STK1000: Løsningsforslag Uke 46

2019

Oppgave 11.1

a) Responsvariabelen er endelig eksamens score (final exam score).

b) *Case*'ene er observasjonene $(y_i, x_{i1}, x_{i2}, \ldots)$ altså kombinasjonen av respons og forklaringsvariabler for hver enkelt observasjon. Vi har totalt $n = 166$ observasjoner.

c) Vi har $p = 7$ forklaringsvariabler.

d) Vi har forklaringsvariablene "math course anxiety", "math test anxiety", "numerical task anxiety", "enjoyment", "self-confidence", "motivation" og "perceived usefulness of the feedback sessions".

Oppgave 11.2

a) Vi har $\hat{y} = -10.8 + 3.2 \cdot 4 + 2.8 \cdot 2 = 7.6$.

b) Nei, det holder at forklaringsvariablene ligger i nærheten av observasjonene i data settet får å gi en fornuftig prediksjon $\hat{y}$.

c) For en fiksert x_1 vil $\hat{y}$ forandre seg med 2.8 enheter for hver enhet x_2 forandrer seg. Så for en forandring i x_2 på 3 enheter, vil $\hat{y}$ forandre seg med $2.8 \cdot 3 = 8.4$ enheter.

Oppgave 14.11

Kapittel 14 er ikke i den trykte utgaven, men en online versjon er tilgjengelig her http://bcs.whfreeman.com/webpub/statistics/ips9e/9781319013387/companionchapters/companionchapter14.pdf

a) Proporsjonen er $\hat{p} = 462/1003 = 0.46$.

b) Vi har odds $= \frac{\hat{p}}{1-\hat{p}} = \frac{0.4606}{1-0.4606} = 0.8539$.

Oppgave 14.13

a) Modellen er $\log(\text{odds}_i) = \log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_i$. Her er $x_i = 1$ hvis alder er mindre eller lik 25, $x_i = 0$ hvis alder er større enn 25, og p_i er sannsynligheten for at individet har brukt mobiltelefonen i en butikk for bla, bla, bla.

b) β_0 er log-oddsen for at personer over 25 år har brukt mobilen (bla, bla), mens β_1 er differansen i log-odds på personer under 25 år. Alternativt kan β_1 forklares med at $\frac{\text{odds}_{x=1}}{\text{odds}_{x=0}} = \exp(\beta_1)$. Altså har personer under 25 odds som er $\exp(\beta_1)$ ganger oddsen til personer over 25.

Oppgave 14.14

a) Vi har nå modellen $\log(\text{odds}_i) = \log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_i$, hvor x_i alder og p_i er sannsynligheten for at individet har brukt mobiltelefonen i en butikk for bla, bla, bla.

b) β_1 er nå gjennomsnittlig forandring i log-odds for hvert år. Det vil si at oddsen for å bruke mobilen (bla, bla) multipliseres med $\exp(\beta_1)$ of en økning på et år.

c) Denne modellen antar at det er et lineært forhold mellom alder og log-odds. Det var ikke tilfellet i Oppgave 14.13 etter som vi bare hadde en enkelt indikator variabel. For å undersøke denne antagelsen kan vi lage et plott tilsvarende det i Example 14.8, med alder langs x-aksen, og se om den estimerte linjen følger punktene. Merk at et slikt plott krever at vi har flere observasjoner for hver alder og at ikke alle observasjonene for en alder har samme respons, ettersom det vil gi en log-odds på $\pm\infty$.

Eksamenoppgaver: 2009 oppg. 2, 2015 oppg. 3, 2016 oppg. 3 (alle disse er fra kap 11)

Løsningsforslag til eksamensoppgavene er på emnets semesterside https://www.uio.no/studier/emner/matnat/math/STK1000/oppgaver/losningsforslag/

I tillegg:

Oppgave: Vis at gjennomsnittet $\bar{x}$ faktisk er en maksimum likelihood-estimator for forventningen μ i en $N(\mu, \sigma)$ fordeling, basert på n uavhengige, identisk fordelte observasjoner $x_1, x_2, \ldots, x_n$. NB! Da trenger du formelen for tetthetsfunksjonen, s. 57 i boken. La for enkelthets skyld σ være kjent.

Løsning:

La $\mathbf{x}$ representere $x_1, x_1, \ldots, x_n$ (for litt mer kompakt notasjon). Vi har da

$$
\begin{aligned}
f(x_i; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)} \\
L(\mu; \mathbf{x}) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma) \\
l(\mu; \mathbf{x}) &= \log(L(\mu; \mathbf{x})) \\
&= \sum_{i=1}^{n} \log(f(x_i; \mu, \sigma)) \\
&= \sum_{i=1}^{n} \left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + (-(x_i-\mu)^2/(2\sigma^2)) \right] \\
l'(\mu; \mathbf{x}) &= \sum_{i=1}^{n} \left[(-2(x_i-\mu)/(2\sigma^2)) \right]
\end{aligned}
$$

$$
\begin{aligned}
l'(\hat{\mu}, \mathbf{x}) &= 0 \\
\sum_{i=1}^{n} \left[(-2(x_i-\hat{\mu})/(2\sigma^2)) \right] &= 0 \\
\sum_{i=1}^{n} (x_i - \hat{\mu}) &= 0 \\
\sum_{i=1}^{n} x_i - n\hat{\mu} &= 0 \\
\sum_{i=1}^{n} x_i &= n\hat{\mu} \\
\frac{\sum_{i=1}^{n} x_i}{n} &= \hat{\mu} \\
\bar{x} &= \hat{\mu}
\end{aligned}
$$