UiT
NORGES ARKTISKE UNIVERSITET

EKSAMENSOPPGAVE

Eksamen i: STA-0001 Brukerkurs i statistikk 1

Dato: Mandag 30. mai 2016

Tid: Kl 09:00-13:00

Sted: Åsgårdvegen 9

Tillatte hjelpemidler: Alle trykte og skrevne, samt kalkulator

Oppgavesettet har 4 sider inkludert forside

Kontaktperson under eksamen: Elinor Ytterstad

Telefon: 776 44015

NB! Det er ikke tillatt å levere inn kladd sammen med besvarelsen

Oppgave 1

Det antaes kjent at en viss type laboratorierotter har en normalfordelt vektøkning Y de første tre måneder etter fødsel med $\mu = 70$ gram og $\sigma = 3$ gram. Dvs $Y \sim N(70, 3)$.

a) Finn følgende sannsynligheter og vis utregninger:
$P(Y > 69)$
$P(Y \leq 64)$
$P(64 \leq Y \leq 70)$
$P(\bar{Y} \leq 69)$, der $\bar{Y}$ er gjennomsnitt av $n = 9$ rotter
Løsningsforslag:
$P(Y > 69) = 1 - P(Y \leq 69) = 1 - G(\frac{69-70}{3}) = 1 - G(-0.33) = 1 - 0.3707 = 0.6293$
$P(Y \leq 64) = G(\frac{64-70}{3}) = G(-2) = 0.0228$
$P(64 \leq Y \leq 70) = P(Y \leq 70) - P(Y < 64) = G(\frac{70-70}{3}) - 0.0228 = 0.5 - 0.0228 = 0.4772$
$P(\bar{Y} \leq 69) = g(\frac{69-70}{\frac{3}{\sqrt{9}}}) = G(-1) = 0.1587$

Det er gjort et eksperiment med $n = 5$ rotter som lever i fullstendig mørke hele døgnet, for å undersøke om vektøkning Y de første tre måneder etter fødsel er mindre enn 70 gram.
Resultatet ble: 71, 65, 65, 66, 68 gram

b) Finn medianen m.
Regn ut gjennomsnittet.
Regn ut standardavviket i utvalget, vis mellomregninger.
Løsningsforslag:
Ordner observasjonene i stigende rekkefølge: 65, 65, 66, 68, 71.
Medianen er den midterste $m = 66$
Gjennomsnitt: $\bar{Y} = \frac{71+65+65+66+68}{5} = 67.0$
Utvalgsvarians: $s^2 = \frac{(71-67.0)^2+(65-67.0)^2+(65-67.0)^2+(66-67.0)^2+(68-67.0)^2}{5-1} = \frac{26.0}{4} = 6.5$
Utvalgsstandardavvik $s = \sqrt{6.5} = 2.55$

Det er uvisst om vektøkning hos rotter som vokser opp i totalt mørke har samme standardavvik som i a).

c) Lag et 99% konfidensintervall for standardavviket σ. (Hint, finn først et konfidensintervall for σ^2.)
Forklar hvordan vi kan bruke dette intervallet til å teste $H_0 : \sigma = 3$ mot $H_0 : \sigma \neq 3$ på 1% signifikansnivå. Du kan besvare dette spørsmålet selv om du ikke har regnet ut konfidensintervallet.
Løsningsforslag:
Et 99% konfidensintervall for σ^2: Bruker kjikvadratfordeling med $n - 1 = 5 - 1 = 4$ frihetsgrader (Tabell E.6). Fordi det er et 99% KI, trenger vi 0.5-prosentilen $\chi_{0.005} = 14.86$ og 99.5-prosentilen $\chi_{0.995} = 0.21$

Nedre grense $N = \frac{(5-1)s^2}{14.86} = \frac{4 \cdot 6.5}{14.86} = 1.75$
Øvre grense $Ø = \frac{(5-1)s^2}{0.21} = \frac{4 \cdot 6.5}{0.21} = 123.81$
Et 99% konfidensintervall for σ blir da $(\sqrt{1.75}, \quad \sqrt{123.81}) \Rightarrow (1.32, \quad 11.13)$
Dersom et 99% konfidensintervall ikke dekker H_0-verdien $\sigma = 3$ forkastes H_0 til fordel for tosidig alternativ på signifikansnivå $(1 - 0.99) = 0.01$. Her inneholder intervallet $\sigma = 3$ og nullhypotesen forkastes ikke på 1% nivå.

d) Anta kjent $\sigma = 3$ og undersøk med hypotesetest om rotter som vokser opp i totalt mørke har lavere vektøkning enn de som vokser opp under normale lysforhold.
Formuler hypoteser.
Hva blir forkastingsområdet?
Gjennomfør testen på 5% signifikansnivå.
Konklusjonen skal formuleres i forhold til rotter og vektøkning.
Løsningsforslag:
$H_0 : \mu \geq 70$ mot $H_1 : \mu < 70$, ensidig alternativ hypotese.
Fordi $\sigma = 3$ antaes kjent skal vi bruke Z testen som sier at H_0 forkastes dersom $Z = \frac{\bar{Y} - 70}{\frac{3}{\sqrt{5}}} \leq -z_{0.05} = -1.645$, dvs forkastingsomårdet: $(-\infty, -1.645]$.
Alternativt: forkast H_0 dersom $\bar{Y} \leq 70 - 1.645 \cdot \frac{3}{\sqrt{5}} = 67.8$
Vi regner ut hva verdien av testobservat Z blir med våre data: $Z = \frac{67.0 - 70}{\frac{3}{\sqrt{5}}} = -2.24$
Konklusjon: H_0 forkastes på 5% nivå fordi $Z \leq -1.645$. Det er grunnlag i data til å hevde at manglende dagslys de tre første måneder gir lavere vektøkning enn hvis rottene hadde hatt normale lysforhold.

Alle fem rotter er utstyrt med en aktivitetsmåler som leses av samtidig med registrering av vektøkning. Aktivitetet er omregnet til et tall medllom 0 og 10, der 10 er høy aktivitet. Resultatet ble:

Vekt Y	71	65	65	66	68
Aktivitet X	2	9	8	8	5

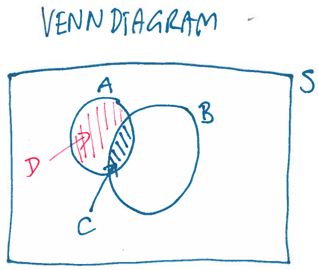
$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = -29.0, \quad \sum(X_i - \bar{X})^2 = 33.2$ og $\sum(Y_i - \bar{Y})^2 = 26.0$

e) Regn ut korrelasjonskoeffisienten for vekt og aktivitet.
Regn ut et estimat for stigningskoeffisienten til regresjonslinja, $\hat{\beta}$.
Regn ut et 98% konfidensintervall for β, du kan bruke at $SE(\hat{\beta}) = 0.082$
Beskriv sammenhengen mellom aktivitet og vekt, bruk maks to setninger.
Løsningsforslag:
Korrelasjonskoeffisient $r = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}} = \frac{-29.0}{\sqrt{33.2 \cdot 26.0}} = -0.987$
$\hat{\beta} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} = \frac{-29.0}{33.2} = -0.873$
Til konfidensintervallet trenger vi 99-prosentilen fra t-fordelingen med $n - 2 = 3$ frihetsgrader, $t_{0.99} = 4.541$
$N = \hat{\beta} - 4.541 \cdot SE(\hat{\beta}) = -0.873 - 0.372 = -1.245$
$Ø = \hat{\beta} + 4.541 \cdot SE(\hat{\beta}) = -0.873 + 0.372 = -0.501$
98% KI for β: (-1.245, -0.501)
Korrelasjonskoeffisienten er ganske nær -1, som betyr at det er en sterk negative sammenheng mellom aktivitet og vektøkning hos rotter de tre første måneder etter fødsel. Høyt aktivitetsnivå gir mindre vektøkning.

Oppgave 2

La A og B være to hendelser hvor $P(A \cap B) > 0$.
La $C = A \cap B$ og $D = A \cap \bar{B}$

a) Tegn hendelsene A og B inn i et Venndiagram og skravér hendelsene C og D
Er hendelsene C og D disjunkte? Begrunn svaret
Er hendelsene C og D uavhengige? Begrunn svaret
Løsningsforslag:



Er hendelsene C og D disjunkte? Ja, fordi $P(C \cap D) = 0$, dvs. hendelsene C og D har ikke noe felles.
C og D er uavhengige dersom $P(C \cap D) = P(C)P(D)$. Her er venstre $P(C \cap D) = 0$ fra forrige spørsmål. Høyre side er lik 0 kun hvis $P(D) = 0$. Altså er
C og D er uavhengige hvis $P(D) = 0$ og
C og D er avhengige hvis $P(D) > 0$.

Norsk barneskole har syv klassetrinn.
La A = elev i 7. klassetrinn, $\bar{A}$=elev i 1. -6. klassetrinn og $P(A) = \frac{1}{7}$
La B = elev er mobbet. $P(B|A) = 0.05$ og $P(B|\bar{A}) = 0.1$
Definisjon av mobbet er at eleven har vært mobbet/plaget minst 3 ganger per uke den siste måneden.

b) Regn ut sannsynligheten for en tilfeldig valgt elev i barneskolen er mobbet.
Regn ut $P(A|B)$
Løsningsforslag:
$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.05 \cdot \frac{1}{7} + 0.1 \cdot \frac{6}{7} = 0.0929$
$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.05 \cdot \frac{1}{7}}{0.0929} = 0.0769$

La X = antall elever som er mobbet, av n tilfeldig valgte 7. klassinger.
Anta X er binomisk fordelt med $p = 0.05$.

c) Regn ut $P(X \leq 3)$ når $n = 10$
Regn ut $P(X \geq 2)$ når $n = 10$
Regn ut $P(2 \leq X < 3)$ uten bruk av binomisk tabell, når $n = 10$
Regn ut $P(X \geq 30)$ når $n = 500$
Løsningsforslag:
$n = 10,\ P(X \leq 3) = 0.999$ Tabell E.1
Eventuelt $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \binom{10}{0}0.05^0 0.95^{10} + \binom{10}{1}0.05^1 0.95^9 + \binom{10}{2}0.05^2 0.95^8 + \binom{10}{3}0.05^3 0.95^7 = 0.999$
$n = 10,\ P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.914 = 0.086$ Tabell E.1
Eventuelt $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - 0.5987 - 0.3151 = 0.0862$
$n = 10,\ P(2 \leq X < 3) = P(X = 2) = \binom{10}{2}0.05^2 \cdot 0.95^8 = 0.0746$

$n = 500$, $P(X \geq 30)$
Bruker normaltilnærming (regel 5.20) fordi $np(1-p) = 500 \cdot 0.05 \cdot 0.95 = 23.75 \geq 5$
Med heltallskorreksjon: $P(X \geq 30) = 1 - G(\frac{29+0.5-500\cdot 0.05}{\sqrt{23.75}}) = 1 - G(0.92) = 1 - 0.8212 = 0.1788$
Uten heltallskorreksjon: $P(X \geq 30) = 1 - G(\frac{29-500\cdot 0.05}{\sqrt{23.75}}) = 1 - G(0.82) = 1 - 0.7939 = 0.2061$

I en undersøkelse om mobbing i barneskolen svarte $X = 30$ av $n = 500$ i 7. klassetrinn at de hadde vært mobbet.
Vi skal gå ut fra at disse 500 elevene er tilfeldig valgte 7.klassinger fra norsk barneskole og skal bruke tallene til å estimere mobbeandelen p blant 7. klassinger.

d) Regn ut estimatet for p fra dette datamaterialet.
Regn ut et 95% KI for p.
Løsningsforslag:
Estimatet for p: $\hat{p} = \frac{X}{n} = \frac{30}{500} = 0.06$
95% KI for p: Her kan vi bruke normaltilnærming fordi $n\hat{p}(1-\hat{p}) = 500 \cdot 0.06 \cdot 0.94 = 28.2 \geq 5$ (Regel 5.20)
Må først regne ut $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.06 \cdot 0.94}{500}} = 0.0106$
$N = \hat{p} - 1.96 \cdot 0.0106 = 0.039$
$\text{Ø} = \hat{p} + 1.96 \cdot 0.0106 = 0.081$
Et 95% KI for p blir da: $(0.039, \quad 0.081)$

e) Undersøk med hypotesetest om mobbeandelen blant 7. klassinger er høyere enn 0.05.
Formulér hypoteser, forklar valg av testobservator og skriv en konklusjon.
Bruk 1 % signifikansnivå.
Regn ut p-verdien.
Løsningsforslag:
$H_0 : p \leq 0.05$ mot $H_1 : p > 0.05$
Bruker normaltilnærming med samme begrunnelse som i forrige punkt. Da er testobservator: $Z = \frac{X - 500 \cdot 0.05}{\sqrt{500 \cdot 0.05(1-0.05)}} = 1.03$
Denne verdien er ikke større enn $z_{0.01} = 2.326$, og nullhypotesen forkastes ikke.
P-verdi=$P(z \geq 1.03) \approx 1 - G(1.03) == 1 - 0.8485 = 0.1515$
Eventuelt kan vi bruke beregninger fra oppgave 2c): P-verdi=$P(X \geq 30) \approx 0.2061$, eller med heltallskorreksjon $P(X \geq 30) \approx 0.1788$.
Det er altså tre ulike svar på dette spørsmålet. Alle er tilnærmede sannsynligheter basert p normalfordeling med eller ute heltallskorreksjon. Den eksakte p-verdien finner vi ved å bruke binomisk fordeling (ikke mulig å gjøre under eksamen), da blir svaret 0.1765.

I samme undersøkelse er 500 elever i 5. klasse spurt om de er mobbet. Resultatet ble:

	5. klasse	7. klasse	
Mobbet	70	30	100
Ikke mobbet	430	470	900
	500	500	1000

f) Undersøk med hypotesetest om mobbeandelen er forskjellig i de to klassetrinnene
Formuler hypoteser, regn ut testobservator og skriv en konklusjon. Bruk 5% signifikansnivå.
Løsningsforslag:

Kjikvadrattest,
H_0 : Mobbeandelen er den samme i begge klassetrinn
H_1 : Mobbeandelen er forskjellig i de to klassetrinnene
Testobservatoren Q er kjikvadratfordelt med $(r-1)(k-1) = 1$ frihetsgrader og vi forkaster nullhypotesen dersom $Q \geq \chi_{0.05} = 3.84$.
Vi regner ut forventede antall $E_{1,1} = 100 \cdot \frac{500}{1000} = 50$, $E_{2,1} = 900 \cdot \frac{500}{1000} = 450$, $E_{1,2} = 100 \cdot \frac{500}{1000} = 50$, $E_{2,2} = 900 \cdot \frac{500}{1000} = 450$, og deretter
$Q = \frac{(70-50)^2}{50} + \frac{(430-450)^2}{450} + \frac{(30-50)^2}{50} + \frac{(470-450)^2}{450} = 17.8$
Nullhypotesen forkastes, mobbeandelen i 7. klassetrinn er signifikant forskjellig fra den i 5. klassetrinn.