UE11 Biomédecine Quantitative

Jeudi 27 Février 2014 de 13h30 à 15h30

Ronéotypeur : Elsa Manzi

Ronéolecteur : Mathilde Berti

# TD n°1

# Sommaire

*1ère partie*

**I. Exercice 1- Suivi d'une épidémie**

**II. Exercice 2- Rôle de la modélisation mathématique dans l'étude de la résistance bactérienne aux antibiotiques. P.Y. Boëlle. La Lettre de l'Infectiologue - Tome XV - n° 7 - septembre 2000**

*2ème partie*

**III. Exercice 3- Facteurs associés au *burnout* en anesthésie–réanimation**

# I. Exercice 1- Suivi d'une épidémie

## 1. Calculer le taux d'attaque

Taux d'attaque : nombre de malades correspondant à la définition de l'épidémie/ population en fin d'épidémie.

Cas index : première personne qui va être identifiée et qui va ainsi permettre de définir l'épidémie.
On classe les patients selon la date d'isolement, le site infecté et selon le fait qu'ils soient certains ou probables.

→ Les cas expectorations sont tous considérés comme certains. Le cas urine est un cas probable car différent de la définition du cas index.

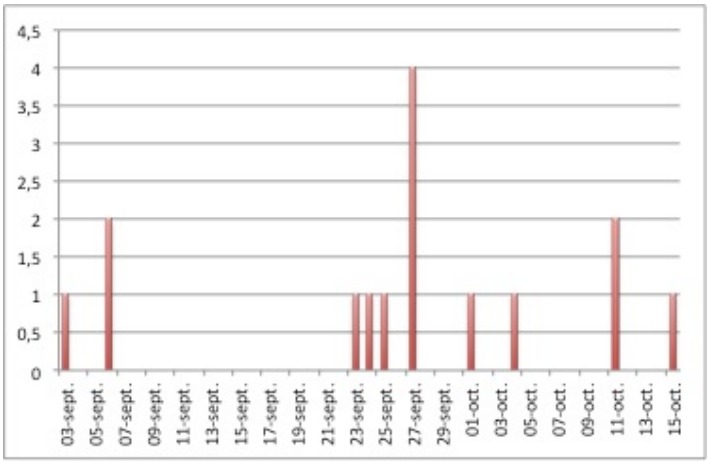
On a ainsi 14 cas certains, 2 cas probables.

- Le taux d'attaque est donc défini par 14/49=28,6% : (les cas probables ne correspondent pas à la définition du cas index, on ne les inclut donc pas dans le nombre de malades correspondant à la définition.)

- Le but du taux d'attaque est de le comparer à une incidence mensuelle. Ce taux d'attaque a été calculé sur deux mois, on peut donc le diviser par deux pour qu'il soit comparable à l'incidence mensuelle. On a en effet un taux d'attaque largement supérieur à l'incidence mensuelle, on peut ainsi affirmer qu'on est en présence d'une épidémie effective.

## 2. Quel outil de base pour le suivi d'une épidémie pouvez-vous construire à partir de ces données ? le construire.

On cherche un outil statistique graphique qui nous permet de définir l'épidémie : c'est la courbe épidémique. Elle est toujours définie par unité de temps et par unité de cas.



3. **Quelles informations sont fournies par cet outil ? ses avantages et inconvénients ?**

La courbe épidémique permet d'avoir :

- le cas index (premier bâtonnet)
- pic épidémique, ici le 27/09
- la progression de l'épidémie en fonction du temps

La courbe épidémique permet de donner un aspect qualitatif de l'épidémie. Le principal défaut de la courbe épidémique est que le recueil des données est a posteriori de l'épidémie. Ainsi, on n'a pas d'idée de l'évolution de l'épidémie dans le temps : aucune notion du dynamisme.

## II. Exercice 2- Rôle de la modélisation mathématique dans l'étude de la résistance bactérienne aux antibiotiques

1. **Selon l'auteur quel est le protocole optimal de prescription d'antibiotiques pour éviter la dissémination de bactéries résistantes dans la communauté ?**

   La modélisation permet d'anticiper plusieurs scénarios au niveau de la résistance des bactéries. Selon l'auteur, le protocole optimal de prescription d'antibiotiques pour éviter la dissémination de bactéries résistantes dans la communauté, est l'utilisation des antibiotiques en combinaison plutôt que séquentielle : « si l'on peut utiliser plusieurs antibiotiques, le même modèle montre que l'emploi de combinaisons d'antibiotiques retarde toujours la dissémination des organismes résistants, lorsque l'on compare ce protocole à une utilisation en alternance » (haut p303).

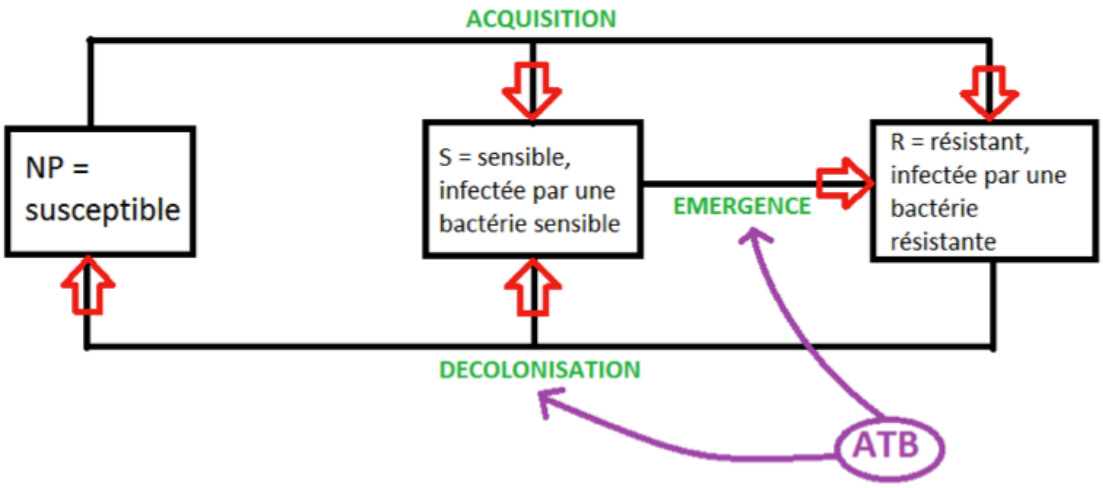
2. **Par analogie avec le modèle SIR et à partir du paragraphe «dissémination de la résistance » proposez un modèle compartimental (modèle interhôte) prenant en compte l'action des antibiotiques.**

   <u>Rappel :</u> Modèle SIR : susceptible infecté retiré. Il permet une description compartimentale de la maladie infectieuse. Une personne est considérée successivement

- « susceptible »: lorsqu'elle n'est pas infectée
- puis « infectieuse »: lorsqu'elle est infectée
- et enfin « retirée » de la chaîne de transmission de l'épidémie: soit on devient vacciné et on ne peut pas redevenir susceptible, soit on ne l'est pas et on peut redevenir susceptible.
  → Le SIR modélise ce passage de « susceptible » à « infectieux » à « retiré ».

Ici, on cherche à modéliser les mouvements de population entre les différents compartiments, avec :

- NP : non porteur, population susceptible
- S : porteur de bactérie sensible
- R : porteur de bactérie résistante




- La population susceptible peut devenir sensible : c'est le phénomène d'acquisition
- La population sensible peut devenir résistante : c'est le phénomène d'émergence
- La population résistante peut devenir susceptible ou sensible : c'est le phénomène de décolonisation.

→ Les antibiotiques peuvent à la fois impacter sur l'émergence et la décolonisation.

3. **Donnez la définition du R0 et rappelez le seuil au-dessous duquel une épidémie est évitée.**

Rappel: Le ratio de reproduction $R_0$ permet de classer les maladies par potentiel épidémique et correspond au nombre de cas secondaires directement infectés par une unique personne infectieuse (=porteuse d'une bactérie sensible ou résistante), placée dans une population totalement susceptible (=tout le monde est non porteur) à la maladie.

- Si $R_0 > 1$ alors chaque individu infecté va être capable de « se reproduire » en infectant plus d'un autre individu, ce qui permettra à la maladie de se répandre dans la population, causant une épidémie.
- Si $R_0 < 1$ alors chaque individu infecté ne transmet pas à plus d'une personne l'infection. Il n'y a pas d'augmentation de l'incidence. On évite donc l'épidémie
- Si $R_0 = 1$ alors une personne est capable d'infecter une seule autre personne.

Dans ce cas-là, on a $R_0$:

$$R_0 = \tau_{p\text{-}s} \cdot P_p \cdot d_p \cdot \tau_{s\text{-}p} \cdot P_s \cdot d_s$$

où $\tau_{p\text{-}s}$ est le taux de contact des patients vers les soignants, $\tau_{s\text{-}p}$ celui des soignants vers les patients, $P_p$ et $P_s$ les pourcentages respectifs de colonisation chez les patients et les soignants, et enfin $d_p$ la durée de séjour des patients et $d_s$ la durée de colonisation des soignants. Pour éviter l'épidémie, il faut que $R_0$ reste inférieur à 1

4. **Proposez une liste de mesures de contrôle d'une épidémie d'une bactérie résistante à l'hôpital et indiquer sur quel paramètre du R0 (du paragraphe « dissémination à l'hôpital ») elles influent principalement.**

Selon la formule du ratio de reproduction, on a donc 6 paramètres qui nous permettent de contrôler la progression de l'épidémie.

- Taux de contact des patients vers les soignants ET des soignants vers les patients (les deux $\tau$).

>Ce taux de contact va être dépendant du nombre de patients, du nombre de soignants mais aussi de l'efficacité de contact ( par exemple, un patient qui présente une expectoration aura une contagiosité différente de l'infirmière, qui est porteuse simplement au niveau des se mains).

Ainsi, on peut agir sur :

→ le nombre de contact :

- mesure d'isolement: lits ou ailes dédiés
- mesure de cohorting: partage des patients entre soignants
- augmentation du ratio entre personnel soignant et patient: diminuer les contacts avec des soignants différents

→ l'efficacité de contact :

- mesures de protection individuelles : masque, blouse…

- Pourcentage de colonisation chez les patients $P_p$ :
  - Antibiothérapie
  - Vaccins
  - Screening : on screene nos patients afin de détecter la présence de certains types de bactéries

- Pourcentage de colonisation chez les soignants $P_s$:
  - Lavage des mains
  - Vaccins

- La durée de colonisation du personnel soignant $d_S$ :
  - Lavage des mains

- La durée de séjours du patient $d_P$ :
  - Pas de mesure de contrôle

*2ème partie*

## III. Exercice 3- Facteurs associés au burnout en anesthésie–réanimation.

### 1. Proposez une méthode permettant d'étudier le lien entre le score d'épuisement émotionnel et l'âge et plus généralement tous les facteurs étudiés.

Dans cette étude, on cherche à regarder quels facteurs sont associés au fait d'avoir un burnout ou pas. On est considéré en « burnout », si on a au moins un des trois scores élevés (déshumanisation, accomplissement personnel, épuisement émotionnel). Quand on veut chercher des facteurs associés à une variable on va la croiser avec d'autres variables potentiellement explicatives (âge, taille, sexe...). Ainsi nous avons une variable à expliquer : le score d'épuisement, et des variables explicatives.

- Âge : variable quantitative continue (car on peut avoir un nombre infini de valeurs)
- Score d'épuisement : variable quantitative discrète (car ne prend que 55 valeurs de 0-54).

Rappel : Variable quantitative variation d'unité de X
Variable qualitative variation de X=1 ou X=0

- Tableau 1 : L'analyse univariée permet de croiser deux variables. On fait un test statistique, soit de comparaison/proportion, soit de comparaison de moyenne. Le résultat du test est présenté dans la dernière colonne avec le degré de signification (la p-value). C'est une comparaison entre des variables une à une. Il peut y avoir plusieurs variables associées au burnout.

- Tableau 2 : L'analyse multivariée détermine les variables indépendamment associées à la variable à expliquer ; c'est-à-dire qu'on met ensemble les variables significatives liées (par exemple le poids est souvent lié à la taille). Il ne reste que les variables qui ne sont pas liées entre elles.

On cherche à savoir l'analyse de la liaison entre ces deux variables qu'on estime a priori indépendantes.

- Pour ce faire, on peut faire un test de Student afin de déterminer si ces deux variables quantitatives sont corrélées.
  - Dans le cadre du coefficient de corrélation, les variables ont un rôle symétrique (on ne cherche pas à expliquer l'un par l'autre).
  - Deux tests:
    - Pearson paramétrique
      - Condition de validité : ce test suppose que les deux variables X et Y soient normalement distribuées (selon la loi Normale).
      - Calcul simplifié
    - Spearman non paramétrique

- On peut aussi faire une régression linéaire.
  - Dans ce cadre, les variables ont un rôle asymétrique.
    - Y : variable à expliquer (score d'épuisement)
    - X : variable explicative (âge).
  - 3 conditions de validité pour la régression linéaire :
      - Indépendance des observations Xi et Yi
      - Linéarité de X et Y (données normalement ditribuées
      - **ε** suit une loi normale de moyenne 0 et de variance **σ**.

<u>Rappel</u> : p-value est la force de conviction du test.

→ p = 1 ou – 1 indique une relation linéaire parfaite

→ p = 0 indique l'absence d'association

ATTENTION : Un coefficient de corrélation nul ne signifie pas qu'il n'y a pas de lien entre les deux variables, mais elle est non linéaire.

Par exemple, on peut avoir un coefficient de corrélation de 0,1 (= faible) avec une p-value significative si on a un échantillon de un million. Plus l'échantillon est grand, plus on a de chance d'avoir quelque chose de significatif. Si on veut mettre en évidence une toute petite différence, il faut un échantillon énorme.

<u>Remarque</u> : à ne pas confondre avec le degré de signification du test ! Il y a une différence entre deux choses (= corrélation non nulle) qui est plus ou moins importante.

**Quel est l'intérêt de cette approche par rapport à celle retenue?**

Ici, c'est un modèle linéaire de régression entre la variable score d'épuisement et l'âge. On obtient un coefficient qui donne une force d'association. Dans l'article, ils ont comparé : burnout (oui/non) et l'âge moyen, ce qui donne juste une p-value.

L'intérêt de l'approche par corrélation, c'est que le modèle linéaire donne la force d'association précise entre les deux, en plus de la p-value.

**2) A propos du tableau 1. Que représentent les p-values données?**

p-values est le degré de signification.

- Si il est ≤ 5% on considère que l'hypothèse nulle H0 est fausse, on la rejette, et on accepte H1. En effet on a peu de chance d'observer cette différence du simple fait du hasard. Plus p-value est proche de 0, plus on a confiance (moins on se trompe).
- Si p > 0,05, c'est non significatif. On ne rejette pas H0 et on ne conclut surtout pas ! On ne dit pas qu'il n'y a pas de différence. En effet soit il n'y a effectivement pas de différence, soit la différence n'a pas été mise en évidence, souvent parce que l'effectif utilisé était trop petit au départ (manque de puissance).

**Que signifie en termes de force du lien entre l'age et la vie en couple le fait d'avoir de p-value respectivement à 0,021 et 0,0049. Donnez le principe des tests utilisés?**

La p-value de 0,0049 pour la vie en couple indique une association avec la burnout : plus on vie en couple, moins on a de chance de faire un burnout (= test du Chi2).

La p-value de 0,021 pour l'âge indique une association avec le burnout : plus on est âgé, plus on a un risque de burnout (= test de Wilcoxon ou Studden pour la comparaison de deux moyenne).

**3) Quel modèle statistique vous permet d'étudier le lien entre l'existence d'un burnout et chacun des facteurs du tableau 1?**

Il s'agit d'une régression logistique, parce qu'il y a une variable à expliquer qualitative (burnout oui/non). La régression linéaire est utilisée quand la variable à expliquer est quantitative.

<u>Rappel</u> :

Le principe de la méthode des moindres carrés, est de trouver la droite de régression Y = **α** + **β**X qui résume le mieux le nuage de point de n couple d'observations indépendantes. On cherche alors la droite dont les points sont les plus proches en distance. Pour chaque couple de point, on calcul la distance à la droite potentielle y = a + bx. La distance d'un point à la droite est la distance verticale entre l'ordonnée du point observé et l'ordonnée du

point correspondant sur la droite. Plutôt que de prendre la valeur absolue, on va mettre la valeur au carré pour n'avoir que des mesures positives. Pour faire une régression d'un nuage de point, le principe est de faire passer une droite au milieu des points afin de voir si la pente est différente de 0 ou pas. Si la pente est différente de 0, alors les deux variables sont liées entre elles. Pour tracer cette droite on utilise donc la méthode des moindres carrées. On calcul la somme du carré des écarts minimum par rapport à la courbe pour chaque point.

**Quels en étaient les implications? Quels en sont les avantages?**

La régression logistique : la variable est binaire (burnout oui/non), alors qu'à l'origine on avait une échelle quantitative: on a donc perdu de l'information. En effet on a « dichotomisé » ; c'est-à-dire transformé une variable continue en classe. L'avantage est de pouvoir faire ensuite de la régression logistique et d'obtenir l'Odds Ratio qui s'interprète quasiment comme un Risque Relatif.

**4) Quelle est la différence entre le tableau 1 et le tableau 2? Quelle est la conséquence pour l'interprétation des OR? Ecrire le modèle logistique multivarié correspondant au tableau 2.**

Dans le tableau 2 : On a pris toutes les variables avec un $p<0,10$ pour une analyse multivariée et le logiciel a conservé uniquement les variables indépendamment associées au burnout. Les autres éliminées étaient déjà liées entre elles.

- La première colonne est l'estimateur qui correspond aux **β** (= coefficient associé à chaque variable dans l'analyse multivarié= logOR). Ici, on a donc plusieurs estimateurs et plusieurs variables.
  B = **β**0 + **β**1 EVA (qualité du travail) + **β**2 EVA (vie privée) + **β**3 EVA (fatigue) + ...
  → **β**1 est la force de l'association entre EVA (qualité du travail) et le burnout.

- La deuxième colonne c'est la p-value qui dit si l'EVA (qualité du travail) est significativement associé au fait de faire un burnout ou pas. (On remarque que tout est significatif).

- La troisième colonne est l'Odds Ratio. C'est l'exponentielle du Beta.

Donc pour la première variable qualité du travail : OR = $e^{-0,135}$ = 0,87. Plus on a une bonne qualité de travail, moins on a de chance de faire un burnout.

Rappel :

- OR > 1 : facteur de risque
- OR < 1 : facteur protecteur
- OR = 1 : comme corrélation à 0

On doit maintenant donner l'intervalle de confiance. P = 0,0002, donc l'intervalle de confiance ne comprendra pas 1, ainsi c'est significatif.

Rappel :

- Quand p est significatif, l'intervalle de confiance ne comprendra pas 1
- Quand p est non significatif l'intervalle de confiance comprendra 1

Dans le tableau 2 l'intervalle de confiance ne comprendra jamais 1, car toutes les p-values sont significatives.
Enfin, l'avantage du modèle de régression logistique est la force de l'association qui se mesure avec l'OR. Car il n'y a pas d'OR dans la régression linéaire.

**Quel est l'OR de développer un burnout pour un patient qui est en conflit avec ses collègues et qui a une EVA qualité du travail à 3/10 par rapport à un patient qui n'est pas en conflit avec ses collègues et qui a une EVA qualité de travail à 9/10, toute chose étant égale par ailleurs?**

Conflit avec les collègues: 1-0
Qualité de travail: conflit 3/10- pas de conflit 9/10

1: patient qui est en conflit avec ses collègues et qui a une EVA qualité du travail à 3/10
2: patient qui n'est pas en conflit avec ses collègues et qui a une EVA qualité de travail à 9/10

$B_1$-$B_2$ = $\beta 0_1$ + $\beta 1_1$ EVA (qualité du travail) + $\beta 2_1$ EVA (vie privée) + $\beta 3_1$ EVA (fatigue) + ... – ($\beta 0_2$ + $\beta 1_2$ EVA (qualité du travail) + $\beta 2_2$ EVA (vie privée) + $\beta 3_2$ EVA (fatigue) + ...)

$B_1$-$B_2$= $\beta 1_1$ EVA (qualité du travail) + $\beta 1_1$ (conflits avec les collègues) – ($\beta 1_2$ EVA (qualité du travail) + $\beta 1_2$ (conflits avec les collègues))

$B_1$-$B_2$=(-0,135)x3 + 0,363x1-((-0,135)x9+(0,363)x0)

$\Delta$B=1,173 soit

OR=$e^{-1,173}$=3,23

**5) Quel est le lien entre l'estimateur et l'OR?**

Le lien entre l'estimateur et l'OR, c'est l'exponentiel. OR = $e^{-\beta}$

**Que signifie un OR à 0.87 pour l'impact de l'EVA qualité de travail sur l'existence d'un burnout?**

OR = 0,87 Un OR, pour une variable continue, est la variation du risque par unité. Cela signifie qu'on multiplie par 0,87 le risque d'avoir un burnout à chaque fois qu'on augmente d'une unité de l'échelle EVA.

**Si ma probabilité d'avoir un burnout est de 10% sans regretter ma spécialité, que devient cette probabilité si je regrette ma spécialité toute chose étant égale par ailleurs?**

$$\log(OR) = \log\left(\frac{\frac{P1}{1-P1}}{\frac{P2}{1-P2}}\right)$$

= Beta(x1-x2)

OR = exp(Beta (x1-x2))

Si p = 1 on regrette

OR = 1,86 (environ = 2)

**OR = P1(1-P0) / P0(1-P1) = 2x0,1/0,9 = P1(1-P1)**
**0,9 P1 = 0,2 - 0,2 P1**
**1,1 P1 = 0,2**
**P1 = 0,18**

**Quelles sont les conditions pour lesquelles OR et RR sont proches?**

L'OR s'interprète comme un RR :

- Si la probabilité de la maladie (= prévalence) est petite
- Si l'OR est proche de 1

**Quelle concordance doit-il y avoir entre l'intervalle de confiance et la p-value?**

Si l'intervalle de confiance ne comprend pas la valeur 1, alors la p-value est significative.

# Fiche ED1-Biomédecine Quantitative

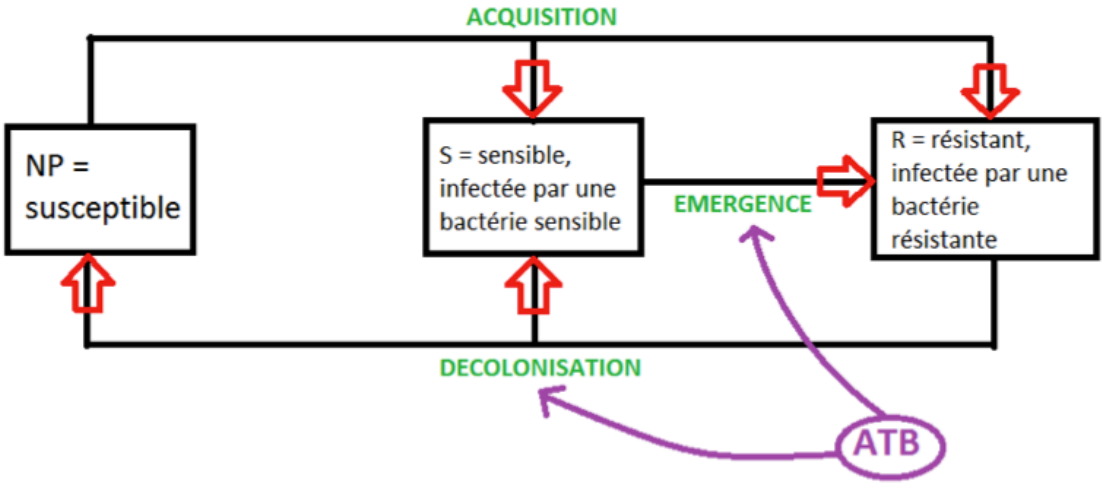
**Rappels :**

- Taux d'attaque : nombre de malades correspondant à la définition de l'épidémie/ population en fin d'épidémie.
- Cas index : première personne qui va être identifiée et qui va ainsi permettre de définir l'épidémie.

**Courbe épidémique**

| Avantages | Inconvénients |
|---|---|
| Visualisation de :<br>– cas index<br>– pic épidémique<br>– la progression de l'épidémie en fonction du temps | Recueil des données a posteriori<br>Pas de dynamisme |

**Modèle SIR** : C'est un outil d'anticipation qui permet une description compartimentale : la personne est d'abord **« susceptible »** (avant d'être infectée), puis **« infectieuse »** (infectée et contagieuse), et enfin **« retirée »** de la chaine de transmission de l'épidémie (guérison, isolement ou décès).




- Ratio de reproduction $R_0$ :nombre de cas secondaires directement infectés par une unique personne infectieuse (=porteuse d'une bactérie sensible ou résistante), placée dans une population totalement susceptible (=tout le monde est non porteur) à la maladie.

- Si $R_0 > 1$ alors chaque individu infecté va être capable de « se reproduire » en infectant plus d'un autre individu, ce qui permettra à la maladie de se répandre dans la population, causant une épidémie.
- Si $R_0 < 1$ alors chaque individu infecté ne transmet pas à plus d'une personne l'infection. Il n'y a pas d'augmentation de l'incidence. On évite donc l'épidémie
- Si $R_0 = 1$ alors une personne est capable d'infecter une seule autre personne.

**Comparer deux variables :**

| | paramétrique | non paramétrique |
|---|---|---|
| 2 variables qualitatives | Chi2 | Fisher (si pop <5 dans un case) |
| 1 qualitative et 1 quantitative | Student (au moins 30 pers) | Wilcoxon (au mois 30 pers) |
| 2 variables quantitatives | Pearson | Spearman |

- Paramétrique **:** la distribution des données suit une loi Normale
- Non paramétrique : la distribution des données ne suit pas une loi Normale