

CAHIER DES CHARGES

STAT-III

*PROJET DE CONSOLIDATION*

*APPRENTISSAGE DU TRAITEMENT ET DES ANALYSES DE DONNEES*

***Deuxième Année ISTOM***

***Mars 2024***

1. CONTEXTE

**Nous avons engagé en 2021-2022 un exercice PEAD (*Première expérience en Analyse de Données*) pour les étudiants 2A, dans le programme STAT3.**

**C ’est un apprentissage qui se veut pluridisciplinaire axé sur l’organisation, le traitement et l’interprétation de** données.

Malgré les imperfections, des résultats sont constatés, l’intérêt des étudiants étaient rendez-vous.

Cette année nous avons la mission de renouveler l’expérience.



Introduction Analyse factorielle.

***Méthodologie Scientifique*** *(Daniel Kalnin) KKALNIN)*

*Année 4*

**EMOA**

**D-M**

**(A**

*Deuxième année,*

S**emestre 3 et Semestre 4**

*Troisième année,*

**Semestre 5**

*Première année*

**Semestre 2**

1. OBJECTIF :

Réaliser une première expérience en Data Mining :

* Pister des données pour répondre à une problématique formuler à partir de la méthodologie scientifique.
* Mobiliser une méthodologie statistique formulée à partir de la méthodologie scientifique pour exploiter des informations.
* Prendre conscience des marges de vigilance nécessaire aux égards des données.

1. PERIMETRE : pays « des Suds » (…)
2. THEMATIQUE : en lien avec les objectifs de formation à l’ISTOM.
3. DIMENSION DONNEES : *inf* (dim) = 5 ( # variables), *min* ( Echantillon) = 100
4. PUBLIC CIBLE**:** semestre 4 (deuxième Année).
5. PREREQUIS **:** Statistique descriptive univariée, boîte de Tukey, asymétrie, statistique bivariée, analyse de corrélation, ajustement, régression, décomposition des séries temporelles, échantillonnage, estimation, lois des probabilités, table statistique, statistique inferentielle, analyse comparative, tests paramétriques.

Logiciel R et packages Rcmdr, corrplot, tseries, factoextra, FactoMineR, MASS, ggplot2, etc…..

1. MATERIELS ET OUTILS :

* *Outils****:*** Logiciel R + package Rcmdr, tseries….
* R-data iris // R-data crabs ( package MASS)
* [http://www.fao.org/faostat/en/#data](http://www.fao.org/faostat/en/)
* <https://comtrade.un.org/data/>
* <https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx?sCS_ChosenLang=en>
* <https://www.imf.org/en/Data>
* <https://data.worldbank.org/>
* <https://ec.europa.eu/eurostat/data/database>

*NB : d’autres bases de données sont possibles après proposition de votre part à la commission (cf. ci-dessous).*

***9-ENCADREURS Sollicités****: IST/DK/SL/AG/MA/MB*

1. ETAPES ET PROCEDURES**:**

* *03- 10 Mars 2024* : construction des groupes (3 étudiants par Groupe) délégués.
* *10 Mars- 25 Avril 2024* recherche de bases de données.
* *29 Avril à 18 heures****:*** dépôt d’un document PPT (4 diapos) d’investigation sur Moodle (thématique choisie et questionnement initial, justification et pertinence, lien thématique et base de données, interrogation de la base de données, démarche statistique proposée).
* *02 Mai 2024* : validation (données-thématiques).
* *03 Mai au 23 Mai 2024: travaux d’analyses- encadrements (C.f. Prérequis, matériels et outils)*
* *25 Mai à 18 h****:*** Fin des travaux, le document (*cf 11- Design document final)* à déposer sur Moodle.

1. DESIGN DOCUMENT FINAL :

Document *structuré* de 04 pages,

* *Page 01 : titre de l’étude, noms et prénoms des auteurs, année universitaire, problématisation*, *questionner la base de données. Proposer une méthode d’analyse statistique* (Statistique descriptive, analyse de corrélation, analyse explicative, tendance et prédiction, analyse comparative et Test statistique). Résumer la démarche en exhibant les principaux socles de l’analyse visée.

**NB**: *cette première page doit être suffisamment attractive pour inciter le lecteur à poursuivre son investigation.*

* Page 02-03 : présentation données, *sources, origine, population cible, variables et* des *livrables* (*tableau de test, graphe, etc*…) et ce, conformément aux étapes mentionnées à la page 01. Discussion sur les limites de l’analyse
* *Page 04 :* conclusion, r*éférences, annexes.*

**NB :** *il n’est pas nécessaire de lister tous les traitements statistiques entrepris ou réalisables. Seules les analyses qui répondent aux attentes de l’objectif principal seront présentées. La pertinence recherchée est l’obtention d’une intégration harmonieuse des objets statistiques tels que (courbes, graphes, nuage, tableaux de test, etc…) ainsi que les commentaires associés. Chacun de ces objets statistiques est numéroté, accompagné d’une légende en italique.*

1. **PERSONNES RESSOURCES :**

MB : coordinateur

Encadrement : IST-MB-AG-DK-MA

**12- BAREME :**

* *Td noté WIMS : 20%*
* PEAD: 25 %
* *Partiel: 55 %*