Statistik Übung 4

Michael Rynkiewicz

27/03/2019

Hilfs-Funktionen

```
groupedQuantile <- function(breaks, distribution, p){
  distribution.cumsum <- cumsum(distribution)
  Np <- sum(distribution) * p

  for(i in 2:length(distribution.cumsum)) {
    if( distribution.cumsum[i-1] <= Np && distribution.cumsum[i] >= Np ){
      e <- breaks[i-1]
      d <- breaks[i] - breaks[i-1]
      f <- distribution[i]
      Fi <- distribution.cumsum[i-1]

      return(e + d / f * (Np - Fi))
    }
  }

  return((breaks[1] + breaks[2]) / 2)
}

NVar <- function(x){
  x.mean <- mean(x)
  nvar <- sum((x - x.mean)^2) / length(x)
  return(nvar)
}

Nsd <- function(x){
  nvar <- NVar(x)
  nsd <- sqrt(nvar)
  return(nsd)
}

NCov <- function(x, y){
  x.mean <- mean(x)
  y.mean <- mean(y)

  ncov <- sum((x - x.mean) * (y - y.mean)) / length(x)
  return(ncov)
}
```

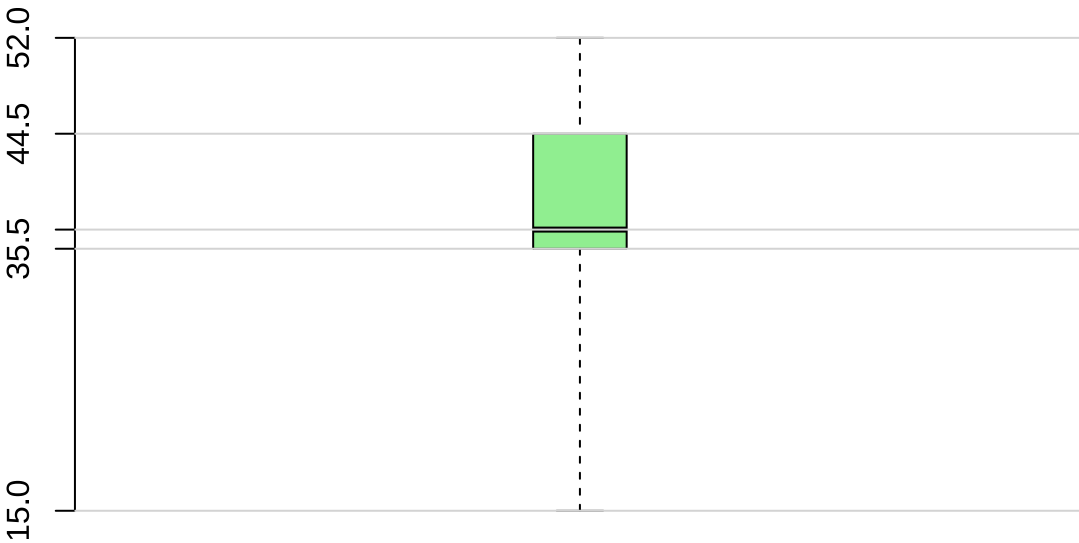
Aufgabe 27

```
traveledDistance <- c(39, 35, 25, 37, 15, 36, 50, 52, 37, 51, 39)
traveledDistance.boxplot <- boxplot(traveledDistance, plot = FALSE, range = 0)
traveledDistance.plotvalues <- c(traveledDistance.boxplot$stats)
names(traveledDistance.plotvalues) <- c("Minimum", "p0.25", "Median", "p0.75", "Maximum")
```

```
traveledDistance.plotvalues
```

```
## Minimum   p0.25  Median   p0.75 Maximum
##    15.0    35.5    37.0    44.5    52.0
```

```
boxplot(traveledDistance, col="lightgreen", axes = FALSE, boxwex=0.2, range = 0)
axis(2, at = traveledDistance.plotvalues)
abline(h = traveledDistance.plotvalues, col = "lightgray")
```



Aufgabe 28

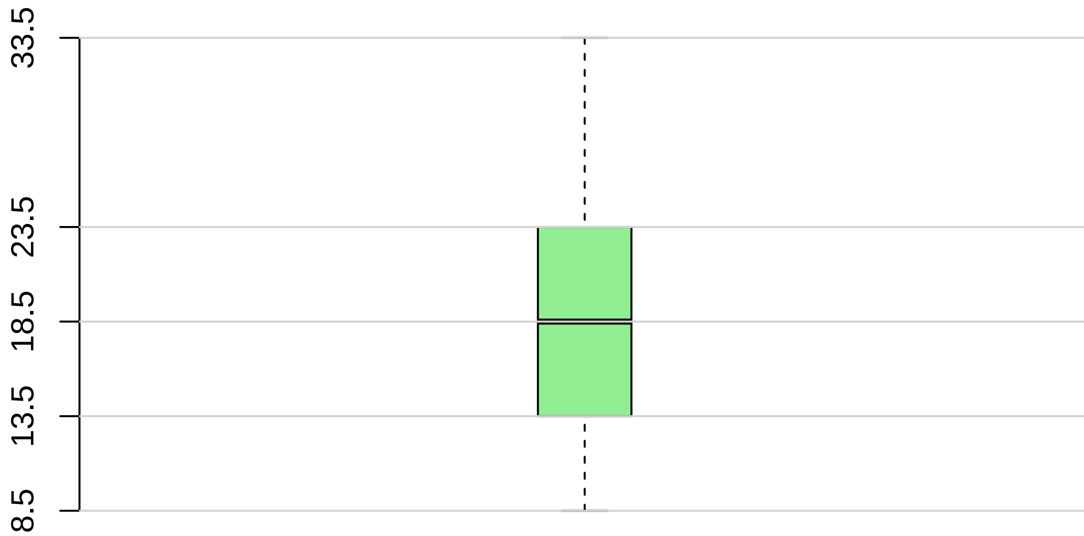
```
generic.breaks <- c(1, 6, 11, 16, 21, 26, 31, 36)
generic.breaks.length <- length(generic.breaks)
generic.c <- c(0, 4, 24, 40, 20, 4, 8)
generic.xi <- (generic.breaks[1:generic.breaks.length-1] + generic.breaks[2:generic.breaks.length]) / 2
generic.boxplot.data <- rep(generic.xi, times = generic.c)

generic.boxplot <- boxplot(generic.boxplot.data, plot = FALSE, range = 0)
generic.boxplot.stats <- c(generic.boxplot$stats)
names(generic.boxplot.stats) <- c("Minimum", "p0.25", "Median", "p0.75", "Maximum")

generic.xi
```

```
## [1]  3.5  8.5 13.5 18.5 23.5 28.5 33.5
```

```
generic.boxplot.stats
```

```
## Minimum   p0.25  Median   p0.75 Maximum
##     8.5    13.5    18.5    23.5    33.5
```

```
boxplot(generic.boxplot.data, col="lightgreen", axes = FALSE, boxwex=0.2, range = 0)
axis(2, at = generic.boxplot.stats)
abline(h = generic.boxplot.stats, col = "lightgray")
```



Aufgabe 29

```
rats.a <- c(99, 103, 106, 93, 98)
rats.b <- c(9.6, 10.2, 10.1, 9.7, 11.6)
rats.a.mean <- sum(rats.a) / length(rats.a)
rats.b.mean <- sum(rats.b) / length(rats.b)
rats.a.var <- sum((rats.a - rats.a.mean)^2) / length(rats.a)
rats.b.var <- sum((rats.b - rats.b.mean)^2) / length(rats.b)
rats.a.sd <- sqrt(rats.a.var)
rats.b.sd <- sqrt(rats.b.var)
rats.a.cv <- rats.a.sd / rats.a.mean
rats.b.cv <- rats.b.sd / rats.b.mean

names(rats.a.cv) <- c("Cv [%] for A")
names(rats.b.cv) <- c("Cv [%] for B")

c(rats.a.cv, rats.b.cv) * 100
```

```
## Cv [%] for A Cv [%] for B
##     4.454130     7.004071
```

Aufgabe 30

```
pupils.original <- rep(74, times = 28)
pupils.after <- c(rep(74, times = 27), 50)

pupils.result <- c(mean(pupils.original), mean(pupils.after))
pupils.result <- c(pupils.result,
                   round(pupils.result[2] / pupils.result[1] * 100, 2))
names(pupils.result) <- c("Davor", "Danach", "Änderung [%]")
```

Durchschnittsgewicht [kg]

```
pupils.result
```

```
##        Davor       Danach Änderung [%]
##     74.00000     73.14286     98.84000
```

Aufgabe 31

```r
radio.breaks <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
radio.2000 <- c(5, 3, 10, 9, 13, 18, 21, 27, 10, 5)
radio.2018 <- c(35, 24 ,13, 8, 9, 4, 2, 0, 0, 2)
radio.xi <- (radio.breaks[1:length(radio.breaks)-1] +
               radio.breaks[2:length(radio.breaks)]) / 2
radio.2000.mean <- sum(radio.2000 * radio.xi) / sum(radio.2000)
radio.2018.mean <- sum(radio.2018 * radio.xi) / sum(radio.2018)

radio.2000.median <- groupedQuantile(radio.breaks, radio.2000, 0.5)
radio.2018.median <- groupedQuantile(radio.breaks, radio.2018, 0.5)

radio.2000.var <- sum((radio.2000 * radio.xi - radio.2000.mean)^2) / sum(radio.2000)
radio.2018.var <- sum((radio.2018 * radio.xi - radio.2018.mean)^2) / sum(radio.2018)

radio.2000.sd <- sqrt(radio.2000.var)
radio.2018.sd <- sqrt(radio.2018.var)

radio.2000.quater <- groupedQuantile(radio.breaks, radio.2000, 0.75)
radio.2018.quater <- groupedQuantile(radio.breaks, radio.2018, 0.75)

radio.2000.boxplot.data <- c(
  min(radio.xi), max(radio.xi), radio.2000.median,
  groupedQuantile(radio.breaks, radio.2000, 0.25),
  groupedQuantile(radio.breaks, radio.2000, 0.75)
)

radio.2018.boxplot.data <- c(
  min(radio.xi), max(radio.xi), radio.2018.median,
  groupedQuantile(radio.breaks, radio.2018, 0.25),
  groupedQuantile(radio.breaks, radio.2018, 0.75)
)

radio.2000.boxplot.stats <- boxplot(
  radio.2000.boxplot.data, plot = FALSE, range = 0)$stats
radio.2018.boxplot.stats <- boxplot(
  radio.2018.boxplot.data, plot = FALSE, range = 0)$stats

radio.data <- data.frame(
  "Radio 2000" = c("Mittelwert" = radio.2000.mean,
                   "Median" = radio.2000.median,
                   "Modale Klasse" = 8,
                   "SD" = radio.2000.sd,
                   "25% liegen über" = radio.2000.quater),
  "Radio 2018" = c("Mittelwert" = radio.2018.mean,
                   "Median" = radio.2018.median,
                   "Modale Klasse" = 1,
                   "SD" = radio.2018.sd,
                   "25% liegen über" = radio.2018.quater)
)
```

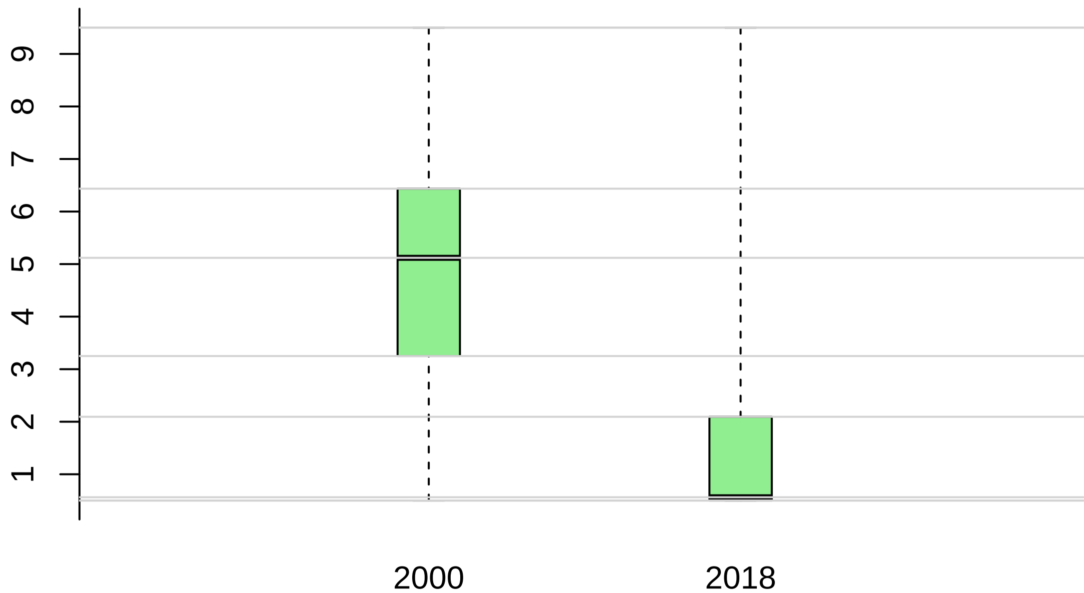
a, b, c, e)

```r
radio.data
```

```
##                   Radio.2000 Radio.2018
## Mittelwert          5.723140   2.149485
## Median              5.119048   0.562500
## Modale Klasse       8.000000   1.000000
## SD                 25.125757   7.354512
## 25% liegen über     6.435185   2.093750
```

d)

```r
boxplot(radio.2000.boxplot.data, col="lightgreen", axes = FALSE, boxwex=0.4,
        range = 0, xlim = c(1, 4), at=2)
boxplot(radio.2018.boxplot.data, col="lightgreen", boxwex=0.4, axes = FALSE,
        range = 0, add = TRUE, at=3)
axis(1, at = c(2, 3), labels = c("2000", "2018"), tick = FALSE)
axis(2, at = 0:(max(c(radio.2000.boxplot.stats, radio.2018.boxplot.stats)) + 1))
      #c(radio.2000.boxplot.stats, radio.2018.boxplot.stats))
abline(h = c(radio.2000.boxplot.stats, radio.2018.boxplot.stats), col = "lightgray")
```



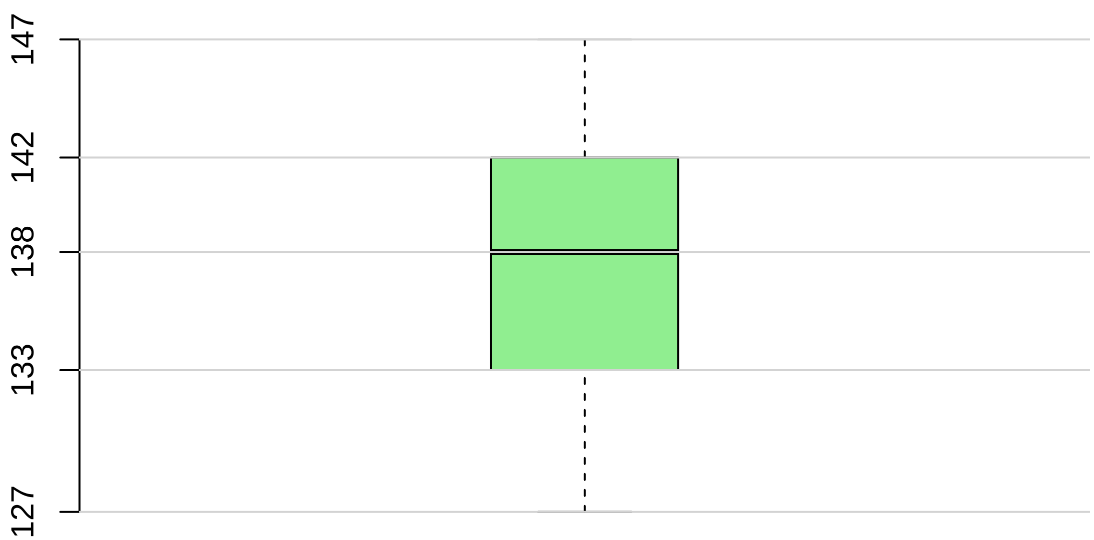
Aufgabe 32

```r
haemoglobin.min <- 127
haemoglobin.max <- 147
haemoglobin.firstquarter <- 133
haemoglobin.thirdquarter <- 142
haemoglobin.median <- 138

haemoglobin.data <- data.frame(
  "Hämoglobin" = c(
    "Min" = haemoglobin.min,
    "Max" = haemoglobin.max,
    "Median" = haemoglobin.median,
```

```r
    "25%" = haemoglobin.firstquarter,
    "75%" = haemoglobin.thirdquarter,
    "60% liegen über" = 136
  )
)

haemoglobin.boxplot.data <- c(
    "Min" = haemoglobin.min,
    "Max" = haemoglobin.max,
    "Median" = haemoglobin.median,
    "25" = haemoglobin.firstquarter,
    "75" = haemoglobin.thirdquarter
)

haemoglobin.data
```

```
##                   Hämoglobin
## Min                      127
## Max                      147
## Median                   138
## 25%                      133
## 75%                      142
## 60% liegen über          136
```

```r
boxplot(haemoglobin.boxplot.data, axes = FALSE, range = 0, boxwex = 0.4, col = "lightgreen")
axis(2, at=haemoglobin.boxplot.data)
abline(h = haemoglobin.boxplot.data, col = "lightgray")
```

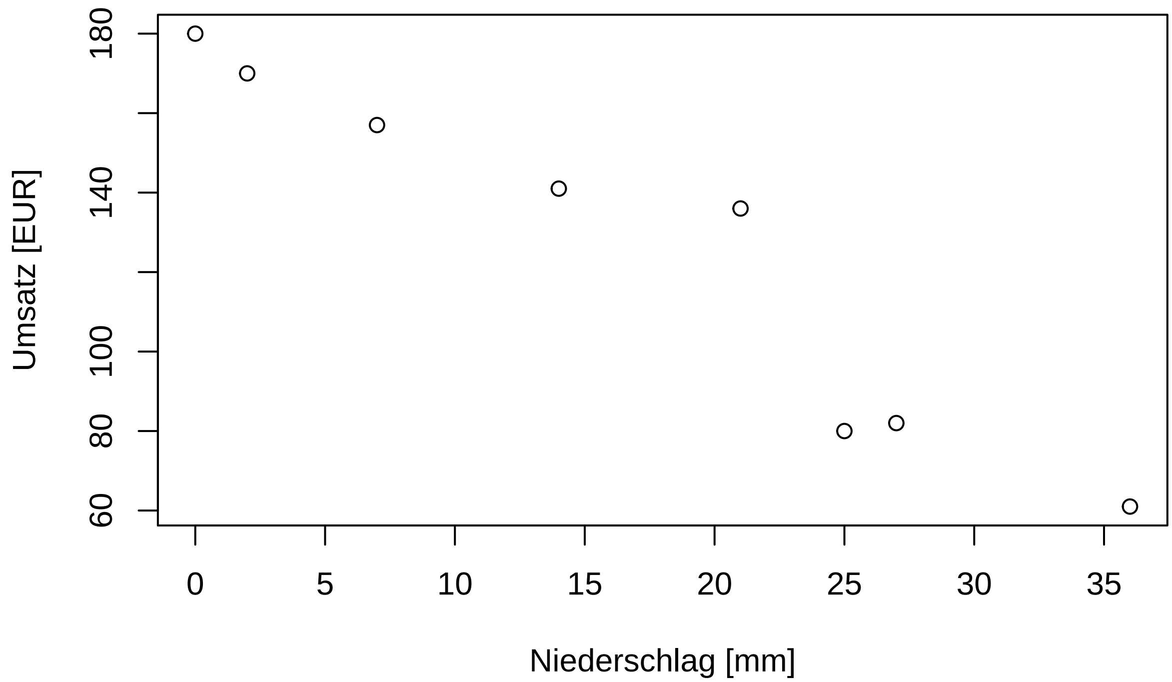


Aufgabe 33

```r
bath.rain <- c(0, 7, 14, 21, 25, 27, 36, 2)
bath.turnover <- c(180, 157, 141, 136, 80, 82, 61, 170)

bath.rain.mean <- mean(bath.rain)
bath.turnover.mean <- mean(bath.turnover)
```

```r
bath.rain.var <- NVar(bath.rain)
bath.turnover.var <- NVar(bath.turnover)

bath.rain.sd <- Nsd(bath.rain)
bath.turnover.sd <- Nsd(bath.turnover)

bath.cov <- NCov(bath.rain, bath.turnover)
names(bath.cov) <- c("Kovarianz")

bath.r <- bath.cov / (bath.rain.sd * bath.turnover.sd)
names(bath.r) <- c("Korrelationskoeffizient")
```

```r
plot(bath.rain, bath.turnover,
     xlab = "Niederschlag [mm]",
     ylab = "Umsatz [EUR]")
```



Je mehr es regnet, desto weniger Personen gehen schwimmen und desto weniger Eis wird verkauft. Daher fällt der Umsatz je mehr Niederschlag vorhanden ist.

```r
c(bath.cov, bath.r)
```

```
##               Kovarianz Korrelationskoeffizient
##            -492.0625000              -0.9618376
```

Aufgabe 34

```r
salary.mean <- 1029
salary.6.mean <- 1234
salary.var <- 927
salary.6.var <- 1519
salary.cov <- 810

salary.b <- salary.cov / salary.var
```

```r
salary.a <- salary.6.mean - salary.b * salary.mean

paste("Intercept: ", salary.a)
```

```
## [1] "Intercept:  334.873786407767"
```

```r
paste("Slope: ", salary.b)
```

```
## [1] "Slope:  0.87378640776699"
```

```r
result <- salary.a + salary.b * 1380
paste(result)
```

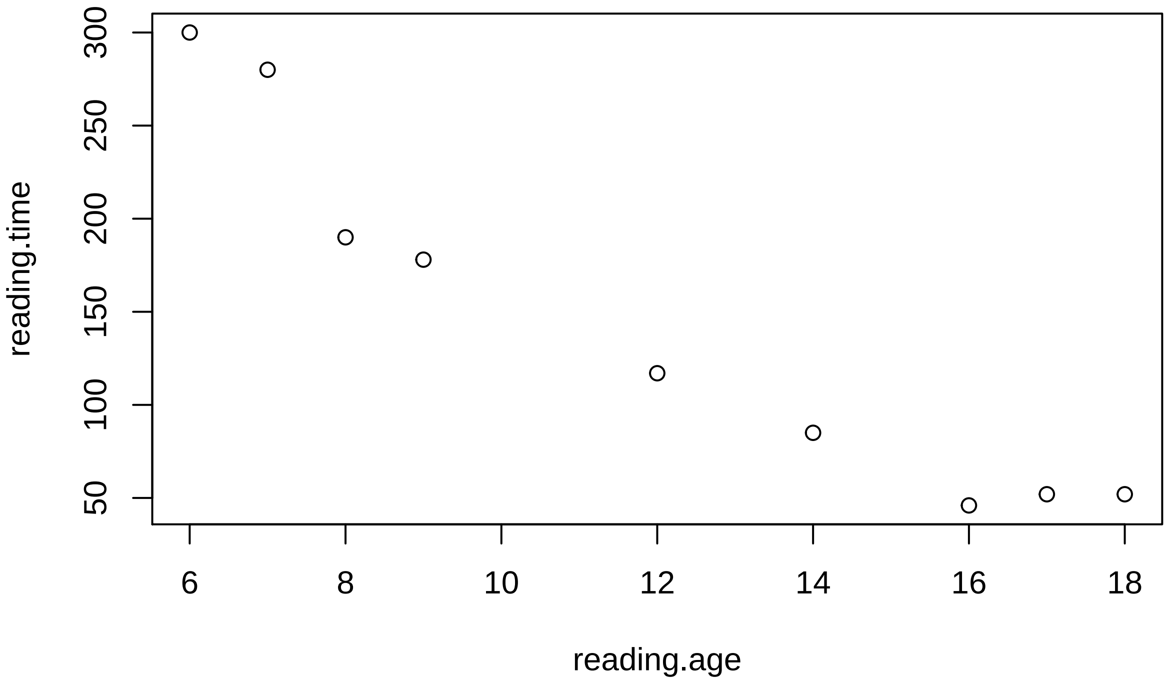
```
## [1] "1540.69902912621"
```

```r
paste(round(result), "€")
```

```
## [1] "1541 €"
```

Aufgabe 35

```r
reading.age <- c(7, 6, 8, 9, 14, 12, 17, 16, 18)
reading.time <- c(280, 300, 190, 178, 85, 117, 52, 46, 52)

reading.r <- NCov(reading.age, reading.time) / Nsd(reading.age)
reading.age.var <- NVar(reading.age)
reading.age.b <- NCov(reading.age, reading.time) / reading.age.var
reading.age.a <- mean(reading.time) - reading.age.b * mean(reading.age)
reading.age.predict <- function(age){
  return(
    reading.age.a + reading.age.b * age
  )
}

plot(reading.age, reading.time)
abline(lm(reading.age ~ reading.time))
```



Korrelationskoeffizient

```r
reading.r
```

```
## [1] -88.33695
```

Lesezeit mit 56 Jahren

```r
reading.age.predict(56)
```

```
## [1] -760.4501
```

Basierend auf dem Regressionsmodell würde ein 56 jähriger Mensch -760 Sekunden brauchen. Man kann daher Annehmen, dass die reale Funktion asymptotisch gegen 0 verläuft. D.h. eine 56 jährige Person braucht nur wenige Sekunden für den einfach gehaltenen Text.

Aufgabe 36

```r
coffee.A <- c(8, 9, 8, 10, 6)
coffee.B <- c(7, 5, 7, 10, 6)
coffee.cov <- NCov(coffee.A, coffee.B)
coffee.A.sd <- Nsd(coffee.A)
coffee.B.sd <- Nsd(coffee.B)
coffee.r <- coffee.cov / (coffee.A.sd * coffee.B.sd)
c("Bravais-Pearson r" = round(coffee.r, 2),
    "A" = paste(coffee.A, collapse = ","),
    "B" = paste(coffee.B, collapse = ","))
```

```
## Bravais-Pearson r                 A                 B
##            "0.54"      "8,9,8,10,6"      "7,5,7,10,6"
```

```r
coffee.A <- rank(c(8, 9, 8, 10, 6))
coffee.B <- rank(c(7, 5, 7, 10, 6))
coffee.cov <- NCov(coffee.A, coffee.B)
coffee.A.sd <- Nsd(coffee.A)
```

```r
coffee.B.sd <- Nsd(coffee.B)
coffee.r <- coffee.cov / (coffee.A.sd * coffee.B.sd)
c("Spearman r" = round(coffee.r, 2),
    "A" = paste(coffee.A, collapse = ","),
    "B" = paste(coffee.B, collapse = ","))
```

```
##      Spearman r               A               B
##          "0.37" "2.5,4,2.5,5,1" "3.5,1,3.5,5,2"
```

Es herrscht eine schwache Korrelation, obwohl die Bewertungen ähnlich sind. Es sind vermutlich zu wenige Daten vorhanden.

Aufgabe 37

```r
speed.kmh <- c(130, 100, 110, 120, 130)
death <- c(4.1, 4.7, 4.3, 5.1, 6.1)
speed.mph <- speed.kmh * 0.621371

speed.kmh.r <- NCov(speed.kmh, death) / (Nsd(speed.kmh) * Nsd(death))
speed.mph.r <- NCov(speed.mph, death) / (Nsd(speed.mph) * Nsd(death))
```

Geschwindigkeiten [km/h]

```r
speed.kmh
```

```
## [1] 130 100 110 120 130
```

R

```r
speed.kmh.r
```

```
## [1] 0.3290597
```

Geschwindigkeiten [mph]

```r
speed.mph
```

```
## [1] 80.77823 62.13710 68.35081 74.56452 80.77823
```

R

```r
speed.mph.r
```

```
## [1] 0.3290597
```

Es liegt eine sehr geringe bis keine lineare Korrelation vor