

TD 1 & 2

1 Rappels informatiques Excel

Pour être efficace sous Excel (et avoir des bonnes notes), il faut absolument maîtriser les points suivants

1.1 Comment fait-on un calcul ?

Tout calcul commence par un “=”.

- Cliquez dans une cellule (celle dans laquelle le résultat final doit apparaître).
- Tapez **=**
- Tapez votre calcul.
- Terminez en appuyant sur **Entrée**.

Application : Calculez **1+1** avec Excel

1.2 Comment fait-on une racine ou une puissance ?

Vous voulez calculer 3^2 ou $\sqrt{5}$. Plusieurs méthodes

- Avec les formules : **=Puissance(3,2)** calcule 3^2 . **=Racine(5)** calcule $\sqrt{5}$.
- Avec les symboles : **^** est le symbole de la puissance **3^2** signifie 3^2 . Pour obtenir le caractère **^**, appuyer sur la touche **Alt Gr** et la touche **9/ζ**, puis sur la touche **Espace**

Application : Calculez $\sqrt{3}$ et 5^4

1.3 Comment fait-on un calcul compliqué avec les cellules ?

Tout l’avantage d’Excel, c’est que l’on peut se servir des résultats d’une cellule dans le calcul d’une autre. Pour calculer $\frac{2^3}{\sqrt{5}}$, on peut soit utiliser plusieurs degrés de parenthèses (mauvaise méthode parce qu’il est facile de faire une erreur) soit utiliser plusieurs cellules (bonne méthode beaucoup plus sûre) :

- Choisissez une cellule
- Tapez **=** (car vous allez commencer un calcul)
- Puis avec la souris, sélectionnez la cellule dans laquelle vous avez calculé 2^3
- Tapez ensuite l’opérateur **/**
- Puis sélectionnez la cellule dans laquelle vous avez calculé $\sqrt{5}$
- Terminez en appuyant sur **Entrée**.

Application : Calculez $\frac{4-1}{5 \times \sqrt{\frac{1}{10} + \frac{1}{8}}}$ SANS utiliser de parenthèses.

Dans 4 cases, inscrivez les chiffres 10 ; 8 ; 2,5 et 3,8. Calculez $\frac{(10-1) \times 2,5 + (8-1) \times 3,8}{10+8-2}$ toujours sans utiliser de parenthèses.

1.4 Comment fait-on une sélection ?

Pour faire une opération sur beaucoup de données, il faut sélectionner une plage (c’est à dire un ensemble de cellules). On peut le faire avec la souris. Mais ca n’est pas très pratique. Pourquoi ? Parce qu’avec la souris, la feuille de calcul se met à défiler très vite et il devient alors très difficile de s’arrêter exactement à la ligne 2783 (par exemple).

Comment faire ? Les touches du clavier permettent d'être plus efficace que la souris. La touche `Shift` (juste au dessus de `Ctrl`) permet de sélectionner. En maintenant `Shift`, vous pouvez régler la sélection avec les touches flèches et sauts de page.

Par ailleurs sous Excel, appuyer sur `Ctrl` et sur les flèches simultanément permet d'aller vers la prochaine cellule non vide jouxtant une cellule vide; si vous êtes au milieu d'un tableau, `Ctrl` + `←` permet d'aller sur la cellule la plus à gauche du tableau, `Ctrl` + `→` va sur la cellule la plus à droite...

La combinaison des deux `Shift`+`Ctrl` plus les flèches permet donc de sélectionner rapidement un grand nombre de donnée.

Application : Ouvrez le fichier `miniESPAD99.csv` et sélectionnez avec les touches du clavier les colonnes 2 et 3.

2 Statistiques, les bases (à la main et sous Excel)

2.1 Définitions

Un groupe de 8 étudiants passent un contrôle de statistique. Les notes sont les suivantes : 2, 15, 8, 11, 13, 13, 9, 18.

1. Quelle est la population considérée?
2. Quels sont les individus?
3. Quelle est, en plus de l'identifiant, la variables en jeu?
4. Quelles sont les modalités de cette variable?
5. Quelles est l'observation du sujet 2?
6. Donnez un exemple de modalité qui est également une observation.
7. Donnez un exemple de modalité qui n'est pas une observation.

2.2 Nature des variables

Dans chacun des cas, déterminez la nature de la variable.

1. Dans un questionnaire de satisfaction, la direction d'un parc d'attraction demande à ses clients leur opinion sur la climatisation des salles de cinéma. La `TEMPERATURE` est-elle Trop froide? Correcte? Trop chaude?
 - (a) Donnez deux exemples de modalités possibles.
 - (b) Vérifiez les propriétés mathématiques de la variable (Peut-on faire des additions? Peut-on l'ordonner?)
 - (c) En conclusion, déterminez la nature de la variable.
2. Dans un questionnaire de satisfaction, la direction d'un parc d'attraction demande à ses clients de noter la qualité de la journée qu'ils viennent de passer entre 0 et 10, 0 étant la pire journée de leur vie et 10 étant la meilleure. La variable considérée est `QUALITEJOURNEE`. Reprenez les trois points (a), (b) et (c) si dessus et déterminez le type de cette variable.
3. Le parc comporte une attraction "illusions visuelles". A la sortie, un expérimentateur demande aux visiteurs de préciser la couleur d'un objet. Il mesure la variable `COULEUR`. Reprenez les trois points (a), (b) et (c) si dessus et déterminez le type de cette variable.

2.3 Analyse univariée, exercice

1. Calculez à la main la moyenne, la variance, la variance corrigée, l'écart type et l'écart type corrigé de $\{1; 2; 3; 4; 5\}$.
2. Calculez les mêmes indices mais avec Excel (fonctions `Var.p()` ; `Ecartype()` ; `Ecartypep()` ; `Moyenne()` ; `Var()`)
3. Calculez les quartiles de $\{5, 10, 6, 5, 8, 5, 4, 1, 3\}$ (fonctions `Quartile()`). Calculez les adhérences. Tracez la boîte à moustache.

3 Initiation à R

R est un logiciel statistique à ligne de commande. Pour commencer, lancer-le en cliquant sur l'icône Rgui (sur le bureau).

3.1 Comment fait-on un calcul ?

Dans la fenêtre **R Console**, on tape le calcul (sans signe égal devant).

Application : Calculez `1+1`

3.2 Comment fait-on une puissance ou une racine ?

Le symbole de la puissance est `^`. Pour la racine, on utilise la fonction `sqrt()`.

Application : Calculez 3^2 et $\sqrt{4}$

3.3 Comment fait-on un calcul compliqué ?

R permet à l'utilisateur de "stocker" ses calculs en mémoires. Chaque case mémoire est désignée par une lettre. Pour mettre un calcul en mémoire, il faut utiliser une flèche : `5 -> A` signifie "Place la valeur 5 dans la case A". Le symbole `->` est composé du signe **moins** (touche `-`) et du signe **supérieure** (touche `>`). A noter, il est possible de faire des affectations dans les deux sens : `5 -> A` est identique à `A <- 5`.

Une fois que A contient une valeur, on peut l'utiliser dans tout type de calcul : tapez `A^2`.
Plus complexe, pour calculez $\frac{4-1}{5 \times \sqrt{\frac{1}{10} + \frac{1}{8}}}$:

- On stocke $4 - 1$ dans *Numerateur*
- On stocke $\frac{1}{10} + \frac{1}{8}$ dans *SousRacine*
- On stocke $5 \times \sqrt{\text{SousRacine}}$ dans *Denominateur*
- On divise *Numerateur* par *Denominateur*

Application : Calculez $\frac{(10-1) \times 2,5 + (8-1) \times 3,8}{10+8-2}$ sans utiliser de parenthèses.

3.4 Comment charger des données ?

Pour ouvrir un fichier de donnée sous R, il faut utiliser la fonction `read.csv2`.

- Créez un répertoire à votre nom dans "Mes documents".
- Enregistrez le fichier **CoupeDuComminge.csv** dans votre répertoire.
- Dans le menu "Fichier", sélectionnez "Changez le répertoire courant". Cliquez alors sur "Browse" et choisissez le nom de votre répertoire.
- Tapez `read.csv2("CoupeDuComminge.csv")`. Cette fonction lit le fichier et l'affiche.
- Pour pouvoir travailler sur les données, il faut les stocker en mémoire :
`read.csv2("CoupeDuComminge.csv") -> Donnees`. A priori, rien ne se passe. C'est simplement parce que R a chargé les données en mémoire. Pour les afficher, il faut maintenant taper `Donnees`.

3.5 Données complètes et colonne seule

On peut travailler sur l'ensemble des données

- Pour obtenir toutes les données, tapez `Donnees`.
- Pour obtenir les noms des colonnes, tapez `names(Donnees)`.
- Pour obtenir la nature des colonnes, tapez `str(Donnees)`. `factor` désigne une variable qualitative, `integer` et `numeric` désignent des quantitatives.

On peut également travailler sur une colonne particulière. Pour cela, on peut soit donner le numéro de la colonne : `Donnees[,2]` , soit ajouter \$ + le nom de la colonne à `Donnees` `Donnees$Points` .

- Par une méthode puis l'autre, affichez les valeurs de la variable `match`.

3.6 Exporter un graphique vers Word

R permet de **copier - coller** ses graphiques vers les autres logiciels.

- Sous **R**, tapez `barplot(c(1,3,2))`
- Avec le bouton droit de la souris, cliquez sur le graphique qui vient de s'ouvrir et choisissez "Copier comme bitmap".
- Ouvrez Word et faites un **Coller**.

4 Analyse univariée, cas facile (sous R)

4.1 Variables qualitatives

R fait assez peu de différences entre les nominales et les ordonnées.

- La fonction `summary(X)` permet d'obtenir les effectifs de X et le nombre de valeurs manquantes.
- `pie(table(X))` donne le camembert de la variable X .
- `barplot(table(X))` donne l'histogramme de la variable X .

Application : Dressez l'analyse univariée de la variable *Commentaire*.

4.2 Variables quantitatives

R fait également assez peu de différences entre les nominales et les ordonnées.

- Pour obtenir le tableau des effectifs, on utilise `table(Donnees(X))`
- La moyenne et les cinq quartiles sont donnés par `summary(X)`
- L'écart type est fournie par `sd(X)` (écart type corrigé). La variance est `var(X)` .
- `boxplot(X)` donne la boîte à moustache de la variable X .

Pour les histogrammes, les variables discrètes et continues sont traitées différemment :

- Pour une variable discrète, `barplot(table(X))` donne un histogramme complet.
- Pour une variable continue, `hist(X)` donne l'histogramme avec regroupement des classes.

Application : Dressez l'analyse univariée de la variable *Points*. Dressez l'analyse univariée de la variable *Match*.

5 Analyse univariée, cas réel (sous R ou sous Excel ou à la main, as you wish)

Note : Tout ce qui précède n'était qu'exercice, le vrai travail commence ici...

L'enquête ESPAD99¹ est un projet Européen visant à étudier la santé des jeunes et leur consommation de drogue (en collaboration avec l'OFDT². L'enquête a été menée auprès de 12000 élèves (de la classe de 4^e à la fin des études secondaires) sur lesquels 564 variables ont été mesurées.

L'équipe INSERM chargée de mener l'enquête en France a ajouté au questionnaire drogues deux questionnaires. Le premier porte sur la pratique sportive. Le second étudie la violence. L'idée est de voir s'il y a des liens entre la pratique sportive et la consommation de drogue et / ou les conduites violentes.

¹European School Survey Project on Alcohol and other Drugs

²Observatoire Français des Drogues et Toxicomanie.

Votre tâche est de partir sur les traces des chercheurs INSERM et d'essayer de trouver des liens qui existeraient entre la pratique sportive, la consommation de drogue et les conduites violentes. Comme vous le savez (parce que ça a été dit en amphi), toute analyse statistique commence par une analyse univariée. Les données ESPAD99 sont dans le fichier **miniESPAD99.csv**. Nous n'avons retenu que 10 des 564 variables. Votre tâche est maintenant de faire l'analyse univariée des 10 variables.

Cette analyse univariée doit être présentée sous Word. Pour chaque variable, vous devez :

- Définir son type
- En fonction de son type, présenter les effectifs, les indices de centralité et / ou de dispersion
- Présenter les représentations graphiques de la variable.
- Quand le cas se présente, dire si la variable suit une loi particulière.

Vous sauvegarderez ensuite votre travail sur le site ftp de l'université.