ECON 2130 EKSAMEN 2006 VÅR

SENSORVEILEDNING

Oppgaven består av 9 delspørsmål som anbefales å veie like mye, Kommentarer og tallsvar er skrevet inn mellom << …, >>,

Oppgave 1

I en by med 10 000 stemmeberettigete innbyggere pågår en heftig debatt om det skal etableres en bomring (dvs bommer på alle innfartsveiene der hver passering mot byen med bil koster en avgift) for å finansiere diverse trafikkprosjekter. På slutten av et debattprogram på lokal-TV ble seerne anmodet om å ringe inn (i løpet av programmet) og stemme for eller imot bomring. Av 5 000 innringere var 4 500 (altså 90%) imot bomring. For enkelthets skyld antar vi at tallet 10 000 bare omfatter byens stemmeberettigete og at bare stemmeberettigete ringte inn til programmet. Vi antar også at alle de stemmeberettigete har en mening for eller imot.

La M betegne motstander og T tilhenger av bomring. La S betegne det å delta i telefonavstemningen i TV og $\bar{S}$ det å ikke delta. Anta at de 10 000 stemmeberettigete er fordelt i de fire kombinasjonene av mening og stemmedeltagelse som angitt i tabell 1.

Tabell 1

	T (for)	M (mot)
S (avgir stemme)	500	4500
$\bar{S}$ (avgir ikke stemme)	2000	3000

a. For en tilfeldig valgt stemmeberettiget, la begivenhetene T, M og S stå for at vedkommende er tilhenger, motstander eller deltok i TV-avstemningen henholdsvis.

(i) Finn sannsynlighetene $P(M)$, $P(M \mid S)$, $P(M \mid \bar{S})$ og $P(M \cup S)$

(ii) Er M og S uavhengige? Begrunn svaret ditt.

<< (i) $P(M) = \frac{7500}{10000} = 0,75$, $P(M \mid S) = \frac{4500}{5000} = 0,90$, $P(M \mid \bar{S}) = \frac{3000}{5000} = 0,60$

$P(M \cup S) = \frac{8000}{10000} = 0,80$ (ii) Nei siden $P(M) \neq `P(M \mid S)$ >>

---

b. Anta vi trekker to personer rent tilfeldig fra populasjonen av stemmeberettigete. La X være antall blant de to som er M (dvs. motstander av bomring). Anta X er binomisk fordelt.

(i) Begrunn antakelsen at X er binomisk fordelt – i hvert fall tilnærmet. Hva blir forventningen til X?
(ii) Beregn sannsynligheten for at de er enige (dvs. begge for eller begge imot bomring).

<< (i) X ~ binomisk (n = 2, p = 0,75). E(X) = 1,5.

(ii) $P(enige) = P(X = 0) + P(X = 2) = (0,25)^2 + (0,75)^2 = 0,625$ >>

---

c. (*Mer krevende*). Tallene i første rad i **tabell 1** (500 og 4500) er kjente fra TV-avstemningen, som impliserer at $P(M \mid S) = 0,9$. Tallene i annen rad (2000 og 3000) er imidlertid bare gjetninger for regneeksemplets skyld. De er i virkeligheten ukjente. Sett $p = P(M)$ som er ukjent og av spesiell interesse. Hvilke tall måtte det ha stått i annen rad i tabell 1 (istedenfor 2000 og 3000) hvis TV-avstemningen hadde vært representativ i betydningen $p = P(M \mid S) = 0,9$?

<< La a være tallet i celle $\bar{S} \cap M$. Vi må da ha $0,9 = P(M) = \dfrac{4500 + a}{10000}$, som gir

$a = 4500$ (og dermed 500 = 5000 – 4500 i celle $\bar{S} \cap T$). >>

Oppgave 2

For å måle graden av over- eller undervekt hos mennesker brukes ofte *kroppsmasseindeksen* (KMI), som definers som $(vekt)/(høyde)^2$ der vekt er målt i kilo (kg) og høyde i meter (m). Normalområdet for KMI regnes fra 20 til 25 kg/m^2. En KMI under 20 regnes som “undervekt” og en KMI over 25 som “overvekt”.

I denne oppgaven betyr “ung kvinne” en kvinne i aldersgruppen 20 – 29 år.

a. La X være KMI for en tilfeldig valgt kvinne fra populasjonen av unge kvinner i Norge. Vi antar at X er normalfordelt med forventning $\mu = 22,4$ og standardavvik $\sigma = 3,4$ (dvs. $X \sim N(22,4\,,\ 3,4)$), der parameterverdiene er hentet fra en norsk kostholdsundersøkelse fra 1997.

(**i**) Vis at sannsynligheten for at en tilfeldig valgt ung kvinne er undervektig (KMI < 20), er 0,239.

(**ii**) Finn nedre desil (d_1) definert ved $P(X \le d_1) = 0,10$.

<< (i) $P(X < 20) = P(X \le 20) = G\left(\dfrac{20-22,4}{3,4}\right) = G(-0,71) = 0,2389$

(ii) $P(X < d_1) = G\left(\dfrac{d_1-22,4}{3,4}\right) = 0,10 \quad \Leftrightarrow \quad \dfrac{Q_1-22,4}{3,4} = -1,28 \quad (-1,282)$

$\Leftrightarrow \quad d_1 = 22,4 - 1,28 \times 3,4 = 18,0$ >>

---

b. Betrakt et tilfeldig utvalg av 4 unge norske kvinner, og la X_i være KMI for kvinne nr. i i utvalget. Anta at X_1, X_2, X_3, X_4 er uavhengige og identisk fordelte med samme fordeling som X i punkt **a**..

(**i**) Finn $P(\bar{X} < 20)$ der $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{4} X_i$.

(**ii**) Hva er sannsynligheten for at minst tre av de fire kvinnene i utvalget har KMI < 20?

<< (i) $\bar{X} \sim N(22,4,\ 3,4/2) = N(22,4,\ 1,7)$ som gir

$P(\bar{X} < 20) = G\left(\dfrac{20-22,4}{1,7}\right) = G(-1,41) = 0.0793$

(ii) U = antall i utvalget med KMI < 20, er ~ Bin(4, p = 0,239).

$P(U = 3) = 4p^3(1-p) = 0,042$, $\quad P(U = 4) = p^4 = 0,003$, som gir

$P(U \ge 3) = 0,045$ >>

---

c. En gruppe på 15 kvinnelige studenter ved en idrettshøyskole diskuterer om det er et urimelig slankepress på kvinnelige idrettsutøvere. Til å begynne med registrerte de sin egen KMI og fant gjennomsnittet for gruppen, $\bar{x} = 21,5$, og empirisk standardavvik, $s = 3,9$. Gruppen tar utgangspunkt i følgende modell:

$X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte med $X_i \sim N(\mu, \sigma)$, der n = 15, og der μ, σ begge antas ukjente. Her representerer X_i KMI for person nr. i i gruppen.

Beregn et 95% konfidensintervall for μ under disse betingelsene.

Drøft kort rimeligheten av modellen og eventuelt tilleggsforutsetninger som trengs for å kunne tolke μ som gjennomsnittlig KMI i populasjonen av yngre kvinnelige idrettsutøvere i Norge.

<< t-intervall med 14 frihetsgrader. Et 95% KI blir

$\bar{X} \pm t_{0,025,14} \dfrac{S}{\sqrt{15}} = 21,5 \pm 2,145 \dfrac{3,9}{\sqrt{15}} = 21,5 \pm 2,2 = [19,3,\ \ 23,7]$. Man må naturligvis forutsette at utvalget er representativt som om det var et tilfeldig utvalg trukket fra populasjonen. Også forutsetningen om uavhengighet kan være i fare om det har vært en sterk gjensidig påvirkning innen gruppen. >>

---

d. Gruppen ønsker å teste $H_0:\ \mu \geq \mu_0$ mot $H_1:\ \mu < \mu_0$ der $\mu_0 = 22,4$ som er brukt i punkt **a**. I håp om å øke utsagnskraften, finner gruppen å kunne forutsette at σ er kjent lik 3,4. Forøvrig er forutsetningene som i punkt **c**. Sett opp en test med signifikansnivå 5% som utnytter at σ er kjent. Gjennomfør testen med dataene gitt i punkt **c.** og formuler en konklusjon. Beregn også testens p-verdi.

<< Alternative ekvivalente formuleringer i tabellen

Testobservator	Kritisk verdi	Observert	Konklusjon
$\bar{X}$	$22,4 - 1,645 \dfrac{3,4}{\sqrt{15}} = 20,96$	$\bar{x}_0 = 21,5$	Ikke forkast H_0
$Z = \dfrac{\bar{X} - 22,4}{3,4} \sqrt{15}$	$-z_{0,05} = -1.645$	$z_o = -1,025$	Ikke forkast H_0

P-verdi: $\hat{\alpha} = P_{\mu_0}(Z \le z_o) = P(Z \le -1,03) = 0,1515$ >>

---

e. Gitt testen i punkt **d**., der antall observasjoner er 15. Hva blir sannsynligheten for feil av type I og sannsynligheten for feil av type II hvis $\mu = \mu_0 = 22,4$? Hva blir de to feilsannsynlighetene hvis $\mu = 21,5$?

<< $\mu = 22,4 \quad \Rightarrow \quad P(\text{feil I}) = 0,05 \quad \text{og} \quad P(\text{feil II}) = 0$

$\mu = 21,5 \quad \Rightarrow \quad P(\text{feil I}) = 0 \quad \text{og}$

$$P(\text{feil II}) = 1 - P(\text{forkast H}_0) = 1 - P_\mu(\bar{X} \le 20,96) = 1 - G\left(\frac{20,96 - 21,5}{3,4}\sqrt{15}\right) =$$

$$= 1 - G(-0,61) = 1 - 0,7257 = 0,2743$$ >>

---

f. La nå antall observasjoner, *n*, være vilkårlig. For øvrig la modell (med kjent σ) og hypoteser være som i punkt **d**. Sett opp testobservator og forkastningområde når *n* er vilkårlig, der signifikansnivået fortsatt er 5%. Sett også opp et uttrykk for styrkefunksjonen, $\gamma(\mu)$. Hvor stor må *n* være for at styrkefunksjonen for testen skal ha en verdi på minst 0,95 hvis den sanne verdien av μ er 21,5?

<< Kriterium: Forkast H_0 hvis $\bar{X} \le \mu_0 - 1,645\frac{3,4}{\sqrt{n}} \quad (= k)$

Styrkefunksjonen:

$$\gamma(\mu) = P_\mu(\bar{X} \le k) = G\left(\frac{\mu_0 - 1.645(3,4/\sqrt{n}) - \mu}{3,4/\sqrt{n}}\right) = G\left(\frac{\mu_0 - \mu}{3,4}\sqrt{n} - 1.645\right)$$

og

$$\lambda(21,5) \ge 0,95 \quad \Leftrightarrow \quad G\left(\frac{0,9}{3,4}\sqrt{n} - 1,645\right) \ge 0,95 \quad \Leftrightarrow$$

$$\frac{0,9}{3,4}\sqrt{n} - 1,645 \ \ge \ 1,645 \quad \Leftrightarrow \quad \sqrt{n} \ge \frac{2(1,645)(3,4)}{0,9} = 12,43 \quad \Leftrightarrow \quad n \ge 154,5$$

dvs. $n \ge 155$ >>
