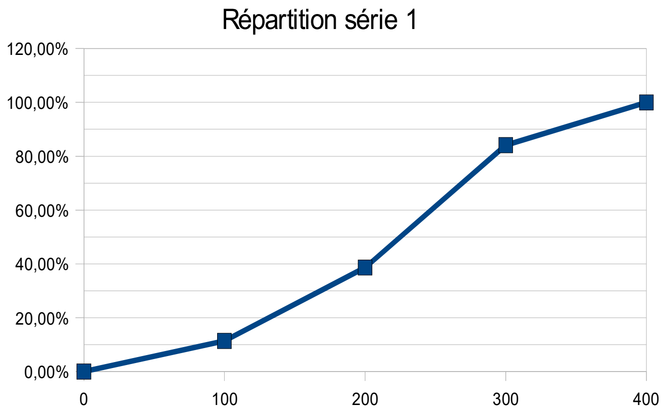
TD n°2 de Statistiques : médiane et quartiles

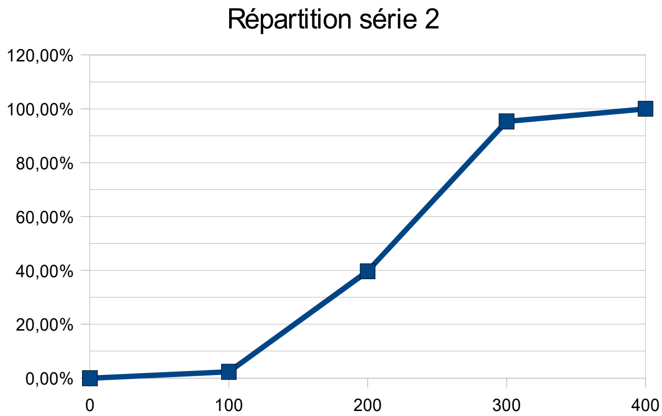
I] Exemple des séries 1 et 2 du cours

a) Compléter ces deux tableaux qui correspondent à deux séries statistiques différentes

valeurs	[0 ; 100[	[100 ; 200[	[200 ; 300[	[300 ; 400[	[0 ; 100[	[100 ; 200[	[200 ; 300[	[300 ; 400[
effectifs	5	12	20	7	6	96	143	12
Cumul croissant des effectifs	5	17	37	44	6	102	245	257
fréquences	11%	27%	45%	16%	2%	37%	56%	5%
Cumul croissant des fréquences	11%	38%	83%	99%	2%	39%	95%	100%

b) Tracer les courbes polygonales donnant les <u>cumuls croissants des fréquences</u> pour ces deux séries. Cette courbe permet de répondre à la question : « quel pourcentage de la population a une valeur inférieure ou égale à *x* ? ». Le tracé polygonal suppose une *répartition homogène* des valeurs dans chaque classe.





c) Lire graphiquement les valeurs de la médiane M, des 1er et 3ème quartiles Q_1 et Q_3 pour les 2 séries.
Série n°1 : M= 225 Q_1= 150 Q_3= 280 ; Série n°2 : M= 220 Q_1= 160 Q_3= 260

d) Calculer, en utilisant le théorème de Thalès, les valeurs précises de la médiane et des quartiles.
D'après les valeurs du tableau pour la Série n°1, le théorème de Thalès conduit à :
$\frac{M-200}{300-200}=\frac{50-39}{95-39}$ soit $\frac{M-200}{100}=\frac{12}{45}$ et donc $M=200+100\times\frac{12}{45}\approx 226{,}67$
De même pour $Q_1=100+100\times\frac{25-11}{27}\approx 151{,}85$ et $Q_3=200+100\times\frac{75-38}{45}\approx 282{,}22$.

D'après les valeurs du tableau pour la Série n°2, le théorème de Thalès conduit à :
$\frac{M-200}{300-200}=\frac{50-39}{96-39}$ soit $\frac{M-200}{100}=\frac{11}{56}$ et donc $M=200+100\times\frac{11}{56}\approx 219{,}64$
De même pour $Q_1=100+100\times\frac{25-2}{37}\approx 162{,}16$ et $Q_3=200+100\times\frac{75-39}{56}\approx 264{,}29$.

e) L'écart inter-quartile *absolu* est Q_3-Q_1 . Il mesure la dispersion absolue des valeurs (exprimée dans la même unité que la valeur). L'écart inter-quartile *relatif* est $\frac{Q_3-Q_1}{Q_2}$. Il mesure la dispersion relative des valeurs (sans unité). Calculer les écarts inter-quartiles relatifs des deux séries, et commentez alors la différence de dispersion des deux séries (lorsqu'une série est très dispersée cet écart est grand, on dit aussi qu'elle est peu concentrée, dispersion et concentration variant en sens inverse).
Série n°1 : $\frac{Q_3-Q_1}{Q_2}=\frac{282{,}22-151{,}85}{226{,}67}\approx 0{,}575$; Série n°2 : $\frac{Q_3-Q_1}{Q_2}=\frac{264{,}29-162{,}16}{219{,}64}\approx 0{,}465$.
Commentaires : La série n°1 est (beaucoup) plus dispersée que la série n°2, l'écart inter-quartile représentant plus de 57% de la médiane dans la série n°1 contre 46% seulement dans la série n°2.

<u>Prolongements</u> : On peut définir les déciles D_1, D_2, …, D_9 de la même manière qu'on a défini les quartiles (ou les centiles C_1, …, C_{99}). Lire graphiquement, puis calculer les déciles D_1 et D_9 pour les deux séries ainsi que les écarts inter-déciles absolus D_9-D_1 et relatifs des deux séries $\frac{D_9-D_1}{D_5}$ (remarque : $D_5=Q_2=M$).
D'après les valeurs du tableau pour la série n°1 on a $D_1=0+100\times\frac{10-0}{11}\approx 90{,}91$ et $D_9=300+100\times\frac{90-83}{16}=343{,}75$ alors que dans la série n°2 on a $D_1=100+100\times\frac{10-2}{37}\approx 121{,}62$ et $D_9=200+100\times\frac{90-39}{56}\approx 291{,}07$. Ces valeurs conduisent aux indicateurs relatifs $\frac{D_9-D_1}{M}=\frac{343{,}75-90{,}91}{226{,}67}\approx 1{,}115$

pour la série n°1 et $\frac{D_9-D_1}{M}=\frac{291,07-121,62}{219,64}\approx 0,771$ pour la série n°1 qui confirment notre remarque précédente, l'écart entre les 2 séries s'amplifiant lorsqu'on considère une plus grande masse centrale (80% de considérés ici entre D_1 et D_9 alors qu'entre Q_1 et Q_3, on ne considère que 50% de la population centrale) : l'écart inter-décile représentant plus de 111% de la médiane dans la série n°1 contre 77% seulement dans la série n°2.

II] Autre exemple

On voudrait comparer les répartitions des salaires mensuels dans deux entreprises A et B données par le tableau ci-contre.

Salaire (en euros)	Fréquence A (%)	Fréquence B (%)
moins de 900	5,4	5,0
De 900 à 1200	15,7	14,0
De 1200 à 1500	32,3	37,0
De 1500 à 1800	22,2	27,0
De 1800 à 2100	15,0	14,0
De 2100 à 2400	7,5	2,5
plus de 2400	1,9	0,5

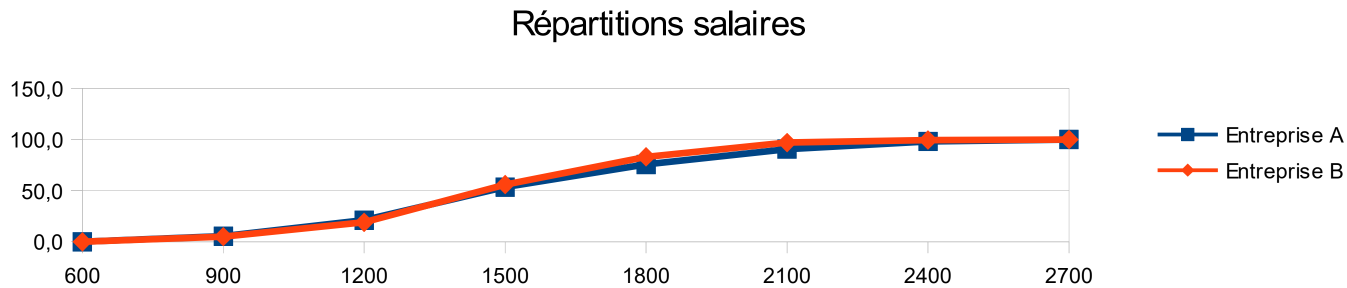
a) Calculer la moyenne des deux séries (prendre pour les deux classes extrêmes une amplitude égale à celle des autres classes).

Pour l'entreprise A, la moyenne des salaires est de 1517,4 euros, alors que pour l'entreprise B, elle est de 1471,5 euros.

Pour les calculs des moyennes, nous avons fait un tableau dans lequel nous avons calculé le total des salaires (somme des produits $f_i \,.\, x_i$ où f_i est la fréquence de la classe i est x_i est le centre de cette classe) que nous avons divisé ensuite par le total des fréquence (100%).

	600-900	900-1200	1200-1500	1500-1800	1800-2100	2100-2400	2400-2700	Total	Moyenne
Entreprise A fi	5,4	15,7	32,3	22,2	15	7,5	1,9	100	
Centres de classe xi	750	1050	1350	1650	1950	2250	2550		
Entreprise A fi . Xi	4050	16485	43605	36630	29250	16875	4845	151740	1517,4
Entreprise B fi	5	14	37	27	14	2,5	0,5		
Entreprise B fi . Xi	3750	14700	49950	44550	27300	5625	1275	147150	1471,5

b)Construire les courbes de répartition des deux séries (polygones des fréquences cumulées en croissant) et déterminer la médiane, les quartiles et l'écart inter-quartile relatif pour chacune de ces séries.



	600	900	1200	1500	1800	2100	2400	2700
Entreprise A	0,0	5,4	21,1	53,4	75,6	90,6	98,1	100,0

D'après les valeurs du tableau pour l'entreprise A, on a $M=1200+300\times\frac{50-21,1}{32,3}\approx 1468,42$

De même pour $Q_1=1200+300\times\frac{25-21,1}{32,3}\approx 1236,22$ et $Q_3=1500+300\times\frac{75-53,4}{22,2}\approx 1791,89$.

On en déduit l'écart inter-quartile relatif pour l'entreprise A $\frac{Q_3-Q_1}{Q_2}=\frac{1791,89-1236,22}{1468,42}\approx 0,378$

	600	900	1200	1500	1800	2100	2400	2700
Entreprise B	0	5	19	56	83	97	99,5	100

D'après les valeurs du tableau pour l'entreprise B, on a $M=1200+300\times\frac{50-19}{37}\approx 1451,35$

De même pour $Q_1=1200+300\times\frac{25-19}{37}\approx 1248,65$ et $Q_3=1500+300\times\frac{75-56}{27}\approx 1711,11$.

On en déduit l'écart inter-quartile relatif pour l'entreprise B $\frac{Q_3-Q_1}{Q_2}=\frac{1711,11-1248,65}{1451,35}\approx 0,319$

c) Conclure.

L'entreprise A a des salaires plus élevés que l'entreprise B (il y a un écart de 45,9 euros entre les salaires moyens de ces entreprises) et la dispersion des salaires dans cette entreprise A est plus grande que dans la B. Cela veut dire qu'il y aura un plus grand écart de salaire entre ceux qui gagne beaucoup (25% gagnent plus de 1791 euros) et ceux qui gagnent moins (25% gagnent moins de 1468 euros) : près de 38% du salaire médian. Dans l'entreprise B cet écart représente moins de 32%.