

Master Santé -S1- Biostatistique

Travaux pratiques (TP1)

Martin LARSEN

Plan d'action

Le TP sera divisé en trois sections.

1. Distribution des données, statistiques descriptives et visualisation des données
2. Intervalle de confiance et Inférence statistique
3. Inférence statistique et corrélation.

Excel - prise en main

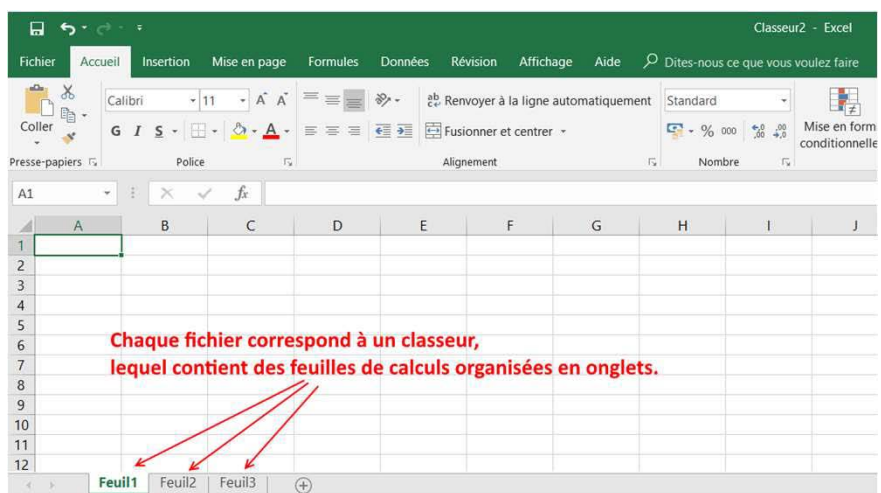
Excel c'est quoi ?

Le logiciel Excel intègre des fonctions :

1. de calcul numérique (à l'aide de formules de calcul) (TP1)
2. de représentation graphique (TP1)
3. d'analyse de données (notamment de tableau croisé dynamique) (TP1-3)
4. et de programmation/macros (Visual Basic for Applications) – **HORS CURSUS**

Excel - prise en main

Un fichier Excel c'est quoi ?



Excel - prise en main

Un tableau Excel c'est quoi ?

Colonne : A, B, C ...G ...

Lignes : 1, 2, 3, 4,16, 17

Mois	CD musicaux	CD jeux	Livres	Magazines	Total par mois
Janvier	1 900	6 520	2 751	514	11 685
Février	1 888	4 522	2 987	687	10 085
Mars	1 457	4 523	1 237	125	7 342
Avril	1 255	9 563	1 588	385	12 790
Mai	957	7 521	2 430	873	11 781
Juin	1 250	6 385	2 100	268	10 004
Juillet	1 278	2 341	1 251	487	5 356
Août	958	4 588	1 955	453	7 953
Septembre	1 752	2 384	1 235	365	5 737
Octobre	1 521	5 367	2 840	569	10 298
Novembre	245	7 851	2 462	235	10 793
Décembre	1 238	4 853	3 100	751	9 943
Total par produit	15 697	66 420	25 936	5 714	113 767

Excel - prise en main

Un tableau Excel c'est quoi ?

Zone de nom

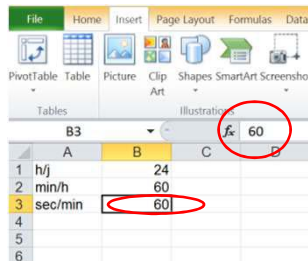
B4

La cellule B4 est la case située à l'intersection de la colonne B et de la ligne 4. Elle contient la valeur : 1900

Mois	CD musicaux	CD jeux	Livres	Magazines	Total par mois
Janvier	1 900				1 900
Février					0
Mars					0
Avril					0
Mai					0
Juin					0
Juillet					0
Août					0
Septembre					0
Octobre					0
Novembre					0
Décembre					0
Total par produit	1 900	0	0	0	1 900

Excel - prise en main

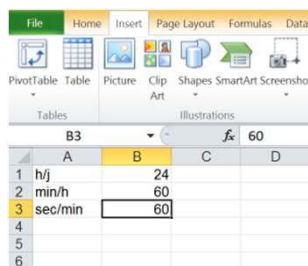
Calcul numérique en Excel



	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4				
5				
6				

Excel - prise en main

Calcul numérique en Excel



	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4				
5				
6				

Formule:

Initié par =

Les opérateurs arithmétiques:

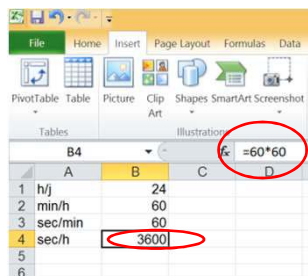
+ - * / ^

Les opérateurs relationnels:

=, <>, <, >, <=, >=

Les opérateurs logiques:

ET, OU, NON (des fonctions en Excel)



	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5				
6				

Excel - prise en main

Calcul numérique en Excel

Formule:
Initié par =
Les opérateurs
arithmétiques:
+ - * / ^

Table 1:

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4				
5				
6				

Table 2 (Formula: =B1*B2*B3):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				

Table 3 (Formula: =B2^2):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				
7				

Ou :
=B2*B3

Excel - prise en main

Calcul numérique en Excel

Formule:
Initié par =
Les opérateurs
arithmétiques:
+ - * / ^
Les opérateurs relationnels:
=, <, >, <=, >=
Les opérateurs logiques:
ET, OU, NON (des
fonctions en Excel)

Table 1 (Formula: =B5>B4):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				
7				

Table 2 (Formula: =AND(TRUE, FALSE)):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				
7				

Table 3 (Formula: =OR(TRUE, FALSE)):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				
7				

Table 4 (Formula: =NOT(TRUE)):

	A	B	C	D
1	h/j	24		
2	min/h	60		
3	sec/min	60		
4	sec/h	3600		
5	sec/j	86400		
6				
7				

Plus grand :
>

Pas égal à:
<>

Excel - prise en main

Des fonctions en Excel

Plage de cellules : Reference 1 : Reference 2 (e.g. A1:A3)
 Decaler(Reference 1, ligne, colonne) (e.g. A3 = Decaler(A1;2;0))

Statistique descriptive:

Moynne(Plage)
 Mediane(Plage)
 Mode(Plage)
 Min(Plage)
 Centile(Plage;0,25) (Q1)
 Centile(Plage;0,75) (Q3)
 Max(Plage)
 Ecartype(Plage)=s
 Var(Plage)=s² (echantillon)

$$s^2 = \frac{n}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Attention!
 Var.P(Plage)=σ²(population)

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Pour info!
 Racine(4) = √4 = 2

Excel - Somme de colonne

Démographie de la population française (Tableau 1):

	A	B	C	D	E
	Year of birth	Age in complete d years	Males	Females	Total
1					
2	2018	0	364,155	347,749	711,904
3	2017	1	370,453	355,472	725,925
4	2016	2	378,518	363,162	741,680
5	2015	3	387,906	372,402	760,308
6	2014	4	399,232	387,042	786,274
7	2013	5	407,611	389,920	797,531
8	2012	6	417,471	396,835	814,306
9	2011	7	418,623	403,349	821,972
10	2010	8	429,919	412,555	842,474
11	2009	9	427,917	408,232	836,149
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					
90					
91					
92	1928	90	55,382	124,322	179,704
93	1927	91	44,797	105,456	150,253
94	1926	92	34,519	91,072	125,591
95	1925	93	27,317	76,447	103,764
96	1924	94	20,525	61,235	81,760
97	1923	95	14,477	48,398	62,875
98	1922	96	10,101	37,882	47,983
99	1921	97	7,239	27,754	34,993
100	1920	98	4,977	19,813	24,790
101	1919	99	2,058	8,273	10,331
102	1918 or before	100 or over	2,976	12,670	15,646
103		Total	32,394,531	34,598,168	66,992,699

- Calculer les sommes des colonnes C, D et E
 Fonction: Somme(Plage)

Source: <https://www.insee.fr/en/statistiques/2382597?sommaire=2382613>

Excel - Diagramme en bâtons (superposés)

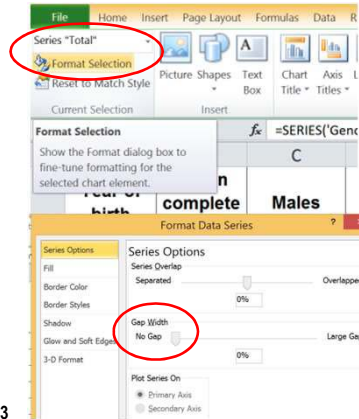
Démographie de la population française (Tableau 1):

	A	B	C	D	E
	Year of birth	Age in complete d years	Males	Females	Total
1					
2	2018	0	364,155	347,749	711,904
3	2017	1	370,453	355,472	725,925
4	2016	2	378,518	363,162	741,680
5	2015	3	387,906	372,402	760,308
6	2014	4	399,232	387,042	786,274
7	2013	5	407,611	389,920	797,531
8	2012	6	417,471	396,835	814,306
9	2011	7	418,623	403,349	821,972
10	2010	8	429,919	412,555	842,474
11	2009	9	427,917	408,232	836,149
92	1928	90	55,382	124,322	179,704
93	1927	91	44,797	105,456	150,253
94	1926	92	34,519	91,072	125,591
95	1925	93	27,317	76,447	103,764
96	1924	94	20,525	61,235	81,760
97	1923	95	14,477	48,398	62,875
98	1922	96	10,101	37,882	47,983
99	1921	97	7,239	27,754	34,993
100	1920	98	4,977	19,813	24,790
101	1919	99	2,058	8,273	10,331
102	1918 or before	100 or over	2,976	12,670	15,646
103		Total	32,394,531	34,598,168	66,992,699

Source: <https://www.insee.fr/en/statistiques/2382597?sommaire=2382613>



1. Affichez le nombre d'individus pour chaque groupe d'âge sous forme de diagramme en bâtons.
2. Formater les données - supprimer l'écart entre des bâtons.



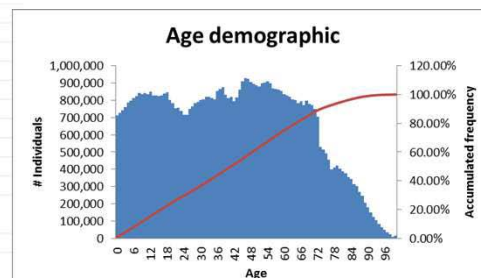
Excel - Diagramme en bâtons superposés

Démographie de la population française (Tableau 1):

	A	B	C	D	E
	Year of birth	Age in complete d years	Males	Females	Total
1					
2	2018	0	364,155	347,749	711,904
3	2017	1	370,453	355,472	725,925
4	2016	2	378,518	363,162	741,680
5	2015	3	387,906	372,402	760,308
6	2014	4	399,232	387,042	786,274
7	2013	5	407,611	389,920	797,531
8	2012	6	417,471	396,835	814,306
9	2011	7	418,623	403,349	821,972
10	2010	8	429,919	412,555	842,474
11	2009	9	427,917	408,232	836,149
92	1928	90	55,382	124,322	179,704
93	1927	91	44,797	105,456	150,253
94	1926	92	34,519	91,072	125,591
95	1925	93	27,317	76,447	103,764
96	1924	94	20,525	61,235	81,760
97	1923	95	14,477	48,398	62,875
98	1922	96	10,101	37,882	47,983
99	1921	97	7,239	27,754	34,993
100	1920	98	4,977	19,813	24,790
101	1919	99	2,058	8,273	10,331
102	1918 or before	100 or over	2,976	12,670	15,646
103		Total	32,394,531	34,598,168	66,992,699

Source: <https://www.insee.fr/en/statistiques/2382597?sommaire=2382613>

1. Affichez le nombre d'individus pour chaque groupe d'âge sous forme de diagramme en bâtons.
2. Formater les données - supprimer l'écart entre des bâtons.
3. Calcule la fréquence accumulée et trace sous forme de ligne. (en cellule F2: $=\text{Sum}(\$E\$2:E2)/\$E\103 et puis copie).
!! Touche de fonction F4 transforme E2 en $\$E\2



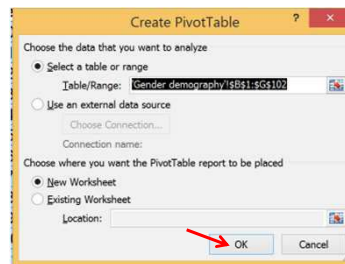
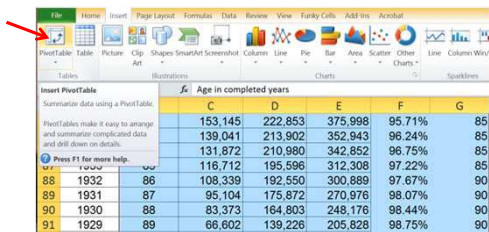
Excel - Tableau Croisé Dynamique (Pivot tables)

Démographie de la population française (Tableau 1):

	A	B	C	D	E	F	G
	Year of birth	Age in complete d years	Males	Females	Total	Freq Acc	<Age
1							
2	2018	0	364,155	347,749	711,904	1.06%	5
3	2017	1	370,453	355,472	725,925	2.15%	5
4	2016	2	378,518	363,162	741,680	3.25%	5
5	2015	3	387,906	372,402	760,308	4.39%	5
6	2014	4	399,232	387,042	786,274	5.56%	5
7	2013	5	407,611	389,920	797,531	6.75%	10
98	1922	96	10,101	37,882	47,983	99.87%	100
99	1921	97	7,239	27,754	34,993	99.92%	100
100	1920	98	4,977	19,813	24,790	99.96%	100
101	1919	99	2,058	8,273	10,331	99.98%	100
102	1918 or before	100+	2,976	12,670	15,646	100.06%	#VALUE!

1. Faire des groupes d'âge avec un intervalle de n années:

$$= n * \text{Plafond} ((B2+1) / n; 1)$$
 Par exemple. n = 5
2. Corrigez les lignes 102 à 100.
3. Sélectionner des données et créer un tableau croisé dynamique

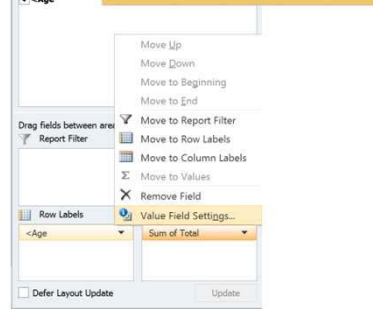
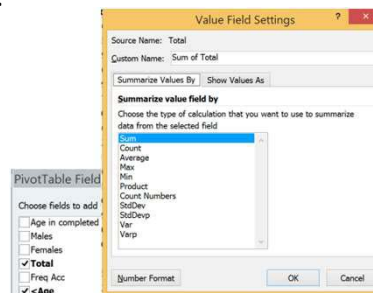


Excel - Tableau Croisé Dynamique (Pivot tables)

Démographie de la population française (Tableau 1):

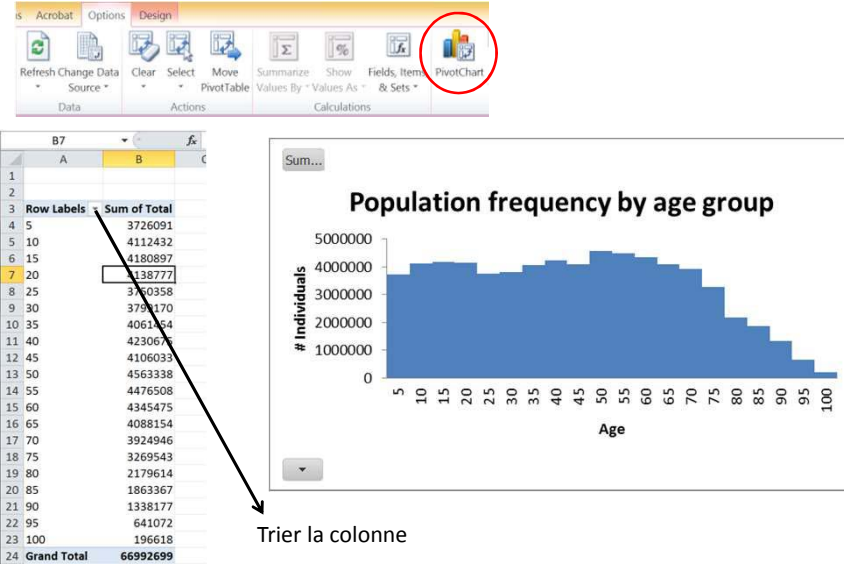


	A	B	C
1			
2			
3	Row Labels	Sum of Total	
4	5	3726091	
5	10	4112432	
6	15	4180897	
7	20	4138777	
8	25	3750358	
9	30	3799170	
10	35	4061454	
11	40	4230675	
12	45	4106033	
13	50	4563338	
14	55	4476508	
15	60	4345475	
16	65	4088154	
17	70	3924946	
18	75	3269543	
19	80	2179614	
20	85	1863367	
21	90	1338177	
22	95	641072	
23	100	196618	
24	Grand Total	66992699	



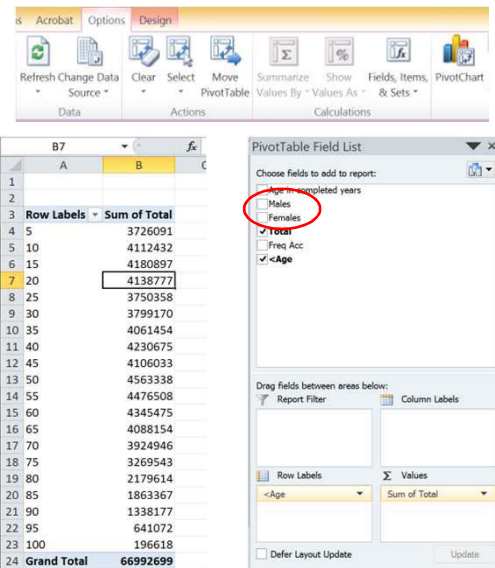
Excel - Tableau Croisé Dynamique (Pivot tables)

Démographie de la population française (Tableau 1):



Excel - Tableau Croisé Dynamique (Pivot tables)

Démographie de la population française (Tableau 1):



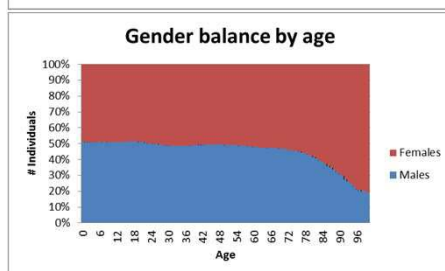
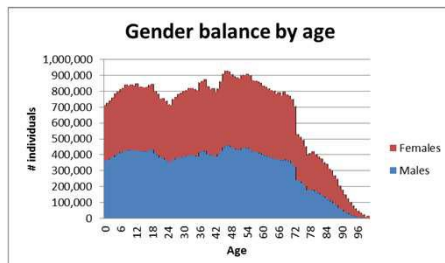
Analyser la parité hommes-femmes stratifiée par «âge» et par «groupes d'âge».

1. Utiliser le jeu de données complet (« age »)
2. Utilisez un tableau croisé dynamique pour effectuer l'analyse stratifiée par groupes d'âge.

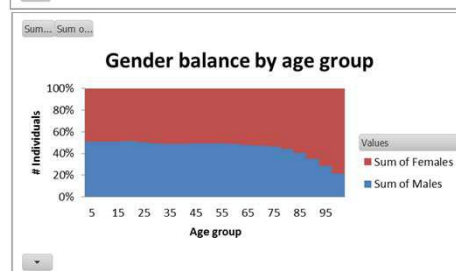
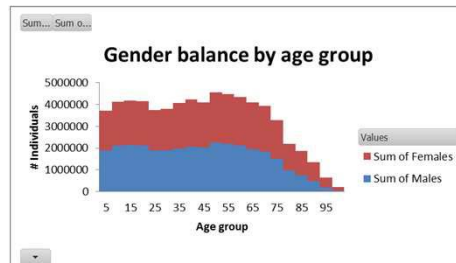
Excel - Tableau Croisé Dynamique (Pivot tables)

Démographie de la population française (Tableau 1):

Stratified by "Age"



Stratified by "Age group"



Conclusion

- Excel permet une analyse simple ainsi que très complexe.
- Excel permet la représentation graphique
 - Le choix de la représentation graphique modifie l'interprétation des données.
- Les tableaux croisés dynamiques sont des outils automatisés pour l'exploration de données.
 - Principaux atouts:
 - Sortie tabulaire et graphique.
 - Modification rapide des critères de stratification.
 - Mises à jour lorsque les données brutes sont modifiées.