

L'Analyse de Variance

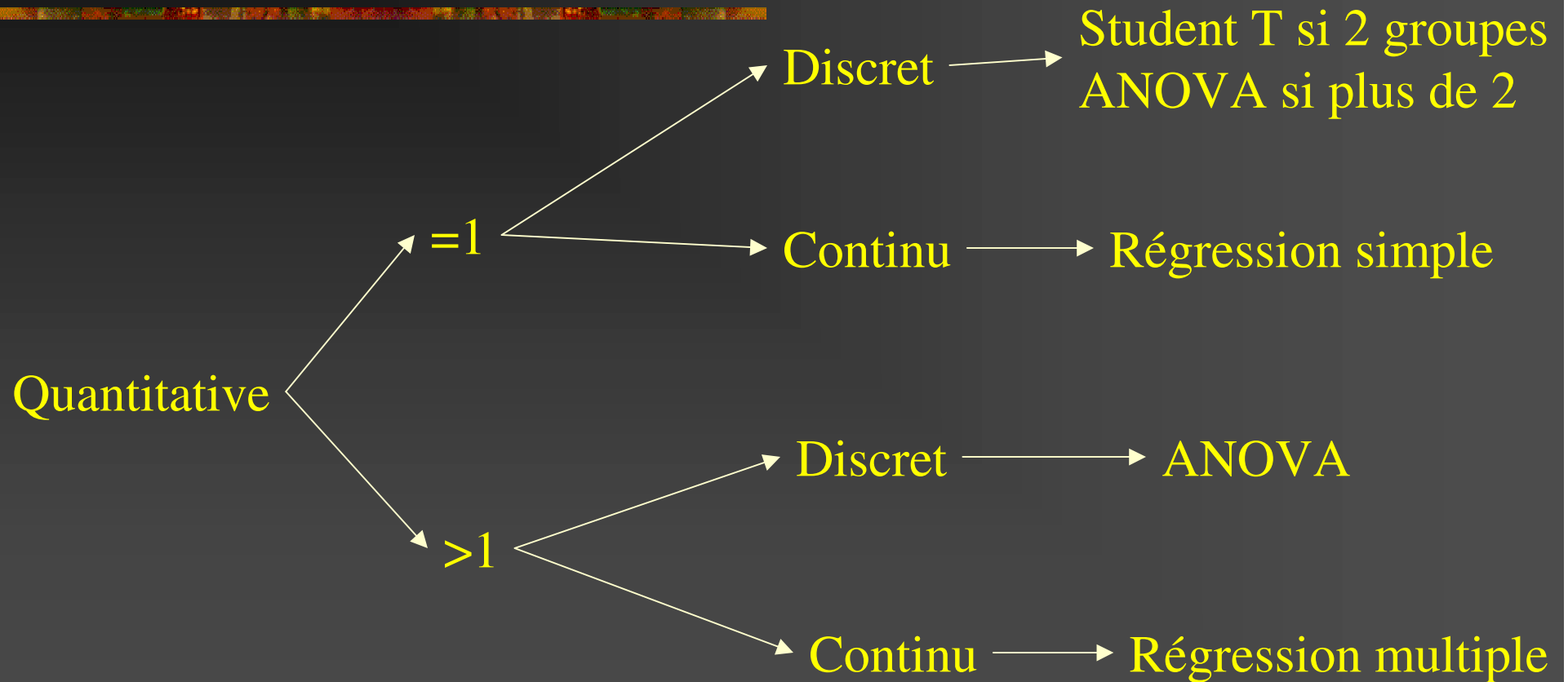
ANOVA

Type de Variable

Nombre de facteurs

Type de facteurs

Type de test

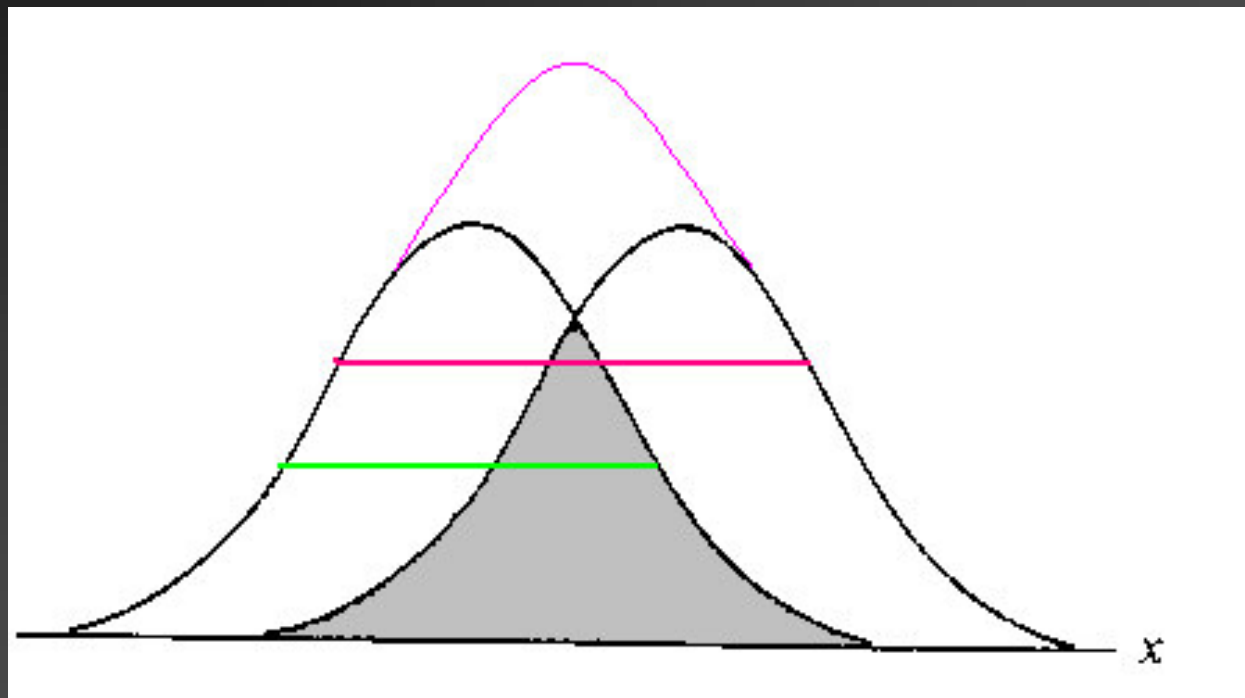


Quand ?

Objectifs de l'ANOVA:

1. Comparer plusieurs moyennes
 2. Détecter les mélanges de distributions
 3. Construire un plan d'expérience
-

Rapport de variances



Basée sur un quotient

$$= \frac{\text{Variance inter-groupes}}{\text{Variance intra-groupes}}$$

Cas particulier à 1 facteur et n égaux.

Dosages en mEq/g	Individu A	Individu B	Individu C	Individu D
Prèlev. 1	15	25	17	10
Prèlev. 2	9	21	23	13
Prèlev. 3	14	19	20	16

Variance des moyennes

$$S^2_{moyennes} = \frac{1}{a-1} \sum_{i=0}^a (\bar{y}_i - \bar{y})^2$$

avec

$$\bar{y} = \frac{1}{a \cdot n} \sum_i^a \sum_j^n y_{ij}$$

$$DDL = a - 1$$

Variance Inter

15	25	17	10
9	21	23	13
14	19	20	16

Comme

$$s^2_{\bar{Y}} = \frac{s^2}{n}$$

et ici $n=3$

$$s^2_{inter} = 3 \times (12.66 + 21.66 + 20 + 13) = 196.33$$

$$DDL1 = 4 - 1$$

Variance intra-groupes

15	25	17	10
9	21	23	13
14	19	20	16

$$S_{\text{intra}}^2 = \frac{1}{a(n-1)} \sum_i^a \sum_j^n (y_{ij} - \bar{y}_i)^2$$

$$= 75.33$$

avec a groupes
et n individus par groupe

$$DDL2 = a(n-1) = 8$$

Carrés moyens et F

15	25	17	10
9	21	23	13
14	19	20	16

Source de variation	Formule	DDL	SDC	Carré moyen	F
Inter groupes	$n \cdot \sum (\bar{y} - \bar{y})^2$	a-1=3	196.33	65.44	6.94
Intra = erreur	$\sum \sum (y - \bar{y})^2$	a(n-1) = 8	75.33	9.41	
Totale	$\sum (y - \bar{y})^2$	an-1 = 11	271.66		

Commentaires

1. Les variations (sommes des carrés) sont additives et donc utilisées de préférence à la variance
2. Les DDL sont aussi additifs
3. La variance intra-groupe est aussi appelée erreur par rapport à l'équation du modèle:

$$y_{ij} = \mu + A_i + \varepsilon_{ij}$$

Les hypothèses

H0 : Pas de différence entre individus

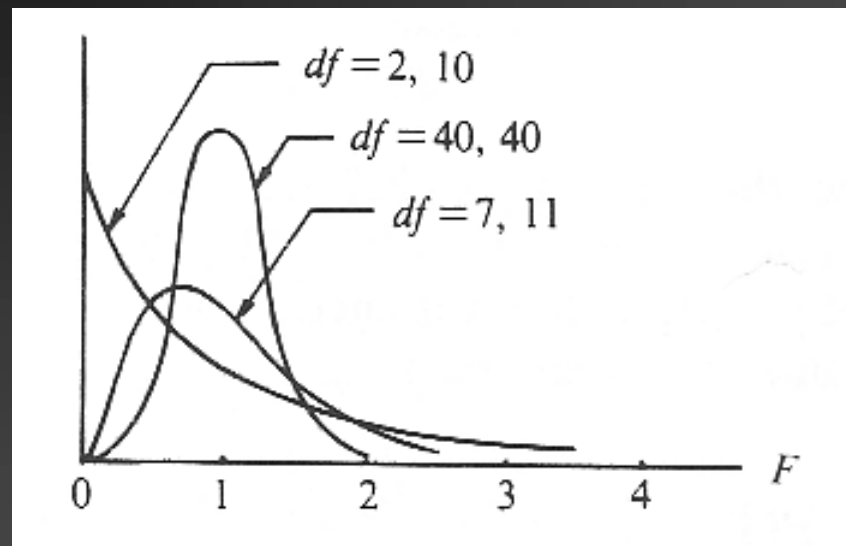
H1 : Il existe une différence

Le F calculé est égal à 6.94

Si $6.94 < F$ théorique au seuil choisi on accepte H0

Si $6.94 > F$ théorique on rejette H0

La distribution F

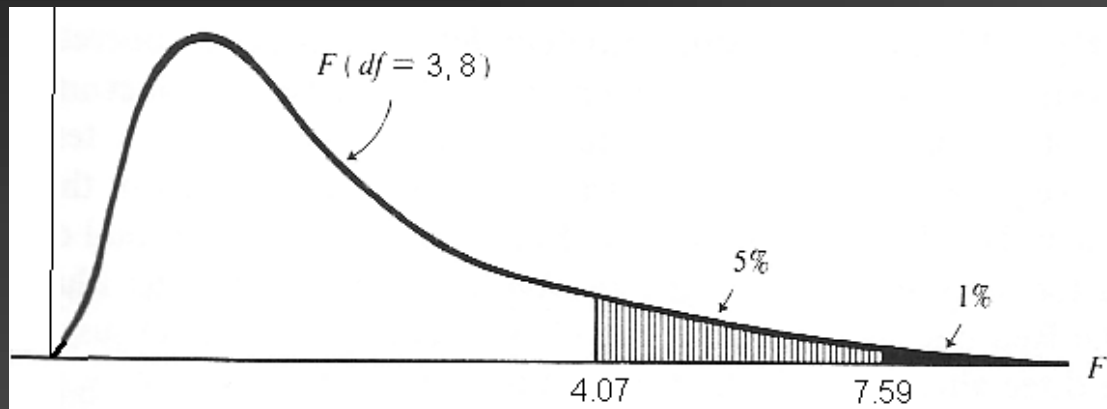


Distribution très polyvalente à 2 DDL. Le X^2 , le t et la loi normale sont des cas particuliers de la distribution F

$$F_{[x, \infty]} = X^2_{[x]}$$

Le F théorique

DDL	1	2	3	4	5
1	161.45	199.50	215.71	224.58	230.16
2	18.51	19.00	19.16	19.25	19.30
3	10.13	9.55	9.28	9.12	9.01
4	7.71	6.94	6.59	6.39	6.26
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97
8	5.32	4.46	4.07	3.84	3.69
9	5.12	4.26	3.86	3.63	3.48
10	4.96	4.10	3.71	3.48	3.33



$$F_{0.05}[3,8]=4.07$$

< 6.94 on rejette H0

Conditions pour l'ANOVA

1. Sélection aléatoire des spécimens
 2. Observations indépendantes
 3. Homogénéité des variances
 4. Normalité de la distribution
-

La variance ajoutée

$$y_{ij} = \mu + \sigma_i + \varepsilon_j$$

Valeur pour
l'individu j du
groupe i

Grande
moyenne

Variance
ajoutée liées
au groupe i

Variabilité
individuelle

Modèle 1 ou modèle 2 ?

Modèle 1 : La variance ajoutée est contrôlée par l'expérimentateur (ex: traitements)

Modèle 2 : La variance ajoutée est liée à une variable aléatoire (ex: le sexe)

Exemple général

Taille (en microns) de tiques provenant de 4 rongeurs

Lapin 1	Lapin 2	Lapin 3	Lapin 4
380	350	354	376
376	356	360	344
360	358	362	342
368	376	352	372
372	338	366	374
366	342	372	360
374	366	362	
382	350	344	
	344	342	
	364	358	
		351	
		348	
		348	

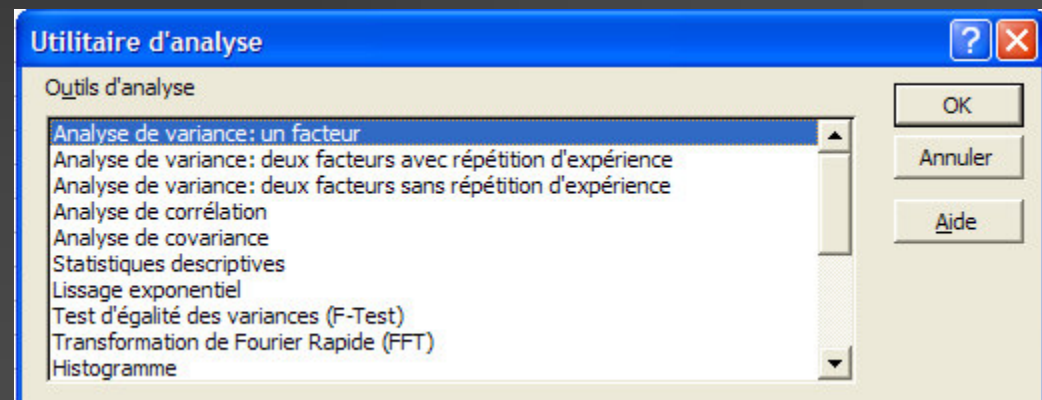
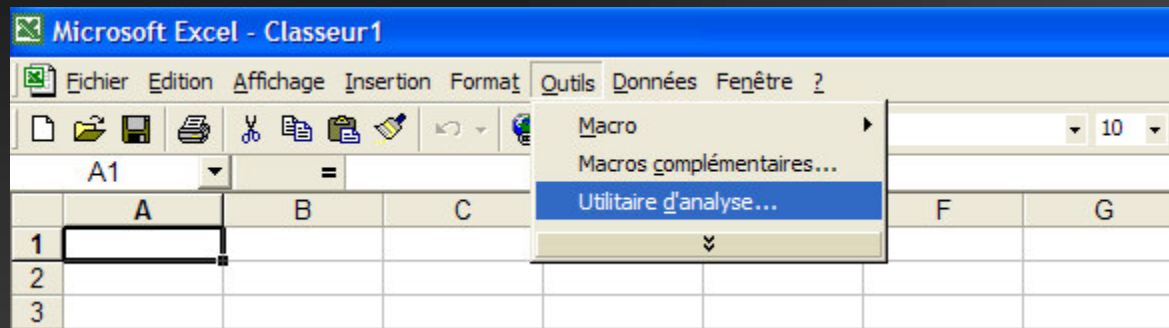
Deux présentations
d'un cas très général

380	1
376	1
360	1
368	1
372	1
366	1
374	1
382	1
350	2
356	2
358	2
376	2
338	2
342	2
366	2
350	2
344	2
364	2
354	3
360	3
362	3
352	3
366	3
372	3
362	3
344	3
342	3
358	3
351	3
348	3
348	3
376	4
344	4
342	4
372	4
374	4
360	4

Partition de la variance

Source de variation	DDL	SDC	Carré moyen	F
Inter groupes	3	a	a/3	$(a/3)/(b/33)$
Intra = erreur	33	b	b/33	
Totale	36	a+b		

Les outils Excel



Résultats

ANALYSE DE VARIANCE						
<i>Source des variations</i>	<i>SDC</i>	<i>DDL</i>	<i>Carrés moyens</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
Entre Groupes	1807.72717	3	602.5757219	5.26336298	0.0044452	2.891567874
Intra Groupes (erreur)	3778.00256	33	114.4849262			
Total	5585.72973	36				

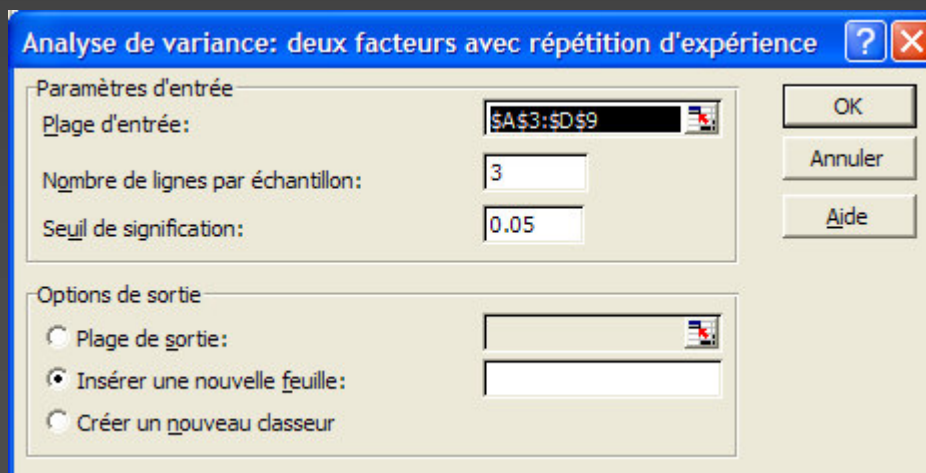
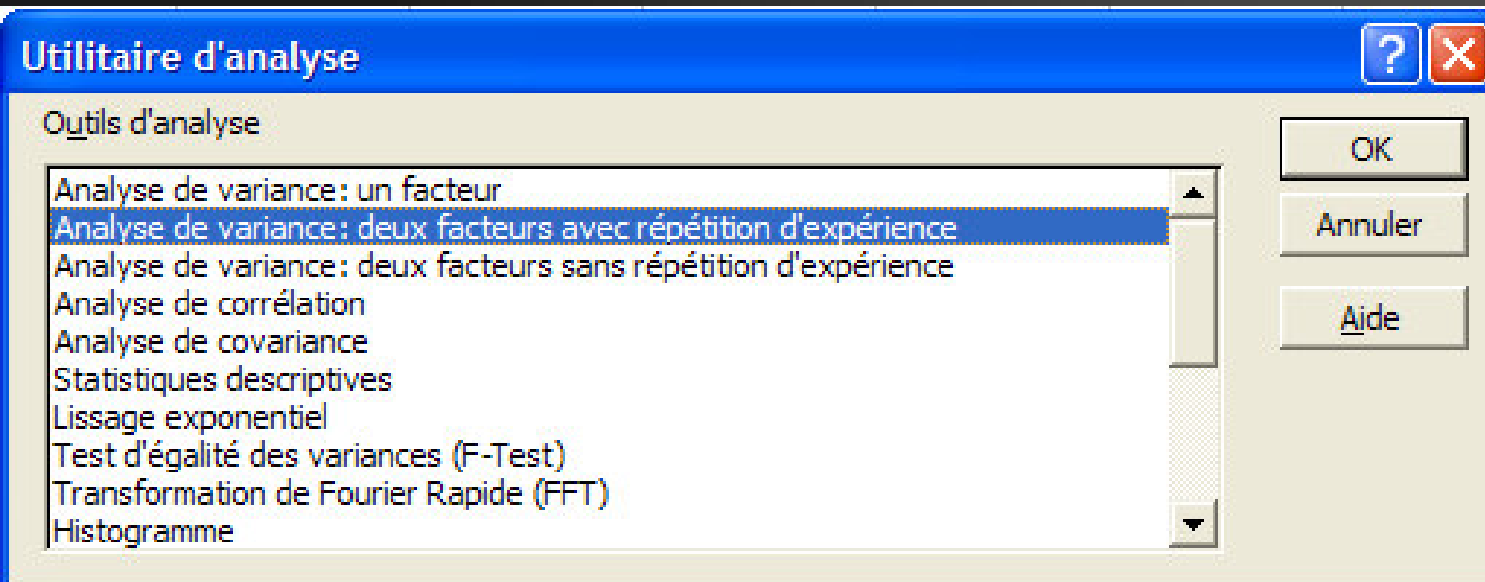
Comme $5.26 > 2.89$ on rejette H_0 selon laquelle l'hôte n'aurait pas d'influence sur la taille du parasite

ANOVA à deux facteurs

Effets de
(1) l'ensoleillement et des
(2) quantités d'engrais sur
la récolte

Soleil / Engrais	Peu	Moyen	Beaucoup
Grand soleil	5	15	21
	10	22	29
	8	18	25
Petit soleil	6	25	55
	9	32	60
	12	40	48

Avec répétition d'expérience



Partition des variations

Source	Formule	DDL
Entre lignes (soleil)	$n \cdot \sum (\bar{l} - \bar{y})^2$	L-1=1
Entre colonnes (engrais)	$n \cdot \sum (\bar{c} - \bar{y})^2$	C-1=2
Interaction	$n \cdot \sum (\bar{y} - \bar{c} - \bar{l} + \bar{y})^2$	(L-1)(C-1)=2
Erreur	$\sum \sum (y - \bar{y})^2$	LC(n-1)=12
Totale	$\sum (y - \bar{y})^2$	LCN-1=17

Résultats Excel

ANALYSE DE VARIANCE						
<i>Source des variations</i>	<i>Somme des carrés</i>	<i>DDL</i>	<i>Moy. des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>F critique</i>
Échantillon	997.5555556	1	997.5555556	43.9022005	2.4443E-05	4.747221283
Colonnes	2952.444444	2	1476.222222	64.9682152	3.6519E-07	3.885290312
Interaction	589.7777778	2	294.8888889	12.9779951	0.00099863	3.885290312
A l'intérieur du groupe	272.6666667	12	22.72222222			
Total	4812.444444	17				

L'interaction

Non additivité
des effets



Potentialisation

Antagonisme

ANOVA à 2 facteurs sans répétition d'expérience

Fusionnés

Facteur 1

Facteur 2

Interaction

Erreur

Total

Exemple des T° du Rot Lake

Profondeur	Jour 1	Jour 2	Jour 3	Jour 4
0	23.8	24	24.6	24.8
1	22.6	22.4	23.2	23.2
2	22.2	22.1	22.1	22.2
3	21.2	21.8	21	21.2
4	18.4	19.3	19	18.8
5	13.5	14.4	14.2	13.8
6	9.8	9.9	10.4	9.6
9	6	6	6.3	6.3
12	5.8	5.9	6	5.8
15	5.6	5.6	5.5	5.6

Facteur A : le jour

Facteur B : la profondeur

Une seule observation par intersection

Résultats Excel

Analyse de variance: deux facteurs sans répétition d'expérience

Paramètres d'entrée

Plage d'entrée:

Intitulé présent

Seuil de signification:

Options de sortie

Plage de sortie:

Insérer une nouvelle feuille:

Créer un nouveau classeur

OK
Annuler
Aide

ANALYSE DE VARIANCE						
<i>Source des variations</i>	<i>Somme des carrés</i>	<i>DDL</i>	<i>Moy. des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>F critique</i>
Lignes	2135.246	9	237.2495556	2172.91	3.1841E-36	2.25013252
Colonnes	0.857	3	0.285666667	2.61635	0.07148279	2.96034841
Erreur	2.948	27	0.109185185			
Total	2139.051	39				

Construire un plan d'expérience

Symétrie et température
chez les moustiques

[Mosquitoes\(AnovaModeles1et2\).xls](#)
