TD n°6.

dé classique à 6 faces de 1 à 6.

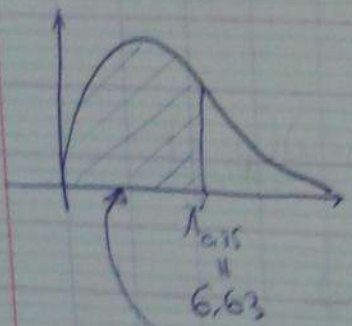
H_0 : "le dé n'est pas truqué".

modèle théorique $p = 1/6$

f_i : fréquence d'apparition de la face i

"distance" χ^2_{obs}

on considère un dé non truqué.



χ^2 théorique

$p(\chi^2 < 6,63) = 0,75$

si $\chi^2_{obs} = 8$: on rejette H_0 $\Leftrightarrow$ le dé est truqué.

p = probabilité d'erreur dans ce cas

si il y a erreur, le dé est pas truqué

$\chi^2_{obs} > \chi^2_{0,75}$ $\quad p < 25\,\%$ d'erreur.

Test d'indépendance :

X et Y, 2 variables sur une population de taille N

X $\quad n$ modalités $x_1 \dots x_n$

Y $\quad p$ modalités $y_1 \dots y_p$

$X \backslash Y$	y_1	...	y_p	
x_1				
x_i				$k_{i.}$
x_n				

k_{ij} effectif de la modalité $X = x_i \cap Y = y_j$

$k_{i.}$ effectif de la modalité x_i

$k_{.j}$ effectif de la modalité y_j

modèle théorique de l'indépendance :

$$\boxed{f'_{ij} = f_{i.} \times f_{.j}}$$

$$\frac{k'_{ij}}{N} = \frac{k_{i.}}{N} \times \frac{k_{.j}}{N} \iff \boxed{k'_{ij} = \frac{k_{i.} \times k_{.j}}{N}}$$

$$\boxed{\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{\left(k_{ij} - k'_{ij}\right)^2}{k'_{ij}}}$$

ν = nb de degré de liberté $= (n-1)(p-1)$

ex pour $\nu = 10$ (3 lignes / 6 colonnes)

① $\chi^2_{obs} = 21$ H_0 : les variables sont indépendantes ; si on rejette l'H_0 :
Probas d'erreur : $p =$ $1\% < p < 2{,}5\%$
$\Rightarrow$ on rejette H_0

② $\chi^2_{obs} = 11$
si on rejette H_0.
$\chi^2_{0,50} < \chi^2_{obs} < \chi^2_{0,75}$
$25\% < p < 50\%$
Acceptation ou pas selon le seuil d'erreur choisi initialement.

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(k_{ij} - k'_{ij})^2}{k'_{ij}}$$

$$f_{ij} = \frac{k_{ij}}{N}$$

$$f'_{ij} = f_{i.} \cdot f_{.j} = \frac{k'_{ij}}{N}$$

$$= \sum \sum \frac{(N f_{ij} - N f_{i.} f_{.j})^2}{N f_{i.} f_{.j}}$$

$$= \sum \sum N^2 \frac{(f_{ij} - f_{i.} \cdot f_{.j})^2}{N f_{i.} f_{.j}}$$

$$\boxed{\chi^2 = N \sum\sum \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}}$$

inertie totale → $R = \sum d_i - 1$

$= Tr(T^*) - 1$ Trace de la matrice qu'on diagonalise

$$R = \sum_{i=1}^{n} \frac{1}{\sqrt{f_{\cdot i} \times f_{\cdot i}}} \cdot \sum_{k=1}^{n} \frac{f_{ki} \cdot f_{ki}}{f_{k\cdot}} - 1$$

changt de notation

$$R = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{f_{ij}^2}{f_{\cdot j} \times f_{i\cdot}} - 1$$

$$\chi^2 = N \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{f_{ij}^2 - 2 f_{ij} f_{i\cdot} f_{\cdot j} + (f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}$$

$$= N \left[\sum_{i=1}^{n} \sum_{j=1}^{p} \frac{f_{ij}^2}{f_{i\cdot} f_{\cdot j}} - 2 \sum\sum f_{ij} + \sum\sum f_{i\cdot} f_{\cdot j} \right]$$

$$= N \left[R + 1 - 2 \times 1 + \underbrace{\sum_{i}^{n} f_{i\cdot}}_{1} \cdot \underbrace{\sum_{j}^{p} f_{\cdot j}}_{1} \right]$$

$$= N [R + 1 - 2 + 1]$$

$$\chi^2 = N \times R$$

ex 1

$$\chi^2_{obs} = 180 \times \left(\left[\frac{\frac{50}{180} - \frac{153}{180} \times \frac{55}{180}}{\frac{153}{180} \times \frac{55}{180}} \right] + \left[\frac{\frac{5}{180} - \frac{27}{180} \times \frac{55}{180}}{\frac{27}{180} \times \frac{55}{180}} \right] + \left[\frac{\frac{47}{180} - \frac{153}{180} \times \frac{61}{180}}{\frac{153}{180} \times \frac{61}{180}} \right] + \left[\frac{\frac{14}{180} - \frac{61}{180} \times \frac{27}{180}}{\frac{61}{180} \times \frac{27}{180}} \right] + \left[\frac{\frac{56}{180} - \frac{153}{180} \times \frac{64}{180}}{\frac{153}{180} \times \frac{64}{180}} \right] + \left[\frac{\frac{8}{180} - \frac{27}{180} \times \frac{64}{180}}{\frac{27}{180} \times \frac{64}{180}} \right] \right) = 4,84$$

H_0 : Il y a indépendance entre la correcteur et l'évaluation

$n = 2$ $\qquad V = 1 \times 2 = 2$

$p = 3$

au seuil de 5 % $\qquad \chi^2_{0,95} = 5,99$

$\chi^2_{obs} < \chi^2_{0,95}$ ⇒ on accepte l'hypothèse H_0 ⇒ il y a indépendance au risque de 5 % $\qquad p > 5\%$

↳ d'erreur (si on rejette)

au seuil de 10 % : $\chi^2_{0,90} = 4,61$

$\chi^2_{obs} > \chi^2_{0,90}$ ⇒ on rejette l'hypothèse H_0 ⇒ il y a dépendance

$p < 10\%$

↳ erreur (si rejette)

ex 3 :

$\chi^2_{obs} = 14,03$ $\qquad V = 12 = 3 \times 4$

$\chi_{0,95} > \chi^2_{obs}$ ⇒ on accepte l'hypothèse

TD 7 :

	E	B	M	m		E	B	M	m		E	B	M	m		E	B	M	m	
20 - 29	80	30	0	5	115	2,3	0,8	0	0,1	3,3	69,6	26,1	0	4,3	100	3,5	3,4	0	4	3,3
30 - 39	295	135	20	20	470	8,5	3,9	0,6	0,6	13,6	62,8	28,7	4,3	4,3	100	13,1	15,1	9,3	16	13,6
40 - 49	750	260	55	30	1095	21,5	7,4	1,6	0,8	31,4	68,5	23,7	5	2,7	100	33,3	29,1	25,6	24	31,4
50 - 59	445	165	35	25	670	12,8	4,7	1	0,7	19,2	66,4	24,6	5,2	3,7	100	19,7	18,4	16,3	20	19,2
60 - 69	315	195	50	15	575	9	5,6	1,4	0,4	16,5	54,8	33,9	8,7	2,6	100	14	21,8	23,6	12	16,5
70 - 79	270	80	35	25	410	7,7	2,3	1	0,7	11,7	65,9	19,5	8,5	6,1	100	12	8,9	16,3	20	11,7
80 - 89	100	30	20	5	155	2,9	0,9	0,6	0,1	4,4	64,5	19,4	12,9	3,2	100	4,4	3,4	9,3	4	4,4
	2255	895	215	125	3490	64,6	25,6	6,1	3,6	100	64,7	25,6	6,1	%	100	100	100	100		100
	EFFECTIF					Freq en %					PFL %					PFC %				

Analyse des PFL :

Sous représentation		Sur-représentation		
E/60-69		E/20-29		
B/70-79	B/80-89	B/60-69	B/30-39	
M/20-29	M/30-39	M/80-89	M/70-79	M/60-69
M/40-49	M/50-59			

fréquence faible → attention de la raison de ce faible effectif :
- petit effectif ⊕ petit effectif tot → gd risque
- // ⊕ gd effectif tot → risque faible.

Analyse des PFC

Sous représentation		Sur représentation	
20/29/M		20-29/m	
30-39/M		30-39/B	30-39/m
40-49/M	40-49/m	60-69/B	60-69/M
50-59/M		70-79/M	70-79/m
60-69/m		80-89/M	
70-79/B			
80-89/B			

Inertie totale.

$R = 0,0122 + 0,0096 + 0,0028 = 0,0246$

Remarque 4 colonnes → 4 valeurs propres dont une triviale qui est égale à 1

$R = \left[\sum \lambda_i\right] - 1$

$\chi^2 = N \times R$

$= 3490 \times 0,0246$

$= 85,854$

$\nu = \cancel{x} \; \cancel{\cdot} \; \cancel{\ldots} \; (n-1)(p-1) = 6 \times 3 = 18$

H_0 : ~~...~~ Il y a indépendance entre l'âge et le résultat :

Frame1
Frame2
du mot de passe()
entrer votre login et votre mot de passe:
identification entrée:(->Système de rése

2) a) $\cancel{\chi^2_{0,95} > \chi^2 > \chi^2_{0,975}}$

~~on accepte l'indépendance avec un risque inférieur à 2,5 %~~

$\chi^2 > \chi^2_{0,995}$

Donc on rejette l'hypothèse H_0 avec un risque presque nul. ($< 0,5\,\%$)

b) $R_1 = \dfrac{0,0122}{0,0246} = 49,6\,\%$

$R_2 = \dfrac{0,0096}{0,0246} = 39\,\%$

$R_3 = \dfrac{0,0028}{0,0246} = 11,4\,\%$

$R_1 + R_2$: 88,6 ; $R_1 + R_2 + R_3$: 100

Remarque : on peut choisir uniquement l'axe 1 et 2 pour faire l'analyse

$qlt = \dfrac{\text{valeur}^2}{\sum \text{valeur}^2}$ $\qquad cr = f_{i.} \times \dfrac{\text{valeur}^2}{d_i}$

$qlt = \dfrac{(0,183)^2}{(-0,183^2 + 0,173^2 + 0,06^2)} = 0,50$

	Axe 1	Axe 2	Axe 3
20 - 29	0,50		

TD 7

Un service de chirurgie orthopédique d'un grand hôpital a effectué des opérations de la main de plusieurs centaines d'individus âgés de 20 à 90 ans et a noté par "mauvais", "moyen", "bon", "excellent" le résultat de l'opération sur chacun des patients.
Le tableau des effectifs (ou de contingence) est le suivant :

Effectif	Excellent	Bon	Moyen	Mauvais	total
20 à 29	80	30	0	5	115
30 à 39	295	135	20	20	470
40 à 49	750	260	55	30	1095
50 à 59	445	165	35	25	670
60 à 69	315	195	50	15	575
70 à 79	270	80	35	25	410
80 à 89	100	30	20	5	155
total	2255	895	215	125	3430

1. Compléter le tableau des effectifs par ses marges et vérifier l'effectif total. Écrire le tableau des fréquences, calculer les fréquences marginales, le tableau des profils-lignes et profils-colonnes, en indiquant les poids de chaque ligne ou colonne. Faire une analyse des PFL et des PFC.
2. Les valeurs propres issues de l'AFC sont les suivantes : 0.0122, 0.0096, 0.0028.
   a. Calculer l'inertie totale et faire un test du χ^2. Que peut-on en déduire ?
   b. Calculer la part d'inertie associée à chaque axe. Que peut-on en déduire ?
3. Le calcul de l'AFC donne les coordonnées suivantes des classes d'âges et de la qualité du résultat de l'opération sur les axes :

	axe 1	axe 2	axe 3
20 à 29	-0,183	0,173	0,06
30 à 39	-0,005	0,086	0,063
40 à 49	-0,067	0,019	-0,055
50 à 59	-0,047	0,012	0,003
60 à 69	0,229	0,048	0,001
70 à 79	-0,044	-0,182	0,082
80 à 89	0,095	-0,272	-0,082

	axe 1	axe 2	axe 3
Excellent	-0,071	-0,011	-0,018
Bon	0,127	0,118	0,019
Moyen	0,268	-0,294	-0,032
Mauvais	-0,101	-0,141	0,254

   a. Calculer les qlt et les cr
   b. Finir l'analyse

52,84	ABS :	629 581
0,83 %	BLANC :	9938
13,72 %	UMP :	163 513
12,44	PS :	148 309
9,75	CC :	116 231
2,89	FG	34 501
2,89	FN	34432
1,88	MODEM	22436
1,39	DLR :	16 611
1,32	EXTG :	15 736
	Total	1 191 318

1	16237	11	79455
2	11572	12	63 456
3	20 119	13	62079
4	17093	14	77967
5	34955	15	130 524
6	25435	16	20039
7	30724	17	87 941
8	22686	18	93 107
9	33936	19	90 691
10	46052	20	103 160

Sur représentation   Sous représentation

	ABS	BLANC	UMP	PS	CC	FG	FN	MOD	DLR	EXTG
	52,6	0,8	13,7	12,4	9,8	2,9	2,9	1,9	1,4	1,3
1	52,6	0,9	10,2	10,5	10,1	2,1	2,5	1,7	1,5	0,7
2	55,1	0,6	10,9	11,5	13,1	2,3	2	2,2	1	1
3	52,3	0,8	10,4	14,5	13,3	3,1	1,8	1,8	0,2	1,1
4	52	0,9	13	13,3	11,3	2,8	2,4	1,9	1,5	1
5	46,9	0,9	15,7	13,9	11,8	3,3	2,6	2,4	1,5	1,1
6	49,7	0,8	20,1	11	9,4	1,5	2,5	2,2	1,5	0,6
7	51,6	0,7	26,5	7,3	6,1	0,8	3	1,9	1,8	0,4
8	56	0,8	24	6	5,7	0,6	3	1,8	1,8	0,4
9	51,6	0,8	14	12,6	11,8	2,5	2,4	2	1,3	0,8
10	52,9	0,7	8,1	14,1	13,4	4,1	2,3	1,7	1	1,7
11	51,3	0,8	9	15,3	12,8	3,9	2,1	1,9	1,2	1,5
12	50,8	0,9	12,1	14,2	10,4	3,2	3,2	2,1	1,6	1,4
13	52,2	1	9,1	15,8	10	3,8	3	2,1	1,3	1,7
14	49,1	0,9	12,6	15,4	10,6	3,2	2,9	2,3	1,6	1,4
15	50,1	0,8	13,1	12	8	1,8	3,3	2,2	1,8	0,9
16	55,8	0,7	27,1	5,1	4,2	0,6	3,2	1,4	1,7	0,3
17	54,1	0,9	13,2	7	7,7	1,6	3,2	1,1	1,5	0,9
18	55	0,7	7,9	13,3	14,5	3,7	3	1,5	1,1	2,1
19	58,4	0,8	7,3	12,9	0,6	3,8	2,7	1,5	0,9	2,1
20	54,9	0,8	6,7	13,9	11,1	4,8	2,3	1,7	1	2,1

ot de passe()

Frame2

⊞ → entrer votre login et votre mot de passe:

⊞ → identification entrée:(->Système de rése

BLANC - MOD - FN très proche du PFL moyen

PS

DLR

EE - ABS

EXTG FG UMP très éloigné du PFL moyen

$N = 1\ 191\ 318$

$R = \Sigma d_i = 0.053$

exe 1 et 2

age = 98 %

$\chi^2 = N \cdot R = 63\ 000 \gg \chi^2_{th}$ on peut donc rejeter l'hypothèse d'indépendance avec de très faible risque

$\nu = 19 \times 9 = 171$

PFL/k → uniquement sur la qualité