

Université de Montréal

Rapport du travail 2

Par
Geneviève Paul-Hus (20037331)
Jean-Claude Desrosiers (20150403)

Faculté art et sciences

Travail présenté à Michalis Famelis
Dans le cadre du cours IFT3913

18 Mars 2022

Table des Matières

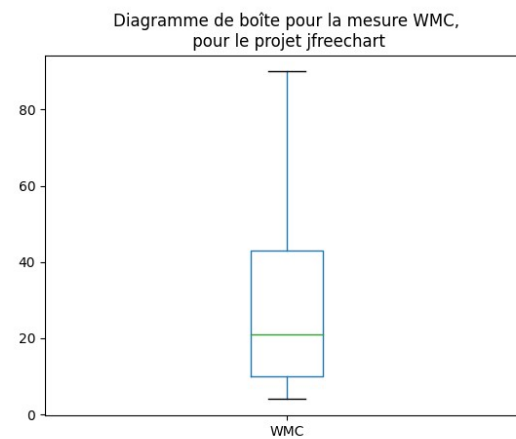
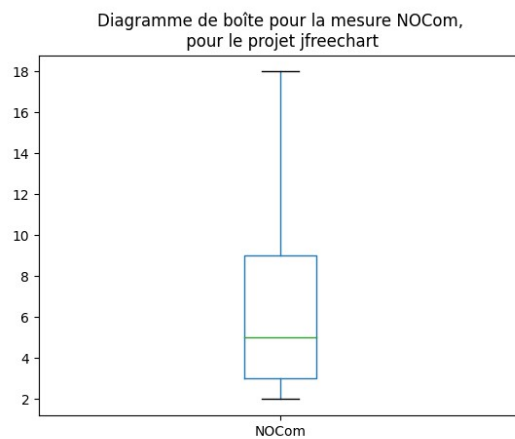
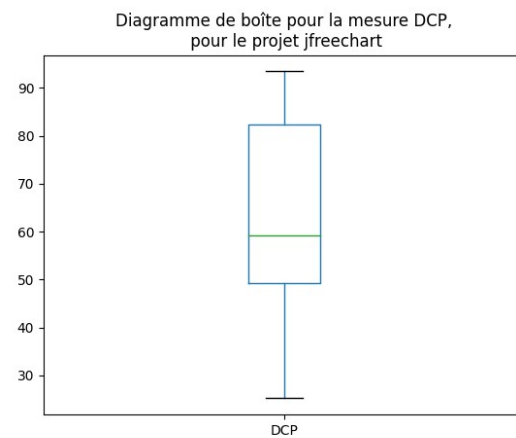
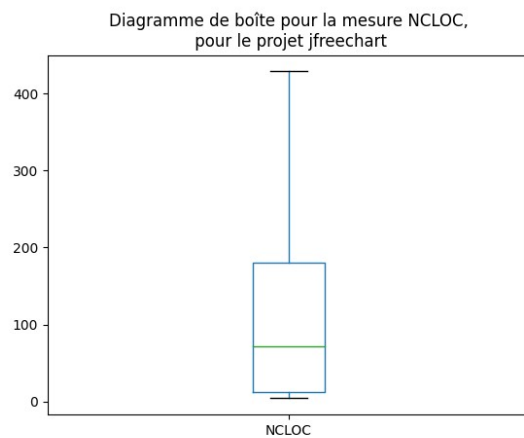
Tâche 1.....	3
Tâche 2.....	5
Tâche 3.....	6
Tâche 4.....	7
WMC et NCLOC.....	8
WMC et DCP.....	8
WMC et NOCom.....	8
Tâche 5.....	9

Tâche 1

En analysant le fichier jfreechart-stats.csv, on obtient les statistiques suivantes:

Statistiques	NCLOC	DCP	NOCCom	WMC
Médiane	71.5	59.21	5	21
Quartile supérieur	180	82.36	8	43
Quartile inférieur	12	49.26	3	10
Longueur	168	33.1	5	33
Limite supérieure	432	132.01	15.5	92.5
Limite inférieure	0	0	1	1

Et on obtient les boîtes à moustache suivantes:



On remarque que les mesures (NCLOC, DCP, NOCom, WMC) ne possèdent pas de distribution normale, elles sont toutes asymétriques. On remarque que 75% des classes ont moins de 180 lignes de code, ce qui indique des classes concises, et que 25% ont plus de 180 lignes de code. Pour le nombre de lignes de code, on observe que la complexité (WMC) est très élevée pour 25% des classes avec une complexité supérieure à 43 ce qui indique soit un très grand nombre de classes, soit beaucoup de boucles, ou les deux. On remarque que 75% des classes ont plus de 50% de commentaires, et qu'elles ont changées 8 ou moins de fois selon le nombre de commits dans l'historique.

Tâche 2

Choix d'étude: Étude de cas

Hypothèse: Les classes du projet jfreechart qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois.

Variables:

- **Indépendante:** Nombre de modifications d'une classe
- **Dépendante:** Densité de commentaires d'une classe

Objets: classes du projet jfreechart

Sujets: calculateurs

Procédure: Prendre le dépôt de code jfreechart, prendre les données que nous avons sur la densité de commentaire et sur le nombre de modifications de classes. Évaluer la relation entre les deux types de données.

Interprétation et généralisation des résultats : Nous avons calculé le coefficient de corrélation de Pearson entre les colonnes DCP et NOCom et avons obtenu un score de -0.487753 ce qui nous indique une faible corrélation. Nous avons aussi calculé des statistiques de base (moyenne géométrique, maximum et minimum) des données que nous avons séparé en deux collections (classes ayant subies 10 modifications ou plus, et les classes ayant subies moins de 10 modifications). Voici les statistiques que nous avons obtenues :

Statistiques	DCP des classes avec NOCom \geq 10	DCP des classes avec NOCom < 10
Moyenne Géométrique	48.24%	65.08%
Minimum	32.01%	25.24%
Maximum	88.19%	93.44%

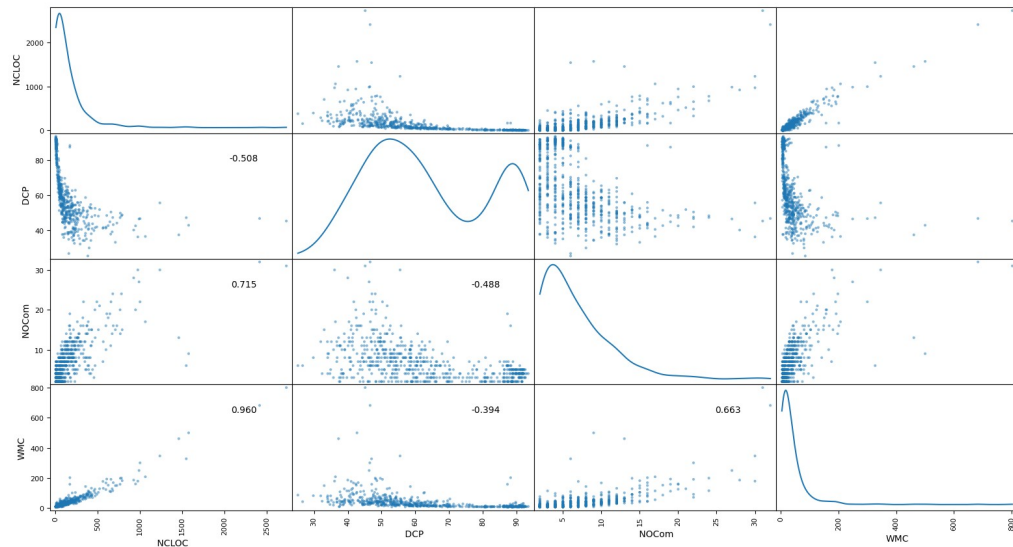
Comme nous remarquons, la moyenne géométrique de la densité de commentaires dans les classes ayant moins de 10 modifications est plus élevée de 16.84% par rapport aux classes ayant 10 modifications et plus. Avec les données que nous avons obtenues, nous pouvons donc infirmer l'hypothèse pour ce cas.

Menaces à la validité:

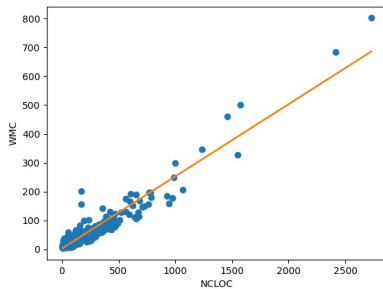
- **Validité de construction**
 - La façon dont est bâti l'étude aurait pu être manipulée pour valider l'hypothèse.
- **Validité interne**
 - Puisque nous sommes nous-mêmes les sujets de notre étude, nous aurions pu ignorer des données importantes non-intentionnellement pour valider l'hypothèse.

Tâche 3

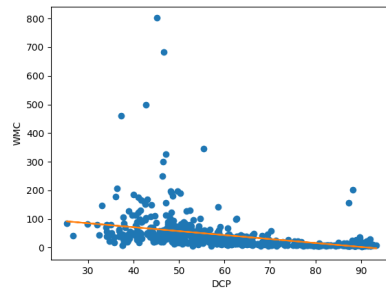
Matrice de corrélation entre différentes mesures, pour le projet jfreechart



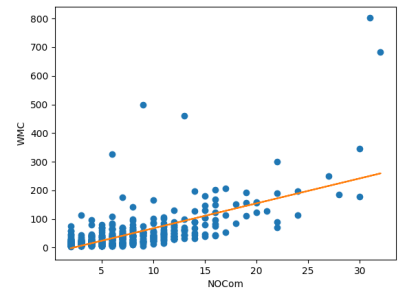
Régression linéaire des données : WMC et NCLOC



Régression linéaire des données : WMC et DCP



Régression linéaire des données : WMC et NOCom



En observant les graphes ci-haut, on remarque qu'il n'y a pas de corrélation significative entre les mesures WMC et DCP, le nombre de commentaires peu être élevé même lorsque que la complexité est relativement basse. Il y a une certaine corrélation entre les mesures WMC et NOCom, bien que quelques points se retrouvent complètement à l'écart. Finalement, la corrélation la plus forte se trouve entre les mesures WMC et NCLOC, on remarque que ces deux mesures croissent au même rythme.

Tâche 4

Pour les trois métriques (NCLOC, DCP, NOCom) on a trois études à faire, avec ces éléments en commun :

Choix d'étude: Étude de cas

Hypothèse: La complexité Weighted Methods per Class des classes du projet jfreechart est une fonction linéaire *d'une des trois métrique*.

Variables:

- **Indépendante:** *L'une des trois métriques*
- **Dépendante:** La complexité Weighted Methods per Class d'une classe

Objets: classes du projet jfreechart

Sujets: calculateurs

Procédure: Prendre le dépôt de code jfreechart, prendre les données que nous avons sur *l'une des trois métriques* et la complexité Weighted Methods per Class des classes. Évaluer la force de relation entre les deux types de données, on considère un score absolu de Pearson supérieur à 0.6 comme étant significatif. Calculer la fonction linéaire qui représente cette relation, par régression linéaire.

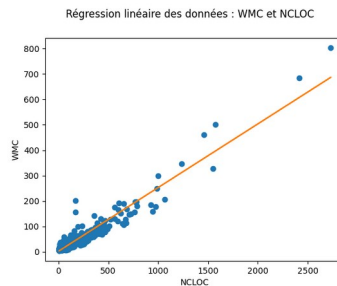
Menaces à la validité:

- **Validité interne**
 - Notre valeur significative (0.6) pour la force d'une relation est peut-être inexacte ou inutile dans le contexte de notre étude.
- **Validité externe**
 - Notre étude ne peut être généralisée pour indiquer la causalité d'une métrique à une autre, étant donné qu'on utilise une mesure de corrélation et pas une de causalité

WMC et NCLOC

Interprétation et généralisation des résultats: On a un score de corrélation Pearson de 0.960 entre les deux métriques ce qui est suffisant pour confirmer la présence d'une relation, notre hypothèse. On calcule ensuite, par régression linéaire, la fonction linéaire de WMC en fonction de NCLOC :

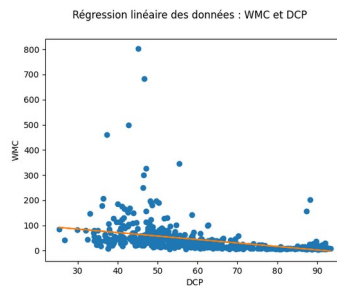
$$\text{WMC} = \text{NCLOC} * 0.25 + 2.49$$



WMC et DCP

Interprétation et généralisation des résultats: On a un score de corrélation Pearson de -0.394 entre les deux métriques ce qui indique une faible (ou pas de) relation entre celles-ci. On infirme donc l'hypothèse qu'il y a une relation linéaire entre es deux. On peut ensuite calculer, par régression linéaire, la fonction linéaire de WMC en fonction de DCP :

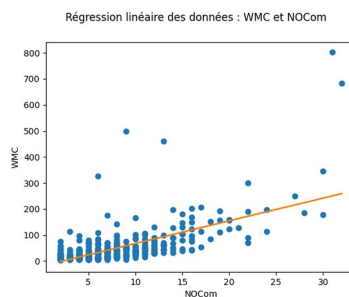
$$\text{WMC} = \text{DCP} * -1.38 + 126.48$$



WMC et NOCom

Interprétation et généralisation des résultats: On a un score de corrélation Pearson de 0.663 entre les deux métriques ce qui est suffisant pour confirmer la présence d'une relation, notre hypothèse. On calcule ensuite, par régression linéaire, la fonction linéaire de WMC en fonction de NCLOC :

$$\text{WMC} = \text{NOCom} * 8.72 - 20.01$$



Tâche 5

Avant de pouvoir commencer, nous avons comptabilisé et analysé quelques statistiques clés sur nos quatre métriques (NCLOC, DCP, NOCom et WMC).

La première hypothèse évaluée a été: «Les classes du projet jfreechart qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois». Étant donnée que, la différence entre la moyenne géométrique des DCP pour les classes avec un NOCom supérieur ou égal à 10 et celle pour les classes avec un NOCom inférieur à 10 n'était que de 16.84%, nous avons infirmé l'hypothèse.

Les trois hypothèses suivantes étaient que la métrique WMC est fonction linéaire de chacune des autres métriques. Nous avons établi un critère qu'un score Pearson absolu de 0.6 ou plus était significatif. Seulement NCLOC et NOCom ont passés ce critère, confirmant ainsi leurs hypothèses respectives. Pour ce qui est de DCP, le score absolu de 0.394 n'est pas assez significatif, on dit donc que WMC n'est pas fonction linéaire de DCP.