

Statistique bivariée

Observation conjointe de deux
variables X et Y

Variable conjointe

- Exemple 1 : On mesure sur des élèves le niveau scolaire X et l'absentéisme en classe Y
- Ces 2 variables X et Y sont observées **simultanément** sur n individus d'un échantillon, pris dans une population
 - La variable X à k modalités : $\{m_1, m_2, \dots, m_k\}$ ici $\{bas, élevé\}$
 - La variable Y à p modalités : $\{m_{1'}, m_{2'}, \dots, m_{p'}\}$ ici $\{rare, moyen, fréquent\}$

- À chaque individu e_i est associée

- une modalité de X notée x_i
- une modalité de Y notée y_i
- Pour l'individu 1, niveau Elevé
Et absentéisme rare

Individus	1	2	...	i	...	n
X	x_1	x_2		x_i		x_n
Y	y_1	y_2		y_i		y_n

- On parle alors de la **variable conjointe** de X et Y
- La **variable conjointe** a $k \times p$ modalités conjointes $\{m_1, \dots, m_k\} \times \{m_{1'}, \dots, m_{p'}\}$

Elève	1	2	...	i	...	n
Niveau	Bas	Elevé		Elevé		Bas
Abscent.	Fréquent	Rare		Moyen		Fréquent

Distribution en effectif

- Une **campagne de mesures présentée par la distribution en effectifs** de la variable conjointe sous la forme d'un tableau de contingence en effectifs
- les k lignes représentent les modalités de X
- les p colonnes représentent les modalités de Y
- les $k \times p$ cellules sont identifiées par leur numéro de ligne et de colonne
- la cellule c_{ij} contient le nombre n_{ij} d'individus ayant à la fois m_i et m_j comme mesures pour X et Y . ie : la modalité conjointe $(m_i ; m_j)$ pour la variable conjointe

X/Y	m'_1	m'_2	...	m'_j	...	m'_p
m_1	n_{11}	n_{12}		n_{1j}		n_{1p}
m_2	n_{21}	n_{22}		n_{2j}		n_{2p}
...						
m_i	n_{i1}	n_{i2}		n_{ij}		n_{ip}
...						
m_k	n_{k1}	n_{k2}		n_{kj}		n_{kp}

X/Y	Rare	Moy en	Fréquent
Bas	7	4	4
Elevé	8	2	2

Test en 3 étapes

Etape 1 : calculer les marges

- Les **marges** qui correspondent aux **distributions des variables** X (15, 6, 6) et Y (15, 12) (somme des effectifs en ligne et colonne).
- Si les deux variables étaient indépendantes, la distribution des valeurs du tableau serait répartie de manière « équilibrée » en ligne et en colonne.

X/Y	Rare	Moy en	Fréquent	Total X
Bas	7	4	4	15
Elevé	8	2	2	12
Total Y	15	6	6	27

En violet : les **marges**

Etape 3 : calculer les fréquences théoriques

- La valeur théorique de chaque case s'obtient en multipliant chaque marge en ligne par la marge en colonne puis en le divisant par la taille de l'échantillon.
- Ainsi, la **1ère case** devrait contenir la valeur $(15 \times 15) / 27$, soit 8,3. Le tableau théorique est donc le suivant :

X/Y	Rare	Moyen	Fréquent	Total X
Bas	8,3	3,3	3,3	15
Elevé	6,6	2,6	2,6	12
Total Y	15	6	6	27

Etape 3 : Calcul de la valeur du Khi2

- Pour évaluer l'écart entre ce tableau et le tableau précédent, on calcule, pour chaque case :

$$\frac{(\text{Valeur théorique} - \text{Valeur réelle})^2}{\text{Valeur théorique}}$$

- Ainsi, la **1ère case** devrait contenir la valeur $(8,3-7)^2/8,3$, soit 0,2

X/Y	Rare	Moy en	Fréquent	Total X
Bas	7	4	4	15
Elevé	8	2	2	12
Total Y	15	6	6	27

X/Y	Rare	Moy en	Fréquent
Bas	8,3	3,3	3,3
Elevé	6,6	2,6	2,6

X/Y	Rare	Moy en	Fréquent	Total
Bas	0,2	0,1	0,1	0,5
Elevé	0,3	0,2	0,2	0,6
Total	0,5	0,3	0,3	1,1

Khi-2 calculé⁸

Interprétation

2 hypothèses

- L'**hypothèse H0** indique qu'il n'y a **pas de lien** entre les variables sexe et stress sur le lieu de travail.
- L'**hypothèse H1** indique qu'il n'y a **un lien** entre les variables sexe et stress sur le lieu de travail.

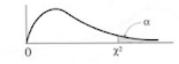
Table du Khi2

- Pour interpréter le Khi2 calculé, on se réfère à la **table du Khi2**.
- Cette table présente les valeurs (cases de la table) ayant une **probabilité** donnée d'être dépassées (en colonne), selon différents **degrés de liberté** (en ligne)
- La **probabilité** est la marge d'erreur que nous nous fixons (en général 5% (0,05) ou 1% (0,01)).
- Le **nombre de degré de liberté** correspond à :

$$ddl = (\text{Nb de lignes} - 1) \times (\text{Nb de colonnes} - 1)$$

Ici : $ddl = (2-1) \times (3-1) = 2$

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2

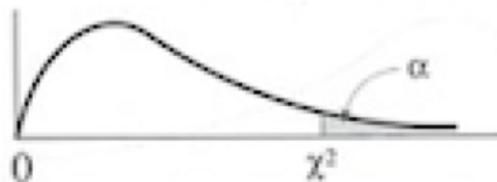


dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

Lecture du seuil critique du Khi-2

- En regardant la case qui correspond à la colonne 0,05 et à la ligne 2, on trouve la valeur **5,99**.

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2



<i>dl</i>	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28

Interprétation

- Avec le risque d'erreur de 5%, comme le Khi-2 calculé (1,1) < seuil critique du Khi-2 (5,99), alors on ne rejette pas l'hypothèse H0.
- Il n'y a **pas de lien** entre l'absentéisme et les niveau.
- On considère donc que **les 2 variables sont indépendantes.**