

N.B : Les étudiants ont le choix du système d'exploitation (linux ou Microsoft Windows) et du tableur (Microsoft Excel, OpenOffice Calc ou Gnumeric), dans la mesure des moyens disponibles dans la salle.

Exercice n°1 (Pirse en main) Présentation succincte d'un tableur :

- **Organisation :**
 - fichier = classeur = plusieurs feuilles de calculs (accessibles par les onglets en bas à gauche)
 - contenu d'une cellule = texte, valeurs numériques, fonction, ...
 - adresse d'une cellule : relative (ex : B3, C1) ou absolue (ex : \$A\$1, B\$3, \$D4)
 - possibilité de donner un nom à une cellule
 - plage de cellules : définie par la cellule en haut à gauche et par celle en bas à droite, les deux adresses étant séparées par un ":" (ex : A2 : B4)
 - copier/coller de cellules : qu'observe-t-on dans les cellules et à l'écran lorsqu'on copie une cellule contenant une valeur ? une formule ?
 - collage spécial : que permet de faire le collage spécial ?
- **Fonctions de base :**
 - fonctions mathématiques : abs, cos, sin, arrondi, max, min, puissance, somme, somme.carres, somme.si, somme.prod, nb, nb.si, ...
 - fonctions logiques : est.pair, ou, et, non, si, vrai, faux, ...
 - fonctions statistiques : moyenne, moyenne.reduite, mediane, var.p, ecartype.p, ecart.moyen, quartile, ...
- **Graphique :** se laisser guider par l'assistant graphique du tableur utilisé tout en restant critique à son égard (car il ne s'agit pas d'un logiciel de statistique, même s'il dispose de certaines fonctionnalités).

Exercice n°2 La note finale de cette UE est une moyenne pondérée donnée par la formule suivante :

$$NF = 0.75(0.3CC + 0.7ET) + 0.25TP .$$

où NF est la note finale, CC la note de contrôle continu, ET la note de l'examen terminal et TP la note de projet.

1. Saisir des notes fictives CC , ET et TP pour environ une vingtaine d'étudiants.
2. Proposer une ou plusieurs méthodes pour le calcul de la note finale.
3. On souhaite changer les coefficients. Proposer une méthode qui permette de tester plusieurs cas de manière efficace.

Exercice n°3 Utiliser les formules et le copier-coller pour retrouver les éléments en italique dans le tableau ci-dessous correspondant à des ventes sur une semaine dans un magasin.

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Total
Livres	99	66	33	44	99	121	462
Disques	77	44	66	88	88	222	585
Vidéos	55	99	55	43	88	155	495
Total	231	209	154	175	275	498	1542
Part de la journée	0.14981	0.13554	0.09987	0.11349	0.17834	0.3226	1

Exercice n°4 On considère le nombre d'accidents par mois survenus dans un département en 2005. Les données sont les suivantes :

2693 2695 2688 2690 2700 2705
2701 2710 2711 2708 2710 2715

1. Représenter les effectifs par un graphique dont l'échelle des ordonnées commence à 0. Commenter ce graphique.
2. Représenter les effectifs par graphique dont l'échelle des ordonnées commence à 2680. Commenter ce graphique.
3. Les différentes interprétations possibles des graphiques ci-dessus sont liées au fait qu'on travaille sur une série temporelle (ou série chronologique), i.e. à une série uni-variée où les individus sont le temps. Dans ce cas, on ne peut pas travailler avec les fréquences. Pour essayer de remédier à cela, on peut préférer travailler sur les indices. Soit $x_0, x_1, \dots, x_t, \dots$ une série temporelle. On appelle indice (élémentaire) de la grandeur observée à la date t par rapport à la date 0, le rapport :

$$I_{t|0} = 100 \times \frac{x_t}{x_0}.$$

Calculer les indices de la série chronologique ci-dessus et les représenter graphiquement. Commenter.

Exercice n°5 Le tableau suivant indique la répartition des familles de l'île de La Réunion selon leur nombre d'enfants :

Nombre d'enfants	Nombre de familles
0	31038
1	54812
2	51252
3	26613
4 ou +	16162

1. Déterminer la population et la variable étudiée.
2. Préciser la nature et les modalités de la variable.
3. Représenter la distribution par diagramme circulaire.
4. à la suite de la question précédente :
 - a. Calculer les effectifs cumulés croissants et décroissants.
 - b. Représenter la fonction de répartition.
5. Combien de familles sont composées de :
 - a. au moins 1 enfant ?
 - b. au plus 2 enfants ?

Exercice n°6 On a relevé les groupes sanguins de vingt individus :

B AB A A O A O A AB A
A B O O A AB A O B A

1. Regrouper les données en utilisant les fonctions de votre tableur.
2. Déterminer le mode de la série statistique.
3. Effectuer deux représentations graphiques : diagramme circulaire (camembert) et diagramme en bâtons.

Exercice n°7 Dans le cadre d'un plan quinquennal d'inventaire aérien de gros gibier, le Service de la Faune du Québec a recensé les ravages d'originaux (*Alces alces*) d'une région du Parc de la Vérendrye. Lors de cet inventaire, 121 ravages furent dénombrés. Pour déterminer la sélection de l'habitat hivernal par cet ongulé, Proulx (1978) numérotait les 121 ravages et tira au sort 60 d'entre eux. Parmi les relevés effectués sur ces ravages, figurent la superficie relative des différents peuplements végétaux, l'orientation exprimée en degré, la présence ou l'absence de plans d'eau, la pente du ravin exprimée en pourcentage (dénivellation en mètres par 100 mètres parcourus sur le plan horizontal), etc. Les résultats relatifs à la pente sont les suivants (résultats en %) :

3.8	19.1	9.4	3.5	5.3	7.6	5.5	3.8	0.0	9.9
4.9	4.9	9.8	11.0	9.6	3.7	9.5	10.8	8.2	3.9
8.1	16.9	0.0	5.9	7.4	13.3	9.7	0.0	5.0	1.2
6.4	8.4	8.1	5.0	14.5	7.0	5.3	7.6	3.5	5.0
2.9	8.3	8.3	3.8	4.9	20.3	16.7	19.0	4.3	5.5
6.1	12.0	0.0	18.1	12.1	5.0	10.7	16.3	11.7	10.9

1. Saisir les données dans une plage et donner un nom à cette plage.
2. Calculer la moyenne, l'écart-type et la médiane des observations.
3. Déterminer le nombre de classes selon la règle de Sturge et selon la règle de Yule. Choisir une règle pour la suite.
4. Utiliser les fonctions appropriées de votre tableur pour effectuer de manière automatique un regroupement des données par classe de sorte que les amplitudes soient égales.
5. Calculer la moyenne, l'écart-type et la médiane des observations regroupées par classe.
6. Est-il possible d'utiliser les fonctions appropriées de votre tableur pour effectuer de manière automatique un regroupement des données par classe de sorte que les effectifs dans chaque classe soient identiques ? Si oui, calculer la moyenne, l'écart-type et la médiane des observations regroupées par classe. Comparer les différentes valeurs obtenues pour les trois caractéristiques numériques.

Exercice n°8 On considère un échantillon de $n = 150$ pièces usinées réparties selon leur diamètre en millimètres.

Classe	Centre	Amplitude	Eff.	Fréq.	Eff. cumulée	Fréq. cumulée	Fréq. cum. décr.
[19.70;19.80[2				
[19.80;19.85[10				
[19.85;19.90[14				
[19.90;19.95[22				
[19.95;20.00[32				
[20.00;20.05[27				
[20.05;20.10[26				
[20.10;20.15[9				
[20.15;20.20[3				
[20.20;20.30[5				
Total							

1. Compléter le tableau ci-dessus. Un calcul approché conduit à $\bar{x} = 20$ mm et à $\sigma_x = 0.10$ mm. Vérifier ces résultats.
Remarque : il est plus simple de créer une colonne pour la borne inférieure et une pour la borne supérieure des classes.
2. Tracer la fonction des fréquences cumulées et celui des fréquences cumulées décroissantes.
3. Situer graphiquement la médiane, puis la calculer.
4. Déterminer la proportion d'observations situées dans les deux intervalles suivants :

$$[\bar{x} - 2\sigma_x; \bar{x} + 2\sigma_x] \quad \text{et} \quad [\bar{x} - 3\sigma_x; \bar{x} + 3\sigma_x] .$$

Exercice n°9 Le traitement statistique des séries temporelles est très particulier. Comme premier outil possible, les moyennes mobiles sont des techniques assez simples d'emploi. Par exemple, la moyenne mobile arithmétique transforme une série temporelle x_0, x_1, \dots, x_n en une autre série temporelle y_q, \dots, y_{n-q} définie par :

$$\forall t \in \{q, \dots, n - q\}, \quad y_t = \frac{1}{2q + 1} \sum_{j=-q}^q x_{t+j},$$

où $q \in \mathbb{N}^*$ est un entier donné. Autrement dit, à chaque instant t , on effectue une moyenne sur les $2q + 1$ observations autour de t . Cette technique présente deux inconvénients : 1) elle tient compte à la fois du passé et de l'avenir ; 2) elle donne un poids identique à chacune des $2q + 1$ observations. C'est pourquoi la technique du lissage exponentiel a été introduite. Le lissage exponentiel simple est un outil de prévision basé sur les observations x_0, x_1, \dots, x_n . L'estimation $\hat{x}_n(k)$ de la valeur future x_{n+k} (inconnue) est donnée par :

$$\forall k \in \mathbb{N}^*, \quad \hat{x}_n(k) = (1 - \beta) \sum_{j=0}^n \beta^j x_{n-j},$$

où $\beta \in]0; 1[$ est appelée la constante de lissage. Si β est proche de 0, alors la prévision sera fortement influencée par les observations les plus récentes et on parle alors de prévision souple. En revanche, si β est proche de 1, alors la prévision prendra en compte des observations plus lointaines et on parle de prévision rigide. Dans un tel cas, la prévision est peu sensible aux fluctuations exceptionnelles (appelées aussi fluctuations conjoncturelles). On peut montrer que :

$$\hat{x}_n(k) = (1 - \beta)x_n + \beta\hat{x}_{n-1}(k) = \hat{x}_{n-1}(k) + (1 - \beta)(x_n - \hat{x}_{n-1}(k)) .$$

Partant de l'une de ces deux équations et en prenant $\hat{x}_1(k) = x_1$, on peut calculer pas à pas les valeurs lissées d'une série temporelle.

Pour les données contenues dans le fichier `EpidemieGrippe.csv` (nombre de cas de gripes déclarés par les médecins en Aquitaine entre la première semaine de 2004 et la neuvième semaine de 2007), représenter graphiquement les séries temporelles lissées par les techniques décrites ci-dessus.

Exercice n°10 Dans cet exercice, on va s'intéresser aux fonctionnalités d'un tableur permettant d'automatiser au maximum des tâches (dont certaines précédemment effectuées dans les exercices antérieures). L'énoncé qui suit a été fait pour Open Office mais tout ceci s'adapte à peu près sans problème à Microsoft Excel.

- Ouvrir le fichier `Etud-0708.xls` : ce fichier contient une colonne correspondant à la formation dans laquelle est inscrit les étudiants qui suivent ce cours.
- Sélectionner la plage de données (y compris le titre de la colonne).
- Aller dans le menu `Données`, puis dans le sous-menu `Pilote de données`, et cliquer sur `Démarrer`. Après avoir cliqué sur `Ok` à la première fenêtre, une seconde fenêtre s'ouvre. Dans celle-ci, faire glisser le titre de la colonne (dans le cas de cet exercice, "Formation") dans la partie `Ligne champs` et à nouveau dans la partie `Champs de données`. Si dans cette partie s'affiche `Somme` avant le titre de la colonne, il faut mettre `Nombre` en choisissant dans les options (bouton du haut). Regarder le résultat obtenu en bas du tableau (il est possible de faire afficher le résultat ailleurs en allant dans les options - bouton du bas).
- Changer la formation du dernier individu de la série. Que se passe-t-il ? Cliquer droit sur une zone du tableau créé automatiquement par le logiciel et cliquer sur `Actualiser`. Que se passe-t-il ?