Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen: Førsteamanuensis Jarle Tufto
Telefon: 99 70 55 19

Bioberegninger, ST1301

30. november 2006
Kl. 9–13
Hjelpemidler: Alle trykte og skrevne hjelpemidler, lommeregner.

Hjelpesider for noen R-funksjoner som er omhandlet nedenfor eller som du vil kunne få bruk for i programmeringsoppgavene følger på side 5.

Oppgave 1

a) Anta at vi definerer følgende vektor i R

```
y <- c(1001,1002,1003,1004)
```

Hva blir da verdien av følgende uttrykk?

```
y[3]
y[3:4]
sum(y[3:4])
y[-2]
sum(y[-2])
i <- 1
y[i:(i+1)]
```

b) Vi definerer følgende funksjon i R.

```
minfunksjon <- function(z) {
  m <- length(z)
  delsum <- 0
  for (j in 1:m) {
     if (z[j]<5)
        delsum <- delsum + z[j]
     else
        delsum <- delsum + 5
  }
  delsum
}
```

Hva blir da verdien av funksjonskallet `minfunksjon(c(3,2,10,7,4))`?

Oppgave 2 Forsuring medfører redusert bestandstørrelse av fisk i mange vassdrag. For å undersøke effekten av kalking gjennomfører vi derfor kalking i et antall vassdrag og ingen kalking i et antall kontrollvassdrag. Så registrerer vi tettheten av lakseyngel samme år (variabelen `n0`) og året etter kalking (`n1`) (antall fisk per 100 m^2). I tillegg registrerer vi fangsten av laks (antall) samme år (`fangst`) som eventuell kalking ble gjennomført. Vi legger dataene inn i filen `kalking.dat` med følgende innhold

```
   elv kalk   n0 fangst   n1
  Mand  nei  9.4     52  0.7
 Audna   ja  7.5     56 13.4
 Lygna   ja  9.4     22  9.6
 Kvina  nei  4.9    251 -1.8
  Sira  nei  8.2     52  2.4
 Sokna  nei 10.6      2 10.6
 Helle  nei  6.4    105  8.6
 Bjerk   ja  9.8     28 14.4
  Ogna   ja  4.5    673  6.2
 Fugle   ja  8.0    141 13.9
 Kvass   ja 17.1    155 24.6
 S.Var  nei 15.0    146 10.2
 N.Var  nei  8.1    441  9.6
  Orre  nei 15.9     28 12.9
Figgjo   ja 13.4    243 16.5
  Imsa   ja 12.8      7 16.1
  Dird   ja 10.9      1 15.7
```

```
 Frafj   ja  9.7     91 15.4
```

Variabelen `kalk` er her en faktor som indikerer hvorvidt kalking er gjort eller ikke (verdiene `ja` og `nei`). Vi leser inn dataene, beregner endring i populasjonsstørrelse (`delta.n`), og tilpasser en lineær modell med `delta.n` som responsvariabel, og faktoren `kalk` og den kontinuerlige variabelen `fangst` som forklaringsvariable som vist nedenfor.

```
> data <- read.table("kalking.dat",header=T)
> attach(data)
> delta.n <- n1-n0
> modell <- lm(delta.n ~ kalk + fangst)
> summary(modell)
Call:
lm(formula = delta.n ~ kalk + fangst)

Residuals:
    Min      1Q  Median      3Q     Max
-5.5487 -2.2823  0.2313  1.6270  5.3585

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.144188   1.293767  -2.430 0.028110 *
kalkja       7.433462   1.534882   4.843 0.000215 ***
fangst      -0.000136   0.004491  -0.030 0.976238
---

Residual standard error: 3.235 on 15 degrees of freedom
Multiple R-Squared:  0.61,Adjusted R-squared: 0.558
F-statistic: 11.73 on 2 and 15 DF,  p-value: 0.0008573
```

a) Innfør matematiske symboler for størrelsene som inngår og skriv et matematisk (algebraisk) uttrykk for modellen som er tilpasset ovenfor. Hva er modellantakelsene?

b) Hvor stor er den estimerte økningen i tettheten av yngel som følge av kalking?

c) Er økningen som følge av kalking statistisk signifikant?

d) Er det noen statistisk signifikant effekt av fangsten av fisk på endringen i tetthet av yngel?

Oppgave 3 En vanlig metode for å estimere den ukjente størrelsen N til en populasjon er såkalt fangst-gjenfangst merkeforsøk. Metoden består i å merke et antall individ m som så slippes løs igjen. En andel m/N av populasjonen er da merket. Så gjøres det gjenfangst med tilbakelegging av et valgt antall individ n og vi observerer antallet X av disse som er merket. Det følger da at X vil være binomisk fordelt med parametere m/N og n. Vi har da at

$$E(X) = n\frac{m}{N} \tag{1}$$

slik at

$$E(\frac{X}{nm}) = 1/N. \tag{2}$$

Altså er $X/(mn) = 1/N$ en forventningsrett estimator av $1/N$. Følgelig er den inverse til denne,

$$\hat{N} = \frac{mn}{X}, \tag{3}$$

en fornuftig estimator av N (populasjonsstørrelsen).

a) Hva blir estimatet $\hat{N}$ av N dersom vi først merker $m = 50$ individ og vi så observerer at $X = 5$ av $n = 10$ individ i gjenfangstutvalget er merket?

b) Har vi noe grunnlag for å si at $\hat{N}$ er en forventningsrett estimator av N (populasjonsstørrelsen)?

c) Programmer en funksjon som gitt n, m, og N simulerer 1000 realisasjoner av utfallet av merkeforsøket (det er variabelen X som er stokastisk), beregner tilhørende verdier av $\hat{N}$ og som så estimerer forventningsverdien til $\hat{N}$ på grunnlag av simulerte verdier.

d) Vis at $\hat{N}$ gitt ved (3) er sannsynlighetsmaksimeringsestimatoren av N.

```
Binomial                package:stats                R Documentation

The Binomial Distribution

Description:

     Density, distribution function, quantile function and random
     generation for the binomial distribution with parameters 'size'
     and 'prob'.

Usage:

     dbinom(x, size, prob, log = FALSE)
     pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
     qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
     rbinom(n, size, prob)

Arguments:

    x, q: vector of quantiles.

       p: vector of probabilities.

       n: number of observations. If 'length(n) > 1', the length is
          taken to be the number required.

    size: number of trials.

    prob: probability of success on each trial.

log, log.p: logical; if TRUE, probabilities p are given as log(p).

lower.tail: logical; if TRUE (default), probabilities are P[X <= x],
          otherwise, P[X > x].

Details:

     The binomial distribution with 'size' = n and 'prob' = p has
     density

                  p(x) = choose(n,x) p^x (1-p)^(n-x)

     for x = 0, ..., n.

     If an element of 'x' is not integer, the result of 'dbinom' is
     zero, with a warning. p(x) is computed using Loader's algorithm,
     see the reference below.

     The quantile is defined as the smallest value x such that F(x) >=
     p, where F is the distribution function.

Value:

     'dbinom' gives the density, 'pbinom' gives the distribution
     function, 'qbinom' gives the quantile function and 'rbinom'
     generates random deviates.

     If 'size' is not an integer, 'NaN' is returned.
```