

**MEDI**

Durée de l'épreuve : 2 heures

Aucun document n'est autorisé. Seule la calculatrice (sans sa documentation) est autorisée.

**Attention** : les exercices (encadrés) sont indépendants. Le barème donné à titre indicatif est sur 60 ; la note finale sera donnée sur 20.

**La page 9 (Dossier Psychophysique et Dossier Papillons) et la page 10 (tables et formules) peuvent être détachées et conservées.** Indiquer les réponses exclusivement sur ce document.

**Dossier Psychophysique (12 points) : 1ère question**  
**Voir présentation des données page 9 : tableaux 1, 2 et 4**

Source : données fournies par Sébastien Georges, équipe enseignante Metus/Medi. UNIC, UPR 2191, CNRS-INAF, Gif-sur-Yvette.

Lors d'une expérience de psychophysique sur la perception visuelle du mouvement, un chercheur a également étudié l'influence de la connaissance préalable des hypothèses de travail sur les résultats. Pour ce faire, il a constitué deux groupes de sujets, choisis au hasard à l'Institut de Psychologie Henri Piéron de Boulogne-Billancourt (Université Paris 5). Le premier groupe ( $n_{g1} = 17$ ) est dit "naïf". Tous les sujets de ce groupe ne connaissent rien des objectifs de l'expérience avant la passation expérimentale. Le second groupe ( $n_{g2} = 8$ ), dit "averti", est parfaitement informé des buts de l'étude. On relève pour chaque sujet son pourcentage de Bonnes Réponses (% BR).

Les résultats obtenus sont les suivants :

$$m_{g1} = 67\% \quad v_{g1} = 17.059$$

$$m_{g2} = 88\% \quad v_{g2} = 6.000$$

**Question initiale** : on se demande si la connaissance préalable des hypothèses de travail peut avoir un effet sur les résultats.

**Critère sémantique** : une différence supérieure ou égale à 20 points de pourcentage sera jugée importante, une différence inférieure ou égale à 10 points de pourcentage, négligeable.

*Tous les résultats seront présentés arrondis à 3 décimales*

**A - Analyse descriptive**

1) Les sujets de cet échantillon sont-ils influencés par "la connaissance préalable des hypothèses de travail"? Calculer un premier indice élémentaire pour répondre à cette question. Indiquer le détail des calculs et commenter :

2) Élaborer une conclusion descriptive :

3) On trouve  $V_{inter} = 95.962$ . Donner la procédure de calcul de cette variance inter et retrouver ce résultat :

$$V_{inter} =$$

4) On trouve  $V_{intra} = 13.520$ . Donner la procédure de calcul de cette variance intra et retrouver ce résultat :

$$V_{intra} =$$

5) Calculer la variance totale de ce protocole :

$$V_{tot} =$$

6) Calculer  $\eta^2$  (indiquer la formule et le détails des calculs) :

7) Interpréter  $\eta^2$  :

### **B - Analyse inférentielle**

Le chercheur se demande s'il pourrait généraliser les résultats obtenus sur son échantillon à une population plus vaste non observée ici (la population parente).

1) Quelle est la population parente pour cet exemple ?

2) Quel test peut-on mettre en oeuvre ? Effectuer le calcul de ce test :

3) Indiquer le résultat du test :

4) Élaborer une conclusion inférentielle :

<p><b>Dossier Psychophysique (10 points) : 2ème question</b> <b>Voir présentation des données page 9 : tableaux 1, 2, 3 et 5</b></p>
--

Dans un second temps, le chercheur a souhaité examiner l'influence du degré de connaissance préalable des hypothèses de travail. Il a alors constitué un troisième groupe de sujets (pris au hasard) et l'a informé succinctement des objectifs de son étude. Les huit sujets de ce groupe (g3) en savent donc plus que les sujets du groupe naïf (g1) mais moins que ceux du groupe averti (g2). Tout comme pour g1 et g2, on recueille le pourcentage de Bonnes Réponses (%BR). Voir les données de ce groupe g3 page 9.

**Question initiale** : le degré de connaissance préalable des hypothèses de travail a-t-elle une influence sur les performances des sujets ?

**Tous les résultats seront présentés arrondis à 3 décimales.**

1) Calculer la moyenne de ce nouveau protocole (indiquer la procédure de calcul):

$$m_{g3} =$$

2) Comparer la moyenne de  $g_3$  aux moyennes obtenues pour  $g_1$  et  $g_2$ . Commenter.

3) Calculer la variance de ce nouveau protocole (indiquer la procédure de calcul) :

$$V_{g3} =$$

4) Calculer la variance inter pour ces 3 groupes ( $g_1$ ,  $g_2$  et  $g_3$ ). Donner la procédure de calcul :

$$V_{inter} =$$

5) Calculer  $\eta^2$  pour ces 3 groupes (indiquer la procédure de calcul) :

6) Élaborer une conclusion descriptive :

### **B - Analyse inférentielle**

1) Afin de généraliser les résultats obtenus sur ces 3 groupes de sujets, le chercheur veut mettre en oeuvre un test d'hypothèse. Quel est ce test ? (donner précisément son nom).

2) La valeur trouvée pour ce test est : 26.820. Indiquer la procédure qui a permis de trouver cette valeur ainsi que le nombre de degrés de liberté qui est associé à ce test :

3) Donner le résultat du test :

4) Élaborer une conclusion inférentielle :

## Dossier Euros (8 points)

On s'intéresse aux comportements induits par l'utilisation des "Euros" chez les personnes âgées. On a donc testé deux fois, dans la même journée, 80 personnes âgées (prises au hasard dans une maison de retraite de Paris). L'expérience de simulation est la suivante : on donne un budget à ces personnes puis on leur propose d'acquérir des articles alimentaires et des articles dits de confort. Le budget ainsi que les achats à effectuer sont exprimés en "Francs" lors du premier test, en "Euros" lors du second test.

Le tableau suivant regroupe le nombre d'achats simulés pour chaque test et pour chaque type d'article :

		<i>2nd Test (Euros)</i>		
		<i>Alimentaire</i>	<i>Confort</i>	
<i>1er Test</i>	<i>Alimentaire</i>	35	20	
<i>(Francs)</i>	<i>Confort</i>	17	8	
				80

**Question initiale** : on se demande si le fait d'utiliser une monnaie inhabituelle peut avoir une influence sur le comportement d'achat des personnes âgées, pour ce qui concerne les articles de confort.

**Critère sémantique** : on jugera un effet faible s'il est inférieur ou égal à 5 points de pourcentage, important s'il est supérieur ou égal à 10 points de pourcentage.

1) Calculer l'effet "Monnaie" sur la fréquence d'achat des articles de confort dans cet échantillon :

2) Élaborer une conclusion descriptive :

3) On désire généraliser les résultats à la population des personnes âgées vivant à Paris.

a) Indiquer le nom du test à utiliser, donner sa formule et développer le calcul :

b) Indiquer le résultat du test :

c) Élaborer une conclusion inférentielle et répondre à la question initiale :

**Dossier Papillons ( 14 points). Voir Tableaux 6 et 7 page 9.**

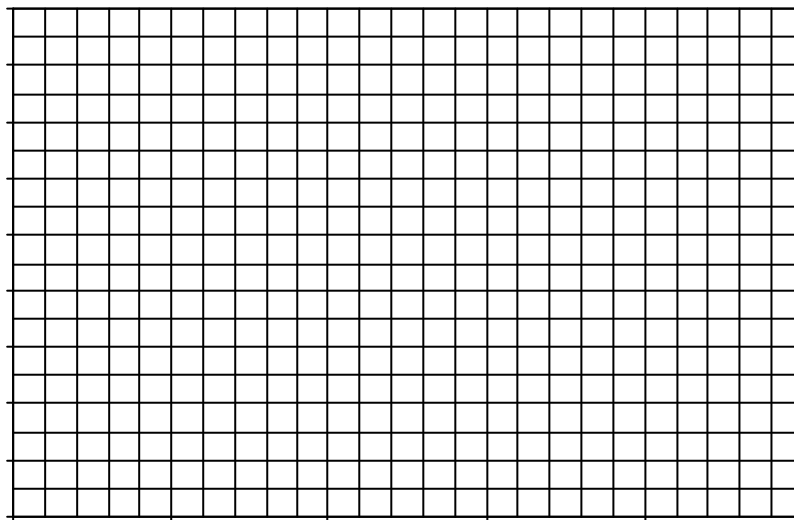
Source : données fournies par Sébastien Georges, équipe enseignante Metus/Medi. UNIC, UPR 2191, CNRS-INAF, Gif-sur-Yvette.

Une équipe de chercheurs en éthologie est partie en mission scientifique en Corrèze pour mener une étude préliminaire sur la relation existant entre l'envergure des ailes de papillon et la fréquence de battements des ailes. Pour ce faire, ils ont marqué temporairement 12 papillons pris au hasard dans la région et dont ils ont mesuré l'envergure des ailes (*Variable X*, exprimée en millimètres : mm). Ces 12 papillons ont ensuite été relâchés dans un espace clos, mais suffisamment grand, et ont été filmés. Le dépouillement des cassettes vidéos a permis de mesurer, chez chacun d'eux, la fréquence maximale de battements des ailes (*Variable Y* exprimée en Hertz : Hz). Le tableau 6 de la page 9 récapitule ces différentes mesures.

**A- PREMIÈRE PARTIE**

**Question initiale** : on aimerait savoir s'il existe une relation entre l'envergure des ailes des papillons et leur fréquence de battements d'ailes et de quel ordre est cette relation.

1) A l'aide du tableau 6 page 9, représenter le nuage de points ci-après (ne pas oublier les légendes) :



2) Commenter ce nuage bivarié :

3) Représenter sur le graphique ci-dessus, le centre de gravité du nuage et indiquer ses coordonnées ci-après :

4) A l'aide de la calculatrice, mettre en oeuvre la procédure appropriée pour pouvoir répondre à la question initiale. Présenter le nom de la statistique, et uniquement le résultat (arrondi à 4 décimales) :

5) Élaborer une conclusion descriptive et répondre à la question initiale :

## **B - DEUXIÈME PARTIE**

**Question initiale** : *peut-on prédire, sans trop de risques d'erreurs, la fréquence des battements d'ailes en fonction de l'envergure ?*

- 1) Déterminer l'équation de la droite de régression de "Fréquence" sur "Envergure" (présentation à 2 décimales) :
- 2) Interpréter la valeur du coefficient de régression (*pente : a*) :
  
- 3) Tracer la droite de régression de "Fréquence" sur "Envergure" sur le graphique de la page précédente.
  
- 4) Quelle est la part de variance de "Fréquence" prise en compte par la régression linéaire de "Fréquence" sur "Envergure" ? (indiquer le nom de l'indice, la procédure de calcul et le résultat).
  
- 5) Commenter la part de variance calculée en 4) et répondre à la question initiale.

## **C - TROISIÈME PARTIE**

**Question initiale** : *peut-on généraliser les résultats obtenus sur cet échantillon de papillons à une population plus vaste non observée (la population parente) ?*

- 1) Définir précisément la population parente :
  
- 2) Les chercheurs mettent en oeuvre le test approprié et trouvent : 3.47. Donner le nom de ce test, sa formule ainsi que le nombre de degrés de liberté qui lui est associé. :
  
- 3) Donner le résultat du test :
  
- 4) Élaborer une conclusion inférentielle et répondre à la question initiale :

## Dossier Allergie (6 points)

Marcel est un parisien allergique à la pollution. Pour se rendre sur son lieu de travail, Boulevard Saint-Germain, il utilise son vélo dans 65% des cas et le métro le reste du temps. Quand la température se situe en dessous de 20°, il n'y a pas de pics de pollution et Marcel n'a pas d'allergie. Par contre, quand il fait chaud, Paris enregistre de forts pics et les crises d'allergie de Marcel se déclenchent alors neuf fois sur dix quand il se déplace à vélo et seize fois sur vingt quand il prend le métro. Aujourd'hui il fait 28° et Marcel a eu une crise d'allergie en se rendant à son travail. *Quel moyen de transport a-t-il le plus probablement utilisé ?*

1) Utiliser la procédure appropriée pour répondre à la question initiale (développer et donner le résultat) :

2) En supposant que le nombre annuel de pics de pollution est de 100, construire le tableau de contingence qui répartit les crises de Marcel selon les deux variables : Moyen de transport et État de santé de Marcel.

			<i>Total</i>
			100

3) Vérifier à partir du tableau construit ci-dessus le résultat trouvé en 1) :

## Dossier Aéronautique (10 points)

*Source : données fournies par Stéphanie Joseph, Rapport de stage pour le DESS d'Ergonomie Informatique, Université Paris 5 (2001).*

Une compagnie aérienne française a voulu analyser les différents incidents survenus lors de vols réguliers depuis les 20 dernières années. Pour ce faire, elle a établi une base de données (*Base 1*) concernant les principales avaries constatées (impacts avec feu, impacts divers, amerrissage, feu, fumée, autres). Cette entreprise dispose également, à titre de référence, d'une autre base de données concernant les mêmes incidents, mais enregistrés par d'autres compagnies aériennes internationales (*Base 2*). Le tableau ci-après résume ces différentes informations :

**Tableau 1 : Effectifs observés**

	Impacts + Feu	Impacts Divers	Amerris- sage	Feu	Fumée	Autres	<i>Total</i>
Base 1	25	20	14	8	6	4	77
Base 2	77	65	14	48	11	16	231
<i>Total</i>	102	85	28	56	17	20	308

**Question initiale** : on aimerait savoir si les types d'incidents constatés par cette compagnie française (*Base 1*) diffèrent globalement de ceux la base de données de référence (*Base 2*).

**A - Analyse descriptive : analyse locale**

1) On se demande quel est le type d'incident le plus fréquemment constaté, indépendamment des bases de données étudiées (les deux Bases confondues).

a) Calculer le pourcentage pertinent pour répondre à cette question (le résultat sera présenté arrondi à l'entier le plus proche) :

b) Commenter ce résultat :

2) Ci-après, le tableau des taux de liaison (Tableau 2) :

**Tableau 2 : Taux de liaison**

	Impacts + Feu	Impacts Divers	Amerris- sage	Feu	Fumée	Autres
Base 1	-0.02	-0.06		-0.43	0.41	-0.20
Base 2	0.01	0.02	-0.33	0.14	-0.14	0.07

a) Calculer le taux de liaison manquant (*Base1/Amerrissage*) et indiquer ci-après le détail des calculs :

b) Élaborer ci-après le graphe des attractions :

c) Commenter ce graphe des attractions :

**B - Analyse descriptive : analyse globale**

1) On trouve  $\phi^2 = 0.046$ . Interpréter :

2) Élaborer une conclusion descriptive et répondre à la question initiale :

**C - Analyse inférentielle**

1) Analyse inférentielle : peut-on généraliser ces résultats à l'ensemble des compagnies françaises ?

a) Indiquer le nom du test que l'on pourrait mettre en oeuvre ainsi que le nombre de degrés de liberté qui lui est associé :

b) Est-il pertinent ici de généraliser ces résultats ? Pourquoi ? (justifier).



**Dossier Psychophysique : présentation des données**

**Tableau 1**

<i>g1 : "Naïf"</i>	
Sujets	%BR
s1	65
s2	59
s3	60
s4	66
s5	62
s6	68
s7	66
s8	75
s9	72
s10	72
s11	67
s12	70
s13	65
s14	65
s15	69
s16	68
s17	70

**Tableau 2**

<i>g2 : "Averti"</i>	
Sujets	%BR
s18	86
s19	84
s20	88
s21	87
s22	89
s23	92
s24	91
s25	87

**Tableau 3**

<i>g3 : "Averti(-)"</i>	
Sujets	%BR
s26	98
s27	87
s28	72
s29	65
s30	66
s31	68
s32	67
s33	69

**Tableau 4**

<u>2 groupes</u> ( <i>g1 et g2</i> )	% BR
<i>Moy</i>	73.720
<i>Var</i>	
<i>V<sub>inter</sub></i>	95.962
<i>V<sub>intra</sub></i>	13.520
<i>h<sup>2</sup></i>	

**Tableau 5**

<u>3 groupes</u> ( <i>g1, g2 et g3</i> )	% BR
<i>Moy</i>	73.788
<i>Var</i>	113.379
<i>V<sub>inter</sub></i>	
<i>V<sub>intra</sub></i>	40.667
<i>h<sup>2</sup></i>	

**Dossier Papillons : présentation des données**

**Tableau 6**

Papillons	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12
X : Envergure (mm)	10	40	15	18	19	46	45	30	32	35	13	25
Y : Fréquence (Hz)	8	4	7	7	6	7	3	5	5	4	8	6

**Tableau 7**

	Variable X	Variable Y
<i>Moy</i>	27.333	5.833
<i>Var</i>	127.355	2.231
<i>Ety</i>	12.002	1.572

- Extrait de la table des valeurs critiques de la variable  $\chi^2$  :

ddl	a	.05	.01	.001
1		3.84	6.63	10.83
2		5.99	9.21	13.82
3		7.81	11.34	16.27
4		9.49	13.28	18.47
5		11.07	15.09	20.52

- Extrait de la table des valeurs critiques de la variable  $T$  de Student :

ddl	a/2	.025	.005	.0005
	a	.05	.01	.001
10		2.228	3.169	4.587
11		2.201	3.106	4.437
12		2.179	3.055	4.318
13		2.160	3.012	4.221
14		2.145	2.977	4.140
15		2.131	2.947	4.073
16		2.120	2.921	4.015
17		2.110	2.898	3.965
18		2.101	2.878	3.922
19		2.093	2.861	3.883
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

Extrait de la table des valeurs critiques de la variable  $F$  de Snedecor :

ddl1	a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
	.001	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.91	3.82	3.75
35	.05	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04	2.01	1.99	1.96
	.01	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64	2.60
	.001	12.9	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.90	3.79	3.70	3.62	3.55
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
	.001	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.55	3.47	3.40

- Quelques formules...  $t_{obs} = \sqrt{n-2} \times \sqrt{\frac{V_{inter}}{V_{intra}}}$        $t_{obs} = \sqrt{n-2} \times \sqrt{\frac{r^2}{1-r^2}}$

$F = \frac{(n-K)}{(K-1)} \times \frac{V_{inter}}{V_{intra}}$        $ddl = K-1$        $ddl2 = n-K$

$r = \frac{Cov(x, y)}{Ety x \cdot Ety y}$        $Cov(x, y) = \frac{\sum (x^i y^i)}{n} - \bar{x} \bar{y}$        $a = \frac{Cov(x, y)}{Var x} = r \times \frac{Ety y}{Ety x}$        $b = \bar{y} - a \bar{x}$

$t^{jk} = \frac{n_{jk} - \hat{n}_{jk}}{\hat{n}_{jk}}$        $F^2 = \sum_{J,K} Cta_{jk}$        $c_{obs}^2 = \frac{(n_{+-} - n_{-+})^2}{n_{+-} + n_{-+}}$

$P(j/k) = \frac{P(j) \times P(k/j)}{\sum_j P(j) \times P(k/j)} = \frac{P(j) \times P(k/j)}{P(k)}$

Les différents calculs ont été réalisés à l'aide des logiciels Statistica et DS3-Win