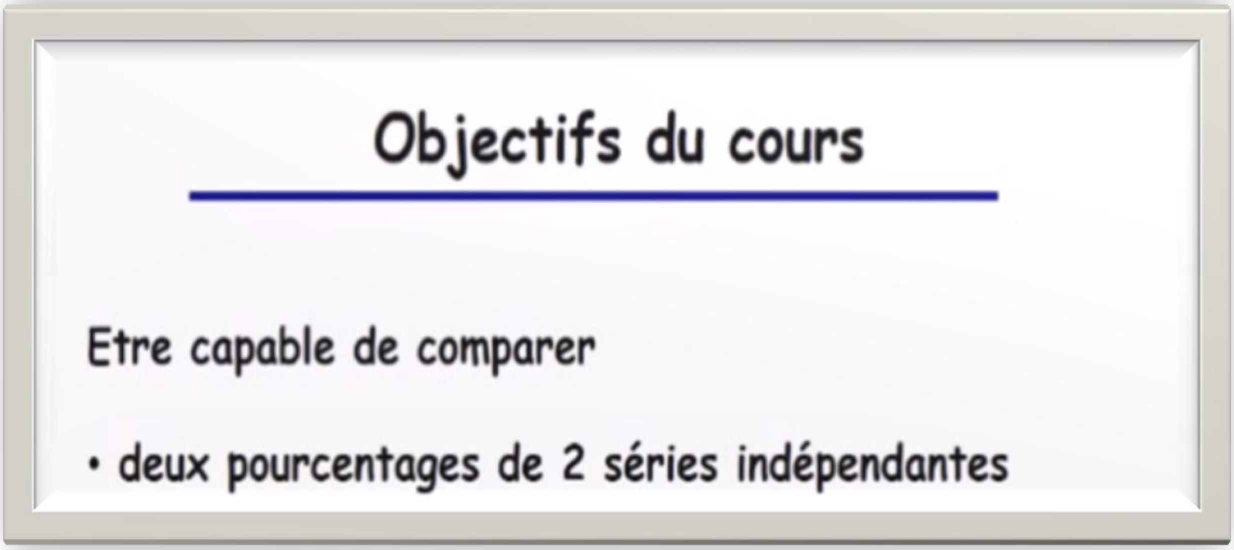
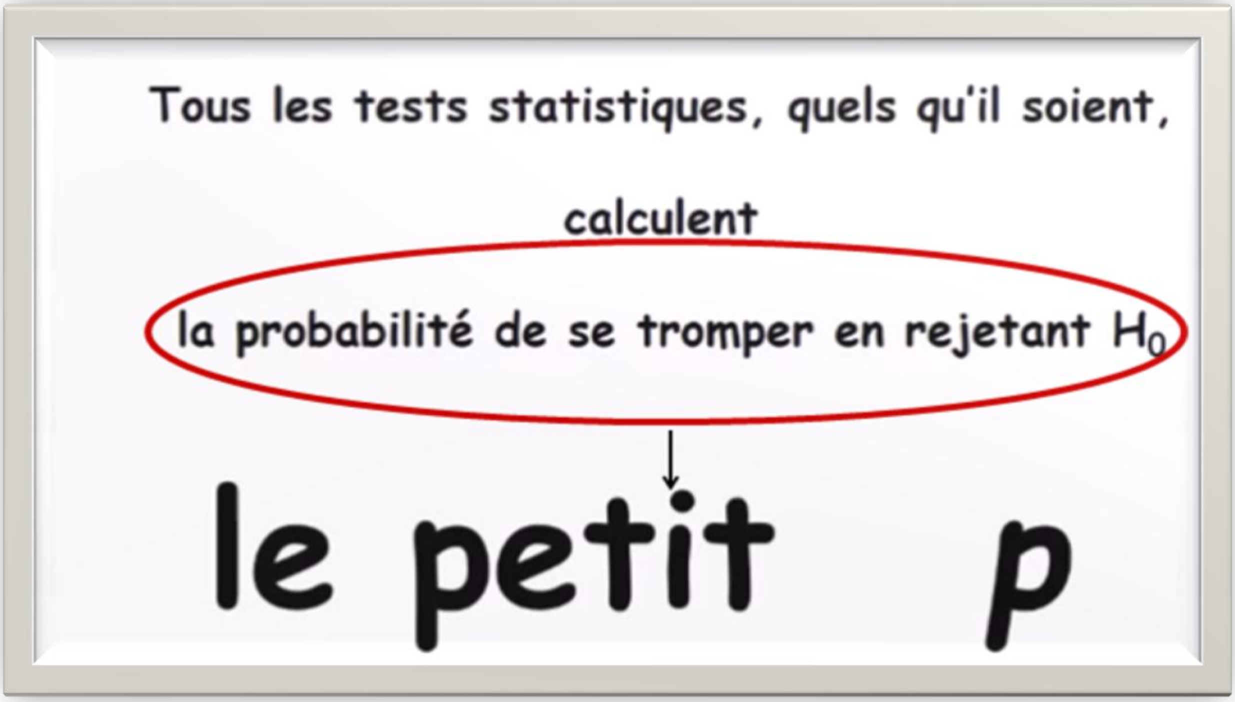
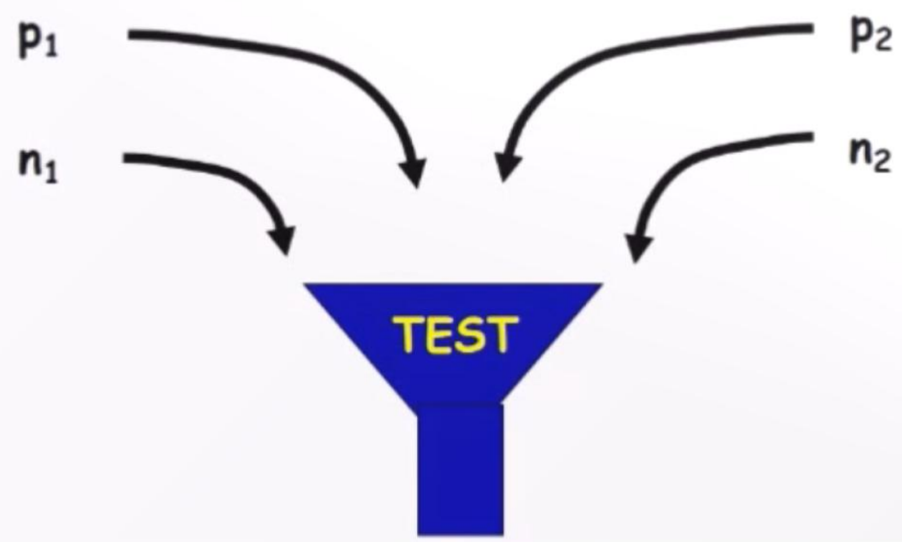
TESTS STATISTIQUES
COMPARAISON DE POURCENTAGES
PR BOUZBID

Faculté de Médecine d'Annaba
Biomathématiques Statistiques
1 ère année Médecine Dentaire





Test statistique



Comment comparer deux pourcentages de deux séries indépendantes?

- Test exact de Fisher
- Test du chi² (χ^2) de Pearson

Exemple 1

Epidémie de gastro-entérite

	Malades	Sains	Taux d'attaque
Huitres +	41	68	37,6 %
Huitres –	13	48	21,3%

Problème ?

Avait-t-on vraiment plus de risques de contracter une gastro-entérite après avoir mangé des huitres?

Y-a-t-il vraiment un taux d'attaque plus élevé chez les mangeurs d'huitres que chez les non-mangeurs ?

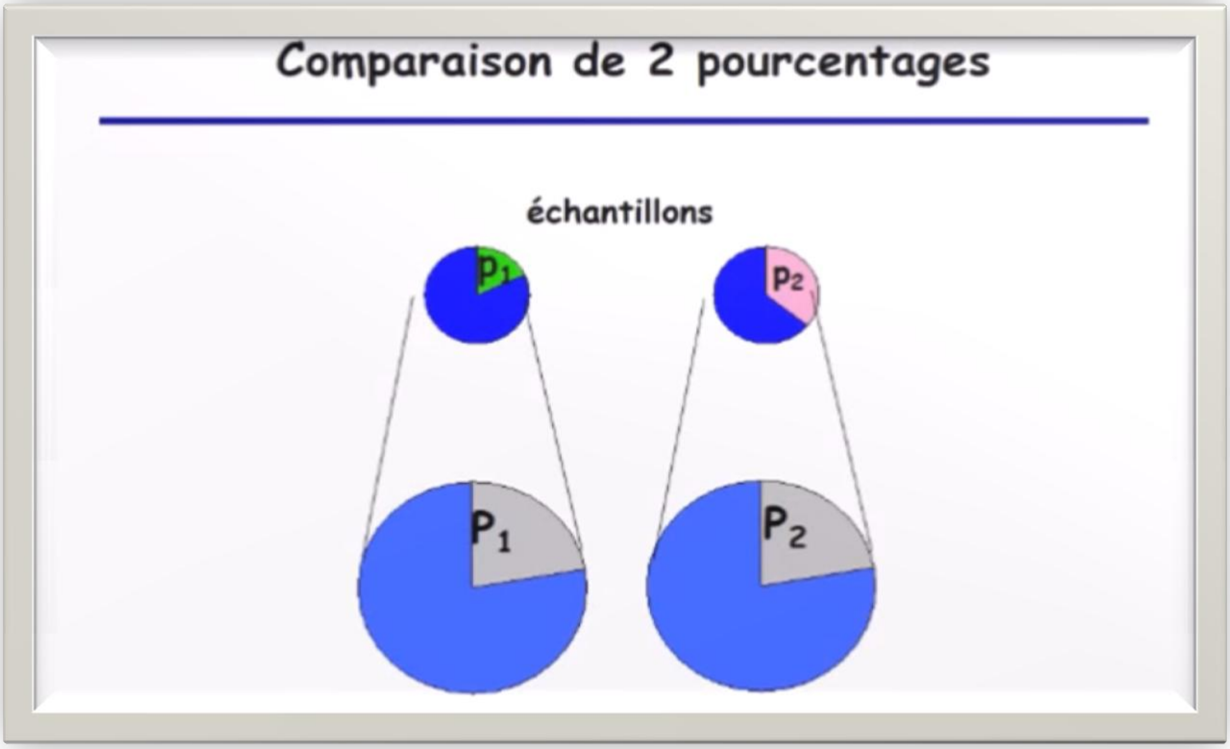
Exemple 1

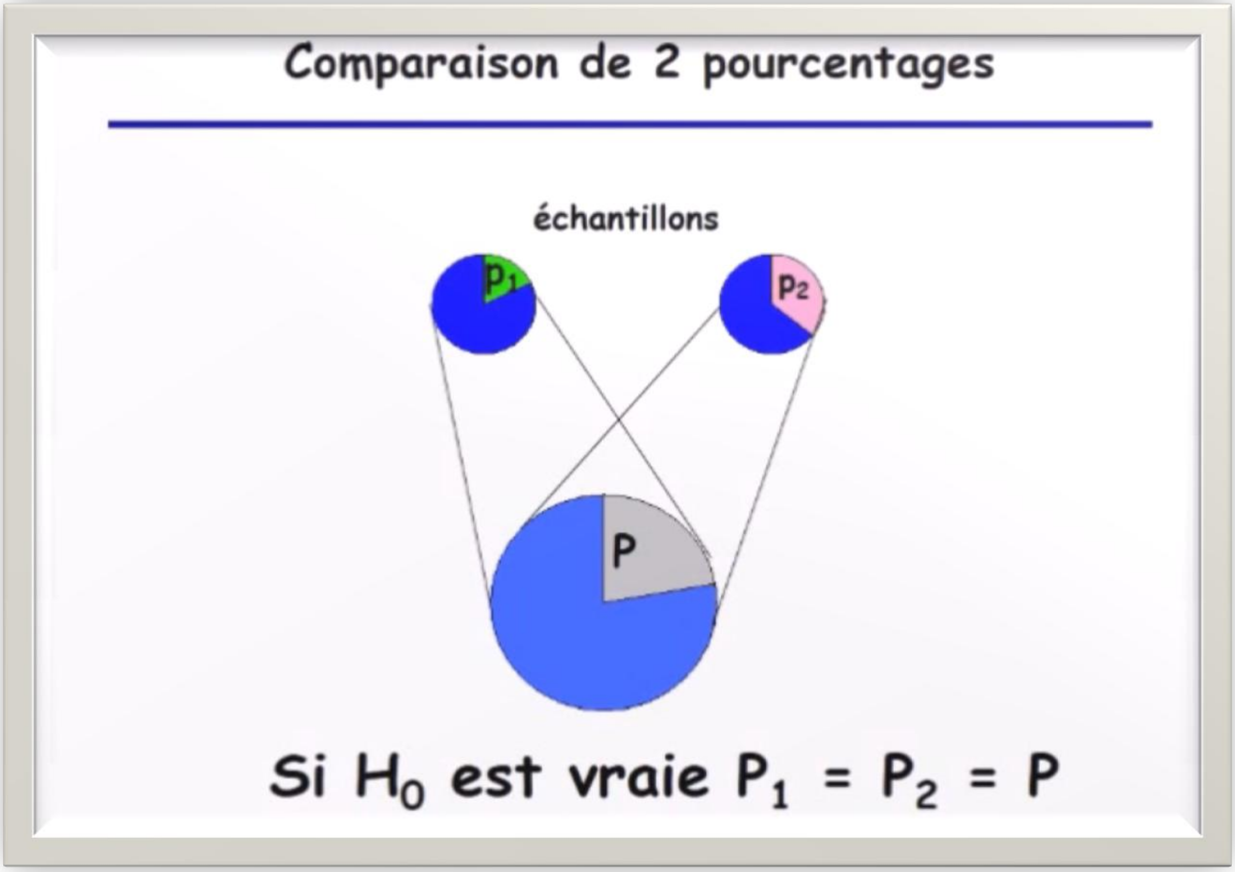
La fréquence de gastro-entérite est-elle plus élevée chez les mangeurs d'huitres ?

Mangeurs d'huitres : taux d'attaque = 37,6 %
Non-mangeurs : taux d'attaque = 21,3 %

H_0 : taux d'attaque$_{mangeurs}$ = taux d'attaque $_{non\text{-}mangeurs}$

H_1 : taux d'attaque$_{mangeurs}$ # taux d'attaque $_{non\text{-}mangeurs}$


Comparaison de 2 pourcentages
échantillons
p1
p2
P1
P2


Comparaison de 2 pourcentages
échantillons
p1
p2
P
Si H0 est vraie P1 = P2 = P

Exemple 1

Epidémie de gastro-entérite

	Malades	Sains	Taux d'attaque
Huitres +	41	68	37,6 %
Huitres -	13	48	21,3%

Si H_0 est vrai,
la différence Δ entre les 2 pourcentages
devrait être proche de zéro

Or Δ = 37,6 % - 21,3 % = 16,3 %

Est-ce que 16,3 est suffisamment éloigné de zéro
pour qu'on rejette H_0 ?

Principe des tests de comparaison de pourcentages

Calculer la probabilité, si H_0 est vrai,
d'obtenir une configuration montrant
une différence au moins aussi grande
que celle qui a été observée

1) Si cette probabilité est faible (< 5%) , alors on considère

- que la configuration observée n'est pas due au hasard
- on rejette H_0
- on accepte H_1

2) Si cette probabilité est élevée, alors

- on ne rejette pas H_0
- on considère que la différence est due à une simple fluctuation d'échantillonage

Test exact de Fisher

pour comparer deux pourcentages

Principe du test exact de Fisher (1)

1) Calculer la probabilité, si H_0 est vrai,
d'obtenir la configuration observée

a	b	t_1
c	d	t_2
n_1	n_2	N

$$p = \frac{n_1!\, n_2!\, t_1!\, t_2!}{a!\, b!\, c!\, d!\, N!}$$

Principe du test exact de Fisher (1)

1) Calculer la probabilité, si H_0 est vrai, d'obtenir la configuration observée

	Malades	Sains	Total
Huitres +	41	68	109
Huitres –	13	48	61
Total	54	116	170

$$\Delta = 37{,}6\ \% - 21{,}3\ \% = 16{,}3\ \%$$

$$p = \frac{54!\ 116!\ 109!\ 61!}{41!\ 68!\ 13!\ 48!\ 170!} = 0{,}012$$

Principe du test exact de Fisher (2)

2) Calculer toutes les probabilités d'obtenir, si H_0 est vrai, des configurations aboutissant à une différence encore plus grande

	M	S	T.A	$\lvert\Delta\rvert$	p
+	41	68	37,6	16,3	0,012
-	13	48	21,3		
+	42	67	38,5	18,8	0,0052
-	12	49	19,7		
+	43	66	39,4	21,4	0,0020
-	11	50	18,0		
+	44	65	40,4	24,0	0,0006
-	10	51	16,4		

Lorsque le taux d'attaque est plus élevé chez les mangeurs

	M	S	T.A	$\lvert\Delta\rvert$	p
+	28	81	25,7	16,9	0,01
-	26	35	42,6		
+	27	82	24,8	19,5	0,005
-	27	34	44,3		

Hypothèse alternative bilatérale

Lorsque le taux d'attaque est plus élevé chez les non mangeurs

Principe du test exact de Fisher (3)

3) Additionner toutes les p d'obtenir, des configurations aboutissant à une différence encore plus grande

	M	S	T.A	|Δ|	p
+	41	68	37,6	16,3	0,012
-	13	48	21,3		
+	42	67	38,5	18,8	0,0052
-	12	49	19,7		
+	43	66	39,4	21,4	0,0020
-	11	50	18,0		

	M	S	T.A	|Δ|	p
+	28	81	25,7	16,9	0,01
-	26	35	42,6		
+	27	82	24,8	19,5	0,005
-	27	34	44,3		
+					
-					

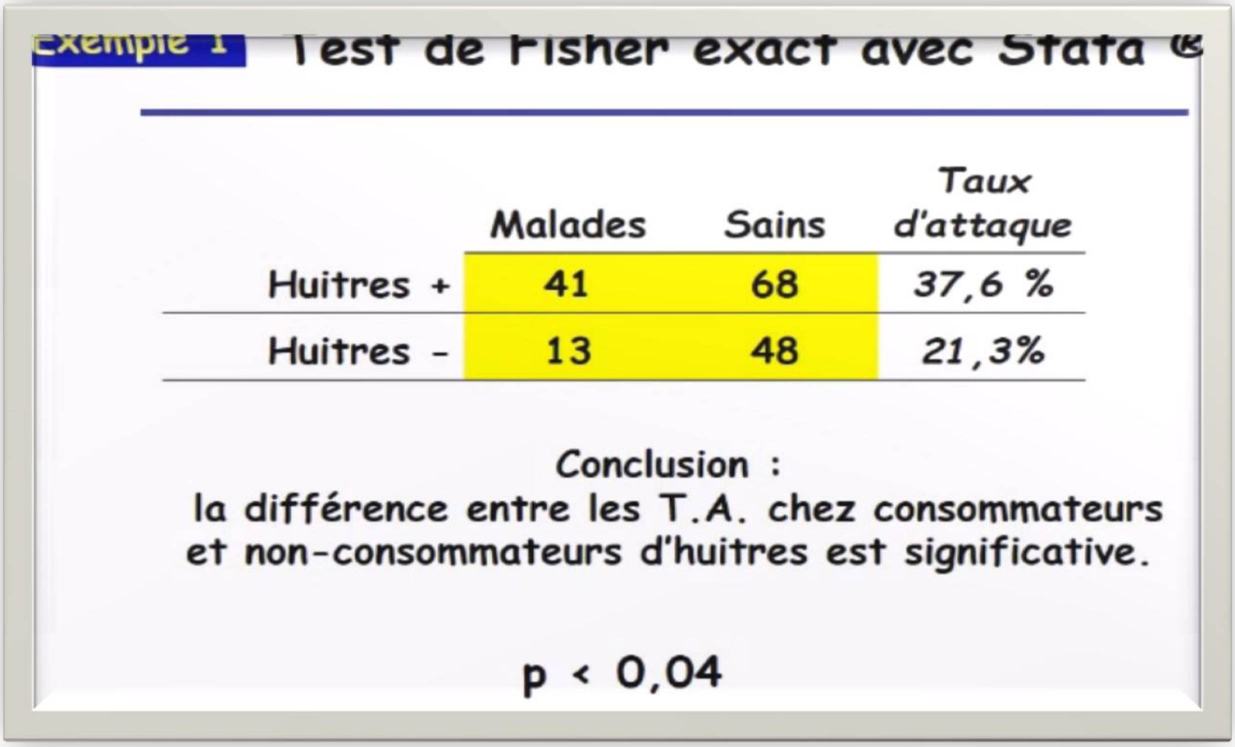
p = 0,12 + 0,0052 + 0,002 +...+ 0.01 + 0,005 +...

p = 0,039

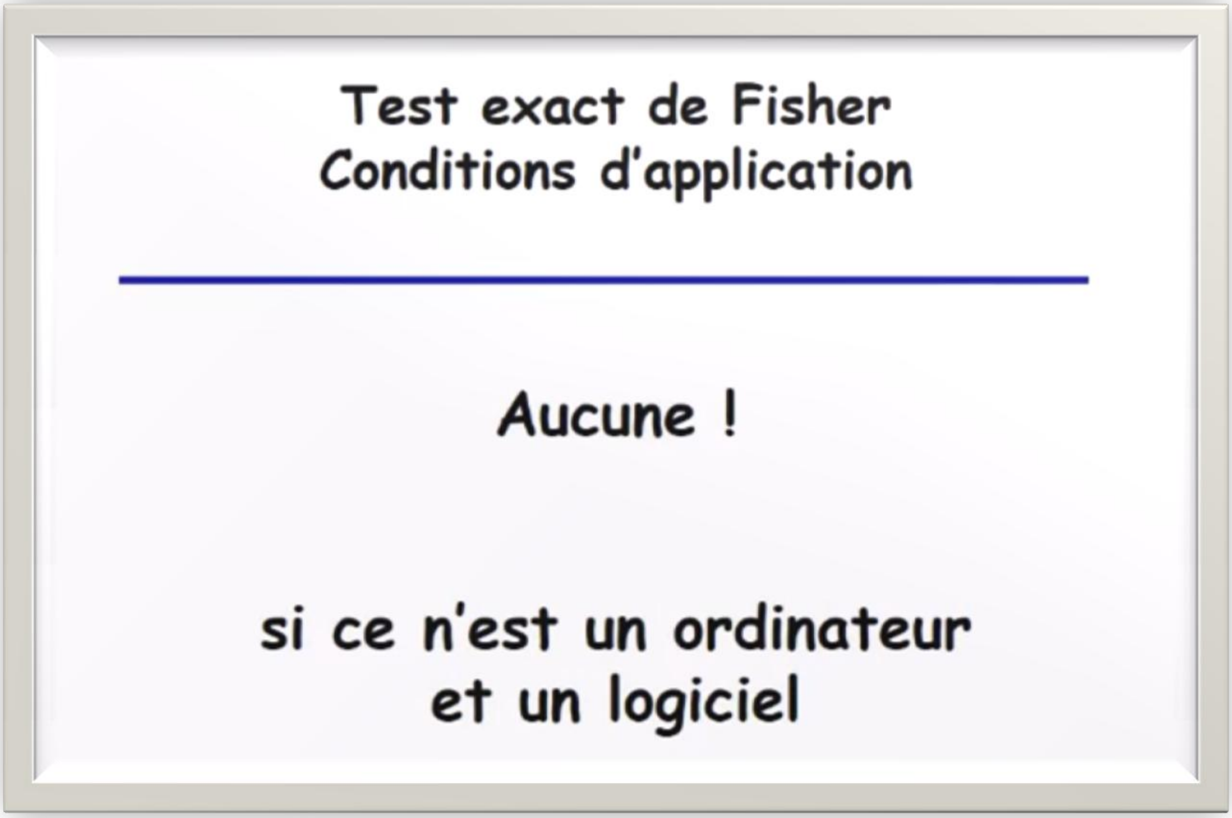
Principe du test exact de Fisher (3)

3) Additionner toutes les probabilités d'obtenir,
si H_0 est vrai,
des configurations aboutissant
à une différence encore plus grande
que celle qui est observée

$$p = \sum \frac{n_1!\, n_2!\, t_1!\, t_2!}{a!\, b!\, c!\, d!\, N!}$$



	Malades	Sains	Taux d'attaque
Huitres +	41	68	37,6 %
Huitres -	13	48	21,3%



Test de χ^2

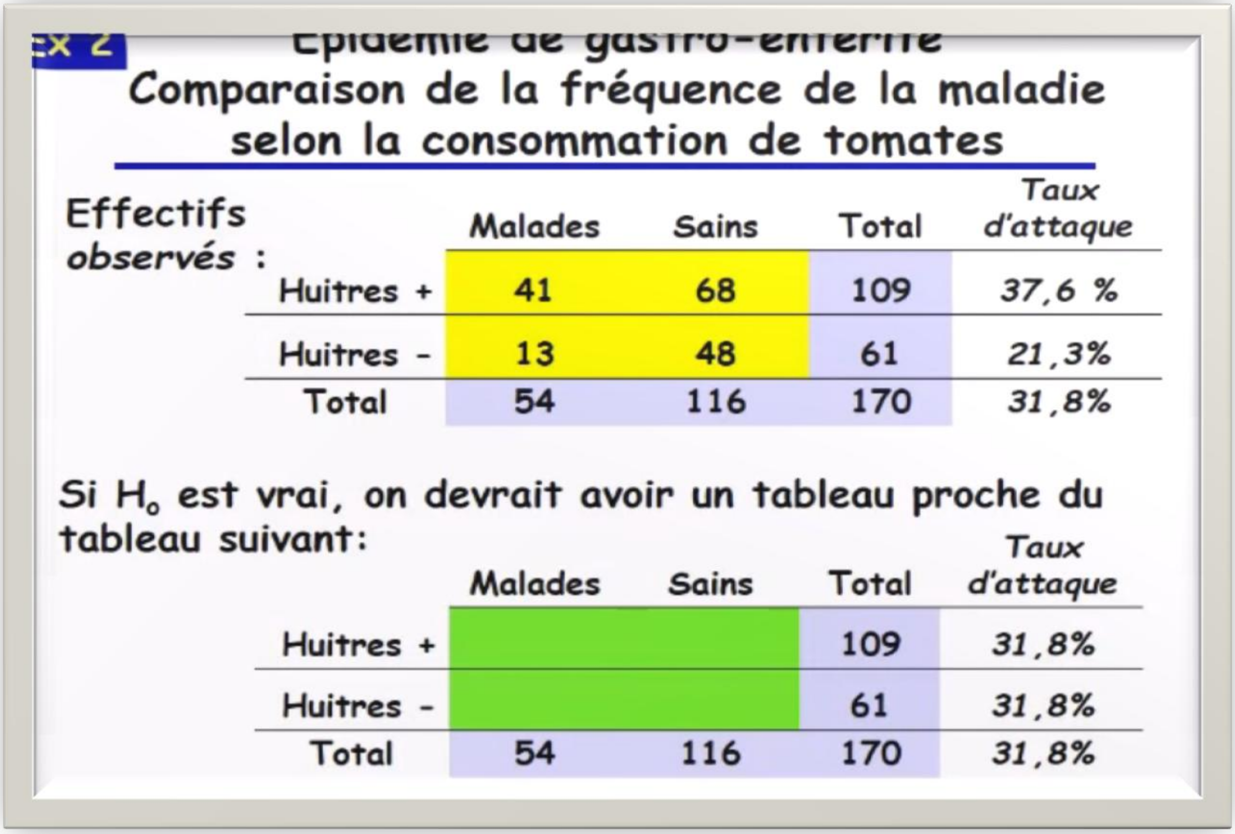
(test de chi2 de Pearson)

Epidémie de gastro-entérite

	Malades	Sains	Total	Taux d'attaque
Huitres +	41	68	109	*37,6 %*
Huitres -	13	48	61	*21,3%*
Total	54	116	170	*31,8%*

Si H_0 est vraie, il devrait y avoir, la même proportion de malades chez consommateurs d'huitres et les non-consommateurs

Si H_0 est vraie, il devrait y avoir
31,8 % de malades chez les 109 consommateurs d'huitres
31,8 % de malades chez les 61 non-consommateurs


Ex 2 Epidémie de gastro-entérite

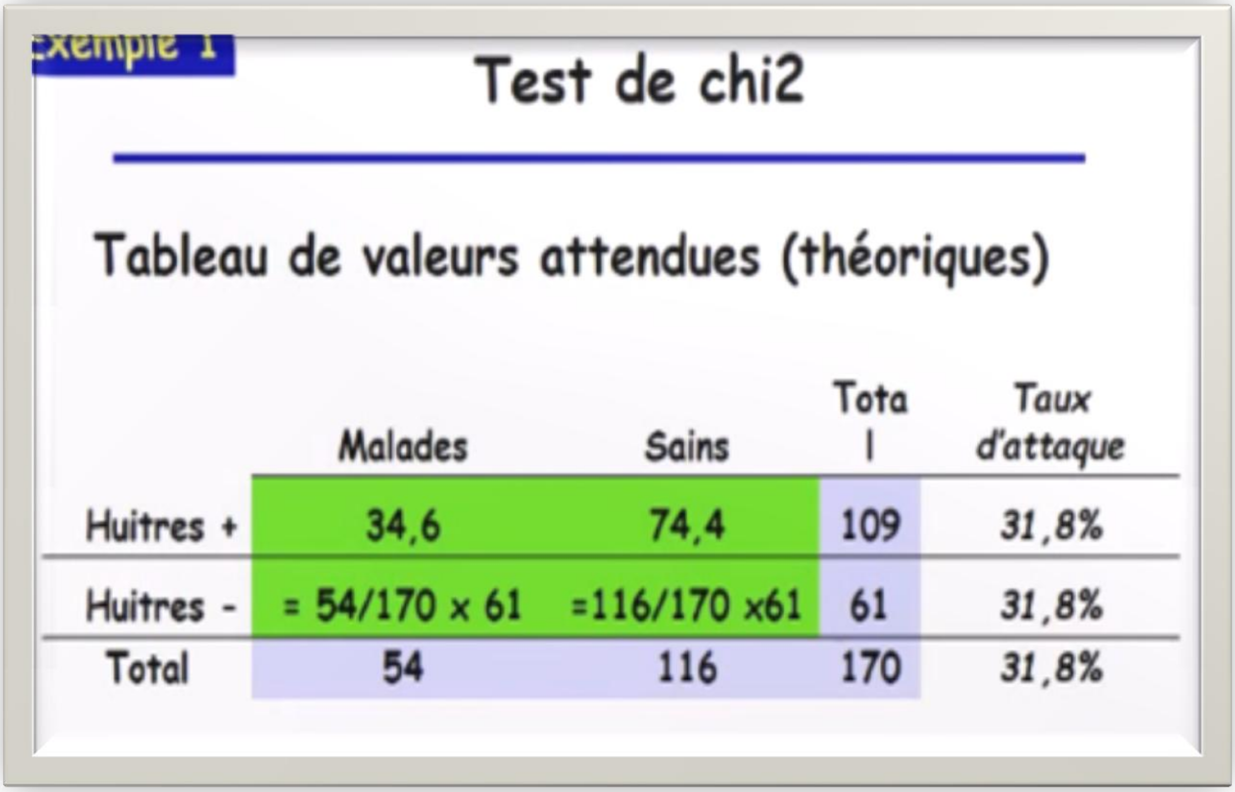
Comparaison de la fréquence de la maladie selon la consommation de tomates

Effectifs observés :

	Malades	Sains	Total	Taux d'attaque
Huitres +	41	68	109	37,6 %
Huitres -	13	48	61	21,3%
Total	54	116	170	31,8%

Si H_0 est vrai, on devrait avoir un tableau proche du tableau suivant:

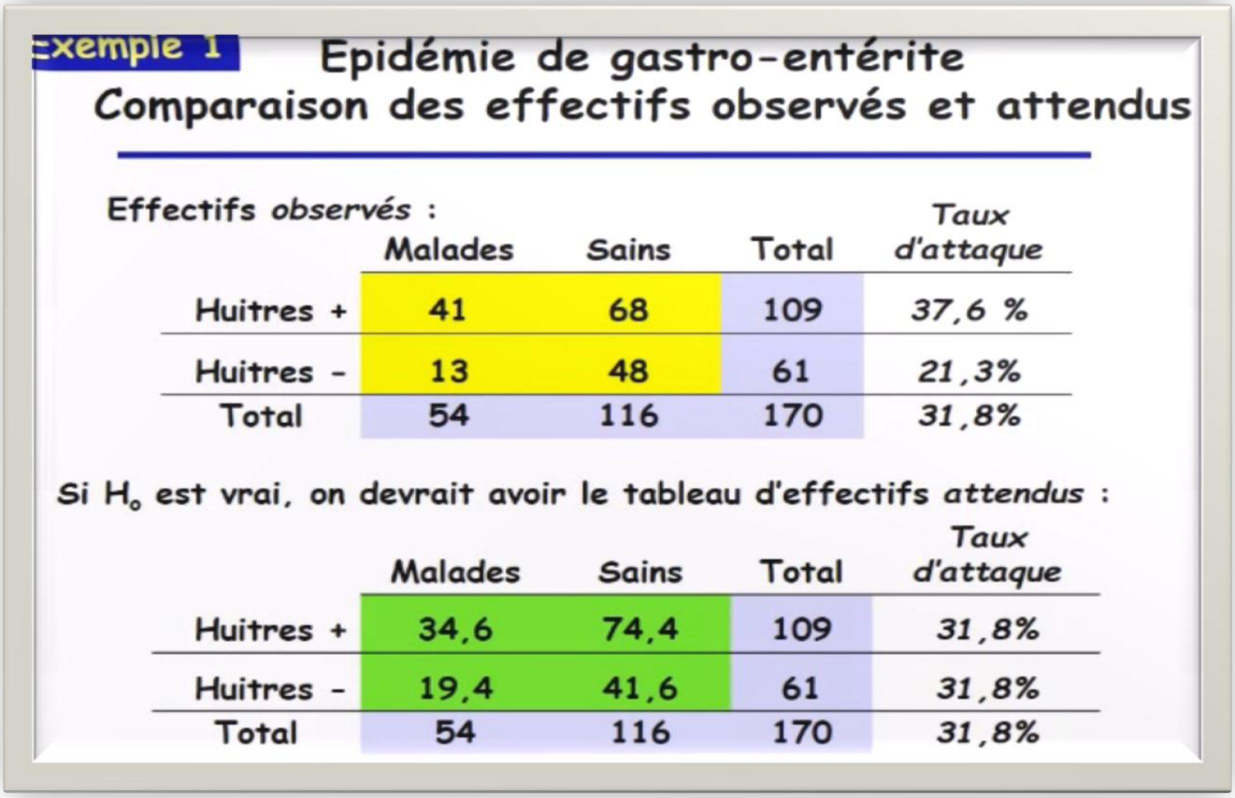
	Malades	Sains	Total	Taux d'attaque
Huitres +			109	31,8%
Huitres -			61	31,8%
Total	54	116	170	31,8%


Exemple 1

Test de chi2

Tableau de valeurs attendues (théoriques)

	Malades	Sains	Tota l	Taux d'attaque
Huitres +	34,6	74,4	109	31,8%
Huitres -	= 54/170 x 61	=116/170 x61	61	31,8%
Total	54	116	170	31,8%



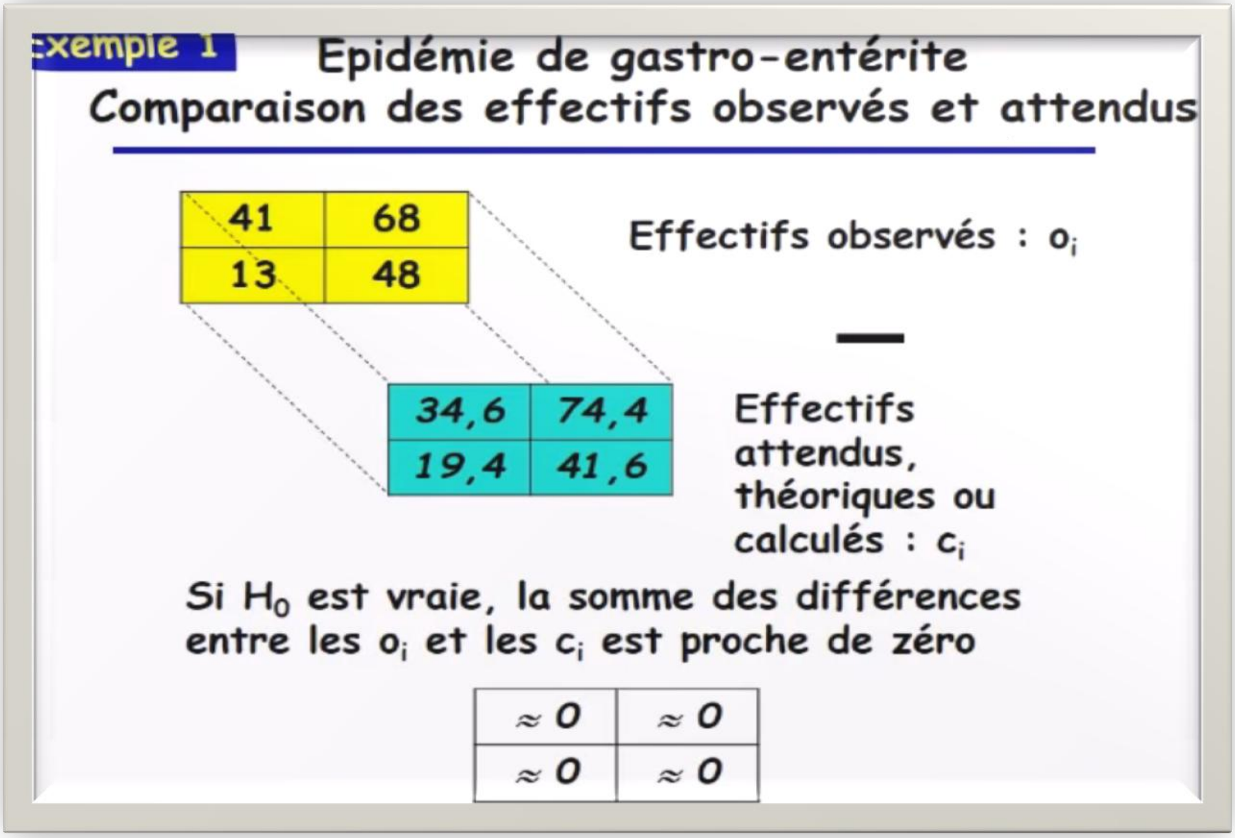
Exemple 1 — Epidémie de gastro-entérite
Comparaison des effectifs observés et attendus

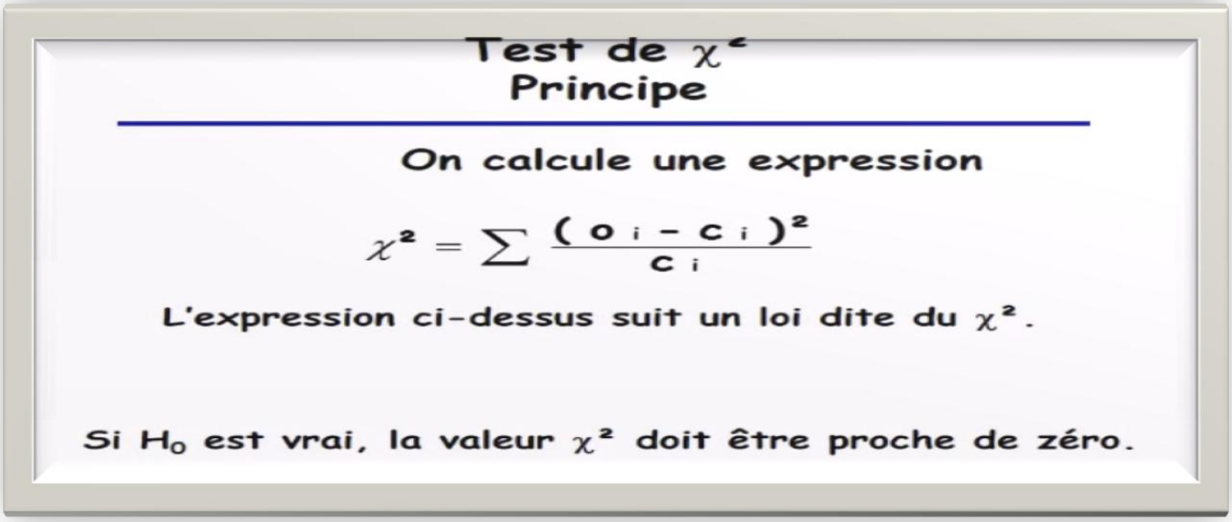
Effectifs *observés* :

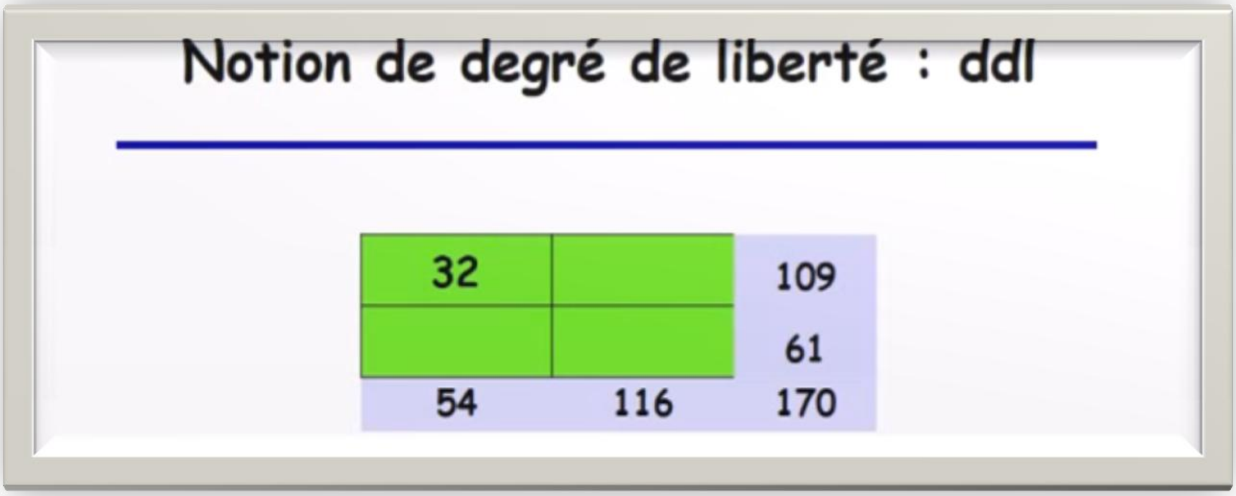
	Malades	Sains	Total	Taux d'attaque
Huitres +	41	68	109	37,6 %
Huitres -	13	48	61	21,3%
Total	54	116	170	31,8%

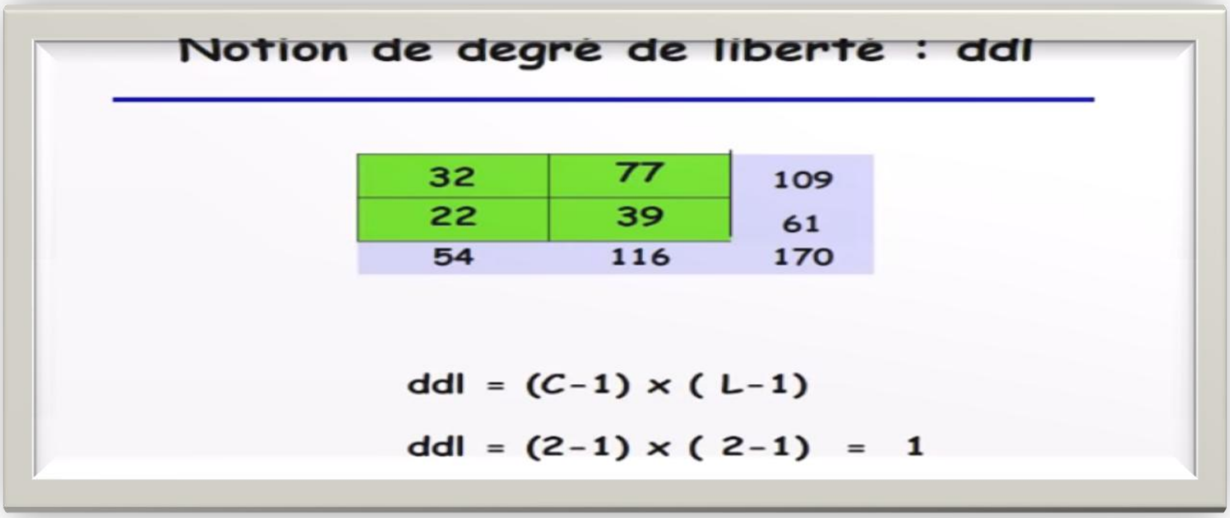
Si H_0 est vrai, on devrait avoir le tableau d'effectifs *attendus* :

	Malades	Sains	Total	Taux d'attaque
Huitres +	34,6	74,4	109	31,8%
Huitres -	19,4	41,6	61	31,8%
Total	54	116	170	31,8%

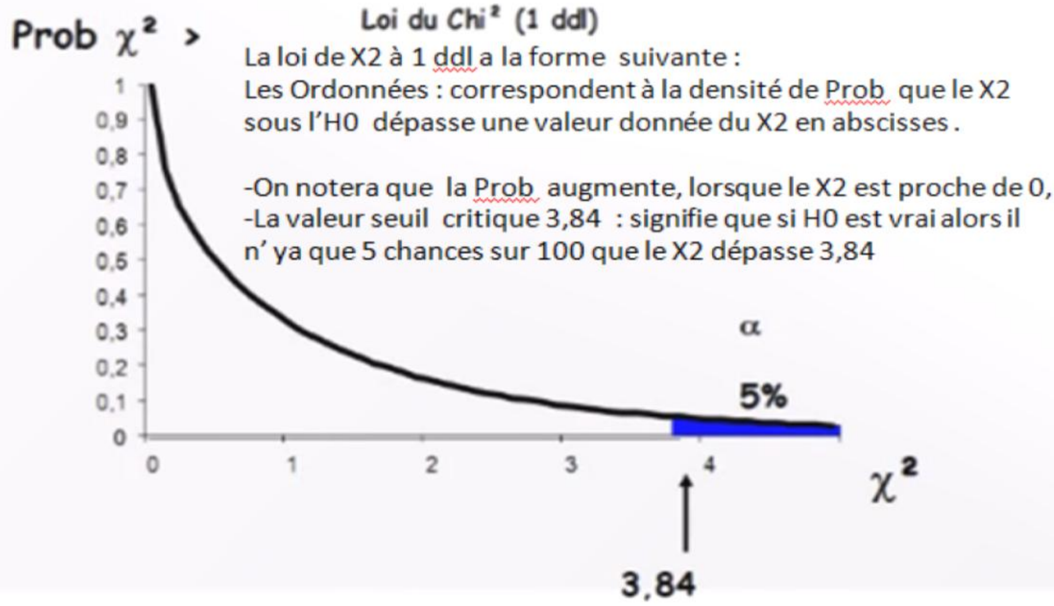



Test de χ^2
Principe
On calcule une expression
$\chi^2 = \sum \frac{(o_i - c_i)^2}{c_i}$
L'expression ci-dessus suit un loi dite du χ^2.
Si H_0 est vrai, la valeur χ^2 doit être proche de zéro.


Notion de degré de liberté : ddl
32
109
61
54
116
170


Notion de degré de liberté : ddl
32 77 109
22 39 61
54 116 170
ddl = (C-1) x (L-1)
ddl = (2-1) x (2-1) = 1

Loi du χ^2



Test de χ^2 à 1ddl
Principe

- Si la valeur observée χ^2_o est inférieure à 3,84 , on ne rejette pas H_0
- Si la valeur observée χ^2_o est supérieure à 3,84, on rejette H_0
- On regarde dans la table de χ^2, la valeur de *p* correspondant à la valeur observée χ^2_o

Exemple 1

Epidémie de gastro-entérite
Calcul du χ^2

o_i

41	68
13	48

c_i

34,6	74,4
19,4	41,6

$$\chi^2 = \sum \frac{(o_i - c_i)^2}{c_i}$$

$$\chi^2 = \frac{(41-34,6)^2}{34,6} + \frac{(68-74,4)^2}{74,4} + \frac{(13-19,4)^2}{19,4} + \frac{(48-41,6)^2}{41,6}$$

$\chi^2 = 4,8$ $\quad > 3,84$

Table de χ^2 , ddl = 1
(alpha bilatéral)

$\chi^2 = 4,8$

$\chi^2 >$	1,64	2,71	3,84	4,22	4,71	5,41	6,63
alpha	20%	10%	5%	4%	3%	2%	1%

$p < 3$ %

Exemple 1

Epidémie de gastro-entérite

	Malades	Sains	Total	Taux d'attaque
Huitres +	41	68	109	37,6 %
Huitres -	13	48	61	21,3%

Conclusion :
la différence entre les taux d'attaque
chez consommateurs et non-consommateurs d'huitres
est significative.

$p < 3\%$

Tests de χ^2 : condition d'application

les effectifs « théoriques » de chaque case

doivent être ≥ 5