Fakultet for naturvitenskap og teknologi

UiT NORGES ARKTISKE UNIVERSITET

EKSAMENSOPPGAVE

Eksamen i:	STA-0001 Brukerkurs i statistikk 1
Dato:	24. september 2019
Klokkeslett:	09:00-13:00
Sted:	Adm.bygget B154
Tillatte hjelpemidler:	Alle trykte og skrevne samt kalkulator
Type innføringsark (rute/linje):	Rute
Antall sider inkl. forside:	**4**
Kontaktperson under eksamen:	**Elinor Ytterstad**
Telefon/mobil:	**775 44015**
Vil det bli gått oppklaringsrunde i eksamenslokalet? Svar: JA Hvis JA: ca. kl.___1030_______	

NB! Det er ikke tillatt å levere inn kladdepapir som del av eksamensbesvarelsen. Hvis det likevel leveres inn, vil kladdepapiret bli holdt tilbake og ikke bli sendt til sensur.

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

Oppgave 1

Ved en taufabrikk ønsker en å undersøke strekkstyrken (belastning i kg) til en bestemt tautype.

Taufabrikken opplyser at denne tautypen har en forventet trekkstyrke på 265 kg, og at 90 % av tauene tåler en belastning på minst 239 kg.

Du skal anta i hele oppgaven at strekkstyrken X er normalfordelt med kjent $\sigma = 20$.

a) Anta taufabrikkens opplysning om forventet strekkstyrke er korrekt. Stemmer det da at 90 % av tauene tåler en belastning på minst 239 kg? Begrunn svaret med nødvendige beregninger (ikke hypotesetest).

Løsningsforslag:
$X \sim N(265, 20)$
$P(X \geq 239) = 1 - G(\frac{239-265}{20}) = 1 - G(-1.3) = 1 - 0.0968 = 0.9032$

Resultatet av belastningsprøver på 10 tilfeldig valgte taulengder ble:

275, 280, 240, 270, 230, 230, 250, 280, 285, 270.

Det oppgis at: $\sum_{i=1}^{10} x_i = 2610$ og $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 4140$

b)
- Finn medianen og gjennomsnittet for dette datamaterialet.
- Regn ut et 99% konfidensintervall for forventet strekkstyrke μ.

Løsningsforslag:
Medianen er gjennomsnittet av de to midterste verdiene, etter at data er ordnet fra minste til største: (270+270)/2

Gjennomsnittet: $\bar{x} = \frac{2610}{10} = 261$

Konfidensintervall for μ (minner om at σ er kjent, $\sigma = 20$):

Nedre grense $= 261 - 2.576 \cdot \frac{20}{\sqrt{10}} = 261 - 16.3 = 244.7$
Øvre grense $= 261 + 16.3 = 277.3$

Et 99% KI for μ: $[244.7, \quad 277.3]$

c) Undersøk med hypotesetest om forventet taustyrke er mindre enn det taufabrikken opplyser (265 kg).

Formulér hypoteser.
Skriv opp testobservator.
Gjennomfør testen på 5% signifikansnivå og konkludér.

Løsningsforslag:
$H_0 : \mu \geq 265$ vs $H_1 : \mu < 265$

Testobservator: $Z = \frac{\bar{X}-265}{\frac{20}{\sqrt{10}}}$
Forkast nullhypotesen dersom $Z \leq -1.645$.

Finner verdien av testobservator for dette datamaterialet:

$$Z = \frac{261 - 265}{\frac{20}{\sqrt{10}}} = -0.63$$

Testobservator er ikke mindre enn -1.645, nullhypotesen forkastes ikke. Det er ikke grunnlag i data til å hevde at strekkstyrken er mindre enn det taufabrikken opplyser.

d)
- Finn standardavvik s.
- Forklar den begrepsmessige forskjellen på σ og empirisk standardavvik s.

Løsningsforslag:

$s = \sqrt{\frac{4140}{9}} = \sqrt{460} = 21.4$

σ er standardavviket i populasjonen, mens s måler standardavviket i utvalget. Nye data gir ny verdi s, mens σ forblir uforandret. σ er som regel ukjent, og da s en god estimator.

Oppgave 2

Sannsynligheten for at en 90 år gammel mann skal dø innen han blir 91 år er 0.165, sannsynligheten for at en 91 år gammel mann skal dø innen han blir 92 år er 0.177.

a)
- Hvor stor andel av 90-åringene (menn) overlever til de fyller 91 år?
- Hva er sannsynligheten for at en 90 år gammel mann skal dø innen han blir 92 år?

Løsningsforslag:

Andel $1 - 0.165 = 0.835$, 83.5% av alle 90 åringene overlever til 91 år.

La A = "dø før 91 år" og B = "dø før 92 år" (forutsetter overlevd til minst 90 år).

Merk at $P(B \cap A) = P(A)$.

Får opplyst at $P(A) = 0.165$ og $P(B|\bar{A}) = 0.177$

Da er $P(B) = P(B \cap A) + P(B \cap \bar{A}) = P(A) + P(B \mid \bar{A})P(\bar{A}) = 0.165 + 0.177 \cdot (1 - 0.165) =$

Oppgave 3

Antall turistskip X som søker om å få anløpe Tromsø en bestemt dag har vist seg å være Poissonfordelt med sannsynlighetsfordeling

$$P(X = x) = \frac{2^x}{x!}e^{-2}, \quad x = 0, 1, 2, \cdots$$

a)
- Finn sannsynligheten for at flere enn tre turistskip søker om å få anløpe Tromsø en bestemt dag?
- Havnen kan maksimalt betjene tre turistskip. De tre første som søker får anløpe havnen, øvrige søkere blir vist til an annen havn nær Tromsø.

  Hvor stor må kapasiteten ved Tromsø havn bygges ut til, for at det skal være 95% sjanse for å kunne betjene samtlige skip som søker?

Løsningsforslag:
Sannsynligheten for at flere enn tre turistskip søker om å få anløpe Tromsø en bestemt dag er:
$P(X > 3) = 1 - P(X \leq 3) = 1 - 0.857 = 0.143$ (Tabell E.2)

Dersom kapasiteten er tre skip, da er sannsynligheten for å betjene samtlige skip som søker: $P(X \leq 3) = 0.857$. Altså mindre enn 95%, og kapasitet på tre skip er ikke tilstrekkelig.

Hvis havnen bygges ut til å betjene fire skip, øker denne sannsynligheten til $P(X \leq 4) = 0.947 \approx 0.95$

Kapasiteten må derfor bygges ut til fire turistskip.

b) I år er det registrert følgende antall turistskip som har ønsket å anløpe Tromsø på ti utvalgte datoer: 3, 4, 0, 2, 2, 5, 3, 4, 3, 4,
dvs. totalt $Y = 30$ turistskip over ti dager.

- Forklar hvorfor Y er Poissonfordelt og la λ betegne den daglige raten.
- Finn et 95% konfidensintervall for λ.

Løsningsforslag:
Y er antall skip som ønsker å anløpe Tromsø over en tidagersperiode. Y er en sum av Poissonforekomster over $t = 10$ dager, $Y \sim Poisson(\lambda{\cdot}10)$.

$\hat{\lambda} = \frac{30}{10} = 3$

Et 95% KI for λ:
$N = 3 - 1.96 \cdot \sqrt{\frac{3}{10}} = 3 - 1.1 = 1.9$ og $\text{Ø} = 3 + 1.1 = 4.1$
Altså $[1.9, \quad 4.1]$

Oppgave 4

Hos kjøpmann Andersen avrundes kundens totale kjøpesum i kassa til nærmeste krone. Det gjelder både kunder som betaler kontant og med bankkort. En kunde som handler for kr 1240.76 betaler kr 1241, mens en som handler for kr 1240.05 betaler kr 1240.

La X være kjøpmannens tap (i kr) for hver kunde, dvs. X = "kjøpesum − betalt beløp". Vi skal anta at X er uniformt fordelt mellom -0.50 og 0.50, dvs. at alle verdier mellom -0.50 og 0.50 er like sannsynlige.

Sannsynlighetstetthetsfunksjonen er da

$$f(x) = \begin{cases} 1 & \text{hvis } x \in [-0.50,\ 0.50) \text{ kr} \\ 0 & \text{ellers} \end{cases}$$

Det oppgis at $\mathrm{Var}(X) = \frac{1}{12}$.

a)
- Vis at $\mathrm{E}(X) = 0$
- Finn $P(X \leq -0.10),\ P(X \geq 0.10)$ og $P(-0.10 < X < 0.10)$

Løsningsforslag:
$\mathrm{E}(X) = \int_{-0.5}^{0.5} x \cdot 1dx = \frac{1}{2}x^2 \mid_{-0.5}^{0.5} = 0$

$P(X \leq -0.10) = \int_{-0.50}^{-0.10} 1dx = 0.40$,
$P(X \geq 0.10) = \int_{0.10}^{0.50} 1dx = 0.40$ og
$P(-0.10 < X < 0.10) = \int_{-0.10}^{0.10} 1dx = 0.20$

Kan også løses ved å gjøre arealbetraktninger av funksjonen f

En dag det var 40 kunder innom butikken, fordelte kjøpmannens tap per kunde seg slik:

Tap per kunde (i kr)	[-0.50 , -0.10)	[-0.10 , 0.10)	[0.10 , 0.50)
Antall kunder	20	8	12

b) Bruk modelltesten til å undersøke om det er korrekt å anta at X er uniform fordelt.
Formulér en nullhypotese med ord, regn ut testobservatoren og forklar hvordan du bruker den til å konkludere i testen. Hva blir din konklusjon, bruk 5% signifikansnivå.

Løsningsforslag:

Modellen (uniform fordelingen) gir oss følgende sannsynligheter i kategoriene i tabellen:
$p_1 = 0.40$, $p_2 = 0.20$ og $p_3 = 0.40$ (beregnet i forrige punkt)

Hypoteser: $H_0: \; p_1 = 0.40, \quad p_2 = 0.20, \quad p_3 = 0.40$ mot
H_1 : to eller alle tre p_i forskjellig fra det som er angitt i nullhypotesen.

Vi kan beregne forventet antall i hver kategori:
$E_1 = 40 \cdot p_1 = 16, \quad E_2 = 40 \cdot 0.20 = 8$ og $E_3 = 40 \cdot 0.40 = 16$

Tap per kunde (i kr)	[-0.50 , -0.10)	[-0.10 , 0.10)	[0.10 , 0.50)
Antall kunder	20	8	12
Forventet antall (E)	16	8	16
$(X - E)^2$	4^2	0	$(-4)^2$

Regner ut testobservatoren: $Q = \frac{(20-16)^2}{16} + \frac{(8-8)^2}{8} + \frac{(12-16)^2}{16} = 2.$

Denne testobservatoren er tilnærmet kji-kvadratfordelt med $k - 1 = 3 - 1 = 2$ frihetsgrader. Dvs. at forkastingsområdet er: $Q > 5.99$

Vår verdi av Q=2 er ikke større enn 5.99, og nullhypotesen skal ikke forkastes.Det er ikke grunnlag for å hevde at uniform modell er feil.

c) En annen dag var det 36 kunder innom butikken.

- Forklar hvorfor kjøpmannens gjennomsnittstap denne dagen, $\bar{X}$, er tilnærmet normalfordelt med forventningsverdi 0 og standardavvik 0.048 (avrundet til tre desimaler).
- Finn $P(\bar{X} > 0.10)$ og $P(-0.10 < \bar{X} < 0.10)$
- Forklar (uten å regne på det) hva som skjer med sannsynligheten $P(\bar{X} > 0.10)$ når gjennomsnittet regnes over alle kunder i løpet av ei uke, en måned, ett år.

Løsningsforslag:
Et gjennomsnitt av mange ($n \geq 30$) uavhengige variabler fra samme fordeling vil være tilnærmet normalfordelt i flg. sentralgrenseteoremtet.

Her er $\mathrm{E}(X) = 0$, $\mathrm{Var}(X) = \frac{1}{12}$ og $n = 36$ da blir
$\mathrm{E}(\bar{X}) = 0$, $\mathrm{Var}(\bar{X}) = \frac{\frac{1}{12}}{36} = 0.0023$
og dermed er standardavviket til $\bar{X}$ lik $\sqrt{0.0023} = 0.048$

$P(\bar{X} > 0.10) = 1 - G(\frac{0.10-0}{0.048}) = 1 - G(2.08) = 1 - 0.9812 = 0.0188$

$P(-0.10 < \bar{X} < 0.10) = G(\frac{-0.10-0}{0.048}) - G(\frac{0.10-0}{0.048}) = G(2.08) - G(-2.08) = 0.9812 - 0.0188 = 0.9624$

n øker når vi utvider observasjonsperioden til uke, måned, år.
Når n øker så avtar standardavviket i fordelingen til gjennomsnittet, og dermed avtar sannsynligheten $P(\bar{X} > 0.10)$