Chapitre 2. Test d'indépendance du χ^2

R2.06 – Techniques quantitatives et représentations 2

D. Fourer

Département TC-Juvisy

Semestre 2 – 2025/2026

Contexte, notations et objectifs

VARIABLES STATISTIQUES

- Variables X, Y qualitatives sur une même population (taille N).
- Notations (valeurs, effectifs, fréquences, etc.) : voir chapitre 1.

Contexte, notations et objectifs

VARIABLES STATISTIQUES

- Variables X, Y qualitatives sur une même population (taille N).
- Notations (valeurs, effectifs, fréquences, etc.) : voir chapitre 1.

IDÉES DIRECTRICES

- Quantifier l'« éloignement » à l'indépendance statistique.
- Évaluer l'indépendance des variables au-delà de la population.

Contexte, notations et objectifs

VARIABLES STATISTIQUES

- Variables X, Y qualitatives sur une même population (taille N).
- Notations (valeurs, effectifs, fréquences, etc.) : voir chapitre 1.

IDÉES DIRECTRICES

- Quantifier l'« éloignement » à l'indépendance statistique.
- Évaluer l'indépendance des variables au-delà de la population.

POINTS ESSENTIELS

- Construction basée sur les distributions marginales.
- Premier exemple de test statistique.

Formule

$$\chi^2(X, Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

Formule

$$\chi^2(X,Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n_{ij}})^2}{\widetilde{n_{ij}}}$$

- $\widetilde{n_{ij}} = \dfrac{n_{i*} \times n_{*j}}{N}$: effectif théorique, valeur de l'effectif conjoint si les variables étaient statistiquement indépendantes.
  - *n_{ij} est dit effectif réel par opposition.*
  - *Indépendance pour les mêmes distributions marginales.*

Formule

$$\chi^2(X, Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

- $\widetilde{n}_{ij} = \dfrac{n_{i*} \times n_{*j}}{N}$: effectif théorique, valeur de l'effectif conjoint si les variables étaient statistiquement indépendantes.
  - *n_{ij} est dit effectif réel par opposition.*
  - *Indépendance pour les mêmes distributions marginales.*
- <3>Contribution au χ^2 du couple (x_i, y_j).

Formule

$$\chi^2(X,Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

- $\widetilde{n}_{ij} = \dfrac{n_{i*} \times n_{*j}}{N}$: effectif théorique, valeur de l'effectif conjoint si les variables étaient statistiquement indépendantes.
  - *n_{ij} est dit effectif réel par opposition.*
  - *Indépendance pour les mêmes distributions marginales.*
- <3>Contribution au χ^2 du couple (x_i, y_j).
- $\chi^2(X,Y) \geqslant 0$, d'où le qualificatif de « distance ».

Formule

$$\chi^2(X,Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

- $\widetilde{n}_{ij} = \dfrac{n_{i*} \times n_{*j}}{N}$: effectif théorique, valeur de l'effectif conjoint si les variables étaient statistiquement indépendantes.
  - *n_{ij} est dit effectif réel par opposition.*
  - *Indépendance pour les mêmes distributions marginales.*
- <3>Contribution au χ^2 du couple (x_i, y_j).
- $\chi^2(X,Y) \geqslant 0$, d'où le qualificatif de « distance ».
- La distance est la somme des contributions.

Formule

$$\chi^2(X, Y) = \sum_{ij} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

- $\widetilde{n}_{ij} = \dfrac{n_{i*} \times n_{*j}}{N}$: effectif théorique, valeur de l'effectif conjoint si les variables étaient statistiquement indépendantes.
  - *n_{ij} est dit effectif réel par opposition.*
  - *Indépendance pour les mêmes distributions marginales.*
- <3>Contribution au χ^2 du couple (x_i, y_j).
- $\chi^2(X, Y) \geqslant 0$, d'où le qualificatif de « distance ».
- La distance est la somme des contributions.
- $\chi^2(X, Y) = 0 \iff X$ et Y statistiquement indépendantes.

Exemple (TD 1 – Exercice 1)

TABLE DES EFFECTIFS RÉELS

	0	**1**	**2**	**3**	***X***
A	0	0	2	3	5
B	1	3	4	2	10
C	4	6	5	0	15
Y	5	9	11	5	30

TABLE DES EFFECTIFS THÉORIQUES (À 10^{-4} PRÈS)

	0	**1**	**2**	**3**	***X***
A	$0,8333$	$1,5$	$1,8333$	$0,8333$	5
B	$1,6667$	3	$3,6667$	$1,6667$	10
C	$2,5$	$4,5$	$5,5$	$2,5$	15
Y	5	9	11	5	30

Exemple (TD 1 – Exercice 1)

TABLE DES CONTRIBUTIONS AU χ^2 (À 10^{-4} PRÈS)

	0	**1**	**2**	**3**
A	$0,8333$	$1,5$	$0,0152$	$5,6333$
B	$0,2667$	0	$0,0303$	$0,0667$
C	$0,9$	$0,5$	$0,0455$	$2,5$

Exemple (TD 1 – Exercice 1)

Table des contributions au χ^2 (à 10^{-4} près)

	0	**1**	**2**	**3**
A	$0,8333$	$1,5$	$0,0152$	$5,6333$
B	$0,2667$	0	$0,0303$	$0,0667$
C	$0,9$	$0,5$	$0,0455$	$2,5$

Distance du χ^2

$$\chi^2(X,Y) = 0,8333 + 0,2667 + 0,9 + 1,5 + 0 + 0,5 + 0,0152 + 0,0303 + 0,0455 + 5,6333 + 0,0667 + 2,5 = 12,291 \text{ à } 10^{-3} \text{ près}.$$

Étapes du test

HYPOTHÈSE NULLE H_0

H_0 = « X et Y sont indépendantes ».

- *On cherche à infirmer cette hypothèse.*

Étapes du test

HYPOTHÈSE NULLE H_0

H_0 = « X et Y sont indépendantes ».

- *On cherche à infirmer cette hypothèse.*

CHOIX DU RISQUE D'ERREUR STATISTIQUE α

- Choix « arbitraire » selon le degré de certitude souhaité.
- Souvent fixé à 5 %.

Étapes du test

HYPOTHÈSE NULLE H_0

H_0 = « X et Y sont indépendantes ».

- *On cherche à infirmer cette hypothèse.*

CHOIX DU RISQUE D'ERREUR STATISTIQUE α

- Choix « arbitraire » selon le degré de certitude souhaité.
- Souvent fixé à 5 %.

CALCULS DES DONNÉES

- Distance du χ^2 : $\chi^2(X, Y)$ comme précédemment.
- Valeur critique : c « seuil de validité »de l'hypothèse.

Étapes du test

Hypothèse nulle H_0

H_0 = « X et Y sont indépendantes ».

- *On cherche à infirmer cette hypothèse.*

Choix du risque d'erreur statistique α

- Choix « arbitraire » selon le degré de certitude souhaité.
- Souvent fixé à 5 %.

Calculs des données

- Distance du χ^2 : $\chi^2(X, Y)$ comme précédemment.
- Valeur critique : c « seuil de validité »de l'hypothèse.

Conclusion

- Si $\chi^2(X, Y) > c$, alors H_0 est rejetée au seuil de risque α.
- Si $\chi^2(X, Y) \leqslant c$, alors H_0 est validée au seuil de risque α.

Calcul de la valeur critique

LECTURE DE LA TABLE DU χ^2

Calcul de la valeur critique

LECTURE DE LA TABLE DU χ^2

DÉTERMINATION DE LA LIGNE

- Degrés de liberté : $\nu = (p-1)(q-1)$.
- p : nombre de valeurs de X.
- q : nombre de valeurs de Y.

Calcul de la valeur critique

Lecture de la table du χ^2

Détermination de la ligne

- Degrés de liberté : $\nu = (p-1)(q-1)$.
- p : nombre de valeurs de X.
- q : nombre de valeurs de Y.

Choix de la colonne

- Degré de confiance : $x = 1 - \alpha$.
- α : risque d'erreur statistique.

Table du χ^2

**Fonction de répartition de
la loi du χ^2 à ν degrés de liberté**

$$F(x;\nu) = \frac{1}{\Gamma(\nu/2)} \int_0^{x/2} t^{\nu/2-1} e^{-t}\, dt$$

ν \ x	0,60	0,70	0,75	0,80	0,85	0,90	0,95	0,98	0,99	0,999
1	0,7083	1,0742	1,3233	1,6424	2,0723	2,7055	3,8415	5,4119	6,6349	10,8276
2	1,8326	2,4079	2,7726	3,2189	3,7942	4,6052	5,9915	7,8240	9,2103	13,8155
3	2,9462	3,6649	4,1083	4,6416	5,3170	6,2514	7,8147	9,8374	11,3449	16,2662
4	4,0446	4,8784	5,3853	5,9886	6,7449	7,7794	9,4877	11,6678	13,2767	18,4668
5	5,1319	6,0644	6,6257	7,2893	8,1152	9,2364	11,0705	13,3882	15,0863	20,5150
6	6,2108	7,2311	7,8408	8,5581	9,4461	10,6446	12,5916	15,0332	16,8119	22,4577
7	7,2832	8,3834	9,0371	9,8032	10,7479	12,0170	14,0671	16,6224	18,4753	24,3219
8	8,3505	9,5245	10,2189	11,0301	12,0271	13,3616	15,5073	18,1682	20,0902	26,1245
9	9,4136	10,6564	11,3888	12,2421	13,2880	14,6837	16,9190	19,6790	21,6660	27,8772
10	10,4732	11,7807	12,5489	13,4420	14,5339	15,9872	18,3070	21,1608	23,2093	29,5883
11	11,5298	12,8987	13,7007	14,6314	15,7671	17,2750	19,6751	22,6179	24,7250	31,2641
12	12,5838	14,0111	14,8454	15,8120	16,9893	18,5493	21,0261	24,0540	26,2170	32,9095
13	13,6356	15,1187	15,9839	16,9848	18,2020	19,8119	22,3620	25,4715	27,6882	34,5282
14	14,6853	16,2221	17,1169	18,1508	19,4062	21,0641	23,6848	26,8728	29,1412	36,1233
15	15,7332	17,3217	18,2451	19,3107	20,6030	22,3071	24,9958	28,2595	30,5779	37,6973
16	16,7795	18,4179	19,3689	20,4651	21,7931	23,5418	26,2962	29,6332	31,9999	39,2524
17	17,8244	19,5110	20,4887	21,6146	22,9770	24,7690	27,5871	30,9950	33,4087	40,7902
18	18,8679	20,6014	21,6049	22,7595	24,1555	25,9894	28,8693	32,3462	34,8053	42,3124
19	19,9102	21,6891	22,7178	23,9004	25,3289	27,2036	30,1435	33,6874	36,1909	43,8202
10	20,9514	22,7745	23,8277	25,0375	26,4976	28,4120	31,4104	35,0196	37,5662	45,3147
21	21,9915	23,8578	24,9348	26,1711	27,6620	29,6151	32,6706	36,3434	38,9322	46,7970
22	23,0307	24,9390	26,0393	27,3015	28,8225	30,8133	33,9244	37,6595	40,2894	48,2679
23	24,0689	26,0184	27,1413	28,4288	29,9792	32,0069	35,1725	38,9683	41,6384	49,7282
24	25,1063	27,0960	28,2412	29,5533	31,1325	33,1962	36,4150	40,2704	42,9798	51,1786
25	26,1430	28,1719	29,3389	30,6752	32,2825	34,3816	37,6525	41,5661	44,3141	52,6197
26	27,1789	29,2463	30,4346	31,7946	33,4295	35,5632	38,8851	42,8558	45,6417	54,0520
27	28,2141	30,3193	31,5284	32,9117	34,5736	36,7412	40,1133	44,1400	46,9629	55,4760
28	29,2486	31,3909	32,6205	34,0266	35,7150	37,9159	41,3371	45,4188	48,2782	56,8923
29	30,2825	32,4612	33,7109	35,1394	36,8538	39,0875	42,5570	46,6927	49,5879	58,3012
30	31,3159	33,5302	34,7997	36,2502	37,9903	40,2560	43,7730	47,9618	50,8922	59,7031

Exemple (TD 1 – Exercice 1)

PARAMÈTRES DE LA TABLE DU χ^2

- Degrés de liberté : $\nu = (3-1) \times (4-1) = 6$.
- Degré de confiance : $x = 1 - \alpha = 0,95$.
- Valeur critique : $c = 12,592$.

Exemple (TD 1 – Exercice 1)

Paramètres de la table du χ^2

- Degrés de liberté : $\nu = (3-1) \times (4-1) = 6$.
- Degré de confiance : $x = 1 - \alpha = 0,95$.
- Valeur critique : $c = 12,592$.

Conclusion du test d'indépendance

- $\chi^2(X, Y) = 12,291 < 12,592 = c$.
- L'hypothèse d'indépendance est validée au seuil de risque de 5 %.

Seuil de risque α et degrés de liberté

SEUIL DE RISQUE α

- Une proportion α de tables indépendantes de degré de liberté ν ont une distance du χ^2 (strictement) supérieure à c.

Seuil de risque α et degrés de liberté

Seuil de risque α

- Une proportion α de tables indépendantes de degré de liberté ν ont une distance du χ^2 (strictement) supérieure à c.
- Définition formelle, risque de première espèce :

$$\alpha = \mathbb{P}(\text{rejeter } H_0 | H_0 \text{ est vraie}).$$

Seuil de risque α et degrés de liberté

Seuil de risque α

- Une proportion α de tables indépendantes de degré de liberté ν ont une distance du χ^2 (strictement) supérieure à c.
- Définition formelle, risque de première espèce :

$$\alpha = \mathbb{P}(\text{rejeter } H_0 \big| H_0 \text{ est vraie}).$$

Degrés de liberté

- Combien de choix pour remplir la table de contingence des fréq. ?

Seuil de risque α et degrés de liberté

Seuil de risque α

- Une proportion α de tables indépendantes de degré de liberté ν ont une distance du χ^2 (strictement) supérieure à c.
- Définition formelle, risque de première espèce :

$$\alpha = \mathbb{P}(\text{rejeter } H_0 | H_0 \text{ est vraie}).$$

Degrés de liberté

- Combien de choix pour remplir la table de contingence des fréq. ?
- Fréq. marginales fixées : une ligne et une colonne imposées.
  - *Donc table de taille $(p-1)(q-1)$ à remplir librement.*

Formulation en terme de p-valeur

CARACTÉRISATION

- p-valeur : valeur de risque associée à $\chi^2(X, Y)$ par la loi du χ^2.
- Signification : probabilité d'obtenir la valeur $\chi^2(X, Y)$ de la distance du χ^2 avec des variables indépendantes.

Formulation en terme de p-valeur

Caractérisation

- p-valeur : valeur de risque associée à $\chi^2(X, Y)$ par la loi du χ^2.
- Signification : probabilité d'obtenir la valeur $\chi^2(X, Y)$ de la distance du χ^2 avec des variables indépendantes.

Conclusion du test d'indépendance avec la p-valeur

- Si la p-valeur est strictement inférieure à α, alors H_0 est rejetée au seuil de risque α.
- Si la p-valeur est supérieure (ou égale) à α, alors H_0 est validée au seuil de risque α.

FIN DU CHAPITRE 2