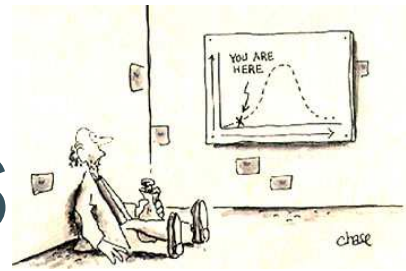


Deuxième Leçon

STATISTIQUES



A Des mathématiques en sciences humaines et sociales

A1 La démarche scientifique en sciences humaines et sociales

Les sciences humaines sont souvent amenées à tirer des enseignements d'observations de certains faits, tout comme les autres sciences. Par exemple, on peut, dans un cadre sociologique, étudier l'affirmation suivante : « les filles réussissent mieux leurs études que les garçons ».

On peut avoir une opinion sur cette affirmation fondée sur son expérience, le sens commun, ses propres valeurs. Il faudrait cependant être capable de l'étudier à partir de faits bien choisis. Quelles informations faut-il obtenir afin de porter un jugement éclairé sur cette affirmation ? Par exemple :

- quelles sont les sources de cette affirmation ?
- A-t-on des données pour le confirmer ?
- Comment ces données ont-elles été collectées ?
- Peut-on comparer les réussites scolaires ?
- etc.

En sciences physiques ou naturelles, par exemple, le scientifique procède souvent ainsi pour analyser un phénomène :

- observer et mettre en évidence sa régularité ;
- le décrire ;
- l'expliquer ;
- le prédire, le maîtriser.

Pour ce qui concerne les sciences humaines, la prédiction est plus difficile car il est délicat de prédire le comportement des individus, de l'économie, etc.

L'utilisation massive de techniques statistiques (graphiques, indices, corrélation, régression, analyse de Fourier) dans les années 1920, appliquées à un objet d'étude omniprésent - le cycle économique -, a conduit à la multiplication des baromètres (et des instituts de conjoncture). Mais leur faillite, perceptible dès 1925, est flagrante en 1929.^a

Et depuis 1929, les absences de prédictions des crises ultérieures s'est perpétuée...

Il est cependant possible de suivre une démarche scientifique en sciences sociales en formulant une problématique, en choisissant ou en élaborant des instruments de mesure, en collectant des données, en les traitant, en les analysant et en les diffusant.

Une différence notable demeure : il est difficile d'adopter la démarche expérimentale des autres sciences. Par exemple, on ne va pas déclencher une guerre en France pour étudier ses effets sur la population ; on ne peut pas non plus diviser le prix du baril de pétrole par 100 pour vérifier certaines hypothèses... On favorise donc l'enquête en sciences sociales.

A2 Les études statistiques

Définir la (les) statistique(s) relève de l'exploit : il existe au moins une centaine de définitions différentes. Nous nous contenterons de parler d'études statistiques dans un cadre mathématique. Nous nous bornerons même en classe de 1^{ère} à étudier la *statistique descriptive* qui a pour objet la représentation de données par des tableaux, des graphiques, des

a. in Michel ARMATTE et Alain DESROSIÈRES : SOUS LA DIRECTION DE JEAN-PIERRE BEAUDET JEAN-GUY PRÉVOST L'ère du chiffre. Presses de l'Université du

mesures. Vous n'aborderez pas la *statistique inférentielle* qui s'occupe de généraliser à toute une population des résultats observés sur un échantillon ou à les projeter dans le futur.

A3 Acquérir un esprit critique

« La quantification constitue un moyen de dépasser l'intuition mais comme elle permet seulement de refléter la réalité sociale de façon parcellaire, il est recommandé d'exercer son sens critique et de s'interroger sur la manière par laquelle les données ont été obtenues ainsi que sur les raisons de leur publication^b ».

On pourra à ce propos lire cet article paru dans le TIME le 26 avril 1993 :

« Damned lies and statistics » : <http://www.time.com/time/magazine/article/0,9171,978314,00.html>

B Les variables

B1 Vocabulaire

Lors une enquête statistique, on recueille un certain nombre de caractéristiques ou de facteurs qu'on nomme en fait **variables** car elles... varient d'une observation à l'autre.

L'individu (personne, objet, entreprise, pays, etc.) sur lequel on mesure la variable est appelé **unité statistique**. L'ensemble de toutes les unités statistiques constitue la **population**. Cet ensemble peut parfois différer de l'ensemble des unités statistiques « interrogées » qui est alors appelé **échantillon**.

B2 Variables qualitatives

Une variable est qualitative si ses différentes formes sont des catégories, des attributs comme par exemple : la spécialité choisie en terminale, le sexe, la couleur des chaussettes, degré de satisfaction à l'égard du professeur de maths, la région d'origine de son arrière grand-père, etc.

Les différentes formes prises par la variable sont appelées **modalités**. Par exemple, la variable « degré de satisfaction à l'égard du professeur de maths » peut prendre quatre modalités : satisfait, très satisfait, extrêmement satisfait, super-mega-plus satisfait.

B3 Variables quantitatives

Une variable est quantitative si elle s'exprime sous la forme d'une valeur numérique : revenu, taille du lob de l'oreille gauche, âge, température, temps consacré quotidiennement à l'étude des mathématiques, etc.

Les variables quantitatives sont dites **continues** si elles peuvent couvrir toutes les valeurs d'un intervalle comme par exemple : la taille (on ne grandit pas d'un centimètre instantanément...), l'âge, etc.

Les autres variables sont dites **discrètes** comme par exemple le nombre de frères et sœurs, nombre de croissants au beurre consommés au petit déjeuner, etc.

C Tableaux et graphiques

C1 Variables qualitatives ou quantitatives discrètes (peu de valeurs distinctes)

C1a Tableau des effectifs/ des fréquences

Le ministre de la justice de Syldavie a emprisonné 75 personnes la semaine dernière car elles n'avaient pas suffisamment déposé d'offrandes au Grand Protecteur Syldave (le GPS). Voici le tableau donnant le sexe des personnes emprisonnées :

b. in AMYOTTE, LUC *Méthodes quantitatives*. ERPI, Québec, 2002

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m | f | m | m | m | f | f | f | m | f | m | m | m | m | f | f | f | m | m | f | f | f | f | m | f |
| f | m | m | m | m | m | f | m | f | f | m | f | f | f | f | m | m | m | f | f | f | m | m | f | f |
| f | f | f | f | f | f | f | f | m | m | m | f | m | m | f | m | f | m | m | m | f | m | m | m | f |

FIGURE 1 – Sexe des 75 personnes ayant été emprisonnées (Source : ministère syldave de la justice - BO n°57 p. 42)

Cette série comporte 75 données qui prennent deux modalités : masculin ou féminin. Ce tableau est peu pratique à interpréter sous cette forme. Mieux vaut dresser un **tableau des effectifs** :

| Sexe | Nombre d'emprisonnés |
|----------|----------------------|
| Masculin | |
| Féminin | |
| Total | 75 |

FIGURE 2 – Répartition des 75 emprisonnés selon le sexe (Source : ministère syldave de la justice - BO n°57 p. 42)

Il peut même être encore plus parlant de dresser le **tableau des fréquences** :

| Sexe | Pourcentage d'emprisonnés (%) |
|----------|-------------------------------|
| Masculin | |
| Féminin | |
| Total | 100 |

FIGURE 3 – Répartition en pourcentage des 75 emprisonnés selon le sexe (Source : ministère syldave de la justice - BO n°57 p. 42)

C1b Représentations graphiques

Les résultats d'un sondage sur l'opinion qu'ont les Syldaves sur leur Grand Protecteur ont été regroupés dans le tableau suivant :

| Degré de satisfaction | Pourcentage de la population |
|-----------------------|------------------------------|
| Extrêmement satisfait | 50,9 % |
| Tout à fait satisfait | 34,3 % |
| Totalement satisfait | 10,9 % |
| Très satisfait | 3,9 % |
| Total | 100 % |

FIGURE 4 – Répartition en pourcentage des Syldaves selon leur degré de satisfaction à l'égard du GPS (Source : ministère syldave de la justice - BO n°58 p. 554)

Diagramme en bâtons En face de chaque modalité représentée régulièrement en abscisses on trace un « bâton » de longueur proportionnelle à la fréquence :

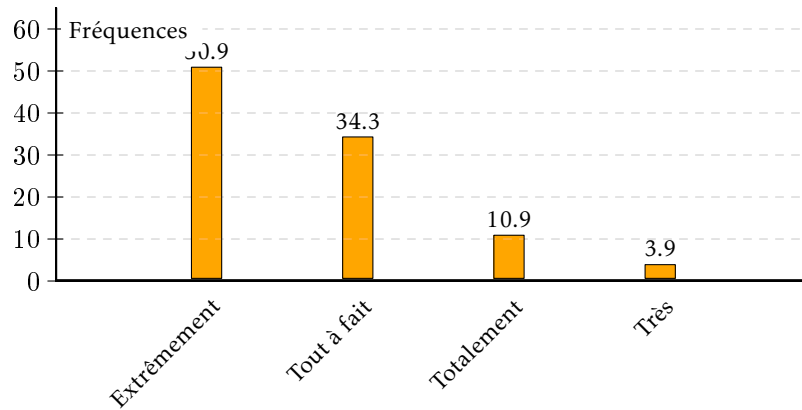


FIGURE 5 – Répartition en pourcentage des Syldaves selon leur degré de satisfaction à l'égard du GPS (Source : ministère syldave de la justice - BO n°58 p. 554) - Diagramme en bâtons

Diagramme à secteurs (« camembert ») On découpe un disque en autant de secteurs qu'il y a de modalités. La part de chaque secteur correspond à la fréquence relative de la modalité qu'il représente, sachant que 100 % correspond bien sûr à 360°.

On remplit alors le tableau suivant :

| Degré de satisfaction | Pourcentage de la population | Angle |
|-----------------------|------------------------------|-------|
| Extrêmement satisfait | 50,9 % | |
| Tout à fait satisfait | 34,3 % | |
| Totalement satisfait | 10,9 % | |
| Très satisfait | 3,9 % | |
| Total | 100 % | 360° |

FIGURE 6 – Répartition en pourcentage des Syldaves selon leur degré de satisfaction à l'égard du GPS (Source : ministère syldave de la justice - BO n°58 p. 554) - Calcul des angles du diagramme circulaire

Et on obtient le camembert suivant :

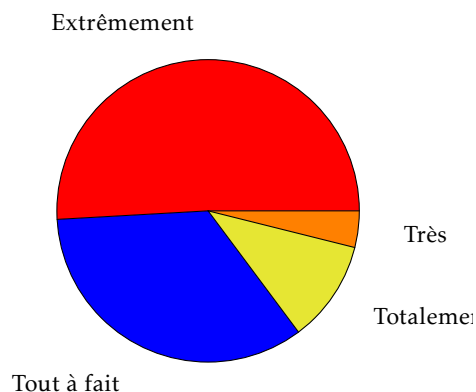


FIGURE 7 – Répartition en pourcentage des Syldaves selon leur degré de satisfaction à l'égard du GPS (Source : ministère syldave de la justice - BO n°58 p. 554) - Diagramme circulaire

On peut préférer d'ailleurs un demi-camembert en remplissant le tableau précédent avec 180° comme angle correspondant à 100 % :

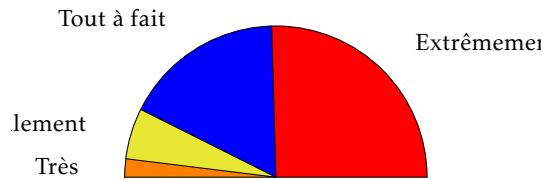


FIGURE 8 – Répartition en pourcentage des Syldaves selon leur degré de satisfaction à l'égard du GPS (Source : ministère syldave de la justice - BO n°58 p. 554) - Diagramme semi-circulaire

C2 Variables quantitatives discrètes (grand nombre de valeurs) ou continues

C2 a Groupement par classes

Voici un tableau donnant le nombre annuel de disques vendus par la femme du Grand Protecteur de la Syldavie de 1960 à 2008 :

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| 163 | 169 | 183 | 224 | 140 | 163 | 172 |
| 104 | 121 | 141 | 139 | 135 | 126 | 163 |
| 127 | 141 | 136 | 95 | 128 | 95 | 133 |
| 92 | 119 | 103 | 150 | 149 | 110 | 148 |
| 153 | 156 | 151 | 109 | 143 | 123 | 75 |
| 189 | 126 | 173 | 171 | 128 | 153 | 142 |
| 148 | 112 | 112 | 150 | 121 | 162 | 120 |

FIGURE 9 – Nombre annuel de disques vendus par la femme du Grand Protecteur de la Syldavie de 1960 à 2008 (Source : ministère syldave de la culture - BO n°51 p. 1254559)

Il est totalement inutile de comptabiliser les occurrences d'un nombre donné de vente. Il est beaucoup plus parlant de regrouper les résultats par classe.

Nous pouvons par exemple ici regrouper les modalités dans des intervalles de largeur 25 en commençant par 75.

Par convention, on utilise des intervalles fermés à gauche et ouverts à droite. Ici, on dénombre 4 modalités m telles que $75 \leq m < 100$. On remplit alors le tableau suivant :

| Ventes annuelles de disques | Nombres d'années |
|-----------------------------|------------------|
| [75; 100[| 4 |
| [100; 125[| |
| [125; 150[| |
| [150; 175[| |
| [175; 200[| |
| [200; 225[| |
| Total | 49 |

FIGURE 10 – Répartition de 49 années selon le nombre annuel de disques vendus par la femme du Grand Protecteur de la Syldavie de 1960 à 2008 (Source : ministère syldave de la culture - BO n°51 p. 1254559)

On peut dresser le même type de tableau avec les fréquences :

| Ventes annuelles de disques | Pourcentage d'années |
|-----------------------------|----------------------|
| [75; 100[| |
| [100; 125[| |
| [125; 150[| |
| [150; 175[| |
| [175; 200[| |
| [200; 225[| |
| Total | 100% |

FIGURE 11 – Répartition en pourcentage de 49 années selon le nombre annuel de disques vendus par la femme du Grand Protecteur de la Syldavie de 1960 à 2008 (Source : ministère syldave de la culture - BO n°51 p. 1254559)

C2 b Histogramme avec classes de même amplitude

Ici, on va représenter dans un repère les données d'un des tableaux précédents. Comme les classes sont de même amplitude, cela ressemblera à un diagramme en bâtons... dont les bâtons sont collés.

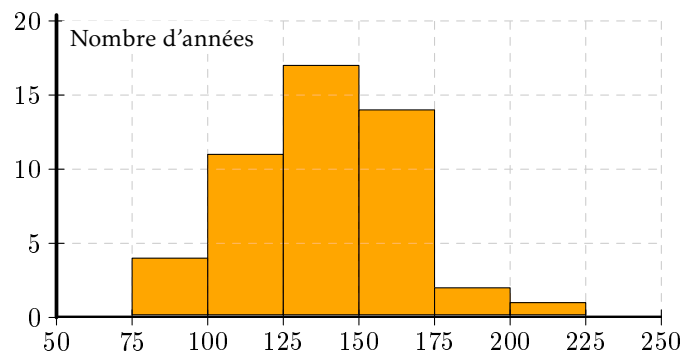


FIGURE 12 – Répartition de 49 années selon le nombre annuel de disques vendus par la femme du Grand Protecteur de la Syldavie de 1960 à 2008 (Source : ministère syldave de la culture - BO n°51 p. 1254559) - Histogramme

C2 c Histogramme avec classes d'amplitudes différentes

Considérez le tableau suivant :

| Âge | Nombre de femmes |
|-------|------------------|
| 15-20 | 202 |
| 20-25 | 204 |
| 25-35 | 359 |
| 35-45 | 338 |
| 45-65 | 304 |
| 65-75 | 10 |
| Total | 1 414 |

FIGURE 13 – Répartition des femmes du harem du GPS selon l'âge (Source : ministère syldave du travail- BO n°7)

Les classes ici ne sont plus de même amplitude. L'usage est de représenter chaque classe par des rectangles d'aires pro-

proportionnelles aux effectifs de chaque classe. Par exemple, la 3^e classe a un plus grand effectif mais est deux fois plus étendue. Elle est donc équivalente à deux classes d'amplitude 5 et d'effectif $\frac{359}{2}$.
On appelle **densité** le rapport

$$\frac{\text{Effectif}}{\text{Amplitude}}$$

Ce résultat est équivalent à

$$\text{Effectif} = \text{Amplitude} \times \text{Densité}$$

L'effectif correspond donc à l'aire du rectangle :

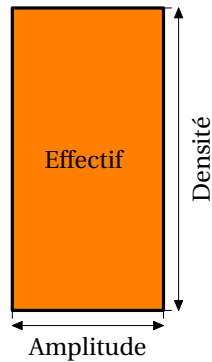


FIGURE 14 – Construction d'un histogramme à pas non constant

On remplit donc le tableau suivant :

| | | | | | | |
|-------------|----------|----------|----------|--|--|--|
| Intervalles | [15; 20[| [20; 25[| [25; 35[| | | |
| Effectifs | 202 | 204 | 359 | | | |
| Amplitudes | 5 | 5 | 10 | | | |
| Densités | 40,4 | | | | | |

FIGURE 15 – Répartition des femmes du harem du Grand Protecteur de la Syldavie selon l'âge (Source : ministère syldave du travail- BO n°7) - Tableau de construction de l'histogramme

On en déduit l'histogramme suivant :

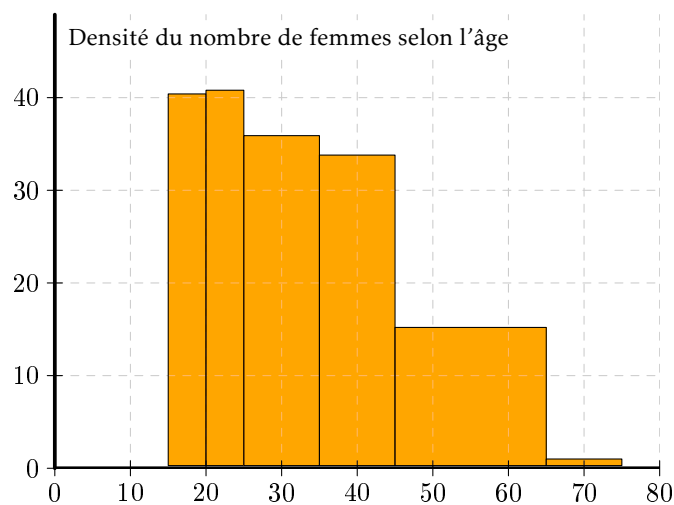


FIGURE 16 – Répartition des femmes du harem du Grand Protecteur de la Syldavie selon l'âge (Source : ministère syldave du travail- BO n°7) -Histogramme

D Tableaux à double entrée

Lorsqu'on effectue une enquête, on peut être amené à recueillir des renseignements sur plus d'une caractéristique des unités statistiques.

On a par exemple étudié la répartition des détenus d'une prison syldave selon l'ancienneté et le sexe :

| Ancienneté \ Sexe | Masculin | Féminin | Total |
|-------------------|----------|---------|-------|
| Moins de 5 ans | 20 | 20 | |
| 5 à 10 ans | 20 | 10 | |
| 10 à 15 ans | 25 | 15 | |
| 15 ans et plus | 75 | 25 | |
| Total | | | |

FIGURE 17 – Répartition des détenus d'une prison syldave selon l'ancienneté et le sexe (Source : ministère syldave des loisirs - BO n°0)

Il est plus parlant de passer aux fréquences :

| Ancienneté \ Sexe | Masculin | Féminin | Total |
|-------------------|----------|---------|-------|
| Moins de 5 ans | | | |
| 5 à 10 ans | | | |
| 10 à 15 ans | | | |
| 15 ans et plus | | | |
| Total | | | 100 % |

FIGURE 18 – Répartition en pourcentage des détenus d'une prison syldave selon l'ancienneté et le sexe (Source : ministère syldave des loisirs - BO n°0)

On aurait pu également répartir les détenus masculins et féminins selon leur ancienneté :

| Ancienneté \ Sexe | Masculin | Féminin |
|-------------------|----------|---------|
| Moins de 5 ans | | |
| 5 à 10 ans | | |
| 10 à 15 ans | | |
| 15 ans et plus | | |
| Total | 100 % | 100 % |

FIGURE 19 – Répartition en pourcentage des détenus d'une prison syldave par sexe selon l'ancienneté (Source : ministère syldave des loisirs - BO n°0)

Ou encore répartir les détenus par catégorie d'ancienneté selon le sexe :

| Ancienneté \ Sexe | Sexe | | Total |
|-------------------|----------|---------|-------|
| | Masculin | Féminin | |
| Moins de 5 ans | | | 100 % |
| 5 à 10 ans | | | 100 % |
| 10 à 15 ans | | | 100 % |
| 15 ans et plus | | | 100 % |

FIGURE 20 – Répartition en pourcentage des détenus d'une prison syldave par catégorie d'ancienneté selon le sexe (Source : ministère syldave des loisirs - BO n°0)

Chacun de ces tableaux met en valeur des phénomènes différents.

E Séries chronologiques

Lorsqu'on étudie l'évolution d'une variable dans le temps, on obtient une série chronologique.

E 1 Mesures sur les séries chronologiques

Considérons par exemple les données suivantes :

| Année | Nombre d'invités à la garden party du GPS |
|-------|---|
| 1987 | 6 491 |
| 1992 | 5 858 |
| 1997 | 5 042 |
| 2002 | 5 418 |
| 2007 | 4 744 |

FIGURE 21 – Évolution du nombre d'invités à la garden party annuelle du GPS (Source : ministère syldave du protocole - BO n°2)

E 1 a Variation absolue



Définition II - 1 : Variation absolue

La variation absolue mesure l'augmentation ou la diminution de la valeur d'une variable entre deux moments donnés. Elle s'exprime dans les mêmes unités que la variable (sauf le cas particulier des pourcentages comme on le verra au E 1 b).

$$\Delta V_a = V(t_2) - V(t_1)$$

Ici, la variation du nombre d'invités entre 1997 et 2007 a été de

$$\Delta V_a = V(2007) - V(1997) =$$

E 1 b Cas particulier des variables exprimées en pourcentage

Voici un tableau donnant l'évolution des suffrages obtenu par le GPS lors des précédentes élections :

| Année | pourcentage des votants ayant voté pour le GPS |
|-------|--|
| 1948 | 99,89% |
| 1968 | 99,91% |
| 1988 | 99,97% |
| 2008 | 99,92% |

FIGURE 22 – Évolution de la proportion des personnes ayant voté pour le GPS lors des 4 dernières élections dans l'ensemble des votants (Source : ministère sylvain des services secrets - BO n°56)

Il serait faux de dire que l'augmentation des suffrages a été de 0,06 % entre 1968 et 1988 car les pourcentages donnés pour ces années correspondent à des nombres de votants sûrement différents. On parle plutôt d'une variations de 0,06 **points**.



variation de pourcentage

Lorsque la variable d'une série chronologique est en pourcentage, la variation s'exprime en POINTS.

E 1 c Variation relative



Définition II - 2 : Variation relative

La variation relative est le pourcentage de diminution ou d'augmentation de la valeur d'une variable entre deux moments donnés par rapport à la valeur initiale :

$$\Delta V_r = \frac{V(t_2) - V(t_1)}{V(t_1)}$$

Elle s'exprime en pourcentage.

Si on reprend le tableau 21 page précédente, la variation relative entre 1997 et 2007 est de :

$$\Delta V_r = \frac{4744 - 5042}{5042} \approx \%$$

On observe donc une baisse de % du nombre d'invités entre 1997 et 2007.



Remarque II - 1 : Variation relative de pourcentages

On peut mesurer des variations relatives de pourcentages car on divise des grandeurs de même nature. Par exemple, si on reprend le tableau 22 et qu'on calcule la variation relative de suffrages entre 1968 et 1988 :

$$\Delta V_r = \frac{99,97 - 99,91}{99,91} \approx \%$$

On observe alors une augmentation relative de % de pourcentage de suffrages en faveur du GPS entre 1968 et 1988.

E 1 d Variation moyenne



Définition II - 3 : Variation moyenne

C'est la variation par unité de temps. Elle s'exprime en unités de la variable par unité de temps. On la calcule grâce à la formule :

$$\Delta V_m = \frac{V(t_2) - V(t_1)}{t_2 - t_1}$$

Reprenons une nouvelle fois le tableau 21 page précédente pour mesurer la variation moyenne du nombre d'invités à la garden party entre 1987 et 2007 :

$$\Delta V_m = \frac{V(2007) - V(1987)}{2007 - 1987} = \frac{4744 - 6491}{20} \approx$$

On observe donc une baisse moyenne de invités par an.

E 1 e Comment commenter une série chronologique ?

On vous demande de commenter le tableau 21 page 9. Voici comment vous pourriez procéder en utilisant ce que nous venons de voir :

En Syldavie (le lieu), entre 1987 et 2007 (la période), le nombre annuel d'invités à la garden party du GPS a diminué de manière importante (le phénomène).

De 6 491 en 1987, le nombre d'invités est passé à 4 744 en 2007 (les valeurs extrêmes).

Ainsi, en 2007, on a dénombré 1 747 invités en moins qu'en 1987 (variation absolue), ce qui représente une baisse moyenne annuelle de 87,4 invités (variation moyenne), soit une baisse de 26,9 % pour cette période (variation relative).

F Mesures de tendance centrale

F 1 Le mode

F 1 a Définition



Définition II - 4 : mode

Mesure qui correspond à la valeur ou à la modalité la plus fréquente. S'il y a plusieurs mode, on dit que la distribution est multimodale.

C'est la mesure la plus simple à évaluer.

F 1 b Données groupées par valeurs ou modalités

Si on reprend l'exemple du tableau 22 page précédente, le mode est « extrêmement satisfait ».

F 1 c Données groupées par classes

Reprenons cette fois le graphique 12 page 6. La classe modale correspond à l'intervalle [125 ; 150[. Si on veut un mode plutôt que la classe modale, on prendra le milieu de la classe modale : ici 137,5.

F 1 d Distribution polymodale

Considérons le nouvel exemple suivant :

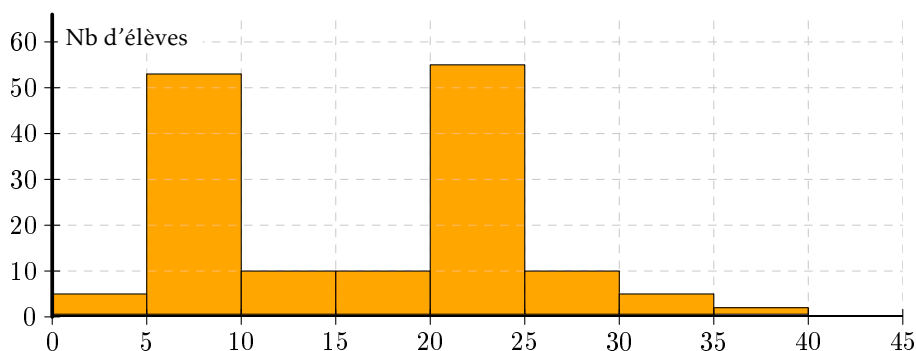


FIGURE 23 – Répartition des prisonniers syldaves selon la durée de leur peine (en années)

Ce graphique présente deux classes modales : $[5; 10[$ et $[20; 25[$.

F2 La médiane

F2 a Conventions

Dans toute la suite, on étudiera une population, notée E , et une variable statistique quantitative X définie sur E .



Exemple II - 1 :

- Si on étudie par exemple le nombre de poupées Barbue que possèdent les élèves de la classe de 1^{ère} ES 1
- E est l'ensemble des élèves de la classe
 - X est la fonction qui, à un élément de E , associe le nombre de poupées Barbue qu'il ou elle possède.

Si E possède n éléments, on notera $V = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ l'ensemble **ordonné par valeurs croissantes** des valeurs prises par X .



Remarque II - 2 :

Notez bien que certains éléments de V peuvent être égaux : en effet, deux élèves différents peuvent avoir le même nombre de poupées Barbue.

F2 b Définition



Définition II - 5 : médiane

La médiane M_e est un nombre tel que :

- **au moins** 50% des éléments de V sont inférieurs à M_e ,
- **au moins** 50% des éléments de V sont supérieurs à M_e .



Exemple II - 2 : calcul de la médiane

- $V = \{10, 20, 20, 50, 50, 80, 80, 90, 370\}$
- ▷ $M_e = 50$
- $V = \{10, 20, 20, 50, 80, 80, 90, 370\}$
- ▷ $M_e = \frac{50 + 80}{2} = 65$

F2 c Cas des données groupées par valeurs

Considérons par exemple la répartition des ministres syldaves selon le nombre d'années qu'ils ont étudié après leur Brevet des Collèges :

| Nombre d'années après le Brevet | Nombre de ministres | Effectifs cumulés croissants |
|---------------------------------|---------------------|------------------------------|
| 0 | 69 | 69 |
| 1 | 31 | 100 |
| 2 | 15 | |
| 3 | 6 | |
| 4 | 3 | |
| 5 | 1 | |

FIGURE 24 – Répartition des ministres syldaves selon le nombre d'années qu'ils ont étudié après leur Brevet des Collèges

Il y a un nombre impair de données (125). La médiane correspond donc à la donnée de rang 63 (62 après, 62 avant). Grâce à la troisième colonne, on trouve que la 63^e valeur vaut 0. Ainsi la médiane $M_e = 0$. On en déduit qu'au moins 50 % des ministres n'ont pas poursuivi d'étude après le Brevet.

Il peut être plus pratique de travailler avec les fréquences :

| Nombre d'années après le Brevet | Nombre de ministres | Pourcentage des ministres | Fréquence | Fréquences cumulées croissantes |
|---------------------------------|---------------------|---------------------------|-----------|---------------------------------|
| 0 | 69 | 69 | | |
| 1 | 31 | 100 | | |
| 2 | 15 | | | |
| 3 | 6 | | | |
| 4 | 3 | | | |
| 5 | 1 | | | |

FIGURE 25 – Répartition en pourcentage des ministres syldaves selon le nombre d'années qu'ils ont étudié après leur Brevet des Collèges

Il suffit alors de regarder à quelle valeur correspond la fréquence cumulée 50 %.

F3 Moyenne

F3a Données non groupées

Soit x_1, x_2, \dots, x_n une série statistique. La moyenne \bar{x} est l'unique valeur que devrait prendre chacune des données pour que la somme des données soit préservée.

En d'autres termes, on cherche \bar{x} tel que

$$x_1 + x_2 + \dots + x_n = \underbrace{\bar{x} + \bar{x} + \dots + \bar{x}}_{n \text{ termes}}$$

On en déduit que :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

On utilise souvent le symbole Σ pour représenter une somme (sigma est la lettre grecque correspondant à notre S).

Ainsi $\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$.

La formule de la moyenne devient donc :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{1} = \frac{1}{n} \sum_{i=1}^n x_i$$

Faut-il vraiment un exemple?...

Bon, voici le relevé des pourcentages de réussite lors du dernier exercice de tir des généraux de l'armée syldave :

12 22 32 2 24 2 5 33

Quel est le pourcentage moyen de réussite ?

F3 b Données groupées par valeurs

Reprenons le tableau 24 page 12. Si nous voulons calculer le nombre moyen d'années d'étude après le brevet des ministres syldaves il faudrait additionner 69 zéros, 31 un, 15 deux, 6 trois, 3 quatre, 1 cinq et diviser le tout par le nombre total de ministres à savoir 125 d'après ce que nous venons de voir.

Nous pouvons simplifier les choses en notant e_i l'effectif de la valeur x_i :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n e_i \cdot x_i$$

Pour les ministres, $\bar{x} = \frac{1}{125} (69 \times 0 + 31 \times 1 + 15 \times 2 + 6 \times 3 + 3 \times 4 + 1 \times 5) = \dots$

F3 c Données groupées par classes

Regrouper des données par classe constitue déjà une approximation mais a le désavantage de nous empêcher d'utiliser nos moyens de calculer la moyenne. Pour y remédier, nous allons effectuer une approximation supplémentaire en prenant comme représentant d'une classe son **milieu**.

Reprenons le tableau 13 page 6 en le complétant :

| Âge | Milieu de classe (m_i) | Nombre de femmes |
|-------|----------------------------|------------------|
| 15-20 | | 202 |
| 20-25 | | 204 |
| 25-35 | | 359 |
| 35-45 | | 338 |
| 45-65 | | 304 |
| 65-75 | | 10 |
| Total | | 1 414 |

FIGURE 26 – Répartition des femmes du harem du Grand Protecteur de la Syldavie selon l'âge avec le milieu des classes

Nous pouvons alors trouver une approximation de l'âge moyen des femmes du Harem :

$$\bar{x} \approx \frac{1}{n} \sum_{i=1}^n m_i e_i = \frac{202 \times 17,5 + \dots}{1414} = \dots$$

G Quartiles et diagrammes en boîte

G1 L'idée

Pour avoir une idée un peu plus précise de la série statistique étudiée, on voudrait séparer notre population en 4 groupes au lieu de 2 comme cela a été fait avec la médiane.

On a donc envie de calculer les médianes des parties basses et hautes.

On a également comme cahier des charges d'avoir au moins 25% des valeurs prises par x inférieures au premier quartile Q_1 et au moins 75% des valeurs prises par x inférieures au troisième quartile Q_3 .

G2 Expérimentons



Exemple II - 3 : huit éléments

$$V = \{10, 20, 30, 40, 50, 60, 70, 80\}$$

On peut séparer l'effectif en quatre groupes de même effectif égal à 25% de l'effectif total donc ça « colle »

$$V = \{10, 20, 30, 40, 50, 60, 70, 80\}$$

On peut prendre

- $Q_1 = \frac{20+30}{2} = 25$ et 25% des effectifs ont une valeur inférieure à Q_1 (10 et 20 c'est-à-dire 2 sur 8)
 - $M_e = \frac{40+50}{2} = 45$ et 50% des effectifs ont une valeur inférieure à M_e (10, 20, 30 et 40 c'est-à-dire 4 sur 8)
 - $Q_3 = \frac{60+70}{2} = 65$ et 75% des effectifs ont une valeur inférieure à Q_3 (6 valeurs sur 8)
- et Q_1 et Q_3 sont bien les médianes respectives des parties basses et hautes.

G3 Définissons



Définition II - 6 : quartiles

Le premier quartile est obtenu en prenant la médiane de la sous-série contenant les observations dont le rang est strictement inférieur à celui de la médiane (*la partie basse*) pour autant qu'au moins 25% des observations soient inférieures ou égales à cette valeur.

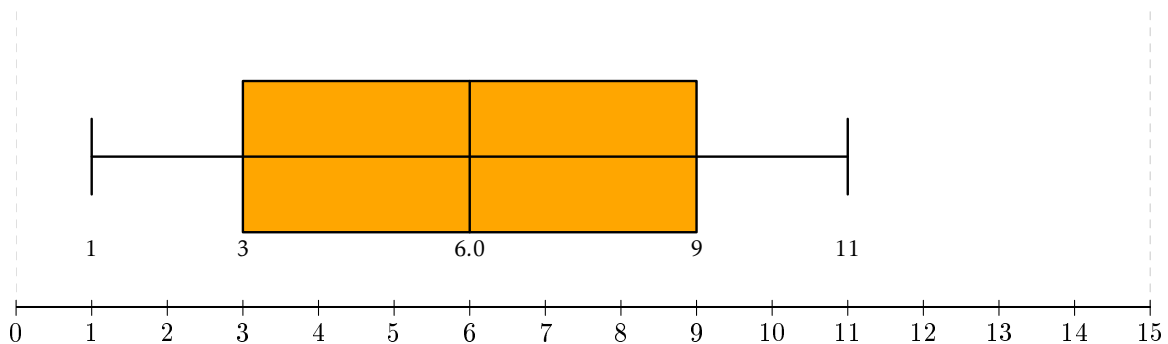
Sinon, il faut inclure la médiane dans la partie basse.

Le troisième quartile est obtenu en prenant la médiane de la sous-série contenant les observations dont le rang est strictement supérieur à celui de la médiane (*la partie haute*) pour autant qu'au moins 75% des observations soient inférieures ou égales à cette valeur.

Sinon, il faut inclure la médiane dans la partie haute.

G4 Boîte à moustache

Après avoir calculé les quartiles, on peut les regrouper dans un tableau. Il est toutefois plus parlant de dresser un diagramme en boîte, ou diagramme de Tukey ou encore boîte à moustache. Reprenons pour cela l'exemple ?? page ??



Ce diagramme est dû au statisticien américain John TUKEY (1915 - 2000) qui a également dit :

Dans un monde où le prix du calcul continue à diminuer rapidement, alors que le prix de la démonstration d'un théorème est stable ou augmente, les principes élémentaires de l'économie indiquent que nous devrions utiliser une part de plus en plus grande de notre temps à faire des calculs.



H Mesures de dispersion

H 1 Écart interquartile

Continuons à exploiter nos quartiles. Environ 50 % de la population a une modalité entre Q_1 et Q_3 : en observant la boîte à moustache décrivant une série statistique, on peut affiner sa description grâce à l'**écart interquartile**.



Définition II - 7 : écart interquartile

La distance entre Q_1 et Q_3 est appelé écart interquartile.

Observons deux exemples :



Exemple II - 4 : comparaison des écarts interquartiles

Voici un tableau donnant les notes au Bac de deux classes :

| Notes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|--------------------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| Effectifs classe A | 0 | 0 | 0 | 1 | 1 | 0 | 2 | 3 | 2 | 9 | 5 | 9 | 5 | 1 | 2 | 1 | 0 | 2 | 1 | 1 | 0 |
| Effectifs classe B | 0 | 5 | 5 | 5 | 5 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 5 | 5 | 0 | 2 | 0 | 5 | 0 | 5 | 0 |

Voici les boîtes à moustaches correspondant :

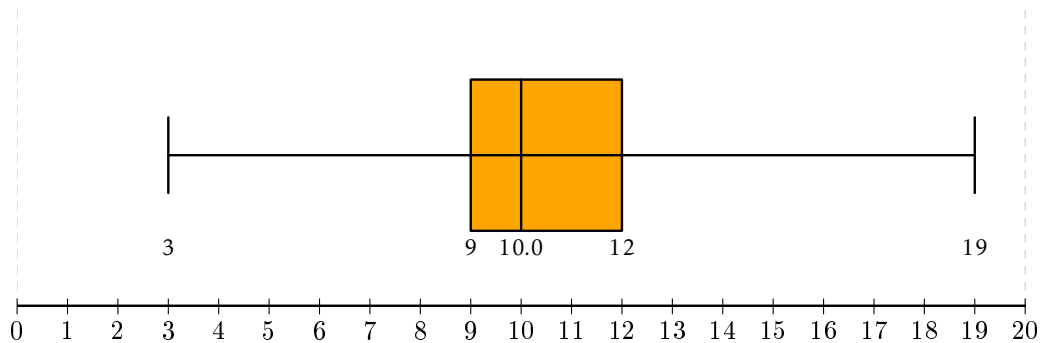


FIGURE 27 – Notes au Bac de la classe A

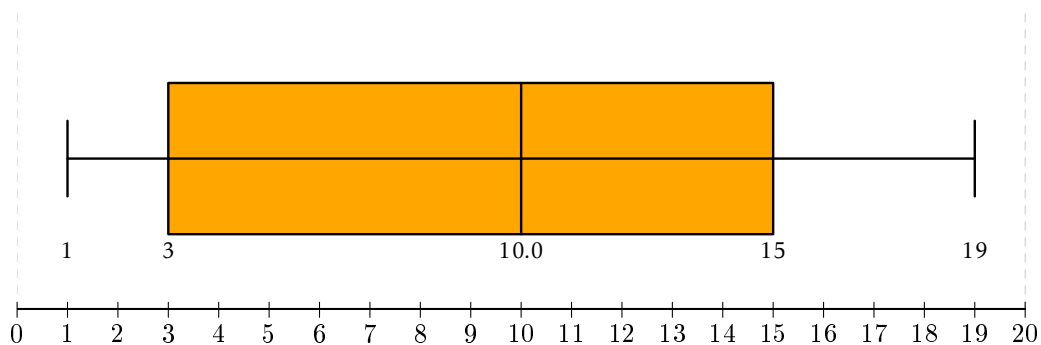


FIGURE 28 – Notes au Bac de la classe B

Dans les deux cas la médiane vaut 10 : cela signifie qu'au moins 50 % des élèves de chaque classe a eu moins de la moyenne.

Cependant, les écarts interquartiles sont nettement différents : $Q_3 - Q_1 = 3$ dans la classe A mais $Q_3 - Q_1 = 12$ dans la classe B. La classe B est nettement plus hétérogène en terme de résultats. En effet, dans la classe A, 50 % environ des élèves ont eu entre 9 et 12 alors que dans la classe B la moitié des notes se situent entre 3 et 15.

H2 Variance et écart-type

La dispersion peut également se mesurer autour de la moyenne.
Considérons une série statistique quelconque :

| | | | | | | |
|-----------------|---|---|---|---|---|---|
| Valeurs x_i | 0 | 1 | 2 | 3 | 4 | 7 |
| Effectifs n_i | 1 | 2 | 2 | 4 | 3 | 2 |

FIGURE 29 – Série statistique quelconque

À partir de ces données, calculez la moyenne \bar{x} puis remplissez le tableau suivant proposant deux façons de « mesurer » pour chaque valeur « l'éloignement » par rapport à \bar{x} .

| | | | | | | |
|---------------------|--|--|--|--|--|--|
| $x_i - \bar{x}$ | | | | | | |
| $(x_i - \bar{x})^2$ | | | | | | |

FIGURE 30 – Essais de mesures de dispersion par rapport à la moyenne

Calculez dans chacun des trois cas l'éloignement moyen, c'est-à-dire la moyenne des écarts...

On peut visualiser les écarts sur le schéma suivant :

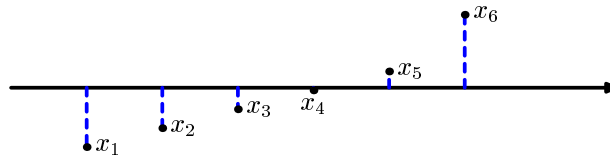


FIGURE 31 – Visualisation de l'écart par rapport à la moyenne

On peut prouver (à titre d'exercice...) que l'écart correspondant à la première ligne du tableau 30 est toujours nul. On préfère donc utiliser la moyenne des écarts de la deuxième ligne que l'on appelle **variance**.



Définition II - 8 : variance

On appelle variance d'une série quelconque à caractère quantitatif discret le nombre :

$$V = \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2 = \sum_{i=1}^p f_i (x_i - \bar{x})^2$$

en notant n_i les effectifs et f_i les fréquences.



Remarque II - 3 : cas d'un regroupement en classe

Dans le cas d'un regroupement en classe, on considère, comme pour le calcul de la moyenne (cf F3 c page 14), le milieu des classes.

La variance est homogène au carré des valeurs x_i : on préfère donc en prendre la racine carrée pour revenir à une grandeur

homogène aux valeurs mesurées. L'ÉCART-TYPE A DONC LA MÊME UNITÉ QUE LA POPULATION!



Définition II - 9 : écart-type

L'écart-type d'une série est la racine carrée de la variance. On le note souvent σ .

Reprenez les séries de l'exemple 4 page 16 et calculez les moyennes et écarts-type. On peut visualiser cette dispersion en traçant un diagramme à bâton et en mettant en évidence l'intervalle $[\bar{x} - \sigma; \bar{x} + \sigma]$:

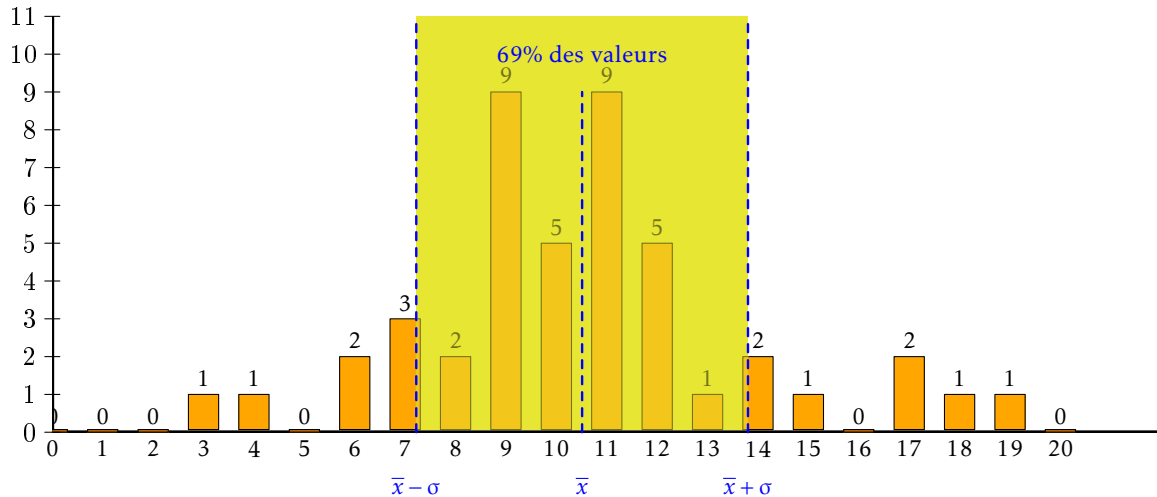


FIGURE 32 – Notes au Bac de la classe A

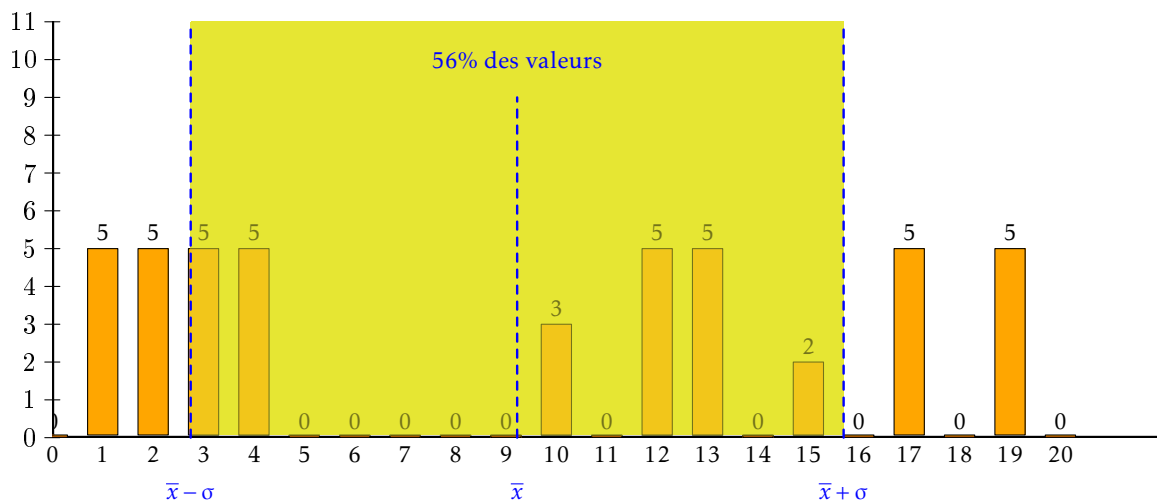


FIGURE 33 – Notes au Bac de la classe B



Propriété II - 1 : Distribution Gaussienne

Dans une distribution Gaussienne (ou normale), de moyenne \bar{x} et d'écart-type σ , on (pas nous) démontre que :

- au moins 68% de la population se trouve dans l'intervalle $[\bar{x} - \sigma; \bar{x} + \sigma]$;
- au moins 95% de la population se trouve dans l'intervalle $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$.

I Utilisation de la calculatrice

11 Casio

Vous rentrez les variables dans la liste 1 et les effectifs ou les fréquences dans la liste 2 puis vous lancez la commande **CALC** en appuyant sur **F2**.

| | List 1 | List 2 | List 3 | List 4 |
|-----|--------|--------|--------|--------|
| SUB | | | | |
| 10 | 6 | | | |
| 11 | 3 | | | |
| 12 | 2 | | | |
| 13 | | | | |

1VAR 2VAR REG SET

Pour vérifier que la machine est bien configurée, il faut aller sur **SET** en appuyant sur **F6**

Il faut placer **List1** devant **1Var XList** et **List2** devant **1Var Freq**.

```
1Var XList :List1
1Var Freq  :1
2Var XList :List1
2Var Freq  :1
2Var XList :List2
2Var Freq  :1
1 LIST
```

On revient à l'écran précédent avec **EXIT**. On lance la commande **1VAR** en tapant sur **F1** et ce qu'on recherche est affiché :

```
1-Variable
Med =3
Q3 =4
maxX =6
Mod =3
Mod:n=1
Mod:F=4
```

et en descendant avec **▼**

```
1-Variable
x̄ =3.16666666
Σx =38
Σx² =142
x̄n =1.34370962
x̄n-1 =1.40345893
n =12
```

12 TI

C'est plus simple...

D'abord, on rentre dans le module de statistique en tapant sur **STAT** et on aperçoit

```

2nd] CALC TESTS
1] Edit...
2] SortA(
3] SortD(
4] ClrList
5] SetUpEditor

```

ensuite on va éditer la liste en appuyant sur **1**.

| L1 | L2 | L3 |
|----------|-------|-------|
| ████████ | ----- | ----- |
| L1(1) = | | |

On entre les variables dans **L1**

| L1 | L2 | L3 |
|---------|----------|-------|
| 30.23 | ████████ | ----- |
| 30.24 | | |
| 30.25 | | |
| 30.26 | | |
| 30.27 | | |
| 30.28 | | |
| 30.29 | | |
| L2(1) = | | |

et les effectifs ou les fréquences dans **L2**. On passe à L2 avec **▶**

| L1 | L2 | L3 |
|---------|----------|----|
| 30.25 | 14 | |
| 30.26 | 22 | |
| 30.27 | 25 | |
| 30.28 | 17 | |
| 30.29 | 12 | |
| 30.3 | 10 | |
| ----- | ████████ | |
| L2(9) = | | |

On quitte ensuite cet écran avec **2nd][QUIT]** puis on retourne dans le module statistique avec **STAT**. Cette fois on va dans **CALC** avec **▶**

```

EDIT 2nd] TESTS
1] 1-Var Stats
2] 2-Var Stats
3] Med-Med
4] LinReg(ax+b)
5] QuadReg
6] CubicReg
7] QuartReg

```

On choisit **1-Var Stats**

```
1-Var Stats █
```

puis on donne la colonne des variables en tapant $\boxed{2nd}[L1]$

```
1-Var Stats L1
```

la virgule pour séparer $\boxed{,}$

```
1-Var Stats L1,
```

puis les fréquences ou les effectifs dans la deuxième colonne $\boxed{2nd}[L2]$

```
1-Var Stats L1,L2
```

et il ne reste plus qu'à taper \boxed{ENTER}

```
1-Var Stats
x̄=30.26608333
Σx=3631.93
Σx²=109924.341
Sx=.0194588915
σx=.0193776432
↓n=120
█
```

et à faire défiler les résultats avec $\boxed{\downarrow}$

```

1-Var Stats
n=120
minX=30.23
Q1=30.25
Med=30.27
Q3=30.28
maxX=30.3

```

J Des exercices de BAC

Exercice 1

1^{re} Partie : Étude de l'évolution du prix du pétrole brut sur une période

La feuille de calcul de l'annexe 2 donne l'évolution entre le 2 mai 2008 (jour de cotation numéro 1) et le 27 juin 2008 (jour de cotation numéro 41) du cours du prix du baril de pétrole brut à New York et du cours de l'euro par rapport au dollar U. S, ce qui correspond à 41 jours de cotation. Les valeurs figurant dans les colonnes B à D sont arrondies au centième.

- Le 2 mai 2008, le prix du baril de pétrole brut à New York était de 112,38 \$.
Le cours de l'euro par rapport au dollar américain était de 1,55; cela signifie qu'un euro valait ce jour là 1,55 dollar américain.
 - Justifier que, le 2 mai 2008, le prix du baril était de 72,50 €.
 - Quelle formule alors saisir dans la cellule D3 pour obtenir, par recopie automatique vers le bas, les prix en euros du baril de pétrole ?
- On répondra aux deux questions suivantes en utilisant les données de l'annexe 2.
Par quel coefficient, arrondi au millième, a été multiplié :
 - le prix du baril de pétrole en dollars entre le 2 mai 2008 et le 27 juin 2008 ? Arrondir à 0,001.
 - le prix du baril de pétrole en euros entre le 2 mai 2008 et le 27 juin 2008 ? Arrondir à 0,001.
- Déduire de la question 2. les pourcentages d'évolution, arrondis au dixième, du prix du baril de pétrole en dollars, puis du prix du baril de pétrole en euros pendant la période du 2 mai 2008 au 27 juin 2008. *Arrondir à 0,1 %*
 - Comment expliquer le fait que ces pourcentages sont différents ?

2^e Partie : Étude statistique :

- On considère la série statistique construite avec les prix du baril de pétrole, arrondis à l'euro, entre le 2 mai 2008 et le 27 juin 2008 (41 journées de cotation).

Le tableau ci-dessous donne la répartition de ces prix arrondis à l'euro :

| Prix | 73 | 76 | 77 | 79 | 80 | 81 | 82 | 83 | 85 | 86 | 87 | 88 | 89 |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Effectifs | 1 | 1 | 1 | 2 | 3 | 6 | 7 | 3 | 3 | 5 | 2 | 4 | 3 |

- Déterminer la médiane, le 1^{er} quartile et le 3^e quartile de cette série.
 - Construire le diagramme en boîtes de cette série sur l'**annexe 1** (à rendre avec la copie).
- Quel était le prix moyen, arrondi à l'euro, du baril de pétrole durant cette période ?

Annexe 2

| | A | B | C | D |
|----|---------------|---------------|--------------|---------------|
| 1 | Journée | Prix du baril | Cours de | Prix du baril |
| 2 | numéro | en \$ | l'euro en \$ | en € |
| 3 | 1 (2/05/08) | 112,38 | 1,55 | 72,50 |
| 4 | 2 | 116,43 | 1,54 | 75,60 |
| 5 | 3 | 120,1 | 1,55 | 77,48 |
| 6 | 4 | 121,89 | 1,55 | 78,64 |
| 7 | 5 | 123,79 | 1,54 | 80,79 |
| 8 | 6 | 124,41 | 1,54 | 80,79 |
| 9 | 7 | 126,16 | 1,55 | 81,39 |
| 10 | 8 | 123,88 | 1,55 | 79,92 |
| 11 | 9 | 125,84 | 1,55 | 81,39 |
| 12 | 10 | 123,88 | 1,55 | 79,92 |
| 13 | 11 | 124,3 | 1,54 | 80,71 |
| 14 | 12 | 126,35 | 1,56 | 80,99 |
| 15 | 13 | 126,83 | 1,55 | 81,83 |
| 16 | 14 | 129 | 1,57 | 82,17 |
| 17 | 15 | 133,89 | 1,58 | 83,41 |
| 20 | 18 | 128,63 | 1,57 | 81,93 |
| 21 | 19 | 130,75 | 1,58 | 82,75 |
| 22 | 20 | 126,54 | 1,55 | 81,64 |
| 23 | 21 | 127,76 | 1,56 | 81,90 |
| 24 | 22 | 1027,6 | 1,55 | 82,32 |
| 25 | 23 | 124,3 | 1,54 | 79,30 |
| 27 | 25 | 128,18 | 1,56 | 82,17 |
| 28 | 26 | 137,86 | 1,58 | 87,25 |
| 29 | 27 | 134,66 | 1,56 | 84,85 |
| 31 | 29 | 136,75 | 1,56 | 87,66 |
| 32 | 30 | 136,72 | 1,54 | 87,48 |
| 33 | 31 | 134,72 | 1,54 | 87,48 |
| 34 | 32 | 133,94 | 1,55 | 86,41 |
| 35 | 33 | 133,8 | 1,55 | 86,32 |
| 36 | 34 | 136,27 | 1,55 | 87,92 |
| 37 | 35 | 131,82 | 1,55 | 88,55 |
| 38 | 36 | 134,84 | 1,56 | 86,44 |
| 39 | 37 | 137,25 | 1,55 | 88,55 |
| 40 | 38 | 137,06 | 1,56 | 87,86 |
| 41 | 39 | 134,6 | 1,57 | 85,73 |
| 42 | 40 | 139,67 | 1,58 | 88,40 |
| 43 | 41 (27/06/08) | 140,7 | 1,58 | 89,05 |

Exercice 2

Un directeur de supermarché décide d'étudier le temps d'attente aux caisses de son établissement pour ajuster le nombre de caisses ouvertes à la demande. Pour cela, il interroge le lundi et le vendredi cent clients et note les temps d'attente approximatifs en minutes entières.

PARTIE A : Étude de l'échantillon du lundi

Le lundi, il obtient la répartition suivante :

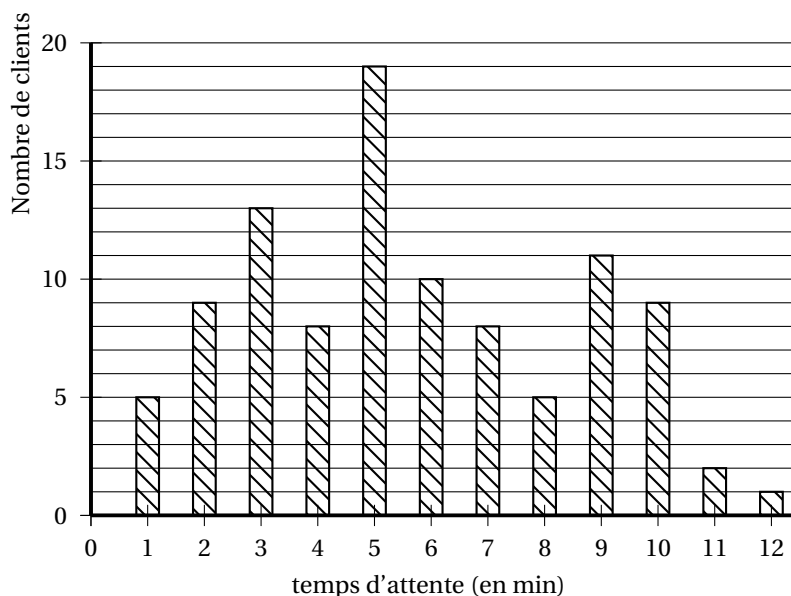
| Temps d'attente en caisse (en min) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------------------------|----|----|----|---|----|---|----|---|---|----|
| Nombre de clients | 14 | 13 | 23 | 9 | 14 | 8 | 12 | 4 | 1 | 2 |

- Calculer le temps moyen d'attente aux caisses du supermarché pour l'échantillon étudié.
- Déterminer la médiane et les quartiles de la série statistique des temps d'attente.
- Construire sur la feuille annexe 2 le diagramme en boîte de cette série.
- a) Son adjoint souhaite ouvrir une caisse supplémentaire si plus de 15 % des clients attendent 7 min ou plus en caisse. Doit-il ouvrir une nouvelle caisse le lundi ? (On justifiera la réponse).
- b) Le directeur décide d'ouvrir une caisse supplémentaire si le temps moyen d'attente aux caisses dépasse 5 min. Doit-il ouvrir une nouvelle caisse le lundi ? (On justifiera la réponse).

PARTIE B : Étude de l'échantillon du vendredi

Le directeur décide de comparer les temps d'attente en début et en fin de semaine. Il a donc relevé le vendredi les temps d'attente aux caisses d'un échantillon de cent clients et obtient les résultats résumés dans le diagramme donné ci-dessous :

Temps d'attente le vendredi



- Par lecture du diagramme, compléter le tableau donné en annexe 2.
- Calculer le temps moyen d'attente aux caisses du supermarché le vendredi pour l'échantillon étudié (arrondi au dixième).

PARTIE C : Comparaison des deux échantillons

On a construit dans l'annexe 2 le diagramme en boîte de la série des temps d'attente aux caisses le vendredi. Dans un questionnaire, les clients qualifient d'acceptable un temps d'attente compris entre 2 et 6 minutes inclus.

Pour chacune des affirmations suivantes, dire si elle est vraie ou fausse en justifiant la réponse.

Affirmation A :

Le vendredi, la moitié des clients attendent cinq min ou plus de cinq min en caisse.

Affirmation B :

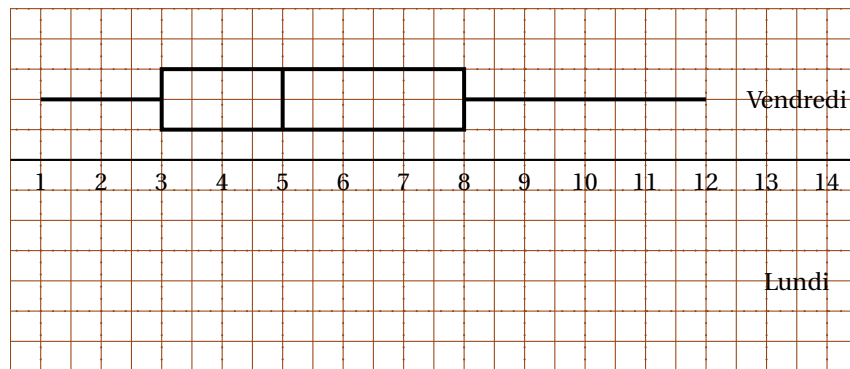
Le vendredi, un quart des clients attend au plus trois minutes en caisse.

Affirmation C :

Ily a autant de clients qui trouvent le temps d'attente acceptable le lundi que le vendredi.

À rendre avec la copie

Diagramme en boîte des séries



Partie B

Tableau de la série du vendredi

| | | | | | | | | | | | | |
|------------------------------------|---|---|---|---|---|----|---|---|---|----|----|----|
| Temps d'attente en caisse (en min) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Nombre de clients | 5 | 9 | | 8 | | 10 | | | | 9 | 2 | 1 |

Exercice 3

Une entreprise qui produit du chocolat, fabrique des tablettes de 100 grammes. Au début de l'année 2010, elle décide de prélever un échantillon dans sa production afin d'en vérifier la masse.

Les résultats sont consignés en **annexe 2**.

1. a) Calculer la masse moyenne μ , exprimée en grammes, des tablettes de cet échantillon. (Arrondir au dixième)
- b) On admet que l'écart-type σ de cette série est environ égal à 1,6.
Déterminer le pourcentage des tablettes de chocolat dont la masse est dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$.
Ce résultat est-il en cohérence avec un modèle gaussien ? Expliquer pourquoi.
2. a) Déterminer la médiane et les quartiles de l'échantillon 2010.
- b) Dessiner le diagramme en boîte correspondant sur **l'annexe 3** en dessous de l'axe.
Vous placerez en extrémités les valeurs minimum et maximum de la série.
- c) Un échantillon de même taille a été prélevé fin 2009, son diagramme en boîte se trouve également en annexe 3.
Donner les valeurs du minimum, du maximum, des quartiles et de la médiane de l'échantillon 2009.

3. Les assertions suivantes sont-elles vraies ou fausses ? Justifier la réponse.

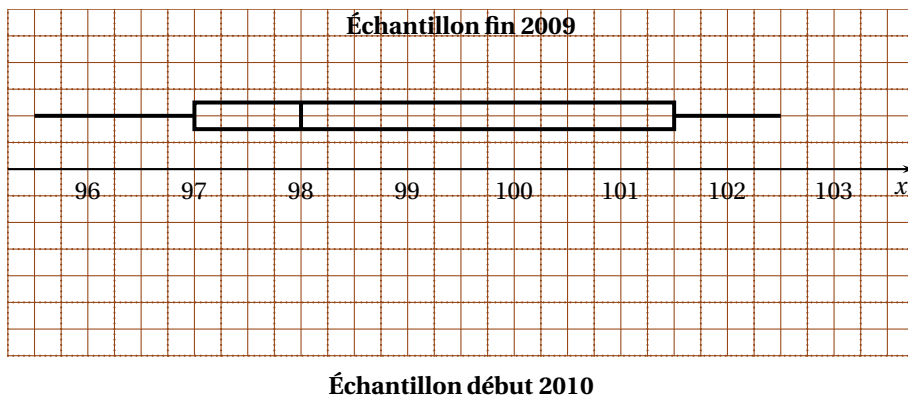
- En fin 2009, environ trois-quarts des tablettes de chocolats avaient une masse supérieure à 98 g.
- L'écart interquartile a été réduit de plus de moitié entre fin 2009 et début 2010.
- Le consommateur qui achète des tablettes produites par cette entreprise en fin 2009 peut se sentir lésé.

ANNEXE 2

Masse des tablettes de chocolat

| | | | | | | | | |
|--------------------|----|----|----|----|-----|-----|-----|-----|
| Masse (en grammes) | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 |
| Effectif | 5 | 6 | 9 | 13 | 32 | 16 | 5 | 4 |

ANNEXE 3



Exercice 4

Un site de vente de livres par Internet désire réaliser une étude statistique de sa clientèle, afin de prévoir l'évolution de ses ventes pour les années à venir.

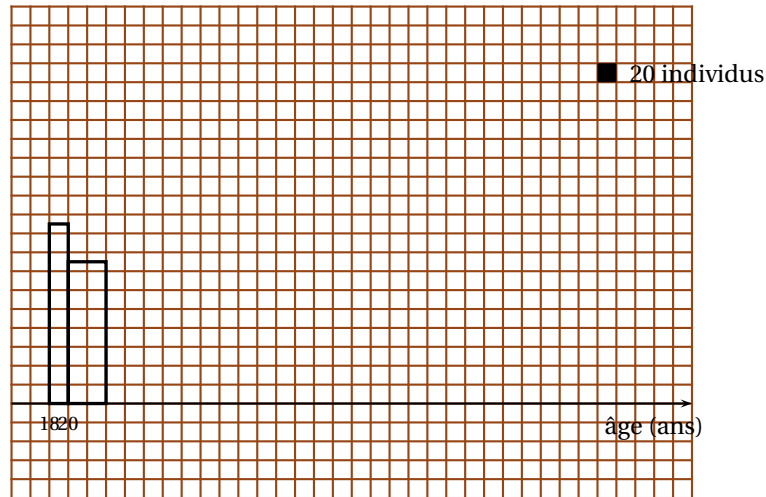
L'étude concerne l'âge de la clientèle. Pour répondre à cette question, les responsables de l'étude utilisent un échantillon de 2 100 clients, parmi les plus réguliers du site.

Les résultats sont résumés dans le tableau ci-dessous :

| Classe d'âge | [18;20[| [20;24[| [24;30[| [30;36[| [36;46[| [46;56[| 56 ans et + | total |
|--------------|---------|---------|---------|---------|---------|---------|-------------|-------|
| Effectif | 190 | 300 | 360 | 450 | 400 | 200 | 200 | 2 100 |

- On assimilera la dernière classe d'âge à l'intervalle [56; 76[.
 - On fera l'hypothèse de l'uniforme répartition de l'effectif dans chaque classe d'âge.
- À l'aide du quadrillage figurant en **annexe 2**, représenter ces données par un histogramme où un carreau représente 20 individus.
 - En utilisant les centres des classes pour valeurs du caractère, déterminer l'âge moyen m et l'écart type s de la série. On ne demande pas de justification, les valeurs seront arrondies au dixième.
 - Hachurer clairement l'histogramme pour faire apparaître l'effectif correspondant à la classe d'âge $[m - s ; m + s]$. Calculer le pourcentage de clients de cette classe d'âge par rapport à l'effectif de l'échantillon (arrondir à 1 %).

Annexe 2 à rendre avec la copie



Aide à la construction de l'histogramme (utilisation non obligatoire et donc tableau non évalué)

| Classe d'âge | [18 ; 20[| [20 ; 24[| [24 ; 30[| [30 ; 36[| [36 ; 46[| [46 ; 56[| [56 ; 76[|
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Effectif | 190 | 300 | 360 | 450 | 400 | 200 | 200 |
| Nombre de carreaux | 9,5 | 15 | | | | | |
| Largeur | 1 | 2 | | | | | |
| Hauteur | 9,5 | 7,5 | | | | | |

Exercice 5

Une association organise chaque année un tournoi comprenant à la fois des épreuves physiques et intellectuelles. Les inscriptions se font par équipe.

PARTIE 1 : étude de deux équipes

L'équipe des « lucioles » participe à cette manifestation ; les âges de ses 24 membres sont :

10 - 10 - 12 - 13 - 14 - 15 - 15 - 17 - 21 - 25 - 33 - 33 - 33 - 34 - 35 - 36 - 40 - 44 - 44 - 44 - 45 - 46 - 59 - 60

Dans cette partie 1, on ne demande pas de justifier.

- Donner l'âge moyen des « lucioles » (arrondir à l'unité).
- Donner la médiane Me , le premier quartile Q_1 et le troisième quartile Q_3 de cette équipe.
- Tracer sur l'**annexe 1** le diagramme en boîte des âges des membres de cette équipe (comme pour le diagramme déjà tracé, les extrémités représentent le minimum et le maximum de la série).
- On a représenté sur l'**annexe 1** le diagramme en boîte des âges des membres de l'équipe des « gazelles ». On considère dans cette question qu'une équipe est « jeune » si au moins 50 % de ses membres ont moins de 30 ans. Pour chacune des propositions suivantes, indiquer si elle est vraie, fausse ou indéterminée (indéterminée signifie ici que les informations données ne permettent pas de conclure si la proposition est vraie ou fausse).
Pour cette question : on attribue 0,5 point par réponse exacte, on retranche 0,25 point par réponse inexacte, l'absence de réponse n'est pas pénalisée. Un éventuel total négatif est ramené à 0.
 - Aucune gazelle n'a plus de 65 ans.
 - Au moins la moitié des gazelles a plus de 25 ans.
 - L'âge moyen des gazelles est compris entre 25 ans et 30 ans.
 - L'équipe « les gazelles » est jeune.
 - L'équipe « les lucioles » est jeune.

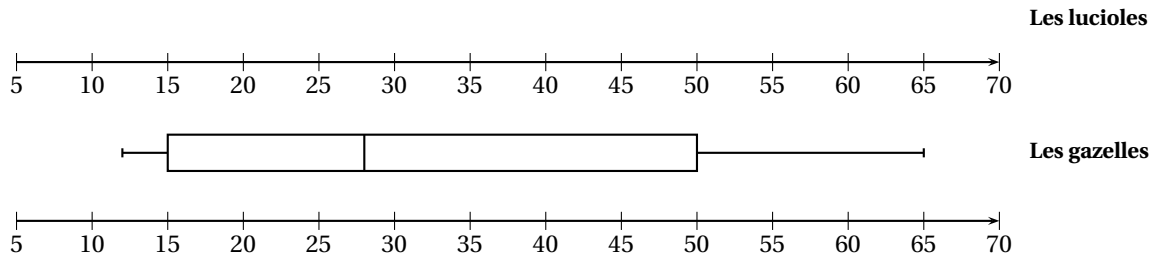
PARTIE 2 : étude de l'ensemble des participants

L'**annexe 2** représente une page automatisée de calcul :

- Le tableau 1 de l'**annexe 2** donne les effectifs des participants en fonction de l'âge et du sexe.
- Le tableau 2 de l'**annexe 2**, calculé automatiquement à partir du tableau 1, indique les pourcentages par rapport au total général.
La plage de cellules C11 : J13 est au format pourcentage arrondi à deux décimales.

1. Compléter la case D11.
2. Donner une formule qui, écrite dans la cellule C11 puis recopiée dans les cellules de la plage C11 : J13 du tableau 2 permet d'obtenir les pourcentages indiqués.
3. a) Calculer le pourcentage des femmes parmi les participants de la tranche 36-45 ans.
b) Donner précisément la signification du pourcentage 12,83 % figurant dans la cellule F12.
4. Dans le tableau 3 de l'**annexe 2**, les cellules de la plage C18 : J20 sont au format pourcentage.
Dans la cellule C18, on entre la formule =C4/C\$6 puis on recopie cette formule dans les cellules de la plage C18 : J20.
Répondre aux questions suivantes sans justifier :
a) Quelle est la signification du pourcentage calculé en C18?
b) Quelle formule obtient-on en D19?

ANNEXE 1



ANNEXE 2

| | A | B | C | D | E | F | G | H | I | J | K |
|----|---|-------|--------|--------|--------|------------------|--------|--------|------------|---------|---|
| 1 | | | | | | | | | | | |
| 2 | | | | | | TABLEAU 1 | | | | | |
| 3 | | | 0-15 | 16-25 | 26-35 | 36-45 | 46-55 | 56-65 | 66 et plus | Total | |
| 4 | | Homme | 75 | 124 | 175 | 251 | 193 | 99 | 25 | 942 | |
| 5 | | Femme | 46 | 83 | 135 | 203 | 73 | 83 | 17 | 640 | |
| 6 | | Total | 121 | 207 | 310 | 454 | 266 | 182 | 42 | 1582 | |
| 7 | | | | | | | | | | | |
| 8 | | | | | | | | | | | |
| 9 | | | | | | TABLEAU 2 | | | | | |
| 10 | | | 0-15 | 16-25 | 26-35 | 36-45 | 46-55 | 56-65 | 66 et plus | Total | |
| 11 | | Homme | 4,74 % | | 11,06% | 15,87% | 12,20% | 6,26% | 1,58% | 59,54% | |
| 12 | | Femme | 2,91% | 5,25% | 8,53% | 12,83% | 4,61% | 5,25% | 1,07% | 40,46% | |
| 13 | | Total | 7,65% | 13,08% | 19,60% | 28,70% | 16,81% | 11,50% | 2,65% | 100,00% | |
| 14 | | | | | | | | | | | |
| 15 | | | | | | | | | | | |
| 16 | | | | | | TABLEAU 3 | | | | | |
| 17 | | | 0-15 | 16-25 | 26-35 | 36-45 | 46-55 | 56-65 | 66 et plus | Total | |
| 18 | | Homme | | | | | | | | | |
| 19 | | Femme | | | | | | | | | |
| 20 | | Total | | | | | | | | | |
| 21 | | | | | | | | | | | |

Exercice 6

Dans cet exercice, les parties A et B sont indépendantes

PARTIE A**4 points**

Marc et Karim préparent leur future course en montagne.

Ils se sont procurés une carte avec courbes de niveau (donnée en **Annexe 1**) sur laquelle ils ont tracé le parcours prévu : le départ se fera du point A, pour aller ensuite jusqu'au point B et arriver en C.

1. Karim affirme : « Là où le trajet est en pente raide, les courbes sont rapprochées ». A-t-il raison ? Justifier votre réponse.
2. Que penser de la pente du terrain entre A et B par rapport à celle entre B et C ?
3. Dans le repère donné en **Annexe 2**, A est le point de départ de coordonnées (0 ; 660). Dans ce repère placer les points A, B et C du parcours de Marc et Karim. Quelle information cette représentation permet-elle d'obtenir sur le trajet ?

PARTIE B

4 points

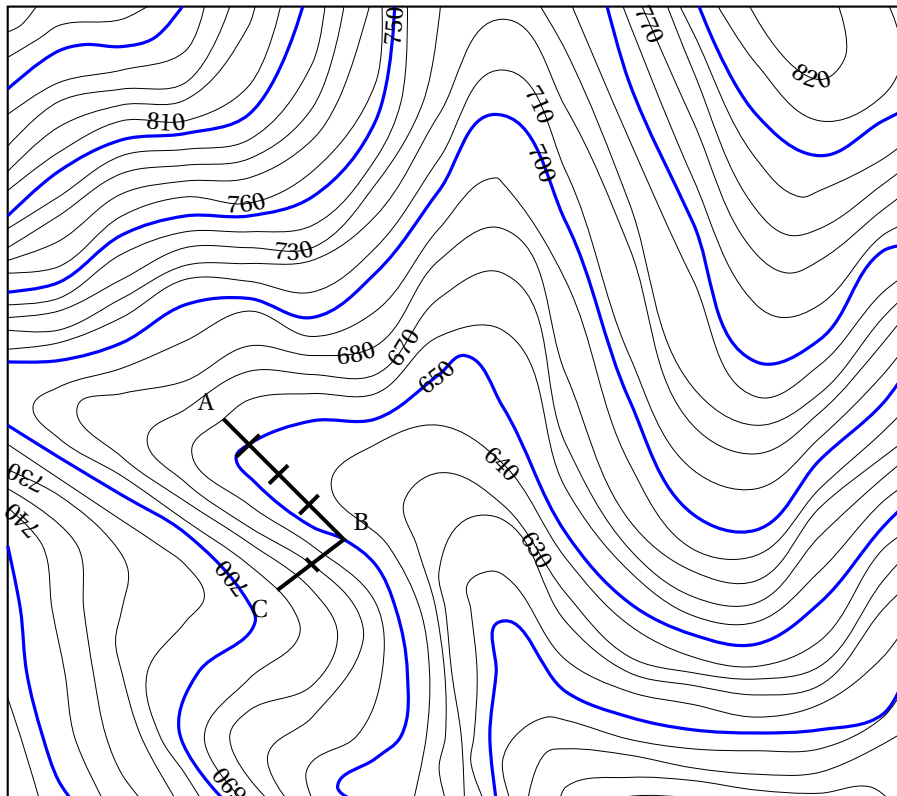
Pendant leur course, Marc et Karim prennent leur rythme cardiaque à intervalles de temps réguliers. Voici le relevé des valeurs obtenues par Karim (rangées dans l'ordre croissant) qui donnent le nombre de battements cardiaques par minute :

59 - 83 - 95 - 98 - 102 - 107 - 116 - 125 - 128 - 131 - 134 - 137 - 140 -
140 - 140 - 145 - 149 - 151 - 152 - 154 - 158 - 162 - 164 - 173 - 176 - 178.

1. Déterminer la médiane, le premier et le troisième quartile de cette série statistique.
2. Construire sur l'**Annexe 2**, (à rendre avec la copie), le diagramme en boîte de cette série statistique.
3. Sachant qu'un entraînement régulier permet de réduire le rythme cardiaque pour un même effort, lequel de Marc et de Karim vous semble le mieux entraîné ? Argumenter la réponse à l'aide des diagrammes en boîte.

ANNEXE 1

Carte servant à préparer la course de Marc et de Karim



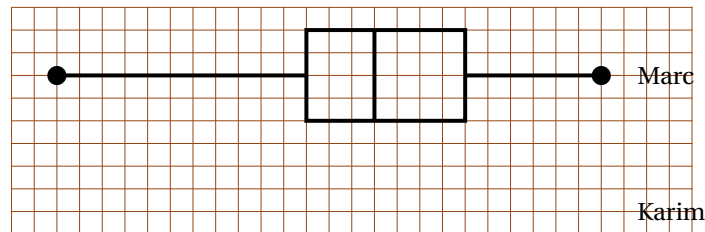
ANNEXE 2

À rendre avec la copie

Repère donnant l'altitude en fonction de la distance parcourue depuis le début de la course



Diagramme en boîte des séries composées de relevés des battements cardiaques lors de la course



Références

- SOUS LA DIRECTION DE JEAN-PIERRE BEAUD ET JEAN-GUY PRÉVOST : L'ère du chiffre. Presses de l'Université du Québec, 2000, Systèmes statistiques et traditions nationales
- AMYOTTE, LUC: Méthodes quantitatives. ERPI, Québec, 2002
- GAUVRIT, NICOLAS: Statistiques Méfiez-vous! Paris: Ellipses, 2007
- HENRY, VALÉRY: Une étude sur les quartiles d'une série statistique univariée. dans Bulletin Vert. APMEP, 2006, n°464
- MALIK, DANIEL: La calculatrice TI-82 Stats.fr. Mai 2008, Lycée La Ramée <URL: http://etablissements.ac-amiens.fr/0020049t/Beespip/rubrique.php3?id_rubrique=129/>
- VERGÈS, DENIS: Annales. 2011, APMEP <URL: <http://www.apmep.asso.fr/-Premiere-L->>
- URBANA-CHAMPAIGN, UNIVERSITY OF ILLINOIS AT: Department of Atmospheric sciences. septembre 2008, Polar Research Group <URL: <http://arctic.atmos.uiuc.edu/cryosphere/>>

Table des matières

| | |
|--|-----------|
| A Des mathématiques en sciences humaines et sociales | 1 |
| A1 La démarche scientifique en sciences humaines et sociales | 1 |
| A2 Les études statistiques | 1 |
| A3 Acquérir un esprit critique | 2 |
| B Les variables | 2 |
| B1 Vocabulaire | 2 |
| B2 Variables qualitatives | 2 |
| B3 Variables quantitatives | 2 |
| C Tableaux et graphiques | 2 |
| C1 Variables qualitatives ou quantitatives discrètes (peu de valeurs distinctes) | 2 |
| C1 a Tableau des effectifs/ des fréquences | 2 |
| C1 b Représentations graphiques | 3 |
| C2 Variables quantitatives discrètes (grand nombre de valeurs) ou continues | 5 |
| C2 a Groupement par classes | 5 |
| C2 b Histogramme avec classes de même amplitude | 6 |
| C2 c Histogramme avec classes d'amplitudes différentes | 6 |
| D Tableaux à double entrée | 8 |
| E Séries chronologiques | 9 |
| E1 Mesures sur les séries chronologiques | 9 |
| E1 a Variation absolue | 9 |
| E1 b Cas particulier des variables exprimées en pourcentage | 9 |
| E1 c Variation relative | 10 |
| E1 d Variation moyenne | 10 |
| E1 e Comment commenter une série chronologique ? | 11 |
| F Mesures de tendance centrale | 11 |
| F1 Le mode | 11 |
| F1 a Définition | 11 |
| F1 b Données groupées par valeurs ou modalités | 11 |
| F1 c Données groupées par classes | 11 |
| F1 d Distribution polymodale | 11 |
| F2 La médiane | 12 |
| F2 a Conventions | 12 |
| F2 b Définition | 12 |
| F2 c Cas des données groupées par valeurs | 12 |
| F3 Moyenne | 13 |
| F3 a Données non groupées | 13 |
| F3 b Données groupées par valeurs | 14 |
| F3 c Données groupées par classes | 14 |
| G Quartiles et diagrammes en boîte | 14 |
| G1 L'idée | 14 |
| G2 Expérimentons | 14 |
| G3 Définissons | 15 |
| G4 Boîte à moustache | 15 |
| H Mesures de dispersion | 15 |
| H1 Écart interquartile | 16 |
| H2 Variance et écart-type | 17 |
| I Utilisation de la calculatrice | 18 |
| I1 Casio | 19 |
| I2 TI | 19 |
| J Des exercices de BAC | 22 |