ISARA - 2ème année - STATISTIQUE
Cours : Partie 6
Mme B. Bottollier-Lemallaz

ANALYSE DE LA VARIANCE

SOMMAIRE

OBJECTIFS

1. Expliquer le principe et l'objectif de l'analyse de la variance.
2. Enumérer les conditions à vérifier avant d'avoir une estimation commune de l'erreur expérimentale. Préciser les méthodes utilisées.
3. Identifier si une étude proposée relève de l'analyse de la variance à 1, 2 ou 3 facteurs avec ou sans répétitions et écrire le modèle théorique correspondant avec la notation indicielle rigoureuse !
4. Enoncer dans une étude proposée quels sont les facteurs étudiés et ceux qui sont mis sous contrôle ainsi que la nature des fluctuations liées à l'erreur expérimentale.
5. Identifier le nombre et la nature des tests statistiques à réaliser dans une étude proposée.
6. Calculer la liste des résidus pour les modèles à 1 et 2 facteurs étudiés.
7. Interpréter l'allure d'un histogramme, une cartographie, les résultats des tests sur les coefficients de Pearson.
8. Réaliser et interpréter un test de Bartlett.
9. Réaliser et interpréter l'analyse de la variance.
10. Réaliser et interpréter les tests de comparaison des moyennes par la méthode ppds.

I – Le principe avec un exemple

L'analyse de la variance permet d'étudier l'influence d'un ou de plusieurs facteurs étudiés sur une variable réponse. On essaie de répondre aux questions suivantes :

- La réponse est-elle la même quelque soit les modalités* du (des) facteur(s) étudié(s)?
- Si non, quelles sont les modalités des facteurs qui donnent des réponses différentes ?

Dans le modèle à 1 facteur, on considère que la variable réponse est le résultat de l'effet d'un seul facteur* et l'étude consiste à vérifier l'existence ou non de cet effet autrement dit : l'effet du facteur étudié est-il significatif ?
Dans le modèle à k facteurs, on considère que la variable réponse est le résultat de l'effet de k facteurs avec ou sans interactions et l'étude consiste à vérifier l'existence ou non de ces effets autrement dit : quels sont les facteurs et / ou interactions qui sont significatifs ?
↳ lorsque l'effet d'un facteur n'est pas le même selon le niveau / la modalité de l'autre facteur.

On poursuivra l'étude par une comparaison des effets des modalités du facteur pour faire des regroupements.

Pommes………  1 Facteur à 4 modalités

L'objectif de l'étude menée ici est de comparer, à la pré récolte, la fermeté d'une variété de pommes cultivée selon 5 conduites différentes et répétées 4 fois. Le dispositif expérimental se déroulera dans un verger de 20 pommiers, sur chacun d'eux on prélèvera 1 pomme à une hauteur et une exposition communes. Sur chaque pomme on mesurera des caractéristiques physiques. A l'aide d'une machine de traction, il est possible de réaliser des tests de compression et de perforation sur les pommes. Lors de ces tests, des données sont enregistrées (force, temps, profondeur…). Nous ne retiendrons ici que la variable 'résistance à la déformation' lorsqu'est appliquée une compression à 20% avec la machine de traction. L'unité de mesure de cette variable est le Newton.

- facteur étudié : la conduite de la culture à 5 niveaux (5 modalités)
- variable réponse : la résistance à la compression à la pré récolte
- individu (unité) statistique : la pomme, il y a 20 unités statistiques

Le dispositif : Randomisation totale avec 4 répétitions (Annexe 1)
↳ répétition au hasard

numéro de parcelle mais [illegible] un pommier

conduite

n° de répétition

301 C1 r3	102 C2 r1	405 C5 r4	404 C4 r4	402 C2 r4
105 C5 r1	303 C3 r3	201 C1 r2	302 C2 r3	403 C3 r4
203 C3 r2	401 C1 r4	305 C5 r3	202 C2 r2	104 C4 r1
101 C1 r1	304 C4 r3	205 C5 r2	103 C3 r1	204 C4 r2

*le facteur peut être un amendement, une ration, une localisation etc.

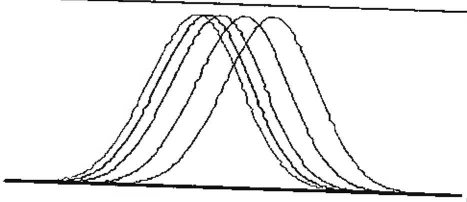
*les modalités d'un facteur sont ses variance : différentes doses, + ou - long ensoleillement...

Les données recueillies (réponses) sont répertoriées dans le tableau suivant :

	Résistance à la compression des pommes selon les 5 conduites C					
modalité / répét.	C1	C2	C3	C4	C5	
r1	296	334	442	374	188	
r2	271	344	415	354	187	
r3	297	351	445	335	197	
r4	276	325	466	331	220	
$\bar{X}_{i.}$	285,000	338,500	442,000	348,500	198,000	$\bar{X}_{..}$ = 322,40

moyenne totale

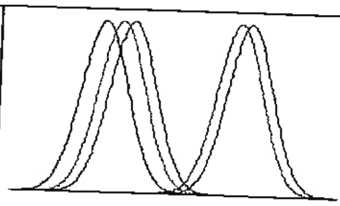
↳ moyenne pour chaque modalité



Graphique 1 :

→ pas de différence quelque soit la conduite

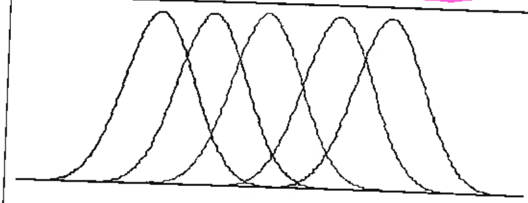
Si les résultats obtenus peuvent se présenter sous la forme ci-dessus, alors on pourra conclure qu'il n'y a pas d'effet de la pratique culturale sur la fermeté des pommes. En effet, les réponses ne sont pas significativement différentes d'une conduite à l'autre. Les courbes représentant la résistance à la compression pour chacune des conduites sont proches les unes des autres et se chevauchent.



Graphique 2 :

→ des conduites sont cools, d'autres pas

Par contre si les résultats obtenus se présentent sous la forme ci-dessus, alors on pourra conclure qu'il y a un effet de la pratique culturale sur la fermeté des pommes. En effet, certaines conduites ne donnent pas les mêmes fermetés moyennes.



Graphique 3 :

→ chaque conduite ~~est différente~~ donne un résultat différent

Toutes les pratiques culturales donnent des fermetés moyennes significativement différentes. Les courbes ne se chevauchent pas toutes.

L'analyse des résultats (Logiciel STATBOX)

1 - ANALYSE DE L'ERREUR : ANALYSE DES RESIDUS

↳ selon fluctuation naturelle des résultats

Histogramme des résidus : → l'histogramme est donné

7		303		
6		403	405	
5		201	202	
4		402	404	103
3	104	105	304	305
2	301	102	101	205
1	203	302	401	204
Effectifs	3 +	7 +	6 +	4 = 20
Bornes	-27,0 à -13,88	-13,88 à -0,75	-0,75 à 12,38	12,38 à 25,5

→ on doit pouvoir dire s'il est en cloche...

Minimum : - 27,000 Maximum : 25,500 Intervalle : 13,125

Indices de normalité (coefficients de K.PEARSON) :

Beta 1 = 0,032 Prob. : 0,727 → pvalue !

Beta 2 = 2,135 Prob. : 0,384

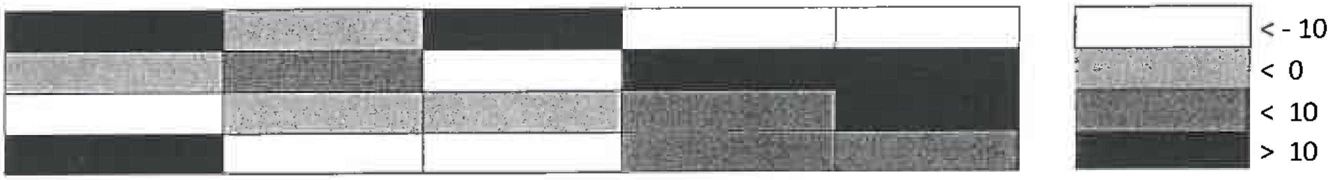
avec 38,4% de chance de tromper, c'est ...

↳ on doit savoir interpreter

Résidus suspects (méthode de GRUBBS) :

Aucun résidu suspect

Cartographie des résidus : → savoir interpreter l'indpdce des individus



Ecart type des résidus : → Test de Barcklet, comparaison de variances

→ on doit savoir donner plein de trucs sur le khi² etc.

Ecarts-types facteur 1 = C	E.T.
1 (C1)	13,441
2 (C2)	11,387
3 (C3)	20,928
4 (C4)	19,740
5 (C5)	15,341

khi² = 1,344 Prob. = 0,85488

SI les VARIANCES sont HOMOGENES ALORS ↓

2 - ANALYSE DE L'"EFFET « VARIETE » : ANALYSE DE LA VARIANCE (ANOVA)

	S.C.E	DDL	C.M.	TEST F	PROBA
Var TOTALE	132594,800	19	6978,674		
Var. FACTEUR 1	128474,800	4	32118,700	116,937	0,000
VAR.RESIDUELLE 1	4120,000	15	274,667		

↳ conclusion : le facteur a ou pas d'effet

3 - COMPARAISON DES VARIETES : REGROUPEMENTS DES MOYENNES PAR LA METHODE PPDS

plus petite différence significatif

Valeur des PPDS (5%) = 24.97

4 Groupes homogènes

Id	Modalité	Moyenne	Groupes homogènes			
3	C3	442,000	A			
4	C4	348,500		B		
2	C2	338,500		B		
1	C1	285,000			C	
5	C5	198,000				D

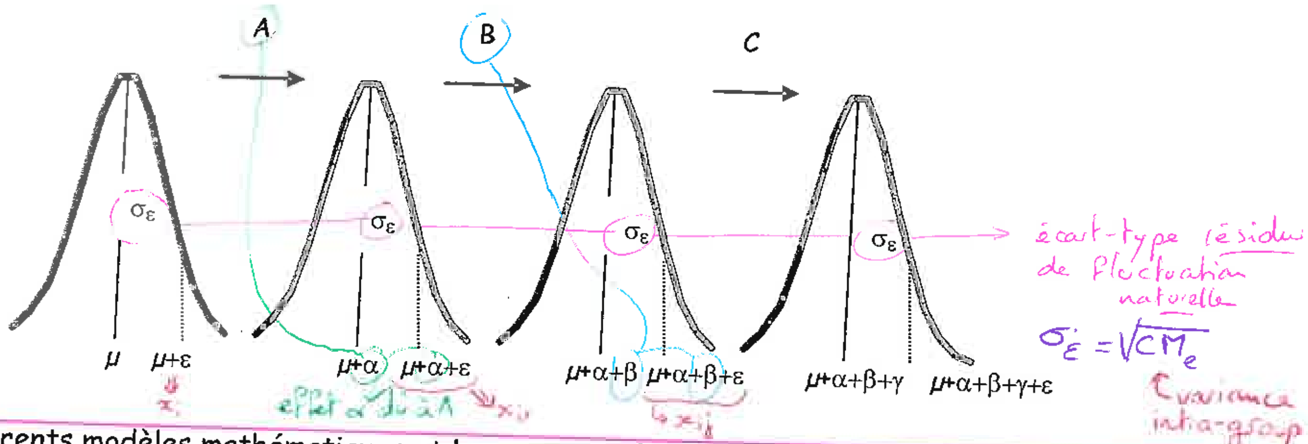
C3 → la meilleure conduite

C4, C2 : conduite au résultat équivalent.

A : toute seule — B : 2 conduites donnent le même résultat — C : autre résultat — D : autre résultat

II – Les modèles et les données

- La réponse (résultat) suit un modèle linéaire additif.
- $L(\varepsilon) = N(0 ; \sigma_\varepsilon)$
- Les effets des facteurs sont certains et ils n'agissent que sur la moyenne.



Les différents modèles mathématiques et leur erreur résiduelle:

Dispositif	Modèle	Erreur (résidu)
1 facteur étudié	$x_{ir} = \mu + \alpha_i + \varepsilon_{ir}$	$e_{ir} = x_{ir} - \bar{X}_{i.}$
2 facteurs étudiés sans répétition	$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$	$e_{ij} = x_{ij} + \bar{X}_{..} - \bar{X}_{i.} - \bar{X}_{.j}$ *
2 facteurs étudiés avec répétitions (pour les interactions)	$x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_{ijr}$	$e_{ijr} = x_{ijr} - \bar{X}_{ij.}$
3 facteurs étudiés sans répétition	$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \varepsilon_{ijk}$	
3 facteurs étudiés avec répétitions	$x_{ijkr} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \alpha_i\beta_j\gamma_k + \varepsilon_{ijkr}$	

Etc.......

Tout ça doit se vérifier test F

x_{ijkr} : réponse observée pour la $r^{ième}$ répétition pour la combinaison $A_iB_jC_k$
μ : niveau moyen
α_i : effet principal du facteur étudié A au niveau i ; i = 1......a
β_j : effet principal du facteur étudié B au niveau j ; j = 1......b
γ_k : effet principal du facteur étudié C au niveau k ; k = 1......c
$\alpha_i\beta_j$: effet de l'interaction du 1er ordre entre les facteurs A et B au niveau ij
$\alpha_i\gamma_k$: effet de l'interaction du 1er ordre entre les facteurs A et C au niveau ik
$\beta_j\gamma_k$: effet de l'interaction du 1er ordre entre les facteurs B et C au niveau jk
$\alpha_i\beta_j\gamma_k$: effet de l'interaction du 2ème ordre entre les facteurs A, B, C au niveau ijk
ε_{ijkr} : effet résiduel

III – Les calculs des variations

x_{ijkr}: $r^{ième}$ valeur observée de la réponse pour la combinaison $A_iB_jC_k$
a = nombre de niveaux du facteur A — i = 1, ..., a
b = nombre de niveaux du facteur B — j = 1, ..., b
c = nombre de niveaux du facteur C — k = 1, ..., c
n = nombre de répétitions pour la combinaison $A_iB_jC_k$ — r = 1, ...,n (si r =1 alors pas de répétition, pas d'indice r)
N =nombre d'unités statistiques = nombre total de réponses
X = somme des N valeurs
μ = niveau moyen = X / N (autres notations selon les dispositifs : $\bar{X}_{..}$ ou $\bar{X}_{...}$ ou $\bar{X}_{....}$)
C = terme de centrage = $X^2 / N = C$
SC = somme des carrés des N valeurs

* retrouvable avec des trucs bizarres

Rappel : SCE : somme des carrés des écarts, un des paramètres pour mesurer la dispersion d'une série

La variation totale de l'ensemble des réponses est égale à la somme entre toutes les variations liées aux facteurs, aux interactions et à l'erreur naturelle (résidus).

$$SCE_{totale} = SCE_{facteurs} + SCE_{interactions} + SCE_{résiduelle}$$

Somme, nombre de valeurs et moyenne par modalité / combinaison Annexe 2	1 facteur avec répétitions	2 facteurs sans répétitions	2 facteurs avec répétitions	3 facteurs sans répétitions	3 facteurs avec répétitions
X_i : somme A_i n_i : nombre de valeurs A_i m_i : moyenne A_i	$X_{i.}$ n $\overline{X}_{i.}$	$X_{i.}$ b $\overline{X}_{i.}$	$X_{i..}$ nb $\overline{X}_{i..}$	$X_{i..}$ bc $\overline{X}_{i..}$	$X_{i...}$ nbc $\overline{X}_{i...}$
X_j : somme B_j n_j : nombre de valeurs B_j m_j : moyenne B_j		$X_{.j}$ a $\overline{X}_{.j}$	$X_{.j.}$ na $\overline{X}_{.j.}$	$X_{.j.}$ ac $\overline{X}_{.j.}$	$X_{.j..}$ nac $\overline{X}_{.j..}$
X_k : somme C_k n_k : nombre de valeurs C_k m_k : moyenne C_k				$X_{..k}$ ab $\overline{X}_{..k}$	$X_{..k.}$ nab $\overline{X}_{..k.}$
X_{ij} : somme A_iB_j n_{ij} : nombre de valeurs A_iB_j m_{ij} : moyenne A_iB_j			$X_{ij.}$ n $\overline{X}_{ij.}$	$X_{ij.}$ c $\overline{X}_{ij.}$	$X_{ij..}$ nc $\overline{X}_{ij..}$
X_{ik} : somme A_iC_k n_{ik} : nombre de valeurs A_iC_k m_{ik} : moyenne A_iC_k				$X_{i.k}$ b $\overline{X}_{i.k}$	$X_{i.k.}$ nb $\overline{X}_{i.k.}$
X_{jk} : somme B_jC_k n_{jk} : nombre de valeurs B_jC_k m_{jk} : moyenne B_jC_k				$X_{.jk}$ a $\overline{X}_{.jk}$	$X_{.jk.}$ na $\overline{X}_{.jk.}$
X_{ijk} : somme $A_iB_jC_k$ n_{ijk} : nombre de valeurs $A_iB_jC_k$ m_{ijk} : moyenne $A_iB_jC_k$					$X_{ijk.}$ n $\overline{X}_{ijk.}$

$SCE_{totale} = SC - \mathfrak{C}$

SCE facteurs :

$SCE_A = (\Sigma X_i^2 / n_i) - \mathfrak{C}$

$SCE_B = (\Sigma X_j^2 / n_j) - \mathfrak{C}$

$SCE_C = (\Sigma X_k^2 / n_k) - \mathfrak{C}$

SCE interactions:

$SCE_{AB} = (\Sigma X_{ij}^2 / n_{ij}) - \mathfrak{C} - SCE_A - SCE_B$

$SCE_{AC} = (\Sigma X_{ik}^2 / n_{ik}) - \mathfrak{C} - SCE_A - SCE_C$

$SCE_{BC} = (\Sigma X_{jk}^2 / n_{jk}) - \mathfrak{C} - SCE_B - SCE_C$

$SCE_{ABC} = (\Sigma X_{ijk}^2 / n_{ijk}) - \mathfrak{C} - SCE_A - SCE_B - SCE_C$

$SCEe = SCE_{totale} - SCE_{facteurs} - SCE_{interactions}$

Pommes…	Somme
1 (C1)	1140
2 (C2)	1354
3 (C3)	1768
4 (C4)	1394
5 (C5)	792
total	6448
SC	2211430

$\mathfrak{C} = 6448^2 / 20$

$SCE_{totale} = 132594{,}8$

$SCE_A = 128474{,}8$

$SCEe = 4120{,}0$

IV- L'analyse de la variance

On va, à partir d'une seule série de données, effectuer plusieurs tests statistiques F.
Pour chaque test, l'hypothèse de base H_0 correspond à l'égalité des moyennes des variantes d'un facteur (ou interaction), donc absence d'effet du facteur (ou interaction).

Chacun des tests est basé sur la comparaison de la variance due au facteur (ou à l'interaction) avec la variance résiduelle.

CM_F = Carré moyen du facteur ; représente la variation de la réponse due aux différentes modalités de ce facteur (CM_A CM_B CM_C ...)
CM_I = Carré moyen de l'interaction ; représente la variation entre les différentes combinaisons (CM_{AB} CM_{AC} CM_{BC} CM_{ABC}...)
CMe = Carré moyen résiduel ; c'est-à-dire la part de variation non expliquée par le(s) facteur(s) (ou interaction(s)) et non prise en compte dans le modèle.

Tester l'effet	H_0	H_1
du facteur A	$CM_A/CMe=1$	$CM_A/CMe>1$
du facteur B	$CM_B/CMe=1$	$CM_B/CMe>1$
du facteur C	$CM_C/CMe=1$	$CM_C/CMe>1$
de l'interaction AB	$CM_{AB}/CMe=1$	$CM_{AB}/CMe>1$
de l'interaction AC	$CM_{AC}/CMe=1$	$CM_{AC}/CMe>1$
de l'interaction BC	$CM_{BC}/CMe=1$	$CM_{BC}/CMe>1$
de l'interaction ABC	$CM_{ABC}/CMe=1$	$CM_{ABC}/CMe>1$

Etc.....

Tableau de l'ANOVA Tableau général ⚠ important

Source de variation	SCE	ddl	CM	F calculé	F théorique $F_{1-\alpha}(\nu_1 ; ddle)$
Facteur A	SCE_A	a - 1	CM_A	$F_A = CM_A / CMe$	$\nu_1 = a - 1$
Facteur B	SCE_B	b - 1	CM_B	$F_B = CM_B / CMe$	$\nu_1 = b - 1$
Facteur C	SCE_C	c - 1	CM_C	$F_C = CM_C / CMe$	$\nu_1 = c - 1$
Interaction AB	SCE_{AB}	(a-1) (b-1)	CM_{AB}	$F_{AB} = CM_{AB} / CMe$	$\nu_1 = (a-1)(b-1)$
Interaction AC	SCE_{AC}	(a-1) (c-1)	CM_{AC}	$F_{AC} = CM_{AC} / CMe$	$\nu_1 = (a-1)(c-1)$
Interaction BC	SCE_{BC}	(b-1) (c-1)	CM_{BC}	$F_{BC} = CM_{BC} / CMe$	$\nu_1 = (b-1)(c-1)$
Interaction ABC	SCE_{ABC}	(a-1) (b-1) (c-1)	CM_{ABC}	$F_{ABC} = CM_{ABC} / CMe$	$\nu_1 = (a-1)(b-1)(c-1)$
résiduelle	SCE_e	ddle (*)	CMe		
totale	SCE_T	N - 1			

(*) ddle = (N - 1) – la somme des autres ddl

conclusions au verso

Si $F_{calculé} < F_{théorique}$: on n'a pas pu mettre en évidence d'effet du facteur testé (ou de l'interaction testée). Même conclusion si p-value > 0,05.

Si $F_{calculé} > F_{théorique}$: on a pu mettre en évidence un effet du facteur testé (ou de l'interaction testée) avec moins de α% de risque d'erreur. Même conclusion si p-value < 0,05.

V - La méthode PPDS (plus petite différence significative)

Pour chacun des facteurs étudiés, si l'analyse de la variance a mis en évidence un effet de ce facteur, cela signifie que les résultats ne sont pas tous égaux selon les modalités du facteur.
Cela ne signifie pas que toutes les modalités du facteur donnent des résultats différents ; on sait uniquement qu'au moins, une modalité du facteur a généré un résultat différent.
Dans la poursuite de l'étude on va s'interroger sur les différences entre les modalités ; ceci, en comparant les moyennes des modalités du facteur entre elles.

Pour comparer ces moyennes, on peut utiliser un test de Student (test de la ppds) ; ce test permet de comparer les moyennes deux à deux. Ce test va nous permettre de constituer des groupes homogènes de moyennes. Les moyennes appartenant à un même groupe sont considérées comme non différentes, c'est à dire que l'on conserve l'hypothèse de non différence entre elles. Par contre lorsque des moyennes sont différentes on considèrera que les traitements sont différents avec α% de risque de se tromper.

Etape 1 : Les différentes moyennes sont classées dans l'ordre croissant (ou décroissant) .

Etape 2 : On calcule la *ppds* (plus petite différence significative)

CM_e, variance intra groupe

$$ppds = t_{1-\alpha/2}^{(\nu)}\sqrt{\frac{CMe}{n_i}+\frac{CMe}{n_i'}} = t_{1-\alpha/2}^{(\nu)}\sqrt{\frac{2CMe}{n_i}}$$

=

$t_{1-\alpha/2}$: représente la valeur lue dans la table de Student
v : nombre de ddl de la variance résiduelle
ni : nombre d'observations pour le calcul de chaque moyenne

Etape 3 : on compare la différence entre les 2 moyennes à comparer $\left|\overline{x_i}-\overline{x_{i'}}\right|$ à la ppds.

Si on a : $\left|\overline{x_i}-\overline{x_{i'}}\right| < ppds$ alors on accepte l'hypothèse d'égalité des 2 moyennes comparées au seuil de confiance $(1-\alpha)$.

Si par contre, on obtient : $\left|\overline{x_i}-\overline{x_{i'}}\right| \geq ppds$ alors on rejette l'hypothèse d'égalité des 2 moyennes comparées au risque α.

Attention !! Si une interaction entre 2 facteurs est significative on appliquera la méthode de la ppds sur les moyennes des combinaisons des 2 facteurs.

VI - L'analyse des résidus

Les analyses réalisées ANOVA + PPDS n'ont de sens que si 3 conditions qui portent sur les résidus (représentants de l'erreur expérimentale ou fluctuations naturelles) sont vérifiées et qu'il faut vérifier en premier !!
Un résidu représente l'écart estimé entre la valeur de la réponse calculée à partir du modèle et la valeur de la réponse réellement observée.
Il faut d'abord établir la liste des résidus calculés à partir du modèle mathématique.
Remarques :

- la somme et la moyenne des résidus sont égales à 0
- SC résidus = SCE des résidus

CONDITION 1 : VERIFICATION DE LA NORMALITE DES RESIDUS :

On groupe la série des résidus en classes et on représente l'histogramme qui doit être symétrique et en forme de cloche.
Pour tester la normalité, on peut utiliser les coefficients d'asymétrie ($\beta_1 = \mu_3^2/\mu_2^3$) et d'aplatissement ($\beta_2 = \mu_4/\mu_2^2$) de Pearson (consulter les cours de 1ère année):

➢ Pour le coefficient d'asymétrie :

$H_0\ \beta_1 = 0$ $H_1\ \beta_1 \neq 0$ (distribution non symétrique)

On a p-value% de risque de se tromper en rejetant l'hypothèse de symétrie des résidus. Si ce risque est trop fort (> 5%), c'est qu'on n'a pas mis en évidence que la distribution n'était pas symétrique.

➢ Pour le coefficient d'aplatissement :

$H_0\ \beta_2 = 3$ $H_1\ \beta_2 \neq 3$ (distribution non mésocurtique)

On a p-value % de risque de se tromper en rejetant l'hypothèse de mésocurtie des résidus. Si ce risque est trop fort (> 5%), c'est qu'on n'a pas mis en évidence que la distribution n'était pas mésocurtique.

➢ Dans ce cas où les 2 hypothèses H_0 sont conservées, on retient l'hypothèse de normalité des résidus. (si on démontre pas le contraire, c'est une loi normale

CONDITION 2 : INDEPENDANCE DES RESIDUS

Les résidus doivent être indépendants, c'est à dire qu'ils doivent être répartis de façon aléatoire sur le terrain.
On peut vérifier cette répartition des résidus à l'aide de la cartographie des résidus.
La cartographie consiste à représenter par un code couleur la valeur de la classe du résidu calculé sur chaque unité expérimentale. Cette représentation est surtout utile en agronomie ou lorsque la mesure faite sur une unité expérimentale peut influer sur la mesure d'une autre unité expérimentale.

CONDITION 3: HOMOSCEDASTICITE (INVARIANCE DES VARIANCES RESIDUELLES)

Pour s'assurer que l'effet d'un facteur n'agit que sur la moyenne, on doit vérifier l'hypothèse d'invariance des variances résiduelles de chaque modalité du facteur étudié (et des combinaisons de facteurs) à l'aide du test de Bartlett qui utilise le critère du X^2 (consulter la partie 5 du cours du semestre 4).

Remarques :

- Il y aura autant de tests à réaliser que d'effets (principaux + interactions) à estimer dans le modèle du départ,
- Si le facteur ne comporte que 2 modalité on fera un test F bilatéral

ANNEXE 1 : LES DISPOSITIFS

- 2 facteurs étudiés sans répétition

→ 12 traitements → 12 unités stat.

A3 B1	A3 B2	A1 B1	A1 B2	A1 B3	A3 B4	A2 B1	A2 B4	A2 B3	A1 B4	A2 B2	A3 B3

- 2 facteurs étudiés avec répétitions

→ 12 traitmts répétés 2 fois → 24 unit. stat

103 A3 B2 r1	203 A3 B2 r2	103 A3 B1 r1	202 A2 B4 r2	201 A1 B3 r2	202 A2 B2 r2	201 A1 B4 r2	203 A3 B1 r2	102 A2 B3 r1	101 A1 B3 r1	202 A2 B3 r2	101 A1 B2 r1
102 A2 B2 r1	101 A1 B1 r1	202 A2 B1 r2	102 A2 B1 r1	101 A1 B4 r1	203 A3 B4 r2	201 A1 B1 r2	102 A2 B4 r1	103 A3 B4 r1	201 A1 B2 r2	103 A3 B3 r1	203 A3 B3 r2

- 3 facteurs étudiés sans répétition

A1 B2 C1	A2 B3 C5	A1 B2 C3	A3 B3 C2	A1 B2 C4	A3 B2 C1	A2 B2 C2	A1 B4 C1	A2 B1 C4	A2 B4 C4
A2 B4 C3	A2 B1 C2	A1 B4 C2	A3 B4 C4	A3 B2 C3	A2 B1 C5	A1 B1 C1	A3 B2 C2	A1 B3 C4	A2 B2 C4
A3 B4 C1	A1 B3 C3	A2 B3 C3	A3 B3 C3	A1 B4 C4	A3 B4 C5	A3 B4 C2	A1 B1 C2	A1 B1 C5	A2 B1 C1
A1 B2 C2	A2 B2 C5	A1 B3 C1	A3 B2 C4	A1 B4 C5	A3 B4 C3	A3 B1 C2	A1 B3 C2	A2 B3 C2	A2 B2 C3
A3 B1 C4	A1 B3 C5	A3 B1 C3	A1 B2 C5	A2 B3 C1	A3 B1 C5	A3 B1 C1	A3 B3 C5	A1 B1 C3	A2 B1 C3
A2 B4 C2	A2 B4 C5	A1 B4 C3	A3 B2 C5	A3 B3 C4	A2 B2 C1	A1 B1 C4	A2 B4 C1	A2 B3 C4	A3 B3 C1

ANNEXE 2 : LES TABLEAUX DE DONNEES

- 1 facteur étudié

x_{ir} → répétition → niveau du facteur

Répétitions	Niveaux du facteur A						
	A1	A2	...	Ai		...	Aa
1	x_{11}	x_{21}	...	x_{i1}		...	x_{a1}
2	x_{12}	x_{22}	...	x_{i2}		...	x_{a2}
.	.	.	.	.		.	.
.	.	.	.	.		.	.
	x_{1r}	x_{2r}	...	x_{ir}		...	x_{ar}
r	.	.		.			.
.	x_{1n1}	.		x_{ini}			.
.		x_{2n2}					.
.							x_{ana}
Sommes	$X_{1.}$	$X_{2.}$		$X_{i.}$			$X_{a.}$
Moyennes	$\overline{X}_{1.}$	$\overline{X}_{2.}$		$\overline{X}_{i.}$			$\overline{X}_{a.}$

une valeur de le groupe modalité

moyenne de modalité i

$\overline{X}_{..} = \mu$, niveau moy

① la différence entre une valeur et la moyenne d'une modalité est le résidu (ici e_{ir})

② différence α_i entre la moy de modalité i en la moyenne totale

- 2 facteurs étudiés sans répétition

	B1	B2	....	Bj	....	Bb	moyennes
A1	x_{11}	x_{12}	....	x_{1j}	....	x_{1b}	$\overline{x}_{1\bullet}$
A2	x_{21}	x_{22}	....	x_{2j}	....	x_{2b}	$\overline{x}_{2\bullet}$
	....	....	....	....	....	....	
Ai	x_{i1}	x_{i2}	....	x_{ij}	....	x_{ib}	$\overline{x}_{i\bullet}$
	....	....	....	....	....	....	
Aa	x_{a1}	x_{a2}	....	x_{aj}	....	x_{ab}	$\overline{x}_{a\bullet}$
moyennes	$\overline{x}_{\bullet 1}$	$\overline{x}_{\bullet 2}$		$\overline{x}_{\bullet j}$		$\overline{x}_{\bullet b}$	$\overline{x}_{\bullet\bullet}$

➢ 2 facteurs étudiés avec répétitions

	B1	B2	....	Bj	....	Bb	moyennes
A1	x_{111}	x_{121}	....	x_{1j1}	....	x_{1b1}	$\bar{X}_{1\bullet\bullet}$
	x_{112}	x_{122}	....	x_{1j2}	....	x_{1b2}	
	...	...	....	...	....	...	
	x_{11r}	x_{12r}	....	x_{1jr}	....	x_{1br}	
	...	...	....	...	....	...	
	x_{11n}	x_{12n}	....	x_{1jn}	....	x_{1bn}	
moyennes	$\bar{x}_{11\bullet}$	$\bar{x}_{12.}$		$\bar{x}_{1j.}$		$\bar{x}_{1b.}$	
A2	x_{211}	x_{221}	....	x_{2j1}	....	x_{2b1}	$\bar{X}_{2\bullet\bullet}$
	x_{212}	x_{222}	....	x_{2j2}	....	x_{2b2}	
	...	...	....	...	....	...	
	x_{21r}	x_{22r}	....	x_{2jr}	....	x_{2br}	
	...	...	....	...	....	...	
	x_{21n}	x_{22n}	....	x_{2jn}	....	x_{2bn}	
moyennes	$\bar{x}_{21.}$	$\bar{x}_{22.}$		$\bar{x}_{2j.}$		$\bar{x}_{2b.}$	
....	....	....	....	....	....	....	
Ai	x_{i11}	x_{i21}	....	x_{ij1}	....	x_{ib1}	$\bar{X}_{i\bullet\bullet}$
	x_{i12}	x_{i22}	....	x_{ij2}	....	x_{ib2}	
	...	...	....	...	....	...	
	x_{i1r}	x_{i2r}	....	x_{ijr}	....	x_{ibr}	
	...	...	....	...	....	...	
	x_{i1n}	x_{i2n}	....	x_{ijn}	....	x_{ibn}	
moyennes	$\bar{x}_{i1.}$	$\bar{x}_{i2.}$		$\bar{x}_{ij.}$		$\bar{x}_{ib.}$	
....	....	....	....	....	....	....	
Aa	x_{a11}	x_{a21}	....	x_{aj1}	....	x_{ab1}	$\bar{X}_{a\bullet\bullet}$
	x_{a12}	x_{a22}	....	x_{aj2}	....	x_{ab2}	
	...	...	....	...	....	...	
	x_{a1r}	x_{a2r}	....	x_{ajr}	....	x_{abr}	
	...	...	....	...	....	...	
	x_{a1n}	x_{a2n}	....	x_{ajn}	....	x_{abn}	
moyennes	$\bar{x}_{a1.}$	$\bar{x}_{a2.}$		$\bar{x}_{aj.}$		$\bar{x}_{ab.}$	
moyennes	$\bar{X}_{\bullet 1\bullet}$	$\bar{X}_{\bullet 2\bullet}$		$\bar{X}_{\bullet j\bullet}$		$\bar{X}_{\bullet b\bullet}$	$\bar{X}_{\bullet\bullet\bullet}$