

Tentamen i Programmering i R, 7.5 hp

Skrivtid: 8-12
Hjälpmedel: Inget tryckt material, dock finns "R reference card v.2" av Matt Baggot tillgängligt elektroniskt.
Betygsgränser: Tentamen omfattar totalt 20 poäng. 12 poäng ger Godkänt, 16 poäng ger Väl godkänt.
Tänk på följande:
Skriv dina lösningar i **fullständig och läsbar kod**.
Lösningen skrivs i en körbar textfil med namnet **Main.R**
Se filen **DocStudent.pdf** för hur tentan ska lämnas in.
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, logik och beräkningar (4p)

- (a) Beräkna $e^\pi + \log_2(4^3)$ **1p**
- (b) Läs in datamaterialet `iris` och skapa en ny logisk variabel i datasetet som du kallar `small_petal` och som anger om variabeln `Petal.Length` är mindre än 4. **1p**
- (c) Beräkna den genomsnittliga `Sepal.Length` för de tre olika orkidéarterna som anges i variabeln `Species`. **1p**
- (d) Plocka ut de orkidéer som har en `Sepal.Width` som är mindre än 3 och en `Petal.Width` som är större än 2. **1p**

2. Kontrollstrukturer (4p)

- (a) Skapa en kod som loopar över värdena 400 till 500 och skriver ut alla värden för vilka kvadratroten blir ett heltal (med `print()`). Exempel på tal som ska skrivas ut är 1, 4, 9, ... **2p**
Obs! Funktionen måste använda en `for`-loop.
- (b) Skapa nu en `while`-loop som, liknande i uppgift a), med en `while`-loop gå igenom värdena 500 till 600 och summera de värden vars kvadratroten är ett heltal. **2p**
Obs! Funktionen måste använda en `while`-loop.

3. Strängar och datum (4p)

- (a) Läs in paketen `lubridate` och `stringr` i R. **0.5p**
- (b) Marie Curie var en av vetenskapshistoriens främsta fysiker. Hon föddes den 7 november 1867 och dog den 4 juli 1934. Hon belönades med två nobelpris, 1903 och 1911. Nobelpriset delas ut den 10 december varje år. Använd `lubridate` för att besvara följande frågor: **1.5p**
- Vilken veckodag föddes Marie Curie?
 - Hur många veckor hade Marie Curie kvar att leva när hon fick sitt andra Nobelpris?
 - Hur många dagar levde Marie Curie?
- (c) Läs in dikten `wilde.txt` i R som `poem`. **0.5p**
- (d) Räkna ut den genomsnittliga längden på orden i dikten (med ord avses textsträngar som avgränsas med mellanslag). **1.5p**

4. Funktioner (4p)

- (a) Skapa en funktion du kalla `approx_e()` som approximerar konstanten e på följande sätt. Funktionen ska generera ett värde på e som är korrekt till 4 decimalen som standard. Annars ska det gå att styra K nedan. **2p**

$$e \approx \sum_{k=0}^K \frac{1}{k!} \text{ där } K \rightarrow \infty$$

```
approx_e()
[1] 2.71825

approx_e(K = 2)
[1] 2.5
```

- (b) Vi ska nu på ett liknande sätt implementera en funktion du kallar `approx_ln()` för att godtyckligt approximera $\ln(z)$. För att approximera använd följande funktion där approximationen blir godtyckligt god när $K \rightarrow \infty$. **2p**

$$\ln(z) = 2 \sum_{k=0}^K \frac{1}{2k+1} \left(\frac{z-1}{z+1} \right)^{2k+1}$$

```
approx_ln(exp(1), K = 2)
[1] 0.998455

approx_ln(exp(1), K = 5)
[1] 0.999992
```

5. Statistik och grafik (4p)

- (a) Skapa en vektor \mathbf{x} genom att upprepa talen 1 till 10, 20 ggr (d.v.s. totalt en vektor av längd 200). Simulera sedan data från vektorn \mathbf{y} på följande sätt: **1p**

$$y_i = 5 + 0.5 \cdot x_i + \epsilon_i$$

där

$$\epsilon_i \sim \mathcal{N}(\mu = 0, \sigma = 2)$$

- (b) Visualisera variabeln x mot y i en scatterplot. **1p**
(c) Anpassa en linjär regressionsmodell mellan x mot y . **1p**
(d) Plocka ut och visualisera residualerna i ett histogram. **1p**

Lycka till!