ECON 2130 EKSAMEN 2012 VÅR

TALLSVAR

Det anbefales at de 9 deloppgavene merket med A, B,… teller likt – uansett variasjon i vanskelighetsgrad. Svarene er gitt i << … >>.

Oppgave 1

Tre ektepar kommer sammen til en quiz-konkurranse. De 6 personene inndeles helt tilfeldig i tre lag der hvert lag består av en kvinne og en mann. Kvinnene betegnes med tallene, 1, 2 og 3, og mennene med M_1, M_2 og M_3, på en slik måte at j er gift med M_j, for $j = 1,2,3$. Siden inndelingen i lag er tilfeldig, kan noen av ekteparene ende opp som lag mens andre ikke gjør det.

Selve uttrekningen av lag skjer på følgende måte: Mennene stilles opp ved siden av hverandre i en fast rekkefølge, M_1, M_2, M_3. Deretter trekkes en tilfeldig rekkefølge av kvinner som stilles opp foran mannerekken. To som står rett overfor hverandre i denne oppstillingen utgjør et lag. Rekkefølgen av kvinner trekkes slik at alle de 6 mulige rekkefølgene er like sannsynlige. I **tabell 1** er to slike oppstillinger angitt. For eksempel, hvis kvinnerekkefølge 1 i tabellen blir trukket ut, blir sammensetningen av de tre lagene, $(3, M_1), (2, M_2)$ og $(1, M_3)$. I denne lagoppstillingen opptrer et ektepar som lag, nemlig ekteparet $(2, M_2)$. Blant lagene basert på kvinnerekkefølge 2 i tabellen er det ingen ektepar.

Tabell 1

Kvinne-rekkefølge	M_1	M_2	M_3	Antall ektepar blant lagene
1	3	2	1	1
2	3	1	2	0
3				
4				
5				
6				

A. **i.** Fyll ut hele tabell 1 med alle de mulige kvinnerekkefølgene. Fyll også ut siste kolonne som viser antall ektepar blant lagene for de forskjellige kvinnerekkefølgene.

ii. La A_j ($j = 1,2,3$) betegne begivenheten at ektepar j (dvs. paret (j, M_j)) blir valgt ut som lag. Vis at $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.

<< **Svar: i:** Utfylt

Kvinne-rekkefølge	M_1	M_2	M_3	Antall ektepar blant lagene
1	3	2	1	1
2	3	1	2	0
3	2	1	3	1
4	2	3	1	0
5	1	2	3	3
6	1	3	2	1

ii. Vi ser at hvert ektepar opptrer to ganger blant de 6 oppstillingene. Dermed

$$P(A_j) = \frac{2}{6} = \frac{1}{3}, \quad j = 1,2,3$$ >>

B. i. Finn sannsynlighetene, $P(A_1 \cap A_2)$, $P(A_1 \cup A_2)$ og $P(A_2 \mid A_1)$.

ii. Finn sannsynligheten for at *enten* ektepar 1 eller 2, men ikke begge, blir valgt ut som lag.

<< **Svar:**

i: $$P(A_1 \cap A_2) = P(\text{rekke } 5) = \frac{1}{6},$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{2}{3} - \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

$$P(A_2 \mid A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{1/6}{2/6} = \frac{1}{2}$$

ii. Begivenheten *enten* A_1 *eller* A_2, inntreffer for kvinnerekkefølge 1 og 6.
Sannsynligheten blir dermed, $2/6 = 1/3$. >>

C. i. La X være antall ektepar som er med blant de uttrukne lagene. Bestem sannsynlighetsfordelingen for X.

ii. Beregn forventningen og variansen til X.

[**Hint:** Hvis du ikke har funnet de riktige sannsynlighetene i fordelingen for X, er det OK om du rett og slett bare velger (gjetter) fritt noen sannsynligheter og gjennomfører beregningen ut fra disse.]

<< **Svar:**

i. |Fordelingen for X følger rett fra den utfylte tabell 1 og blir

x	0	1	3
$P(X=x)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

ii. Finner $E(X)=\sum_x xP(X=x)=1,\quad E\left(X^2\right)=\sum_x x^2P(X=x)=2$ og

$\mathrm{Var}(X)=E\left(X^2\right)-\left[E(X)\right]^2=1$ >>

Oppgave 2

Ifølge elevundersøkelsen 2011 er ca 8.3% av skoleelever i Norge fra 5. skoletrinn og oppover utsatt for gjentatt mobbing på skolen – en prosent som har holdt seg omtrent på samme nivå i de senere årene siden 2007. Mobbeprosenten er høyest på de laveste skoletrinnene (ca 11-12% på 5. skoletrinn) og lavest på videregående (ca 6%).

Elevundersøkelsen arrangeres hvert år i regi av utdanningsdirektoratet og omfatter en rekke spørsmål til elevene om skolesituasjonen. Resultatene bygger på svar på spørreskjemaer fra et stort utvalg av elever (over 50% av elevmassen).

I denne oppgaven skal vi konsentrere oss om 5. skoletrinn, som i alt omfattet $N=62\,000$ elever i 2011. Utvalget fra 5. trinn som besvarte spørreskjemaet, besto av $n=33\,045$ elever. La X betegne antall "mobbeofre" i utvalget (definert som antallet som på spørsmålet: Er du blitt mobbet på skolen de siste månedene?, angir et av svaralternativene: 2 eller 3 ganger i måneden, omtrent en gang i uken, eller flere ganger i uken). Antall mobbeofre i utvalget ble $X_{obs}=3\,866$ (der indeksen obs på en stokastisk variabel indikerer den observerte verdien av variabelen).

For enkelthets skyld antar vi at utvalget er rent tilfeldig slik at X kan antas å være hypergeometrisk fordelt (N,M,n), der M står for antall mobbeofre i populasjonen (5. skoletrinn i 2011). La $p=M/N$ betegne populasjonsandelen av mobbeofre som er ukjent.

A. **i.** Sett opp en forventningsrett estimator ($\hat{p}$) for p basert på X og forklar hvorfor den er forventningsrett.

ii. Forklar hvordan (den teoretiske) standardfeilen til $\hat{p}$, $\sqrt{\frac{p(1-p)}{n}\cdot\frac{N-n}{N-1}}$, følger av formelen for variansen til X: $\quad \mathrm{var}(X) = np(1-p)\frac{N-n}{N-1}$.

iii. Beregn estimatet for p og estimert standardfeil på grunnlag av de oppgitte data.

<< **Svar:** **i:** $\hat{p} = X/n$. Definisjon av den hypergeometriske fordelingen i boka gir $EX = np$, hvorav $E\hat{p} = p$.

ii: Siden $\hat{p}$ er forventningsrett, er standardfeilen lik standardavviket, hvorav

$$SE(\hat{p}) = \sqrt{\mathrm{var}(X/n)} = \sqrt{\mathrm{var}(X)/n^2} = \sqrt{\frac{p(1-p)}{n}\cdot\frac{N-n}{N-1}}$$

iii: Estimat: $\hat{p}_{obs} = 3\,866/33\,045 = 0.117$. Finner korreksjonsfaktoren $\frac{N-n}{N-1} = \frac{62000-33045}{61999} = 0.467$, og dermed estimert standardfeil,

$$SE(\hat{p}) = \sqrt{\frac{0.117*(1-0.117)}{33045}\cdot(0.467)} = 0.0012$$

\>>

B. Beregn et tilnærmet 99% konfidensintervall for p basert på normalfordelingstilnærmelsen for en hypergeometrisk fordeling.

<< **Svar:** For konfidensgrad 99% trenger vi kvantilen, $z_{0.005} = 2.576$ i $N(0,1)$ -fordelingen. Formelen er gitt i KI-oversikten på nettet

$$\left[\hat{p} \pm z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\cdot\frac{N-n}{N-1}}\right]_{obs} = \left[0.117 \pm (2.576)(0.0012)\right] =$$

$$\left[0.117 \pm 0.003\right] = \left[0.114,\ 0.120\right]$$

\>>

C. I resten av oppgaven skal vi forlate den faktiske undersøkelsen referert til ovenfor og tenke oss i stedet en gitt (ikke-eksisterende) kommune som tilsammen har $N = 1200$ elever på 5. skoletrinn i 2011. La p bety mobbeandelen totalt i den aktuelle kommunen der p er ukjent. Anta at man, på grunn av erfaringer fra tidligere år, mistenker at mobbeprosenten (p) faktisk kan ligge over 12% i denne kommunen. Vi ønsker å teste

(1) $\quad H_0 : p \le 0.12$ mot alternativet $H_1 : p > 0.12$.

Anta videre at utvalgsstørrelsen fra kommunen var på $n = 700$, og at antall mobbeofre i utvalget var $X_{obs} = 91$. For øvrig antas at X (antall mobbeofre i et tilfeldig utvalg på n

for den aktuelle kommunen) er hypergeometrisk fordelt, og at forutsetningene for tilnærming til normalfordeling gjelder for fordelingen til *X*.

i. Gjennomfør en test for problemet i (1) med signifikansnivå (tilnærmet) 5% og formuler en konklusjon.
ii. Beregn signifikanssannsynligheten ("p-verdien") for testen din. Si med ord hva denne (p-verdien) er uttrykk for.

<< Svar:

i: Testobservator: $Z = \dfrac{\hat{p} - 0.12}{\sqrt{\dfrac{(0.12)(0.88)}{n} \cdot \dfrac{N-n}{N-1}}}$

Testkriterium (5%): Forkast H_0 hvis $Z > 1.645$.

Estimat: $\hat{p}_{obs} = 91/700 = 0.130$

Korreksjonsfaktor: $\dfrac{N-n}{N-1} = \dfrac{1200-700}{1199} = 0.417$

Standardfeil når $p = 0.12$: $SE(p_0) = 0.0079$

Observert testobservator: $Z_{obs} = 1.26$. Konklusjon: Ikke forkast H_0.

ii: P-verdi = $\hat{\alpha} = P_{p=0.12}(Z > Z_{obs}) = P_{p=0.12}(Z > 1.26) \overset{\text{Tabell}}{\approx} 1 - 0.8962 = 0.104$

P-verdien er uttrykk for ….

>>

D. Innledning om mulig skjevhet i utvalget.

Ikke alle elever som får tilsendt spørreskjema svarer på skjemaet. For eksempel lå den gjennomsnittlige svarprosenten i den faktiske elevundersøkelsen på rundt 85%. Hvis det er sammenheng mellom det å være villig til å svare på skjemaet og tendensen til å bli mobbet, kan dette lede til skjevhet i estimeringen av *p*. Det kan for eksempel tenkes at mobbeofre er mer motivert til å svare på skjemaet enn andre - slik at det er relativt sett flere mobbeofre blant dem som svarer på skjemaet enn blant dem som ikke svarer. Det kan også tenkes at mobbing slår ut den andre veien for noen.

La *M* nå betegne begivenheten at en tilfeldig valgt elev fra kommunen er et mobbeoffer, og *S* begivenheten at vedkommende er villig til å svare på spørreskjemaet. Da blir $P(M) = p$ som vi ønsker å estimere. Hvis svarprosenten i kommunen er $100 \cdot q$ %, kan vi anta $P(S) = q$. Sett $p_1 = P(M \mid S)$ og $p_2 = P(M \mid \overline{S})$, der $\overline{S}$ står for begivenheten "ikke-S". Siden *S* nødvendigvis må ha inntruffet for alle i utvalget, kan det vises at estimatoren $\hat{p}$ blir forventningsrett for $p_1 = P(M \mid S)$ istedenfor *p*. Hvis derfor $p_1 \neq p$, vil ikke $\hat{p}$ lenger være forventningsrett for p. I så fall sies $\hat{p}$ å være en *skjev* estimator for p, og absoluttverdien $|p_1 - p|$ kalles *skjevheten*.

Oppgave.

i. Gjør rede for at $\hat{p} = X/n$ er en skjev estimator for p hvis og bare hvis M og S er stokastisk avhengige begivenheter.

ii. Vis at sammenhengen mellom p og p_1, p_2, q er gitt ved

(2) $$p = q \cdot p_1 + (1-q) \cdot p_2$$

<< **Svar:**

i: $p \neq p_1 \Leftrightarrow P(M) \neq P(M \mid S) \Leftrightarrow P(M \cap S) \neq P(M)P(S)$ (dvs. avhengighet).

ii: (2) betyr det samme som $P(M) = P(S)P(M \mid S) + P(\overline{S})P(M \mid \overline{S})$. Denne følger av å splitte opp M ved $M = (M \cap S) \cup (M \cap \overline{S})$, som er en disjunkt union. Derfor $P(M) = P(M \cap S) + P(M \cap \overline{S}) = P(S)P(M \mid S) + P(\overline{S})P(M \mid \overline{S})$. >>

E. Anta at $q = 0.85$ og at forskjellen i sannsynligheten for å bli mobbet blant dem som er villige til å svare på skjemaet og blant dem som ikke vil det, ikke overskrider 2 prosentenheter (dvs. anta $|p_1 - p_2| \leq 0.02$). Vis at skjevheten i så fall ikke kan bli større enn 0.003.

[**Hint:** Bruk (2) til å vise at $p_1 - p = (1-q)(p_1 - p_2)$]

<< **Svar:** Vi får fra (2)

$$p_1 - p = p_1 - qp_1 - (1-q)p_2 = (1-q)p_1 - (1-q)p_2 = (1-q)(p_1 - p_2)$$

Dermed, hvis $q = 0.85$,

$$|p_1 - p| = (1-q)|p_1 - p_2| \leq (1-q)(0.02) = (0.15)(0.02) = 0.003$$ >>

F. i. Anta nå at estimatoren $\hat{p} = X/n$ *ikke* er skjev (dvs. anta $p_1 = p$). Beregn styrken (tilnærmet) i punktet $p = 0.135$ for testen i punkt **C**. Med andre ord, beregn $P(\text{forkast } H_0)$ hvis den sanne verdien av p er 0.135.

[**Hint:** Utnytt at $\hat{p}$ er tilnærmet normalfordelt,

$$\hat{p} \overset{\text{tilnærmet}}{\sim} N\left(p, \sqrt{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}}\right) \quad \text{hvis} \quad p = 0.135 \quad]$$

ii. Bestem sannsynligheten for å begå feil av type I og sannsynligheten for å begå feil av type II hvis den sanne verdien av *p* er 0.135.

<< **Svar: i:**

Sett $SE_0 = \sqrt{\frac{(0.12)(0.88)}{n} \cdot \frac{N-n}{N-1}} = 0.0079$ og $SE_p = \sqrt{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}} = 0.00834$ for $p = 0.135$. Sett $U = \frac{\hat{p} - p}{SE_p}$. Da er U tilnærmet $\sim N(0,1)$ og $P(U \le u) \approx G(u)$ som er tabulert i tabell D3 i Løvås.

Dermed, hvis sann $p = 0.135$, får vi

$$P(\text{forkast } H_0) = P(Z > 1.645) = P(\hat{p} > 0.12 + (1.645)SE_0) =$$

$$P\left(\frac{\hat{p} - p}{SE_p} > \frac{0.12 + (1.645)SE_0 - p}{SE_p}\right) = P\left(U > \frac{0.12 + (1.645)SE_0 - p}{SE_p}\right) \approx$$

$$1 - G\left(\frac{0.12 + (1.645)SE_0 - p}{SE_p}\right) = 1 - G(-0.23) = 1 - 0.4090 = 0.591$$

ii: Siden $p = 0.135$ hører med under alternativet, H_1, blir $P(\text{feil type I}) = 0$, og $P(\text{feil type II}) = 1 - P(\text{forkast } H_0) = 0.409$
>>
