ch: Qualitatif * Qualitatif

- Chercher une relation numérique approché entre deux caractères qualitatifs.

Tableau de contingence:

Modalités de la variable Y

X \ Y	Y_1	...	Y_j	....	Y_l	Total
X_1	$n_{1,1}$	...	n_{1j}	...	n_{1l}	$n_{1,.}$
⋮						..
						$n_{i,.}$
X_i			n_{ij}		$n_{k,l}$	$n_{k,.}$
X_k					$n_{.,l}$	n
Total	$n_{.,1}$		$n_{.,j}$			

effectifs marginaux pour X : nb des individus ont la modalités $X_1, \dots, X_k$

effectif ou nombre des individus ont modalités X_i et Y_j.

Modalités de la variable X.

eff Marginaux pour Y : nb individu ont la modalité $Y_1 \dots Y_l$

Exp:

Yeux \ cheveux	bruns	chatain	roux	blonds	Total
bleus	11	10	1	8	30
verts	5	8	1	4	18
Marrons	16	22	2	12	52
Total	32	40	4	24	100

18 → ont couleur yeux vert.

32 personnes ont cheveux bruns

100 → Total

- EFFectifs marginaux :

pour X : $n_{i,.} = \sum_{j=1}^{l} n_{ij}$

pour Y : $n_{.,j} = \sum_{i=1}^{k} n_{i,j}$

EFFectifs total :

$$\boxed{n = \sum_{j=1}^{l} n_{.,j} = \sum_{i=1}^{k} n_{i,.} = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{i,j}}$$

totaux des colonnes

totaux des lignes

tous les élément du tableau

. Fréquences Marginales : (m̂ méthode que effectif mais avec les fréquences)

pour X ; $f_{i,.} = \sum_{j=1}^{l} f_{i,j}$

pour Y ; $f_{.,j} = \sum_{i=1}^{k} f_{i,j}$

$$\boxed{\sum_{j=1}^{l} f_{.,j} = \sum_{i=1}^{k} f_{i,.} = \sum_{i=1}^{k}\sum_{j=1}^{l} f_{i,j} = 1}$$

	Y_1	...	Y_j	...	Y_l	Total
X_1	$f_{1,1}$		$f_{1,j}$		$f_{1,l}$	$f_{1,.}$
⋮						
X_i	$f_{i,1}$					$f_{i,.}$
⋮						
X_k	$f_{k,1}$					$f_{k,.}$
total	$f_{.,1}$		$f_{.,j}$		$f_{.,l}$	1

Exemple :

totale $n=100$

Yeux \ Cheveux	bruns	chatagie	roux	blond	Total
bleu	11/100 = 0.11	0.1	0.01	8/100 = 0.08	30/100 = 0.3
vert	5/100 = 0.05	0.08	0.01	0.04	0.18
marron	0.16	22/100 = 0.22	0.02	0.12	0.52
total	0.32	0.4	4/100 = 0.04	0.24	1

Profil ligne et Colonne :

profil ligne : $f_{j/i} = \frac{n_{i,j}}{n_{i,.}}$ (chaque élément tableau est divisé par total du ligne où il existe).

Intérprétation: répartition en fréq de la variable Y dans une sous-populat° definie par une modalité de la variable X.

Exp:

	brun	chatc	roux	blonds	Total
bleu	$\frac{11}{30}$ = 0.37	0,33	0.03	0.27	1
vert	$\frac{5}{18}$ = 0.28	0.44	0.06	0,22	1
Marron	0.31	0.42	0.04	0.23	1

(tj ici les totaux ligne est 1)

vert/brun : profil ligne est :

nb individu ont yeux vert et cheveux brun ← $\frac{5}{18}$ = 0.28 ⇒ alors 28% des personnes ayant les yeux verts ont les cheveux bruns.

18 → nb tous individu qui ont yeux vert

· **Profil colonne** : Répartition en fréquence de la variable X dans une sous-populat° définie par une modalité de Y :

$$\boxed{f_{i|j} = \frac{n_{i,j}}{n_{\cdot,j}}}$$

Exp :

	BRUN	Chataigne	Roux	blonds
bleu	0.34	0.25	0.25	0.33
Vert	0.16	0.2	0.25	0.17
Marron	0.5	0.55	0.5	0.5
Total	1	1	1	[illegible]

individus ont yeux verts et cheveux chataig

$$\frac{8}{40} = 0.2$$ → 20% des chataigne ont les yeux vert.

nb tous ont cheveux chataig

Test Chi_deux:

. Pour mesure le lien de dépendance entre deux Variables qualitatifs X et Y; on calcule la statistique:

$$X^2 = \sum_{i=1}^{K} \sum_{j=1}^{\ell} \frac{(n_{i,j} - t_{i,j})^2}{t_{ij}}$$

Khi_deux: mesure distance entre tableau des eff théorique et tab eff observé

où . $t_{ij} = \frac{n_{i.} \times n_{.j}}{n}$ → effectif théorique de la case (i,j)

. n_{ij} = l'effectif observé dans le tableau pour La case (i,j)

.On compare X^2 avec le seuil donné par la Table de Khi-deux. avec. $d.d.\ell = (s-1)(K-1)$

s et K sont nb des modalités des 2 variable X et Y. → ~~(voir 58)~~

. Interprétation:

. $X^2 >$ seuil ⇒ les 2 variable sont dépendant

. $X^2 <$ seuil ⇒ Indépendantes.