

# INFERENCE STATISTIQUE I :

## TABLEAUX DE COMPTAGES ET TEST DU CHI 2

L'**inférence statistique** consiste à estimer (induire) les caractéristiques inconnues d'une population à partir d'un échantillon aléatoire issu de cette population. Les caractéristiques de l'échantillon, une fois connues, reflètent avec une certaine **marge d'erreur** possible celles de la population. Par exemple, la moyenne d'un échantillon est une estimation de la moyenne de population (souvent dite *espérance*). Quelle est sa marge d'erreur ?

L'inférence statistique est donc un ensemble de méthodes permettant de tirer des conclusions fiables à partir de données d'échantillons statistiques.

Comme l'échantillon est aléatoire, le calcul de la "marge d'erreur" s'appuie sur le calcul des probabilité.

Un grande partie de l'inférence statistique s'appuie sur les tests statistiques. Ce chapitre décrit quelques tests concernant des proportions.

# 1 Comparaison de deux proportions

- On considère deux (sous-)populations  $\mathcal{P}_1$  et  $\mathcal{P}_2$ .
- Leurs individus ont ou n'ont pas le caractère  $C$ .
- On souhaite comparer  $p_1 = P_1(C)$  et  $p_2 = P_2(C)$ , les proportions d'individus avec  $C$  dans  $\mathcal{P}_1$  et  $\mathcal{P}_2$ .

## Echantillonnage et estimation

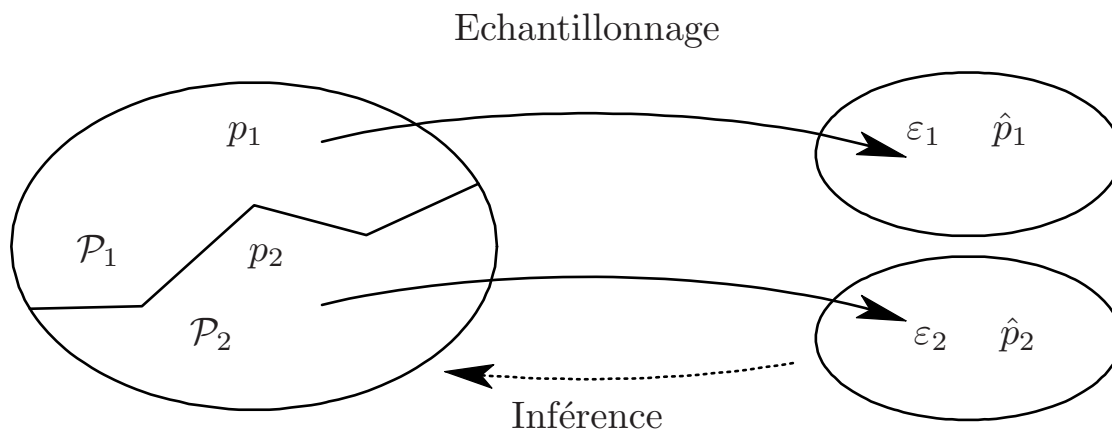
- On prend un **échantillon aléatoire**\*  $\mathcal{E}_1$  de taille  $n_1$ . dans  $\mathcal{P}_1$ ;
- On prend un échantillon aléatoire  $\mathcal{E}_2$  de taille  $n_2$ . dans  $\mathcal{P}_2$ .
- **Table de comptages**:

	$C$	$\bar{C}$	Total
$\mathcal{E}_1$	$n_{11}$	$n_{12}$	$n_1.$
$\mathcal{E}_2$	$n_{21}$	$n_{22}$	$n_2.$

- les fréquences relatives

$$\hat{p}_1 = n_{11}/n_1. \quad \text{et} \quad \hat{p}_2 = n_{21}/n_2.$$

sont des *estimations* de  $p_1$  et  $p_2$ .



“**Inférence**” = transfert des résultats de l'échantillon à la population

\* On dira qu'un échantillon de taille  $n$  est pris de façon aléatoire d'une population si tous les sous-ensembles de taille  $n$  de la population ont les mêmes chances d'être pris.

## Cadre théorique

- Les vraies valeurs de  $p_1$  et  $p_2$  sont inconnues; on formule donc une hypothèse:

$$\mathcal{H}_0 : p_1 = p_2,$$

dite *hypothèse nulle*, et on utilise une *règle de décision* ou *test statistique* pour *rejeter* ou *accepter*  $\mathcal{H}_0$  en s'appuyant sur  $\hat{p}_1$  et  $\hat{p}_2$  et la mesure d'écart standardisé:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}},$$

où  $\hat{p}$  est une estimation de la probabilité de  $C$  commune à  $\mathcal{P}_1$  et  $\mathcal{P}_2$ :

$$\hat{p} = (n_{11} + n_{21}) / (n_{1.} + n_{2.}).$$

- Si on rejette  $\mathcal{H}_0$  on accepte une *hypothèse alternative*  $\mathcal{H}_1$ :

$$\mathcal{H}_1 : p_1 \neq p_2, \quad \text{ou bien} \quad \mathcal{H}_1 : p_1 < p_2, \quad \text{ou encore} \quad \mathcal{H}_1 : p_1 > p_2.$$

- Deux types d'erreurs sont possibles:
  - rejeter une hypothèse nulle bonne: *erreur de type I*,
  - accepter une hypothèse nulle mauvaise: *erreur de type II*.

		R E A L I T E	
		$\mathcal{H}_0$ bonne	$\mathcal{H}_0$ mauvaise
D E C I S I O N	rejeter $\mathcal{H}_0$	erreur type I	OK
	ne pas rejeter $\mathcal{H}_0$	OK	erreur type II

L'erreur de type I peut être contrôlée grâce au résultat théorique suivant.

- *Distribution d'échantillonnage de  $z$*

Si  $p_1 = p_2$ , les valeurs possibles\* de  $z$  ont approximativement une distribution de Gauss standard.

---

\* Les valeurs possibles de  $z$  sont les valeurs de  $z$  calculées à l'aide de tous les échantillons possibles. Noter qu'il ne s'agit pas d'un calcul réel ! La théorie des probabilités nous donne toutefois la distribution de leur population !

### Test *bilatéral*

Hypothèses:  $\mathcal{H}_0 : p_1 = p_2$  contre  $\mathcal{H}_1 : p_1 \neq p_2$

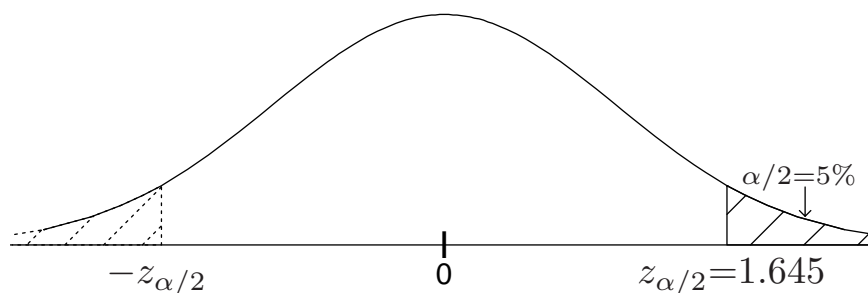
Règle de décision: fixer un *niveau*  $\alpha$  (p.ex. 10% ) et  
rejeter  $\mathcal{H}_0$  si  $z < -z_{\alpha/2}$  ou si  $z > z_{\alpha/2}$ .

La probabilité de rejeter  $\mathcal{H}_0$  si elle est vraie est

$$\text{Prob}(z < -z_{\alpha/2} \text{ ou } z > z_{\alpha/2}) = \alpha$$

ayant défini  $z_{\alpha/2}$  de la façon suivante:

$\alpha$	10%	5%	1%
$z_{\alpha/2}$	1.645	1.960	2.576



Par exemple,  $z_{5\%} = z_{0.05} = 1.645$

La quantité  $z$  est appelée *la statistique de test*.

### Test *unilatéral*

Hypothèses:  $\mathcal{H}_0 : p_1 = p_2$  contre  $\mathcal{H}_1 : p_1 > p_2$

Règle de décision: rejeter  $\mathcal{H}_0$  si  $z > z_\alpha$

La probabilité de rejeter  $\mathcal{H}_0$  si elle est vraie est

$$\text{Prob}(z > z_\alpha) = \alpha$$

ayant défini  $z_\alpha$  de la façon suivante:

$\alpha$	10%	5%	1%
$z_\alpha$	1.282	1.645	2.326

Si  $\mathcal{H}_1$  est bonne, il y a plus de chances de rejeter  $\mathcal{H}_0$  (par rapport au test bilatéral) car  $z_\alpha < z_{\alpha/2}$ ; donc, la probabilité d'erreur de type II est plus faible que dans un test bilatéral.

La probabilité de ne pas commettre une erreur de type II s'appelle la *puissance* du test. Donc, dans le cas d'une alternative unilatérale, un test unilatéral est plus puissant qu'un test bilatéral.

**Exemple 1.** Considérons une étude (fictive) où l'on désire examiner l'association éventuelle entre l'âge maternel et le poids à la naissance de l'enfant. Soit:

$A = \text{âge maternel} \leq 20 \text{ ans}$

$B = \text{poids à la naissance} \leq 2500g.$

On considère deux populations définies par l' "âge de la mère avant la naissance de l'enfant": les mères de moins de 20 ans et les mères de plus de 20 ans. Dans le fichier des naissances d'un grand hôpital on a pris de façon aléatoire 100 mères de moins de 20 ans et 100 mères de plus de 20 ans. On a observé les poids à la naissance de leurs enfants.

Age maternel	Poids à la naissance		Total	Proportion de faibles poids à la naissance
	$\leq 2500g$	$> 2500g$		
$\leq 20$	20	80	100	$0.20 = \hat{p}_1$
$> 20$	10	90	100	$0.10 = \hat{p}_2$
Total	30	170	200	

Tester:  $\mathcal{H}_0 : p_1 = p_2$  contre  $\mathcal{H}_1 : p_1 \neq p_2$  au niveau 5%

$$z = \frac{0.20 - 0.10}{\sqrt{0.15 \times 0.85 \times (1/100 + 1/100)}} = 1.98$$

Comme  $1.98 \notin (-1.96, 1.96)$  on peut rejeter  $\mathcal{H}_0$  au niveau de 5%.

## 2 Logique d'un test statistique

La logique du test présenté ci-dessus est la suivante: pour tester l'hypothèse que  $p_1$  et  $p_2$  sont égales, on s'intéresse à la différence entre les **proportions observées**  $\hat{p}_1$  et  $\hat{p}_2$ . Si cette différence est suffisamment importante, on en conclut que les vraies proportions  $p_1$  et  $p_2$  dans les deux (sous)populations sont différentes. Mais comment décider si la différence que l'on observe dans notre échantillon est suffisamment importante pour conclure que les deux proportions sont effectivement différentes dans les deux populations ? En effet, les deux populations pourraient contenir la même proportion d'enfants de faible poids, et la différence observée ne serait alors due qu'à un hasard malencontreux lors de notre échantillonnage!

Le calcul des probabilités nous permet de déterminer la probabilité que le simple hasard de l'échantillonnage produise une différence au moins aussi importante que la différence observée, dans la situation hypothétique  $\mathcal{H}_0$  où les proportions  $p_1$  et  $p_2$  sont égales.

Voilà qui nous aide dans notre décision: On décide de conclure que les proportions sont différentes si la probabilité que la différence observée soit due au simple hasard de l'échantillonnage est faible.

Le degré de faiblesse requis correspond au **niveau**  $\alpha$  du test. Il est arbitraire et dépendra de l'étude menée et du degré de gravité d'une erreur de type I.

Les règles de décision décrites aux pages 4 et 5 traduisent dans la pratique les considérations ci-dessus:

- Dans le cas du test **bilatéral**, on va rejeter  $\mathcal{H}_0$ , i.e. conclure que les proportions sont *différentes*, si  $z < -z_{\alpha/2}$  ou si  $z > z_{\alpha/2}$ .  $z$  n'est autre qu'une mesure (standardisée) de la différence entre les proportions observées (v. p.2). Si  $z$  est suffisamment grand *en valeur absolue*, on va rejeter  $\mathcal{H}_0$ . Les limites du **domaine de rejet** de  $\mathcal{H}_0$  sont  $-z_{\alpha/2}$  et  $z_{\alpha/2}$ , et elles sont calculées à l'aide de la théorie statistique de façon à atteindre le **but fixé ci-dessus**: si  $z$  dépasse, en valeur absolue,  $z_{\alpha/2}$ , cela signifie que le hasard de l'échantillonnage a une probabilité inférieure à  $\alpha$  de produire une différence au moins aussi importante que la différence observée.

- Dans le cas du test **unilatéral**, on considère a priori qu'il n'est pas possible que  $p_1$  soit inférieur à  $p_2$  (plus de détails au paragraphe 2.1). En conséquence, rejeter  $\mathcal{H}_0$  revient à accepter que  $p_1$  est *supérieur* à  $p_2$  ( $\mathcal{H}_1 : p_1 > p_2$ ). On ne va donc rejeter  $\mathcal{H}_0$  que si  $(p_1 - p_2)$  est *positif*, et suffisamment grand. Le **domaine de rejet** de  $\mathcal{H}_0$  est à présent d'un seul tenant et sa limite  $z_\alpha$  est calculée de sorte que si  $z > z_\alpha$ , le hasard de l'échantillonnage a une probabilité inférieure à  $\alpha$  de produire une différence au moins aussi grande que la différence observée.

NB. Remarquez que, dans ce qui précède, à chaque fois que l'on calcule une probabilité, on le fait conditionnellement à une hypothèse, soit  $\mathcal{H}_0$ , soit  $\mathcal{H}_1$ . Il en va de même de toutes les affirmations que la statistique permet de faire: elles seront toujours de la forme "*Si  $\mathcal{H}_0$  est vraie, alors ...*" ou "*Si  $\mathcal{H}_1$  est vraie, alors ...*".

Attention:

**Le raisonnement du test ne permet pas d'affirmer des choses comme**

*"Il y a une probabilité  $1 - \alpha$  pour que  $\mathcal{H}_0$  soit juste"*.

La logique exposée ci-dessus est la même pour tout test statistique. Seules les hypothèses et la statistique de test (ici  $z$ ) vont changer d'un cas à l'autre.

## 2.1 Test unilatéral ou test bilatéral?

Lorsqu'on fait un test unilatéral, on exclut a priori (i.e. avant l'expérience) que l'une des proportions soit supérieure à l'autre. Par exemple, dans le cas d'un essai clinique, on mène en principe une étude suffisamment poussée *avant* de procéder à l'essai pour pouvoir exclure que le traitement aura un effet plus faible que le placebo. (On ne peut pas exclure la présence d'effets secondaires, mais ce n'est pas sur ceux-ci que porte l'essai.)

Comme signalé plus haut, l'avantage de procéder à un test unilatéral est que la probabilité de commettre une erreur de type II est plus faible que pour un test bilatéral, si le traitement est en effet plus efficace que le placebo. Autrement dit, on aura dans ce cas plus de chance de détecter l'effet du traitement (i.e. de rejeter  $\mathcal{H}_0$ ) que si l'on mène un test bilatéral. Par contre, on renonce complètement à détecter un éventuel effet négatif du médicament.

### 3 Indépendance de deux caractères

- On considère une seule population  $\mathcal{P}$ .
- Les individus ont ou n'ont pas les caractères  $A$  et  $B$ .
- Question: est-ce que  $A$  et  $B$  sont indépendants ?
- On prend un échantillon aléatoire  $\mathcal{E}$  de taille  $n_{..}$ .
- *Table de comptages*

	$B$	$\bar{B}$	Total
$A$	$n_{11}$	$n_{12}$	$n_{1.}$
$\bar{A}$	$n_{21}$	$n_{22}$	$n_{2.}$
	$n_{.1}$	$n_{.2}$	$n_{..}$

- *Estimations*

proportion	estimation de
$\hat{p}_{11} = n_{11}/n_{..}$	$P(A \cap B)$
$\hat{p}_{12} = n_{12}/n_{..}$	$P(A \cap \bar{B})$
$\hat{p}_{21} = n_{21}/n_{..}$	$P(\bar{A} \cap B)$
$\hat{p}_{22} = n_{22}/n_{..}$	$P(\bar{A} \cap \bar{B})$
$\hat{p}_{1.} = n_{1.}/n_{..}$	$P(A)$
$\hat{p}_{2.} = n_{2.}/n_{..}$	$P(\bar{A})$
$\hat{p}_{.1} = n_{.1}/n_{..}$	$P(B)$
$\hat{p}_{.2} = n_{.2}/n_{..}$	$P(\bar{B})$

Soit  $\mathcal{H}_0$ : A et B sont indépendants, donc

$$\begin{aligned}P(A \cap B) &= P(A)P(B), \\P(A \cap \bar{B}) &= P(A)P(\bar{B}), \\P(\bar{A} \cap B) &= P(\bar{A})P(B), \\P(\bar{A} \cap \bar{B}) &= P(\bar{A})P(\bar{B}).\end{aligned}$$

Sous  $\mathcal{H}_0$  on peut estimer les probabilités conjointes (“espérées”) par

$$E_{11} = \hat{p}_1.\hat{p}_{.1}, \quad E_{12} = \hat{p}_1.\hat{p}_{.2}, \quad E_{21} = \hat{p}_2.\hat{p}_{.1}, \quad E_{22} = \hat{p}_2.\hat{p}_{.2}$$

Pour tester  $\mathcal{H}_0$  on comparera les  $E_{ij}$  aux “proportions observées”

$$O_{ij} = \hat{p}_{ij}, \quad i = 1, 2, \quad j = 1, 2,$$

en utilisant la *statistique de test*

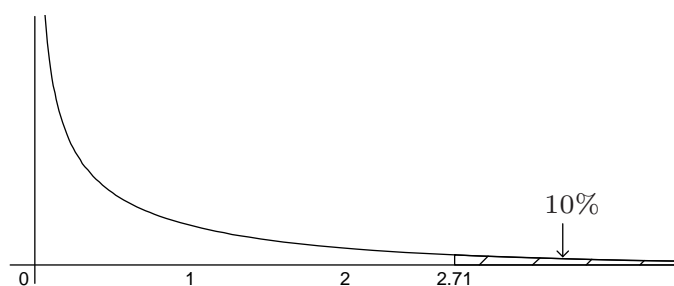
$$\chi^2 = n_{..} \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

On démontre que

$$\chi^2 = \frac{n_{..}(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}.$$

### Résultat théorique

Si  $\mathcal{H}_0$  est vraie,  $\chi^2$  a approximativement une *distribution  $\chi^2$  (à 1 degré de liberté)*.



$\alpha$	10%	5%	1%
$\chi_\alpha^2$	2.71	3.84	6.63

### Test

Hypothèse:  $\mathcal{H}_0$ :  $A$  et  $B$  sont indépendants

Règle de décision: rejeter  $\mathcal{H}_0$  si  $\chi^2 > \chi_\alpha^2$

La probabilité d'erreur de type I est  $\alpha$ .

Le test ci-dessus est appelé *test du chi-carré* et la quantité  $\chi^2$  est appelée *la statistique de test du chi-carré*.

**Exemple 2.** Considérons une étude où l'on désire examiner l'association éventuelle entre l'âge maternel et le poids à la naissance de l'enfant. Soit:

$A = \text{âge maternel} \leq 20 \text{ ans}$

$B = \text{poids à la naissance} \leq 2500g.$

Dans le fichier des naissances d'un grand hôpital on a pris de façon aléatoire un seul échantillon de taille 200. Pour chaque naissance on a observé l'âge maternel et le poids à la naissance.

Age maternel	Poids à la naissance		Total
	$\leq 2500g$	$> 2500g$	
$\leq 20$	10	40	50
$> 20$	15	135	150
Total	25	175	200

Tester  $\mathcal{H}_0$ : "A et B sont indépendants" au niveau 5%.

$$\chi^2 = \frac{200 \times (10 \times 135 - 40 \times 15)^2}{25 \times 175 \times 50 \times 150} = 3.43$$

Comme  $3.43 < 3.84$  on ne peut pas rejeter  $\mathcal{H}_0$  au seuil (niveau)  $\alpha = 5\%$ .

## Remarques

1.  $A$  et  $B$  indépendants signifie aussi  $P(B|A) = P(B|\bar{A})$ .

On peut donc tester  $\mathcal{H}_0$  en utilisant le test **bilatéral** pour la comparaison de deux proportions avec  $p_1 = P(B|A)$  et  $p_2 = P(B|\bar{A})$ . Cette procédure est équivalente au test du  $\chi^2$  et on a:

$$\chi^2 = z^2.$$

De même, pour la règle de décision, on a

$$\chi_\alpha^2 = (z_{\alpha/2})^2.$$

2.  $A$  et  $B$  indépendants signifie aussi  $P(A|B) = P(A|\bar{B})$ .

On peut donc tester  $\mathcal{H}_0$  en utilisant le test **bilatéral** pour la comparaison de deux proportions avec  $p_1 = P(A|B)$  et  $p_2 = P(A|\bar{B})$ .

3. Une étude basée sur un seul échantillon d'individus avec/sans  $A$  et  $B$  est souvent appelée *transversale* (Exemple 2).

4. Si  $A$  est un *facteur antécédent à B*:

- l'étude qui compare un échantillon (d'individus) avec  $A$  à un échantillon sans  $A$  est appelée *prospective* (Exemple 1);
- l'étude qui compare un échantillon avec  $B$  à un échantillon sans  $B$  est appelée *rétrospective* (Exemple 3).

**Exemple 3.** Considérons une étude où l'on désire examiner l'association éventuelle entre l'âge maternel et le poids à la naissance de l'enfant. Soit:

$A = \text{âge maternel} \leq 20 \text{ ans}$

$B = \text{poids à la naissance} \leq 2500g.$

Dans le fichier des naissances d'un grand hôpital on a pris de façon aléatoire 100 naissances d'un poids  $\leq 2500g$  et 100 naissances d'un poids  $> 2500g$ . On a observé l'âge maternel (facteur antécédant).

Poids à la naissance	Age maternel		Total	Proportion de très jeunes mères
	$\leq 20 \text{ ans}$	$> 20 \text{ ans}$		
$\leq 2500g$	40	60	100	.40 [= $\hat{p}(A B)$ ]
$> 2500g$	23	77	100	.23 [= $\hat{p}(A \bar{B})$ ]
Total	63	137	200	

Tester:  $\mathcal{H}_0 : P(A|B) = P(A|\bar{B})$  contre  $\mathcal{H}_1 : P(A|B) \neq P(A|\bar{B})$

$$z = \frac{40/100 - 23/100}{\sqrt{63/200 \times (1 - 63/200) \times (1/100 + 1/100)}} = 2.59$$

$$\chi^2 = \frac{200 \times (40 \times 77 - 23 \times 60)^2}{63 \times 137 \times 100 \times 100} = 6.79 = 2.59^2$$

Comme  $2.59 > 2.575$  ( $6.79 > 6.63$ ) on peut rejeter  $\mathcal{H}_0$  au niveau de 1%.

### *Remarques*

1. Notez que les tableaux des exemples 1, 2 et 3 pourraient représenter des échantillons de la même population (les proportions d'intérêt sont parfaitement cohérentes !). Toutefois, les valeurs de  $\chi^2$  sont différentes.  
La puissance d'un test d'indépendance dépend donc du type d'étude (i.e., de la "méthode d'échantillonnage")!
2. En général, la puissance d'un test dépend aussi de la taille des échantillons. Pour un niveau fixé (par ex.,  $\alpha = 5\%$ ), il est possible de déterminer la taille des échantillons nécessaire pour atteindre une puissance souhaitée (par ex., 90%). (Ce thème n'est pas traité dans ce cours).