

Övningar till R

Datorövning L9MA30, LGMA30, Matematisk statistik, Hösten 17

9 oktober 2017

Examinator: Claes Andersson, mejl andclae@chalmers.se

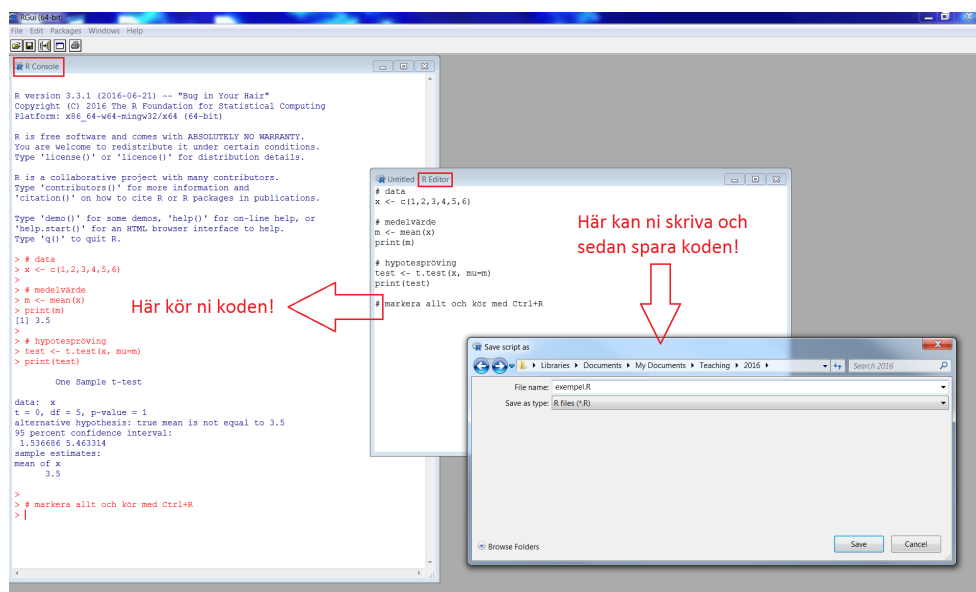
Datorövningen är obligatorisk. Om man inte kan delta vid övningstillfället, måste ett utfyllt övningsblad lämnas in senast på måndag, den 16 oktober 2017.

Extra uppgiften (ej obligatorisk) ger maximalt 0.5 bonuspoäng till tentamen (krävs ≥ 2 poäng). Sista inlämningsdag för extrauppgiften är måndag, den 17 oktober 2016.

I denna datorbaserade övning skall ni bekanta er med programvaran R, som är gratis att ladda ner från <https://cran.r-project.org/> och redan borde finnas installerat på datorerna i datorsalen (under 'all programs'). Börja med att starta programmet och öppna ett nytt skript via 'File' → 'new script' (eller Ctrl+N). Ni kanske vill också byta katalogen: klicka på consol-fönstret → 'File' → 'Change dir'. För att köra koden kan ni antingen markera de relevanta delarna och trycka Ctrl+R eller kopiera dem och klistra in dem i consolen. Ni kommer att använda R-funktionen `t.test`:

```
t.test(x, y = NULL, mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

Det är bra om ni läser igenom hjälpen: skriv ? `t.test` i consol-fönstret för att öppna html-dokumentet.



Övningar till R

Namn:

1) Konfidensintervall för väntevärden

Lös övning 5.9 (sidan 321) med hjälp av R. Du kan läsa in datan med `x <- c(147, 254, 132, 98, 195, 85)`.

- a) Konstruera själv ett 95% konfidensintervall för väntevärdet μ . Du kan använda följande R-funktioner som hjälp om du vill:

`mean()`, `sd()`, `sqrt()`, `length()`, `qt()`

Svar: $KI_{\mu}^a = [\quad , \quad]$

- b) Beräkna konfidensintervallet med `t.test()`; t.ex.

```
test <- t.test(x)
KI <- test$conf.int
```

Svar: $KI_{\mu}^b = [\quad , \quad]$

- c) Beräkna längden av konfidensintervallet; t.ex. med

```
L <- KI[2] - KI[1]
```

Slumpa 100 $\mathcal{N}(150, 64)$ -fördelade tal med

```
y <- rnorm(100, 151, 64)
```

och lägg till dem till `x` (`x112 <- c(x, y)`) och jämför konfidensintervallens längder. Förklara skillnaden.

Svar: $L^{12} = \quad$, $L^{112} = \quad$. Förklaring av skillnaden:

- d) Testa hypotesen $\mu = 165$ med det ursprungliga stickprovet `x` och det nya stickprovet med `12 + 100` observationer, förklara skillnaden.

Svar: Testresultat och förklaring av skillnaden:

2) Konfidensintervall för väntevärdesskillnad

Data för denna uppgift hittar du på hemsidan. Du kan läsa in data med `scan()`. Det finns två dokument, som innehåller data över skördad majs vid användning av två olika gödningsmedel.

- a) Jämför stickproven med `boxplot()`. Tror du att det finns en signifikant skillnad mellan medlens effekt?

Svar: (Lådagram behövs inte.)

Övningar till R

- b) Testa hypotesen att det finns en skillnad mellan avkastningarna på 95% konfidensnivån. Anta samma varians för båda stickprov.

Svar: Hypotesen förkastas ☐ ja, ☐ nej?

- c) Är antagandet att stickproven har samma varians rimligt? Motivera ditt svar.

Svar: Antagandet är rimligt ☐ ja, ☐ nej? Motivering:

- d) Vi har inte en enda gång plottat rådatan hittills. Gör detta med funktionen `plot`. Vad ser du? Tycker du vi analyserat datan på ett rimligt sätt?

3) Stickprov-i-par (fortsättning av GeoGebra övningen)

Erik vill ta reda på hur vilopuls påverkas av meditation. För att göra detta frågar han 15 av sina vänner om de vill delta i hans 'studie'. Han mäter deras vilopuls i början av studien och efter 10 dagar med 30 min meditation/dag. Data för denna uppgift hittar du på hemsidan. Text filen heter *dataR*. För att läsa in data till R skriv `data <- read.table("dataR.txt", header=TRUE)`.

- a) Är det rimligt att anta att observationer kommer från normalfördelade stokastiska variabler?

Ledning: Använd `hist()`.

Svar: Normal-antagningen är rimlig ☐ ja, ☐ nej? Motivation (histogram behövs inte ritas):

- b) Finns det en statistisk signifikant skillnad mellan vilopuls före och efter meditationen? Testa hypotesen på konfidensnivån 95%.

Svar: ☐ ja, ☐ nej?

- c) Prov hypotesen igen, men denna gång glömma att du har stickprov-i-par? Beskriv och förklara vad händer?

Svar:

Övningar till R

4) Extra uppgift (ej obligatorisk och ni kan använda R)

En undersökning av storlek $n = 12$ skolor visade att rund 35% av budgeten gick till modernisering av utrustning; se följande stickprov

$$x = (35.6, 33.7, 37.8, 31.2, 37.2, 34.1, 35.8, 36.6, 37.1, 34.9, 35.6, 34.0).$$

Man antar att x kommer från $N(\mu, 1.8)$ -fördelade stokastiska variabler.

- a) Konstruera ett 95% och 99% konfidensintervall för väntevärdet μ .

1p

- b) Hur och varför ändras intervall längden när man ökar konfidensnivån från 95% till 99%?

1p

- c) Testa hypotesen $\mu = 35$ på 95% konfidensnivån.

1p

- d) Hur många observationer (skolor) behövs, så att längden av konfidensintervallet till 95% konfidensnivån är mindre eller lika med 1.5?

1p