ECON 2130 EKSAMEN 2011 VÅR

TALLSVAR

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

Oppgave 1

En bedrift ønsker å fordele 2 kontrakter i et investeringsprosjekt helt tilfeldig på 3 firmaer, A, B og C. Utvelgelsen skjer ved loddtrekning. Loddtrekningen er slik at hvert av firmaene A, B og C, har en mulighet til få 0, 1 eller begge kontraktene.

La (*i*,*j*) betegne utfallet at kontrakt 1 går til firma *i* og kontrakt 2 til firma *j* (der i og j står for A, B eller C). Utfallet (A, A) betyr for eksempel at A fikk begge kontraktene, (A, B) betyr at A fikk kontrakt 1 og B fikk kontrakt 2, mens (B, A) betyr omvendt at B fikk kontrakt 1 og A kontrakt 2. Anta at loddtrekningen er slik at alle de 9 mulige utfallene av (*i*,*j*) er like sannsynlige.

A. **(i)** Skriv opp listen over de 9 mulige utfallene av (*i*,*j*) – det vil si kompletter listen, $(A, A), (A, B), (A, C), \ldots\ldots$

(ii) Hva er sannsynligheten for at A og B får en kontrakt hver?

(iii) Hva er sannsynligheten for A får minst en kontrakt?

(iv) Anta vi vet at A har fått minst en kontrakt. Hva er da sannsynligheten for at A fikk begge kontraktene?

<< **(i)** $(A, A), (A, B), (A, C), (B, A), (B, B), (B, C), (C, A), (C, B), (C, C)$

(ii) $2/9$ **(iii)** $5/9$

(iv)
$$P(A \text{ får begge } \cap A \text{ får minst en}) / P(A \text{ får minst en}) =$$
$$= P(A \text{ får begge}) / P(A \text{ får minst en}) = 1/5$$

>>

B. La X være antall kontrakter A får, og Y antall kontrakter B får. For eksempel, hvis utfallet av loddtrekningen er (A, A), blir $X = 2$ og $Y = 0$, hvis utfallet er (A, B), blir $X = Y = 1$, osv.

(i) Gjør rede for at den simultane fordelingen for X og Y er gitt ved tabell 1.

[Hint: Lag først en tabell over hvilke verdier X og Y får for hvert av de 9 utfallene i punkt **A(i).]**

Tabell 1 $f(x,y) = P(X = x \cap Y = y)$

	y		
x	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

(ii) Sett opp de marginale fordelingene til X og Y. Er X og Y uavhengige? Er X og Y identisk fordelte?

(iii) Finn $P(X = Y)$ og $P(X \geq Y)$.

<< **(i)**

Utfall	*(A,A)*	*(A,B)*	*(A,C)*	*(B,A)*	*(B,B)*	*(B,C)*	*(C,A)*	*(C,B)*	*(C,C)*
X	2	1	1	1	0	0	1	0	0
Y	0	1	0	1	2	1	0	1	0

(ii)

x	0	1	2
$P(X = x)$	4/9	4/9	1/9

Y har samme fordeling. X og Y er avhengige og identisk fordelte.

(iii) $P(X = Y) = f(0,0) + f(1,1) = 3/9$. Tilsvarende $P(X \geq Y) = 6/9$.

>>

C. **(i)** Beregn $E(X)$, $\text{var}(X)$, $E(Y)$, $\text{var}(Y)$

(ii) Beregn korrelasjonskoeffisienten mellom X og Y.

<< **(i)** $E(X) = 1 \cdot \frac{4}{9} + 2 \cdot \frac{1}{9} = \frac{2}{3}$, $E\left(X^2\right) = 1 \cdot \frac{4}{9} + 4 \cdot \frac{1}{9} = \frac{8}{9}$, $\text{var}(X) = \frac{8}{9} - \left(\frac{2}{3}\right)^2 = \frac{4}{9}$

Y har samme forventning og varians siden X og Y har samme fordeling.

(ii) $E(XY) = \sum_{x,y} xyf(x,y) = \frac{2}{9}$, $\text{cov}(X,Y) = \frac{2}{9} - \frac{2}{3} \cdot \frac{2}{3} = -\frac{2}{9}$. Dermed

$$\rho = \rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{-2/9}{\sqrt{(4/9)^2}} = -\frac{1}{2}$$

>>

D. Anta nå at bedriften, istedenfor 3, hadde 10 firmaer å velge mellom når de to kontraktene skulle fordeles. Firma A er en av disse. Loddtrekningen foregår i to trinn. Først trekkes firmaet som skal få kontrakt 1 rent tilfeldig fra de 10 firmaene. På trinn 2 trekkes firmaet som skal få kontrakt 2 også rent tilfeldig fra de 10 (uansett hva som er utfallet på trinn 1).

Dermed er det fortsatt mulig for ett og samme firma å få begge kontraktene. La X være antall kontrakter som firma A får ved loddtrekningen.

(i) Gjør rede for at X er binomisk fordelt.
(ii) Beregn punktsannsynlighetene i fordelingen for X.

<< **(i)** To uavhengige forsøk, med konstant suksess-sanns.het, $p = 1/10 = 0.1$. Dermed $X \sim \text{bin}(2, 0.1)$.

(ii) $P(X = x) = \binom{2}{x}(0.1)^x(0.9)^{2-x}, \quad x = 0,1,2$ gir

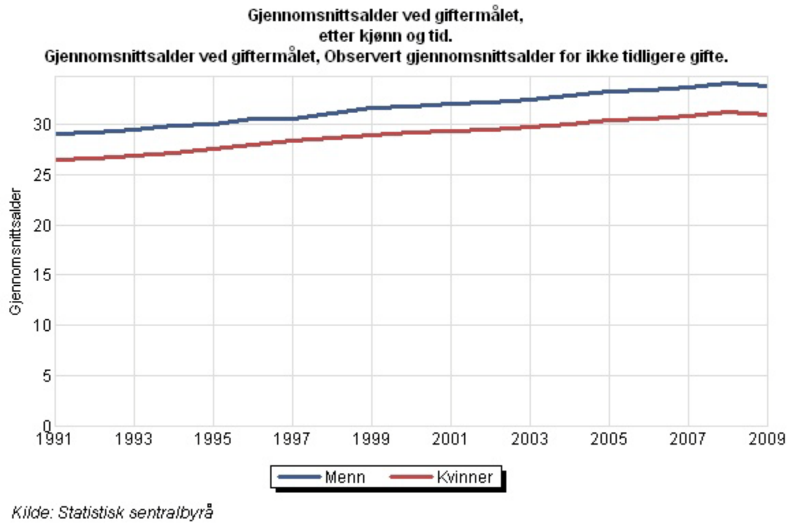
x	0	1	2
$P(X = x)$	0.81	0.18	0.01

>>

Oppgave 2

Det er velkjent at mannen i gjennomsnitt bruker å være noe eldre enn kvinnen når de gifter seg første gang. Figur 1 fra Statistisk sentralbyrå (SSB) bekrefter dette for perioden 1991 til 2009. I tillegg gir figuren inntrykk av at den gjennomsnittlige aldersdifferansen synes å ligge nokså stabilt på rundt 2-3 år. Til tross for inntrykket figuren gir er det ting som tyder på at den gjennomsnittlige aldersdifferansen kan ha økt noe den senere tiden sammenlignet med 90-tallet. Vi skal i denne oppgaven undersøke dette litt nærmere, og vi vil benytte tre forskjellige metodiske tilnærmelser til problemet. Datagrunnlaget for figur 1 er gitt i tabell 2.

Figur 1



Tabell 2 Gjennomsnittsalder ved giftermålet for ikke tidligere gifte.

År	Gj.snittlig alder menn	Gj.snittlig alder kvinner	Differanse (x_i)	År	Gj.snittlig alder menn	Gj.snittlig alder kvinner	Differanse (y_i)
				2000	31.8	29.2	2.6
1991	29.0	26.4	2.6	2001	32.0	29.3	2.7
1992	29.2	26.6	2.6	2002	32.2	29.5	2.7
1993	29.4	26.9	2.5	2003	32.5	29.7	2.8
1994	29.8	27.2	2.6	2004	32.9	30.0	2.9
1995	30.0	27.6	2.4	2005	33.2	30.4	2.8
1996	30.5	28.0	2.5	2006	33.4	30.6	2.8
1997	30.6	28.3	2.3	2007	33.7	30.8	2.9
1998	31.1	28.6	2.5	2008	34.1	31.2	2.9
1999	31.6	28.9	2.7	2009	33.8	31.0	2.8
	Gjennomsnitt ($\overline{x}$)		2.52		Gjennomsnitt ($\overline{y}$)		2.79
	Standardavvik (s_x)		0.1202		Standardavvik (s_y)		0.0994

Til tross for at differansene i tabell 2 er basert på fullstendige tellinger av giftermål i Norge, er det rimelig å tro at det er tilfeldigheter knyttet til dem, siden det sjelden er forutbestemt når og med hvem folk gifter seg. Vi kunne for eksempel forestille oss at de varierer stokastisk rundt en felles forventningsverdi i perioden 91-99, og tilsvarende for periode 00-10.

Som utgangsmodell (for tilnærming 1 og 2) antar vi således at differansene i tabell 2 er observasjoner av stokastiske variable, $X_1, X_2, \ldots, X_9$, for årene 1991-1999, og $Y_1, Y_2, \ldots, Y_{10}$ for 2000-2010. De 19 variablene antas å være uavhengige og normalfordelte med samme ukjente varians, σ^2. I tillegg antas X_i-ene å ha felles forventning, μ_1, og Y_i-ene felles forventning, μ_2, der μ_1 og μ_2 ikke behøver å være like. Problemet er nå å finne ut om det er grunnlag i data for å hevde at $\mu_1 < \mu_2$, eller ekvivalent om $\mu_2 - \mu_1 > 0$.

A. Ved første tilnærming antar vi for enkelthets skyld at det kan anses som kjent at $\mu_1 \leq 2.7$. I så fall koker problemet ned til om det er grunnlag i data til å påstå at $\mu_2 > 2.7$.

Vi skal altså teste $H_0 : \mu_2 \leq 2.7$ mot $H_1 : \mu_2 > 2.7$. Sett opp et test-kriterium og gjennomfør testen basert på informasjonen i tabell 2. Bruk signifikansnivå 1% og formuler en konklusjon.

<< T-test. $\hat{\mu}_2 = \overline{Y}$ er forventningrett og normalfordelt. Testobservator,

$$T = \frac{\hat{\mu}_2 - 2.7}{S_y}\sqrt{10} \sim t_9$$ -fordelt når $\mu_2 = 2.7$. Observert verdi: $T_{obs} = 2.863$ og kritisk verdi, $t_9(0.01) = 2.821$ gir forkastning. Det er sterk evidens i data for at $\mu_2 > 2.7$.

\>>

B. Ved tilnærming 2 aksepteres ikke at hypotesen, $\mu_1 \leq 2.7$, nødvendigvis er sann. I stedet prøver vi å undersøke hva som kan sies om parameteren $\theta = \mu_2 - \mu_1$ ut fra differansene i tabell 2.

(i) Vis at $\hat{\theta} = \bar{Y} - \bar{X}$ er forventningsrett estimator for θ.

(ii) Finn en formel for $\text{var}(\hat{\theta})$ uttrykt ved σ^2.

(iii) Vis ved hjelp av regel[1] 6.13 i Løvås at $\hat{\sigma}^2 = \dfrac{8S_x^2 + 9S_y^2}{17}$ er en forventningsrett estimator for σ^2, der $S_x^2 = \dfrac{1}{8}\sum_{i=1}^{9}(X_i - \bar{X})^2$ og $S_y^2 = \dfrac{1}{9}\sum_{i=1}^{10}(Y_i - \bar{Y})^2$.

<< **(i)** $E(\hat{\theta}) = E\left(\bar{Y}\right) - E(\bar{X}) = \mu_2 - \mu_1 = \theta \quad \Rightarrow \hat{\theta}$ forventningsrett.

(ii) Siden $\bar{X}$ og $\bar{Y}$ er uavhengige, er

$$\text{var}(\hat{\theta}) = \text{var}(\bar{Y}) + \text{var}(\bar{X}) = \frac{\sigma^2}{10} + \frac{\sigma^2}{9} = \sigma^2\left(\frac{1}{10} + \frac{1}{9}\right) = (0.2111)\sigma^2 .$$

(iii) Regel 6.13 gir $E\left(S_x^2\right) = E\left(S_y^2\right) = \sigma^2$, hvorav

$$E\left(\hat{\sigma}^2\right) = \frac{8}{17}E\left(S_x^2\right) + \frac{9}{17}E\left(S_y^2\right) = \left(\frac{8}{17} + \frac{9}{17}\right)\sigma^2 = \sigma^2$$

>>

C. Det kan vises (som du ikke trenger å gjøre) at $\dfrac{\hat{\theta} - \theta}{SE(\hat{\theta})}$ er *t*-fordelt med 17 frihetsgrader, der $SE(\hat{\theta})$ er (den estimerte) standardfeilen for $\hat{\theta}$: $\quad SE(\hat{\theta}) = \hat{\sigma}\sqrt{\dfrac{1}{9} + \dfrac{1}{10}}$.

(i) Vis hvordan dette resultatet kan brukes til å konstruere et 99% konfidensintervall for θ.

(ii) Beregn intervallet ut fra data i tabell 2 og kommenter svaret.

<< **(i)** Den øvre 0.005 fraktilen i t_{17}-fordelingen er $t_{17}(0.005) = 2.898$, som gir

$P\left(-2.898 \leq \dfrac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq 2.898\right) = 0.99$, hvorav

$$P\left(\hat{\theta} - (2.898)SE(\hat{\theta}) \leq \theta \leq \hat{\theta} + (2.898)SE(\hat{\theta})\right) = 0.99$$

(ii) $\hat{\theta}_{obs} = 0.27, \quad \hat{\sigma}_{obs} = \sqrt{(8/17)s_x^2 + (9/17)s_y^2} = 0.10968$, hvorav

$SE(\hat{\theta}) = \hat{\sigma}_{obs}\sqrt{1/10 + 1/9} = 0.050395$, og det observerte konfidensintervallet

---

[1] Regel 6.13 i Løvås sier at hvis $U_1, U_2, \ldots, U_n$ er uavhengige og identisk fordelte (uid) med $E(U_i) = \mu$ og $\text{var}(U_i) = \sigma^2$, er $S_u^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(U_i - \bar{U})^2$ en forventningrett estimator for σ^2.

$(0.27 \pm 0.15) = (0.12,\ 0.42)$, som i sin helhet ligger på den positive aksen, noe som gir sterk evidens (99% konfidens) for at $\theta > 0$.

>>

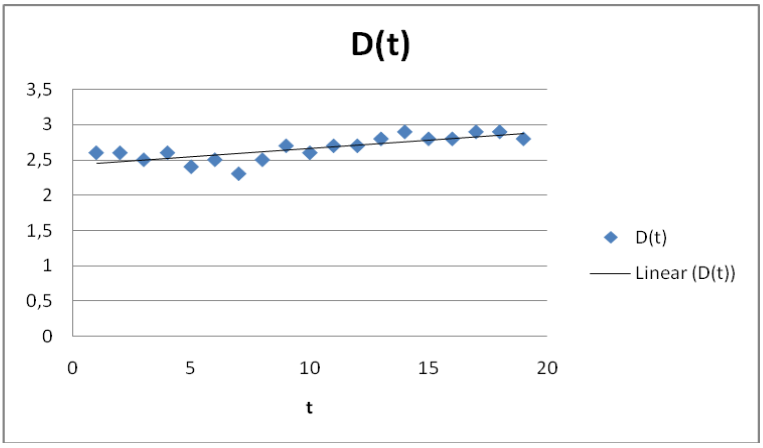
D. **Tilnærming 3.** Det kan innvendes mot modellen brukt ovenfor at den impliserer et sprang i forventet aldersforskjell i overgangen mellom 1999 og 2000. Mange vil føle at det er mer rimelig å forvente en gradvis økning over lengre tid hvis det foreligger en økende tendens. En enkel regresjonsmodell kan ta hensyn til dette:

La nå D_t være aldersdifferansen for år *t*, for $t = 1, 2, \ldots, 19$, og der $t = 1$, svarer til 1991, $t = 2$ til 1992, …., og $t = 19$ til 2009. Anta at $D_1, D_2, \ldots, D_{19}$ er uavhengige og normalfordelte med samme varians, $\operatorname{var}(D_t) = \sigma^2$, og

$$E(D_t) = \alpha + \beta t \quad \text{for} \quad t = 1, 2, \ldots, 19$$

I figur 2 har vi plottet aldersdifferansene mot *t* og tegnet inn minste kvadraters regresjonslinje. Problemet er nå om stigningen vi kan observere i figuren er signifikant.

Figur 2



Sett opp en passende null-hypotese og alternativ hypotese og undersøk om stigningen er signifikant positiv. Bruk signifikansnivå 1%. Formuler en konklusjon.

For å lette regningen er noen mellomresultater gitt i tabell 3.

Tabell 3 Mellomresultater

Formel	Observert verdi
n	19
$\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t$	10

$\bar{D} = \frac{1}{n}\sum_{t=1}^{n} D_t$	2.66
$s_t^2 = \frac{1}{n-1}\sum_{t=1}^{n}(t-\bar{t})^2$	31.6667
$S_d^2 = \frac{1}{n-1}\sum_{t=1}^{n}(D_t-\bar{D})^2$	0.0302
$S_{tD} = \frac{1}{n-1}\sum_{t=1}^{n}(t-\bar{t})(D_t-\bar{D})$	0.7611

<< Vi skal teste $H_0: \beta \le 0$ mot $H_1: \beta > 0$. Testobservatoren er $T = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{17}$-fordelt når $\beta = 0$, der $\hat{\beta} = \frac{S_{tD}}{s_t^2}$ (observert verdi = 0.0240) og $SE(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{(n-1)s_t^2}}$, og der ifølge regresjonsnotat II, $\hat{\sigma}^2 = \frac{SS_E}{n-2} = \frac{n-1}{n-2}\left(S_D^2 - \hat{\beta}^2 s_t^2\right)$ med observert verdi 0.012608, som gir $SE(\hat{\beta}) = 0.004703$ og $T_{obs} = 5.110$. Kritisk verdi er $t_{17}(0.01) = 2.567$ og null-hypotesen forkastes på 1%. Det er altså sterk evidens for å hevde positiv stigning ($\beta > 0$).
>>

E. I kommentarene til aldersdataene i tabell 2 skriver SSB:

> "Alderen er det observerte gjennomsnittet for dem som faktisk giftet seg i perioden/året, og derfor påvirket av variasjoner i størrelsen på årskullene."

La for eksempel antall menn som gifter seg første gang år *t* være n_t. Anta for enkelthets skyld at n_t, $t = 1,2,\ldots,19$, er faste (ikke-stokastiske) størrelser. La Y_t betegne gjennomsnittlig alder for menn som gifter seg første gang år *t*. Figur 1 kan tyde på at regresjonsforutsetningen, $E(Y_t) = \alpha_1 + \beta_1 t$, der α_1 og β_1 er ukjente parametre, er en rimelig forutsetning.

(i) Er det noe som tilsier at også en forutsetning om at Y_t er normalfordelt kan være en rimelig forutsetning? Gi en kort begrunnelse.

(ii) Hva med forutsetningen om at Y_t-ene har konstant varians? Mener du at dette er en rimelig forutsetning sett i lys av SSB's kommentar? Gi en kort begrunnelse.

<< **(i)** Ja. Sentralgrenseteoremet. **(ii)** Nei. $\text{var}(Y_t) = \sigma^2/n_t$ varierer med *t*. På den annen side, hvis årskullene ikke varierer for mye i den aktuelle perioden, behøver ikke forutsetningen om konstant varians være veldig svekket….
>>