



EKSAMENSOPPGAVE I STA-0001

Eksamen i : STA-0001 Brukerkurs i statistikk 1

Eksamensdato : Mandag 12. juni 2006

Tid : Kl 0900-1300

Sted : Åsgårdveien 9

Tillatte hjelpemidler : Alle trykte og skrevne samt kalkulator

Oppgavesettet er på 4 sider inkl. forside

Kontaktperson under eksamen: Alf Harbitz tlf 77609731 / 99345717

Deloppgavene vil telle likt i bedømmelsen

Oppgave 1

Gitt følgende høydedata i cm for $n = 27$ sta-0001 studenter:

Menn: 170 175 175 176 176 178 178 179 180 186

Kvinner: 153 161 161 162 163 164 164 166 168 171 173 174 176 178 180 180 182

- a) Beregn empirisk middelerverdi, standardavvik og median for de 27 studenthøydene.
- b) Velg 5-cm intervaller 151-155, 156-160, 161-165 osv. og lag en frekvenstabell. Hva blir midtpunktene i intervallene? Begrunn svaret.
Beregn gruppert middelerverdi og gruppert median.
- c) Tegn relativ frekvenshistogram for de grupperte dataene fra b). Ser du noe trekk ved formen på histogrammet som avviker fra normalfordelingen? Angi en grunn til at vi her ikke forventer en normalfordeling.

La $x_1, \dots, x_{10}$ betegne høyde på de 10 mannlige studentene, og anta at disse er uavhengige observasjoner fra en normalfordeling $N(\mu_1, \sigma)$. Videre, la $y_1, \dots, y_{17}$ betegne høyden på de 17 kvinnelige studentene, og anta at disse er uavhengige observasjoner fra en normalfordeling $N(\mu_2, \sigma)$. Videre skal vi anta at alle X 'ene og Y 'ene er uavhengige av hverandre.

- d) Test om forventet høyde, μ_1 , på de mannlige studentene er minst 5 cm over forventet høyde, μ_2 , til de kvinnelige studentene, på signifikansnivå $\alpha = 0.05$.

Regnehjelp: $\sum_{i=1}^{10} x_i = 1773$, $\sum_{i=1}^{10} x_i^2 = 314507$, $\sum_{i=1}^{17} y_i = 2876$, $\sum_{i=1}^{17} y_i^2 = 487666$

$$\text{Generelt: } \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n x_i \right)^2$$

Oppgave 2

En hobbysnekker skal legge takpanel. Hver plank skal være minst 4 m lang, fordi han ikke vil ha noen skjøt. Til sammen trenger han 80 plank på minst 4 m. Anta at leverandøren leverer plank med lengde X som er normalfordelt med forventning $\mu = 4.30$ m og standardavvik $\sigma = 0.20$ m.

- a) Hva er sannsynligheten for at en tilfeldig plank er minst 4 m lang?

La n være antall tilfeldige plank hobbysnekkeren kjøper, og la Y betegne hvor mange av disse n som er minst 4 m.

- b) Finn sannsynligheten for at hobbysnekkeren får minst 80 plank som er minst 4 m hvis han kjøper $n = 90$ plank.
- c) Kunden betaler kr. 1,00 pr. 10 cm plank, og totalt kr 200 for frakt av alt materialet. Hvilken fordeling har de totale utgiftene hobbysnekkeren da har ved materialkjøpet når han kjøper 90 plank? Angi både fordelingstype og parameterverdier.

Oppgave 3

En selbestand overvåkes årlig i form av telletokt. Med basis i telletoktene er det utviklet en estimator, $\hat{\mu}$, for hvor stor bestanden er, dvs. hvor mange dyr, μ , den består av. På grunn av indikasjoner den siste tiden på en raskt synkende bestand, har myndighetene bestemt å anvende føre var prinsippet ved å redusere fangstkvoteene dersom ikke data gir klar støtte for at selbestanden er større enn 200 000 dyr. Det anses for en langt alvorligere feil å konkludere at bestanden er større enn 200 000 når den i virkeligheten er mindre enn 200 000 enn å unnlate å konkludere at bestanden er større enn 200 000 når den i virkeligheten er større enn 200 000.

Vi skal i hele oppgaven anta at $\hat{\mu}$ er normalfordelt med forventning μ og standardavvik (ofte også kalt standardfeil, S.E.) $\sigma_e = SD(\hat{\mu}) = 30\,000$.

- a) Hvilken fordeling har testobservatoren $\frac{\hat{\mu} - \mu}{\sigma_e}$?

Inneværende år finner forskerne et estimat på $\hat{\mu} = 210\,000$. En miljøvernorganisasjon, Bluepeace, hevder at forskernes resultat viser at man på et signifikansnivå på 5 % ikke ville forkastet hypotesen at $\mu \leq 200\,000$, og krever at myndighetene derfor bør redusere fangstkvoteene.

- b) Gjennomfør en test på signifikansnivå $\alpha = 0.05$ for å undersøke om Bluepeace har støtte for sin påstand. Benytt testobservatoren fra a), og spesifiser H_0 , H_1 og forkastingsområdet, R , som en del av testen.
- c) Forklar med ord hva som er type I og type II feil under testen i b).

Noen selfangere er uenige i påstanden om at selbestanden er synkende, og hevder at de mener selbestanden ved neste telling vil være på minst 300 000 dyr.

- d) Vis at sannsynligheten er mindre enn 5 % for å unnlate å konkludere at selbestanden er større enn 200 000 dyr basert på neste telleresultat ($\hat{\mu}$) dersom den i virkeligheten er på $\mu = 300\,000$ dyr, med basis i testen i b).

Oppgave 4

En torskeforekomst består av to bestander: Kysttorsk (K) og skrei (S). Hvilken bestand en torsk tilhører bestemmes subjektivt ved visuell tolkning av torskens otolitt (ørestein). La K_{sub} og S_{sub} (subskript sub er forkortelse for subjektivt tolket) angi om torsken tolkes som henholdsvis kysttorsk eller skrei. Anta videre at en sikker identifisering er mulig ved hjelp av genetikk, og at man med basis i forsøk over flere år har funnet følgende simultanfordeling for sann og subjektivt tolket bestand:

		Sann bestand	
		K	S
Subjektivt tolket bestand	K_{sub}	0.36	0.06
	S_{sub}	0.04	0.54

- a) Bestem fra tabellen ovenfor følgende betingede sannsynligheter:
 $P(K_{\text{sub}} | K)$ og $P(K_{\text{sub}} | S)$

La $p = P(K)$ angi sann andel kysttorsk og la $1-p$ angi sann andel skrei i naturen for kommende år, der p generelt vil være forskjellig fra $p = 0.4$ som i tabellen ovenfor, mens sannsynlighetene fra a) antas å være uforandret.

- b) Vis at sannsynligheten, p_{sub} , for at en tilfeldig torsk kommende år skal bli tolket som kysttorsk er gitt som følger:

$$p_{\text{sub}} = P(K_{\text{sub}}) = 0.8p + 0.1$$

Anta at under et bestandstokt kommende år måles n tilfeldige torsk og at X av disse blir subjektivt tolket som kysttorsk. En intuitiv estimator, $\hat{p}_1$, for p vil da være

$$\hat{p}_1 = X / n.$$

Anta at X er binomisk fordelt med parametre n og p_{sub} , der p_{sub} er oppgitt i b).

- c) Hva blir forventningen $E(\hat{p}_1)$ uttrykt som funksjon av p ?
 For hvilken verdi av p blir $\hat{p}_1$ en forventningsrett estimator for p ?

Det kan vises at en estimator, $\hat{p}$, for p som er forventningsrett uavhengig av hvilken verdi p har er gitt som følger:

$$\hat{p} = \frac{X/n - 0.1}{0.8} = \frac{\hat{p}_1 - 0.1}{0.8}$$

- d) Konstruer et tilnærmet 95 % konfidensintervall for p når $x = 3695$ og $n = 9987$.