

Mikro 5, opgave 1

Johan Bysted

2024-11-01

```
# Der er forskellige måder at skrive kode i R. Jeg bruger gerne tidyverse.  
#Det har betydning for hvordan man arbejder med data - laver nye variable, fjerner observationer osv.  
#Spørg mig til ChatGPT hvis du har brug for hjælp
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr   1.5.1  
## v ggplot2    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.1  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# For 3-4 stykker virker csv-filen ikke. Brug i stedet xls-filen, som også er på moodle
```

```
library(readxl)  
raw.data <- read_excel("TASK_mikro5.xls")
```

```
#Man kan lave en variabel for første ciffer ved at definere den ud fra intervaller  
#Her er kun to af intervallerne - ikke alle ni.
```

```
raw.data <- raw.data %>%  
  mutate(occu=case_when(ISCO_NU>=2000&ISCO_NU<3000~2,  
                        ISCO_NU>=3000&ISCO_NU<4000~3))
```

```
table(raw.data$occu)
```

```
##  
##      2      3  
## 389 199
```

```
#Man kan også bruge floor-kommandoen til at skære decimalerne væk efter man har divideret med 1000
```

```
new.dat <- raw.data %>%  
  mutate(one.dig=floor(ISCO_NU/1000))
```

```
table(new.dat$one.dig)
```

```
##
##    0    1    2    3    4    5    6    7    8    9
##    1  68 389 199 109 111    3  58  71 235
```

```
#Housa! Der er rigtig mange 9.
#Prøv at inspicere værdierne.
#Det ser ud til (hos Danmarks Statistik), at værdier over 9700 ikke burde eksistere
```

```
new.dat <- new.dat %>%
  filter(ISCO_NU<9700)

# Divider med antal rækker for at få procent
table(new.dat$one.dig)/nrow(new.dat)
```

```
##
##          0          1          2          3          4          5
## 0.0009363296 0.0636704120 0.3642322097 0.1863295880 0.1020599251 0.1039325843
##          6          7          8          9
## 0.0028089888 0.0543071161 0.0664794007 0.0552434457
```

```
# Værdien 0 observeres også. Det drejer sig om en enkelt værnepligtig, som har sneget sig med i data.
```

```
# I stedet for at printe så gemmer vi data, og så smider vi den værnepligtige ud
# Desuden kalder R automatisk variabelen fra tabellen for "Var1". Det ændres tilbage til one.dig
one.dig.table <- as.data.frame(table(new.dat$one.dig)/nrow(new.dat)) %>%
  filter(Var1!=0) %>%
  rename(one.dig=Var1)
```

```
#Indtaster intervalmidtpunkterne fra dagens tekst i en ny variabel, som jeg kalder pct.auto
one.dig.table <- one.dig.table %>%
  mutate(pct.auto=case_when(one.dig==1~7.5,
                             one.dig==2~10,
                             one.dig==3~30,
                             one.dig==4~55,
                             one.dig==5~30,
                             one.dig==6~22.5,
                             one.dig==7~45,
                             one.dig==8~65,
                             one.dig==9~47.5))
```

```
#Middelværdien for pct auto vægtet efter andele i beskæftigelsen.
weighted.mean(one.dig.table$pct.auto,one.dig.table$Freq)
```

```
## [1] 27.9194
```

```
#TASK er repræsentative for beskæftigede i Danmark når man anvender vægten perugt.
#Hvis vægten for en række er fx 1400 så betyder det, at rækken repræsenterer 1400 personer i arbejde
#Det er vigtigt at bruge vægtene hvis I vil arbejde med TASK fx til bachelorprojektet, men det er ikke
#Når I skal bruge vægtene kan I med fordel bruge pakken survey
```

0.A og 0.B

I ovenstående kode kan svarene på spørgsmål 1 og 2 ses. Svarene fra 0 til 9 er hvor stor en procentdel der findes af de forskellige jobs i Danmark. Lige knap 28 procent af jobs i Danmark forventes at blive automatiseret de næste 2 årtier.

0.C

Teksten beskriver at følgende jobtyper har risiko for at blive automatiseret: - Lowest wage jobs -> "the progression to the next batch of jobs, and the one after that, will roughly follow an upward-income trajectory, with the lowest wage jobs being automated first, and the highest wage jobs being automated last." - Derudover beskriver teksten at disse jobtyper er i højrisiko -> "The study lists two dozen job categories in order of risk, and the breakdown essentially boils down to this: foodservice, production, office and administrative support, farming/fishing/forestry, transportation and material moving, construction and extraction, and installation/maintenance/repair appear to round out the most high-risk occupations." - Derudover så beskriver teksten også, at jobs der involverer social interaktion er meget udsat.

0.D - Størstedelen af kode E i oversigten og jobs involverer jobs der har med social interaktion at gøre, så kode E beskriver i stor grad de karakteristika i forhold til denne jobtype.