UiT Norges arktiske universitet

Kort hjemmeeksamen i:	STA-0001, Brukerkurs i statistikk 1
Dato:	28 september 2021
Tidspunkt:	09:00-13:00 + 30 minutter til innlevering i WISEflow
Kursansvarlig:	Elinor Ytterstad
Antall sider:	5 sider inklusive forsiden.
Support:	Du kan ringe 776 20 880 for support på eksamensdagen.
Vekting av spørsmål, eller annen informasjon:	Alle 10 delspørsmål (a),b), osv.) vektes likt ved bedømming.
Viktig informasjon om sitering og plagiering:	1. Dette er en individuell eksamen som skal besvares uten samarbeid med andre. 2. Alle hjelpemidler er tillatt (egne notater, pdfer fra forelesningene, lærebok, internett etc). 3. Alle eksamener som leveres i WISEflow blir automatisk sjekket for plagiat. Det er ikke tillatt å kopiere medstudenter, nettressurser, kilder, eller litteratur uten referanser.

I oppgaver der det eksplisitt står at en skal vise mellomregninger, vil korrekt svar ikke bli godkjent dersom mellomregninger mangler.

Det brukes desimalpunktum i dette oppgavesettet.
Oppgavesettet består av 10 delpunkter som alle teller likt ved bedømming.

Oppgave 1

Politiske partier levere inn sine valglister ei tid før et stortingsvalg. I ett av de store partiene er det enighet om hvilke kandidater som skal fylle de fem øverste plassene, men intern rekkefølge på toppen (fra nr 1 til nr 5) er de ikke enige om.

a)
- Hvor mange ulike rekkefølger kan en lage av de fem kandidatene som skal bekle de fem øverste plassene?
- Hvor mange ulike måter kan en velge ut tre kandidater til de øverste plassene (uten å ta hensyn til intern rekkefølge) fra gruppen av fem kandidater?

Det er to kvinner og tre menn i topp-fem gruppen.
Partiet har bestemt at rekkefølgen skal enten være "mann, kvinne, mann, kvinne, mann", eller "kvinne, mann, kvinne, mann, mann".

- Hvor mange ulike topp-fem permutasjoner kan en lage, som også oppfyller kravet til kjønnsfordeling?

Løsningsforslag:

- $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$
- $5 \cdot 4 \cdot 3 = 60$
- Antall permutasjoner med kvinne først: $2 \cdot 3 \cdot 1 \cdot 2 \cdot 1 = 12$
  Antall permutasjoner med mann først: $3 \cdot 2 \cdot 2 \cdot 1 \cdot 1 = 12$,
  Totalt $12 + 12 = 24$ permutasjoner.

Vi fortsetter med de to kvinnene og tre mennene i topp-fem gruppen, og at det ikke er stilt krav til kjønnsfordeling i rekkefølgen. Anta også at plassene tildeles tilfeldig.

b)
- Finn sannsynligheten for at begge kvinner er plassert foran mennene?

- Finn sannsynligheten for ingen kvinner i de to øverste plassene, og finn sannsynligheten for ei kvinne og en mann i de to øverste plassene.
  Hva er forventet antall kvinner i de to øverste plassene?

Løsningsforslag:

- $X \sim$ antall kvinner i i to øverste plassene, hypergeometrisk fordelt med $N = 5, M = 2, n = 2$.
  $P(X = 2) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1 \cdot 1}{10} = 0.1$
  Eller vi kan utlede sannsynligheten direkte: $\frac{2 \cdot 1}{5 \cdot 4} = 0.1$
- $P(X = 0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{1 \cdot 3}{10} = 0.3$
  $P(X = 1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{2 \cdot 3}{10} = 0.6$
  $E(X) = 0 \cdot 0.3 + 1 \cdot 0.6 + 2 \cdot 0.1 = 0.8$
  Eller om vi bruker formelen for forventning i hypergeometrisk fordeling: $E(X) = n \cdot \frac{M}{N} = 2 \cdot \frac{2}{5} = 0.8$.

Oppgave 2

a) Anta X er en tilfeldig variabel som har sannsynlighetstetthet:

$$f(x) = \begin{cases} 4x^3 & \text{hvis } 0 \leq x \leq 1 \\ 1 & \text{ellers} \end{cases}$$

og kumulativ fordelingsfunksjon:

$$F(x) = \begin{cases} 0 & \text{hvis } x < 0 \\ x^4 & \text{hvis } 0 \leq x \leq 1 \\ 1 & \text{hvis } x > 1 \end{cases}$$

- Hva er definisjonen av en sannsynlighetstetthet, og vis at $f(x)$ har egenskapene som skal til for å være en sannsynlighetstetthet.
- Finn $P(0.5 < X \leq 1)$. Vis mellomregniger.

Løsningsforslag:

- * Totalt areal under kurve lik 1, $\int_0^1 4x^3 dx = 4 \cdot \frac{1}{4}x^4 \mid_0^1 = 1$,
  * $P(a < X \leq b)$ er lik arealet under kurven fra a til b og
  *$f(x) \geq 0$, alltid. Her er $4x^3$ alltid et positiv tall når $x > 0$.
- $P(0.5 < X \leq 1) = F(1) - F(0.5) = 1 - (0.5)^4 = \frac{15}{16} = 0.9375$

Oppgave 3

Ved en fødeklinikk i Israel ble n=215 førsteganssfødende bedt om å angi største smerte (Y) under fødsel på en skala fra 0 til 100.
Resultatet ble:

	$\bar{Y}$	s	n
Førstegangsfødende	89.95	14.83	215

a)
- Lag et 95 % konfidensintervall for forventningsverdien til Y.
- Hvilke forutsetninger om fordelingen til Y-ene er nødvendig for konfidensintervallet du har beregnet. Begrunn svaret.

Løsningsforslag:

- Et 95 % konfidensintervall for μ, som forutsetter uavhengige og normalfordelte Y-er, baserer seg på t- fordeling med $215 - 1 = 214$ frihetsgrader. Vi leser av nærmeste tabellverdi (100 frihetsgrader) og finner $t_{0.025} = 1.984$:
  $NG = \bar{y} - 1.984 \cdot \frac{14.83}{\sqrt{215}} = 89.95 - 2.01 = 87.94$
  $ØG = \bar{y} + 1.984 \cdot \frac{14.83}{\sqrt{215}} = 89.95 + 2.01 = 91.96$
  Dersom vi kun forutsetter uavhengighet (dvs. ikke normalfordeling), så vil sentralgrenseteoremet si oss at gjennomsnittet $\bar{Y}$ er tilnærmet normalfordelt og konfidensintervallet baserer seg på normalfordelingen:
  $NG = \bar{y} - 1.96 \cdot \frac{14.83}{\sqrt{215}} = 89.95 - 1.98 = 87.97$
  $ØG = \bar{y} + 1.96 \cdot \frac{14.83}{\sqrt{215}} = 89.95 + 1.98 = 91.93$
- Ikke nødvendig å forutsettemer enn uavhengighet. Dvs. Y-ene kan gjerne ha ei fordeling som ligner på sannsynlighetstettheten i a).

La X være varighet av fødsel i minutter. Vi skal anta at X er normalfordelt med forventning $\mu = 420$ (7 timer) og $\sigma = 180$.

b)
- Finn $P(X \leq 390)$.
- Hvor stor andel av fødslene varer i mer enn 600 minutter (10 timer).
- Finn $P(\bar{X} \leq 390)$, der $\bar{X}$ er gjennomsnittet av $n = 215$ fødsler.

Løsningsforslag:

- $P(X \leq 390) = G(\frac{390-420}{180}) = G(-0.17) = 0.4325.$

- $P(X > 600) = 1 - G(\frac{600-420}{180}) = 1 - G(1) = 1 - 0.8415 = 0.1585$. 15.85% av alle fødsler varer lenger enn 10 timer.
- $P(\bar{X} \leq 390) = G(\frac{390-420}{\frac{180}{\sqrt{215}}}) = G(\frac{-30}{12.28}) = G(-2.44) = 0.0073$.

I datamaterialet fra Israel, er også fødselsvarighet (i minutter) fra 215 førstegangfødende og 205 flergangsfødende (kvinner som har født barn tidligere) registrert. Resultatet oppsummeres som følger:

	$\bar{X}$	s	n
Førstegangsfødende	460	264	215
Flergangsfødende	362	190	205

c)
- Vis at fellesestimatet for standardavviket er $s_p = 230.9$.
- Undersøk med hypotesetest om fødselens varighet er lenger hos førstegangsfødende enn hos de som har født tidligere. Bruk 5 % signifikansnivå.
- Formulér konklusjonen med ord relatert til fødselsvarighet.

Løsningsforslag:

- $s_p^2 = \frac{214 \cdot 264^2 + 204 \cdot 190^2}{215+205-2} = 53299.87$
  $s_p = \sqrt{53299.87} = 230.9$
- Hypoteser: $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$, der μ_1 er forventet varighet hos førstegangsfødende.

  Vi bruker t-fordeling med $215 + 205 - 2 = 418$ frihetsgrader. Nærmeste tabellverdi er for 100 frihetsgrader: $t_{0.05} = 1.66$, dvs. forkast H_0 dersom testobervatoren er større enn 1.66.

  Testobservator: $T = \frac{460-362}{230.9\sqrt{\frac{1}{215}+\frac{1}{205}}} = \frac{98}{22.5} = 4.3$

  Verdien av testobservator er mye større enn 1.66 og nullhypotesen forkastes.
- Det er sterke indikasjoner i datamaterialet på at førstegangsfødene bruker (gjennomsnittlig) lenger tid i fødsel enn de som har født tidligere.

Oppgave 4

La X være en binomisk variabel.

a)
- Bruk tabellen i læreboka og finn $P(X \leq 4)$, når $n = 8$ og $p = 0.8$.
- Finn også $P(4 \leq X < 7)$, når $n = 8$ og $p = 0.8$. Vis mellomregninger.

Løsningsforslag:

- $P(X \leq 4) = 0.056$
- $P(4 \leq X < 7) = P(X \leq 6) - P(X \leq 3) = 0.497 - 0.010 = 0.487$

I det israelske datamaterialet med totalt $n = 420$ kvinner, fikk $X = 335$ epidural (smertelindring) under fødselen.
Vi skal anta at X er en binomisk variabel.

b)
- Finn et estimat for p, andel israelske fødende som får epidural. Dvs. hvor utbredt er bruk av epidural i Israel. Avrund estimatet til tre desimaler.
- Vis at $SE(\hat{p}) = 0.020$, dvs estimert standardavvik til $\hat{p}$ avrundet til tre desimaler. Vis mellomregninger.

Løsningsforslag:

- Estimator for p: $\hat{p} = \frac{X}{n}$.
  Verdien av estimatoren (dvs estimatet) er: $\hat{p} = \frac{335}{420} \approx 0.798$.
- $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{420}} = 0.0196 \approx 0.020$

c)
- Lag et 90 % konfidensintervall for p.

Løsningsforslag:

- Bruker at binomisk fordeling kan tilnærmes til normalfordeling fordi $n\hat{p}(1-\hat{p}) = 67.8$ er større enn 5 Da har vi intervallgrensene:
  $NG = \hat{p} - 1.645 \cdot SE(\hat{p}) = 0.798 - 1.645 \cdot 0.020 = 0.798 - 0.033 = 0.765$.
  $ØG = \hat{p} + 1.645 \cdot SE(\hat{p}) = 0.798 + 0.033 = 0.831$.
  Et 90 % KI for p: $[0.765, \quad 0.831]$

d) I Norge er epidural brukt ved 41 % av alle fødsler.

- Undersøk følgende hypoteser i det israelske datasettet:
  $H_0 : p \leq 0.41$ mot $H_1 : p > 0.41$.
- Formulér konklusjonen med ord relatert til epiduralbruk i Israel sammenlignet med Norge.

Løsningsforslag:

- Standardavviket til $\hat{p}$ under H_0 er: $\sqrt{\frac{0.41\cdot(1-0.41)}{420}} \approx 0.024$
  Testobservator: $Z = \frac{0.798-0.41}{0.024} \approx 16.2$
  Den minste p-verdien vi kan lese ut av normalfordelingstabellen er $P(Z > 3.00) = 0.0013$. Siden verdien av testobservatoren her er 16.2, så ervår p-verdi mye mindre enn 0.0013. Det betyr en overbevisende forkasting av nullhypotesen.
- Det er sterke indikasjoner i datamaterialt på at epidural er mer vanlig i Israel enn i Norge.

Kilde til datamaterialet fra Israel er:
"In Pain Thou Shalt Bring Forth Children: The Peak-and-End Rule in Recall of Labor Pain"
Eran Chajut et.al. 2014
DOI: 10.1177/0956797614551004

PS
Det er ingenting i denne kildenhenvisningen som er til hjelp når du skal besvare eksamensoppgavene.