

# R Kodblad

*av Sara Kalucza*

*vt2016*

I detta kodblad finns användbara koder att för att göra enklare analyser i R eller R studio. R kan ni ladda ner här och Rstudio kan ni ladda ner här (välj open source edition).

Syftet med detta kod blad är att ni, efter att ni kommit fram till vilken typ av analys ni vill göra, lätt ska kunna hitta rätt R-kod. Vi i kursen uppmuntrar till att använda R och Rstudio. Det bästa sättet att lära sig är att skriva av koden, men ni kan också kopiera och klistra in.

**Notera:** Eftersom detta är ett *exempelblad* måste ni själva byta ut namn på data och variabler för att koderna ska fungera. För dessa exempelkoder kallar jag min dataframe (mitt datamaterial) *d* och mina variabler var1, var2 osv. alternativet varY, varX när beroende/oberoende är av vikt.

## Grundläggande funktioner

- Installera paket (endast första gången ni ska använda ett paket):

```
install(package)
```

- Hämta in ett paket för att arbeta med det (varje ny session)

```
library(package)
```

- Få tillgång till dokumentation/hjälp om ett paket

```
?package
```

- Ange filken mapp du vill hämta en fil ifrån (exempelvis datamaterial):

```
setwd("~/Documents/Mina filer/krimmetod")
```

Detta kan du även göra i menyn (Rstudio), via Session -> Set working directory

- Ladda in datamaterial (redan sparat i R-format):

```
d <- load(dataname)
```

- Ladda in .csv fil

```
d <- read.csv(file="data.csv", head=TRUE, sep=";")
```

## Studera data / Deskriptiva kommandon

- Ger dig en lista över de variabler som finns i din dataframe:

```
names(d)
```

- Inspektera dataframe (öppnas i ny flik)

```
view(d)
```

- Deskriptiv statistik för en enskild variabel.

```
summary(d$var)
```

- Medelvärde för variabel 1, uppdelat i grupper via variabel 2 (ex. medelinkomst (var 1) uppdelat på kön (var 2))

```
tapply(d$var1, data$var2, mean)
```

- Skapa en frekvenstabell (med antal). t blir namnet på tabellen. Ni kan välja detta helt själva.

```
t <- table(d$var)
t
```

Den första raden skapar tabellen, och den andra (*t2*) visar den.

- Skapa tabell med procent

```
tprop <- table(d$var)
prop.table(tprop)*100
```

- Tvåvägs tabell Rad ett skapar tabellen, rad två visar den som frekvenstabell, rad tre visar i procent, där 1an i (t2,1) representerar på vilken kolumn procenten ska beräknas.

```
t <- table(d$var1, d$var2)
t
prop.table(t,1)*100
```

### Enklare statistiska test / Inferens

- Konfidensintervall

```
t.test(d$var1)
```

- T-test av skillnad i medelvärden, oberoende grupper

```
t.test(d$var1~data$var2)
```

Där var2 är den kategoriska variabeln

- T-test av skillnad i medelvärden, beroende grupper dvs. samma individer, observerade vid två tidpunkter

```
t.test(d$var1,d$var2, paired=TRUE)
```

- Chi2-test / Chi-square test Chi2 test måste göras i två steg: först skapar man en tabell, och sedan testar man om frekvenserna skiljer sig statistiskt från varandra.

```
t <- table(d$var1, d$var2)
chisq.test(t)
```

### Regressionsanalys

- Enkel linjär regressionsanalys Den första raden skapar modellen (m1), och den andra raden ger oss koefficienterna från modellen.

```
m1 <- lm(varY~varX, data=d)
summary(m1)
```

- Multipel linjär regressionsanalys

```
m1 <- lm(varY~varX1+varX2, data=d)
summary(m1)
```

- Logistisk regressionsanalys

```
logm1 <- glm(formula = varY ~ varX, family = binomial(link = "logit"), d = train)
summary(logm1)
```