FORMATIONS AUX STATISTIQUES : MANUEL DU FORMATEUR DE NIVEAU 1

Janvier 1991

SCÉNARIO 7

Tableaux de contingence

INRA

DOCUMENTS "STAGIAIRE"

Tableau des 4 catégories issues des grades

catégorie	grades
scientifiques = S	DRE - DR1 - DR2 - CR1 - CR2 - CC - ASC
ingénieurs = I	IRC - IR1 - IR2 - IE1 - IE2 - AI - 0A - 1A - 2A - 3A - 1B
techniciens = T	ATP - TR1 - TR2 - TR3 - AJT1 - AJT2 - AGT1 - AGT2 - 2B - 3B - 4B - 5B - 6B - 7B
administratifs = A	DR - DGA - SD - AAR1 - AAR2 - SAR1 - SAR2 - SAR3 - AJA1 - AJA2 - ACAS - CONT - 2D - 3D - 4D - 5D - CDAM

TABLE DE NOMBRES AU HASARD

Pour tirer des nombres au hasard, à l'aide de cette table, il faut choisir, par un moyen quelconque, un point de départ, c'est-à-dire une ligne et une colonne, et prélever les chiffres au fur et à mesure de leur rencontre, dans un sens fixé à l'avance.

Exemple :

Si l'on veut tirer au hasard 5 objets dans une liste de 100, numérotés de 0 à 99, on commence par choisir :

- un point de départ : 15e ligne, 41e colonne par exemple,
- un sens de lecture : de gauche à droite (on peut aussi lire de droite à gauche, de haut en bas ...)

On obtient 56626 77422 ...

Si on prend des nombres de deux chiffres, on va donc prélever les objets numérotés 56, 62, 67, 74 et 22.

NOMBRES AU HASARD

	5	10	15	20	25	30	35	40	45	50
	13407	62899	78937	90525	25033	56358	78902	47008	72488	57949
	50230	63237	94083	93634	71652	02656	57532	60307	91619	48916
	84980	62458	09703	78397	66179	46982	67619	39254	90763	74056
	22115	33646	17545	31321	65772	86506	09811	82848	92211	51178
5	68645	15068	56898	87021	40115	27524	42221	88293	67592	06430
	26518	39122	96561	56004	50260	68648	85596	83979	09041	62350
	36493	41665	27871	71329	69212	57932	65281	57233	07732	58439
	77402	12994	59892	85581	70823	53338	34405	67080	16568	00854
	83579	97154	40341	84741	08967	73287	94952	59008	95774	44927
10	71802	39356	02981	89107	79788	51330	37129	31898	34011	43304
	57494	72484	22676	44311	15356	05348	03582	66183	68392	86844
	73364	36416	93128	10297	11419	82937	84389	88273	96010	09843
	14499	83965	75403	18002	45068	54257	18085	92625	60911	39137
	40747	03084	07734	88940	88722	85717	73810	79866	84853	68647
15	42237	59122	92855	62097	81276	06318	81607	00565	56526	77422
	32934	60227	56707	44858	36081	79981	01291	68707	45427	82145
	05764	14284	73069	80830	17231	42936	48472	18782	51646	37564
	32706	94879	93188	66049	25988	46656	35365	13800	83745	40141
	22190	27559	95668	53261	21676	98943	43618	42110	93402	93997
20	81616	15641	94921	95970	63506	22007	29966	38144	62556	07864
	26059	65801	69870	84446	58248	21282	56938	54729	67757	68412
	71874	61692	80001	21430	02305	59741	34262	15157	27545	14522
	08774	29689	42245	41903	69179	96662	91819	60812	47631	50609
	37294	92028	56850	63380	05512	29830	37612	15593	73198	99287
25	33912	37996	78567	57201	66916	73998	54289	07147	84313	51938
	63510	61475	26980	23604	54972	72068	19403	53756	04281	98022
	01570	41701	30262	54647	06077	29354	95704	75928	21811	88274
	24159	77787	38973	82178	46802	90245	01805	23906	95559	06785
	92834	52941	88301	22127	23459	40229	74678	21859	98645	72388
30	16178	60063	59284	16279	48003	44534	08523	32752	40472	05470
	81808	32980	80550	98391	62243	19678	39551	18398	36918	43543
	28628	82072	04854	52809	86608	68017	11120	28638	72850	03650
	62249	65757	12273	91261	96983	15082	83851	77682	81728	52157
	84541	59891	01585	96711	29712	02877	70955	59693	26838	96011
35	89052	39051	99811	69831	47234	93263	47386	17462	18874	74210
	43645	89232	00384	10858	21789	14093	06268	46460	97660	23490
	61618	19275	40744	22482	12424	98601	19069	53166	41836	28205
	68136	05546	04029	47946	19526	27122	42515	55048	23912	81105
	74005	34558	93779	96120	01695	47720	88646	73520	40050	90553
40	54437	88825	07943	81795	31709	13358	04626	64838	92133	44221
	01990	94762	89926	84764	19159	95355	98213	17704	47400	30837
	02404	42408	67981	43684	55467	47030	42545	43920	11199	36521
	59253	71535	26149	35629	87127	45581	00185	01041	46662	98897
	20471	13914	99330	37938	69649	57964	97149	41628	78664	80727
45	55946	60765	74084	22484	49514	69820	41310	19722	07045	28808
	00939	47818	75949	44707	49105	06777	31998	79942	98351	10265
	49952	29123	45950	67578	13524	03023	18046	75287	74989	58152
	17328	70732	46319	26950	19037	02831	36558	82712	05590	64941
	19420	70215	90476	76400	51553	12158	14668	15656	37895	94559
50	19121	41190	49145	05373	00755	17817	22757	76116	75977	94570

STAGIAIRES

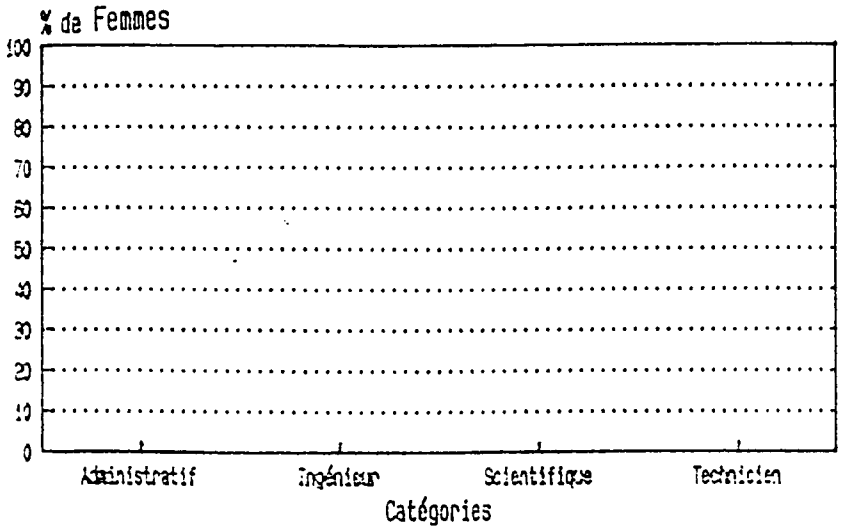
Nombre de COUPLES par stagiaire	n1	n2	n3	n k	Total = 100
F, A					
F, I					
F, S					
F, T					
H, A					
H, I					
H, S					
H, T					

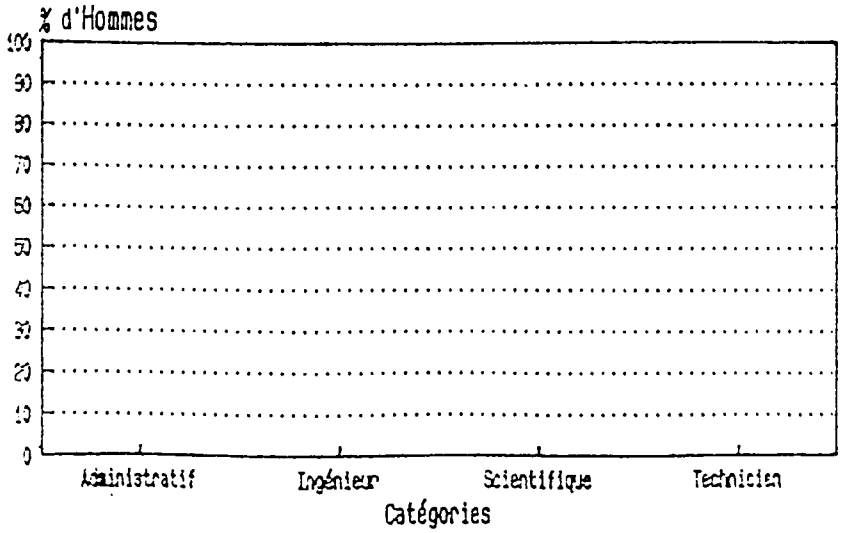
EFFECTIFS

Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					

PROFILS EN LIGNES

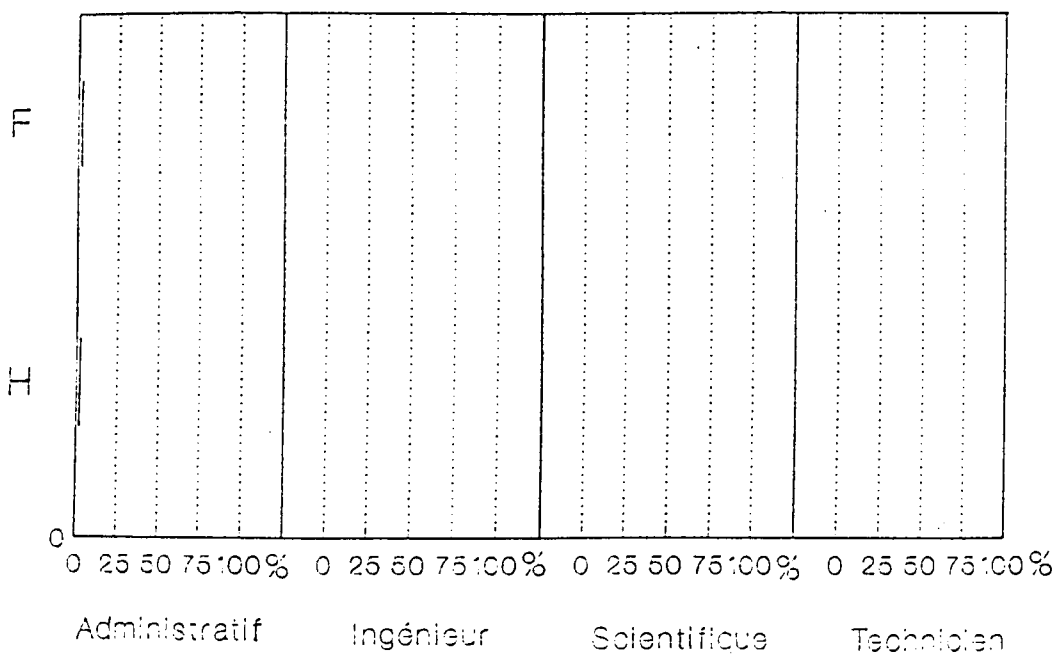
Cat. / Sexe	A	I	S	T	Total
F					100
H					100
Total					





PROFILS EN COLONNES

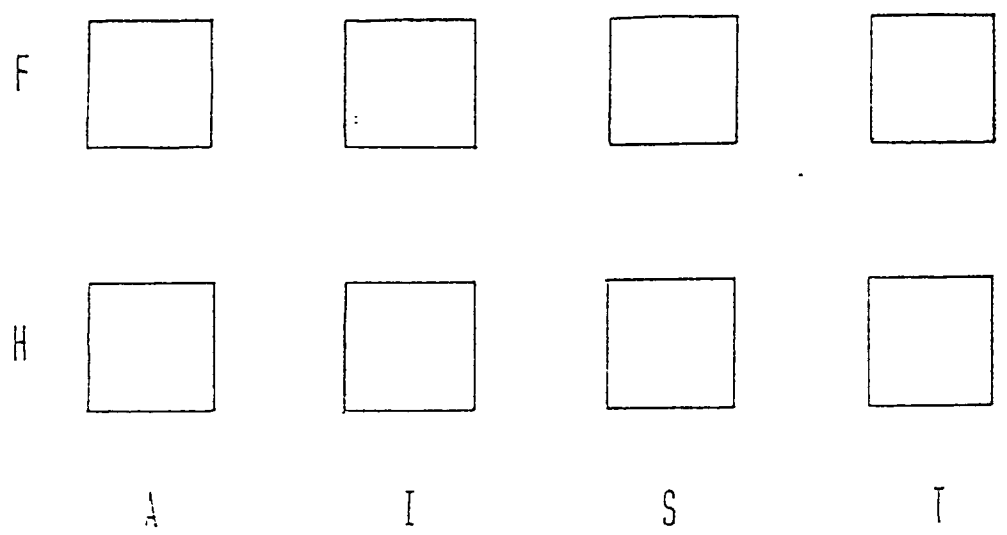
Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					



FRÉQUENCES RELATIVES

(fréquences × 100) = %

Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					100



EFFECTIFS

Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					

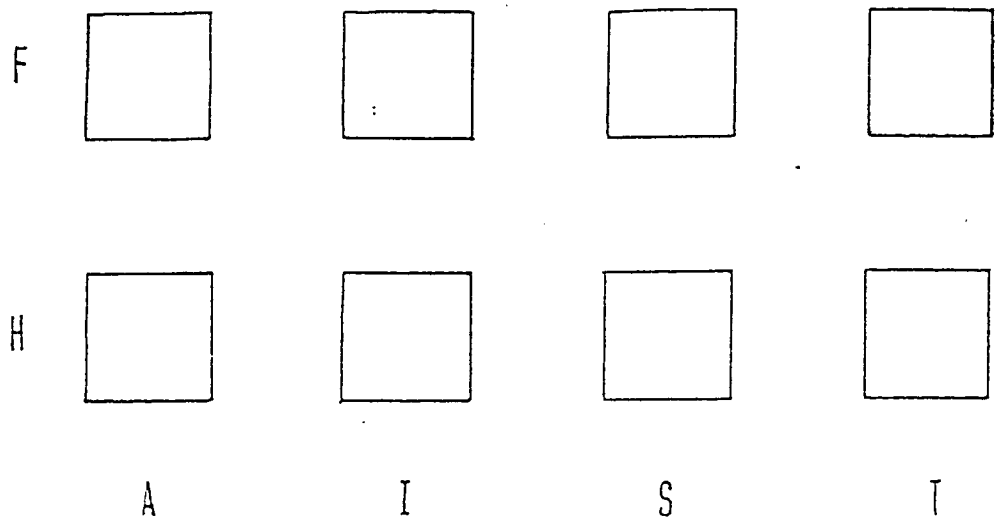
EFFECTIFS AJUSTÉS

Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					

ÉCARTS A L'INDÉPENDANCE
NOMBRE OBSERVÉ - NOMBRE AJUSTÉ

Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					

CONTRIBUTIONS AU KHI2



Cat. / Sexe	A	I	S	T	Total
F					
H					
Total					

Evaluation (ex 1 Snedecor)

Dans ce travail, on étudie le lien entre l'usage du tabac et la moralité. L'échantillonnage porte sur des hommes de 60 à 64 ans, bénéficiaires d'une pension de guerre et classés en fonction de leur usage du tabac. Deux classes sont considérées :

1. Les non-fumeurs
2. Ceux qui déclarent ne fumer que la pipe

Pour chaque pensionné qui meurt dans les six années suivantes, on fait une enquête sur les causes de sa mort. Ainsi les pensionnés peuvent être classés en morts et vivants au bout de six ans.

Non-fumeurs morts	117
Non-fumeurs vivants	950
Fumeurs morts	54
Fumeurs vivants	348

Exemple introductif sur l'Homogénéité (sur population INRA)

Tableau des effectifs

	A	I	S	T	
F	15	9	2	24	50
H	0	5	15	30	50
	15	14	17	54	100

Tableau des profils en lignes

	A	I	S	T	
F	30	18	4	48	100
H	0	10	30	60	100

Tableau des profils en colonnes

	A	I	S	T
F	100	64	12	44
H	0	36	88	56
	100	100	100	100

Tableau des fréquences relatives

	A	I	S	T	
F	15	9	2	24	50
H	0	5	15	30	50
	15	14	17	54	100

Tableau des effectifs ajustés

	A	I	S	T	
F	7.5	7	8.5	27	50
H	7.5	7	8.5	27	50
	15.0	14	17.0	54	100

Tableau des effectifs ajustés

	A	I	S	T
F	0.280	0.021	0.186	0.012
H	0.280	0.021	0.186	0.012

Valeur du χ^2 : 26.7507

Test d'évaluation

Cocher mes case correspondant au type de problème :

Indépendance Homogénéité

1) Le docteur C.H. Richardson a donne les nombres suivants d'aphidiens (*Aphis rumicis.*) morts ou vivants après un traitement avec deux concentrations de solutions d'oleate de sodium

☐ ☐

	morts	vivants	total
c.c. 0.65 %	55	13	68
c.c. 1.10 %	62	3	65
total	117	16	100

Indépendance Homogénéité

2) Pour étudier si le cancer du sein a tendance a être héréditaire, Murphy et Abbey ont :

* recherché la fréquence des cancers du sein trouvés chez les mères de femmes ayant le cancer du sein;
* effectué la comparaison avec un groupe de femmes n'ayant pas le cancer du sein.

Ils ont obtenus les données suivante :

☐ ☐

Cancer du sein chez le sujet	Cancer du sein chez la mère		total
oui	7	193	200
non	3	197	200
total	10	390	400

Indépendance Homogénéité

3) Une enquête portant sur 500 jeunes universitaires révèle que 32 % sont fumeurs, 48 % recoivent moins de 500 francs par semaine comme argent de poche, et 15 % fument et recoivent plus de 500 francs par semaine.

☐ ☐

Indépendance Homogénéité

4) Sur 100 enfants traités par un vaccin anti-grippe, 20 ont fait une poussée grippale durant les deux mois qui ont suivis l'administration du vaccin. Durant la même période, sur 1200 enfants non traités, on a observé 46 cas de grippe.

☐ ☐

Indépendance Homogénéité

5) Le tableau suivant donne le classement de 200 naissances en fonction de la parité et du poids du nouveau né :

☐ ☐

Parité / Poids	Primipares	Multipares	total
< 3 kg	26	20	46
entre 3 et 4 kg	61	63	124
> 4 kg	8	22	30
total	95	105	200

Indépendance Homogénéité

6) Une compagnie d'assurance automobile s'intéresse à la fréquence des accidents selon la date d'obtention du permis de conduire. La compagnie effectue un sondage au hasard sur 600 accidents et elle obtient le tableau suivant :

☐ ☐

Date d'obtention du permis / Fréquence d'accidents	< 1 an	> 1 an	total
au plus 1 par an	35	170	205
2 ou 3 par an	79	193	272
plus de 4 par an	57	66	123
total	171	429	600

Indépendance Homogénéité

7) Une étude comparative de la fréquence des groupes sanguins a été menée sur deux groupes d'individus. Dans le premier groupe, on a examiné 500 Allemands vivant en Allemagne, et dans le second groupe, on a examiné 476 Allemands vivant en Hongrie. Les observatrions sont présentées ci-dessous :

☐ ☐

	Groupe sanguin				total
	AB	A	O	B	
Groupe 1	25	215	200	60	500
Groupe 2	15	207	194	60	476
total	40	422	394	120	976

Indépendance Homogénéité

8) Des médecins ont étudié 141 personnes victimes de tumeur au cerveau (bénigne, maligne, ou autre), ayant des localisations variables (frontale, temporale ou autre). Ils ont obtenus les résultats suivants :

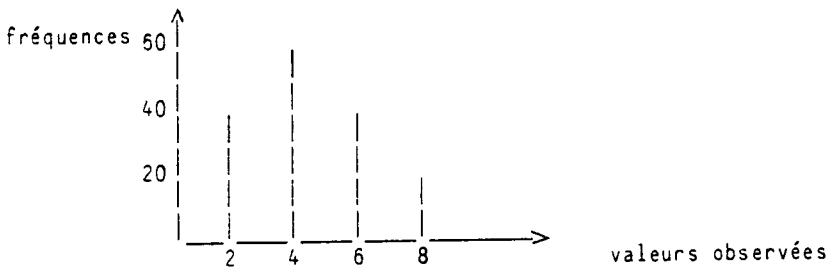
Type de tumeur / Localisation	bégnine	maligne	autre	total
frontale	23	9	6	38
temporale	21	4	3	28
autres zones	34	24	17	75
total	78	37	26	141

Scénario 7

VOCABULAIRE

Ajustement ou calcul de valeur ajustée (fitting, smothing) : A chaque valeur observée correspond une valeur ajustée. Cette valeur ajustée est déterminée par calcul.
Dans le cas d'un tableau de contingence, chaque valeur ajustée d'une case se calcule avec les moyennes de la ligne et de la colonne correspondantes.

Diagramme en bâtons (bar diagram) : cette représentation s'utilise pour une **variable qualitative**. Un diagramme en bâtons est constitué de plusieurs traits verticaux. Chaque trait est parallèle à un axe vertical et coupe l'axe horizontal en une valeur qui représente une valeur observée. La longueur de ce trait est proportionnelle à la fréquence (ou effectif) de la valeur observée.



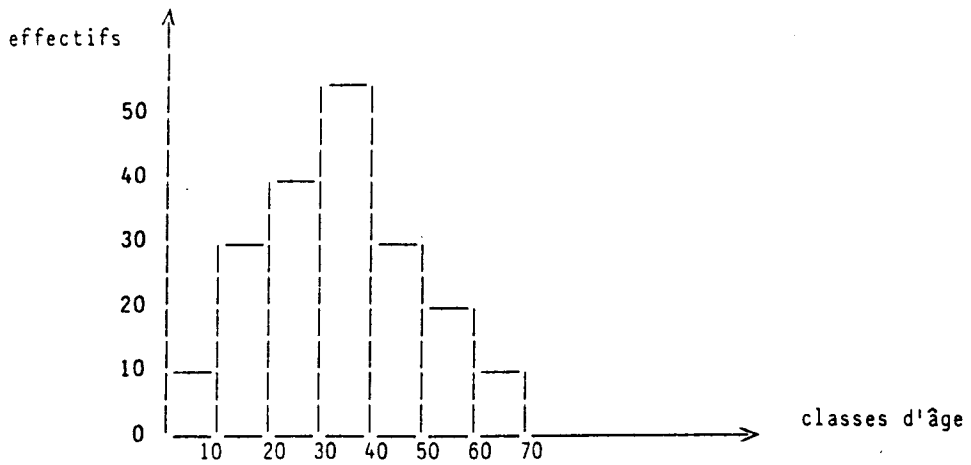
Exemple de lecture : la valeur 6 a été observée 40 fois

Distribution (distribution) : la distribution d'une variable qualitative est la liste des valeurs possibles et de leurs fréquences. On peut la représenter par un diagramme en bâtons.

Fréquence (frequency) : c'est l'effectif (ou nombre d'individus) d'une **modalité** ou d'une **classe**.

Fréquence relative ou proportion (relative frequency) : c'est l'effectif d'une **modalité** ou d'une **classe** divisé par la somme des effectifs de toutes les modalités ou classes.

Histogramme ou histogramme des fréquences (histogram) : l'histogramme se compose de rectangles contigus. Sur l'axe horizontal figurent les **modalités** ou les **classes** de la variable observée (classes d'âge par exemple) et sur l'axe vertical sont représentés les effectifs. La base des rectangles est choisie de même amplitude que la classe et leur hauteur est proportionnelle à l'effectif de la classe correspondante.



Exemple de lecture : la classe des âges 20-30 ans contient 40 individus

homogénéité de deux distributions : c'est l'égalité de deux distributions (ramenées au même effectif). Prenons l'exemple de la qualification du personnel INRA. Si la répartition des grades chez les hommes est la même que celle des femmes c'est que la distribution des grades entre les hommes et celle entre les femmes sont les mêmes.

Indépendance de deux variables qualitatives : mesurer l'indépendance entre deux **variables qualitatives**, c'est montrer qu'il n'y a aucune relation entre les valeurs de ces deuxvariables. Prenons l'exemple de la mortalité et de l'usage du tabac. Il y a indépendance de ces deux variables si la distribution de la moratlié est indépendante de la consommation de tabac, c'est-à-dire si elle est identique pour chaque modalité de consommation de tabac.

Marge colonne : somme des effectifs d'une colonne divisée par l'effectif total (multipliée par 100 si on l'exprime en %).

Marge ligne : somme des effectifs d'une ligne divisée par l'effectif total (multipliée par 100 si on l'exprime en %).

Modalité ou classe (class) : c'est une catégorie ou un état d'une **variable qualitative.**

Profil en colonne : c'est l'histogramme d'une colonne du **tableau de contingence**, ou la distribution des pourcentages de cette colonne.

Profil en ligne : c'est l'**histogramme** d'une ligne du **tableau de contingence**, ou la distribution des pourcentages de cette ligne. **Tableau de contingence ou tableau croisé** (contingency table) : c'est un tableau dont les lignes représentent les **modalités** d'une **variable qualitative**, et dont les colonnes représentent les modalités d'une deuxième variable qualitative. Chaque case de ce tableau contient le nombre d'individus qui ont à la fois :

- la modalité de la ligne correspondante (pour la 1ère variable).
- et la modalité de la colonne correspondante (pour la 2ème variable).

Variable qualitative (qualitative variable) : variable dont les éléments ne peuvent prendre leur valeur que parmi un nombre fini d'états ou **modalités**; états qui peuvent être ordonnés ou pas.
On appelle aussi caractéristique une telle variable.

Varaible quantitative (quantitative variable) : variable dont les éléments peuvent prendre n'importe quelle valeur sur un axe continu.