

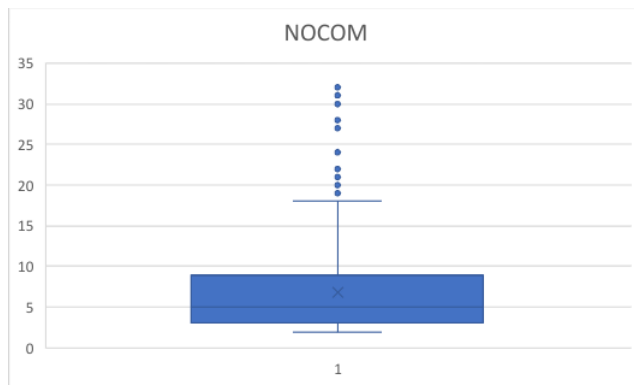
# IFT3913 TP3

Louis-André Brassard : 20182271

Xavier Lapalme : 20187052

18 November 2022

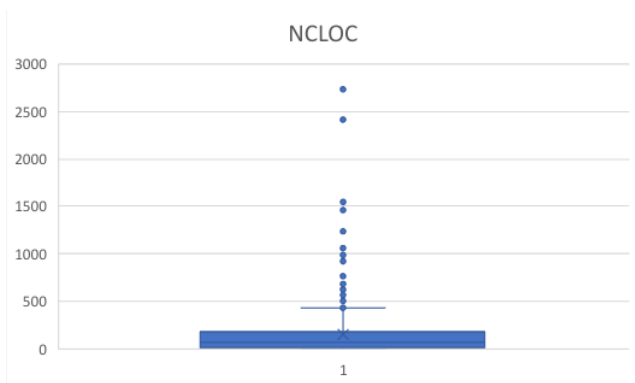
## Tâche 1



Informations	NoCom
Médiane (m)	5
Quartile supérieur (u)	9
Quartile inférieur (l)	3
Longueur (d)	6
Limite supérieur (s)	18
Limite inférieur (i)	2
Valeurs extrêmes	22, 24, 32, 19, 22, 27, 22, 24, 20, 31, 30, 28, 22, 30, 20, 19, 21, 20, 19

On peut voir que NOCOM est généralement bien centré autour de la médiane avec seulement quelques points extrêmes (19), tous supérieurs à la limite maximum dans ce cas-ci. Nous pouvons également voir que la plupart des classes, soit 75% n'ont pas été changées plus de 9 fois comme trouvé par notre quartile supérieur (u). Ces classes peuvent représentées des fonctions secondaires qui n'ont pas besoin d'être modifier souvent et qui une fois développées et fonctionnelles ne posent pas problèmes. Pour ce qui est du reste, le quatrième quartile, il peut s'agir des classes principales de JFreeChart ou encore des interfaces ou des classes complexes qui nécessite beaucoup de travail ou qui ont connus plusieurs bug.

Pour ce qui est de la distribution de NoCom, on peut voir d'après le graph de distribution et la valeur de 344.51 au test de Jarque-Bara que notre distribution n'est pas normal.

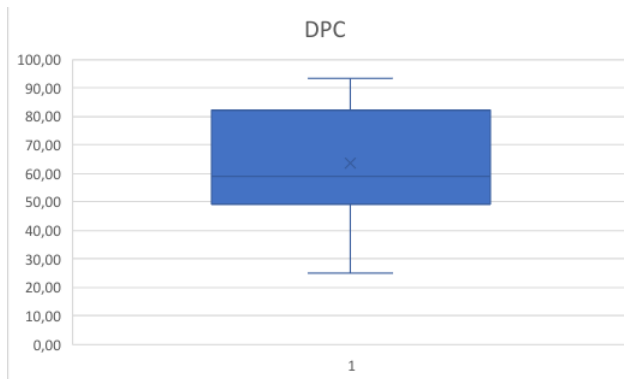


Informations	NCLOC
Médiane (m)	71.5
Quartile supérieur (u)	180
Quartile inférieur (l)	12
Longueur (d)	168
Limite supérieur (s)	432
Limite inférieur (i)	4
Valeurs extrêmes	789, 726, 653, 1063, 561, 565, 597, 670, 767, 1000, 780, 468, 2415, 438, 507, 455, 643, 1547, 656, 988, 677, 764, 2732, 1235, 925, 683, 596, 489, 974, 466, 623, 434, 708, 521, 947, 1575, 469, 1459, 606

La distribution de NCLOC peut sembler plus disparate que celle de NOCOM, cependant il faut apporter une nuance aux résultats obtenus par cette métrique. Comme elle mesure seulement le nombre de ligne de code utile à la compilation ou à l'interprétation de l'ordinateur, on peut dire qu'elle mesure seulement le code "utile". Avec une médiane de 71.5 lignes de codes, on voit que la plupart des classes comportent que très peu de lignes de code. Ceci peut être interpréter par un programme ayant une forte cohésion et un faible couplage. En d'autres mots, ceci indique que les classes, pour la plupart ne font qu'une seule tâche aussi futiles soit-elle. D'autre part, les classes qui ont un score NCLOC plus élevé semblent être des classes ayant une des tâches principales du

programme.

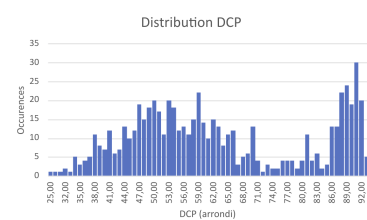
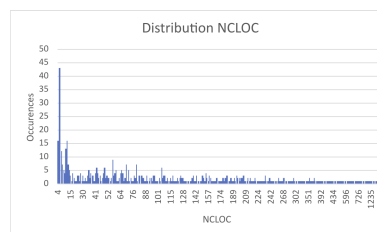
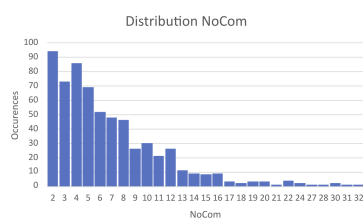
Pour ce qui est de la distribution de NCLOC, on voit une grande concentration des données autour de la médiane et une longue moustache supérieure ce qui est expliqué par les quelques points extrêmes supérieurs très élevés. Contrairement à ce que la boîte à moustache de NCLOC peut laisser penser, nous n'avons pas de classes qui sont 0 lignes de codes. Il s'agit ici d'un effet de perspectives causé encore une fois par les valeurs extrêmes. On peut aussi voir que NCLOC ne suit pas une distribution normale en se basant sur le graph de distribution ainsi qu'à son score au test de Jarque-Bara de 266.21. NCLOC présente elle aussi une distribution biaisée vers la droite



Informations	DCP
Médiane (m)	59.21
Quartile supérieur (u)	82.36
Quartile inférieur (l)	49.26
Longueur (d)	33.1
Limite supérieur (s)	132.01
Limite inférieur (i)	25.24
Valeur extrême	∅

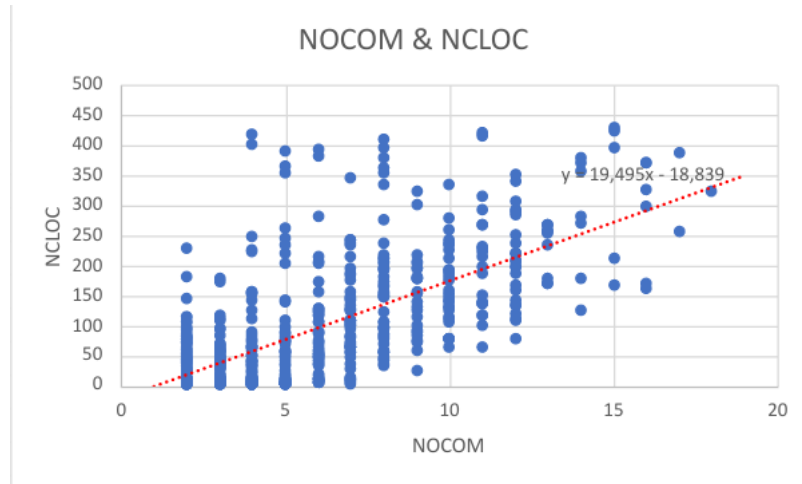
En analysant la distribution de DCP, nous pouvons voir qu'elle semble séparée en deux parties ce qui est caractéristique d'une distribution bimodale. En effet on peut observer un mode autour des  $DCP = 50$  et un autre mode autour de  $DCP = 90$ . Ceci nous démontre qu'il ne s'agit pas d'une distribution normale. Cette observation est renforcée par le résultat du test de Jarque-Bara de 3607.41 qui est très loin de 0 et donc ce qui confirme que DCP n'est pas une distribution normale.

## Distribution



Test Jarque-Bara			
	NoCom	NCLOC	DCP
Skewness	1.88	4.48	5.11
Kurtosis	8.04	30.43	42.05
Jarque-Bara	344.51	2666.21	3607.41

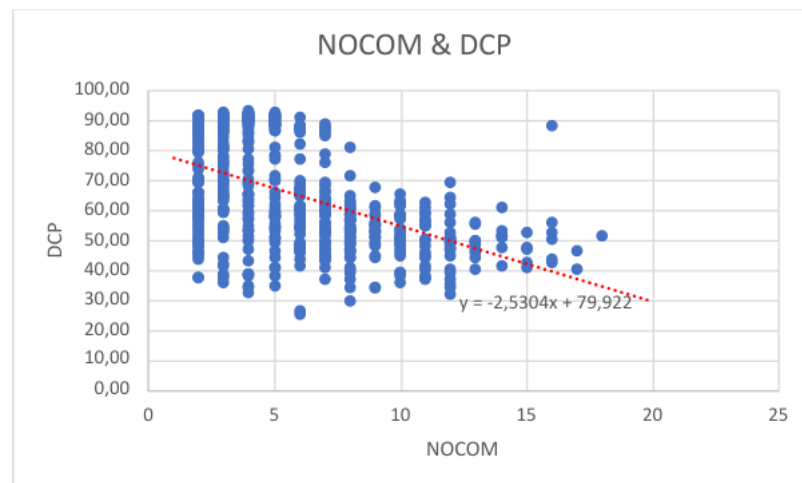
## Tâche 2



Les deux distributions NoCom et NCLOC ne sont pas normalement distribuées donc Pearson ne fait pas de sens comme coefficient de corrélation ici. Et donc nous choisis d'utiliser Spearman et avons obtenu un score de  $\rho = 0.6431$  pour cette relation avec une régression linéaire de  $y = 19,495x - 18,839$ . Pour arriver à ces résultats, nous avons enlevé tous les points extrêmes soit ceux supérieurs à  $s = 18$  et inférieurs à  $i = 2$  pour NoCom et ceux supérieurs à  $s = 432$  et inférieurs à  $i = 4$  pour NCLOC.

Nous avons utilisé cette formule afin d'obtenir le coefficient de Spearman :

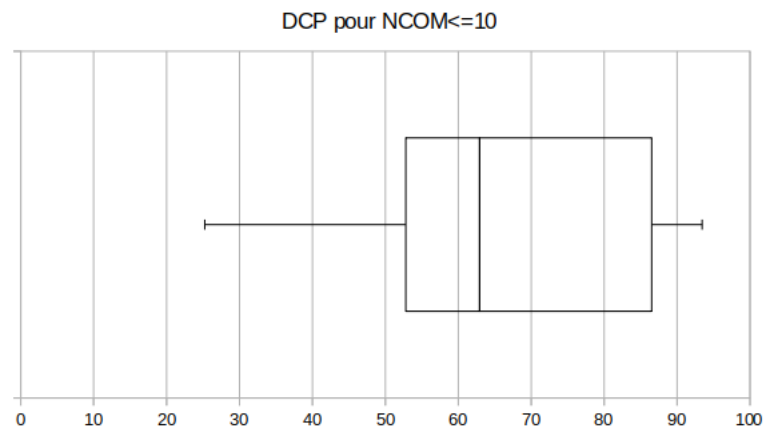
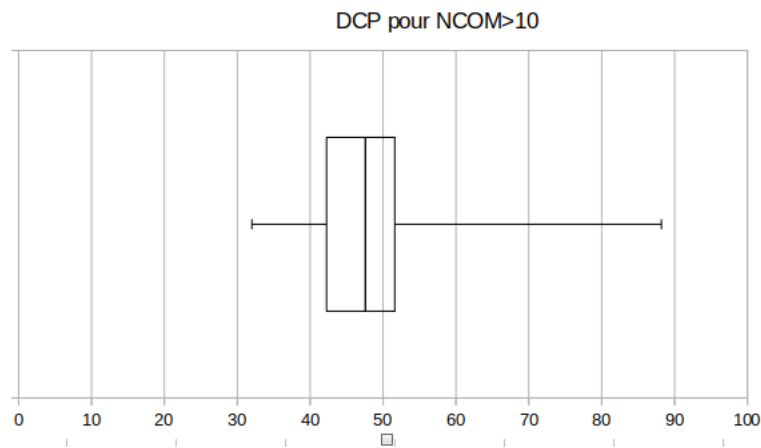
$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$$



Pour la corrélation entre NOCOM et DCP, nous avons encore une fois deux distributions non normales, et donc un coefficient de corrélation de Spearman s'impose. Pour cette relation, nous avons obtenu un coefficient de  $\rho = -0,4846$  et une régression linéaire de  $y = -2.5304x + 79.922$ . Nous avons encore une fois retiré les valeurs extrêmes de NOCOM et DCP soit : ceux supérieurs à  $s = 18$  et inférieurs à  $i = 2$  pour NoCom et ceux supérieurs à  $s = 132.01$  et inférieurs à  $i = 25.24$  pour DCP. Aucun point extrême n'a été retiré pour DCP.

## Tâche 3

Nous avons une hypothèse qui dit que les classes qui ont été modifiées plus de 10 fois sont mieux commentées que les classes pour lesquelles ce n'est pas le cas. Ici nous avons plusieurs classes qui ne s'équivalent pas nécessairement et nous n'avons le contrôle sur aucun de leurs paramètres, nombre de modifications ou densité de commentaire entre autres. Donc, nous devons faire une quasi-expérience. Nos variables sont: le nombre de fois où la classe a été modifiée et la densité de commentaire en pourcentage. Nous avons deux catégories de classes, les classes qui ont été modifiées plus de 10 fois et celles qui ont été modifiées 10 fois ou moins. Voici les deux boîtes à moustaches:



Le premier représente la distribution de la densité de commentaire pour les classes modifiées plus de 10 fois. La deuxième pour celles modifiées 10 fois ou moins. Nous pouvons facilement voir que dans la première boîte à moustache, la 75% est sous 55 alors que, pour la deuxième, c'est au dessus de 50. De plus nous pouvons regarder le graphique de la relation entre NCOM et DCP "NOCOM & DCP" qui démontre bien que, plus la classes à été modifiées souvent, moins la densité de commentaire est élevée. La conclusion est donc que les classes modifiées plus de 10 fois sont moins bien commentées que les autres, donc notre hypothèse était fausse. C'est certain qu'il faut être vigilant de plusieurs choses en voici quelques exemples. Peut-être que c'est la façon dont travail cette équipe spécifiquement et que pour une autre équipe ça serait différent. Peut-être aussi que pour ce projet et ce dont il traite ce résultat est normal et que pour un autre genre de projet ça serait différent.