

1- Boîtes à moustaches de NOCom, NCLOC et DCP:

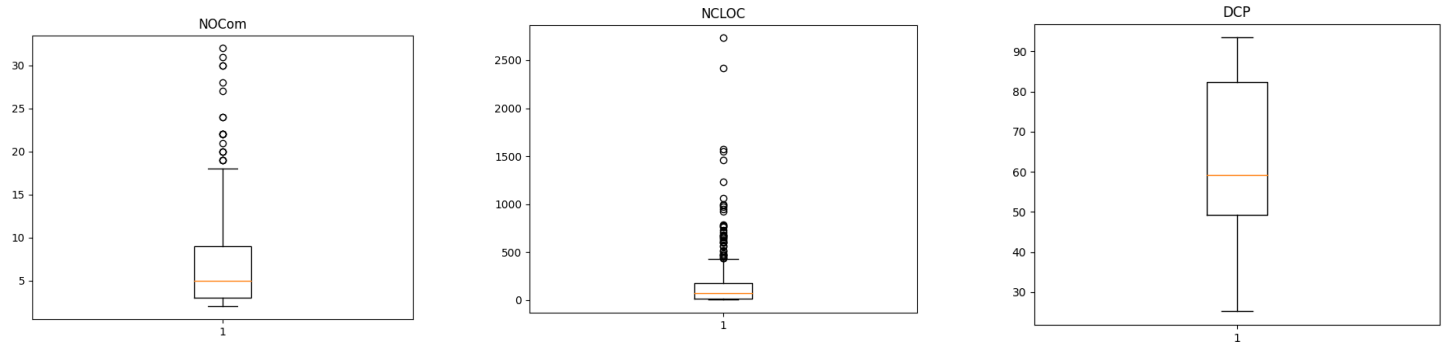
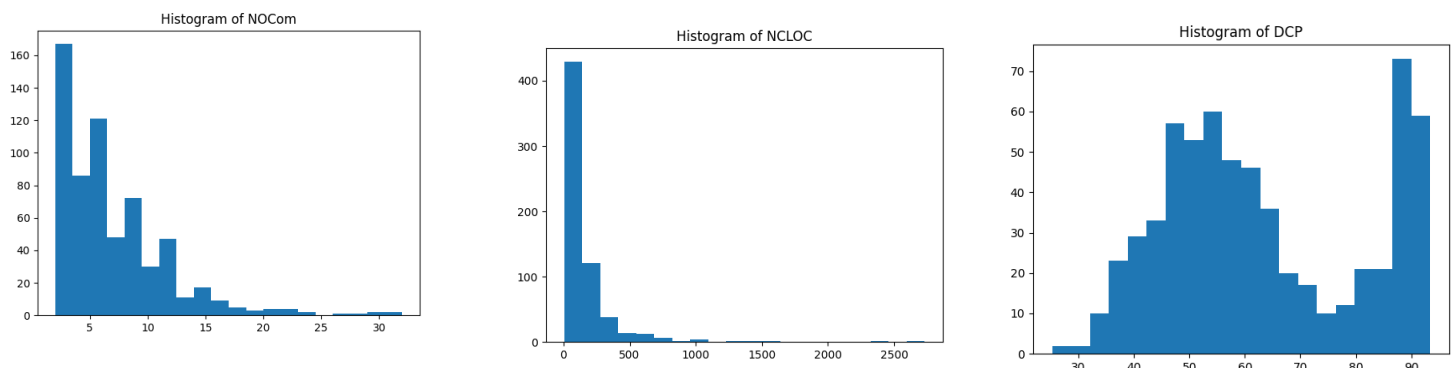


Tableau 1: Données récapitulatives des boîtes à moustaches

	NOCom	NCLOC	DCP
Quartile 1	3	12	49,25
Médiane (Q2)	5	71,5	59,21
Quartile 3	9	180	82,41
Limite supérieure	18	432	132,14999
Limite inférieure	2	4	25.24
Nombre d'éléments dans l'air de donne supérieurs aux limites Min et Max	19	39	0

Toutes les boîtes à moustaches ont une caractéristique commune qui est la non-symétrie de la médiane, celle-ci ne se situe pas au milieu de la boîte. Visuellement, les données sont situées plus vers le côté supérieur et donc une asymétrie. Il est remarquable aussi que les deux boîtes (NOCom et NCLOC) contiennent des données aberrantes, soit des données qui sont en dehors des limites supérieure ou inférieure, dans ce cas ça dépasse la limite supérieure pour NOCom avec 19 éléments et de 39 pour NCLOC. Pour ce qui est de DCP, aucune donnée aberrante n'existe et la médiane gravite vers 59%. La distorsion des boîtes nous indique que les données ne suivent pas une loi normale vue que la médiane ne se retrouve pas au milieu. Pour des amples explications sur le processus de production des boîtes veuillez regarder tache2

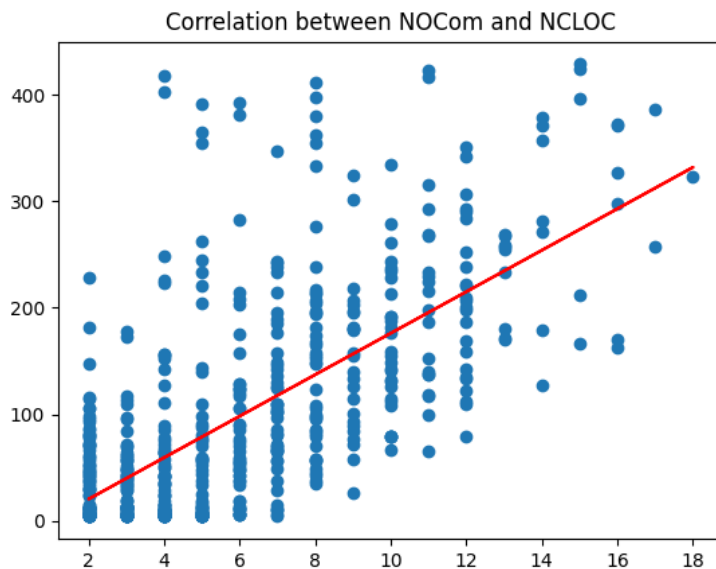
2- Histogrammes de NOCom, NCLOC et DCP:



Pour ce qui est des Histogrammes correspondants a NOCom et NCLOC, ils ont une distribution asymétrique beaucoup plus oriente vers la droite (dépendamment de la direction de la queue), ce qui concorde à l'analyse des boîtes a moustaches ci haut. Pour ce qui est de l'histogramme de DCP, on constate qu'on a plusieurs piques. Et donc, aucun des histogrammes ne contient une répartition normale vue l'absence de la forme de cloche.

3- Graphiques de corrélation :

a. Graphique de corrélation entre NOCom et NCLOC :



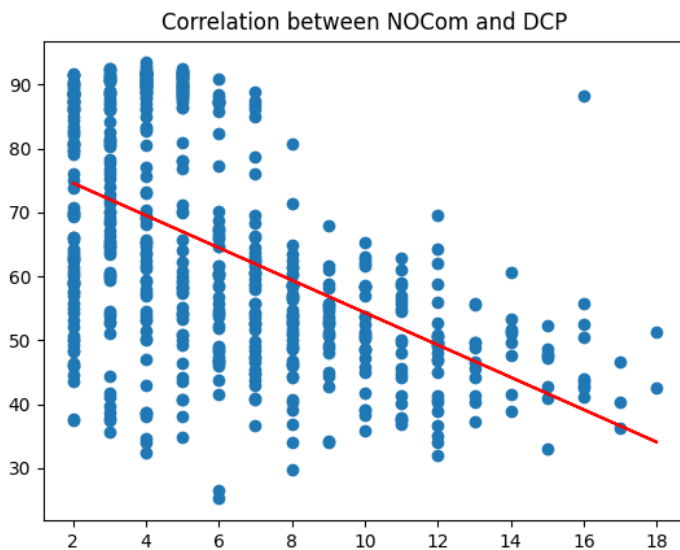
Coefficient de corrélation entre NOCom et NCLOC : 0.6674248320044255

Coefficient de corrélation de spearman entre NOCom et NCLOC : 0.6407751294809536

Droite de régression :

$$y = 19.49452861952861 x + -18.838576116494895$$

b. Graphique de corrélation entre NOCom et DCP :



Coefficient de corrélation entre NOCom et DCP: -0.5212370885327228

Coefficient de corrélation de spearman entre NOCom et DCP: -0.5177069397190578

Droite de régression :

$$y = -2.532879291367214 x + 79.6751667550044$$

4- Discussion :

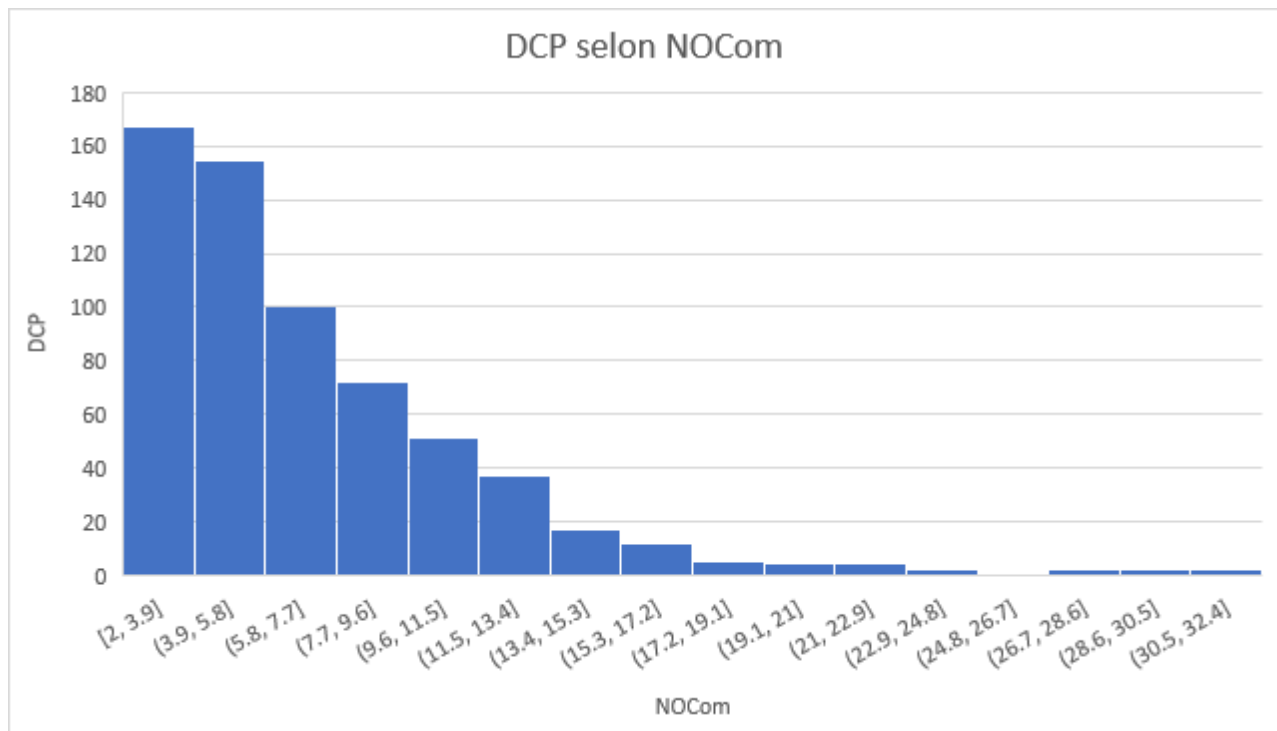
a. Étude des corrélations entre NOCom avec NCLOC et DCP :

Pour les résultats obtenus dans les deux graphiques de corrélations, il est difficile de visualiser s'ils suivent une distribution normale uniquement en se référant aux graphiques. Pour cela, deux types de coefficient de corrélations ont été utilisés, Pearson et Spearman. Il est connu que si le coefficient de Pearson est plus proche de 1 ou -1, alors la corrélation est forte tandis que plus on se rapproche de 0 le contraire est vrai. Dans ce cas, le coefficient de Pearson se rapproche davantage de 1 comparé au coefficient de Spearman alors les données suivent plus ou moins une distribution normale, bien que la corrélation existante soit considérablement faible.

b. Évaluation de l'hypothèse « les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois » :

Pour pouvoir répondre à cette hypothèse, un plan d'une quasi-expérience a été élaboré. Premièrement discutons la raison pour laquelle une quasi-expérience a été sélectionné comme choix de type d'étude. Une quasi-expérience fait partie du type d'étude expérience qui ont certaines caractéristiques distinctes tel que l'affirmation de modèles théoriques ou bien évaluer leurs exactitudes. Concernant le type d'expérience le plus approprié pour l'étude de [jfreechart-stats](#), la quasi-expérience a été sélectionné due aux types de données qu'on traite. Se référant aux notes de cours du chapitre 6, la quasi-expérience a été choisi car on ne peut pas de tester physiquement (de façon concrète) notre hypothèse, vue que toutes les variables liées au code/projet sont sujettes à des changements externes (les personnes qui compose l'équipe, la façon de coder de chaque individu, etc.) L'hypothèse à évaluer est la suivante : « les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois ». Bien évidemment, il est impossible de l'évaluer sans avoir recours à des tests, celui qui a été choisi est un [histogramme](#). Tel-que la définition des variables est la suivante :

- Y (variable indépendante) : représente le nombre de modifications dans une classe (NOCom)
- X (variable dépendante) : représente la densité de commentaire en pourcentage



Selon l'histogramme obtenu, il est marquant que plus le nombre de commit augmente, la densité de commentaire diminue. On peut donc déduire que les classes qui ont été modifié au moins 10 ne sont pas mieux commenté. Toutefois, il est indéniable de reconnaître que les commentaires peuvent ne pas être pertinents ou bien en lien avec les fonctions de la classe et que ce sont des commentaires triviaux. Donc, si on considère que du code mieux commenté correspond à avoir plus de commentaires, alors la réponse à notre hypothèse est non. D'un l'autre côté, si on considère que du code mieux commenté implique une bonne qualité de commentaire alors dans ce cas on ne peut formuler une réponse (oui/non) pour notre hypothèse, la qualité des commentaires et leur pertinence sont des aspects qui ne peuvent pas être mesuré.

Pour conclure, vue le manque d'informations concernant ce que veut dire « mieux commentées » et les critères pour valider les commentaires. On tient compte que des classes mieux commentées sont celles qui ont été modifié au moins 10 fois et qui contiennent moins de commentaire (ou plutôt le rapport code/commentaire). Notre réponse sera donc « non, l'hypothèse est vrai ».