

# **SYS865 Inférence statistique avec programmation R**

Ornwipa Thamsuwan

28 février 2024

# Test de l'indépendance des variables catégorielles

# Test de l'indépendance des variables catégorielles

Le **test du Chi-carré** est une méthode statistique utilisée pour déterminer s'il existe une association significative entre deux variables catégorielles.

Il compare les fréquences observées dans les catégories aux fréquences attendues s'il n'y avait aucune association ( $H_0$ ).

Calculée comme suit :

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- ▶  $O_{ij}$  : Fréquence observée dans la cellule (i, j)
- ▶  $E_{ij}$  : Fréquence attendue dans la cellule (i, j)

Calculée comme suit :

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- ▶  $O_{ij}$  : Fréquence observée dans la cellule (i, j)
- ▶  $E_{ij}$  : Fréquence attendue dans la cellule (i, j)

La **fréquence attendue** pour une cellule est

$$E_{ij} = \frac{(\textit{Total de la ligne}) \times (\textit{Total de la colonne})}{\textit{Total général}}$$

Cela suppose l'indépendance entre les variables.

# Valeur critique du test du Chi-carré

Déterminé par le degré de liberté :

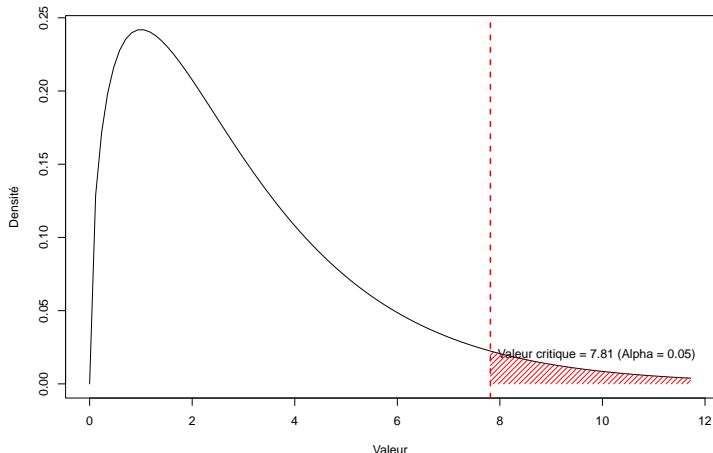
$$df = (\text{Nombre de lignes} - 1) \times (\text{Nombre de colonnes} - 1)$$

# Valeur critique du test du Chi-carré

Déterminé par le degré de liberté :

$$df = (\text{Nombre de lignes} - 1) \times (\text{Nombre de colonnes} - 1)$$

Distribution du Chi-carré, df = 3



Base de données sur R, HairEyeColor

- $H_0$  : Il n'y a pas d'association significative entre la couleur des cheveux et la couleur des yeux.



Base de données sur R, HairEyeColor

- $H_0$  : Il n'y a pas d'association significative entre la couleur des cheveux et la couleur des yeux.

## Tableau de contingence

```
knitr::kable(observe)
```

	Brown	Blue	Hazel	Green
Black	68	20	15	5
Brown	119	84	54	29
Red	26	17	14	14
Blond	7	94	10	16

## Fréquences attendues

```
total <- sum(observe)
attendu <- round(outer(margin.table(observe, 1),
                        margin.table(observe, 2),
                        FUN = "*") / total, 2)
knitr::kable(attendu)
```

	Brown	Blue	Hazel	Green
Black	40.14	39.22	16.97	11.68
Brown	106.28	103.87	44.93	30.92
Red	26.39	25.79	11.15	7.68
Blond	47.20	46.12	19.95	13.73

La fonction `margin.table` permet de calculer les sommes marginales pour le tableau de contingence observe :

- `margin.table(observe, 1)` calcule les sommes des lignes du tableau observe. Cela additionne les comptes sur toutes les colonnes pour chaque ligne, donnant les comptes totaux pour chaque niveau de la couleur des cheveux.

```
margin.table(observe, 1)
```

```
## Hair
## Black Brown    Red Blond
##   108   286    71   127
```

- `margin.table(observe, 2)` calcule les sommes des colonnes du tableau `observe`. Cela additionne les comptes sur toutes les lignes pour chaque colonne, donnant les comptes totaux pour chaque niveau de la couleur des yeux.

```
margin.table(observe, 2)
```

```
## Eye
## Brown  Blue Hazel Green
##   220   215    93    64
```

- ▶ `margin.table(observe, 2)` calcule les sommes des colonnes du tableau `observe`. Cela additionne les comptes sur toutes les lignes pour chaque colonne, donnant les comptes totaux pour chaque niveau de la couleur des yeux.

```
margin.table(observe, 2)
```

```
## Eye  
## Brown  Blue Hazel Green  
##    220    215     93    64
```

La fonction `outer(..., FUN = "*")` prend ces totaux de lignes et de colonnes et calcule le produit extérieur des deux vecteurs.

## Statistique du Chi-carré

```
chi_carre_stat <- sum((observe-attendu)^2 /attendu)  
round(chi_carre_stat, 3)
```

```
## [1] 138.293
```

## Statistique du Chi-carré

```
chi_carre_stat <- sum((observe-attendu)^2 /attendu)  
round(chi_carre_stat, 3)
```

```
## [1] 138.293
```

**Valeur critique** pour  $\alpha = 0.05$  et degrés de liberté

```
df <- (nrow(observe)-1) * (ncol(observe)-1)  
valeur_critique <- qchisq(0.95, df)  
round(valeur_critique, 3)
```

```
## [1] 16.919
```

## Conclusion

`chi_carre_stat > valeur_critique`, rejet  $H_0$ .

Il existe une association significative entre les variables.



## Conclusion

`chi_carre_stat > valeur_critique`, rejet  $H_0$ .

Il existe une association significative entre les variables.

Ou...

En appliquant directement la fonction R

```
chisq.test(observe)
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  observe  
## X-squared = 138.29, df = 9, p-value < 2.2e-16
```

## Résultats de recherche actuelle

Référer au code séparé et non pas encore publié . . .

- ▶ “Score” de Échelle d'Équilibre vs. Historique de Chute
- ▶ “Score” de Échelle d'Équilibre vs. Trouble de Vision