

Glossaire Statistique et épidémiologie

Population

- L'ensemble des individus ou participants potentiels auxquels on s'intéresse dans une étude.

Échantillon

- Sous-ensemble de la population qui est sélectionné pour l'étude, de façon aléatoire, systématique ou volontaire.

Variable

- Caractéristique à laquelle on peut attribuer une valeur numérique ou bien une valeur provenant d'un ensemble restreint.
- Il est possible d'observer, de mesurer ou d'attribuer une valeur de cette variable pour chaque individu dans la population étudiée.

Variable dépendante

- Aussi appelée variable réponse, expliquée, endogène, « outcome », VD, Y.
- Il s'agit de la variable que l'on tente d'expliquer, de prédire à partir de variables indépendantes. Elle est influencée par ces dernières.
- Dans le contexte d'une expérience contrôlée, c'est la variable que l'on mesure chez les participants afin de voir l'impact des variables de l'expérience sur celle-ci.

Variable indépendante

- Aussi appelée variable explicative, expérimentale, prédictive, contrôlée, facteur de risque, VI, X. Il s'agit souvent d'un groupe ou traitement.
- Il s'agit d'une variable que l'on utilise pour expliquer, prédire une variable réponse. Cette variable influence la variable réponse.
- Dans le contexte d'une expérience contrôlée, c'est la variable que l'on manipulera pour voir son effet sur la variable dépendante.

Estimateur ponctuel

- Estimation d'une statistique (par exemple : moyenne, coefficient de corrélation) basée sur les données qui permet de résumer une caractéristique de l'échantillon en une seule valeur.
- C'est un estimateur de la valeur véritable d'un paramètre dans la population.

Intervalle de confiance (IC)

- Intervalle de valeurs qui contient, à une probabilité dépendant de son niveau de confiance, la véritable valeur de la population.
- Permet d'avoir une mesure de la précision associée à l'estimation ponctuelle. Tient compte de la variabilité présente dans l'échantillon.
- L'intervalle de confiance est associé à un niveau de confiance (exemple 95%) qui est l'inverse de l'erreur de Type I (exemple 5%).
- Plus le niveau de confiance est grand, plus l'intervalle sera large. Plus un échantillon est grand, plus l'intervalle sera étroit.
- Un IC à 95% pour une moyenne indique que le niveau de confiance que la moyenne de la population soit dans cet intervalle est de 95%. En autres termes, si on construisait 100 IC à partir de 100 échantillons de la même population, 95 d'entre eux incluraient la véritable moyenne de la population.

Test statistique

- Test permettant de comparer l'hypothèse nulle à l'hypothèse alternative.
- On suppose a priori que l'hypothèse nulle est vraie. Avec suffisamment d'évidence, on peut rejeter l'hypothèse nulle. Par contre, on n'« accepte » jamais l'hypothèse alternative.
- Fournit une statistique ainsi qu'une valeur P afin de déterminer quelle hypothèse est favorisée par les données.

Hypothèse nulle

- Aussi appelée H_0 .
- Habituellement, cette hypothèse correspond à une situation d'égalité.
- Par exemple, lorsque l'on désire vérifier si deux moyennes (\bar{x}) sont égales, $H_0: \bar{x}_1 = \bar{x}_2$, ou bien pour vérifier si une corrélation (ρ) est significativement différente de zéro, $H_0: \rho = 0$.

Hypothèse alternative

- Aussi appelée H_1 ou contre-hypothèse.
- Habituellement, cette hypothèse est une inégalité.
- Hypothèse bilatérale, si on s'intéresse à une différence, peu importe sa direction (ex. : $H_1: \bar{x}_1 \neq \bar{x}_2$ ou $H_1: \rho \neq 0$),
- Hypothèse unilatérale, pour vérifier si une valeur est plus petite, ou bien plus grande qu'une autre (ex. : $H_1: \bar{x}_1 < \bar{x}_2$ ou $H_1: \rho > 0$).
- Si on rejette H_0 , on peut dire par exemple, qu'il y a une différence entre les deux moyennes. Si on ne rejette pas H_0 , alors on n'a pas assez d'évidence pour démontrer l'hypothèse alternative.

Probabilité

- Nombre d'occurrences d'un événement divisé par le nombre de tentatives.
- Valeur entre 0 et 1

Erreur de type I

- Aussi appelée l'erreur alpha, souvent fixée à 5%.
- Cela correspond à rejeter H_0 , alors qu' H_0 est vraie, c'est-à-dire détecter un effet significatif alors qu'il n'y en a pas en réalité. On tente donc de minimiser cette valeur.

Valeur P

- Probabilité d'obtenir le résultat observé strictement à cause du hasard.
- Probabilité de commettre une erreur de Type I.
- Probabilité d'observer un résultat plus extrême que celui que l'on a observé, sous l'hypothèse nulle.
- Si la valeur P est très petite (près de 0), alors il y a de fortes chances que cela ne soit pas dû au hasard et qu'il y ait bel et bien un effet. On tend donc vers H_1 . Si cette valeur est grande, alors cette réalisation est très probable sous H_0 , et donc on ne rejette pas l'hypothèse nulle.
- Lorsque l'on calcule une valeur p dans un test statistique, on compare celle-ci au seuil alpha. Par exemple, lorsque la valeur p est inférieure à 5%, on dira qu'il y a un effet significatif au seuil de 5%.

Pour plus d'informations sur ces concepts :

Rosner, Bernard (2010). Fundamentals of Biostatistics, 7th edition. Cengage Learning. 888 pages.

Types d'études

Essai clinique randomisé

- Type d'étude contrôlée où l'on désire comparer des traitements sur différents groupes.
- Les participants sélectionnés sont répartis aléatoirement par randomisation dans un groupe. Cela permet d'obtenir des groupes ayant des caractéristiques homogènes au départ.
- Il y a aveuglement lorsque les patients ne savent pas quel traitement ils reçoivent ou que le personnel de recherche ne sait pas quel traitement le patient a reçu.

Étude cas-témoins

- Permet de comparer deux groupes, les cas et les témoins, où les patients sont choisis selon la présence ou l'absence de maladie, par exemple.
- Par la suite, on identifie les caractéristiques des patients reliés à leur exposition et aux facteurs de risque.
- Un appariement peut être effectué afin d'avoir des témoins ayant des caractéristiques de base similaires aux cas.
- Souvent utilisée lorsque la maladie étudiée est rare.

Étude de cohorte

- Type d'étude où l'on suit un groupe de patients sur une certaine période afin de voir l'évolution d'un phénomène ou l'apparition de certaines caractéristiques.
- On connaît leur exposition aux facteurs de risque dès leur inclusion dans l'étude.
- Permet de mesurer l'incidence d'un problème de santé sur un échantillon de patients ayant été exposé à un même facteur.

Étude transversale

- Étude où l'on regarde la prévalence d'un événement, d'une caractéristique ou d'une maladie à un moment précis dans le temps.

Étude prospective

- L'exposition des patients à différents facteurs est mesurée au début de l'étude et on voit par la suite le développement du phénomène d'intérêt, comme une maladie.

Étude rétrospective

- L'exposition aux facteurs de risque, ainsi que le développement du phénomène d'intérêt, ont lieu avant le début de l'étude.