- Analyse de variance :
  test F global significatif
  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$ rejetée
  $\Leftrightarrow$ Il existe i et j tels que $\mu_i \neq \mu_j$

- Erreurs indépendantes, identiquement distribuées $N(0, \sigma^2)$

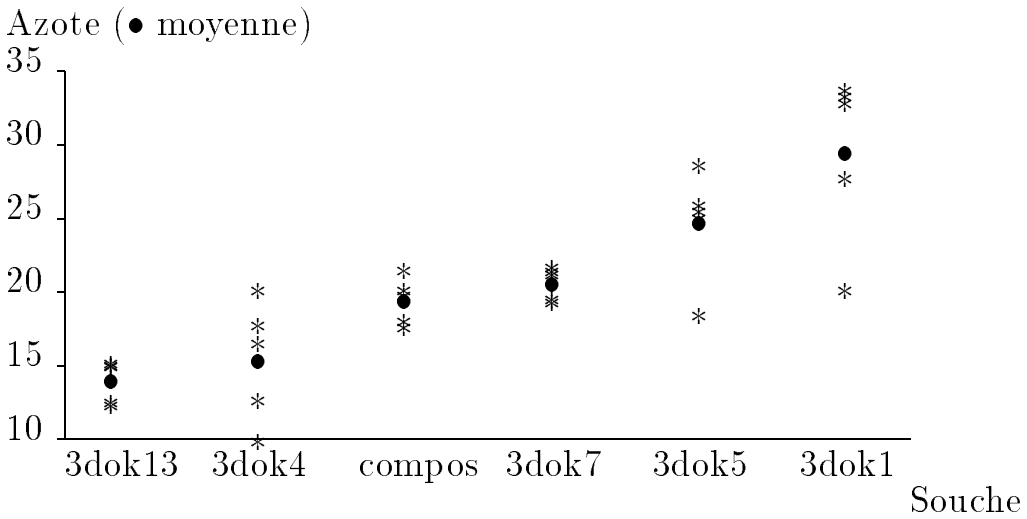
- Explorer les différences entre moyennes

- Former des groupes de moyennes homogènes

- Suite à une analyse de variance ayant déclaré le test global F significatif, l'hypothèse nulle est rejetée.
- Autrement dit, il existe des différences significatives entre les moyennes μ_i correspondant aux niveaux du facteur.
- On a vérifié (graphiquement) les postulats du modèle.
- Pour aller plus loin, que faire ?

INFLUENCE DE 6 SOUCHES DE RHIZOBIUM SUR LA QUANTITE D'AZOTE DE PLANTES DE TREFLE ROUGE

Souche	Azote					moyenne
3dok1	19.4	32.6	27.0	32.1	33.0	28.82
3dok5	17.7	24.8	27.9	25.2	24.3	23.98
3dok4	17.0	19.4	9.1	11.9	15.8	14.64
3dok7	20.7	21.0	20.5	18.8	18.6	19.92
3dok13	14.3	14.4	11.8	11.6	14.2	13.26
compos	17.3	19.4	19.1	16.9	20.8	18.70



Reprenons le problème sur un exemple.

- Facteur étudié : souche de rhizobium
- Variable mesurée : quantité d' azote de la plante.
- Les données sont constituées de 5 plantes pour 1 souche (5 répétitions par niveau du facteur).
- La colonne Moyenne contient la moyenne des 5 répétitions.

Le graphique représente les données individuelles Azote en fonction de la Souche ($*$) et les moyennes par souche ($\bullet$).

Table d'analyse de variance (sortie SAS)

Dependent Variable: AZOTE

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	5	847.05	169.41	14.37	0.0001
Error	24	282.93	11.79		
Total	29	1129.97			

$\boxed{\hat{\sigma}^2 = 11.79}$

Questions :

- Former des groupes de souches de moyennes homogènes ?
- Si "compos" est le témoin, quelles moyennes différent de la moyenne du témoin ?

On effectue une analyse de variance.

Table d'ANOVA : le test F est significatif.

Questions qui restent posées.

Pour comparer 2 moyennes,
méthode classique du t de Student

$$H_0 : \mu_i = \mu_j \text{ rejetée si } | \hat{\mu}_i - \hat{\mu}_j | \geq T(\alpha, \nu) \times s_d$$

α : risque de 1ère espèce
ν : degrés de liberté résiduels
T : quantile de la loi de Student
au seuil α avec ν degrés de liberté
s_d : écart-type de la différence de 2 moyennes
$\mu_i - \mu_j$
en cas d'effectifs égaux $s_d = \hat{\sigma}\sqrt{2/n}$
en cas d'effectifs inégaux $s_d = \hat{\sigma}\sqrt{1/n_i + 1/n_j}$

Rappel (pour formateur)

sous $H_0 : \mu_i - \mu_j = 0$

$$1 - \alpha = Proba(\mid \hat{\mu}_i - \hat{\mu}_j \mid < T(\alpha, \nu) \times s_d)$$

niveau de confiance du test (95 % par exemple).

$$\alpha = Proba(\mid \hat{\mu}_i - \hat{\mu}_j \mid \geq T(\alpha, nu) \times s_d)$$

risque de première espèce : probabilité de conclure que la différence est significative alors qu'elle ne l'est pas (5 % par exemple).

Application à l'exemple Azote / Rhizobium

Cas équirépété $n = 5$

Ecart-type de la différence $\mu_i - \mu_j$

$$s_d = \sqrt{2\hat{\sigma}^2/n} = \sqrt{2 \times 11.79/5} = 2.17$$

Degrés de liberté résiduels $\nu = 24$

Valeur de T au seuil $\alpha = 5\%$ avec $\nu = 24$ d.d.l.

$$T(0.05, 24) = 2.064$$

La différence entre 2 moyennes particulières est significative au seuil 5% lorsqu'elle dépasse

$$T(0.05, 24) \times s_d = 2.064 \times 2.17 = 4.48$$

Cette quantité qui sert de critère est la
Plus Petite Différence Significative (P.P.D.S.)

A FAIRE PLUTOT AU TABLEAU EN LAISSANT LE TRANSPARENT PRECEDENT.

Aller chercher $\hat{\sigma^2} = 11.79$ dans la table d'ANOVA.

Distribuer aux stagiaires une table de T.

P.P.D.S. et Risque global

I niveaux du facteur
$\Rightarrow$ I moyennes à comparer soit
$C_I^2 = \frac{I(I-1)}{2}$ comparaisons de 2 moyennes

α risque de 1ère espèce d'une comparaison
α_g risque de 1ère espèce global
(= probabilité de trouver au moins une différence significative alors qu'il n'y en a pas)
Avec la méthode de la P.P.D.S. :

I	C_I^2	α	α_g
2	1	0.05	0.05
5	10	0.05	0.29
10	45	0.05	0.63
15	105	0.05	0.83

$\boxed{\alpha_g \text{ est supérieur à } \alpha}$

$C_n^k = \frac{n!}{k!(n-k)!}$ nombre de combinaisons de k moyennes prises parmi n.

Risque de 1ère espèce α:
proba. de conclure la différence significative alors qu'elle ne l'est pas.

Risque global α_g :
proba. de trouver au moins une différence significative qui ne l'est pas (les valeurs telles que 0.29 sont obtenues par simulation).
Avec 5 moyennes, on a 29 chances sur 100 d'obtenir au moins une différence significative (sous H_0, alors qu'il n'y en a pas).

Autres méthodes

En contrôlant le risque global α_g,

si $\mid \hat{\mu}_i - \hat{\mu}_j \mid \geq$ valeur critique
les moyennes sont déclarées différentes.

Les valeurs critiques varient selon les méthodes.

Pas de commentaire

t-corrigé de Bonferonni

on choisit α_g , par exemple 5%

pour 1 comparaison $\quad \alpha = \alpha_g / \frac{I(I-1)}{2}$

$$H_0 : \mu_i = \mu_j \text{ rejetée si}$$
$$\mid \hat{\mu}_i - \hat{\mu}_j \mid \geq T(\frac{\alpha_g}{\frac{I(I-1)}{2}}, \nu) \times s_d$$

Intuitivement, méthode la plus simple.

Risque de 1ère espèce de chaque comparaison divisé par le nombre de comparaisons ou plus généralement par le nombre d'hypothèses testées.

Si α devient très petit, c'est-à-dire si le nombre de comparaisons est très grand, le test devient très peu puissant (il aura tendance à mettre en évidence peu de différences réelles).

Ce test est conservatif : le risque de 1ère espèce global réel est inférieur au risque choisi.

SI LE NOMBRE DE REPETITIONS EST CONSTANT

Méthode de Tukey

Intervalles de confiance simultanés pour les différences $\mu_i - \mu_j$

$$H_0 : \mu_i = \mu_j \text{ rejetée si}$$
$$\mid \hat{\mu}_i - \hat{\mu}_j \mid \geq q(\alpha_g, I, \nu) \times s_d/\sqrt{2}$$

$q(\alpha_g, I, \nu)$: quantile de "Studentized range"

Test conservatif

SI LE NOMBRE DE REPETITIONS EST CONSTANT

Méthode de Newman-Keuls

Variante séquentielle de la méthode de Tukey, plus puissante.

Dans un ordre déterminé, on compare la différence entre la plus grande et la plus petite moyenne d'un groupe de p moyennes à

$$\boxed{q(\alpha_g, p, \nu) \times s_d/\sqrt{2}}$$

Détecte des différences vraies plus souvent que Tukey.

Exemple traité par la méthode de Newman-Keuls

Nb de moyennes p	2	3	4	5	6
Etendue critique	4.48	5.42	5.99	6.40	6.71

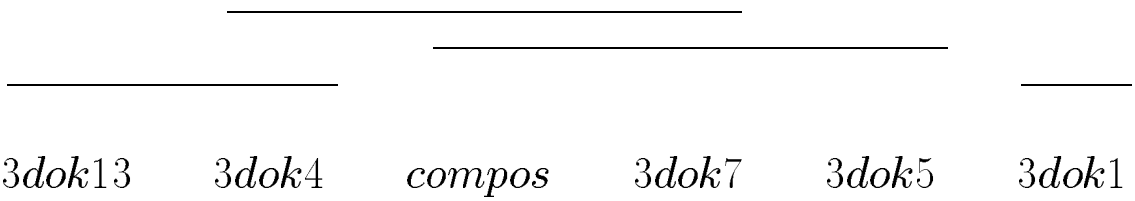
souche	3dok13	3dok4	compos	3dok7	3dok5	3dok1
moyenne	13.26	14.64	18.70	19.92	23.98	28.82

$p = 6$	3dok1-3dok13 > 6.71	significatif
$p = 5$	3dok1-3dok4 > 6.40	significatif
	3dok5-3dok13 > 6.40	significatif
$p = 4$	3dok1-compos > 5.99	significatif
	3dok5-3dok4 > 5.99	significatif
	3dok7-3dok13 > 5.99	significatif
$p = 3$	3dok1-3dok7 > 5.42	significatif
	3dok5-compos < 5.42	non significatif
	3dok7-3dok4 < 5.42	non significatif
	compos-3dok13 > 5.42	significatif
$p = 2$	3dok1-3dok5 > 4.48	significatif
	3dok4-3dok13 < 5.42	non significatif

AZOTE / RHIZOBIUM

- Les valeurs critiques de Newman-Keuls, 1 par nombre de moyennes dans le groupe.
- On classe les moyennes par ordre croissant.
- On commence par le groupe de 6 : la différence entre la plus grande et la plus petite moyenne des 6 est significative. On continue.
- Groupes de 5 : les écarts entre les plus grandes et les plus petites sont significatifs. On continue
- Groupes de 4 : les écarts sont significatifs. On continue.
- Groupes de 3 : 2 écarts sont non significatifs, on peut donc considérer que 2 groupes de 3 sont homogènes. (3dok5,3dok7,compos) et (3dok7,compos,3dok4) sont ces groupes.
- Il ne reste que 2 groupes de 2 à étudier : les 2 groupes aux extrémités.

Construire progressivement au tableau le schéma représentant les groupes :



QUELS QUE SOIENT LES NOMBRES DE REPETITIONS

Méthode de Scheffé

Intervalles de confiance simultanés pour tous les contrastes (combinaisons linéaires de moyennes dont la somme des coefficients est 0), en particulier pour les comparaisons de 2 moyennes

$$H_0 : \mu_i = \mu_j \text{ rejetée si}$$
$$| \hat{\mu}_i - \hat{\mu}_j | \geq \sqrt{(I-1)F(\alpha_g, I-1, \nu)} \times s_d$$

Test conservatif, très général

- Exemple de contraste :

$$\mu_1 - \frac{1}{5}(\mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) = 0$$

  comparaison de μ_1 à la moyenne des autres.

Propriétés :

- Cohérent avec le test F : si le F n'est pas significatif, cette méthode ne donnera aucune différence significative.
- Sa généralité lui fait perdre de la puissance.
- Quand les nombres de répétitions sont différents, les 2 méthodes utilisables sont : Scheffé ou Bonferonni.

COMPARAISONS AU TEMOIN

seulement $(I-1)$ comparaisons

Méthode de Dunnett

$$H_0 : \mu_i = \mu_0 \text{ rejetée si}$$
$$\mid \hat{\mu}_i - \hat{\mu}_0 \mid \geq d(\alpha_g, I-1, \nu) \times s_d$$

μ_0 moyenne du Témoin

Traitements meilleurs que le Témoin :

- test unilatéral à droite
  $H_0 : \mu_i = \mu_0 \quad , H_1 : \mu_i > \mu_0$
  si $\hat{\mu}_i - \hat{\mu}_0 > d(\alpha_g\, I-1, \nu) \times s_d$
  on rejette H_0.
- test unilatéral à gauche
  $H_0 : \mu_i = \mu_0 \quad , H_1 : \mu_i < \mu_0$
  si $\hat{\mu}_i - \hat{\mu}_0 < -d(\alpha_g\, I-1, \nu) \times s_d$
  on rejette H_0.

$d(\alpha_g, I-1, \nu)$: quantile lu dans une table de Dunnett.

Dans certaines expériences, les traitements ne sont intéressants que s'ils sont supérieurs ou inférieurs au témoin.

Test à droite : pour rechercher des moyennes supérieures à celles du groupe Témoin.

Test à gauche : pour rechercher les moyennes inférieures à celles du groupe Témoin.

TABLEAU RECAPITULATIF

Méthode	*Valeur critique*
PPDS	$T(\alpha, \nu)\, s_d$
Bonferonni	$T(\frac{\alpha_g}{I(I-1)/2}, \nu)\, s_d$
Tukey	$q(\alpha_g, I, \nu)\, s_d/\sqrt{2}$
Newman Keuls	$q(\alpha_g, p, \nu)\, s_d/\sqrt{2}$ où $p = 2, 3 \ldots I$
Scheffé	$\sqrt{(I-1)F(\alpha_g, I-1, \nu)}\, s_d$
Dunnett	$d(\alpha_g, I-1, \nu)\, s_d$

I	nombre de moyennes à comparer
α	erreur 1ère espèce, de chaque comparaison 2 à 2
α_g	risque global de toutes les comparaisons 2 à 2
ν	degrés de liberté résiduels
$\hat{\sigma}$	écart-type résiduel
s_d	écart-type de la différence des moyennes
	en cas d'effectifs égaux $s_d = \hat{\sigma}\sqrt{2/n}$
	en cas d'effectifs inégaux $s_d = \hat{\sigma}\sqrt{1/n_i + 1/n_j}$
T	quantile de la loi de Student (table)
q	quantile "studentized range" (table)
F	quantile de la loi de Fisher (table)
d	quantile de Dunnett (table)

Ce tableau récapitule les valeurs critiques des méthodes présentées.

En très gros, la formule est la même : un quantile (à rechercher dans une table) multiplié par s_d l'écart-type de la différence de 2 moyennes.

Toutes les valeurs critiques sauf celle de la PPDS dépendent du risque global α_g choisi par l'utilisateur.

CARACTERISTIQUES DES METHODES

	PPDS	*Tukey*	*Newman Keuls*
Erreur 1ère espèce contrôlée	pour chaque comparaison ou contraste	pour toutes les comparaisons de 2 moyennes	pour toutes les comparaisons de 2 moyennes
Nombre de répétitions	non constant	constant	constant
Utilisations	quelques comparaisons ou contrastes définis a priori	comparaisons de toutes les moyennes	comparaisons de toutes les moyennes
Caracté-ristiques		conservatif	

contraste : combinaison linéaire de moyennes dont la somme des coefficients est 0
exemples : $\mu_1 - \mu_2$, $\mu_1 - \frac{\mu_2+\mu_3}{2}$

test conservatif : test dont le risque de 1ère espèce réel est inférieur au risque choisi (par exemple 5 %).

CONDITIONS D'UTILISATION DES METHODES

- PPDS déconseillée sauf cas particulier où les comparaisons sont prévues à l'avance et en nombre très restreint.

- Exclusivement en cas de nombre de répétitions constant (c'est-à-dire le même nombre de répétitions pour tous les niveaux) : Tukey ou Newman-Keuls qui permet en général de mieux distinguer les groupes que Tukey (plus puissant).

CARACTERISTIQUES DES METHODES - suite

	Bonferonni	*Scheffé*	*Dunnett*
Erreur 1ère espèce contrôlée	pour toutes les comparaisons effectuées	pour tous les contrastes	pour toutes les comparaisons au témoin
Nombre de répétitions	non constant	non constant	constant
Utilisations	comparaisons de moyennes ou contrastes en nombre réduit	contrastes (plus de 2 moyennes) comparaisons de 2 moyennes en grand nombre	comparaisons au témoin recherche des moyennes les plus élevées
Caractéristiques	général conservatif	très général conservatif cohérent avec le test F	bilatéral ou unilatéral

contraste : combinaison linéaire de moyennes dont la somme des coefficients est 0
exemples : $\mu_1 - \mu_2$, $\mu_1 - \frac{\mu_2+\mu_3}{2}$

test conservatif : test dont le risque de 1ère espèce réel est inférieur au risque choisi (par exemple 5 %).

- En cas de nombre de répétitions différents : Bonferonni ou Scheffé.

  Bonferonni plus puissant que Scheffé, c'est à-dire distingue plus de groupes, lorsque le nombre d'hypothèses testées n'est pas trop élevé.

  Si c'est le cas, Scheffé devient plus puissant que Bonferonni.

  De toutes façons, ces 2 méthodes ne sont pas très puissantes.

- Dunnett : en principe le nombre de répétitions n'a pas à être constant MAIS les tables (ou les programmes) ne sont pas prévues dans le cas où le nombre de répétitions n'est pas constant.

EXEMPLE

Méthode	Risque choisi	Valeur critique pour les différences entre 2 moyennes	option SAS
P.P.D.S	$\alpha = 5\%$	4.48	LSD ou T
Tukey	$\alpha_g = 5\%$	6.71	TUKEY
Newman Keuls	$\alpha_g = 5\%$	Nb de moyennes / Etendue 2 — 4.48 3 — 5.42 4 — 5.99 5 — 6.40 6 — 6.71	SNK
Bonferonni	$\alpha_g = 5\%$	7.08	BON
Scheffé	$\alpha_g = 5\%$	7.86	SCHEFFE
Dunnett	$\alpha_g = 5\%$	5.85	DUNNETT ('compos')

Pour l'exemple AZOTE/RHIZOBIUM :
Toutes les méthodes sont utilisables (nombre de répétitions constant).
Le tableau compare les valeurs critiques calculées pour un risque global $\alpha_g = 5\%$.
Les valeurs critiques de Bonferonni et Scheffé sont supérieures à celle de Tukey (7.08 et 7.86 contre 6.71). Bonferonni et Scheffé détectent donc moins de différences que Tukey.
On présente aussi le nom de l'option SAS qui permet d'obtenir le test.
Programme SAS :

```
proc glm;
class souche;
model azote=souche;
lsmeans souche /tdiff; /*differences entre moyennes*/
means souche /lsd tukey snk bon scheffe lines;
means souche /dunnett('compos');
run;
```

RESULTATS

SOUCHE	AZOTE
	moyenne
3dok1	28.82
3dok5	23.98
3dok7	19.92
compos	18.70
3dok4	14.64
3dok13	13.26

DIFFERENCES ENTRE LES MOYENNES

	3dok1	3dok5	3dok7	compos	3dok4
3dok1	0				
3dok5	4.84	0			
3dok7	8.90	4.06	0		
compos	10.12	5.28	1.22	0	
3dok4	14.18	9.34	5.28	4.06	0
3dok13	15.56	10.72	6.66	5.44	1.38

Les moyennes sont classées par ordre croissant.

Les différences entre moyennes sont présentées dans le tableau.
La valeur 9.34 correspond à la différence entre 2 moyennes 3dok5-3dok4 (23.98-14.64).

CLASSEMENTS OBTENUS

P.P.D.S	Newman Keuls	Tukey	Bonferonni	Scheffé	Moyenne	Souche
A	A	A	A	A	28.82	3dok1
B	B	B A	B A	B A	23.98	3dok5
C B	C B	B C	B C	B C	19.92	3dok7
C D	C B	B C	B C	B C	18.70	compos
E D	C D	C	C	C	14.64	3dok4
E	D	C	C	C	13.26	3dok13

Les moyennes portant la même lettre ne sont pas significativement différentes.

DUNNETT

SOUCHE

Comparaison

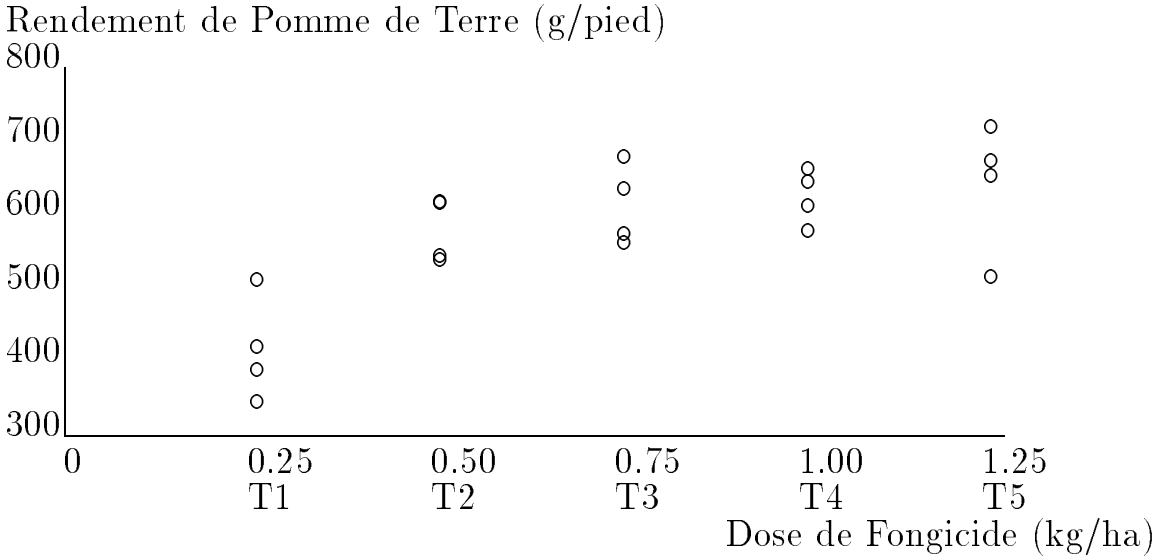
3dok1	-	compos	***
3dok5	-	compos	
3dok7	-	compos	
3dok4	-	compos	
3dok13	-	compos	

Les comparaisons significatives sont repérées par ***

- En comparant les écarts entre moyennes aux valeurs critiques précédemment vues, on peut déclarer quels écarts sont significatifs et quels écarts ne le sont pas.
- 2 moyennes non significativement différentes sont dans le même groupe, un groupe étant repéré par 1 lettre. Exemple : avec Tukey, 3dok1 et 3dok5 sont dans le même groupe (repéré par A), 3dok5 3dok7 et compos sont dans le même groupe (B), 3dok7 compos 3dok4 3dok13 dans le même groupe (C).
- On peut aussi figurer les groupes en reliant par un trait les moyennes non différentes.
- Le test de Dunnett confirme la différence de 3dok1 avec les autres.
- Le test de Newman-Keuls apparaît comme le plus discriminant (exception faite de la PPDS qui ne contrôle pas le risque global).

FACTEURS QUANTITATIFS
Dose, Temps, Température ...

- La différence entre 2 niveaux a un sens numérique
- Comparer toutes les moyennes entre elles ignore la logique de l'expérience
- Décomposer la somme des carrés due au facteur en composantes linéaire, quadratique ...



Les niveaux correspondent par exemple à des doses croissantes d'un substance X.
Il peut être intéressant de savoir si la réponse Y dépend linéairement de X.

On va expliquer la méthode sur un exemple : Rendement de pomme de terre en fonction du traitement (dose) d'un fongicide. On dispose de 4 répétitions par dose de fongicide.

OBS	TRAIT	DOSE kg/ha	RDT g/pied	OBS	TRAIT	DOSE	RDT
1	T1	0.25	377	11	T3	0.75	562
2	T1	0.25	408	12	T3	0.75	667
3	T1	0.25	500	13	T4	1.00	633
4	T1	0.25	333	14	T4	1.00	600
5	T2	0.50	527	15	T4	1.00	650
6	T2	0.50	604	16	T4	1.00	567
7	T2	0.50	606	17	T5	1.25	642
8	T2	0.50	533	18	T5	1.25	708
9	T3	0.75	623	19	T5	1.25	662
10	T3	0.75	550	20	T5	1.25	504

Modèle 1 : Analyse de variance à 1 facteur TRAIT

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$i = 1 \cdots 5$	traitements (facteur TRAIT)
$j = 1 \cdots 4$	répétitions
Y	rendement
μ	moyenne générale
α_i	effet du traitement i
ϵ_{ij}	erreur

Source de variation	d.d.l.	SCE	CM	F	Pr >F
TRAIT	4	133419.2	33354.8	8.81	0.0007
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

L'effet TRAIT est significatif.

On veut déjà savoir s'il y a des différences entre les rendements obtenus pour les 5 traitements : ANALYSE DE VARIANCE.

Interprétation : le test F est hautement significatif. Les 5 traitements ne donnent pas des rendements équivalents.

Comment préciser ce résultat ?

Méthode de Newman-Keuls

$$\alpha_g = 5\%, \quad \hat{\sigma^2} = 3785.6, \quad \nu = 15(d.d.l.)$$

Nombre de moyennes	2	3	4	5
Etendue critique	92.7	113.0	125.4	134.3

Les moyennes portant la même lettre ne sont pas significativement différentes.

	Moyenne	TRAIT
A	629.0	T5
A	612.5	T4
A	600.5	T3
A	567.5	T2
B	404.5	T1

Par un test multiple, par exemple la méthode de Newman-Keuls (équirépétitions).

On peut distinguer 2 groupes : T1 d'une part et T2 T3 T4 T5 d'autre part.

Ceci n'est pas très instructif, à part T1 qui est plus faible, les autres sont considérés comme équivalents.

Modèle 2

$$Y_{ij} = \mu + \underbrace{\beta \times d_i + \alpha'_i}_{\alpha_i} + \epsilon_{ij}$$

d_i dose de fongicide correspondant au niveau i de TRAIT

α'_i effet du facteur TRAIT, écart au modèle de régression linéaire

ϵ_{ij} erreur

Source de variation	d.d.l.	SCE Type I	CM	F	Pr >F
Dose	1	97614.4	97614.4	27.79	0.0001
= Régression linéaire					
TRAIT	3	35804.8	11934.9	3.15	0.056
= Ecart à la Régression linéaire					
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

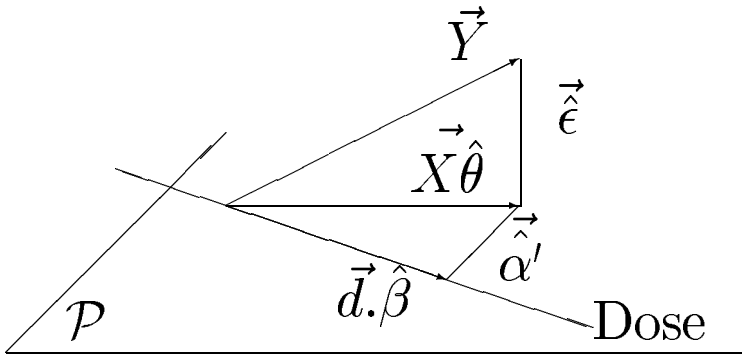
L'effet linéaire Dose est hautement significatif.
L'effet TRAIT est pratiquement significatif à 5 %.

Le deuxième modèle envisagé décompose α_i en

- $\beta \times d_i$: régression linéaire simple en fonction de la dose
- $+\alpha'_i$: ce qui n'est pas expliqué par $\beta \times d_i$ (écart entre α_i et la régression linéaire).

Avec les sommes de carrés type I, on décompose la SCE du terme TRAIT du modèle 1 en SCE expliquée par la régression et SCE expliquée par l'effet TRAIT non pris en compte par la régression.

L'effet TRAIT n'est pas négligeable.



$\mathcal{P}$ engendré par TRAIT

$$\vec{Y} = X\vec{\hat{\theta}} + \vec{\hat{\epsilon}} \qquad \text{projection 1 sur } \mathcal{P}$$

$$X\vec{\hat{\theta}} = \vec{d}.\beta + \vec{\hat{\alpha}'} \qquad \text{projection 2 sur Dose}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_5 \\ \alpha_5 \\ \alpha_5 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_1 \\ d_1 \\ d_1 \\ \vdots \\ d_5 \\ d_5 \\ d_5 \\ d_5 \end{bmatrix} .\beta + \begin{bmatrix} \alpha'_1 \\ \alpha'_1 \\ \alpha'_1 \\ \alpha'_1 \\ \vdots \\ \alpha'_5 \\ \alpha'_5 \\ \alpha'_5 \\ \alpha'_5 \end{bmatrix}$$

Illustration géométrique :

- On projette $\vec{Y}$ sur l'espace $\mathcal{P}$ engendré par les traitements (TRAIT).

- On obtient $\vec{X\hat{\theta}}$.

- On projette $\vec{X\hat{\theta}}$ sur la droite engendrée par le vecteur Dose.

- $\vec{d\hat{\beta}}, \vec{\hat{\alpha}'}, \vec{\hat{\epsilon}}$ sont 2 à 2 orthogonaux donc leurs 3 sommes de carrés s'ajoutent.

Modèle 3

$$Y_{ij} = \mu + \underbrace{\beta \times d_i + \gamma \times d_i^2 + \alpha_i''}_{\alpha_i} + \epsilon_{ij}$$

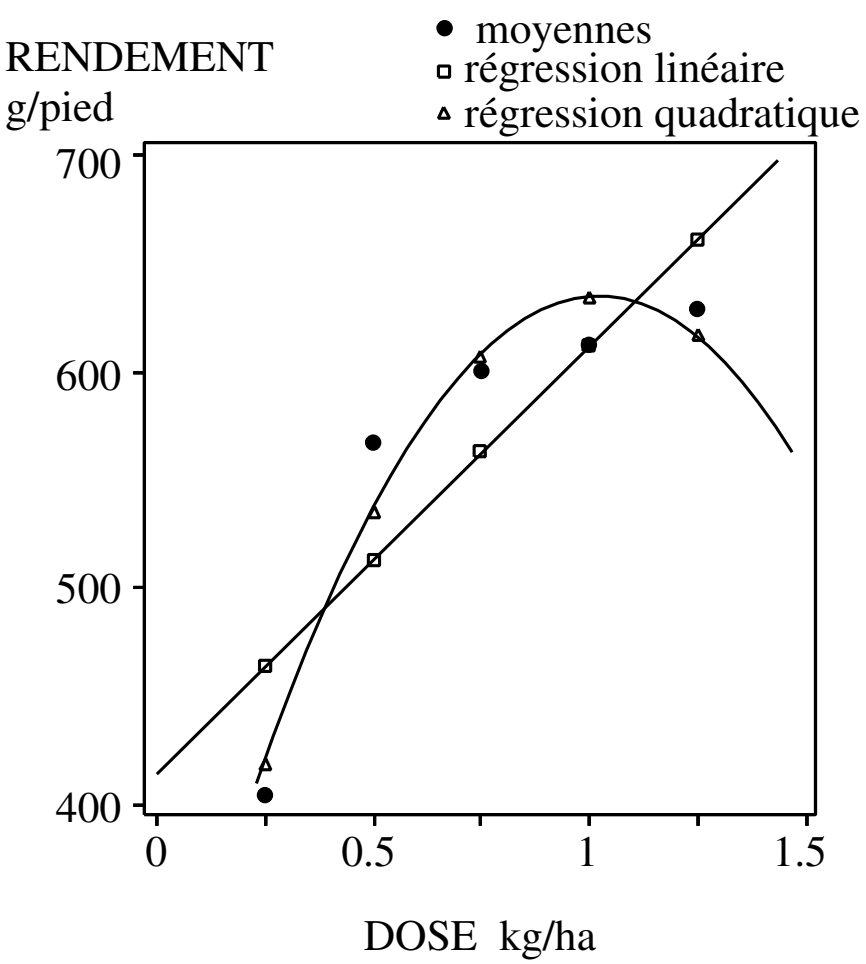
Source de variation	d.d.l.	SCE Type I	CM	F	Pr >F
Dose	1	97614.4	97614.4	27.79	0.0001
Dose^2	1	28170.3	28170.3	7.44	0.0156
TRAIT	2	7634.5	3817.3	1.01	0.3883
= Ecart au modéle quadratique					
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

L'effet quadratique Dose^2 apporte une part d'explication significative.
L'effet TRAIT (écart au modèle quadratique) n'est plus du tout significatif.

On peut s'arrêter là et conclure que la relation entre le rendement et la dose peut être décrite par une régression linéaire simple mais cela semble grossier.

On peut donc envisager un troisième modèle qui décompose l'effet α_i en

- $\beta \times d_i$: régression linéaire
- $+\gamma \times d_i^2$: terme où la valeur de dose est au carré (régression quadratique). Ceci permet de tester si la relation Rendement/Dose présente une courbure.
- $+\alpha_i''$: effet TRAIT non pris en compte ou écart au modèle quadratique.


RENDEMENT
g/pied
moyennes
régression linéaire
régression quadratique
700
600
500
400
0
0.5
1
1.5
DOSE kg/ha

- Le graphique présente les moyennes ($\bullet$) du Rendement par niveau de Dose encadrées par $(+1)\times$ Ecart-type et $(-1)\times$ Ecart-type afin de juger de la variabilité.
- On y a superposé la droite de régression linéaire ($\square$) qui s'ajuste au mieux et la parabole de régression quadratique ($\triangle$) s'ajustant au mieux.
- On comprend ainsi pourquoi le Modèle 3 précédemment vu ne présente plus du tout d'effet TRAIT significatif (écart à la régression quadratique non significatif).
- En revanche le Modèle 2, présentait un effet TRAIT pratiquement significatif à 5 %, la droite ajustée ne décrit que très grossièrement la relation.

CONCLUSION : la parabole approxime de façon correcte la relation Rendement/Dose même si ce graphique indique que ce n'est peut être pas le modèle idéal.