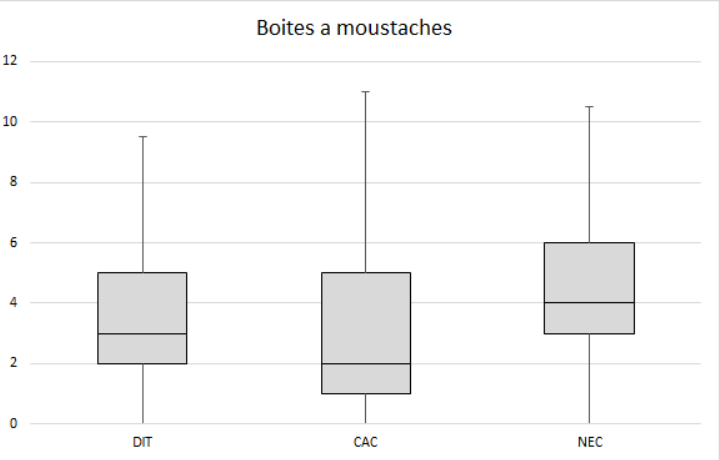
IFT-3913 Qualité de Logiciel et Métriques Julien Lanctôt (20140970)  
Travail Pratique 2 Laurence Fortin (20088891)

T1. **Visualisez** chacune des métriques de l’échantillon en créant les boites à moustaches. Calculez les informations pertinentes et décrivez les distributions.

Ici, nous pouvons voir la représentation des boîtes à moustaches pour DIT, CAC et NEC



s = 9.5

s = 11

s = 10.5

u = 5

u = 5

u = 6

m = 3

m = 2

m = 4

l = 2

l = 1

l = 3

d = 3

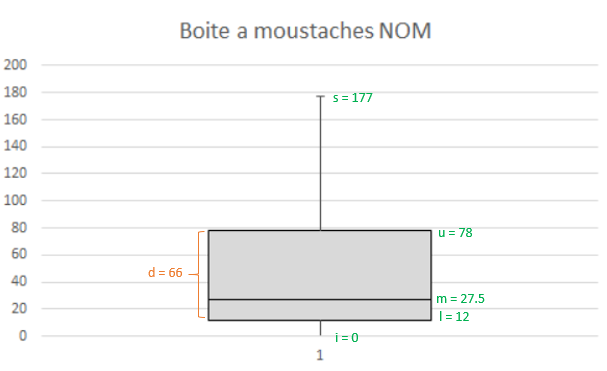
d = 4

d = 3

i = 0

i = 0

i = 0



T2.

Le choix d’étude pour ce cas précis est une étude de cas. Cela s’explique en regardant les données qui nous sont présentées. Nous savons qu’il n’y qu’une seule occurrence d’un phénomène, ce dernier est les 30 classes que nous avons collectées 1 seule fois, en plus d’avoir été obtenu lors de la mesure d’un logiciel hypothétique. Comme nous devons confirmer ou infirmer une théorie, cela nous permet de valider que c’est bel et bien une étude de cas que l’on examine.

En analysant l’hypothèse donnée dans l’énoncé, nous arrivons à l’hypothèse nulle qui est que les classes ayant un nombre de méthodes inférieur ou égal à 30 n’ont pas moins d’erreurs que celles ayant plus de 30 méthodes

Nous pouvons observer que la variable d’état dans notre cas est le nombre de méthodes (NOM) puisque c’est la variable qui peut être manipulée (en augmentant ou diminuant le nombre de méthodes), c’est également la variable qui influence les résultats de l’étude. Puisque des valeurs du nombre d’erreurs (NEC) sont le résultat de notre étude de cas, nous sommes en mesure de conclure que ce dernier est la variable dépendante de notre étude.

Pour évaluer les 2 métriques (NOM & NEC), nous avons choisi la moyenne et la médiane comme outil de comparaison. Lorsque l’on regarde les classes qui ont moins de 30 méthodes (C1 à C16) et qu’on les compare avec les classes de plus de 30 méthodes (C17 à C30) la moyenne et la médiane du nombre de méthodes (NOM) passent respectivement de 15.25 à 90.79 et de 13.31 à 7.79. Pour la moyenne et la médiane du nombre d’erreurs (NEC) qu’en a eu de 13.5 à 83 et de 3.5 à 5.5 (voir [Annexe 1](#_ANNEXE_1_:)). Nous pouvons apercevoir que le nombre d’erreurs augmente bel et bien lorsque le nombre de méthodes augmente aussi. Cela étant dit, cette mesure n’est pas précise, nous ne pouvons pas dire qu’il y a une corrélation directe entre ces 2 métriques seulement avec cette mesure.

Discussion des menaces à la validité

T3. **Étudier les corrélations** individuelles entre les métriques de structure et NEC. Visualisez les données, les droits de régression, etc., et expliquez pourquoi (ou pourquoi pas) ces visualisations sont significatives (ou pas). Dans cette étape, vous ne prenez pas de décisions : vous explorez et vous étudiez l’ensemble de données. Explorez aussi le scénario où vous supprimez les valeurs aberrantes, si cela a du sens.

Tout d’abord, en regardant les valeurs des boites à moustaches, on peut voir que 2 classes possèdent des valeurs aberrantes [les valeurs qui dépassent la valeur de la limite supérieure (s)] dans 2 attributs : C21, qui possède un attribut CAC de 17 (plus grand que 11) et C30, qui possède un NOM de 184 (plus grand que 177). Il sera intéressant d’étudier nos corrélations en tenant compte de ses valeurs et en ne les prenant pas en compte.

En regardant la distribution de la fréquence toutes nos métriques (voir [Annexe 2](#_ANNEXE_2_:)), on remarque qu’aucune d’entre elles n’est normalement distribuée. Cela nous indique qu’il faut utiliser le coefficient de corrélation du rang de Spearman (ρ) pour effectuer nos calculs.

En regardant les corrélations entre les métriques de structure et NEC, nous apercevons les coefficients sont minimes. Que ce soit en considérant ou pas les valeurs aberrantes, les coefficient changent de façon négligeable. Il nous est donc possible de conclure qu’il y a pratiquement aucune corrélation dans notre cas.

|  |  |  |  |
| --- | --- | --- | --- |
|  | NEC et NOM | NEC et CAC | NEC et DIT |
| Spearman Coefficient | 0.314992618 | 0.216015407 | 0.225183507 |

Avec données aberrantes

|  |  |  |  |
| --- | --- | --- | --- |
|  | NEC et NOM | NEC et CAC | NEC et DIT |
| Spearman Coefficient | 0.396937052 | 0.242472242 | 0.282848599 |

Sans données aberrantes

Comme vu plutôt, notre modèle possède 2 données aberrantes, qu’on les prenne en compte ou non, notre modèle reste plutôt similaire. Cela nous indique que peut importe si

En regardant les droites de régressions linéaires des métriques (voir [Annexe 3](#_ANNEXE_3_:)-[4](#_ANNEXE_3_:_1)), nous sommes en mesure de voir qu’aucune de ces dernières semblent linéaire.

T4. **Évaluer les hypothèses suivantes :**

a. Le nombre d’erreurs est une fonction linéaire du NOM

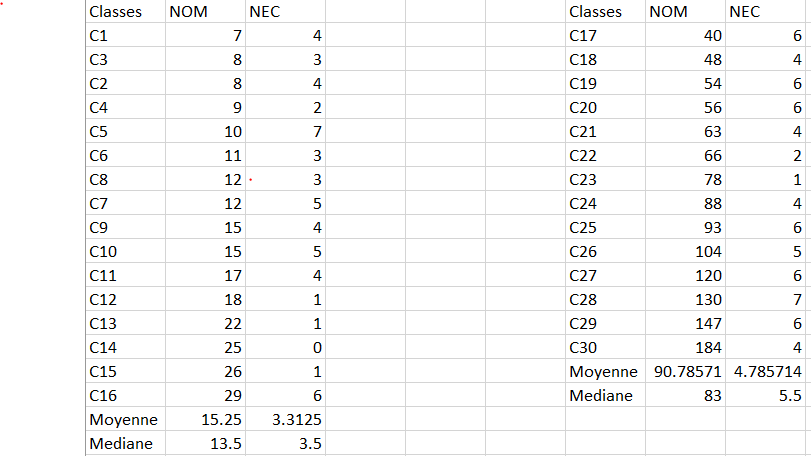
b. Le nombre d’erreurs est une fonction linéaire du DIT

* 1. c. Le nombre d’erreurs est une fonction linéaire du CAC

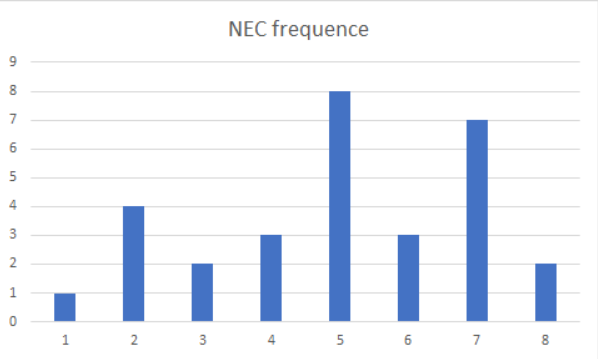
Décrire d’abord la conception de l’étude [comme en T2] et discuter les résultats par la suite.

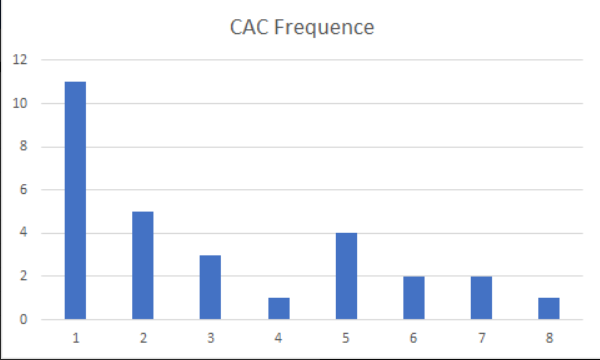
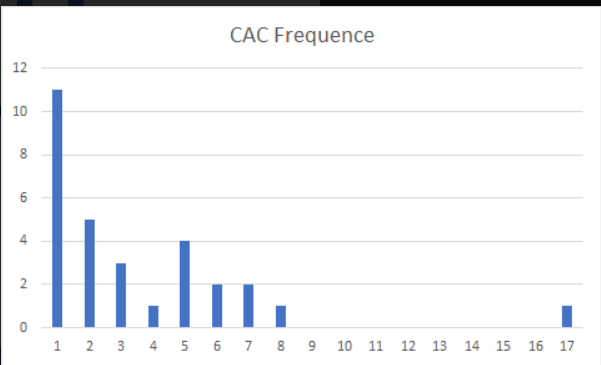
T5. **Décrivez vos conclusions** dans un court paragraphe.

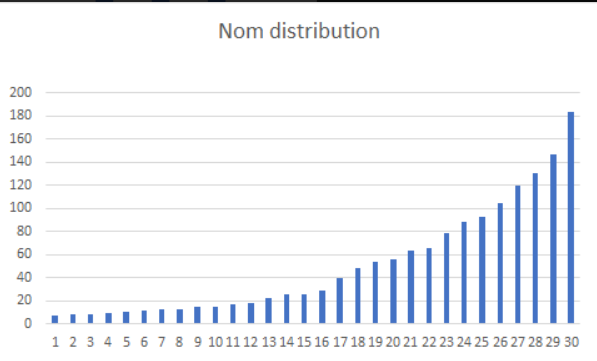
## ANNEXE 1 : Calcul de la moyenne et médiane



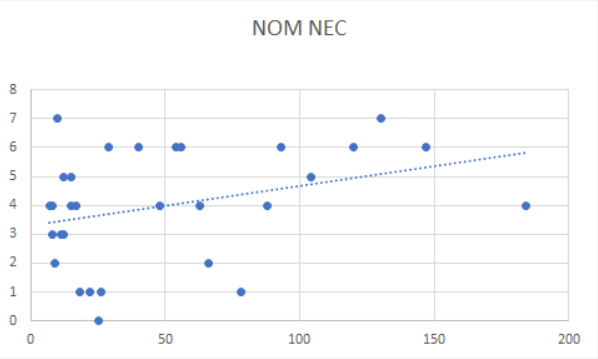
# ANNEXE 2 : Fréquence des métriques pour évaluer la distribution

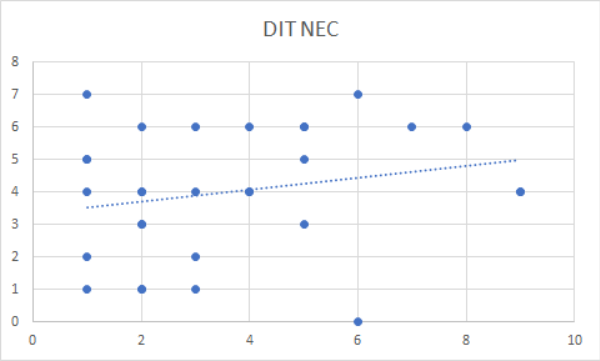


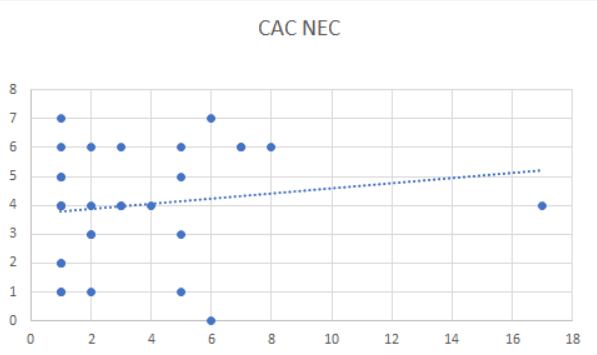


La fréquence de CAC avec et sans la donnée aberrante

# ANNEXE 3 : Droite de régression linéaire avec valeur aberrante







# ANNEXE 4 : Droite de régression linéaire sans valeur aberrante

