

Durée de l'épreuve : 3 heures.

Épreuve sans document. La calculette est autorisée (sans sa documentation).

Les données des dossiers LECTURE1 et LECTURE2, ainsi qu'un extrait de la table du t de Student se trouvent sur les dernières pages. Ces pages peuvent être détachées et conservées.

Les exercices (encadrés) sont indépendants. Le barème (sur 32) est donné à titre indicatif.

Indiquer les réponses exclusivement sur ce document et aux endroits réservés à cet effet (ne pas écrire dans la marge)

Les logiciels Addad, DS3Win et Statistica ont été utilisés pour analyser ces données.

Analyses univariées - Dossier « Lecture 1 »

(29,5 points)

Source : d'après A. Charvillat et P. Colé. Rôle de l'interaction morphologie/lexique dans le traitement syntaxique. (en cours).

Dans une étude portant sur la compréhension durant la lecture, des chercheurs ont utilisé un matériel linguistique incluant 32 phrases expérimentales et 64 phrases de remplissage. Tous les items expérimentaux se composent d'une proposition principale suivie d'une subordonnée relative (ex : *Marc a salué la mère du marié qui était...*). Diverses recherches suggèrent qu'en français, les phrases de ce type sont en moyenne comprises plus rapidement quand la relative est attachée au premier nom (...*la mère du marié qui était élégante*) plutôt qu'au second (...*la mère du marié qui était élégant*). Toutefois, le fait que certains individus s'écartent de ce pattern moyen a conduit les auteurs à rechercher les déterminants de ces différences individuelles.

Les 26 adultes français de l'échantillon (échantillonnage aléatoire) font apparaître les phrases mot après mot sur un écran, à leur propre rythme, en pressant une touche du clavier. Toutes les phrases expérimentales se terminent par un mot cible (adjectif) présenté dans un graphisme différent du reste de la phrase.

Exemple : *Marc a salué la mère du marié qui était très élégante.*

Ce mot cible varie :

- en fréquence, de telle sorte qu'il est tantôt fréquent (f1), tantôt peu fréquent (f2)

- et en genre (masculin/féminin), de telle sorte que la relative s'attache tantôt au premier nom (n1), comme dans l'exemple ci-dessus, tantôt au second (n2 : ...*très élégant*).

Par ailleurs, le contraste phonologique masculin-féminin de l'adjectif cible a été manipulé : pour la moitié des cibles, le genre est phonologiquement contrasté (p1 ; ex. *élégant-élégante* se prononcent différemment) alors que pour l'autre moitié, ce contraste est absent (p2 ; ex : *ahuri-ahurie* ont la même prononciation).

Dès l'apparition de la cible, les sujets doivent décider le plus rapidement possible s'il s'agit ou non d'un mot français en pressant une touche, à droite pour un mot (ex. *élégant*), à gauche pour un non-mot (ex. *élugant*). On recueille les temps de décision lexicale (TDL) en millisecondes (la cible correspondant aux phrases expérimentales est toujours un mot mais, au total, les sujets doivent juger autant de mots que de non-mots).

On a également mesuré le temps de lecture moyen par phrase de chaque sujet de façon à repérer deux types de lecteurs (I1 : lecteurs rapides, I2 : lecteurs lents) dont on envisage de comparer les stratégies de compréhension.

Les données, qui figurent en annexe 1 ne concernent que les décisions lexicales correctes (ex : MOT pour « *élégant* » ; NON-MOT pour « *élugant* »).

Les annexes (pages 12 à 14) peuvent être détachées du reste du fascicule.

A. Structures (On s'aidera du texte ci-dessus et de l'annexe 1)

($\Sigma = 7,5$ pts)

1/ Explicitez la relation entre les facteurs suivants ; utilisez les symboles \leftrightarrow et *, sans omettre de spécifier les indices. **(2,5 pts)**

a. L et F : $L_2 * F_2$ **(0,5 pt)**

b. S et N : $S_{26} * N_2$ **(0,5 pt)**

c. F et P : $S_{26} * P_2$ **(0,5 pt)**

d. S, N et L : $S_{13} < L_2 > * N_2$ **(0,5 pt)**

e. N, P et L $N_2 * P_2 * L_2$ (0,5 pt)

2/ Si l'on fait abstraction du facteur de groupe, combien a-t-on de relations binaires ? (précisez le mode de calcul adopté)

$$C^2_4 = 4! \div (2!) \times (2!) = 6 \quad (1 \text{ pt})$$

3/ Identifiez

a. le(s) G-facteur(s) : L (0,5 pt)

b. le(s) T-facteur(s) : F N P (1,5 pt)

4/ Indiquez la formule

a. du plan le plus riche : $S13 < L2 > * F2 * N2 * P2$ (2 pts)

b. d'un (facteur-)plan complet : $S26 * F2 * N2 * P2$ (2 pts)

B. Protocoles dérivés ($\Sigma = 4.5 \text{ pts}$)

L'Annexe 1 donne les temps de décision lexicale (TDL) des 26 sujets dans chaque condition expérimentale. En vous référant aux colonnes de droite, précisez les PDP permettant d'analyser les effets suivants et détaillez la procédure de calcul pour s1 (retrouvez la valeur exacte à deux décimales). (0,5 pt par PDP + 1 pt par procédure de calcul)

1/ effet de la fréquence de l'adjectif cible : (1,5 pt)

PDP 3

$$s1 = \frac{(2134+1840+2581+2465)}{4} - \frac{(2104+2376+2129+2415)}{4} = -1$$

2/ effet de la phonologie pour les cibles fréquentes : (1,5 pt)

PDP 4

$$s1 = \frac{(2134+2581)}{2} - \frac{(1840+2465)}{2} = 205$$

3/ effet d'interaction entre la position du nom s'accordant avec la cible et le type de lecteur : (1,5 pt)

PDP 2

$$s1 = \frac{(2134+1840+2104+2376)}{4} - \frac{(2581+2465+2129+2415)}{4} = -284$$

C. Analyse descriptive ($\Sigma = 12,5 \text{ pts}$)

1/ Hypothèse 1

Sachant que la lecture est plus lente quand on prononce mentalement les mots que l'on lit, on peut faire l'hypothèse que la lenteur relative de certains lecteurs tient en partie à leur utilisation de la phonologie en lisant. Si tel est le cas, on s'attend à ce que chez les lecteurs lents, la présence ou l'absence d'un contraste phonologique sur la cible ait un effet sur les TDL.

a. Donnez l'expression formelle de l'effet recherché. P/l2 (1 pt)

b. Précisez la procédure (différence de moyennes ou moyenne d'effets individuels) permettant de calculer d_{obs} (l'effet recherché) à partir du PDP, et donnez sa valeur.

Procédure : *moyenne des effets individuels (de P)* (0,5 pt)
 $d_{\text{obs}} = 120.63$ (1 pt)

c. On obtient un effet calibré $EC = 0.91$. Donnez sa formule.

$$ECM = d_{\text{obs}} / s \quad (1 \text{ pt})$$

d. Formulez la conclusion descriptive.

(2 pts)

Chez ces 13 lecteurs lents, les décisions lexicales sont plus lentes quand la cible comporte un contraste phonologique que quand elle n'en comporte pas ($d_{obs} = 121$ ms). Cet effet est important selon le critère psychométrique ($ECM_{obs} = 0.91 > 2/3$)

(portée + sens + importance + VD spécifiée)

2/ Hypothèse 2

La compréhension d'une phrase complète suppose la mémorisation temporaire des mots qui la composent. Mais comme la trace mnésique des mots s'affaiblit au fil du temps, on peut supposer que (dans une certaine mesure) le fait de lire lentement accroît le risque d'oublier les mots les moins récents. On fait donc l'hypothèse que l'effet de la position du nom (n1 ou n2) auquel s'attache la subordonnée relative sera différent chez les lecteurs rapides et chez les lecteurs lents.

a. Donnez l'expression formelle de cet effet N.L (ou bien N/I1 – N/I2)

(1 pt)

b. Le PDP correspondant (PDP2) est le même que celui d'un effet global. Lequel ? (Justifiez) (2 pts)

Même PDP que pour N (effet global de la position du nom) mais on dissocie les sujets rapides (I1) et les sujets lents (I2) (1 + 1)

c. Explicitez la procédure de calcul de d_{obs} (l'effet recherché) à partir du PDP, et donnez sa valeur.

Procédure : différence entre les moyennes N/I1 et N/I2

(1 pt)

$d_{obs} = - 144.75 - 106.48 = - 251.23$ (ms)

(2 pts)

d. On obtient un effet calibré **EC = 1.44**. Donnez sa formule.

$EC_{\underline{G}} = S_{inter} \div S_{intra}$

(1 pt)

D. Inférence ($\Sigma = 5$ pts)

En inférant sur l'effet correspondant à l'hypothèse 2 dans la population parente, on trouve $t_{obs} = 5.20$.

1/ Donnez le nombre de degrés de liberté du test.

$ddl = n - 2 = 24$

(calcul + résultat : 1 pt)

2/ En vous référant à la table du t de Student (Annexe 3), donnez le résultat du test.

$t_{obs} = 5.20 > t_{[24].001} = 3.745$ ou test significatif (au seuil .0005 unilatéral). (1 pt)

3/ Formulez une conclusion générale (descriptive et inférentielle).

(3 pts)

Chez ces 26 sujets, il existe une interaction entre la position du nom auquel s'attache la relative et la vitesse de lecture. Les décisions lexicales sont plus rapides quand la relative est attachée à n1 mais cet effet est moins marqué chez les lecteurs lents que chez les lecteurs rapides ($d_{obs} = - 251.23$ millisecondes). Cet effet d'interaction est important au regard du critère psychométrique ($ECG_{obs} = 1.44 > 2/3$) et il est généralisable à la population parente (test t de Student significatif au seuil unilatéral de .0005)

Calculatrice**(6 points)**

Pour l'effet d'un certain facteur « B », on a calculé les statistiques suivantes sur deux groupes indépendants (g1 et g2).

	n_g	moy	var
B / g1	7	139	477
B / g2	9	- 97	138

A partir de ces valeurs, calculez (précision à 2 décimales) :

$$1/ V_{\text{inter}} = \frac{\sum f_g (m_g - m)^2}{16} = \frac{7 \times (139 - 6.25)^2 + 9 \times (-97 - 6.25)^2}{16}$$

$$= \frac{(7 \times 17622.56) + (9 \times 10660.56)}{16} = \frac{12337.2 + 95945.04}{16} = \underline{\underline{13706.44}} \quad (2 \text{ pts})$$

$$2/ V_{\text{intra}} = \frac{\sum f_g (V_g)}{16} = \frac{(7 \times 477) + (9 \times 138)}{16} = \underline{\underline{286.31}} \quad (2 \text{ pts})$$

$$3/ \text{éta}^2 = 13706.44 \div (13706.44 + 286.31) = 13706.44 \div 7053.95 = \underline{\underline{.98}} \quad (98\%) \quad (2 \text{ pts})$$

Analyses multivariées - Dossier « Lecture 2 »**(16,5 points)**

On s'intéresse à la liaison entre 5 variables numériques (Annexe 2) recueillies avant ou pendant l'expérience décrite précédemment. Il s'agit de :

1. TLEC = temps de lecture moyen par phrase (en millisecondes)
2. COMP = qualité de la compréhension estimée par le nombre total de bonnes réponses à des questions portant sur certaines phrases présentées (score maximum = 45).
3. EMPA = empan de lecture (indicateur de la capacité de mémoire de travail), évalué à partir d'une épreuve administrée avant le début de l'expérience (échelle de 0 à 6).

A ces trois variables on a ajouté deux variables correspondant aux effets individuels de 2 facteurs sur les Temps de Décision Lexicale ou TDL (en millisecondes)

4. PHON = effet de la phonologie calculé dans le sens p1(contraste) – p2 (sans contraste). S'agissant de temps, une valeur positive (p1 > p2) indique un effet inhibiteur du contraste phonologique et une valeur négative (p1 < p2) indique un effet facilitateur.
5. PNOM = effet de la position du nom auquel il faut attacher la relative, calculé dans le sens n1 (éloigné) – n2 (proche). Une valeur positive (n1 > n2) indique une « préférence » pour l'attachement à n2, alors qu'une valeur négative (n1 < n2) indique une « préférence » pour l'attachement à n1.

A. Analyse bivariate ($\Sigma = 9,5 \text{ pts}$)**Tableau 1 : matrice des corrélations** (d'après le logiciel DS3Win)

	TLEC (X1)	COMP (X2)	EMPA (X3)	PHON (X4)	PNOM (X5)
TLEC (X1)	1.00				
COMP (X2)	-.40	1.00			
EMPA (X3)	-.64	.55	1.00		
PHON (X4)	.67	-.34	-.40	1.00	
PNOM (X5)	.71	-.55	-.97	.44	1.00

1/ Interprétez la valeur $r_{12} = -.40$ de la matrice (corrélation X1 X2). **(2 pts)**

Dans l'échantillon, il existe une corrélation négative importante (ou presque = .40) entre TLEC et COMP : plus ces sujets lisent les phrases lentement, mieux ils les comprennent.

2/ Un premier chercheur (A) décide d'inférer sur la liaison entre TLEC et COMP. Il trouve $t_{\text{obs}} = 2.1499$; $p_{\text{obs}} = .021$ (seuil unilatéral). Formulez sa conclusion inférentielle. **(1,5 pt)**

De façon générale (dans la population parente), plus on lit lentement des phrases, mieux on les comprend (test T significatif au seuil de .025 unilatéral)

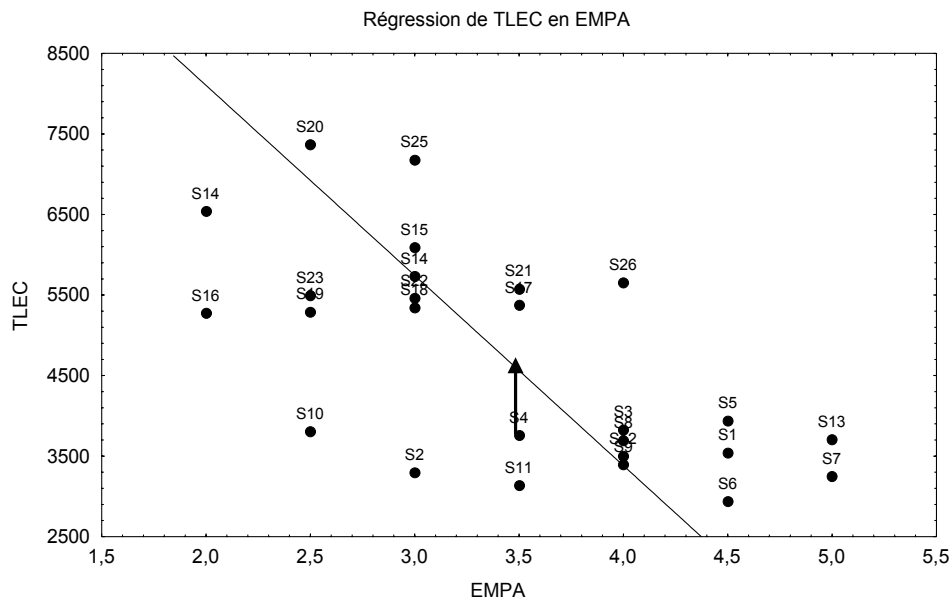
3/ Avant d'inférer, un second chercheur (B) a préféré calculer $r_{12,3}$. Pourquoi ?

Comme X1 et X2 sont bien corrélées avec X3 (-.64 et .55) il se peut qu'une part de leur liaison soit due à leur co-variation (liaison) avec X3 ; il est donc prudent de calculer r_{12} en maintenant X3 constante. (1,5 pt)

4/ Il s'avère que $r_{12,3} = -.07$; $p_{obs} > .10$. Est-on en droit d'affirmer qu'il n'existe pas de liaison $r_{12,3}$ dans la population parente ?

Non : le test n'étant pas significatif ($p_{obs} > .05$), on n'est pas en droit d'affirmer quoi que ce soit (constat d'ignorance). (1,5 pt ; 0 si « non » sans justification)

5/ On cherche à prédire le temps de lecture TLEC, à partir du score d'empan EMPA.



a. Dans le graphique, représentez e_{i4} (l'écart résiduel pour s4) (1 pt)

b. Calculez l'équation de la droite assurant la meilleure prédiction possible de TLEC à partir de EMPA.

$$TLEC_{prédit} = -981,15 EMPA + 8075,68 \quad (2 \text{ pts})$$

B. Régression multiple ($\Sigma = 7$ pts)

Afin de prédire la variable TLEC à partir de COMP, EMPA, PHON et PNOM, on procède à une analyse de régression multiple dont les résultats figurent ci-dessous.

Données : RTDfev.STA 10v * 26					Corrélations (rtdfev.sta)					
VALEURS NUMERIQUE	1	2	3	4	Suite...	COMP	EMPA	PHON	PNOM	TLEC
	TLEC	COM	EMP	PHON	P	COMP	EMPA	PHON	PNOM	TLEC
S1	3540	42	4,5	-37	-	1,0000	,5531	-,3439	-,5516	-,4018
S2	3297	39	3,0	29	-	,55307	1,0000	-,3979	-,9683	-,6447
S3	3826	40	4			-,3439	-,3979	1,0000	4366	6690
S4	3758	39	3							
S5	3941	43	4							
S6	2936	41	4							
S7	3247	42	5							
S8	3694	42	4							
S9	3392	41	4							
S10	3808	43	2							
S11	3137	42	3							
S12	3499	43	4							
S13	3707	42	5,0							
S14	5733	42	3,0							
S15	6091	41	3,0							
S16	5275	35	2,0							
S17	5373	37	3,5	1						
S18	5344	39	3,0	1						
S19	5289	38	2,5	1						
S20	7368	38	2,5	4						
S21	5574	41	3,5	2						

Synthèse Régression de la Var. Dépendante : TLEC						
REGRESS. MULTIPLE	R= ,82538239 R²= ,68125608 R² Ajusté= ,62054296 F(4,21)=11,221 p<,00005 Err-Type de l'Estim.: 810,					
N=26	BETA	Err-Type de BETA	B	Err-Type de B	t(21)	niveau p
OrdOrig.			1027,8	4334,5	,2371	,8149
COMP	,0357	,1500	22,4	94,0	,2383	,8140
EMPA	,5075	,4991	772,3	759,5	1,0169	,3208
PHON	,4325	,1393	4,8	1,6	3,1039	,0054
PNOM	1,0326	,5064	7,7	3,8	2,0392	,0542

No d'Obs.	Valeur Observée	Valeur Prévues	Résidus	Standard Val. Prév.
S1	3540	3074,2	465,79	-1,496
S2	3297	4605,0	-1308,	-,086
S3	3826	4207,0	-381,0	-,453
S4	3758	4208,6	-450,6	-,451
S5	3941	3528,5	412,55	-1,078
S6	2936	3351,3	-415,3	-1,241
S7	3247	3121,8	125,24	-1,453

1/ Reconstituez l'équation de la régression. (1 pt)

$$TLEC_{prédit} = 22.4 COMP + 772.3 EMPA + 4.8 PHON + 7.7 PNOM + 1027.8$$

2/ Commentez la qualité de la prédiction (l'ajustement) de TLEC à partir de l'ensemble des autres variables.

L'ensemble des prédicteurs prend en compte (rend compte de, prédit, éventuellement explique) 68% de la variance de TLEC, ce qui représente une part importante ($R^2 = .6813 \geq .16$). (2 pts)

3/ Interprétez la valeur « ,5075 » dans la colonne BETA.

Quand le score d'empan augmente d'un écart-type et que les autres prédicteurs restent constants, les temps de lecture augmentent de 0.5075 écart-type. (2 pts)

4/ Descriptivement parlant, concluez sur le meilleur prédicteur de la variable dépendante TLEC (justifiez).

Le meilleur prédicteur des temps de lecture est la variable PNOM qui est associée au beta le plus élevé (1.0326) (1 pt)

5/ Peut-on affirmer que, dans la population parente, cette variable permet de prédire TLEC ?

- Non, car le test T n'est pas significatif ($p = .054 > .05$) (1 pt)
- à la rigueur : Oui car le seuil observé ($p = .054$) est très proche de .05 (0.5 pt)

Analyses géométriques – Dossier « Lecture 2 »**(12 points)**

Afin de résumer les données issues des 5 variables initiales (TLEC, COMP, EMPA, PHON, PNOM), on procède à l'analyse géométrique de leur liaison.

1/ Précisez le nom de la procédure adéquate (en justifiant). **(2 pts)**

A.C.P. (Analyse en Composantes Principales) : représentation géométrique de la liaison entre variables numériques **(1 + 1)**

2/ A partir du tableau 2, explicitez le(s) critère(s) justifiant le nombre d'axes qui seront retenus.

Tableau 2 : Valeurs propres

! NUM !	! VAL PROPRE !	! POURC. !	! CUMUL !	! VARIAT. !	! HISTOGRAMME DES VALEURS PROPRES
! 1 !	! 3.27488 !	! 65.498 !	! 65.498 !	! ***** !	! ***** !
! 2 !	! .82831 !	! 16.566 !	! 82.064 !	! 48.932 !	! ***** !
! 3 !	! .61571 !	! 12.314 !	! 94.378 !	! 4.252 !	! ***** !
! 4 !	! .23546 !	! 4.709 !	! 99.087 !	! 7.605 !	! ***** !
! 5 !	! .04565 !	! .913 !	! 100.000 !	! 3.796 !	! ***** !

réponse : on retient 1 seul axe, ($\lambda_1 = 3.27488 > 1$) : il prend en compte une part acceptable (65.498%) de la variance du nuage. **(2 pts)**

3/ Lecture des aides à l'interprétation.

Tableau 3 : Aide à l'interprétation (espace des variables)

! J1 !	! QLT POID !	! INR !	! 1#F !	! COR !	! CTR !	! 2#F !	! COR !	! CTR !	! 3#F !	! COR !	! CTR !	! 4#F !	! COR !	! CTR !	! 5#F !	! COR !	! CTR !	
1! TLEC !	!1000 !	1	200!	-854	730	223!	317	100	121!	-164	27	44!	-378	143	606!	-17	0	6!
2! COMP !	!1000 !	1	200!	676	457	140!	289	83	101!	-674	455	738!	69	5	20!	7	0	1!
3! EMPA !	!1000 !	1	200!	883	781	238!	338	115	138!	257	66	107!	-138	19	81!	141	20	435!
4! PHON !	!1000 !	1	200!	-669	448	137!	683	467	563!	157	25	40!	247	61	259!	-6	0	1!
5! PNOM !	!1000 !	1	200!	-927	860	263!	-252	63	77!	-209	44	71!	89	8	34!	159	25	557!
			1000!			1000!			1000!			1000!			1000!			1000!

a. D'après l'indice pertinent du tableau 3, quelle est la variable la mieux corrélée avec l'axe 1 ?

C'est PNOM : sa corrélation avec l'axe 1 est la plus élevée ($1\#F = -927 \Rightarrow r = -.927$) **(2 pts)**

b. La matrice du tableau 1 révélait une très forte corrélation négative entre EMPA et PNOM. Qu'est-ce qui confirme cette opposition dans le tableau 3 ?
Le fait que ces variables soient en opposition (aient des coordonnées opposées) sur 4 axes (1, 2, 3 et 4) **(1 pt)**

4/ Interprétation des axes.

Tableau 4 : Aide à l'interprétation (nuage des individus)

!	J1	!	QLT	POID	INR!	1#F	COR	CTR!	2#F	COR	CTR!	3#F	COR	CTR!	4#F	COR	CTR!	5#F	COR	CTR!
1!	S1	!	1000	38	45!	2394	970	67!	235	9	3!	170	5	2!	-217	8	8!	-212	8	38!
2!	S2	!	1000	38	16!	-98	5	0!	-1044	519	51!	495	117	15!	865	356	122!	-94	4	7!
3!	S3	!	1000	38	12!	1021	667	12!	-500	160	12!	464	138	13!	-118	9	2!	198	25	33!
4!	S4	!	1000	38	14!	440	106	2!	-1149	723	61!	553	168	19!	-1	0	0!	82	4	6!
5!	S5	!	1000	38	43!	2262	922	60!	408	30	8!	-398	29	10!	-330	20	18!	22	0	0!
6!	S6	!	1000	38	45!	2335	930	64!	-242	10	3!	591	60	22!	71	1	1!	16	0	0!
7!	S7	!	1000	38	70!	2884	909	98!	693	52	22!	570	36	20!	-138	2	3!	-79	1	5!
8!	S8	!	1000	38	24!	1642	882	32!	405	54	8!	-26	0	0!	231	17	9!	-379	47	121!
9!	S9	!	1000	38	16!	1380	900	22!	-296	42	4!	192	17	2!	268	34	12!	119	7	12!
10!	S10	!	1000	38	34!	80	1	0!	-963	208	43!	-1660	619	172!	866	168	122!	-120	3	12!
11!	S11	!	1000	38	22!	1199	508	17!	-812	233	31!	-585	121	21!	607	130	60!	157	9	21!
12!	S12	!	1000	38	28!	1724	827	35!	365	37	6!	-521	75	17!	460	59	34!	-73	1	4!
13!	S13	!	1000	38	58!	2552	867	76!	828	91	32!	452	27	13!	-319	14	17!	71	1	4!
14!	S14	!	1000	38	14!	-560	169	4!	63	2	0!	-1191	766	89!	-337	61	19!	-58	2	3!
15!	S15	!	1000	38	15!	-1040	565	13!	386	78	7!	-726	275	33!	-376	74	23!	-132	9	15!
16!	S16	!	1000	38	93!	-2698	600	86!	-1845	28	158!	1171	113	86!	-50	0	0!	-283	7	67!
17!	S17	!	1000	38	33!	-1396	456	23!	426	42	8!	1461	500	133!	-37	0	0!	87	2	6!
18!	S18	!	1000	38	12!	-1224	964	18!	-93	6	0!	216	30	3!	3	0	0!	-30	1	1!
19!	S19	!	1000	38	29!	-1561	638	29!	394	41	7!	878	202	48!	235	14	9!	634	105	339!
20!	S20	!	1000	38	170!	-3917	695	180!	2413	264	270!	633	18	25!	654	19	70!	-293	4	72!
21!	S21	!	1000	38	26!	-980	289	11!	1463	644	99!	-161	8	2!	400	48	26!	188	11	30!
22!	S22	!	1000	38	17!	-863	341	9!	457	96	10!	-1074	528	72!	243	27	10!	136	9	16!
23!	S23	!	1000	38	42!	-1682	513	33!	-1419	366	94!	560	57	20!	-534	52	47!	-263	13	58!
24!	S24	!	1000	38	76!	-2664	714	83!	-938	89	41!	-1328	177	110!	-314	10	16!	307	9	79!
25!	S25	!	1000	38	35!	-1469	479	25!	695	107	22!	-904	181	51!	-1008	226	166!	-173	7	25!
26!	S26	!	1000	38	11!	239	41	1!	71	4	0!	167	20	2!	-1123	914	206!	169	21	24!
!	!	!	1000!	!	!	1000!	!	!	1000!	!	!	1000!	!	!	1000!	!	!	1000!	!	!

a. A partir des tableaux 3 puis 4, construisez les tableaux dégagant les plus fortes contributions à la variance de la 1^{ère} variable factorielle. (2 pts : 1 pt par tableau)

CTR > 200

1 # F -	1 # F +
PNOM (263)	EMPA (238)
TLEC (223)	

CTR > 38.5

1 # F -	1 # F +
s20 (180)	s7 (98)
s16 (86)	s13 (76)
s24 (83)	s1 (67)
	s6 (64)
	s5 (60)

b. Interprétez l'axe 1 (à partir des contributions dégagées précédemment et des données en Annexe 2).

L'axe 1 oppose

- les lecteurs rapides ayant un fort empan (mnésique) et préférant attacher la relative à n1
- aux lecteurs lents ayant un faible empan et préférant attacher la relative à n2.

(3 pts : 1,5 + 1,5)

Tableaux de contingence	(16 points)
--------------------------------	--------------------

Dans une république bananière, un organisme de sondage procède à une grande enquête d'opinion à propos de Don Biscoto, jeune aristocrate au physique avantageux nourrissant des ambitions politiques. A un mois d'intervalle, on interroge un même échantillon aléatoire de 6000 citoyens majeurs sur son jugement concernant Don Biscoto : en avril (un mois avant sa prise de pouvoir) et en juin (un mois après le début de son action politique).

A. LIAISON OPINION-ÉTAT CIVIL ($\Sigma = 8,5$ pts)

Le tableau A suivant donne l'état de l'opinion en avril selon l'état civil (conjonction du sexe et de la tranche d'âge).

On a groupé les opinions en 5 catégories :

- | | |
|--------------------------------|--------------------------------|
| - très positive = OHO (oh oui) | - négative = NON |
| - positive = OUI | - très négative = AHN (ah non) |
| - mitigée = BOF | |

L'état civil se décompose comme suit :

- | | |
|------------------------------|------------------------------|
| - hommes < 50 ans = HOJ | - femmes < 50 ans = FEJ |
| - hommes \geq 50 ans = HOV | - femmes \geq 50 ans = FEV |

Tableau A

	OHO	OUI	BOF	NON	AHN	TOT	PCT
HOJ	145	218	275	377	435	1450	24.17
HOV	176	243	459	283	189	1350	22.50
FEJ	540	435	135	210	180	1500	25.00
FEV	306	442	493	289	170	1700	28.33
TOT	1167	1338	1362	1159	974	6000	
PCT	19.45	22.30	22.70	19.32	16.23		100

1/ Étant donné le charme personnel de Don Biscoto, on soupçonne que la fréquence des réponses enthousiastes (OHO) dépend du sexe. Donnez les détails du calcul de l'effet permettant de s'en assurer.

$$d_{\text{obs}} = f_{\text{OHO, hommes}} - f_{\text{OHO, femmes}} = \frac{145 + 176}{1450 + 1350} - \frac{540 + 306}{1500 + 1700} = \frac{321}{2800} - \frac{846}{3200} = -\frac{1497}{3200} (15\%) \quad (1,5 \text{ pt})$$

2/ On veut vérifier l'existence éventuelle d'une liaison globale entre état civil et opinion.

Tableau B

	AHO	OUI	BOF	NON	AHN	
HOJ	10.00	15.03	18.97	26.00	30.00	100
HOA	13.04	18.00	34.00	20.96	14.00	100
FEJ	36.00	29.00	9.00	14.00	12.00	100
FEA	18.00	26.00	29.00	17.00	10.00	100
MOY	19.45	22.30	22.70	19.32	16.23	100

a. Dites ce que représente le tableau B et commentez-le succinctement.

- Il s'agit des profils-lignes (des tranches d'âge) (1 pt)
- Commentaire (2 pts : 0,5 par ligne pertinente)
 - les jeunes hommes sont majoritairement défavorables (26+30=59%)
 - les hommes âgés n'ont pas d'opinions très tranchées (fav : 21%, bof : 34%, défav : 35%)
 - les jeunes femmes sont très majoritairement favorables (36+29=65%)
 - les femmes âgés sont plutôt favorables (44%) mais beaucoup s'abstiennent (29%)

b. A partir du tableau C (page suivante) indiquez la valeur du Φ^2 et formulez une conclusion descriptive concernant la liaison globale entre l'état civil et l'opinion sur Don Biscoto.

Tableau C : Contributions absolues au Φ^2 ($C_{t_{jk}}$)

	AHO	OUI	BOF	NON	AHN	TOT
HOJ	0.0111	0.0057	0.0015	0.0056	0.0282	0.0521
HOA	0.0048	0.0019	0.0127	0.0003	0.0007	0.0203
FEJ	0.0352	0.0050	0.0207	0.0037	0.0028	0.0673
FEA	0.0003	0.0017	0.0050	0.0008	0.0068	0.0146
TOT	0.0514	0.0144	0.0398	0.0103	0.0384	0.1543

Conclusion descriptive

$$\Phi^2 = 0.1543 ; \Phi^2_{max} = 3 ; V^2 = 0.1543 \div 3 = .051$$

Dans cet échantillon de 6000 personnes, il existe une liaison entre l'état civil et l'opinion sur Don Biscoto ($\Phi^2 = 0.1543$) ; cette liaison est d'importance moyenne (intermédiaire) ($.04 < V^2 = 0.51 < .16$).
(2 pts)

c. Donnez le nombre de degrés de liberté du tableau, calculez la statistique de test appropriée et formulez une conclusion inférentielle (sachant que $p_{obs} < .00001$).

$$ddl = (L-1) \times (C-1) = 3 \times 4 = 12 \quad (1 \text{ pt})$$

$$\chi^2 = n \times \Phi^2 = 0.1543 \times 6000 = 925.8 \quad (1 \text{ pt})$$

Dans la population parente (de l'ensemble des citoyens majeurs de cette république), il existe une liaison entre l'état civil et l'opinion sur Don Biscoto (test du χ^2 significatif ; $p_{obs} < .00001$)

B. ÉVOLUTION DE L'OPINION ($\Sigma = 7,5$ pts)

On veut savoir si l'opinion initiale des femmes jeunes (FEJ) évolue après avoir vu Don Biscoto à l'œuvre. Le tableau D présente cette évolution entre avril et juin.

« + » = OHO + OUI « - » = NON + AHN

Tableau D : opinion en avril et en juin pour la catégorie FEJ

		Juin		TOT	PCT
		+	-		
Avril	+	215	760	975	71
	-	110	280	390	29
TOT		325	1040	1365	
PCT		24	76		100

1/ Quantifiez et commentez l'évolution de fréquence des réponses favorables entre avril et juin.

$$\bullet d_{obs} = f_{+,avril} - f_{+,juin} = 71\% - 24\% = 47\% \quad (1,5 \text{ pts})$$

• La fréquence des opinions favorables à Don Biscoto chute de 47% entre avril et juin. (1 pt)

Tableau E : profils-lignes

		Juin		TOT
		+	-	
Avril	+	22	78	100
	-	28	72	100
MOY		24	76	100

2/ Commentez la valeur « 78 » du tableau E ci-contre
78% des jeunes femmes favorables à Don Biscoto en avril changent d'avis en juin (après l'avoir vu à l'œuvre)
(2 pts)

3/ On veut inférer sur l'évolution de l'opinion des jeunes femmes entre avril et juin.

a. Précisez le nom du test adéquat et son nombre de degrés de liberté.

$$\chi^2 \text{ de McNemar} \quad ddl = 1 \quad (1 \text{ pt} : 0,5 + 0,5)$$

b. Le calcul de cette statistique de test donne une valeur de 484.14 pour un seuil $p_{obs} < .0005$. Compte tenu de ces éléments, rédigez une conclusion inférentielle.

Chez l'ensemble des jeunes femmes de cette république bananière, la fréquence des opinions favorables à Don Biscoto diminue après sa prise de pouvoir (χ^2 de McNemar significatif au seuil de .0005).

(2 pts)

Annexe 1 : données « Lecture 1 »
(Les valeurs des PDP ont été **arrondies à l'unité**).

	n1f1p1	n1f1p2	n1f2p1	n1f2p2	n2f1p1	n2f1p2	n2f2p1	n2f2p2	PDP 1	PDP 2	PDP 3	PDP 4
s1	2135	1840	2104	2376	2582	2465	2129	2415	-37	-284	-1	205
s2	1522	1654	2036	1137	1252	1651	1533	1785	29	32	-103	-266
s3	830	746	678	845	880	912	820	827	-31	-85	50	26
s4	767	851	862	953	854	863	880	907	-53	-18	-67	-47
s5	1434	1493	1395	1493	1430	1764	1938	1596	-37	-228	-75	-197
s6	1388	1821	1366	1054	1736	1280	1514	2046	-50	-237	61	12
11 s7	926	781	803	913	875	1029	1505	1406	-5	-348	-254	-5
s8	1455	1354	1542	1284	1315	1556	1834	1825	32	-224	-201	-70
s9	636	774	729	709	805	854	878	759	-12	-112	-2	-94
s10	940	1035	1176	1171	882	793	1067	1147	-21	108	-228	-3
s11	1057	981	1107	984	917	1010	994	1290	-47	-21	-103	-9
s12	1277	1236	1581	1242	1704	1903	1145	1224	26	-160	232	-79
s13	990	833	840	938	897	1039	1500	1389	7	-306	-227	8
s14	684	685	750	699	629	592	698	644	35	64	-50	18
s15	917	759	727	647	642	684	831	658	92	59	35	58
s16	1003	1084	1098	1097	811	918	1003	685	33	216	-17	-94
s17	1066	1095	1282	1105	1225	1124	1281	809	180	27	8	36
s18	1471	909	901	826	744	1130	1036	877	102	80	154	88
12 s19	2255	2301	2867	2940	2763	2384	2549	2070	185	149	-181	167
s20	2045	1159	2533	1979	1666	1782	2104	1501	482	166	-366	385
s21	1246	892	1054	1104	952	916	1440	821	240	42	-103	195
s22	1187	1194	1517	1154	986	1147	1379	1102	118	110	-160	-84
s23	967	1037	1104	1297	824	942	1236	920	-16	121	-197	-94
s24	1256	1495	3065	2257	1599	1526	1520	2022	35	352	-747	-83
s25	978	764	960	888	816	911	898	705	96	65	5	60
s26	839	881	1134	874	872	895	986	1236	-14	-65	-186	-33

Annexe 2 : extrait de la table du *t* de Student

α	.05	.01	.001
$\alpha/2$.025	.005	.0005
ddl			
1	12.706	63.657	636.619
2	4.303	9.925	31.599
3	3.182	5.841	12.924
20	2.086	2.845	3.850
21	2.080	2.831	3.819
22	2.074	2.819	3.792
23	2.069	2.807	3.768
24	2.064	2.797	3.745
25	2.060	2.787	3.725
26	2.056	2.779	3.707
27	2.052	2.771	3.690

Annexe 3 : données « Lecture 2 »**Valeurs brutes****Valeurs centrées réduites**

	TLEC	BREP	EMPA	PHON	PNOM	ER TLEC	ER BREP	ER EMPA	ER PHON	ER PNOM
s1	3540	42	4,5	-37	-284	-0,91	0,76	1,25	-0,08	-1,47
s2	3297	39	3,0	29	32	-1,10	-0,65	-0,52	0,23	0,30
s3	3826	40	4,0	-31	-85	-0,67	-0,18	0,66	-0,05	-0,36
s4	3758	39	3,5	-53	-18	-0,73	-0,65	0,07	-0,16	0,02
s5	3941	43	4,5	-37	-228	-0,58	1,23	1,25	-0,09	-1,16
s6	2936	41	4,5	-50	-237	-1,39	0,29	1,25	-0,15	-1,21
s7	3247	42	5,0	-5	-348	-1,14	0,76	1,84	0,07	-1,83
s8	3694	42	4,0	32	-224	-0,78	0,76	0,66	0,25	-1,14
s9	3392	41	4,0	-12	-112	-1,03	0,29	0,66	0,04	-0,51
s10	3808	43	2,5	-21	108	-0,69	1,23	-1,11	-0,01	0,72
s11	3137	42	3,5	-47	-20	-1,23	0,76	0,07	-0,14	0,00
s12	3499	43	4,0	26	-160	-0,94	1,23	0,66	0,22	-0,78
s13	3707	42	5,0	7	-306	-0,77	0,76	1,84	0,13	-1,83
s14	5733	42	3,0	35	64	0,87	0,76	-0,52	0,27	0,47
s15	6091	41	3,0	92	59	1,16	0,29	-0,52	0,54	0,45
s16	5275	35	2,0	33	216	0,50	-2,54	-1,70	0,25	1,33
s17	5373	37	3,5	180	27	0,58	-1,60	0,07	0,97	0,27
s18	5344	39	3,0	102	80	0,55	-0,65	-0,52	0,59	0,56
s19	5289	38	2,5	185	149	0,51	-1,13	-1,11	0,99	0,95
s20	7368	38	2,5	482	166	2,19	-1,13	-1,11	2,43	1,04
s21	5574	41	3,5	240	42	0,74	0,29	0,07	1,26	0,35
s22	5460	42	3,0	118	110	0,65	0,76	-0,52	0,67	0,73
s23	5497	37	2,5	-16	121	0,68	-1,60	-1,11	0,02	0,79
s24	6541	40	2,0	35	352	1,52	-0,18	-1,70	0,26	2,08
s25	7178	41	3,0	96	65	2,04	0,29	-0,52	0,56	0,48
s26	5650	40	4,0	-14	-65	0,80	-0,18	0,66	0,03	-0,25