

Tentamen i Programmering i R, 7.5 hp

Skrivtid: 14-18
Hjälpmedel: Valfritt tryckt material (böcker, anteckningar mm)

Jourhavande lärare: Josef Wilzén & Måns Magnusson
Betygsgränser: Tentamen omfattar totalt 20 poäng. 12 poäng ger Godkänt. 16 poäng ger Väl godkänt.

Skriv dina lösningar i fullständig och läsbar kod.

Lösningen skrivs i en körbar textfil med namnet Main.R.

Kommentera direkt i Main.R filen när något behöver förklaras eller diskuteras.

Eventuella grafer som skapas under tentans gång behöver INTE skickas in för rättning, det räcker med att skicka in den kod som producerar figurerna.

OBS: Glöm inte att spara din fil ofta! Om R krashar kan kod förloras.

1. Datastrukturer

- (a) Beräkna $\sin\left(\frac{\pi}{7}\right) + \left|\log_3 \frac{1}{e}\right| + 2^{\frac{1}{2}}$. **1p**
- (b) Skapa följande vektorer: **1.5p**
 - i. `num`: innehåller talen (0.02, 0.04, 0.06, ..., 1), d.v.s. totalt 50 element.
 - ii. `myText`: som innehåller följande `c('Mon', 'Tue', NA, 'Thu', 'Fri')` upprepade 10 gånger.
 - iii. `cosine`: skapas genom formeln $y = \cos(x)$ där x är vektorn `num` ovan och y är vektorn `cosine` som ska skapas.
- (c) Skapa en dataframe som du kallar `myDF` och som består av variablerna `num`, `myText` och `cosine`. Lägg till en (logisk) variabel i din dataframe som du kallar `missing` och som anger TRUE om `myText` saknar värde (NA), annars FALSE. Räkna ut summan av variabeln `num` för de observationer som saknar värde i `myText`. **1.5p**

2. Kontrollstrukturer

- (a) Skriv en loop som beräknar och skriver ut resten (modulo) från divisionen mellan heltalen 19 till och med 31 (i täljaren) och talet 10 (i nämnaren). Ex. $13 \bmod 10 = 3$. **1p**

- (b) Skapa en textvektor med elementen `c('Sat','Mon','Wed','Tue','Fri')`. Skriv en ny loop som går igenom denna vektor och skriver ut ett meddelande för varje element. Är det en veckodag ska `"Weekday."` skrivas ut till skärmen annars ska `"Week-end!"` skrivas ut. Observera att loopen ska fungera för en godtyckling vektor med veckodagsförkortningar. **2p**
- (c) Skriv en loop som hittar det **minsta** heltal $Z > 1000$ som är jämt delbart med 23 och skriv ut detta tal till skärmen. **1p**

3. Funktioner

- (a) En vanlig operation inom statistik är att multiplicera en matris med dess transponat och sedan ta matrisinversen av denna multiplikation, d.v.s. $B = (X^T X)^{-1}$. Skapa en funktion som kan ta en godtycklig numerisk matris och genomföra denna operation. Kalla funktionen `myFun(X)` och pröva funktionen på följande matris. **2p**

$$X = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

- (b) Skapa en funktion som du kallar `change` som kan ta en textsträng som argument, argumentet ska heta `Text`. Funktionen ska söka igenom textsträngen efter - (bindestreck) och byta ut detta mot ett `" till "`. Pröva din funktion på `"12.00-13.00"`. **2p**

4. Statistik och grafik

För att det ska vara möjligt att reproducera dina resultat vid rättning behöver du bestämma från vilken seed som slumpalen ska genereras. Kör följande kod i R:

```
set.seed(73233)
```

- (a) Generera/simulera två vektorer, x och y , som består av 100 värden från poissonfördelningen med parametern $\lambda_x = 4$ och en vektor $\lambda_y = 25$. Skapa ett scatterplot mellan x och y . **1p**
- (b) Räkna ut medelvärdet och variansen för x och y . Beräkna även kovariansen och korrelationen mellan x och y . **1p**
- (c) Skapa en tom numerisk vektor som du kallar `varSamplingDistribution` av längd 1000. Gör en loop som i varje steg i simulerar x enligt uppgift a) ovan. Därefter ska variansen beräknas för den aktuella simuleringen och sparas som element i i vektorn `varSamplingDistribution`. Skapa sedan ett histogram över `varSamplingDistribution`. **2p**

5. Strängar och datum

- (a) Läs in paketet `lubridate` i R. Din kompis är född den 5 november 1987, hur många hela veckor hade hon levt den 1 januari 2013? **1p**
- (b) Skapa en datumvektor med den 1:a i varje månad under 2013. Skriv ut vektorn på skärmen. **1p**

- (c) Läs in paketet `stringr` i R. Läs in textfilen `BBCarticle201302.txt` i R. Varje mening är ett eget textobjekt. Räkna ut hur många meningar som innehåller landsnamnet “Burma” samt den genomsnittliga teckenlängden i per mening. **2p**

Lycka till!