

# L3 IA – R pour l'Intelligence Artificielle – évaluation 1

Le rendu de ce TP donnera lieu à une note. Ceci est la première évaluation de ce module.

**Vous n'avez pas le droit de communiquer entre vous, ni d'utiliser Internet, ni une quelconque IA.**

---

**Construire un compte-rendu de TP avec, pour chaque question :**

L'intitulé de la question

**Le code R**

La sortie R (valeur(s), tableau ou graphique)

Et un commentaire de la sortie quand c'est utile

**Veiller à laisser dans le script tout le code nécessaire pour construire la réponse à la question.**

---

**Il faudra rendre **ce fichier** et **le fichier de script** dans la boîte de dépôt sur Moodle avant la fin de la séance.**

1. Importer le fichier « sportifs ». Donner le bon type à chacune des variables.  
Décoder la variable Région, sachant que le code « 1 » signifie « Nord », le code « 2 » signifie « Ouest », le code « 3 » signifie « Est », le code « 4 » signifie « Sud ».  
Décoder la variable de La motivation pour le sport dont les codes 1, 2, 3, 4 et 5 signifient respectivement par « Très faible », « Faible », « Modérée », « Forte », « Très forte ».
2. Décrire les données du dataframe. Combien de variables ? De quels types ? Combien d'individus ?
3. Ajouter au dataframe les variables suivantes :
  - Reussite qui calcule le nombre de trophées moyen par compétition pour chaque sportif annuellement
  - Categorie qui prend les valeurs « assidu » ou « très motivé » selon si le temps d'entraînement hebdomadaire est inférieur, ou supérieur ou égal à 5 heures.Afficher les statistiques nécessaires pour vérifier que la création de ces variables s'est bien déroulée.
4. Afficher un résumé statistique clair de l'ensemble des données, permettant notamment de faire un bilan sur les valeurs manquantes.

*Suite page suivante*

5. Effectuer les traitements statistiques des variables suivantes en personnalisant les graphiques.
- |                                      |   |
|--------------------------------------|---|
| a. Le sport pratiqué                 | Tableaux des effectifs et des fréquences<br>Graphique en tuyaux d'orgue   |
| b. La masse musculaire en kg         | Statistiques de tendance centrale et de dispersion<br>Histogramme en imposant des classes bien choisies<br>Tableaux des effectifs et des fréquences selon ces classes |
| c. La satisfaction du sport pratiqué | Tableaux des effectifs et des fréquences<br>Graphique en tuyaux d'orgue   |
| d. Le nombre de trophées annuel      | Statistiques de tendance centrale et de dispersion<br>Tableaux des effectifs et des fréquences<br>Diagramme en bâtons   |
| e. Le sponsor                        | Graphique en secteurs circulaires des fréquences  |
6. On cherche à voir si le temps d'entraînement hebdomadaire en heures varie selon le sponsor. Pour répondre à cette question :
- calculer le tableau des moyennes, écart-types, quartiles conditionnels,
  - tracer le box-plot correspondant en le personnalisant au maximum,
  - tracer un histogramme pour chaque sponsor et les placer dans une même fenêtre graphique.
7. Après avoir commenté toutes les questions précédentes, compléter par tout traitement intéressant selon vous, en diversifiant les types de traitement.