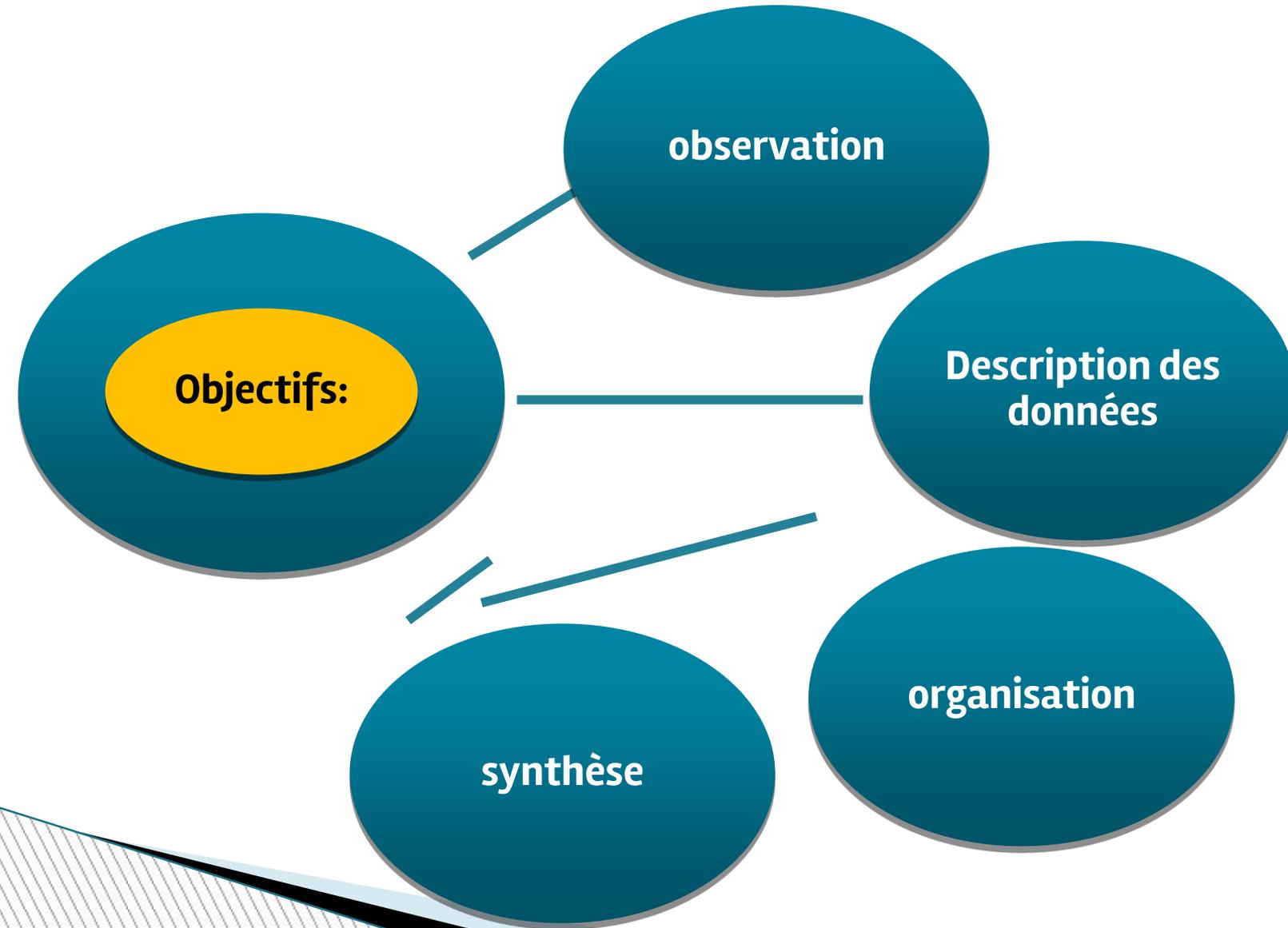


Branches de la statistique:

- **Statistique descriptive**
- **Statistique inférentielle**

Les statistiques descriptives sont des nombres utilisés pour décrire et résumer des données.

STATISTIQUE DESCRIPTIVE



L'observation représente un processus complexe d'identification, mesurage et d'enregistrement des phénomènes de manifestation.

Elle consiste dans le recueil de certaines données se basant sur des règles / critères unitaires chez les unités de la collectivité étudiée.

observation

direct



comptage



expérimentation



interview

indirect



Données enregistrées



Travaux publiés antérieurement

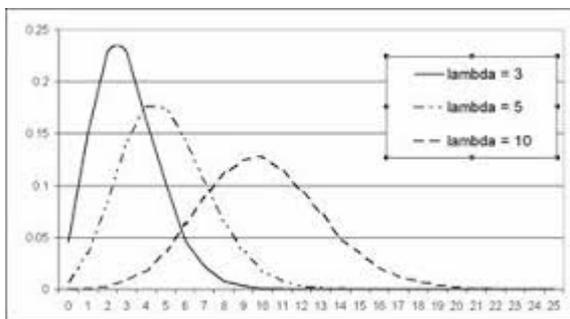
Résultats suite aux observations faites sur les individus d'une sélection; ils peuvent être présentés par:

**Tableaux de données
(tableaux des effectifs)**

Tarași economice în traazie	Economie agricolă (% din PIB)					
	1990	1991	1992	1993	1994	1995
Bulgaria	35,1	34,9	35,0	30,9	28,1	26,1
Republica Ceha	5,7	12,9	15,9	16,9	1,6	11,2
Ungaria	28,6	29,9	32,6	28,5	27,7	26,6
Polonia	19,6	21,5	19,7	18,1	15,2	12,8
România	15,7	11,7	15,7	16,1	17,1	5,1
Slovacia	7,2	12,1	11,5	10,2	14,9	2,8
Croația	22,8	22,9	24,7	22,7	22,8	24,2
Azerbaidjan	21,8	22,7	22,2	21,2	21,0	20,5
Belarus	15,1	16,5	12,2	11,7	10,9	12,2
Estonia	19,9	26,2	25,4	24,1	22,1	21,1
Georgia	21,5	26,0	23,3	21,9	21,7	21,2
Kazakhstan	11,0	15,7	24,9	21,2	24,1	24,2
Lituania	19,8	19,0	24,3	21,9	24,2	23,3
Letonia	11,3	21,6	22,2	21,3	22,2	21,8
Moldavia	18,1	21,1	21,3	21,0	22,7	22,1
Rusia	14,7	22,5	22,8	26,7	26,5	21,6
Turcia	18,7	22,8	22,8	22,9	21,7	22,2
Tadjikistan	11,1	12,8	11,7	10,1	9,1	6,5

**Organisation
des données**

**Représentations
graphiques**



**Tableaux de
distribution de
fréquence**

Table 1	Economie agricolă în România	
	Număr	Procent (%)
0 hectare	20	10%
1 hectare	40	20%
2 hectare	30	15%
3 hectare	20	10%
4 hectare	10	5%
5 hectare	5	2,5%
Total persoane	200	100%

Tableaux de données (tableaux des effectifs)

Nr.crt.	Age (ans)	sexe	Poids (kg)	Taille (cm)
1	6	F	15.70	102
2	9	F	12.80	108
3	11	M	16.10	101
4	16	F	18.5	100
5	8	M	16	106
6	5	M	14.5	104
7	10	F	15	103



- **Un arrangement rectangulaire de données où elles sont distribuées en lignes et colonnes.**
- **Chaque ligne et colonne doit porter une étiquette .**
- **Les lignes et les colonnes peuvent avoir des totaux qui devraient se trouver à la dernière ligne ou sur la dernière colonne.**

Tabelul nr. 2.1

Unitățile colectivității $i=1,n$	Caracteristici observate (X_j)			
	X_1	X_2	...	X_m
1	x_{11}	x_{21}	...	x_{1m}
2	x_{21}	x_{22}	...	x_{2m}
⋮	⋮	⋮	...	⋮
i	x_{i1}	x_{i2}	...	x_{im}
⋮	⋮	⋮	...	⋮
n	x_{n1}	x_{n2}	...	x_{nm}
Total*	$\sum x_{i1}$	$\sum x_{i2}$...	$\sum x_{im}$

Exemple: tableau

Nr. Grt.	x_1 (cm)	x_2 (cm)	f (cm)	f med.(cm)	delta f (cm)	delta f med. (cm)
1	29.2	11.2	8.1	8.584	0.484	0.3112
2	15	17	7.97	8.584	0.614	0.3112
3	17	17.7	8.67	8.584	0.086	0.3112
4	20.5	14.1	8.35	8.584	0.234	0.3112
5	18	15.6	8.36	8.584	0.224	0.3112
6	16.3	19.1	8.79	8.584	0.206	0.3112
7	19.1	17.6	9.16	8.584	0.576	0.3112
8	18.5	16.1	8.61	8.584	0.026	0.3112
9	15.5	21.1	8.54	8.584	0.356	0.3112
10	17.3	18.3	8.89	8.584	0.306	0.3112



Intervale de variație (x_i)	Încadrarea muncitorilor pe intervale de câștiguri salariale	Numărul muncitorilor (n_i)
2,2 - 3,0		5
3,0 - 3,8		7
3,8 - 4,6		9
4,6 - 5,4		12
5,4 - 6,2		8
6,2 - 7,0		6
7,0 - 7,8		3
Total		$\Sigma n_i = 50$

Un **tableau statistique** est plus expressif qu'une masse systématisée de données, à condition que, dans sa construction on respecte quelques règles fondamentales :

- **avoir un titre clair et concis qui** suggère la nature des données présentées, le temps et l'espace auxquels se rapportent les données comprises dans le tableau;
- **indiquer l'unité de mesure.** Si elle est commune pour toutes les données présentées dans le tableau, elle peut être mentionnée dans le titre général du tableau. Si elle n'est pas commune, celle-ci doit être indiquée à chaque cas (dans les titres intérieurs);
- **Mentionner la source des données** (au-dessous du tableau);
- être construit de telle manière qu'il puisse être compris sans explications supplémentaires qui précèdent ou suivent le tableau statistique;
- **Toutes les rubriques du tableau sont complétées avec des chiffres ou des symboles.**
- **Le format des tableaux doit être suggestif**



Répartition de fréquence (Tableau de fréquence)



est une manière de résumer un set de données.



est une évidence qui montre combien de fois chaque valeur (ou un set de valeurs) de la variable en cause apparaît dans l'expérimentation.

Répartitions (distributions) de fréquence

x_i	n_i
x_1	n_1
x_2	n_2
.....
x_k	n_k

x_i - caractéristique observée

n_i - fréquence de l'apparition de la caractéristique x_i , - **fréquence absolue**

$$n_1 + n_2 + \dots + n_k = \mathbf{n}$$

\mathbf{n} est le nombre total des observations

Types de fréquences

► **Fréquence absolue**

La fréquence absolue d'une valeur x d'une série statistique S est le nombre de répétition de la valeur x dans la série S . Donc, la somme des fréquences absolues de toutes les valeurs distinctes d'une série statistique est égalé à la taille de la série.

► **Fréquence cumulée**

► La fréquence absolue cumulée d'une valeur x d'une série statistique S est la somme des fréquences absolues des valeurs de la série plus petites ou égales à x .

► **Fréquence relative**

La fréquence relative d'une valeur x d'une série statistique S est le rapport entre la fréquence absolue de la valeur x et la taille de la série. D'habitude, la fréquence relative est présentée en pourcentage.

► **Fréquence relative cumulée**

La fréquence relative cumulée croissante d'une valeur x d'une série statistique S est le rapport entre la fréquence absolue cumulée croissante de la valeur x et la taille de la série.

► **Fréquence relative en pourcentage**

► **Fréquence relative en pourcentage cumulée**

Comment réalise-t-on une répartition de fréquence?

On suit ces étapes:

1:

On dresse un tableau avec trois colonnes. La première colonne contient toutes les valeurs des données (éventuellement en ordre croissante).

2:

On parcourt la première colonne et on marque avec une ligne chaque apparition de chaque valeur distincte de la première colonne. Lorsqu'on arrive à la cinquième ligne, celle-ci barre les quatre premières. On continuera ce processus jusqu'à ce que toutes les valeurs de données de la liste soient comptées.

3:

On compte les groupes de cinq lignes de la deuxième colonne.

La réalisation d'un tableau à distribution de fréquences

A la suite d'un sondage fait dans un village, on a demandé à 20 fermes le nombre de vaches qu'ils possédaient. Les résultats ont été enregistrés de la manière suivante :

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0

Ordre croissant : 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4

Nombre de vaches (x)	compte	Fréquence(f)
0	<u>IIII</u>	4
1	<u>IIII</u> I	6
2	IIII	5
3	III	3
4	II	2

Comment construire une distribution de fréquence?

Nombre de cochonnets à la II-ème mise bas chez un échantillon de 25 truies:

5, 2, 10, 7, 6, 10, 9, 8, 9, 8, 6, 4, 6, 4, 10, 8, 6, 8, 8, 7, 6, 7, 6, 6, 4, 4

1. **Tableau d'évidence** primaire ou **tableau des effectifs**

Les valeurs se mettent en ordre croissante ou décroissante (croissante):

2. 10, 10, 9, 9, 8, 8, 8, 8, 8, 7, 7, 7, 6, 6, 6, 6, 6, 6, 5, 4, 4, 4, 4, 2

3. On dresse un tableau à deux colans (n_i , f_i)

n	f _a
10	2
9	2
8	5
7	3
6	7
5	1
4	4
3	0
2	1

Fréquence absolue

Valeur	f_a
10	2
9	2
8	5
7	3
6	7
5	1
4	4
3	0
2	1
Total	$\Sigma f_a = 25$

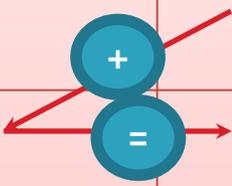
La fréquence absolue d'une valeur x d'une série statistique S est le nombre de répétitions de la valeur x dans la série S .
Donc, la somme des fréquences absolues de toutes les valeurs distinctes dans une série statistique est égale à la taille de la série.

f_a - **Fréquence absolue**

10, 10 9, 9, 8, 8, 8, 8, 8 7, 7, 7, 6, 6, 6, 6, 6, 6,
6, 5, 4, 4, 4, 4, 2

Fréquence cumulée

Valeur	f_a	f_c
10	2	2
9	2	4
8	5	9
7	3	12
6	7	19
5	1	20
4	4	24
3	0	24
2	1	25
Total	$\Sigma f_a = 25$	



La fréquence absolue cumulée croissante d'une valeur x dans une série statistique S est la somme des fréquences absolues des valeurs de la série plus petites ou égale à x .



Fréquence relative

Valeur	f_a	f_c	f_r
10	2	2	0,08
9	2	4	0,08
8	5	9	0,2
7	3	12	0,12
6	7	19	0,28
5	1	20	0,04
4	4	24	0,16
3	0	24	0
2	1	25	0,04
Total	$\Sigma f_a=25$		1

La fréquence relative d'une valeur x dans une série statistique S est le rapport entre la fréquence absolue de la valeur x et la taille de la série. D'habitude, la fréquence relative est présentée en pourcentage.

$$f_r = \frac{f_a}{\Sigma f_a}$$

$$f_{r1} = \frac{2}{25} = 0,08$$

Fréquence relative cumulée

Valeur	f_a	fc	fr	frc
10	2	2	0,08	0,08
9	2	4	0,08	0,16
8	5	9	0,2	0,36
7	3	12	0,12	0,48
6	7	19	0,28	0,76
5	1	20	0,04	0,80
4	4	24	0,16	0,96
3	0	24	0	0,96
2	1	25	0,04	1
Total	$\Sigma f_a = 25$		$\Sigma fr = 1$	

La fréquence relative cumulée croissante d'une valeur x dans une série statistique S est le rapport entre la fréquence absolue cumulée croissante de la valeur x et la taille de la série.

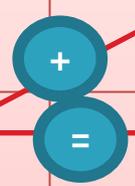


Fréquence relative en pourcentage

Valeur	f_a	f_c	f_r	f_{rc}	$f_{r\%}$
10	2	2	0,08	0,08	8%
9	2	4	0,08	0,16	8%
8	5	9	0,2	0,36	20%
7	3	12	0,12	0,48	12%
6	7	19	0,28	0,76	28%
5	1	20	0,04	0,80	4%
4	4	24	0,16	0,96	16%
3	0	24	0	0,96	0%
2	1	25	0,04	1	4%
Total	$\Sigma f_a=25$		$\Sigma f_r=1$		$\Sigma f_{r\%}=100$

Fréquence relative en pourcentage cumulée

Valeur	f_a	f_c	f_r	f_{rc}	$f_{r\%}$	$f_{r\%c}$
10	2	2	0,08	0,08	8%	8%
9	2	4	0,08	0,16	8%	16%
8	5	9	0,2	0,36	20%	36%
7	3	12	0,12	0,48	12%	48%
6	7	19	0,28	0,76	28%	76%
5	1	20	0,04	0,80	4%	80%
4	4	24	0,16	0,96	16%	96%
3	0	24	0	0,96	0%	96%
2	1	25	0,04	1	4%	100%
Total	$\Sigma f_a=25$		$\Sigma f_r=1$		$\Sigma f_{r\%}=100$	



Tableaux de fréquence - résumé

- On met en ordre les données de façon croissante
 - On détermine la fréquence de chaque valeur
 - On inclus les valeurs distinctes et les fréquences dans un tableau sur deux colonnes
 - On peut dresser des tableaux à plusieurs colonnes, en ajoutant à droite:
 - La fréquence cumulée
 - La fréquence relative
 - La fréquence relative cumulée
 - La fréquence relative en pourcentage
 - La fréquence relative en pourcentage cumulée
- 

Analyse de fréquences groupées

EX: Analysons le pourcentage de graisse dans le lait sur 6850 vaches. Distribution triée de façon croissante.

41 entrées dans la table de fréquence

xi	ni	xi	ni	xi	ni
2,5	1	3,2	233	3,9	198
2,55	1	3,25	337	3,95	104
2,6	1	3,3	373	4	85
2,65	1	3,35	451	4,05	25
2,7	3	3,4	497	4,1	20
2,75	5	3,45	534	4,15	18
2,8	9	3,5	547	4,2	16
2,85	11	3,55	562	4,25	10
2,9	19	3,6	571	4,3	3
2,95	34	3,65	462	4,35	2
3	79	3,7	425	4,4	2
3,05	89	3,75	317	4,45	2
3,1	114	3,8	260	4,5	1
3,15	170	3,85	258	total	6850

Groupement par classes (catégories)

- ▶ **On choisit un nombre de classes, catégories**
 - ❑ On recommande, entre 5 et 15 (conventionnellement)
- ▶ **On choisit la dimension de l'intervalle de classe:**
 - ❑ Principes
 - Tous les intervalles doivent être égaux
 - Les limites des intervalles doivent comprendre toutes les valeurs (entre les limites des intervalles voisins pas de "vides" ou de superpositions)
 - ❑ On applique la procédure suivante ...

Réalisation des classes



On fait la différence entre la valeur la plus grande et la valeur la plus réduite

$$4,5 - 2,5 = 2$$



On divise la valeur obtenue par la dimension possible de l'intervalle de classe (0.2, 0.4, 0.5, ou 1 ou 2) pour réaliser le nombre de classes de la nouvelle distribution

$$2 / 0,05 = 40 \text{ classes (trop)}$$
$$2 / 0,1 = 20 \text{ classes (variante possible)}$$
$$2 / 0,2 = 10 \text{ classes (variante acceptée)}$$



On sélectionne la dimension de l'intervalle qui conduit à un nombre de classes compris entre 5 et 15

On choisira 0,2 parce qu'il produit une distribution à 10 classes qui est plus facile à analyser et à manipuler .



On détermine la limite inférieure du premier intervalle (il doit être un multiple de la dimension de l'intervalle)

On choisit la valeur 2,5 en tant que limite inférieure



On détermine la limite supérieure du premier intervalle

Si la dimension de l'intervalle est 0.2, la limite supérieure sera $2,5 + 0,2 = 2,7$ [2,5, 2,5, 2,6, 2.65, 2,7)



On construit les intervalles de classe pour chaque intervalle

Tableau des fréquences groupées

Limites de la classe	Centre de la classe	Fréquence absolue	Fréquence cumulée	Fréquence relative	Fréquence relative cumulée	Fréquence relative en pourcentage	Fréquence relative en pourcentage cumulée
2,5 - 2,69	2,6	4	4	0,0006	0,0006	0,06%	0,06%
2,7 - 2,89	2,8	28	32	0,0041	0,0047	0,41%	0,47%
2,9 - 3,09	3	221	253	0,0323	0,0369	3,23%	3,69%
3,10 - 3,29	3,2	854	1107	0,1247	0,1616	12,47%	16,16%
3,30 - 3,49	3,4	1855	2962	0,2708	0,4324	27,08%	43,24%
3,5 - 3,69	3,6	2142	5104	0,3127	0,7451	31,27%	74,51%
3,70 - 3,89	3,8	1260	6364	0,1839	0,9291	18,39%	92,91%
3,90 - 4,09	4	412	6776	0,0601	0,9892	6,01%	98,92%
4,10 - 4,29	4,2	64	6840	0,0093	0,9985	0,93%	99,85%
4,30 - 4,5	4,4	10	6850	0,0015	1,0000	0,15%	100,00%
TOTAL		6850		1,0000		100,00%	

Analysant la hauteur au garrot chez 66 vaches, on a obtenu le tableau d'effectifs qui suit :

x_i	n_i	x_i	n_i	x_i	n_i
116	1	122	4	127	3
116,5	1	123	5	128	1
117	1	123,5	6	128,5	1
118	1	124	4	129	2
119	2	124,5	4	129,5	2
119,5	1	125	3	130	1
120	2	125,5	4	131	1
120,5	2	126	2	133	1
121	2	126,5	3	136	1
121,5	2		3	total	66

29 entrées
 $139 - 116 = 40$
 7 classes
 Largeur de la classe = 3

Tableau des fréquences groupées

Limites de la classe	fa	fr	fr%
[116-119)	4	0,061	6%
[119-122)	11	0,167	17%
[122-125)	26	0,394	39%
[125-128)	15	0,227	23%
[128-131)	7	0,106	11%
[131-134)	2	0,030	3%
[134-137)	1	0,015	2%
TOTAL	66	1	100%

Comment fait-on la répartition de fréquence à l'aide des fonctions de EXCEL



Dans EXCEL: Fonction: FREQUENCE

f(x)

Function Arguments

FREQUENCY

Data_array D3:D40 = {2,5;2,55;2,6;2,65;2,7;2,75;2,8;2,85;2,9}

Bins_array L4:L14 = {2,5;2,7;2,9;3,1;3,3;3,5;3,7;3,9;4,1}

= {1;4;4;3;3;4;4;4;4;3;0}

Calculates how often values occur within a range of values and then returns a vertical array of numbers having one more element than Bins_array.

Bins_array is an array of or reference to intervals into which you want to group the values in data_array.

Formula result = 1

[Help on this function](#)

OK Cancel

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

- ▶ **Matrice de données (Data_array)** est une matrice ou une référence à un set de valeurs pour lesquelles vous désirez compter les fréquences. Si data_array ne contient aucune valeur, FREQUENCY retourne une matrice de zéros.
- ▶ **Bins_array** est une matrice ou une référence aux **intervalles dans lesquels vous désirez grouper les valeurs du data_array**. Si bins_array ne contient aucune valeur, FREQUENCY retourne le nombre d'éléments du data_array.

Réponse de la matrice



fx		={FREQUENCY(B1:C13;E2:E11)}				
B	C	D	E	F	G	
10	6			1	0	
10	6			2	1	
9	6			3	0	
9	6			4	4	
8	6			5	1	
8	6			6	7	
8	5			7	3	
8	4			8	5	
7	4			9	2	
7	4			10	2	
7	4				0	
	2					

- ▶ La réponse de la fonction FREQUENCE est une matrice.
- ▶ Le nombre d'éléments de la matrice renversée est d'un plus grand que le nombre d'éléments de bins_array.
- ▶ FREQUENCE ignore les cellules non-complétées et le texte.

Comment procéder?

1. On sélectionne le domaine de la feuille de calcul où on obtiendra la réponse de la matrice.
2. On édite la fonction FREQUENCE
3. Appuyez F2, ensuite CTRL+SHIFT+ENTER.

116	121	122,5	123,5	125	127
116,5	121	122,5	124	125,5	128
117	121,5	123	124	125,5	128,5
118	121,5	123	124	126	129
119	122	123	124	126	129
119	122	123	124,5	126	129,5
119,5	122	123	124,5	126,5	129,5
120	122	123	124,5	126,5	130
120	122,5	123,5	125	126,5	131
120,5	122,5	123,5	125	127	133
120,5	122,5	123,5	125	127	136

	BIN	fréquence
<=116	116	1
<=119	119	5
<=122	122	13
<=125	125	26
<=128	128	12
<=131	131	7
<=134	134	1
>=134		1

=FREQUENCE(B33:B39;G33:G39)

Fonction HISTOGRAMME

Analyse de données



Data Analysis

Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram**

Buttons: OK, Cancel, Help



Bin	Frequency
1	0
2	1
3	0
4	4
5	1
6	7
7	3
8	5
9	2
10	2
More	0

10	6		
10	6		
9	6		
9	6		
8	6		
8	6		
8	6		
8	6		
8	5		
8	4		
7	4		
7			
7			



Histogram

Input

Input Range:

Bin Range:

Labels

Output options

Output Range:

New Worksheet Ply:

New Workbook

Pareto (sorted histogram)

Cumulative Percentage

Chart Output

Buttons: OK, Cancel, Help