

Tentamen i Programmering i R, 7.5 hp

Skrivtid: 8-12

Hjälpmaterial: Inget tryckt material, dock finns referenskort i mappen cheetsheets tillgängligt.

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**.

Spara filen med namnet `tentaX.R` där X är ditt SC-nummer (klientnummer).

Det numret kan du se i studentklienten.

Spara filer i mappen `/home/student_tilde/`

Exempel: om du har SC-nummer SC12345 så ska den heta `tentaSC12345.R`

Notera att du ska lämna in en fil med alla dina lösningar.

Frågor kan ställas till lärare via studentklienten.

När du har loggat ut från datorn är tentan avslutad.

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

Eventuella grafer 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 och beräkningar (4p)

- (a) Beräkna $y = e^{(x-3)} \cdot x^{(3-\log(x))}$ där $x = 2$ och \log är den natuliga logaritmen. Avrunda till fyra decimaler. **1p**
- (b) Skapa en lista med 4 element som innehåller talen 1,2,3 och 4, d.v.s ett numeriskt värde per listelement. Spara listan som `my_list`. **0.5p**
- (c) Återskapa `data.frame` enligt nedan. Spara som `my_df`. **1p**

	numeric	boolean	character
1	6	TRUE	a
2	5	FALSE	b
3	4	TRUE	c

4	3	FALSE	d
5	2	TRUE	e
6	1	FALSE	f

- (d) Skapa följande matris. Glöm inte namnen på rader och kolumner.
Spara som `my_matrix`. **1.5p**

	y1	y2	y3	y4
x1	1	4	7	10
x2	2	5	8	11
x3	3	6	9	12

2. Kontrollstrukturer (4p)

- (a) Skapa en nästlad `for`-loop som skriver ut texten enligt nedan. **2p**

```
[1] "i: 1"
[1] "j: 1"
[1] "j: 2"
[1] "j: 3"
[1] "j: 4"
[1] "i: 2"
[1] "j: 1"
[1] "j: 2"
[1] "j: 3"
[1] "j: 4"
[1] "i: 3"
[1] "j: 1"
[1] "j: 2"
[1] "j: 3"
[1] "j: 4"
```

- (b) Använd nu `break` för att använda din nästlade `for`-loop för att skriva ut följande: **2p**

```
[1] "i: 1"
[1] "j: 1"
[1] "j: 2"
[1] "j: 3"
[1] "j: 4"
[1] "i: 2"
[1] "j: 1"
```

3. Strängar och datum (4p)

- (a) Läs in paketen `lubridate` och `stringr` i R. Läs sedan in datamaterialet "word_data.txt" och spara som vektorn `word_data`. **0.5p**
- (b) Använd funktioner i `stringr` för att: Plocka ut de ord ur `word_data` som innehåller "a" och räkna ut hur många ord det är. Spara antalet ord i variabeln `no_word`. Ta sedan reda på det totala antalet "a" som finns i `word_data`, spara i variabeln `no_a`. **2p**
- (c) Använd funktioner ur R för att svara på följande frågorna nedan. **1.5p**
- Vilken veckodag var det den 11:e november 1918 (Första världskriget tog slut)?
 - Vilket veckonummer var det den 28 februari 1986 (Då Olof Palme sköts till döds)?
 - Koreakriget pågick från den 25 juni 1950 till den 27 juli 1953. Hur många dagar pågick kriget?

4. Funktioner (4p)

- (a) Skapa en function du kallar `wonder_woman()`, som med en eller flera villkorssatser, skriver ut namnet på kvinnorna nedan om deras födelseår har angetts som argumentet `year`. Om inget av de tre födelseåren har anges (eller ett annat år anges) ska programmet skriva ut "unknown". **2 p**
- Amelia Earhart (f. 1897)
 - Ada Lovelace (f. 1815)
 - Vigdis Finnbogadottir (f. 1930)
- (b) Skapa funktionen `meta_data()`. Funktionen har två argument: `x` och `func`. `x` ska vara en lista, som innehåller olika dataset (`data.frames`) och `func` ska vara en godtycklig funktion. Funktionen ska gå igenom listan `x` och applicera funktionen `func` på varje element i listan och spara resultatet som en lista. Den nya listan ska ha samma namn som `x`. Se testfallen för hur funktionen ska fungera. **2 p**

```
a1<-list(a=trees)
meta_data(x = a1,func = colMeans)

$a
  Girth  Height  Volume
13.2484 76.0000 30.1710
```

```

a2<-list(InsectSprays=InsectSprays,trees=trees)
b2<-meta_data(x = a2,func = length)
b2

$InsectSprays
[1] 2

$trees
[1] 3

a3<-list(a=InsectSprays,b=trees,c=ChickWeight)
b3<-meta_data(x = a3,func = nrow)
b3

$a
[1] 72

$b
[1] 31

$c
[1] 578

```

5. Statistik och grafik (4p)

- (a) Simulera två vektorer av längd 300 från $X \sim N(\mu = 5, \sigma = 1)$ och $Y \sim N(\mu = 7, \sigma = 1)$. Sätt slumpseeden till 333 precis innan du simulerar vektorerna. **1p**
- (b) Räkna ut differensen mellan medianen för X och medianen för Y . **0.5p**
- (c) Slå samman X och Y och använd ggplot2 för att visualisera data i ett histogram. Texten på x-axeln ska vara "my_data". **1p**
- (d) Gör ett t-test för att testa om det är en signifikant skillnad i μ mellan X och Y . Använd signifikansnivå alpha=0.01, detta ger ett 99 % konfidensintervall. Spara t-statistikan i variabeln `t_val` och spara konfidensintervall i variabeln `my_conf`. **1.5p**

Kom ihåg: Skriv alla dina lösningar i en körbar **R-fil**. Spara filen med namnet **tentaX.R** där X är ditt SC-nummer (klientnummer) Det numret kan du se i studentklienten. Exempel: om du har SC-nummer SC12345 så ska filen hetा

tentaSC12345.R Lämna sedan in din fil via studentklienten. När du har loggat ut från datorn är tentan avslutad.

Lycka till!