****

**Département de mathématique et d’informatique**

**SDD1002**

**Travail pratique 1**

**Présenté à  
M. Usef Faghihi**

**10 Février 2024**

**Par :** GUESSOUM Abderraouf Zoubir, GUEA80320400

**Table des matières**

1. Introduction
2. Travaux préexistants
3. Relations
   1. Temps de réponse – âge
   2. Temps de réponse – score
   3. Temps de réponse – genre
   4. Corrélation générale
4. Conclusion
5. Introduction

Les données présentées dans ce document ont été recueillies à partir du Narcissistic Personality Inventory (NPI), un questionnaire utilisé pour évaluer le degré de narcissisme d'un individu. Ces données, structurées sur 44 colonnes et 11244 rangées, ont été obtenues via OpenPsychometrics.org et sont disponibles sur [Kaggle](https://www.kaggle.com/datasets/lucasgreenwell/narcissistic-personality-inventory-responses) . Avant de plonger dans l'analyse des données, il est important de comprendre la condition psychiatrique du narcissisme ainsi que le contexte dans lequel le test NPI est utilisé.

Le narcissisme est un trait de personnalité caractérisé par un amour excessif de soi-même, un besoin constant d'admiration et un manque d'empathie envers les autres. Bien que certains niveaux de narcissisme puissent être considérés comme normaux, un narcissisme excessif peut entraîner des conséquences négatives sur les relations interpersonnelles et le fonctionnement global de l'individu.

Le Narcissistic Personality Inventory (NPI) est l'un des outils les plus couramment utilisés pour évaluer le narcissisme. Il se compose de 40 questions conçues pour mesurer différents aspects du narcissisme, tels que la grandiosité, le besoin d'admiration et le sentiment de supériorité. Les réponses des participants à ces questions fournissent des informations sur la prévalence et les caractéristiques du narcissisme dans la population étudiée.

C’est à travers mon intérêt pour le sujet et mon regard sur le set de données présenté aujourd’hui que je me suis demandé comment le temps de réponse aux questions du NPI varie-t-il en fonction de l'âge, du score de celui-ci et du genre des participants, et quelles sont les éventuelles corrélations entre ces quatre éléments ? Pour répondre à cette question, nous verrons en premier lieu les travaux préexistants. S’en suivra alors la mise en évidence, respectivement, des variations de temps de réponse en fonction de l’âge, du genre, et du score. Nous finirons cette analyse en mettant les 4 variables en relation pour constater des corrélations si corrélations il y a.

1. Travaux préexistants

Les données utilisées dans cette étude ont été soigneusement compilées et préparées par Lucas Greenwell, un analyste de données basé à Washington, dont l'expertise et l'engagement dans le domaine de la psychologie se reflètent dans son travail approfondi sur l'analyse de données psychologiques. Fort de son expérience et de sa passion pour ce domaine, Lucas Greenwell a rassemblé ces données il y a environ quatre ans, mettant à profit sa connaissance approfondie des méthodologies de recherche et son engagement envers l'excellence dans l'analyse des données.

Ce jeu de données, qui est disponible depuis plusieurs années, a été mis à disposition via la plateforme Kaggle, offrant ainsi aux chercheurs et aux professionnels de la psychologie un accès pratique à des données riches et pertinentes pour leurs études et leurs recherches.

Au cœur de ces données se trouve le Narcissistic Personality Inventory (NPI), un outil essentiellement utilisé pour évaluer les traits narcissiques chez les individus. Composé de 40 questions soigneusement conçues, le NPI offre un aperçu détaillé des différents aspects du narcissisme, allant de la grandiosité à l'admiration de soi en passant par la recherche de pouvoir et le manque d'empathie. Chaque participant à l'étude répond à ces questions en choisissant l'une des deux options de réponse proposées pour chaque question, chaque option étant notée soit 1 soit 2, fournissant ainsi une évaluation quantitative des tendances narcissiques chez les individus.

En outre, les données comprennent des informations précieuses sur le temps de réponse pour chaque question posée. Ce temps de réponse est calculé comme la différence entre le moment où la question est présentée au participant et le moment où sa réponse est enregistrée. Cette mesure de temps de réponse peut révéler des informations importantes sur la réactivité des participants lorsqu'ils sont confrontés à des questions liées au narcissisme, offrant ainsi des insights supplémentaires sur la dynamique complexe des réponses au NPI.

En ce qui concerne les informations démographiques, les données incluent également des données sur le genre des participants. Utilisant un système de codage simple mais informatif, les participants sont catégorisés comme étant de genre masculin (codé comme 1), de genre féminin (codé comme 2), ou de genre autre (codé comme 3), tandis qu'une valeur de 0 indique que le participant n'a pas choisi de répondre à cette question.

Il est important de souligner que ces données ont été collectées et traitées dans le respect des normes éthiques les plus strictes. Des protocoles de protection des données ont été mis en place pour garantir la confidentialité et l'anonymat des participants, tout en veillant à ce que les résultats de l'étude soient fiables et significatifs.

En somme, ces données fournissent une ressource précieuse pour les chercheurs et les professionnels de la psychologie, offrant un aperçu détaillé des traits narcissiques chez les individus et permettant des analyses approfondies sur la base de données solides et fiables.

Aussi, parmi les autres travaux préexistants, [une étude corrélationnelle entre les réponses aux questions et le score au test](https://www.kaggle.com/code/kerneler/starter-narcissistic-personality-c6c025cc-6) qui, finalement, n’ont révélé aucune conclusion spécifique ou relation entre les réponses et les scores.

On peut le voir à la heatmap suivante :

A blue and yellow squares with a yellow line

Description automatically generated

1. Relations

Avant de m’approfondir dans l’analyse en elle-mëme, un cour travail de nettoyage du set fût effectué par soucis de vraissemblance. Celui-ci était constitué de :

* La suppression des rangées où l’âge correspondait aux valeurs suivantes : 0, 2, et toutes les valeurs égales ou supérieures à 148.
* La suppression des rangées où le temps de réponse est supérieur 5454 secondes, ce qui équivaut à environ 1h30.
  1. Temps de réponse – âge

L'analyse de la variation du temps de réponse en fonction de l'âge revêt une grande importance dans cette étude pour plusieurs raisons. Tout d'abord, elle permet d'explorer comment les différents groupes d'âge réagissent différemment aux questions du Narcissistic Personality Inventory (NPI), offrant ainsi un aperçu des schémas de réponse et des tendances qui pourraient être spécifiques à certaines tranches d'âge. Cette analyse pourrait également révéler des informations sur le développement du narcissisme au fil du temps, en examinant comment les réponses des participants évoluent avec l'âge. En outre, comprendre comment le temps de réponse varie en fonction de l'âge peut fournir des insights sur la cognition et la réflexion des participants lorsqu'ils répondent au questionnaire, ce qui pourrait avoir des implications importantes pour l'interprétation des scores au NPI.

Voici le script python utilisé sur Jupyter Notebook pour obtenir le graphique qui suivra :

*import* pandas *as* pd

*import* matplotlib.pyplot *as* plt

donnees = pd.read\_csv("donnees.csv")

plt.figure(*figsize*=(10, 6))

plt.scatter(donnees['age'], donnees['elapse'], *alpha*=0.5)

plt.title("Variation du temps de réponse en fonction de l'âge")

plt.xlabel("Âge")

plt.ylabel("Temps de réponse (en secondes)")

plt.xlim(0, 120)

plt.ylim(0, 5460)

plt.grid(True)

plt.show()

A graph with blue dots

Description automatically generated

Au-delà de la forte concentration de l’âge des participants entre 17 et 75 ans, le graphique révèle une forte majorité des temps de réponse des participants en dessous de 1000 secondes, indiquant que la majorité des répondants ont répondu relativement rapidement aux questions. Cela suggère une tendance générale vers des temps de réponse courts, avec une proportion significative de participants répondant dans un laps de temps relativement court.

* 1. Temps de réponse – score

Étudier la relation entre le temps de réponse et le score au NPI est essentiel pour mieux comprendre la dynamique des réponses des participants et leur lien avec les traits narcissiques. Cette analyse permettra d'explorer si les individus présentant des niveaux plus élevés de narcissisme ont tendance à répondre plus rapidement ou plus lentement aux questions du NPI, et si cela peut être lié à des différences dans la manière dont ils interprètent ou réagissent aux questions. En examinant cette relation, il est possible d'identifier des schémas de réponse spécifiques qui pourraient être associés à des niveaux plus élevés ou plus faibles de narcissisme, ce qui enrichirait notre compréhension de la façon dont les traits narcissiques se manifestent dans les réponses au NPI.

Voici le script python utilisé sur Jupyter Notebook pour obtenir le graphique qui suivra :

*import* pandas *as* pd

*import* matplotlib.pyplot *as* plt

donnees = pd.read\_csv("donnees.csv")

plt.figure(*figsize*=(10, 6))

plt.scatter(donnees['score'], donnees['elapse'], *alpha*=0.5)

plt.title("Relation entre le temps de réponse et le score au NPI")

plt.xlabel("Score au NPI")

plt.ylabel("Temps de réponse (en secondes)")

plt.grid(True)

plt.show()

A graph with blue dots

Description automatically generated

Une analyse du graphique met en évidence une grande diversité des scores au NPI, s'étendant sur toute la plage de 0 à 40. Cependant, on observe une concentration significative des temps de réponse entre 0 et 1000 secondes. Cette répartition uniforme des scores combinée à une concentration des temps de réponse relativement courts suggère qu'aucune corrélation spécifique entre le score au NPI et le temps de réponse n'émerge de manière évidente à partir de ces données.

* 1. Temps de réponse – genre

Analyser l'impact du genre sur le temps de réponse revêt une importance particulière dans cette étude en raison des différences de comportement et de cognition entre hommes et femmes qui pourraient influencer la manière dont ils répondent aux questions du NPI. Cette analyse permettra de déterminer s'il existe des variations significatives dans les temps de réponse entre hommes et femmes, ce qui pourrait indiquer des différences dans la façon dont ils abordent les questions liées au narcissisme. Comprendre ces différences de comportement entre les genres est crucial pour interpréter correctement les résultats du NPI et pour s'assurer que les évaluations du narcissisme sont sensibles aux différences individuelles et contextuelles.

Voici le script python utilisé sur Jupyter Notebook pour obtenir le graphique qui suivra :

*import* pandas *as* pd

*import* seaborn *as* sns

*import* matplotlib.pyplot *as* plt

donnees = pd.read\_csv("donnees.csv")

plt.figure(*figsize*=(10, 6))

sns.kdeplot(donnees[donnees['gender'] == 1]['elapse'], *label*='Hommes')

sns.kdeplot(donnees[donnees['gender'] == 2]['elapse'], *label*='Femmes')

plt.title("Comparaison du temps de réponse entre hommes et femmes")

plt.xlabel("Temps de réponse (en secondes)")

plt.ylabel("Fréquence")

plt.legend()

plt.grid(True)

plt.show()

A graph with a line graph

Description automatically generated

Une analyse des courbes de densité révèle un pic distinct autour de 300 secondes pour les temps de réponse, à la fois pour les hommes et les femmes, suggérant une tendance commune à répondre dans ce laps de temps pour les deux groupes. De plus, bien que légèrement plus d'hommes que de femmes aient participé à l'étude, aucune tendance significative ne semble émerger en ce qui concerne les différences de temps de réponse entre les deux groupes de genre.

* 1. Corrélation générale

L'exploration des corrélations entre l'âge, le score au NPI, le sexe et le temps de réponse révèle des tendances intéressantes dans les données. Tout d'abord, en analysant les relations entre ces variables, nous pouvons observer si des schémas spécifiques émergent, tels que des différences dans les temps de réponse en fonction de l'âge ou du sexe des participants. Ensuite, cette analyse multicorrélative nous permettra de déterminer si le score au NPI est associé à des temps de réponse particuliers, et si cette association varie en fonction de l'âge ou du sexe. Enfin, en examinant la corrélation entre le temps de réponse et les autres variables, nous pouvons évaluer dans quelle mesure le temps de réponse peut être un indicateur fiable du narcissisme, en tenant compte des différences démographiques et individuelles des participants.

Voici le script python utilisé sur Jupyter Notebook pour obtenir le graphique qui suivra :

*import* pandas *as* pd

*import* seaborn *as* sns

*import* matplotlib.pyplot *as* plt

donnees = pd.read\_csv("donnees.csv")

colonnes\_selectionnees = ['age', 'score', 'gender', 'elapse']

donnees\_selectionnees = donnees[colonnes\_selectionnees]

matrice\_corr = donnees\_selectionnees.corr()

plt.figure(*figsize*=(10, 8))

sns.heatmap(matrice\_corr, *annot*=True, *cmap*='coolwarm', *fmt*=".2f", *linewidths*=.5)

plt.title("Matrice de corrélation entre l'âge, le score au NPI, le genre et le temps de réponse")

plt.show()

A screenshot of a graph

Description automatically generated

Une analyse de la matrice de corrélation révèle des coefficients proches de zéro pour toutes les paires de variables, suggérant qu'il n'existe pas de corrélation linéaire significative entre l'âge, le score au NPI, le genre et le temps de réponse (elapse). Cependant, il convient de noter que l'absence de corrélation linéaire ne préjuge pas de l'absence de relations ou de liens entre ces variables. Il est possible que des relations non linéaires ou des interactions plus complexes existent, mais ne sont pas détectées par la simple corrélation linéaire. Ainsi, pour obtenir une compréhension plus complète des relations entre ces caractéristiques des participants, il pourrait être nécessaire d'explorer d'autres méthodes d'analyse ou de considérer l'inclusion de variables supplémentaires.

1. Conclusion

En conclusion, cette étude visait à explorer les relations entre l'âge, le score au Narcissistic Personality Inventory (NPI), le genre et le temps de réponse dans le cadre d'une analyse des données recueillies. L'analyse des résultats a révélé que, bien que des tendances intéressantes puissent être observées dans la distribution des réponses et des scores au NPI, aucune corrélation significative n'a été mise en évidence entre le temps de réponse et les autres éléments étudiés. En particulier, la matrice de corrélation a montré des coefficients proches de zéro pour toutes les paires de variables, indiquant une absence de corrélation linéaire entre le temps de réponse, l'âge, le score au NPI et le genre des participants. Cela suggère que d'autres facteurs ou mécanismes pourraient être en jeu dans la dynamique des réponses au questionnaire du NPI. Ainsi, des recherches futures pourraient explorer d'autres dimensions ou variables pertinentes pour mieux comprendre les déterminants du narcissisme perçu.