

Andre obligatoriske oppgavesett for STK1000 høsten 2012

Oppgavesettet består av 3 oppgaver. Til mange av deloppgavene 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.

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.

Det er lov til å samarbeide og å bruke hjelpemidler. Hvis du samarbeider med noen må du oppgi hvem du har samarbeidet med på første side av besvarelsen. **Den innleverte besvarelsen skal imidlertid skrives av deg og gjenspeile din forståelse av stoffet.** Du kan altså ikke levere helt identisk besvarelse som en annen student. 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/obligerh12.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 1. november 2012 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. 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). En studie undersøker forskjellige biomarkører for benproduksjon. Nedenfor er OC-målinger av 31 friske kvinner i alderen 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 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 ('bones.MTP').

- Utfør en eksplorativ analyse av dataene (bruk oppsummerende mål, histogram, bokplott, normalfordelingsplott). Beskriv fordelingen.
- Anta at observasjonene er $N(\mu_x, \sigma_x)$ -fordelte med $\sigma_x = 19$. Lag et 95% konfidensintervall for μ_x , forventningen til OC. Bruk først formler og tabeller (ikke MINITAB!), og kontroller deretter svaret ditt ved å finne 95% konfidensintervall ved bruk av MINITAB. Kommenter egnetheten av ditt valg av prosedyre i forhold til det du fant om datasettet i a).

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

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

Kommenter eventuelle forskjeller fra resultatet i (b).

Hint: Her kan du bruke kommandoen **Stat->Basic Statistics->1-Sample t** som du utfører på omtrent samme måte som **1-Sample Z**. Den eneste forskjellen er at du nå ikke behøver å spesifisere standardavviket.

Oppgave 2

Dette er en litt omskrevet versjon av oppgave 7.32 i boken.

Dersom vi øker inntaket av mat, har vekten vår en tendens til å øke. Ernæringsforskere kan beregne omtrent hvilken vektøkning en bestemt økning i kaloriinntak vil kunne medføre. I en studie spiste 16 normalvektige voksne personer mellom 25 og 36 år 1000 kilokalorier (kcal) ekstra per dag i tillegg til kaloriinntaket som var nødvendig for å holde vekten stabil. Personene opprettholdt denne dietten i 8 uker, slik at de konsumerte totalt 56000 ekstra kcal. I følge teorien vil 3500 ekstra kcal medføre en vektøkning på 0,454 kg. Denne teorien forutsier altså at hver av de 16 personene ville legge på seg $56000 \times 0,454/3500 = 7,27$ kg i løpet av de 8 ukene. Vekten til de 16 personene før ('wtb') og etter ('wta') denne 8-ukers dietten finner du i datafilen `weightgain.mtp` (tilgjengelig på bokens hjemmeside og CD-ROM).

- For hver person skal du finne vektøkningen ved å trekke vekten før fra vekten etter. Bruk **Calc->Calculator** og fyll feltet merket `Expression` med 'wta' - 'wtb'. I

feltet `Store result in variable` skriver du for eksempel C4. Lag passende grafiske fremstillinger av vektøkingsvariabelen og kommenter om t-test er passende for denne variabelen.

- b. Finn gjennomsnitt og standardavvik for vektøkningen.
- c. Vi ønsker å teste om forventet vektøkning er forskjellig fra 7,27 kg, som teorien forutsier. Formuler problemstillingen som et hypotesetestingsproblem med passende nullhypotese og alternativhypotese. Vi antar først at standardavviket til vektøkningen er $\sigma = 2,55$. Bruk resultatene fra b) og tabell (ikke MINITAB!) til å regne ut z-observatoren og finne p-verdien for hypotesetesten. Forklar hvordan p-verdien tolkes og hva resultatet på betyr. Bruk signifikansnivå 5%.
- d. Egentlig er standardavviket σ ukjent. Spesifiser antall frihetsgrader, bruk deretter resultatene fra b) og tabell (ikke MINITAB!) til å regne ut t-observatoren og finne p-verdien for hypotesetesten. Sammenlign p-verdien med den du fant i c) da standardavviket var kjent. Hva blir konklusjonen nå?
- e. Utfør nå den samme hypotesetesten som i d) ved å bruke MINITAB. Kontroller at resultatene stemmer overens med det du fant i d).
Hint: Her kan du bruke kommandoen **Stat->Basic Statistics->1-Sample t**
- f. Standardavviket σ er fortsatt ukjent. Beregn feilmarginen for forventet vektøkning μ for konfidensnivå 95%. Lag så et 95% konfidensintervall for forventet vektøkning μ og forklar hva dette konfidensintervallet betyr.
- g. Dersom vi ønsker en feilmargin på maksimalt 0.6 og antall observasjoner $n = 16$ er fast, hva er det største konfidensnivået vi kan bruke?

Oppgave 3

Gjør oppgavene 7.1, 7.3 og 7.6 om leiepriser for leiligheter. Du skal anta at standardavviket σ til månedlig leiepris er ukjent.