UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

Eksamen i: **ECON2130 – Statistikk 1**

Eksamensdag: 03.06.2016 **Sensur kunngjøres: 24.06.2016**

Tid for eksamen: kl. 09:00 – 12:00

Oppgavesettet er på 5 sider

Tillatte hjelpemidler:

- Alle trykte og skrevne hjelpemidler er tillatt. I tillegg kan du ta med lommekalkulator som ikke kan brukes til å kommunisere med andre.

Eksamen blir vurdert etter ECTS-skalaen. A-F, der A er beste karakter og E er dårligste ståkarakter. F er ikke bestått.

ECON2130: EKSAMEN 2016v

Oppgave 1

I en urne er det 10 kuler hvorav 5 er røde, 4 er blå og 1 er hvit. Kulene er ellers like store og tunge. Det trekkes ut et rent tilfeldig utvalg på 2 kuler fra urnen som betyr at alle mulige (ikke-ordnete) utvalg på 2 kuler er like sannsynlige.

Det kan vises (som du ikke trenger å gjøre) at utvalget blir rent tilfeldig hvis kulene trekkes en og en slik at alle de 10 kulene i urnen har samme sjanse for å bli trukket i første trekning, mens alle de 9 gjenværende kulene har samme sjanse for å bli trukket i annen trekning.

A. La, R_1, B_1, H_1 betegne begivenhetene at den første kulen som blir trukket er henholdsvis rød, blå eller hvit. La R_2, B_2, H_2 betegne de tilsvarende begivenhetene for den andre kulen som blir trukket ut. For eksempel, $R_1 \cap H_2$ betyr begivenheten at det blir trukket en rød kule i første trekning og en hvit i andre, mens $B_1 \cap B_2$ betyr at både første og andre kule som blir trukket ut er blå. $\bar{B}_1$ betyr at den første kulen som blir trukket ut ikke er blå.

Finn de 5 sannsynlighetene, $P(B_1)$, $P(\bar{B}_1)$, $P(B_2 | B_1)$, $P(B_2 | \bar{B}_1)$, $P(B_1 \cap B_2)$

B. La Y være antall blå kuler i et rent tilfeldig utvalg på 2 kuler.

i. Forklar hvorfor Y er hypergeometrisk fordelt og sett opp et uttrykk for sannsynlighetsfordelingen.

ii. Beregn sannsynlighetene $P(Y = y)$, $y = 0,1,2$. [**Hint.** Vis først at $\binom{10}{2} = 45$.]

iii. Beregn $E(Y)$, $\text{var}(Y)$ ved, for eksempel, å bruke formlene for forventning og varians i en hypergeometrisk fordeling.

C. La X være antall røde kuler i utvalget. Det kan vises (som du ikke trenger å gjøre) at den simultane fordelingen for X og Y er gitt i tabell 1.

Tabell 1 Den simultane fordelingen for *X* og *Y* gitt ved $f(x, y) = P(X = x \cap Y = y)$

		y			
		0	1	2	Sum
x	0	0	4/45	6/45	10/45
	1	5/45	20/45	0	25/45
	2	10/45	0	0	10/45
	Sum	15/45	24/45	6/45	1

i. Er *X* og *Y* stokastisk uavhengige? Begrunn svaret ditt.
ii. Vis at kovariansen mellom *X* og *Y* er $\text{cov}(X, Y) = -0.3556$.
iii. Finn sannsynlighetene $P(Y = 1 | X = 0)$ og $P(Y = 1 | X = 1)$.

D. Mot å betale en spilleavgift på 10 kr. blir Jens tilbudt et spill som består i å trekke 2 kuler fra urnen som beskrevet i innledningen. For hver blå kule i utvalget får Jens 40 kr, men må betale 25 kr. for hver rød kule i utvalget. Hvis det dukker opp en hvit kule i utvalget, får ikke Jens noe for denne, men slipper også å betale noe. Fortjenesten til Jens, *V*, for et spill kan dermed uttrykkes som $V = 40Y - 25X - 10$. De mulige fortjenestene som kan inntreffe i et enkelt spill er angitt i tabell 2.

Tabell 2 Mulige fortjenester for et spill med utfall (x, y)

		y		
		0	1	2
x	0	---	30	70
	1	–35	5	---
	2	–60	---	---

i. Finn sannsynligheten for at Jens får positiv fortjeneste i et spill (dvs. finn $P(V > 0)$).
ii. Finn forventet fortjeneste i et spill, $E(V)$. Forklar kort hvordan du ville tolke dette resultatet.
iii. Vis at variansen til fortjenesten i et spill er, $\text{var}(V) = 1671.556$.
[**Hint.** Variansen til *X*, som du ikke trenger å beregne her, oppgis til $4/9$]

E. Jens spiller spillet i punkt **D** 30 ganger. La V_i betegne fortjenesten i spill nr. *i*, $i = 1, 2, \ldots, 30$. Den totale fortjenesten er $T = V_1 + V_2 + \cdots + V_{30}$, der $V_1, V_2, \ldots, V_{30}$ antas å være uavhengige og identisk fordelte (*uid*) med samme fordeling som *V*.

i. Finn sannsynligheten tilnærmet for at T blir positiv (dvs. finn $P(T > 0)$ tilnærmet). [**Hint.** Du trenger bl.a. $E(V)$. Hvis du ikke fant denne i punkt **D**, gjett på en verdi og bruk denne verdien i beregningen.]

ii. Anta at sannsynligheten for at et enkelt spill gir positiv fortjeneste er $2/3$. La U være antall ganger i løpet 30 spill V_i blir positiv ($V_i > 0$). Finn sannsynligheten tilnærmet for at U blir minst 20 (dvs. finn $P(U \geq 20)$ tilnærmet).

Oppgave 2

Tabell 3 viser antall trafikkulykker med personskade i Norge for to perioder, mai 2015 (periode 1) og juni-juli 2015 (periode 2)[1].

Tabell 3

Periode	Mai 2015	Juni –juli 2015	Sum
Antall trafikkulykker med personskade	577	1274	1851

La X betegne antall trafikkulykker i mai og Y summen av antall ulykker i juni og juli der X og Y er sett på som stokastiske variable. Anta at X og Y er uavhengige og poissonfordelte med forventninger, λ og 2λ henholdsvis (kort skrevet: $X \sim Pois(\lambda),\ Y \sim Pois(2\lambda)$), der λ tolkes som forventet ulykkesrate pr. måned som antas å være den samme i de tre månedene i perioden mai – juli.

For enkelthets skyld, bortsett fra i punkt **D** nedenfor, antar vi også at de tre månedene er like lange, for eksempel 30 dager hver.

X og Y gir to uavhengige og forventningsrette estimatorer for λ, nemlig $\hat{\lambda}_1 = X$ og $\hat{\lambda}_2 = Y/2$.

A. **i.** Hvis $V \sim Pois(t\lambda)$ for en periode på t måneder, forklar kort hvorfor $\lambda^* = V/t$ er en forventningsrett estimator for λ med varians, $\text{var}(\lambda^*) = \lambda/t$.

ii. Det foreslås to estimatorer for λ basert på $\hat{\lambda}_1$ og $\hat{\lambda}_2$:

$$\hat{\lambda} = \frac{\hat{\lambda}_1 + 2\hat{\lambda}_2}{3} \quad \text{og} \quad \tilde{\lambda} = \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{2}$$

Vis at både $\hat{\lambda}$ og $\tilde{\lambda}$ er forventningsrette og at $\hat{\lambda}$ har minst varians lik $\lambda/3$.

iii. Beregn de to estimatene for λ basert på $\hat{\lambda}$ og $\tilde{\lambda}$ og velg det estimatet du mener er best. Angi hvorfor du mener valget ditt er best.

[1] Kilde SSB.

B. Det kan vises (som du ikke trenger å gjøre) at $\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/3}}$ er tilnærmet standard normalfordelt (N(0, 1)) uansett λ.

i. Bruk dette til å utlede et konfidensintervall for λ med konfidensgrad tilnærmet 0.90, basert på $\hat{\lambda}$.

ii. Beregn det observerte konfidensintervallet basert på data i tabell 3.

C. Vi ønsker å sjekke antakelsen om konstant ulykkesrate i hele tre-månedsperioden mai – juli med en statistisk test. For å få til dette utvider vi modellen litt ved å anta at forventet ulykkesrate for mai er λ_1 og for juni-juli λ_2, der λ_1 og λ_2 ikke behøver å være like. Vi ønsker å teste nullhypotesen at λ_1 og λ_2 er like, dvs.
$H_0: \lambda_2 - \lambda_1 = 0$ mot $H_1: \lambda_2 - \lambda_1 \neq 0$. Innfører vi parameteren $\delta = \lambda_2 - \lambda_1$, kan vi skrive hypotesen vi ønsker å teste som $H_0: \delta = 0$ mot $H_1: \delta \neq 0$

Forventningsrette estimatorer, $\hat{\lambda}_1$ og $\hat{\lambda}_2$, er gitt i innledningen. La σ^2 betegne variansen til $\hat{\delta} = \hat{\lambda}_2 - \hat{\lambda}_1$ som avhenger av de ukjente λ_1 og λ_2. Estimatoren $\hat{\sigma}^2$ er dannet ved å erstatte de ukjente λ_1 og λ_2 med sine respektive estimatorer.

i. Forklar kort hvorfor $\hat{\delta} = \hat{\lambda}_2 - \hat{\lambda}_1$ er tilnærmet normalfordelt.

ii. Det kan vises (som du ikke trenger å gjøre) at $Z = \frac{\hat{\delta}}{\hat{\sigma}}$ er tilnærmet standard normalfordelt, N(0, 1), dersom $\lambda_1 = \lambda_2$ (dvs. $\delta = 0$). Bruk dette til å sette opp et forkastningskriterium for en test for H_0 med signifikansnivå tilnærmet 0.05.

iii. Gjennomfør testen i **ii.** og formuler en konklusjon.

D. Til tross for eventuell forkastning av H_0 i punkt **C.**, går vi nå tilbake til den opprinnelige modellen med konstant forventet ulykkesrate pr. tidsenhet i hele tre-månedsperioden, men ønsker å justere for at månedslengden varierer i perioden. I alt omfatter perioden 92 dager siden mai og juli har 31 dager hver. Istedenfor λ vil vi nå estimere θ definert som forventet ulykkesrate pr. 30 dager uansett måned.

En måte å gjøre dette på er ved å forutsette at $T = X + Y$ er poisson-fordelt med forventning 92μ, der $\mu = \theta/30$ er forventet ulykkesrate pr. dag som antas å være konstant i perioden.

i. Sett opp en forventningsrett estimator, $\hat{\theta}$, for θ og finn et uttrykk for variansen til denne.

ii. Beregn estimatet basert på $\hat{\theta}$ og standardfeilen, $SE(\hat{\theta})$ (som betyr estimert standardavvik for $\hat{\theta}$).

iii. Beregn også et tilnærmet 95% konfidensintervall basert på at $\dfrac{\hat{\theta}-\theta}{SE(\hat{\theta})}$ er tilnærmet standard normalfordelt (som du ikke trenger å begrunne her).