Chapitre 3 : Analyse statistique des données géologiques avec R

3.1 Manipulation des données géologiques dans R

3.1.1 Utilisation de dplyr et tidyr pour manipuler les données

Objectif : Organiser et nettoyer les données géologiques pour les rendre exploitables dans R.

- `dplyr` : filtrage, tri, résumé, transformation des données.
- `tidyr` : restructuration des données (ex. : pivot_longer, pivot_wider).

```
# تحميل مكتبة dplyr لمعالجة البيانات
library(dplyr)
# تحميل مكتبة tidyr لإعادة هيكلة البيانات
library(tidyr)

# تحميل البيانات الجيولوجية من ملف CSV géologiques depuis un fichier CSV
data <- read.csv("donnees_geologiques.csv")

# فلترة الآبار التي يتجاوز عمقها 100 متر avec une profondeur > 100m
data_filtered <- data %>%
  filter(Profondeur > 100)

# حساب متوسط المسامية لكل موقع la moyenne des valeurs de porosité par site
data_# إظهار ملخص إحصائي لقيم المتغير
summary <- data %>%
  group_by(Site) %>%
  summarise(Moyenne_Porosite = mean(Porosite, na.rm = TRUE))
```

3.2 Analyse statistique des données géologiques

3.2.1 Statistiques descriptives

- Moyenne, médiane, écart-type, quartiles.
- Identification des valeurs extrêmes ou manquantes.

```
summary(data$Porosite)
# حساب الانحراف المعياري للقيم
sd(data$Porosite, na.rm = TRUE)
```

3.2.2 # إنشاء نموذج انحدار خطي لدراسة العلاقة بين المتغيرات s et tests statistiques

- Régression linéaire entre variables géologiques.

• Tests statistiques pour comparer des groupes.

```
modele <- lm(Porosite ~ Profondeur, data = data)
summary(modele)
```

3.2.3 Analyses multivariées

تحليل المكونات الرئيسية PCA

PCA (Analyse en Composantes Principales)
• Clustering

```
library(FactoMineR)
library(factoextra)

res_pca <- PCA(data[, c("Porosite", "Permeabilite", "Densite")], graph =
FALSE)
fviz_pca_ind(res_pca, col.ind = "cos2")
```

3.3 Visualisation des résultats géologiques avec # إنشاء رسم بياني باستخدام ggplot2

ggplot2

• Graphiques de dispersion, histogrammes, cartes thématiques.
• Utilisation de couleurs et formes pour distinguer les sites.

```
library(ggplot2)

ggplot(data, aes(x = Profondeur, y = Porosite, color = Site)) +
  geom_point() +
  labs(title = "Porosité selon la profondeur", x = "Profondeur (m)", y =
"Porosité (%)") +
  theme_minimal()
```

3.4 Exemple pratique complet

```
library(dplyr)
library(ggplot2)
library(FactoMineR)
library(factoextra)
```

```r
# إنشاء مجموعة بيانات مثال – بيانات جيولوجية بسيطة
# هنا نعرّف مواقع مختلفة مع قيم العمق والإنفاذية والمسامية

data <- data.frame(
  Site = rep(c("SiteA", "SiteB", "SiteC"), each = 5),
  Profondeur = c(50, 120, 80, 200, 150, 60, 110, 90, 130, 170, 70, 140, 100,
180, 160),
  Porosite = c(12, 15, 13, 18, 16, 11, 14, 12, 17, 15, 13, 16, 14, 19, 17),
  Permeabilite = c(10, 20, 15, 25, 22, 12, 18, 16, 21, 23, 14, 19, 17, 24,
22)
)

# 1. فلترة البيانات – اختيار الحفر التي عمقها أكبر من 100 متر

data_filtered <- data %>% filter(Profondeur > 100)

# 2. حساب متوسط المسامية لكل موقع
# نستخدم group_by لتجميع البيانات حسب الموقع ثم summarise لحساب المتوسط

data_summary <- data_filtered %>%
  group_by(Site) %>%
  summarise(Moyenne_Porosite = mean(Porosite, na.rm = TRUE))
print(data_summary)

# 3. إنشاء نموذج انحدار خطي لدراسة العلاقة بين المسامية والعمق
modele <- lm(Porosite ~ Profondeur, data = data)
summary(modele)

# 4. تحليل PCA لتقليل الأبعاد بين المسامية والإنفاذية
res_pca <- PCA(data[, c("Porosite", "Permeabilite")], graph = FALSE)
fviz_pca_ind(res_pca, col.ind = "cos2", gradient.cols = c("blue", "yellow",
"red"))

# 5. رسم بياني يوضح علاقة المسامية بالعمق
# النقاط ملوّنة حسب الموقع + خط انحدار

ggplot(data, aes(x = Profondeur, y = Porosite, color = Site)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed", color =
"black") +
  theme_minimal()
```