

R2.08 - Outils numériques pour les statistiques descriptives

TP2 - Premiers pas sur Excel

Tom Ferragut, Regis Fleurquin, Lucie Naert

Introduction

Excel est un outil essentiel dans le monde professionnel, simplifiant des tâches telles que la budgétisation et la planification, tout en facilitant l'analyse de données. Les compétences avancées en Excel, notamment la programmation en VBA (Visual Basic for Applications), ouvrent des opportunités dans la programmation et l'automatisation.

C'est pourquoi la maîtrise de cet outil est une compétence nécessaire à développer, pour répondre aux besoins du milieu professionnel actuel.

1 Un petit échantillon des données

Pour commencer, ouvrez le document "TP2 - Premiers pas sur Excel.xlsx" sur **Excel** ou **LibreOffice**. Ce document contient **deux feuilles** de calculs correspondant aux **deux parties** du TP. Dans cette première partie nous allons nous intéresser à un échantillon prélevé des données utilisées dans le TP1, positionnez-vous sur la première feuille "**1 - Échantillon**".

1.1 Recherche d'erreur

Les formules dans Excel sont des expressions mathématiques ou logiques qui effectuent des calculs sur les données d'une feuille de calcul. En utilisant des **références de cellules**, les formules permettent de créer des relations dynamiques entre les données. Par exemple, la formule " $=A1+B1$ " additionne le contenu des cellules A1 et B1. Les fonctions, telles que "SUM", "AVERAGE" peuvent simplifier des calculs complexes.

QUESTION 1. Certaines valeurs de la colonne "Global_Sales" sont fausses, déterminons lesquelles et automatisons cette détection.

- Dans la cellule **L1** écrire "Computed_Global_Sales", cela correspond à une nouvelle variable de nos données.
- Dans la cellule **L2** écrire " $=H2+I2+J2+K2$ " et appuyer sur entrée, puis étendre cette formule à toute la colonne grâce au **petit carré** apparaissant en bas à droite de la cellule **L2** quand elle est sélectionnée. Nous avons re-calculé la somme des ventes de chaque jeu.
- Dans la colonne **M**, que l'on appellera "Différence", calculer de la même façon pour chaque jeu, la différence entre les ventes de la colonne **G** et celles que l'on a calculé de la colonne **L**.

Nous pouvons tout de suite observer les jeux pour lesquels le nombre de vente globale a été mal calculé. Cependant ce n'est pas très visuel, nous allons améliorer cela en utilisant la **fonction conditionnelle** :

SI ("Test"; "Sortie si vrai"; "Sortie si faux"). (1)

QUESTION 2. Dans la colonne N, que l'on appellera "Vérification", écrire une fonction conditionnelle qui renvoie "OK" si la différence est nulle et "ERR" sinon.

Nous pouvons aller plus loin en changeant la couleur d'une cellule en fonction de son contenu.

QUESTION 3. Sélectionner la colonne N. Grâce à l'outil "Mise en forme conditionnelle", colorier en vert le fond des cellules contenant "OK" et en rouge le fond les cellules contenant "ERR".

1.2 Tris et Filtres

Les deux logiciels Excel et LibreOffice possèdent des outils de **filtres** et de **tris** déjà implémentés. Après avoir sélectionné la colonne qui vous intéresse, vous pouvez la trier par ordre alpha-numérique croissant ou décroissant. Attention à bien **étendre** la sélection quand demandé, pour ne pas trier uniquement cette colonne, et pour conserver la structure des données !

QUESTION 4. Grâce à la fonction tris par ordre croissant ou décroissant, donner les jeux ayant le plus de ventes a) aux États-Unis, b) en Europe et les jeux ayant le moins de ventes c) au Japon et d) dans le monde entier.

Après avoir sélectionné la colonne qui vous intéresse, vous pouvez filtrer certaines valeurs à afficher ou masquer.

QUESTION 5. Grâce à l'outil de filtre, réaliser les deux actions suivantes :

- Afficher uniquement les jeux dont les ventes globales ne sont pas correctes (Cf. Q 2).
- Afficher seulement les jeux d'un autre éditeur que Nintendo, combien y en a-t-il ?

1.3 Représentations graphiques

Pour présenter le contenu de nos données, nous allons utiliser des tableaux d'effectifs (ou tableaux agrégés). Pour cela, on utilise l'outil **tableau croisé dynamique** (ou table dynamique) qui permet de réaliser des tableaux reliant plusieurs variables ou des sous-tableaux associés à une seule variable.

Dans le cas de tableaux d'effectifs, attention à bien associer la méthode **comptage** ou **nombre** au paramètre du champs de données (Cela permet de compter le nombre d'occurrence de chaque modalité de la variable dont on souhaite obtenir les effectifs).

QUESTION 6. Représenter graphiquement, avec un diagramme adapté, la distribution :

1. des éditeurs.
2. des genres de jeux.
3. (**BONUS**) des genres de jeux publiés par Nintendo.
4. (**BONUS**) des plateformes de jeux avant 2006.

2 Travail sur l'ensemble des données

Nous étudions maintenant le même ensemble de données sur lequel nous avons travaillé lors du TP1. Ces données sont présentes sur la **deuxième feuille** du tableur, la première étape est de faire comprendre au logiciel que nous allons travailler sur celles-ci. Positionnez-vous sur la deuxième feuille "2 - Ensemble des données"

QUESTION 7. Transformer vos données en tableau de la façon suivante :

- **Sur Excel :** Insertion → Tableau (indiquer les bonnes extrémités du tableau)

- **Sur LibreOffice :**

1. Données → Définir la plage (indiquer les bonnes extrémités du tableau)
2. Données → Sélectionner la plage, puis activer l'AutoFiltre

En utilisant ce tableau ainsi que les notions vus dans la première partie, répondez aux questions suivantes.

QUESTION 8. *Quels sont les jeux dont les ventes globales ne correspondent pas aux ventes ? Comment expliquez-vous cela ?*

QUESTION 9. *Quelle est l'année de publication du premier jeu sorti sur PS2 ? sur GBA ?*

QUESTION 10. *Représenter graphiquement, avec le diagramme le plus adapté, la distribution :*

1. des éditeurs.
2. des genres de jeux.
3. (**BONUS**) des genres de jeux publiés par Nintendo.
4. (**BONUS**) des plateformes de jeux avant 2006.

Ces graphiques sont-ils semblables à ceux réalisés dans la première partie ?

QUESTION 11 (BONUS). *Déterminer le cinquième éditeur ayant réalisé le plus de ventes de jeux vidéos.*