

# ANALYSE DES DONNÉES DU SITE “CHECK YOUR SMILE”

Nadia Yassine-Diab



**Auteur :** Corentin Leroyer

**Contact :** [corentin.leroyer31@orange.fr](mailto:corentin.leroyer31@orange.fr)

**Date :** 12 juillet 2019

Les différents graphiques ont été créés avec le langage de programmation **R** et le logiciel **RStudio**.

Ce rapport a été réalisé à partir des graphiques créés par Mr. **Sébastien Déjean**, ingénieur de recherche en calcul scientifique, à l'Institut de Mathématiques de Toulouse situé à l'Université Paul Sabatier.

## SOMMAIRE

<b>QUESTIONNAIRE Q0</b>	<b>3</b>
Profil des étudiants	3
Estimation du niveau	3
Équipement et jeu	4
<b>QUESTIONNAIRE Q2</b>	<b>6</b>
Jeux	6
<b>SNAPSHOTS</b>	<b>7</b>
Global	7
Profil des étudiants	7
Résultats des étudiants	8
Utilisation de CYS	9
Profil des étudiants	9
CYS (S3) - Snapshot 2 (S3)	10
CYS (S4) - Snapshot 2 (S4)	11
CMI	13
Profil des étudiants	13
CMI - Snapshot 1	13
Langue de TP	17
Profil des étudiants	17
TP (S3) - Snapshot 2 (S3)	18
TP (S4) - Snapshot 2 (S4)	19
CYS(S3) & CMI	21
<b>ANNEXES</b>	<b>23</b>
Boîte à moustaches	23
Diagramme à bandes	25
Test de Wilcoxon	26

## QUESTIONNAIRE Q0

### Profil des étudiants

Voici le nombre d'étudiants qui ont répondu au questionnaire Q0 par semestre :

Nombre d'étudiants par semestre

SEMESTRE	NOMBRE D'ÉTUDIANTS
S3 2017-8	51
S3 2018-9	29
S4 2017-8	46
S4 2018-9	69

Voici le nombre d'étudiants qui ont répondu au questionnaire Q0 par filière :

Nombre d'étudiants par filière

FILIÈRE	NOMBRE D'ÉTUDIANTS
BIOMIP	18
EEA	131
Medecine	46

### Estimation du niveau

Ce tableau représente la répartition de l'estimation faite par les étudiants sur leur niveau en électronique :

Représentation du niveau en électronique

NIVEAU ESTIMÉ	NOMBRE D'ÉTUDIANTS
Très bon	10
Assez bon	89
Passable	80
Mauvais	14

Ce tableau représente la répartition de l'estimation faite par les étudiants sur leur niveau en anglais :

Représentation du niveau en anglais

NIVEAU ESTIMÉ	NOMBRE D'ÉTUDIANTS
Très bon	15
Assez bon	72
Passable	82
Mauvais	23

Regardons le nombre d'étudiants par estimation du niveau en électronique (en lignes) et du niveau en anglais (en colonnes) :

	Assez bon	Mauvais	Passable	Très bon
Assez bon	34	4	36	13
Mauvais	5	3	5	1
Passable	24	15	39	1
Très bon	7	1	2	0

## Équipement et jeu

Voici l'équipement informatique des étudiants :

Équipement informatique des étudiants

ÉQUIPEMENT	OUI	NON
Ordinateur personnel	191	2
Tablette	58	134
Smartphone	189	4

Voici pourquoi les étudiants jouent sur leur ordinateur, tablette ou smartphone :

Raisons des étudiants pour jouer sur leur équipement informatique

RAISON	NOMBRE D'ÉTUDIANTS
Se détendre	110
Se cultiver	43
Apprendre en s'amusant	43
Passer le temps	82
Autre raison	19

Ce tableau donne le temps de jeu estimé par nombre d'étudiants, avec la moyenne d'âge de ces étudiants :

Temps de jeu des étudiants

TEMPS DE JEU	NOMBRE D'ÉTUDIANTS	AGE MOYEN
Ne joue pas	40	20.3
Jusqu'à 2h/semaine	64	20.1
Jusqu'à 5h/semaine	41	20.3
Jusqu'à 10h/semaine	28	20.2
Jusqu'à 20h/semaine	14	20.1

## QUESTIONNAIRE Q2

### Jeux

Regardons la popularité des jeux de CYS :

Popularité des jeux de CYS

JEU	NOMBRE D'ÉTUDIANTS
Pendu	68
Audiowords	32
Check Your Motus	41
QCM	55
Check Your Taboo	28
Flashcards	21

Le jeu du **Pendu** est le jeu le plus populaire, suivi du **QCM**. Le jeu **Flashcards** étant le plus récent, il est le jeu le moins joué, tout comme le jeu **Check Your Taboo**, ce qui pourrait s'expliquer par le fait que ce jeu demande de l'investissement.

## SNAPSHOTS

### Global

### Profil des étudiants

Ce tableau donne la moyenne aux différentes snapshots des étudiants par semestre :

Résumé : Snapshots - Par semestre

FILIÈRE	NOMBRE D'ÉTUDIANTS	MOY SNAP 1	MOY SNAP 2	MOY SNAP 2 + 4M
S3 2017-18	54	7.1	11	10.7
S3 2018-19	188	6.3	8.2	10.6
S4 2016-17	19	10.5	12.1	
S4 2017-18	81	9.9	14.4	
S4 2018-9	96	9.4	8.1	10.7

Ce tableau donne quant à lui la moyenne aux différents snapshots des étudiants par filière :

Résumé : Snapshots - Par filière

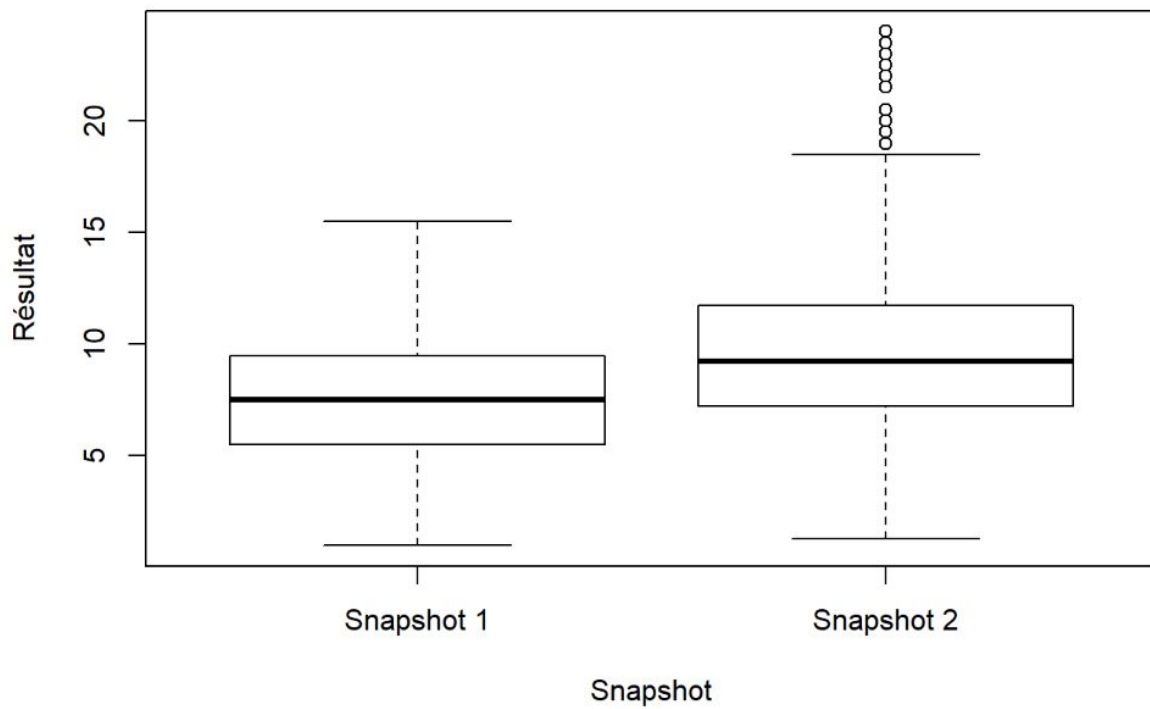
FILIÈRE	NOMBRE D'ÉTUDIANTS	MOY SNAP 1	MOY SNAP 2	MOY SNAP 2 + 4M
BIOMIP	17	9.4		
EEA	373	6.8	8.8	10.7
Medecine	48	10.3	18.3	



## Résultats des étudiants

Représentons la répartition du résultat des étudiants au Snapshot 1 et au Snapshot 2 à l'aide d'une boîte à moustaches :

### Résultats au Snapshot 1 et au Snapshot 2



## Utilisation de CYS

### Profil des étudiants

Ce tableau donne la moyenne au Snapshot 2 (S3) et au Snapshot 2 + 4M (S3) des étudiants selon leur utilisation de CYS au S3 :

Résumé : Snapshots - Utilisation de CYS au S3

UTILISATION CYS S3	NOMBRE D'ÉTUDIANTS	MOY SNAP 2 (S3)	MOY SNAP 2 + 4M (S3)
NON	165	8.5	11.3
OUI	76	9.6	10.2
	1	11.2	

Ce tableau donne la moyenne au Snapshot 2 (S4) et au Snapshot 2 + 4M (S4) des étudiants selon leur utilisation de CYS au S4 :

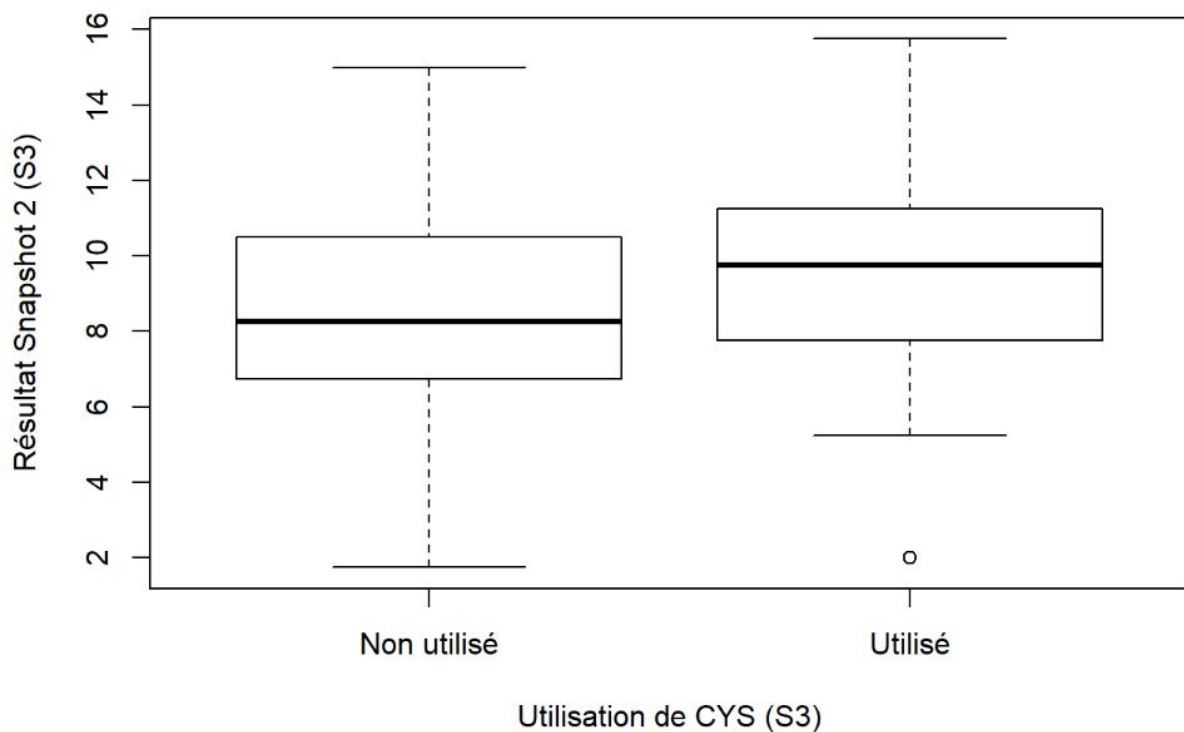
Résumé : Snapshots - Utilisation de CYS au S4

UTILISATION CYS S4	NOMBRE D'ÉTUDIANTS	MOY SNAP 2 (S4)	MOY SNAP 2 + 4M (S4)
NON	96	9.6	
OUI	73	14.5	10.7
	27	10.5	

## CYS (S3) - Snapshot 2 (S3)

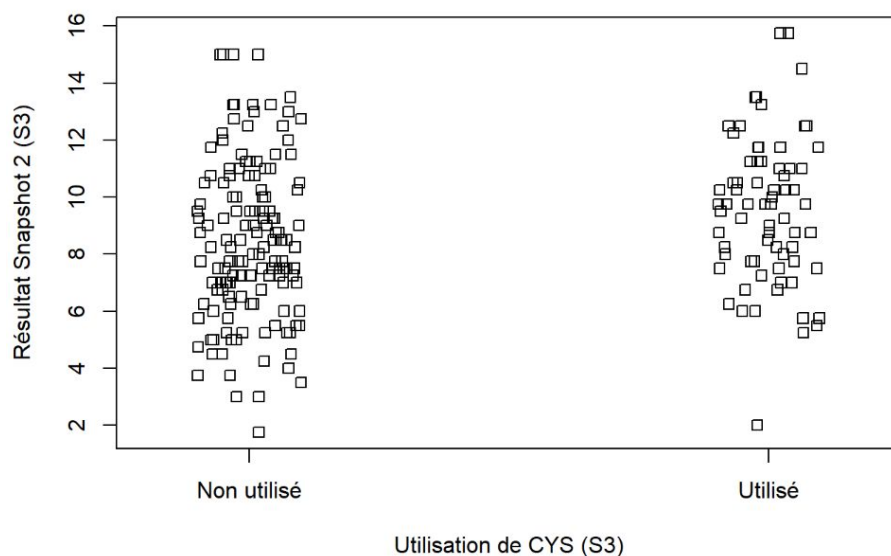
Représentons la répartition du résultat des étudiants au Snapshot 2 (S3) selon leur utilisation de CYS au S3 à l'aide d'une boîte à moustaches :

### Résultats au Snapshot 2 (S3) selon l'utilisation de CYS au S3



Affichons ce graphique autrement en exposant tous les résultats :

### Résultats au Snapshot 2 (S3) selon l'utilisation de CYS au S3



Nous souhaitons vérifier si les résultats des étudiants au Snapshot 2 (S3) sont identiques pour les étudiants ayant utilisé ou non CYS au S3. Pour cela, supposons donc que les résultats au Snapshot 2 (S3) des étudiants ayant utilisé CYS au S3 sont identiques aux résultats au Snapshot 2 (S3) des étudiants n'ayant pas utilisé CYS au S3.

Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

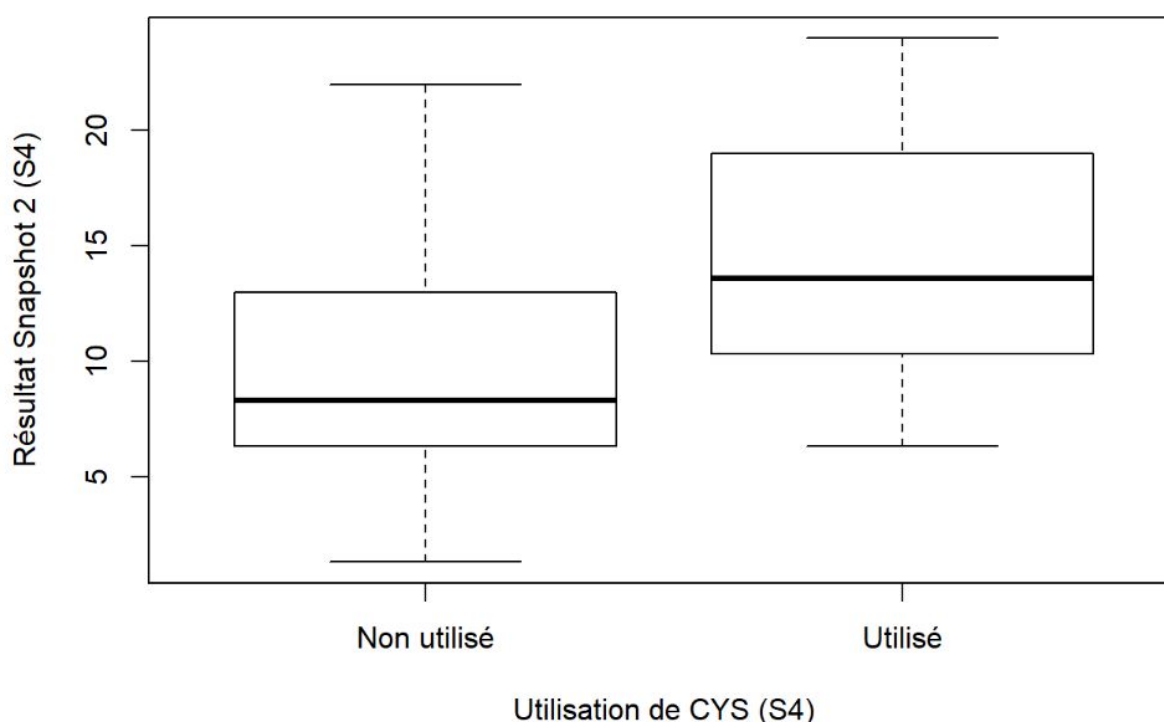
Ce test nous renvoie une **p-value** égale à **0.0047803**. Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 2 (S3) sont identiques selon si les étudiants ont utilisé CYS au S3 ou non.

**Cela veut donc dire que l'utilisation de CYS au S3 aurait un impact sur le résultat au Snapshot 2 (S3) des étudiants.**

### CYS (S4) - Snapshot 2 (S4)

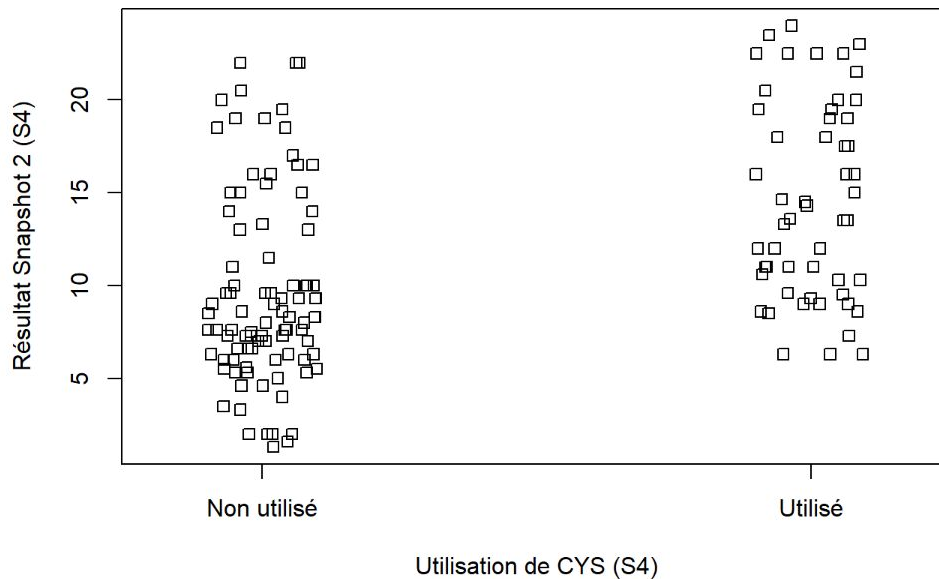
Représentons la répartition du résultat des étudiants au Snapshot 2 (S4) selon leur utilisation de CYS au S4 à l'aide d'une boîte à moustaches :

#### Résultats au Snapshot 2 (S4) selon l'utilisation de CYS au S4



Affichons ce graphique autrement en exposant tous les résultats :

**Résultats au Snapshot 2 (S4) selon l'utilisation de CYS au S4**



Nous souhaitons vérifier si les résultats des étudiants au Snapshot 2 (S4) sont identiques pour les étudiants ayant utilisé ou non CYS au S4. Pour cela, supposons donc que les résultats au Snapshot 2 (S4) des étudiants ayant utilisé CYS au S4 sont identiques aux résultats au Snapshot 2 (S4) des étudiants n'ayant pas utilisé CYS au S4.

Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

Ce test nous renvoie une **p-value** égale à  $1.238285610^{-7}$ . Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 2 (S4) sont identiques selon si les étudiants ont utilisé CYS au S4 ou non.

**Cela veut donc dire que l'utilisation de CYS au S4 aurait un impact sur le résultat au Snapshot 2 (S4) des étudiants.**

Cette conclusion est encore plus significative que la conclusion précédente faite sur l'utilisation de CYS au S3.

## CMI

### Profil des étudiants

Ce tableau donne la moyenne des étudiants selon s'ils sont en CMI ou non :

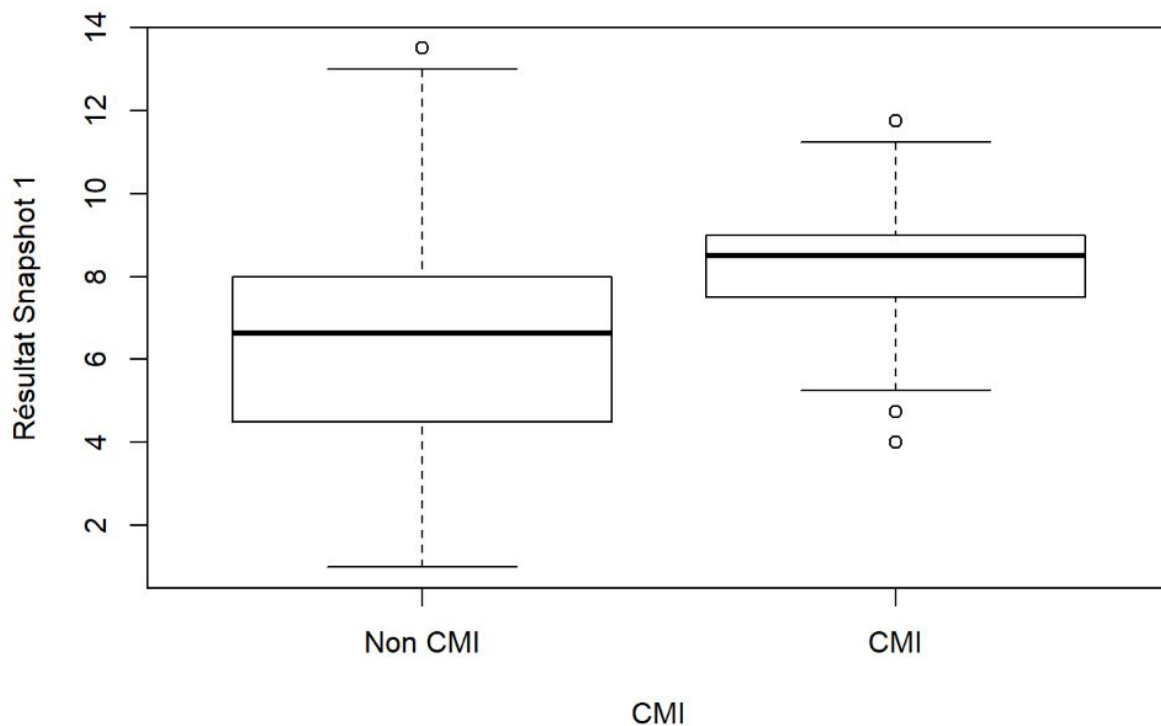
Résumé : Snapshots - Étudiants CMI

CMI	NOMBRE D'ÉTUDIANTS	MOY SNAP 1	MOY SNAP 2	MOY SNAP 2 + 4M
NON	283	6.4	8.2	9.5
OUI	48	8.2	11.2	12
	107	9.4	15.2	10.5

### CMI - Snapshot 1

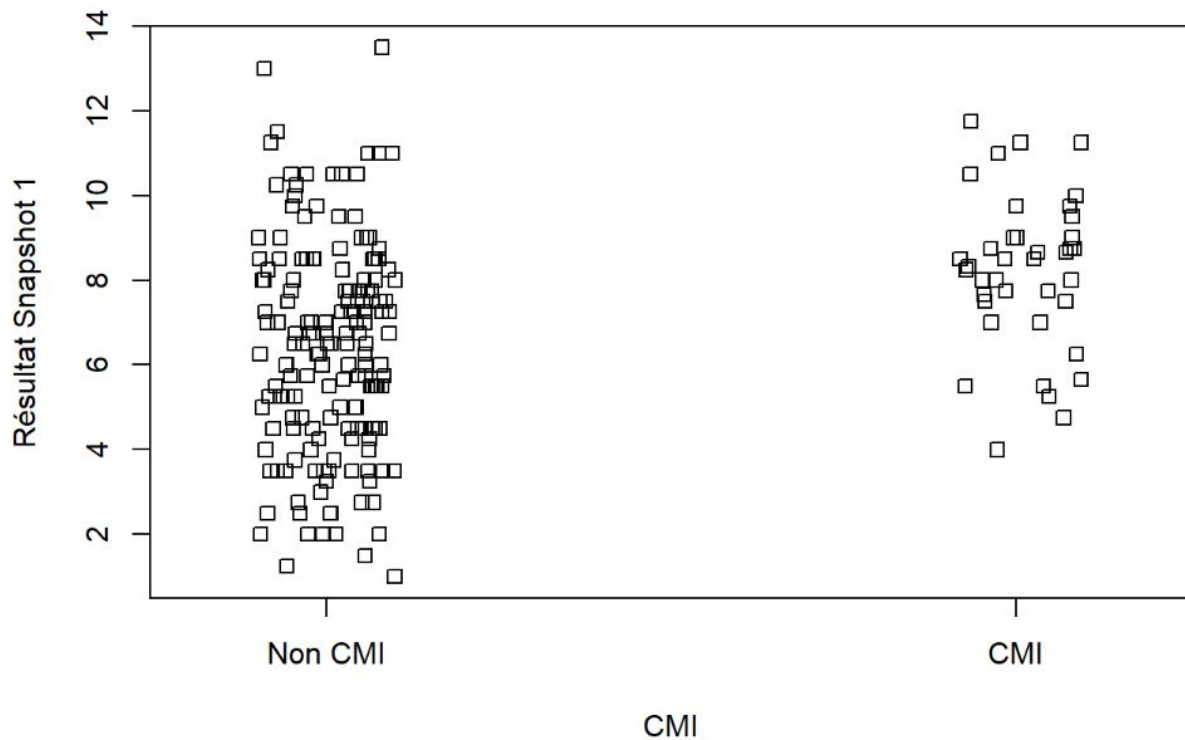
Représentons la répartition du résultat des étudiants au Snapshot 1 selon s'ils sont CMI ou non à l'aide d'une boîte à moustaches :

#### Résultats au Snapshot 1 selon si les étudiants sont CMI ou non



Affichons ce graphique autrement en exposant tous les résultats :

### Résultats au Snapshot 1 selon si les étudiants sont CMI ou non



Nous souhaitons vérifier si les résultats des étudiants au Snapshot 1 sont identiques pour les étudiants CMI ou non. Pour cela, supposons donc que les résultats au Snapshot 1 des étudiants CMI sont identiques aux résultats au Snapshot 1 des étudiants non CMI.

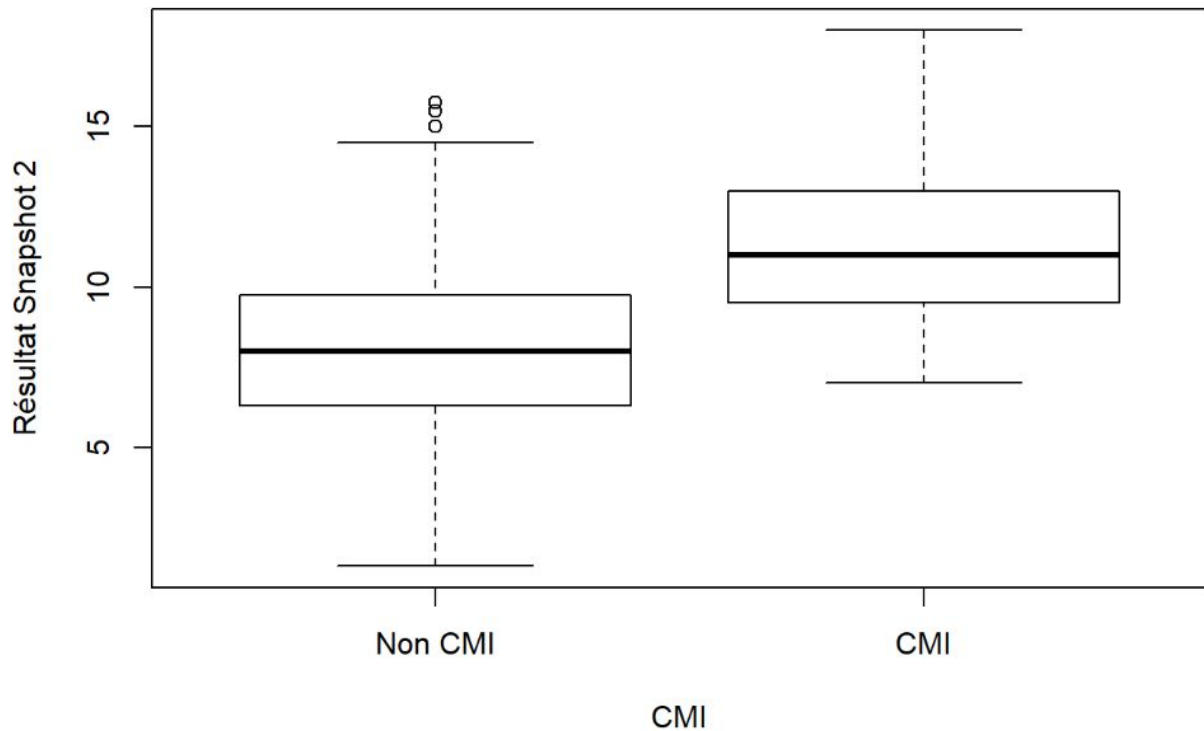
Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

Ce test nous renvoie une **p-value** égale à  $8.697538410 \times 10^{-6}$ . Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 1 sont identiques selon si les étudiants sont CMI ou non.

**Cela veut donc dire que le fait d'être CMI aurait un impact sur le résultat au Snapshot 1 des étudiants.**

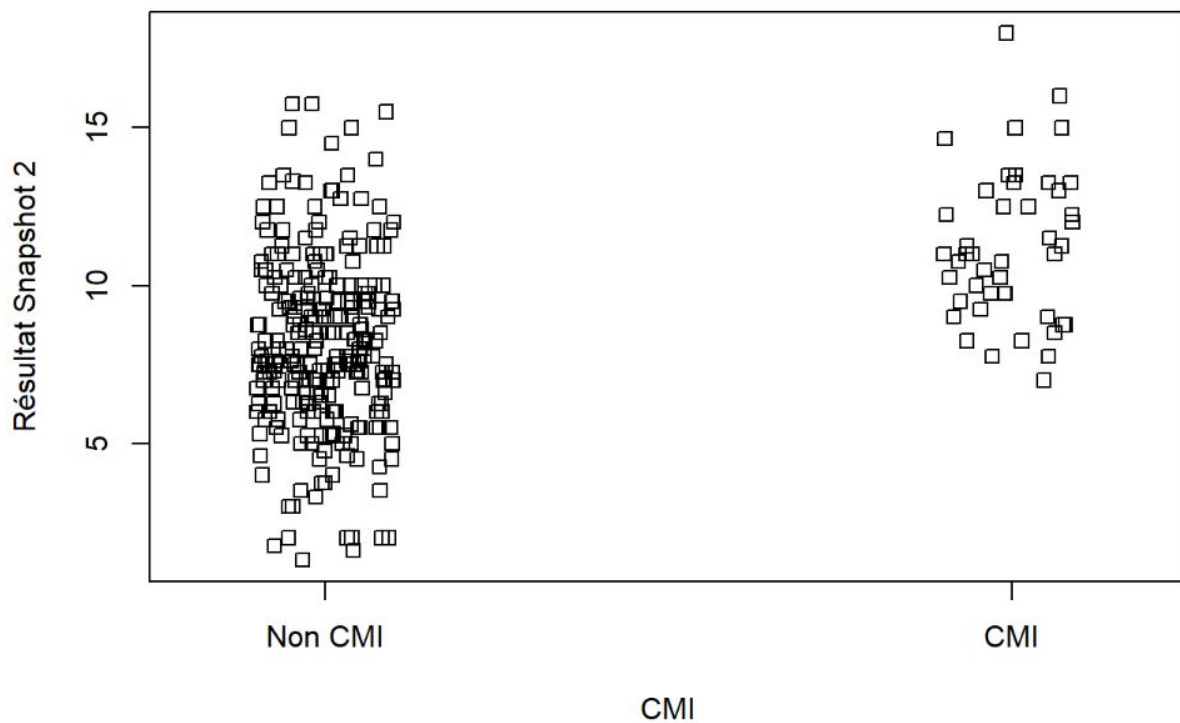
Représentons la répartition du résultat des étudiants au Snapshot 2 selon s'ils sont CMI ou non à l'aide d'une boîte à moustaches :

### Résultats au Snapshot 2 selon si les étudiants sont CMI ou non



Affichons ce graphique autrement en exposant tous les résultats :

### Résultats au Snapshot 2 selon si les étudiants sont CMI ou non





Nous souhaitons vérifier si les résultats des étudiants au Snapshot 2 sont identiques pour les étudiants CMI ou non. Pour cela, supposons donc que les résultats au Snapshot 2 des étudiants CMI sont identiques aux résultats au Snapshot 2 des étudiants non CMI.

Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

Ce test nous renvoie une **p-value** égale à  **$7.312239210^{-11}$** . Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 2 sont identiques selon si les étudiants sont CMI ou non.

**Cela veut donc dire que le fait d'être CMI aurait un impact sur le résultat au Snapshot 2 des étudiants.**

## Langue de TP

### Profil des étudiants

Ce tableau donne la moyenne au Snapshot 2 (S3) et au Snapshot 2 + 4M (S3) des étudiants selon leur langue de TP au S3 :

Résumé : Snapshots - Langue de TP au S3

LANGUE TP S3	NOMBRE D'ÉTUDIANTS	MOY SNAP 2 (S3)	MOY SNAP 2 + 4M (S3)
FR	155	8	8.9
GB	64	10.8	11.3
	23	15	10.2

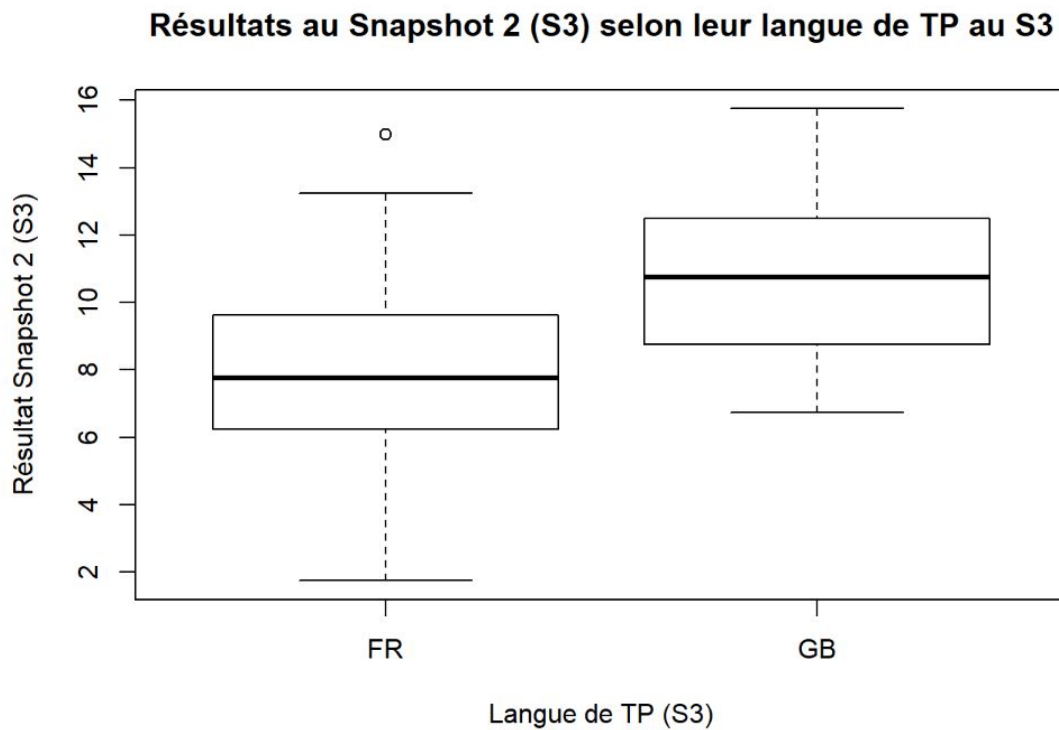
Ce tableau donne la moyenne au Snapshot 2 (S4) et au Snapshot 2 + 4M (S4) des étudiants selon leur langue de TP au S4 :

Résumé : Snapshots - Langue de TP au S4

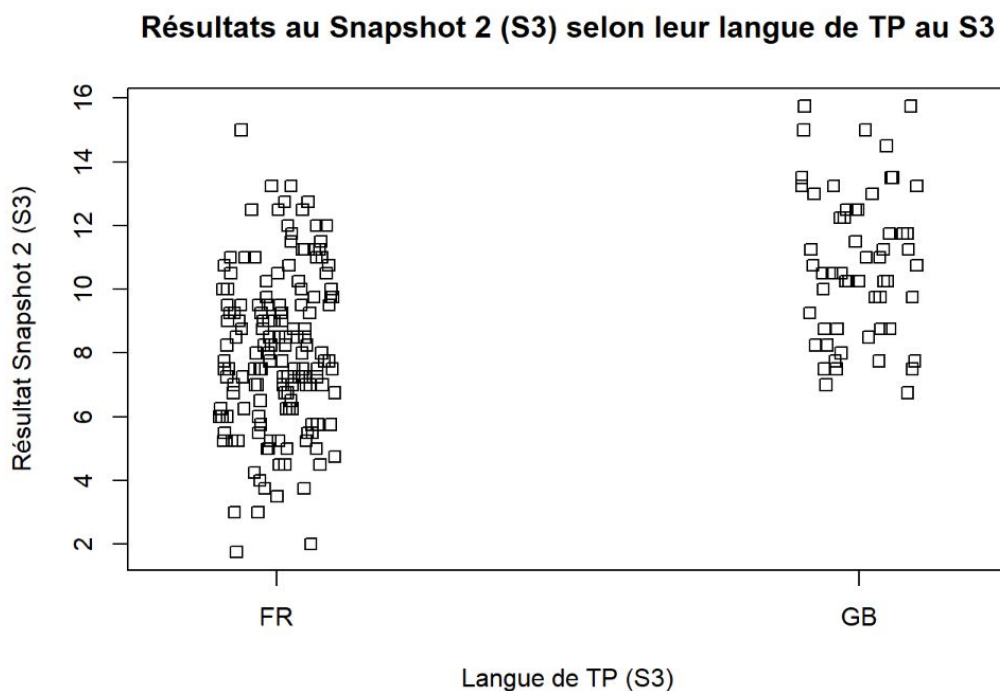
LANGUE TP S4	NOMBRE D'ÉTUDIANTS	MOY SNAP 2 (S4)	MOY SNAP 2 + 4M (S4)
FR	97	7.4	
GB	31	11	10.8
	68	16.9	8.5

## TP (S3) - Snapshot 2 (S3)

Représentons la répartition du résultat des étudiants au Snapshot 2 (S3) selon leur langue de TP au S3 à l'aide d'une boîte à moustaches :



Affichons ce graphique autrement en exposant tous les résultats :



Nous souhaitons vérifier si les résultats des étudiants au Snapshot 2 (S3) sont identiques pour les étudiants ayant effectué leur TP en français ou en anglais au S3. Pour cela, supposons donc que les résultats au Snapshot 2 (S3) des étudiants ayant effectué leur TP en français au S3 sont identiques aux résultats au Snapshot 2 (S3) des étudiants ayant effectué leur TP en anglais au S3.

Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

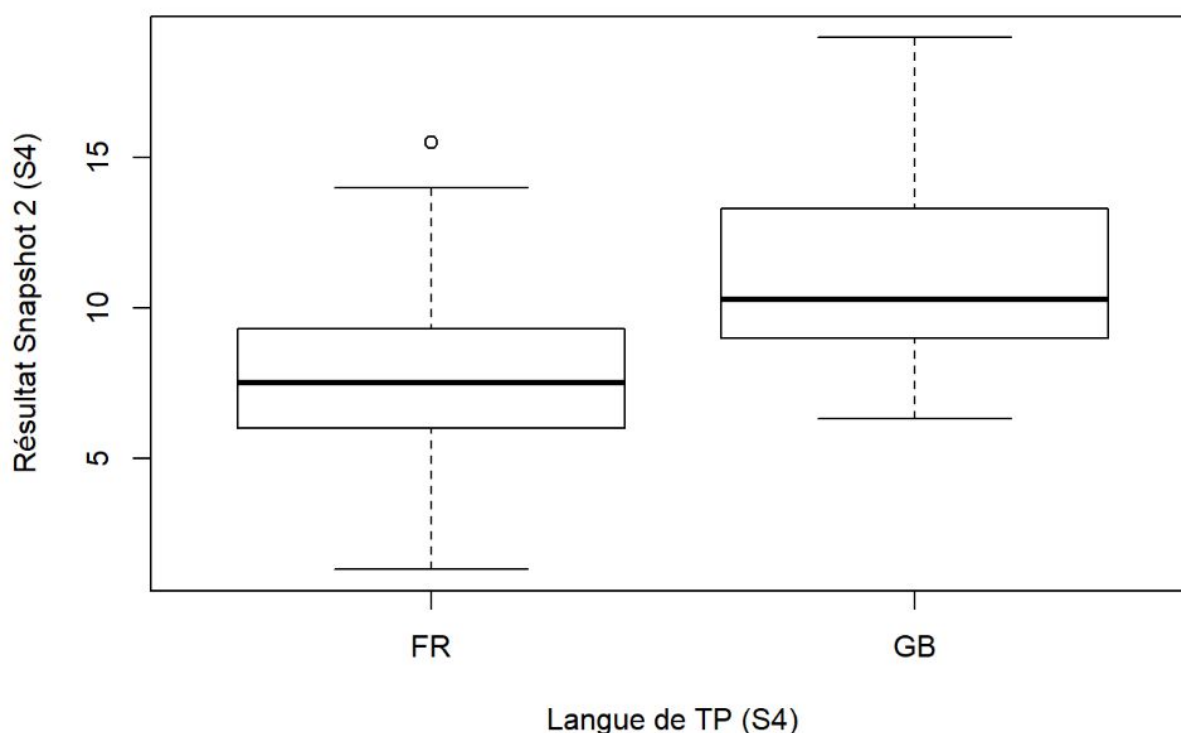
Ce test nous renvoie une **p-value** égale à  **$4.923490810^{-11}$** . Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 2 (S3) sont identiques peu importe la langue de TP au S3.

**Cela veut donc dire que la langue de TP au S3 aurait un impact sur le résultat au Snapshot 2 (S3) des étudiants.**

### TP (S4) - Snapshot 2 (S4)

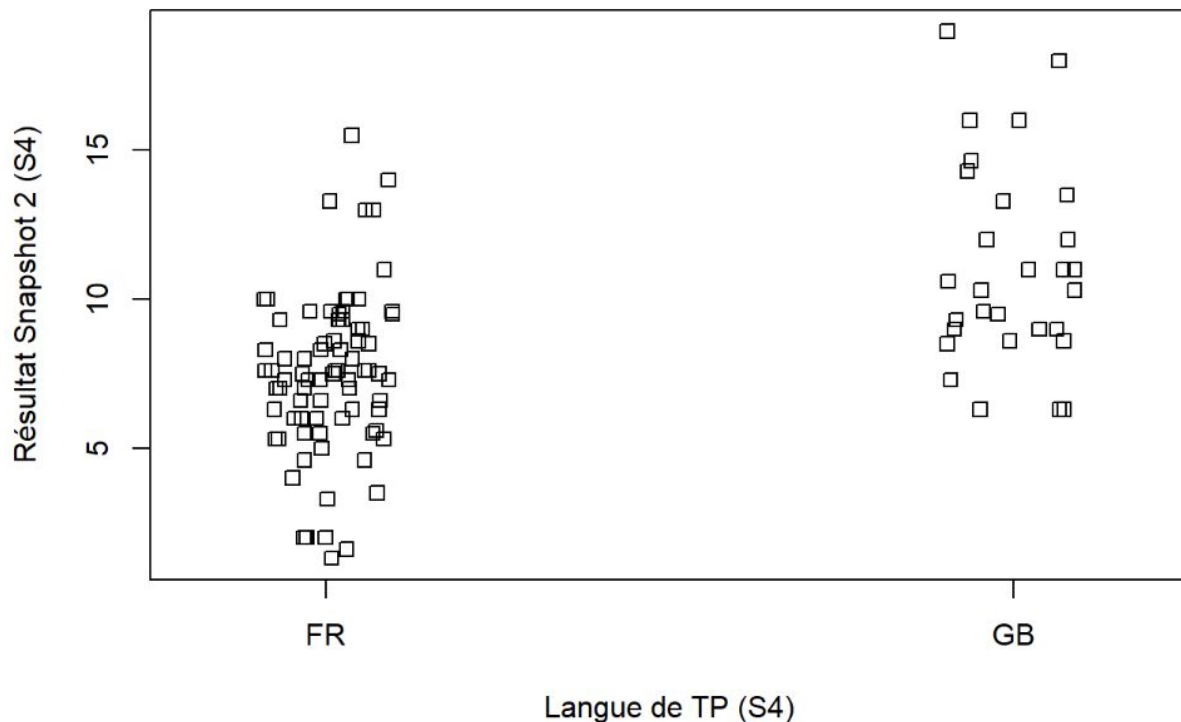
Représentons la répartition du résultat des étudiants au Snapshot 2 (S4) selon leur langue de TP au S4 à l'aide d'une boîte à moustaches :

**Résultats au Snapshot 2 (S4) selon leur langue de TP au S4**



Affichons ce graphique autrement en exposant tous les résultats :

### Résultats au Snapshot 2 (S4) selon leur langue de TP au S4



Nous souhaitons vérifier si les résultats des étudiants au Snapshot 2 (S4) sont identiques pour les étudiants ayant effectué leur TP en français ou en anglais au S4. Pour cela, supposons donc que les résultats au Snapshot 2 (S4) des étudiants ayant effectué leur TP en français au S4 sont identiques aux résultats au Snapshot 2 (S4) des étudiants ayant effectué leur TP en anglais au S4.

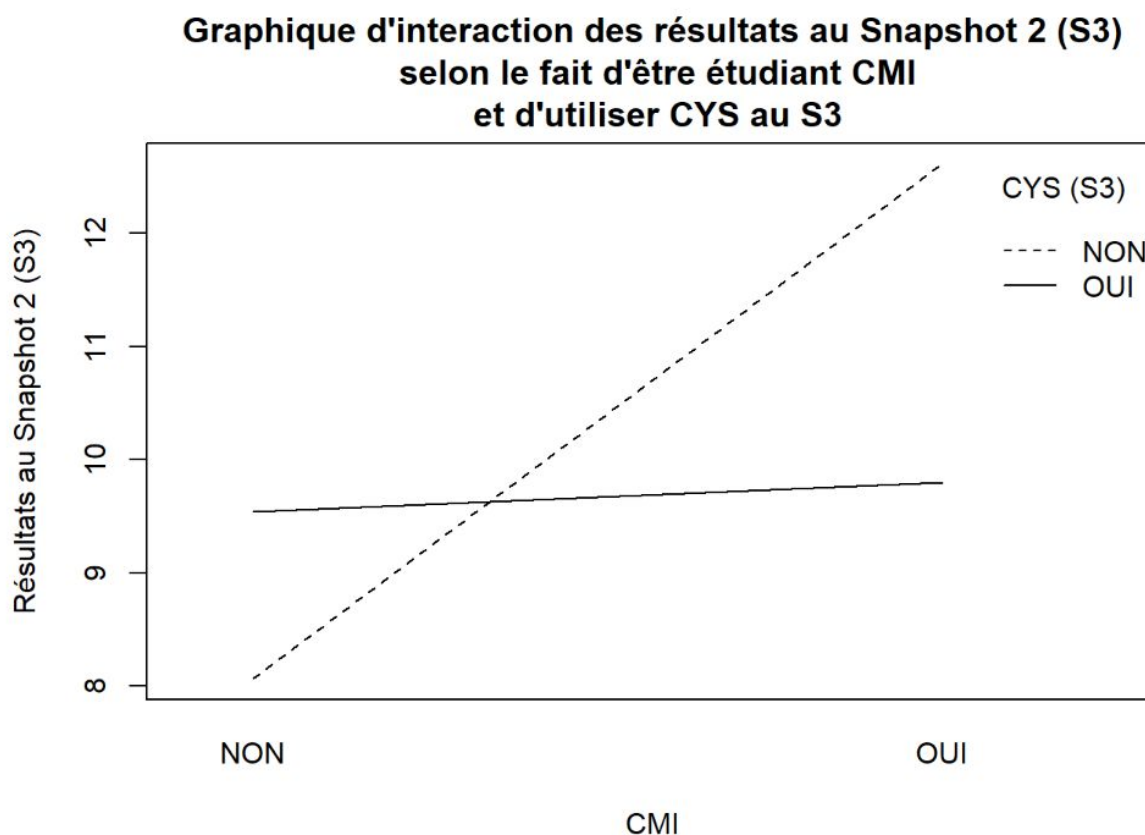
Pour vérifier cette hypothèse effectuons un **test de Wilcoxon**. Fixons le seuil de tolérance à 5%.

Ce test nous renvoie une **p-value** égale à  $1.125610410 \times 10^{-6}$ . Nous trouvons une p-value inférieure à 0.05, nous rejetons donc l'hypothèse que les résultats au Snapshot 2 (S4) sont identiques peu importe la langue de TP au S4.

**Cela veut donc dire que la langue de TP au S4 aurait un impact sur le résultat au Snapshot 2 (S4) des étudiants.**

## CYS(S3) & CMI

Un graphique d'interaction permet d'observer comment varie une variable numérique en fonction d'un premier facteur selon les modalités d'un deuxième facteur.



Le graphique d'interaction ci-dessous nous montre la variation des résultats au Snapshot 2 (S3) selon le fait d'être étudiant CMI et d'utiliser CYS au S3.

La ligne en pointillés représente les résultats au Snapshot 2 (S3) des étudiants n'ayant pas utilisé CYS au S3. Nous retrouvons à l'extrémité gauche de cette ligne la moyenne au Snapshot 2 (S3) des étudiants non CMI n'ayant pas utilisé CYS au S3. À l'extrémité droite, nous retrouvons la moyenne au Snapshot 2 (S3) des étudiants CMI n'ayant pas utilisé CYS au S3.

La ligne noire continue représente les résultats au Snapshot 2 (S3) des étudiants ayant utilisé CYS au S3. Nous retrouvons à l'extrémité gauche de cette ligne la moyenne au Snapshot 2 (S3) des étudiants non CMI ayant utilisé CYS au S3. À l'extrémité droite, nous retrouvons la moyenne au Snapshot 2 (S3) des étudiants non CMI ayant utilisé CYS au S3.

---

**Nous remarquons que le fait d'être étudiant CMI ou non influe peu sur le résultat au Snapshot 2 (S3) si l'étudiant utilise CYS au S3.**

Cependant, nous remarquons une différence significative entre la moyenne au Snapshot 2 (S3) des étudiants non CMI n'ayant pas utilisé CYS au S3 et les étudiants CMI n'ayant pas utilisé CYS au S3, avec une bien meilleure moyenne pour ces derniers.

**Les lignes ne sont pas parallèles, donc le fait d'utiliser CYS au S3 influe bien sur le résultat au Snapshot 2 selon si les étudiants sont CMI ou non.**

Les étudiants non CMI qui n'ont pas utilisé CYS au S3 ont de moins bons résultats que les étudiants non CMI qui ont utilisé CYS au S3.

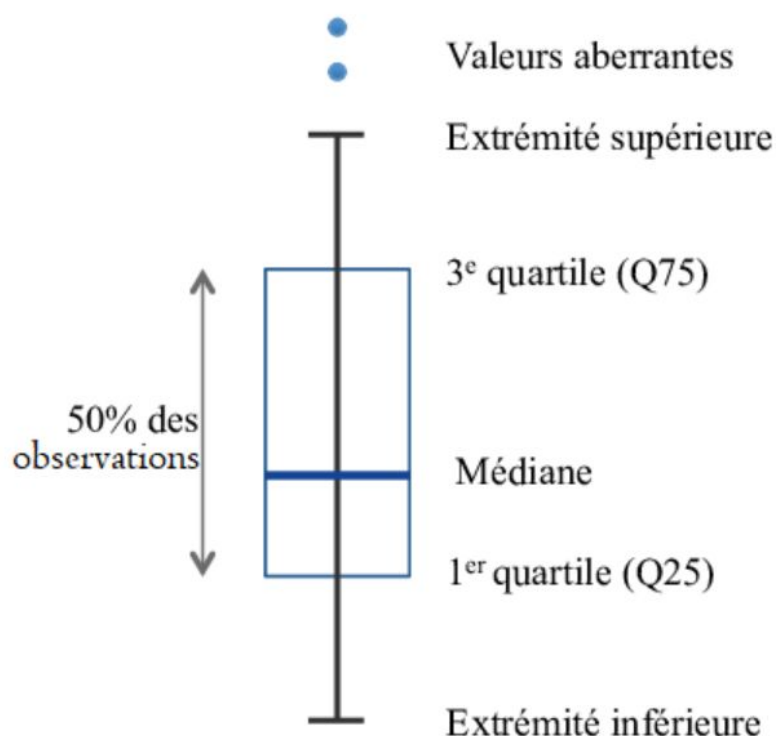
À l'inverse, les étudiants CMI n'ayant pas utilisé CYS au S3 ont de meilleurs résultats que les étudiants CMI ayant utilisé CYS au S3.

**Note :** Tous les étudiants CMI ont utilisé CYS au S4. Nous ne pouvons pas donc faire un graphique d'interaction comme celui-ci dessous pour les résultats au Snapshot 2 (S4) en fonction du fait d'être étudiant CMI et d'utiliser CYS au S4.

## ANNEXES

### Boîte à moustaches

Pour étudier plus en détail la répartition des résultats des étudiants, nous utilisons un graphique appelé **boîte à moustaches** ou **boxplot**.



Une boîte à moustaches est, en statistiques, une représentation graphique d'une variable quantitative continue, c'est-à-dire une variable dont les réalisations sont des nombres reflétant une idée de grandeur.

Ce graphique illustre la variable à l'aide de plusieurs repères :

- **Médiane** : cette valeur sépare nos observations en deux moitiés. Au-dessus de cette valeur, nous retrouvons 50% des observations. En dessous de cette valeur, nous retrouvons les observations restantes.
- **1<sup>er</sup> quartile (Q1)** : 25% des observations sont en dessous de cette valeur.
- **3<sup>e</sup> quartile (Q3)** : 25% des observations sont au-dessus de cette valeur.



L'**écart-interquartile** est la différence entre le 3e quartile et le 1er quartile. Il correspond à la longueur de la boîte sur l'image, qui contient alors 50% des observations.

La largeur de la boîte quant à elle n'a aucune signification, son utilité est purement visuelle.

L'**extrémité supérieure** est la valeur de la plus grande observation inférieure ou égale à  $Q3 + 1.5(Q3 - Q1)$ , tandis que l'extrémité inférieure est la valeur de la plus petite observation supérieure ou égale à  $Q1 - 1.5(Q3 - Q1)$ .

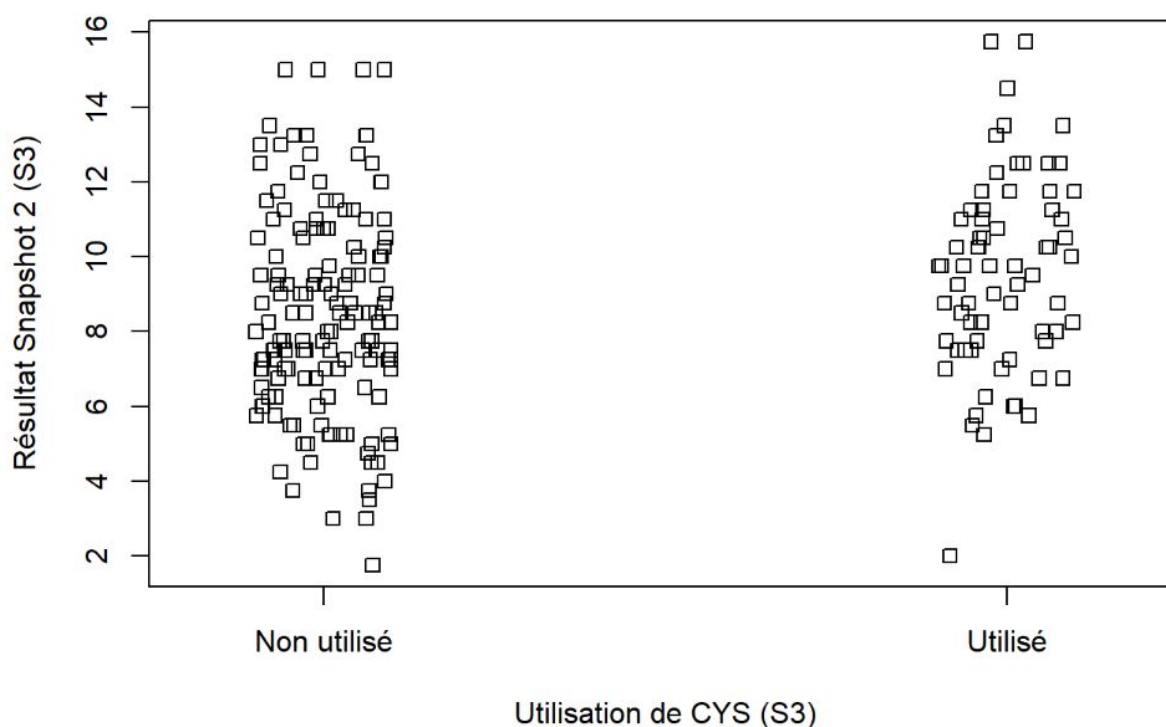
Une représentation comme celle-ci n'est pas pertinente pour un jeu de données très faible, soit avec moins de 10 observations.

## Diagramme à bandes

Un **diagramme à bandes**, ou **stripchart**, est un graphique représentant toutes les observations d'une variable en ordonnée par des marqueurs. Ces observations sont réparties dans différentes bandes selon leur modalité d'une variable en abscisse.

Par exemple, le graphique ci dessus représente les résultats des étudiants au Snapshot 2 (S3) selon leur utilisation de CYS au S3 :

**Résultats au Snapshot 2 (S3) selon l'utilisation de CYS au S3**



Chaque carré représente le résultat d'un étudiant.

Nous retrouvons deux bandes, l'une correspondant aux étudiants n'ayant pas utilisé CYS au S3, l'autre correspondant aux étudiants ayant utilisé CYS au S3.

## Test de Wilcoxon

Le **test de Wilcoxon** est un test non-paramétrique de comparaison des moyennes de deux échantillons ayant un nombre faible de données. Il sert à tester l'hypothèse selon laquelle la distribution des données est la même pour les deux échantillons. Ce test se base sur le rang des valeurs observées.

On pose alors l'hypothèse suivante : **les deux échantillons sont identiques.**

Ici, nous effectuons ce test pour vérifier si une variable (utilisation de CYS, étudiants CMI ...) a une influence sur le résultat des étudiants.

En effectuant ce test sous R, cela nous renvoie une valeur appelée **p-value**. C'est la probabilité de rejeter cette hypothèse alors qu'elle est vraie.

**Si la p-value est inférieure à un seuil de tolérance A**, alors nous **rejetons l'hypothèse**, c'est-à-dire que les deux échantillons seraient différents et donc que le variable aurait une incidence sur le résultat des étudiants.

Si nous trouvons une **p-value supérieure à ce même seuil A**, alors nous **conservons l'hypothèse**.

La valeur A est un seuil de tolérance choisi par le statisticien. **Celui-ci est généralement égal à 5% (0.05).**