ECON2130: EKSAMEN 2014 VÅR - UTSATT PRØVE

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

Oppgave 1

Fra en eldre dansk statistisk årbok (1982) finner vi følgende tabell over familier fordelt etter antall hjemmeboende barn under 26 år pr. 1. januar 1981 fra to områder i Danmark, København og Fredriksborg amtskommune.

Tabell 1. Antall familier med hjemmeboende barn under 26 år.

Antall barn	København	Fredriksborg	Sum
0	322 865	86 323	409 188
1	35 654	24 741	60 395
2	20 927	27 973	48 900
3	4 543	7 227	11 770
4 eller mer	1 138	1 323	2 461
Sum	385 127	147 587	532 714

Tabell 2. Samme tabell i prosent.

Antall barn	København	Fredriksborg	Sum
0	60.6	16.2	76.8
1	6.7	4.6	11.3
2	3.9	5.3	9.2
3	0.9	1.4	2.2
4 eller mer	0.2	0.2	0.5
Sum	72.3	27.7	100

La «**totalområdet**» betegne det kombinerte området der København og Fredriksborg er slått sammen til et område.

A. **(i)** Sett opp en sannsynlighetsmodell (dvs. med utfallsrom og tilhørende sannsynligheter) for et eksperiment som består av å trekke en familie rent tilfeldig fra totalområdet og registrere antall hjemmeboende barn og delområde (København eller Fredriksborg).
(ii) Er modellen en uniform sannsynlighetsmodell?
(iii) Hva er sannsynligheten for at den uttrukne familien ikke har noen hjemmeboende barn?

<< **Svar:** (i, ii): Utfallsrommet består av de 8 sentrale cellene i tabell 1 og 2.
Sannsynlighetene er angitt (i prosent) i tabell 2. Ikke uniform sannsynlighetsmodell.
(iii) Tabell 2 gir $P(\text{Ingen barn}) = 0.768$ >>

B. La K og F betegne begivenhetene at den uttrukne familien kommer henholdsvis fra København og fra Fredriksborg. La B_j være begivenheten at familien har j hjemmeboende barn ($j = 0,1,\ldots,4$), der B_4 betegner «4 eller mer».

(i) Finn $P(F \cap B_0)$ og $P(F \cup B_0)$.

(ii) La C være begivenheten at familien har 3 eller flere hjemmeboende barn. Beregn $P(C\,|\,K),\ P(C\,|\,F)$ og $P(F\,|\,C)$.

<< **Svar:** (i): Tabell 2 gir

$$P(F \cap B_0) = 0.162,\quad P(F \cup B_0) = P(F) + P(B_0) - P(F \cap B_0) = 0.277 + 0.768 - 0.162 = 0.883$$

(Kan også beregnes direkte fra tabell 1.)

(ii) $$P(C\,|\,K) = \frac{P(C \cap K)}{P(K)} = \frac{0.009 + 0.002}{0.723} = 0.015$$

$$P(C\,|\,F) = \frac{P(C \cap F)}{P(F)} = \frac{0.014 + 0.002}{0.277} = 0.058$$

$$P(C) = 0.009 + 0.014 + 0.002 + 0.002 = 0.027 \quad \Rightarrow \quad P(F\,|\,C) = \frac{P(F \cap C)}{P(C)} = \frac{0.016}{0.027} = 0.593$$

Den siste beregnet direkte fra tabell 1: $= \dfrac{8550}{14231} = 0.601$ (forskjell skyldes avrunding).

\>>

C. I dette punktet skal vi konsentrere oss om Fredriksborg og, for enkelthets skyld, anta at alle familier med minst 4 hjemmeboende barn alle har akkurat 4 (hjemmeboende) barn.

(i) Beregn gjennomsnittlig antall hjemmeboende barn pr. familie i Fredriksborg.

(ii) Forklar hvorfor gjennomsnittet beregnet i **(i)** kan tolkes som en forventning, $E(X)$, der X er en stokastisk variabel som gir antall hjemmeboende barn for en rent tilfeldig[1] valgt familie fra Fredriksborg.

<< **Svar:**

(i): Gj.snitt: $$\left[0 \cdot 86323 + 1 \cdot 24741 + \cdots + 4 \cdot 1323\right] / 147587 = \frac{107660}{147587} = 0.73$$

(ii): Sannsynligheten for j barn er $P(X = j) = p_j = \dfrac{N_j}{N},\ j = 0,1,\ldots,4$, der N_j er antall familier med j barn og $N = \sum_{j=0}^{4} N_j$. Dermed

[1] Uttrykket «rent tilfeldig» betyr at alle familier i Fredriksborg har samme sjanse for å bli valgt.

$$E(X) = \sum_{j=0}^{4} j \cdot p_j = \sum_{j=0}^{4} j \frac{N_j}{N} = \left[\sum_{j=0}^{4} j \cdot N_j \right] \Big/ N = \text{Gj.snitt i (i)}$$ >>

D. Anta vi trekker rent tilfeldig et utvalg på 30 familier fra Fredriksborg. Beregn sannsynligheten tilnærmet for at gjennomsnittlig antall hjemmeboende barn pr. familie i utvalget er større enn 1. Beskriv modellen du bruker.

[**Hint.** Det kan vises (som du ikke trenger å gjøre) at variansen til *X* i punkt **C**. er 0.9770. Du kan også anta at de 30 observasjonene i utvalget er uavhengige.]

<< **Svar:** Modell: $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte med

$E(X_i) = 0.73$ og $\mathrm{var}(X_i) = 0.977$. Dermed $\bar{X} \overset{\text{tiln.}}{\sim} N\left(0.73, \sqrt{\dfrac{0.977}{30}}\right) = N(0.73, 0.1805)$

$$\Rightarrow \quad P(\bar{X} > 1) = 1 - P(\bar{X} \le 1) \approx 1 - \mathrm{G}\left(\frac{1 - 0.73}{0.1805}\right) = 1 - G(1.50) = 1 - 0.9332 = 0.0668$$ >>

Oppgave 2

En urne inneholder et stort antall røde og hvite kuler. Arne påstår at det er minst 40% røde kuler i urnen. For å avgjøre dette spørsmålet ble det trukket et rent tilfeldig utvalg på 30 kuler **med tilbakelegging** fra urnen. Det viste seg at utvalget inneholdt 8 røde kuler. Oppgaven dreier seg bl. a. om å teste Arnes påstand (null-hypotesen) basert på dette resultatet.

A. La *X* betegne antall røde kuler i et rent tilfeldig utvalg på *n* = 30 kuler trukket med tilbakelegging fra urnen.

(i) Forklar hvorfor en binomisk fordeling er en rimelig modell for sannsynlighetsfordelingen til *X*.

(ii) Forklar hvorfor *X* kan anses å være tilnærmet normalfordelt hvis den sanne andelen av røde kuler i urnen er akkurat 40%. Hvilken normalfordeling er i så fall aktuell som tilnærmelse til fordelingen for *X*?

<< **Svar:** (i): Hvis andelen av røde kuler i urnen er 100*p*%, er *p* sannsynligheten for at en tilfeldig valgt kule er rød. De to forutsetningene for binomisk fordeling er klart realistisk her pga trekning med tilbakelegging. Så modellen blir $X \sim \mathrm{bin}(n, p)$.

(ii): Hvis *p* = 0.4, blir $\mathrm{var}(X) = 30(0.4)(0.6) = 7.2$, som i følge regel 5.20 er tilstrekkelig for tilnærming til normalfordelingen. Siden $E(X) = 12$, får vi

$X \overset{\text{tilnærmet}}{\sim} N(12, \sqrt{7.2}) = N(12, 2.683)$ >>

B. Testen består i å forkaste Arnes påstand (null-hypotesen) hvis *X* blir tilstrekkelig liten (dvs. hvis $X \leq k$, der *k* er en passende kritisk verdi). Beregn p-verdien tilnærmet for denne testen basert på resultatet i innledningen og formuler en konklusjon. Bruk heltallskorreksjon for å forbedre tilnærmelsen.

<< **Svar:** Siden $X_{obs} = 8$, får vi med heltallskorreksjon

$$\text{p-verdi} = P_{0.4}(X \leq 8) = P_{0.4}(X \leq 8.5) \approx G\left(\frac{8.5-12}{\sqrt{7.2}}\right) = G(-1.30) = 0.0968$$, som ved vanlige nivåer ikke gir grunnlag for forkastning. >>

C. Anta nå i stedet at utvalget på *n* = 30 kuler ble trukket fra urnen **uten tilbakelegging**, og at resultatet ble som før, $X_{obs} = 8$. Anta i tillegg at urnen i alt inneholder 350 kuler. Testen er som i punkt **B**. Hva blir den heltallskorrigerte p-verdien i dette tilfellet?

<< **Svar:** Under de nye omstendighetene blir X hypergeometrisk fordelt. Hvis *p* = 0.4, blir

$$E(X) = 12 \text{ og } \mathrm{var}(X) = 30(0.4)(0.6)\frac{350-30}{349} = 6.6017 = (2.569)^2$$

Betingelsen for tilnærmelse til normalfordelingen i regel 5.20 er fortsatt oppfylt

$$\Rightarrow \quad X \overset{\text{tilnærmet}}{\sim} N(12, 2.569)$$, og tilnærmet p-verdi blir

$$\text{p-verdi} = P_{0.4}(X \leq 8) = P_{0.4}(X \leq 8.5) \approx G\left(\frac{8.5-12}{2.569}\right) = G(-1.36) = 0.0869 .$$

(Samme konklusjon som i **B**). >>

Oppgave 3

I forbindelse med tidsstudieanalyse observerte man sammenhørende verdier av pakketid, y, (i sekunder) for en pakke og dens volum, x, (i kubikkfot) for 9 pakker. Observasjonene er gitt i tabell 3.

Tabell 3.

Volum, x	Pakketid, y
0.72	10.8
0.88	14.4
1.23	19.6
1.12	18.0
0.48	8.4
1.09	17.2
0.63	11.0
0.82	13.3
1.40	23.1

Mellomresultater:

Gjennomsnitt: $\bar{x} = 0.93 \quad \bar{y} = 15.1$

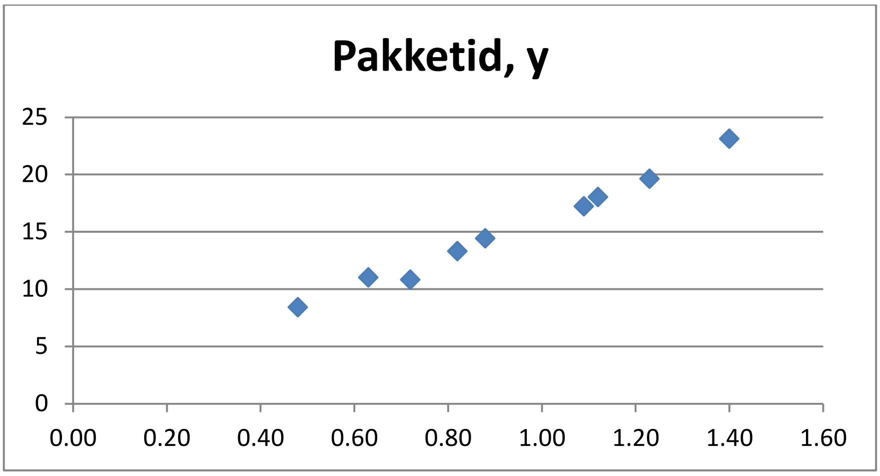
Empirisk standardavvik: $s_x = 0.30 \quad s_y = 4.76$

Empirisk kovarians: $s_{xy} = 1.42$

A. **(i)** Skisser et spredningsplott for observasjonsparene, $(x_i, y_i), i = 1, 2, \ldots, 9$.

(ii) Beregn minste kvadraters regresjonslinje for y med hensyn på x.

<< **Svar:** (i)



(ii): Regresjonslinje: $y = a + bx$, der $b = \dfrac{s_{xy}}{s_x^2} = 15.78$ og $a = \bar{y} - b\bar{x} = 0.43 \quad \Rightarrow$

$y = 0.43 + 15.78 * x$ >>

B. Vi ønsker et 95% konfidensintervall for forventet pakketid for en pakke av størrelse 1 kubikkfot, basert på den regresjonsmodellen for Y med hensyn på x beskrevet i Løvås kapeittel 7. Her betegner Y den stokastiske pakketiden for en pakke med volum x kubikkfot.

(i) Sett opp en forventningsrett estimator for $\mu = E(Y)$ for en pakke med volum $x = 1$ kubikkfot . Begrunn forventningsrettheten. [**Hint.** Du kan utnytte at minste kvadraters estimatorer, $\hat{\alpha}$ og $\hat{\beta}$ for α og β i regresjonsfunksjonen $\mu(x) = \alpha + \beta x$, ut fra teorien, begge er forventningsrette.]

(ii) Beregn et 95% konfidensintervall for μ .

<< **Svar:** (i): Har $\mu = \mu(1) = \alpha + \beta$. $\hat{\mu} = \hat{\alpha} + \hat{\beta}$ er forventningsrett siden $E(\hat{\mu}) = E(\hat{\alpha} + \hat{\beta}) = E(\hat{\alpha}) + E(\hat{\beta}) = \alpha + \beta = \mu$.

(ii): Teorien gir $\operatorname{var}(\hat{\mu}) = \sigma^2 \left(\dfrac{1}{9} + \dfrac{(1-\bar{x})^2}{8 s_x^2} \right)$.

$$\hat{\sigma}^2 = \frac{8}{7}\left(S_Y^2 - \hat{\beta}^2 s_x^2\right) \quad \Rightarrow \quad \hat{\sigma}^2_{obs} = 0.2893$$

$$\frac{1}{9} + \frac{(1-\overline{x})^2}{8s_x^2} = 0.3434 \quad \Rightarrow \quad \mathrm{se}(\hat{\mu})\Big|_{obs} = \hat{\sigma}\sqrt{\frac{1}{9} + \frac{(1-\overline{x})^2}{8s_x^2}}\Bigg|_{obs} = 0.1847 \quad \text{and}$$

$T = \dfrac{\hat{\mu} - \mu}{\mathrm{se}(\hat{\mu})} \sim t(7)$ fordelt . 0.025 kvantilen i $t(7)$-fordelingen er $t_{0.025} = 2.365$, som gir vårt 95% t-intervall: $\left[\hat{\mu} \pm t_{0.025}\mathrm{se}(\hat{\mu})\right]_{obs} = \left[16.2 \pm (2.365)(0.1847)\right] = \left[15.8,\ 16.6\right]$ >>
