

# Oppgaver til Studentveiledning II

## MET 3431 Statistikk

10. april 2012 kl 17.15 - 20.15 i B2

Handelshøyskolen BI

## **Oppgaver**

### **1. Eksamensoppgaver:**

Prøve-eksamen A fra 2010: Oppgave 6 - 7. Prøve-eksamen A fra 2010 er vedlagt.

### **2. Eksamensoppgaver:**

Prøve-eksamen B fra 2010: Oppgave 1 - 6. Prøve-eksamen B fra 2010 er vedlagt.

## **Løsninger**

### **1 Eksamensoppgaver:**

Løsningsforslag til Prøve-eksamen A fra 2010 er vedlagt.

### **2 Eksamensoppgaver:**

Løsningsforslag til Prøve-eksamen B fra 2010 er vedlagt.

**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: **PRØVE EKSAMEN A**

Totalt antall sider:

## Oppgave 1 (5 poeng)

- Definer de fire målenivåene og gi et eksempel for hvert nivå.
- Forklar kort begrepene *populasjon*, *stikkprøve*, *parameter* og *observator*.
- "38% av voksne nordmenn går jevnlig til legen". Dette var konklusjonen til en student etter at hun hadde spurt 530 tilfeldig valgte medstudenter. Hva er galt med konklusjonen hennes?

## Oppgave 2 (6 poeng)

- Hva er forskjellen på et histogram og et Pareto diagram?
- På en prøve svarer en student riktig på 34 av spørsmålene. Det utgjør ca. 42% av spørsmålene. Hvor mange spørsmål var det totalt?
- Det var 4 studenter som tok prøven. Følgende tabell gir alder ( $x$ ) og antall riktige svar ( $y$ ) på prøven.

alder	23	19	19	30
riktige	34	10	50	80

Lag et scatterplott for dataene, med alder på  $x$ -aksen og antall riktige på  $y$ -aksen.

- Hva er medianen for antall riktige svar i stikkprøven i oppgave (c)?
- Lag et histogram for følgende frekvenstabell:

Alder	Antall
$[30, 40 >$	15
$[40, 50 >$	25
$[50, 60 >$	20
$[60, 70 >$	10

### Oppgave 3 (4 poeng)

- (a) Prisen for to typer frokostblanding (A og B) i fire forskjellige butikker er:

	Butikk 1	Butikk 2	Butikk 3	Butikk 4
A	29	29	30	29
B	23	25	30	22

Uten å beregne standardavvikene, avgjør hvilken type, A eller B, som har størst standardavvik i pris. Begrunn svaret.

- (b) Beregn variasjonsbredde/rekkevidde og standardavvik til prisene for type B.
- (c) Merete svarte riktig på 90% av spørsmålene på en eksamen i finans. På eksamen i statistikk var antall riktige svar lik 90 persentilen. Forklar forskjellen på disse to resultatene.

### Oppgave 4 (4 poeng)

- (a) I en klasse er det 13 kvinner og 49 menn. Hva er sannsynligheten for at en tilfeldig valgt student i klassen er kvinne?
- (b) I tabellen ser vi røykevanene til en gruppe astmapasienter.

	Ikke-røyker	Festrøyker	Vanlig røyker	Storrøyker	Total
Menn	339	33	61	34	467
Kvinner	377	32	84	36	529
Total	716	65	145	70	996

En av de 996 astmapasientene velges tilfeldig. Hva er sannsynligheten for at det ikke er en storrøyker?

- (c) Anta at  $x$  er normalfordelt med forventning  $\mu = 60.0$  og standardavvik  $\sigma = 4.0$ . Hva er sannsynligheten for at  $x$  er mindre enn 53.0? Skisser sannsynligheten som areal i en graf.
- (d) Anta at  $\bar{x}$  er gjennomsnittet av fire tilfeldig valgte verdier av  $x$  fra forrige oppgave. Hva er sannsynligheten for at  $\bar{x}$  er mindre enn 53.0?

### Oppgave 5 (3 poeng)

- (a) Et konfidensintervall for populasjons andelen  $p$  blir oppgitt til å være  $0.620 < p < 0.658$ . Finn feilmarginen. (Vanlig notasjon for feilmargin er  $ME$  eller bare  $E$ )
- (b) Av 865 voksne sier 408 at de går jevnlig til tannlegen. Lag et 95% konfidensintervall for andelen voksne som går jevnlig til tannlegen.
- (c) Seks tilfeldig valgte studenter tar en test og poengene legges inn i JMP. Bruk utskriften (a) bakerst til å lage et 99% konfidensintervall for populasjonsgjennomsnittet  $\mu$ .

## Oppgave 6 (7 poeng)

- (a) En markedsanalytiker hevder at mer enn 50% av kundene foretrekker grovbrød framfor loff. Bruk populasjonsparameteren for andel,  $p$ , til å sette opp nullhypotesen  $H_0$  og alternativhypotesen  $H_1$  på symbolsk form.
- (b) I en stikkprøve av 190 kunder, sier 110 at de foretrekker grovbrød. Test på  $\alpha = 0.05$  nivået om markedsanalytikerens påstand i (a) er sann. Skisser testobservator og p-verdi i en graf.
- (c) I en stikkprøve av 100 mangoer fant man at gjennomsnittsomkretsen var  $\bar{x} = 40.5$  cm og at standardavviket var  $s = 2.0$  cm. Test påstanden om at gjennomsnittsomkretsen for alle mangoer er  $\mu = 39.5$  cm. Skisser testobservator og kritisk verdi i en graf.
- (d) For samme stikkprøve som i Oppgave 5(c) foretas en hypotesetest for påstanden om at  $\mu = 80$  poeng. Se JMP utskriften (b) bakerst. Hva blir konklusjonen i en test på  $\alpha = 0.1$  nivået? Skriv svaret i et lettfattelig språk.

## Oppgave 7 (5 poeng)

- (a) Fem tilfeldig valgte studenter oppgav hvor mange timer ( $x$ ) de brukte på forberedelser til en test, og hvilket resultat i poeng ( $y$ ) de fikk på testen. Dataene ble lagt inn i JMP, se utskriften (c) bakerst. Hva er korrelasjonskoeffisienten  $r$ ? Beregn  $r^2$  og tolk resultatet.
- (b) Test på  $\alpha = 0.05$  nivået om det er en positiv lineær korrelasjon mellom antall timer brukt til forberedelser og oppnådd resultat på testen.
- (c) En ny student skal ta testen. Han oppgir at han har brukt fem timer til forberedelse. Bruk regresjonsformelen til å beregne det forventede poengresultatet på testen.

**Test Mean=value**

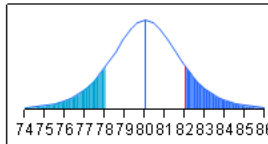
Hypothesized Value 80  
 Actual Estimate 81.9833  
 DF 5  
 Std Dev 4.17632

**t Test**

Test Statistic 1.1633  
 Prob > |t| 0.2972  
 Prob > t 0.1486  
 Prob < t 0.8514

**Moments**

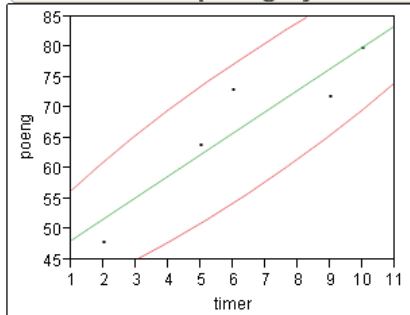
Mean 81.983333  
 Std Dev 4.1763221  
 Std Err Mean 1.7049764  
 Upper 95% Mean 86.366115  
 Lower 95% Mean 77.600552  
 N 6



(a) Oppgave 5c

(b) Oppgave 6d

**Bivariate Fit of poeng By timer**



— Bivariate Normal Ellipse P=0.950  
 — Linear Fit

**Correlation**

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
timer	6.4	3.209361	0.924128	0.0248*	5
poeng	67.4	12.23928			

**Linear Fit**

$poeng = 44.84466 + 3.5242718 * timer$

(c) Oppgave 7

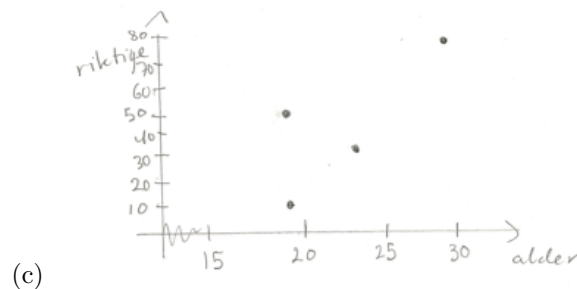
# Fasit Prøveeksamen A, Met 3431

## Oppgave 1

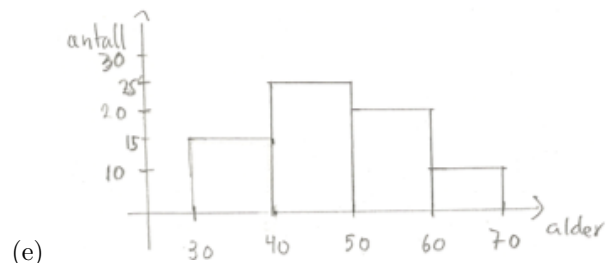
- (a) **Nominalnivå** når dataene faller i kategorier uten noen ordning, som navn, merkelapper Eks: Studieretninger på BI
- Ordinalnivå** når dataene kan ordnes i en bestemt rekkefølge, men avstanden mellom dataverdier ikke har mening. Eksempel: Rangene til militære.
- Intervallnivå** når dataene kan ordnes og avstanden mellom dataverdier er meningsfull. Ikke et naturlig nullpunkt. Eksempel: Årene 1000, 2000, 1776 og 1492.
- Forholdstallnivå** Har i tillegg et naturlig nullpunkt. Brøker har mening. Eks: Prisen på pensumbøker.
- (b) **Populasjon:** Samlingen av *alle* objekter som du ønsker å studere. **Stikkprøve:** Et utvalg av objekter ifra populasjonen. **Parameter:** Et tall som angir et aspekt ved populasjonen, f.eks.  $\mu$ . **Observer:** Et tall som angir et aspekt ved stikkprøven, f.eks.  $\bar{x}$ .
- (c) Konklusjonen omhandler populasjonen av alle voksne. Stikkprøven er hentet fra en annen populasjon, nemlig alle studenter. Det er ikke sikkert at en stikkprøve bestående kun av studenter er representativ for populasjonen av alle voksne.

## Oppgave 2

- (a) Et histogram gir en grafisk fremstilling av kontinuerlige data, slik at vi kan se formen på fordelingen. Pareto diagrammet gir en framstilling av nominale data, slik at vi kan sammenlikne de relative størrelsene på kategoriene.
- (b)  $\frac{34}{0.42} \simeq 81$



(d)  $\tilde{\mu} = (34 + 50)/2 = 42$



### Oppgave 3

(a) Vi ser at prisen til type B varierer mye mer enn for type A. Siden standardavvik er et mål for variasjon/spredning, så vil standardavviket til prisene for B være størst.

(b) Variasjonsbredde =  $30 - 22 = 8$  kr. Standardavviket er

$$s = \sqrt{\frac{(23-25)^2 + (25-25)^2 + (30-25)^2 + (22-25)^2}{3}}$$

$$= \sqrt{\frac{4+0+25+9}{3}} = 3.56 \text{ kr}$$

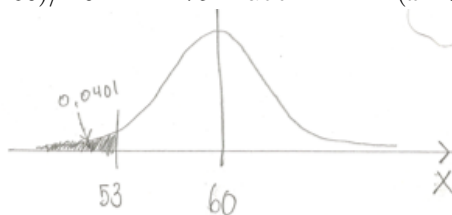
(c) Merete svarte riktig på 90% av spørsmålene i finans, men vi vet ikke hvor god hun var i forhold til resten av klassen. På statistikk var hun bedre enn 90 % av klassen, men vi vet ikke hvor mange spørsmål hun besvarte riktig.

### Oppgave 4

(a)  $\frac{13}{62} \simeq 0.21$

(b) 70 av 996 er storryktere. Sannsynligheten for ikke storryker er  $\frac{996-70}{996} \simeq 0.93$ .

(c) Standardiserer:  $z = (53.0 - 60)/4.0 = -1.75$ . Tabell A-2:  $P(x < 53) =$



$$P(z < -1.75) = 0.0401.$$

(d) Standardiserer:  $z = \frac{53.0-60}{4.0/\sqrt{4}} = -3.50$ . Tabell A-2:  $P(\bar{x} < 53) = P(z < -3.5) = 0.0001$ .

### Oppgave 5

(a)  $ME = \frac{0.658 - 0.620}{2} = 0.019$ .



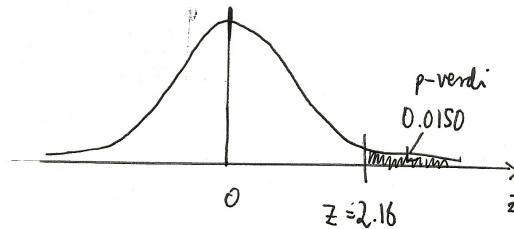
(b)  $ME = 1.96 \cdot \sqrt{\frac{0.472(1-0.472)}{865}} = 0.033$ . Konfidensintervall:  $0.472 \pm 0.033$ .

(c)  $n = 6$ ,  $\bar{x} = 81.983$ ,  $s/\sqrt{6} = 1.705$  og  $df = 5$ . Tabell a-3 gir  $t_{0.005,5} = 4.032$ .  $ME = 4.032 \cdot 1.705 = 6.875$ . Konfidensintervall :  $81.983 \pm 6.875$ , eller  $75.108 < \mu < 88.858$ .

## Oppgave 6

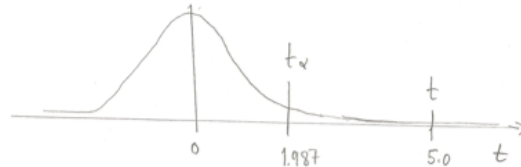
(a)  $H_0 : p = 0.5$  og  $H_1 : p > 0.5$ .

(b) Testobservator  $z = \frac{110/190 - 0.5}{\sqrt{0.5 \cdot (1-0.5)/190}} = 2.18$ . p-verdien er (Tabell A-2) 0.0150. Siden p-verdien er mindre enn 0.05 forkastes  $H_0$ . Det er grunn til å hevde at



andelen er mer enn 50%.

(c) Det står ingenting om signifikansnivå i oppgaven. Vi forutsetter at  $\alpha = 0.05$ . Testobservator  $t = \frac{40.5 - 39.5}{2.0/\sqrt{100}} = 5.0$ . Alternativhypotesen er tosidig, så kritisk verdi fra tabell A-3 er  $t_{0.025,90} = 1.987$  (med 90 df). Siden  $|t| > 1.987$ , så forkastes  $H_0$ . Det er grunn til å hevde at gjennomsnittsvekten ikke er 39.5



cm.

(d) Vi ser at testobservatoren er  $t = 1.1633$ , du kan sammenlikne denne med kritisk verdi  $t_{0.05,5}$  i tabell A-3. Men det er enklere å se i utskriften at p-verdien er 0.2972. Siden denne er større enn  $\alpha = 0.1$  så kan vi ikke forkaste  $H_0$ . Det er ikke grunnlag i dataene til å hevde at populasjons-gjennomsnittet  $\mu$  er forskjellig fra 80.

## Oppgave 7

(a) Ser at  $r = 0.924$  og da er  $r^2 = 0.854$ . 85.4% av variasjonen i oppnådde poeng skyldes variasjon i antall brukte timer til forberedelse.

(b) Testobservator er  $t = \frac{0.924}{\sqrt{(1-0.924^2)/3}} = 4.185$ . Kritisk verdi er  $t_{0.05,3} = 2.353$ . Siden  $t > t_{\alpha/2}$  så forkastes  $H_0 : \rho = 0$ . Det er grunn til å hevde at det er en positiv lineær korrelasjon mellom forberedelse og oppnådde poeng.

(c) poeng =  $44.84466 + 3.5242718 \cdot 5 \simeq 62.5$

**Skriflig 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: **PRØVE EKSAMEN B**

Totalt antall sider:

### Oppgave 1 (6 poeng)

- En aksjemegler sammenligner kursutviklingen til aksje A og aksje B for de siste tre ukene. Er dette en observasjonell studie eller et eksperiment?
- Prisen på et klesplagg settes ned med 20 %. To uker senere settes prisen opp med 20 %. Koster klesplagget nå mer, mindre eller det samme som det gjorde opprinnelig? Begrunn konklusjonen din.
- Lag et enkelt eksempel som viser data i et kakediagram. Vis også de samme dataene i et Pareto diagram. Hvilket diagram synes du er mest effektivt?
- En skoleforsker ønsker en stikkprøve av 100 lærere fra en skolekrets med totalt 800 lærere. Beskriv hvordan dette kan gjøres slik at han får en stratifisert stikkprøve. Beskriv også en måte å ta stikkprøven på slik at den blir av type bekvemmelighet.

### Oppgave 2 (4 poeng)

I en stikkprøve av syv BI studenter er antall timer de jobbet sist uke:

15, 22, 13, 19, 19, 32, 27

- Lag et stamme og blad diagram (stem-and-leaf plot) for stikkprøven. Man kunne også ha laget et histogram for stikkprøven. Nevn en fordel som stamme og blad diagram har framfor et histogram.
- Beregn nedre kvartil, øvre kvartil og medianen for stikkprøven. Lag et boksplokk.

### Oppgave 3 (4 poeng)

- To stikkprøver kan ha samme gjennomsnitt men allikevel være veldig forskjellige. Forklar hvordan og eksemplifiser med to stikkprøver.

- (b) Prisen på en type parfyme varierer fra butikk til butikk. Produsenten oppgir at gjennomsnittsprisen er 148.7 kroner og at standardavviket er 15.3 kroner. Bruk tommelfingerregelen for variasjonsbredde (range rule of thumb) til å anslå en typisk minimums- og maksimumspris.
- (c) I en butikk kostet parfymen 130 kroner. Omregn denne prisen til en  $z$ -verdi. Vil du karakterisere denne prisen som uvanlig lav?

### Oppgave 4 ( 6 poeng)

I denne oppgaven skal vi se på stikkprøven av 62 studenter ved BI Drammen. JMP utskrift (a) bakerst gjelder timer på Facebook per dag.

- (a) Er variabelen *timer Facebook* normalfordelt?
- (b) Hva er variasjonsbredden/rekkevidden i stikkprøven?
- (c) Kjetil et ti timer på Facebook hver dag. Det tilsvarer 90 % persentilen. Hva er sannsynligheten for at en tilfeldig valgt medstudent er mindre på Facebook enn Kjetil?
- (d) Beskriv hva en binomisk forsøksrekke er, og bruk binomialfordelingen til å beregne sannsynligheten for at nøyaktig to av fire tilfeldig valgte medstudenter er mer på Facebook enn Kjetil. Siden en av forutsetningene for binomisk forsøksrekke ikke holder, så blir ikke svaret helt riktig. Hvilken forutsetning er brutt.
- (e) Hva er populasjonen som denne stikkprøven representerer? Oppgi 95 % konfidensintervall for populasjonsgjennomsnittet  $\mu$  og tolk resultatet.

### Oppgave 5 (5 poeng)

I en stikkprøve av 139 menn er 26% røykere. Gjennomsnittsalderen var  $\bar{x} = 45.2$  år og standardavviket var  $s = 13.9$  år.

- (a) Lag et 90 % konfidensintervall for andelen menn som røyker.
- (b) Lag et 90 % konfidensintervall for alderen.
- (c) Finn  $p$ -verdien i en test for påstanden om at mindre enn 30% av alle menn røyker. Skisser testobservator og  $p$ -verdi i en graf.

### Oppgave 6 (4 poeng)

En lege studerer virkningen av et hodepine-middel. Han har en stikkprøve pasienter som får middelet, og en annen stikkprøve der pasientene får en narremedisin (placebo). Det er fire pasienter i hver gruppe. Legen måler for hver pasient hvor mange timer det går før hodepinen er borte.

- (a) Er stikkprøvene relaterte eller uavhengige? Begrunn svaret.

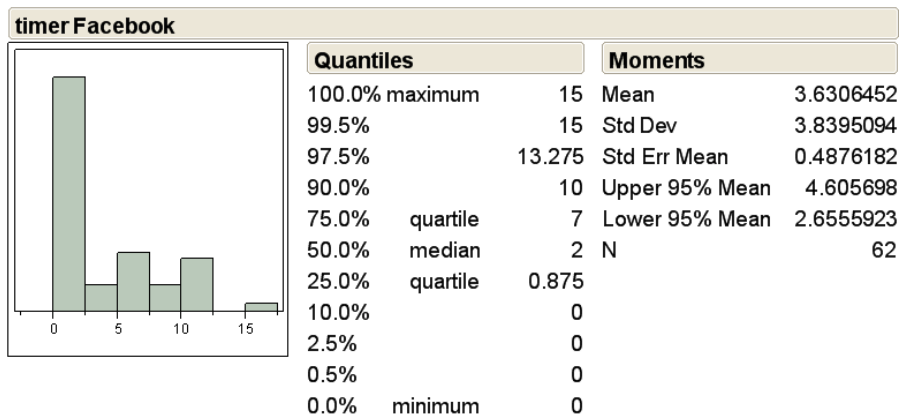
- (b) Tiden det tar før hodepinen er borte er gitt i følgende tabell, sammen med gjennomsnittet og standardavviket:

pasient	1	2	3	4	$\bar{x}$	$s$
hodepine-middel	3.4	3.5	2.1	5.5	3.625	1.403
placebo	4.6	8.3	2.4	6.8	5.525	2.579

Anta at tiden som går før hodepinen er borte er normalfordelt for begge gruppene. Foreta en hypotesetest på  $\alpha = 0.05$  nivået for påstanden: *Hodepine-middelet forkorter tiden det tar før hodepinen er borte.*

## Oppgave 7 (5 poeng)

- (a) I spørreundersøkelsen blant BI studenter våren 2010, oppgav respondentene bl.a. kjønn og hvilken type mobiltelefon de hadde. Resultatene for studiested Nydalen er gitt i JMP i Figur 2 bakerst. Hvor mange studenter er det totalt i stikkprøven, og hvor mange av disse hadde en Samsung? I utskriften er observerte ( $O$ ) og forventede ( $E$ ) verdier gitt for hver celle. Du ønsker å teste på  $\alpha = 0.05$  nivået om det er sammenheng mellom kjønn og type mobiltelefon. Hva er antall frihetsgrader? Hva er kritisk  $\chi^2$  verdi for denne testen? Beregn testobservator og konkluder i et lettfattelig språk.
- (b) Når vi foretar en hypotesetest for om en korrelasjon er signifikant brukes hypotesene  $H_0 : \rho = 0$  og  $H_1 : \rho \neq 0$ . Forklar hva symbolet  $\rho$  betyr. Forklar i et lettfattelig språk hva  $H_0$  og  $H_1$  sier.



Figur 1: Oppgave 4

**Contingency Table**

		kjønn		
		Mann	Kvinne	
mobiltelefon	Count			
	Expected			
	Nokia	164	132	296
		149.966	146.034	
	Sony Ericsson	118	153	271
		137.3	133.7	
	iPhone	142	110	252
		127.674	124.326	
LG	9	9	18	
	9.11954	8.88046		
Samsung	41	79	120	
	60.797	59.203		
Andre	60	37	97	
	49.1442	47.8558		
	534	520	1054	

Figur 2: Oppgave 7





(c) Testobservator

$$z = \frac{0.26 - 0.30}{\sqrt{0.3(1-0.3)}} = -1.03$$

p-verdien blir (tabell a-2) 0.1515. Siden p-verdien er større enn 0.05 forkastes ikke  $H_0$ . Det er ikke grunn til å hevde at andelen røykere er mindre enn 30%.

## Oppgave 6

(a) Stikkprøvene er uavhengige, siden de to stikkprøvene består av ulike personer, det er altså ingen naturlig parvis matching mellom to og to pasienter i placebo gruppen og i hodepine-middel gruppen.

(b) Testobservator

$$t = \frac{3.625 - 5.525}{\sqrt{\frac{1.403^2}{4} + \frac{2.579^2}{4}}} = -1.294$$

Tabell A-3 med 3 frihetsgrader gir kritisk verdi  $t_{0.05;3} = 2.353$ . Vi kan ikke forkaste  $H_0 : \mu_1 = \mu_2$ . Vi har ikke grunnlag nok til å hevde hodepine middelet forkorter tiden det tar før hodepinen er over.

## Oppgave 7

(a) 1054 totalt, 120 hadde Samsung. Det er  $(6-1)(2-1) = 5$  frihetsgrader. Kritisk verdi er  $\chi^2_{0.05;5} = 11.071$ . Testobservator er

$$\chi^2 = \frac{(164 - 149.996)^2}{149.996} + \frac{(132 - 146.034)^2}{146.034} + \frac{(37 - 47.8558)^2}{47.8558} = 29.349$$

Denne verdien er større en kritisk verdi, så vi forkaster  $H_0$ . Vi konkluderer med at det en sammenheng mellom kjønn og type mobiltelefon.

(b)  $\rho$  er en populasjonsparameter, den er den lineære korrelasjonen mellom  $x$  og  $y$ .  $\rho$  måler styrken i korrelasjonen, og  $H_0$  sier at det ikke er noen lineær korrelasjon mellom  $x$  og  $y$ , mens  $H_1$  hevder at det finnes en slik korrelasjon.