

## Estimation ponctuelle et estimation par l'Intervalle de confiance

*Mihaela Iancu, Daniel Leucuța*

### Objectives

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

- ⦿ Savoir déterminer l'intervalle de confiance (IC) pour la moyenne
- ⦿ Savoir interpréter l'intervalle de confiance (IC) pour la moyenne
- ⦿ Savoir déterminer l'intervalle de confiance (IC) pour la fréquence
- ⦿ Savoir interpréter l'intervalle de confiance (IC) pour la fréquence

### Scenario 1

Une étude a été réalisée pour évaluer la relation entre l'**Indice de masse corporelle** (IMC,  $\text{kg/m}^2$ ) et la **Durée du sommeil** chez les jeunes adultes (de 20 à 29 ans).

L'étude a été menée sur un échantillon aléatoire de 248 sujets adultes.

La durée du sommeil par nuit a été évaluée par un questionnaire, chaque sujet inclus dans l'étude étant interrogé : "A quelle heure vous êtes-vous couché la nuit dernière et vous êtes-vous réveillé ce matin ?", la durée du sommeil (mesurée en heures) étant calculée comme la différence entre l'heure à laquelle vous vous réveillez et l'heure à laquelle vous vous couchez, une nuit normale de sommeil étant ainsi définie comme le nombre d'heures passées une même nuit de la semaine au fil du temps.

La durée du sommeil (heures) a été dichotomisée en une variable qualitative définie comme suit : durée du sommeil < 7 heures (durée **réduite**) respectivement durée du sommeil  $\geq 7$  heures (durée **optimale**).

Les données démographiques des sujets comme l'âge, le genre, ont été également enregistrées dans le fichier **BD\_TP08MGFR.xls** dans la feuille nommée **Scenario1**.

### Demands:

1. Créez un nouveau dossier **TP08NP** ou **N = votre nom, P = votre prénom** sur le bureau Bureau (Desktop) de Windows.
2. Enregistrer le fichier Excel **BD\_TP08MGFR.xlsx** dans le dossier **TP08NP**.

### COMMENT SAVOIR SI MES DONNEES SUIVENT UNE DISTRIBUTION NORMALE (GAUSSIENNE)? En utilisant les statistiques descriptives

3. Dans la feuille nommée **Normalité de l'IMC**, copier la variable **IMC**.

4. A cote droite de la variable **IMC**, calculer **les statistiques descriptives**, pour décider si les valeurs de l'IMC suivent la **loi Normale**. (voir **Conseils, pages 3-4**)

**L'INTERVALLE DE FLUCTUATION A 95% : [Moyenne - 1.96 · Ecart-type; Moyenne + 1.96 · Ecart-type]**

5. En supposant que la variable **IMC** suit la **loi Normale**, remplir le **Tableau 1** trouvé sur **page Normalité de l'IMC pour trouver :**
- Quel pourcentage de sujets ayant le taux de l'IMC dans l'intervalle [moyenne-1.96·DS, moyenne+1.96·DS] ou DS = écart type (déviation standard) ?? (voir **Conseils, pages 4-5**)
  - Interprétez cet intervalle. (voir **Conseils, page 5**)

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

6. Dans la page nommée **IC\_Moyenne**, copier la variable **IMC** et puis déterminer l'**intervalle de confiance (95% IC)** pour la **moyenne de l'IMC** pour toute la **population de jeunes adultes ; pour répondre à cette question, utiliser l'option Data Analysis (voir Conseils, pages 5-8 )**.
7. **Interprétez l'intervalle de confiance (95% IC)** obtenu.

#### **ESTIMATION PONCTUELLE DE LA FREQUENCE ET PAR L'INTERVALLE DE CONFIANCE**

8. Dans la page nommée **IC\_Fréquence**, copier la variable **Durée de sommeil** et calculez la fréquence relative (%) de la **durée de sommeil réduite** dans l'échantillon de sujets (voir **Conseils, pages 8-9**).
9. Dans la page nommée **IC\_Fréquence**, remplir le **Tableau 1**, calculer l'**intervalle de confiance (95% IC)** pour la **fréquence de la durée de sommeil réduite** dans l'échantillon (voir **Conseils, pages 10-11**).

#### **FACTEURS QUI PEUVENT INFLUENCER LA LONGUEUR DE L' IC**

On sait que la longueur de l'IC peut être influencer par la taille de l'échantillon (**n**), l'écart type d'échantillonnage (**S**) et le risque d'erreur (**α**).

10. Dans la page nommée **Longueur de l'IC**, remplir le **Tableau 1** en calculant **les intervalles de confiance (IC) de la moyenne** pour des différent valeurs de **n**, **S** et **α**. (voir **Conseils, pages 12-13** ).

#### **Scenario 2**

Une étude rétrospective a été menée pour évaluer les caractéristiques cliniques/paracliniques chez des patients adultes atteints de pneumonie causée par une infection au COVID-19. Les données des patients ont été enregistrées dans le fichier **BD\_TP08MGFR.xls** dans la feuille nommée **Scenario 2**.

## ESTIMATION PONCTUELLE DE LA DIFFERENCE DE MOYENNES (ECHANTILLONS APPARIES) ET PAR L'INTERVALLE DE CONFIANCE

*Objectif de l'étude* : Nous voulons voir si la **moyenne de la vitesse de sédimentation érythrocytaire (VS) à l'admission** à l'hôpital est **significativement différente** de la **moyenne de vitesse de sédimentation érythrocytaire (VS) au moment de la sortie de l'hôpital** chez les patients avec pneumonie causée par COVID-19?

1. Dans la feuille **Scenario 2** créez une nouvelle variable nommée **DifférencesVS** en faisant la différence entre la valeur VS à l'admission et la valeur VS au moment de la sortie de l'hôpital
2. Utiliser l'option Data Analysis pour faire les statistiques descriptives pour la nouvelle variable **DifférencesVS** (voir **Conseils**)
3. **Calculer 95% IC** pour la différence de moyennes. (voir exercice 6, Scenario1 ou **Conseils pages 14-16**)
4. **Interpréter 95% IC** pour la différence de moyennes. (**Conseils page 16**)

### 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'IMC** chez les **sujets ayant une durée réduite de sommeil**.
2. Déterminer l'**intervalle de confiance a 95%** pour la **moyenne de l'IMC** chez les **sujets ayant une durée optimale de sommeil**.
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'IMC chez les jeunes adultes ayant une durée réduite de sommeil et ceux qui ont une durée optimale ? (*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 **Etat de surpoids ou d'obésité** établissant pour chaque sujet s'il a un taux élevé de l'IMC (SI  $IMC \geq 25 \text{ kg/m}^2$ ) ou **non**.
5. Faites un graphique pour la répartition des sujets par rapport au **Sexe** et **Etat de surpoids ou d'obésité**.
6. Calculer la probabilité pour qu'un individu pris au hasard être en surpoids ou obésité.
7. On considère les événements : **M**= {être en surpoids ou obese} et **D**= {avoir une durée réduite de sommeil}. Calculer la probabilité **Pr (M/D)**.
8. Représenter **graphiquement** la **dispersion autour de la médiane de la IMC** par rapport à la **durée de sommeil**.

### Rappelez-vous que....

1. Les mesures descriptives peuvent nous aider à 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·DS; moyenne +1·DS] on trouve ~68,3 % de la population ou DS = écart-type (déviations standard)
  - ⊙ Dans l'intervalle [moyenne -1.96·DS; moyenne +1.96·DS] on trouve ~95 % de la population.
  - ⊙ Dans l'intervalle [moyenne -2·DS; moyenne +2·DS] on trouve ~95,45 % de la population.
  - ⊙ Dans l'intervalle [moyenne -3·DS; moyenne +3·DS] ~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.
8. **NE** pas confondre L'INTERVALLE de FLUCTUATION [MOYENNE -1.96·DS; MOYENNE +1.96·DS] avec L'INTERVALLE DE CONFIANCE !!!!