UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

Utsatt eksamen i: **ECON2130 – Statistikk 1**

Eksamensdag: 15.06.2015 **Sensur kunngjøres senest: 06.07.2015**

Tid for eksamen: kl. 09:00 – 12:00

Oppgavesettet er på 4 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 2015 VÅR - UTSATT PRØVE

Oppgave 1

Vi har 4 spar-kort, hentet fra en vanlig kortstokk, med tallverdier, 5,6,7,8. To av disse, 6 og 8, er partall (dvs. er delelig med 2), og to, 5 og 7, er oddetall.

A. Vi trekker to kort uten tilbakelegging fra disse fire slik at alle utvalg på to kort er like sannsynlige. La A_1 være begivenheten at det første kortet trukket ut har et oddetall som verdi, og A_2 begivenheten at det andre kortet som blir trukket ut har et oddetall som verdi.

i) Finn sannsynlighetene $P(A_1)$, $P(A_1 \cap A_2)$ og $P(A_2)$

ii) Anta vi *vet* at det andre kortet som ble trukket ut viste et oddetall (dvs. at A_2 inntraff), men ikke hva som skjedde i første trekning (dvs. av det første kortet). Hva er sannsynligheten for at det ble trukket et oddetall også i første trekning?

B. To skoleelever, Kari og Per, spiller et spill de kaller «oddetall og partall» basert på de fire kortene beskrevet i innledningen. Et spill består i å trekke to kort uten tilbakelegging fra de fire kortene. Dersom summen av de to uttrukne kortverdiene er et oddetall, vinner Kari, og hvis summen er et partall, vinner Per. De to elevene tror nemlig det er samme sjanse for at summen blir et oddetall som at den blir et partall, slik at begge har samme sjanse for å vinne. De tar imidlertid feil.

Vis at $P(\text{Kari vinner}) = \dfrac{2}{3}$.

[Hint. Det kan lønne seg å lage en tabell over aktuelle summer, $x + y$, for alle kombinasjoner av *x* og *y*, der *x* er verdien på det første kortet som trekkes ut og *y* verdien på det andre. Du kan anta at alle kombinasjoner som er mulige, er like sannsynlige.]

C. La (i et spill) *X* være verdien på det første kortet som trekkes ut og *Y* verdien på det andre kortet som trekkes ut.

i) Vis at $P(X = 5 \cap Y = 6) = \dfrac{1}{12}$

ii) Sett opp en tabell over simultanfordelingen for *X* og *Y*, bestemt ved $f(x, y) = P(X = x \cap Y = y)$ for alle kombinasjoner av *x* og *y*. Du kan, som før, anta at alle kombinasjoner som er mulige, er like sannsynlige.

iii) Er *X* og *Y* stokastisk uavhengige? Begrunn svaret ditt.

D. Kari og Per spiller spillet beskrevet i punkt **B** 24 ganger. Det er klart at Kari vil ha en tendens til å vinne oftere enn Per siden hun har større sannsynlighet for å vinne i hvert spill. La U være antall ganger Kari vinner i løpet av 24 spill.

i) Beregn sannsynligheten tilnærmet for at Kari vinner oftere enn Per i løpet av 24 spill. Bruk heltallskorreksjon.

[**Hint.** Forklar at U er binomisk fordelt. Beregn deretter $P(U > 12)$ tilnærmet.]

ii) La D være forskjellen mellom antall ganger de to vinner (dvs. D er antall ganger Kari vinner minus antall ganger Per vinner i løpet av 24 spill). Beregn forventningen og variansen til D.

iii) Begrunn at D er tilnærmet normalfordelt. I så fall, hvilken normalfordeling?

Oppgave 2

Det er en vanlig erfaring at man må stå i kø foran kassen for å få betalt i supermarkeder. I **tabell 1** er det gitt observasjoner fra 6 forskjellige tilfeldige tilfeller man kom for å betale, målt lørdag ettermiddag ved en bestemt kasse, der x_i angir antall personer som sto foran i køen, og y_i tiden (i minutter) man måtte vente til man ble betjent for betaling (for tilfelle $i = 1, 2, \ldots, 6$).

Tabell 1 **Data**

i	Antall personer foran i køen (x_i)	Ventetid (*) til betjening ved kassen (minutter) (y_i)
1	2	5.5
2	5	9.1
3	8	16.0
4	3	6.9
5	6	12.3
6	4	8.2
Sum	28	58.0

(*) Desimaltallene betyr ti-deler slik at, for eksempel, 5.5 minutter betyr 5 minutter og 30 sekunder.

A. Ventetiden i køen til betjening for betaling, Y, antas generelt å være en stokastisk variabel som er normalfordelt med forventning, βx, og varians, $\sigma^2 x$, der x er antall personer foran i køen når man kommer til kassen for betaling. For enkelthets skyld antas i denne oppgaven at antall personer i køen, x, alltid er et gitt tall (dvs. ikke-stokastisk). β og σ^2 er (ukjente) parametre i modellen.

i) Forklar hvorfor denne modellen sier det samme som å si at ventetiden pr. person i køen, Y/x er normalfordelt med forventning β og varians σ^2/x (der altså x er ikke-stokastisk).

ii) Den samlete tiden en kunde tilbringer ved kassen består av to deler, tiden i kø pluss betjeningstiden ved selve kassen. Forklar kort hvorfor parameteren β kan tolkes som gjennomsnittlig betjeningstid pr. kunde ved kassen (dvs. den gjennomsnittlige tiden det tar å betjene en kunde).

B. I tråd med modellen i punkt **A**, antar vi at ventetidene i **tabell 1** er observasjoner av stokastiske variable, $Y_1, Y_2, \ldots, Y_6$, som antas å være uavhengige og normalfordelte slik at $Y_i \sim N\left(\beta x_i,\ \sigma\sqrt{x_i}\right)$, der x_i angir antall personer i køen foran for tilfelle i $(i = 1, 2, \ldots, 6)$. x_i-ene antas, som før, gitte tall (dvs. ikke-stokastiske). Det blir foreslått to estimatorer for β:

$$\hat{\beta} = \frac{Y_1 + Y_2 + \cdots + Y_6}{x_1 + x_2 + \cdots + x_6} \qquad \text{og} \qquad \tilde{\beta} = \frac{1}{6}\sum_{i=1}^{6}\frac{Y_i}{x_i}$$

i) Vis at både $\hat{\beta}$ og $\tilde{\beta}$ er forventningsrette estimatorer for β.

ii) Beregn estimatene $\hat{\beta}_{obs}$ og $\tilde{\beta}_{obs}$.

[**Hint.** Til hjelp under regningen oppgis $\sum_{i=1}^{6} y_i/x_i = 12.97$]

iii) Beregn variansene, $\text{var}(\hat{\beta})$ og $\text{var}(\tilde{\beta})$, uttrykt ved σ^2. Hvilket av de to estimatene i **ii)** mener du er mest troverdig? Gi en kort begrunnelse for svaret.

[**Hint.** Til hjelp under regningen oppgis $\sum_{i=1}^{6} 1/x_i = 1.58$]

C. Det kan vises (som du ikke trenger å gjøre) at $W = \dfrac{\hat{\beta} - \beta}{\hat{\sigma}\Big/\sqrt{\sum_{i=1}^{6} x_i}} \sim t(5)$ fordelt (dvs. t-fordelt med 5 frihetsgrader), der $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ og $\hat{\sigma}^2 = \frac{1}{5}\sum_{i=1}^{6}\frac{(Y_i - x_i\hat{\beta})^2}{x_i}$.

i) Bruk dette til å finne et 95% konfidensintervall for β.

ii) Beregn konfidensintervallet ut fra data.

[**Hint.** For å lette regningen, oppgis den observerte verdien, $\hat{\sigma}^2_{obs} = 0.2878$.]

D. I dette og neste punkt skal vi benytte modellen i punkt B, men der vi, for enkelthets skyld, antar at parameteren σ^2 er kjent, $\sigma^2 = 0.35$.

i) Forklar hvorfor estimatoren $\hat{\beta}$ er normalfordelt, med $\hat{\beta} \sim N\left(\beta,\ \sqrt{\frac{0.35}{28}}\right)$.

ii) Tidligere har man regnet med en gjennomsnittlig betjeningstid på 1.9 minutter pr. kunde ved kassen lørdag ettermiddag. Tyder data på at gjennomsnittlig betjeningstid har økt (basert på signifikansnivå 5%)? Med andre ord, gjennomfør en 5% test for $H_0 : \beta \leq 1.9$ mot $H_1 : \beta > 1.9$, og formuler en konklusjon.

iii) Beregn p-verdien (basert på data i tabell 1) for testen din i **ii)**.

E. **i)** Sett opp et uttrykk for styrkefunksjonen for testen din i punkt **Dii)**.

ii) Beregn sannsynligheten for å forkaste H_0 hvis den sanne verdien av β er 2.