

STK1000-h14: Første obligatoriske oppgave

Oppgavesettet består av tre oppgaver. Deler av oppgave 1 skal du gjøre for hånd og ved hjelp av lommeregner, deler ved hjelp av MINITAB. I oppgavene 2 og 3 skal du benytte MINITAB. I forbindelse med bruk av MINITAB vil du ha nytte av notatet *Starthjelp i MINITAB*, som blir kalt "innføringsheftet" nedenfor. Det er tilgjengelig på kurshjemmesiden.

Dataene du trenger for oppgave 1 og 3 er tilgjengelige på kurshjemmesiden under "Annet" - "Data".

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 legges ved eller limes inn i besvarelsen. Instruksjoner for utskrift fra MINITAB finner du i avsnitt 10 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/admin/obligatorisk-innlevering/index.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 25. september 2014 kl 14.30

Oppgave 1

Punktene a) - d) skal besvares manuelt, dvs. for hånd eller ved hjelp av lommeregner.

I datasettene `cavendisk.txt` eller `cavendisk.MTV` er det 29 målinger av jordas tetthet.

- Del intervallet fra minste til største observasjon opp i passende delintervaller og tell opp antall observasjoner i hvert delintervall.
- Tegn for hånd et histogram for de 29 observasjonene. Kommenter histogrammets utseende.
- Finn gjennomsnitt og median i datasettet. Hva gir disse to størrelsene uttrykk for?

- d) Finn standardavvik og kvartildifferans ("interquartile range") for målingene av jordklodens tetthet. Hva gir disse to størrelsene uttrykk for?
- e) Skriv eller kopier observasjonene inn i et arbeidsark i MINITAB (jfr. avsnitt 5 i innføringsheftet), og benytt MINITAB til å lage et histogram. Kommandoen er: **Graph - Histogram** ; jfr. avsnitt 8.3 i innføringsheftet. Sammenlign med histogrammet du lagde manuelt i punkt b).
- f) Benytt MINITAB til å beregne gjennomsnitt, median, standardavvik og kvartildifferans. Her kan du bruke kommandoen: **Stat - Basic Statistics - Display Descriptive Statistics**; jfr. avsnitt 8.1 i innføringsheftet. Får du samme resultat som i punktene c) og d)? Kommenter eventuelle avvik.

Oppgave 2

I denne oppgaven skal vi se på noen data som er samlet inn under et innføringskurs i statistikk ved et amerikansk universitet. Datasettet er tilgjengelig i MINITAB som **pulse.MWK**. Du får tak i det ved først å trykke **File - Open Worksheet**. I vinduet som kommer opp trykker du på **Look in Minitab Sample Data Folder**. Rull ned til datasettet **pulse.MWK** og trykk **Open**.

De 92 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). På datafilen er det en linje for hver av de 92 studentene, mens variablene i kolonnene har følgende betydning:

Pulse1: Første pulsmåling (antall slag per minutt).

Pulse2: Andre pulsmåling (antall slag per minutt).

Ran: 1=løp på stedet; 2=satt stille

Smokes: 1= røyker, 2= ikke røyker.

Sex: 1=mann; 2=kvinne.

Height: Høyde i inches (1 inch = 2,54 cm).

Weight: Vekt i pounds (1 pound =0,454 kg).

Activity: Trening: 1= lite, 2= moderat, 3=mye.

Når du har lastet dataene inn i MINITAB, registrerer du dine egne verdier for variablene Sex , Height og Weight i linje 93 i arbeidsarket (husk å regne om til hhv. inches og pounds). 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 Ran (1 hvis du løp, 2 hvis du satt stille).

Med dette modifiserte datasettet:

- a) Beregn enkle oppsummerende mål for variablene Pulse1 og Height. Hvordan ligger dine egne verdier i dette bildet?
- b) Lag histogram og stilk-og-blad plott for de samme variablene, og merk av dine egne verdier. Stilk-og-blad-plott lages med `Graph - Stem-and-Leaf`.
- c) Lag grafiske framstillinger av hvordan variablene Pulse1 og Pulse2 avhenger av løpingen (dvs. av variabelen Ran). Kommenter plottene.
- d) Gjennomfør en regresjonsanalyse med vekt som responsvariabel og høyde som forklaringsvariabel. Forklar hva resultatene av regresjonsanalysen forteller deg. Lag et plott av observasjonene og den tilpassede regresjon-slinja. Kommandoer er her: `Stat - Regression - Regression` og `Stat - Regression - Fitted Line Plot`

Oppgave 3

Datasettet `stride.txt` eller `stride.MTW` er fremkommet ved at man har latt et antall eliteløpere løpe på tredemølle ved bestemte hastigheter ('Speed', som her betraktes som forklaringsvariabel), og målt stegfrekvensen (antall steg per sekund) ved hver hastighet. Responsvariabel er gjennomsnittlig stegfrekvens for henholdsvis kvinnelige og mannlige løpere ('Stride rate'). Menn er kodet 1 og kvinner 2.

- a) Plott først dataene for hastighet og stegfrekvens. Her lager du tre figurer: Først en der dataene plottes i samme figur, men med ulike plottesymboler for menn og kvinner, dernest en der det legges inn en felles regresjonlinje og til slutt en der det legges inn separate regresjon-slinjer. Plottene kan du lage med kommandoen `Graph - Scatterplot`. På bildet som kommer fram kan du spesifisere de tre typene av plott som er beskrevet ovenfor ved `With Groups`, `With Regression` og `With Groups and Regression` henholdsvis.
- b) Anta nå at du fikk dataene uten identifikasjon av kjønn. Beregn koeffisientene for minste kvadraters linje for alle dataene.
- c) Lag plott av residualene fra linjen i punkt b) mot rekkefølgen av observasjonene og mot hastighet. Forklar hvordan det vises i plottet at dataene kommer fra to forskjellige grupper. Kommandoene er her `Stat - Regression - Regression - Fit Regression Model`. Klikk på `Graphs` og be om å få plottet (i) residualene mot rekkefølgen av observasjonene og (ii) residualene mot hastighet.

- d) Beregn så regresjonslinjer separat for hvert kjønn. Hva blir korrelasjonen mellom hastighet og stegfrekvens i hvert av tilfellene? Hva sier disse korrelasjonene om regresjonslinjene?
- e) Lag plott av residualene fra hver av de to regresjonslinjene fra punkt d) mot hastighet. Kommenter plottene!