Labo 4: Paramètres de tendance centrale

Visseho Adjiwanou, PhD.

01 February 2022

# PARTIE A

Voici les résultats obtenus au cours d’une enquête sur l’âge et le statut matrimoniale des répondants.

**Tableau 1: Distribution du statut matrimonial**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Statut | Fréquence | Pourcentage | Pourcentage valide | Pourcentage cumulé |
| Marié | 247 | 29.1 | 29.1 | 29.1 |
| Veuf | 3 | .4 | .4 | 29.4 |
| Divorcé | 36 | 4.2 | 4.2 | 33.6 |
| Séparé | 14 | 1.6 | 1.6 | 35.3 |
| Jamais marié | 550 | 64.7 | 64.7 | 100.0 |
| **Total** | **850** | **100.0** | **100.0** |  |

Répondez aux questions suivantes:

1. Quel est le type de la variable étudiée?
2. Quel est la valeur du mode
3. Si vous pouvez utiliser la médiane, indiquez sa valeur. Sinon, dites que ce n’est pas possible et expliquer votre réponse.
4. Si vous pouvez utiliser la moyenne, indiquez sa valeur. Sinon, dites que ce n’est pas possible et expliquer votre réponse.
5. Quel est le problème avec ce tableau? Quelle solution préconisez-vous?

Au cours de la même enquête, on a collecté les données sur le groupe d’âges des enquêtés. les résultats sont présentés dans le tableau 2.

**Tableau 2: Distribution du groupe d’âges**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Groupe d’âge | Fréquence | Pourcentage | Pourcentage valide | Pourcentage cumulé |
| 15 - 24 | 276 |  |  |  |
| 25 - 34 | 199 |  |  |  |
| 35 - 49 | 263 |  |  |  |
| 50 et plus | 77 |  |  |  |
| Non réponse | 35 |  |  |  |
| **Total** | **850** |  |  |  |

1. Quel est le type de la variable étudiée?
2. Complétez le tableau
3. Quel est la valeur du mode?
4. Si vous pouvez utiliser la médiane, indiquez sa valeur. Sinon, dites que ce n’est pas possible et expliquer votre réponse.
5. Si vous pouvez utiliser la moyenne, indiquez sa valeur. Sinon, dites que ce n’est pas possible et expliquer votre réponse.

## Question 2

Voici les données issues d’une enquête dans une classe

|  |  |
| --- | --- |
| Age | Nombre d’élèves |
| 10 | 5 |
| 11 | 7 |
| 12 | 4 |

1. Quelle est la variable étudiée?
2. Quelle est la valeur de l’âge moyen de la classe?
3. Quelle est la valeur de l’âge modal de la classe?
4. Quelle est la valeur de l’âge médian de la classe?

# PARTIE B

# La solution technologique au changement climatique (exemple tiré de Krieg)

Beaucoup de gens pensent qu’en adoptant de nouvelles technologies, nous pouvons économiser à la fois de l’argent et protéger l’environnement en brûlant moins de combustibles fossiles. Cet exercice est tiré du livre de krieg, “Statistics and data analysis for Social Science”.

## 1. Que pensez-vous de cette assertion?

## 2. En quoi n’est-elle pas valide?

Pour tester cette assertion, nous utilisons les données de 1994 et de 2009 sur les voitures les plus efficients entre les deux périodes. Le tableau suivant présente les vitesses (mile per gallon, mpg) pour les différentes marques de voitures pour leur circulation en ville et sur l’autoroute:

* **Pour 1994**

|  |  |  |
| --- | --- | --- |
| Marque et modèle | Ville (mpg) | Autoroute(mpg) |
| Mazda 626 | 23 | 31 |
| Honda Accord | 22 | 29 |
| Chevrolet Corsica | 22 | 28 |
| Buick Century | 22 | 28 |
| Oldsmobile Cutlass Ciera | 22 | 28 |
| Oldsmobile Achieva | 21 | 32 |
| Pontiac Grand Am | 21 | 32 |
| Infiniti G20 | 21 | 29 |
| Mitsubishi Galant | 21 | 28 |
| Dodge Spirit | 21 | 27 |
| Plymouth Acclaim | 21 | 27 |
| Subaru Legacy | 20 | 28 |
| Toyota Camry | 20 | 27 |
| Hyundai Sonata | 19 | 26 |
| Chrysler LeBaron | 19 | 25 |
| Ford Taurus | 18 | 27 |
| Mercury Sable | 18 | 27 |
| Eagle Vision | 18 | 26 |

* **Pour 2009**

|  |  |  |
| --- | --- | --- |
| Marque et modèle | Ville (mpg) | Autoroute(mpg) |
| Toyota Prius Hybrid) | 48 | 45 |
| Nissan Altima (hybrid) | 35 | 33 |
| Toyota Camry (hybrid) | 33 | 34 |
| Chevrolet Malibu (hybrid) | 26 | 34 |
| Saturn Aura (hybrid) | 26 | 34 |
| Hyundai Elantra | 25 | 33 |
| Kia Spectra | 24 | 32 |
| Nissan Altima | 23 | 32 |
| Saturn Aura | 22 | 33 |
| Kia Optima | 22 | - |
| Hyundai Sonata | 22 | 32 |
| Honda Accord | 22 | 31 |
| Chevrolet Malibu | 22 | 30 |
| Toyota Camry | 21 | 31 |
| Volkswagen Passat | 21 | 31 |
| Mazda 6 | 21 | 30 |
| Chrysler Sebring | 21 | 30 |
| Dodge Avenger | 21 | 30 |
| Ford Fusion | 20 | 29 |
| Mercury Milan | 20 | 29 |
| Mitsubishi Galant | 20 | 27 |
| Subaru Legacy | 20 | 27 |
| Nissan Maxima | 19 | 260 |
| Nissan Altima | 19 | 26 |
| Mercury Sable | 18 | 28 |
| Hyundai Azera | 18 | 26 |
| Buick LaCrosse/Allure | 17 | 28 |

## 3. Quelle est la taille de chaque échantillon

## 4. Éfficacité gagnée en ville

Vous allez calculé le mode, la médiane et la moyenne pour la vitesse en **ville** en 1994 et 2009. Quelle conclusion tirez-vous? A cette étape de l’exercice, je vous demande de faire les calculs à la main.

## 5. Éfficacité sur autoroute

Le calcul que vous venez de faire est trop long. On peut présenter les données précédentes sous forme de données agrégées. C’est quoi encore les données agrégées?

4.1 Regrouper les données de la **vitesse sur l’autoroute** sous forme agrégée. Cela veut dire qu’il faut dénombrer le nombre de voitures pour chaque niveau de vitesse. Faite cela pour les données de 1994 et de 2009.

### 5.1 Présenter dans ce même tableau les fréquences, et les fréquences cumulées

### 5.2 Quelle représentation graphique vous semble la plus appropriée pour ces données?

### 5.3 Calculer à nouveau le mode, la médiane et la moyenne à partir de ses données groupées. Quelle conclusion tirez-vous?

————— PAS NÉCESSAIRE LA QUESTION 6——————-

## 6. Utilisation de R

Maintenant, nous allons utiliser R pour faire le même travail. Voici comment vous allez procéder.

1. Créer la base de données **donnee\_1994** avec les variable suivantes:

* modele
* vitesse\_ville et
* vitesse\_autoroute

Vous comprenez que cette base de données contient donc 18 observations pour 3 variables. Quelle est la nature de chaque variable?

### Réponse 1

1. Calculer la moyenne, la médiane et le mode des deux variables **vitesse\_ville** et **vitesse\_autoroute** à partir des données **donnee\_1994**.

* Commenter vos résultats. Si vous vous rappelez, pour calculer la moyenne et la médiane, il faut utiliser les fonctions **mean** et **mediane**.
* Cependant, il **N’existe PAS** de fonction **mode** pour calculer le mode. Je vous demande de faire quelques recherches et me venir avec une solution. Il est dès fois important de ne pas se focaliser pour comprendre ce que vous faites du moment où ça marche. Donnez-vous le temps de le comprendre plus tard.

1. Il y a plusieurs autres paramètres de tendance centrale que les trois que nous avons vus en classe. Vous avez le minimum, le maximum, le premier quartile, le 3e quartile et plus généralement les **ntiles**. Calculer ces différents paramètres sur les variables vitesse\_ville et vitesse\_autoroute. Commenter vos résultats.
2. La fonction **descr** de summarytools vous permet aussi de calculer ces paramètres de tendance centrale. Utiliser cette fonction pour calculer les paramètres calculer au 2 et 3.
3. Maintenant, refaite la même chose avec les données de 2009.
4. Quelle conclusion tirez-vous sur la solution technologique au changement climatique?