



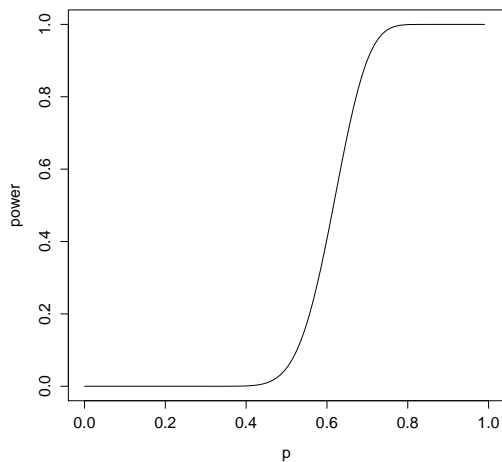
Oppgave 1

a)

```
power <- function(alpha,p0,p,n) {  
  z <- qnorm(alpha,lower.tail=FALSE)  
  pnorm((p-p0)/sqrt(p*(1-p)/n)-z*sqrt(p0*(1-p0)/(p*(1-p))))  
}
```

b) Når første argument p er en vektor som inneholder 0, 0.1, 0.2, ..., 1 vil de enkelte deler av uttrykket som beregnes i linje to i funksjonskroppen også bli vektorer fordi aritmetiske operatører som “-”, “+” o.s.v. og funksjoner som `sqrt` og `pnorm` virker elementvis på vektorer. Dermed vil funksjonen returnere teststyrken for alle disse verdiene av p .

c) Teststyrken vokser fra 0 til 1 med p og er lik α for $p = p_0$.



Oppgave 2

a) Vi kan tenke oss hvert individuelt bjørketre som forekomster i en romlig Poissonprosess med en viss intensitet som kan avhenge av forklaringsvariablene i modellen. Dette forutsetter at antall bjørketrær i disjunkte delareal er uavhengige variable og at ulike individuelle trær ikke kan befinne seg i samme punkt.

Parameteren i Poissonfordelingen (forvetningsverdien) er nødvendigvis ikke-negativ. Log link-funksjon innebærer at vi lar log til denne forvetningsverdien være bestemt av den

lineære prediktoren (et lineært uttrykk av de ulike forklaringsvariablene og regressjonskoeffisientene). Dette sikrer at modellen gir mening for alle parameterverdier og at prediksjoner for forventet antall trær i en gitt sampling-rute blir ikke-negative.

- b) Modellen er at antall trær Y er Poissonfordelt med parameter λ hvor

$$\log \lambda = \alpha_0 + \alpha_{\text{alt}} \text{alt} + \alpha_{\text{area}} \log(\text{area}) + \beta_{\text{time}} \quad (1)$$

hvor time er en faktor som tar verdien 1980 eller 2010.

For de gitt verdiene av forklaringsvariablene og parameterestimaterne blir

$$\log \lambda = -1.2496 - 0.001196 \cdot 1000 + 0.5439 + 1.2357 \cdot \log(200) = 4.6454 \quad (2)$$

og forventet antall trær i ruta

$$\lambda = e^{\log \lambda} = e^{4.6454} = 104.10. \quad (3)$$

- c) Om vi skriver modellen på en litt annen form ser vi at

$$\lambda = \text{area}^{\alpha_{\text{area}}} e^{\alpha_0 + \alpha_{\text{alt}} \text{alt} + \beta_{\text{time}}} \quad (4)$$

En fordobling av arealet medfører dermed at forventet antall trær øker med en faktor på $2^{\alpha_{\text{area}}} = 2^{1.24} = 2.36$.

A priori ville det være rimelig å anta at forventet antall trær er direkte proporsjonalt med arealet av hver rute dersom rutene er plassert helt tilfeldig i terrenget. Standardfeilen til estimatet av α_{area} antyder også at α_{area} ikke er signifikant forskjellig fra 1.

Om vi ønsker å bygge denne antakelsen inn i modellen kan vi i stedet bruke log area som en offset i modellen og tilpasse modellen ved å skrive

```
glm(trees~altitude+time,offset=log(area),family=poisson)
```

- d) Residual devians delt på residualt antall frihetsgrader gir oss et estimat av dispersjonsparameteren $\varphi = 0.59$ som antyder at det underdispersjon i dataene. Observerte residual devians er imidlertid ikke mindre enn nedre 0.025-kvantil lik 4.40 i χ^2 -fordelingen med 12 frihetsgrader så vi må beholde null-hypotesen at det ikke er hverken over eller underdispersjon i dataene.

Dersom individuelle bjørketrær opptrer i klynger (positive kovarians mellom antall trær i disjunkte naboområder) vil dette gi overdispersjon. Det samme gjelder forklaringsvariable som ikke er med i modellen, f.eks. forskjeller i jordsmonn. Konkurransen mellom individuelle trær vil kunne gi negative kovarianser mellom antall trær i disjunkte naboområder og underdispersjon.

Oppgave 3

a)

```
lnL <- function(p,x) {
  a <- p[1]
  b <- p[2]
  -sum(dgamma(x, shape=a, scale=b, log=T))
}
```

b) SMEene av a og b er 2.45 og 8.25.

c) Standardfeilene er 0.32 og 1.21.

d) Forventningen til en gammafordeltvariabel er i følge formelsamlingen

$$\mu = EX = f(a, b) = ab \quad (5)$$

Dermed er et estimat av μ , $\hat{\mu} = \hat{a}\hat{b} = 20.13$ dager.

Delta metoden gir at

$$\text{Var}(\hat{\mu}) = \left(\frac{\partial f}{\partial a}\right)^2 \text{Var}(\hat{a}) + \left(\frac{\partial f}{\partial b}\right)^2 \text{Var}(\hat{b}) + 2 \left(\frac{\partial f}{\partial a}\right) \left(\frac{\partial f}{\partial b}\right) \text{Cov}(\hat{a}, \hat{b}) \quad (6)$$

$$= \hat{b}^2 \text{Var}(\hat{a}) + \hat{a}^2 \text{Var}(\hat{b}) + 2\hat{a}\hat{b}\text{Cov}(\hat{a}, \hat{b}) \quad (7)$$

$$= 8.25^2 \cdot 0.1057 + 2.44^2 \cdot 1.4803 + 2 \cdot 2.44 \cdot 8.25 \cdot (-0.3565) \quad (8)$$

$$= 1.65. \quad (9)$$

slik at standardfeilen blir

$$SE(\hat{\mu}) = \sqrt{1.65} = 1.28. \quad (10)$$

e) Vi bruker at endring i 2 ganger maksimalt log likelihoodet er tilnærmet kjikvadratfordelt med antall frihetsgrader lik forskjell i antall parametere. Eksponentiell modell inneholder én mindre parameter enn gammafordelingen. Kritisk verdi i testen blir dermed øvre 0.05-kvantil i en kji-kvadratfordeling men 1 frihetsgrad, d.v.s. 3.85. Observert endring i 2 ganger maksimalt log likelihood blir

$$2(\ln L_1 - \ln L_0) = 2(-382.71 - -400.6) = 35.78 \quad (11)$$

slik at vi kan forkaste den enklere modellen (eksponentiell fordeling) til fordel for den mer kompliserte modellen (gammafordeling).