



**Høgskolen  
i Innlandet**

Campus Lillehammer

**Kandidatnr: 507**

**Mappe eksamen**

Treningsfysiologi.

IDR4000-1 24H- Kvantitativ metode og statistikk

**22.11.2024**

Antall ord: 18108

# Innholdsfortegnelse

0.1	Forord . . . . .	2
<b>1</b>	<b>Oppgave 1: Reliabilitet og verktøy for reproduserbar vitenskapelig data</b>	<b>3</b>
1.1	Introduksjon . . . . .	3
1.2	Metode og protokoll . . . . .	3
1.3	Resultater . . . . .	4
1.4	Diskusjon . . . . .	8
1.5	Konklusjon . . . . .	9
<b>2</b>	<b>Oppgave 2: Regresjonsmodeller, forutsi fra data</b>	<b>10</b>
2.1	Introduksjon . . . . .	10
2.2	Del 1: Laktatterskel-analyse . . . . .	10
2.3	Del 2: Forutsi størrelser på DNA-fragmenter. . . . .	14
2.4	Del 3: Tolkning av regresjonstabell . . . . .	22
<b>3</b>	<b>Oppgave 3: Reliabilitet og verktøy for reproduserbar vitenskapelig data</b>	<b>26</b>
3.1	Introduksjon . . . . .	26
3.2	Spørsmål 1 . . . . .	27
3.3	Spørsmål 2 . . . . .	29
3.4	Spørsmål 3 . . . . .	29
3.5	Spørsmål 4 . . . . .	30
3.6	Spørsmål 5 . . . . .	31
3.7	spørsmål 6 . . . . .	31
3.8	Spørsmål 7 . . . . .	33
3.9	Spørsmål 8 . . . . .	34
<b>4</b>	<b>Oppgave 4: Study designs</b>	<b>36</b>
4.1	Introduksjon . . . . .	36
4.2	Studie design . . . . .	37
4.3	Statistiske modeller . . . . .	39
4.4	Diskusjon: . . . . .	41
4.5	Konklusjon: . . . . .	43
<b>5</b>	<b>Oppgave 5: Analyzing repeated measures experiments</b>	<b>44</b>
5.1	Introduksjon . . . . .	44
5.2	Metode . . . . .	46

5.3	Data analyse . . . . .	48
5.4	Resultat . . . . .	51
5.5	Diskusjon . . . . .	52
5.6	Konklusjon . . . . .	53
5.7	Statistisk programvare . . . . .	54
5.8	Etiske hensyn: . . . . .	54
<b>6</b>	<b>Oppgave: Filosofihistorie</b>	<b>55</b>
6.1	Introduksjon . . . . .	55
6.2	Spørsmål 1: . . . . .	55
6.3	Spørsmål 3: . . . . .	56
<b>7</b>	<b>Oppgave: Protein Analyse</b>	<b>60</b>
7.1	Introduksjon . . . . .	60
7.2	Innledning . . . . .	60
7.3	Metode . . . . .	60
7.4	Resultater og diskusjon . . . . .	64
7.5	Konklusjon . . . . .	68
	<b>referanser</b>	<b>69</b>

## 0.1 Forord

Denne mappeeksamenen representerer mitt arbeid i faget IDR4000 Kvantitativ metode og statistikk, en del av masterprogrammet i treningsfysiologi. Oppgavene demonstrerer anvendelsen av kvantitative metoder og statistiske analyser innen treningsfysiologi, med fokus på reproduserbarhet i vitenskapelig arbeid. Ikke alle oppgavene klarte jeg å løse, men jeg håper at denne mappeeksamenen likevel demonstrerer min forståelse av kvantitative metoder og statistikk, samt min evne til å anvende disse verktøyene på relevante problemstillinger innen treningsfysiologi.

Gjennom arbeidet med denne mappeeksamenen har jeg tilegnet meg verdifull kunnskap, særlig innen koding som var et helt nytt felt for meg. Selv om jeg støtte på noen tekniske utfordringer underveis som jeg ikke klarte å løse fullt ut, håper jeg at helheten i arbeidet demonstrerer min faglige utvikling og forståelse for kvantitative metoder i treningsfysiologi.

For å sikre full transparens og reproduserbarhet, er all kode, dataanalyse og resultater gjort tilgjengelig via et GitHub-repository. Dette repositoryet inneholder alle R-skript, datasett og supplerende materiale brukt i oppgavene.

GitHub-repositoryet for dette prosjektet kan finnes på:

[Mitt GitHub-repository](#)

# 1 Oppgave 1: Reliabilitet og verktøy for reproduserbar vitenskapelig data

## 1.1 Introduksjon

Dette er oppgave 1: “Reliabilitet og verktøy for reproduserbar vitenskapelig data” i faget IDR4000-1 24H Kvantitativ metode og statistikk for master i treningsfysiologi. Denne oppgaven var opprinnelig en gruppeoppgave, men jeg har rekonstruert den etter tilbakemelding fra fagansvarlig. Hensikten med denne oppgaven er å undersøke reliabiliteten av VO<sub>2</sub>max-tester utført i fysiologilaboratoriet.

For at slike tester skal være nyttige i forsknings- og treningssammenheng, er det avgjørende at de er pålitelige og reproduserbare. I denne oppgaven fokuserer vi spesielt på reliabiliteten av to fysiologiske markører: laktatmaksimum (la.max) og respiratorisk utvekslingsforhold (RER.max). Disse markørene gir verdifull innsikt i den fysiologiske responsen under maksimal anstrengelse. Laktat er en indikator på anaerob energiomsetning, vi måler laktatnivåer i blodet etter svært intens trening for å få en indikasjon på hvor stor syreopphopning (acidose) kroppen har opplevd under treningen (Tanner & Gore, 2012). Mens RER er en måling på mengden karbondioksid (CO<sub>2</sub>) produsert i forhold til mengden oksygen (O<sub>2</sub>) som brukes (Ramos-Jiménez *et al.*, 2008).

## 1.2 Metode og protokoll

Vi har gjennomført gjentatte VO<sub>2</sub>max-tester over fire forskjellige dager. Vi har forsøkt å standardisere flere variabler, inkludert treningsnivå og matinntak dagen før testene, for å minimere påvirkningen av konfunderende faktorer. Som både Halperin *et al.* (2015), Hopkins (2000), og Tanner & Gore (2012) påpeker, er det avgjørende å bruke standardiserte protokoller for å oppnå pålitelige resultater i fysiologiske tester.

Vi gjennomførte en VO<sub>2</sub>maks-test som en trappetrinnstest på ergometersykkel. Testen ble utført over fire forskjellige dager for å vurdere reliabiliteten av målingene. De to første testdagene var påfølgende dager, mens de to siste hadde en hviledag imellom. Før hver test ble utstyret grundig klargjort og kalibrert. Dette inkluderte oppstart av laktatmaskin, klargjøring av munnstykke og slanger, samt kalibrering av oksygenanalysatoren. Gasskalibrering og

volumkalibrering ble utført med krav til nøyaktighet, hvor vi aksepterte en feilmargen på maksimalt 2% for gass og 0,2% for volum.

For hver deltaker opprettet vi en ny profil i Vyntus- og Lode-systemene. Deltakerne ble veid i minimalt med klær (trakk fra 300g for de klærne som var igjen), og sykkelen ble innstilt for hver enkelt deltager. Vi forberedte også nødvendig utstyr for laktatmåling.

Selve testen startet med en individuelt tilpasset wattbelastning basert på deltakerens fysiske form og erfaring med sykling. Motstanden økte deretter med 20W hvert minutt inntil utmattelse, eller når tråkkfrekvensen falt under 60 RPM. Vi registrerte målinger av VO<sub>2</sub>maks hvert 30. sekund gjennom hele testen. Under testen ble deltakerne instruert om å holde en tråkkfrekvens mellom 80-100 RPM. Vi målte kontinuerlig puls og registrerte opplevd anstrengelse med bruk av Borg-skala umiddelbart ved endt test. Ett minutt etter avsluttet test tok vi laktatprøve.

For å sikre standardisering og redusere variasjoner mellom testene, implementerte vi flere tiltak. Deltakerne ble bedt om å opprettholde likt matinntak og koffeinkonsum før hver test. Vi forsøkte også å standardisere trening og søvnmonster i forkant av testene. Dette viste seg utfordrende å kontrollere fullstendig, da vi alle er aktive elever med travel og noe uforutsigbar hverdag.

Hver deltaker hadde samme testleder ved alle fire testene. Dette for å sikre konsistens i verbal instruksjon og oppmuntring, og andre variabler som at kjønn på tester kan påvirke prestasjonen til testdeltager. Etter hver test ble all data nøye registrert og overført til et standardisert Excel-dokument for videre analyse. Dette inkluderte fysiologiske målinger, subjektive vurderinger og miljøfaktorer som temperatur og luftfuktighet i laboratoriet.

Denne detaljerte protokollen ble utviklet og fulgt for å maksimere reliabiliteten i våre målinger, i tråd med anbefalinger fra Halperin *et al.* (2015), Hopkins (2000), og Tanner & Gore (2012) om standardisering i fysiologiske tester.

## 1.3 Resultater

Her presenterer jeg resultatene fra testene våre. Jeg starter med en tabell over demografiske variablene i fra de 4 testene. Deretter en tabell over laktat målinger og RER målinger over de samme testene. Deretter reliabilitetsmål og til slutt individuelle forskjeller.

### 1.3.1 Demografiske data og deltakelse

### 1.3.2 laktat og RER

Antall deltakere varierte fra 14 i de to første testene (t1 og t2) til 11 i t3 og 8 i t4. Gjennomsnittsalderen for deltakerne økte fra 25.7 år (t1 og t2) til 29.1 år (t4), med en merkbar økning i

Deltagerkarakteristikker ved ulike testtidspunkt

Tidsperiode	Deltagerkarakteristikker		
	Alder (år)	Høyde (cm)	Kroppsmasse (kg)
t1 (n = 14)	25.7 ( $\pm 3.6$ )	179.8 ( $\pm 7.6$ )	79.2 ( $\pm 11.9$ )
t2 (n = 14)	25.7 ( $\pm 3.6$ )	179.8 ( $\pm 7.6$ )	79.3 ( $\pm 11.8$ )
t3 (n = 11)	26.8 ( $\pm 3.7$ )	178.7 ( $\pm 7.9$ )	80.4 ( $\pm 11.9$ )
t4 (n = 8)	29.1 ( $\pm 7.2$ )	174.6 ( $\pm 6.2$ )	74.7 ( $\pm 12.3$ )
Gjennomsnitt (n = 11.8)	26.5 ( $\pm 4.5$ )	178.7 ( $\pm 7.5$ )	78.7 ( $\pm 11.7$ )
Verdier er presentert som gjennomsnitt ( $\pm$ standardavvik)			

Tidsperiode	Fysiologiske målinger	
	Laktat (mmol/L)	RER
t1 (n = 14)	12.2 ( $\pm 2.76$ )	1.14 ( $\pm 0.03$ )
t2 (n = 14)	12.3 ( $\pm 2.91$ )	1.15 ( $\pm 0.04$ )
t3 (n = 11)	13.64 ( $\pm 2.5$ )	1.18 ( $\pm 0.05$ )
t4 (n = 8)	13.82 ( $\pm 2.31$ )	1.19 ( $\pm 0.03$ )
Gjennomsnitt (n = 11.8)	12.99 ( $\pm 2.62$ )	1.16 ( $\pm 0.04$ )
Verdier er presentert som gjennomsnitt ( $\pm$ standardavvik)		

standardavvik for alder i t4 ( $\pm 7.2$  år). Gjennomsnittshøyden og -vekten forble relativt stabile over de fire testtidspunktene, med små variasjoner til tross for det ujevne antall deltakere.

### 1.3.2.1 Laktatmålinger

Gjennomsnittlige laktatverdier viste en gradvis økning over testtidspunktene, fra 12.2 mmol/L ( $\pm 2.76$ ) i t1 til 13.82 mmol/L ( $\pm 2.31$ ) i t4. Standardavviket for laktatmålingene varierte mellom 2.31 og 2.91 mmol/L over de fire testene, noe som indikerer en viss variabilitet i målingene.

### 1.3.2.2 RER-målinger

RER-verdiene viste også en svak stigende trend over testtidspunktene, fra 1.14 ( $\pm 0.03$ ) i t1 til 1.19 ( $\pm 0.03$ ) i t4. Standardavviket for RER-målingene var konsistent lavt, og varierte mellom 0.03 og 0.05 over de fire testene.

### 1.3.3 Reliabilitetsmål

For å vurdere reliabiliteten av metoden, beregnet vi variasjonskoeffisienten (CV) for både laktat og RER basert på de to første testene (t1 og t2):

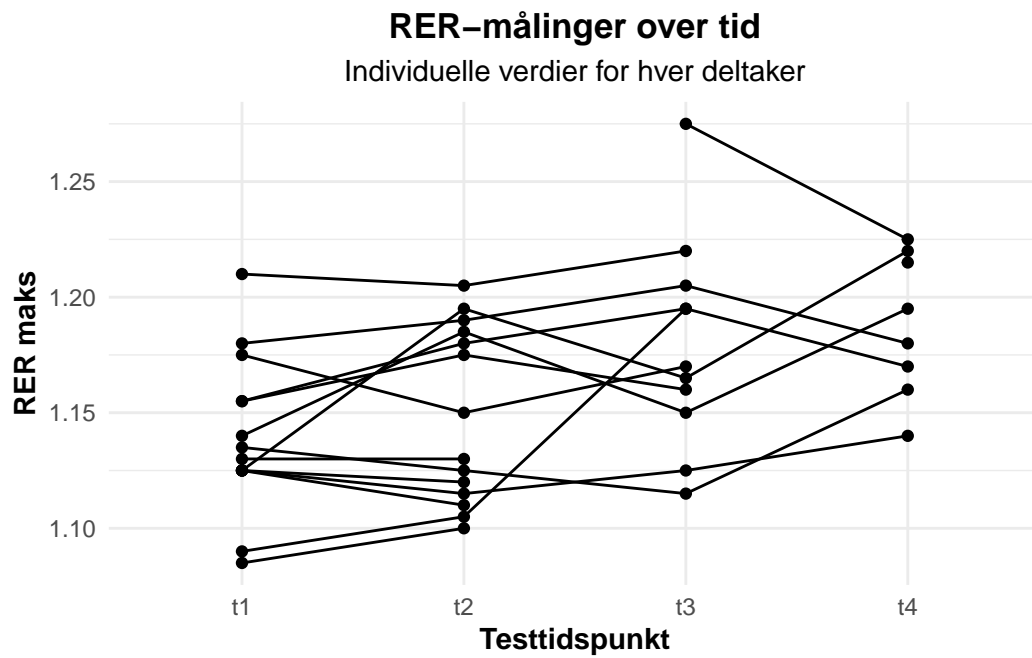
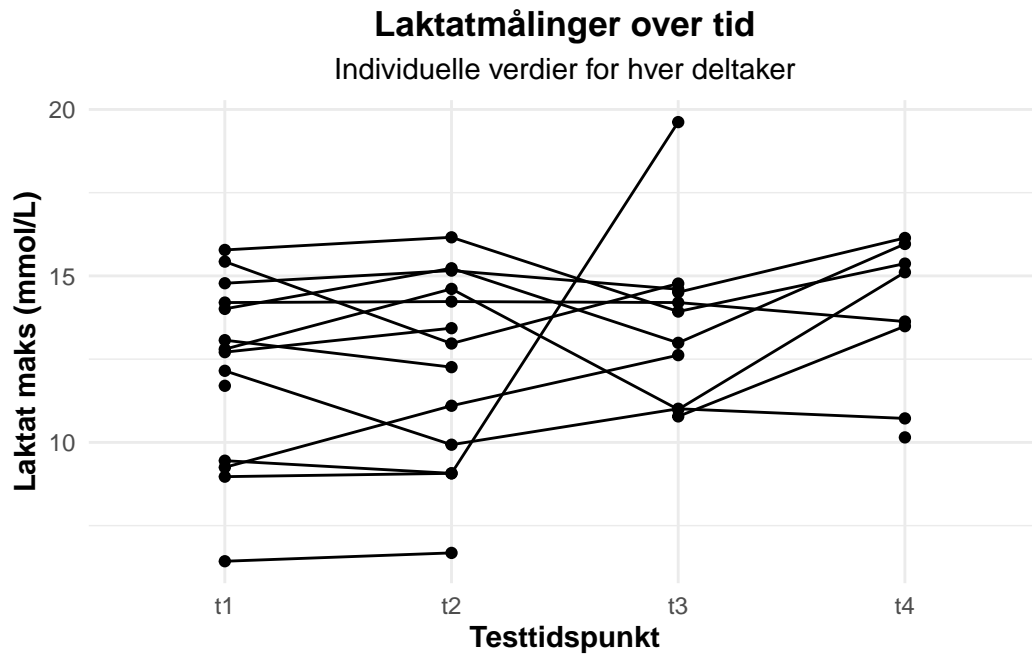
$$\text{Laktat} = 7.6\%$$

$$\text{RER} = 1.6\%$$

Disse resultatene indikerer at RER-målingene viser betydelig høyere reliabilitet sammenlignet med laktatmålingene. Dette er i tråd med forventningene basert på hva Ramos-Jiménez *et al.* (2008) sier om RER målinger, som antyder at RER-målinger generelt er mer stabile enn laktatmålinger under repeterte tester. Hopkins (2000) beskriver CV som en nyttig måte å uttrykke typisk feil på, spesielt når feilen har en tendens til å øke med størrelsen på målingen. Selv om Hopkins (2000) ikke spesifiserer eksakte grenser for hva som er akseptabelt, antyder han at lavere CV-verdier generelt indikerer bedre reliabilitet. I vårt tilfelle vil vi si RER-målingene faller innenfor en lavere CV-verdi, noe som tyder på høyere reliabilitet sammenlignet med laktatmålingene.

Andre mål på reliabilitet som kunne vært brukt inkluderer intraclass correlation coefficient (ICC) og limits of agreement, som diskutert av Hopkins (2000). Disse metodene kunne gitt ytterligere innsikt i konsistensen av målingene over tid og mellom individer.

### 1.3.4 Individuelle forskjeller



Individuelle variasjoner i laktat- og RER-målinger over de fire testtidspunktene er illustrert i grafene over. Grafene viser at mens noen deltakere hadde relativt stabile verdier over tid,



opplevde andre mer markante endringer. For laktatmålinger var det større variabilitet mellom individer og over tid sammenlignet med RER-målingene, som generelt viser hakket mer konsistente mønstre.

## 1.4 Diskusjon

Denne oppgaven undersøkte reliabiliteten av data fra VO<sub>2</sub>max-tester vi har utført, med spesielt fokus på laktatmålinger og respiratorisk utvekslingsforhold (RER) over fire testtidspunkter. Resultatene viser varierende grad av reliabilitet for disse fysiologiske markørene

### 1.4.1 Reliabilitet av laktatmålinger

Variasjonskoeffisienten (CV) for laktatmålinger på 7.6% indikerer en moderat grad av reliabilitet. Hopkins (2000) argumenterer for at CV burde være så lav som mulig og bruker forskjellige eksempler på hvordan CV kan tolkes over og under 5%. Jeg tolker det som at det er ønskelig at CV burde være under 5%. Våre funn for laktat overstiger 5%, noe som tyder på at laktatmålinger kan være mer variable og mindre pålitelige enn ønskelig for presis fysiologisk testing. Dette samsvarer med observasjonen av økende gjennomsnittlige laktatverdier og varierende standardavvik over de fire testene. Den observerte variabiliteten i laktatmålinger kan skyldes flere faktorer.

Halperin *et al.* (2015) påpeker viktigheten av å kontrollere for konfunderende variabler som kan påvirke fysiologiske målinger. Faktorer som ernæringsstatus, treningsstatus dagen før testing, omgivelsene som antall observatører eller kjønn på observatør samt temperatur, luftfuktighet m.m. Det viste seg å være vanskelig å standardisere alle slike konfunderende faktorer, spesielt trening og søvn for flere i gruppen, og jeg vil tro at de andre gruppene også hadde sine utfordringer. Det kan også hende vi som uerfarne studenter har hatt unøyaktige målinger da blodprøve ble tatt manuelt.

### 1.4.2 Reliabilitet av RER-målinger

I motsetning til laktatmålingene viste RER-målingene betydelig høyere reliabilitet, med en CV på 1.6%. Dette faller godt under 5% som Hopkins (2000) bruker som eksempel på god reliabilitet. Den lave CV-verdien for RER indikerer at denne markøren kan være et mer stabilt og pålitelig mål på metabolsk funksjon under maksimal anstrengelse.

Den høye reliabiliteten observert for RER kan delvis forklares av målingens natur. RER er et direkte mål på forholdet mellom produsert CO<sub>2</sub> og konsumert O<sub>2</sub>, og er mindre påvirket av kortvarige fysiologiske svingninger sammenlignet med laktatkonsentrasjoner i blodet (Ramos-Jiménez *et al.*, 2008). I tillegg blir den automatisk målt av en maskin og er ikke utsatt for

menneskelige feil som blodtaking for laktat er. Dette gjør RER til en potensielt mer robust markør for vurdering av metabolsk funksjon under høyintensitets utholdenhetstrening.

### **1.4.3 Begrensninger**

En viktig begrensning i vår oppgave er det synkende antallet deltakere over de fire testtidspunktene, fra 14 i de første to testene til bare 8 i den siste. Dette frafallet kan ha påvirket reliabilitetsestimatene ytterligere, spesielt for de senere testene. CV var bare beregnet ut fra de to første testdagene. Hadde dette vært en reell studie, hadde nok flere av oss blitt ekskludert ut ifra at vi ikke kunne følge protokollen mellom testingen. Selv om vi forsøkte å standardisere testprosedyrene, kan subtile variasjoner i disse faktorene ha bidratt til variabilitet i resultatene. Det var for eksempel andre som brukte de forskjellige laboratoriene parallelt med oss og vi kunne ikke kontrollere trafikken mellom laboratoriene. Dette var en lærerik opplevelse på hvor vanskelig det er å følge protokoller og gjøre det som trengs for å få reliable målinger.

## **1.5 Konklusjon**

Våre funn understreker viktigheten av å vurdere reliabiliteten av fysiologiske markører brukt i idrettsfysiologisk testing. Mens RER viste seg å være en høyst pålitelig markør, i det minste i vårt forsøk, indikerer den moderate reliabiliteten av laktatmålinger i vårt forsøk et behov for forsiktighet ved tolkning av disse verdiene. Ved mer øvelse og mer erfaring fra testerne samt strengere oppfølging av protokoller før testing for de som blir testet, kan nok reliabiliteten til laktat også forbedres.

## 2 Oppgave 2: Regresjonsmodeller, forutsi fra data

### 2.1 Introduksjon

Dette er oppgave 2: “Regresjonsmodeller, forutsi fra data” i faget IDR4000-1 24H Kvantitativ metode og statistikk for master i treningsfysiologi. Oppgaven var opprinnelig en gruppeoppgave, men er nå rekonstruert for mappeeksamen, da jeg så vi hadde flere feil i den originale innleveringen.

Hensikten med denne oppgaven er å utforske og anvende ulike statistiske metoder og regresjonsmodeller innen treningsfysiologi og molekylærbiologi. Oppgaven er delt inn i tre hovedseksjoner: laktatterskel-analyse, prediksjon av DNA-fragmentstørrelser, og tolkning av en regresjonstabell basert på et hypertrofi-datasett. Gjennom disse delene vil vi demonstrere vår forståelse av regresjonsanalyse, datavisualisering og statistisk tolkning innen relevante områder for treningsfysiologi. Målet er å vise hvordan disse metodene kan anvendes for å besvare spesifikke forskningsspørsmål og tolke resultater på en meningsfull måte.

### 2.2 Del 1: Laktatterskel-analyse

#### 2.2.1 Introduksjon

Laktatterskel er et viktig fysiologisk konsept innen idrettsvitenskap og treningslære, som gir verdifull innsikt i utøveres aerobe kapasitet og treningsintensitet. Under kortvarig supramaksimal trening er det ikke tilstrekkelig oksygen tilgjengelig til å ta imot hydrogenet som produseres under glykolysen, og som et resultat dannes laktat (Tanner & Gore, 2012). Tidligere antok man at laktatproduksjon var årsaken til at musklene mistet styrke og kraft under anaerob trening. Nå er det generelt forstått at laktat i seg selv ikke er grunnen til at musklene mister styrke og kraft (Tanner & Gore, 2012). Laktat er også involvert i andre prosesser, som bufferkapasitet, og Tanner & Gore (2012) nevner flere faktorer som påvirker tretthet i musklene. I denne oppgaven vil jeg imidlertid fokusere spesifikt på laktatterskelenes rolle i vurderingen av treningsintensitet og utholdenhetsprestasjon.

Et individ med økt evne til å produsere energi ved anaerob glykolyse, og samtidig motstå de negative effektene som fører til utmattelse, vil demonstrere en økt anaerob kapasitet. Derfor er måling av laktatterskel fortsatt en populær metode for å vurdere effekten av trening, programmere treningsintensitet og forutsi prestasjon (Tanner & Gore, 2012).

Denne analysen fokuserer på å beregne og sammenligne to vanlige laktatterskler: 2 mmol/L og 4 mmol/L. Tanner & Gore (2012) påpeker at mens 2 mmol/L og 4 mmol/L er vanlige referansepunkter, kan den faktiske terskelen variere mellom individer, og noen forskere argumenterer for mer individualiserte terskelverdier.

Målet med oppgaven er å undersøke reliabiliteten til disse tersklene i “cyclingstudy” og vurdere deres praktiske anvendelse i treningssammenheng.

### 2.2.2 Metode

For å gjennomføre denne analysen ble datasettet “cyclingstudy” fra “exscidata”-pakken benyttet. Dette datasettet inneholder informasjon om laktatkonsentrasjoner ved ulike treningsintensiteter for en gruppe syklister. Analysen ble utført ved hjelp av R og flere pakker, inkludert tidyverse, ggplot2 og knitr.

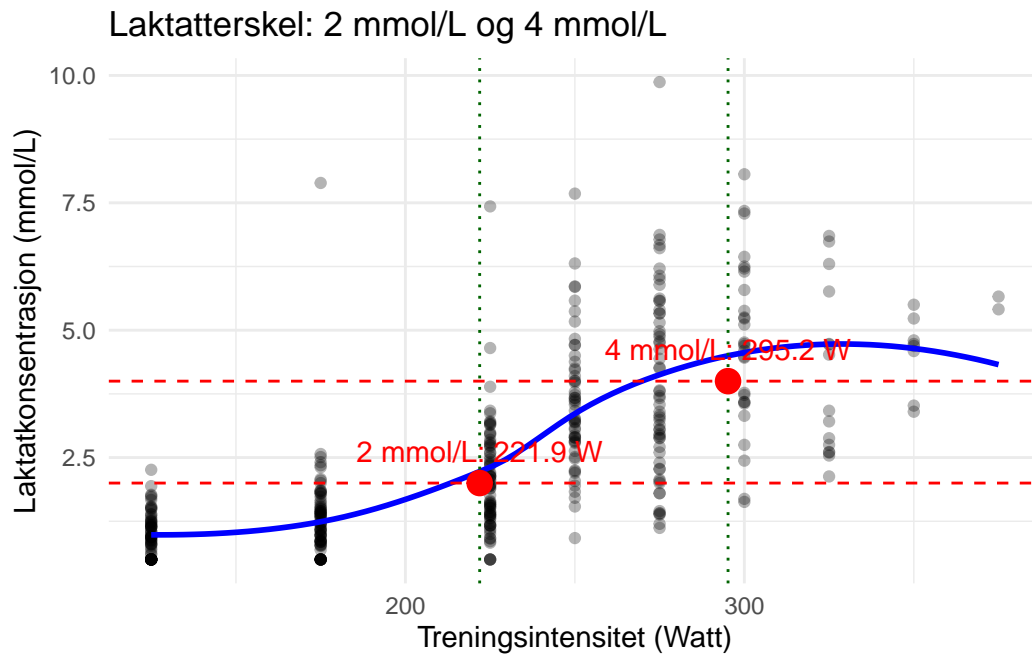
Dataene ble først bearbeidet og omformet fra bredt til langt format ved hjelp av tidyverse-funksjoner. Et scatterplot ble deretter laget med ggplot2 for å visualisere forholdet mellom treningsintensitet (watt) og laktatkonsentrasjon, med interpolerte kurver for å estimere watt-verdiene som korresponderer med 2 mmol/L og 4 mmol/L laktatkonsentrasjoner.

For å beregne gjennomsnittlige terskelverdier og typisk feil ble det utviklet en funksjon som anvendte interpolasjon. Resultatene ble presentert i tabellform ved hjelp av knitr-pakken. Reliabiliteten mellom de to tersklene ble sammenlignet ved å beregne typisk feil som en prosentandel av gjennomsnittsverdien.

En viktig detalj er at to ulike metoder ble brukt for å beregne laktatterskler. For den visuelle representasjonen i plottet, ble gjennomsnittlige laktatverdier for hver watt-verdi beregnet først, etterfulgt av terskelbestemmelse. Dette gir en god visuell representasjon av den generelle trenden og kan være mindre påvirket av ekstreme individuelle verdier.

For den statistiske analysen presentert i tabellene, ble individuelle terskelverdier beregnet først, etterfulgt av gjennomsnittsberegning. Dette tar bedre hensyn til individuelle forskjeller og gir sannsynligvis en mer nøyaktig representasjon av den gjennomsnittlige individuelle terskelen. Denne forskjellen i tilnærming gir litt ulike resultater og belyser viktigheten av å vurdere både gruppetrender og individuelle variasjoner i laktatterskelanalyser.

### 2.2.3 Resultater



Plottet ovenfor illustrerer forholdet mellom treningsintensitet og laktatkonsentrasjon for utvalget i “cyclingstudy”. De grå punktene representerer individuelle målinger, mens den blå kurven viser den generelle trenden i dataene. De røde horisontale stiplede linjene markerer 2 mmol/L og 4 mmol/L laktattersklene. De grønne vertikale prikkede linjene indikerer de estimerte watt-verdiene for disse tersklene, med røde punkter som markerer nøyaktig hvor tersklene krysser den blå kurven. Tilhørende rød tekst viser de spesifikke watt-verdiene for hver terskel.

Som vi ser krysser gruppegjennomsnittet i dette utvalget 2 mmol/L på 221.9 watt og 4 mmol/L på 295.2 watt, med en differanse på 73.3 watt mellom tersklene. Det er viktig å merke seg at disse verdiene er basert på gruppegjennomsnittet og kan avvike fra de individuelle terskelverdiene presentert i tabellene. Den blå kurven viser en eksponentiell trend opp til omtrent 300 watt, noe som er typisk for laktatakkumulering under økende belastning (Tanner & Gore, 2012). Interessant nok flater kurven ut og begynner å synke litt etter ca 350 watt, noe som kan skyldes at vi ser færre målinger ved høyere intensiteter og høyere påvirkning av individuelle forskjeller. Den betydelige spredningen i de individuelle målingene (grå punkter) indikerer stor variasjon mellom deltakerne i deres laktatrespons til økende treningsintensitet. Spredningen av datapunktene ser ut til å være mindre ved lavere intensiteter, noe som kan ha implikasjoner for reliabiliteten av målinger ved ulike laktatnivåer. Dette understreker viktigheten av individuell tilpasning i treningsprogrammer, selv om de gjennomsnittlige terskelverdiene gir en nyttig generell referanse for treningsintensiteter.”

For å kvantifisere disse tersklene og vurdere deres reliabilitet, ble gjennomsnittlige terskelverdier, standardavvik, og typisk feil beregnet for både 2 mmol/L og 4 mmol/L tersklene. Resultatene er presentert i følgende tabeller:

Table 2.1: Laktatterskel Analyse

Terskel	Gjennomsnittlig Watt	Standardavvik	Typisk Feil	Typisk Feil (%)
2 mmol/L	225.4	40.5	28.7	12.7

Table 2.2: Laktatterskel Analyse

Terskel	Gjennomsnittlig Watt	Standardavvik	Typisk Feil	Typisk Feil (%)
4 mmol/L	275.6	30.8	21.8	7.9

Disse tabellene gir en detaljert oversikt over de beregnede verdiene for hver terskel. For 2 mmol/L-terskelen ser vi en gjennomsnittlig watt-verdi på 225,4 W, med et standardavvik på 40,5 W. Den typiske feilen, som er et mål på målevariasjon, er 28,7 W, noe som utgjør 12,7% av gjennomsnittsverdien. For 4 mmol/L-terskelen er den gjennomsnittlige watt-verdien høyere, som forventet, på 275,6 W, med et lavere standardavvik på 30,8 W. Den typiske feilen for denne terskelen er 21,8 W, som utgjør 7,9% av gjennomsnittsverdien.

Denne sammenligningen viser tydelig at 4 mmol/L-terskelen har en lavere typisk feil i prosent (7,9%) sammenlignet med 2 mmol/L-terskelen (12,7%), noe som indikerer en høyere reliabilitet for 4 mmol/L-terskelen.

## 2.2.4 Diskusjon

Resultatene fra denne analysen gir flere interessante innsikter. For det første observerer vi at 4 mmol/L-terskelen oppstår ved en høyere treningsintensitet enn 2 mmol/L-terskelen, noe som er i tråd med vår forståelse av laktatakkumulering under økende fysisk belastning (Tanner & Gore, 2012).

Basert på gruppegjennomsnittet vist i plottet, ser vi at 2 mmol/L-terskelen oppstår ved 221.9 watt, mens 4 mmol/L-terskelen oppstår ved 295.2 watt. Imidlertid, når vi ser på de individuelle beregningene presentert i tabellene, finner vi litt forskjellige verdier. Tabellene viser at basert på individuelle beregninger, er gjennomsnittlig watt-verdi for 2 mmol/L-terskelen 225.4 watt (standardavvik 40.5 watt), mens for 4 mmol/L-terskelen er den 275.6 watt (standardavvik 30.8 watt). Vi ser det er en forskjell men den generelle trenden er lik for begge utregningene.

Et bemerkelsesverdig funn er forskjellen i reliabilitet mellom de to tersklene. Til tross for at spredningen av datapunktene ser ut til å være mindre ved lavere intensiteter i plottet, viser 4

mmol/L-terskelen seg å være betydelig mer reliabel i dette datasettet, med en typisk feil på 7,9% sammenlignet med 12,7% for 2 mmol/L-terskelen. Denne tilsynelatende motsetningen kan skyldes at forholdet mellom treningsintensitet og laktatproduksjon blir mer lineært og forutsigbart ved høyere intensiteter, selv om den absolutte spredningen øker. Individuelle forskjeller i laktatrespons kan også spille en større rolle nær 4 mmol/L-terskelen, da det er færre observasjoner ved høyere intensitet.

## **2.2.5 Konklusjon**

Denne analysen har demonstrert at både 2 mmol/L og 4 mmol/L laktattærskler kan beregnes og anvendes for å vurdere treningsintensitet og aerob kapasitet. Resultatene viser at 4 mmol/L-terskelen har en høyere reliabilitet i dette utvalget, noe som gjør den til et potensielt mer robust verktøy for treningsovervåkning og prestasjonsvurdering. Likevel bør valget av terskel tilpasses individuelle behov og spesifikke treningsmål. Denne type analyse bidrar til en mer presis og individualisert tilnærming til treningsplanlegging og prestasjonsforbedring innen utholdenhetsidretter, og viser viktigheten av å vurdere ulike fysiologiske markører i treningssammenheng.

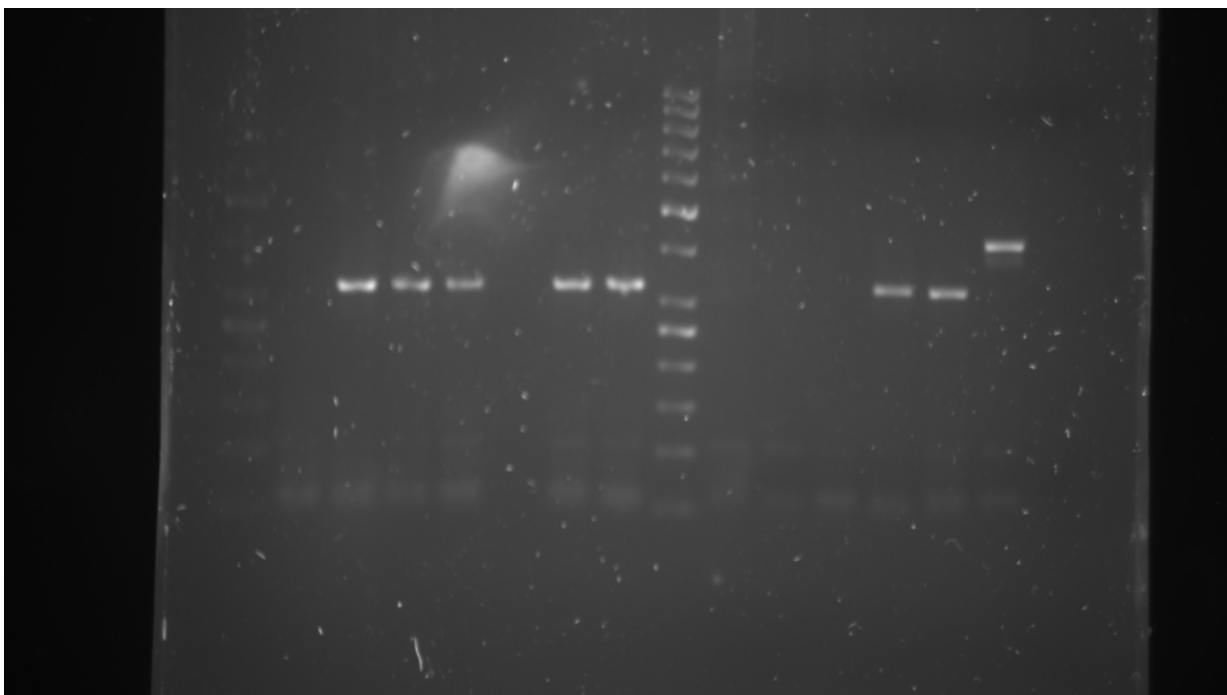
## **2.3 Del 2: Forutsi størrelser på DNA-fragmenter.**

### **2.3.1 Innledning**

I denne oppgaven ble jeg bedt om å analysere størrelsen på DNA-fragmenter som ble amplifisert ved hjelp av PCR (eller RNA med qPCR, jeg valgte å analysere DNA). For å gjøre dette fulgte jeg en protokoll vi fikk tildelt Trainome (n.d.) som inkluderte bruk av agarosegelelektroforese for å visualisere DNA-fragmentene og ImageJ for å måle migrasjonsavstandene til fragmentene. Deretter brukte jeg R for å lage en kalibreringsmodell basert på en DNA-stige og beregne størrelsene på de ukjente DNA-fragmentene fra prøvene mine.

### **2.3.2 Metode**

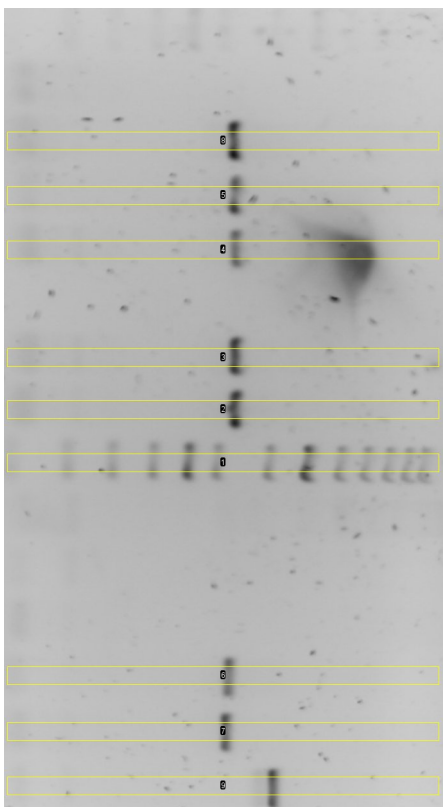
Jeg brukte følgende bilde som viser resultatene for gruppe 3 og 4 fra oppgaven “DNA ekstraksjon og analyse”. Vi skulle se hvilken genotype vi hadde av ACTN3. Som er det genet vi ofte forbinder med raske og trege muskelfibere, eller “Sprint” og “utholdenhet”. Ifølge protokollen vi fulgte i fra forsøket på labben skulle “A R/R genotype produces two bands at 690 and 413 bp. A X/X genotype produces two bands at 690 and 318 bp.”. Vi kan allerede se utifra bildet at alle prøvene som viser noe, bare viser 1 bånd hver og at sannsynligheten er stor for at noe har gått galt i forsøket.



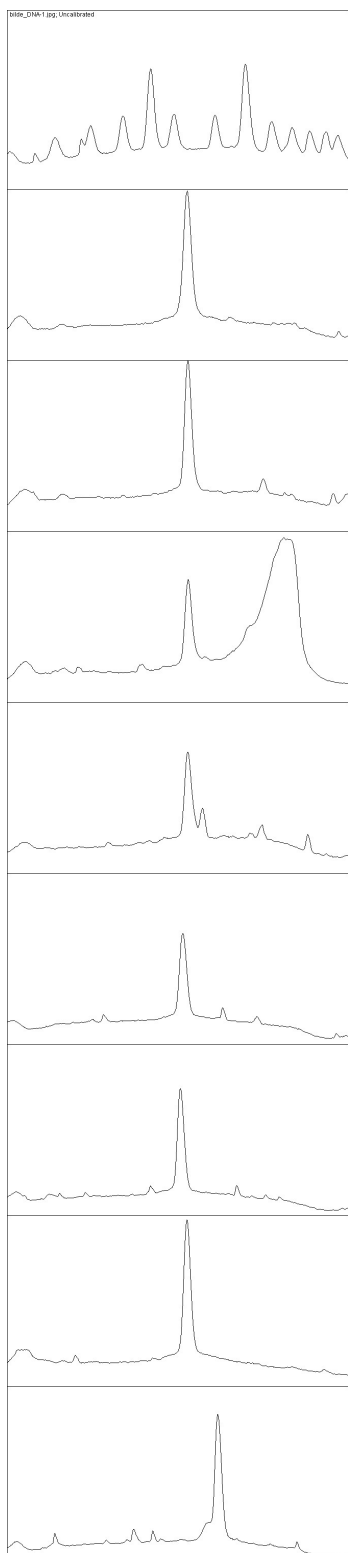
Jeg burde fortsatt kunne få ut hvilket basepar de bandene som er der ligger på, og de 7 første prøvene burde gi likt svar og det siste burde gi et annet svar. Jeg burde da kunne finne ut hvilket band som er vist på bildet. Ligger de første 7 i nærheten av 690 må jeg se om det andre ligger nært 413 eller 318. Ligger de første på 413 eller 318 vet jeg at det som er for seg selv sannsynligvis er det motsatte.

Jeg startet med å åpne bildet av gelen i fiji med ImageJ, hvor jeg inverterte bildet og roterte det slik at brønnene var plassert til venstre. Deretter markerte jeg stigen i midten (brønn 1) ved hjelp av rektangelverktøyet og registrerte intensitetstopper for hvert bånd i stigen. Jeg gjorde det samme for prøvene, der fem prøver foran stigen (brønn 2-5) og tre bak stigen (brønn 6-8) hadde klare bånd. Brønnene er nummerert etter de som gir synlige resultater, ikke kronologisk da flere prøver ikke ga resultat. På grunn utfordringer i håndtering av bildet endte nr rekkefølgen å bli fra toppen til bunnen: 8-5-4-3-2-1(stige)-6-7-9. Dette burde ikke spille noen rolle da prøver fra brønn 2-8 skal være like og bare en prøven som er ifra brønn 9 som er ulik resten. Dette er bildet som viser indelingen.





Resultatet ga dette bildet, hvor hver forhøyning representere piksel intensiteten i bildet fra markeringene. Deretter fikk jeg ut følgende bilde.



Jeg markerte hver forhøyning i bildet med unntak av det jeg tror er støy, Inkludert den store forhøyningen på bilde 4 fra toppen som ser ut som en refleksjon fra da bildet ble tatt. Deretter fortsettet jeg i R.

### 2.3.3 Resultat og analyse i R

Etter at jeg hadde samlet inn data fra ImageJ, eksporterte jeg disse til et regneark og importerte dem til RStudio for videre analyse. Deretter definerte jeg DNA-stigen basert på guiden vi fikk utdelt fra protokollen som var:

Jeg laget deretter en kalibreringsmodell basert på log-transformerte molekylvekter og migrasjonsavstander fra stigen.

Call:

```
lm(formula = log(mw) ~ dist, data = ladder)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.054731	-0.040851	-0.001577	0.034241	0.063743

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.9957281	0.0200403	349.08	< 2e-16 ***
dist	-0.0052316	0.0000712	-73.47	3.69e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0432 on 11 degrees of freedom

Multiple R-squared: 0.998, Adjusted R-squared: 0.9978

F-statistic: 5398 on 1 and 11 DF, p-value: 3.691e-16

Så estimerte jeg størrelsene på DNA-fragmentene fra prøvene. I protokollen står det “Create a spreadsheet with the data collected in ImageJ, alternatively, input your data in R directly (as below). The x-values from ImageJ are the values of interest.”

Jeg har da endret koden fra protokollen

```
unknown <- data.frame(dist = c(470.5, 391.5, 528.5,  
                               421.5, 507.5))
```

til

```
# Definer ukjente prøver
unknown <- data.frame(dist = data$X)

# Estimer molekylvekter
preds <- exp(predict(cal, newdata = unknown))

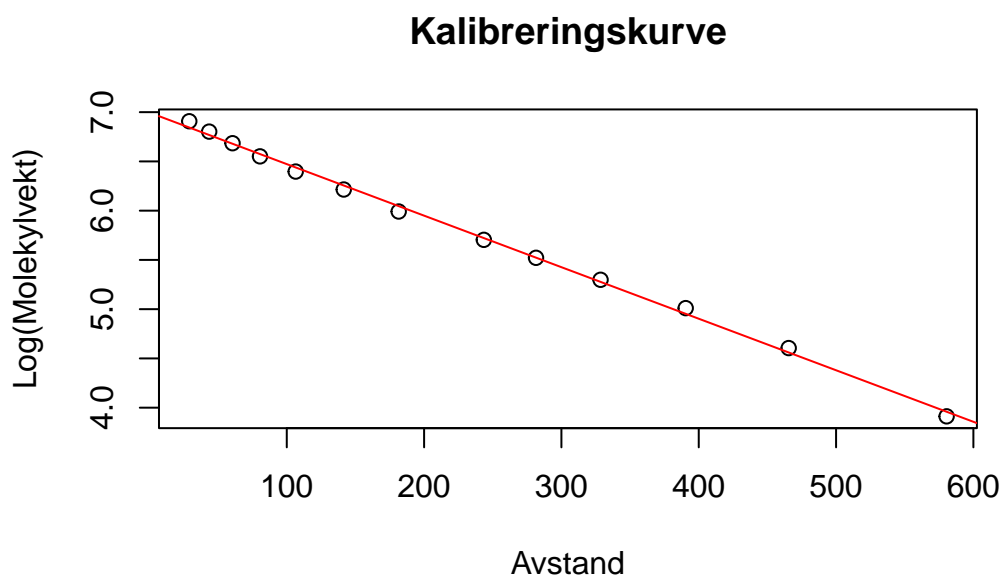
# Lag resultat-dataframe
resultater <- data.frame(Prove = data$X.1, Avstand = data$X, Estimert_Storrelse = round(preds,
```

Deretter laget jeg en tabell som skal gi ut informasjon om prøvene.

Table 2.3: Estimerte DNA-fragmentstørrelser basert på migrasjonsavstander

Prove	Avstand	Estimert størrelse (bp)
1	6.0	1058
2	55.5	817
3	94.5	666
4	164.5	462
5	228.5	330
6	281.5	250
7	326.5	198
8	407.5	130
9	466.5	95
10	519.5	72
11	559.5	58
12	593.5	49
13	626.5	41
14	648.5	37
15	352.5	173
16	353.5	172
17	354.5	171
18	352.5	173
19	343.5	181
20	340.5	184
21	353.5	172
22	411.5	127

Jeg laget også et plott som visualisere hva stigen gjør:



Dette plottet viser kalibreringskurven basert på DNA-stigen vi brukte i eksperimentet. X-aksen representerer migrasjonsavstanden målt i ImageJ, mens Y-aksen viser den naturlige logaritmen av molekylvekten (i basepar). Punktene representerer dataene fra DNA-stigen, mens den røde linjen er den lineære regresjonsmodellen som best passer disse dataene.

Kurven illustrerer det inverse forholdet mellom migrasjonsavstand og molekylvekt: større DNA-fragmenter (høyere molekylvekt) migrerer kortere avstander i gelen. Den lineære trenden i dette log-transformerte plottet bekrefter at vi kan bruke denne modellen til å estimere størrelsen på ukjente DNA-fragmenter basert på deres migrasjonsavstander. Jo tettere punktene ligger til den røde linjen, desto mer pålitelig er vår kalibreringsmodell.

### 2.3.4 Diskusjon

Jeg er faktisk ikke sikker på resultatene. Jeg har forsøkt å finne svar fra ulike kilder, men kan ikke finne en god måte å lese resultatene på. Såvidt jeg har forstått har jeg 8 målinger av prøver + 1 stige. Mens “Estimerte DNA fragmentstørrelsen basert på migrasjonsavstander” gir resultat fra 22 prøver. I tillegg vil jeg forvente at 7 av prøvene som kommer etter hverandre burde ha like, eller veldig nære basepar. Mot slutten av tabellen ser jeg noen likheter jeg kanskje kan tenke meg til er det jeg ser etter. Jeg kan tenke meg flere grunner til at tabellen ikke vises riktig.

Den ene grunnen er at koden jeg brukte til stigen ikke er den samme som den stigen vi brukte i forsøket. Dette kan da ha gjort at resultatet ikke er representativt av hva jeg faktisk forsøker å fremlegge.

Når jeg analyserte dataene, brukte jeg X-verdiene fra datasettet som migrasjonsavstander, basert på protokollens anbefaling om at x-verdiene er av interesse. Dette resulterte i 22 datapunkter, som er mer enn de 8 prøvene pluss 1 stige jeg opprinnelig observerte på gelen. De estimerte DNA-fragmentstørrelsene viser en gradvis nedgang fra 1058 bp til 37 bp for de første 14 prøvene, etterfulgt av en gruppe verdier mellom 127 bp og 184 bp for de siste 8 prøvene. Denne fordelingen stemmer ikke overens med forventningen om å se to distinkte bånd per prøve ved enten 690 og 413 bp eller 690 og 318 bp. Det er mulig at X-verdiene i datasettet representerer noe annet enn de faktiske migrasjonsavstandene, eller at kalibreringen av stigen ikke er optimal for våre spesifikke prøver. Jeg ser imidlertid en uniformitet i de 8 siste prøvene. Den siste har 127 bp og de 7 foregående har mellom 171 og 184 bp.

Jeg har ikke nok kunnskap om denne koden og koding generelt, det er en mulighet til at det er noen feil som har kommet med fra min side. Det var ikke spesifikt skrevet i protokollen hvordan data settet skulle leses, annet enn at det er x-aksen som er av interesse så det er høy sannsynlig at følgende kode ikke er den riktige for å lese datasettet:

Jeg tror dette kanskje er det med høyest sannsynlighet er grunnen til at jeg ikke får riktig data ut.

Jeg jobbet ut ifra et ufullstendig bilde med bare ett bånd per prøve da jeg skulle hatt to bånd per prøve. Dette understreker at bildet jeg jobbet ut ifra sannsynligvis har hatt feil ved seg i forsøket, men jeg antar jeg skulle fortsatt ha klart å se like bånd.

### 2.3.5 Resultater

Jeg har forsøkt å analysere DNA-fragmenter fra ACTN3-genet for å bestemme genotypen relatert til muskelfiber-type. Basert på protokollen fra biologilabben skulle vi forvente å se to bånd, enten ved 690 og 413 bp for R/R-genotypen, eller ved 690 og 318 bp for X/X-genotypen. Jeg hadde bare et bånd å jobbe ut ifra og mine resultater avviker betydelig fra disse forventningene, noe som tyder på at det har oppstått flere utfordringer i løpet av eksperimentet og analysen.

De estimerte DNA-fragmentstørrelsene vi observerte varierte fra 1058 bp til 37 bp, med en gradvis nedgang for de første 14 prøvene og en gruppe av verdier mellom 127 bp og 184 bp for de siste 8 prøvene. Ingen av disse verdiene samsvarer nøyaktig med de forventede båndstørrelsene på 690, 413, eller 318 bp for ACTN3-genotypene. Selv om resultatene viser en logisk trend med avtagende fragmentstørrelser ved økende migrasjonsavstander, er verdiene spredt over et større område enn forventet og inkluderer ikke de spesifikke størrelsene vi lette etter. Dette tyder på at det kan være problemer med enten kalibreringen, datainnsamlingen, eller selve PCR-amplifiseringen.

Det er også verdt å merke seg at jeg fikk resultater for 22 prøver, selv om jeg kun analyserte 8 brønner (9 inkludert stigen). Dette tyder på at det kan ha oppstått feil i dataregistreringen eller analysen.

Det er imidlertid et sted jeg ser en mulig forklaring i resultatet. Den siste prøven har 127 bp og de 7 foregående har mellom 171 og 184 bp som er relativt likt. Det er ikke umulig at de første 14 “prøvene” er stigen, de neste 7 prøvene er de første 7 prøvene som skal i prinsippet være like (men kan ha blitt ulike på grunn av unøyaktighet fra min side) og det siste prøven er den siste prøven som var noe anderledes. Dette er bare ren spekulasjon fra min side, men det virker for meg til å være den mest logiske konklusjonen. Hvis jeg går ut med denne enorme antagelsen i bakgrunn, ville jeg sagt at den siste prøven er X/X siden den har et bånd som skal være på 318 bp som er det laveste, og den har det laveste av de siste 8 prøvene. Jeg vil ikke konkludere med dette. for alt jeg vet har koden snudd skalaen, og jeg gjør mange antagelser.

Gitt disse utfordringene og de avvikende resultatene, kan jeg ikke trekke pålitelige konklusjoner om ACTN3-genotypene i våre prøver. Det jeg kan si er at det kan hende, ut ifra visuell utforsking av det originale bildet er at det er 7 personer som kan ha samme genotype og en som har en annen. Dette kan også være feil da vi bare ser et bånd.

Dette forsøket understreker viktigheten av nøyaktighet og presisjon i alle trinn av molekylær-biologiske eksperimenter, fra prøvepreparering til dataanalyse. Selv om jeg ikke kunne bestemme ACTN3-genotypene som planlagt, har denne erfaringen gitt verdifull innsikt i flere potensielle fallgruver og områder for forbedring i lignende fremtidige studier. Hadde jeg hatt tid ville jeg gjennomgått koden med en faglærer og forsøkt på nytt.

## 2.4 Del 3: Tolkning av regresjonstabell

### 2.4.1 Introduksjon

I denne delen av oppgaven skulle vi finne et forskningsspørsmål basert på “hypertrophy” datasettet i “exscidata” pakken i Rstudio og lage en lineær regresjonsanalyse som skal tolkes.

### 2.4.2 Forskningsspørsmål

basert på data fra “hypertrophy” har jeg formulert følgende spørsmål.

Er det en lineær sammenheng mellom muskeltykkelse (VL\_T1) og styrke i knebøy (SQUAT\_3RM)?”

Muskeltykkelsen målt er da vastus lateralis (VL\_T1) og styrken er den maksimale vekten løftet i knebøy over 3 repetisjoner(SQUAT\_3RM).

### 2.4.3 Metode

For å svare på dette spørsmålet har vi utført en lineær regresjonsanalyse. Resultatene fra analysen er presentert i tabellen nedenfor:

Table 2.4: Regresjonsanalyse: Muskeltykkelse og Knebøystyrke

Variabel	Estimat	Standardfeil	t.verdi	p.verdi
Skjæringspunkt	78.427	19.487	4.025	0.000
Muskeltykkelse (VL)	17.115	6.300	2.717	0.011

### 2.4.4 Tolkning av resultatene (med fokus på standardfeil, t-verdi og p-verdi)

Estimatet representerer den beregnede verdien basert på vårt utvalg, som er en del av den større populasjonen vi ønsker å undersøke. Målet er at dette estimatet skal være en god tilnærming til den sanne verdien i hele populasjonen. Jo større og mer representativt utvalget er, desto mer nøyaktig blir estimatet, forutsatt at utvalgsmetoden er god (Spiegelhalter, 2020). Estimatet for muskeltykkelse er 17,115, som betyr at for hver enhets økning i muskeltykkelse, forventer vi en økning på omtrent 17,115 enheter i knebøystyrke, gitt at alle andre faktorer holdes konstante.

Standardfeilen for muskeltykkelse (VL) er 6,300. Dette tallet representerer usikkerheten i estimatet vårt. Standardfeil er et statistisk mål som forteller oss hvor presist vårt estimat av populasjonsgjennomsnittet er. En lavere standardfeil indikerer et mer presist estimat. I dette tilfellet forteller standardfeilen oss hvor mye estimatet vårt kan variere hvis vi gjentok studien med nye utvalg. Hvis antall observasjoner øker vil standardfeilen bli mindre (Aalen *et al.*, 2018).

t-verdien for muskeltykkelse (VL) er 2,717. Denne verdien beregnes ved å dele estimatet (koeffisienten) på standardfeilen. t-verdien hjelper oss å vurdere om effekten av muskeltykkelse på knebøystyrke er statistisk signifikant. Jo større absoluttverdien av t-verdien er, desto sterkere bevis har vi mot nullhypotesen om at det ikke er noen sammenheng mellom variablene muskeltykkelse og knebøystyrke (Spiegelhalter, 2020).

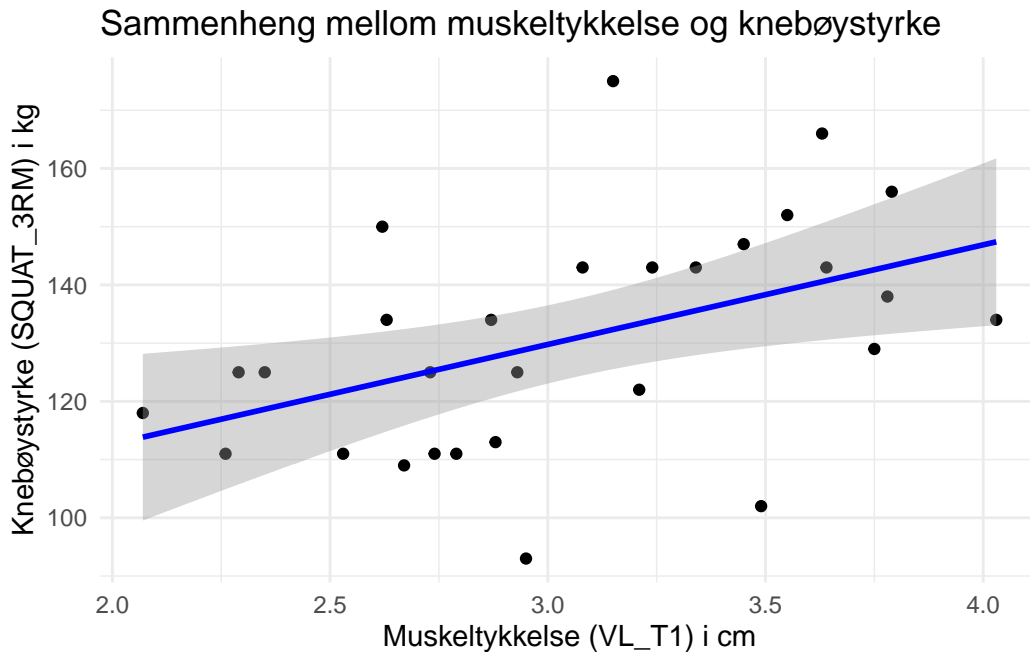
p-verdien er et sentralt konsept i statistisk hypotesetesting. p-verdien for muskeltykkelse (VL) er 0,011. Dette tallet representerer sannsynligheten for å observere en t-verdi like ekstrem som, eller mer ekstrem enn, den vi faktisk observerte, gitt at nullhypotesen er sann. En p-verdi mindre enn det vanlige signifikansnivået på 0,05 indikerer at vi har statistisk signifikante bevis mot nullhypotesen (Hulley *et al.*, 2013).

I vårt tilfelle er p-verdien mindre enn 0,05 som er et vanlig signifikansnivå. Det er viktig å merke seg at p-verdien ikke direkte forteller oss sannsynligheten for at nullhypotesen er sann eller falsk. I stedet gir den oss informasjon om hvor godt dataene våre stemmer overens med



nullhypotesen. En lav p-verdi, som i vårt tilfelle, antyder at dataene våre er lite sannsynlige under nullhypotesen, noe som fører til at vi forkaster nullhypotesen til fordel for den alternative hypotesen om en sammenheng mellom variablene (muskeltykkelse og styrke) (Hulley *et al.*, 2013).

For å gi et visuelt bilde kan vi legge det inn i et spredningsplott:



Spredningsplottet ovenfor visualiserer sammenhengen mellom muskeltykkelse (VL\_T1) og knebøystyrke (SQUAT\_3RM) som vi har analysert i regresjonsmodellen.

Fra plottet kan vi observere den positive lineære trenden som regresjonsanalysen indikerte. Hver prikk representerer en deltaker i studien, og den blå linjen viser den estimerte lineære sammenhengen. Det skyggelagte området rundt linjen representerer 95% konfidensintervallet, som gir oss en visuell indikasjon på usikkerheten i estimatet.

Dette plottet støtter vår tolkning av regresjonskoeffisienten. Vi kan se at når muskeltykkelsen øker, er det en generell tendens til at knebøystyrken også øker. Samtidig illustrerer spredningen av punktene rundt linjen at det er individuell variasjon, noe som understreker viktigheten av å vurdere både estimatet og standardfeilen i vår analyse.

## 2.4.5 Konklusjon

Basert på analysen kan vi konkludere med at det finnes en signifikant lineær sammenheng mellom muskeltykkelse og styrke i knebøy i dette datasettet. Resultatene tyder på at økt

muskeltykkelse er assosiert med høyere styrke i knebøy, noe som kan ha implikasjoner for treningsprogrammer rettet mot styrkeutvikling.

## 3 Oppgave 3: Reliabilitet og verktøy for reproduserbar vitenskapelig data

### 3.1 Introduksjon

Dette er oppgave Oppgave 3: “Reliabilitet og verktøy for reproduserbar vitenskapelig data” i faget IDR4000-1 24H Kvantitativ metode og statistikk for master i treningsfysiologi. Dette var en individuell oppgave jeg har revidert på tilbakemelding fra fagansvarlig.

Vi har fått utdelt noen eksempelkode som simulere statistikk og utregninger i to utvalg av en populasjon. En med utvalgstørrelse på 8, og en med utvalgstørrelse 40. Først blir utregning fra eksempelkodene presentert også svarer jeg på spørsmålene i oppgaven.

#### 3.1.1 Utregning av $m1 = n8$

Call:

```
lm(formula = y ~ 1, data = samp1)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.5322	-1.2523	-0.0883	1.3540	4.8692

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.840	1.251	1.47	0.185

Residual standard error: 3.539 on 7 degrees of freedom

#### 3.1.2 Utregning av $m2 = n=40$

Call:

```
lm(formula = y ~ 1, data = samp2)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.6557	-2.2883	0.2636	2.2549	6.4212

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.5642	0.4774	3.276	0.00221 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.019 on 39 degrees of freedom

## 3.2 Spørsmål 1

Forklar estimatet, standardfeil (standard error = SE), t-verdi og p-verdi fra regresjonsmodellene som blir opprettet her (m1 og m2).

### 3.2.1 Estimater

Er gjennomsnittsverdien i utvalget. Ved å ta et utvalg(en del) av populasjonen(Den gruppen vi ønsker å si noe om), ønsker vi å si at gjennomsnittet i utvalget er representativ for populasjonen. Det vi ønsker er at estimatet skal representere populasjonen så likt som mulig. Estimater blir mer presist (nærmere den sanne verdien i populasjon) når utvalgsstørrelsen øker, gitt at kvaliteten på utvalget er bra. (Spiegelhalter, 2020).

Estimatet er vår beste gjetning på populasjonens sanne gjennomsnitt. I eksemplet er det brukt 8 og 40 som utvalg (utvalget er tilfeldig valgt fra en populasjon på 1000000 i dette tilfellet). For m1 (som har 8 simulerte deltagere) er estimatet 1.84 mens for m2 (som har 40 simulerte deltagere) er det 1.5642 som er nærmere det faktiske populasjons verdien på 1.5.

### 3.2.2 Standardfeilen (SE)

Standardfeil er et statistisk mål som forteller oss hvor presist vårt estimat av populasjonsgjennomsnittet er. Det måler hvor mye vi kan forvente at våre estimer varierer fra utvalg til utvalg, og en mindre standardfeil indikerer at våre estimer er mer pålitelige og sannsynligvis nærmere den sanne verdien i populasjonen. Standardfeilen beregnes ved å dele standardavviket på kvadratroten av utvalgsstørrelsen  $SE = \frac{SD}{\sqrt{n}}$ . Standardfeilen går mot 0 når n vokser, men kan aldri bli nøyaktig 0. Det vil si at estimatet blir stadig mer presist når antall observasjoner

vokser. Usikkerheten avtar bare ved kvadratroten av  $n$ . Det betyr at for å halvere usikkerheten kreves det en firedobling av antall observasjoner (Aalen *et al.*, 2018; Spiegelhalter, 2020).

I vårt eksempel, med utvalg på 8 og 40 fra en populasjon på 1000000, vil utvalget på 40 gi en mindre standardfeil enn utvalget på 8, noe som betyr at estimatet basert på 40 observasjoner sannsynligvis er nærmere det sanne populasjonsgjennomsnittet. Når utvalgsstørrelsen øker, blir standardfeilen mindre, noe som indikerer økt presisjon i våre estimater. Ved bruk av standardfeil kan det hjelpe oss å vurdere hvor sikre vi kan være på våre statistiske konklusjoner og hvor godt våre data representerer den større populasjonen vi studerer (Aalen *et al.*, 2018).

For  $m_1$  er SE 1.251 og for  $m_2$  er SE 0.4774 som er betraktelig lavere enn  $m_1$  siden populasjonutvalget er høyere.

### 3.2.3 t-verdi

Er forholdet mellom estimatet og Standardfeil med formelen  $t = \frac{\text{estimat}}{\text{SE}}$ . t-verdien indikerer hvor mange standardfeil estimatet er fra null. En høyere absoluttverdi av  $t$  indikerer sterkere bevis mot nullhypotesen, at det ikke er en forskjell. En annen måte å si det på er “hvor langt unna er estimatet fra 0, målt i i antall standard feil” (Spiegelhalter, 2020).

Spiegelhalter (2020) sier også at ved store nok utvalg vil en  $t$ -verdi større enn 2 eller mindre enn -2 korrespondere med en  $P < 0.05$ .

Vi ser at  $m_2$  har betraktelig høyere  $t$ -verdi som tyder på en mer statistisk signifikant resultat. For  $m_1$ :  $t = \frac{1.84}{1.251} \approx 1.47$  og for  $m_2$ :  $t = \frac{1.5642}{0.4774} \approx 3.276$

### 3.2.4 p-verdien

Er et mål på avvik mellom data og nullhypotesen. Vi bruker en statistisk test for å finne  $p$ -verdien som er sannsynligheten for å observere en like ekstrem eller mer ekstrem effekt utelukket av tilfeldighet ( $t$ -verdi) (Hulley *et al.*, 2013; Spiegelhalter, 2020).

I følge Hulley *et al.* (2013) så hvis nullhypotesen er sann, og det virkelig ikke er en forskjell i populasjonen. Så er den eneste måten denne studien vi bruker testen på, kunne finne denne effekten i utvalget på, er ved en tilfeldighet. Hvis denne effekten er liten så kan nullhypotesen forkastes til fordel for dens alternativ  $\alpha$ , som er dens forutsagte nivå av statistisk signifikans, som oftest er på 0.05.

Et “ikke signifikant” ( $p$ -verdi over 0.05) resultat betyr ikke det ikke er noen assosiasjon i befolkningen; det betyr bare at resultatet observert i utvalget er lite sammenlignet med hva som kan bli observert av en tilfeldighet alene (Hulley *et al.*, 2013).

For  $m_1$  er  $p$ -verdien 0.185, som er over det vanlige signifikansnivået på 0.05. Dette betyr at vi ikke har tilstrekkelig evidens til å forkaste nullhypotesen basert på dette utvalget. For  $m_2$  er  $p$ -verdien 0.00221, som er betydelig lavere og under 0.05. Dette indikerer sterk evidens mot

(ikke sann) nullhypotesen. Det betyr at det er svært lite sannsynlig at vi ville observere et slikt resultat (eller et mer ekstremt resultat) hvis nullhypotesen var sann. Dette gir oss grunnlag for å forkaste nullhypotesen for  $m_2$ .

### 3.3 Spørsmål 2

Diskuter hva som bidrar til de forskjellige resultatene i de to studiene ( $m_1$  og  $m_2$ ).

$m_2$  har et større utvalgsstørrelse ( $n=40$ ) sammenlignet med  $m_1$  ( $n=8$ ).

Disse resultatene demonstrerer tydelig hvordan en større utvalgsstørrelse fører til mer presise estimater, lavere standardfeil, høyere  $t$ -verdier, og lavere  $p$ -verdier som beskrevet under spørsmål 1.

Dette øker vår evne til å oppdage reelle effekter (økt statistisk styrke) og reduserer sannsynligheten for type II-feil (falske negative). Sannsynligheten for type I-feil (falske positive) forblir konstant på det valgte signifikansnivået (oftest 0,05 eller 5%). Type I-feil skjer når vi ved utregning forkaster en nullhypotese som faktisk er sann i populasjonen. Type II-feil er det som skjer når vi ved utregning ikke forkaster en nullhypotese som faktisk ikke er sann i populasjonen (Hulley *et al.* (2013); Spiegelhalter (2020)).

Med andre ord, når vi øker antall deltakere i studien, får vi mer nøyaktige og troverdige funn. Dette gjør det lettere å oppdage virkelige effekter hvis de eksisterer, men endrer ikke sjansen for å feilaktig konkludere at det finnes en effekt når det egentlig ikke gjør det.

Mindre utvalg som  $m_1$  er mer utsatt for tilfeldige variasjoner og større utvalg som  $m_2$  er mer stabile representasjon av populasjonen.

### 3.4 Spørsmål 3

Hvorfor bruker vi det skraverte området i den nedre og øvre halen av  $t$ -fordelingen?

Det skraverte området i den nedre og øvre halen av  $t$ -fordelingen er nødvendig for å tolke resultatene av en hypotesetest. Dette området representerer de regionene som brukes for å vurdere statistisk signifikans. I praksis beregner vi først en  $t$ -verdi basert på våre observerte data. Denne  $t$ -verdien sammenlignes deretter med  $t$ -fordelingen under nullhypotesen.  $P$ -verdien beregnes som sannsynligheten for å observere en  $t$ -verdi like ekstrem eller mer ekstrem enn den vi faktisk observerte, gitt at nullhypotesen er sann (Spiegelhalter (2020)).

Det skraverte området i figuren illustrerer denne sannsynligheten visuelt. Hvis vår beregnede  $p$ -verdi faller innenfor dette skraverte området, indikerer det at våre observerte data er uvanlige nok til å sette spørsmålstegn ved nullhypotesen. Med andre ord, jo mindre  $p$ -verdien er (jo mer av det skraverte området vår observerte  $t$ -verdi faller innenfor), desto sterkere bevis har vi mot

nullhypotesen. I en to-halet test, som figuren illustrerer, skraverer vi både den øvre og nedre halen. Dette gjør vi fordi vi tester om parameteren er signifikant forskjellig fra nullhypotesen i begge retninger. Summen av arealene i begge haler gir oss den totale p-verdien (Hulley *et al.*, 2013; Spiegelhalter).

Som nevnt noen ganger allerede vil størrelsen på utvalget påvirke estimatet. Med tanke på figuren vil formen på t-fordelingen og dermed også det skraverte området forandre seg. Større utvalg gir en smalere fordeling, noe som kan føre til lavere p-verdier og potensielt sterkere bevis mot nullhypotesen.

#### *Mange studier*

Det er simulert 1000 studier og lagret. Dette er en kopi fra oppgaveteksten som trengs for å besvare de neste spørsmålene.

### **3.5 Spørsmål 4**

Beregn standardavviket (SD) for variabelen estimate, og gjennomsnittet av variabelen standardfeil (SE) for hver av studiens utvalgsstørrelser (8 og 40). Forklar hvorfor disse tallene er veldig like. Hvordan kan du definere standardfeilen (SE) i lys av disse beregningene?

Utvalgsstørrelse 8

Standardavvik (SD) for estimatene: 1.0708432

Gjennomsnittlig Standardfeil (SE): 1.0213745

Utvalgsstørrelse 40

Standardavvik (SD) for estimatene: 0.4838475

Gjennomsnittlig Standardfeil (SE): 0.4696954

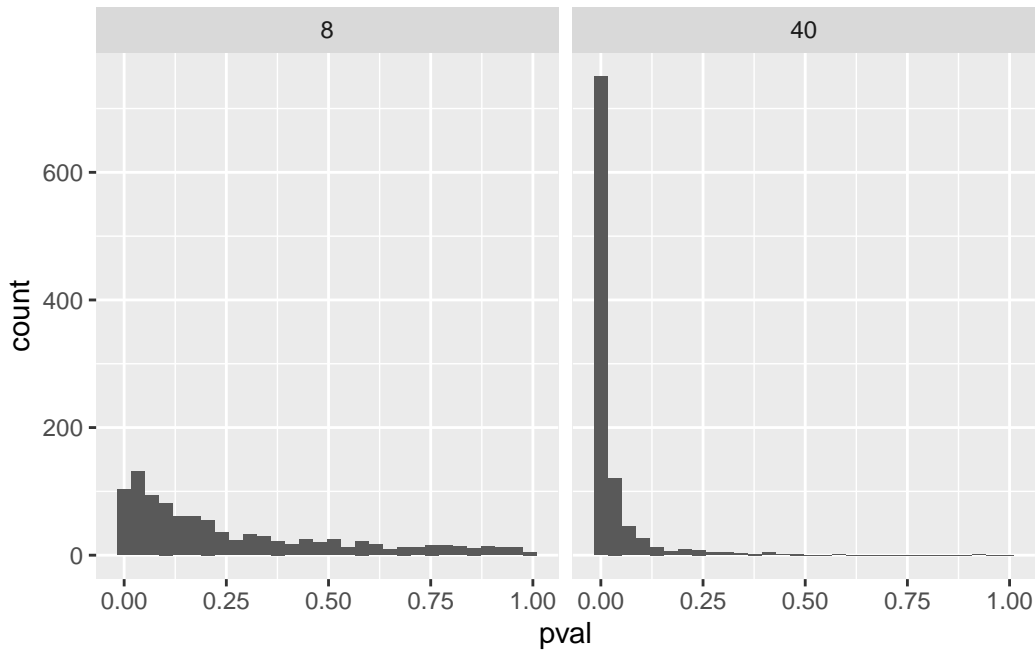
Standardavviket for estimatene viser hvor mye estimatene variere fra gjennomsnittet, mens gjennomsnittet av Standardfeil gir oss en indikasjon på hvor mye usikkerhet det er i disse estimatene. Når tallene er like, kan det tyde på at variasjonen i estimatene er konsistent med usikkerheten som er estimert av Standardfeilen, noe som burde være tilfellet i en god statistisk modell (Spiegelhalter, 2020).

Standardfeilen kan defineres som et mål på hvor mye vi forventer at estimatet vil variere fra det sanne populasjongjennomsnittet. Det gir oss en indikasjon på presisjonen til estimatet; jo lavere SE, desto mer presist er estimatet (Spiegelhalter, 2020).

I tillegg er det viktig å merke seg at den gjennomsnittlige standardfeilen er et estimat på spredningen i utvalgsfordelingen. Utvalgsfordelingen representerer alle mulige beregnede gjennomsnitt fra gjentatte utvalg fra populasjonen (Aalen *et al.*, 2018).

### 3.6 Spørsmål 5

Opprett et histogram (Brukt eksempelkode) av p-verdiene fra hver utvalgsstørrelse. Hvordan tolker du disse histogrammene, hva forteller de deg om effekten av utvalgsstørrelse på statistisk styrke?



Histogrammet  $n=8$  viser en spredt fordeling av p-verdiene. Selv om flere av p-verdiene er nært 0.00, har den også mange som er mye høyere. Som tyder på lavere statistisk styrke. De lave p-verdiene kan være tilfeldige og er ikke nødvendigvis pålitlige (Spiegelhalter, 2020).

Histogrammet  $n=40$  viser en veldig konsentrert fordeling av p-verdiene nært 0.00. Dette indikerer at vi har flere signifikante resultater. Dette er fordi større utvalg redusere usikkerheten i estimatene. Dette gir et klarere bilde av hva som faktisk skjer i populasjonen (Spiegelhalter, 2020).

### 3.7 spørsmål 6

Beregn antall studier fra hver utvalgsstørrelse som erklærer en statistisk signifikant effekt (spesifiser en terskel for  $\alpha$ , ditt signifikansnivå).

Standarden for signifikansnivået i statistiske tester 0.05. og er det vi fikk av eksempelkode.



Table 3.1: Andel signifikante resultater

n	sig_results
8	0.227
40	0.865

Dette betyr at antall studier som erklærer en statistisk signifikant effekt er ( $p < 0.05$ ) for  $n=8$  er 22.7% og for  $n=40$  er det 86.5%. Dette viser tydelig at med større utvalg er det mye mer sannsynlig å oppdage signifikante resultater, noe som bekrefter at større utvalg forbedrer den statistiske styrken

Jeg prøver å senke terskelen for  $\alpha$ , til 0.01. Siden oppgaven spør etter å spesifisere en terskel for  $\alpha$ , mitt signifikansnivå.

Table 3.2: Andel signifikante resultater med terskel 0.01

n	sig_results
8	0.074
40	0.666

Nå ble antall studier som erklærer en statistisk signifikant effekt ( $p < 0.01$ ) for  $n=8$  til 7.4% og for  $n=40$  er det 66.6%. Dette viser tydelig at med strengere krav til hva som definerer signifikant blir også andelen signifikante resultater mindre. Påstanden fra tidligere om at større utvalg forbedrer den statistiske styrken står fortsatt. Hvorfor en velger strengere signifikant kan være blant annet for å redusere falske posetiver for å være enda sikrere i konklusjonen (Aalen *et al.*, 2018).

Jeg gjør det igjen med å øke terskelen for  $\alpha$  til 0.10

Table 3.3: Andel signifikante resultater med terskel 0.10

n	sig_results
8	0.371
40	0.926

Ikke uventet økte antall studier som erklærer en statistisk signifikant effekt til 37.1% for  $n=8$  og 92.6% for  $n=40$ . Det kan være hensiktsmessig å øke signifikansnivået hvis det for eksempel bare skal se om det er en effekt og falske posetiver ikke er farlig for det vi skal se på (Aalen *et al.*, 2018).

Når man endrer signifikansnivået ( $\alpha$ ) i en statistisk test, oppstår det en avveining mellom type I og type II feil. Senking av  $\alpha$  reduserer sannsynligheten for type I-feil (falske positive), men øker samtidig risikoen for type II-feil (falske negative). Motsatt vil en økning av  $\alpha$  føre til større sannsynlighet for type I-feil, men redusere risikoen for type II-feil. Denne avveiningen er uunngåelig i statistisk testing, og valg av signifikansnivå bør reflektere studiens kontekst, inkludert de relative kostnadene ved å gjøre hver type feil. Tradisjonelt foretrekkes det å minimere type I-feil, noe som forklarer de vanlige signifikansnivåene på 1% eller 5%. I spesielle tilfeller, som ved testing av potensielt farlige bivirkninger, kan det imidlertid være nødvendig å prioritere reduksjon av type II-feil over type I-feil (Aalen *et al.*, 2018).

### 3.8 Spørsmål 7

Ved bruk av pwr pakken, kalkuler styrken av en “one-sample t-test” med en effekt størrelse på 1.5/3, den spesifiserte signifikansnivå og utvalgsstørrelse er 8 og 40. Forklar resultatene ut ifra simuleringen.

```
One-sample t test power calculation
```

```
      n = 8
      d = 0.5
sig.level = 0.05
  power = 0.232077
alternative = two.sided
```

```
One-sample t test power calculation
```

```
      n = 40
      d = 0.5
sig.level = 0.05
  power = 0.8693981
alternative = two.sided
```

Resultatet viser:

for utvalgsstørrelse (n): 8

Statistisk styrke (power): 0.232077

og for utvalgsstørrelse (n): 40

Statistisk styrke (power): 0.8693981

Med et utvalgsstørrelse på 8 har testen en statistisk styrke på bare omtrent 23.2%. Med en utvalgsstørrelse på 40 øker den statistiske styrken til omtrent 86.9%.

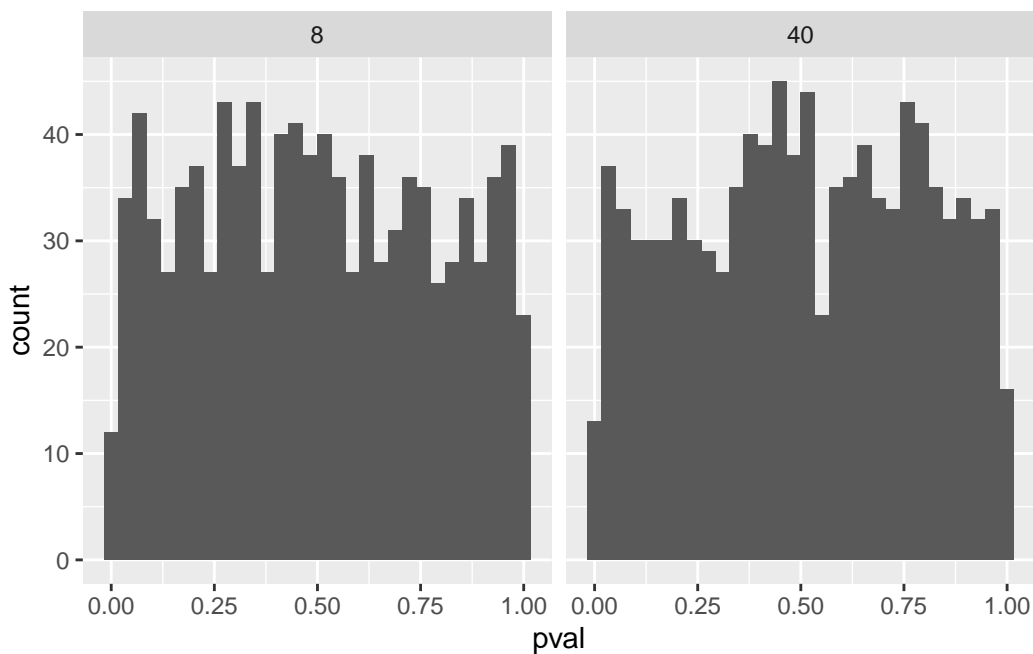
Dette betyr at hvis det faktisk er en reell effekt, er det henholdsvis omtrent 23.2% på  $n=8$  og omtrent 86.9% på  $n=40$  for at testen vil oppdage denne effekten.

Resultatet viser at det er tydelig at utvalgstørrelsen påvirker den statistiske styrken. Dette understreker viktigheten av å ha tilstrekkelig utvalgsstørrelse i studier for å sikre at man kan oppdage reelle effekter når de eksisterer. Lav styrke kan føre til at viktige funn overses, mens høyere styrke gir mer pålitelige resultater (Spiegelhalter, 2020).

#### *Mange studier uten populasjons effect*

Dette er en kopi fra oppgaveteksten som trengs for å besvare de neste spørsmålene.

Ved å bruke den nye datarammen med resultater fra studien med en gjennomsnittlig effect på 0, lag et nytt histogram.



### 3.9 Spørsmål 8

Med et signifikansnivå på 5%, hvor mange studier ville gi deg et 'falskt positivt' resultat hvis du gjorde mange gjentatte studier?

Svar =

[1] 93

Falsk positive resultater er resultater som forkastes selv om nullhypotesen er sann. Siden vi har simulert en populasjon der det ikke skal være en reel effekt mellom behandling og kontroll, skal vi i teorien få ut et tall som er i nærheten av det vi setter signifikante på. Ved et signifikansnivå på 5% (0.05), burde tallet være ca. 50 ved 1000 simuleringer. Jeg fikk imidlertid ut 93, som vil si at jeg har 9.3% falske positive. Siden dette er en simulering så har det i dette tilfellet ført til flere tilfeldige variasjoner som resulterer i falske positive. Dette illustrere et viktig poeng i statistikk, selv når det ikke er en reel effekt, kan man se et betydelig antall falske positive resultater. Dette understreker betydningen av replikasjon i forskning (Hulley *et al.*, 2013; Spiegelhalter, 2020).

Det er også verdt å merke seg at når det ikke er noen reell effekt i populasjonen, vil fordelingen av p-verdier være uniform. Dette betyr at alle p-verdier er like sannsynlige (Spiegelhalter, 2020). Noe som forklarer hvorfor vi ser en jevn fordeling av p-verdier i histogrammet for null-effekt simuleringen.

## 4 Oppgave 4: Study designs

### 4.1 Introduksjon

Dette er oppgave 4: “Study designs” i faget IDR4000-1 24H Kvantitativ metode og statistikk. Dette var en individuell oppgave jeg har gjort små forandringer med basert på tilbakemelding fra fagansvarlig. Vi skal velge oss ut 5 forskningartikler og analysere de ved bruk av QALMRI metoden. Vi fikk instruksjoner om å ikke strukturere rapporten etter QALMRI tabellen, men å bruke QALMRI i beskrivelsene av de valgte studiene i rapporten. Fokus skal være studie design og statistiske modeller eller tester brukt. Avslutningsvis gir jeg en anbefaling om hvordan studier innenfor dette feltet burde settes opp for å svare på lignende spørsmål.

#### 4.1.1 Valg av studier

Jeg har valgt å se på “Sammenhengen mellom fysisk aktivitetsnivå og dødelighet fra alle årsaker i befolkningen”

Fysisk aktivitet(FA) er kjent for å ha en rekke helsefordeler, inkludert forbedret kardiovaskulær helse, bedre mental velvære og redusert risiko for kroniske sykdommer(Stensel *et al.*, 2021). En viktig dimensjon av denne forskningen er hvordan fysisk aktivitetsnivå påvirker dødelighet fra alle årsaker, også kjent som total dødelighet. Dette gir innsikt i hvordan livsstilsintervensjoner kan forbedre folkehelsen(Stensel *et al.*, 2021). Min interesse i dette temaet er drevet av et ønske om å forstå de nyeste funnene innen dette forskningsfeltet og hvordan de kan styrke vår kunnskap om korrelasjonen mellom fysisk aktivitet og dødelighet. Ved å fokusere på total dødelighet, søker jeg å forstå hvordan fysisk aktivitet påvirker generell dødelighet, uavhengig av spesifikke sykdommer som hjerneinfarkt eller diabetes.

For å identifisere relevante studier, utførte jeg et søk i Medline-databasen med søkestrengen “TI (physical activity or exercise) AND TI all-cause mortality AND TI Population”. Jeg valgte å søke i titler for å begrense antall resultater og fokuserte på fagfellevurderte artikler publisert mellom 2019 og 2024. Søkedatoen var 22. oktober 2024, og jeg begrenset språket til engelsk. Dette resulterte i ni treff, hvorav jeg valgte de fem mest relevante basert på egen vurdering.

De fem utvalgte studiene er som følger:

Liang *et al.* (2023) “Joint association of physical activity and sleep duration with risk of all-cause and cause-specific mortality: a population-based cohort study using accelerometry.”

Cho *et al.* (2023) “Aerobic, muscle-strengthening, and flexibility physical activity and risks of all-cause and cause-specific mortality: a population-based prospective cohort of Korean adults”

Lee *et al.* (2023) “Associations of Long-Term Physical Activity Trajectories With All-Cause Mortality in a General Population”

Holtermann *et al.* (2021) “The physical activity paradox in cardiovascular disease and all-cause mortality: the contemporary Copenhagen General Population Study with 104 046 adults”

Li *et al.* (2022) “Long-term physical activity participation trajectories were associated with subsequent cognitive decline, risk of dementia and all-cause mortality among adults aged 50 years: a population-based cohort study”

## 4.2 Studie design

Felles for alle studiene er at de har brukt prospektive kohort-studiedesign. Dette er en observasjonsstudie-design som måler egenskaper hos hver deltaker som kan forutsi fremtidige utfall, og følger deltakerne med målinger av de relevante utfallene (Hulley *et al.*, 2013). I dette tilfellet er det primære utfallet dødelighet, mens fysisk aktivitet (FA) er den hovedprediktoren som studeres. En stor fordel med denne type design er at det tillater kalkulering av insidens - antallet nye tilfeller av en tilstand som oppstår over tid (Hulley *et al.*, 2013).

Ved å måle nivåene av prediktoren (FA) før utfallet (død) skjer, etablerer tidsrekkefølgen for variablene, noe som styrker prosessen med å trekke konklusjoner om årsakssammenhenger. I tillegg vil det prospektive designet forhindre at prediktorene blir påvirket av utfallet, og variablene kan måles mer komplett (Hulley *et al.*, 2013). Blant studiene, er Liang *et al.* (2023) den eneste som målte FA objektivt med akselerometer over 7 dager, noe som gir mer presise data enn selvrappotering, men er mere kostbart (Hulley *et al.*, 2013). De andre studiene (Holtermann *et al.*, 2021; Li *et al.*, 2022; Cho *et al.*, 2023; Lee *et al.*, 2023) baserte seg på selvrappotert FA. I motsetning vil et retrospektivt kohortestudie som ser på dødelighet, måtte de ha rekonstruere informasjon om de avdøde fra indirekte kilder som medisinske journaler eller slektninger (Hulley *et al.*, 2013). Dette gjelder både for FA-data, og for all informasjon om den avdøde, noe som kan påvirke nøyaktigheten av dataene.

Dette studiedesignet har samme svakheter som alle andre observasjonsstudier. Det er utfordrende å trekke slutninger om årsakssammenhenger, og tolkningen blir ofte uklar på grunn av påvirkningen fra konfunderende variabler (Hulley *et al.*, 2013). For studier som ser på FA og dødelighet, kan faktorer som kosthold, sosioøkonomisk status, og generell helsetilstand være viktige konfunderende variabler (Stensel *et al.*, 2021). En konfunderende variabel i denne sammenhengen er en faktor som påvirker både fysisk aktivitetsnivå og dødelighet, men som ikke er en del av den kausale sammenhengen vi ønsker å studere. For eksempel kan sosioøkonomisk status påvirke både muligheten for å drive med fysisk aktivitet på fritiden og tilgangen til helsetjenester, som igjen påvirker dødelighet. Dette kan føre til at vi overvurderer effekten av

fysisk aktivitet på dødelighet hvis vi ikke kontrollerer for sosioøkonomisk status. Mekanismen som gjør konfunderende variabler problematiske i observasjonsstudier er at de kan skape en tilsynelatende sammenheng mellom eksponering og utfall, eller maskere en reell sammenheng, uten at det faktisk er en direkte årsakssammenheng.

En annen svakhet ved prospektive design er at de kan være kostbare og ineffektive for å studere sjeldne utfall (Hulley *et al.*, 2013). I konteksten av disse studiene er dødelighet ikke et sjeldent utfall, men snarere et vanlig endepunkt som observeres over tid i store befolkningsgrupper. Dette gjør prospektive kohortstudier godt egnet for å undersøke sammenhengen mellom FA og dødelighet.

Det er hensiktsmessig å bruke dette studiedesignet for denne typen studier, fremfor intervensjonsstudier som randomiserte kontrollerte studier (RCT) (randomized controlled/clinical trials). Kohortstudier og RCT har ulike styrker og bruksområder innen medisinsk forskning. RCT-er regnes ofte som gullstandarden for å etablere årsakssammenhenger, men de har begrensninger når det gjelder å studere langtidseffekter av livsstilsfaktorer som fysisk aktivitet på dødelighet (Hulley *et al.*, 2013). RCT-er er ofte ressurskrevende både i tid og penger, adresserer ofte snevre forskningsspørsmål, og kan potensielt utsette deltakere for skade (Hulley *et al.*, 2013). Det kan være etisk problematisk å plassere mennesker i en gruppe hvor de skal være fysisk inaktive, spesielt over lengre tid, fordi vi allerede har sterk evidens for de positive helseeffektene av fysisk aktivitet. Å be noen om å være inaktive kan potensielt utsette dem for helserisiko. Dette etiske dilemmaet er en viktig årsak til at RCT-er for fysisk inaktivitet ikke er så vanlige i dette forskningsfeltet. I stedet brukes ofte observasjonsstudier eller intervensjoner som øker aktivitetsnivået fra et baseline-nivå.

Kohortstudier tillater observasjon av et bredere spekter av livsstilmønstre og kan følge deltakere over flere tiår, noe som er essensielt for å studere utfall som dødelighet. Derfor er kohortstudier ofte det mest hensiktsmessige og etisk gjennomførbare designet for å studere sammenhengen mellom fysisk aktivitet og langsiktige helseutfall, til tross for at de ikke kan bevise kausalitet like sterkt som RCT-er (Hulley *et al.*, 2013).

Oppfølgingstiden varierer også betydelig mellom studiene, fra 6 år (Li *et al.*, 2022) til 16,8 år (Lee *et al.*, 2023), noe som kan påvirke styrken av de observerte sammenhengene mellom FA og dødelighet (Hulley *et al.*, 2013). Utvalgsstørrelsen og aldere varierte også betydelig mellom studiene, fra Li *et al.* (2022) med 8,842 deltakere over 50 år i England, til Holtermann *et al.* (2021) som inkluderte hele 104,046 voksne i alderen 20-100 år fra den generelle befolkningen i København.

Studiene representerer en bred geografisk spredning, fra Liang *et al.* (2023) som brukte data fra den generelle britisk populasjonen, Cho *et al.* (2023) som fokuserte på en koreansk voksenpopulasjon, Lee *et al.* (2023) som studerte en taiwansk by sin befolkning, Holtermann *et al.* (2021) som undersøkte innbyggere i København, til Li *et al.* (2022) som analyserte data fra eldre voksne i England. Denne geografiske variasjonen kan påvirke resultatene gjennom forskjeller i livsstil, kosthold, helsetjenester, og kulturelle holdninger til fysisk aktivitet mellom

landene. For eksempel kan den urbane befolkningen i København ha andre muligheter for fysisk aktivitet enn den mer varierte populasjonen i den britiske studien. Videre kan kulturelle forskjeller i arbeids- og fritidsaktiviteter mellom asiatiske og europeiske land påvirke hvordan fysisk aktivitet relaterer seg til helseutfall (Stensel *et al.*, 2021).

#### 4.2.1 De spesefikke studiene

Hver studie fokuserer på spesifikke aspekter ved sammenhengen mellom fysisk aktivitet og dødelighet:

Liang *et al.* (2023): Undersøker sammenhengen mellom akselerometer-målt fysisk aktivitet og søvnvarighet med risiko for dødelighet fra alle årsaker og kardiovaskulære årsaker. Studien inkluderte 92,221 deltakere fra Storbritannia (gjennomsnittlig alder 62.4 år) og hadde en median oppfølgingstid på 7,0 år.

Cho *et al.* (2023): Fokuserer på effekten av aerob og muskelstyrkende fysisk aktivitet, både separat og kombinert, samt fleksibilitetsøvelser, på risikoen for dødelighet fra alle årsaker. Studien omfattet 34,379 koreanske voksne i alderen 20-79 år. Gjennomsnittlig oppfølgingstid 9.2 år (ikke median eller konkret oppfølgingstid som de andre studiene rapporterte).

Lee *et al.* (2023): Analyserer langvarige mønstre av fysisk aktivitet og deres sammenheng med dødelighet fra alle årsaker. Studien inkluderte 21,211 deltakere fra Taipei i Taiwan i alderen 18-90 år, med en median oppfølgingstid på 16.8 år.

Holtermann *et al.* (2021): Utforsker det fysiske aktivitetsparadokset ved å sammenligne effekten av arbeidsrelatert og fritidsrelatert fysisk aktivitet på kardiovaskulær sykdom og allmenn dødelighet. Studien omfattet 104,046 deltakere i alderen 20-100 år fra København i Danmark, med 10 års median oppfølgingstid.

Li *et al.* (2022): Undersøker hvordan langvarig deltakelse i fysisk aktivitet påvirker kognitiv nedgang, risiko for demens, og allmenn dødelighet hos voksne over 50 år. Studien inkluderte 8,842 deltakere fra England og hadde en oppfølgingstid på 6 år.

### 4.3 Statistiske modeller

Alle studiene bruker Cox proportional hazards modeller for å analysere sammenhengen mellom fysisk aktivitet og dødelighet. Denne metoden er spesielt egnet for kohortstudier hvor det kan forekomme betydelig frafall, ulik oppfølgingstid, eller hendelser som hindrer måling av utfallet (Hulley *et al.*, 2013). Cox-modellen håndterer disse utfordringene ved å analysere “tid-til-hendelse” data, hvor hver deltaker bidrar med person-tid fra de blir med i studien til de enten opplever utfallet (død) eller blir “sensurert” på grunn av studieslutt eller frafall (Hulley *et al.*, 2013).



I stedet for å sammenligne rene antall utfall, fokuserer Cox-modellen på insidensrater - antall utfall delt på total person-tid i risiko for hver gruppe. Dette gjør det mulig å justere for ulike oppfølgingstid mellom deltakerne. Modellen estimerer hazard ratios, som ligner på rate ratios, og representerer den relative risikoen for at en hendelse skal inntreffe, gitt ulike nivåer av eksponering (som fysisk aktivitet) (Hulley *et al.*, 2013).

Cox-modellen tillater også multivariat analyse, noe som betyr at den kan justere for flere kovariater eller konfunderende faktorer samtidig. Dette er spesielt viktig i observasjonsstudier hvor man ønsker å isolere effekten av fysisk aktivitet på dødelighet, samtidig som man kontrollerer for andre relevante faktorer (Hulley *et al.*, 2013). Denne egenskapen ved Cox-modellen er særlig verdifull i lys av Hulley *et al.* (2013) nevner at konfunderende variabler er vanskelige å kontrollere for i kohortstudier, og at tolkningen av resultater ofte blir “grumsete” på grunn av innflytelsen fra slike variabler. Ved å bruke multivariat analyse i Cox-modellen, kan forskerne bedre håndtere denne utfordringen og oppnå mer pålitelige estimater av sammenhengen mellom fysisk aktivitet og dødelighet. Selv om jeg vil tro at resultatene fortsatt må tolkes med forsiktighet.

I de gjennomgåtte studiene har forskerne kontrollert for en rekke potensielle konfunderende variabler. For eksempel justerte Holtermann *et al.* (2021) for alder, kjønn, utdanningsnivå, BMI, blodtrykk, røyking, alkoholinntak, kosthold, og kroniske sykdommer som diabetes og KOLS. Liang *et al.* (2023) kontrollerte for faktorer som alder, kjønn, etnisitet, utdanningsnivå, sosioøkonomisk status (målt ved Townsend deprivation index), BMI, kostholdsrelaterte faktorer, røyking, alkoholinntak, og eksisterende sykdommer som hjerte- og karsykdommer, kreft eller diabetes. Ved å inkludere disse variablene i analysene, forsøker forskerne å isolere effekten av fysisk aktivitet på dødelighet.

Li *et al.* (2022) og Lee *et al.* (2023) bruker også gruppebasert trajektoriemodellering (group-based trajectory modeling, GBTM) for å identifisere mønstre i fysisk aktivitet over tid, i tillegg til Cox-modellen. GBTM er en statistisk metode utviklet for å analysere hvordan atferd eller tilstander endrer seg over tid eller alder (Nagin, 2014). Denne metoden er spesielt nyttig for å finne undergrupper i en befolkning som følger ulike utviklingsbaner, noe som er relevant når man studerer hvordan folks fysiske aktivitetsnivå endrer seg over tid. GBTM bruker en type statistisk modellering for å identifisere grupper med lignende utviklingsbaner og viser resultatene i lettforståelige figurer og tabeller. Dette gjør det mulig for forskerne å oppdage og formidle komplekse mønstre i langtidsdata på en måte som er forståelig for både fagfolk og andre (Nagin, 2014). I studier om fysisk aktivitet og dødelighet hjelper GBTM forskerne med å finne ulike grupper av personer basert på hvordan deres fysiske aktivitetsnivå endrer seg over lengre tid. Deretter kan de undersøke hvordan disse ulike aktivitetsmønstrene henger sammen med risikoen for å dø.

Holtermann *et al.* (2021) nevner “The median follow-up time was based on the reverse Kaplan–Meier method”. Det er den eneste gangen artikkelen refererer til denne metoden. Referansen Holtermann *et al.* (2021) gir går til en artikkel om at forskningsartikler bør legge ved hvilken metode de bruker til å finne oppfølgingstid. Gillespie *et al.* (2010) forklarer at reverse Kaplan–Meier metode er en statistisk metode som er spesielt nyttig når man har data hvor noen

verdier er for lave til å måles nøyaktig, eller når man har sensurerte data (left-censored data). I konteksten av studien til Holtermann *et al.* (2021) antar jeg at metoden brukes for å estimere median oppfølgingstid når flere deltakere fortsatt er i live ved studiens slutt. I følge Gillespie *et al.* (2010) er dette en mer presis måte å beregne oppfølgingstid på enn å bruke flere andre metoder som ofte blir brukt, spesielt i studier med lang oppfølgingstid hvor mange deltakere ikke opplever utfallet (i dette tilfellet, død). Holtermann *et al.* (2021) hadde flere oppfølgingstidspunkterpunkter under studien som gikk fra 0.1- 15 år med 10 år som median. Jeg tror reverse Kaplan–Meier method ble brukt for å gi et mer nøyaktig bilde av den faktiske oppfølgingstiden i studien og kan påvirke hvordan resultatene tolkes, i dette tilfellet langtidseffekter av fysisk aktivitet fra enten fritid eller gjennom jobb på dødelighet.

## 4.4 Diskusjon:

Ikke overraskende fant alle studiene konsistent evidens for at regelmessig fysisk aktivitet er assosiert med lavere risiko for dødelighet av alle årsaker. De viser viktigheten av å opprettholde stabile nivåer av fysisk aktivitet over tid. Li *et al.* (2022) understreker at det at det aldri er for sent å begynne med fysisk aktivitet for å oppnå helsefordeler, selv for de som har vært inaktive tidligere i livet. Lee *et al.* (2023) viser at det ikke bare er høyt nivå av fysisk aktivitet som er viktig, men også stabiliteten eller økningen i aktivitetsnivået over tid som har betydning for overlevelse. Dette var spesielt merkbart ved død fra kreft. Liang *et al.* (2023) var den eneste studien som ikke brukte selvrappporterte spørreskjemaer for å måle fysisk aktivitet. Det at de brukte akselortmeter styrker funnene ytterligere. Liang *et al.* (2023) så også på søvnkvalitet og lengde, hvor for mye og for lite sammen med lite fysisk aktivitet var assosiert med høyere dødelighet. De så at høyere fysisk aktivitet var assosiert med lavere dødelighet, og kan redusere den negative effekten assosiert med unormal søvnlengde til en viss grad.

Studiene viser samlet sett også at ulike typer fysisk aktivitet kan ha gunstige effekter. Cho *et al.* (2023) så i tillegg på forskjellige typer aktivitet, aerob, muskelstyrkende og fleksibilitet kan alle ha gunstig effekt. Selv om intensitet og type har varierende effekt. Cho *et al.* (2023) nevner at det var svakheter i spørreskjemaet samt oppfølging som kan ha påvirket resultatet, men er sikre på at det er positive effekter på alle tre typer trening, da i varierende grad. Et interessant funn ved Holtermann *et al.* (2021) er at høyere nivåer av fritidsaktivitet var assosiert med lavere risiko for dødelighet, mens høyere nivåer av arbeidsrelatert fysisk aktivitet var assosiert med økt risiko, uavhengig av hverandre. Som antyder at konteksten for aktiviteten kan være viktig. Selv om kovariabler som arbeidsrelatert fysisk aktivitet er assosiert med lavere sosioøkonomisk status har Holtermann *et al.* (2021) forsøkt å ta hensyn til dette, og mange andre kovariabler ved bruk av Cox regression modeller.

Samlet sett støtter disse funnene sterkt anbefalinger om regelmessig fysisk aktivitet som en viktig faktor for å redusere dødelighetsrisiko og forbedre helse. Studiene gir et nyansert bilde av hvordan ulike aspekter ved fysisk aktivitet - fra type og intensitet til kontekst og livstidsmøn-

stre - påvirker helse og dødelighet. Dette kan ha viktige implikasjoner for både individuell helseadferd og utforming av folkehelsepolitikk.

Studiedesignene som ble brukt i disse undersøkelsene, prospektive kohortstudier, ble valgt for å etablere tidsmessige sammenhenger mellom fysisk aktivitet og dødelighet. Logikken bak dette valget er at ved å måle fysisk aktivitet før utfallet (død) inntreffer, kan man etablere tidsmessige sammenhenger og redusere risikoen for omvendt kausalitet. Imidlertid er det viktig å merke seg at selv med dette designet kan man ikke fullstendig utelukke muligheten for at underliggende helsetilstander påvirker både aktivitetsnivå og dødelighetsrisiko (Hulley *et al.*, 2013).

Studiene brukte hovedsakelig Cox proporsjonal hazard modeller som statistisk metode. Denne modellen ble valgt fordi de kan håndtere data som endrer seg over tid og samtidig justere for flere ulike faktorer som kan påvirke resultatene. Dette valget reflekterer en erkjennelse av at sammenhengen mellom fysisk aktivitet og dødelighet er kompleks og potensielt påvirket av mange faktorer. Bruken av gruppebasert trajektoriemodellering i studiene til Li *et al.* (2022) og Lee *et al.* (2023) representerer et forsøk på å fange opp langsiktige mønstre i fysisk aktivitet, noe som kan gi en mer nyansert forståelse av sammenhengen med dødelighet enn enkle punktmålinger.

Alternative forklaringer som ikke fullt ut kan adresseres av disse designene og modellene inkluderer potensielle uobserverte konfunderende faktorer og seleksjonsbias. For eksempel kan personer som opprettholder høye nivåer av fysisk aktivitet over tid også ha andre helsefremmende atferder eller genetiske faktorer som bidrar til lavere dødelighet. Selv om studiene justerte for mange kjente konfunderende faktorer, kan det fortsatt være uobserverte variabler som påvirker resultatene (Stensel *et al.*, 2021). Seleksjonsbias oppstår når de som blir valgt til utvalget i studien systematisk avviker fra de som ikke blir valgt. I konteksten av disse studiene kan seleksjonsbias oppstå hvis for eksempel de som velger å delta i studien er generelt mer helsebevisste eller har bedre helse enn gjennomsnittet i befolkningen (Stensel *et al.*, 2021). Dette kan føre til en overestimering av de positive effektene av fysisk aktivitet. Forskerne forsøker å minimere seleksjonsbias gjennom nøye utvalgsmetoder og statistiske justeringer, men det er vanskelig å eliminere helt i observasjonsstudier (Stensel *et al.*, 2021).

Videre er det viktig å vurdere begrensningene ved selvrapportert fysisk aktivitet, som ble brukt i de fleste studiene. Liang *et al.* (2023) brukte akselerometer-baserte målinger, som gir mer objektive data, dette kan introdusere andre former for bias. Som at deltagerne endrer atferd når de vet de blir observert. Jeg tror imidlertid at hvis folk overdriver treningen når de blir målt, som ikke er representativt for deres faktiske aktivitetsnivå ville jeg tro denne biasen ville talt imot fysisk aktivitet. Fremtidige studier kunne dra nytte av å kombinere objektive målinger med selvrapportering over lengre perioder for å fange opp både kvantitet og kvalitet av fysisk aktivitet mer nøyaktig.

## 4.5 Konklusjon:

Denne gjennomgangen av fem nyere studier om sammenhengen mellom fysisk aktivitet og dødelighet bekrefter den sterke assosiasjonen mellom regelmessig fysisk aktivitet og redusert risiko for dødelighet av alle årsaker. Studiene, som alle benyttet prospektive kohort-design og Cox proportional hazards modeller, gir et nyansert bilde av hvordan ulike aspekter ved fysisk aktivitet påvirker dødelighetsrisiko.

Nøkkelfunn inkluderer viktigheten av å opprettholde stabile eller økende nivåer av fysisk aktivitet over tid, den potensielle beskyttende effekten av ulike typer fysisk aktivitet (aerob, muskelstyrkende, og fleksibilitet), og det interessante paradokset mellom fritids- og arbeidsrelatert fysisk aktivitet. Studiene understreker også at det aldri er for sent å begynne med fysisk aktivitet for å oppnå helsefordeler, selv for eldre voksne.

Metodologisk sett demonstrerer disse studiene styrken ved å bruke avanserte statistiske metoder. Cox proportional hazards modeller med multivariat analyse ble brukt for å justere for flere kovariater, mens gruppebasert trajektoriemodellering ble anvendt for å fange opp langsiktige mønstre i fysisk aktivitet. Samtidig belyser studiene utfordringene ved å kontrollere for konfunderende faktorer i observasjonsstudier, spesielt gitt den komplekse naturen av sammenhengen mellom fysisk aktivitet og dødelighet.

Basert på min vurdering av disse studiene, mener jeg at fremtidige undersøkelser innen dette feltet bør vurdere å øke bruken av gruppebasert trajektoriemodellering, særlig i studier med store utvalg, store aldersforskjeller og har faktorer som kan dele utvalget i flere “undergrupper”. Etter min oppfatning vil dette kunne bidra til å fange opp mønstre i fysisk aktivitet over tid på en bedre måte, noe jeg anser som viktig for utfallet. Videre ser jeg et behov for at forskningen utforsker paradokset mellom fritids- og arbeidsrelatert fysisk aktivitet mer inngående, med spesielt fokus på spesifikke typer arbeidsrelatert aktivitet og yrker som kan være mest utsatt.

Personlig vil jeg anbefale å kombinere objektive målemetoder, som akselerometri, med selvrapping for å oppnå mer presise og omfattende data om fysisk aktivitet. Jeg er klar over at dette kan medføre utfordringer med tanke på ressurser, men jeg tror likevel at denne kombinasjonen av metoder vil kunne gi et mer nyansert og helhetlig bilde av individers aktivitetsmønstre og deres innvirkning på helse og dødelighet.

## 5 Oppgave 5: Analyzing repeated measures experiments

### 5.1 Introduksjon

Dette er oppgave 5: “Analyzing repeated measures experiments” i faget IDR4000-1 24H Kvantitativ metode og statistikk for master i treningsfysiologi. Dette var en individuell oppgave jeg har gjort små forandringer med basert på tilbakemelding fra fagansvarlig. I denne oppgaven har jeg analysert og rapportert om en studie som undersøker hvordan treningsvolum ved styrketrening påvirker mager kroppsmasse og muskelstyrke. Jeg har brukt data fra `exscidata`-pakken, nærmere bestemt datasettene “`strengthvolume`” og “`dxadata`” som ligger i RStudio.

#### 5.1.1 Bakgrunn og betydning av styrketrening

Styrketrening drives både for å få bedre fysisk prestasjonsevne, forebygge skade, rehabilitering etter skade, bedre velvære, utseende og for å oppnå god helse (Raastad *et al.*, 2010). Styrketrening utføres i dag av mange ulike grupper, som idrettsutøvere, mosjonister og pasienter under opptrening. Frem til begynnelsen av 2000-tallet var det stor skepsis mot tung styrketrening i et helseperspektiv. Det var bare noen idrettsgrupper som vektløftere, styrkeløftere, brytere, kasteutøvere i friidrett og kroppsbyggere som drev med tung styrketrening. Oppfatningen blant andre idrettsutøvere var at slik trening førte til treghet i muskulaturen og redusert bevegelighet. Denne holdningen har heldigvis endret seg radikalt etterhvert som vi har fått bedre kunnskap om effekten av styrketrening (Raastad *et al.*, 2010).

På lik linje med kardiovaskulær form er muskelstyrke et fysiologisk trekk som både blir påvirket av fysisk aktivitet, særlig motstandstrening, og genetiske faktorer. Det er imidlertid, sammenlignet med studier på kardiovaskulær form, få studier som har sett på muskelstyrke og dødelighet (Stensel *et al.*, 2021). Stensel *et al.* (2021) skriver også at i prospektive kohortstudier som har sett på både måling av fysisk aktivitet og målinger av styrke eller kardiovaskulær form, har de to sistnevnte betraktelig større positiv effekt på dødelighet. Stensel *et al.* (2021) sier at dette kan skyldes svakheter i selvrapporteringsmålinger av fysisk aktivitet, men det kan også peke på de indre genetiske variablene for kardiovaskulær form eller muskelstyrke som kan påvirke dødelighetsrisiko uavhengig av fysisk aktivitet.

Studien til Hammarström *et al.* (2020) som vi ser på i denne oppgaven, tar for seg effekten av ulikt treningsvolum på muskelstyrke og muskelvekst i bein hos moderat trente individer.

Siden deltakerne er sin egen kontroll ved at de trener hver fot med ulikt volum, kan vi få et bilde av hvordan muskelstyrke og muskelvekst blir påvirket av volum uavhengig av eventuelle genetiske variabler som Stensel *et al.* (2021) nevner.

### 5.1.2 Treningsvolum og muskulære tilpasninger

Innenfor styrketrening er det primært hvor ofte og hvor hardt hver enkelt muskelgruppe blir trent, som bestemmer treningseffekten. Kombinasjonen av treningsvolum, treningsintensitet, grad av anstrengelse og treningsfrekvens utgjør den totale treningsbelastningen vi utsettes for. Treningsvolumet i hver økt kan grovt regnes som produktet av antall serier  $\times$  antall repetisjoner  $\times$  treningsmotstand for den aktuelle muskelgruppen (Raastad *et al.*, 2010).

Det grunnleggende prinsippet bak et styrke- eller motstandstreningsprogram er overbelastningsprinsippet, som sier at styrkeøkning skjer som et resultat av systematisk og progressiv trening med tilstrekkelig frekvens, intensitet og varighet for å skape tilpasning (Wackerhage, 2014). Å planlegge et treningsprogram for motstandstrening krever derfor nøye valg av volum, belastning, hvile mellom settene og progresjon. I tillegg bør målene med treningen (styrke, kraft, hypertrofi), type utstyr, tilgjengelighet av utstyr og muskelhandling (eksentrisk, konsentrisk, isometrisk, raskt/tregt) vurderes. Det er også viktig at treningsprogrammet tilpasses individet, da ulike programmer er nødvendige for utrente individer, atleter, barn og pasienter som rehabiliteres etter skade (Wackerhage, 2014).

For å få et nøyaktig mål på treningsvolum burde vi ideelt sett måle det indre arbeidet en muskelgruppe utfører for å overvinne ytre motstander. Siden dette i praksis ikke er mulig, kan vi likevel bruke den nevnte metoden for å indikere endringer i treningsvolum over tid, så lenge vi holder oss til de samme øvelsene. Selv om dette ikke gir et fullstendig korrekt bilde av det indre arbeidet en muskelgruppe har utført, gir det en god indikasjon på volumets utvikling (Raastad *et al.*, 2010).

Selv med et optimalt designet treningsprogram vil omfanget av muskelmasse- og styrkeøkning variere dramatisk mellom individer (Wackerhage, 2014). Ifølge Wackerhage (2014) er det, basert på pålitelige vitenskapelige data, mye forvirring rundt hva som utgjør et optimalt treningsprogram, og vi kan egentlig ikke si at et universelt optimalt motstandstreningsprogram finnes. I praksis kan vi si at alt fra 1-5 sett med 6-12 repetisjoner, dersom individet trener til utmattelse i settene, vil føre til betydelige økninger i både muskelstørrelse og styrke. Det er imidlertid viktig å være oppmerksom på at forskjeller i motivasjon, gener, ernæring og individuelle mål vil påvirke responsen på treningen og dermed føre til stor variasjon mellom individene (Wackerhage, 2014).

Forståelsen av sammenhengen mellom muskulære tilpasninger og volum er avgjørende for effektiv trening. Når vi vet hvordan ulike volum påvirker styrke og hypertrofi, kan vi skreddersy treningsprogrammer som maksimerer resultater for ulike individer (Raastad *et al.*, 2010). Dette er spesielt viktig når vi vet at individuelle forskjeller i respons på trening kan være betydelige.

Å ha en solid forståelse av hvordan volumet påvirker muskulære tilpasninger gjør det mulig å optimalisere treningen for spesifikke målsettinger og behov.

### 5.1.3 Kunnskapshull og forskningsbehov

Som nevnt av både Raastad *et al.* (2010) og Stensel *et al.* (2021) er motstandstrening og effekten muskelmasse og styrke har på helsen vår og prestasjon, både i idrett (med unntak av noen få sporter) og i hverdagen til folk, et relativt nytt forskningsområde. Vi har en mye større forståelse av hvordan kardiovaskulær form påvirker helsen vår og hvordan vi bør trene for å forbedre kondisjonen, da det er mye mer forskning på dette området. Vi ser imidlertid at det er flere helsefordeler med motstandstrening, økt styrke og muskelmasse, spesielt ettersom vi blir eldre (ny referanse for å dekke kravet om 5 referanser? Finn en referanse senere, jeg vet jeg har lest det et sted). At vi får mer forskning som ser på hvordan forskjellige former og volum av motstandstrening påvirker vår evne til å bygge mer muskler og/eller styrke, er med på å fylle dette kunnskapshullet.

Da jeg begynte å trene for 19 år siden, var det vanlig “bro science” at høyt volum og moderate vekter var best for å bygge mest muskler, mens lavt volum med høye vekter var best for å bygge mest styrke. Jeg har sett mer forskning de senere årene som tyder på at dette ikke nødvendigvis er helt sant, og at begge treningsmetodene har en større overlapp enn det vi tidligere trodde.

### 5.1.4 Forsknings spørsmål

Hvordan påvirker ulike treningsvolum (ett sett vs. tre sett) muskelhypertrofi i underekstremitetene hos moderat trente individer over en 12-ukers treningsperiode, og er det forskjeller i responsen mellom kjønn?

## 5.2 Metode

Denne studien undersøkte effekten av treningsvolum på muskelhypertrofi og styrke ved å sammenligne ett sett mot tre sett i beinøvelser. Metoden inkluderer beskrivelse av deltakere, studiedesign, treningsprotokoll og datainnsamlingsprosedyrer.

### 5.2.1 Deltagere

Studien rekrutterte opprinnelig 41 deltakere i alderen 18 til 40 år. Inklusjonskriteriene omfattet at deltakerne måtte være ikke-røykere og ha begrenset erfaring med styrketrening, definert som ikke mer enn én ukentlig styrketreningsøkt i løpet av de siste 12 månedene før studien. Eksklusjonskriterier inkluderte intoleranse mot lokalbedøvelse, nedsatt muskelstyrke på grunn

	Alder (år)	Høyde (cm)	Kroppsmasse (kg)
Kvinne, n = 18	22 (1.3)	167.7 (6.9)	64.4 (10.4)
Mann, n = 16	23.6 (4.1)	182.9 (5.9)	75.8 (10.7)

Verdier er presentert som gjennomsnitt (standardavvik)

av tidligere eller nåværende skade, og bruk av reseptbelagte medisiner som kunne påvirke treningsadaptasjoner. I løpet av studien ble syv deltakere (fire kvinner og tre menn) ekskludert fordi de ikke fullførte minst 85% av de planlagte treningsøktene.

Alle inkluderte deltakere rapporterte tidligere erfaring med sportslige aktiviteter. Tjue deltakere oppga at de var engasjert i fysisk trening ved studiestart (median antall økter per uke: 2, spennvidde: 0,5-4), hvorav 10 utførte sporadisk styrketrening, men ingen mer enn én gang i uken.

## 5.2.2 Studiedesign

Denne studien brukte en within-subject design, hvor hver deltaker fungerte som sin egen kontroll. Deltakerne trente ett bein med ett sett (single-set condition) og det andre beinet med tre sett (multiple-set condition) per øvelse. Tildelingen av treningsvolum til hvert bein var randomisert for hver deltaker. Denne designen muliggjør en direkte sammenligning av effekten av treningsvolum på muskelhypertrofi og styrke innenfor samme individ, noe som reduserer påvirkningen av individuelle forskjeller.

### 5.2.2.1 treningsprotokoll

Intervensjonen bestod av 12 uker med helkroppss styrketrening. Treningsprogrammet fokuserte spesielt på beinøvelser, som ble utført unilateralt for å muliggjøre differensiering av treningsvolum innenfor hver deltaker. Beinøvelsene inkluderte unilateral beinpress, leg curl og kneekstensjon. For beinet som var tildelt single-set condition, ble det utført ett sett av hver øvelse, mens for beinet tildelt multiple-set condition ble det utført tre sett. Treningsintensiteten ble gradvis økt gjennom intervensjonen, startende med 10RM de første to ukene, etterfulgt av 8RM i tre uker, og avsluttende med 7RM i syv uker.

### 5.2.3 Datainnsamling

For å besvare forskningsspørsmålet ble det samlet inn data om deltakernes kroppssammensetning og muskelstyrke. Datainnsamlingen fokuserte på to hovedområder: målinger av mager kroppsmasse og vurdering av muskelstyrke gjennom ulike styrketester.



### 5.2.3.1 målinger av kroppsmasse

Mager kroppsmasse ble målt ved hjelp av dual-energy X-ray absorptiometry (DXA) før og etter treningsintervensjonen. DXA-skanningen ble utført i henhold til standardprotokoll, og ga detaljert informasjon om kroppssammensetning, inkludert mager masse i begge bein. Deltakerne ble bedt om å faste i 2 timer og avstå fra hard fysisk aktivitet i 48 timer før målingene for å sikre konsistente forhold.

### 5.2.3.2 Måling av styrke test

Muskelstyrke ble vurdert gjennom flere tester. En repetisjon maksimum (1RM) ble målt i unilateral beinpress og kneekstensjon. Testen startet med en spesifikk oppvarming, etterfulgt av progressive økninger i vekt til deltakeren ikke kunne løfte vekten gjennom full bevegelsesbane. Isokinetisk og isometrisk unilateral kneekstensjonsstyrke ble målt ved hjelp av et dynamometer. Isokinetisk styrke ble testet ved tre vinkelhastigheter (60°, 120° og 240° per sekund), mens isometrisk styrke ble målt ved en knevinkel på 30°.

Disse styrketestene ble utført før intervensjonen startet og etter at den 12-ukers treningsperioden var fullført. For baselinemålingene ble styrketestene utført to ganger, med minst fire dagers mellomrom, og den høyeste verdien ble brukt i analysene.

## 5.3 Data analyse

For å undersøke effekten av treningsvolum på muskelhypertrofi, ble dataene fra Hammarström *et al.* (2020) analysert. Datasettet inneholdt målinger av mager kroppsmasse for venstre og høyre bein hos hver deltaker, før og etter intervensjonen. Dataene ble behandlet ved å beregne endringen i mager kroppsmasse for hvert bein, og deretter kategorisere disse endringene basert på treningsvolumet (ett sett eller tre sett). Den primære statistiske metoden som ble brukt var en paret t-test for å sammenligne endringene i mager kroppsmasse mellom ett og tre sett. Denne metoden ble valgt på grunn av studiens within-subject design, hvor hvert individ fungerte som sin egen kontroll.

For å få et mer detaljert bilde av individuelle responser, ble det laget et lineplot som viser endringer for hver enkelt deltaker:

Figure 5.1 illustrerer de individuelle endringene i mager kroppsmasse for hver deltaker. Hver linje representerer en deltaker, og viser endringen fra ett sett til tre sett. Dette plottet avslører den betydelige variasjonen i individuell respons på treningsvolumet, med noen deltakere som viser større økning ved tre sett, mens andre viser liten forskjell eller til og med større økning ved ett sett.

For å undersøke forholdet mellom pre- og post-intervensjonsverdier og hvordan dette påvirkes av treningsvolum, ble det laget et scatterplot:

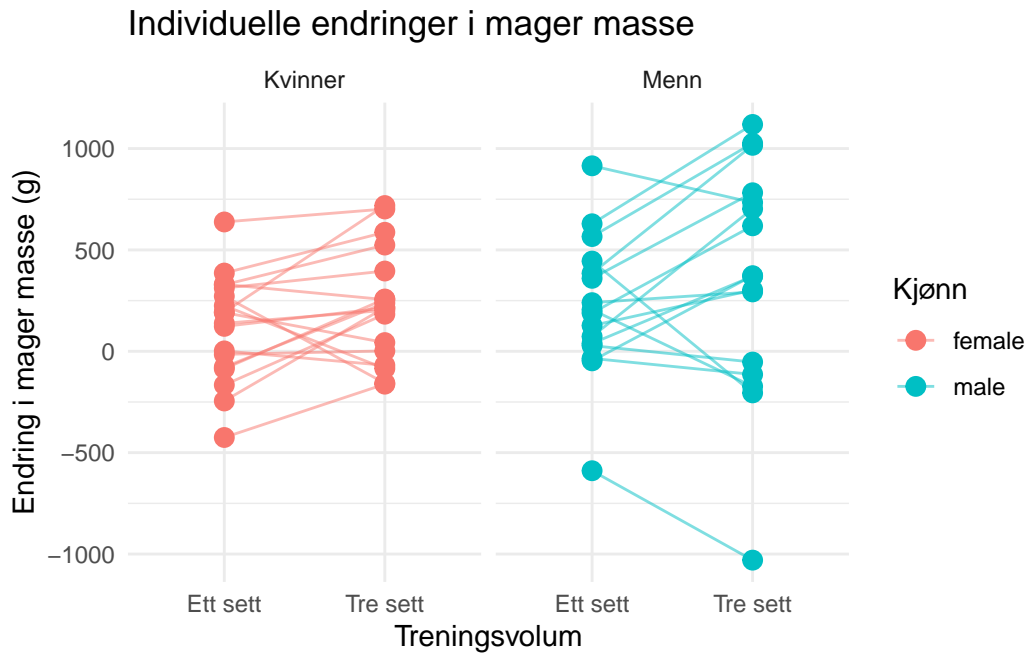


Figure 5.1: Individuelle endringer i mager kroppsmasse fra pre- til post-intervensjon.

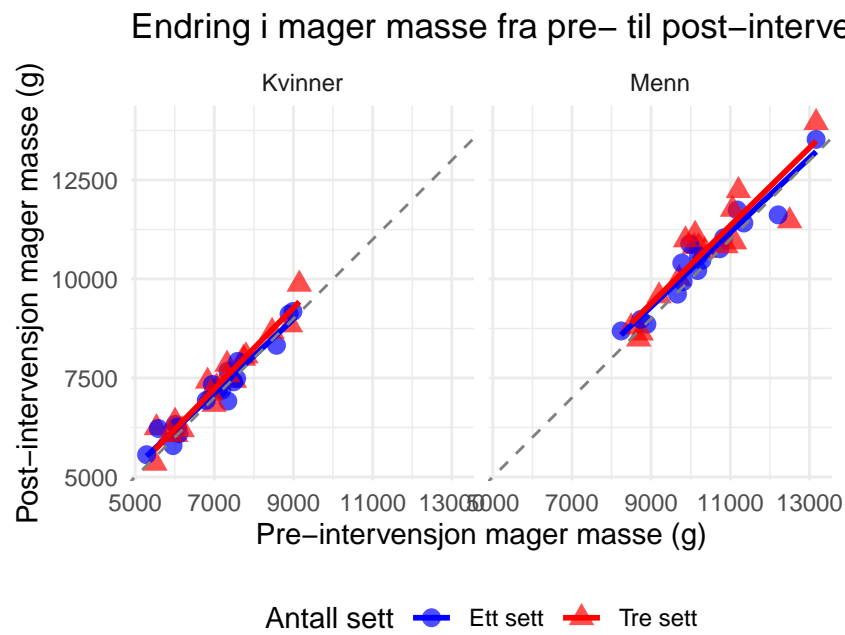


Figure 5.2: Forholdet mellom pre- og post-intervensjonsverdier for mager masse.

Figure 5.2 viser forholdet mellom pre- og post-intervensjonsverdier for mager masse, med separate punkter og regresjonslinjer for ett sett og tre sett. Punkter over den stiplede linjen indikerer en økning i mager masse. Som vi kan se, ligger de fleste punktene over denne linjen, noe som tyder på en generell økning i mager masse for begge treningsvolum. Regresjonslinjen for tre sett ligger generelt høyere enn for ett sett, noe som antyder en større økning i mager masse ved høyere treningsvolum.

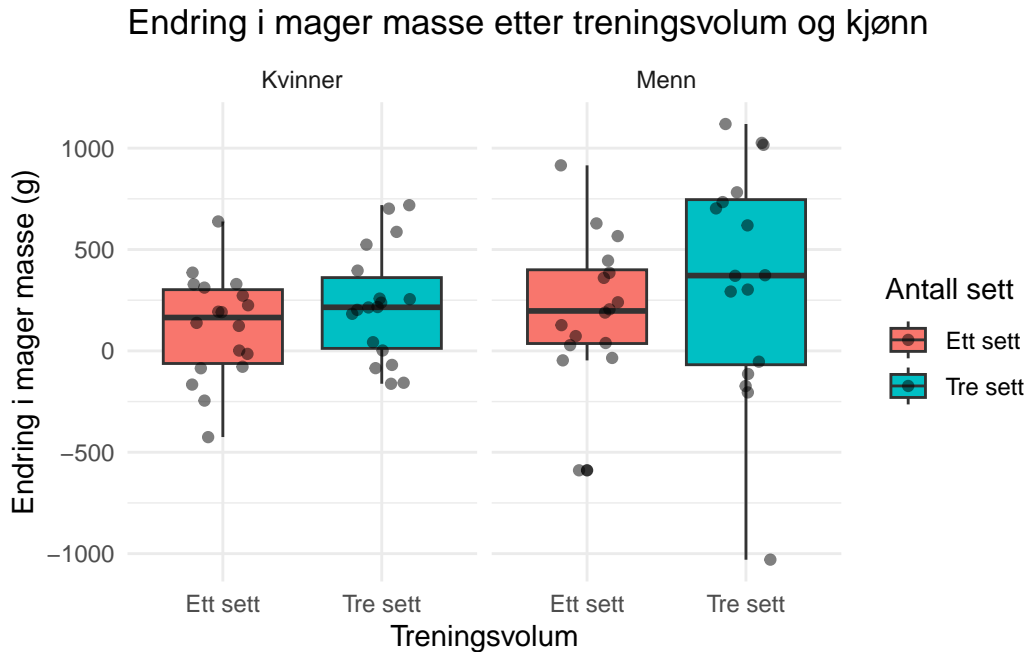


Figure 5.2 viser fordelingen av endringer i mager kroppsmasse for ett sett og tre sett, oppdelt etter kjønn. Boxplottet gir en visuell representasjon av median, kvartiler og spredning i dataene. Som vi kan se, er det en tendens til større økning i mager masse ved tre sett sammenlignet med ett sett, spesielt hos menn. Vi observerer betydelig individuell variasjon hos begge kjønn, noe som vises gjennom spredningen av datapunktene. Til tross for denne variasjonen, ser vi en generell trend mot større økning i mager masse ved tre sett hos både kvinner og menn.

### 5.3.1 t-test

For å kvantifisere disse observasjonene ble en paret t-test utført. Resultatene viste en statistisk signifikant forskjell i endring av mager kroppsmasse mellom settbetingelsene (122.8 (95% KI: [8.6, 237.0], P-verdi = 0.036,  $t_{33} = 2.19$ )). Dette indikerer at tre sett i gjennomsnitt ga 122.8 gram større økning i mager kroppsmasse sammenlignet med ett sett i underekstremitetene.

Et 95% konfidensintervall på [8.6, 237.0] gram for forskjellen mellom ett og tre sett betyr følgende: Hvis vi gjentok denne studien mange ganger med nye utvalg fra samme populasjon,

og beregnet et konfidensintervall hver gang, ville omtrent 95% av disse intervallene inneholde den sanne gjennomsnittlige forskjellen i populasjonen. Det er viktig å merke seg at dette ikke betyr at det er 95% sannsynlighet for at den sanne verdien ligger innenfor dette spesifikke intervallet, eller at alle verdier innenfor intervallet er like sannsynlige. I virkeligheten er verdier nærmere midten av intervallet mer sannsynlige enn verdier nær ytterkantene. Siden intervallet ikke inkluderer 0, er dette konsistent med en statistisk signifikant forskjell mellom ett og tre sett, men det gir oss ikke direkte informasjon om størrelsen eller praktisk betydning av denne forskjellen

P-verdien på 0.036 er under det vanlige signifikansnivået på 0.05, noe som indikerer statistisk signifikans. Denne p-verdien representerer sannsynligheten for å observere en forskjell like ekstrem som eller mer ekstrem enn den vi fant i vårt utvalg, gitt at nullhypotesen er sann (altså at det ikke er noen reell forskjell mellom ett og tre sett i populasjonen). Det er viktig å merke seg at p-verdien ikke forteller oss sannsynligheten for at nullhypotesen er sann eller falsk, og heller ikke direkte sannsynligheten for at resultatet skyldes tilfeldigheter. Den gir oss imidlertid grunnlag for å vurdere styrken av bevisene mot nullhypotesen i lys av våre data og statistiske modell

T-statistikken ( $t(33) = 2.19$ ) representerer forholdet mellom den observerte forskjellen og dens standardfeil. Frihetsgraden 33 er basert på antall par i studien (34) minus 1.

## 5.4 Resultat

Analysen av dataene fra Hammarström *et al.* (2020) avdekket flere interessante funn angående effekten av treningsvolum på muskelhypertrofi og styrke i underekstremitetene.

### 5.4.1 Endringer i mager kroppsmasse

En paret t-test ble utført for å sammenligne endringene i mager kroppsmasse mellom ett sett og tre sett treningsvolum. Resultatene viste en statistisk signifikant forskjell ( $r_{lean.mass.resultater}$ ). Dette indikerer at tre sett i gjennomsnitt ga et estimert gram større økning i mager kroppsmasse sammenlignet med ett sett i underekstremitetene over en 12-ukers treningsperiode.

Figure 5.1 illustrerer de individuelle endringene i mager kroppsmasse for hver deltaker. Plottet avslører betydelig variasjon i individuell respons på treningsvolumet. Mens de fleste deltakere viste større økning ved tre sett, var det noen som hadde liten forskjell eller til og med større økning ved ett sett.

Figure 5.2 viser forholdet mellom pre- og post-intervensjonsverdier for mager masse. De fleste punktene ligger over den stiplede linjen, noe som indikerer en generell økning i mager masse for begge treningsvolum. Regresjonslinjen for tre sett ligger generelt høyere enn for ett sett,

hvilket støtter funnet om større økning i mager masse ved høyere treningsvolum. **?@fig-boxplot** presenterer fordelingen av endringer i mager kroppsmasse etter treningsvolum og kjønn. Det er en tydelig tendens til større økning i mager masse ved tre sett sammenlignet med ett sett, særlig hos menn. Betydelig individuell variasjon observeres hos begge kjønn, men den generelle trenden viser større økning i mager masse ved tre sett for både kvinner og menn.

### 5.4.2 Kjønnforskjeller

Selv om studien ikke var designet for å teste kjønnforskjeller spesifikt, viser **?@fig-boxplot** at menn generelt hadde en større absolutt økning i mager kroppsmasse sammenlignet med kvinner, uavhengig av treningsvolum. Dette er ikke et ukjent fenomen og skyldes nok fysiologiske forskjeller mellom kjønnene. Resultatene viser imidlertid at både menn og kvinner har en tendens til å øke muskelmasse ved 3 sett kontra 1 sett. Men det burde understrekes at dette ikke nødvendigvis er sant for alle individer uavhengig av kjønn.

## 5.5 Diskusjon

Denne studien undersøkte effekten av treningsvolum på muskelhypertrofi i underekstremitetene hos moderat trente individer over en 12-ukers treningsperiode. Resultatene viser en statistisk signifikant forskjell i økning av mager kroppsmasse mellom ett og tre sett, med en gjennomsnittlig større økning på 122,7 gram for tre sett sammenlignet med ett sett.

### 5.5.1 Hovedfunn

Det primære funnet i denne studien er at høyere treningsvolum (tre sett) resulterte i større hypertrofi sammenlignet med lavere volum (ett sett). Dette støtter det grunnleggende prinsippet om overbelastning i styrketrening, som tilsier at økt treningsvolum kan føre til større muskulære tilpasninger (Wackerhage, 2014). Resultatene er i tråd med tidligere forskning som har vist at økt treningsvolum kan føre til større muskelvekst (Raastad *et al.*, 2010). Imidlertid er det viktig å merke seg den betydelige individuelle variasjonen i respons, som illustrert i Figure 5.1. Denne variasjonen understreker kompleksiteten i muskulære tilpasninger og støtter Wackerhage (2014) sin påstand om at det ikke finnes et universelt optimalt motstandstreningsprogram.

### 5.5.2 Kjønnforskjeller

Selv om studien ikke primært var designet for å undersøke kjønnforskjeller, viser **?@fig-boxplot** at menn generelt hadde en større absolutt økning i mager kroppsmasse sammenlignet

med kvinner, uavhengig av treningsvolum. Dette kan skyldes fysiologiske forskjeller mellom kjønnene, som hormonelle profiler og utgangspunkt for muskelmasse. Det er imidlertid viktig å merke seg at både menn og kvinner viste en tendens til større økning i muskelmasse ved tre sett sammenlignet med ett sett.

### **5.5.3 praktiske implikasjoner**

Funnene fra denne studien har flere praktiske implikasjoner for treningsprogrammering og peker også på viktige områder for fremtidig forskning. For individer som søker maksimal muskelvekst, kan det være fordelaktig å inkludere høyere treningsvolum, for eksempel tre sett per øvelse. Den observerte individuelle variasjonen understreker imidlertid behovet for personlig tilpasning av treningsprogrammer. Både menn og kvinner kan forvente fordeler av økt treningsvolum, selv om de absolutte endringene kan variere

### **5.5.4 fremtidig forskning**

For å bygge videre på disse funnene, hadde jeg likt å se studier undersøke interaksjonen mellom treningsvolum og andre variabler som intensitet. Jeg skulle likt å sett for eksempel hva forskjellen mellom ett sett med høyere vekt eller repetisjoner er fra tre sett med lavere vekt eller repetisjoner.

Et annet viktig forskningsområde er å undersøke mekanismene bak individuelle forskjeller i respons på treningsvolum, noe som kan bidra til mer presise og personaliserte treningsanbefalinger. Hvorfor responderer noen dårligere og noen motsatt av trenden? Vi kan gi en generell anbefaling, men det ser ut som det er noen som vil få dårligere respons av å gjøre det som de vil fungere for majoriteten.

Til slutt vil jeg nevne at fremtidige studier kan med fordel fokusere på spesifikke kjønnsforskjeller i respons på ulike treningsvolum. Selv om det ser ut som at i dette tilfellet er en lik trend for begge kjønn selv med en høyere forskjell hos menn, ser vi at det er forskjeller i responsen.

## **5.6 Konklusjon**

Denne studien gir evidens for at høyere treningsvolum (tre sett) fører til større økning i mager kroppsmasse sammenlignet med lavere volum (ett sett) over en 12-ukers periode hos moderat trente individer. Resultatene støtter bruken av høyere treningsvolum for å maksimere muskelhypertrofi, men understreker også viktigheten av individuell tilpasning i treningsprogrammer. Funnene bidrar til å fylle kunnskapshullet rundt optimal treningsprogrammering for muskelvekst og styrker vår forståelse av dose-respons-forholdet i styrketrening.

## 5.7 Statistisk programvare

All databehandling, statistisk analyse og visualisering ble utført i Rstudio (2024.09.1+394 “Cranberry Hibiscus”). Jeg brukte flere R-pakker. Tidyverse ble benyttet for datamanipulering og visualisering, spesielt pakkene dplyr for databehandling og ggplot2 for å lage figurene. GT-pakken ble brukt for å lage den formaterte tabeller. Exscidata-pakken ble brukt for å få tilgang til datasettene fra Hammarström *et al.* (2020). Alle analyser og figurer ble integrert i en reproduserbar rapport ved hjelp av Quarto.

## 5.8 Etiske hensyn:

Studien fulgte etiske retningslinjer og var forhåndsregistrert (ClinicalTrials.gov Identifier: NCT02179307). Den ble godkjent av den lokale etiske komiteen ved Høgskolen i Lillehammer, Avdeling for idrettsvitenskap (nr. 2013-11-22:2). Alle prosedyrer ble utført i samsvar med Helsinkideklarasjonen. Deltakerne ble informert om potensielle risikoer og ubehag forbundet med studien og ga sitt informerte samtykke før de ble inkludert (Hammarström *et al.*, 2020).

## 6 Oppgave: Filosofihistorie

### 6.1 Introduksjon

Dette er et arbeidskrav om “Filosofihistorie” i faget IDR4000-1 24H Kvantitativ metode og statistikk. Dette var en individuell oppgave jeg har gjort små forandringer med og utvidet litt basert på tilbakemelding fra læreren. Vi skulle velge oss ut 2 av 3 spørsmål. Jeg valgte det første og tredje spørsmålet som handlet om Induksjon og Bayesianisme.

### 6.2 Spørsmål 1:

Ifølge Hume er det umulig å rasjonelt begrunne bruken av induksjon. Hva er argumentet for denne konklusjonen? Gi en innvending mot ett av premissene i Humes argument og prøv å svare på denne innvendingen på Humes vegne.

#### 6.2.1 Hva er induksjon, og Hume sitt argument.

Induktiv resonnering går fra det spesifikke til det generelle. En observerer mønstre eller trender i et begrenset antall tilfeller og bruker disse observasjonene til å formulere bredere generaliseringer eller teorier (Okasha, 2016, s. 17). Eksemplet Okasha (2016), bruker er: de første fem eggene i en kartong er gode. Alle eggene i kartongen har samme best-før dato stemplet på seg. Derfor er det sjette egget også god. Selv om de første fem eggene er gode, og alle eggene har samme dato stempel, er det fortsatt fullt mulig at det sjette egget er råttent. Det er logisk mulig at premisset til inferensen er sant, og fortsatt at konklusjonen er falsk. Så inferensen er ikke deduktiv. Deduktiv inferens sier at hvis premisset er sant så må også konklusjonen være sann, som ved eksemplet: alle franskmenn liker rødvin, Pierre er en franskmann derfor liker Pierre rødvin (Okasha, 2016, s. 16).

Ifølge Hume er det umulig å rasjonelt begrunne bruken av induksjon. Argumentet han bruker er at vi forutsetter eller antar «naturens uniformitet» (uniformity of nature). Vi gjør en antagelse av at eggene vi har utforsket er like som de vi ikke har utforsket (Okasha, 2016, s. 20-21). Hume spør hvordan vi kan bevise at en slik uniformitet er sann. Vi kan forestille oss at verden ikke er uniform, og forandrer seg fra dag til dag. I en slik verden kan neste slurk med vann fra springen forgifte oss selv om den aldri før har gjort det. Siden vi kan forestille



oss en slik verden, så kan vi ikke bevise at den uniformelle antagelsen er sann. For hvis vi kunne ville en ikke-uniformell verden være en logisk umulighet (Okasha, 2016, s. 20).

En annen del av argumentet til Hume er at vi heller ikke kan bevise uniformitetsantagelsen induktivt. Hvis vi prøver å rettferdiggjøre induksjon ved å si at den har fungert tidligere, bruker vi nettopp den type *resonnement* vi forsøker å bevise. Dette skaper en sirkulær argumentasjon som ikke gir en gyldig begrunnelse for induksjon (Okasha, 2016, s. 21-22).

### **6.2.2 Innvending mot ett av premissene til Hume og motargument.**

Hume hevder at induksjon forutsetter en tro på “naturens uniformitet”, men dette premisset kan utfordres. Vi kan argumentere for at effektiv induksjon ikke nødvendigvis krever en absolutt tro på naturens uniformitet, men snarere en mer begrenset antagelse om regelmessighet innenfor spesifikke domener eller under visse forhold. For eksempel, i tilfellet med eggene, trenger vi ikke anta at hele naturen er uniform, bare at produksjons- og lagringsforholdene for eggene i én bestemt kartong sannsynligvis er konsistente. Vi trenger ikke at hele naturen skal være uniformell, men innenfor et lite begrenset område av naturen kan vi kontrollere variable nokk til at det vil være rasjonelt å forvente at det neste egget ikke vil være råttent. Denne mer begrensede antagelsen er mindre problematisk enn den universelle påstanden om naturens uniformitet som Hume kritiserer.

Jeg tror Hume svarer at selv om vi begrenser vår antagelse om uniformitet til spesifikke domener eller forhold, står vi fortsatt overfor det samme grunnleggende problemet. Hvordan kan vi rettferdiggjøre selv denne begrensede antagelsen om regelmessighet? Enhver slutning om fremtiden basert på fortiden er i seg selv en induktiv slutning, og dermed sirkulær hvis den brukes til å rettferdiggjøre induksjon (Okasha, 2016, s. 21). Enhver begrunnelse vi gir for å tro at produksjons- og lagringsforholdene for eggene i én kartong er konsistente, vil til syvende og sist hvile på tidligere erfaringer og observasjoner. Men å bruke disse tidligere erfaringene til å rettferdiggjøre vår tro på fremtidig regelmessighet er nettopp den typen induktiv slutning som Hume utfordrer. For hvis du hadde forsøkt å forklare induksjon til noen som ikke stoler på induksjon fra før av ville et slikt argument ikke fungere. For å argumentere at en kan stole på induksjon fordi den har fungert helt til nå er i seg selv å rasjonalisere induktivt (Okasha, 2016, s. 21-22). Vi har ingen garanti for at de spesifikke forholdene vi observerer i dag vil forbli konstante i fremtiden, uansett hvor begrenset domenet er.

### **6.3 Spørsmål 3:**

Forklar de grunnleggende ideene i Bayesianisme, samt hvordan Bayesianske sannsynligheter kan tolkes. Presenter så ett problem med Bayesianisme og evaluer hvor alvorlig problemet er.

### 6.3.1 De grunnleggende ideene i Bayesianisme, og hvordan Bayesianske sannsynlighet kan tolkes.

Bayesianisme er en tilnærming til å forsøke å forstå bevis og bekreftelse i vitenskapen ved hjelp av matematisk sannsynlighetsteori. Den sentrale ideen er at observasjoner kan øke eller redusere sannsynligheten for en hypotese (Godfrey-Smith, 2003, s. 203). Bayesianere bruker Bayes teorem som en formel for å beregne hvordan sannsynligheten for en hypotese endres i lys av nye bevis. Formelen i sin enkleste form kan skrives som:  $P(h|e) = P(e|h)P(h) / P(e)$ . Der  $h$  står for hypotese,  $e$  står for evidens  $P(h|e)$  er sannsynligheten for hypotesen gitt evidensen (posterior),  $P(e|h)$  er sannsynligheten for evidensen gitt hypotesen (likelihood),  $P(h)$  er den tidligere sannsynligheten (prior) for hypotesen og  $P(e)$  er sannsynligheten for evidensen (evidenssannsynlighet) (Godfrey-Smith, 2003, s. 203-205).

De fleste forsøk på å analysere sannsynlighet har forsøkt å måle noen reelle og objektive trekk ved hendelser. Hvor sannsynlighetsverdier måler kjangsen for at noe skal skje (Godfrey-Smith, 2003, s. 205). Godfrey-Smith (2003) fortsetter med å si at majoriteten av filosoferene som ønsker å bruke Bayers teorem for å forstå evidens ser på sannsynlighet som noe subjektivt. «Bayesians claim to give a theory of when a person's total network of degrees of belief is "coherent," or rational. They argue that a coherent set of degrees of belief has to follow the standard rules of the mathematics of probability.» (Godfrey-Smith, 2003, s. 206). Bayesianere hevder å tilby en teori som beskriver når en persons totale nettverk av tro er sammenhengende eller rasjonelle. De argumenterer for at de sammenhengende grader av tro må følge standard-reglene i sannsynlighetsmatematikk. Dette innebærer at troene må være logisk konsistente og justeres i samsvar med nye bevis, slik at de danner et helhetlig system. Dette handler om den subjektive tolkningen, hvor vi rasjonaliserer våre personlige oppfatninger, og er en måte Bayesianske sannsynlighet kan tolkes (Godfrey-Smith, 2003, s. 206). Noen bayesianere ønsker å inkludere en objektiv forståelse av sannsynlighet, men dette krever mer omfattende argumenter (Godfrey-Smith, 2003, s. 208). Siden Godfrey-Smith (2003) ikke utdyper dette mere forholder jeg meg til den subjektive forståelsen.

Regelen for betinget oppdatering er sentral i bayesiansk epistemologi og beskriver hvordan man bør justere sine oppfatninger når ny informasjon mottas (Okasha, 2016, s. 32-33). Bayes' teorem er nøkkelen til å forstå forholdet mellom sannsynligheten for en hypotese før og etter ny evidens presenteres. En rasjonell person vil, ifølge bayesianismen, oppdatere sin tro basert på denne nye informasjonen, slik at den reflekterer den oppdaterte sannsynligheten (Godfrey-Smith, 2003, s. 208). Dette betyr at dagens oppdaterte oppfatning blir grunnlaget for morgendagens vurderinger. Bayesianere hevder at denne kontinuerlige oppdateringsprosessen er essensiell for rasjonell tenkning. Det er viktig å merke seg at i den subjektive tolkningen av sannsynlighet, anses ingen oppfatning som mer korrekt enn en annen, så lenge de følger grunnleggende sannsynlighetsregler (Godfrey-Smith, 2003, s. 208).

### 6.3.2 Et problem og min evaluering av problemet med Bayesianisme.

Et problem ved Bayesiansk teori er at den ikke kan kritisere merkelig starter eller initiale sannsynlighetsfordelinger (Godfrey-Smith, 2003, s. 210). Til tross for dette hevder Bayesianere at selv om priorene kan være rare i starten, vil de “vaskes ut” av ny evidens, så lenge oppdateringen gjøres rasjonelt (Godfrey-Smith, 2003, s. 210). Startpunktet blir mindre viktig etter hvert som mer evidens tas i betraktning. Dette fenomenet kalles konvergens (convergence). Selv om to personer starter med svært ulike priorer, vil deres sannsynligheter for en hypotese bli nærmere hverandre etter hvert som de mottar samme evidens. Det kan bevises at for enhver mengde initial (utgangspunkt) uenighet er det mulig å finne nok evidens til å bringe dem nærmere hverandre. Likevel kan denne konvergens ta lang tid, og det er ikke alltid mulig å få alle til å være enige om sannsynlighetene og disse bevisene har mange antagelser som ikke alltid stemmer i praksis (Godfrey-Smith, 2003, s. 208). Noen Bayesianere har derfor forsøkt å begrense hvordan den initiale sannsynligheter tildeles. Et problem med konvergensargumentet er at det antar at folk er enige om sannsynlighetene for evidensen gitt hypotesene (likelihoods). Men det er ikke alltid rimelig å anta at folk vil ha samme likelihoods når de er uenige om mange andre ting (Godfrey-Smith, 2003, s. 208).

Selv om dette er et problem, virker det som at det er et problem som kan løses, gitt nok tid og ressurser, så lenge man følger de rasjonelle reglene og er logisk konsistente ifølge Godfrey-Smith (2003). Selv om jeg tror at i praksis vil dette være vanskelig å utføre fullt ut. Et eksempel på hvordan dette problemet kan løses innenfor bayesiansk rammeverk, gitt nok tid og ressurser, er gjennom bruken av ikke-informative priorer. I tilfeller hvor vi ikke har tidligere informasjon å basere våre initiale sannsynligheter på, kan vi bruke såkalte “flate” eller “uniforme” priorer. Dette innebærer å gi alle mulige hypoteser like sannsynligheter i utgangspunktet. For eksempel, la oss si at vi introduserer en helt ny teori om universets opprinnelse. Vi har ingen tidligere data å basere våre initiale sannsynligheter på. I stedet for å la dette stoppe oss, kan vi starte med en uniform prior over alle mulige forklaringer. Etter hvert som vi samler inn data og observasjoner, vil vi oppdatere disse sannsynlighetene i henhold til Bayes’ teorem. Over tid, gitt tilstrekkelig med data, vil sannsynlighetene konvergere mot de “sanne” verdiene, uavhengig av vårt initiale utgangspunkt.

Hva med når teorier ikke har en prior å gå ut ifra da teorien eller hypotesen ikke tidligere har blitt tenkt på av forgjengeren? Som er tilfelle med interessante teorier vi har fått av både Copernicus, Newton og Darwin. Dette går imot regelen for betinget oppdatering da den må ha «gammel» informasjon den kan oppdatere. Den vitenskapelige sluttingen som går fra data til en helt ny teori kan ikke bli forstått med bruk av betinget oppdatering (Okasha, 2016, s. 34). Det virker som for meg at dette problemet kan løses ved at enhver teori eller hypotese, uansett hvor bisarr eller rar den er, er gyldig (Godfrey-Smith, 2003, s. 208; Okasha, 2016, s. 35). Så ved tilfeller hvor en ikke har en prior å gå ut ifra kan en nesten «finne opp» (da en tro ikke er mer riktig enn en annen) en prior så lenge de rasjonelle oppdateringene følges etter det (Godfrey-Smith, 2003, s. 210). Dette viser imidlertid at Bayesian synspunkter ikke kan være den hele forklaringen på den vitenskapelige slutningen og forklarer ikke Humes problem med

induksjon. Altså hvordan vi kan begrunne at fremtidige hendelser vil følge samme mønster som tidligere observerte hendelser (Okasha, 2016, s. 35).

# 7 Oppgave: Protein Analyse

## 7.1 Introduksjon

Dette er ett av tre arbeidskrav fra biologilabben i faget IDR4000-1 24H Kvantitativ metode og statistikk. Vi skulle plukke ut en oppgave som skulle presenteres i mappeeksamen, jeg valgte “protein analyse” Dette var en gruppeoppgave jeg har rekonstruert fra den originale elab rapporten. Innholdet er det samme, men den har blitt omstrukturert, utvidet og presentert på en mer ryddig måte.

## 7.2 Innledning

Proteinanalyse av muskelprøver ved hjelp av Western blot er en viktig metode for å undersøke molekulære endringer i skjelettmuskulatur som respons på trening. Dette forsøket fokuserer på å analysere uttrykket og fosforyleringen av proteinet p70 S6 kinase (p70S6K) fra muskelbiopsier tatt fra venstre og høyre vastus lateralis hos en forsøksperson. Høyre bein gjennomgikk en moderat til hard treningsøkt med fokus på kneekstensjon en time før prøvetaking, mens venstre bein forble utrent. Forsøksdeltageren har trent styrke systematisk i flere år.

p70S6K1 er et av de best karakteriserte nedstrømsmålene for mTORC1, en sentral regulator av cellevekst og proteinsyntese. Når mTORC1 aktiveres, fosforylerer den p70S6K1, noe som bidrar til å stimulere proteinsyntese, og som en konsekvens muskelhypertrofi (Neumann & Violette, 2019). Ved å sammenligne fosforyleringsnivåer av p70S6K mellom trent og utrent muskel, søker denne oppgaven å belyse de akutte molekulære responsene på en enkelt treningsøkt i menneskelig skjelettmuskulatur. Gjennom bruk av Western blot-teknikk, vil denne analysen gi innsikt i hvordan akutt trening påvirker signalveier knyttet til muskelvekst og adaptasjon.

## 7.3 Metode

### 7.3.1 Prøvepreparering

Muskelbiopsier ble tatt fra venstre (TSL) og høyre (TSR) vastus lateralis hos forsøkspersonen TS. Høyre bein hadde gjennomgått en moderat til hard treningsøkt med fokus på kneekstensjon

en time før prøvetaking, mens venstre bein forble utrent. Vekten på ferskprøvene ble målt og registrert som følger:

TSL1: 22,4 mg

TSL2: 20,5 mg

TSR1: 21,4 mg

TSR2: 13,1 mg

### 7.3.2 Protokollene

Alle protokoller fikk vi utdelt gjennom canvas.

#### 7.3.2.1 General purpose muscle tissue homogenization protocol

Prøvene ble homogenisert i en lysisbuffer i henhold til “General purpose muscle tissue homogenization protocol” (versjon 2019-11-12).

Homogeniseringsprotokoll

Prøvene ble preparert ved å dissekere bort bindevev og blod. Protease/fosfatase-hemmere ble tilsatt den iskalde lysisbufferen (10 l/ml). Prøvene ble holdt på tørris eller kjøleblokk inntil tilsetning av iskald lysisbuffer. Vevsprøvene ble raskt homogenisert for hånd ved hjelp av en plastpistill til det ikke var synlige biter igjen. Prøvene ble holdt på is og rotert i 20 minutter. Prøvene ble sentrifugert ved 10 000 g i 10 minutter ved 4°C. Supernatanten ble forsiktig overført til et nytt rør uten å forstyrre pelleten. Prøven ble alikvotet for proteinkonsentrasjonsbestemmelse (1:10 fortykning, 4 l til 36 l ddH<sub>2</sub>O).

#### 7.3.2.2 Proteinkonsentrasjonsbestemmelse ved Bradford-assay

Bradford-assay ble brukt for å måle proteinkonsentrasjonen. To separate tester ble utført. Vi la opp brettet for test 1 slik.

Table 7.1: Bradford Assay Plate Oppsett

	1	2	3	4	5	6	7	8	9	10	11	12
A	Std a	Std a	Std a	Std I	Std I	Std I						
B	Std b	Std b	Std b	TSL 1	TSL 1							
C	Std c	Std c	Std c	TSL 2	TSL 2							
D	Std d	Std d	Std d	TSR 1	TSR 1							
E	Std e	Std e	Std e	TSR 2	TSR 2							
F	Std f	Std f	Std f									
G	Std g	Std g	Std g									

	1	2	3	4	5	6	7	8	9	10	11	12
H	Std h	Std h	Std h									

(Når vi testet konsentrasjonen senere i test 2 laget vi nok slik at vi fikk triplikat av prøvene, lagt i B6, C6, D6 og E6)

### 7.3.2.3 Western Blot

Daniel Hammarström Determine protein abundance by immunoblotting System: Bio-Rad Criterion

#### Prøvepreparering

Prøvene ble fortynnet til en sluttkonsentrasjon på 0,24 g/l i Laemmli-buffer. Følgende volumer ble brukt for hver prøve:

TSR1 (original konsentrasjon: 0,55 g/l): 87,3 l prøve, 62,7 l dH<sub>2</sub>O, 50 l samplebuffer

TSR2 (original konsentrasjon: 0,32 g/l): 150,0 l prøve, 0,0 l dH<sub>2</sub>O, 50 l samplebuffer

TSL1 (original konsentrasjon: 0,33 g/l): 145,5 l prøve, 4,5 l dH<sub>2</sub>O, 50 l samplebuffer

TSL2 (original konsentrasjon: 0,41 g/l): 117,1 l prøve, 32,9 l dH<sub>2</sub>O, 50 l samplebuffer

Prøvene ble lastet opp i følgende rekkefølge

Table 7.2: Oppsett av prøver for Western blot

1	2	3	4	5	6	7	8	9	10
Stige	LØ R	TS R	TS L	Stiger	Stige	LØ R	TS R	TS L	Stige

LØ R er en prøve fra en annen gruppe som ble kjørt samtidig med vår. Det ble observert en liten luftboble i brønn 3 og 5, som ble forsøkt fjernet ved hjelp av en pasteurpipette.

Prøvene ble varmet opp til 70°C i 10 minutter før de ble lastet på gelen (ikke 95°C i 5 min som beskrevet i protokollen).

#### Gelelektroforese

Elektroforese ble utført ved konstant spenning på 300 volt. Prosessen ble stanset etter 23 minutter (isteden for 30 min som beskrevet i protokollen) på anbefaling fra læreren.

#### Protein transfer (blotting)

Vi forberedet en “sandwich” med membranene ved følgende prosedyre: La svampene i dH<sub>2</sub>O og fjernet bobler. La membranene (klippet i øvre venstre hjørne for å markere side) i metanol for å aktivere (ristet i 5-10 minutter) Plasserte sandwich i monteringsbrettet – svart side

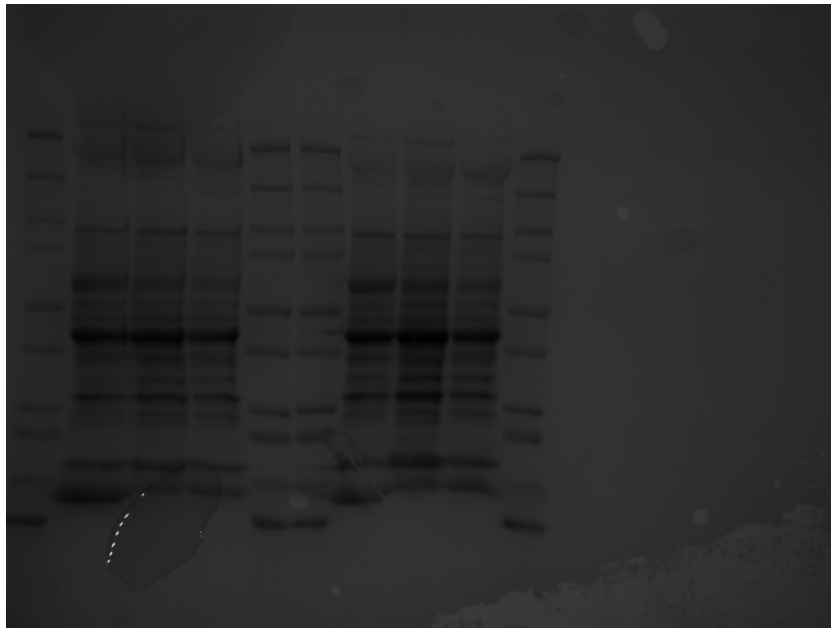
ned. Klemte ut dH<sub>2</sub>O fra svampene og la dem i transferbuffer – fjernet alle bobler. Fuktet 2 filterpapir i buffer, la dem på toppen av gelen og klemte dem forsiktig sammen. Plasserte filterpapirene med gelen på toppen av svampen. Fjernet bobler. La membranen på toppen av gelen. sørget for at det ble lagt i riktig retning, membranene var merket i øvre venstre hjørne. Fjernet bobler. Plasserte den siste svampen på toppen og lukket sandwichen.

Blotting ble gjennomført ved 100 V og 500 mA i 30 minutter (isteden for 300 mA over 3 timer som beskrevet i protokollen).

### **Membranbehandling**

Membranen ble skylt raskt i dH<sub>2</sub>O. MemCode Sensitizer ble tilsatt og plassert på shaker i 2 minutter. MemCode Reversible stain ble tilsatt og plassert på shaker i 1 minutt. Membranen ble avfarget med MemCode destain-løsning. Metanol/destain-løsning ble tilsatt og plassert på rister i 5 minutter. Membranen ble skylt med dH<sub>2</sub>O 4 ganger. Membranen ble avbildet. Membranen ble deretter behandlet med Eraser-løsning på shaker i 10 minutter. Membranen ble skylt i dH<sub>2</sub>O 4 ganger og oppbevart i TBS inntil kutting. Membranen ble kuttet i 2 deler(en del til hver gruppe).

Bilde av membranen



### **Blokkering og antistoffinkubering**

Membranen ble blokkert med 5% melkeløsning i TBS i 1 time ved romtemperatur (ikke 2.5% melkeløsning som protokollen sa). Primære antistoffer ble inkubert over natten i 1% melkeløsning ved 4°C (ikke 5% melkeløsning som protokollen sa).



Lot-nummeret for antistoffene er inkludert for å sikre sporbarhet og reproducerbarhet av resultatene. I tilfelle uventede resultater eller problemer med antistoffene, kan lot-nummeret også brukes til å kontakte produsenten for ytterligere informasjon eller feilsøking:

10 g P-p70 s6 kinase (lot 22, produktnr 9206s) 10 g P-70 s6 kinase (lot 20, produktnr 9202s)

*Dette var så langt jeg var med i protokollen. Læreren ordnet resten uten oss. Dette er resten av protokollen.*

Membranen ble vasket med TBS-T 2x1 minutt + 3x5 minutter. Sekundære antistoffer ble inkubert i 2,5% melk med TBS-T i 1 time ved romtemperatur. Membranen ble vasket med TBS 4x5 minutter.

### **Deteksjon**

ECL (enhanced chemiluminescence) ble brukt for å visualisere proteinbåndene: Arbeidssløsning ble blandet (Thermo super signal). Membranen ble inkubert i 5 minutter ved romtemperatur. Membranen ble dekket med klar plastfolie. Bilder ble tatt med Gel Doc-systemet.

### **Stripping og reinkubering**

Membranen ble vasket 7 ganger med dH<sub>2</sub>O. Membranen ble vasket 3x3 minutter i TBS. Membranen ble strippet med Restore Plus Western Blot Stripping Buffer i 30 minutter ved 37°C. Membranen ble vasket 5x1 minutt i dH<sub>2</sub>O. Membranen ble vasket 3x3 minutter i TBS.

### **Avsluttende kommentar**

Alle trinn i prosedyren ble utført i henhold til standard laboratorieprotokoller, med unntak av de spesifikt nevnte modifikasjonene. Eventuelle avvik fra standardprotokollene ble notert og tatt hensyn til i den påfølgende dataanalysen og tolkningen av resultatene.

## **7.4 Resultater og diskusjon**

### **7.4.1 Proteinkonsentrasjonsmålinger**

Bradford-assay ble brukt for å måle proteinkonsentrasjonen i prøvene. To separate tester ble utført, og resultatene var som følger:

#### **Test 1 (brukt for Western blot):**

TSL1: 329,3903 g/l (SD = 26,9992, CV = 8,1967%)

TSL2: 405,5337 g/l (SD = 2,2046, CV = 0,5436%)

TSR1: 554,6679 g/l (SD = 2,3571, CV = 0,4250%)

TSR2: 323,2049 g/l (SD = 6,4506, CV = 1,9958%)

#### **Test 2 (ekstra test):**

TSL1: 576,3638 g/l (SD = 12,0474, CV = 2,09%)  
TSL2: 603,6164 g/l (SD = 76,1227, CV = 12,61%)  
TSR1: 587,4225 g/l (SD = 56,0197, CV = 9,53%)  
TSR2: 352,2895 g/l (SD = 28,4337, CV = 8,07%)

Som vi ser, er det betydelig variasjon mellom prøvene våre, noe som kan skyldes flere faktorer i vår eksperimentelle prosedyre. Ujevn blanding eller fortynning av prøvene under forberedelsen, pipetteringsfeil og dårlig blanding av reagenser er alle potensielle kilder til disse avvikene. I tillegg kan forskjellen i antall replikater mellom Test 1 (to prøver av hver) og Test 2 (tre prøver av hver) ha påvirket resultatene våre. Flere replikater i Test 2 gir generelt en mer nøyaktig utregning og et bedre bilde av variasjonen mellom prøvene, noe som reflekteres i de høyere standardavvikene (SD) og variasjonskoeffisientene (CV) vi observerer. Samtidig øker flere prøver risikoen for feil i prosessen, da hvert ekstra replikat innebærer ytterligere pipettering, fortynning og håndtering.

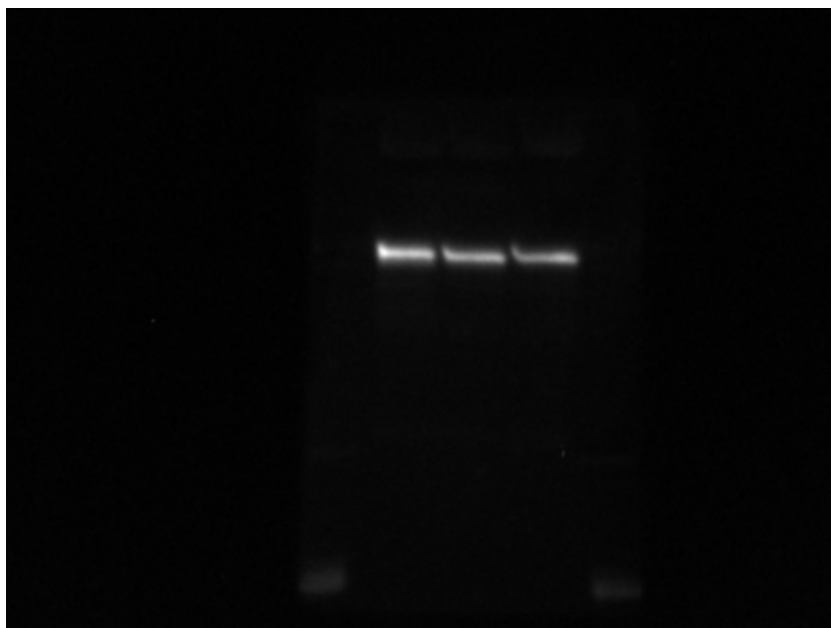
Det er verdt å merke seg at selv med flere replikater i Test 2, ser vi høyere CV-verdier, noe som tyder på større variabilitet i målingene. Dette kan indikere at vi hadde utfordringer med konsistens i prøvehåndteringen, eller at det var mer naturlig variasjon i disse prøvene. Disse observasjonene understreker viktigheten av nøyaktig og konsistent laboratorieteknikk, samt behovet for å optimalisere våre prosedyrer for å redusere variabilitet mellom prøvene. Den eneste grunnen til at vi brukte Test 1 var at vi gikk videre i protokollen, og Test 2 ble utført på et senere tidspunkt når vi hadde litt ekstra tid. De lave CV-verdiene i Test 1 gjorde disse resultatene mest hensiktsmessige å jobbe med, med unntak av "TSL1".

#### **7.4.2 Western blot-resultater**

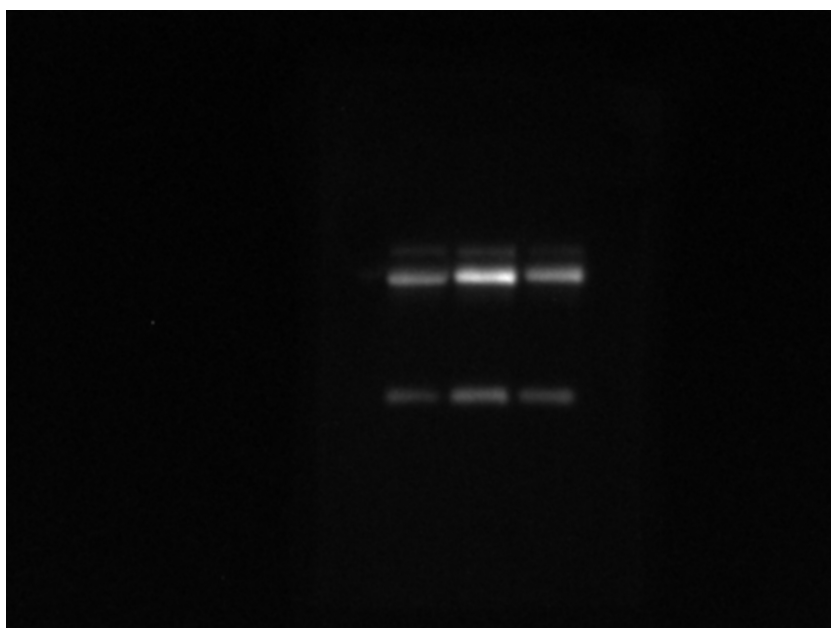
Prøvene var satt opp i rekkefølgen: LØ R - TS R - TS L

Ukjent hva LØ R sin status er men TS R hadde trent før prøvetagning og TS L hadde ikke trent.

**p-p70**



t-p70



Western blot-analysen viste tydelige forskjeller i fosforylering av p70S6K mellom trent og utrent muskel. På p-p70 ser vi at alle tre båndene lyser omtrent like sterkt. Kanskje LØ R lyser hakket sterkere. Dette indikerer at mengden av fosforylert p70 S6 kinase er relativt lik

i alle tre prøvene. Dette kan tyde på at aktivitetsnivået av p70 S6 kinase er sammenlignbart mellom LØ R, TS R (det trente beinet) og TS L (det utrente beinet).

På t-p70 ser vi et mer komplekst mønster. De tre båndene i midten som lyser sterkest, representerer sannsynligvis hovedformen av p70 S6 kinase. De svakere båndene over og under kan være isoformer eller modifiserte versjoner av proteinet. At TSR-prøven (midten) viser sterkere intensitet i alle tre båndene, sammenlignet med de andre prøvene, kan tyde på at det totale uttrykket av p70 S6 kinase er høyere i det trente beinet.

### 7.4.3 Fysiologisk betydning

Økningen av p70S6K i den trente muskelen er i samsvar med tidligere studier som har vist at akutt trening aktiverer mTOR-signalveien, hvor p70S6K er en nøkkelkomponent (Neumann & Viollet, 2019). Fosforylering av p70S6K er kjent for å stimulere proteinsyntese og er en viktig mekanisme i treningsindusert muskelhypertrofi (Neumann & Viollet, 2019). Jeg stusser da over hvorfor vi ikke ser en visuell forskjell på bildet av fosforylert p70S6K.

Tidspunktet for prøvetaking, som var ca. en time etter trening, kan ha påvirket resultatene betydelig. Jacko *et al.* (2022) viser at fosforylering av p70S6K kan vedvare i mer enn 8 timer etter trening. Dette kan forklare hvorfor vi observerte lignende nivåer av fosforylert p70S6K i alle prøvene, inkludert den utrente muskelen. Det er mulig at forskjeller i fosforylering ville vært mer synlige på andre tidspunkter etter trening.

I tillegg, siden forskningsdeltageren (TS) er vant til trening, kan dette ha påvirket responsen. Jacko *et al.* (2022) viser at gjentatt motstandstrening over tid fører til en desensitivisering av mTOR-relatert signalering i skjelettmuskulaturen. Dette kan forklare hvorfor vi ikke så en markant forskjell mellom trent og utrent muskel i vårt forsøk. Artikkelen påpeker også at etter en periode med treningsavbrudd (10 dager) økte fosforyleringen av p70S6K igjen, noe som tyder på en re-sensitivisering av signalveiene (Jacko *et al.*, 2022). Dette understreker viktigheten av å vurdere treningshistorikk og tidspunkt for prøvetaking når man tolker resultater fra slike studier.

Det er verdt å merke seg at vi observerte noen utfordringer i vår Western blot-prosedyre, inkludert luftbobler i noen brønner og variasjon i proteinkonsentrasjonsmålingene. Disse faktorene kan ha påvirket kvaliteten på våre blot og bør tas i betraktning ved tolkning av resultatene. Vår uerfarenhet gjør at vi trenger mere øvelse i utføring av slike protokoller og at feilene mest sannsynlig skyldes menneskelige faktorer og unøyaktighet. Det er ikke usannsynlig at avik har forekommet som vi ikke har registrert har forekommet og dermed ikke har blitt med i rapporten.

## 7.5 Konklusjon

Vårt forsøk viser at en enkelt treningsøkt med kneekstensjon påvirker uttrykket av p70S6K i menneskelig skjelettmuskulatur, med en observert økning i totalt p70S6K-uttrykk i den trente muskelen. Dette samsvarer med tidligere forskning om treningsindusert aktivering av mTOR-signalveien.

Imidlertid fant vi ingen tydelig forskjell i fosforylering av p70S6K mellom trent og utrent muskel, noe som kan skyldes faktorer som prøvetakingstidspunkt og forskningsdeltakerens treningshistorie. Som Jacko *et al.* (2022) påpeker, kan langvarig trening føre til desensitivisering av mTOR-relatert signalering, hvilket kan forklare våre observasjoner.

Våre funn understreker viktigheten av å vurdere flere faktorer i tolkningen av slike forsøk, inkludert prøvetakingstidspunkt, treningshistorie og metodologiske utfordringer. Forsøket bidrar til vår forståelse av akutte molekulære responser på trening, men fremhever også kompleksiteten i disse prosessene.

Jeg erkjenner at min begrensede erfaring med Western blot-teknikken kan ha påvirket resultatenes kvalitet, noe som understreker behovet for grundigere opplæring i avanserte laboratorieteknikker.

## referanser

- Aalen OO, Frigessi A, Moger TA, Scheel I, Skovlund E & Veierød MB (2018). *Statistiske metoder i medisin og helsefag*, 2nd edn. Gyldendal akademisk, Oslo.
- Cho Y, Jang H, Kwon S & Oh H (2023). [Aerobic, muscle-strengthening, and flexibility physical activity and risks of all-cause and cause-specific mortality: A population-based prospective cohort of korean adults](#). *BMC Public Health* **23**, 1148.
- Gillespie BW, Chen Q, Reichert H, Franzblau A, Hedgeman E, Lepkowski J, Adriaens P, Demond A, Luksemburg W & Garabrant DH (2010). [Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator](#). *Epidemiology* **21**, S64–S70.
- Godfrey-Smith P (2003). *Theory and reality: An introduction to the philosophy of science*. University of Chicago Press, Chicago.
- Halperin I, Pyne DB & Martin DT (2015). [Threats to internal validity in exercise science: A review of overlooked confounding variables](#). *International Journal of Sports Physiology and Performance* **10**, 823–829.
- Hammarström D, Øfsteng S, Koll L, Hanestadhaugen M, Hollan I, Apró W, Whist JE, Blomstrand E, Rønnestad BR & Ellefsen S (2020). [Dose-dependent effects of resistance training volume on leg muscle strength and hypertrophy](#). *The Journal of Physiology* **598**, 3077–3088.
- Holtermann A, Schnohr P, Nordestgaard BG & Marott JL (2021). [The physical activity paradox in cardiovascular disease and all-cause mortality: The contemporary copenhagen general population study with 104046 adults](#). *European Heart Journal* **42**, 1499–1511.
- Hopkins WG (2000). [Measures of reliability in sports medicine and science](#). *Sports Medicine* **30**, 1–15.
- Hulley SB, Cummings SR, Browner WS, Grady DG & Newman TB (2013). *Designing clinical research*, 4th edn. Lippincott Williams & Wilkins, Philadelphia, PA.
- Jacko D, Schaaf K, Masur L, Windoffer H, Aussieker T, Schiffer T, Zacher J, Bloch W & Gehlert

- S (2022). [Acute resistance exercise induces different molecular responses in trained and untrained human muscle](#). *International Journal of Molecular Sciences* **23**, 5431.
- Lee C-L, Liu W-J, Chen C-H & Wang J-S (2023). Associations of long-term physical activity trajectories with all-cause mortality in a general population. *International Journal of Public Health*; DOI: [10.3389/ijph.2023.1605332](#).
- Li C, Ma Y, Hua R, Zheng F & Xie W (2022). Long-term physical activity participation trajectories were associated with subsequent cognitive decline, risk of dementia and all-cause mortality among adults aged 50 years: A population-based cohort study. *Age and Ageing*; DOI: [10.1093/ageing/afac071](#).
- Liang YY, Feng H, Chen Y, Jin X, Xue H, Zhou M, Ma H, Ai S, Wing Y-K, Geng Q & Zhang J (2023). [Joint association of physical activity and sleep duration with risk of all-cause and cause-specific mortality: A population-based cohort study using accelerometry](#). *European Journal of Preventive Cardiology* **30**, 832–843.
- Nagin DS (2014). [Group-based trajectory modeling: An overview](#). *Annals of Nutrition and Metabolism* **65**, 205–210.
- Neumann D & Viollet B eds. (2019). *AMP-activated protein kinase signalling*. MDPI, Basel, Switzerland. Available at: <https://www.mdpi.com/books/reprint/1169-amp-activated-protein-kinase-signalling>.
- Okasha S (2016). *Philosophy of science: A very short introduction*, 2nd edn. Oxford University Press, Oxford.
- Raastad T, Paulsen G, Refsnes PE, Rønnestad BR & Wisnes AR (2010). *Styrketrening - i teori og praksis*. Gyldendal undervisning, Oslo.
- Ramos-Jiménez A, Hernández-Torres RP, Torres-Durán PV, Romero-Gonzalez J, Mascher D, Posadas-Romero C & Juárez-Oropeza MA (2008). [The respiratory exchange ratio is associated with fitness indicators both in trained and untrained men: A possible application for people with reduced exercise tolerance](#). *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine* **2**, CCRPM.S449.
- Spiegelhalter D (2020). *The art of statistics: Learning from data*. Penguin Books, London.
- Stensel DJ, Hardman AE & Gill JMR (2021). *Physical activity and health: The evidence explained*. Routledge, London, UK.
- Tanner RK & Gore CJ (2012). *Physiological tests for elite athletes*, 2nd edn. Human Kinetics, Champaign, IL.

Trainome (n.d.). Size determination of PCR products. Available at: [https://trainome.github.io/protocols/sw\\_pcr\\_size\\_determination.html](https://trainome.github.io/protocols/sw_pcr_size_determination.html).

Wackerhage H ed. (2014). *Molecular exercise physiology: An introduction*. Routledge, London.