Numéro d'identifiant à reporter :…………………………………..…………………………

Université Lyon I **Année 2002-2003**

Licence Biologie des Organismes
Contrôle Terminal de BMS
Statistique – 2 heures
1ère session

Tous documents autorisés – Calculatrices autorisées

Question 1

1. Le fichier `toto.txt` se présente comme suit dans le bloc-notes:

```
note.continu note.terminal note.final
          10            15       13,0
           8             8        8,0
          19            14       16,0
          14             8       10,4
```

et comme suit dans R :

```
  note.continu note.terminal note.final
1           10            15       13.0
2            8             8        8.0
3           19            14       16.0
4           14             8       10.4
```

Quelle est l'instruction ayant permis ce résultat ?

`read.table("toto.txt",h=T)` ☐
`read.table("toto.txt")` ☐
`read.table("toto.txt", h=T, dec=",")` ☐

2. On donne l'instruction `rep(c(1,2),c(3,4))`. Le vecteur obtenu est :

`1 1 1 2 2 2 2` ☐
`3 4 4` ☐

3. Cocher l'expression ayant permis d'obtenir la matrice ci-dessous :

```
     [,1] [,2]
[1,]    7    8
[2,]    9   10
[3,]   11   12
```

```
> vecteur <- c(7,8,9,10,11,12)
```

`> matrix(vecteur,nrow=3, byrow=F)` ☐
`> matrix(vecteur,nrow=3, byrow=T)` ☐

4. On construit le vecteur des notes de 11 étudiants `vecnot`. L'expression `sd(vecnot)` est équivalente à une des expressions ci-dessous. Laquelle ?

```
> sqrt(var(vecnot)*11/10)
> sqrt(var(vecnot))
> sqrt(var(vecnot)*10/11)
```

☐ ☐ ☐

5. Un objet data.frame peut-il contenir plusieurs modes possibles, comme par exemple numérique, caractère, complexe… ?

`OUI, plusieurs modes sont possibles` ☐

`NON, un seul mode est possible` ☐

Question 2

2a. Le nombre d'enfants de moins de trois ans dans une famille est modélisé par une variable aléatoire dont la loi de probabilité est la suivante :

Nombre d'enfants	0	1	2	3
Probabilité	p	0.5	0.3	p

Donner, en expliquant la démarche, la probabilité manquante p :………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

2b. Soit une loi binomiale X de paramètres $n = 15$ et $p = 0.32$. On donne les informations suivantes :

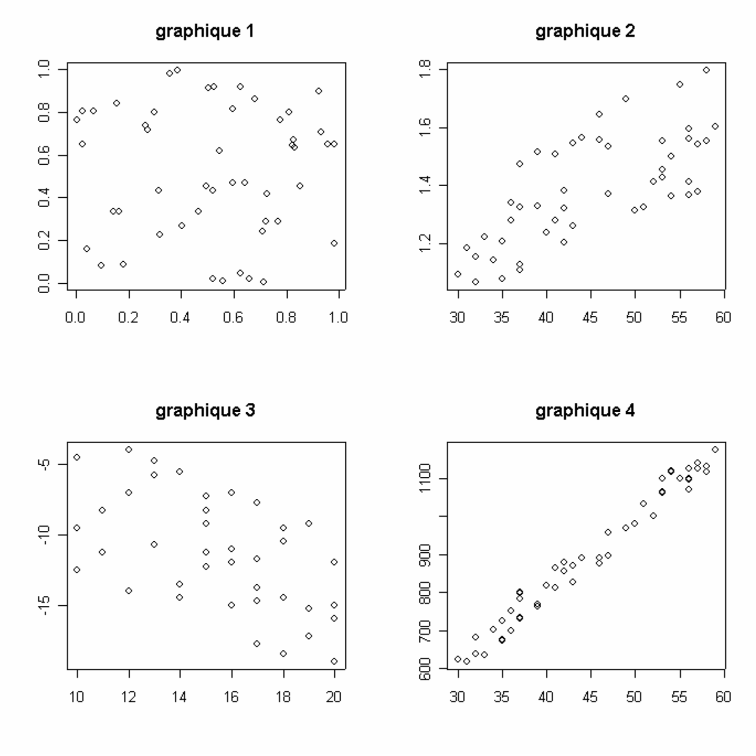
```
> dbinom(0:15,15,0.32)
 [1] 3.074e-03 2.170e-02 7.147e-02 1.457e-01 2.057e-01 2.130e-01 1.671e-01
 [8] 1.011e-01 4.757e-02 1.741e-02 4.916e-03 1.052e-03 1.649e-04 1.791e-05
[15] 1.204e-06 3.778e-08
```

Donner les probabilités $P(X = 6)$ et $P(X \geq 1)$:………………….……………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

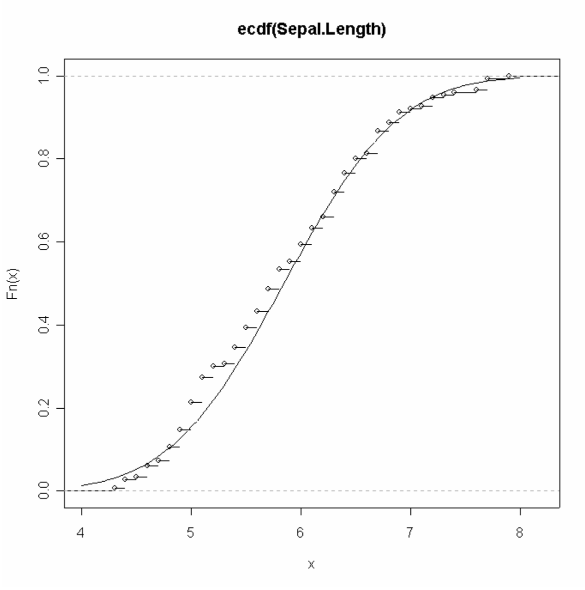
2c. On donne les quatre nuages de points ci-dessous ainsi que les quatre coefficients de corrélations. Cocher le numéro du graphique correspondant au coefficient de corrélation.

	graphiques			
Corrélation	1	2	3	4
-0.5914				
-0.0132				
0.7272				
0.9868				

Numéro d'identifiant à reporter :……………………………………..…………………………



2d. On étudie la longueur du sépale de 150 iris. Donner un titre et une légende à cette représentation graphique.



Titre du graphique :………………………………………………………………………………
………………………………………………………………………………………………
Légende associée au graphe continu …..………………………………………………………
………………………………………………………………………………………………
Légende associée au graphe discontinu…..……………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Question 3

On a les copies d'examen de 108 étudiants de la filière biologie. On s'intéresse à la variable *X* : nombre de fautes d'orthographe dans une copie. A l'aide des informations ci-dessous, on veut vérifier si la variable *X* suit une loi de Poisson de paramètre la moyenne observée.

On donne les valeurs observées de la distribution ainsi que la moyenne.
`> fautes=rep(c(0,1,2,3,4,5),c(18,35,28,16,7,4))` où « `5` » rassemble les copies ayant plus de 5 fautes d'orthographes.

```
> moyenne <- mean(fautes)
[1] 1.731
```

On donne les probabilités et les valeurs théoriques.

```
> probtheo
[1] 0.17702 0.30651 0.26536 0.15315 0.06630 x.xxxxx
> valtheo
[1] 19.118 33.103 28.659 16.541  7.160  3.419
> obs
[1] 18 35 28 16  7  4

> chideux <- sum((obs-valtheo)^2/valtheo)
> chideux
[1] 0.3092
```

```
> 1-pchisq(chideux,6)
[1] 0.9995

> 1-pchisq(chideux,5)
[1] 0.9975

> 1-pchisq(chideux,4)
[1] 0.9892

> 1-pchisq(chideux,3)
[1] 0.9583
```

Hypothèse :……………………………………………………………………………………..

…………………………………………………………………………………………………….

Calcul de la dernière probabilité théorique **`x.xxxxx`** :……………………..…..……..…………

…………………………………………………………………………………………………….

Décision : …………………………………………………………………..

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

Critique (si vous le souhaitez) : ………………………………………………………………..

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

Numéro d'identifiant à reporter :……………………………………..………………………

Question 4.

Proche de l'île de la Tortue, des pirates ont trouvé sur une plage un coffre contenant des pièces en or (1) et en argent (0).

```
coffre=rep(c(1,0),c(80,120))
```

On réalise 10000 échantillons de taille 20, sans remise, et on compte le nombre de pièces d'or dans chacun des échantillons. Le résultat des 10000 nombres de pièces d'or est donné dans la table ci-dessous :

```
> table(result)
result
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
   2   17  106  304  676 1239 1832 1837 1602 1255  662  334  106   22    5    1
```

On fournit également les informations suivantes :

- loi binomiale du tirage de 20 pièces dans un coffre contenant 80 pièces en or et 120 pièces en argent

```
> dbinom(0:20,20,80/120)
 [1] 2.867972e-10 1.147189e-08 2.179659e-07 2.615590e-06 2.223252e-05
 [6] 1.422881e-04 7.114406e-04 2.845762e-03 9.248728e-03 2.466327e-02
[11] 5.425920e-02 9.865310e-02 1.479796e-01 1.821288e-01 1.821288e-01
[16] 1.457030e-01 9.106440e-02 4.285383e-02 1.428461e-02 3.007287e-03
[21] 3.007287e-04
```

- loi hypergéométrique du tirage de 20 pièces dans un coffre contenant 80 pièces en or et 120 pièces en argent

```
> dhyper(0:20,80,120,20)
 [1] 1.825883e-05 2.892488e-04 2.128247e-03 9.670095e-03 3.042826e-02
 [6] 7.047764e-02 1.246656e-01 1.724347e-01 1.893988e-01 1.668099e-01
[11] 1.184350e-01 6.789888e-02 3.137292e-02 1.161802e-02 3.414067e-03
[16] 7.837509e-04 1.372409e-04 1.766398e-05 1.571795e-06 8.620194e-08
[21] 2.190966e-09
```

4a. Donner la probabilité de tirer 6 pièces en or du coffre selon la stratégie de calcul informatique.

…………………………………………………………………………………………………
…………………………………………………………………………………………………

4b. Donner la probabilité de tirer 6 pièces en or du coffre selon la stratégie de calcul mathématique.

…………………………………………………………………………………………………
…………………………………………………………………………………………………

4c. Comparer les résultats liés aux deux stratégies. Quelle conclusion pouvez-vous tirer du lien entre le raisonnement statistique et l'ordinateur ?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Numéro d'identifiant à reporter :……………………………………..…………………………

Question 5.

On étudie la relation entre la taille et l'indice de masse corporelle chez des jeunes filles.

5a. Première étude.

Un premier échantillon conduit à une corrélation linéaire de `-0.2263243` et la réalisation du test sous R nous fournit les résultats suivants :

```
> cor.test(taif,imcf)
        Pearson's product-moment correlation
data:  taif and imcf
t = -1.1143, df = 23, p-value = 0.2766
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5704419  0.1853852
sample estimates:
       cor
-0.2263243
```

Interpréter………………………………………………………………………………………..
………………………………………………………………………………………………...
………………………………………………………………………………………………...
………………………………………………………………………………………………...
………………………………………………………………………………………………...

5b. Deuxième étude.

Un deuxième échantillon conduit à une corrélation linéaire de `-0.2263243 et` la réalisation du test sous R nous fournit les résultats suivants :

```
> cor.test(taif2,imc2)
        Pearson's product-moment correlation
data:  taif2 and imc2
t = -25.9758, df = 12498, p-value = < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2428913 -0.2096252
sample estimates:
       cor
-0.2263243
```

Interpréter………………………………………………………………………………………..
………………………………………………………………………………………………...
………………………………………………………………………………………………...
………………………………………………………………………………………………...
………………………………………………………………………………………………...

5c. Comparaison.

Pourquoi, pour une même corrélation linéaire, le test conduit-il à deux résultats aussi différents ?
………………………………………………………………………………………………...

……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...

Question 6.

L'objectif de cet exercice est de discuter du sens des tests. Pour ce faire, on utilise la fonction `rnorm(n,moyenne,écart-type)` qui tire au hasard `n` individus d'une loi normale.

Etude 1

On extrait deux échantillons d'une population de loi normale $\mathcal{N}(173\ ,15)$.

```
echant1=rnorm(24,173,15)
echant2=rnorm(24,173,15)
t.test(echant1,echant2, var.eq=T)

        Two Sample t-test

data:  echant1 and echant2
t = 0.5878, df = 46, p-value = 0.5596
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
 -5.94318 10.84554
sample estimates:
mean of x mean of y
 178.2490  175.7978
```

Interpréter ce résultat………..……………………………………………………………………..
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...

On réalise le test précédent 1000 fois et on obtient donc 1000 valeurs de t sous l'hypothèse H_0. Au risque de première espèce $\alpha=0.05$, dans combien de cas rejette-t-on H_0 ?

jamais ☐ toujours ☐ autre ☐

Expliquer votre réponse………..…………………………………………………………………..
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...

Etude 2

On extrait deux échantillons l'un d'une population de loi normale $\mathcal{N}(173,15)$ l'autre d'une population de loi normale $\mathcal{N}(168,15)$. On réalise le test de Student 1000 fois. Sous l'hypothèse H_0, au risque de première espèce $\alpha=0.05$, on trouve que le test ne rejette pas H_0 pour 150 cas.

Calculer la probabilité associée et lui donner un sens………………………………………………..

……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...

Etude 3

On extrait deux échantillons l'un d'une population de loi normale $\mathcal{N}(173,15)$ l'autre d'une population de loi normale $\mathcal{N}(168,15)$. On réalise un autre test de comparaison de deux distributions 1000 fois. Sous l'hypothèse H_0, au risque de première espèce $\alpha=0.05$, on trouve que le test ne rejette pas H_0 pour 35 cas.

Calculer la probabilité associée……………………...……………………………………………..
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...

Question finale

Aux études 2 et 3 sont associées deux tests de comparaisons de moyenne. Lequel préférez-vous et pourquoi ?...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...
……………………………………………………………………………………………...