

Statistique descriptive

Gérard Barmarin

Exercices Cours 1

Ex 1.1

Imaginez une étude statistique si possible liée au domaine de votre formation puis:

définissez pour cette statistique

- la population,
- la variable mesurée,
- l'échantillon,
- l'individu.

Ex 1.2

- A. Dans l'exercice précédent, quel est le type de la variable mesurée que vous avez choisie ?
- B. Trouvez une variable nominale non citée dans le cours
- C. Trouvez une variable ordinale non citée dans le cours
- D. Trouvez une variable numérique relative non citée dans le cours
- E. Trouvez une variable numérique discrète non citée dans le cours
- F. Trouvez une variable numérique continue non citée dans le cours

Pour chaque variable, quelles sont les valeurs observables?

Ex 1.3

Voici un tableau d'observations de la couleurs des yeux de 200 étudiants de l'Ifosup

couleur des yeux	bleu	vert	brun	pers	noir
fréquences absolues	66	34	80	15	5
fréquences relatives	33%	17%	40%	7.5%	2.5%

TABLE 2.1 – couleur des yeux d'un échantillon de 200 personnes

De quel type de variable s'agit-il ?

Quel(s) type(s) de diagramme allez-vous utiliser pour représenter cette statistique ?

Dessinez un diagramme possible.

Ex 1.4

Tableau d'observations du nombre d'enfants pour 1000 couples de Wavre :

Nombre d'enfants	0	1	2	3	4	5	6	> 6
fréquence absolue	235	183	285	139	88	67	3	0
fréquence relative	23.5%	18.3%	28.5%	13.9%	8.8%	6.7%	0.3%	0

- De quel type de variable s'agit-il ?
- Quel(s) type(s) de diagramme allez-vous utiliser pour représenter cette statistique ?
- Dessinez un diagramme possible.

Exercices cours 2 :

Vocabulaire

&

représentation graphique

Exercice 2.1

Ex 2.1

Le directeur de fabrication des industries Scheer s'intéresse à un programme qui permettrait de former les employés de maintenance de l'entreprise à la réparation des machines. L'intérêt de cette nouvelle méthode est qu'elle devrait réduire le temps nécessaire pour former les employés. Ce temps est actuellement de 60 jours.

Pour évaluer l'efficacité de la méthode, le GRH l'a utilisée avec 11 employés et a noté le nombre de jours gagnés ou perdus par rapport à l'ancienne méthode. Voici les résultats :

$X(158423) = 5$	$X(261158) = 6$	$X(791581) = 10$	$X(642072) = - 2$
$X(567336) = 3$	$X(243989) = - 5$	$X(412817) = 6$	$X(830975) = - 1$
$X(927955) = - 3$	$X(692147) = - 3$	$X(335834) = 6$	

où X signifie « nombre de jours gagnés » et i est le numéro matricule de l'employé concerné. Donc $X(i)$ signifie « nombre de jours gagnés pour former l'employé qui a le matricule i ». Si ce nombre est positif, cela signifie que la formation avec la nouvelle méthode a duré moins que 60 jours ; s'il est négatif, la formation a duré plus de 60 jours ; s'il est nul, la formation a duré 60 jours, comme avec l'ancienne méthode.

Ex 2.1

$$X(158423) = 5 \quad X(261158) = 6 \quad X(791581) = 10 \quad X(642072) = -2$$

$$X(567336) = 3 \quad X(243989) = -5 \quad X(412817) = 6 \quad X(830975) = -1$$

$$X(927955) = -3 \quad X(692147) = -3 \quad X(335834) = 6$$

- Quels sont les individus observés ?
- Combien y-a-t-il d'observations ?
- De quel type de variable s'agit-il ?
- Quelles sont les valeurs observables de cette variable peut-elle prendre ?
Positive ? Négative ? y-a-t-il un minimum pour cette valeur ? un maximum ?

Ex 2.1

$$X(158423) = 5 \quad X(261158) = 6 \quad X(791581) = 10 \quad X(642072) = -2$$

$$X(567336) = 3 \quad X(243989) = -5 \quad X(412817) = 6 \quad X(830975) = -1$$

$$X(927955) = -3 \quad X(692147) = -3 \quad X(335834) = 6$$

- Faites la liste de toutes les observations rangées ordinalement
- Quelle est la valeur observée le plus souvent (mode) ?
- Quelle est la plus petite valeur observée ?
- Quelle est la plus grande valeur observée ?
- Quelle est la différence entre la plus petite et la plus grande valeur observée (étendue)?
- Recopiez cette distribution dans un tableau adéquat

Ex 2.1

- Faites un graphique représentant les données correctement.
De quel type de graphique s'agira-t-il ?
- Le directeur GRH s'est trompé dans les dates et a inclus un week-end, il faut donc **retirer 2** à chacune des observations ci-dessus :
 - La forme générale du graphique change-t-elle ?
 - La valeur du mode change-t-elle ?
 - La valeur de l'étendue change-t-elle ?
- Peu doué, le directeur GRH s'aperçoit qu'il a fait une autre erreur : il ne s'agit pas de jours mais d'heures gagnées sur le temps de formation.
Même questions que ci-dessus

Les données suivantes ont été collectées en comptabilisant le nombre de salles d'opération en usage dans un hôpital déterminé sur une période de 60 jours :

- Au cours de 6 jours, aucune salle d'opération (sur les sept que compte l'hôpital) n'a été utilisée
- Au cours de 9 jours, une seule salle d'opération a été utilisée
- Au cours de 15 jours, exactement deux salles d'opération ont été utilisées
- Au cours de 12 jours, exactement trois salles d'opération ont été utilisées
- Au cours de 8 jours, exactement quatre salles d'opération ont été utilisées –
- Au cours de 7 jours, exactement cinq salles d'opération ont été utilisées
- Il n'y a pas eu de jour où exactement six salles d'opération ont été utilisées
- Au cours de 3 jours, toutes les sept salles d'opération de l'hôpital ont été utilisées

Soit **X** la variable « nombre de salles d'opération utilisées » :

- De quel type de variable s'agit-il ?
- Que sont les individus observés ?
- Quelles sont les valeurs observables pour cette variable? Positive ? Négative ? y-a-t-il un minimum pour cette valeur ? un maximum ?
- Faites la liste de toutes les observations rangées ordinalement (si cette liste est trop longue, utilisez des pointillés entre les valeurs qui se répètent, et indiquez combien de fois elles se répètent)
- Recopiez cette distribution dans un tableau adéquat avec les effectifs cumulés à gauche
- Quelle est la valeur observée le plus souvent (mode) ?
- Quelle est la plus petite valeur observée ?
- Quelle est la plus grande valeur observée ?
- Quelle est la différence entre la plus petite et la plus grande valeur observée (étendue)?

Ex 2.2

- Faites un graphique pour représenter les données correctement.
De quel type de graphique s'agira-t-il ?
- L'observateur s'est trompé, il faut rajouter 2 jours à toutes les observations :
 - La forme générale du graphique change-t-elle ?
 - La valeur du mode change-t-elle ?
 - La valeur de l'étendue change-t-elle ?
- Idem, mais cette fois, l'opérateur n'a pas vu que les données étaient exprimées en demi-jours, il faut multiplier par deux les observations
- Représentez graphiquement le cumul des effectifs (à gauche).

Exercice 2.3

Une étude réalisée en 2003 sur les abonnés à *The Wall Street Journal* a révélé certaines caractéristiques de ces abonnés, comme leurs responsabilités professionnelles, leur style de vie, leur fortune personnelle et leurs activités d'investissement. Le tableau suivant résume la distribution de 360 abonnés sur la valeur totale de leur portefeuille d'actions (en milliers de dollars) :

Effectif de la classe [0, 50[:	0
Effectif de la classe [50, 100[:	30
Effectif de la classe [100, 140[:	40
Effectif de la classe [140, 200[:	90
Effectif de la classe [200, 300[:	100
Effectif de la classe [300,500[:	60
Effectif de la classe [500,1000] :	40

Ex 2.3

- Quelle est la variable ? De quel type de variable s'agit-il ?
- Quelles valeurs cette variable peut-elle prendre ? Positive ? Négative ? y-a-t-il un minimum pour cette valeur ? un maximum ?
- Combien de classe y-a-t-il ? Sont-elles de largeur constante ?
- Recopiez cette distribution dans un tableau adéquat en indiquant le centre de classe, les effectifs, la densité d'effectif
- Faites un graphique pour représenter les données correctement. De quel type de graphique s'agira-t-il ?
- Quelle est la classe dont la densité d'effectif est la plus grande ? Quel est le milieu de cette classe ?
- L'observateur s'est trompé, il faut rajouter 2 jours à toutes les observations, La forme générale du graphique change-t-elle ?
- Idem, mais cette fois, l'opérateur n'a pas vu que les données étaient exprimées en demi-jours, il faut multiplier par deux les observations
- Quel est le graphique approprié pour représenter le cumul des effectifs pour cette distribution ? Faites ce graphique correctement.

Exercice 2.4

Voici un tableau représentant la distribution sur base de leur superficie (en ares) de 240 terrains agricoles vendus au premier trimestre 2013 à Cracovie:

[35, 80[:	20
[80, 105[:	40
[105, 110[:	50
[110, 120[:	30
[120, 140[:	60
[140, 170[:	40

Ex 2.4

- Quelle est la variable observée ? De quel type de variable s'agit-il ?
- Quelles valeurs cette variable peut-elle prendre ? Positive ? Négative ? y-a-t-il un minimum pour cette valeur ? un maximum ?
- Combien de classe y-a-t-il ? Sont-elles de largeur constante ?
- Que sont les individus ?
- Recopiez cette distribution dans un tableau adéquat en indiquant le centre de classe, les effectifs, la densité d'effectif
- Faites l'**HISTOGRAMME** correspondant
- Quelle est la classe dont la densité d'effectif est la plus grande ? Quel est le milieu de cette classe ?
- L'opérateur n'a pas vu que les données étaient exprimées Hectares et non en ares, il faut donc multiplier les observations par 100. La forme générale du graphique change-t-elle ?
- Quel est le graphique approprié pour représenter le cumul des effectifs pour cette distribution ? Faites ce graphique correctement.

Exercice 2.6

Ex 2.6

Longueur totale du crâne (mm) pour un sous-échantillon de 60 souris sylvestres adultes (I, II et III), tiré d'un échantillon de 122 souris de Landry (2000).



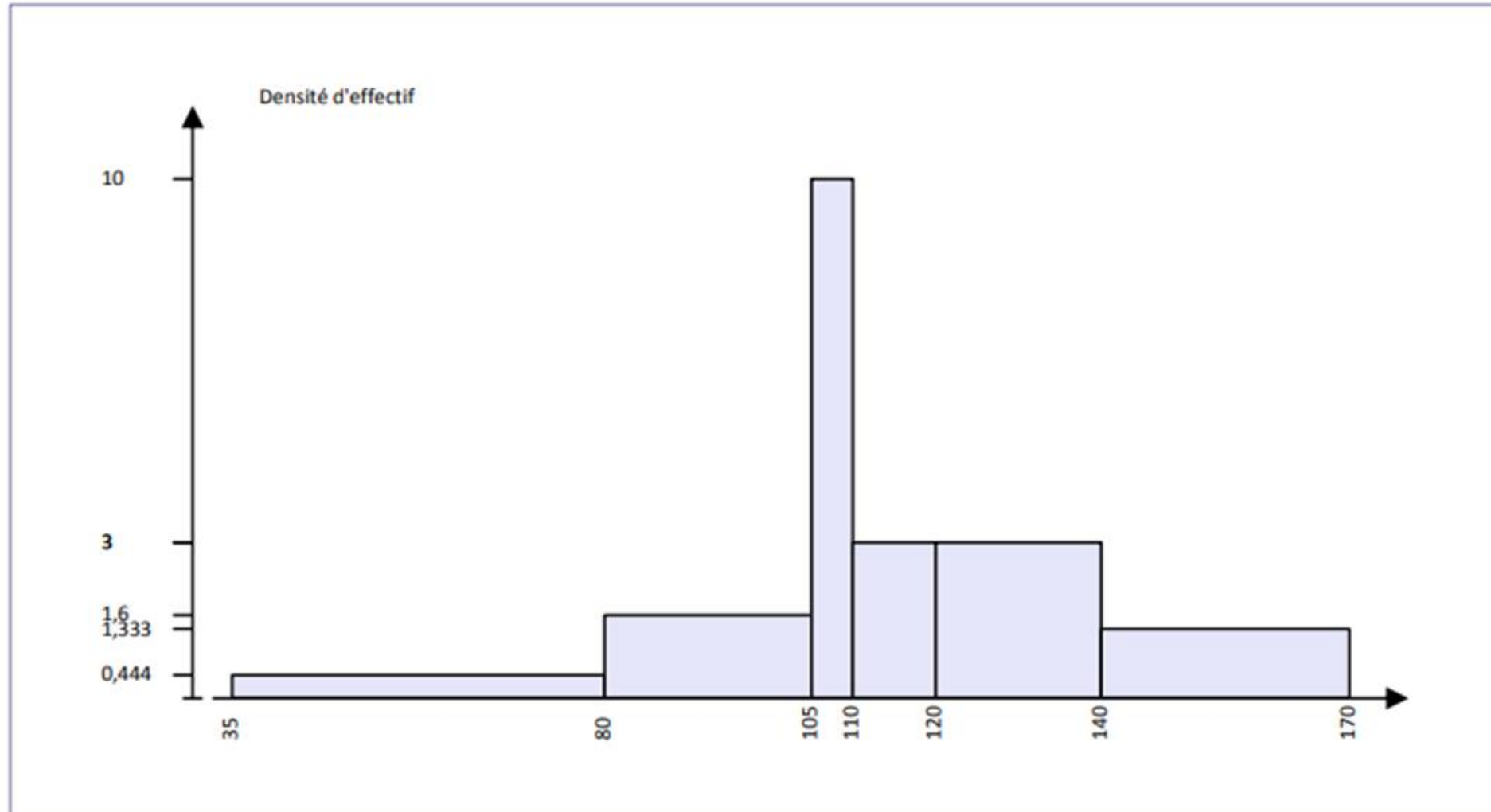
22,28	23,18	23,47	23,72	24,09	24,56
22,56	23,23	23,48	23,48	24,13	24,63
22,57	23,29	23,48	23,48	24,32	24,83
22,60	23,30	23,49	23,49	24,35	24,94
22,69	23,34	23,51	23,51	24,36	24,95
22,73	23,35	23,56	23,56	24,37	25,00
22,78	23,35	23,57	23,57	24,41	25,07
22,91	23,37	23,60	23,60	24,43	25,16
23,05	23,39	23,61	23,61	24,43	25,48
23,14	23,47	23,71	23,71	24,52	25,74

Ex 2.6

- Quelle est la variable ? Que sont les individus ?
- De quel type est-elle ?
- Combien y-a-t-il d'observations ?
- Combien de classe allez-vous faire (selon la règle de Sturge et/ou de Yule) ?
- Quelle sera l'étendue des classes et leurs bornes ?
- Faites un tableau des effectifs (fréquence absolue), des fréquences (fréquence relative), des pourcentages, des densité d'effectifs, des densités de fréquence et des effectifs cumulés.
- Représentez l'histogramme de la distribution de la densité de fréquence relative de la variable
- Faites un polygone cumulé des effectifs à gauche de la variable

Ex 3.6

Voici l'histogramme représentant la distribution sur leur superficie (en ares) de 240 terrains agricoles vendus au premier trimestre 2013 à Cracovie.



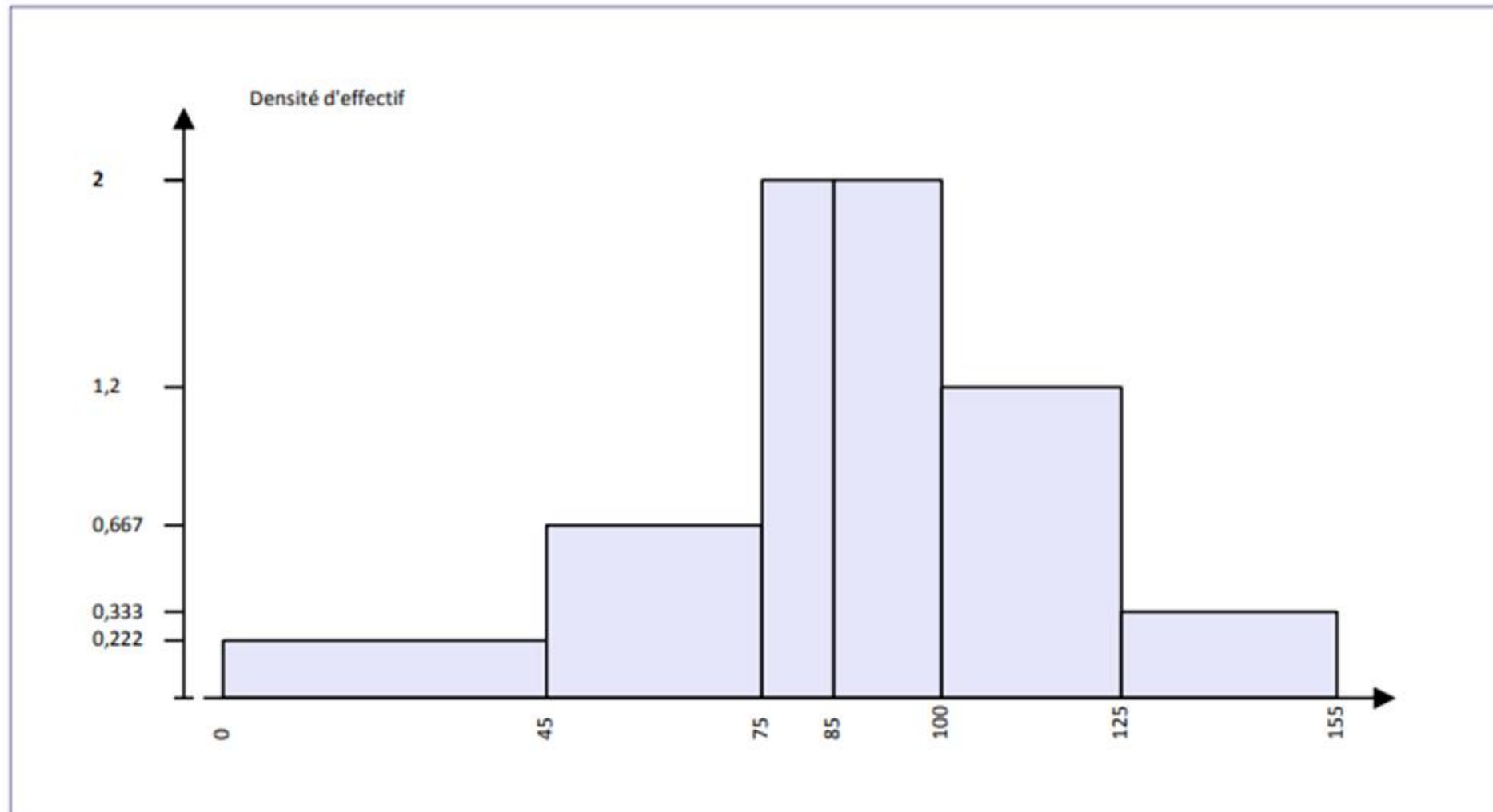
On demande :

- Que sont les individus distribués ?
- Combien y-a-t-il d'individus observés ?
- Combien y-a-t-il d'observations ?
- Pouvez-vous estimer le nombre de valeurs observées différentes ?
- Quelle est la variable étudiée ?
- Pouvez-vous donner la liste de toutes les observations (en utilisant des pointillés si nécessaire) ?
- Quelle est la superficie moyenne de tous ces terrains ?
- Quelle est la variance de cette distribution ?
- Quel est l'écart type ?
- Quel est l'écart absolu moyen ?
- Quelle est l'étendue de cet ensemble d'observations ?
- Quel est la classe modale et le mode de cette distribution ?
- Faites le graphe des effectifs cumulés de cette distribution

Ex 3.6

- Y-a-t-il une valeur telle que $\frac{1}{4}$ des observations sont en-dessous de cette valeur ?
Si oui comment s'appelle-t-elle et pouvez-vous la calculer ?
- Calculer la médiane
- A quoi correspond la médiane ?
- Quelle est la valeur inter-quartile ?
- Quelle est la valeur quantile d'ordre 0,675 ?
- Quelle est la proportion de terrain plus petit que 83,5 ares ?
- Quelle est la proportion de terrains dont la superficie est comprise entre 83,5 et 137 ares.

Voici l'histogramme représentant la distribution sur leur superficie (en ares) de 120 terrains agricoles vendus au premier trimestre 2014 à Cracovie.



On demande :

- Quelle est la superficie moyenne de tous ces terrains ?
- Quelle est la variance de cette distribution ?
- Quel est l'écart type ?
- Quel est l'écart absolu moyen ?
- Quelle est l'étendue de cet ensemble d'observations ?
- Quel est la classe modale et le mode de cette distribution ?
- Faites le graphe des effectifs cumulés de cette distribution
- Y-a-t-il une valeur telle que $\frac{1}{4}$ des observations sont en-dessous de cette valeur ? Si oui comment s'appelle-t-elle et pouvez-vous la calculer ?
- Calculer la médiane
- Quelle est la valeur inter-quartile ?
- Quelle est la valeur quantile d'ordre 0,333 ?
- Quelle est la proportion de terrain plus petit que 35 ares ?
- Quelle est la proportion de terrains dont la superficie est comprise entre 35 et 127,5 ares
- Quelle est la proportion des observations à plus d'un écart type de la moyenne ?

Ex 3.6

- Y-a-t-il une valeur telle que $\frac{1}{4}$ des observations sont en-dessous de cette valeur ?
Si oui comment s'appelle-t-elle et pouvez-vous la calculer ?
- Calculer la médiane
- A quoi correspond la médiane ?
- Quelle est la valeur inter-quartile ?
- Quelle est la valeur quantile d'ordre 0,675 ?
- Quelle est la proportion de terrain plus petit que 83,5 ares ?
- Quelle est la proportion de terrains dont la superficie est comprise entre 83,5 et 137 ares.