2 = \frac{(43,0-40,0)^2}{40,0} + \frac{(31-34,0)^2}{34,0} + \frac{(11-14,0)^2}{14,0} + \frac{(15 -12,0)^2}{12,0} = 1,93.$$

Le dl étant égale à 1, la table du khi-deux montre que la probabilité p correspondante est supérieure à 10%. La différence entre les pourcentages de succès des deux traitements n'est donc pas significative contrairement à ce que laissait supposer un examen superficiel des résultats.

Chapitre II : Fluctuation d'Echantillonnage et Choix de P-Valeur

Comment sont déterminés les résultats d'un sondage (échantillonnage)?

À partir d'une population mère (taille N) on cherche un certain pourcentage (exemple X%), on va extraire un échantillon de taille (n) par une certaine méthode de sondage (la méthode appropriée), on calcule le pourcentage dans cet échantillon, par exemple (Y%), et on attribue à la population totale ce pourcentage trouvé dans l'échantillon.

-On dit qu'on a fait une estimation ponctuelle d'un (X%) à partir du (Y%) obtenu dans un échantillon.

-Dans quelle mesure peut-on faire confiance au résultat obtenu dans cet échantillon ?

-Et la question qui se pose : le résultat aurait-il été le même si on avait pris au hasard, un autre échantillon ?

Définition :

La fluctuation d'échantillonnage.

-Comment minimiser les effets de la fluctuation d'échantillonnage ?

- la réponse c'est l'estimation par intervalle de confiance (IC).

- on choisit un taux de risque (d'erreur) : a, donc le taux de confiance qui reste, est de (1- a), puis on détermine un intervalle pour lequel il y aura une probabilité de (1-a) qu'il contient la vraie valeur cherché dans la population mère (totale).

Exemple : si on fixe un risque d'erreur : a= 5% donc il nous reste une probabilité de 95%, que notre intervalle de confiance (c'est ce que nous appelons un IC à 95%), contient la vraie valeur.

Intervalle de Confiance :

L'intervalle de confiance (IC) d'une estimation (x) (p : parce que en vérité c'est une probabilité), est symétrique autour de la vraie valeur P. pour trouver les deux limites de l'intervalle, il faut résoudre l'équation

$p(x) = p(L_i \leq x \leq L_s) = 1 - a$; L_i : Limite inférieure de l'intervalle, L_s : limite supérieure de l'intervalle.

L'approximation de P (la vraie valeur) = $p \pm z \times$ l'écart type.

Z = l'écart réduit correspondant au risque d'erreur consenti (a).

On sait que la variance = $(p \times q) \div n = (p \times q)/n$.

L'écart type = $\sqrt{\text{variance}} = \sqrt{(p \times q)/n}$; la valeur de z est donnée directement par la table 2

Sur la table on a :

- Au risque consenti (d'erreur) a = 5% (donc l'intervalle est à 95%), Z = 1,96
- Au risque consenti (d'erreur) a = 1%(donc l'intervalle est à 99%), Z = 2,58
- Au risque consenti (d'erreur) a = 1‰ (1 pour mille) (donc l'intervalle est à 99,99%), Z = 3,29...etc.

Remarque :

La définition du Hasard en statistiques : c'est le fait de notre ignorance ou le reflet de notre incapacité à maîtriser tous les degrés de liberté d'une situation.

Aspects Mathématiques

-La probabilité (a) mesure le risque d'erreur de l'intervalle, c'est-à-dire la probabilité que l'intervalle ne contienne pas la vraie valeur du paramètre à estimer (x).

-la probabilité (1-a) donnée que l'on appelle niveau de confiance de l'intervalle.

1. Intervalle de Confiance à 95% (= 1-a) d'une proportion : est déterminé par la relation :

- Borne inférieure : $p - 1,96\sqrt{p(1-p)/n}$; p : la proportion ; n= effectif de l'échantillon.
- Borne supérieure : $p + 1,96\sqrt{p(1-p)/n}$; p : la proportion ; n= effectif de l'échantillon.

2. Intervalle de Confiance d'une densité d'incidence

L'intervalle de confiance d'un taux d'incidence cumulée se calcule de la même manière que celui d'une proportion puisque une incidence cumulée est une fréquence relative.

Les bornes de l'intervalle de confiance à 95% d'une densité d'incidence (DI) peuvent être déterminées par les relations :

- Borne inférieure : $DI - 1,96 \sqrt{(DI/PT)}$
- Borne supérieure : $DI + 1,96 \sqrt{(DI/PT)}$ PT : nombre de personnes-temps à risque.

3. Intervalle de Confiance d'une moyenne :

*Sur un grand échantillon ($N \geq 30$), les deux bornes de l'intervalle de confiance à 95% peuvent être déterminées par la relation :

- Borne inférieure : $m - 1,96 (\delta / \sqrt{N})$
- Borne supérieure : $m + 1,96 (\delta / \sqrt{N})$; δ : l'écart type de la variable quantitative estimé sur l'échantillon.

*Sur un petit échantillon ($N < 30$)

- Borne inférieure : $m - t (\delta / \sqrt{N})$
- Borne supérieure : $m + t (\delta / \sqrt{N})$

t est le nombre de la population caractérisée par une loi de Student de (n-1) degrés de liberté au risque d'erreur choisi, celui-ci étant généralement de 5% (intervalle de confiance à 95%). donc t se détermine à partir de la table (table de t).

Exemple :

On a mesuré chez un échantillon de 17 sujets (donc c'est un petit échantillon) le taux d'urée sanguine, la moyenne et l'écart type de 17 mesures sont respectivement : $m=300\text{mg/l}$ et l'écart type : $\delta = 10\text{mg/l}$ selon la table de loi de student et au risque 5%, la valeur t correspondant au degrés de liberté $dl = (n-1) = 17-1 = 16$ on a sur la table $t = 2,12$ donc les deux bornes de l'intervalle de confiance assigné à m (la moyenne) sont : $300 \pm 2,12 (10/\sqrt{17})$.

Donc $294,9 \leq M$ (la vraie moyenne dans la population mère) $\leq 305,1 \text{ mg}$

- Si on a fixé le risque à 1 pour mille, la valeur de $t = 4,015 \approx 4,02$ donc les deux bornes au risque 1‰ sont : $300 \pm 4,02 (10/\sqrt{17})$ donc : $290,3\text{mg/l} \leq M$ (la vraie moyenne) $\leq 309,7 \text{ mg/l}$

Le Choix de P-Valeur :

Une fois l'hypothèse H_0 et posée et le risque (α : dans certains livres $\alpha = \alpha$) fixé, les données peuvent être collectées. Cette étape terminée, il ne reste plus qu'à retenir ou rejeter H_0 . Ce choix se fait au moyen du calcul d'un paramètre désigné par la lettre « p » (on dit « petit p » ou plus simplement « p » ; les anglo-saxons le dénomment quant à eux « p-value ». si la façon de calculer « p » dépend du type de test envisagé dans l'expérience (s'agit-il de comparer deux pourcentage ; ou plutôt de la comparaison de deux moyennes, ou de deux risques relatifs ?...), le sens de ce paramètre est lui, invariable.

Le paramètre « p » est égal à la probabilité d'observer des résultats au moins aussi en désaccord avec l'hypothèse nulle (H_0) que ceux qui ont été obtenus.

La p-valeur suppose que H_0 est correcte. P-valeur nous dit à quel point il serait rare d'observer une différence aussi grande ou plus grande que celle qui a été observée si H_0 était vraie.

Le scientifique doit répondre à la question : le résultat est-il tellement peu probable que H_0 peut être-écartée ?

Donc le scientifique calcule le p-valeur pour répondre à la question : est ce que notre résultat trouvé dans notre échantillon, est vraiment très proche de la vraie valeur dans la population mère, ou est ce que notre résultat n'est seulement autre qu'une coïncidence (due au Hasard).

Donc p-valeur c'est la probabilité des expériences qui n'ont pas été réalisées, et qui peut être ne vont jamais être réalisées.

Le Seuil de Signification

Souvent on utilise la p-valeur pour rendre compte de la signification statistique.

On peut calculer la p-valeur par : le test t, test de randomisation, test de Mann-Whitney, test khi-2...etc.

La probabilité P (P-valeur) que le hasard puisse expliquer les résultats est donnée par la table Khi-2 (dans certains livres Khi-Deux = χ^2) à la ligne identifiée par le dl.

- Au seuil de signification (α), le résultat est significatif, et les deux variables sont liées (exemple une pathologie et un facteur de risque), Si $p : (p\text{-valeur}) \leq (\alpha)$.

- Au seuil de signification (α), le résultat n'est pas significatif, et les deux variables ne sont pas liées (exemple une pathologie et un facteur de risque), Si $p : (p\text{-valeur}) \geq (\alpha)$.

Exemple :

On a fixé le taux d'erreur $\alpha=5\%$, donc il nous reste un taux de confiance de 95% (IC à 95%).

Si après le calcul de p-valeur avec le test Khi-2, on a trouvé que $(p\text{-valeur}) \leq 5\%$ donc le résultat est significatif (parce que l'effet du hasard est $\leq 5\%$), et les deux variables sont liées, et on va rejeter H_0

Si après le calcul de p-valeur avec le test Khi-2, on a trouvé que $(p\text{-valeur}) \geq 5\%$ donc le résultat n'est pas significatif (parce que l'effet du hasard est $\geq 5\%$), et les deux variables ne sont pas liées, c'est-à-dire il n'y a aucune relation entre les deux variables, et on va garder H_0 (parce que c'est une hypothèse d'absence de relation).

Signification statistique et Signification Biologique ou Clinique

Un test statistique peut déceler une différence significative mais qu'en réalité peu importante du point de vue biologique ou clinique lorsque la comparaison est effectuée sur des populations de taille importante.

Par ailleurs, une différence importante sur le plan biologique ou clinique, mais non significative statistiquement, peut être mise en évidence à partir d'échantillon de faible effectif.

Donc la p-valeur dépend aussi de la taille de l'échantillon. C'est pour cela, que la p-valeur ne doit pas être considérée comme une panacée, en mesure de valider rigoureusement et rapidement la conclusion, relative au rôle d'un facteur donné, mais plutôt comme une incitation à poursuivre la réflexion sur ce rôle.

- Plus l'échantillon est grand, plus l'intervalle de confiance (IC) est étroit, et l'estimation (association ou non) est plus précise.
- Un IC large \Rightarrow un défaut de précision à cause de l'étroitesse de l'échantillon (un petit échantillon).
- Un IC étroit traduit une estimation solide, tandis qu'un IC large peut suggérer une taille échantillonnale insuffisante qui pourrait engendrer une association significative si elle était corrigée.

- L'IC reflète le lien entre signification statistique et signification clinique ou biologique.

La signification statistique implique que la valeur de référence ne soit pas incluse dans l'IC, mais la signification clinique ou biologique (de portée plus pratique), est appréciée par l'éloignement de l'IC vis-à-vis de cette valeur de référence.

Chapitre III : Notion de Biais et de Causalité

I. Notion de Biais

Un biais, ou « erreur systématique » est une cause d'erreurs lors d'une analyse statistique liée à la méthodologie d'une étude. S'ils ne sont pas pris en compte, ou maîtrisés, les biais peuvent entraîner des erreurs dans l'estimation des paramètres (OR, RR...etc.).

Ces erreurs remettent en cause la validité même d'une étude.

Il doit être distingué des fluctuations aléatoires d'échantillonnage qui représentent seulement un défaut de précision de l'estimation.

Il existe trois types de Biais :

1. Biais de sélection :

Les biais de sélection affectent la constitution de l'échantillon d'enquête, c'est-à-dire le processus par lequel les sujets sont choisis au sein de la population. Ils sont à craindre chaque fois que l'échantillon d'enquête n'est qu'une sélection de la population d'étude. Les principales sources de biais de sélection sont :

- La constitution d'un échantillon par un autre moyen que le tirage au sort (sujets volontaires) ;
- Les non-réponses à une enquête ;
- le recrutement de témoins en milieu hospitalier qui ne sont pas représentatifs de la population générale (biais de Berkson) ;
- les sujets perdus de vue dans les enquêtes exposé/non-exposé.

2. Biais de Classement :

Désigne une erreur systématique de mesure de l'exposition ou de la maladie. Ils conduisent à mal classer les sujets en « malades/non malades » ou en « exposés/non exposés ». il peut s'agir d'un biais de classement différentiel entre les deux groupes :

- D'un biais d'enquêteur ou de subjectivité : un enquêteur cherchera avec plus d'acuité une exposition passée chez un cas, s'il a connaissance du statut malade du sujet ;
- D'un biais de mémorisation : un sujet atteint d'une pathologie (cas) se souvient davantage de ces expositions passées qu'un sujet indemne de la pathologie (« témoin »).

Il peut aussi s'agir d'un biais de classement non différentiel (tensiomètre mal étalonné par exemple).

3. Biais de Confusion :

Désigne une erreur systématique dans l'estimation d'une mesure d'association (Odds ratio ou Risque relatif) entre le facteur étudié et la maladie, du fait d'un défaut de prise en compte d'un facteur de confusion. Un facteur de Confusion est un facteur lié à la fois à l'exposition et à la maladie étudiée.

II. Notion de Causalité

La question de la causalité d'une relation entre un facteur et une pathologie intervient après l'objectivation d'une association statistiquement significative. Il s'agit de savoir si l'association mise en évidence correspond à une relation de cause à effet. Une enquête d'observation ne permet pas de conclure à une relation causale (contrairement à un essai randomisé bien conduit). Il faut donc examiner les arguments qui plaident en faveur de la causalité de la relation :

1. Critères internes à l'étude :

- Séquence temporelle : l'exposition doit précéder l'apparition de la pathologie ;
- Force de l'association : RR ou OR élevé ;
- Relation de type « dose-effet » entre l'exposition et la fréquence de la pathologie ;
- Spécificité de la cause et de l'effet : la relation est spécifique si le facteur de risque est présent chez presque tous les malades et seulement chez eux, et ce pour cette seule maladie ;
- Cohérence interne de l'étude : prise en compte la plus complète possible des facteurs de confusion et minimisation des biais de sélection ou de classification.

2. Critères externes à l'étude :

- Constance de l'association et reproductibilité dans diverses situations (différentes périodes, régions ou population) ;
- Cohérence des résultats avec les résultats d'études publiées ;
- Plausibilité biologique ; modèles expérimentaux chez l'animal ;
- Cohérence avec les connaissances générales et les hypothèses qui ont conduit à la réalisation de l'enquête ;

Parallélisme de la distribution (dans l'espace et dans le temps) du facteur de risque et de la pathologie.