

Examen d'analyse de données**Mai 2006 – Durée 2 heures****Documents interdits – Calculatrice autorisée.****Exercice 1: (3.5 points)**

On a effectué le croisement de balsamines blanches avec des balsamines pourpres. En première génération, les fleurs sont toutes pourpres. On obtient en deuxième génération quatre catégories avec les effectifs suivants :

Couleurs	pourpre	rose	blanc lavande	blanc
Effectifs	1790	547	548	213

On veut savoir si l'hypothèse de répartition mendélienne $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, $\frac{1}{16}$ est acceptable.

Analyse avec le logiciel R:

```
> a<-c(9, 3, 3, 1)
> a<-a/sum(a)
> a
[1] 0.5625 0.1875 0.1875 0.0625
> x<-c(1790, 547, 548, 213)
> x
[1] 1790 547 548 213

> chisq.test(x,p=a)
Chi-squared test for given probabilities
data: x
X-squared = 7.063, df = 3, p-value = 0.06992
```

- Quel test est utilisé ? Justifier ce choix.
- Quelle hypothèse est testée ?
- Analyser les résultats du test.
- Quels seraient les effectifs attendus si la répartition mendélienne était respectée ?

(*Probabilités et Statistiques pour biologistes* F. Couty, J. Debord et D. Fredon, 1990, Armand Colin.)

Exercice 2: (5 points)

On donne ci-dessous une suite d'instructions R.

```
> oui
[1] 5.54 6.87 7.26 7.94 7.98 8.11 12.04 12.81 13.04 14.28 14.41 15.47
> non
[1] 3.47 4.59 4.72 5.17 5.30 6.59 7.01 7.25 7.28 7.30 7.60 8.16

> t.test(oui,non)
data: oui and non
t = 3.89, df = 22, p-value = 0.0008
alternative hypothesis: true difference in means is not
```

```

equal to 0
95 percent confidence interval:
1.996 6.556
sample estimates:
mean of x mean of y
10.48      6.203

```

1.
 - a. Quel est le test pratiqué ci-dessus ? Justifier son choix.
 - b. Quelles vérifications sont elles nécessaires pour valider son usage ? Préciser pour chacune d'elles au moins une méthode de vérification.
 - c. Analyser les résultats de ce test.

2. Un deuxième test est utilisé ci-dessous.

```

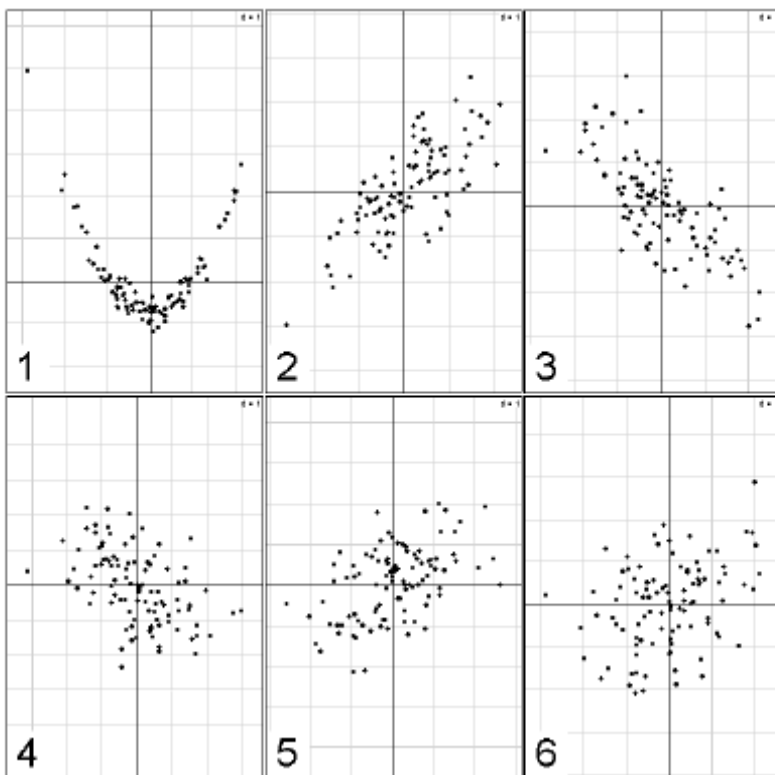
> wilcox.test(oui, non)
data: oui and non
rank-sum statistic W = 202, n = 12, m = 12, p-value = .0018
alternative hypothesis: true mu is not equal to 0

```

- a. Quel est le test utilisé ? Dans quel cas utilise-t-on ce test ?
- b. Analyser le résultat du test.

Exercice 3: (2 points)

Six simulations de statistique bivariable donnent les figures :



Attribuer à chacune d'entre elles son coefficient de corrélation sachant que ces valeurs figurent parmi l'ensemble $\{-1, -0.73, -0.49, -0.04, 0.33, 0.50, 0.74, 1\}$

Exercice 4 : (7.5 points)

Une analyse sensorielle a été organisée simultanément en France et au Pakistan. On a demandé à 150 français et à 163 pakistanais de donner une note d'appréciation à 8 biscuits (0 : je n'aime pas, 10 : j'aime beaucoup). Parmi ces biscuits, 4 sont fabriqués et vendus en France (biscuits notés F1, F2, F3 et F4) et 4 sont fabriqués et vendus au Pakistan (biscuits notés P1, P2, P3, et P4).

L'objectif d'une telle analyse est de comparer les appréciations d'un pays à l'autre.

Un résumé des données est fourni dans le tableau ci-dessous.

	Origine française				Origine pakistanaise			
	F1	F2	F3	F4	P1	P2	P3	P4
Jury français	7,35	6,06	5,08	5,75	4,99	6,03	4,22	5,18
Jury pakistanais	6.29	5.10	5.23	5.38	7.49	6.74	5.56	6.32

1.
 - a. Justifier l'utilisation d'un modèle d'analyse de variance dans cet exercice.
 - b. Quels sont les deux facteurs pris en compte ?
 - c. Proposez un modèle d'analyse de variance pour étudier l'influence de ces deux facteurs.
 - d. Quelles sont les hypothèses à vérifier pour utiliser un modèle d'analyse de variance ?
2. Le listage suivant donne les résultats du modèle d'analyse de la variance sur les seules données du tableau précédent.

	<i>ddl</i>	<i>SCE</i>	<i>CM</i>	<i>f_{obs}</i>	<i>Proba</i>
Nationalité jury	1	0,7439	0,7439	1,2398	0,28732
Origine biscuit	1	0,0053	0,0053	0,0088	0,92698
Nationalité jury : Origine biscuit	1	3,9303	3,9303	0,02503	0,02503
Résiduelle	12	7,2004	0,6000		

Interpréter ce tableau.

3.
 - a. Construire un graphique avec en abscisse les biscuits français et pakistanais et en ordonnée les notes moyennes sur les 4 biscuits par jury. Relier les points d'un même jury.
 - b. Que représente ce graphique?
 - c. Analyser le graphique.

Exercice 5 : (2 points)

- a. Interpréter les résultats du test suivant :

```
> pairwise.t.test(mesure, facteur, P.adj="bonferroni")
```

```
Pairwise comparisons using t tests with pooled SD
```

```
data:  mesure and facteur
```

```
  A      B      C
```

```
B 0.304 -
```

```
C 0.008 0.005 -
```

```
D 0.080 0.012 0.097
```

```
P value adjustment method: holm
```

- b. Les moyennes par modalité sont 5 7 9 11. Retrouver B, C et D sachant que la moyenne de A est 7.