

Mesures descriptives des variables QUALITATIVES

Objectives

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

Savoir calculer les statistiques descriptives (fréquences absolues, fréquences relatives) pour la description d'une seule variable qualitative (dichotomique, nominale)

Savoir calculer les statistiques descriptives (fréquences absolues, fréquences relatives, fréquences relatives cumulées croissantes) pour la description d'une seule variable qualitative ORDINALE

- 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
- Savoir les types de graphiques appropriés pour la description d'une seule variable qualitative (graphique sectorielle, graphique par de colonne/barre)
- Savoir les types de graphiques appropriés pour la description de la relation entre deux variables qualitatives (graphique par de colonne/barre)

Scenario

Une étude a été menée pour évaluer l'association entre la tranche d'âge, la fréquence de brossage dentaire, diabète et la présence de l'abcès dentaire chez les patients adultes traités à la Clinique de Chirurgie buccale-maxillo-faciale de Cluj Napoca L'étude a été réalisée sur un échantillon aléatoire de 101 sujets adultes qui ont été évalués entre le 2 février et le 2 décembre 2021.

Une fois que le traitement dentaire approprié a été effectué, les 101 patients ont été invités à indiquer leur degré d'accord avec la question suivante : « À quelle fréquence consommez-vous des boissons acides ou à forte teneur en sucre ? ». Les réponses ont été codées numériquement sur une échelle de Likert : 1 = Jamais, 2 = Rarement, 3 = Parfois, 4 = Souvent et 5 = Toujours.

Les données recueillies se trouvent dans le fichier **Donnees_TP04.xlsx**.

Demands

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 **Donnees_TP04.xlsx** dans le dossier **TP04NP**.

3. Dans le même fichier, créer plusieurs feuilles de calcul et renommez-les: Colonnes_1_variable, Colonne_2_variables, au lieu de Sheet1, Sheet2, Sheet 3.

Statistiques descriptives et Representation graphique d'une variable QUALITATIVE

4. Dans la feuille nommée **Colonne_1_variable**, copier la variable **Tranche d'age**.
5. Dans la même feuille **Colonne_1_variable**, a coté de la variable **Tranche d'age**, créer une table de la forme suivante puis a l'aide des notions du cours, remplir le tableau :

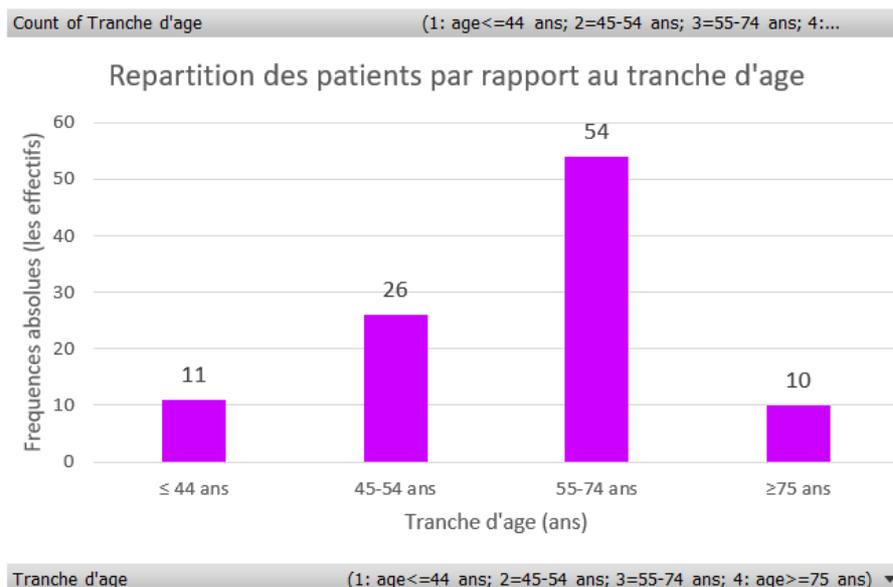
| Variable | Type de la variable (qualitative dichotomique, qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue) |
|---------------|---|
| Tranche d'age | |

6. Dans la même feuille **COLONNE**, créer un autre tableau (**TABLEAU DE FREQUENCE**) contenant les fréquences absolues, en utilisant l'option **PIVOT TABLE**.
7. Formater le tableau croisé dynamique obtenu antérieurement, pour qu'il ressemble à celui de l'image ci-dessous :

| Tranche d'age | 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

8. 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:



9. Au côté droit du tableau de fréquences, **calculer les fréquences relatives et les fréquences relatives cumulées (voir les conseils – page 6 et le cours)**
10. Dans la feuille **Statistiques_descriptives** copier la variable **Consommation boissons acides**.
11. Dans la même feuille nommée **Statistiques_descriptives**, a coté de la variable Consommation des boissons acides, 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) |
|------------------------------|---|
| Consommation boissons acides | |

12. Dans la même feuille nommée **Statistiques_descriptives**, remplir le tableau contenant d'autres statistiques descriptives appropriés aux **VARIABLES ORDINALES** comme : mode, médiane (comme mesures de tendance centrale), le premier quartile, le deuxième quartile, le troisième quartile (comme mesures de position), l'amplitude et l'intervalle interquartile (mesures de dispersion).

13. Créez une nouvelle feuille nommée **Pie** et copier la variable **Type d'abcès dentaire**.

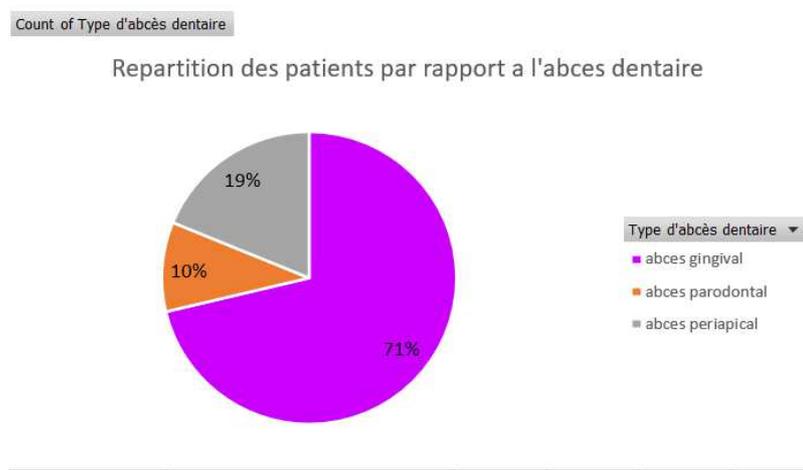
14. Dans la même feuille nommée **Pie**, a coté de la variable **Type d'abcès dentaire**, créer une table de la forme suivante **puis a l'aide des notions du cours, remplir le tableau :**

| Variable | Type de la variable (qualitative dichotomique, qualitative ordinale, qualitative nominale, quantitative discrète, quantitative continue) |
|-----------------------|---|
| Type d'abcès dentaire | |

15. Dans la même feuille Pie, créer un autre tableau (le tableau de fréquence) contenant les fréquences absolues (les effectifs) du Type d'abcès dentaire en utilisant l'option PIVOT TABLE.

16. Réalisez un graphique **CAMEMBERT** pour représenter les données collectées dans le tableau de fréquences créé précédemment.

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



Statistiques descriptives & Représentation graphique de DEUX variables QUALITATIVES

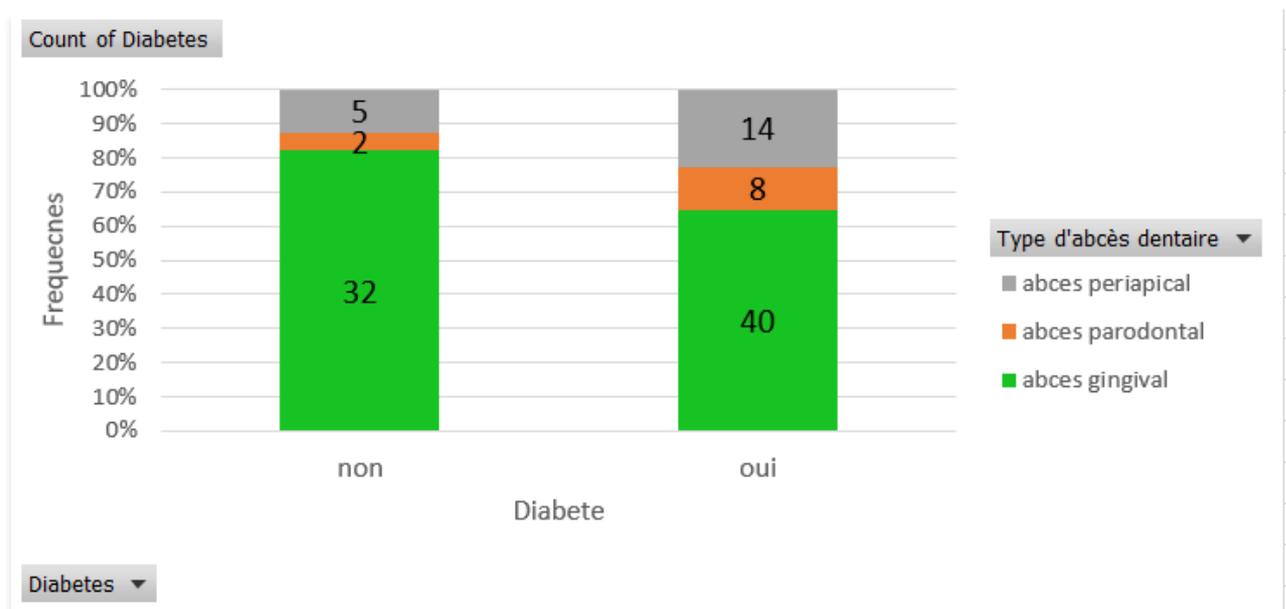
18. Dans la feuille COLONNE_2_variables, copier les variables **Diabète** et **Type d'abcès dentaire**.

19. Dans la même feuille COLONNE_2_variables, à côté des deux variables, créer une table de la forme suivante et 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) |
|-----------------------|---|
| Diabète | |
| Type d'abcès dentaire | |

20. Dans la même feuille COLONNE_2_variables, créer le **TABLEAU DE CONTINGENCE** contenant la répartition des patients par rapport aux deux variables (Diabète et Type d'abcès dentaire) en utilisant l'option **PIVOT TABLE**.

21. Réaliser le graphique par des **COLONNES** empilée OU par des **COLONNES** empilée avec des pourcentages % la répartition des cas (les effectifs) par rapport aux deux variables. La représentation graphique doit ressembler à celle ci-dessous :



22. Peut-on dire que dans l'échantillon étudié, la fréquence de **l'abcès parodontal** est différente chez les diabétiques et les non-diabétiques ? répondez à cette question à l'aide du graphique (interpréter le graphique obtenu).

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 Consommation des boissons acides**.
5. Calculer les mesures de localisation pour la variable **Fréquence de la consommation des sucreries**.

Rappelez-vous que....

1. 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 ;
2. la fonction COUNTIF (voir le TP 03) reste une autre méthode utile pour la création d'un tableau de fréquence ou d'un tableau de contingence ;
3. Le graphique camembert (engl. 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 barres ou colonnes.
5. Pour des variables QUALITATIVES ORDINALES est possible le calcul des autres statistiques descriptives telles que : la médiane, le mode, les quartiles, l'amplitude, IQR.

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)