

- Test du χ^2 d'ajustement
- Formalisation
- H_0 :
- H_1 :
- Statistique de test T
- Loi de T sous H_0

Soit $p_0 \in \mathbb{R}^K$ un vecteur de probabilités
(c.-à-d. $\sum_{i=1}^K (p_0)_i = 1$
et $(p_0)_i \geq 0$ pour tout i)

Soient $(N_1, \dots, N_K)$ des v.a. de loi
multinomiale de paramètre p , p inconnu

On souhaite tester $H_0: p = p_0$
 $H_1: p \neq p_0$

Statistique: $T = \sum_{i=1}^K \frac{(N_i - m(p_0)_i)^2}{m(p_0)_i}$

Sous H_0 , T suit asymptotiquement la loi $\chi^2(K-1)$

Exercice 2

	blond	pas blond	total
AC	$\sim p_{AC}$	$\sim p_{BC}$	$p_{AC} + p_{BC}$
pas AC	$\sim p_{AB}$	$\sim p_{BB}$	$p_{AB} + p_{BB}$

$p = p_{AC} + p_{AB}$
on a accès à
un échantillon de
loi multinomiale $(50, p)$

Comparer p à p_{AC}
estimer $\leftrightarrow$ comparer
ajustement

On veut tester si la probabilité de produire des anticorps est affectée par le fait qu'on ait les cheveux blonds.

On tire $n=50$ individus aux cheveux blonds, on compte N_1 le nombre qui produisant les anticorps
 $N_2 = 50 - 10 = 40$ ne pas

Alors $(N_1, N_2) \sim$ multinomiale $(50, p)$ avec p inconnu.
On souhaite comparer p au vecteur $p_{ref} = (p_1^*, p_2^*)$ de la population dans son ensemble, toutes couleurs de cheveux confondues.

$H_0: p = p_{ref}$ $H_1: p \neq p_{ref}$

$$T = \sum_{i=1}^2 \frac{(N_i - m p_{ref}(i))^2}{m p_{ref}(i)} = \frac{(10 - 5)^2}{5} + \frac{(40 - 45)^2}{45} = \frac{25}{5} + \frac{25}{45} = 5 + \frac{5}{9} = \frac{50}{9} \approx 5.56$$

Sous H_0 , dans la limite $n \rightarrow +\infty$ $T \sim \chi^2(2-1) = \chi^2(1)$

Région de rejet: $R = [t_\alpha, +\infty[$ où t_α est la quantile d'ordre $1-\alpha$ de $\chi^2(1)$

pour un niveau α
Pour $\alpha = 5\%$, $t_\alpha = 3.84$

$T^{obs} > t_\alpha$ donc on rejette H_0 : le fait d'avoir les cheveux blonds a un impact significatif sur le fait qu'on produise des anticorps.

Exercice 3

But: Tester si le séjour dans l'espace affecte le génome des bactéries qui s'y développent.

H_0 : pas d'effet H_1 : effet.
On considère 3 mutations A, B, C, dont on connaît la fréquence sur Terre $\rightarrow p_{ref}(A), p_{ref}(B), p_{ref}(C)$
mutuellement exclusives $(p_{ref}(A) + p_{ref}(B) + p_{ref}(C) = 1)$

On cultive $n=66$ échantillons de bactéries dans l'espace et on compte le nombre exprimant chaque mutation.

Alors $N = (N_A, N_B, N_C) \sim$ Multinomiale (n, p) avec p inconnue.

Hypothèses: $H_0: p = p_{ref}$ $H_1: p \neq p_{ref}$. C'est un test d'ajustement.

Statistique: $T = \sum_{i=1}^3 \frac{(N_i - m p_{ref}(i))^2}{m p_{ref}(i)} = \frac{(12 - 66 \times 0.25)^2}{66 \times 0.25} + \frac{(33 - 66 \times 0.4)^2}{66 \times 0.4} + \frac{(21 - 66 \times 0.35)^2}{66 \times 0.35}$

Sous H_0 et dans la limite $n \rightarrow +\infty$, $T \sim \chi^2(3-1) = \chi^2(2)$

Test du χ^2 d'homogénéité

	A	B	...	J	total
1	N_{11}	N_{12}	...	N_{1J}	$N_{1.}$
2	N_{21}	N_{22}	...	N_{2J}	$N_{2.}$
...	...	...	...	...	...
I	N_{I1}	N_{I2}	...	N_{IJ}	$N_{I.}$

p_A, p_B inconnus.

$H_0: p_A = p_B$ vs $H_1: p_A \neq p_B$

Equivalent à tester

$H_0: \tilde{X} \perp \tilde{Y}$ vs $H_1: \tilde{X} \not\perp \tilde{Y}$
où $(\tilde{X}_i, \tilde{Y}_i)_{1 \leq i \leq m_A} = (X_i^A, A)_{1 \leq i \leq m_A}$

et $(\tilde{X}_i, \tilde{Y}_i)_{m_A+1 \leq i \leq m_A+m_B} = (X_i^B, B)_{1 \leq i \leq m_B}$

$$T = \sum_{i=1}^I \frac{(N_{i1} - m_A \frac{N_{i.} N_{.1}}{m_A + m_B})^2}{m_A \frac{N_{i.} N_{.1}}{m_A + m_B}} + \sum_{i=1}^I \frac{(N_{i2} - m_B \frac{N_{i.} N_{.2}}{m_A + m_B})^2}{m_B \frac{N_{i.} N_{.2}}{m_A + m_B}}$$

Sous H_0 , dans la limite $n \rightarrow +\infty$ $T \sim \chi^2(I-1)$
 $= (2-1)(2-1) = 1$

Région de réjet: Soit $\alpha = 5\%$ le niveau souhaité, $R = [t_\alpha, +\infty[$
 avec t_α la quantile d'ordre $1-\alpha$ de la loi $\chi^2(2)$
 donc $t_\alpha = 5.99$.

$T^{obs} = 3.07 < 5.99 = t_{5\%}$ donc on conserve H_0 .

Le fait d'être cultivées dans l'espace n'a pas d'impact significatif sur le génome des bactéries.

Exercice 4

	Admis	Recalé	Tous résultats
M1	19	25	44
M2	41	35	76
M3	17	33	50
toutes M	77	93	170

On veut tester si la méthode a un impact sur l'admission.
 Notons M_i la méthode choisie par l'individu i
 et R_i son résultat au concours.

On observe $n=170$ couples (M_i, R_i) .

On suppose ces couples iid.

On veut tester $H_0: M \perp R$

$H_1: M \not\perp R$

Test du χ^2 d'indépendance

$$T = \frac{(19 - 170 \cdot \frac{44}{170} \cdot \frac{77}{170})^2}{170 \cdot \frac{44}{170} \cdot \frac{77}{170}} + \frac{(25 - 170 \cdot \frac{44}{170} \cdot \frac{93}{170})^2}{170 \cdot \frac{44}{170} \cdot \frac{93}{170}} + \frac{(41 - 170 \cdot \frac{76}{170} \cdot \frac{77}{170})^2}{170 \cdot \frac{76}{170} \cdot \frac{77}{170}} + \dots$$

Fonction R:

chisq.test (tableau)

Sous H_0 et dans limite $n \rightarrow +\infty$

$$T \sim \chi^2((3-1) \times (2-1)) = \chi^2(2)$$

Région de réjet: $R = [t_\alpha, +\infty[$ avec t_α la quantile de niveau $1-\alpha$ de $\chi^2(2)$ et $\alpha = 5\%$ le niveau souhaité.

$T^{obs} = 4.95 < t_\alpha$ donc on conserve H_0 : la méthode n'a pas d'impact significatif.

	A	R	
M1 + M3	36	58	94
M2	41	35	76
	77	93	

$$T = \frac{(36 - 170 \cdot \frac{94}{170} \cdot \frac{77}{170})^2}{170 \cdot \frac{94}{170} \cdot \frac{77}{170}} + \dots$$

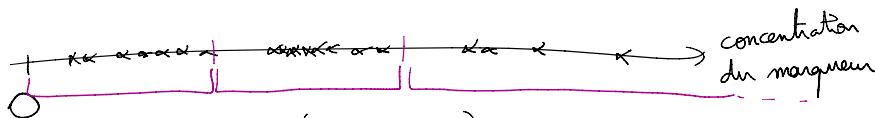
$\sim \chi^2(1)$

donc $R = [3.84, +\infty)$.

$$T^{obs} = 3.55$$

$T^{obs} < t_\alpha = 3.84$: on conserve H_0 .

Exercice 5



On observe n couples (E_i, C_i)
 environnement concentration

$E_i \in \{\text{champ, forêt, bords}$

$C_i \in \{1, 2, 3, 4\}$

On suppose les couples $(E_i, C_i)_{i \in \{1, \dots, n\}}$ iid.

On veut tester

$H_0: E_1 \perp C_1$

$H_1: E_1 \not\perp C_1$

$$T = \sum_{e \in \{E\}} \sum_{c=1}^4 \frac{(N_{ec} - n \frac{N_{e.}}{n} \frac{N_{.c}}{n})^2}{n \frac{N_{e.}}{n} \frac{N_{.c}}{n}}$$

$\sim \chi^2((4-1)(3-1)) = \chi^2(6)$.
 Sous H_0
 limite $n \rightarrow +\infty$

Région de réjet (niveau 5%) = $[12.59, +\infty)$.

2614,703 ~

$$n \frac{N_{e.}}{n} \frac{N_{.c}}{n} \quad \text{avec } n \sim \text{limite } n \rightarrow +\infty$$

Région de rejet (niveau 5%) = $[12,59; +\infty)$.

$$T^{\text{obs}} = 11,77 < 12,59 : \text{on conserve } H_0.$$

L'environnement n'a pas d'impact significatif.

champ
forêt

5	11	37	14
8	23	27	5

test d'homogénéité ou d'indépendance

$$T \sim \chi^2(3)$$

$$R = [7,81; +\infty)$$

$$T^{\text{obs}} = 10,64 > 7,81 : \text{on rejette } H_0.$$