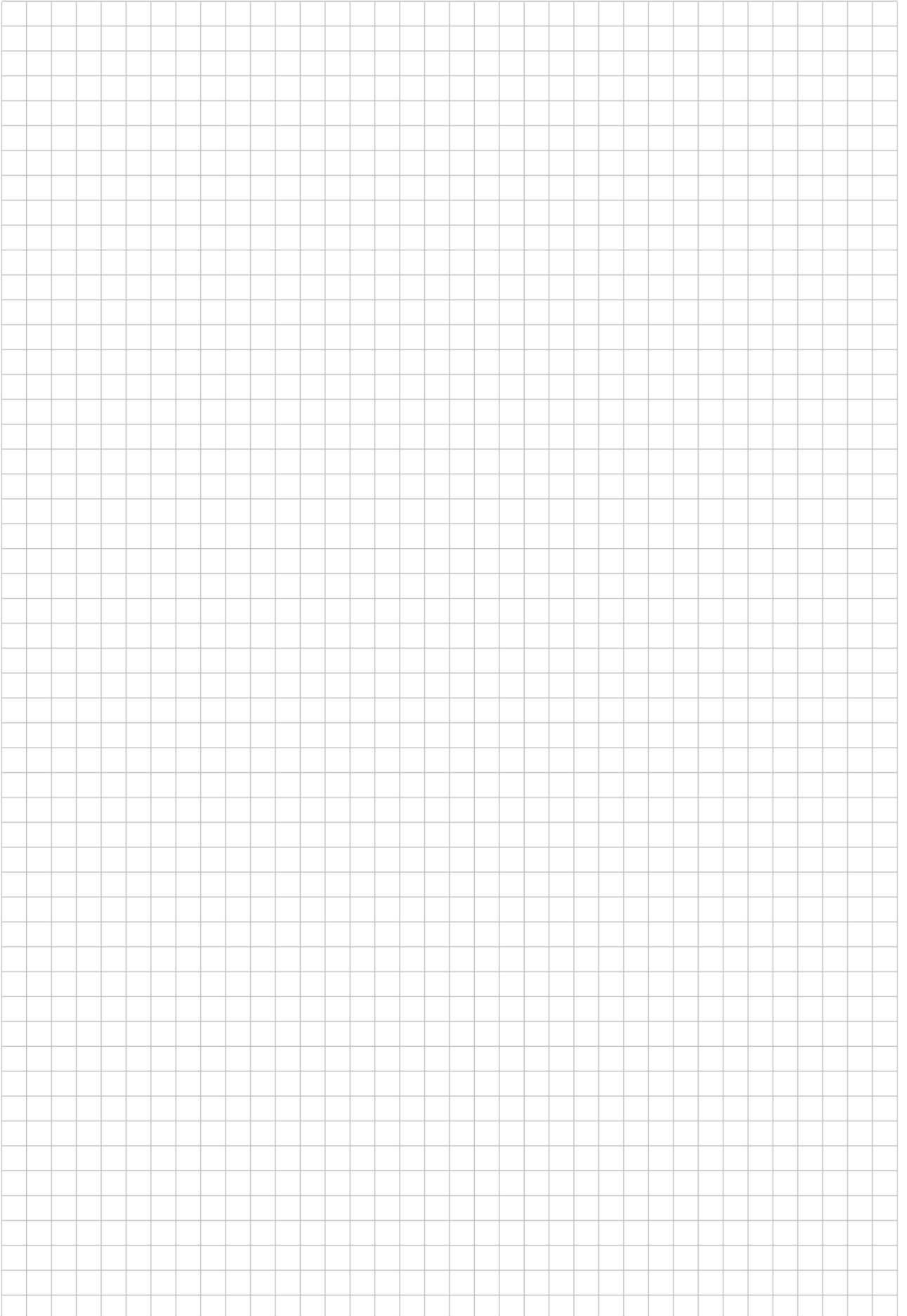


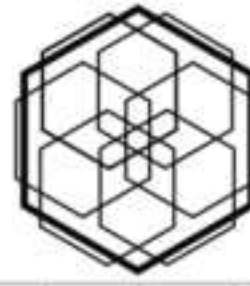
# PrivateTeacher

*Cours Privés de Science*

Statistiques T1

Corrigé Détaillé





HEG 1<sup>er</sup> semestre

Statistiques I

## Test en blanc

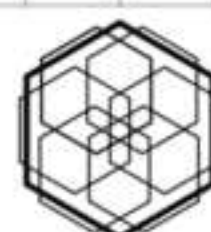
RÉPONSES au QCM

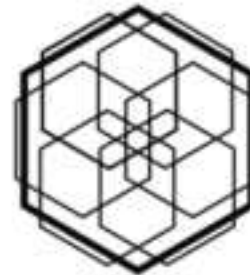
Nom : *Ruppen*

Prénom : *Julien*

*Réponses :*

Question	Choix				
1	A	B	<input checked="" type="radio"/> C	D	E
2	A	B	C	D	<input checked="" type="radio"/> E
3	A	B	<input checked="" type="radio"/> C	D	E
4	A	B	C	D	<input checked="" type="radio"/> E
5	<input checked="" type="radio"/> A	B	C	D	E
6	A	B	C	D	<input checked="" type="radio"/> E
7	<input checked="" type="radio"/> A	B	C	D	E
8	A	B	C	D	E
9	A	B	C	D	E
10	A	B	<input checked="" type="radio"/> C	D	E
11	A	B	C	<input checked="" type="radio"/> D	E
12	A	B	<input checked="" type="radio"/> C	D	E
13	A	B	<input checked="" type="radio"/> C	D	E
14	<input checked="" type="radio"/> A	B	C	D	E
15	A	B	C	D	<input checked="" type="radio"/> E
16	A	B	<input checked="" type="radio"/> C	D	E
17	<input checked="" type="radio"/> A	B	C	D	E
18	A	B	C	D	E
19	A	B	C	D	E
20	A	B	<input checked="" type="radio"/> C	D	E



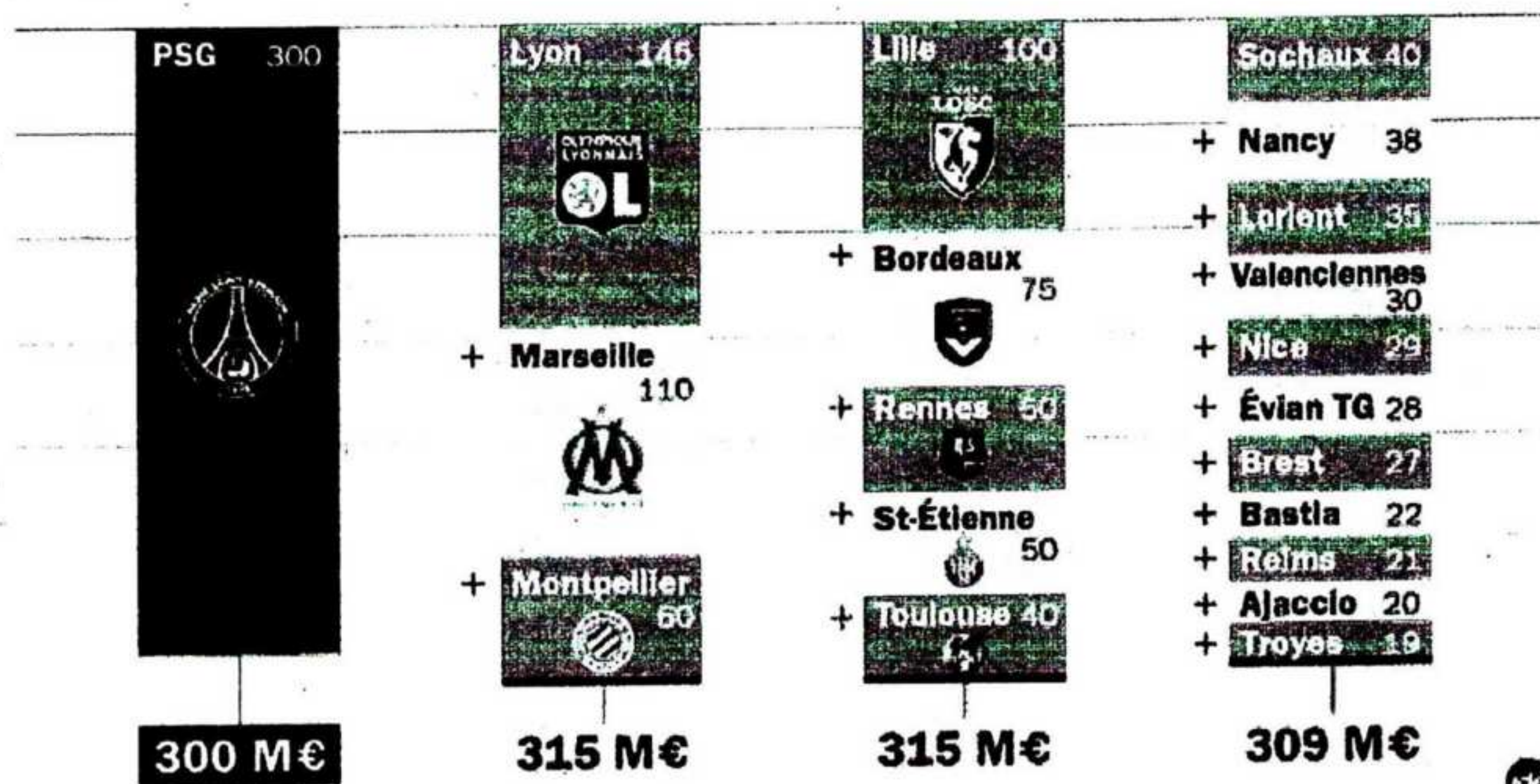


Q1

Le graphique ci-dessous montre la répartition des 20 clubs de Ligue 1 en France en 2012-2013:

### Le PSG, un quart du budget total des clubs de L1

Budget des clubs de L1 pour la saison 2012-2013 (en millions d'euros)



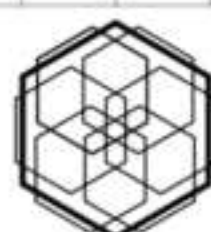
Par quel point passe la courbe de Lorenz ?

- A. -0.30 et +0.15
- B. +0.80 et +0.60
- C. +0.05 et +0.24
- D. +0.95 et +0.24
- E. +0.95 et +0.76

Q2

En décembre 2008 on réalise une enquête auprès de 16 agences immobilières de la ville de Kansas City afin de savoir combien de transactions ont été réalisées au cours du mois. Les réponses ont été les suivantes : 3 0 0 1 1 2 2 0 0 2 1 0 2 1 4 2. Quel est le mode ?

- A. 0
- B. 1
- C. 1.3
- D. 2
- E. La distribution est bi-modale





### V-CRAMER

Q3 Une banque veut savoir si le nombre de transactions par mois d'un client dépend de son état civil. Elle a obtenu les données suivantes (EC = Etat civil et NbT = Nombre de transactions) :

EC	NbT	EC	NbT	EC	NbT	EC	NbT
Célibataire	9	Célibataire	10	Marié	23	Marié	11
Célibataire	8	Célibataire	2	Marié	2	Marié	17
Célibataire	14	Célibataire	32	Marié	7	Marié	1
Célibataire	18	Célibataire	10	Marié	23	Marié	6
Célibataire	2	Célibataire	0	Marié	0	Marié	10
Marié	12	Célibataire	11				

La banque décide de considérer deux catégories de transactions : de 0 à 10 transactions inclus, et plus de 10 transactions.

Le V de Cramer vaut (précision à 2 chiffres après la virgule)

- A. 0.19
- B. 0.94
- C. 0.09
- D. 1.21
- E. aucune des réponses ci-dessus

Q4 Les deux variables sont donc

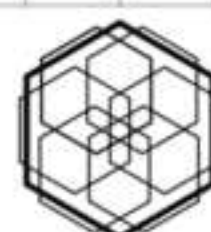
- A. fortement liées
- B. fortement linéairement liées
- C. faiblement linéairement liés
- D. liées
- E. quasi-indépendantes

### TYPE DE VARIABLES

Q5 Une analyse est effectuée sur les voitures dans un parking. Indiquez pour les variables suivantes leur type.

La couleur des voitures, notée 1 pour rouge, 2 pour blanc, 3 pour noir, 4 pour bleu et 5 pour autre couleur.

- A. Qualitative nominale
- B. Qualitative ordinale
- C. Quantitative discrete
- D. Quantitative continue





Q6 On considère l'échantillon  $x$  constitué des six observations suivantes :

1 1 0 -1 0 -1

Quelle valeur obtient-on si on exécute la commande  $sd(x)$  dans R ?

- A. -0.894
- B. 0.800
- C. 0
- D. 0.800
- E. 0.894

Q7

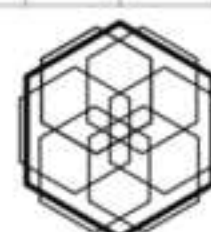
Un fabricant de boisson énergétique a étudié si les consommateurs préféreraient une version naturelle ou aromatisée de sa nouvelle boisson.

	Aromatisée	Naturelle
Hommes	101	97
Femmes	68	134

H : Hommes F : Femmes A : Aromatisée N : Naturelle

Si ces deux facteurs étaient indépendants, quel serait le tableau de contingence ? (HA / HN / FA / FN)

- A. 83.7 / 114.3 / 85.3 / 116.7
- B. 84.5 / 115.5 / 84.5 / 115.5
- C. 99 / 99 / 101 / 101
- D. 100 / 100 / 100 / 100
- E. 101 / 97 / 68 / 134

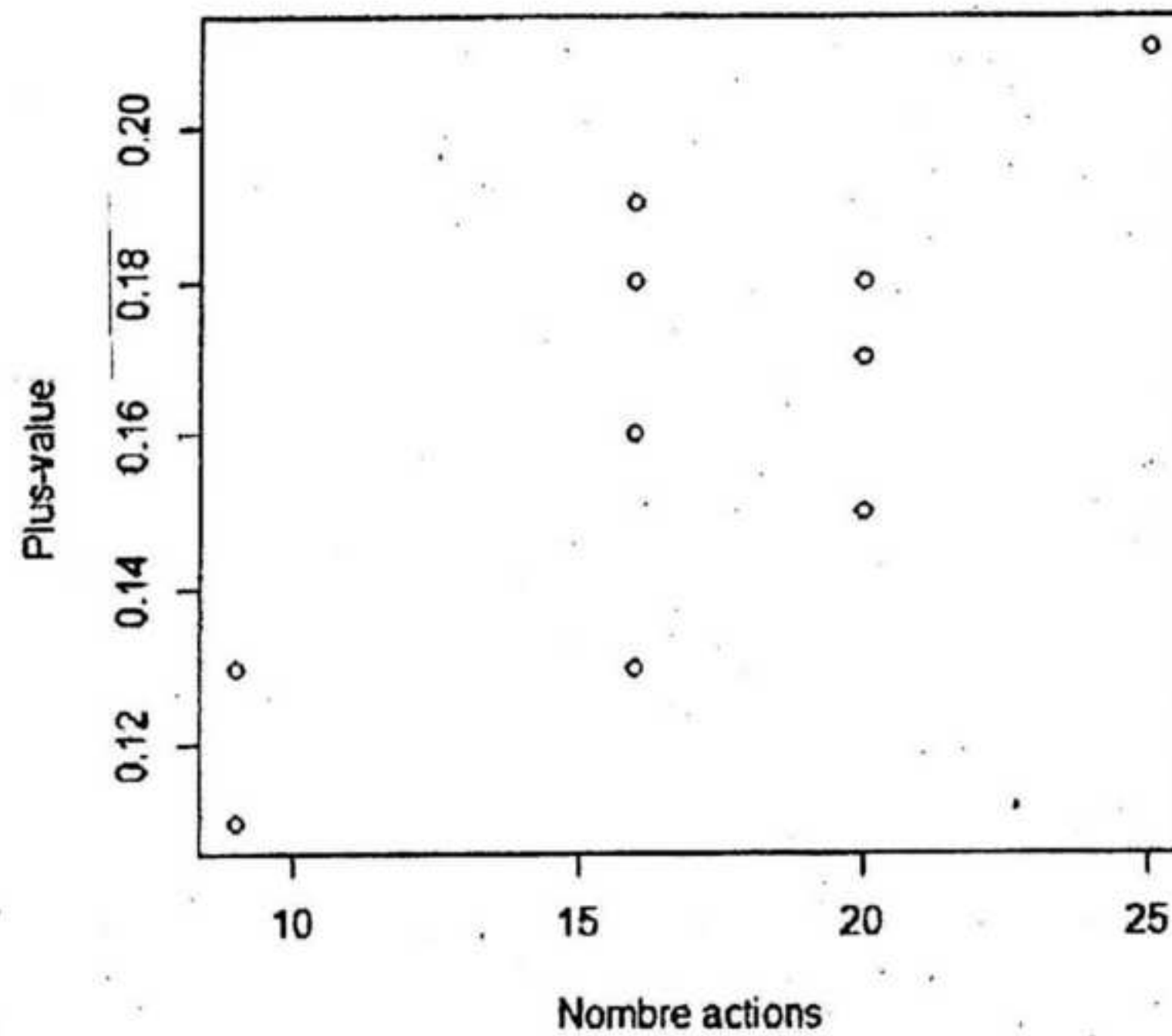




Q8

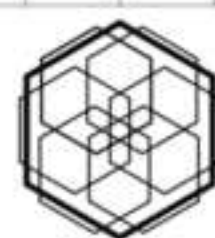
Vous étudiez chez 10 clients de votre banque pris au hasard la corrélation entre le nombre d'actions détenues et le taux de plus-values annuelles.

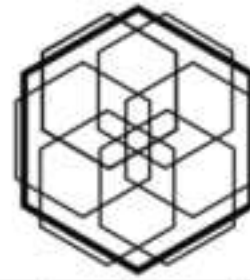
La somme des produits (nombre d'actions) x (plus-values) vaut 27.97.



Vous vous rendez compte que vous avez publié une onzième observation à votre échantillon : (10; 0.16). Sachant que la moyenne du nombre d'actions vaut alors 16.1 et celle des plus-values 0.16, que vaut la covariance entre le nombre d'actions et les plus-values ?

- A. -0.436
- B. -0.123
- C. 0.000
- D. +0.123
- E. +0.436



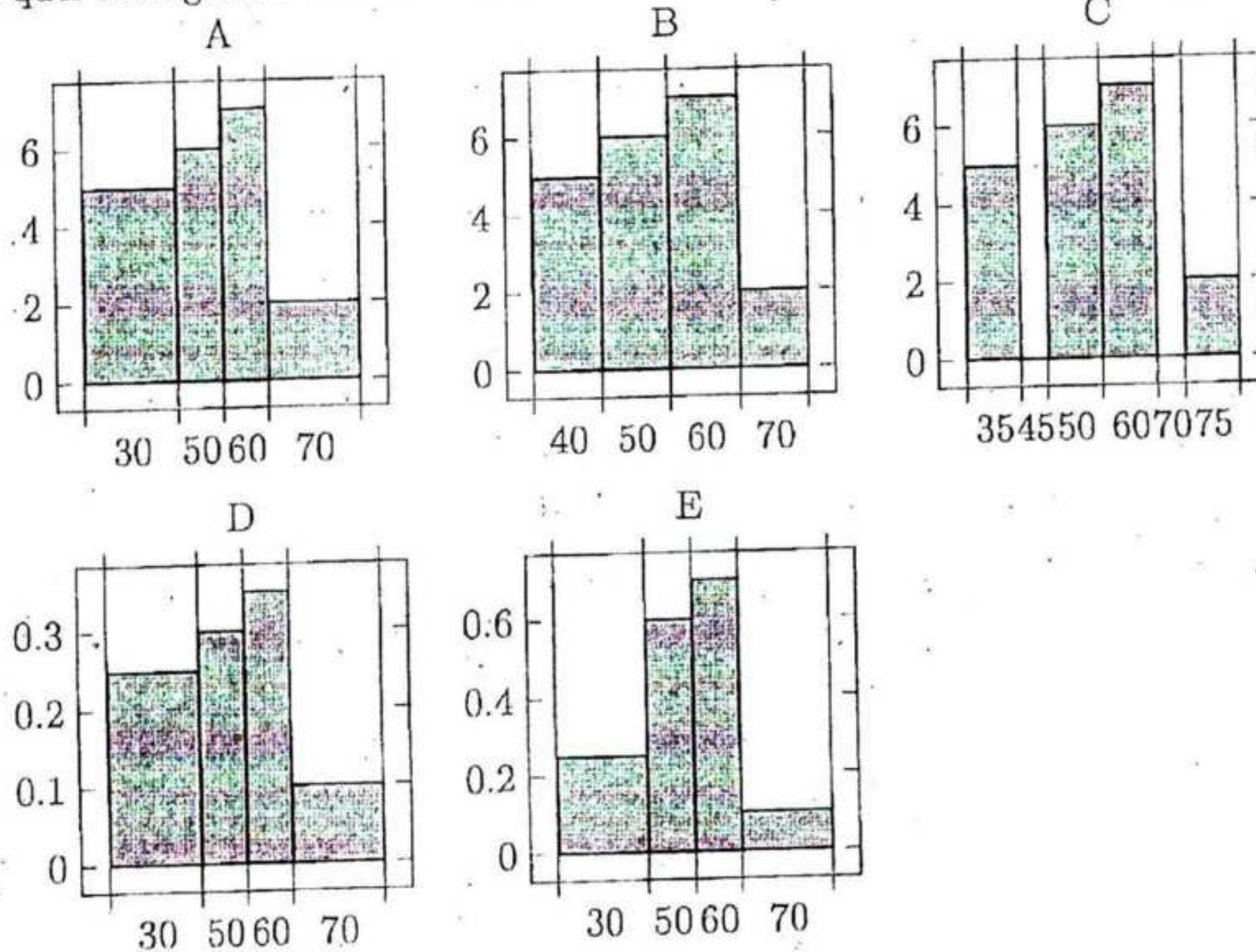


## HISTOGRAMME

Q9

On mesure le poids de 20 adultes. Il y a 5 mesures entre 30 et 50 kg, 6 mesures entre 50 et 60 kg, 7 mesures entre 60 et 70 kg, et 2 mesures comprises entre 70 et 90 kg.

A quel histogramme ces mesures correspondent-elles ?



- A.
- B.
- C.
- D.
- E.

Q10

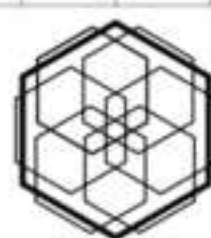
Le service de police effectue un relevé des excès de vitesse mesuré en  $km/h$  par le parc de radars automatiques. Il relève 550 infractions dans la classe  $[0; 5[$ , 256 dans la classe  $[5; 10[$ , 203 dans la classe  $[10; 15[$ , 144 dans la classe  $[15; 20[$ , et 97 dans la classe  $[20; 50[$ .  
Que valent la moyenne des données groupées et le premier quartile ?

- A. 5.9 et 2.8
- B. 5.9 et 5.0
- C. 9.4 et 2.8
- D. 9.4 et 5.0
- E. 12.9 et 5.0

Q11

La moyenne est une mesure de la tendance centrale. Quels sont les avantages associés à cette mesure ?

- A. Elle permet de synthétiser un ensemble de mesures
- B. Elle permet de comparer des groupes entre eux.
- C. Elle permet de comparer un individu à l'ensemble du groupe
- D. Toutes les réponses ci-dessus sont correctes
- E. Les affirmations ci-dessus sont correctes, mais pour la médiane uniquement.







## RÉGRESSION

Q12 Douze personnes sont inscrites à une formation. Au début de la formation, ces stagiaires subissent une épreuve A notée sur 20. A la fin de la formation, elles subissent une épreuve B de niveau identique. Les résultats sont donnés dans le tableau suivant :

Épreuve A	3	4	6	7	9	10	9	11	12	13	15	4
Épreuve B	8	9	10	13	15	14	13	16	13	19	6	19

Calculer la corrélation entre ces 2 variables

- A. 0.214
- B. 0.897
- C. 0.101
- D. 0.456
- E. 0

Q13 Calculer la valeur des paramètres de la régression

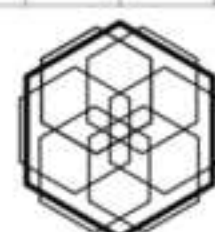
- A. Pente 0.487, Ordonnée à l'origine 3.145
- B. Pente -0.108, Ordonnée à l'origine 14.17
- C. Pente 0.108, Ordonnée à l'origine 11.99
- D. Pente 0.987, Ordonnée à l'origine 23.847
- E. Pente 1.000, Ordonnée à l'origine 11.99

Q14 On utilise une 13<sup>ème</sup> personne qui obtient 14 pour l'épreuve A. Que devrait-elle obtenir pour l'épreuve B d'après votre modèle ?

- A. 13.50
- B. 08.75
- C. 19.87
- D. 9.87
- E. 11.47

Q15 Le coefficient de détermination vaut 0.0103. Cela semble indiquer que le modèle linéaire est

- A. très bon
- B. bon
- C. est indépendant de la variance expliquée
- D. est fortement dépendant de la variance totale
- E. mauvais





Q16

Vous êtes en charge de la gestion budgétaire d'un projet. Les soumissions d'entreprises respectant le cahier des charges, suite à un appel d'offre, sont résumées dans le tableau récapitulatif suivant :

Coût [Frs]	Nombre d'entreprises
[10000; 12000[	8
[12000; 15000[	10
[15000; 18000[	16
[18000; 20000]	6

En considérant les données groupées ci-dessus, que vaut la moyenne des coûts ?

- A. 16600
- B. 16500
- C. 15025
- D. aucune des réponses ci-dessus

Q17

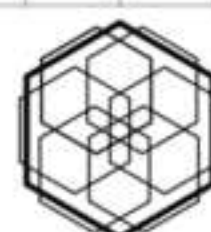
Vous décidez de rejeter le 25% des offres les plus basses (et donc de considérer le 75% des offres restantes). A partir de quel montant les offres sont-elles considérées ?

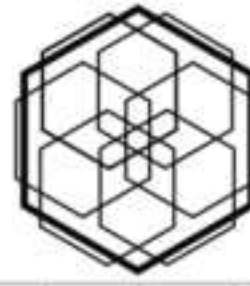
- A. 12600
- B. 12000
- C. 15000
- D. 15025
- E. aucune des réponses ci-dessus

Q18

Calculez, à partir de la méthode vue au cours concernant les tableaux de données (données groupées), le troisième quartile.

- A. 15900
- B. 18200
- C. 15000
- D. 15100
- E. aucune des réponses ci-dessus





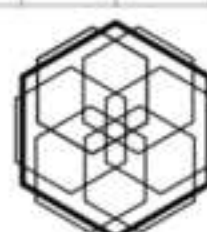
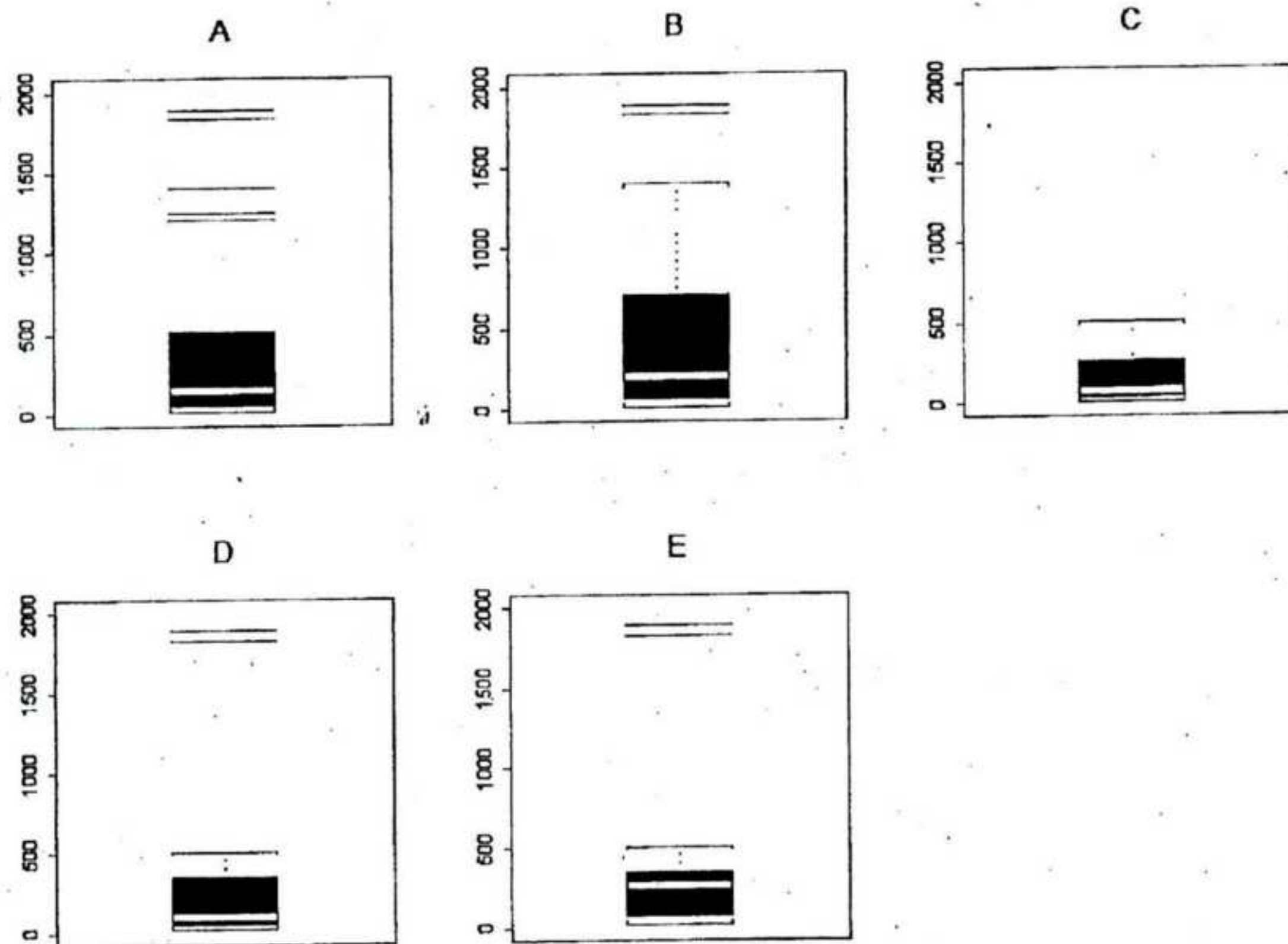
Q19

On mesure un jour donné les 20 cotations des valeurs boursières constituant le Swiss Market index (SMI)

Titre	Prix
ABB LTD	19.060
ACTELION	141.900
ADECCO	68.650
CREDIT SUISSE GROUP	21.750
GEBERIT	338.200
GIVAUDAN	1827.000
JULIUS BAER	47.730
LAFARGEHOLCIM	55.500
NESTLE	75.750
NOVARTIS	88.150
RICHEMONT	75.300
ROCHE GS	268.700
SGS	1892.000
SWATCH GROUP	357.000
SWISS RE	96.800
SWISSCOM	508.000
SYNGENTA	379.000
TRANSOCEAN	14.740
UBS GROUP	19.490
ZURICH INSURANCE	266.200

La médiane de ces valeurs vaut environ 92 et l'intervalle inter-quartile environ 287.

Quel est le boxplot correspondant à ces valeurs ? Remarque : Les potentielles valeurs singulières sont représentées ici sur les boxplots par des barres (au lieu de points ou cercles)

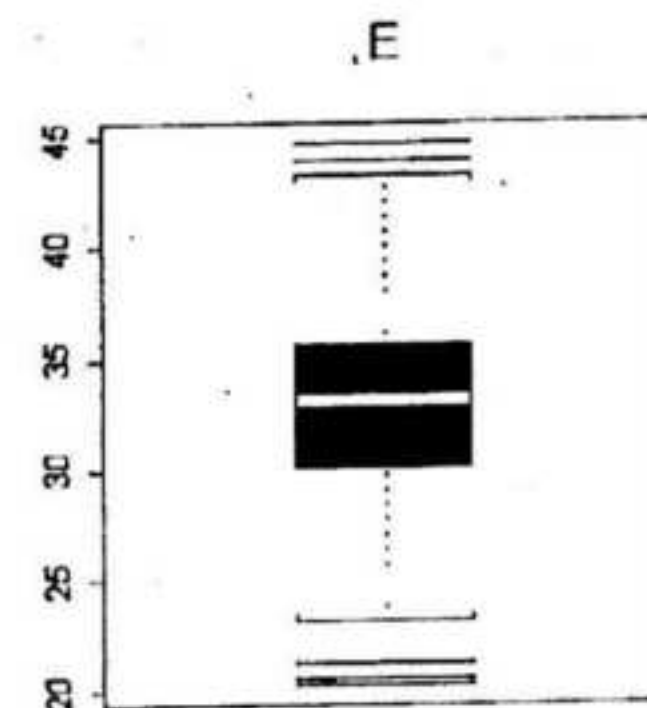
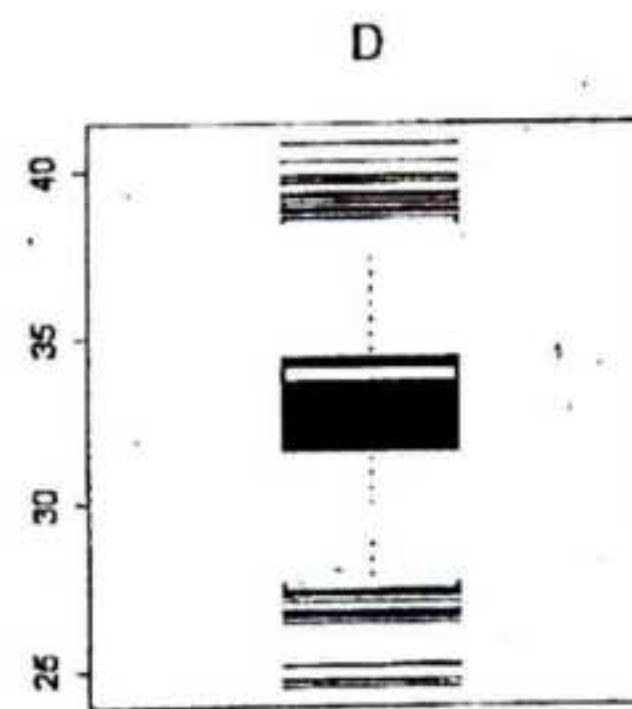
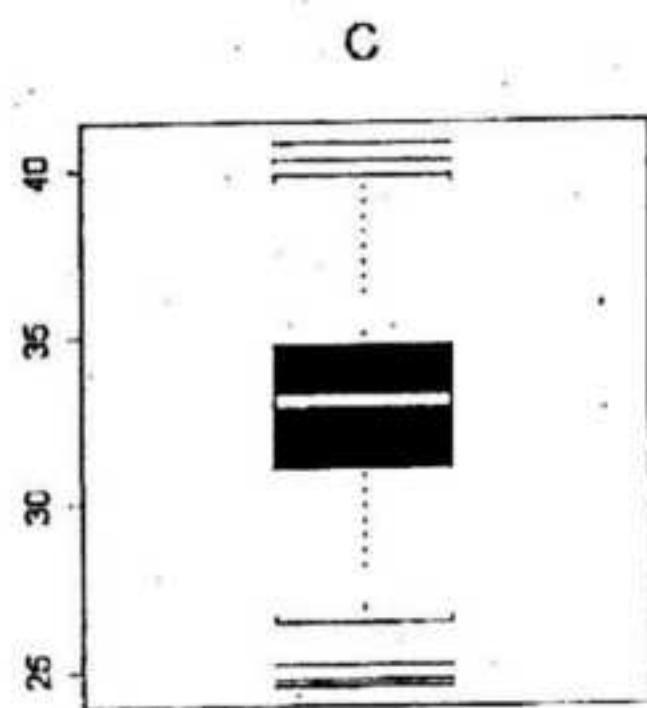
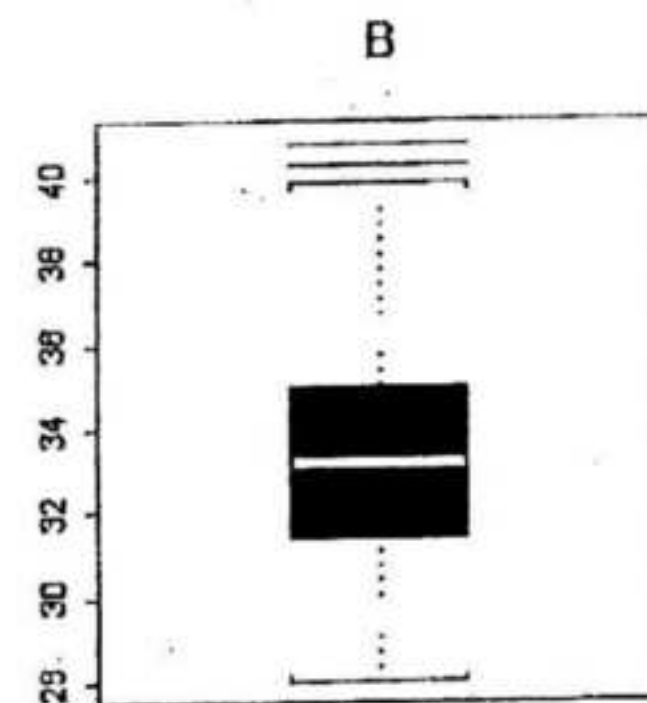
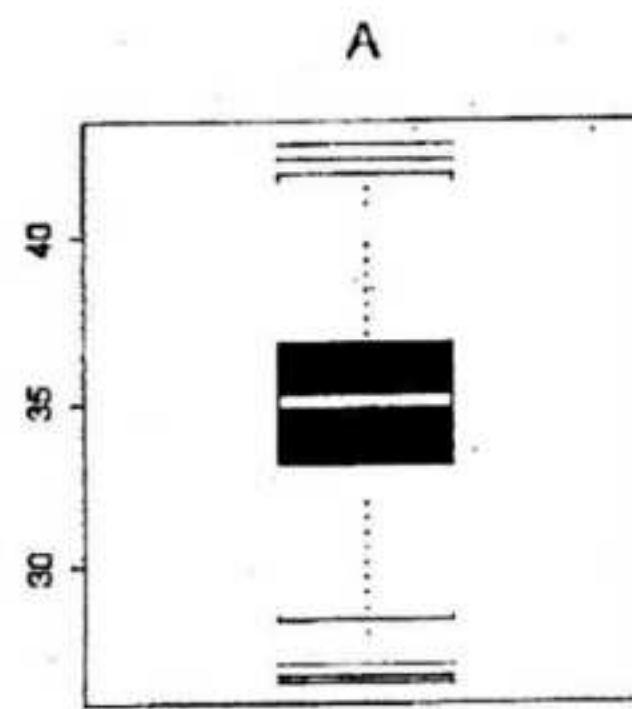
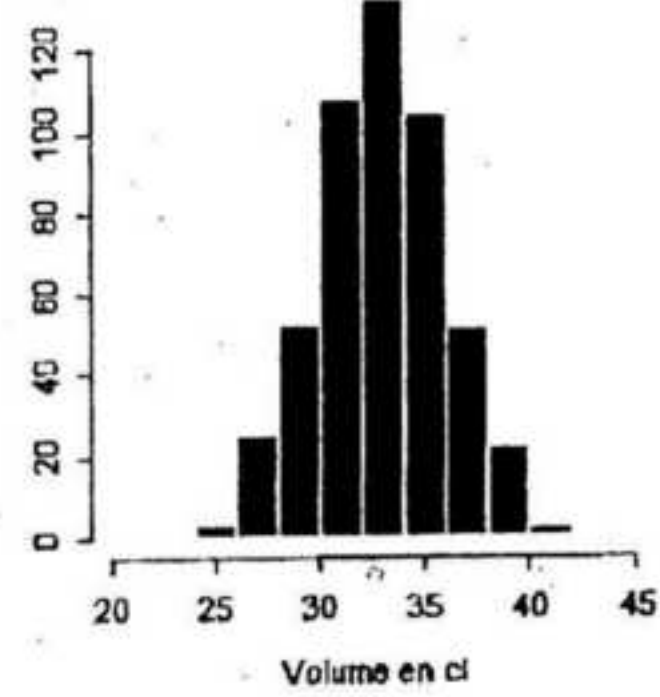




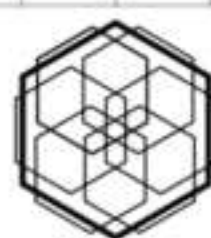
Q20

On mesure le remplissage de 500 cannettes de soda de 33cl choisies au hasard le long d'une chaîne de production. La fréquence des observations est représentée sur le graphique joint.

À quel boxplot cet histogramme correspond il ?



- A.
- B.
- C.
- D.
- E.

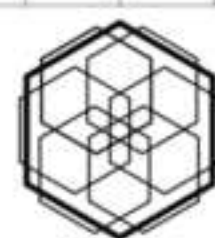
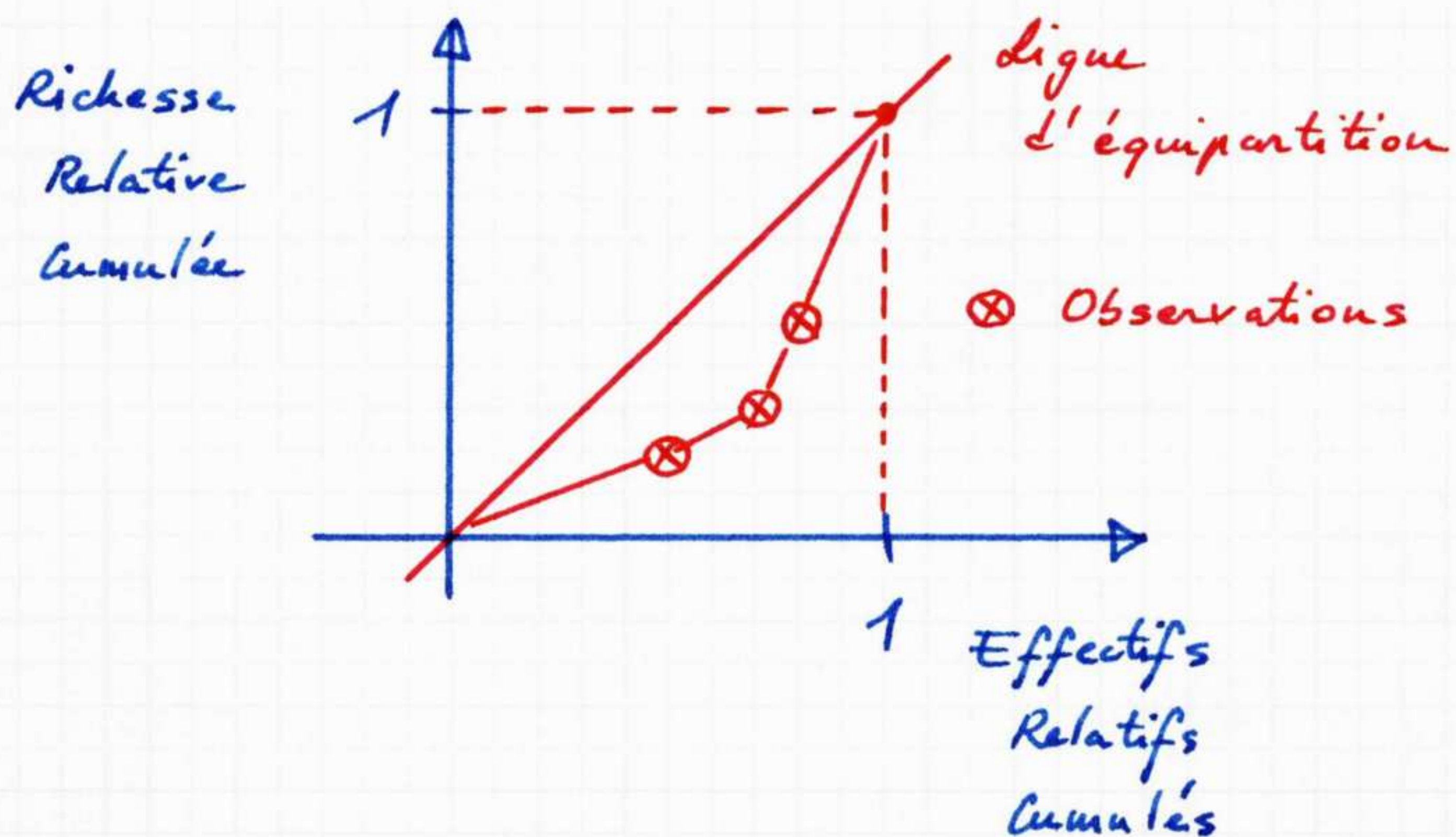




Q1) La courbe de Lorenz est une façon de représenter la répartition de ressources aux sein d'effectifs

Elle s'obtient en traçant, sur l'axe des abscisses, les effectifs relatifs cumulés, et sur l'axe des ordonnées la richesse relative cumulée.

Graphiquement, on a :






Q1) Dans le cas qui nous intéresse ici, les effectifs sont le nombre de chubs dans chaque classe.

On a :

1	3	5	11	Effectifs
1	4	9	20	Effectifs Cumulés
$1/20$	$4/20$	$9/20$	$20/20$	Effectifs Relatifs Cumulés
0.05	0.2	0.45	1	



les ressources (ou richesses) sont les budgets dédiés à chaque classe.

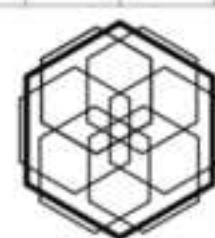
On a :

300	315	315	309	# Richesses
300	615	930	1239	Richesses Cumulées
$300/1239$	$615/1239$	$930/1239$	$1239/1239$	Richesses Relatives Cumulées
0.24	0.49	0.75	1	

Ordonnée

Le point (0.05; 0.24) appartient à la courbe de Lorenz.

Réponse C





Q<sub>2</sub> Le mode est la valeur dont l'observation est la plus fréquente

Observations :

3  
0 0 0 0 0 ←  
1 1 1 1            2 modes  
2 2 2 2 2 ←  
4

La distribution est bimodale

Réponse E

Q<sub>3</sub> Cramers-V est un nombre qui mesure le degrés d'association / le degrés de dépendance / qu'il y a entre deux variable qualitative (catégorielle)

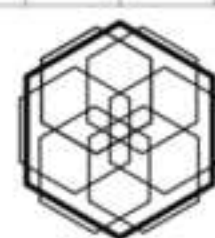
C'est un nombre compris en 0 et 1

$$0 < V < 1$$

0 signifie pas d'association

(pas possible de prédire une variable en connaissant l'autre)

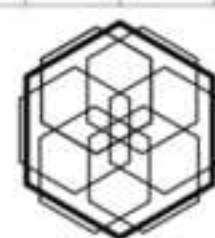
1 signifie association entre les deux variable étudiée





Cramér's  $V$  est un Chi-carré ( $\chi^2$ ) normalisé  
Il se calcule donc à partir de ce dernier  
On a la formule :

$$V = \left( \frac{\chi^2}{N \cdot \min(l-1; c-1)} \right)^{1/2}$$







Q3 Avec :  $\chi^2 = \text{Chi-carré}$

$N$  = Nombre d'observation

$l$  = nombre de lignes

$c$  = nombre de colonnes

$\min(l-1; c-1)$  = le plus petit nombre entre  $l-1$  et  $c-1$

On commence par organiser les observations dans une table

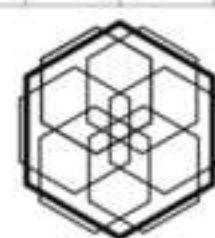
On se propose de classer le nombre de transactions dans deux catégories :

Cat 1 : entre 0 et 10 transactions

Cat 2 : entre 10 et + transactions

On a deux groupes de personne : marié et célibataire

Mariés :		Célib. :	
12		9	Cat 1
23		8	Cat 1
2	Cat 1	14	Cat 2
7	Cat 1	18	Cat 2
23	Cat 2	2	Cat 1
0	Cat 1	10	Cat 1
11	Cat 2	2	Cat 1
17	Cat 2	32	Cat 2
1	Cat 1	10	Cat 1
6	Cat 1	0	Cat 1
10	Cat 2	11	Cat 2
5      6		7      4	





Q3 Les données sous forme de table :

Observations	Echantillon, Population		
	Mariés	Célibataires	Tot
Cat 1 : 0-10	6	7	13
Cat 2 : 10+	5	4	9
Tot.	11	11	22

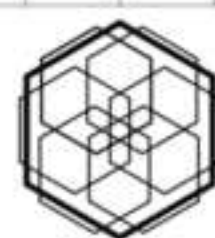
On commence par calculer le  $\chi^2$  qui est une mesure de la déviations des observations par rapport a des effectifs théoriques

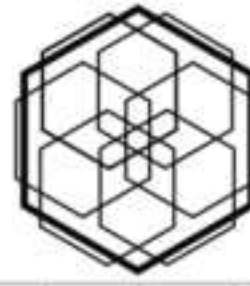
Effectifs théoriques :

	Mariés	Célibs	Tot		Mariés	Célibs	Tot
Cat 1	$\frac{11 \cdot 13}{22}$	$\frac{11 \cdot 13}{22}$	13	→	6.5	6.5	13
Cat 2	$\frac{11 \cdot 9}{22}$	$\frac{11 \cdot 9}{22}$	9		4.5	4.5	9
Tot	11	11	22		11	11	22

Calcul du  $\chi^2$  :

$$\begin{aligned} \chi^2 &= \frac{(6-6.5)^2}{6.5} + \frac{(7-6.5)^2}{6.5} + \frac{(5-4.5)^2}{4.5} + \frac{(4-4.5)^2}{4.5} \\ &= 0.04 + 0.04 + 0.06 + 0.06 \\ &= \underline{0.2} \end{aligned}$$





Calcul de Gamers -  $V$  :

$$V = \left( \frac{\chi^2}{N \cdot \min(l-1; c-1)} \right)^{1/2}$$

$$= \sqrt{\frac{0.2}{22 \cdot 1}}$$

$$= \underline{\underline{0.09}} \quad \underline{\text{Réponse C}}$$

(Q4)  $V$  est très faible  $0 < V < 1$

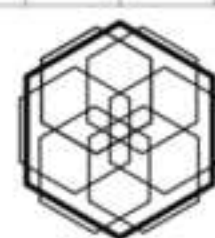
les deux variables sont donc  
quasi-indépendantes.

Réponse E

Attention : le nombre  $V$  va dépendre  
des catégories dans lesquelles  
on range les observations



La valeur d'un nombre, d'une  
statistique, dépend énormément  
de la manière dont on la calcule.





Q6) la commande `sd` de R calcule  
l'écart-type "s" d'un échantillon

Nous utilisons donc la formule suivante :

écart-type d'un échantillon = s

variance d'un échantillon :  $s^2$

$$s^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$$

! la formule n'est pas la même que pour  
l'écart-type et la variance d'un population

$n$  = nombre de mesure

$x_i$  = observation  $i$

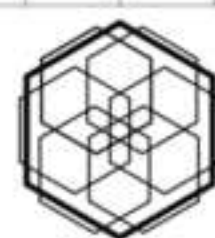
$\bar{x}$  = moyenne des observation

Pour calculer la variance  $s^2$  il faut donc  
d'abord calculer la moyenne  $\bar{x}$

$$\bar{x} = (+1 + 1 + 0 - 1 + 0 - 1) \cdot \frac{1}{6}$$

$$= 0 \cdot \frac{1}{6}$$

$$= 0$$





Q6) Avec les observations suivantes :

+1   +1   0   -1   0   -1

ayant pour moyenne  $\bar{X} = 0$

on calcule la variance comme suit :

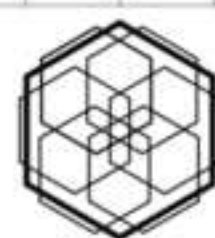
$$\begin{aligned} s^2 &= \frac{1}{6-1} \left( (1-0)^2 + (1-0)^2 + (0-0)^2 + \right. \\ &\quad \left. (-1-0)^2 + (0-0)^2 + (-1-0)^2 \right) \\ &= \frac{1}{5} (1+1+1+1) \\ &= \frac{4}{5} \end{aligned}$$

L'écart-type de cet échantillon, tel que calculé par R avec la commande sd, est donc

$$s = \sqrt{\frac{4}{5}}$$

$$= \underline{\underline{0.894}}$$

Réponse E





Q7 On a :

	A	N	Tot
H	101	97	198
F	68	134	202
Tot	169	231	400

Si les deux facteurs, Aromatisé et Naturel, étaient indépendant, les effectifs observés seraient les mêmes que les effectifs théoriques.

Effectifs théoriques :

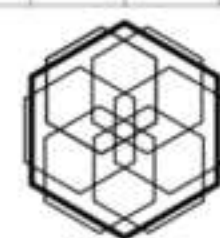
$$H_A : \frac{169 \cdot 198}{400} = 83.65$$

$$H_N : \frac{231 \cdot 198}{400} = 114.34$$

$$F_A : \frac{169 \cdot 202}{400} = 85.3$$

$$F_N : \frac{231 \cdot 202}{400} = 116.65$$

Réponse A





Q10

On commence par organiser les données dans un tableau :

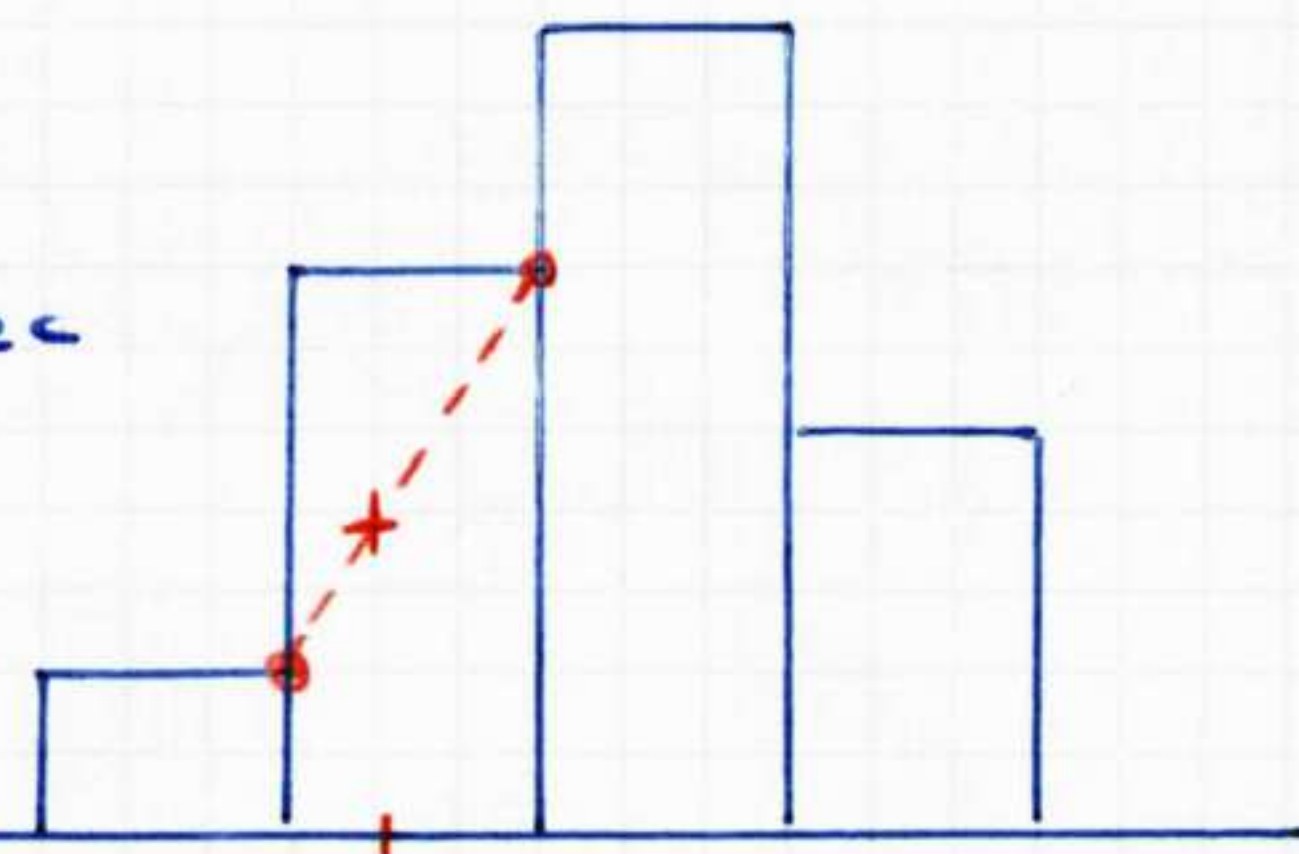
Classe	Valeur Centrale $\{x_i\}$	Nombre d'Observations $\{f(x_i)\}$
0-5	2.5	550
5-10	7.5	256
10-15	12.5	203
15-20	17.5	144
20-50	35	97
		1250

Moyenne groupe  $\bar{x}_g = \frac{1}{N} \sum_i x_i f(x_i)$

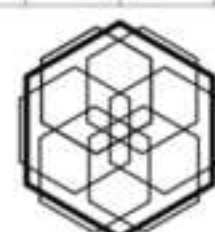
$$\bar{x}_g = \frac{1}{1250} (2.5 \cdot 550 + 7.5 \cdot 256 + 12.5 \cdot 203 + 17.5 \cdot 144 + 35 \cdot 97)$$

= 9.4

Lorsque l'on travail avec des données groupées, on situe d'abord le quantile ou la médiane dans la classe qui la contient. On procède ensuite, au sein de cette classe, à une



Interpolation linéaire





Q10) Le premier quartile ( $Q_1 = 0.25$ ) se situe dans la classe dont la fréquence relative cumulée  $> 0.25$  (25%)

On calcule :

2.5	7.5	12.5	17.5	35	$X_i$
550	256	203	144	97	Effectifs
550	806	1009	1153	1250	Eff. Cumulés
$\frac{550}{1250}$	$\frac{806}{1250}$	$\frac{1009}{1250}$	$\frac{1153}{1250}$	$\frac{1250}{1250}$	} Eff. Relatifs Cumulés
0.44	0.644	0.8	0.92	1	



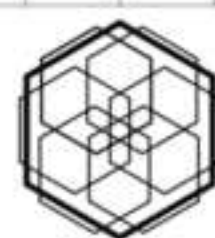
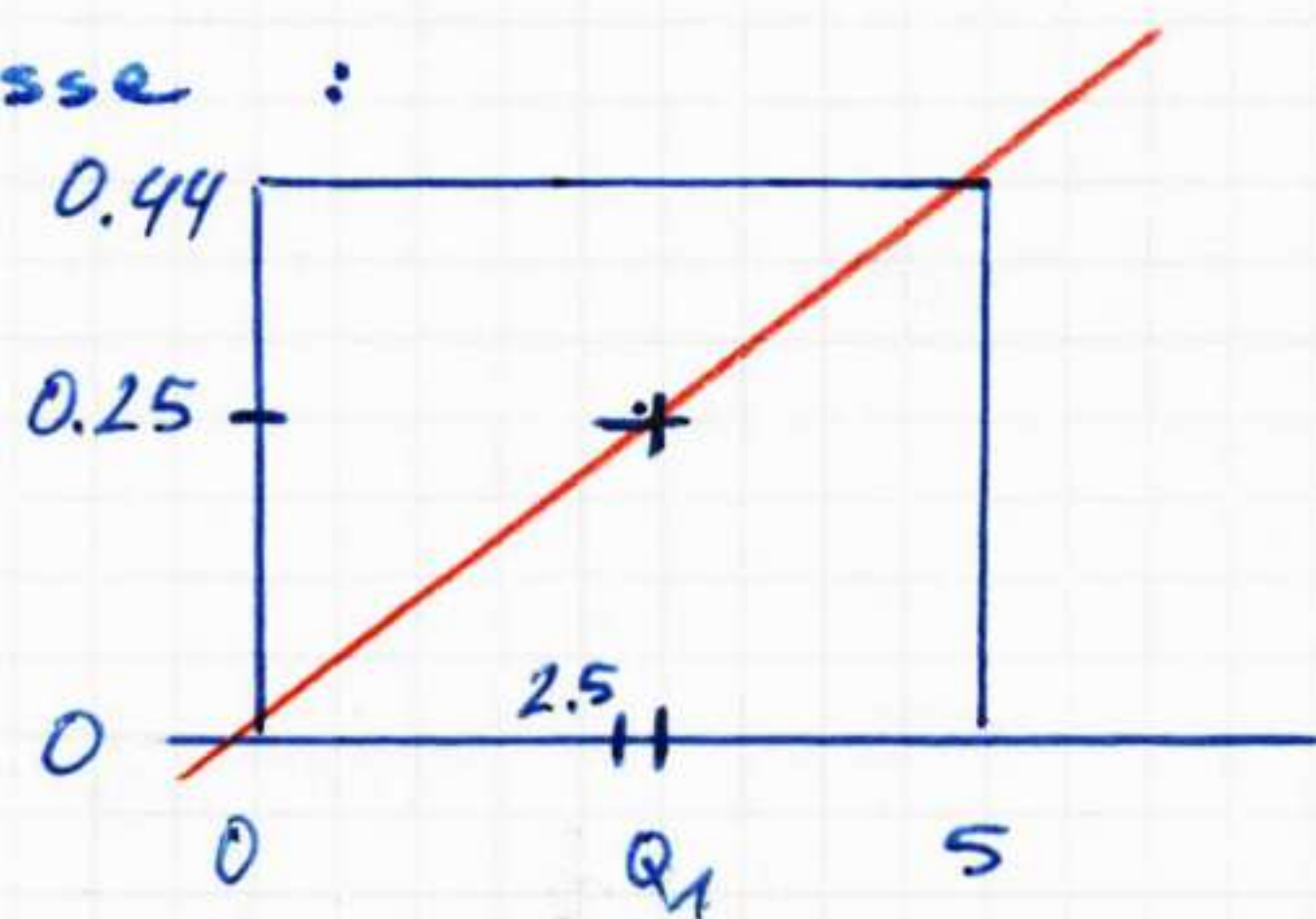
Le 1<sup>er</sup> Quartile est dans cette catégorie

On procède à l'interpolation linéaire au sein de cette classe :

$$\begin{aligned} \text{pente} &= \frac{0.44 - 0}{5 - 0} \\ &= 0.088 \end{aligned}$$

$$\begin{aligned} \text{pente} &= \frac{0.25 - 0}{Q_1 - 0} \\ &= 0.088 \end{aligned}$$

$$\Rightarrow Q_1 = 0.25 / 0.088 = \underline{\underline{2.84}}$$





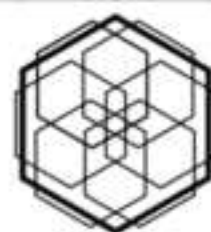


On a donc, pour la moyenne groupée  $\bar{X}_g$   
et le premier quantile  $Q_1$  les valeurs

$$\underline{\bar{X}_g = 9.4}$$

$$\underline{Q_1 = 2.8}$$

Réponse C





Q12) le coefficient de corrélation (Pearson) est une mesure de la dépendance linéaire entre deux variables

Au sein d'une population, il se mesure de la manière suivante :

Coef. corrélation entre X et Y =  $r_{xy}$

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$\sigma_{xy}$  = covariance entre X et Y

$$= \frac{1}{N} \sum_i x_i y_i - \mu_x \mu_y$$

$x_i$  : résultat obtenu pour l'épreuve A

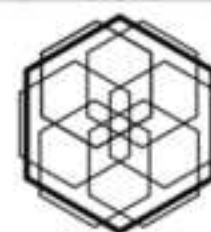
$y_i$  : résultat obtenu pour l'épreuve B

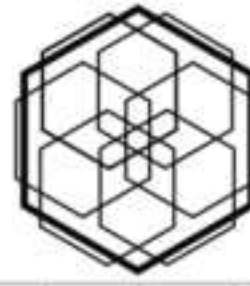
$\mu_x$  = moyenne des  $x_i$

$\mu_y$  = moyenne des  $y_i$

$N$  = nombre de personnes ayant fait les 2 examens

= nombre de "mesures"





Epreuve A :

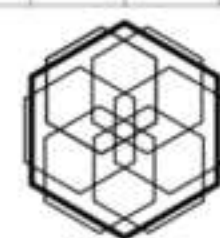
$$\begin{aligned}\mu_x &= \frac{1}{12} (3 + 4 + 6 + 7 + 9 + 10 + 9 + 11 + 12 + \\ &\quad 13 + 15 + 4) \\ &= \frac{1}{12} \cdot 103 \\ &= \underline{8.58}\end{aligned}$$

Epreuve B :

$$\begin{aligned}\mu_y &= \frac{1}{12} \cdot 155 \\ &= \underline{12.91}\end{aligned}$$

$$\begin{aligned}\sum x_i y_i &= 3 \cdot 8 + 4 \cdot 9 + 6 \cdot 10 + 7 \cdot 13 + 9 \cdot 15 + \\ &\quad 10 \cdot 14 + 9 \cdot 13 + 11 \cdot 16 + 12 \cdot 13 + \\ &\quad 13 \cdot 19 + 15 \cdot 6 + 4 \cdot 19 \\ &= \underline{1348}\end{aligned}$$

$$\begin{aligned}\sigma_{xy} &= \frac{1}{N} \sum_i x_i y_i - \mu_x \mu_y \\ &= \frac{1}{12} \cdot 1348 - 8.58 \cdot 12.91 \\ &= \underline{1.47}\end{aligned}$$





Q12

Variance des  $x_i = V(x)$  ! Population !

$$V(x) = \sigma_x^2 = \frac{1}{N} \sum_i (x_i - \mu_x)^2$$

$$= \frac{1}{12} \left( (3-8.58)^2 + (4-8.58)^2 + (6-8.58)^2 + (7-8.58)^2 + (9-8.58)^2 + (10-8.58)^2 + (9-8.58)^2 + (11-8.58)^2 + (12-8.58)^2 + (13-8.58)^2 + (15-8.58)^2 + (4-8.58)^2 \right)$$

$$= \frac{1}{12} \left( 31.13 + 20.97 + 6.65 + 2.49 + 0.18 + 2.01 + 0.18 + 5.86 + 11.69 + 19.54 + 41.22 + 20.97 \right)$$

$$= \underline{13.58}$$

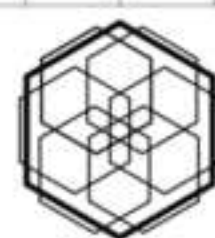
$$V(y) = \sigma_y^2 = \frac{1}{N} \sum_i (y_i - \mu_y)^2$$

$$= \underline{15.41}$$

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

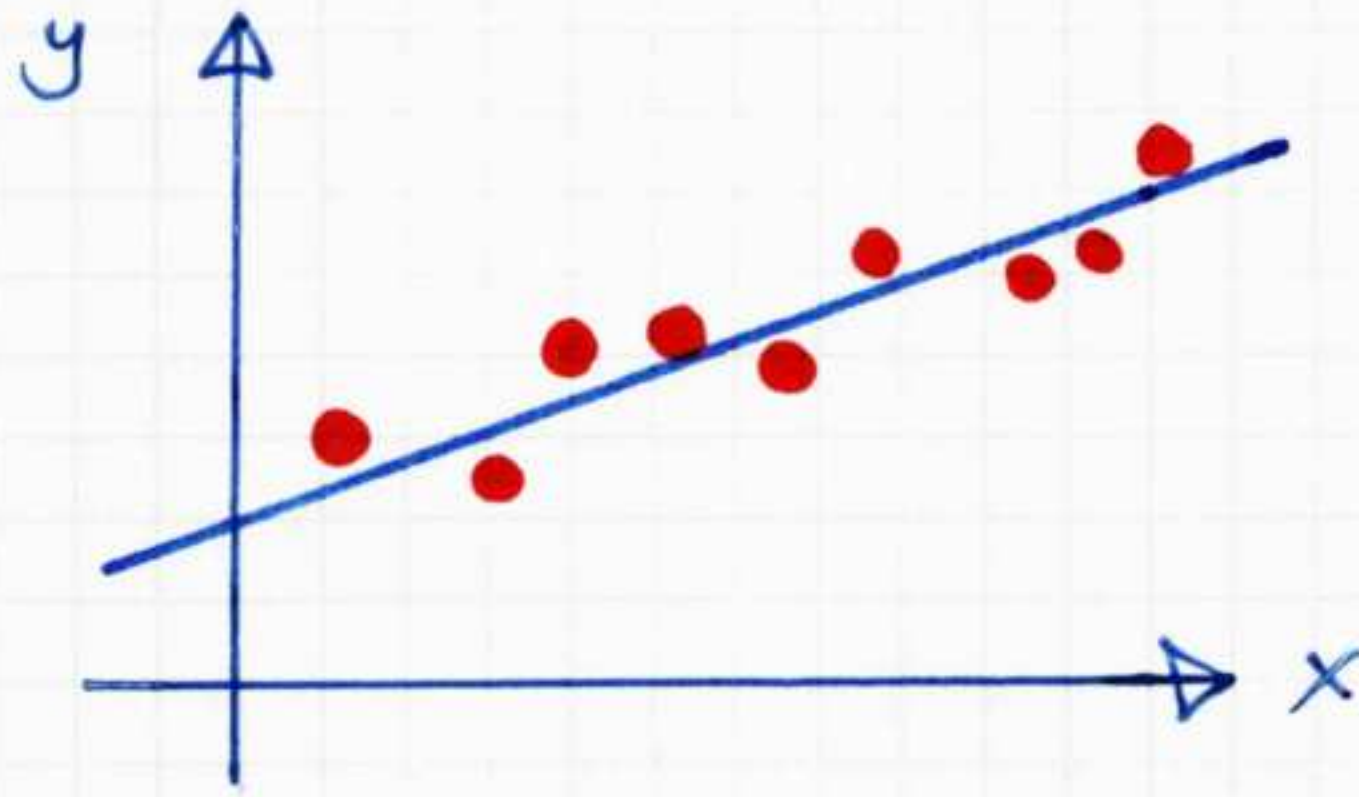
$$= \frac{1.47}{\sqrt{13.58} \cdot \sqrt{15.41}}$$

$$= \underline{0.1} \quad \text{Réponse C}$$

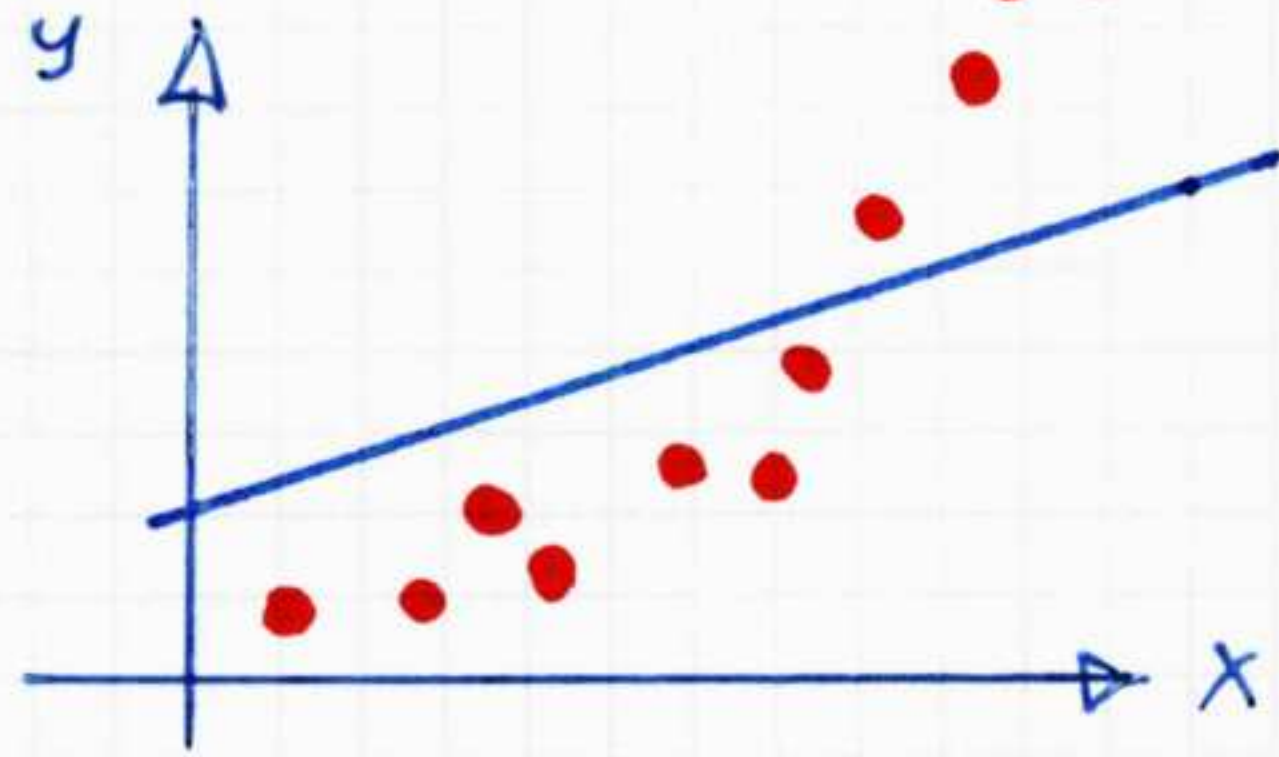




Q13 Le coefficient de corrélation  $r_{xy}$  nous dit à quel point il existe une dépendance linéaire entre deux séries de valeurs.



$r_{xy}$  proche de 1



$r_{xy}$  proche de 0

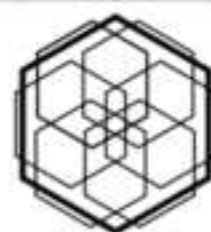
Quelle que soit la valeur de  $r_{xy}$  cependant on peut toujours calculer la droite de régression qui passe au plus près de chaque point.

$r_{xy}$  nous dira simplement si cette droite est utile ou non.

L'équation de la droite de régression est donnée par :

$$y = ax + b$$

Les statistiques nous donnent un moyen de calculer les paramètres  $a$  et  $b$ .





Q13

La pente  $a$  est donnée par :

$$a = \frac{\sigma_{xy}}{\sigma_x^2}$$

L'ordonnée à l'origine  $b$  est donnée par :

$$b = \mu_y - a\mu_x$$

Dans notre cas particulier, on a donc :

$$a = \frac{1.47}{15.41} = \underline{0.095}$$

$$b = 12.91 - 0.095 \cdot 8.58 = \underline{12.09}$$

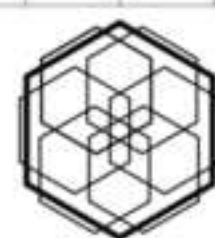
La droite de régression a donc pour équation

$$\underline{y = 0.095x + 12.09} \quad \text{Réponse C}$$

!

Le coef  $r_{xy} = 0.1$  cependant nous dit que la dépendance linéaire entre  $x$  et  $y$  est mauvaise

Il faut donc s'attendre à des résultats imprécis si l'on utilise cette équation.





Q14) En arrondissant, on trouve l'équation suivante :

$$y = 0.1x + 12$$

Si l'on nous donne  $x$ , on peut calculer  $y$ , et inversement.

Ici, on nous donne  $x = 14$

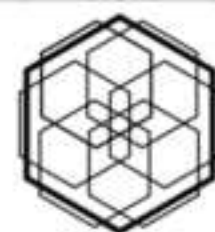
Quel est  $y$  ?

On pose :

$$y = 0.1 \cdot 14 + 12$$

$$= \underline{\underline{13.4}}$$

Réponse A





Q15) Le coefficient de détermination  $R^2$  est une autre mesure (une autre statistique) de la pertinence de la droite de corrélation entre deux variables.

Elle est donnée par le carré du coef. de corrélation linéaire entre les  $x_i$  et les  $y_i$  :

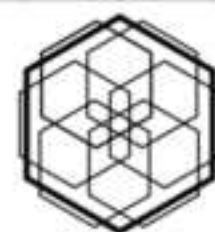
$$\begin{aligned} R^2 &= r_{xy}^2 \\ &= 0.1^2 \\ &= 0.01 \end{aligned}$$

$$0 < R^2 < 1$$

0.01 est proche de 0

$\Rightarrow$  le modèle linéaire est mauvais

Réponse E







Q16) On a affaire à des classes, on va donc calculer une moyenne groupée à l'aide des valeurs centrales :

Classes ( $10^3$ CHF)	Valeurs Centrales	Ocurrence
10 - 12	11	8
12 - 15	13.5	10
15 - 18	16.5	16
18 - 20	19	6

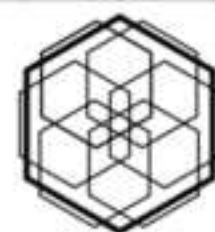
40

Moyenne groupée  $\mu_g$  :

$$\mu_g = \frac{1}{40} ( 11 \cdot 8 + 13.5 \cdot 10 + 16.5 \cdot 16 + 19 \cdot 6 )$$
$$= 15.025$$

Réponse : 15 025

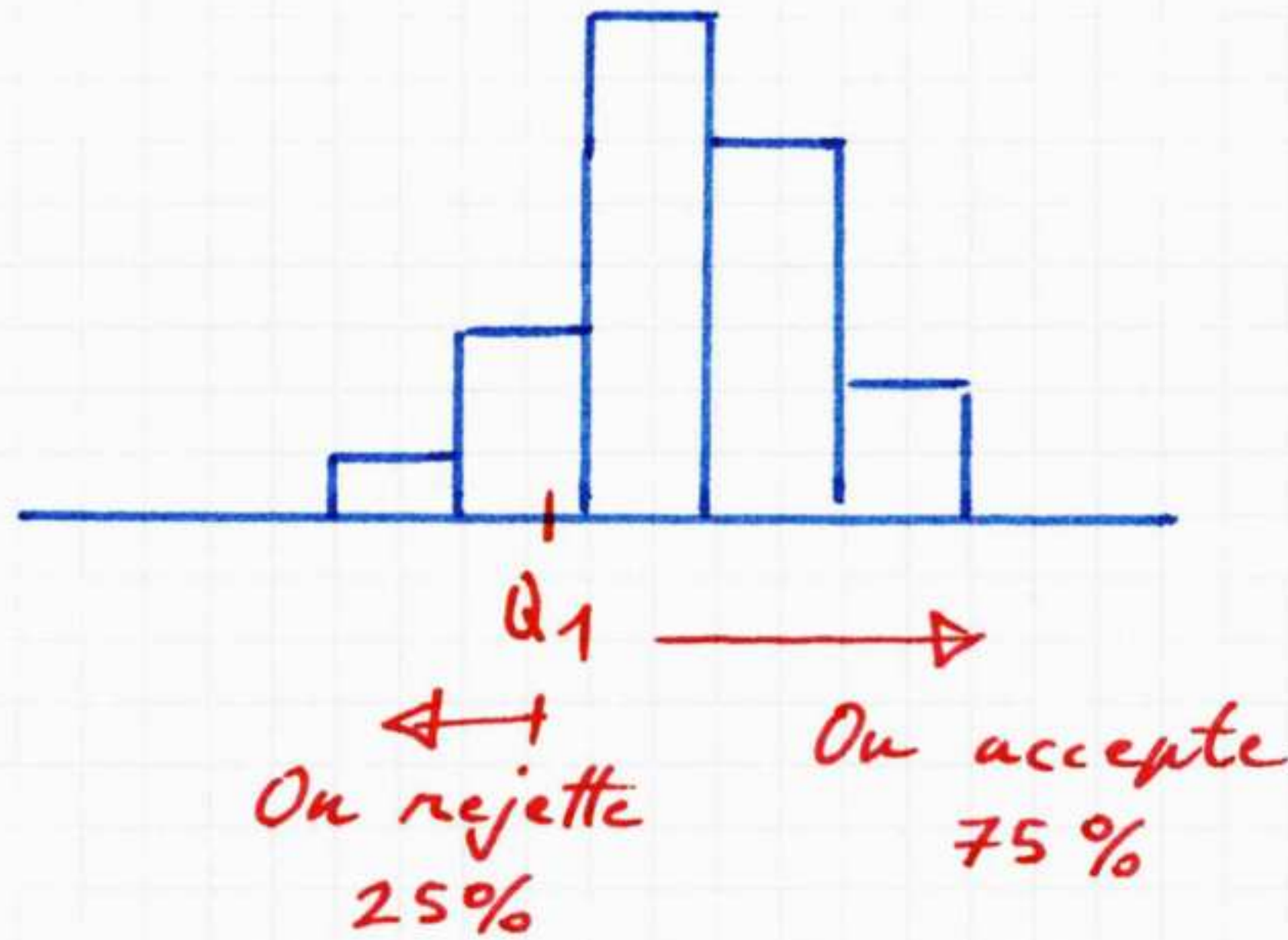
Réponse C





Q17

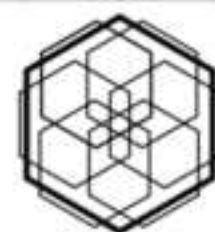
On nous demande de calculer le premier quantile  $Q_1$  afin de situer (déterminer) quel montant délimite le 25% des offres

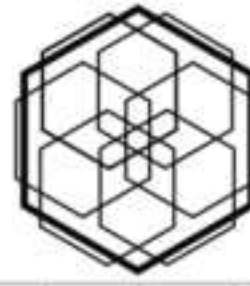


Première étape : dans quelle classe  $Q_1$  se situe-t-il ?

Calcul des effectifs relatifs cumulés :

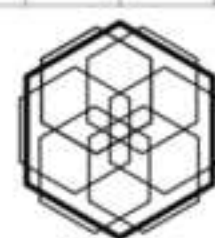
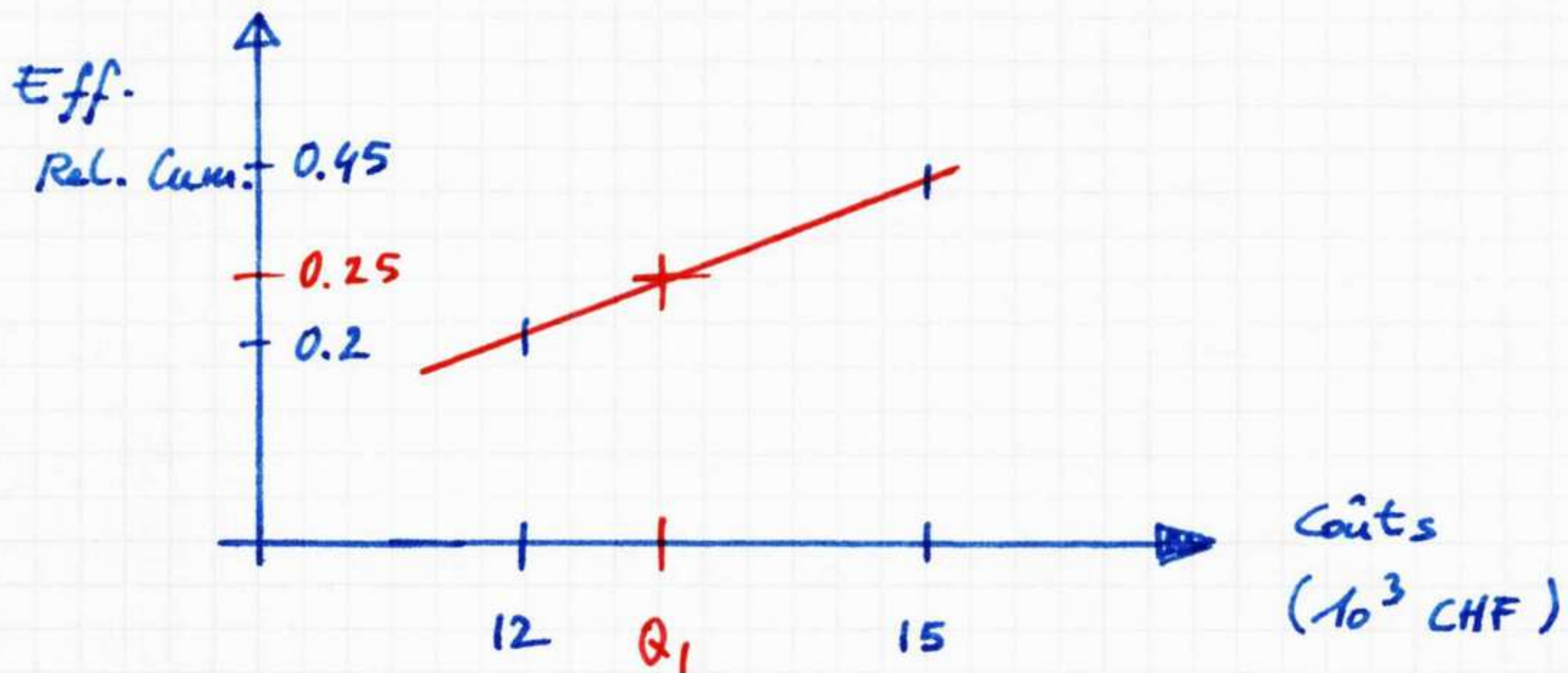
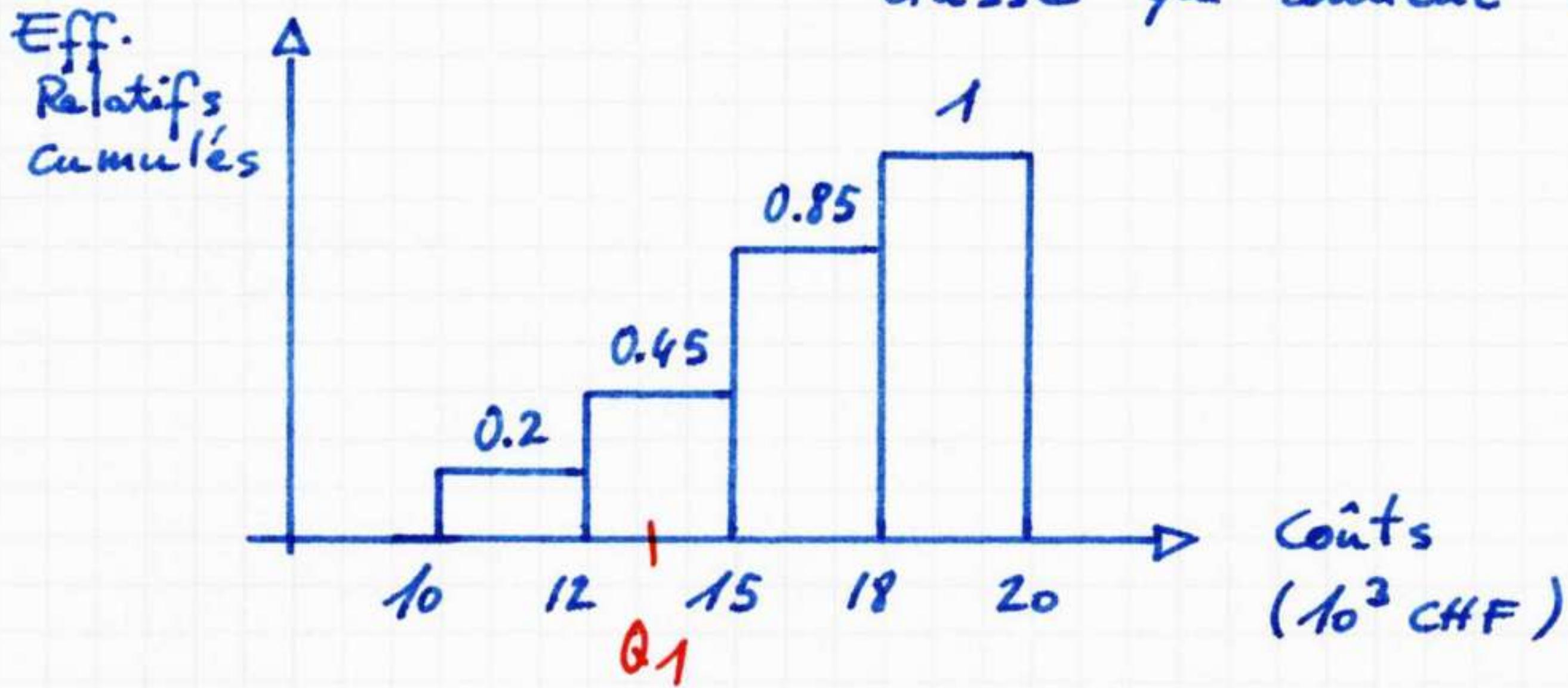
11	13.5	16.5	19	
8	10	16	6	Effectifs
8	18	34	40	Eff. cumulé
8/40	18/40	34/40	40/40	} Eff. Relatifs cumulé
0.2	0.45	0.85	1	
	↑	le 1 <sup>er</sup> quantile ( $Q_1$ ) se situe dans la 2 <sup>e</sup> catégorie.		
	0.25			





Q17

Deuxième étape : On pratique une interpolation linéaire au sein de la classe qui contient  $Q_1$  :





Q17

$$\text{pente} = \frac{0.45 - 0.2}{15 - 12} = 0.08\bar{3}$$

$$\text{pente} = \frac{0.25 - 0.2}{Q_1 - 12} = 0.08\bar{3}$$

$$\Rightarrow (Q_1 - 12) \cdot 0.08\bar{3} = 0.25 - 0.2$$

$$Q_1 \cdot 0.08\bar{3} - 12 \cdot 0.08\bar{3} = 0.25 - 0.2$$

$$Q_1 \cdot 0.08\bar{3} = 0.25 - 0.2 + 12 \cdot 0.08\bar{3}$$

$$Q_1 = \frac{1}{0.08\bar{3}} \cdot (0.25 - 0.2 + 12 \cdot 0.08\bar{3})$$

$$= 12.6$$

Réponse : 12.600      Réponse A



