StatisTICE

Echantillonnage

* Gestion des déchêts d'une ville

Dans la simulation → 3001 bennes, 100 grappes (quartiers). Benne affecté au quartier de façon aléatoire.

↳ ~~dans~~ bennes peuvent provenir de types d'activités différents (habitation, usines, bureaux)

↳ dans les bennes, on a des déchêts ménagers, papier et métal

N° benne	quartier (grappes)	type d'activité (strates)	déchets ménager	déchets papier	déchets métal
0 ⋮ 3000					
			$\mu =$ $\sigma_x =$		

Pour encadrer une moyenne : $P\left[m_n - t_{1-\frac{\alpha}{2}}\,\sigma(m) < \mu < m_n + t_{1-\frac{\alpha}{2}}\,\sigma(m)\right] = 1-\alpha$

⚠ Plusieurs méthodes pour échantillonner

- → en aléatoire, $\sigma(m) = \sqrt{\frac{\sigma_x^2}{n} K}$
- → systématique
- → stratifié simple
- → stratifié prop
- → grappes

} $\sigma(m)$ spécifique à chaque méthode

Dans le logiciel : Nb Rep → nombre d'échantillon

n → taille d'échantillon

"Détails" → sur l'échantillonnage

"Détails de la pop" → d'où viennent les individus prélevés

Ecart-type de la pop

"Détails de l'échantillon" → détail de chaque individu.

La meilleure methode d'échantillonnage en fait, c'est

24/09

MRE

I Module 1

Module 1 AFT

* Dispositif expé. → cf. correction

* Normalité résidus

1) Coef. de formes = de Pearson

↳ d'assymétrie (β_1) $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$ (assymétrie avérée)

↳ d'aplatiss^{nt} (β_2) $H_0 : \beta_2 = 3$
$H_1 : \beta_2 \neq 3$ (aplatis^{nt} avéré) (non mésocurtique)

2)* On a regroupé les résidus ds des classes (cf. vérif° de la normalité)

D'où les "BORNES". Le "centre" = $\frac{\text{borne inf} + \text{borne sup}}{2}$. "EFFECTIF"

"nj'eij" = "centre" x "EFFECTIF"

⚠ $n_j(e_{ij} - e_b)^2 \neq (n_j e_{ij})^2$ et eb = moy des résidus en classes = 0,264

μ_2 = moy des $n_j(e_{ij} - e_b)^2$

3) Vérif' de données out

NB : pas de méthode de Grubbs (cf 1er § p. 12 du poly expe agro zoot)

SC des résidus → 379,408

$\sigma(\varepsilon) = \sqrt{\frac{SC}{N-1}} = \sqrt{\frac{379,408}{20-1}} = 4,469$ (cf liste calculée complète ds correction)

↳ N = nb résidus

* Indépendance des résidus

→ graphe en nuage 

+ calcul de R^2 pour vérifier

formule de ≃ 4 ?

* La subtilité de cette question c'est qu'on considère les résidus dans des classes et non pas "en individuel"

* Homoscédasticité

La subtilité c'est que dans l'exemple on étudie les aliments et là on fait juste pareil mais avec les "blocs d'animaux"

→ tableau pour le test Bartllet and good luck for Bartllet's fuckin' test

* Analyse de la variance

1) Just logic → pour + de précision, + de répét° !
2) Valeur théorique (unilatéral le test F hein!)
3) a- normalité, indépendance, homoscédasticité
   b- cf. exemple.

$X_{i.}$ → somme des poids des A^x au m̂ aliment
$X_{.j}$ → somme des poids des A^x du m̂ bloc

NB saisi sur excel (Alu)

$\bar{X}$ → pareil, mais moyenne

C → terme de centrage ? Formule ?

→ fin baclée

* Comparaison des moyennes

→ bis

I Modele 2

Questions

Construire un plan expé

1) $3 \times 2 \times 4 = 24$ ~~modalités~~ unités exp.

2)
Bloc 1 | V4 A2 | V2 A1 | V3 A2 | V2 A1 | V1 A2 | V3 A1 | V1 A1 | V4 A1 |

Bloc 2

etc.

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

3) $$x_{ghijk} = \mu + \alpha_i + \beta_j + \delta_g + ❀_h + \gamma_k + \alpha_i\beta_j + \alpha_i\delta_g + \alpha_i ❀_h + \beta_j\delta_g + \beta_j ❀_h + \delta_g ❀_h + \alpha_i\beta_j\delta_g + \beta_j\delta_g ❀_h + \delta_g ❀_h \alpha_i + \beta_j ❀_h \alpha_i + \varepsilon_{ijk} + \alpha_i\beta_j\delta_g ❀_h$$

ε_{ijk} → terme résiduel

terme de niveau moyen

effet principal facteur A

effet facteur bloc

effet d'interaction 2e ordre

effet d'interaction 1er ordre

effet d'interaction 3e ordre

$$e_{ijk} = x_{ijk} + \mu - \bar{x}_{..k} - \bar{x}_{ij.}$$

Questions

1)

moy	A_1	A_2	A_3	A_4		
B_1	600	640	540	680	~~2460~~	~~1230~~ 615
B_2	660	660	520	500	~~2340~~	~~1170~~ 585
	~~1260~~ 630	~~1300~~ 650	~~1060~~ 530	~~1180~~ 590	~~4800~~ 600	~~2400~~

moy bloc 1 = 540

moy bloc 2 = 660

2) $e_{312} = x_{312} + \mu - \bar{x}_{..2} - \bar{x}_{31.}$

$e_{312} = 650 + 600 - \cancel{540} - 540$

$e_{312} = \frac{\cancel{4570}}{50}$ 660

Questions

Test sur A → test F (ddl ; ddl$_e$) $\quad F_{calc} = \frac{SCE_{A_1}/\nu_1}{SCE_{A_2}/\nu_2} \quad \nu_1 = \nu_2 = 12 - 1 = 11$

Test sur B → χ^2

$\chi^2_{théo}$ = nb variances − 1 = 4 − 1 = 3

χ^2_{calc}

n_j	v_j	$1/v_j$
8	7	1/7
8	7	1/7
8	7	1/7
24	21	

4 modal. $\sum \frac{1}{\nu_j} - \frac{1}{\nu} = \frac{3}{7} - \frac{1}{21}$

n_j	v_j	$1/v_j$
6	5	1/5
6	5	1/5
6	5	1/5
6	5	1/5
24	20	4/5

test F et χ^2 : H_0 : ~~F calc~~ toutes les variances sont égales

H_1 : ~~toutes les~~ au moins une des variances est ≠

$F_{calc} < F_{théo}$ → on garde H_0.

* Pour l'interaction A×B → χ^2

n_j	v_j	$1/v_j$
3	2	1/2
3	2	1/2
3	2	1/2
3	2	1/2
3	2	1/2
3	2	1/2
3	2	1/2
3	2	1/2
24	16	4

$$\sum \frac{1}{v_j} - \frac{1}{v} = 4 - \frac{1}{16} = \frac{63}{16}$$

$\chi^2_{théo}$ (ddl → 7)

Pour l'anova

$N = a \times b \times r$

$X = \sum x_{ijk}$

$\mu = X/N$

$C = X \times \mu$

* Pour l'effet bloc → χ^2

8	7	1/7
8	7	1/7
8	7	1/7
24	21	3/7

$$\cancel{4}\ \frac{3}{7} - \frac{1}{21} = \frac{9}{21} - \frac{1}{21} = \frac{8}{21}$$

$\chi^2_{théo}$ (ddl → 2)

Questions

1

Facteur 1	SCE	ddl_e	CM	Fcalc	$F_{théo}$
A	~~1450,4~~ 1600,4	2	800,2	937	3,98
B	102,8	3	34,27	40,13	3,587
A×B	2,3	6	0,383	~~0,448~~	
C	1,1	1	1,1	1,288	4,844
rési	9,4	11	0,854		
Tot.	1716	23			

sous Excel, fonction INVERSE.LOI.F

pas d'effet facteur interaction

2

		ddle
X	2 mod.	1
Y	3 mod.	2
Z	4 mod.	3
5 blocs		4
~~résid~~		~~113~~ ~~105~~
~~tot~~		~~124~~ ~~119~~
XY		~~5~~ 2
XZ		~~7~~ 3
YZ		~~11~~ 6
XYZ		6
résid		92
tot		119

Questions Risque 5%

$ppas = q_{1-\alpha} \times \sigma(m)\ k \rightarrow$

Facteur A

$615 - 585 = 30$

$ppas = 3,34 \times \sqrt{\frac{142857}{8}} = 44,63$ ~~ppas > différence~~

⇒ $\hat{m}$ group

Facteur B

m_3	m_4	m_1	m_2
530	590	630	650

60 | 40 | 20

$ppas = 4,26 \sqrt{\frac{142,857}{4}}$

I Module 2

PPAS

ies → 90,95 →

Nb moy comparées	2	3	4	5	6	7	8
ddl = 7	3,34	4,16	4,61	5,06	5,36	5,6	5,82
ppas	89,46	111	125	135	143	149	155

4 3 3 ~~1~~1 2 ~~1~~ 2

2 2 1 1 1 2 2

500 520 540 600 640 660 660 680

20 20 60 40 20 0

160 < 143 → pas m̂ gs

160 < 143 → pas m̂ gs

160 > 143 → m̂ gs

140 < 143 → pas m̂ gs

140 < 143 → pas m̂ gs

$$ppas = 7,68 \times \sqrt{\frac{1428,57}{2}} = \cancel{155,54}\ 205$$

$\neq$

$$ppas = 5,92 \times \sqrt{\frac{1428,57}{2}} = 158$$

I Module 3 Homoscédasticité

	SCE	ν_j	$\frac{1}{\nu_j}$	$\hat{\sigma}_j^2$	$\nu_j \log_{10} \hat{\sigma}_j^2$
L_1	2,813	3	$\frac{1}{3}$	0,938	−0,083
L_2	2,688	3	$\frac{1}{3}$	0,896	−0,143
L_3	1,313	3	$\frac{1}{3}$	0,438	−1,075
L_4	3,563	3	$\frac{1}{3}$	1,188	0,224
		ν 12	$\frac{4}{3}$	3,46	−1,057

→ La correction, plus précise, est de −1,081

$\leftarrow \sum \nu_j \log \hat{\sigma}_j^2$

$\hat{\sigma}^2$ ~~0,288~~

$\hat{\sigma}^2 = \frac{1}{\nu} \sum \nu_j \hat{\sigma}_j^2 = \frac{1}{12} \times 3 \times \sum \hat{\sigma}_j^2 = 0,865$

$\frac{1}{12} \log \hat{\sigma}^2 = -6,487$

−0,045

$$C = 1 + \frac{1}{3(\cancel{16}-1)} \left(\frac{4}{3} - \frac{1}{12} \right) = \frac{\cancel{1,028}}{1,139}$$

4 (nb de variances)

$$\chi^2_{calc} = \left(\frac{2,3026}{\cancel{1,028}} \right) \times \left[-6,487 + 1,057 \right] = -12,162$$

1,139

"−" en fait

$\frac{1}{12} \log \hat{\sigma}^2 = \frac{1}{12} \log(0,865) = -0,00525$ ... je trouve pâââs (dernier essai : $\chi^2_{calc} = 2,1747$

Attendu : 0,65

	SCE_j*	ν_j	$\frac{1}{\nu_j}$	$\hat{\sigma}_j^2$	$\nu_j \log \hat{\sigma}_j^2$
C_1	4,3125	3	1/3	1,4375	0,4728
C_2	2,5625	3	1/3	0,8542	−0,2053
C_3	2,8125	3	1/3	0,9375	−0,0841
C_4	0,6875	3	1/3	0,2292	−1,9193
	10,375	12	4/3		−1,7359

$$C = 1 + \frac{1}{3(4-1)} \left(\frac{4}{3} - \frac{1}{12} \right) = 1,139$$

$\hat{\sigma}^2 = 0,8646$

~~$\frac{1}{12} \log \hat{\sigma}^2 = -0,00526$~~

$12 \log \hat{\sigma}^2 = -0,7582$

$$\chi^2_{calc} = \frac{2,3026}{1,139} \times \left[-0,7582 - (-1,7359) \right]$$

* somme des carrés en fait

Modèle 3 ANOVA

1°) $\nu_1 = 3$ $\nu_2 = ddl_e = 6$ $F_{0,95}(3,6) = 4,76$

2°) $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$

→ Normalité → Homoscédasticité
→ Indpd ...

3°) a) Jolie cloche.

β_1 : on aurait 64% de risque de se tromper en rejetant l'hypothèse de symétrie
β_2 : on aurait 30% mésocurtie

χ^2 : pour $\alpha = 5\%$ et $\nu = 3$ $\chi_{théo} = 12,8$ $\chi_{théo} > \chi_{calc}$ donc on garde H_0.

↳ on a 49% de risque de rejeter H_0 à tort.

Cartographie → rien d'obvious

b)

Δ facteurs	SCE	ddl
Vache		3
Période		3
Traitement		3
Résidu		6
Total		16-1

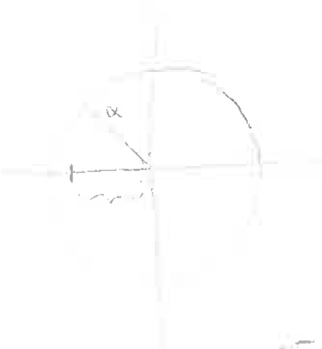
Module 7 16/11/2012

Question 1

Compléter la table

k	2	3	3	4	4	5	5	5	3
N_f	4	8	4	16	8	32	16	8	$\frac{8}{2}=4$
N_a	4	6	6	8	8	10	10	10	6
N_0	4	4	4	4	4	4	4	4	4
α	1,414	1,682	1,414	2	1,682	2,378	2	1,682	1,414
N	12	18	14	18	20	46	30	22	

1,682 = α

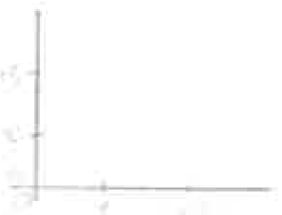


	U1	U2	U3
− α	233,64	5	2,158
− 1	220		3
0	200	10	
+ 1	180		3,842
+ α	166,36	15	5

15 − 10 1,682 → 2

1 → 0,842

CARRE LUCIE



$\frac{x-20}{30-20} = \frac{1}{1,414}$

$x = \frac{30-20}{1,414} + 20 = 27,07$

$\frac{x-150}{200-150} = \frac{1}{1,414}$

$x = \frac{200-150}{1,414} + 150 =$

$\frac{1}{\alpha} =$

1) ~~X~~ 7 colonnes / 16 lignes

~~X~~ 9 colonnes /

→ effet → colonne ID

$$\boxed{\hat{y} = \beta_0 + \sum_{j=1}^{8} \beta_j X_j}$$

Dispositif expé. associé à ce modèle : on a un plan de criblage

comme $k = 8$, $k' = 9$ alors $\mathcal{H}_{12}$ → dispositif : plan de criblage $\mathcal{H}_{12}$

(on va faire 12 essais en partant de la 1ère ligne de signes)

8 facteurs
9 coefficients

⇒ on construit la matrice avec les 11 signes et on négligera les 3 dernières colonnes à la fin

On prend les ddl. Avec une $\mathcal{H}_{12}$, on a 12 expériences → ddl = 11

9 coef. → ddl($\hat{y}$) = 8 $\qquad$ $ddl_e = 11 - 8 = 3$

Matrice expé : [12 ; 8] $\quad$ Matrice X : [12 ; 9] $\quad$ Matrice info dispers $\frac{1}{2}$ B → [9, 9]

Dans la matrice d'info, on aura une diagonale de 12 et des 0 partout ailleurs.

$$\hat{y} = \beta_0 + \sum_{1}^{4} \beta_j X_j + \beta_{14} X_1 X_4 + \beta_{23} X_2 X_3$$

$k = 4$, $k' = 7$ → on devrait faire 2^4 expé à la base → mais on peut faire 2^{4-1} ($\frac{1}{2}$ plan complet) → Générateur :

- $I = 134$? NON car $b_j = b_{12} + b_3$ (contrastes)
- $I = 1234$ NON car $f_j = b_{14} + b_{23}$
- $I = 234$ NON car $f_j = b_{23} + b_4$
- $I = 124$ NON car $b_{14} + b_2$
- $I = 14$ NON car $f_j = b_1 + b_4$
- ... bah en fait on peut pas, faut faire un plan complet.

$N_f = 16$
$N_0 = 2$
$N = 18$

Mat. exp [18, 4]

X [18, 7]

X'X [7, 7]

avec une diagonale 18 → 16 → 16 → etc et zéroooo

et zérooo

⇓

coef. calculables avec la somme algébrique affectée des signes précédents/n

	ddl
$\hat{y}$	6
e	11
ad	10
rep	1
y	17

N_0 → effe

$$\hat{y} = \beta_0 + \sum_{1}^{4} \beta_j X_j + \beta_{14} X_1 X_4 + \beta_{24} X_2 X_4 + \sum^{4} \beta_{jj} X_j^2$$

Dispositif : plan composite centré (plan en étoile) (y a des infos cachées !!)

$N_f = 8$

$N_a = 2 \times 4 = 8$

$N_0 = 4$ (par défaut)

$N = 20$

$k = 4$, $k' = 5$ $\rightarrow 2^4 \rightarrow$ demi plan possible ?

générateur $I = 1234$

↓

demi-plan possible $\rightarrow 2^{4-1} = 8$

	ddl
$\hat{y}$	$11 - 1 = 10$
e	9
ad	6
répét	$4 - 1 = 3$
$y =$	13

	U_1	U_2	U_3	U_4
Point de départ :	7,5	210	375	100

ΔU_1 : 2,5 comp → 6,5 → 1,71

ΔU_2 : 105 comp → 63 → 16,5789

ΔU_3 : 10 comp → 14 → 3,6842

ΔU_4 : 20 comp → 38 ← interdit ! → $\times \frac{10}{38}$ → 10

	U_1	U_2	U_3	U_4
Point de départ	7,5	210	375	100

ΔU_1 : 2,5 comp → 6,5

ΔU_2 : 105

ΔU_3 : 10

ΔU_4 : 20