

## Corrélation. Droites de régression

### Exercice 37 Données Budget

Il s'agit d'un extrait d'une enquête (ONU 1967) sur les budgets-temps (temps passé dans différentes activités au cours de la journée).

Les colonnes comprennent 3 variables numériques, le temps passé en : Profession (PROF), Transport (TRAN) et loisirs (LOIS). Les temps sont notés en centièmes d'heures. Le code suivant est utilisé pour identifier les lignes :

H : hommes, F : femmes, A : actifs, N : non actifs, M : mariés, C : célibataires,

U : USA, W : pays de l'ouest, E : Est sauf Yougoslavie, Y : Yougoslavie.

Budget	PROF	TRAN	LOIS	Budget	PROF	TRAN	LOIS
HAU	610	140	315	FAY	560	105	235
FAU	475	90	305	FNY	10	10	380
FNU	10	0	430	HMY	650	145	358
HMU	615	140	305	FMY	260	52	295
FMU	179	29	373	HCY	615	125	475
HCU	585	115	385	FCY	433	89	408
FCU	482	94	336	HAE	650	142	334
HAW	653	100	330	FAE	578	106	228
FAW	511	70	262	FNE	24	8	398
FNW	20	7	368	HME	652	133	310
HMW	656	97	321	FME	436	79	231
FMW	168	22	311	HCE	627	148	463
HCW	643	105	388	FCE	434	86	380
FCW	429	34	392	Moy	451	86	346
HAY	650	140	365	Ety	223	47	63

1) Représenter le nuage de points correspondant aux variables PROF et TRAN, puis celui correspondant aux variables PROF et LOIS.

2) Calculer la covariance et le coefficient de corrélation pour le couple de variables (PROF, TRAN), puis pour le couple (PROF, LOIS). Dans chacun des deux cas, la corrélation est-elle significative ?

3) Déterminer l'équation de la droite de régression de TRAN selon les valeurs de PROF. Quelle est la part de la variance de TRAN qui est "expliquée" par PROF ?

Réponses : 2)  $Cov(PROF, TRAN) = 9805.12$ ,  $r(PROF, TRAN) = 0.93$ ;  $Cov(PROF, LOIS) = -2651.87$ ,  $r(PROF, LOIS) = -0.19$ . Seule la corrélation entre PROF et TRAN est significative. L'équation de la droite de régression est :  $TRAN = 0.1977 PROF - 3.15$ . La part de la variance de TRAN "expliquée par" PROF est de  $\frac{Var(\widehat{TRAN})}{Var(TRAN)} = r^2 = 0.87$ .

### Exercice 38

Quinze élèves, désignés par les lettres de A à O ont été classés une première fois par une épreuve de français, une seconde fois par une épreuve de mathématiques. Calculer le coefficient décrivant la corrélation entre ces deux classements.

Elèves	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Fran.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Math.	9	3	1	11	2	5	8	13	4	10	7	14	15	6	12

Réponses :  $Cov(F, M) = 9.47$  ;  $\rho = 0.51$ . La corrélation est à peine significative à 5%. Remarquez qu'il s'agit ici d'un coefficient de corrélation des rangs. On pourra consulter le paragraphe Corrélation des rangs de Spearman d'un ouvrage de statistiques. Le coefficient de corrélation peut égelement être obtenu à l'aide de la formule :

$$\rho = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)}$$

dans laquelle  $N$  est le nombre de sujets et  $d_i$  est la différence entre le rang obtenu sur la première variable et celui obtenu sur la seconde.

**Exercice 39**

On mène une étude sur les variations circadiennes de la charge mentale induite par une tâche simple et répétitive. (circadien signifie "sur un cycle de 24 heures").

On considère un échantillon homogène de sujets et on relève, à différents moments de la journée :

- la vitesse d'exécution d'une tâche répétitive simple (nombre d'appuis sur un bouton par minute)
- l'indice de charge mentale induite (mesuré à partir du temps de réaction à un stimulus auditif simple).

On obtient les résultats suivants (moyennes obtenues sur l'ensemble des sujets observés).

Moment	8	10	12	14	16	18	20	22	24
Vitesse	64,54	66,61	71,01	70,10	70,08	68,42	66,63	64,12	63,12
Indice	1,117	1,130	1,171	1,140	1,141	1,129	1,107	1,072	1,052

1) Construire un nuage de points en plaçant en abscisse la variable "moment de la journée", en ordonnée la vitesse, et en choisissant judicieusement les unités.

D'après ce graphique :

- semble-t-il exister une relation entre le moment de la journée et la vitesse ?
- serait-il pertinent de calculer un coefficient de corrélation linéaire pour évaluer l'intensité de cette relation ?

2) Mêmes questions pour les variables vitesse et indice de charge mentale.

3) Calculer la covariance et le coefficient de corrélation linéaire entre les variables vitesse ( $x_i$ ) et indice de charge mentale ( $y_i$ ). La corrélation est-elle significative au seuil de 1% ? Quelle est la part de la variance des  $y_i$  qui est "expliquée" par celle des  $x_i$  ?

*Vu la faible amplitude des variations de l'indice, on aura soin de garder un nombre suffisant de décimales dans les calculs intermédiaires. On utilisera par ailleurs les résultats intermédiaires suivants :*

$$\sum x_i = 604,63 ; \sum y_i = 10,059 ; \sum x_i^2 = 40686,3023, \sum y_i^2 = 11,253289 ; \sum x_i y_i = 676,53294$$

4) Déterminer une équation de la droite de régression de l'indice de charge mentale en fonction de la vitesse. Construire cette droite sur le graphique précédent.

Réponses : 1 et 2) Il semble exister une relation entre le moment de la journée et la vitesse, mais cette relation n'est pas linéaire, et ne peut donc pas être étudiée à l'aide d'un coefficient de corrélation. En revanche, il semble exister une relation linéaire entre la vitesse et l'indice de charge mentale.

3)  $Cov(x_i, y_i) = \frac{676.53294}{9} - \frac{604.63}{9} \times \frac{10.059}{9} = 0.084$ .  $Var(x_i) = \frac{40686.3023}{9} - (\frac{604.63}{9})^2 = 7.39$ .

$\sigma_x=2.72$ . De même,  $\sigma_y = 0.0344$  et finalement,  $\rho = \frac{0.084}{2.72 \times 0.0344} = 0.899$ . Il existe donc une forte corrélation positive entre ces deux variables.

4) Equation de la droite de régression :  $y=0.0114 x + 0.354$

#### Exercice 40 Données Tailles

Le tableau ci-dessous donne la taille de 10 garçons (variable Z) ainsi que la taille de leurs parents (le père X et la mère Y).

	X	Y	Z
i1	160.0	161	165.0
i2	165.0	155	162.5
i3	170.0	155	165.0
i4	172.5	165	175.0
i5	175.0	170	180.0
i6	180.0	166	177.5
i7	185.0	167	180.0
i8	187.5	172	190.0
i9	190.0	175	195.0
i10	195.0	168	187.5

On cherche s'il existe une relation entre la taille du fils et celle de ses parents et, si oui, quelle est la part respective de la mère et du père. Pour cela, on procède à la régression de Z sur X et Y. On donne les résultats intermédiaires suivants :

$$\sum X_i = 1780; \sum Y_i = 1654; \sum Z_i = 1777,5$$

$$\sum X_i^2 = 318012.5, \sum Y_i^2 = 273974; \sum Z_i^2 = 317068.75$$

$$\sum X_i Y_i = 294932.5; \sum X_i Z_i = 317437.5; \sum Y_i Z_i = 294632.5$$

$$Var(X) = 117.25; Var(Y) = 40.24; Var(Z) = 111.81$$

$$Cov(X, Y) = 52.05; Cov(X, Z) = 104.25; Cov(Y, Z) = 63.40.$$

1) Quels sont les coefficients de corrélation des variables prises deux à deux ?

2) On utilise un logiciel de traitement statistique pour déterminer l'équation du plan de régression de Z par rapport à X et Y. On obtient :

$$Z = 0.4455X + 0.9993Y - 66.83.$$

Calculer les valeurs estimées de Z pour chacun des 10 individus statistiques (variable  $\hat{Z}$ ).

On donne par ailleurs :  $\sum \hat{Z}_i^2 = 317060.06$  et  $\sum Z_i \hat{Z}_i = 317054.34$ .

3) Déterminer le coefficient de corrélation multiple.

4) Quelle est la proportion de variance prise en compte par la régression ?

5) Les coefficients de corrélation partielle sont donnés par :  $R_{xz,y} = 0.91$ ;  $R_{yz,x} = 0.95$

Quel est, de la taille du père et de celle de la mère, le meilleur prédicteur de la taille du fils ?

6) Prédire la taille d'un garçon, sachant que son père mesure 188cm et sa mère 171cm.

Réponses : N.B. Calculs exécutés à l'aide d'un logiciel de traitement statistique.

1) Les coefficients de corrélation des variables prises deux à deux sont donnés par :  $r(X, Y)=0.76$ ;  $r(X, Z)=0.91$ ;  $r(Y, Z)=0.95$ .

2) Les valeurs estimées de Z sont données par :

i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
165.34	161.57	163.8	174.90	181.01	179.24	182.47	188.58	192.69	187.92

3) Coefficient de corrélation multiple :  $R=0.991$ .

4) D'où  $R^2 = 0.98$ . Le modèle explique 98% de la variance observée de la variable Z. Cette proportion est très élevée, mais il s'agit de données fictives....

5) Le coefficient de corrélation partielle le plus élevé est celui liant taille de la mère et taille du fils.

6) Taille du fils si  $X=188$  et  $Y=171$  :  $Z = 188$ .

**Exercice 41** Données Evalcour

L'association des étudiants d'une grande université (américaine) a publié une évaluation de plus de cent cours enseignés durant le semestre précédent. Les étudiants de chaque cours avaient rempli un questionnaire d'évaluation portant sur différents aspects du cours ; l'évaluation se faisait sur une échelle en cinq points (1=très mauvais, 5=excellent). Les données figurant dans les deux tableaux 9 et 10 pages 59 et 60 sont les données réelles. Elles représentent les scores moyens enregistrés sur 6 variables pour un échantillon de 50 cours. Ces variables étaient :

Qual-Glob	Pédagogie	Examen	Connaissan	Résultat	Inscriptio
3.4	3.8	3.8	4.5	3.5	21
2.9	2.8	3.2	3.8	3.2	50
2.6	2.2	1.9	3.9	2.8	800
3.8	3.5	3.5	4.1	3.3	221
3	3.2	2.8	3.5	3.2	7
2.5	2.7	3.8	4.2	3.2	108
3.9	4.1	3.8	4.5	3.6	54
4.3	4.2	4.1	4.7	4	99
3.8	3.7	3.6	4.1	3	51
3.4	3.7	3.6	4.1	3.1	47
2.8	3.3	3.5	3.9	3	73
2.9	3.3	3.3	3.9	3.3	25
4.1	4.1	3.6	4	3.2	37
2.7	3.1	3.8	4.1	3.4	83
3.9	2.9	3.8	4.5	3.7	70
4.1	4.5	4.2	4.5	3.8	16
4.2	4.3	4.1	4.5	3.8	14
3.1	3.7	4	4.5	3.7	12
4.1	4.2	4.3	4.7	4.2	20
3.6	4	4.2	4	3.8	18
4.3	3.7	4	4.5	3.3	260
4	4	4.1	4.6	3.2	100
2.1	2.9	2.7	3.7	3.1	118
3.8	4	4.4	4.1	3.9	35
2.7	3.3	4.4	3.6	4.3	32

TAB. 9 – Première partie des données

1. la qualité globale des exposés (Qual-Glob)
2. les aptitudes pédagogiques du professeur (Pédagogie)

Qual-Glob	Pédagogie	Examen	Connaissan	Résultat	Inscriptio
4.4	4.4	4.3	4.4	2.9	25
3.1	3.4	3.6	3.3	3.2	55
3.6	3.8	4.1	3.8	3.5	28
3.9	3.7	4.2	4.2	3.3	28
2.9	3.1	3.6	3.8	3.2	27
3.7	3.8	4.4	4	4.1	25
2.8	3.2	3.4	3.1	3.5	50
3.3	3.5	3.2	4.4	3.6	76
3.7	3.8	3.7	4.3	3.7	28
4.2	4.4	4.3	5	3.3	85
2.9	3.7	4.1	4.2	3.6	75
3.9	4	3.7	4.5	3.5	90
3.5	3.4	4	4.5	3.4	94
3.8	3.2	3.6	4.7	3	65
4	3.8	4	4.3	3.4	100
3.1	3.7	3.7	4	3.7	105
4.2	4.3	4.2	4.2	3.8	70
3	3.4	4.2	3.8	3.7	49
4.8	4	4.1	4.9	3.7	64
3	3.1	3.2	3.7	3.3	700
4.4	4.5	4.5	4.6	4	27
4.4	4.8	4.3	4.3	3.6	15
3.4	3.4	3.6	3.5	3.3	40
4	4.2	4	4.4	4.1	18
3.5	3.4	3.9	4.4	3.3	90

TAB. 10 – Seconde partie des données

3. la qualité des tests et examens (Examen)
4. la connaissance de la matière dont témoigne le professeur, telle qu'elle est perçue par les étudiants (Connaissan)
5. les résultats auxquels s'attendent les étudiants pour ce cours (Résultat, de très bon à insuffisant)
6. le nombre d'inscriptions à ce cours (Inscriptio)

Les résultats de statistiques descriptives concernant les variables précédentes sont donnés dans le tableau 11.

Les coefficients de corrélation des variables prises deux à deux sont donnés dans le tableau 12.

Les coefficients de l'équation de régression multiple de la première variable en fonction des cinq autres sont donnés par le tableau 13.

Ecrire l'équation de régression correspondante, et la vérifier sur l'extrait donné dans le tableau 14.

Enfin, le dernier tableau (tableau 15) donne les coefficients de corrélation partiels entre la variable Qual-Glob et les prédicteurs.

	Qual-Glob	Pedagogie	Examen	Connaissan	Resultat	Inscriptio
Effectif	50	50	50	50	50	50
Moyenne	3.55	3.664	3.808	4.176	3.486	88.0
Variance	0.376429	0.283167	0.2432	0.166351	0.123269	21042.2
Ecart-type	0.613538	0.532135	0.493153	0.407862	0.351097	145.059

TAB. 11 – Statistiques descriptives sur les données

	Qual-Glob	Pedagogie	Examen	Connaissan	Resultat	Inscriptio
Qual-Glob	1	0.8039	0.5956	0.6818	0.3008	-0.2396
Pedagogie	0.8039	1	0.7197	0.5263	0.4691	-0.4511
Examen	0.5956	0.7197	1	0.4515	0.6100	-0.5581
Connaissan	0.6818	0.5263	0.4515	1	0.2242	-0.1279
Resultat	0.3008	0.4691	0.6100	0.2242	1	-0.3371
Inscriptio	-0.2396	-0.4511	-0.5581	-0.1279	-0.3371	1

TAB. 12 – Coefficients de corrélation

Paramètre	Estimation
CONSTANTE	-1.19483
Inscriptio	0.000525491
Examen	0.131981
Resultat	-0.184308
Connaissan	0.488984
Pedagogie	0.763237

TAB. 13 – Coefficients de l'équation de de régression

Ligne	Observé	Ajusté	Résidu
1	3.4	3.77339	-0.373387
2	2.9	2.6592	0.240795
3	2.6	2.54643	0.0535724
4	3.8	3.45119	0.348812
5	3.0	2.74242	0.257584

TAB. 14 – Comparaison des valeurs observées et des valeurs ajustées

Régresseur	Coef
Pedagogie	0,6544
Examen	0,1213
Connaissan	0,4751
Resultat	-0,1656
Inscriptio	0,1990

TAB. 15 – Coefficients de corrélation partielle

La valeur du coefficient de corrélation multiple vérifie :  $R^2 = 0.755$ .

### Exercice 42

Aux élections européennes de juin 1984, les votes pour la liste du Front National ont été très variables dans l'espace et leur comparaison avec d'autres variables fait apparaître un certain nombre de relations. Les variables choisies dans le tableau 16 sont les suivantes :

- LPEN : pourcentage de voix de la liste FN ;
- ETRA : pourcentage d'étrangers dans la population en 1982 ;
- DELI : Nombre pondéré de délinquance pour 100 habitants en 1980 ;
- CRCH : Taux mensuel moyen d'évolution du chômage entre le 31.08.81 et le 30.04.83 ;
- TXCH : Pourcentage de chômeurs dans la population active au 30.09.83 ;
- URBA : Pourcentage de population urbaine en 1982.

Les individus statistiques sont ici les régions de France Métropolitaine (ILEF=Ile de France, CHAM=Champagne-Ardenne, etc.). L'échelle régionale n'est certainement pas la meilleure pour une telle étude et les conclusions valent pour les agrégats spatiaux et non des personnes.<sup>1</sup>

REG	LEPEN	ETRA	DELI	CRCH	TXCH	URBA
ILEF	14.5	13.3	6	0.23	7.1	93.6
CHAM	10.7	5.4	4	0.07	9.5	62.4
PICA	10.8	4.6	4	0.22	9.7	60.7
HNOR	8.9	3.3	4	0.01	11	69.1
CENT	9.3	5.1	3	0.51	7.8	62.9
BNOR	7.6	1.7	4	0.38	9.8	53.4
BOUR	10.1	5.4	3	0.72	8.6	57.9
NORD	9.1	4.8	4	0.21	11.8	86.4
LORR	12.4	8	4	0.51	9.2	72.4
ALSA	12.5	8.1	3	1.25	7.4	73.2
FCOM	12	7.4	4	0.19	8.2	58.8
PAYS	6.8	1.4	3	0.58	9.6	60.1
BRET	6.8	0.7	3	0.84	9.4	55.6
POIT	6.7	1.7	3	0.48	10	50.5
AQUI	8.3	4.6	4	0.85	9.5	64.6
MIDI	8.1	4.8	3	0.54	8.5	59.3
LIMO	4.8	2.7	3	0.57	6.9	50.9
RHON	12.9	9.1	4	0.57	7.5	76.9
AUVE	7.4	4.6	2	0.85	8.3	58.2
LANG	13.2	6.5	4	1.44	11.4	70.7
PROV	19	8.2	6	1.13	10.5	89.6

TAB. 16 – Voix du FN aux élections européennes de 1984

Le tableau 17 donne les valeurs des coefficients de corrélation des variables prises deux à deux. On voit que les votes pour l'extrême droite sont fortement corrélés à trois variables : taux d'urbanisation (URBA), taux de délinquance (DELI) et taux d'étrangers

<sup>1</sup>D'après *Initiation aux méthodes statistiques en Géographie*, Groupe Chadule, Masson Ed., 1994

(ETRA). D'autre part, ces trois variables sont fortement corrélées entre elles, elles sont donc partiellement redondantes.

	LEPEN	ETRA	DELI	CRCH	TXCH	URBA
LEPEN	1	0.81	0.76	0.25	0.05	0.77
ETRA		1	0.62	0.08	-0.35	0.76
DELI			1	-0.14	0.19	0.75
CRCH				1	0.00	0.08
TXCH					1	0.13
URBA						1

TAB. 17 – Corrélations entre les variables

Une première régression multiple est réalisée en utilisant les 5 variables explicatives. Le coefficient de corrélation multiple vaut  $R = 0.934$  et le coefficient de détermination,  $R^2 = 0.872$

Les coefficients de corrélation partielle entre la variable LEPEN et chacune des autres variables sont alors ceux indiqués dans le tableau 18.

	ETRA	DELI	CRCH	TXCH	URBA
LEPEN	0.6910802	0.52652694	0.5494687	0.45550366	-0.2161422

TAB. 18 – Corrélation partielle entre LEPEN et les 5 variables

On peut tester la significativité de ces coefficients de corrélation. Le nombre de degrés de liberté à prendre en compte est  $21 - 6 = 15$ . Au seuil de 5%,  $r_{crit} = 0.4821$ .

On retire alors la variable qui a le plus faible coefficient de corrélation partielle, c'est-à-dire URBA et on réalise une régression multiple de la variable LPEN par rapport aux quatre variables explicatives restantes.

On trouve alors :  $R = 0.930$ ,  $R^2 = 0.865$  et les nouveaux coefficients de corrélation partielle indiqués dans le tableau 19. Notons que  $R$  ne change pratiquement pas : la variable URBA n'apporte pas d'information supplémentaire par rapport aux quatre variables restantes.

	ETRA	DELI	CRCH	TXCH
LEPEN	0.73354021	0.49258622	0.5282723	0.4116398

TAB. 19 – Corrélation partielle entre LEPEN et 4 variables

Pour tester la significativité de ces coefficients, on prend ici  $ddl = 16$  et donc  $r_{crit} = 0.4683$ . Retirons de même la variable qui a le plus faible coefficient de corrélation partielle, c'est-à-dire TXCH.

On trouve alors :  $R = 0.915$ ,  $R^2 = 0.838$  et les coefficients de corrélation partielle du tableau 20.

A ce stade,  $r_{crit} = 0.4683$ , et tous les coefficients sont significatifs. On peut donc dire que les votes pour l'extrême-droite aux élections européennes de 1984, à l'échelle régionale,

	ETRA	DELI	CRCH
LEPEN	0.6829056	0.6825633	0.5516189

TAB. 20 – Corrélation partielle entre LEPEN et 3 variables

ont varié en fonction de trois circonstances : le taux d'étrangers, le taux de délinquance, et à un degré moindre, l'évolution du chômage.

La régression n'a pas été faite dans un but de prévision, et l'équation de régression n'a qu'un intérêt limité :

$$LPEN = 0.52 \text{ ETRA} + 1.69 \text{ DELI} + 2.37 \text{ CRCH} - 0.4$$

## Analyse en composantes principales

### Exercice 43 Données Budget-temps

Il s'agit d'une enquête (ONU 1967) sur les budgets-temps (temps passé dans différentes activités au cours de la journée).

Le tableau suivant comprend 10 variables numériques et 4 variables catégorisées.

Les 10 variables numériques sont : le temps passé en : Profession, Transport, Ménage, Enfants, Courses, Toilette, Repas, Sommeil, Télé, Loisirs.

Les 4 variables catégorisées sont : Le sexe (1=Hommes 2=Femmes), l'activité (1=Actifs 2=Non Act. 9=Non précisé), l'état civil (1=Célibataires 2=Mariés 9=Non précisé), le Pays (1=USA 2=Pays de l'Ouest 3=Pays de l'Est 4=Yougoslavie).

Le code suivant est utilisé pour identifier les lignes : H : Hommes, F : Femmes, A : Actifs, N : Non Actifs(ves), M : Mariés, C : Célibataires, U : USA, W : Pays de l'Ouest sauf USA, E : Est sauf Yougoslavie, Y : Yougoslavie

Les temps sont notés en centièmes d'heures. La première case en haut à gauche du tableau (HAU) indique que les Hommes Actifs des USA passent en moyenne 6 heures et 6 minutes (6 heures + 10/100 d'heure, soit 6 heures et 6mn) en activité PROFESSIONNELLE. Le total d'une ligne (sur ces 10 variables numériques) est 2400 (24 heures).

	PROF	TRAN	MENA	ENFA	COUR	TOIL	REPA	SOMM	TELE	LOIS	SEX	ACT	CIV	PAY
HAU	610	140	60	10	120	95	115	760	175	315	1	1	9	1
FAU	475	90	250	30	140	120	100	775	115	305	2	1	9	1
FNU	10	0	495	110	170	110	130	785	160	430	2	2	9	1
HMU	615	140	65	10	115	90	115	765	180	305	1	9	2	1
FMU	179	29	421	87	161	112	119	776	143	373	2	9	2	1
HCU	585	115	50	0	150	105	100	760	150	385	1	9	1	1
FCU	482	94	196	18	141	130	96	775	132	336	2	9	1	1
HAW	653	100	95	7	57	85	150	808	115	330	1	1	9	2
FAW	511	70	307	30	80	95	142	816	87	262	2	1	9	2
FNW	20	7	568	87	112	90	180	843	125	368	2	2	9	2
HMW	656	97	97	10	52	85	152	808	122	321	1	9	2	2
FMW	168	22	528	69	102	83	174	824	119	311	2	9	2	2
HCW	643	105	72	0	62	77	140	813	100	388	1	9	1	2
FCW	429	34	262	14	92	97	147	849	84	392	2	9	1	2
HAY	650	140	120	15	85	90	105	760	70	365	1	1	9	4
FAY	560	105	375	45	90	90	95	745	60	235	2	1	9	4
FNY	10	10	710	55	145	85	130	815	60	380	2	2	9	4
HMY	650	145	112	15	85	90	105	760	80	358	1	9	2	4
FMY	260	52	576	59	116	85	117	775	65	295	2	9	2	4
HCY	615	125	95	0	115	90	85	760	40	475	1	9	1	4
FCY	433	89	318	23	112	96	102	774	45	408	2	9	1	4
HAE	650	142	122	22	76	94	100	764	96	334	1	1	9	3
FAE	578	106	338	42	106	94	92	752	64	228	2	1	9	3
FNE	24	8	594	72	158	92	128	840	86	398	2	2	9	3
HME	652	133	134	22	68	94	102	763	122	310	1	9	2	3
FME	436	79	433	60	119	90	107	772	73	231	2	9	2	3
HCE	627	148	68	0	88	92	86	770	58	463	1	9	1	3
FCE	434	86	297	21	129	102	94	799	58	380	2	9	1	3

Matrice des corrélations :

	PROF	TRAN	MENA	ENFA	COUR	TOIL	REPA	SOMM	TELE	LOIS
PROF	1.000	0.933	-0.908	-0.870	-0.658	-0.112	-0.455	-0.538	-0.059	-0.190
TRAN	0.933	1.000	-0.869	-0.809	-0.503	-0.079	-0.613	-0.702	-0.044	-0.105
MENA	-0.908	-0.869	1.000	0.861	0.501	-0.035	0.361	0.433	-0.206	-0.113
ENFA	-0.870	-0.809	0.861	1.000	0.543	0.124	0.367	0.277	0.122	-0.109
COUR	-0.658	-0.503	0.501	0.543	1.000	0.593	-0.184	-0.030	0.216	0.235
TOIL	-0.112	-0.079	-0.035	0.124	0.593	1.000	-0.360	-0.217	0.322	0.073
REPA	-0.455	-0.613	0.361	0.367	-0.184	-0.360	1.000	0.817	0.316	-0.040
SOMM	-0.538	-0.702	0.433	0.277	-0.030	-0.217	0.817	1.000	0.018	0.208
TELE	-0.059	-0.044	-0.206	0.122	0.216	0.322	0.316	0.018	1.000	-0.095
LOIS	-0.190	-0.105	-0.113	-0.109	0.235	0.073	-0.040	0.208	-0.095	1.000

Valeurs propres :

Valeur propre	4.5887	2.1198	1.3210	1.1953	0.4684	0.1990
Proportion	0.459	0.212	0.132	0.120	0.047	0.020
Cumulatif	0.459	0.671	0.803	0.922	0.969	0.989

Valeur propre	0.0468	0.0371	0.0239	0.0000
Proportion	0.005	0.004	0.002	0.000
Cumulatif	0.994	0.998	1.000	1.000

Scores des individus sur les quatre premières composantes principales :

	Comp.1	Comp.2	Comp.3	Comp.4
HAU	-1.77294	-0.68605	-1.8713	0.5752
FAU	-0.17159	-2.21532	-0.6608	0.4376
FNU	4.05340	-2.27771	-1.0605	-0.5203
HMU	-1.77937	-0.29267	-1.8851	0.7330
FMU	2.61426	-2.28530	-0.7972	0.1076
HCU	-1.50279	-1.89173	-1.3630	-0.7823
FCU	-0.46524	-2.84430	-1.2964	-0.1476
HAW	-1.17634	2.36768	-1.1166	-0.0458
FAW	0.31200	1.49528	-0.2724	0.9433
FNW	4.32338	1.63256	-0.8903	-0.1438
HMW	-1.12538	2.46392	-1.2856	0.1503
FMW	3.13128	1.98892	-0.5882	0.7346
HCW	-1.37003	2.57197	-0.5263	-1.0228
FCW	1.09911	1.65514	-0.5433	-1.4957
HAY	-2.16270	0.24105	0.7089	-0.2393
FAY	-1.00478	-0.18007	1.6157	2.1323
FNY	3.53743	0.37708	1.6353	-0.5295
HMY	-2.22120	0.21162	0.4832	-0.1096
FMY	1.53997	0.21607	1.6214	1.1658
HCY	-2.13527	-0.58062	1.6098	-2.1775
FCY	-0.33581	-0.41823	1.4906	-1.0249
HAE	-2.14681	0.06942	0.1312	0.3216
FAE	-0.98772	-0.58543	1.3654	2.0399
FNE	3.91870	-0.04945	0.6719	-0.9754
HME	-2.07739	0.17048	-0.4251	0.7923
FME	0.49043	-0.20397	1.1839	2.0125
HCE	-2.52945	-0.14681	1.0625	-1.9634
FCE	-0.05515	-0.80353	1.0024	-0.9678

Qualités des représentations des individus par les deux premières composantes principales ;  
Qualités des représentations des individus par le plan (CP1, CP2)

	Comp.1	Comp.2		QLT
HAU	0.36370	0.05446	HAU	0.4182
FAU	0.004113	0.685605	FAU	0.6897
FNU	0.6578	0.2077	FNU	0.8655
HMU	0.34597	0.00936	HMU	0.3553
FMU	0.5164	0.3946	FMU	0.911
HCU	0.2456	0.3893	HCU	0.6349
FCU	0.01697	0.63414	FCU	0.6511
HAW	0.1658	0.6717	HAW	0.8376
FAW	0.01998	0.45887	FAW	0.4789
FNW	0.8274	0.1180	FNW	0.9454
HMW	0.1383	0.6628	HMW	0.801
FMW	0.6598	0.2662	FMW	0.926
HCW	0.1883	0.6635	HCW	0.8518
FCW	0.1400	0.3174	FCW	0.4574
HAY	0.83446	0.01037	HAY	0.8448
FAY	0.12174	0.00391	FAY	0.1256
FNY	0.755444	0.008584	FNY	0.764
HMY	0.889909	0.008078	HMY	0.898
FMY	0.35148	0.00692	FMY	0.3584
HCY	0.3598	0.0266	HCY	0.3863
FCY	0.03024	0.04690	FCY	0.07714
HAE	0.8926394	0.0009335	HAE	0.8936
FAE	0.13047	0.04583	FAE	0.1763
FNE	0.8752373	0.0001394	FNE	0.8754
HME	0.742612	0.005001	HME	0.7476
FME	0.040219	0.006957	FME	0.04718
HCE	0.543018	0.001829	HCE	0.5448
FCE	0.0008598	0.182535	FCE	0.1834

Saturations des variables ; qualité de la représentation des variables dans le plan (CP1,CP2)

	Comp.1	Comp.2	Comp.3	Comp.4		QLT
PROF	-0.97717	0.12105	-0.084576	0.06694	PROF	0.96952
TRAN	-0.97978	-0.05812	-0.008394	0.04555	TRAN	0.96335
MENA	0.89990	-0.02265	0.362434	0.21417	MENA	0.81034
ENFA	0.87210	-0.17857	0.083731	0.29445	ENFA	0.79244
COUR	0.56359	-0.76062	-0.004560	-0.12104	COUR	0.89617
TOIL	0.07951	-0.81809	-0.302164	-0.06355	TOIL	0.67560
REPA	0.58834	0.66937	-0.426308	0.01414	REPA	0.79419
SOMM	0.64418	0.56930	-0.190853	-0.31249	SOMM	0.73907
TELE	0.09944	-0.19309	-0.930046	0.15125	TELE	0.04717
LOIS	0.09218	-0.11026	0.030219	-0.95745	LOIS	0.02065



- 4) Quelles sont les variables qui sont bien représentées dans le plan (CP1, CP2) ? Comment retrouve-t-on sur le graphique les corrélations entre ces variables ?
- 5) Les variables LOIS et TELE apparaissent proches sur le graphique, alors que leur coefficient de corrélation est faible. Comment peut-on l'expliquer ?
- 6) Les points HMW, HCW, HAW apparaissent très proches sur le graphique. Est-ce le cas dans la réalité ?
- 7) Quelles sont les variables les mieux corrélées à l'axe CP1 ? Quels sont les individus qui participent le plus à l'inertie de cet axe ? Exprimer les résultats relatifs à l'axe CP1 en termes d'oppositions.
- 8) Même question pour l'axe CP2.
- 9) Quelles sont les variables dont l'étude nécessiterait de s'intéresser à CP3, CP4, ... ?

#### Exercice 44 Données Psychométrie

Pour 20 élèves (sujets s1 à s20), on a relevé les notes obtenues à cinq épreuves individuelles : Combinatoire (*Comb*), Probabilités (*Prob*), Logique (*Logi*), notées de 0 à 10, QI verbal (*QI*, notes de 85 à 125) et Mathématiques (*Math*), notée de 0 à 20. Pour chaque sujet, on dispose de deux informations : *Pédagogie* avec deux modalités *p1* (moderne) et *p2* (traditionnelle), *Milieu* avec deux modalités *m1* (favorisé) et *m2* (défavorisé).

	Comb	Prob	Logi	QI	Math	Peda	Milieu
s1	3.9	4.1	6	99	8	p1	m1
s2	5	5	5.2	122	10	p1	m1
s3	5.3	8.5	8.6	108	14	p1	m1
s4	8.3	6.2	7.2	125	18	p1	m1
s5	5.5	6	6.9	108	5	p1	m2
s6	6.6	7.7	5.8	113	7	p1	m2
s7	5.5	3	5.8	94	10	p1	m2
s8	2.2	4.5	3.3	85	9	p1	m2
s9	5.3	4.5	8.3	112	10	p1	m2
s10	5.3	6.4	6.5	125	12	p1	m2
s11	4.6	4.6	5.2	108	14	p1	m2
s12	3.7	4.1	7.2	91	15	p1	m2
s13	4.1	6.7	7.1	91	6	p2	m1
s14	2.7	4.5	3	109	9	p2	m1
s15	6.8	4.5	7.1	125	12	p2	m1
s16	2.7	3.7	6.9	94	13	p2	m1
s17	5.4	8.9	7.3	120	15	p2	m1
s18	6.2	4.7	4.4	112	7	p2	m2
s19	2.5	4.7	7.2	106	11	p2	m2
s20	2.4	4.4	5.2	91	12	p2	m2

#### Analyse en composantes principales

On réalise une ACP normée sur les 4 variables Comb, Prob, Logi et Math. L'essentiel des résultats obtenus est indiqué ci-dessous.

*Matrice des corrélations*

	Comb	Prob	Logi	Math
Comb	1.0000	0.3959	0.3539	0.1512
Prob	0.3959	1.0000	0.3818	0.1114
Logi	0.3539	0.3818	1.0000	0.3476
Math	0.1512	0.1114	0.3476	1.0000

Valeurs propres :

Comp.1	Comp.2	Comp.3	Comp.4
1.8956	0.9533	0.6152	0.5359

Scores, contributions et qualités des individus sur les deux premières composantes :

	QUAL	Comp.1	Ctr	Qual	Comp.2	Ctr	Qual
s1	0.7262	1.0687	0.0301	0.6995	-0.20869	0.0023	0.0267
s2	0.5606	0.50711	0.0068	0.4092	-0.30848	0.0050	0.1514
s3	0.8181	-2.53314	0.1693	0.8175	0.06783	0.0002	0.0006
s4	0.7363	-2.63021	0.1825	0.6665	0.85088	0.0380	0.0698
s5	0.7694	-0.08609	0.0002	0.0020	-1.70358	0.1522	0.7675
s6	0.9575	-0.79544	0.0167	0.1240	-2.06208	0.2230	0.8335
s7	0.2462	0.77573	0.0159	0.2294	0.21026	0.0023	0.0169
s8	0.8419	2.4237	0.1549	0.8417	-0.04369	0.0001	0.0003
s9	0.1687	-0.63765	0.0107	0.1592	0.15502	0.0013	0.0094
s10	0.8293	-0.78804	0.0164	0.8143	-0.10687	0.0006	0.0150
s11	0.5290	0.31604	0.0026	0.0618	0.86867	0.0396	0.4672
s12	0.9548	-0.12761	0.0004	0.0053	1.70023	0.1516	0.9494
s13	0.5104	-0.06279	0.0001	0.0012	-1.32119	0.0915	0.5093
s14	0.8166	2.38174	0.1496	0.8114	-0.19066	0.0019	0.0052
s15	0.3201	-0.87163	0.0200	0.3089	0.16597	0.0014	0.0112
s16	0.8267	0.66734	0.0117	0.1364	1.50142	0.1182	0.6903
s17	0.7030	-2.30108	0.1397	0.7030	0.01647	0.0000	0.0000
s18	0.6412	0.87576	0.0202	0.1953	-1.32301	0.0918	0.4458
s19	0.3743	0.50761	0.0068	0.1042	0.81711	0.0350	0.2701
s20	0.8581	1.30995	0.0453	0.5769	0.91441	0.0439	0.2811

Coefficients des variables :

	Comp.1	Comp.2	Comp.3	Comp.4
Comb	-0.517	-0.354	-0.779	0.000
Prob	-0.518	-0.431	0.521	0.524
Logi	-0.570	0.183	0.322	-0.733
Math	-0.373	0.809	-0.136	0.433

Saturations, contributions et qualité des variables (2 premières composantes)

	QLT	Comp.1	Ctr	Qual	Comp.2	Qual
Comb	0.6265	-0.7121	0.2675	0.5071	-0.3455	0.1252
Prob	0.6861	-0.7133	0.2684	0.5088	-0.4211	0.1860
Logi	0.6480	-0.7848	0.3249	0.6160	0.1791	0.0336
Math	0.8883	-0.5135	0.1391	0.2637	0.7903	0.6551

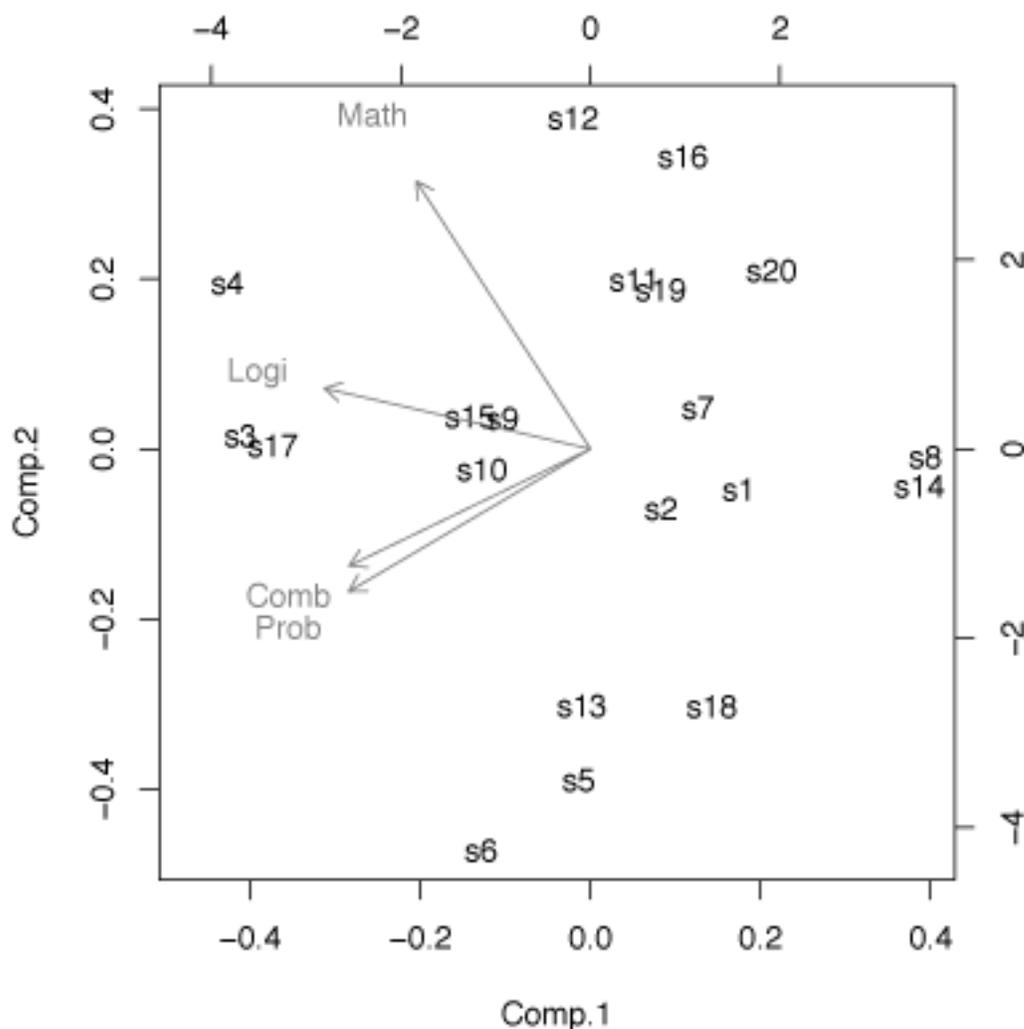


FIG. 2 – Cas Psychométrie

- 1) Quelles sont les coordonnées de l'individu s12 sur les composantes CP1 et CP2. Essayer de placer le point sur le graphique. De même, quelles sont les saturations de la variable Math sur CP1 et CP2. Les retrouve-t-on sur le graphique? Que peut-on penser des graduations des axes fournies par le logiciel?
- 2) Examiner et commenter le tableau des corrélations.
- 3) Les variables Comb et Proba apparaissent proches sur le graphique. Quel est pourtant leur coefficient de corrélation? Comment peut-on l'expliquer?
- 4) Les points s8 et s14 apparaissent très proches sur le graphique. Est-ce le cas dans la réalité? Même question pour s9 et s15.
- 5) Comment les variables contribuent-elles à la formation de l'axe CP1? Comment cet axe classe-t-il les individus?

6) Comment les variables contribuent-elles à la formation de l'axe CP2 ? Décrire cet axe en termes d'oppositions entre variables, en termes d'oppositions entre individus.

7) L'étude limitée aux deux premières composantes vous paraît-elle suffisante ? Comment souhaiteriez-vous poursuivre cette étude ?

## Analyse factorielle des correspondances

### Exercice 45 Données Conjoint

Le tableau ci-dessous rapporte les Circonstances de rencontre de 9675 couples (14 Circonstances) selon leur catégorie socio-professionnelle (16 CSP en lignes). Il s'agit d'un tableau de contingence. Ainsi 9 Agriculteurs de Grande exploitation (AGRD) se sont rencontrés lors de leurs études (ETUD).

#### *Circonstances de rencontre*

Au cours des Etudes (ETUD), sur le lieu de Vacances (VACA), lors d'une Fête entre Amis (FTAM), dans une Association, Sportive par exemple (ASPO), au Travail (TRAV), lors d'une rencontre chez des Particuliers (PART), par Annonce ou agence (ANNO), lors d'une Fête de Famille (FTFM), en Boîte ou Discothèque (BOIT), lors d'une sortie ou Spectacle (SPEC), lors d'une Fête Publique (FTPB), par Voisinage (VOIS), dans un lieu Public (PUBL) ou au Bal (BAL).

#### *Catégories socio-professionnelle (CSP)*

Agriculteurs de Grande (AGRD) ou Petite exploitation (APTI), Ouvriers Non Qualifiés (ONOQ), Ouvriers qualifiés de type Artisan (OART), Ouvriers dans les Transports (OTRA) ou Ouvriers dans l'Industrie (OIND), Artisans (ARTI), Commerçant (COMM), Employés de la Fonction publique (EFCT) ou d'Entreprise (EENT), Professions Intermédiaires de type Contremaître (PCTR) ou de la Santé (PSAN) ou de la Fonction Publique (PFON), Ingénieur ou Cadre d'Entreprise (CENT), Profession Libérale ou Chef d'Entreprise (PLCE), Cadre de la Fonction publique (CFON).

Source : Adapté par l'Université René Descartes, UFR Institut de Psychologie à partir d'une enquête de l'INED, Enquête "Formation des couples", 1984, d'après M. Bozon (1992) Le choix du conjoint in F. de Singly (Ed) La Famille, l'état des savoirs, Ed. La Découverte.

	ETUD	VACA	FTAM	ASPO	TRAV	PART	ANNO	FTFM	BOIT	SPEC	FTPB	VOIS	PUBL	BAL
AGRD	9	8	23	27	17	22	8	12	4	19	15	15	27	150
APTI	12	12	16	28	21	33	0	16	8	43	20	20	39	64
ONOQ	19	5	33	19	86	58	4	23	25	30	30	30	131	191
OART	28	46	63	54	120	103	0	85	65	44	63	63	211	238
OTRA	21	35	44	57	63	67	0	46	39	40	18	18	77	124
OIND	52	32	67	72	67	76	28	57	71	59	64	64	134	217
ARTI	12	29	18	38	70	30	4	12	38	16	13	13	61	110
COMM	14	5	14	11	66	29	8	8	12	12	4	4	42	63
EFCT	34	18	14	29	78	27	0	32	14	17	17	17	69	91
EENT	45	32	38	43	105	74	13	35	24	8	13	13	90	66
PCTR	48	53	84	106	93	107	4	63	36	17	17	17	89	142
PSAN	83	56	72	97	119	59	9	15	49	22	24	24	108	67
PEON	102	45	55	95	103	42	4	22	12	8	8	9	64	37
CENT	93	76	64	59	81	50	0	26	14	9	23	23	72	40
PLCE	60	22	43	22	22	20	0	8	9	0	8	8	16	23
CEON	83	38	34	55	84	30	4	8	17	1	12	12	35	9

## Principaux résultats fournis par Minitab

Axe	Inertie	Proportion	Cumulé	Histogramme
1	0,1203	0,5766	0,5766	*****
2	0,0219	0,1051	0,6818	*****
3	0,0168	0,0807	0,7625	****
4	0,0140	0,0669	0,8294	***
5	0,0112	0,0535	0,8829	**
6	0,0092	0,0443	0,9272	**
7	0,0062	0,0298	0,9570	*
8	0,0034	0,0163	0,9733	
9	0,0028	0,0133	0,9867	
10	0,0020	0,0098	0,9965	
11	0,0005	0,0026	0,9991	
12	0,0002	0,0009	1,0000	
Total	0,2086			

## Contribution des lignes

		--Composante 1--			--Composante 2--			--Composante 3--					
ID	Nom	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr	Coord	Corr	Contr
1	AGRD	0,908	0,037	0,100	-0,532	0,499	0,087	-0,231	0,094	0,090	0,422	0,315	0,390
2	APTI	0,592	0,034	0,060	-0,334	0,305	0,032	-0,296	0,239	0,137	-0,132	0,048	0,036
3	ONOQ	0,824	0,071	0,077	-0,403	0,719	0,095	0,130	0,075	0,054	0,083	0,030	0,029
4	OART	0,866	0,122	0,074	-0,280	0,621	0,079	0,039	0,012	0,008	-0,171	0,233	0,213
5	OTRA	0,573	0,067	0,027	-0,159	0,302	0,014	-0,054	0,035	0,009	-0,141	0,236	0,079
6	OIND	0,635	0,110	0,079	-0,263	0,457	0,063	-0,158	0,164	0,124	0,045	0,013	0,013
7	ARTI	0,430	0,048	0,030	-0,178	0,243	0,013	0,140	0,151	0,043	0,069	0,036	0,013
8	COMM	0,831	0,030	0,035	-0,145	0,086	0,005	0,360	0,531	0,179	0,229	0,214	0,094
9	EFCT	0,320	0,047	0,020	-0,090	0,091	0,003	0,143	0,228	0,044	-0,010	0,001	0,000
10	EENT	0,639	0,062	0,033	0,127	0,146	0,008	0,233	0,490	0,153	-0,021	0,004	0,002
11	PCTR	0,130	0,091	0,041	0,072	0,055	0,004	-0,027	0,008	0,003	-0,079	0,067	0,034
12	PSAN	0,697	0,083	0,046	0,284	0,692	0,056	0,026	0,006	0,002	0,003	0,000	0,000
13	PFON	0,922	0,063	0,106	0,563	0,902	0,165	0,009	0,000	0,000	0,083	0,019	0,025
14	CENT	0,855	0,065	0,087	0,469	0,790	0,119	-0,105	0,040	0,033	-0,084	0,025	0,027
15	PLCE	0,811	0,027	0,080	0,625	0,632	0,088	-0,309	0,154	0,117	0,122	0,024	0,024
16	CFON	0,943	0,044	0,106	0,683	0,923	0,169	0,040	0,003	0,003	0,093	0,017	0,022

## Contribution des colonnes

		---Composante 1---			---Composante 2---			---Composante 3---					
ID	Nom	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr	Coord	Corr	Contr
1	ETUD	0,940	0,074	0,210	0,723	0,882	0,322	-0,118	0,024	0,047	0,143	0,034	0,090
2	VACA	0,812	0,053	0,069	0,449	0,738	0,089	-0,090	0,030	0,020	-0,110	0,044	0,038
3	FTAM	0,729	0,070	0,045	0,265	0,530	0,041	-0,162	0,198	0,084	0,012	0,001	0,001
4	ASPO	0,623	0,084	0,068	0,311	0,570	0,068	-0,093	0,051	0,033	0,017	0,002	0,001
5	TRAV	0,935	0,124	0,076	0,193	0,288	0,038	0,285	0,630	0,458	0,047	0,017	0,016
6	PART	0,237	0,085	0,021	-0,013	0,003	0,000	0,046	0,040	0,008	-0,101	0,193	0,051
7	ANNO	0,392	0,009	0,050	-0,188	0,030	0,003	0,083	0,006	0,003	0,646	0,355	0,220
8	FTFM	0,560	0,048	0,037	-0,154	0,148	0,010	-0,020	0,003	0,001	-0,257	0,409	0,189
9	BOIT	0,179	0,045	0,034	-0,116	0,086	0,005	0,050	0,016	0,005	-0,110	0,077	0,032
10	SPEC	0,647	0,036	0,082	-0,493	0,504	0,072	-0,251	0,131	0,103	-0,078	0,012	0,013

11	FTP	0,587	0,036	0,030	-0,250	0,360	0,019	-0,185	0,196	0,056	-0,074	0,032	0,012
12	VOIS	0,584	0,036	0,029	-0,245	0,354	0,018	-0,184	0,199	0,056	-0,072	0,031	0,011
13	PUBL	0,720	0,131	0,040	-0,131	0,274	0,019	0,142	0,317	0,120	-0,090	0,129	0,063
14	BAL	0,937	0,169	0,206	-0,461	0,831	0,298	-0,028	0,003	0,006	0,162	0,102	0,262

### Diagramme symétrique

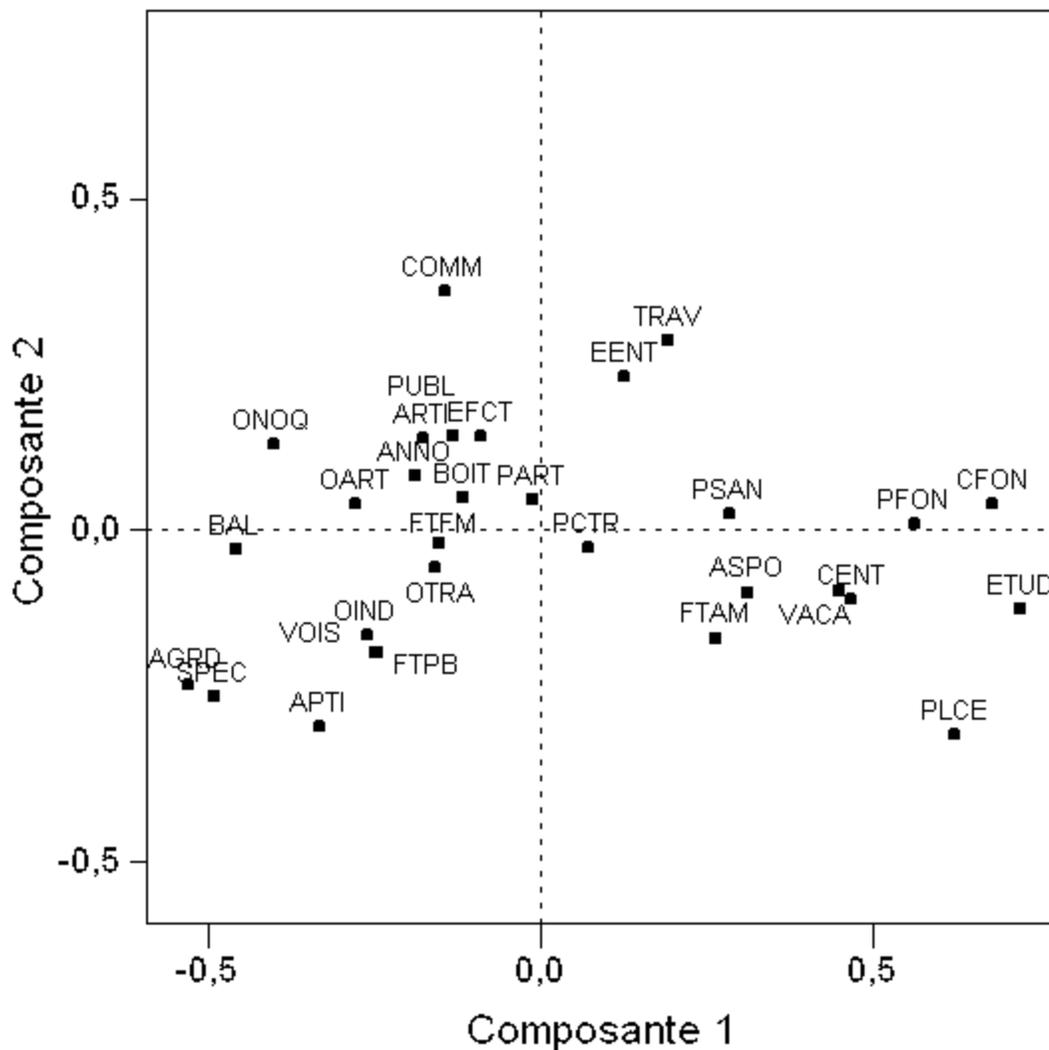


FIG. 3 – Graphe factoriel des points lignes et colonnes

- 1) S'agissant d'un tableau de contingence, indiquer quelle est la méthode factorielle adaptée à ce type de tableau (et mise en oeuvre ici).
- 2) Combien de variables factorielles proposez-vous de retenir pour résumer ces données? Discuter les raisons de ce choix.

- 3) Retrouver approximativement, à l'aide du graphique, la coordonnée factorielle du point CFON sur l'axe 1 (à droite).
- 4) Sur l'axe 1 (cf. Figure) le point AGRD est plus éloigné (coté négatif) du centre de gravité que ne l'est le point ONOQ. Pourtant la contribution de ce dernier à la variance de l'axe est plus forte (cf. Tableau). Expliquer la raison de ce paradoxe apparent.
- 5) Que suggère la position du point PCTR, près de l'intersection des deux axes factoriels ?
- 6) Que suggère, a priori, la proximité des points PFON et CFON sur le graphe factoriel (cf. Figure) ?
- 7) Quelle(s) précaution(s) faut-il prendre avant de conclure définitivement sur la proximité entre ces deux points (PFON et CFON) ? Pourquoi ?
- 8) On considère tout d'abord uniquement le nuage des CSP. A partir du tableau des aides à l'interprétation ci-dessus (cf. Tableau), construire le tableau nécessaire à l'interprétation de la première variable factorielle. Commenter le résultat obtenu.
- 9) Considérant cette fois uniquement les Circonstances (cf. Tableau), procéder de même (construction du tableau et commentaire).
- 10) En se fondant sur les résultats précédents (cf. questions 8 et 9), mettre en correspondance les deux nuages (nuage des CSP et nuage des Circonstances) et indiquer ce que suggère principalement cette analyse de la première variable factorielle.

*Indications de réponses : 1) Une analyse Factorielle des Correspondances (AFC) appelée aussi Analyse des Correspondances (AC).*

*2) On note que les données se situent dans un espace de dimension 13 (plus petite dimension du tableau moins 1). Le pourcentage moyen de variance par axe est 1/13 soit 7.7%. Ceci conduit à retenir les 3 premiers axes. Ce sont ceux dont la contribution à la variance totale est supérieure à 7.7%. Ces trois axes réunissent 76% de la variance totale.*

*3) On trouve environ +0.7 (en fait exactement 0.683)*

*4) Ceci est dû au fait que le poids de ONOQ (71/1000) est plus important que celui de AGRD (37/1000). La contribution à la variance est fonction (du carré) de la distance au centre de gravité, mais aussi du poids relatif.*

*5) Cela suggère que ces sujets (Professions intermédiaires de type Contremaître) ont un profil proche du profil moyen (toutes CSP réunies).*

*6) Cela suggère que les points ont des profils proches (sous la réserve ci-dessous), c'est-à-dire qu'ils rencontrent leurs conjoints dans le même type de lieu.*

*7) Il vaut mieux vérifier en regardant le tableau de données (ou plutôt le tableau des profils). En effet cette proximité peut être la conséquence d'une déformation des distances due à la projection des points sur un plan alors que le nuage se situe en réalité dans un espace de dimension beaucoup plus importante (13).*

*8) On retient les CTR supérieures à 1/16 soit 63/1000.*

-	+
ONOQ (95)	CFON (169)
AGRD(87)	PFON(165)
OART (79)	CENT (119)
OIND (63)	PLCE (88)

On constate que cet axe oppose (coté négatif) les Ouvriers (ONOQ, OART et OIND) et Agriculteurs (AGRD), aux cadres (CFON et CENT) et professions intermédiaires (PFON et PLCE) qu'ils soient fonctionnaires ou en entreprise (coté positif).

9) On retient les CTR supérieures à 1/14 soit 71/1000.

-	+
BAL (298)	ETUD (322)
SPEC (72)	VACA (89)

On constate que cet axe oppose principalement la rencontre dans un Bal (coté négatif) à la rencontre pendant les études (coté positif).

10) Il apparaît pour l'essentiel (sur ce premier axe) que les Ouvriers et Agriculteurs (de grandes exploitations) tendent à rencontrer leur conjoint dans un Bal, alors que les Cadres et Professions intermédiaires rencontrent plutôt leur conjoint pendant leurs Etudes.