Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Brukerkurs i statistikk, ST0103

1. desember, 2011

Løsningsforslag

Oppgave 1

a) La S være hendelsen at en innbygger sykler og B at en innbygger er bilist. DeMorgans setning, komplementsregelen og addisjonssetningen gir da at

$$\begin{aligned} P(\bar{S} \cap \bar{B}) &= P(\overline{S \cup B}) \\ &= 1 - P(S \cup B) \\ &= 1 - P(S) - P(B) + P(S \cap B) \\ &= 1 - 0.7 - 0.4 + 0.2 \\ &= 0.1 \end{aligned} \tag{1}$$

b) Betinget sannsynlighet for å sykle blant bilister blir

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{0.2}{0.7} = 0.286. \tag{2}$$

I følge lov om total sannsynlighet er $P(S) = P(S \cap B) + P(S \cap \bar{B})$ slik at $P(S \cap \bar{B} = P(S) - P(S \cap B) = 0.4 - 0.2 = 0.2$. Dermed blir betinget sannsynlighet for å sykle blant ikke-bilister

$$P(S|\bar{B}) = \frac{P(S \cap \bar{B})}{P(\bar{B})} = \frac{0.2}{1 - 0.7} = 0.667. \tag{3}$$

c) X blir hypergeometrisk fordelt med parametere $p = 0.7$, $N = 110$ og $n = 50$, forventning $EX = np = 35$ og varians $\mathrm{Var}\, X = np(1-p)\frac{N-n}{N-1} = 5.77$.

d)

$$P(X \geq 40) \approx P(X > 39.5) = P(\frac{X - 35}{2.40} > \frac{39.5 - 35}{2.40}) = 1 - G(1.87) = 0.0307 \tag{4}$$

Oppgave 2

a) Interpolert varians $S_P^2 = (S_X^2 + S_Y^2)/2 = 148$. Dette gir

$$T = \frac{\bar{X} - \bar{Y}}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{29.7 - 5.65}{\sqrt{148(\frac{1}{4} + \frac{1}{4})}} = 2.79 \tag{5}$$

Kritisk verdi for ensidig test blir $t_{0.05,4+4-2} = 1.94$. Vi kan dermed forkaste H_0 og konkludere med at DDT innholdet er høyere i oppdrettslaks. Testen forutsetter normalfordelte data og samme varians i begge grupper. De empiriske variansene antyder imidlertid at variansene er forskjellige.

b) 95%-konfidensintervall for differansen i gjennomsnittlig DDT-innholdet blant oppdretts- og villaks blir

$$29.7 - 5.65 \pm -2.44\sqrt{148(\frac{1}{4} + \frac{1}{4})} = (3.06, 45.03) \tag{6}$$

Oppgave 3

a) Likelihoodfunksjonen blir

$$L(\alpha) = \prod_{i=1}^{n} 2\alpha x_i e^{-\alpha x_i^2} = 2^n \alpha^n e^{-\alpha \sum x_i^2} \prod x_i, \tag{7}$$

og log-likelihoodet

$$\ln L(\alpha) = n \ln 2 + n \ln \alpha - \alpha \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} \ln x_i. \tag{8}$$

Likelihoodfunksjonen har sitt maksimum der

$$\begin{aligned} \frac{d}{d\alpha} \ln L &= 0 \\ \frac{n}{\alpha} - \sum_{i=1}^{n} x_i^2 &= 0 \\ \alpha &= \frac{n}{\sum_{i=1}^{n} x_i^2} \end{aligned} \tag{9}$$

Altså er $\hat{\alpha} = n/\sum_{i=1}^{n} X_i^2$ SME av α. Basert på observasjonene får vi estimatet $\hat{\alpha} = 0.00591$.

b) Siden

$$\frac{d}{dx}F(x) = \frac{d}{dx}(1 - e^{-\alpha x^2}) = 2\alpha x e^{-\alpha x^2} = f(x) \tag{10}$$

og $F(x) \to 0$ når $x \to -\infty$ er $F(x)$ kumulativ fordeling til X.

c)

$$P(X > x) = 1 - F(x) = e^{-\alpha x^2} = e^{-0.00591 \cdot 25^2} = 0.0247. \tag{11}$$

d)

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(X^2 \le y) = P(X \le y^{1/2}) = F_X(y^{1/2}) = 1 - e^{-\alpha (y^{1/2})^2} = 1 - e^{-\alpha y} \end{aligned} \tag{12}$$

som er kumulativ fordeling for eksponentiell fordeling med parameter α. Altså er

$$E(X^2) = E(Y) = \frac{1}{\alpha}. \tag{13}$$