

Estimation ponctuelle et estimation par l'Intervalle de confiance

Mihaela Iancu, Andrada Urda

Objectives

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

- ⦿ Savoir identifier si les données suivent la loi Normale de probabilité en utilisant les statistiques descriptives
- ⦿ Savoir identifier si les données suivent la loi Normale de probabilité en utilisant l'histogramme comme la méthode graphique
- ⦿ Savoir déterminer l'intervalle de confiance pour la fréquence d'une variable qualitative
- ⦿ Savoir déterminer l'intervalle de confiance pour une moyenne d'une variable quantitative

Scenario 1

On suppose qu'on a fait une étude pour évaluer la relation entre la **maladie parodontale et plusieurs facteurs sériques (calcium, hémoglobine, ferritine) chez les patients hémodialysés.**

Dans cette enquête clinique, l'échantillon est constitué des 216 patients adultes souffrant d'une maladie rénale chronique qui s'ont présentés à l'Institut clinique d'urologie et de transplantation rénale à Cluj-Napoca dans une certaine période de temps, chaque patient inclus dans l'étude ayant un minimum de six dents, des symptômes d'inflammation parodontale et un minimum de six mois passés depuis l'hémodialyse.

Les données démographiques des patients comme l'âge, le genre, le tabagisme et l'ancienneté en l'hémodialyse ont été enregistrées.

L'examen parodontal a été effectué en mesurant la profondeur de la poche (PD mesurée en mm) et la perte d'attachement clinique (CAL mesurée en mm).

La parodontite a été diagnostiquée selon les critères de l'American Academy of Periodontology: un minimum de deux sites avec $CAL \geq 4\text{mm}$ dans la zone interproximale ou au minimum 2 zones interproximales avec $PD \geq 5\text{mm}$ dans différentes dents.

Les taux sériques de calcium, d'hémoglobine et de ferritine ont été mesurés par un test sanguin de routine.

Les données sont disponibles dans le fichier **BD_TP07MDFR.xls**.

Demandes:

1. Créez un nouveau dossier **TP07NP** ou **N = votre nom, P = votre prénom** sur le bureau Bureau (Desktop) de Windows.
2. Enregistrer le fichier Excel **BD_TP07MDFR.xlsx** dans le dossier **TP07NP**.

COMMENT SAVOIR SI MES DONNEES SUIVENT UNE DISTRIBUTION NORMALE (GAUSSIENNE)?

3. Dans la feuille nommée **Statistiques descriptives**, copier la variable **Calcium**.
4. Remplir le **Tableau 1** contenant **les statistiques descriptives**, pour décider si les valeurs du Calcium suivent la **loi Normale** chez les patients hémodialysée ?
5. Dans une feuille nommée **Histogramme**, copier la variable **Calcium**, **réaliser un histogramme avec 8 classes** et décider si la variable Calcium suit la loi Normale dans l'échantillon d'étude.

COMMENT SAVOIR LA PROBABILITE D'AVOIR UNE CERTAINE VALEUR COMPRISE DANS UN INTERVALLE SI MES DONNEES SUIVENT UNE DISTRIBUTION NORMALE?

6. Si on suppose que la variable Calcium **suit la loi Normale**, remplir le **Tableau 2** trouvé sur **page Statistique descriptives pour trouver** :
 - a. Quel pourcentage de sujets ayant le taux du Calcium ≤ 8 mg/dl ? (**Conseils, page 3**)
 - b. Sachant que l'intervalle de référence (valeurs normales) du calcium est de 8-10 mg/dL, quel pourcentage de sujets ayant le taux du calcium dans l'intervalle 8 – 10 mg/dl ? (**Conseils, page 3**)
 - c. Quel pourcentage de sujets ayant le taux du calcium dans l'intervalle [moyenne-2·DS, moyenne+2·DS] ou DS = écart type (déviation standard)?? (**Conseils, page 3**)

ESTIMATION PONCTUELLE DE LA MOYENNE ET PAR L'INTERVALLE DE CONFIANCE

7. Dans une nouvelle page nommée **ICMoyenne**, copier la variable Calcium et puis déterminer l'intervalle de confiance (95% IC) pour la moyenne du calcium pour toute la population de patients hémodialysés ; pour répondre à cette question, utiliser l'option Data Analysis (**Conseils, pages 3- 5**)

ESTIMATION POCTUELLE DE LA FREQUENCE ET PAR L'INTERVALLE DE CONFIANCE

8. Dans la page nommée **ICFréquence**, copier la variable **Maladie parodontale** et calculez la fréquence relative (%) de la gingivite dans l'échantillon de patients hémodialysés (**Conseils, pages 5-8**).
9. Dans la page nommée **ICfréquence**, remplir le **Tableau 1**, pour calculer l'intervalle de confiance (95% IC) pour la **fréquence de la gingivite** dans l'échantillon (**Conseils, pages 5-8**).

PRESENTER LES RESULTATS DE L'ANALYSE SOUS POWERPOINT

1. Créer une présentation dans Microsoft **PowerPoint** en suivant les exigences ci-dessous :

- a. Ouvrir une nouvelle présentation dans PowerPoint et enregistrez votre présentation sous le nom **Résultats_TP07.pptx**. dans le dossier **TP07NP**.
- b. Choisissez un design (format) pour votre présentation.
- c. La structure de la présentation doit être comme suit :
 - ✚ **Diapositive 1** : Titre: Loi Normale & Intervalle de confiance ; Sous-titres: Votre nom
 - ✚ **Diapositive 2**: Copier le Tableau 1 de la feuille nommée Statistiques descriptives
 - ✚ **Diapositive 4** : Copier l'histogramme du Calcium et interprétez-le
 - ✚ **Diapositive 5** Copier l'intervalle de confiance a 95% de la moyenne du Calcium et interprétez-le.

Probleme récapitulatif/révision

On considère la même base de données du Scenario 1.

1. Déterminer l'**intervalle de confiance a 95% pour la moyenne** de l'hémoglobine **chez les femmes**.
2. Déterminer l'**intervalle de confiance a 95% pour la moyenne** de l'hémoglobine **chez les hommes**.
3. **En utilisant les résultats obtenus au points 1 & 2**, au risque de 5%, on peut affirmer qu'il y a une différence significative entre les moyennes de l'hémoglobine chez les femmes et chez les hommes? (*Conseils: si les intervalles 95% IC sont disjointes \Leftrightarrow différence significative ; au cas contraire, on peut pas dire s'il y a une différence significative*). Ecrivez la réponse pour le point 3 sur votre page de calcul.
4. En utilisant **une fonction prédéfinie** de l'Excel, créer une nouvelle variable appelée **Hypercalcémie** établissant pour chaque sujet s'il a un taux élevé du calcium (SI la concentration du calcium dans le sang ≥ 10 mg/dl) ou **non**.
5. Faites un graphique pour la répartition des sujets par rapport au **Tabagisme** et l'**Hypercalcémie**.
6. Calculer la probabilité pour qu'un individu pris au hasard ait l'**Hypercalcémie**.
7. On considère les événements : **M**= {avoir gingivite} et **F**= {être fumeur}. Calculer la probabilité d'avoir gingivite sachant que le patient pris au hasard est un fumeur ?
8. Vérifier par la méthode graphique si les valeurs de la **Ferritine** suivent une loi NORMALE.
9. Vérifier par les statistiques descriptives si les valeurs de la **Ferritine** suivent une loi NORMALE.
10. Représenter **graphiquement** la **dispersion autour de la médiane de la Ferritine** chez les **femmes** et les **hommes**.

Rappelez-vous que....

1. Les mesures descriptives peuvent nous aider a établir si les valeurs d'une variable QUANTITATIVE suivent la loi NORMALE de probabilité : i) la moyenne \approx la médiane \approx le mode et ii) le coefficient d'aplatissement (engl. kurtosis) et le coefficient d'asymétrie (engl. skewness) sont proches à 0 (dans l'intervalle [-1,1]).
2. La symétrie de l'histogramme peut nous aider à établir si les valeurs d'une variable QUANTITATIVE suivent la loi NORMALE

3. Si les valeurs d'une variable quantitative suivent la LOI NORMALE alors :
 - ⊙ Dans l'intervalle [moyenne $-1 \cdot DS$; moyenne $+1 \cdot DS$] on trouve $\sim 68,3 \%$ de la population ou DS = écart-type (déviatión standard)
 - ⊙ Dans l'intervalle [moyenne $-2 \cdot DS$; moyenne $+2 \cdot DS$] on trouve $\sim 95,45 \%$ de la population.
 - ⊙ Dans l'intervalle [moyenne $-1,96 \cdot DS$; moyenne $+1,96 \cdot DS$] on trouve $\sim 95 \%$ de la population.
 - ⊙ Dans l'intervalle [moyenne $-3 \cdot DS$; moyenne $+3 \cdot DS$] $\sim 99,7 \%$ de la population.
4. Estimation ponctuelle de la moyenne d'une population est la moyenne de l'échantillon
5. L'intervalle de confiance de la moyenne sera utile pour estimer la moyenne d'une population
6. Estimation ponctuelle de la fréquence d'une population est la fréquence de l'échantillon
7. L'intervalle de confiance de la fréquence sera utile pour estimer la fréquence d'une population.