

## 2. obligatoriske oppgave i STK1000

Høst 2010

Oppgavesettet består av 3 oppgaver. For å løse oppgavene trenger du hjelp av MINITAB eller annen statistisk programvare. I forbindelse med bruk av MINITAB kan du ha nytte av notatet *Starthjelp i MINITAB* (kalt "innføringsheftet" nedenfor). Det er tilgjengelig på hjemmesiden til kurset. Datasett brukt i oppgaven vil være direkte tilgjengelig fra hjemmesiden til kurset (under "Informasjon om obligatoriske oppgaver")

I den skriftlige besvarelsen av oppgavene skal du kort forklare hvordan de enkelte punktene er løst. *For å få oppgaven godkjent eller mulighet til å levere på nytt, må du ha prøvd å løse alle deloppgaver.* Merk at det er lov å spørre om hjelp underveis.

Oppgaven skal i utgangspunktet skrives med et tekstbehandlingsprogram. Hvis du velger å skrive for hånd bør du begrunne hvorfor. Der du bruker MINITAB, må relevante utskrifter og figurer legges ved eller limes inn i besvarelsen. Instruksjoner for utskrift fra MINITAB finner du i avsnitt 10 i innføringsheftet.

Obligen skal leveres med en egen forside som du finner under <http://www.math.uio.no/academics/obligforsideMI.pdf>.

Det er lov å samarbeide og å bruke hjelpemidler. Den innleverte besvarelsen skal imidlertid skrives av deg og gjenspeile din forståelse av stoffet. Er vi i tvil om at du virkelig har forstått det du har levert inn, kan vi be deg om en muntlig redegjørelse.

Besvarelsen leveres på instituttkontoret ved Matematisk Institutt i 7. etasje, Niels Henrik Abels hus (Matematikkbygningen).

Frist for innlevering er **torsdag 4. november 2010 kl 14.30.**

### Oppgave 1

Friske benstrukturer blir kontinuerlig fornyet gjennom to prosesser. Gjennom benformering blir nye ben bygget; gjennom celle nedbrytning blir gamle ben fjernet. Hvis en av disse prosessene blir forstyrret (av sykdom, alder etc) kan det resultere i tap av benstrukturer. Osteocalcin (OC) er en biokjemisk markør for benformering: Høye nivåer av benformering er assosiert med høye verdier av OC. En blodprøve blir brukt til å måle OC, noe som er mye enklere og billigere enn direkte måling av benformering. Enheten er milligram av OC per milliliter av blod (mg/ml). Et studie undersøker forskjellige biomarkører for benproduksjon. Nedenfor er OC målinger av 31 friske kvinner i alder 11 til 32 år som deltok i studiet:

68.9 56.3 54.6 31.2 36.4 31.4 52.8 38.4 35.7 76.5 44.4  
40.2 77.9 54.6 9.9 20.6 20.0 17.2 24.2 20.9 17.9 19.7  
15.9 20.8 8.1 19.3 16.9 10.1 47.7 30.2 17.2

Dataene, som vi vil referere til som  $x_1, \dots, x_n$  med  $n = 31$  er også tilgjengelige fra kursets hjemmeside

- (a) Gjør en eksplorativ analyse av dataene (bruk oppsummerende mål, histogram, bokplot, normal kvantilplott). Beskriv fordelingen.
- (b) Anta observasjonene er  $N(\mu_x, \sigma_x)$  med  $\sigma_x = 20$ . Lag et 95% konfidensintervall for  $\mu_x$ , forventningen til OC. Kommentér egnetheten av ditt valg av prosedyre i forhold til disse data.

Hint: Kommandoen **Stat->Basic Statistics->1-Sample Z...** vil være nyttig her.

- (c) Anta igjen observasjonene er  $N(\mu_x, \sigma_x)$ , men betrakt nå den mer realistiske situasjonen der  $\sigma_x$  er ukjent. Lag et 95% konfidensintervall for  $\mu_x$  i dette tilfellet.

Kommentér eventuelle forskjeller fra resultatet i (b).

Hint: Her kan du bruke kommandoen **Stat->Basic Statistics->1-Sample t...** som vil være nesten identisk med **1-Sample Z...** Den eneste forskjellen er at du nå ikke behøver å spesifisere standard avviket.

- (d) Variable som måler konsentrasjoner har ofte høyreskjeve fordelinger. Av den grunn bruker man gjerne å transformere data til en logaritmisk skala (naturlig logaritme) før analyse.

En logaritmisk transformasjon kan utføres i Minitab ved kommandoen

**Calc->Calculator.**

I feltet merket **Expression:** skriver du  $\log_e(C1)$  (hvis dataene dine ligger i kolonnen C1). I feltet **Store result in variable:** skriver du f.eks C2.

Utfør denne transformasjonen og kall disse  $y_1, \dots, y_n$ . Gjenta oppgave (a) for de transformerte data. Kommentér resultatene relativt til de du fikk i (a).

- (e) Anta nå de transformerte observasjonene følger en fordeling  $N(\mu_y, \sigma_y)$  der  $\mu_y = \log(\mu_x)$  dvs logaritmen til  $\mu_x$ .

Lag et 95% konfidensintervall for  $\mu_y$  når vi antar  $\sigma_y$  er ukjent.

- (f) Gitt et konfidensintervall  $[a, b]$  for  $\mu_y$  kan man lage et konfidensintervall for  $\mu_x$  ved  $[e^a, e^b]$ . Her betyr  $e^a$  at du opphøyer konstanten  $e$  i  $a$ . Dette får du til ved **exp** kommandoen på en kalkulator eller ved bruk av Minitab: **Calc->Calculator** og fyll feltet merket **Expression:** med **exp(a)** der **a** er den verdien du vil opphøye  $e$  med.

Bruk dette til nå å lage et alternativt konfidensintervall for  $\mu_x$ .

Sammenlikn dette konfidensintervallet med det du fikk i (c). Kommentér eventuelle forskjeller.

- (g) Å opphøye  $e$  i en verdi er den inverse transformasjonen til log. Dvs hvis du opphøyer  $e$  i  $\log(x)$ , så får du tilbake  $x$  ( $e^{\log(x)} = x$ ). Dermed transformerer du deg da tilbake til opprinnelig skala.

Hvis  $[a, b]$  er et konfidensintervall for  $\mu_y$ , argumenter hvorfor  $[e^a, e^b]$  er et konfidensintervall for  $\mu_x$ .

Hint: Denne oppgaven krever en matematisk argumentasjon som for noen kanskje vil være litt krevende. Start med å skrive intervallet som  $a < \mu_y < b$  og bruk dette til å lage et tilsvarende sett av ulikheter for  $\mu_x$ .

## Oppgave 2

Tabell 7.3 i boka gir nivåer på tre ulike forurensningskilder i eksosen fra 46 tilfeldig valgte kjøretøy av samme type. Vi vil her konsentrere oss om nitrogen oxides (NOX). Dataene er også tilgjengelig fra kursets hjemmeside.

- Utfør en eksplorativ dataanalyse og argumenter for at bruk av en normalfordeling på data vil være rimelig i dette tilfellet.
- Lag et 95% konfidensintervall for forventet NOX nivå.
- Din sjef håper at NOX nivået er mindre enn 1 gram per mile. Du vil måtte fortelle han at dette ikke er tilfelle. Utfør en signifikanstest for å vurdere styrken i konklusjonen at forventet NOX nivå er større enn 1 og skriv så en liten rapport til din sjef basert på dine analyser. Legg her til grunn at din sjef ikke har hørt om P-verdier så du må bruke et forståelig språk.

## Oppgave 3

Oppgave 7.82 i boka.

Lykke til!