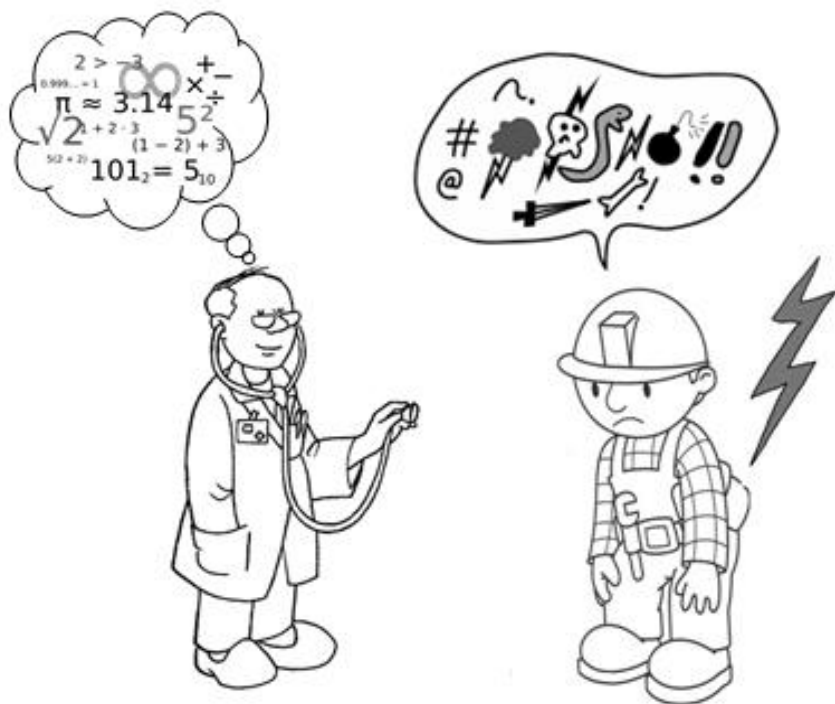


PETIT LIVRET DE STATISTIQUE

SANS DOULEUR



Journées de l'Afmck Montpellier janv 2016.

Sylvain Peterlongo

Introduction

Tu es un des 150 kinésithérapeutes de ce congrès (c'est un effectif). Tu fais donc aussi partie des 3 500 kinés français sur 64 000 à avoir suivi un module A McKenzie (c'est un rapport) soit 1 kiné sur 18 (ratio), autrement dit 5,5% des kinés sont formés (c'est une proportion).

Tu vois déjà le premier usage des statistiques, présenter et résumer une situation. C'est le rôle des **statistiques descriptives**.

Comme on n'est jamais vraiment sûr de rien, tu trouveras après les principales sources d'erreurs de jugement possible. D'abord les erreurs systématiques **-les biais-** puis les erreurs aléatoires et les façons de s'en prémunir grâce aux **statistiques inférentielles**.

Métaphores sportives à l'appui tu devrais refermer ce livret avec une idée plus précise de ces notions nébuleuses que sont les variances, facteurs de risques, petit p et autres biais statistiques. Tu connaîtras aussi l'intérêt des statistiques dans un braquage de banque.

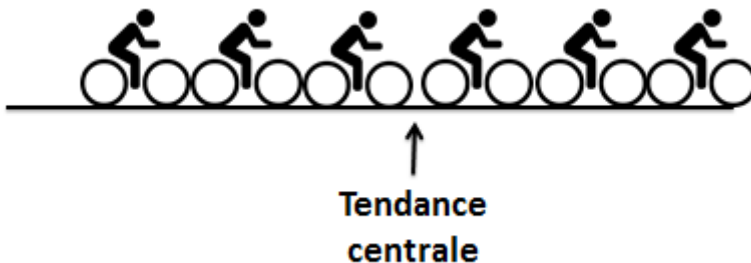
Enfin, une petite séance MDT de Bob le bricoleur te permettra de voir où sont cachées les statistiques dans ta pratique.

Les statistiques descriptives

Les statistiques descriptives permettent de décrire un groupe de personnes par sa position et sa dispersion.

Mesure de tendance centrale

Une mesure centrale résume la localisation des valeurs. Ces valeurs nous donnent une idée d'où se situe « le milieu » du peloton.

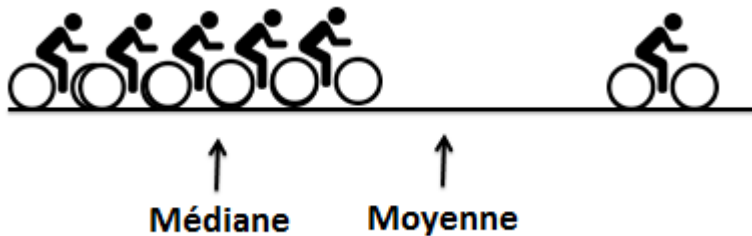


La **moyenne** est la somme des données divisée par le nombre d'individus.

La **médiane** est la valeur qui divise le groupe en deux parties égales.

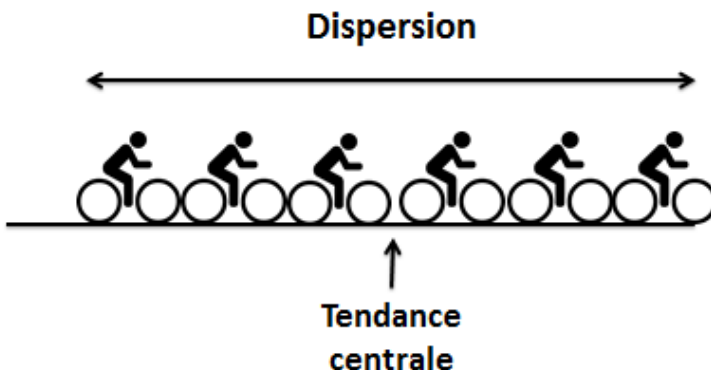
Le **mode** est la valeur qui revient le plus souvent.

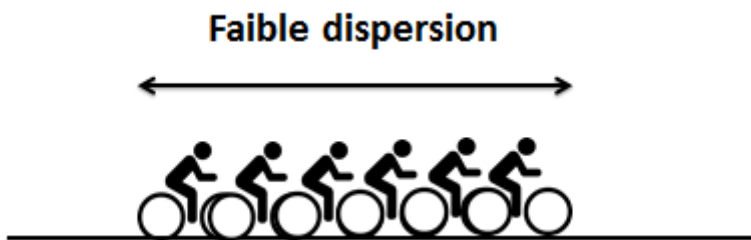
La moyenne est la statistique la plus pratique et la plus intuitive. Son plus grand défaut est d'être très sensible aux valeurs extrêmes. Si un coureur s'échappe loin devant les autres, la moyenne nous donnera ici une image trompeuse du groupe puisqu'un seul coureur dépasse la moyenne. Dans ce cas, la médiane donne une image plus fidèle du groupe.



Mesure de Dispersion

Une mesure de dispersion résume la variabilité des valeurs. Ces valeurs nous donnent une idée de « l'étalement » du peloton.

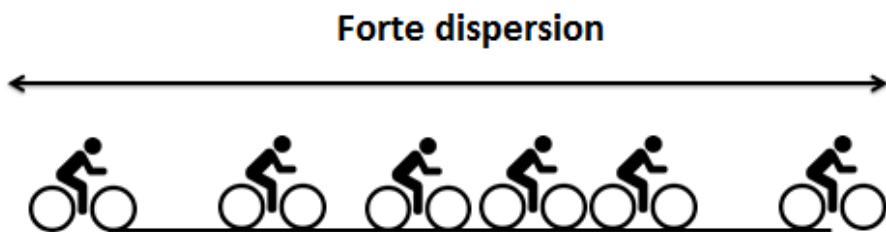




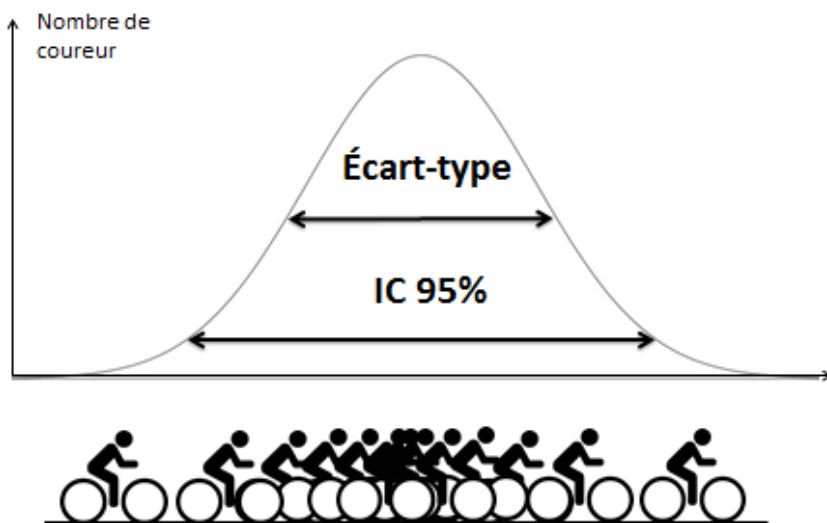
L'**étendue** est l'écart entre les deux valeurs extrêmes. Comme la moyenne, l'étendue est directement influencée par les valeurs extrêmes et présente donc les mêmes inconvénients.

On calcule donc généralement la **variance** et l'**écart-type** (standard-deviation).

Le calcul est moins intuitif mais présente l'avantage de prendre en compte l'ensemble des données et pas seulement les extrêmes. La **variance** est la moyenne des carrés des écarts à la moyenne. Utile pour les calculs, la variance n'est pas très parlante vue qu'elle ne s'exprime pas dans la même unité que les données décrites. C'est l'avantage de l'**écart-type** (la racine carrée de la variance). C'est le paramètre de dispersion le plus utilisé. Les **intervalles de confiances** sont ensuite calculés à partir de l'écart-type.



Dans la plupart des cas, la répartition autour d'une moyenne suit une **loi normale** ou courbe de Gauss. Dans ce cas, l'écart type regroupe 68% de la population autour de la moyenne et **l'intervalle de confiance à 95% = 1,96 fois l'écart-type**. Si la répartition suit une autre loi (la courbe a une forme différente) les calculs seront différents.



Pour illustrer, faisons connaissance avec Colette, **81 ans**, Paulette **78 ans**, Georgette **75 ans**, Ginette **79 ans**, Juliette **82 ans**, Odette **78 ans**, Yvette **80 ans**.



Pour le moment ce ne sont que des données pas des statistiques.

La **moyenne** d'âge est de 79 ans, c'est une statistique. La **médiane** est de 79 ans et le **mode** 78 ans.

7 ans séparent la plus jeune de la plus âgée, c'est l'**étendue**.

Pour calculer la variance, on calcule pour chaque personne la différence de son âge avec la moyenne. Colette a 2 ans de plus que la moyenne, Paulette 1 an de moins que la moyenne, etc etc

Ces valeurs sont élevées au carré. La moyenne de ces valeurs au carré donne la **variance** 4,6 ans². Cette valeur en

année-carré n'est pas très parlante, mais elle est pratique pour les calculs. Pour avoir une valeur qui se compte en année, on utilise la racine carré de la variance : **l'écart-type 2,1 ans**.

Avec une moyenne d'âge de 79 ans et un écart-type de 2,1 ans, on résume donc bien la situation, un groupe de personnes âgées homogène (faible dispersion).

Autre exemple, prenons maintenant la famille de Collette **81 ans**. Son fils a **58 ans**, sa fille **54 ans**, son gendre **51 ans**, ses petits-enfants **25 et 30 ans** et son arrière-petite-fille **2 ans**.



La moyenne d'âge est de 43 ans, l'écart-type est de 19,7 ans. On voit un groupe de personnes de tout âge avec des différences d'âges importantes, et donc une forte dispersion.

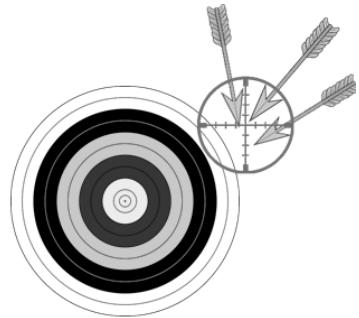
Les biais statistiques

Ou les mille et une façons de manquer sa cible

Biais de sélection

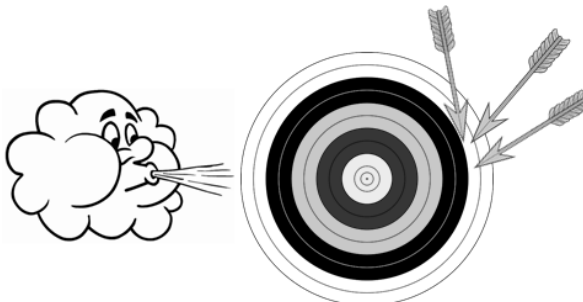
L'étude n'a pas visé la bonne population, la cible est manquée.

(Exemple, une étude de nutrition a montré les effets bénéfiques de l'alimentation fast food... en n'incluant que des patients dénutris)



Biais de confusion

La méthode de l'étude ne prend pas en compte un élément extérieur.



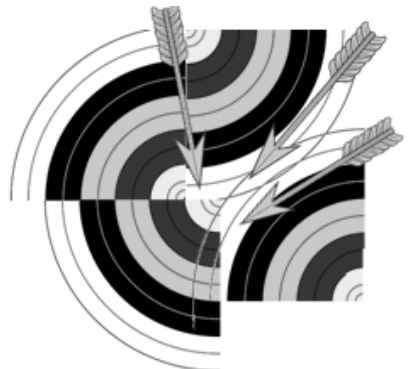
C'est le même processus lorsque l'on confond corrélation et causalité.

Par exemple il existe un lien de corrélation très significatif entre le nombre de noyades et la consommation de glaces à la crème. Une explication logique ? Se baigner après avoir mangé une glace provoque une hydrocution, un malaise et donc une noyade. Convaincu ? C'est oublier un élément extérieur déterminant : le temps qu'il fait ! Il fait beau et chaud, on va manger des glaces et se baigner. Il fait froid et il pleut, plus personne n'a envie ni de l'un ni de l'autre. Un seul élément extérieur provoque deux conséquences, qui n'ont aucune relation entre elle.

Biais d'information

L'étude utilise un mauvais recueil de données. Les évaluations ne sont pas fiables ce qui induit des erreurs de classement initial et/ou des erreurs de mesure des résultats.

Comment départager deux joueurs avec cette cible ?

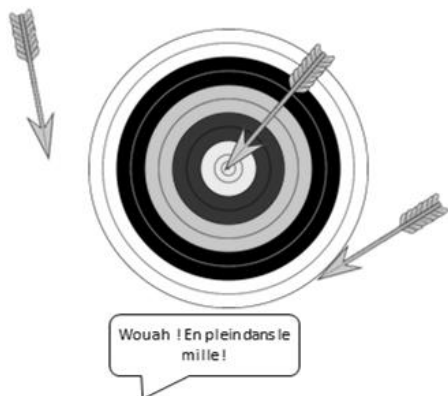


Biais d'interprétation

(Biais de confirmation et

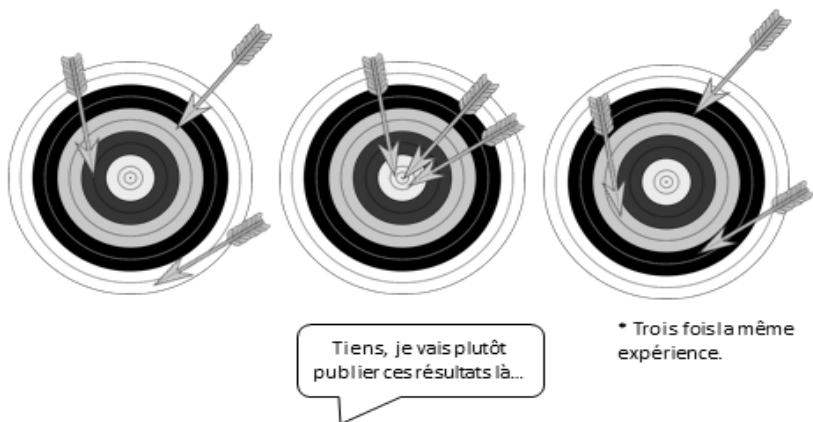
Biais de mémorisation)

On espère toujours qu'une expérience confirme ce que l'on pensait déjà. On a donc tendance à ne voir et ne retenir que la flèche dans le mille.



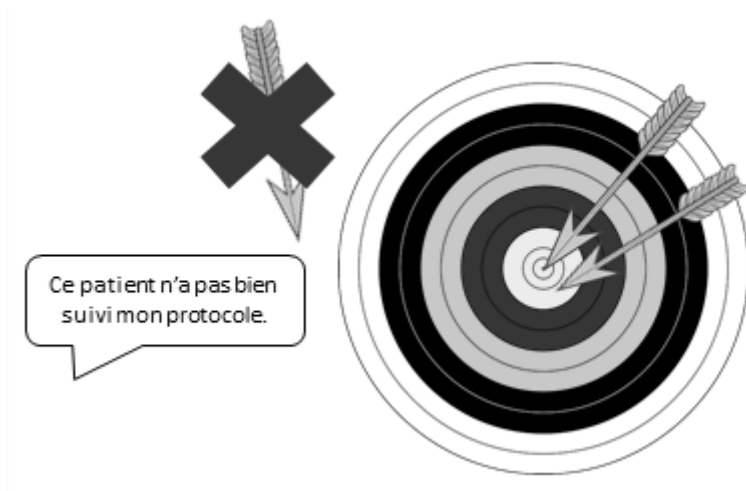
Biais de publication (Effet tiroir)

Il est toujours plus facile et satisfaisant de publier une étude aux résultats significatifs, les autres études restent dans le tiroir. Dans les revues scientifiques comme sur facebook, on parle surtout de ce qui marche bien.



Biais d'attrition

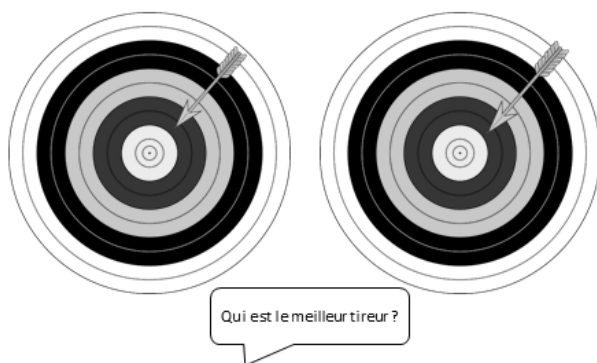
Ce biais apparaît si le chercheur n'analyse que les patients qui ont suivi intégralement le protocole. Or ceux qui ont eu des effets secondaires intolérables n'ont aucune envie de continuer et ne seraient donc pas inclus dans l'analyse. A l'inverse ceux pour qui le traitement a été très vite efficace ne termineront pas non plus leur traitement. Ce biais est maîtrisé par l'analyse dite en intention de traiter, tous les patients inclus au début de l'étude sont analysés, qu'ils aient suivi le traitement ou non. Ce type d'analyse donne une image plus fidèle des conditions réelles de soin.



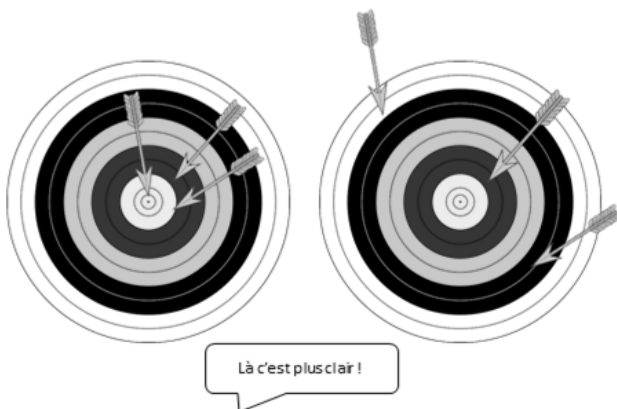
Les erreurs aléatoires

Erreur liés au nombre d'observation

Comme les valeurs fluctuent d'une personne à une autre, il nous faudra un nombre suffisant de personnes dans notre échantillon pour nous faire un avis. Sinon notre étude manque de **puissance**.



On diminue ce risque d'erreur en multipliant les observations. Plus le nombre d'observation sera grand, plus on aura une idée précis de ce qui se passe dans la population.



Estimation du risque d'erreur par p.

Mais même avec un nombre significatif de personnes on prend toujours un **risque à généraliser** nos observations à partir de notre échantillon.

De même chaque individu réagit différemment au traitement, les **résultats fluctuent d'un échantillon à un autre**. Il y a donc toujours un risque que le groupe de patients que l'on observe soit très différent de la population générale. Le meilleur tireur à l'arc du monde vient peut-être tirer devant nous la pire volée de flèche de sa vie.

Ce sont les statistiques dites inférentielles qui nous donnent le risque que l'on prend à généraliser nos résultats. Il s'agit alors de **probabilité**.

On ne cherche pas à valider une hypothèse, mais à démontrer qu'il est très peu probable qu'elle soit fausse. La conclusion d'un article scientifique avec des résultats significatifs revient donc à dire : j'affirme ça, et il est très peu probable que j'ai tort.

La démarche expérimentale consiste d'abord à imaginer que l'on a tort. On dit que l'on pose **l'hypothèse nulle** : mon hypothèse est fausse. Bien que mon hypothèse soit fausse dans la population générale, il est toujours possible que je tombe par hasard sur un échantillon qui confirme mon hypothèse. C'est

cette probabilité qui est exprimé par la **valeur p**. Le seuil $p=0,05$ signifie donc que, si mon hypothèse est fausse, j'ai quand même 5% de chance de tomber sur un échantillon qui confirme mon hypothèse.

Cela veut aussi dire qu'en acceptant ce risque, une étude sur 20 conclura quelque chose à partir de son échantillon qui n'est pas vraie dans le reste de la population.

Ce risque est plus élevé que la probabilité de faire un 12 avec une paire de dés (six et six). C'est peu probable mais loin d'être impossible !



Exemple

Je me demande si un faux monnayeur farceur ne m'a pas donné une pièce avec deux côtés piles.

Mon hypothèse est la suivante : j'ai dans la main une pièce truquée avec deux cotés pile. L'hypothèse nulle serait qu'il s'agit d'une pièce normale, avec un côté pile et un côté face. Pour tester mon hypothèse, je décide de tirer à pile ou face. Je pourrais exclure mon hypothèse si je tombe sur un côté face.

Je tire une première fois ma pièce, je tombe sur pile. Mon hypothèse n'est donc pas exclue. Mais, me direz-vous, j'avais une chance sur 2. Avec un seul lancer, j'ai une chance sur deux de me tromper ($p=0,5$). Ce n'est pas significatif, il nous faut plus d'observation. (On dit alors que l'expérience manque de puissance).



Donc je tire une deuxième fois ma pièce, encore pile. Mon hypothèse n'est toujours pas exclue mais j'avais cette fois une chance sur 4 de tirer deux piles à la suite ($p=0,25$).

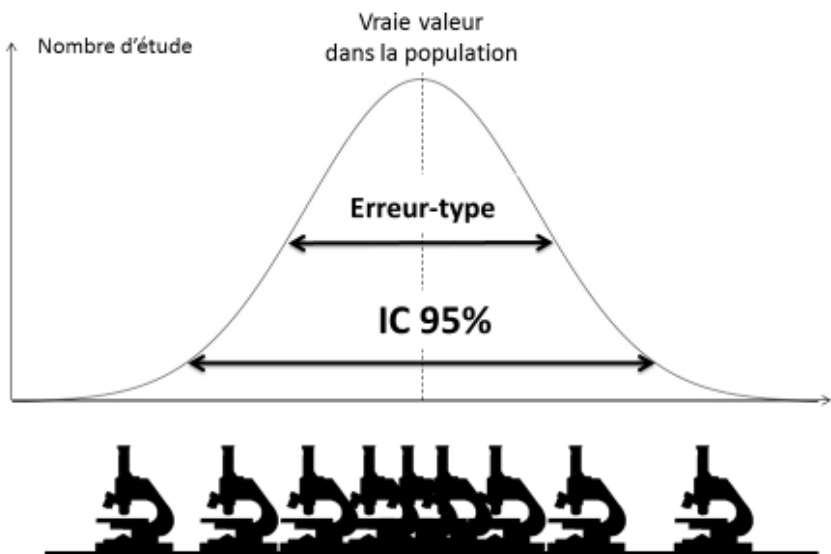
Au cinquième tirage, la probabilité de tirer 5 fois pile est d'une chance sur 32 soit $p<0,05$. C'est à peu près ce risque que l'on accepte de prendre dans les études médicales et paramédicales.

Avec six tirages j'ai une chance sur 64 d'avoir tiré 6 fois pile ($p<0,02$), avec sept tirages $1/128$ ($p<0,01$). Même avec 10 tirages successifs, il y a toujours une chance sur 1024 ($p\approx 0,001$) de n'avoir tiré que des piles. Avec la répétition le risque de se tromper diminue mais ne sera jamais nul. Même avec un nombre infini de répétitions il y a toujours une chance de tomber à chaque fois sur pile.

Estimation du risque d'erreur par Intervalle de confiance (confidence interval)

Il existe une autre façon d'estimer ce risque d'erreur. On a vu que de la même façon que les individus sont différents les uns des autres, les groupes d'individus aussi seront différents les uns des autres. Dans chaque échantillon, les valeurs individuelles sont réparties autour de la moyenne du groupe (comme les coureurs cyclistes autour du milieu du peloton).

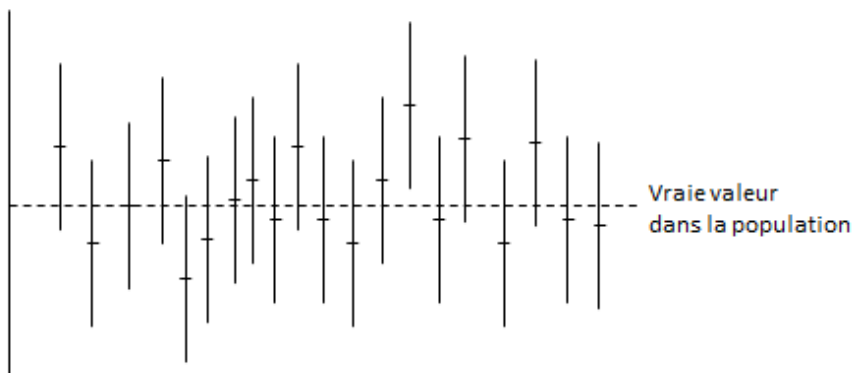
De la même façon, si on répète plusieurs fois une étude avec des échantillons différents, les moyennes de chacun de ces échantillons seront réparties autour de la vraie valeur dans la



population. Si tu as bien lu plus haut, tu sais maintenant que qui dit dispersion, dit écart-type.

C'est **l'erreur-type** (standard error) qui correspond à l'écart-type des moyennes de chaque échantillon. Pour le calculer, pas besoin de faire plusieurs fois la même étude, il est estimé à partir de la dispersion des individus dans un seul échantillon et de la taille de l'échantillon. De même qu'avec l'écart-type d'un échantillon, on peut ensuite calculer l'intervalle de confiance à 95% de notre étude. L'étendue de l'intervalle de confiance est inversement proportionnelle à la taille de l'échantillon. C'est-à-dire que plus il y a de personnes dans mon échantillon, plus mon intervalle de confiance sera resserré, et plus mon estimation sera précise.

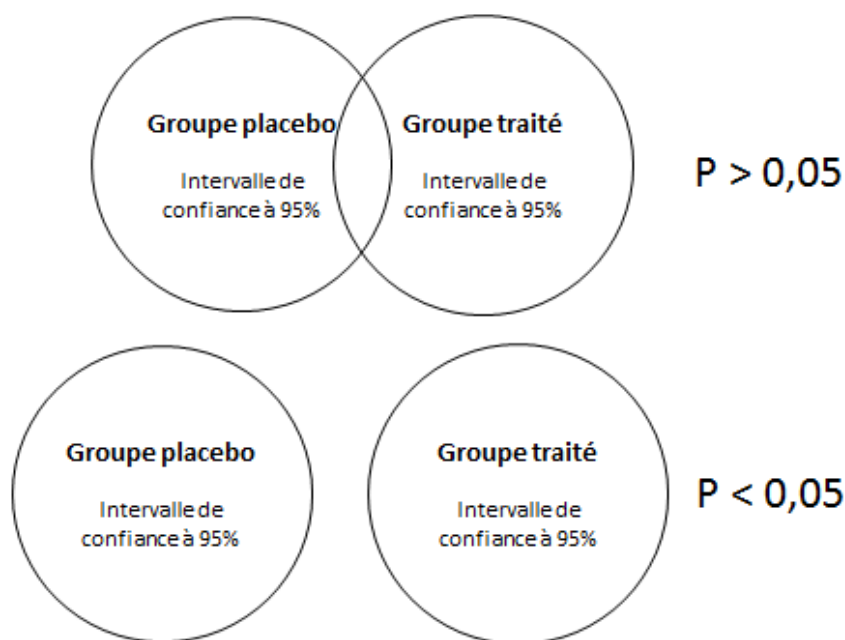
De la même façon qu'en acceptant un risque correspondant à $p=0,05$ en utilisant un intervalle de confiance, une étude sur 20 n'inclue pas dans son intervalle de confiance la vraie valeur.



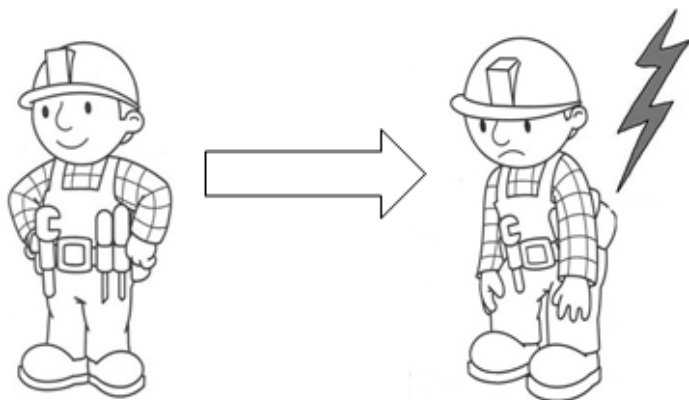
20 échantillons avec intervalle de confiance à 95%

Exemple

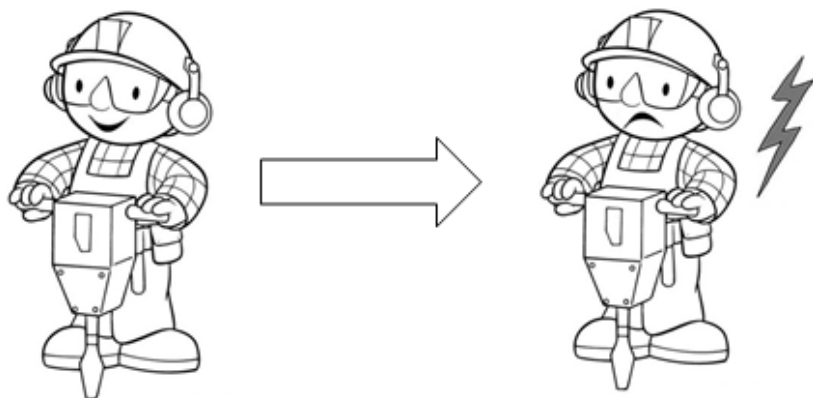
Je veux comparer deux groupes, un groupe traité et un groupe placebo. Si leur intervalle de confiance se recoupe, la différence entre les deux groupes n'est pas statistiquement significative. A l'inverse, si les intervalles de confiance ne se recoupent pas, la différence est significative.



Notion de facteur de Risque



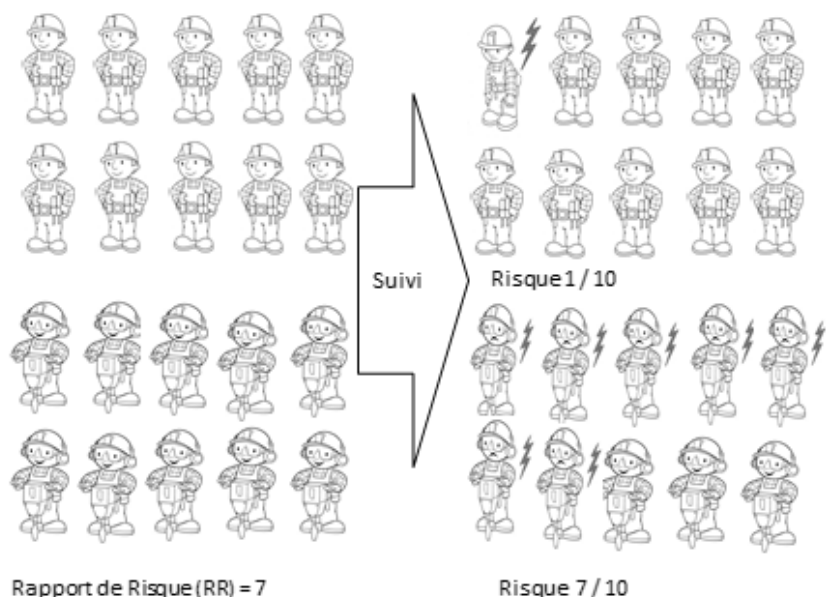
Pourquoi Bob a-t-il mal au dos ? Il travaille tous les jours un marteau-piqueur, est-ce à cause de ça ?



Pour vérifier cette hypothèse, on utilise la notion de **risque**.

Le risque est la **probabilité** qu'un événement survienne dans une période donnée.

Le plus simple pour calculer un risque est d'observer sur le long terme un groupe de personne et de voir ce qui leur arrive. C'est une étude de **cohorte**.



Prenons 20 gars comme Bob, en tout point comparable, en salopette et chemise à carreau. 10 n'utilisent pas de marteau piqueur et les 10 autres l'utilisent quotidiennement. A la fin de mon suivi, parmi ceux qui n'en utilisent pas, 1 seul a eu mal au dos. Le risque dans le groupe « sans marteau-piqueur » est de 1 sur 10.

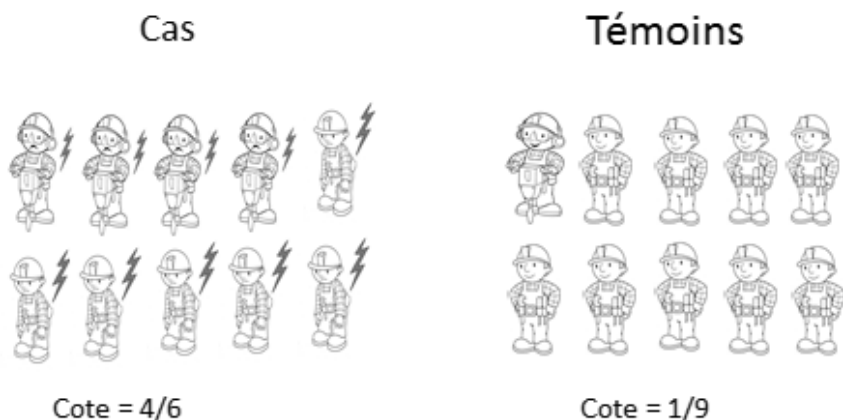
Parmi les 10 qui utilisent un marteau-piqueur, 7 ont eu mal au dos à la fin du suivi. Le risque dans le groupe « avec marteau piqueur » est de 7 sur 10.

Pour savoir si l'utilisation du marteau-piqueur est un facteur de risque, on calcule un **rapport de risque** (RR) groupe exposé/groupe non-exposé. Si $RR=1$, il n'existe pas de lien entre l'exposition et la maladie. Si $RR>1$ l'exposition est un facteur de risque, si $RR<1$, l'exposition est un facteur protecteur.

Ici, le rapport de risque est de 7. Bob a sept fois plus de chance d'avoir mal au dos en utilisant un marteau-piqueur. L'utilisation d'un marteau piqueur augmente la probabilité d'avoir mal au dos : c'est un **facteur de risque**.

Mais procéder de cette façon demande beaucoup de moyens. Il faut évaluer et suivre un grand nombre de personnes qui n'ont à l'origine aucune demande de soin. C'est long et coûteux.

Une autre façon de faire, plus économique mais moins précise est de sélectionner un groupe de patients ayant mal au dos. On recrute alors un autre groupe de personnes, en tout point comparable, mais qui n'ont pas mal au dos. Puis on compare les expositions au facteur de risque présumé dans les deux groupes. C'est une étude de type **cas-témoin**.



Rapport de Cote
(odds ratio) = 6

Pour comparer l'exposition au facteur de risque présumé, on calcule pour chaque groupe le rapport nombre de personnes

exposées/nombre de personnes non-exposées. C'est la **cote** d'exposition. Pour comparer les deux groupes, on calcule ensuite le **rapport de cote** (odd ratio).

Le rapport de cote se lit de la même façon que le rapport de risque : Si $RC=1$, pas d'association entre exposition et maladie. Si $RC>1$ l'exposition est un facteur de risque, si $RC<1$, l'exposition est un facteur protecteur.

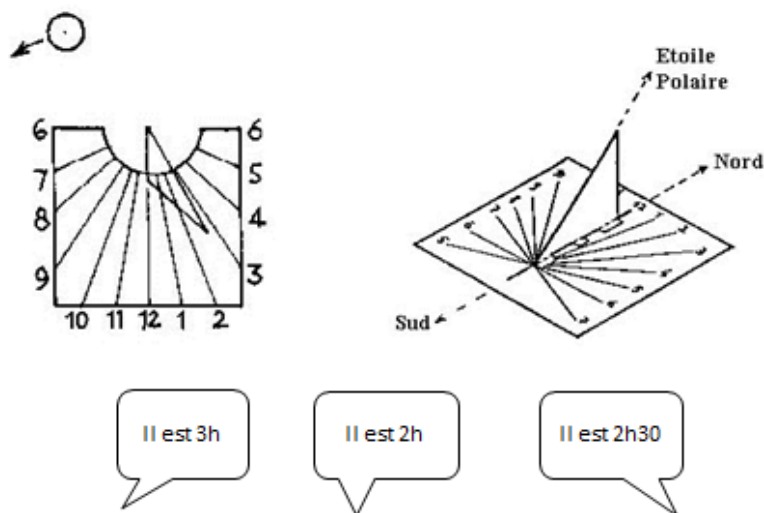
Revenons à Bob, parmi ses amis bricoleurs qui ont mal au dos, 4 ont utilisé un marteau-piqueur et 6 n'en ont pas utilisé. La cote chez les cas est de $4/6$. Parmi les 10 témoins bricoleurs qui n'ont pas mal, 1 seul a utilisé un marteau-piqueur. La cote dans le groupe témoin est de $1/9$. Le rapport de cote est de 6 ($4/6$ divisé par $1/9$).

Qualité d'un test diagnostic

Fiabilité (reliability)

Un bon test diagnostic doit pouvoir mettre tout le monde d'accord. Ce patient rentre dans telles ou telles catégories, auxquelles correspond tel ou tel traitement et tel ou tel pronostic.

La première chose à vérifier, c'est que tout le monde voit bien la même chose avec le même outil. La qualité de l'information dépend beaucoup de la façon dont elle est recueillie.



Il existe souvent plusieurs façons, plus ou moins précises, de recueillir une information.

J'ai le ventre qui gargouille,
Il est midi !



Il est 10h08

La fiabilité d'un recueil d'information se mesure par l'accord entre différentes évaluations. Soit entre deux évaluations faites deux fois de suite par la même personne, **-reproductibilité intra-évaluateur-** soit entre deux évaluations faite par deux personnes différentes **-reproductibilité inter-évaluateur-**.

La reproductibilité se résume par la proportion d'**accord** entre évaluation. C'est la concordance brute. Le problème est qu'il est toujours possible que les évaluations soient cohérentes par hasard.

Par exemple, je demande à deux personnes s'il va faire beau demain ou pas, la première dit beau la deuxième aussi, très bien mais elles avaient une chance sur deux d'être d'accord. Même si les deux personnes donnent leur résultat au hasard, en répétant l'expérience tous les jours, elles seront quand même d'accord une fois sur deux, la concordance brute est au minimum de 50% et pas 0%, pas très lisible...

C'est pourquoi on exprime la concordance avec une valeur qui prend en compte cet accord dû au hasard. C'est le coefficient **kappa**. Grâce à un calcul de l'accord théorique dû au seul hasard, le coefficient kappa est de 0,0 si les résultats sont donnés au hasard. Plus la concordance est bonne, plus le coefficient kappa est proche de 1. Pour une variable continue (comme une mesure de temps, de distance ou de degré) on utilise le coefficient de corrélation intra classe (CCI) qui a, à peu près, la même grille de lecture.

Qualité de la concordance entre observateurs mesurée à partir du coefficient **Kappa**

Niveau de concordance	Kappa
Très bon	> 0,81
Bon	0,80 – 0,61
Modéré	0,60 – 0,41
Médiocre	0,40 – 0,21
Mauvais	0,21 – 0,00
Très mauvais	< 0,00

Admettons que les deux personnes interrogées sur le temps soient d'accord avec un coefficient kappa supérieur à 0,81, leur évaluation est fiable. Mais reste maintenant à vérifier qu'il fasse vraiment beau demain. Ça c'est la validité.

Validité (accuracy)

Ce que l'on attend d'un test diagnostic, c'est qu'il reflète fidèlement la réalité. La fiabilité est une condition nécessaire à la qualité d'un test mais pas suffisante. Une horloge arrêtée est parfaitement fiable -tout évaluateur verra la même heure-, mais absolument pas valide -l'heure obtenue ne correspond pas à la réalité-.

Ce dont on veut être sûr c'est que si mon test est positif, le patient est malade. Si le test est négatif, le patient va bien et inversement. Mais l'information issue d'un test n'est jamais parfaite. Encore une fois, il nous faut passer par les probabilités.

En comparant les résultats de mon test avec un test de référence, je peux en calculer la sensibilité et la spécificité.

La **sensibilité** est la *probabilité* qu'un test soit positif si la caractéristique recherchée est véritablement présente. Si je suis malade, un test sensible sera positif.

La **spécificité** est la *probabilité* qu'un test soit négatif si la caractéristique recherchée est absente. Si je ne suis pas malade, un test spécifique sera négatif.

Exemple

Finalement je me dis que la kiné ne rapporte pas grand-chose et je décide de braquer une banque.



Je commence par remplir mon sac avec tout ce qui brille. Comme ça je suis sûr de ne rien laisser sur place. Le critère « brillant » est sensible. Mais tout ce qui brille n'est pas en or et mon sac est encombré d'objets sans valeur. Le critère « brillant » n'est pas spécifique.

Autre stratégie, je ne prends dans mon sac que les lingots, comme ça je suis sûr de ne prendre que de l'or. Le critère « en lingots » est spécifique. Mais

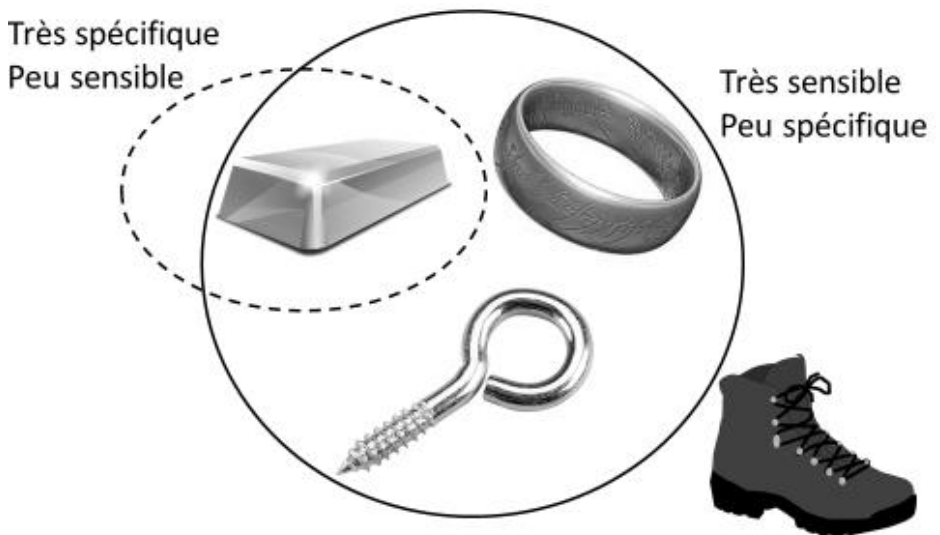


l'or ne se trouve pas forcément en lingots et je vais laisser sur place les autres bijoux. Le critère « en lingots » n'est pas sensible.

En deux mots, avec mon critère sensible j'aurais dans mon sac tout l'or mais pas que de l'or. Tandis qu'avec mon critère spécifique je n'aurai que de l'or mais pas tout l'or.

Le plus malin est d'utiliser une combinaison de critères. Je ne prends pas ce qui ne brille pas, et parmi ce qui brille je ne prends que les lingots, les bagues et les bracelets. Le premier critère sensible, s'il est négatif, me permet d'exclure ce que je ne cherche pas. Le critère spécifique, s'il est positif me permet d'inclure ce que je cherche.

Avec la combinaison de ces trois critères, je ne prends sans doute pas tout l'or, j'aurais sans doute dans mon sac quelques babioles sans valeur mais j'aurais gagné ma journée sans m'être fait prendre parce-que j'avais un sac trop lourd ou que j'aurais passé trop de temps à trier mes bijoux.



Dans ton cabinet Lundi matin...

« C'est bien jolie tous vos chiffres mais moi, dans mon cabinet, mes patients n'en ont rien à faire de leur écart-type. Merci pour la distraction mais j'ai du travail... »

Mais les statistiques sont là, tapies dans l'ombre. Restons à l'affût, on devrait en voir apparaître...

Tu reçois dans ton cabinet Bob qui t'est adressé pour lombalgie.

Il fait partie des 11 millions de français qui déclare avoir souffert des lombaires plus de 30 jours cette année, ce qui représente 17% de la population (54% au moins une journée dans l'année) (c'est la **prévalence**).

Bob souffre depuis 3 mois, il fait partie des 8% de nouveau cas de lombalgie cette année (ça c'est l'**incidence**). Bob fait partie des 56% des personnes qui souffrent du dos depuis plus 3 mois qui cherche un traitement mécanique. (99% prennent des médicaments).

En kinésithérapeute consciencieux, tu te demandes pourquoi ton patient en est arrivé là, qu'est ce qui dans sa vie l'amène à souffrir du dos, bref tu recherches ses **facteurs de risque**.

Ton patient te rapporte qu'il travaille dans les travaux publics avec un marteau piqueur ($RR=7$).

Bob fait de la course à pied. Il pense que c'est la cause de ses douleurs. Mais toi tu sais qu'il n'en est rien ($RR=0$).

Bob a eu très mal la première semaine puis les douleurs ont diminué. Mais depuis 2 mois plus rien ne change (histoire naturelle).

Tu lui fais remplir un diagramme corporel. Ce n'est pas très **fiable** mais toujours utile. De toute façon toute ta prise en charge n'en dépend pas.

Il te dit que d'après une radio du bassin, il a une jambe plus longue que l'autre de 9mm. Tu ne t'affoles pas pour autant car tu sais que **l'intervalle de confiance** de cette mesure est de 25 millimètres, les intervalles de confiance à droite et à gauche se chevauchent immanquablement et pas qu'un peu! Pas la peine non plus de vérifier par toi-même, la **reproductibilité** inter évaluateurs de l'inégalité des membres inférieurs est de $\kappa=0,0$ (du hasard à l'état pur).

Après une observation de ses amplitudes lombaires, tu lui trouves des pertes d'amplitude modérées. Ce n'est pas très reproductible en inter examinateur mais plutôt correct en intra examinateur, ça tombe bien, c'est toi qui devra le réexaminer. Bob a mal en extension debout, ça c'est plutôt reproductible, c'est un bon marqueur.

Bob a un signe de Lasègue négatif (**sensibilité** 88% pour une hernie discale à l'imagerie). Tu peux donc exclure une hernie discale (si Bob en avait une, tu avais 88% de chance que le Lasègue soit positif).

Tu suis sagement ta fiche et fait faire à Bob une série de flexion debout. Produit, empiré. La reproductibilité ? L'étude reste à faire. Passage à plat ventre. 10 extensions en procubitus. Augmenté, Pas Empiré. Réponse mécanique, Amplitudes augmentées (reproductibilité acceptable à condition que tu sois seul juge). Après une série de surpression, tu lui fais donc une mobilisation en procubitus. Produit, ($\kappa=0,4$) centralisé ($\kappa=0,7$).

Comme 50% des patients lombalgiques (**prévalence**) Bob centralise à la première consultation. Sa centralisation te permet d'être sûr de l'origine discale (**spécificité**= 95%). Et de te montrer rassurant pour Bob, le pronostic est bon (Rapport de Cote=9 entre la centralisation et la reprise du travail).

Tu conclus donc un syndrome de dérangement. (Selon le niveau de formation entre certifié MDT **kappa**=0,6 concordance brute=91%; entre kiné en formation niveau A, B, C ou D kappa =0,4 concordance brute 86%).

Tu peux être fier de toi ! Tout en restant modeste. La taille de l'effet du MDT en aigu sur l'EVA est 0,7 sur 10...Pas folichon... Mais toujours mieux que le paracétamol et le décontractant musculaire que Bob prend qui a un effet de 0,0.

Bob n'est pas sauvé pour autant, son risque de récurrence dans l'année est de 75%.