Test du Khi-deux (chi-deux)

# 1.Test d’ajustement du Khi-deux

Le test d’ajustement du correspond à la comparaison d’une distribution de fréquences observées et d’une distribution de fréquences théoriques. Le principe du test est d’estimer à partir d’une loi de probabilité connue ou inférée, les effectifs théoriques pour les différentes modalités du caractère étudié (caractère qualitatif ou quantitatif regroupé en classes) et les comparer aux effectifs observés dans un échantillon.

Problème de test :

Soit une variable aléatoire discrète ou discrétisée, c'est-à-dire divisée en classes de probabilités (calculée à partir d’une loi théorique donnée). Soit un -échantillon de cette variable fournissant les effectifs aléatoires dans chacune de ces classes. L’effectif théorique estimé pour la classe est .

Considérons la statistique (notée aussi ) définie comme suit

est une mesure (distance) de l’écart aléatoire entre les effectifs réalisés (observés) et les effectifs espérés (théoriques). Comme alors dépend de termes (le nombre de degrés de liberté de ).

**Théorème**

La statistique suit asymptotiquement une loi du Khi-deux à degrés de liberté où est le nombre de paramètres estimés.

On admet que si .

**D’où le test du : On rejettera pour les grandes valeurs de . La région critique sera donc de la forme telle que , c.à.d. .**

**Remarque :**

Dans le cas où, on distingue deux cas :

1. on procédera à un regroupement de classes (les classes à regrouper doivent être successives). On regroupera autant de classes que nécessaire pour avoir . Alors le degré de liberté sera avec le nombre de classes après regroupement.
2. : on ne fait pas de regroupement, mais on fait une correction de Yates

avec degré de liberté=1

**Exemple :**

* Pour , .
* Pour , .
* Pour ,

**Exemple : (test de conformité, sur les proportions)**

On sait que dans l’ensemble d’un pays donné, les pourcentages des 4 groupes sanguins s’établissent ainsi : 45% pour O, 35% pour A, 16% pour B et 4% pour AB. Un échantillon de 100 individus est prélevé au hasard, dans une zone montagneuse, 35 sont du groupe O, 35 de A, 20 de B et 10 de AB. Peut-on penser qu’il ya conformité entre ces résultats et ceux établis pour l’ensemble du pays au risque de 5% ?

: groupe sanguin ; on a 4 groupes ,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classe (groupe sanguin) |  |  |  |  |
| O | 35 | 0.45 | **45>5** |  |
| A | 35 | 0.35 | **35>5** |  |
| B | 20 | 0.16 | **16>5** | regrouper |
| AB | 10 | 0.04 | **4<5** |  |
|  |  |  |  |  |

On procède à un regroupement (B+AB) (on obtient 3 classes)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classe (groupe sanguin) |  |  |  |  |
| O | 35 | 0.45 | 45>5 | 2.22 |
| A | 35 | 0.35 | 35>5 | 0 |
| B+AB | 30 | 0.20 | 20>5 | 5 |
|  |  |  |  | 7.22 |

La valeur de la statistique du Khi-deux est

Sous , La statistique du Khi-deux suit une loi .

La région critique est de la forme : .

Pour on a .

Comme alors on rejette (l’échantillon n’est pas représentatif)

**Exemple : Ajustement à une loi binomiale**

1. Est-ce que la distribution du nombre de filles observées dans 320 fratries de 5 enfants suit une loi binomiale de paramètres  ?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nombre de filles | 0 | 1 | 2 | 3 | 4 | 5 |
| Nombre de fratries observé | 18 | 56 | 110 | 88 | 40 | 8 |

Soit la variable aléatoire : Nombre de fille dans une fratrie de 5 enfants. On veut tester v.s. .

Sous , .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre de filles |  |  |  |  |
| 0 | 18 | 0.03125 | 10 | 6.4 |
| 1 | 56 | 0.15625 | 50 | 0.72 |
| 2 | 110 | 0.3125 | 100 | 1 |
| 3 | 88 | 0.3125 | 100 | 1.44 |
| 4 | 40 | 0.15625 | 50 | 2 |
| 5 | 8 | 0.03125 | 10 | 0.4 |
|  |  |  |  | **11.96** |

La valeur de la statistique du  :

On a , alors suit, sous , une Khi-deux à 5 degrés de liberté.

Région critique : .

Pour .

Comme alors on rejette c'est-à-dire le nombre de filles observées dans les fratries de 5 enfants ne suit pas une loi .

1. Comparer, au risque de 5%, la distribution du nombre de filles dans une fratrie de 5 enfants à celle d’une loi binomiale.

On veut tester v.s. .

Comme est inconnu, alors on l’estime par .

Pour calculer les probabilités théoriques, on utilise une , c.à.d.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre de filles |  |  |  |  |
| 0 | 18 | 0.04486342 | 14.35629 | **0.924** |
| 1 | 56 | 0.193017 | 61.76544 | **0.538** |
| 2 | 110 | 0.3321688 | 106.294 | **0.129** |
| 3 | 88 | 0.2858197 | 91.4623 | **0.131** |
| 4 | 40 | 0.1229689 | 39.35005 | **0.01** |
| 5 | 8 | 0.0211621 | 6.771872 | **0.222** |
|  |  |  |  | **1.954** |

La valeur de la statistique du  :

On a , alors suit, sous , une Khi-deux à 4 degrés de liberté

Région critique : .

Pour .

Comme alors on accepte .

**Exemple : (ajustement à une loi de Poisson)**

A partir de la vente de 100 postes de télévision ayant fonctionné le même nombre d’heures, pendant une année, on a pu établir le tableau suivant contenant le nombre de répartition de ces téléviseurs :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre de réparations ( |  |  |  | 3 |
| Nombre de téléviseurs ( | 61 | 30 | 7 | 2 |

1. Peut-on ajuster la distribution observée à une loi de Poisson ?

nombre de réparations

Tester v.s. ne suit pas la loi de Poisson

étant inconnu, on l’estime par

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 61 | 0.606 | 60.6 |  |
|  | 30 | 0.303 | 30.3 |  |
|  | 7 | 0.075 | 7.5 |  |
|  | 2 | 0.012 | 1.2<5  regrouper |  |
|  | 0 | 0.0017 | 0.17<5 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 61 | 0.606 | 60.6 | 0.0026 |
|  | 30 | 0.303 | 30.3 | 0.0029 |
|  | 9 | 0.091 | 9.1 | 0.001 |
|  |  |  |  |  |

La statistique du test suit sous une .

Comme alors on accepte .

1. Peut-on l’ajuster à la loi de Poisson de paramètre ?

**Exemple : (Ajustement à une loi normale)**

On a mesuré la taille en cm de 200 étudiants et on a obtenu les résultats suivants :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Taille |  |  |  |  |  |  |
| Effectif | 16 | 44 | 60 | 44 | 26 | 10 |

Peut-on ajuster cette répartition observée à la loi Normale , au risque de 10% ?

: taille d’un étudiant, .

On veut tester v.s. ne suit pas la loi normale.

Comme sont inconnus, on les estime respectivement par

et

où est le centre de la ième classe.

Sous suit une loi normale centrée réduite et

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classe |  |  |  |  |
|  | 0 |  |  |  |
|  | 16 |  |  |  |
|  | 44 |  |  |  |
|  | 60 |  |  |  |
|  | 44 |  |  |  |
|  | 26 |  |  |  |
|  | 10 |  |  |  |
|  | 0 |  |  |  |
|  |  |  |  |  |