

# TD 1

## UE MRCE

### Mise en situation

Vous travaillez localement à la cellule d'épidémiologie. Un syndrome associant fièvre, diarrhée et vomissements se répand actuellement dans la région, en dehors de toute épidémie virale classique. Une enquête épidémiologique est initiée afin de donner une première description des cas et d'essayer d'identifier de potentiels facteurs de risque ou sources de contamination.

Les cas ont été identifiés par les autorités sanitaires et seront comparés à un témoin familial (membre de la famille du cas ne présentant pas la maladie).

Réfléchissez en groupe aux variables que vous allez recueillir sur les participants à cette étude afin de répondre à la question d'étude. **Noter au moins 5 questions et le type de variables correspondant.**

Ex :

Variable	Type	Sous-type	Valeurs	exemples
Age	Quantitatif	Continu	Min : 0, Max : 120 Unité : année	18, 23, 54, 101
Sexe	Qualitatif	Binaire	femme, homme, autre, NSP	
...	...	...	...	...

### Le fichier des erreurs

Des premières données ont été collectées par un collègue zélé pendant que vous réfléchissiez aux variables à recueillir. Ces données sont disponibles sur l'espace Madoc de l'UE, section ED / ED1, dans le fichier **td1\_erreurs.xls**.

Ouvrez ce fichier. Votre collègue n'est clairement pas formé au traitement des données... Quelles « erreurs » a-t-il commises ? Quelles solutions proposez-vous ?

### Création du masque de saisie

Il vous faudra construire un outil pour saisir les données de chaque individu. Ouvrez un navigateur internet et allez à l'adresse suivante : [wepi.org](http://wepi.org).

Wepi est une application web qui permet de créer des masques de saisie, utiles pour collecter des données de façon systématisée. Contrairement à d'autres outils en ligne du même type (Google Forms, Survey Monkey, etc.), le concepteur de Wepi est agréé pour héberger des données de santé. Cela signifie qu'il présente un niveau de sécurité et de confidentialité nécessaire pour héberger ce type de données sensibles. Toute étude portant sur des données de santé doit selon la loi être hébergée par un organisme agréé.

1. **Créez un compte par groupe**, puis créer une nouvelle étude. Donnez-lui un nom, puis remplissez les données nécessaires au RGPD (responsable : Nantes Université, catégorie principale de données : données fictives, finalité : enseignement).

Un cadre de texte sera ensuite disponible. Copiez-y le texte suivant :

```
*Merci de remplir ce questionnaire pour chaque cas et témoins inclus  
N° d'identification :  
Age :  
Sexe :
```

Sauver le questionnaire et passer en mode visualisation.

Noter que les lignes ne commençant pas par un astérisque sont formatées comme des questions. Cependant, le type de variable choisi par défaut ne convient pas partout. Ouvrez les préférences de chaque question en cliquant sur l'icône « stylo » sur la droite du bloc contenant la question ; modifiez ces préférences (nom de la variable, type de question) pour qu'elles conviennent aux questions posées.

2. **Ajoutez les cinq questions** que vous aviez préalablement notées lors du travail de groupe en ajustant les paramètres en fonction du type de variable. Quelles questions devraient selon vous être obligatoires ?

3. **Repasser à présent en mode texte**. Comme vous le voyez, toutes les questions peuvent être formatées directement dans ce mode. Cela permet d'aller plus vite qu'en mode visualisation, mais nécessite de connaître les marques de formatage. On aimerait dans notre étude savoir si les individus ont déjà été en contact avec un cas. Essayez d'ajouter cette question directement en mode texte.

Si la réponse à cette question est positive, on aimerait connaître la date du dernier contact. **Essayez d'ajouter une question dont l'apparition est conditionnée** (pour cela, vous pouvez repasser en mode visualisation et cliquer sur le stylo pour ouvrir les paramètres de la question à faire apparaître).

4. **Testez votre questionnaire** en cliquant sur l'onglet « Saisir des données » et remplissez quelques questionnaires fictifs pour vous assurer que tout fonctionne correctement. Explorez l'application pour voir comment exporter vos données pour les analyser dans un logiciel de type tableur.

## Analyse de données

Sur l'espace madoc de l'UE MRCE, vous pourrez trouver un jeu de données au format tableur (ods, xls ou csv). **Téléchargez-le et ouvrez-le avec le logiciel adapté** : OpenOffice (pour les fichiers ods et csv) ou Excel (pour les fichiers xls ou csv). Dans la suite du document, nous ferons l'hypothèse que vous utilisez Excel. Les données sont présentes sur la première feuille.

1. **Réaliser les opérations** présentées ci-après : (pour un environnement d'analyse plus propre, utilisez plutôt un nouvel onglet pour cette analyse)

- dans une cellule (« case »), tapez la formule nécessaire pour calculer une moyenne (voir le tableau plus bas) puis cliquez sur l'entête de la colonne correspondant à l'âge dans la feuille des données. Fermez ensuite la parenthèse pour conclure la formule. Appuyer sur Entrée et vous devriez vous apparaître la moyenne d'âge de votre population.

Estimez maintenant l'écart-type, la variance, la médiane et les 1<sup>er</sup> et 3<sup>ème</sup> quartile à l'aide des fonctions adaptées.

## 2. La distribution de l'âge vous semble-t-elle symétrique ?

Afin de s'en assurer, nous allons **construire un histogramme**. Pour cela, cliquez sur l'entête de la colonne de la variable « age » (B), et insérer un histogramme (Insertion > Insérer un graphique statistique).

Graphiquement, la distribution de l'âge semble-t-elle conforme à vos attentes ?

## 3. Nous allons maintenant comparer les caractéristiques des cas et des témoins à l'aide de **tableaux croisés dynamiques**.

Pour cela sélectionnez l'ensemble des données de la feuille 1 et insérer un tableau croisé (voir le tableau à la fin du document). Dans les champs disponibles sélectionnez la variable « cas » comme champ de lignes et la variable à décrire en champ de valeur pour une variable quantitative, ou en champ de colonne pour une variable qualitative (pour les variables qualitatives, il faudra ajouter une variable « à compter » quelconque dans le champ valeur) .

Par défaut, Excel présente comme résultat la somme de la variable qui est dans le champ valeur. Modifier ce calcul en fonction de ce que vous voulez calculer dans les paramètres du champ valeur.

Utiliser ces tableaux comme base pour :

- comparer les âges entre les cas et les témoins,
- construire les tableaux de contingence utilisés pour le calcul l'odds ratio et le risque relatif liés au sexe

## 4. On aimerait enfin voir le lien entre la taille et le poids des individus inclus dans l'étude. Sélectionnez ces deux variables dans la première feuille et insérer un diagramme de dispersion.

Toujours sur la base de ces deux variables, utilisez la formule adaptée pour calculer un coefficient de corrélation. Comment l'interprétez-vous ?

## Tests statistiques

Il est en fait assez rare d'utiliser Excel ou plus généralement des tableurs dans la recherche médicale. On utilise généralement des logiciels dédiés à l'analyse statistique comme R.

Lancer RStudio (ou directement R) et ouvrez le fichier de données au format csv disponible sur l'espace Madoc (td1.csv). Vous pouvez pour cela passer par File > Import Dataset > From Text (base), ou lancer le code suivant :

```
castemoins <- read.csv2("chemin-vers-le-fichier/td1.csv")  
## il faut bien sur remplacer chemin-vers-le-fichier en fonction de  
## l'emplacement de votre fichier.  
## vous pouvez voir le début du jeu de données avec :  
head(castemoins)
```

Faites un test de Fisher pour mesurer le lien entre le sexe et la maladie. Faites également un test de corrélation entre la taille et le poids. Interprétez les résultats.

```
fisher.test(castemoins$sexe, castemoins$cas)  
cor.test(castemoins$taille, castemoins$poids)
```

Formule	Excel	Open Office
Calculer une moyenne	=moyenne(...)	=moyenne(...)
Calculer un écart-type	=ecartype(...) ou = ecartype.standard(...)	=ecartype(...)
Calculer une variance	=var(...) ou =var.s(...)	=var(...)
Calculer une médiane	=mediane(...)	=mediane(...)
Calculer des quartiles	=quartile(... ; type) ou =quartile.exclure(... ; type)  <i>*type* est le quartile souhaité (1 pour premier, 2 pour deuxième, etc.)</i>	=quartile(... ; type)  <i>*type* est le quartile souhaité (1 pour premier, 2 pour deuxième, etc.)</i>

<b>Arrondir au chiffre inférieur</b>	=plancher.math(... ; precision)  La precision permet d'arrondir à l'unité (=1), à la dizaine (=10), au centième (=0,01), etc.	=plancher(... ; precision)  La precision permet d'arrondir à l'unité (=1), à la dizaine (=10), au centième (=0,01), etc.
<b>Insérer un tableau croisé</b>	Insertion > Tableau croisé dynamique	Données > Tableau croisé > Créer
<b>Calculer un coefficient de corrélation</b>	=coefficient.correlation(... ;...)  <i>le premier argument correspond au champ de la première variable, le second au champ de la seconde</i>	=coefficient.correlation(... ;...)  <i>le premier argument correspond au champ de la première variable, le second au champ de la seconde</i>