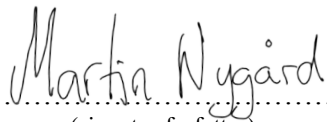




Universitetet
i Stavanger

DET TEKNISK-NATURVITENSKAPELIGE FAKULTET

MASTEROPPGAVE

Studieprogram/spesialisering: Automasjon og Signalbehandling	Vår semesteret, 2015 Åpen oppgave
Forfatter: Martin Anders Nygård	 (signatur forfatter)
Fagansvarlig: Trygve Eftestøl Veiledere: Trygve Eftestøl, Kjersti Engan	
Tittel på masteroppgaven: En analyse av pulsklokker innen NEEDED studien Engelsk tittel: An analysis of pulse watches within the NEEDED project	
Studiepoeng: 30	
Emneord: Nordsjørittet, NEEDED studien, Pulsklokke, Troponin I, Hjertesyke, Matlab, Boksploott, Signalbehandling	Sidetall: 48 + vedlegg/annet: 52 Stavanger, 13. Juni/2015



Universitetet
i Stavanger

En analyse av pulsklokker innen NEEDED-studien

Martin Nygård

Juni 2015

PROJECT / MASTER THESIS

Det teknisk-naturvitenskapelige fakultet

Universitetet i Stavanger

Veileder 1: Professor Trygve Eftestøl

Veileder 2: Professor Kjersti Engan.

Forord

Denne oppgaven markerer slutten på min tid som student ved Universitetet i Stavanger. Det har vært to spennende og lærerike år, både som student og som leder i linjeforeningen ISI (Interesseorganisasjonen for Sivilingeniørstudenter innen Informasjonsteknologi).

Jeg ønsker å takke veilederne mine, Trygve Eftestøl og Kjersti Engan for god hjelp og tilbakemeldinger gjennom hele prosjektet. Jeg ønsker å takke Stein Ørn for å ha involvert meg i en svært interessant studie som omhandler et veldig viktig tema, menneskets helse.

Jeg vil også takke mine flotte medstudenter for en fin studietid, min familie for den moralske og økonomiske støtten jeg har fått og min samboer for hjelp til rapporten.

Sammendrag

Under Nordsjørittet i 2014 ble det avholdt en helseundersøkelse med over 1000 deltagere, hvor samtlige deltagere avla blodprøve, fikk målt blodtrykk og tatt EKG. Motivet med studien var å oppnå mer medisinsk kunnskap om mosjonsidrett. Etter løpet, da resultatet av de medisinske prøvene var klare, viste det seg at 25 av deltagerne hadde hjerteproblemer. Av de over 1000 deltagerne hadde rundt 300 av disse på seg pulsklokker, hvorav åtte av dem hadde hjerteproblemer.

Selv om flere og flere mosjonister bruker pulsklokker, er målingene fra klokken svært lite involvert i medisinske undersøkelser. Derfor er det nå interessant å se nærmere på om de loggede pulsmålingene til deltagerne varierer og om det er mulig å skille mellom de hjertesyke og de friske.

Målet for denne masteroppgaven er å lese inn alle de forskjellige filtypene i Matlab, behandle og slette de som er korrupt, dårlig samplet eller ufullstendige, for så å få en felles struktur på dem. Deltagerne deles så inn i tre forskjellige grupper, de med høye blodverdier som er hjertesyke, de med høye blodverdier som er friske og de med lave blodverdier som er antatt friske. Etter dette ser man nærmere på pulsmålingene fra de forskjellige gruppene og analyserer dem mot hverandre. Først ser man på hele løpet, deretter på mindre, men samtidig mer interessante distanser i løpet. Resultatet er ganske entydig; det er forskjell mellom de hjertesyke og de friske deltagerne. For å kvalitetsikre resultatet ble pulsmålingene normalisert for å neglisjere innvirkningen av alderen på deltagerens puls. Selv etter normaliseringen er det fortsatt stor forskjell mellom gruppene, og ved hjelp av statistisk analyse er det bevist at gruppene er signifikant forskjellige.

Resultatet av studien viser at deltagerne med hjerteproblemer har lavere puls enn de andre. Dette er ikke tilfelle gjennom hele løpet, da pulsen til de hjertesyke deltagerne synker saktere enn pulsen til de friske. Dette blir synlig i de slakkere partiene og nedoverbakkene, hvor det er lavere intensitet.

Innhold

Forord	i
Sammendrag	ii
1 Introduksjon	2
2 Bakgrunnsstoff	3
2.1 Nordsjørippet	3
2.2 Needed studien	4
2.3 Medisinsk teori	6
2.3.1 CT	6
2.3.2 Troponin	6
2.4 Pulsklokkemålinger og data	6
2.4.1 Generell informasjon - Filformat	6
2.4.2 Pulsklokkefiler	7
2.4.3 Informasjon om deltagerne	10
2.5 Boksplott	11
3 Implementering	12
3.1 Innlesing av filer	12
3.1.1 Innlesningsfunksjon og Formatering	13
3.2 Filbehandling	14
3.2.1 Interpolering	14
3.2.2 Tidsformat	16
3.2.3 Sletting og fjerning av ufullstendige filer	16
3.2.4 Distanse	17
3.2.5 Integring	18
3.3 Matlab struktur	20
4 Resultater	22
4.1 Boksplott	22

4.1.1	BoksploTT for hele løpet	22
4.1.2	Analyse fra starten og slutten av løpet	24
4.1.3	Analyse fra en opp- og nedoverbakke	26
4.1.4	BoksploTT av normaliserte hjerteverdier	28
4.2	HR - inndeling	32
4.2.1	Normalisert HR - inndeling	34
4.3	Median og Kvartiler	36
4.3.1	Median og kvartiler i bakkene	38
4.3.2	Normaliserte verdier	40
4.3.3	Normaliserte verdier (kun menn)	41
4.3.4	Hjertemålingene delt på minutter	42
5	Diskusjon	43
6	Konklusjon	46
7	Referanser	47
A	Forklaring av skript	49

Figurer

2.1	Løypeprofilen til Nordsjørippet	4
2.2	Trenings-form og mengde per uke for våren 2013	5
2.3	BoksploTT	11
3.1	Eksempel på sampler	15
3.2	Resultat av interpolering	16
3.3	Flytskjema som viser rekkefølge for de forskjellige prosessene	18
3.4	Flytskjema for hvilke filer som gjenstår	19
3.5	Cellestrukturen satt opp i Matlab	21
4.1	BoksploTT for hele løpet	23
4.2	BoksploTT for start og slutt av løpet	25
4.3	Konfidensintervaller for starten og slutten i løpet	25
4.4	BoksploTT for opp- og nedoverbakke	26
4.5	Konfidensintervall for opp- og nedoverbakke	27
4.6	BoksploTT med normaliserte verdier	29
4.7	Hjerteratefordeling mellom gruppene	33
4.8	Sannsynlighetthetsfunksjon for gruppene	34
4.9	Sannsynlighetthetsfunksjon med normaliserte verdier	36
4.10	Plott av median og kvartiler	37
4.11	Plott av median og kvartiler i bakke 1	38
4.12	Plott av median og kvartiler i bakke 2	39
4.13	Plott av median og kvartiler normalisert	40
4.14	Plott av median og kvartiler normalisert 2	40
4.15	Plott av median og kvartiler (kunn menn)	41
4.16	Plott av median og kvartiler delt på minutter	42
5.1	Fordelingen av alle hjerteratemålingene for gruppene	45

Tabeller

2.1	Oversikt over filtyper benyttet i prosjektet	8
2.2	Antall deltagere i de forskjellige gruppene	10
4.1	Tabell med f- og p-verdier	24
4.2	Resultat av Anova-test	28
4.3	Gjennomsnittsalder i de forskjellige gruppene	29
4.4	Oversikt over konfidensintervall og p-verdier	31
4.5	Fordelingen av hjerterate	32
4.6	Fordelingen av normalisert hjerterate	35
4.7	Oversikt over antall deltagere som oppnår forventet makspuls	35

Kapittel 1

Introduksjon

Det er en økende bruk av pulsklokker blant mosjonister og disse lagrer store mengder data. Pulsklokkene brukes til å kontrollere og optimalisere treningen. Noe de sjelden blir brukt til er et verktøy for diagnostisering av hjertesykdommer. Det er klinisk bevist at ved fysiske belastninger er det en liten, men økt risiko for hjerteinfarkt og hjertestans. Derfor er det ønskelig å identifisere personer med økt risiko **før** en eventuell hendelse inntreffer. Det er beregnet at ett dødsfall inntreffer for hver 100.000 maraton timer. [1]

Utgangspunktet for prosjektet er at antatt friske personer deltar i de 90.6 kilometer lange offroad sykkelløpet Nordsjørittet. Mer enn 1000 presumptivt friske deltagere ble undersøkt ved hjelp av en systematisk innsamling av flere biomarkører, slik som blodprøver, blodtrykk og EKG. Undersøkelsene ble foretatt 24 timer før, samt 3 og 24 timer etter løpet, og resultatet kom som en overraskelse på enkeltindivider. Det viste seg at flere deltagerne hadde kraftig økning i blodprøveverdier, noe som kan indikere skade på hjertemuskelen. Deltagerne med høyere troponinverdier enn 195, ble utredet og det ble påvist alvorlige hjertesykdommer på enkelte av dem. I tillegg til de utførte medisinske undersøkelsene foreligger det pulsklokke-data fra cirka 300 av deltagerne. Formålet med dette prosjektet er, ved hjelp av signalbehandling, se på muligheten for utvikling av nye metoder for å analysere hjertefrekvensdata fra pulsklokker. Et ønske er da å se om det er mulig å skille deltagerne *med* hjerteproblemer fra de uten hjerteproblemer. Om dette er mulig, er det interessant å se om pulsklokke-dataene til de hjertesyke deltagerne kan klassifiseres. Med andre ord, utforske om det er en sammenheng mellom deltagere med hjerteproblemer og deres pulsklokke-data. Da kan man i fremtiden finne ut om enkeltindivid er i risikogruppen, uten medisinske undersøkelser. [2]

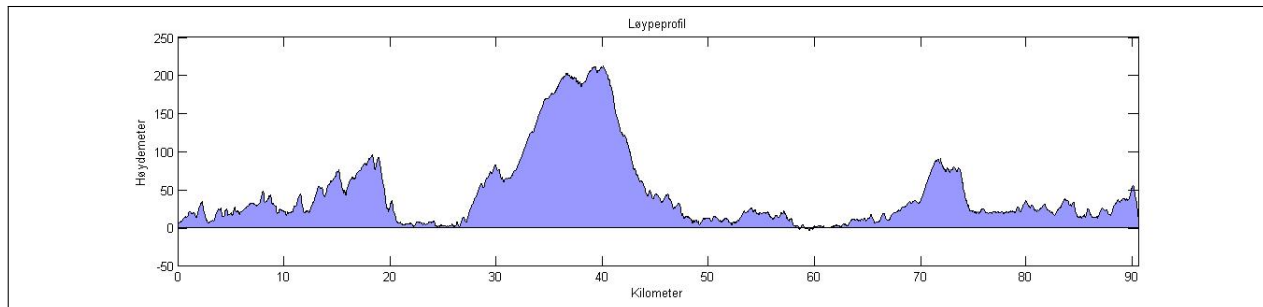
Kapittel 2

Bakgrunnsstoff

2.1 Nordsjørittet

Nordsjørittet er et offroad sykkeløp som har vært arrangert siden 1998. Det starter i Egersund og følger Nordsjøveien til Sandnes, rundt 90 kilometer unna. Underlaget varierer mellom asfalt og grus, og det er forbudt å bruke landeveissykler med bukkehorn eller tempobøyle. 40% av løypen er lagt til turstier og ulendt terreng. Derfor velger de fleste deltagerne terrengsykkel fremfor hybrid. Rittet er Norges nest største sykkelritt med 12.500 deltagere. Det er få eliteutøvere som deltar, flesteparten er ivrige mosjonister. De flinkeste og mest ambisiøse syklistene tilbakelegger distansen på under tre timer, mens gjennomsnittstiden er mellom fire og fem. Løpet er relativt hardt og mange deltagere har dårlig treningsgrunnlag og er i en alder der hjertesykdommer kan forekomme. Det har til nå vært ett tilfelle av hjertedød i forbindelse med løpet.

Løpet starter ved havnivå og ender ved havnivå, med høyeste punkt på over 200 meter. Det er i all hovedsak to bakker i løpet, hvor den første er den største og kommer etter 27 kilometer. Den andre er 80 meter og kommer etter 70 kilometer. Ellers i løpet er det varierende høyde, uten markante stigninger. Løypeprofilen til Nordsjørittet som kan sees [figur 2.1](#) er konstruert i Matlab ved hjelp av høydemålinger fra en pulsklokke. [3]



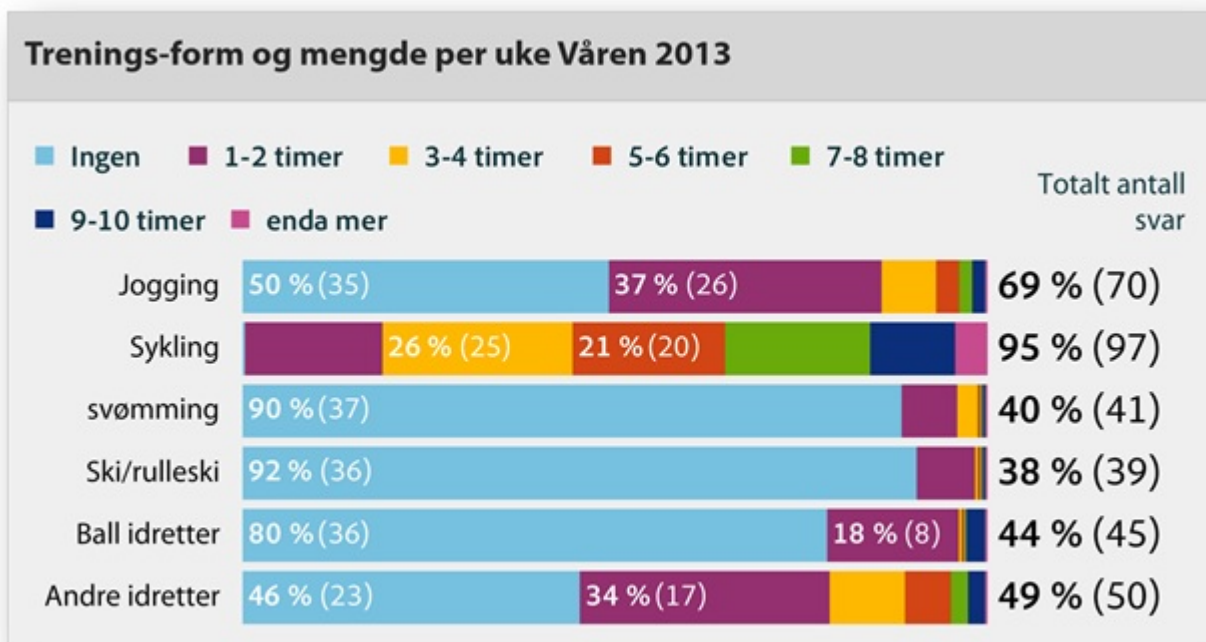
Figur 2.1: Løypeprofilen til Nordsjørippet

2.2 Needed studien

Needed er verdens største helseundersøkelse av antatt friske individer som deltar i en mosjonskonkurranse. Navnet kommer fra The North Sea Race Endurance Exercise Study. Motivet for Needed studien er å oppnå mer kunnskap om mosjonsidrett. Bak prosjektet sitter en gruppe forskere ved Stavanger Universitetsykehus i samarbeid med Nordsjørippet. Hensikten er å øke helsegevinsten for deltagerne i konkurranseidretter for mosjonister.

Første del av prosjektet ble gjennomført i 2013, hvor 97 deltagerne ble undersøkt. Resultatet viste en mulig sammenheng mellom treningstype, intensitet, effekter på kolesterol og fettomsetning, sammen med betennelsereaksjon i kroppen. Prosjektet resulterte også i funn av en blodprøve som kan avsløre trange blodårer til hjertet. Året etter, i 2014 ble over 1041 deltagerne undersøkt, hvorav 1008 av disse fullførte løpet. Dette året brukte en tredjedel av deltagerne pulsklokker. Alle undersøkte deltagerne er over 18 år, har avgitt skriftlig informert samtykke, fylt ut elektronisk spørreskjema og er villige til å møte opp for etterkontroll. Spørsmålene i spørreskjemaet skal kartlegge treningsmengden til deltagerne det siste året, hvor mange år de har trent og om de har vært med i lignende konkurranser tidligere. Resultatet av oversikten over trenings-form og mengde kan sees i [figur 2.2](#).¹

¹Figuren er hentet fra Needed protokollen, med samtykke fra Stein Ørn



Figur 2.2: Trenings-form og mengde per uke for våren 2013

Dette viser at de fleste deltagerne holder seg i god form og tar gjerne sykkelen til hjelp. Etter fullført ritt, når resultatene fra blodprøvene er klare, blir deltagere med høye troponin-verdier innkalt til nærmere medisinsk undersøkelse. Alle med klinisk mistanke eller med betydelig forhøyet TnI-nivå vil få tilbud om koronar utredning. Dette er en undersøkelse av kransarteriene ved hjelp av et røntgenapparat og kontrastmiddel. Det er etter dette en eventuell diagnostisering av en hjertesykdom finner sted. Alle etterundersøkelsene av personer med mistenkt hjertesykdom blir utført av prosjektets leger etter mal anført av Shave et al. [4] [5]

All personinformasjon er konfidensiell og all forskningsdata vil identifiseres med fødselsnummer av hensyn til innhenting av data i den planlagte oppfølgingsperioden. Startnummeret i Nordsjørittet vil være identifikasjonen i studien. Deltagerlisten med all identitetsdata og startnummer blir oppbevart atskilt fra datasettet og slettet etter fullført studieperiode. [3]

2.3 Medisinsk teori

2.3.1 CT

CT, eller computertomografi, er en radiologisk undersøkelse som gir snittbilder av kroppen. Ved CT kan det lages enten tverrsnittbilder eller 3D-bilder av ulike organer. Undersøkelsen kan brukes for å påvise forskjellige sykdommer og skader, for eksempel hjertesykdom. [6]

2.3.2 Troponin

Troponin er et protein som blir skilt ut når hjertemuskulaturen er skadet. Nivået av troponin i blodet varierer alt etter hvor stor skaden er. [7] Det er nivået av Troponin I som er interessant i denne studien, eller TnI som det ofte forkortes til. Nivået av TnI kan fastslås ved hjelp av blodprøver, og høye verdier kan gi svar på om en person har gjennomgått et akutt hjerteinfarkt eller at de er i faresonen. I tillegg må man være klar over at høye verdier også kan ha andre årsaker, som for eksempel nyresykdom, autoimmune tilstander og fysisk anstrengelse. [8] Også andre hjerteproblemer som hjertesvikt og arytmier (forstyrrelser) vil kunne føre til økte verdier av troponin, men her vil ikke det oppstå like raske endringer i verdien som ved et hjerteinfarkt. [9] Når det kommer til fysisk anstrengelse viser noen nyere studier at det ikke bare er ekstreme fysiske anstrengelser som kan gi forhøyede troponinverdier, men at også friske mennesker kan få høye verdier av normal fysisk aktivitet. Dette betyr at friske individ kan fremprovosere høye verdier ved å presse kroppen under fysisk trening. [10] Dette har da ikke nødvendigvis noen sammenheng med skade på hjertemuskulaturen og forklaringen kan være at det også finnes troponin i cytoplasma, som er substansen i cellen. [11] [12]

2.4 Pulsklokkemålinger og data

2.4.1 Generell informasjon - Filformat

Et filformat er en standard for hvordan informasjonen er kodet for lagring i en datafil. Den spesifiserer hvordan bits brukes til å kode informasjon i et digitalt lagringsmedium. Noen filtyper er designet for veldig spesifikke oppgaver, slik som PNG-filer, som kun lagrer punktgrafikk bilder ved hjelp av datakomprimering. Andre filformater kan være designet for lagring av forskjellige typer data, slik som OGG-filer. De kan fungere som en beholder for ulike typer multimedia, inkludert alle kombinasjoner av lyd og video, med eller uten tekst. Tekstfiler kan inneholde hvilke som helst tegn, inkludert kontrolltegn. Filformater slik som HTML, har en definert syntaks som kun tillater dem å bli brukt til spesifikke formål. [13]

Det er ikke alle filtyper som er leselig i standard Windows-programmer. Notisblokk og Notepad++ leser tilsammen de fleste tekstbaserte filtypene. Av erfaring, krever tekstfiler som ikke er lesbare i disse programmene mer spesielle og kanskje egne programmer for lesing av slike typer. Det er en måte for bedrifter og produsenter å prøve og kontrollere markedet på. Lager de et eget filformat til et spesifikt produkt, for eksempel en pulsklokke, ønsker de også å lage programvaren som kan lese nettopp denne filen. Samtidig som de gjør dette, kan de sikre at programvaren deres blir brukt, så lenge det er kun den som kan lese den nevnte filen.

2.4.2 Pulsklokkefiler

Til studien ble det benyttet både nye og gamle pulsklokker fra Polar og Garmin. Det ble derfor mange forskjellige filtyper, flesteparten leselige og noen totalt uleselige. I følge utviklere i Polar har noen av filtypene ikke vært i bruk i pulsklokker siden 2004, så noen av klokkene er svært gamle. I disse klokken logges kun hjerterate og et sammendrag av løpet. Det finnes ingen informasjon om samletid og hvor langt deltagerne har syklet. I nyere modeller er det gjort store fysiske forbedringer og filformatene står i stil til dette. Klokkene logger mye mer data og åpner mer for etterbehandling gjennom produsentens egen software. Heldigvis for studien var det ingen overflod av eldre pulsklokker, så det vart bare et lite antall deltagere som måtte ekskluderes på grunn av dårlige data. Totalt sett var det 304 deltagere med pulsklokker. Et vedlagt Excel-dokument med all informasjon om hver deltager, viser henholdsvis at det er kun 292 deltagere med klokker. Grunnen til manglende informasjon på 12 av deltagere vites ikke. Ingen av de 12 deltagerne er i risikogruppen eller hadde høye blodverdier, så det er ikke et stort savn av data. Som tidligere nevnt vart det mange forskjellige filtyper fra pulsklokkene. Tabellen under gir en god oversikt over filtypene, antallet og hvilken informasjon de lagrer.

Tabell med oversikt over filformat

<i>Filtype</i>	<i>Antall</i>	<i>Tid</i>	<i>Distanse</i>	<i>Fart</i>	<i>Hjerterate</i>	<i>LG</i>	<i>BG</i>	<i>Elevasjon</i>
csv	212	X	X	X	X			
fit	38	X	X	X	X	X	X	X
tcx	15	X	X		X	X	X	X
pxw	21	X	X	X	X	X	X	X
xml	4		X	X	X			X
gpx	5	X				X	X	X
wko	1				X			
hrm	6				X			
lsx	1							
opi	1							

Tabell 2.1: Oversikt over filtyper benyttet i prosjektet
LG står for Lengdegrader og BG står for Breddegrader

Selv om filformatene logger de samme dataene, er de ikke nødvendigvis i samme målenhet eller format. For eksempel, CSV-filene lagrer medgått tid i minutter, mens de andre lagrer tidsstempel i formatet "yyyy-mm-dd hh:mm:ss". All distanse lagres i meter og farten oppgis i km/t. Hjerteraten lagres i bpm² og er kun en oppdatert verdi. Den inneholder ingen informasjon om avstand mellom slagene og bærer seg på at pulsen er jevn. Den utjevnes ved å ta snittverdi over noen sekunder. [14] Lengde- og breddegrader er GPS-koordinater og oppgis i desimaltall. Elevasjonen er høydemålingen som avleses fra GPS-posisjonene eller ved hjelp av et innebygd barometer i klokken. Barometeret er mest nøyaktig, da dette er basert på atmosfærisk trykk. Elevasjonen i klokken oppgis i meter over havet. I de neste delkapitlene ser man nærmere på de forskjellige filformatene.

CSV - Comma Separated Values

All informasjon i *csv*-filer separeres med komma. Disse kan leses ut og legges i celler i Matlab.³ Linjeskiftene i filformatet tas også høyde for ved innlesning. Matlab har en egen innlesningsfunksjon for dette filformatet, *csvread*. Funksjonen henter ut all informasjon og legger den systematisk inn i cellestrukturen identisk med originalfilen. En forutsetning for at innlesningsfunksjonen skal fungere er at filen inneholder kun tall og dette er ikke tilfelle. Dataen i *csv*-filene ligger fordelt i én stor matrise med forklarende overskrifter for hver kolonne. Disse overskriftene må ekskluderes ved innlesningen av filene for å unngå feilmelding og

²beat per minute - hjerteslag i minuttet

³En celle kan inneholde alle typer data, som foreksempel tekststrenger, store matriser eller vektorer.

systemfeil. Funksjonen må derfor endres slik den ikke leser første rad i filen. Tidsformatet i CSV-filene oppgis som desimaltall i minutter og starter på 0. [15]

FIT - Flexible and Interoperable Date Transfer

Fit er ikke leselig i Matlab og er kryptisk med ukjente symboler. Av de totalt 304 tilgjengelige pulsfilene er 38 av dem *fit*-filer. Det er derfor viktig å finne en vei rundt problemet. Det finnes flere forskjellige Windows-baserte programmer og nettsider som kan konvertere filtypen *fit* til *tcx*-filer. Filene kan konverteres siden de inneholder samme type informasjon. Navnene på etiketten i filene er litt ulik, men formateringen er nå identisk og man kan nå enkelt lage en funksjon til innlesing av filene. [16]

GPX - GPS Exchange Format

Gpx er et XML databaseskjema designet for GPS posisjonering i pulsklokker og andre håndholdte enheter. Formatet ble lansert i 2002, åpent og tilgjengelig for alle uten krav til lisens. Som man ser i tabellen ovenfor lagrer *gpx* tid, hjerterate, GPS-koordinater og elevasjon. [16]

TCX - Training Center XML

Tcx er et format utviklet for datautveksling mellom Garmin pulsklokker og pcer. *Tcx* er en videreutvikling av *gpx*-formatet, hvor de nå også logger distansen. Formateringen av filene er veldig lik, men i *tcx*-formatet har de valgt å bruke GPS-posisjonene for å beregne tilbakelagt distanse. Man kunne enkelt regnet ut farten på deltageren, men på grunn av kvalitetsforskjeller på produktene, er ikke alle klokkene utstyrt likt. Tidsformatet i *tcx*-filene er det samme som i *fit*, et tidsstempel med dato og klokkeslett. [17]

WKO

Wko er formatert på annen måte enn de andre filtypene, hjerterefrekvensen er sporadisk og så lavt samplet at dataene blir for unøyaktige. Kombinasjonen av dette, få filer og mangel på logging av tid og distanse gjør at dataene ikke kan brukes i dette prosjektet. Det er fortsatt mulig å lage en funksjon til å lese inn filen, men utbyttet blir for lavt.

2.4.3 Informasjon om deltagerne

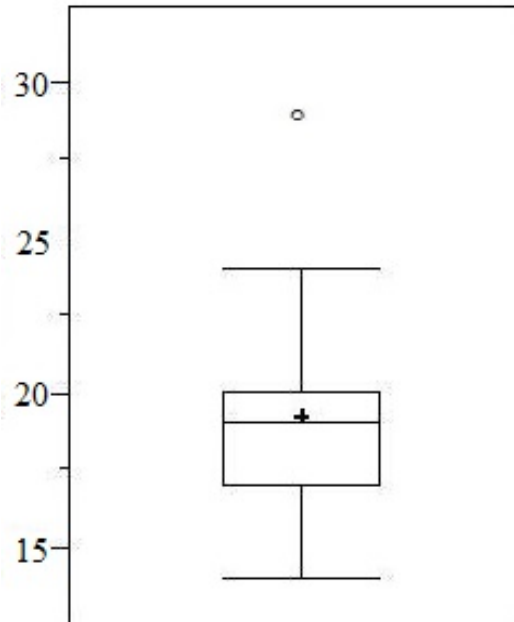
I tillegg til dataene fra pulsklokkene følger det med et regneark med mer detaljert informasjon om hver deltager. I regnearket finner man deltagerens informasjon, sykketid, TnI-verdier, kjønn, anmerkninger fra CT, tiltak og alder. Identifikasjonen som er oppgitt i regnearket er anonymisert og de ansvarlige for studien sitter med krypteringnøkkelen. Analyserer man informasjonen under kolonnen "anmerkninger fra CT", kan deltagerne i studien deles inn i tre forskjellige grupper. En gruppe med deltagerer uten anmerking, en med anmerkningen "Normale forhold" og den siste med alle mulige anmerkninger, unntatt de med "Normale forhold" Grunnen til at noen deltagerer ikke har anmerkninger, er fordi de ikke trengte noen medisinsk etterbehandling. Disse deltagerne danner gruppen med lave TnI-verdier. Deltagerne med anmerkningen "Normale forhold" ble undersøkt etter løpet på grunn av høye TnI-verdier, men ble utskrevet fordi alt så normalt ut. Den siste gruppen, deltagerne med diverse anmerkninger, blir ansett som hjertesyke. Ved etterbehandling og CT-skanning av hjertet, ble det stilt diagnoser til disse deltagerne, som nå utgjør anmerkningene i regnearket. Av de totalt 292 deltagerne med klokker ser fordelingen mellom gruppene slik ut.

Gruppe	Høy troponin <i>med</i> problemer	Høy troponin <i>uten</i> problemer	Lav troponin	Totalt
Antall	8	22	262	292

Tabell 2.2: Antall deltagerer i de forskjellige gruppene

2.5 BoksploTT

BoksploTT kan brukes til presentasjon av store mengder data og illustrere spredningen. Plottet får navnet sitt fra boksen som kan sees i figuren til høyre. Området fra den nederste linjen og ned viser fordelingen for 25 % av dataen, den midterste streken og ned viser 50 % og den øverste streken og ned viser 75 %. Det vil si at 50 % av all dataen ligger inne boksen. De vertikale linjene som avsluttes av horisontale linjer på utsiden boksen kalles børster. Disse viser spredningen for store deler av de resterende dataen, uten at de mest ytterliggående punktene tas med. Sirklene "o" som ligger enda lenger ute viser fordelingen av resten av dataen. I denne figuren er det bare et datapunkt på utsiden av disse børstene. Gjennomsnittsverdien til alle datapunktene er illustrert med "+". [18]



Figur 2.3: BoksploTT

Kapittel 3

Implementering

Deltagerne deles inn i tre grupper, de med lav troponin, de med høy troponin *uten* hjerteproblemer og de med høy troponin *med* hjerteproblemer. På grunn av en lavere populasjon i de to gruppene med høy troponin, må disse tas godt vare på. Dette er fordi de skal gi bredest mulig bidrag til studien. I dette kapittelet ser man nærmere på innlesning og behandling av rådataene fra pulsklokkene.

3.1 Innlesing av filer i Matlab

En viktig del av oppgaven er å lese inn de forskjellige filene. Med totalt 292 tilgjengelige filer, representert med ti forskjellige filtyper er innlesingen svært sentral i dette prosjektet. Her er det viktig å bruke tid, slik at all nødvendig informasjon hentes ut fra de aktuelle filene. Selve innlesingen av filene tar lang tid med en tidsbruk på nærmere 10 minutter. Dette må heldigvis kun gjøres en gang. Etter den første innlesingen kan man benytte seg av en *save*-funksjon i Matlab som gjør det enkelt å lagre dataene i en *.mat*-fil. Disse kan enkelt lastes inn ved oppstart av programmet og er veldig tidsbesparende.

Det er viktig å lage Matlab-funksjonene på en slik måte at man ikke trenger noen form for bearbeiding av dataene i forkant. Filene skal forbli uberørte og det er originalfilene som skal leses inn. Om man får tilgang til enda flere filer i etterkant, skal disse enkelt kunne leses inn ved hjelp av de selvkomponerte funksjonene.

Matlab har ferdige funksjoner for innlesning av CSV og GPX-filer. 2/3 av alle pulsfilene er i CSV-format og cellestrukturen til disse filene er derfor lett å velge som oppsett. GPX-filene mangler hjerterate eller er så dårlig samplet at de ikke tas med. For de andre filtypene må det lages egne innlesningsfunksjoner.

3.1.1 Innlesningsfunksjon og Formatering

Formateringen på flere av filene er som tidligere nevnt svært like. Navnene på etikettene og tidsformatet er det eneste som varierer mellom filtypene. En etikett er en tekst som ligger mellom ulikhetstegnene «"og »", slik som <Trackpoint>. For å hente ut informasjon fra filene brukes noen Matlab kommandoer. Først henter man navnet på folderen hvor filene ligger plassert. Deretter henter man navn og filtype på alle filene i mappen. Filnavnene kan så leses inn og behandles slik at man kun står igjen med ".*filtype*". Man kan da skille filene fra hverandre og legge dem i forskjellige grupper, avhengig av filtype. Dette gjør man siden formateringen i filtypene varierer og de må sendes til forskjellige innlesningsfunksjoner. Når gruppene sorteres, lagres de med fullt navn og peker til folderen. Etter dette starter selve innlesningen av informasjonen fra filene.

Det er kun én innlesningsfunksjon som kjøres om gangen. Den første i listen er innlesningen av *csv*-filene. Dette utgjør to tredjedeler av filene, men til tross for dette er det denne funksjonen som bruker minst tid. Dette skyldes formateringen i filene. Som tidligere nevnt mangler Matlab flere innlesningsfunksjoner for de andre filene.

Grunnprinsippet bak innlesningen av de andre filtypene er svært like. Først bruker man en kommando til å åpne selve filen i Matlab. Når en fil er åpnet bruker man en funksjon til å gå fra linje til linje og lete etter etikettene man søker etter i filen. Et eksempel på hvilke etiketter og hvordan formateringen ser ut for disse filtypene, kan sees i boksen 3.1.1 under. For hente ut informasjonen mellom etikettene må man bruke uttrykket `<(\w+).*>(</\1>` . Det er informasjonen mellom etikettene som er interessant og må hentes ut. For å finne de riktige etikettene må disse defineres øverst i skriptet. Man bruker så en likhetsfunksjon for å se etter match. Utgangen til denne funksjonen er en boolean variabel, som returnerer tallverdien '1' om etiketten og navnet man søker etter stemmer overens. Da hentes informasjonen som ligger mellom etikettene ut og legges i lister. Når man er ferdig med en fil settes listene sammen i større matriser.

I boksen under ser man formateringen til en *fit*-fil konvertert til *tcx*. All informasjon mellom hvert <Trackpoint> hentes ut og lagres og dette skal sikre at vektorene blir like lange. Det er derimot to filer som er inkonsekvent med måten de logger dataen på og dette skyldes utgående pulsklokkemodeller. Her varierer informasjonen som logges mellom hvert <Trackpoint> og vektorene blir av ulik lengde og kan ikke settes sammen i matriser.

```

<Trackpoint>
  <Time>2014-06-14T09:27:16Z</Time>
  <Position>
    <LatitudeDegrees>58.4662186</LatitudeDegrees>
    <LongitudeDegrees>5.9851961</LongitudeDegrees>
  </Position>
  <AltitudeMeters>18.4000000</AltitudeMeters>
  <DistanceMeters>3180.7300000</DistanceMeters>
  <HeartRateBpm>
    <Value>130</Value>
  </HeartRateBpm>
  <Cadence>0</Cadence>
  <Extensions><TPXxmlns="http://www.garmin.com/xmlschemas/ActivityExtension/v2">
    <Speed>6.6190000</Speed></TPX></Extensions>
</Trackpoint>

```

Fra [Tabell 2.1](#) kan man se at filtypene logger forskjellig informasjon. Når disse lastes inn i Matlab er all informasjon utenom "cadence" relevant. Matrisene settes derfor opp på samme måte som i [Tabell 2.1](#), med størrelse $n \times 7$. For at vektorene skal kunne settes sammen til matriser må filtyper som ikke logger all informasjon fylles ut med tomme vektorer. Siden all informasjon i de andre vektorene er av typen *double* bør de tomme vektorene som fylles ut også være det. Et alternativ her er å bruke NaN. Skal man regne ut gjennomsnittlengden av løpene og variansen ser man da bare vekk i fra matriser som inneholder NaN.

3.2 Behandling av filer

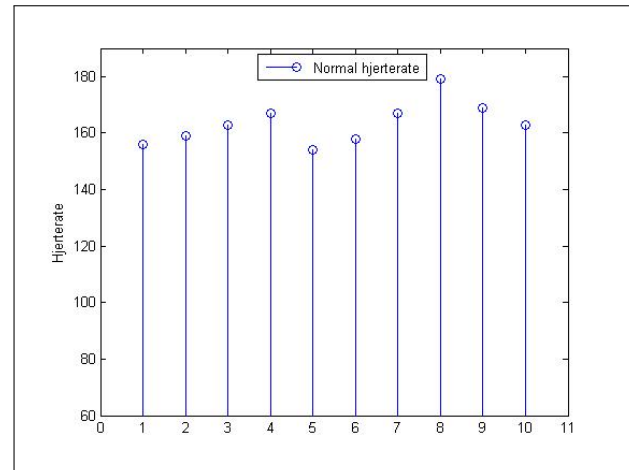
På bakgrunn av at det er benyttet mange forskjellige pulsklokker, nye, gamle og fra forskjellige produsenter, varierer innholdet mye. Dette er alt ifra hvilken informasjon som blir lagret, hvordan den lagres, sampleraten, filformatet og formateringen i selve filen. For at prosjektet best mulig skal la seg gjennomføre må alle filene, uavhengig av sin tidligere form settes sammen og formateres likt.

3.2.1 Interpolering

Når man bruker pulsklokker må de samtidig ha på seg et pulselte som detekterer puls. Dette festes rundt livet og skal være på under hele treningsøkten. Hvis beltet ikke festes riktig og justeres slik det står anført i bruksanvisningen vil klokken slite med å logge korrekte data. Hvis beltet ikke er i kontakt med kroppen på en stund, resulterer dette i at pulsklokken logger 0 i hjerterate. Dette ødelegger for videre arbeid med filene, så enten må filene slettes eller behandles. Det er totalt sett 55 filer hvor dette er tilfelle. Det er ikke så viktig å interpolere alle

disse filene, kun de filene som tilhører deltagere med høye troponin-verdier. [19] Det er flere forskjellige metoder å jobbe rundt dette problemet på. Et alternativ er å slette disse samplene, slette hele filen eller interpolere. Det er viktig å ta vare på flest mulig filer, så de skal ikke slettes. Siden samplingen er lav på flere av filene er det ikke ønskelig å fjerne flere sampler og interpolering er derfor det beste alternativet.

Der er flere forskjellige måter å interpolere på, men en teknikk som gir gode resultater er å bruke en formel som vekter de nærliggende hjerteratene. Denne ble sammenlignet med lineær interpolering og en metode som kun tar høyde for den siste og første gyldige loggede hjerteraten. I figuren til høyre ser man et eksempel med ti sampler av hjerterate. Dette er fiktive tall, i virkeligheten varierer de ikke like mye, men tallene velges slik for at resultatet av interpoleringen skal bli tydelig.

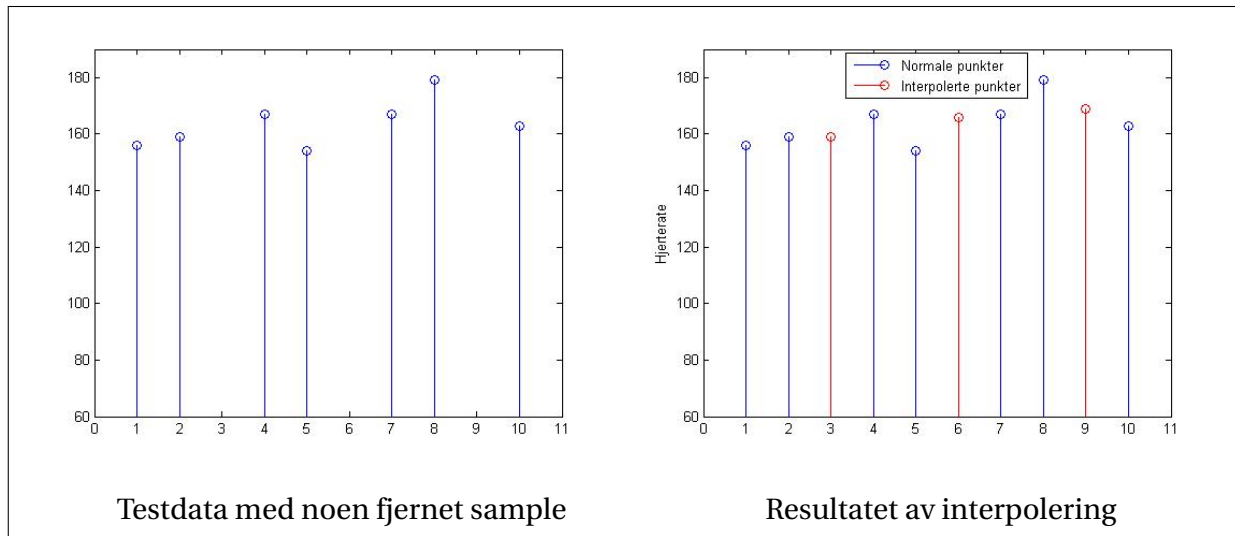


Figur 3.1: Eksempel på sampler

For å vise hvordan interpoleringen fungerer i praksis, må noen sampler fjernes fra figur 3.1 og resultatet kan sees til venstre i figur 3.2. Formelen vekter de nærliggende hjerteratene på en slik måte at de nærmeste vektet mest, se ligning 3.1. Her er x funksjonen, n er nummeret på sampelet og desimaltallene angir vektingen.

$$x_n = 0.1 \cdot x_{n-3} + 0.15 \cdot x_{n-2} + 0.25 \cdot x_{n-1} + 0.25 \cdot x_{n+1} + 0.15 \cdot x_{n+2} + 0.1 \cdot x_{n+3} \quad (3.1)$$

Resultatet av interpoleringen er illustrert til høyre i figur 3.2 og de interpolerte samplene representeres i rødt. Det er litt forskjell mellom de interpolerte og de originale samplene, men i denne oppgaven gir interpoleringen bedre resultat. Dette skyldes at variansen i hjerterefrekvensen til deltagerne er svært liten.



Figur 3.2: Resultat av interpolering

3.2.2 Tidsformat

Tidsformatet i de forskjellige filtypene varierer mye og må konverteres for å bli like. Tidene fra csv-filene oppgis i minutter og multipliseres med 60, mens de andre må behandles i Matlab. Fra boksen 3.1.1 hentes tidsformateringen `<Time>2014-06-14T09:27:16Z</Time>`. Først fjerner man `<Time>`, så må dato og klokkeslett konverteres til sekunder. Da fjerner man bokstavene "T" og "Z" for å kun stå igjen med tall. Deretter bruker man Matlab til å kalkulere forløpt tid mellom hvert trackpoint. Det er viktig å ta vare på det første tidsstempelet, da sampletiden varierer og man ikke kan ta noen antagelser. Resultatet blir da oppgitt i sekunder og erstatter tidsstemplene i listene.

3.2.3 Sletting og fjerning av ufullstendige filer

Det er viktig å slette færrest mulig filer av de deltagerne med høy troponin, som totalt sett er en gruppe på 30 deltagere. For at resultatet skal bli best mulig må denne gruppen tas mest vare på. Det er ingen grunn til å slette unødvendig mye av filene med lav troponin, men på grunn av dårlig sampling, logging av data, kombinert med et høyt antall filer kan disse lettere velges bort. Det er totalt 228 deltagere med lav troponin, så det er muligheter å klippe bort noen av de dårligste filene.

Enkelte filtyper klippes fullstendig bort på grunn av forskjellige årsaker. GPX-filene lagrer kun GPS-posisjon, elevasjon og tid. Et fåtall av disse lagrer hjerterate, da med en sampling så lav at den ikke kan brukes i dette prosjektet og er derfor ikke relevant for oppgaven. WKO-filene lagrer kun hjerterate, ingen tid eller distanse. HRM-filene lagrer kun et sammendrag av løpet og

logger hjerteraten. Den ene personen med hjerteproblemer har brukt en eldre Polar klokke som bruker HRM-formatet. Siden det totalt sett bare er åtte personer med hjerteproblemer er det viktig at alle disse tas med. Ved hjelp av sammendraget kan man lese ut antall sample og tid brukt på løpet. Sampletiden kan da kalkuleres og legges inn. LSX & OPI-filene lagrer kun kryptisk og uleselig informasjon og tas ikke med.

Andre filer som velges vekk er de enkeltfilene som logger data sporadisk. Dette gjør at vektorlengdene i cellene varierer og ikke kan settes sammen til matriser. Ifølge Polar er det en feil ved pulsklokkene som gjør at dette problemet kan oppstå. Dette gjelder kun to filer, og ingen av disse er deltagerer med høy troponin.

Filer som logger 0 i hjerterate over lengre tid og ikke plukkes opp av interpolering metoden må fjernes. Dette er filer som har logget syv sammenhengende 0-ere. Dette er ikke nødvendigvis et problem i seg selv, men i de filene dette oppstår hender dette flere ganger. Det er tolv slike filer. Et alternativ er å kun slette samplene, men siden problemet oppstår hyppig i disse filene, blir samplingen svært dårlig.

3.2.4 Distanse

Et problem er syklet avstand. Ingen av deltagerne har syklet like langt og den nøyaktige avstanden på rittet oppgis å være på mellom 90 og 91 kilometer. Etter å ha vært i kontakt med arrangøren for løpet, oppgis denne å være 90.6 kilometer. Hvis man skal kunne se på hvordan pulsen til deltagerne varierer i oppoverbakker og i perioder av løpet, må avstandsmålingene være presise. Derfor må en terskel settes for å fjerne de deltagerne som har syklet altfor kort og langt. Deltagerne med høy troponin behandles manuelt og de med lav troponin behandles automatisk ved hjelp av funksjoner. Det er derfor smart å se på stigningstallet for distansen. Når stigningstallet er lavt, er ikke rittet startet skikkelig og nullpunktet for starten settes ikke før stigningstallet stiger. For valg av terskel kan det være greit å se litt på hva som er gjennomsnittsverdi og median for avstand syklet. Avstandene under er oppgitt i meter.

Gjennomsnittsverdi:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i = 90.472$$

Varians:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N |X_i - \mu|^2 = 105.872$$

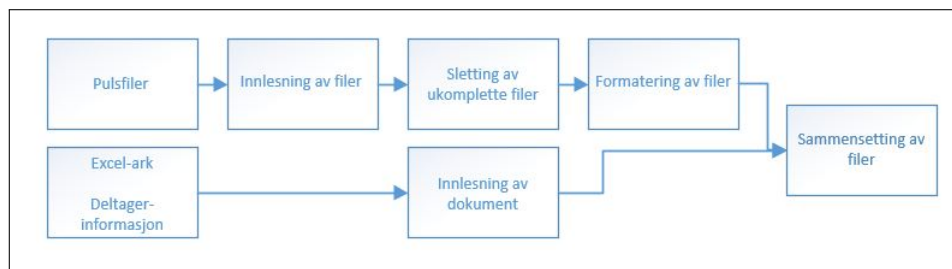
Median:

90.502

For å fjerne deltagere med høy troponin velges ett intervall mellom $89000 < X < 91000$. Dette er en automatisk metode, men selve intervallet er bestemt manuelt basert på observasjoner. Metoden sjekker om den tilbakelagte distansen for enhver deltager er mellom intervallet. Faller den på utsiden fjernes deltageren. Slik fortsetter det til alle deltagerne er sjekket. Det er også en lignende metode for deltagerne med lav troponin. På grunn av en større, men samtidig mindre viktig gruppe, settes intervallet her mye smalere til $89600 < X < 90600$.

3.2.5 Integrering

Det er ikke før alle filene er lest inn og behandlet at man vet hvor mange deltagere man står igjen med. Behandlingsprosessen beskrevet ovenfor er svært nødvendig for det videre arbeidet i prosjektet. Nå er alle filene satt opp med lik struktur og med samme målenheter, uavhengig av deres opprinnelige format. Alle filene settes så sammen i én celle og sorteres. Nå settes deltagerne og pulsklokke-dataene deres sammen med deres egen personlige informasjon fra regnearket. For å få til denne sammenslåingen, må identifikasjonen til hver deltager lagres ved innlesningen av pulsfilene og ved innlesningen fra regnearket. Det er denne identifikasjonen som testes mot hverandre, og ved treff settes de sammen. Rekkefølgen for hvordan de settes sammen ser man i flytskjemaet under.

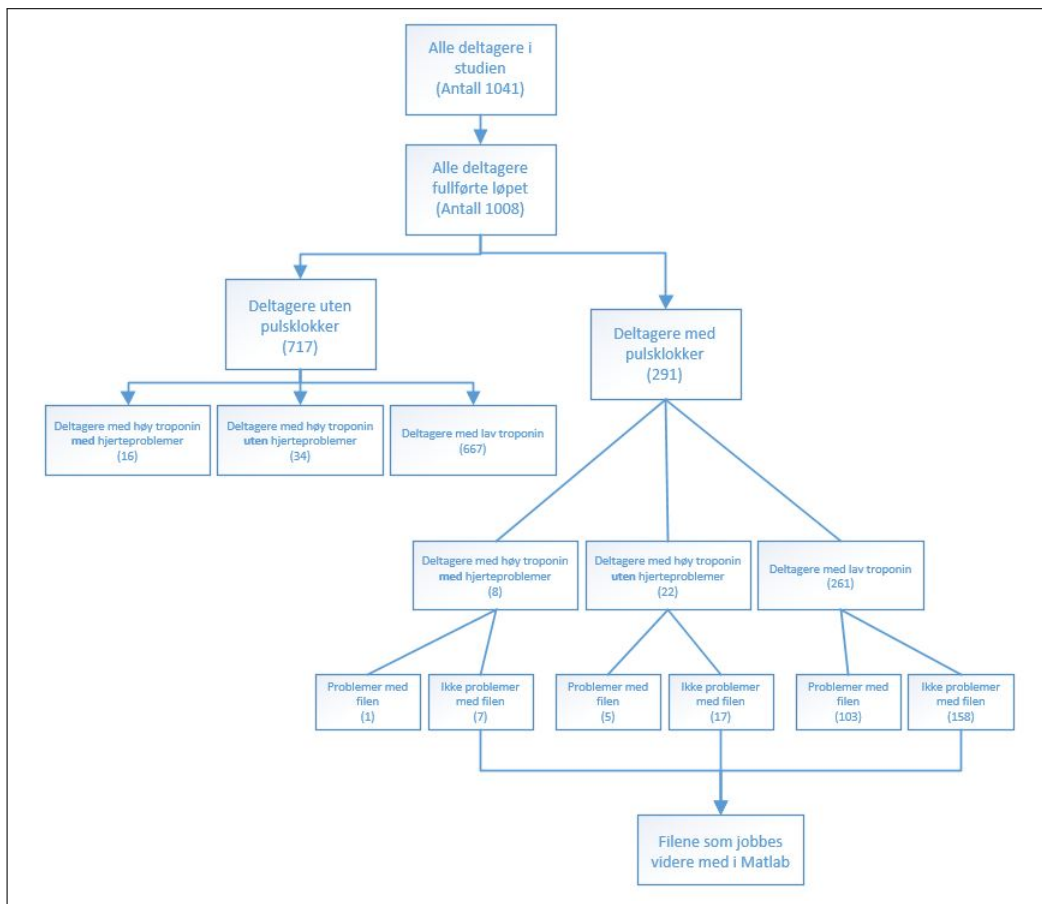


Figur 3.3: Flytskjema som viser rekkefølge for de forskjellige prosessene

Etter dette er det nødvendig med en omstrukturering i cellen, slik de hjertesyke deltagerne havner øverst i cellen. Under dem kommer deltagerne med høye TnI-verdier uten hjerteproblemer, og på bunn, deltagerne med lave verdier. For å skille de enda mer fra hverandre blir alle deltagerne i cellen nummererte. Dette betyr at de som har tilsvarende tall hører til i samme gruppe.

Etter filtreringen og fjerningen av deltagere med korrupte data, feillogginger eller annet, er det en del data som har falt bort. For studien er dette ikke noe stort problem, da de to viktigste gruppene fortsatt er godt representert. I flytdiagrammet under ser man fordelingen av deltagere i gruppene og nøyaktige tall på hvor mange deltagere som er fjernet og hvor mange som gjenstår etter den krevende filtreringsprosessen.

Flytdiagram som viser gruppeinndelingen i Matlab



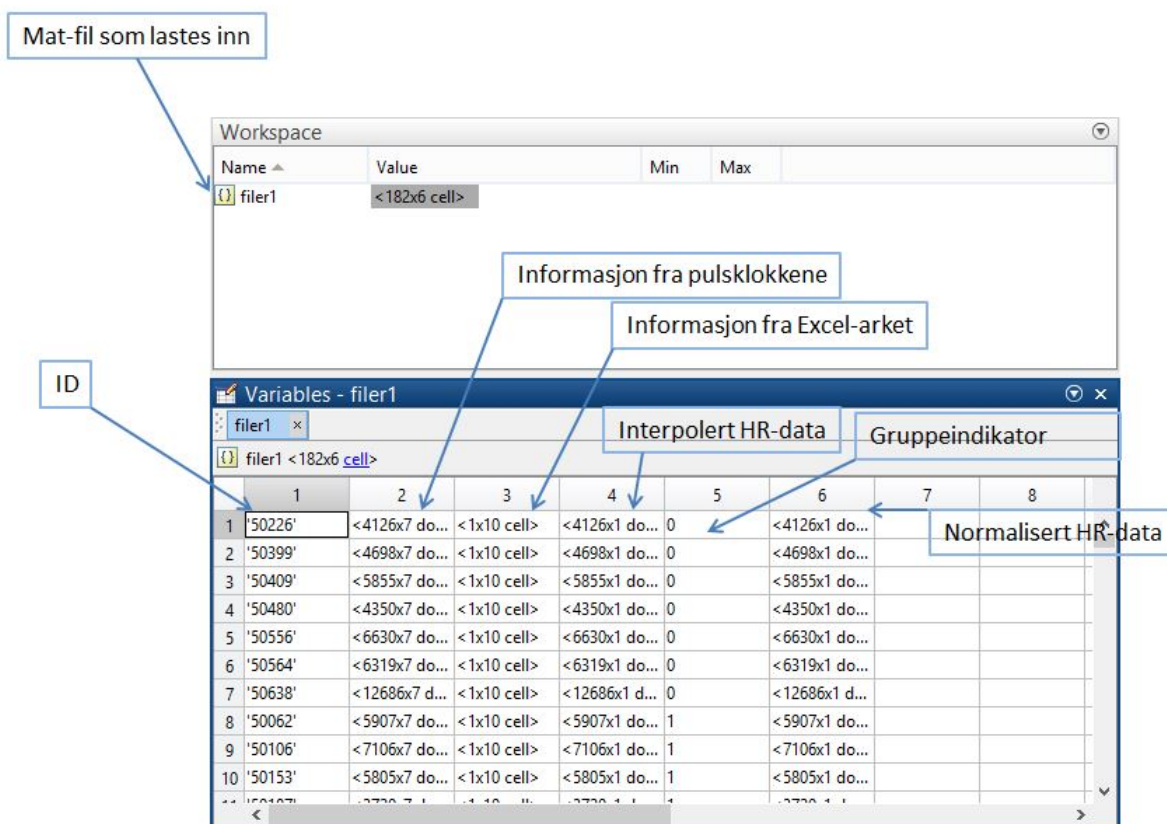
Figur 3.4: Flytskjema for hvilke filer som gjenstår

Antall deltagere med pulsklokker og fordelingen av dem er illustrert i de fire blokkene øverst til høyre i flytdiagrammet. I blokkene under er det kvaliteten på pulsfilene som avgjør om de tas med videre. Resultatet er 182 deltagere med gode pulsmålinger. Fordelingen av deltagerne i de fire boksene til venstre er manuelt utfylt etter informasjon fra de ansvarlige bak Needed-studien. Etter at deltagerne med dårlige pulsdata er valgt bort, står man igjen med nesten to tredjedeler av deltagerne med pulsklokke. Om ønskelig kunne denne gruppen vært enda større. Man må da bruke enda mer tid på analysering av start- og stoppunkt til hver deltager og velge

større intervaller for hvilke distanser tilbakelagt i løpet som aksepteres.

3.3 Matlab struktur

Flere av innlesningsfunksjonene i Matlab er tidskrevende og bruker mye datakraft når de kjører. Dette gjelder også etterbehandlingen som må gjøres i hver fil. Det er derfor ønskelig å redusere antall ganger dette må gjøres, og et alternativ er da som tidligere forklart, å lagre alt i en *.mat*-fil. Selv om filen nå inneholder mye informasjon, tar det mellom ett og to sekunder å laste den inn. Det er ikke bare innlesning og etterbehandlingen som tar mye tid, pulsfilene må også kobles sammen med riktig informasjon fra regnearket. For å skille gruppene fra hverandre fikk de også hvert sitt tall som gruppeindikator. I denne studien er det interessant å se på normaliserte hjerteverdier, så disse lages én gang og settes sammen med resten. Det er viktig å sette de sammen og ikke overskrive de, dette for å bevare de originale hjerteratene. Av samme grunn blir de interpolerte hjerteratene lagt til for å ikke erstatte originaldataene. All denne informasjonen legges til i en og samme celle. Derfor ble ordet cellestruktur nevnt i kapittel 2, som et oppsett av hvordan informasjonen ligger lagret. Det er nemlig viktig at man har en god struktur som man kjenner igjen og er lett anvendelig. I [figur 3.5](#) ser man cellestrukturen som er benyttet i dette prosjektet.



Figur 3.5: Cellestrukturen satt opp i Matlab

I toppen av figuren ser man mat-filen, filer1. Cellestrukturen i denne filen er det man ser underst i figuren. I kolonnen til venstre ser man den anonymiserte identifikasjonen på hver deltager. All gjeldende data for hver deltager ligger i samme rad, til høyre for ID-en. I kolonne nummer to ligger all data innlest fra pulsklokkene. Dette er matriser som er satt opp på samme måte som i [tabell 2.1](#). Kolonne tre inneholder all informasjon om deltagerne innlest fra regnearket. De interpolerte hjerteratene finner man i kolonne fire, mens gruppeindikatoren er i kolonne fem og de normaliserte hjerteratene ligger i kolonne seks. Flere av de selvlagede funksjonene for analysing av hjerteratene krever en input med en identisk cellestruktur som er forklart her. Funksjonene baserer seg på å hente ut alle hjertermålingene fra enten kolonne nummer fire eller seks. Disse funksjonene krever også en ekstra parameter, hvor man kan velge mellom normaliserte eller vanlige interpolerte hjertermålinger. Når man har kommet så langt at man har laget denne cellen på nytt, eller lastet inn mat-filen som medfølger kan man begynne analysen av målingene og se på resultatene.

Kapittel 4

Resultater

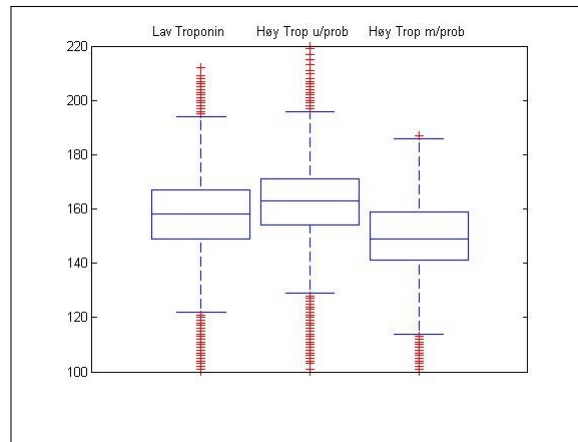
I dette kapitlet ser man nærmere på hvilke metoder og funksjoner som er produsert i Matlab for analyse av pulsmålingene. Før man ser på resultatene av hver metode, får man en liten innføring i tanken bak hver metode og hvordan man bruker dem. Når resultatene fremstilles medfølger informasjon om hvilke parametre som er benyttet. Det vil også gis en personlig tolkning av resultatet veid opp mot forventet resultat, da begrunnet i lignende studier.

4.1 Boksplott

Boksplot benyttes for å skape et overblikk over spredningen av hjerteraten til deltagerne. Dette kan gi en rask indikasjon på om de forskjellige gruppene skiller seg fra hverandre. Som tidligere forklart ser man medianen, øvre og nedre kvartil og ytterpunkter fra deltagerne. Slike plot kan genereres for hele løpet eller enkelte distanser.

4.1.1 Boksplott for hele løpet

I figuren under ser man et boksplott generert ut fra alle loggede hjerterateverdier for de tre gruppene gjennom hele løpet.



Figur 4.1: Boksplottet viser hjerteratefordelingen på de tre forskjellige gruppene over hele løpet

Som man ser i figuren over er det markante forskjeller mellom de tre gruppene. Medianen til deltagerne med hjerteproblemer indikerer at de har 9 til 14 slag mindre i minuttet enn de hjertefriske deltagerne. Ved hjelp av statistiske analyser kan man se nærmere på den virkelige ulikheten mellom gruppene. Det finnes i hovedsak to forskjellige tester som kan benyttes, parametrisk- eller ikke-parametrisk metode. Et eksempel på en parametrisk metode er Anova, mens Kruskal Wallis er en ikke-parametrisk metode. Forutsetningen for valg av parameterisk metode er at dataene i gruppene må være normalfordelte, variasjonsbredden er omtrent like stor og medianene ligger i variasjonsområdet.[20] I boksplottet ser man at alt bortsett fra normalfordelingen stemmer. Ved hjelp av Matlab kan man kjøre en normalfordelingstest på datasettene. Alle tre gruppene bestod denne, dette til tross for at fordelingen i hjerteraten til gruppen med de hjertesyke var bimodal, se figur 4.8. Dette er ikke et problem, da sentralgrenseteoremet sier at et resultat av Anova er gyldig uansett hvilken fordeling enkeltdataene kommer fra, bare de kan betraktes som identisk fordelte tilfeldige variable.[20] Resultatet av Anova-analysen gir god grunn til å forkaste nullhypotesen¹ på grunn av en høy f- og lav p-verdi. F-verdien er et mål på hvor stor forskjell det er mellom gruppene. Denne kalkuleres ved å dele variansen mellom gruppene på variansen innad i gruppen. Så lenge man analyserer datasett som er større enn 10, vil f-verdier over 5 bety at det er en statistisk signifikans i dataene. P-verdien er et tall mellom 0 - 1, og er et mål på hvor like dataene mellom gruppene er. Punktene under avgjør om gruppene klassifiseres som like eller ulike.

- $p\text{-verdi} > 0.1 \Rightarrow$ Ikke signifikant
- $p\text{-verdi} \leq 0.1 \Rightarrow$ Marginalt signifikant
- $p\text{-verdi} \leq 0.05 \Rightarrow$ Signifikant

¹Forkaster man en nullhypotese, sannsynliggjør man at en alternativ hypotese er sann

- $p\text{-verdi} \leq 0.01 \Rightarrow$ Veldig signifikant

I [tabell 4.1](#) kan man se signifikansen mellom de tre gruppene.

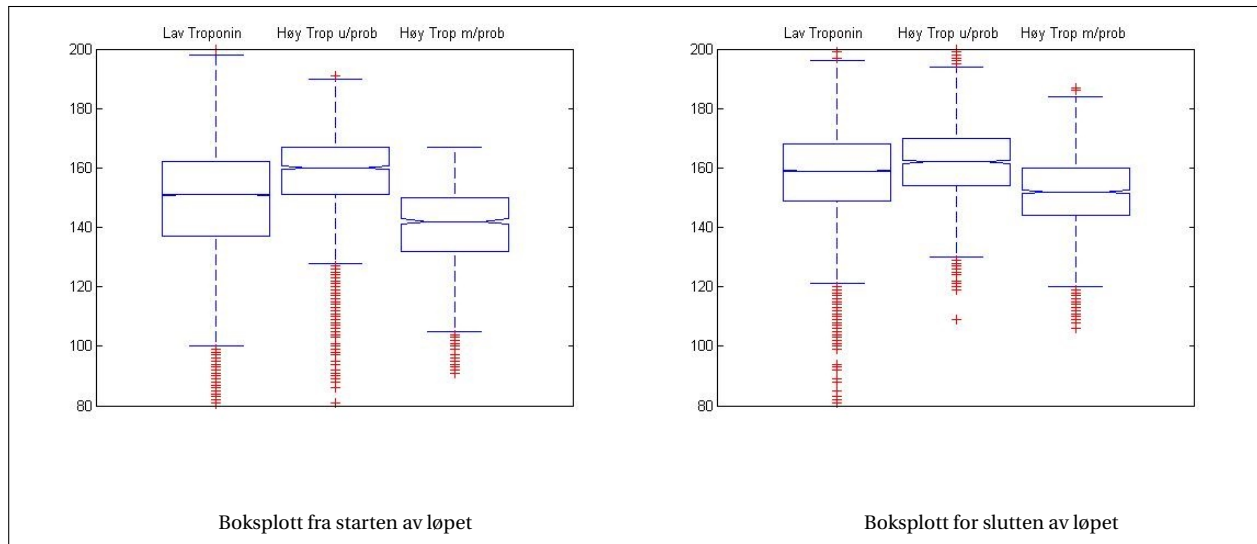
<i>Grupper</i>	<i>F-verdi</i>	<i>p-verdi</i>
Høy troponin m/problemer - Høy troponin u/problemer	42.02	$5.9763 \cdot 10^{-110}$
Høy troponin m/problemer - Lav troponin	67.32	$3.309 \cdot 10^{-123}$
Høy troponin u/problemer - Lav troponin	16.02	$2.8181 \cdot 10^{-79}$

Tabell 4.1: Tabell med f- og p-verdier

Dette betyr at desto høyere f- og lavere p-verdien er, desto større er forskjellen mellom gruppene. Ser man nærmere på resultatet i tabellen, ser man at gruppen som skiller seg mest ut er gruppen deltagerer med hjertesykdommer. F-verdiene her er en del høyere enn den mellom de to friske gruppene. Det betyr at det er størst forskjell mellom den hjertesyke gruppen og de to andre. Samtidig er det mindre forskjell mellom de to gruppene med friske deltagerer. Denne forskjellen er også synlig i boksplottet fra [figur 4.1](#).

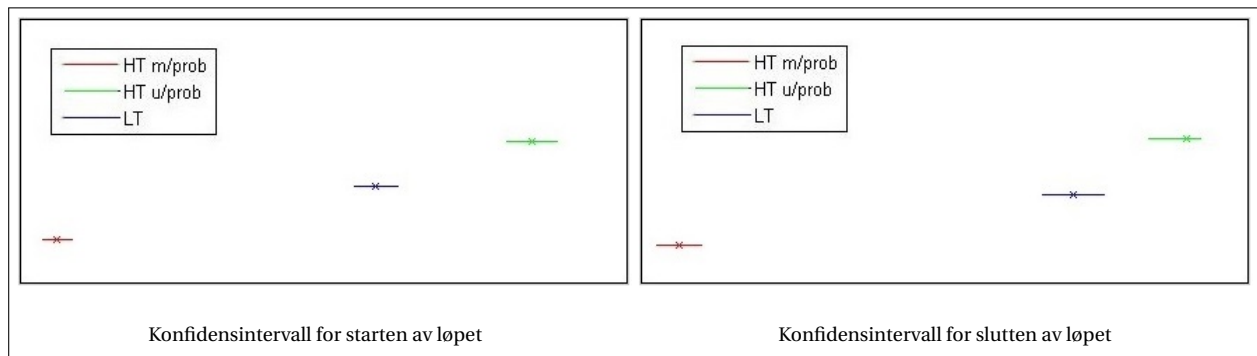
4.1.2 Analyse fra starten og slutten av løpet

Dataen bak boksplottet og tabellen ovenfor er basert på hjertemålingene for alle deltagerne gjennom hele løpet. Det som er interessant er å se nærmere på korte intervaller i løpet og analysere endringene i hjerteraten. Dette gjelder i hovedsak pulsendingene i starten og slutten av løpet, og i opp- og nedoverbakkene. Eksempler på forskjellen mellom starten og slutten kan sees under i [figur 4.2](#).



Figur 4.2: Boksplottet viser hjerteratefordelingen i starten og slutten av løpet

Som man ser i begge figurene ovenfor er det forskjeller mellom de tre gruppene i både starten og slutten av løpet. Den ikke-parametriske metoden, Kruskal Wallis, kan brukes på disse datasettene. Denne metoden fungerer til både normaliserte og ikke-normaliserte data, og kan derfor anvendes her. Resultatet av metoden gir p-verdiene $1.807 \cdot 10^{-214}$ og $3.652 \cdot 10^{-138}$. Dette betyr at gruppene er signifikant forskjellige siden begge verdiene er lavere enn 0.05. Basert på p-verdiene ser man at det er størst forskjell mellom gruppene ved starten av løpet. Sammen med p-verdiene får man konfidensintervallene til gruppene, og disse er illustrert under i [figur 4.3](#).



Figur 4.3: Konfidensintervaller for starten og slutten i løpet

Hvordan man tolker disse resultatene er som følger:

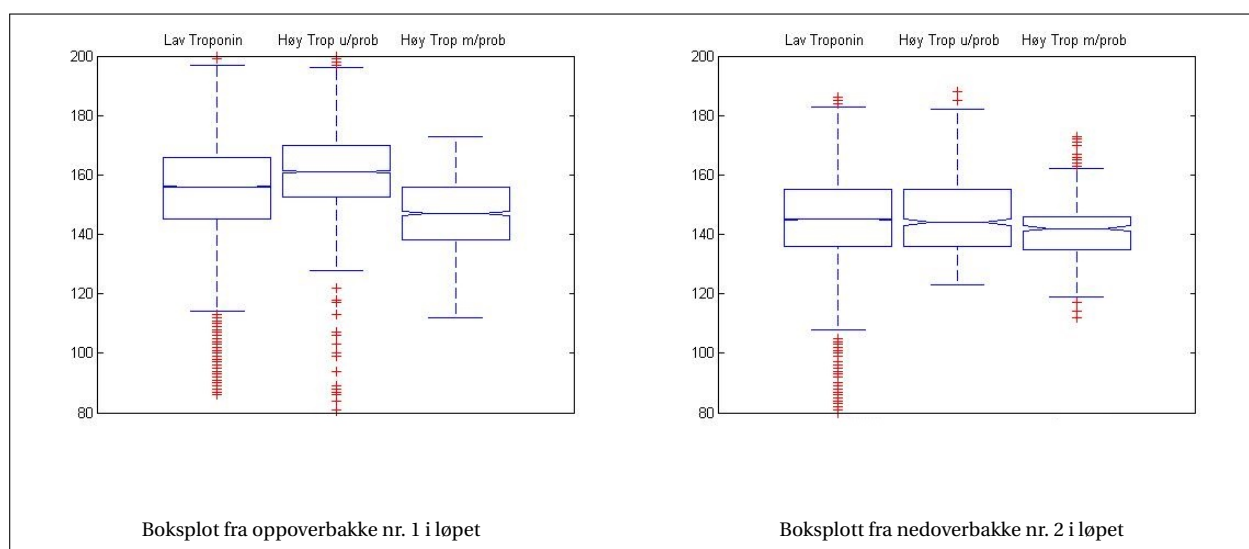
- De er signifikant forskjellige om intervallene ikke overlapper.
- Det er ikke signifikant forskjellige om der er mye overlapping.

- Ved små overlapp må man undersøke nærmere for å avgjøre.

Resultatet av konfidensintervallene ovenfor gir det samme resultatet som p-verdiene. I figuren til høyre ser man at konfidensintervallene til deltagerne uten hjerteproblemer ligger nærmere hverandre. Dette stemmer overens med resultatet fra boksploTTene i figur 4.2. Det er vanskelig å gi en faglig begrunnelse for akkurat disse resultatene. Utslagene kan skyldes at antall deltagere i gruppen med hjertesyke er for liten og resulterer i mangel på data for å kunne generalisere informasjonen.² Det vil si at resultatene gitt i både boksploTTene og konfidensintervallene ovenfor ikke nødvendigvis hadde blitt identisk om gruppene med høy troponin hadde vært større.

4.1.3 Analyse fra en opp- og nedoverbakke

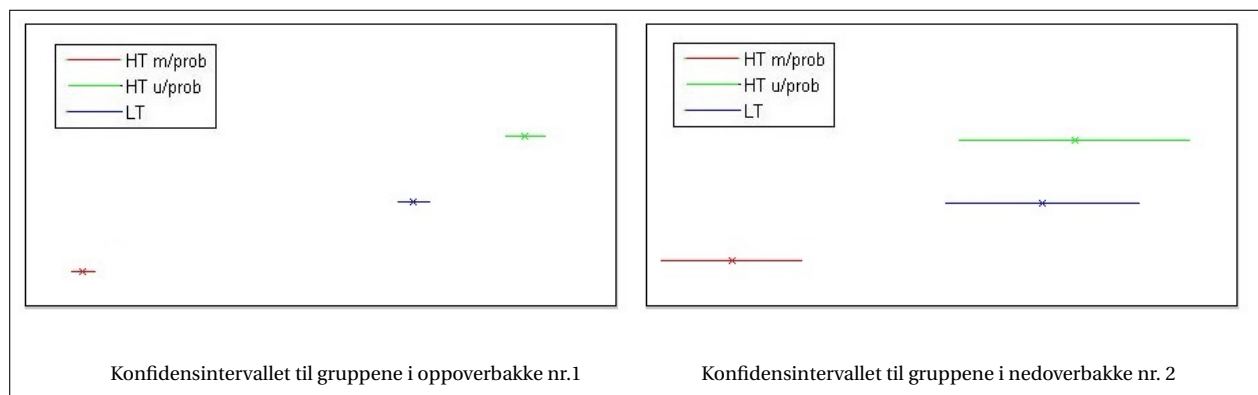
Et annet scenario som er interessant å se nærmere på er fordelingen av hjerterateverdiene i opp- og nedoverbakker. Det er spesielt interessant å se om det oppstår en endring i spredningen mellom de tre gruppene i nedoverbakken. Dette er fordi hjertermålingene til deltagerne kan variere i restitusjonsfasen av løpet. Ifølge Stein Ørn, overlege ved hjerteavdelingen ved Helse Stavanger, er deselerasjonsfasen til deltagere med hjerteproblemer lenger enn hos de friske. Dette gjør at personer med hjerteproblemer har en lavere gradient på pulsen ved restituering og at pulsen deres derfor synker saktere enn hos friske. For å se nærmere på dette og om pulsdataene gir de samme utslagene, hentes målinger ut fra en opp- og en nedoverbakke. Resultatet av dette er illustrert med to boksploTT i figuren under.



Figur 4.4: BoksploTTet viser hjerteratefordelingen i en opp- og nedoverbakke i løpet

²Generalisere - "kunne trekke konklusjoner til en større gruppe enn akkurat selve utvalget" [21]

Resultatet er interessant, og det er synlige forskjeller mellom de to plottene. Den største variasjonen er mellom gruppene i plottet til venstre. Det er nesten så stor forskjell mellom gruppene at nedre kvartil til gruppen "høy TnI uten problem" er på linje med øvre kvartil til deltagerne med problem. Samtidig ser man at der er en del pulsmålinger på utsiden av børstene til deltagerne uten problemer. Dette er ikke tilfelle for deltagerne med problemer fordi målingene deres ligger mer samlet. Ser man på figuren til høyre er forskjellen mellom gruppene mindre. Området mellom øvre og nedre kvartil er svært lite og viser da at 50% av de samlede loggingene ligger svært nære hverandre. Samtidig er det en del målinger på utsiden av børstene, noe som muligens er målingene til en eller to deltagere. Ved hjelp av Kruskal Wallis metoden blir p-verdien for oppoverbakken er $1.4713 \cdot 10^{-269}$ og nedoverbakken $3.356 \cdot 10^{-11}$. Dette viser da at gruppene er signifikant forskjellige. Ser man på p-verdien mellom de to friske gruppene i boksplottet til høyre er denne hele 0.27, altså mye høyere enn 0.05-grensen for å være signifikant forskjellige. Dette kommer mer til syne i konfidensintervallene som er illustrert i figuren under.



Figur 4.5: Konfidensintervallene viser forskjellen mellom gruppene i en opp- og en nedoverbakke

Her ser man at det er stor forskjell mellom intervallene til venstre i figuren og vesentlig mindre forskjell mellom intervallene til høyre. Her kommer det frem at de friske deltagergruppene ikke bare overlapper litt, men veldig mye. Samtidig er det ikke langt unna at intervallet til deltagerne med hjerteproblemer overlapper litt med de andre. Selv om det er ganske synlig at det er lite forskjeller mellom gruppene i boksplottet til høyre i figur 4.4, kan det være greit med en kryssvalidering av resultatet. Siden det finnes både parametriske og ikke-parametriske metoder, kan man se at begge metodene gir samme resultat. Resultatet av Anova-testen ser man i tabellen under.

<i>Grupper</i>	<i>F-verdi</i>	<i>p-verdi</i>
Høy troponin m/problemer - Høy troponin u/problemer	9.35	$1.4029 \cdot 10^{-10}$
Høy troponin m/problemer - Lav troponin	8.44	$3.4274 \cdot 10^{-8}$
Høy troponin u/problemer - Lav troponin	2.13	$0.242 \cdot 10^{-1}$

Tabell 4.2: Resultatet av en Anova-test som sjekker signifikansen mellom gruppene i nedoverbakken

Tallene fra tabellen gir det samme resultatet som løsningen fra den ikke-parametriske metoden. Gruppen med de hjertesyke skiller seg ut fra de andre, på grunn av F-verdiene 8.44 og 9.35. Begge er over 5, som ut fra den statistiske analysen konkluderer med at de er signifikant forskjellige. F-verdien på 2.13, bekrefter at deltagerne uten hjertesykdom ikke er signifikant forskjellige, akkurat som resultatet fra Kruskal Wallis metoden.

4.1.4 Boksplokk av normaliserte hjerteverdier

Lest ut fra figurene ovenfor ser man at det er store variasjoner i hjerteraten på de utvalgte segmentene i løpet, noe som gir grunn til nærmere undersøkning ved hjelp av andre funksjoner. I oppoverbakken ligger forskjellen mellom de med hjerteproblemer og de uten på 11 og 12 slag i minuttet. I nedoverbakken er forskjellen en del mindre og ligger på 2 og 3 slag.

En mulig feilkilde til dette er at alderen på deltagerne spiller inn. Det er klinisk bevist at desto eldre man blir, desto lavere blir forventet makspuls, samtidig som risikoen for hjertesykdom øker. Menn er også mer utsatt for hjertesykdommer enn kvinner, da kvinnes østrogen beskytter dem.^[22] I følge en lignende studie ^[23], kan også deltagernes BMI ³, kolesterolnivå, blodtrykk og om de røyker, avgjøre om personer klarer å oppnå makspuls. På grunn av manglende data, er det i denne studien kun tatt høyde for alder og kjønn.

Hvis en av de tre gruppene inneholder enten mange unge eller gamle deltagere vil dette virke inn på resultatet. Derfor er det viktig å sjekke om det er en ujevn aldersfordeling mellom gruppene.

³BMI - Body Mass Index, et forhold mellom høyde og vekt på mennesker

Gruppe	Gjennomsnittlig alder
Høy troponin m/prob	50.4 år
Høy troponin u/prob	39.7 år
Lav troponin	44.4 år

Tabell 4.3: Gjennomsnittsalder i de forskjellige gruppene

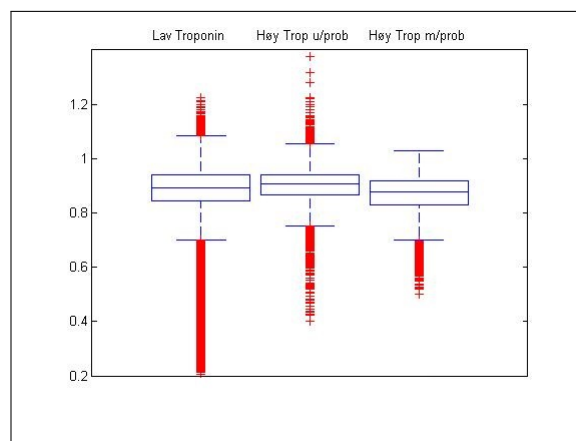
Dette viser at den gjennomsnittlige alderen til de hjertesyke deltagere er en del høyere. Dette kan være en av grunnene til at denne gruppen har lavere puls enn de andre. Det er derfor viktig å se på de normaliserte hjertemålingene. Den maksimale forventede hjerteraten til en gjennomsnittlig person er gitt ved formelen:

$$\text{Makspuls} = 208 - 0.7 \cdot \text{alder (år)} \quad [24] \quad (4.1)$$

Bruker man formelen 4.1 vil hjerteraten normaliseres og alderen til deltageren blir ubetydelig. Normaliseringen gjøres ved å dele deltagerens hjerterate på utregnet makspuls ved hjelp av formelen ovenfor. Deltagernes alder leses ut fra det vedlagte regnearket, bestående av informasjon om hver deltager. Formelen for normalisert hjerterate blir da seende slik ut:

$$\text{Normalisert hjerterate} = \frac{\text{Hjerterate}}{\text{Makspuls}} \quad (4.2)$$

Etter man har laget en metode for å hente ut alle hjertemålingene til hver deltager og delt dem på deres forventede makspuls, får man boksplottet som man ser i figuren under. Dette boksplottet er generert ut fra alle loggede hjerterateverdier for alle gruppene gjennom hele løpet.



Figur 4.6: Boksplottet viser normalisert hjerteratefordelingen på hele løpet

Boksplottet i figuren ovenfor ligner litt på de i [figur 4.1](#). Det er ikke overraskende at det er så mye punkter utenfor børstene da alle pulsmålingene, uansett tidspunkt, deles på forventet makspuls. Det er fortsatt tydelig at deltagerne med hjerteproblemer ikke får like høy puls som de andre. Samtidig har de få eller ingen målinger som overskrider forventet makspuls (verdier over 1).

Grunnen til at forskjellen mellom gruppene nå er blitt mindre, kan enkelt forklares. I følge [tabell 4.3](#) er gjennomsnittsalderen til de syke mye høyere, noe som gir en lavere forventet makspuls. Siden makspuls står i nevneren i [ligning 4.2](#) blir deres normaliserte hjerterate høyere. Selv om det nå er blitt vanskeligere å se forskjellen mellom gruppene, gjør Kruskal Wallis fortsatt jobben og konkluderer med at p-verdien er 0, blank. Noe som betyr at det fortsatt er store forskjeller mellom gruppene. I stedet for å se nærmere på flere plott, lages heller en tabell som viser både konfidensintervallene og p-verdiene for de interessante strekningene i løpet. Tabellen inneholder resultatene med både originale hjerterverdier (lysegrå) og normaliserte hjerterverdier (mørkegrå). Denne tabellen er illustrert under og de forskjellige distansene i løpet er som følger:

- 0-2 km er starten av løpet
- 35-37 km er første oppoverbakke
- 42-44 km er første nedoverbakke
- 70-72 km er andre oppoverbakke
- 72.8-73.5 km er andre nedoverbakke
- 88-90 km er slutten av løpet

<i>Hjerterate</i>	<i>Distanse i løpet</i>	<i>Gruppe</i>	<i>Konfidensintervall</i>	<i>p-verdi</i>
Vanlige	0 - 2 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-5074 , -4197> <4185 , 5456> <8705 , 10208>	$1.8 \cdot 10^{-214}$
Normaliserte	0 - 2 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-3903 , -3027> <2439 , 3710> <5788 , 7290>	$9.6 \cdot 10^{-109}$
Vanlige	35 - 37 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-2816 , -2129> <2705 , 3693> <5086 , 6257>	$2.8 \cdot 10^{-219}$
Normaliserte	35 - 37 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-1912 , -1223> <1362 , 2351> <2839 , 4009>	$7.2 \cdot 10^{-45}$
Vanlige	42 - 44 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-2948 , -2193> <1967 , 3072> <4439 , 5740>	$1.8 \cdot 10^{-85}$
Normaliserte	42 - 44 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-1638 , -884> <-412 , 692> <750 , 2051>	$2.8 \cdot 10^{-14}$
Vanlige	70 - 72 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-2196 , -1336> <4387 , 5565> <6033 , 7451>	$4.7 \cdot 10^{-111}$
Normaliserte	70 - 72 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-295 , 567> <1637 , 2872> <1438 , 2857>	$1.2 \cdot 10^{-18}$
Vanlige	72.8 - 73.5 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-347 , -130> <617 , 1269> <660 , 1443>	$3.4 \cdot 10^{-11}$
Normaliserte	72.8 - 73.5 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<136 , 612> <-295 , 357> <-774 , 49>	$1.2 \cdot 10^{-3}$
Vanlige	88 - 90 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<-1467 , -594> <4259 , 5470> <5169 , 6620>	$2.9 \cdot 10^{-87}$
Normaliserte	88 - 90 km	Høy troponin m/problemer Høy troponin u/problemer Lav troponin	<377 , 1250> <1020 , 2232> <87 , 1538>	$8.6 \cdot 10^{-13}$

Tabell 4.4: Tabell med konfidensintervall og p-verdier

Som man ser her, er forskjellene mellom gruppene alltid mindre når normaliserte hjertemålinger benyttes, og dette gjelder for samtlige testede distanser i løpet. Ser man nærmere på konfidensintervallene, overlapper de ikke før sent ut i løpet. Overlappingen skjer

ved de siste kilometerne i løpet og i siste opp- og nedoverbakke.

4.2 HR - inndeling

For å se mer nøyaktig på hvordan pulsen til deltagerne i gruppene er fordelt lager man et histogram. Man velger da antall intervall og hvordan man ønsker dem inndelt. Alle samplene til deltagerne samles før inndelingen og resultatet er at noen av deltagerne bistår med flere sampler enn andre. Det er ingen overdådig representasjon av sampler fra enkelte deltagerne, så en vektning av dataen er ikke nødvendig, da resultatet blir mye det samme. Det ble gjort en slik vektning i studien, ved å dele hjertemålingene til hver deltager på totalt antall sample. Dette gjør at hver deltager bidrar med akkurat like mye informasjon til studien, og dette fungerer som en normalisering. Resultatet av denne normaliseringen gav svært like plott og tilførte ikke studien mer bredde. Problemet er også at de nye hjertemålingene blir i desimaltall og kan ikke direkte anvendes mot normaliseringen av hjerterateverdiene.

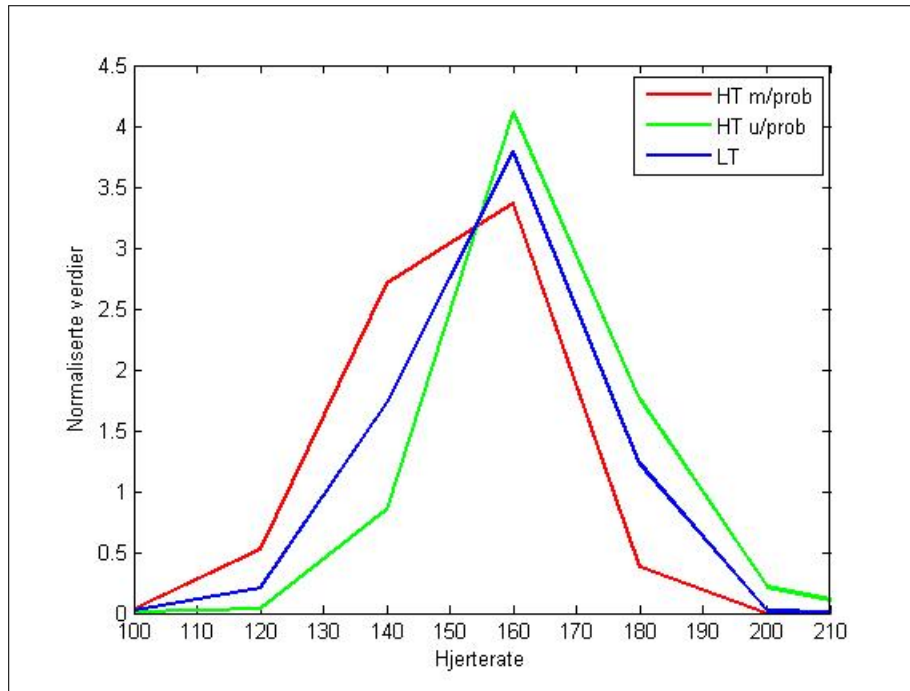
I tabellen under ser man nærmere på fordelingen av hjerterate med intervaller på 20.

HR-fordeling	< 99	110-119	120-139	140-159	160-179	180-199	200 <
Lav Troponin *	5486 0.39%	8811 0.75%	102654 8.83%	502082 41.9%	475446 42.19%	50754 4.91%	780 0.03%
Høy Troponin u/prob *	217 0.21%	171 0.16%	3417 2.49%	39660 32.9%	60533 55.78%	6422 7.77%	495 0.69%
Høy Troponin m/prob *	30 0.08%	448 1.28%	9335 19.5%	24319 50.52%	10460 28.36%	72 0.25%	0 0%

Tabell 4.5: Tabell med fordelingen av hjerterate

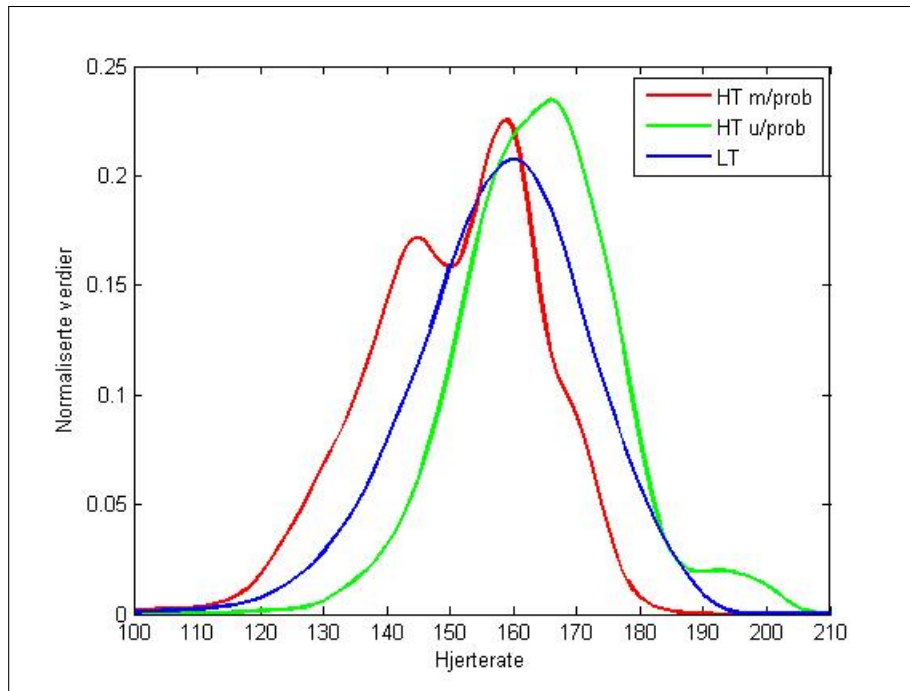
Med tanke på hvordan boksplottet så ut er resultatet fra tabellen som forventet. Mesteparten av samplene for deltagerne uten hjerteproblemer er fordelt mellom 140 og 180 i puls. Dette gjelder de to øverste gruppene i tabellen, de med lav troponin og de med høy troponin uten problemer. Samtidig er mellom 5 og 8% av de loggede pulsverdiene over 180. Denne tendensen ser man ikke hos de med hjerteproblemer. Deres pulsverdier er noe lavere fordelt og ligger hovedsaklig mellom 120 og 180. Til sammenligning er deltagerne *uten* hjerteproblemer lite representert i kolonnen 120-139. Plotter man fordelingen til de representative gruppene med samme intervaller som ovenfor, blir resultatet som vist i figuren

under.



Figur 4.7: Fordelingen av hjerteratene fordelt mellom de tre forskjellige gruppene

Ved å plote fordelingen synliggjør man hvor stor forskjell det er mellom de med hjerteproblemer og de uten. Reduseres lengden av intervallene, vil oppløsningen på grafene forbedres og bli mye glattere. Dette kan gjøres i Matlab ved hjelp av det vedlagte skriptet, "HR-inndeling". Her kan man enkelt velge andre intervaller og plote sannsynlighetstetthetsfunksjonen for de forskjellige gruppene. I figuren under er det valgt lavere intervaller, slik at funksjonene får enda flere sampler og oppløsningen på grafene økes.



Figur 4.8: Sannsynlighetstetthetsfunksjonene til de tre forskjellige gruppene

I realiteten er ikke plottet vist ovenfor så glatt, men for det visuelle benyttes en glattefunksjon som fikser litt på dette. Som man ser, er det størst avvik mellom deltagerne med høye TnI-verdier uten problemer og de med problemer. Det er et interessant resultat, da begge deltagergruppene har høy troponin til felles. Det var forventet større avvik mellom deltagerne med problemer og deltagerne med lav troponin. Dette kan på lik linje som tidligere skyldes den gjennomsnittlige alderen til gruppene. Deltagere med høye TnI-verdier uten hjerteproblemer har lavest gjennomsnittlig alder og en normalisering av hjertemålingene kan endre på resultatet som man ser i figuren ovenfor.

4.2.1 Normalisert HR - inndeling

Som tidligere forklart er det viktig å ta høyde for alderen på deltagerne, da dette gir andre resultater. Selv om bokplottet med de normaliserte hjertemålingene i [figur 4.6](#) viser at det fortsatt er stor forskjell mellom gruppene, ser man i [tabell 4.4](#) at forskjellen mellom gruppene er blitt mindre. Samtidig viser tabellen at det er markante forskjeller mellom gruppene, noe følgende tabell underbygger.

<i>HR-fordeling</i>	<i>< 0.5</i>	<i>0.5-0.6</i>	<i>0.6-0.7</i>	<i>0.7-0.8</i>	<i>0.8-0.9</i>	<i>0.9-1.0</i>	<i>1.0 <</i>
Lav Troponin	0.24%	0.22%	1.12%	9.08%	42.96%	39.86%	6.52%
Høy Troponin u/prob	0.13%	0.10%	0.23%	3.73%	41.12%	49.82%	4.87%
Høy Troponin m/prob	0%	0.14%	1.28%	12.74%	50.25%	35.39%	0.2%

Tabell 4.6: Tabell med fordelingen av de normaliserte hjerteratene

Ser man i tabellen ovenfor ser man at de hjertesyke deltagerne sliter med å oppnå makspuls til tross for alderen. Om dette skyldes at deltagerne føler på kroppen at den ikke fungerer optimalt og holder tilbake av den grunn, eller om hjertet setter begrensningene, er usikkert. Det er bare 0.2% av alle samplene til de hjertesyke deltagerne som oppnår eller har høyere verdier enn makspuls. Samtidig ser man de samme tendensene som i [tabell 4.5](#), at hovedvekten av samplene til de hjertesyke deltagerne ligger lavere enn hos de andre.

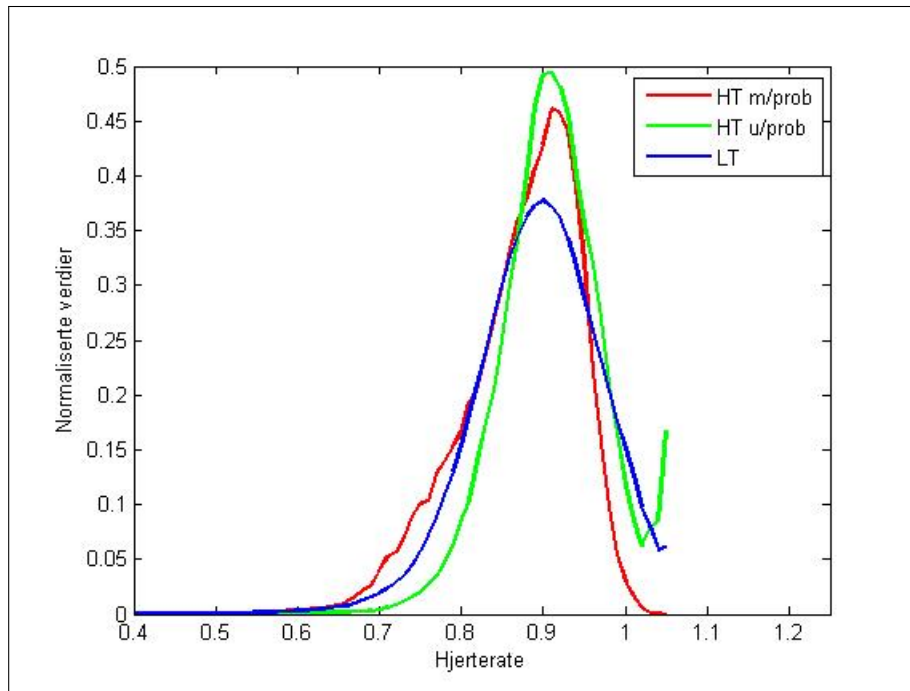
Ser man nærmere på en studie utført om hjerterespons under trening, ser man at 21% av totalt 1575 deltagere klarte ikke oppnå 85% av deres forventede makspuls. Dette var før de tok høyde for alder, BMI, røyking, diabetes og diverse andre faktorer. Selv etter denne justeringen forble det en del deltagerne som fortsatt ikke nådde 85%. Av disse var det dessverre ingen tydelig sammenheng mellom deltagerens tilstand og deres evne til å oppnå makspuls. Resultatene her indikerer vertfall at det er forskjell mellom de hjertesyke og de friske deltagerne. Dette er basert på resultat fra både de vanlig hjertemålingene og de normaliserte. Begge gir grunnlag for å kunne si at de hjertesyke deltagerne har lavere snitt- og makspuls. I tabellen under kan man se hvor mange av deltagerne som faktisk oppnår sin forventede makspuls.

<i>Navn</i>	<i>Oppnår makspuls</i>	<i>Oppnår ikke makspuls</i>	<i>Totalt</i>
Gruppe	Antall deltagere	Antall deltagere	Antall deltagere
Lav Troponin	77 (48.7%)	81 (51.3%)	158
Høy Troponin u/prob	12 (70.6%)	5 (29.4%)	17
Høy Troponin m/prob	4 (57.1%)	3 (42.9%)	7

Tabell 4.7: Oversikt over antall deltagere som oppnår forventet makspuls

Som man ser i tabellen ovenfor oppnår omtrent 51% av alle deltagerne sin forventede makspuls. Igjen, overraskende nok er det størst andel av deltagere i gruppen med høye TnI-verdier som oppnår forventet makspuls.

I figuren under ser man nærmere på selve fordelingen av de normaliserte verdiene.



Figur 4.9: Sannsynlighetstetthetsfunksjonene til de tre forskjellige gruppene med normaliserte hjertemålinger

Her får man veldig like resultat som i [tabell 4.6](#). Grafen til deltagerne med hjerteproblemer (rød) stiger stiger før de andre, og synker før de andre. Likevel er hovedvekten i fordelingen av hjertemålingene på samme sted som de andre gruppene. Denne sannsynlighetsfordelingen gjelder for alle de tre gruppene gjennom hele løpet. Det er derfor interessant å dele opp løpet i mindre deler og se nærmere på dem, slik som det ble gjort i [delkapittel 4.1](#).

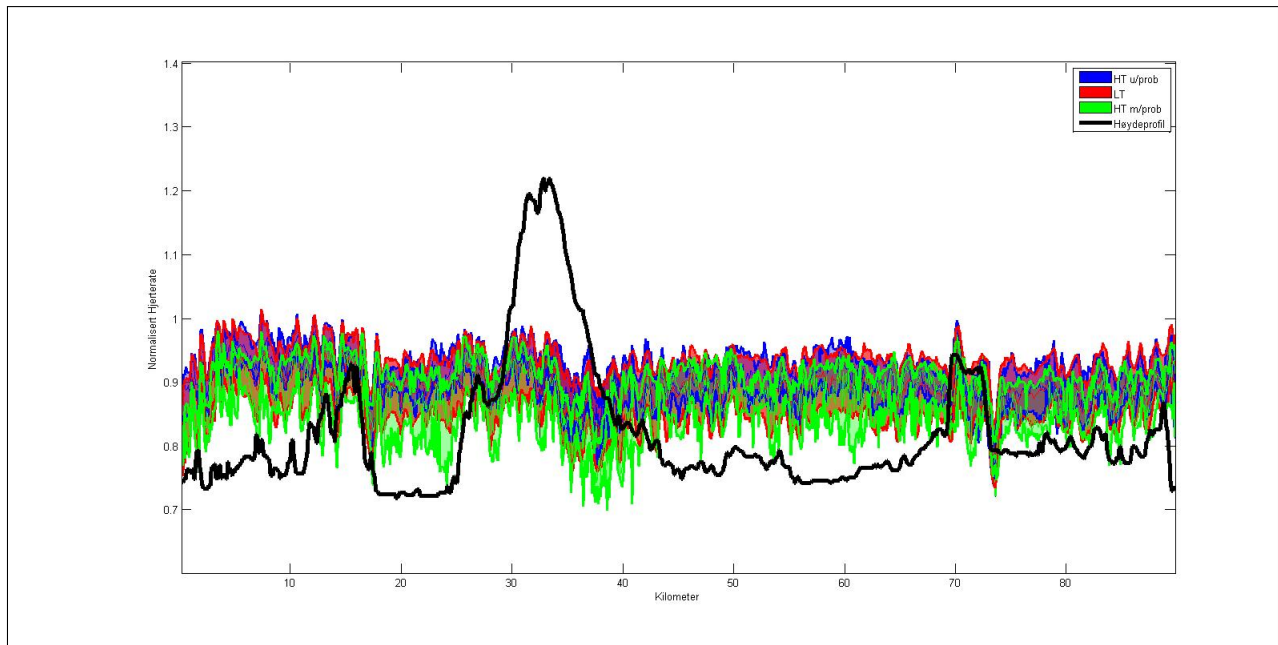
4.3 Median og Kvartiler

Ved hjelp av boksploot kan man enkelt se fordelingen av hjerteraten til deltagerne. Begrensningene til funksjonen er at det blir en snittmåling over lengre intervaller. Intervallene kan deles mer opp, men nøyaktigheten uteblir. Hensikten bak bruken av funksjonen er å se etter tendenser i pulsen som skiller gruppene. Dette gjelder spesielt rundt bakkene.

For en mer nøyaktig fremstilling av dataene må andre metoder benyttes. Her bruker man det vedlagte skriptet 'Kvartil'. Denne funksjonen er laget slik at man kan velge avstanden mellom hver gang kvartilene skal kalkuleres og hvor nær denne avstanden loggingen må ligge for å tas med. På grunn av at det tar lang tid å kalkulere dataene for hvert punkt, kan man velge større avstander mellom hver logging. Resultatet blir en trade-off mellom tidsbesparelse og en høyere samplet graf. Skal man gå dypt i detaljer og studere oppførselen til pulsen nøye, vil det være

smart å velge avstander på rundt 15-20 meter. Et annet alternativ er å snevre ned området man søker på, slik det blir færre punkter å kalkulere dataene for. Her er det allerede forhåndsregret fem forskjellige distanser å velge mellom. Fire av dem gjelder for opp- og nedoverbakkene og det femte, for et flatt parti. Grunnen til at denne funksjonen er så viktig, er at man nå kan se på pulsvariasjonen til de tre gruppene, meter for meter gjennom hele løpet.

For plottingen av dataen brukes en fyllfunksjon som setter farge mellom øvre og nedre kvartil. Medianen plottes med en helstrukket linje i samme farge. Hver gruppe plottes med hver sin gjennomsiktige farge sammen med høydekurven fra løpet. I figuren under ser man et slikt plott for hele løpet.

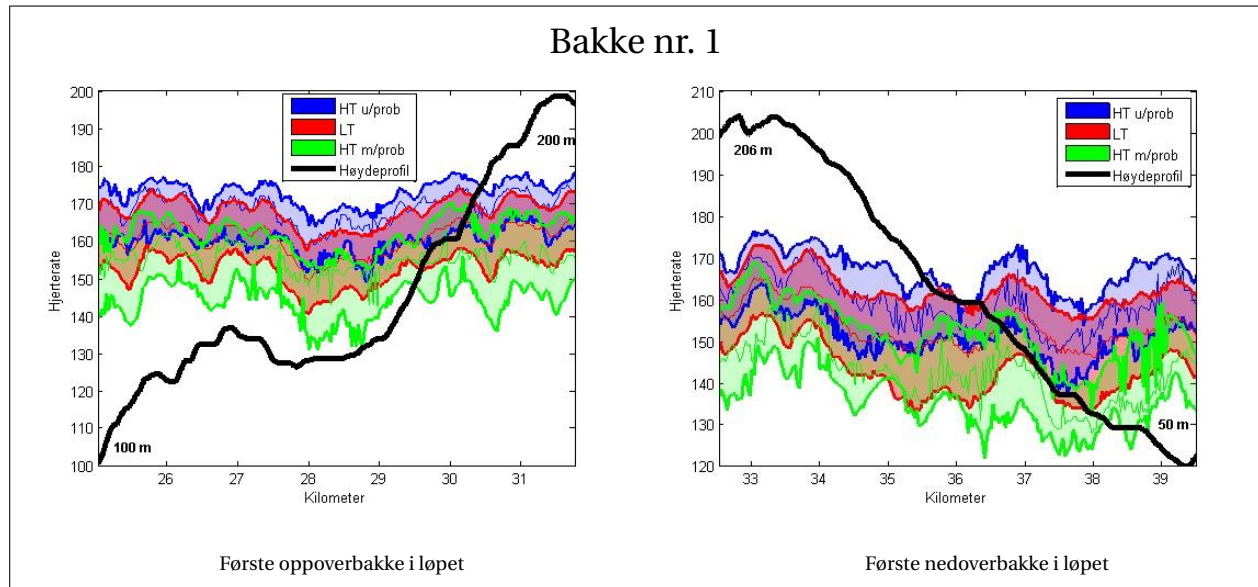


Figur 4.10: Figuren ovenfor viser medianen, øvre og nedre kvartil til gruppene lav troponin og høy troponin *med* og *uten* hjerteproblemer. Den svarte linjen viser høydeprofilen til løpet.

Som man ser langs y-aksen er det normaliserte hjerterverdier i plottet ovenfor. Samtidig ser man at det ikke er noen markant pulsøking i oppoverbakkene, og at pulsen synker litt i nedoverbakkene. Dette viser at det er en viss korrelasjon mellom høydekurven og hjerteraten til deltagerne. Siden pulsendingene ikke er så store i løpet, er det da rimelig å anta at deltagerne presser seg hardt under hele økten. Det er nok derfor de fleste normaliserte målingene ligger mellom 0.80 og 0.95. Resultatet fra [tabell 4.6](#) underbygger dette og viser at mellom 80 og 90% av målingene ligger i dette intervallet.

4.3.1 Median og kvartiler i bakkene

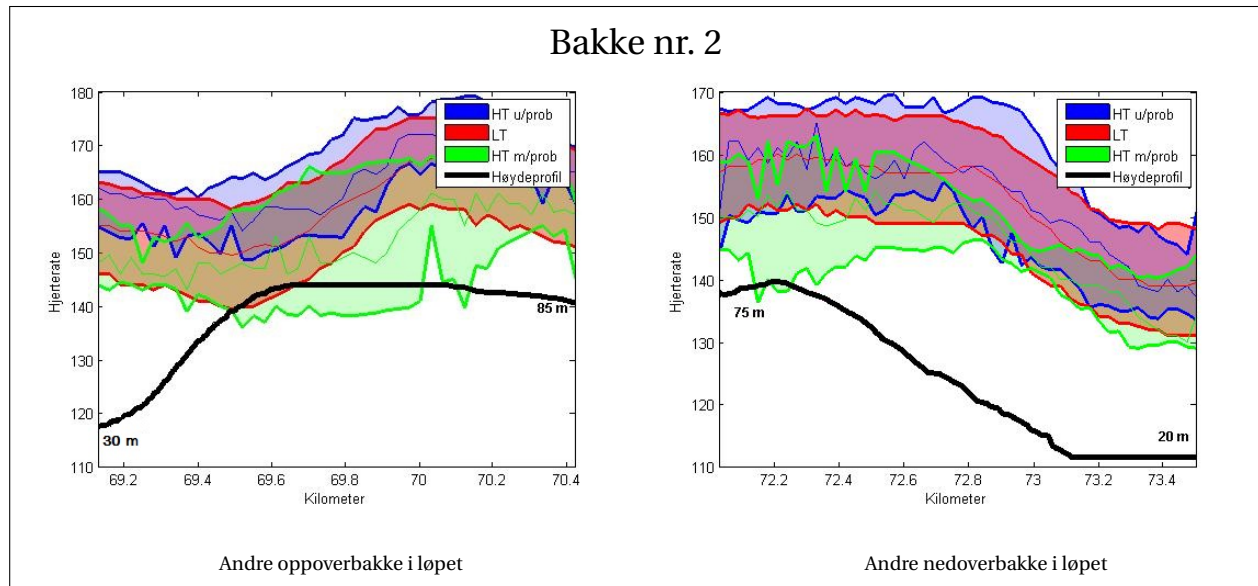
Som tidligere forklart har hjertesyke mennesker en lengre deselerasjonsfase enn de friske, noe resultatene fra boksplottet i figur 4.4 bekreftet. Det er derfor interessant å se på hvordan spredningen av de loggede hjertemålingene for hver deltagergruppene forandrer seg over tid. I figur 4.11 ser man resultatet av denne metoden i første bakke av løpet.



Figur 4.11: Plotting av median, øvre og nedre kvartil for hver gruppe i bakke nr. 1 i løpet

Som man ser i figuren ovenfor starter motbakken på omtrent 100 høydemeter, før den stiger ytterligere 100 nye høydemeter. Selve motbakken er på 6 kilometer, noe som gir en gjennomsnittlig stigning på nesten 2%. I det hardeste partiet, mot slutten av motbakken, er stigningen oppe i over 3%. Dette gjør denne enkeltdistansen til en av de delene i løpet med høyest målte pulsverdier. Som man ser i figuren til venstre er det også en liten nedoverbakke. Det er her forskjellen mellom de tre forskjellige gruppene er minst. Dette er veldig interessant, og det er igjen synlig at pulsen til de hjertesyke deltagerne synker saktere ved lavere intensitet. Ellers i oppoverbakken er pulsen til deltagerne med hjerteproblemer (grønn), nesten konsekvent minimum 10 slag i minuttet lavere enn deltagerne i de andre gruppene.

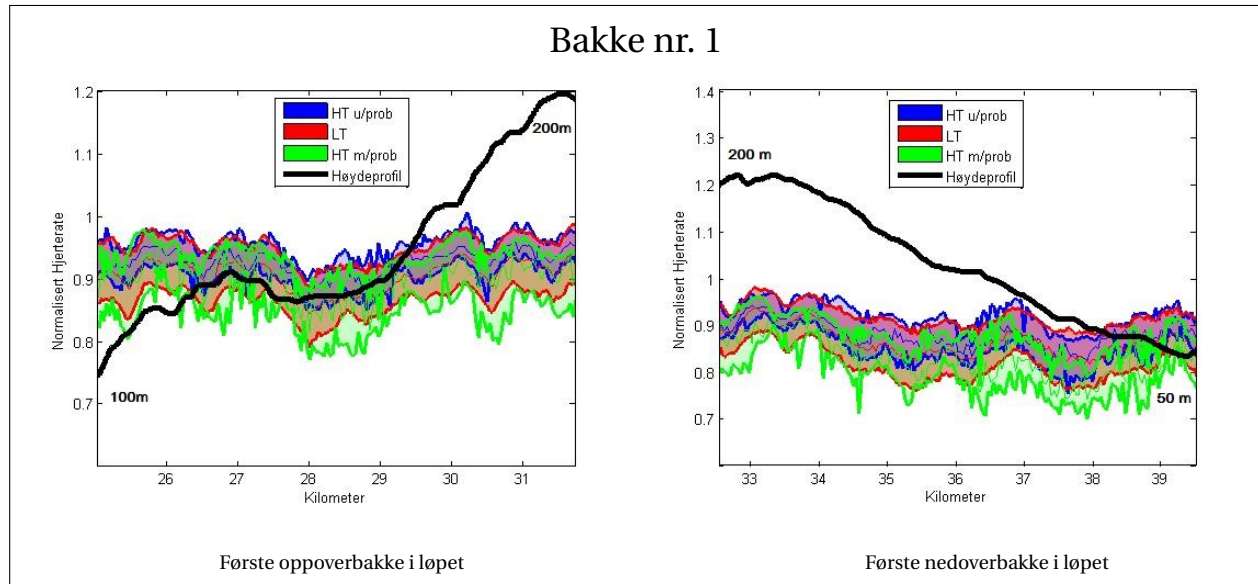
Ser man på figuren til høyre, ser man at det er flere perioder hvor de hjertesyke deltagerne har lik puls med deltagerne med lav TnI-verdi. Siden nedoverbakken er over 6 kilometer, vil pulsen til alle deltagerne falle en god del, inkludert pulsen til de hjertesyke. Det er derfor litt vanskelig å få frem de helt store forskjellene mellom gruppene her. Disse kommer mer til syne i figur 4.12.



Figur 4.12: Plotting av median, øvre og nedre kvartil for hver gruppe i bakke nr. 2 i løpet

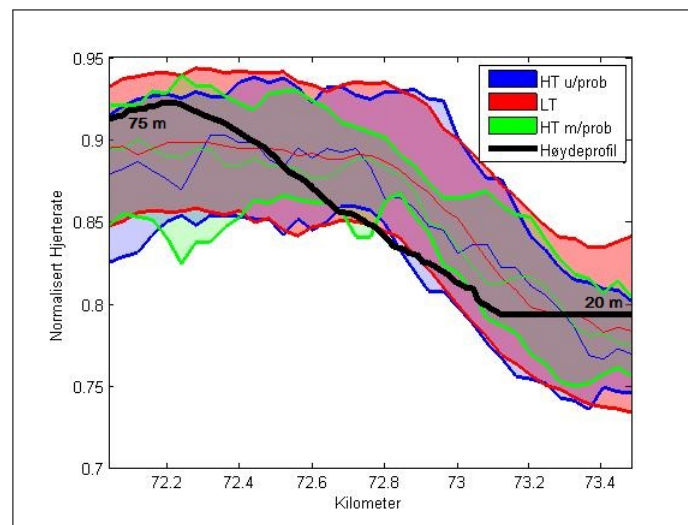
Ser man på høydeprofilen i figuren ovenfor, starter bakken på 30 høydemeter og stiger til rundt 85. På grunn av lavere verdier langs x-aksen, blir høydeprofilen dratt i horisontal retning og det er ikke lenger synlig at stigningen er på hele 13%. I starten av bakken har de friske deltagerne like lav puls som de hjertesyke, men mot toppen endrer dette seg. Da ser man at pulsen til de friske deltagerne øker mest og igjen når et høyere nivå enn de hjertesyke. I plottet til høyre ser man at deltagerne med hjerteproblemer har en del lavere puls enn de andre før nedoverbakken. Når man begynner på nedoverbakken og deltagerne kan puste ut, daler pulsen deres. I dette plottet er det svært synlig at pulsen til deltagerne med hjerteproblemer (grønn) synker saktere og legger seg på de samme pulsverdiene som resten av deltagerne.

4.3.2 Normaliserte verdier



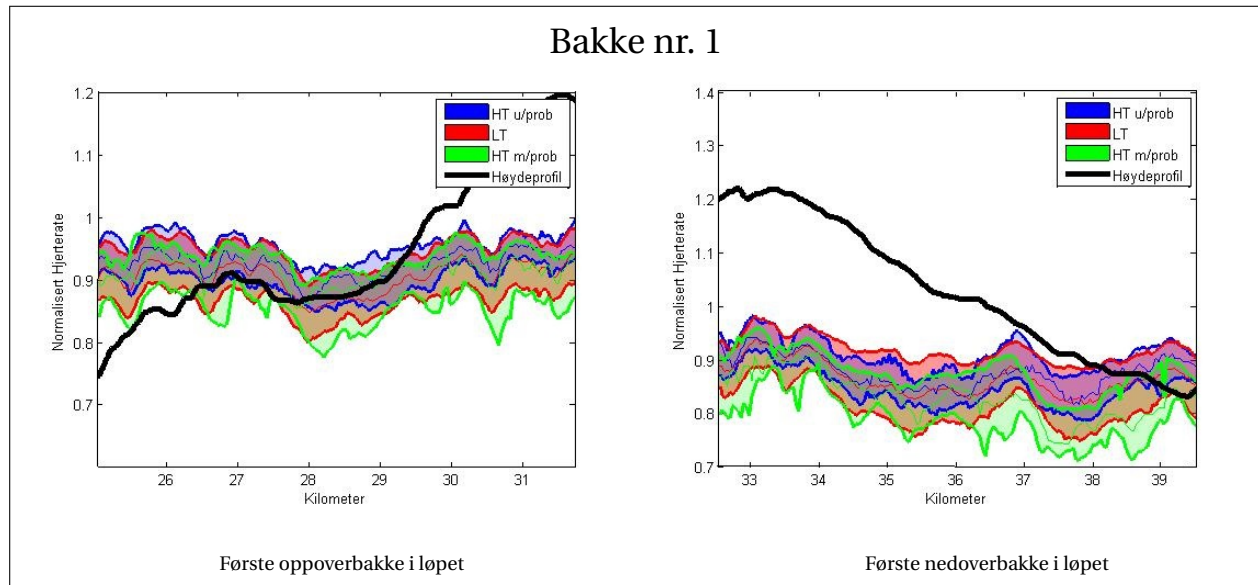
Figur 4.13: Plotting av median, øvre og nedre kvartil for hver gruppe i bakke nr. 1 i løpet

Selv etter man har normalisert hjertemålingene og neglisjert innvirkningen av alderen på målingene, ser man det at de hjertesyke deltagerne fortsatt har lavere puls. Forskjellen mellom gruppene er nå blitt mindre, slik som den statistiske analysen forutså. Ved flere anledninger er det liten eller ingen pulsforskjell mellom gruppene, og dette oppstår oftest i nedoverbakkene eller i de små hengene av løpet. Ser man på [figur 4.14](#) under, ser man at det ikke lenger er de hjertesyke deltagerne som har de laveste pulsmålingene.



Figur 4.14: Denne figuren viser tydelig at pulsen til deltagerne med hjerteproblemer synker saktere enn pulsen til de friske

4.3.3 Normaliserte verdier (kun menn)



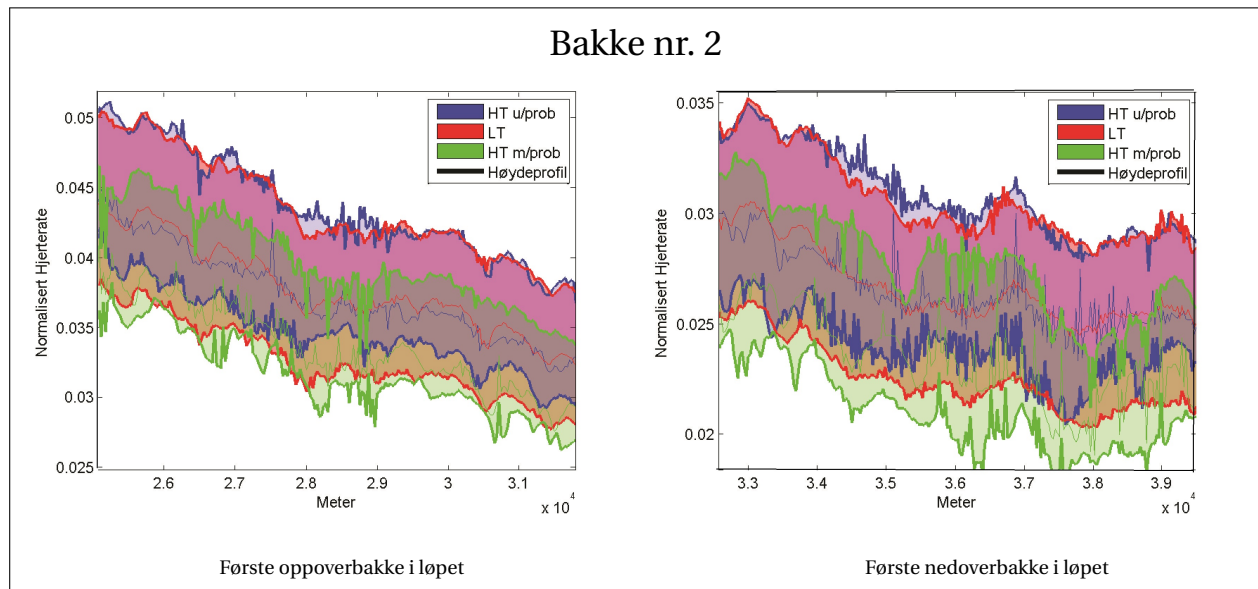
Figur 4.15: Plotting av median, øvre og nedre kvartil for hver gruppe (kun menn) i bakke nr. 1 i løpet.

Siden sjansen for å få hjerteproblemer er litt større hos menn enn kvinner, og rundt 80% av deltagerne i Nordsjørittet er menn, er det interessant å se hvordan resultatet blir når man fjerner kvinnene fra studien. Alle de åtte hjertesyke deltagerne er menn, så denne gruppen er fortsatt sterkt representert. Det er derimot ikke den andre gruppen med høye TnI-verdier. Her forsvant syv av de totalt 17 deltagerne, noe som betyr at kvinner utgjorde over 41% av denne gruppen. Det er en svært høy prosentandel, og det er vanskelig å gi en god forklaring på hvorfor så mange kvinner hadde høye TnI-verdier uten å ha hjerteproblemer. Etter man har fjernet alle kvinnene, står man igjen med 151 mannlige deltagere. Pulsmålingene til alle mennene er plottet i figur 4.15 ovenfor. Fortsatt ser man de samme tendensene, at de hjertesyke deltagerne har lavere puls. Forskjellen mellom gruppene er nå blitt enda mindre, noe som indikerer at kvinnene bidro med høye pulsmålinger. Dette kan man blant annet se til høyre i begge plottene i figur 4.13 og figur 4.15. Siden kvinnene var sterkt representert i gruppen høy TnI uten problemer (blå), følger man utviklingen av det blå området mellom figurene. Da ser man at pulsen er mye lavere og at det blå området nå ikke ligger i overkant av plottet.

Nå som man vet at kvinnene bidrar med høye pulsmålinger til studien, kan dette kanskje forklare hvorfor de var så sterkt representert i gruppen med høye TnI-verdier. Som det ble forklart i teorien om troponin i delkapittel 2.3.2, kan fysiske anstrengelser føre til høye TnI-verdier.

4.3.4 Hjertemålingene delt på minutter

For å eksperimentere litt mer og prøve å skille gruppene enda mer fra hverandre, deler man pulsmålingene til hver deltager på medgått tid. Dette vil gi en synkende kurve utover hele løpet siden medgått tid øker, men det er uansett kun aktuelt å se på mindre distanser. Bruker deltagerne lang tid, vil nevneren øke og hjertemålingene vil bli lavere. Resultatet av denne undersøkelsen ser man i figuren under.



Figur 4.16: Plotting av median, øvre og nedre kvartil for hver gruppe i bakke nr. 1 i løpet. Her deler man på medgått tid for å se om de tre gruppene kan skilles mer fra hverandre

I figuren ovenfor ser man at forskjellen mellom de tre gruppene ikke er blitt noe særlig større. Dette betyr at deltagerne i de forskjellige gruppene bruker tilsynelatende lik tid på løpet og at tiden ikke virker inn på resultatet.

Kapittel 5

Diskusjon

Det er interessant at deltagerne med hjerte problemer har lavere puls enn de andre. En forklaring på dette kan være at hjertet ikke fungerer optimalt og ikke er i stand til å pumpe fortere. Alle deltagerne som var med i denne studien krysset av på at kroppen deres kjentes bra ut både før, under og etter løpet. Siden symptomer på hjerte problemer oppleves veldig forskjellig, kan det hende at deltagerne følte et lite ubehag i kroppen, men ute av stand til å vite om det var hjerte relatert. Dette kan muligens gjøre at deltagere enten bevisst eller ubevisst unngår å presse kroppen for mye. Til nå finnes det svært få studier hvor man har sammenlignet makspuls til hjertesyke og friske personer. Det er derfor vanskelig å komme med en begrunnelse for dette resultatet, men en teori er at hjertet ikke får nok blodtilførsel og derfor fungerer dårligere. Dette kan da resultere i at det ikke er i stand til å generere like høy hjerterate.

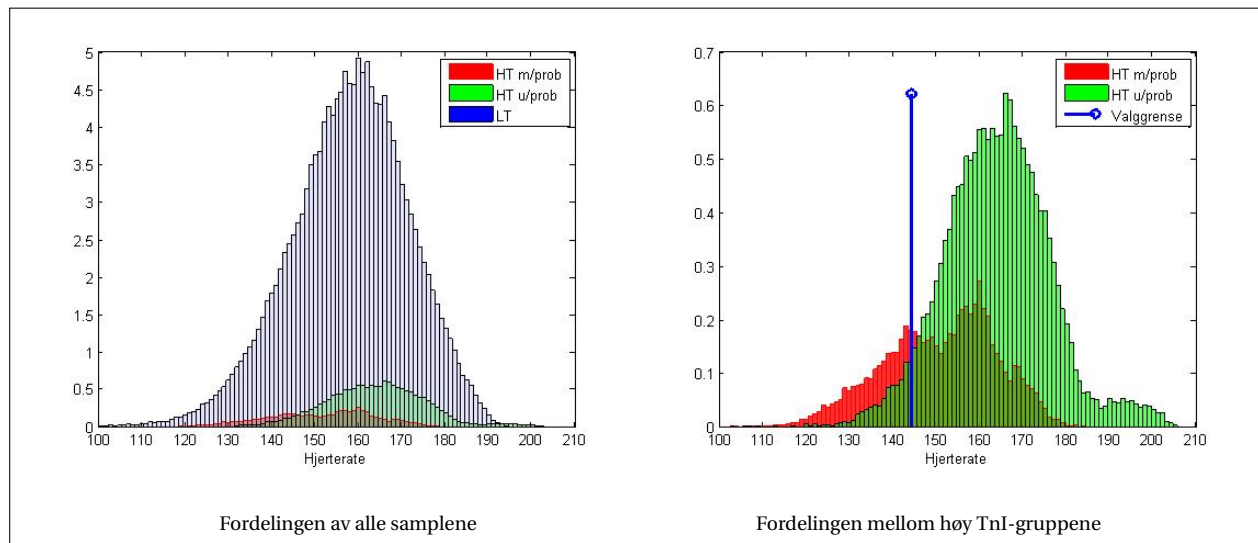
Ser man nærmere på resultatet fra [tabell 4.4](#) ser man at forskjellen mellom gruppene blir mindre og mindre utover i løpet. Dette kan man også se i boksplokkene i [figur 4.2](#). Pulsmålingene til de friske deltagerne er uforandret ved starten og slutten av løpet, mens pulsen til de syke deltagerne øker med nesten 10 bpm. Etter samtaler med Stein Ørn, overlege på hjerteavdelingen ved Helse Stavanger, er dette resultatet vanskelig å begrunne. Det er mulig at forklaringen er enkel, og at de syke deltagerne jobber hardere på sykkelen mot slutten av løpet.

I lignende studier ser man at de tar høyde for om deltagerne røyker, har høy BMI, lite treningsgrunnlag eller har andre sykdommer. Dette er fordi det kan svekke helsen og resultatet av prestasjonene til deltagerne. Dette er ikke blitt gjort i dette prosjektet på grunn av manglende informasjon fra deltagerne.

Om en slik studie skal gjennomføres igjen, sparer man mye tid på om alle deltagerne har lik pulsklokke, eller pulsklokke fra samme produsent. Dette vil gjøre at alle deltagerne har samme filtype og at man da trenger bare én innlesningsfunksjon i Matlab. Filene vil da logge de samme dataene, være likt samplet og formatert. Dette gjør at man ikke trenger å etterbehandle dataene før videre arbeid. Det har gått med mye tid på denne prosessen, noe som gjør at det blir mindre tid til analyse av dataene. Etter informasjon fra en av personene bak studien kom det frem at alle pulsfilene kan konverteres til *csv*-format. Da vil man også kun trenge en innlesningsfunksjon.

Det kunne være interessant å måle nivået av hemoglobin i blodet på deltagerne for å se om det er en mangel på røde blodceller. Oppgaven til de røde blodcellene er å frakte oksygenet rundt i kroppen og en mangel på disse kan føre til at hjertet ikke fungerer optimalt under fysisk aktivitet. Dette vil da sette begrensninger for treningskapasiteten til deltageren og ved veldig lave blodverdier er det en risiko å trene hardt. Mangel på røde blodceller assosieres med økt odds for lav hjerteratevariabilitet, som er direkte sammenkoblet med økt risiko for kardiologiske problemer. Det vil derfor være interessant å se om det er en sammenheng mellom hemoglobin i blodet og biomarkøren troponin I. Dette er et felt som er lite undersøkt, så fremtidige studier vil besvare dette.

Som man ser i rapporten er det forskjell mellom de tre forskjellige gruppene. Det var derfor interessant å se om det var mulig å lage en automatisk metode som klarte å skille mellom disse tre. Det ble forsøkt å lage en slik klassifiserer, men dette bydde på et problem; Apriori sannsynligheten for å være hjertesyk er svært liten. Til venstre i figuren under ser man alle de loggede samplene til alle deltagerne og fordelingen av de. Her ser man at gruppen med lav TnI overdøyer de andre på grunn av mye større apriori sannsynlighet. I figuren til høyre er gruppen med lav troponin fjernet og man ser nå på fordelingen mellom de to andre gruppene. Apriori sannsynligheten for å være hjertesyk er nå $\frac{8}{30} = 0.27$. Ser man på plottet, ser man at selv etter å ha fjernet gruppen med lav troponin vil det fortsatt oppstå mye feilklassifisering. Valggrensen i figuren ble utregnet ved hjelp av "Bayes Decision Rule". Men på grunn av dårlig statistikk og mye feilklassifiseringer ble denne metoden vraket.



Figur 5.1: Fordelingen av alle hjerteratemålingene for gruppene

Kapittel 6

Konklusjon

Målet med denne oppgaven var å se om det er mulig å skille mellom hjertesyke og friske personer ved hjelp av pulsklokke data. Alle personene har gjennomgått medisinske undersøkelser på forhånd av dette prosjektet, så tilstanden til hver deltager er kjent. Personene kan derfor deles inn i tre forskjellige grupper, de med høy TnI-verdi som er syke, de med høy TnI-verdi som er friske og de med lav TnI-verdi som er antatt friske. Disse tre gruppene analyseres mot hverandre for å se om det er noen forskjell mellom de. Her skiller gruppen med de hjertesyke deltagerne seg klart ut fra de andre ved å ha en lavere median og snittpuls gjennom nesten hele løpet. Selv etter at man har normalisert hjertemålingene med tanke på alder og kjønn, skiller gruppen med de hjertesyke deltagerne seg fortsatt ut. Når man ser nærmere på mindre, men samtidig mer interessante distanser i løpet, gjør man oppsiktsvekkende observasjoner. Forskjellen mellom gruppene minsker i de mindre krevende partiene og denne tendensen ser man gjennom hele løpet.

På grunn av de fremlagte resultatene og innholdet i denne rapporten ansees oppgaven derfor som fullført.

Kapittel 7

Referanser

- [1] Kaminsky LA, Arena R, Beckie TM et al. The importance of cardiorespiratory fitness in the United States: The need for a national registry: A policy statement from the American Heart Association. *Circulation*, published online January 7., 2013.
- [2] Stein Ørn, (2014), NEEDED_2014_Protokoll_Stein, Hjerterefrekvens og hjertemuskelskade ved mosjonskonkurranse – en analyse av data fra pulsklokker til deltakere i Nordsjørippet 2014 [3] [http://www.nordsjorippet.no/om-oss/\(15/05/2015\)](http://www.nordsjorippet.no/om-oss/(15/05/2015))
- [4] Shave R, Baggish A, George K et al. Exercise-induced cardiac troponin elevation. Evidence, mechanisms, and implications. *J Am Coll Cardiol*. 2010;56:169-176
- [5] <http://www.unn.no/utredning/category26049.html> (11/02/15)
- [6] Dilsizian, Vasken, Pohost, Gerald, (2010), *Cardiac CT, PET and MR*, Second Edition, Wiley-Blackwell (11/02/15)
- [7] Jonas Hallén, (2011), *Cardiac troponin I and electrocardiogram for estimation of infarct size and for risk stratification in patients with STElevation myocardial infarction treated with primary percutaneous coronary intervention*, Doctoral Thesis (10/02/15)
- [8] <http://tidsskriftet.no/pdf/pdf2014/1348.pdf> (10/02/15)
- [9] http://brukerhandboken.no/index.php?action=showtopic&topic=99003a0a12cc3d6dbdf5&book_request=biokjemi (11/02/15)
- [10] <http://tidsskriftet.no/pdf/pdf2011/2226.pdf> (10/02/15)
- [11] http://brukerhandboken.no/index.php?action=showtopic&topic=99003a0a12cc3d6dbdf5&book_request=biokjemi (11/02/15)
- [12] <http://legeforeningen.no/pagefiles/197014/hjerteforum%204-2014%209%20esc%20i%20barcelona%20-%20referater.pdf> (06/06/2015)
- [13] http://en.wikipedia.org/wiki/File_format (15/03/2015)
- [14] Petter Andersen - Spesialist ved Polar Elektro Norge AS - Kontaktet via telefon og mail
- [15] Adaszewski, Stanislaw, (2014), *Mynodbcsv: Lightweight zero-config database solution*

for handling very large CSV files

- [16] <http://ridewithgps.com/help/gpx-tcx-fit-and-klm-files> (15/03/2015)
- [17] http://en.wikipedia.org/wiki/Training_Center_XML (15/03/2015)
- [18] <https://www.matematikkfessor.dk/lessons/boksplot-og-kassediagram-247>
(15/03/2015)
- [19] John G. Proakis, Dimitris K. Manolakis, (2014), Digital Signal Processing, Fourth Edition, Pearson New International Edition (15/03/2015)
- [20] Gunnar G. Løvås, (2004) Statistikk for universiteter og høyskoler, 2. utgave, Universitetsforlaget, ISBN 978-82-15-00224-8
- [21] <http://folk.uio.no/hanssto/old/kvantmet/Statistiskgeneralisering.pdf> (06/06.2015)
- [22] <http://www.portamedicus.no/ArticleDisplay.aspx?MenuId=4260&ContentId=5982>
(01/06/2015)
- [23] Michael S. Lauer, Peter M Okin, Martin G. Larson, Jone C. Evans, Daniel Levy, Impaired Heart Rate Response to Graded Exercise (09/06/2015).
- [24] Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol. 2001;37:153–156.

Tillegg A

Forklaring av skript

For å lage filer1-cellen I slutten av dette prosjektet ble det oppdaget at alle filtypene kan konverteres til csv-filer. Nå er skriptene laget slik at man bruker først `Hovedfil.m` for å lese inn alle csv-filene. Så kjører man `readingDoc.m` for å lese inn all informasjon fra Excel-arket. Her deler man deltagerne inn i de tre gruppene. Så bruker man `SamlerExcelogKlokkene.m` Denne samler informasjonen fra Excel-arket og pulsfilene. Så bruker man skriptet `allHRandInterpol.m` som interpolerer og normaliserer hjertemålingene. Så må man kjøre skriptet `sletteDeltagere.m` som fjerner alle deltagerne som har syklet for kort eller langt. Etter dette er cellen komplett. Det er fortsatt anbefalt å bruke den vedlagte *filer1*-cellen, da deler av denne er manuelt fikset på, for å ta vare på de små deltagergruppene.

Hovedfil.m Dette er hovedskriptet som tar for seg innlesningen. Først finner den alle pulsfiler med navn og filtype. Disse sorteres så inn i grupper, før de sendes videre til sine respektive innlesningsfunksjoner. Alle input som trengs i de forskjellige innlesningsfunksjonene blir laget i dette skriptet.

typelist.m Dette er en funksjon som blir automatisk kjørt inne i skriptet *hovedfil.m*. Funksjonen ser slik ut `d=typelist(pwdir,d)` og krever to input.

pwdir - Gir filplasseringen til alle pulsdataene.

d - En vektor med alle filene

Det er denne funksjonen som henter navnet på alle filene i mappen og filformatet og legger dem i en matrise. Denne matrisen er utgangsvariabelen fra funksjonen.

readingPwx.m Dette er innlesningsfunksjonen for alle *pwx*-filene. Funksjonen ser slik ut

`pxer=readingPwx(d21,type,sorter,pwdir)` og krever da fire input

d21 - En celle med navnet på alle filene og filtypene

type - En vektor med navnet på filformatene

sorter - En celle som inneholder grupperingen av filformatene.

pwdir - Peker til mappen hvor pulsklokkefilene ligger

Funksjonen åpner hver fil, går linje for linje gjennom hele og henter ut all informasjon. Denne legges inn i forskjellige vektorer. Så endres tidsformatet i filene. Alle loggingene i pwx-formatet må konverteres fra string til tall. Når dette er ferdig settes vektorene sammen i en stor matrise som blir utgangsvariabelen fra funksjonen.

readingXml.m Dette er innlesningsfunksjonen for alle *xml*-filene. Funksjonen ser slik ut

```
xmler=readingXml(d21,type,sorter,pwdir) og har de samme inputene som readingPwx.
```

Funksjonen åpner hver fil, går linje for linje gjennom hele og henter ut all informasjon. Denne legges inn i forskjellige vektorer. Tidsformatet i disse filene trengs ikke å endres, da de allerede er logget i sekunder. Alle loggingene i xml-formatet må konverteres fra string til tall. Når dette er ferdig settes vektorene sammen til en stor matrise som blir utgangsvariabelen fra funksjonen.

readingCsv.m Dette er innlesningsfunksjonen for alle *csv*-filene. Funksjonen ser slik ut

```
csver=readingCsv(d21,type,sorter,pwdir) og har de samme inputene som readingPwx.
```

Funksjonen benytter seg av en ferdiglaget funksjon i Matlab som leser inn all informasjon og legger den i matriser. Dette blir utgangsvariabelen fra funksjonen.

readingTcx.m Dette er innlesningsfunksjonen for alle *tcx*-filene. Funksjonen ser slik ut

```
tcxer=readingTcx(d21,type,sorter,pwdir) og har de samme inputene som readingPwx.
```

Funksjonen åpner hver fil, går linje for linje gjennom hele og henter ut all informasjon. Denne legges inn i forskjellige vektorer. Så endres tidsformatet i filene. Alle loggingene i tcx-formatet må konverteres fra string til tall. Når dette er ferdig settes vektorene sammen i en stor matrise som blir utgangsvariabelen fra funksjonen.

readingHrm.m Dette er innlesningsfunksjonen for alle *hrm*-filene. Funksjonen ser slik ut

```
hrmer=readingHrm(d21,type,sorter,pwdir) og har samme inputene som readingHrm.
```

Funksjonen gjør det samme som de andre innlesningsfilene. Hrm-filene logger kun hjerteraten og et sammendrag av løpet. Metoden ble laget fordi en av de åtte hjertesyke hadde dataene i dette filformatet. Selv etter innlesning ble deltagerne med hrm-data

slettet. Denne metoden er laget slik at man kan kalkulere samplefrekvensen i klokken, men man kan ikke sette den i sammenheng med distanse i løpet, og er derfor verdiløs.

readingDoc.m Dette skriptet er kjørbart slik det er og leser all informasjon i det vedlagte regnearket. Etter all informasjon er lest inn, deles det inn i tre grupper, avhengig av hva som står i kolonnen "Anmerkninger".

histog.m Denne funksjonen fremstiller histogram og plot av hjerteratefordelingen i de forskjellige gruppene. Funksjonen ser slik ut $[\mu, \sigma, db]=\text{histog}(ta, bw, str)$ og krever tre input

ta - Dette må være mat-filen *filer1*

bw - Velger størrelsen på intervaller man vil se på i hjerteraten / oppløsningen

str - Velger man 'N' normaliseres hjertereverdiene til deltagerne. Alle andre bokstaver eller kombinasjoner gir vanlige hjerterateverdier.

I tillegg til plottene regner funksjonen ut gjennomsnittsverdien og variansen til gruppene. Den kalkulerer også valggrensen mellom de to gruppene med høy troponin.

hoydeBehandling.m Denne funksjonen starter flere andre funksjoner i Matlab. Parametrene funksjonen krever er:

filer - Dette må være mat-filen *filer1*

valg - Velger hvilke distanser man ønsker å se på i løpet

val - Velger hvor stor avstand mellom hvert generert punkt

diff - Hvor nær denne avstanden punktet må ligge

ta - Velger $ta='N'$ om man ønsker å bruke normaliserte hjertemålinger

maksHR.m Denne funksjonen kjøres fra *hoydeBehandling.m* og deler opp alle pulsmålingene i forskjellige intervaller.

y - En celle utregnet i *hoydeBehandling.m*

str - Velger hvilken funksjon som kjøres. Kan velge mellom 'HRinndeling' og 'HRinndelingNorm'

pf - Man kan dele opp histogrammene for å se på fordelingen av hjertemålingene. Man kan velge alt i fra 'abcdefghi'. Velger man 'cdefg' ser man kun på de midterste histogrammene.

FungerendeDistanseTabell.m Denne funksjonen også kjøres fra *hoydeBehandling.m*. Denne deler opp og finner hjertemålingene til hver deltager i nærheten av de valgte punktene *diff*.

y - En celle utregnet i *hoydeBehandling.m*

val - Se *hoydeBehandling.m*

diff Se `hoydeBehandling.m`

string Velger hvilken case som skal kjøres i funksjonen. '1' - velges om man skal kun kjøre denne filen. '2' - velges om man kjører filen fra `hoydeBehandling.m`

c - Starten av intervallet man ser på. Denne kalkuleres i funksjonen `hoydeBehandling.m`

kvartil1.m Denne plotter median, øvre og nedre kvartil til alle gruppene. Denne funksjonen kjøres også fra `hoydeBehandling.m`

HR - Her brukes utgangsvariabelen fra `FungerendeDistanseTabell.m`

print - Velger hvilken case som skal kjøres inne i funksjonen. '1' plotter median og kvartiler til alle gruppene. '2' plotter kun median til alle gruppene. '3' plotter de hjertesyke mot de friske.

val - Se `val` fra `hoydeBehandling.m`

plotter.m Dette er ingen funksjon, men et skript med flere deler. Første delen plotter høydekurven til løpet. De neste tre plotter boksploottene på forskjellige måter. Kjører man del nr 3 i dette skriptet kan man bruke utgangsvariablene i skriptet under.

statistiskanalyse.m Dette er ingen funksjon, men et skript som fortar en statistisk analyse av utgangsvariablene fra boksploottene ovenfor.

sletteDeltagere.m Dette er ingen funksjon, men en samling av skript for å slette deltagere som har syklet for kort eller langt.

SamlerExcelogKlokkene.m Dette er ingen funksjon, men et skript som samler informasjonen fra regnearket og pulsklokkene.

allHRandInterpol.m Er ikke en funksjon, men et skript som interpolerer og normaliserer hjertemålingene

sletteDeltagere.m Dette er et skript som sletter alle deltagerne som har enten syklet kortere enn 89 km eller lengre enn 91 km. Disse parametrene kan man endre på selv.