

FORELESNING 22

MET 3431

EIVIND ERIKSEN

APR 17 2012

STATISTIKK

PLAN:

- ① Oppgaver Kji-kvadrat test
- ② Repetisjon: Gjennomgang av Eksamen 06/2010

① Eksamen 06/2011, Oppgave 7

a) χ^2 -kvadrat test siden det er sammenhengen mellom to kategoriske variable

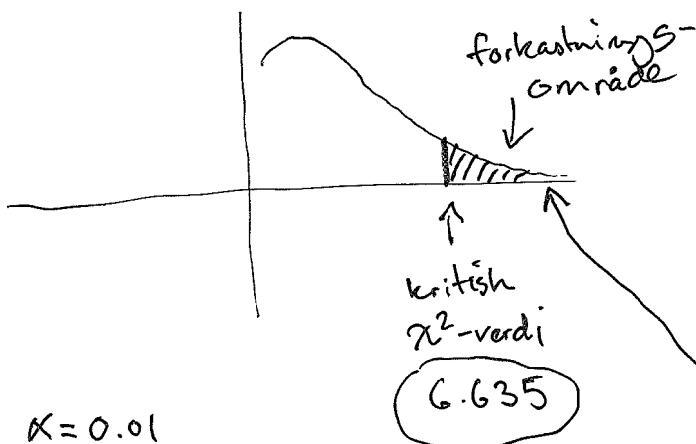
b) Utfører χ^2 -test:

H_0 : ingen sammenheng

H_1 : det er en sammenheng

} testobservator
 χ^2

c) Forkastningsområde:



$$\alpha = 0.01$$

$$df = (2-1) \times (2-1) = 1$$

Tabell A4
(χ^2 verdier)

(b) Testobservator:

$$\begin{aligned}\chi^2 &= \sum_{i=1}^4 \frac{(O_i - \bar{E}_i)^2}{E_i} \\ &= \frac{(92 - 79.67)^2}{79.67} + \frac{(69 - 81.33)^2}{81.33} \\ &\quad + \frac{(52 - 64.33)^2}{64.33} + \frac{(78 - 65.67)^2}{65.67} \\ &\approx \underline{\underline{8.46}}\end{aligned}$$

Konklusjon: Vi forkaster H_0 .

På grunnlag av de gitte dataen, har vi grunn
tel å konkludere at det er en sammenheng på 1%
nivået.

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.

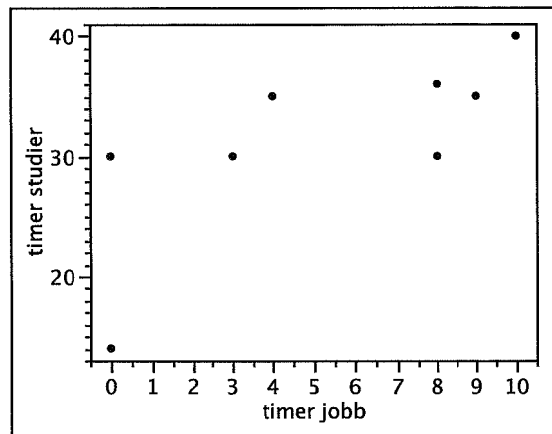
Oppgave 7

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.

$$\alpha = 0.01$$

K²i-kvadrat
test



Figur 3: Oppgave 6

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

Observerte (with arrow pointing to the circled observed counts 92, 69, 52, and 78)

Figur 4: Oppgave 7

Met 3431 Statistikk: Pensum og forslag til oppgaver.

Lærebok: *Essentials of Statistics*, 4th edition, av Triola.

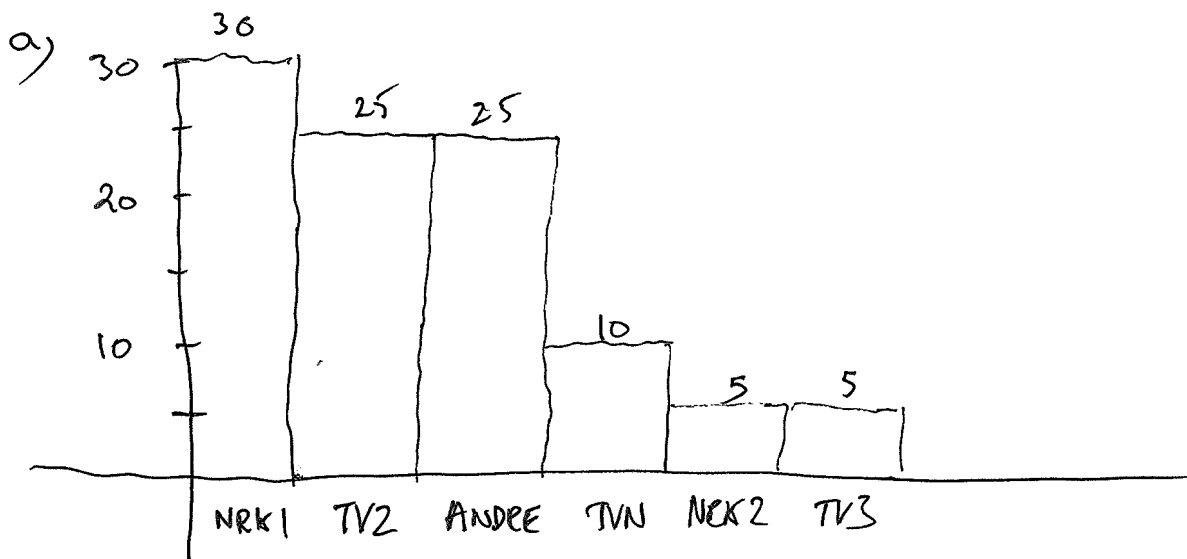
Kapittel	Tema/Nøkkelord	Oppgaver
1.1 Review and preview	Populasjon. Utvalg/stikkprøve.	
1.2 Statistical thinking	Frivillig respons stikkprøve. Utvalgsmetode	1, 3, 5,
1.3 Types of data	Parameter. Observator. Diskrete og kontinuerlige data. Målingsnivå: nominal, ordinal, intervall og ratio.	5, 7, 9, 13, 15, 23, 25, 27
1.4 Critical thinking	Selvseleksjon. Effekter av små og skjeve utvalg. Prosentregning.	9, 15, 21, 23, 27
1.5 Collecting Sample Data	Observasjonelle og eksperimentelle studier. Confounding. Typer av stikkprøver (tilfeldig, enkel tilfeldig, systematisk, osv.)	9, 11, 13, 15
		Chapter quick quiz.
2.1 Review and Preview	Senter, variasjon, fordeling og outlier.	
2.2 Frequency distributions	Frekvenstabell. Relative frekvenser. Normalfordelingen.	3, 5, 9, 29
2.3 Histograms	Histogram.	1,5, 7
2.4 Statistical graphics	Kakediagram. Paretdiagram. Koordinatsystem. Scatterplott.	13
2.5 Bad Graphs	Misvisende grafer	3, 7, 9 Review excs: 5. Cum rev excs: 1,2,3,4
3.1 Review and Preview		
3.2 Measures of Center	Gjennomsnitt. Median.	3, 5, 29, 33
3.3 Measures of Variation	Varians. Standardavvik.	1, 3, 15, 31
3.4 Measures of Relative Standing and Boxplots	z-verdier. persentiler, kvartiler. boksplott. outlier.	1,3, 5, 11, 13, 37
		Review exc: 10.
4.1 Review and Preview	Rare event rule	
4.2 Basic Concepts of Probability	Hendelser og utfallsrom. Sannsynlighet.	1, 3, 15, 31
5.1 Review and Preview		
5.2 Random variables	Tilfeldig variabel, diskret og kontinuerlig. Sannsynlighetsfordeling. Raritetsregelen og range rule of thumb.	1, 5
5.3 Binomial probability distr	Binomialfordelingen.	1,3, 29
6.1 Review and Preview	Normalfordelingen	
6.2 The standard normal distribution	Uniform fordeling. Tetthetskurve. Standard normalfordeling. Finne sannsynligheter gitt z verdier.	3, 9, 11, 37, 39
6.3 Applications of normal distribution	Standardisere til z-verdi.	1, 5, 7, 27, 31
6.4 Sampling distributions and estimators	Fordelingen til en observator, f.eks. andel og forventning. Variansen til en observator.	1, 5, 7
6.5 The central limit theorem	Sentralgrenseteoremet. Standardfeil til gjennomsnittet.	1, 3, 5, 7,19
7.1 Review and Preview		
7.2 Estimating a population proportion	Punktestimering og konfidensintervall for andelen. Feilmargin. Konfidensnivå.	1, 3, 5, 7, 35, 37
7.4 Estimating a population mean: σ not known	Konfidensintervall for gjennomsnittet. Students <i>t</i> -fordeling.	23, 27, 29

② Eksamen 06/2010

Oppgave 1:

- a) 1) Pop: Alle kunder
Stikkp: 3
Hvordan repr: Lite utvalg, ikke tilfeldig (bekjent)
- 2) Pop: Alle voksne nordmenn
Stikkp: 1000
Repr: Mange nok, tilfeldig utvalg
- b) Feil: Frivillig respons (ikke representativt)
- c) Ordinalnivå ^{co} (kategorisk, _{ordning} mellom alternativer)
(rang-)

Oppgave 2:



b) Ser ikke normalfordelt ut (skjev)

Ca 58

$$\frac{58}{151} \cdot 100\% \approx \underline{38\%}$$

Oppgave 3:

$$a) \bar{x}_A = \frac{13+21+12+34+31+13+22+22}{8} = 21$$

$$b) s_B = \sqrt{\frac{(20 - \bar{x}_B)^2 + (18 - \bar{x}_B)^2 + (14 - \bar{x}_B)^2 + (19 - \bar{x}_B)^2}{3}} \quad \left. \begin{array}{l} \bar{x}_B = \frac{20+18+14+19}{4} \\ = 17.75 \end{array} \right\}$$

3
↑
n-1

$$= \sqrt{\frac{(20-17.75)^2 + (18-17.75)^2 + (14-17.75)^2 + (19-17.75)^2}{3}}$$

$$\approx \underline{\underline{2.63}}$$

Median: 20, 18, 14, 19

Volumer: 14, 18, 19, 20

median = 18.5

c) For B:

$$\bar{x} = \bar{x}_B = 17.75$$

$$\tilde{x} = \text{median} = 18.5$$

$$s = s_B \approx 2.63$$

$$I = \frac{3 \cdot (\bar{x} - \text{median})}{s} = \frac{3 \cdot (17.75 - 18.5)}{2.63}$$

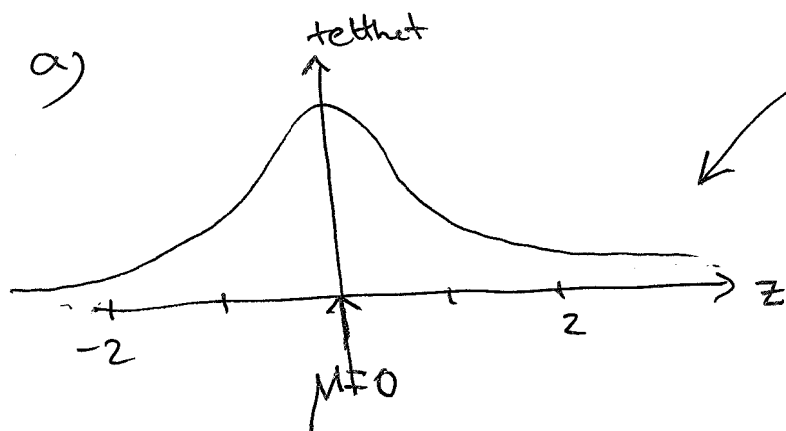
$$\approx \underline{\underline{-0.86}}$$

$$d) \quad z = \frac{x - \mu}{\sigma} = \frac{38 - 20.5}{8.0} = \frac{17.5}{8.0} \approx \underline{\underline{2.2}}$$

Mer enn 2 std. avvik unna gjennomsnittet

\Downarrow
uventlig

Oppgave 4:



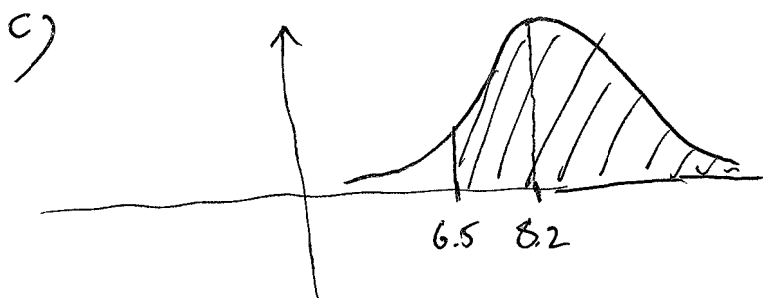
graf til sannsynlighetstettheten til z

$\mu = 0, \sigma = 1$

b) Antall utfall: $6 \times 6 = \underline{36}$

Sum = 3: $1+2$ eller $2+1 \Rightarrow \underline{2}$ utfall

$p(\text{sum} = 3) = \frac{2}{36} \approx 5.56\% \Rightarrow \underline{\underline{\text{ikke uventlig}}}$



z-verdi: $\frac{6.5 - 8.2}{1.1} = \frac{-1.7}{1.1} \approx \underline{\underline{-1.545}}$

$$\begin{aligned}
 & p(X \geq 6.5) \\
 &= p(Z \geq -1.545) \\
 &= 1 - p(Z \leq -1.545) \\
 &= 1 - 0.0606 \\
 &= 0.9394 \approx \underline{\underline{94\%}}
 \end{aligned}$$

Skriftlig eksamen i: MET 34311 Statistikk

Eksamensdato: 02.06.10, kl. 09.00-14.00

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

Innføringsark: Ruter

Totalt antall sider: 4

Oppgave 1

- (a) I hver av de følgende to situasjoner, hva er stikkprøven og hva er populasjonen? Angi også om stikkprøven er representativ for populasjonen.
- En baker spør sine tre første kunder om de foretrekker rugbrød framfor loff.
 - 1000 tilfeldig valgte voksne nordmenn ble spurt om de drikker mer enn to liter vann til dagen.
- (b) Et spørreskjema om holdning til bompenger sendes til 10000 tilfeldig valgte personer. 3000 svarer på skjemaet, og 1000 av dem sier at de er imot bompenger. På bakgrunn av spørreundersøkelsen konkluderer man med at en av tre personer er imot bompenger. Hva er feil med konklusjonen?
- (c) En bedrift rangerer produktkvaliteten som 1=dårlig, 2=middele og 3=bra. Hvilket målenivå måles produktkvaliteten på?

Oppgave 2

- (a) En uke i 2009 var markedsandelene i prosent til de forskjellige TV-kanalene som følger:

NRK1	NRK2	TV2	TVN	TV3	Andre
30	5	25	10	5	25

Lag et Pareto diagram for markedsandelene.

- (b) JMP utskriften i Figur 1 bakerst gjelder antall studietimer i uka for en stikkprøve studenter ved BI Stavanger. Ser antall studietimer ut til å være normalfordelt? Ca hvor mange av studentene i stikkprøven studerer mellom fem og ti timer i uka? Hvor mange prosent utgjør dette av hele stikkprøven?

pop:
Alle voksne
nordmenn
stikkprøve:
1000
ja:
stor nok
tilfeldig valgt

populasjon:
Alle kunder
stikkprøve:
tre kunder
Nei:
- liten stikkprøve
- ikke tilfeldig
utvalg

Oppgave 3

I et parkeringsfirma jobber parkeringsvaktene i to avdelinger. Det ble tatt tilfeldige stikkprøver av åtte vakter i avd. A og av fire vakter i avd. B. Vaktene i stikkprøve A skrev ut 13, 21, 12, 34, 31, 13, 22 og 22 bøter i løpet av formiddagen. For vaktene i stikkprøve B var antallet 20, 18, 14 og 19 bøter.

- (a) Beregn gjennomsnittet for antall bøter i avdeling A.
- (b) Beregn standardavviket s for antall bøter i avdeling B.
- (c) I en avansert statistikkbok står følgende formel for skjevhet:

$$I = \frac{3(\bar{x} - \text{median})}{s}$$

Bruk formelen til å regne ut I for stikkprøve B.

- (d) Gjennomsnittet og standardavviket for hele avdeling A er $\mu = 20.5$ og $\sigma = 8.0$. En av vaktene i avdeling A som ikke var med i stikkprøven skrev ut 38 bøter. Beregn den standardiserte z -verdien til dette antallet og kommenter hvorvidt denne vaktens skrev ut 'uvanlig' mange bøter i forhold til resten av avdelingen.

Oppgave 4

- (a) Skisser grafen/tetthetskurven til en standard normalfordelt variabel z . Hva er verdien til populasjonsparametrene μ og σ for variabelen z ?
- (b) Vi definerer en hendelse som 'uvanlig' dersom sannsynligheten for at den inntreffer er mindre enn 0.05. Du triller to terninger. Hvor mange utfall er det totalt? Og i hvor mange av disse utfallene er summen tre? Er det uvanlig at summen av terningene blir tre?
- (c) Levetiden på nye TV apparat er normalfordelt med gjennomsnitt $\mu = 8.2$ år og standardavvik $\sigma = 1.1$ år. Hva er sannsynligheten for at et tilfeldig valgt TV apparat varer mer enn 6.5 år? Skisser sannsynligheten som arealet under en graf.

Oppgave 5

I en stikkprøve våren 2010 oppgir 183 av 878 markedsføringsstudenter ved BI at de har en iPhone mobiltelefon. Se JMP utskrift i Figur 2 bakerst.

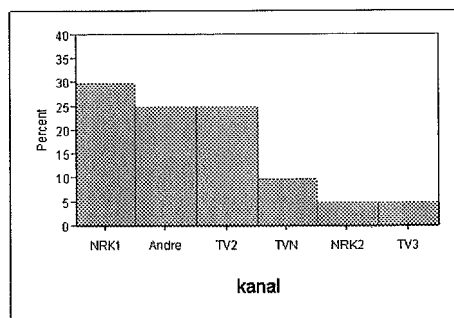
- (a) Lag et 95 % konfidensintervall for andelen studenter på markedsføring som har en iPhone.
- (b) Gjør en hypotesetest for å avgjøre om andelen Nokia brukere på markedsføringsstudiet er mer enn 25 %. Bruk signifikansnivå $\alpha = 0.05$. Skisser p -verdien i en graf.

Sensorveiledning i:	MET 34311 Statistikk
Eksamensdato:	02.06.2010, kl 09.00-14.00
Tillatte hjelpemidler:	Alle hjelpemidler + eksamenskalkulator TEXAS INSTRUMENTS BA II Plus
Totalt antall sider:	3

Oppgave 1 (4 poeng)

- (a) 1. Populasjon: Alle kunder. Stikkprøve: Tre kunder. Ikke representativ: for liten stikkprøve og heller ikke tilfeldig utvalg.
2. Populasjon: Alle voksne nordmenn. Stikkprøve: 1000 voksne nordmenn. Representativ.
- (b) Frivillig respons er problemet. Gruppen som svarer på spørreskjema er ikke lik gruppen som ikke svarer på spørreskjema. Dermed er stikkprøven ikke representativ for populasjonen.
- (c) Ordinalnivå, siden det er snakk om tre kvalitetsnivå med en naturlig ordning.

Oppgave 2 (4 poeng)



- (a)
- (b) Ikke normalfordelt, men usymmetrisk med lang høyrehale. Det er ca 58 av totalt 151 studenter. I prosent: $\frac{58}{151} \cdot 100 \simeq 38\%$.

Oppgave 3 (6 poeng)

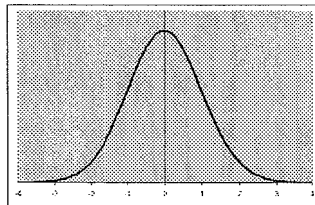
- (a) $\bar{x}_A = 21$

$$(b) s_B = \sqrt{\frac{(20-17.75)^2+(18-17.75)^2+(14-17.75)^2+(19-17.75)^2}{3}} = 2.63$$

$$(c) I_B = \frac{3(17.75-18.5)}{2.63} = -0.86$$

(d) $z = (38 - 20.5)/8.0 = 2.19$. Siden 38 er mer enn to standardavvik over gjennomsnittet 20.5, kan vi kalle det 'uvanlig'.

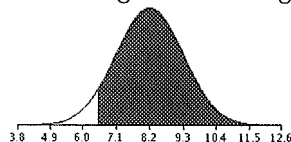
Oppgave 4 (5 poeng)



(a) $\mu = 0$ og $\sigma = 1$.

(b) Totalt antall utfall er $6 \cdot 6 = 36$. Bare $(1, 2)$ og $(2, 1)$ gir sum tre. Da er sannsynligheten for summen tre lik $\frac{2}{36} \approx 0.056$. Ergo er det ikke 'uvanlig' å få summen tre.

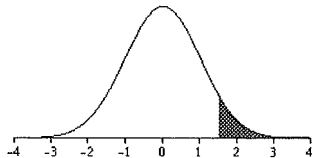
(c) $z = (6.5 - 8.2)/1.1 = -1.55$. Tabell A2 gir da sannsynligheten $1 - 0.0606 = 0.9394$. Denne sannsynligheten kan skisseres som arealet under den originale x -fordelingen slik som angitt her, eller som arealet under z -fordelingen.



Oppgave 5 (6 poeng)

(a) $ME = 1.96 \cdot \sqrt{\frac{0.208(1-0.208)}{878}} = 0.027$ som gir konfidensintervallet $0.182 < p < 0.235$

(b) $H_0 : p = 0.25$ vs $H_1 : p > 0.25$. Testobservator $z = \frac{0.2722-0.25}{\sqrt{0.25 \cdot 0.75/878}} = 1.52$. Tabell A2 gir p -verdi 0.0643, som er større enn α . Vi kan ikke forkaste H_0 .



(c) 90 % konfidensintervallet er kortest. Et 99 % konfidensintervall inneholder populasjonsparameteren med større sikkerhet enn et 90% intervall. Da må det være lengre enn et 90% konfidensintervall.