

Les tests statistiques en pratique : comparaison de deux moyennes

- Comparer deux moyennes avec un test t de Student

- Comparer deux moyennes avec un test t de Student
- Conditions de validité :
 - $n > 30$ ou distribution normale, en fait ces deux conditions sont à considérer ensemble
 - Variances égales (sinon approximation de Welch)

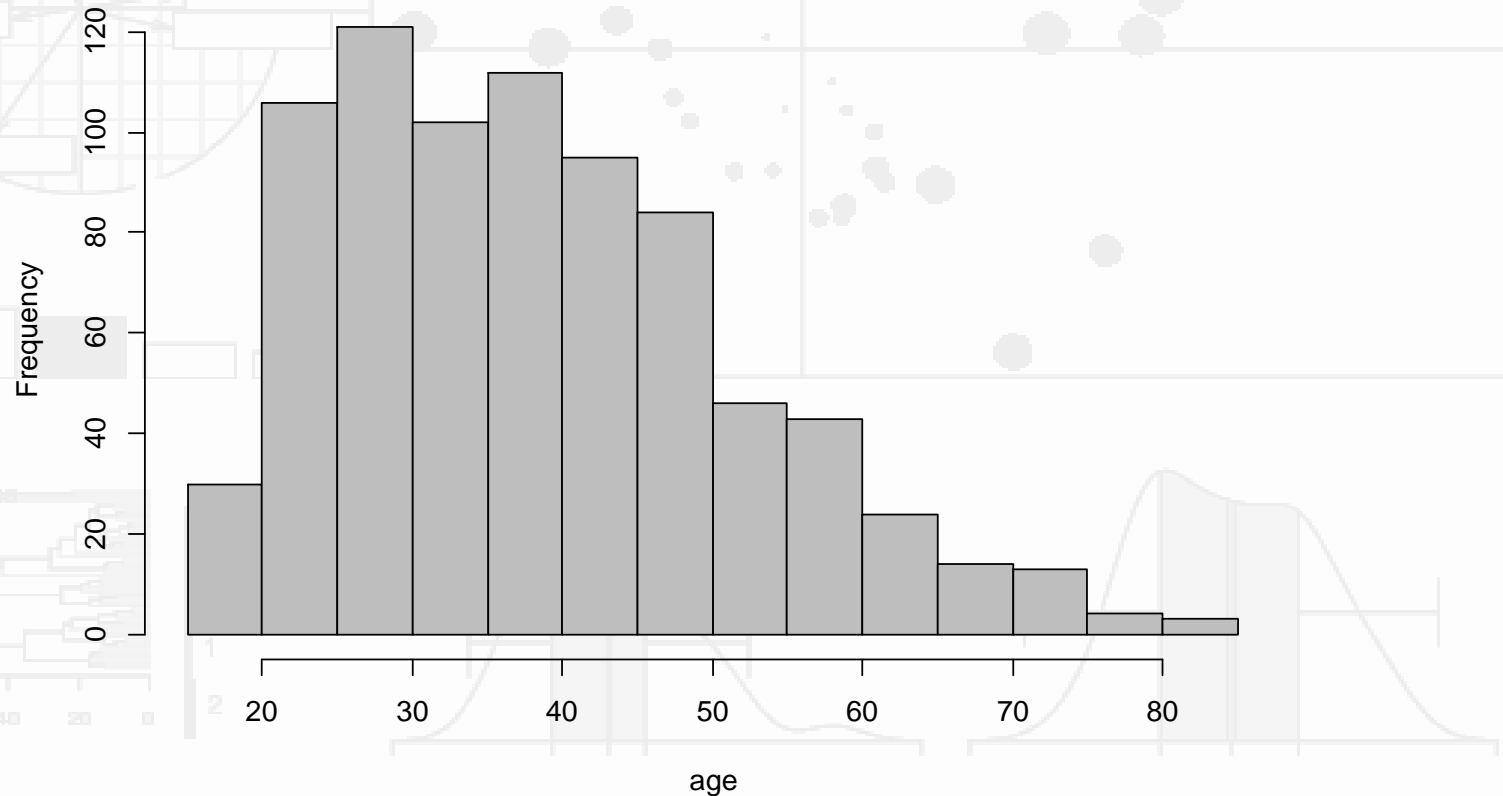
Existe-t-il, en moyenne, une différence significative d'âge entre les détenus ayant un haut niveau d'évitement du danger et les autres ?

Test t de Student : normalité de la variable

Introduction à la statistique avec R > Comparaison de 2 moyennes



```
> hist(smp.c$age)
```



- L'inspection visuelle de l'histogramme est informative, mais pas très rigoureuse

- L'inspection visuelle de l'histogramme est informative, mais pas très rigoureuse
- Tester la normalité de la variable ?
 - Ces tests existent
 - Mais ils ne répondent pas à la question !
 - Les tests de normalité sont fait pour affirmer qu'une distribution n'est pas normale, et pas le contraire...

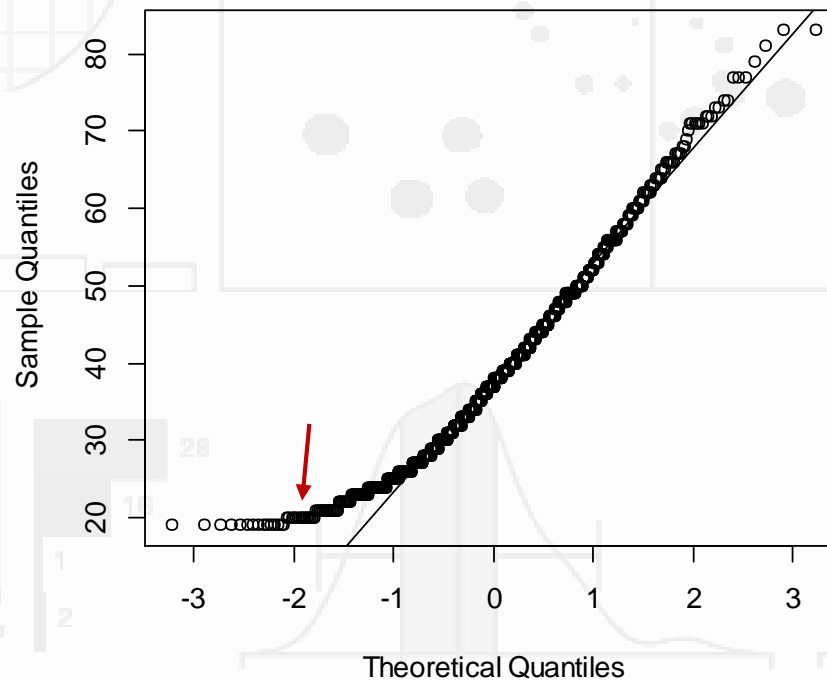
Test t de Student : normalité de la variable

Introduction à la statistique avec R > Comparaison de 2 moyennes



```
> qqnorm(smp.c$age); qqline(smp.c$age)
```

Normal Q-Q Plot



Test t de Student : égalité des variances

Introduction à la statistique avec R > Comparaison de 2 moyennes



```
> by(smp.c$age, smp.c$ed.b, sd, na.rm=TRUE)
smp.c$ed.b: 0
[1] 13.38593
-----
smp.c$ed.b: 1
[1] 13.29636
```

```
> t.test(smp.c$age~smp.c$ed.b, var.equal=TRUE)
```

Two Sample t-test

```
data: smp.c$age by smp.c$ed.b
```

```
t = 1.7142, df = 690, p-value = 0.08694
```

```
alternative hypothesis: true difference in means is not equal  
to 0
```

```
95 percent confidence interval:
```

```
-0.2710524  4.0005138
```

```
sample estimates:
```

```
mean in group 0 mean in group 1
```

```
39.46383
```

```
37.59910
```

```
> t.test(smp.c$age~smp.c$ed.b, var.equal=TRUE)
```

Two Sample t-test

```
data: smp.c$age by smp.c$ed.b
```

```
t = 1.7142, df = 690, p-value = 0.08694
```

```
alternative hypothesis: true difference in means is not equal  
to 0
```

```
95 percent confidence interval:
```

```
-0.2710524  4.0005138
```

```
sample estimates:
```

```
mean in group 0 mean in group 1
```

```
39.46383
```

```
37.59910
```

```
> t.test(smp.c$age~smp.c$ed.b, var.equal=TRUE)
```

Two Sample t-test

```
data: smp.c$age by smp.c$ed.b
```

```
t = 1.7142, df = 690, p-value = 0.08694
```

```
alternative hypothesis: true difference in means is not equal  
to 0
```

```
95 percent confidence interval:
```

```
-0.2710524  4.0005138
```

```
sample estimates:
```

```
mean in group 0 mean in group 1
```

```
39.46383      37.59910
```

Le test de Mann-Whitney (Wilcoxon)

Introduction à la statistique avec R > Comparaison de 2 moyennes



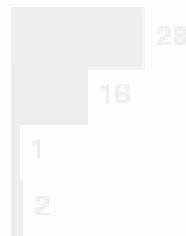
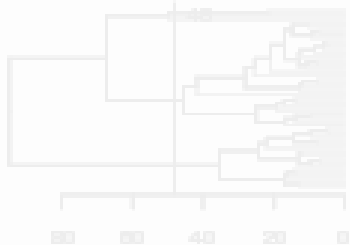
```
> wilcox.test(smp.c$age~smp.c$ed.b)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: smp.c$age by smp.c$ed.b
```

```
W = 56769.5, p-value = 0.06091
```

```
alternative hypothesis: true location shift is not equal to 0
```



Conclusion

Introduction à la statistique avec R > Comparaison de 2 moyennes



```
qqnorm(smp.c$age); qqline(smp.c$age)
by(smp.c$age, smp.c$ed.b, sd, na.rm=TRUE)
t.test(smp.c$age~smp.c$ed.b, var.equal=TRUE)
wilcox.test(smp.c$age~smp.c$ed.b)
```