

## STK1000 H15: Andre obligatoriske oppgave

Oppgavesettet består av to oppgaver. Begge forutsetter bruk av MINITAB. I den forbindelse vil du ha nytte av notatet *Starthjelp i MINITAB*, som blir kalt "innføringsheftet" nedenfor. Det er tilgjengelig på kurshjemmesiden.

Datasettet du trenger for Oppgave 2 er et av de samme som i første obligatoriske oppgave og er dessuten tilgjengelig på kurshjemmesiden.

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!  
Besvarelsen skal leveres med en **egen forside** som du også finner der.

*Frist for innlevering er torsdag 29. oktober 2015 kl 14.30*

### Oppgave 1

Tenk deg at du har 50 tilfeldige utvalg, hver av størrelse 25, fra en  $N(20,5)$ -fordelt populasjon. På grunnlag av hvert av de 50 utvalgene kan du finne et 90% konfidensintervall for forventningsverdien (som vi her vet at er lik 20) ved å bruke formelen på side 359-360 i læreboka.

a) Hvor mange av de 50 konfidensintervallene venter du at vil inneholde den riktige verdien 20?

b) La  $Y$  være en tilfeldig variabel som angir hvor mange av de 50 konfidensintervallene som vil inneholde den riktige verdien 20. Hvilken fordeling har  $Y$ ? Svaret ditt må begrunnes!

Du kan bruke MINITAB til å simulere den situasjonen som er beskrevet innledningsvis i oppgaven:

Først trekker du 25x50 tilfeldige variable som er uavhengige og  $N(20,5)$ -fordelte. Kommando: `Calc - Random Data - Normal`. Generer 25 rader med data og ta vare på dem i kolonnene C1-C50. Tallene i en kolonne gir deg nå de 25 observasjonene i et tilfeldig utvalg, mens de 50 kolonnene gir deg de 50 utvalgene. Så finner du et 90% konfidensintervall for hvert av de 50 utvalgene. Kommando: `Stat - Basic Statistics - 1-Sample Z` og tast OK.

c) Utfør kommandoene ovenfor og angi de 50 konfidensintervallene du får. Hvor mange av dem inneholder den riktige forventningen 20? Kommenter resultatet i lys av punkt a).

## Oppgave 2

Vi er interesserte i å finne ut om det er forskjell mellom kjønnene når det gjelder forventet hvilepuls. Til å undersøke dette skal du bruke ditt modifiserte puls-datasett fra Oppgave 2 i det første obligatoriske oppgavesettet (se den oppgaven for detaljer). Siden vi er interesserte 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 om at pulsnivået for kvinner er forskjellig fra pulsnivået for menn, vil vi først anta at gjennomsnittet av Pulse1 for menn (Sex=1) faktisk gir den sanne forventning for menn. Anta videre at standardavviket for begge grupper er 10.0 (slag per minutt).

Formuler problemstillingen som et hypotesetestingsproblem (for gruppen av kvinner) med nullhypotese og alternativ hypotese. Bruk resultatene fra punkt a) til å beregne teststatistikken slik den er gitt på side 383 i læreboka.

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 kvinner (Sex=2). 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. Også i punktene f) og g) nedenfor bruker du den delen som svarer til kvinner.

f) Lag et konfidensintervall for forventningen til Pulse1 for kvinner. 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 kvinner 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?

Hva blir dine konklusjoner nå?

h) Antagelsen vi gjorde om at forventningen til Pulse1 for menn var kjent, 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.

Slutt