

# Mesures de l'occurrence du phénomène de santé

Brice Leclère

Ce document a pour but de résumer les principales informations à retenir pour ce cours. Les exemples et discussions supplémentaires donnés en cours n'y apparaissent donc pas. Pour une meilleure compréhension des concepts vus dans ce cours, il est conseillé de lire en regard vos notes de cours.

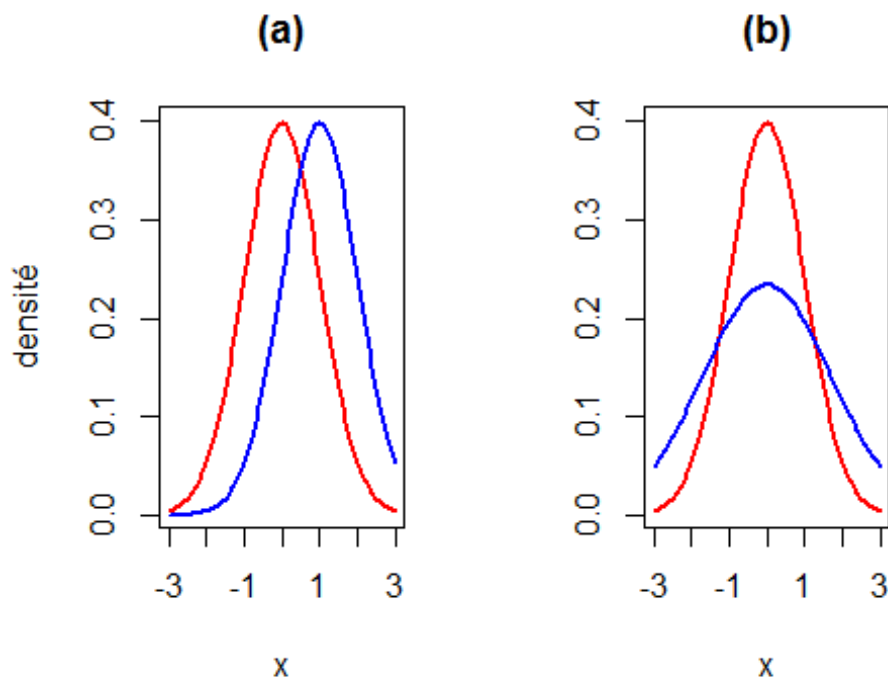
## Mesures continues et comptage

Les mesures continues sont souvent utilisées dans le domaine de la santé, que ce soit pour étudier les maladies ou les expositions (indice de masse corporelle, pression artérielle, cholestérolémie, mais aussi âge, taille...). Ces mesures sont dites « continues » car le nombre de mesures différentes est extrêmement important — potentiellement infini — si on augmente la précision de la mesure. Les comptages (par exemple nombre de cas d'une maladie donnée, nombre de partenaires sexuels) sont un type particulier de variable mais peuvent être décrits comme des variables continues lorsque le nombre de valeurs possibles est potentiellement grand.

Pour catégoriser ces mesures, deux paramètres sont nécessaires : un paramètre de position et un paramètre de dispersion, le premier reflétant une tendance centrale, le second une mesure de l'hétérogénéité dans la population ou l'échantillon.

## Paramètres classiques : moyenne et variance

Ces sont les paramètres les plus fréquemment utilisés pour catégoriser une variable continue. Les formules permettant de les calculer sont connues et ne seront pas rappelées dans ce document. Je vous conseille de vous retourner vers les cours de l'UE de biostatistique ou vers un ouvrage introductif en statistiques en cas de besoin (par exemple *Statistique Épidémiologie* de T. Ancelle, disponible à la BU).



Bien qu'incontournables, ces paramètres ont toutefois leurs limites, notamment lorsque la distribution de la variable étudiée dans la population d'étude est très asymétrique, c'est-à-dire présentant un certain nombre de valeurs très extrêmes d'un côté de la moyenne. C'est le cas notamment des délais de séjours hospitaliers en MCO (médecine-chirurgie-obstétrique-odontologie) la plupart des séjours durent quelques jours, mais certains peuvent être très longs, de plusieurs semaines ou plusieurs mois. Dans ce cas, la durée moyenne de séjour va être « déplacée » vers les valeurs extrêmes et reflètera de façon moins pertinente la tendance centrale.

### Les indices non paramétriques : médiane, quartiles, percentiles...

Lorsque la distribution de la variable continue étudiée est très asymétrique (on dit aussi « biaisée vers la droite/la gauche »), il est préférable d'utiliser comme indice de la tendance centrale la médiane. Cette dernière correspond à la valeur au centre de l'étendue des valeurs dans la population étudiée ; ainsi, 50 % des valeurs restantes sont au dessus, et 50 % en dessous.

#### Exemple

Voici la durée d'hospitalisation dans un service de court séjour d'une file active de 12 patients.

Patient n°	durée d'hospitalisation
1	1
2	3
3	3
4	4
5	4

7	5
8	5
9	14
10	15
11	17
12	21

**Durée moyenne d'hospitalisation = 8,4 jours**

**Durée médiane d'hospitalisation = 5 jours**

La médiane semble bien ici une meilleure représentation de ce qu'il se passe réellement dans le service ?

Ce principe peut être étendu au-delà du seuil de 50 % : on parle alors de quartiles pour les quatre quarts de l'étendue (25-50-75-100 %) ou de percentiles pour chaque centième. La médiane correspond donc au second quartile et au cinquantième percentile. Ces indices supplémentaires sont utiles pour refléter la dispersion des valeurs dans une distribution asymétrique, et on utilise fréquemment dans ce cas l'espace inter-quartile (espace compris entre le premier et le troisième quartile).

## Une tendance centrale rarement utilisée : le mode

Le mode représente la valeur la plus fréquente dans la population étudiée. Bien qu'intéressante, cette statistique est en fait rarement utilisée en pratique. Dans le cas d'une distribution symétrique, mode, médiane et moyenne sont confondus.

## Comptage d'observations

Il ne s'agit plus ici de faire un comptage *variable par observation* (par exemple le nombre de journées d'hospitalisation pour chaque patient) mais un comptage *observation par variable* (par exemple le nombre de sujets ayant présenté la maladie étudiée). Il existe alors plusieurs façon d'analyser ces variables.

### Comptage brut

La méthode la plus simple pour rendre compte de ce type de variable est de présenter le chiffre brut. Par exemple, on estime à 148 000 le nombre de décès par cancer survenus en France en 2012. Ce chiffre brut est parlant quant au fardeau de la maladie ou de l'exposition étudié, en revanche, il est difficilement comparable : 148 000 décès ne représente pas le même poids dans la population française ou dans la population luxembourgeoise, par exemple.

### Pourcentage, prévalence et incidence

On rapporte donc en général ce nombre brut à un dénominateur afin de l'exprimer par un rapport (sous forme de pourcentage ou « pour mille », la plupart du temps). La statistique calculée dépend de la nature du numérateur et du dénominateur utilisé. Les deux statistiques principales utilisées en épidémiologie sont :

1. **la prévalence**, qui représente le **poids** du phénomène dans la population ; c'est le nombre total de cas rapporté à la population totale à un instant donné (prévalence *ponctuelle*) ou sur une période donnée
2. **l'incidence** représente quant à elle le **risque** de présenter le phénomène étudié (par exemple de développer une maladie) ; c'est le nombre de **nouveaux cas** rapporté à la **population à risque** sur une période donnée (ensemble des personnes saines au début de l'étude).

## La baignoire de l'épidémiologiste

Une analogie fréquemment utilisée est la suivante : si notre population est représentée par une baignoire, l'incidence est l'eau s'écoulant du robinet et la prévalence est le niveau d'eau dans la baignoire. L'eau peut aussi s'écouler par la bonde, représentant les décès ou les guérisons. Par cette analogie, on comprend que la prévalence de la maladie dépend directement de l'incidence et de la durée de la maladie. Ainsi, pour une maladie rare et courte, la prévalence est proche de l'incidence, alors que pour une maladie chronique, la prévalence sera toujours supérieure à l'incidence.

## les taux ou densité

La notion d'incidence implique un suivi de la population dans le temps afin de mesurer les nouveaux cas. l'incidence dans son acception classique présentée ci-dessus – aussi appelée proportion d'incidence, incidence cumulée ou risque — ne doit être calculé que si l'ensemble de la population ou de l'échantillon étudié a pu être suivi, en d'autres termes s'il n'y a pas de perdus de vue — ou d'attrition. Dans le cas contraire, il est impossible de savoir si les perdus de vue ont développé ou non la maladie et l'incidence cumulée est donc faussée. La solution la plus fréquemment utilisée dans ce cas est d'utiliser un taux : chaque individu est alors pris en compte uniquement sur la période durant laquelle il a contribué à l'étude, c'est-à-dire le temps écoulé entre l'inclusion et soit 1) le développement de la maladie, soit 2) la sortie de l'étude. Cette durée est exprimée en personne-temps : une personne ayant été suivie pendant un an a le même poids que deux personnes suivies pendant six mois (soit une personne-année). Le **taux d'incidence** (ou **densité d'incidence**) utilise la somme des personnes-temps comme dénominateur, le numérateur restant le nombre de nouveaux cas survenus durant la période. Le taux d'incidence est à envisager comme une **vitesse de propagation** du phénomène étudiée dans la population. Quand il n'y a pas de perdus de vue, taux et proportion d'incidence se rejoignent.

## Un premier pas vers la comparaison de mesure : les risques conditionnels et cotes

Quand il s'agit de comparer les mesures entre plusieurs groupes, comme c'est le cas notamment dans les études étiologiques ou les essais randomisés, les proportions (prévalence, risque, taux...) seront calculés indépendamment dans ces différents groupes. C'est ce qu'on appelle les risques conditionnels, qui sont utilisés pour le calcul du risque relatif (ou du rapport de taux), vu dans le cours suivant. Il est utile ici d'introduire une nouvelle façon de calculer une proportion : **la cote**. Ici, le comptage n'est pas rapporté à l'ensemble de la population ou à la population à risque mais à la

population restante. Ainsi, la cote d'une maladie est le nombre de personnes présentant cette maladie divisé par le nombre de personnes ne l'ayant pas présentée. Si la prévalence et le risque variait entre 0 et 1, la cote varie entre 0 et l'infini.