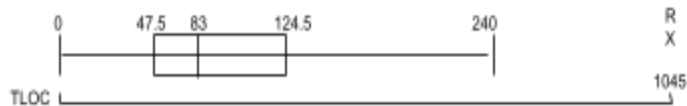
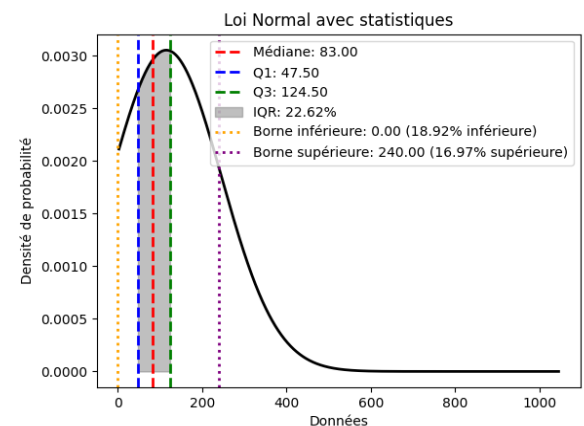


T1TLOC

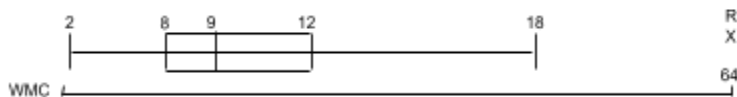


En observant la distribution de TLOC, il est possible de voir que seulement 22.62% des valeurs se retrouvent entre le quartile supérieur et le quartile inférieur. Cela veut dire que la distribution n'est pas normale étant donné qu'environ 50% des valeurs devraient se retrouver dans le IQR. Alors, il est possible de d'assumer que les valeurs de TLOC présentent beaucoup de points extrêmes, plus précisément, presque 42% des valeurs sont inférieures ou supérieures à i ou s . Avec un point extrême se trouvant à 1045.

$l =$	47.5
$m =$	83
$u =$	124.5
$d = u - l =$	77
$s = u + 1,5d =$	240
$i = l - 1.5d =$	0 (-68)

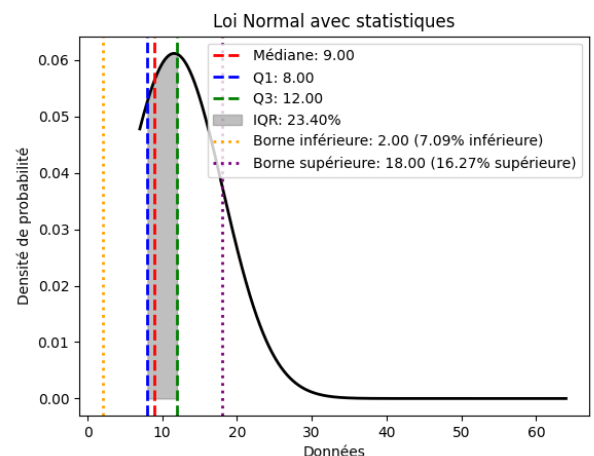


WMC

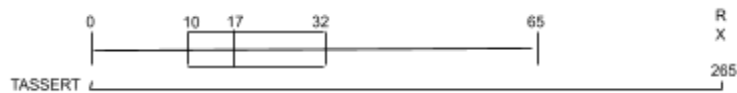


En observant la distribution de WMC, il est possible de voir que seulement 23.40% des valeurs se retrouvent dans le IQR. Les mêmes observations que le TLOC sont faites, mais avec des valeurs encore moins bien distribuées que TLOC. Plus de 54% des valeurs étant des points extrêmes. De plus, il est possible de constater que peu de valeurs se trouvent dans la borne inférieures (7,09%). Dans une distribution normale, la borne supérieure et inférieure devrait contenir respectivement 24.6% des valeurs.

$l =$	8
$m =$	9
$u =$	12
$d = u - l =$	4
$s = u + 1,5d =$	18
$i = l - 1.5d =$	2

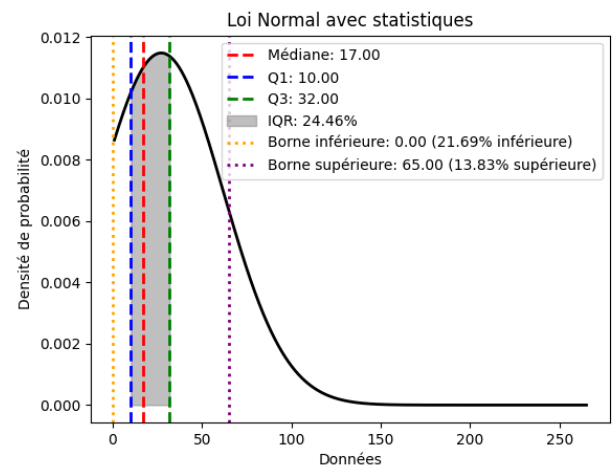


TASSERT



$I =$	10
$m =$	17
$u =$	32
$d = u - I =$	22
$s = u + 1,5d =$	65
$i = I - 1.5d =$	0 (-23)

En observant la distribution de TASSERT, il est possible de voir que seulement 24.46% des valeurs se retrouvent dans le IQR. Les mêmes observations que le TLOC sont faites, mais avec des valeurs un peu mieux distribuées. Plus de 40% des valeurs étant des points extrêmes. La distribution reste très similaire aux 2 autres métriques, mais est légèrement mieux distribuée.



T2

Étude des corrélations :

Comme nous pouvons le voir visuellement,

Tloc / Tassert

Voici les calculs pour Tloc et Tassert

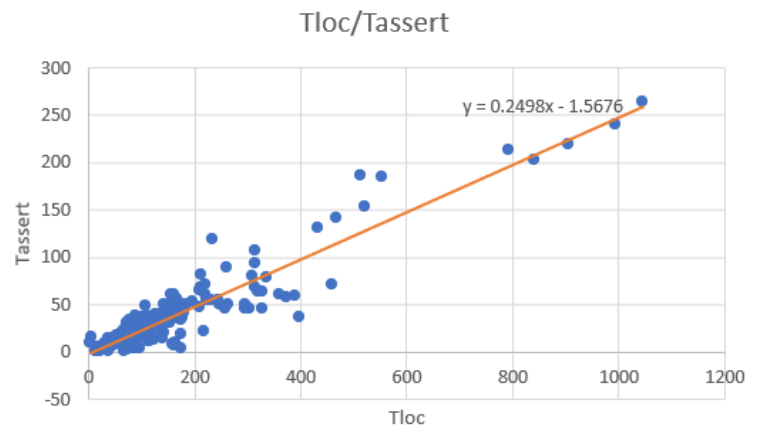
Écarts-types Tloc = 130.88

Écarts-types Tassert = 34.79

Covariance = 4278.76

$r = 0.94$

Comme nous pouvons le voir, selon le test de corrélation de Pearson, nous obtenons un $r = 0.94$. Cette valeur nous permet de dire que ces deux données se rapprochent beaucoup d'une corrélation linéaire positive parfaite. De plus visuellement, nous pouvons remarquer que dans le graphique, nous avons que les valeurs suivent parfaitement la régression linéaire



Voici les calculs pour Wmc et Tassert

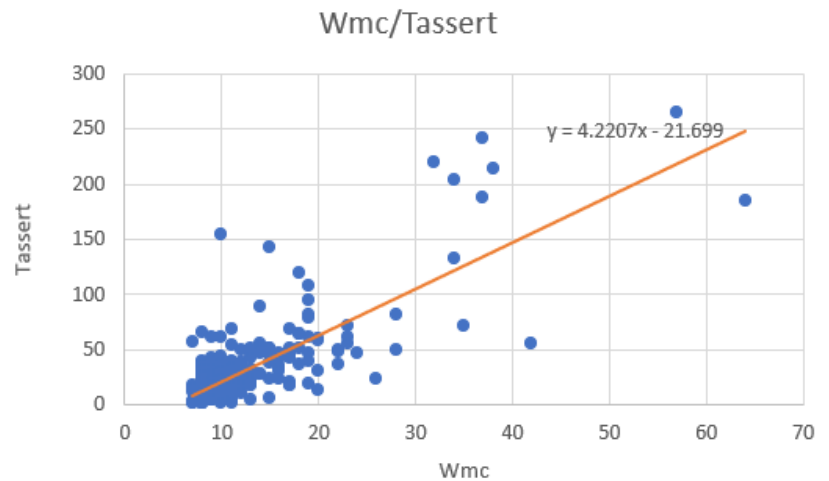
Écart-types Wmc = 6.53

Écart-types Tassert = 34.79

Covariance = 180.16

$r = 0.79$

Pour cette relation, nous avons un $r = 0.79$. Encore une fois, ceci signifie une corrélation positive. Cette fois si, nous sommes un peu plus loin de la valeur de 1 mais sans être alarmant. Même chose pour le graphique, les points suivent relativement la régression linéaire. Nous pouvons donc quand même conclure qu'il y a une corrélation positive.



T3

Description de la quasi-expérience:

Hypothèses :

Hypothèse principale: Les classes avec plus de 20 assertions sont plus complexes

Hypothèse nulle: Il n'y a pas de différence significative de complexité entre les classes avec plus de 20 assertions et celles avec moins.

Nos variables sont les suivantes :

Variable indépendante: Nombre d'assertions dans une classe.

Variable dépendante: Complexité des classes.

Résultats obtenus :

Donc pour commencer, vu que nos tests ne sont pas normalement distribués. Nous avons donc décidé de faire des tests de Mann-Whitney. Selon le test tloc et tassert, nous obtenons un score z de 18.39314, avec une valeur $p < .00001$, se qui rend le résultat significatif ($p < 0.05$).

<<https://www.socscistatistics.com/tests/mannwhitney/default2.aspx>>

Nous pouvons donc dire que ce test suggère fortement que les classes contenant plus de 20 assertions sont statistiquement différentes en termes de complexité par rapport à celles qui en contiennent moins. Pour wmc et tassert, nous obtenons un score z de -9.56023 , avec une valeur $p < .00001$, se qui rend le résultat significatif ($p < 0.05$).

<<https://www.socscistatistics.com/tests/mannwhitney/default2.aspx>>

Nous obtenons un renforcement sur l'idée que les classes contenant plus de 20 assertions sont statistiquement différentes en termes de complexité.

Selon nos graphiques calculés plus haut, nous avons deux graphiques qui mettent en relation le nombre de lignes de code et le nombre de assert et la relation entre le poids des méthodes par classe et le nombre de assert. Comme il est possible de remarquer, les points sur le graphique suivent presque parfaitement la régression linéaire. On peut donc dire que le nombre de assert est linéairement. Ce qui renforce l'idée que plus il y a d'assertions, plus il y a de la complexité.

En conclusion, on peut dire que selon nos résultats, nous obtenons bel et bien que les classes qui contiennent plus de 20 assertions sont plus complexes que celles contiennent moins de 20 assertions.

Menaces de validité :

Il est possible que le test Mann-Whitney ne soit pas le test le plus approprié pour nos données, il pourrait y avoir un test qui serait plus utile. Il est aussi possible que d'autres facteurs, hors du nombre d'insertions , influencent la complexité des classes, comme la qualité du code. Il se peut aussi qu'il y ait une raison autre que tloc et wmc qui peut engendrer la complexité des classes.

Les graphiques ont été créés à l'aide d'un script python et de Excel. Le code se retrouve sur le repository git.