

ÉCOLE POLYTECHNIQUE DE L'UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS

Département Informatique

64 avenue Jean Portalis

37200 Tours, France

Tél. +33 (0)2 47 36 14 14

[www.polytech.univ-tours.fr](http://www.polytech.univ-tours.fr)

**Projet Recherche & Développement  
2015-2016**

**Reconstitution virtuelle de la  
morphologie d'une omoplate malade**

**Tuteurs académiques**

Mohamed SLIMANE

Julien BERHOUET

Christian PROUST

**Étudiants**

Théo VOISIN (DI5)

# Liste des intervenants

Nom	Mail	Qualité
Théo VOISIN	<a href="mailto:theo.voisin@etu.univ-tours.fr">theo.voisin@etu.univ-tours.fr</a>	Étudiant DI5
Mohamed SLIMANE	<a href="mailto:mohamed.slimane@univ-tours.fr">mohamed.slimane@univ-tours.fr</a>	Tuteur académique, Département infomatique
Julien BERHOUET	<a href="mailto:julien.berhouet@gmail.com">julien.berhouet@gmail.com</a>	Tuteur académique
Christian PROUST	<a href="mailto:christian.proust@univ-tours.fr">christian.proust@univ-tours.fr</a>	Tuteur académique, Département infomatique

# Avertissement

Ce document a été rédigé par Théo VOISIN susnommé l'auteur.

L'école polytechnique de l'université François Rabelais de Tours est représentée par Mohamed SLIMANE, Julien BERHOUEY et Christian PROUST susnommés les tuteurs académiques.

Par l'utilisation de ce modèle de document, l'ensemble des intervenants du projet acceptent les conditions définies ci-après.

L'auteur reconnaît assumer l'entière responsabilité du contenu du document ainsi que toutes suites judiciaires qui pourraient en découler du fait du non respect des lois ou des droits d'auteur.

L'auteur atteste que les propos du document sont sincères et assument l'entière responsabilité de la véracité des propos.

L'auteur atteste ne pas s'approprier le travail d'autrui et que le document ne contient aucun plagiat.

L'auteur atteste que le document ne contient aucun propos diffamatoire ou condamnable devant la loi.

L'auteur reconnaît qu'il ne peut diffuser ce document en partie ou en intégralité sous quelque forme que ce soit sans l'accord préalable des tuteurs académiques.

L'auteur autorise l'école polytechnique de l'université François Rabelais de Tours à diffuser tout ou partie de ce document, sous quelque forme que ce soit, y compris après transformation en citant la source. Cette diffusion devra se faire gracieusement et être accompagnée du présent avertissement.

## Pour citer ce document :

Théo VOISIN, *Reconstitution virtuelle de la morphologie d'une omoplate malade*, Projet Recherche & Développement, Ecole Polytechnique de l'Université François Rabelais de Tours, Tours, France, 2015-2016.

```
@mastersthesis{
  author={VOISIN, Théo},
  title={Reconstitution virtuelle de la morphologie d'une omoplate malade},
  type={Projet Recherche \& Développement},
  school={Ecole Polytechnique de l'Université François Rabelais de Tours},
  address={Tours, France},
  year={2015-2016}
}
```

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>I Les états de l'art nécessaires et suffisants</b>	<b>2</b>
<b>1 Présentation de l'objet du problème : l'omoplate</b>	<b>3</b>
1 Généralités .....	3
2 La cavité glénoïde .....	4
3 Pathologies liées.....	4
4 Solutions proposées .....	5
<b>2 Modélisation de la glène : la régression linéaire</b>	<b>8</b>
1 Régression linéaire simple.....	8
1.1 Présentation .....	8
1.2 Estimation des paramètres .....	9
1.3 Analyse des résultats .....	10
1.4 Mise en pratique manuelle .....	12
1.5 Mise en pratique avec R.....	15
2 Régression linéaire multiple .....	17
2.1 Présentation .....	17
2.2 Estimation des paramètres .....	18
2.3 Analyse des résultats .....	18
2.4 Mise en pratique avec R.....	19
3 Régression linéaire multiple en mode pas à pas.....	21
3.1 Pas à pas ascendant.....	23
3.2 Pas à Pas descendant.....	23
3.3 Pas à pas progressif.....	23
3.4 Les critères et mise en pratique.....	24
4 Précisions sur l'analyse des résultats .....	25
4.1 La distance de Cook.....	25
4.2 Le COVRATIO .....	26
4.3 Test de Fisher - F-value.....	26

<b>3 Les outils</b>	<b>28</b>
1 Le langage R.....	28
1.1 Réalisation de scripts.....	28
1.1.1 Chargement de fichiers externes.....	29
1.1.2 Syntaxe de base.....	29
1.1.3 Typage des données.....	29
1.1.4 Matrices.....	30
1.1.5 Fonctions.....	30
1.1.6 Structures de contrôle.....	30
2 L'environnement RStudio.....	31
3 Blender.....	32
<b>4 Le travail de David Boas</b>	<b>33</b>
1 Mesures d'omoplates.....	33
2 La solution mise en œuvre.....	37
2.1 Les difficultés de mise en œuvre.....	37
2.2 Les scripts développés.....	37
2.3 Les modèles produits.....	37
2.4 Conclusions tirées et travail restant.....	38
<b>II Travail accompli et à venir</b>	<b>39</b>
<b>5 Vérifications</b>	<b>40</b>
1 Vérification.....	40
<b>6 Apprentissage de Shiny</b>	<b>42</b>
1 Architecture d'une application Shiny.....	43
1.1 Partie Client.....	43
1.2 Partie Serveur.....	43
<b>7 Ouverture sur le travail à venir</b>	<b>45</b>
<b>Conclusion - Partie Recherche</b>	<b>46</b>
<b>III Partie développement</b>	<b>47</b>
<b>8 Méthodologie de suivi et de gestion du projet</b>	<b>48</b>
<b>9 Mise en œuvre : outil de calculs des distances entre les points</b>	<b>50</b>
<b>10 Mise en œuvre : Tests sur les données réelles</b>	<b>52</b>
1 Structure de données.....	52
2 Comparaison d'un premier modèle avec celui de David.....	53

<b>11 Mise en œuvre : Production de modèles</b>	<b>54</b>
1 Points sur l'omoplate .....	54
2 Rappels sur les modèles .....	56
3 Quelles distances endogènes utiliser ? .....	56
4 Quelles distances utiliser pour chaque point exogène ? .....	58
5 Implémentation des tests statistiques .....	59
5.1 Test de Breusch-Pagan .....	59
5.2 Test de Breusch-Godfrey .....	60
<b>12 Analyse des résultats obtenus</b>	<b>62</b>
1 Glène totalement atteinte .....	63
1.1 Equations du point X1 .....	63
1.2 Equations du point X2 .....	63
1.3 Equations du point X3 .....	63
1.4 Equations du point X4 .....	63
1.5 Equations du point X14 .....	64
1.6 Equations du point X15 .....	64
1.7 Equations du point X16 .....	64
1.8 Equations du point X17 .....	64
1.9 Equations du point X18 .....	64
1.10 Equations du point X19 .....	65
1.11 Résultats .....	65
2 Glène partiellement atteinte : partie haute dite "saine", premier test .....	65
2.1 Equations du point X2 .....	65
2.2 Equations du point X3 .....	66
2.3 Equations du point X4 .....	66
2.4 Equations du point X14 .....	67
2.5 Equations du point X15 .....	67
2.6 Résultats .....	68
3 Glène partiellement atteinte : partie haute dite "saine", deuxième test .....	68
3.1 Equations du point X2 .....	68
3.2 Equations du point X3 .....	68
3.3 Equations du point X4 .....	68
3.4 Equations du point X14 .....	69
3.5 Equations du point X15 .....	70
3.6 Résultats .....	70
4 Test des équations .....	70
4.1 Tests sur le premier modèle .....	70
4.2 Tests sur le deuxième modèle .....	71
4.3 Comparaison des deux modèles .....	71
<b>13 Reproductibilité</b>	<b>74</b>
1 Préparation des données .....	74
2 Création de la matrice de distances .....	74
3 Utilisation des scripts .....	75
3.1 Application Shiny .....	76

14 Conclusion	77
Annexes	78

# Table des figures

## 1 Présentation de l'objet du problème : l'omoplate

1	Représentation d'une omoplate .....	3
2	L'inclinaison de la glène sur le plan frontal.....	4
3	Stades de Walch .....	5
4	Les implants utilisés.....	6

## 2 Modélisation de la glène : la régression linéaire

1	Disposition normale des résidus par rapport à la variable à expliquer .....	10
2	Disposition anormale : valeurs atypiques.....	11
3	Disposition anormale : non-linéarité .....	11
4	Les données utilisées pour l'exemple.....	12
5	Graphique du rendement en fonction de la température .....	13
6	Les données enrichies.....	13
7	Ajout de la droite calculée.....	14
8	Graphique des résidus.....	15
9	Le format de données du fichier .dat chargé .....	16
10	Résumé d'une régression sous R.....	17
11	Le fichier "voitures.dat" .....	20
12	Résultat de la régression multiple sous R.....	20
13	Résultat de la régression avec une sélection de variables.....	21
14	Résultat de la régression avec une sélection de variables.....	21
15	Résultat de la régression avec une sélection de variables.....	22
16	Représentation graphique d'un maximum local .....	24
17	Résultat de la commande summary sur le jeu de données Prostate .....	25

<b>3 Les outils</b>	
1 Le logo du langage R .....	28
2 Le logo de l'environnement RStudio.....	31
3 La fenêtre de RStudio .....	31
4 Logo de Blender.....	32
5 Scène de test sous Blender.....	32
<b>4 Le travail de David Boas</b>	
1 Les individus constituant la population .....	35
2 Points de l'omoplate - Vue latérale.....	35
3 Points de l'omoplate - Vue longitudinale.....	36
4 Les angles définis sur l'omoplate .....	36
5 Les modèles proposés par M. Boas.....	38
<b>5 Vérifications</b>	
1 Résultat de la recherche d'individus aberrants .....	40
2 Aperçu de la matrice de corrélation de la matrice V7 .....	41
<b>6 Apprentissage de Shiny</b>	
1 Aperçu des possibilités de Shiny (ici une carte présentant les revenus moyens.....	42
<b>8 Méthodologie de suivi et de gestion du projet</b>	
1 Diagramme de Gant du projet .....	48
2 Logo de Git .....	49
<b>9 Mise en œuvre : outil de calculs des distances entre les points</b>	
1 Schéma de principe de l'outil de calcul de distances .....	50
2 Extrait de la matrice de distances produite, au format CSV et ouvert par Microsoft Excel....	51
<b>10 Mise en œuvre : Tests sur les données réelles</b>	
1 Tableau de comparaison des équations produites .....	53
<b>11 Mise en œuvre : Production de modèles</b>	
1 Croquis de l'omoplate et de ses points (endogènes), vue postérieure.....	54
2 Croquis de l'omoplate et de ses points (endogènes), vue latérale .....	55
3 Croquis de l'omoplate (glène) et de ses points (exogènes), vue latérale .....	55
4 Rappel sur le principe .....	56
5 Exemple de fichier produit .....	57
6 Rappel sur le principe .....	58
7 Tracé des résidus pour la première régression .....	60
8 Tracé des résidus pour la deuxième régression .....	60

**12 Analyse des résultats obtenus**

1	Le fichier de configuration pour l'analyse .....	62
2	Capture d'écran de l'application réalisée avec le framework Shiny.....	66
3	Le fichier de configuration de la première analyse avec glène partiellement saine .....	67
4	Le fichier de configuration pour l'analyse avec glène partiellement saine, deuxième version .....	69
5	La matrice de résidus du premier modèle (valeurs en mm).....	71
6	La matrice de résidus du deuxième modèle (valeurs en mm).....	72
7	Le graphique en surface des résidus du modèle 1 .....	73
8	Le graphique en surface des résidus du modèle 2.....	73

**13 Reproductibilité**

1	Fichier d'un individu.....	75
2	Capture d'écran de l'application réalisée avec le framework Shiny.....	76

**Conclusion**

1	Table du test de Fisher avec alpha à 5 pourcent .....	78
2	Equations obtenues pour le point 1, glène totalement atteinte.....	79
3	Equations obtenues pour le point 2, glène totalement atteinte.....	80
4	Equations obtenues pour le point 3, glène totalement atteinte.....	81
5	Equations obtenues pour le point 4, glène totalement atteinte.....	82
6	Equations obtenues pour le point 14, glène totalement atteinte.....	83
7	Equations obtenues pour le point 15, glène totalement atteinte.....	84
8	Equations obtenues pour le point 16, glène totalement atteinte.....	85
9	Equations obtenues pour le point 17, glène totalement atteinte.....	86
10	Equations obtenues pour le point 18, glène totalement atteinte.....	87
11	Equations obtenues pour le point 19, glène totalement atteinte.....	88
12	Equations obtenues pour le point 2, glène partiellement atteinte .....	89
13	Equations obtenues pour le point 3, glène partiellement atteinte .....	90
14	Equations obtenues pour le point 4, glène partiellement atteinte .....	91
15	Equations obtenues pour le point 14, glène partiellement atteinte .....	92
16	Equations obtenues pour le point 15, glène partiellement atteinte .....	93



# Introduction

Ce rapport a pour but de présenter le Projet de Recherche et Développement auquel j'ai pu prendre part lors de ma 5<sup>e</sup> année d'étude au département informatique de Polytech Tours.

Ce projet s'inscrit dans un contexte plus large, nommé projet "Epaule" dont le but final est de pouvoir proposer aux patients atteints d'une omoplate abîmée (âge, maladie ...) une reconstruction mais également la pose d'une prothèse de glène adéquate. Les véritables enjeux de ce projet seront bien sûr présentés plus tard mais on peut dès à présent se rendre compte de la complexité de la reconstruction d'un élément sans en connaître la forme d'origine.

Dans ce projet plus vaste donc, s'insère la tâche qui m'a été confiée qui fut de mettre en place une série de tests afin de pouvoir produire les modèles mathématiques permettant la modélisation de la glène originelle à partir de mesures sur l'omoplate abîmée.

Je n'ai jamais été très attiré par ce côté mathématique que certains domaines de l'informatique comportent. J'ai pourtant choisi ce sujet qui m'attirait tout particulièrement. En effet, en dépit des moyens à mettre en œuvre, c'est surtout la finalité de ce projet qui m'a attiré. J'ai en effet toujours aimé le domaine médical à tel point que j'ai longtemps hésité entre un cursus médical et celui que je suis actuellement. Ainsi, choisir un des trois sujets impliqués dans le projet Epaule semblait être un choix évident pour moi.

Ce projet est en réalité divisé en deux parties distinctes. En effet, dans un premier temps, il s'agissait d'effectuer une partie de recherche, dite d'état de l'art tandis qu'une deuxième partie était consacrée au développement et à la mise en œuvre de la solution proposée. Ainsi, ce projet aura pour but, dans un premier temps, de présenter les outils (technologiques et mathématiques) qui seront utilisés tout du long du projet ainsi que ses enjeux. C'est seulement par la suite que nous nous attacherons à présenter la mise en œuvre de la solution proposée.

## Première partie

# Les états de l'art nécessaires et suffisants

Cette première partie a pour but de présenter les travaux existants sur ce sujet. En effet, ce projet est la suite logique du Projet de Fin d'Etudes réalisé par David Boas l'an passé. Cette partie aura également une fonction de présentation des outils qui seront utilisés pour ce projet (tant technologiques que théoriques). Ces présentations s'appuient évidemment sur des documents qui m'ont été fournis (ou trouvés lors de mes recherches) et qui seront cités en temps voulu.

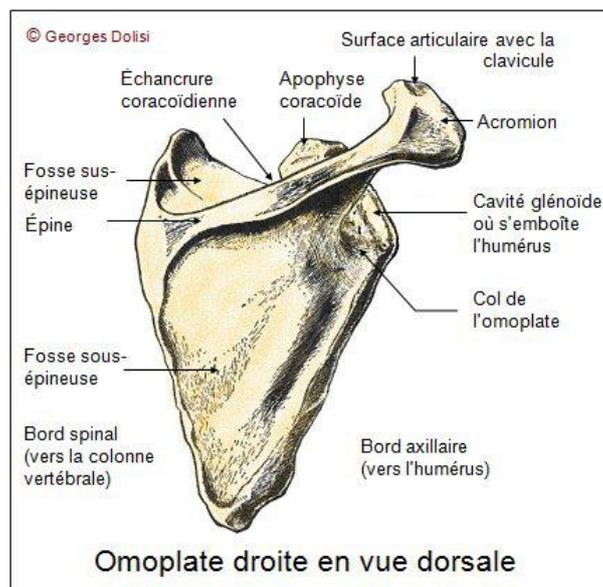
# 1

## Présentation de l'objet du problème : l'omoplate

Il est en effet nécessaire de bien comprendre la morphologie de l'omoplate afin d'obtenir un recul suffisant sur le problème auquel nous allons être confrontés.

### 1 Généralités

L'omoplate (ou scapula) est un os plat de forme globalement triangulaire situé au niveau de l'épaule (sur la partie supérieure et postérieure du thorax). Reliée à la clavicle via l'articulation acromio-claviculaire et à l'humérus qui s'insère dans sa cavité glénoïde (articulation scapulo-humérale), l'omoplate représente une des articulations les plus complexes du corps humain, permettant à l'Homme d'exécuter d'amples mouvements avec ses membres supérieurs.[3]



Source : <http://www.bio-top.net/Terminologie/G/gleno.htm>

Figure 1 – Représentation d'une omoplate

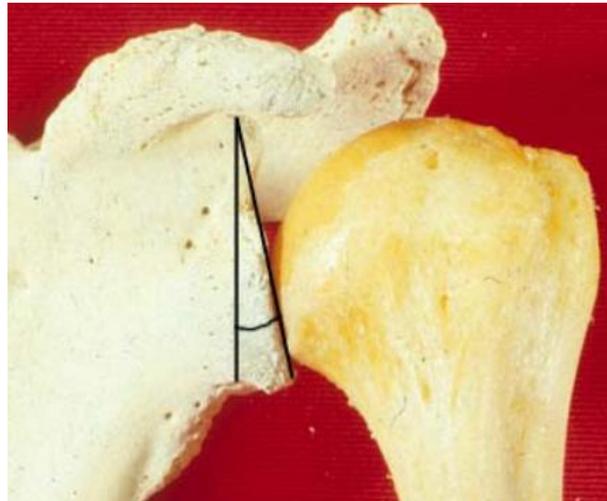
## 2 La cavité glénoïde

La **Figure 1** permet de se rendre compte de la localisation de la partie de l'omoplate qui nous intéresse réellement ici : la cavité glénoïde. Cette cavité qui est destinée à recevoir la tête de l'humérus, possède une forme globalement concave de manière à permettre la rotation de la tête humérale qui possède elle, une forme sphérique.

La tête humérale sera alors maintenue par un ensemble de quatre muscles nommé *coiffe des rotateurs*. Pour assurer une lubrification suffisante aux mouvements du bras et ainsi limiter autant que possible les frottements, des bourrelets (cartilage) sont présents sur la tête humérale ainsi que sur la cavité glénoïde.

La cavité glénoïde est également appelée glène. La morphologie de cette dernière peut varier énormément d'un individu à l'autre comme le montre la présentation "La glène osseuse - Aspect normal, anomalies morphologiques et pathologies à connaître" [5]. En effet, ce document indique que 71% des glènes sont piriformes (laissant les 29% restants avec une forme elliptique). De plus, une étude sur 412 scapula a montré également de grandes disparités de taille : entre 31.2mm et 50.1mm de hauteur pour 22.6mm à 41.5mm de largeur.

Mais les différences entre les individus ne s'arrêtent pas à la forme et à la taille de la glène. En effet, un autre paramètre est à prendre en compte : l'inclinaison de la glène.

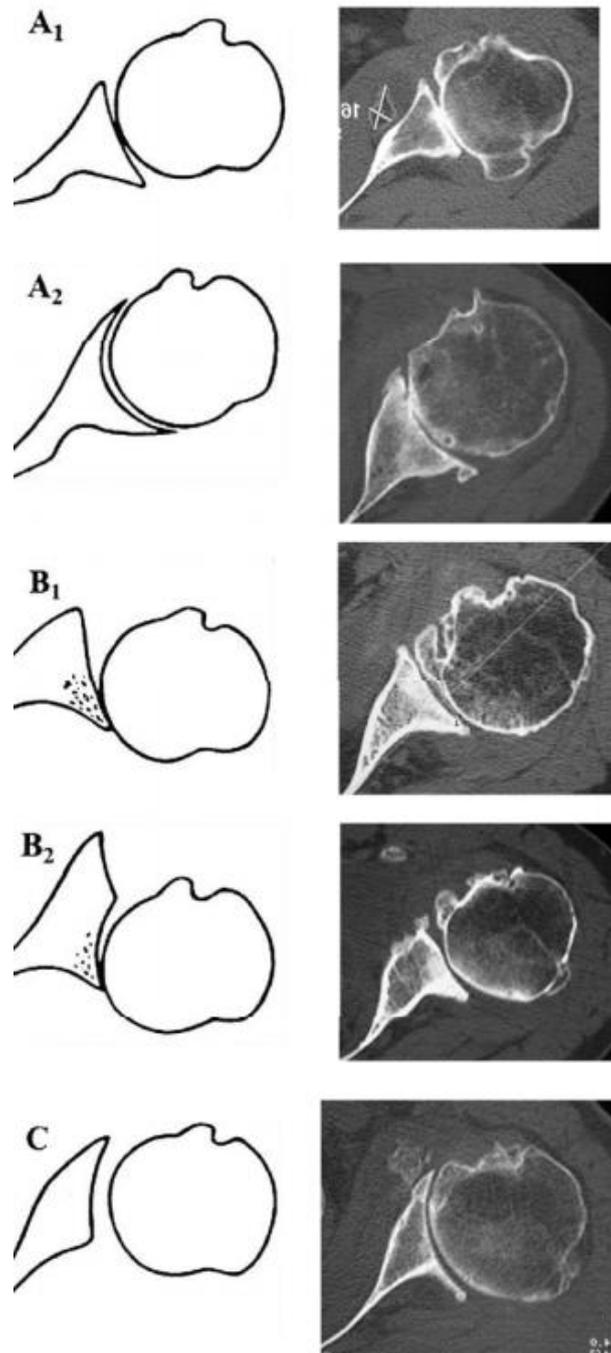


Source : La glène osseuse - Aspect normal, anomalies morphologiques et pathologies à connaître [5]

**Figure 2** – L'inclinaison de la glène sur le plan frontal

## 3 Pathologies liées

De nombreuses pathologies peuvent affecter la glène mais celle qui nous intéresse ici est l'omarthrose : une usure des cartilages de la glène (et généralement de la tête humérale) entraînant un contact direct entre l'omoplate et la tête humérale et donc des douleurs. En 1999, Walch a classifié cette usure en trois types comportant éventuellement plusieurs stades :



Source : La glène osseuse - Aspect normal, anomalies morphologiques et pathologies à connaître [5]

Figure 3 – Stades de Walch

Comme le montre la **Figure 3** trois types d'usure de la glène ont été répertoriés.

- Type A : L'usure de la glène est centrale et la tête humérale reste centrée dans la cavité.
- Type B : Usure postérieure de la glène, entraînant un décalage de la tête humérale. Dans le type B2, on observe la formation d'un second hémisphère.
- Type C : L'usure a entraîné un renversement de la glène.

## 4 Solutions proposées

L'usure de la glène précédemment présentée est une affection douloureuse, pour remédier à cela, des implants peuvent être posés. Lors de ces opérations, ce sont généralement deux implants qui sont utilisés :

l'un pour remplacer la glène disparue tandis que l'autre remplacera la tête humérale comme le montre la figure **Figure 4** :



Source : Chirurgie prothétique de l'épaule [[WWW9](#)]

**Figure 4** – *Les implants utilisés*

L'idée est d'implanter une pièce en plastique qui viendra se fixer au niveau de la cavité glénoïde, remplaçant l'ancienne glène disparue. Dans le même temps, la tête humérale osseuse sera réduite et l'humérus percé de manière à insérer une prothèse de tête humérale en titane, recréant alors la tête humérale abîmée. Ce type de prothèse est dit "anatomique".

En effet, d'autres solutions existent comme l'utilisation de prothèses inversées où, dans ce cas, la boule est implantée au niveau de l'omoplate tandis que la cavité recrée en lieu et place de l'ancienne tête humérale.

Les résultats d'une telle pose de prothèse sont encourageants, générant le plus souvent une disparition totale des douleurs ainsi que, après plusieurs mois de rééducation, une mobilité similaire à celle présente avant l'opération. Mais, cette opération bien que généralement peu risquée, est souvent un échec sur le long terme. Deux facteurs entrent en jeu ici. Le premier est l'usure des matériaux utilisés. En revanche, le deuxième facteur nous concerne directement puisqu'il s'agit de la qualité de la forme de la prothèse de glène ainsi que de la précision de sa pose.

Le problème est que cette prothèse doit reproduire le plus fidèlement possible les caractéristiques morphologiques que la glène originelle possédait lorsqu'elle était encore saine. De plus, la précision de la pose important énormément, le praticien sera d'autant plus efficace s'il connaît l'aspect général que la glène possédait à l'origine. Ceci est primordial pour la réussite de l'opération sur le long terme entraînant, le cas échéant, de nombreuses déconvenues comme par exemple la présence d'un risque accru que l'articulation scapulo-humérale se déboîte. Tout le problème réside donc ici, car si cette opération chirurgicale fonctionne très bien pour l'articulation de la hanche, elle est moins maîtrisée dans le cas de la glène.

Or, comment réussir à modéliser au mieux quelque chose qui n'existe plus ?

# 2

## Modélisation de la glène : la régression linéaire

*Cette présentation de la régression linéaire a été rédigée à la suite de la lecture de "Régression Linéaire avec R" [1]. D'autres sources ont bien sûr été utilisées et seront citées en temps voulu.*

Pour pallier ce problème, il a été choisi de créer un modèle permettant, grâce à des données connues (variables explicatives ou endogènes), de prédire les valeurs de variables inconnues (à expliquer ou exogènes). Ceci était l'objet du travail de David Boas lors de son Projet de Fin d'Etudes de l'an passé. Ce projet, dont le rapport [2] sera étudié lors de la section suivante, a fait appel à une technique bien connue en modélisation, la régression. Cette section s'attachera donc à décrire le fonctionnement de cette technique, étape indispensable afin de comprendre le travail effectué par M. Boas. Nous allons décrire la technique de modélisation par régression linéaire de manière incrémentielle, en partant de sa forme la plus simple avant d'arriver à celle finalement utilisée par M. Boas.

Précision sur la notation utilisée quant à la taille des ensembles manipulés :

- $n$  représente le nombre total d'individus présent dans notre population.
- $m$  représente le nombre total de variables endogènes à notre disposition (excluant donc l'exogène).
- $p$  représente le nombre de variables endogènes retenues pour le modèle. Cet ensemble est donc un sous ensemble, à un temps donné, de notre ensemble total contenant les  $m$  variables endogènes. A tout moment,  $p \leq m$ .

### 1 Régression linéaire simple

Une rapide présentation de la régression linéaire simple ainsi que l'exécution manuelle puis aidée du logiciel R vont nous permettre d'appréhender les concepts de base de la régression.

#### 1.1 Présentation

Le but d'une régression linéaire simple est d'expliquer une variable  $y$  (exogène, à expliquer) en fonction d'une variable  $x$  (endogène, explicative). Il s'agit d'une régression linéaire, le but est donc de trouver une formule du type

$$f(x) = \beta_0 + \beta_1 x$$

Ce qui donne donc

$$y = \beta_0 + \beta_1 x$$

On dégage alors une notion de causalité avec  $y$  dépend linéairement de  $x$ . Cette relation peut être

mesurée. En effet une régression linéaire n'a de sens que si  $x$  et  $y$  sont en relation, il peut donc être utile de tester cela grâce au calcul d'un coefficient de corrélation noté  $\rho$  :

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

Que l'on interprétera :

- $\rho = 1 \Rightarrow x$  et  $y$  proportionnelles
- $\rho = -1 \Rightarrow x$  et  $y$  inversement proportionnelles
- $\rho = 0 \Rightarrow x$  et  $y$  non corrélées

Il est important de noter que  $x$  et  $y$  ne sont pas forcément corrélés. En effet, une corrélation implique une relation symétrique où  $x$  dépend de  $y$  et inversement, ce qui n'est pas le but de la régression où il s'agit d'une relation asymétrique.

Il faut également ajouter à ceci la notion d'erreur  $\epsilon$  représentant l'erreur de la modélisation. Cette erreur peut venir de deux choses :

- D'une erreur de mesure sur  $x$
- Du fait que  $x$  seule est insuffisante pour expliquer  $y$

Donnant finalement :

$$y = \beta_0 + \beta_1 x + \epsilon$$

L'objectif va alors être, à partir d'un jeu de valeurs de variables endogènes  $x$  et exogènes  $y$ , de déterminer  $\beta_0$  et  $\beta_1$  de manière à prédire au mieux les futurs  $y$ .

Il est important de noter que dans le cas présent :

- les individus de la population sont indépendants (i.e. ils ne s'influencent pas entre eux)
- $\beta_0$  et  $\beta_1$  sont constants (i.e. ils ne changent pas d'un couple  $x, y$  à l'autre).
- $\epsilon \hookrightarrow N(0, \sigma^2)$

### 1.2 Estimation des paramètres

A partir du jeu de données fourni, le calcul de  $\beta_0$  et de  $\beta_1$  va pouvoir s'effectuer en cherchant à minimiser l'expression suivante :

$$\sum_{i=1}^n l(y_i - f(x_i))$$

Où  $n$  est le nombre de données,  $y_i$  la variable exogène,  $f(x_i)$  la prédiction de  $y_i$  et  $l()$  la fonction de coût. Ce qui revient donc à minimiser le coût total de l'erreur entre la valeur réelle et celle prédite à l'aide de nos  $\beta_0$  et  $\beta_1$ .

La fonction de coût se traduit généralement par l'utilisation des moindres carrés. Le but étant de trouver une droite minimisant l'erreur, la méthode des moindres carrés minimise la distance verticale de tous les points à la droite. Ceci se traduit par :

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Nous devons alors trouver des réalisations de nos estimateurs  $\widehat{\beta}_0$  et  $\widehat{\beta}_1$ . Nous posons :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

On peut alors poser les réalisations des estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  :

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Ce qui amène finalement à une estimation de  $y$  comme suit :

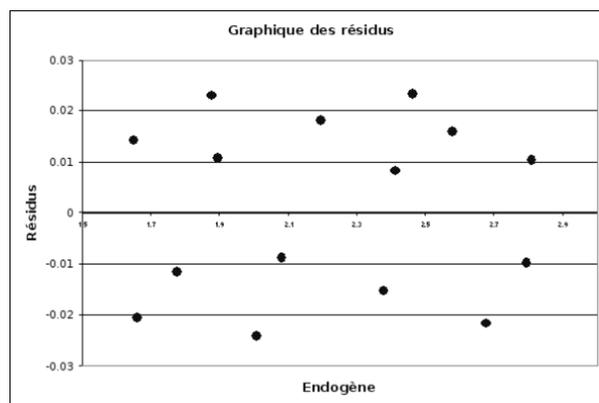
$$\hat{y}_i = b_0 + b_1 x_i$$

### 1.3 Analyse des résultats

La première chose qu'il est possible de faire est l'étude des résidus. Les résidus représentent en effet l'erreur d'estimation par rapport à la réalité, définie par l'expression suivante :

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

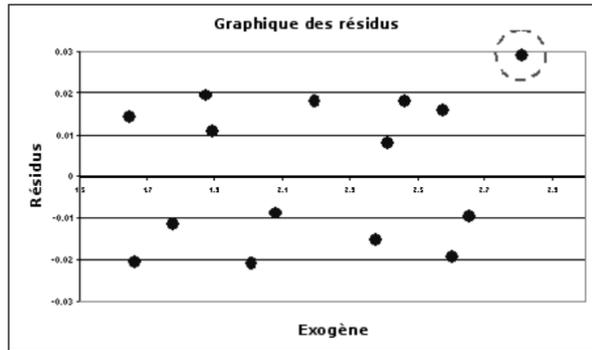
De là, une étude simplement graphique de ces résidus permet de rapidement détecter les potentiels problèmes.



Source : Pratique de la régression linéaire multiple - Diagnostic et sélection de variables [9]

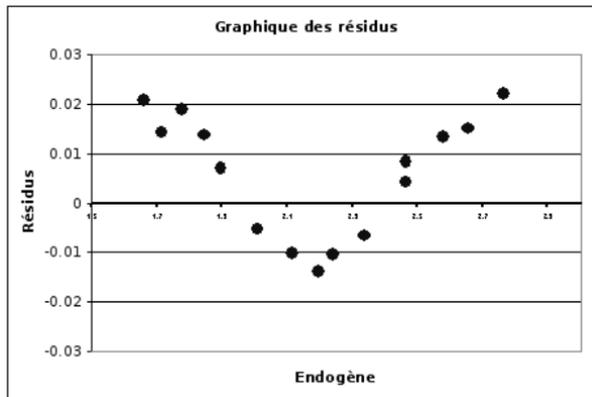
**Figure 1** – Disposition normale des résidus par rapport à la variable à expliquer

La **Figure 1** montre bien une disposition des résidus aléatoire et homogène, et ce quelle que soit la région de valeurs de l'endogène. Mais ces derniers peuvent ne pas être distribués de cette manière (localement divergents (**Figure 2**) ou encore manifestant une non-linéarité (**Figure 3**)).



Source : Pratique de la régression linéaire multiple - Diagnostic et sélection de variables [9]

Figure 2 – Disposition anormale : valeurs atypiques



Source : Pratique de la régression linéaire multiple - Diagnostic et sélection de variables [9]

Figure 3 – Disposition anormale : non-linéarité

En effet, la **Figure 2** amène à penser que le résidu entouré est induit par un individu incohérent ou même par une erreur de mesure. Il est alors possible, après vérification de la donnée, de supprimer cet individu.

De la même manière, la **Figure 3** présente elle une distribution des résidus laissant penser, de par sa forme générale, que la régression à utiliser n'est pas du type linéaire mais plutôt polynomiale ici.

Une donnée numérique peut également résumer la pertinence du modèle engendré : le coefficient de détermination noté  $R^2$ . Ce dernier se détermine grâce à trois analyses de variances : SCT, SCE et SCR :

$$SCT = SCE + SCR$$

$$SCE = \sum_{i=1}^n (\widehat{y}_i - \bar{y})^2$$

$$SCR = \sum_{i=1}^n (y_i - \widehat{y}_i)^2$$

Où, en d'autres termes, SCE représente la variabilité induite par le modèle, SCR la variabilité des erreurs du modèle et enfin SCT, la variabilité de la variable explicative.

On peut alors calculer  $R^2$  avec la formule :

$$R^2 = 1 - \frac{SCR}{SCT}$$

On observe donc que  $R^2 \in [0, 1]$  où :

- $R^2$  proche de 0 indique que le modèle est mal adapté
- $R^2$  proche de 1 indique que le modèle explique de manière convenable le phénomène.

On cherche donc évidemment à obtenir une valeur  $R^2$  proche de 1.

## 1.4 Mise en pratique manuelle

Dans cette section, nous allons essayer de mettre en pratique ce qui a été présenté ci-dessus de manière manuelle dans le tableur Excel (ceci m'a été fortement inspiré par le cours de Monsieur Rakotomalala et dont les [8] données d'exemple sont également tirées).

Cet exemple présente l'étude du rendement d'une expérience en fonction de la température ambiante.

Température	Rendement
X	Y
100	45
110	51
120	54
130	61
140	66
150	70
160	74
170	78
180	85
190	89

Figure 4 – Les données utilisées pour l'exemple

Traçons un graphique de ces données :

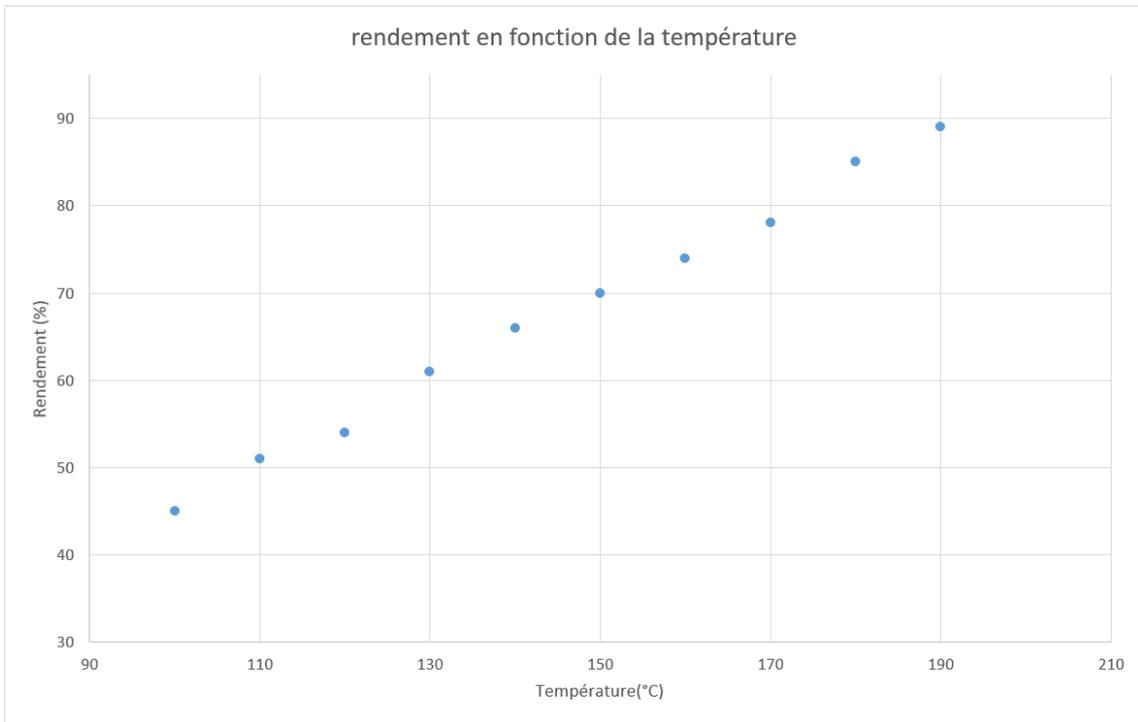


Figure 5 – Graphique du rendement en fonction de la température

On peut assez rapidement remarquer la linéarité des mesures affichées ici d'où la possibilité d'utiliser une régression linéaire pour modéliser le phénomène. Nous allons donc ajouter les colonnes  $S_{xy}$  et  $S_x^2$  représentant respectivement  $S_{xy}$  et  $s_x^2$  ainsi qu'un calcul de moyenne de X et Y représentant respectivement  $\bar{x}$  et  $\bar{y}$  :

Température	Rendement		
X	Y	$S_{xy}$	$S_x^2$
100	45	1003,5	2025
110	51	570,5	1225
120	54	332,5	625
130	61	94,5	225
140	66	6,5	25
150	70	13,5	25
160	74	100,5	225
170	78	267,5	625
180	85	619,5	1225
190	89	976,5	2025
Moyenne:		Somme:	
145	67,3	3985	8250

Figure 6 – Les données enrichies

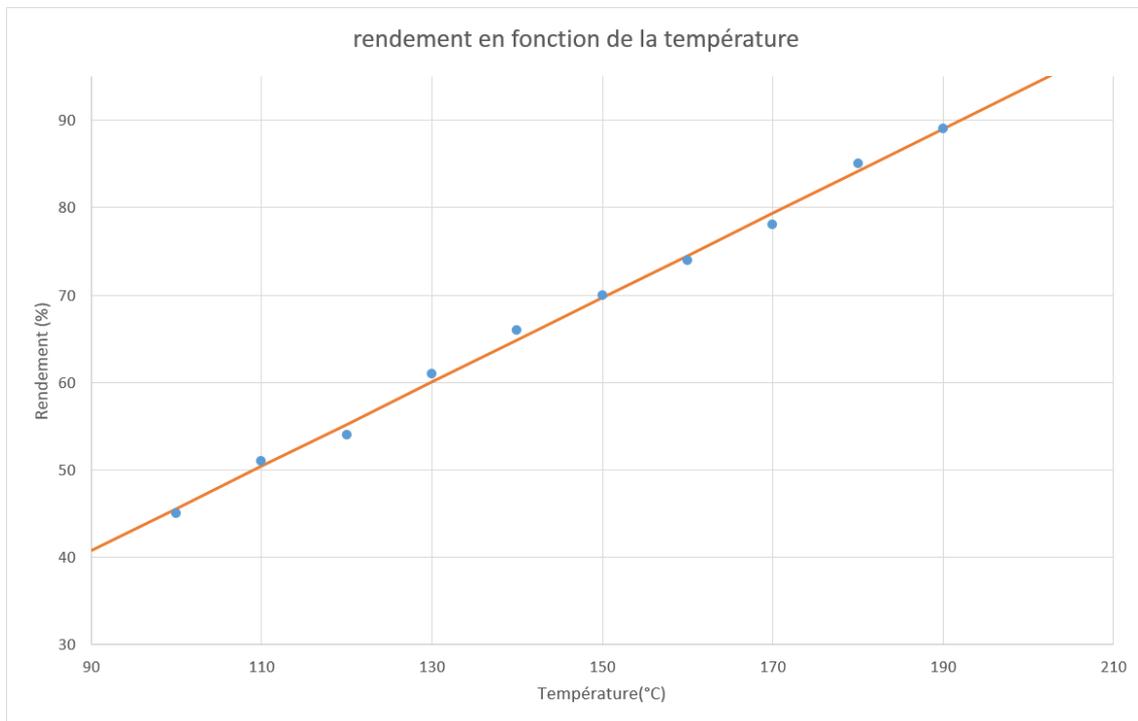
Les données ajoutées dans la Figure 6 vont nous permettre de produire des estimations pour  $\beta_0$  et  $\beta_1$  :

$$\beta_1 = \frac{S_{xy}}{S_x^2} = \frac{3985}{8250} \approx 0.48$$

$$\beta_0 = \bar{y} - \beta_1 \times \bar{x} = 67.3 - \beta_1 \times 145 \approx -2.74$$

Nous avons donc notre équation de droite découlant de cette régression où :

$$y = \beta_0 + x\beta_1 \approx 0.48 - 2.74x$$



**Figure 7** – Ajout de la droite calculée

En analysant de manière simplement visuelle la droite générée par le modèle, on peut d'ores et déjà obtenir une intuition de la qualité de la régression qui semble bonne ici. Une analyse également visuelle des résidus permet, elle, de vérifier l'homogénéité de la population :

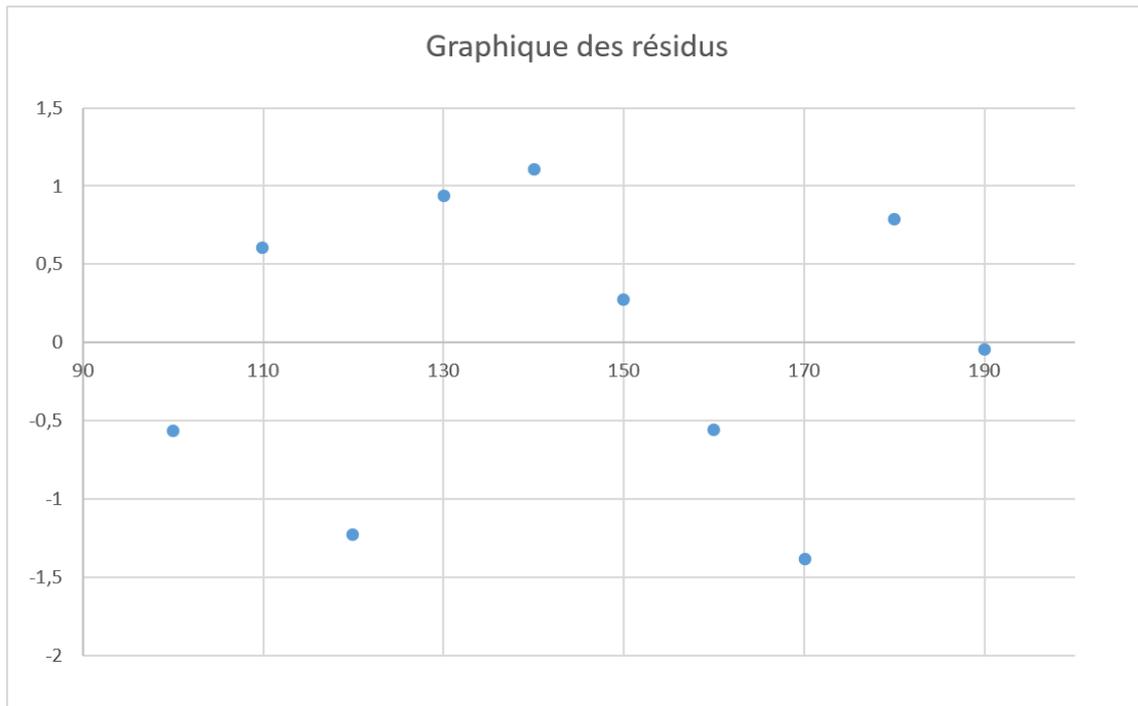


Figure 8 – Graphique des résidus

Sur ce graphe des résidus engendrés, aucune formation spécifique ne se dégage et les points semblent être répartis aléatoirement tout en restant globalement groupés. Ces observations tendent à laisser penser que la régression est cohérente avec le phénomène à modéliser et que, de plus, aucun individu ne fait figure de réelle exception.

Mais, comme vu précédemment, il existe une valeur numérique capable de quantifier la justesse du modèle : le coefficient de détermination  $R^2$  :

$$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \approx 45614.20$$

$$SCR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \approx 7.22$$

$$SCT = SCE + SCR \approx 45621.42$$

D'où :

$$R^2 = 1 - \frac{SCR}{SCT} \approx 0.99$$

Nous obtenons donc un coefficient de détermination extrêmement proche de 1, indiquant que la modélisation est plus que convenable.

## 1.5 Mise en pratique avec R

Cette section sera consacrée à l'utilisation du logiciel R, sur le même exemple, afin de constituer une première approche de cet outil.

La première étape est de sélectionner le working directory approprié :

```
1 setwd("C:\\Users\\theov\\Desktop\\")
```

Puis de charger le fichier contenant les données. Ce dernier est ici dans le format suivant (première ligne du nom des variables et suivantes constituées des individus) :

```
X Y
exp1 100 45
exp2 110 51
exp3 120 54
exp4 130 61
exp5 140 66
exp6 150 70
exp7 160 74
exp8 170 78
exp9 180 85
exp10 190 89
```

Figure 9 – Le format de données du fichier .dat chargé

La commande pour charger le fichier est :

```
1 data = read.table('tempRendement.dat')
```

Nous possédons maintenant une variable contenant les données de notre fichier. Il est possible de vérifier cela en écrivant simplement le nom de la variable dans l'interpréteur de commande. Il est également utile d'ajouter la variable au chemin de recherche de R pour une utilisation plus commode :

```
1 attach(data)
```

Cela fait, nous pouvons lancer la régression pour obtenir rapidement nos  $\beta_0$  et  $\beta_1$  :

```
1 regression=lm(Y~X)
```

Indiquant alors que la variable Y est à expliquer en fonction de la variable X. Le résultat de la régression est stocké dans la variable "regression". Un appel à la commande "coef" permet alors d'obtenir les coefficients calculés :

```
1 coef(regression)
```

Produisant alors les mêmes résultats que précédemment :

$$(intercept) \approx 0.48$$

$$X \approx -2.74$$

Où (*intercept*) représente donc notre ordonnée à l'origine et X le facteur de la droite de régression. Nous savons maintenant que la commande *lm* nous a permis d'obtenir les mêmes résultats que par un travail manuel ; il est donc évident que, comme montré précédemment, la régression trouvée est de qualité. Mais essayons tout de même d'obtenir cette information à partir de R :

Ainsi, l'utilisation de la commande suivante permet d'afficher un graphique des résidus comme présenté plus tôt :

```
1 plot(residuals(regression))
```

Un affichage des données complètes de la régression passe, lui, par la commande :

```
1 summary(regression)
```

Affichant

alors

:

```

Call:
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-1.3758 -0.5591  0.1242  0.7470  1.1152

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.73939    1.54650   -1.771    0.114
X             0.48303    0.01046  46.169 5.35e-11 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9503 on 8 degrees of freedom
Multiple R-squared:  0.9963,    Adjusted R-squared:  0.9958
F-statistic: 2132 on 1 and 8 DF,  p-value: 5.353e-11

```

Figure 10 – Résumé d'une régression sous R

De nombreuses informations sont affichées ici mais nous pouvons tout de même apercevoir des informations sur les résidus, les coefficients calculés ainsi que des statistiques sur la régression en général. Mais ce qui nous intéresse ici est la valeur de "Adjusted R-squared" qui est en réalité le coefficient de détermination précédemment calculé :  $\approx 0.99$ .

## 2 Régression linéaire multiple

Dans cette deuxième section, nous allons nous attacher à faire une présentation de la régression linéaire multiple agrémentée, comme précédemment, de la réalisation d'une régression sur un jeu de données de test. Peu d'informations supplémentaires seront données ici puisque la régression linéaire multiple n'est qu'une généralisation de la régression linéaire simple.

### 2.1 Présentation

La régression linéaire multiple est basée sur le même principe que la régression linéaire simple. A la différence près que pour expliquer notre variable exogène, nous n'avons plus une mais plusieurs variables endogènes. Ainsi, dans l'exemple précédent, nous étudions l'évolution du rendement d'une expérience selon la température ambiante en essayant d'en déduire un modèle. La température n'étant qu'une des variables possibles représentant les conditions de l'expérience, pourquoi alors ne pas ajouter d'autres variables rendant compte de ces conditions ? Comme la pression ou encore l'humidité ? En effet, intuitivement nous pourrions avancer que plus nous avons de variables endogènes, plus la justesse du modèle sera assurée. Même si cela est globalement vrai, nous verrons plus tard qu'il n'en est pas toujours le cas.

Précédemment, le calcul d'un modèle de régression linéaire simple était au final le calcul de l'équation d'une droite. Maintenant, nous avons  $m$  variables, le but est donc de trouver une équation décrivant un hyperplan. Ainsi, le but est cette fois-ci de déterminer les  $m$  coefficients ( $m$  coefficients sous-entendant qu'aucune sélection sur les variables endogènes à utiliser n'a encore été effectuée)  $\beta$  tels que :

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon$$

Posons les matrices  $Y$ ,  $X$ ,  $\beta$  et  $\epsilon$  respectivement matrice des valeurs de la variable exogène, des valeurs des variables endogènes, des coefficients et enfin des erreurs ( $m$  variables et  $n$  observations) :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$X = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{m1} \\ X_{12} & X_{22} & \cdots & X_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{mn} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}$$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

## 2.2 Estimation des paramètres

Encore une fois, l'estimation des paramètres  $\beta_m$  passe par l'utilisation d'estimateurs des moindres carrés. Ainsi, le but étant de trouver dans un premier temps les valeurs de la matrice  $\beta$ , nous posons par application des moindres carrés [WWW2] :

$$\beta = (X'X)^{-1}X'Y$$

Et donc l'estimation de  $Y$ ,  $\widehat{Y}$  :

$$\widehat{Y} = X\beta$$

Mais il manque à cela l'erreur  $\epsilon$  présentée plus tôt. Cette dernière est facilement calculable :

$$\epsilon = Y - \widehat{Y} = Y - X\beta$$

## 2.3 Analyse des résultats

L'analyse des résultats d'une régression multiple fonctionne sur le même principe que la régression linéaire simple. L'idée est en effet encore une fois d'analyser les résidus de la régression.

Un graphe des résidus permet dans un premier temps de repérer les individus anormaux exactement de la même manière que présenté plus tôt (subsection 1.3).

Par la suite, le calcul du coefficient de détermination  $R^2$  utilise les analyses de variance également déjà présentées : SCT, SCE et SCR. Or ce coefficient de détermination a tendance à augmenter lorsque le nombre de variables endogènes  $m$  (ou  $p$  si une sélection sur les endogènes a été effectuée) augmente. Ceci représente un problème car comme nous l'avons vu, un  $R^2$  significativement grand est synonyme d'une qualité de régression correcte. Or, si ce nombre dépend également du nombre de variables endogènes utilisées, les conclusions sur la qualité de la régression s'en trouvent faussées.

C'est pourquoi nous définissons le coefficient de détermination ajusté : le  $R^2$  ajusté qui permet de s'affranchir de ce problème, avec  $m > 1$  le nombre de variables endogènes :

$$SCT_a = SCE + SCR$$

$$SCE_a = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m-1}$$

$$SCR_a = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{m-1}$$

Amenant alors à :

$$R_a^2 = 1 - \frac{SCR_a}{SCT_a}$$

## 2.4 Mise en pratique avec R

Pour la mise en pratique de ce que nous venons de voir, j'ai choisi d'utiliser un exemple disponible ici [[WWW7](#)]. Cet exercice met à disposition un jeu de données sur un ensemble de voitures, comme leur prix, leur cylindrée, puissance etc.

Le but de notre régression sera de déterminer le prix d'une voiture.

On peut d'ores et déjà remarquer que la variable "finit" est un facteur discret. En effet, à l'instar de ce que nous pourrions rencontrer avec les omoplates d'homme ou de femme (modificateur binaire), la variable finit devra se traduire par un modificateur ternaire de facteur car  $finit \in \{0, 1, 2\}$ . Il nous faut donc transformer cette variable "finit" en facteur :

```
1 finit=factor(finit)
```

Lancer une régression multiple avec R utilise la même commande "lm". En revanche, il faut cette fois-ci spécifier plusieurs variables endogènes :

```
1 model=lm(prix~cyl+puiss+long+larg+poids+vites+finit)
```

Donnant le résultat :

Plusieurs choses sont à noter ici :

- Les différents coefficients ainsi que l'ordonnée à l'origine
- Le facteur "finit" fonctionne de manière particulière, comme annoncé. C'est à dire que, si on omettait les autres facteurs, l'équation serait du type :

$$y = \text{fact1} \times 3548.52 + \text{fact2} \times 9979.70$$

Où, selon le cas, (fact1 = 1 ET fact2 = 0) OU (fact1 = 0 ET fact2 = 1)

```

      cyl  puiss long larg poids vites finit prix
Alfasud-ti-135 1350 79 393 161 870 165 1 30570
Audi-100-1     1588 85 468 177 1110 160 2 39990
Simca-1307-gls 1294 68 424 168 1050 152 0 29600
Citroen-gc-club 1222 59 412 161 930 151 0 28250
Fiat-132-1600gls 1585 98 439 164 1105 165 1 34900
Lancia-beta-1300 1297 82 429 169 1080 160 2 35480
Peugeot-504    1796 79 449 169 1160 154 1 32300
R16-t1        1565 55 424 163 1010 140 1 32000
R30-ts        2664 128 452 173 1320 180 2 47700
Toyota-corolla 1166 55 399 157 815 140 0 26540
Alfetta-1.66   1570 109 428 162 1060 175 2 42395
Princess-1800-hl 1798 82 445 172 1160 158 2 33990
Datsun-2001    1998 115 469 169 1370 160 2 43980
Taurus-2000-gl 1993 98 438 170 1080 167 1 35010
Rancho        1442 80 431 166 1129 144 2 39450
Mazda-9295    1769 83 440 165 1095 165 0 27900
Opel-rekord-1 1979 100 459 173 1120 173 1 32700
Lada-1300     1294 68 404 161 955 140 0 22100

```

Figure 11 – Le fichier "voitures.dat"

```

call:
lm(formula = prix ~ cyl + puiss + long + larg + poids + vites +
    finit)

Residuals:
    Min       1Q   Median       3Q      Max
-4628  -1374    339    1636   3691

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 35982.865  32658.557   1.102  0.29914
cyl           2.849     4.490   0.634  0.54158
puiss        75.547    136.315   0.554  0.59294
long         49.590     95.098   0.521  0.61463
larg        -261.573    319.632  -0.818  0.43429
poids         1.194     18.302   0.065  0.94942
vites        18.206    161.311   0.113  0.91262
finit1       3548.519    2167.987   1.637  0.13610
finit2       9979.695    2799.231   3.565  0.00607 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3116 on 9 degrees of freedom
Multiple R-squared:  0.881,    Adjusted R-squared:  0.7752
F-statistic: 8.328 on 8 and 9 DF,  p-value: 0.002277

```

Figure 12 – Résultat de la régression multiple sous R

$$\text{fact} = i \Rightarrow \text{fact}_j = 0 \forall j \neq i \text{ et } \text{fact}_i = 1$$

- Le coefficient de détermination  $R^2$  est supérieur au  $R_a^2$  ce qui est logique d'après ce que nous avons vu.

### 3 Régression linéaire multiple en mode pas à pas

La régression multiple permet d'utiliser plusieurs variables endogènes pour déterminer une exogène. Mais, pour reprendre notre exemple précédent, toutes ces données sur les voitures sont-elles réellement nécessaires? Que se passerait-il si nous n'utilisions qu'une partie de ces données. Le but de cette manœuvre étant bien évidemment d'augmenter la qualité de notre régression. Une illustration de cela peut être faite en lançant la régression suivante :

```
1 model=lm(prix~puiss+finit)
```

Produisant :

```
Call:
lm(formula = prix ~ puiss + finit)

Residuals:
    Min       1Q   Median       3Q      Max
-4975.4 -1463.2   278.1  1420.5  3294.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 17485.39   2948.70   5.930 3.67e-05 ***
    puiss      141.03    40.51   3.481 0.003670 **
    finit1     3463.88   1772.04   1.955 0.070882 .
    finit2     9220.82   1992.84   4.627 0.000392 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2660 on 14 degrees of freedom
Multiple R-squared:  0.8651,    Adjusted R-squared:  0.8362
F-statistic: 29.92 on 3 and 14 DF,  p-value: 2.403e-06
```

Figure 13 – Résultat de la régression avec une sélection de variables

On remarque assez rapidement que, malgré un nombre de variables endogènes réduits, le coefficient de détermination a augmenté (notamment le coefficient ajusté). La sélection des variables adéquates a donc un rôle important dans une régression linéaire multiple et c'est justement ce mode pas à pas qui permet de sélectionner les variables les plus intéressantes.

Avant de commencer la présentation d'une régression en mode pas à pas, j'aimerais présenter la manière avec laquelle j'ai choisi les deux variables. En effet, j'ai utilisé la matrice de corrélation des données pour choisir ces variables. La matrice de corrélation s'obtient de cette manière sous R :

```
1 cor(voit)
```

Produisant :

```
      cyl    puiss    long    larg    poids    vites    finit    prix
cyl  1.000000  0.7966277  0.7014619  0.6297572  0.7889520  0.6649340  0.4438897  0.6385812
puiss 0.7966277  1.0000000  0.6413624  0.5208320  0.7652930  0.8443795  0.6200582  0.7987004
long  0.7014619  0.6413624  1.0000000  0.8492664  0.8680903  0.4759285  0.5626258  0.6437569
larg  0.6297572  0.5208320  0.8492664  1.0000000  0.7168739  0.4729453  0.5674615  0.5466494
poids 0.7889520  0.7652930  0.8680903  0.7168739  1.0000000  0.4775956  0.6327987  0.7532948
vites 0.6649340  0.8443795  0.4759285  0.4729453  0.4775956  1.0000000  0.4217126  0.5817597
finit 0.4438897  0.6200582  0.5626258  0.5674615  0.6327987  0.4217126  1.0000000  0.8633155
prix  0.6385812  0.7987004  0.6437569  0.5466494  0.7532948  0.5817597  0.8633155  1.0000000
```

Figure 14 – Résultat de la régression avec une sélection de variables

Pour une représentation plus visuelle nous pouvons utiliser le package "corrplot" :

```
1 corrplot(cor(voit), type="full", method="number")
```

Produisant :

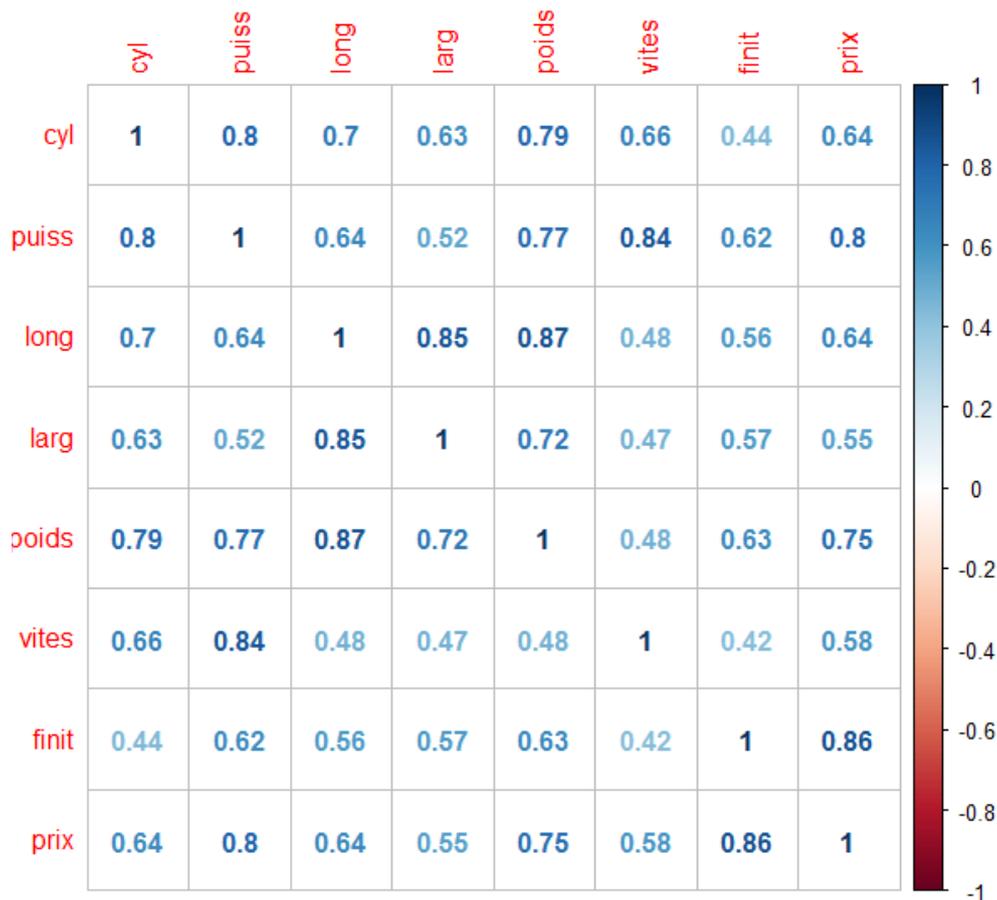


Figure 15 – Résultat de la régression avec une sélection de variables

Il devient alors assez intuitif d’analyser ces corrélations et donc de remarquer rapidement que la puissance et la finition semblent avoir une grande importance dans l’établissement du prix. Mais comment alors savoir s’il faut ajouter également le poids qui semble être pourtant non négligeable ?

Une solution serait de tester toutes les combinaisons de variables et de garder celle maximisant un certain critère (le coefficient de détermination ajusté par exemple). Ceci est faisable dans notre exemple, mais devient beaucoup plus problématique sur des jeux de données plus grands. C’est pourquoi la technique de régression pas à pas est utile. Cette méthode est présentée dans le livre de Monsieur Rakotomalala, *Pratique de la régression linéaire multiple*[7] mais nous a également été présentée par M. Cardot dans le cours de reconnaissance des formes.

De fait, les problématiques de reconnaissance des formes reposent également sur la sélection des variables décrivant la forme. Trop peu de variables descriptives étant évidemment insuffisant. Mais à l’inverse, et ce de manière contre-intuitive, trop de variables peut également nuire aux résultats (ce que nous allons pouvoir observer à la fin de cette section).

La méthode de régression pas à pas comporte plusieurs algorithmes qui peuvent être utilisés selon les cas. Pour décrire ces algorithmes, nous allons utiliser la notation suivante :

$X$ , l’ensemble des variables endogènes de départ (qui contient donc  $m$  variables)

$Z$ , l’ensemble vide au départ, qui contiendra les variables endogènes sélectionnées en sortie d’algorithme

$u$ , une variable endogène avec à tout moment :  $\forall u, u \in X \cup u \in Z$

C, le critère de choix

### 3.1 Pas à pas ascendant

Dans la méthode ascendante nous partons d'un ensemble  $X$  contenant toutes les variables endogènes. A chaque pas nous ajoutons une variable. Cette variable est choisie de manière à maximiser le gain sur le critère de choix :

```

1      X := Ensemble contenant les m variables endogènes
2      Z := Ensemble vide
3      tant que (il existe u tel que Z + u améliore C(Z)) faire:
4          choisir le meilleur u dans X;
5          retirer u de X;
6          ajouter u dans Z;
7          recalculer C(Z);
8      fin tant que
9      Z := Ensemble contenant les p variables endogènes sélectionnées

```

### 3.2 Pas à Pas descendant

Cette méthode fonctionne exactement de manière inverse. En effet, l'initialisation des ensembles est toujours la même, mais nous allons cette fois-ci conserver en sortie l'ensemble  $X$ . A chaque pas, nous retirons la variable de  $X$  qui induit le meilleur gain sur le critère de choix :

```

1      X := Ensemble contenant les m variables endogènes
2      Z := Ensemble vide
3      tant que (il existe u tel que X - u améliore C(X)) faire:
4          choisir le meilleur u dans X;
5          retirer u de X;
6          ajouter u dans Z;
7          recalculer C(X);
8      fin tant que
9      X := Ensemble contenant les p variables endogènes sélectionnées

```

### 3.3 Pas à pas progressif

L'idée de cette méthode est de combiner les deux précédentes. A chaque tour, nous ajoutons la variable qui maximise le critère de sélection et, dans le même tour, nous supprimons une variable, s'il en existe une, qui permettrait d'améliorer le critère de sélection. Ceci permet d'éviter le cas où, une variable était "utile" seule mais ne l'est plus en présence d'autres variables et donc d'éviter un phénomène de maximum local :

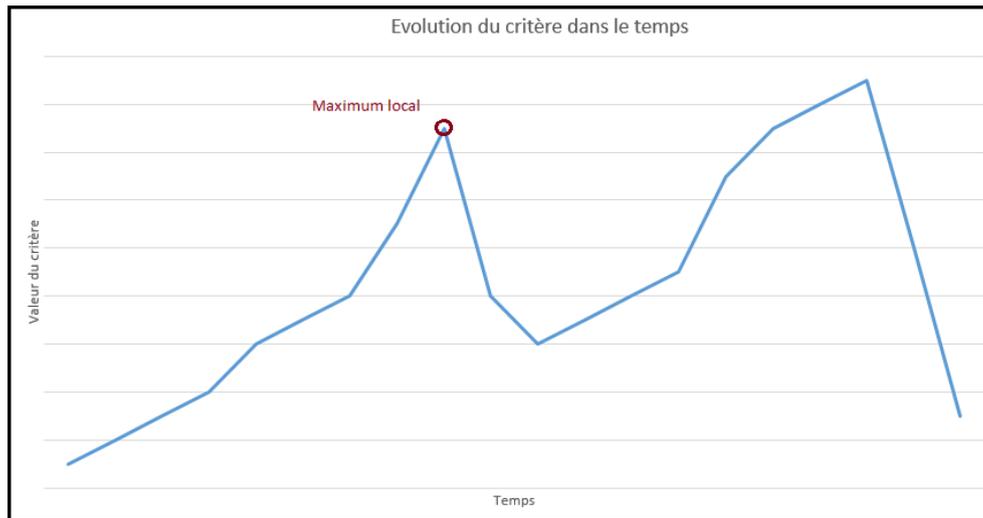


Figure 16 – Représentation graphique d'un maximum local

En effet, dans le graphique ci-dessus, on voit bien que la méthode ascendante se serait arrêté au maximum local. Tandis que la méthode progressive a plus de chances d'obtenir un résultat supérieur. Cette dernière se traduit simplement par :

```

1      X := Ensemble contenant les m variables endogènes
2      Z := Ensemble vide
3      tant que (il existe u tel que Z + u améliore C(Z)) faire:
4          choisir le meilleur u dans X;
5          retirer u de X;
6          ajouter u dans Z;
7          recalculer C(Z);
8      tant que (il existe u tel que Z - u améliore C(Z)) faire:
9          choisir le meilleur u dans Z;
10         retirer u de Z;
11         ajouter u dans X;
12         recalculer C(Z);
13     fin tant que
14 fin tant que
15     Z := Ensemble contenant les p variables endogènes sélectionnées

```

On comprend assez rapidement que cette méthode est plus lourde que les deux précédentes mais présente l'avantage de potentiellement obtenir de meilleurs résultats. Une variante existe néanmoins, ne supprimant qu'une seule variable par pas.

### 3.4 Les critères et mise en pratique

Nous avons jusqu'ici abordé la régression en mode pas à pas en expliquant que le but était de minimiser le nombre de variables endogènes ainsi que de, bien sûr, maximiser l'exactitude du modèle. C'est pourquoi nous *choisissons le meilleur u dans X*. Tout le problème réside donc dans la sélection du *meilleur*. Or pour faire cette sélection, il faut pouvoir comparer les modèles entre eux.

A cette fin, plusieurs critères ont été mis au point. Les plus connus sont le AIC et le BIC. Afin de les mettre en pratique, j'ai suivi un cours (disponible ici [WWW3]) proposé par l'université de Toulouse.

Il s'agit donc dans un premier temps de rapatrier et d'installer le package *lasso2* disponible ici : [WWW6]. Dans ce package, se trouve le jeu de données qui nous intéresse : *Prostate*. Ce dernier contient les données d'une étude ayant mis en relation le taux d'un antigène spécifique avec les mesures physiques sur 97 individus étant sur le point de subir une prostatectomie :

```
> summary(Prostate)
      lcavol      lweight      age      lbph
Min.   :-1.3471   Min.    :2.375   Min.   :41.00   Min.   :-1.3863
1st Qu.: 0.5128   1st Qu.:3.376   1st Qu.:60.00   1st Qu.: -1.3863
Median : 1.4469   Median :3.623   Median :65.00   Median : 0.3001
Mean   : 1.3500   Mean    :3.653   Mean   :63.87   Mean   : 0.1004
3rd Qu.: 2.1270   3rd Qu.:3.878   3rd Qu.:68.00   3rd Qu.: 1.5581
Max.   : 3.8210   Max.    :6.108   Max.   :79.00   Max.   : 2.3263

      svi      lcp      gleason      pgg45
Min.   :0.0000   Min.   :-1.3863   Min.   :6.000   Min.   : 0.00
1st Qu.:0.0000   1st Qu.: -1.3863   1st Qu.:6.000   1st Qu.: 0.00
Median :0.0000   Median : -0.7985   Median :7.000   Median :15.00
Mean   :0.2165   Mean    :-0.1794   Mean   :6.753   Mean   :24.38
3rd Qu.:0.0000   3rd Qu.: 1.1787   3rd Qu.:7.000   3rd Qu.:40.00
Max.   :1.0000   Max.    : 2.9042   Max.   :9.000   Max.   :100.00

      lpsa
Min.   :-0.4308
1st Qu.: 1.7317
Median : 2.5915
Mean   : 2.4784
3rd Qu.: 3.0564
Max.   : 5.5829
```

Figure 17 – Résultat de la commande `summary` sur le jeu de données `Prostate`

Nous pouvons appliquer les algorithmes vus précédemment, combinés avec le critère AIC de manière assez simple ici par la commande :

```
1 modselect_b=stepAIC(modlin, ~., trace=TRUE)
```

Par défaut, l'algorithme utilisé est celui dit du *pas à pas progressif* mais il est possible de modifier cela en ajoutant le paramètre :

```
1 direction=c("forward") (pour un pas à pas ascendant)
2 direction=c("backward") (pour un pas à pas descendant)
3 direction=c("both") (pour un pas à pas progressif)
```

## 4 Précisions sur l'analyse des résultats

L'analyse d'une régression linéaire ne passe pas forcément par une simple analyse du coefficient de détermination. En effet, comme nous avons pu le voir, des informations supplémentaires sont données par R après une régression et il est important de les comprendre.

### 4.1 La distance de Cook

Nous avons fait référence plusieurs fois à l'analyse des résidus permettant de détecter des individus influençant de manière anormalement forte le modèle : [subsection 1.3](#). Or cette analyse est purement visuelle et informelle puisque reposant sur la seule appréciation de l'évaluateur. La distance de Cook permet elle de corriger cela en apportant un élément numérique quantifiant cette influence. Le principe de base est assez simple puisqu'il s'agit dans un premier temps d'effectuer la régression avec l'ensemble des données à notre disposition ce qui représentera notre modèle de référence B. Ensuite, pour chaque individu  $i$  nous allons calculer la distance entre le modèle de référence B et celui privé de la variable  $i$  :  $B \setminus \{i\}$ .

Le calcul de cette distance s'effectue grâce à la formule suivante :

$$D_i = \frac{\sum_{j=1}^n (\widehat{Y}_j - \widehat{Y}_{j(i)})^2}{p \times \text{MSE}}$$

Où  $i$  est l'individu retiré du modèle,  $\widehat{Y}_j$  le modèle de référence,  $\widehat{Y}_{j(i)}$  le modèle privé de l'individu  $i$ ,  $n$  le nombre d'individus du modèle, MSE l'erreur quadratique moyenne du modèle et enfin  $p$  le nombre de variables endogènes utilisées.

L'interprétation de la valeur de la distance de Cook varie. En effet, le moyen le plus simple est de décréter qu'un individu  $i$  est anormalement influent dès lors que :

$$D_i > 1$$

Mais, elle est jugée trop permissive (i.e. cette limite à tendance à laisser passer des individus qui devraient être considérés comme anormalement influents (ou en tout cas nécessitant une vérification) [9] d'où la proposition :

$$D_i > \frac{4}{n-p-1}$$

Un individu montré influent par la distance de Cook devra alors faire l'objet d'une vérification toute particulière afin de déterminer s'il est pertinent de le garder au sein du modèle.

## 4.2 Le COVRATIO

Le COVRATIO permet lui de déterminer l'influence de l'individu  $i$  sur la variance des résultats estimés et ainsi avoir une idée de la précision qu'apporte l'ajout d'un individu :

$$\text{COVRATIO}_i = \frac{\text{var}(\widehat{y} - i)}{\text{var}(\widehat{y})}$$

Ainsi, si le COVRATIO est supérieur à 1, alors la variance a diminué, indiquant donc que la mesure est plus précise. A l'inverse, un résultat inférieur à 1 indique plutôt un accroissement de la variance et donc que l'individu était "utile". Mais, comme précisé par M. Rakotomalala [9], une diminution excessive de la variance peut également indiquer que l'individu est anormalement influent.

## 4.3 Test de Fisher - F-value

Le test de Fisher s'articule autour des hypothèses suivantes :

$H_0$  = Le modèle de base et celui privé des coefficients (sauf "intercept") sont égaux

$H_1$  = Le modèle de base et celui privé des coefficients (sauf "intercept") sont différents

Basiquement ce test permet donc de déterminer si l'utilisation des coefficients autres que "intercept" est utile. Ainsi, tandis que le coefficient de détermination estime la qualité de la régression, le test de Fisher permet de donner une validité statistique à cette estimation. La F-value est donnée automatiquement par R à l'issue d'une régression, mais voici comment l'obtenir (nous allons réutiliser SCE et SCR définis plus tôt) avec  $n$  le nombre d'individus et  $p$  le nombre de variables endogènes utilisées (basé sur la présentation faite par M. D. Jost [WWW5]) :

$$\text{DFM} = p - 1 (\text{degrés de liberté corrigé du modèle})$$

$$\text{DFE} = n - p (\text{degrés de liberté de l'erreur})$$

$$\text{MSM} = \frac{\text{SCE}}{\text{DFM}} \text{ (Moyenne des carrés du modèle)}$$

$$\text{MSE} = \frac{\text{SCR}}{\text{DFE}} \text{ (Moyenne des carrés de l'erreur)}$$

Et enfin la valeur de notre F-value :

$$\text{F-value} = \frac{\text{MSM}}{\text{MSE}}$$

Finalement, en lisant la table disponible en annexe 1 ([Figure 1](#)) nous pouvons déterminer l'intervalle de confiance à 95%. Il faut pour cela poser  $v_1 = \text{DFM}$  et  $v_2 = \text{DFE}$ . La valeur  $x$  ainsi obtenue place l'intervalle de confiance de F-value à  $[0, x]$ . Ainsi si  $\text{F-value} \notin [0, x]$ , alors nous rejetons notre hypothèse  $h_0$  (à 95% de confiance).

Reste alors à chiffrer la probabilité de rejet de  $h_0$ . Cela se fait par le calcul de la P-value associée au modèle. C'est cette valeur qui apparaît à droite de la F-value en résultat de régression sur R ([Figure 13](#)).

# 3

## Les outils

Même si dans cette première partie de mon projet, il s'agissait essentiellement d'effectuer des recherches sur les outils mathématiques entrant en jeu ainsi que de s'imprégner du travail précédemment effectué sur le sujet, il m'a tout de même été demandé de prendre en main certains outils. Ces outils seront donc présentés dans ce chapitre.

### 1 Le langage R



Figure 1 – Le logo du langage R

R est basé sur le plus ancien langage S (mais aussi sémantiquement inspiré par Scheme). Développé dans un premier temps dans les années 1990 par Ross Ihaka et Robert Gentleman, il est principalement destiné à permettre de nombreuses opérations mathématiques et plus précisément statistiques [WWW4]. L'avantage étant un regroupement et une manipulation des données simplifiées. La visualisation graphique de ces données est également facilitée par les outils directement inclus.

Sa distribution sous licence GPL (gratuit et code source accessible) a permis la formation d'une importante communauté participant activement à son amélioration via des mises à jour régulières. La communauté est également très présente du côté de la publication de packages qui permettent de faciliter toute sorte de traitements ainsi que d'améliorer la visualisation graphique dans certains cas d'utilisation.

De part son positionnement résolument orienté vers le traitement de données (data mining etc.) le langage R s'est rapidement imposé dans la communauté scientifique.

#### 1.1 Réalisation de scripts

Lors de cette première phase de recherches, j'ai pu prendre connaissance du travail réalisé lors du projet de M. David Boas. Il s'est avéré que ce dernier avait logiquement créé des scripts R afin d'automatiser

certaines tâches redondantes. C'est pourquoi j'ai cru bon de m'initier à l'écriture de scripts R. Cette étape est en effet indispensable pour comprendre le travail réalisé par le passé mais également pour réaliser mes propres scripts quand arrivera la partie "développement" de ce projet. Ce qui va suivre s'attachera donc à présenter succinctement le chargement de fichiers externe dans R ainsi que la syntaxe et les structures de contrôle disponibles dans le langage R. Cette présentation ne se veut pas être un cours exhaustif sur le langage mais simplement un rapide compte rendu des possibilités de ce langage.

Pour cette présentation, j'ai choisi de m'inspirer du cours proposé par Vincent Goulet dans son ouvrage *Introduction à la programmation en R* [4]

### 1.1.1 Chargement de fichiers externes

Charger et exécuter un script externe en langage R est très simple, il suffit en effet d'exécuter la commande suivante :

```
1 source("script.R")
```

Cette commande chargera le fichier "script.R" se situant dans le répertoire de travail courant puis l'exécutera.

### 1.1.2 Syntaxe de base

La syntaxe de R ne diffère pas beaucoup de tout ce que nous avons pu voir par le passé en programmation impérative. Toutefois, quelques différences restent notables. En effet, les affectations utilisent '<-' :

```
1 x <- 2
2 a <- 5 + x
```

De plus, lors de la rédaction d'une affectation, il est possible de forcer l'affichage du calcul (ce qui n'est pas fait par défaut) en entourant l'expression de '(' et ')' :

```
1 x <- 2
2 (a <- 5 + x)
```

Où ici, la deuxième ligne affectera '7' à 'a' tout en affichant '7' sur la sortie standard.

### 1.1.3 Typage des données

En R, tout est objet. Ce que l'on appelait "type" d'un objet par le passé est ici nommé "mode". Il existe plusieurs modes en R :

Mode	Description
numeric	nombre réel
complex	nombre complexe
logical	booléen
character	chaîne de caractères
function	fonction
list	contenu général
expression	expression non-évaluée

Le mode d'un objet n'a pas à être spécifiée par l'utilisateur, il est automatiquement mis à jour lors d'une affectation. Il est toutefois possible de connaître le mode d'un objet en écrivant la ligne suivante :

```
1 mode(variable)
```

De même, en R, tout est représenté sous forme de vecteur (ainsi, un simple nombre est en réalité un vecteur de taille 1). Il est donc également possible de connaître la longueur de ce vecteur grâce à la ligne qui suit :

```
1 length(variable)
```

### 1.1.4 Matrices

Comme nous l'avons présenté, nous allons travailler sur une matrice de données. Il est donc intéressant de voir comment travailler avec ce type de données. Cela se fait tout naturellement en R. En effet, comme tout est représenté sous forme de vecteur, une matrice (et plus généralement un tableau de  $n$  dimensions) n'est au final qu'un vecteur possédant un attribut *dim* > 1. Ainsi, pour déclarer une nouvelle matrice il suffit d'écrire :

```
1 matrix(1:20, nrow=4, ncol=5)
```

### 1.1.5 Fonctions

Il est possible de définir des fonctions. Cela se fait de la manière suivante :

```
1 maFonction <- function(X,Y)
2 {
3   return sqrt((X-Y)^2)
4 }
5
6 ##PLUS LOIN DANS LE CODE##
7
8 d = maFonction(3,2)
```

### 1.1.6 Structures de contrôle

Les structures de contrôle ressemblent (encore une fois) beaucoup à ce que l'on connaît déjà. Les blocs sont délimités par l'utilisation d'accolades ouvrantes et fermantes :

```
1 i <- 0
2 j <- 0
3 while(i < 100)
4 {
5   if(i %% 2)
6   {
7     j <- j + 1
8   }
9   else
10  {
11    j <- j / 2
12  }
13
14  i <- i + 1
15 }
```

On aurait également pu remplacer la boucle "while" par une boucle for :

```

1     i <- 0
2     for(i in 1:100)
3     {
4         ##CODE ICI##
5     }

```

## 2 L'environnement RStudio

R n'est "qu'un" langage. Une fois installé, il dispose de son propre environnement d'exécution en console. Mais la console n'est pas toujours la manière la plus pratique d'utiliser R.

L'utilisation d'un environnement graphique est donc conseillée. J'ai choisi d'utiliser RStudio d'ailleurs que c'est cet environnement qui m'avait été (brièvement) présenté lors du cours de Software Analysis/Data Mining auquel j'avais pu assister lors de mon semestre Erasmus à Nijmegen (Pays-Bas) l'an passé.



Figure 2 – Le logo de l'environnement RStudio

C'est donc au travers de cet outil que nous pourrons exécuter nos commandes R (directement depuis la console ou via des scripts écrits au préalable).

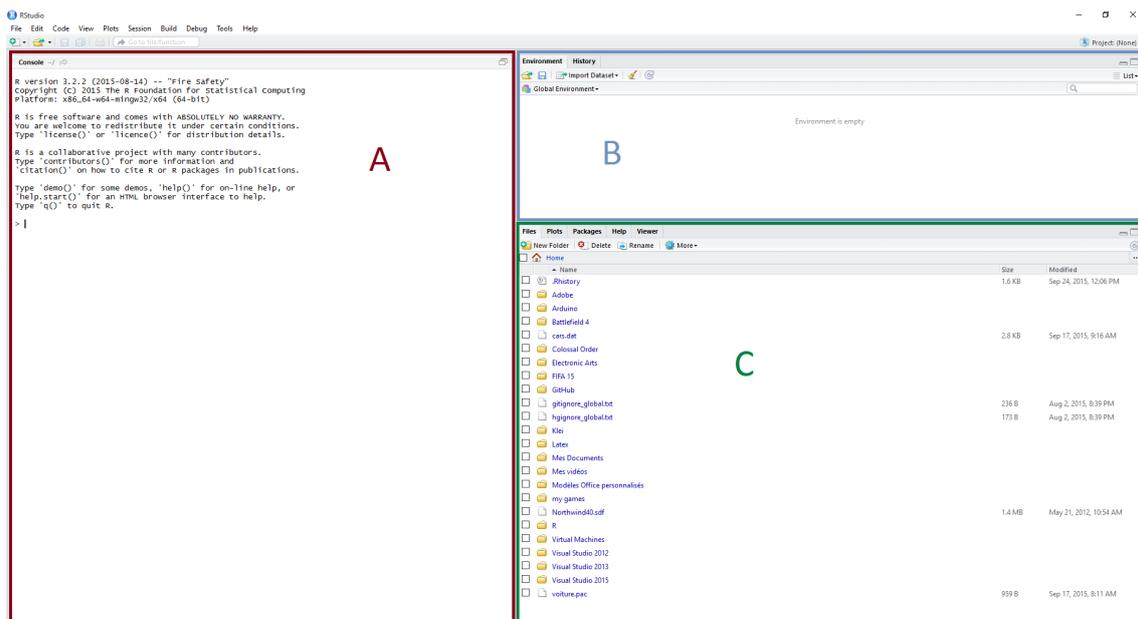


Figure 3 – La fenêtre de RStudio

La Figure 3 montre l'affichage d'une fenêtre type sous RStudio :

- A : La console. C'est ici que les commandes à exécuter sont écrites et les scripts chargés.
- B : Les variables. C'est dans ce panneau que les variables déclarées et créées sont affichées. Elles peuvent être parcourues ou encore supprimées.
- C : Affichage de répertoires. C'est ici que nous pouvons parcourir l'arborescence de notre système de fichier. La sélection du répertoire de travail se fait également ici. Lorsqu'un élément graphique doit être affiché, c'est également dans cette zone qu'il sera dessiné.

### 3 Blender

Même si le travail que j'ai eu à effectuer jusqu'ici ne m'a pas amené à utiliser Blender, il m'a été demandé de m'y former brièvement. J'ai donc suivi un tutoriel sur internet a fin d'acquérir les bases du fonctionnement de ce logiciel [WWW1].



Figure 4 – Logo de Blender

Ayant décidé de ne passer qu'une seule journée à prendre en main ce logiciel, je me suis cantonné à n'effectuer que les deux premiers chapitres de ce cours. Les deux premières parties étant suffisantes pour prendre en main les raccourcis, la création, rotation, transformation et translation d'objets. J'ai également pu expérimenter l'ajout de texture et la mise en place d'éclairages.

J'ai ainsi pu créer une petite scène à partir du modèle de la mascotte de Blender, Suzanne :

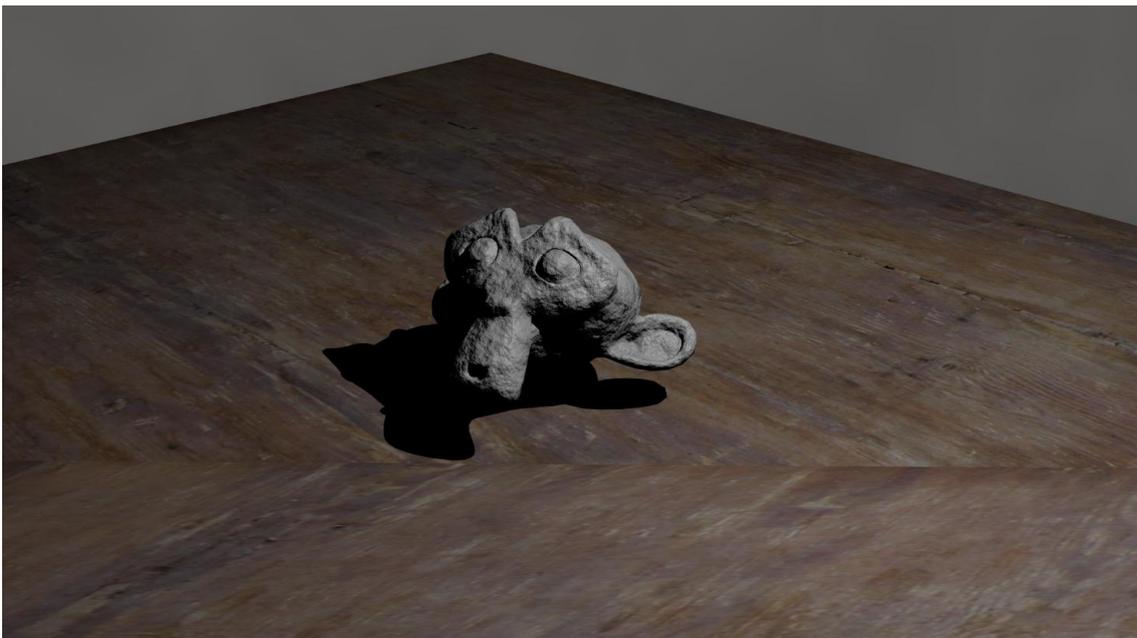


Figure 5 – Scène de test sous Blender

# 4

## Le travail de David Boas

Le but de ce chapitre est de présenter le travail que David Boas a effectué durant son Projet de Fin d'Etudes, l'an passé. Ce dernier a en effet étudié, comme nous venons de le faire, la régression linéaire afin de l'appliquer au problème de modélisation de glène. Il est important de présenter son travail afin de bien cerner dans quel contexte viendront se greffer mes travaux. Toutes les informations que je vais présenter ici ont été obtenues à partir du rapport qu'il a délivré à la fin de son projet [2]. Comme présenté plutôt, le défi réside dans la modélisation d'une glène à son état originel. Ainsi, la marche à suivre était dans un premier temps d'obtenir des données sur un jeu d'omoplates saines puis, par la suite mettre en place une série de régressions linéaires afin de créer des modèles. Ces modèles ayant pour but de prédire la position et la taille de la glène sur l'omoplate.

### 1 Mesures d'omoplates

En dehors de l'étude de la régression linéaire, le premier travail de M. Boas fût l'obtention de mesures à partir d'omoplates saines. Ces omoplates proviennent d'un jeu de 43 individus que j'aurai à ma disposition dans les prochaines semaines. Le tableau en [Figure 1](#) présente, pour chaque individu, le nom de fichier, le côté d'où provient l'omoplate et enfin le sexe de l'individu.

Un ensemble de points a alors été fixé de manière à caractériser avec précisions les omoplates. Ces derniers sont visibles en [Figure 2](#) et [Figure 3](#).

Ces points ont tous un nom associé à leur numéro qui permet de les identifier plus facilement :

- $P_1$ , point supérieur de la glène.
- $P_2$ , point inférieur de la glène.
- $P_3$ , point le plus postérieur de la glène.
- $P_4$ , point le plus antérieur de la glène.
- $P_5$ , point le plus haut de l'écaille.
- $P_6$ , point le plus bas de l'écaille.
- $P_7$ , point le plus latéral de l'écaille.
- $P_8$ , point du milieu de la base de l'épine projeté sur la glène.
- $P_9$ , point le plus postérieur de l'acromion.
- $P_{10}$ , point le plus antérieur de la pointe de l'acromion.
- $P_{11}$ , point de la pointe coracoïde.
- $P_{12}$ , point au fond de l'échancrure coracoïdienne.
- $P_{13}$ , point de réunion des deux faces en bas.
- $P_{14}$ , point postéro-supérieur de pilier projeté sur la glène.
- $P_{15}$ , point antéro-supérieur du pilier projeté sur la glène.
- $P_{16}$ , point inférieur de la base de l'épine projeté sur la glène.
- $P_{17}$ , point supérieur de la base de l'épine projeté sur la glène.

- $P_{18}$ , point arrière de la base de la coracoïde projeté sur la glène.
- $P_{19}$ , point avant de la base de la coracoïde projeté sur la glène.
- $P_{20}$ , point du milieu de la base de la coracoïde projeté sur la glène.

En accord avec M. Berhouet (qui a également réalisé les mesures) des distances ont été définies sur ces omoplates. Ces distances ne présentent pas une liste exhaustive des possibilités offertes par le total de 20 points que nous avons à notre disposition mais plutôt une sélection qui permet de garder les distances les plus pertinentes. Elles sont classées en deux groupes :

- Les distances  $X_i$  : distances dont les deux points ne se trouvent pas sur la glène. Ces distances formeront nos variables endogènes puisque même une fois la glène malade, elles resteront identiques.
- Les distances  $Y_i$  : distances dont au moins un point se situe sur la glène. Ces distances formeront nos variables exogènes. Une fois qu'un modèle aura été défini pour chacune de ces exogènes, la prédiction de la forme et de la position de la glène dans son entièreté pourra s'effectuer. (Ce travail de modélisation de glène est effectué en parallèle de mon projet par Min Jiang [6] tandis que le travail de positionnement est lui étudié, toujours en parallèle par Minghui Zhang) [11].

Ainsi nous noterons les distances sous la forme  $X_{i,j}$  pour un distance de type X s'appliquant sur les points  $P_i$  et  $P_j$  (idem pour  $Y_{i,j}$ ). Voici la liste des 44 distances sélectionnées :

- $Y_{1,2}, Y_{3,4}, Y_{1,7}, Y_{2,7}, Y_{3,7}, Y_{4,7}, Y_{7,8}, Y_{7,14}, Y_{7,15}, Y_{7,16}, Y_{7,17}, Y_{7,18}, Y_{7,19}, Y_{7,20}, Y_{2,13}, Y_{13,14}, Y_{13,15}, Y_{6,14}, Y_{6,15}, Y_{1,12}, Y_{2,12}, Y_{3,12}, Y_{4,12}, Y_{8,12}, Y_{12,14}, Y_{12,15}, Y_{12,16}, Y_{12,17}, Y_{12,18}, Y_{12,19}, Y_{12,20}$ .
- $X_{5,6}, X_{5,7}, X_{5,10}, X_{5,11}, X_{6,7}, X_{6,9}, X_{6,11}, X_{7,9}, X_{7,10}, X_{7,12}, X_{9,10}, X_{10,11}, X_{11,12}$ .

Trois angles ont également été ajoutés afin de caractériser au mieux une omoplate :

- $\hat{a} = \widehat{P_1 P_2 P_6}$  : angle axillo-glénoïdien.
- $\hat{e} = \widehat{P_2 P_6 P_5}$  : angle scapulo-axillaire.
- $\hat{f}g = \widehat{P_5 P_7 P_6}$  : angle spino-suprasinal.

Ces angles ont été matérialisés en [Figure 4](#). On remarque qu'il s'agit en somme des angles proposés par la méthode de Martin, présentée dans l'article sur l'omoplate humaine de Henri V. Vallois [10].

Pour finir sur les mesures d'omoplates, il est important de noter qu'une grande partie du travail de M. Boas fut consacrée à vérifier les mesures effectuées. Ces mesures l'ont amené à produire plusieurs versions de la matrice des mesures, huit au total. C'est sur cette huitième version que je serai amené à travailler dans les prochaines semaines. Cette matrice, de la forme  $M_{43,47}$  contient les 43 individus et les 47 variables. Parmi ces 47 variables, 14 (dont un angle) sont endogènes (distances entre deux points non-situés sur la glène) et 34 (dont deux angles) exogènes (distances dont au moins un des deux points est situé sur la glène).

Individu	Côté	Sexe	Individu	Côté	Sexe
Individu 1	Droit	M	Individu 1	Gauche	M
Individu 2	Droit	M	Individu 2	Gauche	M
Individu 3	Droit	F	Individu 3	Gauche	M
Individu 4	Droit	M	Individu 4	Gauche	M
Individu 5	Droit	M	Individu 5	Gauche	M
Individu 6	Droit	M	Individu 6	Gauche	M
Individu 7	Droit	F	Individu 7	Gauche	F
Individu 8	Droit	M	Individu 8	Gauche	M
Individu 9	Droit	M	Individu 9	Gauche	F
Individu 10	Droit	M	Individu 10	Gauche	M
Individu 11	Droit	M	Individu 11	Gauche	F
Individu 12	Droit	M	Individu 12	Gauche	M
Individu 13	Droit	F	Individu 13	Gauche	M
Individu 14	Droit	M	Individu 14	Gauche	M
Individu 15	Droit	F	Individu 15	Gauche	M
Individu 16	Droit	M	Individu 16	Gauche	M
Individu 17	Droit	M	Individu 17	Gauche	M
Individu 18	Droit	M	Individu 18	Gauche	M
Individu 19	Droit	F	Individu 19	Gauche	M
Individu 20	Droit	M	Individu 20	Gauche	F
			Individu 21	Gauche	M

Figure 1 – Les individus constituant la population

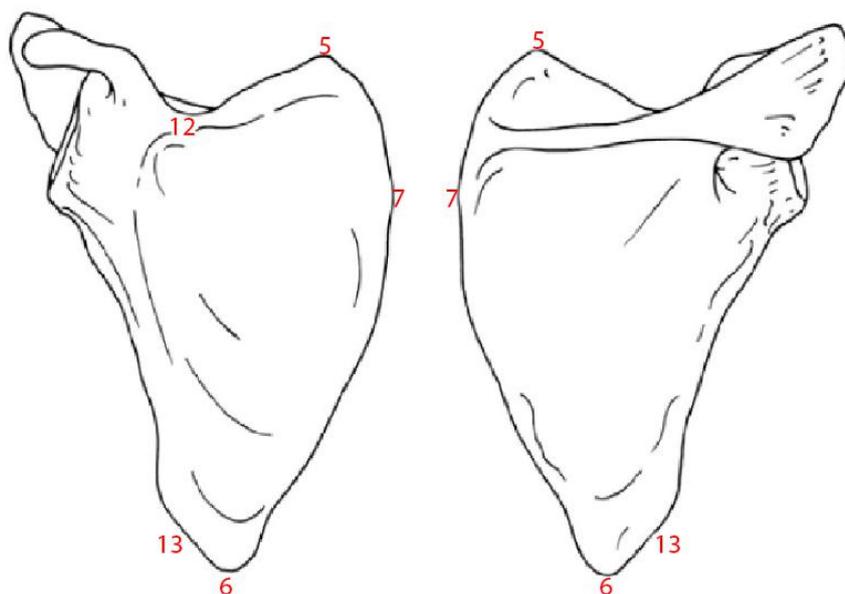


Figure 2 – Points de l'omoplate - Vue latérale

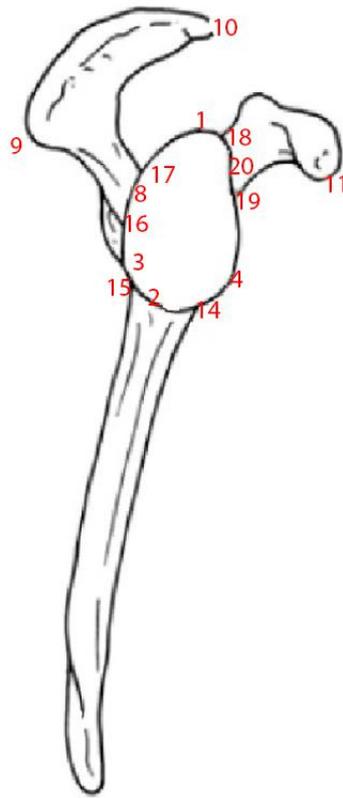


Figure 3 – Points de l'omoplate - Vue longitudinale

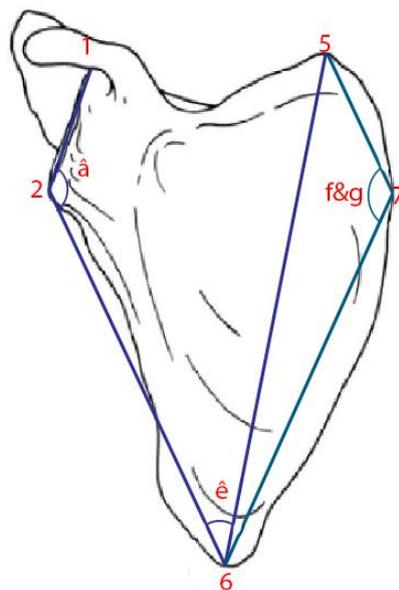


Figure 4 – Les angles définis sur l'omoplate

## 2 La solution mise en œuvre

Par la suite, M. Boas s'est attaché à effectuer une série de tests de régression linéaire sur la matrice de données à sa disposition. Plusieurs choses en sont sorties.

### 2.1 Les difficultés de mise en œuvre

En effet, la première des choses qui apparaît lorsque l'on étudie la partie du rapport présenté est la difficulté que M. Boas a pu rencontrer en utilisant la régression linéaire. Car, travaillant sur une très petite population, les précautions qu'il a dû prendre quant à l'analyse des résultats furent nombreuses. D'où la nécessité d'utiliser de nombreux tests (qui n'ont pas tous été présentés ici pour l'instant).

De plus, ces tests ont permis de déceler des individus incohérents et ont amené à la vérification des données entraînant la création de plusieurs versions de la matrice.

Concernant le travail que j'aurai à faire, cela implique qu'une énième vérification des données ne sera pas superflue. Bien que dispensable à la vue du travail précédemment effectué, nous pouvons légitimement penser que vérifier les données ne représentera pas du temps de perdu.

### 2.2 Les scripts développés

Afin d'automatiser les tâches répétitives de modification de données ainsi que de calculs et tests sur ces données, notamment après un changement de matrice, David Boas a développé une série de scripts qu'il détaille dans son rapport. La présentation que j'en ferai ici restera succincte du fait que je ne possède pas les dits scripts mais seulement la présentation qui en est faite dans le rapport. Néanmoins, cela permet d'obtenir un bon aperçu du travail à réaliser cette année.

- Script "makeDistances" : Créé la matrice des distances.
- Script "matrices" : Permet d'obtenir les matrices de covariance et de corrélation de la matrice de distances.
- Script "ACP" : Réalise une ACP (Analyse en Composantes Principales) des données. (i.e. une dé-corrélation des variables afin d'obtenir une analyse comportant moins de variables).
- Script "plotHist" : Affiche un certain nombre d'histogrammes permettant une comparaison de populations (densité, moyenne).
- Script "MakeModelsWithCriterion" : Ce script comporte une fonction notable qui permet de trouver les modèles valides (la validité des modèles est déterminée grâce à un seuil appliqué sur la significativité des variables explicatives).
- Script "makeAllModels" : Construit l'ensemble des modèles demandés grâce au script précédent.
- Script "testNorm" : Script permettant de tester si l'échantillon suit bien une loi normale.
- Script "testVariance" : Script permettant de tester si les hypothèses sur la variance de l'échantillon sont bien vérifiées.
- Script "testNonParam" : Ce fichier contient les fonctions utiles dans le cas où l'échantillon ne suit pas une loi connue.
- Script "evaluateFormula" : Permet d'automatiser la plupart des tâches décrites plus tôt : traçage d'histogrammes, calcul des modèles etc.

### 2.3 Les modèles produits

A la suite de nombreuses vérifications et de nombreux tests, M. Boas a été en mesure de produire une série de modèles permettant d'expliquer certains points situés sur la glène. Car, tous les points n'ont pas pu être expliqués, cela vient principalement du manque d'équations viables et du fait que dans notre espace à trois dimensions, il faut quatre équations de distances pour localiser un point.

Ci dessous, vous pourrez trouver un extrait des équations générées par David Boas qui est affiché dans son rapport disponible en [Figure 5](#).

Numéro	Equations : Globales	R <sup>2</sup> ajusté	Moyenne écarts
1	$Y_{2,7} = 0.20493X_{5,6} + 0.21845X_{5,11} - 0.14519X_{6,9} + 0.32906X_{7,9} + 44.30713$	0.75	2,207208681
2	$Y_{2,7} = 0.16677X_{5,6} - 0.21552X_{6,9} + 0.409X_{7,9} + 0.20905X_{10,11} + 0.37066X_{11,12} + 43.65133$	0.74	2,277961791
3	$Y_{7,19} = 0.19232X_{7,9} + 0.68508X_{7,12} + 0.2721X_{10,11} - 2.84573Cote : Gauche + 20.46473$	0.70	2,830316559
4	$Y_{7,20} = -0.19232X_{5,6} + 0.38258X_{5,7} + 0.32238X_{5,11} + 0.3768X_{6,11} + 22.47661$	0.80	2,067701798
5	$Y_{7,20} = 0.34367X_{5,6} + 0.35101X_{5,11} - 0.50902X_{6,7} + 0.41118X_{6,11} - 0.39408AngleF.G + 65.88903$	0.79	2,004535864
Equations : Droites			
6	$Y_{6,15} = 0.2884X_{5,6} + 0.6886X_{10,11} + 55.4357$	0.60	4,85480406
7	$Y_{7,15} = 0.5912X_{6,11} + 0.7243X_{7,12} + 0.7593AngleA - 78.8028$	0.79	3,307961385
8	$Y_{7,18} = 0.37517X_{5,7} + 0.36305X_{5,11} + 0.25027X_{7,9} + 22.00914$	0.78	2,332626996
Equations : Gauches			
9	$Y_{6,15} = 0.6325X_{5,10} - 0.5173X_{5,11} + 0.5497X_{6,7} + 44.7412$	0.68	5,100982419
10	$Y_{2,7} = 0.33915X_{5,10} + 6.50045Sexe : M + 72.06428$	0.69	2,621014844
11	$Y_{2,7} = 0.26879X_{6,11} + 0.43752X_{7,12} + 33.89137$	0.71	2,417446834
12	$Y_{4,7} = 0.36682X_{5,10} + 6.93717Sexe : M + 65.95487$	0.70	2,774849672
13	$Y_{7,14} = 0.34765X_{5,10} + 6.08704Sexe : M + 67.52523$	0.64	3,213130291
14	$Y_{7,18} = 0.49845X_{7,10} - 0.30714AngleA + 52.0671$	0.89	2,266570769
15	$Y_{7,19} = 0.4635X_{7,12} + 0.5267X_{11,12} + 5.1708Sexe : M + 39.2793$	0.67	3,007456304
16	$Y_{12,14} = 0.1848X_{5,10} + 3.83626Sexe : M + 26.65591$	0.46	2,534198598
Equations : Hommes			
17	$Y_{6,15} = 0.25173X_{6,7} + 0.39499X_{7,9} + 54.16319$	0.40	4,829307601
18	$Y_{1,7} = 0.3816X_{7,9} + 0.5704X_{9,10} + 34.4586$	0.50	3,193075208
19	$Y_{1,7} = 0.265X_{5,6} - 0.2809X_{6,9} + 0.5162X_{7,9} + 50.4346$	0.42	3,332367091
20	$Y_{1,7} = -0.19132X_{6,7} + 0.48875X_{7,9} - 0.65296AngleA + 104.41555$	0.53	3,549652313
21	$Y_{7,20} = 0.31385X_{7,9} + 0.5103X_{9,10} + 0.44196AngleE + 30.11052$	0.58	2,834658829

Source : Typologie d'une omoplate - Rapport de projet de fin d'études [2]

Figure 5 – Les modèles proposés par M. Boas

## 2.4 Conclusions tirées et travail restant

Même si le travail précédemment effectué est encore incomplet du point de vue du nombre de points expliqués par les modèles proposés, nous pouvons d'ores et déjà en tirer quelques conclusions.

La première d'entre elle est le fait que le travail statistique sur un si petit nombre d'individus n'est pas aisé. Il faut donc redoubler de prudence quant à l'analyse des résultats.

Ensuite, si tous les points n'ont pas été expliqués, on peut tout de même remarquer une différence dans les modèles en fonction du sexe de l'individu. En revanche, la disparité concernant la latéralité de l'omoplate m'interpelle, je pense qu'il pourrait être intéressant d'obtenir de connaître si l'individu est gaucher ou droitier. Comment expliquer une telle différence sinon ? Du fait de ce problème de latéralité mon idée première à la présentation du projet, qui était de "s'aider" de l'omoplate opposée encore saine semble quelque peu compromise (mais peu tout de même être testée).

Pour conclure sur le travail restant, il est bien sûr indispensable d'aborder ce qu'il m'a été demandé de faire à l'origine : effectuer peu ou prou le même travail que M. Boas mais cette fois-ci en considérant que certains points de la glène sont encore sains. De ce fait, nous aurons accès à des modèles expliquant plus de points et de manière, nous l'espérons, plus précise.

## Deuxième partie

# Travail accompli et à venir

Le but de cette partie est de présenter le début de travail sur les données que j'ai pu réaliser dès lors que j'ai pu y avoir accès. Malgré le fait que nous sommes encore dans la phase de recherche de ce Projet de Recherche et Développement, j'ai trouvé intéressant de présenter ce travail effectué dans le sens où il permet une mise en application de tout ce qui vient d'être présenté.

Je profiterai également de cette partie pour présenter le travail qui sera le miens lors du prochain semestre, c'est à dire pendant la partie développement du projet.

# 5

## Vérifications

A partir de la 9<sup>ème</sup> semaine du projet, j'ai pu avoir accès aux données finalement rendues par David Boas. Ces données sont l'ensemble des distances présentées plus-tôt pour chacun des 43 individus.

### 1 Vérification

Mon premier travail fût de vérifier l'absence d'individus aberrants dans la dernière version de la matrice (numéro 7).

Pour ce faire j'ai donc réalisé un rapide classeur Excel [7\_verif.xls] qui permet d'obtenir rapidement les individus qui pourraient être problématiques. Le principe pour cela est assez simple : il suffit de, pour chaque colonne de la matrice (chaque distance mesurée), sortir l'individu maximum et l'individu minimum. Une fois cela-fait, il ne reste qu'à compter le nombre de fois ou l'individu  $x$  à été maximum ou minimum.

Après avoir appliqué ce cheminement sur les données de la matrice V7, nous obtenons le graphique suivant :

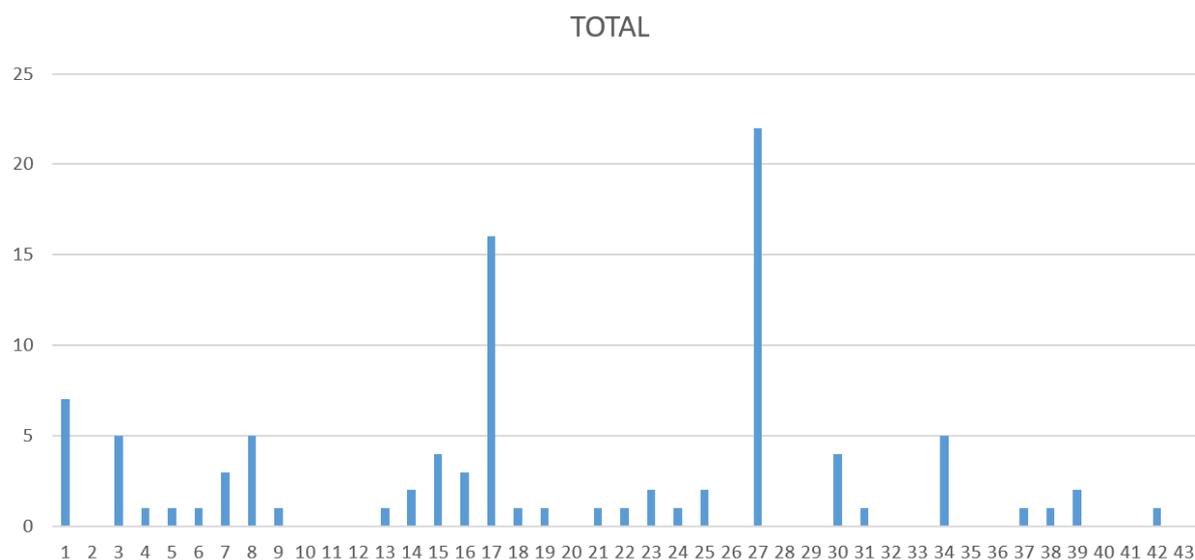


Figure 1 – Résultat de la recherche d'individus aberrants

Ainsi, on voit clairement que les individus 17 et 27 sont régulièrement hors normes. (Pour préciser, il s'avère que l'individu 17 est très souvent maximum tandis que l'individu 27 est très souvent minimum).

Obtenir ces informations s'effectue donc via un processus très simpliste. En revanche, le problème réside dans l'analyse de ce résultat. En effet, comment déterminer si ces deux individus aberrants proviennent d'erreurs de mesures ou bien tout simplement du fait qu'ils soient les plus petits (ou grands) de l'échantillon à notre disposition.

Dans le même temps, j'ai réalisé une matrice de corrélation des variables, disponible également dans le classeur Excel et dont voici un aperçu ici :

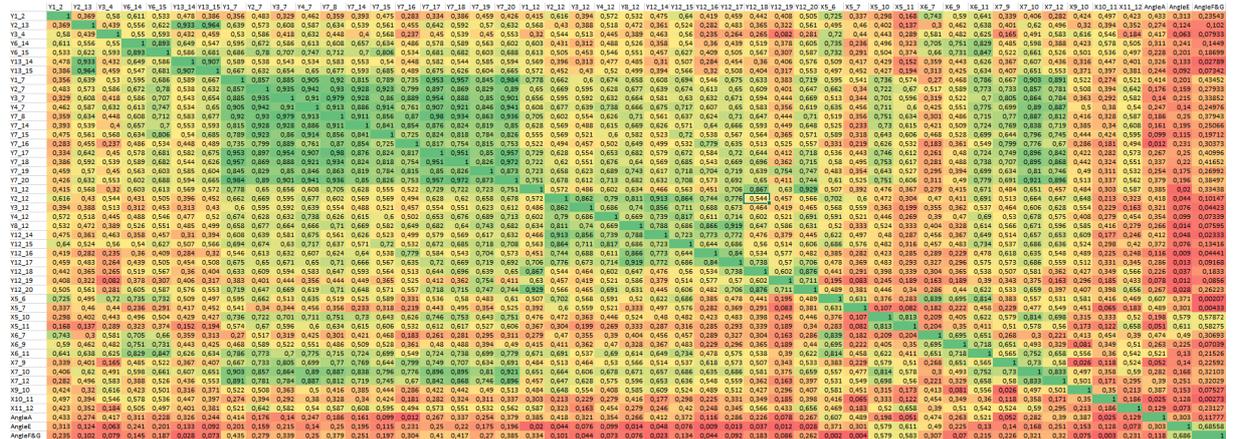


Figure 2 – Aperçu de la matrice de corrélation des variables, disponible également dans le classeur Excel et dont voici un aperçu ici :

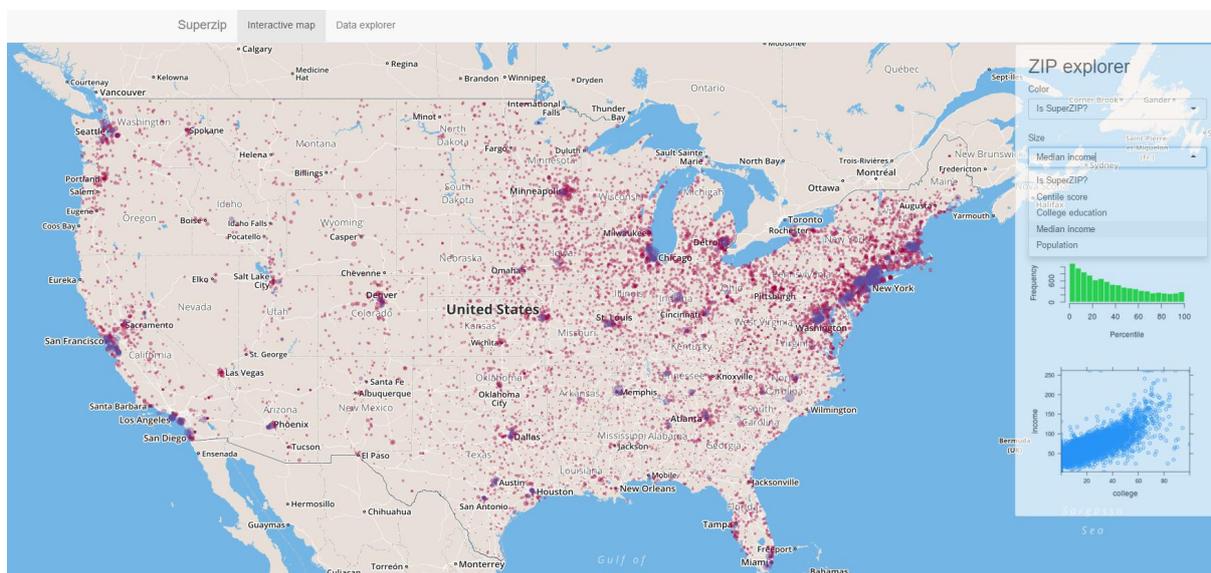
Même s'il ne s'agit ici que d'un aperçu, nous pouvons tout de même voir rapidement les variables fortement corrélées, en nuances de vert, tandis que les variables peu corrélées sont affichées en nuances de rouge.

# 6

## Apprentissage de Shiny

En vue du travail qui devra être réalisé plus tard (présenté au prochain chapitre), j'ai décidé d'apprendre à utiliser Shiny.

Il s'agit d'une technologie proposée par RStudio et qui permet de rendre l'utilisation de scripts R beaucoup plus agréable pour les utilisateurs. En effet, une fois l'application développée, une page web est proposée au client en lieu et place du chargement (manuel) de scripts R dans l'environnement. Cela permet donc également de présenter les résultats de manière plus simple, plus graphique, comme le montre la figure suivante, tirée de la galerie Shiny [WWW10] :



Source : <http://shiny.rstudio.com/gallery/superzip-example.html>

Figure 1 – Aperçu des possibilités de Shiny (ici une carte présentant les revenus moyens)

On voit alors rapidement l'intérêt d'une telle technologie dans l'optique de rendre un livrable utilisable. Mais il ne s'agit là que d'un soucis d'affichage des données car, comme nous allons le voir dans la section suivante, Shiny reste basé sur des scripts R.

## 1 Architecture d'une application Shiny

L'architecture d'une application Shiny est relativement simple dans le sens où nous utilisons la classique architecture Client/Serveur très utilisée dans le secteur du développement Web. Ainsi, le client sera développé grâce à l'utilisation de composants définis par Shiny. De l'autre côté le serveur sera lui principalement développé avec des scripts R. Dans la suite, nous allons observer au travers un exemple très simple (extrait de mes divers tests de développement) comment mettre en œuvre une telle application.

### 1.1 Partie Client

La première chose à faire lors du développement d'une application Shiny est de développer la partie client. C'est à dire l'interface qui sera affichée sur un navigateur Web. Ici, nous allons simplement mettre en œuvre un panneau latéral permettant de charger un fichier ainsi qu'un panneau principal permettant lui, d'afficher les informations de ce fichier :

```

1     library(shiny)
2     shinyUI(pageWithSidebar(
3
4         # Titre de l'application
5         headerPanel("Data analyzer"),
6
7         #Panneau latéral
8         sidebarPanel(
9             fileInput('file_input', 'Choose TSV File', ↵
10                 accept=c('text/tab-separated-values', '.tsv', '.txt'))
11         ),
12
13         mainPanel(
14             tabsetPanel(
15                 tabPanel('Data', tableOutput('data_table')),
16                 tabPanel('Correlation', imageOutput('correlation_table'))
17             )
18         ))

```

L'analyse de ce code peut être faite très rapidement puisqu'on peut observer successivement :

- La définition d'une interface Shiny
- La mise à jour du titre de la page
- La création d'un panneau latéral qui contiendra l'objet fileInput, paramétrable (ici, nous permettons le chargement de TSV (tab separated values).
- La création d'un panneau principal qui lui même contiendra un panneau d'onglets. Ces onglets sont eux-mêmes définis afin de contenir les objets tableOutput et imageOutput. C'est sur ces objets que le serveur travaillera afin d'afficher les résultats sur l'interface utilisateur que nous venons de créer.

### 1.2 Partie Serveur

La partie serveur n'est pas plus difficile à réaliser. En effet, ici, nous nous appuyons sur les variables et les output précédemment définies afin de retourner les informations à l'utilisateur. Ainsi :

- Nous renvoyons simplement le tableau produit par read.csv à notre output datatable.
- Nous renvoyons la corrélation à notre output correlationtable.
- On observe également le fait que les variables de l'interface sont accessibles directement ici (comme par exemple dans le cas du fichier dans inputFile).

```

1     library(shiny)
2     library(datasets)
3     library(corrplot)

```

```
4
5   mpgData <- mtcars
6   mpgData$am <- factor(mpgData$am, labels = c("Automatic", "Manual"))
7
8   shinyServer(function(input, output) {
9     #data_tab
10    output$data_table <- renderTable({
11      inputFile <- input$file_input
12      if(is.null(inputFile))
13        return(NULL)
14      read.csv(inputFile$datapath, header = TRUE, sep = "\t")
15    })
16    #correlation_tab
17    output$correlation_table <- renderPlot({
18      inputFile <- input$file_input
19      if(is.null(inputFile))
20        return(NULL)
21      data <- read.table(inputFile$datapath)
22      plot(cor(data))
23    })
24  })
```

# 7

## Ouverture sur le travail à venir

Le but du projet est bien sûr d'obtenir les équations nécessaires à Min Jiang pour transformer les glènes génériques en glènes parfaitement adaptées au patient. C'est donc sur la mise en œuvre de la régression que je devrai me concentrer. Ainsi, mon travail futur consistera à la mise en place de scripts permettant au minimum d'égaliser ce qu'avait obtenu David Boas. Dans cette optique, la mise en place d'une application avec Shiny peut paraître quasi-optionnelle mais je pense qu'il peut s'agir d'un développement intéressant permettant de produire un livrable.

## Conclusion - Partie Recherche

Nous arrivons donc à la conclusion de ce rapport. Cette première partie de mon projet de Recherche et Développement fût très intéressante dans le sens où elle m'a permis de découvrir et de me former à de nouveaux outils, tant mathématiques que technologiques.

Je dois avouer que la régression linéaire mise en oeuvre ici n'est pas forcément quelque chose qui m'attire quand à ma future carrière professionnelle, tout comme l'utilisation du langage R et de son IDE RStudio. Toutefois, l'utilisation qui m'a été donnée d'en faire n'en est pas désagréable pour autant. De plus, c'est surtout la finalité de ce projet qui me semble intéressante.

La présentation de tous les outils qui à été faite ici, couplée à la lecture du rapport de David Boas me donne une vision assez lointaine de ce que sera mon travail du semestre prochain. En effet, contrairement à certains projets de Recherche et Développement, il s'agira pour moi de continuer ce que M. Boas avait pu faire l'an passé et tenter de l'améliorer. Cette amélioration passera notamment par une réflexion sur d'éventuels ajouts de variables extérieures. Ajoutons à cela l'ajout de possibles points afin de, notamment, ajouter de nouvelles distances  $Y$  dans la liste des variables endogènes.

Quoiqu'il en soit, le développement de scripts R et la recherche des meilleures équations possibles pour chacun des points spécifiés sur la glène s'annonce comme une entreprise nécessitant le nombre d'heures que nous consacrons à ce projet chaque semaine. De plus, comme présenté à la fin de ce rapport, j'aspire à développer une interface en utilisant Shiny de manière à présenter les résultats de manière plus agréable pour l'utilisateur.

## Troisième partie

# Partie développement

Cette partie s'intéressera aux développements exigés par ce projet. Elle peut être considérée comme une mise en pratique des techniques et outils présentés plutôt. Durant la période consacrée au développement, plusieurs choses m'ont été demandé et je m'attacherai ici à les présenter autant que possible de manière chronologique.

# 8

## Méthodologie de suivi et de gestion du projet

Avant de commencer le développement à proprement parlé, nous avons défini une marche à suivre afin de respecter les objectifs initiaux du projet. Il s'agissait en effet dans un premier temps de produire des scripts R permettant d'avoir peu ou prou les mêmes résultats que David précédemment. L'idée était de pouvoir, par la suite, améliorer ces équations pour obtenir des versions définitives de ces dernières pour un maximum d'équations. Et enfin, d'ajouter les considérations d'omoplate partiellement abîmée introduites plus tôt dans ce rapport. Tout ceci est résumé dans le diagramme de Gant présent ci-dessous :

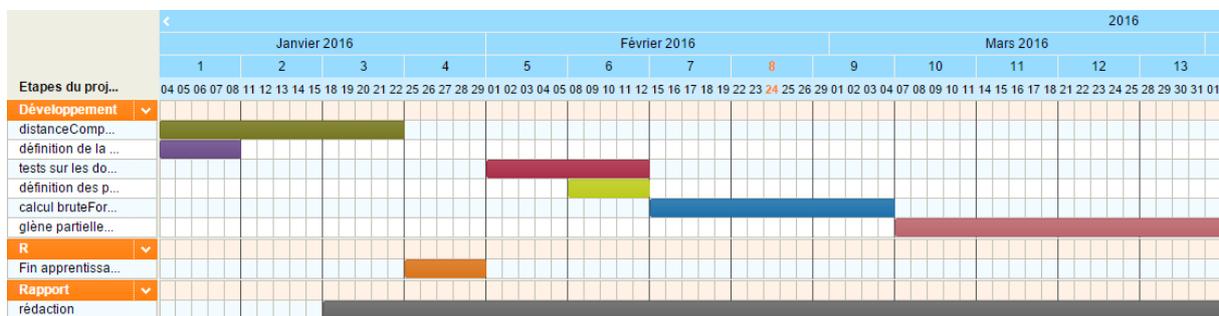


Figure 1 – Diagramme de Gant du projet

Ce projet à une forme particulière, s'éloignant des classiques cycles de réflexion/développement que nous avons pu voir jusqu'ici. En effet, dans ce projet, le développement en soi n'était pas forcément compliqué et la quantité de script à écrire n'était rien en comparaison à d'autres projets que j'avais pu faire. Il s'agissait en effet ici plus de réflexion sur la manière de conduire le projet jusqu'au bout, le développement n'étant qu'un outil d'automatisation. Cependant, que ce soit pour le rapport ou le développement, j'ai mis en place sur une machine personnelle (et donc privée) un serveur Git de sorte à garder privées toutes les informations à ma disposition (plutôt que sur GitHub par exemple). J'ai ainsi pu versionner le code et le rapport sur deux dépôts distincts (chose qui me fût très utile du fait d'une panne de disque dur au cours du projet).



**Figure 2** – *Logo de Git*

# 9

## Mise en œuvre : outil de calculs des distances entre les points

Nous avons vu cela précédemment, pour mener à bien la reconstitution de la glène et afin de mettre en œuvre la régression linéaire, une série de points ont été définis sur l'omoplate.

Ainsi, pour chacun des individus dont nous disposons, un fichier CSV à été créé, donnant les coordonnées X, Y et Z de chaque point. A partir de ces points, nous avons besoin de calculer les distances précédemment définies (voir : [Section 1](#) (Chapitre 4)). Le but final étant d'obtenir la matrice des distances pour chacun des individus.

Le calcul manuel de ces données représentant un temps considérablement long, il a été décidé de développer en Java un outil permettant d'automatiser cela. Ce dernier devait être capable de charger chacun des fichiers CSV des individus afin de produire la matrice de distances, également sous forme de CSV. De plus, les points et les distances étant amenés à être modifiés, l'ajout d'un fichier de configuration permettant de spécifier les distances à calculer semblait des plus utiles. L'outil développé fonctionne selon la [Figure 1](#) :

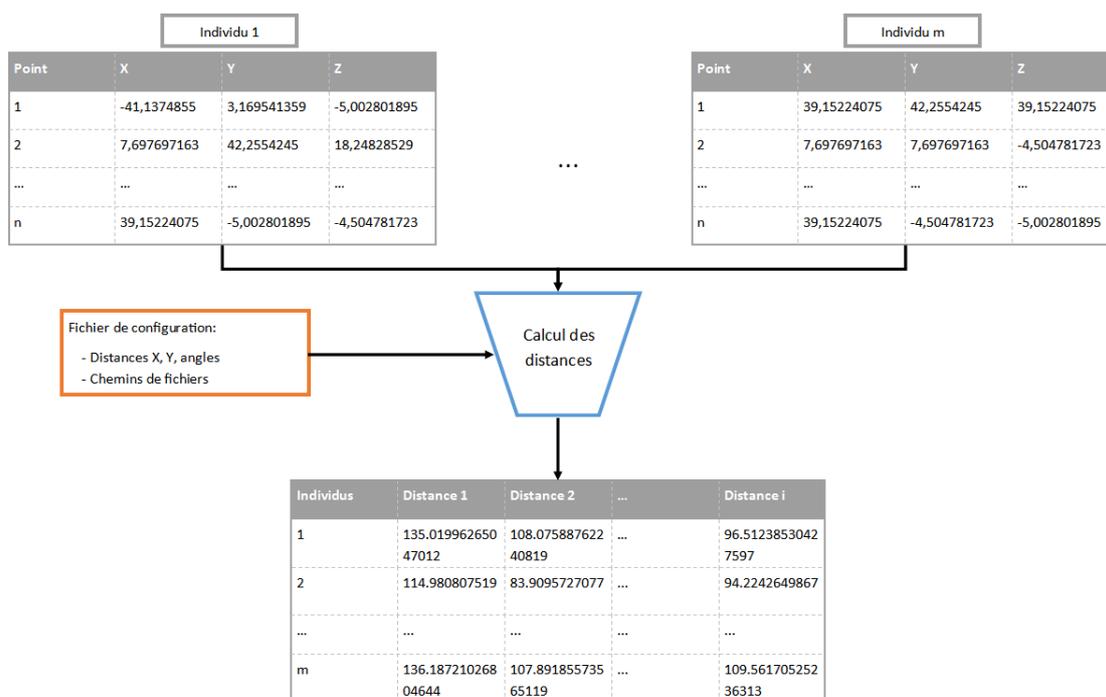


Figure 1 – Schéma de principe de l'outil de calcul de distances

Le fichier de configuration se présente sous la forme suivante :

```

1 FILE :
2 (chemin d'accès au répertoire parent des fichiers individus)
3 (chemin d'accès au répertoire contenant les individus droitiers relatif au chemin ↗
4 d'accès parent)
5 (chemin d'accès au répertoire contenant les individus gauchers relatif au chemin ↖
6 d'accès parent)
7 (caractère séparateur du CSV)
8 Y :
9 <liste des distances Y (sous la forme idPoint1/idPoint2), séparées un saut de ligne>
10 X :
11 <liste des distances X (sous la forme idPoint1/idPoint2), séparées un saut de ligne>
ANGLE :
<liste des angles (sous la forme NOM/idPoint1/idPoint2/idPoint3 pour l'angle ↗
formé par (idPoint1,idPoint2) et (idPoint2,idPoint3) en idPoint2), ↖
séparés un saut de ligne>
    
```

Ainsi, à partir des données en entrée et du fichier de configuration, l'outil calcule les distances euclidiennes et produit la matrice de distances voulue dont voici un extrait ci-dessous :

individu	Y1_2	Y2_13	Y3_4	Y6_14	Y6_15	Y13_14	Y13_15	Y1_7	Y2_7	Y3_7	Y4_7	Y7_8	Y7_14
Droit	39.18327370084	106.5741008050	30.29773474083	133.5587463535	134.1210021354	108.0282573507	108.7332409167	112.4934332308	119.7795393451	111.28132242308	113.7375848663	109.2815907814	115.1245619087
Droit	36.57546109744	90.38828941925	31.03002744044	134.4921729768	133.8862142367	91.18639901718	90.49496286804	111.2751607021	114.9797347764	110.5143895742	114.5966892260	107.8557798627	110.699425807
Droit	31.59129449544	91.44934937094	23.25589276363	119.8207118046	108.0140936959	90.38538166120	79.98281962586	97.69060917987	99.86835505782	96.51238530427	93.37958933426	94.21540832364	97.52638289263
Droit	38.197182807621	107.817677046	30.97575229090	142.3653331740	145.5529765476	108.8181383433	108.0758876224	114.8975764878	120.6750967219	115.5730072875	114.8754771442	114.5230326586	117.1229937984
Droit	33.91178284405	114.9808075190	28.84808071158	132.4429156128	136.6756299072	111.9507209328	116.6767041735	116.7503783818	116.3321452524	111.9819962681	114.3306091169	113.4434182195	117.9486904286
Droit	32.93727625451	100.26494755082	28.84808071158	132.028290416	132.7314510241	106.2875243824	101.2314701776	106.8180050513	111.1512406125	100.9131801709	106.0793939415	99.31142402071	101.2560740996
Droit	36.08815039543	106.6061445712	32.80111280073	129.4965744193	130.3556993352	105.7913041152	105.7900161584	104.6781805340	107.8622113896	102.6102663506	102.9913554713	101.3647856594	102.0483582390
Droit	38.79926165503	108.5769532031	35.12073284503	138.4144367044	143.3739205185	104.5298524940	109.8008662050	112.2447162603	113.1385522527	102.5005016732	112.1346445868	102.8929169092	102.3402641865
Droit	36.70870425170	111.9008228869	29.86433173330	147.3752114956	140.8306066485	116.8865108011	110.3092570432	113.3506728935	114.5576226908	109.5617052523	113.2660823375	108.6341056488	110.0905363538
Droit	39.30203395953	97.27428186615	33.89743274797	131.2672975536	135.0199626504	92.39012839928	95.94289026275	104.4892301398	112.7542333836	106.3426368238	106.5420123177	105.0690949632	99.46210817043
Droit	40.18350730153	103.6934645554	29.93361193882	134.4305476027	134.2367875248	105.7614785205	103.196873518	101.6320906882	109.8885120017	104.6242549207	105.1964413310	101.6900850842	106.3049641619
Droit	40.35137168586	107.1711212637	28.78568619534	134.6435206698	128.0100417969	111.4155153142	104.2967141425	110.9615743811	115.1192168047	107.2586116629	109.8122597976	107.0762609623	110.2103590081
Droit	35.53241742020	99.41109027462	28.97151879828	131.7161700106	136.1872102680	100.4887727916	105.2939204792	117.5769181655	122.5514301108	115.3936527388	116.9438298188	114.4445457444	118.2407802635
Droit	29.36490686527	82.78605059999	23.69487987147	112.3331177075	114.1870121075	78.91185437891	80.91189657679	97.81525320028	102.7120407390	97.16892631908	98.59886916077	94.10603811787	97.78344514018
Droit	30.15386583179	93.17297200481	30.73639171593	123.3004198520	126.2859739559	91.97713417208	94.56332646382	110.0037750118	110.6717285105	105.8279277195	111.2194106183	104.1129542956	105.6936260458
Droit	33.45390834053	88.23177297669	25.71735215609	121.3151568146	122.9677192833	81.79246789815	83.90957270773	99.68700995283	105.4877173327	99.44757029362	102.1049965430	97.93688156168	97.46748015447
Droit	41.25099564539	96.99008403576	29.25149199730	144.9249982826	145.9538824787	95.17311688686	96.76634662803	118.8411420437	124.2901628701	117.8589834010	119.9943464694	116.9506270691	119.9486983940
Droit	41.87121779459	97.1770096464	29.72283801895	130.8582773297	130.5846354958	99.01525245250	98.00045290807	115.5341519400	111.1981721417	106.5567134663	112.1568717119	106.3288171459	106.944048332
Droit	36.63930614093	93.39316714381	25.85193860247	125.9918300761	124.1096889362	94.02452609646	91.88784943551	103.8059231810	107.0978193421	100.2406779687	103.6441820497	98.14525791128	101.3543667653
Droit	35.95506913785	99.97641240677	26.08419688849	132.5283404180	128.5988049714	104.6605522095	101.1077703601	106.8768164997	110.3428006409	107.1070359416	108.1734983458	105.8503450813	108.4291063648
Droit	40.64420688685	98.11332903776	26.67245447792	127.8811829879	129.4266496861	96.85778103237	97.75433976722	102.1908239344	108.7581102119	104.2621922379	109.1917336753	101.0664835445	103.7461309973
Gauche	41.36679719058	92.50143889164	31.30482739200	130.4520675280	129.8359750917	89.0380699031	88.9294076129	106.3246761785	115.0468621019	105.3460414076	113.5889618052	102.3934058215	109.5166905541
Gauche	39.2867773959	106.1481298555	29.57447108020	121.5342757449	121.0497600066	107.7979057616	97.79246906766	101.90736652973	110.7062099193	116.6603545308	102.3273129077	110.0700284877	101.4486946427
Gauche	44.35835062060	90.29808634796	31.03826924247	137.3120900528	133.3117076717	97.72469067567	97.91397652973	100.7062099193	116.6603545308	102.3273129077	110.0700284877	101.4486946427	108.7398732141
Gauche	44.69651046678	110.2308765647	33.95870350592	148.2493951517	142.3062373910	116.3479428501	110.3368476685	113.5866677904	106.0174136889	111.0828988165	110.9703622961	108.6794230242	
Gauche	40.9086170185	99.83665805782	32.26963022300	133.1073852755	122.7650201615	105.4994300700	95.35806337062	106.0422072150	106.8476506031	99.41212938007	109.4122930665	99.78912100749	101.3857246315
Gauche	38.59571369660	107.3372910202	32.13245156861	150.3649961227	149.5228075455	112.8820231924	112.5152088042	106.0065876310	117.4317657350	110.5428414773	115.4153289738	109.4884645715	113.3062400949
Gauche	36.53018248830	89.26967875816	27.90708612241	115.9223786353	112.6490406392	93.44485453908	90.91628789021	90.88585757688	101.0095663742	92.88240718842	98.19821761050	92.83262053904	97.08276106645
Gauche	33.68017458917	80.73253461597	26.79070428901	129.4376188607	123.2901421442	86.05155119063	79.51911144480	98.99340220360	102.99462336742	94.27501288979	99.87528271980	97.88578392998	97.17617964413
Gauche	40.81483028631	95.04181898004	28.40819571215	143.8814089305	137.7973048959	101.8109686585	96.27662495336	107.4653682033	112.2840662883	106.8962899284	107.0503440286	106.4289213741	110.85513572255
Gauche	33.97471966955	72.29238442858	24.95079930178	120.6258310630	111.2940302367	77.21527645362	68.64295294325	95.67452293524	101.8515576341	95.93363014301	97.12848426640	93.66052389256	98.48871704733

Figure 2 – Extrait de la matrice de distances produite, au format CSV et ouvert par Microsoft Excel

# 10

## Mise en œuvre : Tests sur les données réelles

### 1 Structure de données

Le but de la partie développement est avant tout d'améliorer les équations précédemment trouvées et d'en trouver de nouvelles. Cependant, il semble important, avant toute chose, d'essayer de produire les modèles avec les données qui sont à ma disposition (i.e. sans chercher de nouveaux points ou encore de nouvelles distances). C'est pourquoi j'ai réalisé quelques scripts R pour calculer les distances exogènes demandées. Nous sommes donc ici pleinement dans une application des concepts introduits plus tôt et nous nous attacheront ici à expliciter les scripts créés pour ensuite présenter les premiers résultats obtenus.

Il s'agissait dans un premier temps de lire le fichier CSV et de découper la matrice en fonction des parties qui nous intéressent :

- Nom de l'individu, sexe et latéralité
- Distances X
- Distances Y
- Angles

Ce qui nous permet d'obtenir quatre matrices distinctes, respectivement de nom :

- individusMatrix
- endogeneMatrix
- exogeneMatrix
- angleMatrix

Cette découpe de la matrice est réalisée dans le script loadmatrix.R. Par la suite nous utilisons ces matrices de la manière suivante :

```
1      var <- 5
2
3      individusMatrix <- individusMatrix[, 2:3]
4      individusMatrix[,1] <- factor(individusMatrix[,1])
5      individusMatrix[,2] <- factor(individusMatrix[,2])
6
7      data <- cbind(endogeneMatrix, individusMatrix)
8      lmFit <- lm(exogeneMatrix[,var] ~ (.), data = data)
9      model <- stepAIC(lmFit, data, direction = "both")
10     print(summary(model))
```

Où l'on observe la transformation des variables "sexe" et "latéralité" en facteurs ainsi que la concaténation de ces variables avec la matrice des variables endogènes. La variable exogène à modéliser est choisie grâce à la variable var.

In-fine, nous obtenons donc une matrice "data" contenant l'entièreté des variables endogènes, composée des distances X, des angles mais également du sexe et de la latéralité de l'individu.

## 2 Comparaison d'un premier modèle avec celui de David

Il m'a été demandé de comparer une des équations de David avec une de celles que je pouvais construire. A cette fin, nous avons décidé d'utiliser la première qu'il fournit dans son rapport, à savoir :

$$Y1\_5 = 0.205 \times X5\_6 + 0.218 \times X5\_11 - 0.145 \times X6\_9 + 0.329 \times X7\_9 + 44.307$$

Cette équation a été produite grâce à la méthode de régression multiple en mode pas à pas. Cependant, afin de vérifier que la régression produit le même résultat, nous allons ici forcer la régression à utiliser les mêmes endogènes (i.e. Régression linéaire multiple *standard*). Cependant, cela produit un résultat un peu différent de celui de David :

$$Y1\_5 = 0,211 \times X5\_6 + 0,201 \times X5\_11 - 0,104 \times X6\_9 + 0,313 \times X7\_9 + 40,069$$

Comme nous pouvons le voir, les coefficients diffèrent peu, mais diffèrent tout de même. Nous avons donc cherché à savoir la cause de ces différences. A cette fin, nous avons vérifié deux choses :

- Comparaison des matrices produites automatiquement : nos scripts respectifs utilisent leurs propres matrices construites automatiquement. Il était donc raisonnable de comparer ces matrices. Ceci fût fait grâce à l'observation de leurs différences sous Microsoft Excel. Nous avons donc mit en lumière quelques différences principalement dues à des arrondis mais également car ma version de la matrice possède plus de chiffres significatifs.
- Comparaison des équations : nous l'avons vu, les coefficients diffèrent, mais quelles sont les différences d'estimations ? Nous avons donc quantifié cette différence en comparant les estimations des deux équations aux observations sur les cinq premiers individus :

Observations - Estimations					
Id	Observation	Estimation David	Estimation Theo	Y2_7 - David	Y2_7 - Theo
1	119,77954	119,57209	118,817911	0,2074493	0,9616283
7	111,15124	110,484143	111,0295844	0,6670976	0,1216562
22	108,75811	109,4267078	109,8522254	0,6685976	1,0941152
31	112,28407	111,240912	111,4661993	1,0431543	0,8178669
44	107,41863	109,1506557	108,9797957	1,732024	1,561164
			Moyenne	0,8636646	0,9112861

Figure 1 – Tableau de comparaison des équations produites

On remarque donc que, même avec une équation différent quelque peu, les résultats sont très semblables, différant seulement de cinq centièmes de millimètres sur la moyenne des observations.

En conclusion de cela, la seule chose qui permet d'expliquer ces variations sont les différences entre les deux matrices, ce qui est vérifié par l'utilisation de la matrice de David avec ma méthode.

# 11

## Mise en œuvre : Production de modèles

### 1 Points sur l'omoplate

Une fois le travail préparatoire effectué (Calcul automatique des distances et mises en place des structures de données destinées à accueillir les variables endogènes et exogènes) il s'agissait de fixer, une fois pour toutes, les points que nous allons utiliser sur l'omoplate :

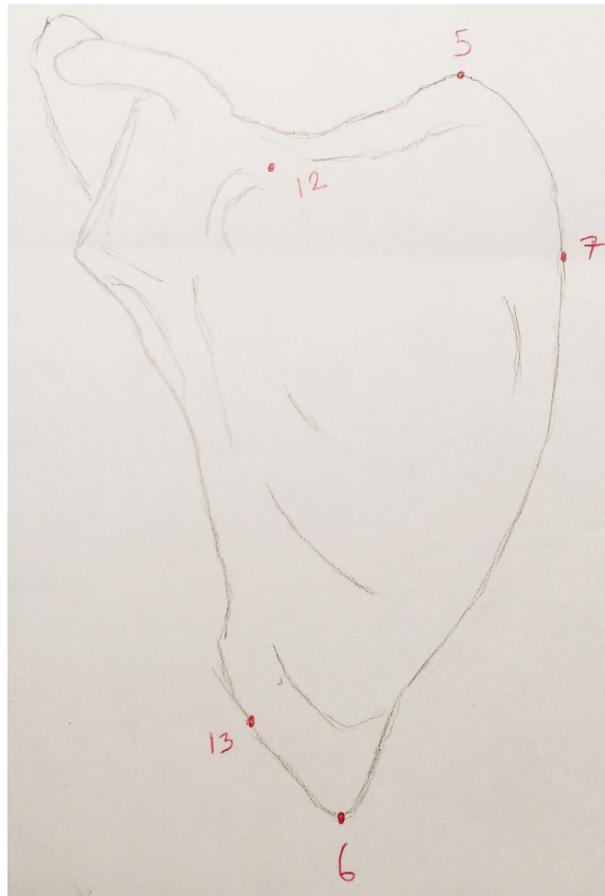
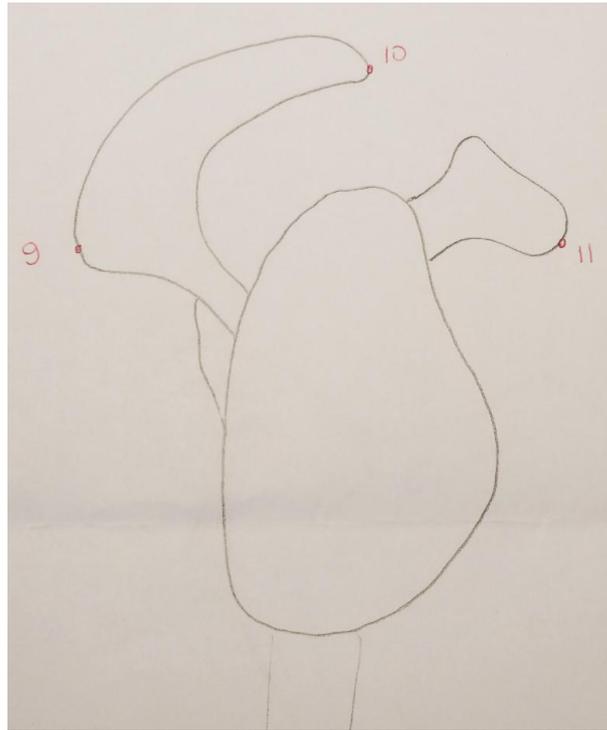
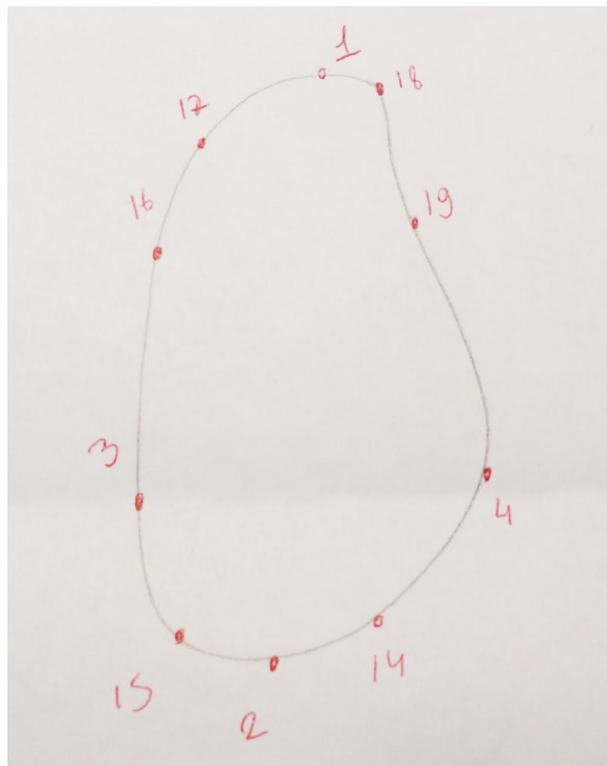


Figure 1 – Croquis de l'omoplate et de ses points (endogènes), vue postérieure



**Figure 2** – Croquis de l'omoplate et de ses points (endogènes), vue latérale



**Figure 3** – Croquis de l'omoplate (glène) et de ses points (exogènes), vue latérale

On peut remarquer que les points sur la glène 8 et 20 ont été supprimés car trop redondants avec d'autres vis à vis de leur positionnement. Mais il a également été décidé d'en ajouter un qui permettrait de rendre compte de la courbure de l'échelle de l'omoplate. Ceci sera explicité plus tard dans ce rapport.

## 2 Rappels sur les modèles

Pour rappel, chaque point situé sur la glène nécessite d'être expliqué par ses coordonnées  $x$ ,  $y$  et  $z$ . A cette fin, et puisque nous nous situons dans un espace à 3 dimensions, 4 distances à partir de points connus de coordonnées connues nous permettent de déterminer les coordonnées de ce point. Ainsi, pour chaque point, un modèle local au point sera composé de quatre équations de distances :

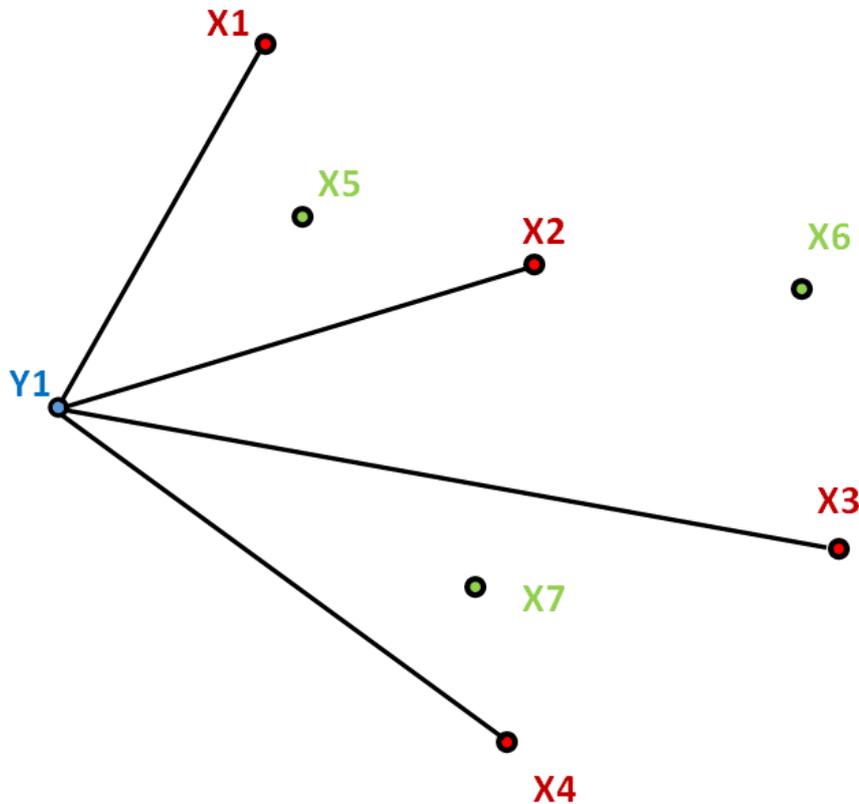


Figure 4 – Rappel sur le principe

On peut ainsi voir ici que nous cherchons à déterminer les coordonnées du point Y1 grâce à l'estimation de distances  $Y1/X1$ ,  $Y1/X2$ ,  $Y1/X3$ ,  $Y1/X4$ . Chacune de ces distances donnant lieu à la production d'une équation et c'est l'ensemble de ces équations qui formera le modèle local au point.

Comment déterminons nous les distances citées précédemment ? En utilisant la régression linéaire de manière à estimer, par exemple, la distance  $Y1/X1$  en fonction de distances connues (exemples :  $X1/X2$ ,  $X1/X3$ ,  $X5/X3$ ,  $X6/X7$  etc.).

A partir de là, deux questions se posent :

- Pour chacune des distances à estimer, quelles distances connues utiliser ?
- Pour chaque point sur la glène, nous devons déterminer quatre distances à utiliser pour les caractériser. Comment choisir ces distances ?

Et c'est précisément à ces deux questions que nous allons tenter de répondre ici.

## 3 Quelles distances endogènes utiliser ?

Dans un premier temps nous allons donc commencer par tenter de déterminer les distances endogènes à utiliser. Pour cela nous allons mettre en place un script permettant de calculer automatiquement toutes les combinaisons de distances endogènes possibles et garder la meilleure.

Notons que nous aurions pu utiliser une régression linéaire multiple, implémentée dans R, mais cette dernière, basée sur le critère AIC ou BIC ne semble pas tenir compte des p-value de chacun des coefficients. En effet, après de nombreux tests, les modèles retournés par ces méthodes contiennent de nombreux coefficients non déterminants. Ajoutons à cela le fait que ces modèles contiennent bien souvent de très nombreuses variables endogènes tandis que sur demande de mes encadrants, je dois autant que faire se peut, limiter les équations à un maximum de quatre variables endogènes utilisées.

L'idée était donc de générer tous les modèles possibles et de les filtrer en fonction de :

- Leur coefficient de détermination R2 *le plus haut possible, mais à minima de 0.75*
- Leur p-value *devant être inférieur à 0.05*
- D'autres tests statistiques qui seront détaillés plus tard.

Ce filtrage permettant de produire un fichier texte, composé des meilleurs équations possibles pour une distance exogène donnée. Un extrait type de fichier texte produit est donné ci-dessous. Il ne présente bien évidemment pas toutes les équations créés mais permet de se rendre compte de la manière dont sont affichées ces dernières.

```
Y1_5 = (0.2923873 X5_6) + (0.2052971 X5_10) + (-0.3578566 AngleA) + (-2.059598 AngleE) + 320.6733
```

```
Call:
```

```
lm(formula = exogene ~ ., data = endogenes)
```

```
Residuals:
```

```
  Min      1Q  Median      3Q      Max
-5.279 -1.265 -0.061  1.164  5.098
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 320.67332   31.88801  10.056 1.64e-12 ***
X5_6         0.29239    0.04378   6.678 5.28e-08 ***
X5_10        0.20530    0.06228   3.296 0.002061 **
AngleA      -0.35786    0.08708  -4.110 0.000191 ***
AngleE      -2.05960    0.21085  -9.768 3.78e-12 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.288 on 40 degrees of freedom
Multiple R-squared:  0.9482, Adjusted R-squared:  0.9431
F-statistic: 183.2 on 4 and 40 DF, p-value: < 2.2e-16
```

```
=====  
=====  
Y1_5 = (0.2447583 X5_6) + (0.2399251 X6_11) + (-0.3399449 AngleA) + (-2.472675 AngleE) + 366.4971
```

```
Call:
```

```
lm(formula = exogene ~ ., data = endogenes)
```

```
Residuals:
```

```
  Min      1Q  Median      3Q      Max
-3.7090 -1.6147  0.0583  1.1158  5.1280
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 366.49711   18.68064  19.619 < 2e-16 ***
X5_6         0.24476    0.04526   5.408 3.2e-06 ***
X6_11        0.23993    0.05585   4.296 0.000108 ***
AngleA      -0.33994    0.08155  -4.168 0.000160 ***
AngleE      -2.47268    0.11185  -22.108 < 2e-16 ***

```

Figure 5 – Exemple de fichier produit

Ainsi, toutes les équations de l'endogène (ici Y 1,5) présentées ici respectent les paramètres de filtrage spécifiés en paramètres de la fonction *bruteForce*. En effet, la fonction *bruteForce* prend en paramètres les données (*data*), la colonne de l'exogène (*exogeneToUse*, la taille de la combinatoire (*combSize* >= 2), le coefficient de détermination minimum requis (*rsquared*), la p-value maximum requise (*pvalue*) :

```
1  bruteForce <- function(data, exogeneToUse, combSize, rsquared, pvalue){
2    ptm <- proc.time()
3    cat("BRUTE FORCE STARTED (combSize=", combSize, ")\n")
4
5    upperCombBound <- dim(data)[2]
6    modelList <- list()
7    for(i in 2:combSize){
```

```

8      cat("  started comb=", i, "\n")
9      comb <- combn(1:upperCombBound, i, simplify=TRUE)
10
11     for(combCols in 1:dim(comb)[2]){
12       model <- regressionMultiple(exogeneToUse, data[,comb[,combCols]])
13
14       if(isModelValid(model, rsquared, pvalue) == TRUE){
15         modelList <- append(modelList, list(model))
16       }
17     }
18   }
19
20   cat("BRUTE FORCE ENDED (", (proc.time() - ptm)[3], "s)\n")
21   return(modelList)
22 }

```

Ainsi, ce script génère toutes les équations viables selon la combinatoire fournie. Les tests sont eux effectués dans la fonction *isValidModel* que nous allons décrire plus tard.

#### 4 Quelles distances utiliser pour chaque point exogène ?

Nous venons de le voir, nous savons maintenant, pour une distance trouver les équations les plus pertinentes. En revanche, nous avons besoin de quatre de ces distances pour chaque point. Toujours dans une optique d'automatisation, nous avons développé des scripts permettant de calculer automatiquement, pour chaque point et pour chaque distance associée, les équations répondants aux critères de finis plus tôt. Ainsi, après une exécution du script, nous obtenons l'arborescence de fichiers suivante :

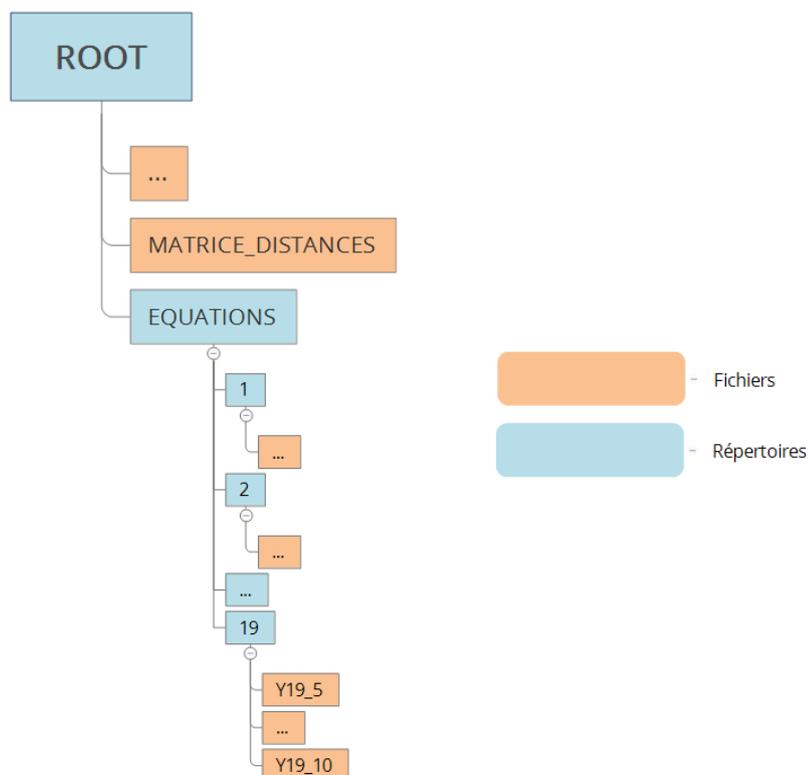


Figure 6 – Rappel sur le principe

On observe ainsi qu'en sortie, chaque point possède un fichier par distance viable. Chacun de fichiers une ou plusieurs équations.

Pour arriver à ce résultat, nous avons dû créer une nouvelle matrice de distance contenant toutes les combinaisons possibles d'un point endogène aux exogènes. Ceci fut réalisé très rapidement grâce à l'outil de calcul précédemment réalisé et à une modification du fichier de configuration associé.

Une fois cette matrice créée, il ne restait plus qu'à créer une boucle autour de la fonction *bruteForce* de manière à itérativement lui donner en argument chacune des exogènes. On obtient alors le script suivant :

```

1  timeStart <- proc.time()
2  cat("<===== STARTED =====>\n")
3  nbDistancesByExogene <- 7
4  pointCount <- 0
5  modelList <- list()
6  unlink("EQUATIONS", recursive=TRUE)
7  dir.create("EQUATIONS")
8  dirName <- ""
9  outputFile <- ""
10 for(exogeneName in names(exogeneMatrix)){
11     print(exogeneName)
12
13     if(pointCount == 0){
14         dirName <- substring(strsplit(exogeneName, "_")[[1]][1], 2)
15         dir.create(strcat(c("EQUATIONS/", dirName)))
16     }
17
18     exogeneToUse <- exogeneMatrix[exogeneName]
19     modelList <- bruteForce(data, exogeneToUse, 4, 0.70, 5e-02)
20
21     outputFile <- exogeneName
22     writeListOfModels(modelList, strcat(c("EQUATIONS/", dirName, "/", outputFile, ←
23         ".txt")))
24
25     if(pointCount == nbDistancesByExogene){
26         pointCount <- 0
27     }
28     else{
29         pointCount <- pointCount + 1
30     }
31 }
32 cat("<===== ENDED =====>  TIME ELAPSED: \n", (proc.time() - timeStart)[3], "s)\n")

```

Dans ce cas, la génération prend évidemment du temps, tout autant que l'écriture des résultats. En tout une vingtaine de minute. Mais il est bien sûr possible de réduire la génération à un point particulier.

## 5 Implémentation des tests statistiques

Lorsque nous effectuons une régression linéaire, plusieurs choses doivent être vérifiées. Bien entendu, dans un premier temps, le coefficient de détermination doit être suffisamment grand et la p-value de chaque coefficient suffisamment proche de 0 (p-value < 0.05).

Mais cela n'est pas tout. Il est en effet nécessaire de vérifier plusieurs autres choses et la première d'entre elles est un test sur la variance des résidus. Ce sont ces tests qui sont implémentés dans la fonction *isModelValid* et qui permettront de mettre en œuvre les hypothèses indispensables à une régression linéaire.

### 5.1 Test de Breusch-Pagan

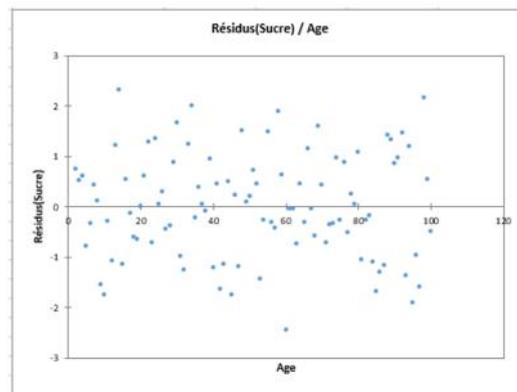
En régression linéaire, nous produisons des estimations de variables grâce à l'équation proposée. Il est possible d'effectuer la différence entre l'estimation et la variable observée et cela s'appelle le "résidu". Le test de Breusch-Pagan permet de vérifier que la variance de ces résidus n'évolue pas en fonction des variables explicatives, cas que l'on appelle homoscédasticité, l'inverse étant l'hétéroscédasticité.

Ainsi, si la p-value issue du test de Breusch-Pagan est supérieure à 0,05, on admet l'homoscédasticité des résidus.

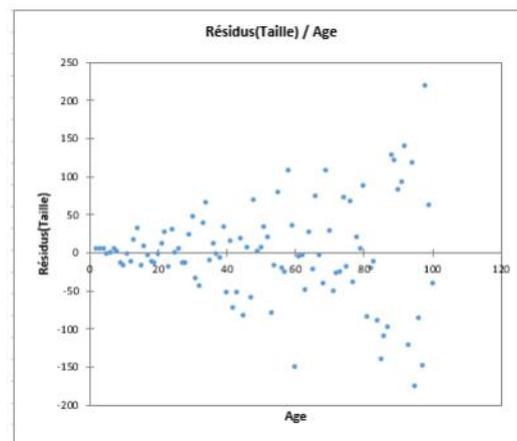
Pour expliquer par l'exemple ce phénomène, nous pouvons présenter les résultats proposés par la page internet disponible ici [WWW11]. Cette dernière présente en effet le cas de deux régressions sur une même population :

- Taux de sucre en fonction de l'âge
- Taille en fonction de l'âge.

Le tracé des résidus en fonction de l'âge donne, pour la première régression la **Figure 7** tandis que la deuxième la **Figure 8**.



**Figure 7** – Tracé des résidus pour la première régression



**Figure 8** – Tracé des résidus pour la deuxième régression

Ainsi, même si les résidus restent centrés, on observe une variabilité relativement constante dans le cas du taux de sucre en fonction de l'âge. Ceci est vérifié par le test Breusch-Pagan qui donne une p-value de 0,332. On en déduit donc que le taux de sucre des individus ne diffère pas plus ou moins selon l'âge. Ce qui n'est pas le cas de la taille où cette fois-ci la p-value est bien inférieure à 0,05.

Dans notre cas il est important d'avoir l'homoscédasticité des résidus. En effet, nous ne voulons pas que ces derniers varient plus (ou moins) selon une variable explicative. C'est pourquoi nous ajoutons à la fonction de test des modèles un test sur la fonction R permettant le test Breusch-Pagan :

```
1 bptest(model)
```

## 5.2 Test de Breusch-Godfrey

Nous allons ici tester l'autocorrélation des résidus. Chose qui ne peut être faite seulement si il y a l'homoscédasticité des résidus. Dans les faits il y a autocorrélation si on peut trouver un coefficient non

nul permettant d'expliquer une relation linéaire entre les résidus. Cette autocorrelation, dans notre cas, pourrait provenir d'un élément non pris en compte par le modèle qui amènerait à une modélisation faussée.

Encore une fois, pour valider la non-autocorrelation des résidus, la p-value retournée par le test de Breusch-Godfrey doit être inférieure à 0,05.

# 12

## Analyse des résultats obtenus

Une fois tous les outils mis en place, il s'agissait d'analyser les équations obtenues de manière à produire, pour chaque point, 4 équations de distances à partir de points connus. Les outils mis en place ont permis de grandement accélérer le travail puisqu'en filtrant au maximum selon la qualité des équations, le nombre de ces dernières s'en trouvait très réduit.

Le fichier de configuration de la matrice de distances était fait de manière à obtenir toutes les distances nécessaires, comme présenté en [Figure 1](#).

```
FILE:
./Final/ 15/5      X:
D/        15/6      5/6
G/        15/7      5/7
;         15/9      5/10
Y:        15/10     5/11
1/5       15/11     6/7
1/6       15/12     6/9
1/7       15/13     6/11
1/9       16/5      7/9
1/10      16/6      7/10
1/11      16/7      7/12
1/12      16/9      9/10
1/13      16/10     10/11
2/5       16/11     11/12
2/6       16/12     ANGLE:
2/7       16/13     A/1/2/6
2/9       17/5      E/2/6/5
2/10      17/6      FG/5/7/6
2/11      17/7
2/12      17/9
2/13      17/10
3/5       17/11
3/6       17/12
3/7       17/13
3/9       18/5
3/10      18/6
3/11      18/7
3/12      18/9
3/13      18/10
4/5       18/11
4/6       18/12
4/7       18/13
4/9       19/5
4/10      19/6
4/11      19/7
4/12      19/9
4/13      19/10
14/5      19/11
14/6      19/12
14/7      19/13
14/9
14/10
14/11
14/12
14/13
```

Figure 1 – Le fichier de configuration pour l'analyse

Ci-dessous, nous pourrions trouver les tables recensant pour chaque point les quatre meilleures équations de distance.

## 1 Glène totalement atteinte

### 1.1 Equations du point X1

Equation	Rsquared
$Y1.5 = (0.244 X5.6) + (0.239 X6.11) + (-0.339 \text{ AngleA}) + (-2.472 \text{ AngleE}) + 366.497$	0.95
$Y1.6 = (0.201 X6.9) + (0.674 X6.11) + (0.382 X10.11) + (0.525 \text{ AngleE}) + -76.875$	0.94
$Y1.7 = (-0.479 X5.10) + (0.383 X6.11) + (0.715 X7.10) + (-1.099 \text{ AngleE}) + 151.761$	0.86
$Y1.10 = (0.447 X5.10) + (0.410 \text{ AngleA}) + (0.657 \text{ AngleE}) + (0.245 \text{ AngleFG}) + -141.296$	0.75

### 1.2 Equations du point X2

Equation	Rsquared
$Y2.5 = (0.500 X5.7) + (0.589 X6.7) + (-2.890 \text{ AngleE}) + 408.586$	0.96
$Y2.6 = (0.330 X5.10) + (0.290 X6.9) + (1.809 \text{ AngleE}) + (-0.541 \text{ AngleFG}) + -173.832$	0.94
$Y2.7 = (-0.682 X5.10) + (0.403 X6.7) + (0.918 X7.10) + (-1.901 \text{ AngleE}) + 276.369$	0.91
$Y2.10 = (0.247 X6.9) + (0.485 X9.10) + (-0.455 \text{ AngleE}) + (0.238 \text{ AngleFG}) + 45.285$	0.73

### 1.3 Equations du point X3

Equation	Rsquared
$Y3.5 = (0.373 X5.7) + (0.382 X5.10) + (0.238 X6.9) + (-1.187 \text{ AngleE}) + 160.582$	0.87
$Y3.6 = (-0.296 X5.11) + (0.640 X6.7) + (0.341 X7.9) + (-0.735 \text{ AngleFG}) + 91.232$	0.79
$Y3.7 = (-0.622 X5.10) + (0.261 X6.7) + (0.953 X7.10) + (-1.616 \text{ AngleE}) + 236.267$	0.82
$Y3.11 = (0.201 X5.7) + (0.271 X5.10) + (0.231 X6.9) + (-0.332 X7.9) + 18.987$	0.66

### 1.4 Equations du point X4

Equation	Rsquared
$Y4.5 = (0.300 X5.7) + (0.509 X6.7) + (-2.110 \text{ AngleE}) + (-0.384 \text{ AngleFG}) + 329.723$	0.91
$Y4.6 = (-0.395 X5.11) + (0.595 X6.7) + (0.445 X7.10) + (-0.791 \text{ AngleFG}) + 88.917$	0.79
$Y4.7 = (0.255 X5.10) + (0.961 X7.12) + (0.314 \text{ AngleA}) + (0.742 \text{ AngleE}) + -109.212$	0.88
$Y4.9 = (0.356 X6.9) + (0.315 X9.10) + (-0.700 \text{ AngleE}) + (0.310 \text{ AngleFG}) + 64.842$	0.73

## 1.5 Equations du point X14

Equation	Rsquared
$Y14.5 = (0.399 X5.7) + (0.429 X6.7) + (0.240 X6.11) + (-2.592 \text{ AngleE}) + 348.634$	0.96
$Y14.6 = (0.540 X5.10) + (0.412 X10.11) + (2.350 \text{ AngleE}) + (-0.493 \text{ AngleFG}) + -243.052$	0.71
$Y14.7 = (0.321 X5.6) + (-0.218 X5.7) + (-0.251 X6.9) + (0.580 X7.9) + 39.728$	0.72
$Y14.11 = (0.214 X5.7) + (0.276 X5.10) + (0.199 X6.9) + (-0.290 X7.9) + 20.720$	0.62

## 1.6 Equations du point X15

Equation	Rsquared
$Y15.5 = (0.419 X5.7) + (0.299 X5.11) + (0.430 X6.7) + (-1.987 \text{ AngleE}) + 274.890$	0.92
$Y15.6 = (-0.490 X5.10) + (0.520 X6.7) + (0.741 X7.10) + (-0.718 \text{ AngleFG}) + 58.538$	0.73
$Y15.7 = (-0.827 X5.10) + (0.447 X6.7) + (1.106 X7.10) + (-2.131 \text{ AngleE}) + 291.169$	0.76
$Y15.9 = (0.173 X5.7) + (0.398 X6.9) + (-0.967 \text{ AngleE}) + (0.388 \text{ AngleFG}) + 95.403$	0.65

## 1.7 Equations du point X16

Equation	Rsquared
$Y16.5 = (0.282 X5.7) + (0.426 X5.10) + (0.278 X6.7) + (-1.439 \text{ AngleE}) + 199.377$	0.87
$Y16.6 = (0.375 X5.10) + (0.359 X6.9) + (1.794 \text{ AngleE}) + (-0.427 \text{ AngleFG}) + -181.797$	0.83
$Y16.7 = (0.221 X6.7) + (-0.338 X6.9) + (0.783 X7.9) + (-0.260 \text{ AngleFG}) + 53.564$	0.72
$Y16.11 = (0.181 X5.6) + (0.171 X5.10) + (-0.139 X7.9) + (0.260 X10.11) + 13.033$	0.70

## 1.8 Equations du point X17

Equation	Rsquared
$Y17.5 = (0.290 X6.7) + (0.361 X7.10) + (-0.436 \text{ AngleA}) + (-2.200 \text{ AngleE}) + 329.887$	0.87
$Y17.6 = (-0.369 X5.11) + (0.700 X6.7) + (0.416 X7.10) + (-0.789 \text{ AngleFG}) + 89.183$	0.84
$Y17.7 = (0.306 X5.6) + (0.181 X5.11) + (-0.255 X6.7) + (0.314 X7.9) + 32.674$	0.76
$Y17.10 = (-0.195 X5.6) + (-0.114 X5.11) + (0.254 X6.7) + (0.433 X7.10) + -10.42796$	0.71

## 1.9 Equations du point X18

Equation	Rsquared
$Y18.5 = (0.247 X5.7) + (0.426 X6.7) + (-2.181 \text{ AngleE}) + (-0.458 \text{ AngleFG}) + 346.215$	0.91
$Y18.6 = (0.552 X6.9) + (0.493 X9.10) + (0.430 X10.11) + (-0.474 \text{ AngleA}) + 46.789$	0.89
$Y18.7 = (0.341 X5.6) + (0.197 X5.11) + (-0.272 X6.7) + (0.238 X7.9) + 37.793$	0.77
$Y18.10 = (0.136 X5.10) + (0.476 X7.10) + (-0.446 X7.12) + (0.228 \text{ AngleFG}) + -24.988$	0.71

### 1.10 Equations du point X19

Equation	Rsquared
$Y_{19.5} = (0.386 X_{5.11}) + (0.214 X_{6.9}) + (0.432 X_{9.10}) + (-1.241 \text{ AngleE}) + 165.414$	0.84
$Y_{19.6} = (-0.485 X_{5.11}) + (0.641 X_{6.7}) + (0.514 X_{7.10}) + (-0.827 \text{ AngleFG}) + 92.234$	0.82
$Y_{19.7} = (-0.524 X_{5.10}) + (0.347 X_{6.11}) + (0.713 X_{7.10}) + (-1.107 \text{ AngleE}) + 161.645$	0.69
$Y_{19.10} = (0.458 X_{5.10}) + (0.353 \text{ AngleA}) + (1.149 \text{ AngleE}) + -183.2155$	0.57

### 1.11 Résultats

Ainsi, et de manière générale, on remarque que les points hors glène les plus *utiles* sont les points X5 et X6. C'est en effet ces deux points qui ressortent systématiquement dans les modèles définissant chaque point. Les deux autres équations de chaque modèle étant plus variables. Les équations présentées ici ne présentent pas l'entièreté des résultats statistiques générés. Ces derniers sont disponibles en annexe.

## 2 Glène partiellement atteinte : partie haute dite "saine", premier test

Nous avons dans un premier temps considéré que la glène était totalement atteinte, c'est à dire que tous les points sur la glène sont considérés comme "à déterminer". Cependant dans les faits, cette assumption n'est pas forcément vraie. Cela implique que nous pouvons tenter de générer des modèles beaucoup plus précis en admettant certains points sur glènes comme endogènes. C'est ce que nous allons tenter de faire ici. Cela passe donc par la modification du fichier de configuration de l'outil de calcul automatique de la matrice de distance de manière à mettre en place ce que nous venons de dire. Nous prenons maintenant en compte les points en vert ci-dessous comme points connus :

Nous faisons ici l'hypothèse que la partie haute de la glène est saine. Dans les faits cependant, ce n'est pas toujours cette partie qui le sera. En effet, si l'on observe les différents types d'atteinte de la glène présentés succinctement plus tôt, on remarque par exemple que le type B entraîne une détérioration de la glène sur la partie postérieure. Dans ce cas là, on considèrerait plutôt les points 1, 18, 19, 4 et 14 comme sains. Quoiqu'il en soit, il s'agit ici de montrer l'intérêt de considérer cette hypothèse.

Ces ajouts se traduisent par le fichier de configuration de la matrice de distances disponible en [Figure 3](#).

Nous allons donc répertorier, comme précédemment, les équations obtenues pour les points restant à déterminer.

### 2.1 Equations du point X2

Equation	Rsquared
$Y_{2.5} = (0.575 X_{5.6}) + (0.056 X_{7.10}) + (-2.253 \text{ AngleE}) + 319.215$	0.996
$Y_{2.6} = (0.496 X_{6.11}) + (0.379 X_{10.11}) + (0.962 \text{ AngleE}) + (-0.436 \text{ AngleFG}) + -78.001$	0.85
$Y_{2.7} = (-0.682 X_{5.10}) + (0.403 X_{6.7}) + (0.918 X_{7.10}) + (-1.901 \text{ AngleE}) + 276.369$	0.91
$Y_{2.10} = (0.247 X_{6.9}) + (0.485 X_{9.10}) + (-0.455 \text{ AngleE}) + (0.238 \text{ AngleFG}) + 45.285$	0.73

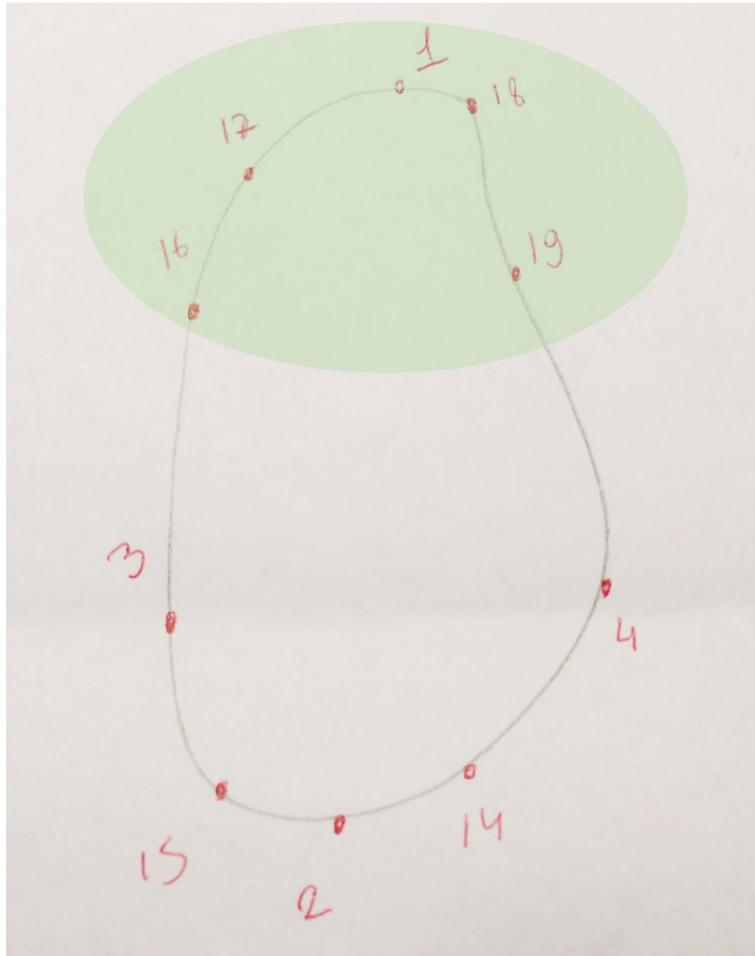


Figure 2 – Capture d'écran de l'application réalisée avec le framework Shiny

## 2.2 Equations du point X3

Equation	Rsquared
$Y_{3.5} = (0.586 X_{5.6}) + (0.842 X_{5.10}) + (-0.377 X_{6.7}) + (-0.4707572 X_{7.10}) + 23.104$	0.90
$Y_{3.6} = (0.860 X_{6.11}) + (0.455 \text{ AngleA}) + (0.655 \text{ AngleE}) + -114.524$	0.80
$Y_{3.7} = (-0.622 X_{5.10}) + (0.261 X_{6.7}) + (0.953 X_{7.10}) + (-1.616 \text{ AngleE}) + 236.267$	0.82
$Y_{3.11} = (0.147 X_{5.6}) + (0.191 X_{5.10}) + (-0.150 X_{7.9}) + (0.547 X_{16.19}) + 16.441$	0.71

## 2.3 Equations du point X4

Equation	Rsquared
$Y_{4.5} = (0.475 X_{5.6}) + (-0.190 X_{5.7}) + (0.418 X_{7.12}) + (-1.875 \text{ AngleE}) + 256.583$	0.94
$Y_{4.6} = (0.610 X_{6.11}) + (0.371 X_{10.11}) + (0.829 \text{ AngleE}) + (-0.352 \text{ AngleFG}) + -74.268$	0.86
$Y_{4.7} = (0.144 X_{7.9}) + (0.183 X_{7.10}) + (0.502 X_{7.12}) + (0.567 X_{16.19}) + 13.775$	0.91
$Y_{4.9} = (0.291 X_{6.9}) + (0.687 X_{16.19}) + (-0.653 \text{ AngleE}) + (0.273 \text{ AngleFG}) + 67.427$	0.73

```

FILE:
./Final/      14/10
D/           14/11
G/           14/12
;            14/13
Y:           14/1
2/5         15/5
2/6         15/6
2/7         15/7
2/9         15/9
2/10        15/10
2/11        15/11
2/12        15/12
2/13        15/13
2/1         15/1
3/5         X:
3/6         5/6
3/7         5/7
3/9         5/10
3/10        5/11
3/11        6/7
3/12        6/9
3/13        6/11
3/1         7/9
4/5         7/10
4/6         7/12
4/7         9/10
4/9         10/11
4/10        11/12
4/11        16/19
4/12        ANGLE:
4/13        A/1/2/6
4/1         E/2/6/5
14/5        FG/5/7/6
14/6        T/16/1/19
14/7
14/9

```

Figure 3 – Le fichier de configuration de la première analyse avec glène partiellement saine

## 2.4 Equations du point X14

Equation	Rsquared
$Y_{14.1} = (0.158 X_{6.9}) + (0.276 X_{9.10}) + (2.225 \text{ sexeM}) + (0.072 \text{ AngleT}) + -11.906$	0.69
$Y_{14.5} = (0.399 X_{5.7}) + (0.429 X_{6.7}) + (0.240 X_{6.11}) + (-2.592 \text{ AngleE}) + 348.634$	0.96
$Y_{14.6} = (0.540 X_{5.10}) + (0.412 X_{10.11}) + (2.350 \text{ AngleE}) + (-0.493 \text{ AngleFG}) + -243.052$	0.71
$Y_{14.7} = (-0.201 X_{6.9}) + (0.312 X_{6.11}) + (0.461 X_{7.9}) + (0.228 X_{10.11}) + 22.951$	0.78

## 2.5 Equations du point X15

Equation	Rsquared
$Y_{15.5} = (0.378 X_{5.7}) + (0.566 X_{6.7}) + (-2.150 \text{ AngleE}) + (-0.419 \text{ AngleFG}) + 334.263$	0.93
$Y_{15.6} = (-0.490 X_{5.10}) + (0.520 X_{6.7}) + (0.741 X_{7.10}) + (-0.718 \text{ AngleFG}) + 58.538$	0.73
$Y_{15.7} = (-0.827 X_{5.10}) + (0.447 X_{6.7}) + (1.106 X_{7.10}) + (-2.131 \text{ AngleE}) + 291.169$	0.76
$Y_{15.9} = (0.173 X_{5.7}) + (0.398 X_{6.9}) + (-0.967 \text{ AngleE}) + (0.388 \text{ AngleFG}) + 95.403$	0.65

## 2.6 Résultats

Si l'on observe les résultats, on se rend compte que les points ajoutés n'apparaissent que très peu. C'est assez étonnant puisque situés au plus près des points à déterminer, l'intuition voudrait que nous trouvions des résultats sensiblement meilleures.

Il s'agit cependant du premier test effectué. En effet, la définition des nouvelles distances à utiliser (définies plus haut) était un premier jet améliorable. C'est précisément ce que nous allons faire par la suite.

## 3 Glène partiellement atteinte : partie haute dite "saine", deuxième test

En effet, la définition des distances n'était pas des plus optimales. Car, si l'on se penche sur les résultats présentés au tout début, on peut se rendre compte que les points 5, 6 et 7 reviennent très régulièrement et permettent d'obtenir de très bons résultats. C'est pourquoi cette fois-ci nous définirons pour chaque nouveau point connu sur la glène Z (c'est à dire 1, 16, 17, 18, 19), les distances YZ.5, YZ.6 et YZ.7. D'où le nouveau fichier de configuration disponible en [Figure 4](#)

Comme précédemment, nous allons maintenant présenter les équations obtenues. Nous remarquerons que ces dernières sont beaucoup plus intéressantes que précédemment :

### 3.1 Equations du point X2

Equation	Rsquared
$Y2.5 = (0.578 X5.6) + (0.0738 X17.7) + (-2.254 \text{ AngleE}) + 318.767$	0.996
$Y2.6 = (0.549 X17.6) + (0.377 X19.6) + (0.407 \text{ AngleA}) + -27.184$	0.95
$Y2.7 = (0.219 X6.7) + (0.800 X1.7) + (0.493 \text{ AngleA}) + -24.922$	0.94
$Y2.9 = (0.317 X5.6) + (0.368 X6.9) + (-0.567 X17.5) + (-1.772 \text{ AngleE}) + 236.052$	0.72

### 3.2 Equations du point X3

Equation	Rsquared
$Y3.5 = (0.334 X5.7) + (0.288 X6.7) + (0.536 X1.5) + (-1.248 \text{ AngleE}) + 171.124$	0.95
$Y3.6 = (-0.297 X5.6) + (0.502 X16.5) + (0.884 X17.6) + (1.265 \text{ AngleE}) + -168.419$	0.93
$Y3.7 = (0.141 X6.7) + (1.048 X17.7) + (-0.314 X18.5) + (-0.723 \text{ AngleE}) + 105.440$	0.94
$Y3.9 = (-0.129 X5.11) + (0.619 X6.9) + (-0.426 X16.6) + (-0.386 \text{ AngleE}) + 57.224$	0.78

### 3.3 Equations du point X4

Equation	Rsquared
$Y4.5 = (0.449 X5.6) + (-0.157 X5.7) + (0.377 X18.7) + (-1.875 \text{ AngleE}) + 251.036$	0.95
$Y4.6 = (0.771 X18.5) + (0.572 X19.6) + (0.592 \text{ AngleA}) + (1.834 \text{ AngleE}) + -291.933$	0.91
$Y4.7 = (0.143 X6.7) + (0.585 X7.12) + (0.526 X18.7) + (0.307 \text{ AngleA}) + -23.12627$	0.95
$Y4.9 = (0.291 X6.9) + (0.687 X16.19) + (-0.653 \text{ AngleE}) + (0.273 \text{ AngleFG}) + 67.427$	0.73

```

FILE:      X:
./Final/  5/6
D/        5/7
G/        5/10
;         5/11
Y:        6/7
2/5      6/9
2/6      6/11
2/7      7/9
2/9      7/10
2/10     7/12
2/11     9/10
2/12     10/11
2/13     11/12
3/5      16/19
3/6      1/5
3/7      1/6
3/9      1/7
3/10     16/5
3/11     16/6
3/12     16/7
3/13     17/5
4/5      17/6
4/6      17/7
4/7      18/5
4/9      18/6
4/10     18/7
4/11     19/5
4/12     19/6
4/13     19/7
14/5     ANGLE:
14/6     A/1/2/6
14/7     E/2/6/5
14/9     FG/5/7/6
14/10
14/11
14/12
14/13
15/5
15/6
15/7
15/9
15/10
15/11
15/12
15/13

```

Figure 4 – Le fichier de configuration pour l'analyse avec glène partiellement saine, deuxième version

### 3.4 Equations du point X14

Equation	Rsquared
$Y_{14.5} = (0.442 X_{5.6}) + (0.317 X_{17.5}) + (-1.544 \text{ AngleE}) + 218.613$	0.97
$Y_{14.6} = (-0.334 X_{5.6}) + (0.410 X_{16.5}) + (0.936 X_{17.6}) + (1.095 \text{ AngleE}) + -147.427$	0.96
$Y_{14.7} = (0.269 X_{6.7}) + (1.004 X_{17.7}) + (-0.429 X_{18.5}) + (-1.150 \text{ AngleE}) + 165.262$	0.86
$Y_{14.11} = (0.207 X_{5.7}) + (0.320 X_{5.10}) + (0.896 X_{16.19}) + (-0.381 X_{17.7}) + 30.63603$	0.67

### 3.5 Equations du point X15

Equation	Rsquared
$Y_{15.5} = (0.287 \times X_{5.6}) + (0.813 \times X_{1.5}) + (0.470 \text{ AngleA}) + -36.191$	0.93
$Y_{15.6} = (0.357 \times X_{1.7}) + (0.800 \times X_{18.6}) + (0.585 \text{ AngleA}) + -62.163$	0.91
$Y_{15.7} = (0.252 \times X_{6.7}) + (0.492 \times X_{7.12}) + (0.678 \times X_{18.7}) + (0.612 \text{ AngleA}) + -58.446$	0.86
$Y_{15.9} = (0.173 \times X_{5.7}) + (0.398 \times X_{6.9}) + (-0.967 \text{ AngleE}) + (0.388 \text{ AngleFG}) + 95.403$	0.65

### 3.6 Résultats

Ainsi, on observe une nette amélioration du coefficient de détermination. Cela correspond à l'intuition que nous pouvions avoir puisqu'il s'agit bien là d'obtenir des points connus au plus proche de la glène. A l'inverse du test précédent où les résultats n'étaient pas concluants, les résultats complets obtenus en sortie des scripts seront disponibles en annexe (à l'instar des résultats standards).

Cette intuition est confirmée lorsque l'on remarque que les points ajoutés sont très régulièrement utilisés. Cette démarche de considération de glène partiellement abîmée est donc concluante puisque l'amélioration des équations est notable. Nous n'avions pas le temps de tenter d'autres expériences avec d'autres parties de glène ici. Le but était de montrer que cela avait un intérêt. Notons cependant que le gain obtenu en qualité de régression devrait être mit en opposition avec les éventuelles approximations de jugement du praticien. En effet, pour que ces équations soient utilisables, nous avons besoin de poser la certitude que les-dits points sont bel et bien sains. Si le praticien faisait une erreur d'appréciation en, par exemple, déclarant un point sain alors qu'il ne l'est pas, les modèles produits seraient alors faussés. Comme précédemment, on peut retrouver les détails de chaque équation générée en annexe.

## 4 Test des équations

Nous allons maintenant vérifier la justesse des équations en se basant sur notre base d'individus. En effet, obtenir des équations abstraites comme ici en ne se basant que sur le coefficient de détermination n'est pas suffisant. Il s'agit maintenant de quantifier l'erreur moyenne obtenue, c'est à dire la moyenne des résidus (*Observations - Estimations*). Nous commencerons par ce que nous appellerons désormais le premier modèle, c'est à dire celui considérant la totalité de la glène comme malade.

### 4.1 Tests sur le premier modèle

Afin de réaliser ce qui a été expliqué précédemment, nous allons utiliser un tableur afin d'obtenir les résidus. Il suffisait alors en effet d'entrer les équations obtenues et de générer les estimations afin d'en déduire les résidus. Dans le cas du premier modèle, une matrice des résidus à pu être construite, visible en [Figure 5](#). Notons que nous n'avons pas utilisé toutes les distances composant le modèle pour obtenir ces chiffres. Nous nous sommes en effet limité à une distance par point.

Cette matrice de résidus nous permet d'obtenir une moyenne de résidus. Cette dernière s'élève ici, pour la sélection de distance choisie, à une erreur d'environ 2,4mm. Il s'agit bien d'une moyenne ici puisque certaines combinaisons de distances/individus descendent à une erreur de 4.e-3 mm alors que d'autres s'élèvent à 12,56mm. On peut alors se poser la question de la pertinence d'un tel modèle puisque l'erreur maximale est clairement rédhibitoire. Nous parlons ici d'une erreur de plus d'un centimètre là où nous traitons des distances de l'ordre du décimètre. Toujours est-il que les résultats sont, en général assez satisfaisant.

Y1_5	Y2_6	Y3_7	Y4_9	Y14_5	Y15_6	Y16_7	Y17_10	Y18_5	Y19_6
2,4041824	5,526769	2,430951	0,206794	1,96148	3,505411	2,113653	0,641354	5,060223	1,263772
1,3745818	2,58769	2,47812	0,590819	0,250941	2,491871	0,084135	0,597597	1,312655	1,963848
1,1158047	7,155977	1,514926	0,2051	3,611469	12,56048	6,872429	2,403441	1,980553	1,823204
3,2935177	0,830112	5,698177	4,002166	2,741683	2,825197	1,565457	2,666403	2,483162	0,75015
1,697129	6,989197	2,499841	1,29816	0,342937	6,073065	9,568958	2,762285	2,908484	7,421187
0,7520924	3,201575	1,863384	0,352512	0,322926	1,647884	0,719923	0,747163	3,11653	3,81458
1,6146765	3,186898	5,878013	1,387406	1,113691	4,890271	2,289189	1,138538	2,022019	6,007901
2,6915508	0,104218	3,657017	1,879512	0,511681	2,024462	1,063075	0,554484	0,313723	10,57101
3,2134957	2,467044	2,957027	2,049022	0,49383	3,515778	11,3285	0,766261	2,366333	1,261596
2,4914425	1,548863	1,302377	4,273602	0,01227	1,279769	2,879787	2,992381	0,714295	0,059852
0,9158447	1,738816	1,548658	1,233451	3,325369	0,729225	2,178799	1,685929	1,999621	2,492758
1,9390132	4,517537	4,832712	3,449522	0,980472	5,656845	5,163308	0,105779	2,419433	1,671487
2,2831329	2,494796	3,472065	1,136667	3,517552	4,532429	0,549522	3,365537	2,25652	0,777718
1,0762347	2,10581	0,022142	3,173286	0,576025	1,41961	0,52032	0,686379	3,196037	6,647013
0,0583155	1,474814	0,080666	0,730831	2,094746	5,307965	1,832788	1,551298	2,183556	2,34194
0,2265827	0,547819	0,283426	4,250096	0,594225	4,143762	1,428623	2,382883	0,670395	0,895468
1,4786686	3,56131	0,145314	0,725353	2,589775	2,995057	2,388032	2,763568	2,791329	8,185998
1,1257632	3,411561	2,846611	0,420097	1,391279	1,884232	0,438358	1,212625	2,908081	0,658644
0,4527306	6,007218	0,330956	2,066609	1,035631	6,039366	0,960537	1,617237	2,647586	3,417494
0,7742113	1,000421	0,739252	1,556728	2,16886	0,635059	1,478687	0,352581	0,18311	1,270244
0,1158954	2,240219	0,06563	6,408783	1,48203	7,950864	4,971871	2,656709	3,29704	3,924577
2,8141038	2,516621	0,60868	1,334636	2,241177	2,160699	0,16114	0,684135	0,843644	1,25652
2,9423325	0,864405	2,700671	1,240281	0,813919	4,74633	0,532788	2,670416	1,295416	2,695466
1,4718154	11,34173	0,763687	3,82282	1,859468	11,70731	3,453315	0,415856	0,877028	7,790749
3,7090067	2,197481	4,438164	2,801538	0,854576	3,524251	2,948183	1,228267	2,737805	5,55234
0,4855376	3,876936	0,16859	5,105357	2,417918	0,85727	4,004579	1,382495	2,472797	4,650113
2,6854796	2,082136	1,323261	3,718657	4,279068	0,262346	2,944215	1,49039	1,848943	2,369401
4,3407831	6,73168	5,344636	1,645113	0,031727	11,78086	6,295745	3,659849	3,180197	11,05846
2,3330579	2,554568	0,98972	3,608905	1,854968	4,255628	1,018639	2,140855	2,288641	4,159462
5,1279728	3,274743	3,116176	2,461077	0,676084	1,632988	0,280264	3,949939	0,365813	5,357688
1,053121	3,911711	3,265343	4,501977	0,557243	4,735058	1,179623	2,854785	4,754262	0,124478
1,8681864	0,27823	0,933499	1,459168	0,122386	4,664694	3,891195	1,467481	2,695919	1,153341
0,5616048	0,044946	0,191095	0,252845	0,086573	2,122092	0,858404	1,087991	5,145198	0,666971
0,3531685	1,999822	2,790925	2,28645	1,032587	0,405342	1,780771	1,108652	1,836953	4,351406
1,0488636	0,494984	0,578786	1,200714	4,147911	4,126132	4,293248	1,15093	0,663905	2,59653
0,9479412	2,238663	0,445298	0,027737	0,688887	1,416491	2,000761	0,853005	0,329827	0,021413
0,3959028	0,396749	1,874131	0,974204	1,439258	1,286543	1,526009	3,131413	0,319244	1,422395
0,2522732	2,530247	0,004329	1,226397	0,581091	1,532708	4,037234	0,939891	0,170555	3,396217
2,9972534	0,92222	0,083236	0,528552	0,474783	7,171445	3,842269	1,756945	0,612987	1,004543
1,5494355	3,15964	4,176498	1,69504	0,570716	8,931373	2,497619	3,54792	2,047338	1,067732
1,5100877	4,740194	1,28931	1,266961	2,560545	2,125625	6,27215	1,758185	3,964969	1,231459
0,6972663	0,182106	1,003117	2,564701	2,7298	1,482099	5,297767	1,729745	6,077647	3,488929
1,0220813	6,231323	1,235529	3,255821	0,492273	3,07017	3,799191	0,882472	1,071053	0,848959
0,0589593	4,462233	0,606431	0,091519	0,629607	4,507944	1,285984	1,146875	2,233638	1,653933
2,0850373	0,124645	0,590551	1,440869	2,673818	1,015223	2,424499	1,694986	2,448426	0,841171

Figure 5 – La matrice de résidus du premier modèle (valeurs en mm)

## 4.2 Tests sur le deuxième modèle

Pour le deuxième modèle nous avons appliqué exactement la même procédure que précédemment. A ceci-près que moins de points étaient à expliquer (glène partiellement saine). Pour chaque point nous avons cependant repris les mêmes distances utilisées (parmi les quatre disponibles donc) afin de pouvoir comparer par la suite les deux modèles. Pour ce modèle, nous obtenons la matrice de résidus disponible en Figure 6.

Cette fois-ci, la moyenne des résidus s'élève à 1.66 mm ce qui est quasiment deux fois mieux que précédemment. De plus, si l'erreur minimale est inchangée, la maximale est également divisée par deux puisqu'elle s'élève ici à 6.5 mm ce qui est beaucoup plus acceptable dans notre cas.

## 4.3 Comparaison des deux modèles

Ainsi, et comme nous avons commencé à le laisser entendre lors de la présentation des équations, l'ajout de points connus sur la glène est bel-et-bien utile, à ceci-près qu'il faut choisir les distances pertinentes à

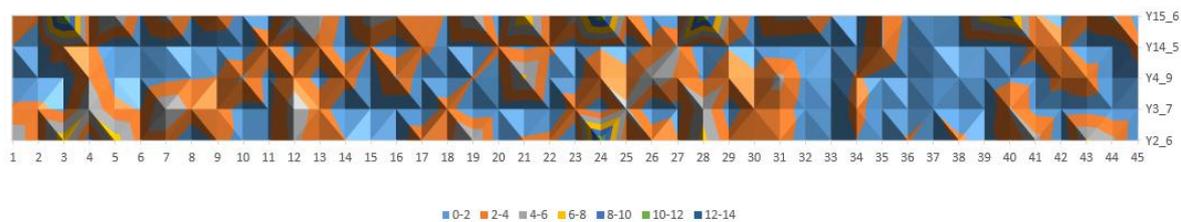
Y2_6	Y3_7	Y4_9	Y14_5	Y15_6
2,378431	1,423528	1,040108	0,141395	1,083191
0,522526	2,089705	1,423868	0,685274	0,754677
0,168761	0,759314	1,918391	2,12471	6,246329
0,361977	1,862773	3,071483	0,875109	1,529967
0,323155	1,087027	0,963905	0,516919	1,234966
0,208549	1,9799	0,130343	1,200126	2,050077
2,5061	1,051323	0,910182	2,287599	0,412356
0,74225	0,038812	3,322773	1,612929	1,367327
1,159816	3,413186	1,110445	1,818667	1,807051
1,317883	0,264475	4,534924	2,049923	0,004206
2,926167	0,141013	0,771952	5,594565	1,565733
0,262844	2,534662	4,727959	1,749781	1,33379
1,311714	2,192051	0,249409	2,435863	4,009399
1,687807	0,420235	1,406741	0,17819	3,702724
0,212028	0,226962	2,118489	0,388242	3,291726
0,487735	1,000564	3,707221	0,720173	0,885478
3,071139	2,0913	1,903716	0,294217	2,931921
2,139893	1,32146	0,108377	0,614473	3,519865
0,977392	0,25539	2,734124	0,141611	1,442158
0,744994	1,349407	2,008686	2,125286	0,015707
1,251036	0,801529	5,066914	1,305531	2,095761
0,842612	1,279169	1,746401	0,025181	1,469262
1,470721	0,227859	0,669969	2,015195	5,365963
2,194679	0,072091	2,59364	1,612222	5,178698
3,454442	1,118376	1,202147	0,174558	0,341885
4,20019	0,685644	3,128844	0,368712	1,191828
1,82047	1,948827	3,285842	2,117867	6,499422
4,615478	0,406752	2,497832	2,515508	4,60806
2,267022	1,894722	0,225972	1,303699	1,546653
2,122399	1,137553	1,441813	3,351436	0,634618
2,842378	0,336892	3,877152	0,608614	1,827955
0,975723	0,682524	2,885568	0,696126	2,512523
3,51691	0,185263	0,977154	1,453044	3,719951
0,750549	1,430995	5,731013	0,862819	1,040011
2,171506	1,00173	1,244081	2,335302	2,472214
0,437794	0,007164	0,902346	0,076029	4,722817
0,933774	1,707236	1,642339	2,094672	2,991956
1,559811	2,631292	0,638125	1,464745	2,52213
0,461642	0,48772	0,247637	0,042773	2,365308
0,293902	1,694035	3,659127	1,490776	4,986526
2,394487	1,312525	0,012541	2,676535	1,766117
0,30281	1,087288	2,797086	1,69888	0,094108
0,109526	0,125199	2,495204	0,285664	0,930249
0,247641	0,093175	0,776731	0,656361	1,684231
0,772254	2,585202	3,493687	1,246718	4,021498

Figure 6 – La matrice de résidus du deuxième modèle (valeurs en mm)

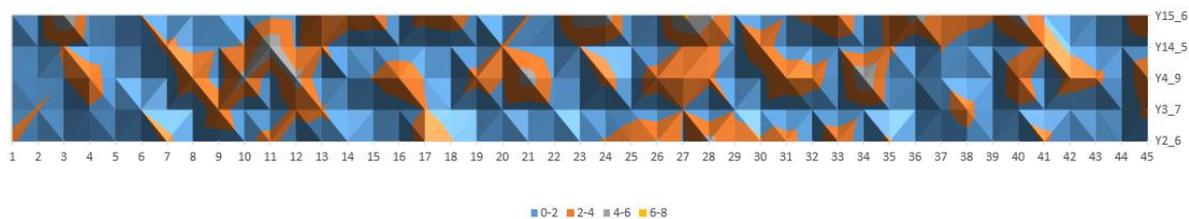
ajouter à la matrice. Les chiffres des résidus parlent d’eux même et pour s’en convaincre encore plus, nous avons tracé ces chiffres sous forme de surface. L’axe des abscisses représentant les individus, celui des ordonnées les distances et l’axe de profondeur (représenté sous forme de couleurs) les résidus.

Les graphiques permettent d’avoir une évaluation visuelle de la qualité des équations obtenues. Les échelles sont dans les deux cas les mêmes. On remarque donc que pour chacune des distances, le modèle 2 est à favoriser puisque présentant moins d’erreurs. Rappelons cependant les réserves ajoutées précédemment sur la validité de ce modèle (erreurs d’appréciations).

Résidus en fonction des individus et des distances - Modèle 1

**Figure 7** – *Le graphique en surface des résidus du modèle 1*

Résidus en fonction des individus et des distances - Modèle 2

**Figure 8** – *Le graphique en surface des résidus du modèle 2*

# 13

## Reproductibilité

Nous l'avons vu, nous avons pu obtenir des résultats intéressants. En revanche, les données à notre dispositions peuvent être amenées à changer dans un futur proche et ce, de plusieurs manières :

- Modification de la population
- Ajout ou suppression de nouveaux points sur l'omoplate
- Ajout ou suppression de nouveaux points sur la glène
- Ajout ou suppression de distances endogènes
- Ajout ou suppression de distances exogènes

Ces changements pourraient conduire à de meilleurs résultats mais surtout amènent avec eux le problème de la reproductibilité du travail présenté jusqu'ici, c'est pourquoi cette partie se concentrera sur la présentation de l'utilisation des différents outils développés afin de produire de nouvelles équations.

### 1 Préparation des données

Dans un premier temps, il faut disposer de fichiers au format CSV contenant les coordonnées X, Y et Z des individus (un fichier par individu). Ces fichiers, tels qu'ils m'étaient fournis avaient le format suivant :

Ainsi, chaque individu possédant son fichier, il faut désormais les classer dans les répertoires *Gaucher* ou *Droitier* de manière à ajouter l'information de latéralité.

### 2 Création de la matrice de distances

Le but ici est de produire la matrice de distances, pour ce faire, il suffit d'utiliser l'outil java développé, *distanceComputer* donc les sources sont disponibles mais également une archive *JAR* qu'il suffit de lancer pour créer la matrice de distances. Il est cependant préalablement nécessaire de configurer le fichier *config.txt* de manière à indiquer au programme :

- Les répertoire de localisation des fichiers individus
- Les distances exogènes à calculer
- Les distances endogènes à calculer
- Les angles à calculer

Cette configuration du fichier est décrite dans la partie [Chapitre 9](#).

Une fois lancé, le fichier créé *output.csv* contient la matrice de distances qui sera utilisable directement par les scripts R développés.

Point\XYZ	X	Y	Z
Point1	-41,1374855	-8,88639069	40,8505707
Point2	-50,3928261	-3,42675352	3,16954136
Point3	-41,798748	8,41907692	19,5835781
Point4	-43,8635521	-21,7501888	17,7114887
Point5	30,7641411	-26,0823402	59,8150864
Point6	25,9083252	-6,37960815	-105,371277
Point7	68,6281433	-5,0028019	16,5360279
Point8	-39,5534859	7,69769716	25,3620415
Point9	-50,627903	35,6369553	19,3181782
Point10	-65,4500885	-0,24096394	64,882431
Point11	-48,8364449	-39,7110252	42,2554245
Point12	-7,04327297	-19,3283806	39,1522408
Point13	1,62727928	-4,50478172	-89,8400574
Point14	-45,6227684	5,28527165	6,81237268
Point15	-47,7976799	-17,6814842	6,11027145
Point16	-41,4900246	8,83558846	18,2482853
Point17	-39,2617416	5,45848942	31,4801712
Point18	-33,6310539	-0,44393349	40,2407608
Point19	-40,0698319	-16,6319103	34,7524643
Point20	-41,3524513	-10,4891424	40,6319199

Figure 1 – Fichier d'un individu

### 3 Utilisation des scripts

Les scripts développés ont pour but d'automatiser la génération des équations pour chacune des distances. Dans ce but, il s'agit simplement de lancer le scripts principal via une fonction R. Il reste cependant la question des paramètres à établir pour obtenir les résultats voulus. Avant de présenter cela, nous allons lister les scripts développés :

- *main* : le script à lancer qui s'occupera d'appeler les scripts suivants
- *loadmatrix* : le script qui permet de charger la matrice de distances spécifiée et qui se charge de la découper de manière à avoir une utilisation plus simple des données par la suite
- *regressionscript* : le script comprenant tout le nécessaire pour le lancement des calculs de régression linéaire multiple avec, notamment, la présence de la fonction permettant de générer toutes les équations disponibles pour une distance donnée.
- *outputsript* : scripts permettant d'afficher ou d'écrire les résultats dans un fichier.

Ainsi, les réglages à effectuer se trouvent dans *main.R*. Ces derniers sont les suivants :

- *DATAPATH* : variable contenant le nom de fichier de la matrice de distances et qui sera utilisé par la suite.
- *Rsquared* : la valeur minimum requise du coefficient de détermination pour que l'équation soit retenue.
- *pval* : la valeur maximum requise de la p-value pour que l'équation soit retenue.
- *maxCoefs* : le nombre maximum de coefficients (c'est à dire du nombre de distances endogènes) à utiliser dans l'équation.
- *nbDistancesByExogene* : le nombre de distances endogène disponible pour chaque distance exogène.
- *points* : la liste des points à calculer. Cela s'avère très pratique lorsque l'on ne veut calculer que certains points et non pas tous ceux disponible, d'autant que le calcul prend du temps.

Ainsi, un simple réglage de ces paramètres suivi d'une exécution du scripts permet de générer les résultats d'équations qui seront présenté sous la forme d'une arborescence, comme vu précédemment.

### 3.1 Application Shiny

Cependant, nous avons annoncé plus tôt vouloir aller un peu plus loin dans le développement sous R de manière à mettre en place une application plus facile d'accès à un utilisateur standard afin que ce dernier puisse modifier ces paramètres sans avoir à modifier le code du script. C'est pourquoi nous avons pris sur notre temps pour mettre en place une application utilisant le framework Shiny (présenté plus tôt). Cette dernière permet la modification de tous les paramètres présentés plus tôt directement sur une page web. Pour le reste du fonctionnement, cela reste basé sur la même chose, c'est à dire que les équations seront toujours stockées sous la forme d'une arborescence.

Ci-dessous, nous pouvons voir une capture d'écran de ladite application, où on y retrouve tous les réglages énoncés plus tôt :

EPAULE

**View data**

Ensure that your data is inside the working directory and named MATRDX.bt

View Data

**Computation parameters**

Maximum size of combinations  
4

Maximum p-value  
0.05

Minimum R-squared  
0.8

Type the points id that need to be computed, separated with a coma: x,y,...z

Points to compute  
1,2,3,4

Number of distances by exogène  
7

Start

Matrice

individu	cote	sexe	Y1_5	Y1_6	Y1_7	Y1_9	Y1_10	Y1_11	Y1_12	Y1_13	Y2_5	Y2_6	Y2_7	Y2_8	Y2_10	Y2_11	Y2_12
1 ArchambaultS_Omo_D.csv	Droit	M	76.32	160.88	112.49	50.36	35.26	31.80	35.70	137.58	101.53	132.71	119.78	42.27	63.60	53.35	56.44
2 AubertA_Omo_D.csv	Droit	M	86.56	160.42	111.28	48.79	29.39	36.79	32.01	119.93	103.59	134.24	114.98	52.01	58.74	55.12	47.81
3 AvelineO_Omo_D.csv	Droit	F	72.06	146.00	97.69	36.73	26.40	32.38	32.50	118.20	86.68	120.95	99.87	50.99	53.45	48.34	44.81
4 BaschouaiD_Omo_D.csv	Droit	M	102.26	174.22	114.90	42.07	30.74	35.38	33.44	138.65	120.27	145.57	120.68	56.31	59.91	53.49	46.81
5 BaudettL_Omo_D.csv	Droit	M	86.30	159.64	116.75	40.99	27.41	32.00	38.12	141.23	102.43	135.62	116.33	45.56	56.47	50.44	55.81
6 BlanchetA_Omo_D_(fracture_pointe).csv	Droit	M	57.61	139.58	93.71	43.25	27.80	30.98	27.88	113.21	79.92	119.86	102.63	42.71	51.36	49.23	41.81
7 BiotC_Omo_D.csv	Droit	F	73.14	165.03	106.82	45.09	33.77	36.72	34.39	135.40	99.68	133.41	111.15	54.42	63.61	53.93	46.81
8 BouchereauV_Omo_D.csv	Droit	M	62.99	156.10	104.68	42.77	23.51	28.93	37.04	133.45	81.71	131.13	107.86	45.12	53.81	42.71	44.81
9 BoulegueB_Omo_D.csv	Droit	M	90.40	174.98	112.24	45.15	29.39	41.09	36.27	142.84	110.88	143.21	113.14	58.21	64.18	59.27	55.81
10 BourreauMC_Omo_D.csv	Droit	M	76.10	171.50	113.35	42.98	28.36	37.08	34.61	141.99	98.54	142.57	114.56	45.11	56.97	52.11	50.81
11 ChapinC_Omo_D.csv	Droit	M	76.00	164.67	104.49	42.27	29.91	35.50	32.91	127.33	99.25	136.69	112.75	53.49	62.61	40.36	52.81
12 ClermontF_Omo_D.csv	Droit	M	74.76	161.85	101.63	42.62	27.37	36.83	33.88	136.23	97.66	133.08	109.89	52.30	61.25	52.92	47.81
13 DeJesusC_Omo_D.csv	Droit	M	90.64	159.89	110.96	45.93	32.97	35.86	36.17	138.63	109.25	131.05	115.12	57.92	61.94	57.57	46.81
14 DeleryG_Omo_D.csv	Droit	F	96.25	157.35	117.58	42.00	29.99	40.88	38.29	127.18	114.24	130.97	122.55	51.92	58.13	51.70	55.81
15 ExcoffonA_Omo_D.csv	Droit	M	79.13	136.58	97.82	32.81	25.19	32.08	30.96	104.51	96.49	116.21	102.71	40.42	50.57	43.60	44.81
16 GuilbanV_Omo_D.csv	Droit	F	84.50	143.81	110.00	41.59	28.14	33.63	33.41	114.17	100.27	125.02	110.67	49.93	51.06	51.61	46.81
17 KozrakV_Omo_D.csv	Droit	M	73.58	148.93	99.69	37.51	28.06	34.81	28.61	110.86	94.38	127.73	105.49	46.61	50.74	42.33	44.81
18 LagneauK_Omo_D.csv	Droit	M	91.90	176.17	118.84	53.18	35.15	38.15	39.71	128.83	113.43	146.69	124.29	56.57	64.81	54.26	55.81
19 MeunierS_Omo_D.csv	Droit	M	87.42	166.41	115.53	44.52	27.53	35.61	37.39	135.66	105.22	129.89	111.20	50.00	62.78	57.74	55.81
20 PandaJ_Omo_D.csv	Droit	F	74.84	152.60	103.81	37.00	25.28	33.74	31.52	122.56	91.96	125.79	107.10	43.19	54.98	53.40	41.81
21 RodotO_Omo_D.csv	Droit	M	77.39	152.39	106.88	37.01	34.52	33.47	33.17	126.94	93.22	127.90	110.34	46.58	59.17	50.78	44.81
22 TouzeauT_Omo_D.csv	Droit	M	61.08	158.46	102.19	47.05	32.59	31.69	31.28	128.21	87.20	129.50	108.76	48.97	60.18	53.32	55.81
23 BarrautF_Omo_G.csv	Gauche	M	73.61	164.80	106.32	42.52	29.45	36.02	35.12	125.93	97.84	134.26	115.05	59.11	63.38	53.32	55.81
24 BaschouaiD_Omo_G.csv	Gauche	M	86.07	150.01	115.99	44.65	33.59	38.41	33.64	136.49	111.56	119.58	118.31	59.34	60.26	55.28	55.81
25 BiotC_Omo_G.csv	Gauche	M	81.73	164.66	100.71	49.06	34.49	42.59	31.16	126.57	112.06	131.85	116.66	58.15	61.96	54.27	55.81
26 BoulegueB_Omo_G.csv	Gauche	M	84.98	178.52	110.34	48.09	28.39	40.08	36.28	148.44	109.75	142.62	113.59	67.96	64.88	56.39	55.81
27 BrennerS_Omo_G.csv	Gauche	M	84.91	158.41	106.04	45.56	24.74	31.03	33.05	133.15	105.17	128.06	106.85	50.43	60.13	57.04	55.81

Figure 2 – Capture d'écran de l'application réalisée avec le framework Shiny

# 14

## Conclusion

Nous en arrivons donc à la conclusion de ce projet. Il est toujours difficile de conclure presque six mois de travaux d'autant que ce projet avait une forme un peu particulière.

En effet, il s'inscrivait comme la base d'un projet plus vaste, destiné à reproduire le plus fidèlement les glènes. La part de mon apport ici se limitait à la production de modèles destinés à retrouver les points disparus d'une glène. Retrouver ces points est donc primordial pour les éléments suivant du projet (reconstitution 3D puis affichage en réalité augmentée) et en ce sens il s'agissait de quelque chose de très intéressant. Cependant, alors que le projet se termine, je me rend compte du côté frustrant que cette division du travail apporte. J'aurai en effet pouvoir observer le résultat de mon présent travail en modélisation 3D notamment et comparer visuellement (en sortant des critères numériques) la qualité des modèles obtenus.

Quoiqu'il en soit, il s'agissait ici de me limiter à ce qui m'était demandé, les autres parties étant déjà en développement par Min et Minghui. Ce petit bémol n'enlève cependant rien à la finalité du projet qui à elle seule aurait suffi à me motiver.

Car malgré tout, nous avons tout de même réussi à produire des outils d'automatisation permettant de calculer rapidement de nouvelles équations si les données (nouveaux points ou nouveaux individus) étaient amenées à changer. En effet, si durant toute la durée de mon projet, j'ai travaillé sur une seule et même matrice, je suis parti du principe que l'ajout de nouveaux individus était quelque chose de plausible dans le futur. Il s'agissait donc de prévoir l'arrivée de ces nouvelles données afin de pouvoir rapidement les prendre en compte. Et c'est chose faite ici. Car si j'ai souvent eu la sensation de produire un travail redite de ce que David Boas avait pu faire l'année passée, de par l'outil de calcul de la matrice de distance mais également par les scripts automatiques de calcul, je suis heureux de pouvoir présenter cette plus-value améliorant le travail de l'an passé.

Ajoutons à cela l'obtention d'équations assez satisfaisantes dans la majorité des cas vérifiées par rapport aux observations ainsi que le travail de recherche quant aux considérations de glènes partiellement atteintes et produisant des équations encore meilleures, et nous obtenons un projet qui, à mon sens, est une réussite.

Il me tarde donc de voir ces équations à l'œuvre lors de la modélisation 3D de la glène reconstituée. Ce projet m'a permis de me plonger dans un domaine qui m'intéresse tout particulièrement, celui de la médecine. Il a d'ailleurs été déterminant dans le choix de mon option *imagerie médicale* et pourrait par ailleurs orienter mes futurs choix professionnels.

Enfin et pour finir, je tiens à remercier mes encadrants pour leurs conseils d'une aide précieuse et leur disponibilité face à mes demandes.

En vous remerciant d'avoir lu ce rapport,

Théo Voisin.

# Annexes

$\alpha=0.05$																
$v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	
$v_2$	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65
	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55
	$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46

Source : Applications of statistics in measurement and testing [WWW8]

Figure 1 – Table du test de Fisher avec alpha à 5 pourcent

```

|Y1_5 = (0.2447583 X5_6) + (0.2399251 X6_11) + (-0.3399449 AngleA) + (-2.472675 AngleE) + 366.4971
Breusch-Godfrey test | P-value = 0.6899041624732528
Breusch-Pagan test | P-value = 0.37376545814466
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.7090 -1.6147  0.0583  1.1158  5.1280
Coefficients:
(Intercept) 366.49711  18.68064  19.619 < 2e-16 ***
X5_6        0.24476   0.04526   5.408  3.2e-06 ***
X6_11      0.23993   0.05585   4.296  0.000108 ***
AngleA     -0.33994   0.08155  -4.168  0.000160 ***
AngleE     -2.47268   0.11185  -22.108 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.134 on 40 degrees of freedom
Multiple R-squared:  0.955, Adjusted R-squared:  0.9505
F-statistic: 212 on 4 and 40 DF, p-value: < 2.2e-16

|Y1_7 = (-0.4798803 X5_10) + (0.3833128 X6_11) + (0.7158278 X7_10) + (-1.099954 AngleE) + 151.7611
Breusch-Godfrey test | P-value = 0.111959651803557
Breusch-Pagan test | P-value = 0.319227207771759
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.8367 -2.0688  0.0193  1.0000  8.5179
Coefficients:
(Intercept) 151.76115  33.30916  4.556  4.81e-05 ***
X5_10      -0.47988   0.11684  -4.107  0.000192 ***
X6_11      0.38331   0.07350   5.215  5.95e-06 ***
X7_10      0.71583   0.09502   7.534  3.42e-09 ***
AngleE     -1.09995   0.25396  -4.331  9.68e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.592 on 40 degrees of freedom
Multiple R-squared:  0.8757, Adjusted R-squared:  0.8633
F-statistic: 70.48 on 4 and 40 DF, p-value: < 2.2e-16

|Y1_6 = (0.2012593 X6_9) + (0.6740268 X6_11) + (0.3826262 X10_11) + (0.5254944 AngleE) + -76.87529
Breusch-Godfrey test | P-value = 0.21563297783829
Breusch-Pagan test | P-value = 0.449171026574256
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.626 -1.559 -0.039  1.569  7.250
Coefficients:
(Intercept) -76.87529  18.15827  -4.234  0.000131 ***
X6_9         0.20126   0.04752   4.235  0.000130 ***
X6_11       0.67403   0.05451  12.365  3.02e-15 ***
X10_11      0.38263   0.06777   5.646  1.49e-06 ***
AngleE      0.52549   0.12641   4.157  0.000165 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.669 on 40 degrees of freedom
Multiple R-squared:  0.9491, Adjusted R-squared:  0.944
F-statistic: 186.5 on 4 and 40 DF, p-value: < 2.2e-16

|Y1_10 = (0.4470809 X5_10) + (0.4109172 AngleA) + (0.6576005 AngleE) + (0.2451707 AngleF) + -141.2962
Breusch-Godfrey test | P-value = 0.8475368089076
Breusch-Pagan test | P-value = 0.371601508598987
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.9771 -1.3581 -0.0681  1.4285  3.2627
Coefficients:
(Intercept) -141.29621  22.88382  -6.175  2.69e-07 ***
X5_10       0.44708   0.03788  11.803  1.32e-14 ***
AngleA      0.41092   0.06713   6.121  3.19e-07 ***
AngleE      0.65760   0.14083   4.669  3.37e-05 ***
AngleF      0.24517   0.06135   3.996  0.000269 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.88 on 40 degrees of freedom
Multiple R-squared:  0.7792, Adjusted R-squared:  0.7571
F-statistic: 35.29 on 4 and 40 DF, p-value: 1.257e-12

```

Figure 2 – Equations obtenues pour le point 1, glène totalement atteinte

```

|y2_5 = (0.5804365 X5_7) + (0.5894362 X6_7) + (-2.89036 AngleE) + 408.5866
Breusch-Godfrey test | P-value = 0.286415921283218
Breusch-Pagan test | P-value = 0.806927977890255
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.9390 -1.5569  0.2672  1.4321  3.5246
Coefficients:
(Intercept) 408.58658  13.02059  31.38 <2e-16 ***
X5_7        0.58044  0.03279  15.26 <2e-16 ***
X6_7        0.58944  0.02569  22.94 <2e-16 ***
AngleE     -2.89036  0.10207  -28.32 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.915 on 41 degrees of freedom
Multiple R-squared:  0.9629, Adjusted R-squared:  0.9602
F-statistic: 355.2 on 3 and 41 DF, p-value: < 2.2e-16

|y2_6 = (0.3306997 X5_10) + (0.2904502 X6_9) + (1.809847 AngleE) + (-0.5410745 AngleFG) + -173.832
Breusch-Godfrey test | P-value = 0.860868394021972
Breusch-Pagan test | P-value = 0.820943365250578
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-11.3417 -2.5166  0.5478  2.2402  6.9892
Coefficients:
(Intercept) -173.83203  43.02495 -4.040 0.000235 ***
X5_10       0.33070  0.08907  3.713 0.000625 ***
X6_9       0.29045  0.06605  4.397 7.89e-05 ***
AngleE     1.80985  0.33012  5.482 2.52e-06 ***
AngleFG   -0.54107  0.13069 -4.140 0.000174 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.923 on 40 degrees of freedom
Multiple R-squared:  0.8162, Adjusted R-squared:  0.7978
F-statistic: 44.41 on 4 and 40 DF, p-value: 3.345e-14

|y2_7 = (-0.6828073 X5_10) + (0.4031751 X6_7) + (0.9187141 X7_10) + (-1.901659 AngleE) + 276.3695
Breusch-Godfrey test | P-value = 0.809771030009891
Breusch-Pagan test | P-value = 0.340237960240643
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.2707 -0.9124 -0.4259  1.0029  5.3423
Coefficients:
(Intercept) 276.36952  30.63101  9.023 3.44e-11 ***
X5_10       -0.68281  0.09606 -7.108 1.33e-08 ***
X6_7        0.40318  0.03964 10.171 1.18e-12 ***
X7_10       0.91871  0.07580 12.121 5.70e-15 ***
AngleE     -1.90166  0.23501 -8.092 5.94e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.77 on 40 degrees of freedom
Multiple R-squared:  0.9226, Adjusted R-squared:  0.9149
F-statistic: 119.2 on 4 and 40 DF, p-value: < 2.2e-16

|y2_10 = (0.247702 X6_9) + (0.4851283 X9_10) + (-0.4557821 AngleE) + (0.2381742 AngleFG) + 45.28515
Breusch-Godfrey test | P-value = 0.713036283799918
Breusch-Pagan test | P-value = 0.249457915230863
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.3201 -2.1403  0.7734  1.5487  3.5944
Coefficients:
(Intercept) 45.28515 18.13158  2.498 0.01672 *
X6_9       0.24770  0.03026  8.187 4.42e-10 ***
X9_10     0.48513  0.06356  7.633 2.50e-09 ***
AngleE    -0.45578  0.16211 -2.812 0.00760 **
AngleFG   0.23817  0.07858  3.031 0.00426 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.331 on 40 degrees of freedom
Multiple R-squared:  0.7564, Adjusted R-squared:  0.7321
F-statistic: 31.06 on 4 and 40 DF, p-value: 8.71e-12

```

Figure 3 – Equations obtenues pour le point 2, glène totalement atteinte

```

|Y3_5 = (0.3737541 X5_7) + (0.3822917 X5_10) + (0.2380053 X6_9) + (-1.187415 AngleE) + 160.5824
Breusch-Godfrey test | P-value = 0.520518676162298
Breusch-Pagan test | P-value = 0.182819905181536
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-7.6938 -2.2014 -0.1112  2.3058  7.7241
Coefficients:
(Intercept) 160.58237  39.18063  4.099 0.000197 ***
X5_7        0.37375   0.06384   5.855 7.57e-07 ***
X5_10       0.38229   0.08269   4.623 3.90e-05 ***
X6_9        0.23801   0.05720   4.161 0.000153 ***
AngleE     -1.18741   0.27703  -4.286 0.000111 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.452 on 40 degrees of freedom
Multiple R-squared:  0.8837, Adjusted R-squared:  0.8721
F-statistic: 75.99 on 4 and 40 DF, p-value: < 2.2e-16

|Y3_6 = (-0.2967399 X5_11) + (0.6408192 X6_7) + (0.3411711 X7_9) + (-0.7353851 AngleFG) + 91.23267
Breusch-Godfrey test | P-value = 0.678719545641151
Breusch-Pagan test | P-value = 0.0598517565243556
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-8.8132 -2.6899 -0.0419  2.3946 11.5523
Coefficients:
(Intercept)  91.23267  14.52427  6.281 1.90e-07 ***
X5_11       -0.29674   0.10826  -2.741 0.009110 ***
X6_7         0.64082   0.06415   9.989 1.99e-12 ***
X7_9         0.34117   0.08183   4.169 0.000159 ***
AngleFG     -0.73539   0.15067  -4.881 1.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.381 on 40 degrees of freedom
Multiple R-squared:  0.8101, Adjusted R-squared:  0.7911
F-statistic: 42.65 on 4 and 40 DF, p-value: 6.427e-14

|Y3_7 = (-0.6226786 X5_10) + (0.2615605 X6_7) + (0.9530643 X7_10) + (-1.616094 AngleE) + 236.2671
Breusch-Godfrey test | P-value = 0.339696454481282
Breusch-Pagan test | P-value = 0.18122783930632
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.8780 -1.2355 -0.0832  1.3024  5.6982
Coefficients:
(Intercept) 236.26713  45.47340  5.196 6.33e-06 ***
X5_10       -0.62268   0.14261  -4.366 8.68e-05 ***
X6_7         0.26156   0.05885   4.445 6.81e-05 ***
X7_10        0.95306   0.11252   8.470 1.84e-10 ***
AngleE     -1.61609   0.34888  -4.632 3.79e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.628 on 40 degrees of freedom
Multiple R-squared:  0.8359, Adjusted R-squared:  0.8173
F-statistic: 50.2 on 4 and 40 DF, p-value: 4.528e-15

|Y3_11 = (0.2011632 X5_7) + (0.2718382 X5_10) + (0.2318764 X6_9) + (-0.3328429 X7_9) + 18.98787
Breusch-Godfrey test | P-value = 0.357987678139375
Breusch-Pagan test | P-value = 0.124979382295783
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.8305 -1.9518  0.4003  1.8127  6.7815
Coefficients:
(Intercept) 18.98787  6.06948  3.128 0.003274 **
X5_7         0.20116   0.04801   4.190 0.000149 ***
X5_10        0.27184   0.04601   5.908 6.37e-07 ***
X6_9         0.23188   0.04513   5.138 7.61e-06 ***
X7_9        -0.33284   0.06640  -5.013 1.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.841 on 40 degrees of freedom
Multiple R-squared:  0.6929, Adjusted R-squared:  0.6622
F-statistic: 22.56 on 4 and 40 DF, p-value: 8.282e-10

```

Figure 4 – Equations obtenues pour le point 3, glène totalement atteinte

```

|V4_5 = (0.3003542 X5_7) + (0.5095748 X6_7) + (-2.110799 AngleE) + (-0.384493 AngleFG) + 329.7239
Breusch-Godfrey test | P-value = 0.302569173928805
Breusch-Pagan test | P-value = 0.251404296822935
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.2353 -1.7258  0.2769  2.0702  4.8769
Coefficients:
(Intercept) 329.72385  21.99777  14.989 < 2e-16 ***
X5_7         0.30035  0.04724  6.359 1.48e-07 ***
X6_7         0.50957  0.03543  14.384 < 2e-16 ***
AngleE      -2.11080  0.19351 -10.908 1.49e-13 ***
AngleFG     -0.38449  0.08472  -4.538 5.09e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.629 on 40 degrees of freedom
Multiple R-squared:  0.9219, Adjusted R-squared:  0.9141
F-statistic: 118.1 on 4 and 40 Df, p-value: < 2.2e-16

|V4_6 = (-0.3958667 X5_11) + (0.5953625 X6_7) + (0.4458263 X7_10) + (-0.7911668 AngleFG) + 88.91776
Breusch-Godfrey test | P-value = 0.471775069376813
Breusch-Pagan test | P-value = 0.621615923986852
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-7.2106 -3.2991 -0.3857  2.4050  14.6988
Coefficients:
(Intercept) 88.91776  15.60539  5.698 1.26e-06 ***
X5_11       -0.39589  0.10798  -3.666 0.000710 ***
X6_7        0.59536  0.06540  9.104 2.69e-11 ***
X7_10       0.44583  0.09058  4.922 1.52e-05 ***
AngleFG     -0.79117  0.15175  -5.214 5.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.379 on 40 degrees of freedom
Multiple R-squared:  0.8111, Adjusted R-squared:  0.7922
F-statistic: 42.94 on 4 and 40 Df, p-value: 5.759e-14

|V4_7 = (0.2550248 X5_10) + (0.9616215 X7_12) + (0.314395 AngleA) + (0.7425003 AngleE) + -109.2129
Breusch-Godfrey test | P-value = 0.759290528396364
Breusch-Pagan test | P-value = 0.930062273218501
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.588 -1.247 -0.417  1.425  6.455
Coefficients:
(Intercept) -109.21292  26.34577  -4.145 0.000171 ***
X5_10        0.25502  0.05120  4.981 1.26e-05 ***
X7_12        0.96162  0.09119  10.545 4.09e-13 ***
AngleA       0.31439  0.08102  3.880 0.000381 ***
AngleE       0.74250  0.14805  5.015 1.13e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.198 on 40 degrees of freedom
Multiple R-squared:  0.893, Adjusted R-squared:  0.8823
F-statistic: 83.49 on 4 and 40 Df, p-value: < 2.2e-16

|V4_9 = (0.3560735 X6_9) + (0.3159914 X9_10) + (-0.7008121 AngleE) + (0.3101575 AngleFG) + 64.84216
Breusch-Godfrey test | P-value = 0.4744075719595602
Breusch-Pagan test | P-value = 0.213502459114845
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.4109 -1.3896  0.0256  1.6430  5.1032
Coefficients:
(Intercept) 64.84216  20.61889  3.145 0.0003131 **
X6_9        0.35607  0.03441  10.349 7.11e-13 ***
X9_10       0.31599  0.07228  4.372 8.53e-05 ***
AngleE     -0.70081  0.18435  -3.802 0.000481 ***
AngleFG     0.31016  0.08936  3.471 0.001258 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.651 on 40 degrees of freedom
Multiple R-squared:  0.7563, Adjusted R-squared:  0.7319
F-statistic: 31.03 on 4 and 40 Df, p-value: 8.812e-12

```

Figure 5 – Equations obtenues pour le point 4, glène totalement atteinte

```

Y14_5 = (0.3999755 X5_7) + (0.4297512 X6_9) + (0.2406509 X6_11) + (-2.59244 AngleE) + 348.6347
Breusch-Godfrey test | P-value = 0.751151086349987
Breusch-Pagan test | P-value = 0.73526447338565

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2790 -0.8545  0.0865  1.0357  4.1480

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  348.63468    17.42024   20.013 < 2e-16 ***
X5_7         0.39998     0.04073    9.821 3.25e-12 ***
X6_9         0.42975     0.04002   10.737 2.39e-13 ***
X6_11        0.24065     0.04593    5.239 5.51e-06 ***
AngleE      -2.59244     0.12084  -21.454 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.964 on 40 degrees of freedom
Multiple R-squared:  0.9636, Adjusted R-squared:  0.9599
F-statistic: 264.6 on 4 and 40 DF, p-value: < 2.2e-16

Y14_6 = (0.5408343 X5_10) + (0.4129013 X10_11) + (2.350063 AngleE) + (-0.4931689 AngleFG) + -243.0524
Breusch-Godfrey test | P-value = 0.436807190293923
Breusch-Pagan test | P-value = 0.682048380051287

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-10.1232 -3.0016 -0.6436  3.0735  10.1091

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -243.05240    49.83510  -4.877 1.75e-05 ***
X5_10        0.54083     0.09786    5.527 2.18e-06 ***
X10_11       0.41290     0.13403    3.081 0.00373 **
AngleE       2.35006     0.35324    6.653 5.73e-08 ***
AngleFG     -0.49317     0.16347   -3.017 0.00443 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.015 on 40 degrees of freedom
Multiple R-squared:  0.741, Adjusted R-squared:  0.7151
F-statistic: 28.61 on 4 and 40 DF, p-value: 2.92e-11

Y14_7 = (0.3218083 X5_6) + (-0.218767 X5_7) + (-0.2519026 X6_9) + (0.5807454 X7_9) + 39.72833
Breusch-Godfrey test | P-value = 0.206656196394182
Breusch-Pagan test | P-value = 0.901191844288465

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-8.5818 -2.0203  0.1654  1.6196  6.5752

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  39.72833    7.22485    5.499 2.39e-06 ***
X5_6         0.32181     0.06928    4.645 3.65e-05 ***
X5_7        -0.21877     0.07981   -2.741 0.00911 **
X6_9        -0.25190     0.07862   -3.204 0.00266 ***
X7_9         0.58075     0.07347    7.905 1.06e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.384 on 40 degrees of freedom
Multiple R-squared:  0.7463, Adjusted R-squared:  0.721
F-statistic: 29.42 on 4 and 40 DF, p-value: 1.936e-11

Y14_11 = (0.2148798 X5_7) + (0.2767603 X5_10) + (0.1997917 X6_9) + (-0.2906282 X7_9) + 20.72099
Breusch-Godfrey test | P-value = 0.585370249123949
Breusch-Pagan test | P-value = 0.192204043147106

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5263 -1.6302  0.2931  1.8692  7.0563

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.72099    6.57834    3.150 0.003087 **
X5_7         0.21488     0.05204    4.129 0.000180 ***
X5_10        0.27676     0.04987    5.549 2.03e-06 ***
X6_9         0.19979     0.04891    4.085 0.000206 ***
X7_9        -0.29063     0.07196   -4.039 0.000237 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.079 on 40 degrees of freedom
Multiple R-squared:  0.6535, Adjusted R-squared:  0.6188
F-statistic: 18.86 on 4 and 40 DF, p-value: 8.758e-09

```

Figure 6 – Equations obtenues pour le point 14, glène totalement atteinte

```

Y15_5 = (0.4192924 X5_7) + (0.2990987 X5_11) + (0.430265 X6_7) + (-1.987062 AngleE) + 274.8908
Breusch-Godfrey test | P-value = 0.372080637924113
Breusch-Pagan test | P-value = 0.58595185656592
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-7.5877 -1.1437 -0.1004  1.8054  4.2759
Coefficients:
(Intercept) 274.89078  34.42075  7.986  8.25e-10 ***
X5_7        0.41929  0.04651  9.016  3.51e-11 ***
X5_11      0.29910  0.07110  4.207  0.000142 ***
X6_7       0.43027  0.05109  8.422  2.14e-10 ***
AngleE     -1.98706  0.24243  -8.196  4.29e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.684 on 40 degrees of freedom
Multiple R-squared:  0.9291, Adjusted R-squared:  0.922
F-statistic: 130.9 on 4 and 40 DF, p-value: < 2.2e-16

Y15_6 = (-0.4900719 X5_10) + (0.5204356 X6_7) + (0.741909 X7_10) + (-0.7187124 AngleF6) + 58.53842
Breusch-Godfrey test | P-value = 0.447575790556025
Breusch-Pagan test | P-value = 0.592677449546045
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-12.5605 -3.0702  0.8573  2.8252  11.7809
Coefficients:
(Intercept) 58.53842  16.30494  3.590  0.000893 ***
X5_10      -0.49007  0.14909  -3.287  0.002113 ***
X6_7       0.52044  0.07277  7.151  1.16e-08 ***
X7_10     0.74191  0.15072  4.922  1.52e-05 ***
AngleF6   -0.71871  0.17176  -4.184  0.000152 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.139 on 40 degrees of freedom
Multiple R-squared:  0.7511, Adjusted R-squared:  0.7263
F-statistic: 30.18 on 4 and 40 DF, p-value: 1.329e-11

Y15_7 = (-0.8279634 X5_10) + (0.4474301 X6_7) + (1.106198 X7_10) + (-2.13117 AngleE) + 291.1693
Breusch-Godfrey test | P-value = 0.377686211031375
Breusch-Pagan test | P-value = 0.537705552855151
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-14.576 -2.180  0.671  1.564  8.054
Coefficients:
(Intercept) 291.16930  63.57166  4.580  4.46e-05 ***
X5_10      -0.82796  0.19937  -4.153  0.000167 ***
X6_7       0.44743  0.08227  5.439  2.90e-06 ***
X7_10     1.10620  0.15731  7.032  1.69e-08 ***
AngleE    -2.13117  0.48773  -4.370  8.60e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.674 on 40 degrees of freedom
Multiple R-squared:  0.7818, Adjusted R-squared:  0.76
F-statistic: 35.83 on 4 and 40 DF, p-value: 9.958e-13

Y15_9 = (-0.1734441 X5_7) + (0.398352 X6_9) + (-0.96761 AngleE) + (0.3889494 AngleF6) + 95.40304
Breusch-Godfrey test | P-value = 0.562657154461364
Breusch-Pagan test | P-value = 0.641891974481929
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-9.2854 -1.7083  0.1704  1.7505  7.8803
Coefficients:
(Intercept) 95.40304  28.93847  3.297  0.002057 **
X5_7       0.17344  0.06402  2.709  0.009881 **
X6_9      0.39835  0.04621  8.620  1.17e-10 ***
X7_10    0.96761  0.25905  -3.735  0.000585 ***
AngleE   -0.96761  0.12167  3.197  0.002715 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.556 on 40 degrees of freedom
Multiple R-squared:  0.6829, Adjusted R-squared:  0.6512
F-statistic: 21.54 on 4 and 40 DF, p-value: 1.548e-09

```

Figure 7 – Equations obtenues pour le point 15, glène totalement atteinte

```

|Y16_5 = (0.2829582 X5_7) + (0.4261597 X5_10) + (0.2782309 X6_7) + (-1.439692 AngleE) + 199.3776
Breusch-Godfrey test | P-value = 0.44689784811442
Breusch-Pagan test | P-value = 0.447709159352502
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-14.8371 -1.4887  0.4084  1.7276  5.5871
Coefficients:
(Intercept) 199.37764  51.05634  3.905 0.000354 ***
X5_7        0.28296  0.06794  4.105 0.00161 ***
X5_10       0.42616  0.09253  4.606 4.12e-05 ***
X6_7        0.27823  0.07046  3.949 0.000310 ***
AngleE     -1.43969  0.36360 -3.960 0.000300 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.514 on 40 degrees of freedom
Multiple R-squared:  0.6825, Adjusted R-squared:  0.8708
F-statistic: 75.14 on 4 and 40 DF, p-value: < 2.2e-16

|Y16_6 = (0.3750512 X5_10) + (0.3595801 X6_9) + (1.794491 AngleE) + (-0.4275761 AngleF6) + -181.7978
Breusch-Godfrey test | P-value = 0.213834654935915
Breusch-Pagan test | P-value = 0.940526271327165
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
 -8.3323 -1.7136 -0.5351  2.3665  9.5802
Coefficients:
(Intercept) -181.79779  42.34828 -4.293 0.000109 ***
X5_10       0.37505  0.08767  4.278 0.000114 ***
X6_9        0.35958  0.06501  5.531 2.15e-06 ***
AngleE      1.79449  0.32493  5.523 2.21e-06 ***
AngleF6    -0.42758  0.12863 -3.324 0.001906 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.861 on 40 degrees of freedom
Multiple R-squared:  0.8488, Adjusted R-squared:  0.8337
F-statistic: 56.16 on 4 and 40 DF, p-value: 6.971e-16

|Y16_7 = (0.2214111 X6_7) + (-0.3381554 X6_9) + (0.7838069 X7_9) + (-0.2609212 AngleF6) + 53.56443
Breusch-Godfrey test | P-value = 0.49897526036818
Breusch-Pagan test | P-value = 0.0764331300618794
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-11.3285 -2.1137  0.0841  2.1788  9.5690
Coefficients:
(Intercept) 53.56443 11.07517  4.836 1.99e-05 ***
X6_7        0.22141  0.07740  2.861 0.006689 ***
X6_9       -0.33816  0.09247 -3.657 0.000735 ***
X7_9        0.78381  0.08727  8.981 3.89e-11 ***
AngleF6    -0.26092  0.10276 -2.539 0.015108 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.933 on 40 degrees of freedom
Multiple R-squared:  0.7491, Adjusted R-squared:  0.7241
F-statistic: 29.86 on 4 and 40 DF, p-value: 1.557e-11

|Y16_11 = (0.1818602 X5_6) + (0.1713791 X5_10) + (-0.1394386 X7_9) + (0.2602913 X10_11) + 13.03325
Breusch-Godfrey test | P-value = 0.791262136047161
Breusch-Pagan test | P-value = 0.449820691933455
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
 -5.7486 -1.6109 -0.2339  1.3993  6.9528
Coefficients:
(Intercept) 13.03325  5.84606  2.229 0.031468 *
X5_6        0.18186  0.03381  5.379 3.51e-06 ***
X5_10       0.17138  0.04610  3.718 0.000616 ***
X7_9       -0.13944  0.05387 -2.588 0.013374 *
X10_11      0.26029  0.07137  3.647 0.000757 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.738 on 40 degrees of freedom
Multiple R-squared:  0.7339, Adjusted R-squared:  0.7073
F-statistic: 27.58 on 4 and 40 DF, p-value: 4.979e-11

```

Figure 8 – Equations obtenues pour le point 16, glène totalement atteinte

```

Y17_5 = (0.2900287 X6_7) + (0.3615061 X7_10) + (-0.4363095 AngleA) + (-2.200669 AngleE) + 329.8873
Breusch-Godfrey test | P-value = 0.812407136991159
Breusch-Pagan test | P-value = 0.912426859461307

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.9800 -1.9961  0.1975  2.8807  5.6069

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  329.88727    31.13050   10.597 3.53e-13 ***
X6_7         0.29003    0.05419    5.352 3.84e-06 ***
X7_10        0.36151    0.06451    5.604 1.70e-06 ***
AngleA      -0.43631    0.11777   -3.705 0.000639 ***
AngleE     -2.20067    0.19409  -11.338 4.57e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.392 on 40 degrees of freedom
Multiple R-squared:  0.886, Adjusted R-squared:  0.8746
F-statistic: 77.74 on 4 and 40 DF, p-value: < 2.2e-16

Y17_6 = (-0.3692147 X5_11) + (0.7006764 X6_7) + (0.4169939 X7_10) + (-0.7898577 AngleF) + 89.18389
Breusch-Godfrey test | P-value = 0.23227791546678
Breusch-Pagan test | P-value = 0.140345876435663

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-7.0939 -2.6471 -0.8204  2.7180  9.2368

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.18389    14.90084    5.985 4.96e-07 ***
X5_11       -0.36921    0.10310   -3.581 0.000917 ***
X6_7         0.70068    0.06244   11.221 6.29e-14 ***
X7_10        0.41699    0.08649    4.821 2.09e-05 ***
AngleF      -0.78986    0.14490   -5.451 2.79e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.182 on 40 degrees of freedom
Multiple R-squared:  0.8541, Adjusted R-squared:  0.8395
F-statistic: 58.54 on 4 and 40 DF, p-value: 3.456e-16

Y17_7 = (0.3065301 X5_6) + (0.1811857 X5_11) + (-0.2559564 X6_7) + (0.3145607 X7_9) + 32.67402
Breusch-Godfrey test | P-value = 0.3539900378020548
Breusch-Pagan test | P-value = 0.764011370540925

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.2638 -2.0122  0.0415  1.8948  7.7066

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.67402    6.74609    4.843 1.95e-05 ***
X5_6         0.30693    0.06401    4.795 2.27e-05 ***
X5_11        0.18119    0.05758    3.147 0.003115 **
X6_7        -0.25596    0.07017   -3.647 0.000756 ***
X7_9         0.31456    0.06086    5.169 6.90e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.172 on 40 degrees of freedom
Multiple R-squared:  0.7854, Adjusted R-squared:  0.7639
F-statistic: 36.59 on 4 and 40 DF, p-value: 7.199e-13

Y17_10 = (-0.1958152 X5_6) + (-0.114408 X5_11) + (0.2544495 X6_7) + (0.4332581 X7_10) + -10.42796
Breusch-Godfrey test | P-value = 0.446897365228111
Breusch-Pagan test | P-value = 0.88900209558153

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-3.6598 -1.4904  0.4159  1.4675  3.9499

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.42796    4.81302   -2.167 0.036278 *
X5_6        -0.19582    0.04977   -3.935 0.000323 ***
X5_11       -0.11441    0.03994   -2.865 0.006618 **
X6_7         0.25445    0.04984    5.106 8.45e-06 ***
X7_10        0.43326    0.05200    8.331 2.83e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.085 on 40 degrees of freedom
Multiple R-squared:  0.7331, Adjusted R-squared:  0.7064
F-statistic: 27.47 on 4 and 40 DF, p-value: 5.267e-11

```

Figure 9 – Equations obtenues pour le point 17, glène totalement atteinte

```

|Y18_5 = (0.2479953 X5_7) + (0.4267189 X6_7) + (-2.181798 AngleA) + (-0.4584328 AngleFG) + 346.2156
Breusch-Godfrey test | P-value = 0.681060817016774
Breusch-Pagan test | P-value = 0.546482015285478
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.0776 -2.1835  0.1831  2.2336  5.1452
Coefficients:
(Intercept) 346.21565  22.75539  15.215 < 2e-16 ***
X5_7         0.24800    0.04886   5.075 9.31e-06 ***
X6_7         0.42672    0.03665  11.645 2.01e-14 ***
AngleA      -2.18180    0.20017 -10.900 1.52e-13 ***
AngleFG     -0.45843    0.08764  -5.231 5.65e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.72 on 40 degrees of freedom
Multiple R-squared:  0.9188, Adjusted R-squared:  0.9107
F-statistic: 113.1 on 4 and 40 DF, p-value: < 2.2e-16

Y18_6 = (0.5523057 X6_9) + (0.4934559 X9_10) + (0.4307119 X10_11) + (-0.474728 AngleA) + 46.78945
Breusch-Godfrey test | P-value = 0.454418487598806
Breusch-Pagan test | P-value = 0.597229590713339
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.3977 -2.4007 -0.4434  1.9296  7.7566
Coefficients:
(Intercept) 46.78945  13.08118  3.577 0.000928 ***
X6_9         0.55231    0.04962  11.132 8.03e-14 ***
X9_10        0.49346    0.11022  4.477 6.16e-05 ***
X10_11       0.43071    0.09750  4.418 7.41e-05 ***
AngleA      -0.47473    0.12338  -3.848 0.000419 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.651 on 40 degrees of freedom
Multiple R-squared:  0.9001, Adjusted R-squared:  0.8901
F-statistic: 90.1 on 4 and 40 DF, p-value: < 2.2e-16

|Y18_7 = (0.3410233 X5_6) + (0.1975687 X5_11) + (-0.2726835 X6_7) + (0.2388734 X7_9) + 37.79319
Breusch-Godfrey test | P-value = 0.461960099607173
Breusch-Pagan test | P-value = 0.775354815458484
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.7984 -1.7454 -0.5281  1.6261  6.7981
Coefficients:
(Intercept) 37.79319  6.31756  5.982 5.01e-07 ***
X5_6         0.34102    0.05994  5.689 1.29e-06 ***
X5_11        0.19757    0.05392  3.664 0.000721 ***
X6_7         -0.27268    0.06572  -4.149 0.000169 ***
X7_9         0.23887    0.05699  4.191 0.000149 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.971 on 40 degrees of freedom
Multiple R-squared:  0.7934, Adjusted R-squared:  0.7727
F-statistic: 38.4 on 4 and 40 DF, p-value: 3.39e-13

|Y18_10 = (0.136838 X5_10) + (0.4762699 X7_10) + (-0.4460141 X7_12) + (0.2286251 AngleFG) + -24.98811
Breusch-Godfrey test | P-value = 0.68268803720962
Breusch-Pagan test | P-value = 0.203549081903503
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-7.4757 -1.1321  0.5345  1.0733  5.1307
Coefficients:
(Intercept) -24.98811  7.91118  -3.159 0.003015 **
X5_10        0.13684    0.06722  2.036 0.048453 *
X7_10        0.47627    0.09061  5.257 5.21e-06 ***
X7_12       -0.44601    0.12021  -3.710 0.000629 ***
AngleFG      0.22863    0.07111  3.215 0.002581 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.371 on 40 degrees of freedom
Multiple R-squared:  0.741, Adjusted R-squared:  0.7151
F-statistic: 28.62 on 4 and 40 DF, p-value: 2.91e-11

```

Figure 10 – Equations obtenues pour le point 18, glène totalement atteinte

```

Y19_5 = (0.3862749 X5_11) + (0.21440339 X6_9) + (0.4323178 X9_10) + (-1.241517 AngleF) + 165.4146
Breusch-Godfrey test | P-value = 0.262738795198083
Breusch-Pagan test | P-value = 0.839512361568276
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-10.506  -1.621   0.501   2.211   6.155
Coefficients:
(Intercept) 165.41459  36.38237  4.547 4.96e-05 ***
X5_11       0.38627   0.09116  4.237 0.000129 ***
X6_9       0.21403   0.05736  3.731 0.000592 ***
X9_10     0.43232   0.09880  4.375 8.44e-05 ***
AngleF    -1.24152   0.25777  -4.816 2.12e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.521 on 40 degrees of freedom
Multiple R-squared:  0.8588, Adjusted R-squared:  0.8446
F-statistic: 60.81 on 4 and 40 DF, p-value: < 2.2e-16

Y19_6 = (-0.4853084 X5_11) + (0.6417719 X6_7) + (0.5140298 X7_10) + (-0.8270787 AngleF) + 92.23406
Breusch-Godfrey test | P-value = 0.803060225240252
Breusch-Pagan test | P-value = 0.501905735866975
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-8.1860  -1.8232  -0.0214  1.9639  11.0585
Coefficients:
(Intercept)  92.23406  15.34233  6.012 4.55e-07 ***
X5_11       -0.48531  0.10616  -9.572 4.59e-05 ***
X6_7        0.64177  0.06429  9.982 2.04e-12 ***
X7_10      0.51403  0.08906  5.772 9.89e-07 ***
AngleF     -0.82708  0.14919  -5.544 2.07e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.306 on 40 degrees of freedom
Multiple R-squared:  0.8359, Adjusted R-squared:  0.8195
F-statistic: 50.94 on 4 and 40 DF, p-value: 3.549e-15

Y19_7 = (-0.5247495 X5_10) + (0.3473546 X6_11) + (0.7139001 X7_10) + (-1.107068 AngleF) + 161.6457
Breusch-Godfrey test | P-value = 0.273728470487807
Breusch-Pagan test | P-value = 0.659985960697273
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-11.1292  -1.5629   0.3388   2.1840   8.3597
Coefficients:
(Intercept) 161.6457  49.5751  3.261 0.00228 **
X5_10      -0.5247  0.1739  -3.018 0.00442 **
X6_11      0.3474  0.1094  3.175 0.00288 **
X7_10      0.7139  0.1414  5.048 1.02e-05 ***
AngleF    -1.1071  0.3780  -2.929 0.00559 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.858 on 40 degrees of freedom
Multiple R-squared:  0.7159, Adjusted R-squared:  0.6874
F-statistic: 25.19 on 4 and 40 DF, p-value: 1.802e-10

Y19_10 = (0.4580668 X5_10) + (0.3531564 AngleA) + (1.149461 AngleF) + -183.2155
Breusch-Godfrey test | P-value = 0.196355943061934
Breusch-Pagan test | P-value = 0.339748013320475
Call:
lm(formula = as.matrix(exogene) ~., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.5801  -1.4081  -0.0957  1.9784  6.7880
Coefficients:
(Intercept) -183.21551  36.69098  -4.993 1.15e-05 ***
X5_10       0.45807  0.05952  7.696 1.75e-09 ***
AngleA      0.35316  0.10966  3.220 0.00251 **
AngleF      1.14946  0.20573  5.587 1.67e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.075 on 41 degrees of freedom
Multiple R-squared:  0.5961, Adjusted R-squared:  0.5666
F-statistic: 20.17 on 3 and 41 DF, p-value: 3.454e-08

```

Figure 11 – Equations obtenues pour le point 19, glène totalement atteinte

```

|Y2_5 = (0.578585 X5_6) + (0.07388978 X17_7) + (-2.254008 ANGLEE) + 318.7678
Breusch-Godfrey test | P-value = 0.056029781122071
Breusch-Pagan test | P-value = 0.342720479469377

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-1.54268 -0.28307 -0.05969  0.33364  1.15833

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  318.767786   5.450086   58.489 < 2e-16 ***
X5_6         0.578585    0.009171   63.091 < 2e-16 ***
X17_7       0.0738890    0.020105    3.675 0.000682 ***
ANGLEE     -2.254008    0.034987  -64.423 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5978 on 41 degrees of freedom
Multiple R-squared:  0.9964, Adjusted R-squared:  0.9961
F-statistic: 3770 on 3 and 41 Df, p-value: < 2.2e-16

|Y2_6 = (0.5493551 X17_6) + (0.3770359 X19_6) + (0.4074787 ANGLEA) + -27.18474
Breusch-Godfrey test | P-value = 0.7053535288225017
Breusch-Pagan test | P-value = 0.143036014122417

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2002 -0.9774 -0.2628  1.1598  4.6155

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -27.18474    7.27929  -3.735 0.000573 ***
X17_6       0.54935    0.07830    7.016 1.56e-08 ***
X19_6       0.37704    0.08177    4.611 3.88e-05 ***
ANGLEA      0.40748    0.06657    6.122 2.91e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.946 on 41 degrees of freedom
Multiple R-squared:  0.9537, Adjusted R-squared:  0.9503
F-statistic: 281.3 on 3 and 41 Df, p-value: < 2.2e-16

|Y2_7 = (0.2190522 X6_7) + (0.8004768 X1_7) + (0.493842 ANGLEA) + -24.92254
Breusch-Godfrey test | P-value = 0.604086063766307
Breusch-Pagan test | P-value = 0.202649451991614

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3555 -0.5983  0.1320  0.6508  3.5937

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -24.92254    6.19669  -4.022 0.000242 ***
X6_7         0.21905    0.02004   10.932 1.01e-13 ***
X1_7         0.80048    0.03576   22.384 < 2e-16 ***
ANGLEA      0.49384    0.05331    9.263 1.32e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.465 on 41 degrees of freedom
Multiple R-squared:  0.9456, Adjusted R-squared:  0.9417
F-statistic: 237.8 on 3 and 41 Df, p-value: < 2.2e-16

|Y2_9 = (0.312258 X5_6) + (0.3686669 X6_9) + (-0.5676645 X17_5) + (-1.772584 ANGLEE) + 236.0529
Breusch-Godfrey test | P-value = 0.792865706776313
Breusch-Pagan test | P-value = 0.519296992054727

Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)

Residuals:
    Min       1Q   Median       3Q      Max
-6.4997 -1.9550 -0.2027  2.2549  7.2045

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  236.05294    57.38706   4.113 0.000189 ***
X5_6         0.31272    0.08628   3.577 0.000694 ***
X6_9         0.36867    0.05229   6.984 1.97e-08 ***
X17_5       -0.56766    0.15423  -3.681 0.000666 ***
ANGLEE     -1.77258    0.41097  -4.313 0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.101 on 40 degrees of freedom
Multiple R-squared:  0.7405, Adjusted R-squared:  0.7146
F-statistic: 28.54 on 4 and 40 Df, p-value: 3.028e-11

```

Figure 12 – Equations obtenues pour le point 2, glène partiellement atteinte

```

Y3_5 = (0.3349888 X5_7) + (0.2887637 X6_7) + (0.5366058 X1_5) + (-1.248789 AngleE) + 171.1247
Breusch-Godfrey test | P-value = 0.666179167510715
Breusch-Pagan test | P-value = 0.286947671554012
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.1584 -1.5173 -0.1641  1.8936  4.0208
Coefficients:
(Intercept) 171.12475  39.54485  4.327  9.80e-05 ***
X5_7        0.33499   0.04874  6.873  2.81e-08 ***
X6_7        0.28876   0.04747  6.082  3.62e-07 ***
X1_5        0.53661   0.08544  6.280  1.91e-07 ***
AngleE     -1.24879   0.28196 -4.429  7.15e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.105 on 40 degrees of freedom
Multiple R-squared:  0.9567, Adjusted R-squared:  0.9524
F-statistic: 221.2 on 4 and 40 DF, p-value: < 2.2e-16

Y3_6 = (-0.2973695 X5_6) + (0.5026999 X16_5) + (0.8848539 X17_6) + (1.265644 AngleE) + -168.4199
Breusch-Godfrey test | P-value = 0.80218256129346
Breusch-Pagan test | P-value = 0.81527459637387
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.7975 -1.3638  0.2734  1.4017  6.9387
Coefficients:
(Intercept) -168.41987  37.06925 -4.543  5.01e-05 ***
X5_6        -0.29797  0.06758 -4.409  7.60e-05 ***
X16_5       0.50270  0.10038  5.008  1.15e-05 ***
X17_6       0.88485  0.07354  12.033  7.18e-15 ***
AngleE      1.26564  0.26551  4.767  2.48e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.447 on 40 degrees of freedom
Multiple R-squared:  0.9407, Adjusted R-squared:  0.9348
F-statistic: 158.7 on 4 and 40 DF, p-value: < 2.2e-16

Y3_7 = (0.1416253 X6_7) + (1.040179 X17_7) + (-0.3144897 X18_5) + (-0.7237942 AngleE) + 105.4406
Breusch-Godfrey test | P-value = 0.383115910286234
Breusch-Pagan test | P-value = 0.285246303420721
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-3.4132 -1.0870 -0.0932  1.0513  2.6313
Coefficients:
(Intercept) 105.44065  24.10480  4.374  8.47e-05 ***
X6_7        0.14163   0.03184  4.448  6.73e-05 ***
X17_7       1.04818   0.05730  18.293  < 2e-16 ***
X18_5      -0.31449   0.07264 -4.330  9.73e-05 ***
AngleE     -0.72379   0.17637 -4.104  0.000194 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.483 on 40 degrees of freedom
Multiple R-squared:  0.9471, Adjusted R-squared:  0.9418
F-statistic: 179.1 on 4 and 40 DF, p-value: < 2.2e-16

Y3_9 = (-0.1299561 X5_11) + (0.6195938 X6_9) + (-0.4269121 X16_6) + (-0.3860609 AngleE) + 57.2245
Breusch-Godfrey test | P-value = 0.352769032866785
Breusch-Pagan test | P-value = 0.101022075629158
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.491 -1.942  0.194  1.832  4.035
Coefficients:
(Intercept)  57.22450  24.71234  2.316  0.0258 *
X5_11       -0.12996  0.06180 -2.103  0.0418 *
X6_9        0.61959  0.05363  11.554  2.56e-14 ***
X16_6      -0.42691  0.07567 -5.642  1.51e-06 ***
AngleE     -0.38606  0.17619 -2.191  0.0343 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.377 on 40 degrees of freedom
Multiple R-squared:  0.8016, Adjusted R-squared:  0.7817
F-statistic: 40.4 on 4 and 40 DF, p-value: 1.524e-13

```

Figure 13 – Equations obtenues pour le point 3, glène partiellement atteinte

```

|y4_5 = (0.449193 X5_6) + (-0.1577094 X5_7) + (0.3771865 X18_7) + (-1.875816 AngleA) + 251.0364
Breusch-Godfrey test | P-value = 0.795983054551189
Breusch-Pagan test | P-value = 0.323776641411136
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-4.3076 -1.2266  0.1573  1.2292  3.8473
Coefficients:
(Intercept) 251.03644 19.25838 13.035 5.52e-16 ***
X5_6        0.44919   0.03358 13.375 2.38e-16 ***
X5_7       -0.15771   0.04439 -3.553 0.000994 ***
X18_7      0.37719   0.07583  4.974 1.29e-05 ***
AngleA     -1.87582   0.12085 -15.522 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.03 on 40 degrees of freedom
Multiple R-squared:  0.9535, Adjusted R-squared:  0.9488
F-statistic: 204.8 on 4 and 40 DF, p-value: < 2.2e-16

|y4_6 = (0.7712982 X18_5) + (0.5727466 X19_6) + (0.592584 AngleA) + (1.834984 AngleE) + -291.9334
Breusch-Godfrey test | P-value = 0.116261067567713
Breusch-Pagan test | P-value = 0.836027542597952
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.6565 -2.1761  0.1735  2.1833  5.3336
Coefficients:
(Intercept) -291.93345 60.83254 -4.799 2.24e-05 ***
X18_5        0.77130   0.15407  5.006 1.16e-05 ***
X19_6        0.57275   0.08793  6.513 8.98e-08 ***
AngleA       0.59258   0.12514  4.736 2.74e-05 ***
AngleE       1.83498   0.39239  4.676 3.30e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.865 on 40 degrees of freedom
Multiple R-squared:  0.9191, Adjusted R-squared:  0.9111
F-statistic: 113.7 on 4 and 40 DF, p-value: < 2.2e-16

|y4_7 = (0.143942 X6_7) + (0.5859862 X7_12) + (0.5268743 X18_7) + (0.307699 AngleA) + -23.12627
Breusch-Godfrey test | P-value = 0.079214775250324
Breusch-Pagan test | P-value = 0.43894222153963
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-2.6611 -0.7248 -0.0891  0.9357  3.3723
Coefficients:
(Intercept) -23.12627  5.92222 -3.905 0.000354 ***
X6_7        0.14394   0.01953  7.372 5.73e-09 ***
X7_12       0.58599   0.08754  6.694 5.02e-08 ***
X18_7      0.52687   0.07328  7.190 1.02e-08 ***
AngleA     0.30770   0.05082  6.055 3.96e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.404 on 40 degrees of freedom
Multiple R-squared:  0.9563, Adjusted R-squared:  0.952
F-statistic: 219 on 4 and 40 DF, p-value: < 2.2e-16

|y4_9 = (0.2916838 X6_9) + (0.687766 X16_19) + (-0.6538534 AngleA) + (0.2736808 AngleF6) + 67.42742
Breusch-Godfrey test | P-value = 0.954961482989652
Breusch-Pagan test | P-value = 0.67748793894641
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.7310 -1.4418  0.1084  1.7464  4.7280
Coefficients:
(Intercept) 67.42742 20.55592  3.280 0.002155 **
X6_9        0.29168   0.03654  7.982 8.35e-10 ***
X16_19     0.68777   0.15504  4.436 7.00e-05 ***
X16_19     -0.65385   0.18141 -3.604 0.000857 ***
AngleA      0.27368   0.08764  3.123 0.003324 **
AngleF6     0.27368   0.08764  3.123 0.003324 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.638 on 40 degrees of freedom
Multiple R-squared:  0.7586, Adjusted R-squared:  0.7344
F-statistic: 31.42 on 4 and 40 DF, p-value: 7.309e-12

```

Figure 14 – Equations obtenues pour le point 4, glène partiellement atteinte

```

|Y14_5 = (0.4422833 X5_6) + (0.3171618 X17_5) + (-1.544303 AngleE) + 218.6139
Breusch-Godfrey test | P-value = 0.259607575110017
Breusch-Pagan test | P-value = 0.78767514347854
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-5.5946 -0.8629  0.1416  1.3055  2.6765
Coefficients:
(Intercept) 218.61393  32.93606  6.638 5.37e-08 ***
X5_6        0.44228  0.04864  9.094 2.21e-11 ***
X17_5      0.31716  0.08733  3.632 0.000775 ***
AngleE     -1.54430  0.23403  -6.599 6.09e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.788 on 41 degrees of freedom
Multiple R-squared:  0.969, Adjusted R-squared:  0.968
F-statistic: 427.8 on 3 and 41 Df, p-value: < 2.2e-16

|Y14_6 = (-0.3340882 X5_6) + (0.4107298 X16_5) + (0.9360078 X17_6) + (1.095448 AngleE) + -147.427
Breusch-Godfrey test | P-value = 0.388222052696479
Breusch-Pagan test | P-value = 0.585413174727983
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-9.435 -1.808  0.028  1.927  7.622
Coefficients:
(Intercept) -147.42701  54.55824  -2.702  0.01006 *
X5_6        -0.33409  0.09946  -3.359 0.00173 **
X16_5      0.41073  0.14774  2.780 0.00824 ***
X17_6      0.93601  0.10823  8.648 1.07e-10 ***
AngleE     1.09545  0.39077  2.803 0.00776 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.602 on 40 degrees of freedom
Multiple R-squared:  0.8664, Adjusted R-squared:  0.853
F-statistic: 64.85 on 4 and 40 Df, p-value: < 2.2e-16

|Y14_7 = (0.2695529 X6_7) + (1.004603 X17_7) + (-0.4291524 X18_5) + (-1.150655 AngleE) + 165.2622
Breusch-Godfrey test | P-value = 0.1310808347730393
Breusch-Pagan test | P-value = 0.833085421110383
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-9.5165 -0.9929 -0.1131  1.2922  3.8364
Coefficients:
(Intercept) 165.26216  39.35202  4.200 0.000145 ***
X6_7        0.26955  0.05198  5.186 6.53e-06 ***
X17_7      1.00460  0.09355  10.739 2.37e-13 ***
X18_5     -0.42915  0.11858  -3.619 0.000821 ***
AngleE    -1.15066  0.28794  -3.996 0.000269 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.421 on 40 degrees of freedom
Multiple R-squared:  0.8701, Adjusted R-squared:  0.8571
F-statistic:  67 on 4 and 40 Df, p-value: < 2.2e-16

|Y14_11 = (0.2075594 X5_7) + (0.3209107 X5_10) + (0.896765 X16_19) + (-0.3816264 X17_7) + 30.63603
Breusch-Godfrey test | P-value = 0.2458317544697987
Breusch-Pagan test | P-value = 0.10829551770278
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.5162 -1.5029  0.1578  1.3964  4.8042
Coefficients:
(Intercept) 30.63603  7.24993  4.226 0.000134 ***
X5_7        0.20795  0.05662  3.673 0.000702 ***
X5_10      0.32091  0.05812  5.522 2.22e-06 ***
X16_19     0.89676  0.17262  5.195 6.35e-06 ***
X17_7     -0.38163  0.11577  -3.296 0.002060 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.844 on 40 degrees of freedom
Multiple R-squared:  0.7044, Adjusted R-squared:  0.6749
F-statistic: 23.83 on 4 and 40 Df, p-value: 3.912e-10

```

Figure 15 – Equations obtenues pour le point 14, glène partiellement atteinte

```

|Y15_5 = (0.287341 X5_6) + (0.8135025 X1_5) + (0.4703896 AngleA) + -36.19151
Breusch-Godfrey test | P-value = 0.66884908090578875
Breusch-Pagan test | P-value = 0.090355185625888
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-8.4568 -1.4241  0.1183  1.4419  4.4336
Coefficients:
(Intercept) -36.19151  9.33729 -3.876  0.000376 ***
X5_6        0.28734  0.83487  8.241  3.10e-07 ***
X1_5        0.81350  0.04472  18.193  < 2e-16 ***
AngleA     0.47039  0.09690  4.854  1.79e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.539 on 41 degrees of freedom
Multiple R-squared:  0.9349, Adjusted R-squared:  0.9302
F-statistic: 196.3 on 3 and 41 DF, p-value: < 2.2e-16

|Y15_6 = (0.3573075 X1_7) + (0.8001399 X18_6) + (0.5851105 AngleA) + -62.16341
Breusch-Godfrey test | P-value = 0.688139429677185
Breusch-Pagan test | P-value = 0.424376757958487
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-6.4994 -1.7661  0.9302  1.8070  4.6081
Coefficients:
(Intercept) -62.16341  12.29331 -5.057  9.34e-06 ***
X1_7        0.35731  0.08979  3.979  0.000275 ***
X18_6      0.80014  0.05948  13.452  < 2e-16 ***
AngleA     0.58511  0.10573  5.534  1.99e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.023 on 41 degrees of freedom
Multiple R-squared:  0.9118, Adjusted R-squared:  0.9053
F-statistic: 141.2 on 3 and 41 DF, p-value: < 2.2e-16

|Y15_7 = (0.2526077 X6_7) + (0.4924973 X7_12) + (0.678321 X18_7) + (0.6123392 AngleA) + -58.44626
Breusch-Godfrey test | P-value = 0.2067535521012673
Breusch-Pagan test | P-value = 0.0516260248381904
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-9.4082 -0.9454  0.3463  1.8621  3.8162
Coefficients:
(Intercept) -58.44626  11.91122 -4.907  1.59e-05 ***
X6_7        0.25261  0.03927  6.432  1.17e-07 ***
X7_12       0.49250  0.17607  2.797  0.00789 ***
X18_7      0.67832  0.14738  4.603  4.16e-05 ***
AngleA     0.61234  0.10221  5.991  4.87e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.824 on 40 degrees of freedom
Multiple R-squared:  0.8711, Adjusted R-squared:  0.8582
F-statistic: 67.58 on 4 and 40 DF, p-value: < 2.2e-16

|Y15_9 = (0.1734441 X5_7) + (0.398352 X6_9) + (-0.96761 AngleE) + (0.3889494 AngleFG) + 95.40304
Breusch-Godfrey test | P-value = 0.562657154461364
Breusch-Pagan test | P-value = 0.641891974481929
Call:
lm(formula = as.matrix(exogene) ~ ., data = endogenes)
Residuals:
    Min       1Q   Median       3Q      Max
-9.2854 -1.7083  0.1704  1.7505  7.8803
Coefficients:
(Intercept) 95.40304  28.93847  3.297  0.002057 **
X5_7        0.17344  0.06402  2.709  0.009881 **
X6_9        0.39835  0.04621  8.620  1.17e-10 ***
X6_9        0.39835  0.04621  8.620  1.17e-10 ***
AngleE     -0.96761  0.25905 -3.735  0.000585 ***
AngleFG    0.38895  0.12167  3.197  0.002715 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.556 on 40 degrees of freedom
Multiple R-squared:  0.6829, Adjusted R-squared:  0.6512
F-statistic: 21.54 on 4 and 40 DF, p-value: 1.548e-09

```

Figure 16 – Equations obtenues pour le point 15, glène partiellement atteinte

# Comptes rendus hebdomadaires

## Compte rendu n°1 du 20/09/2015

Travail effectué :

- Lecture du rapport de PFE 2015 de David Boas
- Lecture du rapport de projet de Jessé
- Début d'apprentissage sur la régression
- Installation de R (et RStudio), installation de Blander
- Premiers essais de régression (simple et multiple) sur des jeux de données automobile trouvés sur internet
- Début de rédaction du rapport pour prendre en main le nouveau modèle envoyé par M. Aupetit
- Rédaction d'une première partie d'exposition du problème : omoplate / glène / pathologies / solutions chirurgicales etc.

Travail à venir :

- Continuer de prendre en main R (tracés de plot, options sur la régression(?))
- Apprentissage de la lecture des résultats statistiques de regression
- Continuer à rédiger le rapport sur la régression

## Compte rendu n°2 du 27/09/2015

Travail effectué :

- Poursuite de l'apprentissage sur la régression linéaire simple
- Premier test "a la main" concluant visuellement et statistiquement
- Premier test du même exemple sur R concluant après comparaison avec le précédent
- Poursuite de la rédaction du rapport sur la régression linéaire simple

Travail à venir :

- Continuer de prendre en main R via des exercices de régression
- Finir la partie rendant compte du premier test sur R
- Commencer régression multiple : apprentissage et rédaction du rapport

## Compte rendu n°3 du 4/10/2015

Travail effectué :

- Fin de rédaction sur la régression linéaire simple.
- Apprentissage de la régression linéaire multiple.
- Premiers contacts avec les techniques pas à pas.
- Reconfiguration d'un éditeur en local Latex.

Travail à venir :

- Approfondir la présentation de l'omoplate
- Continuer le rapport qui à peu avancé cette semaine notamment sur la régression linéaire multiple.
- Poursuivre sur les techniques de pas à pas.

#### Compte rendu n°4 du 11/10/2015

Travail effectué :

- Fin de la rédaction sur la régression linéaire multiple (avec exemple sur R).
- Apprentissage de la méthode pas à pas.
- Rédaction de la partie sur le pas à pas
- Remodelage de la bibliographie/webographie.
- Début d'apprentissage Blender, comme demandé.

Travail à venir :

- Approfondir la présentation de l'omoplate.
- Après relecture, la comparaison de modèles ne semble pas forcément se faire seulement sur les coefficients de détermination ajustés, à approfondir.
- Rédaction d'une petite partie sur le travail fait sur Blender.
- Rédaction d'une petite partie "outils utilisés" pour présenter plus formellement R/Rstudio.
- Poursuivre sur la présentation du travail de David l'année dernière pour finalement amener sur le travail que j'aurai à faire cette année.

#### Compte rendu n°5 du 18/10/2015

Jeudi 22/10 à partir de 16h je serai en entretien chez Sopra Tours pour mon stage de fin d'études. Un autre entretien devrait survenir après la pause pédagogique (je n'ai pas encore la date exacte).

Travail effectué :

- Ajout de tests sur les individus et sur les résultats de la régression.
- Rédaction d'une petite partie « outils utilisés » pour présenter un peu plus formellement R/Rstudio.
- Suivi d'un tuto Blender.
- Rédaction d'une petite partie sur le travail fait sur Blender.

Travail à venir :

- Approfondir la présentation de l'omoplate et du problème en général.
- Poursuivre sur la présentation du travail de David l'année dernière pour finalement amener sur le travail que j'aurai à faire cette année.

#### Compte rendu n°6 du 22/10/2015

Travail effectué :

- Début de la rédaction sur le travail de David.
- Correction du rapport d'après les notes ajoutées par M. Proust sur le RapportV0.
- Fin de la lecture « approfondie » du rapport de David.
- Poursuite de la rédaction sur le travail de David.

Travail à venir :

- Approfondir la présentation de l'omoplate et du problème en général.
- ???

#### Compte rendu n°7 du 08/11/2015

Travail effectué :

- Ajout d'une rapide mise en pratique de la régression en mode pas à pas sur R
- Poursuite de la rédaction sur le travail de David.
- Beaucoup de problèmes avec la mise en page de certaines images sur latex...

Travail à venir :

- Approfondir la présentation de l'omoplate et du problème en général.
- Approfondir la présentation sur le travail de David qui est beaucoup trop succincte à mon gout (tout en essayant de ne pas retomber dans un simple copié/collé)

#### Compte rendu n°8 du 15/11/2015

Travail effectué :

- Ajout d'une courte partie de description des scripts réalisés l'an passé.
- Poursuite de la rédaction sur le travail de David.
- Lecture d'un cours sur la programmation R (je m'étais jusque là concentré seulement sur l'utilisation de R pour la régression linéaire).
- Rédaction dans la foulée d'un début de présentation du langage R.

Travail à venir :

- Approfondir la présentation de l'omoplate et du problème en général.
- Continuer la lecture du cours sur la programmation en R.
- Continuer la rédaction sur la programmation en R.
- Approfondir la présentation sur le travail de David qui est beaucoup trop succincte à mon gout (tout en essayant de ne pas retomber dans un simple copié/collé)
- Peut-être commencer à faire des tests sur les données ? (il me faudrait dans ce cas les données/matrice de distance).

#### Compte rendu n°9 du 22/11/2015

Travail effectué :

- Relecture du rapport et quelques corrections.
- Discussion avec M. Proust concernant les matrices et notamment le travail à effectuer sur la V7.
- Création d'une feuille excel permettant d'identifier les individus aberrants de la matrice. Ainsi, sauf ajout de variables au problème, les individus aberrants ressortiront tout de suite si nouvelle matrice il y a.

Travail à venir :

- Approfondir la présentation de l'omoplate.
- Continuer la lecture du cours de programmation en R.
- Continuer la rédaction sur la programmation en R.
- Approfondir la présentation du travail de David Boas.

#### Compte rendu n°10 du 29/11/2015

Travail effectué

- Mise en place de la nouvelle version du modèle Latex
- Correction du rapport en suivant les recommandations de M. Proust.
- Fin de la lecture du cours de programmation en R.
- Fin de la rédaction sur la programmation en R.

Travail à venir

- Certains points restent à corriger (notamment des explication à rajouter)
- Approfondir la présentation du travail de David Boas.
- Approfondir la présentation de l'omoplate.

#### Compte rendu n°11 du 6/12/2015

Travail effectué

- Fin des corrections du rapport
- Questionnement avec Min sur l'ajout d'éventuels points sur la glène
- Correction de la feuille Excel de vérification des données
- Ajout de la matrice de corrélation sur la matrice Excel

Travail à venir

- Certains points restent à corriger (notamment des explication à rajouter)
- Approfondir la présentation du travail de David Boas.
- Approfondir la présentation de l'omoplate.

- Commencer à faire quelques régressions de test sur les données que j'ai désormais à ma disposition.

### Compte rendu n°12 du 13/12/2015

#### Travail effectué

- Découverte de Shiny (apprentissage pour les développements futurs)
- Création du poster de présentation du projet

#### Travail à venir

- Certains points restent à corriger (notamment des explication à rajouter)
- Approfondir la présentation du travail de David Boas.
- Approfondir la présentation de l'omoplate.
- Commencer à faire quelques régressions de test sur les données que j'ai désormais à ma disposition.
- Rédiger une partie dans "Outils" sur Shiny

### Compte rendu n°13 du 20/12/2015

#### Travail effectué

- Apprentissage un peu plus poussé de Shiny.
- Fin de rédaction du rapport (ajout de la partie sur Shiny ainsi qu'une conclusion).

#### Travail à venir

- Relire le rapport / le corriger.
- Préparer le diaporama de soutenance.
- Commencer à faire quelques régressions de test sur les données que j'ai désormais à ma disposition.

### Compte rendu n°14 du 28/01/2016

Définition avec M. Proust de la marche à suivre qui est, pour le moment :

- Génération de modèles et comparaisons avec ceux de David Boas.
- Redéfinition des distances que nous prendrons en compte.
- Ajouter un point permettant de rendre compte de la courbure de l'écaille.

A la suite de cela, travail effectué :

- Développement d'un outil java permettant le calcul automatique de distances à partir de chacun des fichiers décrivant les points des individus et rédaction d'un petit chapitre sur cet outil et son fichier de configuration dans le rapport.
- Développement de scripts R permettant de générer une équation expliquant une distance Y en fonction de distances X à partir de la matrice générée par l'outil Java (soit saisies manuellement, soit pas régression en mode pas à pas).
- Cependant, un problème est apparu, pour une même sélection de variables X définissant une variable Y, je ne trouve pas la même équation que David. Investigations sur ce point :
  - Comparaison des deux matrices calculées : DAVIDV7 et THEOV7 sont les mêmes (quelques différences à 10<sup>-12</sup> qui viennent des arrondis lors du calcul mais au vu de la précision dont l'on dispose c'est clairement négligeable). Régression linéaire en step by step sur Y<sub>2,7</sub> avec ma méthode : sur DAVIDV7 et THEOV7 : mêmes résultats. Nous pouvons donc en conclure que les matrices sur lesquelles nous travaillons sont les mêmes, le problème ne vient pas de là.
  - Comparaison des équations :
    - Méthode de David : Créer le modèle avec toutes les variables en utilisant lm(). Passer ce modèle dans du step by step en utilisant stepwise (backward/forward et AIC/BIC). Pour chaque modèle créé, vérifier sa validité statistique
    - Nous prenons donc l'équation n°1 de DAVID créée avec la méthode ci-dessus qui est :  $Y_{2,7} = 0.20493 X_{5,6} + 0.21845 X_{5,11} - 0.14519 X_{6,9} + 0.32906 X_{7,9} + 44.30713$  ( $R^2 = 0.75$ )
    - Nous prenons l'équation générée (pas en step by step, mais en forçant l'utilisation de ces endogènes) par ma méthode (qui utilise un simple lm) et qui donne :  $Y_{2,7} = 0.2113668 X_{5,6} + 0.201372 X_{5,11} - 0.104582 X_{6,9} + 0.3126708 X_{7,9} + 40.06934$  ( $R^2 = 0.79$ )

- Dans le fichier ComparaisonEquation.xlsx, non allons comparer sur 5 individus la différence des estimateurs produits par nos deux modèles : les différences d'estimations sont minimales.
- En conclusion, nous obtenons donc deux équations différentes pour un même jeu de données avec les mêmes variables utilisées. Cependant les différences sont minimales.

## Travail à venir

- Continuer à chercher l'éventuel problème ?
- Définir les distances à utiliser avec M. Proust (prévu mercredi prochain).
- Trouver un moyen de calculer la distance permettant de rendre compte de la courbure de l'écaillage (moins prioritaire pour l'instant).

**Compte rendu n°15 du 07/02/2016**

## Travail effectué

- Fin des investigations sur les différences entre les résultats de David et moi.
- Définition de la marche à suivre avec M. Proust pour un point donné Y :
  - Calculer toutes les équations disponibles avec les distances entre Y et tous les X.
  - Garder les quatre meilleures
  - (Par la suite on reproduira la méthode en limitant le nombre de variables endogènes dans l'équation).
- Début de scripts sur le point précédent.

## Travail à venir

- Rédiger le rapport sur tout ce qui a été fait depuis la dernière fois.
- Continuer le script pour générer les équations.
- Pour plus tard : considérations de glènes partiellement atteintes seulement.

**Compte rendu n°16 du 14/02/2016**

## Travail effectué

- Ajout de la description du travail effectué dans le rapport
- Suivi de la marche à suivre en implémentant les scripts R permettant :
  - De charger la matrice produite par l'outil automatique Java générant les distances souhaitées.
  - De calculer, pour un exogène donné, toutes les équations en fonction des combinaisons disponibles d'endogènes. La taille de combinaison maximum est réglable.
  - Ceci m'a permis, pour la variable Y1 5, de produire des fichiers de sorties contenant les équations ainsi que la description (summary) des équations produites.
  - Ces équations sont filtrées de manière à ne pas passer sous un R2 donné ainsi que au-dessus d'une p-value donnée.
- Vous pourrez donc trouver ci-joint, une archive contenant les résultats, classés par dossier de la forme : BRUTE NBmaxCombinatoire R2 Pvalue.

## Travail à venir

- Rédiger le rapport sur les scripts implémentés et décrits juste au-dessus.
- Améliorer les scripts notamment de manière à mettre en place des tests statistiques sur les modèles afin de filtrer encore un peu plus ces derniers.
- Pour plus tard (et seulement si le temps le permet) Apprendre le développement de scripts blender pour le calcul du point convenu avec M. Proust.
- Pour plus tard : considérations de glènes partiellement atteintes seulement.

**Compte rendu n°17 du 28/02/2016**

## Travail effectué :

- Complétion du rapport.
- Ajout de tests statistiques pour filtrer les modèles valides.
- Présentation du travail à M. Proust.

## Travail à venir :

- Continuer l'ajout de tests statistiques et leur description dans le rapport.
- Faire en sorte de sélectionner les 4 distances à utiliser pour chaque point à expliquer.

- Pour plus tard : considérations de glènes partiellement atteintes seulement.
- Pour plus tard (et seulement si le temps le permet) Apprendre le développement de scripts blender pour le calcul du point convenu avec M. Proust.

#### Compte rendu n°18 du 06/03/2016

##### Travail effectué

- Rapport continué.
- Mise en place d'une automatisation totale => tous les points sont calculés (un répertoire par point et un fichier par distance contenant toutes les équations viables). Les résultats actuellement obtenus sont en pièce jointe avec le rapport.
- Début de développement Shiny.

##### Travail à venir

- Pour plus tard (et seulement si le temps le permet) Apprendre le développement de scripts blender pour le calcul du point convenu avec M. Proust.
- Pour plus tard : considérations de glènes partiellement atteintes seulement.

#### Compte rendu n°19 du 13/03/2016

##### Travail effectué

- Analyse des équations obtenues la semaine dernière, ajout dans le rapport des meilleures. Nous avons maintenant 4 équations de distance par point.
- Lancement de tests en considérant les points hauts de la glène comme sains. Premier test non concluant, ajouté au rapport, mais qui sera complété par un deuxième qui devrait être plus intéressant la semaine prochaine. (changement des distances utilisées en conséquence).
- Rédaction du rapport sur ces parties.
- Fin de l'application Shiny, utilisable et utilisée pour lancer les tests.

##### Travail à venir

- Pour plus tard (et seulement si le temps le permet) Apprendre le développement de scripts blender pour le calcul du point convenu avec M. Proust.
- Analyser les équations obtenues avec les nouvelles distances et ajouter cela au rapport.
- Rédiger le rapport sur ces nouveaux éléments .

#### Compte rendu n°20 du 20/03/2016

##### Travail effectué

- Correction du rapport.
- Fin des calculs avec la glène partiellement saine.
- Analyse des résultats.
- Ajout des résultats complets en annexe pour l'analyse standard.
- Tests sur les données.
- Rédaction de tout cela dans le rapport.

##### Travail à venir

- Ajout de la distance de cook et du covratio.
- Rédaction du rapport / conclusion.
- Finalisation des livrables / rapport / résultats.
- Préparation de la soutenance.

# Webographie

- [WWW1] "AERODARK". Sous la dir. d'OPENCLASSROOMS. URL : <https://openclassrooms.com/courses/debutez-dans-la-3d-avec-blender>.
- [WWW2] BESSE. *Introduction à la régression linéaire multiple*. Université de Toulouse. URL : <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-1-inf-intRegmult.pdf>.
- [WWW3] BESSE. *Scénario : sélection de modèles avec R (prostate)*. URL : <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-scenar-reg-penal-prostate.pdf>.
- [WWW4] The R FOUNDATION. *What is R?* URL : <https://www.r-project.org/about.html>.
- [WWW5] Steve D. JOST. *The F test for linear regression*. Sous la dir. de DePaul UNIVERSITY. URL : <http://facweb.cs.depaul.edu/sjost/csc423/documents/f-test-reg.htm>.
- [WWW6] Bill Venables JUSTIN LOKHORST et Berwin TURLACH. *lasso2 : L1 constrained estimation aka 'lasso'*. URL : <https://cran.r-project.org/web/packages/lasso2/index.html>.
- [WWW7] Anatoli LOUDITSKI. *TP3 : Régression linéaire multiple*. URL : <http://ljk.imag.fr/membres/Anatoli.Iouditski/cours/M1MAI/dm-tp3.pdf>.
- [WWW8] Prof. Leslie PENDRILL. *Applications of statistics in measurement and testing*. URL : <https://metrology.files.wordpress.com/2009/09/fdist95.gif>.
- [WWW9] Dr Jean-Marie PHILIPPEAU. *Chirurgie prothétique de l'épaule*. URL : <http://www.chirurgien-orthopedique.net/chirurgie-anatomie-epaule/chirurgie-prothese-epaule/>.
- [WWW10] RSTUDIO. *Galerie de présentation des possibilités de Shiny*. URL : <http://shiny.rstudio.com/gallery/>.
- [WWW11] XLSTAT. *Tests d'hétéroscédasticité de Breusch-Pagan et de White avec XLSTAT*. URL : [https://help.xlstat.com/customer/fr/portal/articles/2062433-tests-d-h%C3%83%C2%A9t%C3%83%C2%A9roscedasticit%C3%83%C2%A9-de-breusch-pagan-et-de-white-avec-xlstat?b\\_id=9283](https://help.xlstat.com/customer/fr/portal/articles/2062433-tests-d-h%C3%83%C2%A9t%C3%83%C2%A9roscedasticit%C3%83%C2%A9-de-breusch-pagan-et-de-white-avec-xlstat?b_id=9283).

# Bibliographie

- [1] Sébastien BALLESTEROS. *Le modèle linéaire avec R*. Muséum national d'histoire naturelle, 2008. URL : [http://rug.mnhn.fr/semin-r/PDF/semin-R\\_lm\\_SBallesteros\\_110308.pdf](http://rug.mnhn.fr/semin-r/PDF/semin-R_lm_SBallesteros_110308.pdf).
- [2] David BOAS. *Typologie d'une omoplate - Rapport de projet de fin d'études*. Polytech Tours, 2015.
- [3] Dr CAPTIER. *Ostéologie du membre supérieur - Présentation*. Laboratoire d'anatomie de Montpellier, 2010.
- [4] Vincent GOULET. *Introduction à la programmation en R*. Quatrième Edition. URL : [https://cran.r-project.org/doc/contrib/Goulet\\_introduction\\_programmation\\_R.pdf](https://cran.r-project.org/doc/contrib/Goulet_introduction_programmation_R.pdf).
- [5] S Sethom H CHIAVASSE-GANDOIS. *La glène osseuse - Aspect normal, anomalies morphologiques et pathologies à connaître*. 2010. URL : <http://pe.sfrnet.org/Data/ModuleConsultationPoster/pdf/2010/1/498dfe86-54e8-4d4b-8917-38a2e64b1530.pdf>.
- [6] Min JIANG. *Projet de Recherche et Développement - Reconstitution d'une glène artificielle*.
- [7] Pierre-André Cornillon Eric MATZNER-LOBER. *Régression avec R*. Springer, 2011.
- [8] Ricco RAKOTOMALALA. *Régression linéaire simple - Cours*. Laboratoire ERIC. URL : [http://eric.univ-lyon2.fr/~ricco/cours/cours/Regression\\_Lineaire\\_Simple.pdf](http://eric.univ-lyon2.fr/~ricco/cours/cours/Regression_Lineaire_Simple.pdf).
- [9] Ricco RAKOTOMALALA. *Pratique de la régression linéaire multiple - Diagnostic et sélection de variables*. Université Lumière Lyon 2, 2015. URL : [http://eric.univ-lyon2.fr/~ricco/cours/cours/La\\_regression\\_dans\\_la\\_pratique.pdf](http://eric.univ-lyon2.fr/~ricco/cours/cours/La_regression_dans_la_pratique.pdf).
- [10] Henri V. VALLOIS. « L'omoplate humaine (suite) ». fr. In : *Bulletins et Mémoires de la Société d'anthropologie de Paris* 3.1 (1932), p. 3–153. ISSN : 0037-8984. DOI : [10.3406/bmsap.1932.9315](http://www.persee.fr/doc/bmsap_0037-8984_1932_num_3_1_9315). URL : [http://www.persee.fr/doc/bmsap\\_0037-8984\\_1932\\_num\\_3\\_1\\_9315](http://www.persee.fr/doc/bmsap_0037-8984_1932_num_3_1_9315).
- [11] Minghui ZHANG. *Projet de Recherche et Développement - Suivi de glène et réalité augmentée*.

# Reconstitution virtuelle de la morphologie d'une omoplate malade

**Théo VOISIN**

**Encadrement : Mohamed SLIMANE, Julien BERHOUE et Christian PROUST**

## Objectifs

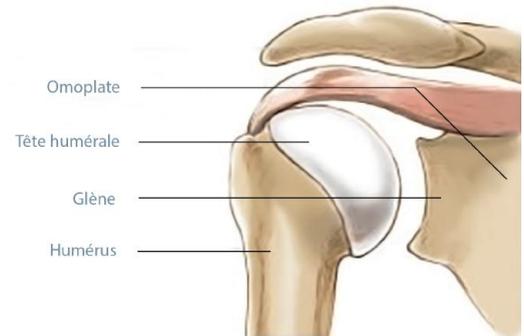
- Venir en aide aux personnes atteintes d'une affection de la glène (arthrose, usure naturelle).
- Permettre la création d'une prothèse reproduisant au mieux la glène telle qu'elle était avant l'apparition des problèmes.
- Première étape du projet EPAULE.

## Mise en œuvre

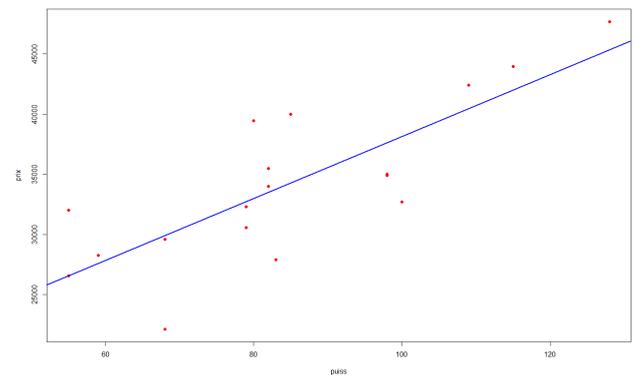
- Détermination de points stratégiques sur l'omoplate.
- Utilisation de la régression linéaire multiple sur des omoplates saines pour trouver des équations permettant de déterminer les points originaux de la glène sur des omoplates malades.

## Résultats attendus

- Utiliser les équations obtenues pour modéliser la prothèse en 3D (Min Jiang)
- Utiliser le modèle 3D sur des lunettes de réalité augmentée utilisées par le chirurgien (Minghui Zhang)



*La glène et ses alentours dans le corps humain*



*Le prix de voitures en fonction de leur puissance (en rouge) et l'équation calculée (en bleu)*



*Visualisation en réalité augmentée de l'omoplate et de la glène à reconstituer*

# Reconstitution virtuelle de la morphologie d'une omoplate malade

## Résumé

Il s'agit du projet de recherche et développement mené en 5<sup>ème</sup> année d'école d'ingénieur informatique à Polytech Tours. Dans un premier temps nous présentons la morphologie de glène ainsi que les problèmes liés. Ensuite nous nous attachons à expliquer le fonctionnement de la régression linéaire afin de l'appliquer, plus tard, aux données à notre disposition pour modéliser au mieux les glènes à partir de glènes malades. Enfin, nous présenteront le développement de l'automatisation destinée à produire des équations pour ensuite présenter les résultats finalement obtenus.

## Mots-clés

omoplate, glène, régression linéaire, modélisation, R

## Abstract

This is the research and development project undertaken during the 5<sup>th</sup> year of computer science engineering school Polytech Tours. First we present the scapula, its socket and its diseases. Then we explain the way the linear regression works in order to apply it, later, to the data we have to model as well as possible the sockets from sick sockets. Finally, we will show the development intended to produce equations and present the results we had.

## Keywords

scapula, linear regression, model, R

## Tuteurs académiques

Mohamed SLIMANE

Julien BERHOUE

Christian PROUST

## Étudiants

Théo VOISIN (DI5)