

Mesures descriptives et représentations graphiques des variables QUALITATIVES

Objectives

À la fin du TP, l'étudiant devra posséder les connaissances suivantes :

- Savoir les statistiques descriptives (fréquences absolues, fréquences relatives) pour la description d'une seule variable qualitative
- Savoir les types de graphiques appropriés pour la description d'une seule variable qualitative (graphique sectorielle, graphique par de colonne/barre)
- Savoir déterminer le tableau de contingence pour la distribution de deux variables qualitatives
- Savoir les types de graphiques appropriés pour la description de la relation entre deux variables qualitatives (graphique par de colonne/barre)
- Créer des tableaux de fréquences à l'aide de la fonction COUNTIF
- Créer des tableaux de fréquences à l'aide de l'option PIVOT TABLE (fr. tableau croisée dynamique)
- Créer des tableaux de contingence à l'aide de l'option PIVOT TABLE

Principes pour la réalisation des graphiques

Toute représentation graphique doit avoir:

- ✓ titre clair, concis et précis
- ✓ définitions des axes, sans abréviations (à l'exception des unités de mesure)
- ✓ unités de mesure (ou est le cas)
- ✓ légende (s'il faut)
- ✓ toute symboles, des abréviations ou lettres utilisées dans la figure doivent être expliqué clairement dans la légende

Scenario

Une étude a été menée pour évaluer l'association entre la présence de l'hypertension, le tabagisme, et le risque d'infection au COVID-19 chez les patients atteints de cancer. L'étude a été réalisée sur un échantillon aléatoire de 400 sujets adultes atteints de différents types de cancer, patients extraits du registre régional du cancer de la région du nord-ouest de la Roumanie qui ont été testés pour l'infection au COVID-19 entre le 2 février et le 2 juin 2020. Les données recueillies se trouvent dans le fichier **Patients.xlsx**.

Demandes

1. Créez un nouveau dossier nommé **TP04NP** ou **N = votre nom, P = votre prénom** sur le Bureau (Desktop) de Windows.
2. Enregistrer le fichier Excel **Patients.xlsx** dans le dossier **TP04NP**.

3. Dans le même fichier, créer plusieurs feuilles de calcul et renommez-les: Pie (Camemebert), Colonne_1_variable, Colonne_2_variables, Barres, au lieu de Sheet1, Sheet2, Sheet 3.

Statistiques descriptives et Representation graphique d'une **variable QUALITATIVE**

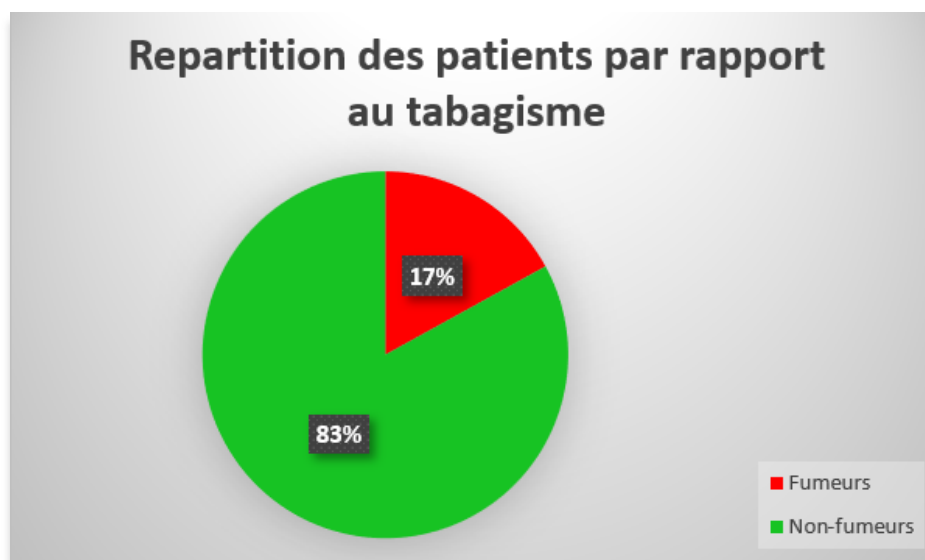
4. Dans la feuille **Pie**, copier la variable **Tabagisme**.
 5. Dans la même feuille nommée **Pie**, a coté de la variable **Tabagisme**, créer une table de la forme suivante **puis a l'aide des notions du cours, remplir le tableau** :

Variable	Type de la variable (qualitative dichotomique/qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue)
Tabagisme	

6. Dans la même feuille Pie, créer un autre tableau (tableau de fréquence) contenant les fréquences absolues (les effectifs) du Tabagisme en utilisant la fonction **COUNTIF**. Le tableau doit avoir la forme suivante :

Tabagisme	Fréquences absolues (Nombre de cas)
fumeur (=1)	
Non-fumeur (=0)	

7. Réalisez un graphique **CAMEMBERT** pour représenter les données collectées dans le tableau de fréquences créé précédemment.
 8. Formater le graphique pour qu'il ressemble à celui de l'image ci-dessous:



9. Au côté droit du tableau de fréquences, calculer le rapport Fumeurs/Non-fumeurs dans l'échantillon d'étude.
 10. Dans la feuille nommée **Colonne_1_variable**, copier la variable **Tranche d'âge**.

11. Dans la même feuille **Colonne_1_variable**, à côté de la variable **Tranche d'âge**, créer une table de la forme suivante **puis à l'aide des notions du cours, remplir le tableau** :

Variable	Type de la variable (qualitative dichotomique/qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue)
Tranche d'âge	

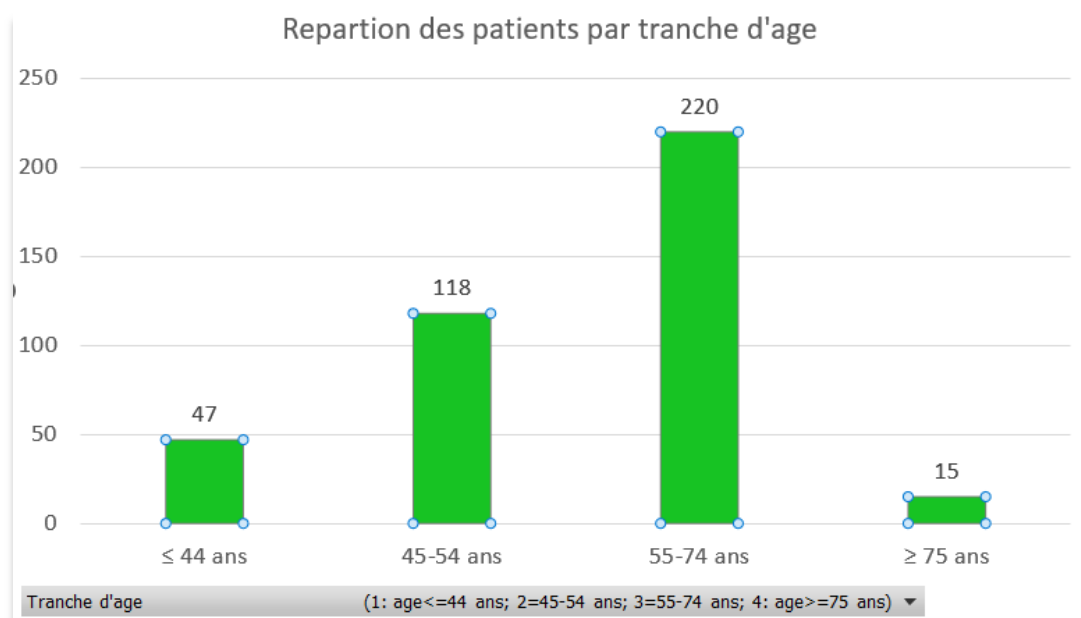
12. Dans la même feuille **COLONNE**, créer un autre tableau (**TABLEAU DE FREQUENCE**) contenant les fréquences absolues, en utilisant l'option **PIVOT TABLE**.

13. Formater le tableau croisé dynamique obtenu antérieurement, pour qu'il ressemble à celui de l'image ci-dessous :

Tranche d'âge	Fréquences absolues (Nombre de cas)
≤44 ans	
45 – 54 ans	
55 – 74 ans	
≥ 75 ans	
Total	

Les fréquences absolues sont les valeurs obtenues par l'option PIVOT TABLE

14. Réalisez un graphique par **COLONNES** pour représenter les données collectées dans le tableau de fréquences créé précédemment. Le graphique créé devrait ressembler à celui ci-dessous :



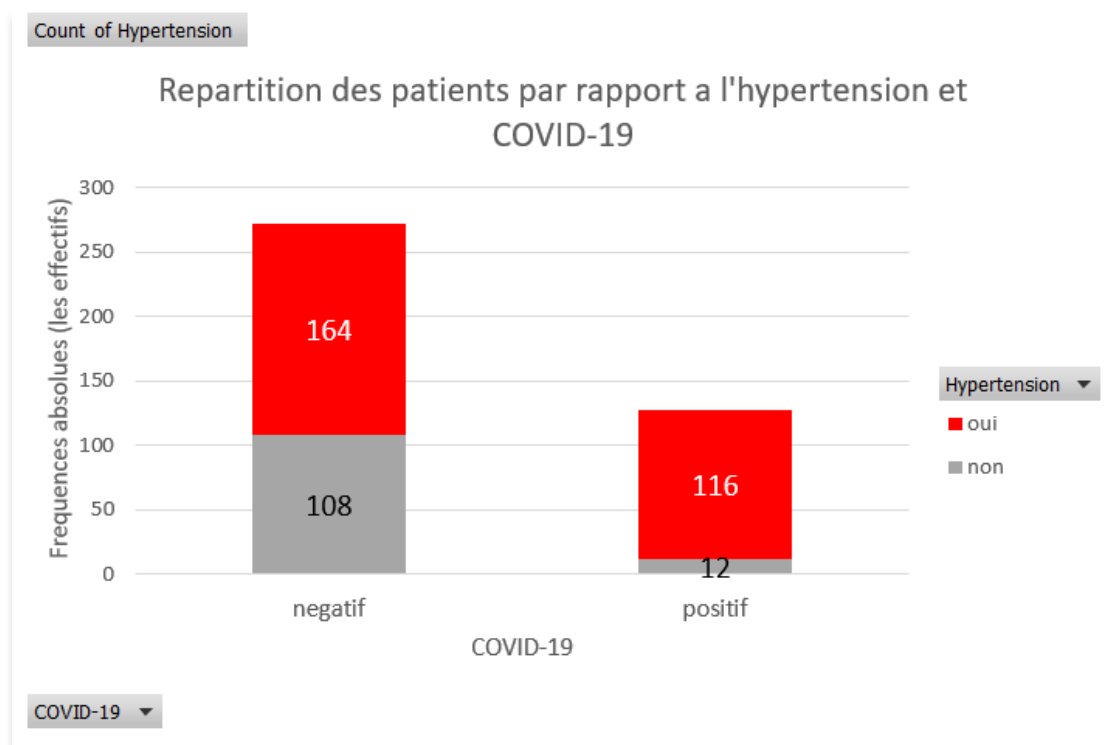
15. Au côté droit du tableau de fréquences, **calculer les fréquences relatives et les fréquences relatives cumulées** (**voir les conseils – page 4 et le cours**)

Statistiques descriptives & Representation graphique de **DEUX variables QUALITATIVES**

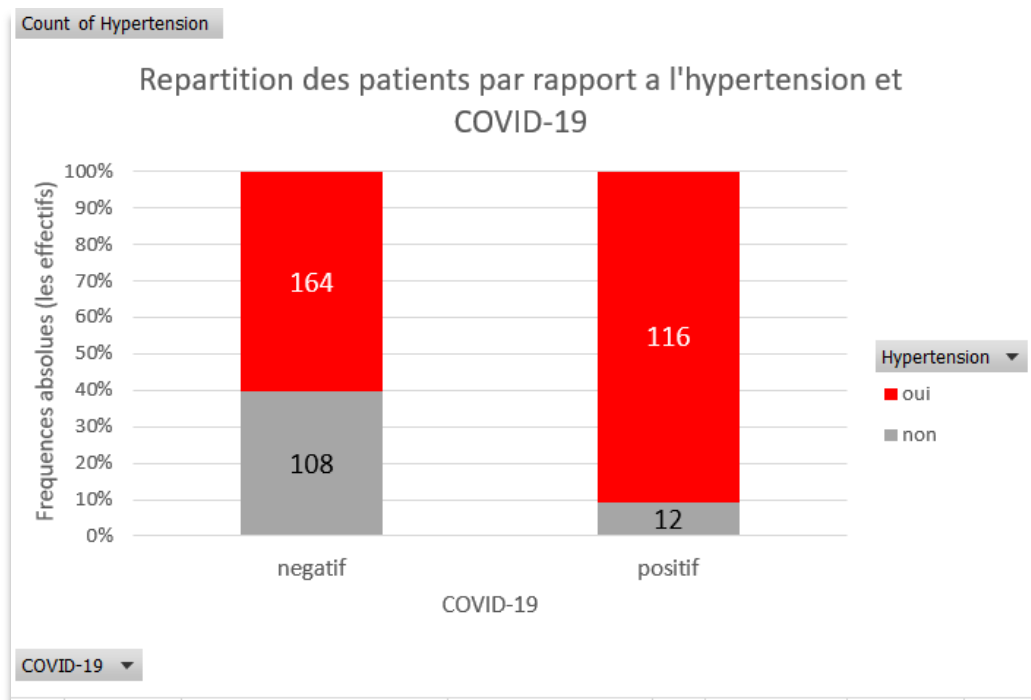
16. Dans la feuille **COLONNE_2_variables**, copier les variables **Hypertension** et **COVID-19**.
 17. Dans la même feuille **COLONNE_2_variables**, a cote des deux variables, créer une table de la forme suivante et puis a l'aide des notions du cours **remplir le tableau** :

Variable	Type de la variable (qualitative dichotomique/qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue)
Hypertension	
COVID-19	

18. Dans la même feuille **COLONNE_2_variables**, créer le **TABLEAU DE CONTINGENCE** contenant la répartition des patients par rapport par rapport aux deux variables (Hypertension et COVID-19) en utilisant l'option **PIVOT TABLE**.
 19. Réaliser le graphique par des **COLONNES** empilée la répartition des cas (les effectifs) par rapport aux deux variables (le tableau de contingence).
 20. Formater le graphique pour qu'il ressemble à celui de l'image ci-dessous:



21. En utilisant le même tableau de contingence que précédemment, créez une colonne 100% empilée. La représentation graphique doit ressembler à celle ci-dessous:



22. Peut-on dire que dans l'échantillon étudié, La fréquences de l'hypertension est différente chez les malades de COVID-19 et les non-malades ? répondez à cette question a l'aide du graphique (interpréter le graphique obtenu).

23. Dans la feuille **BARRES**, copier les variables **Comté du patient et COVID-19**.

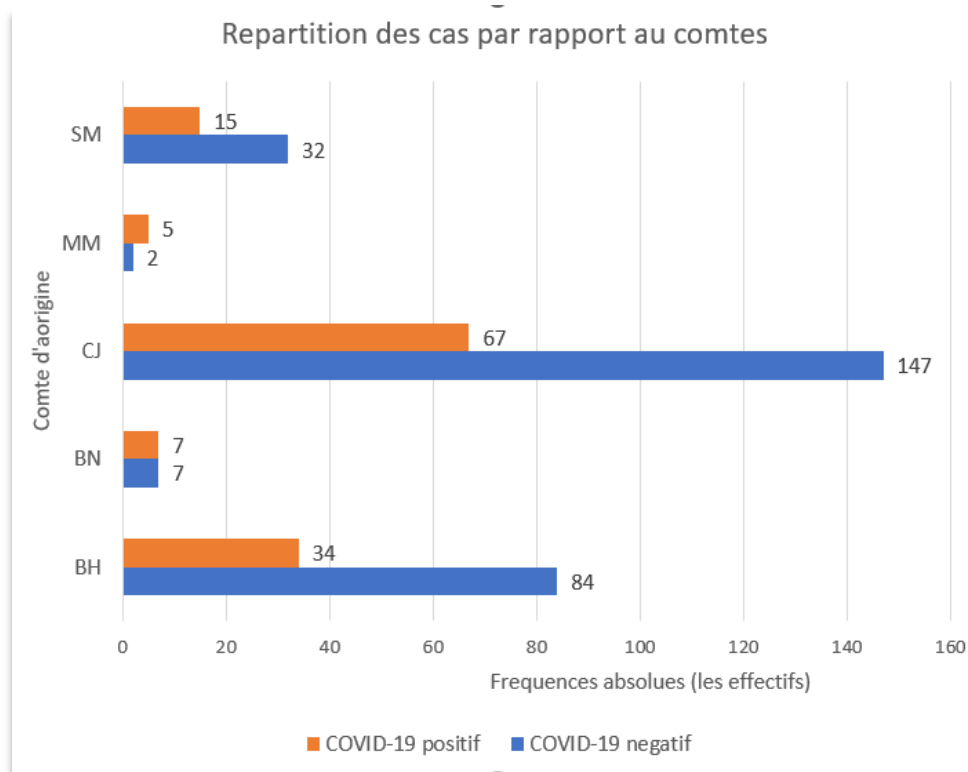
24. Dans la même feuille **BARRES**, a cote des deux variables, créer une table de la forme suivante et puis a l'aide des notions du cours **remplir le tableau** :

Variable	Type de la variable (qualitative dichotomique/qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue)
Comté du patient	
COVID-19	

25. Dans la même feuille **BARRES**, créer le tableau de contingence contenant la répartition des patients par rapport par rapport aux deux variables (Comté et COVID-19) en utilisant l'option **PIVOT TABLE**.

26. Réaliser le graphique par des **BARRES** la répartition des cas par rapport aux deux variables (**Comté du patient et COVID-19**).

27. Formater le graphique pour qu'il ressemble à celui de l'image ci-dessous:



28. Dans la nouvelle feuille **DONNEES ORDONNEES**, copier les variables **Sexe** et **Hypertension**
29. Trier les données par rapport au **Sexe** (voir les Conseils, page 10)
30. Réaliser le **tableau de contingence** en utilisant la fonction **COUNTIF** (voir les Conseils, page 11).

Probleme récapitulatif/révision

On considère la même base de données.

1. Réaliser le tableau de fréquence pour la variable **Sexe**.
2. Créer/Réaliser ensuite un graphique qui montre le pourcentage des patients par rapport au sexe dans l'échantillon d'étude.
3. Insérez une nouvelle variable appelée **Obésité**. En utilisant une **fonction prédéfinie**, établir pour chaque sujet s'il est obèse (SI l'IMC ≥ 30 kg/m²) ou pas-obèse (IMC < 30 kg/m²).
4. Réaliser un graphique approprié pour la répartition des patients par rapport aux **Obésité et COVID-19**.

Rappelez-vous que....

1. La fonction **COUNTIF** sera utilisée pour compter combien de fois une modalité/critère d'une variable qualitative est répétée.

2. L'option PIVOT TABLE sera utilisée pour compter les effectifs des patients par rapport à 2 ou plusieurs variables qualitatives. Cette option peut être également utilisée pour compter les effectifs associés à une seule variable qualitative.
3. Le graphique camembert (angl. Pie) est approprié pour la description d'une SEULE variable QUALITATIVE. Le graphique à barres ou colonnes peut également être utilisé.
4. La relation entre DEUX variables QUALITATIVES sera décrite par un graphique à barres ou colonnes.

A la fin du TP....

Le fichier sera envoyé à l'enseignant du TP par la manière/méthode établie de celle-ci mais aussi à votre adresse e-mail (cela peut être celle sur Outlook Teams `votre_nom_utilisateur@elearn.umfcluj.ro`)