

## Statistique Inférentielle - Examen du 15/05/2015

Durée : 1h30 - Tous documents autorisés

Consignes : la qualité de la rédaction sera prise en compte dans la note. Sauf avis contraire, tous les intervalles de confiance seront réalisés avec un niveau de confiance de 95% et les tests avec un risque de première espèce de 5%.

### Exercice 1

L'Université Lyon 2 réalise une enquête sur l'employabilité des étudiants qu'elle forme. Pour ce faire, un questionnaire est envoyé aux étudiants de différents masters, une année après l'obtention de leur diplôme. Une des questions demande le nombre de mois de recherche d'emploi avant l'obtention d'un premier emploi. Bien entendu tous les diplômés n'ont pas répondu, et le nombre de réponses diffère d'un Master à l'autre. Voici les réponses obtenues pour deux masters :

Master A	2	0	5	3	1	0	9																								
Master B	5	8	1	0	3	5	4	2	5	3	4	3	2	2	6	5	3	6	3	5	1	4	2	1	4	5	4	2	2	3	0

Pour vous aider dans vos calculs, j'ai exécuté sous le logiciel R le code suivant :

```
> A=c(2,0,5,3,1,0,9)
> B=c(5,8,1,0,3,5,4,2,5,3,4,3,2,2,6,5,3,6,3,5,1,4,2,1,4,5,4,2,2,3,0)
> print(mean((A-mean(A))^2))
8.979592
> print(mean((B-mean(B))^2))
3.379813
```

1. Donner une estimation, pour chaque master, du temps moyen (en mois) avant de trouver un premier emploi ainsi que de l'écart-type de ce temps (pour chaque Master).
2. L'Université Lyon 2 s'attend à ce que les étudiants de ses masters trouvent un emploi avant 5 mois. Pouvez-vous confirmer cette hypothèse pour ces 2 Masters ?
3. Peut-on affirmer que l'un des deux master permet de trouver un emploi plus vite que l'autre ? Là encore je vous aide avec le code R suivant (attention, les indications que vous donne ce code ne doivent pas être utilisées dans les questions précédentes) :

```
> shapiro.test(A)
data: A
W = 0.8724, p-value = 0.1949
> var.test(A,B)
data: A and B
F = 2.9996, num df = 6, denom df = 30, p-value = 0.04079
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
1.046378 15.193898
sample estimates:
ratio of variances
2.999648
```

### Exercice 2

Afin de préparer les étudiants au TOEIC, un test de langue, les enseignants de langue de Lyon 2 ont mis au point 4 techniques d'apprentissage de l'anglais. Pour évaluer ces techniques, les 40 étudiants d'une promotion sont répartis en 4 groupes, utilisant chacun une des 4 techniques d'apprentissage. En fin d'année, ils passent tous un test TOEIC. Les résultats ( $T_1$  à  $T_4$  correspondant aux scores TOEIC avec les techniques d'apprentissage 1 à 4) sont donnés ci-après.

$T_1$	810	770	800	730	720	890	900	660	890	900
$T_2$	820	780	820	720	760	690	770	690	790	910
$T_3$	770	650	700	600	720	870	700	960	690	720
$T_4$	860	750	590	830	600	880	880	660	810	930

De leur côté, les enseignants de statistique assurent que le niveau d'anglais dépend des connaissances en statistiques. Les notes à l'examen de statistique ( $S_1$  à  $S_4$ ) de ces étudiants sont donc également relevées.

$S_1$	13	15	15	15	17	16	16	11	15	17
$S_2$	17	16	14	14	16	16	15	11	16	18
$S_3$	15	11	13	13	14	15	14	14	13	15
$S_4$	15	14	12	14	17	14	19	12	16	18

Pour vous aider dans vos calculs, j'ai exécuté sous le logiciel R le code suivant :

```
> T1=c(810,770,800,730,720,890,900,660,890,900)
> T2=c(820,780,820,720,760,690,770,690,790,910)
> T3=c(770,650,700,600,720,870,700,960,690,720)
> T4=c(860,750,590,830,600,880,880,660,810,930)
> print(colMeans(cbind(T1,T2,T3,T4)))
T1 T2 T3 T4
807 775 738 779
> print(mean(c(T1,T2,T3,T4)))
774.75
> print(var(cbind(T1,T2,T3,T4)))
      T1          T2          T3          T4
T1 7512.222 2661.111 -1673.333 7352.222
T2 2661.111 4472.222 -2877.778 1861.111
T3 -1673.333 -2877.778 11151.111 -1402.222
T4 7352.222 1861.111 -1402.222 15121.111
> print(var(c(T1,T2,T3,T4)))
9446.09
> S1=c(13,15,15,15,17,16,16,11,15,17)
> S2=c(17,16,14,14,16,16,15,11,16,18)
> S3=c(15,11,13,13,14,15,14,14,13,15)
> S4=c(15,14,12,14,17,14,19,12,16,18)
> print(var(c(S1,S2,S3,S4)))
3.717308
> print(cov(c(T1,T2,T3,T4),c(S1,S2,S3,S4)))
102.1218
```

1. En supposant que cette promotion est représentative des étudiants de Lyon 2, pouvez-vous affirmer que le score moyen au TOEIC est inférieur à 800 ?
2. Les 4 techniques d'apprentissage sont-elles équivalentes pour préparer au TOEIC ?
3. Montrer qu'il existe un lien entre les notes de statistique et les scores TOEIC.
4. Question bonus : construire un modèle permettant de prédire la note au test du TOEIC en fonction de la note en statistique, et utiliser ce modèle et votre dernière note de statistique pour prédire le score que vous auriez au TOEIC.