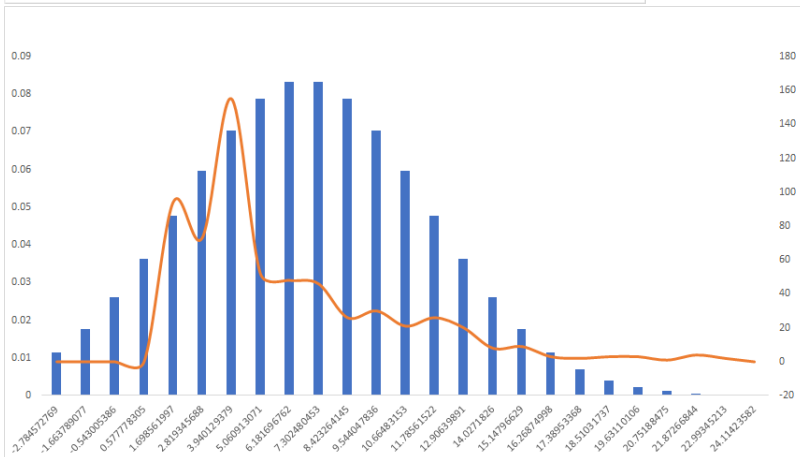
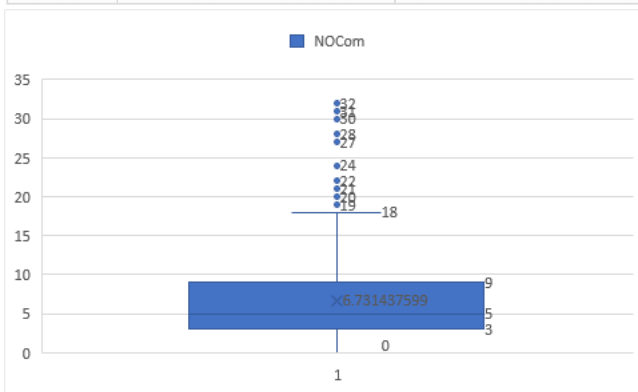


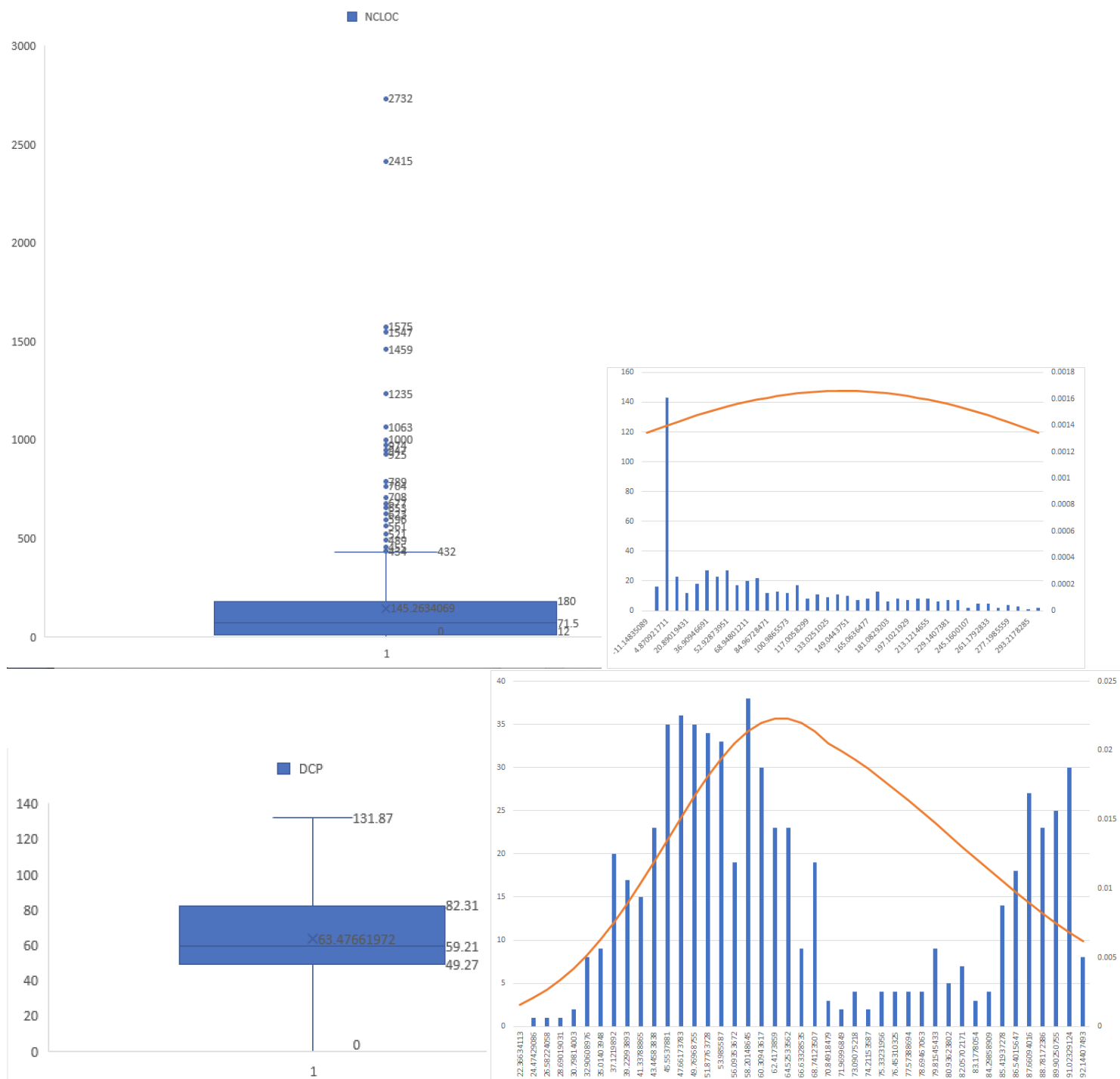
# Rapport TP3

Rosine Rolande Simo Tegninko, 20183729  
Yu Deng, 20151659

## Tâche 1

	NCom	NCLOC	DCP
l	3	12	49.27
m	5	71.5	59.21
u	9	180	82.31
d	6	168	33.04
s	18	432	131.87
i	0	0	0





Le tableau contient les informations nécessaires pour représenter la boîte à moustaches. À partir de la courbe de distribution, la distribution de NoCom n'est pas très proche de la distribution normale, tandis que la courbe normale de NCLOC est la meilleure, mais l'écart type est plus grand, la courbe normale de DCP est très proche de la courbe normale standard, et l'écart type est acceptable.

## Tâche 2

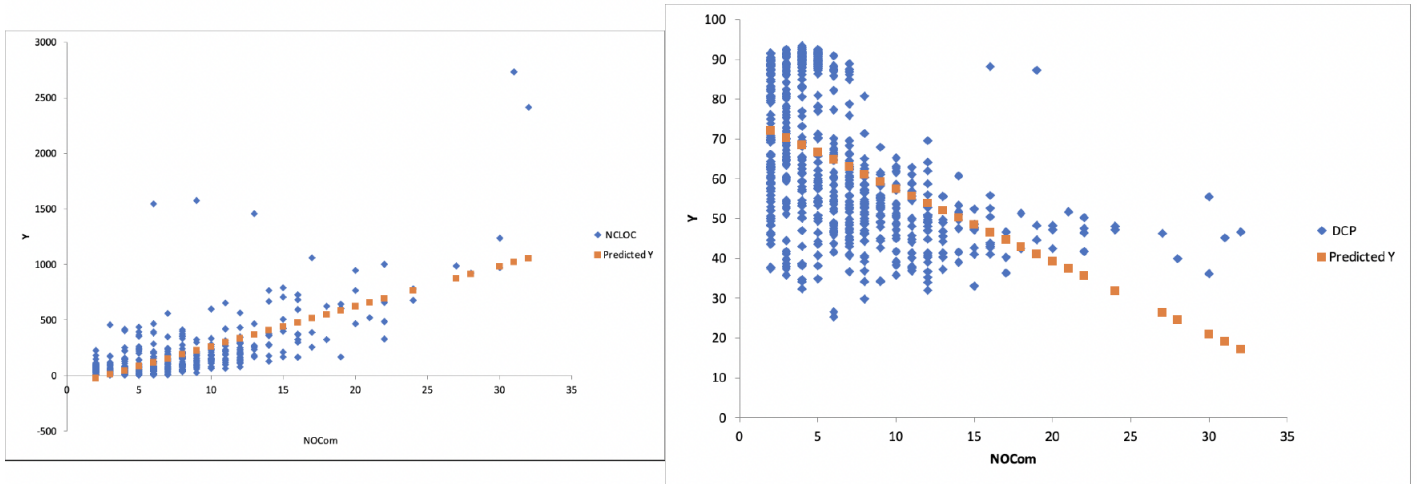
Coefficient de corrélation de Pearson ( $r$ ):

$$r(\text{NoCom}, \text{NCLOC}) : 0.71457166 \quad r(\text{NoCom}, \text{DCP}) : -0.487753$$

Coefficient de corrélation de rang de Spearman ( $\rho$ ):

$$\rho(\text{NoCom}, \text{NCLOC}) : 0.7098611 \quad \rho(\text{NoCom}, \text{DCP}) : -0.4267282$$

Courbe de régression:



Nous pouvons trouver à partir des trois indicateurs différents du coefficient de corrélation de Pearson ( $r$ ), du coefficient de corrélation de rang de Spearman ( $\rho$ ) et de la courbe de régression qu'il existe une forte corrélation positive entre NoCom et NCLOC, et une corrélation relativement forte entre NoCom et DCP. Corrélation négative faible.

### Tâche 3

Evaluer l'hypothèse

#### 1. Choisissez le type d'étude

Quasi-expérience

Nous savons déjà que le nombre de modifications de classe est positivement corrélé avec les lignes de code et négativement corrélé avec la densité des commentaires. Mais nous ne connaissons toujours pas la relation entre le nombre de modifications de classe et de bonnes révisions, nous voulons donc étudier la relation entre le nombre de modifications de classe et de meilleures révisions à travers des quasi-expériences.

#### 2. Définir et énoncer les hypothèses

Nous faisons 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.

#### 3. Définir et étudier les variables

NoCom\_over10 : défini comme une classe avec plus de 10 modifications. ( $\text{NoCom} > 10$ )

NoCom\_down10 : défini comme des classes avec moins de 10 modifications. ( $\text{NoCom} < 10$ )

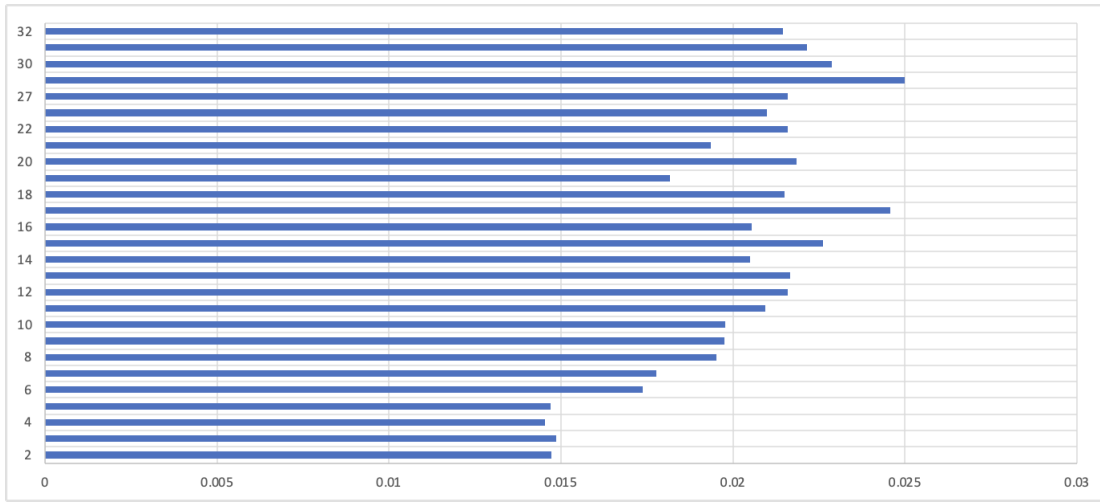
Good\_comments : définis comme de bons commentaires, les bons commentaires doivent être étroitement liés au code, de sorte qu'il n'y aura pas trop de commentaires redondants, et la densité de commentaires des bons commentaires doit être faible. ( $1/\text{DCP}$ )

Avg\_over10 : défini comme la moyenne de good\_comments pour les classes avec plus de 10 modifications. ( $\text{AVERAGE}(\text{NoCom} > 10, \text{Good\_comments})$ )

avg\_down10 : défini comme la moyenne de good\_comments pour les classes avec moins de 10 modifications. ( $\text{AVERAGE}(\text{NoCom} < 10, \text{Good\_comments})$ )

#### 4. Interpréter et généraliser les résultats

	Avg_over10	avg_down10
Good_comments	0.02138612	0.01599458



(L'axe des ordonnées de l'histogramme est le nombre de modifications de classe NoCom, et l'axe des abscisses est le bon commentaire correspondant)

Dans le tableau, nous pouvons voir que le Good\_comments des classes avec plus de 10 modifications est de 0,02138612 est supérieur à 0,01599458 pour les classes avec moins de 10 modifications. Dans le même temps, nous pouvons également constater à partir de l'histogramme que les Good\_comments des classes avec plus de 10 modifications sont fondamentalement plus que celles avec moins de 10 modifications.

On peut donc supposer que 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" est correcte.

### 5. discussion des menaces à la validité

Quant à la métrique subjective des "bons commentaires", nous n'utilisons la métrique de densité de commentaires que quantitativement, en ignorant la quantité de code de la classe et la qualité réelle des commentaires. Par conséquent, pour cette étude quasi-expérimentale, nous avons affirmé l'hypothèse proposée sur la base des indicateurs définis par nous-mêmes, et avons également eu une certaine influence sur la validité en ignorant les facteurs subjectifs et certains indicateurs potentiels internes.