

## Table des matières

<b>Introduction</b>	<b>3</b>
<b>Objectifs</b>	<b>4</b>
<b>Matériel et Méthodes</b>	<b>4</b>
Composition et structuration de l'échantillon	4
Analyses statistiques	4
<b>Résultats</b>	<b>6</b>
Présentation des données	6
Objectif 1	6
Objectif 2	8
Objectif 3	9
<b>Perspectives et Discussion</b>	<b>10</b>
<b>Annexes</b>	<b>11</b>
<b>Références</b>	<b>16</b>

## Table des figures

1	Structuration des données . . . . .	5
2	Diagramme de décision selon les critères du test Kim [4] . . . .	5
3	Description de la dispersion des variables . . . . .	6
4	Dispersion de notre échantillon . . . . .	7
5	Etudes de facteurs de variabilité : Anova complet . . . . .	8
6	Etudes de facteurs de variabilité : Anova ciblé . . . . .	8
7	Boxplot des delta des fumeurs VS non fumeurs . . . . .	9
8	Protocole de test de comparaison . . . . .	9
9	Code des tests de comparaison de proportion . . . . .	10
10	Résultat du test du Khi2 . . . . .	10
11	Résulta du Test exact . . . . .	10
.12	Dispersion des échantillons de Pulse1 en fonction de la taille des participants et de leur genre . . . . .	12
.13	Code Objectif 1 . . . . .	13
.14	Extrait sortie Code R . . . . .	14
.15	Code Objectif 3 . . . . .	15

# Analyse statistique de l'influence des composantes physiologiques et comportementales sur la fréquence cardiaque.

Ismaël Hachin Loita & Maxime Latry<sup>a</sup>,

*Responsables : M. Jacques Bénichou et Mme Camille Le Clézio*

*<sup>a</sup>Rapport d'analyse dans le cadre de l'Unité d'enseignement Biostatistique appliquées à la médecine personnalisée à l'Université de médecine pharmacie de Rouen.*

---

## Abstract

Dans le cadre d'une étude sur la modulation de la fréquence cardiaque lors d'un exercice et selon les comportements individuels, nous avons été mis en charge de l'analyse d'un jeu de données. Nous avons pu montrer une corrélation fortement positive de l'exercice sur l'augmentation de la fréquence cardiaque ( $p\text{-value} < 2.10^{-16}$ ) mais aussi de la fréquence de consommation tabagique ( $p\text{-value} = 0.025$ ) ainsi que de l'âge ( $p\text{-value} = 0.056$ ). Cependant, la comparaison de moyenne du pouls au repos des fumeurs et des non fumeurs n'a pas permis de montrer de différence significative ( $p\text{-value} = 0.197$ ). Par ailleurs, il n'a pas été possible de montrer de biais d'autosélection concernant la proportion d'hommes ou de fumeurs ayant couru avant ou après le changement de protocole.

---

## Introduction

Dans ce rapport nous étudierons un jeu de données issu de l'expérience du Professeur John Eccleson et le Dr Richard Wilson de l'Université de Queensland. Dans l'expérience l'étu-

diant prend son pouls puis effectue un pile ou face. Si la pièce tombe sur face ils doivent courir sur place pendant 1min, sinon il reste assis pendant une minute. Ensuite, il reprend son pouls et note ses données physiologiques et habitudes comportementales.

---

*Email address: [hachinloitaismael@gmail.com](mailto:hachinloitaismael@gmail.com) & [maxime.latry@gmail.com](mailto:maxime.latry@gmail.com)*  
(Ismaël Hachin Loita & Maxime Latry )

tales. Après quelques expériences, Richard Wilson était inquiet qu'un biais ai pu être introduit dans la proportion de ceux qui courent. En effet, avant 1995 il suspecte que les étudiants fausser l'hypothèse d'un tirage aléatoire équitable entre les coureurs et les non coureurs. Cet éventuel biais d'auto-sélection a été corrigé par une affectation en début d'exercice par le professeur à partir de 1995. Par ailleurs, à partir de 1995, le nombre de coureur n'était pas correctement contrôlé puisque tous les formulaires n'ont pas tous été retournés.

## Objectifs

Au cours de notre analyse, nous tenterons d'évaluer la corrélation entre différents paramètres ainsi que l'influence des composantes physiologiques et comportementales sur la fréquence cardiaque. Voici nos interrogations

1. Quels paramètres influence la fréquence cardiaque dans cette expérience ?
2. Existe-t-il une différence du pouls au repos selon le mode de vie ?
3. L'instauration de la nouvelle méthode change-elle la proportions de coureurs après 1995 ?

# Matériel et Méthodes

## Composition et structuration de l'échantillon

Le logiciel de statistique R a été utilisé pour chaque étude statistique et le code a été produit spécifiquement pour l'étude.

Pour l'importation et la structuration des données (voir fig. [1 page suivante](#)) : nous avons modifié les sujets 102 et 106 qui ont mal été saisis. Exclusion du sujet 77 dont les pouls sont manquants. Ajout d'une colonne contenant les variables delta (différence entre le pouls avant et après exercice) et l'IMC =  $\left(\frac{Poids}{Taille^2}\right)$ . Voir fig. [3 page 6](#). « Exercice » décrit le degré d'activité physique des sujets au quotidien et « Ran » correspond à un exercice effectué pendant 1min ou non.

## Analyses statistiques

Nous décrirons dans cette partie les hypothèses, tests et conditions d'applications. Tous les tests seront effectués à risque  $\alpha = 5\%$

### *Etude des facteur de variabilité*

Pour étudier les facteurs de variabilité de la variance d'un paramètre nous effectuerons le *test d'Anova*.

```

url <- "http://www.statsci.org/data/oz/ms212.txt"
eval7 <- read.delim(url)
download.file(url, "eval7.txt")
eval7$Weight
eval7$Weight[106]<- 72
eval7$Height[102]<- 168
eval7$Height[106]<- 193
eval7$Gender[eval7$Gender == 1] <- "Male"
eval7$Gender[eval7$Gender == 2] <- "Female"
eval7$Gender
eval7$Smokes[eval7$Smokes == 1] <- "Yes"
eval7$Smokes[eval7$Smokes == 2] <- "No"
eval7$Smokes
eval7$Alcohol[eval7$Alcohol == 1] <- "Yes"
eval7$Alcohol[eval7$Alcohol == 2] <- "No"
eval7$Alcohol
eval7$Exercise[eval7$Exercise == 1] <- "Hight"
eval7$Exercise[eval7$Exercise == 2] <- "Moderate"
eval7$Exercise[eval7$Exercise == 3] <- "Low"
eval7$Exercise
eval7$Ran[eval7$Ran == 1] <- "Ran"
eval7$Ran[eval7$Ran == 2] <- "Sat"
eval7$Ran

eval7[106,] <- eval7 %>%
  mutate(IMC = Weight/((Height/100)^2)) %>%
  mutate(delta = Pulse2-Pulse1)

```

FIGURE 1: Structuration des données

L'hypothèse nulle correspond au cas où les distributions suivent la même loi normale. L'hypothèse alternative est qu'il existe au moins une distribution dont la moyenne s'écarte des autres moyennes. La forme générale de l'analyse de variance repose sur le test de Fisher et donc sur la normalité des distributions et l'indépendance des échantillons.

### Variabilité des échantillons

Pour effectuer les tests de comparaisons, nous supposons que les échantillons se comportent de façon normale. En effet, les effectifs des

échantillons testés seront tous supérieurs à 30. Il est possible d'effectuer un test de Shapiro pour tester l'hypothèse de normalité. Nous vérifierons l'égalité des variances par un test de Fisher (conditions d'application : normalité des échantillons).

### Etude comparative

En fonction des résultats précédents, trois tests peuvent alors être effectués :

- Le test de Student, dans le cas d'une normalité des deux échantillons comparés et de variances non significativement différentes. `t.test` [5]
- Le test de Welch, dans le cas où les variances seraient significativement différentes avec normalité des échantillons.
- Le test de Wilcoxon (ou Mann-Witney), dans le cas d'une non normalité des échantillons peu importe les variances. Ce test est non paramétrique car il ne se base pas sur la distribution statistique. `wilcox.test` [6], Hollander et al. [3, 2], Bauer [1]

Normalité ?	Egalité des variances ?	Test à effectuer	Test paramétrique ?
Oui	Oui	Student	Oui
Non	Non	Welch	Oui
	Ignoré	Wilcoxon	Non

FIGURE 2: Diagramme de décision selon les critères du test Kim [4]

```
> summary(eval7)
```

Height	Weight	Age	Gender	Smokes	Alcohol
Min. : 68.0	Min. : 27.00	Min. : 18.00	Min. : 1.000	Min. : 1.0	Min. : 1.000
1st Qu.: 165.2	1st Qu.: 56.25	1st Qu.: 19.00	1st Qu.: 1.000	1st Qu.: 2.0	1st Qu.: 1.000
Median : 172.5	Median : 63.00	Median : 20.00	Median : 1.000	Median : 2.0	Median : 1.000
Mean : 171.6	Mean : 66.33	Mean : 20.56	Mean : 1.464	Mean : 1.9	Mean : 1.382
3rd Qu.: 180.0	3rd Qu.: 75.00	3rd Qu.: 21.00	3rd Qu.: 2.000	3rd Qu.: 2.0	3rd Qu.: 2.000
Max. : 195.0	Max. : 110.00	Max. : 45.00	Max. : 2.000	Max. : 2.0	Max. : 2.000

Exercise	Ran	Pulse1	Pulse2	Year
Min. : 1.000	Min. : 1.000	Min. : 47.00	Min. : 56.0	Min. : 93.00
1st Qu.: 2.000	1st Qu.: 1.000	1st Qu.: 68.00	1st Qu.: 72.0	1st Qu.: 95.00
Median : 2.000	Median : 2.000	Median : 76.00	Median : 84.0	Median : 96.00
Mean : 2.209	Mean : 1.582	Mean : 75.69	Mean : 96.8	Mean : 95.63
3rd Qu.: 3.000	3rd Qu.: 2.000	3rd Qu.: 82.00	3rd Qu.: 125.0	3rd Qu.: 97.00
Max. : 3.000	Max. : 2.000	Max. : 145.00	Max. : 176.0	Max. : 98.00
		NA's : 1	NA's : 1	

FIGURE 3: Description de la dispersion des variables

# Résultats

## Objectif 1

### Présentation des données

Pour comprendre nos données et introduire d'éventuelles corrélations entre différentes variables, la dispersion des résultats des delta est présentée sur le graphique de la fig. page suivante et celle de pulse1 en fonction de la taille Annexe : 2 page 12. Nous pouvons observer une certaine influence de la taille sur le delta grâce à une régression linéaire simple. Aussi, il semblerait que les boxplot soient bien séparés entre ceux qui ont courus et ceux qui sont restés assis. À partir de ces observations, nous répondrons aux questions présentées page 4 par la démonstration.

Quels paramètres influencent Delta ?

*Test utilisé et conditions*

Test d'Anova dont les conditions d'application sont décrites page 4.

*Hypothèses*

$H_0$  : la variabilité de delta suit la même loi normal que les autres variables.

$H_1$  : au moins une distribution dont la moyenne s'écarte des autres moyennes

*Calculs*

Après avoir cherché à évaluer la variabilité intrinsèque des différents paramètres (voir fig. 5 page 8), nous avons sélectionné les trois

## Dispersion des échantillons

Echantillons d'Hommes et de femmes en fonction de la course ou non, de leur taille et de leur consommation d'alcool

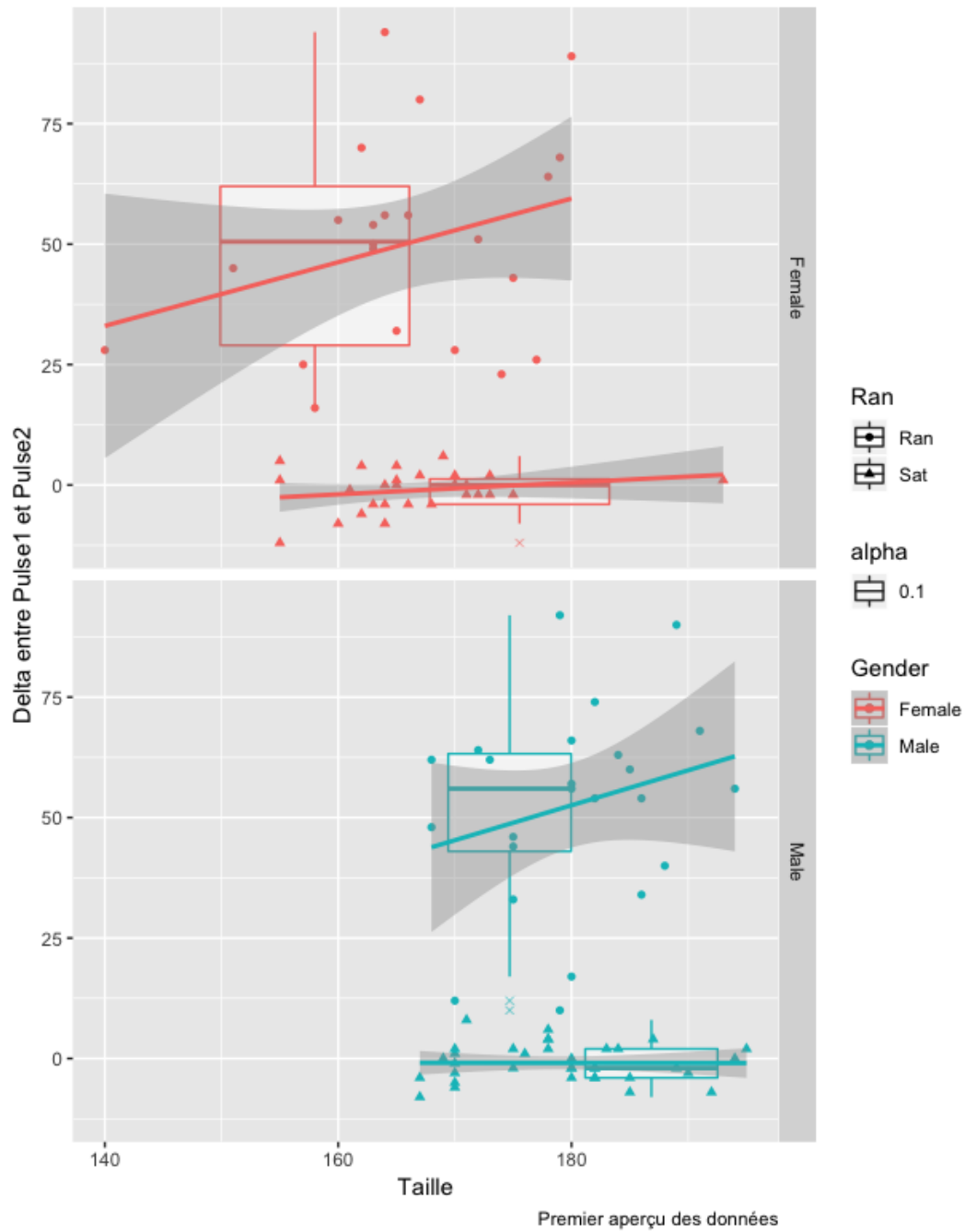


FIGURE 4: Dispersion de notre échantillon

meilleurs facteurs de variabilité après avoir successivement retiré les moins pertinents (voir fig. 6). Trois valeurs montrent une variabilité intrinsèque significative avec des p-value proche de 0.05 (Age) ou franchement inférieur à 0.05 : Smokes et Ran. Ran étant le facteur de variabilité incontestablement le plus fort comme en témoigne sa p-value à  $2.10^{-16}$  et son coefficient 10 fois supérieur à celui de Smokes. Par ailleurs, la taille

### Conclusion

On rejette  $H_0$  avec un risque  $\alpha = 5\%$ , il existe au moins une distribution dont la moyenne s'écarte des autres moyennes.  $pvalue_{Ran} = 2.10^{-16}$   $pvalue_{Smokes} = 2,5.10^{-2}$   $pvalue_{age} = 5,6.10^{-2}$

```
> breaks.aov <- aov(delta ~ Gender+Age+Smokes+Alcohol+Exercice+
+ Ran+Height*Weight*IMC)
> anova(breaks.aov)
Analysis of Variance Table

Response: delta
      Df Sum Sq Mean Sq F value Pr(>F)
Gender  1      9      9    0.0445 0.83335
Age      1    735    735    3.6891 0.05780 .
Smokes   1   1041   1041    5.2249 0.02451 *
Alcohol  1    459    459    2.3010 0.13265
Exercice 2    102    51    0.2568 0.77409
Ran      1  70789  70789  355.2119 < 2e-16 ***
Height   1    437    437    2.1952 0.14179
Weight   1     79     79    0.3961 0.53066
IMC      1   1029   1029    5.1610 0.02538 *
Height:Weight 1    162    162    0.8146 0.36907
Height:IMC  1     13     13    0.0664 0.79714
Weight:IMC  1    307    307    1.5413 0.21752
Height:Weight:IMC 1     52     52    0.2586 0.61229
Residuals 94  18733    199
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> coef(breaks.aov)
      (Intercept)      GenderMale      Age      SmokesYes
-7.371845e+02    -3.598712e+00    -5.291122e-02    -4.031906e-01
      AlcoholYes      ExerciceLow      ExerciceModerate      RanSat
-3.435030e+00    3.611400e-01    1.426139e-01    -5.278275e+01
      Height      Weight      IMC      Height:Weight
4.905612e+00    -4.287486e+00    1.773674e+01    -1.831361e-02
      Height:IMC      Weight:IMC      Height:Weight:IMC
1.120485e-05    -1.847290e-01    1.252194e-03
```

FIGURE 5: Etudes de facteurs de variabilité : Anova complet

```
> breaks.aov <- aov(delta ~ Age+Smokes+Ran)
> anova(breaks.aov)
Analysis of Variance Table

Response: delta
      Df Sum Sq Mean Sq F value Pr(>F)
Age      1      744      744    3.7334 0.05603 .
Smokes   1   1024   1024    5.1377 0.02546 *
Ran      1   71257  71257  357.6208 < 2e-16 ***
Residuals 105  20922    199
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> coef(breaks.aov)
      (Intercept)      Age      SmokesYes      RanSat
52.70434314    -0.05863287    -2.01570592    -52.22453574
```

FIGURE 6: Etudes de facteurs de variabilité : Anova ciblé

## Objectif 2

Quels paramètres influencent le pouls au repos ?

### Test utilisé et conditions

Tests de comparaisons Voir condition Diagramme page 5

### Hypothèses

$H_0 : \mu_1 = \mu_2$  : la moyenne des pouls 1 chez les fumeurs et la moyenne du pouls1 chez les non fumeurs sont égales

$H_1 \mu_1 \neq \mu_2$  : les moyennes sont différentes chez les fumeurs et chez les non fumeurs

### Calculs

Dans un premier temps on effectue un test qui suit un loi de Student à 109ddl en supposant la normalité de l'échantillon. Cependant, le test de fisher indique que l'hypothèse d'égalité des variance n'est pas vérifiée ( $pvalue = 0,78$  voir fig) On effectue alors un test de Welch. Voir page 14



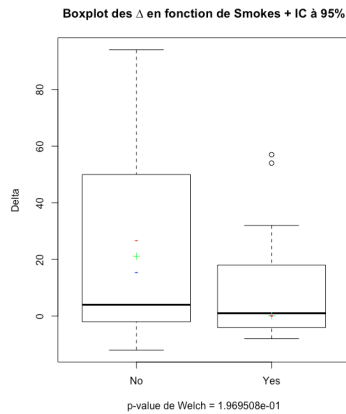


FIGURE 7: Boxplot des delta des fumeurs VS non fumeurs

## Conclusion

On n'a pas mis en évidence de différence significative entre la moyenne observée du pouls au repos chez les fumeurs et celle chez les non fumeurs : on conserve  $H_0$  sans connaître le risque de second espèce

Par ailleurs, la conclusion est identique en comparant les populations de consommateurs réguliers d'alcool  $p - value = 0.4629$ .

## Objectif 3

Trouver une population ayant introduit des biais d'auto-sélection.

### Test utilisé et conditions

Test du Khi deux pour la Comparaison de pulse1 avant et après 95 et Test de Fisher pour Comparaison de pulse1 avant et après 95 chez les fumeurs car conditions d'application du Khi2 non respecté (effectif calculé sous  $H_0 : A_{ij} \geq 5$ )

### Hypothèses

$H_0$  : les variables sont indépendantes

$H_1$  : Les variables sont liées

```
> delta_SmokesYes = filter(eval7, eval7$Smokes == "Yes")$Pulse2-
+ filter(eval7, eval7$Smokes == "Yes")$Pulse1
> delta_SmokesNo = filter(eval7, eval7$Smokes == "No")$Pulse2-
+ filter(eval7, eval7$Smokes == "No")$Pulse1
> sd(delta_SmokesYes)
[1] 24.29553
> sd(na.omit(delta_SmokesNo))
[1] 29.92463
> t.test(delta_SmokesYes, na.omit(delta_SmokesNo))
```

Welch Two Sample t-test

```
data: delta_SmokesYes and na.omit(delta_SmokesNo)
t = -1.3565, df = 13.655, p-value = 0.197
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -27.786438  6.287366
sample estimates:
mean of x mean of y
11.45455 22.20408
```

(a) Résultats Test de Welch (Mann-Whitney)

```
> var.test(delta_SmokesYes, delta_SmokesNo)
```

F test to compare two variances

```
data: delta_SmokesYes and delta_SmokesNo
F = 0.9332, num df = 66, denom df = 41, p-value = 0.7893
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5245769 1.6003353
sample estimates:
ratio of variances
0.9331991
```

(b) Test de Fisher d'égalité des variance

FIGURE 8: Protocole de test de comparaison

## Calculs

```
#Comparaison de la proportion d'Homme ayant couru avant et après 95
P1B <- filter(eval7, eval7$Year <= 95, Gender=="Male")[8]
P1BAS <- nrow(filter(P1B, Ran == "Sat"))
P1BAR <- nrow(filter(P1B, Ran == "Ran" ))

P1H <- filter(eval7, eval7$Year > 95, Gender=="Male")[8]
P1HAS <- nrow(filter(P1H, Ran == "Sat"))
P1HAR <- nrow(filter(P1H, Ran == "Ran" ))
#khi 2
prop.test(c(P1BAS,P1HAS),c(P1BAS+P1BAR,P1HAS+P1HAR))

#Comparaison de la proportion de fumeurs ayant couru avant et après 95
P1B <- filter(eval7, eval7$Year %in% c(93,95), eval7$Smokes=="Yes")[8]
P1BAS <- nrow(filter(P1B, Ran == "Sat"))
P1BAR <- nrow(filter(P1B, Ran == "Ran" ))

P1H <- filter(eval7, eval7$Year > 95, eval7$Smokes=="Yes" )[8]
P1HAS <- nrow(filter(P1H, Ran == "Sat"))
P1HAR <- nrow(filter(P1H, Ran == "Ran" ))

MP <- rbind(c(P1BAS,P1BAR),c(P1HAS,P1HAR))
# Fisher
fisher.test(MP)
```

FIGURE 9: Code des tests de comparaison de proportion

```
> prop.test(c(P1BAS,P1HAS),c(P1BAS+P1BAR,P1HAS+P1HAR))

2-sample test for equality of proportions with continuity correction

data:  c(P1BAS, P1HAS) out of c(P1BAS + P1BAR, P1HAS + P1HAR)
X-squared = 0.35867, df = 1, p-value = 0.5492
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1951133  0.4559829
sample estimates:
 prop 1  prop 2 
0.6521739 0.5217391
```

FIGURE 10: Résultat du test du Khi2

```
> MP <- rbind(c(P1BAS,P1BAR),c(P1HAS,P1HAR))
> # Fisher
> fisher.test(MP)

Fisher's Exact Test for Count Data

data:  MP
p-value = 0.5455
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1087056 234.7561625
sample estimates:
odds ratio 
2.97405
```

FIGURE 11: Résulta du Test exact

## Conclusion

Nous n'avons pas pu mettre en évidence une liaison significative

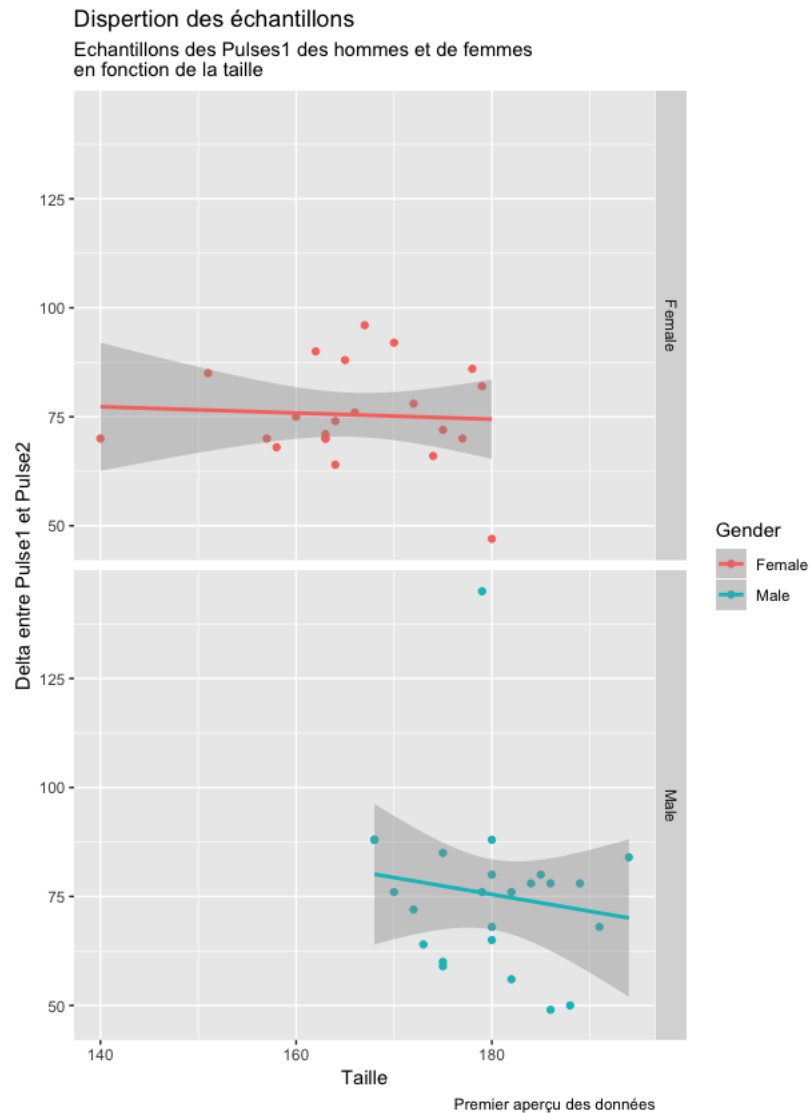
entre pulse1 et aucuns des sous-groupe de population avant et après le changement de protocole. Donc on conserve H0 sans connaitre le risque  $\beta$  de second espèce.

## Perspectives et Discussion

Malgré les inquiétudes du Docteur, il semblerait qu'aucun biais d'autoselection n'ai finalement été introduit. Nous avons travaillé sur un jeu de données peu qualitatif. Il n'était pas évident d'orienter une stratégie d'analyse là où peu de corrélation ou différence significative pouvaient émerger. Cependant, malgré ce que nous avait laissé croire le bon sens au vu des graphiques nous n'avons pas pu démontré la corrélation de la taille sur la fréquence cardiaque. Aussi bien sur le delta que sur le pouls au repos. Evidemment, nous avons pu démontrer de façon très significative que l'effort entraine une hausse de la fréquence cardiaque. Cette différence si significative aurait pu nous permettre de fixer un seuil de positivité au delà duquel nous pouvons affirmer à 100% que le sujet ait couru. Aussi, les analyse de corrélation pourrait nous permettre de prédire le delta d'un sujet en fonction des trois facteurs identifiés, à savoir Smokes, Age et Ran.



# Annexes



(a) Dispersion de Pulse1 des sujets ayant couru

FIGURE .12: Dispersion des échantillons de Pulse1 en fonction de la taille des participants et de leur genre

```
##### Objectif 1 #####
delta = eval7$delta
IMC = eval7$Weight/((eval7$Height/100)^2)
Gender <- eval7$Gender
Age <- eval7$Age
Smokes <- eval7$Smokes
Alcohol <- eval7$Alcohol
Exercise <- eval7$Exercise
Ran <- eval7$Ran
Height <- eval7$Height
Weight <- eval7$Weight
Pulse1 <- eval7$Pulse1
Pulse2 <- eval7$Pulse2

breaks.aov <- aov(delta ~ IMC+Gender+Age+Smokes+Alcohol+Exercise+Ran+Height+Weight)
anova(breaks.aov)
coef(breaks.aov)

#L'age, fumer et courir sont facteur de variabilité de delta
breaks.aov <- aov(delta ~ Age+Smokes+Ran+Height)
anova(breaks.aov)
coef(breaks.aov)

#Faire de l'exercice et avoir un IMC sont facteurs de variabilité de Pulse1
breaks.aov <- aov(Pulse1 ~ IMC+Exercise+Gender+Age+Smokes+Alcohol+Ran+Height+Weight)
anova(breaks.aov)
coef(breaks.aov)

breaks.aov <- aov(Pulse1 ~ Exercise)
anova(breaks.aov)
coef(breaks.aov)
```

FIGURE .13: Code Objectif 1

```
#### Objectif 2
delta_SmokesYes = filter(eval7, eval7$Alcohol == "Yes")$Pulse2-
  filter(eval7, eval7$Alcohol == "Yes")$Pulse1
delta_SmokesNo = filter(eval7, eval7$Alcohol == "No")$Pulse2-
  filter(eval7, eval7$Alcohol == "No")$Pulse1
sd(delta_SmokesYes)
sd(na.omit(delta_SmokesNo))
var.test(delta_SmokesYes,delta_SmokesNo)
t.test(delta_SmokesYes,na.omit(delta_SmokesNo),var.equal=T,alternative="two.sided")
t.test(delta_SmokesYes,na.omit(delta_SmokesNo),var.equal=F,alternative="two.sided")
wilcox.test(delta_SmokesYes,na.omit(delta_SmokesNo))
#On récupère la p value
pvalue <- t.test(delta_SmokesYes,na.omit(delta_SmokesNo),var.equal=F,alternative="two.sided")[3]

boxplot(delta ~ Smokes, main = paste("Boxplot des Δ en fonction de Smokes + IC à 95%"),
  xlab=paste("p-value de Welch =",format(as.numeric(pvalue),scientific=T)),
  ylab = "Delta")
smokesB = gsub("Yes",1,Smokes)
smokesB = gsub("No",0,smokesB)

a=t.test(delta, conf.level=0.95)
inta=round(a$estimate-a$conf.int[1],2)
b=t.test(as.numeric(smokesB), conf.level=0.95)
intb=round(b$estimate-b$conf.int[1],2)
points(1, a$estimate, col = "Green",pch = 3)
a$estimate
points(1, a$conf.int[1], col = "blue",pch = "-")
a$conf.int[1]
points(1, a$conf.int[2], col = "red",pch = "-")
a$conf.int[2]

points(2, b$estimate, col = "green",pch = 3)
b$estimate
points(2, b$conf.int[1], col = "blue",pch = "-")
b$conf.int[1]
points(2, b$conf.int[2], col = "red",pch = "-")
b$conf.int[2]
```

(a) Boxplot et code pour comparaison

```
> t.test(delta_SmokesYes,na.omit(delta_SmokesNo),var.equal=T,alternative="two.sided")

Two Sample t-test

data: delta_SmokesYes and na.omit(delta_SmokesNo)
t = 0.74332, df = 107, p-value = 0.4589
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.207322 15.854799
sample estimates:
mean of x mean of y
 22.77612  18.45238

> t.test(delta_SmokesYes,na.omit(delta_SmokesNo),var.equal=F,alternative="two.sided")

Welch Two Sample t-test

data: delta_SmokesYes and na.omit(delta_SmokesNo)
t = 0.7374, df = 84.942, p-value = 0.4629
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.334561 15.982038
sample estimates:
mean of x mean of y
 22.77612  18.45238
```

```

> ##### Objectif 3 #####
> #Comparaison de la proportion de sujet ayant couru avant et après 95
> P1B <- filter(eval7, eval7$Year <= 95)[8]
> P1BAS <- nrow(filter(P1B,Ran == "Sat"))
> P1BAR <- nrow(filter(P1B,Ran == "Ran" ))
> P1H <- filter(eval7, eval7$Year > 95)[8]
> P1HAS <- nrow(filter(P1H,Ran == "Sat"))
> P1HAR <- nrow(filter(P1H,Ran == "Ran" ))
> #khi 2
> prop.test(c(P1BAS,P1HAS),c(P1BAS+P1BAR,P1HAS+P1HAR))

      2-sample test for equality of proportions with continuity correction

data:  c(P1BAS, P1HAS) out of c(P1BAS + P1BAR, P1HAS + P1HAR)
X-squared = 1.0055, df = 1, p-value = 0.316
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.0885655  0.3157160
sample estimates:
   prop 1    prop 2 
0.6458333 0.5322581

> #Comparaison de la proportion de fumeurs ayant couru avant et après 95
> P1B <- filter(eval7, eval7$Year %in% c(93,95), eval7$Smokes=="Yes")[8]
> P1BAS <- nrow(filter(P1B,Ran == "Sat"))
> P1BAR <- nrow(filter(P1B,Ran == "Ran" ))
> P1H <- filter(eval7, eval7$Year > 95 ,eval7$Smokes=="Yes" )[8]
> P1HAS <- nrow(filter(P1H,Ran == "Sat"))
> P1HAR <- nrow(filter(P1H,Ran == "Ran" ))
> MP <- rbind(c(P1BAS,P1BAR),c(P1HAS,P1HAR))
> # Fisher
> fisher.test(MP)

```

#### Fisher's Exact Test for Count Data

```

data:  MP
p-value = 0.5455
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  0.1087056 234.7561625
sample estimates:
odds ratio
  2.97405

```

FIGURE .15: Code Objectif 3

# Références

- [1] Bauer, D. F., 1972. Constructing confidence sets using rank statistics. Journal of the American Statistical Association 67 (339), 687–690.  
URL <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1972.10481279> :10.1080/01621459.1972.10481279
- [2] Hollander, M., A. Wolfe, D., Chicken, E., Hollander, M., A. Wolfe, D., Chicken, E., 2015. The One-Sample Location Problem. In : Nonparametric Statistical Methods. John Wiley & Sons, Inc., pp. 39–114, dOI : 10.1002/9781119196037.ch3.  
URL <http://onlinelibrary.wiley.com/doi/10.1002/9781119196037.ch3/summary> 5
- [3] Hollander, M., A. Wolfe, D., Chicken, E., Hollander, M., A. Wolfe, D., Chicken, E., 2015. The Two-Sample Location Problem. In : Nonparametric Statistical Methods. John Wiley & Sons, Inc., pp. 115–150, dOI : 10.1002/9781119196037.ch4.  
URL <http://onlinelibrary.wiley.com/doi/10.1002/9781119196037.ch4/summary> 5
- [4] Kim, T. K., Dec. 2015. T test as a parametric statistic. Korean Journal of Anesthesiology 68 (6), 540–546.  
URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667138/> 2, 5
- [5] t.test(), open source. R : Student’s t-Test.  
URL <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/t.test.html> 5
- [6] wilcox.test(), open source. R : Wilcoxon Rank Sum and Signed Rank Tests.  
URL <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html> 5