Oppgaver i STK3100.

Oppgave 13

Den "standardiserte" Cauchy-fordelingen har tetthet

$$f(y) = \frac{1}{\pi(1+z^2)} \qquad \text{for } -\infty < y < \infty$$

a) Plott tettheten $f(y)$.

b) Vis at den kumulative til $f(y)$ gis ved

$$F(y) = \frac{\arctan(y)}{\pi} + \frac{1}{2}$$

c) Det er implementert en simuleringsrutine `rcauchy` for Cauchy-fordelte variable `R`. Sammenlign plottet i punkt a) med et histogram over $n = 10000$ simulerte Cauchy-fordelte variable. (Det kan være lurt å se bort fra observasjoner som i absoluttverdi er større enn si 20).

d) Fordelingen til et gjennomsnitt av standardiserte Cauchy-fordelte variable har den forbløffende egenskap at de selv er standard Cauchy fordelt. Forklar hvorfor dette ikke er i motsetning til store talls lov.

e) Beregn $n = 1000$ gjennomsnitt av 100 standard Cauchy-fordelte variable.

f) Kjenner du et annen betegnelse på standard Cauchy-fordelingen?

Oppgave 14

I denne oppgaven skal vi se på noen data fra en studie av dødeligheten av lam i løpet av lammets første sommer. Vi skal konsentrere oss om å se på hvilken betydning lammets vekst fra fødselen til det slippes på sommerbeite har for dødeligheten. Variablen `tilvekst` kan sees som et mål på lammets helstilstand og styrke.

På filen

`http://www.uio.no/studier/emner/matnat/math/STK3100/h07/undervisningsmateriale/datasett/lam`

(dvs. på STK3100-hjemmesiden) ligger det informasjon om 179 lam fra en besetning et sted på Østlandet. Hver linje inneholder opplysninger om ett lam, mens kolonnene inneholder følgende opplysninger:

1. Tilvekst i gram pr. dag fra fødselen til lammet slippes på sommerbeite.

2. Indikator for død av sykdom.
   (0 = ikke død av sykdom, 1 = død av sykdom)

3. Indikator for drept av rovdyr.
   (0= ikke drept av rovdyr, 1 = drept av rovdyr)

4. Indikator for drept ved ulykke.
   (0= ikke drept ved ulykke, 1= drept ved ulykke)

5. Indikator for død av ukjent årsak.
   (0= ikke død av ukjent årsak, 1= død av ukjent årsak)

Da det tilsammen bare er observert 17 døde lam i løpet av sommeren, vil vi i denne oppgaven nøye oss med å studere den totale dødelighet for lammene uansett dødsårsak.

a) Hvorfor er logistisk regresjon egnet til å studere samenhengen mellom total dødelighet og tilvekst. Tilpasss modellen med totaldødelighet som respons og "tilvekst" som forklaringsvariabel.

b) Beregn og sammenlign Pearson- og deviansresidualene fra modelltilpasningen i a).

c) Legg inn kvadrat av `tilvekst` i modellen og undersøk på denne måten om den (logit-) lineære modellen i a) er rimelig.

d) Lag også en kategorisk variabel av tilvekst, f.eks. delt opp i grupper < 150 gram, 150-250 gram og > 250 gram. Tilpass også en modell med denne kategoriske variablen. Sammenlign med modellene i a) og c).