

## 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  $\pi^e + \log_2 3$ . **1p**
- (b) Skapa följande vektorer: **1.5p**
  - i. `x`: innehåller talen (2, 4, 6, 8, ..., 50)
  - ii. `myText`: som innehåller följande `c("Mon", "Tue", "Wed", "Thu", "Fri")` upprepad 5 gånger.
  - iii. `y`: skapas genom formeln  $y = 10 + \log_e x$  där `x` är vektorn `x` ovan.
- (c) Skapa en dataframe som du kallar `myDF` och som består av variablerna `x`, `myText` och `y`. Lägg till en (logisk) variabel i din dataframe som du kallar `check` och som anger TRUE om  $x > y$ , annars FALSE. Räkna ut i R hur många rader som är FALSE. **1.5p**

### 2. Kontrollstrukturer

- (a) Skriv en loop som skriver ut alla heltal mellan 45 och 61 till skärmen. **1p**
- (b) Skriv en ny loop som går igenom talen 139 till 177 och i varje iteration undersöker (1) om talet är delbart med 7 - skriv då ut "X can be divided with 7." till skärmen, (2) om talet är delbart med 11 - skriv då ut "X can be divided with 11." till skärmen. Om talet både är delbart med både 7 och 11 ska den skriva ut "X is my favorite number!". Om varken (1) eller (2) är uppfyllt ska den inte skriva ut något till skärmen. **2p**

- (c) Skriv en loop som hittar det **största** heltal  $Z$  som uppfyller olikheten  $\log_e Z \leq 6$  och skriver ut detta till skärmen. Börja med att  $Z = 1000$ . **1p**

### 3. Strängar och datum

- (a) Läs in paketen `lubridate` och `stringr` i R. Din kompis är född den 3 maj 1989, vilket datum kommer hon fylla 10000 dagar? **1p**
- (b) Generera en vektor med de datum hon fyller 2000, 4000 o.s.v. t.o.m. 10000 dagar. Skriv ut vektorn på skärmen. **1p**
- (c) Läs in textfilen `BBCarticle.txt` i R. Räkna det totala antalet tecken i artikeln. **1p**
- (d) Räkna antalet gånger ordet "the" förekommer i texten med funktionerna. Alla ord ska räknas, oavsett versaler/gemener eller om ordet är i början av en mening eller i början av ett citat. Tänk dock på att inte få med ord som "other". **1p**

### 4. Funktioner

- (a) Talet  $e$  är ett gränsvärde som kan beräknas på följande sätt:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad \text{för alla } x$$

Skapa en egen funktion `myExp()` som beräknar  $e$  upphöjt till valfritt värde  $x$ . I formeln ovan är serien oändlig, men ju även vid en begränsat antal summationer blir approximationen bra. Lägg därför till ett argument `N` som anger  $n = 0, 1, 2, \dots, N$ . Jämför med `exp(1)`, hur stort behöver `N` vara för att approximationen av  $e$  ska vara samma tal upp till den 5:e decimalen? **2p**.

- (b) Skapa en funktion som kan ta ett personnummer som en textsträng på formatet `YYMMDD-NNNN`. Om den tredje kontrollsiffran är jämn ska funktionen returnera `Female`, annars returneras `Male`. Prova din funktion på 811218-9876. **2p**

### 5. Statistik och grafik

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

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

- (a) Generera en vektor av 1000 värden från weibullfördelningen med parametrarna `shape=10` och `scale=10`. Skapa ett histogram över fördelningen. **1p**
- (b) Simulera fram en vektor av storlek 400 som du kallar  $x$  från den uniforma fördelning  $[0, 1]$ . Simulera sedan fram en vektor  $y \sim \mathcal{N}(\mu_i, \sigma)$ , normalfördelad med  $\mu_i = 10 \cdot x_i$  och  $\sigma = 1$ , där  $x_i$  är elementen i vektorn  $x$  och  $\mu_i$  är medelvärdet i normalfördelningen. Skapa ett histogram över  $y$ . **1.5p**
- (c) Gör en linjär regression med  $y$  som beroende variabel och  $x$  som hjälpvariabel (regressor), d.v.s. modellen  $y = \alpha + \beta x + \epsilon$ . Är  $\alpha$  och  $\beta$  signifikanta? Varför, varför inte? **1.5p**

*Lycka till!*