

**IFT1169 – TRAVAIL PRATIQUE #1 – 25 septembre 2022****« Science des données »****1<sup>re</sup> partie****Mohamed Lokbani**

---

**Équipes :** le travail est à faire en monôme (une seule personne).

**Remise :** une seule remise est à effectuer par voie électronique le samedi 15 octobre 2022, 23h59 au plus tard, sans possibilités de prorogation.

**Conseils :** n'attendez pas le dernier jour avant la remise pour engager votre travail. Vous n'aurez pas le temps nécessaire pour le réaliser.

---

**But :** ce TP a pour but de vous faire pratiquer la conception orientée objet, la surcharge des opérateurs, la gestion des flux et l'utilisation d'un makefile.

Ce travail est la première partie d'un projet de session à réaliser en 3 étapes. Vu que les étapes se suivent, il est donc important de réaliser chacune d'elle avant de pouvoir passer à la suivante.

**Énoncé :** le diabète a un impact sur la santé. Il se manifeste par des lésions d'organes tels la rétine, les reins, etc. Nous allons examiner une étude sur la progression de la maladie pour 442 malades. Cette étude contient 442 enregistrements (un enregistrement par malade), 10 caractéristiques et une cible.

Les 10 caractéristiques représentent différentes mesures relevées durant cette étude. Elles sont :

- âge : l'âge du malade en années
- Sexe : le sexe du malade, 1 pour les hommes, 2 pour les femmes.
- bmi : l'indice de la masse corporelle (poids en kg / (taille en m)<sup>2</sup>)
- bp : pression artérielle moyenne
- s1 : tc, cholestérol total dans le sérum
- s2 : ldl, lipoprotéines à faible densité
- s3 : hdl, lipoprotéines à haute densité
- s4 : tch, cholestérol total dans HDL
- s5 : ltg, le log du niveau de triglycérides dans le sérum
- s6 : glu, le taux de sucre

La cible :

Elle représente le niveau de progression sur une année par rapport à une mesure référentielle. Une progression de 200 et plus est considérée comme étant élevée.

Ces données ont été utilisées dans l'étude décrite dans cet article :

[https://hastie.su.domains/Papers/LARS/LeastAngle\\_2002.pdf](https://hastie.su.domains/Papers/LARS/LeastAngle_2002.pdf)

Vous allez devoir d'abord lire les données, les organiser dans une approche orientée objet et finalement d'afficher certaines caractéristiques.

**L'utilisation de la librairie STL n'est pas permise à ce stade.** Par contre, l'utilisation de la classe « string » est autorisée.

**Lecture des données :** commencer par lire le contenu du fichier « diabetes.tab.txt ». La première ligne contient le nom des 11 paramètres (les 10 caractéristiques + la cible) de l'étude. Chaque paramètre est séparé par une tabulation.

Le nom du fichier n'est pas figé. Il est fourni au programme, en ligne de commandes comme suit :

```
tp1.exe nom_fichier_entree
```

Si l'argument n'est pas disponible, on sort du programme avec un message d'erreur. Ce message signale la nécessité d'avoir le nom du fichier en ligne de commandes.

**Approche orientée objet :** organiser la structure des classes pour représenter au mieux ces données ainsi que le calcul demandé. On vous laisse choisir la représentation. Cette dernière doit-être expliquée dans le rapport final accompagnant ce travail pratique. Le contenu du rapport est décrit à la fin de ce document.

**Propriétés des données :** calculer la moyenne et la déviation standard de chaque caractéristique.

**Affichage :** après avoir chargé les données et effectué les différents calculs demandés précédemment, place maintenant à l'affichage en sortie.

Vous allez afficher le menu suivant :

Sélectionner un nombre de la liste suivante :

- [1] Description de l'ensemble de données
- [2] L'entête des données
- [3] Queue des données
- [4] Propriétés des données
- [5] L'ensemble des données

1. Description de l'ensemble de données : une description de l'étude ainsi que ses propriétés.
2. L'entête des données : les 5 premières lignes ainsi que la ligne d'entête.
3. Queue des données : les 5 dernières lignes ainsi que la ligne d'entête.
4. Propriétés des données : la moyenne et la déviation standard de chaque caractéristique.
5. L'ensemble des données : les 442 enregistrements.

Chacun des affichages demandés est fourni sur le site web du travail pratique. Il vous est demandé d'obtenir exactement le même affichage. Vous allez utiliser pour cela les propriétés de l'affichage

nouvellement introduites dans le standard C++20. Ces propriétés ont été décrites dans la séance de démonstration #02.

**Aide** : compléter le programme avec une option d'aide.

tp1.exe --help ou tp1.exe -h

Ces deux options font la même chose. Elles affichent une courte description comment utiliser votre programme.

### **Fichiers à remettre :**

Vous devez regrouper l'ensemble de ces fichiers dans « TP1\_A22.zip » et le remettre par l'entremise du système Studium, dans le dépôt « tp01 ».

Le fichier « TP1\_A22.zip » va contenir :

L'ensemble des programmes (\*.h/\*.cpp) relatifs à ce travail pratique.

Le makefile permettant de compiler simplement ce travail pratique.

« TP1\_A22 \_Rapport.pdf » : le rapport décrivant le travail effectué. Vous allez trouver sur le site web du cours, dans la section « Foire aux questions », les détails qui doivent figurer dans un rapport.

« TP1\_A22 \_feuille2route.pdf » : une auto-évaluation du travail réalisé.

La correction va se faire sur un poste de la DESI en utilisant le compilateur « g++ » (12.x). Vérifiez que le code que vous avez produit (devant normalement fonctionner correctement chez vous) fonctionne aussi bien sur les ordinateurs de la DESI.

**Barème** : ce TP1 est noté sur **11 points**. Les points sont répartis comme suit :

Affichage : 5 points (affichage en sortie, aide, gestion de l'erreur)

P.O.O : 3 points

Rapport : 1.5

Avis global : 1.5

En plus du précédent barème, vous risquez de perdre des points dans les cas suivants ....

- La non-remise électronique (volontaire ou par erreur) est sanctionnée par la note 0.
- La non-remise du fichier « TP1\_A22\_feuille2route.pdf » : -1.
- Un programme qui ne compile pas : 0.
- Un programme qui compile, mais ne réalise pas les choses prévues dans la spécification : 0.
- Les avertissements (warnings) non corrigés : cela dépend de la quantité! À partir de -0.25 et plus.
- Aberration dans le codage : même si tous les chemins mènent à Rome, faites l'effort nécessaire pour

éviter de prendre le plus long!

Des questions à propos de ce TP? Une seule adresse : [dift1169@iro.umontreal.ca](mailto:dift1169@iro.umontreal.ca)

Pour faciliter le traitement de votre requête, inclure dans le sujet de votre courriel, au moins la chaîne [IFT1169] et une référence au tp01.

**Mise à jour**

25-09-2022 diffusion de l'énoncé.

---

**Rapport**

Vous devez remettre un rapport complet avec ce travail. Ce rapport va expliquer votre démarche. Quelles sont les difficultés rencontrées et comment les avez-vous résolues? Le rapport ne doit pas contenir votre programme! Telle ligne fait telle opération. Les commentaires doivent être intégrés au code.

Le rapport a deux objectifs :

1- aider une personne qui cherche à faire la même chose que vous, à mieux comprendre votre programme. En lisant votre code, bien commenté, et votre rapport, bien rédigé, elle aura une vision globale de votre travail. Elle peut mieux comprendre votre programme pour l'utiliser tel quel ou lui intégrer plus facilement de nouvelles fonctionnalités.

2- aider le correcteur à mieux comprendre votre démarche. Vous avez décidé d'une démarche, à première vue ce choix peut paraître étrange. Votre rapport va permettre de mieux appuyer le pourquoi de ce choix. Ce dernier sera ainsi mieux compris.