Tests Paramétriques

Plan

1. Rappels: principe des tests
2. 2 moyennes observées
3. 2 pourcentages observés
4. Corrélation
5. Plusieurs moyennes observées
6. Plusieurs pourcentages observés

Mode

1. Théorie
2. Exercice Dirigé, *logiciel R*
3. Exercice Individuel, *logiciel R*

V. Comparaison de plusieurs moyennes observées

Ex: Effet de la gravité sur la TAS, Ech. de 32 sujets

data<-read.csv2("D:\\BIOSTAT\\TABAC.csv", header=TRUE)

- Importer le fichier de données *TABAC.xls*
- Moyenne globale de la TAS

  m= **140,8** mmHg *mean(TAS)*
- Variance globale de la TAS

  s^2= **252,9** mmHg^2 *var(TAS)*
- Graphiques *boxplot(TAS, col="blue")*

Variable GRAV: 3 classes (0;1;2)

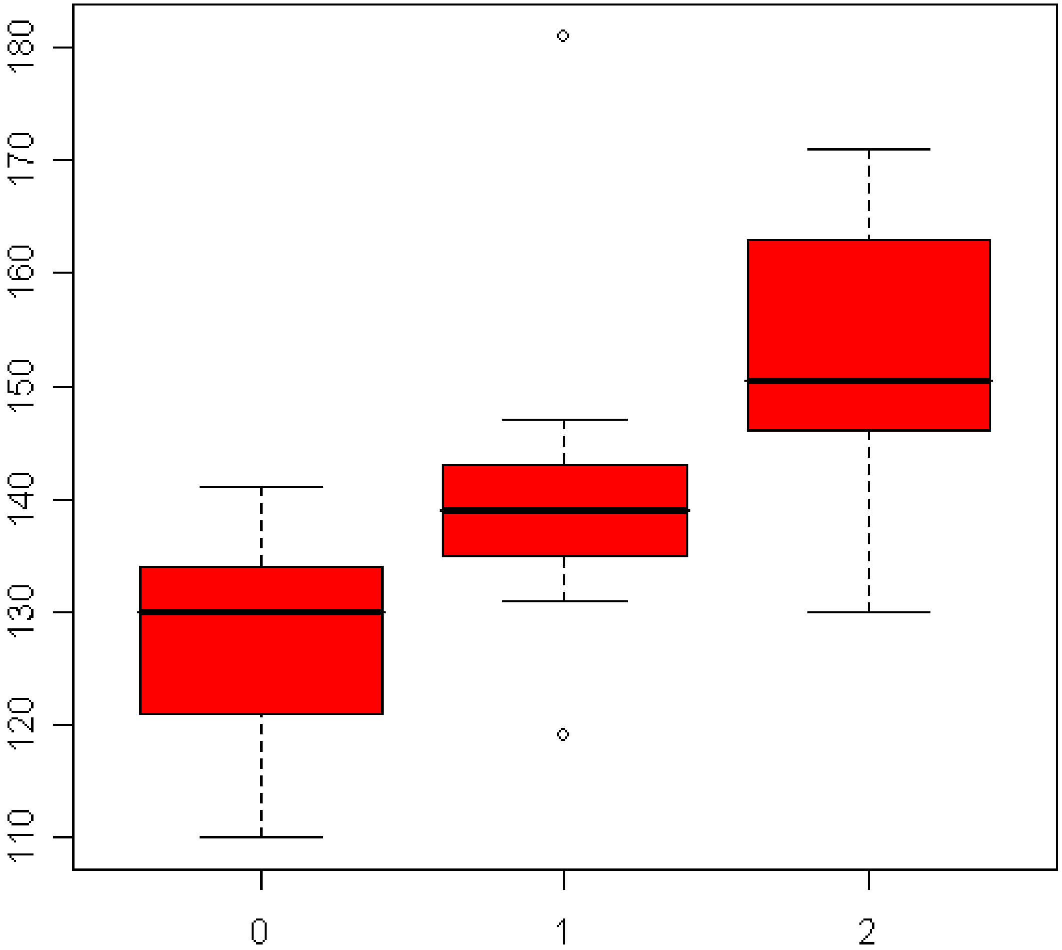
length(TAS[GRAV==0])

mean(TAS[GRAV==0])

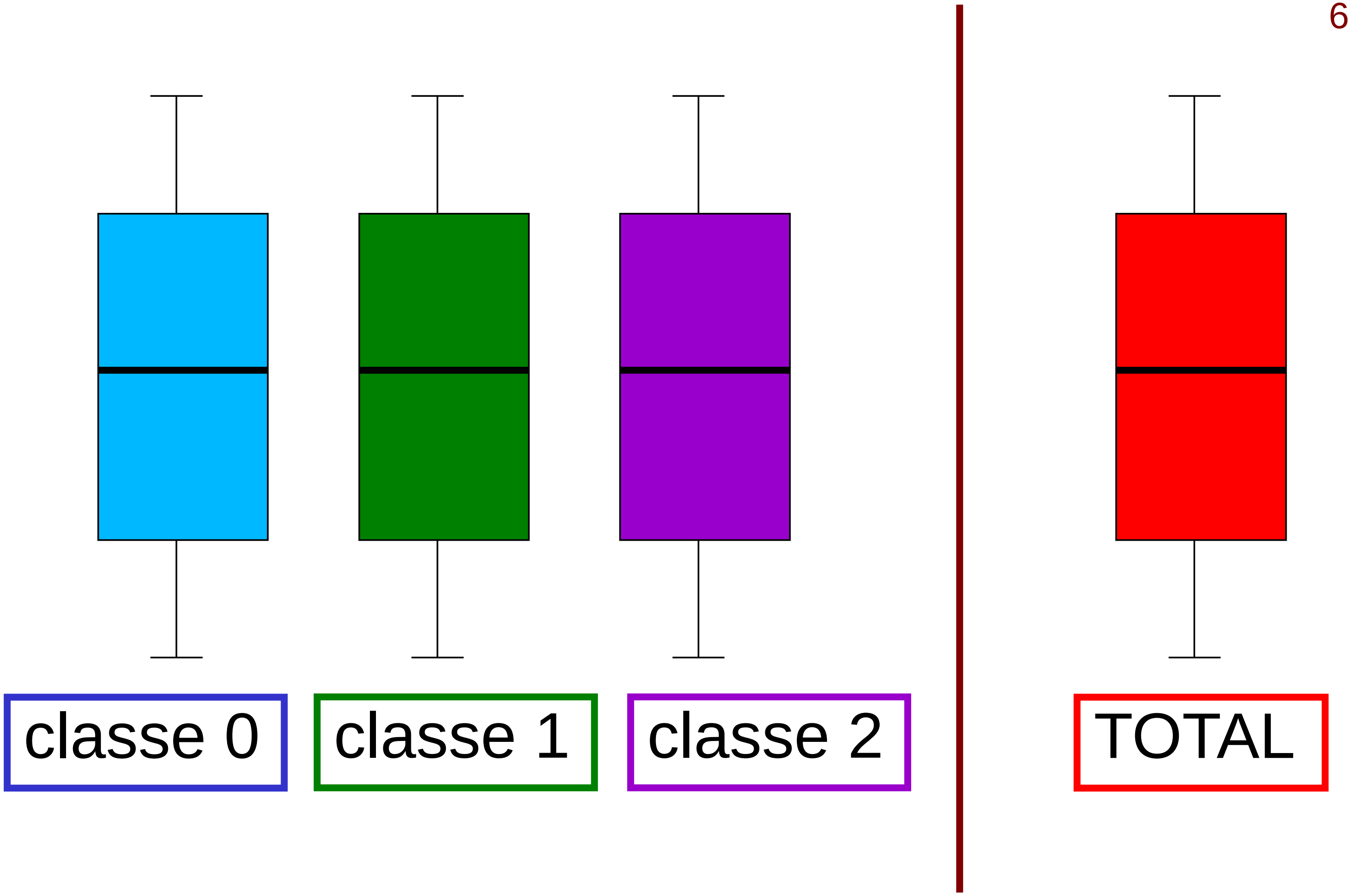
var(TAS[GRAV==0])

	nombre	moyenne	variance
classe 0	**9**	**127,1**	**111,6**
classe 1	**13**	**140,8**	**197,8**
classe 2	**10**	**153,1**	**152,8**

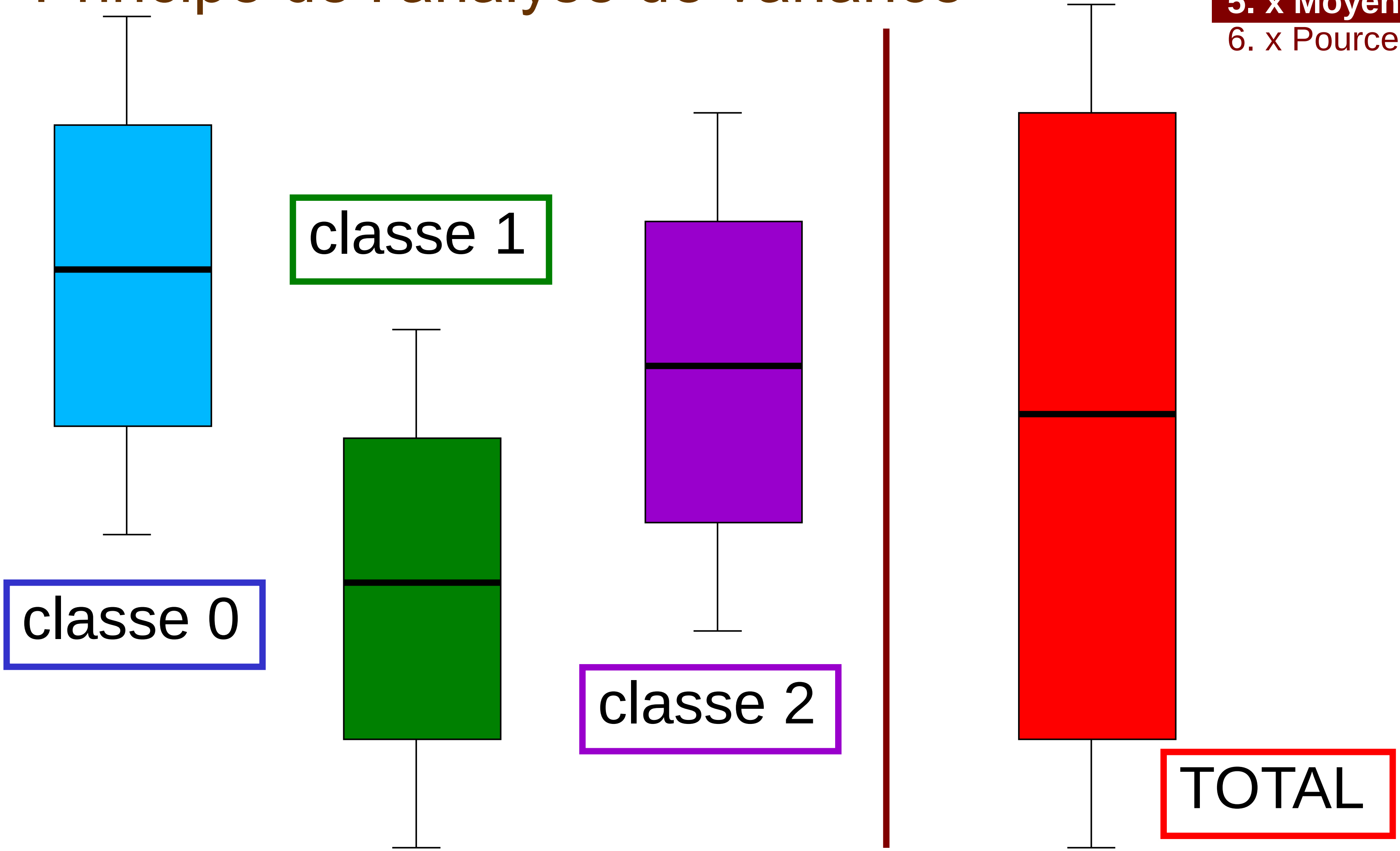
boxplot(TAS~GRAV, col="red")



Principe de l’analyse de variance

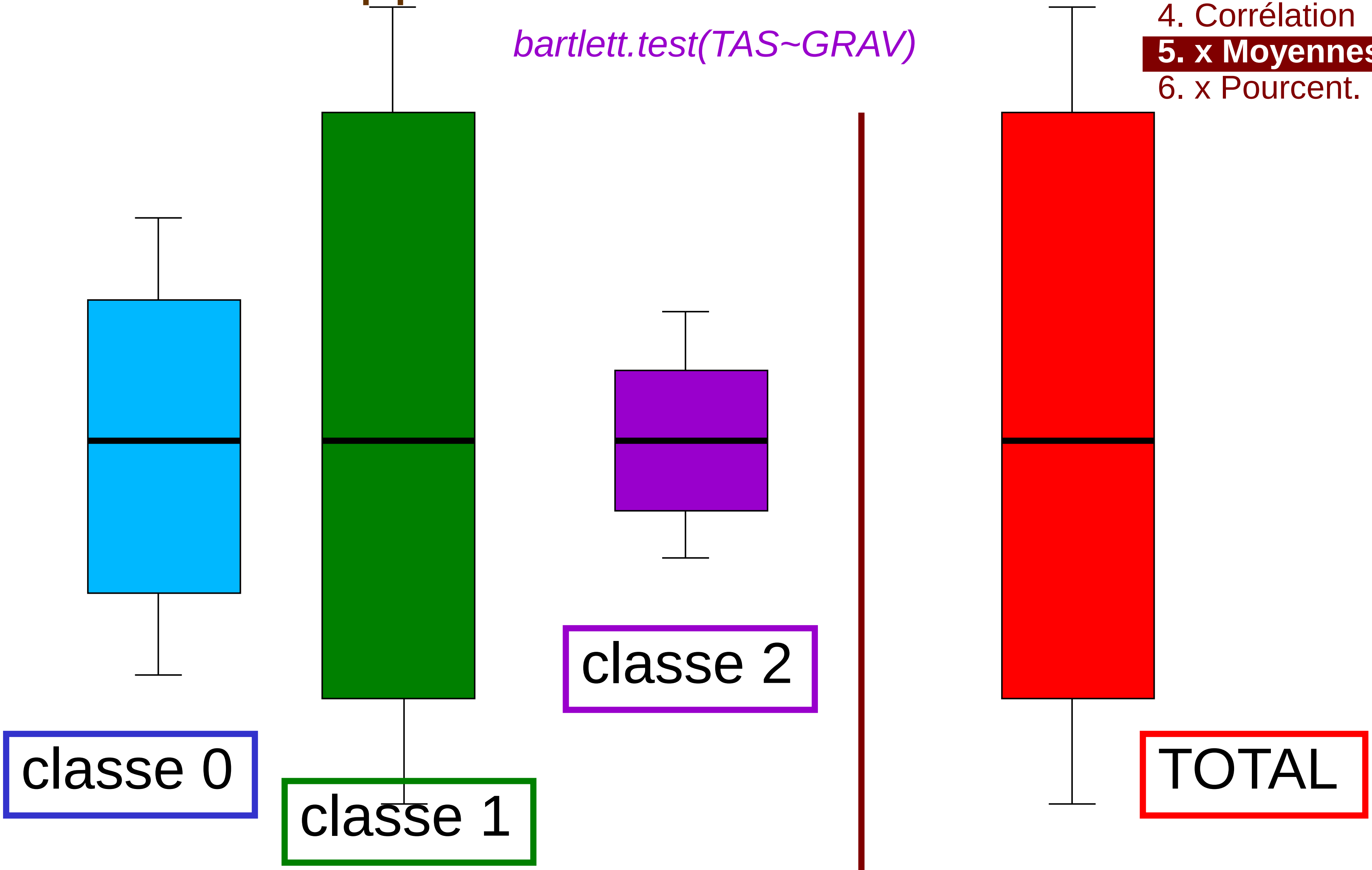


Principe de l'analyse de variance

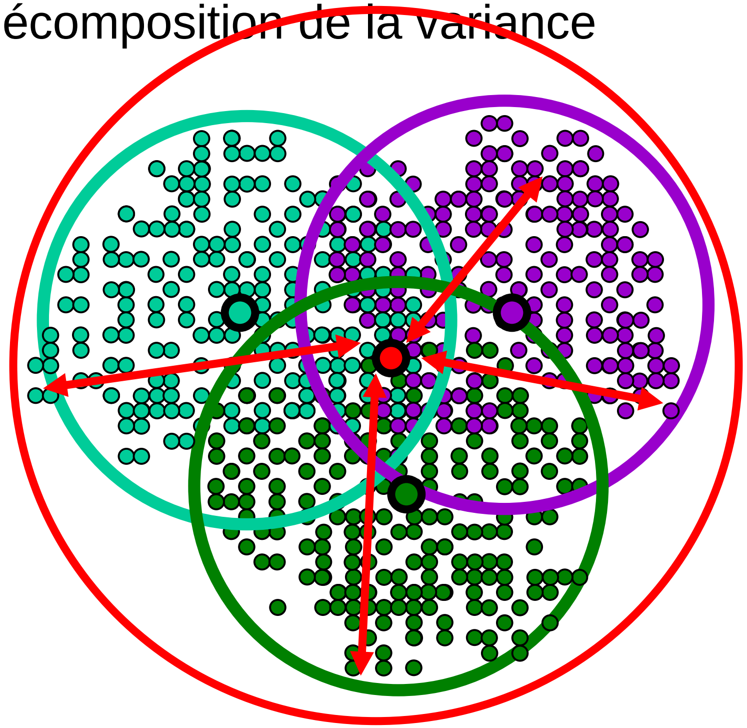


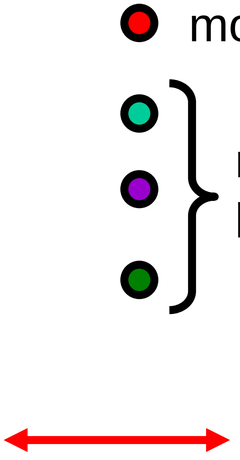
Condition d'application

bartlett.test(TAS~GRAV)

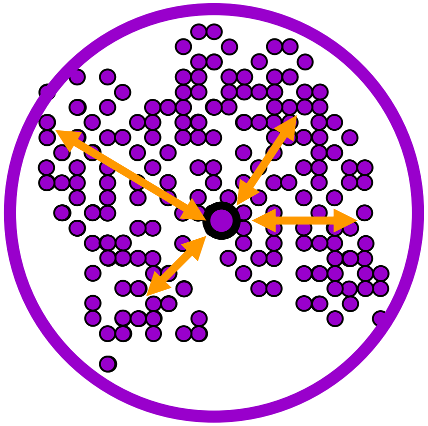


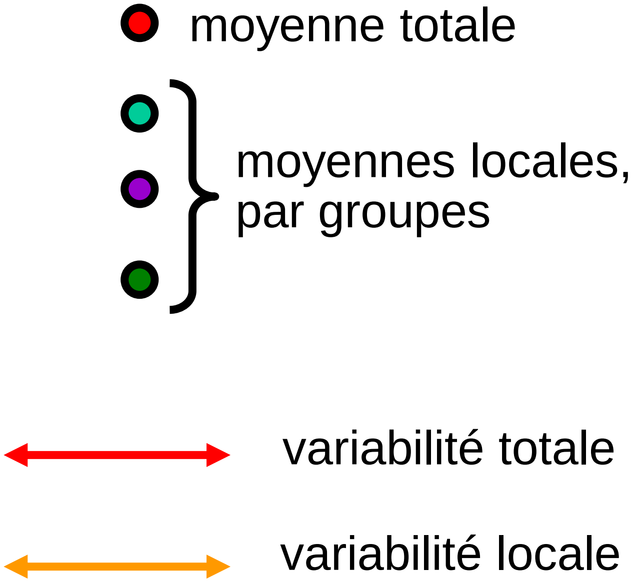
Décomposition de la variance



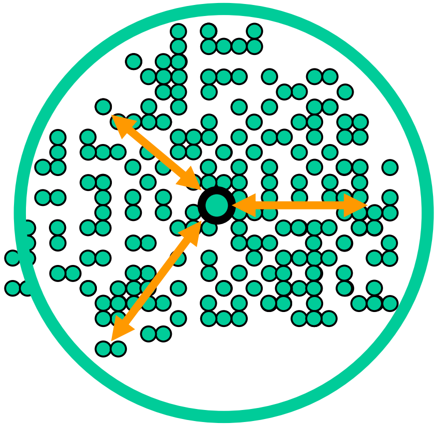


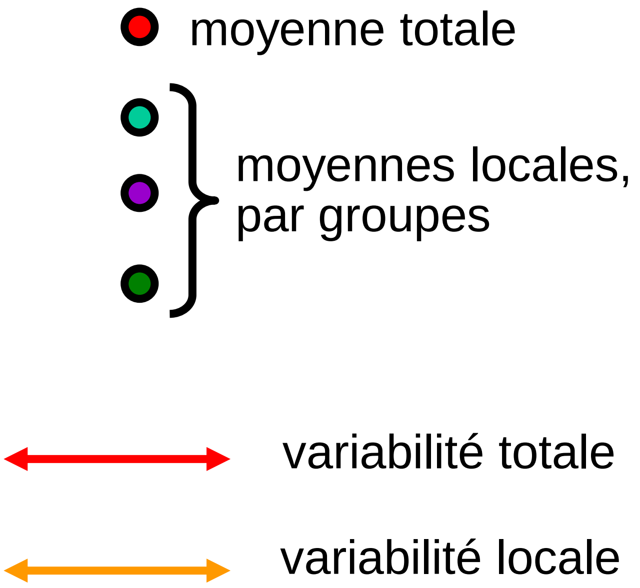
Décomposition de la variance



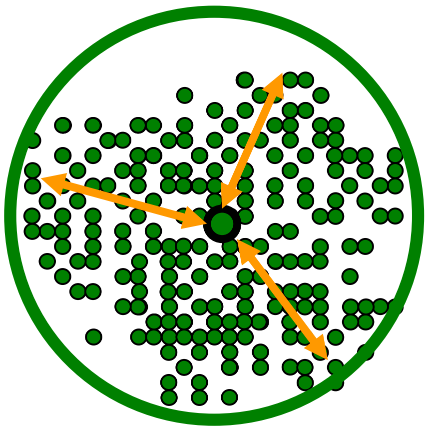


Décomposition de la variance





Décomposition de la variance



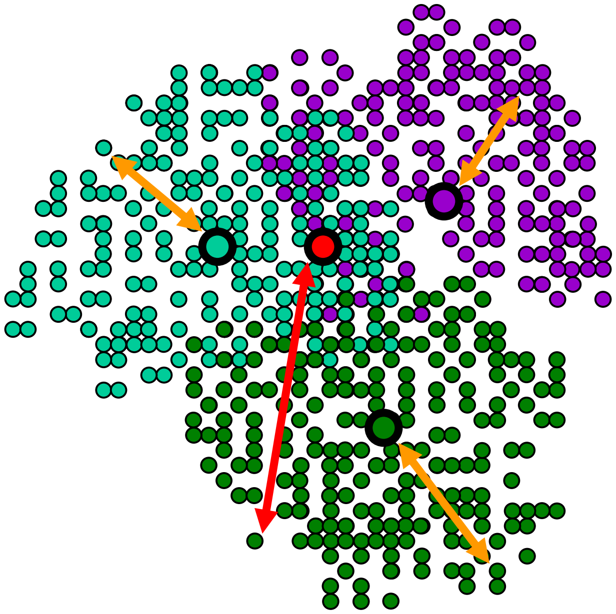
moyenne totale

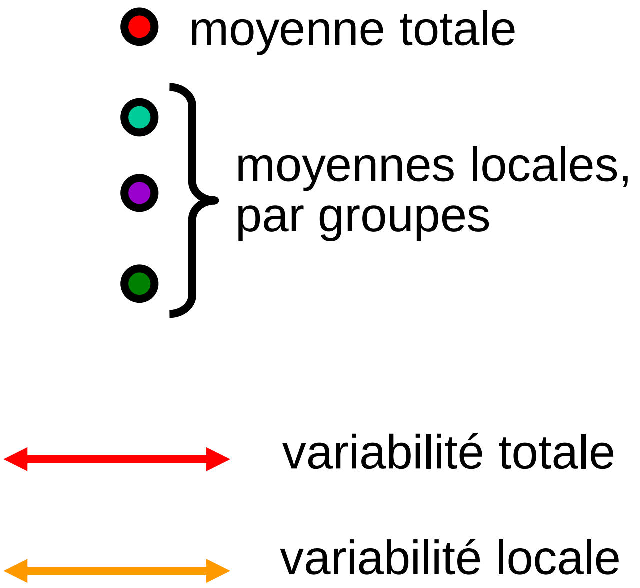
moyennes locales, par groupes

variabilité totale

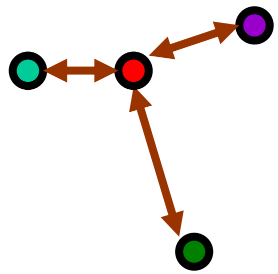
variabilité locale

Décomposition de la variance





Décomposition de la variance



moyenne totale

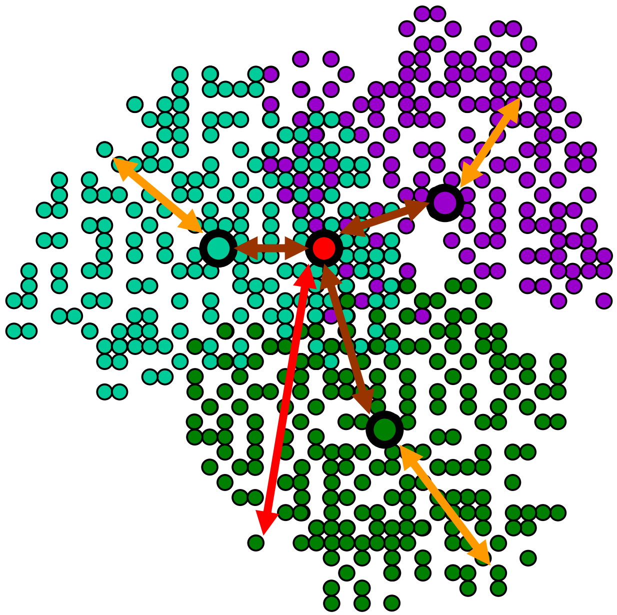
moyennes locales,
par groupes

variabilité totale

variabilité locale

variabilité entre-groupes

Décomposition de la variance



moyenne totale

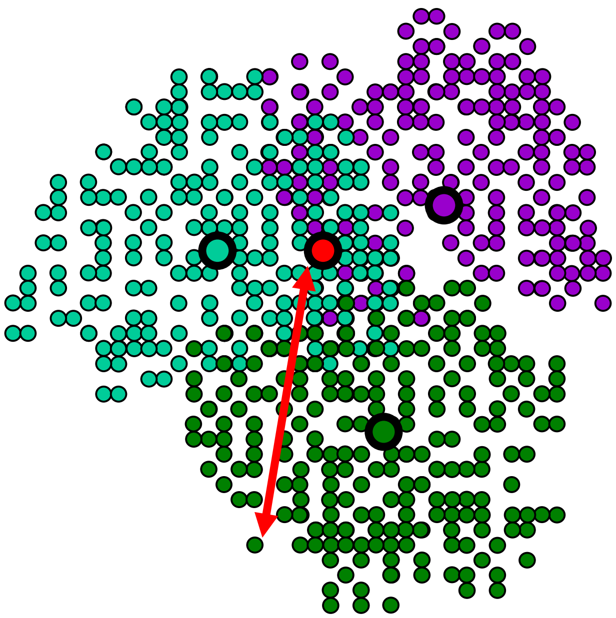
moyennes locales, par groupes

variabilité totale

variabilité locale

variabilité entre-groupes

Décomposition de la variance ⟷ variabilité **totale**



Somme des carrés des écarts, **totale**

$$SCE_T = \sum_{i=1}^{n} \left(x_i - m_T\right)^2$$

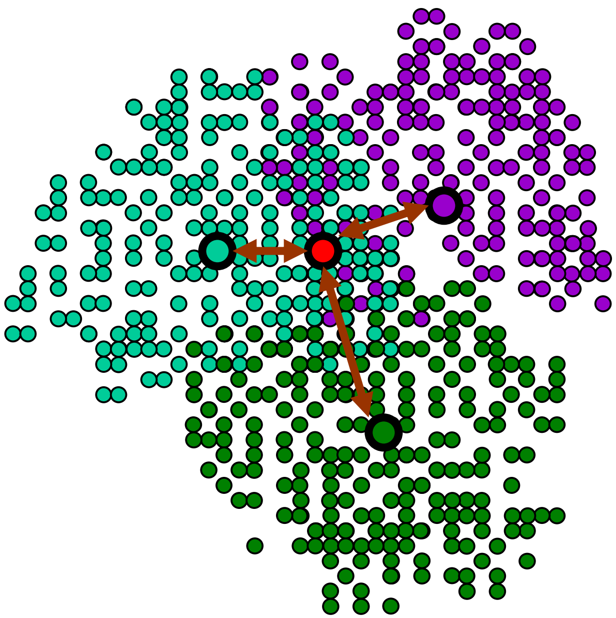
Variance **totale**

$$\sigma_T^2 = \frac{\sum_{i=1}^{n} \left(x_i - m_T\right)^2}{n-1} = \frac{SCE_T}{n-1}$$

m_T moyenne totale

m_1, m_2, m_3 moyennes locales, par groupes m_j

Décomposition de la variance ⟷ variabilité **entre-groupes**



Somme des carrés des écarts, **entre-groupes**

$$SCE_A = \sum_{j=1}^{k} n_j \left(m_j - m_T\right)^2$$

Variance **entre-groupes**

$$\sigma_A^2 = \frac{\sum_{j=1}^{k} n_j \left(m_j - m_T\right)^2}{k-1}$$

m_T moyenne totale

$\left.\begin{matrix} m_1 \\ m_2 \\ m_3 \end{matrix}\right\}$ moyennes locales, par groupes m_j

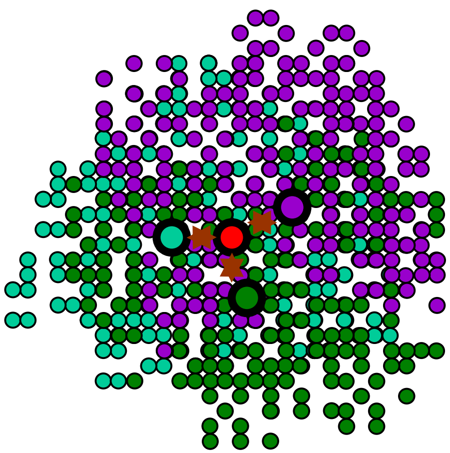
Décomposition de la variance ⟷ variabilité **entre-groupes**

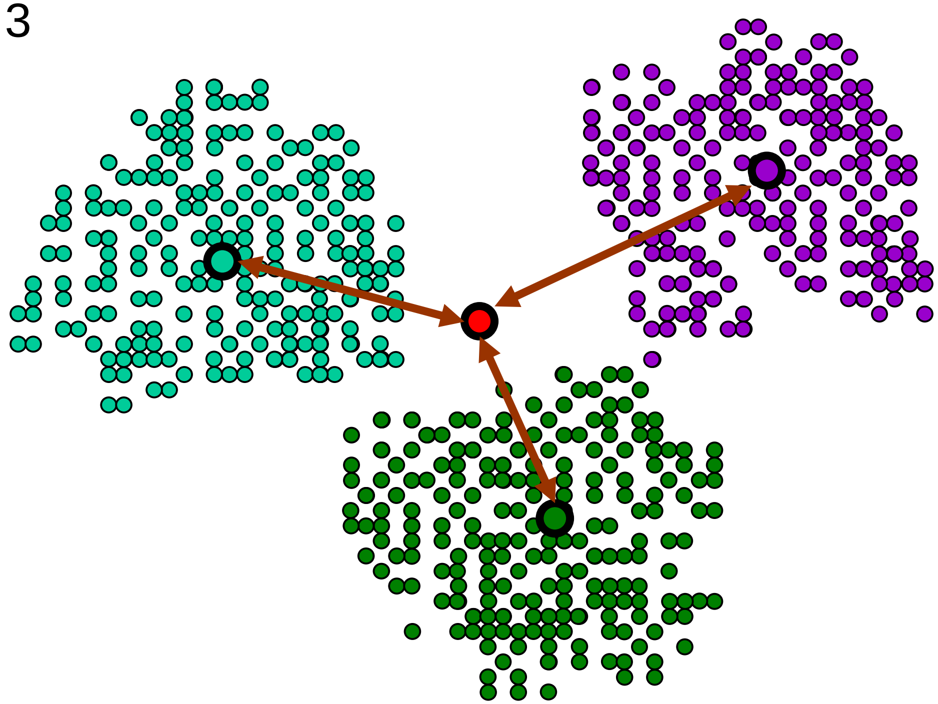
Facteur étudié A

ex: 3 classes
- gravité 0
- gravité 1
- gravité 3

Effet ++ du facteur A

Pas d'effet du facteur A

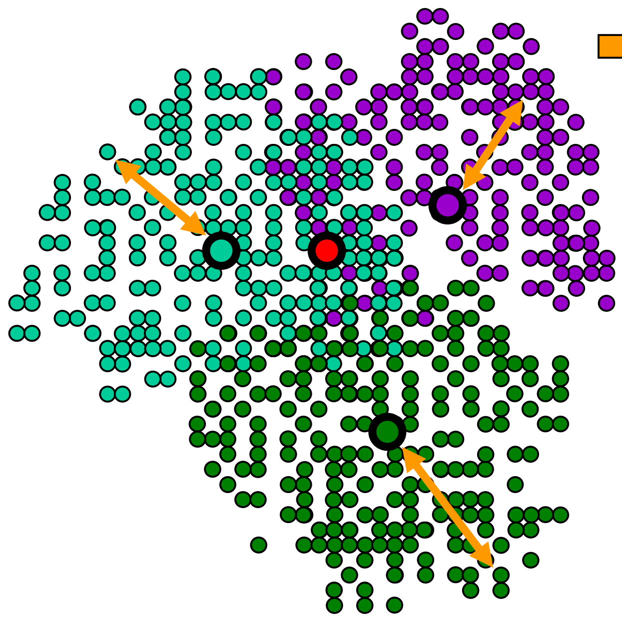




$$SCE_A = \sum_{j=1}^{k} n_j \left(m_j - m_T\right)^2 \downarrow\downarrow$$

$$SCE_A = \sum_{j=1}^{k} n_j \left(m_j - m_T\right)^2 \uparrow\uparrow$$

Décomposition de la variance ⟷ variabilité **locale**



Pour chaque groupe j

Somme des carrés des écarts, **locale**

$$SCE_j = \sum_{i=1,j}^{n_j} \left(x_{i,j} - m_j\right)^2$$

Variance **locale**

$$\sigma_j^2 = \frac{\sum_{i=1,j}^{n_j} \left(x_{i,j} - m_j\right)^2}{n_j - 1}$$

Somme des carrés des écarts, **Résiduelle**

$$SCE_R = \sum_{j=1}^{k} SCE_j = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - m_j\right)^2$$

m_T moyenne totale

m_1, m_2, m_3: moyennes locales, par groupes m_j

variance **Résiduelle** $\sigma_R^2 = SCE_R / (n-k)$

Tableau d'analyse de variance

Source de variation	ddl	SCE	variance	statistique F
Entre-groupe, Facteur A	k-1	$SCE_A = \sum_{j=1}^{k} n_j \left(m_j - m_T\right)^2$	$\sigma_A^2 = \frac{SCE_A}{k-1}$	$\boxed{F_{n-k}^{k-1} = \frac{\sigma_A^2}{\sigma_R^2}}$
Résiduelle	n-k	$SCE_R = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - m_j\right)^2$	$\sigma_R^2 = \frac{SCE_R}{n-k}$	
Totale	n-1	$SCE_T = \sum_{i=1}^{n} \left(x_i - m_T\right)^2$	$\sigma_T^2 = \frac{SCE_T}{n-1}$	

1. Hypothèses

H0: $\mu_1 = \mu_2 = \mu_3$, pas d'effet de la gravité sur TAS

H1: une égalité au moins est fausse, il existe un effet de la gravité sur la TAS

2. Prédiction sous H0

sous H0 et si les conditions d'application sont respectées

$$F = \frac{\sigma_A^2}{\sigma_R^2} \rightarrow F_{n-k}^{k-1}$$

Loi de Fisher

1. Hypothèses
2. Prédiction sous H0:

Conditions d’applications

- Loi de X Normale dans chaque groupe
(ou effectifs >30 dans chaque groupe)

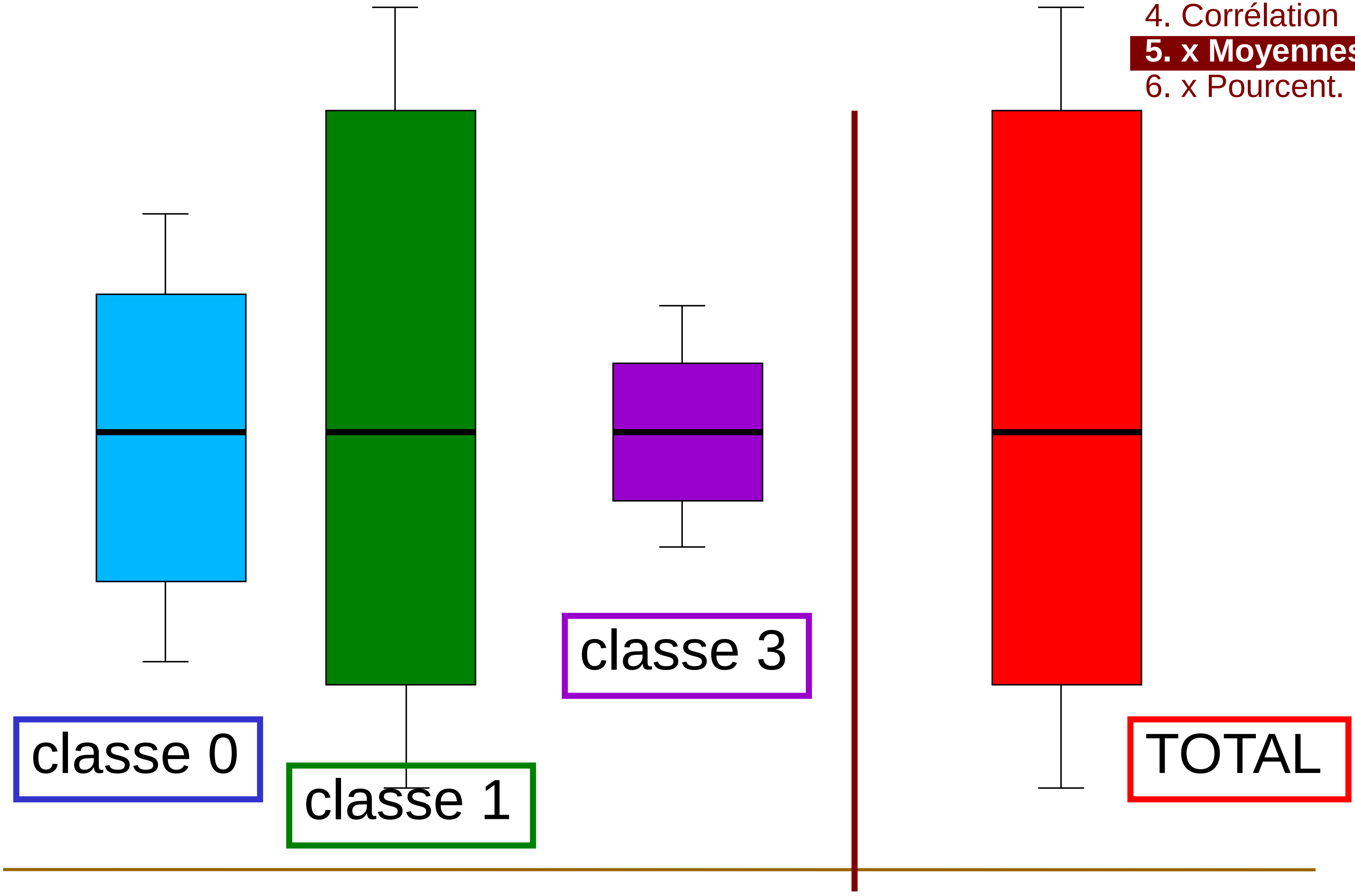
ET

- Variance de X constante

ET

- Independence des individus

Variances constantes



1. Hypothèses
2. Prédiction sous H0
3. Confrontation

GRAV2<-factor(GRAV)
anova<-aov(TAS~GRAV2)
summary(anova)

conditions d'applications

- Loi de X Normale dans chaque groupe — L(X/Y)→N
- Variance de X constante — var(X/Y) cst
- Independence des individus

RAPPELS: Conditions d'applications

- Test de Student
  - Indépendance des sujets
  - Distribution de X, par groupe, Normale

    $$L(X/Y) \to N$$

  - Variances égales

    $$\operatorname{var}(X/Y) \quad cst$$

- Correlation
  - Indépendance des sujets
  - Distribution de X, par valeurs de Y, Normale

    $$L(X/Y) \to N$$

  - Variances égales

    $$\operatorname{var}(X/Y) \quad cst$$

  - Relation linéaire

- ANOVA
  - Indépendance des sujets
  - Distribution de X, par groupe, Normale

    $$L(X/Y) \to N$$

  - Variances égales

    $$\operatorname{var}(X/Y) \quad cst$$

Conditions d'applications

- égalité des variances: *bartlett.test(TAS~GRAV2)*

H0: les variances sont égales
H1: une variance au moins est différente

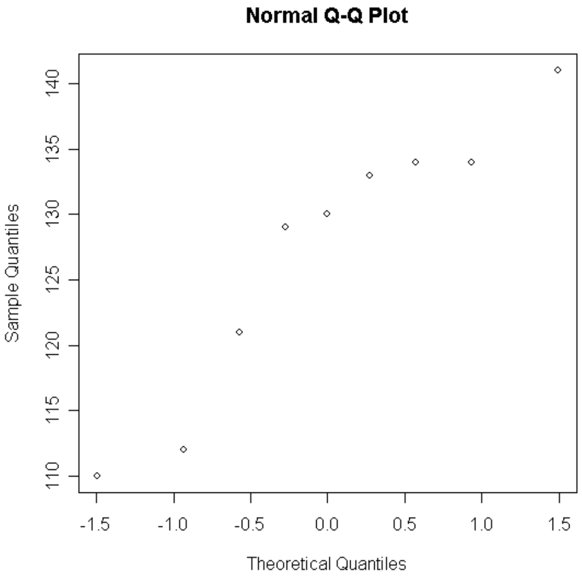
```
Bartlett test of homogeneity of variances
data: TAS by GRAV2
Bartlett's K-squared = 0.7269, df = 2, p-value = 0.6953
```

- p>0,05
- Test non significatif
- Non rejet de H0 au risque β
- On ne met pas en évidence de différence entre les variances

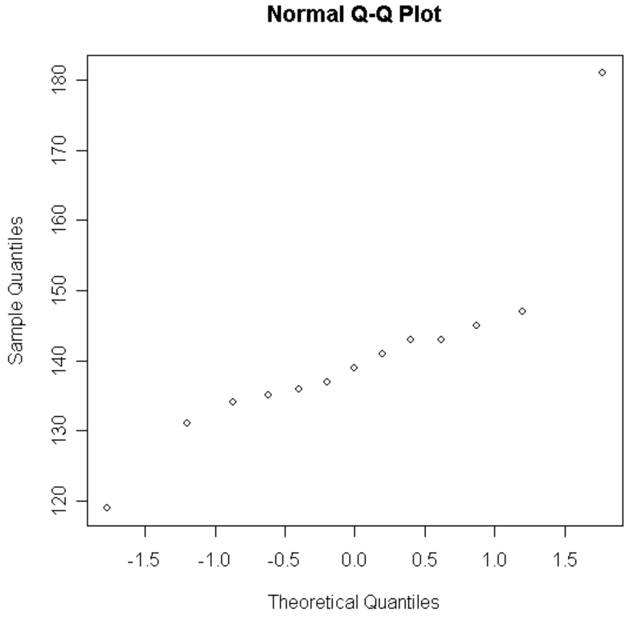
Conditions d'applications

Normalité des distributions dans chaque groupe

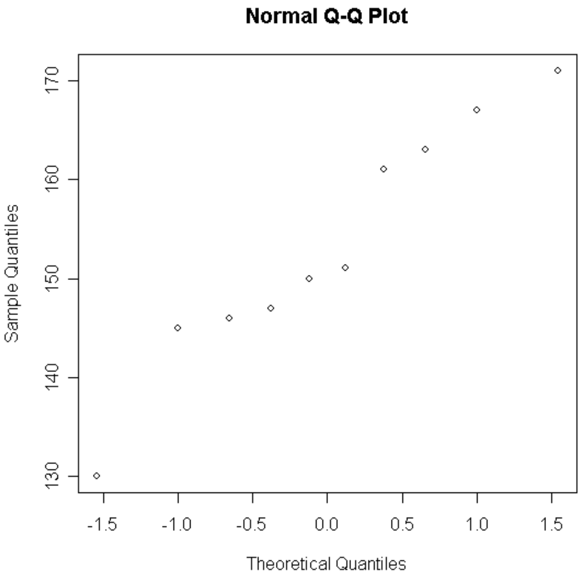
qqnorm(TAS[GRAV==0])



qqnorm(TAS[GRAV==1])



qqnorm(TAS[GRAV==2])



1. Hypothèses

probleme df et le reste du calcul

2. Prédiction sous H0

3. Confrontation

GRAV2<-factor(GRAV)
anova<-aov(TAS~GRAV2)
summary(anova)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
GRAV2	2	3199.4	1599.7	9.995	0.000499 ***
Residuals	29	4641.5	160.1		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sources de variations

Degrés de liberté

SCE

Variances

Statistique F

« petit p »

1. Hypothèses
2. Prédiction sous H0
3. Confrontation
4. Interprétation
   - $p<0,05$
   - Test Significatif
   - Rejet de H0 au risque α
   - Il existe un lien entre la gravité et la TAS
   - Moyennes de TAS:

127,1[118,9-135,2] 140,8[132,4-149,4] 153,1[144,3-161,9]

ATTENTION

1. On rejette l'hypothèse nulle $\mu_1 = \mu_2 = \mu_3$

$\Rightarrow$ une égalité au moins est fausse

$\Rightarrow$ il y a au moins une différence

$\Rightarrow$ mais on n'a pas testé laquelle:

$\mu_1 \neq \mu_2$ ou $\mu_1 \neq \mu_3$ ou $\mu_2 \neq \mu_3$ **???**

!

2. On ne peut pas tester **ENSUITE** les moyennes 2 à 2 sinon **α** ↑↑

inégalité de Bonferroni

$$\alpha_{total} \leq \sum_{i=1}^{k\ tests} \alpha_i$$

$\Rightarrow$ test post-hoc

Exercice

- fichier *TABAC.csv*
- Y a-t-il une différence entre la moyenne de la TAS chez les sujets ayant des antécédents différents ?

- Echantillon de 32 sujets
- 3 groupes

attach(data)

ATCD 0

m_0= 127,4 mmHg
s_0^2= 133,2 mmHg^2

mean(TAS[ATCD==0])
var(TAS[ATCD==0])

ATCD 1

m_1= 148,5 mmHg
s_1^2= 194,3 mmHg^2

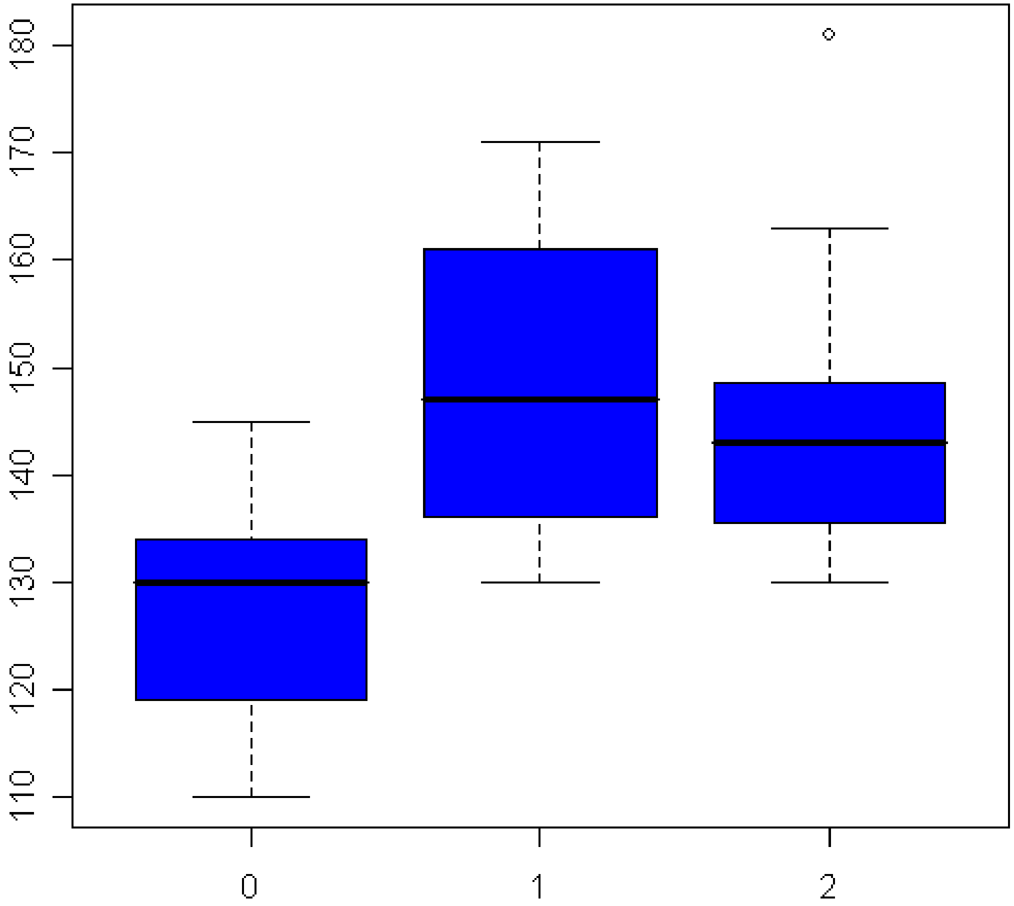
mean(TAS[ATCD==1])
var(TAS[ATCD==1])

ATCD 2

m_2= 145,6 mmHg
s_2^2= 202,8 mmHg^2

mean(TAS[ATCD==2])
var(TAS[ATCD==2])

boxplot(TAS~ATCD,col="blue")



1. Hypothèses

H0: $\mu_1 = \mu_2 = \mu_3$, pas d'effet des antécédents sur TAS

H1: une égalité au moins est fausse, il existe un effet des antécédents sur la TAS

2. Prédiction sous H0

sous H0 et si les conditions d'application sont respectées

$$F = \frac{\sigma_A^2}{\sigma_R^2} \rightarrow F_{n-k}^{k-1}$$

1. Hypothèses
2. Prédiction sous H0:

Conditions d'applications

- Loi de X Normale dans chaque groupe

(ou effectifs >30 dans chaque groupe)

ET

- Variance de X constante

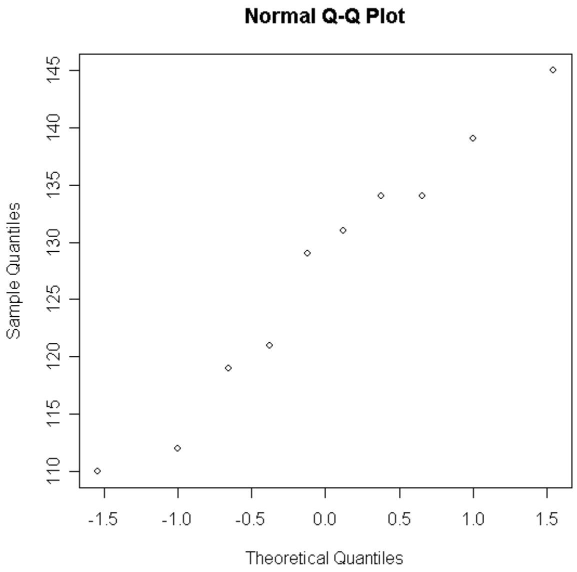
ET

- Independence des individus

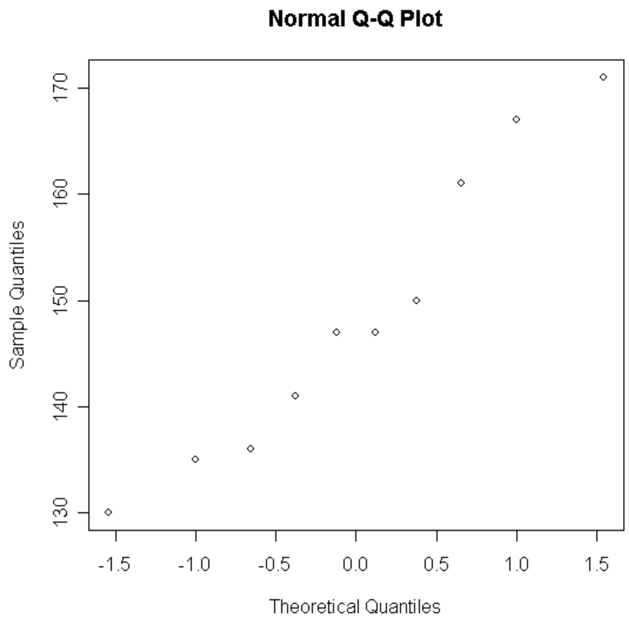
Conditions d’applications

➡ Normalité des distributions dans chaque groupe

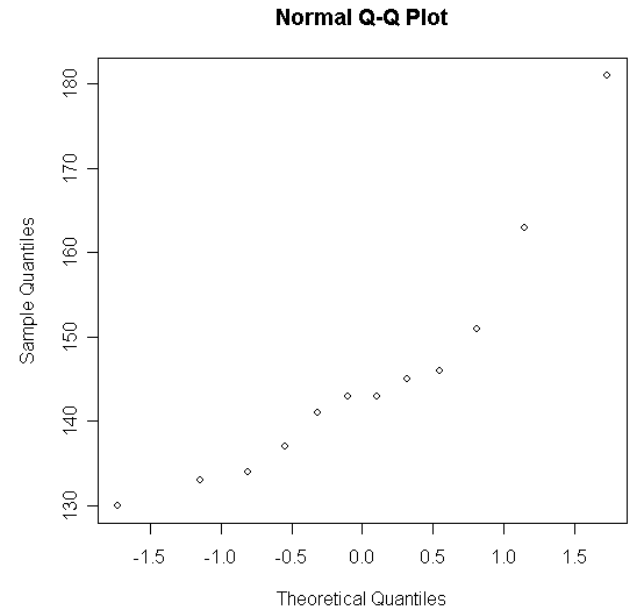
qqnorm(TAS[ATCD==0])



qqnorm(TAS[ATCD==1])



qqnorm(TAS[ATCD==2])



Conditions d'applications

- égalité des variances: *bartlett.test(TAS~ATCD)*

H0: $\sigma_0^2 = \sigma_1^2 = \sigma_2^2$ les variances sont égales
H1: une variance au moins est différente

```
Bartlett test of homogeneity of variances
data: TAS by ATCD
Bartlett's K-squared = 0.457, df = 2, p-value = 0.7957
```

- p>5%
- Test non significatif
- Non rejet de H0 au risque β
- On ne met pas en évidence de différence entre les variances

1. Hypothèses

probleme df et le reste du calcul

2. Prédiction sous H0

3. Confrontation

ATCD2<-factor(ATCD)
anova<-aov(TAS~ATCD2)
summary(anova)

```
            Df   Sum Sq  Mean Sq   F value   Pr(>F)
ATCD2        2   2663.1   1331.5    7.4577   0.002437 ***
Residuals   29   5177.8    178.5
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sources de variations

Degrés de liberté

SCE

Variances

Statistique F

« petit p »

1. Hypothèses
2. Prédiction sous H0
3. Confrontation
4. Interprétation
   - p<0,05
   - Test Significatif
   - Rejet de H0 au risque α
   - Il existe un lien entre antécédents et TAS
   - Moyennes de TAS:

127,4[119,2-135,7]   148,5[138,5-158,5]   145,61[136,5-154,6]

Références

- Jean Bouyer: *Méthodes statistiques, Médecine-Biologie*, éditions INSERM

- Coll. (CIMES): *Biostatistiques*, éditions Omnisciences

Contact

jean.gaudart@univ-amu.fr

http://sesstim.univ-amu.fr

Faculté de Médecine de Marseille

