

Traitement et Analyse de Données Biologiques – L2S4

Examen de la Session 1

Ma 10 février 2015

- 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)

Nous nous intéressons dans cet examen à une étude sur la pollinisation de plantes à baies comestibles menée dans une région d'Europe du Nord ainsi qu'aux répercussions d'épandages dans les champs sur la recherche de fleurs à butiner par les abeilles.



« Si l'abeille disparaissait de la surface du globe,
l'homme n'aurait plus que quatre années à vivre »
Albert Einstein

Figure 1

1^{ère} partie

Q1/ Définissez, **dans ce contexte**, une variable aléatoire continue, une variable aléatoire discrète ainsi qu'une variable qualitative. Vous avez entière liberté pour le choix (et 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/ Donnez, **dans le contexte de la figure 1**, montrant une abeille (*Apis mellifera*) qui butine une fleur, une variable qui pourrait suivre une loi normale et une variable qui pourrait suivre une loi de Binomiale.

Q3/ QCM > Indiquez sur votre copie si chacune des propositions suivantes est vraie ou fausse :

- a/** L'abeille sur la figure 1 est un individu au sens statistique du terme [VRAI ou FAUX ?]
- b/** L'abeille sur la figure 1 appartient à la fois à un échantillon et à une population [VRAI ou FAUX ?]
- c/** La masse de pollen récolté par une abeille est une variable aléatoire continue [VRAI ou FAUX ?]
- d/** La masse de pollen récolté par une abeille est une variable qualitative [VRAI ou FAUX ?]
- e/** La masse de pollen récolté par une abeille est une variable binomiale [VRAI ou FAUX ?]



Figure 2

Un minuscule émetteur radio placé sur leur thorax, permet de mesurer le déplacement des abeilles dans un rayon allant jusqu'à 10 km autour des ruches.

Les distances maximales de butinage (par rapport à leur ruche, qui a la cote zéro, c'est-à-dire pour laquelle : distance=0) ont été enregistrées par un doctorant sur 401 abeilles appartenant à 5 ruches d'une même zone. Ces distances ont été acquises 15 jours avant l'épandage de produits chimiques dans le champ attenant aux ruches. Le jeune doctorant a réalisé la figure 3 avec le logiciel R.

Q4/ On considère dans cette question le graphe présenté en [figure 3](#)

- a/** Quelle est la variable aléatoire impliquée dans l'histogramme présenté en figure 3 ?
Est-elle discrète ou continue ?
- b/** La distribution présentée en figure 3 est-elle associée à une population ou à un échantillon ?
- c/** Combien d'abeilles, environ, parcourent-elles entre 1,5 et 2 km ?
- d/** Calculez une valeur approchée de la moyenne de la distribution.
- e/** Sachant que l'écart-type calculé à partir de cette distribution vaut 0,7 km , dessinez à main levée la loi normale qui approche la distribution de la figure 3 (on ne vous demande pas d'en faire un tracé précis)
- f/** En utilisant cette loi normale, déterminez la distance dépassée par seulement 5% des abeilles
- g/** Donnez l'intervalle de confiance de la moyenne de la population d'origine au seuil de confiance de 95%

Q5/ On considère dans cette question le graphe présenté en [figure 4](#)

- a/** Sur la figure 4, sont représentées, à l'aide de boîtes de Tukey, les distributions des distances parcourues par les abeilles provenant de 2 ruches, l'une exposée aux vents d'ouest (ruche A), l'autre à l'abri du vent en lisière d'une forêt (ruche B). Les résultats d'un test effectué sur les 2 séries avec le logiciel R sont présentés ci-dessous (copie d'écran). Commentez ces résultats en 5 à 6 lignes maximum (*on ne vous demande aucun calcul*).

```
> t.test(dist_rucheA, dist_rucheB)

      Welch Two Sample t-test

data:  distance and distrib
t = 5.3025, df = 745.651, p-value = 1.506e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1345832 0.2928200
sample estimates:
mean of x mean of y
 1.754377  1.540675
```

- b/** Aurait-on pu effectuer un test du Chi-deux à la place du test effectué en Q5-a?
(justifiez votre réponse en une ligne)

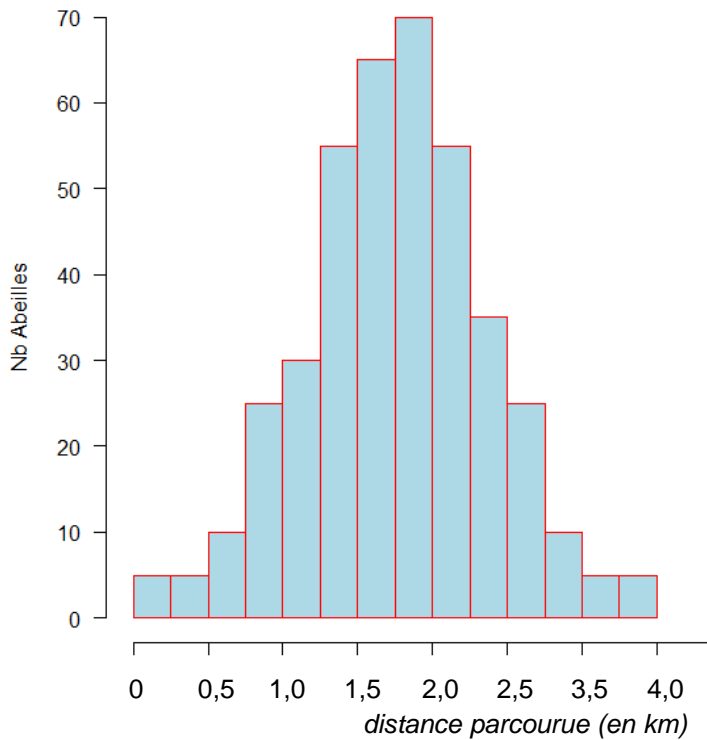


Figure 3 Histogramme de la distribution des distances

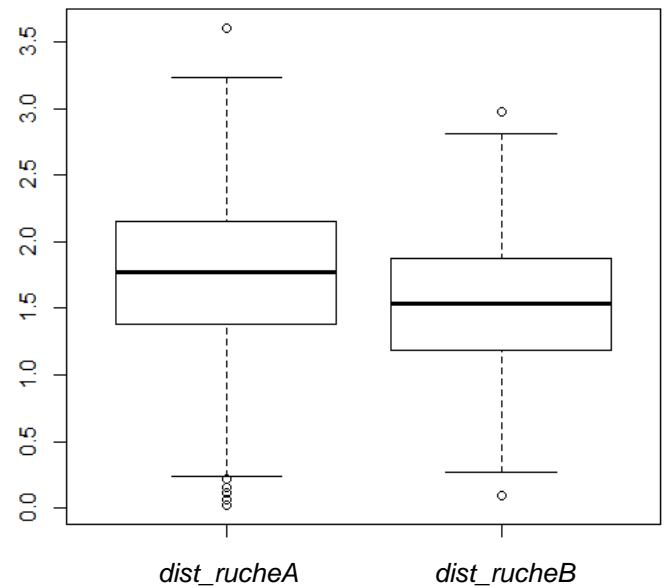


Figure 4 Boîtes de Tukey des distributions des distances parcourues par les abeilles provenant des ruches A et B

2^{ème} partie

Nous nous intéressons maintenant, dans cette 2^{ème} partie, à la pollinisation des groseilliers par les abeilles domestiques.

Selon un protocole bien établi (que nous ne détaillerons pas) des observateurs ont noté, dans des parcelles de 2 m² exposés aux pollinisateurs, les visites des abeilles sur les fleurs de certaines tiges clairement identifiables. Quelques semaines plus tard, les fruits récoltés sur ces tiges ont été pesés au laboratoire avec une balance de précision (au mg près). On a ainsi pu mener une étude quantitative avec le logiciel R.

Q6/ Interprétez de la façon la plus complète possible le résultat de l'analyse réalisée avec le logiciel R dont une copie d'écran figure ci-dessous (définissez bien la variable réponse et la variable explicative).

La variable *poidsfruit* contient la masse des fruits et la variable *ovfec* contient le nombre d'ovules fécondés.

```
> summary(lm(poidsfruit~ovfec))
```

Call:

```
lm(formula = poidsfruit ~ ovfec)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.85616	-0.18775	-0.00639	0.19934	0.79492

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.263525	0.029153	43.34	<2e-16 ***
poidsfruit	0.029125	0.001036	28.12	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2812 on 382 degrees of freedom

Multiple R-squared: 0.6742, Adjusted R-squared: 0.6734

F-statistic: 790.7 on 1 and 382 DF, p-value: < 2.2e-16

Q7/ Sur quelles données pourrions-nous faire un test de Shapiro-Wilk et à quelle fin? Justifiez votre réponse en une à deux lignes maxi.

N° d'anonymat :

N° de table :

(Feuille annexe à rendre dûment complétée)

Q8/ Ajoutez le tracé du modèle testé à la question Q6/ sur le graphique de la figure 5 que vous rendrez avec votre copie (ne prenez pas une trop grande précision, arrondissez judicieusement). Vous complétez ce graphique en y ajoutant également un titre et les intitulés manquant des deux axes

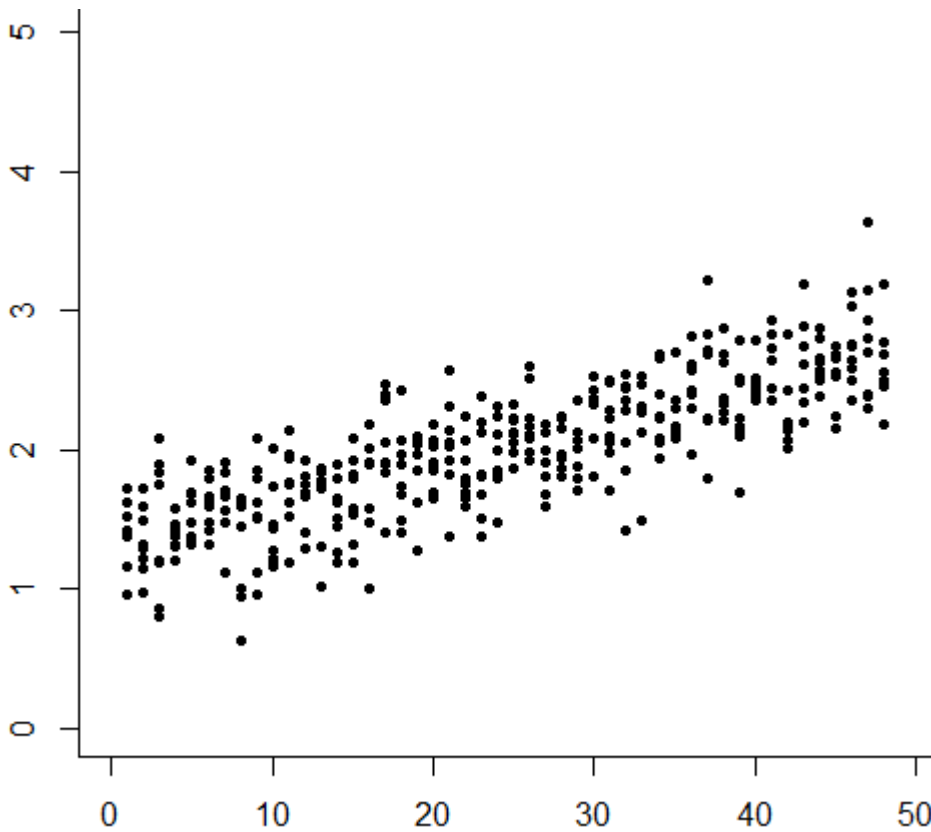


Figure 5 *Impact de la pollinisation des groseilliers rouges*

3^{ème} partie

Nous nous intéressons dans cette 3^{ème} partie, aux répercussions d'épandages dans les champs sur la recherche de fleurs à butiner par les abeilles. Il s'agit plus précisément d'étudier le déplacement des abeilles (*Apis mellifera*) avant, pendant et après l'épandage dans les champs avoisinant les ruches.

```
temps;dist;etdist;nbees
0;0.96;0.10;5
1;1.37;0.15;15
2;2.42;0.22;10
3;3.52;0.41;12
4;3.99;0.42;8
5;3.77;0.36;9
6;2.62;0.24;11
7;1.52;0.15;15
8;1.34;0.25;9
```

Le fichier '**bees.stat**', dont le contenu figure dans l'encadré ci-contre, rassemble les mesures relevées sur 94 abeilles d'une même ruche par une jeune technicienne.

La date relative au jour de l'épandage est indiquée en jours en 1^{ère} colonne du fichier (le jour de l'épandage, temps=0); en 2^{ème} colonne est reportée la moyenne des distances (en km) parcourues par les abeilles au jour indiqué en 1^{ère} colonne ; l'écart-type des distances parcourues est porté (en km) sur la 3^{ème} colonne ; enfin, le nombre d'abeilles considérées à chaque jour d'étude est en 4^{ème} colonne.

La figure 6, qui a été obtenue à partir des données de ce fichier, représente la distance moyenne parcourue (en km) aux différents temps (en jours) de l'étude.

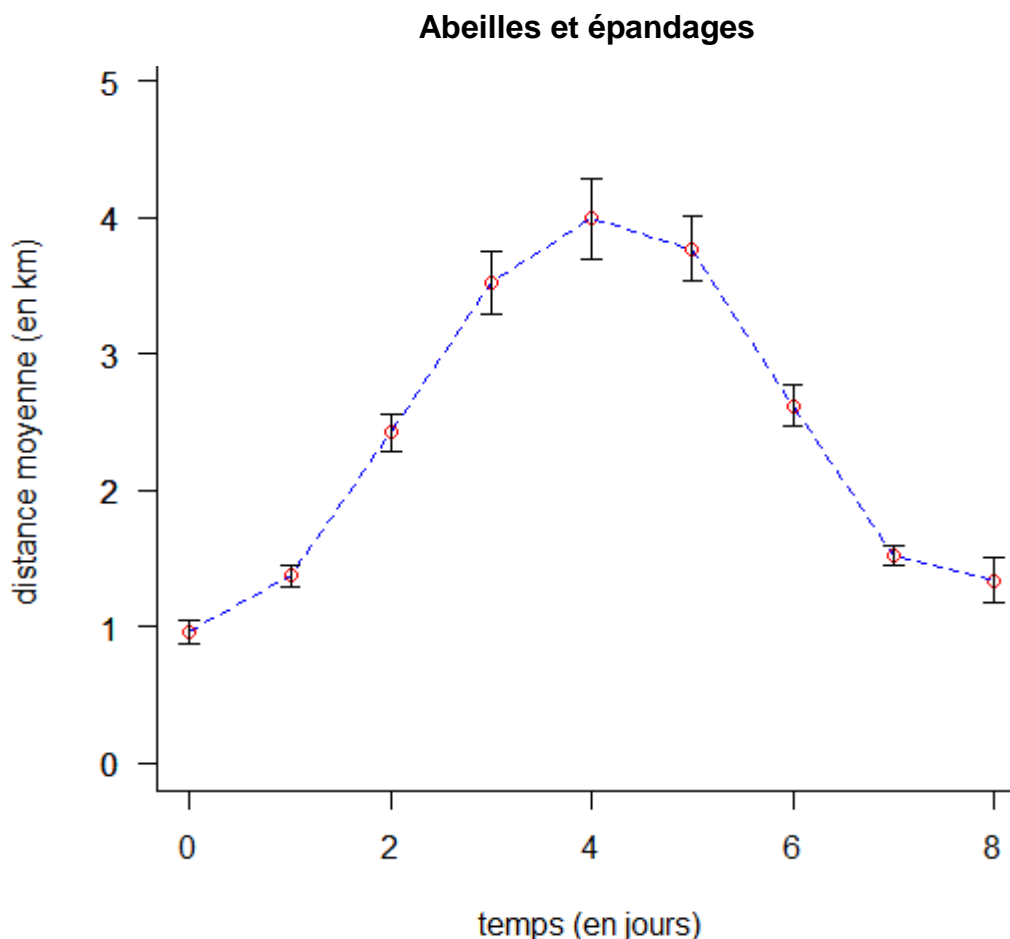


Figure 6 Distance moyenne parcourue aux différents temps de l'étude

Q9/ Expliquez, en 6 à 7 lignes maxi, comment le graphe reporté en figure 6 a été obtenu à partir du fichier '**bees.stat**' (points, barres d'erreur)

Dans ce qui suit, pour faciliter les énoncés, X désigne le temps en jours et Y, la distance parcourue en km

Q10/ QCM : Indiquez sur votre copie si chacune des propositions suivantes, concernant le graphe en figure 6, est vraie ou fausse.

- a/** On peut dire que cette courbe décrit un phénomène stochastique liant les variables X et Y, pour X compris entre 0 et 8 jours et Y compris entre 0 et 5 km. [VRAI ou FAUX ?]
- b/** Il n'est pas logique que la courbe ne passe pas par l'origine (0,0) du graphe ; ceci constitue une erreur. [VRAI ou FAUX ?]
- c/** Il y a 9 points représentés sur le graphique, associés à 9 variables aléatoires à considérer pour la dimension portée par l'axe des ordonnées. [VRAI ou FAUX ?]
- d/** Si les barres d'erreurs sur Y sont les seules qui sont représentées c'est parce qu'on ne dispose pas de données permettant d'évaluer celles sur X ; il y a également une incertitude statistique sur les valeurs de X. [VRAI ou FAUX ?]
- e/** Si l'on disposait de 70 valeurs mesurées de Y pour X=4 jours (correspondant aux distances parcourues par 70 abeilles), on pourrait représenter les variations de Y pour X=4 jours à l'aide d'un histogramme. [VRAI ou FAUX ?]