EKSAMEN I: STA100 SANNSYNLIGHETSREGNING OG STATISTIKK

Universitetet i Stavanger

VARIGHET: 4 TIMER DATO: 29. AUGUST 2018

HJELPEMIDLER: Godkjent enkel kalkulator.

Eksamen består av 3 oppgaver på 3 sider og 5 sider vedlegg, total 8 sider.

EMNEANSVARLIG: Jan Terje Kvaløy TELEFON: 51 83 22 55

Oppgave 1

Tvillingpar er enten eneggede eller toeggede. Sannsynligheten for at det ved en tvillingfødsel blir født eneggede tvillinger er i Nord-Europa omtrent 30%, og vi vil benytte denne sannsynligheten for enegget tvillingfødsel i denne oppgaven.

De to tvillingene i et enegget tvillingpar har alltid samme kjønn. Vi antar at ved enegget tvillingfødsel er sannsynligheten for to jenter 48.6%. Vi antar videre at for toeggede tvillinger er kjønnet til den ene tvillingen uavhengig av kjønnet til den andre, og at for hver enkelt av tvillingene er sannsynligheten for jente 48.6%.

a) Hva er sannsynligheten for to jenter ved en toegget tvillingfødsel?

Hva er sannsynligheten for en gutt og en jente ved en tvillingfødsel?

b) Hva er sannsynligheten for at begge tvillingene i et tvillingpar er gutter?

Dersom begge tvillingene i et tvillingpar er gutter, hva er sannsynligheten for at de er eneggede?

Oppgave 2

Vestlandschips AS lager chips som pakkes i pakker som skal inneholde en viss mengde chips. I praksis er det vanskelig å få produksjonsutstyret til å porsjonere ut nøyaktig samme mengde chips i hver pakke så det vil være en viss variasjon i mengden chips fra pakke til pakke. Anta at mengden chips i en pakke er normalfordelt med forventning μ gram og standardavvik $\sigma = 10$ gram.

Vi skal i første omgang se på en situasjon der de produserer pakker som skal inneholde 250 gram chips. Anta i punkt a) at produksjonsutstyret er innstilt slik at $\mu = 250$ gram.

a) Finn sannnsynligheten for at mengden chips i en pakke er mindre enn 225 gram.

Finn også sannsynligheten for at mengden chips er mellom 225 og 275 gram.

En grossist foreslår ovenfor Vestlandschips at de burde justere produksjonen slik at forventet mengde chips i en pakke, μ, blir så høy at de kan garantere at minst 95% av pakkene inneholder 250 gram chips eller mer. Regn ut hvilken verdi μ eventuelt må justeres til for at Vestlandschips skal kunne overholde en slik garanti.

Vestlandschips bestemmer seg for å kjøre en kampanje hvor de i en periode har 20% mer chips i pakkene men selger dem til samme pris. Dvs de ønsker å justere produksjonen slik at forventet mengde chips per pakke, μ, blir 300 gram. De er imidlertid usikre på om de har klart å justere μ til rett verdi, så for å sjekke dette undersøker de mengden chips i 50 tilfeldig valgte pakker. De finner at gjennomsnittlig mengde chips i de 50 pakkene er 296.1 gram. Anta fremdeles at $\sigma = 10$ gram.

b) Bruk informasjonen over til å regne ut et 95% konfidensintervall for μ.

Hvor mange målinger må Vestlandschips gjøre dersom de ønsker at lengden på 95% konfidensintervallet skal være maksimalt 5 gram?

Et forbrukermagasin får et tips fra en kunde som mener at forventet mengde chips i kampanjepakkene er for lav og bestemmer seg for å undersøke dette. Siden magasinet ikke kjenner standardavviket σ må vi nå anta at det er ukjent. Forbrukermagasinet sjekker vekten på 20 tilfeldig valgte pakker og finner et gjennomsnitt på 293.9 gram og et utvalgsstandardavvik på 11.7 gram.

c) Formuler kunden sin påstand som en hypotesetest, og utfør testen. Bruk 5% signifikansnivå. Si hva utfallet av testen i praksis betyr.

Etter kampanjeperioden justerer Vestlandschips igjen produksjonen slik at forventet mengde chips per pakke er 250 gram. Vi skal i resten av oppgaven anta at $\mu = 250$ og $\sigma = 10$.

Vestlandschips bestemmer seg også for å begynne å bruke statistisk prosesskontroll for å holde kontroll med om produksjonen er stabil. De vil med jevne mellomrom ta ut en tilfeldig stikkprøve på n pakker fra produksjonen og veie disse, men lurer på hvor strenge kontrollgrenser de skal bruke. La $\overline{X}$ betegne gjennomsnittsvekten av en stikkprøve med n pakker.

La Y betegne antall stikkprøver til første gang man får en stikkprøve med en gjennomsnittsvekt som ligger *utenfor* intervallet $[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$.

d) Vis at $P(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$.
Hvilken fordeling har Y? Begrunn svaret?
Forklar hvorfor ARL $= \mathrm{E}(Y) = 1/\alpha$.
Regn ut $P(Y > 100)$ når $\alpha = 0.01$.

Vestlandschips bestemmer ser for å ta en stikkprøve med $n = 6$ pakker hver time, og de ønsker et kontrolldiagram hvor forventet antall stikkprøver mellom hver falske alarm er ARL $= 500$.

Resultatet av 14 stikkprøver er gitt i tabellen under, hvor $\overline{x}_i$ angir gjennomsnittetsvekten i stikkprøve nr i.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\overline{x}_i$	252.4	243.5	249.1	252.6	257.7	259.9	261.1	264.7	258.2	247.0	249.4	251.1	244.3	253.3

e) Regn ut kontrollgrensene for et $\overline{x}$-diagram med ARL $= 500$ (husk at verdiene for μ og σ angitt før punkt d) gjelder fremdeles).

Plott verdiene i tabellen over inn i kontrolldiagrammet og kommenter resultatet.

Når man skal beregne kontrollgrensene for et $\overline{x}$-diagram er i praksis ofte μ og σ ukjente og må estimeres fra data. Hvilken betydning har dette for ARL-verdien til diagrammet?

Oppgave 3

I evalueringen av STA100 både våren 2016 og våren 2018 ble studentene spurt om de støtter en omlegging av forelesningene i faget til rene videoforelesninger. Studenten ble bedt om å svare på om de er for eller mot en slik omlegging, og resultatene er gitt i tabellen under.

	antall for	antall mot
evaluering 2016	69	145
evaluering 2018	41	112

a) Bruk en kjikvadrattest til å avgjøre om undersøkelsen gir grunnlag for å konkludere at det var signifikant forskjell mellom studentene i 2016 og studentene i 2018 i holdningen til videoforelesninger. Bruk 5% nivå.

Er forutsetningene for testen oppfylte? Kommenter kort.

En annen måte å vurdere resultatene på er ved å estimere forskjell i andel studenter som var for omlegging i hhv 2016 og 2018. La p_1 betegne andelen av alle studenter i STAT100 våren 2016 som var for omlegging og la p_2 betegne andel av alle studenter i STA100 våren 2018 som var for omlegging. Da er $p_1 - p_2$ forskjellen i andel studenter som var for omlegging de to årene.

La videre X_1 være antall studenter blant de n_1 som besvarer evalueringen som er for omlegging i 2016 og X_2 være antall studenter blant de n_2 som besvarer evalueringen som er for omlegging i 2018. Vi skal anta (som en tilnærming) at $X_1 \sim \text{Bin}(n_1, p_1)$ og $X_2 \sim \text{Bin}(n_2, p_2)$. Vi antar også at X_1 og X_2 er uavhengige.

En rimelig estimator for $p_1 - p_2$ er da $\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$.

b) Bruk tallene i tabellen i starten av oppgaven til å regne ut estimater for p_1 og p_2.

Vis at $\hat{p}_1 - \hat{p}_2$ er forventningsrett.

Regn ut et uttrykk for $\text{Var}(\hat{p}_1 - \hat{p}_2)$.

Forklar hvorfor $\hat{p}_1 - \hat{p}_2 \approx \text{N}(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}})$.

I punkt c) under kan du bruke følgende tilnærminger som hjelp:
$\frac{p_1(1-p_1)}{n_1} \approx \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}$ og $\frac{p_2(1-p_2)}{n_2} \approx \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$.

c) Bruk resultater fra punkt b) over som utgangspunkt og utled et 95% konfidensintervall for $p_1 - p_2$. Regn ut intervallet for tallene gitt i starten av oppgaven.

Tyder intervallet på at det har vært en endring i andel studenter som er for omlegging? Kommenter kort.

Er antagelsene om fordelingene til X_1 og X_2 gjort før punkt b) nødvendigvis rimelig? Kommenter kort og forklar hva eventuelle avvik har å si for resultatet av estimatet og konfidensintervallet.