

## IFT3913 QUALITÉ DE LOGICIEL ET MÉTRIQUES – AUTOMNE 2020 – TRAVAIL PRATIQUE 2

Dans ce TP, vous allez analyser un ensemble de données obtenues lors de la mesure d'un logiciel hypothétique.

### CONDITIONS DE RÉALISATION

- Travail à remettre le **vendredi 13 novembre 23h59** via StudiUM. Aucun retard ne sera accepté.
- Travaillez en groupes de 2, Le membre de l'équipe qui ne soumet pas le fichier Zip principal de la soumission doit soumettre un zip avec un fichier contenant les noms de l'équipe, comme dans TP1.
- Vous pouvez utiliser n'importe quel logiciel qui appuie les opérations statistiques requises (Excel, R, Octave, SPSS, Python, Java, ...). Cependant, nous n'offrons pas de conseils sur l'utilisation des logiciels statistiques. Ajoutez votre code/script/fichiers dans le Zip de votre soumission.
- Le livrable le plus important est le **rapport**, en format PDF. Assurez-vous de communiquer clairement.

### TRAVAIL À RÉALISER

Ce travail a pour objectif **d'étudier** l'influence de la structure d'une classe sur le *nombre d'erreurs de conception* (NEC) attribuables à cette classe. Pour mener cette étude, vous devez utiliser l'échantillon de données fourni dans la section 5. Cet échantillon donne les valeurs de 3 des métriques de structures collectées sur un ensemble de classes : NOM (nombre de méthodes locales et héritées de la classe), DIT (profondeur de la classe dans l'arbre d'héritage) et CAC (Nombre d'associations (incluant les agrégations) locales/héritées auxquelles participe une classe). Il donne ainsi les valeurs de la métrique du nombre d'erreurs (NEC). Pour cet échantillon, on suppose que les valeurs des métriques NEC, NOM, DIT, CAC sont normalement distribuées. Les tâches à effectuer :

- T1. (15%) **Visualisez** chacune des métriques de l'échantillon en créant les boîtes à moustaches. Calculez les informations pertinentes et décrivez les distributions.
- T2. (25%) **Évaluer l'hypothèse** selon laquelle les classes ayant un nombre de méthodes inférieur ou égal à 30 ont moins d'erreurs que celles ayant plus de 30 méthodes. Décrire d'abord la conception de l'étude et discuter par la suite les résultats. Suivez les étapes d'une étude empirique (choix d'étude, énoncé des hypothèses, définition des variables, interprétation et généralisation des résultats, discussion des menaces à la validité).
- T3. (15%) **Étudier les corrélations** individuelles entre les métriques de structure et NEC. Visualisez les données, les droits de régression, etc., et expliquez pourquoi (ou pourquoi pas) ces visualisations sont significatives (ou pas). Dans cette étape, vous ne prenez pas de décisions: vous explorez et vous étudiez l'ensemble de données. Explorez aussi le scénario où vous supprimez les valeurs aberrantes, si cela a du sens.
- T4. (30%) **Évaluer les hypothèses suivantes** :
  - a. Le nombre d'erreurs est une fonction linéaire du NOM
  - b. Le nombre d'erreurs est une fonction linéaire du DIT
  - c. Le nombre d'erreurs est une fonction linéaire du CACDécrire d'abord la conception de l'étude (comme en T2) et discuter les résultats par la suite.
- T5. (5%) **Décrivez vos conclusions** dans un court paragraphe.

La qualité du rapport vaut 10%.

## ÉCHANTILLON DE DONNÉES

Classes	NOM	DIT	CAC	NEC
C1	7	2	2	4
C2	8	1	1	4
C3	8	2	2	3
C4	9	1	1	2
C5	10	1	1	7
C6	11	2	2	3
C7	12	1	1	5
C8	12	5	5	3
C9	15	2	3	4
C10	15	1	1	5
C11	17	4	4	4
C12	18	2	2	1
C13	22	1	1	1
C14	25	6	6	0
C15	26	2	1	1
C16	29	2	2	6
C17	40	3	3	6
C18	48	4	1	4
C19	54	7	7	6
C20	56	5	5	6
C21	63	9	17	4
C22	66	3	1	2
C23	78	3	5	1
C24	88	3	3	4
C25	93	4	7	6
C26	104	5	5	5
C27	120	5	1	6
C28	130	6	6	7
C29	147	8	8	6
C30	184	9	1	4