

TD 06 – PRESENTATION DES DONNEES STATISTIQUES (2).

Réaliser les tables de fréquence pour les données groupées.

Utiliser la fonction FREQUENCE et l'option Histogramme de l'analyse de données.

Objectifs :

- I. **Réaliser la table de fréquence pour les données groupées**
- II. **Calculs :**
Fréquences absolues ou fréquences d'apparition(f_a)
Fréquences relatives(f_r)
Fréquences cumulées (f_c)
- III. **Présentation graphique des données groupées**
Construire, lire et interpréter un graphique
- I. **Réaliser la table de fréquence en utilisant la fonction FREQUENCE et l'option Histogramme des données groupées.**

Exigences :

1. Partant des données du tableau Table_TD05, réalisez les tableaux de fréquence pour les variables : glycémie, cholestérol, âge, poids et taille par le groupement des données par classes :

- Age (5 classes)
- Glycémie (5 classes),
- Cholestérol (5 classes),
- Poids (5 classes),
- Taille (5 classes),

1. Calculez pour ces variables les fréquences cumulées et les fréquences relatives.
2. Réalisez l'histogramme à base de tableaux de fréquence pour les variables ci-dessus.
3. Utilisez la fonction FREQUENCE pour réaliser la table de fréquence.
4. Utilisez l'option Histogramme de l'analyse de données pour réaliser la table de fréquence et l'histogramme.

<https://www.youtube.com/watch?v=asEuFvWGJDs>

Indications :

1. Dans le cas où la variable prise en étude a plus de 10 valeurs distinctes on recommande le groupement des valeurs en classes.

Prenons en considération la variable âge. Cette variable a 20 valeurs distinctes. L'obtention des valeurs distinctes peut se faire « manuellement », (copiez la colonne âge dans une colonne annexée, ordonnez de manière croissante, effacez les valeurs qui se répètent et ensuite regroupez-les) ou en utilisant les outils offerts par Excel 2010, *Grouper et Dissocier* du menu *Données*.

Après avoir obtenu la série de valeurs distinctes de la variable, suit la constitution du nombre des classes et le groupement des valeurs par classes.

Dans les exigences on précise les 5 classes pour la variable âge. Pour établir les classes on procède de la manière suivante :

- On fait la différence entre la valeur la plus grande et la valeur la plus réduite ($67-24=43$)
- Si on établit 5 classes, on divise l'écart (43) par le nombre de classes ($43/5=8,6$), et on établit par exemple la largeur de la classe valeur entière, par exemple 8.
- On établit les classes en partant de la valeur la plus réduite (24), en ajoutant la largeur de la classe (8). On obtient la première classe (intervalle) entre 24 et 32 (inclusivement 32 ou ≤ 32) etc.
- On établit ensuite les autres classes (≤ 32 , 32-40, 40-48, 48-56, 56-64, >64)

- On observe ainsi qu'il y a en fait 6 classes au lieu de 5 que nous nous sommes proposées, chose peu importante. Si l'on prenait une largeur de 9 classe, on obtiendrait 6 classes.

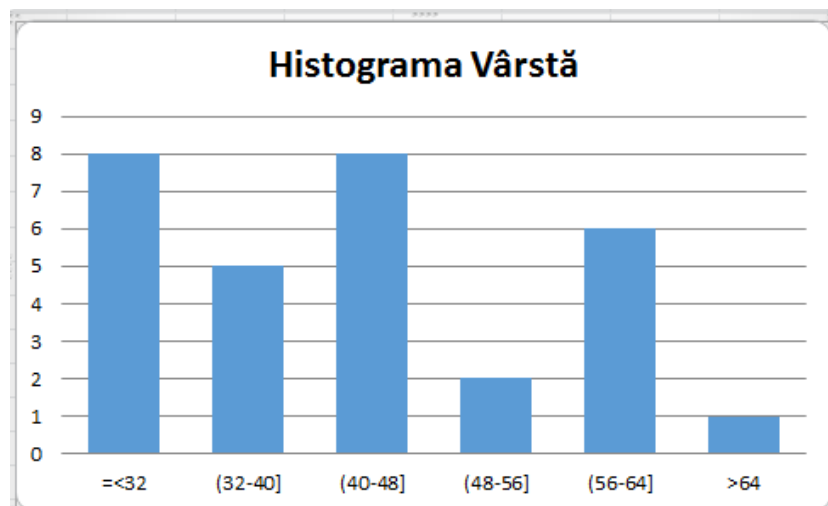
Continuez par réaliser la répartition de fréquence, c'est-à-dire établir combien de valeurs observées se retrouvent dans chaque classe ; on comptera le nombre des valeurs qui sont dans l'intervalle 24-32 (< ou égales à 32), puis celles de l'intervalle 32-40 (**intervalle ouvert à gauche et fermé à droite**) etc.

Le tableau de gauche peut être remplacé avec :

Clasa	f_a
24-32	8
32-40	5
40-48	8
48-56	2
56-64	6
64-72	1

Clasa	f_a
≤ 32	8
$(32-40]$	5
$(40-48]$	8
$(48-56]$	2
$(56-64]$	6
> 64	1

2. A la réalisation de l'histogramme dans le cas des données groupées, sur l'axe Ox apparaîtront les classes.



On observe que, dans l'histogramme Age, les colonnes sont séparées par gap (intervalles), à cause du fait que la variable âge est considérée comme variable discrète. Pour les autres variables, par exemple Poids, Taille, Cholestérol ou Glycémie, ces variables peuvent être considérées comme variables continues et, dans ce cas, on recommande que les barres ne soient pas séparées par gap.

3. L'utilisation de la fonction FREQUENCE pour réaliser la table de fréquence.

L'obtention de la répartition de fréquence peut se faire à l'aide de la fonction FREQUENCE de « Insérer une Fonction f_x » de la barre de formules. Cette fonction se trouve dans la catégorie des fonctions statistiques.

La fonction FREQUENCE calcule la fréquence de l'apparition de certaines valeurs dans un intervalle de valeurs, le résultat est une matrice verticale de nombres.

Cette fonction a 2 arguments obligatoires :

FREQUENCE (tableau_données; matrice_intervalles)

Tableau_données (Data_array) est une matrice ou une référence à un set de valeurs pour lesquelles on désire compter les fréquences. Ce premier argument de la fonction comprendra le domaine de la feuille de calcul avec toutes les valeurs observées.

Matrice_intervalles (Bin_array) est une matrice ou une référence aux zones où l'on désire grouper les valeurs de tableau_données. Ce deuxième argument de la fonction comprendra le domaine de la feuille de calcul où sont introduites les valeurs qui représentent la limite de droite des classes où les observations ont été distribuées. Par exemple, pour les valeurs d'au-dessus matrice_intervalles contiendra les valeurs :

Bin	Classe	f _a
32	=<32	8
40	(32-40]	5
48	(40-48]	8
56	(48-56]	2
64	(56-64]	6
	>64	1

	A	B	C	D	E	F
1	24	34	43	58	Bin	
2	25	34	44	60	32	8
3	26	35	45	62	40	5
4	26	35	45	64	48	8
5	27	39	46	64	56	2
6	27	41	49	64	64	6
7	30	41	49	67		1
8	31	41				

Observations

Dans la cellule F2 on a introduit la formule =FREQUENCE(A1:D8;E2:E6)

Domaine A1:D8 contient toutes les observations (30 observations) FREQUENCE ignore les cellules non-complétées et le texte.

Domaine E2:E6 contient les classes (Bin). Cet argument est obligatoire à l'utilisation de la fonction FREQUENCE.

L'obtention des valeurs de la colonne classes (Bin) se fait ainsi :

L'exigence du problème est :

Age (6 classes 24-32, 32-40, 40-48, 48-56, 56-64, 64-72)

Les classes fixées sont :

=<32; (32-40]; (40-48]; (48-56]; (56-64]; >64.

La colonne Classes (BIN) est complétée par des valeurs de droite de l'intervalle de chaque classe ;

bin Les 5 valeurs de la colonne Classes (BIN) définissent les 6 intervalles conformément aux classes fixées.

- On choisit la cellule de droite de la première valeur de la colonne BIN.

- On appelle la fonction FREQUENCE. Le premier argument représente des domaines de la feuille de calcul qui contiennent toutes les valeurs de la variable Age (dans ce cas le domaine A1:D8, (les valeurs peuvent apparaitre aussi sur une seule colonne)), le deuxième argument représente le domaine de la feuille de calcul qui contient les valeurs BIN. Attention ! On ne choisit que les valeurs numériques et non pas l'étiquette !

- Dans la cellule F2 apparaitra une première valeur de la table de fréquence. Pour obtenir les fréquences pour les autres classes on procède de la manière suivante : on

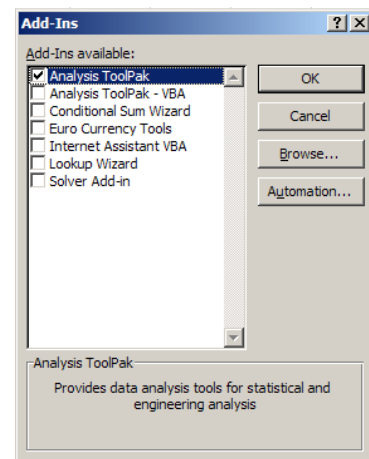
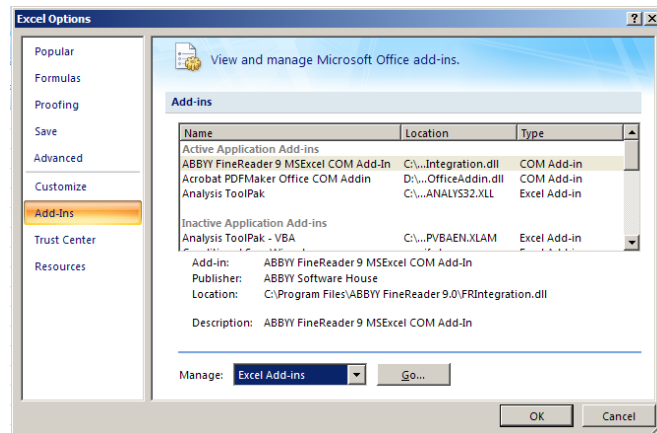
choisit le domaine F2:F7 (un de plus que la matrice Bin, obligatoirement adjacent au domaine Bin) et on tape : **CTRL+SHIFT+ENTRER**.

- Le nombre d'éléments de la matrice renversée est d'un plus grand que le nombre d'éléments de matrice_intervals.
- On obtient la matrice aux fréquences absolues.

3. Utilisez l'option Histogramme de Analyse des données (Data Analysis) pour réaliser la table de fréquence et l'histogramme

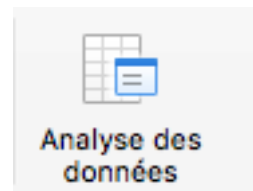
La réalisation de la répartition de fréquence peut se faire avec l'option Histogramme de Analyse de données (Menu Données –Office 2007, 2010, 2013). Dans le cas où elle n'existe pas il faudra l'installer. Manière de l'installer :

Accéder le bouton Office (situé à gauche-en haut de la fenêtre Excel) > Options > Complément excel > Add-Ins



On choisit de la liste MANAGE Excel Add-Ins et on tape le bouton Go...

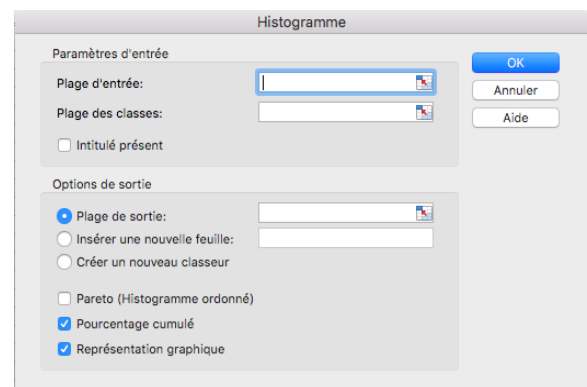
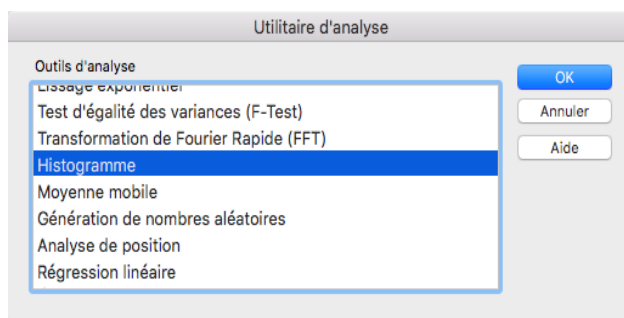
De la liste affichée dans la fenêtre nouvellement apparue on coche « Utilitaire d'analyse » (Analysis ToolPak) , on tape le bouton OK et....on attend qu'il s'installe. ☺



Ultérieurement, dans la section de menus Données on retrouvera l'option Analyse de données.

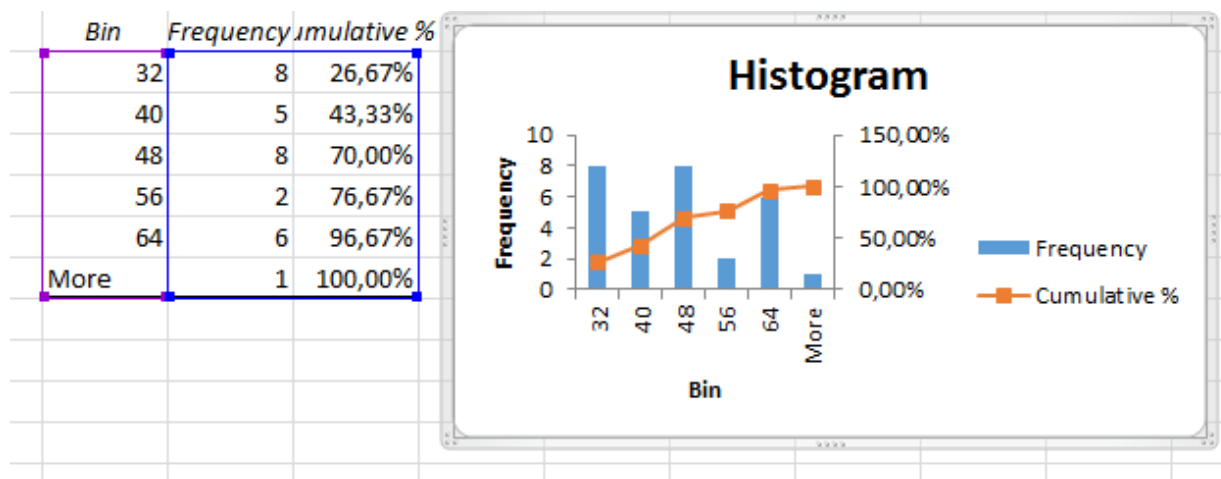
Si l'option **Analyse de données** est présente, on fait le pas suivant.

De la fenêtre de dialogue qui apparait, on choisit **Histogramme**. Cliquer **OK**.



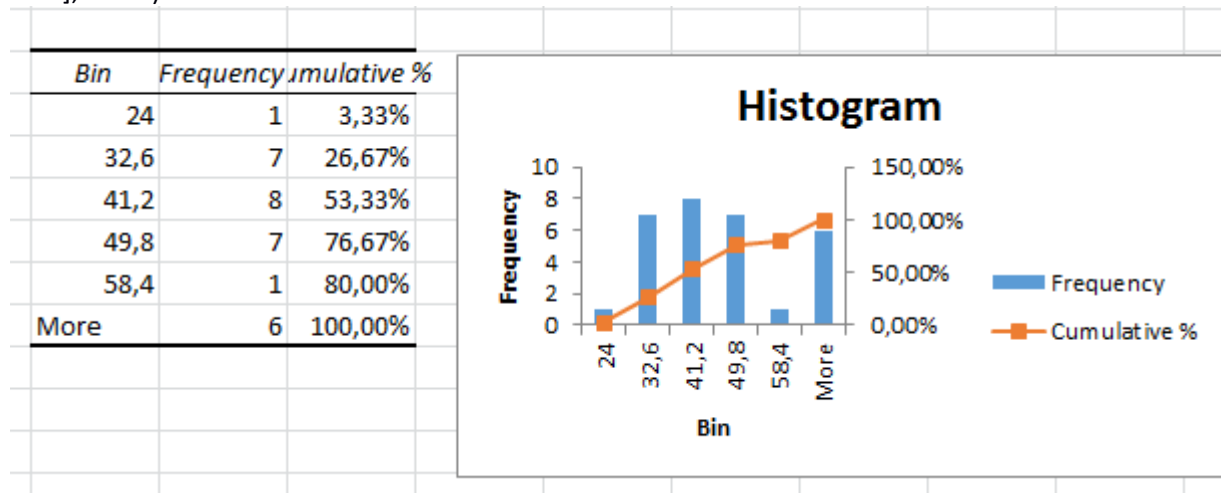
- Dans la fenêtre qui apparaît on va compléter à « Plage d'entrée » (Input Range) le domaine de la feuille de calcul qui contient toutes les valeurs de la variable pour laquelle on désire dresser la table de fréquence (pas forcément ordonnée).
- A « Plage des classes » (Bin Range) on précisera le domaine de la feuille de calcul qui contient les valeurs de la caractéristique x_i pour laquelle on va calculer les fréquences d'apparition de la même façon que pour la fonction FREQUENCE. (Dans le cas des données groupées, les valeurs de droite des intervalles des classes). Dans le cas où ce champ d'édition reste non-complété, l'application réalise un groupement des données selon d'autres critères.
- Si les domaines sélectionnés ci-dessus contiennent aussi des étiquettes on coche « Intitulé présent » (Labels)
- A « Plage de sortie » (Output Range) on précise la cellule où l'on désire faire apparaître le résultat (dans la même fiche de calcul).
- Il est nécessaire de cocher au moins une des 3 cases à cocher (check box) qui précise ce qu'on va calculer à cet appel de l'option.
- On calcule la fréquence d'apparition, la fréquence relative cumulée et la fréquence absolue cumulée. Cocher « Représentation graphique » (Chart Output) conduit à la construction de l'histogramme.

L'appellation de la fonction génère les résultats suivants :



Dans le cas où on ne sélectionne rien à Plage des classes (Bin range), on affiche la répartition de fréquence suivante :

On observe que Excel a établi d'autres classes (<24, (24, 32.6], (32.6, 41.2], (41.2, 49.8], (49.8, 58.4], >58.4)



On observe que l'application établit des valeurs pour lesquelles elle calcule la fréquence d'apparition. On interprète ainsi :

- Valeurs plus petites ou égales à 24,
- Ensuite valeurs plus grandes que 24 et plus petites que 32.6, etc.

Dans le cas où c'est nous qui établissons les valeurs pour lesquelles on calcule la fréquence de l'apparition, nous les mettons dans une colonne dans la fiche de calcul.



Obtenez les tableaux de fréquence (fréquences absolues) et calculez les fréquences relatives et cumulées pour la variable âge, Glycémie (5 classes), Cholestérol (5 classes), poids (5 classes) et taille (5 classes).

Résolvez les exercices ci-dessous en utilisant les fonctions Histogramme et/ou Fréquence.

Exercice 1



Les notes obtenues par 40 étudiants sont :

8; 10; 4; 9; 6; 8; 10; 7; 8; 3; 9; 6; 5; 4; 8; 7; 10; 9; 6; 5; 4; 3; 6; 9; 10; 8; 7; 7; 7; 6; 5; 5; 6; 7; 9; 10; 7; 6; 3; 4;

1. Dressez la répartition de fréquence.
2. Calculez les fréquences relatives.
3. Calculez les fréquences cumulées (d'apparition et relatives).
4. Dressez l'histogramme.

Exercice 2



Les valeurs suivantes représentent le jour d'incubation suite à un possible contact étiologique jusqu'à la manifestation d'une maladie infectieuse: 7, 3, 5, 9, 10, 6, 8, 4, 5, 3, 7, 6, 5, 4, 8, 8, 7, 10, 10, 3, 3, 5, 6, 7, 8.

Laquelle de ces valeurs correspond à **une fréquence relative égale** à 0,16 ?

Indication :

On réalise la répartition de fréquence.

On calcule les fréquences relatives.

Exercice 3



Les valeurs suivantes représentent le jour de l'incubation à la suite d'un possible contact étiologique jusqu'à la manifestation d'une maladie infectieuse :

7, 3, 5, 9, 10, 6, 8, 4, 5, 3, 7, 6, 5, 4, 8, 8, 7, 10, 10, 3, 3, 5, 6, 7, 8.

Laquelle de ces valeurs correspond à une **fréquence relative cumulée** égale à 68% ?

Indications :

On réalise la répartition de fréquence.

On calcule les fréquences relatives.

On calcule fréquences relatives cumulées.

Exercice 4



Les valeurs suivantes représentent le jour d'incubation à la suite d'un possible contact étiologique jusqu'à la manifestation d'une maladie infectieuse :

7, 3, 5, 9, 10, 6, 8, 4, 5, 3, 7, 6, 5, 4, 8, 8, 7, 10, 10, 3, 3, 5, 6, 7, 8.

Laquelle de ces valeurs correspond à une **fréquence absolue cumulée** égale à 10 ?

Indications :

On réalise la répartition de fréquence.

On calcule les fréquences absolues cumulées.

Exercice 5

Considérons la distribution suivante de fréquences des températures quotidiennes :

Classe de température (°C)	Fréquences (jours)
$(-5^{\circ}) - (-3^{\circ})$	14
$(-2^{\circ}) - 0^{\circ}$	27
$1^{\circ} - 3^{\circ}$	3
$4^{\circ} - 6^{\circ}$	5
$7^{\circ} - 9^{\circ}$	1
Total	50

En combien de jours la température a été plus élevée que (pas égale avec) le point de gel ?

- 41;
- 36;
- 27;
- 9;
- ne peut pas être déterminée à base d'informations disponibles.

Indications :

On calcule les fréquences absolues cumulées.

Exercice 6

Considérons la distribution suivante de fréquences des températures quotidiennes :

Classe de température (°C)	Fréquence (jours)
$(-5^{\circ}) - (-3^{\circ})$	14
$(-2^{\circ}) - 0^{\circ}$	27
$1^{\circ} - 3^{\circ}$	3
$4^{\circ} - 6^{\circ}$	5
$7^{\circ} - 9^{\circ}$	1
Total	50

Se basant sur les données de la répartition de fréquence ci-dessus, précisez en combien de jours la température a été inférieure (et non pas égale) au point de gel ?

- 9;
- 41;
- 27;
- 14;
- ne peut pas être déterminée à base des informations disponibles.

Indications :

On calcule les fréquences absolues cumulées.

Exercice 7

Le tableau suivant présente les niveaux de cholestérol de 50 poulets nourris avec de l'aluminium dans la ration:

107	138	134	90	138
105	105	129	128	105
109	102	135	115	104
111	99	126	114	110
120	109	128	111	111
127	111	104	112	121
98	128	98	110	127
115	125	88	99	128
114	123	95	89	130
118	102	105	105	135

1. Réalisez la répartition de fréquence.
2. Calculez les fréquences relatives.
3. Calculez les fréquences absolues cumulées.
4. Calculez les fréquences relatives cumulées.
5. Réalisez l'histogramme. Tracez sur le même graphique la courbe des fréquences cumulées.