

Contrôle 1 de Bio-statistique et Informatique

**PARTIE BIOSTATISTIQUE**

I. Les quartiles de la variable (d'âge) dans un échantillon de  $N$  participants à un essai clinique étaient respectivement de 27, 41 et 59. La moyenne d'âge des participants de moins de 27 ans est de 14 ans avec un écart typ de 6 ans alors que la moyenne d'âge des plus de 27 ans est de 50 ans avec un écart type de 4,5.

01) Indiquer la bonne réponse dans ce qui suit :

- a) la moitié des individus avait moins de 27 ans
- b) la moitié des individus avait 41 ans
- c) la moitié des individus avaient entre 27 et 59 ans
- d) les trois quarts des individus avait plus de 59 ans
- e) aucun individu n'a un âge compris entre 0 et 27 ans

02. La moyenne d'âge dans l'échantillon est égale à :

- a) 32ans
- b) 27 ans
- c) 20 ans
- d) 35ans
- e) 41 ans

03) Laquelle des affirmations suivantes est juste :

- a) La tranche d'âge des moins de 27 ans est moins dispersée que celle des plus de 27 ans
- b) la tranche d'âge des moins de 27 ans est assez homogène
- c) la tranche d'âge des plus de 27 ans est assez homogène
- d) la tranche d'âge des plus de 27 ans est hétérogène
- e) l'écart interquartile est égal à 30

II. Une étude sur une variable statistique  $X$  continue a donné les résultats suivants :

$X_i$	$[0-a[$	$[a-b[$	$[b-9[$
$n_i$	$x$	40	$y$

La taille de l'échantillon est égale à 120.

04. Si les classes sont de mêmes amplitudes et le premier quartile  $Q_1=2$  alors la valeur de  $x$  vaut :

- a) 80
- b) 40
- c) 45
- d) 20
- e) 30

05) Si les classes sont de mêmes amplitudes et le premier quartile  $Q_1=2$ , le nombre d'individus ayant une valeur comprise entre  $a$  et 7 est approximativement égal à :

- a) 100
- b) 51
- c) 115
- d) 120
- e) 105

06. Si cette fois ci les classes avaient les mêmes effectifs, le troisième décile  $D_3=3,6$  et le 60<sup>ème</sup> centile  $C_{60}=7$ , les valeurs de  $a$  et  $b$  valent :

- a)  $a=3,5$  et  $b=8,5$
- b)  $a=5$  et  $b=8$
- c)  $a=4$  et  $b=5$
- d)  $a=3$  et  $b=6$
- e)  $a=4$  et  $b=7,75$

II. Dans une ville de 100000 habitants, une épidémie s'est déclenchée. On enregistre chaque jour le nombre de cas qui se sont déclarés. Au septième jour le tableau suivant a été obtenu :

$X_i$	1	2	3	4	5	6	7
$Y_i$	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$	$Y_7$

$X_i$  représente le jour et  $Y_i$  représente le nombre de cas enregistrés. Après calculs, nous avons obtenu les résultats suivants :  $\bar{Y} = 625$ , si nous considérons la variable  $Z = \ln Y$  alors  $\bar{Z} = 4,7$  et  $\sum_{i=1}^7 X_i Z_i = 162,166$

07. La moyenne marginale et la variance de  $X$  valent :

- a)  $\bar{X}_M = 4$  et  $Var(X) = 4$
- b)  $\bar{X}_M = 28$  et  $Var(X) = 16$
- c)  $\bar{X}_M = 7$  et  $Var(X) = 4$
- d)  $\bar{X}_M = 4$  et  $Var(X) = 16$
- e)  $\bar{X}_M = 28$  et  $Var(X) = 140$

08. La fonction exponentielle qui s'ajuste le mieux à cette série est :  
 a)  $Y=0,33(1,09)^X$     b)  $Y=1,4(3)^X$     c)  $Y=4(1,3)^X$     d)  $Y=3(1,4)^X$     e)  $Y=1,09(0,33)^X$
09. La fonction exponentielle ajustée sera utile pour :  
 a) donner la droite d'ajustement de Y en fonction de X    c) tracer le nuage de points    e) dresser le tableau  
 b) calculer la covariance entre X et Y    d) faire des prévisions
10. Si aucune mesure n'est prise pour stopper l'épidémie, le nombre de jours nécessaire pour que tous les habitants de la ville soient atteints est :  
 a) 8    b) 9    c) 10    d) 11    e) 15

**PARTIE INFORMATIQUE**

I.

11. Vous écrivez un texte donné à l'intérieur de la cellule « A1 ». Pour commencer une nouvelle ligne de texte dans la même cellule (insérer un saut de ligne), vous appuyez sur:  
 a) « Shift+Entrée ».    c) « Alt+Entrée ».    e) « Alt+Shift+Entrée ».  
 b) « Ctrl+Entrée ».    d) « Entrée ».
12. La cellule « E5 » contient la formule suivante « =SC1 ». Le glissement (en utilisant la *poignée de copie*) de la cellule « E5 » vers la gauche implique :  
 a) Le changement de la ligne et de la colonne dans C1.    c) Le changement de la ligne dans C1.  
 b) Le changement de la colonne dans C1.    d) Aucun changement.  
 e) La disparition de symbole dollar « \$ ».

13. Les plages «A1:A10» et «B1:B10» contiennent des nombres aléatoires. Parmi les fonctions suivantes, une fonction n'est pas correctement écrite, laquelle ? :  
 a) =B3+2\*A1  
 b) =MOYENNE(A1:A10) - 2\*B5  
 c) =MOYENNE(A1:A10) - MOYENNE(B1:B10)/2  
 d) =A1=B3  
 e) =MOYENNE(A1:A10) - =MOYENNE(B1:B10) +A1

II. De nos jours, les données sont l'un des contributeurs les plus importants à la résolution de problèmes complexes. Dans le domaine de la santé, de nombreuses données sont collectées pour être utilisées dans le processus de prévention et de diagnostic. Par conséquent et dans ce contexte, les données peuvent jouer un rôle important dans la lutte contre la pandémie de COVID-19.  
 Dans ce but, nous avons collecté quelques données utiles liées au Covid-19, notamment : le sexe, l'âge, la saturation en oxygène (SO%), la toux et la fatigue.

	A	B	C	D	E	F
1	ID	Sexe	Age	SO (%)	Toux	Fatigue
2	1	f	29	80,00	1	Oui
3	2	m	25	99,00	0	Non
4	3	m	71	73,00	0	Oui
5	4		39	97,00	0	Non
6	5		25		0	Non
7	6	m	80	60,00	1	Oui
8	7	f	22	99,00	0	Non
9	8	f	21	97,00	1	Non
10	9	m	46	70,00	0	Oui
11	10	m	50	95,00	0	Non
12	11	f	47	97,00	0	Non
13	12		29	97,00	1	Non
14	13	m	90	60,00	0	Oui
15	14	m	65		1	Oui
16	15	f	27	96,00	0	Non
17	16	m	27	88,12	1	Oui
18	17	f	76	72,00	1	Oui
19	18	m	34	98,00	0	Non
20	19	m	28	97,00	0	Non
21	20	m	36	81,00	0	Oui

14. Comme vous pouvez le voir, chaque colonne contient des valeurs qui varient selon le type de variable. Cependant, nous avons quelques cellules vides qui peuvent poser des problèmes dans le processus d'analyse et d'exploitation des données. Alors, parmi les fonctions suivantes, quelle fonction nous devrions utiliser pour calculer le nombre de cellules vides ? :  
 a) NB.SI    c) NB.VIDE    e) NBVIDE  
 b) NB    d) NBVAL

15. Pour restreindre la saisie et éviter de tels problèmes à l'avenir, nous vous suggérons d'effectuer un processus de validation avant d'écrire les données. Par exemple, pour assurer que seules les valeurs valides sont entrées dans la plage de cellules D2:D21:

- Données → Outils de données → Validation des données → Autoriser : Nombre entier → comprise entre → taper entre 0 et 100
- Données → Outils de données → Validation des données → Autoriser : Décimal → liste → taper entre 0 et 100
- Données → Outils de données → Validation des données → Autoriser : Décimal → supérieure à → taper 0
- Données → Outils de données → Validation des données → Autoriser : Nombre entier → inférieure à → taper 100
- Données → Outils de données → Validation des données → Autoriser : Décimal → comprise entre → taper entre 0 et 100

16. Afin de traiter les valeurs manquantes ("D6" et "D15") dans la plage de cellules D2:D21, nous devons remplacer ces valeurs vides par:

- =VAR(D2:D21)
- =MAX(D2:D21)
- =MOYENNE(D2:D21)
- =RACINE(D2:D21)
- =ECARTYPE(D2:D21)

17. Pour calculer le nombre de patients souffrant de fatigue et de saturation en oxygène inférieure ou égale à 80,00%, il faut utiliser la formule suivante:

- =NB.SI.ENS(F2:F21;"Oui";D2:D21;<=80)
- =NB.SI.ENS(F2:F21;"Oui";D2:D21;"<=80")
- =NB.SI.ENS(F2:F21;"Oui";C2:C21;"<=80")
- =NB.SI.ENS(F2:F21;"Oui";D2:D21;"<=80")
- =NB.SI.ENS(B2:B21;"Oui";D2:D21;">=80")

18. Le plus âgé patient (e) est âgé (e) de:

- =MAX(C2:C21)
- =MIN(C1:C20)
- =MAX(A2:A21)
- =QUARTILE(A2:A21; 1)
- =MAX(C2:C20)

III.

19. D'après la courbe présentée à droite, il est clair qu'il existe une relation linéaire entre l'âge du patient et le pourcentage de saturation en oxygène. Pour s'assurer que cette relation existe, on calcule le coefficient de corrélation en utilisant la formule suivante :

- =COEFFICIENT.CORRELATION(C2:C21;D2:D21)
- =COEFFICIENT.CORR(C2:C21;D2:D21)
- =CORRELATION(C2:C21;D2:D21)
- =COEFFICIENTCORRELATION(C2:C21;D2:D21)
- =COEFFICIENTC(C2:C21;D2:D21)

20. Le calcul du coefficient de corrélation a révélé une forte relation linéaire négative égale à (-0,82) entre l'âge du patient et le pourcentage d'oxygène. Cela nous a beaucoup encouragé à calculer les paramètres « a » et « b » à l'aide de la formule suivante (Note : En supposant que les valeurs vides sont traitées):

- =DROITEREG(C2:C21;D2:D21;VRAI;VRAI)
- =DROITEREG(D2:D21;C2:C21;VRAI;VRAI)
- =COVARIANCE(C2:C21;D2:D21)/VAR.P(C2:C21)
- =DROITER(C2:C21;D2:D21;VRAI;VRAI)
- =MOYENNE(D2:D21)-(16\*MOYENNE(C2:C21))



