IUT Lannion
BUT Informatique - 1e année

Année 2021-2022

Statistiques : Devoir Surveillé
20 mai 2022

Nom Prénom : Corrigé **Groupe :**

L'épreuve dure **1h15**.
Les calculatrices sont autorisées.
Bon travail !

Exercice 1. *(7.5 points)*

Des éleveurs de dinosaures présentent leurs plus beaux sujets à un concours, au cours duquel les animaux reçoivent une note sur 20, qui sera notre variable statistique X. Le tableau ci-dessous est le tableau statistique de X, il présente les effectifs des dinosaures ayant eu chaque note :

Note	0	9	10	11	12	13	14	15	16
Effectifs	14	50	55	70	74	71	60	10	4
Effectifs cumulés	14	64	119	189	263	334	394	404	408

Dans cet exercice, bien poser tous les calculs que vous faites (ne donnez pas seulement le résultat) : ce sera la moitié des points de chaque question.

1. Compléter la dernière ligne du tableau statistique.

2. Quel est l'effectif total de X ?

408

3. Calculer la moyenne (avec 2 chiffres après la virgule) des notes obtenues par les dinosaures.

$$\overline{X} = \frac{1}{408}\left(14\times 0 + 50\times 9 + 55\times 10 + 70\times 11 + 74\times 12 + 71\times 13 + 60\times 14 + 10\times 15 + 4\times 16\right)$$

$$= \frac{1}{408} \times 4635 \simeq 11.36$$

4. Quel est le mode de cette série ? Justifier votre réponse.

12 car l'effectif (74) de 12 est le plus élevé.

5. Quelle est la médiane de cette série ? Justifier votre réponse.

12 car si on classe les dinosaures par note, le 204e et le 205e ont 12 vu que l'effectif cumulé de 11 est 189, et celui de 12 est 263.

6. Donner le 1e et le 3e quartile de la série.

Le 1e quartile est 10 , et le 3e quartile est 14

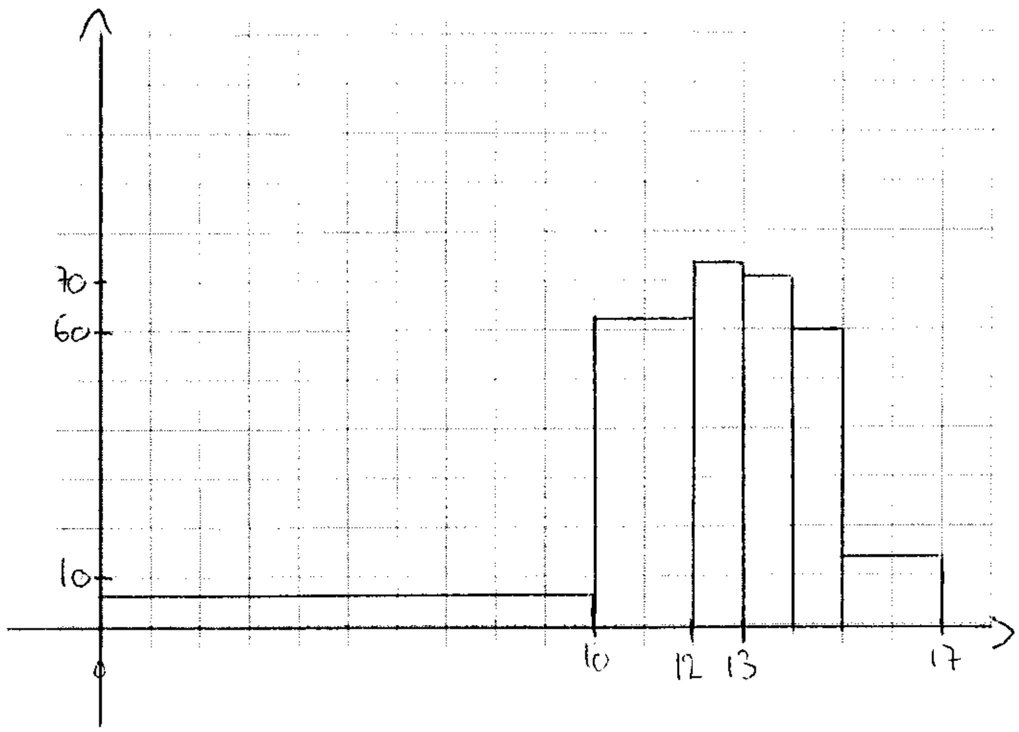
On choisit maintenant de répartir les dinosaures en groupes selon les modalités suivantes

$$[0, 10[, [10, 12[, [12, 13[, [13, 14[, [14, 15[, [15, 17[$$

7. Compléter le tableau statistique suivant, associé à ces modalités :

Note	[0,10[	[10,12[	[12,13[	[13,14[	[14,15[	[15,17[
Effectifs	64	125	74	71	60	14
Effectifs corrigés	6,4	62,5	74	71	60	7

8. Tracer ci-dessous l'histogramme associé à ces modalités :



9. Quel est le mode de X pour ces modalités ?

Le mode est [12,13[, car son effectif corrigé est le + élevé.

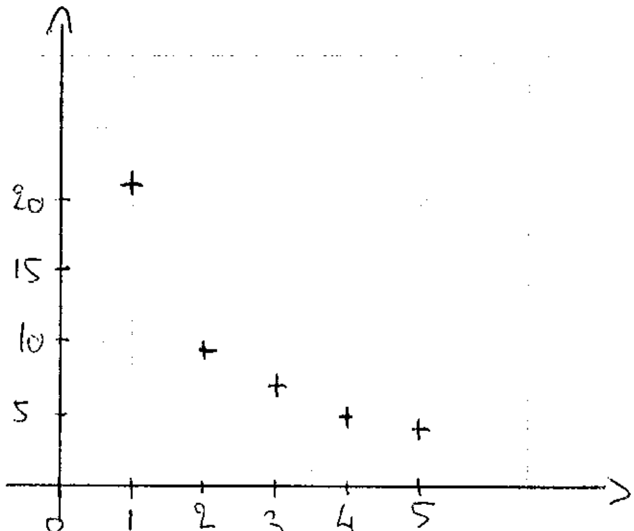
Nom Prénom : **Groupe :**

Exercice 2. *(8.5 points)*

On s'intéresse aux variables statistiques X et Y suivantes, on se demande s'il existe un lien de colérration entre elles :

Période (X)	1	2	3	4	5
Observations (Y)	21	9.5	7	4.9	4

1. Tracer ci-dessous le nuage de points correspondant, avec X en abscisse et Y en ordonnée.

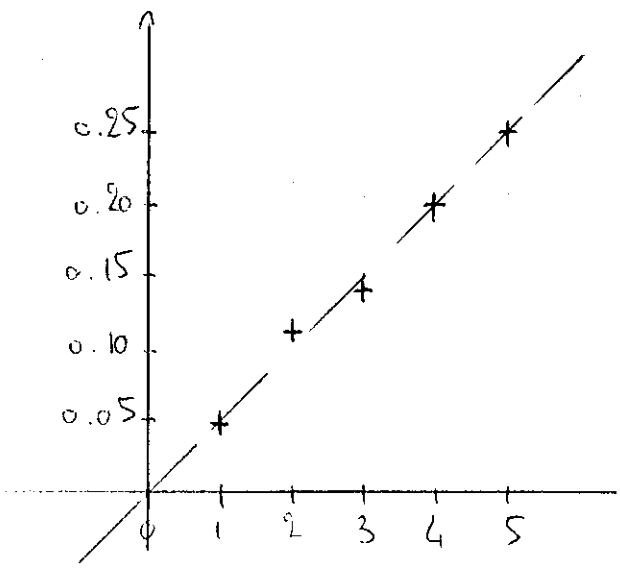


En voyant ce nuage de points, on peut soupçonner qu'il existe une relation de corrélation entre X et Y, mais pas sous la forme d'une droite.
On introduit la nouvelle variable statistique $Z = \frac{1}{Y}$ afin d'observer s'il y a une corrélation linéaire entre X et Z.

2. Compléter le tableau ci-dessous avec les valeurs de Z pour les différentes périodes :

Période (X)	1	2	3	4	5
Observations (Y)	21	9.5	7	4.9	4
Z $(= 1/Y)$	0.048	0.105	0.143	0.204	0.25

3. Tracer ci-dessous le nouveau nuage de points correspondant, avec X en abscisse et Z en ordonnée.



4. Calculer la covariance $Cov(X, Z)$ entre X et Z.

Pour cette question et les suivantes, bien poser tous les calculs que vous faites (ne donnez pas seulement le résultat) : ce sera la moitié des points de chaque question.

$$\overline{X} = \frac{1}{5}(1+2+3+4+5) = 3$$

$$\overline{Z} = \frac{1}{5}(0.048+0.105+0.143+0.204+0.250) = 0.15$$

$$\overline{XZ} = \frac{1}{5}(1\times 0.048+2\times 0.105+3\times 0.143+4\times 0.204+5\times 0.250) \simeq 0.551$$

$$Cov(X,Z) = \overline{XZ} - \overline{X}\,\overline{Z} \simeq 0.551 - 3\times 0.15 \simeq 0.101$$

5. Calculer le coefficient de corrélation $Cor(X, Z)$ entre X et Z.

$$Var(X) = \overline{X^2} - \overline{X}^2 = \frac{1}{5}(1^2+2^2+3^2+4^2+5^2) - 3^2 = 11 - 9 = 2$$

$$Var(Z) = \overline{Z^2} - \overline{Z}^2 = \frac{1}{5}(0.048^2+0.105^2+0.143^2+0.204^2+0.25^2) - 0.15^2$$

$$\simeq 0.028 - 0.0225 \simeq 0{,}005$$

$$Cor(X,Z) = \frac{Cov(X,Z)}{\sigma(X)\,\sigma(Z)} \simeq \frac{0.101}{\sqrt{2}\times\sqrt{0.005}} \simeq 1{,}002$$

(Les arrondis donnent un résultat > 1, ce qui est normalement impossible)

6. Calculer les coefficients a et b de la droite d'ajustement linéaire de Z sur X, d'équation $z = ax + b$.

$$a = \frac{Cov(X,Z)}{Var(X)} \simeq \frac{0.101}{2} \simeq 0{,}05$$

$$b = \overline{Z} - a\overline{X} \simeq 0.15 - 0.05\times 3 \simeq 0$$

7. Tracer cette droite d'ajustement linéaire sur le graphique représentant Z et X.

8. A l'aide de l'équation de la droite trouvée ci-dessus, calculer une prévision de l'observation Y pour la période 6.

Pour la période 6, on aurait environ $z = 0.05\times 6 + 0 = 0.3$ et donc $y = \frac{1}{z} = 3.33$.

Nom Prénom : **Groupe :**

Exercice 3. *Un peu de Python (4 points)*

Dans cet exercice vous devez écrire du code Python sur papier : vous ne serez pas pénalisé.e par d'éventuelles petites fautes de syntaxe.

On s'intéresse à 2 séries statistiques X et Y codées en Python, comme par exemple les suivantes (dont on ne se servira pas) :

```
X=np.array([20,5,5,40,30,35,5,5,15,40])
Y=np.array([5,1,2,7,8,9,3,2,5,8])
```

On crée la fonction `covar` suivante, qui calcule la covariance de 2 séries statistiques `X` et `Y` de même longueur :

```
def covar(X,Y):
    N=len(X)        #on présuppose que X et Y sont de même longueur
    Xm=sum(X)/N     #Xm contient la moyenne de X
    Ym=sum(Y)/N     #Ym contient la moyenne de Y
    return sum((X-Xm)*(Y-Ym))/N
```

1. La fonction `covar` ci-dessus calcule la covariance à partir de la formule de la définition de la covariance. Ecrire une fonction `covarK` qui calcule la covariance à partir de la formule de Koenig sur la covariance.

```
def covarK(X,Y):
    N = len(X)
    Xm = sum(X)/N
    Ym = sum(Y)/N
    XYm = sum(X*Y)/N
    return XYm - Xm * Ym
```

On a fait en TP une fonction `Correl(X,Y)` et une fonction `CoeffDroite(X,Y)` qui prennent en paramètres 2 tableaux X et Y et renvoient respectivement le coefficient de corrélation $Cor(X,Y)$ et les coefficients a et b de la droite d'ajustement linéaire de Y en X :

Pour 2 variables statistiques X et Y, on souhaite tester si la corrélation sous forme de droite est la corrélation la plus forte entre X et Y, ou bien si la corrélation entre X et l'une des fonctions de Y suivantes est plus forte :

- $\frac{1}{Y}$ (comme dans l'exercice 2)
- le logarithme népérien de Y (`np.log(Y)` en Python)
- l'exponentielle de Y (`np.exp(Y)` en Python).

2. En utilisant les fonction `Correl` et `CoeffDroite`, créer une fonction `FormeCorrel(X,Y)`, qui prend en paramètres 2 tableaux X et Y (comme dans la question 1) et renvoie un triplet `(Forme,a,b)`, où

   - `Forme` est un entier qui vaut 0 si la corrélation la plus forte est avec Y, 1 si la corrélation la plus forte est avec $\frac{1}{Y}$, 2 si c'est avec le logarithme de Y, et 3 si la corrélation la plus forte est avec l'exponentielle de Y.
   - `a,b` sont les coefficients de la droite d'ajustement linéaire entre X et la forme de Y dont la corrélation est la plus forte.

```
def FormeCorrel(X,Y):
    Forme=0
    Cmax=Correl(X,Y)
    (a,b)=CoeffDroite(X,Y)
    c1=Correl(X,1/Y)
    if c1>Cmax:
        Cmax=c1
        Forme=1
        (a,b)=CoeffDroite(X,1/Y)
    c2=Correl(X,np.log(Y))
    if c2>Cmax:
        Cmax=c2
        Forme=2
        (a,b)=CoeffDroite(X,np.log(Y))
    c3=Correl(X,np.exp(Y))
    if c3>Cmax:
        Cmax=c3
        Forme=3
        (a,b)=CoeffDroite(X,np.exp(Y))
    return (Forme,a,b)
```