

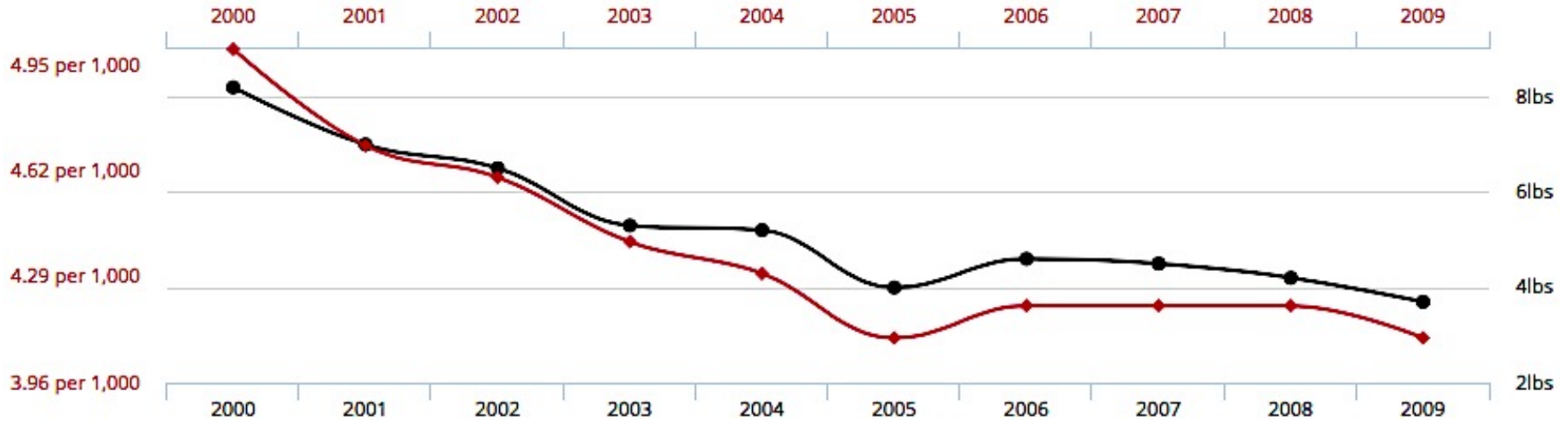
Remédiation

Analyser des documents scientifiques
en biologie

Les corrélations

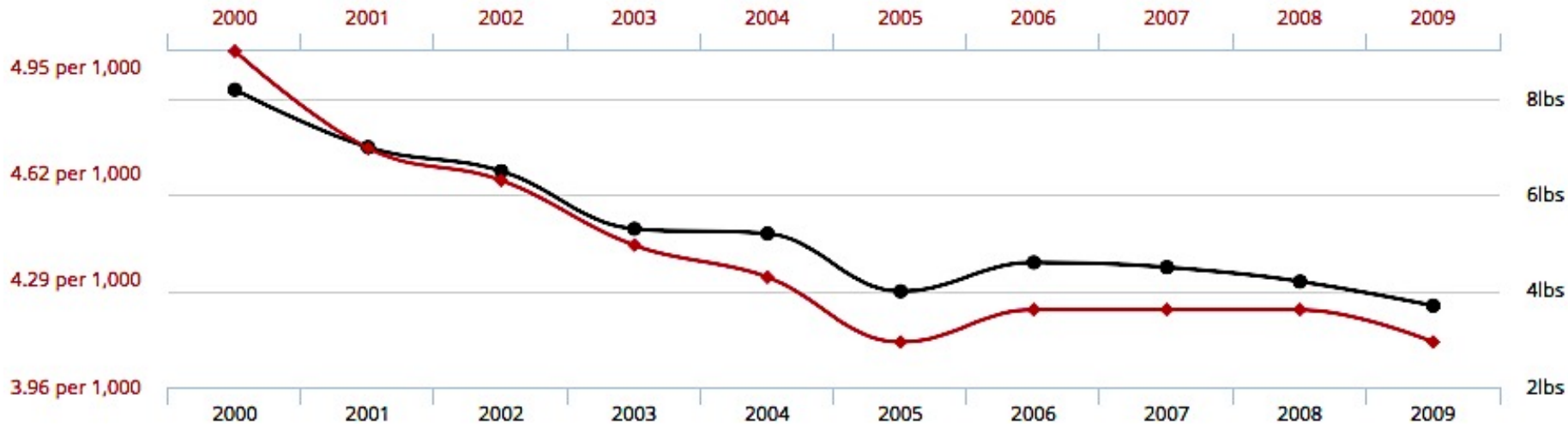
- Attention aux courbes qui se ressemblent...

Correlation: 99.26% ($r=0.992558$)



Les deux courbes suivent la même évolution : leur décroissance est parallèle et le coefficient de corrélation est élevé.

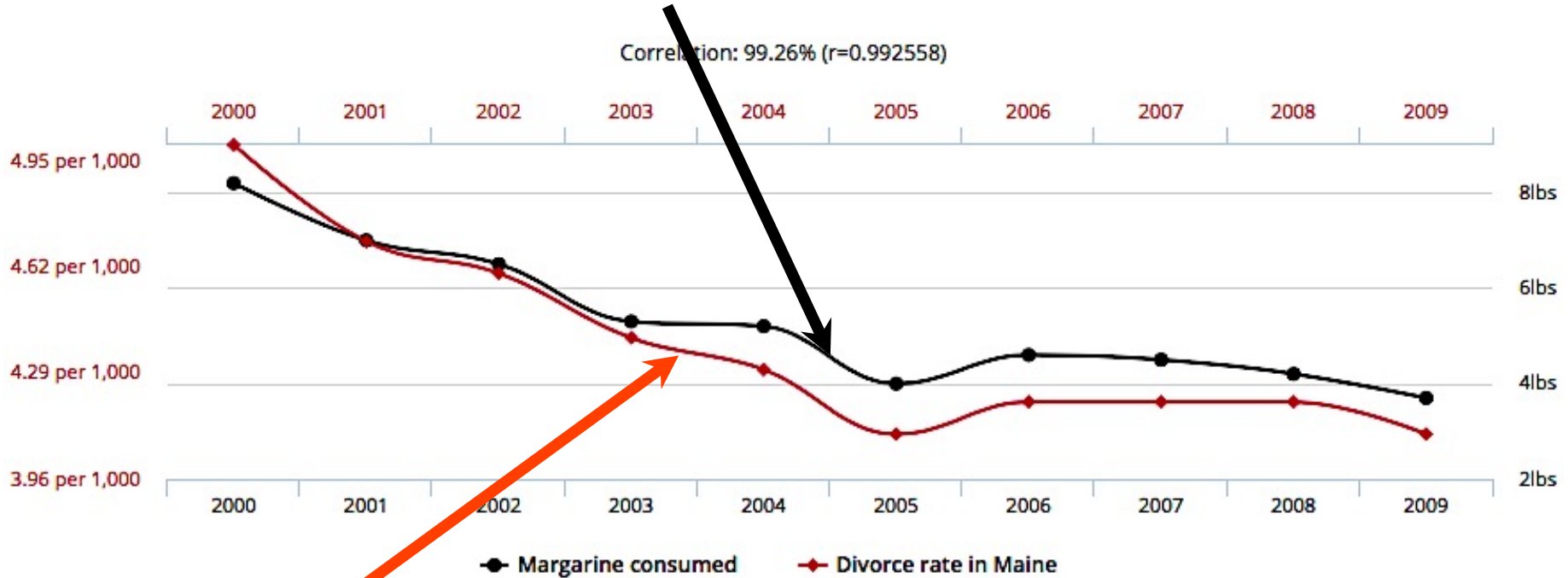
Correlation: 99.26% (r=0.992558)



Il est tentant d'écrire :

Les deux paramètres sont liés et leur évolution est due à un même phénomène ou bien l'un des paramètres a une incidence sur l'autre.

Consommation de margarine aux Etats-Unis



Taux de divorce dans l'état du Maine (USA)
C'est l'effet cigogne

**Une telle corrélation a été une base de travail de recherche (1981) pour établir un lien entre la consommation de café et le cancer du pancréas...
L'hypothèse s'est avérée fausse.**

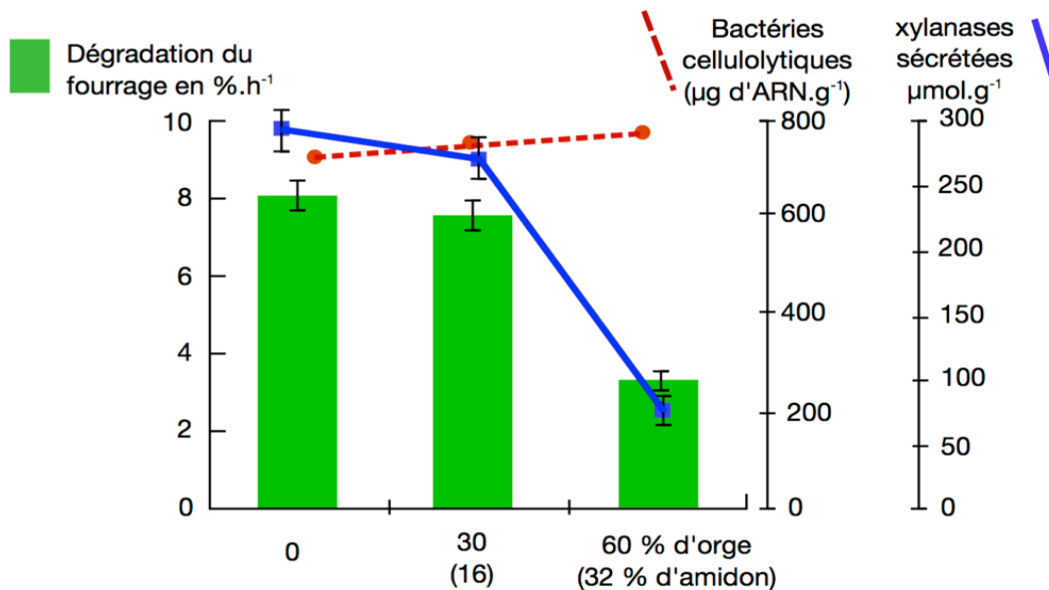
Une corrélation permet de faire une hypothèse qui est ensuite testée grâce à des expériences. Elle ne permet pas de conclure avec certitude.

Chiffrer une analyse

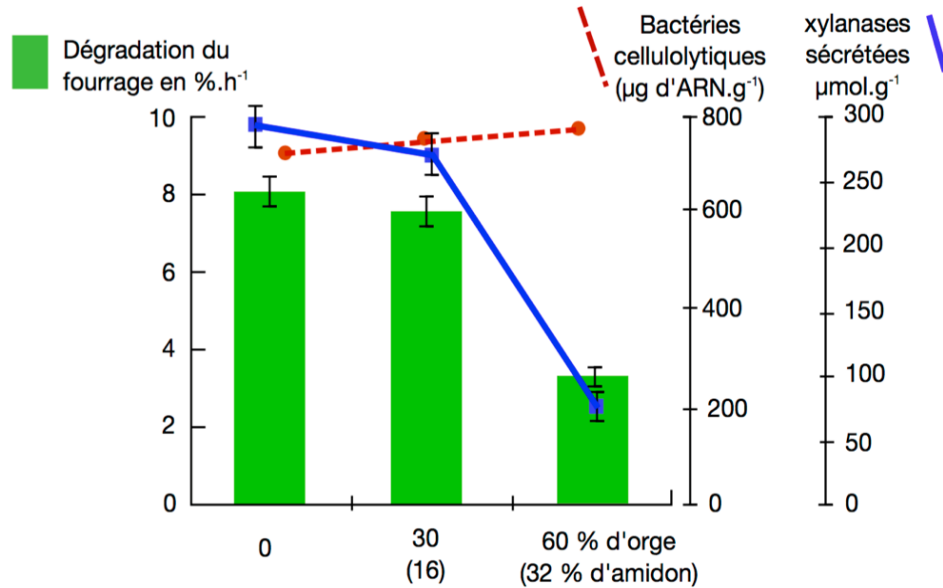
Décrire l'évolution d'un paramètre nécessite rigueur et réflexion.

- Chiffrer les analyses
- Utiliser des facteurs multiplicateurs ou des pourcentages.

La différence est significative entre 30 et 60% d'orge.
Cette différence doit être chiffrée.



Ne pas écrire : la sécrétion de xylanases a baissé de 200 μmol·g⁻¹. Cela ne représente pas la même chose selon que la valeur de départ était de 280 ou de 250 000 !



Écrire :

la sécrétion de xylanases

– a été divisée par 3,5

ou

– ne représente plus que 35% de la valeur initiale

Comparer des valeurs

Les différences statistiques

Données statistiques traitées de diverses manières :

- Indication des écarts-types
- Indication des écarts à la moyenne
- Box-plot

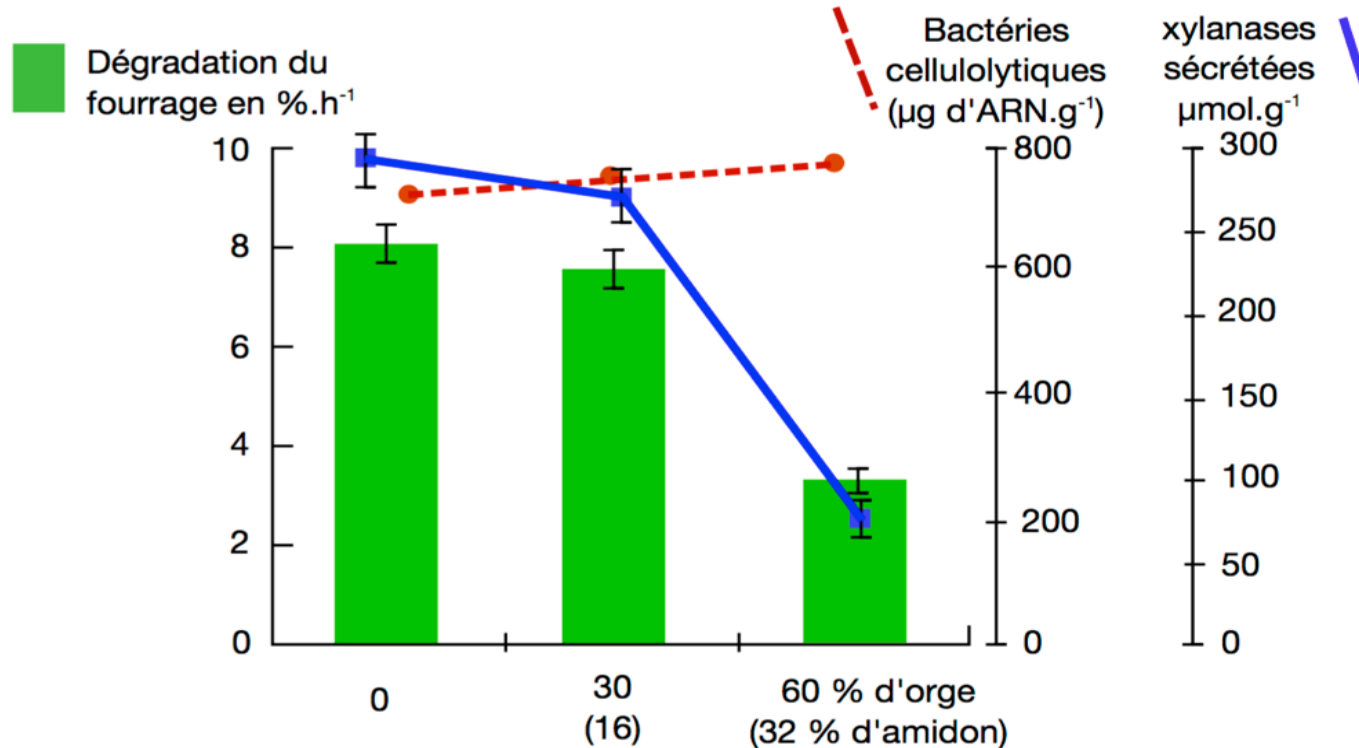
Attention à l'importance de l'échantillon.

Les écart-types et écarts à la moyenne

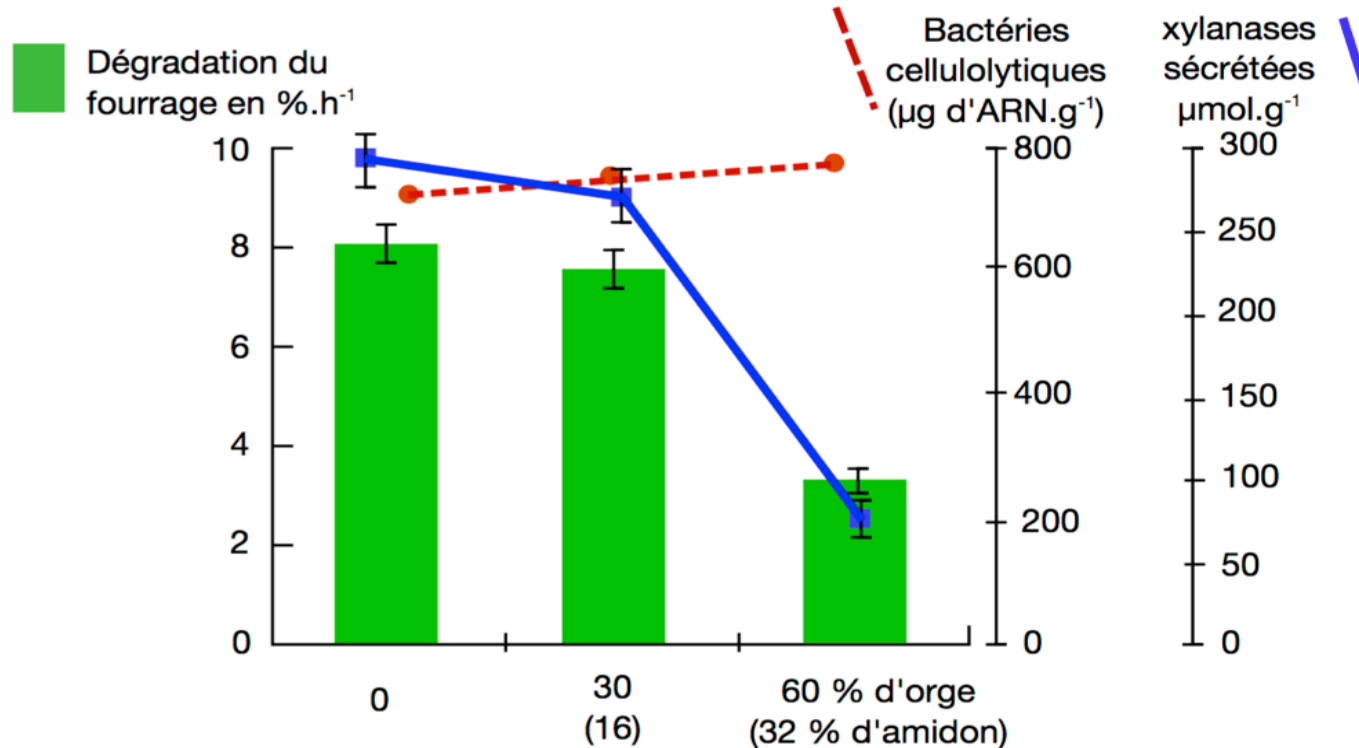
$$\text{Écart à la moyenne} = \frac{\text{Écart-type}}{\sqrt{\text{nombre d'échantillon}}}$$

Les barres ne doivent pas se chevaucher.

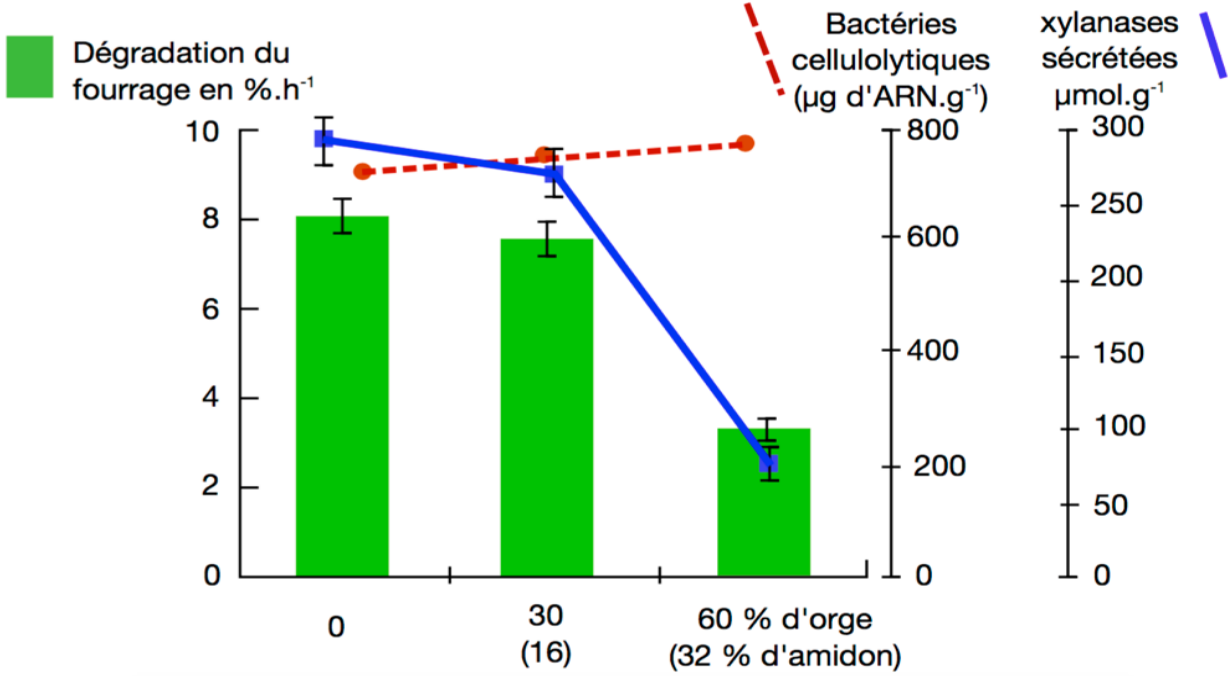
Les valeurs sont significativement différentes si et seulement si leurs barres ne se chevauchent pas.



=> pas de différence significative entre 0 et 30% d'orge pour la dégradation du fourrage et la sécrétion des xylanases



Ne pas écrire : «on ne peut pas conclure entre 0 et 30% d'orge» mais écrire : « il n'y a pas de différence significative entre 0 et 30% d'orge donc la proportion d'orge, jusqu'à 30%, n'a pas d'incidence sur la dégradation du fourrage ».

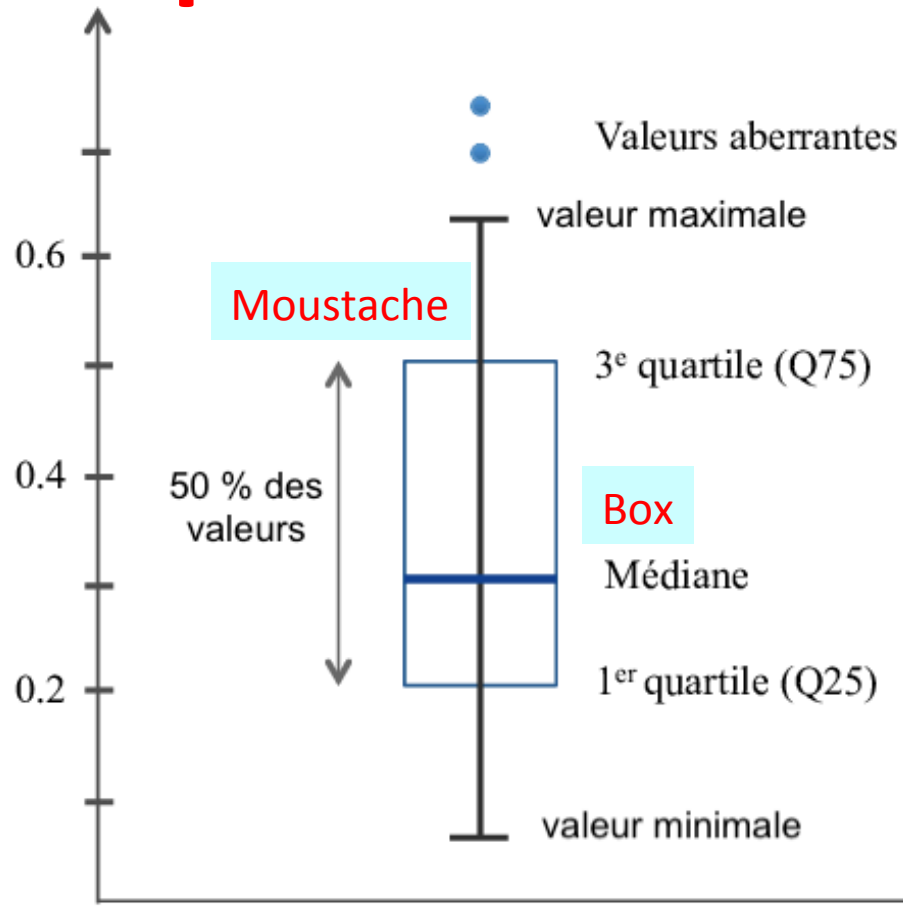


Les box-plots

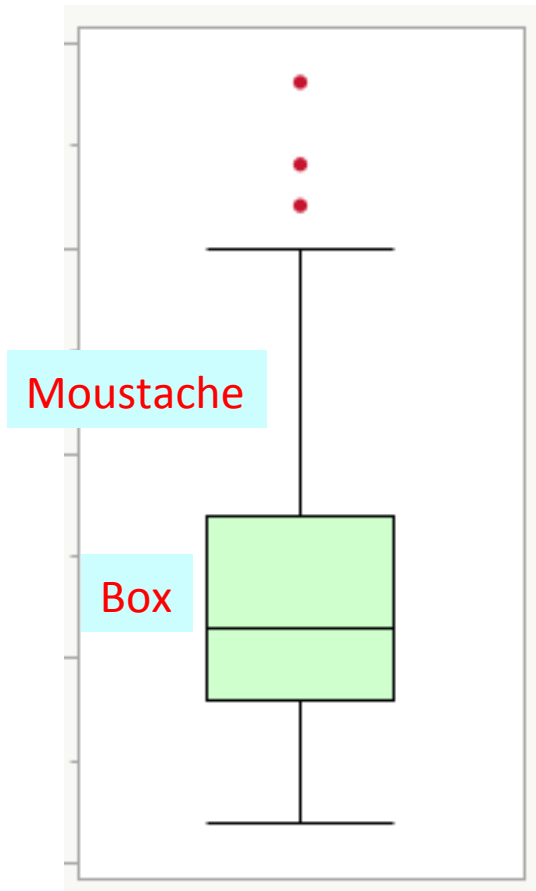
... ou boîtes à moustaches...

... ou diagrammes en boîtes...

Une représentation statistique

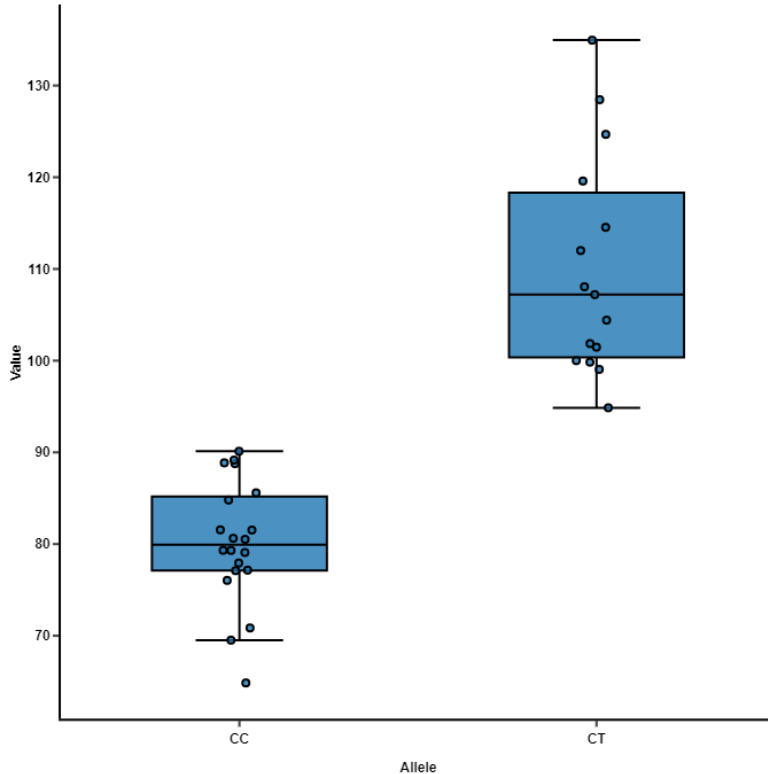


Une représentation statistique



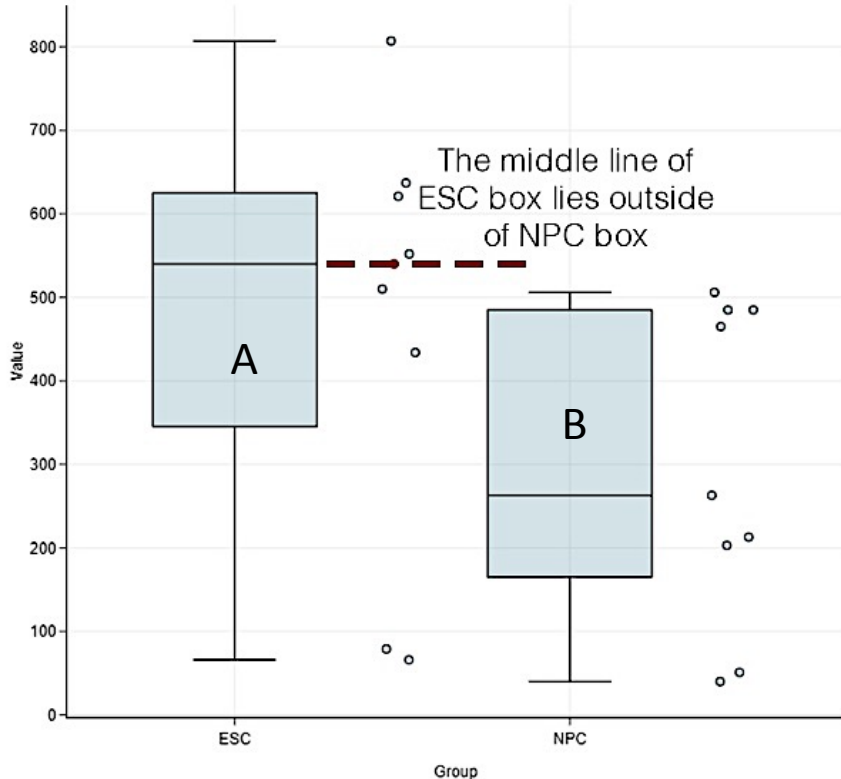
Valeurs aberrantes car éloignées de la médiane d'une distance de plus de 1,5 fois de la taille de la box

Comparaison des box-plots



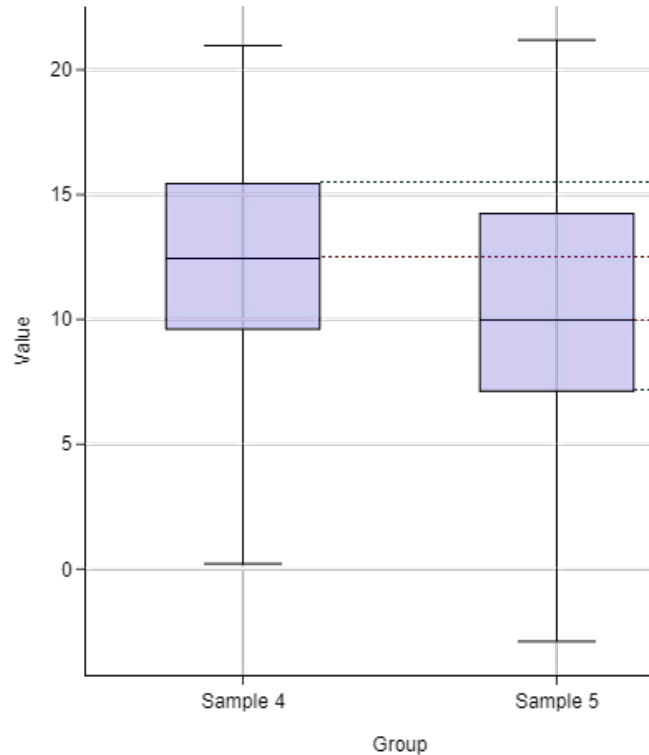
Si 2 box ne se recouvrent pas l'une l'autre, par exemple si la box A est complètement au-dessus ou au-dessous de la box B, alors il y a une différence entre les deux groupes.

Comparaison des box-plots



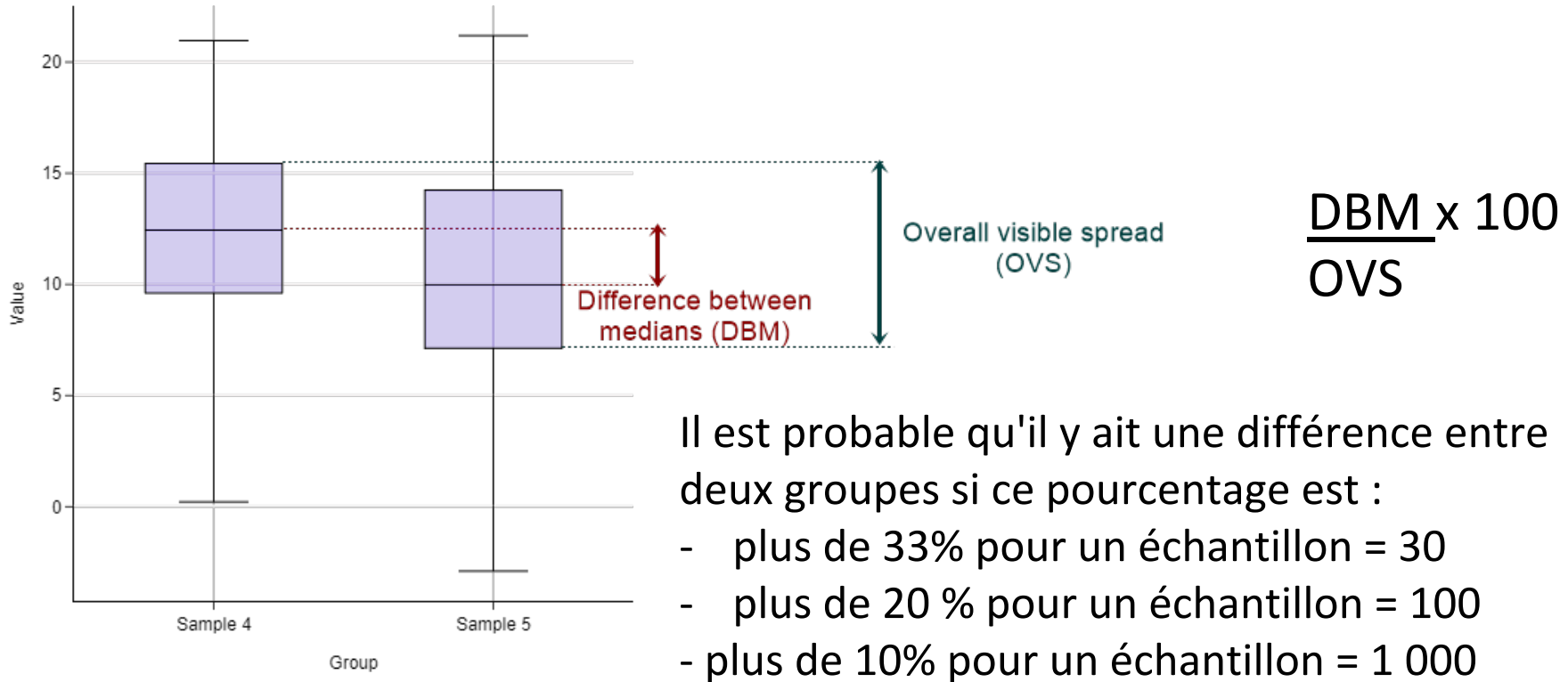
Si la ligne médiane de la box A se trouve entièrement à l'extérieur de la boîte B, il est probable qu'il y ait une différence entre les 2 groupes.

Comparaison des box-plots



Si les deux lignes médianes se trouvent à l'intérieur du chevauchement entre deux box, nous devons franchir une autre étape pour parvenir à une conclusion sur leurs groupes.

Comparaison des box-plots



Pas à faire en DS

En pratique : les tests statistiques

Différents tests statistiques permettent de déceler si une différence est significative ou non.

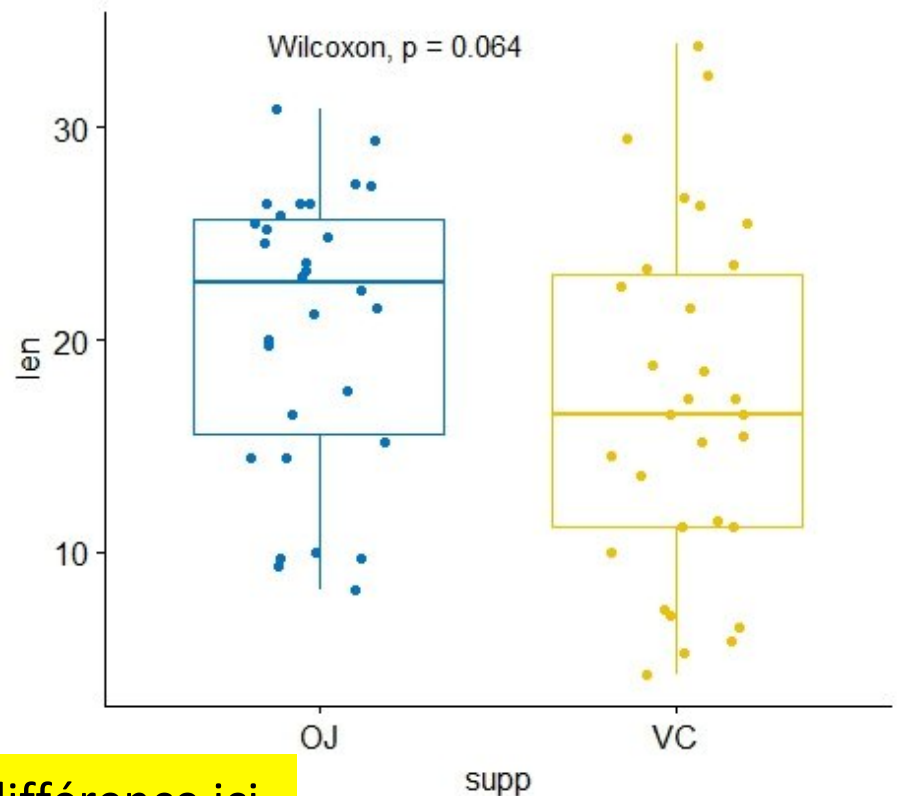
- test t de Student
- test de Wilcoxon
- test de Mann Whitney

On en déduit des valeurs de p.

En pratique : les tests statistiques

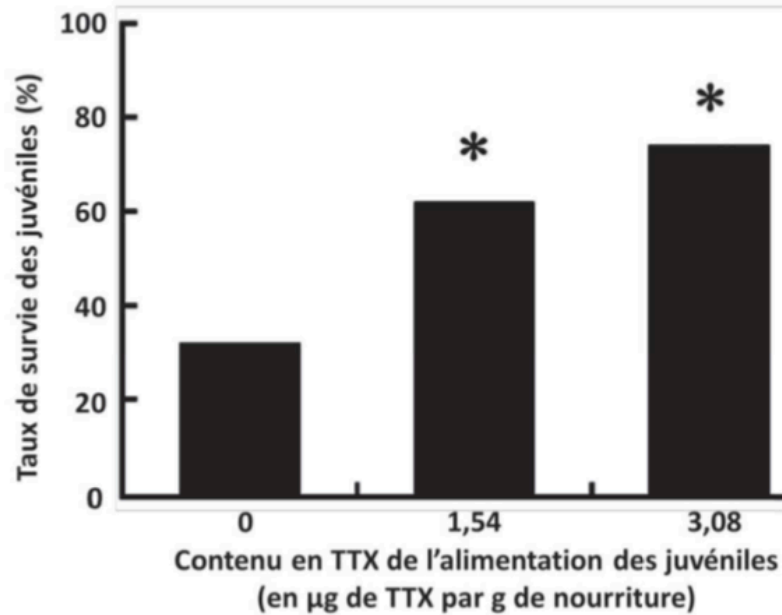
La différence est significative si $p < 0,05$.

Cela signifie qu'il y a moins de 5 % de chances que la différence obtenue soit le fait du hasard.

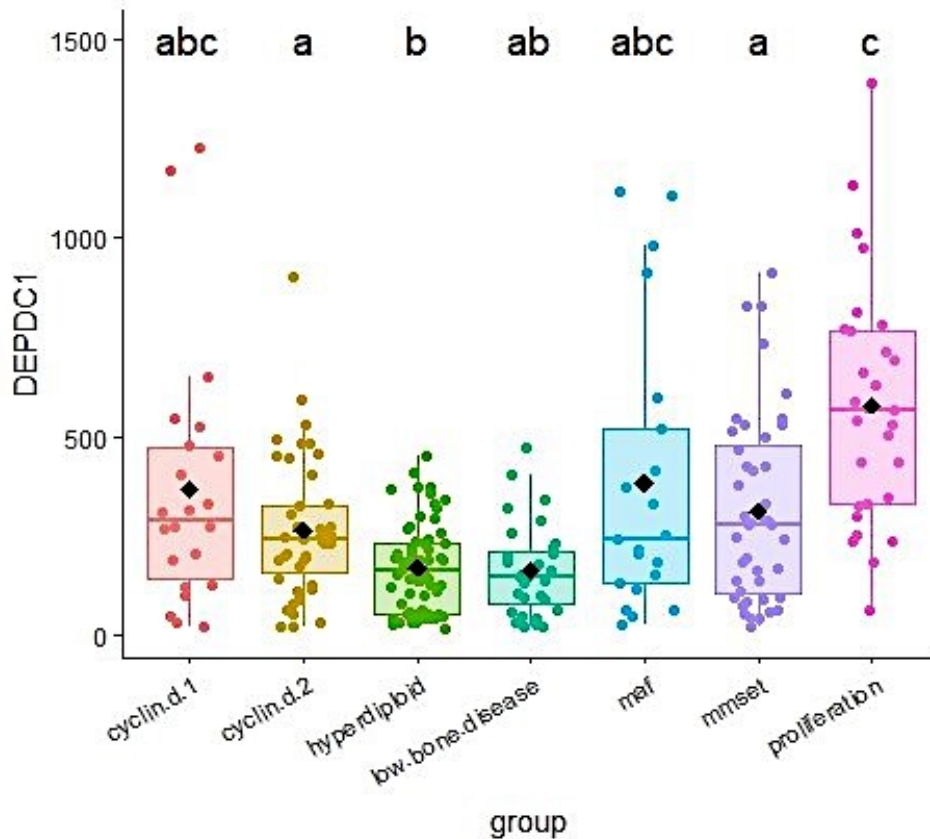


Pas de différence ici.

En pratique : les astérisques



En pratique : les lettres

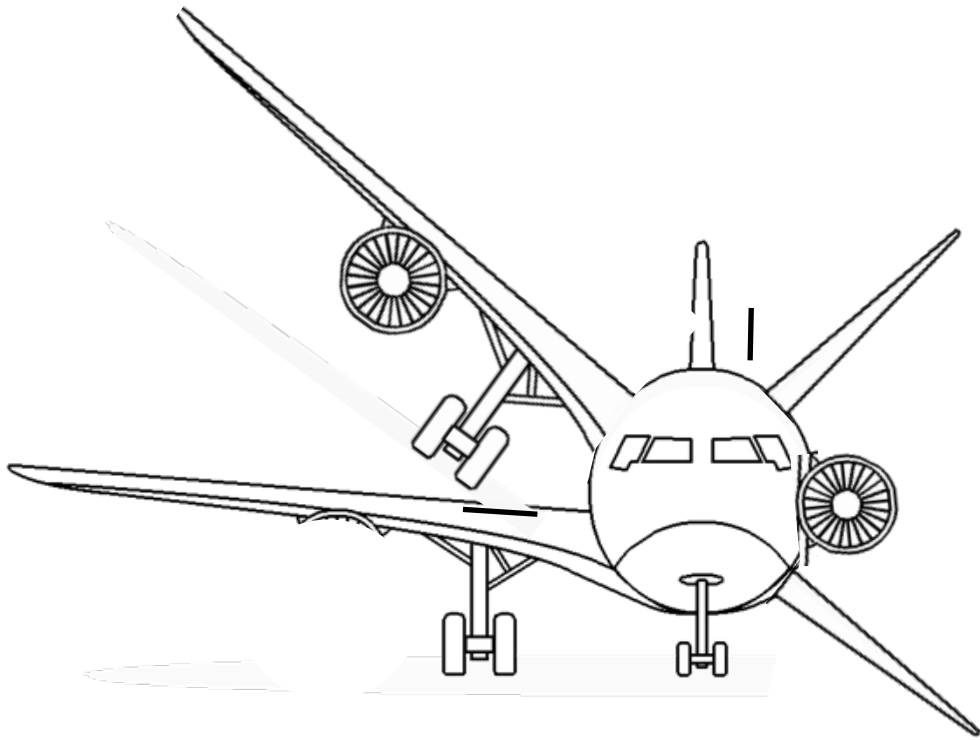


La convention veut que si deux moyennes partagent une même lettre alors elles ne sont pas significativement différentes, et au contraire, si deux moyennes ne partagent pas une même lettre alors elles sont significativement différentes

La rigueur

comment une imprécision peut conduire à une erreur

**Mais si, il est bien !
il a deux ailes, deux ailerons, deux réacteurs...**



impossible ? oui
mais...

