EISTI - DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE COMPUTATIONNELLE

XX janvier 2011 – **DURÉE 3h00**

La consultation des documents et l'échange des documents et des calculatrices est interdit.
L'utilisation des 3 feuilles manuscrites recto-verso, format A4 est autorisée

- **Ne pas détacher les feuilles.**
- **Utiliser l'espace blanc pour vos réponses et le verso pour brouillon. Si vous avez besoin des feuilles supplémentaires pour brouillon, vous êtes priés de demander aux surveillant(e)s.**
- **Pensez à indiquer votre nom sur chaque feuille**

NOM : ....

NOTE 20

DÉTAIL

	a	b	c	
Exercice 1.	0.5	0.5		1
Exercice 2.	1	1		2
Exercice 3.	2	1	1	4
Exercice 4.				3
Exercice 5.	1	2	1	4
Exercice 6.				4
Exercice 7.				2

Le corrigé se trouve sur `http ://sifoci.eisti.fr`***, rubrique Logique computationnelle – Prolog.***

LA BASE DE DONNÉES

Les données "iris" sont des données proposées en 1933 par le statisticien Ronald Aylmer Fisher et elles correspondent à 3 types (classes) de fleurs, à savoir iris setosa, iris virginica et iris versicolor.

Pour chaque classe il y a 50 observations, donc au total il y a 150 observations.

Chaque observation contient les mesures de quatre variables (variables quantitatives) qui sont, dans l'ordre, la longueur et la largeur des sépales (LOSE et LASE), la longueur et la largeur des pétales (LOPE et LAPE) et aussi le numéro de la classe à laquelle appartient la fleur observée (variable ordinale). Toutes les variables quantitatives sont exprimées en millimètres.

Les trois lignes suivantes présentent un échantillon du fichier de données, chaque ligne correspondant à une classe différente.

```
Long. s\'ep.  Larg. S\'ep   Long. p\'et.  Larg. p\'et. Classe
  5.100         3.500         1.400         0.200        1
  7.000         3.200         4.700         1.400        2
  6.300         3.300         6.000         2.500        3
```

On se propose de faire une étude complète de cet ensemble de données et, en particulier, l'influence de la longueur des pétales (3e variable quantitative) à l'ensemble des observations.

LES RÉSULTATS DES CALCULS PRÉLIMINAIRES

Sur l'ensemble de mesures, on a effectué plusieurs calculs dont les résultats sont donnés ci après :

En utilisant l'ensemble des observations (150 observations, 4 variables quantitatives, 1 variable qualitative (numéro de classe)), on a calculé :

1. moyenne et écart-type non biaisé (table 1)

	LOSE	LASE	LOPE	LAPE
Moyenne	5.84	3.06	3.76	1.20
Écart-type non biaisé	0.83	0.43	1.76	0.76

TABLE 1 – Grandeurs statistiques de la base de données totale

2. matrice des variances-covariances (table 2)

	LOSE	LASE	LOPE	LAPE
LOSE	0.6811	-0.0422	1.2658	0.5128
LASE	-0.0422	0.1887	-0.3275	-0.1208
LOPE	1.2658	-0.3275	3.0955	1.2870
LAPE	0.5128	-0.1208	1.2870	0.5771

TABLE 2 – Matrice des variances-covariances de la base de données totale

3. matrice des corrélations (table 3)

	LOSE	LASE	LOPE	LAPE
LOSE	1.0000	-0.1176	0.8718	0.8179
LASE	-0.1176	1.0000	-0.4284	-0.3661
LOPE	0.8718	-0.4284	1.0000	0.9629
LAPE	0.8179	-0.3661	0.9629	1.0000

TABLE 3 – Matrice des corrélations de la base de données totale

La table 4 fournit la moyenne et l'écart-type non biaisé pour la variable "longueur pétales" selon les trois classes des fleurs

	Classe 1	Classe 2	Classe 3
Moyenne	1.46	4.26	5.55
Écart-type non biaisé	0.18	0.47	0.55

TABLE 4 – Grandeurs statistiques pour la variable "longueur pétales" par classe

TRAVAIL À FAIRE

1. Parmi les 150 fleurs il y a 30 dont la longueur des pétales (LOPE) est dans l'intervalle $[5.0, 5.9]$.
   (a) On cherche à estimer la probabilité qu'une fleur prise au hasard a une longueur de pétale dans cette intervalle. Calculer l'estimation qui correspond au résultat du test.
   SOL.- Soit X_i la variable aléatoire qui prend deux valeurs : 1 si $5.0 \leq LOPE \leq 5.9$ et 0 sinon. Donc X_i suit la loi de Bernoulli $B\,(1, \widehat{p})$. L'estimateur $\widehat{p}$ est donné par la moyenne $\overline{X} = \frac{1}{150}\sum_{i=1}^{150} X_i = 30/150 = 0.20$.
   (b) Trouver la loi suivie par cet estimateur. Est-il possible d'utiliser l'approximation de la loi normale pour effectuer les calculs numériques ? Justifier votre réponse.
   SOL.- La loi suivie est la loi binomiale car nous avons deux résultats possibles avec probabilités $p = 0.2$ et $1 - p = 0.8$ respectivement.
   D'autre part, on a $n = 150 > 30$, $np = 30 \;\; > 15$ et $np\,(1-p) = 24 > 15$. Donc on peut appliquer l'approximation par la loi normale. On a donc $\overline{X} \sim \mathcal{N}(np, \sqrt{np(1-p)}) = \mathcal{N}(0.2, \sqrt{24})$.
2. On considère pour la suite que les distributions des valeurs de toutes les variables suivent une distribution normale $\mathcal{N}(\mu, \sigma)$. Proposez et calculez des estimateurs de la variable "longueur des pétales" (c'est la troisième variable dans le tableau de données) pour
   (a) la moyenne ;
   SOL.- On prend la moyenne empirique $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 3.76$
   (b) la variance non biaisée
   SOL.- On prend la variance non biaisée empirique $s^2_{1,X} = 1.76^2$
3. Nous voulons comparer les fleurs de la classe 2 avec celles de la classe 3 par rapport à la variable "longueur des pétales".
   (a) Calculer, avec un risque de $\alpha = 0.01$, un intervalle de confiance pour la moyenne et la variance de la variable "longueur des pétales" pour les fleurs de la 2e classe.
   SOL.- Calcul de l'I.D.C. pour la moyenne. La variance est inconnue et la taille de l'échantillon est $n = 50 > 30$. Donc l'estimateur suit la loi normale $\mathcal{N}(\mu, \sigma) = \mathcal{N}(4.26, 0.47)$. On a $z_{0.005} = 3.57$. Donc I.D.C. = $\left[4.26 - 3.57\frac{0.47}{\sqrt{50}}, 1.26 + 3.57\frac{0.47}{\sqrt{50}}\right] = [4.02, 4.5]$
   Calcul de l'I.D.C. pour la variance. La moyenne est inconnue. Donc l'estimateur de la variance suit la loi du χ^2_{n-1} avec $n - 1 = 49$ ddl. Nous avons $a = \chi^2_{49,0.995} = 28$ et $b = \chi^2_{49,0.005} = 79.5$. Donc I.D.C. $= \left[\frac{50 \cdot 0.47^2}{79.5}, \frac{50 \cdot 0.47^2}{28}\right] = [0.14, 0.39]$.
   (b) Vérifier si la taille de l'échantillon est suffisante pour l'estimation de la variance.
   SOL.- On a $R_\Theta = 0.39 - 0.14 = 0.25$. Donc on a $\frac{R_\Theta}{s^2_X} \cdot \frac{1}{\frac{1}{a} - \frac{1}{b}} = \frac{0.25}{0.22} \cdot \frac{1}{\frac{1}{28} - \frac{1}{79.5}} = 1.14 \times 43.22 = 49.11$ d'où $n = 50$, donc la taille de l'échantillon est suffisante.
   (c) En comparant les intervalles, que vous venez de calculer, avec la moyenne et la variance de fleurs de la 3e classe (cf. table 4), quelle est la conclusion que vous pouvez établir, concernant le pouvoir de séparation de cette variable pour les deux classes ? Justifier votre réponse.
   SOL.- La moyenne de la 3e classe est égale à 5.55, donc à l'extérieur de l'I.D.C. de la moyenne de la 2e classe. La variance non biaisée de la 3e classe est égale à 0.30, c'-à-d. presque 1.5 fois la variance ($= 0.22$) de la classe 2. Donc, bien que les moyennes de deux classes sont bien séparées, du fait que variance de la 3e classe est plus importante, il y a risque que des fleurs de la 3e classe soient considérées comme appartenant à la 2e classe au vu de la la valeur de la mesure de la longueur des pétales.
4. Nous allons essayer de consolider la conclusion précédente. Pour cela, on doit établir si, avec un risque $\alpha = 0.01$, il y a une différence significative pour la longueur de pétales entre la 2e et la 3e classe des iris. Effectuer les calculs et donner, en la justifiant, votre conclusion.
   SOL.- Comparaison de deux produits. Test bilatéral, sur les moyennes des populations normales. Variances inconnues et inégales. On applique l'algorithme

(1) $H_0 : \mu_x - \mu_y = 0$, $H_1 : \mu_x \neq \mu_y$, où on a noté X la variable aléatoire qui représente la longueur des pétales des fleurs de la 2e classe et Y celle des fleurs de la 3e classe.

(2) $\alpha = 0.01$, $n_x = n_y = 50$

(3) Calcul de la taille de l'échantillon

$$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n+1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m+1}} - 2 = \frac{\left(\frac{0.22}{50} + \frac{0.3136}{50}\right)^2}{\frac{\left(\frac{0.22}{50}\right)^2}{51} + \frac{\left(\frac{0.3136}{50}\right)^2}{51}} - 2 = \frac{0.0001096}{0.0000012} - 2 = 94.95 - 2 = 92.95$$

d'où $\nu = 93 < 100$ et la taille de l'échantillon est suffisante.

(4) Calcul de la valeur du critère

$$C = T_{\alpha\nu}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} = 2.67 \cdot 0.125 = 0.33$$

(5) Décision $\overline{X} - \overline{Y} = 5.55 - 4.26 = 1.29 > 0.33$, l'hypothèse H_0 est rejetée.

On conclut que les moyennes de deux classes sont différentes avec un risque de α = 1%, donc la conclusion précédente est confirmée.

5. Nous voulons maintenant étudier l'influence de la variable "longueur des pétales" sur la séparation en classes des iris en comparant les moyennes de cette variable dans chacune des classes à l'aide de l'analyse de variance.

   (a) Indiquer, de façon schématique, le tableau de données à utiliser pour faire l'analyse de variance. De façon schématique veut dire qu'il s'agit d'un tableau à double entrée et vous indiquerez ce que représentent ses lignes et ses colonnes, comme suit

	colonne 1	$\cdots$
ligne 1	$\cdots$	
$\cdots$		

SOL.- Un facteur contrôlé. Le tableau est formé comme suit :

	observation 1	observation 2	$\cdots$
Classe 1	$\cdots$	$\cdots$	$\cdots$
Classe 2	$\cdots$	$\cdots$	$\cdots$
Classe 3	$\cdots$	$\cdots$	$\cdots$

   (b) Vérifier, en établissant le tableau de l'analyse de variance, si les moyennes de la variable "longueur des pétales" dans les trois classes sont équivalentes avec une confiance de 99%.

On donne : $\overline{x}_T = 3.13, SS_T = 523.17, SS_A = 496.11$

SOL.- Hypothèse $H_0 : \mu_1 = \mu_2 = \mu_3$, les trois moyennes sont équivalentes.

On a : nombre de modalités $p = 50$, $n_T = 150, n_1 = n_2 = n_3 = 50$. De la table 4 on obtient $\overline{x}_1 = 1.46, \overline{x}_2 = 4.26, \overline{x}_3 = 5.55$. En faisant les calculs, on obtient la table 5b :

On a $F_A = F_A = \frac{SS_A\,/\,\nu_A}{SS_R\,/\,\nu_R} = 37.49$ et $F_{\alpha;\nu_A,\nu_R} = F_{0.01;49,100} = 1.83$.

Décision : On rejette H_0 puisque $F_A > F_{\alpha;\nu_A,\nu_R}$. Sur la base des observations, on décide que les moyennes de trois classes pour la longueur des pétales sont différentes et donc la longueur des pétales d'une fleur dépend de la classe de la fleur.

Tableau de l'analyse de variance		
Variance	*Somme des carrés*	*Degrés de liberté*
Facteur A	$SS_A = 496.11 \simeq 496$	$\nu_A = p - 1 = 49$
Résiduelle	$SS_R = SS_T - SS_A \simeq 27$	$\nu_R = n_T - p = 100$
Totale	$SS_T = 523.17 \simeq 523$	$\nu_T = n_T - 1 = 149$

(c) Donnez vos commentaires pour le résultat de l'analyse de la variance. Est-il possible de corroborer ce résultat par des analyses supplémentaires et lesquelles ?

SOL.- Les résultats indiquent que les moyennes sont différentes. Néanmoins la remarque de nous incite à la prudence. En effet, on sait que si un seul couple de classes parmi le trois possibles, ne valide pas H_0 pour que cette hypothèse soit rejetée. Il faut donc examiner plus en détail la situation en effectuant une analyse des contrastes.

6. Nous voulons, en utilisant la régression multilinéaire sur la totalité de la base de données, établir une relation linéaire entre la valeur de la variable “longueur des pétales” et les autres variables.

   Le listing de l'annexe 1 (à la fin du sujet) fournit les résultats de toutes les régressions possibles.

   Établir, en la justifiant, la meilleure relation linéaire.

   SOL.- Algorithme pas à pas. Avec une variable explicative, la meilleure régression est en fonction de la largeur de pétales (LAPE). Avec deux variables explicatives la meilleure régression, qui améliore la précédente, est obtenue avec les variables LAPE et LOSE. Ensuite avec les trois variables on a encore une amélioration de la régression.

7. En utilisant la totalité du tableau, on procède à une analyse en composantes principales dont les résultats sont donnés à l'annexe 2.

   Commenter les graphiques obtenus et en particulier la formation des axes dans le graphique des variables.

   SOL.- Le graphique des variables indique qu'il y a une grande corrélation entre LOPE et LAPE en premier lieu, et aussi entre ces deux variables et LOSE en deuxième lieu. De plus, Le premier axe est formé par l'opposition entre LASE d'un côté et les autres variables de l'autre côté.

   Le graphique des points montre une bonne séparation entre les iris de la classe 1 et les iris de deux autres classes.

ANNEXE 1
RÉSULTATS DES TOUTES LES RÉGRESSIONS

```
==================================================================================

ANALYSE STATISTIQUE DU TABLEAU DES DONNEES

Nombre d'observations utilisees    150
Nombre de variables utilisees        4

===================================================================================================================
    IRIS  :  Analyse n° 1
===================================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        2
Variable \'a expliquer              LOPE
   1e variable explicative       LOSE

-------------------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       352.8662 Carres moyens ....       352.8662
Somme des carres des ecarts SCE ...........       111.4592 Carres moyens ....         0.7531
-------------------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......          0.8718
Carre du coef de correlation multiple R^2 ..          0.7600
Valeur du test F de signification de R^2 ...        468.5502   \'a    1    et  148    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................         0.7431
             Expliquee ....................         2.3524
             Totale .......................         3.0955
-------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle       Erreur      Test      Proba   Test t  148 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LOSE       1.8584  [    1.689    2.028]      0.0859    468.5502  0.01      21.6460         0.01
Terme constant      -7.101443
-------------------------------------------------------------------------------------------------------------------

===================================================================================================================
    IRIS  :  Analyse n° 2
===================================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        2
Variable \'a expliquer              LOPE
   1e variable explicative       LASE

-------------------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE
```

```
Variation totale ..........................       464.3254
Somme des carres des estimations ..........        85.2320 Carres moyens ....        85.2320
Somme des carres des ecarts SCE ...........       379.0934 Carres moyens ....         2.5614
------------------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......         0.4284
Carre du coef de correlation multiple R^2 ..         0.1836
Valeur du test F de signification de R^2 ...        33.2750   \'a    1    et  148    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................        2.5273
             Expliquee ....................        0.5682
             Totale .......................        3.0955
------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  148 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LASE      -1.7352  [   -2.330   -1.141]      0.3008     33.2750  0.01       5.7684         0.01
Terme constant       9.063151
------------------------------------------------------------------------------------------------------------------

==================================================================================================================
    IRIS  :  Analyse n° 3
==================================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        2
Variable \'a expliquer              LOPE
   1e variable explicative       LAPE

------------------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       430.4806 Carres moyens ....       430.4806
Somme des carres des ecarts SCE ...........        33.8448 Carres moyens ....         0.2287
------------------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......         0.9629
Carre du coef de correlation multiple R^2 ..         0.9271
Valeur du test F de signification de R^2 ...      1882.4524   \'a    1    et  148    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................        0.2256
             Expliquee ....................        2.8699
             Totale .......................        3.0955
------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  148 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LAPE       2.2299  [    2.128    2.332]      0.0514   1882.4524  0.01      43.3872         0.01
Terme constant       1.083558
------------------------------------------------------------------------------------------------------------------

==================================================================================================================
    IRIS  :  Analyse n° 4
```

```
==========================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        3
Variable \'a expliquer              LOPE
   1e variable explicative       LOSE
   2e variable explicative       LASE

----------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       402.8887 Carres moyens ....       201.4443
Somme des carres des ecarts SCE ...........        61.4367 Carres moyens ....         0.4179
----------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......         0.9315
Carre du coef de correlation multiple R^2 ..         0.8677
Valeur du test F de signification de R^2 ...       481.9968   \'a    2    et  147    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................        0.4096
             Expliquee ....................        2.6859
             Totale .......................        3.0955
----------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  147 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LOSE       1.7756  [    1.648    1.903]      0.0644    760.0586  0.01      27.5692         0.01
LASE      -1.3386  [   -1.580   -1.097]      0.1224    119.6889  0.01      10.9402         0.01
Terme constant      -2.524762
----------------------------------------------------------------------------------------------------------

==========================================================================================================
    IRIS  :  Analyse n° 5
==========================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        3
Variable \'a expliquer              LOPE
   1e variable explicative       LOSE
   2e variable explicative       LAPE

----------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       440.4236 Carres moyens ....       220.2118
Somme des carres des ecarts SCE ...........        23.9018 Carres moyens ....         0.1626
----------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......         0.9739
```

```
Carre du coef de correlation multiple R^2 ..          0.9485
Valeur du test F de signification de R^2 ...       1354.3397   \'a    2    et  147    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................         0.1593
             Expliquee ....................         2.9362
             Totale .......................         3.0955
-------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  147 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LOSE       0.5423  [    0.405    0.679]      0.0693     61.1509  0.01       7.8199          0.01
LAPE       1.7481  [    1.599    1.897]      0.0753    538.4926  0.01      23.2054          0.01
Terme constant      -1.507138
-------------------------------------------------------------------------------------------------------------------

===================================================================================================================
    IRIS  :  Analyse n° 6
===================================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        3
Variable \'a expliquer               LOPE
   1e variable explicative       LASE
   2e variable explicative       LAPE

-------------------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       433.5704 Carres moyens ....       216.7852
Somme des carres des ecarts SCE ...........        30.7550 Carres moyens ....         0.2092
-------------------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......          0.9663
Carre du coef de correlation multiple R^2 ..          0.9338
Valeur du test F de signification de R^2 ...       1036.1716   \'a    2    et  147    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................         0.2050
             Expliquee ....................         2.8905
             Totale .......................         3.0955
-------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  147 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard   partiel F  d'erreur non nullite    d'erreur
LASE      -0.3550  [   -0.538   -0.172]      0.0924     14.7683  0.01       3.8430          0.01
LAPE       2.1556  [    2.051    2.260]      0.0528   1664.9585  0.01      40.8039          0.01
Terme constant       2.258164
-------------------------------------------------------------------------------------------------------------------

===================================================================================================================
    IRIS  :  Analyse n° 7
===================================================================================================================

REGRESSION MULTIPLE

Nombre d'observations utilisees    150
Nombre de variables utilisees        4
```

```
Variable \'a expliquer              LOPE
   1e variable explicative      LOSE
   2e variable explicative      LASE
   3e variable explicative      LAPE

-------------------------------------------------------------------------------------------------------------------

RESULTATS DE LA REGRESSION MULTILINEAIRE

                ANALYSE DE LA VARIANCE

Variation totale ..........................       464.3254
Somme des carres des estimations ..........       449.4725 Carres moyens ....       149.8242
Somme des carres des ecarts SCE ...........        14.8529 Carres moyens ....         0.1017
-------------------------------------------------------------------------------------------------------------------

   MESURES GLOBALES DE LA QUALITE DE LA REGRESSION

Coefficient de correlation multiple R ......         0.9839
Carre du coef de correlation multiple R^2 ..         0.9680
Valeur du test F de signification de R^2 ...       1472.7262   \'a    3    et  146    ddl   Proba d' erreur   0.01
Variances :  residuelle ...................        0.0990
             Expliquee ....................        2.9965
             Totale .......................        3.0955
-------------------------------------------------------------------------------------------------------------------

       COEFFICIENTS DE LA REGRESSION

                         Intervalle        Erreur      Test      Proba   Test t  146 ddl  Proba
Var       Coef      de confiance \'a 0.95   standard    partiel F  d'erreur non nullite    d'erreur
LOSE       0.7291  [    0.614    0.844]      0.0583    156.3121  0.01      12.5025          0.01
LASE      -0.6460  [   -0.781   -0.511]      0.0685     88.9473  0.01       9.4312          0.01
LAPE       1.4468  [    1.313    1.580]      0.0676    457.9047  0.01      21.3987          0.01
Terme constant      -0.262711
-------------------------------------------------------------------------------------------------------------------
```

ANNEXE 2

ANALYSE EN COMPOSANTES PRINCIPALES

IRIS

```
================================================================
RESULTATS
```

Valeurs propres

Axes	Val. propres	Information en pc
axe no 1	2.9185	72.9624
axe no 2	0.9140	22.8508
axe no 3	0.1468	3.6689
axe no 4	0.0207	0.5179

Correlations entre les variables et les axes principaux

	Axe 1	Axe 2	Axe 3	Axe 4
LOSE	0.8902	-0.3608	-0.2757	0.0376
LASE	-0.4601	-0.8827	0.0936	-0.0178
LOPE	0.9916	-0.0234	0.0544	-0.1153
LAPE	0.9650	-0.0640	0.2430	0.0754

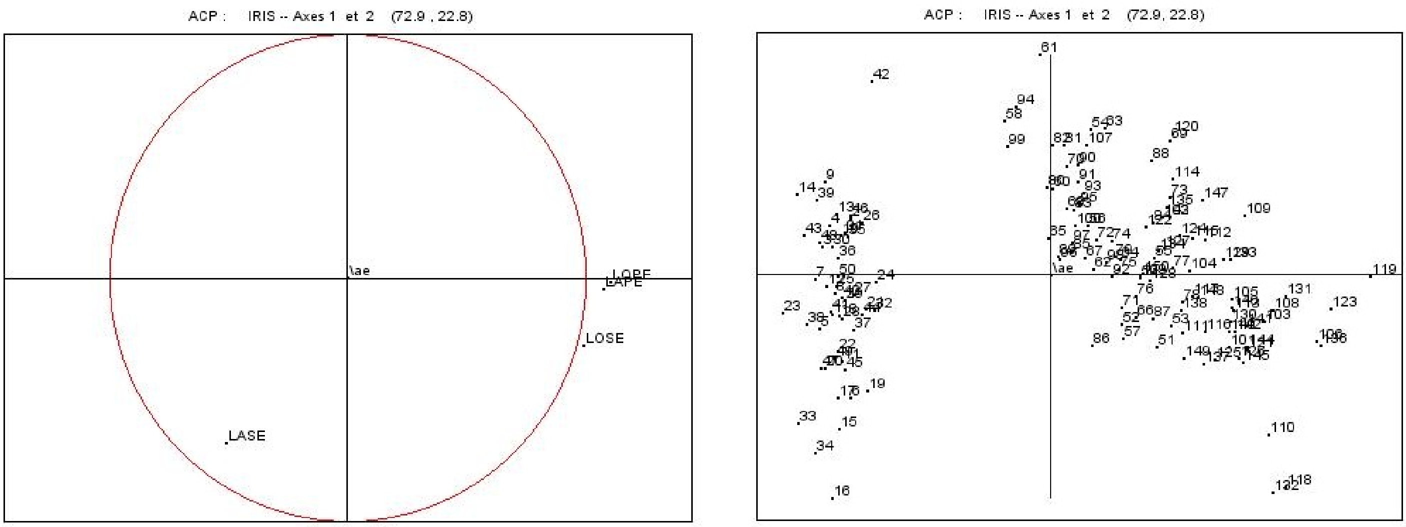

FIGURE 1 – Projection du nuage des variables et des points sur les deux premiers axes principaux de l'ACP.

ANNEXE 3
ANALYSE FACTORIELLE DES CORRESPONDANCES

```
    IRIS
===============================================================
RESULTATS

Valeurs propres
      Axes              Val. propres       Information en pc
    axe no 1               0.0613            95.5854
    axe no 2               0.0022              3.4317
    axe no 3               0.0006              0.9829

Tableau des variables reconstituees
1e ligne : Coordonnees factorielles
2e ligne : Contributions absolues (cosinus carre, caracteristique de la qualite de la representation)
3e ligne : Contributions relatives (\‘a l inertie expliquee par l axe)

           Axe 1       Axe 2       Axe 3
LOSE      0.0252      0.0014      0.0006
          0.0044      0.0004      0.0002
          0.0006      0.0000      0.0000
LASE      0.0762     -0.0023     -0.0006
          0.0209      0.0005      0.0001
          0.0053      0.0000      0.0000
LOPE     -0.0672      0.0013     -0.0007
          0.0200      0.0002      0.0002
          0.0042      0.0000      0.0000
LAPE     -0.1067     -0.0052      0.0009
          0.0161      0.0010      0.0001
          0.0095      0.0000      0.0000
```

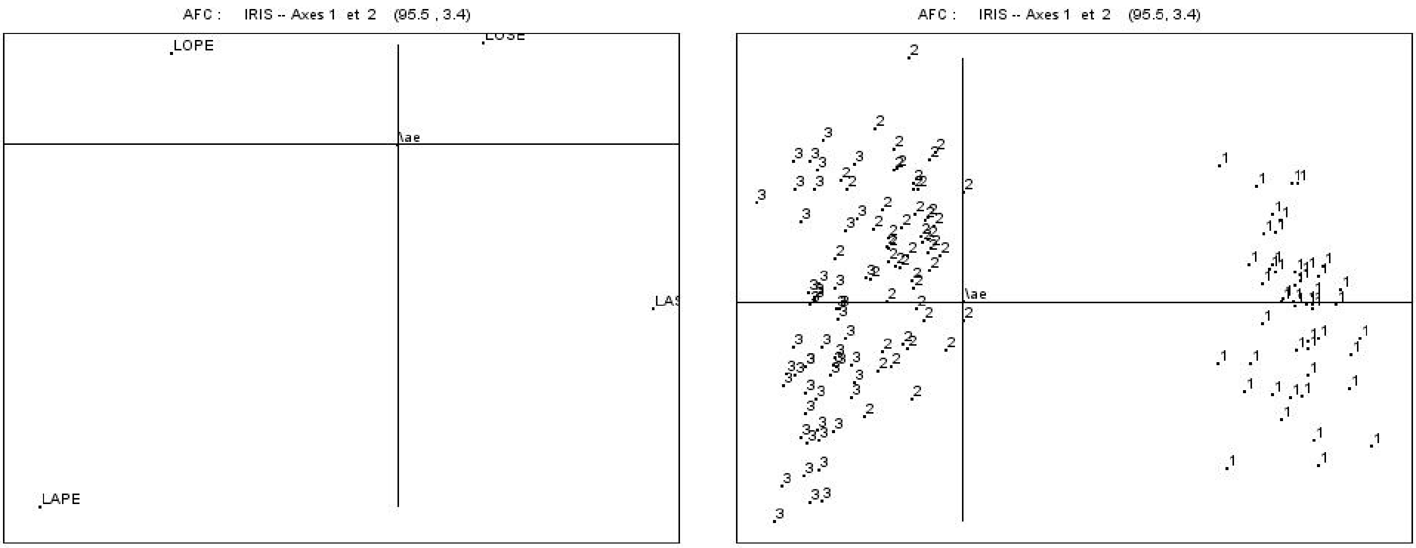

FIGURE 2 – Projection du nuage des variables et des points (suivant les numéros de classe) sur les deux premiers axes principaux de l’AFC.