

TP 2 echantillon

2024-11-19

Instructions: Un compte-rendu du TP rédigé sous la forme d'un fichier pdf est à rendre par binôme d'étudiants. Vous devrez déposer votre compte rendu sous moodle. Chaque binôme remettra un document pdf ayant pour nom *nomsetudiants_tptest.pdf* avant le 17 décembre.

Exercice 1

L'objectif de cet exercice est d'utiliser des tests statistiques pour comparer différents aliments en fonction d'indicateurs relatifs aux impacts environnementaux de leur production agricole.

Nous allons travailler sous R avec des données extraites de la base de données Agribalyse®. Agribalyse® est une base de données publique française qui fournit des informations sur les impacts environnementaux des produits agricoles et alimentaires. Pour en savoir plus, vous pouvez télécharger la documentation à l'adresse suivante : <https://doc.agribalyse.fr/documentation/acces-donnees>.

Le fichier *Agri_conv_TP.xlsx* sur lequel nous allons travailler dans ce TP a été réalisé à partir des données extraites du tableur pour les produits agricoles bruts conventionnels (à la sortie de la ferme) de la base de données Agribalyse 3.1. Vous trouverez plus d'informations sur la modification des noms des variables et la préparation des données dans l'onglet "Info" du fichier *Agri_conv_TP.xlsx*. Ce fichier a été élaboré par Julie Charles lors de son stage de BUT SD, en deuxième année.

Nous disposons ici de valeurs de consommation en CO₂ (ou équivalent) par kg pour 258 aliments. De plus, pour chacun de ces aliments, nous avons les informations suivantes :

- Catégorie : S'agit-il d'un produit d'origine animale ou végétale ?
- Groupe : Quel est le sous-groupe du produit (par exemple, pour les produits animaux : bœuf, mouton, etc.) ?

0. Charger les données *Agri_conv_TP.csv*.

1. On s'intéresse au *Co2_eq*. Faire une étude représentation univariée de cette variable.

2. On s'intéresse ici au lien entre la catégorie de produit (animale ou végétale) et le *Co2*.

a. Etudier de manière descriptive le lien entre la variable catégorie et le *Co2* (avec un graphique pertinent)

b. On veut maintenant tester si la moyenne de *Co2* émis est la même pour les aliments animaux et les aliments végétaux. Préciser dans quelle situation vous vous trouvez (taille de l'échantillon, normalité égalité des variances, ...). Puis faites le test et concluez.

c. Si on voulait montrer que les produits animaux ont une espérance plus élevée que les produits végétaux que faudrait il changer (refaites la question précédente dans ce sens). Faites une conclusion.

3. On veut mettre en évidence au risque 5% que la proportion d'aliment végétal qui nécessitent plus d'un kilo d'équivalent CO₂ pour un kilo de production est inférieur chez les végétaux que chez les animaux. Faites le test et donner la conclusion adéquate.

Exercice 2 : l'amour ouf fait le buzz.

L'amour Ouif a été décrit comme très populaire auprès des 15-25 ans. Afin d'étudier ce phénomène, l'Institut Vertigo (un institut d'études spécialisé dans la recherche marketing pour le marché des loisirs) a mené une enquête sur l'âge des spectateurs à la sortie de certains films en 2024. Les résultats obtenus sont les suivants :

```
data.frame(Film = c( "Amour Ouf", "Dead_pool","Dune2","Monte_Cristo", "Emilia_Perez"), Age_15_25= c(0.2
```

```
##           Film Age_15_25 Age_Autres
## 1    Amour Ouf      580      1420
## 2    Dead_pool      540      1460
## 3      Dune2       500      1500
## 4 Monte_Cristo      400      1600
## 5 Emilia_Perez      220      1780
```

1. Au risque de 5%, pouvez-vous dire que la proportion de 15-25 ans allant voir un film varie en fonction du film ?
2. Au risque de 5%, pouvez-vous dire que la proportion de spectateurs âgés de 15-25 ans ayant vu “Amour Ouf” est significativement supérieure à celle des autres films étudiés (tous films confondus) ?
3. En 2024, 20% des spectateurs de cinéma ont entre 15 et 25 ans.
 - a. Si 20% des spectateurs de “Amour Ouf” avaient eu entre 15 et 25 ans, combien aurait-on observé théoriquement de personnes de cette tranche d’âge parmi les spectateurs interrogés ayant vu ce film ?
 - b. Proposez un test permettant de comparer les effectifs observés dans la salle de “Amour Ouf” à l’effectif théorique si 20% des spectateurs avaient eu entre 15 et 25 ans (test du chi carré d’adéquation à une loi).