RWorksheet_Sobusa#3b.Rmd

Nexon Sobusa

2024-09-30

1. Create a data frame based on the table in the instructions.

df <- data.frame(Respondent = c(1, 2, 3, 4, 5), Sex = c('Male', 'Female', 'Male', 'Female', 'Female'), FatherOccupation = c('Farmer', 'Driver', 'Others', 'Farmer', 'Driver'), SiblingsAttending = c(3, 6, 5, 2, 4), TypesOfHouses = c('Wood', 'Concrete', 'Semi-Concrete', 'Wood', 'Concrete'))

Describe the data structure.

str(df) summary(df)

c. Check if the mean number of siblings attending is 5.

mean_siblings <- mean(df$SiblingsAttending) mean_siblings == 5

d. Extract the 1st two rows and all columns.

df[1:2,] # Corrected: Using indexing instead of subset

e. Extract the 3rd and 5th rows with the 2nd and 4th columns.

df[c(3, 5), c(2, 4)]

f. Select the variable types of houses and store it as types_houses.

types_houses <- df$TypesOfHouses types_houses

g. Select all males whose father occupation is farmer.

subset(df, Sex == "Male" & FatherOccupation == "Farmer")

h. Select females with 5 or more siblings attending school.

subset(df, Sex == "Female" & SiblingsAttending >= 5)

2. Create an empty dataframe.

df_empty <- data.frame(Ints = integer(), Doubles = double(), Characters = character(), Logicals = logical(), Factors = factor(), stringsAsFactors = FALSE) print("Structure of the empty dataframe:") str(df_empty)

3. Create a .csv file of the respondent data frame and save it as "HouseholdData.csv".

write.csv(df, "HouseholdData.csv", row.names = FALSE) # Save df instead of respondent_data

a. Import the csv file into the R environment.

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

Check the structure of the imported data to ensure columns are correct

str(imported_data) print(names(imported_data)) # Print column names to verify them

b. Convert the Sex column into factor and change it to integer (Male = 1, Female = 2).

imported_data$Sex $< -factor(imported_data$Sex, levels = c("Male", "Female"), labels = c(1, 2)) print(imported_data$Sex) # Display the converted column

c. Convert the TypesOfHouses into factor and change it into integer (Wood = 1, Concrete = 2, Semi-Concrete = 3).

imported_data$TypesOfHouses $< -factor(imported_data$TypesOfHouses, levels = c("Wood", "Concrete", "Semi-Concrete"), labels = c(1, 2, 3)) print(imported_data$TypesOfHouses) # Display the converted column

d. Factor the Father's occupation (Farmer = 1, Driver = 2, Others = 3).

imported_data$FatherOccupation $< -factor(imported_data$FatherOccupation, levels = c("Farmer", "Driver", "Others"), labels = c(1, 2, 3)) print(imported_data$FatherOccupation) # Display the converted column

e. Select all female respondents whose father is a driver.

female_driver <- subset(imported_data, Sex == 2 & FatherOccupation == 2) print("Female respondents whose father is a driver:") print(female_driver)

f. Select respondents who have greater than or equal to 5 siblings attending school.

siblings_five_or_more <- subset(imported_data, SiblingsAttending >= 5) print("Respondents with >= 5 siblings attending school:") print(siblings_five_or_more)