

# Traitement et Analyse de Données Biologiques – L2S4

## Examen de la Session 1

Ma 16 février 2016

- Durée : 2 h 00

Paris Sud - Orsay

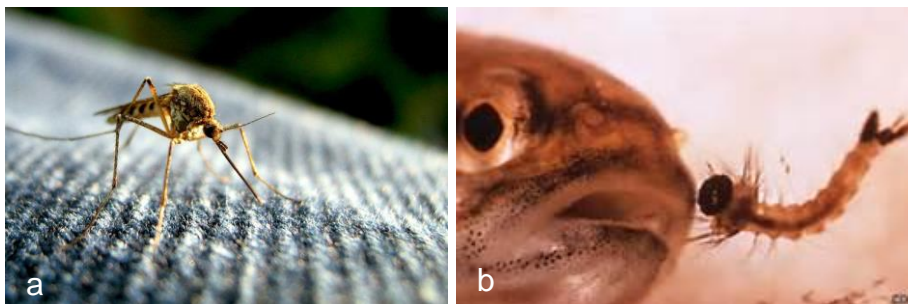
- Les questions posées sont, pour la plupart, indépendantes
- Tous les documents sont autorisés
- Téléphones portables éteints
- N'oubliez pas de reporter votre N° d'anonymat sur les intercalaires
- Reportez également votre N° d'anonymat sur la page 4 que vous devez rendre
- Le texte de cet examen est composé de 6 pages. La page 4 est à rendre (complétée).

Le terme QCM est employé pour désigner une question à choix multiple. Il vous faut dans ce cas indiquer, pour chaque proposition, si elle est vraie ou si elle est fausse en reportant le numéro de la question et la réponse choisie sur votre copie (par ex : Q12-a : FAUX; Q12-b : FAUX; Q12-c : VRAI; Q12-d : FAUX; Q12-e : VRAI)

Environ 3500 espèces de moustiques, réparties en 111 genres, ont été inventoriées dont 200 seulement piquent l'homme. Présents dans tous les continents, sauf en Antarctique, les moustiques ont un rôle dans les écosystèmes (les larves qui contribuent notamment à filtrer et nettoyer les nappes d'eau sont, par ailleurs, une importante source de nourriture). Ils sont également le plus important groupe de vecteurs d'agents pathogènes transmissibles à l'homme (Plasmodium, virus, ...).

Dans cet examen nous nous intéressons à une étude sur l'émergence d'un virus récemment apparu en Afrique occidentale qui est transmis par les moustiques.

La plupart des questions peuvent être traitées indépendamment les unes des autres.



**Figure 1 :**

**a/** Moustique s'appêtant à piquer à travers un vêtement pour prendre son repas de sang.

**b/** Poisson sur le point de capturer une larve de moustique.

### 1<sup>ère</sup> partie

**Q1/** Définissez, dans ce contexte, une variable aléatoire continue qui pourrait suivre une loi normale, une variable aléatoire discrète qui pourrait suivre une loi de binomiale ainsi qu'une variable qualitative. Vous avez entière liberté pour le choix (il est vaste) de ces variables. Vous indiquerez, pour chacune d'entre elles, si elle est définie dans une population ou un échantillon (population ou échantillon que vous définirez également, en 2 ou 3 mots).

**Q2/ QCM** > Indiquez sur votre copie si chacune des propositions suivantes est vraie ou fausse :

- a/ Le moustique sur la figure 1a est un individu au sens statistique du terme [ VRAI ou FAUX ? ]
- b/ Le moustique en figure 1a appartient à la fois à un échantillon et à une population [ VRAI ou FAUX ? ]
- c/ Le volume de sang prélevé par jour par un moustique est une variable aléatoire continue [ VRAI ou FAUX ? ]
- d/ L'espèce du moustique est définie par une variable binomiale [ VRAI ou FAUX ? ]
- e/ Le nombre d'individus piqués en 1 nuit par ce moustique est une variable qualitative [ VRAI ou FAUX ? ]
- f/ Le nombre de larves capturées en un jour par un poisson suit une loi de Poisson [ VRAI ou FAUX ? ]

**Q3/** Le temps de vie des moustiques de l'espèce *Aedes aegypti* suit une loi normale de moyenne 21 jours et d'écart-type 5 jours.

- Calculez la probabilité qu'un moustique pris au hasard de la population puisse vivre plus de 4 semaines.
- Quelle est la probabilité de trouver des individus dont le temps de vie est distant de plus de deux écarts-types de la moyenne de 21 jours ?

## **2<sup>ème</sup> partie**

On s'intéresse à la propagation dans 5 zones différentes (littoral, forêt, villes moyennes, capitale et savane) d'un nouveau virus émergent présentant 3 mutations sur le déterminant antigénique majeur (protéine à la surface de la capsid reconnue par un récepteur de l'hôte). La figure 2 (page 3) représente le résultat d'une analyse réalisée dans un pays d'Afrique occidentale à partir de 385 cas déclarés de la maladie provoquée par le virus (malades accueillis dans des dispensaires).

**Q4/** De quel type d'analyse s'agit-il ?

**Q5/** Que cherche-t-on à établir ?

**Q6/** Combien de variables sont prises en compte dans cette analyse ?

**Q7/** Quel(/s) est(/sont) le(/s) type(/s) de ces variables ?

**Q8/** Quel test proposez-vous pour étudier si ce que vous cherchez à établir est ou non significatif ?

*(on ne vous demande pas de réaliser ce test)*

**Q9/** *A priori*, sans plus de précision que ce qui est donné dans le court texte introductif, est-on ici en situation d'observation ou bien en situation d'expérience ?

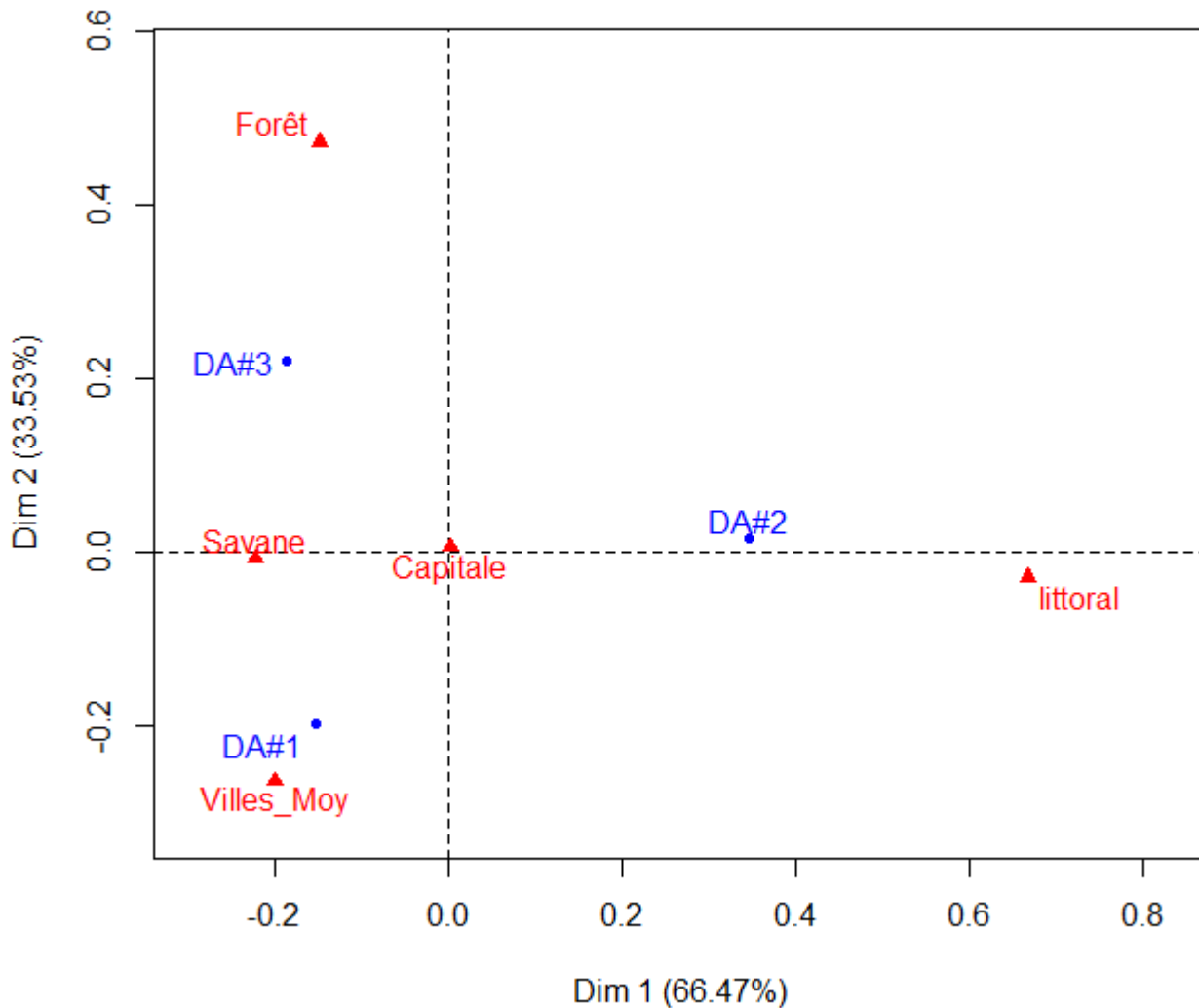
**Q10/** Selon les éléments dont vous disposez, combien y-a-t-il de degrés de liberté ?

**Q11 /** La p-value obtenue à l'issue du test que vous avez proposé en Q6 est de 0,000034. Qu'en concluez-vous ?

**Q12/** Sans en calculer la valeur, avez-vous toutes les données pour mesurer l'intensité de ce que vous cherchez à révéler ?

**Q13/** Proposez une synthèse **de 7 à 8 lignes, tout au plus**, qui expose les **principaux** résultats que vous pouvez extraire de l'analyse de la figure 2 (n'oubliez pas de commenter les valeurs portées entre parenthèse près des axes).

## CA factor map



**Figure 2** - Analyse des données collectées en Afrique occidentale sur la propagation dans 5 zones identifiées (littoral, forêt, villes moyenne, capitale, savane, indiquées en rouge sur la figure) d'un nouveau virus émergent dont le déterminant antigénique majeur (protéine à la surface de la capsid) présente 3 mutations répertoriées (identifiées en bleu par DA#1, DA#2, DA#3). 385 cas déclarés de la maladie ont été inclus dans cette analyse.

### 3<sup>ème</sup> partie

**Q14/** Dans une étude complémentaire, on s'est intéressé aux vecteurs de cette maladie. Quatre vecteurs potentiels ont été surveillés : *Aedes aegypti*, *Aedes albopictus*, *Anopheles gambiae* et *Culex Pipiens*

Des moustiques ont été piégés dans des dispositifs spéciaux (disposés au hasard dans la zone d'étude) puis on a recherché chez chacun d'eux la présence de virus (grâce à un test rapide très récemment mis au point). Ce travail a révélé que *Anopheles gambiae* et *Culex Pipiens* ne sont pas de bons vecteurs du virus ; seuls *Aedes aegypti* et *Aedes albopictus* (moustique tigre) transmettent réellement le nouveau virus.

Les résultats sont les suivants :

164 moustiques *Aedes aegypti* capturés dont 8 portaient le virus ; 102 moustiques *Anopheles gambiae* capturés, aucun ne portait le virus ; 211 moustiques *Aedes albopictus* capturés dont 19 portaient le virus et 94 moustiques *Anopheles gambiae* capturés, aucun ne portait le virus.

- Proposez (sans le réaliser ; on ne vous demande pas de réaliser des calculs) un test **unilatéral** permettant de comparer ces statistiques.
- Combien d'échantillons et de populations doit-on considérer dans ce test?
- Que cherche-t-on à établir ?
- Posez l'hypothèse nulle  $H_0$  ainsi que l'hypothèse alternative  $H_1$  (chacune en 1 ligne seulement)
- Quelle est la variable dépendante ?
- Quel est l'estimateur entrant dans la statistique de test ?
- Quelle est la statistique de test utilisée ?
- La p-value obtenue est 0,1828. Que permet de conclure ce test ?
- Quelle critique (en 5 à 6 mots) pouvez-vous apporter à cette analyse ?

**Q15/** On a également réalisé sous R le test de Student portant sur l'analyse bivariée suivante :

variable d'échantillonnage : temps de vie terrestre moyen ;

variable qualitative : espèce du moustique (*Aedes aegypti* et *Aedes albopictus*)

*(rq :le temps de vie aquatique concerne les larves);*

```
> t.test(Tvie_Aedes_aegypti,Tvie_Aedes_albopictus,var.equal=TRUE)
```

```
Two Sample t-test
```

```
data: Tvie_Aedes_aegypti and Tvie_Aedes_albopictus
```

```
t = -0.7275, df = 283, p-value = 0.4675
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-1.5876565  0.7307445
```

```
sample estimates:
```

```
mean of x mean of y
```

```
19.88341  20.31187
```

Essayer de reconstruire le fil du test, à savoir :

- D'après vous et en toute logique, que contiennent les 2 variables *Tvie\_Aedes\_aegypti* et *Tvie\_Aedes\_albopictus* utilisées dans le test ci-dessus (effectué avec le logiciel R)?
- Quelle est la variable dépendante ?
- Que cherche-t-on à établir ?
- Combien d'échantillons et de populations sont sous-jacentes ?
- Posez l'hypothèse nulle (en 1 ligne seulement)
- Quelle est la statistique de test ?
- Que le permet de conclure ce test ?

**Q16/** On s'intéresse, pour finir, aux répercussions sur la santé humaine et animale de l'épandage d'un insecticide utilisé pour éradiquer le moustique porteur du virus. Donnez dans ce contexte un exemple d'analyse bivariée que vous pourriez réaliser ; citez la variable dépendante considérée

*(on ne vous demande aucun calcul dans cette question).*

### Table de la loi normale centrée réduite

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

