

## Comparaison de deux variances, F de Fisher

### Exercice 1

Deux methodes de dosage de l'azote ont ete repetees, a partir d'un m<sup>^</sup>eme echantillon, 25 fois avec la methode A, 30 fois avec la methode B. Les resultats sont rassembles dans les tableaux ci-dessous.

| Methode A    |       |
|--------------|-------|
| $x_i$ (en g) | $n_i$ |
| 37           | 1     |
| 39           | 2     |
| 40           | 2     |
| 41           | 4     |
| 42           | 7     |
| 43           | 4     |
| 44           | 2     |
| 46           | 2     |
| 47           | 1     |
| Total        | 25    |

| Methode B    |       |
|--------------|-------|
| $x_i$ (en g) | $n_i$ |
| 39           | 2     |
| 40           | 1     |
| 41           | 6     |
| 42           | 9     |
| 43           | 8     |
| 44           | 3     |
| 45           | 1     |
| Total        | 30    |

1) Tester l'hypothese : "les valeurs moyennes obtenues par les deux methodes sont egales". (Autrement dit, les methodes sont-elles exactes ?)

2) Comparer les variances des echantillons traites avec les deux methodes. (Autrement dit, les deux methodes ont-elles la m<sup>^</sup>eme precision ?)

Reponses : 1) Les parametres de statistiques descriptives sont donnees par :

|                    | Methode A | Methode B |
|--------------------|-----------|-----------|
| Moyenne            | 42.08     | 42.10     |
| Variance           | 4.95      | 1.89      |
| Variance corrigees | 5.16      | 1.96      |

Le test de comparaison des deux moyennes (groupes independants) conduit a :  $t_{obs} = 0.04$ , evidemment non significatif aux seuils traditionnels. On ne peut donc pas refuser l'hypothese  $H_0$  d'egalite des moyennes.

2) La statistique de test suit une loi de Fisher a  $ddl_1 = 24$  et  $ddl_2 = 29$  degres de liberte. On obtient :  $F_{obs} = 2.63$ . Au seuil de 1% unilatere, on a  $F_{crit} = 2.49$ . On conclut donc a une difference des variances.

### Exercice 2

Au cours de certaines experiences, on est amen a mesurer le temps de reaction (TR) des sujets. C'est le temps qui s'ecoule entre la presentation d'un stimulus (par exemple, une lampe qui s'allume devant le sujet) et la reaction que ce stimulus doit declencher (par exemple, presser un bouton).

Premiere experience. | Le tableau 1 fournit les TR d'une personne qui a reagi 20 fois a l'allumage d'une lampe rouge. On constate que ces 20 TR ne sont pas egaux. Ces variations d'un moment a l'autre sont imprevisibles a partir des informations dont on dispose dans l'experience.

Deuxieme experience. | Le sujet voit maintenant s'allumer devant lui une lampe qui peut etre rouge, verte ou jaune. il doit reagir si la lampe est rouge, mais ne doit pas reagir dans

les deux autres cas. Le tableau 1 fournit 20 TR mesures dans ces conditions. On observe de nouveau des variations imprévisibles d'un moment à l'autre.

Troisième expérience. | Les conditions sont les mêmes que dans la première expérience (une seule lampe) avec une seule référence : au lieu d'être rouge, la lampe donnant le signal de la réaction est verte. La troisième ligne du tableau donne les résultats. Les temps sont de nouveau différents entre eux.

|                                     |    |    |    |    |    |    |    |    |    |    |
|-------------------------------------|----|----|----|----|----|----|----|----|----|----|
| Numero d'ordre des 20 presentations | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
| 1ere experience                     | 20 | 15 | 18 | 25 | 17 | 32 | 18 | 17 | 19 | 23 |
| 2e experience                       | 32 | 40 | 33 | 37 | 35 | 29 | 42 | 62 | 50 | 39 |
| 3e experience                       | 16 | 18 | 19 | 18 | 15 | 18 | 17 | 32 | 23 | 19 |

|                                     |    |    |    |    |    |    |    |    |    |    |
|-------------------------------------|----|----|----|----|----|----|----|----|----|----|
| Numero d'ordre des 20 presentations | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1ere experience                     | 19 | 21 | 15 | 22 | 17 | 17 | 21 | 19 | 17 | 23 |
| 2e experience                       | 45 | 47 | 52 | 37 | 38 | 39 | 40 | 41 | 42 | 39 |
| 3e experience                       | 23 | 20 | 18 | 25 | 15 | 15 | 17 | 23 | 17 | 19 |

La dispersion des TR est-elle la même dans chacune des trois conditions expérimentales ? Pour répondre à cette question, comparer deux à deux les variances des trois séries de données.

Reponses : Les variances des trois séries sont données par :

|                 | Variance | Variance corrigée |
|-----------------|----------|-------------------|
| 1ere experience | 14.89    | 15.67             |
| 2e experience   | 53.85    | 56.68             |
| 3e experience   | 16.23    | 17,08             |

Pour  $ddl_1 = 19$  et  $ddl_2 = 19$  et un seuil de 5%, on a :  $F_{crit} = 3:00$ . Ici,  $F_{2;1;obs} = 3:61$ ,  $F_{2;3;obs} = 3:31$ ,  $F_{3;1;obs} = 1:09$ . Pour les expériences 1 et 3, l'hypothèse nulle (même variance) peut être retenue. En revanche, l'expérience 2 conduit à une variance différente de celles des deux autres.

### Exercice 3 Dossier \pedago"

Lors d'une expérience pédagogique, on s'intéresse à l'efficacité et compare de deux pédagogies des mathématiques chez deux groupes de 10 sujets :

- { pédagogie traditionnelle ( $p_1$ )
- { pédagogie moderne ( $p_2$ )

On note la performance à une épreuve de combinatoire.

| p1<br>traditionnelle |     | p2<br>moderne |     |
|----------------------|-----|---------------|-----|
| s1                   | 5.0 | s11           | 4.0 |
| s2                   | 4.0 | s12           | 5.5 |
| s3                   | 1.5 | s13           | 4.5 |
| s4                   | 6.0 | s14           | 6.5 |
| s5                   | 3.0 | s15           | 4.5 |
| s6                   | 3.5 | s16           | 5.5 |
| s7                   | 3.0 | s17           | 1.0 |
| s8                   | 2.5 | s18           | 2.0 |
| s9                   | 1.5 | s19           | 4.5 |
| s10                  | 2.5 | s20           | 4.5 |

1) Verifier que les parametres des deux echantillons sont donnees par :

|                    | $\mu_1$ | $\mu_2$ |
|--------------------|---------|---------|
| Moyenne            | 3.250   | 4.250   |
| Ecart-type         | 1.365   | 1.553   |
| Variance           | 1.863   | 2.413   |
| Ecart-type corrige | 1.439   | 1.637   |
| Variance corrige   | 2.069   | 2.681   |

2) Avant d'appliquer un test de comparaison de moyennes, on veut s'assurer que l'on peut supposer les variances egales dans les populations parentes. Proceder a un test de comparaison de variances permettant de s'en assurer.

Reponses : 2) On obtient  $F_{obs} = 1:30$ . Or, pour  $ddl_1 = 9$ ,  $ddl_2 = 9$  et un seuil de 5%, on lit dans la table :  $F_{crit} = 3:18$ . L'hypothese  $H_0$  (egalite des variances) est donc retenue.

Exercice 4

1) Pour  $ddl_1 = 2$ ;  $ddl_2 = 4$ , la densite  $f$  de la loi de Fisher-Snedecor est donnee, pour  $x \geq 0$  par :

$$f(x) = \frac{8}{(2+x)^3}$$

Construire point par point la courbe de la fonction  $f$ .

2) Pour  $ddl_1 = 4$ ;  $ddl_2 = 4$ , la densite  $g$  de la loi de Fisher-Snedecor est donnee pour  $x \geq 0$  par :

$$g(x) = \frac{6x}{(1+x)^4}$$

Construire point par point la courbe de la fonction  $g$ .

Analyse de la variance a un facteur (ANOVA) : comparaison de  $k$  moyennes sur des groupes independants

Exercice 5

Un editeur veut choisir entre trois couvertures possibles pour une revue. A cet effet, il a fait noter chaque couverture par un groupe de 5 sujets. Les trois groupes ainsi constitues sont independants. Les notes obtenues sont les suivantes :

|  | Couv. 1 | Couv. 2 | Couv. 3 |
|--|---------|---------|---------|
|  | 14      | 16      | 14      |
|  | 6       | 14      | 16      |
|  | 12      | 8       | 14      |
|  | 10      | 8       | 14      |
|  | 8       | 14      | 12      |

Le test indique-t-il une différence significative entre les trois couvertures ?

Reponses : Exercice traite en CM. Rappel des resultats.

Calcul des sommes de carres :

|                     | C1   | C2   | C3   | Total |
|---------------------|------|------|------|-------|
| $T_j$               | 50   | 60   | 70   | 180   |
| $T_j^2$             | 2500 | 3600 | 4900 |       |
| $n_j$               | 5    | 5    | 5    | 15    |
| $\frac{T_j^2}{n_j}$ | 500  | 720  | 980  | 2200  |
| $\sum x_{ij}^2$     | 540  | 776  | 988  | 2304  |

$$SC_1 = 2200 - \frac{180^2}{15} = 40 ; SC_2 = 2304 - 2200 = 104 ; SC_T = 144$$

| Source     | SC  | ddl | CM   | F                |
|------------|-----|-----|------|------------------|
| A          | 40  | 2   | 20   | $F_{obs} = 2:31$ |
| Residuelle | 104 | 12  | 8:67 |                  |
| Total      | 144 | 14  |      |                  |

Au seuil de 5%,  $F_{crit}(2; 12) = 3:89$ . La différence entre les groupes n'est donc pas significative.

Complement : Modele de score. Chaque observation  $x_{ij}$  peut s'interpreter comme la somme de trois termes :

$$x_{ij} = \mu + a_j + e_{ij}$$

avec les regles suivantes :

- |  $\mu$  est la moyenne de la variable X etudiee (la m<sup>e</sup>me, quel que soit l'individu ou le groupe) ;
- |  $a_j$  est un e et d<sup>u</sup> au groupe (le m<sup>e</sup>me pour tous les individus d'un groupe), nul en moyenne ;
- |  $e_{ij}$  est une variation due au hasard, speci que a chaque observation, de moyenne nulle dans chaque groupe.

Sur l'exemple traite, cette decomposition s'ecrit :

$$\begin{array}{r}
 0 \quad 6 \quad 14 \quad 16 \quad 1 \quad 0 \quad 12 \quad 12 \quad 12 \quad 1 \quad 0 \quad 2 \quad 0 \quad 2 \quad 1 \quad 0 \quad 4 \quad 2 \quad 2 \quad 1 \\
 14 \quad 16 \quad 14 \quad \quad \quad 12 \quad 12 \quad 12 \quad \quad \quad 2 \quad 0 \quad 2 \quad \quad \quad 4 \quad 4 \quad 0 \\
 B \quad 12 \quad 8 \quad 14 \quad C = B \quad 12 \quad 12 \quad 12 \quad C + B \quad 2 \quad 0 \quad 2 \quad C + B \quad 2 \quad 4 \quad 0 \quad C \\
 B \quad 10 \quad 8 \quad 14 \quad C \quad B \quad 12 \quad 12 \quad 12 \quad C \quad B \quad 2 \quad 0 \quad 2 \quad C \quad B \quad 0 \quad 4 \quad 0 \quad C \\
 B \quad 8 \quad 14 \quad 14 \quad C \quad B \quad 12 \quad 12 \quad 12 \quad C \quad B \quad 2 \quad 0 \quad 2 \quad C \quad B \quad 2 \quad 2 \quad 2 \quad C
 \end{array}$$

@                    A                    @                    A                    @                    A                    @                    A

Les sommes des carres inter-groupes et intra-groupes se retrouvent alors comme sommes des carres des elements des deux dernieres matrices (par exemple :  $40 = 5 (2)^2 + 5 (2)^2$ ).

## Exercice 6

Dans un établissement scolaire, on a reparti les élèves en trois classes de troisième ; les notes ci-dessous sont celles obtenues par les élèves en mathématiques au Brevet des Collèges. Peut-on dire que ces trois classes sont équivalentes ? Si oui, quelles seraient les caractéristiques de la population résultant de la fusion des trois groupes ?

| G1 | G2 | G3 |
|----|----|----|
| 14 | 8  | 7  |
| 15 | 18 | 8  |
| 20 | 3  | 11 |
| 7  | 12 | 11 |
| 8  | 15 | 20 |
| 13 | 8  | 14 |
| 10 | 7  | 13 |
| 1  | 11 | 13 |
| 12 | 8  | 10 |
| 16 | 14 | 12 |
| 17 | 14 | 12 |
| 17 | 9  | 13 |
| 11 | 9  | 12 |
| 6  | 9  | 14 |
| 16 | 10 | 8  |

| G1 | G2 | G3 |
|----|----|----|
| 8  | 14 | 13 |
| 10 | 15 | 12 |
| 11 | 14 | 8  |
| 11 | 13 | 8  |
| 7  | 10 | 11 |
| 10 | 12 | 15 |
| 11 | 10 | 8  |
| 12 | 12 | 14 |
| 11 | 12 | 16 |
| 8  | 11 | 13 |
|    | 10 | 12 |
|    | 10 | 15 |
|    | 10 |    |
|    | 12 |    |

Vérifier l'exactitude des tableaux ci-dessous et conclure.

|                 | G1      | G2      | G3      | Totaux   |          |
|-----------------|---------|---------|---------|----------|----------|
| $n_j$           | 25      | 29      | 27      | 81       |          |
| $T_j$           | 282     | 320     | 323     | 925      | 10563,27 |
| $\sum x_{ij}^2$ | 3600    | 3782    | 4091    | 11473    |          |
| $T^2/n_j$       | 3180,96 | 3531,03 | 3864,04 | 10576,03 |          |
| Inter           | 12,76   |         |         |          |          |
| Total           | 909,73  |         |         |          |          |

| Sources de variations | Sommes des carrés | DDL | Carres moyens | F    |
|-----------------------|-------------------|-----|---------------|------|
| Inter                 | 12,76             | 2   | 6,38          | 0,55 |
| Intra                 | 896,97            | 78  | 11,50         |      |
| Total                 | 909,73            | 80  |               |      |

Reponses : Au seuil de 5%,  $F_{crit}(2; 78) = 3:1$ . La différence entre les groupes n'est donc pas significative. De plus, l'obtention d'un  $F_{obs}$  inférieur à 1 semblerait indiquer (sans pour autant le montrer) que les classes n'ont pas été constituées au hasard, mais qu'elles ont, au contraire, été rendues artificiellement homogènes : on a composé les trois classes de façon qu'elles soient de niveau équivalent.

## Exercice 7

On reprend les données d'un exercice vu au premier semestre (dossier pédagogie).

Lors d'une expérience pédagogique, on s'intéresse à l'effet et compare de deux pédagogies des mathématiques chez deux groupes de 10 sujets :

{ pedagogie traditionnelle ( $p_1$ )

{ pedagogie moderne ( $p_2$ )

On note la performance a une epreuve de combinatoire.

| p1<br>traditionnelle |     | p2<br>moderne |     |
|----------------------|-----|---------------|-----|
| s1                   | 5.0 | s11           | 4.0 |
| s2                   | 4.0 | s12           | 5.5 |
| s3                   | 1.5 | s13           | 4.5 |
| s4                   | 6.0 | s14           | 6.5 |
| s5                   | 3.0 | s15           | 4.5 |
| s6                   | 3.5 | s16           | 5.5 |
| s7                   | 3.0 | s17           | 1.0 |
| s8                   | 2.5 | s18           | 2.0 |
| s9                   | 1.5 | s19           | 4.5 |
| s10                  | 2.5 | s20           | 4.5 |

1) Verifier que les parametres des deux echantillons sont donnes par :

|                    | p1    | p2    |
|--------------------|-------|-------|
| Moyenne            | 3.250 | 4.250 |
| Ecart-type         | 1.365 | 1.553 |
| Variance           | 1.863 | 2.413 |
| Ecart-type corrige | 1.439 | 1.637 |
| Variance corrigee  | 2.069 | 2.681 |

2) Ces donnees experimentales permettent-elles d'affirmer que la pedagogie a un effet sur les resultats a l'epreuve de combinatoire ?

a) Comparer les moyennes des deux groupes a l'aide d'une analyse de variance.

b) Comparer les resultats avec ceux obtenus au premier semestre, a l'aide de la statistique T.

Reponses :

Les calculs intermediaires sont resumes dans le tableau suivant :

|                 | Peda1   | Peda2   | Totaux |        |
|-----------------|---------|---------|--------|--------|
| $n_j$           | 10      | 10      | 20     |        |
| $T_j$           | 32.5    | 42.5    | 75     | 281.25 |
| $\sum x_{ij}^2$ | 124.25  | 204.75  | 329    |        |
| $T_j^2/n_j$     | 105.625 | 180.625 | 286.25 |        |
| Inter           | 5.00    |         |        |        |
| Total           | 47.75   |         |        |        |

Le tableau d'analyse de variance est donc :

| Sources de variation | Sommes des carres | DDL | Carres moyens | F    |
|----------------------|-------------------|-----|---------------|------|
| Inter                | 5,0               | 1   | 5,0           | 2,11 |
| Intra                | 42,75             | 18  | 2,375         |      |
| Total                | 47,75             | 19  |               |      |

Au seuil de 5%,  $F_{crit}(1; 18) = 4.41$ . Hypothese  $H_1$  rejetee.

Comparaison possible avec l'exercice vu au premier semestre :  $t_{obs}^2 = (1.45)^2 = 2.10$ , c'est-a-dire la valeur de  $F$ .

Enonce 8 Donnees Bransfor

On reprend une experience de Bransford et al. (1972), dans laquelle on demande a des sujets d'ecouter le texte suivant :

"Si les ballons eclatent, le son ne portera pas puisque tout sera bien trop loin du bon etage. Une fenetre fermee emp^echera egalement le son de porter, surtout depuis que les immeubles recents sont correctement isoles. Comme l'essentiel de l'operation depend d'une arrivee correcte d'electricite, un I casse causerait bien des problemes. Evidemment, le type peut hurler. Mais la voix humaine n'est pas assez puissante pour porter bien loin. Un probleme supplementaire serait qu'une corde casse sur l'instrument. Alors il serait im-possible d'accompagner le message. C'est clair que la meilleure situation im-pliquerait la plus petite distance. Alors, il y aurait bien moins de problemes potentiels. Avec un contact en face a face, un bien petit nombre de choses pourrait g^ener."

Le but vise par Bransford et al. est de montrer l'importance du contexte dans la comprehension et la memorisation d'un texte. Pour ce faire, ils utilisent quatre groupes experimentaux :

1. Un groupe "sans contexte" entend simplement le texte.
2. Le groupe "avec contexte avant" regarde une gure suggerant un contexte approprie pendant qu'il entend le texte.
3. Le groupe "avec contexte apres" entend le texte puis regarde la gure precedente.
4. Le groupe "avec contexte partiel" regarde une gure suggerant un contexte inapro-prie pendant qu'il entend le texte.

A proprement parler cette etude comprend un groupe experimental (le groupe 2 : contexte pendant) et trois groupes contr^oles (les groupes 1, 3 et 4). Les groupes contr^oles doivent permettre d'eliminer des explications concurrentes (en particulier, e et facilitateur sur la memoire de l'imagerie, de l'aspect concret du materiel, etc.). L'experimentateur s'attend, donc, a observer une performance pour le groupe 2 superieure aux trois autres groupes. Il choisit de mesurer le comportement des sujets par deux Variables Dependantes : une note de comprehension donnee par les sujets (de 0 a 7, avec 0 indiquant l'incomprehension totale), et le nombre d'idees correctement rappelees (Bransford decoupe le texte en 14 idees, essayez de les retrouver !). Quoique cette derniere Variable Dependante souleve de delicats problemes de codage (e.g., a partir de quel moment une idee est presente...), elle re ete clairement l'inter^et des auteurs de cette experimentation.

Dans cette experience, on utilise vingt sujets repartis en quatre groupes. Les resultats, pour la Variable Dependante "nombre d'idees rappelees" (maximum 14) se trouvent ci-dessous (mais avant, faites ce que doit faire un bon experimentateur : prenez une feuille et detaillez les cinq premieres etapes du test statistique avant de partir a la p^eche aux resultats) :

Figure 1 { Contexte inapproprié (a) et approprié (b) pour l'expérience de Bransford

| Resultats de l'experience |     |     |     |     |
|---------------------------|-----|-----|-----|-----|
|                           | G.1 | G.2 | G.3 | G.4 |
|                           | 3   | 5   | 2   | 5   |
|                           | 3   | 9   | 4   | 4   |
|                           | 2   | 8   | 5   | 3   |
|                           | 4   | 4   | 4   | 5   |
|                           | 3   | 9   | 1   | 4   |
| $T_{.j}$                  | 15  | 35  | 16  | 21  |
| $n_j$                     | 5   | 5   | 5   | 5   |
| $\frac{T_{.j}}{n_j}$      | 3   | 7   | 3.2 | 4.2 |
| $\sum x_{ij}^2$           | 47  | 267 | 62  | 91  |

Justifiez les calculs et le tableau d'ANOVA suivants :

Table d'ANOVA :

| Source | ddl | SC    | CM    | $F_{cal}$ | $P r(F > F_{cal})$ |
|--------|-----|-------|-------|-----------|--------------------|
| A      | 3   | 50.95 | 16.98 | 7.22 **   | .00288             |
| S(A)   | 16  | 37.60 | 2.35  |           |                    |
| Total  | 19  | 88.55 |       |           |                    |

Si on utilise la procédure des valeurs critiques :

\*\*  $F_{critique} = 5.29$ , au seuil =  $.01$  ;  $F_{cal} > F_{critique}$ . On rejette  $H_0$ .

Les cinq étapes du test sont évidemment :



1. Formulation des hypotheses statistiques  $H_0$  et  $H_1$ . Ici :  
 $H_0$  : dans les 4 conditions, les moyennes dans la population parente sont egales  
 $H_1$  : les 4 moyennes ne sont pas toutes egales.
2. Choix du test : ici, une analyse de variance a un facteur. Statistique : F .
3. Distribution de la statistique de test : ici, le F de Fisher Snedecor avec  $ddl_1 = 3$  (nombre de groupes { 1) et  $ddl_2 = 16$  (nombre d'observations { nombre de groupes).
4. Seuil de signi cation choisi : ici, = 1%.
5. Regle de decision : determination des zones d'acceptation et de rejet de  $H_0$ . Ici, :  
 { Si  $F_{cal} \leq 5:29$ , on accepte  $H_0$  (egalit des moyennes)  
 { Si  $F_{cal} > 5:29$ , on refuse  $H_0$  et on accepte  $H_1$ .

L'etude pourrait ^etre poursuivie a l'aide de la methode des contrastes orthogonaux (que nous ne detaillerons pas).

La premiere etape consiste opposer le groupe 2 aux trois autres groupes en testant l'hypothese nulle :  $3 \mu_2 = \mu_1 + \mu_3 + \mu_4$ . On calcule :  $L_1 = 3\bar{x}_2 - \bar{x}_1 - \bar{x}_3 - \bar{x}_4 = 10:6$  ;

$$P_{aj^2} = 3^2 + (1)^2 + (1)^2 + (1)^2 = 12 ; \quad SC_{\text{contraste1}} = \frac{nL^2}{P_{aj^2}} = 46:81$$

Dans la formule precedente, n est le nombre d'observations par groupe. Ici, n = 5. Le F de Fisher associe a ce contraste est obtenu en divisant  $SC_{\text{contraste1}}$  par le carre moyen residuel 2.35 ; il vaut 19.92. Les degres de libert sont 1 et 16. Le resultat est donc signi catif d'un comportement du groupe 2 different de celui des autres groupes.

La methode peut ^etre poursuivie en opposant le groupe 4 aux groupes 1 et 3 (coe cients appliques aux quatre moyennes : 1, 0, 1, -2) puis en opposant les groupes 1 et 3 (coe cients appliques : 1, 0, -1, 0).

Pourquoi s'agit-il de contrastes orthogonaux ?

Reponse : Les "vecteurs" associes aux coe cients des trois contrastes, a savoir  $V_1 = (1; 3; 1; 1)$ ,  $V_2 = (1; 0; 1; 2)$ ,  $V_3 = (1; 0; 1; 0)$  sont deux a deux orthogonaux (par exemple,  $V_1 \cdot V_2 = 1 \cdot 1 + 3 \cdot 0 + 1 \cdot 1 + 1 \cdot 2 = 0$ ), ce qui garantit l'indpendance des resultats des trois tests.

Une autre grandeur interessante est le coe cient (souvent note  $\eta^2$ ) d'estimation de l'intensit de l'effet et de la variable independante. Dans le cas d'une analyse de variance a un facteur, il est defini par :

$$\eta^2 = \frac{SC_{\text{inter}}}{SC_{\text{total}}}$$

Il vaut donc ici :  $\eta^2 = 0:58 = 58\%$ .

Signi cation : 58% de la variance de la Variable Dependante est expliquee par la Variable Independante (les differentes conditions experimentales).

$\eta^2$  est aussi le carre d'un coe cient de correlation. peut en ^etre obtenu comme coe cient de la correlation entre l'ensemble des donnees observees d'une part, et la serie de donnees obtenue en remplaçant chaque observation par la moyenne de son groupe d'autre part. Sur notre exemple, soit U la serie des donnees observees et V la serie des donnees du modele ainsi obtenu.

|       |   |   |   |   |   |   |   |   |   |   |     |     |     |     |     |     |     |     |     |     |
|-------|---|---|---|---|---|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $u_j$ | 3 | 3 | 2 | 4 | 3 | 5 | 9 | 8 | 4 | 9 | 2   | 4   | 5   | 4   | 1   | 5   | 4   | 3   | 5   | 4   |
| $v_i$ | 3 | 3 | 3 | 3 | 3 | 7 | 7 | 7 | 7 | 7 | 3.2 | 3.2 | 3.2 | 3.2 | 3.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 |

On obtient :  $r(U; V) = 0.7585$  et  $r^2(U; V) = 0.575$ .

### Enonce 9 Donnees Loftus

Elisabeth Loftus (Loftus et Palmer 1974) a dans une serie d'experimentations sur le theme du temoignage desire mettre en evidence l'influence de la tournure d'une question sur la reponse de temoins. Pour ce faire, elle montre a ses sujets, un film decrivant un accident de voiture. Elle pose, ensuite, une serie de questions aux sujets. Parmi celles-ci se trouve une des cinq versions d'une question relative a la vitesse des vehicules. Voici ces versions :

- 1) HIT : About how fast were the cars going when they hit each other ? (A environ quelle vitesse allaient les voitures quand elles se sont "percutees").
- 2) SMASH : About how fast were the cars going when they smashed each other ? (To smash : ecraser, heurter avec violence).
- 3) COLLIDE : About how fast were the cars going when they collided each other ? (To collide : entrer en collision, s'emboutir).
- 4) BUMP : About how fast were the cars going when they bumped each other ? (To bump : cogner, frapper).
- 5) CONTACT : About how fast were the cars going when they contacted each other ? (To contact : entrer en contact).

Les sujets repondaient en indiquant une vitesse exprimee en miles (nous sommes aux U.S.A). Voici les resultats obtenus (lors d'une replique de l'experience) :

| HIT | SMASH | COLLIDE | BUMP | CONTACT |
|-----|-------|---------|------|---------|
| 22  | 38    | 43      | 47   | 27      |
| 29  | 40    | 39      | 29   | 24      |
| 33  | 50    | 32      | 58   | 46      |
| 50  | 45    | 44      | 34   | 37      |
| 19  | 48    | 29      | 36   | 31      |
| 37  | 56    | 44      | 43   | 37      |
| 33  | 52    | 45      | 25   | 34      |
| 43  | 47    | 33      | 58   | 18      |
| 40  | 39    | 48      | 24   | 28      |
| 34  | 40    | 37      | 31   | 26      |

Apres avoir identifi les variables dependante(s) et independante(s), vous tirerez les conclusions de cette experimentation.

Pour vous aider voici quelques statistiques pour chaque groupe :

|       | $\sum x_j$ | $\bar{x}_j = \frac{\sum x_j}{n_j}$ | $\sum x_j^2 = n_j \bar{x}_j^2$ | $\sum x_j^2$ |
|-------|------------|------------------------------------|--------------------------------|--------------|
| Gr. 1 | 340        | 34.0                               | 11560                          | 12338        |
| Gr. 2 | 455        | 45.5                               | 20702.5                        | 21043        |
| Gr. 3 | 394        | 39.4                               | 15523.6                        | 15894        |
| Gr. 4 | 385        | 38.5                               | 14822.5                        | 16241        |
| Gr. 5 | 308        | 30.8                               | 9486.4                         | 10060        |
| Total | 1882       |                                    | 72095                          | 75576        |

La Variable Dependante est évidemment la vitesse exprimée en miles. La Variable Indépendante est le type de verbe utilisé pour poser la question sur la vitesse des voitures. Manifestement, E. Loftus veut montrer que les "sous-entendus" des verbes sont pris en compte par les sujets dans leur décision sur la vitesse (e.g., les sujets utilisent la signification implicite des verbes comme une source d'information). Le point d'importance dans cette expérience est de remarquer que E. Loftus désire généraliser ses résultats à l'ensemble des verbes signifiant quelque chose comme "entrer en contact". Quoiqu'elle n'ait pas, proprement parler, sélectionné ses verbes au hasard, elle les juge représentatifs de l'ensemble des verbes de mouvement. Le problème ici est de décider si le facteur expérimental est fixe ou aléatoire. Si l'on admet que les verbes choisis par Loftus représentent un échantillon représentatif, on décidera que le facteur est aléatoire (cf. La polémique initiée par Clark 1973). Si l'on juge que les modalités sont choisies en fait arbitrairement, on décidera que le facteur est fixe, et les conclusions de l'étude se restreignent aux modalités effectivement présentes dans l'expérimentation. Quelle que soit la décision prise, elle sera critiquable.

Ici, le distinguo entre facteur fixe et aléatoire peut paraître sans importance car la décision (rejet ou non de l'hypothèse nulle) sera identique dans les deux cas. Ce ne sera plus le cas dans des plans d'expérience plus complexes. En fait, l'essentiel de l'argument de Clark (1973) est de montrer qu'une partie des recherches utilisant du matériel linguistique aboutit à des conclusions scientifiques erronées du fait de la confusion entre facteurs fixes et aléatoires (cf. aussi les réponses de Wike et Church 1976). Clark défend l'idée qu'une partie des conclusions de la psychologie du langage est invalide pour avoir cru que des facteurs aléatoires étaient fixes. A cette attaque répond Chastaing (1986) qui démontre méthodologiquement qu'une autre partie de la psychologie du langage est invalide d'avoir cru que des facteurs fixes étaient aléatoires !

Dans le cas présent, le choix entre les deux modèles n'a pas d'influence sur les résultats de l'analyse statistique : on aboutit à des conclusions statistiques identiques (mais pas à des interprétations psychologiques identiques !). L'analyse de variance permet de conclure en tout cas à un effet et sur la vitesse estimée, du type le verbe utilisé pour poser la question. On obtient le tableau d'analyse de variance suivant :

| Source        | ddl | SC      | CM     | $F_{cal}$ | $P r(F > F_{cal})$ |
|---------------|-----|---------|--------|-----------|--------------------|
| Expérimentale | 4   | 1256.52 | 314.13 | 4.06 **   | .0069              |
| Erreur        | 45  | 3481.00 | 77.36  |           |                    |
| Total         | 49  | 4737.52 |        |           |                    |

Ainsi, le type de verbe employé pour interroger les sujets sur la vitesse des véhicules, influence l'estimation qu'ils donnent ( $F_{cal}(4; 45) = 4.06; p < .05$ ). On remarque la vitesse élevée induite par "to smash". Nous pourrions poursuivre cet exemple en essayant d'apprécier les différences entre ces différents verbes les uns par rapport aux autres).

#### Énoncé 10 Données Besançon

On fait subir à 30 élèves d'une école de Besançon une épreuve de "précision perceptive" qui consiste à évaluer un nombre de points sur une diapositive projetée pendant un temps relativement court (une demi-seconde). Les auteurs de cette expérience pensent que la présence d'un témoin peut influencer la performance des sujets dans cette tâche perceptive. Pour vérifier cette idée, les expérimentateurs divisent leur échantillon en trois groupes [ chaque enfant étant affecté à un groupe en utilisant une "table de nombres au hasard". Dans le premier groupe (A1) l'expérience est effectuée sans témoin ; dans le second groupe

(A2) l'enfant accomplit sa tâche en compagnie d'un témoin présent par l'expérimentateur comme un spécialiste ; dans le troisième groupe (A3), le témoin est présent comme un simple curieux. On répète 10 fois pour chaque sujet 10 vingt-cinq fois l'expérience. Et l'on retient pour chaque sujet la moyenne des écarts absolus (i.e. en ignorant le signe) entre l'estimation fournie et le nombre exact de points.

Les expérimentateurs s'attendent à trouver des différences entre les trois conditions expérimentales ; mais, plus précisément, entre la condition "sans témoin" et la condition "témoin simple curieux" (cette différence leur permettrait de contredire un de leurs collègues qui avançait dans une expérience voisine que le témoin n'agissait que parce que les enfants le jugeaient spécialiste). Les auteurs veulent, également vérifier l'existence d'une et spécifique à la condition "témoin spécialiste".

Questions :

Pourquoi les expérimentateurs décident-ils de prendre l'écart absolu et non pas 10 l'écart relatif. Tout de même, pourquoi retiennent-ils la moyenne des vingt-cinq essais, plutôt qu'un seul essai ?

Quelle est la (les) variable(s) indépendante(s), la (les) variable(s) dépendante(s) ?

Après avoir traduit en termes statistiques les objectifs des expérimentateurs, peut-on penser que ces objectifs sont atteints ? Voici les résultats obtenus :

|              |     |     |     |     |     |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Condition A1 | 140 | 124 | 118 | 115 | 110 | 110 | 108 | 104 | 102 | 90  |
| Condition A2 | 170 | 164 | 161 | 158 | 156 | 148 | 143 | 140 | 130 | 126 |
| Condition A3 | 136 | 120 | 112 | 104 | 102 | 96  | 92  | 84  | 81  | 75  |

Éléments de réponses. Calculs intermédiaires :

|                 | A1       | A2       | A3       | Totaux   |           |
|-----------------|----------|----------|----------|----------|-----------|
| $n_j$           | 10       | 10       | 10       | 30       |           |
| $T_j$           | 1121     | 1496     | 1002     | 3619     | 436572.03 |
| $\sum x_{ij}^2$ | 127303   | 225746   | 103582   | 456637   |           |
| $T_j^2/n_j$     | 125664.1 | 223801.6 | 100400.4 | 449866.1 |           |
| Inter           | 13294.07 |          |          |          |           |
| Total           | 20064.97 |          |          |          |           |

Le tableau d'analyse de variance est donné par :

| Source        | ddl | SC      | CM      | $F_{cal}$ |
|---------------|-----|---------|---------|-----------|
| Inter-groupes | 2   | 13294.1 | 6647.03 | 26.51 **  |
| Intra-groupes | 27  | 6770.9  | 250.77  |           |
| Total         | 29  | 20065   |         |           |

Les trois groupes ne sont donc pas équivalents. La méthode peut être poursuivie en décomposant la variation intra-groupes selon les deux contrastes orthogonaux suggérés par l'énoncé :

$$L_1 = \bar{x}_1 - \bar{x}_3 = 11:9$$

$$L_2 = 2 \bar{x}_2 - \bar{x}_1 - \bar{x}_3 = 86:9$$

On obtient alors :

$$SC_{\text{contraste1}} = 708.05; F = 2.82; P r(F) = 0.10$$

$$SC_{\text{contraste2}} = 12586.02; F = 50.19; P r(F) = 1.310^{-7}$$

L'expérience ne met donc pas de différence en évidence entre les conditions "sans témoin" et "témoin simple curieux" mais par contre, montre un comportement différent dans la condition "témoin spécialiste".

#### Énoncé 11

Un chercheur a soumis quatre groupes de cinq élèves à un apprentissage de "résolutions de problèmes mathématiques". Chaque groupe apprend avec une méthode pédagogique propre : le premier avec une méthode uniquement verbale, le second avec une méthode écrite, le troisième avec un schéma annoté, le quatrième avec une série de schémas annotés. L'apprentissage dure une heure pour chaque groupe, et le même contenu est présent. Deux jours après l'apprentissage, les sujets sont soumis à un test de raisonnement mathématique. Ce test provient des travaux d'autres chercheurs qui ont étalonné ce test sur une population comparable à celle dont provient l'échantillon d'enfants utilisé ici ; le résultat de ce test est une note (de 0 à 35 : plus la note est élevée, meilleur est le résultat).

Quelle est la Variable Indépendante, la Variable Dépendante ? Comment l'expérimentateur traitera-t-il les résultats de son expérience (souvenez-vous qu'il faut pouvoir répondre à cette question avant de recueillir les résultats !)

En outre, l'auteur a mis au point cette expérience pour vérifier certaines hypothèses précises :

1. La méthode verbale diffère-t-elle de l'ensemble des autres méthodes
2. La méthode écrite diffère-t-elle des méthodes avec schémas (un ou plusieurs) ?
3. Le nombre de schémas a-t-il une influence décelable sur la performance ?

L'auteur peut-il répondre simultanément à ces différentes questions, et quelles seront les réponses ? Interprétez | en vous justifiant | les résultats obtenus et concluez. Voici les résultats :

| groupe expérimental |    |    |    |
|---------------------|----|----|----|
| A1                  | A2 | A3 | A4 |
| 6                   | 14 | 22 | 23 |
| 13                  | 10 | 11 | 19 |
| 16                  | 14 | 19 | 25 |
| 14                  | 19 | 19 | 24 |
| 14                  | 25 | 23 | 25 |

Éléments de réponses. Calculs intermédiaires :

|                      | A1     | A2     | A3     | A4     | Totaux |         |
|----------------------|--------|--------|--------|--------|--------|---------|
| $n_j$                | 5      | 5      | 5      | 5      | 20     | 6301.25 |
| $\sum_j x_{ij}^2$    | 63     | 82     | 94     | 116    | 355    |         |
| $\sum_{ij} x_{ij}^2$ | 853    | 1478   | 1856   | 2716   | 6903   |         |
| $T_j^2 = n_j$        | 793.8  | 1344.8 | 1767.2 | 2691.2 | 6597   |         |
| Inter                | 295.75 |        |        |        |        |         |
| Total                | 601.75 |        |        |        |        |         |

Le tableau d'analyse de variance est donné par :

| Source        | ddl | SC     | CM     | $F_{cal}$ | $Pr(F > F_{cal})$ |
|---------------|-----|--------|--------|-----------|-------------------|
| Inter-groupes | 3   | 295.75 | 98.58  | 5.15      | 0.011             |
| Intra-groupes | 16  | 306    | 19.125 |           |                   |
| Total         | 19  | 601.75 |        |           |                   |

La methode peut ^etre poursuivie en decomposant la variation inter-groupes selon les trois contrastes orthogonaux suggeres par l'enonc :

$$L_1 = 3 \frac{\bar{x}_1}{\bar{x}_2} - \frac{\bar{x}_2}{\bar{x}_3} - \frac{\bar{x}_3}{\bar{x}_4} = 20:6$$

$$L_2 = 2 \frac{\bar{x}_2}{\bar{x}_3} - \frac{\bar{x}_3}{\bar{x}_4} = 9:2$$

$$L_3 = \bar{x}_3 - \bar{x}_4 = 4:4$$

On obtient alors :

$$SC_{\text{contraste1}} = 176:82; F = 9:24; P r(F) = 0:0067$$

$$SC_{\text{contraste2}} = 70:53; F = 3:69; P r(F) = 0:07$$

$$SC_{\text{contraste3}} = 48:40; F = 2:53; P r(F) = 0:13$$

## Exercice 12

Ci-dessous gure un extrait d'un ouvrage de statistiques relatif a un test statistique qui n'a pas et etudie en cours, le test H de Kruskal et Wallis.

Liaison entre un caractere quantitatif et un caractere qualitatif a k classes ( $k > 2$ ).

C'est le probleme appele, dans les chapitres precedents, "comparaison de plusieurs moyennes" et traite par analyse de variance. Le test non parametrique correspondant le plus usuel est le test H de Kruskal et Wallis.

On classe les observations de l'ensemble des k series, comme on le faisait pour les deux series dans les tests precedents, puis on calcule les rangs moyens  $\bar{W}_1; \bar{W}_2; \dots; \bar{W}_k$  et le

rang moyen  $\bar{W}$ , ce dernier valant  $\frac{N+1}{2}$  si N represente le nombre total d'observations. 2

Dans l'hypothese nulle,  $\bar{W}_1; \bar{W}_2; \dots; \bar{W}_k$  ne doivent pas trop s'ecarter de  $\bar{W}$ , de sorte

que les quantites  $W_i - \bar{W}$  ne doivent pas ^etre trop grandes. On montre que, sous l'hypothese nulle, la statistique :

$$H = \frac{\sum_{i=1}^k (\bar{W}_i - \bar{W})^2}{N} \frac{N-1}{(N+1)=12}$$

suit approximativement une loi du  $\chi^2$  a  $k - 1$  degres de libert .

Dans cette expression, les  $n_i$  designent les effectifs des diverses series.

L'approximation n'est valable que s'ils atteignent tous la dizaine, a la rigueur 5.

1) Dans quelles situations ce test doit-il ^etre prefer e a une analyse de variance ?

2) Premiere application. Une variable A comporte trois modalites  $a_1; a_2; a_3$ . Pour chaque modalite, on dispose de 2 ou 3 observations d'une variable numerique. Ces observations sont rassemblees dans le tableau ci-dessous.

| $a_1$ | $a_2$ | $a_3$ |
|-------|-------|-------|
| 14    | 12    | 14    |
| 16    | 15    | 18    |
| 13    |       | 14    |

Construire sur cet exemple le protocole des rangs ( $W_i$ ) et calculer les rangs moyens et la statistique H. (Vu le faible nombre d'observations, on s'abstiendra ici d'effectuer le test).

3) Deuxieme application. A n de constituer un groupe su samment important en vue d'une recherche, un chercheur teste 3 groupes de 10 sujets. Les rangs moyens observes sur

les trois groupes sont les suivants :

$$\bar{W}_1 = 9;8 ; \bar{W}_2 = 24;05 ; \bar{W}_3 = 12;65:$$

Peut-on considerer que les trois groupes testes sont issus d'une m<sup>e</sup>me population ?

Reponses : 1) Un test non parametrique doit etre prefer lorsque la variable est ordinale, ou lorsque l'on ne peut pas faire d'hypothese concernant la normalite des distributions dans les populations parentes.

2) Le protocole des rangs est donne par :

| a1   | a2  | a3   |
|------|-----|------|
| 4    | 1   | 4    |
| 7    | 6   | 8    |
| 2    |     | 4    |
| 4.33 | 3.5 | 5.33 |

Dans ce cas,  $\bar{W} = \frac{36}{8}$  ;  $H = \frac{1}{8} \frac{3(4:33 \ 4:5)^2 + 2(3:5 \ 4:5)^2 + 3(5:33 \ 4:5)^2}{9=12} = 0:6944$

3) Dans ce cas,  $H_{obs} = 14:67$ ,  $ddl = 2$  et, au seuil de 1%,  $\chi^2_c = 9:21$ . On conclut donc a l'heterogene te des groupes.

## Plans d'expériences, interactions

### Énoncé 13

Déterminez VI et VD dans les hypothèses suivantes :

- a) Un individu est d'autant plus attaché à une opinion qu'il s'est davantage engagé à la défendre.
- b) Le degré de violence d'un événement modifie sa mémorisation.
- c) Je l'aime plus qu'hier et bien moins que demain.
- d) Les individus agressifs assurent plus souvent le leadership dans un groupe, mais ils en sont moins satisfaits que les membres des groupes non agressifs.
- e) On retient plus facilement un matériel significatif qu'un matériel dépourvu de sens.
- f) C'est dans les vieux pots que l'on fait les meilleures confitures.
- g) On restitue d'autant mieux une information que celle-ci est rappelée dans le même contexte que celui où elle a été apprise.
- h) Les aînés sont plus anxieux que les jeunes.
- i) L'importance d'un discours est d'autant plus importante que l'orateur possède du prestige.
- j) Le nombre de conversations au cours d'un repas dépend étroitement de la disposition des individus autour de la table.

Indications de réponses : a) VI = intensité de l'engagement, VD = une mesure numérique liée à l'attachement à l'opinion.

b) VI = degré de violence d'un événement, VD = mesure numérique liée à la mémorisation de l'événement.

c) VI = Date, VD = mesure du degré d'amour.

d) Ici, deux hypothèses combinées :

VI = agressivité, VD liée à la prise de leadership

VI = agressivité, VD = mesure de la satisfaction des membres du groupe.

e) VI = significativité du matériel, VD = évaluation numérique de la performance mnésique. f) VI = âge du pot, VD = mesure de la qualité de la confiture.

g) VI1 = contexte d'apprentissage, VI2 = contexte de rappel (ou lien (même / différent) entre contexte de rappel et contexte d'apprentissage), VD = évaluation numérique de la performance mnésique.

h) VI = rang de naissance, VD = mesure de l'anxiété.

i) VI = prestige de l'orateur, VD = mesure du degré d'importance.

j) VI = disposition autour de la table, VD = nombre de conversations.

### Énoncé 14 Données Schizo

Dans une série d'expériences destinées à éclaircir la notion de "maladie mentale" on soumet des sujets diagnostiqués comme schizophrènes et des sujets normaux à une épreuve de "formation de concept". Tous les sujets retenus pour participer à l'expérience doivent posséder un Q.I. compris entre 100 et 105. Pourquoi ?

On compte pour chaque sujet le nombre d'essais nécessaires pour arriver à former un nouveau concept.

Dans cette expérience on utilise deux ensembles de stimuli : Le premier contient des images illustrant l'approbation sociale, le second des images illustrant la désapprobation sociale. L'auteur de cette expérience émet les prédictions suivantes (qui découlent de certaines théories de la personnalité et des performances intellectuelles) :

- a) Les sujets normaux devront arriver plus rapidement que les schizophrènes à accomplir l'épreuve et ce, indépendamment de la nature des images ;



- b) Les sujets normaux ne seront pas influencés par la nature des stimuli ;  
 c) Les schizophrènes devront réussir moins facilement les épreuves comportant des images exprimant la désapprobation sociale que les épreuves décrivant l'approbation sociale. Quel est le plan d'expérience utilisé ? Les prédictions du chercheur se traduisent par des prédictions sur les hypothèses statistiques, lesquelles ?

Éléments de réponses : Facteurs : sujet S, maladie M<sub>2</sub>, nature des images I<sub>2</sub>. L'effet du facteur QI a été éliminé par le choix initial des sujets. Plan : S < M<sub>2</sub> > I<sub>2</sub>. La combinaison des hypothèses a) et c) se traduit par une interaction : les sujets réussiront moins facilement que les sujets normaux, mais cette différence sera plus marquée dans la condition i<sub>2</sub> que dans la condition i<sub>1</sub>.

#### Énoncé 15

Une expérience a été menée en utilisant un plan de la forme S<sub>8</sub> < A<sub>3</sub> B<sub>2</sub> > C<sub>4</sub>.

- 1) a) Combien de groupes indépendants de sujets a-t-on constitué ?  
 b) Combien de sujets différents ont été utilisés pour cette expérience ?  
 Combien d'observations différentes d'un même sujet a-t-on effectué ?
- 2) a) Dériver ce plan d'expérience et donner les différentes sources de variation.  
 b) Proposer un modèle de score.

Réponses : 1) a) Les sujets sont imbriqués dans < A<sub>3</sub> B<sub>2</sub> >. On a 6 combinaisons de modalités et donc 6 groupes de sujets.

b) L'indice dans S<sub>8</sub> < A B > montre qu'il y a 8 sujets dans chaque groupe, soit en tout : 6 × 8 = 48 sujets.

c) Les sujets sont croisés avec C<sub>4</sub>, on a donc fait 4 observations de chaque sujet.

2) a) Les facteurs élémentaires sont ici : A, B, C, S(AB).

Termes d'interaction d'ordre 1 : AB, AC, BC, CS(AB).

Terme d'interaction d'ordre 2 : ABC.

b) Le modèle de score peut s'écrire :

$$Y_{abcs} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_{s(ab)} + \theta_{ab} + \eta_{ac} + \zeta_{bc} + \psi_{cs(ab)} + \epsilon_{cs(ab)}$$

#### Énoncé 16

Dans une tâche de dénomination de figures géométriques, l'auteur étudie l'évolution du temps de réaction verbale en fonction de la discriminabilité des figures.

Dans un premier temps, on présente aux sujets une série de figures. Pour la moitié d'entre eux, la série est constituée de 2 figures, pour l'autre moitié, de 4 figures. Dans chacun des cas, la série est constituée soit de figures facilement discriminables (triangle, carré, . . .) soit de figures plus complexes (octogone, décagone, . . .).

Dans un deuxième temps, on demande à chaque sujet de nommer une figure tirée au hasard dans la série précédente et on mesure le temps de réaction verbale du sujet.

48 sujets répartis en 4 groupes de 12 ont participé à l'expérience.

Les moyennes des temps de réaction mesurés en millisecondes observés sur chacun des quatre groupes sont indiqués dans le tableau suivant :

| Incertitude | Discriminabilité |        |
|-------------|------------------|--------|
|             | Forte            | Faible |
| 2 figures   | 460              | 510    |
| 4 figures   | 559              | 864    |

- 1) Définir la variable dépendante et les variables indépendantes prises en compte. Quel est le plan d'expérience utilisé ?

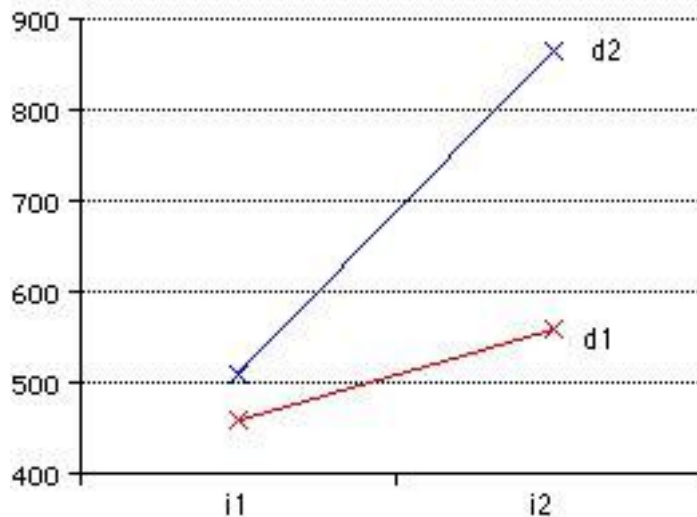
2) Au vu du tableau precedent, indiquer s'il semble y avoir une interaction entre les deux facteurs etudies. Construire un graphe d'interaction. Commenter ce graphe en redigeant une phrase exprimant comment se traduit l'e et d'interaction.

3) a) Deriver ce plan d'experience et donner les differentes sources de variation.

b) Proposer un modele de score.

Elements de reponses : 1) Le plan utilise est ici  $S_{12} < I_2 D_2 >$ .

2) Le temps de reaction augmente lorsque la discriminabilite est plus faible. Mais cet e et est d'autant plus important que l'incertitude est elev .



3) a) Les facteurs elementaires sont ici : D, I, S(DI).

Terme d'interaction d'ordre 1 : DI.

b) Le modele de score peut s'ecrire :  $Y_{dis} = \mu + d + i + di + s_s(di)$ .

#### Enonce 17 Donnees Conrad

Dans une reprise d'une experience de Conrad (1971), on veut mettre en evidence l'hypothese de recherche suivante : "les enfants jeunes n'utilisent pas un codage phonologique en memoire a court terme". Pour ce faire, on selectionne cinq enfants de 5 ans et 5 enfants de 12 ans (Variable A, avec deux modalites). On montre a chaque enfant un certain nombre de paires d'images representant des objets dont on s'est assure auparavant qu'ils sont nommes d'une seule maniere par les enfants. On montre les images aux enfants. Puis on retourne les images (les enfants ne voient plus que le dos des images). Ensuite, on donne aux enfants une paire d'images identiques a celles retournees. En n, on leur demande de placer ces nouvelles images comme les images retournees sur la table. Pour la moitie des paires d'images les noms des objets se ressemblent (e.g., noix et doigt). Pour l'autre moitie, les noms des objets ne se ressemblent pas (e.g., maison et cheval). Conrad predit que les enfants les plus vieux reussiront dans l'ensemble mieux que les enfants les plus jeunes, mais egalement que les enfants les plus vieux utiliseront un codage phonologique comme mnemonique (i.e., "la parole interieure"). De ce fait, les enfants les plus vieux devront commettre plus d'erreurs lorsque les noms se ressemblent acoustiquement que lorsque les noms different. On presente a chaque enfant cinquante paires d'images correspondant a la modalite b1 (dissemblance acoustique), et cinquante paires d'images correspondant a la modalite b2 (ressemblance acoustique) ; la Variable Dependante choisie est le nombre de paires d'images correctement reconstituees. L'ordre de presentation est "aleatorise" pour chaque passation (Pourquoi cette precaution ?).

Essayer de traduire l'hypothese de recherche en prediction sur les sources de variation de l'analyse de variance.

Vous avez d'au conclure que, d'une part, on s'attend a un e et principal de l'age (qui est trivial), et, d'autre part, a un e et d'interaction : c'est le point d'importance, ou si vous preferez, le point crucial de la theorie. On retrouve, ici, le r^ole essentiel de l'interaction \comme test de theorie ".

Enonce 18 Donnees Cochran

Les donnees suivantes, adaptees d'une experience de Cochran et Cox, illustrent un pa-radigme experimental extr^emement courant : la comparaison de deux conditions avec contrebalancement des ordres.

Il s'agissait de comparer l'e cacit de deux types de machines a calculer m1 et m2 : on supposera ici que 10 sujets, s1 a s10, ont execut la m^eme sequence de calculs, successive-ment sur chacune des deux machines m1 et m2. Les sujets s1 a s5 ont travaille d'abord (essai e1) avec la machine m1, puis (essai e2) avec la machine m2 ; les sujets s6 a s10 ont travaille dans l'ordre inverse (m2 a l'essai e1, puis m1 a l'essai e2). Les resultats (temps d'execution du calcul, en unites conventionnelles) sont les suivants :

|    | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 | s9 | s10 |
|----|----|----|----|----|----|----|----|----|----|-----|
| m1 | 30 | 22 | 29 | 12 | 23 | 21 | 22 | 18 | 16 | 23  |
| m2 | 14 | 5  | 17 | 14 | 8  | 21 | 13 | 13 | 7  | 24  |

Comme facteurs decrivant le protocole, nous prendrons d'abord : S (sujets : dix modalites, s1 a s10) ; M ou M2 (Machines : deux modalites m1 et m2) ; E ou E2 (essais : deux modalites e1 et e2). A ce facteur nous adjoindrons, pour des raisons qui appara^tront plus loin, le facteur ordre O ou O2 avec :

o1 : machine m1 passee a l'essai e1 et machine m2 passee a l'essai e2 ;

o2 : machine m2 passee a l'essai e1 et machine m1 passee a l'essai e2 ;

N.B. : le tableau precedent correspond a la description des donnees selon le plan S\*M2 Mais on pourrait egalement presenter ces donnees selon le tableau suivant :

|    | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 | s9 | s10 |
|----|----|----|----|----|----|----|----|----|----|-----|
| e1 | 30 | 22 | 29 | 12 | 23 | 21 | 13 | 13 | 7  | 24  |
| e2 | 14 | 5  | 17 | 14 | 8  | 21 | 22 | 18 | 16 | 23  |

Ce nouveau tableau correspondrait a la description selon le plan S\*E2.

En introduisant le facteur ordre, ces m^emes donnees pourront encore ^etre decrites selon l'un des plans S<O2>\*M2 et S<O2 >\*E2

Question principale : y a-t-il une di erence d'e cacit entre les machines ? Question secondaire : y a-t-il une di erence entre les deux essais ?

Du point de vue des objectifs de la recherche, le facteur Machine sera donc consider comme principal, et le facteur Essai comme secondaire (ce qui n'implique nullement que, lors de la plani cation de l'experience, on s'attendait a ce que l'e et du facteur Essai soit peu important ; les deux ordres ont et contrebalances precisement a n de parer a l'eventualit d'un e et m^eme important du facteur Essai).

Reponses : Il faut ici bien comprendre que le facteur \essai" represente l'interaction entre les facteurs \machine" et \ordre" ; de m^eme, le facteur \machine" represente l'interaction entre les facteurs \ordre" et \essai".

L'analyse, au niveau descriptif, de l'interaction entre les facteurs "ordre" et "essai" (c'est-à-dire l'analyse de l'e et "machine") pourra être faite à partir du tableau suivant obtenu à partir des moyennes calculées dans chacune des conditions  $e_{101}$ ,  $e_{102}$ ,  $e_{201}$ ,  $e_{202}$  :

|       | $o_1$ | $o_2$ | Moy. | Di . |
|-------|-------|-------|------|------|
| $e_1$ | 23.2  | 15.6  | 19.4 | 7.6  |
| $e_2$ | 11.6  | 20.0  | 15.8 | -8.4 |
| Moy.  | 17.4  | 17.8  | 17.6 | -0.4 |
| Di .  | 11.6  | -4.4  | 3.6  | 16   |

Dans ce tableau, 3.6 représente deux fois l'e et "essai", 16 représente deux fois l'e et "machine". L'interaction apparaît clairement sur un graphe d'interaction.

On pourra répondre aux deux questions posées à l'aide de comparaisons de moyennes sur des groupes appariés, en ignorant le troisième facteur. La comparaison des moyennes obtenues pour  $M = m_1$  et  $M = m_2$  aboutit à  $T_{obs} = 3:52$ , valeur significative d'une différence entre machines au seuil de 1%. La comparaison des moyennes obtenues pour  $E = e_1$  et  $E = e_2$  aboutit à  $T_{obs} = 1:09$ . La différence n'est donc pas significative. L'analyse de variance permet ici une étude plus neuve. Mais, le tableau d'analyse de variance est assez complexe, car il s'agit d'un plan à mesures partiellement répétées (sujets imbriqués dans un facteur et croisés avec l'autre facteur). On obtient par exemple :

| Source      | ddl | SC    | CM    | $F_{cal}$ | Pr     |
|-------------|-----|-------|-------|-----------|--------|
| O           | 1   | 0.8   | 0.8   | 0.018     | 0.89   |
| S<O>        | 8   | 357   | 44.6  |           |        |
| E           | 1   | 64.8  | 64.8  | 3.08      | 0.11   |
| Interaction | 1   | 320   | 320   | 15.2      | 0.0045 |
| Residu      | 8   | 168.2 | 21.02 |           |        |
| Total       | 19  | 910.8 |       |           |        |

On voit que l'e et du facteur "essai" est peu significatif (niveau de significativité de 11%) alors que l'interaction (c'est-à-dire l'e et du facteur "machine") est quant à lui très significative. Notez que ce tableau pourrait tout aussi bien être donné sous la forme suivante :

| Source      | ddl | SC    | CM    | $F_{cal}$ | Pr     |
|-------------|-----|-------|-------|-----------|--------|
| O           | 1   | 0.8   | 0.8   | 0.018     | 0.89   |
| S<O>        | 8   | 357   | 44.6  |           |        |
| M           | 1   | 320   | 320   | 15.2      | 0.0045 |
| Interaction | 1   | 64.8  | 64.8  | 3.08      | 0.11   |
| Residu      | 8   | 168.2 | 21.02 |           |        |
| Total       | 19  | 910.8 |       |           |        |

## Plans S A

### Énoncé 19 Données pharma

Dans une expérimentation de psychopharmacologie, on veut vérifier l'e et de deux drogues de type "amphétamines" sur le temps de réaction à une épreuve de psychomotricité. Afin de contrôler une source possible de perturbations, on décide de prendre les six mêmes

sujets exposes a trois traitements experimentaux di erents : Drogue A, Drogue B, Placebo. L'experimentation est construite en "double aveugle". La variable dependante sera le temps de reaction mesur en ms.

Mais, avant tout, remplissez les six etapes du test :

1. Hypotheses Statistiques.
2. Choix du test.
3. Distribution d'echantillonnage.
4. Seuil de Signification.
5. Region de Rejet et Regle de decision.
6. Resultats et decision.

Voici les resultats :

| Sujets         | Condition experimentale |         |          | Total |
|----------------|-------------------------|---------|----------|-------|
|                | Drogue A                | Placebo | Drogue B |       |
| s <sub>1</sub> | 165                     | 231     | 217      | 613   |
| s <sub>2</sub> | 172                     | 219     | 217      | 608   |
| s <sub>3</sub> | 109                     | 199     | 243      | 551   |
| s <sub>4</sub> | 197                     | 219     | 160      | 576   |
| s <sub>5</sub> | 199                     | 247     | 162      | 608   |
| s <sub>6</sub> | 193                     | 245     | 191      | 629   |
| Total          | 1035                    | 1360    | 1190     | 3585  |

$$Q1 = 165 + 172 + \dots + 191 = 3585$$

$$Q2 = 165^2 + \dots + 191^2 = 735819$$

$$Q3 = (1035^2 + \dots + 1190^2) = 6 = 722820:8$$

$$Q4 = (613^2 + \dots + 629^2) = 3 = 715371:7$$

$$Q5 = 3585^2 = 18 = 714012:5$$

$$Q6 = SC_T = Q2 \quad Q5 = 21806:5$$

$$Q7 = SC_A = Q3 \quad Q5 = 8808:3$$

$$Q8 = SC_S = Q4 \quad Q5 = 1359:2$$

$$Q9 = SC_{AS} = Q2 \quad Q4 \quad Q3 + Q5 = 11639$$

Tableau d'ANOVA

| Source | ddl | SC      | CM      | F <sub>cal</sub> | Pr(F <sub>cal</sub> ) |
|--------|-----|---------|---------|------------------|-----------------------|
| A      | 2   | 8808.3  | 4404.15 | 3.784            | .0592                 |
| S      | 5   | 1359.2  | 271.84  |                  |                       |
| AS     | 10  | 11639.0 | 1163.90 |                  |                       |
| Total  | 17  | 21806.5 |         |                  |                       |

Avec la procedure des valeurs critiques : F<sub>critique</sub> = 4:10 (avec 1 = 2; 2 = 10 au seuil = :05). F<sub>cal</sub> < F<sub>critique</sub> on ne peut pas rejeter H<sub>0</sub>.

Enonce 20 Donnees Inhibit

Dans une experimentation sur l'inhibition proactive, des sujets apprennent une liste de dix paires de mots, puis doivent se rappeler ces paires deux jours plus tard. Apres le rappel, les sujets doivent apprendre une deuxieme liste de dix paires dont ils devront se rappeler deux jours plus tard, le rappel de la deuxieme liste est suivie de l'apprentissage d'une

troisième, etc., jusqu'à la sixième liste. La variable indépendante sera la position ordinaire de la liste (e.g., première, seconde, . . . , sixième). La variable dépendante sera le nombre de paires correctement rappelées. Les auteurs de l'expérience prédisent que le rappel se détériorera à mesure que l'on progresse dans la position ordinaire (prédiction qui traduit simplement l'effet et de l'inhibition proactive. . .). Voici les résultats :

| Sujet | Position ordinaire de la liste |     |     |    |    |    | Total |
|-------|--------------------------------|-----|-----|----|----|----|-------|
|       | 1                              | 2   | 3   | 4  | 5  | 6  |       |
| S1    | 17                             | 13  | 12  | 12 | 11 | 11 | 76    |
| S2    | 14                             | 18  | 13  | 18 | 11 | 12 | 86    |
| S3    | 17                             | 16  | 13  | 11 | 15 | 14 | 86    |
| S4    | 18                             | 16  | 11  | 10 | 12 | 10 | 77    |
| S5    | 17                             | 12  | 13  | 10 | 11 | 13 | 76    |
| S6    | 16                             | 13  | 13  | 11 | 11 | 11 | 75    |
| S7    | 14                             | 12  | 10  | 10 | 10 | 10 | 66    |
| S8    | 16                             | 17  | 15  | 11 | 13 | 11 | 83    |
|       | 129                            | 117 | 100 | 93 | 94 | 92 | 625   |

Justifier le tableau d'ANOVA suivant :

| Source | ddl | SC     | CM    | $F_{cal}$ | $Pr(F_{cal})$ |
|--------|-----|--------|-------|-----------|---------------|
| A      | 5   | 146.85 | 29.37 | 10.32**   | .000005       |
| S      | 7   | 52.48  | 7.50  |           |               |
| AS     | 35  | 99.65  | 2.85  |           |               |
| Total  | 47  | 298.98 |       |           |               |

### Énoncé 21 Données Écoute

On brouille l'écoute

Dans une étude sur l'effet et du bruit sur la discrimination perceptive, on utilise six sujets. On mesure pour chaque sujet le nombre d'erreurs commises dans une tâche de discrimination perceptive. Les sujets sont soumis à trois conditions. Dans la première, les sujets accomplissent la tâche en l'absence de bruit ; dans la seconde, le bruit est présent de façon intermittente (i.e., bruits d'avions) ; dans la dernière, le bruit est présent de façon continue (bruits de "marteau piqueur" ) On obtient les résultats suivants :

| Sujets | Absence de bruit | Bruit intermittent | Bruit continu |
|--------|------------------|--------------------|---------------|
| 1      | 117              | 119                | 127           |
| 2      | 130              | 126                | 131           |
| 3      | 122              | 118                | 129           |
| 4      | 123              | 117                | 134           |
| 5      | 126              | 120                | 137           |
| 6      | 116              | 120                | 128           |

Après avoir identifié la ou les variable(s) indépendante(s), dépendante(s), vous répondrez à la question | classique | :

"La (ou les) Variable(s) Independante(s) in ue(nt) elle(s) sur la (les) Variable(s) Dependante(s) ?"

La condition "absence de bruit" di ere-t-elle des conditions "avec bruit" (qu'il soit continu ou intermittent) ?

Les deux conditions "avec bruit" sont elles equivalentes ?

Au vu des resultats, le chercheur remarque la moyenne obtenue dans la condition "bruit intermittent". Il voudrait savoir si cette valeur di ere de la condition "temoin". Comment repondra-t-il a cette question ? Et quelle sera la reponse ?

Reponses. Le tableau d'analyse de variance est donne par :

| Source | ddl | SC     | CM     | $F_{cal}$ | $P r(F_{cal})$ |
|--------|-----|--------|--------|-----------|----------------|
| Bruit  | 2   | 403.11 | 201.56 | 19.98**   | .0003          |
| Sujets | 5   | 164.44 | 32.89  |           |                |
| Residu | 10  | 100.89 | 10.09  |           |                |
| Total  | 17  | 668.44 |        |           |                |

On peut comparer la condition "absence de bruit" aux deux autres conditions en calculant  $L_1 = 2\bar{x}_1 - \bar{x}_2 - \bar{x}_3$  (methode des contrastes orthogonaux, cf. exercice 8). On peut aussi comparer la moyenne de ce protocole deriv a 0. On obtient alors  $x = 6:33$  et  $s_c = 6:74$  d'o  $t_{obs} = 2:30$ . Le resultat n'est pas signi catif d'une di erence dans le cas d'un test bilateral au seuil de 5%. Il l'est dans le cas d'un test unilateral.

On peut comparer les deux conditions "avec bruit" a l'aide d'un test de comparaison de moyennes sur groupes appareilles. Le protocole deriv des di erences individuelles a une moyenne de 11.0, et un ecart type corrige de 5.02. D'o  $t_{obs} = 5:37$ . On conclut donc, au seuil de 5%, que les conditions "avec bruit" ne sont pas equivalentes.

On peut de m^eme comparer les conditions "bruit intermittent" et "temoin". On obtient alors  $t_{obs} = 1:33$ , qui n'est pas signi catif d'une di erence entre les deux conditions

#### Enonce 22 Donnees Craik

Dans une reprise de l'experience de Craik et Tulving (1975), on desire veri er l'hypothese selon laquelle "la profondeur de traitement de l'information" in uence la memorisation. Pour ce faire, on constitue trois groupes de dix sujets. Le premier groupe correspond a la condition "traitement de l'information en profondeur" (i.e., on demande au sujet si le mot present est un synonyme de "jeu"). Au second groupe, on demandera un "traitement de l'information acoustique" (i.e., Le mot present rime-t-il avec "table"). Le troisieme groupe n'e ectuera qu'un traitement super ciel (i.e., le mot est-il ecrit en majuscules ou minuscules).

Chaque mot est present deux fois dans la m^eme condition. On soumet cinquante mots a chaque sujet. On compte pour chaque sujet le nombre de mots retenus apres lui avoir demand de compter "a reculons" de trois en trois a partir de 120 (pourquoi cette derniere precaution ?).

- | Combien y a-t-il de variables independantes ? de variables dependantes ? Identifiez-les.
- | Comment le chercheur traduira-t-il son "hypothese de recherche" en hypothese statistique ?
- | Comment traitera-t-il son experience ?

1) Voici les resultats obtenus (on donne le nombre de mots retenus) ; traitez cette experience.

|    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|
| G1 | 29 | 30 | 33 | 33 | 34 | 34 | 36 | 36 | 40 | 40 |
| G2 | 6  | 10 | 10 | 15 | 15 | 15 | 17 | 17 | 18 | 24 |
| G3 | 1  | 1  | 1  | 2  | 3  | 6  | 7  | 7  | 9  | 10 |

2) Repondez aux m<sup>^</sup>emes questions, mais en admettant qu'une colonne correspond aux resultats d'un m<sup>^</sup>eme sujet soumis aux di erentes conditions.

Reponses :

1) Le tableau d'analyse de variance obtenu est donne par :

| Source | ddl | SC      | CM      | F <sub>cal</sub> |
|--------|-----|---------|---------|------------------|
| Groupe | 2   | 4600,27 | 2300,13 | 135.39           |
| Residu | 27  | 458,70  | 16,99   |                  |
| Total  | 29  | 5058.97 |         |                  |

Le F obtenu est tres signi catif d'une di erence de comportement entre les trois groupes

2) Remarquez que les donnees fournies sont fort peu realistes dans ce cas (il est hautement improbable que les sujets soient ranges exactement dans le m<sup>^</sup>eme ordre pour les trois conditions experimentales).

Dans le cas de groupes appareilles (cf. infra d'autres situations de ce genre), la variation intra-groupes se decompose en une variation due aux sujets et un residu. On obtient le tableau d'analyse de variance suivant :

| Source | ddl | SC      | CM      | F <sub>cal</sub> |
|--------|-----|---------|---------|------------------|
| Groupe | 2   | 4600,27 | 2300,13 | 1024.81          |
| Sujet  | 9   | 418,30  | 46.48   |                  |
| Residu | 18  | 40.40   | 2.24    |                  |
| Total  | 29  | 5058.97 |         |                  |

Le F obtenu est evidemment tres signi catif d'une di erence de comportement selon la condition experimentale.

## Plans S < A B >

Enonce 23 Dossier \Geometrie"

Dans une t<sup>^</sup>ache de denomination de gures geometriques, l'auteur etudie l'evolution du temps de reaction verbale en fonction de la discriminabilite des gures.

Dans un premier temps, on presente aux sujets une serie de gures. Pour la moitie d'entre eux, la serie est constituee de 2 gures, pour l'autre moitie, de 4 gures. Dans chacun des cas, la serie est constituee soit de gures facilement discriminables (triangle, carre, . . .) soit de gures plus complexes (octogone, decagone. . .).

Dans un deuxieme temps, on demande a chaque sujet de nommer une gure tiree au hasard dans la serie precedente et on mesure le temps de reaction verbale du sujet.

48 sujets repartis en 4 groupes de 12 ont participe a l'experience.

Les moyennes des temps de reaction mesures en millisecondes observes sur chacun des quatre groupes sont indiquees dans le tableau suivant :

| Incertitude | Discriminabilite |        |
|-------------|------------------|--------|
|             | Forte            | Faible |
| 2 gures     | 460              | 510    |
| 4 gures     | 559              | 864    |



- 1) Déterminer la variable dépendante et les variables indépendantes prises en compte. Quel est le plan d'expérience utilisé ?
- 2) Au vu du tableau précédent, indiquer s'il semble y avoir une interaction entre les deux facteurs étudiés. Construire un graphe d'interaction. Commenter ce graphe en rédigeant une phrase exprimant comment se traduit l'effet et d'interaction.
- 3) Le tableau d'analyse de variance se présente ainsi :

| Sources de var. | ddl | SC      | CM     | F     |
|-----------------|-----|---------|--------|-------|
| Discriminalité  | 1   | 3858.3  | 3858.3 | 45,06 |
| Incertitude     | 1   | 6238.3  | 6238.3 | 72,85 |
| Interaction     | 1   | 1885.4  | 1885.4 | 22,02 |
| Residu          | 44  | 3767.6  | 85.6   |       |
| Total           | 47  | 15666.7 |        |       |

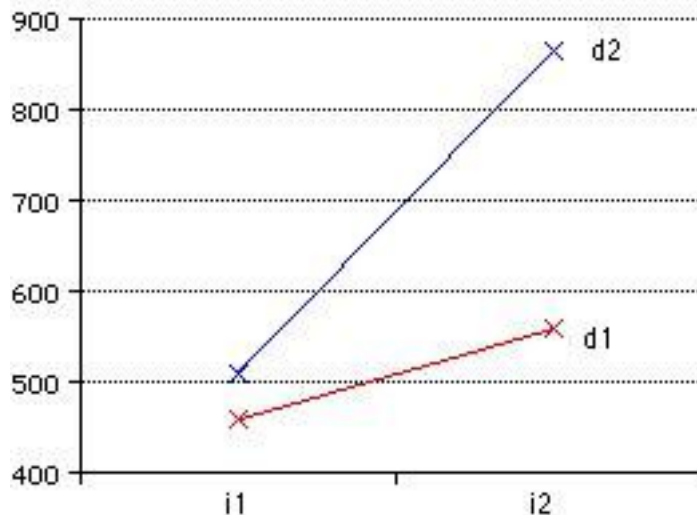
Préciser comment ont été obtenues :

- la valeur 85.6 dans la ligne "residu" ;
- la valeur 45.06 dans la ligne "discriminalité".

Utiliser la table de la loi de Fisher-Snedecor pour indiquer si les effets principaux et l'effet d'interaction sont significatifs au seuil de 1%.

- 4) Aurait-on pu (au moins partiellement) traiter ces données par des méthodes de comparaison de moyennes ?

- 2) Le temps de réaction augmente lorsque la discriminalité est plus faible. Mais cet effet est d'autant plus important que l'incertitude est élevée.



- 3)  $85,6 = \frac{3767,6}{44} ; 45,06 = \frac{3858,3}{85,6}$ . Au seuil de 1%,  $F_{crit}(1; 44) = 7,2$ . Les effets principaux et l'effet d'interaction sont donc significatifs.

- 4) Les effets principaux auraient pu être facilement étudiés par des méthodes de comparaison de moyennes, puisque les facteurs ne comportent que deux modalités. En revanche, il aurait été difficile d'étudier l'interaction.

#### Énoncé 24 Données Tulving

On demande aux sujets de mémoriser des listes comportant 12, 24 ou 48 mots (facteur A, avec trois modalités). Ces mots peuvent se regrouper par paires en catégories (par exemple pomme et orange se regroupent en "fruits"). On demande aux sujets d'apprendre

les mots, et on leur montre le nom des categories a ce moment en leur precisant qu'ils n'ont pas a apprendre le nom de ces categories. Au moment de l'epreuve de rappel | qui a lieu immediatement apres l'apprentissage | on cree deux conditions. Dans un cas, on presente aux sujets la liste des categories. Dans l'autre cas, on ne leur presente pas cette liste (facteur B : presentation de la liste des categories au moment de l'apprentissage versus absence de presentation). Dans cette reprise d'une experience de Tulving et Pearlstone (1966), la variable dependante sera le nombre de mots rappelés. En examinant les deux variables independantes, la premiere (nombre de mots de la liste) est, clairement, triviale : il semble superfluo de construire une experimentation pour montrer que plus une liste de mots est longue, plus on peut en retenir. Cette remarque indique que les auteurs de cette experience s'interessaient d'emblee a un e et d'interaction.

On interroge dix sujets par condition experimentale. Voici les resultats :

Resultats d'une reprise de l'experience de Tulving et Pearlstone.

|                | Facteur A : Nombre de mots par liste |    |                     |    |                     |    |
|----------------|--------------------------------------|----|---------------------|----|---------------------|----|
| Facteur B      | a <sub>1</sub> : 12                  |    | a <sub>2</sub> : 24 |    | a <sub>3</sub> : 48 |    |
| b <sub>1</sub> | 10                                   | 6  | 13                  | 15 | 17                  | 16 |
|                | 8                                    | 11 | 18                  | 13 | 20                  | 23 |
|                | 12                                   | 10 | 19                  | 9  | 22                  | 19 |
|                | 8                                    | 9  | 13                  | 8  | 13                  | 20 |
|                | 7                                    | 9  | 8                   | 14 | 21                  | 19 |
| b <sub>2</sub> | 12                                   | 10 | 12                  | 13 | 31                  | 29 |
|                | 12                                   | 12 | 20                  | 12 | 30                  | 32 |
|                | 7                                    | 10 | 19                  | 13 | 26                  | 24 |
|                | 9                                    | 7  | 14                  | 15 | 29                  | 24 |
|                | 9                                    | 12 | 16                  | 6  | 28                  | 27 |

On peut resumer ces resultats dans une matrice des moyennes :

|                  | Facteur A : Nombre de mots par liste |                     |                     |       |
|------------------|--------------------------------------|---------------------|---------------------|-------|
| Facteur B        | a <sub>1</sub> : 12                  | a <sub>2</sub> : 24 | a <sub>3</sub> : 48 | Marge |
| b <sub>1</sub> : |                                      |                     |                     |       |
| Moyenne          | 9                                    | 13                  | 19                  | 13.47 |
| Total            | 90                                   | 130                 | 190                 | 410   |
| b <sub>2</sub> : |                                      |                     |                     |       |
| Moyenne          | 10                                   | 14                  | 28                  | 17.33 |
| Total            | 100                                  | 140                 | 280                 | 520   |
| Marge :          |                                      |                     |                     |       |
| Moyenne          | 9.5                                  | 13.5                | 23.5                | 15.5  |
| Total            | 190                                  | 270                 | 470                 | 930   |

Avant de commencer les calculs, construire un graphe d'interaction entre les deux variables independantes considerees. L'examen de la gure suggere l'existence d'un e et du nombre de mots de la liste, d'un e et de la presentation d'indices lors du rappel, et egalement d'un e et d'interaction : l'e et facilitateur des indices se manifeste essentiellement pour la liste de grande taille. Les calculs permettent d'evaluer ces e ets par une approche inferentielle. Justifier le tableau d'ANOVA suivant :

| Source | ddl | SC      | CM      | $F_{cal}$ | $P r(F_{cal})$ |
|--------|-----|---------|---------|-----------|----------------|
| A      | 2   | 2080.00 | 1040.00 | 115,57 ** | $<10^{-10}$    |
| B      | 1   | 201,67  | 201,67  | 22,41 **  | .000029        |
| AB     | 2   | 213,33  | 106,67  | 11,85 **  | .000074        |
| S(AB)  | 54  | 486.00  |         |           |                |
| Total  | 59  | 2981    |         |           |                |

L'analyse de variance permet de mettre en évidence un effet et trivial par ailleurs | du nombre de mots de la liste à mémoriser ( $F_{cal}(2; 54) = 115,57$ ;  $p < :01$ ) sur le nombre de mots retenus. La présentation d'indices lors du rappel améliore la performance des sujets ( $F_{cal}(1; 54) = 22,41$ ;  $p < :01$ ). Mais surtout, on note une interaction significative entre les deux Variables Indépendantes ( $F_{cal}(2; 54) = 11,85$ ;  $p < :01$ ). Cette interaction pouvant s'attribuer pour l'essentiel au fait que l'effet est facilitateur des indices ne se manifeste que pour les longues listes (48 mots). De ce fait, une expérimentation construite pour montrer l'effet et des indices au moment du rappel avec uniquement des listes courtes, ne pourrait { probablement } pas rapporter un effet et significatif de ce facteur.

#### Énoncé 25 Dossier "Eysenck"

Le modèle de la mémorisation proposé par Craik et Lockhart (1972) stipule que le degré auquel un sujet se rappelle un matériel verbal est fonction du degré auquel ce matériel a été traité lors de sa présentation initiale. Eysenck (1974) voulait tester ce modèle et examiner s'il pouvait contribuer à expliquer certaines différences relevées entre des sujets jeunes et âgés concernant leur aptitude à se rappeler du matériel verbal. L'étude qu'il a menée incluait 50 sujets dont l'âge se situait entre 18 et 30 ans et 50 sujets compris dans la tranche d'âge 55-65 ans. Dans chacune des tranches d'âge, Eysenck a réparti les 50 sujets dans cinq groupes. Le premier devait lire une liste de mots et se contenter de compter le nombre de lettres de chacun d'eux. Le deuxième groupe devait lire chaque mot et lui trouver une rime. Le troisième groupe devait donner un adjectif qui aurait pu être utilisé pour modifier chaque mot de la liste. Le quatrième devait essayer de se former une image précise de chaque mot. Aucun de ces quatre groupes ne savait qu'il faudrait se rappeler les mots ultérieurement. Enfin, le cinquième groupe, ou groupe d'apprentissage intentionnel, devait lire la liste et mémoriser tous les mots. Après avoir passé trois fois en revue la liste de 27 mots, les sujets devaient retranscrire tous les mots dont ils se souvenaient. Le nombre de mots rappelés par chacun des 100 sujets est indiqué par le tableau ci-dessous :

|                          | Gr. 1 | Gr. 2 | Gr. 3 | Gr. 4 | Gr. 5 |
|--------------------------|-------|-------|-------|-------|-------|
| Sujets <sup>^</sup> ages | 9     | 7     | 11    | 12    | 10    |
|                          | 8     | 9     | 13    | 11    | 19    |
|                          | 6     | 6     | 8     | 16    | 14    |
|                          | 8     | 6     | 6     | 11    | 5     |
|                          | 10    | 6     | 14    | 9     | 10    |
|                          | 4     | 11    | 11    | 23    | 11    |
|                          | 6     | 6     | 13    | 12    | 14    |
|                          | 5     | 3     | 13    | 10    | 15    |
|                          | 7     | 8     | 10    | 19    | 11    |
|                          | 7     | 7     | 11    | 11    | 11    |
| Sujets jeunes            | 8     | 10    | 14    | 20    | 21    |
|                          | 6     | 7     | 11    | 16    | 19    |
|                          | 4     | 8     | 18    | 16    | 17    |
|                          | 6     | 10    | 14    | 15    | 15    |
|                          | 7     | 4     | 13    | 18    | 22    |
|                          | 6     | 7     | 22    | 16    | 16    |
|                          | 5     | 10    | 17    | 20    | 22    |
|                          | 7     | 6     | 16    | 22    | 22    |
|                          | 9     | 7     | 12    | 14    | 18    |
|                          | 7     | 7     | 11    | 19    | 21    |

1) a) Quelles sont les variables independantes (ou facteurs de variation) prises en compte ? Quel est le nombre de niveaux de chacun des facteurs ?

b) Quelle est la variable dependante ? Quel est son domaine de variation ? c) Ecrire le plan d'experience correspondant.

2) a) On veut etudier chez les sujets ag<sup>^</sup>es s'il existe une difference de performance entre le groupe 2 (traitement syntaxique) et le groupe 3 (traitement semantique), en faveur de ce dernier. Le calcul permet d'obtenir les resultats de statistiques descriptives suivants :

|                    | Groupe 2 | Groupe 3 |
|--------------------|----------|----------|
| Moyenne            | 6.9      | 11.0     |
| Ecartif            | 10       | 10       |
| Ecart-type         | 2.02     | 2.37     |
| Ecart-type corrige | 2.13     | 2.49     |

b) Etudier de m<sup>^</sup>eme s'il existe une difference de performance due a l'<sup>^</sup>age parmi les sujets du groupe 2.

3) Etude de l'interaction entre les facteurs. Le tableau suivant indique les moyennes observees pour chacune des deux tranches d'<sup>^</sup>age, dans chacun des groupes.

|                          | Gr. 1 | Gr. 2 | Gr. 3 | Gr. 4 | Gr. 5 |
|--------------------------|-------|-------|-------|-------|-------|
| Sujets <sup>^</sup> ages | 7.0   | 6.9   | 11.0  | 13.4  | 12.0  |
| Sujets jeunes            | 6.5   | 7.6   | 14.8  | 17.6  | 19.3  |

Realiser un graphe illustrant une eventuelle interaction entre l'<sup>^</sup>age et la condition de memorisation (groupe). Commenter le diagramme ainsi obtenu.

4) Analyse de variance. Le tableau d'analyse de variance relatif aux donnees observees se presente ainsi :

| Sources de var. | ddl | SC      | CM     | F   |
|-----------------|-----|---------|--------|-----|
| Age             | 1   | 240.25  | 240.25 | ... |
| Groupe          | 4   | 1514.94 | ...    | ... |
| Age Groupe      | 4   | 190.30  | ...    | ... |
| Residu          | 90  | 722.30  | 8.026  |     |
| Total           | 99  | 2667.79 |        |     |

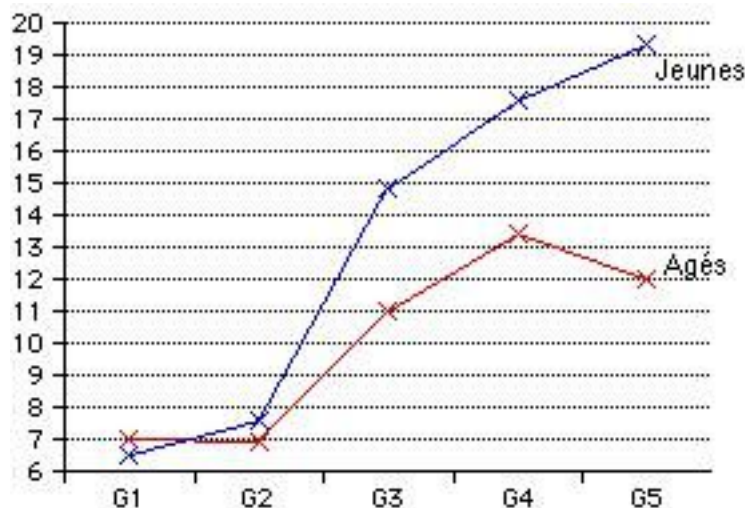
- a) Compléter ce tableau en calculant les carrés moyens et les statistiques F de Fisher qui sont remplacés par "...".
- b) Parmi les différentes sources de variation, quelles sont celles qui sont significatives au seuil de 1% ?
- c) Quelles conclusions Eysenck peut-il tirer de cette expérience ?

1) Outre le facteur sujet, les facteurs étudiés sont l'âge (2 niveaux), et le groupe expérimental (5 niveaux), selon un plan  $S_{10} < G_5 A_2 >$ . La variable dépendante est le nombre de mots retranscrits, son domaine de variation est  $[0; 27]$ .

2) a) Il s'agit d'une comparaison de moyennes sur deux groupes indépendants. Avec les données fournies, on obtient  $t_{obs} = 3:80$  alors que, pour un test bilatéral au seuil de 5%, on obtient  $t_{crit} = 2:10$  (ddl = 18). Il existe donc une différence significative entre le traitement syntaxique et le traitement sémantique.

b) Les sujets du groupe 2 forment deux sous-groupes indépendants du point de vue de l'âge. On obtient  $t_{obs} = 0:766$ . La différence de performance n'est pas significative dans ce cas.

3) Le graphe d'interaction est donné par :



D'après ce graphique, il semble que l'effet de l'âge soit plus marqué lorsque le traitement de l'information est effectué en profondeur.

4) Le tableau d'ANOVA complet est donné par :

| Sources de var. | ddl | SC      | CM     | F     |
|-----------------|-----|---------|--------|-------|
| Age             | 1   | 240.25  | 240.25 | 29.93 |
| Groupe          | 4   | 1514.94 | 378.74 | 47.19 |
| Age Groupe      | 4   | 190.30  | 47.58  | 5.93  |
| Residu          | 90  | 722.30  | 8.026  |       |
| Total           | 99  | 2667.79 |        |       |

Au seuil de 1%, on a :  $F_{crit}(1; 90) = 6:9$  ;  $F_{crit}(4; 90) = 3:5$ . Les effets du groupe, de l'âge et l'effet de l'interaction sont donc tous trois significatifs.

Ainsi, les sujets plus jeunes se rappellent davantage d'éléments que les sujets plus âgés. Les tâches impliquant un traitement plus approfondi permettent une mémorisation plus efficace que celles qui impliquent un traitement plus superficiel. Cependant, l'effet significatif de l'interaction montre que les sujets plus âgés ne réalisent pas d'aussi bonnes performances que les plus jeunes dans les tâches qui impliquent un traitement approfondi, mais réalisent des performances pour ainsi dire équivalentes à celles des sujets plus jeunes lorsque la tâche n'implique qu'un traitement réduit.

#### Énoncé 26 Dossier "Locus"

Un psychologue s'intéresse à la relation entre le sexe (variable X), le statut socio-économique (variable C) et le "locus of control" perçu. Il a pris huit adultes de chaque combinaison sexe-statut socio-économique et leur a administré une échelle portant sur le "locus of control" ; un score élevé indique que le sujet estime contrôler sa vie quotidienne.

|        | statut socio-économique |       |       |
|--------|-------------------------|-------|-------|
|        | Bas                     | Moyen | Élevé |
| Hommes | 10                      | 16    | 18    |
|        | 12                      | 12    | 14    |
|        | 8                       | 19    | 17    |
|        | 14                      | 17    | 13    |
|        | 10                      | 15    | 19    |
|        | 16                      | 11    | 15    |
|        | 15                      | 14    | 22    |
|        | 13                      | 10    | 20    |
| Femmes | 8                       | 14    | 12    |
|        | 10                      | 10    | 18    |
|        | 7                       | 13    | 14    |
|        | 9                       | 9     | 21    |
|        | 12                      | 17    | 19    |
|        | 5                       | 15    | 17    |
|        | 8                       | 12    | 13    |
|        | 7                       | 8     | 16    |

- 1) a) Quelles sont les variables indépendantes (ou facteurs de variation) prises en compte ? Quel est le nombre de niveaux de chacun des facteurs ?  
 b) Quelle est la variable dépendante ? Quel est son domaine de variation ? c) Écrire le plan d'expérience correspondant.
- 2) On veut étudier, pour les sujets de statut socio-économique moyen, s'il existe une différence de "locus of control" entre les hommes et les femmes.  
 Réaliser un test de comparaison de moyennes permettant d'apporter une réponse à la question posée (seuil choisi : 5%).
- 3) Étude de l'interaction entre les facteurs. Calculer les moyennes des scores observés sur chacun des 6 groupes. Réaliser un graphique illustrant une éventuelle interaction entre le sexe et le statut socio-économique. Commenter le diagramme ainsi obtenu.
- 4) Analyse de variance. Le tableau d'analyse de variance relatif aux données observées se présente ainsi :

| Sources de var. | ddl | SC     | CM    | F   |
|-----------------|-----|--------|-------|-----|
| Sexe            | 1   | 65.33  | 65.33 | ... |
| Statut soc-eco  | 2   | 338.67 | ...   | ... |
| X C             | 2   | 18.67  | ...   | ... |
| Residu          | 42  | 355.0  | 8.45  |     |
| Total           | 47  | 777.67 |       |     |

- a) Compléter ce tableau en calculant les carrés moyens et les statistiques F de Fisher qui sont remplacés par "." dans le tableau ci-dessus.
- b) Parmi les différentes sources de variation, quelles sont celles qui sont significatives au seuil de 1% ?

5) Quelles conclusions le psychologue peut-il tirer de cette expérience ? Réponses.

1) On étudie ici les facteurs "sexe" (variable X à deux niveaux) et "statut socio-économique" (variable C à trois niveaux). La variable dépendante est une échelle évaluant le "locus of control" des sujets. Les valeurs observées se situent dans l'intervalle [5; 22]. Il s'agit d'un plan  $S_8 \times X_2 \times C_3$ .

2) Il s'agit d'une comparaison de deux moyennes sur des groupes indépendants. On obtient :  $t_{obs} = 1:29$ , avec 14 ddl. Pour un test unilatéral au seuil de 5%,  $t_{crit} = 1:76$ . On conclut donc à une différence selon les sexes.

3) Le tableau d'analyse de variance complet est donné par :

| Sources de var. | ddl | SC     | CM     | F       |
|-----------------|-----|--------|--------|---------|
| Sexe            | 1   | 65.33  | 65.33  | 7.73**  |
| Statut soc-eco  | 2   | 338.67 | 169.33 | 20.03** |
| X C             | 2   | 18.67  | 9.33   | 1.10 NS |
| Residu          | 42  | 355.0  | 8.45   |         |
| Total           | 47  | 777.67 |        |         |

On conclut donc à des effets significatifs du sexe et du statut socio-économique sur le "locus of control". En revanche, il ne semble pas y avoir d'interaction entre ces deux facteurs : chez les femmes, la mesure du "locus of control" fournit des résultats inférieurs à ceux des hommes, et l'amplitude de cette différence est la même, quel que soit le statut socio-économique.

Énoncé 27 Dossier "Multimedia"

Le multimedia offre certaines potentialités pour induire l'imagerie mentale. Selon certains auteurs, on devrait donc s'attendre à une efficacité supérieure du multimedia en termes d'apprentissage, comparativement aux produits traditionnels. Dubois et al. (1998) ont mené une expérimentation visant à identifier les effets des différents formats de présentation de l'information sur l'apprentissage d'une langue étrangère.

Ils s'attendaient à ce que les sujets produisent un meilleur rappel lorsque l'information verbale est accompagnée d'une information graphique. En revanche, la simple présence de multiples sources d'informations devrait provoquer un partage de l'attention. On s'attend donc à de moins bonnes performances si la présentation, en ajoutant une simple illustration, ne permet pas d'intégration des informations entre elles.

Pour leur expérience, les auteurs ont utilisé 60 sujets auxquels était proposée une tâche de mémorisation d'un vocabulaire russe. Quatre groupes de 15 sujets ont été constitués selon les quatre modes de présentation de l'information suivants :

- { En condition contr<sup>o</sup>le P<sub>1</sub>, seuls le mot russe et sa traduction etaient donnees ;
- { En condition P<sub>2</sub>, une illustration representant le mot etait ajoutee ;
- { En conditions P<sub>3</sub> et P<sub>4</sub>, une methode suscitant une imagerie mentale imposee a et utilisee, selon la technique du mot-cle. La phrase contenant ce mot-cle etait presentee de facon uniquement orale en P<sub>3</sub>, et a l'ecrit sur l'ecran en P<sub>4</sub>.

Le rappel du vocabulaire consistait a presenter aux sujets un mot russe pour lequel ils devaient trouver et ecrire la traduction. Au sein de chaque groupe, trois sous-groupes homogenes ont et constituees, selon le mode de rappel utilise ; le mot est donne :

- { soit seulement a l'ecrit (rappel visuel Rv),
- { soit seulement a l'oral (rappel auditif Ra)
- { soit a l'ecrit et a l'oral (rappel audiovisuel Rav).

La performance des sujets est mesuree par le nombre de traductions correctes fournies (score de 0 a 19).

Dans une reprise de cette experience, les resultats observes sont les suivants :

|     | P <sub>1</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>4</sub> |
|-----|----------------|----------------|----------------|----------------|
| Rv  | 0              | 0              | 0              | 2              |
|     | 2              | 3              | 0              | 0              |
|     | 4              | 8              | 0              | 1              |
|     | 6              | 6              | 0              | 6              |
|     | 0              | 6              | 3              | 8              |
| Ra  | 7              | 1              | 10             | 4              |
|     | 6              | 4              | 13             | 7              |
|     | 3              | 8              | 17             | 8              |
|     | 0              | 4              | 15             | 9              |
|     | 3              | 6              | 18             | 13             |
| Rav | 5              | 0              | 4              | 4              |
|     | 12             | 5              | 7              | 12             |
|     | 7              | 2              | 15             | 5              |
|     | 10             | 0              | 12             | 10             |
|     | 7              | 3              | 14             | 9              |

- 1) a) Quelles sont les variables independantes (ou facteurs de variation) prises en compte ? Quel est le nombre de niveaux de chacun des facteurs ?
  - b) Quelle est la variable dependante ?
  - c) Ecrire le plan d'experience correspondant.
- 2) a) Calculer les moyennes correspondant aux 12 conditions experimentales de nies par les combinaisons des variables P et R.
  - b) Realiser un graphe illustrant une eventuelle interaction entre les variables P et R.
- 3) Analyse de variance.
- Le tableau d'analyse de variance relatif aux donnees observees se presente ainsi :



| Sources de var. | ddl | SC    | CM   | F   |
|-----------------|-----|-------|------|-----|
| P               | 3   | 198.6 | 66.2 | ... |
| R               | 2   | 301.9 | ...  | ... |
| P R             | 6   | 402.1 | ...  | ... |
| Residu          | 48  | 446.8 | 9.31 |     |
| Total           | 59  | 1349  |      |     |

a) Compléter ce tableau en calculant les carrés moyens et les statistiques F de Fisher qui sont remplacés par \. . ." dans le tableau ci-dessus.

b) En utilisant un seuil de 5%, répondre aux questions suivantes :

{ La variable "mode de présentation" a-t-elle un effet ?

{ La variable "mode de rappel" a-t-elle un effet ?

{ L'interaction entre ces variables est-elle significative ?

4) Comparaisons de moyennes

On donne les résultats intermédiaires suivants :

|               | Rv      | Ra      | Rav     |
|---------------|---------|---------|---------|
| $\sum x_i$    | 55      | 156     | 143     |
| $\sum x_i^2$  | 315     | 1722    | 1401    |
|               |         |         |         |
| Var. corrigée | 8.61842 | 26.5895 | 19.9237 |

a) Sans tenir compte des modalités de la variable P, effectuer un test de comparaison de deux moyennes visant à montrer que les 20 sujets soumis à la modalité Rav obtiennent de meilleurs résultats que ceux soumis à la modalité Rv.

b) Comparer de même les sujets soumis à la modalité Ra à ceux soumis à la modalité Rv.

5) À partir des données et des éléments d'étude développés ci-dessus, justifier les conclusions suivantes formulées par les auteurs :

"On constate l'importance de certaines modalités de présentation de l'information sur la mémorisation. (...) L'ajout d'une image à un corpus sonore et textuel peut constituer une aide notable pour les sujets sous certaines conditions."

"La quantité d'information à traiter par le sujet n'apparaît pas induire de partage attentionnel limitant les effets d'apprentissage selon l'hypothèse de surcharge cognitive dès lors que les différentes sources d'information sont intégrées."

"Pour les trois situations de rappel, le résultat le plus intéressant à noter est la moindre performance en situation uniquement visuelle."

"Dans la situation de rappel auditif, on observe d'une manière générale un meilleur apprentissage lorsque les informations sont intégrées sous forme auditive."

Reponses.

1) Outre le facteur sujet, les facteurs étudiés sont le mode de présentation de l'information (facteur P à 4 niveaux notés P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub> et P<sub>4</sub>) et le mode de rappel (facteur R à trois niveaux notés R<sub>v</sub>, R<sub>a</sub> et R<sub>av</sub>). La variable dépendante est le nombre de traductions correctes fournies. L'expérience a été menée selon le plan : S<sub>5</sub> < P<sub>4</sub> R<sub>3</sub> >.

2) Les moyennes correspondant aux 12 conditions experimentales sont donnees par :

|                 | P <sub>1</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>4</sub> |
|-----------------|----------------|----------------|----------------|----------------|
| R <sub>v</sub>  | 2.4            | 4.6            | 0.6            | 3.4            |
| R <sub>a</sub>  | 3.8            | 4.6            | 14.6           | 8.2            |
| R <sub>av</sub> | 8.2            | 2              | 10.4           | 8              |

Le graphe d'interaction a l'allure suivante :

3) Le tableau d'analyse de variance complet donne :

| Sources de var. | ddl | SC    | CM     | F     |
|-----------------|-----|-------|--------|-------|
| P               | 3   | 198.6 | 66.2   | 7.11  |
| R               | 2   | 301.9 | 150.95 | 16.21 |
| P R             | 6   | 402.1 | 67.02  | 7.20  |
| Residu          | 48  | 446.8 | 9.31   |       |
| Total           | 59  | 1349  |        |       |

Pour le facteur P, les nombres de degres de libert a prendre en compte sont  $ddl_1 = 3$  et  $ddl_2 = 48$ . Au seuil de 5%, la table du F de Fisher donne  $F_{crit} = 2:84$ . L'e et du facteur P est donc signi catif.

Pour le facteur R, les nombres de degres de libert a prendre en compte sont  $ddl_1 = 2$  et  $ddl_2 = 48$ . Au seuil de 5%, la table du F de Fisher donne  $F_{crit} = 3:23$ . L'e et du facteur R est donc signi catif.

Pour l'interaction P R, les nombres de degres de libert a prendre en compte sont  $ddl_1 = 6$  et  $ddl_2 = 48$ . Au seuil de 5%, la table du F de Fisher donne  $F_{crit} = 2:34$ . L'e et d'interaction est donc signi catif.

4) a) Il s'agit ici d'une comparaison de moyennes sur deux groupes independants. On obtient, en utilisant les resultats fournis,  $t_{obs} = 3:68$ . Ici,  $ddl = 38$  ; pour un seuil de 5% unilatéral, la table du T de Student donne  $t_{crit} = 1:686$ . On obtient donc de meilleurs resultats en modalite R<sub>av</sub> qu'en modalite R<sub>v</sub>.

b) La methode est identique. On obtient ici  $t_{obs} = 3:81$  et une conclusion analogue.

5) On a montré un effet du facteur P, ce qui justifie en partie la première phrase de conclusion. Une étude complémentaire devrait montrer que les différences constatées entre les 4 présentations se font au bénéfice de P<sub>2</sub> et P<sub>4</sub>.

La deuxième phrase reprend l'une des hypothèses de recherche. Les moins bonnes performances ont effectivement été observées lorsque l'illustration est simplement ajoutée, alors que la situation P<sub>4</sub>, dans laquelle l'illustration est intégrée aux autres sources conduit à des résultats généralement supérieurs à ceux de la situation contrôle.

La troisième phrase fait référence aux conclusions trouvées dans la question 4 : nous y avons montré que la modalité R<sub>v</sub> obtenait des résultats inférieurs à chacune des deux autres modalités.

La dernière phrase traduit l'interaction entre les facteurs P et R. C'est effectivement le groupe soumis à P<sub>3</sub> et R<sub>a</sub> qui obtient le meilleur résultat absolu.

## Plans $S < A > B$

### Enonce 28 Donnees Conrad

Dans une reprise d'une experience de Conrad (1971), on veut mettre en evidence l'hypothese de recherche suivante : "les enfants jeunes n'utilisent pas un codage phonologique en memoire a court terme". Pour ce faire, on selectionne cinq enfants de 5 ans et 5 enfants de 12 ans (Variable A, avec deux modalites). On montre a chaque enfant un certain nombre de paires d'images representant des objets dont on s'est assure auparavant qu'ils sont nommes d'une seule maniere par les enfants. On montre les images aux enfants. Puis on retourne les images (les enfants ne voient plus que le dos des images). Ensuite, on donne aux enfants une paire d'images identiques a celles retournees. En n, on leur demande de placer ces nouvelles images comme les images retournees sur la table. Pour la moitie des paires d'images les noms des objets se ressemblent (e.g., noix et doigt). Pour l'autre moitie, les noms des objets ne se ressemblent pas (e.g., maison et cheval). Conrad predit que les enfants les plus vieux reussiront dans l'ensemble mieux que les enfants les plus jeunes, mais egalement que les enfants les plus vieux utiliseront un codage phono-logique comme mnemonique (i.e., "la parole interieure"). De ce fait, les enfants les plus vieux devront commettre plus d'erreurs lorsque les noms se ressemblent acoustiquement que lorsque les noms different. On presente a chaque enfant cinquante paires d'images correspondant a la modalite b1 (dissemblance acoustique), et cinquante paires d'images correspondant a la modalite b2 (ressemblance acoustique) ; la Variable Dependante choisie est le nombre de paires d'images correctement reconstituees. L'ordre de presentation est "aleatorise" pour chaque passation (Pourquoi cette precaution ?).

Essayer de traduire l'hypothese de recherche en prediction sur les sources de variation de l'analyse de variance.

Vous avez d'ailleurs conclu que, d'une part, on s'attend a un effet principal de l'age (qui est trivial), et, d'autre part, a un effet d'interaction : c'est le point d'importance, ou si vous preferez, le point crucial de la theorie. On retrouve, ici, le role essentiel de l'interaction "comme test de theorie".

| Resultats de l'experience |      |      |       |
|---------------------------|------|------|-------|
|                           | a1b1 | a1b2 | Somme |
| s1                        | 15   | 14   | 29    |
| s2                        | 23   | 20   | 43    |
| s3                        | 12   | 11   | 23    |
| s4                        | 16   | 17   | 33    |
| s5                        | 14   | 13   | 27    |
|                           | 80   | 75   | 155   |
|                           | a2b1 | a2b2 | Somme |
| s6                        | 40   | 33   | 73    |
| s7                        | 38   | 23   | 61    |
| s8                        | 31   | 21   | 52    |
| s9                        | 36   | 26   | 62    |
| s10                       | 30   | 22   | 52    |
|                           | 175  | 125  | 300   |
|                           | 255  | 200  | 455   |

Calcul en 13 points

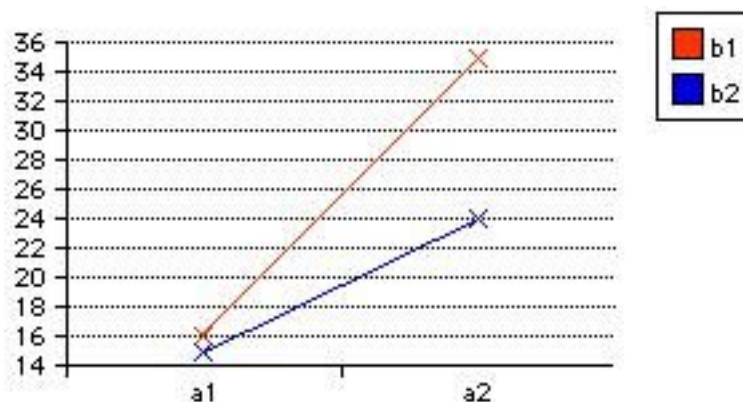
$$\begin{aligned}
 Q1 &= 15 + \dots + 22 = 455 \\
 Q2 &= 15^2 + \dots + 22^2 = 11945 \\
 Q3 &= (15^2 + 300^2) = 10 = 11402:5 \\
 Q4 &= (255^2 + 200^2) = 10 = 10502:5 \\
 Q5 &= (80^2 + \dots + 125^2) = 5 = 11655 \\
 Q6 &= (29^2 + \dots + 52^2) = 2 = 11669:5 \\
 Q7 &= 455^2 = 20 = 10351:25 \\
 Q8 &= SC_A = Q3 \quad Q7 = 1051:25 \\
 Q9 &= SC_B = Q4 \quad Q7 = 151:25 \\
 Q10 &= SC_{S(A)} = Q6 \quad Q3 = 267 \\
 Q11 &= SC_{AB} = Q5 \quad Q3 \quad Q4 + Q7 = 101:15 \\
 Q12 &= SC_{BS(A)} = Q2 \quad Q5 \quad Q6 + Q3 = 23 \\
 Q13 &= SC_{total} = Q2 \quad Q7 = 1593:75
 \end{aligned}$$

| Source  | ddl | SC      | CM      | F <sub>cal</sub> | P r(F <sub>cal</sub> ) |
|---|-----|---------|---------|------------------|------------------------|
| Entre les sujets  |     |         |         |                  |                        |
| A   | 1   | 1051.25 | 1051.25 | 31.5 **          | .00050                 |
| S(A)  | 8   | 267.00  | 33.38   |                  |                        |
| Dans les sujets   |     |         |         |                  |                        |
| B   | 1   | 151.25  | 151.25  | 52.6 **          | .00009                 |
| AB  | 1   | 101.25  | 101.25  | 35.2 **          | .00035                 |
| BS(A)   | 8   | 23.00   | 2.86    |                  |                        |
| Total   | 19  | 1593.75 |         |                  |                        |
| ns : suspension du jugement au seuil .05<br>* : p inferieur a .05 ; ** : p inferieur a .01. |     |         |         |                  |                        |

Ainsi, les predictions de Conrad se realisent. On note un e et principal attribuable a l'age ( $F_{cal}(1; 8) = 31:5$  ; p inferieur a .01), une interaction entre l'age et la ressemblance acoustique ( $F_{cal}(1; 8) = 35:2$  ; p inferieur a .01). On obtient egalement un e et principal de la ressemblance  $F_{cal}(1; 8) = 52:6$ , p inferieur a .01), mais cet e et est d'interpretation delicate du fait de l'interaction.

Retrouver les resultats precedents a l'aide de methodes de comparaison de moyennes sur des protocoles derives convenablement choisis.

Représenter graphiquement l'interaction entre les deux facteurs.



Le resultat concernant l'age peut etre retrouv a l'aide d'un test de comparaison de moyennes sur deux groupes independants, en utilisant le protocole deriv des sommes ou des moyennes par sujet.

Le resultat relatif a l'interaction peut etre retrouv en construisant le protocole des differences individuelles et en faisant une comparaison des moyennes de ce protocole sur les deux groupes.

#### Enonce 29 Donnees Bahrick

Dans une reprise partielle d'une experimentation de Bahrick (1984), on demande a dix sujets (cinq etudiants et cinq etudiantes) de participer a l'experience suivante :

On montre aux sujets vingt portraits en noir et blanc (composes de dix portraits d'hommes et dix portraits de femmes). On demande aux sujets d'essayer de "memoriser" ces vingt portraits a n de pouvoir les reconnaitre lors d'un test ulterieur. Les sujets accomplissent ensuite pendant environ une demi-heure diverses taches. Puis on leur presente vingt paires de photographies composees d'un portrait vu pendant la phase d'apprentissage et d'un portrait inconnu des sujets ; et on leur demande d'identifier dans chaque paire de photographies le portrait connu.

On donne ci-dessous le nombre de portraits correctement identies en fonction des sujets et du "sexe" des portraits. (Les sujets sont identies par un prenom).

| Nom du sujet | Portrait masculin | Portrait feminin |
|--------------|-------------------|------------------|
| Albert       | 6                 | 6                |
| Henri        | 6                 | 6                |
| Jules        | 5                 | 5                |
| Paul         | 5                 | 5                |
| Octave       | 5                 | 6                |
| Albertine    | 6                 | 8                |
| Henriette    | 7                 | 8                |
| Julie        | 6                 | 6                |
| Paule        | 7                 | 7                |
| Octavie      | 6                 | 6                |

Indications de solution. L'experience est menee selon le plan  $S_5 \times X_2 \times P_2$ . Le tableau d'analyse de variance est donne par :

| Source  | ddl | SC   | CM   | $F_{cal}$ | $Pr(F_{cal})$ |
|---|-----|------|------|-----------|---------------|
| Entre les sujets  |     |      |      |           |               |
| X   | 1   | 7.2  | 7.2  | 10.28 *   | .0124         |
| S(X)  | 8   | 5.6  | 0.7  |           |               |
| Dans les sujets   |     |      |      |           |               |
| P   | 1   | 0.8  | 0.8  | 3.2 NS    | .11           |
| X P   | 1   | 0.2  | 0.2  | 0.8 NS    | .40           |
| PS(X)   | 8   | 2    | 0.25 |           |               |
| Total   | 19  | 15.8 |      |           |               |
| ns : suspension du jugement au seuil .05<br>* : p inferieur a .05 ; ** : p inferieur a .01. |     |      |      |           |               |

### Enonce 30 Dossier \King"

En 1986, King a etudi l'activite motrice chez le rat apres injection d'un medicament appel midazolam. La premiere injection du medicament entra^ne generalement une dimi-nution nette de l'activite motrice. Mais une certaine tolerance se developpe rapidement. King souhaitait savoir si cette tolerance acquise pouvait s'expliquer sur la base d'une tolerance conditionnee.

Il a utilise trois groupes et n'a recueilli les donnees (presentees dans le tableau ci-dessous) que le dernier jour, jour du test. Durant le pre-test, deux groupes d'animaux ont recu a plusieurs reprises des injections de midazolam reparties sur plusieurs jours, tandis que le groupe temoin recevait des injections d'une solution saline physiologique.

Le jour du test, un groupe (le groupe \m^eme") a recu une injection de midazolam dans le m^eme environnement qu'auparavant. Le groupe \di erent" a egalement recu une injection de midazolam, mais dans un environnement di erent. En n, le groupe temoin a recu, pour la premiere fois, une injection de midazolam. Ce groupe temoin devrait donc manifester la reaction initiale classique au medicament (comportement ambulateur reduit), tandis que le groupe \m^eme" devrait presenter l'e et normal de tolerance. Par contre, si King a raison, le groupe \di erent" devrait reagir de la m^eme facon que le groupe temoin ; en e et, ces animaux allaient cette fois recevoir l'injection dans un environnement di erent, et les elements necessaires pour susciter une tolerance conditionnee ne seraient pas presents. La variable dependante du tableau ci-dessous est une mesure du comportement ambulateur, en unites arbitraires.

Comme le medicament se metabolise sur une periode d'environ 1 heure, King a enregistr ses donnees par blocs (ou intervalles) de 5 minutes. Le tableau 1 donne les valeurs observees pour les 6 premiers blocs de donnees.

1) a) Quelles sont les variables independantes (ou facteurs de variation) prises en compte ? Quel est le nombre de niveaux de chacun des facteurs ?

b) Quelle est la variable dependante ?

c) Ecrire le plan d'experience correspondant.

2) a) Calculer les moyennes correspondant aux 18 conditions experimentales de nies par les combinaisons des variables \groupe" et \intervalle".

b) Realiser un graphe illustrant une eventuelle interaction entre ces variables. Commenter le graphe obtenu.

3) Analyse de variance.

Le tableau d'analyse de variance relatif aux donnees observees se presente ainsi :

| Sources de var.  | ddl | SC      | CM     | F   |
|------------------|-----|---------|--------|-----|
| Entre les sujets |     |         |        |     |
| Groupes          | 2   | 285815  | 142907 | ... |
| S(G)             | 21  | 384722  | 18320  |     |
| Dans les sujets  |     |         |        |     |
| Intervalles      | 5   | 399736  | ...    | ... |
| I G              | 10  | 80820   | ...    | ... |
| Residu           | 105 | 281199  | 2678   |     |
| Total            | 143 | 1432293 |        |     |

a) Completer ce tableau en calculant les carres moyens et les statistiques F de Fisher qui sont remplaces par \. . ." dans le tableau ci-dessus.

|                    | Intervalles |     |     |     |     |     |
|--------------------|-------------|-----|-----|-----|-----|-----|
|                    | 1           | 2   | 3   | 4   | 5   | 6   |
| Temoin             | 150         | 44  | 71  | 59  | 132 | 74  |
|                    | 335         | 270 | 156 | 160 | 118 | 230 |
|                    | 149         | 52  | 91  | 115 | 43  | 154 |
|                    | 159         | 31  | 127 | 212 | 71  | 224 |
|                    | 159         | 0   | 35  | 75  | 71  | 34  |
|                    | 292         | 125 | 184 | 246 | 225 | 170 |
|                    | 297         | 187 | 66  | 96  | 209 | 74  |
|                    | 170         | 37  | 42  | 66  | 114 | 81  |
| M <sup>^</sup> eme | 346         | 175 | 177 | 192 | 239 | 140 |
|                    | 426         | 329 | 236 | 76  | 102 | 232 |
|                    | 359         | 238 | 183 | 123 | 183 | 30  |
|                    | 272         | 60  | 82  | 85  | 101 | 98  |
|                    | 200         | 271 | 263 | 216 | 241 | 227 |
|                    | 366         | 291 | 263 | 144 | 220 | 180 |
|                    | 371         | 364 | 270 | 308 | 219 | 267 |
|                    | 497         | 402 | 294 | 216 | 284 | 255 |
| Di erent           | 282         | 186 | 225 | 134 | 189 | 169 |
|                    | 317         | 31  | 85  | 120 | 131 | 205 |
|                    | 362         | 104 | 144 | 114 | 115 | 127 |
|                    | 338         | 132 | 91  | 77  | 108 | 169 |
|                    | 263         | 94  | 141 | 142 | 120 | 195 |
|                    | 138         | 38  | 16  | 95  | 39  | 55  |
|                    | 329         | 62  | 62  | 6   | 93  | 67  |
|                    | 292         | 139 | 104 | 184 | 193 | 122 |

Table 1 { Donnees King

b) En utilisant un seuil de 5%, etudier quelles sont les sources de variation dont l'e et est signi catif.

#### 4) Comparaisons de moyennes

a) Determiner le protocole deriv obtenu en calculant le score moyen observ sur les intervalles 2 a 6 pour chacun des sujets des groupes \temoin" et \di erent".

b) Comparer le comportement des deux groupes a l'aide d'un test sur les moyennes de ce protocole deriv .

#### Enonce 31 Dossier \Termites"

Dans une etude experimentale (d'apres Catherine Venturelli : la dynamique du creusement chez Reticulitermes santonensis, 1990), on etudie le comportement de creusement de 72 groupes de termites (d'ou le facteur groupe G a 72 modalites). Chaque groupe comprend 50 termites. Ces 72 groupes sont repartis dans 12 conditions experimentales di erentes ; d'ou le facteur Condition C a 12 modalites. A chacune des conditions experimentales on a ecte 6 groupes, chacun des 72 groupes etant a ecte a une seule condition experimentale. Pour chaque groupe on observe le nombre de centimetres de galeries creuses par le groupe en 12 heures. L'experience se deroule sur 15 jours, avec 2 releves par jour correspondant



a deux periodes d'activite : un relev le soir, ou l'on observe le nombre de centimetres creuses pendant la journee (periode p1), un relev le matin, ou l'on observe le nombre de centimetres creuses pendant la nuit (periode p2). D'ou le facteur Jour J a 15 modalites et le facteur Periode d'activite P a 2 modalites (p1 et p2).

Les donnees suivantes concernent une partie des observations, les resultats de 12 groupes de termites. Six groupes ont et places dans du sable humide (h1), 6 groupes dans du sable peu humide (h2). D'ou le facteur Humidite a 2 modalites. Il s'agit des deux releves (p1 et p2) du premier jour.

Donnees "Termites" et protocoles derives

|                     |     | p1   | p2    | p2-p1 | (p1+p2)/2 |
|---------------------|-----|------|-------|-------|-----------|
| G(h1)<br>humide     | g1  | 0.0  | 48.80 | 48.80 | 24.40     |
|                     | g2  | 36.5 | 79.3  | 42.80 | 57.90     |
|                     | g3  | 16.2 | 46.3  | 30.10 | 31.25     |
|                     | g4  | 17.8 | 69.3  | 51.50 | 43.55     |
|                     | g5  | 34.4 | 96.8  | 62.40 | 65.60     |
|                     | g6  | 29.4 | 81.2  | 51.80 | 55.30     |
| G(h2)<br>peu humide | g7  | 0.0  | 0.0   | 0.00  | 0.00      |
|                     | g8  | 9.8  | 51.0  | 41.20 | 30.40     |
|                     | g9  | 9.4  | 64.0  | 54.60 | 36.70     |
|                     | g10 | 13.1 | 76.9  | 63.80 | 45.00     |
|                     | g11 | 2.2  | 30.7  | 28.50 | 16.45     |
|                     | g12 | 4.4  | 7.3   | 2.90  | 5.85      |

Moyennes

|    | p1    | p2    |
|----|-------|-------|
| h1 | 22,38 | 70,28 |
| h2 | 6,48  | 38,32 |

E ets intra

| H*P             | p1 (jour) | p2 (nuit) | di p1,p2 |
|-----------------|-----------|-----------|----------|
| h1 (humide)     | 22,38     | 70,28     | -47,90   |
| h2 (peu humide) | 6,48      | 38,32     | -31,84   |
| di h1,h2        | 15,90     | 31,96     | -16,06   |

- 1) En moyenne, jour et nuit confondus, les termites creusent-ils plus en sable humide ou en sable peu humide ?
- 2) Le jour, les termites creusent-ils plus en sable humide ou en sable peu humide ?
- 3) Les termites creusent-ils plus la nuit que le jour ?
- 4) Lorsqu'ils sont places en sable peu humide, les termites sont-ils influences par le jour et la nuit ?
- 5) La difference entre le jour et la nuit est-elle la m<sup>e</sup>me, quelle que soit l'humidite du sable ?

Indications de solutions.

Les donnees fournies correspondent au plan d'experience  $S_6 < H_2 > P_2$ .

1) On raisonne ici sur le protocole deriv des moyennes par individu. Au niveau descriptif :  $\bar{x}_1 = 46:33$ ,  $\bar{x}_2 = 22:4$ ,  $s_{c;1} = 16:1352$ ,  $s_{c;2} = 17:8343$ . Les termites semblent creuser plus en condition  $H_1$  qu'en condition  $H_2$ .

Un test de comparaison de deux moyennes sur des groupes independants donne :  $t_{obs} = 2:43$ , et  $ddl = 10$ , valeur signi cative au seuil de 5% unilateral.

2) On procede ici comme dans la question precedente, mais en utilisant la premiere co-colonne de donnees. On obtient :  $\bar{x}_1 = 22:38$ ,  $\bar{x}_2 = 6:48$ ,  $s_{c;1} = 13:8070$ ,  $s_{c;2} = 5:0598$ .  $t_{obs} = 2:65$ , et  $ddl = 10$ , valeur signi cative au seuil de 5% unilateral.

Un autre parametre descriptif interessant est l'e et calibre du facteur H de ni par :

$$EC = q \frac{\bar{x}_1 - \bar{x}_2}{s^2_{1;c} + s^2_{2;c}} = 1:08. \text{ Un ecart calibre superieur a 1 indique un e et important.}$$

3) On utilise ici le protocole deriv des e ets individuels (colonne des di erences). On a :  $\bar{d} = 39:87$  et  $s_c = 20:99$ . Une comparaison de moyennes sur deux groupes appareilles donne :  $t_{obs} = 6:57$  et  $ddl = 11$ , valeur signi cative au seuil de 5% unilateral.

4) On procede comme dans la question precedente, en se limitant aux individus statistiques du deuxieme groupe. On a :  $\bar{d} = 31:83$  et  $s_c = 26:42$ . Une comparaison de moyennes sur deux groupes appareilles donne :  $t_{obs} = 2:95$  et  $ddl = 5$ , valeur signi cative au seuil de 5% unilateral.

5) Il s'agit ici d'etudier l'interaction entre le facteur P et le facteur H. On pourra tracer un diagramme d'interaction a partir du tableau des moyennes donne dans l'enonc .

On obtient :  $Moy((P_2 - P_1)=H_1) = 47:90$ ,  $s_{c;1} = 10:79$ ,  $Moy((P_2 - P_1)=H_2) = 31:83$ ,  $s_{c;2} = 26:42$ . On peut aussi calculer l'e et calibre (cf. question 2). Ici,  $EC = 0:563$ , ce qui est assez faible.

Une comparaison de moyennes sur les deux groupes correspondant aux modalites  $h_1$  et  $h_2$  du facteur H donne :  $t_{obs} = 1:38$ , valeur non signi cative aux seuils traditionnels. Remarque. Le tableau d'analyse de variance est ici le suivant :

| Source         | ddl | SC    | CM    | $F_{cal}$ | Pr     |
|----------------|-----|-------|-------|-----------|--------|
| H              | 1   | 3437  | 3437  | 5.94      | .035   |
| S(H)           | 10  | 5784  | 578.4 |           |        |
| P              | 1   | 9536  | 9536  | 46.84     | .00005 |
| Interaction HP | 1   | 387.2 | 387.2 | 1.90      | .20    |
| Residu         | 10  | 2036  | 203.6 |           |        |
| Total          | 23  | 21180 |       |           |        |

On retrouve ainsi les resultats des questions 1, 3 et 5, en remarquant que  $F_{cal} = t^2_{obs}$ . En particulier,  $t_{obs} = 2:43$  correspond a  $F_{cal} = 5:94$  (question 1), tandis que  $t_{obs} = 1:38$  correspond a  $F_{cal} = 1:90$  (question 5).

## Autres plans

### Enonce 32 Dossier \Neglige"

Une recherche a porte sur la \pseudo-neglige" qu'on observe chez des sujets normaux. Ce nom provient des similarites qu'elle presente avec l'hemineglige (atteinte de la moitie du champ visuel) de sujets atteints d'une lesion cerebrale. La t^ache des sujets consiste a determiner le milieu subjectif d'une baguette de 24cm avec la seule aide d'in-formations kinesthesiques. La pseudo-neglige se traduit par une deviation systematique

vers la droite (pour les droitiers) de ce milieu subjectif par rapport au milieu objectif de la baguette.

Les données portent sur 24 femmes droitieres (facteur S) réparties selon 2 conditions (12 sujets pour chacune) : active (c1) ou le sujet peut librement déplacer son doigt pose sur un curseur mobile le long de la baguette ; ou passive (c2) ou le sujet commande un moteur déclenchant le mouvement de la baguette dans un sens ou dans l'autre, alors que son doigt ne bouge pas (facteur C). Chaque sujet exécute cette tâche dans 6 situations expérimentales obtenues par le croisement de la main utilisée, gauche (m1) ou droite (m2) et l'orientation du regard, 30N° a gauche (o1), 0N° (o2) ou 30N° a droite (o3) (facteurs M et O). Pour chaque sujet et chaque situation on mesure la déviation en cm entre le milieu subjectif et le milieu objectif de la baguette. Une déviation a droite est notée par une valeur positive, a gauche par une valeur négative.

L'objectif principal de l'expérience est l'étude de l'e et de la condition sur la déviation, et des possibles variations de cet e et selon l'orientation.

Les données sont présentées dans le tableau ci-dessous.

Données "Negligence" et protocoles dérivés (PD1 a PD3)

|       | m1o1  | m1o2  | m1o3  | m2o1  | m2o2  | m2o3  | PD1   | PD2   | PD3   |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| s1c1  | 1.95  | 0.95  | 0.55  | 0.15  | -0.80 | -0.65 | 0.36  | 1.80  | 1.58  |
| s2c1  | 3.00  | 3.10  | 1.55  | -0.10 | -0.30 | -0.60 | 1.11  | 3.10  | 2.88  |
| s3c1  | 1.00  | 1.20  | -0.65 | -1.25 | -2.20 | -2.15 | -0.68 | 2.25  | 2.38  |
| s4c1  | 0.25  | 0.20  | 0.85  | -0.85 | 0.05  | -0.05 | 0.08  | 1.10  | 0.72  |
| s5c1  | 1.15  | 0.55  | 0.95  | -0.10 | -0.40 | -1.95 | 0.03  | 1.25  | 1.70  |
| s6c1  | 1.85  | 0.75  | -1.65 | -0.35 | 0.15  | -0.35 | 0.07  | 2.20  | 0.50  |
| s7c1  | 2.05  | 1.75  | -1.50 | 1.05  | 0.05  | -0.90 | 0.42  | 1.00  | 0.70  |
| s8c1  | 1.75  | -0.50 | 0.25  | -0.25 | -0.05 | 0.90  | 0.35  | 2.00  | 0.30  |
| s9c1  | 0.40  | 1.85  | -0.10 | -0.45 | -0.40 | -1.05 | 0.04  | 0.85  | 1.35  |
| s10c1 | -0.80 | 4.10  | 3.00  | -1.15 | -0.20 | -2.25 | 0.45  | 0.35  | 3.30  |
| s11c1 | 2.50  | -0.75 | 0.30  | 0.20  | -1.25 | -1.20 | -0.03 | 2.30  | 1.43  |
| s12c1 | 1.80  | 1.65  | 0.55  | 1.00  | -1.00 | 1.30  | 0.88  | 0.80  | 0.90  |
| s13c2 | -0.30 | -0.10 | -0.55 | 1.30  | -1.90 | 0.75  | -0.13 | -1.60 | -0.37 |
| s14c2 | 1.40  | -1.00 | 0.95  | -0.20 | 0.00  | -0.50 | 0.11  | 1.60  | 0.68  |
| s15c2 | 0.25  | 0.75  | -0.70 | 0.75  | 0.55  | 0.60  | 0.37  | -0.50 | -0.53 |
| s16c2 | 0.75  | 1.10  | 1.40  | 0.25  | -1.55 | 0.05  | 0.33  | 0.50  | 1.50  |
| s17c2 | -0.30 | -0.70 | -0.80 | 0.85  | -0.10 | -1.15 | -0.37 | -1.15 | -0.47 |
| s18c2 | -2.10 | 3.45  | -1.85 | 0.95  | 1.50  | 2.80  | 0.79  | -3.05 | -1.92 |
| s19c2 | 1.85  | -0.55 | 2.25  | -0.05 | 1.55  | -0.30 | 0.79  | 1.90  | 0.78  |
| s20c2 | 1.65  | -0.75 | 0.05  | -0.60 | -1.55 | -1.75 | -0.49 | 2.25  | 1.62  |
| s21c2 | -0.75 | 1.25  | -0.25 | -0.20 | -0.05 | 2.35  | 0.39  | -0.55 | -0.62 |
| s22c2 | 1.80  | 1.15  | 1.95  | 0.80  | 0.85  | 0.90  | 1.24  | 1.00  | 0.78  |
| s23c2 | -0.95 | 0.15  | 0.80  | -0.15 | -0.30 | -1.05 | -0.25 | -0.80 | 0.50  |
| s24c2 | 0.20  | -0.45 | -0.80 | 0.65  | 2.20  | -0.10 | 0.28  | -0.45 | -1.27 |

Ref : Chokron, Imbert (1993) - Egocentric reference and asymmetric perception of space. *Neuropsychologia* 31, 3, 267-275. D'après J.M. Bernard (1994) - Structure des données, données planifiées Mathématiques, Informatique et Sciences humaines n° 126 7-18.

1) a) Pour chacun des 4 cas ci-dessous, écrire au moyen des symboles <> et \* la relation entre les deux facteurs S et C ; M et O ; C et M ; C et O.

- b) Ecrire au moyen des m<sup>es</sup> symboles la relation entre les trois facteurs S, M et O.  
 2) a) On calcule la moyenne de toutes les valeurs observees. On trouve 0.256. Que signi e cette valeur. Que peut-on en conclure ?  
 b) On s'interesse aux e ets moyens et intra du facteur C.

|          | m1     | m2     | Moyennes |
|----------|--------|--------|----------|
| c1       | +0.996 | -0.483 | 0.256    |
| c2       | +0.285 | +0.226 | 0.256    |
| Moyennes | 0.640  | -0.128 | 0.256    |

Commenter l'e et moyen du facteur C.

Commenter l'e et observ du facteur C lorsque c'est la main gauche (m1) qui est utilisee. Calculer la valeur de cet e et

Construire un graphe representant l'interaction entre les facteurs C et M.

3) Les trois dernieres colonnes du tableau general (PD1 a PD3) indiquent trois protocoles pouvant ^etre derives du protocole de base. Indiquer, pour chacun des e ets ci-dessous, quel est le protocole deriv pertinent et indiquer quel calcul a permis d'obtenir la premiere valeur de ce protocole deriv :

E et de C ; e et de M ; e et de M.C ; e et de M/o1 .

4) On s'interesse a l'e et de la condition (C) avec la main gauche (m1) et avec l'orientation o2. Le protocole deriv pertinent est la deuxieme colonne du tableau principal (m1o2). A l'aide d'un test de comparaison de moyennes, determiner si cet e et est signi catif.

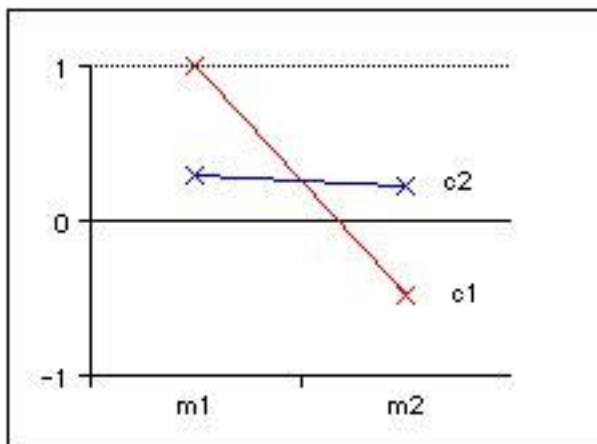
Reponses : 1) a)

|        |      |
|--------|------|
| S et C | S<C> |
| M et O | M*O  |
| C et M | C*M  |
| C et O | C*O  |

1) b) Entre S, M et O : S\*M\*O.

2) a) La moyenne generale represente l'e et de la pseudo-negligeance, independamment des e ets des autres facteurs.

b) L'e et moyen du facteur C est nul. En revanche, l'e et observ du facteur C dans la modalite m1 est 0:996 0:285 = 0:711. Le graphe d'interaction pourra ^etre represent par :



c)

|              | PD  | Calcul                                   |
|--------------|-----|--|
| E et de C    | PD1 | (1:95 + 0:95 + 0:55 + 0:15 0:80 0:65)=6  |
| E et de M    | PD3 | (1:95 + 0:95 + 0:55)=3 (0:15 0:8 0:65)=3 |
| E et de M.C  | PD3 | id.                                      |
| E et de M/o1 | PD2 | 1:95 0:15                                |

3) Il s'agit d'une comparaison de moyennes sur groupes independants. On obtient les resultats suivants :

|    | Moyenne | ecart type | ecart type cor. |
|----|---------|------------|-----------------|
| c1 | 1.238   | 1.333      | 1.392           |
| c2 | 0.358   | 1.211      | 1.265           |

$t_{obs} = 1:6191$  et  $ddl = 22$ . Au seuil de 5%, la difference n'est pas significative.

Il s'agit ici d'un plan  $S < C_2 > M_2 O_3$ . Les elements vus en cours ne permettent pas de prevoir la structure du tableau d'analyse de variance, qui est ici assez complexe :

| Source   | ddl | SC      | CM     | $F_{cal}$ | Pr    |
|----------|-----|---------|--------|-----------|-------|
| C2       | 1   | 0.000   | 0.000  | 0.00      | 0.997 |
| S(C2)    | 22  | 31.658  | 1.439  |           |       |
| M2       | 1   | 21.275  | 21.275 | 13.50     | 0.001 |
| C2 M2    | 1   | 18.169  | 18.169 | 11.53     | 0.003 |
| M2 S(C2) | 22  | 34.677  | 1.576  |           |       |
| O3       | 2   | 5.342   | 2.671  | 2.25      | 0.118 |
| C2 O3    | 2   | 3.236   | 1.618  | 1.36      | 0.267 |
| O3 S(C2) | 44  | 52.348  | 1.190  |           |       |
| M2 O3    | 2   | 1.357   | 0.678  | 0.57      | 0.571 |
| C2 M2 O3 | 2   | 0.495   | 0.248  | 0.21      | 0.814 |
| Residu   | 44  | 52.586  | 1.195  |           |       |
| Total    | 143 | 221.142 |        |           |       |

Ce tableau peut ^etre obtenu a l'aide de Minitab. On saisit les donnees, a raison d'une colonne pour chaque facteur et d'une colonne pour la variable dependante. Ces colonnes sont nommees \Sujet", \Condition", \Main", \Orientation". On utilise le menu Stats - Anova - Modele lineaire generalis . On complete ensuite le dialogue en indiquant dans la zone d'edition \Modele" :

Condition Sujet(Condition) Main Main\*Condition Main\*Sujet(Condition)  
 Orientation Orientation\*Condition Orientation\*Sujet(Condition)  
 Main\*Orientation Main\*Orientation\*Condition et en indiquant Sujet comme facteur aleatoire.

Ce tableau ne fait d'ailleurs que confirmer ce que nous avons etabli par ailleurs : seuls le facteur \Main" et l'interaction \Main Condition" ont des effets significatifs.

### Enonce 33 Dossier \Tapping"

Une t^ache de \Tapping" consiste a demander a des sujets droitiers d'appuyer avec l'index sur un bouton le plus rapidement possible durant une periode limitee (20s). On note le nombre d'appuis effectues durant cette periode. Huit garcons (sexe l) et huit filles (sexe 2) ont passe une experience dans laquelle on comparait les performances selon que la t^ache etait effectuee avec la main gauche (m1) ou avec la main droite (m2), et selon

que les sujets avaient à résoudre une tâche concurrente (c1) ou non (c2). Chaque sujet est confronté à chacune des 4 conditions de tâches par le croisement des facteurs M et C. Les données présentées dans le tableau 2 correspondent à la moyenne des performances obtenues sur 5 essais.

Critères d'importance de l'e et :

Critère sémantique : On considérera qu'un e et est faible s'il est inférieur à 1, important s'il est supérieur à 2.

Critère psychométrique : On prendra les critères habituels ( $d = s < 1=3$  et  $d = s > 2=3$ ) comme limites d'un e et faible et d'un e et important.

N.B. : l'e et d'un facteur est la différence des scores observés pour les deux niveaux du facteur. Dans les notations précédentes, s désigne l'écart type de la série des e ets

individuels. De manière classique, la quantité  $\overline{s}(\overline{d})$  est appelée e et calibre.

|       | m1c1  | m2c1  | m1c2  | m2c2  |
|-------|-------|-------|-------|-------|
| s1x1  | 70.25 | 79.50 | 61.75 | 68.00 |
| s2x1  | 57.50 | 65.00 | 62.25 | 71.00 |
| s3x1  | 71.00 | 91.00 | 60.00 | 68.00 |
| s4x1  | 63.75 | 76.25 | 58.25 | 70.50 |
| s5x1  | 72.50 | 60.25 | 61.25 | 60.50 |
| s6x1  | 61.25 | 80.00 | 64.25 | 87.00 |
| s7x1  | 58.00 | 59.75 | 62.75 | 69.00 |
| s8x1  | 57.50 | 64.25 | 74.25 | 85.00 |
| s9x2  | 62.00 | 66.50 | 66.25 | 72.50 |
| s10x2 | 74.00 | 78.25 | 74.25 | 87.25 |
| s11x2 | 63.25 | 77.50 | 61.50 | 86.50 |
| s12x2 | 88.25 | 91.50 | 93.50 | 83.25 |
| s13x2 | 60.75 | 67.25 | 60.75 | 69.25 |
| s14x2 | 70.75 | 84.00 | 73.25 | 79.50 |
| s15x2 | 54.50 | 59.00 | 64.50 | 68.25 |
| s16x2 | 48.50 | 51.00 | 62.50 | 70.25 |

Table 2 { Données \Tapping"

1) Pour chacun des 5 e ets X, M, C, X:M, M:C indiquer :

{ quel est, parmi les 4 protocoles dérivés du tableau 3, le protocole dérivé pertinent pour l'étude de cet e et

{ comment a été calculée la première valeur (sujet s1) de ce protocole dérivé pertinent. 2)

On s'intéresse à l'e et du facteur Sexe (X), c'est-à-dire aux différences de performances obtenues par les garçons (x1) et les filles (x2). On trouve une différence de performance  $d_{obs} = 3:086$  en faveur des filles. On veut tester l'hypothèse  $H_0$  d'une absence d'e et parent pour cette comparaison.

{ Indiquer la formule du test T de Student à utiliser dans ce cas.

{ On trouve  $t_{obs} = 0:788$ . Indiquer le nombre de degrés de liberté et le résultat du test. Donner une conclusion inférentielle sur l'e et de X.

3) On s'intéresse à l'e et du facteur C, c'est-à-dire à l'e et de la tâche concurrente sur les performances au tapping. En moyenne, on trouve  $m_{c1} = 68:27$  et  $m_{c2} = 70:53$  d'où

|       | PD1    | PD2    | PD3     | PD4     |
|-------|--------|--------|---------|---------|
| s1x1  | -3.00  | 7.750  | -10.000 | 69.8750 |
| s2x1  | 1.25   | 8.125  | 5.375   | 63.9375 |
| s3x1  | -12.00 | 14.000 | -17.000 | 72.5000 |
| s4x1  | -0.25  | 12.375 | -5.625  | 67.1875 |
| s5x1  | 11.50  | -6.500 | -5.500  | 63.6250 |
| s6x1  | 4.00   | 20.750 | 5.000   | 73.1250 |
| s7x1  | 4.50   | 4.000  | 7.000   | 62.3750 |
| s8x1  | 4.00   | 8.750  | 18.750  | 70.2500 |
| s9x2  | 1.75   | 5.375  | 5.125   | 66.8125 |
| s10x2 | 8.75   | 8.625  | 4.625   | 78.4375 |
| s11x2 | 10.75  | 19.625 | 3.625   | 72.1875 |
| s12x2 | -13.50 | -3.500 | -1.500  | 89.1250 |
| s13x2 | 2.00   | 7.500  | 1.000   | 64.5000 |
| s14x2 | -7.00  | 9.750  | -1.000  | 76.8750 |
| s15x2 | -0.75  | 4.125  | 9.625   | 61.5625 |
| s16x2 | 5.25   | 5.125  | 16.625  | 58.0625 |

Table 3 { Protocoles derives des donnees \Tapping"

 $u_{obs}$ 

= 2:26. On souhaite savoir si l'e et de la t^ache concurrente est di erent pour les garcons (C/x1) et les lles (C/x2). Pour cela, on derive le protocole des moyennes de support X\*C represent ci-dessous.

|    | c1     | c2     |
|----|--------|--------|
| x1 | 67.984 | 67.734 |
| x2 | 68.563 | 73.328 |

{ Pour chacun des deux e ets C/x1 et C/x2, calculer les valeurs des e ets, puis conclure sur l'importance des e ets en utilisant le critere semantique.

{ Construire un graphique representant l'interaction entre C et X.

4) La premiere colonne du tableau 3 (protocole PD1) est un sous-protocole de structure S8<X2>.

{ Calculer sur ce protocole les moyennes  $m_{x1}$  et  $m_{x2}$ , les variances corrigees ( $s^2_{1;c}$  et  $s^2_{2;c}$ ) et les ecarts-types corriges de chacun des deux groupes x1 et x2.

{ Calculer la variance corrigee intra et l'ecart-type corrige intra.

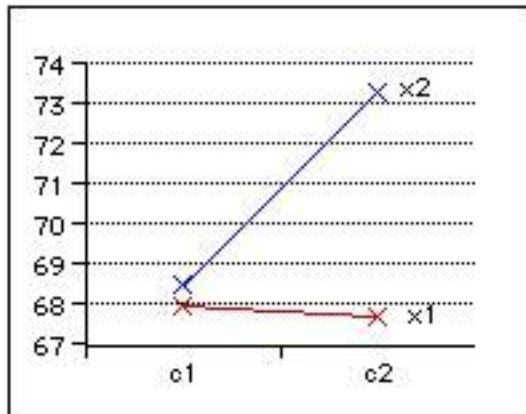
5) Pour les m^emes donnees que la question precedente, de structure S8<X2>, on peut calculer la variance inter (variance des moyennes des deux groupes x1 et x2) et la variance intra (moyenne des variances des deux groupes).

{ Indiquer comment augmenter la variance intra sans modi er la variance inter. { Indiquer comment augmenter la variance inter sans modi er la variance intra. Indications de reponses : 1)

| E et | Protocole | Calcul                                   |
|------|-----------|--|
| X    | PD4       | $(70:25 + 79:50 + 61:75 + 68)=4$         |
| M    | PD2       | $[(79:50 \ 70:25) + (68:00 \ 61:75)] =2$ |
| C    | PD3       | $[(61:75 + 68:00) \ (70:25 + 79:50)] =2$ |
| X.M  | PD2       |  |
| M.C  | PD1       | $(68:00 \ 61:75) \ (79:50 \ 70:25)$      |

2) Comparaison de moyennes sur groupes independants. ddl = 14. Pas d'e et signi catif au seuil de 5%.

3) Le graphe d'interaction est donne par :



4) On obtient :  $m_{x1} = 1:25$ ,  $m_{x2} = 0:9025$ ,  $s^2_{1;c} = 46:66$ ,  $s^2_{2;c} = 64:570$ ,  $s_{1;c} = 6:83$ ,  $s_{2;c} = 8:04$ .

La variance corrigée intra est la demi-somme des variances précédentes (car il s'agit de groupes équilibrés). D'où  $s^2_{intra} = 55:615$  et  $s_{intra} = 7:45$ .

5) Pour augmenter la variance intra sans modifier la variance inter, on modifie les scores de façon à augmenter la dispersion dans l'un des groupes sans modifier les moyennes des groupes. Par exemple, on diminue de 10 points le score du sujet 3 et on augmente de 10 points celui du sujet 5.

Pour augmenter la variance inter sans modifier la variance intra, on peut, par exemple, augmenter d'une même quantité tous les scores du groupe présentant la moyenne la plus élevée.

Enonce 34 Dossier \TR"

TR= Temps de reaction

Les données présentées ci-dessous sont extraites d'une expérience de temps de réaction (Holender et Bertelson, 1975). La tâche du sujet (adulte) est de réagir le plus rapidement possible à la présentation d'un stimulus. Deux facteurs expérimentaux sont en jeu : le facteur F, fréquence du stimulus, à 2 modalités : f1 : stimulus fréquent (de fréquence 0.75) f2 : stimulus rare (de fréquence 0.25) et le facteur D, durée de la période préparatoire (délai entre la présentation d'un signal avertisseur et celle du stimulus), à 2 modalités également : d1 : période courte (0.5 secondes) d2 : période longue (5 secondes). On introduit ici un facteur supplémentaire (non présent dans l'expérience originale) en supposant que les sujets sont répartis en 2 groupes de 4 sujets chacun, d'où le facteur Groupe,  $G=fg1, g2g$ . Le facteur sujet, S, a 8 modalités.

Chaque sujet effectue plusieurs essais dans chacune des 4 conditions expérimentales correspondant aux modalités du croisement  $F2 * D2$ . Le protocole de base présenté ci-dessous donne pour chaque sujet, et chacune des conditions expérimentales, la moyenne des temps



de reaction pour les di erents essais (en millisecondes). D'ou un protocole numerique de 32 observations.

|       |    | Protocole de base |      |      |      | Protocoles derives |      |        |
|-------|----|-------------------|------|------|------|--------------------|------|--------|
|       |    | f1d1              | f2d1 | f1d2 | f2d2 | F/d1               | F/d2 | Moy.   |
| S(g1) | s1 | 387               | 435  | 416  | 473  | 48                 | 57   | 427.75 |
|       | s2 | 321               | 336  | 343  | 368  | 15                 | 25   | 342    |
|       | s3 | 333               | 362  | 358  | 390  | 29                 | 32   | 360.75 |
|       | s4 | 344               | 430  | 352  | 393  | 86                 | 41   | 379.75 |
| S(g2) | s5 | 368               | 432  | 432  | 504  | 64                 | 72   | 434    |
|       | s6 | 357               | 367  | 394  | 411  | 10                 | 17   | 382.25 |
|       | s7 | 336               | 346  | 340  | 421  | 10                 | 81   | 360.75 |
|       | s8 | 387               | 454  | 438  | 496  | 67                 | 58   | 443.75 |

Tableau des moyennes (m) et des ecart types corriges (s) :

|      |   | Protocole de base |         |         |         | Protocoles derives |        |        |
|------|---|-------------------|---------|---------|---------|--------------------|--------|--------|
|      |   | f1d1              | f2d1    | f1d2    | f2d2    | F/d1               | F/d2   | Moy.   |
| g1   | m | 346.250           | 390.750 | 367.250 | 406.000 | 44.500             | 38.750 | 377.56 |
|      | s | 28.745            | 49.406  | 33.079  | 46.036  | 30.795             | 13.817 | 36.84  |
| g2   | m | 362.000           | 399.750 | 401.000 | 458.000 | 37.750             | 57.000 | 405.19 |
|      | s | 21.307            | 51.461  | 45.092  | 48.778  | 32.066             | 28.296 | 40.08  |
| Ens. | m | 354.125           | 395.25  | 384.125 | 432     |                    |        |        |

Facteur F - Facteur D

|                  | f2    | f1    | f2 - f1 |
|------------------|-------|-------|---------|
| s1               | 454   | 401.5 | 52.5    |
| s2               | 353   | 332   | 20      |
| s3               | 376   | 345.5 | 30.5    |
| s4               | 411.5 | 348   | 63.5    |
| Moy.             |       |       | 41.625  |
| s <sub>1;c</sub> |       |       | 19.90   |
| s5               | 468   | 400   | 68      |
| s6               | 389   | 375.5 | 13.5    |
| s7               | 383.5 | 338   | 45.5    |
| s8               | 475   | 412.5 | 62.5    |
| Moy.             |       |       | 47.375  |
| s <sub>2;c</sub> |       |       | 24.53   |
| Moy.             |       |       | 44.5    |
| s <sub>c</sub>   |       |       | 20.91   |

|                  | d2    | d1    | d2 - d1 |
|------------------|-------|-------|---------|
| s1               | 444.5 | 411   | 33.5    |
| s2               | 355.5 | 328.5 | 27      |
| s3               | 374   | 347.5 | 26.5    |
| s4               | 372.5 | 387   | -14.5   |
| Moy.             |       |       | 18.125  |
| s <sub>1;c</sub> |       |       | 21.98   |
| s5               | 468   | 400   | 68      |
| s6               | 402.5 | 362   | 40.5    |
| s7               | 380.5 | 341   | 39.5    |
| s8               | 467   | 420.5 | 46.5    |
| Moy.             |       |       | 48.625  |
| s <sub>2;c</sub> |       |       | 13.28   |
| Moy.             |       |       | 33.75   |
| s <sub>c</sub>   |       |       | 23.42   |

- 1) Quel est le plan d'experience utilise ?
- 2) Etudier l'e et du facteur G (groupe).
- 3) Etudier l'e et principal du facteur F.
- 4) Etudier l'e et principal du facteur D.
- 5) Construire un graphe illustrant l'interaction entre les les facteurs D et F. Le commenter.

6) Construire un graphe illustrant l'interaction entre les facteurs D et G. Quel est le protocole deriv pertinent pour etudier cette interaction ? Cette interaction est-elle signi-cative ? Conclure a l'aide d'un test de comparaison de moyennes.

Indications de correction.

1) Le plan d'experience est  $S_4 < G_2 > D_2 F_2$ .

2) Le protocole deriv pertinent est ici celui des moyennes par sujet (premier tableau). En utilisant les donnees calculees dans le second tableau, on obtient  $t_{obs} = 1:015$ , resultat non signi catif aux seuils traditionnels.

3) L'e et principal du facteur F peut ^etre etudi a l'aide d'une comparaison de moyennes sur deux groupes appareilles, en utilisant les donnees du tableau 3. On obtient :  $t_{obs} = 4:03$  et  $ddl = 7$ , resultat signi catif au seuil de 5% unilaterial.

4) L'e et principal du facteur D peut ^etre etudi a l'aide d'une comparaison de moyennes sur deux groupes appareilles, en utilisant les donnees du tableau 4. On obtient :  $t_{obs} = 6:01$  et  $ddl = 7$ , resultat signi catif au seuil de 5% unilaterial.

5) On utilise les moyennes gurant dans la derniere ligne du tableau 2. Le graphe ne montre aucune interaction entre D et G.

6) On utilise ici le tableau 4. L'interaction peut ^etre etudiee a l'aide d'une comparaison de deux moyennes sur des groupes independants. On obtient :  $t_{obs} = 1:67$  et  $ddl = 6$ , resultat non signi catif au seuil de 5% unilaterial.

Enonce 35 Donnees \Figures"

Les donnees qui suivent sont empruntees a une experience de segregation perceptive sur des gures de GOTTSCHALDT. La t^ache du sujet consiste a retrouver et a retracer les contours d'une gure simple dans des gures complexes. Le materiel comportait plu-sieurs planches de di culte croissante. Pour l'analyse qui suit, on retiendra seulement les resultats relatifs aux planches p1 et p2.

Les sujets de l'experience ont ete affectes au hasard a l'une des deux conditions d'appren-tissage a1 et a2 decrites plus loin. A l'interieur de chaque condition, les sujets passent toutes les planches, dans tous les ordres possibles, donc s'agissant des planches p1 et p2 : soit dans l'ordre 1-2, soit dans l'ordre 2-1.

Les conditions a1 et a2 sont de nies de la maniere suivante :

- condition a1 : les sujets e ectuent la t^ache apres avoir proced a un apprentissage prolonge de gures simples ;
- condition a2 : les sujets e ectuent la t^ache sans apprentissage prealable.

La variable dependante de base est le temps (en secondes) mis pour repasser une gure simple (temps moyen par sujet et par planche).

Pour l'analyse qui suit, on a retenu le sous-protocole suivant relatif a 24 sujets (6 sujets pour chacun des 4 groupes) et aux deux planches p1 et p2.

| condition a1     |    |            |            | condition a2     |    |            |            |
|------------------|----|------------|------------|------------------|----|------------|------------|
|                  |    | planche p1 | planche p2 |                  |    | planche p1 | planche p2 |
| g11<br>ordre 1-2 | 1  | 15         | 38         | g12<br>ordre 1-2 | 13 | 6          | 27         |
|                  | 2  | 9          | 40         |                  | 14 | 7          | 15         |
|                  | 3  | 17         | 20         |                  | 15 | 25         | 24         |
|                  | 4  | 8          | 13         |                  | 16 | 26         | 30         |
|                  | 5  | 16         | 41         |                  | 17 | 10         | 31         |
|                  | 6  | 9          | 35         |                  | 18 | 6          | 10         |
| g21<br>ordre 2-1 | 7  | 7          | 27         | g22<br>ordre 2-1 | 19 | 6          | 22         |
|                  | 8  | 7          | 36         |                  | 20 | 10         | 27         |
|                  | 9  | 12         | 35         |                  | 21 | 15         | 18         |
|                  | 10 | 4          | 24         |                  | 22 | 10         | 22         |
|                  | 11 | 7          | 36         |                  | 23 | 9          | 25         |
|                  | 12 | 7          | 30         |                  | 24 | 9          | 5          |

1) Chacun des sujets de l'expérience a et a ect au hasard a l'un des quatre groupes g11, g21, g12, g22. Par ailleurs chaque sujet passe les deux planches p1 et p2. On appellera :

P2 = fp1, p2g le facteur "Planches" ;

G4 = fg11, g21, g12, g22g le facteur "Groupe"

; S le facteur "sujets".

Donner une formule verbale et une justification de chacune des écritures suivantes :  $S < G4 >$  ;  $S * P2$  ;  $S < G4 > * P2$

2) On considère les facteurs :

Apprentissage A2 = fa1, a2g avec a1 : apprentissage préalable, a2 : pas d'apprentissage préalable.

Ordre O2 = fo1, o2g avec o1 : planche p1 puis planche p2, o2 : planche p2 puis planche p1

Donner une formulation verbale et une justification de l'écriture suivante :  $O2 * A2$ .

Donner en n une formulation verbale de l'écriture suivante :  $S < O2 * A2 > * P2$

3) A l'intérieur de chacun des 4 groupes, on a calculé la moyenne et la variance corrigée relatives à chaque planche. Dans le tableau suivant, on trouvera les résultats relatifs à la planche 1 (les résultats relatifs à la planche 2 n'étant pas nécessaires pour les calculs demandés par la suite) ; dans chaque case, la première valeur est la moyenne, la deuxième la variance corrigée.

| condition a1 |               |            | condition a2 |               |            |
|--------------|---------------|------------|--------------|---------------|------------|
|              | planche p1    | planche p2 |              | planche p1    | planche p2 |
| g11          | 12,3<br>16,67 |            | g12          | 13,3<br>91,07 |            |
| g21          | 7,3<br>6,67   |            | g21          | 9,8<br>8,57   |            |

L'objectif principal de l'expérience est d'examiner s'il y a une différence entre les conditions a1 et a2. On se bornera ici à cet examen pour la planche p1 seulement, à l'intérieur de chacun des ordres de passation. Donc, on examinera les 2 comparaisons g11, g12 et g21, g22.

a) Que suggère l'examen à vue des données, indépendamment de toute procédure d'inférence statistique ?

b) Examiner d'un point de vue inferentiel chacune des 2 comparaisons indiquees ; on pourra notamment proceder a un test de comparaison de 2 moyennes. Commenter brievement les resultats obtenus.

Reponses : 1)  $S < G4 >$  : chaque sujet est affecte a un seul groupe ; le facteur "sujet" est embo^te dans le facteur "groupe".

$S * P2$  : chaque sujet passe par les deux niveaux du facteur "planche". Le facteur "sujet" est croise avec le facteur "planche".

2)  $O2 * A2$  : chaque niveau du facteur "apprentissage" est combine avec chaque niveau du facteur "ordre". Les facteurs  $O2$  et  $A2$  sont croises.

$S < O2 * A2 > * P2$  : Chaque sujet est affecte a un ordre donne et un type d'apprentissage donne. Chaque sujet passe par les deux modalites du facteur "planche".

3) b) On compare  $g_{11}$  et  $g_{12}$  a l'aide d'une comparaison de moyennes sur des groupes independants.  $t_{obs} = 0.24$ , qui n'est pas significatif d'une difference entre les deux groupes.

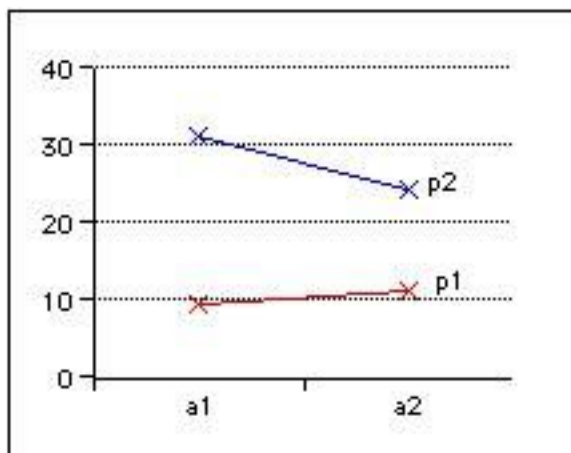
De m^eme, la comparaison des groupes  $g_{21}$  et  $g_{22}$  aboutit a  $t_{obs} = 1.56$ , dont le niveau de significativite n'est que de 7.4% .

Remarques. On obtiendrait des resultats plus significatifs en etudiant les scores obtenus sur la planche p2.

Le tableau complet d'analyse de variance donne ici :

| Source  | ddl | SC      | CM      | $F_{cal}$ | Pr   |
|---------|-----|---------|---------|-----------|------|
| O2      | 1   | 96.33   | 96.33   | 2.75      | 11%  |
| A2      | 1   | 200.08  | 200.08  |           |      |
| O2A2    | 1   | 2.08    | 2.08    |           |      |
| S(O2A2) | 20  | 1457.51 | 72.88   |           |      |
| P2      | 1   | 2914.08 | 2914.08 | 89.22     | **   |
| O2P2    | 1   | 24.08   | 24.08   | 12.5      | 0.2% |
| A2P2    | 1   | 408.33  | 408.33  |           |      |
| O2A2P2  | 1   | 16.33   | 16.33   |           |      |
| Residu  | 20  | 653.17  | 32.66   |           |      |
| Total   | 47  | 5772    |         |           |      |

L'effet du facteur "planche" est tres significatif. L'interaction entre les facteurs "apprentissage" et "planche" est egalement tres significatif et peut ^etre illustre par un graphe d'interaction.



Enonce 36 Dossier "Syssau"

Dans un article publie en 1996, A. Syssau et D. Brouillet etudient le r^ole de la nature et

de la valeur affective d'un texte dans la récupération du souvenir chez les personnes âgées. L'expérience a concerné 80 sujets répartis en 2 groupes : 40 sujets présentant un déficit mnésique et 40 sujets non déficitaires. Quatre textes ont été utilisés, correspondant au croisement des deux modalités du facteur "type de texte" (narratif v/s descriptif) et des deux modalités du facteur "connotation" (affective v/s neutre). On a constitué 8 groupes de 10 sujets : chaque texte est proposé à un groupe de sujets déficitaires et un groupe de sujets non-déficitaires.

La performance des sujets est mesurée par le nombre de propositions correctement rappelés (score de 0 à 20).

Les auteurs font les hypothèses suivantes :

Quel que soit le groupe, les textes narratifs seront mieux restitués que les textes descriptifs, avec des performances, dans le groupe non déficitaire, supérieures à celles du groupe déficitaire. De plus, la charge affective facilitera la restitution des textes, dans le groupe déficitaire. Ainsi, l'association "charge affective et structure narrative" aurait pour conséquence une atténuation des différences de performances entre le groupe déficitaire et non déficitaire.

Les résultats observés dans une reprise de cette expérience sont rassemblés dans le tableau 4.

1) a) Quelles sont les variables indépendantes (ou facteurs de variation) prises en compte ? Quel est le nombre de niveaux de chacun des facteurs ?

b) Quelle est la variable dépendante ?

c) Écrire le plan d'expérience correspondant.

2) a) Compléter le tableau 6 en calculant les carrés moyens et les statistiques F de Fisher qui sont remplacés par ". . ." dans l'énoncé .

b) Utiliser ce tableau pour déterminer si les effets principaux des facteurs de déficit mnésique, type de texte et connotation affective sont significatifs.

3) Étude des interactions

a) Calculer les scores moyens des 4 groupes obtenus en croisant les deux facteurs de déficit et type de texte, indépendamment de la connotation affective du texte.

b) Réaliser un graphique montrant l'absence d'interaction entre ces deux facteurs.

c) Réaliser une étude analogue montrant l'interaction entre le déficit mnésique et la connotation affective du texte.

d) Utiliser le tableau d'analyse de variance (tableau 6) pour conclure au niveau inférentiel les résultats précédents.

4) Les auteurs affirment :

Quel que soit le texte, lorsque la connotation affective est présente, les sujets déficitaires ont des performances améliorées.

a) Justifier cette affirmation, pour les textes descriptifs, à l'aide d'un test unilatéral de comparaison de moyennes.

Les auteurs poursuivent :

Ils ont alors des performances comparables à celles des sujets non déficitaires.

b) Justifier cette affirmation en s'appuyant sur l'un des tableaux d'analyse de variance fournis.

Les auteurs affirment également :

Quel que soit le texte, si la connotation est neutre, les sujets non de citaires obtiennent des résultats supérieurs aux sujets de citaires.

c) L'un des tableaux d'analyse de variance permet de justifier en partie cette conclusion. Lequel ? Pourquoi la justification n'est-elle que partielle ?

5) Au vu des résultats obtenus, les hypothèses des auteurs sont-elles vérifiées ?

|                 | texte narratif     |                    | texte descriptif   |                    |
|-----------------|--------------------|--------------------|--------------------|--------------------|
|                 | connotation active | connotation neutre | connotation active | connotation neutre |
| de citaires     | 0                  | 2                  | 1                  | 1                  |
|                 | 2                  | 0                  | 3                  | 0                  |
|                 | 1                  | 5                  | 0                  | 0                  |
|                 | 8                  | 9                  | 4                  | 1                  |
|                 | 2                  | 1                  | 1                  | 0                  |
|                 | 4                  | 0                  | 2                  | 1                  |
|                 | 1                  | 0                  | 2                  | 1                  |
|                 | 0                  | 7                  | 1                  | 1                  |
|                 | 6                  | 3                  | 0                  | 0                  |
|                 | 9                  | 0                  | 0                  | 0                  |
| non de citaires | 0                  | 0                  | 0                  | 6                  |
|                 | 4                  | 2                  | 0                  | 4                  |
|                 | 2                  | 3                  | 3                  | 2                  |
|                 | 5                  | 6                  | 4                  | 0                  |
|                 | 1                  | 3                  | 2                  | 0                  |
|                 | 2                  | 8                  | 0                  | 1                  |
|                 | 6                  | 6                  | 0                  | 1                  |
|                 | 2                  | 4                  | 2                  | 3                  |
|                 | 0                  | 3                  | 1                  | 0                  |
|                 | 2                  | 5                  | 0                  | 5                  |

Table 4 { Données observées

|                 |           | texte narratif          |                       | texte descriptif        |                       |
|-----------------|-----------|-------------------------|-----------------------|-------------------------|-----------------------|
|                 |           | connotation<br>a ective | connotation<br>neutre | connotation<br>a ective | connotation<br>neutre |
| de citaires     | $\bar{X}$ | 3.3                     | 2.7                   | 1.4                     | 0.5                   |
|                 | s         | 3.13                    | 3.10                  | 1.28                    | 0.50                  |
|                 | $S_c$     | 3.30                    | 3.27                  | 1.35                    | 0.53                  |
| non de citaires | $\bar{X}$ | 2.4                     | 4.0                   | 1.2                     | 2.2                   |
|                 | s         | 1.91                    | 2.19                  | 1.40                    | 2.09                  |
|                 | $S_c$     | 2.01                    | 2.31                  | 1.48                    | 2.20                  |

Table 5 { moyennes et ecart types par groupe

| Sources de var. | ddl | SC      | CM    | F    |
|-----------------|-----|---------|-------|------|
| De cit D        | 1   | 4.513   | 4.513 | ...  |
| Type texte T    | 1   | 63.012  | ...   | ...  |
| Connotation C   | 1   | 1.513   | ...   | ...  |
| Interaction D T | 1   | 1.513   | ...   | ...  |
| Interaction D C | 1   | 21.013  | ...   | ...  |
| Interaction T C | 1   | 1.013   | ...   | ...  |
| Inter. D T C    | 1   | 0.112   | 0.112 | 0.02 |
| Residu          | 72  | 360.700 | 5.010 |      |
| Total           | 79  | 453.388 |       |      |

Table 6 { Analyse de variance (ensemble des facteurs)

| Sources de var. | ddl | SC     | CM    | F    | Prob. |
|-----------------|-----|--------|-------|------|-------|
| De cit D        | 1   | 3.02   | 3.02  | 0.64 | 0.43  |
| Type texte T    | 1   | 24.02  | 24.02 | 5.07 | 0.03  |
| Interaction D T | 1   | 1.22   | 1.22  | 0.26 | 0.61  |
| Residu          | 36  | 170.50 | 4.74  |      |       |
| Total           | 39  | 198.77 |       |      |       |

Table 7 { Analyse de variance (connotation a ective)

| Sources de var. | ddl | SC     | CM    | F    | Prob. |
|-----------------|-----|--------|-------|------|-------|
| De cit D        | 1   | 22.50  | 22.50 | 4.26 | 0.046 |
| Type texte T    | 1   | 40.00  | 40.00 | 7.57 | 0.009 |
| Interaction D T | 1   | 0.40   | 0.40  | 0.08 | 0.785 |
| Residu          | 36  | 190.20 | 5.28  |      |       |
| Total           | 39  | 253.10 |       |      |       |

Table 8 { Analyse de variance (connotation neutre)

## Correlation. Droites de regression

### Exercice 37 Donnees Budget

Il s'agit d'un extrait d'une enquete (ONU 1967) sur les budgets-temps (temps passe dans differentes activites au cours de la journee).

Les colonnes comprennent 3 variables numeriques, le temps passe en : Profession (PROF), Transport (TRAN) et loisirs (LOIS). Les temps sont notes en centiemes d'heures. Le code suivant est utilise pour identifier les lignes :

H : hommes, F : femmes, A : actifs, N : non actifs, M : maries, C : celibataires,  
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 ?

Reponses : 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  $\frac{Var(TRAN)}{Var(TRAN)} = r^2 = 0.87$ .

### 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.

| Elevés | 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 |



Reponses :  $Cov(F; M) = 9,47$  ;  $r = 0,51$ . La corrélation est a peine significative a 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 également être obtenu a l'aide de la formule :

$$r = 1 - \frac{6 \sum d_i^2}{N^2}$$

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, a 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 a partir du temps de réaction a 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 :

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

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.

Reponses : 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 a 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)  $Cov(x_i; y_i) = \frac{676,53294}{9} - \frac{604,63}{9} \cdot \frac{10,059}{9} = 0,084$ .  $Var(x_i) = \frac{40686,3023}{9} - \left(\frac{604,63}{9}\right)^2 = 7,39$ .

$\bar{x}=2.72$ . De même,  $\bar{y} = 0.0344$  et finalement,  $r = \frac{0.084}{2.72 \cdot 0.0344} = 0.899$ . Il existe donc une forte corrélation positive entre ces deux variables.

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

#### 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 :

$$\sum X_i = 1780 ; \sum Y_i = 1654 ; \sum Z_i = 1777 ; n = 10$$

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

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

$$s^2(X) = 117.25 ; s^2(Y) = 40.24 ; s^2(Z) = 111.81$$

$$\text{Cov}(X, Y) = 52.05 ; \text{Cov}(X, Z) = 104.25 ; \text{Cov}(Y, Z) = 63.40$$

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 Z).

On donne par ailleurs :  $\sum Z_i = 1777$  et  $\sum Z_i^2 = 317068.75$ .

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.

Reponses : 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 | Pedagogie | Examen | Connaissances | Resultat | Inscription |
|-----------|-----------|--------|---------------|----------|-------------|
| 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          |

Table 9 { Première partie des données

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

| Qual-Glob | Pedagogie | Examen | Connaissan | Resultat | 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         |

Table 10 { Seconde partie des donnees

3. la qualite des tests et examens (Examen)
4. la connaissance de la matiere dont temoigne le professeur, telle qu'elle est percue par les etudiants (Connaissan)
5. les resultats auxquels s'attendent les etudiants pour ce cours (Resultat, de tres bon a insu sant)
6. le nombre d'inscriptions a ce cours (Inscriptio)

Les resultats de statistiques descriptives concernant les variables precedentes sont donnees dans le tableau 11.

Les coe cients de correlation des variables prises deux a deux sont donnees dans le tableau 12.

Les coe cients de l'equation de regression multiple de la premiere variable en fonction des cinq autres sont donnees par le tableau 13.

Ecrire l'equation de regression correspondante, et la verifier sur l'extrait donne dans le tableau 14.

En n, le dernier tableau (tableau 15) donne les coe cients de correlation partiels entre la variable Qual-Glob et les predicteurs.

|            | Qual-Glob | Pedagogie | Examen   | Connaissan | Resultat | Inscriptio |
|------------|-----------|-----------|----------|------------|----------|------------|
| E ectif    | 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    |

Table 11 { Statistiques descriptives sur les donnees

|            | 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          |

Table 12 { Coe cients de correlation

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

Table 13 { Coe cients de l'equation de de regression

| Ligne | Observ | Ajuste  | Residu    |
|-------|--------|---------|-----------|
| 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  |

Table 14 { Comparaison des valeurs observees et des valeurs ajustees

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

Table 15 { Coe cients de correlation partielle

La valeur du coefficient de corrélation multiple vérifiée :  $R^2 = 0,755$ .

#### 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 |

Table 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

1. 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    |

Table 17 { Correlations 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.

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

Table 18 { Corrélation partielle entre LEPEN et les 5 variables

On peut tester la signification 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 LEPEN 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.

| LEPEN | ETRA       | DELI       | CRCH      | TXCH      |
|-------|------------|------------|-----------|-----------|
|       | 0.73354021 | 0.49258622 | 0.5282723 | 0.4116398 |

Table 19 { Corrélation partielle entre LEPEN et 4 variables

Pour tester la signification 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 |

Table 20 { Correlation partielle entre LEPEN et 3 variables

ont varie en fonction de trois circonstances : le taux d'etrangers, le taux de delinquance, et a un degr moindre, l'evolution du ch^omage.

La regression n'a pas et faite dans un but de prevision, et l'equation de regression n'a qu'un inter^et limite :

$$LEPEN = 0:52 ETRA + 1:69 DELI + 2:37 CRCH \quad 0:4$$

## Analyse en composantes principales

### Exercice 43 Donnees Budget-temps

Il s'agit d'une enqu^ete (ONU 1967) sur les budgets-temps (temps passe dans di erentes activites au cours de la journee).

Le tableau suivant comprend 10 variables numeriques et 4 variables categorisees.

Les 10 variables numeriques sont : le temps passe en : Profession, Transport, Menage, Enfants, Courses, Toilette, Repas, Sommeil, Tele, Loisirs.

Les 4 variables categorisees sont : Le sexe (1=Hommes 2=Femmes), l'activite (1=Actifs 2=Non Act. 9=Non precise), l'etat civil (1=Celibataires 2=Maries 9=Non precise), le Pays (1=USA 2=Pays de l'Ouest 3=Pays de l'Est 4=Yougoslavie).

Le code suivant est utilise pour identi er les lignes : H : Hommes, F : Femmes, A : Actifs, N : Non Actifs(ves), M : Maries, C : Celibataires, U : USA, W : Pays de l'Ouest sauf USA, E : Est sauf Yougoslavie, Y : Yougoslavie

Les temps sont notes en centiemes d'heures. La premiere case en haut a 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 activite PROFessionnelle. Le total d'une ligne (sur ces 10 variables numeriques) 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   |
| FN Y | 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   |
| FM Y | 260  | 52   | 576  | 59   | 116  | 85   | 117  | 775  | 65   | 295  | 2   | 9   | 2   | 4   |
| HC Y | 615  | 125  | 95   | 0    | 115  | 90   | 85   | 760  | 40   | 475  | 1   | 9   | 1   | 4   |
| FC Y | 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 |

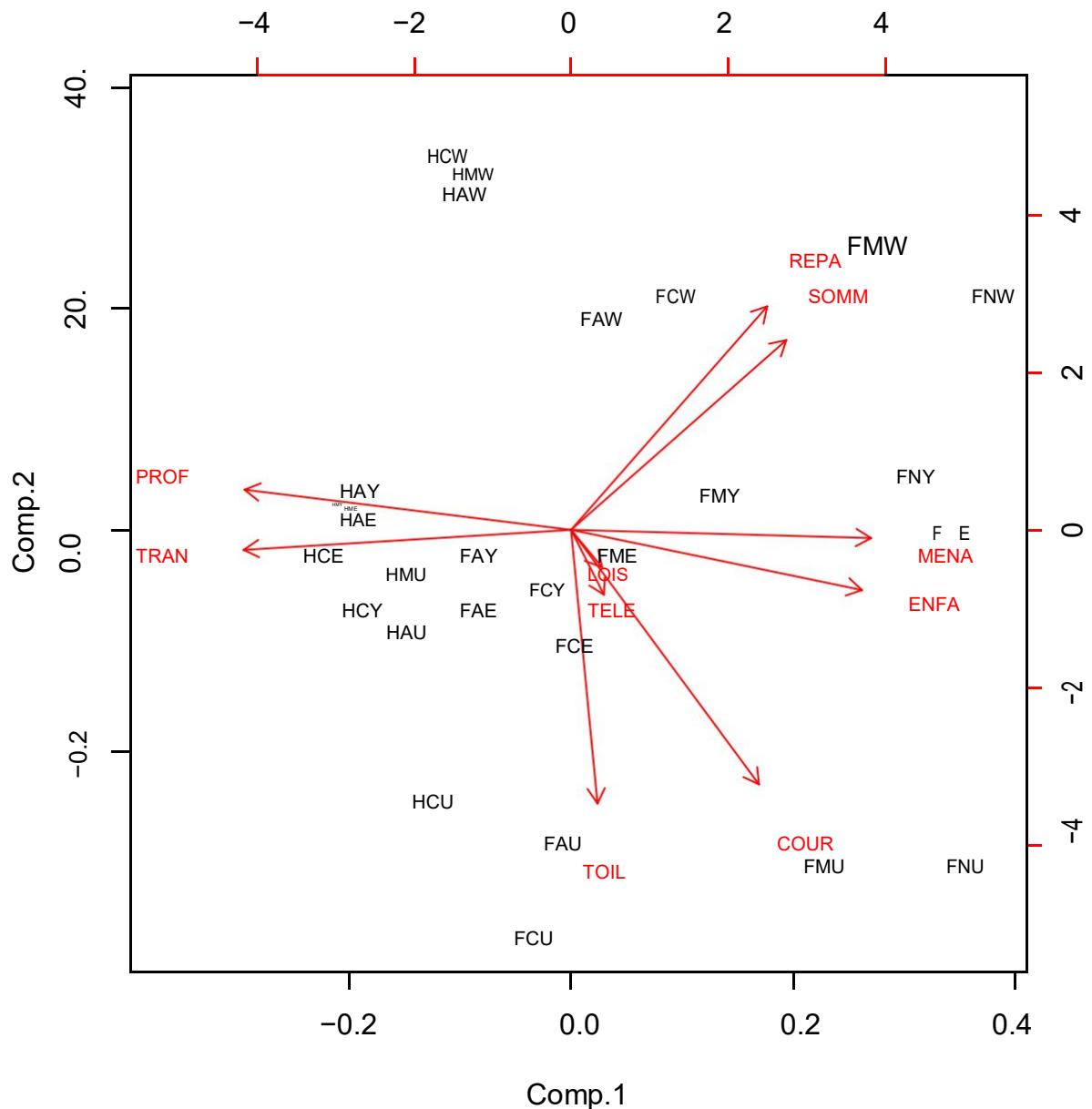
Qualites des representations des individus par les deux premieres composantes principales ; Qualites des representations des individus par le plan (CP1, CP2)

|     | 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 ; qualite de la representation 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 |

Projection du nuage selon les deux premières composantes principales



- 1) Combien de composantes principales proposez-vous pour resumer ces donnees ?
- 2) Quelles sont les coordonnees de l'individu HCU sur les composantes CP1 et CP2. Essayer de placer le point sur le graphique. De m<sup>e</sup>me, quelles sont les saturations de la variable PROF sur CP1 et CP2. Les retrouve-t-on sur le graphique ? Que peut-on penser des graduations des axes fournies par le logiciel ?
- 3) Examiner le tableau des correlations. Quelles variables apparaissent fortement correlees entre elles ?

- 4) Quelles sont les variables qui sont bien representees dans le plan (CP1, CP2) ? Comment retrouve-t-on sur le graphique les correlations entre ces variables ?
- 5) Les variables LOIS et TELE apparaissent proches sur le graphique, alors que leur coefficient de correlation est faible. Comment peut-on l'expliquer ?
- 6) Les points HMW, HCW, HAW apparaissent tres proches sur le graphique. Est-ce le cas dans la realite ?
- 7) Quelles sont les variables les mieux correlees a l'axe CP1 ? Quels sont les individus qui participent le plus a l'inertie de cet axe ? Exprimer les resultats relatifs a l'axe CP1 en termes d'oppositions.
- 8) M<sup>e</sup>me question pour l'axe CP2.
- 9) Quelles sont les variables dont l'etude necessiterait de s'interesser a CP3, CP4, . . . ?

#### Exercice 44 Donnees Psychometrie

Pour 20 eleves (sujets s1 a s20), on a relev les notes obtenues a cinq epreuves individuelles : Combinatoire (Comb), Probabilites (Prob), Logique (Logi), notes de 0 a 10, QI verbal (QI, notes de 85 a 125) et Mathematiques (Math), notee de 0 a 20. Pour chaque sujet, on dispose de deux informations : Pedagogie avec deux modalites p1 (moderne) et p2 (traditionnelle), Milieu avec deux modalites m1 (favorise) et m2 (defavorise).

|     | 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 realise une ACP normee sur les 4 variables Comb, Prob, Logi et Math. L'essentiel des resultats obtenus est indique ci-dessous.

#### Matrice des correlations

|      | 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 qualites des individus sur les deux premieres 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 qualite des variables (2 premieres 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 |

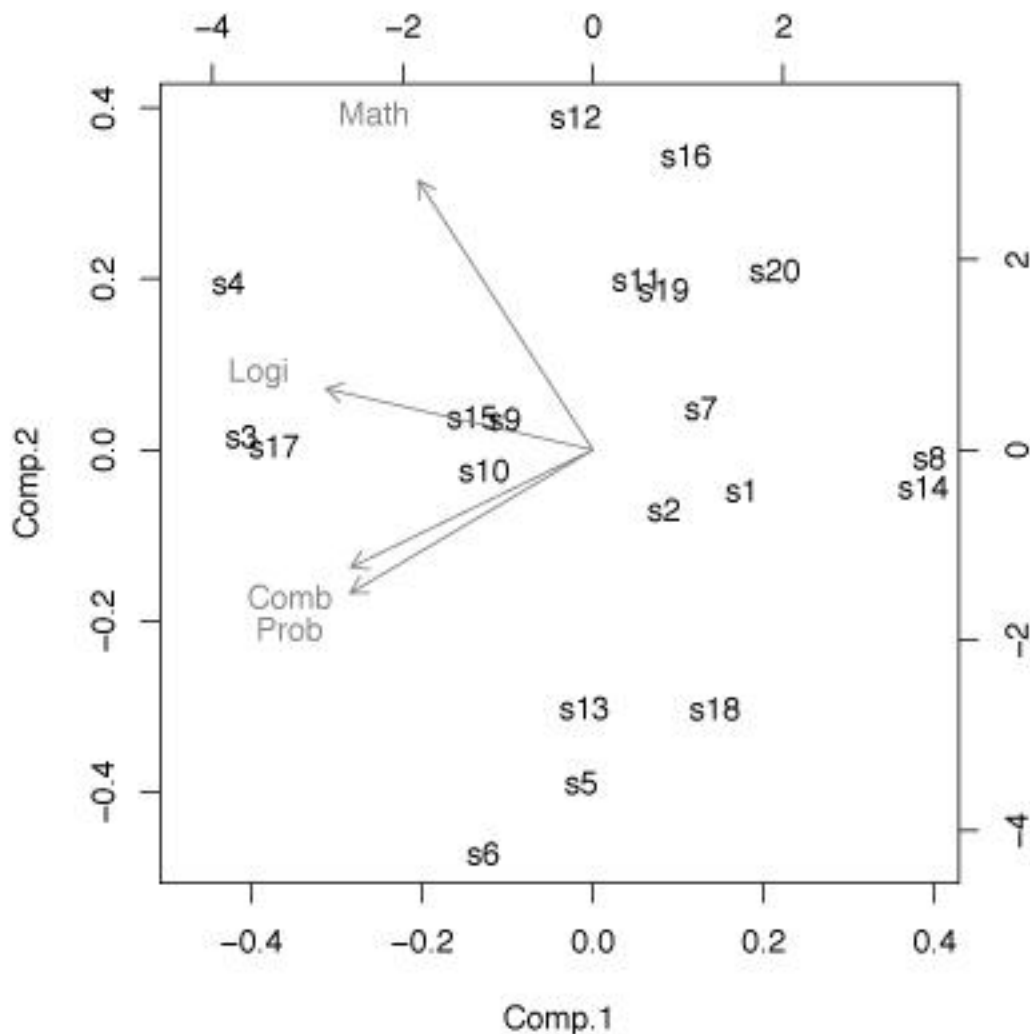


Figure 2 { Cas Psychometrie

- 1) Quelles sont les coordonnées de l'individu s12 sur les composantes CP1 et CP2. Es-sayer de placer le point sur le graphique. De m<sup>^</sup>eme, 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 correlations.
- 3) Les variables Comb et Proba apparaissent proches sur le graphique. Quel est pourtant leur coefficient de correlation ? Comment peut-on l'expliquer ?
- 4) Les points s8 et s14 apparaissent tres proches sur le graphique. Est-ce le cas dans la realite ? M<sup>^</sup>eme question pour s9 et s15.
- 5) Comment les variables contribuent-elles a 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 sur la santé ? 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 Études (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 resultats fournis par Minitab

| Axe   | Inertie | Proportion | Cumule | 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 | FTPB | 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 |

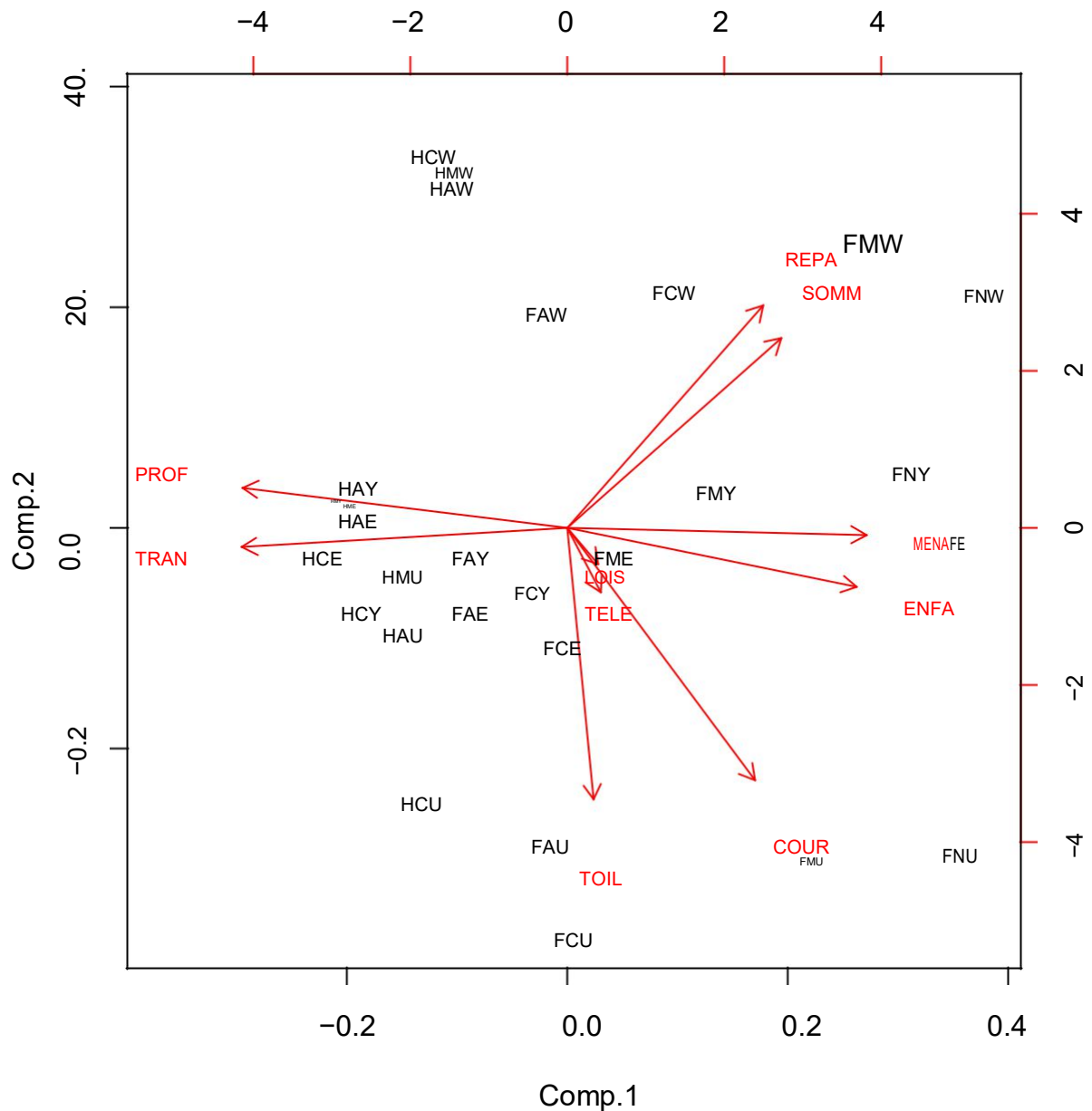


Figure 3 { Graphe factoriel des points lignes et colonnes

1) S'agissant d'un tableau de contingence, indiquer quelle est la methode factorielle adaptee a ce type de tableau (et mise en oeuvre ici).

- 2) Combien de variables factorielles proposez-vous de retenir pour resumer ces donnees ? Discuter les raisons de ce choix.
  - 3) Retrouver approximativement, a l'aide du graphique, la coordonnee factorielle du point CFON sur l'axe 1 (a droite).
  - 4) Sur l'axe 1 (cf. Figure) le point AGRD est plus eloigne (cote negatif) du centre de gravite que ne l'est le point ONOQ. Pourtant la contribution de ce dernier a la variance de l'axe est plus forte (cf. Tableau). Expliquer la raison de ce paradoxe apparent.
  - 5) Que suggere la position du point PCTR, pres de l'intersection des deux axes factoriels ?
  - 6) Que suggere, a priori, la proximite des points PFON et CFON sur le graphe factoriel (cf. Figure) ?
  - 7) Quelle(s) precaution(s) faut-il prendre avant de conclure de nitivement sur la proximite entre ces deux points (PFON et CFON) ? Pourquoi ?
  - 8) On considere tout d'abord uniquement le nuage des CSP. A partir du tableau des aides a l'interpretation ci-dessus (cf. Tableau), construire le tableau necessaire a l'interpretation de la premiere variable factorielle. Commenter le resultat obtenu.
  - 9) Considerant cette fois uniquement les Circonstances (cf. Tableau), proceder de m<sup>e</sup>me (construction du tableau et commentaire).
  - 10) En se fondant sur les resultats precedents (cf. questions 8 et 9), mettre en correspondance les deux nuages (nuage des CSP et nuage des Circonstances) et indiquer ce que suggere principalement cette analyse de la premiere variable factorielle.
- Indications de reponses :
- 1) Une analyse Factorielle des Correspondances (AFC) appelee aussi Analyse des Correspondances (AC).
  - 2) On note que les donnees 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 a retenir les 3 premiers axes. Ce sont ceux dont la contribution a la variance totale est superieure a 7.7%. Ces trois axes reunissent 76% de la variance totale.
  - 3) On trouve environ +0.7 (en fait exactement 0.683)
  - 4) Ceci est d<sup>u</sup> au fait que le poids de ONOQ (71/1000) est plus important que celui de AGRD (37/1000). La contribution a la variance est fonction (du carre) de la distance au centre de gravite, mais aussi du poids relatif.
  - 5) Cela suggere que ces sujets (Professions intermediaires de type Contrema<sup>^</sup>tre) ont un pro I proche du pro I moyen (toutes CSP reunies).
  - 6) Cela suggere que les points ont des pro Is proches (sous la reserve ci-dessous), c'est-a-dire qu'ils rencontrent leurs conjoints dans le m<sup>e</sup>me type de lieu.
  - 7) Il vaut mieux veri er en regardant le tableau de donnees (ou plut<sup>^</sup>ot le tableau des pro Is). En e et cette proximite peut ^etre la consequence d'une deformation des distances due a la projection des points sur un plan alors que le nuage se situe en realit dans un espace de dimension beaucoup plus importante (13).
  - 8) On retient les CTR superieures a  $1/16$  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 (cote negatif ) les Ouvriers (ONOQ, OART et OIND) et Agriculteurs (AGRD), aux cadres (CFON et CENT) et professions intermediaires (PFON et PLCE) qu'ils soient fonctionnaires ou en entreprise (cote positif ).

9) On retient les CTR superieures a 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 (cote negatif ) a la rencontre pendant les etudes (cote positif ).

10) Il appara^t pour l'essentiel (sur ce premier axe) que les Ouvriers et Agriculteurs (de grandes exploitations) tendent a rencontrer leur conjoint dans un Bal, alors que les Cadres et Professions intermediaires rencontrent plut^ot leur conjoint pendant leurs Etudes.

## Classification ascendante hiérarchique

### Exercice 46 Données Conjoint

1) L'histogramme des indices de niveau ci-dessus (cf. Figure 4) suggère une partition des CSP en 3 classes. Indiquer les numéros de ces 3 classes.

2) Quelle est la part de variance dont rend compte cette partition en trois classes ?

3) À partir du dendrogramme ci-dessus (cf. Figure 5) indiquer quelles sont les CSP qui composent chacune des 3 classes principales que l'on nommera arbitrairement A, B et C.

4) Expliquer, à partir des tableaux des données et des profils (cf. dernière feuille), pourquoi il n'est guère surprenant que ARTI et EFCT appartiennent à la même classe : Indications de réponses.

1) c27, c28 et c29 (c28 et c29 étant deux sous-classes de la c30)

2) Cette partition en 3 classes rend compte de 58.4

3) Classe A OIND et OART

Classe B CENT, PEON, CFON, PCTR, PSAN

Classe C ONOQ, OTRA, EENT, PLCE, ARTL EFCT, COMM. APTL AGRD.

4) Ils ont des profils relativement proches (relativement aux autres). Si l'on considère les pourcentages de rencontre dans les différents lieux on trouve, respectivement pour ces deux CSP, 3% et 7% (au cours des Études), 6% et 4% (Vacances), 4% et 3% (lors d'une fête entre Amis), 8% et 6% (dans une association Sportive), 15% et 17% (au Travail), 6% et 6% (rencontre chez des particuliers), 1% et 0% (par Annonce ou Agence).