

Oppgaver til Studentveiledning 3

MET 3431 Statistikk

24. april 2012 kl 17.15 - 20.15 i B2

Handelshøyskolen BI

Oppgaver

1. Eksamensoppgaver:

Eksamen 01/06/2011: Oppgave 1 - 7. Eksamensoppgaven fra 06/2011 er vedlagt.

Løsninger

1 Eksamensoppgaver:

Løsningsforslag til Eksamen 06/2011 er vedlagt.

Skriftlig eksamen i: MET 34311 Statistikk
 Eksamensdato: 01.06.11, kl. 09.00-14.00
 Tillatte hjelpemidler: Alle + BI-definert eksamenskalkulator :
 TEXAS INSTRUMENTS BA II Plus TM
 Innføringsark: Ruter
 Totalt antall sider: 5

Oppgave 1

I løpet av en dag har et massasjelinstitut 131 klienter. Sekretæren noterer følgende informasjon for hver klient: *Dato*, *Navn*, *Kjønn*, *Personnummer*, *Behandlingsform* og *Beløpet* som pasienten skal betale. Det er fire behandlingsformer: klassisk massasje, idrettsmassasje, triggerpunktsmassasje og bindevevs-massasje.

- På hvilket målenivå er variabelen *Behandlingsform*?
- På slutten av dagen legger sekretæren inn all innsamlet informasjon i en ny JMP datafil. Hvor mange rader og søyler(kolonner) har datafila?
- Histogrammet i figur 1 på side 4 viser fordelingen til *Beløp*. Omtrent hvor mange klienter betalte mindre enn 300 kroner?
- Gjennomsnittsbeløpet er 314.2 kroner mens medianbeløpet er 293 kroner. Kan du kort forklare ut fra histogrammet hvorfor gjennomsnittet er såpass mye større enn medianen?
- Sekretæren setter opp en kryss-tabell for sammenhengen mellom behandlingsform og kjønn:

	Behandlingsform			
	A	B	C	D
Mann	22	3	15	36
Kvinne	12	24	9	?

Hvor mange kvinner fikk behandlingsform D? Hvor mange prosent av klientene var menn?

- Sekretæren sender et spørreskjema til tre tilfeldig valgte klienter fra datafila. Hva er sannsynligheten for at alle de tre utvalgte klientene er menn?

Oppgave 2

For å få et innblikk i kundenes alderssammensetning spør ekspeditøren i en butikk hver tiende kunde hvilken alder hun/han har. Resultatet ble:

11, 15, 23, 26, 19, 22, 15, 25, 24

- Hvilket type utvalg er det ekspeditøren har framskaffet?
- Regn ut gjennomsnitt og standardavvik til alderen basert på stikkprøven. Vil du si at en kunde på 12 år har en 'uvanlig' alder? Begrunn svaret kort.
- Beregn nedre og øvre kvartil (Q_1 og Q_3), og finn interkvartilbredden $Q_3 - Q_1$ til alderen. Forklar også kort i hvilke situasjoner interkvartilbredden kan være et bedre mål for variasjon enn standardavviket.

Oppgave 3

- Du er i BI Nydalen og skal ta heisen. Anta at *antall sekunder* som går fra du trykker på heisknappen til heisen kommer er normalfordelt med $\mu = 22$ og $\sigma = 8$ sekunder. Hva er sannsynligheten for at du må vente mellom 25 og 30 sekunder på heisen? Skisser sannsynligheten som arealet under en graf.
- Anta at *antall minutter* en BI student bruker på å finne en ledig parkeringsplass er normalfordelt med gjennomsnitt μ og standardavvik σ . Foreslå verdier for μ og σ som du tror kan passe for ditt BI studiested. Hva er sannsynligheten for at en tilfeldig valgt student må bruke mer tid enn to standardavvik over gjennomsnittet for å finne parkeringsplass? Skisser denne sannsynligheten som arealet under en graf.

Oppgave 4

Direktørene for de forskjellige BI studiestedene ønsker å vite hvor mye deres studenter jobber per uke utenom studiene.

- I en stikkprøve på 52 studenter ved BI Nydalen finner man at gjennomsnittlig antall timer på jobb er $\bar{x} = 9.74$ timer med et standardavvik på $s = 6.86$ timer. Konstruer et 90 % konfidensintervall for populasjonsgjennomsnittet μ .
- Direktøren ved BI Stavanger spør 11 tilfeldig valgte studenter på sitt studiested hvor mye de jobber. Basert på disse svarene konstruerer han et konfidensintervall for μ . Diskuter kort om det er noen svakheter ved direktørens metode. Hvilken betingelse må være oppfylt for at direktørens konfidensintervall skal være gyldig?
- Direktøren ved BI Drammen henter inn en tilfeldig stikkprøve på 63 studenter. Han beregner konfidensintervallet for ukentlig antall timer på jobb:

$$5.263 < \mu < 11.658$$

Finn stikkprøvegjennomsnittet \bar{x} og feilmarginen E som ble brukt for å konstruere dette konfidensintervallet.

Oppgave 5

- (a) I hypotesetesting, hva vil det si å begå en type I feil? Anta at du tester på signifikansnivå $\alpha = 0.05$, og at H_0 er sann. Hva er sannsynligheten for å begå en type I feil?
- (b) I en stikkprøve på 157 studenter ved BI Bergen våren 2011 hadde 13 studenter mobiltelefon av merket Samsung. Test påstanden om at mindre enn ti prosent av studentene ved BI Bergen har Samsung mobiltelefon. Skriv opp nullhypotesen og alternativhypotesen. Foreta testen på et $\alpha = 0.05$ signifikansnivå og formuler konklusjonen i et lettfattelig språk.
- (c) Du har tilgang til stikkprøver av ukentlige timer på jobb for 60 mannlige studenter og for 99 kvinnelige studenter. Er de to stikkprøvene med menn og kvinner uavhengige eller relaterte? Test hypotesen om at mannlige studenter gjennomsnittlig jobber mer enn kvinnelige studenter basert på de to stikkprøvene vist i Figur 2. Skriv opp nullhypotesen og alternativhypotesen og bruk signifikansnivå $\alpha = 0.1$. Skisser p-verdien som arealet under en graf.

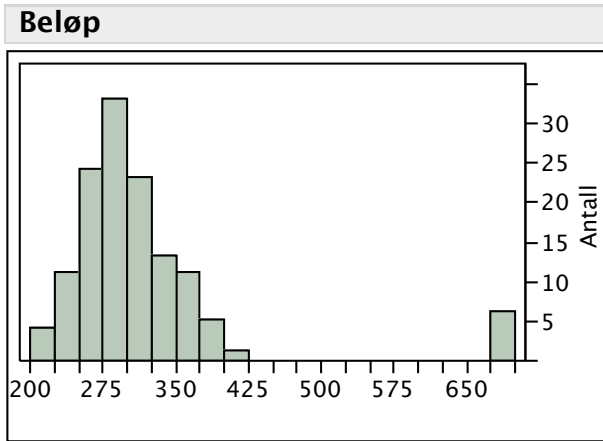
Oppgave 6

- (a) Figur 3 viser et scatterplott av timer på jobb vs studietimer per uke for en tilfeldig valgt stikkprøve av åtte BI studenter våren 2011. Hvor mange studenter i stikkprøven oppga at de studerte 30 timer i uka? Hvor mye jobbet hver av disse studentene i uka?
- (b) Korrelasjonskoeffisienten for dataene i Figur 3 er $r = 0.73$. Test på $\alpha = 0.05$ nivået om det er en korrelasjon mellom timer på jobb og timer brukt på studier. Skriv opp H_0 og H_1 .
- (c) Gi et eksempel på to kontinuerlige variable som du tror er signifikant negativt korrelerte.

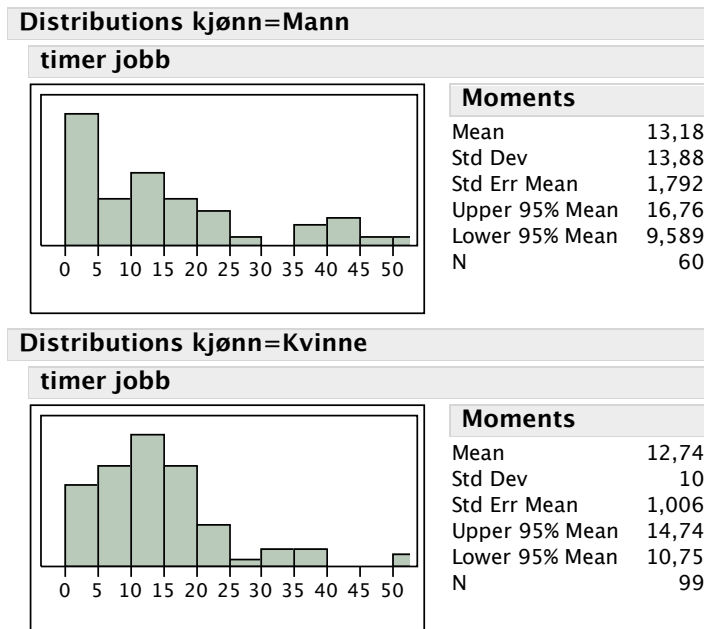
Oppgave 7

Figur 4 gir krysstabellen for sammenhengen mellom kjønn og studium i en tilfeldig stikkprøve av 291 studenter fra BI Trondheim. I hver celle oppgis det faktiske og det forventede antallet.

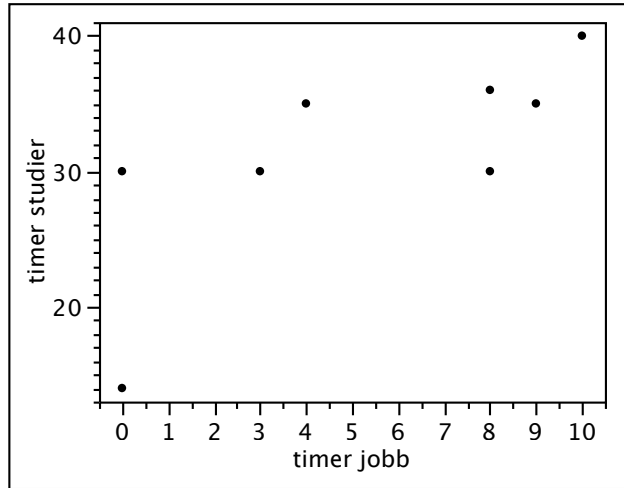
- (a) Hvilken hypotesetest kan du bruke for å avgjøre om det er en sammenheng mellom kjønn og studium i populasjonen av studenter ved BI Trondheim?
- (b) Foreta en test på 1% nivået og konkluder i et lettfattelig språk.



Figur 1: Oppgave 1



Figur 2: Oppgave 5



Figur 3: Oppgave 6

		kjønn		
Count		Mann	Kvinne	
Expected				
Studium	Økonomi	92	69	161
		79,6701	81,3299	
Markedsføring		52	78	130
		64,3299	65,6701	
		144	147	291

Figur 4: Oppgave 7

Sensorveiledning i:	MET 34311 Statistikk
Eksamensdato:	01.06.11, kl. 09.00-14.00
Tillatte hjelpemidler:	Alle + BI-definert eksamenskalkulator : TEXAS INSTRUMENTS BA II Plus TM
Totalt antall sider:	4

Oppgave 1 (6 poeng)

- (a) Nominal.
- (b) 131 rader (en for hver klient) og 6 søyler (en for hver variabel).
- (c) Det er fire stolper for beløp mindre enn 300. Høydene er omtrent 34, 25, 11 og 5. Så det er omtrent $34+25+11+5=75$ klienter.
- (d) I situasjoner der histogrammet har en tung høyrehale vil gjennomsnittet være større enn medianen. Her har vi 6 ekstreme observasjoner (outliers), som har betalt beløp over 675 kroner. Det er disse som drar gjennomsnittet opp, mens medianen ikke påvirkes nevneverdig av disse.
- (e) Siden det er totalt 131 klienter, så må det være $131 - (22 + 3 + 15 + 36 + 12 + 24 + 9) = 10$ kvinner som fikk behandlingsform D.
 $\frac{22+3+15+36}{131} \cdot 100 = 58.02\%$ av klientene var menn.
- (f) Det er 76 menn, så vi får:

$$P(\text{Alle tre er menn}) = \frac{76}{131} \cdot \frac{75}{130} \cdot \frac{74}{129} = 0.192$$

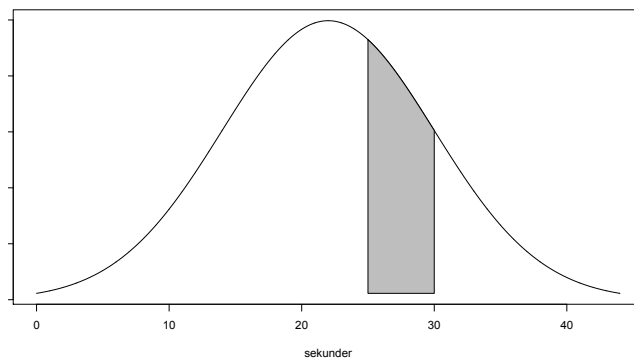
Oppgave 2 (5 poeng)

- (a) Systematisk utvalg.
- (b) $\bar{x} = 20$ og $s = \sqrt{\frac{(11-20)^2 + (15-20)^2 + \dots + (24-20)^2}{8}} = 5.27$
En tolvåring er åtte år yngre enn gjennomsnittet, mens to standardavvik er omtrent 10.6 år. Ergo er ikke 12 år uvanlig.
- (c) Det finnes ulike måter å beregne kvartiler på, men her bruker vi Figur 3-5 i boka: $L_1 = 0.25 * 9 = 2.25$, som ikke er heltall. Vi runder opp til $L = 3$ og finner nedre kvartil $Q_1 = 15$. Øvre kvartil $L_2 = 0.75 * 9 = 6.75$, som ikke er heltall. Vi runder opp til $L = 7$ og finner øvre kvartil $Q_3 = 24$. Interkvartilbredden er $24 - 15 = 9$.

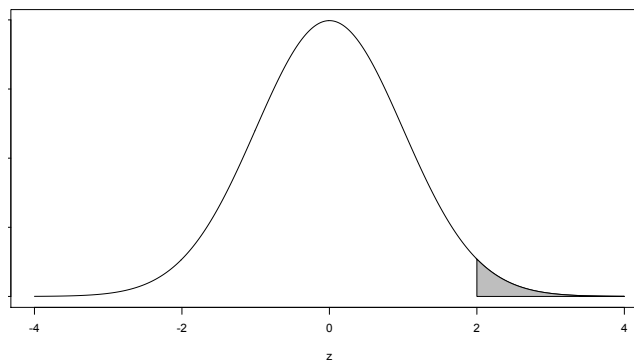
I situasjoner med outliers blir standardavviket svært påvirket av disse, i motsetning til interkvartilbredden. I slike situasjoner er interkvartilbredden å foretrekke som mål for variasjon i dataene.

Oppgave 3 (4 poeng)

(a) $P(25 < X < 30) \simeq P(0.38 < z < 1.00) = 0.8413 - 0.6480 = 0.1933$



(b) Spørsmålet handler om sannsynligheten til at den standardiserte z -verdien er større enn 2. $P(z > 2) = 1 - 0.9772 = 0.0228$. Man kan tegne arealet under tetthetskurven til $N(\mu, \sigma)$ fra $\mu + 2\sigma$ og utover, der μ og σ er studentens valg. Alternativt bruke standard normalfordelingen:



Oppgave 4 (4 poeng)

(a) Feilmarginen $E = 1.676 \cdot \frac{6.86}{\sqrt{52}} = 1.59$. Så vi får konfidensintervallet $8.15 < \mu < 11.33$.

(b) En stikkprøve med størrelse $n = 11$ kan være for liten. Normalt anbefaler man at $n > 30$. Men hvis antall timer på jobb er normalfordelt, så vil intervallet allikevel være gyldig.

(c) Intervallbredden er $11.658 - 5.263 = 6.395$. Feilmarginen er da $E = 6.395/2 = 3.1975$. Gjennomsnittet er midtpunktet i intervallet: $\bar{x} = 5.263 + 3.1975 = 8.461$.

Oppgave 5 (6 poeng)

- (a) Type I feil er når man feilaktig forkaster en nullhypotese som er sann. Sannsynligheten for å forkaste en sann nullhypotese er 0.05.
- (b) $H_0 : p = 0.1$ vs $H_1 : p < 0.1$. Testobservator er

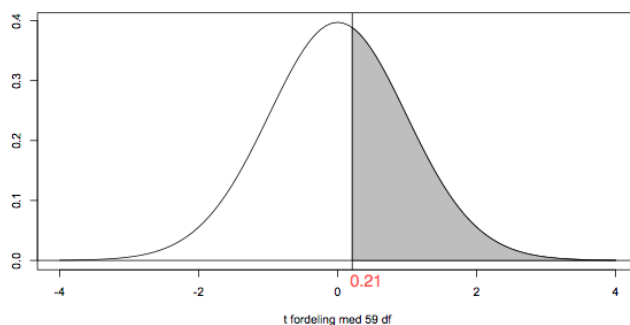
$$z = \frac{\frac{13}{157} - 0.1}{\sqrt{\frac{0.1 \cdot 0.9}{157}}} = -0.72$$

Så p -verdien er 0.2358, og vi forkaster ikke H_0 . Det er ikke tilstrekkelig grunnlag til å hevde at andelen Samsung brukere er mindre enn 10 % ved BI Bergen.

- (c) De to stikkprøvene er uavhengige. $H_0 : \mu_M = \mu_K$ vs. $H_1 : \mu_M > \mu_K$. Testobservator er

$$t = \frac{13.18 - 12.74}{\sqrt{13.88^2/60 + 10^2/99}} \simeq 0.21$$

mens kritisk verdi fra tabell (runder ned fra 59 til 50 frihetsgrader) er $t_{0.1, df=50} \simeq 1.299$. Vi forkaster ikke H_0 .



Oppgave 6 (4 poeng)

- (a) Tre studenter. De jobbet 0, 3 og 8 timer.
- (b) $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$. Testobservatoren er

$$t = \frac{0.73}{\sqrt{\frac{1-0.73^2}{6}}} = 2.62$$

Med 6 frihetsgrader er kritisk verdi $t_{0.025,6} = 2.447$. Vi forkaster H_0 .

- (c) Prisen på en vare og etterspørselen etter varen.

Oppgave 7 (3 poeng)

(a) Kji-kvadrat testen.

(b) Testobservator er

$$\chi^2 = \frac{(92 - 79.6701)^2}{79.6701} + \frac{(69 - 81.3299)^2}{81.3299} + \frac{(52 - 64.3299)^2}{64.3299} + \frac{(78 - 65.6701)^2}{65.6701} = 8.46.$$

Kritisk verdi med 1 frihetsgrad fra tabell er 6.635. Vi forkaster H_0 . Det er grunnlag i dataene til å hevde at det er en sammenheng mellom kjønn og studium.