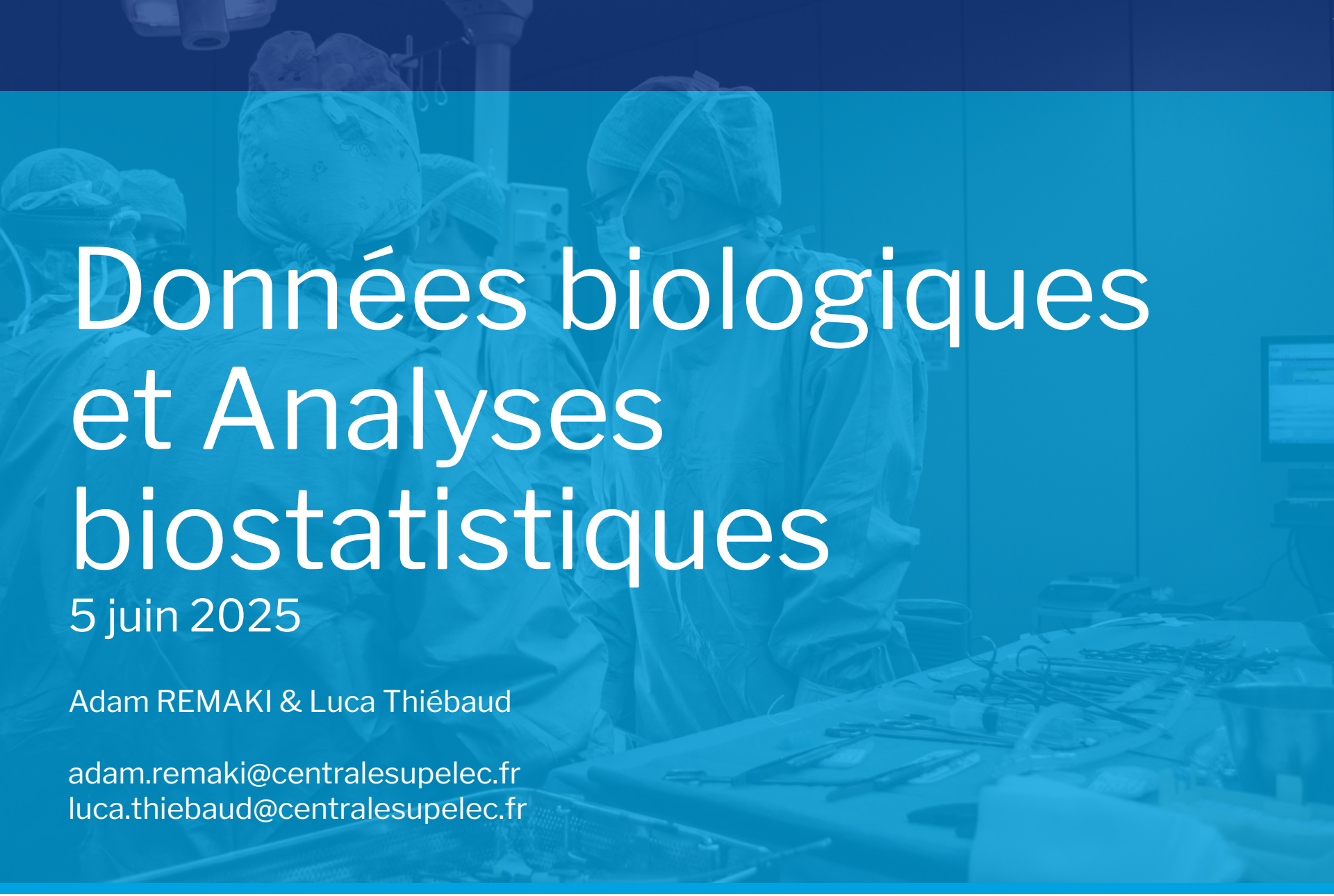


Données biologiques et Analyses biostatistiques

5 juin 2025

Adam REMAKI & Luca Thiébaud

adam.remaki@centralesupelec.fr
luca.thiebaud@centralesupelec.fr

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Plan

- **Les données de biologie à l'AP-HP**
  - Exemple d'intégration d'une analyse biologique
  - Les différents référentiels de concepts
  - L'uniformisation des mesures via bio_clean
- **Rappels de biostatistiques**
- **Application 4 : Analyses biostatistique sur l'effet du médicament B**

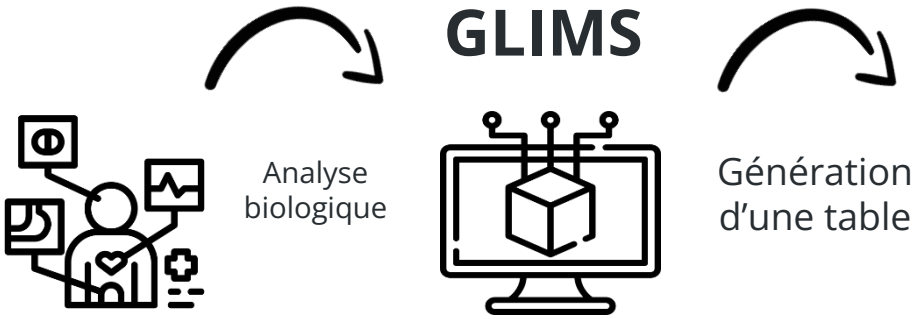
1.

Les données de biologie à l'AP-HP

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Les données de biologie à l'AP-HP

Intégration « type » d'une analyse de laboratoire

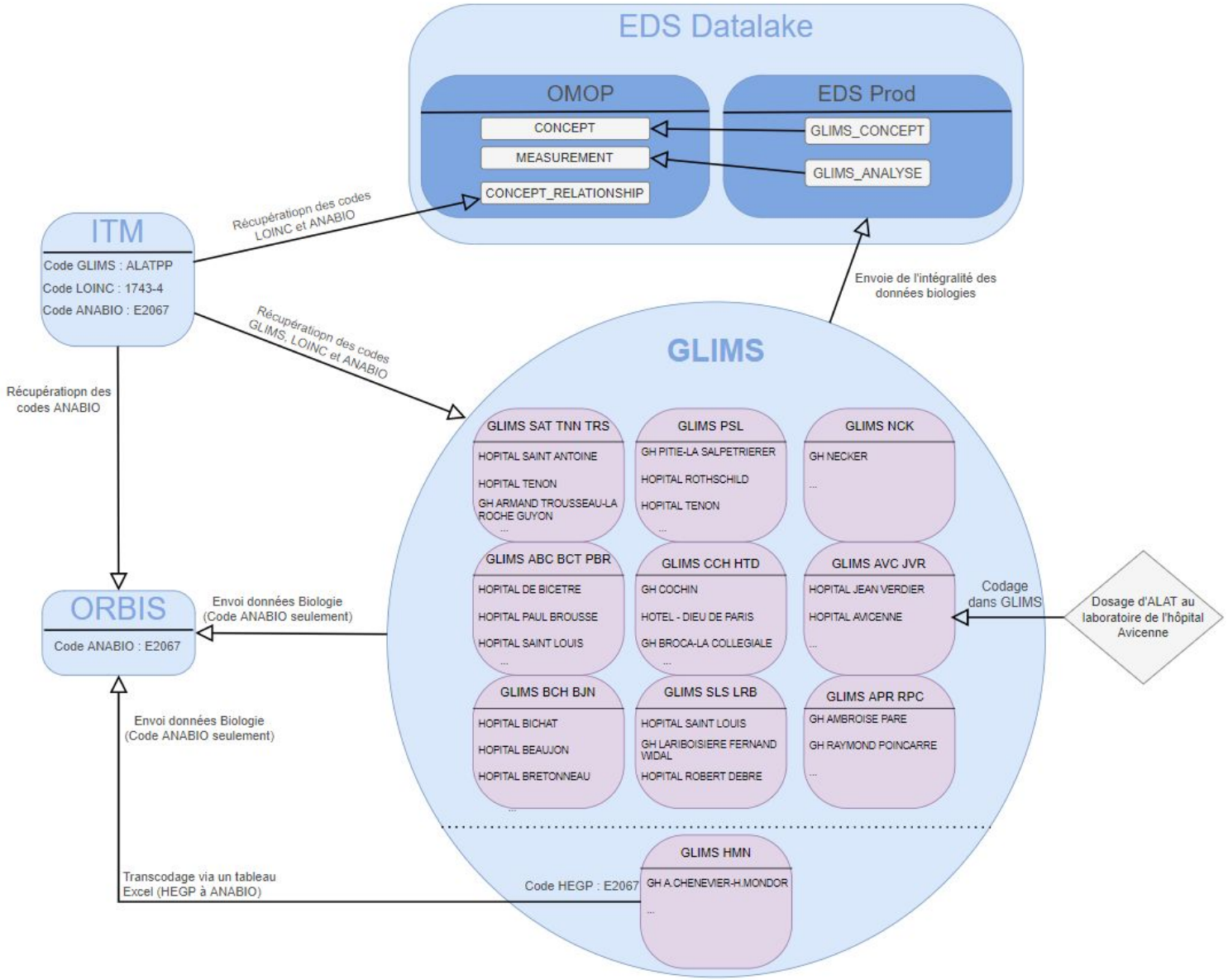


id	concept_id	value	unit
1	XXX_YY_1	13	g/dL
2	XXX_ZZ_1	1.5	mmol/L

En pratique :

- Autant de logiciels GLIMS que d'hôpitaux
- Transcodage vers d'autres référentiels (ANABIO et LOINC)
- Besoin d'agréger les concepts similaires en méta-concepts et de les uniformiser.

Les différents référentiels de concepts



Les données de biologie à l'AP-HP

- **Uniformisation via *bio_clean***
  - Regroupement de concepts correspondant à la même méta-analyse
    - *Ex : Hb : A0163, I7893, Urée : A0286, G3350*
  - Uniformisation des unités
  - Seuillage pour limiter les valeurs aberrantes
  - Mise au format df_bio

- **Mais travail lourd et manuel :**
  - 400M analyses biologiques
  - 50k concepts dans le référentiel Anabio

- **Dans le cadre du TD et du projet, on considère que le pré-traitement (conversion et seuillage) a déjà été effectué, et que vous disposez de la table *df_bio* directement exploitable.**

2.

Rappels de statistiques

Comparaison de deux populations

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Test d'hypothèse

Idée : Comparer deux populations distinctes selon un critère de jugement

1. Le critère est une variable qualitative (donnée discrète)
   Test du Chi2

2. Le critère est une variable quantitative (donnée continue)
   Test de Student

Critère de jugement est une variable qualitative

Exemple : antécédent de diabète – OUI/NON

1. Dénombrement du critère de jugement dans chaque population

	Population 1	Population 2
Présence de diabete	40	50
Absence de diabete	60	50

Le taux de personnes diabétique est-il plus important dans la population 2 que dans la population 1?

- **Critère de jugement est une variable quallitative**

Exemple : antécédent de diabète – OUI/NON

2. On pose les 2 hypothèses

H0 : Le taux de personnes diabétique le même dans les 2 populations

H1 : Le taux de personnes diabétique différent dans les 2 populations

Si H0 est vraie : les fluctuations observées sont uniquement dues au hasard
Si H0 est rejetée : les populations sont différentes

- **Critère de jugement est une variable quallitative**

Exemple : antécédent de diabète – OUI/NON

3. On calcule le taux de personnes diabétiques théorique sous H0 (populations semblables)

	Population 1	Population 2	Population 1+2
Présence de diabete	*45*	*45*	90 (45%)
Absence de diabete	*55*	*55*	110 (55%)
TOTAL	100	100	20

Critère de jugement est une variable qualitative

Exemple : antécédent de diabète – OUI/NON

4. On compare les taux théoriques aux taux réels

Réel	Population 1	Population 2
Présence de diabète	40	50
Absence de diabète	60	50

Théorique	Population 1	Population 2
Présence de diabète	*45*	*45*
Absence de diabète	*55*	*55*

--> On pose une VA qui suit une loi du Chi2 :

Effectifs réels

Effectifs théorique

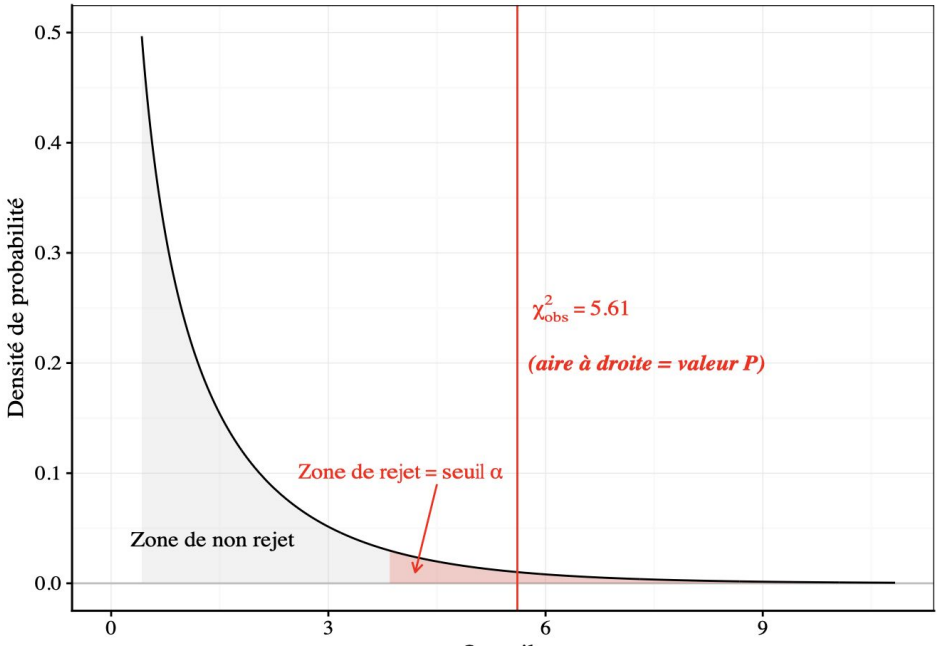
$$X^2_{dl} = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

Critère de jugement est une variable qualitative

Exemple : antécédent de diabète – OUI/NON

5. On choisi le seuil a partir duquel on rejette l'hypothèse nulle

Densité de probabilité d'une loi du Chi2



Pour α = 0.05, la borne vaut 3,84

- **Critère de jugement est une variable qualitative**

Exemple : antécédent de diabète – OUI/NON

6. On conclut pour $\alpha = 0.05$

Si X > 3,84 alors on rejette H0

Si X < 3,84 alors on accepte H0

Application a notre exemple :

$$\chi = \frac{(40-45)^2}{45} + \frac{(60-55)^2}{55} + \frac{(50-45)^2}{45} + \frac{(50-55)^2}{55}$$
$$= 2.02$$

- **Critère de jugement est une variable qualitative**

Exemple : antécédent de diabète – OUI/NON

7. On calcule le seuil α minimum pour rejeter H0 selon notre valeur de χ

Avec χ = **2.02,** α = 0.18

On appelle p-value le seuil α

3.

Application 4

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Application 4 – Analyses biostatistiques

- **Pull le repo GitHub pour mettre à jour le projet**
  - Lien URL : https://github.com/Aremaki/edstuto_2025

- **Ouvrir le notebook *exercises/exercise-4***
  - Contexte : Explorer les facteurs de risque de décès dans une sous cohorte ayant pris le médicament B.
  - Objectif : Analyses biostatistiques pour comparer les populations.

4.

Projet

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Projet – Analyses biostatistiques & article

Objectifs du jour

- Rassembler tous les facteurs de risque dans une unique table d'analyse
- Réaliser les analyses biostatistiques répondant à la question scientifique
- Finalisation du notebook
- Ecriture de l'article

Données

- Base de données disponible dans le dossier *final_project/data*

Reproductibilité : vos notebooks seront lancés sur un set de données de test non fourni et doivent fonctionner tels quels.

Compétences évaluées :

- **Compétence C4** : Avoir le sens de la création de valeur
- **Compétence C6** : Être opérationnel, responsable et innovant dans le monde numérique
- **Compétence C7** : Savoir convaincre
- **Compétence C8** : Mener un projet en équipe