Zied GHARBI

Université Lumière Lyon2

TD 6 : Loi usuelles continues
Licence 1 MIASHS

Rappel test d'indépendance (basé sur la loi kh-deux):

Le test du chi-carré d'indépendance est un test statistique utilisé pour déterminer si deux variables catégorielles sont indépendantes l'une de l'autre. Il est basé sur les fréquences observées dans une table de contingence et les fréquences attendues si les variables étaient indépendantes.

1. Formulation des Hypothèses

Les hypothèses à tester sont les suivantes :

- **Hypothèse nulle (H_0)** : Les deux variables sont indépendantes.
- **Hypothèse alternative (H_1)** : Les deux variables ne sont pas indépendantes.

2. Calcul de la Statistique du Test

2.1 Fréquences Observées (O)
Les fréquences observées O_{ij} sont les valeurs réelles que vous obtenez dans votre échantillon et qui sont placées dans chaque cellule de la table de contingence.

2.2 Fréquences Attendues (E)
Les fréquences attendues E_{ij} sont les valeurs que l'on s'attend à observer dans chaque cellule si les deux variables étaient indépendantes. Elles sont calculées à l'aide de la formule suivante :

$$E_{ij} = \frac{(Total\,ligne_i) \times (Total\,colonne_j)}{Total\,gnral}$$

Où :

- $Total\,ligne_i$ est le total des éléments dans la ligne i,
- $Total\,colonne_j$ est le total des éléments dans la colonne j,
- $Total\,gnral$ est le nombre total d'observations.

3. Statistique χ^2
La statistique du test du chi-carré χ^2 est calculée en utilisant la formule suivante :

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

Où :

- O_{ij} est la fréquence observée dans la cellule ij,
- E_{ij} est la fréquence attendue dans la cellule ij.

Le calcul est effectué pour chaque cellule, puis les résultats sont additionnés pour obtenir la statistique χ^2 totale.

4. Degrés de Liberté
Le nombre de degrés de liberté pour un test du chi-carré d'indépendance est donné par :

$$df = (r-1) \times (c-1)$$

Où :

- r est le nombre de lignes de la table de contingence,
- c est le nombre de colonnes de la table de contingence.

5. Critères de Décision

Une fois que vous avez calculé la statistique χ^2, vous la comparez à la **valeur critique** du chi-carré, qui dépend des degrés de liberté et du niveau de signification α choisi.

- Si χ^2 calculé est plus grand que la valeur critique, vous rejetez l'hypothèse nulle.
- Si χ^2 calculé est plus petit que la valeur critique, vous ne rejetez pas l'hypothèse nulle.

La valeur critique du chi-carré peut être obtenue dans une table de chi-carré pour les degrés de liberté calculés et le niveau de signification α.

Tableau de Chi-Carré Critique

$$\text{Exemple : pour } df = 2,\ \alpha = 0.05, \text{ la valeur critique} = 5.991$$

Remarques Importantes

- Les variables doivent être **catégorielles**.
- La taille de l'échantillon doit être suffisamment grande, et chaque fréquence attendue doit être supérieure ou égale à 5. Si certaines cellules ont des fréquences attendues inférieures à 5, le test du chi-carré peut ne pas être approprié. Dans ce cas, il est recommandé de regrouper les catégories ou d'utiliser un autre test comme le test exact de Fisher.

Exercice 1 : application d'un test de dépendance avec la loi χ^2

Une entreprise de marketing souhaite savoir si le choix d'un produit (Produit A, Produit B) est indépendant de l'âge des consommateurs. Un échantillon de 200 personnes a été interrogé, et leurs réponses ont été réparties selon les catégories d'âge et les produits choisis. Voici les résultats obtenus :

Âge \Produit	Produit A	Produit B	Total
Moins de 30	50	30	80
30-50 ans	40	30	70
Plus de 50	20	30	50
Total	110	90	200

On veut tester l'hypothèse suivante :

- **H0** : Il n'y a pas de relation entre le choix du produit et l'âge des consommateurs (les deux variables sont indépendantes).
- **H1** : Il y a une relation entre le choix du produit et l'âge des consommateurs (les deux variables sont dépendantes).

1. Calcul des effectifs attendus

Les effectifs attendus sont calculés sous l'hypothèse d'indépendance des variables. Pour chaque cellule du tableau, l'effectif attendu E_{ij} est donné par la formule suivante :

$$E_{ij} = \frac{(\text{Total ligne i}) \times (\text{Total colonne j})}{\text{Total général}}$$

Nous allons maintenant calculer les effectifs attendus pour chaque cellule.

$$E_{11} = \frac{80 \times 110}{200} = 44 \quad (\text{Produit A, Moins de 30 ans})$$

$$E_{12} = \frac{80 \times 90}{200} = 36 \quad \text{(Produit B, Moins de 30 ans)}$$

$$E_{21} = \frac{70 \times 110}{200} = 38.5 \quad \text{(Produit A, 30-50 ans)}$$

$$E_{22} = \frac{70 \times 90}{200} = 31.5 \quad \text{(Produit B, 30-50 ans)}$$

$$E_{31} = \frac{50 \times 110}{200} = 27.5 \quad \text{(Produit A, Plus de 50 ans)}$$

$$E_{32} = \frac{50 \times 90}{200} = 22.5 \quad \text{(Produit B, Plus de 50 ans)}$$

Les effectifs attendus sont donc :

Âge \Produit	Produit A	Produit B	Total
Moins de 30	44	36	80
30-50 ans	38.5	31.5	70
Plus de 50	27.5	22.5	50
Total	110	90	200

2. Calcul du Khi-deux observé

Le test du Khi-deux est calculé avec la formule suivante :

$$\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

où O_{ij} est l'effectif observé dans chaque cellule et E_{ij} est l'effectif attendu. Nous allons maintenant calculer la statistique de test.

$$\chi^2 = \frac{(50-44)^2}{44} + \frac{(30-36)^2}{36} + \frac{(40-38.5)^2}{38.5} + \frac{(30-31.5)^2}{31.5} + \frac{(20-27.5)^2}{27.5} + \frac{(30-22.5)^2}{22.5}$$

Calculons chaque terme :

$$\chi^2 = \frac{6^2}{44} + \frac{(-6)^2}{36} + \frac{(1.5)^2}{38.5} + \frac{(-1.5)^2}{31.5} + \frac{(-7.5)^2}{27.5} + \frac{(7.5)^2}{22.5}$$

$$\chi^2 = \frac{36}{44} + \frac{36}{36} + \frac{2.25}{38.5} + \frac{2.25}{31.5} + \frac{56.25}{27.5} + \frac{56.25}{22.5}$$

$$\chi^2 \approx 0.818 + 1 + 0.058 + 0.071 + 2.045 + 2.5$$

$$\chi^2 \approx 6.492$$

3. Degrés de liberté et valeur critique

Les degrés de liberté (df) sont donnés par la formule :

$$\text{Degrés de liberté} = (n_{\text{lignes}} - 1) \times (n_{\text{colonnes}} - 1)$$

Dans ce cas, nous avons $n_{\text{lignes}} = 3$ et $n_{\text{colonnes}} = 2$, donc les degrés de liberté sont :

$$\text{df} = (3-1) \times (2-1) = 2$$

On consulte une table de Khi-deux pour déterminer la valeur critique à 2 degrés de liberté et un seuil de signification de 0,05. La valeur critique pour un seuil de 0,05 et 2 degrés de liberté est :

$$\chi^2_{\text{critique}} = 5.991$$

4. Conclusion

La valeur observée de χ^2 est 6.492, ce qui est supérieur à la valeur critique de 5.991. Par conséquent, nous rejetons l'hypothèse nulle H_0.

$\hookrightarrow$ Il existe une relation significative entre le choix du produit et l'âge des consommateurs. Les deux variables ne sont pas indépendantes.

Rappel test Student pour une moyenne (Test t de Student):

Le test de Student est un test statistique utilisé pour comparer la moyenne d'un échantillon à une moyenne hypothétique ou pour comparer les moyennes de deux échantillons afin de déterminer s'il existe une différence statistiquement significative entre elles. Le test le plus couramment utilisé est le test t de Student pour un échantillon.

1. Hypothèses

- **Hypothèse nulle** (H_0) : La moyenne de la population est égale à une valeur hypothétique μ_0,

$$H_0 : \quad \mu = \mu_0$$

.

- **Hypothèse alternative** (H_1) : La moyenne de la population est différente de μ_0 (test bilatéral).

$$H_1 = \mu \neq \mu_0$$

2. Calcul de la Statistique t

La formule pour la statistique t est la suivante :

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Où :

- $\bar{x}$ est la moyenne de l'échantillon.
- μ_0 est la moyenne hypothétique de la population.
- s est l'écart-type de l'échantillon.
- n est la taille de l'échantillon.

3. Degrés de liberté(df)

$$df = n - 1$$

avec n est la taille de l'échantillon.

4 Décision

Une fois la statistique t calculée, il faut la comparer à la valeur critique de t pour un certain niveau de signification α (souvent 0.05) et avec df degrés de liberté.

- Si $|t| > t_{\alpha/2,df}$ (test bilatéral), vous rejetez l'hypothèse nulle H_0.
- Si $|t| \leq t_{\alpha/2,df}$ (test bilatéral), vous ne rejetez pas l'hypothèse nulle H_0.

Exercice 2 : application d'un test de Student

Un chercheur en éducation souhaite savoir si la moyenne des scores obtenus par un groupe d'élèves dans un test de mathématiques est différente de la moyenne nationale, qui est de 75 points. A disposition un échantillon de 40 élèves avec une moyenne de 78 points et un écart-type de 10 points.

Objectif :

Effectuer un test t bilatéral pour vérifier si la moyenne des scores de l'échantillon est significativement différente de la moyenne nationale de 75 points, avec un niveau de signification $\alpha = 0.05$.

Supposons que nous soyons intéressé par la moyenne des scores obtenus par un groupe d'élèves dans un test de mathématiques. La moyenne nationale des scores est de 75 points. Nous avons un échantillon de 40 élèves, dont la moyenne des scores est de 78 points et l'écart-type est de 10 points. Nous voulons vérifier si la moyenne des scores de cet échantillon est significativement différente de la moyenne nationale de 75 points.

1. Hypothèses

- **Hypothèse nulle (H_1)** : La moyenne des scores de l'échantillon est égale à la moyenne nationale, soit $\mu = 75$.
- **Hypothèse alternative (H_1)** : La moyenne des scores de l'échantillon est différente de la moyenne nationale, soit $\mu \neq 75$ (test bilatéral).

2. Données de l'échantillon

- Moyenne de l'échantillon : $\bar{x} = 78$
- Moyenne hypothétique de la population : $\mu_0 = 75$
- Écart-type de l'échantillon : $s = 10$
- Taille de l'échantillon : $n = 40$
- Niveau de signification : $\alpha = 0.05$

3. Formule de la Statistique t

La statistique t est calculée par la formule suivante :

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Où :

- $\bar{x}$ est la moyenne de l'échantillon,
- μ_0 est la moyenne hypothétique de la population,
- s est l'écart-type de l'échantillon,
- n est la taille de l'échantillon.

4. Calcul de la statistique t

Substituons les valeurs dans la formule de t :

$$t = \frac{78 - 75}{\frac{10}{\sqrt{40}}} = \frac{3}{\frac{10}{6.324}} = \frac{3}{1.581} \approx 1.90$$

5. Degrés de liberté

Les degrés de liberté sont calculés par :

$$df = n - 1 = 40 - 1 = 39$$

6. Valeur Critique

Pour un test bilatéral avec un niveau de signification $\alpha = 0.05$ et $df = 39$, la valeur critique de $t_{\alpha/2,df}$ (en consultant une table de Student) est environ 2.022.

7. Décision

La valeur calculée de t est 1.90, qui est inférieure à la valeur critique de 2.022. Par conséquent, nous ne rejetons pas l'hypothèse nulle H_0.

8. Conclusion

Nous ne pouvons pas conclure qu'il y a une différence significative entre la moyenne des scores de l'échantillon (78) et la moyenne nationale (75) au niveau de signification $\alpha = 0.05$.