

Premix

The dream team

2024-02-11

Contents

1	Introduction	1
2	Chargement des données, des packages R et des fonctions utiles	1
3	Visualisation des trajectoires pour le marqueur CD57	2
3.1	Analyse de la variance	3
3.2	Test de Kruskal-Wallis	5
3.3	Anova à mesures répétées	5
3.4	Test de Friedman	6
4	Conclusion	7

1 Introduction

Dans un premier temps, on cherche à visualiser l'évolution des marqueurs biologiques au cours du temps. On cherchera ensuite à voir si le vieillissement des réactifs premix impacte les mesures.

2 Chargement des données, des packages R et des fonctions utiles

```
#Packages
library(readxl)
library(ggplot2)
library(reshape2)
library(nlme)
library(tidyverse)
library(ggpubr)
library(kableExtra)
library(rstatix)
```

```
#####
# Import Datasets #
#####
A = as.data.frame(read_excel("stab_premix_at.xlsx", sheet = 9, col_names = T))
rownames(A) = A[, 1]
A = A[, -1]
```

3 Visualisation des trajectoires pour le marqueur CD57

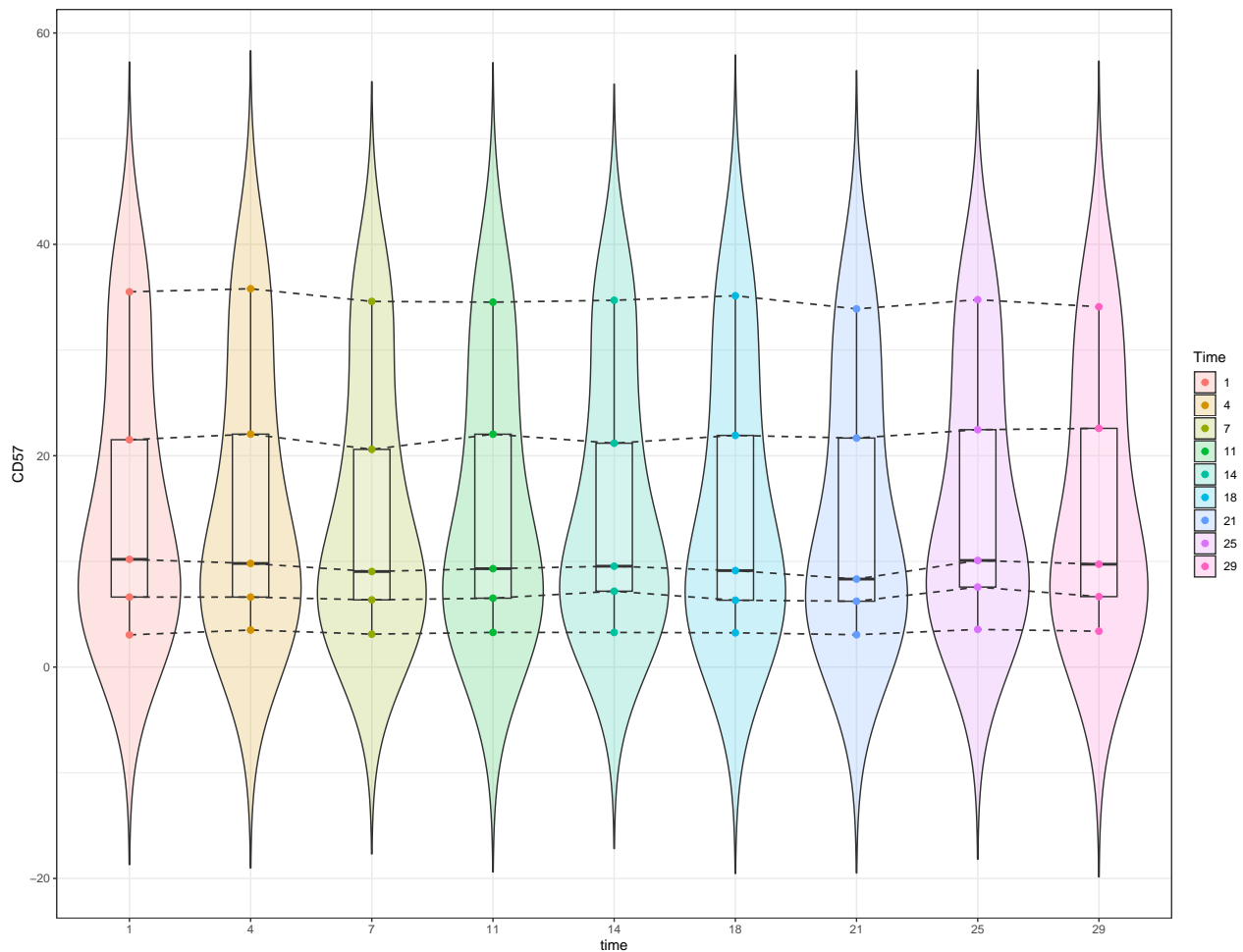
Dans un premier temps, on va visualiser le décours temporel de la variable CD57 pour chacun des patients. Pour ce faire on commence par mettre en forme les données

```
B = A[grep("CD57", rownames(A)), ]
df = data.frame(melt(B), as.factor(rep(1:5, 9)))
```

```
## No id variables; using all as measure variables
```

```
colnames(df) = c("Time", "CD57", "id")
```

```
p <- ggplot(data = df, aes(x = Time, y = CD57)) +
  geom_violin(trim = FALSE, alpha = 0.2, aes(fill = Time)) +
  geom_boxplot(alpha = 0.2, show.legend = FALSE, width = 0.3) +
  geom_line(linetype = "dashed", linewidth = .6, aes(group = id),
            color = "gray20") +
  geom_point(aes(colour = Time), alpha = 1, size = 2) +
  xlab("time") + ylab("CD57") + theme_bw()
p
```



On constate une stabilité du réactif CD57 au cours du temps qu'il s'agit maintenant de tester à l'aide de méthodes statistiques appropriées. Nous allons ici tester quelques méthodes.

3.1 Analyse de la variance

On réalise une analyse de la variance en utilisant la fonction `aov()`. L'objectif de l'analyse de la variance est double : (i) Tester si la moyenne de CD57 au temps 1 est significativement différente de la moyenne de CD57 aux temps suivants. (ii) Tester si globalement le temps a un effet sur le niveau moyen de CD57

Par curiosité, on calcule la moyenne de CD57 aux différents temps.

```
tapply(df$CD57, df$Time, mean)
```

```
##      1      4      7     11     14     18     21     25     29
## 15.3792 15.5550 14.7440 15.1358 15.1790 15.1522 14.6362 15.6898 15.2954
```

Les résultats de l'analyse de la variance permettant de tester globalement, le lien entre le facteur temps et CD57. Les résultats sont reportés ci-dessous.

```
aov_out = aov(CD57~Time, data = df)
summary(aov_out)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Time       8      5      0.59   0.003    1
## Residuals  36    6076   168.77
```

Comme attendu, on ne rejette pas l'hypothèse nulle puisque la p-value est à égale à 1.

On peut malgré tout s'intéresser aux différences pairwise.

```
tukey_out = TukeyHSD(aov_out)
print(tukey_out)
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = CD57 ~ Time, data = df)
##
## $Time
##           diff           lwr           upr p adj
## 4-1      0.1758 -26.91404 27.26564    1
## 7-1     -0.6352 -27.72504 26.45464    1
## 11-1    -0.2434 -27.33324 26.84644    1
## 14-1    -0.2002 -27.29004 26.88964    1
## 18-1    -0.2270 -27.31684 26.86284    1
## 21-1    -0.7430 -27.83284 26.34684    1
## 25-1     0.3106 -26.77924 27.40044    1
## 29-1    -0.0838 -27.17364 27.00604    1
## 7-4     -0.8110 -27.90084 26.27884    1
## 11-4    -0.4192 -27.50904 26.67064    1
## 14-4    -0.3760 -27.46584 26.71384    1
## 18-4    -0.4028 -27.49264 26.68704    1
## 21-4    -0.9188 -28.00864 26.17104    1
## 25-4     0.1348 -26.95504 27.22464    1
## 29-4    -0.2596 -27.34944 26.83024    1
## 11-7     0.3918 -26.69804 27.48164    1
## 14-7     0.4350 -26.65484 27.52484    1
## 18-7     0.4082 -26.68164 27.49804    1
## 21-7    -0.1078 -27.19764 26.98204    1
## 25-7     0.9458 -26.14404 28.03564    1
## 29-7     0.5514 -26.53844 27.64124    1
## 14-11    0.0432 -27.04664 27.13304    1
## 18-11    0.0164 -27.07344 27.10624    1
## 21-11   -0.4996 -27.58944 26.59024    1
## 25-11    0.5540 -26.53584 27.64384    1
## 29-11    0.1596 -26.93024 27.24944    1
## 18-14   -0.0268 -27.11664 27.06304    1
## 21-14   -0.5428 -27.63264 26.54704    1
## 25-14    0.5108 -26.57904 27.60064    1
## 29-14    0.1164 -26.97344 27.20624    1
## 21-18   -0.5160 -27.60584 26.57384    1
## 25-18    0.5376 -26.55224 27.62744    1
## 29-18    0.1432 -26.94664 27.23304    1
## 25-21    1.0536 -26.03624 28.14344    1
## 29-21    0.6592 -26.43064 27.74904    1
## 29-25   -0.3944 -27.48424 26.69544    1
```

On constate qu'aucune différence de moyenne de CD57 entre t_1 et $t_j, j = \{4, \dots, 29\}$ n'est significative.

3.2 Test de Kruskal-Wallis

On pourrait nous reprocher de ne pas avoir vérifié les conditions d'utilisation de l'analyse de la variance. Son pendant non paramétrique, le test de Kruskal-Wallis, est ici mise en oeuvre.

```
kruskal_out = kruskal.test(CD57~Time, data = df)
print(kruskal_out)

##
##  Kruskal-Wallis rank sum test
##
## data:  CD57 by Time
## Kruskal-Wallis chi-squared = 0.93449, df = 8, p-value = 0.9986
```

Le test de Kruskal-Wallis conduit à la même conclusion, à savoir le non rejet de l'hypothèse nulle ($p_value = 0.9986$).

3.3 Anova à mesures répétées

Enfin, pour tenir compte du caractère longitudinal des données (mesures répétées), on réalise une analyse de la variance à mesures répétées.

```
aov_out_repeated = aov(CD57 ~ Time + Error(id), data = df)
summary(aov_out_repeated)

##
## Error: id
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  4   6070     1517
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Time       8   4.698   0.5873   3.144 0.0097 **
## Residuals 32   5.978   0.1868
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comme l'atteste la p-value, l'anova à mesures répétées ne conduit pas aux mêmes conclusions. Examinons donc plus finement les résultats en réalisant des comparaisons pairwise.

```
# pairwise comparisons using rstatix
pwc <- df %>%
  pairwise_t_test(
    CD57 ~ Time, paired = TRUE,
    p.adjust.method = "bonferroni"
  )
kable(as.data.frame(pwc))
```

.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
CD57	1	4	5	5	-1.0429325	4	0.356	1.000	ns
CD57	1	7	5	5	2.7695341	4	0.050	1.000	ns
CD57	1	11	5	5	0.8149687	4	0.461	1.000	ns
CD57	1	14	5	5	0.7747354	4	0.482	1.000	ns
CD57	1	18	5	5	0.8862820	4	0.426	1.000	ns
CD57	1	21	5	5	1.7675492	4	0.152	1.000	ns
CD57	1	25	5	5	-0.9483951	4	0.397	1.000	ns
CD57	1	29	5	5	0.2013738	4	0.850	1.000	ns
CD57	4	7	5	5	3.5676780	4	0.023	0.842	ns
CD57	4	11	5	5	1.8397977	4	0.140	1.000	ns
CD57	4	14	5	5	1.3081169	4	0.261	1.000	ns
CD57	4	18	5	5	3.5804741	4	0.023	0.835	ns
CD57	4	21	5	5	2.8387266	4	0.047	1.000	ns
CD57	4	25	5	5	-0.4117083	4	0.702	1.000	ns
CD57	4	29	5	5	0.6822941	4	0.533	1.000	ns
CD57	7	11	5	5	-1.4655417	4	0.217	1.000	ns
CD57	7	14	5	5	-3.2622778	4	0.031	1.000	ns
CD57	7	18	5	5	-1.6432136	4	0.176	1.000	ns
CD57	7	21	5	5	0.3309634	4	0.757	1.000	ns
CD57	7	25	5	5	-3.1664968	4	0.034	1.000	ns
CD57	7	29	5	5	-1.3485010	4	0.249	1.000	ns
CD57	11	14	5	5	-0.1732941	4	0.871	1.000	ns
CD57	11	18	5	5	-0.1089638	4	0.918	1.000	ns
CD57	11	21	5	5	3.5460585	4	0.024	0.860	ns
CD57	11	25	5	5	-3.5388201	4	0.024	0.864	ns
CD57	11	29	5	5	-0.9368375	4	0.402	1.000	ns
CD57	14	18	5	5	0.0938281	4	0.930	1.000	ns
CD57	14	21	5	5	1.7796019	4	0.150	1.000	ns
CD57	14	25	5	5	-2.4550362	4	0.070	1.000	ns
CD57	14	29	5	5	-0.3228799	4	0.763	1.000	ns
CD57	18	21	5	5	2.3380656	4	0.080	1.000	ns
CD57	18	25	5	5	-1.9265568	4	0.126	1.000	ns
CD57	18	29	5	5	-0.4602930	4	0.669	1.000	ns
CD57	21	25	5	5	-4.7396961	4	0.009	0.325	ns
CD57	21	29	5	5	-2.9571529	4	0.042	1.000	ns
CD57	25	29	5	5	2.1730420	4	0.096	1.000	ns

Ces différences globales détectées par l'anova à mesures répétées ne résistent pas aux comparaisons pairwise.

3.4 Test de Friedman

Enfin, pour le plaisir, on peut mettre en oeuvre le test de Friedman (vu comme une version non paramétrique de l'anova à mesures répétées).

```
friedman.test(as.matrix(B))
```

```
##
## Friedman rank sum test
##
## data:  as.matrix(B)
## Friedman chi-squared = 21.493, df = 8, p-value = 0.005946
```

Le test de Friedman renvoie les mêmes conclusions que l'anova à mesures répétées... mais ne passe toujours pas l'épreuve des comparaisons multiples comme l'atteste le tableau ci-dessous.

```
# pairwise comparisons using rstatix
pwc <- df %>%wilcox_test(
  CD57 ~ Time, paired = TRUE,
  p.adjust.method = "bonferroni"
)
kable(as.data.frame(pwc))
```

.y.	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
CD57	1	4	5	5	4	0.438	1	ns
CD57	1	7	5	5	14	0.125	1	ns
CD57	1	11	5	5	10	0.625	1	ns
CD57	1	14	5	5	11	0.438	1	ns
CD57	1	18	5	5	10	0.625	1	ns
CD57	1	21	5	5	12	0.312	1	ns
CD57	1	25	5	5	4	0.438	1	ns
CD57	1	29	5	5	8	1.000	1	ns
CD57	4	7	5	5	15	0.062	1	ns
CD57	4	11	5	5	15	0.062	1	ns
CD57	4	14	5	5	12	0.312	1	ns
CD57	4	18	5	5	15	0.062	1	ns
CD57	4	21	5	5	15	0.062	1	ns
CD57	4	25	5	5	5	0.625	1	ns
CD57	4	29	5	5	10	0.625	1	ns
CD57	7	11	5	5	1	0.125	1	ns
CD57	7	14	5	5	0	0.062	1	ns
CD57	7	18	5	5	1	0.125	1	ns
CD57	7	21	5	5	10	0.625	1	ns
CD57	7	25	5	5	0	0.062	1	ns
CD57	7	29	5	5	3	0.312	1	ns
CD57	11	14	5	5	6	0.812	1	ns
CD57	11	18	5	5	10	0.625	1	ns
CD57	11	21	5	5	15	0.062	1	ns
CD57	11	25	5	5	0	0.062	1	ns
CD57	11	29	5	5	4	0.438	1	ns
CD57	14	18	5	5	8	1.000	1	ns
CD57	14	21	5	5	13	0.188	1	ns
CD57	14	25	5	5	0	0.062	1	ns
CD57	14	29	5	5	7	1.000	1	ns
CD57	18	21	5	5	15	0.062	1	ns
CD57	18	25	5	5	2	0.188	1	ns
CD57	18	29	5	5	5	0.625	1	ns
CD57	21	25	5	5	0	0.062	1	ns
CD57	21	29	5	5	0	0.062	1	ns
CD57	25	29	5	5	14	0.125	1	ns

4 Conclusion

Les différentes approches utilisées conduisent à des résultats consistants.

Naturellement, on peut réaliser des analyses similaires sur les autres variables (“CD2”, “CD7”, “CD56”, “CD19”, “CD4_”, “CD3”, “CD5_”, “Ly”, “TRBC”, “CD8”, “HLADR”) qui devraient conduire toutes aux mêmes résultats. A vous de vérifier...