COMP. K GROUPES

ANOVA et assimilés

L3 STE 2012/2013

Introduction : comparaisons multiples

Analyse de variance à 1 facteur

1. **Principe de l'ANOVA 1 facteur**
   - 1.1 Définitions : dispersion, variance
   - 1.2 Sources de variabilité
   - 1.3 F-ratio et test de H0
2. **Formalisation de l'ANOVA 1 facteur**
   - 2.1 Système de notation
   - 2.2 Décomposition de la variance
   - 2.3 Test de H0
3. **Conditions d'application**
4. **Deux modèles d'ANOVA**
5. **Tests *a posteriori***

Alternatives non-paramétriques

Comparaison de plus de 2 moyennes

Exemple : 3 groupes

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3$$

Option 1 : multiple *t*-tests

$$\begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 \\ \mathrm{H}_0 : \mu_1 = \mu_3 \\ \mathrm{H}_0 : \mu_2 = \mu_3 \end{cases}$$

Problèmes :

- laborieux : $k(k-1)/2$ comparaisons avec *k* groupes
- **<u>inadapté</u> : augmentation du risque** α ($\alpha = 0.14$ avec 3 groupes)

Comparaison de plus de 2 moyennes

Exemple : 3 groupes

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Option 1 : multiple *t*-tests

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_0 : \mu_1 = \mu_3 \\ H_0 : \mu_2 = \mu_3 \end{cases}$$

Correction du seuil de significativité

Correction pour comparaisons multiples

Tests multiples :

- *m* hypothèses nulles
- chacune est déclarée soit significative soit non-signicative

	H_0 vraie	H_0 fausse	TOTAL
déclarée significative	V	S	R
déclarée non-significative	U	T	$m - R$
Total	m_0	$m - m_0$	m

m : nombre total d'hypothèses testées

m_0 : nombre d'hypothèses nulles vraies

V : nombre de faux positifs (erreur de type I)

S : nombre de vrais positifs

U : nombre de vrais négatifs

T : nombre faux négatifs (erreur type II)

R : nombre d'hypothèses nulles rejetées

Q : proportion de fausses découvertes

$$Q = \frac{V}{R}$$

Correction pour comparaisons multiples

Quantités à contrôler :

- *FWER :* familywise error rate
- *FDR = E[Q] :* false discovery rate

	H_0 vraie	H_0 fausse	TOTAL
déclarée significative	V	S	R
déclarée non-significative	U	T	$m - R$
Total	m_0	$m - m_0$	m

$$FWER = \Pr(V \geq 1) \leq \alpha$$

- contrôle de la probabilité de faire au moins 1 erreur de type I à un risque α

Correction pour comparaisons multiples

Si les tests sont indépendants, alors

$$\alpha = 1 - (1-\gamma)^m$$

γ : le risque d'erreur de type I de chaque test

si γ = 5% et m = 3 alors α = 0.14

Si il existe une dépendance entre les tests, alors

$$\alpha < m \times \gamma$$

si γ = 5% et m = 3 alors α = 0.15

Procédures de contrôle du *FWER* et correction du seuil de significativité de chaque test:

Bonferroni : $\gamma = \alpha / m$

Šidák : $\gamma = 1 - (1-\alpha)^{1/m}$

Tukey HSD : voir cours

Holm (Bonferroni sequentiel): $\gamma_i = \alpha / (m - i + 1)$

Comparaison de plus de 2 moyennes

Exemple : 3 groupes

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3$$

Option 1 : multiple *t*-tests

$$\begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 \\ \mathrm{H}_0 : \mu_1 = \mu_3 \\ \mathrm{H}_0 : \mu_2 = \mu_3 \end{cases}$$

+ *Correction seuil de significativité*

Option 2 : Analyse de variance (ANOVA)

$$\begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \\ \mathrm{H}_1 : \text{tous les moyennes ne sont pas} \end{cases}$$

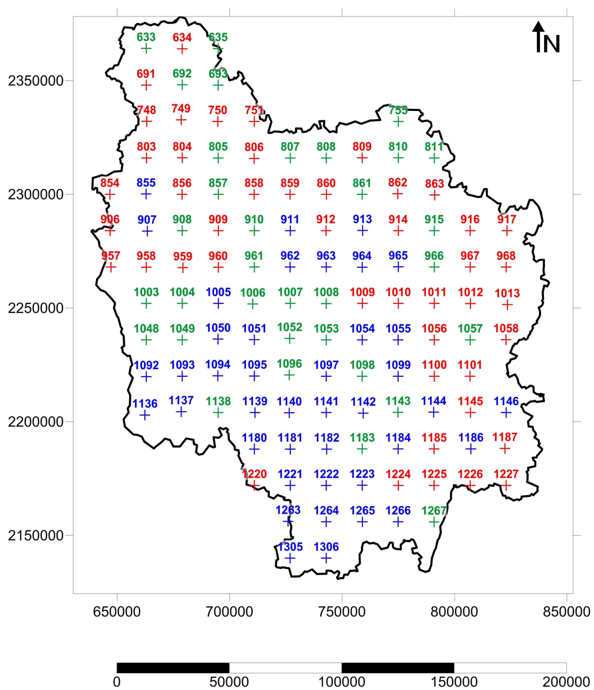
ANOVA

Analyse of Variance

ANOVA 1 facteur / One-way ANOVA

Principe de l'ANOVA

Exemple : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)



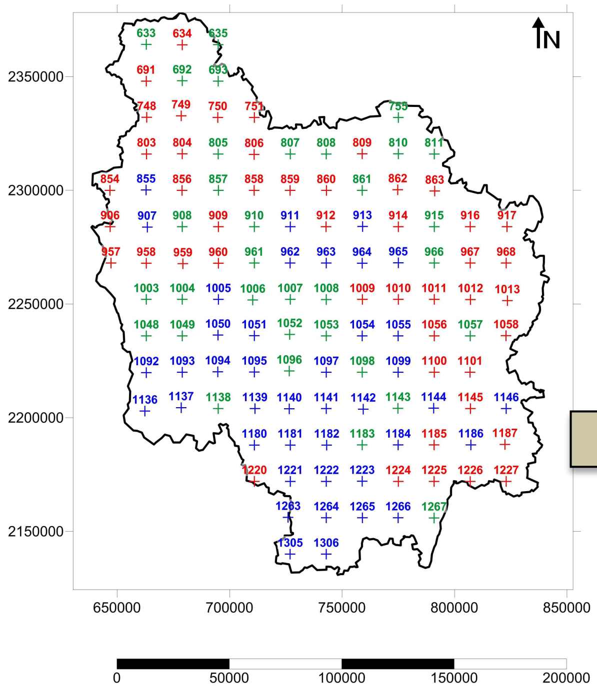
90 sols échantillonnés

3 types d'occupation : Prairie, Champs, Forêt

Example : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)

Principe de l'ANOVA

Exemple : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)



90 sols échantillonnés

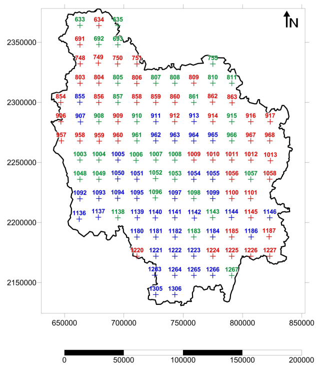
3 types d'occupation : Prairie, Champs, Forêt

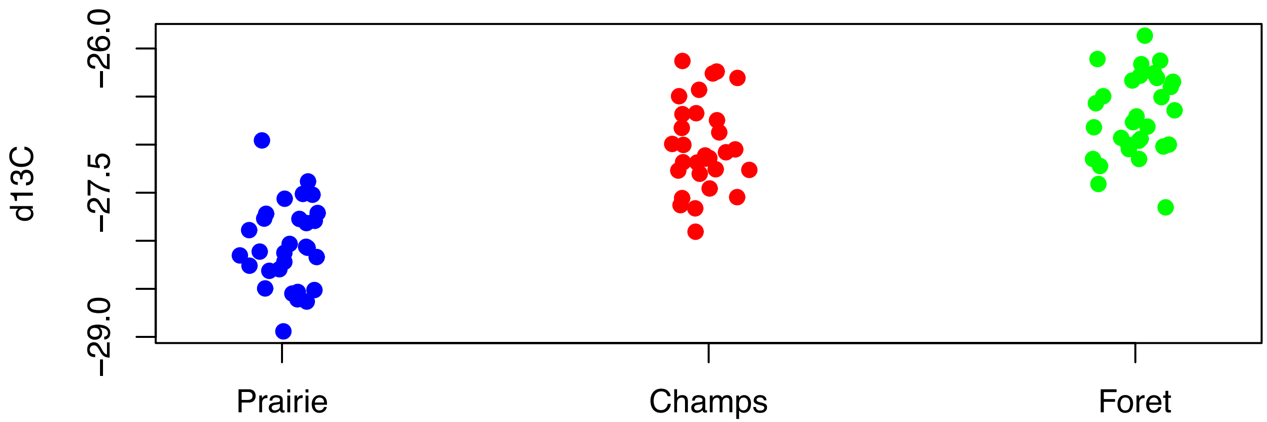
Example : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)

Prairie	**Champs**	**Forêt**
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
-27.76	-27.00	-27.01
-28.03	-26.42	-26.94
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

Principe de l’ANOVA

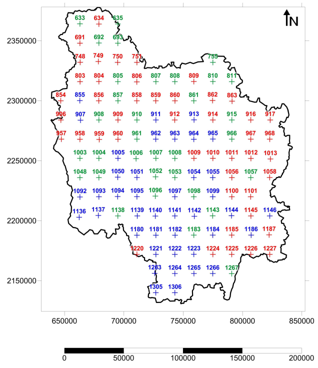
Exemple : Effet de la nature de l’occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)

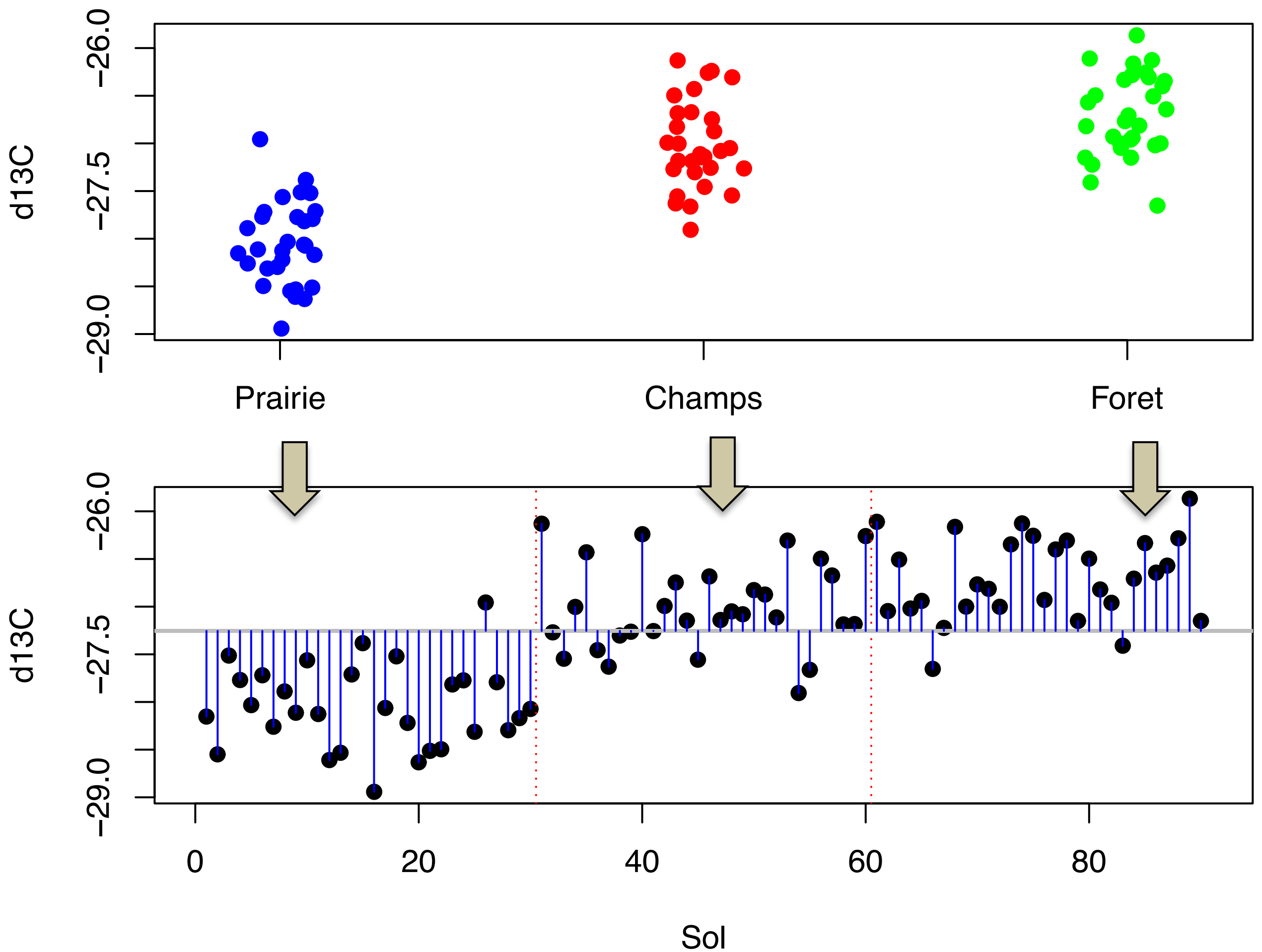




Principe de l'ANOVA

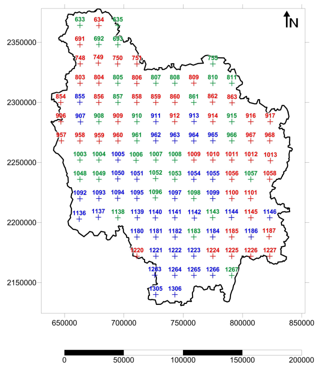
Exemple : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)

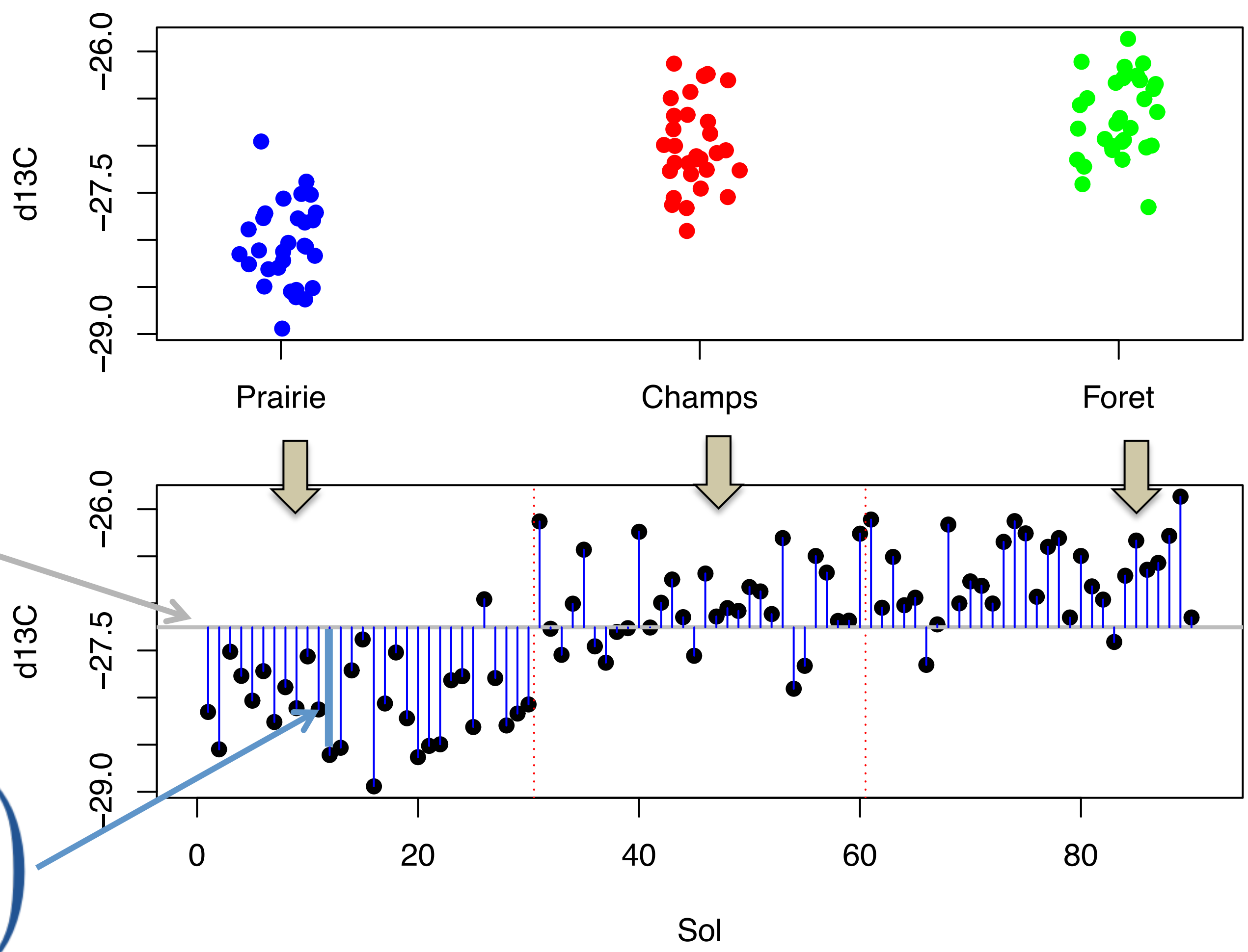




Principe de l'ANOVA

Exemple : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)

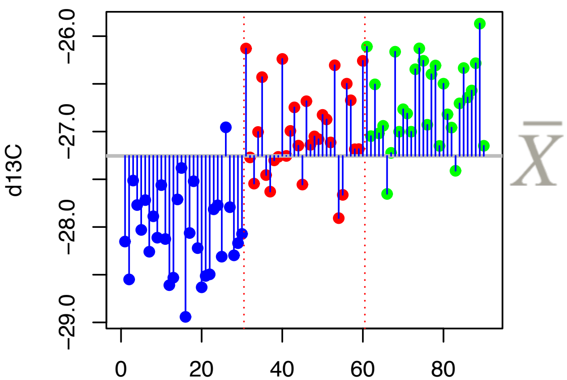




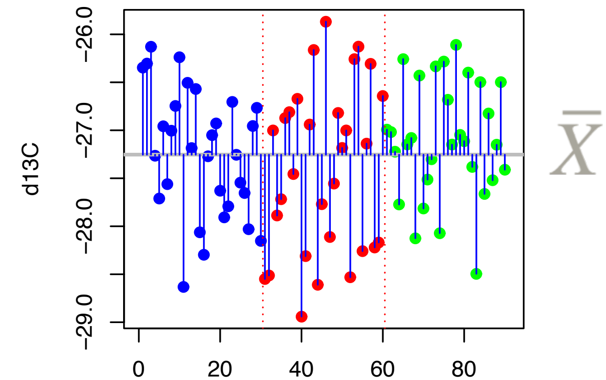
$$\bar{X}$$

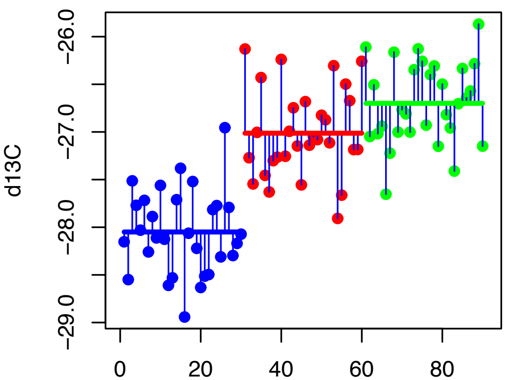
$$\left(X_i - \bar{X}\right)$$

ANOVA 1 facteur – Principe

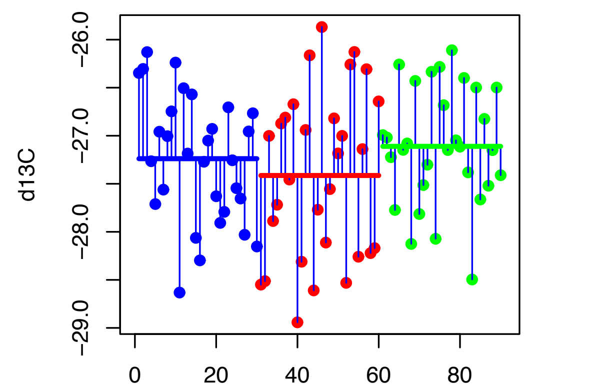


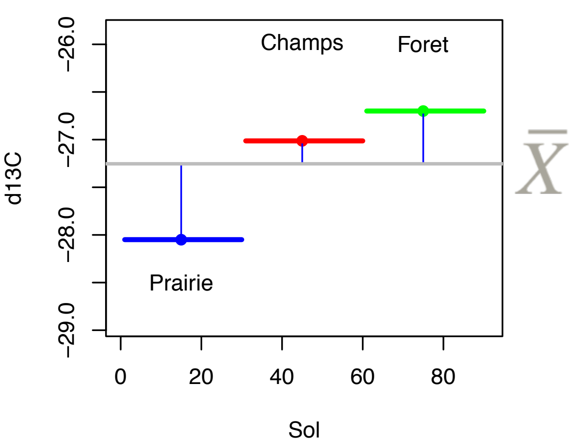
Dispersion totale



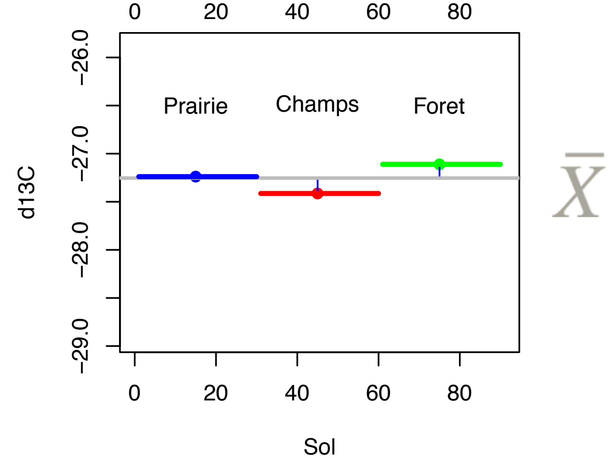


Dispersion intra-groupe





Dispersion inter-groupe



ANOVA 1 facteur – Principe / Source de variabilité

Impossible

de comparer directement les dispersions intra-groupe et inter-groupe

Comparaison des variances : « Carrés moyens »

Carré moyen = Somme des carrés / degré de liberté

Rappel :

Nombre de **degrés de liberté** = nombre d’unités d’information utilisées pour calculer une statistique

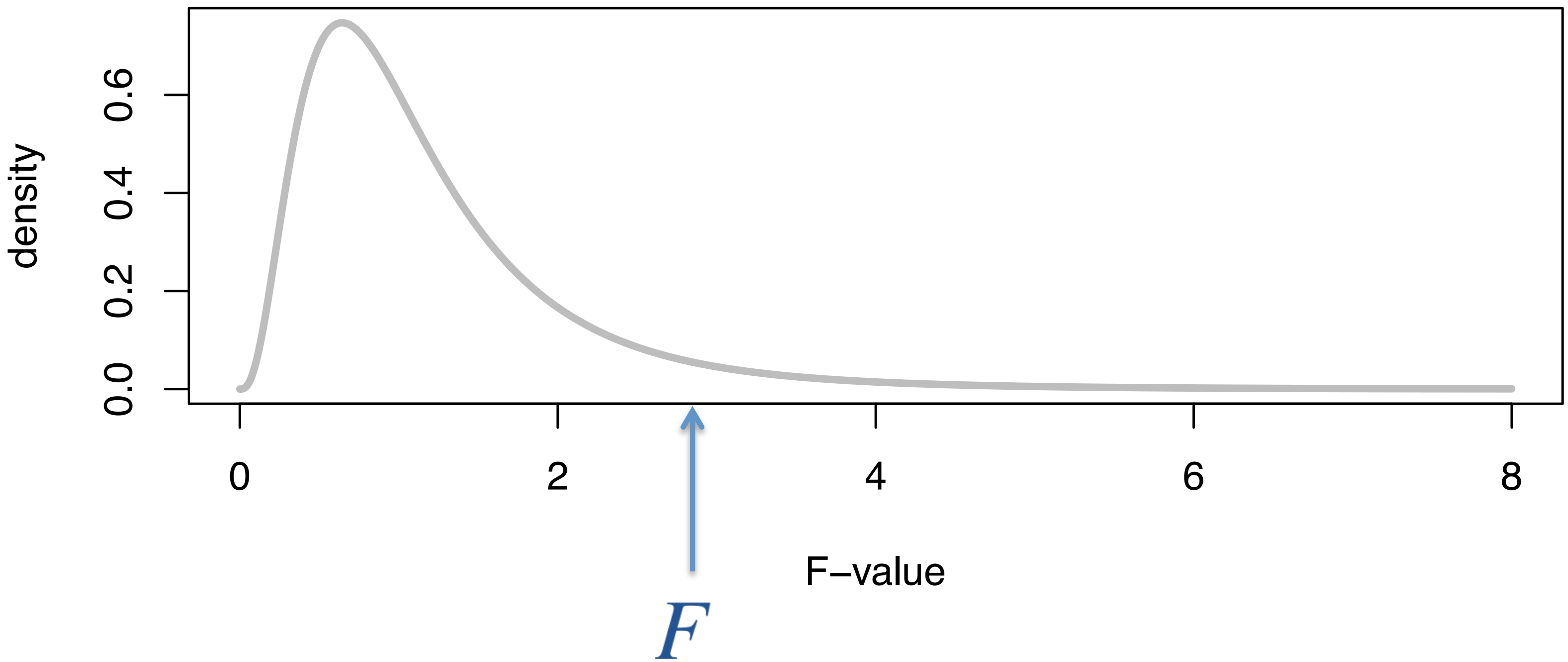
Test de l'hypothèse nulle

Comparaison des variances : « Carrés moyens »

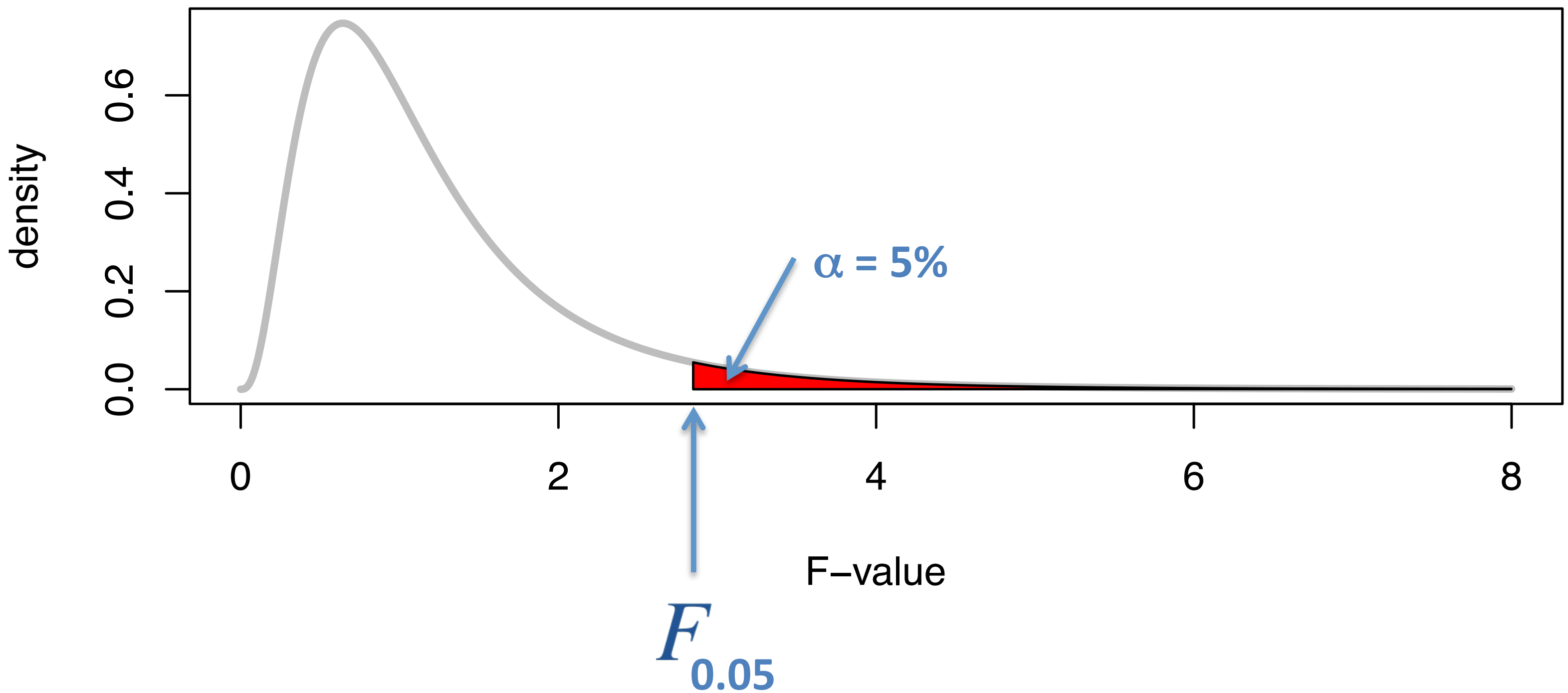
⇩

Calcul du F-ratio: $F = \dfrac{V_I}{V_E}$

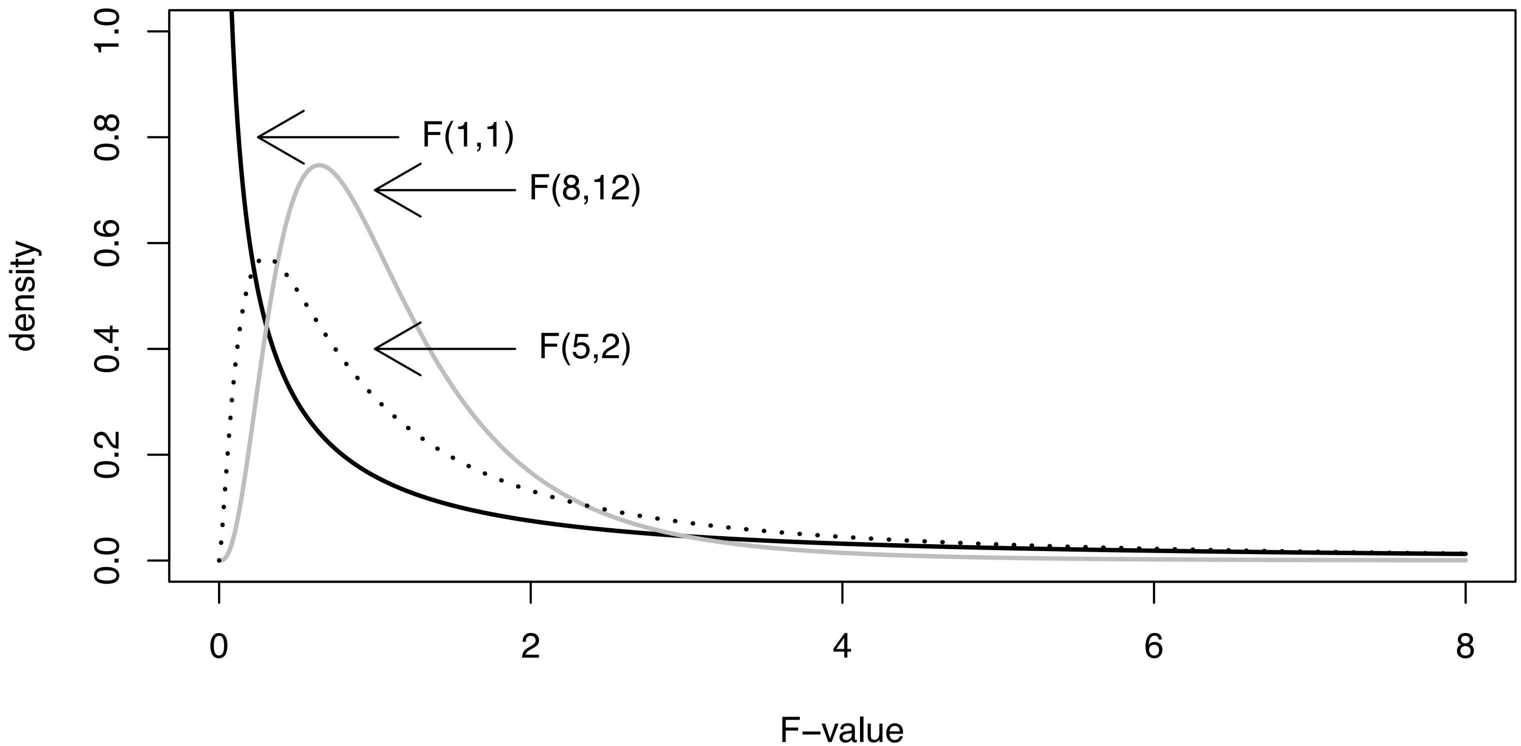
Distribution du F



Distribution du F



Distribution du F



Elle dépend des nombres de degrés de liberté associés à ν_I et ν_E

ANOVA 1 facteur

Formalisation

Système de notation / Décomposition de la variance / Test de H0

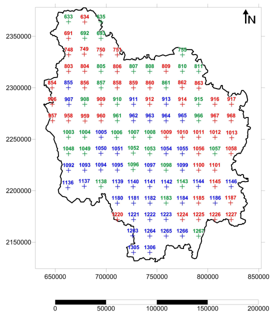
Système de Notation

Groupes

obs.	1	2	$\cdots$	j	$\cdots$	k
1	$x_{1,1}$	$x_{1,2}$	$\cdots$	$x_{1,j}$	$\cdots$	$x_{1,k}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$x_{2,j}$	$\cdots$	$x_{2,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	$x_{i,1}$	$x_{i,2}$	$\cdots$	$x_{i,j}$	$\cdots$	$x_{i,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n_j	$x_{n_1,1}$	$x_{n_2,2}$	$\cdots$	$x_{nj,j}$	$\cdots$	$x_{n_1,k}$
Effectifs	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k
Totaux	T_1	T_2	$\cdots$	T_j	$\cdots$	T_k
Moyennes	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_j$	$\cdots$	$\bar{x}_k$

ANOVA 1 facteur – Formalisation / Système de notation

<u>Exemple</u> : Comparaison de la composition isotopique du carbone de 3 types d'occupation du sol en Bourgogne



	Prairie	Champs	Forêt	
	-28.15	-26.13	-26.11	
	-28.54	-27.27	-27.04	
	-27.51	-27.54	-26.50	
	-27.76	-27.00	-27.01	
	-28.03	-26.42	-26.94	
	-27.71	-27.45	-27.65	
	-28.25	-27.62	-27.22	
	-27.88	-27.30	-26.16	
n_j	8	8	8	
T_j	-223.83	-216.73	-214.63	**-655.19**
$\bar{x}_j$	-27.98	-27.09	-27.83	**-27.30**

Calcul des variances intra- et inter-groupes

obs.	Groupes 1	2	$\cdots$	j	$\cdots$	k
1	$x_{1,1}$	$x_{1,2}$	$\cdots$	$x_{1,j}$	$\cdots$	$x_{1,k}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$x_{2,j}$	$\cdots$	$x_{2,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	$x_{i,1}$	$x_{i,2}$	$\cdots$	$x_{i,j}$	$\cdots$	$x_{i,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n_j	$x_{n_1,1}$	$x_{n_2,2}$	$\cdots$	$x_{nj,j}$	$\cdots$	$x_{n_1,k}$
Effectifs	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k
Totaux	T_1	T_2	$\cdots$	T_j	$\cdots$	T_k
Moyennes	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_j$	$\cdots$	$\bar{x}_k$

<u>Dispersion totale :</u>

$$SCT = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{X}\right)^2$$

<u>ou</u>

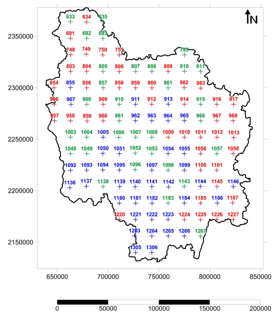
$$SCT = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}^2 - \frac{T^2}{n}\right)$$

<u>Variance totale :</u>

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1} = \frac{SCT}{n-1}$$

ANOVA 1 facteur – Formalisation / Décomposition de la variance

Exemple : Comparaison de la composition isotopique du carbone de 3 types d'occupation du sol en Bourgogne



Prairie	Champs	Forêt
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

$$\bar{X} = -27.30$$

Dispersion totale

$$SCT = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{X} \right)^2$$

Calcul de la variance intra-groupe

Dispersion intra-groupe:

	Groupes					
obs.	1	2	$\cdots$	j	$\cdots$	k
1	$x_{1,1}$	$x_{1,2}$	$\cdots$	$x_{1,j}$	$\cdots$	$x_{1,k}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$x_{2,j}$	$\cdots$	$x_{2,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	$x_{i,1}$	$x_{i,2}$	$\cdots$	$x_{i,j}$	$\cdots$	$x_{i,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n_j	$x_{n_1,1}$	$x_{n_2,2}$	$\cdots$	$x_{nj,j}$	$\cdots$	$x_{n_1,k}$
Effectifs	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k
Totaux	T_1	T_2	$\cdots$	T_j	$\cdots$	T_k
Moyennes	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_j$	$\cdots$	$\bar{x}_k$

$$SCE = \sum_{i=1}^{n_1} \left(x_{i1} - \bar{x}_1\right)^2 + \sum_{i=1}^{n2} \left(x_{i2} - \bar{x}_2\right)^2 + \ldots + \sum_{i=1}^{nk} \left(x_{ik} - \bar{x}_k\right)^2$$

$$SCE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$$

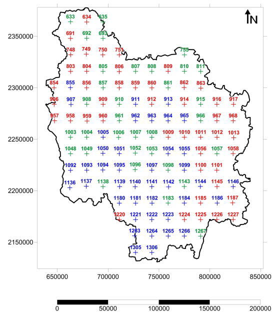
ou

$$SCE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^{k} \frac{T_j^2}{n_j}$$

Variance intra-groupe : $$V_E = \frac{SCE}{\nu_E} = \frac{SCE}{n-k}$$

ANOVA 1 facteur – Formalisation / Décomposition de la variance

<u>Exemple</u> : Comparaison de la composition isotopique du carbone de 3 types d'occupation du sol en Bourgogne



Prairie	Champs	Forêt
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
$\bar{x}_1 = -27.98$	$\bar{x}_2 = -27.09$	$\bar{x}_3 = -26.83$
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

<u>Dispersion intra-groupe</u>

$$SCE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$$

Calcul de la variance inter-groupe

Dispersion inter-groupe:

	Groupes					
obs.	1	2	$\cdots$	j	$\cdots$	k
1	$x_{1,1}$	$x_{1,2}$	$\cdots$	$x_{1,j}$	$\cdots$	$x_{1,k}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$x_{2,j}$	$\cdots$	$x_{2,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	$x_{i,1}$	$x_{i,2}$	$\cdots$	$x_{i,j}$	$\cdots$	$x_{i,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n_j	$x_{n_1,1}$	$x_{n_2,2}$	$\cdots$	$x_{nj,j}$	$\cdots$	$x_{n_1,k}$
Effectifs	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k
Totaux	T_1	T_2	$\cdots$	T_j	$\cdots$	T_k
Moyennes	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_j$	$\cdots$	$\bar{x}_k$

$$SCI = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{X}\right)^2$$

ou

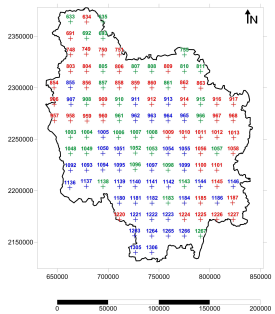
$$SCI = \sum_{j=1}^{k} \frac{T_j^2}{n_j} - \frac{T^2}{n}$$

Variance inter-groupe :

$$V_I = \frac{SCI}{\nu_I} = \frac{SCI}{k-1}$$

ANOVA 1 facteur – Formalisation / Décomposition de la variance

Exemple : Comparaison de la composition isotopique du carbone de 3 types d'occupation du sol en Bourgogne



Prairie	Champs	Forêt
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

$$\bar{X} = -27.30$$

$$\bar{x}_1 = -27.98 \qquad \bar{x}_2 = -27.09 \qquad \bar{x}_3 = -26.83$$

Dispersion inter-groupe

$$SCI = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{X}\right)^2$$

Tableau d'ANOVA

Source	**SS** (dispersion)	**ddl**	**MS** (Variance)
Inter-groupe (Facteur)	$\sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{X}\right)^2$	$k-1$	$\sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{X}\right)^2 \Big/ (k-1)$
Résiduelle (Erreur; Intra-group)	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$	$n-k$	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2 \Big/ (n-k)$
Totale	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{X}\right)^2$	$n-1$	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{X}\right)^2 \Big/ (n-1)$

Rappel :
Nombre de **degrés de liberté** associé à 1 calcul est le nombre de ses composantes indépendantes, i.e. le nombre de composantes de base du calcul moins le nombre de relations (paramètres) qui lient celles-ci

Tableau d'ANOVA

Source	**SS** (dispersion)	**ddl**	**MS** (Variance)
Inter-groupe (Facteur)	$SCI = 5.81$	$k-1=2$	$MS_{InterGpe} = 2.90$
Résiduelle (Erreur; Intra-group)	$SCE = 4.83$	$n-k=21$	$MS_{Erreur} = 0.23$
Totale	$SCT = 11.10$	$n-1=23$	$s^2 = 0.46$

Tableau d'ANOVA

Source	**SS** (dispersion)	**ddl**	**MS** (Variance)
Inter-groupe (Facteur)	$SCI = 5.81$	$k-1=2$	$MS_{InterGpe} = 2.90$
Résiduelle (Erreur; Intra-group)	$SCE = 4.83$	$n-k=21$	$MS_{Erreur} = 0.23$
Totale	$SCT = 11.10$	$n-1=23$	$s^2 = 0.46$

Important !!

$$SCI + SCE = SCT$$

$$\nu_I + \nu_E = \nu_T$$

MAIS !!

$$MS_I + MS_E \neq s^2$$

Expected Mean Squares

Source	MS	$E[MS_{\text{Facteur à effet fixe}}]$
Inter-groupe (Facteur)	$MS_I = SCI/(k-1)$	$\sigma_\varepsilon^2 + \sum_{j=1}^{k} n_j \frac{\alpha_j^2}{k-1}$
Résiduelle	$MS_E = SCE/(n-k)$	σ_ε^2

ANOVA 1 facteur – Formalisation / Test de H0

Hypothèses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu \\ H_1 : \text{tous les moyennes} \\ \text{sont pas égales} \end{cases} \qquad \alpha_i = \mu_i - \mu : \begin{cases} H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0 \\ H_1 : \text{tous les niveaux ont} \\ \text{sont pas le même effe} \end{cases}$$

Si 1) <u>normalité des données</u> et 2) <u>homogénéité des variances</u> entre les *k groupes*

Alors, si H0 est vraie (même moyenne dans tous les populations), la variance globale σ^2 de la population peut être estimée de 2 façons :

1) Variance intra-groupe (MS_E) = moyenne pondérée des variances des *k* groupes

2) Variance inter-groupe (MS_I)

Rapport des variances *F*

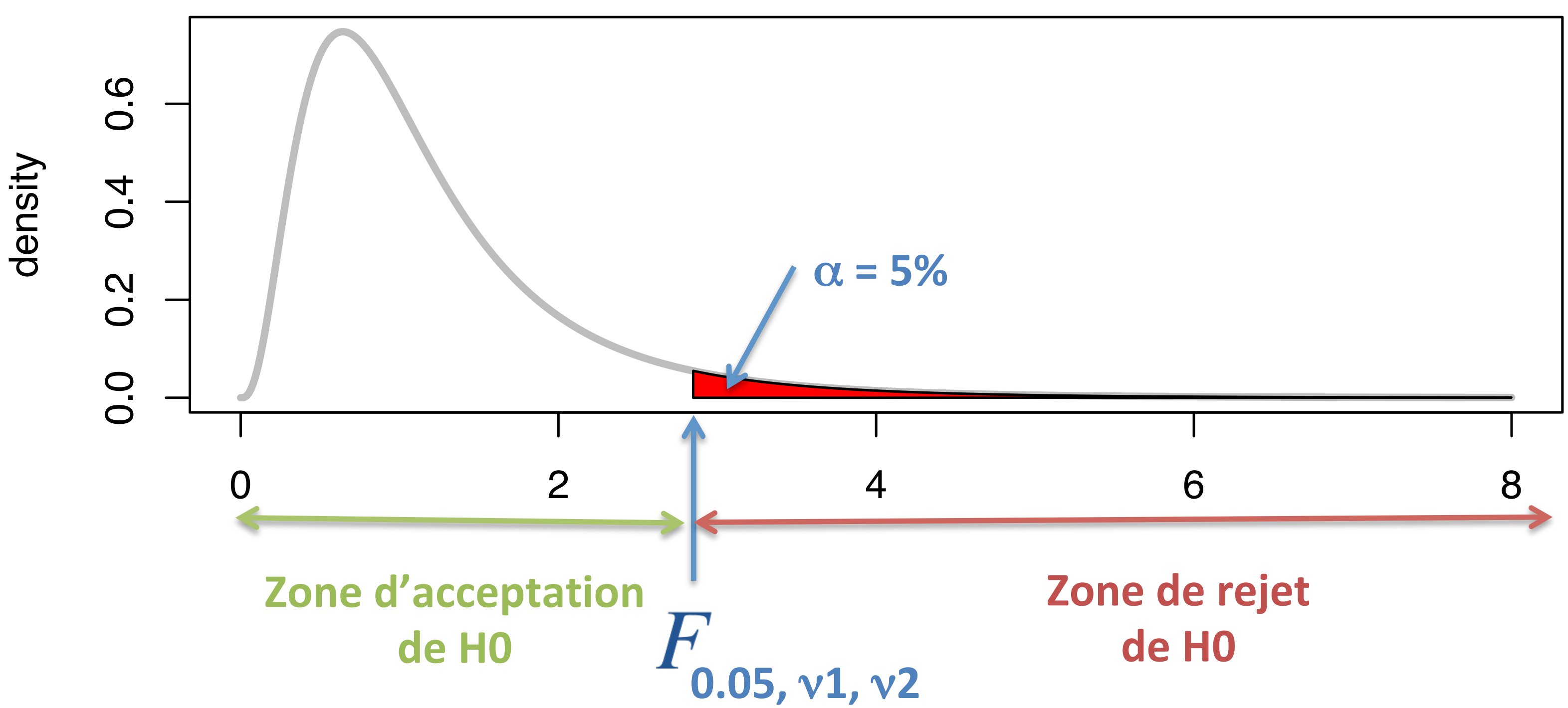
Ainsi, puisque $\mathrm{E}\left[MS_I\right] = \sigma_\varepsilon^2 + \sum_{j=1}^{k} n_j \frac{\alpha_j^2}{k-1}$ et $\mathrm{E}\left[MS_E\right] = \sigma_\varepsilon^2$

si H0 vraie → $MS_I = MS_E$ $\qquad F = \frac{MS_I}{MS_E} \approx 1$

si H0 fausse → $MS_I > MS_E$ $\qquad F = \frac{MS_I}{MS_E} > 1$

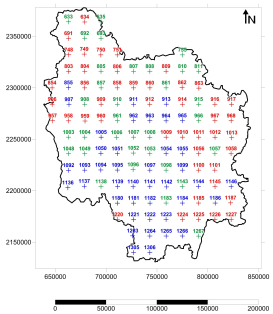
F suit une loi de Fisher-Snedecor à ν_I et ν_E degrés de liberté

Distribution du F et zone de rejet de H0



ANOVA 1 facteur – Formalisation / Test de H0

Exemple : Comparaison de la composition isotopique du carbone de 3 types d'occupation du sol en Bourgogne



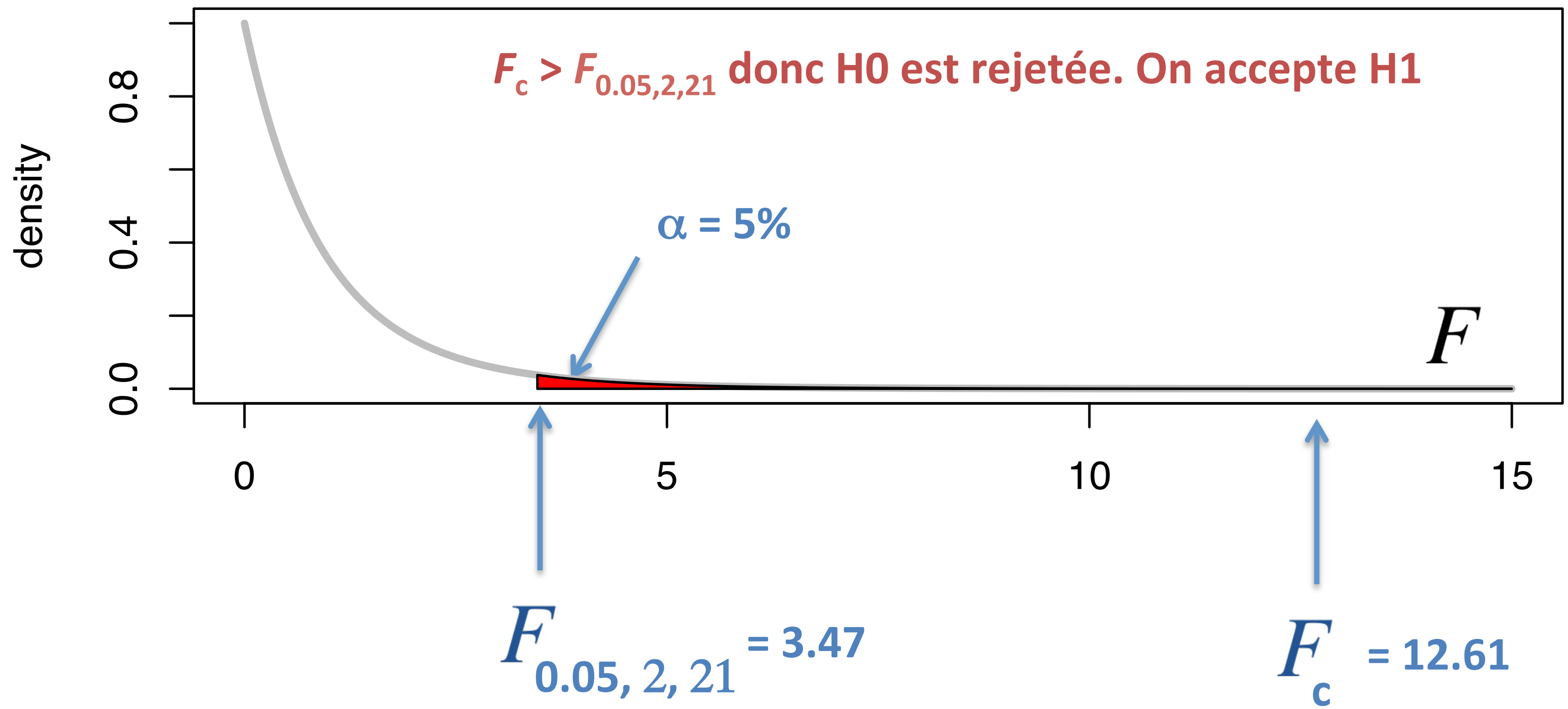
Prairie	Champs	Forêt
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
-27.76	-27.00	-27.01
-28.03	-26.42	-26.94
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

Calcul du F-ratio

Source	**SS** (dispersion)	**ddl**	**MS** (Variance)
Inter-groupe (Facteur)	$SCI = 5.81$	$k-1=2$	$MS_{InterGpe} = 2.90$
Résiduelle (Erreur; Intra-group)	$SCE = 4.83$	$n-k=21$	$MS_{Erreur} = 0.23$
Totale	$SCT = 11.10$	$n-1=23$	$s^2 = 0.46$

$$F_c = \frac{MS_I}{MS_E} = \frac{2.90}{0.23} = 12.61$$

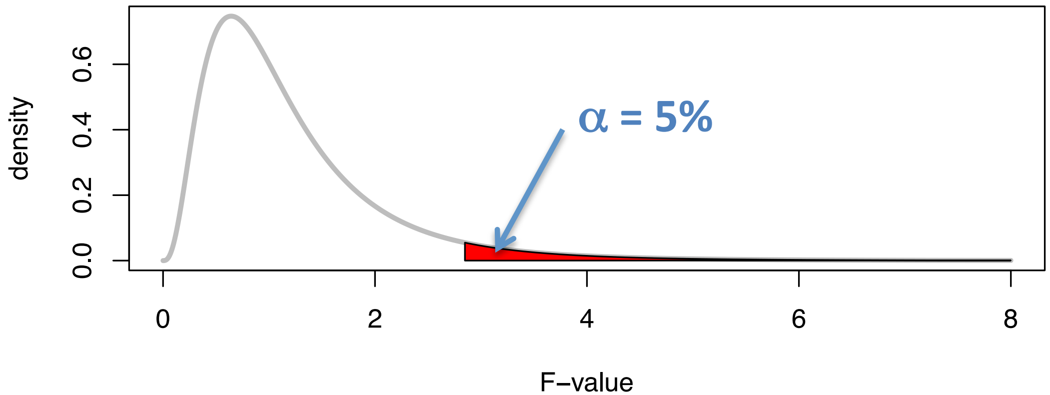
Test de H0



Les compositions isotopiques du carbone des 3 types d'occupation du sol en Bourgogne <u>ne sont pas toutes égales</u>

Remarques :

1) L'ANOVA est toujours un test unilatéral



2) L'ANOVA n'est pas un test de comparaison des variances

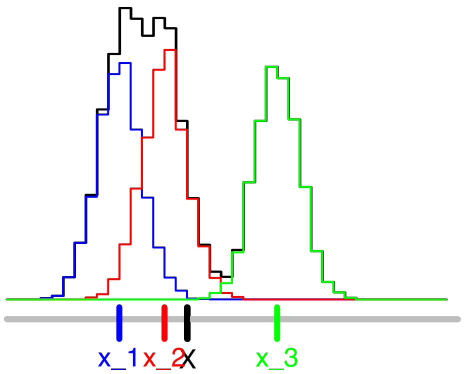
$$H_0 \text{ neq} : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

ANOVA 1 facteur

Conditions d’application

ANOVA 1 facteur – Conditions d'application

Avant de faire une ANOVA…

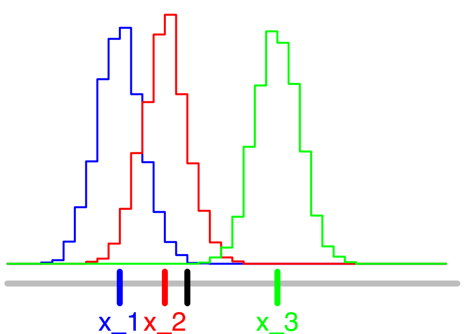


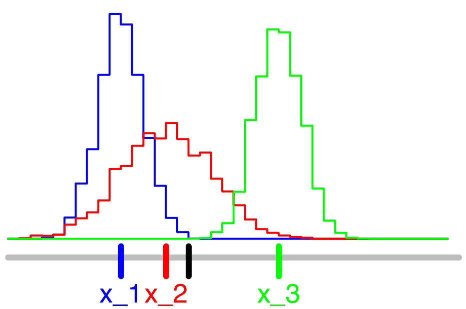
1) Variable dépendante quantitative

2) Indépendance des observations

3) Distribution normales à l'intérieur des k groupes

4) Variances des k groupes équivalentes (homoscédasticité)





Avant de faire une ANOVA…

Conditions pas toujours toutes vérifiées, **MAIS** :

ANOVA robuste :

- Variances hétérogènes, mais n_j égaux ou proches
  sinon, probabilité erreur type I $\neq \alpha$

- Non normalité, mais grands échantillons
  sinon, puissance du test modifié

Alternatives :

- Transformation des données
- Procédure modifiée d'ANOVA
- Test non-paramétrique (Kruskal-Wallis)

Plusieurs types d'ANOVA

Modèle I

Modèle avec effet fixe

"fixed-effects model"

Facteur contrôlé (fixe)

Niveaux du facteur complètement contrôlés

Choix des seuls niveaux d'intérêts

Modèle II

Modèle avec effet aléatoire

"random-effects model"

Facteur aléatoire

Niveaux du facteur choisis au hasard

ANOVA 1 facteur

Tests *a posteriori*

Une fois H0 rejetée…

ANOVA
$$\begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu \\ \mathrm{H}_1 : \text{tous les moyennes ne sont pas égales} \end{cases}$$

H0 acceptée

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3$$

H0 rejetée

$$\mathrm{H}_0 : \mu_1 = \mu_2 \neq \mu_3$$

$$\mathrm{H}_0 : \mu_1 \neq \mu_2 = \mu_3$$

$$\mathrm{H}_0 : \mu_1 \neq \mu_2 \neq \mu_3$$

$$\mathrm{H}_0 : \mu_1 = \mu_3 \neq \mu_2$$

Tests *a posteriori* :

- Test de Tukey
- Test de Newman-Keuls
- Test de Scheffé
- Test de Duncan…

Test de Tukey

Test de Tukey $\begin{cases} \mathrm{H}_0 : \mu_A = \mu_B \\ \mathrm{H}_1 : \mu_A \neq \mu_B \end{cases}$

1) On classe les moyennes des *k* groupes par ordre croissant

2) On calcule des différences de moyennes, en commençant par la plus grande et la plus petite, la plus grande et la 2nd plus petite, …, la 2nd plus grande et la plus petite, la 2nd plus grande et la 2nd plus petite,…

3) On calcule la variable auxiliaire *q* :

$$q = \frac{\bar{X}_B - \bar{X}_A}{SE} \text{ avec } SE = \sqrt{\frac{MS_E}{n}} \text{ ou } SE = \sqrt{\frac{MS_E}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} \text{ si } n_A \neq n_B$$

Test de Tukey

Test de Tukey $\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases}$

1) On classe les moyennes des *k* groupes par ordre croissant

2) On calcule des différences de moyennes, en commençant par la plus grande et la plus petite, la plus grande et la 2^{nd} plus petite, …, la 2^{nd} plus grande et la plus petite, la 2^{nd} plus grande et la 2^{nd} plus petite,…

3) On calcule la variable auxiliaire *q* :

4) Compare *q* à la valeur seuil :$q_{\alpha,\nu,k}$: $\text{si } q \geq q_{\alpha,\nu,k} \Rightarrow H_0 \text{ est rejetée}$

α *:* seuil de significativité = proba. de commettre au moins 1 erreur de type I au cours des comparaisons multiples

ν *:* ddl associé à MS_E dans l'analyse de variance

k : nombre total de moyennes comparées

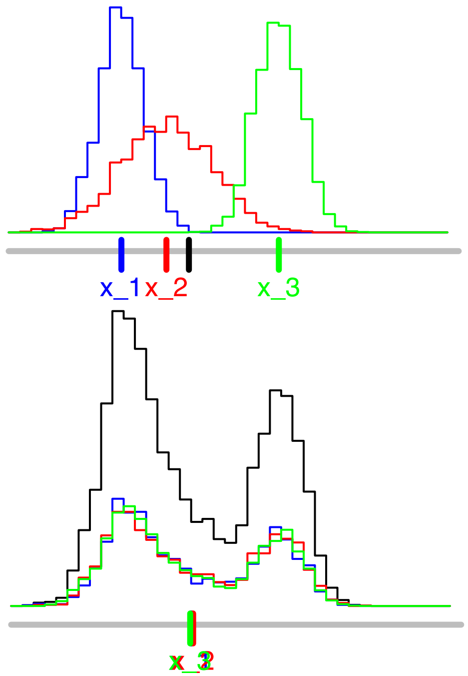
ANOVA 1 facteur

et si les postulats ne sont pas vérifiés…

ANOVA 1 facteur – Alternatives non-paramétriques

1 – Transformation des données

2 – Tests de permutation



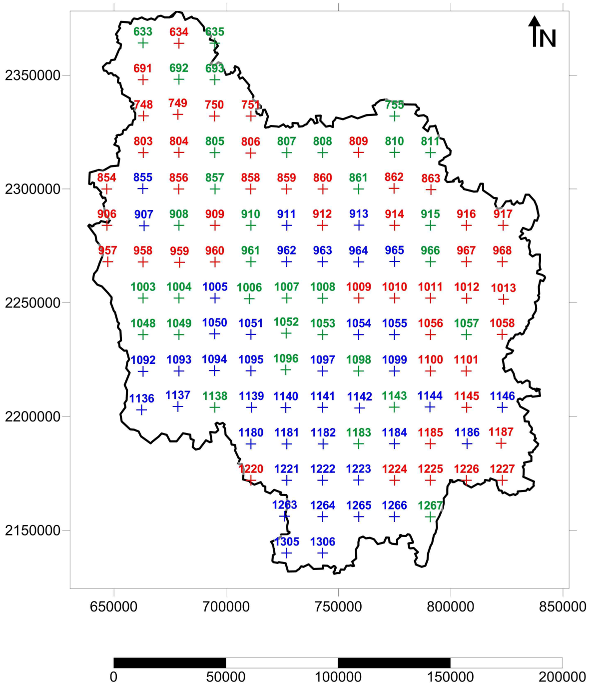
3 – Tests non-paramétriques : Kruskal-Wallis

KRUSKAL-WALLIS

Comparaison des rangs

ANOVA 1 facteur – Test de Kruskal Wallis

Exemple : Effet de la nature de l'occupation des sols en Bourgogne sur leur composition isotopique ($\delta^{13}C$)



90 sols échantillonnés

3 types d'occupation : Prairie, Champs, Forêt

Prairie	Champs	Forêt
-28.15	-26.13	-26.11
-28.54	-27.27	-27.04
-27.51	-27.54	-26.50
-27.76	-27.00	-27.01
-28.03	-26.42	-26.94
-27.71	-27.45	-27.65
-28.25	-27.62	-27.22
-27.88	-27.30	-26.16

distribution des données non normale

<u>MAIS</u> **et/ou**

variances hétérogènes

Les hypothèses

H_0 : Il n'y a aucune différence entre les *k* groupes

<u>ou</u> les k groupes sont issus de la même population statistique

<u>ou</u> les k groupes constituent un ensemble homogène

H_1 : Il existe au moins une différence entre 2 groupes

<u>ou</u> les k groupes ne proviennent pas de la même population statistique

<u>ou</u> les k groupes constituent un ensemble hétérogène

<u>Important</u> ! :

$$\cancel{\begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu \\ \mathrm{H}_1 : \text{tous les moyennes ne sont pas égales} \end{cases}}$$

Comparaison de plusieurs échantillons indépendants

Variable quantitative
normalité des données
ET
Homogénéité des variances

ANOVA

H_0 acceptée

Toutes les moyennes sont égales

H_0 rejetée

Test *a poteriori*
(ex. Tukey)

Variable quantitative
non-normalité des données
ET/OU
Hétérogénéité des variances
Variable semi-quantitative

Test de Kruskal-Wallis

H_0 acceptée

Aucune différence entre les *k* groupes

H_0 rejetée

Test *a poteriori*

Quels sont les groupes qui diffèrent