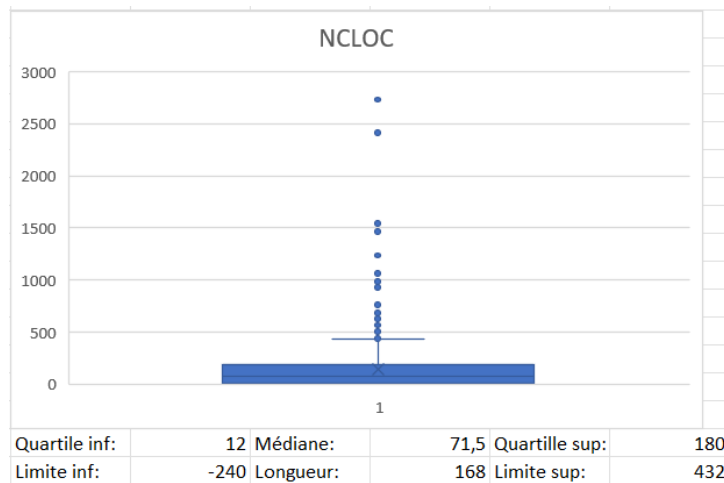


Guillaume Gagnon (20191696), Pierre-Paul Hamon (20160518)

Date : 18 Mars 2022

TP2 IFT3913

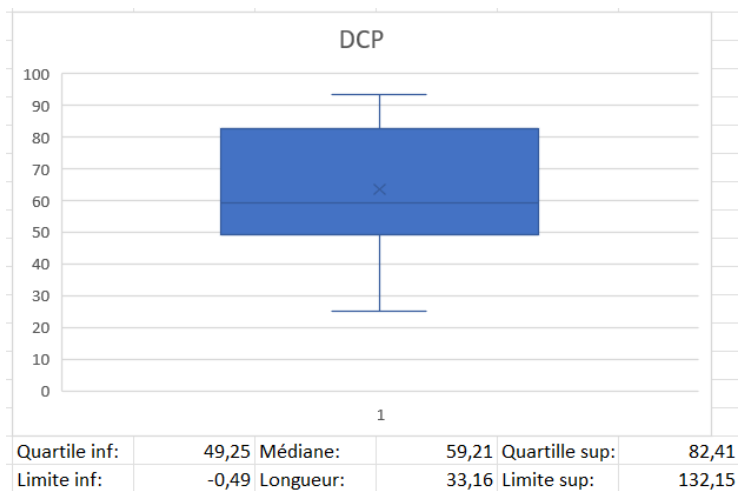
1. Les informations pertinentes dans cette partie ont été calculé à l'aide du fichier Excel tp2#1



NCLOC :

$$71,5 - 12 = 59,5$$

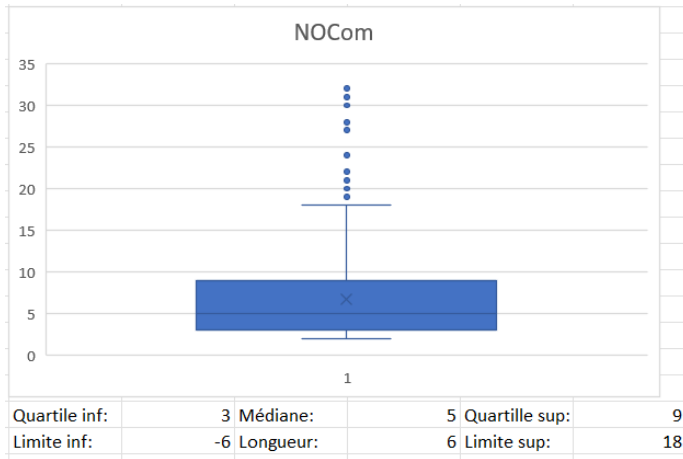
$$59,5/168 = 0,35 = 35\%$$



DCP :

$$59,21 - 49,25 = 9,96$$

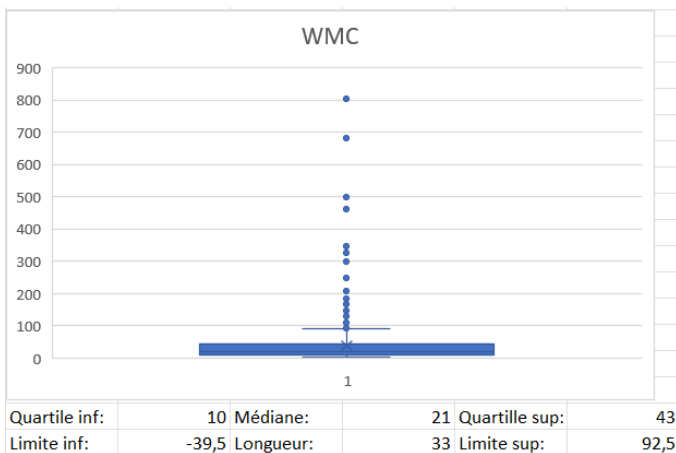
$$9,96/33,16 = 0,30 = 30\%$$



NOCom :

$$5 - 3 = 2$$

$$2/6 = 0,33 = 33\%$$



WMC :

$$21 - 10 = 11$$

$$11/33 = 0,33 = 33\%$$

De ce qu'on peut voir, aucune des métriques ci-dessus est distribuées normalement, puisque, pour toutes les métriques, les quartiles inférieures et supérieures représentent environ 33% et 66% de la longueur de la boîte, respectivement.

Pour DCP, la limite inférieure semble aller plus loin que la limite supérieure, contrairement aux autres métriques. Ceci est probablement dû au fait que la valeur peut seulement être entre 0 et 100, et que la médiane supérieure se trouve proche de 85%, laissant donc moins de place à la limite. Aussi, DCP est la seule métrique à ne pas avoir de points extrêmes.

2. Afin d'évaluer l'hypothèse, on a besoin de comparer deux groupes d'une population. Ayant accès à un échantillon, il fait du sens de choisir l'étude d'ensemble pour cette étude.

L'hypothèse peut donc être énoncé tel quel : « Les classes dont le NOCom est plus élevé que 10 ont un meilleur DCP que les classes dont le NOCom est inférieure à 10 »

Comme variable indépendantes, on a : le nombre de classes dont le NOCom > 10 et le nombre de classes dont le NOCom < 10

Pour les variables dépendantes, on a : la moyenne du DCP des classes dont le NOCom > 10 et la moyenne du DCP des classes dont le NOCom < 10

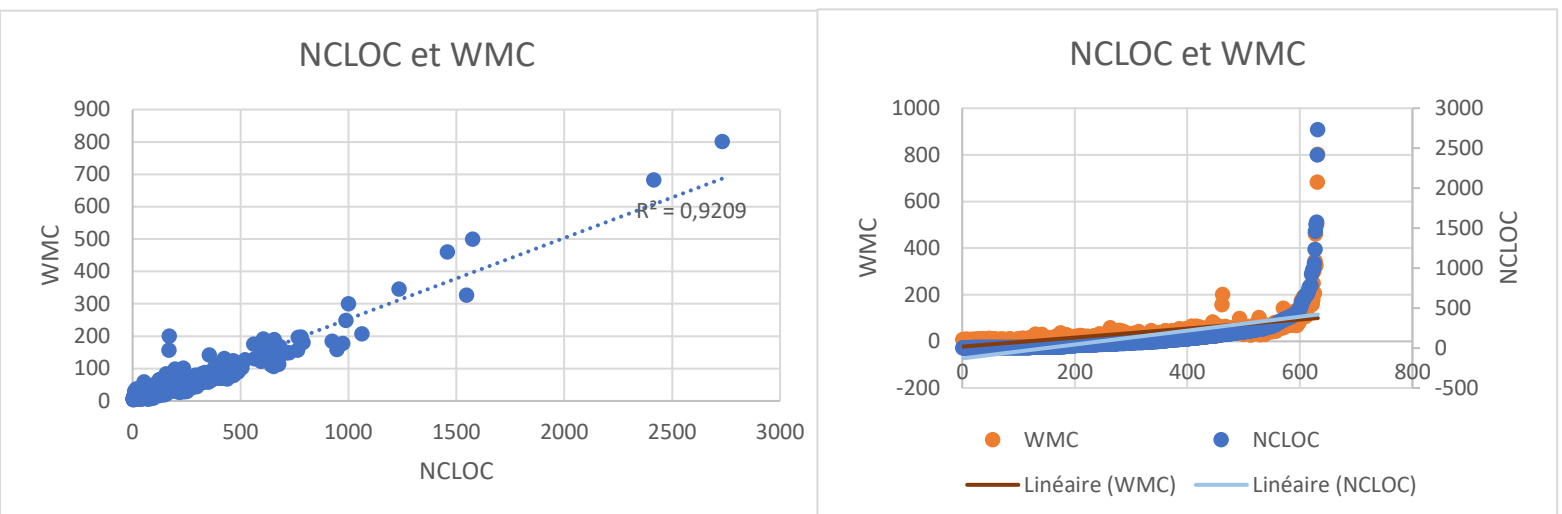
Les résultats de cette étude pourraient confirmer l'hypothèse ci-dessus pour la population des classes jfreechart en général. Ils pourraient aussi indiquer une possible corrélation entre les métriques DCP et NOCom.

Il se pourrait que la validité de cette étude soit atteinte par un problème de variables confondantes. Il se peut aussi que la taille d'un des deux groupes évalués soit trop petite, ce qui pourrait causer des problèmes de régression vers la moyenne.

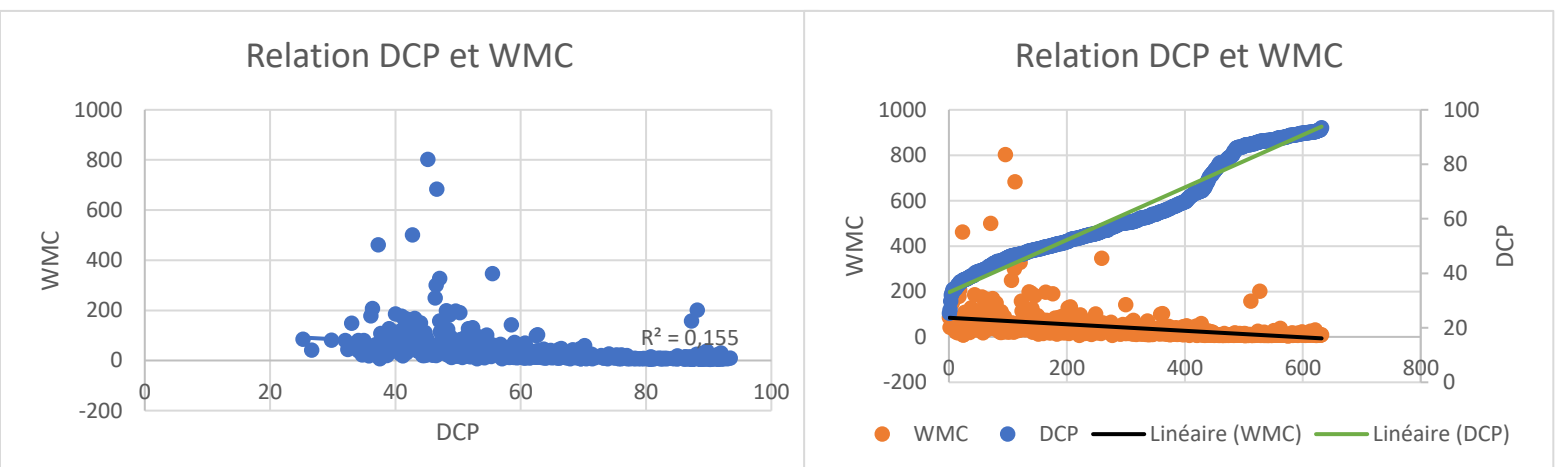
Le sondage a été fait à l'aide d'un programme appelé EtudeMetrique, qui a écrit les résultats dans un fichier appelé tp2#2.txt

Selon tp2#2, la moyenne de la DCP des classes ayant un NOCom inférieure à 10 est 67.51805, alors que la moyenne de la DCP des classes ayant un NOCom supérieure à 10 est 48.18528, ce qui va directement à l'encontre de l'hypothèse posé. Il se pourrait que la plupart des commentaires soit créé plus tôt durant le développement du code, au lieu d'être fait au fur et à mesure, ce qui donnerait donc des résultats tel que celui-ci.

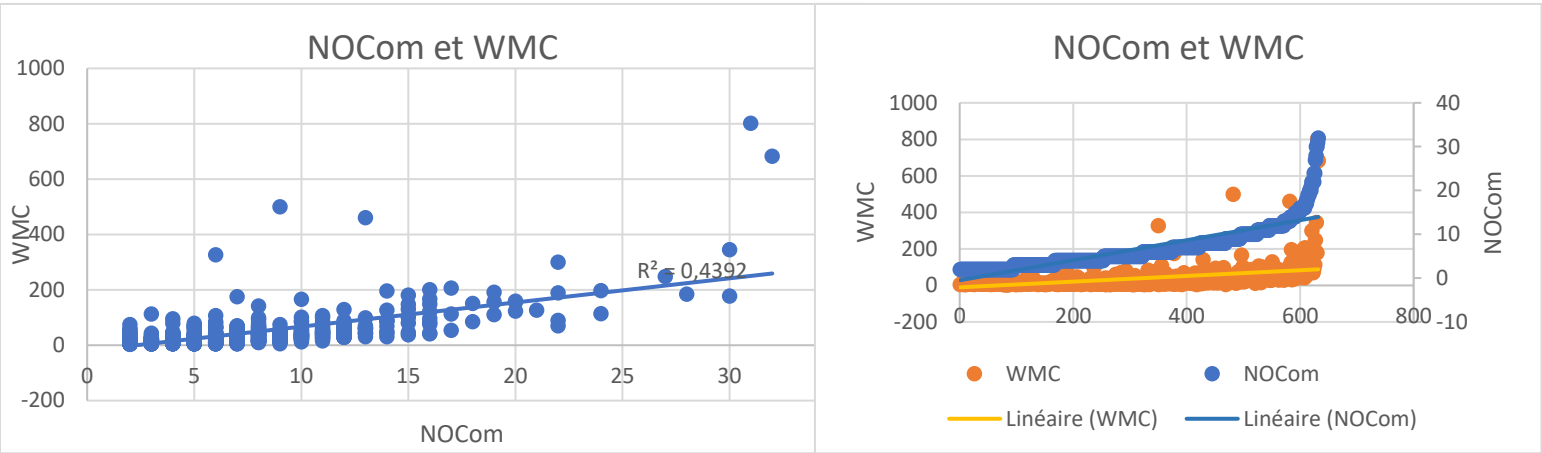
3.



NCLOC et WMC : À partir de ce graphique, on peut observer une certaine corrélation entre les données. Autrement dit, ce graphique semble significatif : il montre que la métrique de complexité « Weighed Methods per Class » semble être en corrélation direct avec le nombre de lignes de code. Autrement dit, plus il y a de ligne de code, plus la classe sera complexe! Le deuxième graphique nous montre la même chose (données triées du plus petit au plus grand selon NCLOC).



DCP et WMC : Ici, c'est un peu moins clair, à première vue. On remarque quand même une tendance : plus la densité de commentaire est haute, moins la classe semble complexe. Cependant, les données ne sont pas très significatives. Autrement dit, la corrélation étant assez faible, il y a de bonnes chances qu'il n'y pas de lien significatif fort entre les deux mesures. On remarque sensiblement la même chose dans le deuxième graphique (données triées selon DCP, du plus petit au plus grand).



NOCom et WMC : Ici, à l'aide des deux graphiques, on remarque une corrélation moyennement forte suggérant que plus il y a de « commit » dans l'historique Git de la classe, plus celle-ci aura tendance à être complexe. Sans entrer dans les détails, il est naturel de penser que cette corrélation puisse être vraie. En effet, normalement, quand on « commit » c'est car on a modifié l'original, soit en optimisant ou en corrigeant des problèmes, soit en ajoutant des fonctionnalités. Cependant, on ne peut rien conclure directement à partir de ces graphiques et de la corrélation observée.

4.

Afin d'évaluer l'hypothèse, on a besoin d'évaluer la corrélation des échantillons de deux populations. Ainsi, on va se servir d'une étude d'ensemble afin d'obtenir les coefficients de Spearman dont on a besoin.

L'hypothèse peut donc être énoncée tel que : « Les coefficients de Spearman entre WMC et NCLOC, WMC et DCP ainsi que WMC et NOCom sont significatifs »

Comme variable indépendante, on a : la taille de l'échantillon

Pour les variables dépendantes, on a : les coefficients de Spearman entre WMC et NCLOC, WMC et DCP ainsi que WMC et NOCom

Les résultats de cette étude pourraient indiquer une corrélation directe entre les métriques ci-dessus pour la population des classes jfreechart en général.

Il se pourrait que la validité de cette étude soit atteinte par un problème de variables confondantes. Il se peut que les résultats aient été directement impactés par les outils utilisés ou un manque de fiabilité des mesures.

Cette étude a été faite à l'aide d'une feuille de calcul Excel, appelé T4.

Selon tp2#2, la moyenne de la DCP des classes ayant un NOCom inférieure à 10 est 67.51805, alors que la moyenne de la DCP des classes ayant un NOCom supérieure à 10 est 48.18528, ce qui va directement à l'encontre de l'hypothèse posée. Il se pourrait que la plupart des commentaires soit créés plus tôt durant le développement du code, au lieu d'être fait au fur et à mesure, ce qui donnerait donc des résultats tels que celui-ci.

#a (WMC est une fonction linéaire du NCLOC) :

On procède au calcul du coefficient de corrélation de Spearman (r) et on obtient $r = 0.9192822 \approx 0.92$

Donc, on remarque une très forte corrélation entre NCLOC et WMC et on peut donc conclure que WMC est une fonction linéaire du NCLOC. Autrement dit, plus il y a un nombre élevé de lignes de code, plus la classe sera complexe.

#b (WMC est une fonction linéaire du DCP) :

On procède au calcul du coefficient de corrélation de Spearman (r) et on obtient $r = -0.77077053 \approx -0.77$

Donc, on remarque une forte corrélation **négative** entre DCP et WMC. Plus la densité de commentaire est élevée, moins la complexité de classe le sera.

#c (WMC est une fonction linéaire du NOCom) :

On procède au calcul du coefficient de corrélation de Spearan (r) et on obtient $r = 0.67367731 \approx 0.67$

Donc, on remarque une moyenne-forte corrélation entre NOCom et WMC.

Plus le nombre de commit est élevé, plus la complexité de la classe le sera aussi.

5. Ainsi, durant cette étude, nous avons démontrer que les métriques NCLOC, DCP, NOCom ont une distribution non-normale. Aussi, on a vu que les classes dont le NOCom > 10 sont généralement moins bien commentés que celles dont le NOCom < 10. Enfin, on a démontrer des corrélations entre WMC et NCLOC, WMC et DCP ainsi que WMC et NOCom.