FORELESNING 23

MET 3431

EIVIND ERIKSEN

APR 19 2012

STATISTIKK

PLAN:

1. Eksamen 06/2010 Oppg. 5-7
2. ~~Oppsummering~~
3. ~~Eksamen 11/2010~~

Eksamen 06/2010

Oppg. 5:

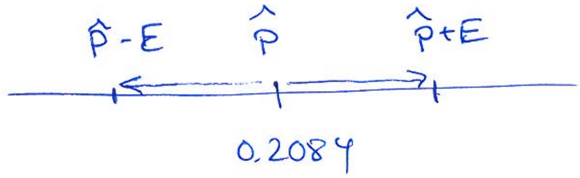
a) Andelen med iPhone : p

$$\hat{p} = \frac{183}{878} \simeq 0.2084$$

Konfidensintervall: $(\hat{p} - E, \hat{p} + E)$

$$1.96 = z_{\alpha/2}$$

$$\sqrt{\frac{0.2084 \cdot 0.7916}{878}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



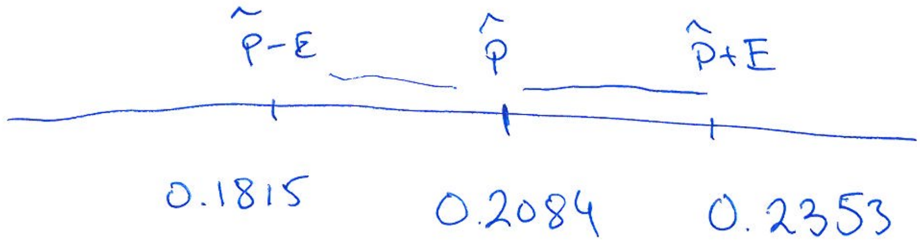
$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \cdot 0.0137 \simeq 0.0269$$

95% konfidensintervall:

$$\alpha = 0.05 \implies z_{\alpha/2} = 1.96$$

Konklusjon: $\hat{p} = \frac{\text{antall}}{n}$ 95% $\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$

$\hat{P} = 0.2084$ $E = 1.96 \cdot 0.0137 = 0.0269$



Konfidensintervall:

$$\underset{=\hat{P}}{0.2084} \pm \underset{=E}{0.0269}$$

eller

$$(0.1815,\ 0.2353)$$

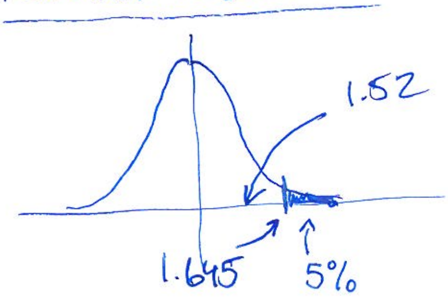
b) Hypotese-test: p : andelen Nokia-bruker

H_0 : $p = 0.25$

H_1 : $p > 0.25$

Test-observator: Z

i) Forkastningsområde:



ii) Regn ut test-obs. Z:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p\cdot(1-p)}{n}}} = \frac{0.2722 - 0.25}{\sqrt{\frac{0.25 \cdot 0.75}{878}}}$$

$$= \frac{0.0222}{0.0146}$$

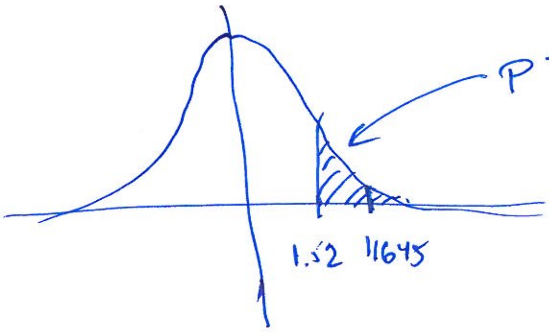
$$\simeq \underline{1.52}$$

$Z <$ kritisk z-verdi

$\parallel$ $\qquad\qquad$ $\parallel$

1.52 $\qquad\qquad$ 1.645

Vi kan <u>ikke forkaste H_0</u> .



p-verdi = areal (fra 1.52 og opp)

$> 5\%$

c) 90%

99%

99% gir lengst konfidensintervall

Større konfidensnivå svarer til at vi er sikrer, altså må vi ta med flere mulige verdier

d) <u>Konfidensintervall:</u>

Et intervall av mulige verdier for en parameter, basert på en stikkprøve, slik at det er p% sikkert at den riktige verdien havner innenfor intervallet (p = konfidensnivå) .

Oppgave 6

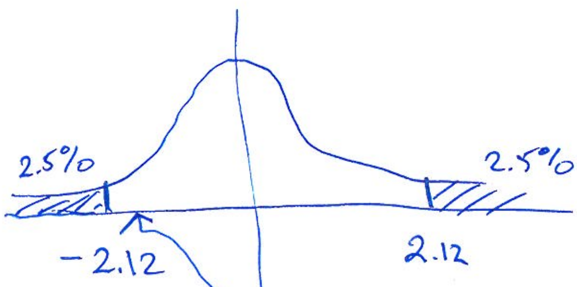
a) μ_1 : gjennomsnittslønn i by I

$H_0: \mu_1 = 30.000$

$H_1: \mu_1 \neq 30.000$

testobservator: t $(df = 16 = n_1 - 1)$

i) Forkastningsområde:



ii) Regn ut t:

$$t = \frac{23198 - 30.000}{14000/\sqrt{17}} = \frac{-6802}{3395} = -2.00$$

$$t = \frac{\bar{X}_1 - \mu_1}{s_1/\sqrt{n_1}}$$

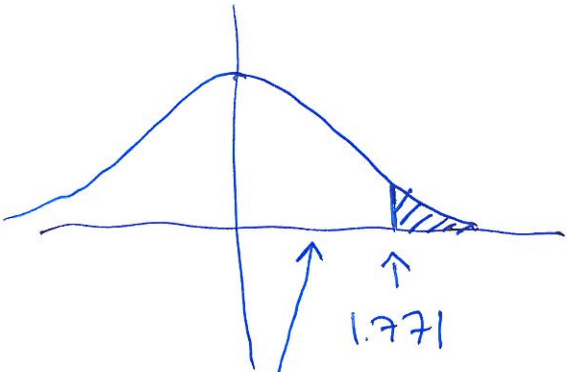
Vi har ikke grunnlag for å forkaste H_0.

$\Rightarrow$ Det er grunnlag for å hevde at $\mu_1 = 30.000$ med signifikansnivå $\alpha = 0.05$.

b) Vanskelig.

$$H_0: \quad \mu_2 = \mu_1$$
$$H_1: \quad \mu_2 > \mu_1$$
$\Big\}$ t-fordeling

i) <u>Forkastningsområde:</u>



1.771

($\alpha = 0.05$, ensidig,
df $= 13$)

ii) <u>Regn ut t:</u>

$$t = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}}$$

$$= \frac{26899 - 23198}{\sqrt{\frac{14000^2}{17} + \frac{13000^2}{14}}}$$

$$= \frac{3701}{4858}$$

$$\simeq 0.76$$

<u>Ikke</u> grunnlag til å <u>forkaste H0</u>

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:

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

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 vakten 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.

(c) I denne oppgaven skal du ikke beregne noe. Men tenk deg at du lager et 99 % konfidensintervall og et 90 % konfidensintervall basert på den samme stikkprøven. Hvilket intervall er lengst? Begrunn svaret kort.

(d) Forklar kort med egne ord hva et konfidensintervall er for noe, og hvordan man skal tolke det.

Oppgave 6

Anta at månedslønn i by 1 er normalfordelt. Vi tar en stikkprøve med $n_1 = 17$ personer med gjennomsnittslønn $\overline{x}_1 = 23198$ kroner og standardavvik $s_1 = 14000$ kroner.

(a) Test påstanden om at i by 1 er gjennomsnittslønnen lik 30000 kroner. Bruk signifikansnivå $\alpha = 0.05$. Skriv opp hypotesene på symbolsk form, skisser kritisk verdi og testobservator i en graf, og formuler konklusjonen i et lettfattelig språk.

(b) Også i by 2 er månedslønnen normalfordelt. Der tar vi en stikkprøve med $n_2 = 14$ personer med gjennomsnittslønn $\overline{x}_2 = 26899$ kroner og standardavvik $s_2 = 13000$ kroner. Er stikkprøvene fra by 1 og 2 relaterte eller uavhengige? Test påstanden om at gjennomsnittslønnen i by 2 er høyere enn i by 1. Bruk signifikansnivå $\alpha = 0.05$. Skriv opp hypotesene på symbolsk form og formuler konklusjonen i et lettfattelig språk.

Oppgave 7

Ernst & Young og SR-Bank gjorde i 2009 en undersøkelse om norske oljeleverandører. De fant følgende tall for antall ansatte (i tusen) og omsetning (i milliarder kroner) for ulike regioner:

	Rogaland	Møre	Hordaland	Agder	Kongsberg
Ansatte	31	10	12	5	6
Omsetning	90	45	32	36	18

(a) Lag et spredningsdiagram (scatterplott) for dataene med ansatte på x-aksen og omsetning på y-aksen. Ser det ut til å være en lineær korrelasjon mellom antall ansatte og omsetningen og er den evt. positiv eller negativ?

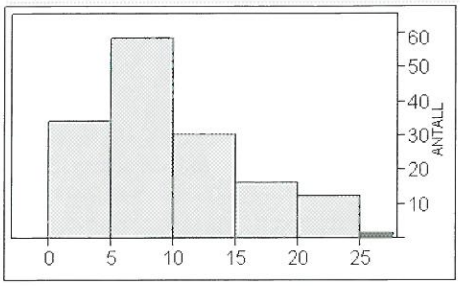
(b) Korrelasjonskoeffisienten er $r = 0.935$. Beregn r^2 og gi en tolkning av dette tallet.

(c) Du ønsker å bruke kji-kvadrattesten for å teste om det er en sammenheng mellom kjønn og holdning (for/mot) til bompenger. Dataene er oppsummert i følgende krysstabell:

	Mann	Kvinne	
For	3	13	16
Mot	2	15	17
	5	28	33

Hvilken viktig forutsetning for å kunne bruke kji-kvadrattesten er ikke tilstede?

timer studier

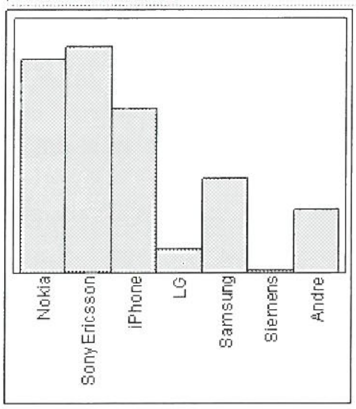


Moments

Mean	8.4403974
Std Dev	5.5226585
Std Err Mean	0.4494276
Upper 95% Mean	9.3284237
Lower 95% Mean	7.552371
N	151

Figur 1: Studietimer i uka ved BI Stavanger.

mobiltelefon



Frequencies

Level	Count	Prob
Nokia	239	0.27221
Sony Ericsson	254	0.28929
iPhone	183	0.20843
LG	25	0.02847
Samsung	106	0.12073
Siemens	1	0.00114
Andre	70	0.07973
Total	878	1.00000

N Missing 2

7 Levels

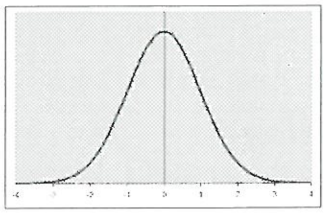
Figur 2: Type mobiltelefon på markedsføringsstudiet

(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'.

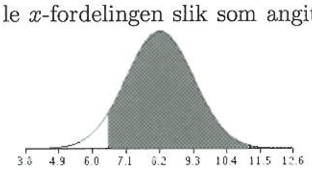
Oppgave 4 (5 poeng)



(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 $\frac{2}{36} \simeq 0.056$. Ergo er det ikke 'uvanlig' å få summen tre.

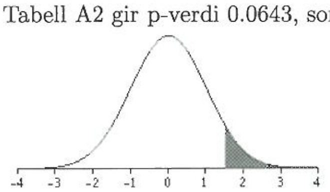
(c) $z = (6.5-8.2)/1.1 = -1.55$. Tabell A2 gir da sannsynligheten $1-0.0606 = 0.9394$. Denne sannsynligheten kan skisseres som arealet under den opprinnelige x-fordelingen slik som angitt her, eller som arealet under z-fordelingen.



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.

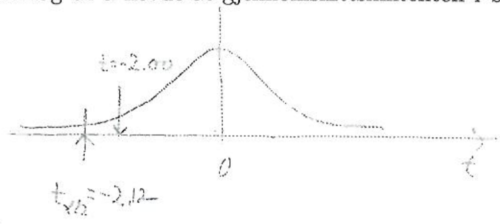


(c) 90 % konfidensintervallet er kortest. Et 99 % konfidensintervall inneholder populasjonsparameteren med større sikkerhet enn et 90% intervall. Da må det være lengre enn et 90% konfidensintervall.

(d) Et konfidensintervall er et intervall som er beregnet ifra en stikkprøve. Det inneholder en populasjonsparameter med en viss sikkerhet, kalt konfidensnivået. Hvis f.eks. konfidensnivået er 95 %, så er vi 95 % sikre på at intervallet inneholder populasjonsparameteren.

Oppgave 6 (4 poeng)

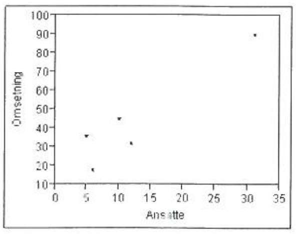
(a) $H_0 : \mu_1 = 30000$ vs $H_1 : \mu_1 \neq 30000$. Testobservator $t = \frac{23198-30000}{14000/\sqrt{17}} = -2.00$. Kritisk verdi er $t_{0.025,16} = 2.12$. Vi kan ikke forkaste H_0. Det er ikke tilstrekkelig grunnlag til å hevde at gjennomsnittsinntekten i by 1 er



forskjellig fra 30000.

(b) Uavhengige stikkprøver, siden de to stikkprøvene har forskjellig størrelse. $H_0 : \mu_1 = \mu_2$ og $H_1 : \mu_1 < \mu_2$. Testobservator $t = \frac{26899-23198}{\sqrt{13000^2/14+14000^2/17}} = 0.76$. Kritisk verdi er $t_{0.05,13} = 1.771$. Vi kan ikke forkaste H_0. Det er ikke tilstrekkelig grunnlag i dataene til å hevde at by 2 har høyere gjennomsnittsinntekt enn by 1.

Oppgave 7 (4 poeng)

(a) 

Det er en ganske klar positiv korrelasjon.

(b) $r^2 = 0.874$. 87.4 % av variasjonen i omsetning skyldes variasjon i antall ansatte.

(c) Den forventede verdien E skal være mer enn 5 i mer enn 80 % av cellene. Her er $E < 5$ i to av fire celler. For eksempel er den forventede verdien av menn som er for lik $E = 16 \cdot 5/33 \simeq 2.4$.