

# Obligatorisk oppgave nr. 2 i STK1000

Høsten 2011

Oppgavesettet består av 3 oppgaver. For å løse oppgavene trenger du MINITAB eller annen statistisk programvare. I forbindelse med bruk av MINITAB kan du ha nytte av innføringsheftet *Starthjelp i MINITAB*. Det er tilgjengelig på hjemmesiden til kurset. Datasett til denne obligen vil du også finne på hjemmesiden.

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 (utenom 3h)*. Merk at det er lov å spørre om hjelp underveis.

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.

Konsulter <http://www.mn.uio.no/math/studier/obligerh11.html> for regler for obligatoriske innleveringer før du setter i gang! Obligen skal leveres med en egen forside som du også finner der. Der du bruker MINITAB, må relevante utskrifter og figurer legges ved eller limes inn i besvarelsen. Instruksjoner for utskrift fra MINITAB finner du i innføringsheftet.

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

Frist for innlevering er **torsdag 3. november 2011 kl 14.30**.

## Oppgave 1

Friske benstrukturer blir kontinuerlig fornyet gjennom to prosesser. Gjennom benformering blir nye ben bygget; gjennom cellenedbrytning blir gamle ben fjernet. 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). En studie undersøker forskjellige biomarkører for benproduksjon. Nedenfor er OC-målinger av 31 friske kvinner i alder 11 til 32 år som deltok i studien:

68,9 56,3 54,6 31,2 36,4 31,4 52,8 38,4 35,7 76,5 44,4  
40,2 57,3 54,6 9,9 20,6 20,0 17,2 24,2 40,9 37,9 19,7  
25,9 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 (bones.MTP)

(a) Utfør en eksplorativ analyse av dataene (bruk oppsummerende mål, histogram, boksplott, normalfordelingsplott). Beskriv fordelingen.

(b) Anta at observasjonene er  $N(\mu_x, \sigma_x)$  med  $\sigma_x = 18$ . Lag et 95% konfidensintervall for  $\mu_x$ , forventningen til OC. Bruk formel og tabell først, og kontroller svaret ditt ved å deretter gjøre det samme i MINITAB. Kommentér egnetheten av ditt valg av prosedyre i forhold til datasettet.

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

(c) Lag også 90% og 99% konfidensintervaller for  $\mu_x$  og forklar hvordan lengden av intervallene forandrer seg.

(d) Anta igjen at 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: Dette er stoff fra kap. 7.1. Her kan du bruke kommandoen **Stat->Basic Statistics->1-Sample t...** som du kan bruke på samme måte som **1-Sample Z...** Den eneste forskjellen er at du nå ikke skal spesifisere standardavviket.

## Oppgave 2

Oppgave 7.24 i boka om estimering av forventet antall venner på Facebook. Datafilen heter Facebookfriends.MTP.

## Oppgave 3

Vi er interesserte i å finne ut om det er forskjell mellom kjønnene når det gjelder hvilepuls. Til å undersøke dette skal du bruke ditt modifiserte puls-datasett fra oppgave 2 i det første obligatoriske oppgavesettet (Pulse.MTW, se den oppgaven for detaljer). Siden vi er interessert i hvilepuls, skal du i hele oppgaven konsentrere deg om variabelen Pulse1. Husk å legge inn 1 eller 2 for kjønn i variabelen Sex i linje 93, som representerer deg selv.

(a) Bruk kommandoen Stat > Basic Statistics > Display Descriptive Statistics til å finne gjennomsnittlig hvilepuls og empirisk standardavvik separat for kvinner og for menn.

(b) Lag boksplott av pulsmålingene for kvinner og for menn i samme figur, og forklar hvilke av størrelsene fra utskriften i punkt a) du kan finne igjen i boksplottene.

(c) For å svare på problemstillingen gitt først i oppgaven, vil vi for enkelthets skyld anta at den deskriptive oppsummeringen av Pulse1 for kvinner (Sex=2) gir den sanne forventningen for denne gruppen. Anta videre at standardavviket for hvilepuls for menn er 10.0 (slag per minutt). Formuler problemstillingen som et hypotesetestingsproblem med nullhypotese og alternativ hypotese. Bruk resultatene fra punkt a) til å beregne testobservatoren slik den er gitt på side 367 i læreboka.

Presisering: Du skal teste om forventningen for menn er lik det observerte gjennomsnittet for kvinner uten å ta hensyn til usikkerhet i gjennomsnittet for kvinner. Til sist i oppgaven (under frivillig punkt h) skal du også ta hensyn til denne usikkerheten.

- (d) Finn P-verdien for testen. Forklar hvordan denne tolkes og hva resultatet av testen betyr.
- (e) Gjør testen direkte ved hjelp av kommandoen Stat > Basic Statistics > 1-Sample z. Kontroller at resultatene blir de samme som de du fikk i punktene c) og d). Du vil her bli nødt til å lage et eget arbeidsark eller "worksheet" som kun inneholder data for gruppen bestående av menn (Sex=1). Dette kan du gjøre ved kommandoen Data > Split Worksheet. I boksen for "By variable" skriver du 'Sex'. Du får da delt opp arbeidsarket i to deler, en for kvinner og en for menn. I punktene f) og g) nedenfor bruker du den delen som svarer til menn.
- (f) Lag et konfidensintervall for forventet hvilepuls for menn. Ved å bruke sammenhengen mellom tester og konfidensintervall, hva blir resultatet av testen da?
- (g) Å anta at standardavviket er kjent, er urealistisk. Bruk teorien i avsnitt 7.1 i læreboka til å utføre en test på forventningen for gruppen av menn når du nå ikke forutsetter kjennskap til standardavviket. (Du kan her bruke kommandoen Stat > Basic Statistics > 1-Sample t i MINITAB)
- Hvor mange frihetsgrader får denne testen?
- Sammenlign konfidensintervallet du får når det ikke forutsettes kjent standardavvik med konfidensintervallet i punkt f). Hva blir dine konklusjoner nå?
- (h) FRIVILLIG, men lærerik! Antagelsen vi gjorde om at forventningen til hvilepuls for kvinner var kjent (og identisk med gjennomsnittet), er ikke riktig. Det er imidlertid mulig å teste om de to gruppene har forskjellig forventning uten å gjøre antagelser om at den ene er kjent. Teorien bak dette er beskrevet i avsnitt 7.2 i læreboka. I MINITAB kan du bruke kommandoen Stat > Basic Statistics > 2-Sample t for å utføre en slik test. Du må nå gå tilbake til det opprinnelige arbeidsarket som inneholder dataene for både kvinner og menn. Det får du fram ved å klikke på Window i hovedmenyen og deretter markere arbeidsarket. I MINITAB har du en mulighet til å velge å anta at standardavvikene for de to gruppene er like eller ikke. Utfør testen ved hjelp av denne kommandoen både når du antar at standardavvikene er like og når du ikke gjør det. Diskuter resultatene og spesielt forskjeller mellom de to testene.

Lykke til!