Université Cheikh Anta Diop de Dakar



Ecole Supérieure Polytechnique de DAKAR

Département Génie INFORMATIQUE

**Fouille de Données : Rapport de Projet**

Membres du groupe :

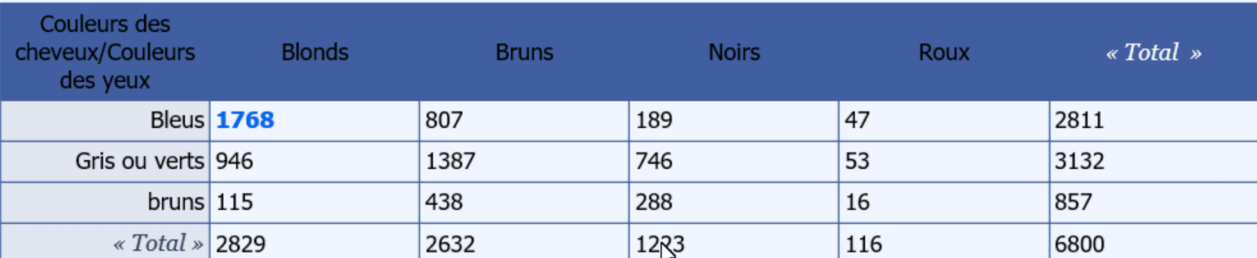
Saliou **NIANE**

Moussa **DIALLO** Professeur : Mr **CAMARA**

Mahomed **CISSOKHO**

Exercice 1 : **Test khi-Carre**

**Enonce** : Nous voulons savoir s'il existe une liaison, c’est-à-dire une relation statistique entre la couleur des cheveux et la couleur des yeux ou bien si ces deux caractères sont indépendants c'est-à-dire n'ont aucune relation ou lien entre eux.



**Hypotheses**

**H0 :** Les deux variables (couleur des cheveux et couleur des yeux) sont indépendantes.

**H1 :** Il existe une relation entre les deux variables.

**Le degre de signification** est fixe a 5% soit **0.05**

#### **Calculons les valeurs théoriques ou valeurs attendues : la couleur des cheveux et des yeux**

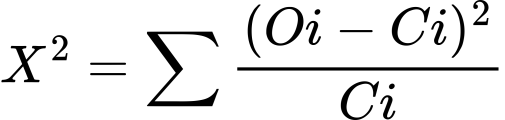
**Valeur theorique de chaque case =** (Total de sa ligne \* total de sa colonne) / N

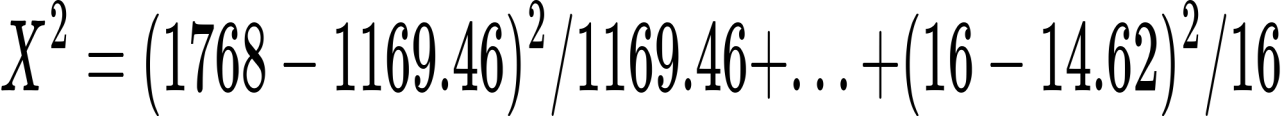
N : nombre d’echantillon

|  |  |
| --- | --- |
| **Valeurs Observees** | **Valeurs théoriques (attendues)** |
| 1768 | 1169.46 |
| 807 | 1088.02 |
| 189 | 505.57 |
| 47 | 47.95 |
| 946 | 1303 |
| 1387 | 1212.27 |
| 746 | 563.30 |
| 53 | 53.43 |
| 115 | 356.54 |
| 438 | 331.71 |
| 288 | 154.13 |
| 16 | 14.62 |

On remarque que toutes les valeurs theoriques sont superieures à 5 donc on peut bel et bien appliquer le test **khi-Carre.**

#### **Calculons la valeur de test du Khi-Carre donnee par la formule suivante :**

****

****

**X2** = 306.34 + 72.58 + 198.23 + 0.019 + 97.81 + 25.18 + 59.25 + 0.00346 + 163.63 + 163.63 + 34.45 + 116.27 + 0.13 = **1073,89**

**X2 = 1073,89**

* La valeur du khi-deux est de **1073,89**
* Nous allons maintenant calculer le degre de liberte combine a la valeur du degre de signification(5%) nous permettra de connaitre la valeur seuil .

#### **Calculons le degre de liberte ddl**

**ddl = (nb\_lignes - 1) (nb\_colonnes - 1)**

**AN : ddl =** (3-1)(4-1) = 6

* La **valeur** du degre de liberte est 6 .
* D’après la table du khi-deux, la statistique critique pour un **ddl de 6** et un degre de signification de **0.05** est de **12.592**
* **1073.89 >> 12.592**
* La statistique critique etant largement inferieure a la statisque obtenue, ce qui nous permet de rejeter l’hypothese nulle ( H0)
* Ce test demontre une forte correlation statistique entre la **couleur des yeux** et la **couleur des cheveux** d’ou l’on peut affirmer que **les deux variables sont etroitement liees (H1)**.