# EKSAMENSOPPGAVER

| | | |
|---|---|---|
| Institutt: | **IKBM** | |
| Eksamen i: | **STAT100** | **STATISTIKK** |
| Tid: | Torsdag 12. des 2013 | 09.00-12.30 (3.5 timer) |
| Emneansvarlig: | Solve Sæbø | |

**Tillatte hjelpemidler: C3: alle typer kalkulator, alle andre hjelpemidler**

Oppgaveteksten er på: 12

*antall sider inkl. vedlegg*

Alle deloppgaver teller likt. For hver oppgave er det 5 svaralternativer. Kun ett svar-alternativ er riktig. Du får ett poeng for riktig svar, null poeng for feil svar. Maksimal score er da 40 poeng.

**Alle svar føres i svarskjemaet på side 12 og denne siden er den ENESTE som skal leveres når eksamen er slutt. Husk å skrive kandidatnummer på svarskjemaet!**

***Oppdrett av gås (Oppgave 1-3)***
Kjøttkvaliteten på gås er best når gåsa veier mellom 4 og 5 kilo. En oppdretter av gjess vil derfor slakte gjessene sine dersom han er sikker på at gjennomsnittsvekta på gjessene er på over 4.5 kilo. Han tar derfor et tilfeldig utvalg på 10 gjess og måler vektene deres og får:

Vekt i kilo (X) : 4.7, 4.2, 4.9, 5.1, 4.0, 5.5, 4.9, 4.5, 4.8, 5.0
Tilleggsopplysninger: $\bar{x} = 4.76$ og $s = 0.44$

Anta at $X \sim N(\mu, \sigma)$

**Oppgave 1**
Hva er verdien på den t-fordelte testobservatoren for å teste $H_0 : \mu = 4.5$ mot $H_1 : \mu > 4.5$?

a) 2.03 b) 1.87 c) 1.64 d) 2.45 e) -0.52

**Oppgave 2**

Hvor mange frihetsgrader er knyttet til denne testobservatoren?

a) 9 b) 10 c) 1 d) 8 e) $\infty$

**Oppgave 3**

Bonden er usikker på om han vil slakte nå, så han utvider utvalget av gjess til totalt å bli 32. Gjennomsnittsvekten i utvalget er 4.7, og utvalgsstandardavviket er 0.4. Hva blir et 95% konfidensintervall for $\mu$ i dette tilfellet OG hva kan du råde bonden til?

a) [4.40, 4. 60]. Bonden bør ikke slakte siden intervallet inneholder verdien 4.5
b) [4.51, 4.89]. Bonden bør ikke slakte siden intervallet ikke inneholder verdien 4.5
c) [4.51, 4.89]. Bonden kan slakte siden intervallet ikke inneholder 4.5
d) [4.56, 4.84]. Bonden kan slakte siden nedre intervallgrense er større enn 4.5
e) [4.61, 4.79]. Bonden bør ikke slakte siden intervallet ikke inneholder verdien 4.5

***Bensinpriser (Oppgave 4 - 5)***

Det hevdes at bensinprisene er lavere på mandag morgen enn på mandag kveld. For å sjekke denne påstanden ble $n$=8 tilfeldige bensinstasjoner valgt ut en mandag og prisene ble registrert morgen og kveld. De registrerte prisene ble:

| Bensinstasjon | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Mandag kveld | 14.45 | 14.09 | 14.5 | 14 | 14 | 13.5 | 13.99 | 13.59 |
| Mandag morgen | 12.99 | 13.5 | 13.69 | 12.5 | 13.09 | 13.11 | 14.01 | 12.64 |
| Differanse Kveld - morgen | 1.46 | 0.59 | 0.81 | 1.5 | 0.91 | 0.39 | -0.02 | 0.95 |

Tilleggsopplysninger:

```
        mean     sd        n
kveld  14.02    0.35       8
morgen 13.19    0.51       8
Diff    0.82    0.51       8
```

**Oppgave 4**

Bruk en parvis test for å teste om forventet pris på bensin om kvelden er høyere enn om morgenen på mandager. Hva blir verdien på den t-fordelte testobservatoren?

a) 4.55 b) 0.82 c) 1.61 d) 3.57 e) 1.46

**Oppgave 5**

Hvis du heller vil teste om prisforskjellen (kveld minus morgen) er på mer enn 50 øre (dvs 0.5 kr), hva blir da verdien på testobservatoren?

a) 7.21 b) 2.34 c) 3.16 d) 1.77 e) 2.01

***Høydevekst på juletrær og avstand mellom trærne (Oppgave 6-16)***

En juletreprodusent har gjort et forsøk for å finne ut hva som er den beste avstanden mellom trærne for å få best vekst. Han vet at dersom de står for tett, vil de hemme hverandres vekst fordi de konkurrerer om ressurser som næring, lys og plass. Men, han har også en mistanke om at for stor avstand kan være negativt. Derfor sådde han trær med fire ulike avstander A, B, C og D, der A er korteste avstand og D er største avstand. For hver avstand sådde han 6 trær (3 trær i 2 rader). Dessverre så var det noen trær som ble ødelagt av frost første vinteren slik at det bare ble igjen 4 trær i to av avstandskategoriene. Etter noen år målte han høyden på de gjenlevende trærne, og noen resultater er gitt i Tabell 1

**Tabell 1: Resultater fra R Commander**

| | mean | sd | n |
|---|---|---|---|
| A | 136.25 | 25.62 | 4 |
| B | 166.67 | 20.90 | 6 |
| C | 183.33 | 15.38 | 6 |
| D | 157.50 | 27.54 | 4 |

**Oppgave 6**

Produsenten trodde først at han ikke burde bruke data fra A og D siden det var så få observasjoner igjen for disse to avstandene, så han ville først kun sammenlikne avstandene B og C. Anta at begge grupper har samme populasjonsstandardavvik $\sigma$. Hva blir $S_p$, dvs *pooled* estimat for $\sigma$basert på data fra avstand B og C?

a) 18.35 b) 17.54 c) 18.14 d) 20.90 e) 15.38

**Oppgave 7**

Hvor mange frihetsgrader er knyttet til $S_p$ ?

a) 5 b) 12 c) 10 d) 6 e) 8

**Oppgave 8**

Juletreprodusenten ville teste nullhypotesen $H_0 : \mu_2 - \mu_3 = 0$ mot et tosidig alternativ, der $\mu_2$ er forventet høyde ved avstand B og $\mu_3$ er forventet høyde ved avstand C. Han kjente ikke til muligheten for å regne ut $S_p$ (som i oppgave 6), så han brukte bare utvalgsstandardavviket på 20.90 fra gruppe B som et estimat for $\sigma$. Hvilken verdi fikk han da på testobservatoren sin som han skulle bruke i hypotesetesten?

a) -2.32 b) -1.38 c) -0.67 d) 1.83 e) -1.95

Juletreprodusenten prøvde likevel å bruke alle avstandsgruppene i en ANOVA-modell analyse. Anova-tabellen fra analysen i R Commander er gitt nedenfor i Tabell 2. Noen tall er erstattet med * i tabellen.

**Tabell 2**

```
> Anova(LinearModel)
          Df Sum Sq Mean Sq F value  Pr(>F)
Avstand    * 5533.3 1844.44  3.8777 0.02933
Residuals  * 7610.4  475.65
```

**Oppgave 9**
Hvor mange frihetsgrader er knyttet til SSG, dvs kvadratsummen til gruppevariabelen Avstand?

a) 4 b) 19 c) 2 d) 20 e) 3

**Oppgave 10**
Hvor mange frihetsgrader er knyttet til støykvadratsummen SSE (Residuals)?

a) 4 b) 16 c) 3 d) 20 e) 17

**Oppgave 11**
Anta at det er et felles populasjonsstandardavvik for høyder på juletrær i alle avstandsgruppene. Hva blir estimatet for denne basert på alle fire grupper?
*(Presisering: Vi skal her fram til det estimatet som er basert på kvadratroten av den beste forventningsrette estimatoren for populasjonsvariansen $\sigma^2$)*

a) 19.6 b) 21.8 c) 27.5 d) 22.4 e) 23.2

**Oppgave 12**
Det ble definert en kontrast for å sammenlikne forventningen $\mu_4$ for avstand D med forventningen $\mu_3$ for avstand C på følgende måte: $\theta = \mu_4 - \mu_3$. Hva er det forventningsrette estimatet for $\theta$?

a) -21.67 b) -23.33 c) 157.50 d) -25.83 e) 27.54

**Oppgave 13**
Hva blir standardfeilen til estimatoren $\hat{\theta}$ for kontrasten definert i foregående oppgave?

a) 21.8 b) 475.65 c) 23.78 d) 13.68 e) 14.08

**Oppgave 14**
Juletreprodusenten vil ved hjelp av kontrasten $\theta$ teste om det var en forventet reduksjon i høyde på trærne dersom han gikk fra avstand C til avstand D mellom trærne. Han regnet ut den vanlige testobservatoren for å teste $H_0 : \theta = 0$ mot $H_1 : \theta \leq 0$ og fikk en verdi på denne på -1.84. Det er gitt at denne testobservatoren er t-fordelt med 16 frihetsgrader. I hvilket intervall vil p-verdien for testen ligge?

a) [0.005, 0.01] b) [0.01, 0.025] c) [0.025, 0.05] d) [0.05, 0.1] e) [0.1, 0.25]

**Oppgave 15**
Hva blir verdien av determinasjonskoeffisenten $R^2$ i analysen av juletre-dataene?

a) 0.42 b) 0.53 c) 0.92 d) 0.78 e) 0.65

**Oppgave 16**
Hvis man ser på tallet $(1 - R^2)$, hva er en riktig tolkning av den?

a) Den er et mål på variasjonen i juletrehøyder for en gitt avstand mellom trærne.
b) Den sier hvor stor del av variasjonen i høyder på juletrær som kan forklares ved andre faktorer enn avstanden mellom trærne
c) Det er populasjonsstandardavviket for juletrehøyder for en gitt avstand mellom trærne

d) Den angir den kvadrerte korrelasjonen mellom høyde og avstand.
e) Den er et mål på hvor store del av variasjonen i avstand mellom trærne som kan forklares ved at de har ulike høyder.

***Produksjonsformer på gårdsbruk (Oppgave 17-20)***

I en kartleggingsstudie over produksjonsformer på gårdsbruk ble 113 tilfeldige gårder trukket ut og det ble registrert størrelse og type produksjon. Dataene er gjengitt i Tabell 3. «Blandet» drift betyr at gården har både dyrehold og planteproduksjon. Noen analyse-resultater fra R Commander er gitt i Tabell 4. Man ønsker å teste nullhypotesen $H_0$ : Gårdsstørrelse er uavhengig av produksjonsform, mot alternativ hypotese at de ikke er uavhengige.

**Tabell 3**

| | Liten gård | Middels stor gård | Stor gård | Sum |
|---|---|---|---|---|
| Dyrehold | 8 | 16 | 2 | 26 |
| Blandet | 10 | 20 | 5 | 35 |
| Plantedyrking | 27 | 15 | 10 | 52 |
| Sum | 45 | 51 | 17 | 113 |

**Tabell 4**

```
> .Test$expected # Expected Counts
              Liten        Middels        Stor
Dyr          10.35398      11.73451   3.911504
Blandet      13.93805      15.79646   5.265487
Planter      20.70796             *   7.823009

> round(.Test$residuals^2, 2) # Chi-square Components
             Liten    Middels    Stor
Dyr          0.54     1.55          *
Blandet      1.11     1.12       0.01
Planter      1.91     3.06       0.61

> assocstats(.Table)
                    X^2 df P(> X^2)
Pearson          10.838  * 0.028444
```

## Oppgave 17

Hvor mange frihetsgrader er knyttet til den kji-kvadratfordelte testobservatoren for denne testen?

a) 6 b) 9 c) 3 d) 4 e) 2

## Oppgave 18

Hva er estimatet for det forventede antall middels store gårder som driver kun med planteproduksjon dersom nullhypotesen er riktig?

a) 15 b) 23.5 c) 20.2 d) 13.6 e) 25.4

**Oppgave 19**
Hva er bidraget til testobservatoren Q fra store gårder som kun driver med dyrehold?

a) 1.25 b) 11.35 c) 5.21 d) 0.93 e) 0.62

**Oppgave 20**
Siden det er få observasjoner i noen av kategoriene i Tabell 3 kan det være lurt å slå sammen noen nivåer av variablene for å få høyere observerte antall. Anta at du i Tabell 3 adderer sammen de to første radene (Dyrehold og Blandet) til en ny kategori som du kaller «Dyr/Blandet». Anta at det fortsatt er nullhypotesen om uavhengighet mellom rad- og kolonnevariabelen som skal testes. Hva blir da den estimerte sannsynligheten $p_{13}$ under $H_0$ for at en tilfeldig gård er stor og har driftsform Dyr/Blandet?

a) 0.62 b) 0.32 c) 0.08 d) 0.13 e) 0.02

***Vannforbruk i husholdninger* (Oppgave 21-30)**

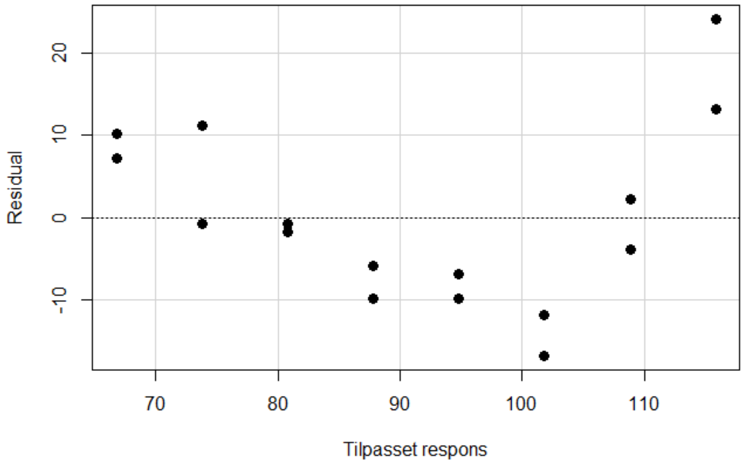
I et område ble $n$=16 husholdninger undersøkt med hensyn til hvor mye vann som forbrukes pr husholdning (liter pr dag og pr person) og hvordan dette kan relateres til antall personer i husholdningen. En regresjonsanalyse ble kjørt med forbruk som responsvariabel og antall personer som forklaringsvariabel (Tabell 5). Deretter ble modellantagelsene sjekket med et residualplott som vist i Figur 1.

**Tabell 5**

```
lm(formula = ForbrukPrPerson ~ Personer, data = Vannforbruk)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  122.839      6.175  19.893 1.16e-11 ***
Personer      -7.006      1.223  -5.729 5.21e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

s: 11.21 on 14 degrees of freedom
Multiple R-squared: 0.701,
Adjusted R-squared: 0.6797
F-statistic: 32.83 on 1 and 14 DF,  p-value: 5.209e-05
```




**Figur 1** Residualplott

**Oppgave 21**

Hva er riktigst å si ut fra residualplottet i Figur 1?

a) Figuren viser et klart brudd på antagelsen om konstant varians for støyleddet i regresjons-modellen.

b) Fordelingen til støyleddet ser ut til å være binomisk fordelt og ikke normalfordelt som an-tatt i modellen.

c) Modellantagelsene om lineæritet, konstant varians og normalfordeling ser ut til å være opp-fylte.

d) Det er en ikke-lineær sammenheng mellom forbruk og antall personer i husholdningene.

e) Det ser ut til at de tilpassede verdiene opptrer i par, noe som er et brudd på antagelsen om uavhengige observasjoner.

De som stod for undersøkelsen var ikke helt fornøyd med modellen, blant annet fordi $R^2$ ikke var på mer enn 0.70. De omdefinerte derfor forklaringsvariabelen fra å være antall personer til å bli $x = 1/\text{Antall personer}$ (altså den inverse av personantallet). Denne forklaringsvariabelen ble så brukt i en vanlig regresjonsmodell med Forbruk per person som responsvariabel $y$. Modellen som ble brukt var da:

$y_i = \alpha + \beta x_i + \epsilon_i$, der $\epsilon_i \sim N(0, \sigma)$ og alle uavhengige.

En utskrift fra analysen er gitt i Tabell 6.

**Tabell 6**

```
lm(formula = ForbrukPrPerson ~ invPersoner, data = Vannforbruk)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   68.203      1.813       *       *
invPersoner   68.022      4.150       *       *

s: 4.562 on 14 degrees of freedom
Multiple R-squared: *,
Adjusted R-squared: *
F-statistic: 268.6 on 1 and 14 DF,  p-value: 1.567e-10

> Anova(LinearModel)
            Df Sum Sq Mean Sq F value    Pr(>F)
invPersoner  1 5590.1  5590.1  268.64 1.567e-10 ***
Residuals   14  291.3    20.8
```

**Oppgave 22**

Hvor stor er determinasjonskoeffisienten for denne modellen?

a) 0.99 b) 0.87 c) 19.19 d) 0.95 e) 0.79

**Oppgave 23**

Hva er riktig tolkning av $\hat{\beta} = 68.022$ ?

a) Ca 68 % av variasjonen i Forbruk kan forklares ved den lineære modellen
b) Når x øker med én enhet vil det estimerte standardavviket øke med 68.022
c) Dersom det inverse personantallet øker med én enhet vil det estimerte endringen i Forbruk pr person bli på 68.022 liter pr dag.
d) Når forbruket øker med én enhet vil den estimerte endringen i x være på 68.022
e) Det er det forventede forbruket når x=0

**Oppgave 24**

Dersom man vil teste om det er lineær sammenheng mellom x og y i denne modellen, hva blir verdien av den t-fordelte testobservatoren?

a) 2.36 b) 9.75 c) 37.62 d) 13.42 e) 16.39

**Oppgave 25**

Finn et 99% konfidensintervall for $\beta$

a) [58.76, 77.29] b) [64.72, 71.32] c) [55.67, 80.38] d) [61.84, 74.20] e) [49.50, 86.55]

**Oppgave 26**

Hvordan kan vi tolke et 99% konfidensintervall for $\beta$?

a) Det er 99% sannsynlig at den estimerte effekten av forklaringsvariabelen x på responsen y ligger i dette intervallet.
b) Sannsynligheten for at forventet forbruk for gitt verdi av x skal ligge i dette intervallet er 99%
c) Det er 99% sannsynlig at verdien av $\alpha + \beta x$ skal ligge i dette intervallet.
d) Det er 99% sannsynlig at dette intervallet dekker den sanne verdien av $\beta$
e) Sannsynligheten for at den estimerte $\hat{\beta}$ ligger i intervallet er 99%

**Oppgave 27**

Hva er estimert forventet forbruk pr person hvis antall personer i husstanden er 4?

a) 85.2 b) 300.2 c) 120.3 d) 80.3 e) 93.4

**Oppgave 28**

Det anslåtte daglige forbruket i en tilfeldig husholdning med 1 person er i følge den estimerte modellen 136.2 liter. Hva blir da et 95% prediksjonsintervall for y når x=1? (Tilleggsopplysning: Gjennomsnittlig x-verdi er $\bar{x} = 0.34$).

a) [127.8 , 144.6]
b) [124.8 , 137.6]
c) [124.5 , 147.9]
d) [105.6 , 116.8]
e) [129.8 , 142.6]

**Oppgave 29**

Hvorfor blir et prediksjonsintervall for en tilfeldig ny observasjon av responsen bredere enn et konfidensintervall for forventet y, når begge er beregnet for x=1 og med samme verdi av $\alpha$?

a) Fordi usikkerheten knyttet til en ny observasjon er større enn usikkerheten til den sanne modellen.
b) Fordi det er veldig usikkert om den tilfeldig valgte husholdningen virkelig har 1 person.
c) Fordi et prediksjonsintervall også må ta hensyn til den ekstra usikkerheten som ligger i at den nye observasjonen vil varierer rundt sin forventning i tillegg til usikkerheten knyttet til den estimerte modellen.
d) Fordi det er vanskeligere anslå en forventning en enkelt ny observasjon.
e) Fordi prediksjonsintervaller som regel har høyere konfidensnivå og dermed bredere intervall.

**Oppgave 30**

Hvor stor er den estimerte endringen i forbruk dersom *antall personer* i husstanden øker fra 1 til 2 personer

a) 136.0 b) 0.64 c) -17.0 d) -45.3 e) -34.0

***Bakterienivå i vakuumpakket laks (Oppgave 31 - 34)***

En produsent av vakuumpakket laks vil undersøke holdbarheten av produktet sitt ved å måle bakterienivået (y) (målt på logaritmisk skala) i *n*=15 pakker som er lagret i ulikt antall dager (x). Produsenten antok følgende modell:

$y_i = \alpha + \beta x_i + \epsilon_i$, der $\epsilon_i \sim N(0, \sigma)$

En utskrift fra analysen er gitt i Tabell 7

**Tabell 7**

```
lm(formula = Bakterier ~ Dager, data = Bakteriedata)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.51640    0.56764   0.910   0.3795
Dager         0.22984    0.08292   2.772   0.0159 *

s: 0.584 on * degrees of freedom
Multiple R-squared: 0.3715,
Adjusted R-squared: 0.3231
F-statistic: 7.683 on 1 and 13 DF,  p-value: 0.01586
```

**Oppgave 31**

Hva er den estimerte forventede endringen i bakterieverdien over en lagringsperiode på 5 dager?

a) 1.67 b) 0.52 c) 2.30 d) 0.23 e) 1.15

**Oppgave 32**

Et konfidensintervall for $\beta$ ble beregnet til [ 0.051 , 0.409 ]. Hva er konfidensnivået til dette intervallet?

a) 80% b) 90% c) 95% d) 98% e) 99%

**Oppgave 33**
Hvordan tolker vi parameteren $\sigma$ i regresjonsmodellen?

a) Den er et mål på den observerte totalvariasjonen i bakterieverdier i de 15 målingene som ble gjort
b) Hvis denne er positiv, betyr det at bakterieverdiene øker med lagringstiden.
c) Den er et mål på hvor mye av variasjonen i bakterieverdier som kan forklares ved den lineære sammenhengen til antall lagringsdager.
d) Den er et mål på variasjonen i bakterieverdier for en gitt lagringstid, og denne variasjonen skal være den samme for alle lagringstider.
e) Den er et mål på graden av lineær sammenheng mellom lagringstid og bakterieverdier.

**Oppgave 34**
Anta at du vil teste $H_0 : \beta = 0$ mot $H_1 : \beta > 0$. Hva blir p-verdien for testen?

a) 0.0159 b) 0.0080 c) 0.398 d) 0.05 e) 0.0318

***Om korrelasjonen mellom kontinuerlige variable (Oppgave 35 - 36)***

**Oppgave 35**
Gitt er følgende størrelser basert på *n*=5 observasjoner av to variabler x og y:
$S_{xy} = 1.98$, $S_x = 0.85$ og $S_y = 2.63$. Hva er korrelasjonen *r* mellom x og y?

a) 0.40 b) 0.77 c) 0.93 d) 0.66 e) 0.89

**Oppgave 36**
Hvilket av følgende utsagn om korrelasjonen *r* mellom to variabler x og y er IKKE korrekt:

a) Dersom *r* <0 er det en negativ sammenheng mellom variablene.
b) Korrelasjonen *r* måler styrken av den lineære sammenhengen mellom to variabler.
c) Korrelasjonen *r*=0 dersom $\hat{\beta} = 0$ fra regresjonsmodellen $y = \alpha + \beta x + \epsilon$
d) Korrelasjon er alltid et tall mellom -1 og 1
e) Korrelasjonen mellom x og y er ikke lik korrelasjonen mellom y og x (rekkefølgen er viktig).

***Om sopp i tobakk-planter (Oppgave 37 – 40)***
På en tobakkplantasje ville de undersøke graden av soppinfeksjon hos plantene. Totalt ble *n*=100 planter høstet i undersøkelsen, hvorav *Y*=80 var infisert med sopp.

**Oppgave 37**
Hvilken sannsynlighetsfordeling har *Y*?

a) $B(100, 80)$ b) $B(100, p)$ c) $N(\mu, \sigma/\sqrt{100})$ d) $N(0.8, p(1-p))$ e) $B(80, \sigma)$

**Oppgave 38**
Finn ved normaltilnærming et 95% konfidensintervall for sannsynligheten *p* for at en tilfeldig plante på plantasjen er soppinfisert.

a) [0.72, 0.88] b) [0.68, 0.92] c) [0.75, 0.85] d) [0.78, 0.88] e) [0.5, 1]

**Oppgave 39**
Plantasjeeierne syntes konfidensintervallet var for bredt/langt og ville samle inn et nytt utvalg planter for å få sikrere svar. De ville på forhånd finne ut hvor mange planter som trengtes for at 95% intervallet (tilsvarende oppgave 38) skulle ha en lengde $L$ som var mindre enn 0.1. Basert på den første undersøkelsen brukte de $\hat{p} = 0.8$ i beregningene. Hvor mange planter trengte de?

a) 310 b) 456 c) 246 d) 203 e) 157

**Oppgave 40**
På en naboplantasje testet de ut et nytt soppmiddel på en del av plantasjen. Deretter samlet de tilfeldig inn 230 planter hvorav 110 planter var fra den delen som ikke var behandlet og 120 planter var fra den delen som hadde fått soppmiddel. Dataene kan oppsummeres i en tabell:

| | Behandlet del | Ubehandlet del | Sum |
|---|---|---|---|
| Soppinfisert plante | 70 | 80 | 150 |
| Frisk plante | 50 | 30 | 80 |
| Sum | 120 | 110 | 230 |

Utfra dataene beregnet de en testobservator Q=5.24. Hva er det laveste av følgende testnivåer som gir forkastning av en hypotese om at infeksjonssannsynligheten er uavhengig av behandling?

a) 0.005 b) 0.01 c) 0.025 d) 0.05 e) 0.1