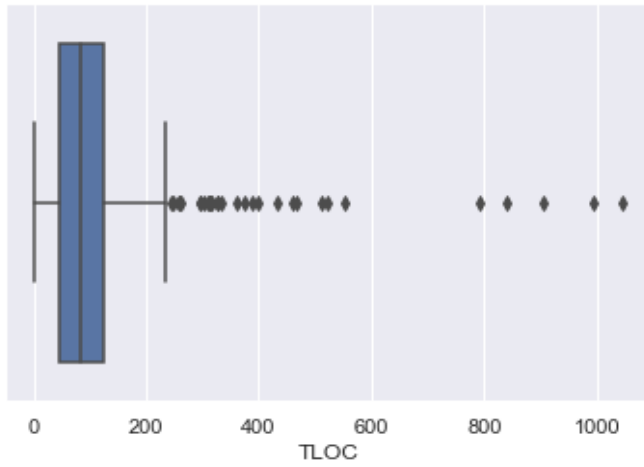
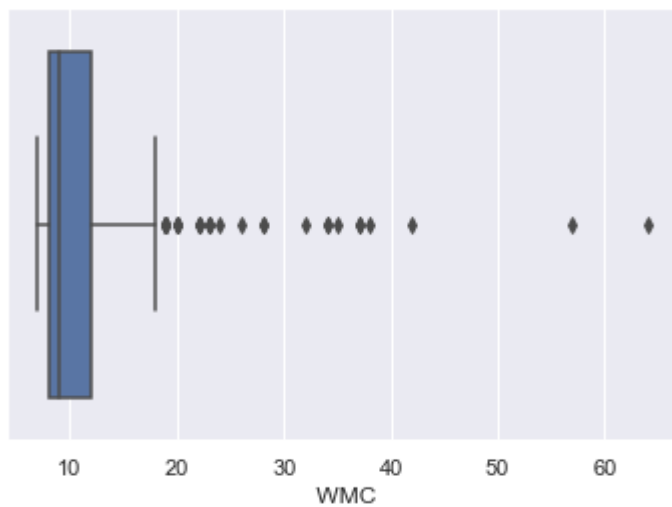


TP 3 – Rapport

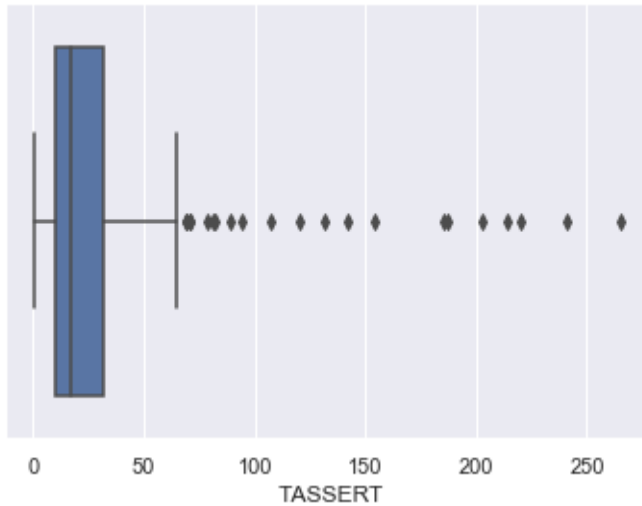
Tache 1:



m	u	l	d	s	i
83	125	46	79	244	3



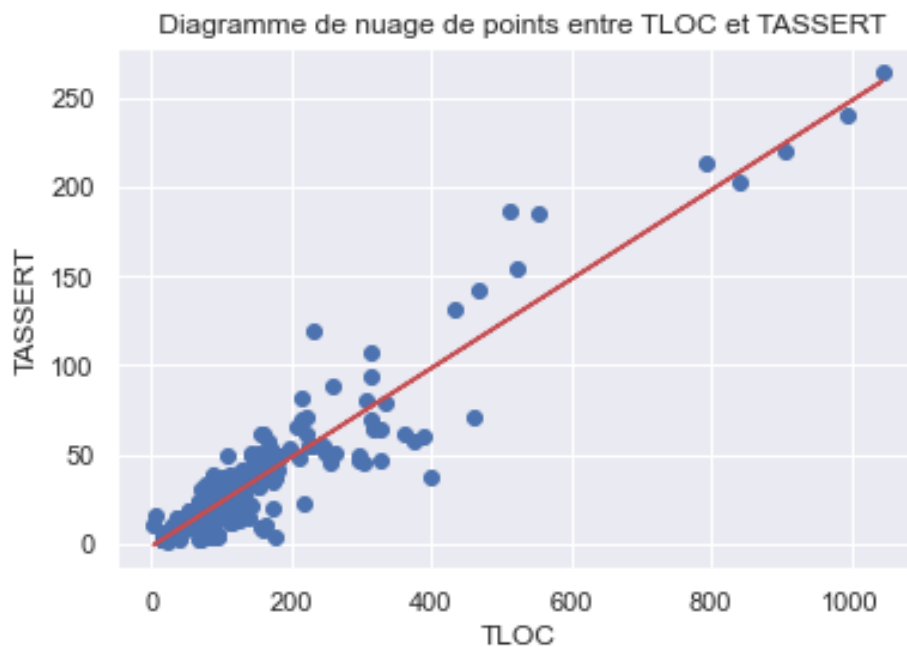
m	u	l	d	s	i
9	12	8	4	18	2



m	u	l	d	s	i
17	33	10	23	68	1

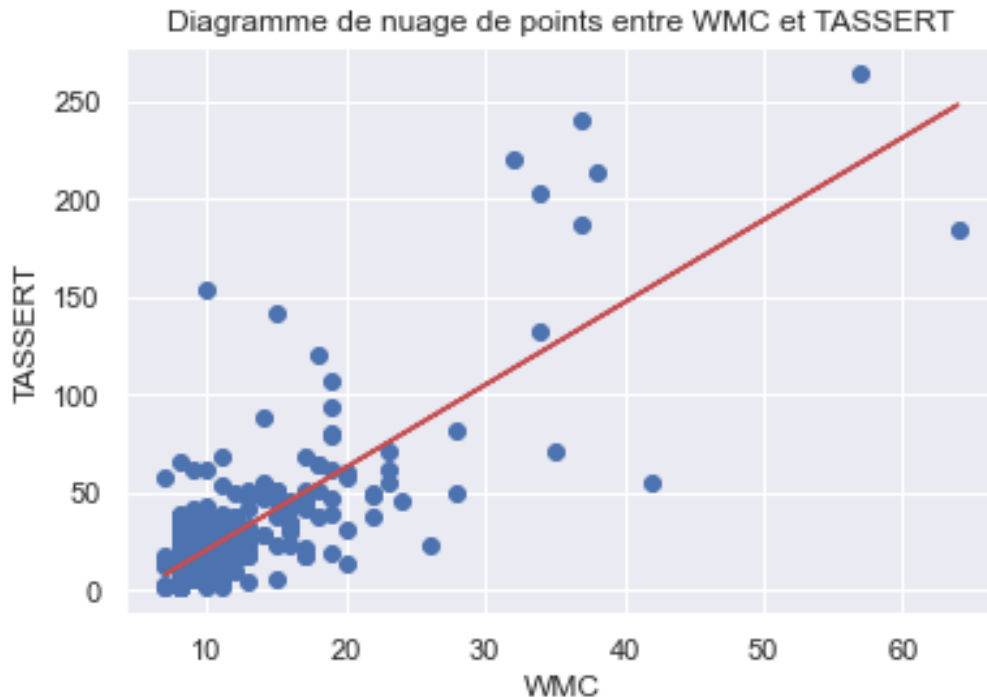
Descriptions : Les trois distributions ont beaucoup de points extrêmes vers la droite (positives) et donc ils ne sont pas symétriques.

Tache 2:



- La droite de régression : **$0.25x - 1.57$**
- Le coefficient de corrélation de Pearson (r) : 0.94

- Le coefficient de corrélation de Spearman (ρ) : 0.83
- Le fait que la distribution n'est pas normale, on prend en considération Le coefficient de corrélation de Spearman ($\rho = 0.83$) qui indique que une forte corrélation positive entre TLOC (Nombre de lignes de code non-vides qui ne sont pas de commentaires) et TASSERT (nombre d' assertions) .



- La droite de régression : $4.22x - 21.7$
- Le coefficient de corrélation de Pearson (r) : 0.79
- Le coefficient de corrélation de Spearman (ρ) : 0.61
- Le fait que la distribution n'est pas normale, on prend en considération Le coefficient de corrélation de Spearman ($\rho = 0.83$) qui indique que une corrélation positive modérée entre WMC(Méthodes Pondérées par Classe) et TASSERT (nombre d' assertions) .

Tâche 3:

Choix d'étude : Quasi-expérience

Affectation aux groupes contrôlée :

Nous allons diviser l'ensemble de données (test) en deux groupes :

- Les classes qui ont plus de 20 assertions (TASSERT > 20)
- Les classes qui ont moins de 20 assertions (TASSERT < 20)

Énoncé des hypothèses :

H_0 : « Les classes qui contiennent plus de 20 assertions ont une valeur de WMC plus grande que celles qui ont moins de 20 assertions ».

Définition des variables :

Variables indépendantes : WMC

Variables dépendantes : TASSERT

Nous utilisons WMC pour mesurer la complexité des classes, et nous utilisons TASSERT pour diviser les données en deux groupes : celles qui ont plus de 20 assertions et celles qui ont moins de 20 assertions. TLOC ne mesure ni les assertions ni la complexité, nous ne le prenons donc pas en considération.

Interprétation et généralisation des résultats :

```
1 #calcul de la moyenne de wmc pour les deux groupes de classes (plus et moins 20 assertions)
2 plus20_assertion=jfree_donnees[jfree_donnees[' TASSERT'] >= 20]
3 moins20_assertion=jfree_donnees[jfree_donnees[' TASSERT'] < 20]
4 wmc_plus_moyen=plus20_assertion[' WMC'].mean()
5 wmc_moins_moyen=moins20_assertion[' WMC'].mean()
6 print("la moyenne de wmc pour les classes qui ont plus que 20 assertions : ",wmc_plus_moyen)
7 print("la moyenne de wmc pour les classes qui ont moins que 20 assertions : ",wmc_moins_moyen)
```

```
la moyenne de wmc pour les classes qui ont plus que 20 assertions : 15.25531914893617
la moyenne de wmc pour les classes qui ont moins que 20 assertions : 9.119047619047619
```

On constate que la moyenne de WMC des classes qui ont plus de 20 assertions est plus élevée que celle des classes qui en ont moins de 20. De plus, si l'on souhaite généraliser, lors de l'étape précédente, nous avons calculé le coefficient de corrélation de Spearman entre WMC et TASSERT et avons observé que $p = 0.61$, indiquant une corrélation positive entre les deux variables. Malgré cela, nous avons identifié des contre-exemples où des classes ayant plus de 20 assertions étaient moins complexes (en termes de valeur de WMC) que des classes ayant moins de 20 assertions.

```
1 wmc_plus_min=plus20_assertion[' WMC'].min()
2 wmc_moins_max=moins20_assertion[' WMC'].max()
3 print("le valeur min de wmc pour les classes qui ont plus de 20 assertions : ",wmc_plus_min)
4 print("le valeur max de wmc pour les classes qui ont moins de 20 assertions : ",wmc_moins_max)
```

```
le valeur min de wmc pour les classes qui ont plus de 20 assertions : 7
le valeur max de wmc pour les classes qui ont moins de 20 assertions : 20
```

Avec ces constatations, on en conclut qu'avoir plus d'assertions dans une classe ne signifie pas toujours qu'elle est plus complexe et donc on ne peut ni accepter ni refuser cette hypothèse.

Menace à la validité :

- Validité interne :

La validité interne est potentiellement compromise, car nous n'avons pas pris en considération toutes les variables qui pourraient influencer nos données.

Validité externe :

Nous aurions besoin d'appliquer notre hypothèse sur d'autres projets pour pouvoir mener une étude comparative plus approfondie et tirer des conclusions générales.