

## ASDP4 (UE ADN)

Durée de l'épreuve : 1h30

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

Les exercices (encadrés) sont indépendants. Le barème (sur 32) est donné à titre indicatif. La note finale sera donnée sur 16.

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

### Régression multiple : Dossier TMT (10 points)

Le test des points à relier (Trail Making Test : *TMT*) est un test servant à observer la capacité attentionnelle du sujet. Il s'agit d'un test simple mais qui demande l'inhibition d'une routine de conduite.

On demande au sujet de relier par un trait 13 chiffres et 13 lettres (présentés éparpillés sur une feuille) dans l'ordre de la série des chiffres et dans l'ordre alphabétique ; ce qui nécessite d'intercaler une lettre avec un chiffre.

On note le temps de réalisation du test (en secondes). Dix sujets ont participé à ce test. Ils ont également été sollicités pour la passation du test de la Tour de Londres : *TOURL* (noté sur 12) et pour une épreuve de résolution de problèmes : *PROB* (notée sur 20) pour laquelle on dispose également du temps : *TPPROB* (en secondes) mis pour résoudre les problèmes.

**Question initiale :** *Quel est le meilleur prédicteur du temps réalisé au test TMT ?*

Le logiciel *Statistica* donne les sorties suivantes :

**Tableau 1 : Matrice des corrélations**

REGRESS. MULTIPLE	TMT	TOURL	PROB	TPPROB
TMT	1.00	-.76	-.33	.22
TOURL	-.76	1.00	.83	-.60
PROB	-.33	.83	1.00	-.65
TPPROB	.22	-.60	-.65	1.00

1) Commenter les corrélations entre les variables prédictrices et la variable à prédire :

**Tableau 2 : Synthèse de la régression multiple**

REGRESS. MULTIPLE	R= .93386925 R <sup>2</sup> = .87211178 R <sup>2</sup> Ajusté= .80816767 F(3,6)=13.639 p<.00435 Err-Type de l'Estim.: 7.0321					
N=10	BETA	Err-Type de BETA	B	Err-Type de B	t(6)	niveau p
OrdOrig.			131.4602	19.28111	6.81809	.000488
TOURL	-1.56094	.261228	-10.3864	1.73819	-5.97541	.000986
PROB	.85057	.274780	3.3895	1.09497	3.09547	.021239
TPPROB	-.16159	.193141	-.0526	.06281	-.83665	.434840

2) Indiquer ce que représentent les coefficients *R* et *R*<sup>2</sup> et interpréter leurs valeurs :

3) Indiquer ci-après l'équation de la régression multiple :

4) Le sujet 1 a obtenu 11.00 au test "Tour de Londres" (*TOURL*), 17.00 pour la résolution de problèmes (*PROB*) et a exécuté les problèmes en 105.00 secondes (*TPPROB*).

a) Calculer la valeur prévue par la régression multiple pour le test *TMT* :

b) Quelle est la valeur du résidu pour ce sujet 1, sachant qu'il a obtenu 69.00 au test *TMT* ?  
 Indiquer le calcul qui a permis de trouver ce résultat.

5) Interpréter pour la variable "Tour de Londres" (*TOURL*) :

a) la valeur "*Bêta*" :

b) le résultat du test *T* de Student :

c) formuler une conclusion inférentielle pour cette variable prédictrice :

6) En se fondant sur les résultats contenus dans le tableau 2 (page précédente), répondre à la question initiale (justifier la réponse) :

### ACP Standard : Dossier Psychométrie (11 points)

*Source* : Rouanet, H, & Le Roux, B. (1993). *L'analyse des données multidimensionnelles*. Paris : Dunod.

Les données "Psychométrie" présentées ci-après, sont inspirées d'une recherche sur la pédagogie des mathématiques. Pour 20 sujets, on a les notes aux cinq épreuves individuelles suivantes : Combinatoire (*COMB*), Probabilités (*PROB*), Logique (*LOG*), notées de 0 à 10, *QI* verbal (dont les notes s'échelonnent ici de 85 à 125), Mathématiques (*MATH*), notée de 0 à 20. On aimerait étudier la structure des relations entre ces variables ainsi que les profils des 20 individus. Pour cela, on a effectué une ACP standard. L'analyse effectuée avec le logiciel *Addad* conduit aux résultats suivants :

**Tableau 1 : Valeurs propres**

! NUM !	VAL PROPRE !	POURC. !	CUMUL !	VARIAT. !	! * !	HISTOGRAMME DES VALEURS PROPRES
! 1 !	2.35638 !	47.128 !	47.128 !	***** !	* !	***** !
! 2 !	1.12229 !	22.446 !	69.574 !	24.682 !	* !	***** !
! 3 !	.73187 !	14.637 !	84.211 !	7.808 !	* !	***** !
! 4 !	.53531 !	10.706 !	94.917 !	3.931 !	* !	***** !
! 5 !	.25414 !	5.083 !	100.000 !	5.623 !	* !	*** !

**Tableau 2 : Aides à l'interprétation pour le "nuage des variables"**

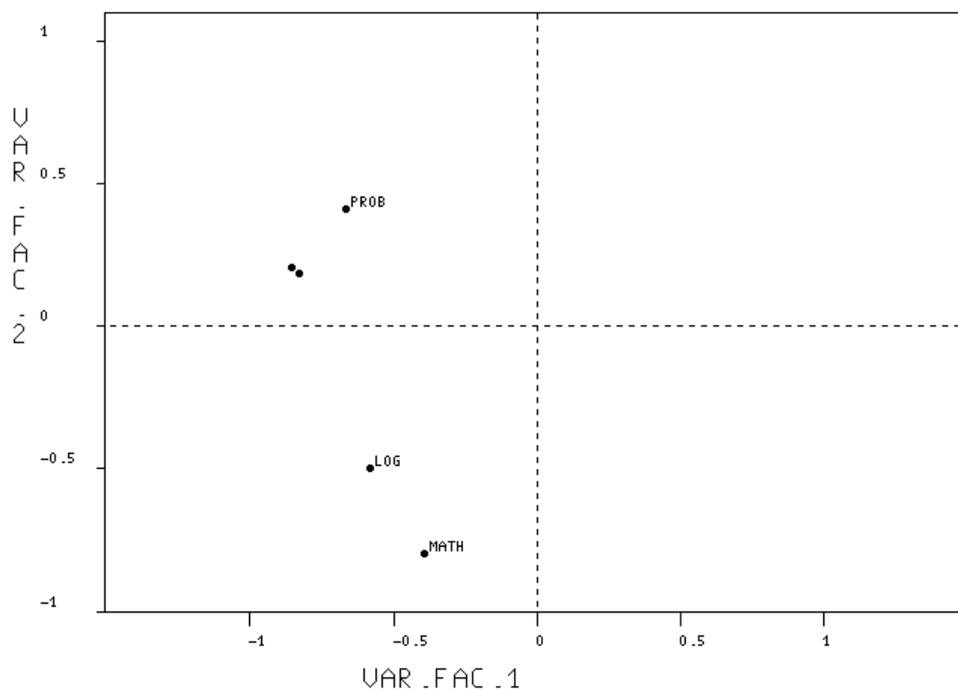
! J1 !	QLT	POID	INR !	1#F	COR	CTR !	2#F	COR	CTR !	3#F	COR	CTR !
1!COMB!	796	1	200!	-857	735	312!	204	42	37!	141	20	27!
2!PROB!	739	1	200!	-666	443	188!	412	170	151!	-356	127	173!
3!LOG !	917	1	200!	-582	338	144!	-496	246	219!	-577	332	454!
4!QI !	880	1	200!	-828	686	291!	183	33	30!	400	160	219!
5!MATH!	878	1	200!	-393	154	66!	-795	632	563!	304	92	126!
! !			1000!			1000!			1000!			1000!

1) A l'examen des tableaux 1 et 2 page précédente, on a décidé de retenir 3 axes factoriels pour cette analyse. Pourquoi ?

2) Quelle est la variable la moins bien représentée par l'ensemble de ces 3 axes ? Donner le nom et la valeur de l'indice correspondant, indiqués tableau 2 :

Ci-après le graphique factoriel pour le plan 1-2 (logiciel EyeLID2) :

Figure 1 : "Nuage des variables" ; Plan 1-2



3) A l'aide du tableau 2, reporter précisément les deux étiquettes manquantes sur le graphique ci-dessus.

4) On s'intéresse maintenant à la représentation géométrique des corrélations entre les variables pour **le plan 1-2** (indépendamment des axes factoriels). Compléter le graphique ci-dessus de manière à en permettre l'interprétation.

5) "Les variables COMB et QI présentent une corrélation forte et positive". Justifier cette affirmation :

6) On désire maintenant interpréter les 2 premiers axes factoriels.

a) Indiquer les variables qui contribuent au premier axe factoriel :

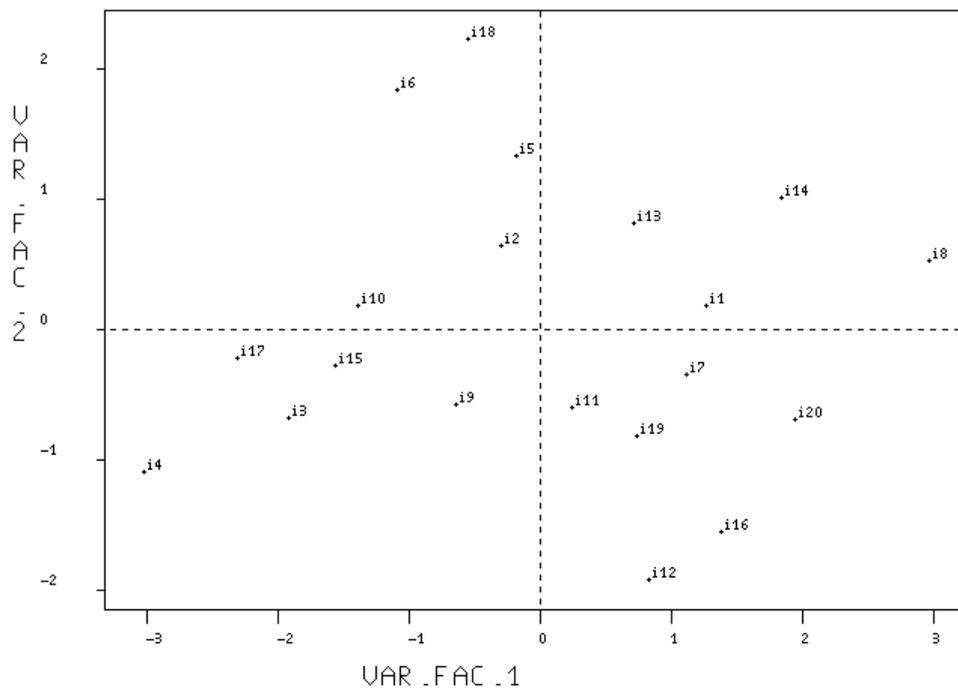
b) Interpréter ce premier axe factoriel :

c) Indiquer les variables qui contribuent au second axe factoriel :

d) Interpréter ce second axe factoriel :

Ci-après (figure 2), le nuage des individus :

**Figure 2 : Nuage des individus ; plan 1-2**



7) Compte-tenu de l'analyse qui a été faite de la première et de la deuxième variable factorielle, indiquer ce que suggère l'ACP concernant les profils des individus i17 (à gauche), i8 (à droite) et i18 (en haut).

8) Pour analyser plus finement les profils de groupes d'individus, quelle analyse pourrait être mise en œuvre après cette ACP ? Indiquer le nom complet de cette analyse, ainsi que son abréviation usuelle.

## Analyse d'un tableau de contingence : Dossier Environnement (11 points)

Source : d'après Lebart, L. (1993). Analyse exploratoire de fichiers d'enquêtes ; apports de l'analyse des correspondances. In : *Développements récents dans l'analyse de grands ensembles de données*. Eurostat, Luxembourg.

Les données suivantes ont été recueillies (cf. tableau 1) pour étudier la relation entre la Catégorie Socio-Professionnelle (CSP) et la principale source d'information sur les problèmes d'environnement : 1283 personnes classées selon leur CSP ont été interrogées. Le tableau suivant donne pour chaque CSP le nombre de personnes qui ont pour principale source d'information l'une des 6 sources suivantes : télévision, journaux, radio, livre, associations, et mairie. Les 7 CSP étudiées sont : agriculteur, cadre supérieur, cadre moyen, employé de bureau, ouvrier, retraité, chômeur.

**Question initiale** : existe-t-il une liaison entre les catégories socio-professionnelles et la principale source d'information ?

**Tableau 1 : Catégories socio-professionnelles / Sources d'information**

	Télé	Journal	Radio	Livre	Association	Mairie	Total
Agriculteur	26	18	9	5	4	6	68
Cadre supérieur	19	49	4	16	5	3	96
Cadre Moyen	44	87	4	39	14	3	191
Employé	83	87	13	24	5	1	213
Ouvrier	181	107	16	31	7	7	349
Retraité	167	95	29	15	7	7	320
Chômeur	27	9	4	2	2	2	46
<b>Total</b>	<b>547</b>	<b>452</b>	<b>79</b>	<b>132</b>	<b>44</b>	<b>29</b>	<b>1283</b>

1) Si l'on veut obtenir une représentation factorielle de ces données, quelle est la méthode appropriée ? Donner son nom complet et son abréviation usuelle.

2) On procède à cette analyse avec le logiciel *Addad*. On obtient le tableau des valeurs propres suivant :

**Tableau 2 : Valeurs propres**

```
-----
!NUM ! VAL PROPRE ! POURC. ! CUMUL !VARIAT.!! HISTOGRAMME DES VALEURS PROPRES
-----
! 1 ! .09151 ! 75.099! 75.099!*****!*!*****!*****!*****
! 2 ! .02193 ! 18.001! 93.100! 57.098!*!*****
! 3 ! .00642 ! 5.269! 98.369! 12.732!*!****
! 4 ! .00175 ! 1.437! 99.806! 3.832!*!*
! 5 ! .00024 ! .194!100.000! 1.243!*!
```

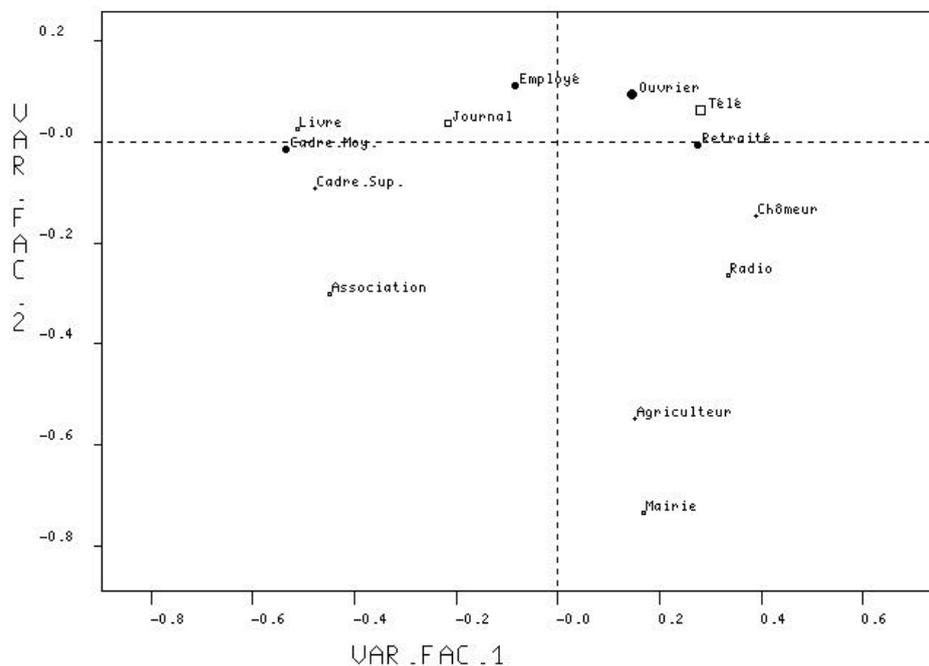
3) On trouve  $\phi^2 = 0.1218$ . Indiquer comment retrouver cette valeur à l'aide du tableau 2 :

4) On se demande si la liaison entre la CSP et la principale source d'information est importante dans cet échantillon. Calculer l'indice qui permet de répondre à cette question et conclure.

5) On a décidé ici de retenir les 2 premiers axes factoriels. Pourquoi ?

La représentation simultanée des lignes et des colonnes du tableau donne la représentation géométrique suivante :

**Figure 1 : Représentation simultanée des modalités lignes et des modalités colonnes**



On donne également le tableau des taux de liaison :

**Tableau 3 : Taux de liaison**

	Télé	Journal	Radio	Livre	Association	Mairie
Agriculteur	-0.10	-0.25	1.15	-0.29	0.72	2.90
Cadre supérieur	-0.54	0.45	-0.32	0.62	0.52	0.38
Cadre Moyen	-0.46	0.29	-0.66	0.98	1.14	-0.31
Employé	-0.09	0.16	-0.01	0.10	-0.32	-0.79
Ouvrier	0.22	-0.13	-0.26	-0.14	-0.42	-0.11
Retraité	0.22	-0.16	0.47	-0.54	-0.36	-0.03
Chômeur	0.38	-0.44	0.41	-0.58	0.27	0.92

6) Illustrer à partir des points "Agriculteur" et "Mairie" (en bas, à droite) que ce graphe factoriel est bien une représentation géométrique des taux de liaison (cf. figure 1 et tableau 3 ci-dessus) :

7) Inférence : dans la mesure où l'on peut considérer que ces 1283 personnes ont été échantillonnées au hasard dans la population française, on cherche à conclure sur l'existence d'une liaison entre les deux variables dans cette population parente.

a) Indiquer la statistique de test appropriée, et une formule de cette statistique qui permette de calculer rapidement sa valeur (156.27) à partir des résultats précédents :

b) Quel est le nombre de degrés de liberté (ddl) associé à ce test ?

c) Sachant que ce test est significatif au seuil .001, rédiger une conclusion inférentielle :