

# Oppgaver til Studentveiledning 4

## MET 3431 Statistikk

8. mai 2012 kl 17.15 - 20.15 i B2

Handelshøyskolen BI

## **Oppgaver**

### **1. Eksamensoppgaver:**

Eksamen 22/11/2011: Oppgave 1 - 7. Eksamensoppgaven fra 11/2011 er vedlagt.

## **Løsninger**

### **1 Eksamensoppgaver:**

Løsningsforslag til Eksamen 11/2011 er vedlagt.



---

<b>Skriftlig eksamen i:</b>	<b>MET 34311 Statistikk</b>
Eksamensdato:	22.11.11, kl. 14.00-19.00
Tillatte hjelpemidler:	Alle hjelpemidler + eksamenskalkulator TEXAS INSTRUMENTS BA II Plus™
<u>Innføringsark:</u>	Ruter
Totalt antall sider:	4

---

## Oppgave 1

En bokhandel ved et BI studiested foretar en spørreundersøkelse av studentene idet de forlater bokhandelen. Følgende informasjon samles inn:

- Hvor mange minutter man brukte i bokhandelen
  - Hvor mange bøker man kjøpte
  - Hvor tilfreds man er med bokhandelen på en skala fra 1 til 4
  - Hvorvidt man tidligere har handlet pensumlitteratur i en nettbokhandel (ja/nei)
- (a) Angi målenivået (nominal, ordinal, intervall, forholdstall) for hver av de fire variablene.
- (b) Beskriv kort populasjonen i denne undersøkelsen.
- (c) Bokhandelen ønsker en stikkprøve på 45 studenter. Forklar hvordan man kan gå fram for å få en tilfeldig stikkprøve. Gi deretter et eksempel på en utvalgsmetode som gir en bekvemmelighetsstikkprøve, og drøft kort om denne stikkprøven vil gi et feilaktig bilde av populasjonen.
- (d) Stikkprøven gir følgende frekvenstabell for tilfredshet:

Tilfredshet	1	2	3	4
Antall	0	5	25	15

Hva er minimum, maksimum og median for tilfredshet i stikkprøven? Er disse tallene observatorer eller parametre?

- (e) 20% av studentene i stikkprøven har kjøpt pensumlitteratur over nettet. Bokhandelen ønsker å velge ut to av disse for et dybdeintervju. Hvor mange slike utvalg av to studenter som har kjøpt pensumlitteratur over nettet finnes det?

- (f) Figur 1 bakerst viser histogrammet for tid brukt i bokhandelen. Ser tid ut til å være en normalfordelt variabel? Bruk histogrammet til å gi et anslag for gjennomsnittlig tid brukt i bokhandelen, samt for nedre kvartil  $Q_1$ .

## Oppgave 2

Et stort konsulentfirma har kontorer over hele landet. 500 konsulenter ble spurt hvor mange timer i uka de sitter i kundemøter.

- (a) Resultatet for hele stikkprøven vises i Figur 2 bakerst. Er antall timer i kundemøter en normalfordelt variabel?
- (b) En konsulent sier at hun sitter 34 timer i uka i møter med kunder. Regn ut den standardiserte  $z$ -verdien til dette tallet. I lys av stikkprøven, vil du karakterisere 34 timer som 'uvanlig' mye? Forklar.
- (c) I en by oppgir fire konsulenter følgende antall timer i kundemøter: 22, 28, 28 og 30. Regn ut standardavviket for denne stikkprøven. I nabobyen jobber det 5 konsulenter, som har 28, 28, 28, 28 og 28 timer i kundemøter. Forklar hva standardavviket blir for nabobyen uten å beregne noe.
- (d) På østlandet er gjennomsnittet og standardavviket  $\bar{x}_1 = 27$  og  $s_1 = 3$ , mens tilsvarende tall for vestlandet er  $\bar{x}_2 = 28$  og  $s_2 = 1$ . Hva er mest sannsynlig: at en tilfeldig valgt konsulent på østlandet har mer enn 30 møtetimer, eller at en tilfeldig valgt konsulent på vestlandet har mer enn 30 møtetimer?

## Oppgave 3

- (a) Du triller en terning 20 ganger og noterer for hver gang hvilket tall (fra 1 til 6) som terningen viser. Er dette en binomisk forsøksrekke?
- (b) Du triller en terning 20 ganger og noterer for hver gang om du fikk tallet 6 (ja/nei). Er dette en binomisk forsøksrekke?
- (c) Hvor mange 6'ere forventer du å få dersom du triller terningen 20 ganger? Hva er sannsynligheten for å få nøyaktig tre 6'ere dersom du triller terningen 20 ganger?

## Oppgave 4

I en kommune vet man at husholdningenes strømforbruk i juli er normalfordelt med gjennomsnitt  $\mu = 1050$  kWt og med standardavvik  $\sigma = 218$  kWt. Det er totalt 1600 husholdninger i kommunen.

- (a) Regn ut sannsynligheten for at en tilfeldig valgt husholdning bruker mindre enn 800 kWt. Hvor mange husholdninger bruker mindre enn 800 kWt?
- (b) Regn ut hvor mange husholdninger som har et strømforbruk som er lavere enn 30 % persentilen  $P_{30}$ .

- (c) Finn persentilen  $P_{45}$ , dvs. strømforbruket som skiller de laveste 45% fra de høyeste 55%.
- (d) Hva er sannsynligheten for at fire tilfeldig valgte husholdninger i gjennomsnitt bruker mindre enn 1024  $kWh$ ?

## Oppgave 5

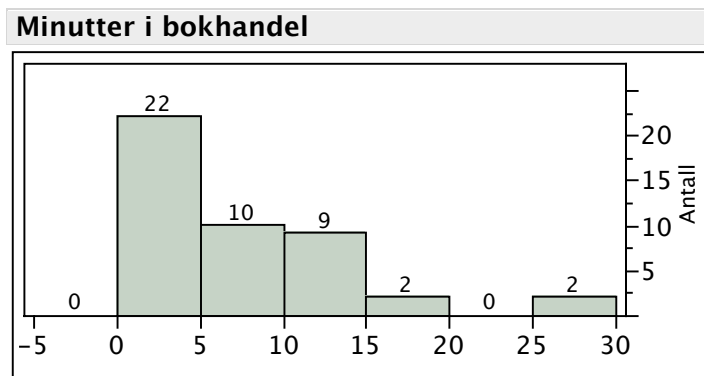
- (a) I en meningsmåling av 865 stemmeberettigede svarer 408 at de ønsker å stemme ved kommunevalget. Lag et 95 % konfidensintervall for andelen velgere som ønsker å stemme.
- (b) Tenk deg at meningsmålingen var utført på en mindre gruppe, for eksempel med bare 100 stemmeberettigede. Hvis du lagde et 95 % konfidensintervall basert på denne mindre gruppen, ville du forvente et bredere eller smalere intervall enn i oppgave (a)?
- (c) Du ønsker å teste om mindre enn halvparten av alle stemmeberettigede i kommunen ønsker å stemme. Skriv opp nullhypotesen og alternativhypotesen. Bruk informasjonen i oppgave (a) til å utføre testen og konkluder i et lettfattelig språk. Bruk signifikansnivå  $\alpha = 0.05$ .

## Oppgave 6

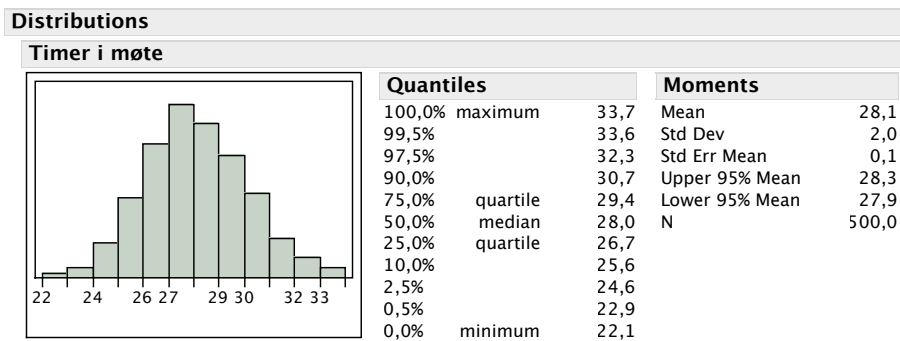
- (a) Tretti tilfeldig valgte studenter i markedsføringsklassen tar en statistikktest. Gjennomsnittresultatet ble 95 poeng med standardavvik 9.6 poeng. Lag et 99 % konfidensintervall for gjennomsnittresultatet for alle markedsføringsstudenter.
- (b) Du ønsker å teste hypotesen om at gjennomsnittresultatet for alle markedsføringsstudenter er over 90 poeng. Skriv opp hypotesene og utfør testen på signifikansnivå  $\alpha = 0.1$ . Skisser testobservator og p-verdien i en graf.
- (c) Samme statistikk test tas også av 40 tilfeldig valgte studenter i økonomiklassen. Gjennomsnittresultatet ble 98 poeng med standardavvik 6.6 poeng. Test påstanden om at økonomistudentene har høyere gjennomsnittresultat enn markedsføringsstudentene. Bruk signifikansnivå  $\alpha = 0.05$ .

## Oppgave 7

I en stikkprøve av 17 studenter var korrelasjonen mellom tid brukt på studier og reisetid mellom studiestedet og hjem lik  $\hat{\rho} = -0.01$ . Hvordan vil du tolke denne korrelasjonskoeffisienten? Tror du den er signifikant? Drøft kort uten å gjøre noen beregninger.



Figur 1: Oppgave 1



Figur 2: Oppgave 2

**Sensorveiledning i: MET 34311 Statistikk**

Eksamensdato: 22.11.11, kl. 14.00-19.00

Tillatte hjelpemidler: Alle hjelpemidler + eksamenskalkulator  
TEXAS INSTRUMENTS BA II Plus™

Totalt antall sider: 3

### Oppgave 1 (7 poeng)

- (a) Minutter i bokhandelen: Forholdstall. Bøker kjøpt: Forholdstall. Tilfredshet: Ordinal. Handlet pensum på nettet: Nominal.
- (b) Populasjonen består av alle studenter ved studiestedet som besøker bokhandelen.
- (c) For å få en tilfeldig stikkprøve bør man velge ut flere dager og tider på døgnet. Deretter velges studentene tilfeldig når de kommer ut av butikken, f.eks. ved å kaste kron/mynt. Bekvemmelighet: Man spør for eksempel 45 studenter på mandag morgen, rett etter en forelesning. Da får man uforholdsmessig mange studenter som går i klassene som nettopp har hatt forelesning. Det gir et skjevt bilde av populasjonen.
- (d) Minimum: 2, Maksimum: 4, Median: 3. Disse tallene er observatorer - de er beregnet fra en stikkprøve.
- (e) 20 % av 45 er 9. Vi skal velge ut 2 studenter fra 9. Antall slike utvalg er  $\binom{9}{2} = \frac{9 \cdot 8}{2!} = 36$ .
- (f) Ikke normalfordelt, fordi den ikke er symmetrisk men høyreskjev. Gjennomsnittlig tid brukt kan anslås ved å bruke midtpunktene:

$$\bar{x} \approx \frac{22 \cdot 2.5 + 10 \cdot 7.5 + 9 \cdot 12.5 + 2 \cdot 17.5 + 2 \cdot 27.5}{45} \approx 7.4$$

Nedre kvartil  $Q_1$  befinner seg i posisjon  $L = 45 \cdot 0.25 = 11.25$  som vi runder opp til  $L = 12$ . De 22 første observasjonene ligger mellom 0 og 5 minutter. Vi ekstrapolerer for å anslå verdien til observasjon nr  $L = 12$ :  $Q_1 \approx \frac{12}{22} \cdot 5 \approx 3$ .

### Oppgave 2 (6 poeng)

- (a) Ser ut til å være normalfordelt, siden grafen er symmetrisk med en topp i midten.
- (b) Standardisert verdi til 34 timer er  $z = \frac{34 - 28.1}{2.0} = 2.95$ . Altså mer enn to standardavvik over gjennomsnittet. Dette regnes som 'uvanlig'.

- (c) Standardavviket er

$$s = \sqrt{\frac{(22 - 27)^2 + (28 - 27)^2 + (28 - 27)^2 + (30 - 27)^2}{4 - 1}} = \sqrt{\frac{25 + 1 + 1 + 9}{3}} \approx 3.5$$

For nabobyen har alle konsulentene samme tid i møter. Det er ingen variasjon. Derfor vil  $s = 0$  i nabobyen.

- (d) På østlandet er z-verdien til 30 timer lik  $z_\phi = \frac{30-27}{3} = 1$ . På vestlandet er z-verdien til 30 lik  $z_v = \frac{30-28}{1} = 2$ . Mao, på østlandet tilsvarer tretti timer ett standardavvik over gjennomsnittet, mens på vestlandet tilsvarer det to standardavvik over gjennomsnittet. Det er derfor mest sannsynlig at en tilfeldig valgt konsulent på østlandet har mer enn tretti timer.

### Oppgave 3 (4 poeng)

- (a) Utfallet i hvert forsøk kan ha 6 verdier. I en binomisk forsøksrekke er utfallet i hvert forsøk binomisk, dvs. ja/nei eller suksess/ikke-suksess. Derfor er ikke denne forsøksrekken binomisk.
- (b) Denne forsøksrekken er binomisk. Det er identiske uavhengige forsøk med binomisk (ja/nei) utfall i hvert forsøk.
- (c) Vil forvente  $20 \cdot \frac{1}{6} \approx 3$  seksere. Sannsynligheten for nøyaktig tre seksere gis av binomialfordelingen:

$$P(X = 3) = \binom{20}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{1}{6}\right)^{17} \approx 0.24$$

### Oppgave 4 (7 poeng)

- (a) Vi standardiserer først 800:  $z = \frac{800-1050}{218} = -1.15$ . Da får vi fra tabellen for standard normalfordeling at

$$P(X < 800) = P(z < -1.15) = 0.1251.$$

Det tilsvarer  $0.1251 \cdot 1600 = 200$  husstander.

- (b) Per definisjon må 30 % av husstandene ha et forbruk som er lavere enn  $P_{30}$ . Det tilsvarer  $0.3 \cdot 1600 = 480$  husstander.
- (c) Ved å lete i tabellen for standard normalfordeling ser vi at  $P(z < -0.12) \approx 0.45$ . Den tilsvarende persentil verdien for strømforbruk er da  $P_{45} = 1050 - 0.12 \cdot 218 = 1024$ .
- (d) Gjennomsnittet  $\bar{x}$  av fire tilfeldige verdier er normalfordelt med  $\mu_{\bar{x}} = 1050$  og standardavvik  $\sigma_{\bar{x}} = \frac{218}{\sqrt{4}} = 109$ . Dermed har vi

$$P(\bar{x} < 1024) = P\left(z < \frac{1024 - 1050}{109}\right) = P(z < -0.24) = 0.4052$$



## Oppgave 5 (5 poeng)

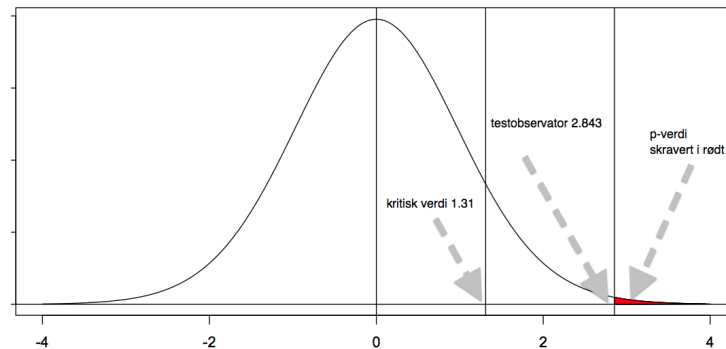
- (a) Feilmarginen  $E = 1.96 \sqrt{\frac{\frac{408}{865}(1-\frac{408}{865})}{865}} = 0.044$  så konfidensintervallet blir  $0.438 < p < 0.505$ .
- (b) Med en mindre stikkprøve forventes et bredere intervall. Vi har mindre informasjon, så konfidensintervallet blir mindre presist, altså bredere.
- (c)  $H_0 : p = 0.5$  og  $H_1 : p < 0.5$ . Testobservator

$$z = \frac{\frac{408}{865} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{865}}} = -1.67$$

Kritisk verdi for ensidig test er  $z = -1.645$ . Vi forkaster  $H_0$ . Det er tilstrekkelig støtte i dataene til å hevde at mindre enn halvparten av de stemmeberettigede ønsker å stemme.

## Oppgave 6 (6 poeng)

- (a) Feilmargin  $E = t \cdot s / \sqrt{n} = 2.756 \cdot 9.6 / \sqrt{30} = 4.83$ . Konfidensintervallet blir da  $95 \pm 4.83$ .
- (b)  $H_0 : \mu = 90$  og  $H_1 : \mu > 90$ . Testobservatoren er  $t = \frac{95-90}{9.6/\sqrt{30}} = 2.853$  Kritisk verdi fra tabell er  $t_{\alpha=0.1} = 1.311$ . Vi forkaster  $H_0$ .



- (c)  $H_0 : \mu_1 = \mu_2$  vs.  $H_1 : \mu_1 < \mu_2$ . Testobservator er

$$t = \frac{98 - 95}{\sqrt{\frac{6.6^2}{40} + \frac{9.6^2}{30}}} = 1.471$$

kritisk verdi er  $t_{29,0.05} = 1.697$ . Vi forkaster ikke  $H_0$ .

## Oppgave 7 (2 poeng)

Korrelasjonskoeffisienten er nesten lik 0, noe som tyder på at vi ikke har signifikant korrelasjon. For at  $\hat{\rho} = -0.01$  skal være signifikant, så må stikkprøven være svært stor. Men her stikkprøven liten ( $n = 17$ ).