

# Compte rendu TP3 - Tests Statistiques

Laroussi Labid Bachri, M1 BBS

15-09-2025

## Contents

<b>Tests de Chi2</b>	<b>1</b>
Tests de Chi2 de conformité . . . . .	1
Tests de Chi2 d'indépendance . . . . .	2
<b>Tests paramétriques d'adéquation et tests d'homogénéité (sur une variable quantitative)</b>	<b>3</b>
<b>Test d'adéquation à la loi normale (<math>H_0</math> = "les données suivent la loi Normale")</b>	<b>4</b>
<b>Test non paramétriques d'homogénéité (sur une variable quantitative)</b>	<b>6</b>
Test d'adéquation à la loi normale . . . . .	6

## Tests de Chi2

### Tests de Chi2 de conformité

```
freq_exp = c(1/4,1/2,1/4) # fréquences relatives théoriques F2 $
freq_obs = c(22,53,25) # fréquences observées écrits sous forme réelle.
chisq.test(freq_obs, p=freq_exp) ## Valeur de la stat de Chi2 , degré de liberté est la p-value.

##
## Chi-squared test for given probabilities
##
## data:  freq_obs
## X-squared = 0.54, df = 2, p-value = 0.7634
```

Question : les proportions observées sont-elles conformes aux proportions mendéliennes?

$H_0$  les données observées sont conformes aux données théoriques. - L'hypothèse nulle de ce test est que les fréquences observées et les fréquences théoriques sont différentes. - La valeur de p-value est de 0.7634 qui est bien supérieure a seuil alpha 0.05 donc on ne rejette pas  $H_0$ . P value: probabilité de rejeter  $H_0$  a tort. Risque alpha.

**Conclusion : Les frequences observées sont conformes a la Loi de Mendel ( fréquences attendues)**

## Tests de Chi2 d'indépendance

Notre exemple : la résistance ou sensibilité à un pathogène est-elle indépendante de l'écotype d'*Arabidopsis thaliana* ? Autrement dit y a-t-il indépendance entre phénotype et écotype ?

```
#Tableau de contingence
```

```
mat = matrix( c( 15,5,3,0,19,16), nrow=2 ,ncol=3, byrow = T, dimnames = list(c("R","S"), c("Col","Ws","Can"), mat
```

```
##      Col Ws Can
## R    15  5   3
## S     0 19  16
```

```
chisq.test(mat, correct = F) # Correct: correction des effectifs lorsqu'ils sont trop faibles ( pas imp
```

```
##
## Pearson's Chi-squared test
##
## data:  mat
## X-squared = 30.901, df = 2, p-value = 1.949e-07
```

On fait un premier calcul d'effectif théorique.

```
# (15 + 0) x 23 / total
```

```
effectif_théorique_COL_R = 15*23/58
round(effectif_théorique_COL_R)
```

```
## [1] 6
```

Ensemble de R et Col plus somme des R divisé par total. H0: la résistance à un pathogène est indépendante de l'écotype H1 : La résistance dépend de l'écotype.

X = 30.901 très loin de 0 donc p value très petite. On rejette H0 et accepte hypothèse alternative. Il y'a une dépendance entre les variables.

Le degré de liberté, 3 genotypes avec des fréquences. On enlève un car si on est sûr des fréquences la somme est de 100%. Il suffit d'en connaître 2 pour déduire l'autre

S'il y a une relation entre phénotype et écotype, comment se traduit-elle ?

```
obs=chisq.test(mat, correct =F)$observed
exp=chisq.test(mat, correct =F)$expected
obs-exp
```

```
##           Col           Ws           Can
## R  9.051724 -4.517241 -4.534483
## S -9.051724  4.517241  4.534483
```

On compare nos tableaux des effectifs observés à nos effectifs théoriques grâce au \$.

Si différence positive : excès d'effectifs, Col résiste, Ws S et Can s. si négatif : déficit d'effectifs

**Conclusion :** La résistance est dépendante de l'écotype mais particulièrement pour l'écotype Col.

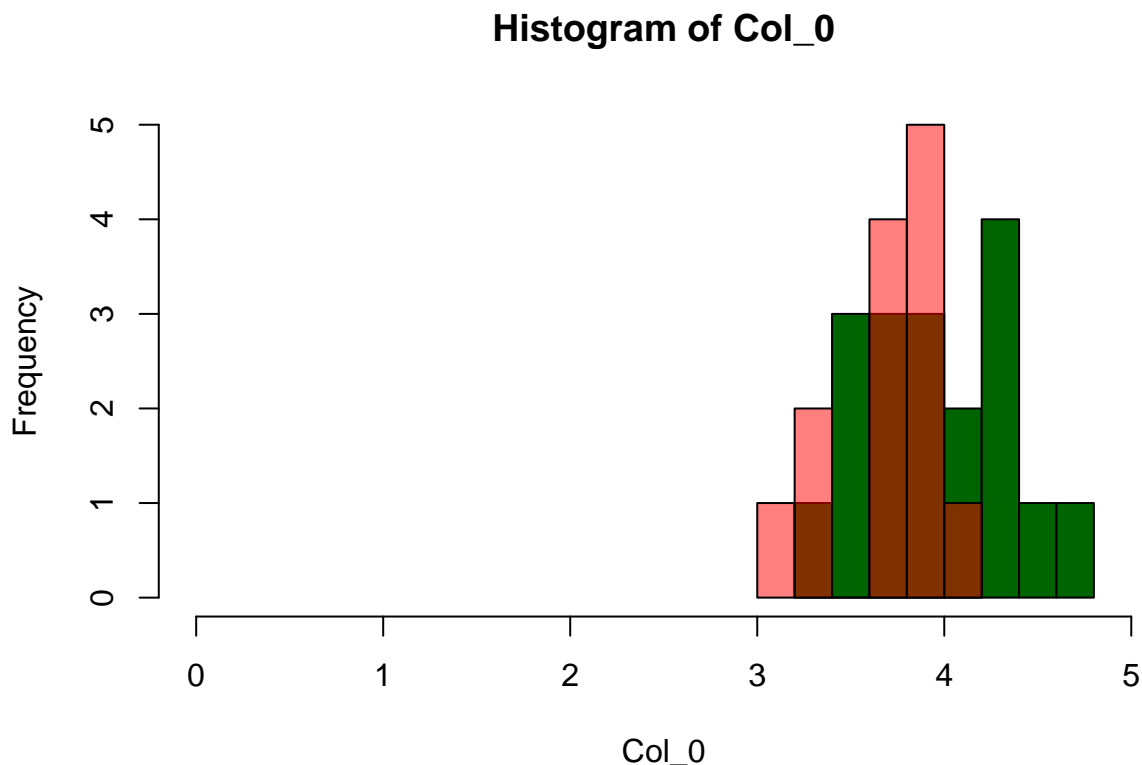
## Tests paramétriques d'adéquation et tests d'homogénéité (sur une variable quantitative)

Dans cet exercice, nous allons chercher à comparer les valeurs d'une variable quantitative mesurée dans 2 échantillons 1 et 2, à l'aide d'un test de Student de comparaison de moyennes de 2 échantillons. La question posée est : la taille moyenne (à un temps de donné) de plantules d' *A. thaliana* de génotype "sauvage" (Col\_0) et d'un mutant pour le gène X1 (Mut\_X1) est-elle la même (hypothèse H0) ?

```
# Vecteurs des tailles des plantes
```

```
Col_0=c(4.30,4.25,3.50,3.35,4.30,3.75,3.55,4.10,3.95,4.55,4.25,3.75,3.85,4.15,3.55,4.75,3.95,3.65)
```

```
Mut_X1=c(3.06,4.05,3.95,3.40,3.80,3.95,3.65,4,3.85,3.95,3.65,3.75,3.4)
```



```
## La taille moyenne des plantes wild est de 3.972222 et la taille des plantes mutants est de 3.727692
```

On peut les représenter par des histogrammes.

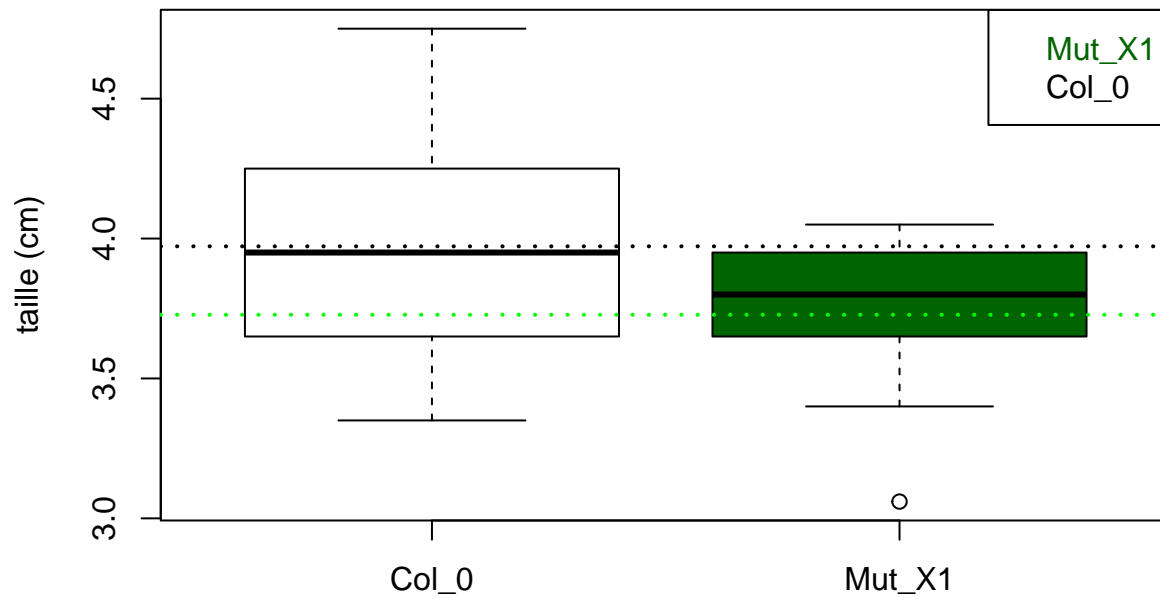
```
# Représentation graphique des données
```

```
boxplot(Col_0,Mut_X1,names=c("Col_0","Mut_X1"),col=c("white","darkgreen"),ylab="taille (cm)")
```

```
abline(h=mean(Col_0),col="black",lty=3,lwd=2)
```

```
abline(h=mean(Mut_X1),col="green",lty=3,lwd=2)
```

```
legend("topright", legend = c("Mut_X1","Col_0"), text.col = c("darkgreen","black"))
```



Mais aussi avec des boxplots. On observe une plus grande dispersion chez les groupe Col\_0

**Test d'adéquation à la loi normale ( $H_0$  = “les données suivent la loi Normale”)**

```
shapiro.test(Col_0)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Col_0
## W = 0.9701, p-value = 0.7996
```

```
shapiro.test(Mut_X1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Mut_X1
## W = 0.88924, p-value = 0.09525
```

```
var.test(Col_0, Mut_X1)
```

```
##
##  F test to compare two variances
##
## data:  Col_0 and Mut_X1
## F = 1.7592, num df = 17, denom df = 12, p-value = 0.3234
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```
## 0.5622863 4.9695146
## sample estimates:
## ratio of variances
## 1.759191
```

Les 2 series suivent une loi normal car les 2 p value sont superieures à 0.05 donc on accepte H0.

Rapport de variance superieur à 1. Donc une des 2 varainces est 2 fois plus forte que l'autre.

rmq : cela dépend de l'échantillon et du degré de liberté. On a une p-value de 0.3234 donc on accepte H0.

On peut passer au test de comapraison des moyennes. # Test d'homogénéité (de comparaison) des moyennes (H0 = "les moyennes sont égales")

```
t.test(Col_0,Mut_X1,var.equal = T)
```

```
##
## Two Sample t-test
##
## data: Col_0 and Mut_X1
## t = 1.9173, df = 29, p-value = 0.0651
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01631732 0.50537715
## sample estimates:
## mean of x mean of y
## 3.972222 3.727692
```

```
t.test(Col_0,Mut_X1,var.equal = T, alternative = "greater")
```

```
##
## Two Sample t-test
##
## data: Col_0 and Mut_X1
## t = 1.9173, df = 29, p-value = 0.03255
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.02782437 Inf
## sample estimates:
## mean of x mean of y
## 3.972222 3.727692
```

```
t.test(Col_0,Mut_X1,var.equal = T, alternative = "less")
```

```
##
## Two Sample t-test
##
## data: Col_0 and Mut_X1
## t = 1.9173, df = 29, p-value = 0.9675
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
## -Inf 0.4612355
## sample estimates:
## mean of x mean of y
## 3.972222 3.727692
```

La p-value est supérieure à 0.05 donc on ne rejette pas l'hypothèse nulle. Les moyennes ne présentent pas de différence significative

avec greater, on voit que la p-value est inférieure à 0.05 donc la moyenne de Col\_0 est supérieure à la moyenne de Mut\_X1.

Les hypothèses alternatives on peut choisir qu'elle soit unilatérale ou bilatérale.

Conclusion biologique : le gène pourrait avoir un impact sur la croissance et donc le mutant on réduit la taille des plantes. Autre hypothèse : Germination plus tardive car gène muté.

## Test non paramétriques d'homogénéité (sur une variable quantitative)

```
non_infected=c(0.021,0.15,0.023,0.03,0.022,0.05,0.035,0.1,0.03)
infected=c(1.22,1.12,1.06,1.04,0.86,1.24,1.96,0.9,2.5)
mean(non_infected)
```

```
## [1] 0.05122222
```

```
mean(infected)
```

```
## [1] 1.322222
```

## Test d'adéquation à la loi normale

### Test de Shapiro

```
shapiro.test(infected)
```

```
##
## Shapiro-Wilk normality test
##
## data: infected
## W = 0.77769, p-value = 0.01133
```

```
shapiro.test(non_infected)
```

```
##
## Shapiro-Wilk normality test
##
## data: non_infected
## W = 0.72241, p-value = 0.002591
```

### Test de Mann & Whitney (équivalent non paramétrique du test de Student)

```
wilcox.test(infected,non_infected) # bilatéral
```

```
## Warning in wilcox.test.default(infected, non_infected): cannot compute exact
## p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: infected and non_infected
## W = 81, p-value = 0.0004095
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(infected,non_infected,alternative="greater") # unilatéral
```

```
## Warning in wilcox.test.default(infected, non_infected, alternative =
## "greater"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: infected and non_infected
## W = 81, p-value = 0.0002047
## alternative hypothesis: true location shift is greater than 0
```