

T1

TLOC

$$\text{Médiane}(m) = 83$$

$$\text{Premier quartile}(l) = 47$$

$$\text{Troisième quartile}(u) = 124$$

$$\text{Longueur}(d) = u - l = 124 - 47 = 77$$

$$\text{Limite supérieure}(s) = u + 1.5d = 124 + 1.5(77) = 239.5$$

$$\text{Limite inférieure}(i) = l - 1.5d = 47 - 1.5(77) = -68.5$$

Pour la métrique TLOC, la longueur de la boîte est 77, ce qui signifie qu'il y a une grande variabilité dans le nombre de lignes de code entre le premier et le troisième quartile. La limite supérieure est 239.5, ce qui signifie qu'au-delà de cette valeur, les classes sont considérées comme des points aberrants. La limite inférieure est -68.5, ce qui signifie qu'en dessous de cette valeur, les classes sont également considérées comme des points aberrants.

WMC

$$\text{Médiane}(m) = 9$$

$$\text{Premier quartile}(l) = 8$$

$$\text{Troisième quartile}(u) = 12$$

$$\text{Longueur}(d) = u - l = 12 - 8 = 4$$

$$\text{Limite supérieure}(s) = u + 1.5d = 12 + 1.5(4) = 18$$

$$\text{Limite inférieure}(i) = l - 1.5d = 8 - 1.5(4) = 2$$

Pour la métrique WMC, la longueur de la boîte est 4, ce qui signifie qu'il y a une faible variabilité dans le nombre de méthodes pondérées entre le premier et le troisième quartile. La limite supérieure est 18, ce qui signifie qu'au-delà de cette valeur, les classes sont considérées comme des points aberrants. La limite inférieure est 2, ce qui signifie qu'en dessous de cette valeur, les classes sont également considérées comme des points aberrants.

TASSERT

$$\text{Médiane}(m) = 17$$

$$\text{Premier quartile}(l) = 10$$

$$\text{Troisième quartile}(u) = 32$$

$$\text{Longueur}(d) = u - l = 32 - 10 = 22$$

$$\text{Limite supérieure}(s) = u + 1.5d = 32 + 1.5(22) = 65$$

$$\text{Limite inférieure}(i) = l - 1.5d = 10 - 1.5(22) = -23$$

Pour la métrique TASSERT, la longueur de la boîte est 22, ce qui signifie qu'il y a une grande variabilité dans le nombre d'assertions entre le premier et le troisième quartile. La limite supérieure est 65, ce qui signifie qu'au-delà de cette valeur, les classes sont considérées comme des points aberrants. La limite inférieure est -23, ce qui signifie qu'en dessous de cette valeur, les classes sont également considérées comme des points aberrants.

T2

TLOC et TASSERT

Pour TLOC et TASSERT, nous avons trouvé un coefficient de 0.93957 (calculé avec PEARSON dans Excel), ce qui signifie qu'il y a une corrélation positive très forte entre le nombre de lignes de code et le nombre d'assertions. Cela signifie que plus une classe a de lignes de code, plus elle a tendance à avoir d'assertions, et vice versa. Le nuage de points montre que les points sont très proches de la droite de régression, ce qui confirme la forte relation linéaire (voir dans fichier Excel T2)

WMC et TASSERT

Pour WMC et TASSERT, nous avons trouvé un coefficient de 0.79247 (calculé avec PEARSON dans Excel), ce qui signifie qu'il y a une corrélation positive modérée entre le nombre de méthodes pondérées et le nombre d'assertions. Cela signifie que plus une classe a de méthodes pondérées, plus elle a tendance à avoir d'assertions, mais pas de façon aussi marquée que pour TLOC et TASSERT. Le nuage de points montre que les points sont assez proches de la droite de régression, mais il y a plus de dispersion que pour TLOC et TASSERT (voir dans fichier Excel T2)

T3

Choix d'étude : Nous avons réalisé une quasi-expérience, qui est un type d'étude qui vise à tester une hypothèse causale entre deux variables, mais qui ne respecte pas tous les critères d'une expérimentation. Nous avons utilisé une variable préexistante, le nombre d'assertions dans les classes, pour définir les groupes. Nous avons comparé la moyenne de la complexité des classes entre les groupes, en utilisant un test t pour échantillons indépendants. Nous avons mesuré la complexité des classes en utilisant un outil d'analyse statique du code, et le nombre d'assertions en utilisant une formule Excel. Nous avons choisi ce type d'étude car il nous permet de tester notre hypothèse de manière simple et efficace.

Hypothèses : Notre hypothèse nulle est qu'il n'y a pas de différence entre les groupes. Notre hypothèse alternative est que les classes qui contiennent plus de 20 assertions sont plus complexes que celles qui contiennent moins de 20 assertions.

Définition des variables : La variable indépendante est le nombre d'assertions dans les classes, qui est une variable catégorielle avec deux modalités : TASSERT > 20 et TASSERT < 20. La variable dépendante est la complexité des classes, qui est une variable quantitative mesurée par un indicateur nommé Moyenne. Nous avons mesuré le nombre d'assertions dans les classes en utilisant la formule suivante : =NB.SI(E2:E352;"A") pour les classes qui contiennent plus de 20 assertions, et =NB.SI(E2:E352;"B") pour les classes qui contiennent moins de 20 assertions. Nous avons ensuite classé les classes en deux groupes selon le seuil de 20 assertions, qui est un choix arbitraire.

Interprétation et généralisation des résultats : Nous avons eu du mal à calculer le test de Student, qui est un test statistique qui permet de comparer les moyennes de deux groupes de données. Ce test nous aurait permis de calculer la valeur de p, qui est la probabilité d'obtenir une telle différence entre les moyennes par hasard. En comparant la valeur de p avec un seuil de significativité, par exemple 0,05, nous aurions pu conclure si la différence entre les moyennes est significative ou pas. Si la valeur de p était inférieure au seuil, nous aurions pu rejeter l'hypothèse nulle, qui suppose qu'il n'y a pas de différence entre les groupes, et accepter l'hypothèse alternative, qui suppose que les classes qui contiennent plus de 20 assertions sont plus complexes que celles qui contiennent moins de 20 assertions. Si la valeur de p était supérieure au seuil, nous n'aurions pas pu rejeter l'hypothèse nulle, et nous aurions dû conclure qu'il n'y a pas de différence significative entre les groupes. De plus, nous avons calculé la moyenne de la complexité (WMC) pour chaque groupe de classes, en utilisant la fonction MOYENNE.SI dans Excel. Nous avons obtenu une moyenne de 15,50 pour les classes qui contiennent plus de 20 assertions, et une moyenne de 9,14 pour les classes qui contiennent moins de 20 assertions. Ces moyennes nous auraient permis de faire le test de Student, en utilisant la fonction T.TEST ou la formule manuelle. Nous aurions pu ainsi comparer les moyennes des deux groupes et tester notre hypothèse.

Discussion des menaces à la validité : Nous avons identifié et analysé les facteurs qui peuvent remettre en cause la qualité et la crédibilité de notre étude, tels que les biais, les erreurs, etc. Voici quelques exemples de menaces à la validité de notre étude :

La quasi-expérience ne permet pas d'établir un lien de causalité entre le nombre d'assertions et la complexité des classes, car il peut y avoir des variables confondantes qui influencent les deux variables. Par exemple, les classes qui contiennent plus de 20 assertions peuvent appartenir à des domaines plus difficiles, ou avoir des auteurs différents, ou utiliser des langages de programmation différents, etc. Ces facteurs peuvent biaiser les résultats et rendre l'interprétation plus délicate. Pour contrôler ces variables confondantes, il faudrait utiliser des techniques statistiques avancées.

Le choix du seuil de 20 assertions pour définir les groupes est arbitraire et peut être remis en question. Pourquoi pas 15, 25 ou 30 assertions ? Il faut justifier ce choix en se basant sur des critères objectifs, tels que la distribution des données, la littérature scientifique, le contexte du problème, etc. Il faut aussi vérifier si les résultats sont sensibles au choix du seuil, c'est-à-dire s'ils changent significativement si on modifie le seuil. Pour cela, il faudrait faire des analyses de sensibilité, qui consistent à tester plusieurs valeurs du seuil et à comparer les résultats obtenus.

RÉFÉRENCES

[D7_correction.pdf](#)

[Fonction MOYENNE.SI - Support Microsoft](#)

<https://support.microsoft.com/en-au/office/create-a-box-plot-10204530-8cdf-40fe-a711-2eb9785e510f>

<https://www.scribbr.com/frequently-asked-questions/calculate-pearson-correlation-coefficient-in-excel/#:~:text=You%20can%20use%20the%20PEARSON,the%20significance%20of%20the%20correlation>

.