

## Test du chi-2 d'indépendance

### Exemple

Un enquête a été réalisée auprès d'un échantillon de 250 personnes au sujet de l'abaissement à 16 ans du droit de vote. Les réponses ont été classées suivant le niveau d'instruction des individus:

X/Y	pour	contre
Brevet	10	15
BAC	20	85
Bac +2 et plus	20	100

Peut-on accepter, au risque d'erreur  $\alpha = 1\%$ , l'hypothèse d'une liaison entre l'opinion d'une personne et son niveau d'instruction ?

**Tableau des effectifs théoriques attendus sous l'hypothèse d'indépendance :**

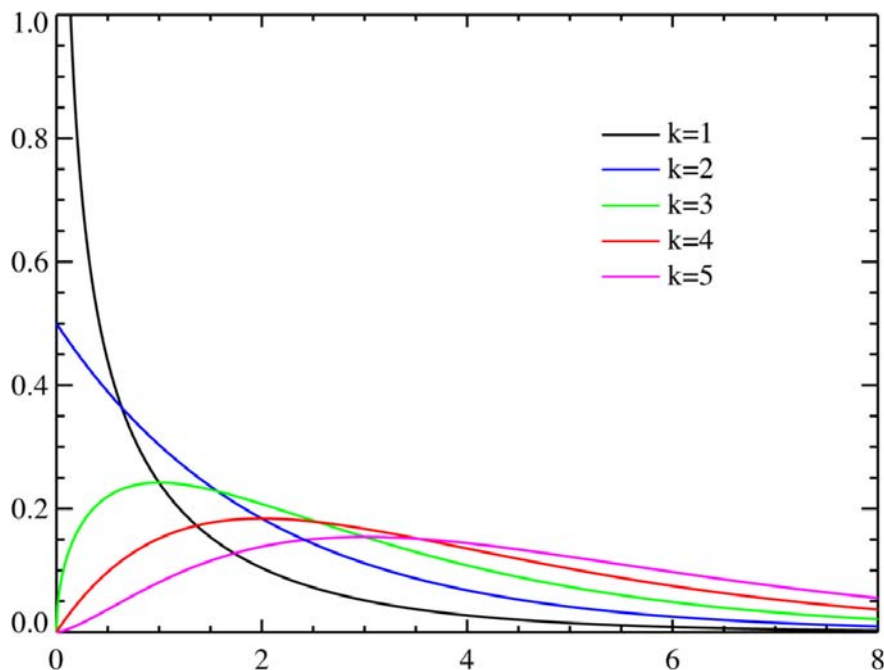
X/Y	pour	contre
BEP	$\frac{25 \times 50}{250} = 5$	$\frac{25 \times 200}{250} = 20$
BAC	$\frac{105 \times 50}{250} = 21$	$\frac{105 \times 200}{250} = 84$
Universitaire	$\frac{120 \times 50}{250} = 24$	$\frac{120 \times 200}{250} = 96$

**Distance du khi-deux**

$$Q_n^2 = \sum_{i=1}^l \sum_{j=1}^m \frac{(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{n})^2}{\frac{N_{i\bullet} N_{\bullet j}}{n}}$$

Sous l'hypothèse nulle  $H_0$ , la v.a.  $Q_n^2$  suit approximativement une loi  $\chi^2((l-1) \times (m-1))$  dès que  $n \geq 30$  et que les effectifs théoriques sont supérieurs ou égaux à 5.

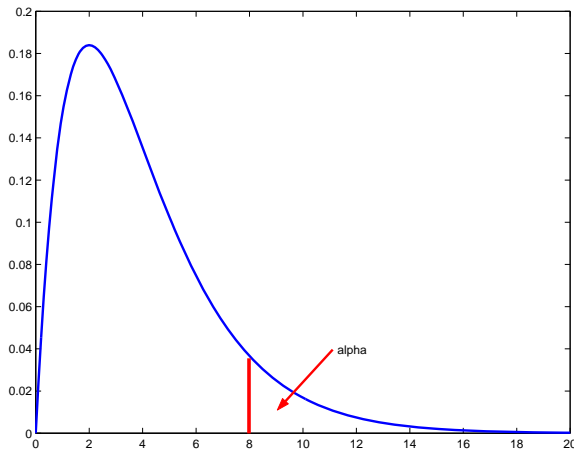
**Exemples de loi du khi-2 à  $k$  degrés de liberté:**



## Intervalle d'Acceptation ou Région Critique

$$IA = \{0 \leq Q_n \leq q_\alpha\}$$

$$RC = \{Q_n \geq q_\alpha\}$$



## Exercices

**Exercice 1** On considère un échantillon de 280 salariés d'une grande entreprise, classés selon deux critères: le niveau hiérarchique  $X$  et l'origine sociale  $Y$ . Les effectifs observés sont les suivants:

$X/Y$	ouvrier, employé	cadre
ouvrier, employé	145	64
cadre	14	57

Peut-on, au risque d'erreur 5%, accepter l'hypothèse d'une liaison entre l'origine sociale et le niveau hiérarchique dans l'entreprise ?

**Exercice 2** On veut déterminer s'il existe une relation entre l'allégeance politique des Français et leur attitude face à un récent projet de loi. A partir d'un échantillon de 800 personnes, on a recueilli les résultats suivants:

$X/Y$	Favorable	Contre	Indécis
Parti A	100	150	150
Parti B	150	100	150

Au niveau 5%, ces résultats confirment-ils l'existence de cette relation ?

**Exercice 3** La distribution suivante a été dressée par Haberman (1978) à partir de données fournies par le National Opinion Research Center de l'Université de Chicago. Les variables sont le nombre d'années de scolarité  $X$  et l'attitude face à l'avortement  $Y$ .

$X/Y$	Pour	Indifférent	Contre
Moins de 8 ans	31	23	56
Entre 9 et 12 ans	171	89	177
Plus de 12 ans	116	39	74

Peut-on considérer, au niveau 5%, que l'attitude face à l'avortement dépend du niveau d'études ?

**Exercice 4** On reporte dans le tableau ci-dessous la gravité de l'accident  $X$  et l'instant de l'accident  $Y$  pour des personnes accidentées dans un département français en 1998.

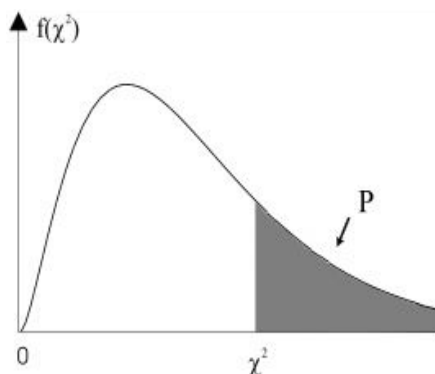
$X/Y$	Tués	Blessés graves	Blessés légers
Jour	79	232	482
Nuit	51	137	253

Peut-on affirmer, au seuil de 2%, qu'il existe un lien entre la gravité et l'instant de l'accident ?

**Exercice 5** Reprendre les données du cours et celles de l'exo 1 - Lecture des résultats sur logiciel SAS et R

### 11.3 Loi du Khi-deux $\chi^2(\nu)$

Valeurs de  $\chi^2$  ayant la probabilité **P** d'être dépassées



$\nu$	<b>P = 0,995</b>	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,00004	0,0002	0,001	0,0039	0,0158	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,965
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,39	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672

Note : L'entier  $\nu$  est le **nombre de degrés de liberté** de  $\chi^2$ .

## Distributions jointe, marginales, conditionnelles Exemple de sortie SAS

On considère un échantillon de 280 salariés d'une grande entreprise, classés selon deux critères: le niveau hiérarchique X et l'origine sociale Y. Les effectifs observés sont les suivants:

X: origine \ Y:niveau	ouvrier, employé	cadre
ouvrier, employé	145	64
cadre	14	57

Peut-on, au risque d'erreur 5%, accepter l'hypothèse d'une liaison entre l'origine sociale et le niveau hiérarchique dans l'entreprise ?

The FREQ Procedure

Fréquence Pourcentage Pctage en ligne Pctage en col.	Table de niveau par origin			
	niveau	origin		Total
		ouv-emp	cadre	
<b>ouv-emp</b>		145	64	209
		51.79	22.86	74.64
		69.38	30.62	
		91.19	52.89	
<b>cadre</b>		14	57	71
		5.00	20.36	25.36
		19.72	80.28	
		8.81	47.11	
<b>Total</b>		159	121	280
		56.79	43.21	100.00

## Test du chi2 avec SAS

Fréquence Attendu Khi-2 par cellule Pourcentage Pctage en ligne Pctage en col.	Table de niveau par origin			
	niveau	origin		
	ouv-emp	cadre	Total	
	<b>ouv-emp</b>	145	64	209
		118.68	90.318	
		5.836	7.6688	
		51.79	22.86	74.64
		69.38	30.62	
		91.19	52.89	
	<b>cadre</b>	14	57	71
		40.318	30.682	
		17.179	22.574	
		5.00	20.36	25.36
		19.72	80.28	
		8.81	47.11	
	<b>Total</b>	159	121	280
		56.79	43.21	100.00

### Statistics for Table of niveau by origin

Statistique	DDL	Valeur	Prob
<b>Chi-Square</b>	1	53.2584	<.0001
<b>Likelihood Ratio Chi-Square</b>	1	54.9843	<.0001
<i>Autres statistiques .....</i>			

## Test du chi 2 avec R (Rcmdr)

Une enquête a été réalisée auprès d'un échantillon de 250 personnes au sujet de l'abaissement à 16 ans du droit de vote. Les réponses ont été classées suivant le niveau d'instruction des individus:

X / Y	pour	contre
BEP	10	15
BAC	21	84
Universitaire	20	100

Peut-on au niveau  $\alpha=1\%$ , accepter l'hypothèse d'une liaison entre l'opinion d'une personne et son niveau d'instruction?

```
>.Table # Counts
```

```
      Pour CContre
BEP  10   15
BAC  21   84
Univ 20  100
```

```
> .Test <- chisq.test(.Table, correct=FALSE)
```

```
      Pearson's Chi-squared test
```

```
data: .Table
```

```
X-squared = 6.9547, df = 2, p-value = 0.03089
```

## Exemple Chi 2 avec regroupement

Une enquête a été réalisée auprès d'étudiants de licence sur les intentions de vote pour les élections présidentielles. On distingue les variables Intention et Année d'étude.

Les premiers résultats sont les suivants :

Fréquence Attendu Pourcentage Pctage en ligne Pctage en col.	Table de annee par intention					
	annee	intention				Total
		UMP	PS	Autre	Indecis	
<b>1A</b>	9	38	14	29	3	93
	7.479	48.776	11.706	22.437	2.6014	
	3.15	13.29	4.90	10.14	1.05	32.52
	9.68	40.86	15.05	31.18	3.23	
	39.13	25.33	38.89	42.03	37.50	
<b>2A</b>	8	36	22	26	5	97
	7.8007	50.874	12.21	23.402	2.7133	
	2.80	12.59	7.69	9.09	1.75	33.92
	8.25	37.11	22.68	26.80	5.15	
	34.78	24.00	61.11	37.68	62.50	
<b>3A</b>	6	76	0	14	0	96
	7.7203	50.35	12.084	23.161	2.6853	
	2.10	26.57	0.00	4.90	0.00	33.57
	6.25	79.17	0.00	14.58	0.00	
	26.09	50.67	0.00	20.29	0.00	
<b>Total</b>	23	150	36	69	8	286
	8.04	52.45	12.59	24.13	2.80	100.00

Statistics for Table of annee by intention

Statistique	DDL	Valeur	Prob
Chi-Square	8	51.3833	<.0001

Suite à l'examen des résultats, on décide de regrouper les données. Les résultats sont alors présentés ci-dessous :

Fréquence Attendu Pourcentage Pctage en ligne Pctage en col.	Table de annee par intention				
	annee	intention			
	UMP	PS	Autre	Indecis-	
<b>1A</b>	9	38	14	32	93
	7.479	48.776	11.706	25.038	
	3.15	13.29	4.90	11.19	32.52
	9.68	40.86	15.05	34.41	
	39.13	25.33	38.89	41.56	
<b>2A</b>	8	36	22	31	97
	7.8007	50.874	12.21	26.115	
	2.80	12.59	7.69	10.84	33.92
	8.25	37.11	22.68	31.96	
	34.78	24.00	61.11	40.26	
<b>3A</b>	6	76	0	14	96
	7.7203	50.35	12.084	25.846	
	2.10	26.57	0.00	4.90	33.57
	6.25	79.17	0.00	14.58	
	26.09	50.67	0.00	18.18	
<b>Total</b>	23	150	36	77	286
	8.04	52.45	12.59	26.92	100.00

Statistics for Table of annee by intention

Statistique	DDL	Valeur	Prob
Chi-Square	6	49.1568	<.0001