Tests de comparaison de pourcentages

Docteur **Alexandrine Lambert**

> Faculté de Pharmacie

Comparer deux pourcentages

- Pourcentage
  - Variable qualitative dichotomique

  (Présence/Absence, Malades/Non malades, Décès/Survie, …)

  - π est le pourcentage (inconnu) d'individus présentant la caractéristique dans la **population**
  - π est estimé par le pourcentage **p** observé sur un **échantillon** de taille n dont k individus présentent la caractéristique

$$p = \frac{k}{n}$$

Comparer deux pourcentages

Comparaison de deux pourcentages dans le cas des grands échantillons.

- Comparaison d'un pourcentage observé à un pourcentage théorique
- Comparaison de deux pourcentages observés
  - Échantillons indépendants
  - Échantillons appariés

Comparer deux pourcentages

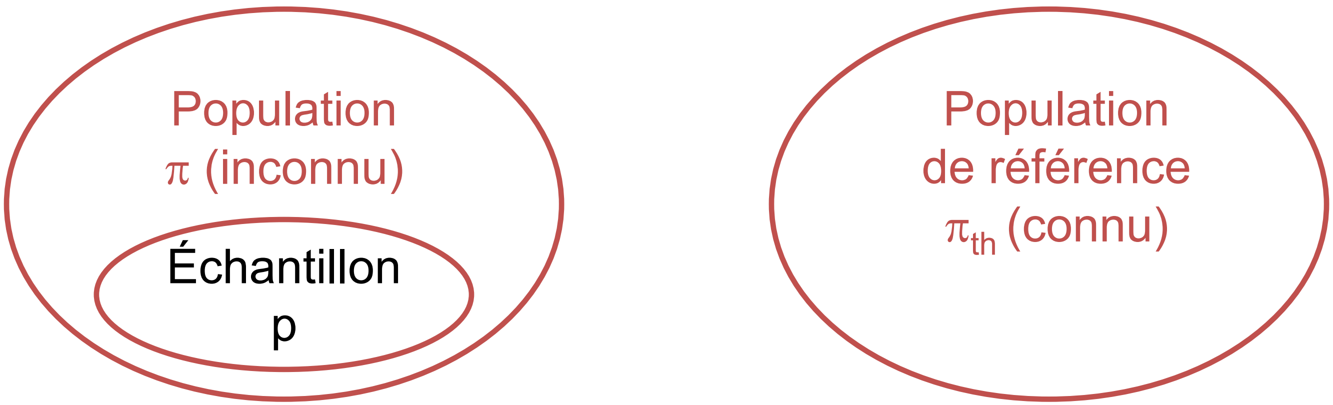
Comparaison de deux pourcentages dans le cas des grands échantillons.

- Comparaison d'un pourcentage observé à un pourcentage théorique
- Comparaison de deux pourcentages observés
  - Échantillons indépendants
  - Échantillons appariés

Comparer un pourcentage à une valeur théorique

- <u>Problème</u> : déterminer si un pourcentage observé p sur un échantillon de taille n est différent d'une valeur théorique π_{th}

➲ Comparer π à π_{th}



Comparer un pourcentage à une valeur théorique

- Formuler une hypothèse
  - Hypothèse nulle H_0
    - $\pi = \pi_{th}$ où π est le pourcentage de la population dont est issu l'échantillon
  - Hypothèses alternatives H_1
    - Test bilatéral : $\pi \neq \pi_{th}$
    - Test unilatéral à gauche ou à droite : $\pi < \pi_{th}$ ou $\pi > \pi_{th}$

Comparer un pourcentage à une valeur théorique

- Fixer le risque α
- Choisir la statistique
  - Test du χ^2 de conformité (loi du X^2)
  - Test z (loi normale)
- Conditions d'application :
  - $n.\,\pi_{th} \geq 5$ et $n.(1- \pi_{th}) \geq 5$ (cas des grands échantillons)

Comparer un pourcentage à une valeur théorique

➲ Test du χ^2 de conformité

- Calculer la valeur χ^2 prise par la statistique du test
  - Tableau

Effectifs observés	$O_1 = n.p$	$O_2 = n.(1-p)$	n
Effectifs calculés	$C_1 = n.\pi_{th}$	$C_2 = n.(1-\pi_{th})$	n

  - Conditions d'application : $C_1 \geq 5$ et $C_2 \geq 5$
  - $\chi^2 = \dfrac{(O_1 - C_1)^2}{C_1} + \dfrac{(O_2 - C_2)^2}{C_2}$
  - Sous H_0 la statistique suit une loi du X^2 à 1 ddl

Comparer un pourcentage à une valeur théorique

➲ **Test du χ^2 de conformité**

- Confronter χ^2 à la valeur seuil $\chi^2_{1,\alpha}$
  - Lecture de la valeur seuil dans la table de la loi du χ^2
  - Test bilatéral : on rejette H_0 au risque α si $\chi^2 \geq \chi^2_{1,\alpha}$
    - En pratique, si α = 5%, $\chi^2_{1,\alpha} = 3{,}84$
      - Si $\chi^2 \geq 3{,}84$: rejet de H_0
      - Si $\chi^2 < 3{,}84$: non rejet de H_0

Comparer un pourcentage à une valeur théorique

- **Exemple 1**
  - En France, 7% des personnes hospitalisées contractent une infection nosocomiale dans l'établissement où elles sont soignées.
  - Sur un échantillon de 250 personnes soignées à l’hôpital H, 28 ont contracté une infection nosocomiale.
  - Le pourcentage observé sur l’échantillon diffère-t-il de la référence nationale au risque $\alpha = 5\%$?

Comparer un pourcentage à une valeur théorique

➲ Test du χ^2 : exemple 1

- Données : $n = 250 \; ; \; p = \frac{28}{250} = 0{,}112 \; ; \; \pi_{th} = 0{,}07$
- Hypothèses : $H_0 : \pi = 0{,}07$; $H_1 : \pi \neq 0{,}07$
- Calcul

	Infection nosocomiale OUI	Infection nosocomiale NON	Total
Effectifs observés	$O_1 = 28$	$O_2 = 222$	250
Effectifs calculés	$C_1 = 250 \times 0{,}07 = 17{,}5$	$C_2 = 250 \times 0{,}93 = 232{,}5$	250

– Conditions d'application vérifiées : $C_1 \geq 5$ et $C_2 \geq 5$

– $\chi^2 = \frac{(28 - 17{,}5)^2}{17{,}5} + \frac{(222 - 232{,}5)^2}{232{,}5} = 6{,}77$

Comparer un pourcentage à une valeur théorique

➲ Test du χ^2 : exemple 1

- Lecture

Table de χ^2

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (ddl)

ddl \ P	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266

- $\chi^2 = 6{,}77 \geq \chi^2_{1,5\%} = 3{,}84$: rejet de H_0

  On montre, au risque 5%, une différence significative entre le pourcentage de personnes hospitalisées contractant une infection nosocomiale à l’hôpital H et dans l’ensemble du pays ($p < 0{,}01$).

Comparer un pourcentage à une valeur théorique

➲ Test z

- Calculer la valeur z prise par la statistique Z
  - $z = \dfrac{p - \pi_{th}}{\sqrt{\dfrac{\pi_{th}.(1 - \pi_{th})}{n}}}$
  - Sous H_0, Z suit une loi normale centrée réduite
  - Conditions d'application : $n.\pi_{th} \geq 5$ et $n.(1- \pi_{th}) \geq 5$

Équivalence entre χ^2 et test z :
$\chi^2 = z^2$

χ^2 à 1 ddl est le carré d'une loi normale centrée réduite

Équivalent à C1 et C2 ≥ 5

Comparer un pourcentage à une valeur théorique

➲ Test z

- Confronter z à la valeur critique z_{α}
  - Test bilatéral : on rejette H_0 si $|z| \geq z_{\alpha}$
  - Test unilatéral :
    - si H_1 s'écrit $\pi > \pi_{th}$, on rejette H_0 si $z \geq z_{2\alpha}$
    - si H_1 s'écrit $\pi < \pi_{th}$, on rejette H_0 si $z \leq -z_{2\alpha}$

Pour un même risque d'erreur, les valeurs seuil du χ^2 sont donc les carrés des valeurs seuil de z : $\chi^2_{1,\alpha}=(z_{\alpha})^2$ (3,84 =1,96^2)

Comparer un pourcentage à une valeur théorique

➲ Test z : exemple 1

- Données : $n = 250 \; ; \; p = \frac{28}{250} = 0{,}112 \; ; \; \pi_{th} = 0{,}07$
- Hypothèses : $H_0 : \pi = 0{,}07$; $H_1 : \pi \neq 0{,}07$
- Calcul
  - $z = \dfrac{0{,}112 - 0{,}07}{\sqrt{\dfrac{0{,}07.(1-0{,}93)}{250}}} = 2{,}60$
  - Conditions d'application vérifiées : $250 \times 0{,}07 \geq 5$ et $250 \times 0{,}93 \geq 5$

Comparer un pourcentage à une valeur théorique

➲ Test z : exemple 1

- Lecture

Table de l'écart-réduit (loi normale)

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860

…

Table pour les petites valeurs de la probabilité

α	1.10^{-3}	1.10^{-4}	1.10^{-5}	1.10^{-6}	1.10^{-7}	1.10^{-8}	1.10^{-9}
ε_α	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

- $z = 2{,}60 \geq z_{0,05} = 1{,}96$: rejet de H_0 *(même conclusion que test précédent)*

  Degré de signification lu dans la table : $p < 0{,}01$

Comparer deux pourcentages

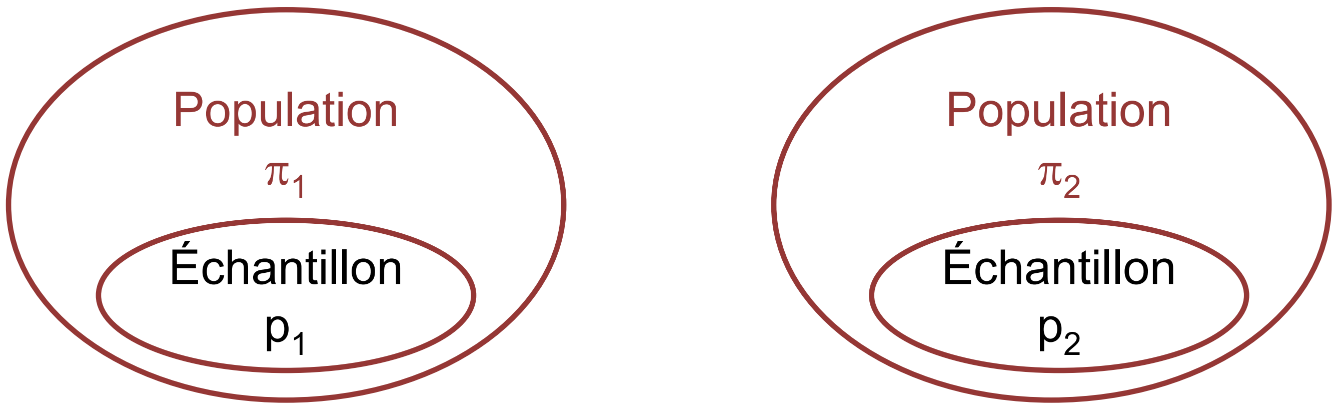
Comparaison de deux pourcentages dans le cas des grands échantillons.

- Comparaison d'un pourcentage observé à un pourcentage théorique
- Comparaison de deux pourcentages observés
  - Échantillons indépendants
  - Échantillons appariés

Comparer 2 pourcentages observés
- Échantillons indépendants -

- <u>Problème</u> : comparer 2 proportions (p_1 et p_2) dans 2 groupes indépendants de tailles n_1 et n_2

  ➲ Comparer π_1 à π_2



Comparer 2 pourcentages observés - Échantillons indépendants -

- Formuler une hypothèse
  - Hypothèse nulle H_0
    - Les 2 échantillons sont issus de la même population ayant comme pourcentage π_0

      $\pi_1 = \pi_2$ (= π_0) où π_1 et π_2 pourcentages de la population dont sont issus les échantillons 1 et 2
  - Hypothèses alternatives H_1
    - Test bilatéral : $\pi_1 \neq \pi_2$
    - Test unilatéral : $\pi_1 < \pi_2$ ou $\pi_1 > \pi_2$

Comparer 2 pourcentages observés
- Échantillons indépendants -

- Fixer le risque α
- Choisir la statistique :
  - Test du χ^2 (loi du χ^2)
  - Test z (loi normale)
- Conditions d'application :
  - $n_1 . \pi_0 \geq 5$ et $n_1 .(1- \pi_0) \geq 5$
  - $n_2 . \pi_0 \geq 5$ et $n_1 .(1- \pi_0) \geq 5$

Comparer 2 pourcentages observés - Échantillons indépendants -

➲ Test du χ^2

- Calculer la valeur χ^2 prise par la statistique
  - Tableau de contingence (tableau à 4 cases)

	Succès	Échec	Total
Groupe 1	O_{11} (C_{11})	O_{12} (C_{12})	n_1
Groupe 2	O_{21} (C_{21})	O_{22} (C_{22})	n_2
Total	n'_1	n'_2	N

Effectifs calculés sous H_0 :

$$C_{ij} = \frac{n'_i \, n_j}{N}$$

  - Conditions d'application : $C_{ij} \geq 5$
  - $$\chi^2 = \sum_{i,j} \frac{(O_{ij} - C_{ij})^2}{C_{ij}}$$
  - Sous H_0 la statistique suit une loi du X^2 à 1 ddl

Comparer 2 pourcentages observés
- Échantillons indépendants -

➲ Test du χ^2 : Remarques

- Dans le cas des tableaux de contingence à 4 cases, il est possible d'utiliser la correction de continuité, surtout lorsque les valeurs attendues sont faibles (en pratique $C_{ij} < 5$)

$$\chi^2 = \sum_{i,j} \frac{\left(\left|O_{ij} - C_{ij}\right| - 0,5\right)^2}{C_{ij}}$$

- Petits échantillons : test exact de Fisher (hors programme)

Comparer 2 pourcentages observés
- Échantillons indépendants -

➲ **Test du χ^2**

- Confronter χ^2 à la valeur critique $\chi^2_{1,\alpha}$
  - Lecture de la valeur seuil dans la table
  - Test bilatéral :
    - Si $\chi^2 \geq \chi^2_{1,\alpha}$: rejet de H_0
    - Si $\chi^2 < \chi^2_{1,\alpha}$: non rejet de H_0

Comparer 2 pourcentages observés - Échantillons indépendants -

- **Exemple 2**
  - On désire comparer l'efficacité de deux traitements T1 et T2 sur 100 patients atteints d'une maladie M.
  - On tire au sort 2 deux groupes de 50 patients, un groupe est soumis à T1, le second à T2.
  - Le pourcentage de guérison chez les patients soumis à T1 est de 30%, chez ceux soumis à T2 de 40%.
  - Le taux de guérison est-il différent entre les 2 traitements ?

Comparer 2 pourcentages observés - Échantillons indépendants -

➲ Test du χ² : exemple 2

- Données : $n_1 = 50\,;\ p_1 = 0{,}3$ et $n_2 = 50\,;\ p_2 = 0{,}4$
- Hypothèses : $H_0 : \pi_1 = \pi_2$; $H_1 : \pi_1 \neq \pi_2$
- Calcul

	Guéris	Non guéris	Total
Groupe T1	15 (17,5)	35 (32,5)	50
Groupe T2	20 (17,5)	30 (32,5)	50
Total	35	65	100

  - Conditions d'application vérifiées : $C_{ij} \geq 5$
  - $$\chi^2 = \frac{(15-17{,}5)^2}{17{,}5} + \frac{(20-17{,}5)^2}{17{,}5} + \frac{(35-32{,}5)^2}{32{,}5} + \frac{(30-32{,}5)^2}{32{,}5} = 1{,}10$$

Comparer 2 pourcentages observés - Échantillons indépendants -

➲ Test du χ^2 : exemple 2

- Lecture

Table de χ^2

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (ddl)

ddl \ P	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266

- $\chi^2 = 1,10 < \chi^2_{1,5\%} = 3,84$: H_0 acceptable.

On ne met pas en évidence, au risque 5%, de différence significative entre les taux de guérison avec les 2 traitements

Comparer 2 pourcentages observés
- Échantillons indépendants -

➲ Test z

- Calculer la valeur z prise par la statistique Z
  - $z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_0.(1-p_0)}{n_1} + \dfrac{p_0.(1-p_0)}{n_2}}}$ avec $p_0 = \dfrac{n_1.p_1 + n_2.p_2}{n_1 + n_2}$ *$X^2 = (z)^2$*
  - p_0 est l'estimation de la proportion commune π_0
  - Z suit une loi normale centrée réduite
  - Conditions d'application :
    - $n_1.\ \pi_0 \geq 5$ et $n_1.(1-\ \pi_0) \geq 5$
    - $n_2.\ \pi_0 \geq 5$ et $n_2.(1-\ \pi_0) \geq 5$

X^2 à 1 ddl est le carré d'une loi normale centrée réduite

$C_{ij} \geq 5$

Comparer 2 pourcentages observés - Échantillons indépendants -

➲ **Test z**

- Confronter z à la valeur critique z_{α}
  - Test bilatéral : on rejette H_0 si $|z| \geq z_{\alpha}$
  - Test unilatéral :
    - si H_1 s'écrit $\pi_1 > \pi_2$, on rejette H_0 si $z \geq z_{2\alpha}$
    - si H_1 s'écrit $\pi_1 < \pi_2$, on rejette H_0 si $z \leq -z_{2\alpha}$

Comparer 2 pourcentages observés
- Échantillons indépendants -

➲ Test z : exemple 2

- Données : $n_1 = 50$; $p_1 = 0,3$ et $n_2 = 50$; $p_2 = 0,4$
- Hypothèses : $H_0 : \pi_1 = \pi_2$; $H_1 : \pi_1 \neq \pi_2$
- Calcul

$$p_0 = \frac{50x0,3 + 50x0,4}{50 + 50} = 0,35$$

$$z = \frac{0,3 - 0,4}{\sqrt{\frac{0,35x0,65}{50} + \frac{0,35x0,65}{50}}} = 1,05$$

  – Conditions d'application vérifiées : 50 x 0,35 ≥ 5 et 50 x 0,65 ≥ 5

Comparer 2 pourcentages observés
- Échantillons indépendants -

➲ Test z : exemple 2

- Lecture

Table de l'écart-réduit (loi normale)

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860

…

- $z = 1{,}05 < z_{0,05} = 1{,}96$: H_0 acceptable

(*Même conclusion que le test précédent*)

Comparer deux pourcentages

Comparaison de deux pourcentages dans le cas des grands échantillons.

- Comparaison d'un pourcentage observé à un pourcentage théorique
- Comparaison de deux pourcentages observés
  - Échantillons indépendants
  - Échantillons appariés

Comparer 2 pourcentages observés
- Séries appariées -

- Variable aléatoire qualitative dichotomique
- Cas des grands échantillons
- Individus de 2 échantillons liés
  - Présence d'une caractéristique sur les mêmes sujets
  - Présence d'une caractéristique chez des sujets appariés
- <u>Problème</u> : on s'intéresse aux taux de guérison chez des sujets ayant reçus un traitement T1 et des sujets **appariés** ayant reçus un traitement T2 : on cherche à comparer p_1 et p_2 les taux de guérison avec T1 et T2.

Comparer 2 pourcentages observés
- Séries appariées -

- Formuler une hypothèse
  - Hypothèse nulle H_0
    - $\pi_1 = \pi_2$ où π_1 et π_2 pourcentages inconnus des 2 populations d'où sont issus les échantillons
  - Hypothèses alternatives H_1
    - Test bilatéral : $\pi_1 \neq \pi_2$
    - Test unilatéral : $\pi_1 < \pi_2$ ou $\pi_1 > \pi_2$

Comparer 2 pourcentages observés - Séries appariées -

- Tableau des valeurs
  - Pour tenir compte de l'appariement, il faut faire apparaître quels sont les sujets qui appartiennent aux mêmes paires.
  - Pour chaque paire d'individus, on peut observer, selon s'il y a présence (+) ou absence (-) du caractère étudié, l'une des 4 configurations possibles.

Échantillon 1	Échantillon 2	Nombre de paires
+	+	a
+	-	b
-	+	c
-	-	d

Comparer 2 pourcentages observés - Séries appariées -

Éch. 2

		+	-	Total
Éch. 1	+	a	b	a+b
	-	c	d	c+d
	Total	a+c	b+d	**n**

– Les paires concordantes n'apportent pas d'information sur la liaison entre le traitement et la guérison. On doit donc se fonder sur la répartition des paires discordantes.
– Si l'hypothèse H_0 est vraie, il doit y avoir autant de paires discordantes du type +- que de type -+
– Tester H_0 revient donc à tester si le pourcentage observé de paires -+ est significativement différent de la valeur théorique 0,5.

Comparer 2 pourcentages observés - Séries appariées -

- Fixer le risque α
- Choisir la statistique
  - Test du χ^2 de McNemar (loi du X^2)
  - Test z (loi normale)
  - Conditions d'application

Comparer 2 pourcentages observés
- Séries appariées -

➲ Test de McNemar (χ^2)

- Calculer la valeur χ^2 prise par la statistique

	+-	-+	
Effectifs observés	b	c	b+c
Effectifs calculés	$\frac{b+c}{2}$	$\frac{b+c}{2}$	b+c

  - Conditions d'application : $\frac{b+c}{2} \geq 5$
  - $\chi^2 = \frac{\left(b - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} + \frac{\left(c - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c}$
  - La statistique suit une loi du X^2 à 1 ddl

Comparer 2 pourcentages observés
- Séries appariées -

➲ Test de McNemar (χ^2) : remarques

- Il est possible d'utiliser la correction de continuité, surtout lorsque les valeurs attendues sont faibles

$$\chi_0^2 = \frac{\left(\left|b - \frac{b+c}{2}\right| - 0{,}5\right)^2}{\frac{b+c}{2}} + \frac{\left(\left|c - \frac{b+c}{2}\right| - 0{,}5\right)^2}{\frac{b+c}{2}} = \frac{\left(\left|b-c\right|-1\right)^2}{b+c}$$

Comparer 2 pourcentages observés - Séries appariées -

➲ **Test de McNemar (χ^2)**

- Confronter χ^2 à la valeur critique $\chi^2_{1,\alpha}$
  - Lecture de la valeur seuil dans la table
  - Test bilatéral : on rejette H_0 si
    - Si $\chi^2 \geq \chi^2_{1,\alpha}$: rejet de H_0
    - Si $\chi^2 < \chi^2_{1,\alpha}$: non rejet de H_0

Comparer 2 pourcentages observés
- Séries appariées -

➲ **Test z**

- Calculer la valeur z prise par la statistique Z
  - $z = \dfrac{b - \dfrac{b+c}{2}}{\sqrt{(b+c).0{,}5.0{,}5}} = \dfrac{b-c}{\sqrt{b+c}}$

$X^2 = (z)^2$

  - Z suit une loi normale centrée réduite

X^2 à 1 ddl est le carré d'une loi normale centrée réduite

  - Conditions d'application : b+c ≥ 10

Comparer 2 pourcentages observés
- Séries appariées -

➲ **Test z**

- Confronter z à la valeur critique $z_{\boldsymbol{\alpha}}$
  - Test bilatéral : on rejette H_0 si $|z| \geq z_\alpha$
  - Test unilatéral :
    - si H_1 s'écrit $\pi_1 > \pi_2$, on rejette H_0 si $z \geq z_{2\alpha}$
    - si H_1 s'écrit $\pi_1 < \pi_2$, on rejette H_0 si $z \leq -z_{2\alpha}$

Comparer 2 pourcentages observés - Séries appariées -

- **Exemple 3**
  - On désire comparer l'efficacité de deux traitements T_1 et T_2 chez 100 patients atteint d'une maladie M.
  - Les deux traitements sont administrés aux patients. L'ordre d'administration des 2 traitements est tiré au sort en ménageant une période dite de *wash-out* entre les 2 administrations.
  - Les résultats sont les suivants :

		T_1	
		Succès	Échec
T_2	Succès	24	16
	Échec	6	54

  - Le taux de guérison est-il différent entre les deux traitements ?

Comparer 2 pourcentages observés - Séries appariées -

➲ Test de McNemar : exemple 3

- On cherche à comparer les pourcentages observés :

$$p_1 = \frac{24+6}{100} = 0{,}3 \qquad p_2 = \frac{24+16}{100} = 0{,}4$$

- Hypothèses : $H_0 : \pi_1 = \pi_2$; $H_1 : \pi_1 \neq \pi_2$
- Conditions d'application vérifiées : nombre de paires discordantes = 16 + 6 = 22 ≥ 10
- $\chi^2 = \frac{(16-6)^2}{(16+6)} = 4{,}55$

Comparer 2 pourcentages observés
- Séries appariées -

➲ Test de McNemar : exemple 3

- Lecture

Table de χ^2

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (ddl)

ddl \ P	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266

- $\chi^2 = 4{,}55 \geq \chi^2_{1,5\%} = 3{,}84$: H_0 rejetée

  On montre, au risque 5%, une différence significative entre les taux de guérison avec les 2 traitements (p<0,05).

L1 SANTE

Nancy-Université
Université Henri Poincaré

ministère éducation nationale enseignement supérieur recherche

RÉPUBLIQUE FRANÇAISE