RWorksheet_Arenal#3b.Rmd

Jo Carlo Arenal

2024-10-06

```
# 1
# a
data <- data.frame(
  Respondent = 1:20,
  Sex = c(2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 1, 2, 1, 2, 2, 1, 1, 1, 2, 2),
  Fathers_Occupation = c(1, 3, 3, 1, 2, 3, 3, 2, 2, 2, 1, 3, 3, 3, 3, 3, 1, 2, 1, 1),
  Persons_at_Home = c(5, 7, 3, 8, 5, 6, 7, 8, 7, 4, 7, 3, 8, 7, 5, 3, 7, 11, 7, 6),
  Siblings_at_School = c(6, 4, 3, 2, 2, 5, 6, 3, 4, 2, 4, 3, 7, 5, 3, 4, 3, 3, 3, 2),
  Types_of_Houses = c(1, 2, 3, 1, 2, 2, 1, 2, 1, 3, 3, 2, 3, 1, 3, 3, 1, 3, 1, 2)
)

print(data)
```

```
##    Respondent Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   2                  1               5                  6
## 2           2   2                  3               7                  4
## 3           3   2                  3               3                  3
## 4           4   1                  1               8                  2
## 5           5   2                  2               5                  2
## 6           6   2                  3               6                  5
## 7           7   2                  3               7                  6
## 8           8   2                  2               8                  3
## 9           9   2                  2               7                  4
## 10         10   1                  2               4                  2
## 11         11   1                  1               7                  4
## 12         12   2                  3               3                  3
## 13         13   1                  3               8                  7
## 14         14   2                  3               7                  5
## 15         15   2                  3               5                  3
## 16         16   1                  3               3                  4
## 17         17   1                  1               7                  3
## 18         18   1                  2              11                  3
## 19         19   2                  1               7                  3
## 20         20   2                  1               6                  2
##    Types_of_Houses
## 1                1
## 2                2
## 3                3
## 4                1
## 5                2
## 6                2
## 7                1
## 8                2
```

```
## 9                1
## 10               3
## 11               3
## 12               2
## 13               3
## 14               1
## 15               3
## 16               3
## 17               1
## 18               3
## 19               1
## 20               2
```

```
# b
str(data)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondent        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex               : num  2 2 2 1 2 2 2 2 2 1 ...
##  $ Fathers_Occupation: num  1 3 3 1 2 3 3 2 2 2 ...
##  $ Persons_at_Home   : num  5 7 3 8 5 6 7 8 7 4 ...
##  $ Siblings_at_School: num  6 4 3 2 2 5 6 3 4 2 ...
##  $ Types_of_Houses   : num  1 2 3 1 2 2 1 2 1 3 ...
```

```
summary(data)
```

```
##    Respondent          Sex       Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00       Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:1.00   1st Qu.:1.00       1st Qu.: 5.0
##  Median :10.50   Median :2.00   Median :2.00       Median : 7.0
##  Mean   :10.50   Mean   :1.65   Mean   :2.15       Mean   : 6.2
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00       3rd Qu.: 7.0
##  Max.   :20.00   Max.   :2.00   Max.   :3.00       Max.   :11.0
##  Siblings_at_School Types_of_Houses
##  Min.   :2.00       Min.   :1
##  1st Qu.:3.00       1st Qu.:1
##  Median :3.00       Median :2
##  Mean   :3.70       Mean   :2
##  3rd Qu.:4.25       3rd Qu.:3
##  Max.   :7.00       Max.   :3
```

```
# c
mean_siblings <- mean(data$Siblings_at_School)
is_mean_five <- mean_siblings == 5
mean_siblings
```

```
## [1] 3.7
```

```
is_mean_five
```

```
## [1] FALSE
```

```
# d
first_two_rows <- data[1:2, ]
print(first_two_rows)
```

```
##   Respondent Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1   2                  1               5                  6
```

```
## 2          2   2                  3                7                  4
##   Types_of_Houses
## 1               1
## 2               2
```

```
# e
subset_rows <- data[c(3, 5), c(2, 4)]
print(subset_rows)
```

```
##   Sex Persons_at_Home
## 3   2               3
## 5   2               5
```

```
# f
types_houses <- data$Types_of_Houses
print(types_houses)
```

```
##  [1] 1 2 3 1 2 2 1 2 1 3 3 2 3 1 3 3 1 3 1 2
```

```
# g
males_farmers <- subset(data, Sex == 1 & Fathers_Occupation == 1)
print(males_farmers)
```

```
##    Respondent Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4           4   1                  1               8                  2
## 11         11   1                  1               7                  4
## 17         17   1                  1               7                  3
##    Types_of_Houses
## 4                1
## 11               3
## 17               1
```

```
# h
females_with_siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
print(females_with_siblings)
```

```
##    Respondent Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   2                  1               5                  6
## 6           6   2                  3               6                  5
## 7           7   2                  3               7                  6
## 14         14   2                  3               7                  5
##    Types_of_Houses
## 1                1
## 6                2
## 7                1
## 14               1
```

```
# 2
df <- data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE
)

# a
```

```r
print("Structure of the empty dataframe:")
```

```
## [1] "Structure of the empty dataframe:"
```

```r
str(df)
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

```r
# 3
# a
data <- read.csv("/cloud/project/HouseholdData.csv")

head(data)
```

```
##   Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1   Male                  1               5                  2
## 2           2 Female                  2               7                  3
## 3           3 Female                  3               3                  0
## 4           4   Male                  3               8                  5
## 5           5   Male                  1               6                  2
## 6           6 Female                  2               4                  3
##   Types.of.Houses
## 1            Wood
## 2        Concrete
## 3        Concrete
## 4            Wood
## 5   Semi-concrete
## 6   Semi-concrete
```

```r
# b
data$Sex <- factor(data$Sex, levels = c("Male", "Female"), labels = c(1, 2))

data$Sex <- as.integer(data$Sex)

head(data)
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1   1                  1               5                  2
## 2           2   2                  2               7                  3
## 3           3   2                  3               3                  0
## 4           4   1                  3               8                  5
## 5           5   1                  1               6                  2
## 6           6   2                  2               4                  3
##   Types.of.Houses
## 1            Wood
## 2        Concrete
## 3        Concrete
## 4            Wood
## 5   Semi-concrete
## 6   Semi-concrete
```

```r
# c
data$Types.of.houses <- factor(data$Types.of.Houses,
                               levels = c("Wood", "Concrete", "Semi-concrete"),
                               labels = c(1, 2, 3))

head(data)
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1   1                  1              5                  2
## 2           2   2                  2              7                  3
## 3           3   2                  3              3                  0
## 4           4   1                  3              8                  5
## 5           5   1                  1              6                  2
## 6           6   2                  2              4                  3
##   Types.of.Houses Types.of.houses
## 1            Wood               1
## 2        Concrete               2
## 3        Concrete               2
## 4            Wood               1
## 5   Semi-concrete               3
## 6   Semi-concrete               3
```

```r
str(data)
```

```
## 'data.frame':    10 obs. of  7 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : int  1 2 2 1 1 2 2 1 2 1
##  $ Fathers.Occupation: int  1 2 3 3 1 2 2 3 1 3
##  $ Persons.at.Home   : int  5 7 3 8 6 4 4 2 11 6
##  $ Siblings.at.School: int  2 3 0 5 2 3 1 2 6 2
##  $ Types.of.Houses   : chr  "Wood" "Concrete" "Concrete" "Wood" ...
##  $ Types.of.houses   : Factor w/ 3 levels "1","2","3": 1 2 2 1 3 3 1 3 3 2
```

```r
# d
data$Fathers.Occupation <- factor(data$Fathers.Occupation,
                                  levels = c( 1, 2 , 3),
                                  labels = c( "Farmer", "Driver", "Others"))

head(data)
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1   1             Farmer              5                  2
## 2           2   2             Driver              7                  3
## 3           3   2             Others              3                  0
## 4           4   1             Others              8                  5
## 5           5   1             Farmer              6                  2
## 6           6   2             Driver              4                  3
##   Types.of.Houses Types.of.houses
## 1            Wood               1
## 2        Concrete               2
## 3        Concrete               2
## 4            Wood               1
## 5   Semi-concrete               3
## 6   Semi-concrete               3
```

```r
str(data)
```

```
## 'data.frame':    10 obs. of  7 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : int  1 2 2 1 1 2 2 1 2 1
##  $ Fathers.Occupation: Factor w/ 3 levels "Farmer","Driver",..: 1 2 3 3 1 2 2 3 1 3
##  $ Persons.at.Home   : int  5 7 3 8 6 4 4 2 11 6
##  $ Siblings.at.School: int  2 3 0 5 2 3 1 2 6 2
##  $ Types.of.Houses   : chr  "Wood" "Concrete" "Concrete" "Wood" ...
##  $ Types.of.houses   : Factor w/ 3 levels "1","2","3": 1 2 2 1 3 3 1 3 3 2
```

```r
# e
female_drivers <- data[data$Sex == 2 & data$Fathers.Occupation == 2, ]

print(female_drivers)
```

```
## [1] Respondents        Sex                Fathers.Occupation Persons.at.Home
## [5] Siblings.at.School Types.of.Houses    Types.of.houses
## <0 rows> (or 0-length row.names)
```

```r
# f
siblings_ge_5 <- data[data$Siblings.at.School >= 5, ]

print(siblings_ge_5)
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1             Others               8                  5
## 9           9   2             Farmer              11                  6
##   Types.of.Houses Types.of.houses
## 4            Wood               1
## 9   Semi-concrete               3
```