

Partiel de Statistiques descriptives avec R – mai 2024

Durée 2 heure, sans documents ni calculatrice.

Exercice 1

Partie 1

1. Quels types de variables existent en R ?
2. Utiliser des instructions R pour répondre aux questions suivantes:
 - a. Créer le vecteur `c(0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0)` de trois manières différentes.
 - b. Créer un vecteur contenant une fois la lettre A, deux fois la lettre B, etc., 26 fois la lettre Z.
 - c. Quelle est la longueur de cette suite ? (Utiliser la chaîne `LETTERS` prédéfinie).

Partie 2

1. Écrire le résultat de l'opération suivante:

```
u = c(10, 20, 30)
v = c(1, 2, 3, 4, 5, 6, 7, 8, 9)
u+v
```

2. Écrire le résultat du code suivant:

```
v = c(2,4,1,9)
v[3]
v[-1]
v[c(T,T,F,F)]
v > 4
v[v>4]
which(v>4)
```

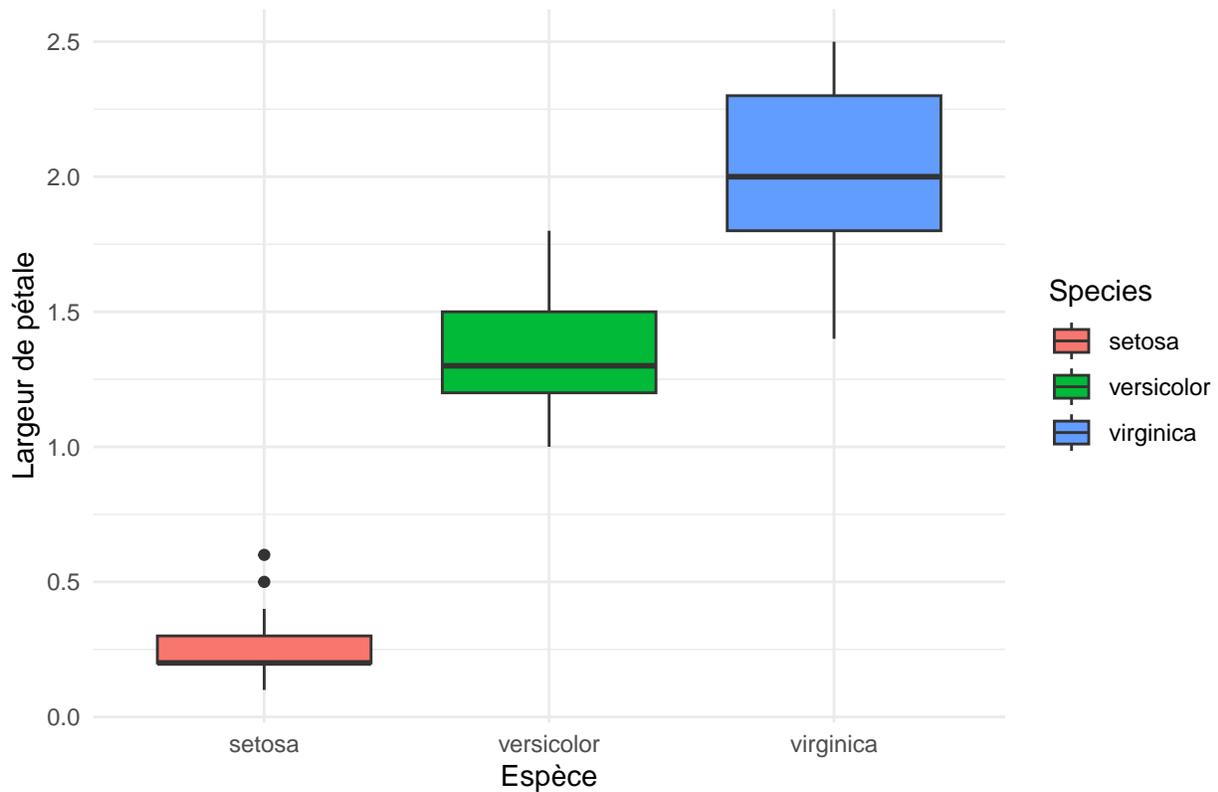
Partie 3

```
data("iris")
x<-iris$Sepal.Length
trim<-0.01
n<-length(x)
k<-round(n*trim)
x.trimmed <-sort(x)[(k+1):(n-k)]
mu.hat<-mean(x.trimmed)
sd.hat<-sqrt(var(x.trimmed))
```

1. Que calcul le code précédent ?
2. Quels avantages présentent `mu.hat` et `sd.hat` par rapport aux estimateurs classiques de la moyenne et l'écart type ?
3. Ecrire le code sous forme de fonction qui prend `x` en entrée et retourne `mu.hat` et `sd.hat`.

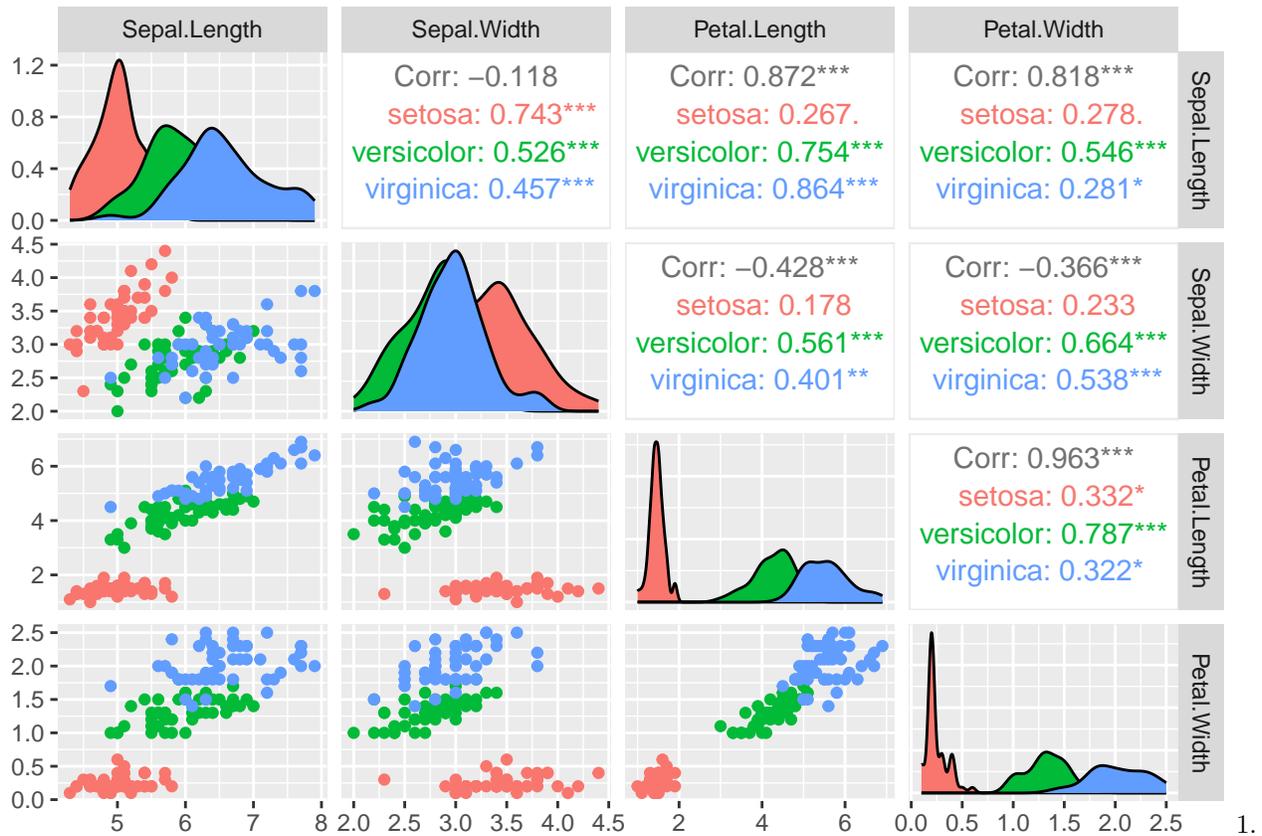
Exercice 2

Boxplot de la largeur de pétale des iris



1. Dans un boxplot que représentent les trois traits horizontaux?
2. Quel pourcentage de l'échantillon se trouve entre les traits inférieurs et supérieurs de la boîte ?
3. Commenter en une phrase les 3 boxplots.

Exercice 3



Comment s'appelle ce type de graphe ? 2. Qu'est-ce qui est représenté sur la diagonale ? 3. Que peut on dire sur les 3 espèces de fleurs à la vue du graphique ?

Exercice 4

Soit le tableau suivant décrivant les relations entre deux variables aléatoires X et Y

	$Y = m_1$	$Y = m_2$	$Y = m_3$
$X = l_1$	n_{11}	n_{12}	n_{13}
$X = l_2$	n_{21}	n_{22}	n_{23}

Nous noterons

- $n_{i\bullet} = \sum_j n_{ij}$ (marge en ligne)
- $n_{\bullet j} = \sum_i n_{ij}$ (marge en colonne)
- $n = \sum_{ij} n_{ij} = \sum_i n_{i\bullet} = \sum_j n_{\bullet j}$, le nombre total d'individus de l'échantillon.

Partie 1

1. Comment s'appelle ce type de tableau ?
2. Donner une estimation de la probabilité $P(X = l_1, Y = m_2)$.
3. Donner une estimation de la probabilité $P(Y = m_2)$.
4. Donner une estimation de la probabilité $P(X = l_1 | Y = m_2)$.
5. Si les deux variables X et Y étaient indépendantes comment pourrait on estimer la probabilité jointe $P(X = l_i, Y = m_j)$?

Partie 2

Nous considérerons Y comme une variable qualitative prenant deux modalités, *fumeur* et *non-fumeur* et X une variable qualitative décrivant le niveau d'étude, *Bac et moins*, *Licence*, *Master et plus*:

	Bac et moins	Licence	Master et plus
fumeur	113	52	22
non-fumeur	80	58	30

1. Calculer les marges du tableau.
2. Utiliser des statistiques descriptives pour décrire X .
3. Représenter X par un diagramme en bâtons.



4. Expliquer comment est produite la mosaïque suivante:
5. Commenter cette représentation.
6. Calculer le tableau des effectifs théoriques sous hypothèse d'indépendance.