RWorksheet_Sobusa#3b.Rmd

Nexon Sobusa

2024-09-30

1. Creating a Data Frame

a. Write the codes

data <- data.frame(Respondent = c(1, 2, 3, 4, 5), Sex = c("Male", "Female", "Male", "Female", "Male"), Fathers_Occupation = c("Farmer", "Driver", "Farmer", "Others", "Driver"), Number_of_Siblings = c(5, 6, 3, 4, 2), Types_of_Houses = c("Wood", "Concrete", "Semi-Concrete", "Wood", "Concrete"))

b. Structure or summary of the data

str(data) summary(data)

c. Is the mean number of siblings attending school 5?

mean(data$Number_of_Siblings) == 5

d. Extract the 1st two rows and all columns

subset_1 <- data[1:2,] print(subset_1)

e. Extract 3rd and 5th row with 2nd and 4th column

subset_2 <- data[c(3, 5), c(2, 4)] print(subset_2)

f. Select the variable 'Types of Houses' and store as types_houses

types_houses <- data$Types_of_Houses print(types_houses)

g. Select all Male respondents whose father's occupation is 'Farmer'

male_farmer <- subset(data, Sex == "Male" & Fathers_Occupation == "Farmer") print(male_farmer)

h. Select all Female respondents with >=5 number of siblings attending school

female_siblings <- subset(data, Sex == "Female" & Number_of_Siblings >= 5) print(female_siblings)

2. Creating an empty data frame

df <- data.frame(Ints = integer(), Doubles = double(), Characters = character(), Logicals = logical(), Factors = factor(), stringsAsFactors = FALSE)

print("Structure of the empty dataframe:") str(df)

3. Creating a CSV file and saving it as HouseholdData.csv

write.csv(data, "HouseholdData.csv")

a. Import the CSV file into the R environment

imported_data <- read.csv("HouseholdData.csv") print(imported_data)

b. Convert 'Sex' into factor and change it into integer

imported_data$Sex <- -as.integer(factor(imported_data$Sex, levels = c("Male", "Female"))) print(imported_data$Sex)

c. Convert 'Types of Houses' into factor and change it into integer

imported_data$Types_of_Houses <- -as.integer(factor(imported_data$Types_of_Houses, levels = c("Wood", "Concrete", "Semi-Concrete"))) print(imported_data$Types_of_Houses)

d. Convert 'Father's Occupation' into factor with Farmer=1, Driver=2, Others=3

imported_data$Fathers_Occupation <- -as.integer(factor(imported_data$Fathers_Occupation, levels = c("Farmer", "Driver", "Others"))) print(imported_data$Fathers_Occupation)

e. Select all Female respondents whose father's occupation is Driver

female_driver <- subset(imported_data, Sex == 2 & Fathers_Occupation == 2) print(female_driver)

f. Select respondents with >=5 siblings attending school

siblings_5plus <- subset(imported_data, Number_of_Siblings >= 5) print(siblings_5plus)