

# **Relation entre variables catégorielles = analyse de fréquence**

Pierre-Yves Henry  
henry@mnhn.fr



# Quelle analyse ?

Y \ X	Var. catégorielle	Var. continue
	Var. catégorielle	Var. continue
Var. catégorielle	Analyse de fréquence	Faire des catégories  Transformer en variable quantitative
Var. continue	Comparaison de moyenne ANOVA	Corrélation Régression



# Test de Khi2

**Question: la proportion de jeunes adultes (2A) dans la population varie-t'elle entre deux sites ?**

	Site 1	Site 2	Total
Nb A (+2A)	45	15	60
Nb B (2A)	11	32	43
Total	56	47	103

# Test de Khi2

## Nb d'évènements observés

	Site 1	Site 2	Total
Nb A (+2A)	45	15	60
Nb B (2A)	11	32	43
Total	56	47	103

### Conditions d'application:

N attendu / catég. > 5  
N observé / catég. > 10  
N total > 1000

$H_0$  = l'âge-ratio est le même dans les deux sites

## Nb d'évènements attendus sous $H_0$

	Site 1	Site 2	Total		Site 1	Site 2
Nb A (+2A)	45	15	60	Nb A (+2A)	32.6	27.4
Nb B (2A)	11	32	43	Nb B (2A)	23.4	19.6
Total	56	47	103			

$$\frac{56 \times 60}{103}$$

**Statistique = somme des carrés des écarts entre valeurs attendues et valeurs observées divisées par les valeurs attendues**

Si  $H_0$  vrai,  $Khi2 = 0$

# Test de Khi2

**Nombre de degrés de libertés =**  
**(nb de valeurs de X – 1) x (nb de valeurs de Y – 1)**

2 classes d'âge dans 2 sites

ddl = 1

2 classes d'âge dans 4 sites

ddl = 3

**Probabilité de  $H_0$**

**=TEST.KHIDEUX(plage de valeurs observées; plages de valeurs attendues)**

I.KHIDEUX		=TEST.KHIDEUX(C5;D6;C11:D12)		
A	B	TEST.KHIDEUX(plage réelle; plage attendue)		
Test de Khi2 pour tester des différences de proportions				
Effectifs observés				
		Site 1	Site 2	Total
	Nb A (+2A)	45	15	60
	Nb B (2A)	11	32	43
	Total	56	47	103
Données attendus (sous l'hypothèse d'absence de différence)				
		Site 1	Site 2	
	Nb A (+2A)	32.6	27.4	
	Nb B (2A)	23.4	19.6	



# Test de Khi2

**Nombre de degrés de libertés =**  
**(nb de valeurs de X – 1) x (nb de valeurs de Y – 1)**

2 classes d'âge dans 2 sites

ddl = 1

2 classes d'âge dans 4 sites

ddl = 3

**Probabilité de  $H_0$**

=TEST.KHIDEUX(plage de valeurs observées;plages de  
valeurs attendues)

**Probabilité de  $H_0$**

=LOI.KHIDEUX(cellule avec statistique Khi2;cellule avec nb  
ddl)



# Test de Khi2

Et si les conditions d'application ne sont pas respectées ?

Si au moins un N attendu < 5  
(et N < 1000)

Utilisez le **test G** au lieu du test de Khi2 normal

Composants du test G (quand un effectif attendu < 5)									
Statistique de G =	13.308	NB: si pour une catégorie N = 0 -> le G-test ne peut pas être calculé							
Nb de ddl =	6								
Prob. =	0.038								

Si au moins un N observé < 10

Utilisez le **test du Khi2 avec correction de Yates**

Composants du test de Khi2 avec correction de Yates (quand un effectif < 10)									
Statistique de Khi2 corrigée =	11.055								
Nb de ddl =	6								
Prob. =	0.087								

*Une précaution pour éviter de se tromper: préférer le test G qui aura tendance à être plus "conservatif". Le Khi avec correction de Yates tend à être trop conservatif (n'indique que rarement une différence)*