Homework 1

Course 0214 - Statistics for Economics with R

Paula Natascha Gillessen () Julian Sandalek ()
Julian Hofmaninger (12413122, h12413122@s.wu.ac.at)

2025-10-25

1. Basic R Operations

Task 1

```
numseq <- rep(seq(1,7,by=3), each=4) * c(-1,1)
complexseq <- c(rep(c(4,2,0), 2), rep(c(7,14,21), 3))
charpattern <- LETTERS[seq(2,26, by=2)]
numseq;complexseq;charpattern
```

```
##  [1] -1  1 -1  1 -4  4 -4  4 -7  7 -7  7
```

```
##  [1]  4  2  0  4  2  0  7 14 21  7 14 21  7 14 21
```

```
##  [1] "B" "D" "F" "H" "J" "L" "N" "P" "R" "T" "V" "X" "Z"
```

Task 2

```
vec <- 0:12
vec
```

```
##  [1]  0  1  2  3  4  5  6  7  8  9 10 11 12
```

```
vec[vec %% 2 == 1] <- NA
na_idx <- which(is.na(vec))
vec[na_idx] <- vec[na_idx-1]+vec[na_idx+1]
na_idx;vec
```

```
## [1]  2  4  6  8 10 12
```

```
##  [1]  0  2  2  6  4 10  6 14  8 18 10 22 12
```

Task 3

```
set.seed(1)
randvec <- rnorm(50, 0, 1)
subsetvec <- randvec[randvec > .5]
sum <- sum(randvec > 1)
randvec;subsetvec;sum
```

```
##  [1] -0.62645381  0.18364332 -0.83562861  1.59528080  0.32950777 -0.82046838
##  [7]  0.48742905  0.73832471  0.57578135 -0.30538839  1.51178117  0.38984324
## [13] -0.62124058 -2.21469989  1.12493092 -0.04493361 -0.01619026  0.94383621
## [19]  0.82122120  0.59390132  0.91897737  0.78213630  0.07456498 -1.98935170
## [25]  0.61982575 -0.05612874 -0.15579551 -1.47075238 -0.47815006  0.41794156
## [31]  1.35867955 -0.10278773  0.38767161 -0.05380504 -1.37705956 -0.41499456
## [37] -0.39428995 -0.05931340  1.10002537  0.76317575 -0.16452360 -0.25336168
## [43]  0.69696338  0.55666320 -0.68875569 -0.70749516  0.36458196  0.76853292
## [49] -0.11234621  0.88110773
```

```
##  [1] 1.5952808 0.7383247 0.5757814 1.5117812 1.1249309 0.9438362 0.8212212
##  [8] 0.5939013 0.9189774 0.7821363 0.6198257 1.3586796 1.1000254 0.7631757
## [15] 0.6969634 0.5566632 0.7685329 0.8811077
```

```
## [1] 5
```

Task 4

```
mat <- matrix(1:25, nrow=5, byrow=TRUE)
mat[4,];mat[,ncol(mat)];dim(mat)
```

```
## [1] 16 17 18 19 20
```

```
## [1]  5 10 15 20 25
```

```
## [1] 5 5
```

2. Analysis of Data on US States

Task 1

```
state <- as.data.frame(state.x77)
names(state) <- make.names(names(state))
variable_names <- names(state)
variable_names
```

```
## [1] "Population" "Income"     "Illiteracy" "Life.Exp"   "Murder"
## [6] "HS.Grad"    "Frost"      "Area"
```

Using the command `str(state)` we can explore the structure of the dataset. The output of this command also tells us the datatypes of the variables. After inspecting the output we can see that all variables are numerical, hence none is a categorical variable.

Task 2

```
LowEdu <- ifelse(state$HS.Grad < 48, TRUE, FALSE)
HighEdu <- ifelse(state$HS.Grad > 60, TRUE, FALSE)
MixedEdu <- ifelse(state$HS.Grad >= 48 & state$HS.Grad <= 60, TRUE, FALSE)
```

Task 3

```
LowEduNum <- ifelse(LowEdu, 1, 0)
EduStatus <- MixedEdu + 2 * HighEdu
table(EduStatus)
```

```
## EduStatus
##  0  1  2
## 13 29  8
```

Task 4

```
EduStatus <- as.factor(EduStatus)
```

```
levels(EduStatus) <- c("Low Education", "Mixed", "High Education")
state <- transform(state, Education_category=EduStatus)
head(state[1:2, c(1,ncol(state))])
```

```
##         Population Education_category
## Alabama       3615      Low Education
## Alaska         365     High Education
```

Task 5

```
state <- transform(state, Education_category=EduStatus)
head(state[1:2, c(1,ncol(state))])
```

```
##         Population Education_category
## Alabama       3615      Low Education
## Alaska         365     High Education
```

Task 6

```
low_education_data <- state[state$Education_category
                            == "Low Education", c("Illiteracy", "Life.Exp")]
head(low_education_data,2)
```

```
##          Illiteracy Life.Exp
## Alabama         2.1    69.05
## Arkansas        1.9    70.66
```