# Stats Partie 6

I L'analyse de la variance sert dans la recherche et dvlp et plein d'autres trucs. On l'appelle ANOVA parce que c'est plus court.

II Les logiciels modélisent tout seuls en gros. Mais l'ingénieur doit bien comprendre les modèles.

- La réponse suit un modèle linéaire additif
  - expl : force sur une pomme → bah oui, linéaire
  - nbre bactéries → euh.. bas c'est plutôt exponentiel ⇒ alors on prend log.

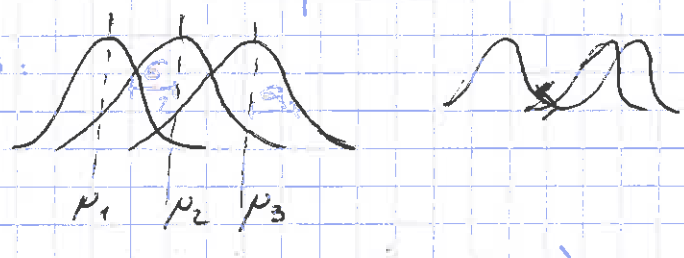
Si on teste une dose d'engrais on effet est positif ou negatif, mais s'ajoute

L'invariance de la variance : homoscédasticité des variances résiduelles
⇒ le facteur doit influe la moyenne, mais JAMAIS l'effet de variance unique : le cm

Donc : Comment comparer plusieurs moyennes ?
- existe-t-il une différence entre les moyennes ? → test ANOVA*
- lesquelles vont ensemble ? constitu° des groupes → ppds

↳ on procède à l'analyse des résidus :
- normalité
- homosédasticité
- independance

* ANOVA
On peut retrouver des constitutions :



on s'intéresse à la comparaison de variance intergroupe et variance intra groupe.

Mais il faut $\sigma_2 = \sigma_3 = \sigma_1$ qu'on écrira alors $CM_r$ ($=\sigma^2$)
↳ test homosédasticité

Ensuite, on compare $\dfrac{\text{var. inter}}{\text{var intra}} = F$

CM est une variance étudiée : $\dfrac{SCE}{ddl}$ c'est le carré moyen

exemple : avec un facteur $x_{ir} = \mu + \alpha_i + \varepsilon_{ir}$
Facteur étudié A, de a modalité et a moyennes

| répét | $A_1$ | $A_2$ | ... | $A_i$ | ... | $A_a$ | |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | | | | | | | |
| ⋮ | | | | | | | |
| r | | | | $x_{ir}$ | | | |
| ⋮ | | | | | | | |
| n | | | | | | | |
| $\bar{x}_{i.}$ | $\bar{x}_{1.} = \dfrac{X_1}{n}$ | $\bar{x}_{2.} = \dfrac{X_2}{n}$ | ... | $\bar{x}_{i.} = \dfrac{X_i}{n}$ | ... | $\bar{x}_{a.} = \dfrac{X_a}{n}$ | $\bar{x}_{..} = \mu$ |

NB : $X = \sum_1^a x_{i.}$

$N = \sum_1^a n_{i.} = n \times a$

$CM_A = \dfrac{SCE_A}{a-1}$

$SCE_A = \sum n(\bar{x}_{i.} - \bar{x}_{..})^2$

$= \sum \dfrac{X_{i.}^2}{n} - X\bar{x}_{..}$

$= \left(\dfrac{X_{1.}^2}{n} + \dfrac{X_{2.}^2}{n} \dots + \dfrac{X_{a.}^2}{n}\right) - \dfrac{X^2}{N}$

c'est C, le terme de centrage.

La variation totale des éléments est dûe à la fluctuation naturelle (des pommes traitées pareil ne sont pas égales) et la variation du traitement.

On écrit : variat. totale [illegible] [illegible]

$$SCE_{TOT} = SCE_A + SCE_e$$

donc $SCE_{e} = SCE_{TOT} - SCE_A$

or $SCE_{TOT} = \sum x_{ir}^2 - C$

(ça se compense chez les résidus)

et $SCE_e = \sum (e_{ir} - \bar{e})^2 = \sum e_{ir}^2$ $\left(\text{car } \sum \frac{e_{ir}}{e} = 0\right)$

→ Lorsqu'on a 2 facteurs (sans répét) : $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

le tableau donne :

| | $B_1$ | $B_2 \cdots B_j \cdots$ | $B_b$ | somme |
|---|---|---|---|---|
| $A_1$ | $x_{11}$ | $x_{12} \cdots$ | | $X_{1.}$ |
| $A_2$ | $x_{21}$ | | | $X_{2.}$ |
| $\vdots$ | | | | |
| $A_i$ | | | | $X_{i.}$ |
| $\vdots$ | | | | $\vdots$ |
| $A_a$ | | | | $X_{a.}$ |
| somme | $X_{.1}$ | $X_{.2} \cdots X_{.j} \cdots$ | $X_{.b}$ | $X$ |

alors $SCE_{Tot} = SCE_A + SCE_B + SCE_e$

variation totale, entre les facteurs A, inter fact B, intragroupe

$$CM_A = \frac{SCE_A}{a-1} \qquad CM_B = \frac{SCE_B}{b-1} \qquad CM_e = \frac{SCE_e}{ddl_e}$$

$$SCE_A = \sum_1^a \frac{X_{i.}^2}{b} - C$$

← pas [illegible] erreur ! bien regarder !

$$SCE_B = \sum_1^b \frac{X_{.j}^2}{a} - C$$

« Toujours pareil : somme des carrés moins terme de centrage. »

$$SCE_{TOT} = \sum_{i,j} x_{ij}^2 - C$$

→ Lorsqu'on a 2 facteurs (AVEC répét) : $x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_{ijr}$

on a $a \times b \times n$ unités stat (n → nb de répét)

$a \times b$ traitements différents

| | $B_1$ | $B_2$ | somme ↓ |
|---|---|---|---|
| $A_1$ | $x_{111}$<br>$x_{112}$<br>$x_{113}$<br>⋮<br>$x_{11n}$<br>$X_{11\cdot}$ | etc | $X_{1\cdot\cdot}$ |
| ⋮ etc | | | |
| somme | $X_{\cdot 1\cdot}$ | | |

alors $SCE_{TOT} = SCE_A + SCE_B + \underset{\substack{\uparrow \\ \text{variation due aux interactions}}}{SCE_{AB}} + SCE_e$

Il y aura 3 test F à faire :

$X = \sum x_{ijn}$
$N = abn$

$\dfrac{CM_A}{CM_e}$ et $\dfrac{CM_B}{CM_e}$ $\qquad \dfrac{CM_{AB}}{CM_e}$

avec $CM_A = \dfrac{SCE_A}{a-1}$ $\qquad CM_B = \dfrac{SCE_B}{b-1}$

$$CM_{AB} = \frac{SCE_{AB}}{(a-1)(b-1)}$$

$$SCE_A = \sum_{i=1}^{a} \frac{X_{i\cdot\cdot}^2}{bn} - C$$

$$SCE_B = \sum_{i=1}^{b} \frac{X_{\cdot j\cdot}^2}{an} - C$$

$$SCE_{TOT} = \sum x_{ijn}^2 \cancel{/abn} - C$$

$$SCE_{AB} = \sum \frac{X_{ij\cdot}^2}{n} - SCE_A - SCE_B$$

→ Avec 3 facteurs étudiés sans répéta :

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \varepsilon_{ijk}$$

(avec répéta on ajout $\alpha_i\beta_j\gamma_k$)

donc sans répét :

$$SCE_C = \sum_{1}^{c} \frac{X_{\cdot\cdot k}^2}{ab} - C$$

$$SCE_{\cancel{A}BC} = \sum_{j,k}^{bc} \frac{X_{\cdot jk}^2}{a} - C - SCE_B - SCE_C$$

$$SCE_{TOT} = \sum_{i,j,k}^{abc} x_{i,j,k}^2 - C$$

$$SCE_e = SCE_{TOT} - SCE_A - SCE_B - SCE_C - SCE_{AB} - SCE_{AC} - SCE_{BC}$$

# VI L'analyse des résidus

L'analyse de la variance permettant d'affirmer ou non des facteurs ou interactions.

Avant de procéder à cette analyse, il faut étudier des hypothèses :

exemple : l'erreur liée à une fluctuation doit garder le même ordre de grandeur.
Le CME doit être fiable

→ en fait, on vérifie des hypothèses sur les résidus.

Hypothèses à vérifier sur les résidus

1 → suivent une loi normale
2 → sont indpd des résidus
3 → présente une homoscédasticité sur les variances résiduelles

## 1) Normalité ?

Reprenons :

1 facteur

$$x_{ir} = \mu + \underset{\substack{\uparrow \\ \text{variance dûe au} \\ \text{facteur A à la modalité } i}}{\alpha_i} + \underset{\text{fluctuation naturelle}}{\varepsilon_{ir}}$$

$$x_{ir} = \mu + (\bar{x}_i - \mu) + \varepsilon_{ir}$$

$$- x_{ir} + \mu + \bar{x}_i - \mu = -\varepsilon_{ir}$$

$$\bar{x}_i - x_{ir} = -\varepsilon_{ir} \quad \text{soit } \varepsilon_{ir} = x_{ir} - \bar{x}_i$$

et c'est logique : résidu = réel − impact des facteurs attendus

2 facteurs sans interactions

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

$$x_{ij} = \mu + (\bar{x}_{i.} - \mu) + (\bar{x}_{.j} - \mu) + \varepsilon_{ij}$$

Ainsi $\varepsilon_{ij} = x_{ij} + \mu - \bar{x}_{i.} - \bar{x}_{.j}$

2 facteurs avec interaction

$\varepsilon_{ijr} = x_{ijr} - \bar{x}_{ij.}$ (c'est censé être intuitif)

Il y a autant de résidu que d'unités statistiques (cases du tableau)
On aura donc un tableau spécial résidus ! ouais ! !!

On étudie la normalité des résidus. L'histogramme nous sera donné (étudié en 1A)
Il faudra déterminer si il est normal :

- premier aperçu avec la courbe (en cloche et symétrique)
- approfondissement avec les coef. de Pearson $\beta_1$ et $\beta_2$ (donnés aussi car étudiés en 1A)

Les tests sur ces coef. donnent des p-values (données aussi)
On doit savoir conclure sur la p-value !

## ② Indépendance

$SCE_e = SC_e$ 
case $\Sigma(e_i - \bar{e})^2$ mais $\bar{e} = 0$

« $SCE_e = SC_e$ est une propriété magique »

La fluctuation naturelle est elle bien indpd des cobayes ?
- sur les animaux, on peut pas vérifier
- sur les parcelles, on peut vérifier avec une cartographie.

Si la répartition des résidus semble aléatoire sur l'ensemble de la carte, alors RAS.

## ③ Homoscédasticité

On a 2 facteurs avec répét

$e_{111}$
$e_{11r}$
$e_{11n}$

| | $B_1$ ... | $B_j$ ... | $B_b$ ... | |
|---|---|---|---|---|
| $A_1$ | $\sigma^2_{11\cdot}(\varepsilon)$ $e_{111}$ $e_{118}$ $e_{11n}$ | | | $\sigma^2_{1\cdot\cdot}(\varepsilon)$ |
| $A_i$ | | | | $\sigma^2_{i\cdot\cdot}(\varepsilon)$ |
| $A_a$ | | | | $\sigma^2_{a\cdot\cdot}(\varepsilon)$ |
| | $\sigma^2_{\cdot 1\cdot}(\varepsilon)$ | $\sigma^2_{\cdot i\cdot}(\varepsilon)$ | $\sigma^2_{\cdot b\cdot}(\varepsilon)$ | |

Il y a autant de test d'homoscédasticité que de tests F dans l'ANOVA.

Les tests d'homoscédasticité sont de 2 sortes :
- $\chi^2$, unilatéral, plusieurs variances
- F bilatéral (si 2 variances à tester seulmt)

Chaque variance résiduelle est la $SCE_{e_{i\cdot\cdot}}$ divisé par $(n_{i\cdot\cdot} - 1)$ (pour les lignes)

*Exemple : exercice 3*

facteur A : appareils, 4 modalités $a = 4$
facteur B : temps, 2 modalités $b = 2$
} 8 traitements donc 8 $\sigma^2(\varepsilon)$
répétitions : $n = 3$

Pour
- A : + de 2 facteurs → test $\chi^2$ $ddl = a - 1 = 3$
- B : 2 facteurs → test F $F_{1-\frac{\alpha}{2}}(11,11)$
- répet : + de 2 facteurs → test $\chi^2$ $ddl = 8 - 1 = 7$

$\chi^2$ → Au moins une des variances résiduelles est plus grande que les autres ? $H_0$ : toutes les $\sigma$ sont $=$ $H_1$

F → $H_0 = \frac{\sigma_1^2(\varepsilon)}{\sigma_2^2(\varepsilon)}$ les variances sont différentes ?

| Pour $\chi^2_{calc}$ | $\nu_j$ | $\nu$ | $\frac{1}{\nu}$ | $\Sigma \frac{1}{\nu_j}$ |
|---|---|---|---|---|
| $\chi^2$ A | $6-1=5$ | $4 \times 5 = 20$ | $\frac{1}{20}$ | $\frac{4}{5}$ |
| $F_{calc}$ B | | | | |
| $\chi^2$ A×B | $3-1=2$ | $8 \times 2 = 16$ | $\frac{1}{16}$ | $\frac{8}{2}$ |

STAGE : faire tous les SCE α(1, 2, 3, 4, 5) SCEα et SCE, et SCE ...

| | |
|---|---|
| isaralyon<br>Une école d'ingénieurs au coeur de la vie | **ISARA – 2ème année – STATISTIQUE**<br>**Exercices : Partie 6**<br>**Mme B. Bottollier-Lemallaz** |

# ANALYSE DE LA VARIANCE

Pour chaque exercice vous préciserez le modèle mathématique et définirez les termes du modèle.

## Exercice 1

On a réparti au hasard 15 cobayes en 3 lots afin de tester leur prise pondérale dans 3 situations de température d'environnement. Les gains de poids moyens au cours d'une même période sont exprimés en grammes, on admet que le gain de poids d'un animal suit une loi normale.

Moy 1 → 43,5
2 → 49,78
3 →

| T1 (5°C) | T2 (15°C) | T3 (22°C) | som |
|---|---|---|---|
| 42.5 | 48.2 | 58.3 | 149 |
| 44.6 | 54.2 | 59.1 | 157,9 |
| 42.1 | 43.1 | 56.9 | 142,1 |
| 45.1 | 48.3 | 52.2 | 145,6 |
| 43.2 | 55.1 | 54.1 | 152,4 |
| 217,5 | 248,9 | 2[illegible]6 | ≈ 747 X |

moyenne → 49,8

1°) Décrivez le contexte de l'étude (modèle, nombre facteur étudié, identifier des facteurs contrôlés, nombre d'unité, nombre de traitement, sources des fluctuations naturelles ...)
2°) Vérifier l'hypothèse d'homoscédasticité
3°) Réaliser l'analyse de la variance
4°) Comparer les moyennes et les grouper

## Exercice 2

Le projet d'étude proposé a pour principal objectif d'évaluer les différences dans la composition de la flore mésophile aérobie (FAM) présente sur les poissons dans 5 étangs pour 2 saisons (été, hiver). Les étangs suivis sont situés dans des zones géographiques variées de la Dombes. Les sites sur lesquels les carpes ont été prélevées ont fait l'objet d'un suivi de la température de l'eau en continu.
Sur le terrain, dans chaque étang, 5 poissons ont été capturés au moyen de filets maillants de type araignée. Aussitôt après la capture, au laboratoire les poissons ont été tué, un prélèvement de mucus a été effectué par raclage avec une spatule en conditions aseptiques sur le côté gauche du poisson vivant : flanc et nageoires. Pour avoir une quantité suffisante, les 5 mucus des 5 poissons d'un même étang ont été déposés dans un même pot à prélèvement numéroté et pesé.

1°) D'après ce contexte donnez la listes des facteurs étudié (et leur nombre de modalités) et ceux mis sous contrôle. Proposer des facteurs non contrôlés responsables de l'erreur expérimentale.

2°) Les résultats ci dessous ont été obtenus.

2 facteurs → 2 tests d'homoscédasticité

| Etang | Saison | FAM | ln |
|---|---|---|---|
| E1 | Ete | 1640000 | 14,310 |
| E1 | Hiver | 1374000 | 14,133 |
| E2 | Ete | 52200 | 10,863 |
| E2 | Hiver | 15900 | 9,674 |
| E3 | Ete | 156900 | 11,963 |
| E3 | Hiver | 118000 | 11,678 |
| E4 | Ete | 896000 | 13,706 |
| E4 | Hiver | 325100 | 12,692 |
| E5 | Ete | 268000 | 12,499 |
| E5 | Hiver | 41600 | 10,636 |

saison → 2 modalités → test F
étang → 5 modalités → test $\chi^2$

1

La variable FAM ne suit pas une loi normale car il s'agit du comptage de cellules qui se multiplient par scissiparité. Pour palier à cet inconvénient ont va utiliser le LN (log népérien) du résultat avec une précision au 1/1000ème. Dans ce cas on a pu vérifier la normalité des résidus.

a) Vérifier l'hypothèse d'homoscédasticité (résultats au 1/100$^{ème}$)
b) Réaliser l'analyse de la variance (résultats au 1/100$^{ème}$)
c) Comparer et grouper les moyennes.

## Exercice 3

Les résultats suivants, exprimés en µA, ont été obtenus en analysant par polarographie une substance herbicide à 2 températures différentes, avec 4 appareils de marques différentes et, à chaque fois avec 3 répétitions. On admet que la variable mesurée et la série des résidus suivent une loi normale.

| | A1 | A2 | A3 | A4 |
|---|---|---|---|---|
| 25° C | 3.58 | 3.47 | 3.50 | 3.54 |
| | 3.53 | 3.44 | 3.48 | 3.57 |
| | 3.48 | 3.50 | 3.52 | 3.60 |
| 45° C | 3.55 | 3.59 | 3.59 | 3.62 |
| | 3.56 | 3.55 | 3.58 | 3.63 |
| | 3.58 | 3.63 | 3.57 | 3.60 |

$\sigma(\varepsilon)$ $\Sigma(e)^2$

1°) Vérifier toutes les hypothèses d'homoscédasticité. On donne déjà :

ECARTS-TYPES FACTEUR 1 = A

| 1 (A1) | 2 (A2) | 3 (A3) | 4 (A4) |
|---|---|---|---|
| 0,0331 | 0,0316 | 0,0141 | ? |

KHI2 = ?

ECARTS-TYPES INTER F1*2 = A T

| | 1 (A1) | 2 (A2) | 3 (A3) | 4 (A4) |
|---|---|---|---|---|
| 1 (T1) | 0,0500 | 0,0300 | 0,0200 | 0,0300 |
| 2 (T2) | 0,0153 | 0,0400 | 0,0100 | 0,0153 |

KHI2 = 6.29 PROBA =.5079

2°) Réaliser l'analyse de la variance.
3°) Regrouper les moyennes, conclure

## Exercice 4

Dans une expérience sur le rendement de vaches laitières, on a choisi 40 animaux aussi identiques que possible, et on les a répartis en 8 groupes (notés de 1 à 8) de 5 ; chaque groupe a été soumis à une alimentation différente. Les facteurs testés sont la nature de l'aliment (A1 : paille ; A2 : foin ; A3 : herbe, A4 : aliments ensilés) et la dose de l'aliment (D1 : dose forte ; D2 : dose faible). L'ensemble des mesures est rassemblé dans le tableau ci-après. On admettra que la variable étudiée suit une loi normale.

| A1 | A1 | A2 | A2 | A3 | A3 | A4 | A4 |
|---|---|---|---|---|---|---|---|
| D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 |
| 8 | 8 | 12 | 10 | 12 | 11 | 14 | 17 |
| 10 | 9 | 13 | 12 | 10 | 9 | 17 | 19 |
| 11 | 10 | 11 | 10 | 13 | 11 | 13 | 17 |
| 10 | 8 | 14 | 11 | 12 | 11 | 14 | 16 |
| 7 | 9 | 10 | 7 | 14 | 12 | 17 | 21 |

1°) Quel est le modèle associé à cette expérimentation ? Préciser chacun des termes du modèle.
2°) Combien de tests d'homoscédasticité doit on réaliser ? Préciser les critères statistiques et les ddl.
3°) On suppose que les hypothèses de normalité des résidus et d'homoscédasticité sont vérifiées. Construisez le tableau de l'ANOVA, posez les hypothèses et concluez au risque $\alpha = 0.05$
4°) Quel type d'alimentation et quelle dose proposeriez vous pour maximiser le rendement ?

## Exercice 5

Les résultats suivants proviennent du dosage du $C^{14}$ effectué à 3 dates différentes (i; ii; iii), considérées comme aléatoire, et dans 2 substances (a; b), associées chacune à 4 solvants (1; 2; 3; 4). On admettra que la variable étudiée suit une loi normale. Procéder à l'analyse de ces résultats au seuil de 5%.

| subst. et solvant | date (i) | date (ii) | date (iii) |
|---|---|---|---|
| a1 | 15 | 16 | 16 |
| b1 | 18 | 18 | 18 |
| a2 | 15 | 16 | 14 |
| b2 | 13 | 15 | 14 |
| a3 | 17 | 19 | 15 |
| b3 | 18 | 18 | 17 |
| a4 | 17 | 17 | 13 |
| b4 | 15 | 16 | 17 |

1°) Quel est le modèle associé à cette expérimentation ? Préciser chacun des termes du modèle.
2°) On suppose que les hypothèses de normalité des résidus et d'homoscédasticité sont vérifiées. Construisez le tableau de l'ANOVA, posez les hypothèses et concluez au risque $\alpha = 0.05$
3°) Identifiez les solvants équivalents.

4 solvants → test $\chi^2$ } 8 combinaisons (solv × subs)
2 substances → test F
~~3 dates~~ → test
3 dates

Pour les interactions → toujours $\chi^2$

Résidus
4 solv → 4 $\sigma^2(\varepsilon)$ }
2 subs → 2 $\sigma^2(\varepsilon)$ } N = 24
3 dates → 3 $\sigma^2(\varepsilon)$ }

solv × sub → 8 combinaisons — 8 $\sigma^2(\varepsilon)$
solv × date → 6 — 6
subs × date → 12 — 12

Δ test F →

| Pour $\chi^2$ calc | $\nu_j$ | $\nu = \sum \nu_j$ | $\frac{1}{\nu}$ | $\sum \frac{1}{\nu_j}$ | ddl pour $\chi^2$ théo |
|---|---|---|---|---|---|
| solv | 6 - 1 = 5 | 4×5 = 20 | $\frac{1}{20}$ | $\frac{4}{5}$ | 3 |
| subs | 12 - 1 = 11 | | | | (11 ; 11) |
| dates | 8 - 1 = 7 | 3×7 = 21 | $\frac{1}{21}$ | $\frac{3}{7}$ | 2 |
| solv × subs | 3 - 1 = 2 | 8×2 = 16 | $\frac{1}{16}$ | $\frac{8}{2}$ | 7 |
| subs × date | 4 - 1 = 3 | 6×3 = 18 | $\frac{1}{18}$ | $\frac{6}{3}$ | 5 |
| solv × date | 2 - 1 = 1 | 12×1 = 12 | $\frac{1}{12}$ | $\frac{12}{1}$ | 11 |

Examen- Mai 2010 (durée 1h)

Le tableau suivant reprend les rendements observés ($x_{ij}$), en tonnes par hectare, au cours d'un essai en blocs(*) aléatoires complets (j = 1 à 4) destiné à comparer six variétés de froment (i = 1 à 6) et définir la (les) plus rentable(s).
(*) Ici, on appelle « bloc » un « agriculteur » qui dispose de 6 parcelles pour réaliser cette expérimentation.

| | Blocs | | | | |
|---|---|---|---|---|---|
| variétés | Bloc 1 | Bloc 2 | Bloc 3 | Bloc 4 | somme |
| 1 | 5.02 | 5.37 | 5.41 | 5.54 | 21.34 |
| 2 | 4.92 | 5.28 | 5.00 | 5.16 | 20.36 |
| 3 | 5.28 | 4.84 | 5.52 | 5.01 | 20.65 |
| 4 | 4.71 | 4.83 | 4.67 | 4.79 | 19.00 |
| 5 | 4.55 | 5.07 | 5.14 | 4.97 | 19.73 |
| 6 | 4.77 | 5.07 | 4.98 | 4.99 | 19.81 |
| somme | 29.25 | 30.46 | 30.72 | 30.46 | 120.89 |

On donne la somme des carrés des 24 valeurs = 610.5341
1°) (2 pts) Combien y a-t-il de facteurs ? ..................de traitements ? ..............................d'unités statistiques ?.............................
Ecrivez le modèle théorique supposé associé à l'étude, préciser la signification de chaque terme.
2°) (4 pts) A partir des éléments donnés ci-dessous, démontrer qu'on ne peut refuser l'homogénéité des variances résiduelles, vous conclurez à partir de l'encadrement des « p-value » , posez les hypothèses soumises au test.

*On donne pour les variétés : $X^2$ calculé = 8,315*
*On donne aussi :*

| Blocs | SCEe.j |
|---|---|
| 1 (B1) | 0,1655 |
| 2 (B2) | 0,1710 |
| 3 (B3) | 0,1505 |
| 4 (B4) | 0,0652 |

3°) (2 pts) Que conclure à partir de ces résultats donnés par STATBOX ?

| | |
|---|---|
| Beta 1 = 0,074 | Prob. : 0,565 |
| Beta 2 = 2,889 | Prob. : 0,904 |

4°) (1 pt) En donnant le détail, calculez la valeur de $e_{4,3}$ (utiliser l'arrondi au 1/1000ème pour les calculs intermédiaires):
Pour la suite on admettra que toutes les conditions portant sur les résidus sont respectées.
5°) (5 pts) Le(s) facteur(s) ont-ils des effets significatifs ?. Hypothèses, conclusions et calculs au 1/1000ème.
6°) (5 pts) A l'aide de la méthode de Newman Keuls, pour un risque d'erreur de 5%, constituer les groupes de variétés ? (calculs intermédiaires au 1/10000ème).
7°) (1 pt) Conclusion de l'étude

# Examen- Mai 2011

**Exercice d'application (30 min) :**

On vous propose d'apporter une réponse à la question posée : « Quelle orientation des ruches et quel type de pollen faut-il mettre à disposition des abeilles pour obtenir une quantité maximale de miel ? ».
Pour cela on dispose de 8 serres et de 24 ruches. On choisit 2 types d'orientation (O 1 et O 2) et 4 types de pollen (P 1 à P 4).Trois ruches sont disposées, selon une orientation définie, dans 1 serre à l'intérieur de laquelle on a disposé les fleurs artificielles en tissu remplies d'un type de pollen. On a utilisé des pelotes d'un pollen que l'on a enrichi avec une substance à tester pour former les 4 types de pollen. A la fin de l'expérimentation on mesure la quantité de miel produite dans chaque ruche.

1) Compléter les propositions suivantes à l'aide de nombres :

   Cette étude est une étude à ........facteurs étudiés avec........répétitions. Il y a ..........traitements à tester à répartir sur un total de ............ unités expérimentales.

2) Les résultats des tests sur les coefficients de Pearson sont résumés dans le tableau suivant :

| coefficient | p-value | Formuler les hypothèses et la conclusion relative au résultat de béta 2 uniquement |
|---|---|---|
| Beta 1 | 0,806 | |
| Beta 2 | 0,45 | |

3) Les résultats des tests du $X^2$ de l'étude sont résumés dans le tableau suivant :

| | $X^2$ calculé | Quelles sont les hypothèses et la conclusion par rapport au résultat de O*P pour $\alpha= 0{,}05$? Justifier. |
|---|---|---|
| O | 1,917 | |
| P | 1,723 | |
| O*P | 2,8 | |

4) Citez les 3 hypothèses qui ont du être validées pour pouvoir réaliser une ANOVA.
5) Etablissez le tableau de l'ANOVA. Hypothèses et conclusions pour alpha = 5%.

On donne :

| somme des répétitions | O 1 | O2 |
|---|---|---|
| P 1 | 67 | 97 |
| P 2 | 37 | 106 |
| P 3 | 73 | 106 |
| P 4 | 69 | 101 |

Somme des carrés de toutes les valeurs = 19492
Somme de toutes les valeurs = 656

6) Mettez en œuvre une méthode statistique qui permette de répondre à la question posée au début de l'étude pour un risque d'erreur de 5%. Conclusion.

Excel

→ Options, → compléments → Atteindre → Analys Tool pack

Puis → Données → Utilitaire d'analyse

Saisir le tableau sous forme

| | D1 | D2 |
|---|---|---|
| A1 | | |
| A1 | | |
| A1 | | |
| A1 | | |
| A1 | | |
| A2 | | |
| A2 | | |
| A2 | | |
| A2 | | |
| A2 | | |

Sélectionner le tableau

puis → Utilitaire d'analyse

**Ecart type des résidus :**

| Ecarts-types facteur 1 = T | E.T. |
|---|---|
| 1 (T25) | 0,029 |
| 2 (T45) | 0,020 |

*$khi^2$ = 1,539 Prob. = 0,212*

$X^2_{0,95(1)}$ = 3,84

| Ecarts-types facteur 2 = A | E.T. |
|---|---|
| 1 (A1) | 0,033 |
| 2 (A2) | 0,032 |
| 3 (A3) | 0,014 |
| 4 (A4) | 0,021 |

$khi^2$ = 3,728 Prob. = 0,29182

$X^2_{0,95(3)}$ = 7,81

| Ecarts-types inter f1*2 = T * A | 1 (A1) | 2 (A2) | 3 (A3) | 4 (A4) |
|---|---|---|---|---|
| 1 (T25) | 0,050 | 0,030 | 0,020 | 0,0300 |
| 2 (T45) | 0,015 | 0,040 | 0,010 | 0,0150 |

*$khi^2$ = 6,285 Prob. = 0,5079*

$X^2_{0,95(7)}$ = 14,1

# Exercice 4

**Histogramme des résidus :**

→ comme ça on s'dit "Hum. c'est un genre de bimodal ça non ?"

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | | | 504 | | 308 | 501 | | |
| 7 | | | 304 | | 202 | 303 | | |
| 6 | | 104 | 502 | | 206 | 101 | 204 | |
| 5 | | 306 | 102 | | 508 | 203 | 302 | |
| 4 | | 506 | 405 | | 207 | 208 | 106 | |
| 3 | | 305 | 205 | | 401 | 103 | 403 | |
| 2 | | 406 | 107 | 105 | 301 | 201 | 108 | |
| 1 | 408 | 307 | 402 | 505 | 407 | 507 | 503 | 404 |

**Résidus suspects (méthode de GRUBBS) :**

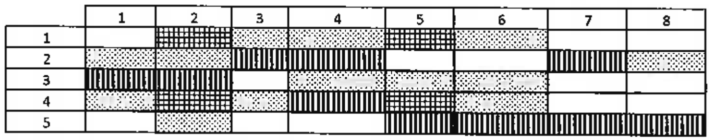
Aucun résidu suspect

**Indices de normalité (coefficients de K.PEARSON) :**

Beta 1 = 0,004 Prob. : 0,868 → on dirait symétrique oui

Beta 2 = 2,236 Prob. : 0,297 → proba > à 5% donc on garde la mésocurtie MAIS méfiance...

**Cartographie des résidus :**



Légende :
- < - 0,981
- < 0,000
- < 0,981
- < 999999,0

**Ecart type des résidus :**

| Ecarts-types facteur 1 = A | E.T. |
|---|---|
| 1 (A1) | 1,229 |
| 2 (A2) | 1,633 |
| 3 (A3) | 1,229 |
| 4 (A4) | 1,826 |

*$khi^2$ = 2,091 Prob. = 0,55739*

$X^2_{0,95(3)}$ = 7,81

| Ecarts-types facteur 2 = D | E.T. |
|---|---|
| 1 (D1) | 1,515 |
| 2 (D2) | 1,407 |

*$khi^2$ = 0,101 Prob. = 0,74887*

| Ecarts-types inter f1*2 = A * D | 1 (D1) | 2 (D2) |
|---|---|---|
| 1 (A1) | 1,643 | 0,837 |
| 2 (A2) | 1,581 | 1,871 |
| 3 (A3) | 1,483 | 1,095 |
| 4 (A4) | 1,871 | 2,000 |

*$khi^2$ = 3,781 Prob. = 0,80603*

$X^2_{0,95(7)}$ = 14,1



# Exercice5

| Moyennes facteur 1 = Sub | Moyenne |
|---|---|
| 1 (sub a) | 15,833 |
| 2 (sub b) | 16,417 |

| Moyennes inter f1*2 = Sub * Sol | 1 (sol 1) | 2 (sol 2) | 3 (sol 3) | 4 (sol 4) |
|---|---|---|---|---|
| 1 (sub a) | 15,667 | 15,000 | 17,000 | 15,667 |
| 2 (sub b) | 18,000 | 14,000 | 17,667 | 16,000 |

| Moyennes facteur 2 = Sol | Moyenne |
|---|---|
| 1 (sol 1) | 16,833 |
| 2 (sol 2) | 14,500 |
| 3 (sol 3) | 17,333 |
| 4 (sol 4) | 15,833 |

| Moyennes inter f2*3 = Sol * D | 1 (d1) | 2 (d2) | 3 (d3) |
|---|---|---|---|
| 1 (sol 1) | 16,500 | 17,000 | 17,000 |
| 2 (sol 2) | 14,000 | 15,500 | 14,000 |
| 3 (sol 3) | 17,500 | 18,500 | 16,000 |
| 4 (sol 4) | 16,000 | 16,500 | 15,000 |

| Moyennes facteur 3 = D | Moyenne |
|---|---|
| 1 (d1) | 16,000 |
| 2 (d2) | 16,875 |
| 3 (d3) | 15,500 |

| Moyennes inter f1*3 = Sub * D | 1 (d1) | 2 (d2) | 3 (d3) |
|---|---|---|---|
| 1 (sub a) | 16,000 | 17,000 | 14,500 |
| 2 (sub b) | 16,000 | 16,750 | 16,500 |

**Indices de normalité (coefficients de K.PEARSON) :**

Symétrie (valeur idéale théorique = 0) : Beta 1 = 0,000 Prob. : 1,000

Aplatissement (valeur idéale théorique = 3) : Beta 2 = 2,330 Prob. : 0,466

→ loi normale

**Résidus suspects (méthode de GRUBBS) :**

Aucun résidu suspect

| | | | | |
|---|---|---|---|---|
| 8 | | 107 | 103 | |
| 7 | | 106 | 123 | |
| 6 | | 109 | 113 | |
| 5 | | 111 | 116 | |
| 4 | 118 | 102 | 122 | 117 |
| 3 | 101 | 115 | 119 | 114 |
| 2 | 108 | 110 | 104 | 120 |
| 1 | 112 | 121 | 105 | 124 |
| Effectifs | | | | |
| | 4 | 8 | 8 | 4 |
| Bornes | | | | |
| | -1,12 | -0,56 | 0,0 | 0,56 |

**Ecart type des résidus :**

| Ecarts-types facteur 1 = Sub | E.T. |
|---|---|
| 1 (sub a) | 0,600 |
| 2 (sub b) | 0,600 |

*$khi^2$ = 0,000 Prob. = 0,99*

| Ecarts-types inter f1*2 = Sub * Sol | 1 (sol 1) | 2 (sol 2) | 3 (sol 3) | 4 (sol 4) |
|---|---|---|---|---|
| 1 (sub a) | 0,781 | 0,361 | 0,439 | 1,023 |
| 2 (sub b) | 0,781 | 0,361 | 0,439 | 1,023 |

*$khi^2$ = 4,447 Prob. = 0,7289*

| Ecarts-types facteur 2 = Sol | E.T. |
|---|---|
| 1 (sol 1) | 0,698 |
| 2 (sol 2) | 0,323 |
| 3 (sol 3) | 0,393 |
| 4 (sol 4) | 0,915 |

*$khi^2$ = 6,093 Prob. = 0,10545*

| Ecarts-types inter f2*3 = Sol * D | 1 (d1) | 2 (d2) | 3 (d3) |
|---|---|---|---|
| 1 (sol 1) | 0,884 | 0,354 | 1,237 |
| 2 (sol 2) | 0,295 | 0,589 | 0,295 |
| 3 (sol 3) | 0,648 | 0,589 | 0,059 |
| 4 (sol 4) | 1,237 | 0,354 | 1,591 |

*$khi^2$ = 8,205 Prob. = 0,696*

| Ecarts-types facteur 3 = D | E.T. |
|---|---|
| 1 (d1) | 0,635 |
| 2 (d2) | 0,367 |
| 3 (d3) | 0,770 |

*$khi^2$ = 3,343 Prob. = 0,18574*

| Ecarts-types inter f1*3 = Sub * D | 1 (d1) | 2 (d2) | 3 (d3) |
|---|---|---|---|
| 1 (sub a) | 0,685 | 0,397 | 0,832 |
| 2 (sub b) | 0,685 | 0,397 | 0,832 |

*$khi^2$ = 2,698 Prob. = 0,74888*

**ISARA - 2ème année - STATISTIQUE**
**Compléments exercices : Partie 6**
Mme B. Bottollier-Lemallaz

# ANALYSE DE LA VARIANCE

## Exercice 1

**Indices de normalité (coefficients de K.PEARSON) :** ① $\beta_1 \simeq 0$ → symétrique
② $\beta_2$ : si on rejete $H_0$, → a risque de se tromper de → mésocurtique

BETA 1 = 0.05 PROBA = 0.6858
BETA 2 = 2.81 PROBA = 0.8668

③ c'est à peu près symétrique, n'infirme pas les théories ci-dessus.

**Histogramme des résidus :**

nombre de résidus

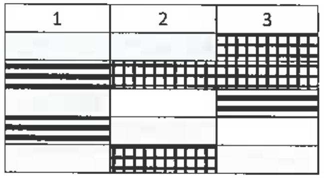
| | | | | |
|---|---|---|---|---|
| 5 | | 403 | 302 | |
| 4 | | 301 | 402 | |
| 3 | | 303 | 502 | 203 |
| 2 | 503 | 101 | 401 | 202 |
| 1 | 102 | 501 | 201 | 103 |

différentes classes

**Cartographie des résidus :**

Légende :

< - 2,132
< 0,000
< 2,132
> 2,132



Méthode de GRUBBS (si une valeur est out, la valeur est extrême et fausse la mesure, faut remplacer)

**Résidus suspects (méthode de GRUBBS) :**
Aucun résidu suspect

## Exercice 2

**Résidus suspects (méthode de GRUBBS) :**
Aucun résidu suspect

**Histogramme des résidus :**

| | | | | |
|---|---|---|---|---|
| 3 | 110 | | | 108 |
| 2 | 105 | 104 | 107 | 102 |
| 1 | 109 | 106 | 101 | 103 |

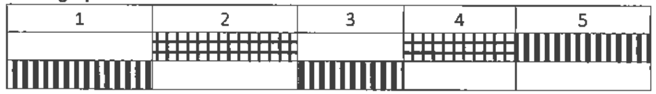
→ symétrique mais pas en cloche distribution en ().
en même temps, avec seulement 10 valeurs, c'est moyen représentatif

**Indices de normalité (coefficients de K.PEARSON) :** → $\beta_1 \simeq 0$ et si on rejète → 100% chance de se tromper
Beta 1 = 0,000 Prob. : 1,000
Beta 2 = 1,724 Prob. : 0,339

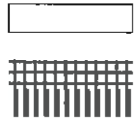
⇒ symétrique
$\beta = 0$? si on dit non, on a 33% chances de se tromper
⇒ méso curtique



**Cartographie des résidus :**



Légende :



< - 0,222
< 0,000
< 0,222
>0,222

| Ecarts-types facteur 1 = E | |
|---|---|
| | E.T. |
| 1 (E1) | 0,515 |
| 2 (E2) | 0,200 |
| 3 (E3) | 0,439 |
| 4 (E4) | 0,077 |
| 5 (E5) | 0,677 |

*$khi^2$ = 2,763 Prob. = 0,60121*
$\chi^2_{0,95(4)}$ = 9,49

| Ecarts-types facteur 2 = S | |
|---|---|
| | E.T. |
| 1 (Eté) | 0,347 |
| 2 (Hiver) | 0,347 |

*$khi^2$ = 0,000 Prob. = 0,99*
$\chi^2_{0,95(1)}$ = 3,84

## Exercice 3

**Indices de normalité (coefficients de K.PEARSON) :** → $\beta_1 \simeq 0$ [...] → symétrie
Beta 1 = 0,000 Prob. : 1,000
Beta 2 = 2,725 Prob. : 0,765

$\beta_2$, 76% de chance de se tromper en rejetant $H_0$
⇒ mésocurtique

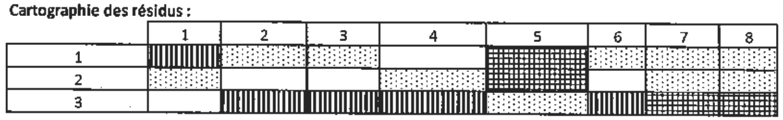
**Résidus suspects (méthode de GRUBBS) :**
Aucun résidu suspect

⇒ Normalité

**Histogramme des résidus :**

| | | | | |
|---|---|---|---|---|
| 11 | | 107 | | |
| 10 | | 307 | | |
| 9 | | 203 | | |
| 8 | | 103 | | |
| 7 | | 301 | | |
| 6 | | 303 | | |
| 5 | | 105 | 306 | |
| 4 | 302 | 102 | 207 | 108 |
| 3 | 305 | 308 | 304 | 106 |
| 2 | 101 | 205 | 202 | 204 |
| 1 | 104 | 208 | 206 | 201 |

**Cartographie des résidus :**



< - 0,017
< 0,000
< 0,017
> 0,017

# Stats - Part 6 Exercices

## Exercice 1

→ Bartlett
→ à partir de qd on travaille ln ?

1) Contexte de l'étude : Modèle à 1 facteur, 3 modalités | ⇔ : 3 traitements, 5 répétitions → 15 unités d'expérience

Réponse mesurée : gain de poids

2) Test d' homoscédasticité... (~~= analyse des résidus ?~~)
→ invariance des variances

Bref, on va faire ça au feeling hein.

$$\mathbb{C} = \frac{(\text{somme de tout})^2}{\text{effectif}}$$

Données :

| T1 | T2 | T3 |
|---|---|---|
| 42,5 | 48,2 | 58,3 |
| 44,6 | 54,2 | 59,1 |
| 42,1 | 43,1 | 56,9 |
| 45,1 | 48,3 | 52,2 |
| 43,2 | 55,1 | 54,1 |

sommes → 217,5 + 248,9 + 280,6 = 747

$$\mathbb{C} = \frac{747^2}{15} = 37\,200,6$$

3) *ANOVA (étude du facteur "fac")

| Sources de variation | SCE | ddl | CM | Fcalc | Fthéo *2 |
|---|---|---|---|---|---|
| facteur "fac" | 398,164 ① | 2 ② | 199,082 ③ | ⑧ 17,36 | 3,89 |
| résidus | 137,656 ④ | 12 *1 ④ | 11,471 ⑤ | | |
| TOTAL | 535,82 ⑥ | 15 ⑦ | | | |

*1 $ddl_{tot} - ddl_{fac}$ ?

① $$SCE_{fac} = \left(\frac{(\Sigma \text{ modalités } 1)^2}{\text{effectif}} + \frac{(\Sigma \text{ modalité } 2)^2}{\text{effectif}} + \frac{(\Sigma \text{ modalité } 3)^2}{\text{effectif}}\right) - \mathbb{C}$$

$$SCE_{fac} = \left(\frac{217,5^2}{5} + \frac{248,9^2}{5} + \frac{280,6^2}{5}\right) - 37\,200,6$$

$$SCE_{fac} = 398,164$$

② $ddl_{fac}$ = nb modalités − 1

③ $$CM = \frac{SCE}{ddl} = \frac{398,164}{2} = 199,082$$

④ ces données se retrouvent par différence "total - facteur"

⑤ $$CM = \frac{SCE}{ddl}$$

⑥ $\Sigma(\text{chaque résultat}^2) - \mathbb{C} = SCE_{tot} = 37\,736,42 - 37\,200,6 = 535,82$

⑦ nb unités stat − 1 = $ddl_{tot}$

⑧ $$F_{calc} = \frac{CM_{fac}}{CM_e} = \frac{199,082}{11,471} = 17,36$$

*2 $F_{1-\alpha}\left(ddl_{fac}, ddl_e\right)$

* Avant l'ANOVA faut faire l'analyse des résidus (cf verso)

## Analyse des résidus (oubliée avant l'ANOVA)

- Histogramme → oui, cloche + symétrique (= normal peut-être)
- Coef. de Pearson → $\beta_1 \simeq 0$ (= symétrique)
  $\beta_2 \to$ ? (= mésocurtique)
- Cartographie → useless

→ on a une loi normale

## TEST BARTLETT

$$\chi^2_{cal} = \left(\frac{2{,}3026}{C}\right) \times \left[\nu \log\right.$$

Tableau des résidus

Moyenne : 49,8

| | | |
|---|---|---|
| ~~-4,7 3~~ | ~~-1,6~~ | ~~-8,5~~ |
| ~~-5,2~~ | ~~-4,4~~ | ~~-9,3~~ |
| ~~7,7~~ | ~~6,7~~ | ~~-7,1~~ |
| ~~4,7~~ | ~~1,5~~ | ~~-2,4~~ |
| ~~6,6~~ | ~~-5,3~~ | ~~-4,5~~ |

NON → on utilise les ~~résidus~~ moyennes de modalité !

$\nu_j = 14$

$\nu = 14$

① $\hat{\sigma}^2_j = \frac{SCE}{\nu}$

$\hat{\sigma}^2 = \frac{1}{\cancel{12}} \times$

| Ech | $n_j$ | SCE | $\nu_j$ | $\hat{\sigma}^2_j$ ① | $\frac{1}{\nu_j}$ | $\nu_j \log \hat{\sigma}^2$ |
|---|---|---|---|---|---|---|
| $T_1$ | 5 | 6,82 | 4 | 1,705 | 0,25 | 0,9269 |
| $T_2$ | 5 | 97,148 | 4 | 24,287 | 0,25 | 5,5415 |
| $T_3$ | 5 | 33,688 | 4 | 8,422 | 0,25 | 3,7017 |
| sommes : | | | (12) $\nu$ | | 0,75 | 10,1701 * |

### 4) PPDS

$$ppds = t_{1-\frac{\alpha}{2}} \sqrt{\frac{CM_e}{n_j} \times 2}$$

$$ppds = 2{,}179 \sqrt{\frac{11{,}471}{15} \times 2}$$

$$ppds = 4{,}667$$

Moyennes et effectifs :

$T_1 = 43{,}5$ $\quad T_2 = 49{,}78$ $\quad T_3 = 56{,}12$

* DONC $C = 1 + \frac{1}{3 \times (3-1)} \times \left(0{,}75 - \frac{1}{12}\right) = 1{,}11$

$\nu = 12$ $\quad$ nb éch.

$$\chi^2 = \frac{2{,}3026}{C} \times (\nu \times \log \hat{\sigma}^2$$

Exercice 3

① Contexte : modèle : $x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + e_{ijr}$

2 facteurs → machine, 4 modalités
↳ température, 2 modalités
} → 8 traitements à 3 répét
24 unités statistiques

② Résidus
↳ histogramme → symétrique en cloche
↳ coef. de Pearson → $\beta \simeq 0$ → symétrique
↳ $\beta' \simeq 3$ → mésocurtique
↳ cartographie inutile
↳ Bartlett

0,025825

| modalité | $n_j$ | $m_j$ | SCE |
|---|---|---|---|
| 25°C | 12 | 3,5475 | 0,026 |
| 45°C | 12 | 3,5875 | 0,008825 |

Tableau résidus $e = x_{ijr} - \bar{x}_{ijr}$

| | A1 | A2 | A3 | A4 |
|---|---|---|---|---|
| 25°C | 3,53 | 3,47 | 3,5 | 3,57 |
| 45°C | 3,56 | 3,53 | 3,58 | 3,62 |

| | A1 | A2 | A3 | A4 |
|---|---|---|---|---|
| 25°C | 0,05 | 0 | 0 | 0,03 |
| | 0 | 0,03 | 0,02 | 0,0 |
| | 0,05 | 0,03 | 0,02 | 0,03 |
| 45°C | 0,01 | 0 | 0,01 | 0 |
| | 0 | 0,04 | 0 | 0,01 |
| | 0,02 | 0,04 | 0,01 | 0,02 |

→ tout ça ça sert à rien, on compare les $\chi^2$ qu'on nous donne.

$$\mathbb{C} = \frac{85{,}26^2}{24} = 302{,}88615$$

## ③ → ANOVA

| source | SCE | | ddl | CM | Fcalc | Fthéo =0,95 |
|---|---|---|---|---|---|---|
| Totale | 0,064 | ④ | 23 | 0,0027 | | |
| facteur T | 0,029 | ① | 1 | 0,029 | 36,25 | 4,49 |
| Facteur A | 0,012 | ② | 3 | 0,00406 | 5,075 | 3,24 |
| Facteur AT | 0,009 | ③ | 3 | 0,003 | , | 3,24 |
| résidu | 0,014 | | 16 | $\underline{0{,}0008}$ | | |

① $SCE_T = \left(\frac{1781}{12} + \frac{1853}{12}\right) - \mathbb{C} = 148{,}4 + 154{,}44 - \mathbb{C} = 0{,}029$

② $SCE_A = \left(\frac{452{,}83}{6} + \frac{448{,}59}{6} + \frac{451{,}14}{6} + \frac{464{,}83}{6}\right) - \mathbb{C} = 0{,}01218$

③ $SCE_{AT} = \left(10{,}53^2 + 10{,}41^2 + 10{,}5^2 + 10{,}71^2 + 10{,}69^2 + 10{,}77^2 + 10{,}74^2 + 10{,}85^2\right)\frac{1}{3} - \mathbb{C}$

$= 908{,}8094 \, \frac{1}{3} - \mathbb{C}$

④ $SCE_{tot} = 302{,}9502 - \mathbb{C} = 0{,}06405$

## ④ → PPDS

→ 25°C moyenne : 3,5175 } 0,07
↳ 45°C moyenne : 3,5875
↳ A1 : 3,55
↳ A2 : 3,53
↳ A3 : 3,54
↳ A4 : 3,59

$PPDS_T = 1{,}746 \times \sqrt{\frac{0{,}0008}{12} \times 2} = 0{,}02016$

$PPDS_A = 1{,}746 \times \sqrt{\frac{0{,}0008}{6} \times 2} = 0{,}028$

0,07 > 0,02 les T ne sont pas du m̃ groupe

| | | |
|---|---|---|
| A1 | A | |
| A2 | A | |
| A3 | A | |
| A4 | | B |

ça me strrrresse !

Etude de 1 facteur (plusieurs modalités, plusieurs parcelles)

modalité ← parcelles →

$A_1$
$A_2$
⋮
$A_i$
⋮
$A_a$

→ pour 1 modalité, ça varie selon les parcelles (fluctuations naturelles)

$F_{calc} = \frac{CM_a \text{ (inter)}}{CM_e \text{ (intra)}} > 1 ?$

si oui, y a quelque chose à comprendre

Ex1 15 cobayes qui sont exposé à 3 T° différentes pr tester leur prise pondérale.

Contexte d'étude :
- Nb facteur(s) → 1 (T°)
- modalité → 3
- Traitement → 3
- Unité expérience → 15
- Réponse mesurée → gain de poids

Ex2
- Nb facteurs → 2 (saison, étang)
- Modalités → saison : 2 / étang : 5
- Traitement → 2 x 5 = 10
- Unité expérim. → 10 (chacune de mucus cumulé de 5 poissons)
- Réponse mesurée → FAM, flore aérobie mésophile

Ex 3
- Nb facteurs → 2 (T° et appareil)
- → 2 T° et 4 appareils
- → 8
- → 8 x 3 = 24
- → micro ampère

Ex 4 (faire)

Ex 5 (21 min)
- Nb facteurs → 3
- 2 ( ), 4 ( ), 3 ( )
- Traitement → 2 x 4 x 3 = 24
- Unité expé → 24

# Ex 1

cas de $x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$CM$ = carré moyen $= \frac{SCE}{ddl} = \frac{\Sigma(x_i - \bar{x})^2}{ddl}$

| Sommes de variations ↓ | SCE | ddl | CM | Fcalc |
|---|---|---|---|---|
| facteur A | $SCE_A$ | $a-1$ | $CM_A$ | $\frac{CM_A}{CM_e} = F_{calc}$ 17,355 |
| résiduelle | $SCE_e$ | $ddl_e$ | $CM_e$ | |
| totale | $= SCE_{tot}$ | $= N-1$ | | |

$F_{1-\alpha}(a-1 ; ddl_e)$
3,89

On pose $H_0 = \frac{CM_A}{CM_e} = 1$
$H_1 = \frac{CM_A}{CM_e} > 1$

On garde $H_0$ → on a pas pu prouver d'effet de facteur (23 min)
On prend $H_1$ → on a prouvé un effet de facteur.

# Ex 2

| ΔΘ | SCE | ddl | CM | F calc |
|---|---|---|---|---|
| Facteur A | $SCE_A$ | $a-1$ | $CM_A$ | $F_A = \frac{CM_A}{CM_e}$ → 13,452 |
| Facteur B | $SCE_B$ | $b-1$ | $CM_B$ | $F_B = \frac{CM_A}{CM_e}$ → 8,521 |
| résiduel | $SCE_e$ | $ddl_e$ | $CM_e$ | |
| Tot | $= SCE_{TOT}$ | $= N-1$ | | |

32 min 20

$H_0 = \frac{CM_A}{CM_e} = 1$ et $H_0 = CM_B / CM_e = 1$

$H_1 =$ " $> 1$ et $H_1 =$ " $= 1$

$F_{1-\alpha}((a-1) ; ddl_e)$ $(F_{1-\alpha}(b-1 ; ddl_e))$

# Ex 3

| ΔΘ | SCE | ddl | CM | Fcalc |
|---|---|---|---|---|
| fact A (T) | $SCE_A$ | $a-1$ | $CM_A$ | $F_A = \frac{CM_A}{CM_B}$ 34,252 |
| fact B (A) ~~fact~~ | $SCE_B$ | $b-1$ | $CM_B$ | $F_B = \frac{CM_A}{CM_e}$ 5,308 |
| interact. AB | $SCE_{AB}$ | $(a-1)(b-1)$ | $CM_{AB}$ | $F_{AB} = \frac{CM_{AB}}{CM_e}$ 2,645 |
| résiduel | $SCE_e$ | $ddl_e$ | $CM_e$ | |

F théoriques
→ T → 4,49
→ A → 3,24
→ A x T → 3,24

# Ex 4

| | | |
|---|---|---|
| subst | 1,447 | 3,33 |
| solv | 7,105 | 4,76 |
| date | 2,337 | 5,14 |
| subs x solv | 2,137 | 4,26 |
| subs x date | 2,305 | 5,14 |
| solv x date | | |

# Stats

3/05/2012

## Méthodologie

① → on comprend le modèle

② → on calcule l'erreur expérimentale $CM_e$ en étudiant les résidus (normalité, indépendance, homoscédasticité)
Si l'erreur expé est "régulière", on peut faire une ~~expé~~ estimation de $CM_e$ commun $\hat{\sigma}_\varepsilon^2 = CM_e$

③ → ANOVA, on fait un test global

④ → ppds*: méthode de regroup$^{nt}$ des moyennes → Dans les modalités, quelles sont celles qui donnent des résultats proches et quelles sont celles qui se démarquent

$$ppds = t_{1-\alpha}(\nu)\sqrt{CM_e\left(\frac{1}{n_i}+\frac{1}{n'_i}\right)}$$

par exemple : on a k moyennes
→ on les range par ordre croissant
→ $CM_e$ est donné avec ANOVA

on étudie les moyennes 2 par 2
→ $n_i$ correspond à l'effectif (nombre de valeurs utilisées pour calculer) $m_i$
→ $n'_i$ correspond à l'effectif $m'_i$

$H_0 : \mu_i = \mu'_i$
$H_1 : \mu_i \neq \mu'_i$

si $|m_1 - m'_1| < ppds$ alors $m_i$ et $m'_i$ sont du m̂ groupe
si $|m_1 - m'_1| > ppds$ alors $m_i$ et $m'_i$ sont de différ. groupe

## Exercices exemples

**Ex 1** ANOVA : p-value du facteur T° = 0,000 → effet facteur T°
$CM_e = 11{,}471$ ddle = 12

| | $T_1$ | $T_2$ | $T_3$ |
|---|---|---|---|
| $m_i$ | 43,5 | 49,78 | 56,12 |
| $n_i$ | 5 | 5 | 5 |

→ les moyennes sont déjà dans l'ordre croissant

$t_{0,975}(12) = 2{,}179$

$$ppds = 2{,}179\sqrt{11{,}471\left(\frac{1}{5}+\frac{1}{5}\right)} = 4{,}668$$

$|49{,}78 - 43{,}5| = 6{,}28 > 4{,}668$ → $T_1$ et $T_2$ sont pas du m̂ groupe
$|56{,}12 - 49{,}78| = 6{,}34 > 4{,}668$ → $T_2$ et $T_3$ sont pas du m̂ groupe

⇓

$T_1$ et $T_3$
m̂ pas la peine d'essayer !

→ Gain de poids optimal à 22°C
minimal à 5°C

* il existe plein d'autres méthodes. cf l'année prochaine

**Ex2** On étudie les facteurs étang ($p$ value $= 0{,}007$)
saison ($p$ value $= 0{,}044$)

$CM_e = 0{,}241$ $\quad ddl_e = 4$ $\quad t_{0,975}(4) = 2{,}776$

$$ppds = 2{,}776\sqrt{0{,}241\left(\frac{1}{2}+\frac{1}{2}\right)} = 1{,}363$$

→ ON TRIE LES MOYENNES :
l'effectif $n_i$ est à chaque fois 2,

| $E_2$ | $E_5$ | $E_3$ | $E_4$ | $E_1$ |
|---|---|---|---|---|
| 10,268 | 11,567 | 11,821 | 13,199 | 14,222 |

1,299 ; 0,254 ; 1,378 ; 1,023

1,563

voilà comment on schématise

| | groupes | | |
|---|---|---|---|
| $E_2$ | A | | |
| $E_5$ | A | B | |
| $E_3$ | | B | |
| $E_4$ | | | C |
| $E_1$ | | | C |

↳ $E_2$ et $E_5$ du m̂ groupe
$E_5$ et $E_3$ du m̂ groupe
Mais $E_2$ et $E_3$ pas au m̂ groupe

⟹ Si on veut le − de flore possible on prend $E_2$ ou $E_5$ (mais surtt $E_2$)
si on veut le + de flore possible on prend $E_4$ ou $E_1$.

**Ex3** facteur T ⟶ $p$ value $= 0{,}000$ effet fact T ← 2 moy
facteur A ⟶ $p$ value $= 0{,}009$ effet fact A ← 4 moy
facteur A×T − $p$ value $= 0{,}086$ pas d'effet d'interaction

on peut calculer $t_{calc} = 5{,}852$

$CM_e = 0{,}0008583$ $\quad ddl_e = 16$ $\quad t_{0,975}(16) = 2{,}12$ $\quad p$

$$ppds = 2{,}12\sqrt{\frac{0{,}0008583 \times 2}{6}} = 0{,}039$$

→ ON TRIE

| | $A_2$ | $A_3$ | $A_1$ | $A_4$ |
|---|---|---|---|---|
| $n_j$ | 6 | 6 | 6 | 6 |
| $m_j$ | 3,53 | 3,54 | 3,547 | 3,593 |

0,01 ; 0,007 ; 0,046

0,017

1 groupe ; 1 autre groupe

| | | |
|---|---|---|
| $A_2$ | A | |
| $A_3$ | A | |
| $A_1$ | A | |
| $A_4$ | | B |

## Ex4

facteur
- → A  pvalue = 0,000 → effet fact A → 4 moyennes
- → D  pvalue = 0,696 → pas d'effet D
- → AxD  pvalue = 0,006 → effet d'interaction → 8 moyennes

$CM_e = 2,538$  $ddl_e = 32$  $t_{0,975}(32) \simeq t_{0,975} = 1,96$

→ on ne travaille plus sur la prise ou non, mais bu. sur traitement, avec dose si prise

$ppds = 1,96 \sqrt{2,538 \times \frac{2}{5}} = 1,975$

| | $A_1D_2$ | $A_1D_1$ | $A_2D_2$ | $A_3D_2$ | $A_2D_1$ | $A_3D_1$ | $A_4D_1$ | $A_4D_2$ |
|---|---|---|---|---|---|---|---|---|
| $n_j$ | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| $m_j$ | 8,8 | 9,2 | 10 | 10,8 | 12 | 12,2 | 15 | 18 |

0,4  0,8  0,8  1,2  0,2  2,8  3

1,2  1,4

1,6

p

| | | | | | |
|---|---|---|---|---|---|
| $A_1D_2$ | A | | | | |
| $A_1D_1$ | A | B | | | |
| $A_2D_2$ | A | B | | | |
| $A_3D_2$ | | B | C | | |
| $A_2D_1$ | | | C | | |
| $A_3D_1$ | | | C | | |
| $A_4D_1$ | | | | D | |
| $A_4D_2$ | | | | | E |

→ si on veut η max : $A_4D_2$

→ déconseillé : $A_1$ quelque soit la dose
$A_2D_2$

## Ex5

| facteur | pvalue |
|---|---|
| subst | 0,26 |
| solv | 0,022 |
| date | 0,129 |
| substx solv | 0,197 |
| substx date | 0,180 |
| solvx date | 0,766 |

$CM_e = 1,319$

$ddle = 6$

$t_{0,975}(6) = 2,447$

$ppds = 2,447 \sqrt{1,319 \times \frac{2}{6}} = 1,623$

| | solv2 | solv4 | solv1 | solv3 |
|---|---|---|---|---|
| $n_j$ | 6 | 6 | 6 | 6 |
| $m_j$ | 14,5 | 15,833 | 16,833 | 17,333 |

1,33  1  0,5

2,33

1,5

| | | |
|---|---|---|
| S2 | A | |
| S4 | A | B |
| S1 | | B |
| S3 | | B |

La fiche RNCP ☺☺
les compétences