Exercice 1. Pour chacune des situations suivantes déterminer la population de l'étude, l'échantillon étudié, la variable et son type.

1. On effectue un sondage auprès de 200 étudiants louvanistes pour connaitre leur bière préférée.

   > **Solution:** (a) Population : l'ensemble des étudiants de LLN (b) Échantillon : 200 étudiants de LLN sélectionnés (c) Variable considérée : Type de bière préférée (Qualitative nominale)

2. Afin de déterminer le profil socioéconomique des ménages de la villes de Montréal, on a noté le nombre d'enfants par ménage pour un échantillon de 380 ménages.

   > **Solution:** (a) Population : l'ensemble des familles de la ville de Montréal (b) Échantillon : 380 ménages sélectionnés (c) Variable considérée : Nombre d'enfants par famille (Quantitative discrète)

3. Selon les données du recensement, 58% de la population parlait le flamand à la maison, 27% le français, 2% l'allemand et 13% une autre langue.

   > **Solution:** (a) Population : l'ensemble de la population belge (b) Échantillon : aucun car recensement (c) Variable considérée : Langue parlée à la maison (Qualitative nominale)

4. Lors d'une étude, la sécurité Routière Belge s'est intéressée à classer un échantillon de 150 accidents selon le type de dégâts observés : "Léger", "Important", "Mortel".

   > **Solution:** (a) Population : l'ensemble des accidents de Belgique (b) Échantillon : 150 accidents sélectionnés (c) Variable considérée : Gravité de l'accident (Qualitative ordinale)

Exercice 2. Dans un communiqué de presse datant du 10 janvier 2010 intitulé *"Les Belges et leurs voitures"*[1], la Direction Générale Statistique et Information économique a voulu

1. Rapport complet téléchargeable sur `http://statbel.fgov.be/fr/modules/pressrelease/statistiques/circulation_et_transport/les_belges_et_leurs_voitures_2010.jsp`

connaître les habitudes des belges vis à vis de l'utilisation de leurs véhicules. Ils comparent le résultat belge avec les autres pays européens. Ainsi le tableau suivant donne le nombre de voitures particulières en 2004 pour 1000 habitants de nos voisins européens.

Pays	Liechten-stein	Luxem-bourg	France	Belgique	Islande	Allema-gne	Pays-Bas
Nombre voitures	692	659	491	467	599	546	429

1. Nommer et donner le type de la variable d'étude.

> **Solution:** On s'interresse ici aux *Nombres de voitures pour 1000 habitants* (Variable *quantitative discrète*).

2. Calculer et interpréter la médiane de cet échantillon. Que peut-on dire sur le résultat de la Belgique vis-à-vis de l'ensemble de l'Europe.

> **Solution:** **Etape 1 :** On ordonne les valeurs de l'échantillon par ordre croissant.
>
> 429 467 491 546 599 659 692
>
> **Etape 2 :** Soit n le nombre d'observations, puisque $n = 7$ est impair alors :
>
> $$\mathsf{Med}(X) = X_{\left(\frac{n+1}{2}\right)} = X_{\left(\frac{7+1}{2}\right)} = X_{(4)} = 546$$
>
> La médiane vaut 546 c'est à dire qu'il y a autant d'observations supérieurs à 546 et inférieurs à 546.

3. Calculer la moyenne et l'écart-type. Interpréter les résultats.

> **Solution:** On travaille sur un échantillon. Alors,
>
> $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{7}(692 + 659 + 491 + 467 + 599 + 546 + 429) = 554.71$$
>
> $$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}[x_j - \bar{x}]^2 = \frac{1}{7-1}\left[(692 - 554.71)^2 + \cdots + (429 - 554.71)^2\right] = 9886.24$$
>
> Donc $s = \sqrt{9886.24} = 99.43$. On peut finalement conclure que le nombre de voitures pour 1000 habitants dans la plupart (68%) des pays européens est compris entre $[\bar{x} - s; \bar{x} + s] = [455.28; 654.14]$. De la même manière, on peut conclure

que le nombre de voitures pour 1000 habitants dans 95% des pays européens est compris entre $[\bar{x} - 2s; \bar{x} + 2s] = [355.86; 753.57]$. Enfin, on peut conclure que le nombre de voitures pour 1000 habitants dans presque tous (99.7%) les pays européens est compris entre $[\bar{x} - 3s; \bar{x} + 3s] = [256.43; 853]$.

4. Imaginons maintenant que l'on veuille rajouter la Norvège (429) dans notre échantillon, les valeurs calculées précédemment vont-elles changer ? Déterminer maintenant la médiane dans notre nouvel échantillon.

Solution: Ainsi maintenant on considère

Pays	Liecht.	Lux.	Fr.	Bel.	Isl.	All.	P-B	Norv
Nombre voitures	692	659	491	467	599	546	429	429

Très clairement toutes les mesures statistiques (moyenne, variance, etc.) vont changer. Pour le calcul de la nouvelle médiane on effectue la même démarche que précédemment,

Etape 1 : On ordonne les valeurs de l'échantillon par ordre croissant.

$$429 \quad 429 \quad 467 \quad 491 \quad 546 \quad 599 \quad 659 \quad 692$$

Etape 2 : Soit m le nouveau nombre d'observations, puisque $m = 8$ est *pair* alors :

$$\mathsf{Med}(X) = \frac{X_{(\frac{m}{2})} + X_{(\frac{m}{2}+1)}}{2} = \frac{X_{(4)} + X_{(5)}}{2} = \frac{491 + 546}{2} = 518.5$$

Exercice 3. Dans un rapport datant de 2007, le ministère de l'Éducation du Québec a relevé l'age de l'ensemble de ses enseignants. Le tableau suivant résume les résultats obtenus :

Age (en année)	Pourcentage
Moins de 30 ans	8.5%
$[30, 40[$	23.4%
$[40, 50[$	29.5%
$[50, 60[$	32%
60 ans et plus	6.6%

1. Calculer la moyenne et l'écart-type. Interpréter les résultats.

Solution: On est dans le cas d'une population. Alors, $\mu = \sum_{i=1}^{J} x_i \cdot f_i$ et $\sigma = \sqrt{\sum_{i=1}^{J} (x_i - \mu)^2 \cdot f_i}$. Cependant on va devoir ramener la classe *"Moins de 30 ans"* à la classe $[20; 30[$ et la classe *"Plus de 60 ans"* à $[60; 70[$.

$$\mu = \sum_{i=1}^{5} x_i \cdot f_i = 25 \cdot 0.085 + 35 \cdot 0.234 + 45 \cdot 0.295 + 55 \cdot 0.32 + 65 \cdot 0.066 = 45.48$$

$$\sigma = \sqrt{\sum_{i=1}^{5} (x_i - \mu)^2 \cdot f_i} = \sqrt{(25 - 45.48)^2 \cdot 0.085 + \cdots + (65 - 45.48)^2 \cdot 0.066} = 10.75$$

En conclusion on peut dire que :

- la plupart (68%) des enseignants en 2007 au Québec avait un âge compris entre $[\mu - \sigma; \mu + \sigma] = [34.73; 56.23]$ années
- 95% des enseignants en 2007 au Québec avait un âge compris entre $[\mu - 2\sigma; \mu + 2\sigma] = [23.98; 66.98]$ années
- presque tous (99.7%) les enseignants en 2007 au Québec avait un âge compris entre $[\mu - 3\sigma; \mu + 3\sigma] = [13.23; 77.73]$ années

2. Calculer le coefficient de variation de notre distribution.

Solution: Le coefficient de variation est la mesure statistique qui permet de vérifier l'homogénéité des données. Il est définit par $CV = \sigma/\mu$ où μ est la moyenne de la population et σ l'écart-type de la population. De manière générale on dit que :

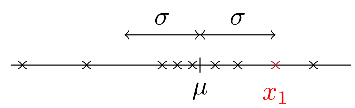
- la population est homogène si $CV < 15\%$
- la population est hétérogène si $CV > 15\%$

Dans notre situation $CV = 10.75/45.48 = 23.6\% > 15\%$ donc la population est hétérogène.

3. Déterminer et interpréter la cote z d'un professeur qui a 22 ans.

Solution: La cote z est une valeur qui permet placer un individu par rapport à ses compères. On peut la définir de la manière suivante $z = \dfrac{Valeur - \mu}{\sigma} = \dfrac{22 - 45.48}{10.75} = -2.18$. On peut dire que le professeur de 22 ans se situe à 2.18 distances d'écart-type de la moyenne (vers la gauche de la distribution puisque z

est négatif).
Dans le cas général, dire que l'observation en x_1 se situe à une distance d'écart-type de la moyenne (vers la droite) signifie :



De plus on dira que :

- l'événement est assez rare si la cote z est plus grande que 2 ou plus petite que -2.
- l'événement est très rare si la cote z est plus grande que 3 ou plus petite que -3.

Dans notre cas on peut remarquer que le fait d'avoir 22 ans et d'être professeur est assez rare (puisque $z < -2$).

4. Déterminer la médiane (50%), le premier décile (10%) et le premier quartile (25%).

Solution: A l'aide des fréquences cumulées, il nous faut tout d'abord identifier qu'elle sera la classe abritant la médiane.

Age (en année)	Fréq	Fréq Cum.
$\ominus$ 30 ans	.085	.085
$[30, 40[$	.234	.319
$[40, 50[$	.294	.613
$[50, 60[$	.32	.939
$\oplus$ 60 ans	.066	1

Ainsi on peut remarquer que la médiane Q_2 se trouve dans la classe $[40, 50[$ ce qui signifie qu'on peut la décomposer comme $Q_2 = 40 + x$.

On conclut que $x = \frac{10(0.5-0.319)}{0.294} = 6.16$ ce qui implique qu'on peut approximer la médiane par la valeur $Q_2 = 40 + 6.16 = 46.16$. Cela signifie que 50% des enseignants avait moins (ou plus) de 46.16 ans.

En effectuant exactement le raisonnement que pour la médiane, on peut approcher le premier décile D_1 par $30 + \frac{10(0.1-0.085)}{0.234} = 30.64$ et le premier quartile par $Q_1 = 30 + \frac{10(0.25-0.085)}{0.234}$=37.05. C'est à dire que 10% de la population (respectivement 90% de la population) avait moins (respectivement plus) de 30.64 ans et 25% de la population (respectivement 75% de la population) avait moins (respectivement plus) de 37.05 ans

EXERCICE 4. Lors du dernier 20km de Bruxelles, un groupe d'amis tous licenciés dans le même club ont bouclé la course selon les temps suivants (en minutes) :

80 111 100 86 86 119 101 109 96 109 113 103

1. Grace à la table de Sturges (Table 1), regrouper les données en classes

Nombre de données	Nombres de classes
Entre 10 et 22	5
Entre 23 et 44	6
Entre 45 et 90	7
Entre 91 et 180	8
Entre 181 et 360	9
Entre 361 et 720	10

TABLE 1 – Table de Sturges

Solution: Comment construire nos classes ? La principale question que l'on doit se poser est : Quelle amplitude de classe doit-on choisir ? Il nous faut suivre la démarche suivante :
Etape 1 : Fixer le nombre de classes grâce à la table de Sturges. Dans notre cas on dispose de 12 données. Ainsi on regroupera nos données en 5 classes.
Etape 2 : Calculer l'étendue de la série E :

$$E = X_{max} - X_{min} = 119 - 80 = 39$$

Etape 3 : Calculer l'amplitude A :

$$A = \frac{Etandue}{Nombre\ Classes} = \frac{39}{5} = 7.8$$

Les classes considérées seront [80 ;88[, [88 ;96[, [96 ;104[, [104 ;112[et [112 ;120[.

Classe	Effectif (n_j)	Fréquence (f_j)	Fréq. Cum.
$[80; 88[$	3	0.25	0.25
$[88; 96[$	0	0	0.25
$[96; 104[$	4	0.33	0.58
$[104; 112[$	3	0.25	0.83
$[112; 120[$	2	0.17	1
Total	12	1	

2. A l'aide du tableau d'effectif de la question 1., calculer la moyenne et l'écart-type.

> **Solution:** On est dans le cas d'un échantillon d'où
>
> $$\begin{aligned}\bar{x} &= \tfrac{1}{n}\sum_{j=1}^{J} n_j x_j = \tfrac{1}{12}\sum_{j=1}^{5} n_j x_j = \frac{1}{12}\left(3\cdot 84 + 0\cdot 92 + \cdots + 2\cdot 116\right)\\ &= 100.67\\ s^2 &= \frac{1}{n-1}\sum_{j=1}^{J}\left[n_j\cdot(x_j-\bar{x})^2\right]\\ &= \frac{1}{12-1}\left[3\cdot(84-100.67)^2+\cdots+2\cdot(116-100.67)^2\right]\\ &= 133.33\end{aligned}$$
>
> alors l'écart-type vaut $\sqrt{133.33} = 11.55$. Ce qui signifie que le temps de course de :
>
> - la plupart (68%) de nos amis est compris entre $[\bar{x}-s;\bar{x}+s] = [89.12; 112.22]$ minutes
> - 95% de nos amis est compris entre $[\bar{x}-2s;\bar{x}+2s] = [77.57; 123.77]$ minutes
> - presque tous (99.7%) nos amis est compris entre $[\bar{x}-3s;\bar{x}+3s] = [66.02; 135.32]$ minutes

3. Déterminer la classe modale

> **Solution:** La classe modale correspond à la classe regroupant le plus d'individus. En utilisant le tableau de fréquence on conclut que la classe modale correspond à la classe $[96; 104[$.

4. A l'aide du tableau d'effectif déterminer la médiane et le troisième quartile.

> **Solution:** En utilisant la méthode de l'interpolation linéaire (idem que l'exercice 3) on arrive à $Q_2 = 96 + \frac{8(0.50-0.25)}{0.33} = 102.06$ et $Q_3 = 104 + \frac{8(0.75-0.58)}{0.25} = 109.44$.

Exercice Supplémentaire 1. Pour vérifier l'efficacité d'une nouvelle pile, une association de consommateurs teste un échantillon composé de 20 de ces nouvelles piles. Chacune est soumise à un usage identique jusqu'à son extinction. Les durées de vie mesurée (en min) sont :

65.1	58.4	64.9	76.0	67.8	75.1	76.7	64.2	74.9	77.6
58.0	68.0	73.3	75.4	76.0	59.4	65.4	74.7	76.6	81.3

1. Regrouper les données en classes.

Solution: On dispose de $n = 20$ données, à partir de la table de Sturges (cfr. Table 1) je vais devoir regrouper mes données en 5 classes d'amplitude

$$A = \frac{81.3 - 58.0}{5} = 4.66 \approx 5.$$

Les classes considérées seront $[58, 63[$, $[63, 68[$, $[68, 73[$, $[73, 78[$, $[78, 83[$

Classe	[58,63[	[63,68[	[68,73[	[73,78[	[78,83[	Total
Effectif (n_i)	3	5	1	10	1	20
Fréquence (f_i)	0.15	0.25	0.05	0.50	0.05	1
Fréq. Cum. (F_i)	0.15	0.4	0.45	0.95	1	
Centre (x_i)	60.5	65.5	70.5	75.5	80.5	

2. A partir de données de l'énoncé, calculer la moyenne. Comparer le résultat obtenu à la moyenne calculée à partir de données regroupées en classes de la question précédente.

Solution: A partir de données de l'énoncé, $\bar{x}_1 = \frac{\sum_{i=1}^{n} x_i}{n}$ où x_i correspond à la durée de vie de la i^{eme} pile de l'échantillon. Donc :

$$\begin{aligned} \bar{x}_1 &= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{65.1 + 58.4 + \cdots + 76.6 + 81.3}{20} \\ &= 70.44 \end{aligned}$$

A partir des classes de la question précédente, $\bar{x}_2 = \frac{\sum_{i=1}^{J} x_i \cdot n_i}{n} = \sum_{i=1}^{J} x_i \cdot f_i$ où x_i représente le centre de la i^{eme} classe, n_i l'effectif de la i^{eme} classe, f_i la fréquence de la i^{eme} classe et J correspond au nombre de classes. Alors,

$$\begin{aligned} \bar{x}_2 &= \sum_{i=1}^{n} x_i \cdot f_i = 60.5 \cdot 0.15 + 65.5 \cdot 0.25 + 70.5 \cdot 0.05 + 75.5 \cdot 0.5 + 80.5 \cdot 0.05 \\ &= 70.75 \end{aligned}$$

Le fait de regrouper les classes implique une perte de précision du calcul de la moyenne. Cette erreur diminue lorsque que le nombre de classes augmente.

3. Effectuer le même raisonnement pour l'écart-type.

Solution: A partir de données de l'énoncé, $s_1 = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_1)^2}{n-1}}$ où x_i correspond à la durée de vie de la i^{eme} pile de l'échantillon. Donc :

$$\begin{aligned} s_1 &= \sqrt{\frac{(65.1 - 70.75)^2 + (58.4 - 70.75)^2 + \cdots + (81.3 - 70.75)^2}{20-1}} \\ &= \sqrt{\frac{958.768}{19}} = 7.11 \end{aligned}$$

A partir des classes de la question précédente, $s_2 = \sqrt{\frac{\sum_{i=1}^{J} n_i (x_i - \bar{x}_2)^2}{n-1}} = \sqrt{\frac{n}{n-1}\sum_{i=1}^{J} f_i(x_i - \bar{x}_2)^2}$ où x_i représente le centre de la i^{eme} classe, n_i l'effectif de la i^{eme} classe, f_i la fréquence de la i^{eme} classe et J le nombre de classes. Alors,

$$\begin{aligned} s_2 &= \sqrt{\frac{n}{n-1}\sum_{i=1}^{J} f_i(x_i - \bar{x}_2)^2} \\ &= \sqrt{\frac{20}{19}\left(0.15 \cdot (60.5 - 70.75)^2 + \cdots + 0.05 \cdot (80.5 - 70.75)^2\right)} \\ &= \sqrt{\frac{20 \cdot 38.6875}{19}} = 6.38 \end{aligned}$$

On remarque le même phénomène que lors du calcul de la moyenne : Le fait de regrouper les classes implique une perte de précision pour l'écart-type.

4. Même raisonnement pour la médiane.

Solution: Pour déterminer la médiane à partir des données de l'énoncé il nous faut tout d'abord ordonner (dans l'ordre croissance) les données.

58.0	58.4	59.4	64.2	64.9	65.1	65.4	67.8	68.0	73.3
74.7	74.9	75.1	75.4	76.0	76.0	76.6	76.7	77.6	81.3

On dispose d'un nombre pair de données ($n/2$) donc la médiane sera donnée par $\frac{x_{(n/2)} + x_{(n/2+1)}}{2}$ où $x_{(n)}$ représente la n^{eme} valeur ordonnée de la série (ainsi remarquons que $x_{(1)}$ correspond au minimum de l'échantillon). Finalement :

$$\text{Médiane} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2} = \frac{x_{(10)} + x_{(11)}}{2} = \frac{73.3 + 74.7}{2} = 74$$

A partir des données en classe, on cherche tout d'abord la classe dans laquelle se trouve la médiane. Pour se faire on regarde les fréquences cumulées dans la première question :

Classe	[58,63[	[63,68[	[68,73[	[73,78[	[78,83[
Fréquence (f_i)	0.15	0.25	0.05	0.50	0.05
Fréq. Cum. (F_i)	0.15	0.4	0.45	0.95	1

Ainsi on sait que la médiane se situe dans la classe $[73, 78[$ car grâce au tableau ci-dessus on sait que 45% des données sont plus petites que 73 (strictement). De la même façon que pour l'exercice 3, on trouve que la médiane est égale à $73 + \dfrac{5(0.5 - 0.45)}{0.5} = 73.5$.

Exercice Supplémentaire 2. Lors du dernier championnat d'Europe de Basket-Ball en Septembre 2013, on a relevé la taille de l'ensemble des joueurs de la compétition. [2] Les données groupées (en m) sont présentées dans le tableau suivant :

Classe	[1.75, 1.82[	[1.82, 1.89[	[1.89, 1.96[	[1.96, 2.03[	[2.03, 2.10[	[2.10, 2.17[	[2.17, 2.24[
Fréquence	.021	.096	.258	.254	.254	.107	.010

1. Déterminer la classe modale.

Solution: Au vu du tableau de fréquence ci-dessus, la classe modale est $[1.89, 1.96[$ ce qui signifie que la majorité des joueurs avait entre 1.89m et 1.96m.

2. Calculer la moyenne et l'écart-type.

Solution: On travaille sur une population d'où

$$\mu = \sum_{j=1}^{J} f_j \cdot x_j = [1.785 \cdot 0.021 + 1.855 \cdot 0.096 + 1.925 \cdot 0.258 + 1.995 \cdot 0.254$$
$$+2.065 \cdot 0.254 + 2.135 \cdot 0.107 + 2.205 \cdot 0.010] = 1.994m$$

En moyenne un joueur participant au championnat d'Europe 2013 mesurait

2. Les données sont disponibles via le lien suivant `http://www.eurobasket2013.org`

1.994m. Pour l'écart-type on a :

$$\begin{aligned}\sigma^2 &= \sum_{j=1}^{7} f_j \cdot (x_j - \mu)^2 \\ &= \left[(1.785 - 1.994)^2 \cdot 0.021 + \cdots + (2.205 - 1.994)^2 \cdot 0.010\right] \\ &= 0.0082\end{aligned}$$

On s'intéresse à l'écart-type donc $\sigma = \sqrt{0.0082} = 0.091$.

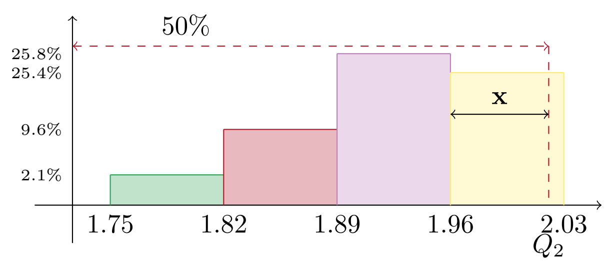
3. Déterminer la médiane et le quantile à 10%.

Solution:

Classe	Fréquence	Fréquence Cumulée
$[1.75; 1.82[$	0.021	0.021
$[1.82; 1.89[$	0.096	0.117
$[1.89; 1.96[$	0.258	0.375
$[1.96; 2.03[$	0.254	0.629
$[2.03; 2.10[$	0.254	0.883
$[2.10; 2.17[$	0.107	0.990
$[2.17; 2.24[$	0.010	1

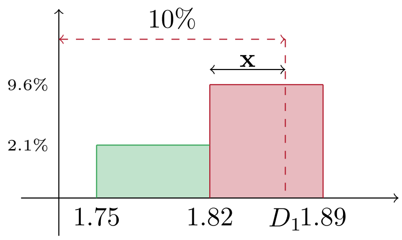
Au vu des données de l'énoncé, on ne dispose que de données en classe et non pas de l'ensemble des données (comme dans les exercices précédents) ainsi on ne peut pas déterminer exactement la médiane. Cependant on peut l'approximation en utilisant la méthode dite de "l'interpolation linéaire". On va tout d'abord rechercher la classe où l'on franchit le cap des 50% des observations. A partir du tableau ci-contre on sait que la médiane va se trouver dans la classe $[1.96; 2.03[$.

On peut donc décomposer la médiane comme $Q_2 = 1.96 + \dfrac{0.07(0.5 - 0.375)}{0.254}$.



Finalement $Q_2 = 1.96 + 0.034 = 1.994$. Ce qui signifie qu'il y a autant de personnes mesurant plus de 1.994m que moins.

On s'intéresse maintenant au premier décile notée D_1 (i.e. 10% de la population).



Au vu du graphique ci-contre, le premier décile va se trouve dans la classe $[1,82; 1.89]$.

On peut donc le décomposer comme $D_1 = 1.82 + \dfrac{0.07(0.10 - 0.021)}{0.096}$. D'où $D_1 = 1.82 + 0.058 = 1.878m$. Autrement dit il y a 10% (90%) de la population qui mesure moins (plus) de $1.878m$.

Exercice Supplémentaire 3. On dispose de 2 échantillons de taille $n_1 = 30$ (respectivement $n_2 = 70$) et de moyenne $\bar{x}_1 = 100$ (respectivement $\bar{x}_2 = 110$). On regroupe maintenant les 2 échantillons, calculer la moyenne de ce nouvel échantillon.

Solution: En partant de la formule de la moyenne on a $\bar{x}_1 = \dfrac{\sum_{i=1}^{n_1} x_i}{n_1}$ ce qui équivalent à $\sum_{i=1}^{n_1} x_i = \bar{x}_1 n_1$. D'où avec $n = n_1 + n_2$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} x_i}{n_1 + n_2} = \frac{30 \cdot 100 + 70 \cdot 110}{30 + 70} = 107$$

Intuitivement, il est logique que :

- la moyenne des 2 echantillons combinés se trouve entre les 2 moyennes individuelles
- la moyenne des 2 échantillons combinés se trouve plus proche (107 est plus proche de 110 que de 100) de la moyenne individuelle qui comporte le plus d'observations car plus de poids est attribué à la moyenne de l'échantillon où n est le plus grand

Exercice Supplémentaire 4. On dispose de toutes les données d'une population composée de 8 individus. L'écart-type de cette série est de 19.7. On dispose également d'une autre série statistique identique à la première mais qui elle est relative à un échantillon extrait d'une population plus grande. Quel est l'écart-type de cette deuxième série, qui permettra d'estimer l'écart-type de la population correspondante ?

Solution: En partant de la formule de la l'écart-type d'une population on a : $\sigma = \sqrt{\frac{1}{n}\sum(x-\mu)^2}$ et pour l'écart-type d'un échantillon on a : $s = \sqrt{\frac{1}{n-1}\sum(x-\bar{x})^2}$. Finalement on arrive à (p. 62 dans le syllabus, point 5.5) :

$$\begin{aligned} s^2 &= \sigma^2 \frac{n}{n-1} \\ &= 19.7^2 \frac{8}{8-1} = 443.53 \\ s &= \sqrt{443.53} = 21.06 \end{aligned}$$

Exercice Examen (Janvier 2014). Le Centre Sportif du Blocry a publié le nombre de calories brulées pour l'ensemble des sports pouvant être pratiqués dans leurs locaux. Les résultats du tableau ci-dessous sont exprimés en kilojoules (KJ) et représentent le nombre de calories brulées pour une pratique d'environ 30 min :

Sport	Marche sportive	Tennis	Natation	Running	Squash	Sretching	Golf
Nombre de KJ	632	758	888	1139	1520	318	569

1) Calculer la moyenne et la variance de la population considérée ci-dessus. Interpréter les résultats.

Solution: On est dans le cas d'une population. On pose que n_i représente le nombre de calories (en KJ) brulées pour 30 min de pratique du *i-ème* sport.

$$\begin{aligned} \mu &= \frac{1}{n}\sum_i n_i = \frac{569+632+758+888+1139+1520+318}{7} = 832 \\ \sigma^2 &= \frac{1}{n}\sum_i (n_i-\mu)^2 = \frac{(569-832)^2+\cdots+(318-832)^2}{7} = 135652.9 \end{aligned}$$

La majorité des données vont se situer dans l'intervalle :

$$\begin{aligned} [\mu-\sigma, \mu+\sigma] &= [832-368.31, 832+368.31] \\ &= [463.69; 1200.31] \end{aligned}$$

2) Déterminer la médiane de la population.

Solution: On ordonne tout d'abord notre série :

$$318 \quad 569 \quad 632 \quad 758 \quad 888 \quad 1139 \quad 1520$$

Puisqu'on dispose d'un nombre impair de valeur la médiane correspond à la 4^{eme} valeur, c'est à dire 758.

3) D'après cette même étude, on sait que la pratique du squash permet de bruler 1520KJ pour 30 min de pratique. Calculer la mesure statistique qui permet de classer ce sport par rapport aux autres sports. Interpréter et commenter les résultats obtenus.

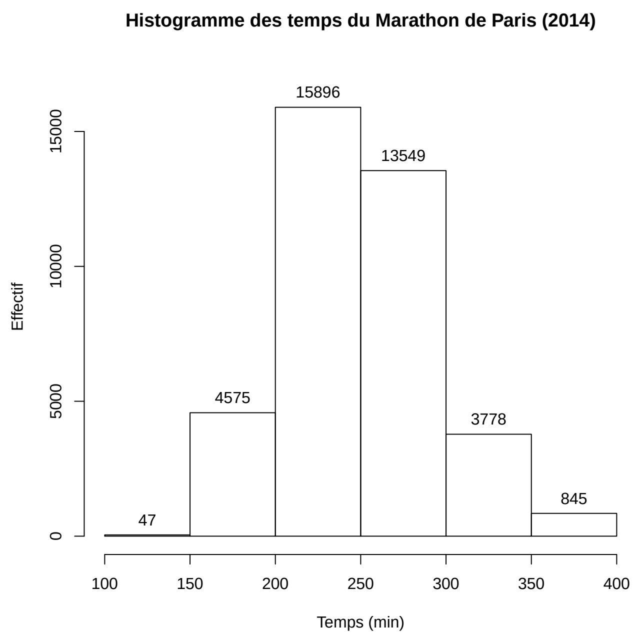
Solution: On calcule la cote z de cette observation : $z = \dfrac{1520 - 832}{\sqrt{135652.9}} = 1.87$.

De plus on dit que :

- l'événement est assez rare si la cote z est plus grande que 2 ou plus petite que -2.
- l'événement est très rare si la cote z est plus grande que 3 ou plus petite que -3.

Dans notre cas on peut remarquer que faire du squash pendant 30 min n'est pas considéré comme une dépense d'énergie rare/extrême (puisque $z < 2$).

Exercice Examen (Juin 2014). Lors du dernier marathon de Paris en avril 2014 sur les 39967 coureurs partant seuls 38690 ont franchi la ligne d'arrivée (les autres ont abandonné). Les résultats ont été regroupés en 6 classes d'amplitude 50 min et sont représentés dans l'histogramme ci-dessous représentant les temps du marathon de Paris 2014. On a noté au-dessus de chaque classe son effectif.



1. A partir de l'histogramme, dresser le tableau de distribution du temps des participants du dernier Marathon de Paris (Effectif et Fréquence).

Solution: Tableau de distribution

Classe	[100,150[	[150,200[	[200,250[	[250,300[	[300,350[	[350,400[	Total
Effectif	47	4575	15896	13549	3778	845	38690
Fréquence	0.0012	0.1182	0.4109	0.3502	0.0976	0.0218	1
Fréq. Cum.	0.0012	0.1194	0.5303	0.8805	0.9781	1	

2. Peut-on déterminer la valeur exacte de médiane ? Si oui la calculer sinon, déterminer une valeur approchée en expliquant votre raisonnement.

Solution: En utilisant une approximation linéaire, on a :

$$\mathsf{Med} = 200 + \frac{50 \cdot (0.50 - 0.1194)}{0.4109} = 246.31$$

3. Calculer la variance de la population.

Solution: On doit tout d'abord calculer la moyenne de la population

$$\mu = \frac{47 \cdot 125 + \cdots + 845 \cdot 375}{38690} = 249.52 \text{ min}$$

Puis finalement la variance vaut :

$$\sigma^2 = \frac{47 \cdot (125 - 249.52)^2 + \cdots + 845 \cdot (375 - 249.52)^2}{38690} = 2050.07$$

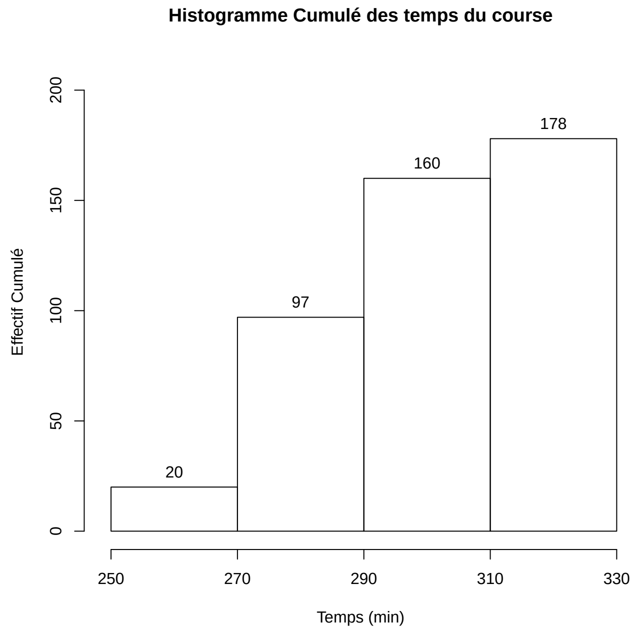
4. Le vainqueur de la course a parcouru les 42.195 km en 2h05, calculer une mesure statistique qui permet de classer cet individu parmi les autres concurrents. Interpréter le résultat.

Solution: On doit calculer la cote z de cet individu : $z = \frac{125-249.52}{\sqrt{2050.07}} = -2.75$.
De plus on dit que :

- l'événement est assez rare si la cote z est plus grande que 2 ou plus petite que -2.
- l'événement est très rare si la cote z est plus grande que 3 ou plus petite que -3.

Dans notre cas on peut remarquer que compléter le marathon de Paris en 2h05 est rare (puisque $z < -2$).

Exercice Examen (Septembre 2014). Lors du dernier Tour de France en juillet 2014, on a noté les temps de courses de l'ensemble des 178 coureurs qui ont participé à la 3ème étape. Les temps ont été regroupés en 4 classes d'amplitude 20 min et sont représentés dans le graphique ci-dessous. On a noté au-dessus de chaque classe son **effectif cumulé**.



1. A partir de la Figure ci-dessus, dresser le tableau de la distribution du temps de course des coureurs de la troisième étape du Tour de France 2014 (Effectif et Fréquence).

Solution: Tableau de distribution

Classe	[250,270[	[270,290[	[290,310[	[310,330[	Total
Effectif Cum.	20	97	160	178	
Effectif	20	77	63	18	178
Fréquence	0.112	0.433	0.354	0.101	1
Fréq. Cum.	0.112	0.545	0.899	1	

2. Peut-on déterminer la valeur exacte de la médiane ? Pourquoi ? Si oui la calculer sinon déterminer une valeur approchée.

Solution: En utilisant une approximation linéaire, on a :

$$\mathsf{Med} = 270 + \frac{20 \cdot (0.5 - 0.112)}{0.433} = 287.92$$

3. Le vainqueur de la course a parcouru les 194 km de l'épreuve en 4h15, calculer une mesure statistique qui permette de classer cet individu parmi les autres concurrents.

Interpréter le résultat.

Solution: On doit calculer la cote z de cet individu, cependant il nous faut tout d'abord connaître la moyenne et l'écart-type de la population. Pour la moyenne de la population on a :

$$\mu = \sum_i f_i x_i = 260 \cdot 0.112 + 280 \cdot 0.433 + 300 \cdot 0.354 + 320 \cdot 0.101 = 288.88 \text{ min}$$

Puis finalement la variance vaut :

$$\sigma^2 = \sum_i f_i (x_i - \mu)^2 = 0.112 \cdot (260 - 288.88)^2 + \cdots + 0.101 \cdot (320 - 288.88) = 269.52$$

Donc l'écart-type vaut $\sigma = \sqrt{269.52} = 16.42$

Finalement $z = \frac{255 - 288.88}{16.42} = -2.06$

De plus on dit que :

- l'événement est assez rare si la cote z est plus grande que 2 ou plus petite que -2.
- l'événement est très rare si la cote z est plus grande que 3 ou plus petite que -3.

Dans notre cas on peut remarquer que terminer l'épreuve en 4h15 est considéré comme rare (puisque $z < -2$).