



FEETME

Rapport de stage :
Analyse des données de la marche chez des
patients atteints de troubles de mobilité

Par Hack SOKA, supervisé par Damien JACOBS

Rapport de stage :
Analyse des données de la marche chez des patients atteints de
troubles de mobilité

Par ¹ Hack SOKA, ²
supervisé par Damien JACOBS. ³

30 septembre 2019

1. Élève ingénieur en dernière année de Mathématiques Appliquées et Calcul Scientifique (MACS3)
2. **mailto : hacksoka@gmail.com/hack.yacoubou.soka@umontreal.ca**
3. Data scientist - FeetMe, Paris.

Table des matières

Remerciements	iv
Introduction	v
Propos	v
Plan	v
Présentation de l'entreprise	vi
Bref historique	vi
Présentation des produits	vii
A) FeetMe One®	vii
B) FeetMe Monitor®	vii
C) FeetMe Insole® (3ème génération)	viii
Organisation de l'entreprise	ix
I PFE : Recherche bibliographique sur les méthodes statistiques utilisées dans l'analyse de la marche en milieu écologique	x
Outils statistiques utilisés pour l'analyse de la marche en vie réelle	xi
1) La marche : définition	xii
A) Cycle de la marche	xii
a) La phase d'appui	xii
b) La phase d'oscillation	xii
2) Méthodes statistiques pour l'analyse de la marche en vie réelle	xiii
A) Test d'ANOVA à un facteur	xiii
a) La normalité	xiii
b) L'homogénéité des variances	xiii
α) Le test de Levene	xiv
c) L'indépendance des observations	xiv
B) Test de MANOVA	xiv
C) Test de Chi-deux	xiv
a) Exemple : Test de conformité du χ^2	xiv
D) Le d de Cohen	xiv
E) Test de Kolmogorov-Smirnov	xv
F) Test de Lilliefors	xv
G) Test de Kruskal-Wallis	xv
H) Mean Absolute Percentage Error (MAPE)	xv
a) Régression avec MAPE : définition	xv
I) Correction de Bonferroni	xvi
J) Coefficients de corrélation	xvi

a)	Corrélation de Pearson	xvi
b)	Corrélation de Spearman	xvi
K)	Test de Shapiro-Wilk	xvii
L)	Test de Student	xvii
M)	Test post-hoc	xvii
N)	Variabilité des données	xvii
a)	Coefficient de variation	xvii
b)	Variabilité inter-groupes	xviii
c)	Variabilité intra-groupe	xviii
d)	Relation fondamentale de l'ANOVA	xviii
e)	La statistique de test ou F-statistique	xviii

II Étude des données du pied diabétique xix

Écriture du poster - FeetMe sur le pied diabétique xx

3)	Gradation du risque d'ulcération des pieds chez le patient diabétique	xx
A)	Classification de WAGNER	xx
B)	Classification de PEDIS	xx
C)	Classification de l'UT(Université du Texas)	xxi
D)	Classification du consensus international sur le pied diabétique (IWGDF)	xxi
4)	Description de l'étude	xxi
A)	Démographie de la donnée d'étude	xxi
B)	Métriques utilisées	xxii
5)	Méthodes	xxii
6)	Résultats	xxiii
7)	Conclusion	xxiii

Écriture de l'article - FeetMe sur l'étude du pied diabétique xxiv

8)	Variabilité intra-patient	xxiv
9)	Distribution des données	xxv
10)	Différences significatives entre les différents grades	xxvi
11)	Description statistique de la donnée	xxvii
12)	Adimensionnement/Adimensionnalisation	xxix
13)	Conclusion	xxxii

Conclusion xxxiii

Références xxxv

Liste des tableaux et figures

1	FeetMe One®.	vii
2	FeetMe Monitor®.	viii
3	FeetMe Insole® (3ème génération).	viii
4	Organigramme de l'entreprise - Mai 2019.	ix
5	Les deux phases principales du cycle de la marche.	xii
6	Le cycle (complet) de la marche - par Viel.	xiii
7	Démographie de la population.	xxii
8	Évolution des pressions maximales à chaque pas chez un patient de grade 2.	xxv
9	Histogrammes des pressions moyennes.	xxv
10	Histogrammes des pics de pression.	xxvi
11	Histogrammes des intégrales pression-temps.	xxvi
12	Différences significatives avec correction de Bonferroni.	xxvi
13	Description statistique des métriques en fonction des grades.	xxvii
14	Comparatif entre description statistique et histogrammes.	xxviii
15	Gaussian mixture appliqué aux grades 1.	xxviii
16	Comparatif des histogrammes des données avec dimension vs sans dimension.	xxx
17	Descriptions statistiques des données avec dimension vs sans dimension.	xxx
18	Outliers.	xxx
19	Illustration des outliers	xxx
20	Comparatif des moyennes entre outliers et ensemble de la donnée (avec dimension vs sans dimension).	xxx

Remerciements

Mes remerciements vont premièrement à l'endroit de Damien JACOBS, mon tuteur de stage, qui m'a non seulement accordé sa confiance en me donnant l'occasion d'effectuer mon stage au sein de la société FeetMe, mais aussi permis d'enrichir mon bagage de connaissances en analyse de données et en data science plus généralement.

Je remercie particulièrement Leila FARID, pour son soutien lors de mes déplacements au CHU de Lille et aussi pour sa formation sur le terrain.

J'aimerais également remercier Sabine FERRÉ et Leila Farid pour leurs conseils en graphisme, qui m'auront servi dans la réalisation du poster.

Enfin, merci à toute l'équipe FeetMe.

Introduction

DANS le présent rapport, je fais état du déroulement de mon stage de fin d'études. Tout au long du récit, je vous raconte de façon plus ou moins concise le travail que j'ai pu effectuer les six mois passés au sein de la société FeetMe. Ma mission était principalement axée sur l'analyse, l'exploitation des données du pied diabétique, dans le but de trouver des biomarqueurs numériques d'aide au diagnostic et à la prévention des plaies chez les patients diabétiques.

L'étude sur le pied diabétique que j'ai menée durant mon stage, a abouti à deux principaux évènements : la **publication d'un poster** ainsi que la **publication d'un article** sur le pied diabétique, tous deux basés sur les résultats obtenus lors de l'étude menée pendant mon stage. Dans ce rapport, je vais donc naturellement vous présenter le travail accompli pour aboutir à ces deux publications, en commençant par l'écriture du poster, pour finir par l'écriture de l'article.

Dans le cadre de mon projet de fin d'études (PFE), j'ai également été amené à faire une étude bibliographique des méthodes mathématiques (statistiques plus précisément) utilisées dans l'analyse de la marche en milieu écologique.

Propos

CE rapport s'inscrit dans le cadre de ma dernière année d'études d'ingénierie en mathématiques appliquées et calcul scientifique, il suppose que le lecteur a des solides notions de probabilités, de tests statistiques, d'analyse de données mais aussi de programmation. Toute la programmation a été réalisée en langage Python.

Plan

LA structure du rapport est la suivante :

Avant tout, je fais la présentation de l'entreprise, de l'historique à l'organisation en passant par la présentation des produits, puis :

- ① dans la première partie, je décris les différentes méthodes issues de ma recherche bibliographique (la liste est non exhaustive).
- ② Dans la deuxième partie,
 - je commence par relater l'écriture du poster sur le pied diabétique.
 - Enfin, je termine par l'écriture de l'article sur le pied diabétique.

Présentation de l'entreprise

Sommaire

Bref historique	vi
Présentation des produits	vii
A) FeetMe One®	vii
B) FeetMe Monitor®	vii
C) FeetMe Insole® (3ème génération)	viii
Organisation de l'entreprise	ix

FEETME est une jeune start-up innovante du domaine médical en quête de la résolution des problèmes de troubles de mobilité. À l'heure actuelle, on compte plus de 100 millions de personnes à l'échelle mondiale souffrant de troubles de mobilité. Ces troubles sont dus en partie à la vieillesse, à des maladies chroniques ou encore à des traumatismes.

Expérimentée dans l'analyse de la marche, l'entreprise s'est alors donnée pour mission de développer des semelles connectées, écologiques, intelligentes et ergonomiques ainsi que des applications (mobiles et web), afin d'accompagner les patients dans le suivi et la gestion de leur maladie chronique par le biais de biomarqueurs numériques et d'outils de suivi des études en milieu écologique pour les patients, les médecins et les groupes pharmaceutiques. Mais aussi, dans le but de simplifier et accélérer l'évaluation et la réhabilitation de ces troubles de mobilité notamment le diabète, la maladie de Parkinson, la sclérose en plaques et le handicap après accident vasculo-cérébral.

Aujourd'hui, FeetMe est en pleine croissance et a déjà commercialisé deux de ses produits (2 générations de semelles) d'aide au diagnostic pour les professionnels de santé. L'heure est maintenant à la préparation de la 3^e version du produit destinée aux patients.

Bref historique

LA jeune start-up naît de la décision de jeunes polytechniciens Alexis MATHIEU et Andrey MOSTOVOV en 2013. Ils ont l'ambition de développer un outil connecté, capable d'effectuer un suivi en temps réel de l'évolution du traitement d'un patient atteint de la maladie de Parkinson ou d'une sclérose en plaques. L'idée est de créer une paire de semelles dotées de 70 capteurs de pression permettant de déterminer avec exactitude le mouvement d'un patient, dans le but de lui apporter les solutions nécessaires à l'amélioration de sa marche.

C'est le début d'une belle aventure, soldée par plusieurs récompenses telles que : le prix de la meilleure start-up d'innovation de l'année ou encore de l'objet connecté de l'année, ce qui leur vaudra des financements pour le lancement du produit. Ils travaillent alors en collaboration avec plusieurs laboratoires, hôpitaux et centres d'études cliniques.

L'entreprise se développe et connaît l'été 2015 sa première certification européenne (marquage CE) suite à une multitude de tests. S'ensuit alors, la commercialisation du produit **FeetMe One®**¹ auprès des professionnels de santé. Puis l'entreprise décide d'élargir son champ d'application au

1. Dispositif médical

marché du sport, visant ainsi les jeux olympiques de 2024 qui auront lieu en France, avec pour ambition de développer une solution qui permettrait d'identifier et alerter les coureurs d'une éventuelle blessure. Mais l'idée finit par être abandonnée, l'entreprise souhaitant se concentrer sur le marché du médical.

L'équipe a remporté de nombreux concours et aides publiques (Concours Mondial de l'Innovation, Création Développement, FUI notamment) et a réalisé une levée de fonds lui permettant d'assurer son développement. Elle a également officialisé une collaboration avec les laboratoires Novartis pour améliorer la prise en charge de la sclérose en plaques.

Présentation des produits

FEETME dispose de 3 types de semelles à son actif. Deux premiers types (**FeetMe One®** et **FeetMe Monitor®**) commercialisés dont les cibles sont les podologues, les cliniques et les centres hospitaliers. Et un autre produit (**FeetMe Insole® (3ème génération)**) pas encore sur le marché, qui sera commercialisé très prochainement. Ces produits contiennent une paire de semelles accompagnée de son logiciel et d'une application mobile. Le premier produit aide les podologues à effectuer leurs diagnostics. Le second produit s'adresse aux patients atteints de pathologies chroniques pour évaluer leur mobilité en vie réelle.

A) FeetMe One®

FeetMe One® est une semelle connectée de la société FeetMe qui permet de mesurer plusieurs paramètres spatio-temporels de la marche dont les pressions plantaires, le centre de mouvement et de variabilité de pression, l'index de force médiale latérale, la symétrie de marche et une analyse d'allure. Le logiciel embarqué propose également des mesures d'analyse statique comme la position du centre de gravité, la vitesse et l'accélération, la surface et la longueur de déplacement. La semelle analyse aussi de manière intelligente les données de pression pour le diagnostic et la prévention d'ulcères du pied diabétique. Ces semelles connectées ont un ensemble de 63, 68 ou 70 capteurs capacitifs selon la taille : 63 pour les tailles allant du 37 au 39, 68 du 40 au 42 et 70, du 43 au 45. Ces capteurs sont reliés à un smartphone via une connexion sans fil.

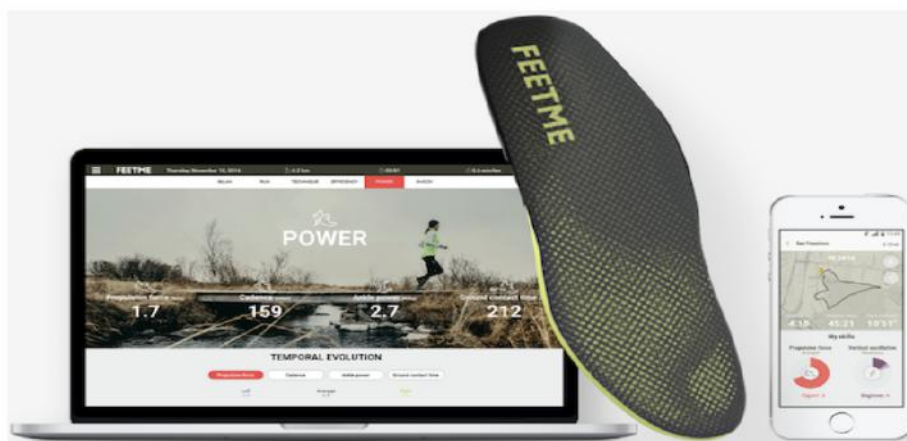


FIGURE 1 – FeetMe One®.

B) FeetMe Monitor®

FeetMe Monitor® est un dispositif composé d'une paire de semelles connectées permettant de mesurer la pression plantaire et les paramètres spatio-temporels de la marche en temps réel et d'une application mobile FeetMe Android application (FTM-AA) qui permet de donner un retour en temps réel et d'enregistrer les données de pressions plantaires et de mouvement.

Les mesures faites par les semelles **FeetMe Monitor®**, sont réalisées grâce à une centrale inertielle couplée à 19 capteurs capacitifs. La centrale inertielle mesure l'accélération et la vitesse angulaire dans 3 directions (x, y, z). Mesure fournie par STMicroelectronics. Les capteurs capacitifs sont produits en interne et les mesures de capacités sont directement lues par un processeur ATMEL. La mesure de la pression plantaire est transmise via une connexion sans fil avec le logiciel.



Semelles FeetMe et interface de l'application.

FIGURE 2 – FeetMe Monitor®.

C) FeetMe Insole® (3ème génération)

Plus fine, plus légère et toujours aussi précise avec un capteur en moins (18 capteurs), la 3è génération de semelle FeetMe offre une solution d'analyse ambulatoire de la marche qui combine des capteurs de pression, des centrales inertielles et une capacité de calcul embarquée pour une mesure en temps réel des paramètres de marche.

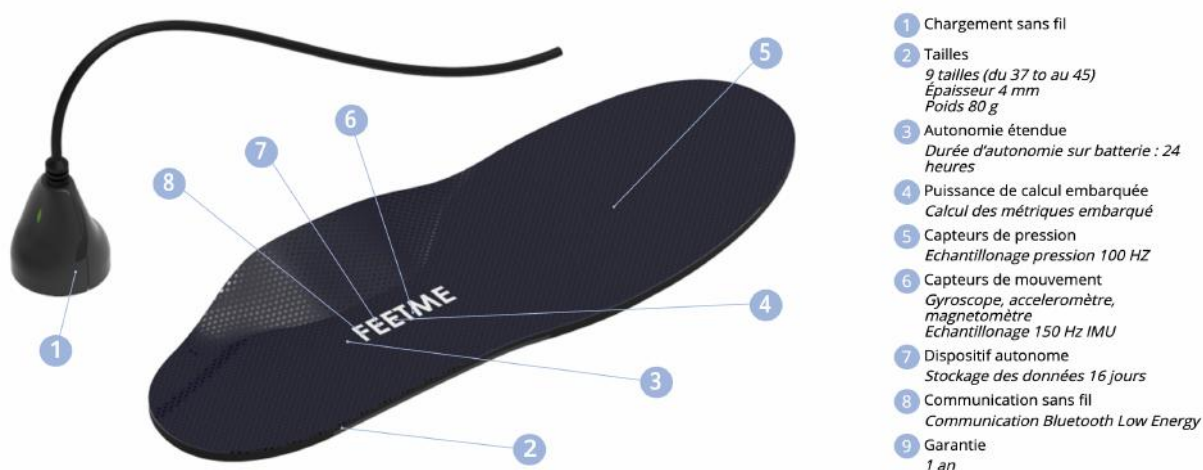


FIGURE 3 – FeetMe Insole® (3ème génération).

Organisation de l'entreprise

L'entreprise FeetMe est composée d'une vingtaine d'ingénieurs, développeurs et chefs de projet. Comme indiqué précédemment, FeetMe travaille en collaboration avec plusieurs laboratoires pharmaceutiques et académiques. L'entreprise est conseillée par un board stratégique et un board médical. L'équipe FeetMe est répartie en plusieurs pôles, à savoir :

- Un pôle R&D data science composé d'ingénieurs en imagerie, en TI santé, en mathématiques appliquées et en génie des matériaux, s'occupant à la fois de la détection des biomarqueurs numériques, de la qualité des capteurs, et de l'analyse des données cliniques.
- Un pôle mécanique et électronique qui s'occupe de la mécanique et l'électronique régies dans la semelle.
- Un pôle systèmes embarqués qui s'occupe de la partie intelligence de la semelle.
- Un pôle software composé de développeurs qui gèrent les applications mobiles et web.
- Un pôle qualité et management du produit.
- Un pôle industrialisation.
- Un pôle commercialisation.

L'entreprise est en pleine croissance, de nouveaux postes sont en cours de création, de fait, ces pôles sont amenés à être réorganisés. Ci-dessous un organigramme de l'entreprise.

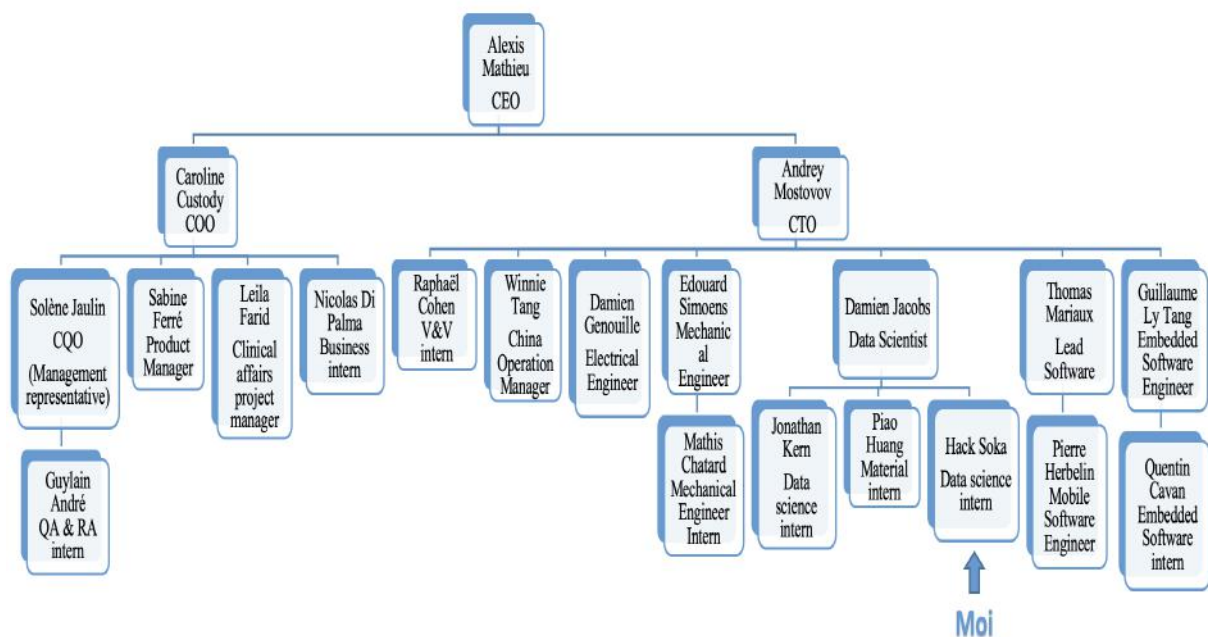


FIGURE 4 – Organigramme de l'entreprise - Mai 2019.

Première partie

PFE : Recherche bibliographique sur les méthodes statistiques utilisées dans l'analyse de la marche en milieu écologique

Outils statistiques utilisés pour l'analyse de la marche en vie réelle

Sommaire

1)	La marche : définition	xii
A)	Cycle de la marche	xii
a)	La phase d'appui	xii
b)	La phase d'oscillation	xii
2)	Méthodes statistiques pour l'analyse de la marche en vie réelle	xiii
A)	Test d'ANOVA à un facteur	xiii
a)	La normalité	xiii
b)	L'homogénéité des variances	xiii
α)	Le test de Levene	xiv
c)	L'indépendance des observations	xiv
B)	Test de MANOVA	xiv
C)	Test de Chi-deux	xiv
a)	Exemple : Test de conformité du χ^2	xiv
D)	Le d de Cohen	xiv
E)	Test de Kolmogorov-Smirnov	xv
F)	Test de Lilliefors	xv
G)	Test de Kruskal-Wallis	xv
H)	Mean Absolute Percentage Error (MAPE)	xv
a)	Régression avec MAPE : définition	xv
I)	Correction de Bonferroni	xvi
J)	Coefficients de corrélation	xvi
a)	Corrélation de Pearson	xvi
b)	Corrélation de Spearman	xvi
K)	Test de Shapiro-Wilk	xvii
L)	Test de Student	xvii
M)	Test post-hoc	xvii
N)	Variabilité des données	xvii
a)	Coefficient de variation	xvii
b)	Variabilité inter-groupes	xviii
c)	Variabilité intra-groupe	xviii
d)	Relation fondamentale de l'ANOVA	xviii
e)	La statistique de test ou F-statistique	xviii

1) La marche : définition

La marche est le mode de locomotion naturel de l'humain. A priori, la marche consiste à mettre un pied devant l'autre de façon alternée et répétée, cela nous semble simple dit comme cela mais en réalité la marche est un phénomène complexe, c'est bien pour cela qu'elle s'acquière au bout de quelques années chez l'Homme. En effet, la marche nécessite une coordination de mouvements des parties du corps afin de maintenir l'équilibre de ce dernier lors de son déplacement. Ce maintien de l'équilibre debout est combiné à la propulsion des deux membres inférieurs vers l'avant. La marche est caractérisée par une succession de double appuis, d'appuis unipodaux, le corps restant en permanence en contact avec le sol par au moins un appui unilatéral (cf figure 6).

A) Cycle de la marche

Le cycle de la marche est standard et reconnu à l'échelle internationale comme base de description d'une marche normale ou pathologique. Il est constitué des différentes phases suivantes :

a) La phase d'appui

Le pied est en contact avec le sol. Cette phase débute par le contact initial du pied avec le sol et se termine par le décolllement des orteils du sol.

b) La phase d'oscillation

Le pied n'est plus en contact avec le sol, la propulsion du membre inférieur est enclenchée. La phase d'oscillation débute au décolllement des orteils et prend fin lorsque le même pied recontacte le sol (talon au sol).

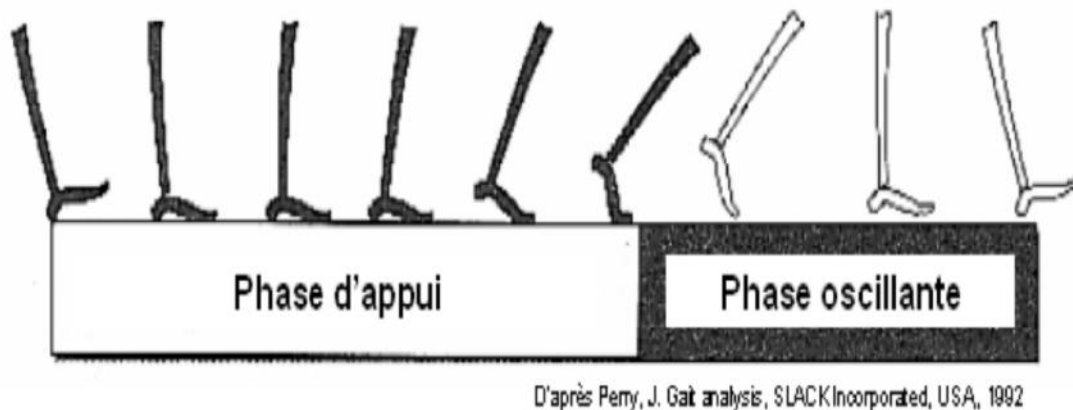


FIGURE 5 – Les deux phases principales du cycle de la marche.

α) Le test de Levene

- Si p-valeur est > 0.005 , alors les groupes ont des variances homogènes (égales).
- Sinon, on a inhomogénéité des variances.

c) L'indépendance des observations

Une manière de s'en assurer est de tracer un nuage de points des données ou encore de calculer les coefficients d'autocorrélation. Dans le premier cas, il faut que le nuage de points ne dégage aucune tendance et dans le second les coefficients d'autocorrélation ne doivent pas être très différents de 0. Les conditions d'expérience que l'on choisit nous déterminent aussi si nous sommes dans un cas d'indépendance. On s'assure donc que lors de l'échantillonnage, les données soient récoltées de manière indépendante.

Pour avoir des résultats précis au test d'ANOVA, il est préférable de travailler avec des groupes de même taille : on parle d'expérience équilibrée.

B) Test de MANOVA

Il s'agit du test d'ANOVA multivarié.

C) Test de Chi-deux

Il s'agit d'un test statistique dont la statistique de test suit une loi du χ^2 sous l'hypothèse nulle. Il permet de vérifier si un échantillon d'une variable aléatoire X donne des observations comparables à celles d'une loi de probabilité P qui est présumée être la loi de probabilité de X . L'hypothèse nulle H_0 est alors : la variable X suit la loi de probabilité P .

a) Exemple : Test de conformité du χ^2

Soit des observations d'un échantillon de données x_1, \dots, x_n d'une variable aléatoire X prenant un nombre fini N de valeurs. L'hypothèse nulle à tester est la suivante : la probabilité que X prenne la valeur i est p_i pour i allant de 1 à N , avec $\sum_{i=1}^N p_i = 1$. Soit \hat{p}_i la probabilité empirique que X prenne la valeur i , i.e. le nombre d'observations qui prennent la valeur i dans l'échantillon divisé par le nombre total d'observations n .

$$\hat{p}_i = \frac{1}{n} \sum_{k=1}^n [x_k = i], \text{ avec } [x_k = i] = \begin{cases} 1 & \text{si } x_k = i \\ 0 & \text{sinon} \end{cases}$$

La statistique de test est alors définie comme suit :

$$T = \sum_{i=1}^N \frac{n(\hat{p}_i - p_i)^2}{p_i}$$

Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi du χ^2 à $N - 1$ degré(s) de liberté.

D) Le d de Cohen

Encore appelé d' , il permet de caractériser la taille d'effet associée à une population par rapport à une hypothèse nulle.

Un d de Cohen $\simeq 0.2$ est considéré comme « faible » alors que $d \simeq 0.5$ est considéré comme « moyen » et $d \simeq 0.8$ est dit « fort ».

E) Test de Kolmogorov-Smirnov

Ce test est basé sur la distance entre la fonction de répartition empirique \hat{F} et la fonction de répartition que l'on veut tester, ici Φ . Pour un échantillon x_1, \dots, x_n de loi normale, on s'attend à avoir $\hat{F}(t) \simeq \Phi(t), \forall t \in \mathbb{R}$.

La statistique de test est définie par :

$$D = \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - \Phi(t)|$$

Le test de Kolmogorov rejette la normalité si D est trop grande.

F) Test de Lilliefors

Ou plutôt test de Kolmogorov-Lilliefors, c'est une adaptation des tests classiques de Kolmogorov-Smirnov. La statistique de test est :

$$D = \max |F_n^*(x) - S_n^*(x)|$$

Valeur maximale des écarts entre la fonction de répartition $F_n^*(x)$ de la loi normale ajustée aux observations, avec la distribution cumulative $S_n^*(x)$ définie par :

$$S_n^*(x) = \begin{cases} 0 & \text{pour } x < x_1 \\ \frac{i}{n} & \text{pour } x_1 < x < x_{i+1} \\ 1 & \text{pour } x > x_n \end{cases}$$

G) Test de Kruskal-Wallis

C'est une alternative non paramétrique au test d'ANOVA d'ordre 1. Le test de Kruskal-Wallis permet de comparer au moins trois échantillons et tester l'hypothèse nulle selon laquelle les échantillons à comparer sont issus de la même distribution ou de distributions de même médiane. À la différence d'une ANOVA d'ordre 1, le test de Kruskal-Wallis est basé sur les rangs et pas sur les moyennes.

H) Mean Absolute Percentage Error (MAPE)

C'est une mesure de prédiction de la précision des méthodes de prévision en statistique. Par exemple, dans l'estimation de tendance, aussi utilisée comme fonction de perte pour des problèmes de régression en machine learning et d'évaluation des modèles car c'est une interprétation très intuitive en terme d'erreur relative.

$$M = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \text{ où } F_t \text{ est la prédiction de } A_t.$$

a) Régression avec MAPE : définition

Soit le couple $Z = (X, Y)$ avec $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}$ et n copies i.i.d $(X_1, Y_1), \dots, (X_n, Y_n)$ de (X, Y) , le modèle de régression consiste à trouver une fonction mesurable g « proche » de Y .

En régression classique, la « proximité » de g avec Y est mesurée par le risque en norme L^2 aussi appelé Mean Squared Error (MSE) :

2. indépendantes et identiquement distribuées

$$MSE(\hat{\theta}) \stackrel{\text{def}}{=} \mathbb{E}[(\hat{\theta} - \theta)^2], \quad \hat{\theta} : \text{estimateur de } \theta$$

Mais dans le contexte de la régression avec MAPE, on détermine g telle que :

$$g_{MAPE}(x) = \underset{g \in \mathcal{G}}{\operatorname{argmin}} \mathbb{E} \left[\left| \frac{g(x) - Y}{Y} \right| \middle| X = x \right]$$

Où \mathcal{G} est la classe des modèles considérés (ex : modèles linéaires).

I) Correction de Bonferroni

Elle permet de corriger le seuil de significativité lors des comparaisons de plusieurs groupes notamment pour réduire le nombre de faux positifs.

Principe

Soit H_1, \dots, H_m une famille d'hypothèses et p_1, \dots, p_m leurs p-valeurs correspondantes. Soit m le nombre total d'hypothèses nulles et m_0 le nombre de vraies hypothèses nulles. La FamilyWise Error Rate (FWER) est la probabilité de rejeter au moins une H_i vraie \iff faire au moins une erreur de type I (rejeter à tort l'hypothèse nulle). La correction de Bonferroni rejette l'hypothèse nulle pour chaque $p_i \leq \frac{\alpha}{m}$ contrôlant ainsi la FWER au seuil α .

$$\begin{aligned} FWER &= \mathbb{P} \left[\bigcup_{i=1}^{m_0} (P_i \leq \frac{\alpha}{m}) \right] \leq \sum_{i=1}^{m_0} \left[\mathbb{P}(P_i \leq \frac{\alpha}{m}) \right] \\ &= m_0 \frac{\alpha}{m} \leq m \frac{\alpha}{m} = \alpha \end{aligned}$$

J) Coefficients de corrélation

a) Corrélation de Pearson

Elle permet de réfuter ou non l'indépendance des variables. Le coefficient de corrélation entre deux variables aléatoires X et Y ayant chacune une variance finie, est donné par :

$$r = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Où $\operatorname{Cov}(X, Y)$ désigne la covariance de X et Y et σ_X et σ_Y les écarts-types respectifs de X et Y .

b) Corrélation de Spearman

Le coefficient de corrélation de Spearman est le coefficient de corrélation de Pearson mais sur les variables de rang.

$$r = \frac{\operatorname{Cov}(\operatorname{rg}_X, \operatorname{rg}_Y)}{\sigma_{\operatorname{rg}_X} \sigma_{\operatorname{rg}_Y}}$$

K) Test de Shapiro-Wilk

Ce test est basé sur l'hypothèse nulle selon laquelle un échantillon x_1, \dots, x_n est issu d'une population normalement distribuée. La statistique du test s'écrit :

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Où

- $x_{(i)}$ désigne la i ème statistique d'ordre, i.e. le i ème plus petit nombre dans l'échantillon.
- \bar{x} représente la moyenne de l'échantillon, $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$
- la constante a_i est donnée par :

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \quad \text{avec } m = (m_1, \dots, m_n)^T$$

Les m_i sont les espérances des statistiques d'ordre d'un échantillon de variables i.i.d de loi normale, V est la matrice de variance-covariance de ces statistiques d'ordre.

On interprète la statistique W comme le coefficient de détermination (le carré du coefficient de corrélation) entre la série des quantiles générés à partir de la loi normale et les quantiles empiriques obtenus à partir des données. Plus W est élevé, plus la compatibilité avec la loi normale est crédible. La région critique, rejet de la normalité, s'écrit :

$$R.C. : W < W_{crit}$$

Les valeurs seuils W_{crit} pour différents risques α et effectifs n sont lues dans la table de Shapiro-Wilk³.

L) Test de Student

Encore appelé test t, ce test est un ensemble de tests statistiques paramétriques où la statistique de test :

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n^*}$$

suit une loi de Student à $n-1$ degrés de liberté lorsque l'hypothèse nulle est vraie. S_n^* est l'estimateur de la variance :

$$S_n^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

M) Test post-hoc

Permet pour un test d'ANOVA de montrer où sont les différences significatives entre les groupes.

N) Variabilité des données

a) Coefficient de variation

Il permet d'avoir une idée de l'étendue (amplitude) relative des données, c'est le rapport entre la moyenne et l'écart-type.

3. <http://www.educnet.education.fr/rnchimie/math/benichou/tables/tshapiro/tshapiro.htm>

$$c = \frac{m}{\sigma}$$

b) Variabilité inter-groupes

C'est la somme des carrés moyens inter-groupes. Elle permet de quantifier la variabilité entre les groupes, on la note : CM_{inter} .

$$CM_{inter} = \frac{SC_{inter}}{\underbrace{k-1}_{\text{degré de liberté}}}$$

SC_{inter} est la somme des carrés des écarts inter-groupes (entre les moyennes des groupes et la moyenne globale) :

$$SC_{inter} = N \sum_{i=1}^k (\bar{X}_i - \bar{X})^2$$

N est le nombre d'individus dans chaque échantillon.

c) Variabilité intra-groupe

Somme des carrés moyens intra-groupe.

$$CM_{intra} = \frac{SC_{intra}}{n-k}, \quad SC_{intra} = \sum_{i=1}^k \sum_{j=1}^N (X_{ji} - \bar{X}_i)^2$$

d) Relation fondamentale de l'ANOVA

$$SC_{totale} = \sum_{i=1}^N \sum_{j=1}^k (X_{ij} - \bar{X})^2 = SC_{inter} + SC_{intra}$$

Remarque : Cette relation ne s'applique pas aux variabilités i.e.

$$CM_{totale} \neq CM_{inter} + CM_{intra}$$

e) La statistique de test ou F-statistique

Elle est calculée à l'aide de la formule suivante :

$$F = \frac{CM_{inter}}{CM_{intra}}$$

Deuxième partie

Étude des données du pied diabétique

Écriture du poster - FeetMe sur le pied diabétique

Sommaire

3)	Gradation du risque d'ulcération des pieds chez le patient diabétique .	xx
A)	Classification de WAGNER	xx
B)	Classification de PEDIS	xx
C)	Classification de l'UT(Université du Texas)	xxi
D)	Classification du consensus international sur le pied diabétique (IWGDF) .	xxi
4)	Description de l'étude	xxi
A)	Démographie de la donnée d'étude	xxi
B)	Métriques utilisées	xxii
5)	Méthodes	xxii
6)	Résultats	xxiii
7)	Conclusion	xxiii

Dans cette étude, nous cherchions à montrer si l'on pouvait identifier le degré de risque clinique associé au pied diabétique. Il existe plusieurs classifications de la maladie du diabète, mais nous avons travaillé avec celle de l'IWGDF⁴ qui est décrite ci-dessous :

3) Gradation du risque d'ulcération des pieds chez le patient diabétique

Le degré de la maladie (grade) est représentatif du degré de risque de plaie. Plus il est élevé, plus il y a de risque de plaie(s) et donc d'éventuelle(s) amputation(s) chez le patient. Il existe plusieurs classifications de la maladie du diabète, à savoir :

A) Classification de WAGNER

De grade 0 (Pas de lésion ouverte, mais présence possible d'une déformation osseuse ou d'hyperkératose) à grade 5 (Gangrène massive du pied associée à des lésions nécrotiques et à une infection des tissus mous).

B) Classification de PEDIS

De grade 1 (Pas de symptômes ou de signes d'artériopathie périphérique du pied atteint avec artères pédieuses etc) à grade 4(sepsis sévère).

4. International Working Group on Diabetic Foot

C) Classification de l'UT(Université du Texas)

Du grade 0 - stade A (lésion épithéliale - pas d'infection, pas d'ischémie) au grade 3 - stade D (atteinte de l'os ou de l'articulation - infection et ischémie). Le stade représente le pourcentage estimé d'infection et d'ischémie).

D) Classification du consensus international sur le pied diabétique (IWGDF)

D'après la classification de l'IWGDF, les grades de la maladie sont répartis de la manière suivante :

- **Grade 0** : Absence de neuropathie sensitive.
- **Grade 1** : Neuropathie sensitive⁵ isolée.
- **Grade 2** : Neuropathie sensitive associée
 - ➔ à une artériopathie des membres inférieurs⁶ et/ou
 - ➔ à une déformation du pied.⁷
- **Grade 3** : Antécédent
 - ➔ d'ulcération du pied évoluant depuis plus de 4 semaines et/ou
 - ➔ d'amputation des membres inférieurs.

4) Description de l'étude

Des études ont montré que des pressions plantaires élevées pouvaient entraîner une augmentation du risque d'ulcères au niveau du pied diabétique (Diabetic Foot Ulcers : DFU). Développer des outils de mesure de la pression plantaire est un facteur important dans l'amélioration de la prévention des DFU. Le but de l'étude était :

- ① dans un premier temps, d'évaluer les pressions renvoyées par de nouvelles semelles connectées pendant la marche, afin de ressortir des différences significatives entre les grades (1,2 et 3).
- ② Puis dans un second temps, il était question de faire des comparaisons parmi les patients diabétiques de niveau 3 (grade 3). D'abord pour les patients amputés, nous avons comparé les zones du pied qui ont été amputées vs les autres zones, ensuite pour les patients uniquement blessés (qui n'ont pas subi d'amputation), les zones de plaies vs les autres zones.

A) Démographie de la donnée d'étude

L'étude a porté sur une population de 120 patients de l'hôpital la Pitié Salpêtrière et des hospices civils de Lyon, dont 87 hommes. Ces patients ont été sujets à une expérience de marche avec les semelles **FeetMe One®** dans des chaussures standardisées.

Voici ci-après un tableau qui résume la démographie de la donnée :

-
5. Définie par l'anomalie du test au monofilament de Semmes-Wenstein (10g)
 6. Définie par l'absence d'au moins un des deux poulx du pied ou par un IPS < 0.9
 7. Hallux valgus, orteil en marteau ou en griffe, proéminence de la tête des métatarsiens, déformations post-chirurgicales ou liées à une neuro-arthropathie (pied de Charcot)

	Grade 1 N = 9	Grade 2 N = 52	Grade 3 N = 59
Age (years)	69±6	70±9	62±13
Sex ratio (M/F)	5/4	42/10	44/15
Weight (Kg)	94 ± 20	88 ± 18	90 ± 21
BMI (Kg/m²)	35±6	30±5	30±6
Type 2 diabetes	9 (100)	50 (96)	50 (85)
Diabetes duration	16±12	19±11	20±13
Insulin use	7 (78)	33 (63)	47 (80)
Retinopathy	2 (22)	23 (44)	38 (64)
Nephropathy	5 (55)	30 (58)	28 (47)
Lower limb arteriopathy	0 (0)	23 (43)	19 (32)

FIGURE 7 – Démographie de la population.

Ce qui est à remarquer dans cette démographie, c'est le nombre de grades 1 qui est faible comparé aux grades 2 et 3, mais également la variation de poids au sein de la population (autour de ± 20 kg), car le poids est un élément essentiel dans la mesure des pressions.

B) Métriques utilisées

Trois métriques sont entrées en jeu dans l'analyse de données, à savoir : le **pic de pression**, l'**intégrale pression-temps** et la **pression moyenne**.

- ➔ **Pic de pression** : sélection spatiale des pressions.
- ➔ **Intégrale pression-temps** : aussi une sélection spatiale des pressions.
- ➔ **Pression moyenne** : sélection temporelle des pressions (sélection des pressions en fonction du temps).

5) Méthodes

➔ Analyse comparative 1 :

Deux méthodes ont alors été appliquées, la première consistait à prendre le maximum de la pression renvoyée par chaque capteur sur tous les pas, méthode que l'on a baptisée AS. La seconde consistait à prendre la plus grande valeur renvoyée par tous les capteurs sur tous les pas, méthode que l'on a baptisée OS.

- ➔ **Analyse comparative 2** : dans cette deuxième analyse, seules la pression maximale et l'intégrale pression-temps ont été étudiées. Cette fois-ci, deux différentes approches ont été utilisées selon le paramètre. Pour la pression maximale, nous avons pris le maximum de chaque capteur sur tous les pas et pour l'intégrale pression-temps, nous avons pris la moyenne de chaque capteur sur tous les pas.

6) Résultats

TABLE 1 – Your caption here

	Amputated vs Non-amputated			Wound vs Non-wound		
Metrics	P-value	Median	Std	P-value	Median	Std
Peak pressure	0.050	0.11	0.23	0.020	0.11	0.13
Pressure-time integral	1.0	0.19	0.22	0.046	0.12	0.19

7) Conclusion

Les résultats ont confirmé que les patients de grade 3 avaient des plus grandes intégrales pression-temps que ceux de grade 2 en appliquant la méthode AS ou OS. D'après les résultats, on pouvait confirmer les différences significatives entre les différentes zones sauf, dans le cas de la comparaison des zones amputées vs non-amputées pour l'intégrale pression-temps. Le poster a été présenté lors d'un congrès au Pays-Bas et nous avons eu droit à des retours positifs et des encouragements qui nous ont poussés à aller plus loin dans la recherche.

Écriture de l'article - FeetMe sur l'étude du pied diabétique

De nombreuses pathologies telles que le diabète requièrent l'analyse de la marche du patient et de la force normale appliquée au sol. Les médecins utilisent cette analyse pour quantifier et évaluer l'efficacité des traitements dans différents domaines des troubles neurologiques telles que la sclérose en plaques, la maladie de Parkinson et l'hémiplégie.

Le diabète mellitus, plus connu en France sous le nom de diabète sucré, peut conduire à des complications telles que la neuropathie périphérique dont l'une des conséquences est l'ulcération du pied. Cette ulcération du pied est une cause majeure de morbidité et de mortalité et représente un coût financier important. Les patients diabétiques avec neuropathie ont des ulcérations dues au manque de sensibilité induit par la neuropathie.

Les zones à risque du pied peuvent être évaluées à l'aide des mesures de pression, que l'on peut réaliser via des outils tels que :

- Des plaques de force qui font une mesure 3D des forces de réactions au sol, associées à des systèmes de capture de mouvement.
- Des plaques baropodométriques qui permettent de mesurer la distribution de la pression sous le pied et peuvent être utilisées en analyse statique, c'est-à-dire lorsque le patient est debout ou encore en analyse dynamique, lorsque le patient est en marche.

Chacun de ces systèmes comporte des points positifs et des points négatifs. Il existe des articles dont l'objet d'étude était l'évaluation de ces outils dans le but de s'assurer que les médecins choisissent des outils qui correspondent vraiment à leurs besoins.

Depuis plusieurs années maintenant, les systèmes embarqués peuvent être utilisés hors laboratoire pour évaluer des paramètres baropodométriques. Cependant, à notre connaissance, aucun de ces systèmes n'est capable de mesurer les paramètres spatio-temporels des patients en conditions écologiques (de vie réelle) et pendant une longue période de temps. En outre, ces outils doivent être extrêmement précis et non chronophages. De fait, les outils portables, associés à une analyse de données automatique peuvent accroître la quantité d'informations utilisées par les cliniciens pour diagnostiquer et traiter leurs patients.

FeetMe (Paris, France) a développé un outil : **FeetMe One®** afin de répondre à ces exigences.

8) Variabilité intra-patient

La figure suivante montre l'évolution des pressions maximales à chaque pas pour chaque métrique chez un patient de grade 2 choisi de manière aléatoire :

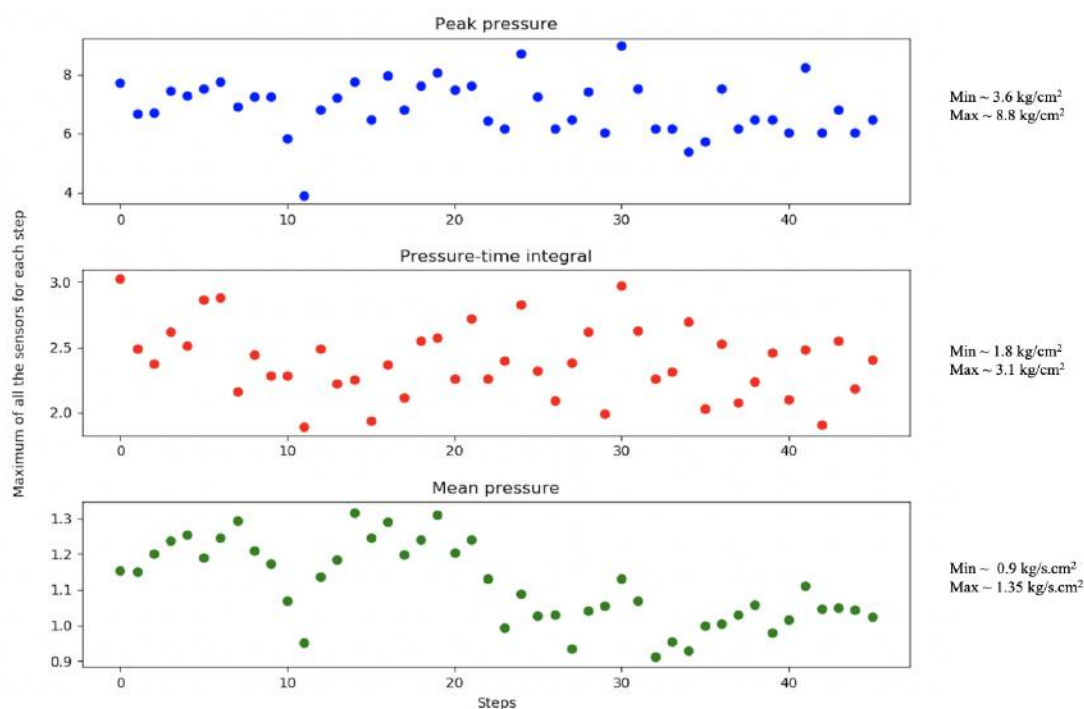


FIGURE 8 – Évolution des pressions maximales à chaque pas chez un patient de grade 2.

À travers cette figure, on remarque à quel point on a des variations intra-sujet. Autrement dit, d'un pas à l'autre on a des valeurs de maximum très différentes, cela laisse penser que les patients ont une marche telle que la répétabilité du pas est très rare.

9) Distribution des données

Pour voir de quelle manière sont distribuées nos données, nous avons affiché les histogrammes de chaque métrique en fonction du grade de la maladie.

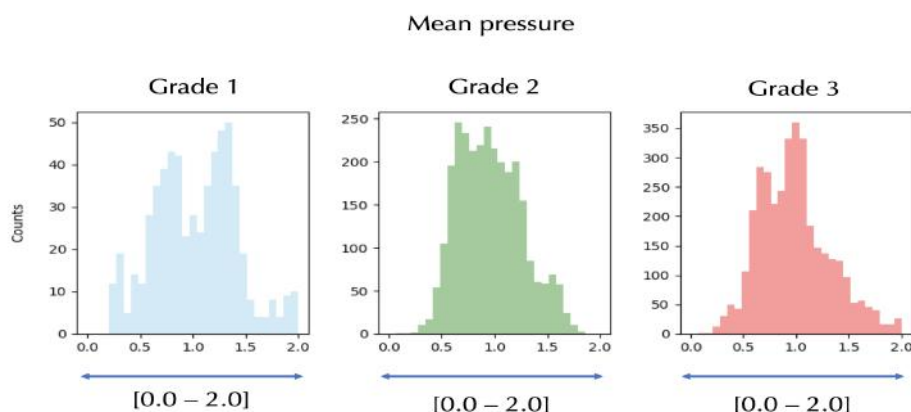


FIGURE 9 – Histogrammes des pressions moyennes.

Ici on remarque que quelque soit le grade, les valeurs de pressions pour la métrique « mean pressure » sont comprises entre 0 et 2.

Cependant, lorsqu'on regarde les histogrammes pour les deux autres métriques « pressure-time integral » et « peak pressure » (figure ci-dessous), on s'aperçoit qu'il y a une certaine variabilité dans les échantillons, la borne supérieure des intervalles croît à mesure que le grade augmente.

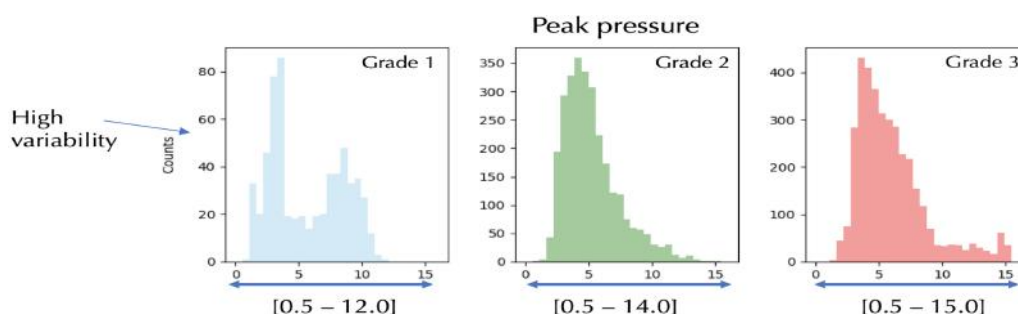


FIGURE 10 – Histogrammes des pics de pression.

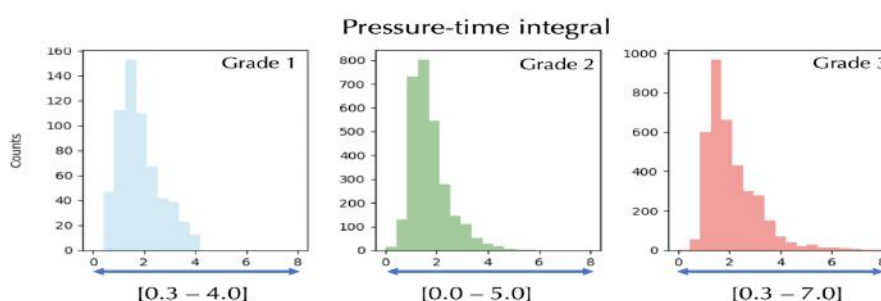


FIGURE 11 – Histogrammes des intégrales pression-temps.

10) Différences significatives entre les différents grades

Afin de mieux comprendre la donnée, nous avons voulu voir si oui ou non il existe des différences significatives entre les différents échantillons, voir à quel point elles sont différentes entre-elles. Mais cette fois-ci nous avons utilisé la correction de Bonferroni dans le but de limiter le risque de faux positifs. Voici ci-dessous, le tableau des p-valeurs obtenues en appliquant des tests statistiques sur nos 3 différents échantillons (grades 1, 2 et 3).

Approach	Grades	P-values with Bonferroni adjustment		
		Peak pressure	Pressure-time integral	Mean pressure
Maximum of all the sensors for each step (with dimension)	G1-G2	0.0366	0.0364	0.0013
	G1-G3	0.4207	0.0001	0.0407
	G2-G3	0.0023	0.0	0.2644

FIGURE 12 – Différences significatives avec correction de Bonferroni.

De ces analyses statistiques, il ressort que les grades 1 et 3 n'ont pas de différences significatives en tenant compte de la métrique « peak pressure ». Il en est de même pour les grades 2 et 3 concernant la métrique « mean pressure ».

Cela a éveillé notre curiosité et nous a poussés à voir le problème de plus près. Nous avons donc décrit les échantillons plus en détail :

11) Description statistique de la donnée

With dimension									
	Peak pressure			Pressure-time integral			Mean pressure		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Mean	5.64	5.29	5.77	1.83	1.74	2.04	1.04	0.96	0.99
Standard deviation	2.89	2.27	2.55	0.81	0.79	0.97	0.40	0.31	0.36
Confidence interval	[5.41, 5.87]	[5.11, 5.47]	[5.57, 5.97]	[1.77, 1.89]	[1.67, 1.80]	[1.97, 2.12]	[1.00, 1.07]	[0.94, 0.99]	[0.96, 1.01]





-  The mean values of each metric and the confidence interval are lower for grade 2 than grade 1.
-  The high standard deviation of peak pressure for grade 1 comparing to grade 3 can explain the non-significant difference between grade 1 and grade 3.
-  The very close mean values between grade 2 and grade 3 can explain the non-significant differences observed for the mean pressure metrics.
-  Recovery in confidence intervals

FIGURE 13 – Description statistique des métriques en fonction des grades.

Dans ce tableau, on fait l'analyse suivante :

- En **rouge**, on observe une sorte d'anomalie, dans le sens où les moyennes ne respectent pas un certain ordre, c'est-à-dire que contrairement à ce à quoi on s'attend, les moyennes des grades 1 quelle que soit la métrique, sont plus grandes que celles des grades 2.
- En **jaune**, on remarque que l'écart-type des grades 1 est plus grand que celui des grades 3. Cela pourrait expliquer le fait que l'on ait pas observé des différences significatives entre ces deux groupes.
- En **vert**, on peut remarquer que les valeurs des moyennes des grades 2 et 3 sont assez proches, ce qui pourrait justifier la non-significativité des différences observées entre ces deux groupes.
- En **bleu**, on constate un certain recouvrement dans les intervalles de confiance, les intervalles se chevauchent.

Lorsqu'on s'intéresse encore de plus près à l'état de la donnée, on remarque les observations suivantes :

With dimension									
	Peak pressure			Pressure-time integral			Mean pressure		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Mean	5.64	5.29	5.77	1.83	1.74	2.04	1.04	0.96	0.99
Standard deviation	2.89	2.27	2.55	0.81	0.79	0.97	0.40	0.31	0.36
Confidence interval	[5.41, 5.87]	[5.11, 5.47]	[5.57, 5.97]	[1.77, 1.89]	[1.67, 1.80]	[1.97, 2.12]	[1.00, 1.07]	[0.94, 0.99]	[0.96, 1.01]

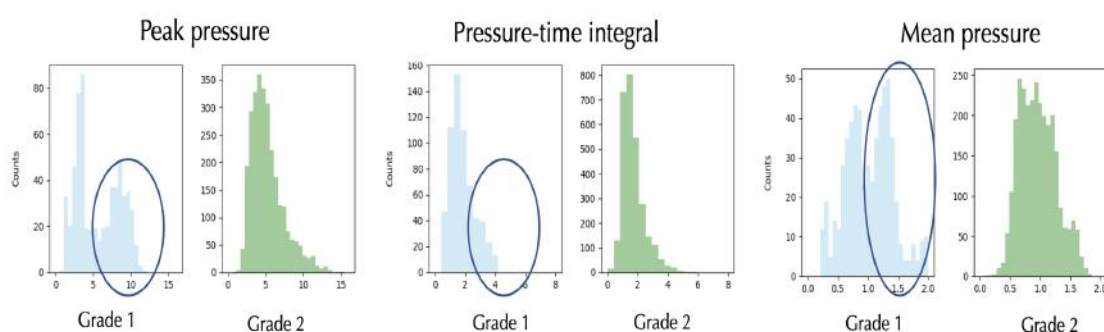


FIGURE 14 – Comparatif entre description statistique et histogrammes.

Les distributions des échantillons comportent une sorte de double distribution notamment très prononcée chez les grades 1, qui serait à l'origine des anomalies observées. Pour nous en convaincre, nous avons appliqué un modèle statistique de gaussian mixture pour estimer paramétriquement la distribution de nos données, voici ce que cela a donné :

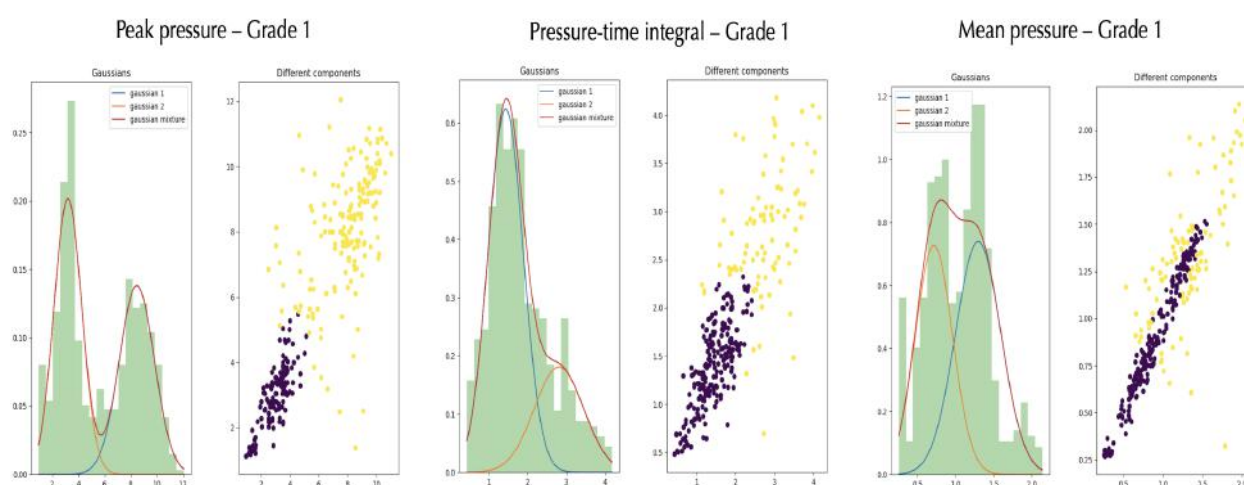


FIGURE 15 – Gaussian mixture appliqué aux grades 1.

On remarque bien l'existence de doubles distributions dans les données. En s'interrogeant sur l'origine de ces anomalies, nous sommes arrivés à la conclusion qu'il fallait exploiter la donnée de manière différente. Rappelez-vous, nous avons dit dans (A) que le poids est un élément important dans les mesures de pressions. En effet, la pression est une force appliquée sur une surface :

$$p = \frac{F}{S}, \quad F \text{ en Newton(s)}, S \text{ en } cm^2 \text{ et } p \text{ en } N/cm^2$$

Mais la pression peut aussi s'exprimer en bar ($1bar \simeq 1kg/cm^2$). Voici une explication : pour donner un exemple de vie courante, considérons notre peau comme une surface, ce qui est tout à fait le cas, chaque cm^2 de la surface que représente notre peau supporte une force de la part de l'atmosphère qui représente environ $1kg$ de son poids : c'est la pression atmosphérique au niveau de la mer. En fait, le kilogramme par centimètre carré ($kg/cm^2 \equiv kgf/cm^2$) est une mesure de pression égale à un kilogramme-force par centimètre carré. Ainsi, les pressions récoltées dans nos données sont exprimées en kg/cm^2 l'équivalent d'une masse (ce qu'on considère comme poids par abus de langage) sur une surface.

L'idée derrière cette nouvelle façon d'exploiter la donnée est d'adimensionner la donnée, car en effet, le poids du patient induit un biais dans la mesure des pressions. Mais également la surface des capteurs qui représente un facteur au carré dans la formule de la pression.

On pourrait se demander par exemple :

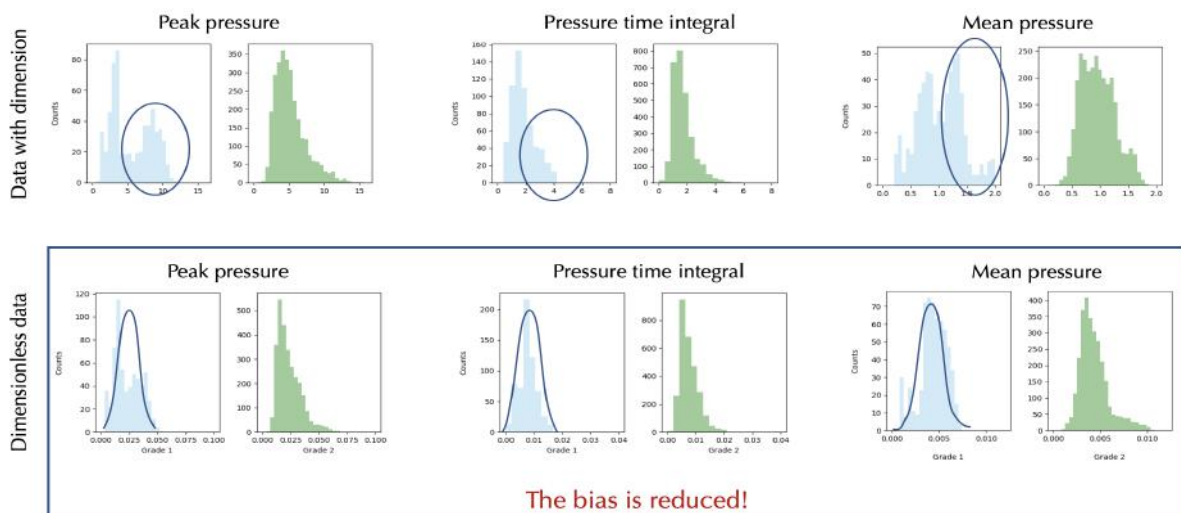
Pourquoi les personnes obèses n'ont pas de risque de plaie(s) comme c'est le cas pour les patients diabétiques ?

Plusieurs facteurs entrent en ligne de compte pour expliquer ce phénomène, parmi lesquels : les forces de cisaillement, la temporalité ainsi que les pressions normales. Mais a priori, les obèses exercent leur poids sur plusieurs capteurs et donc naturellement, la pression est telle que la force exercée est divisée par autant de fois les surfaces impliquées au carré. Ce qui laisse penser que même la surface induit un biais dans les mesures de pressions, d'autant qu'il y a des espaces entre capteurs qui pourraient fausser la mesure.

D'où l'intérêt d'adimensionner la donnée.

12) Adimensionnement/Adimensionnalisation

Nous avons donc divisé les pressions par le poids (la masse en fait) de chaque patient et multiplié ces valeurs par la surface des capteurs. De cette manière, nous avons des données sans dimension. En affichant de nouveau les histogrammes des échantillons, nous avons observé une nette amélioration, l'effet de double distribution est considérablement réduit et donc le biais aussi comme en témoignent les figures suivantes :



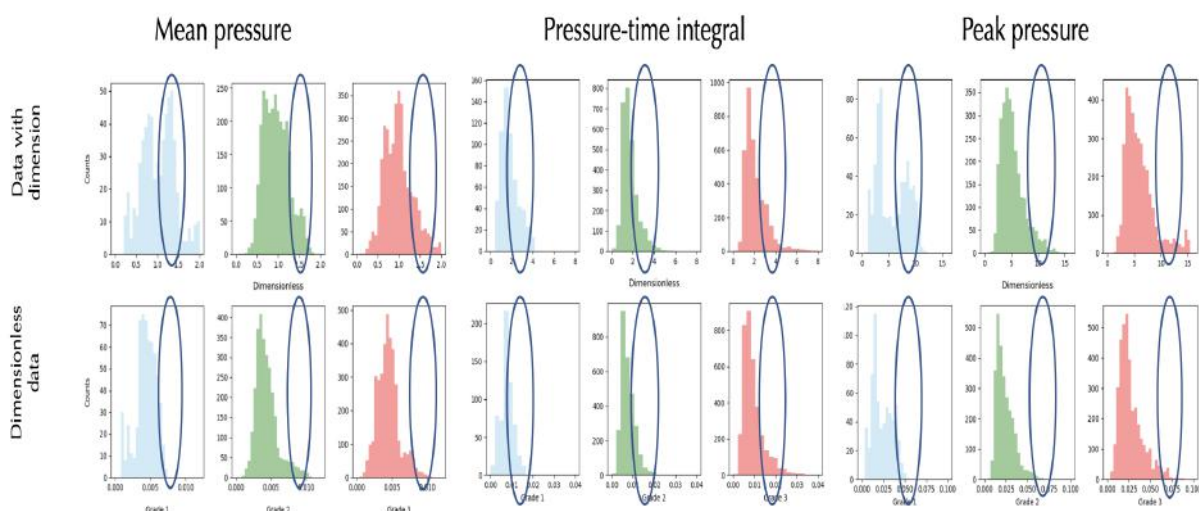


FIGURE 16 – Comparatif des histogrammes des données avec dimension vs sans dimension.

Lorsqu'on regarde la description statistique des données, on se rend compte que les valeurs moyennes sont dans un ordre croissant, autrement dit, plus le grade est élevé plus la moyenne est grande au sein d'une même métrique. En outre, pour les « peak pressure » l'écart-type des grades 1 est maintenant plus petit que celui des grades 3.

With dimension									
	Peak pressure			Pressure-time integral			Mean pressure		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Mean	5.64	5.29	5.77	1.83	1.74	2.04	1.04	0.96	0.99
Standard deviation	2.89	2.27	2.55	0.81	0.79	0.97	0.40	0.31	0.36
Confidence interval	[5.41, 5.87]	[5.11, 5.47]	[5.57, 5.97]	[1.77, 1.89]	[1.67, 1.80]	[1.97, 2.12]	[1.00, 1.07]	[0.94, 0.99]	[0.96, 1.01]
Dimensionless									
	Peak pressure			Pressure-time integral			Mean pressure		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Mean	0.0236	0.0239	0.0267	0.00772	0.00776	0.00936	0.00435	0.00438	0.00442
Standard deviation	0.0115	0.0107	0.0133	0.00305	0.00329	0.00489	0.00144	0.00166	0.00155
Confidence interval	[0.0227, 0.0245]	[0.0231, 0.0248]	[0.0256, 0.0277]	[0.0074, 0.0080]	[0.0075, 0.0080]	[0.0090, 0.0098]	[0.0042, 0.0045]	[0.00425, 0.00451]	[0.00429, 0.00454]

The mean values are now in ascending order, from grade 1 to grade 3 for each metric.

FIGURE 17 – Descriptions statistiques des données avec dimension vs sans dimension.

En voyant cela, nous avons eu une autre idée d'approche, celle de s'intéresser aux outliers (valeurs aberrantes des échantillons ou encore valeurs extrêmes) afin de voir s'ils pouvaient être des indicateurs d'appartenance à un échantillon. Commençons d'abord par définir ce que sont les outliers. Les outliers sont définis comme les valeurs de l'échantillon qui sont très différentes du reste des valeurs de l'échantillon. Ce sont les valeurs plus éloignées de la moyenne que le reste des valeurs (« normalement mesurées »), elles peuvent apparaître par hasard dans n'importe quelle distribu-

tion, mais elles indiquent souvent soit une erreur de mesure soit que la population est distribuée suivant une loi de probabilité à queue lourde.

Le schéma suivant permet de mieux visualiser ce qu'est un outlier, il s'agit de toute valeur supérieure à 2.698σ ou inférieure à -2.698σ

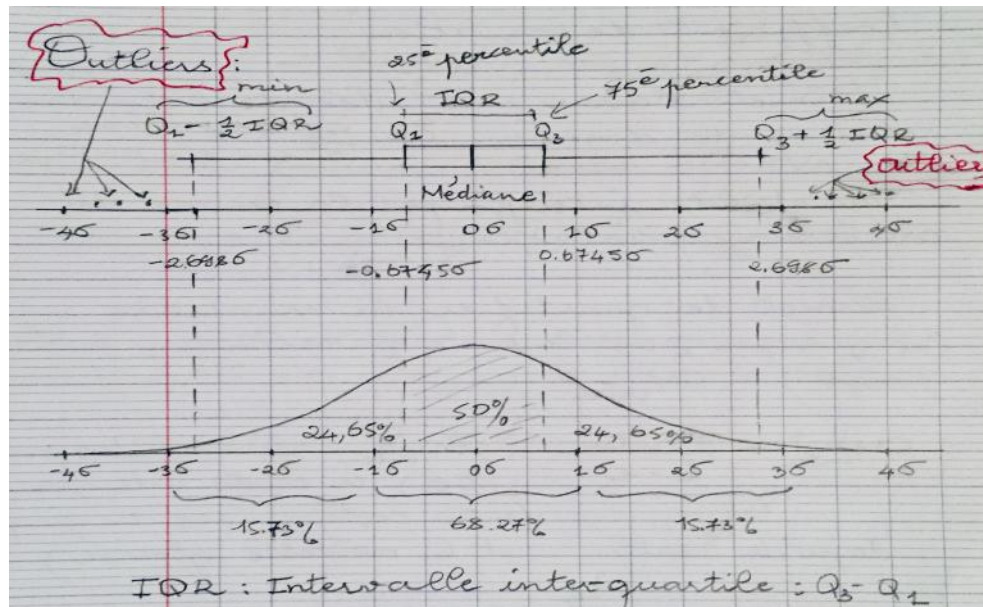


FIGURE 18 – Outliers.

Voici quelques illustrations d'échantillons avec outliers pour étayer le schéma précédent :

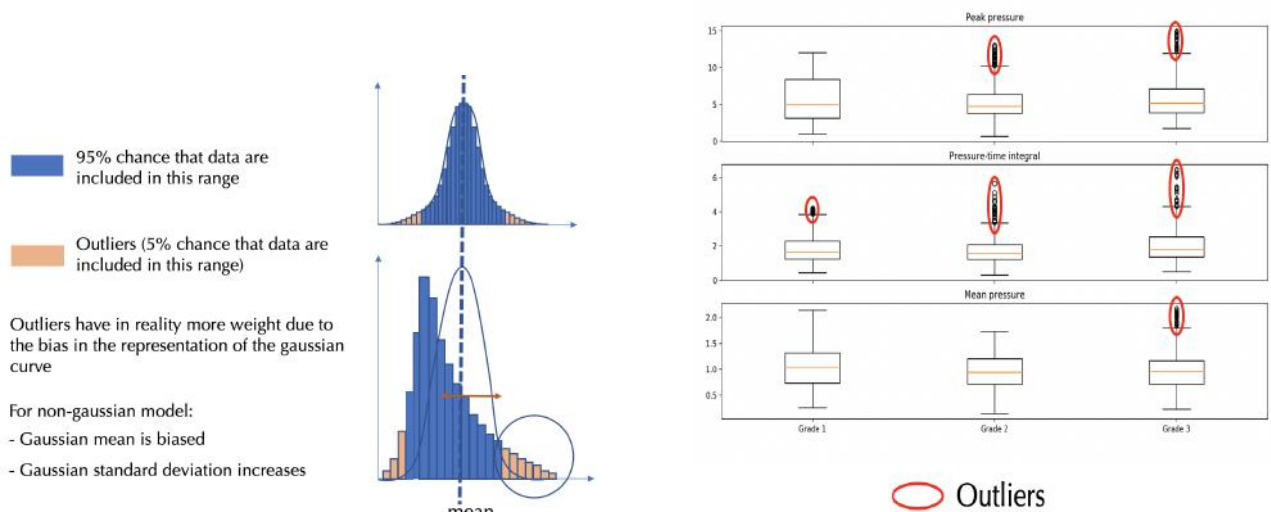


FIGURE 19 – Illustration des outliers

Ce qu'on peut remarquer avec les outliers de nos données, c'est que la différence des moyennes est beaucoup plus prononcée et qu'il y a donc bien une hiérarchie entre les différents grades de la maladie.

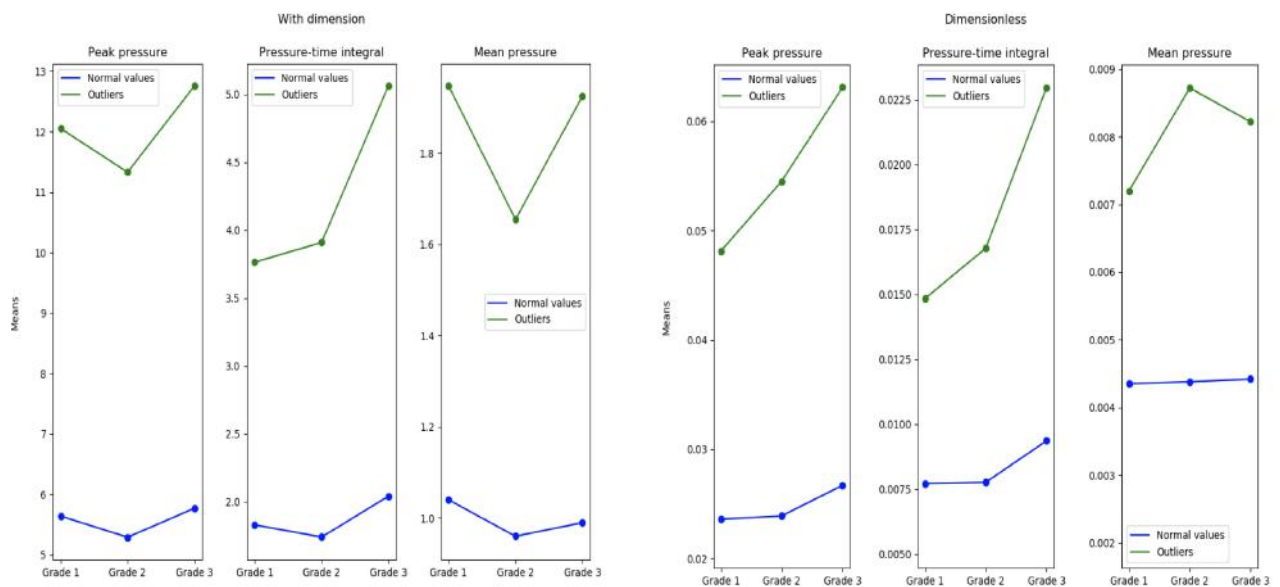


FIGURE 20 – Comparatif des moyennes entre outliers et ensemble de la donnée (avec dimension vs sans dimension).

13) Conclusion

À l'heure actuelle, l'article est encore en cours de rédaction, mais les résultats montrent une certaine avancée dans l'analyse de la donnée. L'approche utilisée est très prometteuse et ouvre des portes à une analyse plus approfondie pouvant débouchant sur la trouvaille d'un biomarqueur.

Conclusion

En résumé, mon stage au sein de FeetMe aura été une expérience très enrichissante, tant sur le plan cognitif que sur le plan humain. J'ai eu la chance de travailler aux côtés d'ingénieurs dotés d'un bon savoir faire. Cela m'a permis de progresser à tous les points de vue. Durant ces six mois de stage, j'ai pu aiguiller mon choix de carrière, je suis plus que jamais fasciné par le monde de la donnée et j'aimerais exercer dans ce domaine.

Bibliographie

- [1] A. SOLTANI, H. DEJNABADI, M. S. K. A. Real-world gait speed estimation using wrist sensor : A personalized approach. *IEEE Journal of Biomedical and Health Informatics* (2018).
- [2] ALESSANDRA MORANO, JINANE FATTOUCH, M. A. S. C. M. F. L. M. B. A. V. M. M. A. T. G. C. D. B. Perampanel as adjunctive therapy in highly refractory epilepsies : Real-world data from an italian tertiary care epilepsy centre.
- [3] ANDREAS EJUPI, STUDENT MEMBER, I. M. B. M. I. S. R. L. J. A. S. J. R. S. M. I., AND DELBAERE, K. Wavelet-based sit-to-stand detection and assessment of fall risk in older people using a wearable pendant device. *IEEE Transactions on Biomedical Engineering* (2016).
- [4] CIRA FUNDARÒ, ANNA GIARDINI, R. M. S. T. M. B. R. C. Motor and psychosocial impact of robot- assisted gait training in a real-world rehabilitation setting : A pilot study. *PLOS ONE* (February 2018).
- [5] FARID, L. Validation du système feetme monitor®. *Mémoire de stage de 3è année* (2018).
- [6] FRANCISCO ALEJANDRO RODRIGUEZ-LEAL, ROCCO HAASE, K. T. J. C. E. U. P. T. S. R. K., AND ZIEMSEN, T. Fampridine response in ms patients with gait impairment in a real-world setting : Need for new response criteria? *MULTIPLE SCLEROSIS JOURNAL MSJ* (June 2017).
- [7] KI HUN CHO, W. H. L. Effect of treadmill training based real-world video recording on balance and gait in chronic stroke patients : A randomized controlled trial. *Gait & Posture* (2013 September).
- [8] LINDSAY P. TOTH, SUSAN PARK, C. M. S. M. D. F. J. A. S., AND BASSETT, D. R. Video-recorded validation of wearable step counters under free-living conditions. *Official Journal of the American College of Sports Medicine* (January 2018).
- [9] MATTHEW M. ENGELHARD, STEPHEN D. PATEK, J. C. L. M. D. G. Real-world walking in multiple sclerosis : Separating capacity from behavior. *Gait & Posture* (<https://doi.org/10.1016/j.gaitpost.2017.10.015>).
- [10] MYLA D. GOLDMAN, MD, M. R. W. M. P. J. S. M. J. H. P. M. J. J. S. P. D. C. M. Clinically meaningful performance benchmarks in ms. *American Academy of Neurology* (November 2013).
- [11] SARA PIZZAMIGLIO, HASSAN ABDALLA, U. N. L. T. Neural predictors of gait stability when walking freely in the real-world. *Journal of NeuroEngineering and Rehabilitation* (2018).
- [12] SIDDHARTHA KHANDELWAL, N. W. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the marea gait database. *Gait & Posture* (2016 September).
- [13] WANG, W., AND ADAMCZYK, P. G. Analyzing gait in the real world using wearable movement sensors and frequently repeated movement paths. *Sensors* (April 2019).

Références

- [1] <http://www.thelin.net/laurent/plongee/pression.html>
- [2] https://fr.wikipedia.org/wiki/Donnée_aberrante
- [3] http://www.infectio-lille.com/Fichiers_infectio-lille/congres/os-2007/tsirstikolou.pdf
- [5] <http://nuxeo.edel.univ-poitiers.fr/nuxeo/site/esupversions/79a24c44-5587-4172-abb4-2eb7651d6ce5>
- [6] <http://www2.mat.ulaval.ca/fileadmin/Cours/STT-2300/Chap4H12.pdf>
- [7] <https://statistique-et-logiciel-r.com/anova-a-un-facteur-quand-les-hypotheses-ne-sont-pas-satisfaites/>
- [8] <https://perso.univ-rennes1.fr/valerie.monbet/ExposesM2/2013/anova.pdf>
- [9] https://en.wikipedia.org/wiki/Mean_absolute_percentage_error
- [10] <https://feetme.fr/fr/actualites>
- [11] https://fr.wikipedia.org/wiki/Test_de_Shapiro-Wilk
- [12] https://fr.wikipedia.org/wiki/Test_de_Student
- [13] http://www.numdam.org/article/RSA_1972__20_2_5_0.pdf
- [14] [https://fr.wikipedia.org/wiki/Corrélation_\(statistiques\)](https://fr.wikipedia.org/wiki/Corrélation_(statistiques))
- [15] https://fr.wikipedia.org/wiki/Corrélation_de_Spearman