

Projet d'Ingénierie des Données

Contexte : Analyse de survie de patients atteints de myélome. Analyse de survie et coût efficacité

Documents :

Diapositives sur l'analyse de survie

Librairie Python : lifelines ou lifetimes

Travail à faire

Etudier les fonctionnalités de la librairie « lifelines » et lifetimes qui permettent de faire de l'analyse de survie.

Dresser la liste des fonctions avec un bref descriptif

Fournir une interface (Web) permettant d'afficher les résultats de l'analyse selon les recommandations ci-dessous.

Fichier des données :

Ajouter une nouvelle variable Temps jusqu'à l'hospitalisation

Créer une nouvelle variable tranche d'âge avec les modalités suivantes : (Ages < 50 ; Ages 50-64 ; Ages 65+)

Ajouter des lignes dans le fichier de données avec des valeurs fictives (comparables aux données existantes). Il faut atteindre environ 300 patients.

Vérifier dans le fichier de données qu'un patient n'est présent qu'une fois dans le fichier. Supprimer les lignes en doublons pour les patients avec un « Event=0 ».

Fonctionnalités de l'application

Gestion des données manquantes. Remplacer les données manquantes selon le cas (suppression de ligne, de colonne, remplacement par une valeur moyenne ou médiane, ...)

Estimation et affichage de la fonction de survie (Kaplan-Meier) avec et sans intervalle de confiance

Estimation de la fonction risque ou Hazard rates (Nelson-Aalen, modèle de Weibull)

Prédiction : Estimation de la survie pour quelques jours.

Estimation et affichage de la fonction de survie en fonction du sexe, de l'âge, du TtoMM1, de l'anémie, de l'hypercalcémie, de

Observer la différence et comparer les données de survie (par exemple H, F, ou tranches d'âge, ...) à l'aide de tests statistiques.

Présentation de la page Web

Une barre latérale contenant les fonctionnalités de sélection des variables et des sous-groupes d'individus.

Une barre horizontale (en haut de la page) contenant les onglets des analyses à faire :

- Lecture des données
- Traitement des données manquantes
- Statistiques descriptives
- Représentations graphiques des variables
- Probabilités de survie et courbes de survie
- Prédiction de survie d'un individu
- Modèle de régression de Cox
- Analyse coût-efficacité

Facultatif

Pour plus de convivialité, vous pouvez utiliser la librairie « Streamlit » qui permet de faciliter l'affichage sur une page Web et ajouter des boutons, des listes déroulantes,

Une autre librairie peut être utilisée : « plotly » couplée à la librairie « Dash ».

Vous pouvez explorer et utiliser d'autres librairies utiles pour le projet.

Variables à utiliser en tant que filtres :

Gender

Body Mass Index

Insurance Plan

Anemia

Fragility

FISH(del17p1)

Subclasification by the platform

ISS by the platform

Multiple Myeloma Treatment 1

Multiple Myeloma Treatment 2

Travail à faire à 3 personnes maximum