

Université Saad Dahlab Blida1

Faculté de Médecine

Département de Pharmacie

5^{eme} Année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

**Statistiques descriptives : Définition préliminaires (effectif, pourcentage, ratio, taux) Ensemble/Population/Echantillon/Elément/Individu/Echantillonnage
(Cours à l'usage des étudiants en sciences médicales)**

Dr BENILHA.S

Statistiques descriptives : Définition préliminaires (effectif, pourcentage, ratio, taux/Ensemble/Population/Echantillon/Elément/Individu/Echantillon

I-Définition de la statistique descriptive :

Ensemble des méthodes qui permettent de décrire les unités statistiques qui composent une population.

Buts de la statistique descriptive :

Toute série d'observations comporte un certain nombre de données relatives à un ou plusieurs caractères ou encore **variables**.

Le but des statistiques descriptives est de décrire un ensemble d'observations à l'aide de quelques éléments caractéristiques (tableaux, graphiques, numériques).

II-Concepts de base en statistique :

Population :

Désigne l'ensemble des entités ou des individus qui font l'objet de l'étude.

Echantillon :

Une partie (sous -ensemble) de la population dont l'effectif (la taille) est très réduit et qui permet d'effectuer l'étude expérimentale.

Individu ou unité statistique :

Unité de base sur laquelle l'observation réalise un certain nombre de mesures (une personne, un ménage, une ville,).

III- Le caractère en statistique :

Contrairement à une constante -caractéristique ayant la même valeur pour tous les individus- **une variable** comporte nécessairement plus d'une modalité, les modalités sont les différentes catégories que peut présenter une variable.

Exemple : le « **groupe sanguin dans le système ABO** » est **une variable** car il comporte les modalités : groupe O, groupe A, groupe B, groupe AB.

Donc, chaque individu d'un échantillon ou d'une population sera décrit relativement à un ou plusieurs **caractères** (on dit aussi **variables**).

Types des caractères :

1-Variables qualitatives :

Une variable qualitative est un caractère dont les modalités s'expriment par des qualités et non pas par des valeurs numériques.

On dit d'un caractère qu'il est qualitatif, si ses diverses modalités ne sont pas mesurables.

Le « **groupe sanguin** dans le système ABO », le « **sexe** » ou « genre » avec ses deux modalités : masculin et féminin sont deux variables qualitatives,
Les variables qualitatives à **deux modalités** sont souvent désignées sous le vocable de « variables **dichotomiques** » parce que de type « l'un ou l'autre ».
Tandis que les variables comportant plus de deux modalités, comme le « groupe sanguin » sont appelées **multichotomiques**.

Classification des caractères qualitatifs :

***Nominale :** Ordre n'existe pas (ex: couleur des yeux,...).

Les variables multichotomiques dont les modalités ne sont pas soumises à **un tel ordre** sont appelées variables **nominales**.

Exemples :

- groupe sanguin** dans le système ABO (O, A, B, AB).
- statut matrimonial** (célibataire, marié, divorcé, veuf).

***Ordinale :** Peuvent- être **ordonnées**

Exemples :

- Stade de la maladie au moment du diagnostic : stades I à IV pour la maladie de Hodgkin.
- Gravité des lésions engendrées par les accidents de la route : mineure, modérée, sérieuse, sévère, critique.
- Intensité de la douleur : faible, modérée, importante.
- Niveau d'instruction : analphabète, primaire, secondaire, supérieur.

2-Caractère quantitatif :

Un caractère est dit quantitatif, si ses diverses modalités s'expriment par des valeurs numériques (**mesurables**).

A chacune de ses modalités, on peut associer un nombre, ce nombre est appelé **variable statistique**.

Exemples : l'âge, le poids, la taille, la glycémie, le nombre de personnes par ménage, le temps, etc.

Classification des variables quantitatives :

***Variables discrètes :** quand la variable ne peut prendre que des valeurs isolées (nombre entier).

Exemple: nombre d'enfants par famille, nombre d'accidents du travail par ateliers, nombre de globules par unité de volume sanguin »,

***Variables continues :** Quand la variable est susceptible de prendre toute valeur appartenant à son intervalle de variation.

Une variable quantitative continue a des modalités en nombre infini qui se situent à un point quelconque d'une échelle numérique,

Exemples :

La glycémie, l'heure du jour, l'âge, la taille », la tension artérielle, le taux d'urée sanguine, le taux de cholestérol, le temps depuis l'instauration d'un traitement » ...

IV-Effectif :

L'effectif ou fréquence **absolue** est le nombre d'individus par classe.

Ce dénombrement donne lieu à une représentation des données sous forme de tableau.

Exemples :

L'effectif ou fréquence absolue est le nombre d'individus correspondant à une modalité donnée d'une variable.

Exemple 1 :

La distribution de 50 malades selon le sexe, parmi ces 50 malades, 15 sont de sexe masculin et 35 de sexe féminin, les effectifs correspondant à chacune des deux modalités sont 15 et 35.

Exemple 2 :

On a dénombré sur un ensemble de 180 sujets, les individus qui appartenaient aux différents groupes sanguins (tableau).

Tableau : Description de l'échantillon des groupes sanguins.

<i>A+</i>	<i>A-</i>	<i>B+</i>	<i>B-</i>	<i>AB+</i>	<i>AB-</i>	<i>O+</i>	<i>O-</i>
80	10	20	5	5	2	50	8

V-Fréquence relative :

On peut définir les fréquences relatives (pourcentage ou proportion) qui sont, pour chaque classe, le rapport de son effectif au nombre total d'individus de la série des mesures.

Le rapport entre l'effectif d'une modalité de variable et l'effectif total de la série sur laquelle cette variable est mesurée. Le numérateur fait obligatoirement partie du dénominateur.

La somme des fréquences relatives est égale à 1.

Parfois, les résultats sont exprimés en pourcentage, chacune des fréquences relatives étant multipliée par 100. La fréquence relative s'exprime généralement en pourcentage parce que celui-ci est plus expressif.

Tableau : fréquences relatives d'un échantillon selon le groupe sanguin

A	B	AB	O
50	14	4	32

VI-Fréquences cumulées (relatives et absolues)

Les fréquences cumulées sont utilisées pour les données ordinales qui présentent des classes ordonnées.

Exemple :

Sur un échantillon de 500 malades cancéreux, on a noté le stade de la maladie.

On peut résumer ou présenter ces données par des fréquences relatives.

Les résultats obtenus sont présentés par le tableau.

Tableau : Répartition selon le stade de la maladie.

<i>Stade</i>	<i>Nombre de malades</i>	<i>Fréquence relative (%)</i>	<i>Fréquence relative cumulée (%)</i>
<i>1</i>	<i>350</i>	<i>70</i>	<i>70</i>
<i>2</i>	<i>110</i>	<i>22</i>	<i>92</i>
<i>3</i>	<i>30</i>	<i>6</i>	<i>98</i>
<i>4</i>	<i>10</i>	<i>2</i>	<i>100</i>

Interprétation : cette présentation permet de dire, par exemple, que **92%** des sujets examinés ont un stade inférieur ou égal à 2.

VII- Ratio :

Le Ratio : est le rapport entre les effectifs de deux modalités (classes) d'une même variable. Donc le numérateur et le dénominateur sont de même nature, mais exclusifs l'un de l'autre. C'est un nombre sans unité.

Le sex-ratio : est le rapport numérique des sexes à la naissance, le sex-ratio = (effectif des garçons) / (effectif des filles).

Si le sex-ratio est égal à 1.06 (ou 106%), cela signifie que pour 100 naissances féminines, il y a 106 naissances masculines.

Lorsqu'on n'a plus affaire à des naissances, on peut parler de ratio hommes/femmes.

Exemple :

Dans une population de 100 individus, on observe 49 hommes et 51 femmes.

Quel est le ratio H/F (Sexe ratio) ? $49/51 = 0,96$

VIII -Taux :

Le Taux : est un rapport qui prend en compte la notion de temps.

C'est La probabilité de survenue d'un évènement au cours du temps.

Le numérateur(N) : les individus ayant subis un évènement pendant une période de temps déterminé

Dénominateur (D): ensemble des individus susceptibles de connaître cet évènement pendant cette période.

Au (N) figurent des individus ayant subi un événement pendant une période de temps déterminé et au (D) figure l'ensemble des individus susceptibles de connaître l'événement pendant cette période (la population à risque).

Exemple :

Dans une population de 500 personnes, on a relevé au cours de l'année 2002,74 cas d'infarctus aigue du myocarde, le taux d'incidence est donc de 148/1000.

-la cantine d'une école recevant 250 enfants a été le siège d'une toxi-infection alimentaire collective.52 enfants ont présenté l'affection.

Le taux d'attaque est de $52/250 = 208$ pour 1000.

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

Présentation tabulaire de l'information

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

Présentation tabulaire de l'information

Introduction :

- Un tableau est utilisé pour mettre en ordre des données répétitives.
- C'est une unité d'information autonome se suffisant à elle-même, il doit être compréhensible en soi-même, indépendamment du texte.
- C'est pourquoi, il doit contenir toutes les indications utiles à sa compréhension:

Il doit être simple, compréhensible, avec

- un titre
- un numéro et
- une référence

Exemple :

Tableau. Prévalence des néphropathies vasculaires dans 14 régions françaises

Région	Cas prévalents dialysés par million d'habitants	Âge médian des cas prévalents (ans)	Pourcentage de néphropathies vasculaires* (%)
Auvergne	428	69,9	44,9
Basse Normandie	402	68,9	31,4
Bourgogne	418	69,4	38,1
Bretagne	387	70,7	23,4
Centre	493	70,7	35,9
Champagne-Ardennes	491	69,3	42,9
Ile de France	437	63,0	35,7
Languedoc-Roussillon	606	71,4	38,0
Limousin	374	70,6	37,9
Lorraine	505	70,3	40,1
Midi-Pyrénées	513	71,2	36,9
Nord-Pas de Calais	702	66,7	39,9
Provence-Alpes-Côte d'Azur	608	70,1	44,5
Rhône-Alpes	486	68,8	42,0

1-Titre :

- Un tableau ne peut être conçu sans un titre.
- Le titre définit le contenu du tableau, il est explicite quant à l'objet et la population concernée avec les caractéristiques de personnes, de temps et de lieu.
- En d'autres termes, un titre doit répondre à la question : “ **quoi, qui, où, et quand** ?”.

2- Autres indications :

- Des indications complémentaires sont portées en notes au bas du tableau.
- Il peut s'agir de la source des données, des unités de mesure, de la définition d'une abréviation, d'explications complémentaires, de mesures statistiques ou épidémiologiques testant la liaison entre deux variables...

3-Types du tableau :

a. Tableaux à une variable :

- Le tableau correspondant est dit à une **seule variable** ou tableau statistique à **simple entrée** ou encore tableau à une **seule dimension** (unidimensionnelle).

-Exemples :

Le tableau 1 contenait la distribution de malades selon le sexe.

Sexe	Effectif	%
Masculin	15	30.0
Féminin	35	70.0
Ensemble	50	100.0

b-Tableaux à deux variables :

- C'est un tableau à deux variables (à double entrée ou à deux dimensions).
- Ce tableau contient les données des tableaux à une variable de chacun des deux caractères.
- Un tel tableau peut avoir un intérêt purement descriptif mais va permettre surtout de tester la liaison entre deux variables.

*Exemple : Echantillon de 176 enfants répartis selon la variable "**diarrhée**" et la variable "**qualité de l'eau**".

Tableau : Répartition de 176 enfants selon qu'ils ont contracté éventuellement une diarrhée et la qualité de l'eau dont ils disposent

Diarrhée	Eau mauvaise	Eau bonne	Ensemble
Diarrhéiques	54	17	71
Non diarrhéiques	55	50	105
Ensemble	109	67	176

Il s'agit dans cet exemple de tester la relation entre la diarrhée (modalités : diarrhéiques et non diarrhéiques) et la qualité de l'eau (modalités : bonne et mauvaise). Puisque les deux variables sont qualitatives, il s'agit en fait de comparer des pourcentages.

c-Tableaux à trois variables :

- Un tableau peut présenter simultanément les données de trois variables, donc c'est un tableau à trois dimensions.
- Le tableau à trois variables contient l'information des tableaux à deux variables correspondants.

- Ce tableau à trois variables peut avoir un intérêt purement descriptif, comme, par exemple, contenir la répartition de malades selon l'âge, le sexe et la commune de résidence.

Exemple :

- L'investigation d'une intoxication alimentaire incriminait deux aliments simultanément (mayonnaise et œufs durs),
- Il conviendrait de construire un tableau à trois variables (intoxication, mayonnaise et œufs) pour départager entre les deux aliments incriminés

Tableau : étude de la relation entre l'intoxication et la consommation des œufs et la mayonnaise.

		Œufs oui	Œufs non	Total
Mayonnaise oui	Malades	13	3	16
	Non malades	3	0	3
	Subtotal	16	3	19
Mayonnaise non	Malades	0	2	2
	Non malades	1	9	10
	Subtotal	1	11	12
Ensemble	Malades	13	5	18
	Non malades	4	9	13
	Total	17	14	31

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

Représentations graphiques de l'information
(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

Représentations graphiques de l'information

I-Introduction :

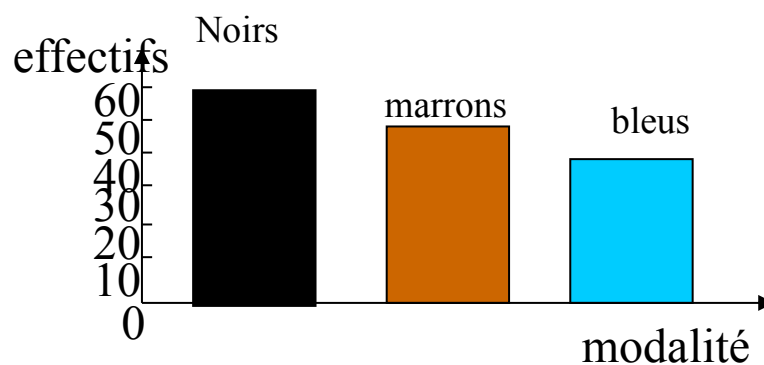
La représentation graphique est la première synthèse de l'information contenue dans un tableau statistique.

Le choix d'un type de graphe dépend de la nature des données du problème posé.

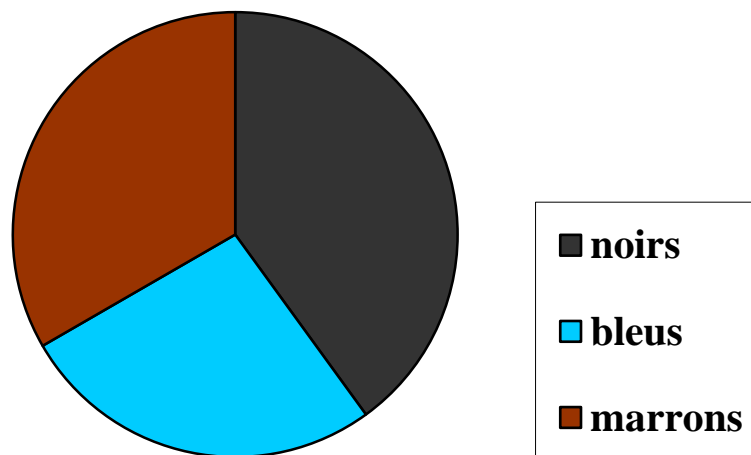
II-Représentation graphique :

1- Caractère qualitatif :

La représentation graphique par tuyaux d'orgue



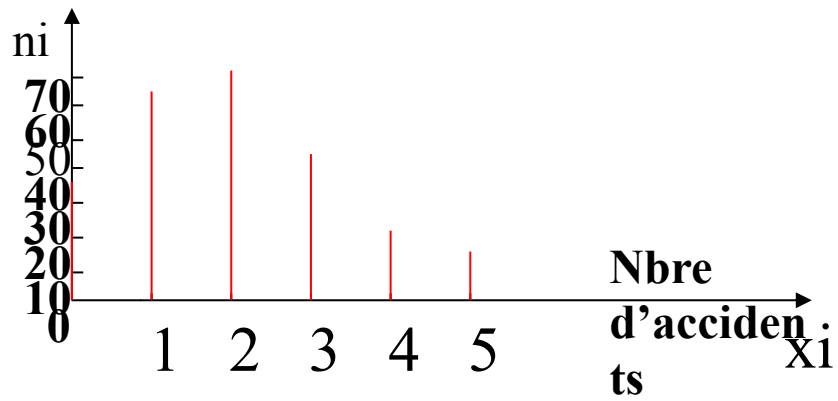
La représentation par secteur circulaire (nombre de modalités < 5)



2- Caractère quantitatif

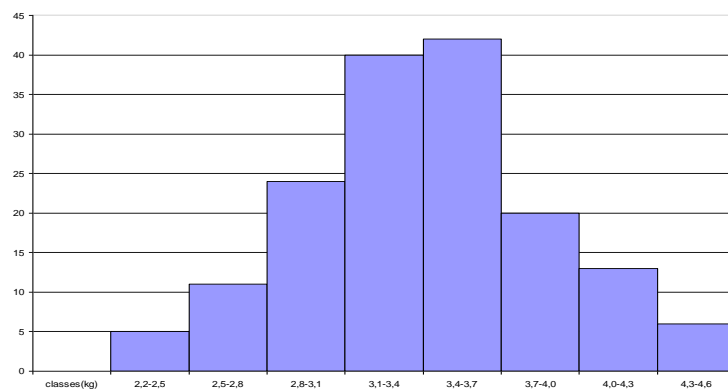
a-Variable statistique discrète : Diagramme en bâtons

Distribution des consultants en fonction du nombre d'accidents du travail



b-Variable statistique continue : Histogramme

La répartition du poids de 161 nouveau-nés



Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S

Module d'Epidémiologie et Méthodologie de la recherche

PARAMETRES DE REDUCTION

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

Paramètres de réduction

Introduction :

Afin de résumer une série statistique d'une façon simple tout en conservant au mieux le contenu informationnel en limitant au maximum la perte d'information, on utilise:

A- Les paramètres de tendance centrale (de position):

Elles permettent d'obtenir une idée de l'ordre de grandeur des valeurs de la série et indiquent la position où semble se rassembler les valeurs de la série.

B- Les paramètres de dispersion:

Elles quantifient les fluctuations des valeurs observées et leur étalement.

Les paramètres de tendance centrale :

a-mode:

Le mode est la variable qui a l'effectif le plus grand:

1-Données non groupées :

Ex: série : 3,5,7,15,**16,16,16**,17,17,30 Mode = 16

2-Données groupées: (Variable quantitative discrète)

Le mode correspond à l'effectif maximal. Mode = 4

Nombre d'enfant	Nombre de famille
0	4
1	5
2	10
3	16
4	18
5	14
6	7

3-Données groupées: (variable quantitative continue):

La classe modale correspond à l'effectif le plus élevé. Le mode correspond au centre de la classe modale. La classe modale : 3.5- 4

Le mode = $(3.5+4)/2 = 3.75$

Mode = 3.75 kg

POIDS en KG	effectif
2 - 2.5	2
2.5 - 3	4
3 - 3.5	6
3.5 - 4	30
4 - 4.5	8

B-médiane:

La médiane est une valeur de variable qui divise l'ensemble des observations en 2 parties égales, 50% de l'effectif se situe en dessous de la médiane et 50% de l'effectif se situe au dessus.

1-Données non groupées:

Nombre d'observation est impair :

Ex: 7, 9, 13, 45, 70, 101, 115 Médiane = 45

Nombre d'observation est pair :

Ex: 2, 5, 9, 10, 12, 14, 20, 22 Médiane = $(10+12)/2=11$

2-Données groupées :(variable quantitative discontinue),

La médiane est la valeur de la variable qui occupe le $(n/2)$ ème rang

N=80 le $(n/2)$ ème rang = 40

Nombre d'enfant	Nombre de famille	Effectif cumulé
0	4	4
1	5	9
2	10	19
3	16	35
4	18	53
5	14	67
6	7	74
7	6	80

Université Saad Dahlab Blida1
Faculté de Médecine
Département de pharmacie
Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche
5^{ème} année de Pharmacie

Mesures d'Association épidémiologique
Interprétation de l'intervalle de confiance de RR et OR à 95%

Dr BENILHA.S

Mesures d'Association épidémiologique

Interprétation de l'intervalle de confiance de RR et OR à 95%

I-Introduction

L'étude de l'association entre un facteur d'exposition et une maladie est l'une des étapes majeures de la recherche des facteurs étiologiques des maladies.

* Pour identifier un facteur d'exposition, il faut :

- mettre en évidence une liaison entre ce facteur et la maladie avec un test statistique approprié (le test du χ^2 ,...);

- mesurer la force de cette liaison.

II- Notion de risque et mesures d'association :

Notion de risque : Elle est définie par la probabilité de survenue de la maladie ou tout autres événements de santé (complication, décès...) dans une population à un moment donné ou pendant un intervalle de temps donné suite à l'**exposition** à des facteurs de risque.

Facteur de risque : désigne la source de risque qui augmente la probabilité pour un individu de développer une maladie, il peut être soit :

Individuel : biologique(hérédité), comportement (alimentation, habitudes toxique).

Communautaire : Hygiène de milieu.

III- Mesures d'association dans les études épidémiologiques :

- Afin d'établir clairement les mesures d'association entre la maladie et les facteurs de risque, on doit réaliser les études analytiques ont pour objectif de comparer des malades et des non malades selon leur niveau d'exposition à un ou plusieurs facteurs de risque.
- On distingue selon la chronologie du recueil des données :
 - ✓ Etude cas témoins
 - ✓ Etude Cohorte

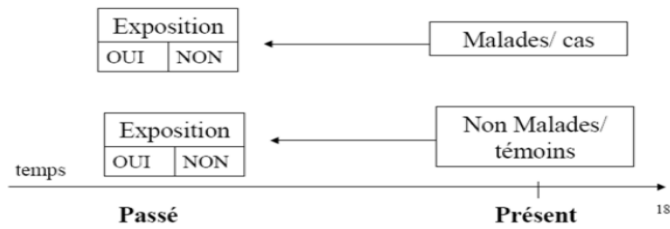
A-Etude Cas Témoins :

1-Principe :

Elle consiste à comparer la fréquence d'exposition au facteur parmi un groupe de malade et parmi un groupe de sujets non malades choisis comme témoins.

Si la survenue de la maladie est liée à l'exposition, on doit observer un pourcentage d'exposition plus élevé chez les cas que chez les témoins.

Schéma d'étude :



3-Presentation des données et mesures d'association :

A- Fréquence d'exposition

	cas	témoins	Total
Exposé	a	b	a+b
Non exposé	c	d	c+d
total	a+c	b+d	a+b+c+d

-Fréquence d'exposition chez les cas (EXP_c):

$$EXP_c = \frac{a}{a+c}$$

-Fréquence d'exposition chez les témoins (EXP_t):

$$EXP_t = \frac{b}{b+d}$$

Ces cotes en tant que telles ne servent à rien, mais leur rapport permet d'évaluer la liaison entre l'exposition et la maladie, ce rapport est appelé Rapport des Cotes (RC) ou Odds Ratio (OR).

B - Les cotes d'exposition :

*La cote d'exposition chez les cas : a/c

*La cote d'exposition chez les témoins : b/d

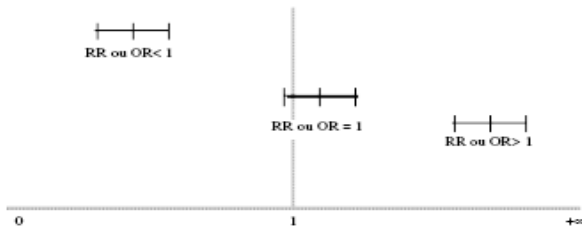
c-L' Odds ratios (OR) :

L'OR est la mesure la plus utilisée dans une étude cas témoins, l'OR est le rapport de la cote d'exposition chez les cas sur la cote d'exposition chez les témoins :

$$OR = \frac{a/c}{b/d}$$

4-Interprétation de l'OR :

L'OR est un nombre sans unité compris entre 0 et l'infini



OR = 1 : absence de risque, Intervalle de confiance à 95% inclut la valeur 1 : risque non mis en évidence (non significatif)

OR > 1, le facteur étudié est un facteur de risque

OR < 1 et le facteur étudié est un facteur protecteur

Exemple :

OR = 4, signifie que les sujets exposés ont un risque multiplié par 4 de contracter la maladie par rapport aux sujets non exposés.

Intervalle de Confiance De l'OR :

-Par définition les Cas et les témoins sont issus d'une population.

-L'OR est donc une variable aléatoire qui subit des fluctuations d'échantillonnage.

-on calcule donc un intervalle de confiance (IC) à 95% de RR

$$\text{IC à 95\%} = \text{RR}^{1 \pm (1,96 / \sqrt{x^2})}$$

Si l'IC inclut la valeur 1 ca signifie l'absence d'association entre le facteur et la maladie.

Comparaison OR et X^2 :

L'intervalle de confiance à la même signification que le résultat de p (degré de signification)

du test X^2 :

*si p est supérieur à **0,05** (test non significatif), alors, l'IC à 95% du RR contient la valeur 1, le test n'est pas significatif, le facteur étudié n'est pas un facteur de risque.

*si p est inférieur à **0.05**, alors l'IC à 95% exclut la valeur 1, et le facteur étudié est un facteur de risque.

Donc l'intérêt du calcul de l'OR est de donner la force le sens et le degré de signification de l'association lors le test de chi deux ne donne que le degré de signification de l'association.

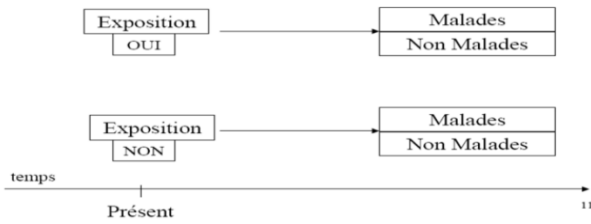
B-Etude de Cohorte :

1-Principe :

Une cohorte se définit comme un groupe de sujets suivis dans le temps, si l'événement observé est la survenue d'une maladie, on mesure, à la fin de l'étude, le nombre de sujets atteints par la maladie pendant la période d'étude.

Une étude de cohorte est parfois appelée étude exposés/non exposés.

2-Schémas de l'étude:



Une étude de Cohorte est une étude comparant une cohorte de sujets exposés à une cohorte de sujets non exposés. A l'issue d'une étude de cohorte, on compare donc le taux d'incidence entre exposés (**Ie**) et non exposés(**Ine**).

3-Présentation des données et Mesures d'association

a-Des taux d'incidences dans chaque groupe de comparaison. Ces taux d'incidence sont assimilés aux probabilités ou risque de survenue de la maladie

	malade	Non malade	Total
Exposé	a	b	a+b
Non exposé	c	d	C+d
total	a+c	b+d	a+b+c+d

Ie : incidence de la maladie chez les sujets exposés :
$$Ie = \frac{a}{a+b}$$

Ine : incidence de la maladie chez les sujets non exposés
$$Ine = \frac{c}{c+d}$$

b-La différence de risque: entre exposés et non exposés :

***C'est l'excès de maladie dû à la présence du facteur de risque.**

*c'est le risque de la maladie attribuable au facteur de risque chez les exposés.

$$DR=Ie-Ine$$

*Il représente la partie du risque exclusivement liée au facteur étudié.

C- Le risque relatif (RR) :

Le RR est le rapport entre l'incidence chez les exposés surs

L'incidence chez les non exposés, cet indicateur est appelé **encore le ratio de risque** ou

rapport de risque (RR).

$$RR= Ie/Ine$$

-Il est utilisé très fréquemment en recherche étiologique.

-Il exprime la force de l'association entre un facteur et la maladie.

Interprétation du risque relatif(RR) :

***un risque nul n'a pour valeur 1** : signifie l'absence d'association entre le facteur et la maladie.

***un risque inférieur à 1** : signifie que le facteur étudié est un facteur protecteur

***un risque plus de 1** : signifie que le facteur étudié est un facteur de risque, et plus le RR est éloigné de 1 plus l'association entre la survenue de la maladie et le facteur est forte.

Intervalle de Confiance du risque relatif :

-Une étude de cohorte est rarement réalisée sur l'ensemble d'une population à risque d'une maladie donnée.

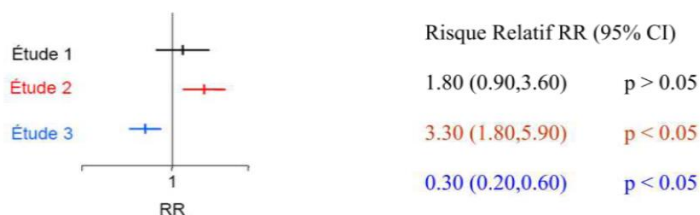
-Elle est effectuée sur un échantillon représentatif de cette population.

- Le RR est donc une variable aléatoire qui subit des fluctuations d'échantillonnage.

-on calcule donc un intervalle d confiance à 95% de RR.

$$IC \text{ à } 95\% = RR \pm (1,96 / \sqrt{x^2})$$

Exemples :



L'intérêt de calculer le **RR** et de son **intervalle de confiance** est de donner **la force, le sens et le degré de signification de l'association.**

Comparaison RR et X^2 :

On pourrait comparer les taux d'incidence dans les deux groupes : exposés et non exposés, par un simple test de X^2 , ce test permettra d'affirmer qu'il existe une différence significative entre les incidences observées :

*si p est supérieur à **0,05** (test non significatif), alors, l'IC à 95% du RR contient la valeur 1, le test n'est pas significatif, le facteur étudié n'est pas un facteur de risque.

*si p est inférieur à **0.05**, alors l'IC à 95% exclut la valeur 1, et le facteur étudié est un facteur de risque.

Conclusion :

La notion de risque est très importante en Epidémiologie.

*Elle se base sur les enquêtes épidémiologiques analytiques.

* L'enquête de cohorte reste la plus précise

*En pratique courante les enquêtes cas témoins et les enquêtes transversales sont les plus utilisées.

*La réduction des problèmes de santé repose sur la prévention et nécessite donc une connaissance rigoureuse des risques encourus d'où l'intérêt des études épidémiologiques.

Université Saad Dahlab Blida1
Faculté de Médecine
Département de pharmacie
5^{ème} année de Pharmacie
Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

Principe des tests d'hypothèse
(Cours à l'usage des étudiants en sciences médicales)

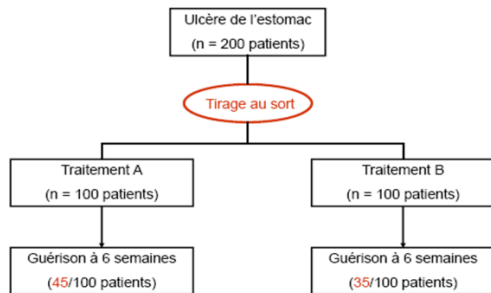
Dr BENILHA.S

Principes des tests d'hypothèse

Introduction :

- ❑ Lorsqu'on effectue une comparaison entre deux ou plusieurs séries de données on observe toujours une différence entre les paramètres mesurés.
- ❑ Le but du test est de déterminer si la différence observée est simplement due au hasard, (fluctuations d'échantillonnage), ou au contraire la différence observée est bien réelle.

Exemple : **Lequel de deux traitements est efficace ?**



Que conclure ?

Les tests statistiques d'hypothèse permettent de se fixer une règle de décision objective

- ❑ Un test n'a de sens que s'il teste une hypothèse **préalablement posée** afin de répondre **à une question**.
- ❑ Tout test statistique doit donc avoir pour objectif de **vérifier une hypothèse justifiée**.

Observation → hypothèse → test

Principe des tests :

- ❑ Servent à comparer des séries de données entre elles ; deux situations :
 - comparer un échantillon observé à une population de référence.
 - comparer deux ou plusieurs échantillons entre eux.
- ❑ Le principe général d'un test est de regarder si la différence qu'on observe est due au hasard ou non.

Quelle que soit la nature d'un test son principe et sa chronologie sont toujours les mêmes

1- Etablir l'hypothèse nulle(H0):

Proposer H0 c'est de supposer que la différence observée provient seulement **des fluctuations d'échantillonnage** donc la **différence n'est pas significative**.

2- Proposer une hypothèse alternative :

On appelle hypothèse **alternative H1** l'hypothèse qui sera retenue au cas où les résultats du test aboutirait à **rejeter** l'hypothèse nulle H0, donc lorsque **la différence est significative**.

3-calcul statistique :

Calculer une quantité mathématique X exprimant l'écart entre les paramètres ou les distributions (chaque test est sa formule).

Confronter cette quantité à un modèle de distribution théorique X` (**table statistique de chaque test**).

4- résultat d'un test

deux situations :

- la valeur de $\bar{X} < X'$ (**table**). On en conclut que la différence observée entre les paramètres étudiées **n'est pas significative** (fluctuations d'échantillonnage).

H0 est pas retenue

-la valeur de $\bar{X} \geq X'$, **donc** la différence observée entre les paramètres étudiées est significative.

on rejette l'hypothèse nulle H0 et on accepte l'hypothèse H1.

5-Choix du risque d'erreur :

a) **le risque α** :

c'est le risque de se tromper en rejetant H0.

On l'appelle risque de première espèce ou risque α

=**probabilité de rejeter H0 si H0 est vraie**

on lui assigne communément la valeur **5%**

b) **le risque β** :

risque β de deuxième espèce

= **probabilité de ne pas rejeter H0, si H1 est vraie.**

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S

Module d'Epidémiologie et Méthodologie de la recherche

Comparaison de pourcentages (Test de l'écart réduit (Z)

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

Comparaison de pourcentages : (Test de l'écart réduit (Z))

1-Introduction :

Pourcentages : Variable qualitative dichotomique (Présence/Absence, Malades/Non malades,).

P: le pourcentage (inconnu) d'individus présentant la caractéristique dans la **population**.

p: le pourcentage observé sur un **échantillon** de taille n dont k individus présentent la caractéristique.

Types de Comparaison :

Il existe trois types de comparaison :

- Comparaison d'un pourcentage observé à un pourcentage théorique.
- Comparaison de deux pourcentages observés : (Échantillons indépendants)
- Comparaison de deux pourcentages observés : (Échantillons appariés)

2. Comparaison d'un pourcentage observé à un pourcentage théorique :

Intérêt : déterminer si un pourcentage **p** observé sur un échantillon de taille n est pas différent d'une valeur théorique **P**.

Démarche :

⇒ Comparer **p** à **P**.

⇒ 1-Formuler les hypothèses :

⇒ Hypothèse nulle H_0 :

$$p = P$$

P est le pourcentage de la population dont l'échantillon est issu.

⇒ Hypothèses alternatives H_1 : $p \neq P$

2-Fixer le risque α à 5 %

3-Choisir la statistique:

Test z « Test de l'écart réduit » (loi normale centrée réduite)

4-Conditions d'application :

$$n \cdot P \geq 5 \text{ et } n \cdot q \geq 5 \quad (\text{N.B: } q = 1 - P)$$

N.B:

P : pourcentage de la population

$$Z = \frac{|p - P|}{\sqrt{\frac{P \cdot (1-P)}{n}}}$$

5-Calculer la statistique **Z**:

6-Comparer la valeur calculé de z avec la valeur critique de z_{α} (1,96):

- _ si $|z| < z_{\alpha}$: la différence n'est pas **significative** au risque α et donc **H_0 est retenue**.
- _ si $|z| \geq z_{\alpha}$: la différence est **significative** au risque α et donc **H_0 est rejetée , H_1 est retenue**.

Exemple :

En France, 7% des personnes hospitalisées contractent une infection nosocomiale dans l'établissement où elles sont soignées.

Sur un échantillon de 250 personnes soignées à l'hôpital H, 28 ont contracté une infection nosocomiale.

Le pourcentage observé sur l'échantillon diffère-t-il de la référence nationale au risque $\alpha = 5\%$?

Démarches:

1. Poser les hypothèses :

H_0 : Le pourcentage observé sur l'échantillon (**p**) ne diffère pas de la référence nationale (**P**),

c.à.d : **p = P**

H_1 : Le pourcentage observé sur l'échantillon diffère de la référence nationale, c.à.d. : **p \neq P**

2. Détermination du risque α à 5 %:

3. Choix du test : Test de l'écart réduit (test Z)

4. Vérification des conditions d'application:

Sachant que : $n = 250$, **P = 0.07** $q = 1 - P = 1 - 0.07 = 0.93$

np = 250 x 0,07 = 17,5 \geq 5 , **nq = 250 x 0,93 = 232,5 \geq 5**

5. Calcul de la statistique Z:

$$z = \frac{|p - P|}{\sqrt{\frac{P * (1-P)}{n}}} = z = \frac{0,112 - 0,07}{\sqrt{\frac{0,07 * (1-0,07)}{250}}} = 2,625$$

6. Comparaison et conclusion:

$$z = 2,625 \geq z_{0,05} = 1,96 : \text{rejet de } H_0.$$

On montre, au risque 5%, une **différence significative** entre le pourcentage de personnes hospitalisées contractant une infection nosocomiale à l'hôpital H et dans l'ensemble du pays (**p < 0,01**).

2. Comparaison de deux pourcentages observés: Échantillons indépendants

Intérêt : comparer 2 proportions (p_1 et p_2) dans 2 groupes indépendants de tailles n_1 et n_2 :

Démarches:

1. Formuler une hypothèse:

Hypothèse nulle H_0 :

$p_1 = p_2$ (p_1 et p_2 , pourcentages de la population dont sont issus les échantillons 1 et 2)

Hypothèses alternatives H_1 :

$$p_1 \neq p_2$$

2. Fixer le risque α à 5 %

3. Choisir la statistique : Test de l'écart réduit (test z)

4. Conditions d'application :

- $n_1 \cdot p_0 \geq 5$ et $n_1 \cdot (1 - p_0) \geq 5$
- $n_2 \cdot p_0 \geq 5$ et $n_2 \cdot (1 - p_0) \geq 5$ Avec p_0

$$p_0 = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}$$

p_0 : pourcentage commune

5. Calculer la statistique Z :

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_0 \cdot (1 - p_0)}{n_1} + \frac{p_0 \cdot (1 - p_0)}{n_2}}} \quad p_0 = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}$$

6. Conclusion et prise de décision:

Confronter la valeur de z calculée à la valeur critique z_α :

– $Z < Z_\alpha$: H_0 est retenue la différence n'est pas significative au risque α .

– $Z \geq Z_\alpha$: H_0 n'est pas retenue, donc H_1 est retenue la différence est significative au risque α .

Exemple

On désire comparer l'efficacité de deux traitements T1 et T2 sur 100 patients atteints d'une maladie M.

On tire au sort 2 deux groupes de 50 patients, un groupe est soumis à T1, le second à T2.

Le pourcentage de guérison chez les patients soumis à T1 est de 30%, chez ceux soumis à T2 de 40 %.

Le taux de guérison est-il différent entre les 2 traitements ?

Démarches :

1. Poser l'hypothèse nulle :

- H0: Le taux de guérison n'est pas différent entre les 2 traitements donc $p_1 = p_2$.
- H1: Le taux de guérison est différent entre les 2 traitements donc $p_1 \neq p_2$.

2. Détermination du risque $\alpha=5\%$.

3. Choix du test : Test de l'écart réduit (Test Z).

4. Vérification des conditions d'application :

$$n_1=50; p_1=0,3 \text{ et } n_2=50; p_2=0,4 \quad p_0 = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2} \quad p_0 = \frac{50 \times 0,3 + 50 \times 0,4}{50 + 50} = 0,35$$

$$q_0 = 1 - p_0 = 1 - 0,35 = 0,65$$

$$n_1 \cdot p_0 = 50 \times 0,35 = 17,5 \geq 5 \text{ et } \dots\dots\dots$$

$$n_2 \cdot q_0 = 50 \times 0,65 = 32,5 \geq 5$$

5. Calcul de la variable testée:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_0 \cdot (1 - p_0)}{n_1} + \frac{p_0 \cdot (1 - p_0)}{n_2}}} \quad z = \frac{0,3 - 0,4}{\sqrt{\frac{0,35 \times 0,65}{50} + \frac{0,35 \times 0,65}{50}}} = 1,05$$

$$z = 1,05 < z_{0,05} = 1,96 : H_0 \text{ est retenue}$$

6. Conclusion : il y'a pas de différence significative entre les taux de guérison avec les 2 traitements, au risque 5%.

3. Comparer deux pourcentages observés - Séries appariées -

Intérêt : on s'intéresse aux taux de guérison chez des sujets ayant reçus un traitement T1 et des sujets **appariés** ayant reçus un traitement T2.

on cherche à comparer p_1 et p_2 (les taux de guérison avec T1 et T2).

		Éch. 2		
		+	-	Total
Éch. 1	+	a	b	a+b
	-	c	d	c+d
Total		a+c	b+d	n

Les paires concordantes n'apportent pas d'information sur la liaison entre le traitement et la guérison. On doit donc se fonder sur la répartition des **paires discordantes**.

Conditions d'application : $b+c \geq 10$

Calcul de la statistique Z:

$$z = \frac{b - c}{\sqrt{b + c}}$$

Comparaison et conclusion :

_ $Z < Z_\alpha$: H_0 est retenue la différence n'est pas significative au risque α .

_ $Z \geq Z_\alpha$: H_0 n'est pas retenue, donc la différence est significative au risque α .

Exemple

On désire comparer l'efficacité de deux traitements T₁ et T₂ chez 100 patients atteint d'une maladie M.

Les deux traitements sont administrés aux patients. L'ordre d'administration des 2 traitements est tiré au sort en ménageant une période dite de *Wash-out* entre les 2 administrations.

Les résultats sont les suivants :

		T ₁	
		Succès	Échec
T ₂	Succès	24	16
	Échec	6	54

Le taux de guérison est-il différent entre les deux traitements ?

On cherche à comparer les pourcentages observés :

$$p_1 = \frac{24 + 6}{100} = 0,3 \quad p_2 = \frac{24 + 16}{100} = 0,4$$

1. Hypothèses :

H_0 : Le taux de guérison n'est pas différent entre les deux traitements ($p_1 = p_2$)

H_1 : Le taux de guérison est différent entre les deux traitements ($p_1 \neq p_2$)

2. fixer le risque à 5 %

3.choix du test :test de l'écart réduit (séries appariées)

4.conditions d'application vérifiées :

nombre de paires discordantes = $16 + 6 = 22 \geq 10$

$$Z = \frac{b - c}{\sqrt{b + c}} = \frac{16 - 6}{\sqrt{16 + 6}} = 2.13$$

6.Conclusion:

$Z=2.13 > 1,96$ donc

On montre, au risque 5%, une différence significative entre les taux de guérison avec les 2 traitements ($p < 0,05$).

Travaux dirigés :

Exercice 01 :

L'étude expérimentale d'un médicament a été pratiquée sur **100**malades divisés par tirage au sort en deux groupes A et B.

Le groupe A composé de **60** malades a absorbé le médicament étudié.

Le groupe B composé de **40** malades n'a absorbé qu'un produit inactif << placebo>> extérieurement identique au médicament.

Les résultats sont les suivants :

Groupe A : 40 malades guéris.

Groupe B : 20 malades guéris.

Peut- on conclure à l'efficacité de médicament ?

Exercice 02 :

On a observé dans une maternité sur un échantillon de **400** naissances dont les mamans font un travail pénible que le nombre des naissances prématurées est de **52**.

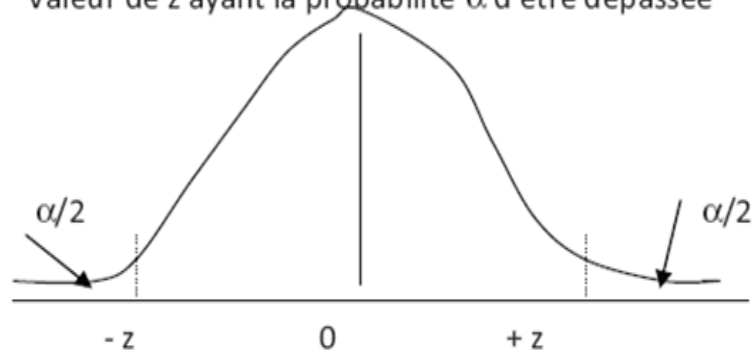
On admet que la proportion des naissances prématurées est de **8%**.

Le travail pénible est-il cause des naissances prématurées ?

Table de l'écart réduit z

Table des aires limitées par la courbe N (0, 1)

Valeur de z ayant la probabilité α d'être dépassée



α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.751	1.695
0.1	1.645	1.598	1.555	1.514	1.476	1.440	1.405	1.372	1.341	1.311
0.2	1.282	1.254	1.227	1.200	1.175	1.150	1.126	1.103	1.080	1.058
0.3	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.4	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.5	0.674	0.659	0.643	0.628	0.613	0.598	0.583	0.568	0.553	0.539
0.6	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.426	0.412	0.399
0.7	0.385	0.372	0.358	0.345	0.332	0.319	0.306	0.292	0.279	0.266
0.8	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.9	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013

Exemple : si $\alpha(z) = 0.23$, $z = -1.200$

Table pour les petites valeurs de $\alpha(z)$

α	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
z	3.291	3.891	4.417	4.892	5.327	5.731	6.109

Université Saad Dahlab Blida1
Faculté de Médecine
Département de pharmacie
5^{ème} année de Pharmacie
Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

**COMPARAISON DES DISTRIBUTIONS DE CARACTERES
QUALITATIFS LE TEST DU CHI-DEUX (χ^2)**

(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

COMPARAISON DES DISTRIBUTIONS DE CARACTERES QUALITATIFS LE TEST DU CHI-DEUX (χ^2)

A-Définition :

Formulations équivalentes : Test de chi- deux = Test de chi- carré =
Test de Pearson

Le principe général d'un test est de regarder si la différence qu'on observe est due au hasard ou si au contraire cette différence est telle qu'il est fort peu probable de l'observer par hasard

Quelle que soit la nature d'un test, son principe et sa chronologie sont toujours les

Mêmes.

B-Principe du test :

Le test de χ^2 permet de tester la liaison entre deux ou plusieurs distributions de caractères qualitatifs

C-Categories du test :

1-Comparaison d'une distribution observée et théorique

Il sert à comparer une distribution observée sur un échantillon à une distribution connue dans la population ou à une distribution théorique

Il s'agit d'un test statistique qui étudie l'écart entre la distribution théorique et la distribution observée

Exemple :

Dans une maternité, sur 100 naissances, on a observé 44 garçons et 56 filles ; cette observation est- elle compatible avec la statistique nationale donnant les proportions de naissances de garçons et de filles de respectivement 53% et 47% ?

Tableau de contingence:

	Garçon	Fille	Total
Statistique nationale	0.53	0.47	1
Effectif théorique (Ci/Ti)	53	47	100
Effectif observé	44	56	100

Démarche à suivre :

1- **Ho** : la distribution par sexe observée à la maternité est conforme à la distribution nationale

2-choix du test → un test de χ^2

3-vérification des conditions d'application :

Tous les $T_i \geq 5$ (53 et 47 sont > 5)

4-Calcul de la statistique (somme des écarts) :

$$\chi^2 = \sum (C_i - O_i)^2$$

Ci

$$\chi^2 = (53-44)^2 / 53 + (47-56)^2 / 47 = 3.25$$

5-on fixe le seuil de signification au risque $\alpha = 5\%$

6-on compare le χ^2 calculé au χ^2 de la table au risque $\alpha = 5\%$ et avec un nombre de degré de liberté $ddl = (l - 1)(c - 1)$

(k = nombre de modalités de la variable étudiée) $\rightarrow k = 2-1=1$

7-conclusion :

χ^2 calculé < χ^2 de la table (3.25 < 3.84)

$\rightarrow H_0$ n'est pas rejetée au risque $\alpha = 5\%$

\rightarrow La distribution observée est **conforme** à la répartition nationale par sexe des naissances

2-Comparaison entre plusieurs distributions observées :

Il sert à comparer deux ou plusieurs distributions observées sur des échantillons

Exemple :

Deux médicaments (A et B) ont été testés sur deux groupes de malades.

A l'issue de l'essai, on a observé les résultats suivants :

	Disparition des symptômes	Persistance des symptômes	aggravation	Réaction secondaire	Total
A	100	40	20	30	190
B	220	80	70	40	410
Total	320	120	90	70	600

Peut-on dire que ces deux traitements ont les mêmes effets ?

Correction :

1-H0 :

les deux traitements ont les mêmes effets \rightarrow

il n'y a pas de différence entre les deux distributions \rightarrow les différences observées résulteraient des seules fluctuations d'échantillonnage.

2-tableau de contingence :

	Disparition des symptômes	Persistance des symptômes	aggravation	Réaction secondaire	Total
A	100 / 101.33	40/ 38	20/ 28.5	30/ 22.16	190
B	220/ 218.66	80/ 82	70/ 61.5	40/ 47.83	410
Total	320	120	90	70	600

3-choix du test \rightarrow un χ^2

4-vérification des conditions d'application :

Tous les $T_i \geq 5$

Calcul des T_i :

Ti = (total de la ligne × total de la colonne) / total general

(101.33- 218.66- 38- 82 - 28.5- 61.5- 22.16 et 47.83) sont tous > 5

5-Calcul de la statistique (somme des écarts) :

$$\chi^2 = \sum \frac{(C_i - O_i)^2}{C_i}$$

$$\chi^2 = (100 - 101,33)^2 / 101,33 + (220 - 218,66)^2 / 218,66 + (40 - 38)^2 / 38 + \\ (80 - 82)^2 / 82 + (20 - 28,5)^2 / 28,5 + (70 - 61,5)^2 / 61,5 + \\ (30 - 22,16)^2 / 22,16 + (40 - 47,83)^2 / 47,83$$

$$\chi^2 = 0,0175 + 0,0082 + 0,105 + 0,048 + 2,535 + 1,174 + 2,773 + 1,28$$

$$\chi^2 = 7,94$$

6-on fixe le seuil de signification au risque $\alpha = 5\%$

7-on compare le χ^2 calculé au χ^2 de la table au risque $\alpha = 5\%$ et avec un nombre de degré de liberté

$$ddl = (C-1) \times (L-1)$$

C= nombre de colonnes

L= nombre de lignes

$$ddl = (4-1) \times (2-1) = 3$$

8-conclusion:

χ^2 calculé > χ^2 de la table (7.94 > 7.81)

→ Ho est rejetée

→ La différence est significative au risque $\alpha = 5\%$

→ les deux distributions **ne sont pas homogènes**. Ou

Les deux traitements ont des effets différents au risque de 5%.

$$\text{Avec } \alpha = 5 \% \text{ et d.d.l} = 3 \quad \chi^2 = 7,815$$

3-Etude de la liaison entre les distributions de deux variables :

Il sert à étudier sur un même échantillon la liaison entre les distributions de deux variables

Exemple :

Un épisode d'intoxication alimentaire collective (TIAC) est survenu parmi les stagiaires d'un atelier, Le docteur chargé de l'enquête a dressé le tableau suivant croisant l'information sur la consommation de glace au chocolat, l'un des desserts proposés au cours du dernier repas pris en commun par les stagiaires, et le statut malade / non malade.

1-

Tableau de contingence :

Ho : il n'y a pas de liaison entre la consommation de glace au chocolat et la survenue de la gastro- entérite.

	Malade	Sain	Total
Glace au chocolat	270 (206.5)	7 (70.5)	277
Pas de glace	26 (89.5)	94 (30.5)	120
Total	296	101	397

Le principe et le calcul du test sont identiques

χ^2 calculé > χ^2 de la table (253.7 > 3.84)

- Ho est rejetée
- La différence est significative au risque $\alpha = 5\%$
- Il existe une liaison statistique entre la consommation de glace au chocolat et la survenue de la gastro-entérite au risque de 5%.

Table de la loi du X^2

Cette table donne la probabilité p d'observer une valeur supérieure à une valeur de X^2 en fonction du nombre de degrés de liberté (dl)

dl / p	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.001
1	0.000	0.001	0.004	0.016	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.12	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.47	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.13
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.27	19.67	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.84	32.00	39.25
17	6.41	7.56	8.67	10.08	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.80	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.32
21	8.90	10.28	11.59	13.24	29.61	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.41	39.37	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	12.84	15.38	17.29	35.56	38.88	41.92	45.64	54.05
27	12.88	13.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

Comparaison de moyennes
(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

COMPARAISON DE DEUX MOYENNES

1- INTRODUCTION :

Deux moyennes d'un caractère quantitatif sont-elles réellement différentes ? Une question qui préoccupe souvent l'expérimentateur.

L'examen plus fin montre que deux situations peuvent se présenter :

- ⇒ **Comparaison d'une moyenne observée à une moyenne théorique** et dont l'objectif est de :
 - a- Vérifier l'hypothèse de l'appartenance de l'échantillon étudié à la population mère.
 - b- Confirmer ou infirmer l'hypothèse de la compatibilité des résultats observés avec les résultats théoriques, appelée encore l'analyse de la conformité.

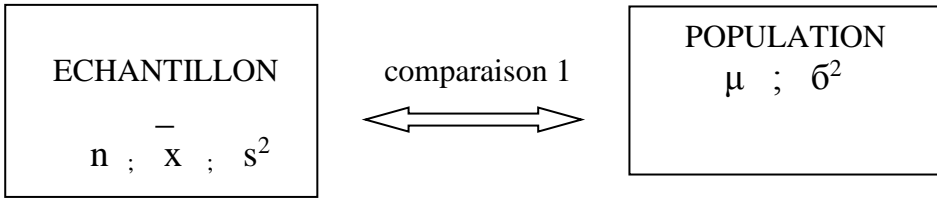
- ⇒ **Comparaison de deux moyennes observées** dont l'objectif est de :
 - a- Vérifier l'hypothèse de l'homogénéité du caractère dans les deux échantillons.
 - b- Vérifier l'hypothèse d'appartenance des deux échantillons à la population mère.

2- PRINCIPE :

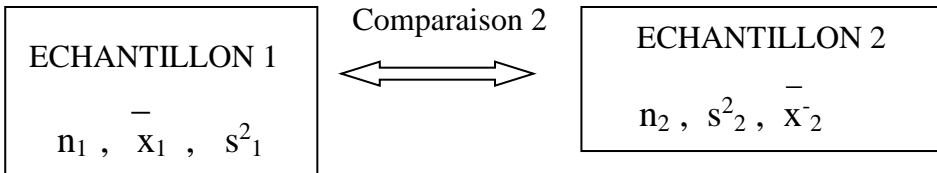
Soit deux échantillons 1 et 2, de paramètres respectifs n_1 , s^2_1 , \bar{x}_1 et n_2 , s^2_2 , \bar{x}_2 dont on veut vérifier l'appartenance à la population mère de moyenne μ et d'écart type σ .

Comparaison de moyennes

Le premier cas



Le deuxième cas



Le principe du test dans les deux situations I et II, est l'analyse de l'écart numérique apparent entre les deux moyennes, en appréciant jusqu'à *quelle limite* la différence des paramètres comparés est imputable aux fluctuations fortuites de l'échantillonnage.

**Hypothèse émise sur la population.
Loi suivie par la variable dans la population.**

a) Comparaison d'une moyenne observée à une moyenne théorique

Dans une population donnée, on extrait tous les échantillons de tailles n , dans lesquels on calcule une nouvelle variable : la moyenne \bar{x}_i , prenant les valeurs:

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots$$

La distribution de l'échantillonnage des moyennes montre que la moyenne des moyennes $\bar{\bar{x}}$ et l'écart type des moyennes \bar{x}_i sont respectivement :

$$E(\bar{x}_i) = \mu$$

Comparaison de moyennes

$$\sigma(\bar{x}_i) = \sigma / \sqrt{n}$$

Quand la taille des échantillons est relativement élevée (pratiquement $n \geq 30$), la distribution des moyennes \bar{x}_i suit une loi normale de moyenne $E(\bar{x}_i) = \mu$ et d'écart type $\sigma(\bar{x}_i) = \sigma / \sqrt{n}$, d'après le Théorème Central Limite.

Selon la théorie de l'échantillonnage, si la variable x_i suit une loi normale $N(\mu, \sigma)$ dans la population, alors la distribution des moyennes \bar{x}_i suit une loi normale $N(\mu, \sigma / \sqrt{n})$ quelle que soit la valeur de n .

Résumons ces conséquences fondamentales :

- Quelle que soit la loi suivie par la variable x_i dans la population, si $n \geq 30$, alors la variable « moyenne » \bar{x}_i est gaussienne et le test de l'écart réduit est applicable.
- Si la variable x_i obéit à une loi normale dans la population, alors \bar{x}_i est gaussienne quelque soit la taille n de l'échantillon, l'application du test de l'écart réduit est valide.

Calculons dans ces conditions la variable centrée réduite de la variable \bar{x}_i :

Variable – moyenne	
$t =$	_____
	Ecart type

$$\text{alors } t = (\bar{x}_i - \mu) / (\sigma / \sqrt{n}),$$

qui n'est que la variable testée dans le test de comparaison d'une moyenne observée avec une moyenne théorique. En tenant compte de la valeur absolue, on obtient :

$\varepsilon =$	$\frac{ \bar{X} - \mu }{\sigma / \sqrt{n}}$
-----------------	---

Dans le cas où l'écart type σ de la population n'est pas connu, on peut l'estimer par les paramètres de l'échantillon, soit : $\hat{\sigma} = S / \sqrt{n-1}$, qui est un estimateur sans biais, selon la théorie de l'estimation.

Alors la variable testée devient :

$$\varepsilon = \frac{\bar{X} - \mu}{s / \sqrt{n-1}}$$

qui est appelée à être comparée à la valeur limite (critique) tirée de la table de l'écart réduit.

b- comparaison de deux moyennes observées ($\bar{X}_1 - \bar{X}_2$)

Selon la théorie de l'échantillonnage, la distribution des différences des moyennes ($\bar{X}_1 - \bar{X}_2$) considérée comme variable, a comme moyenne :

$$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2$$

et comme écart type $\sigma(\bar{x}_1 - \bar{x}_2) = \sqrt{\sigma^2(\bar{x}_1) + \sigma^2(\bar{x}_2)}$

Alors $\sigma(\bar{x}_1 - \bar{x}_2) = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$

Calculons la variable centrée réduite de la variable ($\bar{x}_1 - \bar{x}_2$) :

$$t = \frac{\text{Variable} - \text{moyenne}}{\text{Écart type}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

c'est la variable testée utilisée dans le test de comparaison de deux moyennes observées et qui sous l'hypothèse nulle H_0 , suppose $\mu_1 = \mu_2$ alors $\mu_1 - \mu_2 = 0$, on obtient :

$$\varepsilon = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

En tenant compte de la valeur absolue $I \in I$, la variable testée devient finalement :

$$\varepsilon = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Si les variances σ_1^2 et σ_2^2 ne sont pas connues, l'estimateur non biaisé donne :

$$\frac{\sigma_1^2}{n_1} = \frac{S_1^2}{n_1 - 1} \quad \text{et} \quad \frac{\sigma_2^2}{n_2} = \frac{S_2^2}{n_2 - 1}$$

par conséquent :

$$\varepsilon = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}}$$

Il s'agit de la variable testée qui est comparée, dans la conclusion du test, avec la valeur limite critique ε_α tirée de la table de l'écart réduit.

2.3 - PROCEDURE :

a- Comparaison d'une moyenne observée avec une moyenne théorique :

- 1) - Détermination du risque α
- 2) - Condition : $n \geq 30$
- 3) - Construction de H_0 et H_1
- 4) - Calcul de la variable testée :

$$\varepsilon = \frac{I\bar{X} - \mu I}{\frac{s}{\sqrt{n-1}}}$$

- 5) - Conclusion :

Comparaison de moyennes

- Si $\varepsilon \geq \varepsilon_\alpha \Rightarrow$ la différence est significative au risque α , H_0 est rejetée.
- Si $\varepsilon < \varepsilon_\alpha \Rightarrow$ la différence n'est pas significative au risque α , H_0 est retenue.

b- Comparaison de deux moyenne observées :

- 1) - Détermination du risque α
- 2) - Condition : $n_1 \geq 30$ et $n_2 \geq 30$
- 3) - Construction de H_0 et H_1
- 4) - Calcul de la variable testée :

$$\varepsilon = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- 5) - Conclusion :
 - Si $\varepsilon \geq \varepsilon_\alpha \Rightarrow$ la différence est significative au risque α , H_0 est rejetée.
 - Si $\varepsilon < \varepsilon_\alpha \Rightarrow$ la différence n'est pas significative au risque α , H_0 est retenue.

2.4- EXEMPLES :

a- Comparaison d'une moyenne observée à une moyenne théorique

Dans une région, l'étude du poids des nouveau-nés sur un échantillon de taille 103 enfants a révélé les résultats suivants : une moyenne de 3,45 kg avec un écart type de 0,30 kg.

On admet que le poids moyen des nouveau-nés dans la population générale est égal à 3,50kg.

Au taux de sécurité de 95%, peut on dire que les données collectées dans cette région sont représentatives de la population générale ?

Solution :

Il s'agit d'un caractère quantitatif où on compare une moyenne observée avec une moyenne théorique : on applique le test de l'écart réduit dont les étapes sont :

- 1) $\alpha = 5\% \Rightarrow \varepsilon_\alpha = 1,96$
- 2) Condition : $n = 103 > 30$
- 3) H_0 : on peut dire que les données de l'échantillon sont représentatives de celles de la population quant au poids des nouveau-nés
- 4) La variable testée :

$$\varepsilon = \frac{\overline{IX} - \mu I}{\frac{s}{\sqrt{n-1}}} = \frac{I 3,45 - 3,50 I}{\frac{0,30}{\sqrt{103-1}}}$$

- 5) Conclusion :

$\varepsilon < \varepsilon_\alpha$ ($1,68 < 1,96$) \Rightarrow la différence n'est pas significative au risque de 5%, H_0 est retenue.

b- Comparaison de deux moyennes observées

On a relevé des taux de glycémie en gramme par litre de plasma chez deux groupes de sujets et à jeun depuis 4 heures.

Les résultats sont :

Sur le premier échantillon de 100 sujets on a obtenu comme moyenne 0,97 g/l avec un écart type de 0,28 g/l.

Sur le deuxième échantillon de 120 sujets on a trouvé 1,35 g/l comme moyenne et 0,30 g/l comme écart type .

Au risque de 5 % , tester l'homogénéité du taux de glycémie dans les deux groupes.

SOLUTION

$$n_1 = 100$$

$$\overline{X}_1 = 0,97 \text{ g/l}$$

$$S_1 = 0,28 \text{ g/l}$$

et

$$n_2 = 120$$

Comparaison de moyennes

$$\bar{X}_2 = 1,35 \text{ g/l}$$

$$S_2 = 0,30 \text{ g/l}$$

Procédure :

- 1) $\alpha = 5 \%$ $\Rightarrow \varepsilon_\alpha = 1,96$
- 2) Validité $n_1 = 100 > 30$ et $n_2 = 120 > 30$
- 1) H_0 : Le taux de glycémie est homogène dans les deux groupes.
- 2) calcul de la variable testée :

$$\varepsilon = \frac{I \bar{X}_1 - \bar{X}_2 I}{\sqrt{S_1^2/(n_1-1) + S_2^2/(n_2-1)}}$$

$$\varepsilon = \frac{I 0,97 - 1,35 I}{\sqrt{0,28^2/99 + 0,30^2/119}}$$

$$\varepsilon = 9,6$$

3) Conclusion:

$\varepsilon > \varepsilon_\alpha$ ($9,6 > 1,96$) \Rightarrow la différence est significative au risque de 5 % donc H_0 est rejetée.

Le taux de glycémie n'est pas homogène dans les deux groupes.

Chapitre 23

CAS DES PETITS ECHANTILLONS (Comparaison de moyennes)

1 - INTRODUCTION :

Quand la taille de l'échantillon étudié est inférieure à 30 (comparaison de \bar{X} avec μ), ou au moins un des deux échantillons (comparaison de \bar{X}_1 avec \bar{X}_2) est petit ($n < 30$), la distribution des variables :

$$t = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} \quad \text{et} \quad t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

suit plutôt la loi dite de Student. Les tests précédents de l'écart réduit relatifs aux moyennes ($n \geq 30$) ne sont plus applicables. La variable testée de Student est comparée à la valeur limite critique tirée de la table à double entrée (risque, degré de liberté) de Student. La loi de Student est une loi de probabilité symétrique et qui se confond avec la loi normale centrée réduite quand la taille de l'échantillon devient importante.

2 – PROCEDURE :

2.1 *Test de Student : comparaison d'une moyenne observée avec une moyenne théorique.*

- 1) Détermination du risque α
- 2) Condition $n < 30$
- 3) Construction de H_0 et H_1
- 4) Degré de liberté d.d.l = $v = n - 1$ et $t_{\alpha v}$ de la table

$$5) \text{ Variable testée : } t = \frac{|\bar{X} - \mu|}{\frac{s}{\sqrt{n-1}}}$$

- 6) Conclusion

Si $t \geq t_{\alpha} \Rightarrow$ la différence est significative au risque α , H_0 est rejetée.

Si $t < t_{\alpha} \Rightarrow$ la différence n'est pas significative au risque α , H_0 est retenue.

2.2– *Test de Student: Comparaison de deux moyennes observées.*

Ce test est d'une large utilisation, en particulier quand la variable est gaussienne dans la population et la taille des échantillons ne dépassant pas la trentaine. Pour comparer deux moyennes observées, les étapes du test de Student à suivre sont :

- 1) le risque α
- 2) Condition $n_1 < 30$ et / ou $n_2 < 30$
- 3) Hypothèse H_0

4) Calcul de la variance commune :

$$Sc^2 = \frac{n_1.S_1^2 + n_2.S_2^2}{(n_1-1) + (n_2 - 1)} = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

Sous l'hypothèse H_0 , les 2 échantillons sont supposés issus de la même population dont la variance estimée est donnée en fonction des paramètres des deux échantillons.

- 5) d.d.l = $v = n_1 + n_2 - 2$
- 6) Détermination de $t_{\alpha,v}$ de la table
- 7) Variable de Student : t

$$\frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{Sc^2}{n_1} + \frac{Sc^2}{n_2}}}$$

8) Conclusion :

Si $t \geq t_{\alpha} \Rightarrow$ la différence est significative au risque α , H_0 est rejetée.

Si $t < t_{\alpha} \Rightarrow$ la différence n'est pas significative au risque α , H_0 est retenue.

3 – EXEMPLES :

3.1– *Comparaison d'une moyenne observée à une moyenne théorique*

Dans une population, le taux moyen d'uricémie est de 62 mg/l de plasma. On extrait un échantillon de 20 sujets au hasard. Les résultats observés de l'uricémie sur cet échantillon sont :

Taux moyen = 59 mg/l

Ecart type = 12 mg/l

Au risque de 5%, peut – on penser que les résultats observés de l'uricémie sont compatibles avec ceux de la population ?

Solution :

Il s'agit de comparer une valeur observée avec une valeur théorique dans le cas d'un caractère quantitatif dont la taille de l'échantillon est inférieure à 30 : test de Student (petits échantillons).

- 1) $\alpha = 5\%$
- 2) Condition $n = 20 < 30$
- 3) H_0 : on peut penser que les résultats observés sont compatibles avec ceux de la population.
- 4) Le degré de liberté : d.d.l = $v = n - 1 = 20 - 1 = 19$
Valeur critique $t_{\alpha,v} = t_{5\%,19} = 2,093$
- 5) Variable testée :

$$\frac{|\bar{X} - \mu|}{\frac{s}{\sqrt{n-1}}} = \frac{|59 - 62|}{\frac{12}{\sqrt{20-1}}}$$

$$t = \quad = \quad = 1,09$$

6) Conclusion

$t < t_{\alpha}$ (1,09 < 2,093) \Rightarrow la différence n'est pas significative au risque de 5 %, H_0 est retenue.

3.2 Comparaison de deux moyennes observées ($n_1 < 30$ et/ou $n_2 < 30$)

On a mesuré sur un échantillon de 12 élèves d'une école et d'un même niveau, la variable « poids » en kg. Les mêmes mesures ont été effectuées sur un échantillon de 10 élèves d'une autre école et du même niveau. Les résultats sont :

Ecole1(poids en kg): 30, 30, 32, 33, 33, 36, 36, 38, 39, 39, 40,40.

Ecole2 (poids en kg) : 28, 28, 29, 29, 29, 31, 31, 33, 33, 38.

Tester l'homogénéité de la variable « poids » dans les deux écoles, en considérant un taux de sécurité de 95%.

Solution :

Comparaison de deux mesures observées (moyennes) dans le cas des petits échantillons \Rightarrow test de Student.

- 1) $\alpha = 5\%$
- 2) Condition $n_1 = 12 < 30$ et $n_2 = 10 < 30$
- 3) Hypothèse nulle H_0 :

La variable « poids » est homogène dans les deux échantillons ou bien la différence entre les poids moyens observés n'est pas significative.

- 4) Calcul de la variance commune :

$$S_c^2 = \frac{n_1.S_1^2 + n_2.S_2^2}{(n_1-1) + (n_2 - 1)} = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$\Rightarrow S_c^2 = \frac{12. 3,62 + 10.2,95}{12 + 10 - 2}$$

$$\Rightarrow S_c^2 = 3,65$$

- 5) Degré de liberté = d.d.l = $v = n_1 + n_2 - 2 = 12+10-2 = 20$

- 6) Valeur critique de Student

$$t_{\alpha,v} = t_{5\%, 20} = 2,086$$

- 7) Variable testée :

Comparaison de moyennes

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{Sc^2}{n_1} + \frac{Sc^2}{n_2}}}$$

$$t = \frac{|35,5 - 30,9|}{\sqrt{\frac{3,65}{12} + \frac{3,65}{10}}}$$

$$t = 7,62$$

8) Conclusion

$t > t_\alpha$ (7,62 > 2,086) \Rightarrow la différence est significative au risque de 5 %, H_0 est rejetée.

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

**Analyse de la variance
(Cours à l'usage des étudiants en sciences médicales)**

Dr BENILHA.S

ANALYSE DE LA VARIANCE :

1-Position du problème :

Comparaison de deux moyennes

→ test de l'écart réduit ou test T

Comparaison de plusieurs moyennes ?

Solution :

a-Comparaison de moyennes deux à deux :

Inconvénients : - méthode longue

-Vue d'ensemble flou

b-Analyse de la variance (terme souvent abrégé par le terme **ANOVA** : *ANalysis Of VAriance*) : bon reflet de dispersion autour de la moyenne.

➤ **Test approprié** : test de **Fisher-Snedecor**

2-Principe de test d'ANOVA :

Il repose sur la **décomposition de la variation de la variable X**

a) La dispersion entre les classes (variation inter-groupe)

(D inter) Chaque classe est caractérisée par sa moyenne (X_1, \dots, X_k) qui s'écarte plus ou moins de la moyenne générale (\bar{X}).

b) La dispersion à l'intérieur de chaque classe (intra-groupe ou celle due au hasard)(D intra)

On teste le rapport de deux variances :

1- variance **inter-groupes**(variance factorielle **Vf**)

2- variance **intra-groupes**(variance résiduelle **Vr**)

-Ho: teste donc l'hypothèse de l'homogénéité des k moyennes.(la différence entre les moyennes n'est pas significative)

-On dit aussi que le facteur (à partir duquel on a construit les k groupes) n'a pas d'influence sur la variable X

Formule pratique $D_{inter} = \frac{n_1(\bar{X}_1)^2 + n_2(\bar{X}_2)^2 + \dots + n_k(\bar{X}_k)^2}{N} - N(\bar{X})^2$

\bar{X}_i = les moyennes respectives des K échantillons

\bar{X} = moyenne générale = $\frac{\sum X_i}{n_1+n_2+n_3+\dots+n_k}$

$N = n_1 + n_2 + n_3 + \dots + n_k$

1-variance factorielle Vf

Vf = Dinter / k-1

K: nombre d'échantillons

Dinter: Dispersion inter-groupes: analyse la variabilité **entre** les groupes

$$D_{inter} = n_1 (X_1 - \bar{X})^2 + n_2 (X_2 - \bar{X})^2 + n_3 (X_3 - \bar{X})^2 + \dots + n_k (X_k - \bar{X})^2$$

2- variance résiduelle Vr :

$$V_r = D_{intra} / N - k$$

N= somme des tailles de tous les échantillons

Dintra: Dispersion intra-groupes :analyse la variabilité **à l'intérieur** des groupes

$$D_{intra} = \sum (X_{i1} - \bar{X}_1)^2 + \sum (X_{i2} - \bar{X}_2)^2 + \sum (X_{i3} - \bar{X}_3)^2 + \dots + \sum (X_{ik} - \bar{X}_k)^2$$

Formule pratique :

$$D_{intra} = \sum X^2_i - (n_1 \cdot \bar{X}_1^2 + n_2 \cdot \bar{X}_2^2 + \dots + n_k \cdot \bar{X}_k^2)$$

$$\sum x^2_i = \sum x_i^2_1 + \sum x_i^2_2 + \sum x_i^2_3 + \dots + \sum x_i^2_k$$

4-Procédure du test

- 1) Détermination du risque α
- 2) Formulation des hypothèses H_0 et H_1
- 3) vérification des conditions d'application
- 4) Calcul de la dispersion inter –groupe
- 5) Calcul de la dispersion intra –groupe
- 6) Calcul de la variance factorielle V_f
- 7) Calcul de la variance résiduelle V_r

Conditions d'utilisation

1- L'ensemble des N individus est réparti au hasard (randomisation).

2- Normalité de la distribution des mesures.

3- L'égalité des variances.

Exercice :

5 milieux de culture de BCG (K)

10 tubes par milieu de culture ($n_1 = n_2 = \dots = 10$; $N = 50$)

on observe le nombre de colonies par tubes (x_i)

Ces milieux sont-ils équivalents?

Milieu de culture	1	2	3	4	5
	10	11	7	12	7
	12	18	14	9	6
	8	12	10	11	10
	10	15	11	10	7
	6	13	9	7	7
	13	8	10	8	5
	9	15	9	13	6
	10	16	11	14	7
	8	9	7	10	9
	9	3	9	11	6
Moyenne	9.5	13.0	9.7	10.5	7.0

Corrigé :

1) $\alpha = 5\%$

2) H_0 : les milieux sont équivalents

3) $D_{inter} = n_1(x_1)^2 + n_2(x_2)^2 + \dots - N(\bar{x})^2 = 10(9.5)^2 + 10(13)^2 + \dots - 50(9.94)^2 = 185.7$

4) $D_{intra} = \sum x_i^2 - (n_1 \cdot x_1^2 + n_2 \cdot x_2^2 + \dots + n_k \cdot x_k^2) = 5083 - (10 \cdot 9.5^2 + 10 \cdot 13^2 + \dots + 10 \cdot 7^2) = 225.1$

5) $V_f = 185.7 / 5 - 1 = 46.4$

6) $V_r = 225.1 / 50 - 5 = 5$

7) conclusion : $V_f / V_r = 46.4 / 5 = 9.3$

8) F de la table = **2.56** (ddl1=4 et ddl2=45)

le rapport est supérieur à F critique ; H_0 est rejetée.

Lois de Fisher–Snédécour (0, 05)

4-Comparaison de deux variances

Intérêt :

- 1) comparer deux moyennes dans le cas des petits échantillon (la variable étudiée doit être normale et les variance des deux populations ne doivent pas être significativement différentes)
- 2) comparer la précision de deux méthodes

Principe:

Comparaison du rapport $F = S^2_A / S^2_B$ (S^2_A est supérieure à S^2_B) et F de la table de Fisher.

F critique: intersection de ddl=nA-1 et ddl=nB-1

Procédure:

Les étapes à suivre sont:

- 1) Détermination du risque α
- 2) construction de l'hypothèse nulle
- 3) Calcul de la variance testée $F = S^2_A / S^2_B$
- 4) Détermination de la valeur critique F de la table de Fischer, en fonction du risque α et les degrés de liberté $V_1 = n_1 - 1$ et $V_2 = n_2 - 1$

5) Conclusion :

5) $F_c \geq F_t$ la différence est significative au risque α et H_0 est rejetée

6) $F_c < F_t$ la différence n'est pas significative au risque α et H_0 est retenue

Exemple:

	A	B
	37	38
	36	39
	37	38
	38	37
	37	36
		39
		39
		38
		37
Moyenne	37	38
Variance	0.5	1.11

F calculé = $1,11 / 0,5 = 2,22$

F table = **6.04** (intersection de 8 et 4)

$F_c < F_t$ pas de différence significative entre les deux variances

Lois de Fisher-Snedecor (0, 05)

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10
1	161	200	216	225	230	234	239	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35

Lois de Fisher–Snedecor (0, 05)

$\nu_2 \backslash \nu_1$	1	2	3	4
1	161	200	216	225
2	18,5	19,0	19,2	19,2
3	10,1	9,55	9,28	9,12
4	7,71	6,94	6,59	6,39
5	6,61	5,79	5,41	5,19
6	5,99	5,14	4,76	4,53
7	5,59	4,74	4,35	4,12
8	5,32	4,46	4,07	3,84
9	5,12	4,26	3,86	3,63
10	4,96	4,10	3,71	3,48
11	4,84	3,98	3,59	3,36
12	4,75	3,89	3,49	3,26
13	4,67	3,81	3,41	3,18
14	4,60	3,74	3,34	3,11
15	4,54	3,68	3,29	3,06
16	4,49	3,63	3,24	3,01
17	4,45	3,59	3,20	2,96
18	4,41	3,55	3,16	2,93
19	4,38	3,52	3,13	2,90
20	4,35	3,49	3,10	2,87
22	4,30	3,44	3,05	2,82
24	4,26	3,40	3,01	2,78
26	4,23	3,37	2,98	2,74
28	4,20	3,34	2,95	2,71
30	4,17	3,32	2,92	2,69
40	4,08	3,23	2,84	2,61
50	4,03	3,18	2,79	2,56
60	4,00	3,15	2,76	2,53

Université Saad Dahlab Blida1
Faculté de Médecine
Département de Pharmacie
5^{ème} année de Pharmacie
Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

LES INDICATEURS DE MORBIDITE
(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

LES INDICATEURS DE MORBIDITE

I/ INTRODUCTION :

Les concepts d'indicateurs ont été utilisés au départ en économie, avec les notions d'indicateurs sociaux qui étaient destinés à « décrire et à mesurer les résultats pour l'action à entreprendre ».

L'appropriation du terme d'indicateur de santé par les professionnels de santé (Francophones) s'est faite à partir des années 1970 dans la Revue d'épidémiologie et de santé publique.

les indicateurs (1981) comme « **des variables qui aident à mesurer, directement ou indirectement, les changements dans la situation sanitaire et à apprécier dans quelle mesure les objectifs et cibles d'un programme sont atteints** » OMS.

Les indicateurs de santé sont donc des paramètres de mesure qui reflètent diverses composantes de l'état de santé. Ils ont pour intérêt de décrire l'état de santé d'une population, planifier et évaluer les actions de santé et sont perçus comme des instruments de la surveillance épidémiologique.

Selon le phénomène observé, maladie ou décès, on distingue :

- Les indicateurs de morbidité qui décrivent la fréquence des maladies
- Les indicateurs de mortalité qui décrivent la fréquence des décès

II/ LES MESURES DE BASE :

Toute mesure en épidémiologie doit être précédée par une définition des deux termes du rapport (N/D) :

- Définition d'un cas au numérateur (N)
- Définition de la population d'étude au dénominateur (D)

1-Proportion :

Dans une proportion, le numérateur est une part du dénominateur. Le (N) et le (D) sont donc de même nature ; $P = a / (a+b)$

Une proportion s'exprime sous forme d'un nombre compris entre 0 et 1, ou bien sous forme d'un pourcentage

Exemple : Dans une population de 7500 enfants de moins de cinq ans, on constate que 5300 sont correctement vaccinés contre la rougeole. La proportion d'enfants vaccinés est de $P = a / (a+b) = 5300/7500 = 0,707 = 70,7\%$. Cette proportion est communément appelée « couverture vaccinale ».

2-Ratio :

Un ratio représente le rapport entre les effectifs de deux classes d'une même variable. Le (N) et le (D) sont de même nature, mais exclusifs l'un de l'autre.

Un ratio s'exprime par un nombre sans unités

Exemple : Dans une population de 100 individus, on observe 49 hommes et 51 femmes. Le ratio h/f (sex ratio en anglais) = $49/51 = 0,96$ (0,96 homme pour une femme).

3-Indice :

Un indice est le rapport de deux effectifs qui sont de nature différente. On l'utilise surtout comme indicateurs de fonctionnement, notamment en économie de la santé.

Exemple : Dans un hôpital on dispose de 850 lits et 10 médecins. On calcule l'indice lits d'hôpital / médecin ou le nombre de lits par médecins = $850/10 = 85$ (85 lits par médecins).

4-Taux :

Un taux mesure la probabilité de survenue d'un événement au cours du temps.

Au (N) figurent des individus ayant subi un événement pendant une période de temps déterminé et au (D) figure l'ensemble des individus susceptibles de connaître l'événement pendant cette période (la population à risque).

III/ LES INDICATEURS DEMOGRAPHIQUES :

Les taux bruts sont calculés sur l'ensemble de la population et les taux spécifiques sont calculés pour une sous population définie par tranche d'âge, sexe ou par catégorie socio professionnelle.

1-Taux brut de natalité (TBN) :

TBN = $\frac{\text{Nombre de naissances vivantes d'une région donnée pendant une période donnée}}{\text{Population moyenne de la même région et durant la même période}} \times 1000$

2-Taux d'accroissement naturel :

Ce taux représente la différence entre le taux de natalité et le taux de mortalité

3-Taux de fécondité :

$\frac{\text{Nombre de naissances (filles et garçons) durant l'année} \times 1000}{\text{Nombre de femmes en âge de procréer (15-49) durant la même année}}$

4- Espérance de vie à la naissance : vie moyenne

C'est le nombre moyen d'années qu'un nouveau né peut espérer vivre

IV/ LES INDICATEURS DE MORBIDITE :

1- La prévalence :

La prévalence est un indice important largement utilisé pour déterminer les besoins médicaux et sociaux surtout dans le cas des maladies chroniques. Elle mesure la présence d'une maladie.

Il peut s'agir d'une prévalence instantanée (à un moment donné) ou de période (pendant une période de temps).

Taux de prévalence = $\frac{\text{Nombre total de cas d'une maladie (anciens et nouveaux cas) dans une région donnée et pendant une période donnée}}{\text{Population moyenne de la même région et de la même période}} \times 100$

2- L'incidence :

L'incidence est un indice important des besoins en soins préventifs, utile pour les maladies aiguës et chroniques. L'incidence permet d'évaluer l'efficacité des mesures de contrôle d'une maladie à caractère de masse et exprime donc la vitesse d'apparition d'une maladie dans une population.

Taux d'incidence = $\frac{\text{Nombre de nouveaux cas d'une maladie dans une région donnée et pendant une période donnée}}{\text{Population moyenne de la même région et de la même période}} \times 100$

Population moyenne de la même région et de la même période

Exemple :

Dans une population de 500 personnes relevés au cours de l'année 2002, 74 cas ont présenté un infarctus aigu du myocarde. Le taux d'incidence = $74/500 = 0,148 = 148$ cas pour 1000 personnes

2-1/ Taux d'attaque :

Ce taux désigne la proportion de sujets atteints au cours d'une période définie et déterminée. On l'utilise en règle à la suite d'une exposition de courte durée, par exemple lors d'une épidémie.

Exemple :

La cantine d'une école recevant 250 enfants a été le siège d'une toxi-infection alimentaire collective. 52 enfants ont présenté l'affection. Le taux d'attaque est de $52/250 = 208$ cas pour 1000.

2-2/ Densité de l'incidence (Id) :

Densité de

l'incidence = $\frac{\text{Nombre de nouveaux cas d'une maladie pendant une période de temps}}{\text{Nombre total de personnes années d'exposition au facteur étudié pendant la même période}}$

Exemple : **une** personne exposée pendant **3 ans** à un facteur étudié =
3 personnes années d'exposition

3- Relation entre le taux d'incidence et le taux de prévalence:

Le taux de prévalence (P) dépend à la fois du taux d'incidence et de la durée de la maladie

P = taux d'incidence \times durée moyenne de la maladie
(I) **(D)**

4- Létalité :

La létalité représente la proportion de malades atteints d'une affection qui décèdent de cette affection durant une période donnée. Cet indicateur témoigne de la gravité de la maladie et de la qualité des soins.

Un taux élevé de létalité implique un problème dans la prise en charge de la maladie.

Taux de létalité = $\frac{\text{Nombre de décès attribuable à une maladie donnée au cours d'une période donnée}}{\text{Nombre total de cas de cette maladie au cours de la même période}}$

5- Relation entre mortalité, incidence et létalité :

Létalité = $\frac{\text{Mortalité}}{\text{Incidence}}$

Mortalité = Incidence \times Létalité

V/ BIBLIOGRAPHIE :

Mesli MF, Bouziani M. Epidémiologie Objet et Méthodes. Laboratoire de biostatistiques. Faculté de médecine d'Oran. Mai 2007

Abdelouahab A. Cours à l'usage des étudiants en médecine .Les indicateurs de santé. Année universitaire 2005-2006

Dr.R.TALHI – Pr. MF. MESLI, Cours à l'usage des étudiants en sciences médicales LES INDICATEURS DE SANTE

Belateche F. Cours à l'usage des étudiants en médecine. Les indicateurs de santé. Année universitaire 2004-2005

Mesli MF, Mokhtari A. Biostatistique description et analyse des données en médecine et en biologie. Les éditions universitaires le fennec 2002

Ancelle T. Statistique Epidémiologie. Edition 2002

Bezzaoucha A. Epidémiologie et biostatistiques. Office des publications universitaires decembre 1996

Université Saad Dahlab Blida1
Faculté de Médecine
Département de pharmacie
5^{ème} année de Pharmacie
Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

LES INDICATEURS DE MORTALITE
(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

LES INDICATEURS DE MORTALITE

I/ INTRODUCTION :

Les concepts d'indicateurs ont été utilisés au départ en économie, avec les notions d'indicateurs sociaux qui étaient destinés à « décrire et à mesurer les résultats pour l'action à entreprendre ».

L'appropriation du terme d'indicateur de santé par les professionnels de santé (Francophones) s'est faite à partir des années 1970 dans la Revue d'épidémiologie et de santé publique.

Les indicateurs (1981) comme « **des variables qui aident à mesurer, directement ou indirectement, les changements dans la situation sanitaire et à apprécier dans quelle mesure les objectifs et cibles d'un programme sont atteints** » OMS.

Les indicateurs de santé sont donc des paramètres de mesure qui reflètent diverses composantes de l'état de santé. Ils ont pour intérêt de décrire l'état de santé d'une population, planifier et évaluer les actions de santé et sont perçus comme des instruments de la surveillance épidémiologique.

Selon le phénomène observé, maladie ou décès, on distingue :

- Les indicateurs de morbidité qui décrivent la fréquence des maladies
- Les indicateurs de mortalité qui décrivent la fréquence des décès

II-LES INDICATEURS DE MORTALITE :

1-Taux brut de mortalité (TBM) :

$$\text{TBM} = \frac{\text{Nombre de décès au cours de l'année}}{\text{Population totale au milieu de la même année}} \times 1000$$

2-Taux de mortalité spécifique par âge :

En raison de l'influence déterminante de l'âge sur la mortalité, il faut calculer le taux de mortalité pour chaque tranche d'âge.

$$\frac{\text{Nombre de personnes d'âge particulier décédées au cours de l'année}}{\text{Population de cet âge au milieu de la même année}} \times 1000$$

3-Taux de mortalité spécifique selon le sexe :

$$\frac{\text{Nombre de femmes ou d'hommes décédés au cours de l'année}}{\text{Population moyenne du même sexe au milieu de la même année}} \times 1000$$

4-Taux de mortalité proportionnelle (TMP) :

Ce taux représente la proportion de mortalité qui peut être attribuée à une cause spécifique

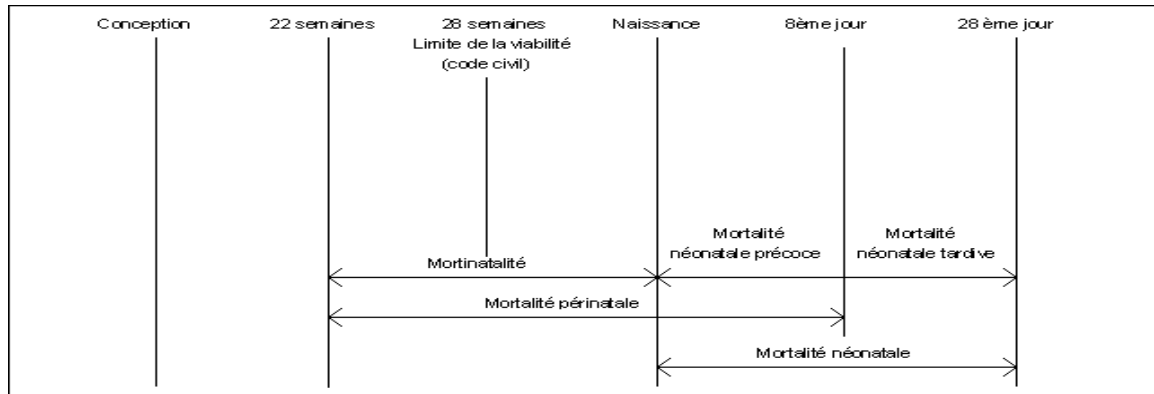
$$\text{TMP} = \frac{\text{Nombre de décès attribuable à une maladie donnée au cours de l'année}}{\text{Nombre de décès dans la population durant la même année}} \times 1000$$

5-Taux de mortalité infantile (TMI) :

Ce taux mesure surtout le niveau médical et social de la collectivité

$$\text{TMI} = \frac{\text{Nombre de décès d'enfants de moins de 1 ans (0 à 11 mois et 29 jours) durant l'année}}{\text{Nombre de naissances vivantes de la même année}} \times 1000$$

Ce taux de mortalité infantile peut se décomposer comme suit :



5-1/ Taux de mortalité néonatale précoce :

$$\frac{\text{Nombre de décès des nouveaux nés de 0 à 07 jours durant l'année}}{\text{Nombre de naissances vivantes de la même année}} \times 1000$$

5-2/ Taux de mortalité néonatale tardive :

$$\frac{\text{Nombre de décès des nouveaux nés de 8 à 28 jours durant l'année}}{\text{Nombre de naissances vivantes de la même année}} \times 1000$$

5-3/ Taux de mortalité post natale :

$$\frac{\text{Nombre de décès de nourrissons de 29 jours à 11 mois et 29 jours durant l'année}}{\text{Nombre de naissances vivantes de la même année}} \times 1000$$

6-Taux de mortalité maternelle :

$$\frac{\text{Nombre de décès des mères dus à l'accouchement, aux complications de la grossesse et suites de couches durant l'année}}{\text{Nombre de naissances vivantes de la même année}} \times 1000$$

V/ BIBLIOGRAPHIE :

Mesli MF, Bouziani M. Epidémiologie Objet et Méthodes. Laboratoire de biostatistiques. Faculté de médecine d'Oran. Mai 2007

Abdelouahab A. Cours à l'usage des étudiants en médecine .Les indicateurs de santé. Année universitaire 2005-2006

Dr.R.TALHI – Pr. MF. MESLI, Cours à l'usage des étudiants en sciences médicales LES INDICATEURS DE SANTE

Belateche F. Cours à l'usage des étudiants en médecine. Les indicateurs de santé. Année universitaire 2004-2005

Mesli MF, Mokhtari A. Biostatistique description et analyse des données en médecine et en biologie. Les éditions universitaires le fennec 2002

Ancelle T. Statistique Epidémiologie. Edition 2002

Bezzaoucha A. Epidémiologie et biostatistiques. Office des publications universitaires decembre 1996

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

5^{ème} année de Pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

Structure des études épidémiologiques
(Cours à l'usage des étudiants en sciences médicales)

Dr BENILHA.S

Structure des études épidémiologiques :

Structure des études descriptives et analytiques, Etude descriptive, transversale, cas-témoins, cohorte, Structure incomplète, échantillonnage et biais

1-Introduction :

L'épidémiologie :

Raisonnement et méthode appliqués à la description des phénomènes de santé, à l'explication de leur étiologie et à la recherche des méthodes d'intervention les plus efficaces.

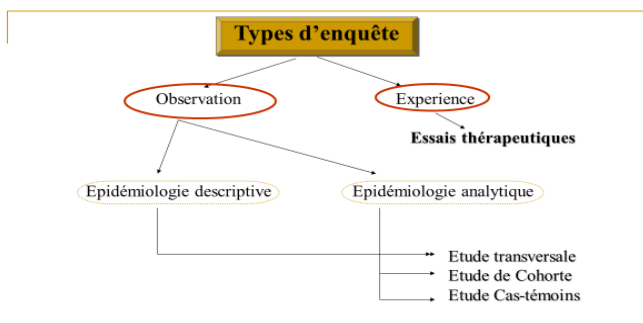
On distingue l'épidémiologie :

-**Descriptive** : Il y a plus de... chez les...

-**Explicative** : Etiologique (analytique) : Les gens exposés à... sont plus atteints par... que les non exposés.

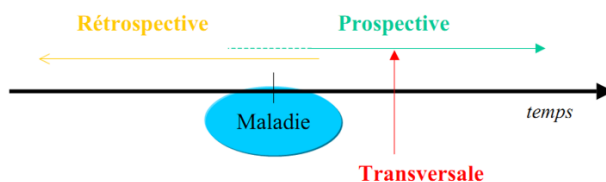
-**Evaluative** : (d'intervention) : Quand je donne... il y a moins de...

2-Structures des études épidémiologie :

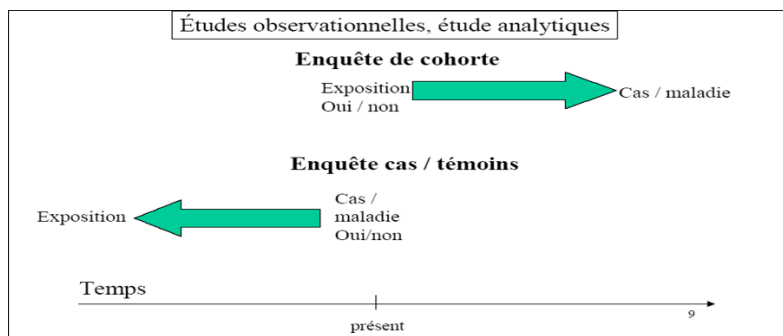


Types d'études selon la chronologie :

Classification chronologique :



Types d'études selon le recueil des données:



1- Enquêtes transversales (Observation)

- A un instant donné un enquêteur cherche à connaître le nombre de personnes qui présentent un caractère donné dans une population.
 - Les Enquêtes transversales permettent de mesurer le nombre de cas (prévalence) du problème ciblé mais aussi de recueillir les facteurs associés : caractéristiques socio-économiques, niveau d'éducation, conditions environnementales, style de vie...
 - Possibilité de déduction d'une hypothèse étiologique.
 - **Exemple :** Enquête de prévalence des infections Nosocomiales dans un hôpital – au Maroc cette prévalence était de 7,7 % dans les hôpitaux régionaux en 1994
- Avantages :**

- Facilité de mise en œuvre

Inconvénients :

- Biais de sélection (patients sortis de la cohorte car malades par ex)
- Relation temporelle exposition – maladie ?

2- Enquêtes analytiques :

- • Elle s'intéresse donc à la recherche des déterminants c.à.d. origines ou étiologies des phénomènes de santé.
- • Elle vise à comparer les fréquences d'une maladie dans différents groupes afin de mettre en évidence les « facteurs de risque »

Principe

- Etudier la relation entre un phénomène de santé (M) et des facteurs susceptibles de l'influencer (E)

– M = pathologie ou comportements de santé

– E = facteur biologique, comportemental, environnemental, pathologie... (exposition)

• Exemple :

– association entre contraception orale (E) et cancer du sein (M)

Buts :

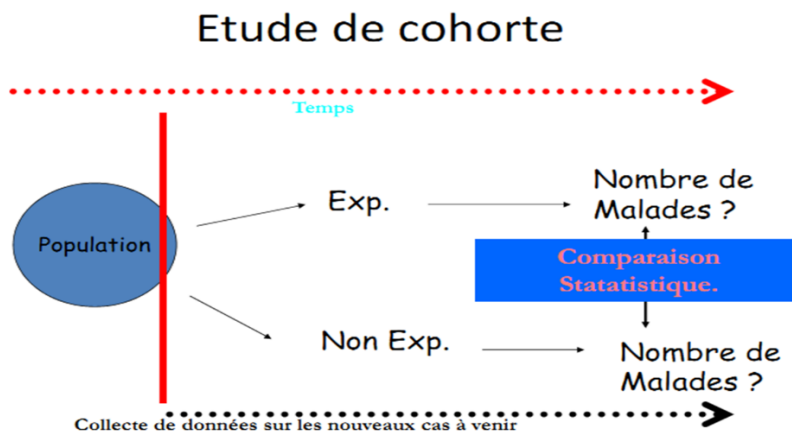
- mettre en évidence l'association entre une exposition (E) et une maladie (M)
- Les gens exposés à (E) sont plus atteints par (M) que les non exposés.
- Comparaison +++
- Notion de risque

a-Enquêtes de cohorte :

Cohorte = groupe de sujets suivi dans le temps

Les sujets sont sélectionnés sur la base de leur exposition au facteur étudié et sont suivis de manière prospective pour voir s'ils développent la maladie.

Schémas d'étude :



Avantages :

- ✓ Étude du risque de plusieurs maladies (Ex : exposition à l'efavirenz et risque de tératogénicité et de troubles psychiatriques).
- ✓ Expositions rares (Ex: exposition à l'atazanavir et apparition de lithiases)
- ✓ Séquence chronologique exposition / maladie plus facile à établir
- ✓ Données d'incidence et donc meilleure estimation du risque (RR)

Inconvénients

- Souvent coûteuses et longues
- Perdus de vue
- Peu performantes pour les maladies rares, ou à temps de latence long

B- Enquêtes cas-témoins

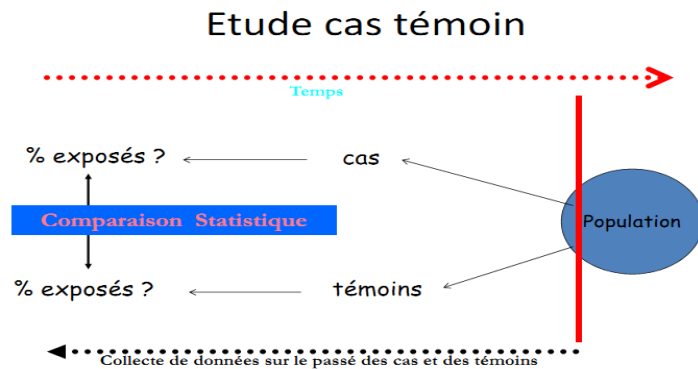
Définition

- Étude rétrospective
- Sélection de la population en fonction de la présence ou non du critère de la maladie
- Recherche rétrospective de l'exposition aux facteurs de risque

• Observation :

• L'enquêteur choisit les groupes étudiés sur la base de leur statut malade (cas) et non malade (témoins). Ce sont des études rétrospectives, c'est-à-dire que l'enquêteur va chercher dans le passé des individus des deux groupes l'exposition au facteur étudié.

Schémas d'étude :



Avantages :

• Intérêt

- Résultats rapides
- Coût faible

• Indications

- Maladies dont les périodes de latence sont longues
ex: cancer du poumon et tabac
- Maladies rares
- Étude de plusieurs facteurs de risque pour une seule maladie
ex: cancer de l'œsophage et exposition au tabac et à l'alcool

Inconvénients :

Non adapté pour l'étude

- De plusieurs maladies
- Des expositions rares
- De la relation chronologique entre exposition et maladie
- Ne permettent pas un calcul direct des taux d'incidence de la maladie chez les patients exposés et non exposés
 - Estimation du risque par OR (Odds Ratio)
- Les risques de biais sont importants particulièrement ceux de sélection et de mémoire (recueil rétrospectif)

c- Essai clinique :

- Groupe avec et sans intervention
- Mesure de l'effet de l'intervention
- 2 types de conclusions:
 - jugement de signification (statistique)
 - jugement de causalité (qualité de l'essai)

Schémas d'étude :



- Groupes identiques
 - même type de patients
 - même stade de la maladie, etc.
- qui ne diffèrent que par le traitement appliqué
- Si, à la fin, il existe une différence, celle-ci n'est due qu'au traitement

Université Saad Dahlab Blida1

Faculté de Médecine

Département de pharmacie

Responsable de l'enseignement : Dr BENILHA.S



Module d'Epidémiologie et Méthodologie de la recherche

5^{ème} année de Pharmacie

L'essai thérapeutique contrôlé
(Cours à l'usage des étudiants en sciences médicales)

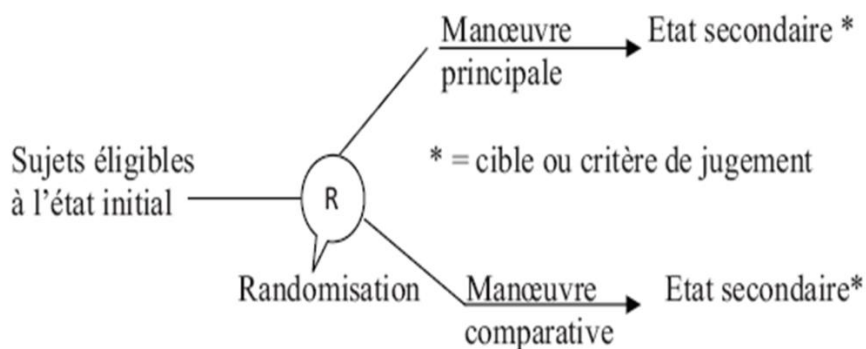
Dr BENILHA.S

L'essai thérapeutique contrôlé

I-Structure d'un essai thérapeutique contrôlé:

- ❑ Essai thérapeutique randomisé est un cas particulier de l'étude cohorte, il s'agit d'une expérience réalisée pour évaluer l'utilité de traitements, de vaccins ou de toutes autres manœuvres préventives ou thérapeutique.
- ❑ La répartition des sujets éligibles entre les différentes cohortes est réalisée par un tirage au sort, ce qui assure dès le départ la comparabilité de celles-ci.
- ❑ L'essai randomisé se distingue de l'étude cohorte par deux caractéristiques essentielles :
 - Les sujets éligibles de l'étude sont sélectionnés par tirage au sort pour former la cohorte principale et la cohorte de comparaison, on dit aussi que les sujets sont **randomisés** ;
 - L'application des manœuvres (traitements, vaccins...) est **contrôlée directement** par l'investigateur ou le chercheur en charge de l'étude.
 - Il pourrait même être mieux indiqué de comparer les examens en les réalisant chez un même sujet selon le dessin d'une étude randomisée en cross-over : chaque sujet étant son propre témoin.

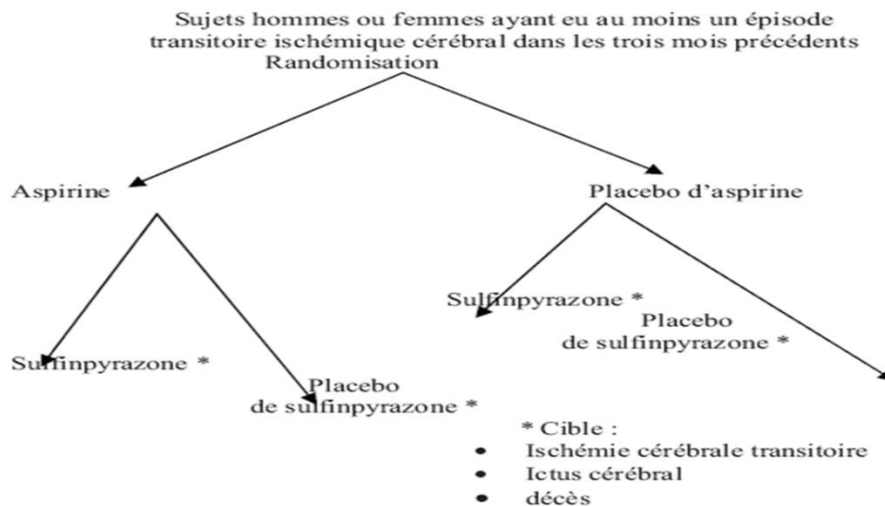
Le dessin (ou l'architecture) d'un essai randomisé peut être schématisé comme suit :



- Un sujet devrait avoir la même probabilité de subir l'une ou l'autre des manœuvres. Cela constitue le respect du principe d'ambivalence qui justifie le tirage au sort.
- La cible peut être mesurée par une densité d'incidence, un pourcentage ou une moyenne. La comparaison de ces paramètres s'effectue par un test statistique.

- Les résultats de l'essai, s'il a été correctement réalisé, peuvent être attribués à la manœuvre car le tirage au sort, qui répartit les différents facteurs pronostiques de manière identique dans les deux groupes, a constitué deux groupes comparables à l'état initial.

Le dessin de l'essai peut être schématisé comme suit :



Cet essai a envisagé d'étudier l'utilité de l'aspirine et du sulfinyprazone et de leur association dans la prévention des accidents cérébraux, :

L'essai a été réalisé chez des sujets ayant présenté des accidents ischémiques cérébraux transitoires.

Ainsi, les sujets éligibles de l'étude furent soumis à l'un des quatre régimes thérapeutiques suivants possibles : aspirine seul, sulfinyprazone seul, aspirine + sulfinyprazone, ou deux placebos.

L'organisation d'un essai selon un plan factoriel permet d'étudier l'interaction entre les deux traitements. Les résultats de l'essai ont été les suivants :

- l'aspirine a réduit le risque d'accidents ischémiques cérébraux récidivants, d'ictus et de décès ;
- le sulfinyprazone n'a pas d'utilité.

II-Substituts au tirage au sort :

Des méthodes proposées pour remplacer le tirage au sort, sont incorrectes car elles introduisent un biais susceptible d'entraver la validité des résultats de l'étude. Ces méthodes sont :

1-Méthode des témoins historiques ou approche historique : cette méthode repose sur l'idée que les malades sont comparables d'une époque à une autre. En fait, l'amélioration

des conditions diagnostiques et les modifications des contextes hospitaliers aboutissent à la constitution de groupes non comparables.

2- Comparaison des malades de deux hôpitaux ou approche géographique : les malades peuvent différer de façon considérable d'un hôpital à un autre.

3- Observation simple ou approche systématique : les groupes comparés risquent fort de n'être pas comparables, la diversité thérapeutique traduit une diversité initiale même en présence d'un dossier médical informatisé. Cette approche doit faire l'objet de plus de méfiance.

4- Cohorte de comparaison constituée par les sujets refusant la manœuvre : le refus peut être dicté par des raisons sociales, psychologiques ou médicales liées au traitement.

5- Répartition systématique : la distribution des malades dans chaque cohorte est réalisée par un moyen alphabétique, par la date de naissance... Cette technique peut être plus compliquée que le tirage au sort, n'offre aucune garantie éthique et cache des biais.

6-Techniques adaptatives : la distribution des sujets dans l'une ou l'autre des deux cohortes est fonction des résultats par l'accumulation des cas. Le plus grand nombre possible de malades est affecté au traitement le plus efficace en

administrant toujours le même traitement tant qu'un échec ne survient pas (**play**

the winner ». La qualité éthique est en réalité plus apparente que réelle.

III. Echantillonnage dans une expérimentation

- Le choix de la population dépend de la nature des interventions qui peuvent être des méthodes de prévention primaire (empêcher la survenue de la maladie) ou des méthodes de prévention secondaire (traitement des maladies) et tertiaire (réduction des séquelles).
- Les populations peuvent selon le cas être définies sur des bases administratives, professionnelles ou recrutées dans des organismes de soins.
- Il est en effet parfois plus commode que la cohorte de la manœuvre principale et celle de la manœuvre comparative ne soient pas composées d'individus mais de groupes entiers. Mais il faut évidemment s'assurer que les deux cohortes ne diffèrent pas par des caractéristiques inégalement réparties d'une cohorte à l'autre.
- Par ailleurs, les essais peuvent porter sur des individus ou des groupes d'individus appelés unités collectives.

- On peut, par exemple, attribuer la population d'une école (ou d'un quartier) à une cohorte et celle d'une autre école (ou d'un autre quartier) à une autre cohorte en vue de la réalisation d'un essai sur un vaccin.
- Il est plus sûr cependant de disposer de plusieurs paires de groupes et d'attribuer les groupes de chaque paire au hasard à la cohorte principale et à la cohorte comparative. Le choix d'unités collectives est obligatoire lorsque :
- Les interventions sont appliquées de façon collective (effet sur la fréquence des caries dentaires par la fluoration de l'eau, campagnes d'éducation sanitaire réalisées par la voie des media...);

IV. Modalités pratiques de tirage au sort

- Le tirage au sort doit être effectué le plus tard possible après l'inclusion du malade dans l'essai juste avant de commencer le traitement. Le système de randomisation doit être disponible au moment où se présente un malade.
- Il s'agit de constituer à partir d'un groupe donné deux cohortes soumises à deux manœuvres différentes si on veut, par exemple, comparer un traitement A à un traitement B.
- Le système le plus simple est celui des enveloppes cachetées classées dans un ordre chronologique (malade n°1, malade n°2, etc.) dans lesquelles sont contenues les exigences du traitement A ou du traitement B.
- Ces exigences sont préalablement randomisées grâce à l'utilisation d'une

table de **nombre au hasard**.

- Pour effectuer la randomisation, deux techniques sont possibles : tirage au sort simple et raffinement.

Le tirage au sort simple:

- il consiste à utiliser une seule colonne de la liste des nombres au hasard (table statistique 4). Les chiffres 0, 1, 2, 3, 4 peuvent être attribués aux sujets de la cohorte A et les chiffres 5, 6, 7, 8, 9 sont attribués aux sujets de la cohorte B.
- Le premier chiffre est désigné au hasard dans la colonne. S'il s'agit d'un 8 par exemple, la première enveloppe doit contenir les exigences relatives au traitement B.
- Si le chiffre suivant est 2, l'enveloppe correspondante doit contenir les instructions du traitement A et ainsi de suite.
- On peut généraliser aisément à plus de deux traitements.

- Pour trois traitements A, B, C, on peut établir la correspondance suivante : 0, 3, 6 sont attribués à A ; 1, 4, 7 peuvent désigner B et 2, 5, 8 sont attribués à C. Le chiffre 9 est tout simplement ignoré.
- Si le rapport de randomisation est de 2/1 en faveur d'un traitement, les chiffres 0, 1, 2, 3, 4, 5 peuvent désigner A tandis que les chiffres 6, 7, 8 sont attribués à B (le chiffre 9 est ignoré).
- Le tirage au sort simple aboutit à des effectifs à peu près égaux, mais

des inégalités restent possibles surtout si les effectifs sont faibles.

Le raffinement (tirage au sort équilibré):

- Il permet d'assurer l'égalité numérique des cohortes subissant chacune une des manœuvres de l'essai.
- Le raffinement sur le groupe total et le raffinement par séries sont les techniques utilisées pour réaliser l'équilibrage des cohortes.
- Le raffinement sur le groupe total est effectué si l'essai est conduit dans un seul centre et si l'effectif total des sujets éligibles est déterminé d'emblée. Pour répartir de façon équilibrée un groupe de 50 patients (numérotés de 00 à 49) en deux cohortes comprenant chacune 25 sujets, on utilise deux colonnes de la table des nombres randomisés.
- Les 25 premiers chiffres seront associés, par exemple, au traitement A tandis que les 25 suivants désigneront le traitement B (les chiffres 50 à 99 sont ignorés).
- **Le raffinement par séries:**
 - est la technique usuelle, il est toujours effectué dans les essais randomisés multicentriques impliquant plusieurs centres.
 - Pour deux traitements, l'équilibrage est réalisé par séries de 4, 6, 8 ou 10 patients. Le nombre usuel de malades dans une série est de 6.
 - A cette fin, on utilise une table de permutations au hasard à 6 éléments

(en annexe de ce chapitre).

- On lit au hasard (en fermant les yeux et en pointant un crayon) une première permutation, suite de six chiffres de 1 à 6.
- Cette suite peut être, par exemple, 246315. Si les chiffres 1, 2, 3 sont associés au traitement A et les chiffres 4, 5, 6 au traitement B, on obtient la séquence ABBAAB qui détermine le contenu des six premières enveloppes de randomisation.

- Les permutations suivantes de la table, lues dans un sens quelconque, déterminent la suite du tirage au sort.
- Dans un essai comparant trois traitements, on peut désigner le

traitement A par les chiffres 1 et 2, le traitement B par les chiffres 3 et 4 et le traitement C par les chiffres 5 et 6.

- Si le rapport de randomisation est 2/1, le traitement A peut être associé aux chiffres 1, 2, 3, 4 et le traitement B aux chiffres 5 et 6.
- Si l'essai est multicentrique, le plus simple est d'établir une liste pour un centre et de reprendre la même liste pour les autres centres en commençant par une autre suite, quitte à revenir au début si la liste est insuffisante.

Elaboration d'un protocole d'essai randomisé

- Un essai randomisé obéit aux mêmes règles qui régissent la réalisation de toute étude épidémiologique.
- C'est ainsi que le protocole guide l'exécution de l'essai et assure à tous les participants une exécution identique de ses différentes phases.
- Le protocole d'un essai comporte obligatoirement :

- la définition de la maladie,
- la définition des malades,
- la définition des manœuvres,
- le choix et la définition des cibles.

Toutes les précautions relatives à la définition de la maladie, des malades et des manœuvres ont pour but d'assurer la comparabilité et l'homogénéité des cohortes tout au long de l'essai.

- Il convient, en matière de cibles, de retenir le moins possible de critères et seulement ceux utilisables avec profit au moment de l'analyse.
- Un critère très utilisé est le délai d'apparition d'un événement.
- Dans tous les cas, les critères retenus doivent tendre vers l'objectivité (examen bactériologique...) mais certains critères restent essentiellement subjectifs (appréciation de l'intensité d'une douleur...).
- Les techniques à l'aveugle sont susceptibles de réduire cette subjectivité.
- Le résultat du travail d'élaboration du protocole se concrétise par trois documents :

- le protocole proprement dit ;

- le questionnaire d'enregistrement avec cinq grandes rubriques :

- o renseignements généraux et identification ;

- o description du malade et de la maladie ;

- o facteurs éventuellement pronostiques ;

- o description des traitements reçus ;

- o cibles.

6. Analyse des résultats

- L'analyse d'un essai randomisé est la comparaison des résultats, obtenus dans les cohortes, des critères de jugement retenus dans le protocole (comparaison de pourcentages, de densités d'incidence, de moyennes).
- Une étape obligatoire dans cette analyse est la description des sujets inclus :

- description des caractéristiques des sujets ;

- description des écarts au protocole.

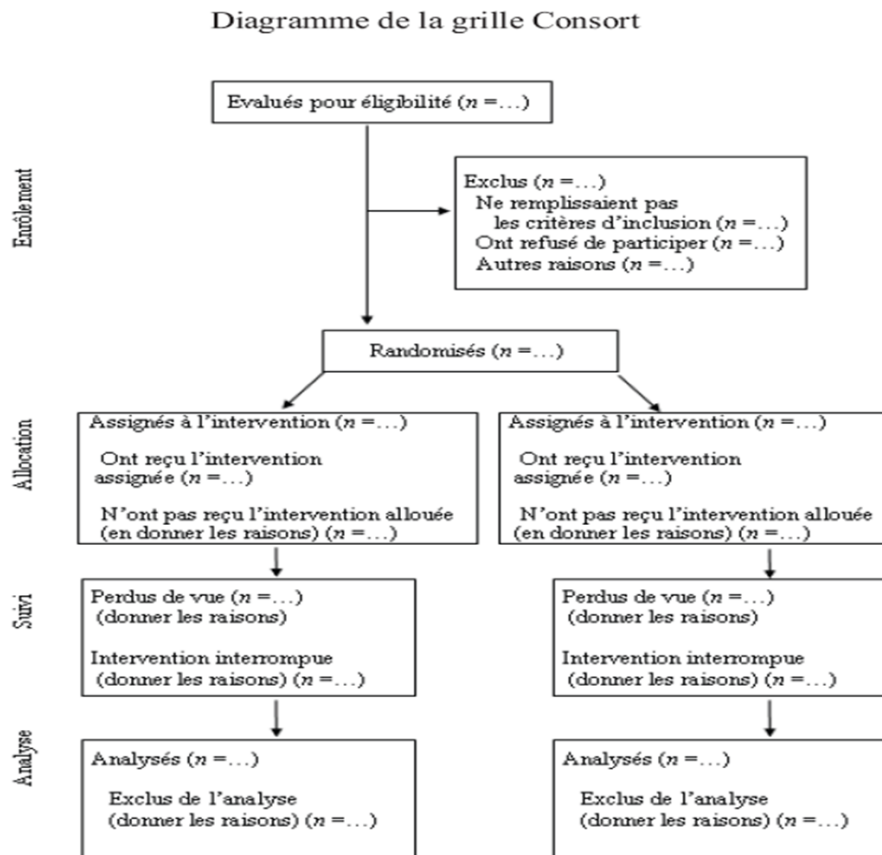
- Les caractéristiques des sujets inclus, les traitements administrés, les modalités de détection des cibles doivent être en principe l'exact reflet de ce qui est exigé dans le protocole.
- L'existence d'écarts au protocole est cependant inévitable en pratique sujets inclus à tort, écarts au protocole de traitement, perdus de vue (biais).
- Les écarts au protocole concernent les sujets qui n'ont pas respecté les conditions d'administration de la manœuvre telles qu'elles ont été définies dans le protocole. Si les écarts au protocole dépendent du traitement, un biais d'information survient.
- Dans la plupart des cas, les écarts au protocole de traitement ne sont pas dus à des raisons identiques dans les deux cohortes.
- Si l'un des deux traitements entraîne des effets indésirables sévères, les deux groupes seront ainsi déséquilibrés étant donné que les malades qui ont reçu un tel traitement seront plus nombreux à l'abandonner par rapport à ceux qui ont reçu l'autre traitement.
- Dans de telles conditions, l'exclusion de la comparaison finale de malades dont la manœuvre s'écarte trop de celle qui leur a été attribuée est une attitude erronée et illicite. D'une cohorte à une autre, les malades qui s'écartent de leur traitement

peuvent avoir une attitude si différente vis-à-vis de l'atteinte des cibles qu'une comparaison effectuée sur les seuls malades restants sera lourdement biaisée.

- L'analyse des résultats doit toujours se faire en intention de traiter, avec tous les malades déclarés bons pour l'essai et inclus dans l'essai tels qu'ils ont été affectés par tirage au sort. L'analyse par traitement réellement reçu (analyse per-protocole) doit toujours être prudente sinon rejetée si ses résultats ne coïncident pas avec ceux de l'analyse en intention de traiter.
- L'analyse en intention de traiter n'augmente pas le risque. Par contre, le risque augmente et engendre une perte de puissance de l'étude. La probabilité de ne pas mettre en évidence une différence qui existe réellement augmente évidemment lorsque les sujets changent de groupe.
- Dans certains cas, le nombre de ces écarts a une importance telle qu'il conduira à mettre en doute l'applicabilité du protocole et la possibilité d'une analyse à moins de reformuler les objectifs de manière pertinente.
- Les perdus de vue sont les sujets pour lesquels le résultat de la cible n'a pu être obtenu. Les perdus de vue ne pouvant être pris en compte dans l'analyse, le test statistique est moins puissant que celui initialement prévu et un biais de répartition peut survenir.
- Dans les essais où la cible est le délai d'apparition d'un événement, les perdus de vue sont pris en compte jusqu'au moment de leur disparition grâce aux méthodes de survie.
- L'analyse d'un essai est effectuée le plus souvent lorsqu'on dispose du résultat de la cible pour le nombre de sujets nécessaire déterminé dans le protocole (analyse en fin d'essai).
- La comparaison doit :
 - porter sur les critères de jugement définis dans le protocole ;
 - comparer les « bons groupes », c'est-à-dire sans exclusion illicite ;
 - déterminer les mesures d'association statistique et épidémiologiques appropriées.
 - Une analyse intermédiaire est toute analyse effectuée avant le recueil du résultat de la cible pour le nombre de sujets nécessaire déterminé dans le protocole.
 - Il est d'usage d'apprécier la comparabilité des groupes de l'essai et le maintien de cette comparabilité tout au long de l'essai (depuis l'inclusion des patients jusqu'à

l'analyse en passant par la randomisation, l'attribution des manœuvres et le suivi) par un organigramme appelé diagramme de flux (flowchart).

- La figure suivante est le diagramme de la grille Consort fréquemment utilisé.



Ethique

- Les problèmes éthiques, en matière d'essais randomisés, dominent
- tous les autres. Les difficultés éthiques dépendent :
 - de la gravité de la maladie ;
 - du type de traitement proposé ;
 - de la spécificité du traitement vis-à-vis de la maladie.
 - Le problème éthique est minimum si la maladie est bénigne, s'il existe un traitement classique de cette maladie et si l'on veut tester les effets symptomatiques de ce traitement.
 - Le problème éthique est maximum si la maladie est grave, s'il n'existe jusqu'alors aucun traitement efficace et si l'on cherche à mettre en évidence un traitement curatif.
 - Des conventions internationales fixent des règles strictes en ce qui concerne les recherches biomédicales portant sur l'être humain qui sont plus ou moins respectées.

- En santé communautaire, des populations considérables peuvent être impliquées. Ce qui, du point de vue éthique, amplifie chaque erreur.

Annexe : Table de permutation au hasard à 6 éléments

421536	154623	345162	513642	132654	324615
213564	124536	543216	364521	254631	452316
243516	143256	521643	235641	135246	531246
154623	635241	263541	642315	623514	132654
142635	134562	316425	514362	426135	241356
615243	126345	523614	156324	246315	541326
653412	425361	213645	523641	425361	126435
351624	156423	625431	245613	621345	641532
364215	461325	524361	216543	152346	651234
451326	162534	564231	263514	456213	264513
245361	456132	531246	563214	241563	435162
164253	432165	463152	126354	536142	213465
632514	321645	426153	451632	432165	462513
351642	314625	142635	653142	352146	536241
526341	215346	642315	142356	461325	153642
253164	132546	314526	263415	142635	614325
214365	561423	126534	523461	123564	352164
461532	423165	612453	436251	241635	615243
615324	632514	254163	235614	135642	324156
345162	426531	365124	362154	132465	531624