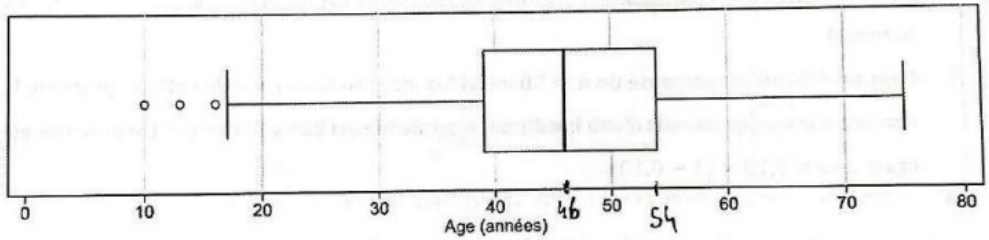
Statistiques descriptives – Exercices – Devoirs

QCM 1 corrigé disponible

Dans une étude, les auteurs décrivent la distribution d'âge de la population étudiée à l'aide du tableau et de la figure ci-dessous :

	N = 500
Age (années)	
moyenne (écart type)	46,0 (11,4)
médiane [Q1, Q3]	46 [39 ; 54]
Age en classes (n, %)	
< 46 ans	248 (49,6%)
46 à 65 ans	231 (46,2%)
> 65 ans	21 (4,2%)

N = effectif total ; Q1 = 1er quartile ; Q3 = 3ème quartile



Indiquer si les propositions suivantes sont vraies ou fausses.

A. 75% de la population d'étude a un âge supérieur à 54 ans

B. La moitié de la population d'étude a un âge entre 39 ans et 54 ans

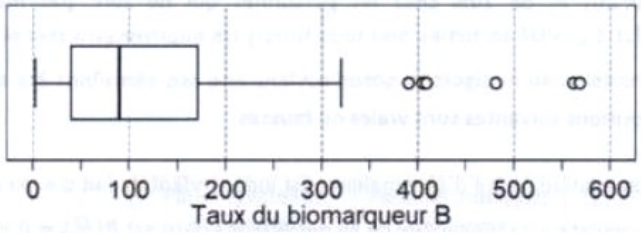
C. Le minimum de l'âge dans la population d'étude est égal à 17 ans

D. L'âge en classes correspond à une variable qualitative ordinale

E. On pourrait représenter la distribution de l'âge avec un histogramme

QCM 2 corrigé disponible

Dans une étude réalisée auprès de 100 patients, on décrit la distribution du taux d'un biomarqueur B à l'aide de la figure ci-dessous :



Indiquer si les propositions suivantes sont vraies ou fausses.

A. Le taux du biomarqueur B est une variable quantitative

B. Il s'agit d'une figure en box-plot ou « boîte à moustaches »

C. On pourrait utiliser le même type de graphique pour représenter la distribution d'une variable qualitative nominale

D. Le 3ème quartile de la distribution du taux du biomarqueur B est inférieur à 100

E. On peut considérer que le taux du biomarqueur B suit approximativement une loi normale

QCM 3 corrigé disponible

A propos des statistiques descriptives, indiquer si les propositions suivantes sont vraies ou fausses.

A. Dans une distribution parfaitement symétrique, la moyenne et la médiane prennent la même valeur

B. Dans le cas d'une variable qualitative nominale, il n'est pas possible de déterminer le mode de la distribution

C. L'étendue d'une distribution correspond à l'intervalle entre la plus petite et la plus grande valeur

D. La variance de X est homogène au carré de X

E. La variance de X est égale à la somme des carrés des écarts des valeurs de X à la valeur moyenne de X

QCM 4 corrigé disponible

On a recueilli l'âge de 200 femmes ayant accouché une semaine donnée dans une maternité privée. Les résultats sont rapportés dans le tableau suivant (par tranches d'âge d'un an, numérotées $i = \{1, \dots, 22\}$) :

i	Age x_i	Effectif n_i
1	15	6
2	16	7
3	17	12
4	18	15
5	19	15
6	20	19
7	21	15
8	22	14
9	23	13
10	24	13
11	25	15
12	26	10
13	27	4
14	28	9
15	29	7
16	30	7
17	31	5
18	32	4
19	33	5
20	34	1
21	35	2
22	36	2

Indiquer si les propositions suivantes sont vraies ou fausses.

A. L'âge maternel est une variable qualitative

B. Dans cet échantillon, le mode de l'âge maternel est égal à 19

C. Dans cet échantillon, la moyenne de l'âge maternel est égale à $\frac{\sum_{i=1}^{22} x_i n_i}{200}$

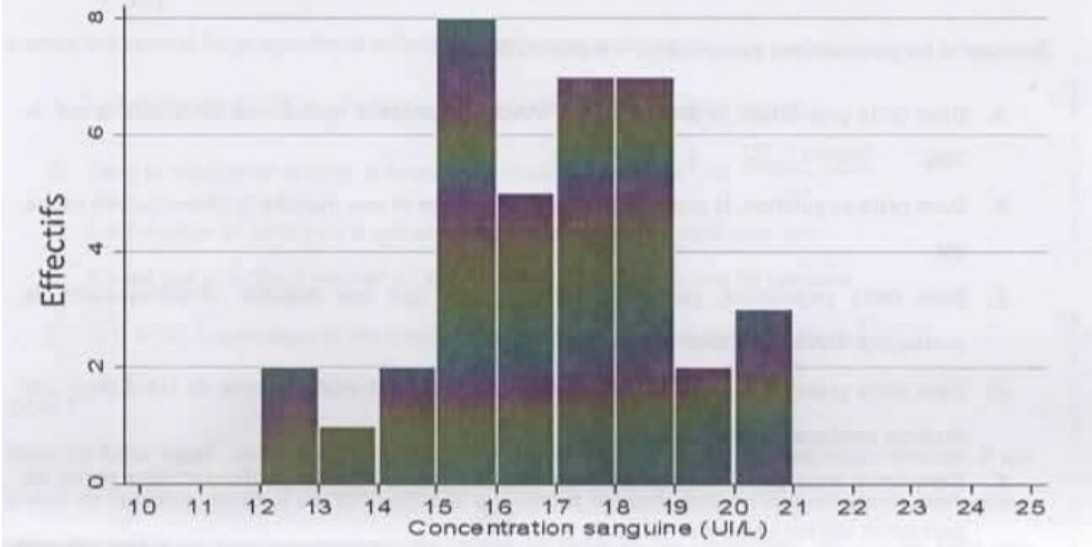
D. 10% des femmes de l'échantillon ont 33 ans ou plus

E. Dans cet échantillon, le premier quartile de l'âge maternel est égal à 19

QCM 5 corrigé disponible

On a dosé un paramètre biologique X chez 37 patients. Dans cet échantillon, la concentration sanguine de X était comprise entre 12,27 UI/L et 20,80 UI/L.

La distribution des résultats dans l'échantillon est représentée par le graphique suivant :



Indiquer si les propositions suivantes sont vraies ou fausses.

A. La variable X est une variable quantitative

B. Il s'agit d'un graphe à composantes superposées

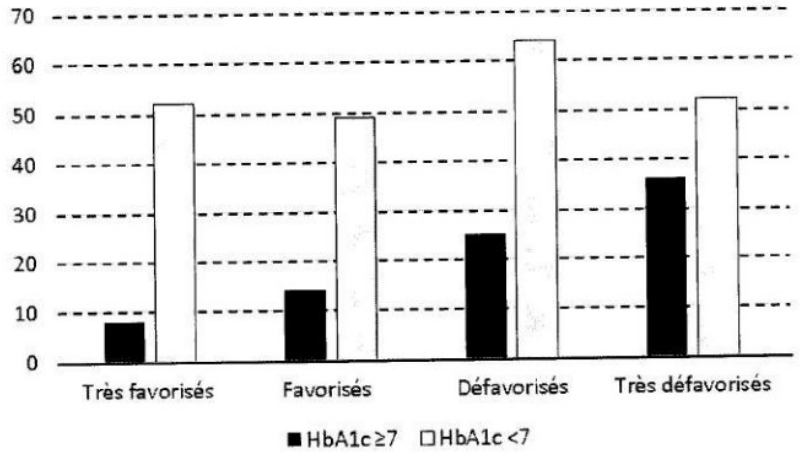
C. Dans cet échantillon, l'étendue est de 20,80 UI/L

D. Dans cet échantillon, la médiane est comprise entre 17 et 18 UI/L

E. Dans cet échantillon, la classe modale est l'intervalle [20 ; 21[

QCM 6 corrigé disponible

On a recueilli la position socio-économique et le résultat du dosage de l'hémoglobine glyquée (HbA1c) dans un échantillon de 300 patients diabétiques. Les effectifs de chaque catégorie sont décrits par la figure ci-dessous.

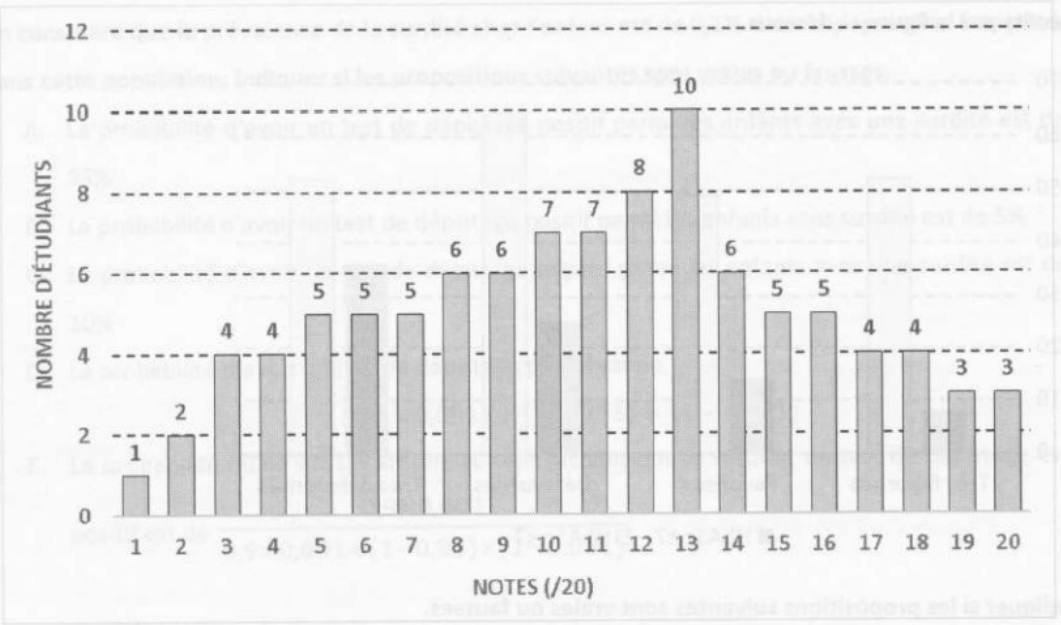


Indiquer si les propositions suivantes sont vraies ou fausses.

A. La figure est un histogramme
B. La variable « position socio-économique » telle que représentée par la figure est une variable quantitative
C. La variable « HbA1c » telle que représentée par la figure est une variable qualitative
D. Les pourcentages sont représentés en ordonnée
E. L'HbA1c médian dans l'échantillon de patients est égal à 7

QCM 7 corrigé disponible

Soit le graphique suivant, représentant la répartition des notes (de 1 à 20) de cent étudiants à une épreuve de statistique :

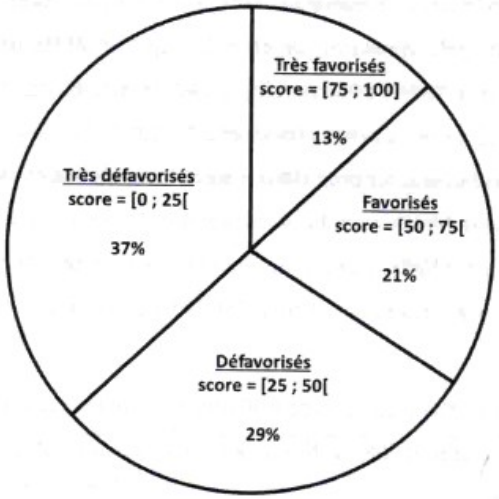


Indiquer si les propositions suivantes sont vraies ou fausses.

A. Il s'agit d'un boxplot (boîte à moustache)
B. La distribution de la variable « note » suit une loi normale
C. Le mode de la variable est 10
D. Une majorité d'étudiants a obtenu une note supérieure ou égale à 10
E. Le premier quartile est 5

QCM 8 corrigé disponible

On a recueilli la position socio-économique de 500 patients pris en charge pour une Broncho-Pneumopathie Chronique Obstructive. La position socio-économique a été mesurée par un score allant de 0 à 100, sachant que plus les individus sont favorisés, plus le score est élevé. Le score a ensuite été transformé en une variable à 4 classes, dont les fréquences respectives sont représentées par la figure ci-dessous.



Indiquer si les propositions suivantes sont vraies ou fausses.

A. La figure est un histogramme
B. La variable « position socio-économique » telle que représentée par la figure est une variable quantitative continue
C. La variable « position socio-économique » a été transformée en classes en fonction des quartiles du score dans l'échantillon
D. Le score médian est égal à 50
E. Au moins 300 personnes de l'échantillon sont considérées comme « défavorisées » ou « très défavorisées »

QCM 9 corrigé disponible

Soit un jeu de valeurs représentant les âges de onze patients suivant :

21 – 45 – 21 – 66 – 53 – 31 – 57 – 49 – 36 – 19 – 42

On précise que le total de ces valeurs est de 440.

Indiquer si les propositions suivantes sont vraies ou fausses.

A. La médiane est égale à 42
B. La médiane est supérieure à la moyenne
C. La moyenne est égale à 21
D. L'étendue de la distribution est égale à 66
E. Le mode est égal à 21

QCM 10 corrigé disponible

On a recueilli l'âge et le statut vis-à-vis de l'asthme dans un échantillon de 300 personnes. Les effectifs sont décrits dans le tableau ci-dessous, en fonction de l'âge et du statut asthmatique.

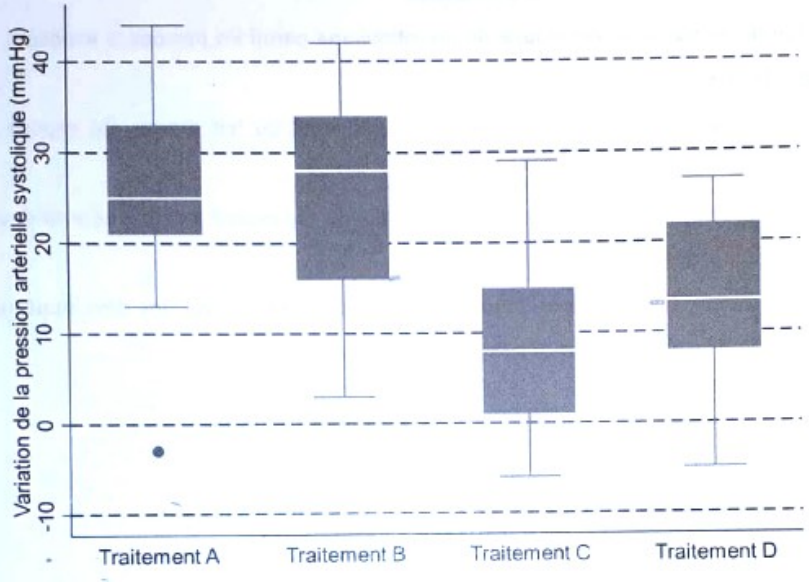
Indiquer si les propositions suivantes sont vraies ou fausses.

Age (en années)	Asthmatiques	Non Asthmatiques	Total
[0;5[	8	52	60
[5;10[	14	49	63
[10;15[	25	64	89
15 et plus	36	52	88
Total	83	217	300

A. La variable « âge », telle que rapportée dans le tableau, est une variable qualitative ordinale
B. Le mode de la variable « âge » du groupe « asthmatiques » est identique au mode de la variable « âge » du groupe « non asthmatiques »
C. L'âge médian, dans l'échantillon total, est supérieur ou égal à 10
D. Il est possible de calculer l'étendue de l'âge dans chaque groupe à partir des données fournies par le tableau
E. Il est possible de calculer l'âge moyen dans chaque groupe à partir des données fournies par le tableau

QCM 11 corrigé disponible

Quatre traitements A, B, C et D sont suspectés de présenter un effet secondaire indésirable en induisant une augmentation de la pression artérielle systolique (PAS). On mesure donc la variation de la PAS d'un groupe de patients suite à l'administration de chacun des traitements A, B, C et D. Les résultats obtenus sont résumés dans la représentation graphique suivante.



Indiquer si les propositions suivantes sont vraies ou fausses.

A. Les médianes de la variation de la PAS avec le traitement A et B sont supérieures à 20mmHg
B. Le premier quartile de la variation de la PAS avec le traitement B est inférieur à la médiane de la variation de la PAS avec le traitement D
C. L'intervalle interquartile de la variation de la PAS avec le traitement B est inférieur à l'intervalle interquartile de la variation de la PAS avec le traitement A
D. La borne supérieure de l'intervalle interquartile de la variation de la PAS avec le traitement C est inférieure à la borne inférieure de l'intervalle interquartile de la variation de la PAS avec le traitement A
E. L'intervalle interquartile de la variation de la PAS avec le traitement D est inférieur à 10 mmHg

QCM 12 corrigé disponible

A propos des différents types de variables, indiquer si les propositions suivantes sont vraies ou fausses.

A. Le nombre de femmes ayant accouché à la maternité Paule de Viguier en 2011 est une variable quantitative discrète
B. La kaliémie (dosage du potassium sanguin en mmol/l) est une variable quantitative continue
C. L'origine géographique de naissance (Europe, Asie, Afrique, Amérique, Océanie) est une variable qualitative non ordonnée
D. L'origine géographique de naissance (Europe, Asie, Afrique, Amérique, Océanie) est une variable qualitative binaire
E. L'origine géographique de naissance (Europe, Asie, Afrique, Amérique, Océanie) peut être discrétisée

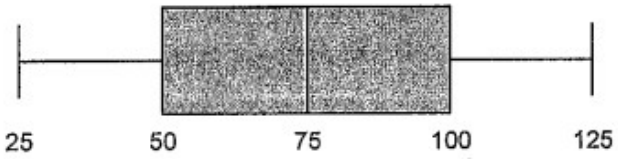
QCM 13 corrigé disponible

On mesure le temps hebdomadaire passé à faire du sport (en heures) dans un échantillon de 10 étudiants en 1ère année de médecine. Les valeurs de temps passé à faire du sport (en heures), ordonnées par ordre croissant, sont les suivantes : 0 ; 0 ; 0 ; 0,5 ; 1 ; 1 ; 1 ; 2 ; 2 ; 2. Indiquer si les propositions suivantes sont vraies ou fausses

A. La valeur du dixième percentile est égale à 0 heure
B. La valeur du 90ème percentile est égale à 2 heures
C. L'intervalle interquartile est de [0 heure - 2heures]
D. La médiane est égale à 1 heure
E. La valeur du 1er quartile est égale à 1 heure

QCM 14 corrigé disponible

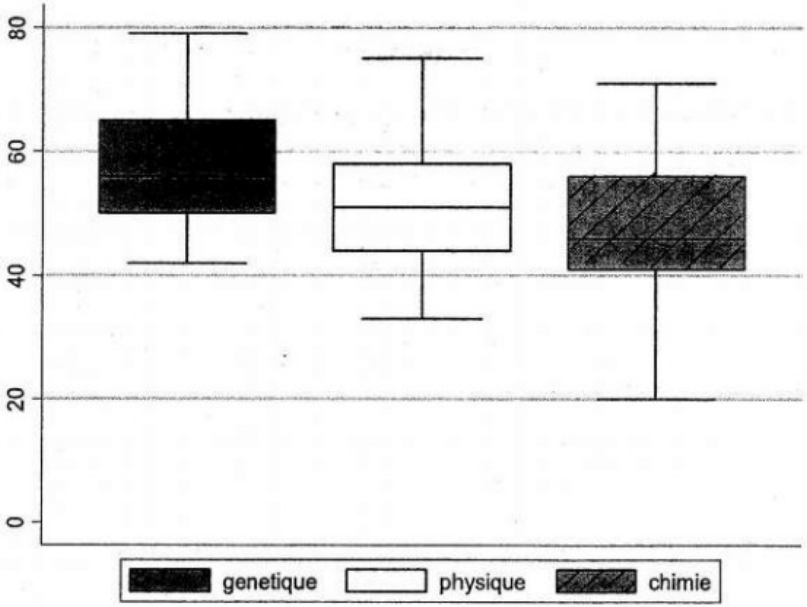
On donne la figure ci-dessous, indiquer si les propositions suivantes sont vraies ou fausses.



A. La figure ci-dessus est un histogramme
B. Le 1ème quartile vaut 50
C. Le 3ème quartile vaut 75
D. Le 1er quartile ne vaut ni 25, ni 75, ni 125
E. L'intervalle interquartile est égal à [25-75]

QCM 15 corrigé disponible

On relève les notes (sur 80) de 200 élèves dans trois matières (génétique, physique, chimie). On réalise la représentation graphique ci-dessous.



Indiquer si les propositions suivantes sont vraies ou fausses.

A. Il s'agit d'un diagramme en bâtons
B. La note médiane de génétique est supérieure à la note médiane de chimie
C. Les bornes de l'intervalle interquartile de physique sont toutes les deux supérieures au premier quartile de génétique
D. Le premier quartile des notes de physique est inférieur à 40
E. Le troisième quartile des notes de physique est supérieur à la médiane des notes de chimie

QCM 16 corrigé disponible

Indiquer si les propositions suivantes sont vraies ou fausses.

A. Le polygone cumulé des fréquences tend vers la densité de la loi de probabilité quand la taille de l'échantillon augmente
B. L'intervalle inter-quartile est une mesure de la dispersion des valeurs
C. La moyenne est un bon indicateur de tendance centrale pour une distribution bimodale
D. Pour construire un diagramme en secteurs (camembert), on divise la surface d'un disque en secteurs dont la surface est proportionnelle aux fréquences respectives des modalités de la variable étudiée
E. Le coefficient de variation est un paramètre de dispersion donné par le rapport de la moyenne sur l'écart-type

QCM 17 corrigé disponible

Un échantillon de patients a été randomisé en deux groupes. Les caractéristiques de ces groupes sont décrites dans le tableau suivant :

	Groupe A (N=30)		Groupe B (N=26)	
Sexe (effectif, %)				
Femme	11	36,7	10	38,5
Homme	19	63,3	16	61,5
Age en années (moyenne, écart-type)	59,0	13,4	64,3	9,1
Score de sévérité (effectif, %)				
0 à 3	15	50,0	12	46,1
4 à 6	3	10,0	4	15,4
7 à 10	12	40,0	10	38,5

A partir des données décrites dans ce tableau, indiquer si les propositions suivantes sont vraies ou fausses.

A. La variable « score de sévérité » est traitée comme une variable quantitative
B. La dispersion des âges est plus grande dans le groupe A que dans le groupe B dans cet échantillon
C. Dans cet échantillon, les patients du groupe B sont en moyenne plus âgés que les patients du groupe A
D. La valeur « 13,4 » pour la variable âge est un indicateur de tendance centrale
E. Le score de sévérité médian du groupe B est compris entre 4 et 6