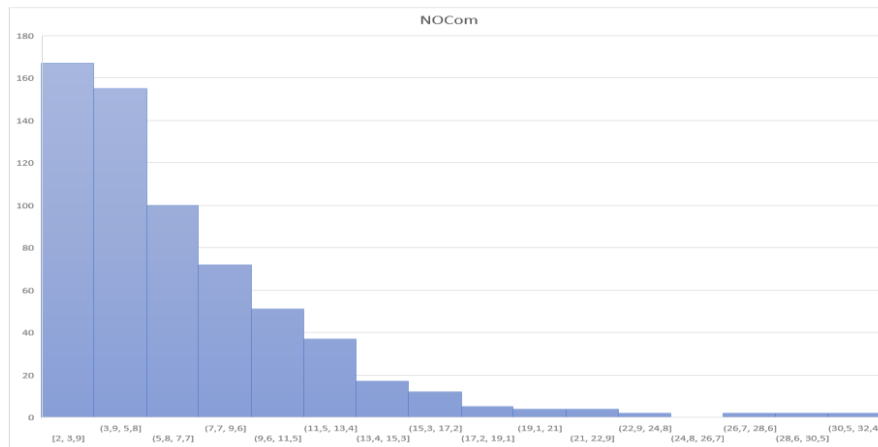
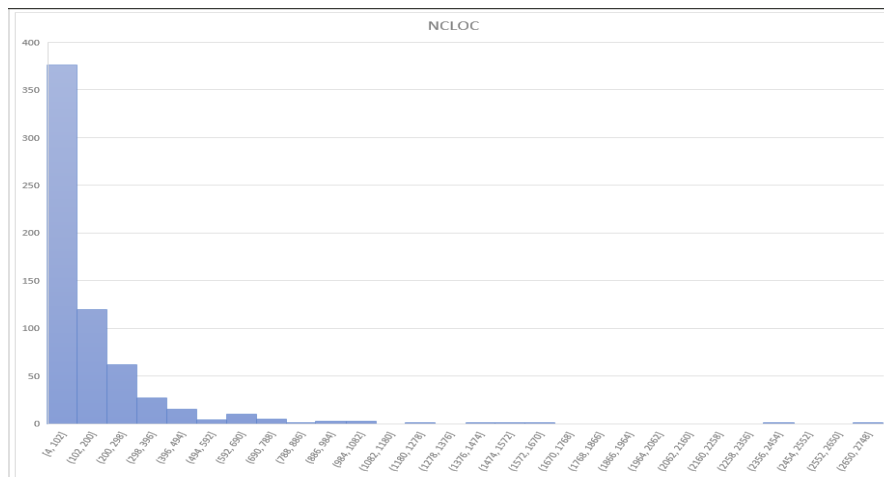


Tache #1:

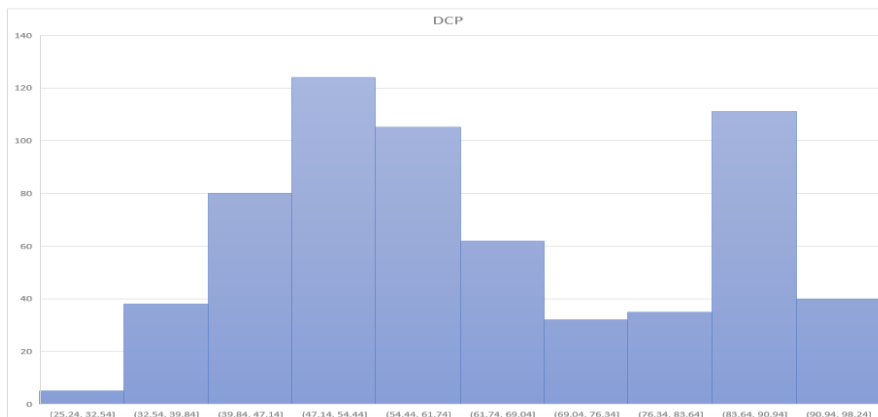
1. Histogramme NoCom

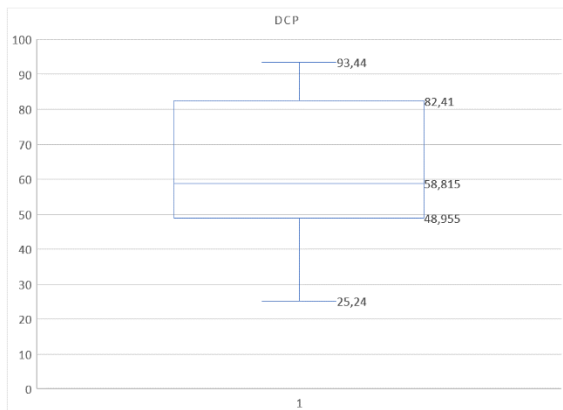
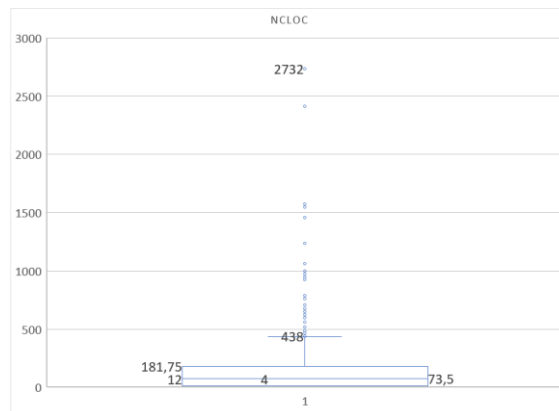
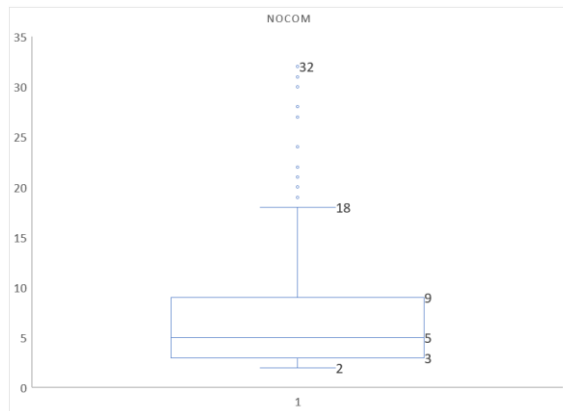


2. Histogramme NCLOC



3. Histogramme DCP





Dans ces boîtes à moustaches les valeurs sont disposées comme cela: La première valeur à partir du bas est la valeur minimum. Ensuite la 2^e, qui est aussi le début de la boîte, est le quartile inférieur. Suit par la médiane. Suit elle-même par le quartile supérieur, qui représente aussi la fin de la boîte. Les valeurs au-dessus de la boîte sont présente en titre indicatives. Et la valeur maximum étant la plus haute de tous.

Tache #2 :

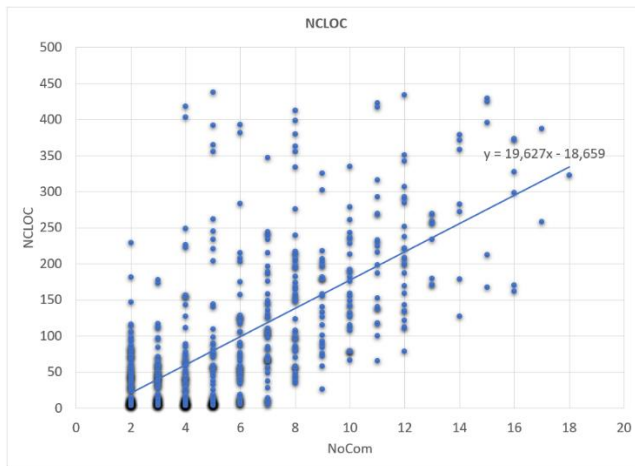
Tout d'abord, il nous fallait savoir les distributions des données, donc nous avons produit trois histogrammes à partir des données fournies. D'après ces histogrammes, voir ci-dessus, il est impossible de conclure que les données sont normalement distribuées. Toutefois, on remarque que NoCom suit une loi normale repliée. NCLOC pour sa part affiche une décroissance exponentielle et DCP ne suit pas sensiblement rien. Pour cette raison, le coefficient de corrélation de Spearson est la technique d'analyse la plus approprié. Nous obtenons les coefficients de Spearson r suivant :

$$R_1 : \text{NoCom et NCLOC} = 0,659$$

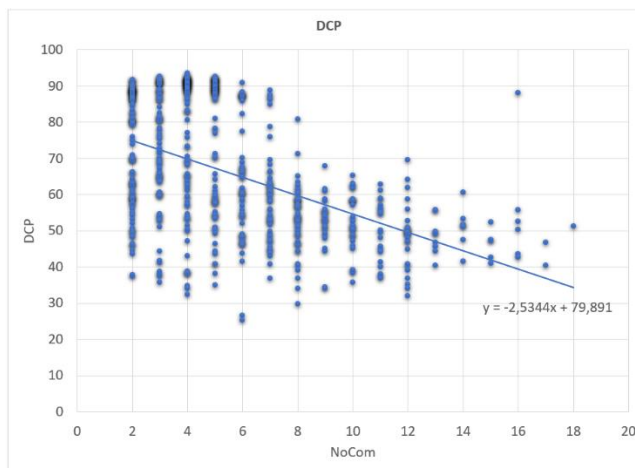
$$R_2 : \text{NoCom et DCP} = -0,514$$

R_1 indique que la quantité de ligne de code va dans le même sens que la quantité de commit. Ce qui était le résultat attendu. R_2 indique une direction opposée, donc plus la quantité de commit augmente, moins le pourcentage de commentaire est élevé. Ce qui correspond aux pentes des graphiques 1 et 2. On peut donc en conclure que ces deux pentes sont valables.

Graphique 1



Graphique 2



Tache #3 :

Pour cette analyse, nous voulons évaluer l'hypothèse suivante : « les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois ». Pour ce faire nous devons déterminer une quasi-expérience.

Le choix d'une quasi-expérience comme étude se justifie très bien par le fait que notre hypothèse prend en considération un échantillon des variables indépendantes et l'on veut une généralisation des résultats, donc l'expérience est le meilleur choix. Ensuite, nos groupes sont définis à partir nos variables, donc on fait une affectation contrôlée et le traitement des cas n'est pas contrôlé. Ce qui résulte avec le choix d'une quasi-expérience.

Hypothèse : les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois.

Variables : Indépendantes : NoCom, NCLOC, DCP

Sujets/Objets : Les classes

Procédure : Nous regardons la corrélation entre NoCom et NCLOC ainsi que la corrélation entre NoCom et DCP pour chaque classe. Ces relations vont nous donner des pentes reliant notre de commits et la documentation. Si la pente de la relation NoCom sur DCP est positive alors le nombre de commentaires augmente avec le nombre de commits et donc notre hypothèse sera vérifiée, sinon on l'hypothèse est fausse et la documentation n'est pas meilleure avec le nombre de modifications.

Menace à la validité : Dans notre cas particulier, peut-être que la manière dont les développeurs écrivent leurs commentaires peut nuire à nos résultats. Par exemple, si lors de la création d'une nouvelle classe, ils écrivent ce qu'il reste à faire dans la classe, la manière dont ils ont prévu de le faire, ou n'importe qu'elle autre information qui sera enlevés au fur et à mesure que la classe est modifiée.

Évaluation de l'hypothèse : En regardant les graphiques de la tâche 2, on voit que la pente du NoCom en fonction de DCP est négative. Cela veut dire qu'en générale, la moyenne de la quantité de documentation d'une classe diminue avec le nombre de modification de celle-ci. De plus, si on regarde le graphique numéro 1, on voit que le nombre de ligne de code par classe augmente avec le nombre de commits, ce qui est logique. Alors si on prend on considération nos données analysées, on peut réfuter la véracité de notre hypothèse car les classes avec plus de 10 commits n'ont pas plus de documentation que les autres.

Notre conclusion est la suivante : Les classes qui ont été modifiées plus de 10 fois ne sont pas mieux commentées que celles qui ont été modifiées moins de 10 fois.