Datorlaboration 4

Josef Wilzén och Måns Magnusson

12 februari 2016

Instruktioner

- Denna laboration ska göras **en och en**.
- Det är tillåtet att diskutera med andra, men att plagiera eller skriva kod åt varandra är **inte tillåtet**.
- Deadline för labben framgår på **LISAM**
- Laborationen ska lämnas in via **LISAM**.
- Använd inte `å`, `ä` eller `ö` i variabel- eller funktionsnamn.
- Utgå från laborationsfilen, som går att ladda ned **här**, när du gör inlämningsuppgifterna. Spara denna som labb[no]_[liuID].R , t.ex. labb1_josad732.R om det är labb 1. Ta inte med hakparenteser i filnamnet. Denna fil ska laddas upp på LISAM och ska **inte** innehålla något annat än de aktuella funktionerna, namn- och ID-variabler och ev. kommentarer. Alltså **inga** andra variabler, funktionsanrop för att testa inlämningsuppgifterna eller anrop till markmyassignment-funktioner.
- Laborationen består av två delar:
  - Datorlaborationen
  - Inlämningsuppgifter
- I laborationen finns det extrauppgifter markerade med *. Dessa kan hoppas över.

Innehåll

Del I

Datorlaboration

Kapitel 1

R-paket

R-paket är extra moduler/biblitek som läses in i R för att skapa extra funktionalitet i form av nya funktioner eller nya data. De flesta funktioner som används i R finns i olika paket. Några få paket läses automatiskt in i R när vi startar R, medan andra paket måste vi läsa in aktivt för att få tillgång till funktionaliteten. Den stora mängd personer som bidrar till R gör det genom att utveckla nya funktioner som de sedan släpper som paket.

Paket är något som skiljer R från andra statistikprogram är att den mesta funktionaliteten inte kommer med från början. I andra programmeringsspråk är denna form av **modularisering** betydligt vanligare. Den stora fördelen med detta är att vi bara behöver läsa in de paket vi verkligen har behov av just nu.

För att kunna använda ett paket behöver vi gå igenom två steg:

- Paketet måste först installeras på din aktuella dator.
- Paketet måste sedan läsas in i den aktuella sessionen för att användas - eller anropas explicit.

Alla paket har olika versioner och generellt följer de kriterierna för **semantisk versionshantering**.

1. Först måste vi installera ett paket. Detta kan antingen göras genom CRAN (Comprehensive R Archive Network) på internet där de flesta paket ligger uppe. Detta görs med funktionen `install.packages()`. Pröva att installera `stringr` och `lubrudate` på detta sätt.

```
install.packages("stringr")
install.packages("lubridate")
# eller om du ska installera något paket i PC-pularna:
install.packages("stringr",lib="sökväg till en mapp i din hemkatalog")
install.packages("lubridate",lib="sökväg till en mapp i din hemkatalog")
```

2. För att se vilka paket som är installerade så körs:

```
installed.packages()
# eller
x<-installed.packages()
View(x)
```

3. En annan server där det finns mycket paket är på github. Det är vanligt att paket som fortfarande utvecklas aktivt finns på både CRAN och github.com då github underlättar enormt för så kallad collaborativ utveckling där flera personer hjälps åt med utvecklingen. För att installera från github direkt behöver först paketet `devtools` installeras. Pröva att installera paketet `pxweb` på detta sätt med följande kod.

```
install.packages("devtools")
devtools::install_github("pxweb","rOpenGov")
library(pxweb)
```

4. För att läsa in ett paket (d.v.s. för att använda paketet i den aktuella sessionen) används funktionen `library()`.

```
library(pxweb)
# eller om du har installerat ett paket på en egenvald plats:
library(markmyassignment,lib="sökväg till en mapp i din hemkatalog")
```

5. Om flera paket har samma funktionsnamn kan vi bestämma exakt från vilket paket vi ska använda en given funktion med `::`. Då behöver vi inte först läsa in paketet. Detta är särskilt bra om vi bara vill använda en enskild funktion från ett paket. Vi kan då snabbt se i vår kod var paketet används.

```
lubridate::ymd("19990101")
```

6. För att ta bort ett paket från en aktuell R-session används `detatch()`.

```
detach("package:pxweb", character.only = TRUE)
```

7. I R-Studio kan vi studera vilka paket som finns installerade och vilka som är inlästa i R under "Packages". Här kan vi också lägga till eller ta bort paket från vår aktiva session. Undersök om du har `ggplot2` installerat och vilken version du har av `ggplot2`. Om `ggplot2` inte är installerat pröva att installera det.
8. Ta reda på hur paket kan avinstalleras i R genom att söka på webben eller i R:s dokumentation.

Kapitel 2

Mer om funktioner

Funktioner är en central del i R. Allt som "gör" något i R är en funktion och allt som "är" något är ett objekt. Av detta följer att varje enskild funktion är ett objekt i sig.

För en mer fördjupad om det som gås igenom här rekommenderas **kapitlet om funktioner** i *Advanced R programming* av Hadley Wickham.

2.1 Tilldelning

Tilldelning (assignment) kan ske på några olika sätt i R. Ni har redan använd det vanligste "`<-`".

1. För att tilldela i den närmste lokala miljön ska "`<-`" användas, men det finns varianter. Testa koden nedan. "`->`" bör inte användas. "`=`" används för att definiera defaultargument i funktionsdefinitioner och när funktioner anropas.

```
x<-1:4
2:6->y
z=5:9
?"<-"
```

2. För att tilldela i den globla miljön används "`<<-`" eller "`->>`". Testa koden nedan. Ofta vill vi undvika att tilldela variabler i den globla miljön inifrån funktioner, och därför är det bäst att använda den vanliga "`<-`".

```
rm(list=ls())
y<-"hej!"
y
h<-funciton(x,y){
x<-x+10
y<<-y+20
print(x)
print(y)
}
h(2,3)
y
```

3. En annan tilldeningsoperator är `assign()`, testa att köra `?assign()`. Testa koden nedan.

```
assign("x",12)
assign("abc",TRUE)
assign("myVar",rep("hej",10))
```

4. `assign()`kan användas för att skapa nya variabler med hjälp av elementen i en lista. Testa koden nedan.

```
rm(list=ls())
ls()
my_list<-list(x1=c(1,2,3),x2=matrix(1:4,2,8),x3=sin(pi^2),x4=LETTERS)
no_elements<-length(my_list)
for(i in 1:no_elements){
  assign(x = names(my_list)[i],value = my_list[[i]])
}
ls()
```

2.2 Delar i en R-funktion

Varje funktion består av tre huvudsakliga delar. Funktionens argument (eller formals), kropp (body) och lokala miljö (environment). Det är dessa tre som tillsammans utgör en funktion.

1. Vi börjar med att skapa följade funktion i R.

$$f(x, y) = x^2 + y^2 - 1$$

```
f <- function(x,y){
        fxy <- x^2 + y^2 - 1
        return(fxy)
}
```

2. Vi kan nu använda oss av funktionen `formals()` för att plocka ut funktionens argument.

```
formals(f)
```

3. På samma sätt kan vi studera funktionens kropp med `body()`.

```
body(f)
```

4. Till sist kan vi också undersöka funktionens miljö. Detta säger ingeting just nu, men vi kommer strax återkomma till funktioners lokala miljöer.

```
environment(f)
```

2.2.1 Primitiva funktioner

Det finns ett undantag från dessa regler och det gäller primitiva funktioner. Dessa funktioner är skrivna direkt i C-kod och anropar C-kod direkt. Dessa funktioner har varken `environment()`, `body()` eller `formals()`.

```
c

function (..., recursive = FALSE)  .Primitive("c")
```

```
formals(c)

NULL

body(c)

NULL

environment(c)

NULL
```

2.3 Krav på funktioner

Vi börjar med att se vilka delar som är nödvändiga för en funktion i R. Vi börjar med följande enkla funktion.

```
g <- function(a,b){
        c <- a + b
        return(c)
}
```

1. Funktioner behöver inte ha ett `return()`-steg. Om returnsteget saknas måste det som ska returneras "skrivas ut". Se exemplet nedan:

   ```
   g <- function(a,b){
           c <- a + b
           c
   }
   ```

2. Det gör att vi kan förenkla funktionen ytterligare på följande sätt. Pröva att funktionen fortfarande fungerar.

   ```
   g <- function(a,b){
           a + b
   }
   ```

3. Det är också möjligt att uttrycka en funktion på en och samma rad. Då behöver vi inte heller {}. Se exempel nedan:

   ```
   g <- function(a,b) a + b
   ```

4. Pröva att använda `formals()` och `body()` på funktionen `g()` ovan.

2.4 Defaultvärden och argumentordning

I vissa fall kan det vara så att vi vill att ett givet värde ska vara det värde som används som standardvärde för en särskild funktion. Exempelvis funktionen `mean()` har argumentet `rm.na` satt till `FALSE` som standard.

För att skapa ett standardvärde använder ställer vi in detta i funktionsdefinitionen på följande sätt `function(a, b=10)`. Vi kan i princip ha vad vi vill som standardvärden.

1. Nedan är ett exempel på standardvärden i funktioner.

```
g <- function(a, b = 10){
        res <- a + b
        return(res)
}
```

2. Pröva att köra funktionen både genom att ange argumentet `b` och utan att ange det. Upprepa funktionen men sätt standarvärdet för `a` till 5 och `b` till 15. Pröva att anropa funktionen utan att ange några värden alls.

3. Pröva att använda funktionen `formals()` på funktionen `g()` ovan. Framgår defaultvärdet?

4. Skapa en ny funktion på följande sätt:

$$h(x, y) = x^y - y$$

där y sätts till 1 som standard.

5. När vi anropar en funktion kan vi välja att ange namnet på argumentet eller inte. Anger vi namnet på argumenten så spelar ordningen ingen roll. Anger vi däremot inte argumentens namn utgår R från att argumenten följer samma ordning som vi skapade argumenten i (och som vi ser med `formals()`)

6. Pröva lite olika värden på `x` och `y`. Pröva att använda argumentnamnen och byt ordning på `x` och `y` i funktionsanropet (d.v.s `h(y=10, x=100)`).

7. Pröva följande kod. En funktion kan bara ha ett argument med ett givet argumentnamn.

```
k <- function(a, a = 10){
res <- a^2
return(res)
}
```

2.5 Funktioner i funktioner

Ibland kan det vara så att vi vill ge en hel funktion som ett argument till en annan funktion. Ett exempel på detta är om vi vill integrera en funktion numeriskt.

1. Vi ska nu pröva att beräkna en integral numeriskt i R. Börja med att skapa följande funktion i R och kalla den för `f`.

$$f(x) = \frac{1}{3}x^2$$

2. För att integrera numeriskt i R använder vi funktionen `integrate()`. Vi behöver då ange funktionen vi vill integrera genom att ange denna funktion som ett argument till `integrate()`. Vi behöver också ange från vilka värden vi vill utföra integralen. Exempelvis följande integral:

$$\int_0^3 f(x)dx = \int_0^3 \frac{1}{3}x^2 dx$$

kan beräknas på följande sätt i R:

```
integrate(f=f, 0, 3)

3 with absolute error < 3.3e-14
```

3. Pröva nu att beräkna följande integral för `f`

$$\int_{-3}^{9} f(x)dx$$

4. Vi kan också skicka med argument till den funktion vi skickar med. Skapa täthetfunktionen för en exponentialfördelad variabel med argumenten `x` och `lambda` och kalla den `exp_pdf()` i R. Täthetsfunktionen ges nedan:

$$f_X(x, \lambda) = \lambda e^{-\lambda x}$$

5. För att skicka med både en funktion och argument som ska skickas vidare används `...` i funktionen `integrate()`. För att beräkna följande integral

$$\int_{0}^{1} f_X(x, \lambda = 1)dx$$

gör vi på följande sätt:

```
integrate(f=exp_pdf, 0, 1, lambda = 1)

0.63212 with absolute error < 7e-15
```

2.5.1 Ellipsis (...)

Ovan var ett exempel på `...` som kallas ellipsis. Ellipsis är ett sätt att kunna skicka ett godtyckligt antal argument (och godtyckligt namngivna) till en funktion och sedan skicka vidare dessa till en ny funktion. På detta sätt behöver inte varje funktion vi konstruerar ta hänsyn till alla möjliga funktioner.

1. Använd hjälpen och titta på dokumentationen till funktionen apply(). apply() har dels ett argument `FUN` där vi anger en funktion vi använder och ... för att kunna skicka godtyckliga argument till den funktion vi angett under `FUN`.

2. Vi ska nu pröva att skapa en funktion på följande sätt.

```
apply_my_function_on_x <- function(x, FUN) FUN(x)
```

3. Pröva att skapa en numerisk vektor och prova lite olika funktioner som argument `FUN`. Ex. `mean()`, median() och kontrollera att det fungerar.

4. Pröva nu att byta ut ett element i din numeriska vektor till `NA` och pröva återigen lite funktioner som `mean()` och `median()`. Nu får vi `NA` som resultat och det finns inget sätt att skicka med `na.rm=TRUE` till ex. `mean()`. Antingen är `na.rm=TRUE` eller `na.rm=FALSE`. Det går inte att från vår funktion `apply_my_function_on_x()`.

5. Skapa nu följande funktion där vi använder `...` . Observera att vi behöver ange det både som argument i vår funktion och som argument i den funktion vi vill kunna skicka vidare argument till.

6. Pröva nu att styra `na.rm` i `mean()` direkt från vår `apply_my_function_on_x()`-funktion och kontrollera att resultaten fungerar som de ska.

2.5.2 do.call()

`do.call()` är en mycket kraftfull funktion när det gäller hantering av funktioner i funktioner.

1. Kör `?do.call`, testa sen koden nedan:

```r
# kolla upp vad argumentet heter i summary:
?summary
x<-list(object=iris)
do.call(what = summary,args = x)
g<-function(x,a1,a2,a3){
  y<-a1*x^2+a2*x+a3
return(y)
}
mylist<-list(x=10,a1=2,a2=-4,a3=10)
do.call(g,mylist)
mylist$x<-20
do.call(g,mylist)
```

2. Skriv en egen funktion med minst 2 argument som du testar att anropa med `do.call()`.
3. Testa nu den omgjorda funktionen av `apply_my_function_on_x`. Pröva nu att styra `na.rm` i `mean()` på vektorn `x<-1:300`.

```r
apply_my_function_on_args <- function(FUN,args) do.call(FUN,args)
```

4. Det blir lätt problem om vi har flera funktioner i huvudfunktionen som använder `...` , se exemplet nedan:

```r
apply_my_function_on_xy <- function(x,y, FUN1,FUN2, ...){
        x_new<-FUN1(x,...)
        y_new<-FUN2(y,...)
        return(list(x=x_new,y=y_new))
}
apply_my_function_on_xy(x = c(NA,12,64,NA,2),y=letters,FUN1 = mean,FUN2 = length)
apply_my_function_on_xy(x = c(NA,12,64,NA,2),y=letters,FUN1 = mean,FUN2 = length,na.rm=TRUE)
```

5. I fallet ovan är det bra att använda `do.call()`, testa den modifierade funktionen nedan:

```r
apply_my_function_on_xy <- function(FUN1,FUN2,arg1,arg2){
        x_new<-do.call(FUN1,args=arg1)
        y_new<-do.call(FUN2,args=arg2)
        return(list(x=x_new,y=y_new))
}
a<-list(x = c(NA,12,64,NA,14),na.rm=TRUE)
b<-list(x=letters)
apply_my_function_on_xy(FUN1 = mean,FUN2 = length,arg1=a,arg2=b)
```

2.6 Globala och lokala miljöer i R

Alla funktioner i R skapar egna lokala miljöer när funktionerna anropas där initialt bara argumenten finns. Fördelen med detta är att det inte finns några risker att olika objekt skulle krocka om de skulle ha samma namn. Ett bra sätt att tänka är att R startar en helt ny R-session varje gång en funktion anropas och att koden i funktionen körs i denna miljö. Det gör att våra variabler som vi har i den globala miljön inte påverkas och att vi behöver inte oroa oss för vad vi använder oss av för variabler inuti funktioner.

1. Kör följande kod. Vad förväntar du dig att ska hända?

```
mitt_x <- function(){
        x <- 15
        print("x:")
        print(x)
}
x <- 10
mitt_x()
x
```

En central del när det gäller funktioner (och objekt) i R är den så kallade sökvägen till objekt. Det handlar om hur R väljer vilken funktion eller objekt den ska returnera om vi anger ett objektsnamn.

R gör detta baserat på sökvägen till objektsnamnen. Med funktionen `search()` kan vi se hur R söker efter en funktion eller objekt om vi anger ett objektnamn.

```
search()
```

```
[1] ".GlobalEnv"        "package:knitr"     "package:stats"
[4] "package:graphics"  "package:grDevices" "package:utils"
[7] "package:datasets"  "Autoloads"         "package:base"
```

Det är i denna ordning som R kommer söka om vi exempelvis anropar funktionen `mean()`. Först kommer den se om det finns en funktion som heter `mean()` i den globala miljön. Hittar den inte denna funktion där kommer den gå vidare och leta efter `mean()` i de olika paketen i den ordning som anges ovan. Finns det ingen `mean()`-funktion i något paket kommer den tillslut titta i base-paketet där `mean()` ligger och anropa funktionen.

Nu återkommer vi till funktionen environment() som vi prövade tidigare. Med denna funktion kan vi se i vilken miljö en funktion har skapats.

Detta sätt att leta reda på funktioner kallas för **lexical scoping**.

1. Vi ska nu titta på funktionen `mean()`. Börja med att studera var denna funktion ligger men `environment()`.

   ```
   environment(mean)
   ```

2. Skapa följande funktion och pröva denna kod:

   ```
   mean <- function(x){
           100
   }
   ```

3. Den funktion vi brukar använda för att räkna ut medelvärden heter också `mean()`. Vilken funktion är det som används om du anropar funktionen `mean` nu för en numerisk vektor.

4. Pröva att se vilken miljö `mean()` nu ligger i.

5. Pröva nu att anropa den gamla `mean()`-funktionen med hjälp av `::` direkt från den namespace den gamla funktionen `mean()` ligger i.

   ```
   base::mean(1:10)
   mean(1:10)
   ```

6. Ta nu bort din `mean()`-funktion med `rm(mean)`. Pröva följande kod igen.

```
base::mean(1:10)
mean(1:10)
```

7. Skapa nu följande funktionen nedan. Vilken variabel är en så kallad fri variabel?

```
f <- function(x){
        (x + y)^2 - 1
}
```

2.6.1 Fria variabler och dynamic lookup

Är det så att en variabel används i en funktion men inte skapas i funktionen är det en så kallad fri variabel. I dessa fall kommer R söka efter denna variabel **i den miljö där funktionen skapades.**

R använder också så kallad dynamic lookup. Det innebär att R anropar fria variabler när funktionen körs, inte när funktionen skapas.

1. Kör denna kod i R: Pröva lite olika värden på `fri_variabel`.

```
fri_variabel <- 10
ny_fun <- function(){
  a <- c(1, 3, 5)
  b <- a + fri_variabel
  return(b)
}
ny_fun()

[1] 11 13 15

fri_variabel <- 10
ny_fun()

[1] 11 13 15

rm(fri_variabel)
ny_fun()

Error in ny_fun(): object 'fri_variabel' not found
```

Det vi sett ovan är hur R letar upp en funktion (eller objekt). När det gäller fria variabler i funktioner fungerar det på samma sätt.

- Först försöker R hitta variabeln i den lokala miljön för funktionen.
- Hittar R inte funktionen där letar den vidare i den miljö där funktionen skapades.
- Hittar den inte variabeln där fortsätter den upp till dess att den kommer till den globala miljön.
- Finns variabeln inte i den globala miljön söker den vidare i de inlästa paketen. Finns funktionen inte där så returnerar R ett felmeddelande om att den fria variabeln saknas.

1. Skapa koden nedan, vad gör funktionen? Använder den `y <- 5` eller `y <- 10`?

```r
y <- 5
g <- function(x){
  y <- 10

  f <- function(a) a^2

  print(environment(f))
  output <- f(y)
  return(output)
}
```

Kapitel 3

Högnivåfunktioner (*apply)

Vi ska nu använda så kallade de så kallade *apply-funktionerna i R. Dessa funktioner är så kallade högnivåfunktioner som vi använder om vi vill applicera en funktion mer generellt.

Det finns flera olika högnivåfunktioner.

Högnivåfunktion	Anropa funktionen FUN för...
apply()	varje rad eller kolumn i en matris
vapply()	varje element i en vektor
tapply()	varje grupp eller id
lapply()	varje element i en lista
mapply()	olika uppsättningar av argument till FUN

Exempelvis vapply() och lapply() blir på detta sätt ett alternativ till att använda loopar som ibland kan vara snabbare.

1. Vi börjar med funktionen tapply(). tapply() används för att använda en funktion per grupp (över en så kallad "Ragged array" eller vektorer av olika längd). Detta är ofta av intresse i praktiken. Vi börjar med att läsa in datsetet ChickWeight.

```
data(ChickWeight)
```

2. Vi ska nu pröva tapply() som har argumenten X, INDEX, FUN och simplify. X anger variabeln (eller datasetet) vi vill använda funktionen på, INDEX anger vilken gruppvariabel som ska användas och FUN anger vilken funktion som ska användas per grupp. Ett exempel på hur vi kan beräkna den genomsnittliga vikten per kyckling ser ut på följande sätt:

```
tapply(X=ChickWeight$weight, INDEX=ChickWeight$Chick, FUN=mean)
```

3. Pröva att på ett liknande sätt beräkna standardavvikelsen för varje kyckling samt antalet observationer (längden av vektorn) och kvantilerna med quantile().

4. Pröva nu att skicka argument till quantile (med ...) för att räkna ut percentiler för varje kyckling.

5. *apply-funktioner är särskilt smidiga att använda tillsammans med så kallade anonyma funktioner (d.v.s. funktioner som skapas "on the fly". Pröva koden nedan. Vad gör den?

```
tapply(X=ChickWeight$weight, INDEX=ChickWeight$Chick, FUN=function(x) sum(x)^2)
```

6. Pröva att på liknande sätt skapa en funktion som räknar ut skillnaden mellan det första värdet och det sista värdet för varje kyckling med en anonym funktion.

7. Pröva nu att göra om uppgiften ovan, men i `tapply()` ange `simplify = TRUE`. Vad är skillnaden?

8. Vi ska nu studera `lapply()`. `lapply()` använder en funktion `FUN` på varje element i en lista `X`. Kör koden nedan: Vad har du skapat för lista?

```
myList <- split(x = ChickWeight, f = ChickWeight$Diet)
```

9. Räkna ut medelvärde, varians och percentiler för varje weight i varje element i `myList` med `lapply()`.

10. Skapa nu en funktion som kan ta ett dataset för en kyckling och räknar ut skillnaden i vikt mellan tidpunkt 0 och tidpunkt 10, om värden saknas för tidpunkt 0 eller 10 ska `NA` returneras. Använd denna funktion tillsammans med `lapply()` för att beräkna den skillnaden mellan tidpunkt 0 och 10 för alla kycklingar.

11. Testa nu koden nedan med `apply()`. Vad betyder `MARGIN = 2` eller `MARGIN = 1`? Kolla i hjälpen!

```
data("trees")
data("iris")
apply(X = trees,MARGIN = 2,FUN = mean)
apply(X = trees,MARGIN = 1,FUN = mean)
apply(X = iris[,-5],MARGIN = 2,FUN = mean)
?quantile
apply(X = iris[,-5],MARGIN = 2,FUN = quantile)
apply(X = iris[,-5],MARGIN = 2,FUN = quantile,probs=c(0.1,0.5,0.9))
```

Kapitel 4

Dokumentation av funktioner - ROxygen

När vi arbetar med att utveckla funktioner finns det ofta ett behov av att dokumentera de funktioner vi skapar. Vi dokumenterar inte bara funktioner för andras skull utan också för vår egen skull. Det kan många gånger vara nog så svårt att komma ihåg hur vi tänkte för fyra månader när vi skrev en funktion.

Det är viktigt att vi dokumenterar en funktion tillsammans med funktionen. Annars är risken stor att vi kanske ändrar en funktion och glömmer sedan bort att ändra i dokumentationen när vi ändrar i vår funktion.

Det standardsätt att dokumentera funktioner i R kallas R-documentation och utgör den dokumentation som vi får upp med `help()` eller `?` för enskilda funktioner. Dessa dokument är skrivna i LaTeX och är separata dokument som är kopplade till funktioner i R-paket. Vi kan således bara använda detta för funktioner i paket.

Vill vi dokumentera våra funktioner utan att skapa egna paket använder vi ROxygen. Det är ett format för att dokumentera funktioner direkt i R. Har vi väl dokumenterat våra funktioner med Roxygen kan vi generera R documentation-filer automatiskt om vi väljer att lägga in funktionen i ett paket.

Ett exempel på hur roxygen-dokumentation framgår nedan:

```
#' @description
#' En funktion som kvadrerar argumenten i x och y och summerar dem.
#'
#' @param x
#' Den numeriska variabel x som ska kvadreras
#' @param y
#' Den numeriska variabel y som ska kvadreras
#'
#' @return
#' Funktionen returnerar en numerisk vektor
#'
f <- function(x, y) x^2 + y^2
```

Observera att roxygendokumentation inleds med `#'`.

I dokumentationen ovan dokumenteras funktionens argument, vad funktionen gör och vad funktionen returnerar för värde.

Följande delar i dokumentationen är vanliga att använda.

roxygendel	Innehåll
`@title`	Anger titel för dokumentet
`@description`	En beskrivning vad funktionen gör
`@details`	Detaljer om funktionen, ex. speciella argument
`@param`	Argument till funktionen
`@return`	Vad funktionen returnerar
`@references`	Eventuella referenser av intresse
`@seealso`	Andra funktioner som kan vara aktuella
`@examples`	Exempel på hur funktionen kan användas

För ett exempel på dokumentation i Roxygen och hur det ser ut i R är funktionen `interactive_pxweb()` i paketet `pxweb`.

Funktionen med tillhörande dokumentation finns **[här]**. Nedan finns kod för att studera hur dokumentationen ser ut i ett R-paket.

```
install.packages("pxweb") # Om paketet inte installerats tidigare
library(pxweb)
?interactive_pxweb
```

Kapitel 5

Kodstil

Kolla på de olika stilguiderna:

- Hadley Wickhams
- Googles

Notera att dessa två stilguider kan ge något olika svar på samma fråga.
Svara nu på frågorna nedan:

1. Är `dayone` ett bra variabel namn?
2. Hur ska jag skriva ”`x+y`” eller ”`x + y`”?
3. Ska jag välja alt. 1 eller alt. 2?

```
# alt. 1
if (y == 0) {
log(x)
} else {
y ^ x
}

# alt. 2
if (y == 0) {
log(x)
}
else
{
y ^ x
}
```

4. Hur många tecken ska en rad med kod maximalt vara?
5. Hur bör jag kommentera min kod?
6. Vilken operator ska jag använda vid vanlig tilldelning? (Dvs jag vill jag variabeln `x` ska ha värdet 5.)
7. Hur ser ett bra filnamn ut?
8. Att diskutera: Varför är det viktigt att ha en bra kodstil?

Del II

Inlämningsuppgifter

Tips!

Inlämningsuppgifterna innebär att konstruera funktioner. Ofta är det bra att bryta ned programmeringsuppgifter i färre små steg och testa att det fungerar i varje steg.

1. Lös uppgiften med vanlig kod direkt i R-Studio (precis som i datorlaborationen ovan) utan att skapa en funktion.
2. Testa att du får samma resultat som testexemplen.
3. Implementera koden du skrivit i 1. ovan som en funktion.
4. Testa att du får samma resultat som i testexemplen, nu med funktionen.

Automatisk återkoppling med `markmyassignment`

Som ett komplement för att snabbt kunna få återkoppling på de olika arbetsuppgifterna finns paketet `markmyassignment`. Med detta är det möjligt att direkt få återkoppling på uppgifterna i laborationen, oavsett dator. Dock krävs internetanalsutning.

Information om hur du installerar och använder `markmyassignment` för att få direkt återkoppling på dina laborationer finns att tillgå **här**.

Samma information finns också i R och går att läsa genom att först installera `markmyassignment`.

```
install.packages("markmyassignment")
```

Om du ska installera ett paket i PC-pularana så behöver du ange följande:

```
install.packages("markmyassignment",lib="sökväg till en mapp i din hemkatalog")
```

Tänk på att i sökvägar till mappar/filer i R i Windowssystem så används ‘‘\\’’, tex ‘‘C:\\Users\\Josef’’.

Därefter går det att läsa information om hur du använder `markmyassignment` med följande kommando i R:

```
vignette("markmyassignment")
```

Det går även att komma åt vignetten **här**. Till sist går det att komma åt hjälpfilerna och dokumentationen i `markmyassignment` på följande sätt:

```
help(package="markmyassignment")
```

Lycka till!

Kapitel 6

Inlämningsuppgifter

För att använda `markmyassignment` i denna laboration ange:

```
library(markmyassignment)
```

```
Loading required package: methods
Loading required package: yaml
Loading required package: testthat
Loading required package: httr
```

```
lab_path <-
"https://raw.githubusercontent.com/STIMALiU/KursRprgm/master/Labs/Tests/d4.yml"
suppressWarnings(set_assignment(lab_path))
```

```
Assignment set:
D4 : Statistisk programmering med R: Lab 4
```

6.1 `hilbert_matrix()`

Vi ska nu skapa en funktion för att skapa godtyckligt stora kvadratiska Hilbertmatriser. Varje element i en Hilbertmatris är definierat som

$$H_{ij} = \frac{1}{i + j - 1}$$

För mer information om Hilbertmatrsier finns [här]. Skapa en funktion `hilbert_matrix()` med argumenten `nrow` och `ncol`. Funktionen ska returnera en Hilbertmatris enligt ovan. Nedan finns exempel på hur funktionen ska fungera.

Tips! Använd en nästlad for-loop.

```
hilbert_matrix(nrow = 2, ncol = 4)
```

```
     [,1]    [,2]    [,3] [,4]
[1,]  1.0 0.50000 0.33333 0.25
[2,]  0.5 0.33333 0.25000 0.20
```

```
hilbert_matrix(3, 3)
```

```
        [,1]    [,2]    [,3]
[1,] 1.00000 0.50000 0.33333
[2,] 0.50000 0.33333 0.25000
[3,] 0.33333 0.25000 0.20000
```

6.2 `toeplitz_matrix()`

Ännu en specialmatris som finns är den så kallade Toeplizmatrisen. Den kallas också ibland för diagonalkonstant matris. Samtliga diagonaler har samma värde. Mer information om Toeplitzmatriser finns [här].

Nedan är ett exempel på en toeplitzmatris.

$$\begin{pmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & a \end{pmatrix}$$

Vi ska nu skapa en funktion som skapar en Toeplitzmatris av godtycklig storlek. Funktionen ska ta en godtycklig vektor `x` och returnera en toeplizmatris baserat på denna vektor. Funktionen ska först kontrollera att vektorn är av en udda längd och annars returnera "`x not of odd length`".

Tips! Använd en nästlad `for`-loop.

```
toeplitz_matrix(x="Ett element")

     [,1]
[1,] "Ett element"

toeplitz_matrix(x=1:5)

     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    1    2
[3,]    5    4    1

toeplitz_matrix(c("a","b", "c", "d", "e", "f", "g"))

     [,1] [,2] [,3] [,4]
[1,] "a"  "b"  "c"  "d"
[2,] "e"  "a"  "b"  "c"
[3,] "f"  "e"  "a"  "b"
[4,] "g"  "f"  "e"  "a"

toeplitz_matrix(1:4)

x not of odd length.

toeplitz_matrix(c(1,0,0,0,0))

     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

6.3 Väntevärde och varians i en dimension

I denna uppgift ska du skapa funktioner som kan beräkna teoretiskt väntevärde och varians för en slumpvariabel i en dimension. Totalt ska du skapa tre funktioner `E_discrete()`, `V_discrete()` och `moment_func()`.

Det teoretiska väntevärdet för diskret slumpvariabel defineras som:

$$\mu = E(X) = \sum_{i=1} x_i \cdot p(x_i) \tag{6.1}$$

Här är x_1 till x_N alla möjliga utfall av slumpvaraiabeln X, och $p(x)$ är sannolikhetsfördelningen för alla möjliga utfall. Väntevärdet för en linjärkombination $a \cdot X + b$ av en diskret slumpvariabel ges av:

$$\mu_{linj} = E\left(a \cdot X + b\right) = b + \sum_{i=1} a \cdot x_i \cdot p\left(x_i\right) \tag{6.2}$$

Nu ska du skapa en funktionen som kan beräkna väntevärde enligt (6.1) och (6.2). Funktionen ska heta `E_discrete(density_matrix=matrix(c(1,1),ncol = 2),trans=c(1,0))`. Notera argumentnamn och defaultvärden.

Argumenten är:

- `density_matrix=` En numerisk matris med två kolumner, den första representerar de olika utfallen på slumpvariabeln X och den andra är sannolikheterna för de olika utfallen $p\left(x\right)$. Den andra kolumnen ska summera till 1.
- `trans=` är en numerisk vector av längd två, där första elementet är a och andra elementet är b i (6.2).

Funktionen ska kontrollera att villkoren är uppfyllda, om inte så ska den avbryta och ge felmedelanden enligt exemplen nedan:

```
y<-1:3
py<-c(0.2,0.7,0.1)
y_mat<-cbind(y,py)
E_discrete(density_matrix = matrix("hej"))

Error in E_discrete(density_matrix = matrix("hej")): density_matrix is not numeric!

E_discrete(trans = matrix("hej"))

Error in E_discrete(trans = matrix("hej")): trans is not numeric!

E_discrete(density_matrix = cbind(y,c(0.4,0.6,0.8)))

Error in E_discrete(density_matrix = cbind(y, c(0.4, 0.6, 0.8))): Probabilities do not sum
to 1!

E_discrete(trans = c(1,2,3))

Error in E_discrete(trans = c(1, 2, 3)): trans has not length 2!
```

Testa nu följande fungerande fall:

```
x<-1:10
px<-rep(1/10,10)
x_mat<-cbind(x,px)

E_discrete()

[1] 1

E_discrete(trans = c(2,0))

[1] 2

E_discrete(trans = c(2,5))

[1] 7

E_discrete(density_matrix = x_mat)

[1] 5.5

E_discrete(density_matrix = y_mat)

[1] 1.9
```

```
E_discrete(density_matrix = y_mat,trans = c(4,0))

[1] 7.6

E_discrete(density_matrix = y_mat,trans = c(1,5))

[1] 6.9

E_discrete(density_matrix = y_mat,trans = c(0,5))

[1] 5
```

Det teoretiska variansen för diskret slumpvariabel defineras som:

$$\sigma^2 = Var\left(X\right) = \sum_{i=1}^{N} \left(x_i - \mu\right)^2 \cdot p\left(x_i\right) \tag{6.3}$$

Här är x_1 till x_N alla möjliga utfall av slumpvaraiabeln X, och $p\left(x\right)$ är sannolikhetsfördelnignen för alla möjliga utfall. Variansen kan även beräknas som:

$$Var\left(X\right) = E\left(X^2\right) - E\left(X\right)^2 \tag{6.4}$$

Varinasen för en linjärkomination $a \cdot X + b$ av en diskret slumpvariabel ges av:

$$\sigma^2_{linj} = Var\left(a \cdot X + b\right) = a^2 \cdot Var\left(X\right) \tag{6.5}$$

Nu ska du skapa en funktionen som kan beräkna varians enligt (6.3) eller (6.4) och (6.5). Funktionen ska heta `V_discrete(density_matrix=matrix(c(1,1),ncol = 2),trans=c(1,0))`. Notera argumentnamn pch defaultvärden.

Argumenten är:

- `density_matrix=` En numerisk matris med två kolumner, den första representerar de olika utfallen på slumpvariabeln X och den andra är sannolikheterna för de olika utfallen $p\left(x\right)$. Den andra kolumnen ska summera till 1.
- `trans=` är en numerisk vector av längd två, där första elementet är a och andra elementet är b i (6.5).

Funktionen ska kontrollera att villkoren är uppfyllda, om inte så ska den avbryta och ge felmedelanden enligt exemplen nedan. **Tips**: använd funktionen `E_discrete` när du beräknar variansen.

```
y<-1:3
py<-c(0.2,0.7,0.1)
y_mat<-cbind(y,py)
V_discrete(density_matrix = matrix("hej"))

Error in V_discrete(density_matrix = matrix("hej")): density_matrix is not numeric!

V_discrete(trans = matrix("hej"))

Error in V_discrete(trans = matrix("hej")): trans is not numeric!

V_discrete(density_matrix = cbind(y,c(0.4,0.6,0.8)))

Error in V_discrete(density_matrix = cbind(y, c(0.4, 0.6, 0.8))): Probabilities do not sum
to 1!

V_discrete(trans = c(1,2,3))

Error in V_discrete(trans = c(1, 2, 3)): trans has not length 2!
```

Testa nu följande fungerande fall:

```
x<-1:10
px<-rep(1/10,10)
x_mat<-cbind(x,px)

V_discrete()

[1] 0

V_discrete(trans = c(2,0))

[1] 0

V_discrete(trans = c(2,5))

[1] 0

V_discrete(density_matrix = x_mat)

[1] 8.25

V_discrete(density_matrix = y_mat)

[1] 0.29

V_discrete(density_matrix = y_mat,trans = c(4,0))

[1] 4.64

V_discrete(density_matrix = y_mat,trans = c(1,5))

[1] 0.29

V_discrete(density_matrix = y_mat,trans = c(0,5))

[1] 0
```

Nu ska du skapa en funktion som ska heta `moment_func(densities,moment=NULL)`. Funktionen syfte är att ta kunna ta flera diskreta sannolikhetsfördelningar, med eventualla linjärkombinationsparametrar och beräkna olika moment (väntevärde och varians). Argument:

- `densities=` Är en lista. Varje element ska i sin tur vara en lista med elementen `density_matrix` och `trans`, som motsvarar argumenten för funktionerna `E_discrete` och `V_discrete`.
- `moment=` Är en av följande funktioner: `E_discrete` och `V_discrete`.

Se nedan för exempel på kraven på argumenten:

```
x<-1:10
px<-rep(1/10,10)
x_mat<-cbind(x,px)
y<-1:3
py<-c(0.2,0.7,0.1)
y_mat<-cbind(y,py)
first_list<-list(density_matrix=x_mat,trans=c(2,3))
second_list<-list(density_matrix=y_mat,tran=c(5,1))
final_list1<-list(first_list,second_list)
moment_func(densities = final_list1,moment = E_discrete)
```

Det första som funktionen ska göra är att kontrollera argumenten. Kontrollera först att `densities` är en lista, stoppa annars funktionen med meddelandet "`densities is not a list!`". Kontrollera sedan att `moment` är en funktion, stoppa annars funktionen med meddelandet "`No function supplied!`". Tips: `is.function()`. Funktionen ska nu gå igen alla element i `densities` och givet elemeten `density_matrix`

och `trans` anropa funktionen `moment`. Resultatet ska sparas i en lista, med elementnamn på formen `density1`, `density2`, ... Se även exemplen nedan. Listan ska sedan returneras. Föslag på lösning:

- Kontrollera argumenten.
- Sätt upp en tom lista för resutlaten. Tips: `vector()` kan användas för att skapa tomma listor av önskad längd.
- Gå igenom alla element i `densities` och anropa `moment` genom att använda `do.call()`.

Testa nu din funktion och se om den fungerar enligt exemplen nedan:

```
x<-1:10
px<-rep(1/10,10)
x_mat<-cbind(x,px)
y<-1:3
py<-c(0.2,0.7,0.1)
y_mat<-cbind(y,py)
z<-seq(1,by = 3,length=5)
pz<- c(0.1,0.2,0.3,0.2,0.2)
z_mat<-cbind(z,pz)

test1<-list(density_matrix=x_mat)
test2<-list(density_matrix=y_mat)
test3<-list(density_matrix=z_mat)
test4<-list(density_matrix=x_mat,trans=c(4,2))
test5<-list(density_matrix=y_mat,trans=c(3,0))

test_list1<-list(test1)
test_list2<-list(test1,test2)
test_list3<-list(test1,test2,test3)
test_list4<-list(test1,test4,test5)

moment_func(densities = TRUE,moment = E_discrete)

Error in moment_func(densities = TRUE, moment = E_discrete): densities is not a list!

moment_func(densities = test_list1,moment = 1:10)

Error in moment_func(densities = test_list1, moment = 1:10): No function supplied!

moment_func(densities = test_list1,moment = E_discrete)

$density1
[1] 5.5

moment_func(densities = test_list2,moment = V_discrete)

$density1
[1] 8.25

$density2
[1] 0.29

moment_func(densities = test_list3,moment = E_discrete)

$density1
[1] 5.5

$density2
[1] 1.9

$density3
[1] 7.6
```

```
moment_func(densities = test_list4,moment = E_discrete)

$density1
[1] 5.5

$density2
[1] 24

$density3
[1] 5.7

moment_func(densities = test_list4,moment = V_discrete)

$density1
[1] 8.25

$density2
[1] 132

$density3
[1] 2.61
```

Grattis! Nu är du klar!