

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1000 – Innføring i anvendt statistikk

Eksamensdag: 13. januar 2017

Tid for eksamen: 14.30 – 18.30

Oppgavesettet er på 3 sider

Tillatte hjelpemidler: Godkjent kalkulator, ordliste for STK1000, og lærebok (alle utgaver, og det er lov å notere i læreboka)

Kontroller at oppgavesettet er komplett  
før du begynner å besvare spørsmålene. Lykke til!

### Oppgave 1

I en norsk studie av 95 personer med medfødt beinskjørhet (Osteogenesis Imperfecta, OI), samles det inn store mengder data.

- Når forskerne skal publisere en tabell med bakgrunnsdata, har de både kategoriske data, skjevfordelte numeriske data og symmetrisk fordelte numeriske data som skal oppsummeres. Hvilke oppsummeringstall bør forskerne oppgi i tabellen? Begrunn svaret.
- I denne studien fant man ut at 47 personer hadde høyere utdanning. Gi et estimat for andelen personer med OI-diagnose som har høyere utdanning, og beregn så et 95% konfidensintervall for det samme. Bruk dette til å avgjøre om det er en signifikant høyere andel av OI-pasienter som har høyere utdanning enn i den generelle befolkningen, der den tilsvarende andelen er 36%.

### Oppgave 2

I en studie av beinhelse i et tilfeldig utvalg av 39 kvinner var det tre variabler som var av spesiell interesse: Alder, beintetthet (bmd) målt i lårhalsen, og antall beinbrudd de siste ti årene. Beintetthet (bmd) ble målt på et sykehuslaboratorium, og opplysninger om antall beinbrudd ble hentet fra sykehusets journaler. R ga følgende deskriptive statistikk for de tre variablene:

```
alder
Min. 1st Qu. Median Mean 3rd Qu. Max. sd
30.170 36.760 52.330 51.190 63.520 76.300 13.692
bmd
Min. 1st Qu. Median Mean 3rd Qu. Max. sd
```

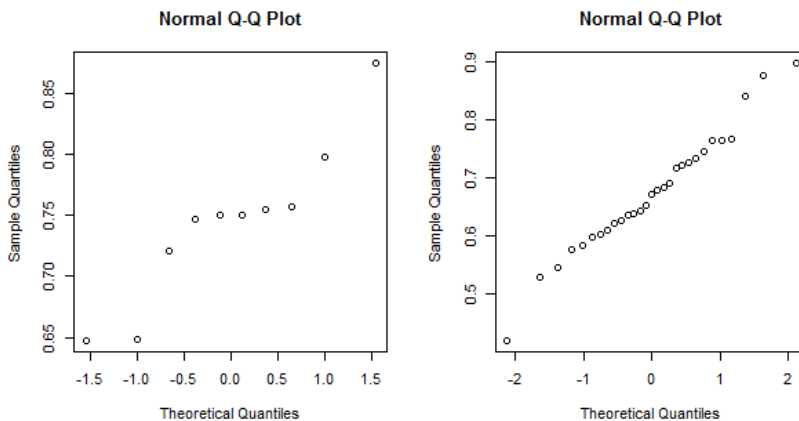
```

0.4191  0.6303  0.6911  0.6919  0.7524  0.8960  0.0999
beinbrudd
  Min.   1st Qu. Median   Mean   3rd Qu.  Max.    sd
0.0000  0.5000  1.0000  1.9230  2.0000 11.0000  2.5994

```

Det viste seg at 10 av kvinnene hadde ingen beinbrudd i løpet av de siste ti årene, mens 29 kvinner hadde hatt ett eller flere beinbrudd. En sammenligning av disse to gruppene viste at gjennomsnittlig bmd blant dem uten brudd var 0.67 (SD = 0.104), men gjennomsnittlig bmd blant dem med brudd var 0.76 (SD = 0.066).

- a) Sett opp hypoteser, og gjennomfør en hypotesetest med nivå 0.05 for å undersøke om det er statistisk signifikant forskjell på gruppene. Forklar hvorfor figuren under kan brukes til å begrunne valg av test.



- b) Beregn et 95% konfidensintervall for bmd for kvinner med brudd, og et 95% konfidensintervall for bmd for kvinner uten brudd.  
c) Kommenter svarene du fikk i oppgave 2 b) i lys av det du fant i oppgave 2 a). (Tips: Fikk du det du forventet? Hvis ikke, kan det forklares?)

Det er kjent at beintettheten varierer med alder, og man vet en del om mekanismene bak dette. Flere studier har funnet at beintettheten øker i ungdomsårene, når sitt høyeste nivå i siste halvdel av tjuårene, for så å avta. Forskerne ønsket å undersøke om den kjente sammenhengen mellom alder og bmd kunne reproduseres i dette datamaterialet og gjorde følgende regresjonsanalyse:

Call:

```
lm(formula = bmd ~ alder)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-0.239353 -0.048540  0.006791  0.062128  0.190653

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.838102    0.058411  14.348  <2e-16 ***
alder        -0.002855    0.001103  -2.588  0.0137 *
---

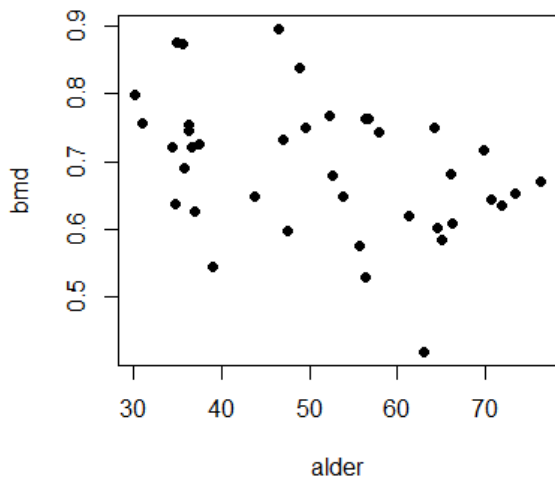
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 0.09313 on 37 degrees of freedom
Multiple R-squared:  0.1533,    Adjusted R-squared:  0.1304
F-statistic: 6.698 on 1 and 37 DF,  p-value: 0.01371

```



- d) Formuler regresjonsmodellen og forklar hvordan regresjonsparameterne tolkes.
- e) Gi et estimat for sammenhengen mellom alder og beintetthet, og beregn et 95% konfidensintervall for det samme. Er den observerte sammenheng mellom de to variablene som forventet? Begrunn svaret.

Etter diskusjoner mellom forskerne, ble følgende multiple regresjonsanalyse også gjort:

```
Call:
lm(formula = bmd ~ alder + beinbrudd)

Residuals:
    Min       1Q   Median       3Q      Max
-0.120725 -0.056434 -0.005432  0.042895  0.166144

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.8217694  0.0488065  16.837 < 2e-16 ***
alder        -0.0017087  0.0009593  -1.781  0.083316 .
beinbrudd    -0.0216755  0.0052053  -4.164  0.000186 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07756 on 36 degrees of freedom
Multiple R-squared:  0.4285,    Adjusted R-squared:  0.3968
F-statistic: 13.5 on 2 and 36 DF,  p-value: 4.226e-05
```

- f) Forklar hvorfor eller hvorfor ikke denne regresjonsanalysen er det vi bruker for å vurdere om variabelen beinbrudd er en konfunderende variabel for sammenhengen mellom alder og bmd, og gi din vurdering av om den er det.
- g) Måling av beintetthet krever at personen som skal måles, må gjennom en undersøkelse på sykehuslaboratoriet. Det hadde vært fint om vi kunne brukt regresjonsanalyse til å predikere beintettheten basert på alderen. Beregn et 95% prediksjonsintervall for beintettheten for en 45 år gammel kvinne, og bruk opplysningene i resten av oppgaven til å vurdere om svaret du får vil være nyttig i praktisk bruk.

**SLUTT**