Introduksjon til R

STK1110

G. Storvik

Matematisk institutt
Universitet i Oslo

Høst 2023

Outline

Hva er R?

- Et program for å utføre statistisk analyse
- Grafisk verktøy
- Programmeringsspråk
  - Høynivå
  - Objektorientert
- Kommandolinjebasert
- Mulig med meny-styring (Rstudio)
- Gratis
- Også sky-baserte muligheter

Hvordan starte R

- Linux
  - Gi kommandoen `R` i shell-vindu
  - Gi kommandoen `ESC-x R` i emacs
- Windows
  - Klikk på R-ikon
  - Finn R under **programmer**
- Starter opp et interaktivt program

Enkle kommandoer.

- `rnorm(15)`

```
 [1] -1.40965671 -1.34995532 -0.66079685 -1.06171239 -1.51795142 -0.86903622
 [7]  2.41334598  0.39782654  0.17715629 -0.35065906 -0.08771934  1.21074545
[13]  1.19167291  0.82251500 -1.48620141
```

- `plot(rnorm(1000))`

Tilordning.

- Tilordning av verdi 2 til *variabel* `x`:

```
x <- 2
x = 2
```

- Gir samme resultat
- Vektorer
  - `weight = c(60,72,57,90,95,72)`
  - `height = c(1.75,1.80,1.65,1.90,1.74,1.91)`
  - `bmi = weight/height^2`
  - `xbar = sum(weight)/length(weight)`
  - `weight - xbar`

Grafikk

- `plot(height,weight)`
- `hh = seq(1.65,1.90,0.05)`
- `lines(hh,22.5*hh^2)`
- `boxplot(height)`

Statistikk

- `x = rnorm(50)`
- `mean(x)`
- `sd(x)`
- `var(x)`
- `quantile(x)`
- `summary(x)`

Ett-utvalgs T-test

```
d=c(5260,5470,5640,6180,6390,6515,
    6805,7515,7515,8230,8770)
```

- `t.test(d,mu=7725)`

```
One Sample t-test

data:  d
t = -2.8208, df = 10, p-value = 0.01814
alternative hypothesis: true mean is not equal to 7725
95 percent confidence interval:
 5986.348 7520.925
sample estimates:
mean of x
 6753.636
```

To-utvalgs T-test

- `data(energy)`
- `t.test(expend~stature,data=energy)`

```
Welch Two Sample t-test

data:  expend by stature
t = -3.8555, df = 15.919, p-value = 0.001411
alternative hypothesis: true difference in means is
95 percent confidence interval:
 -3.459167 -1.004081
sample estimates:
 mean in group lean mean in group obese
          8.066154           10.297778
```

Sammenlikning av varianser

- `var.test(expend~stature,data=energy)`

```
F test to compare two variances

data:  expend by stature
F = 0.7844, num df = 12, denom df = 8, p-value = 0.
alternative hypothesis: true ratio of variances is
95 percent confidence interval:
 0.1867876 2.7547991
sample estimates:
ratio of variances
          0.784446
```

Enkel regresjon

```
> data(thuesen)
> lm(short.velocity~blood.glucose,data=thuesen)
Call:
lm(formula = short.velocity ~ blood.glucose, data =

Coefficients:
  (Intercept)  blood.glucose
      1.09781         0.02196
```

Trekke ut informasjon

```
> fit = lm(short.velocity~blood.glucose,data=thuese
> names(fit)
 [1] "coefficients"  "residuals"     "effects"
 [5] "fitted.values" "assign"        "qr"
 [9] "na.action"     "xlevels"       "call"
[13] "model"

> fit$coef
  (Intercept) blood.glucose
   1.09781488    0.02196252
```

Objektorientering

```
> d = c(5260,5470,5640,6180,6390,6515,
        6805,7515,7515,8230,8770)
> fit = lm(short.velocity~blood.glucose,
           data=thuesen)
> summary(d)+
> summary(fit)+
> plot(d)+
> plot(fit)+
```

Reading data

- Direkte inntasting

```
> weight = c(60,72,57,90,95,72)
> height = c(1.75,1.80,1.65,1.90,1.74,1.91)
```

- Innlesning fra fil (hver variabel som kollonne)

```
> d = matrix(scan("foo.dat"),ncol=2,byrow=T)
> d = read.table("foo.dat",
                 col.names=c("weight","height")
```

Kjøring av skript

- Fil `foo.R`

```
d = read.table("foo.dat",
               col.names=c("weight","height"))
fit = lm(weight~height,data=d)
plot(fit)
```

- `source("foo.R")`

Biblioteker

- Mange tilleggsbiblioteker
- `search()`
- `library(Rcmdr)`
- `search()`
- `install.packages("ISwR")`
- `install.packages("abind",lib="~/Rlib")`
- `library(abind,lib.loc="~/Rlib")`

Rcmdr

- Grafisk grensesnitt til R
- Kan utføre “vanlige” kommandoer med menyer
- Viser frem kommandoene