

Første obligatoriske oppgavesett for STK1000 høsten 2012

Oppgavesettet til første obligatoriske innlevering i STK1000 høsten 2012 består av tre oppgaver. For å løse de to første 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.

I den skriftlige besvarelsen av oppgavene skal du kort forklare hvordan de enkelte punktene er løst. Det er valgfritt om du vil skrive besvarelsen for hånd eller om du vil bruke et tekstbehandlingsprogram. Der du bruker MINITAB, må relevante utskrifter og plott settes inn i besvarelsen. Instruksjoner for utskrift fra MINITAB finner du i innføringsheftet.

Det er lov 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.

Innlevering: Besvarelsen leveres på instituttkontoret ved Matematisk Institutt i 7. etasje, Niels Henrik Abels hus (Matematikkbygningen). Se på websiden <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. **Frist for innlevering er torsdag 27. september 2012 kl 14.30.**

Oppgave 1

Det klages over at det regnet mye i Oslo i juli. I denne oppgaven skal vi se på observasjoner av daglig nedbør målt på Blindern i juli 2012 (hentet fra www.yr.no/place/Norway/Oslo/Oslo/Oslo/detailed_statistics.html). Datasettet finner du i Minitab-filen www.uio.no/studier/emner/matnat/math/STK1000/h12/nedbor.mtp.

- Lag et stilk-og-blad-plott for de 31 observasjonene av daglig nedbør i juli på Blindern for hånd. Kommenter formen på fordelingen.
- Bruk MINITAB til å lage et histogram over observasjonene. Sammenlign dette med stilk-og-blad-plottet du lagde for hånd (kommando: **Graph->Histogram**; jfr. avsnitt 8.3 i innføringsheftet).
- Etter å ha sett på fordelingen grafisk, hvilke numeriske oppsummeringer burde man bruke for å beskrive denne fordelingen? Begrunn svaret. Finn så disse numeriske oppsummeringene ved hjelp av MINITAB (kommando: **Stat -> Basic Statistics -> Display Descriptive Statistics**; jfr. avsnitt 8.1 i innføringsheftet). Velg ut det du trenger fra MINITAB-utskriften.

- d. Lag et boxplott over daglig nedbør i juli på Blindern ved hjelp av MINITAB. Forklar hva plottet viser deg. Beregn kriteriet for hvilke observasjoner som blir stjernemerket, og angi hvilke dager disse stjernemerkede observasjonene ble gjort.

Oppgave 2

I denne oppgaven skal vi se på data som er samlet inn under et innføringskurs i statistikk ved et amerikansk universitet. Studentene i kurset gjennomførte et enkelt forsøk. Hver student noterte sin høyde og vekt og målte pulsen sin (under hvile). Så kastet hver av studentene en mynt. De som fikk krone løp på stedet i ett minutt, mens de som fikk mynt ble sittende stille i ett minutt. Så målte alle pulsen en gang til (for de som satt stille, er dette en måling til av pulsen under hvile). Dataene for de 92 studentene finnes i MINITAB-filen

http://www.uio.no/studier/emner/matnat/math/STK1000/h12/puls_h12.mtw.

I datafilen er det en linje for hver av de 92 studentene, der variablene i kolonnene har følgende betydning:

- *Puls1*: Første pulsmåling (antall slag per minutt)
- *Puls2*: Andre pulsmåling (antall slag per minutt)
- *Løp*: 1=løp på stedet; 2=satt stille
- *Kjønn*: 1=mann; 2=kvinne
- *Høyde*: Høyde i meter (m)
- *Vekt*: Vekt i kg

Når du har lastet dataene inn i MINITAB, registrerer du dine egne verdier for variablene *Kjønn*, *Høyde* og *Vekt* i linje 93 i arbeidsarket. Mål pulsen din under hvile og registrer den som *Puls1*. Kast så et kronestykke. Hvis det viser krone, løper du på stedet i ett minutt. Hvis ikke sitter du stille ett minutt. Så måler du pulsen din igjen og registrer den som *Puls2* i linje 93 i arbeidsarket. Skriv også inn din verdi for *Løp* (1 hvis du løp, 2 hvis du satt stille). Med dette modifiserte datasettet:

- Lag histogram for variabelen *Puls1* og merk av din egen verdi. Lag også et kvantilplott og avgjør på bakgrunn av plottet og histogrammet om fordelingen er tilnærmet normal.
- Beregn enkle oppsummerende mål for den samme variabelen. Hvordan ligger din egen verdi i dette bildet?
- Lag et kryssplot med *Puls1* på *x*-aksen og *Puls2* på *y*-aksen, med forskjellige symboler for de som løp og de som ikke løp. Forklar hva plottet viser. Lag andre grafiske fremstillinger som viser effekten av aktivitet på *Puls2*.
- Du skal nå gjøre en regresjonsanalyse med *Vekt* som responsvariabel (*y*) og *Høyde* som forklaringsvariabel (*x*). Finn \bar{x} , s_x , \bar{y} , s_y , og r ved hjelp av MINITAB, og regn deretter ut for hånd skjæringspunkt og stigningstall for minste kvadraters regresjonslinjen. Skriv så ned ligningen for den tilpassede regresjonslinjen, og forklar hvordan den skal tolkes.

- e. Du møter en person som er 1.76 meter høy. Bruk den tilpassede regresjonslinjen til å predikere vekten til denne personen.
- f. Gjennomfør så den samme regresjonsanalysen i MINITAB, med *Vekt* som responsvariabel og *Høyde* som forklaringsvariabel. (Kommandoer: **Stat -> Regression -> Regression** og **Stat -> Regression -> Fitted Line Plot**)
- g. Lag et residualplott ved hjelp av **Stat -> Regression -> Regression**, klikk på *Graphs* og be om å få plottet residualer versus forklaringsvariabelen *Høyde*. Kommenter hvordan plottet ser ut og hvordan dette tolkes.
- h. Forklar hva R-Sq (r^2) i utskriften betyr. Hva er korrelasjonen mellom *Høyde* og *Vekt*?
- i. Gi en ny kolonne navnet *BMI*. Bruk MINITABs kalkulator til å regne ut BMI og legge resultatet i kolonnen *BMI*, ved å bruke formelen $BMI = \frac{Vekt}{Høyde^2}$ (kommando: **Calc -> Calculator**, velg *BMI* i boksen *Store result in variable*. I boksen *Expression* legger du formelen '*Vekt (kg)*'/(*Høyde(m)*'***Høyde (m)*'). Bruk så **Data-> Split Worksheet** til å lage ett datasett for de som løp, og ett for de som ikke løp. Velg bare de som løp, og tilpass en regresjonslinje til sammenhengen mellom *Puls2* (responsvariabel) og *BMI* (forklaringsvariabel) ved hjelp av **Stat -> Regression -> Fitted Line Plot**. Beskriv og forklar hva resultatene av regresjonsanalysen forteller, inkludert R-Sq (r^2).

Oppgave 3

Dette er litt modifisert versjon av oppgave 1.145 i læreboken. Lengden på svangerskap fra unnfangelse til fødsel for mennesker er normalfordelt med forventning $\mu = 266$ dager og standardavvik $\sigma = 16$ dager. Bruk tabell A i læreboken til å løse oppgavene under. Du må vise alle relevante mellomregninger samt forklare hvordan du bruker tabell A for hver av deloppgavene.

- a. Hvor stor andel av svangerskap varer mindre enn 243 dager (tilsvarer ca 8 måneder)?
- b. Hvor stor andel av svangerskap varer mellom 243 og 274 dager (tilsvarer ca mellom 8 og 9 måneder)?
- c. Hvor lenge varer de 15% lengste svangerskapene?