



ECOLE D'EDUCATION PHYSIQUE ET DE SPORT

STATISTIQUES

Genève : 2009

Cours dispensés par : Guy Haller

Prise de notes : Yann Bernardini

Révision : Claudine Bernardini

Conception et réalisation : Yann Bernardini

Genève, le 12 mai 2009

STATISTIQUES

Analyse de l'article sur l'électrostimulation :

Il est primordial de choisir les patients en fonction d'une liste à numéro randomisé par ordinateur.

De même, il faudrait que les enveloppes soient opaques.

Si les participants sont randomisés et qu'ils partent de l'étude, on doit les prendre en compte !

Le score de qualité sur les 10 questions d'analyse, nous permet de constater la pertinence et la qualité de l'étude.

Si l'on a plus de la moitié des points ($> 5/10$), l'étude peut être qualifiée de bonne.

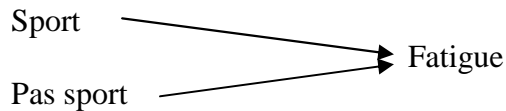
Pour une bonne cohorte, on se doit d'avoir un environnement semblable, du même bassin de population.

Plus le NNT est haut, plus c'est mauvais. Car l'on doit exposer plus de personne pour avoir de résultats.

STATISTIQUES

Vu qu'il y a une idée de temporalité, la cohorte correspond bien.

A la place de questionnaire, on peut demander aux parents ou à des collaborateurs d'observer les participants. De même, on pourrait calculer le taux d'acide lactique dans les muscles pour mesurer la fatigue.



Il va y avoir des facteurs confondants (différent de celui observé) :

- Âge
- Sexe
- Niveau d'entraînement
- Constitution
- Maladie

Qui doivent disparaître ou être égal dans les deux groupes. Cela est possible grâce au hasard.

Une analyse univariée prendra un seul critère (ex. fatigue), alors qu'une analyse multivariée prendra en compte plusieurs critères (âge, sexe, niveau d'entraînement, constitution, maladie).

Lors d'une étude multivariée, on tire les deux droites et on fait la moyenne entre les deux (âge et performance physique).

Le chiffre magique pour une cohorte est de 80 %. Si l'on perd plus que 20 %, ce n'est plus très fiable pour la cohorte.

C'est une étude bien faite au niveau du design, mais qui manque un peu de détails et n'est pas très fiable au niveau du questionnaire, des facteurs confondants et des différences de classes d'âges. Les résultats trouvés sont donc difficilement quantifiables.

Une étude de cas témoin se dit : POCE

- Participants
- Outcome (malades)
- Contrôle
- Exposés (cause)

Un biais de mesure est caractérisé par une approximation de données.

Les biais de mémoire sont fréquents chez les personnes n'étant pas malade, alors que celles qui le sont beaucoup sont plus précises (recherche de causes).

Le matching est le fait de prendre des sujets du même groupe avec les mêmes caractéristiques, ce qui limite les biais.

La formule d'Odds ratio est $= a/c \cdot b/d = ad/bc$

On fait des Odds ratio et non des risques relatifs car c'est nous qui faisons le choix sur les participants.

3/1 \Rightarrow 3 chances de perdre et si l'on gagne on multiplie par 3 les gains.

STATISTIQUES

Analyse du texte sur l'hernie discale :

Cas témoin :

- P : Population : les travailleurs de l'Oise
- O : Outcome : les hernies discales (HDL)
- C : Contrôle : ceux qui n'ont pas eu d'hernie discale
- E : Exposé (causes) :
 - o Type d'activité
 - o Antécédent
 - o Ambiance au travail
 - o Donnée démographique (âge, sexe)
 - o Posture
 - o profession

Le confondant est une donnée alternative du facteur de risque qui n'est pas celle qui est prévu dans l'étude. (Ici = tabac).

Une étude cas témoin permet de vérifier une quantité de facteurs de risque mais une seule maladie.

Une étude de cohorte permet d'étudier un large pannel d'Outcome, mais qu'un facteur de risque.

Cette étude n'est pas totalement liée au hasard, car le fait de prendre un patient sur trois à une certaine heure reprend des gens de la même fonction. Le mieux étant le hasard par liste à l'ordinateur.

Par le fait de prendre les patients chez le médecin, on fait un biais de sélection avec ceux ayant la maladie la plus grave. (Biais de sélection)

Il y a deux grandes familles de biais :

- biais de mesure : mesure l'exposition ou la maladie (biais de mémoire => questionnaire)
- biais de sélection : sélection de l'échantillon

Le but de l'analyse multi variée est de voir l'importance de chaque cause par rapport à la maladie.

L'intervalle de confiance se situe au niveau du tableau 5.

Si la différence est par exemple de (1,2 – 4,9), on est très large et c'est très significatif donc pas très intéressant.

Si la différence est par exemple de (0,7 – 3,00), ce n'est pas significatif car on est d'un côté protecteur (< 1,00) et de l'autre facteur de risque (> 1,00).

Dans les études transversales, on ne fait que le PEO. Il n'y a donc pas de contrôle.

On étudie les expositions et les causes.

- Participant
- Exposition
- Outcome (causes)

Le problème de l'étude transversale est le fait que j'étudie tout et que je tire des pourcentages dont je ne peux rien faire.

Les taux d'incidence si les études transférables sont bien faites peuvent être crus. Il faut néanmoins faire attention avec les facteurs de risque car ceux-ci ne font que décrire, mais ne pourrait pas être retrouvé dans une autre étude.

Si l'on veut mesurer l'exposition, on a donc recours à des études :

- Randomisée
- Cohorte
- Cas témoin

Dans les études transversales, on parle de proportion.

L'échantillon de commodité est le fait de prendre des gens en fonction des besoins (Sport => classe de sport), alors qu'un échantillon représentatif est l'ensemble de la population et sélectionner des personnes au hasard.

STATISTIQUES

Statistiques

Statistique descriptive :

L'objectif est de reconnaître le type de variable. Derrière un chiffre il y a toujours une réalité qu'il faut savoir décrire.

Statistiques :

- Récolte des données

Est-ce une variable :

- Catégorique : valeur qui désigne des objets non-définis (nationalité, sexe)
 - o Nominale : qui n'ont aucune valeur l'une avec l'autre
 - o Ordinale : ont un rapport l'une avec l'autre dans un ordre défini (Score de Glasgow)
- Numérique : ce sont des chiffres définis depuis le départ, échelle
 - o Discrète : aucune valeur l'une avec l'autre
 - o Continue : taille, poids et température (Celsius)

Les proportions sont une manière typique de faire de la distribution (groupe, tendance, diffusion).

Avec un chiffre il y a toujours une unité de mesure : code (kg, mml, etc.)

Pour tout ce qui est des données catégoriques, on utilise le mode de pourcentage et la notion de groupe car les données n'ont aucun rapport les uns avec les autres.

Un histogramme s'utilise uniquement pour les données continues (barres collées).

Les diagrammes en secteur sont peu utilisés dans la presse scientifique car ils sont trompeurs (presse laïque).

La tendance centrale résume le tout du groupe (taille).

La médiane est une manière de calculer la moyenne. Si l'on a un chiffre impair de données, on prend celui qui se trouve au milieu. Si l'on a un chiffre pair, on prend la moyenne des deux chiffres à la médiane.

Le diagramme à point est identique que l'histogramme sauf que l'on colle les données (barres).

L'étendue de la distribution est le maximum ./ le minimum.

Le Quartile est la variable entre la moyenne, le 25 et le 75.

Le percentile est la valeur où il reste 10% en-dessus et en-dessous.

La variance nous permet de nous rendre compte à quelle distance nous sommes de la moyenne (+ X ou - X). Les chiffres sont ensuite mis au carré pour enlever les chiffres négatifs. Ils sont ensuite additionnés et divisés par la moyenne des chiffres. On trouve donc le chiffre de la variance moyenne qu'il faut mettre à la racine 2.

X_i = chiffre

\bar{X} = moyenne des chiffres

Écart-type ou déviation standard sont des synonymes.

Une distribution normale est une distribution en cloche.

Les courbes peuvent être tirées vers la droite ou la gauche en fonction des valeurs que l'on obtient et qui ne sont pas représentatives du groupe.

La distribution numérique est normale dès 30 personnes (médiane et moyenne au même endroit), théorème de la limite centrale.

STATISTIQUES

Statistiques

Une variance est une mesure de l'erreur, une mesure de la variabilité.

Le paramètre correspond à l'Univers (l'ensemble de la population). Ecrit en lettres grecques.

L'estimateur correspond à l'Echantillon (la partie de la population choisie). Ecrit en lettres latines.

Une distribution normale ou de Gauss se traduit par une courbe en cloche :



1 déviation standard sont égales à 68,27 % de la population.

2 déviations standards sont égales à 95,45 % de la population.

3 déviations standards sont égales à 99,73 % de la population.

Plus l'échantillon est grand, meilleur sera le résultat et la représentation de la population.

Une erreur aléatoire est une erreur de précision.

Une erreur systématique est une erreur de validité. Cela entraîne un biais (différence). On pourrait prendre comme exemple d'une erreur systématique une balance qui affiche toujours 3 kg de moins !

STATISTIQUES

Une hypothèse nulle est un paramètre que l'on souhaite tester.

Lorsqu'on définit une hypothèse, on peut ensuite classer les résultats dans deux zones :

- Zone d'acceptation
- Zone de rejet

Une hypothèse nulle (H_0) est lorsqu'on pense obtenir un résultat identique en fonction de deux groupes ou de deux produits (homme vs femme / échantillon A vs échantillon B).

Une hypothèse H_1 est lorsqu'on pense obtenir un résultat différent (échantillon A vs échantillon B / homme vs femme).

La zone d'acceptation est à 95 % des sujets. Ce qui se trouve en-dehors est la zone de rejet.

La puissance est la capacité à détecter une erreur dans une hypothèse. Plus le nombre de sujets est élevé, plus les chances d'erreurs sont faibles.

Si P est plus grand que 0.05, on se trouve dans la zone d'acceptation de H_0 et on ne peut pas rejeter l'hypothèse.

Si P est plus petit que 0.05, on se trouve dans la zone de rejet de H_0 et on doit rejeter l'hypothèse.

Un petit échantillon est autour d'une dizaine d'individus.

Un grand échantillon est un nombre au-dessus de 100.

STATISTIQUES

Présentation :

- Analyse sur le résumé de l'étude
- Critique face à l'étude
- Présentation orale (tenue, etc.)
- Questions du prof

Si l'on peut comparer statistiquement les résultats d'une revue systématique, on peut faire une méta-analyse.

La méta-analyse est l'analyse de l'analyse.

Une revue systématique est la prise en compte de toutes les analyses traitant d'un sujet et une quantification des résultats par rapport aux deux possibilités (vrai / faux).

La méta-analyse se base sur un test statistique (résultats), alors que la revue systématique se base sur l'ensemble des études peu importe les critères mais dont le sujet est le même.

Dans les méta-analyses, on effectue le PICO (Participants, Intervention, Contrôle, Outcome).

En prenant un standard de critère qualité plus bas, on sera plus représentatif du Monde, mais on va s'éloigner du modèle de laboratoire. Si on fait le contraire, on va avoir peu d'études qui correspondent à la nature de la société, mais une haute valeur de laboratoire.

L'abréviation américaine GIGO donne Garbage In – Garbage Out qui dit que lorsqu'on prend des mauvaises études, notre étude sera mauvaise.

La précision est analysable à l'intervalle de confiance.

STATISTIQUES

Faire examen en blanc pour le 18.05.2009

Test diagnostique :

A travers l'idée de test, il y a une idée de mesure (podomètre, cardiofréquencemètre).

La prévalence est le nombre de malade sur la population totale.

La sensibilité est la capacité de mon test à trouver une information (malades). Capacité d'échapper au contrôle.

La spécificité est le nombre de malades qui n'ont pas été décelés (le contraire de sensibilité). Capacité à être considéré malade alors que l'on n'est pas.

Une bonne spécificité et une bonne sensibilité sont des bons tests (HIV : 99 %).

Valeur prédictive + (valeur du test à prédire) est le nombre de malades sur le nombre de tests positifs.

Valeur prédictive – (valeur du test à prédire) est le nombre de non-malades sur le nombre de tests négatifs.

La sensibilité et la spécificité de la maladie ne change pas avec la prévalence alors que la prédictivité positive et négative oui.

STATISTIQUES

Lors de la lecture ou de l'analyse d'un article, il y a 4 étapes importantes :

- Introduction
- Méthodes
- Résultats
- Discussion

Études randomisées contrôlées : (on fait quelque chose)

- *Étude où on applique un traitement*
- *Aucune connaissance des sujets*
- *Sélection de la population (échantillon)*
- *Lié au hasard (tri par ordinateur) :*
 - o *Groupe intervention*
 - o *Groupe contrôle*
- *Étude longue et coûteuse*
- *PICO :*
 - o *Participants*
 - o *Intervention*
 - o *Contrôle*
 - o *Outcome (issue)*

Les risques relatifs doivent être supérieurs à 1, sinon le placebo est plus efficace (!).

NNT (Number Needed to Treat) : plus il est petit, mieux c'est. C'est le nombre de patients à traiter pour obtenir un résultat. Calcul : $1 - (\text{résultat des exposés} / \text{total}) / (\text{résultat non-exposé} / \text{total})$

Risque relatif : plus il est grand mieux c'est

Il existe plusieurs biais :

- Biais de sélection (lié au hasard => ordi)
- Biais de performance (double aveugle)
- Biais d'analyse (suivi des « malades » / sujets à garder où laisser)
- Biais de détection (contrôle des résultats)

	Effet	Ø Effet
Traitement	a	b
Ø Traitement	c	d

Réduction de risque absolue=CER-EER

$$NNT = 1 / CER - EER$$

Effet (risque) si exposé (EER) ou risque absolu = $a/a+b$

Effet (risque) si non-exposé (CER) = $c/c+d$

Différence de risque = $EER - CER$

Risque relatif = EER / CER

Cohorte : (on observe)

- *Temps nécessaires (temporalité)*
- *Source de données identiques*
- *Méthode standard et reproductible*
- *Long et coûteux*
- *Sensible à la perte de suivi (20 % de perte maximum)*
- *Analyse de causalité*
- *Très bonne validité*
- *Facteur confondant (âge, sexe, entraînement, constitution, etc.)*
- *PECO :*
 - o *Participants*
 - o *Exposition*
 - o *Contrôle*
 - o *Outcome (issue)*

Une analyse univariée prend en compte un seul et unique critère.

Une analyse multivariée prend en compte plusieurs critères.

Cas-témoin : (on observe)

- *Biais de mémoire*
- *Approximation des données*
- *Matching des personnes de même groupe*
- *Odds ratio (=/= risque relatif), car c'est nous qui faisons le choix*
- *Restriction à une seule et unique maladie*
- *Caractéristiques proches*
- *Biais de mesures, données manquantes*
- *Utile pour les maladies rares*
- *Très sensible aux biais*
- *Difficulté de trouver une causalité*
- *POCE :*
 - o *Participants*
 - o *Outcome*
 - o *Contrôle*
 - o *Exposés (causes)*

Les études de cohorte permettent de trouver plusieurs maladies avec un seul facteur de risque.

Les études de cas-témoin permettent de trouver une seule maladie avec plusieurs facteurs de risque.

Un facteur confondant est la donnée alternative du facteur de risque pas pris en compte dans l'étude.

Il existe plusieurs biais :

- Biais de mesure : mesure de l'exposition des malades (biais de mémoire)
- Biais de sélection : sélection de l'échantillon

Un échantillon de commodité est le fait de prendre un groupe facile à sélectionner, pas de sélection au hasard (ordi).

Un chiffre en dessus de 1 est synonyme de facteur de risque.

Un chiffre en dessous de 1 est synonyme de protection.

Etudes transversales : (on observe)

- Proportion
- Pourcentage
- Exposition : Nombre de sujet exposé / nombre de sujet total
- Outcome : Nombre de sujet malade / nombre de sujet total
- Mesure de prévalence
- Hypothèse générée
- Peu de temps et pas cher
- Pas de séquence temporelle
- Pas d'association
- Biais de sélection et de mesure
- PEO :
 - o Participants
 - o Exposition
 - o Outcome

Variables catégoriques : valeur qui désigne des objets non-définis

- Nominale : nationalité, sexe
- Ordinale : 1, 2, 3

Variables numériques : chiffres qui définissent une échelle

- Discrète : nombre enfant dans la salle
- Continue : taille, poids, température

Un quartile est des tranches de 25 (0,25).

Un percentile est des tranches de 10 (0,10).

La variance est la distance moyenne du côté positif et négatif (mis au carré). Cela mesure l'erreur.

Il existe plusieurs façons de calculer :

- Moyenne (entier divisé par 2)
- Médiane (chiffre se trouvant au milieu)
- Mode (le chiffre le plus souvent dans l'échantillon)
- Diffusion (le chiffre le plus grand – le chiffre le plus petit = X (diffusion))
- Variance (1. 68,27 % (-1 / +1) ; 2. 95,45 % (- 1.96 / + 1.96) ; 3. 99,73 % (- 2.58 / + 2.58)

Un paramètre est la dimension de l'Univers (toute la population).

Un estimateur est l'échantillon (une partie de la population).

Une distribution normale où « de Gauss » est une courbe : 

Une erreur aléatoire est une erreur de précision.

Une erreur systématique est une erreur de validité (+ 3kg sur la balance).

Une revue systématique : tout ce qui a été fait dans ce domaine d'étude, que l'on peut trouver dans la littérature.

Une méta analyse : regroupement de toutes les statistiques de la revue systématique en un seul tableau (fusion des données)

- Aspect systématique
- Combinaison de plusieurs études
- Meilleure généralisation
- Sensible aux biais de sélection
- Dépend de la qualité de l'étude

Hypothèse 0 (H0) : obtenir le même résultat pour 2 groupes (hommes vs femmes)

Hypothèse 1 (H1) : obtenir un résultat différent pour 2 groupes (hommes vs femmes)

La zone d'acceptation (≠/≠ zone de rejet) se situe à 95 %.

La puissance est la capacité à détecter une erreur. Plus l'échantillon est grand, moins il y a de chance d'avoir des erreurs.

Un petit échantillon compte au minimum 10 participants.

Un grand échantillon compte au minimum 100 participants.

Un Test est une mesure.

La prévalence est le nombre de malade / la population totale.

La sensibilité : capacité de mon test à trouver l'information (maladie ?)

La spécificité : capacité des non-malades à ne pas être dans le test

Si le test est bon, il a une excellente sensibilité et spécificité (HIV 99 %).

La valeur prédictive positive : nombre de malades / nombre de tests positifs

La valeur prédictive négative : nombre de non-malades / nombre de tests positifs

Les valeurs prédictives changent avec la prévalence.

Une cohorte est un objet auquel on a été exposé (Hiroshima) et on observe les symptômes.

Une étude cas-témoin est l'observation des malades / blessés (Outcome) afin de rechercher les causes.

Lorsqu'on compare la population, c'est une valeur catégorique : Chi 2

Lorsqu'on compare des moyennes, c'est une valeur numérique : T test

L'écart type se calcule par la variance. On calcule la différence entre chaque nombre et la moyenne, puis on met tous les chiffres au carré. On divise ensuite le total par le nombre de variables (chiffre mis au carré) - 1. Une fois le résultat obtenu, on effectue une racine carré... !

Calcul de la sensibilité, spécificité, valeur prédictive positive et négative :

Sensibilité et Spécificité ↓		Dopé : oui	Dopé : non
	Test positif	34	20
	Test négatif	116	9830
	→ Prédictivité positive et négative		

Sensibilité : 34 / 150

Spécificité : 9830 / 9850

Valeur prédictive positive : 34 / 54

Valeur prédictive négative : 116 / 9946