A Médiane et quartiles

A.1 Questions de cours

1 Donner une définition des éléments suivants :

1. la médiane d'une série statistique ;
2. le premier quartile d'une série statistique.

A.2 Faire ses gammes

2 Pour chacune des séries statistiques suivantes, déterminer la médiane, les quartiles et l'écart interquartile.

1. 1 ; 2 ; 5 ; 5 ; 10 ; 15 ; 29 ; 36 ; 43.

Les valeurs sont déjà rangées dans l'ordre croissant.
L'effectif total est 9.
9 est impair donc la médiane est la $\frac{9+1}{2}$-ième, soit la 5-ième valeur.
Ainsi : $M = 10$.
$\frac{9}{4} = 2{,}25$, donc Q_1 est la 3-ième valeur, soit 5.
$\frac{3\times 9}{4} = 6{,}75$, donc Q_3 est la 7-ième valeur, soit 29.
Ainsi : $EI = Q_3 - Q_1 = 29 - 5 = 24$.

2. 3 ; 8 ; 9 ; 12 ; 13 ; 15 ; 17 ; 20 ; 23 ; 25.

Les valeurs sont déjà rangées dans l'ordre croissant.
L'effectif total est 10.
10 est pair.
$\frac{N}{2} = \frac{10}{2} = 5$.
Donc la médiane est la moyenne entre la 5-ième et la 6-ième valeurs.
Ainsi : $M = \frac{13+15}{2} = 14$.
$\frac{10}{4} = 2{,}5$, donc Q_1 est la 3-ième valeur, soit 9.
$\frac{3\times 10}{4} = 7{,}5$, donc Q_3 est la 8-ième valeur, soit 20.
Ainsi : $EI = Q_3 - Q_1 = 20 - 9 = 11$.

3. 18 ; 11 ; 2 ; 2 ; 3 ; 9 ; 21 ; 7 ; 13.

On commence par ranger les valeurs dans l'ordre croissant :

2 ; 2 ; 3 ; 7 ; 9 ; 11 ; 13 ; 18 ; 18

L'effectif total est 9.
9 est impair donc la médiane est la $\frac{9+1}{2}$-ième, soit la 5-ième valeur.
Ainsi : $M = 9$.
$\frac{9}{4} = 2{,}25$, donc Q_1 est la 3-ième valeur, soit 3.
$\frac{3\times 9}{4} = 6{,}75$, donc Q_3 est la 7-ième valeur, soit 13.
Ainsi : $EI = Q_3 - Q_1 = 13 - 3 = 10$.

4. 10 ; 5 ; 3 ; 8 ; 6 ; 10 ; 14 ; 2 ; 14 ; 13 ; 12 ; 11.

On commence par ranger les valeurs dans l'ordre croissant :

2 ; 3 ; 5 ; 6 ; 8 ; 10 ; 10 ; 11 ; 12 ; 13 ; 14 ; 14

L'effectif total est 12.
12 est pair.
$\frac{N}{2} = \frac{12}{2} = 6$.
Donc la médiane est la moyenne entre la 6-ième et la 7-ième valeurs.
Ainsi : $M = \frac{10+10}{2} = 10$.
$\frac{12}{4} = 3$, donc Q_1 est la 3-ième valeur, soit 5.
$\frac{3\times 12}{4} = 9$, donc Q_3 est la 9-ième valeur, soit 12.
Ainsi : $EI = Q_3 - Q_1 = 12 - 5 = 7$.

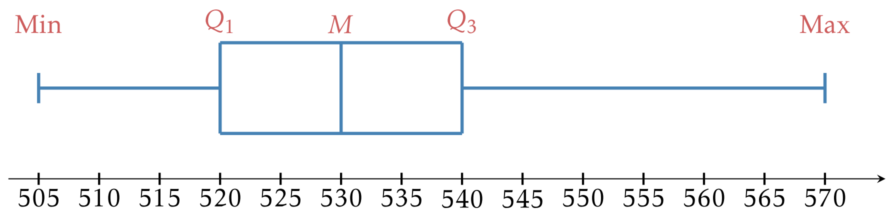
3 La masse nette inscrite sur des paquets de barres énergétiques est de 160 grammes.
Afin de vérifier que la production est conforme à cette valeur, le service de qualité prélève un échantillon de 20 paquets et pèse chacun de ces paquets.
Voici la liste ordonnée des valeurs obtenues (en grammes) :
157,5 ; 158 ; 158 ; 160,5 ; 161 ; 161,5 ; 161,5 ; 162 ; 162 ; 162 ; 162 ; 162,5 ; 162,5 ; 163 ; 163 ; 163,5 ; 164 ; 164,5 ; 164,5 ; 164,5.
Les critères de conformité sont les suivants : la médiane ne doit pas s'écarter de plus de 2 grammes de la masse inscrite sur les boîtes, l'écart interquartiles ne doit pas dépasser 2 % du poids inscrit sur les boîtes et l'étendue (max-min) ne doit pas dépasser 5 % du poids inscrit sur les boîtes. La production est-elle conforme ?

Masse	157,5	158	160,5	161	161,5	162	162,5	163	163,5	164	164,5
Eff.	1	2	1	1	2	4	2	2	1	1	3
ECC	1	3	4	5	7	11	13	15	16	17	20

N est pair.
$\frac{N}{2} = 10$. La médiane est la moyenne entre la $10^{\text{ème}}$ et la $11^{\text{ème}}$ valeur.
Ainsi : $M = \frac{162+162}{2} = 162$.
La masse inscrite sur les boîtes est 160, la médiane ne s'écarte donc pas de plus de de 2 grammes de la masse inscrite.
$\frac{N}{4} = 5$, donc Q_1 est la $5^{\text{ème}}$ valeur, soit 161.
$\frac{3N}{4} = 15$, donc Q_3 est la $15^{\text{ème}}$ valeur, soit 163.
Ainsi ; $EI = Q_3 - Q_1 = 163 - 161 = 2$.
Que représente 2 % de la masse inscrite ?
$160 \times \frac{2}{100} = 3{,}2$.
Donc l'écart interquartiles est bien inférieur à 2 % de la masse inscrite.
Étendue $= \max - \min = 164{,}5 - 157{,}5 = 7$.
Que représente 5 % de la masse inscrite ?

$160 \times \frac{5}{100} = 8$.
Donc l'étendue ne dépasse pas 5 % de la masse inscrite.
Tous les critères étant vérifiés, on en déduit que la production est bien conforme.

4 Le kendo est un art martial traditionnel japonais qui se pratique principalement avec un shinai, un sabre de bambou.
Ci-dessous, on a tracé le diagramme en boîte portant sur la masse (en g) d'un échantillon de shinais fabriqués par Kumamoto.



1. Lire la valeur de la médiane puis en donner une interprétation.
2. Écrire trois phrases commençant par "Environ 50 % des shinais ont une masse comprise entre ...".
3. Calculer l'écart interquartiles.
4. Pour être homologué, la masse d'un shinai doit dépasser 510 g. Que penser de la production de Kumamoto ?

1. $M = 530$. Au moins 50 % des shinais ont une masse inférieure ou égale à 530 g et au moins 50 % ont une masse supérieure ou égale à 530 g.
2. Environ 50 % des shinais ont une masse comprise entre 505 et 530 g.
   Environ 50 % des shinais ont une masse comprise entre 520 et 530 g.
   Environ 50 % des shinais ont une masse comprise entre 530 et 570 g.
3. $EI = Q_3 - Q_1 = 540 - 520 = 20$.
4. On sait que $Q_1 = 520$. Donc on peut affirmer qu'au moins 75 % des shinais ont une masse supérieure ou égale à 520 g, et donc par extension on peut aussi affirmer qu'au moins 75 % des shinais ont une masse supérieure ou égale à 510 g et sont donc homologués.

A.3 Exercices d'entraînement

5 Voici le pourcentage de femmes dans les parlements des 23 pays es plus peuplés du monde en 2018.
Thaïlande : 5 % ; Nigéria : 6 % ; Iran : 6 % ; Japon : 10 % ; Congo : 11 % ; Brésil : 11 % ; Inde : 12 % ; Égypte : 15 % ; Turquie : 15 % ; Russie 16 % ; U.S.A : 19 % ; Indonésie : 20 % ; Bengladesh : 20 % ; Pakistan : 22 % ; Chine : 25 % ; Vietnam : 27 % ; Philippines : 29 % ; Allemagne : 31 % ; Royaume-Uni : 32 % ; Italie : 35 % ; Éthiopie : 39 % ; France : 39 % ; Mexique : 42 %.
Source : http://archive.ipu.org/wmn-f/classif.htm.)

1. Calculer la médiane Me de cette série.
2. Interpréter ce résultat en utilisant les mots : femmes - pays - parlement - moins - moitié - %.
3. Calculer les quartiles Q_1 et Q_3.
4. Interpréter chacun de ces deux résultats.
5. Calculer l'écart interquartiles.

1. On dresse le tableau des effectifs cumulés croissants.

%	5	6	10	11	12	15	16	19	20	22	25	27	29	31	32	35	39	42
Eff.	1	2	1	2	1	2	1	1	2	1	1	1	1	1	1	1	2	1
ECC	1	3	4	6	7	9	10	11	13	14	15	16	17	18	19	20	22	23

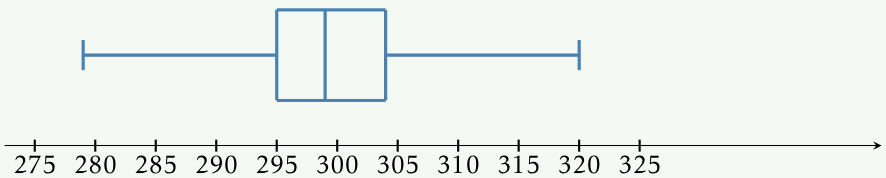
   N est impair. $\frac{N+1}{2} = \frac{23+1}{2} = 12$.
   Donc la médiane est la $12^{\text{ème}}$ valeur, soit 20.
2. Au moins la moitié des pays ont un parlement constitué de moins de 20 % de femmes.
3. $\frac{N}{4} = \frac{23}{4} = 5{,}75$.
   Donc Q_1 est la $6^{\text{ème}}$ valeur, soit 11.
   $\frac{3N}{4} = \frac{23 \times 3}{4} = \frac{69}{4} = 17{,}25$.
   Donc Q_3 est la $18^{\text{ème}}$ valeur, soit 31.
4. Au moins 25 % des pays ont un parlement constitué de 20 % de femmes ou moins.
   Au moins 75 % des pays ont un parlement constitué de 31 % de femmes ou moins.

6 Une fabrique de céréales pour le petit-déjeuner confectionne des paquets de 300 g.
Les machines prévues pour le remplissage sont testées régulièrement.
Voici les résultats d'un test portant sur un échantillon de 997 paquets.

Masse (g)	279	280	281	282	283	284	285	286	287	288	289
Effectif	1	5	5	1	5	7	8	9	16	15	22
Masse (g)	290	291	292	293	294	295	296	297	298	299	300
Effectif	25	23	33	34	33	51	45	58	55	50	61
Masse (g)	301	302	303	304	305	306	307	308	309	310	311
Effectif	50	46	50	51	39	29	31	23	24	15	27
Masse (g)	312	313	314	315	316	317	318	319	320	321	
Effectif	21	8	5	5	3	4	2	1	1	0	

1. À l'aide de la calculatrice, déterminer le minimum, la médiane, les 1er et 3e quartiles, le maximum et l'écart interquartile de cette série.
2. Dresser le diagramme en boîte de cette série.
3. Voici les critères retenus par la fabrique pour ses machines :
   (a) $299 \leq M \leq 301$
   (b) $295 \leq Q_1$
   (c) $Q_3 \leq 305$
   (d) $EI \in [0;9]$
   (e) moins de 1,5 % des valeurs en dehors de $[Q_1 - 1{,}5EI; Q_3 + 1{,}5EI]$.

   La machine testée réussit-elle le test ?

1. Sur la calculatrice, on va dans Stats→Modifier, puis on remplit la colonne L1 avec toutes les valeurs de 279 à 321. Dans la colonne L2 on indique tous les effectifs.
   On revient dans le menu Stats puis on suit CALC→Stats 1 Var.
   On indique bien L1 dans Xliste et L2 dans ListeFréq, puis on clique sur Calculer.
   On obtient :
   - $Q_1 = 295$ • $M = 299$ • $Q_3 = 304$
   - min = 279 et max = 320.
   - $EI = Q_3 - Q_1 = 304 - 295 = 9$.
2. 

3. Les 4 premiers critères sont vérifiés.
   Combien représente 1,5 % de 997 ?
   $997 \times \frac{1{,}5}{100} = 14{,}955$.
   $[Q_1 - 1{,}5EI; Q_3 + 1{,}5EI] = [295 - 1{,}5 \times 9; 304 + 1{,}5 \times 9] = [281{,}5; 317{,}5]$.
   Vérifions donc si nous avons bien moins de 14,955 paquets (c-à-d 14 ou moins) qui n'appartiennent pas à cet intervalle.
   Il y en a $1 + 5 + 5 = 11$ dont la masse est inférieure ou égale à 281.
   Il y en a $2 + 1 + 1 = 4$ dont la masse est supérieure ou égal à 318.
   Il y a donc 15 paquets dont la masse est en dehors de l'intervalle, soit un peu plus de 1,5% des valeurs.
   La machine ne réussit donc pas le test.

B Moyenne, variance, écart-type

B.1 Faire ses gammes

7 Une SCOP (Société coopérative ouvrière de production) a dégagé des bénéfices cette année.
Pour l'an prochain, elle décide d'augmenter tous les salaires mensuels de 10%, puis de les augmenter de 100 €.
Cette année le salaire moyen était de 1 700 €.

1. Quel sera le salaire moyen l'an prochain ?
2. Estimer comment va évoluer l'écart type ?

1. $1\ 700 \times 1{,}1 = 1870$.
   $1\ 870 + 100 = 1\ 970$.
   Donc le nouveau salaire moyen sera de 1 970 €.
2. L'écart-type va augmenter, car multiplier les salaires par 1,1 (c'est à dire les augmenter de 10 %) creuse les inégalités. En effet un salaire plus élevé sera plus augmenté qu'un salaire plus bas.

8 Une personne compte le nombre de mail reçus chaque jour pendant un an. Les résultats sont donnés dans le tableau ci-dessous.

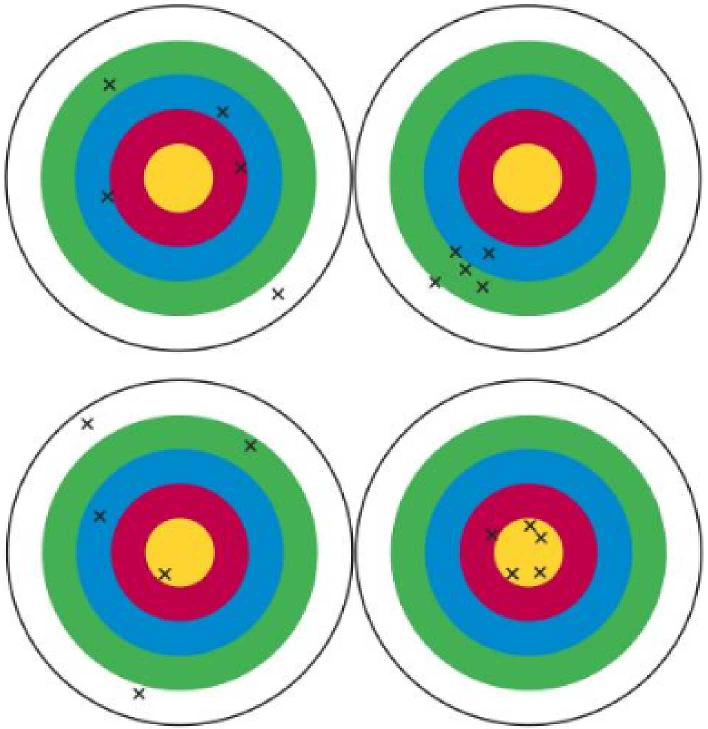
x_i	1	2	3	4	5	6	7	8	9	10
n_i	24	18	32	52	65	78	31	27	25	13

1. Calculer une valeur approchée de la moyenne $\overline{x}$ et de l'écart type σ de cette série.
2. Si leur effectif augmentait, donner les valeurs de x_i qui feraient :
   (a) baisser en même temps $\overline{x}$ et σ.
   (b) augmenter en même temps $\overline{x}$ et σ.

1. Avec la calculatrice, on trouve : $\overline{x} \approx 5{,}33$ et $\sigma \approx 2{,}26$.
2. (a) Il faut augmenter l'effectif d'une valeur inférieure à la moyenne, et qui est assez proche de la moyenne (moins d'un écart-type).
   Les valeurs 4 et 5 conviennent.
   (b) Pour augmenter à la fois la moyenne et l'écart-type, il faut choisir une valeur plus élevée que la moyenne, et qui soit assez éloignée de celle-ci (de plus d'un écart-type).
   Les valeurs 8, 9 et 10 conviennent.

B.2 Exercices d'entraînement

9 On a représenté ci-dessous quatre cibles avec l'impact de cinq flèches tirées à l'arc. Chaque couronne a un rayon de 1 unité.



Pour chaque tireur, on considère la distance de chaque flèche au centre de la cible.

1. Associer chaque couple (moyenne; écart type) à chaque cible : (2,85;0,97); (0,79;0,19); (3,33;0,55); (3,21;1,42).
2. Associer chacun des quatre tireurs à une cible : un tireur expérimenté, deux tireurs débutants et un tireur ayant mal réglé son viseur.

1. La cible en bas à droite correspond au couple (0,79;0,19). Les impacts sont au centre et tous proches.
   La cible en haut à droite correspond au couple (3,33;0,55). Les impacts sont éloignés du centre et tous proches.
   La cible en haut à gauche correspond au couple (2,85;0,97). Les impacts sont assez éloignés du centre et modérément éloignés de la moyenne.
   La cible en bas à gauche correspond au couple (3,21;1,42). Les impacts sont assez éloignés du centre et plus éloignés de la moyenne que la cible précédente.
2. Tireur expérimenté : cible en bas à droite.
   Tireur ayant mal réglé son viseur : cible en haut à droite.
   Tireurs débutants : cibles de gauche.

10 Une société a en charge l'entretien de distributeurs automatiques.
Elle a observé durant une année le nombre d'interventions réalisées sur chacun des distributeurs.

1. Déterminer le nombre moyen d'interventions $\overline{x}$, ainsi que l'écart type σ.
2. Le responsable de la société considère qu'il faut changer les distributeurs si l'intervalle $[\overline{x}-2\sigma;\overline{x}+2\sigma]$ contient moins de 95 % des valeurs de la série.
   Quelle va être sa décision ?
3. Il s'aperçoit qu'il a oublié de compter un distributeur sur lequel on a relevé 3 pannes.
   Cela change-t-il sa décision ?

Nombre d'interventions	Nombre de machines
1	10
2	12
3	17
4	44
5	78
6	94
7	83
8	49
9	36
10	16

1. À l'aide de la calculatrice : $\overline{x} \approx 6{,}09$ et $\sigma \approx 1{,}97$.
   On trouve aussi : $N = 439$.
2.
$$\begin{aligned}[\overline{x}-2\sigma;\overline{x}+2\sigma] &\approx [6{,}09-2\times1{,}97;6{,}09+2\times1{,}97]\\ &= [2{,}15;10{,}03]\end{aligned}$$

   On regarde conbien de machines ont subies entre 3 et 10 interventions inclus.
   Il y en a $439-(10+12)$, soit 417 (on enlève celles qui ont subi 1 ou 2 interventions).
   Quel pourcentage cela représente-t-il ?
   $\frac{417}{439} \approx 0{,}9499 < 0{,}95$.

Donc l'intervalle contient moins de 95% des valeurs de la série (même si c'est minime).
La décision serait alors de changer les distributeurs.

3. En changeant l'effectif associé à la valeur 3 de 17 à 18, l'écart-type reste le même à 10^{-2} près, et on a $\overline{x} \approx 6{,}08$.
L'intervalle contient toujours toutes les valeurs comprises entre 3 et 10.
Cette fois-ci, il y en a 418, sur un total de 440.
Or : $\frac{418}{440} = 0{,}95 = 95\%$.
La décision aurait alors été de ne pas changer les distributeurs.

11 Chicago et Rome sont situées à la même latitude.
Voici les relevés des températures moyennes à la mi-journée des deux villes en 2017.

Mois	J	F	M	A	M	J	J	A	S	O	N	D
Rome en ^oC	11	16	18	20	23	28	34	36	25	23	16	12
Chicago en ^oC	0	4	5	14	15	24	28	26	22	14	4	-3

1. Calculer la moyenne et l'écart-type pour chacune des deux séries afin de comparer les températures de Rome et de Chicago en 2017.

$\overline{x}_R \approx 21{,}8^oC$ et $\sigma_R \approx 7{,}64^oC$.
$\overline{x}_C \approx 12{,}75^oC$ et $\sigma_R \approx 10{,}22^oC$.
Il a donc fait plus chaud en moyenne à Rome, et les températures étaient plus homogènes.

2. Comment expliquer ces différences pour deux villes situées à la même latitude ?

Les différences de température entre la côte est nord-américaine et l'Europe de l'ouest sont dues aux vents et aux conditions géographiques différentes.

C Synthèse

C.1 Exercices d'entraînement

12 On a résumé une série statistique dans le tableau ci-dessous.

Valeur	-5	0	1	2	6
Effectif	2	5	4	7	8

1. Prévoir, sans calcul, si la moyenne de cette série sera supérieure, inférieure ou égale à la médiane ? *Justifier*.
2. Vérifier par le calcul. *Arrondir à* 10^{-2}.
3. Déterminer les quartiles et l'écart interquartile.

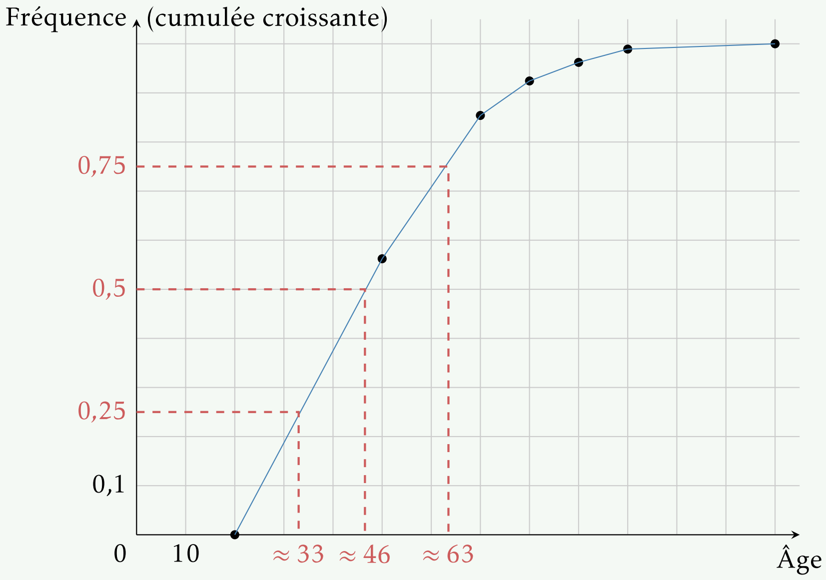
1. La moyenne devrait être supérieure à la médiane. La médiane semble se situer entre 0 et 2, tandis que les nombreuses valeurs extrêmes hautes laissent penser que la moyenne sera supérieure à 2.
2. On trouve $M = 2$ et $\overline{x} \approx 2{,}15$.
3. $Q_1 = 0$ et $Q_3 = 6$.
Ainsi : $EI = Q_3 - Q_1 = 6$.

13 Les résultats d'un contrôle de vitesse en agglomération son consignés dans le tableau ci-dessous.

Vitesse en km/h	Effectif	Centre
[20; 50]	104	35
]50; 70]	54	60
]70; 80]	13	75
]80; 90]	7	85
]90; 100]	5	95
]100; 130]	2	115

1. On suppose que dans chaque classe, les élément sont répartis de manière uniforme.
Estimer la vitesse moyenne enregistrée.
2. (a) En dressant la courbe des fréquences cumulées croissantes, estimer la médiane et les quartiles de cette série.
(b) Pour chaque résultat, faire un phrase d'interprétation.

1. $\overline{x} \approx 49{,}49$.
2. (a)



Ainsi :

- $M \approx 46$
- $Q_1 \approx 33$
- $Q_3 \approx 63$

(b)

- Au moins 25% des vitesses enregistrées sont inférieures ou égales à 33km.h^{-1}.
- Au moins 50% des vitesses enregistrées sont inférieures ou égales à 46km.h^{-1}.
- Au moins 75% des vitesses enregistrées sont inférieures ou égales à 63km.h^{-1}.
  ou
  Au moins **25%** des vitesses enregistrées sont **supérieures** ou égales à 63km.h^{-1}.