# Première partie

On représente l’ensemble des données dans un tableau et on calcule le rang de chaque observation.

8 12 3 9 11 7 10 1 4 5 2

variable "B" "B" "A" "B" "B" "B" "B" "A" "A" "A" "A"

value "2" "3" "4" "5" "5" "7" "7" "8" "8" "8" "9"

rank " 1.0" " 2.0" " 3.0" " 4.5" " 4.5" " 6.5" " 6.5" " 9.0" " 9.0" " 9.0" "11.0"

La fonction rank() de R permet de calculer les rangs des observations.

On peut donc ensuite calculer les coefficients W. Pour A, la somme des rangs des 5 valeurs vaut 41 :

Pour B la somme des rangs des 6 valeurs vaut 25 :

On observe donc une différence entre le coefficient pour A et pour B, indiquant une différence entre les distributions A et B.

On peut vérifier ces résultats avec la fonction wilcox.test() de R :

> wilcox.test(A,B)

Wilcoxon rank sum test with continuity correction

data: A and B

W = 26, p-value = 0.05193

alternative hypothesis: true location shift is not equal to 0

> wilcox.test(B,A)

Wilcoxon rank sum test with continuity correction

data: B and A

W = 4, p-value = 0.05193

alternative hypothesis: true location shift is not equal to 0

Mais avec une valeur p supérieure à 0.05, nous ne sommes pas en mesure de rejeter l’hypothèse nulle. Nous ne pouvons donc pas conclure à une différence significative entre les deux distributions.

# Seconde partie

Les variables sont mesurées sur les mêmes individus, on peut donc les apparier.

Pour faire un test de wilcox apparié, on lance la commande suivante :

> wilcox.test(alldata$Var4, alldata$Var6, paired=T)

Wilcoxon signed rank test with continuity correction

data: alldata$Var4 and alldata$Var6

V = 983.5, p-value = 0.9637

alternative hypothesis: true location shift is not equal to 0

La valeur p étant supérieure à 0.05, nous ne pouvons déclarer que les distributions des variables Var4 et Var6 sont différentes.

Le résultat pour les variables Var3 et Var 5 est le suivant :

> wilcox.test(alldata$Var3, alldata$Var5, paired=T)

Wilcoxon signed rank test with continuity correction

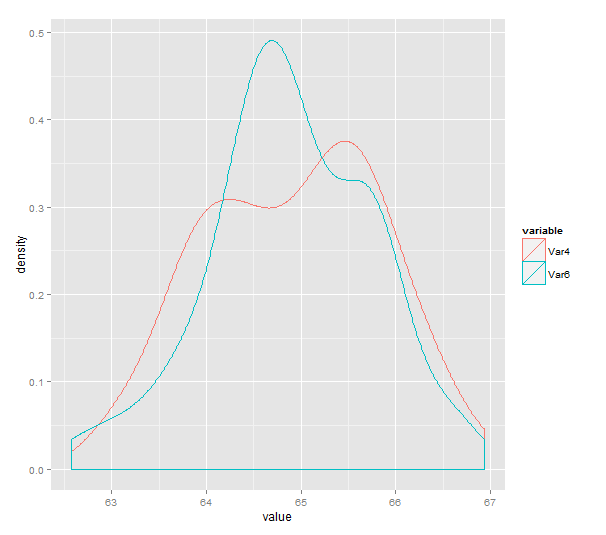
data: alldata$Var3 and alldata$Var5

V = 2080, p-value = 3.608e-12

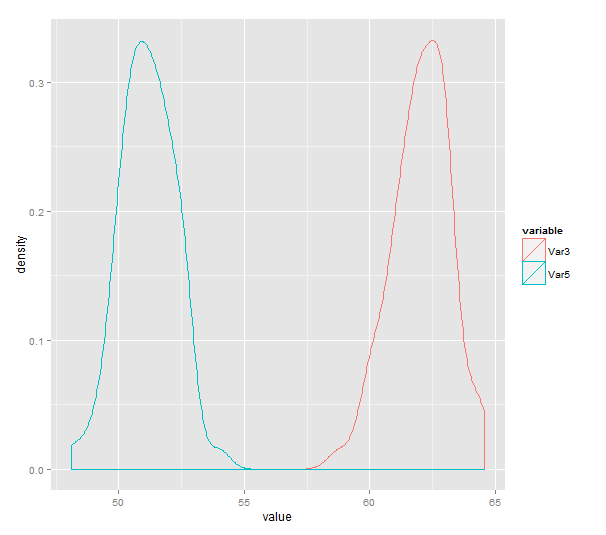
alternative hypothesis: true location shift is not equal to 0

Ici, nous avons une valeur p inférieure à 0.05, nous pouvons donc rejeter l’hypothèse nulle et conclure à une différence significative entre les distributions des deux variables.

Ci-dessous les distributions des variables Var4 et Var6 :



Ci-dessous les distributions des variables Var3 et Var5 :



Ces deux graphiques corroborent les calculs précédents.