



## SPIPOLL

*À vos appareils photos pour devenir paparazzi des pollinisateurs !*

Proposition d'activité pédagogique



© Steed - Spipoll



## Impact de la météo sur la fréquentation des insectes

Exploitation statistique d'une base de données avec un tableur pour évaluer l'Impact des conditions météo sur la fréquentation des insectes

### Disciplines concernées - lycée :

- Sciences de la Vie et de la Terre
- Mathématiques
- Enseignement scientifique
- Physique - Chimie

### Auteurs :

-  Sébastien Turpin  
(enseignant de SVT)
-  Philippe Dutarte  
(IA IPR de Mathématiques)
-  Tristan Briant  
(maître de conférence)





## En quoi consiste Vigie-Nature École ?

Vigie-Nature École est un programme de sciences participatives qui vise à suivre la réponse de la biodiversité aux activités humaines et aux changements globaux (urbanisation, intensification de l'agriculture et changement climatique).

Pour y parvenir, nous proposons aux enseignants de mettre en place avec leurs élèves des protocoles permettant de suivre plusieurs groupes d'êtres vivants. Toutes les observations faites sont ensuite envoyées aux chercheurs du Muséum pour qu'ils puissent s'en servir dans leurs recherches.

## En quoi consiste cette activité ?

### Niveau

Première voie générale.

### Contenus liés aux mathématiques

- Croisement de variables catégorielles.
- Tableau croisé d'effectifs.
- Écart type.
- Variabilité d'échantillonnage.
- Fréquence. Approche fréquentiste de la notion de probabilité.

### Contenus liés au SVT

- Identifier, quantifier et comparer la biodiversité.
- Mettre en œuvre des protocoles d'échantillonnage statistique permettant des descriptions rigoureuses concernant la biodiversité.
- Suivre une campagne d'études de la biodiversité et y participer.
- Extraire et organiser des informations, issues de l'observation directe sur le terrain, pour savoir décrire les éléments et les interactions au sein d'un système. Comprendre l'importance de la reproductibilité des protocoles d'échantillonnage pour suivre la dynamique spatio-temporelle d'un système.

### Origine des données

Suivi photographique des insectes pollinisateurs.  
<http://www.spipoll.org/>





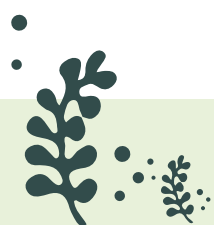
Le fichier « base\_SPIPOLL.xlsx » (format Microsoft Excel) ou « base\_SPIPOLL.ods » (format OpenOffice ou LibreOfficeCalc) comporte 69 481 lignes de données. Chaque ligne correspond à une observation d'insecte (espèce ou groupe d'espèce appelé taxon). Ces données ont été collectées dans le cadre d'un protocole de sciences participatives appelé le Suivi Photographique des Insectes pollinisateurs (Spipoll) . Ce protocole demande aux élèves de photographier tous les insectes se posant sur une fleur durant 20 minutes. À l'issue de ces 20 minutes d'observation, chaque insecte est nommé grâce à une clé de détermination.

Le fichier comporte 6 variables :

- « **collection** » correspondant au numéro de la collection, c'est-à-dire à une participation au protocole Spipoll. Plusieurs insectes pouvant être vus en 20 minutes, il peut y avoir plusieurs lignes avec le même numéro de collection, chaque ligne représentant un insecte ;
- « **mois** » correspondant au mois de l'année, de janvier (1) à décembre (12) ;
- « **ciel** » indiquant la nébulosité du ciel ;
- « **vent** » indiquant la nature du vent et
- « **taxon** » correspondant au nom du groupe taxonomique (groupe d'espèces) d'insectes observé.

	A	B	C	D	E	F
1	collection	mois	ciel	température	vent	taxon
2	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Les Coccinelles <Coccinellidae>
3	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Les Halictes (femelles) <Halictus, Lasioglossum et autres>
4	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	L'Eristale des fleurs <Myathropa florea>
5	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Le Drap mortuaire <Oxythyrea funesta>
6	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Les Fourmis difficiles à déterminer <Formicidae>
7	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Taxon inconnu de la clé
8	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Taxon inconnu de la clé
9	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	Les Chrysanthies et autres <Chrysanthia, Ischnomera>
10	46	5	75 à 100 %	10 - 20°C	faible, irrégulier	La Valgue hémiptère <Valgus hemipterus>
11	64	5	0 à 25 %	10 - 20°C	faible, irrégulier	Les Cantharides <Cantharis>
12	64	5	0 à 25 %	10 - 20°C	faible, irrégulier	Les larves de Punaises <Heteroptera>

En explorant ces données à l'aide d'un tableur, il s'agit d'examiner l'impact de la météo sur la diversité (nombre de taxons observés) de ces insectes.



## Partie A : étude de la biodiversité selon la température

### 1. On souhaite croiser les deux variables « collection » et « température » de façon à visualiser le nombre de taxons observés.

- Avec Excel, cliquer dans une cellule des données et faire « Insertion / Tableau croisé dynamique » et choisir sur une « Nouvelle feuille de calcul ». Sélectionner les variables en les cochant dans la boîte de dialogue, puis les glisser en « Étiquettes de lignes » ou «  $\Sigma$  Valeurs ». Examiner l'évolution dynamique du tableau selon vos choix de glissement afin d'obtenir le résultat désiré.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3	Nombre de taxon	Étiquettes de colonnes									
4	Étiquettes de lignes	+ de 30°C	10 - 20°C	10°C et moins	20 - 30°C	Total général					
5	46		9			9					
6	64		13			13					
7	141		5			5					
8	170				9	9					
9	209		23			23					
10	220		4			4					
11	360				6	6					
12	366		13			13					
13	367	6			18	24					
14	376		23			23					
15	678				2	2					
16	722				27	27					
17	832		6			6					
18	850		16			16					
19	863		1			1					
20	909		1			1					
21	926		5			5					
22	976				13	13					
23	1004				16	16					
24	1019				4	4					
25	1044		14			14					

Liste de champs de tableau croisé dynamique

Choisissez les champs à inclure dans le rapport :

collection  
 mois  
 ciel  
 température  
 vent  
 taxon

Faites glisser les champs dans les zones voulues ci-dessous:

Filtrer le rapport

Étiquettes de colonnes: température

Étiquettes de lignes: collection

$\Sigma$  Valeurs: Nombre de taxon

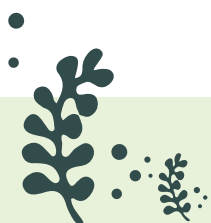
Différer la mise à jour de la disposition

Mettre à jour

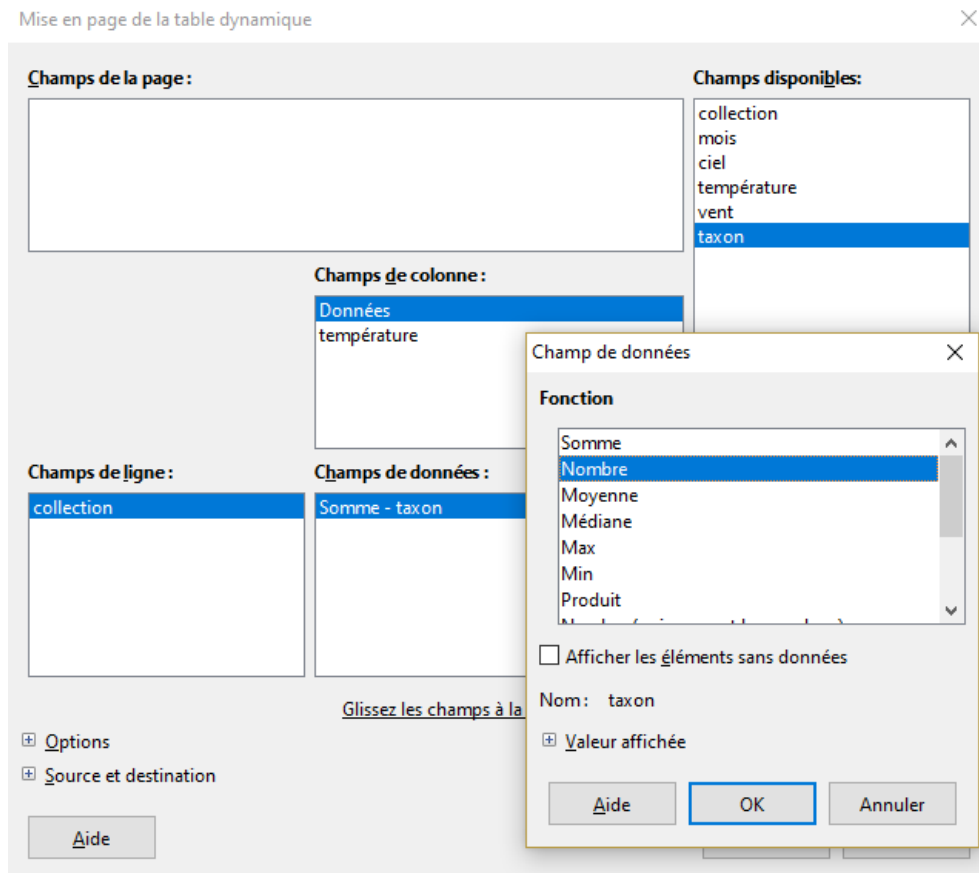
Par un clic droit en cellule B4, choisir « Déplacer » pour modifier l'ordre des colonnes afin que les classes de températures apparaissent dans l'ordre croissant comme ci-dessous.

	A	B	C	D	E	F
1						
2						
3	Nombre de taxon	Étiquettes de colonnes				
4	Étiquettes de lignes	10°C et moins	10 - 20°C	20 - 30°C	+ de 30°C	Total général
5	46		9			9
6	64		13			13
7	141		5			5
8	170			9		9
9	209		23			23
10	220		4			4

### Interpréter les valeurs figurant à la ligne 6.



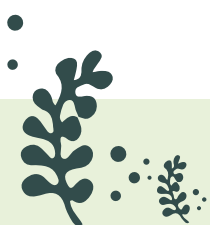
- Avec OpenOffice ou LibreOffice Calc, cliquer dans une cellule des données et faire « Insertion / Table dynamique... » puis glisser, parmi les « Champs disponibles », la variable « collection » en « Champs de ligne », la variable « température » en « Champs de colonne » et la variable « taxon » en « Champs de données ».



Effectuer un double clic sur la variable en « Champs de données » pour afficher la boîte de dialogue permettant de choisir la fonction « Nombre ». On obtient le résultat suivant.

	A	B	C	D	E	F
1	Compter - taxon	Données				
2	collection ▼	+ de 30°C	10 - 20°C	10°C et moins	20 - 30°C	Total Résultat
3	46		9			9
4	64		13			13
5	141		5			5
6	170				9	9
7	209		23			23
8	220		4			4
9	360				6	6
10	366		13			13
11	367	6			18	24
12	376		23			23

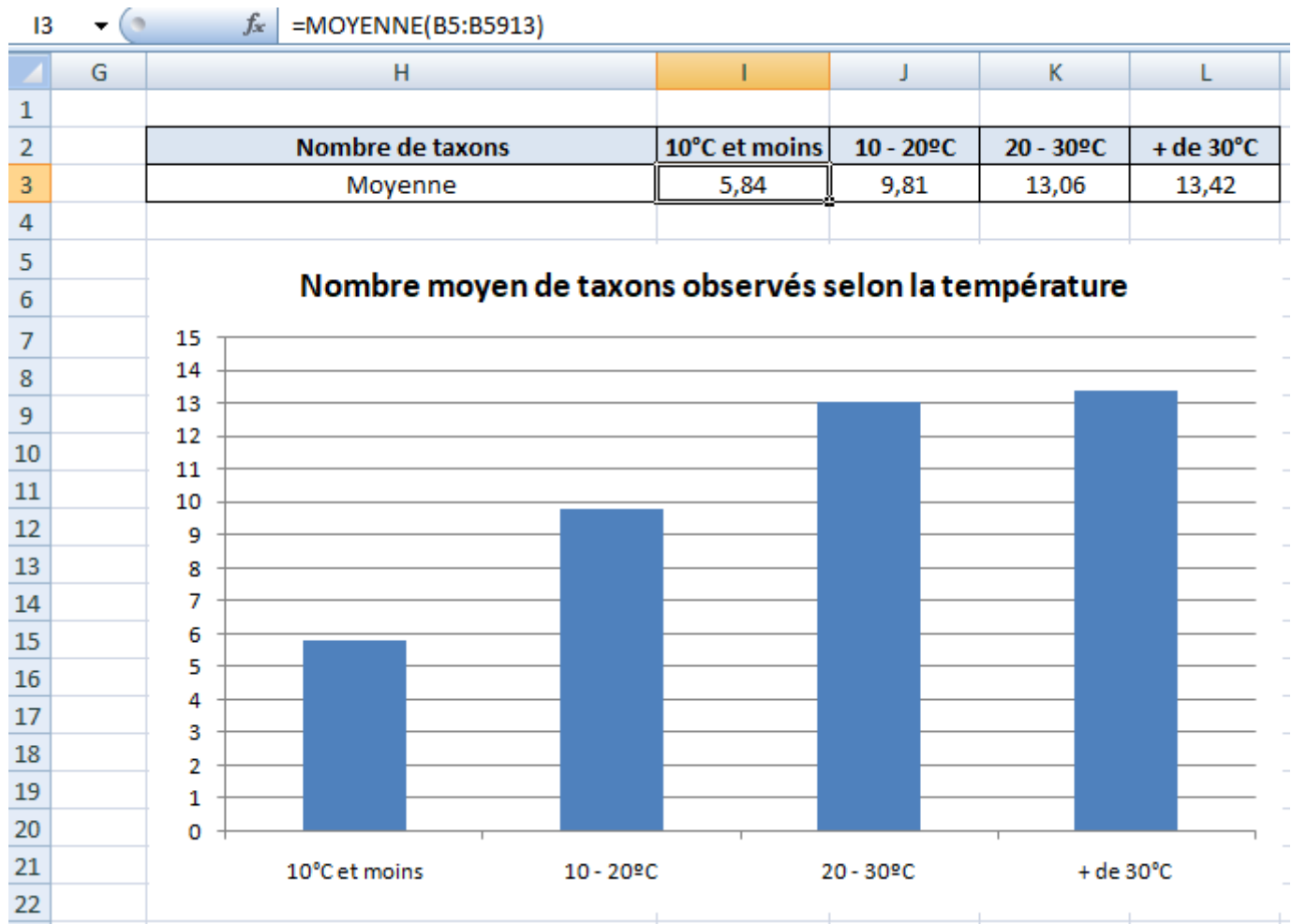
Interpréter les valeurs figurant à la ligne 4.



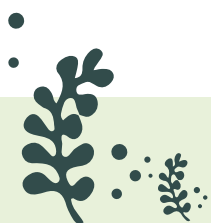
## 2. Représenter le nombre moyen de taxons observés par séance de 20 minutes selon la température (voir l'image d'écran suivante).

Sur Excel, le nombre moyen de taxons observés pour une température inférieure ou égale à 10°C est donné par la formule =MOYENNE(B5:B5913).

Sur Calc, le nombre moyen de taxons observés pour une température inférieure ou égale à 10°C est donné par la formule =MOYENNE(D3:D5911).



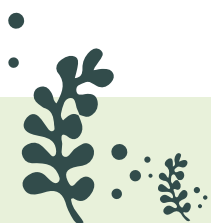
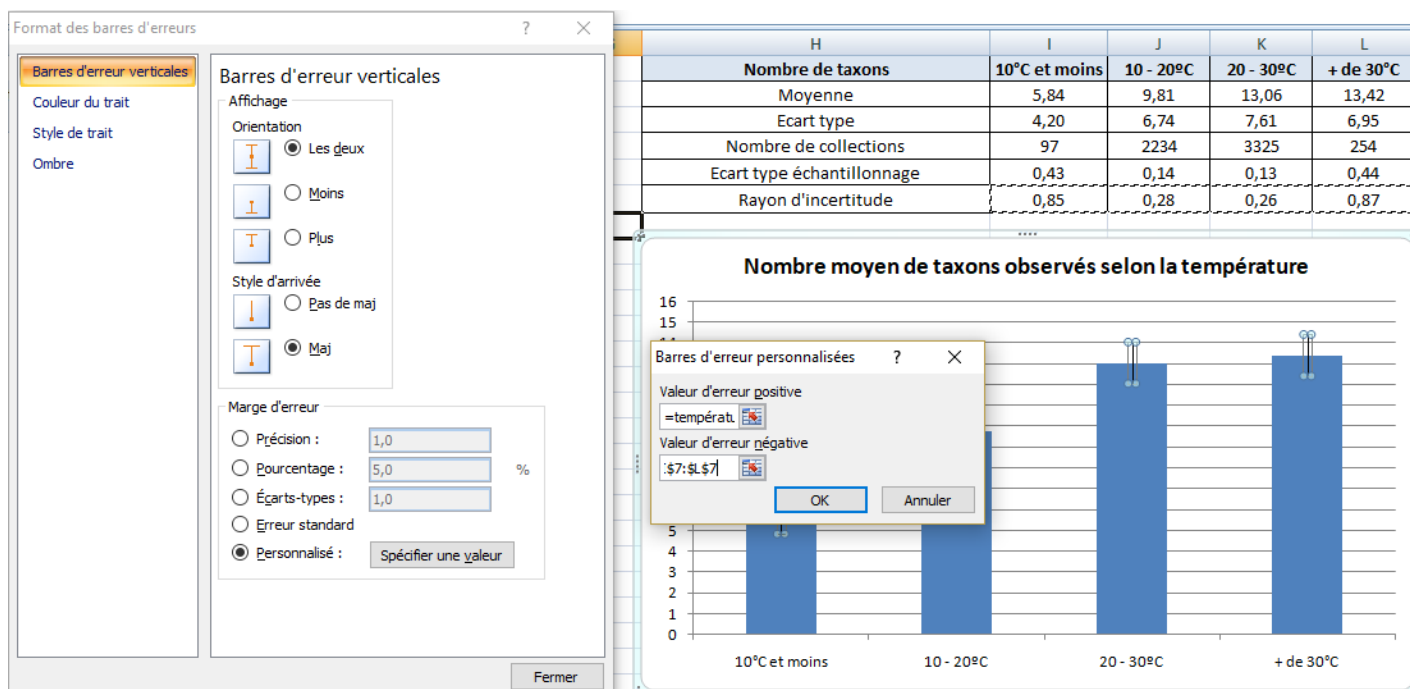
3. Le nombre de collections utilisées pour calculer ces quatre moyennes n'est pas le même pour chacune. Or plus on utilise d'observations pour calculer une moyenne, plus celle-ci est précise, c'est-à-dire représentative de la réalité. On sait en effet, d'après l'étude de la fluctuation d'échantillonnage effectuée en mathématiques, qu'en prenant un échantillon de taille  $n$ , on divise l'écart type mesurant la dispersion par  $\sqrt{n}$ . **On souhaite accompagner chacune des moyennes précédentes par un intervalle mesurant la précision de chacune.**





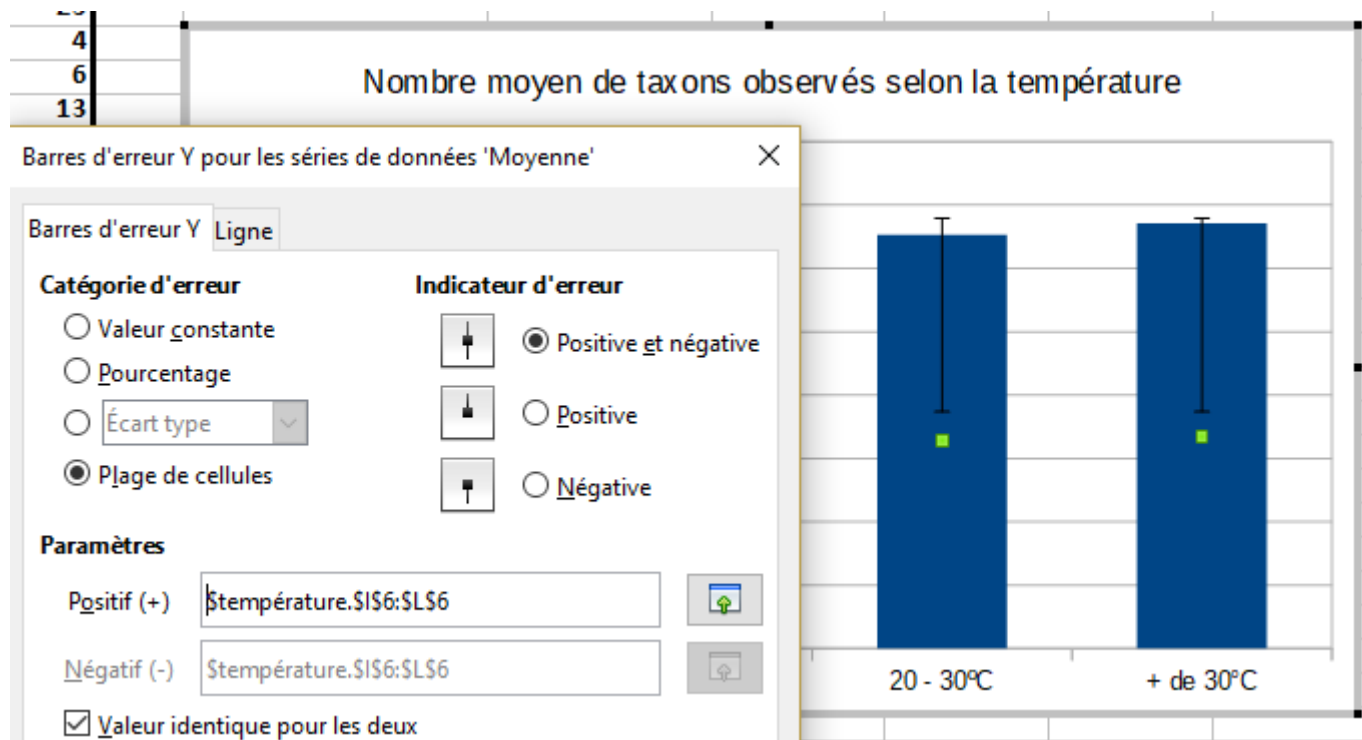
- Calculer l'écart type de chacune des quatre séries correspondant au nombre de taxons observés selon la température en utilisant la formule =ECARTYPEP(plage) ou, sur certaines versions récentes d'Excel, =ECARTYPE.PEARSON(plage). Pour les observations à 10°C et moins, la plage de cellules du tableau dynamique Excel est B5:B5913 et la plage de cellules du tableau dynamique Calc est D3:D5911.
- Calculer, pour chacune des quatre classes de températures, le nombre de collections, c'est-à-dire la taille de chacun des quatre échantillons, en utilisant la formule =NBVAL(plage) qui compte le nombre de cellules non vides d'une plage de cellules.
- Calculer une estimation de l'écart type d'échantillonnage (écart type de la distribution du nombre moyen de taxons sur des collections de taille  $n$ ) obtenue en divisant l'écart type obtenu à la question a. par la racine de  $n$ .
- Pour une distribution normale (on dit aussi « gaussienne »), environ 95 % des valeurs s'écartent de la moyenne de moins de deux écarts types. On peut supposer que la distribution du nombre moyen de taxons sur une collection de taille  $n$  est normale. On définit ainsi un rayon d'incertitude autour de la moyenne observée, au niveau de confiance de 95 %, en multipliant par deux l'écart type d'échantillonnage. Calculer ce rayon d'incertitude pour chacune des quatre nombres moyens de taxons observés pour chaque classe de température.
- On peut matérialiser ces barres d'incertitude sur le graphique en ajoutant ce que le tableur nomme « barre d'erreur ».

Avec Excel, cliquer sur l'histogramme et, dans l'onglet « Disposition », choisir « Barres d'erreur » puis « Autres options de barres d'erreur ». Dans la boîte de dialogue « Format des barres d'erreurs », choisir « Personnalisé » et cliquer sur « Spécifier une valeur ».



Sélectionner alors les rayons d'incertitude sur la feuille de calcul.

Avec Calc, cliquer sur l'histogramme puis sur « Insérer barre d'erreur Y... ». Compléter



alors la boîte de dialogue comme ci-dessous.

- f. On considère que la différence entre deux mesures est « significative » lorsque les « barres d'erreurs » correspondantes sont disjointes. Analyser le graphique obtenu pour le nombre moyen de taxons observés selon la température.

## Partie B : étude de la biodiversité selon la nébulosité et le mois de l'année

1. Reprendre la méthodologie de la partie A pour étudier l'impact de la nébulosité sur la fréquentation des insectes.
  - a. Construire un tableau croisé dynamique comptant le nombre de taxons en croisant les variables « collection » et « ciel ».
  - b. Calculer le nombre moyen de taxons pour chaque classe de nébulosité ainsi que le rayon d'incertitude correspondant. Représenter ces moyennes avec des « barres d'erreurs ».
  - c. Interpréter le graphique obtenu.





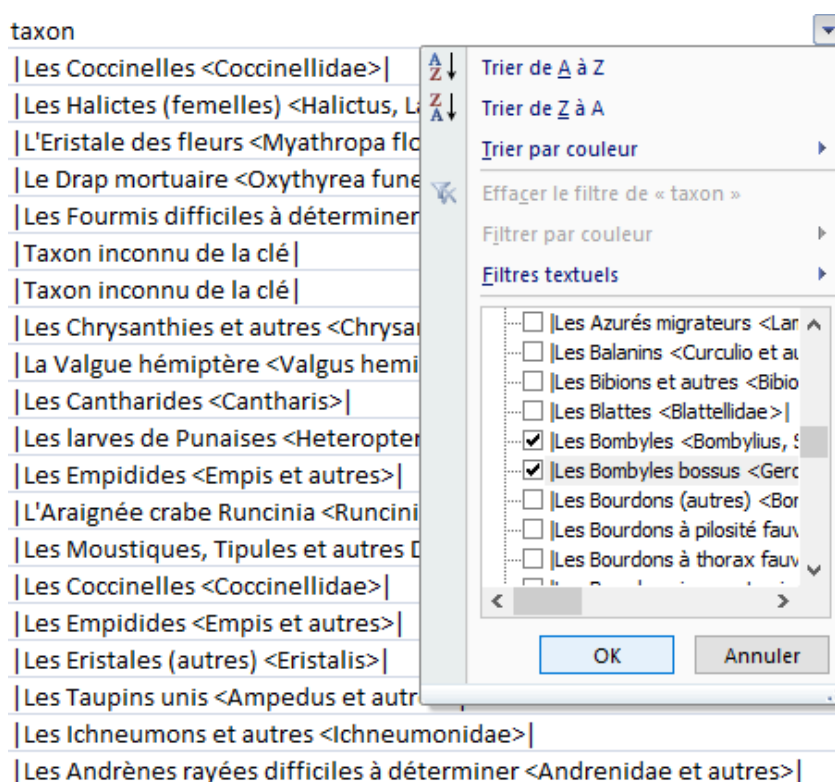
## 2. Reprendre la méthodologie de la partie A pour étudier l'impact du mois de l'année sur la fréquentation des insectes.

- Construire un tableau croisé dynamique comptant le nombre de taxons en croisant les variables « collection » et « mois ».
- Calculer le nombre moyen de taxons pour chaque mois ainsi que le rayon d'incertitude correspondant. Représenter ces moyennes avec des « barres d'erreurs ».
- Interpréter le graphique obtenu.

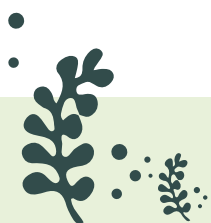
## Partie C : estimation de la probabilité d'observer des Bombyles selon les mois de l'année

On cherche à estimer la probabilité d'observer des Bombyles selon les différents mois de l'année. Pour cela, compte tenu du grand nombre de collections que possède la base de données, on va calculer la fréquence d'observation des Bombyles dans les collections de la base selon les différents mois. Vu le grand nombre d'observations, ces fréquences seront considérées comme de bonnes estimations des probabilités correspondantes.

### 1. Aller dans l'onglet des données et filtrer la variable « taxon » : cliquer sur la flèche



à droite en cellule F1, désélectionner tout puis sélectionner dans le menu déroulant les deux taxons « Les Bombyles » et « Les Bombyles bossus ».



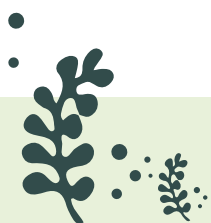
Sélectionner et copier les données filtrées et les coller dans une nouvelle feuille de calcul.

H3    fx    =NB.SI(\$B2:\$B158;H2)

	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
2		1	2	3	4	5	6	7	8	9	10	11	12
3	Nombre de collections avec Bombyles	0											
4	Nombre de collections	88	257	358	344	446	554	1085	1132	745	494	289	117
5	Fréquence des collections avec Bombyles												

**2. Préparer un tableau comme ci-dessous. Le nombre de collections pour chaque mois de l'année a été obtenu à la question B.**

- a. Déterminer le nombre de collections avec des Bombyles selon les mois de l'année en entrant en H3 la formule =NB.SI(\$B2:\$B158;H2) puis en la recopiant vers la droite. Expliquer la formule utilisée.
- b. Estimer les probabilités d'observer des Bombyles selon les différents mois de l'année.





## Éléments de réponse

A.

1. La collection numéro 64 comporte 13 taxons observés alors que la température est comprise entre 10 et 20°C.

2. Image d'écran donnée dans l'énoncé.

3. a. Remarque pour le professeur : veiller à bien utiliser la fonction ECARTYPEP du tableur qui calcule l'écart type des valeurs d'une plage de cellules. La fonction ECARTYPE du tableur calcule une estimation de l'écart type de la population dont la plage de cellules constitue un échantillon. Compte tenu du nombre de données, les deux valeurs sont cependant ici très proches.

b. On obtient la taille  $n$  de chacun des quatre échantillons.

c. **Remarque pour le professeur** : on obtiendrait une meilleure estimation (« sans biais ») de l'écart type de la distribution d'échantillonnage du nombre moyen de taxons sur des échantillons de taille  $n$  en utilisant l'expression :

$$\frac{S_{n-1}}{\sqrt{n}}$$

où  $S_{n-1}$  est l'estimation de l'écart type du nombre de taxons obtenu à la question 3.a. par la formule ECARTYPE. Cependant, on ne peut pas expliquer la notion d'estimateur sans biais à ce niveau d'étude et la différence obtenue peut être ici négligée.

d. Remarque pour le professeur : le coefficient multiplicatif 1,96 serait plus précis que le coefficient 2. Le rayon de l'intervalle de confiance à 95 % est ainsi donné par la formule :

$$1,96 \frac{S_{n-1}}{\sqrt{n}}$$

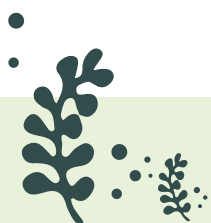
On peut constater que la différence est faible en utilisant, sous Excel, la formule =INTERVALLE.CONFIANCE(0,05;ECARTYPEP(B5:B5913);NBVAL(B5:B5913)) qui fournit un rayon environ égal à 0,84 au lieu de 0,85 selon les calculs menés dans cette activité.

Il nous semble essentiel, plutôt que d'utiliser une formule « boîte noire », que les élèves retiennent :

- qu'en prenant une moyenne sur un échantillon de taille  $n$ , on réduit l'écart type en multipliant par un facteur égal à :

$$\frac{1}{\sqrt{n}}$$

- que, pour une distribution normale (hypothèse fréquemment rencontrée), environ 95 % des valeurs sont comprises dans un intervalle de deux écarts types autour de la moyenne.



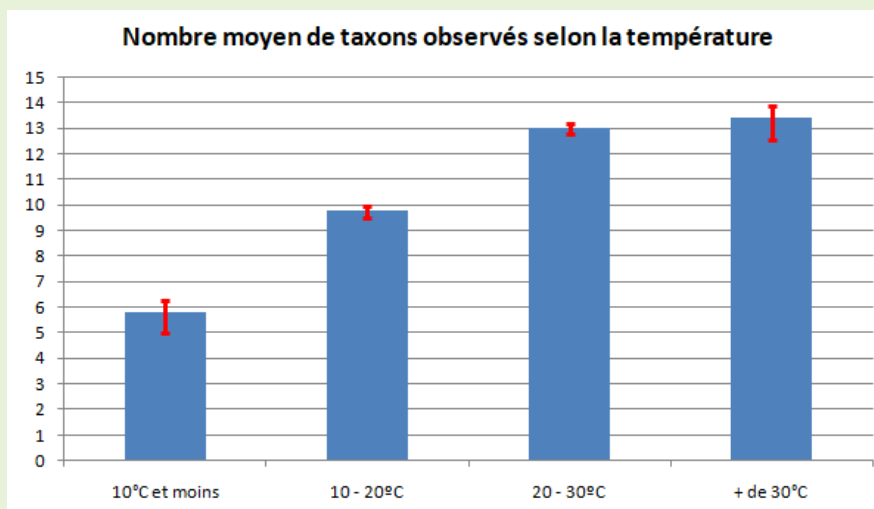


On obtient le tableau suivant :

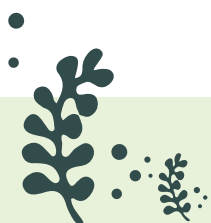
	H	I	J	K	L
1					
2	<b>Nombre de taxons</b>	<b>10°C et moins</b>	<b>10 - 20°C</b>	<b>20 - 30°C</b>	<b>+ de 30°C</b>
3	Moyenne	5,84	9,81	13,06	13,42
4	Ecart type	4,20	6,74	7,61	6,95
5	Nombre de collections	97	2234	3325	254
6	Ecart type échantillonnage	0,43	0,14	0,13	0,44
7	Rayon d'incertitude	0,85	0,28	0,26	0,87

e. **Remarque pour le professeur** : la terminologie, largement pratiquée, de « barre d'erreur » est un peu dangereuse car elle laisse supposer qu'elle correspond à l'erreur maximale alors que le niveau de confiance n'est que de 95 %. Cette « barre d'erreur » est en fait un intervalle de confiance au niveau de confiance de 95 %. C'est pourquoi l'activité parle plutôt de « rayon d'incertitude ». En métrologie, particulièrement en physique et en chimie, on parlerait plutôt « d'incertitude type » correspondant à l'écart type de la distribution d'échantillonnage. La « barre d'erreur » correspond alors à « deux incertitudes types ».

On obtient le graphique suivant :



f. Le nombre moyen de taxons observés augmente avec la température. La différence entre ces nombres moyens entre chaque classe de température est significative sauf pour les deux dernières classes. On ne possède pas assez d'observations avec une température de plus de 30°C (le rayon d'incertitude est assez grand) pour affirmer que le nombre moyen de taxons observés lorsque la température est de plus de 30°C est supérieur à ce qu'il est lorsque la température est comprise entre 20 et 30°C.





B.  
1. On obtient les résultats suivants :

	A	B	C	D	E	F	G	H	I
1									
2									
3	Nombre de taxon	Étiquettes de colonnes							
4	Étiquettes de lignes	0 à 25 %	25 à 50 %	50 à 75 %	75 à 100 %	Total général			
5	46				9	9			
6	64	13				13			
7	141		5			5			
8	170				9	9			
9	209		23			23			
10	220				4	4			
11	360				6	6			
12	366	13				13			
13	367		18		6	24			
14	376		23			23			
15	678			2		2			
16	722	27				27			
17	832		6			6			
18	850		16			16			
19	863				1	1			
20	909				1	1			
21	926				5	5			
22	976				13	13			
23	1004			16		16			
24	1019		4			4			

Liste de champs de tableau croisé dynamique

Choisissez les champs à inclure dans le rapport :

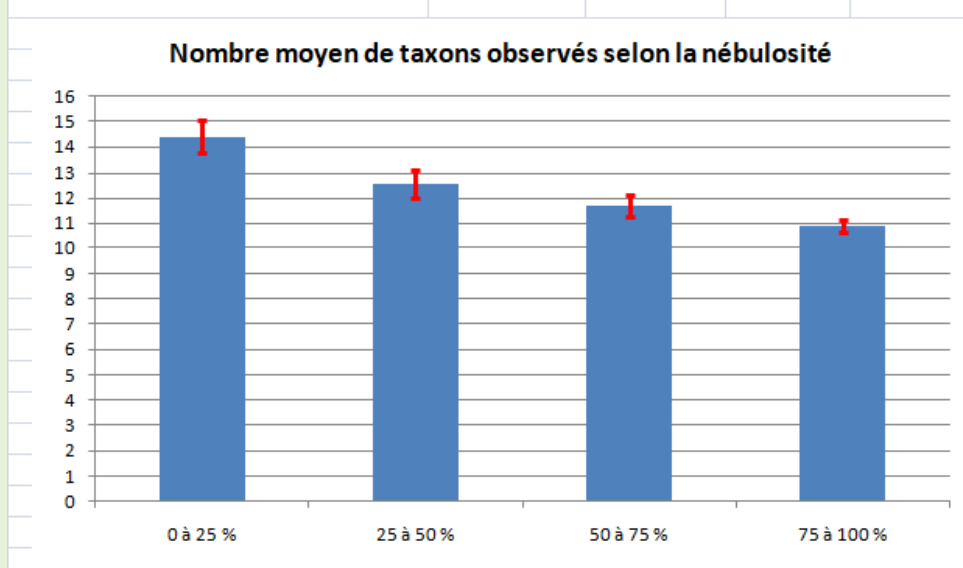
- collection
- mois
- ciel
- température
- vent
- taxon

Faites glisser les champs dans les zones voulues ci-dessous :

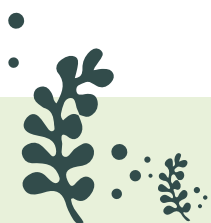
Filtre du rapport       Étiquettes de colonnes  
 Étiquettes de lignes       Valeurs

Différer la mise à jour de la disposition

Nombre de taxons	0 à 25 %	25 à 50 %	50 à 75 %	75 à 100 %
Moyenne	14,40	12,55	11,68	10,88
Ecart type	9,09	7,82	7,06	6,80
Nombre de collections	782	820	1083	3225
Ecart type échantillonnage	0,33	0,27	0,21	0,12
Rayon d'incertitude	0,65	0,55	0,43	0,24



Le nombre moyen de taxons observés diminue lorsque la nébulosité augmente. La différence entre les différentes classes de nébulosité est relativement significative (il y a une très petite intersection entre les intervalles d'incertitude correspondant à une nébulosité comprise entre 25 % et 50 % et une nébulosité comprise entre 50 % et 75 %).



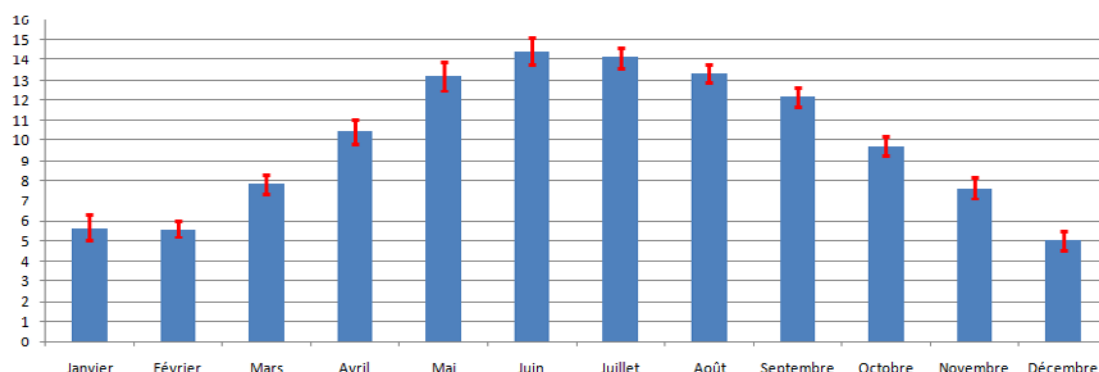


2. On obtient les résultats suivants :

Étiquettes de lignes	1	2	3	4	5	6	7	8	9	10	11	12	Total général
46					9								9
64					13								13
141				5									5
170					9								9
209					23								23
220					4								4
360				6									6
366					13								13
367					24								24
376					23								23
678					2								2
722					27								27
832					6								6
850			16										16
863					1								1
909					1								1
926					5								5
976					13								13
1004					16								16
1019					4								4

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
<b>Nombre de taxons</b>	1	2	3	4	5	6	7	8	9	10	11	12
Moyenne	5,66	5,60	7,83	10,42	13,19	14,44	14,11	13,33	12,18	9,73	7,62	5,04
Ecart type	3,05	3,09	4,50	5,57	7,49	7,97	8,64	7,67	6,61	5,13	4,48	2,71
Nombre de collections	88	257	358	344	446	554	1085	1132	745	494	289	117
Ecart type échantillonnage	0,33	0,19	0,24	0,30	0,35	0,34	0,26	0,23	0,24	0,23	0,26	0,25
Rayon d'incertitude	0,65	0,38	0,48	0,60	0,71	0,68	0,52	0,46	0,48	0,46	0,53	0,50

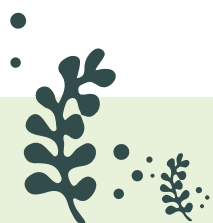
Nombre moyen de taxons observés selon les mois de l'année



Le nombre moyen de taxons observés pendant les mois d'été est plus important que pendant les mois d'hiver, avec une différence significative. La différence n'est pas toujours significative pour deux mois consécutifs.

C.

2. b. La formule =NB.SI(\$B2:\$B158;H2) compte le nombre de cellules de la plage B2:B158 (correspondant à la variable « mois » des collections comportant des Bombyles) dont la valeur est égale à celle de la cellule H2 (qui vaut 1 correspondant au mois de janvier). Cette formule compte donc le nombre de collections où on a observé des Bombyles en janvier (à savoir 0).







c. On obtient les résultats suivants :

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
	1	2	3	4	5	6	7	8	9	10	11	12
Nombre de collections avec Bombyles	0	0	42	41	22	14	16	11	9	2	0	0
Nombre de collections	88	257	358	344	446	554	1085	1132	745	494	289	117
Fréquence des collections avec Bombyles	0	0	0,117	0,119	0,049	0,025	0,015	0,010	0,012	0,004	0	0

Les Bombyles s'observent essentiellement en mars et avril. On peut estimer la probabilité d'observer un Bombyle durant une observation de 20 minutes durant ces mois à environ 0,12. La taille des échantillons correspondant est environ 350. En considérant une incertitude d'environ  $1 / \sqrt{350}$  c'est-à-dire 0,05, on obtient comme intervalle de confiance à 95 % de cette probabilité : [0,07 ; 0,17].

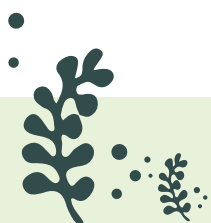
**Remarque pour le professeur :** on utilise là les résultats de la fluctuation d'échantillonnage d'une fréquence observés, par simulation, en classe de seconde générale et technologique. Si l'on observe une fréquence  $f$  sur un échantillon de taille  $n$  (assez grand), on peut estimer que la probabilité  $p$  (avec un niveau de confiance de 95 %) correspondante est comprise entre :

$$f - \frac{1}{\sqrt{n}} \text{ et } f + \frac{1}{\sqrt{n}}$$

En effet, on observe que l'écart type de la distribution d'échantillonnage des fréquences obtenues sur des échantillons de taille  $n$  est environ égal à

$$\frac{1}{2\sqrt{n}}$$

et on construit un intervalle de rayon égal à deux écarts types (la distribution d'échantillonnage est considérée comme normale).





## Le SPIPOLL en bref !

- Un protocole pour étudier les insectes pollinisateurs
- Accessible dès le cycle 3 jusqu'au lycée
- Réalisable toute l'année

## Nos observatoires



## Nous suivre, nous contacter



[vigenature-ecole.fr](http://vigenature-ecole.fr)



[vne@mnhn.fr](mailto:vne@mnhn.fr)



[VigieNatureEcole](https://www.facebook.com/VigieNatureEcole)



[@VigieNatureEcole](https://twitter.com/VigieNatureEcole)

## Un programme du



*Avec l'appui de*



## Partenaire co-fondateur de l'observatoire



*Avec le soutien de*

**AGENCE FRANÇAISE  
POUR LA BIODIVERSITÉ**  
ÉTABLISSEMENT PUBLIC DE L'ÉTAT

