**PSY4016 – Plan du projet – Fadila Moussaoui**

**Description des données :**

Cette base de données contient des informations sociodémographiques, des informations sur les habitudes de vies et sur les traits de personnalité.

**Variables utilisées :**

Pour cette étude, les groupes seront formés en fonction de la variable dichotomique « Meditation ». Il y aura deux groupes, ceux qui pratiquent la méditation et ceux qui ne la pratiquent pas. La méditation est donc la variable indépendante qui sera analysée en fonction des variables dépendantes suivantes : « Empathy.Agreeableness » et « Disinhibition ». Pour le plan B, la variable « Sex » sera prise en compte.

**Type des variables :**

Les variables utilisées, soit « Meditation », « Empathy.Agreeableness », « Disinhibition » et « Sex », sont des variables de type chaîne (str) dans le fichier csv. Afin d’effectuer les analyses, il faudra coder la « Meditation » et le « Sex » chacun en deux groupes distincts de type integer (int). Il faudra aussi transformer les variables « Empathy.Agreeableness » et « Disinhibition » en chiffres à virgule (float).

**Hypothèses :**

Selon la littérature sur ce sujet, je crois que :

1. Les individus pratiquant la méditation auront de plus hauts niveaux d’empathie.
2. Les individus pratiquant la méditation auront de plus faibles niveaux de désinhibition.

**Exemple des graphiques pour chaque hypothèse :**

Hypothèse 1 : Hypothèse 2 :



**Les défis attendus :**

Certains défis que j’appréhende pour le projet sont d’importer le fichier csv dans jupyter et de transformer les variables de type chaîne à des variables de type integer ou float. De plus, je crois qu’il sera un défi de gérer les données manquantes et les scores extrêmes afin de réaliser des analyses statistiques de qualité. Je devrai diminuer le seuil alpha pour mon deuxième test T afin de gérer le problème des comparaisons multiples.

**Pseudo-code, les étapes d'analyse :**

**Étape 1 : Importer la base de données et imprimer le tableau**

import pandas as pd

df = pandas.read\_csv("/Users/fadilamoussaoui/Desktop/Session H22/PSY4016 –

Programmation/Projet/Base\_de\_données\_travail.csv", sep = ";")

df

**Étape 2 : Supprimer les colonnes inutiles au projet et vérifier les colonnes restantes**

df\_projet = df.drop(columns =['Age','Salary','Study\_Level','Mood\_Disorder','BMI','Sport','Music',

'Extraversion','Conscientiousness','Instability.Neuroticism','Openness\_to\_Experience.Intellect', 'Honesty.Humility','Detachment',

'Psychoticism','Negative\_Affect','Antagonism'])

**Étape 3 : Trouver les données manquantes**

for value in df\_projet["Sex"]:

print(value)

for value in df\_projet["Meditation"]:

print(value)

for value in df\_projet["Empathy.Agreeableness"]:

print(value)

for value in df\_projet["Disinhibition"]:

print(value)

**Étape 4 : Gérer les données manquantes**

Avec sklearn, seaborn et impute

Calculer la moyenne des variables avec la commande ".mean" et créer une colonne "mean"

sklearn.impute.SimpleImputer(missing\_values = np.nan, strategy = "mean")

**Étape 5 : Vérifier les types de variables**

print(type("Sex"))

print(type("Meditation"))

print(type("Empathy.Agreeableness"))

print(type("Disinhibition"))

**Étape 6 : Transformer les variables de chaîne à int ou float**

Avec la commande : sklearn.preprocessing.LabelEncoder() et imputer dans une variable X

Faire ".fit" sur cette variable X

**Étape 7 : Créer les scores Z**

Avec la commande : sklearn.preprocessing.StandardScaler()

**Étape 8 : Gérer les scores extrêmes**

Avec la commande : sklearn.preprocessing.PowerTransformer()

**Étape 9 : Test T pour hypothèse 1**

Test T indépendant entre la pratique de la méditation (VI) et le niveau d’empathie (VD).

Avec la commande : scipy.stats.ttest\_ind()

**Étape 10 : Graphique pour hypothèse 1**

Avec la commande : seaborn.catplot

sns.catplot(data=tips, x = 'Meditation', y = 'Empathy. Agreeableness', kind='box')

set\_axis\_labels('Meditation', 'Empathy')

**Étape 11 : Test T pour hypothèse 2**

Test T indépendant entre la pratique de la méditation (VI) et le niveau de désinhibition (VD).

Avec la commande : scipy.stats.ttest\_ind()

**Étape 12 : Graphique pour hypothèse 2**

Avec la commande : seaborn.catplot

sns.catplot(data=tips, x = 'Meditation', y = 'Disinhibition', kind='box')

set\_axis\_labels('Meditation', 'Disinhibition')

**Étape 13 : Refaire les test T pour les hypothèses 1 et 2 en contrôlant pour le sexe (si besoin du plan B)**

On ajoute la variable « Sex » en covariable.

**Description de la méthode statistique :**

Je vais utiliser un test T indépendant pour répondre à chacune de mes hypothèses. Ce test a été choisi, car il y a deux groupes indépendants et une variable dépendante continue pour chacune des hypothèses.

**Présentation des résultats attendus :**

Nous prédisons une différence significative entre les individus qui font de la méditation et ceux qui n’en font pas pour le niveau d’empathie, où ceux qui pratique la méditation ont des niveaux plus élevés. Nous prédisons une différence significative entre les individus qui font de la méditation et ceux qui n’en font pas pour le niveau de désinhibition, où ceux qui pratique la méditation ont des niveaux plus faibles.

**Calendrier, échéancier et temps alloué :**

|  |  |  |
| --- | --- | --- |
| **Semaine** | **Travail à faire** | **Temps alloué** |
| 28 février | Transformer variables en integer/float, gérer données manquantes | 3h |
| 7 mars | Créer les scores Z, gérer les scores extrêmes | 3h |
| 14 mars | Code pour faire les statistiques, code pour faire les graphiques | 3h |
| 21 mars | Code pour AA supervisé, code pour graphique AA supervisé | 3h |
| 28 mars | Code pour AA non-supervisé, code pour graphique AA non-supervisé | 3h |
| 4 avril | M’assurer que j’ai tous les éléments attendus dans mon code | 3h |
| 11 avril | Rapport et description des résultats | 3h |
| 18 avril | Révision et remise en fichier .py sur github | 2h |

**Description du groupe de personne attendu :**

Mon étude sera présentée à la population générale et sera accessible à tous.

**Plan A :**

Déterminer la relation entre la pratique de méditation et l’empathie, puis la pratique de méditation et la désinhibition.

**Plan B :**

Déterminer si les résultats sont différents lorsqu’on contrôle pour le sexe.