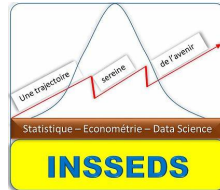


REPUBLIQUE DE CÔTE D'IVOIRE



Union – Discipline – Travail

*Ministère de l'Enseignement Supérieur et de
la Recherche Scientifique*



Institut Supérieur de Statistique, d'Econométrie
et de Data Science

MINI-PROJET : **STATISTIQUE INFERENTIELLE**

Pour le diplôme d'Ingénieur Statisticien et Data

SUJET :

**ETUDE DE LA SATISFACTION DES
PATIENTS DES HOPITAUX PARISIENS
AVEC PYTHON**

Rédigé par :

SAVANE MORY

Promotion 2023-2024

Ingénierie Statistique et Data Science

Période de rédaction : Janvier 2024

ENCADREUR PEDAGOGIQUE :

M. Didier Martial Akposso

Directeur Fondateur de l'INSEDS

Economiste-Statisticien

Data-Scientist

Avant-Propos

Ce travail s'inscrit dans le cadre de la formation au cycle d'ingénieur Statisticien et Data à l'Institut Supérieur de Statistique, d'Économétrie et de Data Science (INSEDS). L'objectif de ce mini-projet est d'évaluer la compréhension du cours portant sur la statistique inférentielle. Il est important de souligner que ce projet est strictement réalisé à des fins académiques et n'engage en aucun cas la responsabilité de l'école.

En tant qu'auteur, nous reconnaissons humblement que notre travail ne prétend pas avoir exploré tous les aspects nécessaires pour conférer à cette étude le caractère d'une recherche véritablement approfondie. Il s'agit plutôt d'une exploration limitée dans le cadre des enseignements dispensés au sein de notre formation.

Nous souhaitons ainsi offrir une contribution modeste à la compréhension des concepts étudiés tout en restant conscient de la complexité et de l'étendue du domaine de la statistique inférentielle. Ce travail ne prétend pas être exhaustif, mais plutôt constituer une étape préliminaire dans notre parcours académique.

Nous exprimons notre gratitude envers nos enseignants et toute l'équipe pédagogique de l'INSEDS pour leur encadrement et leurs enseignements enrichissants qui ont contribué à la réalisation de ce projet.

Enfin, nous espérons que ce travail pourra susciter des réflexions et des discussions fructueuses au sein de la communauté académique et être une source d'apprentissage continu pour tous ceux qui s'intéressent à ce passionnant domaine.

Liste des illustrations

Listes des figures

Figure 1: Informations initiales du jeu de données	12
Figure 2: Visualisation des données manquantes avant traitement	14
Figure 3: Résultat après imputation.....	15
Figure 4: Traitement des doublons.....	16
Figure 5: Boîte à moustache avant Winzoration.....	17
Figure 6: Boîte à moustache après Winzoration.....	17
Figure 7: Boxplots des variables quantitatives.....	19
Figure 8: Diagramme en barre des variables qualitatives.....	21
Figure 9: Tableau des effectifs des variables qualitatives	22
Figure 10: Diagramme groupé de la repartition des recommandations par sexe	25
Figure 11: Résultat du test de Khy-deux	26
Figure 12: Boîte à Moustache des sous-groupes	27
Figure 13: Résumé statistiques du score relation par genre	27
Figure 14: Histogramme du score relation par genre	28
Figure 15: Résultats du test de Shapiro	28
Figure 16: Résultat du test de Wilcoxon	29
Figure 17: Boîte à Moustache du ScoreInfo par profession	31
Figure 18: Résumé statistiques du ScoreInfo par profession	31
Figure 19: Histogramme du score relation par profession	32
Figure 20: Résultats du test de Shapiro	33
Figure 21: Résultat du test de Kruskal Walli.....	33
Figure 22: Boîte moustache du score relation par service.....	35
Figure 23: Résumé statistiques du score.relation par service	36
Figure 24: Histogrammes du ScoreInfo par service fréquenté	37
Figure 25: Résultats du test de Shapiro	37
Figure 26: Résultat du test de Kruskal Wallis	37
Figure 27: Résultats des tests de Shapiro Wilk	37
Figure 28: Nuage de point de la variable age et du score.relation.....	39
Figure 29: Histogramme de la variable âge.....	40
Figure 30: qq-plot des variables age et score.relation	40
Figure 31: résultats du test de Shapiro wilk sur les variables age et score.relation.....	41
Figure 32: Resultat du test de Spearman	41

Liste des tableaux

Tableau 1: Dictionnaire des données.....	13
Tableau 2: Tableau des données manquantes par variables	14
Tableau 3: Résumé statistiques	20
Tableau 4: Tableau de contingence des effectifs empiriques des hommes et des femmes	25

Sommaire

Avant-Propos.....	1
Sigles	Erreur ! Signet non défini.
Liste des illustrations.....	2
Listes des figures	2
Liste des tableaux	2
Sommaire	Erreur ! Signet non défini.
Résumé	4
Abstract	5
Introduction Générale.....	9
PARTIE I : PRETRAITEMENT ET STATISTIQUE DESCRIPTIVE	11
I. PRETRAITEMENT DU JEU DE DONNEES.....	12
Introduction	12
Conclusion.....	18
II. STATISTIQUE DESCRIPTIVE UNIVARIÉE	19
Introduction	19
1. Analyse univariée des variables quantitatives	19
2. Analyse univariée des variables qualitatives	21
Conclusion.....	22
PARTIE II : STATISTIQUE INFERENTIELLE	23
I- TEST DE COMPARAISON DE DEUX POPULATIONS INDEPENDANTES.....	24
Introduction	24
1- La proportion de sujets recommandant le service dans lequel ils sont passés est-il le même quel que soit le genre ?	24
2- La moyenne du score relation est-il significativement différent chez les hommes et chez les femmes ?	26
Conclusion.....	29
II- TEST DE COMPARAISON DE PLUSIEURS POPULATIONS INDEPENDANTES.....	30
Introduction	30
1- Peut-on affirmer que la qualité de l'information reçue est la même quelle que soit la profession ?	30
2- La qualité des relations avec le personnel soignant est-il fonction du service ayant accueilli le patient ?	35
Conclusion.....	38
III- TEST DE LIAISON OU D'INDEPENDANCE : Le score de relation est-il significativement corrélé à l'âge ?	39
Introduction	39
Conclusion générale	42
Annexe	44
Bibliographie et Webographie.....	49

Résumé

Cette étude vise à évaluer la qualité de la relation et la quantité d'information reçue par le patient lors de son séjour à l'hôpital. L'objectif principal est donc de tester si les patients qui fréquentent les hôpitaux de la région parisienne sont globalement satisfaits ou pas et le ou les facteurs contribuant le plus à leur satisfaction.

L'analyse a montré que les patients sont majoritairement des cadres d'entreprises et adultes d'âge moyen, 57 ans.

En ce qui concerne le résultat des tests statistiques, il ressort que :

- **Les patients masculins comme féminin recommandent les services** par lesquels ils sont passés à leur entourage. Ce qui signifie que **le personnel soignant ne fait pas de distinction entre les genres** dans le traitement des patients.
- **Les patients sont satisfaits de la qualité de la relation** avec le personnel soignant,
- La quantité d'information reçue du personnel soignant est la même quelque soit la profession du patient,
- La qualité de la relation avec le patient ne dépend pas du service fréquenté,
- La qualité de la relation ne dépend pas de l'âge du patient.

Ces résultats prouvent que les hôpitaux enquêtés sont globalement justes dans le traitement des patients.

Pour finir, nous félicitons le personnel soignant pour son professionnalisme et l'encourageons à continuer dans ce sens de respect de la dignité humaine.

Mots clés : Statistique inférentielle, test statistique, Satisfaction, évaluation, patients

Abstract

This study aims to assess the quality of the relationship and the amount of information received by patients during their hospital stay. The main objective is to determine whether patients attending hospitals in the Parisian region are generally satisfied or not and to identify the factors contributing most to their satisfaction.

The univariate analysis revealed that the majority of patients are executives from businesses and middle-aged adults, with an average age of 57 years.

As for the results of the statistical tests, it is evident that:

- Both male and female patients recommend the services they have experienced to their acquaintances, indicating that healthcare personnel do not differentiate between genders in patient treatment.
- Patients are satisfied with the quality of the relationship with healthcare staff.
- The amount of information received from healthcare personnel is consistent regardless of the patient's profession.
- The quality of the relationship with the patient does not depend on the specific hospital department visited.
- The quality of the relationship is not influenced by the age of the patient.

These results demonstrate that the surveyed hospitals are generally fair in their treatment of patients. In conclusion, we commend the professionalism of healthcare personnel and their commitment to respecting human dignity.

Keywords: Inferential statistics; statistical test, satisfaction, evaluation, patient.

Introduction Générale

Selon l'Organisation Mondiale de la Santé : « L'évaluation de la qualité des soins est une démarche qui permet de garantir à chaque patient des actes diagnostiques et thérapeutiques assurant le meilleur résultat en termes de santé, conformément à l'état actuel de la science médicale, au meilleur coût pour le meilleur résultat, au moindre risque iatrogène¹ et pour sa plus grande satisfaction en termes de procédures, de résultats et de contacts humains à l'intérieur du système de soins »².

La satisfaction du patient est donc une composante essentielle de l'évaluation de la qualité des soins. La mesure de cette satisfaction s'inscrit dans un contexte global du positionnement de l'utilisateur dans l'organisation du système de santé et la montée des démarches qualité dans toutes les entreprises et services publics.

Dans cette optique, des données ont donc été collectées dans certains hôpitaux de la région parisienne auprès de 534 patients. L'objectif principal de cette collecte était de vérifier la satisfaction des patients au sujet de la qualité de la relation et de la quantité d'information reçues du personnel soignant.

Ainsi, au cœur de notre réflexion réside la question centrale de savoir si *les usagers des hôpitaux de la région parisienne son globalement satisfait du personnel de santé*. Cette interrogation constitue le fil conducteur de notre étude.

La réponse à cette problématique nécessite de répondre aux questions suivantes :

- La proportion de sujets recommandant le service dans lequel ils sont passés est-il le même quel que soit le genre ?
- La moyenne du score de relation est-il significativement différent chez les hommes et chez les femmes ?
- Le score de relation est-il significativement corrélé à l'âge ?
- Peut-on affirmer que la qualité de l'information reçue est la même quelle que soit la profession ?

¹ Se dit d'un trouble, d'une maladie provoqués par un acte médical ou par les médicaments, même en l'absence d'erreur du médecin. (larousse.fr)

² <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6211818/#cit0001>

- La qualité des relations avec le personnel soignant est-il fonction du service ayant accueilli le patient ?

Pour mener à bien ce travail, nous poserons des hypothèses (Hypothèse nulle et Alternative) pour chaque question, comme l'impose la méthodologie de la statistique inférentielle.

Ainsi, afin de ressourdre efficacement cette étude, nous l'avons érigé essentiellement en deux parties, d'une part, prétraitement et analyse descriptive du jeu de donnée et d'autres parts, la statistique inférentielle proprement dite.

PARTIE I : PRETRAITEMENT ET STATISTIQUE DESCRIPTIVE

I. PRETRAITEMENT DU JEU DE DONNEES

Introduction

Le nettoyage des données ou data cleansing est une étape essentielle de l'analyse statistique et la Data Science. Il s'agit de corriger ou supprimer des enregistrements inexacts dans des jeux de données afin de pouvoir les exploiter par la suite.

C'est un processus qui vise, de manière générale, à améliorer **la qualité des données**. En pratique, il consiste à **importer les données** dans un logiciel statistique, dans notre cas Jupyter Notebook du logiciel Python. Ensuite, faire une **première visualisation** pour avoir une idée succincte de la structure de nos données et enfin, identifier et corriger les valeurs manquantes, valeurs aberrantes, valeurs extrêmes afin de qu'elles puissent être plus cohérentes et sans erreurs.

1. Présentation des données

Les données utilisées dans ce rapport sont issues de la plateforme dédiée à la Data Science, **Kaggle**³.

Elles portent sur l'évaluation de la qualité de relation et la quantité d'information reçue par le patient lors de son séjour à l'hôpital. Les patients ont été enquêtés dans plusieurs hôpitaux de la région parisienne sur 9 variables récapitulées ci-dessous.

Figure 1: Informations initiales du jeu de données

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 534 entries, 0 to 533
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   service                              534 non-null    int64
1   sexe                                534 non-null    int64
2   age                                  528 non-null    float64
3   profession                           427 non-null    float64
4   amelioration.sante                   376 non-null    float64
5   amelioration.moral                   383 non-null    float64
6   recommander                          405 non-null    float64
7   score.relation                       349 non-null    float64
8   score.information                    358 non-null    float64
dtypes: float64(7), int64(2)
memory usage: 37.7 KB
```

³ <https://www.kaggle.com/datasets/samawel97/satisfaction-hospital/code>

Le dictionnaire des données se présente comme suit :

Tableau 1: Dictionnaire des données

Variables	Nature/Type	Description	Modalités
service	category (Qualitative)	Numéro des services ayant accueilli le patient	De 1 à 8
sexe	category (Qualitative)	Le genre du patient	homme et femme
age	int64 (Quantitative)	Age en année	Age compris entre 27 et 85
profession	category (Qualitative)	Le travail exercé par les patients enquêtés	'agriculteur', 'artisan', 'cadre', 'prof_intermédiaire', 'employé', 'ouvrier', 'sans emploi', 'autre'
<i>Source : Réalisé par l'auteur avec Python</i>			
amelioration.sante	category (Qualitative)	Les quantités d'articles vendues	Numérique
amelioration.moral	category (Qualitative)	Nombre de produits en promotion	Numérique
recommander	category (Qualitative)	Niveau de recommandation à l'entourage	0 : Non ; 1 : oui, probablement ; 2 : oui, sûrement
score.relation	int64 (Quantitative)	Notation de la qualité de la relation avec le personnel soignant pendant le séjour	De 10 à 40
ScoreInfo	int64 (Quantitative)	Notation de la qualité de l'information recue pendant le séjour	De 10 à 40

Source : Réalisé par l'auteur

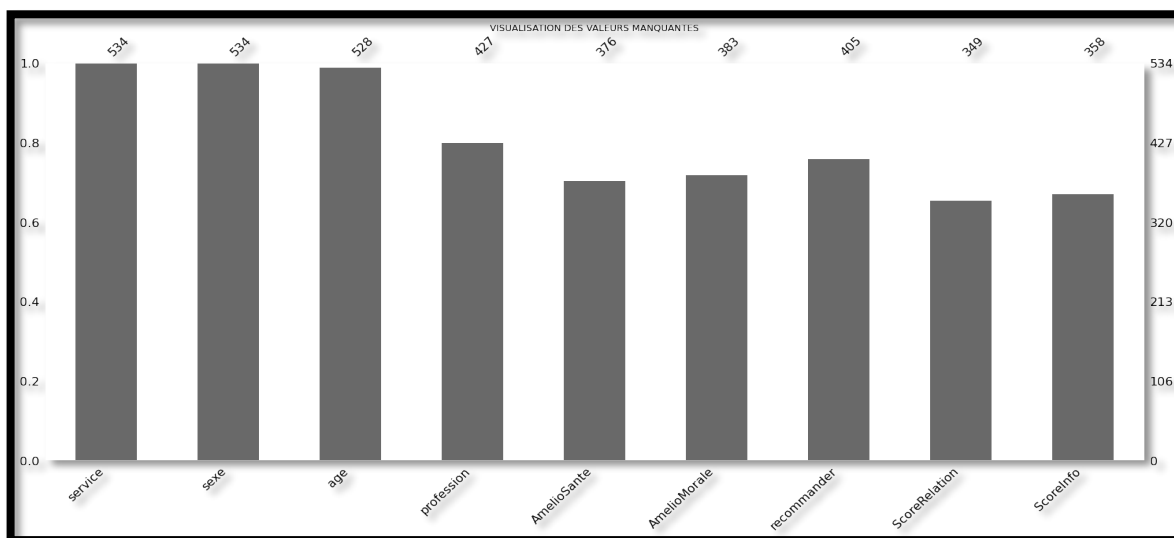
En résumé, la lecture des informations de bases de nos données nous montre des anomalies telles que : l'existence de valeurs manquantes et un mauvais typage de certaines variables. Dans la suite de notre analyse, nous allons traiter les valeurs manquantes, les valeurs aberrantes et recoder certaines variables si nécessaire.

2. Traitement des valeurs manquantes et doublons

➤ Traitement des valeurs manquantes

Nous allons visualiser les valeurs manquantes (NaN) grâce au diagramme des données manquantes avant et après traitement.

Figure 2: Visualisation des données manquantes avant traitement



Source : Réalisé par l'auteur avec Python

Les tailles des six dernières variables montrent la présence de valeurs manquantes dans celles-ci. Le tableau ci-dessous récapitule le nombre de NaN par colonne.

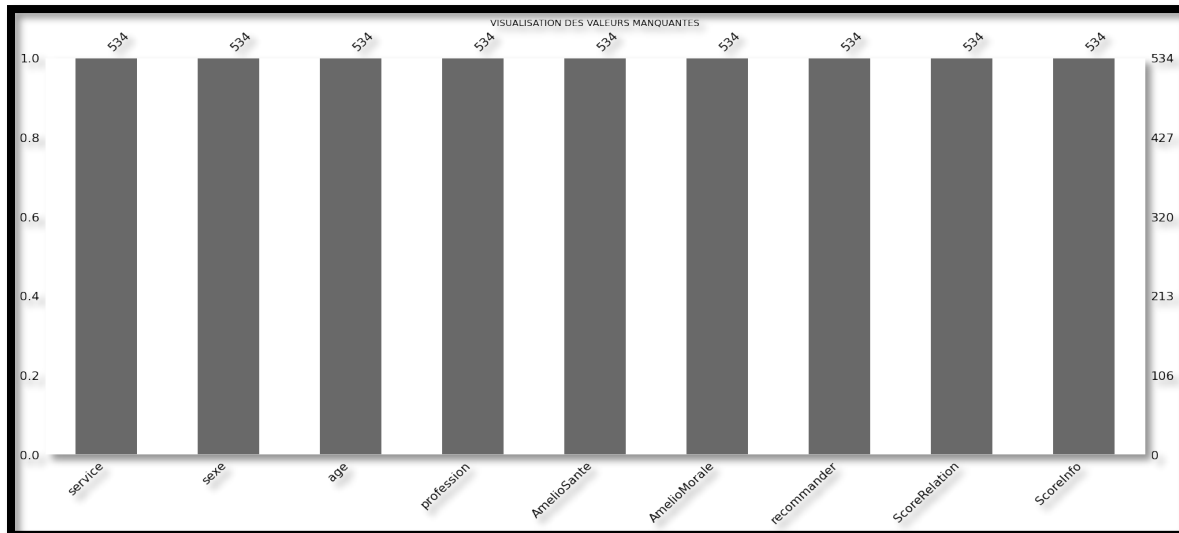
Tableau 2: Tableau des données manquantes par variables

Variables	Nombre de NaN
service	0
sexe	0
age	6
profession	107
AmelioSante	158
AmelioMorale	151
recommander	129
score.relation	185
ScoreInfo	176

Source : Réalisé par l'auteur

Nous pouvons pertinemment constater l'existence de données manquantes dans toutes les variables exceptées les colonnes **sexe** et **service**. Nous allons donc procéder par imputation avec la méthode des K-NN (K plus proches voisins).

Figure 3: Résultat après imputation



Source : Réalisé par l'auteur avec Python

Le diagramme ci-dessus nous permet de constater qu'effectivement l'imputation a fonctionné et que nos données sont complètes.

Nous pouvons donc passer à la vérification des doublons.

➤ Traitement des doublons

Selon la plateforme de formation en ligne DataScientest, la présence de données redondantes est un problème très fréquent dans la plupart des organisations.

Les doublons peuvent provenir d'une collecte d'informations identiques provenant de différentes sources, d'erreurs humaines, ou encore de données ajoutées au lieu d'être mises à jour.

Les doublons peuvent fausser tout type d'analyse de données, voire entraîner indirectement de mauvaises prises de décision.

Les données redondantes peuvent également coûter très cher à l'entreprise si elles sont nombreuses. Heureusement, il existe sur Python, par exemple, des fonctions (comme `drop_duplicates` de Pandas) permettant de se débarrasser de doublons très facilement. (DataScientest 2023)

Nous constatons la présence de 42 doublons dans les données soumises à notre analyse. La fonction `drop_duplicates` de python nous permet de nous en débarrasser.

En voici le résultat :

Figure 4: Traitement des doublons

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 492 entries, 0 to 533
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   service                492 non-null    category
1   sexe                  492 non-null    category
2   age                   492 non-null    int64
3   profession            492 non-null    category
4   amelioration.sante    492 non-null    category
5   amelioration.moral    492 non-null    category
6   recommander           492 non-null    category
7   score.relation        492 non-null    int64
8   ScoreInfo             492 non-null    int64
dtypes: category(6), int64(3)
memory usage: 19.6 KB
```

Source : Réalisé par l'auteur avec Python

Nous pouvons constater que l'effectif des données passe de 534 à 492 suite à la suppression des individus en double.

3. Traitement des valeurs aberrantes

En statistiques, une valeur aberrante est une valeur qui diffère grandement de la distribution d'une variable. Il s'agit d'une observation anormale, qui s'écarte de données par ailleurs bien structurées.

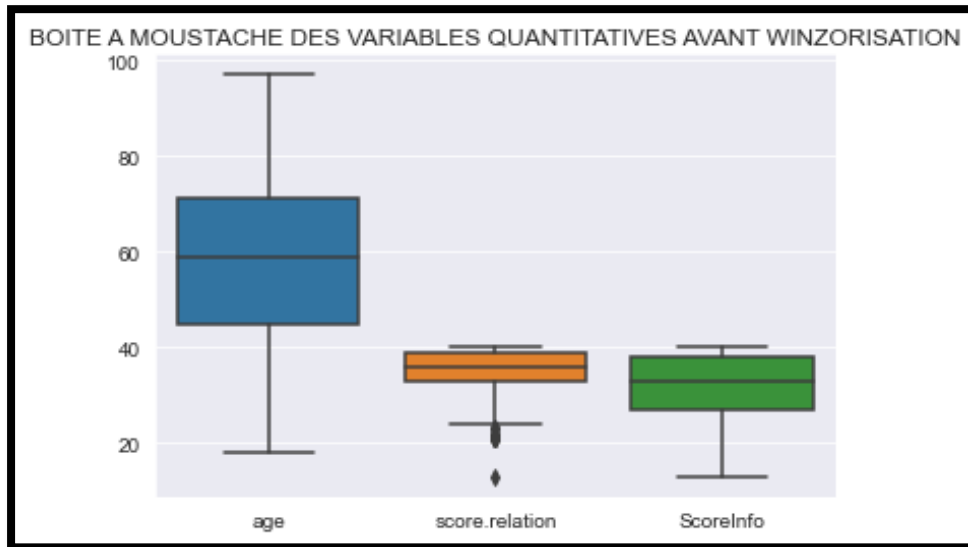
La détection des valeurs aberrantes ou des anomalies est l'un des problèmes fondamentaux de l'exploration des données. L'expansion émergente et continue des données nous fait repenser la façon dont nous abordons les anomalies et les cas d'utilisation qui peuvent être construits en examinant ces anomalies.

En Data Science, les valeurs aberrantes peuvent affecter certains paramètres statistiques, comme la moyenne. Si les Outliers ne sont pas détectés, cela peut fausser notre compréhension d'un jeu de données et nous conduire à émettre des hypothèses erronées sur ce dernier.

Une autre raison pour laquelle il est important de s'intéresser aux valeurs aberrantes est qu'une majorité des algorithmes de Machine Learning sont très sensibles aux données sur lesquelles ils sont entraînés, ainsi qu'à leurs distributions.

Avoir des Outliers dans le jeu d'entraînement d'un modèle de Machine Learning peut rendre la phase d'entraînement plus longue et potentiellement biaisée.

Par conséquent, le modèle de prédiction produit sera moins performant ou moins précis. (DataScientest 2023)

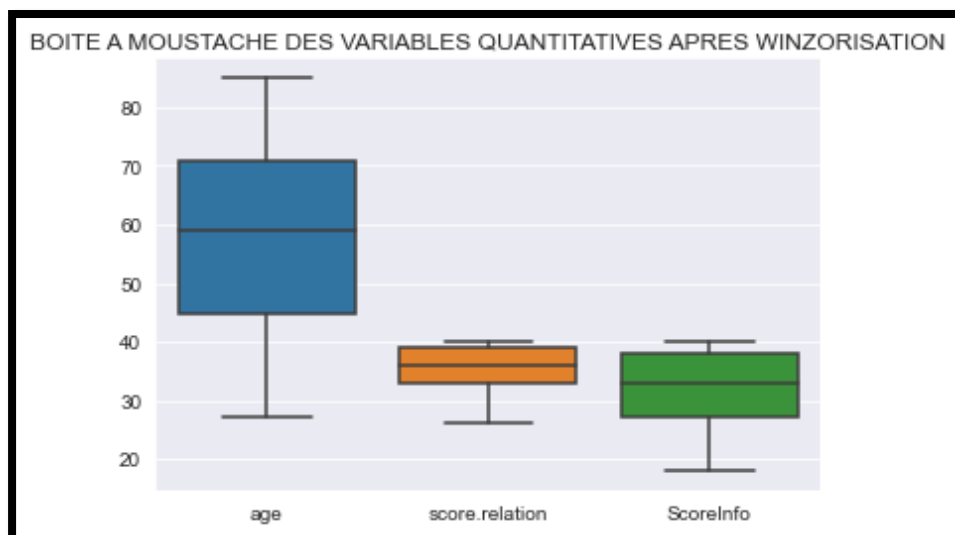
Figure 5: Boîte à moustache avant Winzoration

Source : Réalisé par l'auteur avec Python

Les boîtes à moustache ci-dessous permettent de constater que la variable **score.relation** est la seule qui présente des valeurs aberrantes au niveau de la moustache inférieure.

La technique de Winzoration⁴ consiste à remplacer les valeurs extrêmes par des valeurs moins extrêmes afin de réduire leur impact sur les analyses statistiques.

La figure ci-après nous montre qu'effectivement nos données sont maintenant exemptes de valeurs aberrantes.

Figure 6: Boîte à moustache après Winzoration

Source : Réalisé par l'auteur avec Python

⁴ Une technique de transformation des données utilisée en statistiques pour atténuer l'impact des valeurs extrêmes, des valeurs aberrantes ou des points de données influents dans un ensemble de donnée.

Conclusion

En conclusion, Cette première phase de prétraitement a permis de traiter les valeurs manquantes, qui étaient relativement élevées dans notre jeu de données, ainsi que les doublons qui étaient au nombre de 42. De plus, les valeurs extrêmes et aberrantes ont été découvertes au niveau de la variable **score.relation** et ont été ramenées à la borne inférieur.

Ce travail préalable a permis de rendre nos données beaucoup plus aptes à l'analyse. La seconde phase consistera à faire l'analyse descriptive univarié de nos variables d'intérêt.

II. STATISTIQUE DESCRIPTIVE UNIVARIÉE

Introduction

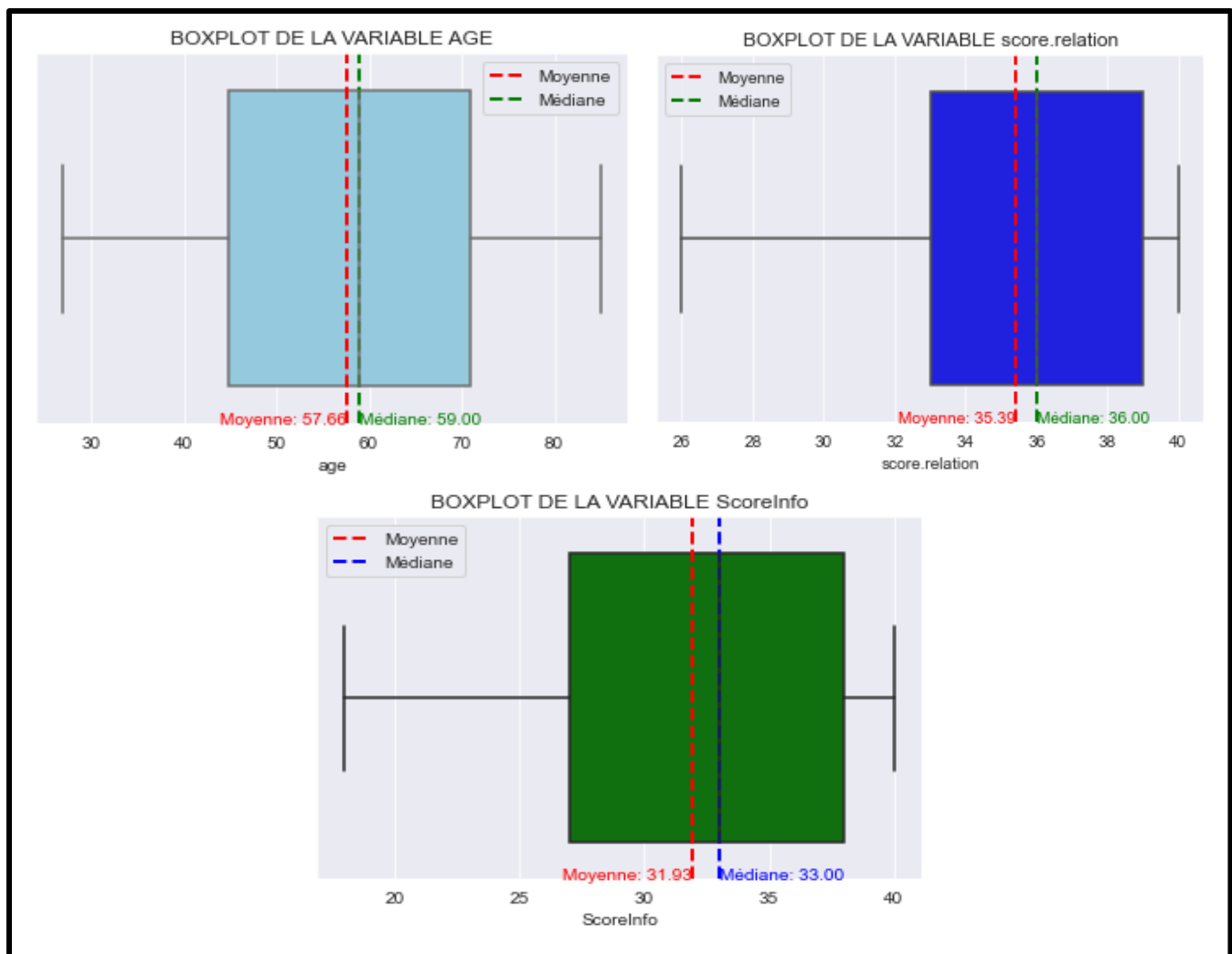
Selon Bernard PY, dans son livre *Statistique descriptive, nouvelle méthode pour bien comprendre et réussir* (éditions Economica) : « La statistique [descriptive] est un ensemble de méthodes permettant de décrire et d'analyser, de façon quantifiée, des phénomènes repérés par des éléments nombreux, de même nature, susceptibles d'être dénombrés et classés. » (MAZEROLLE 2006).

Cette partie nous permettra de mieux cerner les spécificités de nos données grâce aux différents indicateurs statistiques et aux graphiques.

1. Analyse univariée des variables quantitatives

Les variables quantitatives sont au nombre de trois dans notre jeu de données, ce sont les variables : **sexe**, **ScoreInfo** et **score.relation**. Visualisons tout d'abord quelques graphiques et appuyons cela par des indicateurs statistiques.

Figure 7: Boxplots des variables quantitatives



Source : Réalisé par l'auteur avec Python

La visualisation des graphiques ci-dessus met en exergue les boîtes à moustache des trois variables avec les principales caractéristiques qui sont les moyennes et les médianes.

Nous constatons que pour ces trois variables, les médianes sont relativement supérieures par rapport aux moyennes. Ce qui indique que les moyennes sont beaucoup influencées pour la queue inférieure de la distribution, c'est-à-dire par les petites valeurs qui sont dans la distribution.

Le tableau ci-dessous nous permettra de clarifier cette situation.

Tableau 3: Résumé statistiques

Indicateurs	age	score.relation	ScoreInfo
INDICATEUR DE TENDANCE CENTRALE ET DE POSITION			
Minimum	27	26	18
Maximum	85	40	40
Moyenne	57.66	35.39	31.93
Ecart-type	16.99	4.22	6.64
Mode	27	40	40
INDICATEUR DE FORME			
Skewness	-0.15	-0.77	-0.54
Kurtosis	-0.99	-0.46	-0.77

Source : Réalisé par l'auteur avec Python

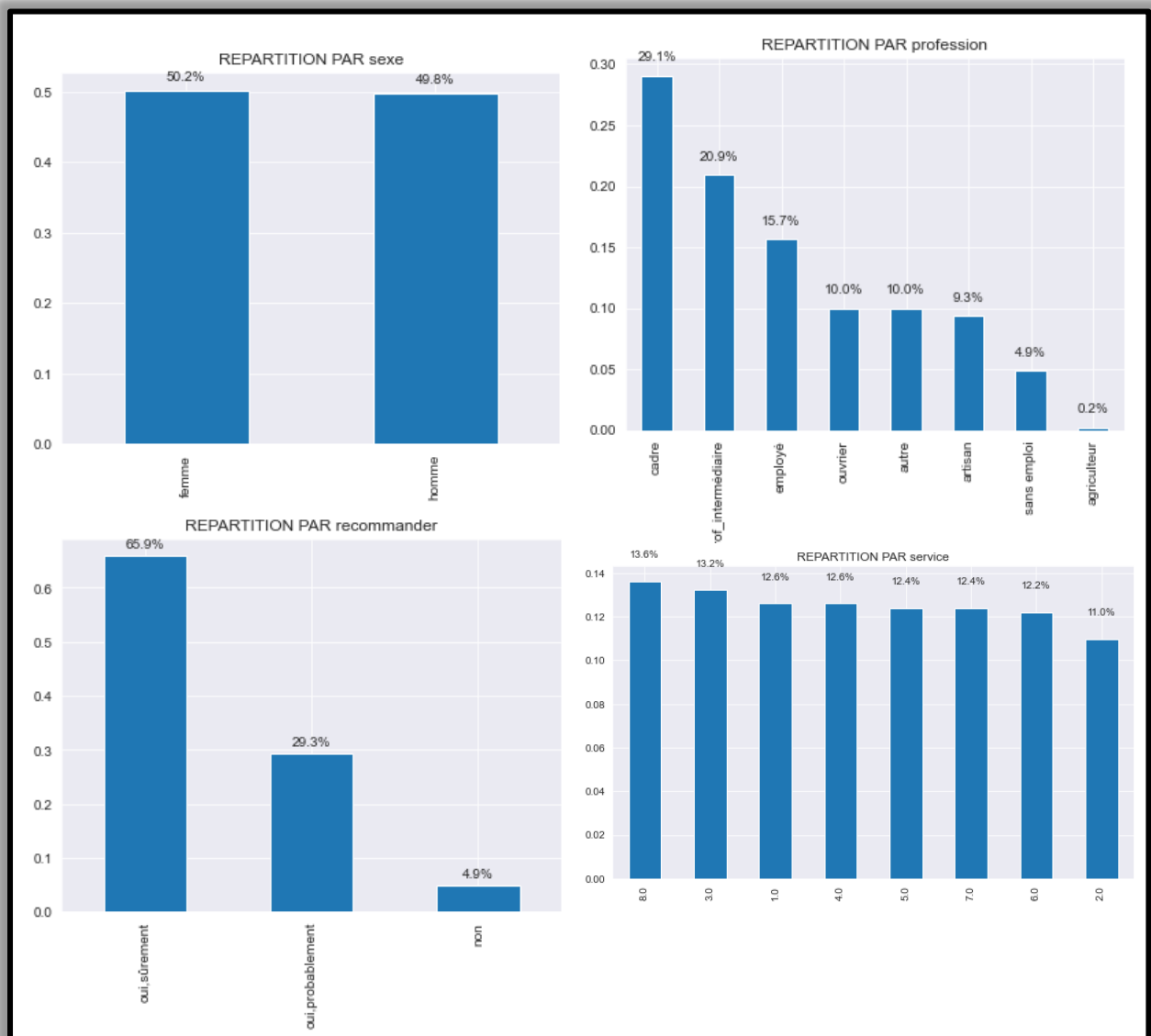
Le tableau ci-dessus vient conforter l'analyse graphique. Il nous permet d'affirmer **qu'en moyenne les patients sont âgés d'environ 58 ans avec un écart type de 17 ans**, cependant, l'âge minimal et modal est de 27 ans.

L'analyse des indicateurs de tendance centrale des deux autres variables montre que **les scores moyens** attribués par les patients concernant la qualité de la relation avec le personnel soignant et la qualité de l'information reçue sont respectivement **d'environ 35 et de 32** sur une échelle de 10 à 40 ce qui montre qu'ils sont **globalement satisfaits**.

Les valeurs des indicateurs de forme montrent que les distributions sont pour la plupart étalées à gauche et platikurtique.

2. Analyse univariée des variables qualitatives

Figure 8: Diagramme en barre des variables qualitatives



Source : Réalisé par l'auteur sous Python

Les graphiques ci-dessus met en évidence que 50,2% des enquêtés sont de sexe masculin. Pratiquement **66% des patients recommandent les services** par lesquels ils sont passés durant le séjour à l'hôpital, seulement 5% environ ne recommandent pas.

De plus, nous constatons que la majorité des patients sont des cadres (29%) suivi des professions intermédiaires (20%) et en dernière position les agriculteurs (0,2%).

Figure 9: Tableau des effectifs des variables qualitatives

L'effectif par sexe est:

femme 247

homme 245

Name: sexe, dtype: int64

L'effectif par profession est:

cadre 143

prof_intermédiaire 103

employé 77

ouvrier 49

autre 49

artisan 46

sans emploi 24

agriculteur 1

Name: profession, dtype: int64

L'effectif par recommander est:

oui,sûrement 324

oui,probablement 144

non 24

Name: recommander, dtype: int64

L'effectif par service est:

8.0 67

3.0 65

1.0 62

4.0 62

5.0 61

7.0 61

6.0 60

2.0 54

Name: service, dtype: int64

Source : Réalisé par l'auteur sous Python

Conclusion

Au terme de cette première partie, il ressort que les patients ont un profil de cadre d'entreprise de 57 ans en moyenne. Ils ont globalement apprécié la relation avec le personnel soignant en leur attribuant un score de 35/40 en moyenne. De plus, ces patients ont jugé avoir reçu le maximum d'information durant leur passage à l'hôpital.

La partie suivante nous permettra de tester la pertinence des résultats de ces analyses simplement descriptive.

PARTIE II : STATISTIQUE INFÉRENTIELLE

I- TEST DE COMPARAISON DE DEUX POPULATIONS INDEPENDANTES

Introduction

L'étude présentée s'inscrit dans le cadre du test de comparaison de deux populations indépendantes, une méthodologie statistique puissante visant à évaluer les différences significatives entre deux groupes distincts au sein d'une population. Notre analyse se focalise sur la relation entre le personnel soignant des hôpitaux de la région parisienne et les patients, en explorant les nuances de cette interaction en fonction du genre. Le choix de ce test vise à fournir des insights précieux sur la dynamique des relations entre le personnel soignant et les patients masculins et féminins, tout en permettant une évaluation approfondie des facteurs contribuant à la satisfaction globale des patients.

1- La proportion de sujets recommandant le service dans lequel ils sont passés est-il le même quel que soit le genre ?

Cette question nous emmène à vérifier si les avis des patients sont différents selon qu'ils soient de sexe masculin ou de sexe féminin. En d'autres termes, les avis de recommandation dépendent-ils du genre de la personne qui recommande ?

Il s'agira donc de tester s'il y a égalité significative de proportion entre deux groupes, homme et femme qui recommandent les services par lesquels ils sont passés.

➤ Formulation des hypothèses

Les hypothèses statistiques sont formulées comme suit :

H_0 : la proportion de femmes ayant recommandée le service par lequel elles sont passées est la même que celle des hommes.

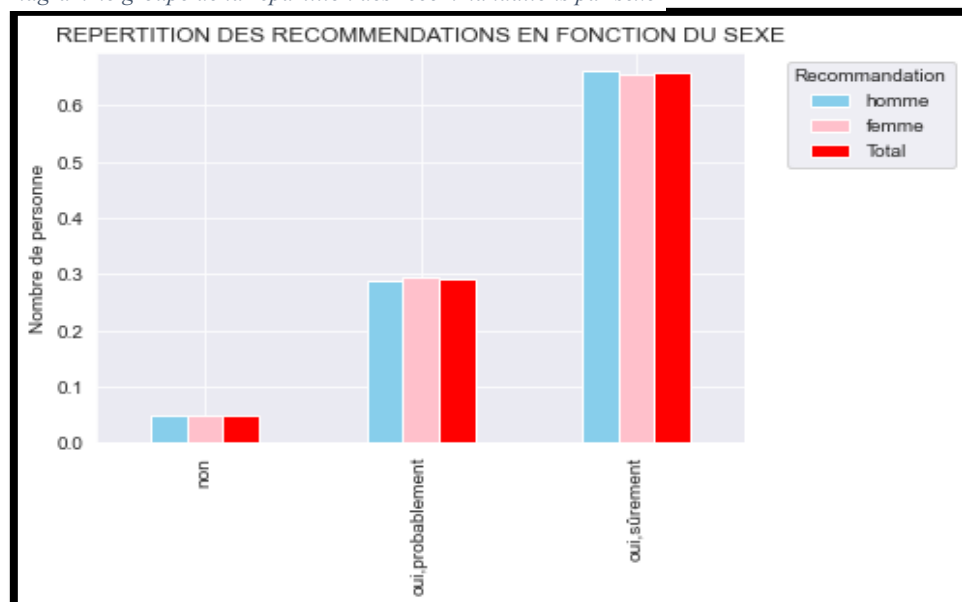
H_1 : la proportion de femme ayant recommandée le service par lequel elles sont passées n'est pas la même que celle des hommes.

La vérification de ces hypothèses se fait sous le logiciel Python.

➤ Vérification des hypothèses

Cette étape consiste à calculer des indicateurs et faire des graphiques pour les deux sous-groupes afin de faire le test statistique qui permet de trancher.

Figure 10: Diagramme groupé de la repartition des recommandations par sexe



Source : Réalisé par l'auteur sous Python

Ce graphique nous montre clairement que les proportions de recommandation chez les hommes et chez les femmes ne varient pas grandement. Les bars des diagrammes ont pratiquement les mêmes tailles quelques soient les genres.

Tableau 4: Tableau de contingence des effectifs empiriques des hommes et des femmes

	Homme	Femme	Total	Homme	Femme	Total
Effectif empirique				Effectif théorique		
non	12	12	24	11.95	12.04	24
oui, probablement	71	73	144	71.70	72.29	144
oui, sûrement	162	162	324	161.34	162.65	324
Total	245	247	492	245	247	492

Source : Réalisé par l'auteur avec Python

Les tableaux ci-dessus peuvent nous permettre d'affirmer que toutes les valeurs théoriques sont supérieures à 5 et qu'il n'y a pas grande différence entre effectif observé dans l'échantillon et l'effectif calculé du test de Khy-deux. Nous soupçonnons donc une indépendance des réponses des femmes de celles des hommes.

➤ Décision

Suite à l'implémentation du test de Khy-deux, il ressort que les deux populations sont statistiquement indépendantes.

Figure 11: Résultat du test de Khy-deux

```

Statistique de test du Chi-2 : 0.019648021151780368
P-value : 0.9999998431385899
Degrés de liberté : 6
Tableau des effectifs attendus :
[[ 11.95121951  12.04878049  24.          ]
 [ 71.70731707  72.29268293 144.          ]
 [161.34146341 162.65853659 324.          ]
 [245.          247.          492.          ]]
On ne rejette pas l'hypothèse nulle. Les variables sont indépendantes.

```

Source : Réalisé par l'auteur sous Python

➤ Conclusion

Le test de Khy-deux nous permet de conclure que nous acceptons l'hypothèse nulle qui stipule que **la proportion de femme qui recommande le service où elle a séjourné n'est pas statistiquement différente de celle des hommes**. En d'autres termes, les recommandations ne varient pas en fonction du genre des patients.

Cela voudrait dire que les patients apprécient les services du personnel soignant de ces différents hôpitaux sont globalement corrects.

2- La moyenne du score relation est-il significativement différent chez les hommes et chez les femmes ?

La question suggère que nous souhaitons déterminer s'il existe une différence significative dans les scores de relation entre deux groupes, à savoir les hommes et les femmes. En d'autres termes, le score relatif à la qualité de la relation avec le personnel soignant lors du séjour est-il significativement différent d'un sexe à l'autre ?

Cela peut être interprété comme une comparaison des moyennes des scores de relation entre ces deux groupes.

➤ Formulation des hypothèses

Les hypothèses statistiques sont formulées comme suit :

H_0 : la moyenne du score relation chez les hommes est statistiquement égale à celle des femmes.

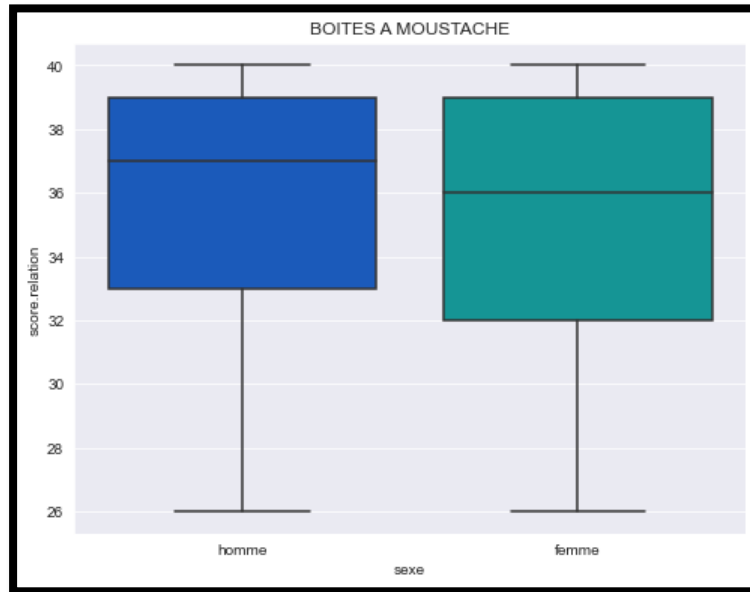
H_1 : la moyenne du score relation chez les hommes est statistiquement différente de celle des femmes.

La vérification de ces hypothèses se fait sous le logiciel Python.

➤ Vérification des hypothèses

• Etape 1 : analyse graphique des sous-populations

Figure 12: Boîte à Moustache des sous-groupes



Source : Réalisé par l'auteur sous Python

Les formes des boîtes à moustache des scores relations chez les hommes et chez les femmes ne semblent pas varier grandement même si la moyenne chez les femmes est un peu plus basse que celle des hommes.

La suite de notre analyse nous permettra de clarifier ce fait.

• Etape 2 : Statistiques de base des sous populations

Figure 13: Résumé statistiques du score relation par genre

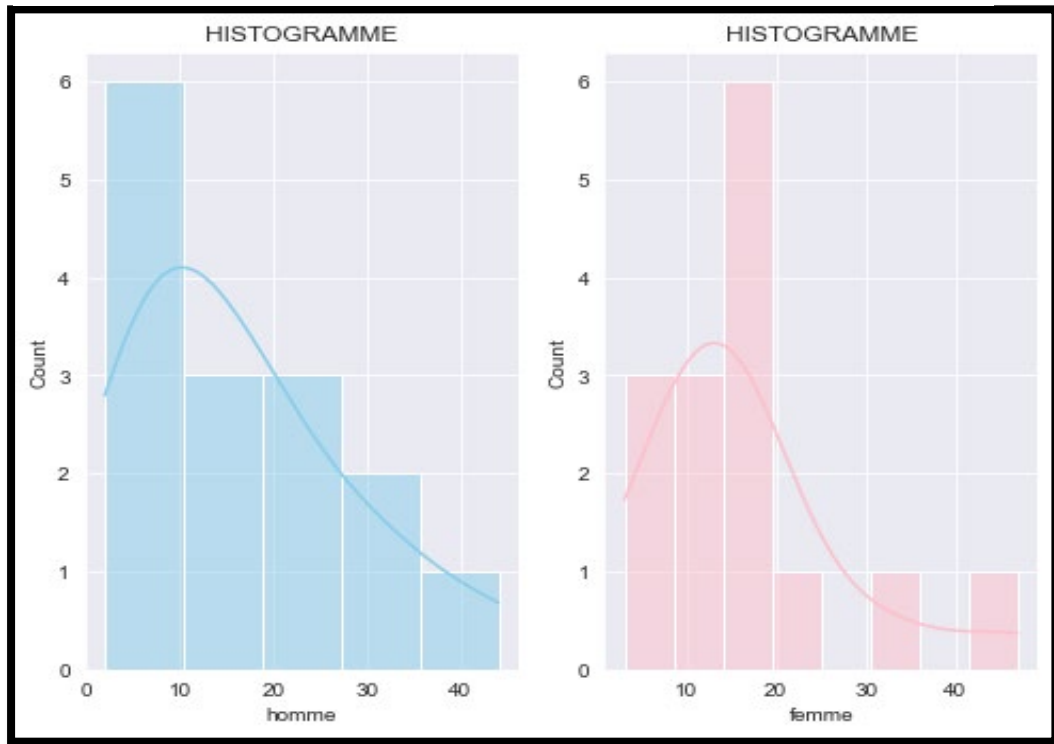
	count	mean	std	min	25%	50%	75%	max
sexe								
homme	245.0	35.555102	4.138239	26.0	33.0	37.0	39.0	40.0
femme	247.0	35.226721	4.302251	26.0	32.0	36.0	39.0	40.0

Source : Réalisé par l'auteur sous Python

Ce tableau présente une grande similitude entre les deux sous populations, les caractéristiques de tendance centrale et les effectifs sont très voisines.

• Etape 3 : Test de normalité dans les sous populations

Figure 14: Histogramme du score relation par genre



Source : Réalisé par l'auteur sous Python

Les formes des histogrammes ci-dessus ne sont pas semblables à la forme typique en cloche de la distribution gaussienne. Elles sont plus étalées à droite, par conséquent, nos deux sous populations ne semblent pas suivre la loi normale.

Nous utiliserons le test de Shapiro-Wilk pour vérifier cette assertion.

Figure 15: Résultats du test de Shapiro

```
Shapiro test pour la modalité *homme*: Statistic=0.88098, P-value=0.00000
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec l'un des tests non-parametriques de wilcox.test ou kruskal.test

Shapiro test pour la modalité *femme*: Statistic=0.89627, P-value=0.00000
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec l'un des tests non-parametriques de wilcox.test ou kruskal.test
```

Source : Réalisé par l'auteur sous Python

Le test de Shapiro Wilk confirme **la non-normalité des distributions des deux modalités** : homme et femme.

Le test d'égalité des proportions se fera donc grâce au test non-paramétrique des rangs de Wilcoxon ou test U de Mann-Whitney.

➤ Décision

L'implémentation du test de Wilcoxon nous donne le résultat ci-dessous :

Figure 16: Résultat du test de Wilcoxon

```
Statistique U de Mann-Whitney : 106.0  
P-valeur : 0.8032050994203768  
Ne pas rejeter l'hypothèse nulle :  
Il n'y a pas de différence significative entre les échantillons.
```

Source : Réalisé par l'auteur sous Python

Nous constatons clairement que le test des rangs de Wilcoxon autorise à accepter l'hypothèse nulle avec une P-value de 0.80 qui est largement supérieur à 5%.

➤ Conclusion du test

À un niveau de signification de 5%, les résultats suggèrent qu'il n'y a pas de différence significative entre les hommes et les femmes en ce qui concerne le score moyen de relation.

En d'autres termes, sur la base de l'échantillon de données que nous avons, il n'y a pas suffisamment de preuves pour affirmer que la moyenne du score de relation avec le personnel soignant est significativement différente entre les hommes et les femmes.

Par conséquent, nous pouvons affirmer qu'**en moyenne la relation entre le personnel soignant et les patients est la même quel que soit le sexe.**

Conclusion

En somme, l'attitude du personnel soignant au regard des patients, qu'ils soient de sexe masculin ou féminin, est très acceptable sur la base de notre échantillon.

Ces résultats prouvent que le personnel soignant joue bien son rôle en ce qui concerne la qualité de la relation avec les patients. Les patients étant satisfaits, recommandent alors les services par lesquels ils sont passés.

II- TEST DE COMPARAISON DE PLUSIEURS POPULATIONS INDEPENDANTES

Introduction

L'analyse de notre étude se déploie également dans le cadre du test de comparaison de plusieurs populations indépendantes, une méthodologie statistique éclairante visant à évaluer les distinctions significatives entre divers groupes au sein d'une population donnée. Notre examen s'est concentré sur le personnel soignant des hôpitaux de la région parisienne, en évaluant son attitude envers les patients en fonction de critères tels que l'âge, la profession, et le genre. Ce test de comparaison de plusieurs populations indépendantes offre une perspective riche et nuancée sur les variations potentielles dans l'interaction entre le personnel soignant et les patients, permettant ainsi une compréhension approfondie des dynamiques interpersonnelles au sein du contexte hospitalier.

1- Peut-on affirmer que la qualité de l'information reçue est la même quelle que soit la profession ?

La présente étude vise à déterminer si la qualité de l'information perçue varie en fonction de la profession des individus. L'objectif est d'analyser les différences éventuelles dans la perception de la qualité de l'information entre divers groupes professionnels.

Nous examinerons si les patients de domaines différents ont des évaluations similaires de la qualité de l'information reçue ou s'il existe des disparités significatives. Cette analyse contribuera à éclairer la relation potentielle entre la profession des individus et leur perception de la qualité de l'information reçue du personnel soignant.

➤ Formulation des hypothèses

Les hypothèses statistiques sont formulées comme suit :

H_0 : la qualité de l'information reçue est la même selon les professions

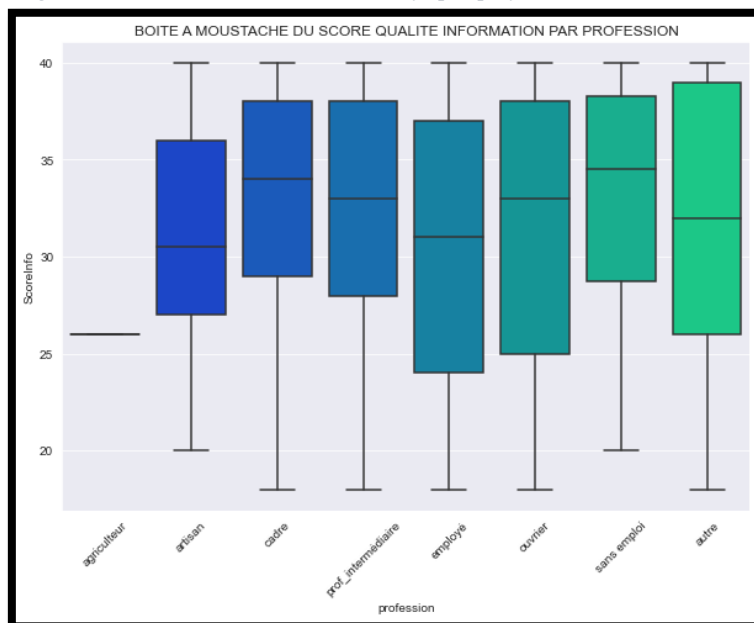
H_1 : la qualité de l'information reçue est significativement différente selon les professions

La vérification de ces hypothèses se fait sous le logiciel Python.

➤ Vérification des hypothèses

• Etape 1 : analyse graphique des sous-populations

Figure 17: Boîte à Moustache du ScoreInfo par profession



Source : Réalisé par l'auteur sous Python

Les formes des boîtes à moustache des scores informations (notation de la qualité de l'information) par profession ne semblent pas varier grandement même si les moyennes représentées sur chaque boîte varient selon les différentes professions.

La suite de notre analyse nous permettra de clarifier ce fait.

• Etape 2 : Statistiques de base des sous populations

Figure 18: Résumé statistiques du ScoreInfo par profession

profession	count	mean	std	min	25%	50%	75%	max
agriculteur	1.0	26.000000	NaN	26.0	26.00	26.0	26.00	26.0
artisan	46.0	31.239130	5.839272	20.0	27.00	30.5	36.00	40.0
cadre	143.0	33.195804	5.957902	18.0	29.00	34.0	38.00	40.0
prof_intermediaire	103.0	32.029126	6.409944	18.0	28.00	33.0	38.00	40.0
employé	77.0	30.103896	7.602950	18.0	24.00	31.0	37.00	40.0
ouvrier	49.0	31.693878	6.907737	18.0	25.00	33.0	38.00	40.0
sans emploi	24.0	33.708333	5.614261	20.0	28.75	34.5	38.25	40.0
autre	49.0	31.040816	7.664744	18.0	26.00	32.0	39.00	40.0

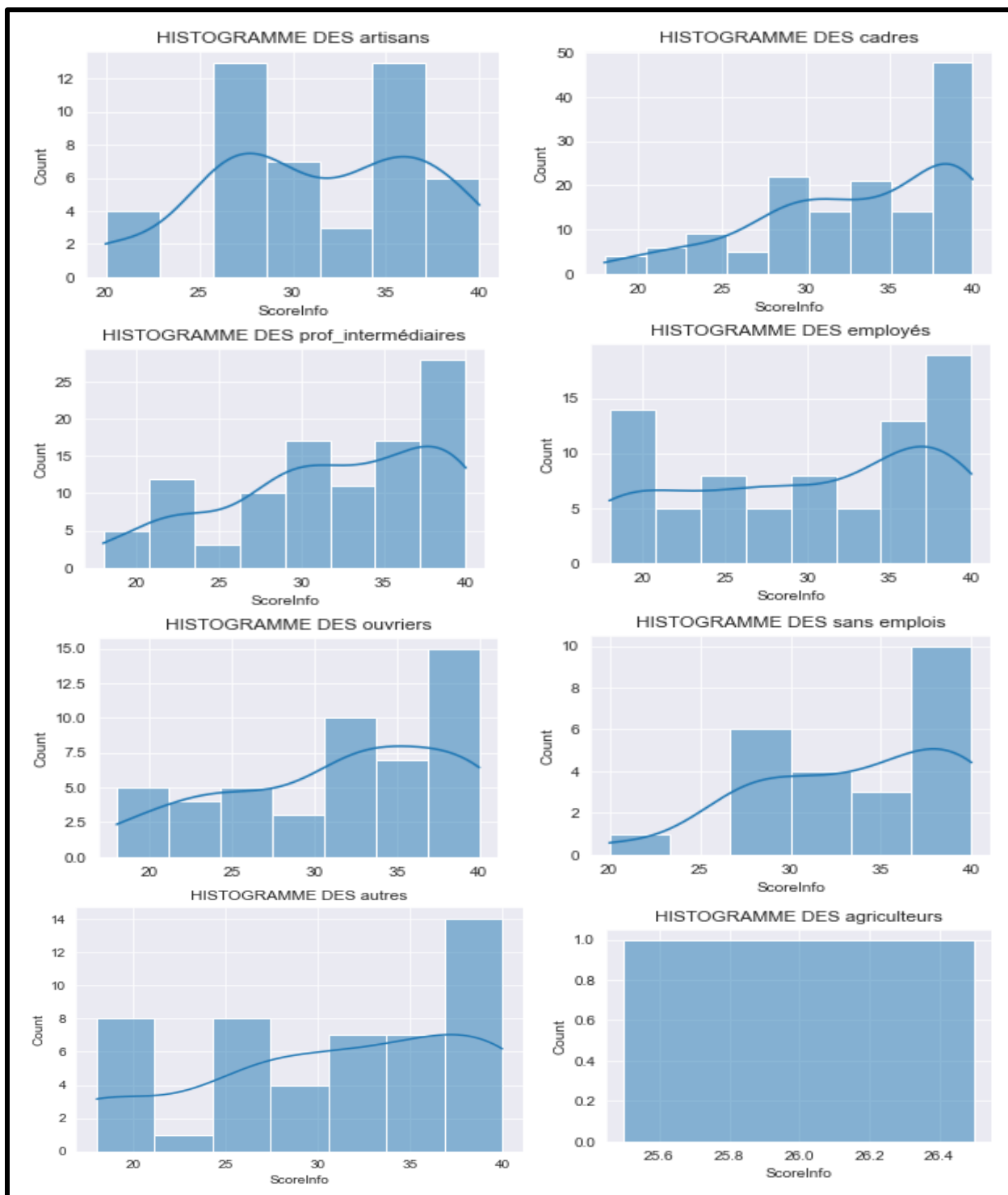
Source : Réalisé par l'auteur sous Python

Ce tableau présente une grande similitude entre ces sous populations, les caractéristiques de tendance centrale et les effectifs sont très voisines.

• Etape 3 : Test de normalité dans les sous populations

Les formes des histogrammes ci-après ne sont pas semblables à la forme typique en cloche de la distribution gaussienne. Elles sont de formes assez irrégulières, par conséquent, nos sous populations ne semblent pas suivre la loi normale. Nous utiliserons le test de Shapiro-Wilk pour vérifier cette assertion.

Figure 19: Histogramme du score relation par profession



Source : Réalisé par l'auteur sous Python

Figure 20: Résultats du test de Shapiro

```

Shapiro test pour la modalité **prof_intermédiaire**: Statistic=0.92520, P-value=0.00002
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **autre**: Statistic=0.89010, P-value=0.00027
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **artisan**: Statistic=0.93258, P-value=0.01044
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **cadre**: Statistic=0.91345, P-value=0.00000
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **employé**: Statistic=0.90044, P-value=0.00002
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **ouvrier**: Statistic=0.91493, P-value=0.00175
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité **sans emploi**: Statistic=0.90585, P-value=0.02869
L'échantillon ne semble pas suivre une distribution normale (on rejette H0),
alors on continue avec le test non-paramétrique de kruskal.test

Le groupe agriculteur a moins de 3 observations,
le test de Shapiro ne peut pas être effectué.

```

Source : Réalisé par l'auteur sous Python

Le résultat du test de Shapiro Wilk ci-dessus confirme **la non-normalité des distributions des six modalités des professions**.

Le test d'égalité des proportions se fera donc grâce au test non-paramétrique de Kruskal Wallis.

➤ Décision

L'implémentation du test de Kruskal Wallis nous donne le résultat ci-dessous :

Figure 21: Résultat du test de Kruskal Walli

```

Statistique de Kruskal : 12.929684972425948
P-valeur : 0.07384024953504864
Ne pas rejeter l'hypothèse nulle :
Il n'y a pas de différence significative entre les groupes.

```

Source : Réalisé par l'auteur sous Python

Nous constatons clairement que le test de Kruskal-Wallis nous autorise à accepter l'hypothèse nulle avec une P-value de 0.07 qui est largement supérieur à 5%.

➤ Conclusion

La p-value résultante du test de Kruskal-Wallis est supérieure au niveau de signification choisi ($0.07 > 0.05$ dans notre cas), cela suggère que nous ne pouvons pas rejeter l'hypothèse nulle.

Ce qui signifie que, sur la base du test global, il n'y a pas suffisamment de preuves pour conclure qu'il y a une différence significative entre les groupes.

Par conséquent, **la qualité de l'information reçue est la même selon les professions.** En d'autres termes, le personnel soignant ne fait pas de distinction selon la profession quant à l'information donnée aux patients.

2- La qualité des relations avec le personnel soignant est-il fonction du service ayant accueilli le patient ?

Cette étude cherche à déterminer si la qualité des relations avec le personnel soignant varie en fonction du service dans lequel le patient a été accueilli.

Il s'agira donc, de tester les éventuelles différences dans la qualité des relations entre les patients et le personnel soignant au sein de différents services.

Nous chercherons à établir si le service d'accueil du patient a une incidence significative sur la perception de la qualité des relations avec le personnel soignant. Cette analyse contribuera à approfondir notre compréhension des facteurs influençant la qualité des interactions entre les patients et le personnel soignant.

➤ Formulation des hypothèses

Les hypothèses statistiques sont formulées comme suit :

H_0 : la qualité de la relation avec les patients est la même quel que soit le service.

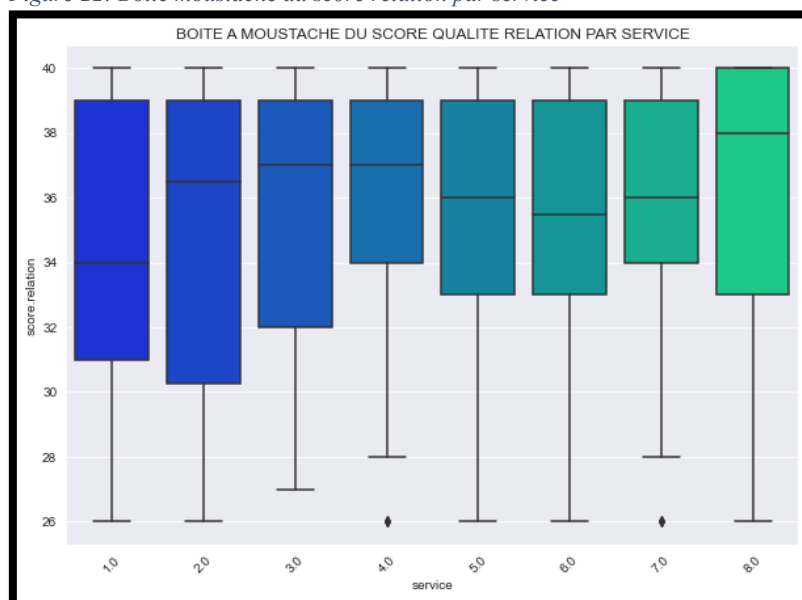
H_1 : la qualité de la relation avec les patients est significativement différente selon le service.

La vérification de ces hypothèses se fait sous le logiciel Python.

➤ Vérification des hypothèses

- **Etape 1 : analyse graphique des sous-populations**

Figure 22: Boîte moustache du score relation par service



Source : Réalisé par l'auteur sous Python

Les formes des boîtes à moustache des scores relations (notation de la qualité de la relation avec le personnel soignant) par service ne semblent pas varier grandement même si les moyennes représentées sur chaque boîte varient selon les différents services.

La suite de notre analyse nous permettra de mieux comprendre cet état de fait.

- **Etape 2 : Statistiques de base des sous-populations**

Figure 23: Résumé statistiques du score.relation par service

	count	mean	std	min	25%	50%	75%	max
service								
1.0	62.0	34.354839	4.838699	26.0	31.00	34.0	39.0	40.0
2.0	54.0	34.407407	5.045149	26.0	30.25	36.5	39.0	40.0
3.0	65.0	35.769231	4.006905	27.0	32.00	37.0	39.0	40.0
4.0	62.0	36.129032	3.721818	26.0	34.00	37.0	39.0	40.0
5.0	61.0	35.229508	4.124696	26.0	33.00	36.0	39.0	40.0
6.0	60.0	35.016667	4.184279	26.0	33.00	35.5	39.0	40.0
7.0	61.0	35.868852	3.442264	26.0	34.00	36.0	39.0	40.0
8.0	67.0	36.134328	4.096915	26.0	33.00	38.0	40.0	40.0

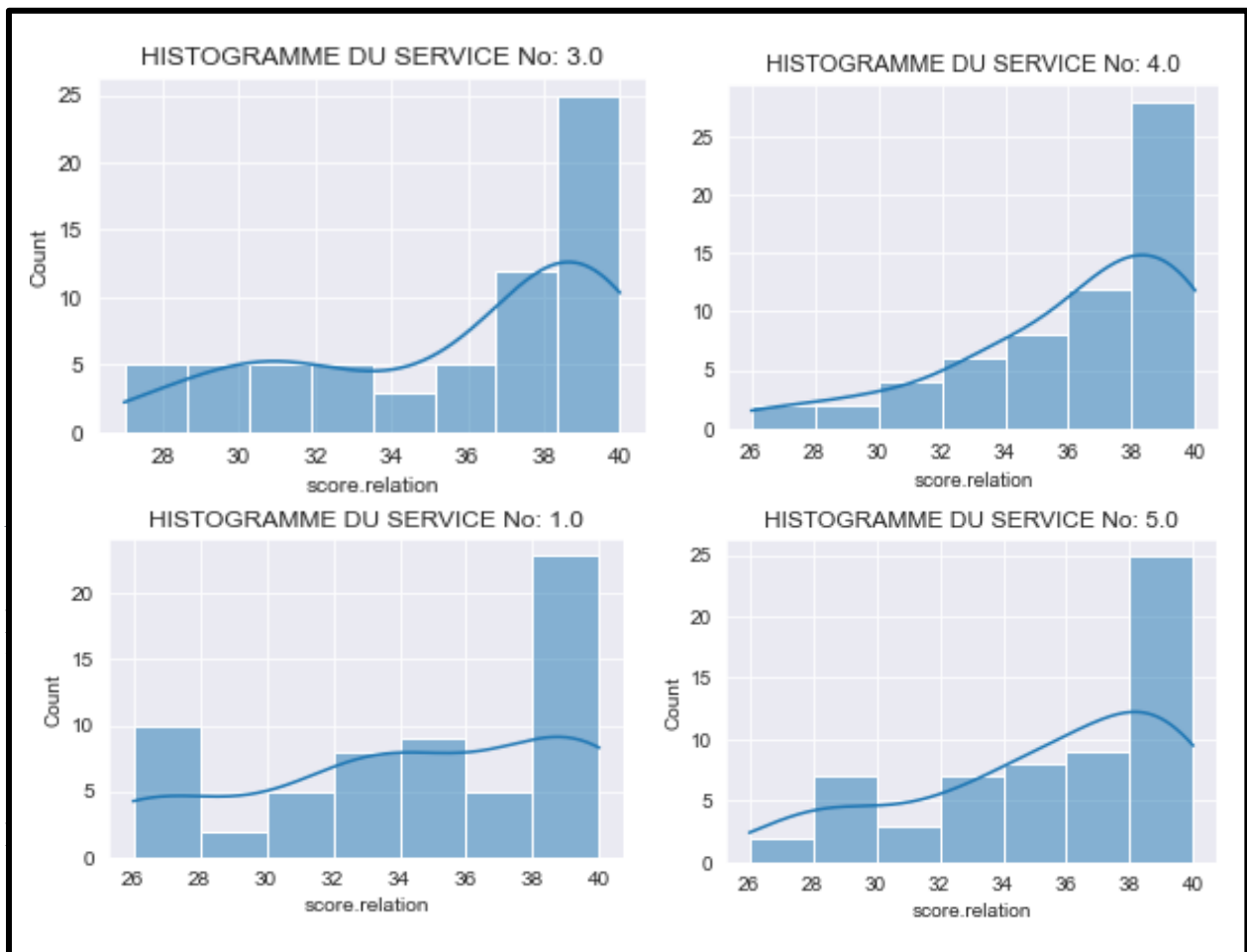
Source : Réalisé par l'auteur sous Python

Ce tableau présente une grande similitude entre ces sous populations, les caractéristiques de tendance centrale comme la moyenne du score qui tourne autour de 34/40 avec un écart type de score de 4 et les effectifs sont très voisines, en moyenne 60 individus par service.

- **Etape 3 : Test de normalité dans les sous populations**

Les formes des histogrammes ci-après ne sont pas semblables à la forme typique en cloche de la distribution gaussienne. Elles sont de formes assez irrégulières, par conséquent, nos sous populations ne semblent pas suivre la loi normale. Cette observation est confirmée par le test de Shapiro Wilk.

Figure 24: Histogrammes du ScoreInfo par service fréquenté



Source : Réalisé par l'auteur sous Python

Figure 27: Résultats des tests de Shapiro Wilk

Shapiro test pour la modalité -3.0-: Statistic=0.85657, P-value=0.00000
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -4.0-: Statistic=0.87742, P-value=0.00002
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -1.0-: Statistic=0.89332, P-value=0.00006
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -5.0-: Statistic=0.89909, P-value=0.00011
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -6.0-: Statistic=0.89903, P-value=0.00012
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -7.0-: Statistic=0.91293, P-value=0.00036
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -2.0-: Statistic=0.86197, P-value=0.00002
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Shapiro test pour la modalité -8.0-: Statistic=0.85237, P-value=0.00000
L'échantillon ne semble pas suivre une distribution normale (on rejette H_0), alors on continue avec le test non-paramétrique de kruskal.test

Source : Réalisé par l'auteur sous Python

➤ Décision

la P-value est : 0.23, alors on ne rejette pas H_0 ,
il n'y a pas de différence significative entre les groupes.

Source : Réalisé par l'auteur sous Python

Nous constatons clairement que le test de Kruskal-Wallis nous autorise à accepter l'hypothèse nulle avec une P-value de 0.23 qui est largement supérieur à 5%.

➤ Conclusion

Pour conclure, nous acceptons l'hypothèse nulle qui stipule **qu'il n'y a pas de différence significative entre la qualité des relations avec le personnel soignant en fonction du service ayant accueilli le patient.**

Ce signifierait que les différents services entretiennent de bonnes relations avec les patients qu'ils accueillent.

Conclusion

Pour conclure, **la qualité de l'information reçue est la même selon les professions.**
En d'autres termes, le personnel soignant ne fait pas de distinction selon la profession quant à l'information donnée aux patients.

De plus, il n'y a pas de différence significative entre la qualité des relations avec le personnel soignant en fonction du service ayant accueilli le patient.

III- TEST DE LIAISON OU D'INDEPENDANCE :

Le score de relation est-il significativement corrélé à l'âge ?

Introduction

Cette dernière question vise à déterminer l'existence d'une corrélation significative entre le score de relation et l'âge des patients.

En d'autres termes, nous testerons si les notes attribuées par les patients sont corrélées à leur âge ?

Pour ce faire, nous allons utiliser le coefficient de corrélation de Pearson ou de Spearman pour évaluer la force et la direction de la relation entre ces deux variables.

➤ Formulation des hypothèses

Les hypothèses statistiques sont formulées comme suit :

H_0 : le score de la relation attribué par les patients n'est pas significativement corrélé à leur âge

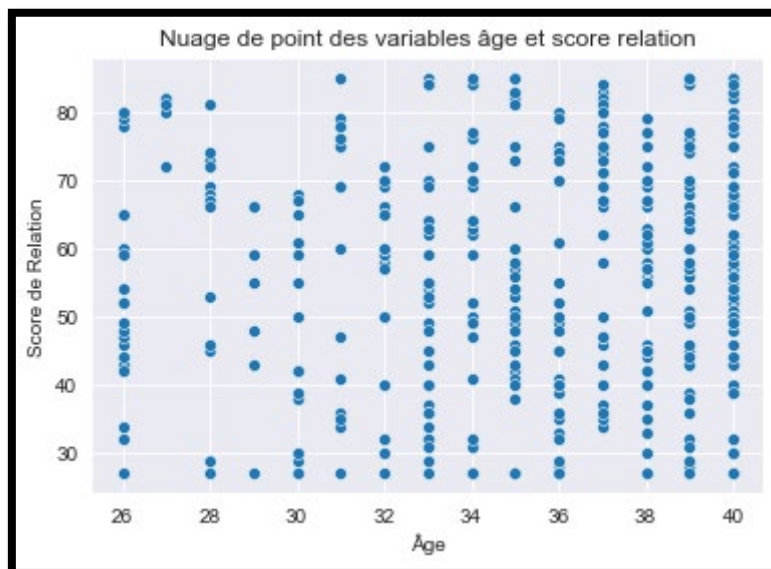
H_1 : le score de la relation attribué par les patients est significativement corrélé à leur âge

La vérification de ces hypothèses se fait sous le logiciel Python.

➤ Vérification des hypothèses

- **Etape 1** : analyse graphique des sous-populations

Figure 28: Nuage de point de la variable age et du score.relation



Source : Réalisé par l'auteur sous Python

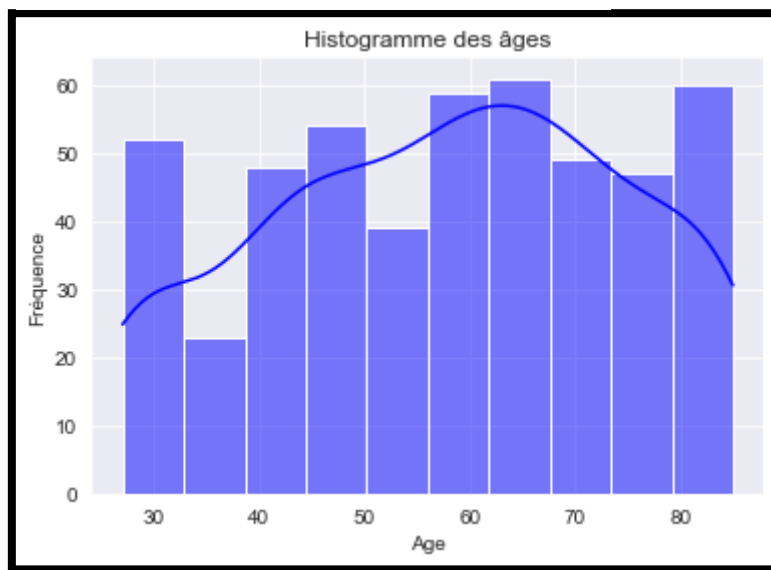
La répartition des points sur le graphique ci-dessus ne montre pas une corrélation explicite entre les deux variables.

La suite de notre analyse nous permettra de mieux comprendre cet état de fait.

- **Etape 2 : Test de normalité dans les sous populations**

La forme de l'histogramme n'est pas semblable à la forme typique en cloche de la distribution gaussienne. Elle a une forme assez irrégulière, par conséquent, la variable *age* ne semble pas suivre la loi normale.

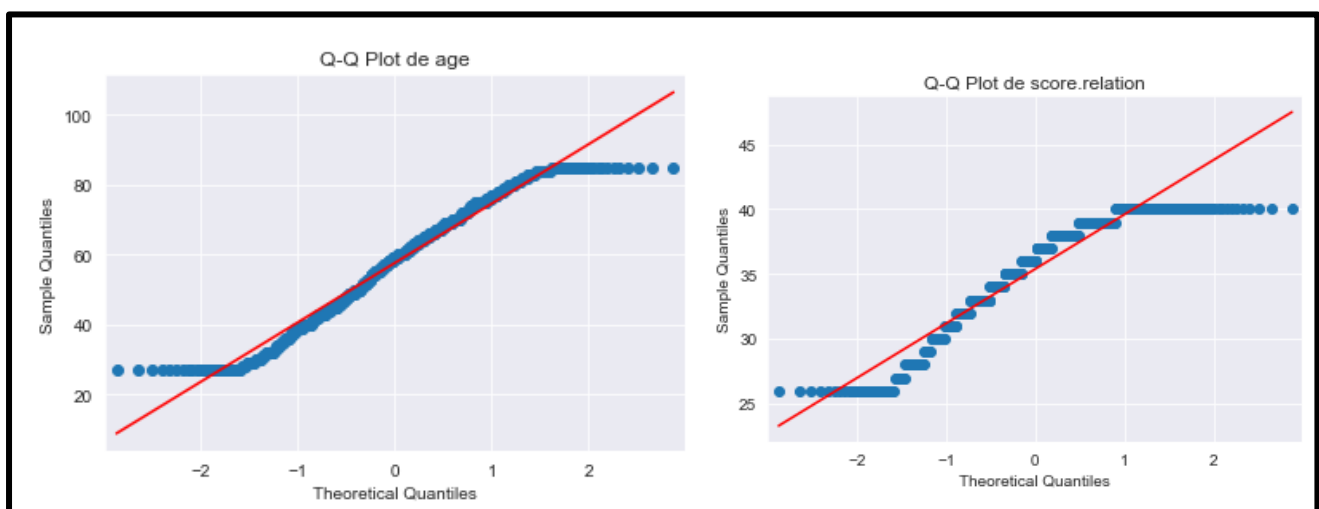
Figure 29: Histogramme de la variable âge



Source : Réalisé par l'auteur sous Python

Les QQplot des deux variables nous montrent aussi que les variables ne suivent pas une loi normale. En voici les figures :

Figure 30: qq-plot des variables age et score.relation



Source : Réalisé par l'auteur sous Python

Les de Shapiro Wilk nous permet de conclure à la non-normalité des variables **age** et **score.relation**.

Figure 31: résultats du test de Shapiro wilk sur les variables age et score.relation

```
Test de Shapiro-Wilk pour age: Statistique = 0.9616654515266418, p-value = 5.07744568611912e-10
Le test a rejeté l'hypothèse nulle. Les données ne semblent pas suivre une distribution normale.

Test de Shapiro-Wilk pour score.relation: Statistique = 0.889987587928772, p-value = 2.826735474215731e-18
Le test a rejeté l'hypothèse nulle. Les données ne semblent pas suivre une distribution normale.
```

Source : Réalisé par l'auteur sous Python

➤ Décision

Les coefficients de corrélation de Pearson et de Spearman donnent respectivement la valeur 0,05 et 0,067 cette valeur est très faible. Cela traduit un manque de corrélation entre les deux variables.

Figure 32: Resultat du test de Spearman

```
Coefficient de corrélation de Spearman : 0.0666111804142539
P-valeur : 0.14010819936316002
La corrélation n'est pas statistiquement significative.
```

Source : Réalisé par l'auteur sous Python

Le test sur ce coefficient nous montre que l'âge est très faiblement corrélé aux score.relation attribué par les patients.

Conclusion

Au terme de cette partie, nous retenons que nous acceptons l'hypothèse nulle de la non corrélation significative des âges et la valeur du score de la relation avec le personnel soignant.

Conclusion générale

Notre étude visait à déterminer si les patients qui fréquentent les hôpitaux de la région parisienne sont satisfaits de la qualité de la relation et des informations reçues du personnel médical.

L'analyse descriptive des données recueillies auprès de 534 patients des hôpitaux de cette région a révélé que ceux-ci sont des cadres d'entreprise de 57 ans en moyenne. Ces derniers ont apprécié la relation avec le personnel soignant en attribuant un score moyen de 35/40. De plus, les patients ont jugé avoir reçu le maximum d'information durant leur passage à l'hôpital.

Les résultats obtenus de cette première analyse seront testés grâce à des tests statistiques adaptés à cet effet.

La seconde partie portant sur la statistique inférentielle des questions issues de l'objectif principal montre que **le personnel soignant des hôpitaux de la région parisienne ne fait de discrimination selon le genre, l'âge et la profession des patients**. Les usagers sont globalement satisfaits, car plus de 66% des enquêtés *recommandent sûrement* les services par lesquels ils sont passés.

Les résultats sont très satisfaisants, ils montrent que le personnel de ces hôpitaux joue correctement leurs rôles. Cependant, nous ne pouvons nous empêcher de faire des recommandations afin d'améliorer davantage ces services.

Alors, sur la base de ces résultats, nous recommandons :

- *Renforcement de la Communication avec les Patients* : améliorer la communication entre le personnel soignant et les patients contribue à une meilleure compréhension des soins et à une relation de confiance, ce qui peut avoir un impact significatif sur la satisfaction des patients.
- *Formation Continue pour le Personnel Soignant* : la formation continue garantit que le personnel dispose des compétences nécessaires pour offrir des soins de haute qualité. Cela inclut des aspects tels que la communication, la sensibilité culturelle et la gestion du stress.

- *Évaluation Continue de la Satisfaction des Patients* : la rétroaction régulière des patients permet de comprendre les besoins changeants et d'apporter des ajustements continus pour répondre aux attentes des patients et maintenir un haut niveau de satisfaction.
- *Analyse Approfondie des Recommandations* : comprendre en détail les raisons derrière les recommandations des patients permet d'identifier les forces et les faiblesses spécifiques du système de soins, orientant ainsi les efforts d'amélioration.
- *Sensibilisation à la Non-Discrimination* : bien que les résultats indiquent une absence de discrimination, maintenir une sensibilisation constante et des formations régulières peut prévenir d'éventuelles situations de discrimination, garantissant un traitement équitable pour tous les patients.

Ces recommandations visent à améliorer la qualité des soins, à renforcer la confiance des patients et à favoriser un environnement de soins de santé positif dans les hôpitaux de la région parisienne.

Annexe

Importation des packages

```
#chargement des packages necessaire
```

```
import matplotlib.pyplot as plt
import pandas as pd
import missingno as msno
```

```
# Importation des donnees
```

```
hopital = pd.read_csv("satisfaction_hopital.csv", sep= ";")
```

Prétraitement des données

```
#transformation de la variable 'score,information' en 'ScoreInfo'
df.rename(columns={'score,information': 'ScoreInfo'}, inplace=True)
import seaborn as sns
```

```
import matplotlib.pyplot as plt
# Création de la heatmap pour visualiser les valeurs manquantes
sns.heatmap(df.isnull(), cbar=False, cmap='YlGnBu')
plt.title('Visualisation des Valeurs Manquantes dans le Jeu de Données')
plt.show()
```

```
# Utilisation de la fonction pour déterminer K
variables_a_imputer = ['age', 'profession', 'amelioration.sante',
'amelioration.moral', 'recommander', 'score.relation', 'ScoreInfo']
trouver_k_optimal(df, variables_a_imputer)
import pandas as pd
from sklearn.impute import KNNImputer
```

```
# Liste des variables à imputer
variables_a_imputer = ['service', 'sexe', 'age', 'profession',
'amelioration.sante', 'amelioration.moral', 'recommander',
'score.relation', 'ScoreInfo']
```

```
# Imputation des valeurs manquantes avec K-plus proche voisin
imputer = KNNImputer(n_neighbors=1) # Vous pouvez ajuster le nombre
de voisins selon vos besoins
```

```
df_imputed =
pd.DataFrame(imputer.fit_transform(df[variables_a_imputer]),
columns=variables_a_imputer)
# df_imputed contient maintenant les valeurs imputées
```

```
#Suppression des doublons
df_imputed=df_imputed.drop_duplicates()
```

```
#visualisation et traitement des valeurs aberrantes
import seaborn as sns
```

```
## Afficher la boîte à moustaches
sns.set_style('darkgrid')
sns.boxplot(data= df_imputed)
```

```
plt.title('BOITE A MOUSTACHE DES VARIABLES QUANTITATIVES AVANT
WINZORISATION')
plt.show()

#transformations des variable

df_propre['sexe'] = df_propre['sexe'].replace({0.0: 'homme', 1.0:
'femme'})
df_propre['recommander'] = df_propre['recommander'].replace({0.0:
'non', 1.0: 'oui,probablement', 2.0: 'oui,sûrement'})
df_propre['AmelioSante'] =
df_propre['amelioration.sante'].replace({0.0: 'aggrravée', 1.0:
'aucune amélioration', 2.0: 'peu améliorée', 3.0: 'nettement
améliorée'})
df_propre['AmelioMorale'] =
df_propre['amelioration.moral'].replace({0.0: 'aggrravé', 1.0:
'aucune amélioration', 2.0: 'peu amélioré', 3.0: 'nettement
amélioré'})
df_propre['profession'] = df_propre['profession'].replace({1.0:
'agriculteur', 2.0: 'artisan', 3.0: 'cadre', 4.0 :
'prof_intermediaire', 5.0: 'employé', 6.0: 'ouvrier', 7.0: 'sans
emploi', 8.0: 'autre'})
```

Analyse univariée

```
# Créer un boxplot avec seaborn
sns.boxplot(data=df_propre, x='age', color='skyblue')

# Ajouter la moyenne en tant que ligne rouge
mean_value = np.mean(df_propre['age'])
plt.axvline(mean_value, color='red', linestyle='dashed',
linewidth=2, label='Moyenne')

# Afficher les valeurs caractéristiques
plt.text(mean_value, 0.5, f'Moyenne: {mean_value:.2f}', color='red',
verticalalignment='baseline', horizontalalignment='right')

# Ajouter la médiane
median_value = np.median(df_propre['age'])
plt.axvline(median_value, color='green', linestyle='dashed',
linewidth=2, label='Médiane')

# Afficher les valeurs caractéristiques
plt.text(median_value, 0.5, f'Médiane: {median_value:.2f}',
color='green', verticalalignment='baseline',
horizontalalignment='left')

# # Ajouter d'autres caractéristiques si nécessaire (médiane, etc.)
# min_value = np.min(df_propre['age'])
# plt.axvline(min_value, color='blue', linestyle='dashed',
linewidth=2, label='Minimum')

# Personnaliser le graphique
plt.title('BOXPLOT DE LA VARIABLE AGE')
plt.legend()
```

```
# Afficher le graphique
plt.show()# Representation des ventes
sns.set_style("darkgrid")
plt.figure(figsize=(15, 5))
plt.subplot(111)
plt.plot(sales, marker = 'o')
plt.title("EVOLUTION DES VENTES PAR MOIS")
plt.xlabel("Dates")
plt.show()

# Isolement des variables qualitatives

quali=[] # Initialisation de quali en tant que liste vide

for col in df_propre.columns:
    if df_propre[col].dtype == 'category':
        quali.append(col)

print(quali)
```

Prevision: Construction du modele

```
df_train = sales.iloc[:-12]
df_test = sales.iloc[-12:]
model = HWES(df_train, seasonal_periods=12, trend='add',
seasonal='add')
fitted = model.fit()
print(fitted.summary())
sales_forecast = fitted.forecast(steps=12)
print(df_test, sales_forecast)
fig = plt.figure()
fig.suptitle('EVOLUTION DES VENTES DANS LES EPICERIES FAVORITA 2013-
2017 ')
past, = plt.plot(df_train.index, df_train, 'b.-', label='Ventes
Hystoriques')
future, = plt.plot(df_test.index, df_test, 'r.-', label='Ventes
Actuelles')
predicted_future, = plt.plot(df_test.index, sales_forecast, 'g.-',
label='Prévision des ventes')
plt.legend(handles=[past, future, predicted_future])
plt.show()

# Effectif des variables
colQuali = ['sexe', 'profession', 'recommander', 'service']
for var in colQuali:
    print(f"L'effectif par {var} est:\n",
round(df_propre[var].value_counts(), 2),"\n")
```

Statistique inférentielle

```
## 1- *La proportion de sujets recommandant le service dans lequel
ils sont passés est-il le même quel que soit le genre ?*

# Test d'Indépendance pour les variables catégorielles
```

```

# Effectuer le test du Chi-2
chi2_stat, p_value, dof, expected = chi2_contingency(TabEff1)

# Imprimer les résultats
print(f"Statistique de test du Chi-2 : {chi2_stat}")
print(f"P-value : {p_value}")
print(f"Degrés de liberté : {dof}")
print("Tableau des effectifs attendus :")
print(expected)

# Interpréter les résultats
alpha = 0.05
if p_value < alpha:
    print("On rejette l'hypothèse nulle. Les variables ne sont pas
indépendantes.")
else:
    print("On ne rejette pas l'hypothèse nulle. Les variables sont
indépendantes.")

### 2- *La moyenne du score de relation est-il significativement
différent chez les hommes et chez les femmes ?*
data = df_propre[['age', 'score.relation']].dropna()
# Effectuer le test de corrélation de Spearman
correlation_coefficient, p_value = spearmanr(data['age'],
data['score.relation'])
# Afficher les résultats
print(f"Coefficient de corrélation de Spearman :
{correlation_coefficient}")
print(f"P-valeur : {p_value}")
# Interprétation du résultat
alpha = 0.05
if p_value < alpha:
    print("La corrélation est statistiquement significative.")
else:
    print("La corrélation n'est pas statistiquement significative.")

### 4) Peut-on affirmer que la qualité de l'information reçue est la
même quelle que soit la profession?
# Effectuer le test de Wilcoxon sur les rangs
statistique, p_valeur =
kruskal(*[df_propre['ScoreInfo'][df_propre['profession'] == prof]
for prof in df_propre['profession'].unique()])

# Afficher les résultats
print("Statistique de Kruskal :", statistique)
print("P-valeur :", p_valeur)

# Vérifier si le résultat est statistiquement significatif
alpha = 0.05
if p_valeur < alpha:
    print("Rejeter l'hypothèse nulle : Il y a une différence
significative entre les groupes.")
else:
    print("Ne pas rejeter l'hypothèse nulle :\nIl n'y a pas de
différence significative entre les groupes.")

```

```
### 5) La qualité des relations avec le personnel soignant est-il
fonction du service ayant accueilli le patient?
# Test de Kruskal-Wallis
result_kruskal =
kruskal(*[df_propre['score.relation'][df_propre['service'] == prof]
for prof in df_propre['service'].unique()])

# Interprétation
alpha = 0.05
if result_kruskal.pvalue < alpha:
    print(f"la P-value est : {round(result_kruskal.pvalue, 2)},
    alors on rejette H0,\nil y a une différence significative entre au
    moins deux groupes.")
else:
    print(f"la P-value est : {round(result_kruskal.pvalue, 2)},
    alors on ne rejette pas H0,\nil n'y a pas de différence significative
    entre les groupes.")
```

Consultez le code entier su Git Hub :

[https://github.com/Morys1/Mini-Projets-](https://github.com/Morys1/Mini-Projets-INSEDS/blob/main/MiniProjet3/Code_MiniProjet_StatInferentielle.ipynb)
[INSEDS/blob/main/MiniProjet3/Code_MiniProjet_StatInferentielle.ipynb](https://github.com/Morys1/Mini-Projets-INSEDS/blob/main/MiniProjet3/Code_MiniProjet_StatInferentielle.ipynb)

Bibliographie et Webographie

Akposso, Didier Martial. «PREDICTION DES SERIES TEMPORELLES.» *LISSAGE EXPONENTIEL SIMPLE-DOUBLE-HOLT WINTERS AVEC R*. 2022.

ChatGPT

DataScientest. *DataScientest*. 10 Octobre 2023. <https://datascientest.com/data-quality-erreurs> (accès le Janvier 17, 2023).

Köseoğlu, Buse. *Medium*. 15 07 2023. <https://buse-koseoglu13.medium.com/guide-to-time-series-analysis-with-python-2-moving-average-process-784328325e5f> (accès le 12 12, 2023).

Louis. *STATOSCOPI*. 04 Mai 2021. <https://blog.statoscop.fr/timeseries-1.html> (accès le 12 12, 2023).

MAZEROLLE, Fabrice. *Statistique Descriptive*. Marseille: Gualino Editeur, 2006.

Table des matières

Avant-Propos.....	1
Sigles.....	Erreur ! Signet non défini.
Liste des illustrations.....	2
Listes des figures.....	2
Liste des tableaux.....	2
Sommaire.....	Erreur ! Signet non défini.
Résumé.....	4
Abstract.....	5
Introduction Générale.....	9
PARTIE I : PRETRAITEMENT ET STATISTIQUE DESCRIPTIVE.....	11
I. PRETRAITEMENT DU JEU DE DONNEES.....	12
Introduction.....	12
1. Présentation des données.....	12
2. Traitement des valeurs manquantes et doublons.....	14
3. Traitement des valeurs aberrantes.....	16
Conclusion.....	18
II. STATISTIQUE DESCRIPTIVE UNIVARIÉE.....	19
Introduction.....	19
1. Analyse univariée des variables quantitatives.....	19
2. Analyse univariée des variables qualitatives.....	21
Conclusion.....	22
PARTIE II : STATISTIQUE INFERENTIELLE.....	23
I- TEST DE COMPARAISON DE DEUX POPULATIONS INDEPENDANTES.....	24
Introduction.....	24
1- La proportion de sujets recommandant le service dans lequel ils sont passés est-il le même quel que soit le genre ?.....	24
2- La moyenne du score relation est-il significativement différent chez les hommes et chez les femmes ?.....	26
Conclusion.....	29
II- TEST DE COMPARAISON DE PLUSIEURS POPULATIONS INDEPENDANTES.....	30
Introduction.....	30

1- Peut-on affirmer que la qualité de l'information reçue est la même quelle que soit la profession ?	30
2- La qualité des relations avec le personnel soignant est-il fonction du service ayant accueilli le patient ?	35
Conclusion	38
III- TEST DE LIAISON OU D'INDEPENDANCE :Le score de relation est-il significativement corrélé à l'âge ?	39
Introduction	39
Conclusion générale	42
Annexe	44
Bibliographie et Webographie	49