Project 2 Part 1

Kaushal Khatiwada

2024-03-18

PROBLEM: Minimum value 0

```r
covid<- read.csv("covnep_252days.csv")
summary(covid$totalCases)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       2     963   13376   19341   77816
```

We get minimum value of Total Cases as 0

Subsetting based on Total cases greater than 1

```r
new_covid<-subset(covid,covid$totalCases>=1)
summary(new_covid$totalCases)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1     108   11754   17465   24956   77816
```

Get frequencies of q01, q03, q06 & q08

Compute frequencies, percentage, valid percentage and cumulative percentage

```r
library(foreign)
data <- read.spss("SAQ8.sav")
```

For Q1 "Statistics makes me cry"

```r
q01 <- data$q01           # Get the q01 column
q01_level <- levels(q01)      # Get the categorical levels
q01_freq <- as.numeric(table(q01))  # Get the frequency to each category
q01_perc <- as.numeric(round(prop.table(q01_freq) * 100, 1))   # Calculate the Percentage
q01_valid_per <- as.numeric(round(prop.table(q01_freq) * 100, 1))   #Calculate valid percentage
q01_cum_per <- cumsum(q01_perc)         #Calculate Cumulative Percentage

q1_df <- data.frame(                    #Create Data Frame
```

```
    Levels = q01_level,
    Frequency = q01_freq,
    Percent = q01_perc,
    Valid_Percent = q01_valid_per,
    Cumulative_Percent = q01_cum_per
)
q1_total = c("Total",sum(q1_df$Frequency),sum(q1_df$Percent),sum(q1_df$Valid_Percent),"")
q1<-rbind(q1_df,q1_total)     # Row bind the Total of frequency, percentage and valid percentage in dat
q1
```

```
##               Levels Frequency Percent Valid_Percent Cumulative_Percent
## 1     Strongly agree       270    10.5          10.5               10.5
## 2              Agree      1338      52            52               62.5
## 3            Neither       735    28.6          28.6               91.1
## 4           Disagree       187     7.3           7.3               98.4
## 5 Strongly disagree        41     1.6           1.6                100
## 6       Not answered         0       0             0                100
## 7              Total      2571     100           100
```

For Q3 "Standard deviations excite me"

```
q03 <- data$q03
q03_level <- levels(q03)
q03_freq <- as.numeric(table(q03))
q03_perc <- as.numeric(round(prop.table(q03_freq) * 100, 1))
q03_valid_per <- as.numeric(round(prop.table(q03_freq) * 100, 1))
q03_cum_per <- cumsum(q03_perc)

q3_df <- data.frame(
    Levels = q03_level,
    Frequency = q03_freq,
    Percent = q03_perc,
    Valid_Percent = q03_valid_per,
    Cumulative_Percent = q03_cum_per
)
q3_total = c("Total",sum(q3_df$Frequency),sum(q3_df$Percent),sum(q3_df$Valid_Percent),"")
q3<-rbind(q3_df,q3_total)
q3
```

```
##               Levels Frequency Percent Valid_Percent Cumulative_Percent
## 1     Strongly agree       497    19.3          19.3               19.3
## 2              Agree       672    26.1          26.1               45.4
## 3            Neither       878    34.2          34.2               79.6
## 4           Disagree       448    17.4          17.4                 97
## 5 Strongly disagree        76       3             3                100
## 6              Total      2571     100           100
```

For Q6 "I have little experience of computers"

```r
q06 <- data$q06
q06_level <- levels(q06)
q06_freq <- as.numeric(table(q06))
q06_perc <- as.numeric(round(prop.table(q06_freq) * 100, 1))
q06_valid_per <- as.numeric(round(prop.table(q06_freq) * 100, 1))
q06_cum_per <- cumsum(q06_perc)

q6_df <- data.frame(
  Levels = q06_level,
  Frequency = q06_freq,
  Percent = q06_perc,
  Valid_Percent = q06_valid_per,
  Cumulative_Percent = q06_cum_per
)
q6_total = c("Total",sum(q6_df$Frequency),sum(q6_df$Percent),sum(q6_df$Valid_Percent),"")
q6<-rbind(q6_df,q6_total)
q6
```

```
##              Levels Frequency Percent Valid_Percent Cumulative_Percent
## 1    Strongly agree       702    27.3          27.3               27.3
## 2             Agree      1127    43.8          43.8               71.1
## 3           Neither       344    13.4          13.4               84.5
## 4          Disagree       252     9.8           9.8               94.3
## 5 Strongly disagree       146     5.7           5.7                100
## 6             Total      2571     100           100
```

For Q8 "I have never been good at mathematics"

```r
q08 <- data$q08
q08_level <- levels(q08)
q08_freq <- as.numeric(table(q08))
q08_perc <- as.numeric(round(prop.table(q08_freq) * 100, 1))
q08_valid_per <- as.numeric(round(prop.table(q08_freq) * 100, 1))
q08_cum_per <- cumsum(q08_perc)

q8_df <- data.frame(
  Levels = q08_level,
  Frequency = q08_freq,
  Percent = q08_perc,
  Valid_Percent = q08_valid_per,
  Cumulative_Percent = q08_cum_per
)
q8_total = c("Total",sum(q8_df$Frequency),sum(q8_df$Percent),sum(q8_df$Valid_Percent),"")
q8<-rbind(q8_df,q8_total)
q8
```

```
##              Levels Frequency Percent Valid_Percent Cumulative_Percent
## 1    Strongly agree       383    14.9          14.9               14.9
```

```
## 2            Agree   1487   57.8        57.8              72.7
## 3          Neither    482   18.7        18.7              91.4
## 4         Disagree    147    5.7         5.7              97.1
## 5 Strongly disagree     72    2.8         2.8              99.9
## 6            Total   2571   99.9        99.9
```

INCOME FREQUENCY

Create Table using inco1, inco2, inco3, inco4, inco5, inco6 and inco7 Has multiple response 0 as NO and 1 as YES)

```r
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
mr_drug <- read_xlsx("MR_Drugs.xlsx")    # Retrieve data from excel file
```

Select inco1,. . .,inco7 column and Calculate the frequency of each income source. Calculate percentage for the income source and for the cases.

```r
incomes_freq_table <- select(mr_drug, inco1, inco2, inco3, inco4, inco5, inco6, inco7)
has_income <- colSums(incomes_freq_table)
Total_has_income <- sum(has_income)          # Total frequency of observation having income source
Total_observation <- colSums(!is.na(incomes_freq_table))    # Total observation

# Percentage of having each income source to total income source frequency i.e (226/1761) * 100
Perc_per_income <- round(as.numeric((has_income/Total_has_income)*100),2)

# percentage of having the observation from the individual income source i.e(226/972)*100
perc_per_case <- round(as.numeric(has_income/Total_observation*100),2)
```

```r
income_frequencies<-data.frame(                      # Create Data Frame
  Income_Response = names(incomes_freq_table),
  N = has_income,
  Percent = Perc_per_income,
  Percent_of_cases = perc_per_case
)

total <- c("Total",Total_has_income, sum(Perc_per_income), sum(perc_per_case))
income_frequencies <- rbind(income_frequencies, total)      # Row bind the Total of frequency, percent
row.names(income_frequencies) <- NULL        # Remove the rownames.
income_frequencies
```

```
##   Income_Response    N Percent Percent_of_cases
## 1          inco1  226   12.83            23.25
## 2          inco2  607   34.47            62.45
## 3          inco3  293   16.64            30.14
## 4          inco4   50    2.84             5.14
## 5          inco5   82    4.66             8.44
## 6          inco6  151    8.57            15.53
## 7          inco7  352   19.99            36.21
## 8          Total 1761     100           181.16
```