

ANDRE OBLIGATORISKE OPPGAVE
STK1000 HØSTEN 2009

INNLEVERINGSFRIST: 30. OKTOBER KL 14.30

Følgende obligatoriske innleveringsoppgave har 3 oppgaver. De delene av oppgaven som er Minitab-oppgaver som skal tekstbehandles,¹ med relevante Minitab-utskrifter limt inn i besvarelsen på en *fornuftig* måte.

For å få godkjent **må man ha over 60 % riktig** på obligen. Men, hvis oppgavebesvarelsen ikke er tilfredsstillende oversiktlig presentert, eller hvis man for eksempel har limt inn drøssevis av tall og plott fra Minitab som leseren ikke får noe ut av, kan gruppelæreren underkjenne besvarelsen. Man må da levere inn en mer oversiktlig versjon som kan godkjennes, og dette bruker opp en av de to forsøkene man har på obligen.

Husk at man bare har to sjanser på å få godkjent, og at man må ha godkjent på begge obligene for å gå opp til eksamen! Hvis man får underkjent på første forsøk får man et nytt forsøk kun hvis gruppelæreren anser den første besvarelsen som et ærlig forsøk. Det vil si at hvis man for eksempel leverer blankt (eller nesten) får man altså ikke mulighet til å levere en ny besvarelse.

Det er dessuten lov å samarbeide til vedlagt regelverk. Bruk dessuten gruppetimene godt hvis dere står fast.

Lykke til!

Steffen Grønneberg

Regelverk for obligatoriske oppgaver ved matematisk institutt. For at studenten skal kunne fremstille seg til emnets siste eksamen må alle obligatoriske oppgaver være bestått. Utover dette teller obligatoriske oppgaver ikke med i vurderingen.

Det er tillatt å bruke alle hjelpemidler. Det er lov å samarbeide, men studenten skal selv ha formulert og skrevet den besvarelsen som leveres inn, og den skal gjenspeile studentens forståelse av stoffet. Studenten kan bli bedt om å redegjøre muntlig for innholdet i den obligatoriske oppgaven.

Besvarelsen må leveres instituttets ekspedisjon innen oppgitt tidsfrist. Dette er en absolutt frist og besvarelsen tas ikke imot etter denne fristen med mindre annet er avtalt med studieadministrasjonen på forhånd.

Dersom det på grunn av sykdom eller liknende er behov for utsettelse på oppgaven, må det søkes om dette til studieadministrasjonen. Legeattest eller annen dokumentasjon på forholdet må følge søknaden. Det kan gis utsettelse med inntil én uke. Hvis

¹Her er Microsoft Word, som finnes på windowsmaskinene ved UiO et opplagt valg, men den frie office-suiten Open Office er et jevngodt valg som anbefales for hjemmearbeid. Se <http://why.openoffice.org/> For de informatikkinteresserte kan selvsagt L^AT_EX også benyttes, så lenge man kan nok til at man får besvarelsen like fin som om man hadde skrevet den i Word eller Open Office Writer.

det på grunn av langvarig sykdom ikke er mulig å levere innen denne dato, kan det i særskilte tilfeller gis fritak for obligatoriske oppgaver. Undervisningpersonell kan ikke gi utsettelse eller fritak.

Ved underkjennelse av obligatorisk oppgave, kan kandidaten hvis det er vist et reelt forsøk på å løse oppgaven, levere en ny besvarelse innen gitt frist.

Det er bare hvis du har fått godkjent alle obligatoriske oppgaver i et emne tidligere at du ikke trenger å levere om igjen. Hvis du har levert obliger (og disse har blitt godkjent), men ikke har tatt eksamen i emnet tidligere, må du vise frem dine godkjente obliger til studieadministrasjonen ved Matematisk institutt for at dette skal bli registrert.

Obligatoriske oppgaver fra andre/lignende emner godtas ikke som erstatning for innlevering av obligatoriske oppgaver.

Oppgave 1. Gjør oppgave 1 på STK1000-eksamen høst 2006 som er å finne på http://www.math.uio.no/academics/eks/ST/STK1000/STK1000_2006_2.pdf

Hint til b): En estimator \hat{p} som estimerer p er *forventningsrett*, hvis $\mu_{\hat{p}} = p$.

Oppgave 2. Vi fortsetter med datasettet til oppgave 2 i første oblig. Vi er nå 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 (se den oppgaven for detaljer). Siden vi er interessert i hvilepulsen, 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 \mapsto **Basic Statistics** \mapsto **Display Descriptive Statistics**

til å finne gjennomsnittlig hvilepuls og empirisk standardavvik separat for kvinner og for menn.

b). Lag boksplot 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 først 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 gitt først i oppgaven som et hypotesetestingsproblem med nullhypotese og alternativ hypotese. Bruk resultatene fra punkt a) til å beregne testobservatoren slik den er gitt på side 410 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 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 \mapsto **Basic Statistics** \mapsto **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 \mapsto **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 \mapsto **Basic Statistics** \mapsto **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). Følgende oppgave er frivillig, siden den krever teori fra avsnitt 7.2 i læreboken som vi nok ikke rekker igjennom før obliginnlevering! Det er nemlig et problem med analysene vi gjorde over, nemlig at antagelsen om at forventningen til hvilepuls for kvinner var kjent (og identisk med gjennomsnittet) ikke er 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** \downarrow **Basic Statistics** \downarrow **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.

Oppgave 3. Også i denne obligen er oppgave 3 et simuleringsforsøk, denne gangen for å se hvordan konfidensintervaller oppfører seg. Hovedpoenget med de følgende simuleringene er å se nærmere på hva det vil si at man har et 90% konfidensintervall. Et 90% konfidensintervall kan sees på som en maskin som fungerer i 90% av gangene *når man bruker den veldig mange ganger*. Med ”fungerer” mener vi at den klarer å inkludere den faktiske forventningen. Vi skal her bruke ”konfidensintervall-maskinen” 50 ganger, og se hvor mange som inkluderer den faktiske verdien. Dette kan vi sjekke

siden *vi simulerer datasettet selv*, og derfor kjenner den sanne verdien til forventningen. I virkeligheten vet man aldri hva den sanne forventningsverdien er. Derfor kan simuleringer være veldig illustrerende!

Tenk deg at du har et tilfeldig utvalg (SRS) av størrelse 25 fra en $N(10, 5)$ -fordelt populasjon, altså en populasjon som er normalfordelt, med forventning lik 10 og standardavvik lik 5.

a). Hvilken fordeling får gjennomsnittet i et slikt tilfeldig utvalg? Er det svaret du gir eksakt eller tilnærmet? Begrunn. Tenk deg så at du har 50 slike tilfeldige utvalg, hvert på størrelse 25. På grunnlag av hvert av de 50 utvalgene er det mulig å finne 50 stykk 90% konfidensintervall for forventningsverdien i populasjonen (som vi her vet at er lik 10) ved å bruke formelen på side 388 i læreboka (du skal ikke gjøre dette foreløpig).

b). Hvor mange av de 50 konfidensintervallene forventer du vil inneholde den riktige verdien 10?

c). La Y være en stokastisk variabel som angir hvor mange av 50 slike konfidensintervall som inneholder den riktige verdien 10. Hvilken fordeling har Y ? Svaret ditt må begrunnes skikkelig!

d). Du kan bruke Minitab til å simulere den situasjonen som er beskrevet innledningsvis i oppgaven:

- (1) Først trekker du 50x25 stokastiske variable som er uavhengige og $N(10, 5)$ -fordelte. Kommando:

Calc \mapsto **Random Data** \mapsto **Normal**.

Generer 25 rader med data og ta vare på dem i kolonnene C1-C50. Tallene i én kolonne gir deg nå de 25 observasjonene i et tilfeldig utvalg, mens de 50 kolonnene gir deg de 50 utvalgene.

- (2) Så finner du et 90% konfidensintervall for hvert av de 50 utvalgene. Kommando:

Stat \mapsto **Basic Statistics** \mapsto **1-Sample Z**.

Utfør kommandoene over, og angi de 50 konfidensintervallene du får. Hvor mange av dem inneholder den riktige forventningen 10? Kommenter resultatet i lys av punkt c).

e). Bruk de samme simulerte dataene til å finne 50 99% konfidensintervall for forventningsverdien i populasjonen. Forklar hvorfor disse intervallene blir lenger enn intervallene i d). Hvor mange av de 50 intervallene inneholder nå den riktige forventningen 10? Hvor mange hadde du forventet?