

Université Toulouse 1 Capitole École d'économie de Toulouse

Année universitaire 2017–2018

Session 1

Semestre 5

Licence 3 mention Économie

Epreuve : Analyse des Données

Date de l'épreuve : Mardi 12 décembre 2017

Durée de l'épreuve : 1h30

Liste des documents autorisés : aucun.

Liste des matériels autorisés : calculatrice non programmable.

Nombre de pages (y compris page de garde) : 7.

Partie I : Tests statistiques

Merci de répondre à la **Partie I : Tests statistiques** et à la **Partie II : Analyse des données sur deux feuilles distinctes**.

Pour tous les tests statistiques demandés, on prendra un risque α de 5 %. On écrira les hypothèses nulle et alternative testées, on donnera le nom du test utilisé, sa statistique de test, sa règle de décision et on donnera ses conditions de validité (à vérifier quand c'est possible). Quand un test est significatif, on commentera la différence ou la liaison mise en évidence.

Les résultats des calculs seront donnés avec 2 décimales après la virgule.

Barème indicatif sur 10 : 3 – 4 – 3.

Exercice A (3 points)

Pendant le cours de mathématiques, Paul joue avec la touche “rnd” (pour “random”) de sa calculatrice, qui est censée lui donner à chaque appui une réalisation d'une variable aléatoire de loi uniforme sur $[0,1]$.

En appuyant 100 fois de suite sur la touche, il obtient les résultats suivants, où les nombres affichés par sa calculatrice ont été rangés par classes de valeurs et où la deuxième ligne correspond à l'effectif observé :

$[0;0,1[$	$[0,1;0,2[$	$[0,2;0,3[$	$[0,3;0,4[$	$[0,4;0,5[$	$[0,5;0,6[$	$[0,6;0,7[$	$[0,7;0,8[$	$[0,8;0,9[$	$[0,9;1]$
9	11	11	12	12	11	10	8	6	10

Paul se demande si sa calculatrice est fiable, c'est-à-dire si la touche “rnd” fait bien ce qu'elle est censée faire. Répondre à cette question par un test statistique.

Exercice B (4 points)

Un échantillon de 31 ampoules électriques de qualité A a donné une durée de vie moyenne de 1400 h et d'écart-type estimé $s_1 = 100$ h.

Un échantillon de 21 ampoules électriques de qualité B a donné une durée de vie moyenne de 1200 h et d'écart-type estimé $s_2 = 80$ h.

1. On suppose que la durée de vie des ampoules suit une loi normale.
Peut-on considérer que les deux qualités d'ampoules sont équivalentes? On répondra à cette question à l'aide de deux tests statistiques.
2. Donner le nom du test qu'il faudrait effectuer si on ne supposait plus une loi normale pour la durée de vie des ampoules.

Exercice C (3 points)

Une entreprise qui conditionne des céréales de petit-déjeuner hésite entre quatre emballages différents pour leur nouveau produit. Pour permettre de trancher, dix points de vente de chiffres d'affaires semblables sont choisis. On affecte par tirage au sort un type d'emballage à chaque magasin et on relève pendant un mois les chiffres de vente du nouveau produit. Le but est, à l'aide d'un modèle d'analyse de variance, de déterminer si le type d'emballage (variable **type**) a une incidence sur les ventes de céréales (variable **ventes**).

On donne les sorties de R suivantes :

```
> tapply(ventes,type,mean)
 1  2  3  4
14 13 19 23
```

```
> leveneTest(ventes,type)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 3  1.2463  0.373
      6
```

```
> anova(aov(ventes~as.factor(type)))
Analysis of Variance Table
```

Response: ventes

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(type) ..	160		0.1639
Residuals ..	132			

1. Donner le tableau d'analyse de variance et le compléter.
2. A l'aide d'un test statistique, déterminer si le type d'emballage a un effet significatif sur les ventes de ce nouveau produit.

On rappelle que les quantiles de la loi de Student pour un degré de liberté supérieur ou égal à 30 sont proches de ceux de la loi normale, à savoir $z_{0,05} = 1,645$ et $z_{0,025} = 1,96$.

On donne les quantiles suivants des lois de χ^2 : $\chi_{1;0,05}^2 = 3,84$, $\chi_{2;0,05}^2 = 5,99$, $\chi_{5;0,05}^2 = 11,07$, $\chi_{6;0,05}^2 = 12,59$, $\chi_{7;0,05}^2 = 14,07$, $\chi_{8;0,05}^2 = 15,51$, $\chi_{9;0,05}^2 = 16,92$, $\chi_{10;0,05}^2 = 18,31$, $\chi_{11;0,05}^2 = 19,68$.

On donne les quantiles suivants des lois de Fisher : $F_{20;30;0,95} = 1,93$, $F_{21;31;0,95} = 1,91$, $F_{31;21;0,95} = 2,00$, $F_{30;20;0,95} = 2,04$, $F_{20;30;0,975} = 2,20$, $F_{30;20;0,975} = 2,35$, $F_{21;31;0,975} = 2,16$, $F_{31;21;0,975} = 2,30$.

Partie II : Analyse des données

Merci de répondre à la Partie I : Tests statistiques et à la Partie II : Analyse des données sur deux feuilles distinctes.

Questions de cours (2 points)

1. A quel type de données est adaptée une analyse en composantes principales (ACP) ?
2. Donner le nom de la fonction de R permettant de réaliser une ACP, ainsi que le nom du package R nécessaire pour utiliser cette fonction.
3. Donner le nombre maximum de valeurs propres calculées dans une ACP.
4. En ACP, donner la formule de la contribution d'une variable à une composante principale ainsi que la formule de la contribution d'un individu à une composante principale.

Analyse factorielle discriminante (AFD) (4 points)

Dans le cadre d'une étude de marché pour un constructeur automobile, un échantillon de trente automobilistes est interrogé. Celui-ci est séparé en deux groupes : ceux qui comptent changer de véhicule dans les deux ans à venir (changement = oui) et ceux qui n'ont aucune intention de le faire (changement = non). On pose aux répondants des questions personnelles (âge et nombre de personnes dans le foyer) et des questions liées à leur automobile : nombre de km par an, âge du véhicule et budget annuel hors carburant (garagiste, accessoires, assurance...).

Utiliser les sorties du logiciel R ci-dessous pour répondre aux questions suivantes.

1. Expliquer la raison pour laquelle il y a un seul axe discriminant et interpréter cet axe.
2. Commenter la qualité de la discrimination.
3. Donner la valeur moyenne de l'axe discriminant dans chacun des deux groupes.
4. Calculer le taux d'automobilistes mal classés.
5. Prédire le groupe d'affectation pour un nouvel automobiliste de 40 ans, dont la taille du foyer est de 3 personnes, ayant une voiture de 3 ans, parcourant 30 000 km par an et dont le budget voiture annuel hors carburant est de 2500 euros.

Group means:

	km.an	age	taille_foyer	budget	age_voiture
non	22600	44.00000	2.600000	2453.333	2.400000
oui	39000	43.66667	2.733333	3080.000	4.266667

```
apply(tabchang[,-6],2,mean)
```

	km.an	age	taille_foyer	budget	age_voiture
	30800.00000	43.833333	2.666667	2766.66667	3.333333

Coefficients of linear discriminants:

	LD1
km.an	4.421462e-05
age	3.368177e-03
taille_foyer	2.632003e-01
budget	-1.426217e-04
age_voiture	6.120583e-01

```

> round(cor(voitures2),digits=2)
          km.an  age taille_oyer budget age_voiture  LD1
km.an      1.00 0.20      -0.01  0.61      0.19 0.68
age        0.20 1.00      -0.23  0.26     -0.12 -0.02
taille_oyer -0.01 -0.23      1.00  0.02     -0.13 0.09
budget      0.61 0.26      0.02  1.00      0.26 0.47
age_voiture 0.19 -0.12     -0.13  0.26      1.00 0.82
LD1        0.68 -0.02      0.09  0.47      0.82 1.00

> summary(lm(LD~voitures$changement))$r.squared
[1] 0.4679956

> tapply(LD,tabchang$voitures$changement,mean)
      non      oui
-0.9061114      ?

> matconfusion
      prediction
      non oui
non  13  2
oui   5 10

```

Analyse factorielle des correspondances (AFC) (4 points)

Vous disposez d'un fichier issu d'une enquête emploi et concernant 684 individus. On s'intéresse aux deux variables suivantes :

- le type d'emploi occupé (variable `emploi` à 5 modalités : C si chômage, I si emploi CDI, D si emploi CDD, A si emploi aidé et N si inactivité),
- le niveau d'études (variable `annetud` à 8 modalités : 1 si sans diplôme, 2 si CAP ou BEP, 3 si Bac, 4 si Bac+2, 5 si Bac+3, 6 si Bac+4, 7 si Bac+5 et 8 si Doctorat).

Utiliser les sorties du logiciel R ci-dessous pour répondre aux questions suivantes.

1. L'AFC est-elle justifiée ?
2. Combien d'axes sont à retenir ? Justifier votre réponse.
3. Quels sont les profils ayant fortement contribué à l'inertie des axes retenus ? Justifier.
4. Quels sont les profils bien représentés sur les axes retenus ? Justifier.
5. Commenter le graphique simultané de l'AFC.

```
> chisq.test(table(emploi,annetud))
```

```
Pearson's Chi-squared test
data:  table(emploi,annetud)
```

```
X-squared = 122.1316, df = 28, p-value = 1.005e-13
```

```
> resu$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.093830096          66.459291          66.45929
dim 2 0.035482535          25.132066          91.59136
dim 3 0.008023450           5.682961          97.27432
dim 4 0.003848235           2.725682          100.00000
```

```
> resu$row
$coord
      Dim 1      Dim 2      Dim 3      Dim 4
A -0.3931826 -0.10535532  0.08186483 -0.157097972
C  0.7337677 -0.08548624 -0.10531619 -0.070927268
D -0.1079110 -0.12436424 -0.02067462  0.039450531
I -0.1202719  0.35685136 -0.05644606 -0.003221574
N  0.3624829  0.10576455  0.22308822  0.035353239
```

```
$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
A 16.138570  3.064199  8.181865 62.82004406
C 59.563097  2.137864 14.349321 13.56960127
D  6.150823 21.603296  2.640327 20.04415127
I  2.997656 69.783959  7.721499  0.05244091
N 15.149854  3.410682 67.106988  3.51376249
```

```
$cos2
      Dim 1      Dim 2      Dim 3      Dim 4
A 0.78443940 0.05632274 0.03400679 1.252311e-01
C 0.95829798 0.01300696 0.01974120 8.953851e-03
D 0.40023290 0.53158420 0.01469116 5.349174e-02
I 0.09975764 0.87819801 0.02197278 7.157379e-05
N 0.67869354 0.05778019 0.25707036 6.455905e-03
```

```
> resu$col
$coord
      Dim 1      Dim 2      Dim 3      Dim 4
1  0.73822656  0.12408343  0.038783800 -0.04438873
2  0.10934377 -0.18093513 -0.059103720  0.01809837
3  0.04068152 -0.05519779  0.080265205  0.06703267
4 -0.25734075 -0.12699599  0.071087647 -0.06290878
5 -0.14826957 -0.00775514 -0.208877662  0.04273730
6 -0.09366664  0.38808671 -0.050369989 -0.01515841
7 -0.31493591  0.28628610  0.001595878 -0.01915547
8 -0.14556643  0.16240593  0.222802798  0.27606259
```

```
$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
1 66.2331892  4.94824812  2.137854174  5.8387822
2  3.0924162 22.39149656 10.566240087  2.0657095
3  0.2088724  1.01685164  9.508727707 13.8273957
4 16.2001307 10.43299286 14.456750487 23.6050207
5  1.9524533  0.01412483 45.314959826  3.9552254
6  0.7791956 35.37216674  2.635125536  0.4975815
7 10.9724711 23.97662994  0.003294888  0.9897516
8  0.5612716  1.84748930 15.377047294 49.2205332
```

\$cos2

	Dim 1	Dim 2	Dim 3	Dim 4
1	0.96653143	0.0273063884	2.667702e-03	0.003494476
2	0.24644364	0.6748003181	7.200442e-02	0.006751626
3	0.10583332	0.1948369997	4.119863e-01	0.287343393
4	0.72484605	0.1765259865	5.531169e-02	0.043316275
5	0.32568515	0.0008909906	6.463651e-01	0.027058768
6	0.05410639	0.9288298783	1.564668e-02	0.001417052
7	0.54643011	0.4515343503	1.403104e-05	0.002021510
8	0.12218869	0.1520940794	2.862529e-01	0.439464374

