

Rapport TP 3 - IFT3913

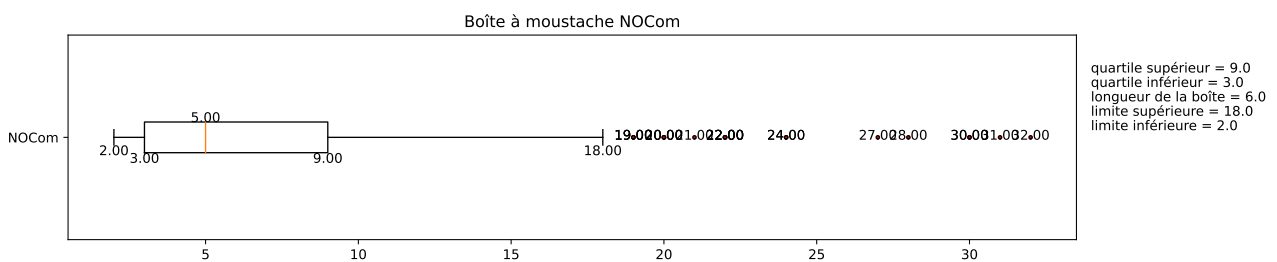
Loïc Daudé Mondet
20243814 – Programmes d'échanges - 1er c.(Échange)

Alaa edlin Yacoub

18/11/2022

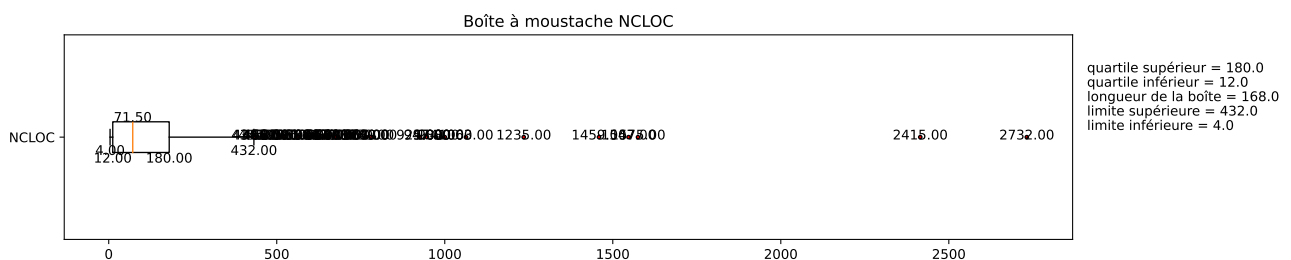
I – Boîtes à moustache et distribution des données

I.1 NOCom



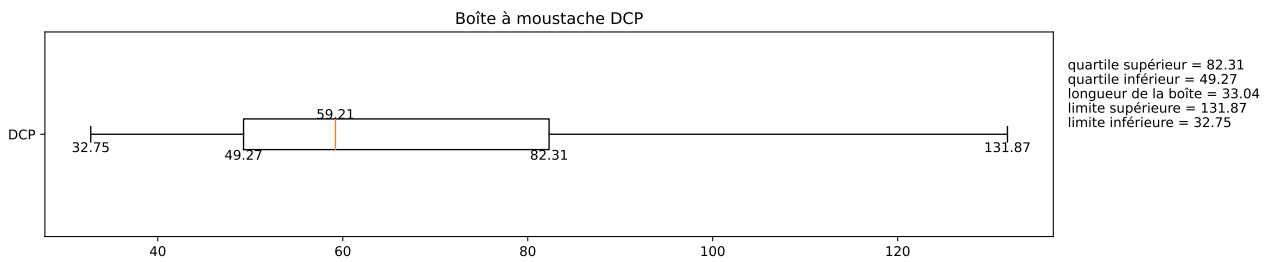
La boîte n'est pas symétrique ce qui signifie que les variables ne sont pas normalement distribuées, conclusion renforcée par le grand nombre de valeurs extrêmes.

I.2 NCLOC



Ici aussi, la boîte n'est pas symétrique et les variables ne sont donc encore une fois pas normalement distribuées, de plus on observe un nombre encore plus élevé de valeurs extrêmes.

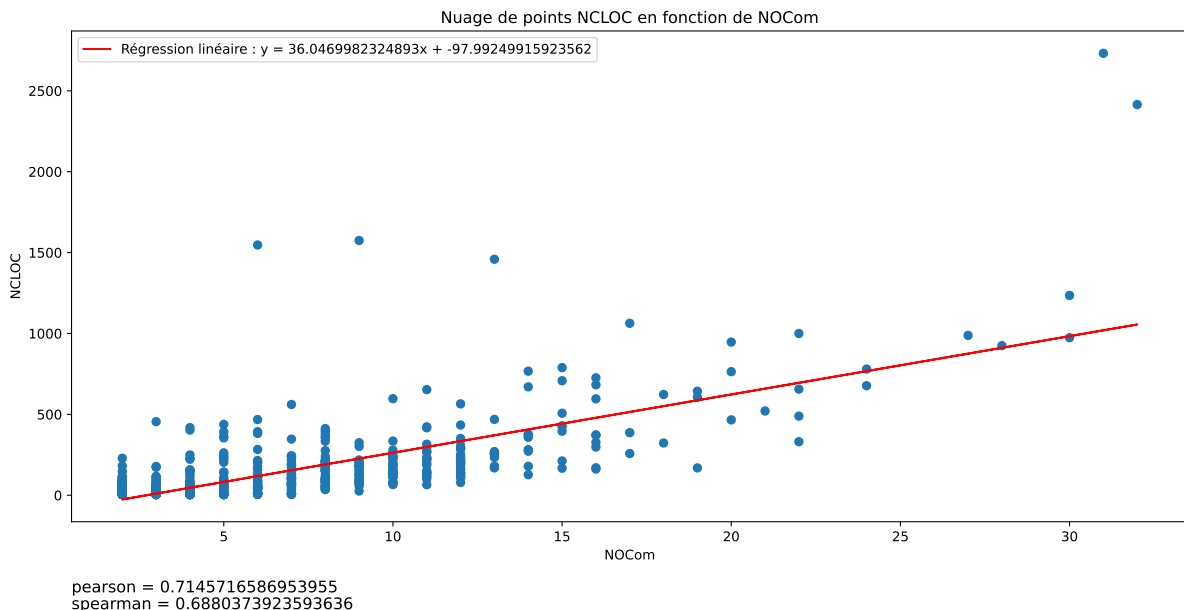
I.3 DCP



Pour DCP, la boîte bien que toujours asymétrique, les longueurs des moustaches sont sensiblement plus proches que dans les cas précédents et le placement de la médiane dans la boîte est également plus centré, on note également l'absence de valeurs extrêmes, cela traduit donc une distribution des variables plus proche d'une distribution suivant la loi normale.

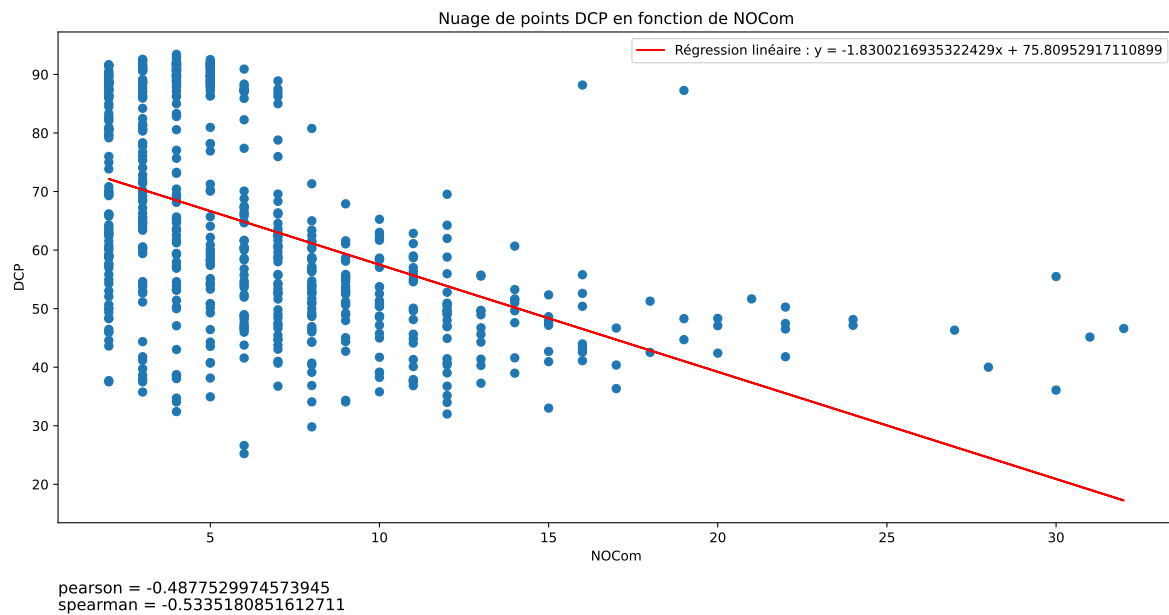
II – Correlations

II.1 NCLOC en fonction de NOCom



On remarque que la plupart des points du graphique représentant NCLOC en fonction de NOCom semblent se répartir autour et en suivant une droite. Comme les variables ne sont pas normalement distribuées (cf la partie précédente), on utilise le coefficient de corrélation de Spearman et on trouve 0.688 ce qui témoigne d'une certaine relation relativement linéaire entre ces deux métriques puisque cette valeur est plus proche de 1 que de 0. On trouve alors en effectuant la régression linéaire l'équation suivante $y = 36,05x - 97,99$ décrivant cette relation.

II.2 DCP en fonction de NOCom



Ici, les points ne semblent pas vraiment suivre une même droite, cette observation est confirmée par la valeur de -0,53 du coefficient de Spearman qui bien que légèrement plus proche de -1 que de 0 ne constitue pas une valeur témoignant d'une relation très proche de la linéarité entre les deux métriques. Si on essaye toutefois d'obtenir une régression linéaire entre ces deux variables, on trouve l'équation $y = -1,830x + 75,80$.

III – 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 »