

Analyse Statistique des Données – Niveau 2 (ASD2)

Aucun document n'est autorisé. La calculatrice n'est pas autorisée.
Les différents exercices (encadrés) sont indépendants.

Le barème, indicatif, est sur 60 ; la note finale sera donnée sur 20.

**INDIQUER LES REPONSES EXCLUSIVEMENT SUR CE DOCUMENT.
NE RIEN ECRIRE DANS LES MARGES.**

De nombreuses questions portent sur le dossier CATEG présenté sur la dernière page.

Les calculs et les graphiques ont été réalisés avec le logiciel SES-Pegase.

1 Dossier CATEG – Types de variables (3 points)

(cf. présentation des données CATEG sur la dernière page)

**1.1 Indiquer ci-dessous le type de chacune des six variables suivantes,
en deux catégories (Nominale / Numérique)**

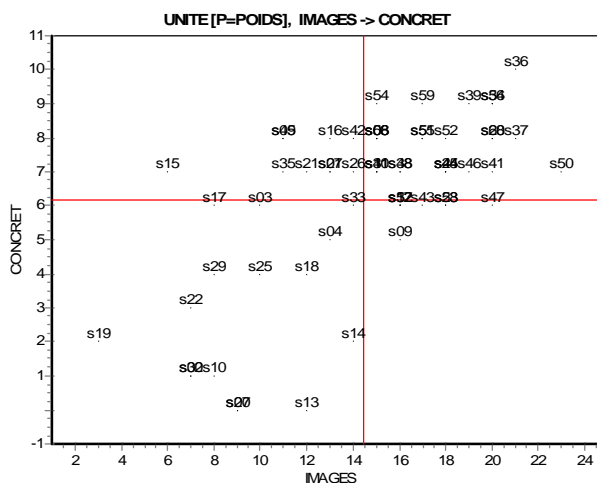
CLASSE :	Nominale	AGE :	Numérique
SEX :	Nominale	IMAGES :	Numérique
CONCRET :	Numérique	ABSTRAIT :	Numérique

1.2 Indiquer ci-dessous quelles sont, parmi ces variables, les variables de rapport :
Concret, Age, Images, Abstrait

2 Dossier CATEG – Préviation (17 points)

(cf. présentation des données CATEG sur la dernière page)

On s'intéresse à la liaison entre le score de catégorisation d'images (IMAGES) et le score de catégorisation de mots concrets (CONCRET). On obtient le graphique suivant :



Diplôme : Licence 2 - Psychologie

CORRIGÉ

Session mai 2011 - Semestre 2
ASD2 - Examen final
Année universitaire 2010 / 2011

Durée de l'épreuve : 1 h 30

Coefficient : 5

Responsable : Denis CORROYER

Nombre de pages : 1 / 8

2.1 Donner trois synonymes (français ou anglais) utilisés pour désigner ce type de graphique :

Nuage bivarié, Nuage des individus, Graphique de corrélation, Scatterplot, Diagramme de corrélation, Diagramme de dispersion, Nuage de points, Diagramme cartésien

2.2 On s'intéresse à la forme de la liaison entre ces deux variables.

a. Pourquoi ce graphique n'est-il pas le graphique idéal pour répondre à cette question ?

Parce que la présence de points superposés risque de cacher la nature de la liaison.

b. Indiquer un autre type de graphique plus adéquat (décrire en quoi consiste ce graphique) :

Il est préférable d'utiliser le graphique pondéré où la taille d'un point est proportionnelle au nombre d'individus en ce point.

2.3 On considère que la liaison est de type linéaire. Pourquoi peut-on qualifier cette liaison linéaire de positive ? Donner deux arguments s'appuyant sur :

a. L'ajustement du nuage par une droite :

Si on ajuste ce nuage par une droite, celle-ci a une pente positive (ascendante, de la gauche vers la droite)

b. La répartition des sujets dans les quatre quadrants :

*On trouve beaucoup de sujets dans deux des quatre quadrants :
- en bas à gauche
- en haut à droite*

2.4 On calcule deux indices de liaison linéaire, la covariance (notée Cov) et le coefficient de corrélation linéaire de Bravais-Pearson (noté r).

a. Quel est l'inconvénient de la covariance - et l'avantage de la corrélation - pour évaluer la force de la liaison ?

*La covariance (Cov)
a des bornes (minimum et maximum) spécifiques à chaque couple de variables,
précisément fonction des écarts-type des deux variables.
La corrélation (r)
a des bornes fixes (-1, +1).*

b. Quelle est la relation entre ces deux indices (indiquer la formule permettant de passer de l'un à l'autre) :

$$r = \frac{\text{Cov}(X,Y)}{\text{Ety}(X) \times \text{Ety}(Y)}$$

c. Expliquer pourquoi les deux indices Cov et r sont nécessairement de même signe :

*Un écart-type étant toujours positif, le dénominateur de la formule précédente est toujours positif.
La corrélation (r) est donc de même signe que le numérateur (Cov).*

Diplôme : Licence 2 - Psychologie		CORRIGÉ	
Session mai 2011 - Semestre 2 ASD2 - Examen final Année universitaire 2010 / 2011		Durée de l'épreuve : 1 h 30 Coefficient : 5	
Responsable : Denis CORROYER		Nombre de pages : 2 / 8	

2.5 On se demande s'il est possible de prédire le score de catégorisation de mots concrets (CONCRET) connaissant le score de catégorisation d'images (IMAGES).

On trouve l'équation de régression suivante : $CONCRET_{pred} = 0.4 \times IMAGES + 0.9$

a. Pourquoi devait-on s'attendre à un coefficient de régression (0.4) positif ?

Parce que (cf. graphique) :

- la liaison entre ces deux variables (Images et Concret) est positive,

- la droite qui ajuste le nuage bivarié a une pente positive.

b. Que prédit le coefficient 0.9 ?

Si quelqu'un a un score nul en catégorisation d'images, il aura un score de 0.9 en catégorisation de mots concrets.

c. Quel score en catégorisation de mots concrets (CONCRET) peut-on prédire à un sujet qui aurait un score de 10 à la catégorisation d'images (IMAGES) ? Indiquer le détail des calculs et la valeur obtenue :

$$Concret_{pred} = 0.4 \times 10 + 0.9 = 4.9$$

d. On trouve $R^2 = .41$. Rédiger une conclusion descriptive sur la qualité de la prédiction :

Pour ces 60 enfants provenant de 3 écoles d'une ville de la région parisienne,

le score en catégorisation d'images

permet de prédire 41 % de la variance

des scores en catégorisation de mots concrets,

ce qui est une proportion importante

($R^2 = 41 \% > 16\%$)

3 Dossier CATEG - Comparaison des classes (4 points)

(cf. présentation des données CATEG sur la dernière page)

On fait l'hypothèse que les CM1 réussissent mieux que les CP la tâche de catégorisation des mots concrets (CONCRET).

On considère, pour cette variable, qu'une différence est faible si inférieure à 1 point (sur 20) et importante au-delà de 1.5 points.

On trouve les résultats suivants :

- moyennes : $m_{CP} = 4.83$ et $m_{CM1} = 7.50$.

- test t de Student : $t = 4.88$, $ddl = 58$, $p < .001$

- Intervalle de Confiance sur la différence parente : $IC (.05) = [1.56 ; 3.77]$

Rédiger une conclusion inférentielle détaillée :

Chez les enfants de CP et CM1 de cette ville de la région parisienne,

le nombre de bonnes réponses

à la tâche de catégorisation des mots concrets

est plus élevé chez les enfants de CM1 que chez les CP

($t = 4.88$, $ddl = 58$, $p < .001$)

La différence peut être considérée comme importante

car supérieure à 1.5 mots

($IC(.05) = [1.56 ; 3.77]$)

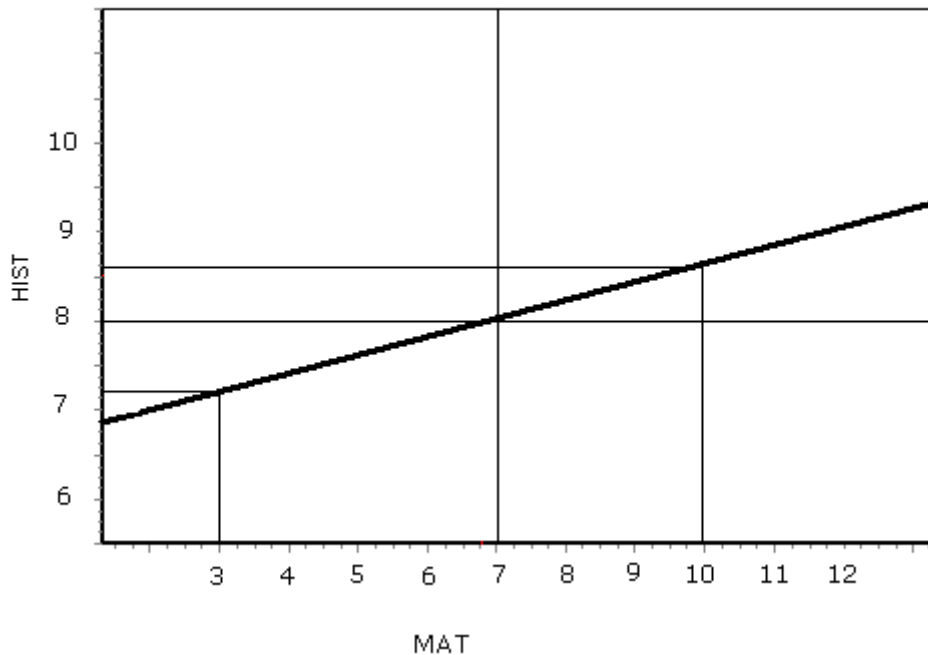
Diplôme : Licence 2 - Psychologie	CORRIGÉ
Session mai 2011 - Semestre 2 ASD2 - Examen final Année universitaire 2010 / 2011	Durée de l'épreuve : 1 h 30 Coefficient : 5
Responsable : Denis CORROYER	Nombre de pages : 3 / 8

4 Graphe de corrélation (6 points)

Soit les données bivariées numériques suivantes : notes en MATH (sur 20) et notes en HISTOIRE (sur 10).

Sujet	s1	s2	s3	s4	s5
MATH	3	12	8	5	7
HIST	7	10	7	10	6

a. Préparer ci-dessous le graphe de corrélation pour ces données (indiquer les échelles sur les deux axes). On ne demande PAS de positionner les sujets.



b. Tracer les deux axes moyens (indiquer ci-dessous les calculs effectués) :

$$\text{Moy (MATH)} = (3 + 12 + 8 + 5 + 7) / 5 = 35 / 5 = 7$$

$$\text{Moy (HIST)} = (7 + 10 + 7 + 10 + 6) / 5 = 40 / 5 = 8$$

c. Tracer, sur ce graphique précédent, la droite de régression, d'équation :
 $\text{HIST} = 0.2 \text{ MATH} + 6.6$ (indiquer ci-dessous les calculs effectués)

Par exemple :

- si $\text{MATH} = 2$, $\text{HIST} = 0.2 \times 2 + 6.6 = 7.0$

- si $\text{MATH} = 10$, $\text{HIST} = 0.2 \times 10 + 6.6 = 8.6$

(Le calcul d'un seul point suffit car on sait que la droite passe par le point moyen)

Diplôme : Licence 2 - Psychologie

CORRIGÉ

Session mai 2011 - Semestre 2
ASD2 - Examen final
Année universitaire 2010 / 2011

Durée de l'épreuve : 1 h 30
Coefficient : 5

Responsable : Denis CORROYER

Nombre de pages : 4 / 8

5 Dossier CATEG – Comparaison des taux de réussites (10 points)

(cf. présentation des données CATEG sur la dernière page)

5.1 Le chercheur fait l'hypothèse que le taux de réussite pour certains mots concrets très simples, ne progresse pas, ou peu, du CP au CM1. On s'intéresse ici à la question portant sur les animaux (ANIM). On observe la répartition suivante :

	1	0	Total
CP	23	7	30
CM1	28	2	30
Total	51	9	60

a. On trouve, pour le test du χ^2 : $\chi^2 = 3.27$ et $p = .07$. Indiquer le résultat du test (Significatif / Non Significatif) en justifiant :

*Non significatif
car $p > .05$*

b. Que conclure par rapport à l'hypothèse du chercheur ?

*Ce test est non significatif
mais il ne permet pas au chercheur de confirmer son hypothèse
car un test non significatif ne permet pas de conclure à une absence de différence,
ni même à une différence faible.*

5.2 On s'intéresse maintenant à la comparaison des deux classes du point de vue de la réussite à la question portant sur les fleurs (FLEU). On trouve la répartition suivante :

	1	0	Total
CP	20	10	30
CM1	27	3	30
Total	47	13	60

On constate que, dans cet échantillon, les enfants de CM1 réussissent mieux que les enfants de CP. Pour évaluer l'ampleur de l'écart entre les deux groupes du point de vue de leur réussite, on peut calculer différents indices.

Indiquer, pour chacun des indices suivants, le calcul à effectuer pour en trouver la valeur (développer les calculs mais il n'est pas demandé de calculer le résultat final) :

a. Différence de pourcentages :

$$d = (27 / 30) - (20 / 30) = 7 / 30 (= 0.23)$$

b. Différence relative des pourcentages :

$$dr = (27/30 - 20/30) / (20/30) = (27 - 20) / 20 = 7 / 20 (= 0.35)$$

c. Rapport des pourcentages ou risque relatif :

$$rr = (27/30) / (20 / 30) = 27 / 20 (= 1.35)$$

ou

$$rr = dr + 1 = 0.35 + 1 = 1.35$$

Diplôme : Licence 2 - Psychologie

CORRIGÉ

Session mai 2011 - Semestre 2
ASD2 - Examen final
Année universitaire 2010 / 2011

Durée de l'épreuve : 1 h 30

Coefficient : 5

Responsable : Denis CORROYER

Nombre de pages : 5 / 8

d. Odds ratio (*or*). Indiquer la procédure de calcul des deux odds, ou chances, puis le calcul de l'odds ratio, ou rapport des chances :

$$odds_{CP} = 20 / 10 (= 2)$$

$$odds_{CM1} = 27 / 3 (= 9)$$

$$or = (20 / 10) / (27 / 3) = 2 / 9 (= 0.22)$$

ou

$$or = (27 / 3) / (20 / 10) = 9 / 2 (= 4.5)$$

6 Dossier CATEG - Comparaison de mesures répétées (3 points)

(cf. présentation des données CATEG sur la dernière page)

Le chercheur fait l'hypothèse que les écarts de pourcentages de réussite peuvent être importants entre certains items abstraits. Il compare en particulier les taux réussite à l'item SENS (Goût et Odorat) et QUAL (Beauté et Gentillesse). On observe la répartition suivante des 60 enfants selon leur réussite aux deux items.

		QUAL		Total
		1	0	
SENS	1	3	13	16
	0	0	44	44
Total		3	57	60

6.1 Indiquer le mode de calcul des deux pourcentages de réussite (le calcul du résultat final n'est pas demandé) :

- pour l'item SENS, $p_{SENS} = 16 / 60 (= .27 = 27\%)$

- pour l'item QUAL, $p_{QUAL} = 3 / 60 (= .05 = 5\%)$

6.2 Indiquer le nom précis du test permettant de se prononcer sur l'existence d'une différence entre ces deux pourcentages dans la population :

Test du χ^2 de McNemar

7 Procédures de calcul (14 points)

On a les scores (x) d'un échantillon de 10 sujets répartis en deux groupes g1 et g2 d'effectifs $n_1 = 6$ et $n_2 = 4$.

Tableau 1

Tableau 2

	g1						g2			
	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10
X	3	6	0	1	0	2	8	4	9	7
$x-m$	-1	2	-4	-3	-4	-2	4	0	5	3
$x-m_g$	1	4	-2	-1	-2	0	1	-3	2	0

	n_g	m_g	v_g
g1	6	2	4.33
g2	4	7	3.50

On a reporté quelques résultats :

- Dans le tableau 1, les écarts à la moyenne générale ($x-m$) et les écarts intra-groupes ($x-m_g$)
- Dans le tableau 2, les effectifs (n_g), les moyennes (m_g), et les variances (v_g) de chacun des deux groupes.
- On sait par ailleurs que la moyenne globale, $m = 4$, $V_{inter} = 6.0$ et $V_{intra} = 4.0$

Diplôme : Licence 2 - Psychologie	CORRIGÉ
Session mai 2011 - Semestre 2 ASD2 - Examen final Année universitaire 2010 / 2011	Durée de l'épreuve : 1 h 30 Coefficient : 5
Responsable : Denis CORROYER	Nombre de pages : 6 / 8

a. La moyenne générale des 10 scores est égale à 4. Indiquer comment retrouver cette moyenne générale à partir des valeurs du tableau 2 (détailler les calculs jusqu'au résultat).

$$m = 4 = \frac{(6 \times 2) + (4 \times 7)}{10} = \frac{12 + 28}{10} = \frac{40}{10} = 4.0$$

b. Indiquer comment calculer la variance totale à partir des valeurs du tableau 1 (indiquer seulement le principe du calcul sans calculer le résultat final) :

$$V_{\text{totale}} = \frac{(-1)^2 + 2^2 + \dots + 3^2}{10}$$

c. Indiquer un autre moyen d'obtenir la variance totale. Indiquer la formule utilisée et le résultat final.

$$V_{\text{totale}} = V_{\text{inter}} + V_{\text{intra}} = 6.0 + 4.0 = 10.0$$

d. Indiquer deux procédures pour calculer la variance intra-groupes (notée V_{intra}) à partir des valeurs du tableau 1, puis à partir des valeurs du tableau 2 (indiquer seulement le principe du calcul) :

$$V_{\text{intra}} = \frac{1^2 + 4^2 + (-2)^2 + \dots + 0^2}{10} \quad \left(= \frac{1 + 16 + 4 + \dots + 0}{10} = \frac{40}{10} = 4 \right)$$

$$V_{\text{intra}} = \frac{(6 \times 4.33) + (4 \times 3.50)}{10}$$

e. Indiquer comment calculer la variance inter-groupes, notée V_{inter} (développer les calculs jusqu'au résultat) :

$$V_{\text{inter}} = \frac{6 \times (2 - 4)^2 + 4 \times (7 - 4)^2}{10} = \frac{(6 \times 4) + (4 \times 9)}{10} = \frac{24 + 36}{10} = 6.0$$

ou

$$V_{\text{inter}} = V_{\text{totale}} - V_{\text{intra}} = 10 - 4 = 6$$

f. Calculer le rapport de corrélation, Eta^2 (indiquer la formule et le détail des calculs) :

$$Eta^2 = \frac{V_{\text{inter}}}{V_{\text{totale}}} = \frac{V_{\text{inter}}}{V_{\text{inter}} + V_{\text{intra}}} = \frac{6}{6 + 4} = 0.60 = 60\%$$

8 Formules (3 points)

Démontrer l'équivalence des deux formules de F pour la régression :

$$F = (n - 2) \frac{R^2}{1 - R^2}$$

$$F = (n - 2) \frac{Var(Y_{\text{pred}})}{Var(Y_{\text{res}})}$$

$$R^2 = \frac{Var(Y_{\text{pred}})}{Var(Y_{\text{obs}})}$$

$$\text{d'où : } 1 - R^2 = 1 - \frac{Var(Y_{\text{pred}})}{Var(Y_{\text{obs}})} = \frac{Var(Y_{\text{obs}}) - Var(Y_{\text{pred}})}{Var(Y_{\text{obs}})} = \frac{Var(Y_{\text{res}})}{Var(Y_{\text{obs}})}$$

$$\text{d'où : } \frac{R^2}{1 - R^2} = \frac{Var(Y_{\text{pred}}) / Var(Y_{\text{obs}})}{Var(Y_{\text{res}}) / Var(Y_{\text{obs}})} = \frac{Var(Y_{\text{pred}})}{Var(Y_{\text{res}})}$$

$$\text{d'où : } F = (n - 2) \frac{R^2}{1 - R^2} = (n - 2) \frac{Var(Y_{\text{pred}})}{Var(Y_{\text{res}})}$$

Diplôme : Licence 2 - Psychologie

CORRIGÉ

Session mai 2011 - Semestre 2
ASD2 - Examen final
Année universitaire 2010 / 2011

Durée de l'épreuve : 1 h 30

Coefficient : 5

Responsable : Denis CORROYER

Nombre de pages : 7 / 8

Données CATEG - Présentation

On a recueilli ces données sur 60 enfants : 30 enfants de CP et 30 enfants de CM1 (cf. colonne CLASSE) tirés au hasard parmi les trois écoles d'une ville de la région parisienne. On a relevé leur Age en mois (cf. colonne AGE) et leur Sexe (cf. colonne SEX).

Ces données comportent les résultats de deux épreuves :

- une épreuve de catégorisation d'images,
- une épreuve de catégorisation de mots.

Pour la catégorisation d'images, on a relevé un score global (cf. colonne IMAGES)

Pour la catégorisation de mots, il s'agissait de répondre à des questions du type « En quoi un chat et un chien se ressemblent ? » (mots concrets) ou « En quoi l'hiver et l'été se ressemblent (mots abstraits).

Dix questions portaient sur des mots concrets et dix questions portaient sur des mots abstraits.

Une bonne réponse est notée 1 ; une mauvaise réponse est notée 0.

Le tableau de données ci-dessous ne comporte qu'une partie des résultats :

- les résultats (0/1) pour quatre couples de mots concrets : Chien et Chat (cf. colonne (ANIM), Rose et Marguerite (cf. colonne FLEU), Sapin et Pommier (cf. colonne ARBR), Marteau et Hache (cf. colonne OUTI).
- le nombre total de bonnes réponses pour les 10 mots concrets (cf. colonne CONCRET),
- les résultats (0/1) pour quatre couples de mots abstraits : Hiver et Été (cf. colonne SAIS), SENS (Goût, Odorat), Beauté et Gentillesse (cf. colonne QUAL), Samedi et Dimanche (cf. colonne JOUR),
- le nombre total de bonnes réponses pour les 10 mots abstraits (cf. colonne ABSTRAIT).

Source : *inspiré de Rozencwajg, P. (2007). Cognitive Process Development as Measured by an Adapted Version of Wechsler's Similarities Test. Learning and Individual Differences, 17-4, 298-306.*

SUJET	CLASSE	AGE	SEX	IMAGES	ANIM	FLEU	ARBR	OUTI	(...)	CONCRET	SAIS	SENS	QUAL	JOUR	(...)	ABSTRAIT
s01	CP	83	F	13	1	1	1	1		7	1	0	0	1		3
s02	CP	76	G	7	0	0	0	0		1	0	0	0	0		0
s03	CP	86	F	10	1	1	1	1		6	1	0	0	1		5
s04	CP	83	G	13	1	1	0	1		5	0	0	0	1		2
s05	CP	83	G	11	1	1	1	1		8	0	0	0	1		4
s06	CP	80	G	15	1	1	1	1		8	1	0	0	1		3
s07	CP	76	G	9	0	0	0	0		0	0	0	0	1		1
s08	CP	86	F	15	1	1	1	1		8	0	0	0	0		1
s09	CP	86	F	16	1	1	1	0		5	1	0	0	0		1
s10	CP	79	G	8	1	0	0	0		1	0	0	0	1		1
s11	CP	77	F	15	1	1	1	0		7	1	0	0	1		4
s12	CP	82	F	16	1	1	1	0		6	0	0	0	0		1
s13	CP	77	G	12	0	0	0	0		0	0	0	0	1		2
s14	CP	86	F	14	0	0	1	0		2	0	0	0	1		1
s15	CP	84	G	6	1	1	1	0		7	0	0	0	1		2
s22	CP	75	F	7	1	0	0	0		3	0	0	0	1		1
s23	CP	76	G	18	1	0	1	1		6	1	0	0	1		2
s24	CP	83	G	18	1	1	1	1		7	1	1	0	1		4
(...)																
s52	CM1	117	F	18	1	1	1	1		8	1	0	0	1		5
s53	CM1	113	F	15	1	1	0	1		8	1	0	0	1		5
s54	CM1	116	F	15	1	1	1	1		9	1	1	0	1		5
s55	CM1	119	F	17	1	1	1	1		8	1	1	0	1		8
s56	CM1	121	G	20	1	1	1	1		9	1	1	1	1		9
s57	CM1	119	G	16	1	0	0	1		6	1	0	0	0		3
s58	CM1	121	G	18	1	1	1	0		6	1	1	0	1		8
s59	CM1	117	G	17	1	1	1	1		9	1	1	0	1		8
s60	CM1	117	G	17	1	1	1	1		9	1	1	0	1		8

Diplôme : Licence 2 - Psychologie

CORRIGÉ

Session mai 2011 - Semestre 2
ASD2 - Examen final
Année universitaire 2010 / 2011

Durée de l'épreuve : 1 h 30

Coefficient : 5

Responsable : Denis CORROYER

Nombre de pages : 8 / 8