ECON2130: EKSAMEN 2016 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

En konferanse om aksjer hadde en av dagene to parallelle sesjoner rett etter lunsj, som innebar at de to sesjonene gikk samtidig og at deltakerne måtte velge hvilken de ville delta på. Ingen kunne få med seg begge sesjonene. Den ene sesjonen dreide seg om portefølje-administrasjon og hadde 40% av deltakerne på konferansen, mens temaet for den andre var grafiske metoder og hadde 50% av deltakerne.

Senere om kvelden var det en fellessesjon med tittel «Er Random-walk-modellen[1] død?». 80% av konferansens deltakere deltok på denne.

Vi trekker en deltaker rent tilfeldig slik at alle deltakerne på konferansen har samme sjanse for å bli trukket. La Pf, Gr, Rw indikere begivenhetene at den uttrukne deltakeren deltok henholdsvis på sesjonen om portefølje-administrasjon (kalt portefølje-sesjonen), på sesjonen om grafiske metoder (kalt grafikk-sesjonen) og på sesjonen om random-walk-modellen (kalt random-walk-sesjonen).

Følgende tre ting antas om disse tre begivenhetene:

(1) Pf og Gr er disjunkte,
(2) Pf og Rw er stokastisk uavhengige,
(3) 85% av dem som deltok på grafikk-sesjonen deltok også på random-walk-sesjonen.

Disse tre antakelsene gjelder i hele oppgaven.

A. **i.** Hva er sannsynligheten for at den uttrukne deltakeren deltok på minst en av portefølje-sesjonen og grafikk-sesjonen?

ii. Hva er sannsynligheten for at den uttrukne deltakeren deltok på minst en av portefølje-sesjonen og random-walk-sesjonen?

<<< **Svar:** **i.** Disjunksjonen i (1) gir

$$P(Pf \cup Gr) = P(Pf) + P(Gr) = 0.4 + 0.5 = 0.9$$

ii. Uavhengigheten i (2) gir

[1] Random-walk-modellen er en type modell som av og til er brukt for å analysere aksjemarkeder.

$$P(Pf \cup Rw) = P(Pf) + P(Rw) - P(Pf \cap Rw) = P(Pf) + P(Rw) - P(Pf)P(Rw)$$
$$= 0.4 + 0.8 - (0.4)(0.8) = 0.88$$

>>>

B. **i.** Er *Gr* og *Rw* uavhengige begivenheter? Begrunn svaret ditt.
ii. Hva er sannsynligheten for at den uttrukne deltakeren deltok på minst en av grafikk-sesjonen og random-walk-sesjonen?

<<< **Svar:** **i.** $P(Rw) = 0.8$ er forskjellig fra $P(Rw \mid Gr) = 0.85$, som impliserer avhengighet.

ii. Finner $P(Rw \cap Gr) = P(Gr)P(Rw \mid Gr) = (0.5)(0.85) = 0.425$, hvorav
$P(Gr \cup Rw) = P(Gr) + P(Rw) - P(Gr \cap Rw) = 0.5 + 0.8 - 0.425 = 0.875$
>>>

C. La *X* være antall sesjoner blant de tre (portefølje-, grafikk- og random-walk-sesjonen) som den uttrukne deltakeren deltok på.

i. Vis at $P(X = 2) = 0.745$.
ii. Sett opp sannsynlighetsfordelingen for *X* når du får oppgitt (som du ikke trenger å vise her) at $P(X = 0) = 0.045$.
iii. Beregn $E(X)$ og $\text{var}(X)$.

<<< **Svar:** **i.** Vi har $(X = 2) = (Pf \cap Rw) \cup (Gr \cap Rw)$, som er en disjunkt union. Dermed
$P(X = 2) = P(Pf \cap Rw) + P(Gr \cap Rw) = (0.4)(0.8) + (0.5)(0.85) = 0.745$

ii. Siden
$1 = P(X = 0) + P(X = 1) + P(X = 2) = 0.045 + P(X = 1) + 0.745 = P(X = 1) + 0.790$, blir
$P(X = 1) = 1 - 0.790 = 0.210$.
Dermed fordelingen for *X*:

x	0	1	2
P(X=x)	0.045	0.210	0.745

iii. $E(X) = 0 + 1 \cdot 0.21 + 2 \cdot 0.745 = 1.7$
$E(X^2) = 0 + 1 \cdot 0.21 + 4 \cdot 0.745 = 3.19$
$\text{var}(X) = E(X^2) - [E(X)]^2 = 3.19 - (1.7)^2 = 0.3$
>>>

D. Konferansen hadde 200 deltakere i alt. Av disse deltok 80 på portefølje-sesjonen, 100 på grafikk-sesjonen og 160 på random-walk-sesjonen. Hvor mange som deltok både på portefølje-sesjonen og på random-walk-sesjonen var ikke kjent. Heller ikke hvor mange som deltok på både random-walk- og grafikk-sesjonen eller hvor mange som kun deltok på random-walk-sesjonen.

Anslå de tre ukjente antallene ved å bruke antakelsene i innledningen, dvs. anslå hvor mange som deltok på både random-walk- og portefølje-sesjonen, hvor mange som deltok på både random-walk- og grafikk-sesjonen og hvor mange som kun deltok på random-walk-sesjonen.

[**Hint.** Bruk for eksempel forventet antall i de tre kategoriene.]

<<< **Svar:** Vi bruker sannsynlighetene,

$$P(Pf \cap Rw) = (0.4)(0.8) = 0.32 \text{ og } P(Gr \cap Rw) = (0.5)(0.85) = 0.425$$

Hvis U_1, U_2 er antallet av $Pf \cap Rw$ og antallet av $Gr \cap Rw$ henholdsvis blant deltakerne, og U_3 er antall som kun deltok på random-walk-sesjonen, finner vi

$$E(U_1) = 200 \cdot P(Pf \cap Rw) = 200 \cdot 0.32 = 64,$$
$$E(U_2) = 200 \cdot P(Gr \cap Rw) = 200 \cdot 0.425 = 85$$
$$E(U_3) = 160 - E(U_1) - E(U_2) = 11 \quad (\text{siden } U_1 + U_2 + U_3 = 160)$$

Dette kan bl.a. begrunnes ved å vise til at U_1, U_2 begge er hypergeometrisk fordelte der sannsynlighetene er andeler i populasjonen av konferansedeltakere.
>>>

E. **i.** Skisser et Venn-diagram over begivenhetene *Pf*, *Gr* og *Rw*.

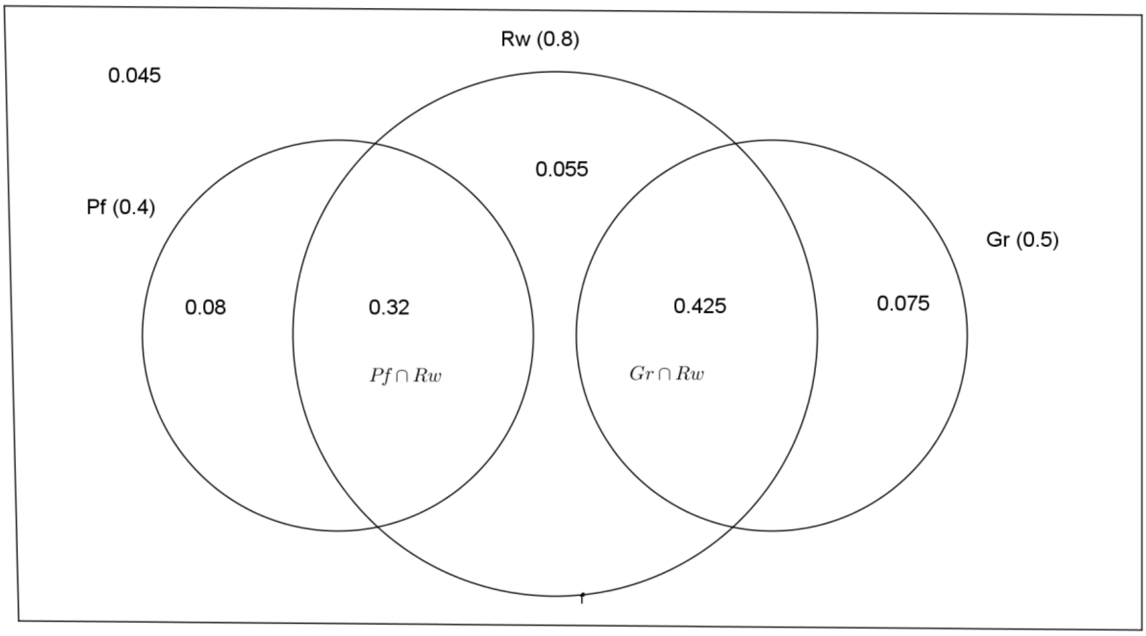
ii. Venn-diagrammet deler utfallsrommet i 6 disjunkte deler. For eksempel, to av disse delene er $Gr \cap Rw$ og $Gr \cap \overline{Rw}$, som tilsammen blir *Gr*. Skriv inn i diagrammet sannsynlighetene for alle de 6 disjunkte delene.

iii. Forklar hvorfor $P(X = 0) = 0.045$ som angitt i punkt **C.ii.** ovenfor.

<<< **Svar:**

i. og ii.:

Utfallsrommet



iii. Begivenheten $(X = 0)$ inntreffer hvis og bare hvis den uttrukne deltakeren havner utenfor de tre sirklene, som har sannsynlighet 1- (0.8+0.08+0.075)=0.045.

>>>

Oppgave 2

Dataene er hentet fra et eksperiment (1981 i England) for å undersøke effekten av et svakt stimulerende stoff som koffein på ytelsen ved en enkel fysisk oppgave. Den fysiske oppgaven i eksperimentet var å tromme med en finger på bordet så fort man kunne og der hastigheten ble målt som antall slag pr. minutt.

Eksperimentet besto i først å trene 30 mannlige college studenter en tid i å tromme med fingeren uten at studentene hadde inntatt noen stimulerende midler. Gjennomsnittlig ytelse for studentene ved denne innledende treningen var 245 slag pr. minutt.

Deretter ble studentene ved loddtrekning inndelt i 3 grupper på 10 hver. Alle studentene måtte drikke en kopp kaffe. Gruppe 1 hadde fått 0 ml koffein i koppen, gruppe 2 hadde fått 100 ml. koffein i koppen mens gruppe 3 fikk 200 ml koffein i koppen. Ingen av studentene visste hvilken gruppe de tilhørte. Etter ti minutter ble studentene bedt om å tromme med fingeren og hastigheten ble registrert.

Det viste seg at gjennomsnittshastigheten blant dem som hadde fått 0 ml koffein ikke var signifikant forskjellig fra 245 slag pr. minutt, noe som gav øket evidens for at en eventuell signifikant endring i hastighet blant dem som fikk koffein skyltes koffeinen først og fremst.

I denne oppgaven skal vi kun se på gruppe 2 som hadde fått 100 ml koffein (og ignorere gruppe 3) og undersøke om tallene gir grunnlag for å konkludere at koffein (100 ml) øker ytelsen. Hastighetene, $x_1, x_2, \ldots, x_{10}$ for gruppe 2, er gitt i tabell 1:

Tabell 1 Antall slag pr. minutt for studenter som fikk 100 ml koffein i koppen.

248	246	245	247	248	250	247	246	243	244

For å lette regningen i oppgaven, oppgis $\sum_{i=1}^{10} x_i = 2464$ og $\sum_{i=1}^{10} x_i^2 = 607\,168$.

La X_i betegne den stokastiske variabelen bak målingen x_i for $i = 1, 2, \ldots, 10$. Vi antar at $X_1, X_2, \ldots, X_{10}$ uavhengige og identisk normalfordelte med forventning $E(X_i) = \mu$ og varians, $\mathrm{var}(X_i) = \sigma^2$. Vi ønsker blant annet å teste $H_0 : \mu \le 245$ mot $H_1 : \mu > 245$.

A. Vi antar i dette punktet for enkelthets skyld at variansen til X_i er kjent lik 5 (dvs. $\sigma^2 = 5$).

- **i.** En test består i å forkaste H_0 hvis $\bar{X} > 246$, der $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$. Beregn signifikansnivået for denne testen.
- **ii.** Beregn p-verdien basert på dataene i tabell 1 og en test som består i å forkaste H_0 hvis $\bar{X}$ blir stor nok.
- **iii.** Formuler en konklusjon basert på resultatet i delpunkt **ii.**

<<< **Svar: i.:** Nivået er gitt ved

$$P_{\mu=245}(\text{forkast } H_0) = P_{\mu=245}(\bar{X} > 246) = 1 - P_{\mu=245}\left(Z \le \frac{246-245}{\sqrt{5/10}} \right) = 1 - G(1.41) = 1 - 0.9207 = 0.0793$$

Nivå 7.93%

ii.: Observert $\bar{X}_{obs} = 246.4$. P-verdien blir

$$\hat{\alpha} = P_{\mu=245}\left(\bar{X} > \bar{X}_{obs}\right) = P_{\mu=245}\left(\bar{X} > 246.4\right) = 1 - G\left(\frac{246.4-245}{\sqrt{5/10}} \right) 1 - G(1.98) = 1 - 0.9761 = 0.0239$$

iii: En p-verdi på 2.4% gir forkastning på 2.5% som som oftest betraktes som tilstrekkelig evidens for å forkaste H_0.

>>>

B. Vi antar ikke lenger at variansen, σ^2, er kjent. Dette betyr at σ^2 er ukjent og må estimeres hvis den inngår i testen.

- **i.** Sett opp en test for H_0 med signifikansnivå 5% basert på $X_1, X_2, \ldots, X_{10}$.
- **ii.** Gjennomfør testen basert på data i tabell 1 og formuler en konklusjon.

<<< **Svar:** **i:** T-test siden $X_i - ene$ er uid og normalfordelte $N(\mu,\sigma)$ med μ,σ ukjente. Testobservator $T = \frac{\bar{X} - 245}{S}\sqrt{10} \overset{\text{eksakt}}{\sim} t(9) -$ fordelt hvis $\mu = 245$. 5%-kvantilen i t(9)-fordelingen er 1.833 som blir kritisk verdi for *T*. Forkast H_0 hvis $T > 1.833$.

ii. Finner $9S_{obs}^2 = \sum x_i^2 - 10\bar{x}^2 = 607168 - 10(246.4)^2 = 38.4 \quad \Rightarrow \quad S_{obs}^2 = 4.2667$ hvorav

$T_{obs} = \frac{246.4 - 245}{\sqrt{4.2667}}\sqrt{10} = 2.143$ som gir forkastning av H_0 på 5%. Sterk evidens for at koffein virker.

>>>

C. Vi ønsker et 90% konfidensintervall for økningen i forventet hastighet, $\theta = \mu - 245$.

i. Begrunn følgende regel: Hvis $[A, B]$ er et $1-\alpha$ konfidensintervall for μ, er $[A - 245,\ B - 245]$ et $1-\alpha$ konfidensintervall for $\theta = \mu - 245$.

ii. Sett opp og beregn et 90% konfidensintervall for $\theta = \mu - 245$ basert på data i tabell 1 når variansen σ^2 er ukjent.

<<< **Svar:**

i. Åpenbart

$$1-\alpha = P(A \le \mu \le B) = P\left(A - 245 \le \mu - 245 \le B - 245\right) = P\left(A - 245 \le \theta \le B - 245\right)$$

ii. Basert på at $W = \frac{\bar{X} - \mu}{S}\sqrt{10} \sim t(9)$-fordelt for alle μ og 1.833 er 5% kvantilen i t(9), får vi $P(-1.833 \le W \le 1.833) = 0.90$, som gir 90% konf.intervallet $\bar{X} \pm 1.833\frac{S}{\sqrt{10}}$ for μ, eller observert $246.4 \pm 1.833\sqrt{\frac{4.2667}{10}} = 246.4 \pm 1.197 = [245.2, \quad 247.6]$

Og for økningen θ ifølge punkt **i.:** **[**0.2**,** 2.6**]**

>>>

Oppgave 3

Anne har hatt en bil noen år og er interessert i å finne ut hvor mange liter bensin bilen bruker pr. mil ved landeveiskjøring[2]. Hun har foretatt tre målinger gitt i tabell 2:

[2] 1 norsk mil = 10 kilometer

Tabell 2

Antall mil kjørt (x_i)	11	16.1	21
Liter bensin brukt (y_i)	8.1	15.1	19.2

La den stokastiske variabelen Y_i representere antall liter bensin brukt på en kjøretur på x_i mil der x_i antas ikke-stokastisk $(i = 1,2,3)$. Vi antar dessuten at Y_1, Y_2, Y_3 er stokastisk uavhengige med forventning og varians, $E(Y_i) = x_i\beta$ og $\mathrm{var}(Y_i) = x_i\sigma^2$, $i = 1,2,3,$ der β og σ^2 er ukjente parametre.

i. Forklar hvorfor β kan tolkes som forventet antall liter bilen bruker pr mil.

ii. Vis at estimatorene $\hat{\beta}$ og $\tilde{\beta}$ begge er forventningsrette, der

$$\hat{\beta} = \frac{Y_1 + Y_2 + Y_3}{x_1 + x_2 + x_3}, \qquad \tilde{\beta} = \frac{1}{3}\left(\frac{Y_1}{x_1} + \frac{Y_2}{x_2} + \frac{Y_3}{x_3}\right)$$

iii. Finn uttrykk for $\mathrm{var}(\hat{\beta})$ og $\mathrm{var}(\tilde{\beta})$ og bestem hvilken av dem som er minst (for x_i-ene i tabell 2).

iv. Estimer β ut fra målingene i tabell 2 ved å bruke den beste av de to estimatorene.

<<< **Svar:**

i: Forventet antall liter pr mil blir $E(Y_i/x_i) = E(Y_i)/x_i = (x_i\beta)/x_i = \beta$

ii:
$$E(\hat{\beta}) = E\left(\frac{\sum Y_i}{\sum x_i}\right) = \frac{\sum E(Y_i)}{\sum x_i} = \frac{\sum x_i\beta}{\sum x_i} = \beta$$

$$E(\tilde{\beta}) = \frac{1}{3}\sum E\left(\frac{Y_i}{x_i}\right) = \frac{1}{3}\sum\left(\frac{EY_i}{x_i}\right) = \frac{1}{3}\sum\left(\frac{x_i\beta}{x_i}\right) = \frac{1}{3}\sum\beta = \beta$$

iii:
$$\mathrm{var}(\hat{\beta}) = \frac{1}{\left(\sum x_i\right)^2}\sum \mathrm{var}(Y_i) = \frac{1}{\left(\sum x_i\right)^2}\sum x_i\sigma^2 = \frac{\sigma^2}{\sum x_i} = \frac{\sigma^2}{48.1} = 0.0208\,\sigma^2$$

$$\mathrm{var}(\tilde{\beta}) = \frac{1}{9}\sum\frac{\mathrm{var}(Y_i)}{x_i^2} = \frac{1}{9}\sum\frac{x_i\sigma^2}{x_i^2} = \left(\frac{1}{9}\sum\frac{1}{x_i}\right)\sigma^2 = 0.0223\,\sigma^2$$

som viser at $\hat{\beta}$ har minst varians.

iv: Estimat $\hat{\beta}_{obs} = \dfrac{\sum y_i}{\sum x_i} = \dfrac{42.4}{48.1} = 0.88$

>>>
