

***** Epreuve d'Analyse des Données (3H ) *****
---

---

**Consignes :**

1. l'épreuve comporte deux pages, s'assurer que c'est bien le cas
  2. les cas des tricheries remarqués sur les copies feront l'objet de sanction.
  3. Documents non autorisés.
- 

***** Problème 1 (10 points) *****
------------------------------------

***Répondre aux questions suivantes en justifiant votre démarche :***

1. Sur quels types de données peut-on réaliser une ACP ?
2. Quels sont les objectifs d'une ACP ?
3. Lorsque nous sommes dans une étude à  $n$  individus et  $p$  variables, dans quel espace vivent les variables ? Dans quel espace vivent les individus ?
4. Donnez une définition de l'inertie totale du nuage des points.
5. Que peut-on dire de la somme des valeurs propres lorsque les données sont centrées et réduites en ACP.
6. En ACP, l'inertie totale est égale à la somme des variances de tous les individus :  
(a) vrai  
(b) faux
7. Quel est l'intérêt essentiel du (des) graphique(s) des variables réalisé(s) à l'issue d'une ACP ?
8. Quel est l'objectif principal d'une A.F.C. ? Quelle précaution doit-on prendre avant de réaliser une A.F.C. ?

***** Problème 2 (10 points) *****
------------------------------------

***Étude de cas :***

On cherche à mettre en oeuvre une stratégie de prédiction du poids utile du poulpe, c'est-à-dire son poids éviscéré, à partir de son poids non éviscéré. C'est en effet le poulpe éviscéré qui est commercialisé. Pour cela, un échantillon de poulpes a été collecté en 2003 lors des opérations de pêche dans les eaux mauritaniennes. Vu l'importante différence de poids entre les poulpes mâles et les poulpes femelles, on étudie ici uniquement les données concernant 240 poulpes femelles.

L'ensemble de ces données est représenté dans la figure 1.

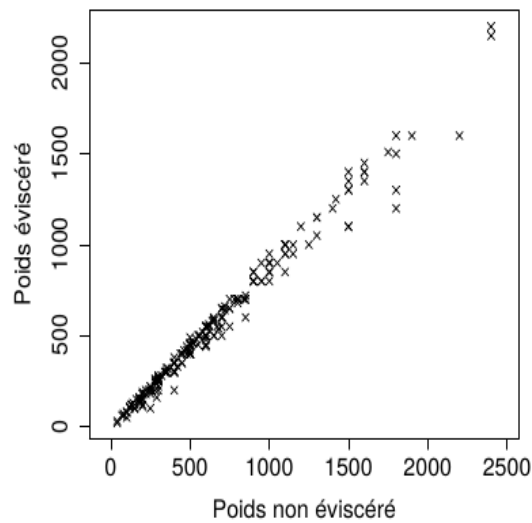


FIGURE 1 – Poids de poulpe éviscéré en fonction du poids non éviscéré (en grammes).

1. Proposer un modèle reliant le poids éviscéré et le poids non éviscéré d'un poulpe. Donner les hypothèses de votre modèle.
2. Rappeler les formules des estimateurs des paramètres du modèle.
3. Que peut-on dire de la qualité d'ajustement de la droite des moindres carrée aux données ? Justifier votre réponse.
4. A partir de la sortie R, donner les estimations numériques des paramètres du modèle.
5. Que représente la valeur 0.698 dans la sortie R ? L'interpréter.
6. Au vu de cette valeur 0.698, proposer un autre modèle reliant les poids éviscéré et non éviscéré.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.312146	5.959670	-0.388	0.698
Poids non éviscéré	0.853169	0.007649	111.545	<2e-16

Residual standard error: 52.73 on 238 degrees of freedom  
Multiple R-Squared: 0.9812, Adjusted R-squared: 0.9812  
F-statistic: 1.244e+04 on 1 and 238 DF, p-value: < 2.2e-16

FIGURE 2 – sortie R.