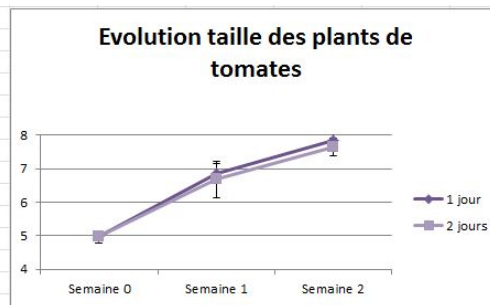
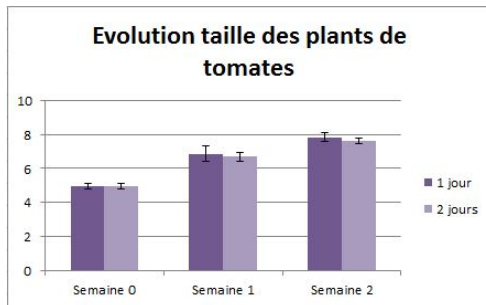


# Représenter ses résultats de TIPE

Selon votre sujet et vos expériences, vous pouvez avoir différents types de résultats. Si tout s'est bien passé, il se peut que vous ayez beaucoup de valeurs...et bien sur on ne va pas mettre dans son rapport un énorme tableau de données brutes....on ne peut pas non plus mettre tout sur un graphique... alors comment faire???

**Exemple:** Voici un tableau de données brutes et deux façons de les représenter. Le contexte: on s'interroge sur le lien entre l'arrosage et la croissance de plants de tomates. On considère 2 jardinières contenant chacune 5 plants de tomates. On arrose la première jardinière avec 20cL d'eau tous les jours et la deuxième jardinière tous les 2 jours avec 20cL aussi. On mesure la taille des plants de tomate tous les mardis, la semaine 0 correspondant à la taille initiale des plants avant arrosage.

	A	B	C	D	E
1	<b>Evolution de la taille en cm de 5 plants de tomate selon arrosage</b>				
2					
3	<b>Données :</b>				
4		Semaine 0	Semaine 1	Semaine 2	
5	20 cl /jour	4,8	7,2	8	
6		5	6,8	8,1	
7		4,7	6,9	8	
8		5,1	7,5	7,5	
9		5,2	5,9	7,6	
10	20 cl /2 jours	4,8	6,9	7,5	
11		5	6,2	7,6	
12		4,7	6,7	8	
13		5,1	7	7,5	
14		5,2	6,6	7,6	



? Que voit-on sur ces figures?

À gauche: une hauteur par semaine par condition et un intervalle l'encadrant

À droite: Un point par semaine par condition et un intervalle l'encadrant

► le point/la hauteur: correspond à la **moyenne** de la hauteur des 5 plants de la jardinière

► l'intervalle: encore appelé **barre d'erreur** : ici il représente un intervalle de confiance à 95 % associé à nos mesures, c'est-à-dire qu'on estime qu'il y a 95 % de chance que la 'vraie' hauteur moyenne soit dans cet intervalle (pour n'importe quel plant de même type, dans les mêmes conditions).

## I. Différents types de graphiques possibles

Deux principaux types de représentation étudiés ici sont les histogrammes et les courbes de tendance, comme présentés dans l'exemple précédent. Selon les données dont on dispose, on privilégiera l'un ou l'autre. On peut également faire un nuage de points si on cherche des relations entre les variables (régression linéaire ou autre). Les autres types de représentation présentent des avantages mais sont rarement appropriés à une étude mathématique.

Quelques exemples:

- Évolution temporelle avec plusieurs mesures par condition (ex: hauteurs de plants, on a plusieurs plants par jardinière) ~> les deux sont possibles
- Évolution temporelle avec une seule mesure par condition (ex: taux de survie de crevettes, un seul aquarium par condition) ~> les deux sont possibles
- Comparaison de différents échantillons, indépendamment du temps, dans différentes conditions avec une seule mesure par échantillon (ex: nombre de graines germées au bout d'une semaine selon l'arrosage) ~> plutôt des histogrammes

**Remarque :** Plus les résultats sont nombreux, plus on réussira à exploiter les graphiques dans la partie interprétation.

**⚠ Règle fondamentale :** Bien noter tous les résultats bruts relevés dans un tableur (et mettre des titres de lignes/colonnes pour savoir à quoi ils correspondent, avec les unités, la date et toute information capitale pour savoir à quoi ils correspondent...) et sauvegarder le fichier sur différents supports: ordinateur personnel (de plusieurs membres du groupe), clé usb, stockage en ligne,...

**⚠ Règle fondamentale :** ne jamais modifier les données brutes relevées: les garder et exploiter les résultats à part, dans une autre feuille de calcul.

## II. Barres d'erreur sur un graphique

### 1) Barres d'erreur?

En général, pour interpréter les résultats obtenus sur une série de données, un graphique d'évolution seul ne suffit pas: on a besoin de faire apparaître une marge d'erreur pour les valeurs obtenues, ce qu'on appelle couramment barre d'erreur, ce qu'on appelle plus mathématiquement intervalle de confiance.

Vous pouvez avoir plusieurs types de résultats possibles :

- Si vous avez répété plusieurs fois la même expérience, ou bien si vous avez fait plusieurs mesures, alors vous avez plusieurs données et pouvez faire une moyenne. Or la

moyenne seule ne suffit pas à dire quelque chose sur une série statistique, il faut également utiliser l'écart type pour étudier la dispersion des données autour de la moyenne. Plus précisément pour pouvoir "dire" quelque chose sur une population de manière plus générale, on utilise des intervalles de confiance. ("dire" avec des pincettes car cela reste des statistiques...)

- Si vous avez une seule mesure par condition, et que cette mesure représente une proportion (taux de survie, ...) on peut faire un intervalle de confiance de proportion
- Si vous avez réalisé une seule fois l'expérience et avez une mesure, vous pouvez comme barre d'erreur mettre les incertitudes de mesure si vous les connaissez. La méthode d'ajout sera la même.

## 2) Différents types de barres d'erreur

Il existe plusieurs intervalles différents qu'on peut mettre sur un graphique. En voici quelques exemples, les détails de ces intervalles sont donnés au paragraphe suivant

### ► Intervalles "résumant" les données l'échantillon:

- Intervalle du min/max  
↪ On sait que toutes les valeurs de l'échantillon sont dans cet intervalle
- Intervalle moyenne/écart-type seul  
↪ Cet intervalle résume l'échantillon en étant centré sur la moyenne des valeurs, et de demi-largeur l'écart-type, qui indique la dispersion des valeurs autour de la moyenne (petit écart-type = valeurs resserrées autour de la moyenne, grand écart-type = valeur très dispersée)

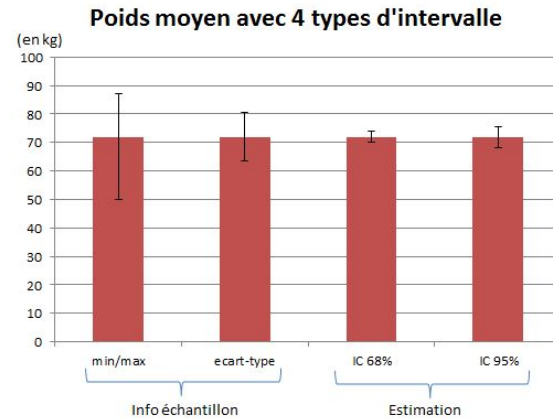
### ► Intervalles permettant de faire une estimation:

- Intervalle de confiance à 68% d'une moyenne
- Intervalle de confiance à 95% d'une moyenne
- Intervalle de confiance à 95% d'une proportion
- Intervalle lié au comptage de colonies en microbiologie
- Intervalle lié à l'incertitude de mesure de la machine

↪ Quand on parle d'intervalle de confiance en mathématique, on est en train de faire une estimation. Le pourcentage donné est lié à la notion de probabilité. On utilise ce type d'intervalle quand on ne connaît pas la vraie moyenne de la population entière et qu'on utilise un échantillon d'individus pour avoir une idée de cette valeur. Ce que dit cet intervalle c'est qu'il y a 95% de chance que la vraie moyenne soit dans cet intervalle (d'autres pourcentages sont possibles).

**⚠ Règle fondamentale:** lorsque l'on met des barres d'erreur sur un graphique, toujours indiquer leur type.

Voici un exemple de comparaison des 4 premiers types d'intervalle sur une série de mesures de poids en kg de 20 personnes.



Bien sûr, ces intervalles sont plus ou moins larges, on imagine qu'un intervalle à 95 % de confiance sera plus grand qu'un intervalle à 68 %... et plus on a de données, c'est-à-dire plus  $n$  est grand, plus les intervalles de confiance sont petits...

## 3) Déterminer les valeurs des barres d'erreurs

**⚠ Règle fondamentale:** Comme les données brutes sont rentrées dans un tableur, utiliser les formules du tableur pour faire vos calculs (et surtout pas la calculatrice!!)

### a) Intervalle du min/max

Valeur minimale d'une série de données situées des cases A1 à A10 : `=MIN(A1:A10)`

Valeur maximale d'une série de données situées des cases A1 à A10 : `=MAX(A1:A10)`

### b) Intervalles liés à une moyenne

Pour ces intervalles, on a besoin de deux grandeurs : la moyenne empirique  $m$  et une estimation de l'écart-type empirique notée  $\hat{s}$  dans la suite.

On considère une série de  $n$  données notées  $x_1, x_2, \dots, x_n$  pour les formules mathématiques, situées dans les cases A1 à An pour les formules tableur (attention  $n$  devra être remplacé par une valeur numérique)

	Moyenne empirique $m$	Écart-type empirique $\hat{s}$
<b>Maths</b>	$\frac{1}{n} \sum_{i=1}^n x_i$	$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2}$
<b>Excel</b>	<code>=MOYENNE(A1:An)</code>	<code>=ECARTYPE.STANDARD(A1:An)</code>
<b>OpenOffice</b>	<code>=MOYENNE(A1:An)</code>	<code>=ECARTYPE(A1:An)</code>

Voici les 3 types d'intervalles que l'on peut obtenir à partir de ces données:

• Intervalle sur une moyenne avec l'écart-type seul:  $[m - \hat{s}, m + \hat{s}]$

• **Intervalle de confiance à 68% d'une moyenne:**  $\left[ m - \frac{\hat{s}}{\sqrt{n}}, m + \frac{\hat{s}}{\sqrt{n}} \right]$

(aussi appelé erreur standard à la moyenne)

• **♥ Intervalle de confiance à 95% d'une moyenne :**  $\left[ m - \alpha \frac{\hat{s}}{\sqrt{n}}, m + \alpha \frac{\hat{s}}{\sqrt{n}} \right]$ ,

où  $\alpha$  est une constante dépendant du nombre de mesures effectuées et du pourcentage de confiance de l'intervalle (elle est liée à des lois de probabilités, ici la loi de Student). Différentes valeurs possibles sont données ci-dessous pour le niveau de confiance à 95%.

nb de mesures effectuées	2	3	4	5	6	7	8	9	10
valeur de $\alpha$	12.71	4.3	3.18	2.78	2.57	2.45	2.36	2.31	2.26

↪ Si vous avez un nombre différent de mesures :venez me voir!

↪ Si vous voulez être plus ou moins exigeant sur la confiance (99%, 90%,...): venez me voir!

**i** Si vous avez plus de 30 mesures, vous pouvez utiliser la valeur  $\alpha = 1,96$  pour le niveau de confiance à 95% (qui correspond cette fois à la loi normale...)

**Remarque:** Cet intervalle est centré sur la moyenne. On s'écarte de chaque côté de la même valeur, ici  $\alpha \frac{\hat{s}}{\sqrt{n}}$ : donc plus cette valeur est petite, plus l'intervalle est petit donc resserré autour de la moyenne (et donc mieux pour en conclure quelque chose). Si vous avez peu de mesures, les valeurs de  $\alpha$  du tableau sont grandes et  $\sqrt{n}$  petit donc l'intervalle risque d'être assez grand...(d'où l'intérêt en pratique d'avoir beaucoup de valeurs...)

**Un peu de vocabulaire:** la valeur  $\frac{s}{\sqrt{n}}$  est appelée écart standard à la moyenne. Plus votre nombre de mesures est grand, plus cet écart diminue et donc plus votre intervalle est petit.

### c) Intervalle lié à une proportion, $n \geq 30$

**Cas d'utilisation:** Lorsque notre échantillon est grand (de taille supérieure ou égale à 30) et lorsque l'on a une seule mesure par échantillon, parce que l'on regarde la proportion d'un caractère de notre échantillon.

**Exemple pratique:** Lorsqu'on étudie un taux de mortalité. On part d'une population de 50 crevettes dans un aquarium que l'on met dans des conditions de stress et on regarde chaque semaine la proportion d'individus morts. On a donc une seule mesure par semaine. Un intervalle de confiance de notre taux de mortalité est donné ci-dessous.

**Intervalle à :95%**  $\left[ p - 1.96 \sqrt{\frac{p(1-p)}{n}}; p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right]$  avec

★  $n$  la taille de l'échantillon (= nombre total d'individus)

★  $p$  la proportion mesurée dans notre échantillon  $\frac{\text{nb individus ayant le caractère}}{n}$

**Remarque:** Le coefficient 1,96 est à relier à la loi normale, cf cours de mathématiques de TB2...

### d) Autres intervalles

Je vous invite à consulter votre cours de micro-biologie ou de physique.

## III. En pratique, comment faire un graphique?

### 1) Préliminaires

Comme nous l'avons vu précédemment, on ne représente pas forcément les données brutes sur notre graphique, notamment lorsqu'elles sont nombreuses où l'on va privilégier la moyenne. De plus, les "barres d'erreur" que nous allons mettre sur notre graphique n'ont pas forcément la même largeur pour chaque point du graphique, donc il faudra également calculer cette largeur à chaque fois. Ainsi on va construire un tableau, à part des données brutes (soit à côté, soit sur une autre feuille de calcul) qui va regrouper toutes les valeurs qui vont nous servir pour notre graphique. Dans les tableaux suivants, erreur représente la moitié de la longueur de l'intervalle à déterminer : par exemple pour l'intervalle  $\left[ m - \alpha \frac{\hat{s}}{\sqrt{n}}, m + \alpha \frac{\hat{s}}{\sqrt{n}} \right]$ , l'erreur est  $\alpha \frac{\hat{s}}{\sqrt{n}}$ .

**⚠** Ne surtout pas effacer les données brutes !!

**⚠** Et bien sûr, pour remplir les lignes des tableaux lorsqu'il s'agit de calculer des moyennes, écart-types, ... on utilise les formules du tableur !

Voici quelques exemples de tableau dans divers cas de figure (non exhaustif), IC désigne Intervalle de Confiance :

IC sur la proportion:

Temps	$t_0$	$t_1$	...
Proportion			
$n$			
Erreur			

IC sur la moyenne:

Temps	$t_0$	$t_1$	...
Moyenne			
$n$			
Erreur			

J'ai choisi de mettre une ligne avec la taille  $n$  de l'échantillon dans le tableau car cette taille peut varier au cours du temps pour diverses raisons (prélèvement d'individus effectués, individus morts au cours du temps, individus nés au cours du temps, ...)

**i** Concrètement, les deux premières lignes du tableau vont nous servir pour tracer le graphique (abscisse/ordonnée) et la dernière ligne va nous servir pour mettre les barres d'erreur.

## 2) Tracé du graphique

Nous allons utiliser les deux premières lignes de notre tableau des préliminaires pour tracer le graphique (courbe ou histogramme) correspondant à la moyenne, la proportion (ou toute autre valeur que vous avez choisi de représenter)

### • Avec Excel :

On sélectionne les lignes (ou colonnes) de données puis on va dans *insertion* et on choisit son type de graphique : *colonne* pour histogramme 2D et *lignes* pour les courbes 2D

Ensuite lorsque le graphique est sélectionné on a accès à 3 onglets spéciaux qui permettent de modifier, améliorer le graphique. En particulier l'onglet *disposition* permet de mettre un titre, des légendes, ...

Pour modifier la légende, si ce n'est pas la première valeur de la ligne, aller sur la légende et cliquer droit, puis *sélectionner des données*, puis cliquer sur la série concernée et faire modifier et mettre la légende souhaitée.

Pour modifier l'échelle, cliquer sur l'axe et changer les valeurs minimales et maximales

### • Avec OpenOffice :

On sélectionne les lignes (ou colonnes) de données puis on va dans *insertion* puis *diagramme*. Un graphique s'affiche et on peut suivre les instructions sur la fenêtre affichée pour modifier le type, ajouter un titre, etc.

Pour modifier l'échelle, cliquer sur l'axe et changer les valeurs minimales et maximales

### **i** Quelques règles de présentation du graphique:

Mettre un titre, une légende, des titres sur les axes, numéroter vos graphiques s'ils sont nombreux pour bien les identifier.

## 3) Ajout des barres d'erreur

Une fois que le graphique est tracé avec la valeur centrale de notre échantillon (moyenne, proportion, ...), on va ajouter les barres d'erreur correspondant à l'intervalle calculé dans le tableau, pour soit donner plus d'informations sur notre échantillon, soit pour faire une estimation.

**⚠ Rappel:** il faudra toujours préciser à côté de votre graphique (sur la fiche synthèse ou dans votre présentation orale) quel est le type d'intervalle que vous avez choisi pour les barres d'erreur.

Concrètement, nous allons utiliser la dernière ligne de notre tableau des préliminaires.

• **Avec Excel :** Cliquer sur le graphique sur la courbe correspondant à la série regardée ou une barre de la série. Aller dans l'onglet *disposition* et aller dans *barre d'erreurs, autres options de barre d'erreur*. Vous pouvez personnaliser les valeurs que vous mettez en erreur, pour utiliser les erreurs que vous avez calculées. Pour cela aller dans *personnaliser : spécifier une valeur* et utiliser l'icône permettant de sélectionner la ligne ou vous avez calculé vos erreurs.

• **Avec Openoffice :** Clic droit sur la série de données représentée sur le graphique et choisir *insérer des barres d'erreurs Y*. Dans la fenêtre qui s'affiche, choisir *plage de cellule* et en cliquant sur l'icône de sélection, aller ensuite sélectionner les erreurs positives puis négatives (même chose) dans le tableau pour les ajouter.

## 4) Exemple

Reprenons l'exemple d'introduction. Les barres d'erreur sur les graphiques correspondant à un intervalle de confiance à 95 % pour la moyenne, avec 5 mesures. Voici les tableaux correspondants et en-dessous les formules excel de la partie traitement de données (Ne pas tout taper mais écrire une seule formule et la copier/coller!) Les graphiques obtenus ont été donnés en page 1.

Tableau des données brutes:

	A	B	C	D	E
1	Evolution de la taille en cm de 5 plants de tomate selon arrosage				
2					
3	Données :				
4		Semaine 0	Semaine 1	Semaine 2	
5	20 cl./jour	4,8	7,2	8	
6		5	6,8	8,1	
7		4,7	6,9	8	
8		5,1	7,5	7,5	
9		5,2	5,9	7,6	
10	20 cl./2 jours	4,8	6,9	7,5	
11		5	6,2	7,6	
12		4,7	6,7	8	
13		5,1	7	7,5	
14		5,2	6,6	7,6	

Tableau des préliminaires (valeurs):

	A	B	C	D	E
1	Evolution de la taille en cm de 5 plants de tomate selon arrosage				
16	Traitement des données :				
17					
18	20 cl./jour		Semaine 0	Semaine 1	Semaine 2
19	Moyenne empirique		4,96	6,86	7,84
20	Ecart-type empirique		0,21	0,60	0,27
21	n		5,00	5,00	5,00
22	Erreur 95%		0,26	0,75	0,34
23					
24	20 cl./2jour		Semaine 0	Semaine 1	Semaine 2
25	Moyenne empirique		4,96	6,68	7,64
26	Ecart-type empirique		0,21	0,31	0,21
27	n		5,00	5,00	5,00
28	Erreur 95%		0,26	0,39	0,26

Tableau des préliminaires (formules tapées):

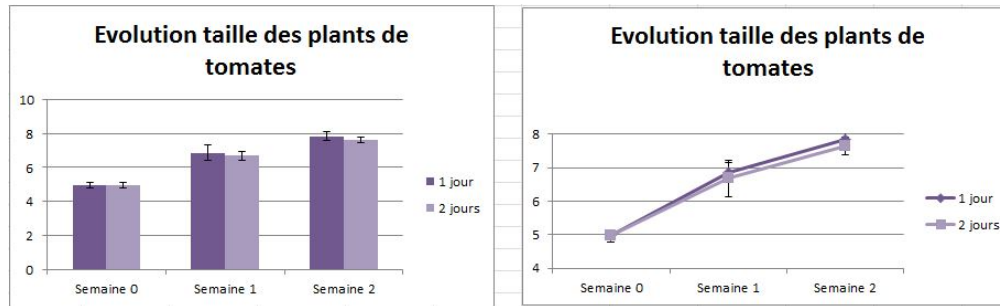
	A	C	D	E
16	Traitement des données :			
17				
18	20 cl./jour	Semaine 0	Semaine 1	Semaine 2
19	Moyenne empirique	=MOYENNE(B5:B9)	=MOYENNE(C5:C9)	=MOYENNE(D5:D9)
20	Ecart-type empirique	=ECARTYPE.STANDARD(B5:B9)	=ECARTYPE.STANDARD(C5:C9)	=ECARTYPE.STANDARD(D5:D9)
21	n	5	5	5
22	Erreur 95%	=2,78*C20/RACINE(C21)	=2,78*D20/RACINE(D21)	=2,78*E20/RACINE(E21)
23				
24	20 cl./2jour	Semaine 0	Semaine 1	Semaine 2
25	Moyenne empirique	=MOYENNE(B10:B14)	=MOYENNE(C10:C14)	=MOYENNE(D10:D14)
26	Ecart-type empirique	=ECARTYPE.STANDARD(B10:B14)	=ECARTYPE.STANDARD(C10:C14)	=ECARTYPE.STANDARD(D10:D14)
27	n	5	5	5
28	Erreur 95%	=2,78*C26/RACINE(C27)	=2,78*D26/RACINE(D27)	=2,78*E26/RACINE(E27)

## IV. Interpréter les résultats

Pour réussir à interpréter les résultats, on a en général besoin d'intervalles de confiance petits, et donc d'avoir un grand nombre de valeurs. Mais même si dans votre TIPE vous avez peu de mesures, donc de grands intervalles, l'intérêt est de s'entraîner ici à la démarche statistique...mais il s'agira de rester honnêtes dans vos conclusions...

- **Cas des intervalles résumant l'échantillon:** Si l'intervalle moyenne/écart-type obtenu est grand : les valeurs sont très dispersées autour de la moyenne (échantillon hétérogène), sinon elles sont très proches ( échantillon homogène )
- **Cas des intervalles de confiance:** lorsque l'on compare des conditions de culture différentes: si les intervalles sont disjoints, la différence entre les deux conditions de culture est significative. Sinon on ne peut rien conclure.

Interprétation de l'exemple d'introduction :



Ici les moyennes sont proches et les intervalles de confiance se chevauchent donc on ne peut pas conclure sur une différence significative entre les 2 façons d'arroser.

Amélioration : faire plus de mesures (deux fois par semaine), utiliser plus de plants, ou bien mesurer plus longtemps pour avoir plus de points et peut-être mettre en évidence une meilleure tendance...