

# **Mappeeksamen IDR4000**

KANDIDATNR

2024-11-05

# Table of contents

<b>Forord</b>	<b>5</b>
<b>1 Reliabilitet</b>	<b>6</b>
1.1 Introduksjon . . . . .	6
1.2 Metode . . . . .	6
1.2.1 Standardisering før test . . . . .	6
1.2.2 Utstyr . . . . .	7
1.2.3 Protokoll . . . . .	7
1.2.4 Databehandling etter test . . . . .	8
1.3 Resultat . . . . .	8
1.4 Diskusjon . . . . .	10
<b>2 Predikere data</b>	<b>11</b>
2.1 Introduksjon . . . . .	11
2.2 Metode . . . . .	11
2.2.1 Del 1: Predikert laktatterskel . . . . .	11
2.2.2 Del 2: Predikert hellning på qPCR kalibreringskurve . . . . .	11
2.2.3 Del 3: Tolke en regresjonstabell . . . . .	11
2.3 Resultat . . . . .	12
2.3.1 Del 1: Predikert laktatterskel . . . . .	12
2.3.2 Del 2: Predikert hellning på qPCR kalibreringskurve . . . . .	13
2.3.3 Del 3: Tolke en regresjonstabell . . . . .	14
2.4 Diskusjon . . . . .	15
2.4.1 Diskusjon - del 1 . . . . .	15
2.4.2 Diskusjon - del 2 . . . . .	15
2.4.3 Diskusjon - del 3 . . . . .	15
<b>3 Statistisk inferens</b>	<b>16</b>
3.1 Simulering . . . . .	16
3.2 Forklaring av “estimate”, “Standardfeil”, “t-verdi” og “p-verdi” fra m1 og m2 .	16
3.3 Hvorfor forskjellig resultat i m1 og m2? . . . . .	16
3.4 Hvorfor bruker vi det skraverte området i nedre og øvre del av en t-fordeling? .	17
3.5 Repeterte studier . . . . .	17
3.6 Hvorfor er sandardavviket til estimert gjennomsnitt så lik standardfeilen i dette tilfellet? . . . . .	17

3.7	Histogram med p-verdier fra simuleringene, hva sier dette om statistisk styrke?	18
3.8	Filtrer resultater med en gitt p-verdi . . . . .	18
3.9	Test av statistisk styrke . . . . .	19
3.10	Hvor mange studier gir falsk positiv når signifikansnivået er satt til 5%? . . . .	19
<b>4</b>	<b>Studiedesign</b>	<b>25</b>
4.1	Introduksjon . . . . .	25
4.2	Metode . . . . .	25
4.3	Resultat . . . . .	26
4.4	Diskusjon . . . . .	26
4.4.1	Studiedesign og valg av metode . . . . .	26
4.4.2	Statistiske modeller . . . . .	27
4.4.3	Varighet og oppfølging . . . . .	27
4.4.4	Konklusjon . . . . .	27
<b>5</b>	<b>Styrketrening på bein med moderat volum har større effekt på maksimal styrke og muskelmass enn lavt volum</b>	<b>28</b>
5.1	Introduksjon . . . . .	28
5.2	Metode . . . . .	28
5.2.1	Forsøkspersoner og studiedesign . . . . .	28
5.2.2	Dataanalyse og statistikk . . . . .	30
5.3	Resultater . . . . .	30
5.3.1	Muskelmasse . . . . .	30
5.3.2	Maksimal styrke . . . . .	30
5.4	Diskusjon . . . . .	32
<b>6</b>	<b>Filosofi</b>	<b>33</b>
6.0.1	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. . . . .	33
6.0.2	Gi en kort beskrivelse av falsifikasjonisme og si litt om hvorfor Karl Popper var motivert til å utvikle denne teorien. Presenter så ett problem med teorien og vurder hvorvidt problemet kan løses. . . . .	34
<b>7</b>	<b>Labrapport</b>	<b>36</b>
7.1	Introduksjon . . . . .	36
7.2	Metode . . . . .	37
7.2.1	Materialer . . . . .	37
7.2.2	Statistisk analyse . . . . .	38
7.3	Resultater . . . . .	38
7.3.1	Primereffektivitet . . . . .	38
7.4	Myosin Heavy Chain . . . . .	40

7.5	Diskusjon . . . . .	40
	<b>Kilder</b>	<b>42</b>

# Forord

Tusen takk til Lone, Lars og Daniel for svært godt samarbeid! Lone, du har vært en ekte helt på ImageJ, takk for det. Lars du guidet oss trygt gjennom  $\dot{V}O_{2max}$ -testene i testlabben, takk for det. Og takk til Daniel for artige, frustrerte ansiktsuttrykk på snap, R er ingen lek!

Takk til Ole og Eskil for veiledning med kod.

Til slutt rettes en takk til Jens Konrad Daniel Hammarström, for eminent undervisning i emnet.

Dennne oppgaven hadde ikke vært mulig å fått til uten de nevnte.

# 1 Reliabilitet

## 1.1 Introduksjon

Hensikten med denne rapporten er å presentere estimer for reliabilitet av data samlet i fysiologisk testlab ved Høgskolen i Innlandet. Hovedfokuset i eksperimentet er å vise til god standardisering og gjennomtenkt metode for å minske støy, dette er viktig for å øke reliabiliteten (Halperin *et al.*, 2015). Ifølge (Hopkins, 2000a) er reliabilitet reproduserbarheten av verdier fra en test, analyse, eller annen måling ved gjentatte forsøk på samme individ. Vi har gjennomført  $\dot{V}O_{2max}$ -tester på sykkel der 16 deltakere har gjennomført 2 eller 4 tester i løpet av tre uker. De to første testene ble gjennomført med ca. 24 timers mellomrom i uke en. De to siste testene ble gjennomført med ca. 48 timers mellomrom i uke tre. Vi har brukt Rstudio for å analysere dataen og få svar på om våre tester er reliable.

## 1.2 Metode

Vi gjennomførte en test-retest i vår studie hvor vi samlet inn data fra flere  $\dot{V}O_{2max}$ -tester. 16 deltakere gjennomførte to til fire tester i løpet av en periode på tre uker. De to første testene ble gjennomført med 24 timers mellomrom i løpet av den første uken. De to siste testene ble gjennomført med 48 timers mellomrom i løpet av den tredje uken.

### 1.2.1 Standardisering før test

Resultatene fra en  $\dot{V}O_{2max}$ -test påvirkes av ulike fysiologiske og miljømessige faktorer, så det er viktig å sette visse regler som deltakeren må følge i dagene før testen. En veiledning ble sendt til deltakerne noen dager før deres første test med informasjon om hvordan de bør forberede seg til testdagen. Denne veiledningen bør inkludere standardiseringsrutiner som hjelper til med å isolere deltakernes sanne aerobe ytelse og eliminerer eksterne faktorer som kan påvirke testresultatene. Biologiske faktorer som kontrolleres for er fysisk aktivitet, hydrering, søvn, koffein- og kaloriinntak, alkoholforbruk og tidspunktet på dagen testen gjennomføres. Deltakeren må følge den samme rutinen før sin neste test for å sikre nøyaktighet, pålitelighet og sammenlignbarhet av testresultatene.

### 1.2.2 Utstyr

Spesifikt utstyr er nødvendig for å utføre en  $\dot{V}O_{2max}$ -test. Deltakerne utførte testen på en elektromagnetisk bremsset sykkelergometer (Lode Excalibur Sport, Lode. B.V., Groningen, Nederland) for å utfordre deltakernes aerobe kapasitet. Vi brukte også et system kalt Vyntus CPX mixing chamber (Vyntus CPX JaegerCareFusion, UK) for å måle  $\dot{V}O_{2max}$ . For å måle hjertefrekvensen brukte vi en bryststropp fra Polar. For å måle deltakernes laktatnivå brukte vi en maskin kalt Biosen c-line (EKF diagnostic GmbH, Barleben, Tyskland). Det ble også benyttet en stoppeklokke, en ekstra datamaskin for vårt Excel-ark og en vekt for å måle kroppsvekt. Hele oppsettet gir oss mulighet til å overvåke og måle nødvendige data for å bestemme deltakernes  $\dot{V}O_{2max}$ .

### 1.2.3 Protokoll

Testprotokollen for å utføre en  $\dot{V}O_{2max}$ -test beskriver hvordan testen skal gjennomføres hver gang, da standardisering er viktig for å få nøyaktige og pålitelige resultater. En deltaker bør derfor ha samme testleder dersom en test gjentas. Under testene våre fulgte vi følgende protokoll, der et Excel-ark ble klargjort av testlederen før ankomst til laboratoriet. Ved ankomst tar testlederen på seg labfrakk og sørger for at Biosen (laktatmåling) er slått på. Maskinen må deretter kalibreres, og resultatet skal være 12 mmol/l. Ved kalibreringsfeil, som kan føre til unøyaktige data, byttes standardvæsken. Vyntus må deretter kalibreres ved hjelp av gasskalibrering og volumkalibrering. Gasskalibreringen må være innenfor 2,0 i differanse, og volumkalibreringen må være innenfor 0,2. Her må gassbeholderen åpnes, og ventilasjonsvolumtransduseren (Triple-V-turbinen) festes til Vyntus. Mens kalibreringen pågår, monterer testlederen munnstykket sammen med nesklemmen og kobler det til miksekompartimentet via en slange. Deretter må ergometersykkelen kalibreres ved å posisjonere krankarmen vertikalt. Denne kalibreringen gjøres på datamaskinen koblet til ergometersykkelen, da vi bruker et annet program her enn Vyntus. Krankarmen kontrolleres for å sikre at den er 172,5 mm lang og at riktig pedaltipe er montert. Til slutt samles utstyret for å måle laktat, og alt er klart for deltakerens ankomst.

En deltakerprofil må opprettes både i Vyntus og ergometersykkelens program. Profilen inkluderer deltakerens navn, fødselsdato, kjønn, høyde og vekt. Det første vi gjør når deltakeren ankommer, er å måle vekt. Vekten reduseres med 300g p.g.a. klær. Deretter justeres ergometersykkelen slik at deltakeren sitter komfortabelt. Innstillingene på sykkelen lagres deretter i sykkelprogrammet for fremtidige tester. Oppvarmingen varer i 5 minutter, der deltakeren skal ha en gradvis økning på Borgs skala for opplevd anstrengelse (RPE) fra 10 til 13. Hvis deltakeren ønsker det, slås viften på og rettes mot dem. Testlederen gir deretter informasjon om testen og Borg-skalaen.

Vyntus settes deretter til "måling," og testlederen sørger for at pulsmåleren er koblet til Vyntus, samt at innstillingene er satt til 30 sekunders målinger og størrelse medium for munnstykket.

Gassbeholderen lukkes, og Triple-V kobles fra Vyntus og festes til miksekommeret. Oppvarmingen er ferdig, og testen starter i Vyntus. Selve testen begynner etter 1 minutt i Vyntus, og  $\dot{V}O_{2max}$ -protokollen startes i sykkelprogrammet. En stoppeklokke startes også og plasseres foran deltakeren.

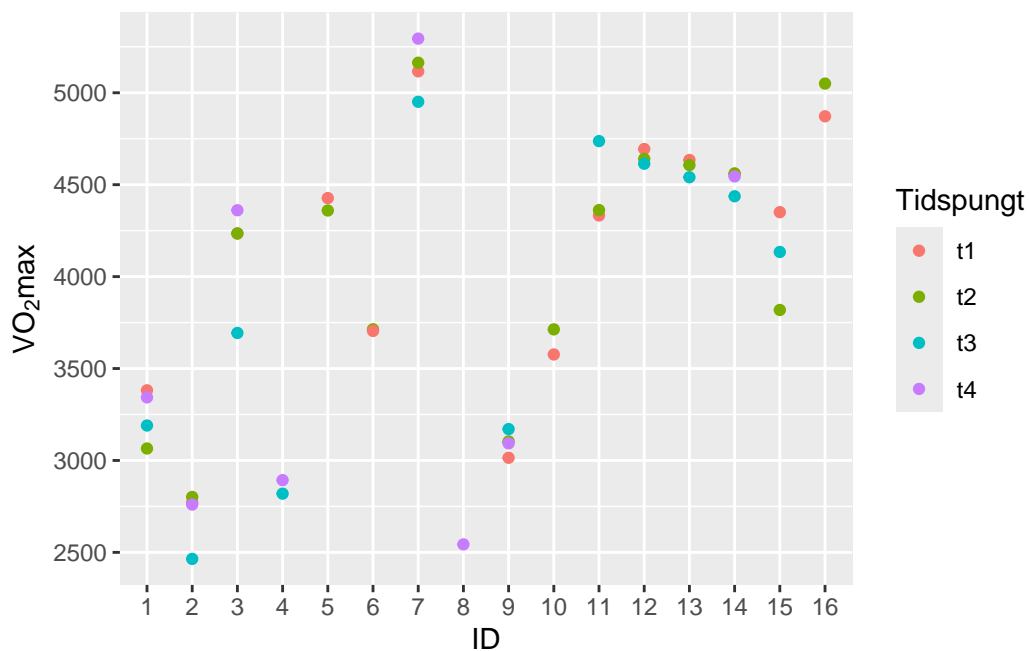
Under testen informerer testlederen om økninger i motstand (W), normalt 20W for kvinner og 25W for menn, og oppmuntrer deltakeren til å trække til utmattelse. Testen avsluttes når deltakeren gir seg eller RPM faller under 60. Testlederen noterer makspuls, sluttid, watt ved endt test og Borg-score på slutten av testen i excel-arket. Laktat tas ett minutt etter testslutt ved å tørke av fingeren til deltakeren, stikke et hull, tørke av den første bloddråpen og deretter fylle opp røret. Blodet analyseres deretter i Biosen. Testen er ferdig for deltakeren. Testlederen avslutter testen i både Vyntus og sykkelprogrammet. Sykkelen, munnstykket, slangen og pulsmåleren vaskes, og viften rettes mot miksekommeret for å la det tørke.

#### 1.2.4 Databehandling etter test

Korrekt innsamling av data fra testen er avgjørende, da det gir oss innsikt i deltakerens aerobe kapasitet. Rapporten generert av Vyntus lagres på en minnepinne for senere opplasting til OneDrive. Excel-arket fylles deretter med verdier fra rapporten. Disse verdiene inkluderer absolutt  $\dot{V}O_{2max}$ , uttrykt i milliliter oksygen per minutt (ml/min), som gir oss det totale oksygenforbruket til deltakeren og er nødvendig for å vurdere den aerobe kapasiteten. Maksimal mengde karbondioksid produsert ( $vco2_{max}$ ) viser hvor mye CO<sub>2</sub> som utåndes under testen. Respiratoriskutvekslingsforhold (RER) estimerer forholdet mellom produsert karbondioksid (CO<sub>2</sub>) og forbrukt oksygen (O<sub>2</sub>) under metabolisme. Maksimal minuttventilasjon (VE<sub>max</sub>) er det høyeste volumet av luft deltakeren kan bevege inn og ut av lungene per minutt. Maksimal pustefrekvens (BF<sub>max</sub>) er det maksimale antallet pust per minutt. Til slutt samles laktatmålingene fra Biosen. Deretter fikk vi data fra de andre gruppene og samlet alt i ett excel-ark. Dataene ble deretter eksportert til RStudio for videre analyse av resultatene.

### 1.3 Resultat





Figur 1.1: Deltakere og testresultat

Tabell 1.1: Deltakerenes VO<sub>2</sub>max-verdier i ml/min

ID	Test 1	Test 2	Test 3	Test 4
ID1	3381.5	3065.0	3190.0	3343.0
ID2	2771.0	2801.5	2464.5	2760.0
ID3	4234.5	4235.0	3693.5	4361.0
ID4	NA	NA	2819.5	2893.0
ID5	4427.0	4359.5	NA	NA
ID6	3704.5	3713.5	NA	NA
ID7	5116.5	5163.5	4951.0	5294.5
ID8	NA	NA	NA	2543.5
ID12	4694.0	4640.5	4614.0	NA
ID9	3014.5	3103.5	3170.5	3093.0
ID10	3576.5	3713.0	NA	NA
ID11	4332.5	4362.0	4737.0	NA
ID13	4634.5	4606.5	4540.5	NA
ID14	4556.5	4561.5	4437.0	4545.0
ID15	4350.5	3818.5	4134.0	NA

Resultatene fra vår utrekning av målefeil fra  $\dot{V}O_{2max}$ -testene viste at mellom t1 og t2 var 3.16 % variasjonskoeffesienten og mellom t3 og t4 var 4.2 % variasjonskoeffesienten. Dette tilsvarer en målefeil på 129.8ml mellom t1 og t2 og 170.3ml mellom t3 og t4.

## 1.4 Diskusjon

Variasjonskoeffesientene 3.16% og 4.2% er et mål på den typiske prosentvise feilen i målingene. Det er både svakheter og styrker ved våre gjennomføringer av testene. Dette var for mange, både som deltaker og testleder, det første møte med fysiologisk testlab og  $\dot{V}O_{2max}$ -test. Selv om formålet med rapporten var å estimere reliabilitet, var hensikten med testgjennomføringene at deltakerne skulle lære seg å gjennomføre en  $\dot{V}O_{2max}$ -protokoll. Dette førte til svakheter som at det ble brukt 16 forskjellige testledere og assistenter, dataen ble samlet inn av fire forskjellige grupper, antall observatører varierte og innsatsen var ikke alltid maksimal da motivasjonen ikke var på hos alle deltakerne. På en annen side var utstyret det som regnes som gullstandarden ved test av  $\dot{V}O_{2max}$ . Læreren som underviste hadde mye erfaring og høy kompetanse.

I følge (Hopkins, 2000a) er observert endring reliabel om den er større enn den typiske feilen. I vårt tilfelle vil det tilsvare en større endring enn 129.8ml og 170.3ml noe vi synes å være i overkant mye. De nevnte svakhetene ved gjennomføringene vil operere som forstyrrende variabler og kan dermed virke negativt på reliabiliteten (Halperin *et al.*, 2015).

Basert på vår estimering av reliabilitet fra testene synes vi at variasjonskoeffesienter på 3.16% og 4.2% er for høye. Det vil da kreve i overkant mye å skulle bekrefte endring ved ny test. Dette kommer mest sannsynlig som følge av en for dårlig og variert gjennomføring av metoden.

## 2 Predikere data

### 2.1 Introduksjon

Hensikten med denne rapporten er å predikere data ved hjelp av regresjon gjort i Rstudio, samt tolke en regresjonstabell. Rapporten inneholder tre deler. Ved å bestemme laktatterskel ved blodlaktatverdiene 2 og 4 mmol L<sup>-1</sup> analyserer vi forholdet mellom prestasjon i watt og treningsintensitet. Vi analyserte hellningen til en qPCR kalibreringskurve, og tolket en regresjonstabell om forholdet mellom 3RM squat og tverrsnittsareal til type II muskelfibrer.

### 2.2 Metode

#### 2.2.1 Del 1: Predikert laktatterskel

I del 1 av rapporten ble datasettet cyclingstudy fra (Sylta *et al.*, 2016) brukt til å predikere to bestemte blodlaktatterskler ved 2 og 4 mmol L<sup>-1</sup>. Det ble testet ut fire modeller for estimering av blodlaktatterskler for ID10, en rett linje modell (m1), andregradspolynom (m2), tredjegradspolynom (m3) og fjerdegradspolynom (m4). Behandlingen av dataen ble gjort i (R Core Team, 2021).

#### 2.2.2 Del 2: Predikert hellning på qPCR kalibreringskurve

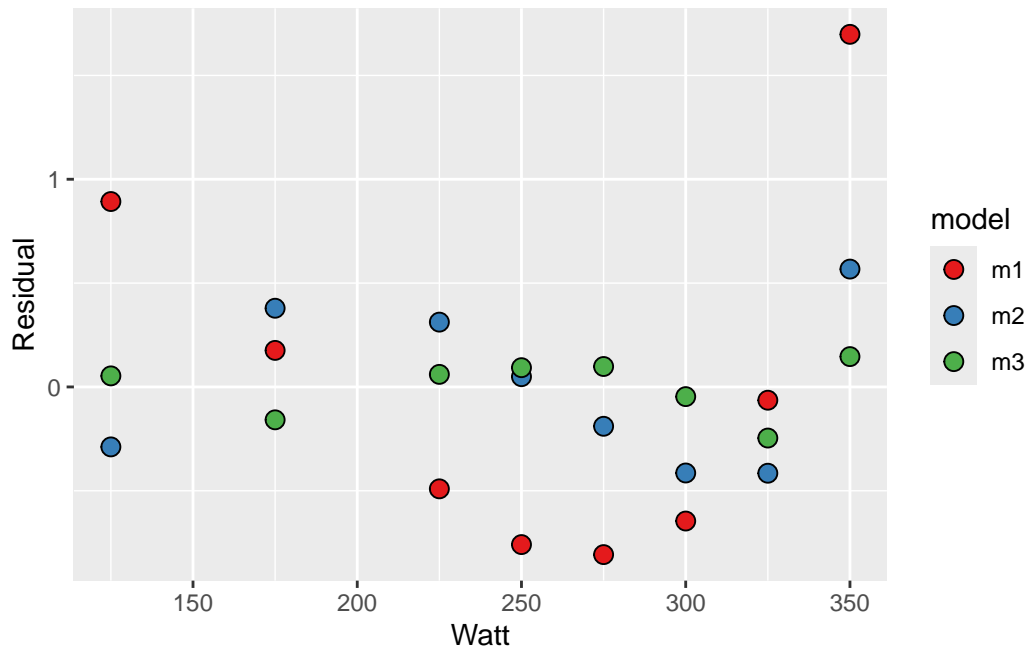
I del 2 av rapporten har vi ved hjelp av (Schindelin *et al.*, 2012) analysert et bilde av qPCR som ble hentet fram fra forsøket (Anon). Analysen av bildet gav oss data som vi anvendte i (R Core Team, 2021) for å predikere hellningen til qPCR kalibreringskurven.

#### 2.2.3 Del 3: Tolke en regresjonstabell

I del 3 av rapporten har vi gjort en statistisk analyse av forholdet mellom “Type II (FAST) fibers cross sectional area (micrometer<sup>2</sup>)” ved baseline (FAST\_CSA\_T1) og “Squat 3 repetition maximum load (kg)” ved baseline (SQUAT\_3RM) fra datasettet til (Haun *et al.*, 2018) og (Haun *et al.*, 2019) for å undersøke om det var et linjert forhold.

## 2.3 Resultat

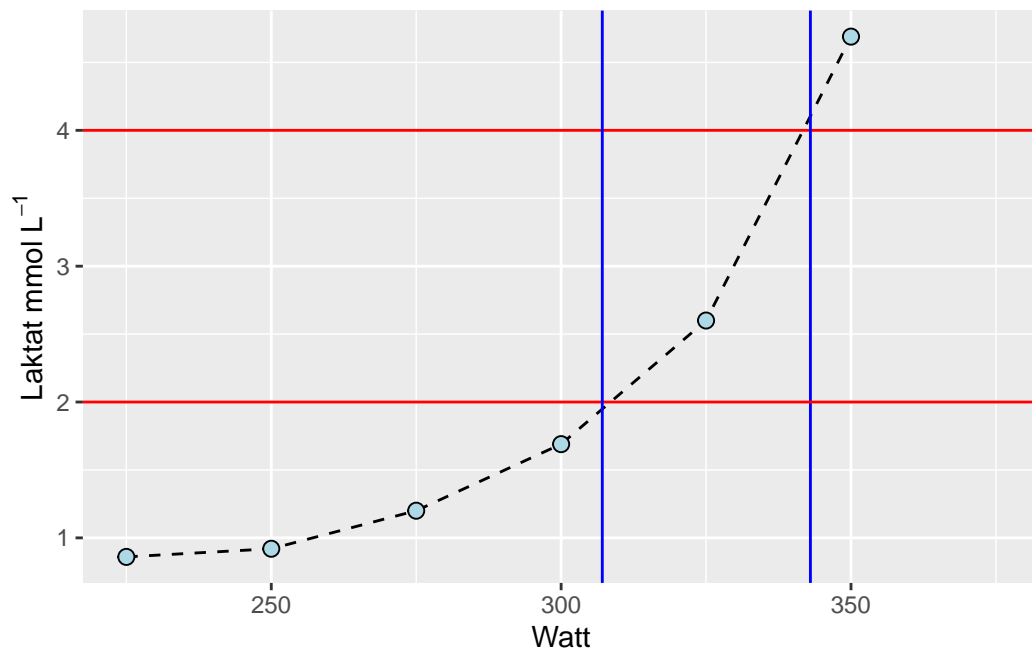
### 2.3.1 Del 1: Predikert laktatterskel



Figur 2.1: Tre forskjellige polynomiale regressionsmodeller.

Vi fant ut at en tredjegradspolynom (m3) var den beste modellen for å estimere laktatverdier da residualene fra denne modellen varierte minst fra observerte verdier, se Figure 2.1.

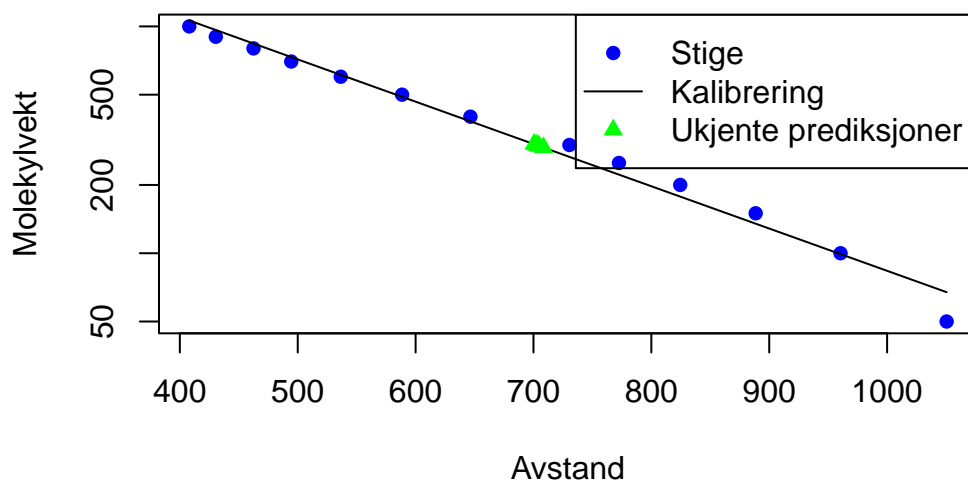
Vi har kalkulert blodlaktatterskel ved  $4\text{mmol L}^{-1}$  til å være 343 watt, og blodlaktatterskel ved  $2\text{mmol L}^{-1}$  til å være 307 watt for subject “10” ved timepoint “pre”. Se Figure 2.2.



Figur 2.2: Figur som viser 2 predikerte blodlaktatterskler for ID10

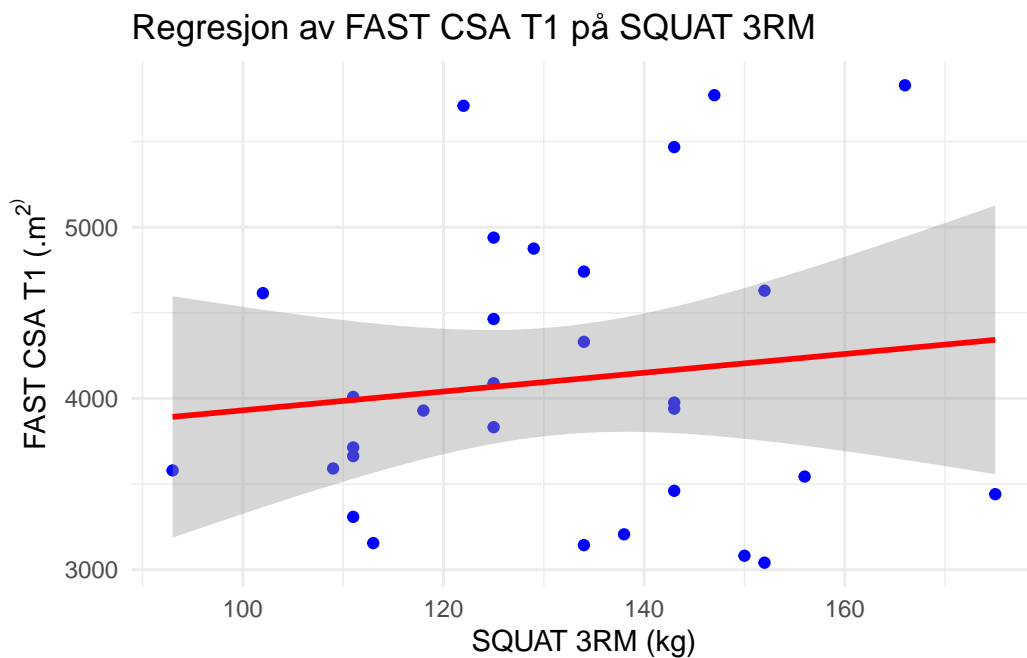
### 2.3.2 Del 2: Predikert hellning på qPCR kalibreringskurve

#### Kalibreringsmodel: Avstand vs. Molekylvekt



Kalibreringsmodellen beskriver forholdet mellom avstand og molekylvekt. Etter log-transformering av molekylvekten, blir forholdet mellom avstand og molekylvekt tilnærmet lineært, noe som gjør den lineære regresjonsmodellen passende. R-kvadratverdien på 0.98 indikerer hvor godt modellen passer til dataene da den er tilnærmet 1. De forutsagte molekylvektene for de ukjente prøvene er basert på den tilpassede kalibreringsmodellen. Predikasjonene for de ukjente prøvene gir et estimat av deres molekylvekt basert på deres migrasjonsavstand i gelen. QQ-plottet forteller oss om forskjellene mellom observert og predikert molekylvekt følger en normalfordeling. Ideelt sett bør punktene i QQ-plottet falle langs referanselinjen noe de tilsynelatende gjør. Det vil si at modellen er godt spesifisert, og fanger forholdet mellom migrasjonsavstand og molekylvekt.

### 2.3.3 Del 3: Tolke en regresjonstabell



Figur 2.3

Resultatene viste ingen sammengeng mellom SQUAR\_3RM og FAST\_CSA\_T1 (Estimat: 5.483 m<sup>2</sup>, SE= 8.032, T= 0.683, p= 0.50). Se Figure 2.3.

## 2.4 Diskusjon

### 2.4.1 Diskusjon - del 1

Rapporten viser hvordan man kan kalkulere seg fram til wattbelastning på sykkel ved bestemte laktatverdier på 2 og 4 mmol L<sup>-1</sup> som beskriver intensiteten. Vi kom fram til at terskelwatten ved 2 mmol L<sup>-1</sup> var 307, og at terskelwatten ved 4mmol L<sup>-1</sup> var 343 for subject “10” ved time-point “pre”. Beregningen ble gjort ved å bruke en tredjegradspolynomial modell da estimatene fra denne passet bedre enn en linjeær og andregradspolynomial modell. En tredjegradspolynomial modell har tidligere blitt bevist å passe bra for estimering av blodlaktat-kurve (Newell *et al.*, 2007).

### 2.4.2 Diskusjon - del 2

Kalibreringsmodellen viser seg å være effektiv til å beskrive forholdet mellom avstand og molekylvekt, spesielt etter at molekylvekten ble log-transformert. Denne transformasjonen bidro til å gjøre forholdet tilnærmet lineær, noe som bekrefter at en lineær regresjonsmodell er en passende tilnærming. Med en R-kvadratverdi på 0.98 ser vi at modellen gir en nesten perfekt tilpasning til dataene, som styrer modellens prediksjonsnøyaktighet. For de ukjente prøvene baseres prediksjonene av molekylvekt på modellen og gir troverdige estimater basert på migrasjonsavstand i gelen. QQ-plottet støtter også modellens robusthet, ettersom punktene faller nær referanselinjen, som indikerer at restene følger en normalfordeling. Dette antyder at modellen er godt spesifisert og gir en pålitelig beskrivelse av sammenhengen mellom migrasjonsavstand og molekylvekt. Samlet bekrefter resultatene at den utviklede modellen fanger opp de sentrale aspektene ved dataene på en god måte.

### 2.4.3 Diskusjon - del 3

Resultatene viser ingen sammenheng mellom FAST CSA T1 og SQUAT 3RM (Estimat: 5.483 m<sup>2</sup>, SE= 8.032, T= 0.683, p= 0.50). P-verdien på 0.50 tilsier i dette tilfellet at vi vil se et likt eller mer ekstremt resultat i 50% av repeterte studier gitt at nullhypotesen er sann. T-verdien er lav og indikerer at differansen mellom gjennomsnittet i utvalget og gjennomsnittet i populasjonen sannsynligvis er liten (Spiegelhalter, 2019). Den lave t-verdien sammen med den høye p-verdien indikerer at det ikke er statistisk grunnlag for å hevde at forskjellen mellom gjennomsnittet i utvalget og populasjonen er signifikant. Standardfeil (SE) forklarer hvor mye gjennomsnittet fra vårt utvalg forventes å avvike fra det sanne gjennomsnittet i populasjonen (Spiegelhalter, 2019). Oppsummert gir dette grunnlag for å si at det ikke noen sammenheng mellom økning i kilo i SQUAT 3RM og økning i m<sup>2</sup> FAST CSA T1.

## 3 Statistisk inferens

### 3.1 Simulering

I denne rapporten benyttes (R Core Team, 2021) til å simulere en populasjon på 1 million med et gjennomsnitt på 1.5 og et standaravvik på 3. For å kunne reprodusere resultatene benyttes `set.seed(1)`. Deretter skapes det to studier hvor det trekkes to tilfeldige utvalg, et på 8 (`samp1`) og et på 40 (`samp2`). Videre lages en lineær modell for `samp1` og `samp2` som lagres som `m1` og `m2`.

Alle oppgavene er utført i (R Core Team, 2021) og koden vil bli vist på siste side.

### 3.2 Forklaring av “estimate”, “Standardfeil”, “t-verdi” og “p-verdi” fra `m1` og `m2`

Estimatet er en kalkulert teoretisk verdi av snittet til `samp1` (1.84) og `samp2` (1.564). Standard feil er estimert variasjon i utvalgets gjennomsnitt. Dette forteller oss hvor presist gjennomsnittet fra utvalget estimerer populasjonens sanne gjennomsnitt. I dette tilfellet 1.251 (`m1`) og 0.4774 (`m2`). T-verdiene er 1.47 (`m1`) og 3.276 (`m2`). `m1` i dette tilfellet sier at gjennomsnittet (1.840) er 1.47 standardfeil unna null noe som forteller oss at snittet ikke er signifikant (når signifikansnivå =  $p < 0.05$ ) forskjellig fra nullhypotesen. P-verdien på 0.185 (`m1`) og 0.002 (`m2`) viser sannsynligheten for å få en t-verdi like ekstrem eller mer ekstrem enn den observerte t-verdien, gitt at nullhypotesen er sann. P-verdien regnes ut fra t-verdien og antall frihetsgrader og sier i dette tilfellet (`m1`) at det er 18.5% sannsynlighet for å få en t-verdi på 1.47 eller mer ekstrem gitt at nullhypotesen er sann. I tilfellet til (`m1`) kastes derfor ikke nullhypotesen.

### 3.3 Hvorfor forskjellig resultat i `m1` og `m2`?

Størrelsen på utvalget gir forskjellig avvik fra gjennomsnittet. `m1` med et utvalg på 8 har større standardfeil da det er større sannsynlighet for at det er større variasjon i utvalget. Slik jeg forstår det vil en “uteligger” i en studie hvor  $N = 8$  øker variansen i gjennomsnittet i større grad og derfor gi større utslag på standardfeilen. Dette fører igjen til en høyere p-verdi.



### 3.4 Hvorfor bruker vi det skraverte området i nedre og øvre del av en t-fordeling?

Det skraverte området i en t-fordeling visualiserer sannsynligheten for å få en verdi dersom nullhypotesen er sann. Dette brukes i hypotesetesting for å si at et resultat er signifikant forskjellig fra null.

P-verdien beregnes som sannsynligheten i halen(e) av t-fordelingen basert på teststatistikken ( t-verdien). Et gitt signifikansnivå, ofte satt til  $p < 0.05$ , avgjør om resultatet anses som statistisk signifikant. For en tosidig test betyr dette at t-verdien må falle utenfor de sentrale 95 % av t-fordelingen (dvs. i de ytterste 2.5 % i hver hale) for å bli ansett som signifikant forskjellig fra null.

### 3.5 Repeterte studier

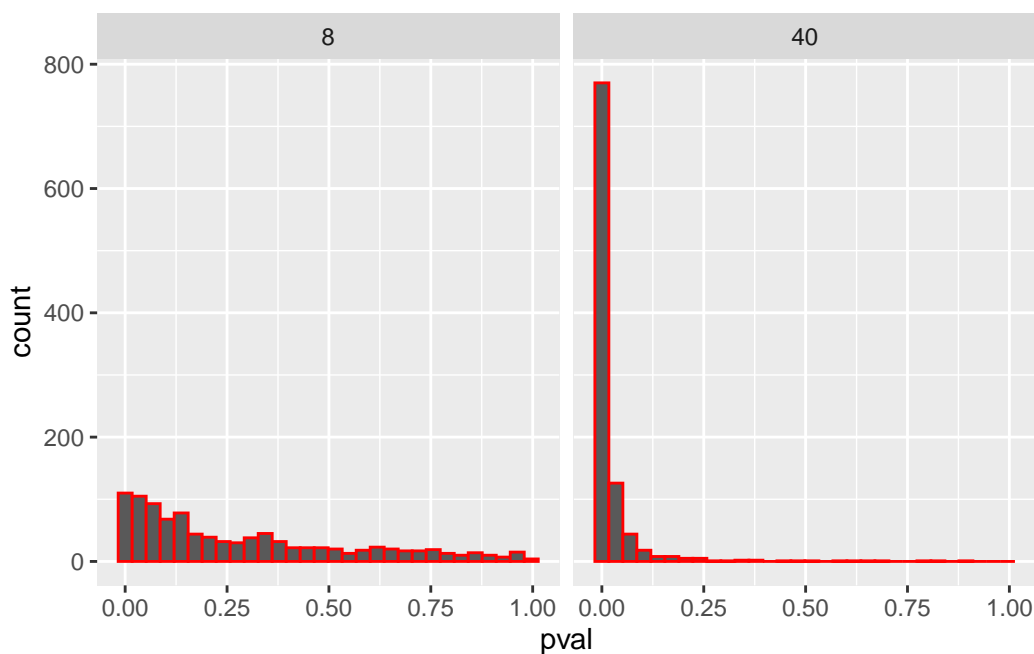
I de neste oppgavene simuleres studiene 1000 ganger, og lagres deretter som `results_8` og `results_40`.

### 3.6 Hvorfor er standardavviket til estimert gjennomsnitt så lik standardfeilen i dette tilfellet?

Standardavviket til variabelen `estimate`, i simuleringen hvor  $n = 8$  (`sim8`) og  $n = 40$  (`sim40`), er 1.078 og 0.454. Gjennomsnittlig standardfeil for `sim8` og `sim40` er 1.023 og 0.469. Standardavviket samsvarer godt med standardfeilen fordi studiene nå er gjort 1000 ganger med tilfeldige utvalg fra populasjonen. Dette fanger opp mer av populasjonens egenskaper og standardavviket vil nærmer seg en teoretisk fordeling.

Grunnen til at standardavviket og standardfeilen er større ved små utvalg er at variasjonen mellom gjennomsnittene er større.

### 3.7 Histogram med p-verdier fra simuleringene, hva sier dette om statistisk styrke?



Histogrammene viser fordelingen av p-verdier fra studiene. P-verdiene fra studiene gjort med et utvalg på 8 har en bredere fordeling enn studiene gjort med utvalg på 40. Mindre utvalg er mer følsomme for variasjon i dataen, uteliggere og større spredning viser seg som større spredning i p-verdier. Mindre utvalg vil i tillegg gi lavere presisjon og redusere sannsynligheten for signifikante resultater.

Større utvalg vil derimot oftere føre til lavere p-verdi. Det større utvalget vil også gi en lavere standardfeil da variasjonen i gjennomsnittet reduseres. Et større utvalg er også mindre følsomt for uteliggere og vil gi en høyere t-verdi selv med lik effektstørrelse. Dette kan leses som større statistisk styrke.

### 3.8 Filtrer resultater med en gitt p-verdi

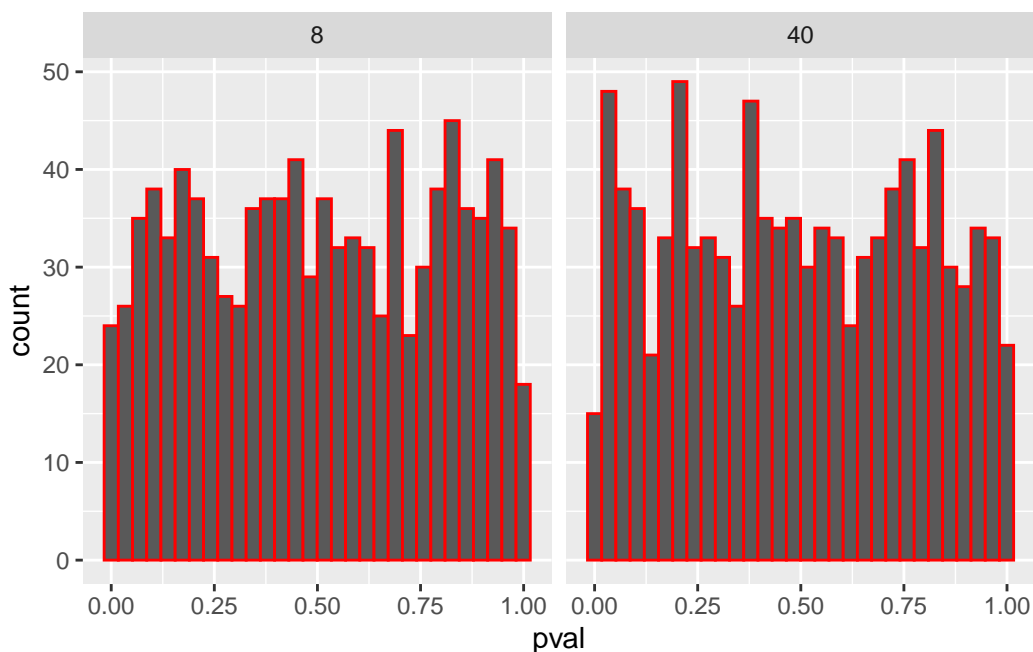
Av 1000 studier simulert med et utvalg på 8 hadde 213 studier  $p < 0.05$ . Med et utvalg på 40 var det 895 studier med  $p < 0.05$ .

### 3.9 Test av statistisk styrke

Testing av statistisk styrke er gjort med pakken “pwr”.

Styrken til studiene gjort med utvalg på 40 er 87% og et utvalg med 8 er 23%. Dette vil si at studiene gjort med utvalg på 40 har 87% sannsynlighet for å oppdage opp en effekt om det er en effekt. En høy (87%) statistisk styrke vil også si at man har lav (13%) sannsynlighet for type-2 feil, å ikke oppdage en effekt dersom den finnes.

### 3.10 Hvor mange studier gir falsk positiv når signifikansnivået er satt til 5%?



Hvis man gjør mange repeterte studier med utvalg på 40 kan man anta at 59 er falsk positiv og 49 for utvalg på 8 dersom signifikansnivået er satt til 5%. Disse tallene er derfor å forvente etter 1000 simulerte studier da  $5\% = 50$ . Med signifikansnivå satt til 5% vil det si at vi tillater 5% sannsynlighet for å gjøre type-1 feil, altså å kaste nullhypotesen selv om den er sann.

```

#Simulering

library(tidyverse)

# Muliggjør reproduksjon av resultater
set.seed(1)

# Simuler en populasjon
population <- rnorm(1000000, mean = 1.5, sd = 3)

# Lag en studie med et utvalg på 8 hvor den avhengige variabelen heter y
samp1 <- data.frame(y = sample(population, 8, replace = FALSE))

# Lag en studie med et utvalg på 40 hvor den avhengige variabelen heter y
samp2 <- data.frame(y = sample(population, 40, replace = FALSE))

# Lag en lineær modell fra utvalg 1 og lagre den som m1
m1 <- lm(y ~ 1, data = samp1)
# Lag en lineær modell fra utvalg 2 og lagre den som m2
m2 <- lm(y ~ 1, data = samp2)

est_8 <- coef(m1)[1]
est_40 <- coef(m2)[1]

```

```

#Repeterte studier

set.seed(1)
# Create data frames to store the model estimates
results_8 <- data.frame(estimate = rep(NA, 1000),
                          se = rep(NA, 1000),
                          pval = rep(NA, 1000),
                          n = 8)

results_40 <- data.frame(estimate = rep(NA, 1000),
                          se = rep(NA, 1000),
                          pval = rep(NA, 1000),
                          n = 40)

# A for loop used to sample 1000 studies, each iteration (i) will draw a new sample
# from the population.

```

```

for(i in 1:1000) {

  # Draw a sample
  samp1 <- data.frame(y = sample(population, 8, replace = FALSE))
  samp2 <- data.frame(y = sample(population, 40, replace = FALSE))

  # Model the data
  m1 <- lm(y ~ 1, data = samp1)
  m2 <- lm(y ~ 1, data = samp2)

  # Extract values from the models
  results_8[i, 1] <- coef(summary(m1))[1, 1]
  results_8[i, 2] <- coef(summary(m1))[1, 2]
  results_8[i, 3] <- coef(summary(m1))[1, 4]

  results_40[i, 1] <- coef(summary(m2))[1, 1]
  results_40[i, 2] <- coef(summary(m2))[1, 2]
  results_40[i, 3] <- coef(summary(m2))[1, 4]

}

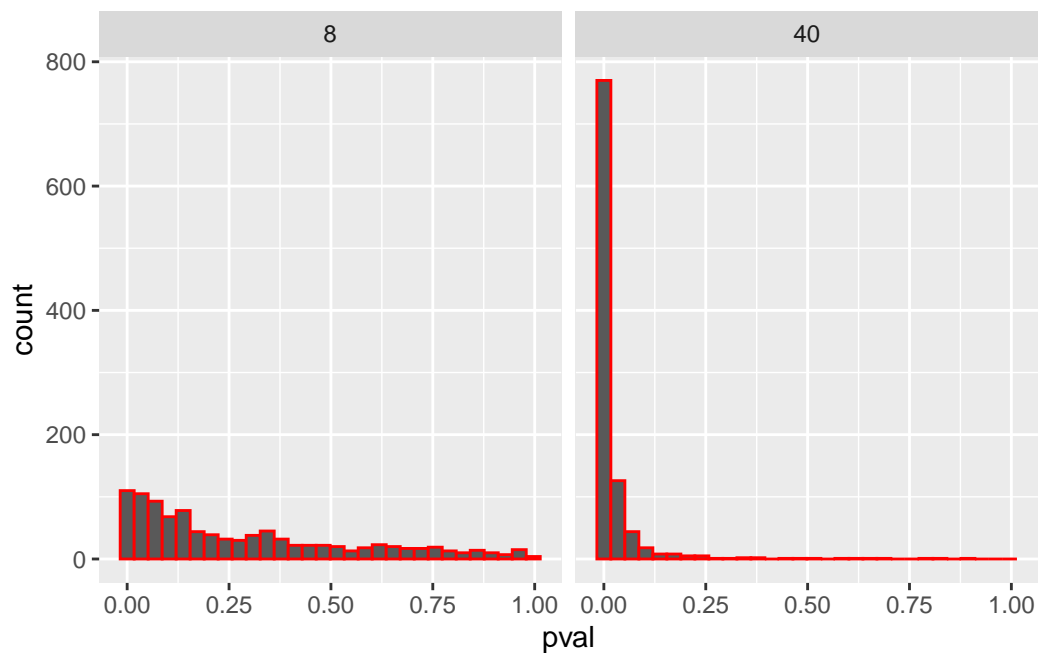
# Save the results in a combined data frame

results <- bind_rows(results_8, results_40)

# Utregning standardavvik av estimert gjennomsnitt
sd_res_8 <- sd(results_8$estimate)
sd_res_40 <- sd(results_40$estimate)

# Utregning gjennomsnitt av standardfeil
se_mean_8 <- mean(results_8$se)
se_mean_40 <- mean(results_40$se)

```



```
# Filtrer resultater med en gitt p-verdi
```

```
n_sig_8 <- results %>%
  filter(pval < 0.05) %>%
  filter(n != 40) %>%
  nrow()
```

```
n_sig_40 <- results %>%
  filter(pval < 0.05) %>%
  filter(n == 40) %>%
  nrow()
```

```
# Test av statistisk styrke
```

```
library(pwr)
```

```
pwr_40 <- pwr.t.test(n = 40, sig.level = 0.05, d = 1.5/3, type = "one.sample")
pwr_8 <- pwr.t.test(n = 8, sig.level = 0.05, d = 1.5/3, type = "one.sample")
```

```
numeric_pwr_40 <- pwr_40[["power"]]
pwr_pros_40 <- numeric_pwr_40 * 100
```

```
numeric_pwr_8 <- pwr_8[["power"]]
pwr_pros_8 <- numeric_pwr_8 * 100
```

```

# Hvor mange studier gir falsk positiv når signifikansnivået er satt til 5%?
set.seed(1)

population <- rnorm(1000000, mean = 0, sd = 3)

# Create data frames to store the model estimates
results_8 <- data.frame(estimate = rep(NA, 1000),
                        se = rep(NA, 1000),
                        pval = rep(NA, 1000),
                        n = 8)

results_40 <- data.frame(estimate = rep(NA, 1000),
                        se = rep(NA, 1000),
                        pval = rep(NA, 1000),
                        n = 40)

# A for loop used to sample 1000 studies, each iteration (i) will draw a new sample
# from the population.

for(i in 1:1000) {

  # Draw a sample
  samp1 <- data.frame(y = sample(population, 8, replace = FALSE))
  samp2 <- data.frame(y = sample(population, 40, replace = FALSE))

  # Model the data
  m1 <- lm(y ~ 1, data = samp1)
  m2 <- lm(y ~ 1, data = samp2)

  # Extract values from the models
  results_8[i, 1] <- coef(summary(m1))[1, 1]
  results_8[i, 2] <- coef(summary(m1))[1, 2]
  results_8[i, 3] <- coef(summary(m1))[1, 4]

  results_40[i, 1] <- coef(summary(m2))[1, 1]
  results_40[i, 2] <- coef(summary(m2))[1, 2]
  results_40[i, 3] <- coef(summary(m2))[1, 4]

}

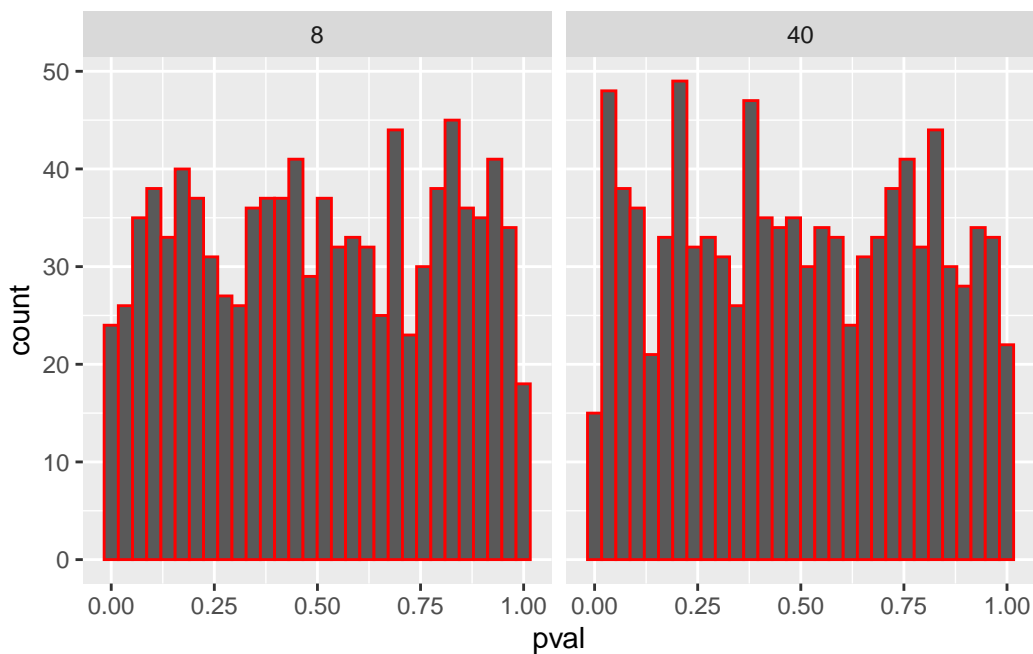
```

```
# Save the results in a combined data frame

results_null <- bind_rows(results_8, results_40)

false_positive_resnull <- sum(results_null$pval < 0.05)

results_null %>%
  ggplot(aes(pval)) +
  geom_histogram(color = "red") +
  facet_wrap(~ n)
```



```
n_sig_rnull_40 <- results_null %>%
  filter(pval < 0.05) %>%
  filter(n == 40) %>%
  nrow()

n_sig_rnull_8 <- results_null %>%
  filter(pval < 0.05) %>%
  filter(n != 40) %>%
  nrow()
```



## 4 Studiedesign

Hensikten med denne rapporten er å sammenligne fem studier sine studiedesign og metoder for statistisk analyse.

### 4.1 Introduksjon

I dagens idrettsvitenskapelige forskningsfelt spiller studiedesign og statistiske modeller en sentral rolle for innhenting og analyse av data. Statistisk analyse blir ofte brukt som et forsøk på å trekke meningsfulle konklusjoner fra en mengde innsamlede data. Dette være seg til hypotesetesting, forutsi fremtidige utfall eller avdekke årsakssammenhenger. Statistiske modeller kan være svært komplekse, noe som kan gjøre det vanskelig for forskere å vurdere hvilke analyser som egner seg best til sine innsamlede data (Atkinson & Nevill, 2001). Hvilke statistiske modeller som skal bli brukt på innsamlede data burde derfor ses i sammenheng med studiedesign (Hulley *et al.*, 2006).

Et studiedesign er strukturen til en studie, en beskrivelse av studiens gjennomføring og alle aspekter av datainnsamlingen (Knight, 2010). Studier har derfor forskjellig struktur for å best fange opp den aktuelle forskningens formål (Creswell, 2008). I studier hvor formålet er å undersøke fysiologiske responser på trening brukes ofte et randomisert-kontrollert studiedesign (RCT) da dette blir sett på som gullstandarden innen idrettsvitenskap (Bouchard *et al.*, 2012). Dette vil si at forsøkspersonene fordeles tilfeldig i en eksperimentell gruppe som får en behandling, her som blokkperiodisering av trening (BP) og en kontroll-gruppe som ikke får behandling, eventuelt en behandling til sammenligning, her som tradisjonell trening (TRAD) (Hulley *et al.*, 2006). RCT-designet balanserer deltakernes egenskaper mellom gruppene gjennom randomisering, noe som er med på å gjøre det mulig å tilskrive eventuelle forskjeller i utfall til intervensjonen som studeres (Hariton & Locascio, 2018).

### 4.2 Metode

Analysene bygger på QUALMRI-metoden (Brosowsky *et al.*, 2020) og tar for seg fem studier som undersøker effekten av blokkperiodisering på utholdenhetstrening med et kriterium om at den fysiologiske variabelen VO<sub>2</sub>max er blitt målt. Analysen har fokusert på likheter, ulikheter, styrker og svakheter ved studienes design og bruk av statistiske modeller. Dette er studiene som er analysert;

- “5-week block periodization increases aerobic power in elite cross-country skiers” (Rønnestad *et al.*, 2016)
- “Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists” (Rønnestad *et al.*, 2014*b*)
- “Block periodization in alpine skiing: effects of 11-day HIT on VO<sub>2</sub>max and performance” (Breil *et al.*, 2010)
- “Effect of 12 weeks block periodization on performance and performance indices in well-trained cyclist” (Rønnestad *et al.*, 2014*a*)
- “Performance changes in world-class kayakers following two different training periodization models” (García-Pallarés *et al.*, 2010)

## 4.3 Resultat

Alle studiene indikerer at BP gir større økning i VO<sub>2</sub>max enn TRAD. Analysen viser til en gjennomgående bruk av RCT, med én overkryssningsstudie (García-Pallarés *et al.*, 2010), som studiedesign med et utvalg på 10 - 20 forsøkspersoner totalt. Utvalget varierte fra godt trente til utøvere i verdensklasse. Videre ble det gjennomført pre- og post-intervensjonstesting av både fysiologiske variabler og variabler som mål på prestasjon. Studiene varierte med intervensjoner som pågikk fra 11 dager - 2 år.

Studiene benyttet seg av variansanalyser (ANOVA og t-test) for å måle forskjeller over tid mellom grupper. Én studie brukte kovariansanalyse (ANCOVA) og kun én viste til statistisk styrkeberegning.

## 4.4 Diskusjon

### 4.4.1 Studiedesign og valg av metode

I idrettsvitenskap anses et RCT-design som gullstandarden. Ved å bruke RCT-design, som de analyserte studiene gjør, skapes et godt grunnlag for å kontrollere for eksterne faktorer. Dette igjen øker sannsynligheten for at resultatene kan forklares på bakgrunn av intervensjonen. En av de inkluderte studiene benytter en overkryssningsstudie. Dette muliggjør sammenligning av to ulike treningsformer innad i samme individ, men kan være utfordrende med tanke på “carryover”-effekter – det vil si at effektene fra den første perioden påvirker resultatene i den andre perioden (Senn, 2002*a*).

Hva gjelder utvalg er det fordeler og ulemper ved bruk av eliteutøvere og utøvere i verdensklasse. På den ene siden gir eliteutøvere et høyt prestasjonsnivå noe som gjør det enklere å observere mindre, men meningsfulle forandringer i målte variabler (Hopkins, 2000*b*). På en annen side er utvalgsstørrelsen liten, noe som kan begrense studiens evne til å generalisere samt påvirke den statistisk styrken (Spiegelhalter, 2019).

#### 4.4.2 Statistiske modeller

I den statistisk analysen av dataene benyttes variansanalyse (ANOVA og t-test) i de fleste studiene. Disse analysemetodene er gode for å undersøke forskjeller over tid mellom grupper. Dette er også passende modeller for gjentatte målinger hvor forskerne ønsker å sammenligne en gruppe med seg selv og med en kontrollgruppe. En av studiene benytter ANCOVA. Denne kovariansanalysen justerer for variasjoner i pre-intervensjonsverdier (regression to the mean) og kan potensielt sett gi et bedre bilde av BP sin effekt. At kun én studie rapporterer statistisk styrke er en betydelig svakhet da dette begrenser forståelsen av eksperimentets pålitelighet (Spiegelhalter, 2019).

Statistisk styrke er særlig relevant da små utvalg øker risikoen for type II-feil (falske negative). Altså at man ikke finner en effekt der en effekt faktisk eksisterer (Spiegelhalter, 2019). Den lave bruken av beregning av statistisk styrke gjør det utfordrende å konkludere med at BP faktisk er signifikant bedre enn TRAD. En anbefaling, slik jeg ser det, vil være å inkludere beregning av statistisk styrke som en del av metoden i idrettsvitenskapelige studier.

#### 4.4.3 Varighet og oppfølging

Treningsintervensjonene i studiene varierer alt fra 11 dager til 2 år. Dette gir et variert bilde av BP sin effekt på kort og lang sikt. Variasjonen i intervensjonslengde gjør det mulig å sammenligne kortsiktige og langsiktige responser, men det kan også føre til utfordringer i hvordan vi tolker resultatene. Studier med lengre varighet kan for eksempel vise effekter som utvikles over tid, men det kan da være vanskeligere å kontrollere for de potensielt eksterne faktorene som påvirker utviklingen utover intervensjonen (Bacon *et al.*, 2013).

#### 4.4.4 Konklusjon

Samlet sett peker analysen mot at BP har større effekt på VO<sub>2</sub>max enn TRAD i de aktuelle studiene, men svakheter som små utvalg og mangel på statistisk styrke begrenser grunnlaget for å kunne gi noen generelle anbefalinger. For fremtidig forskning anbefales det å fokusere på metodologiske forbedringer, som å øke størrelsen på utvalget, gjøre flere longitudinelle studier med bedre standradisering og oppfølging. Hva gjelder statistikk vil implementering av styrkeberegning og standardiserte analyser øke studienes kredibilitet. Disse anbefalingene vil trolig gi et mer virkelighetsnært bilde av BP sin effekt på VO<sub>2</sub>max.

## 5 Styrketrening på bein med moderat volum har større effekt på maksimal styrke og muskelmass enn lavt volum

### 5.1 Introduksjon

Styrketrening er godt etablert for å være en effektiv metode for å øke maksimal styrke og muskelmasse, og det er per dags dato mange studier som kommer frem til denne konklusjonen (Currier *et al.*, 2023). Variabler slik som antall repetisjoner per sett, antall sett, treningsstatus og treningsfrekvens viser seg å ha en innvirkning på effekten av styrketrening (American College of Sports Medicine, 2009). Disse variablene er undersøkt med varierende resultat, da særlig spørsmålet om volum sin betydning for økning i maksimal styrke og muskelmasse.

Så hvis man ser til studier som ser på økning i maksimal styrke og muskelmasse, hvorpå få antall sett versus et moderat antall sett er sammenlignet, kan det vise seg å være manglende konsensus på feltet. I en studie gjort av (Cannon & Marino, 2010) og en gjort av (Mitchell *et al.*, 2012), som sammenlignet ett sett mot tre sett, så man ingen forskjell i hverken maksimal styrke eller muskelmasse. På en annen side ble det gjort en meta-analyse av (Krieger, 2009) som fant en større effekt på maksimal styrke ved å utføre to til tre sett mot for å bare utføre ett sett.

Formålet med denne rapporten er derfor å undersøke effekten av styrketrening med lavt volum og moderat volum på maksimal styrke og muskelmasse. Min hypotese er at moderat volum gir større økning i nevnte parametre enn lavt volum.

### 5.2 Metode

#### 5.2.1 Forsøkspersoner og studiedesign

34 ikke-røykende deltakere mellom 18 - 40 år gjennomførte studien, se Table 5.1. Eksklusjonskriteriene for deltakelse var; mer enn én styrketreningsøkt i uka siste året, begrenset kraftutvikling i muskulatur som følge av skade eller bruk av medisiner som kan påvirke effekten av trening. Alle deltakerne rapporterte at de hadde erfaring med idrettslige aktiviteter. 20 deltakere

rapporterte at de var fysisk aktive ved intervensjonsstart, 10 av disse gjorde sporadisk styrketrening.

Intervensjonen bestod av 12 uker med helkropps-styrketrening. Beintreningen ble utført unilateralt og hvert bein ble tilfeldig tildelt en protokoll. Det vil si at hver deltaker utførte begge protokollene. Det være seg ett sett per øvelse, lavt volum (1-sett), versus tre sett per øvelse, moderat volum (3-sett). Dette gjorde det mulig å se på forskjellen i treningsvolum innad i hver deltaker. Maksimal styrke ble målt ved baseline (pre) og etter intervensjonen (post). Kroppssammensetning ble målt før og etter intervensjonen.

Tabell 5.1: Beskrivelse av deltakere

Kjønn	Alder (år)	Høyde (cm)	Vekt (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)

### 5.2.1.1 Treningsprotokoll

I forkant av alle treningsøktene gjennomførte deltakerne en standardisert oppvarming. Oppvarmingen bestod av fem minutter sykling på ergometersykel (med en opplevd anstrengelse på 13-14 på Borg' skala), etterfulgt av 10 repetisjoner i forskjellige øvelser (armhevinger, magebøyninger, rygghev og knebøy) med kroppsvekt som motstand og til slutt ett sett å 10 repetisjoner på ~50% av en repetisjon maksimum (1RM) i hver stryketreningsøvelse for bein. Øvelsene for trening av bein ble gjennomført slik: ensidig beinpress, lårcurl og lårspark, gjort med enten 1-sett eller 3-sett, avhengig av hvilket bein som ble tildelt den respektive protokollen. Videre gjennomførte deltakerne to sett med benkpress, nedtrekk og enten skulderpress eller sittende roing (de to sistnevnte øvelsene ble variert mellom økter). Hvile mellom sett var 1.5 - 3min. Treningsintensiteten økte gjennom intervensjonen, fra 10RM de to første ukene til 8RM i tre uker og 7RM de siste sju ukene. I gjennomsnitt ble 91% av utførte økter gjort med oppfølging av en fra prosjektet. Deltakere som utførte økter uten oppfølging måtte avgi rapport slik at forskerne kunne sikre en eventuelt ønsket fremgang og riktig gjennomføring i henhold til protokoll. Etter den niende treningsøkten ble det innført at en av tre ukentlige treningsøkter skulle gjennomføres med submaksimal belastning. Dette tilsvarte 90% av belastningen fra forrige økt med samme antall repetisjoner. Treningøkter gjort med maksimal innsats ble gjort med 48-timers mellomrom og treningsøkter gjort med submaksimal belastning ble gjort med minst 24-timers mellomrom fra andre økter. For å bedre kvaliteten på restitusjonen fikk deltakerne en standardisert drikk etter hver treningsøkt med individualisert tilpasset mengde protein, fett og karbohydrater.

### 5.2.1.2 Mål på maksimal styrke og kroppssammensetning

Maksimal styrke ble målt som 1RM i ensidig beinpress og lårspark. Oppvarming før test ble gjennomført med 10, 6, og 3 repetisjoner på henholdsvis 50, 70 og 85% av forventet 1RM. Belastningen ble så økt helt til utførelsen ikke ble gjort i henhold til øvelsens bevegelsesbane. 1RM ble fastsatt på høyest mulig belastning gjort riktig i henhold til øvelsens bevegelsesbane. Deltakerne fikk fire til seks forsøk. Maksimal styrke ble testet to ganger pre intervensjon med minst fire dagers mellomrom. Høyeste 1RM fra beinpress ble brukt i senere analyse. Både pre- og post-test ble gjort med minst 48timers mellomrom før neste eller fra siste treningsøkt.

Kroppssammensetning (muskelmasse) ble målt ved bruk av dual-energy X-ray absorptiometry (DXA) (Lunar Prodigy, GE Healthcare, Oslo, Norway) i henhold til standard protokoll. I forkant av DXA-skanning ble deltakerne bedt om å møte fastende (to timer i forkant) og å avstå fra moderat til hard fysisk aktivitet 48 timer før skann.

## 5.2.2 Dataanalyse og statistikk

Alle deskriptive data er presentert som gjennomsnitt og standardavvik (gjennomsnitt (SA)). For å undersøke effekten av 1-sett versus 3-sett på maksimal styrke (beinpress) og fettfri masse ble det utført en parett-test. Dette fordi denne statistiske testen egner seg til å måle forskjeller innad i samme individ over tid (pre - post) (Cohen, 1988). Tallene i resultatdelen viser gjennomsnittlig forskjell mellom protokollene med tilhørende konfidensintervall (CI), *p*-verdi og *t*-verdi. Signifikansnivå ble satt til  $p < 0.05$ .

Den statistiske analysen ble gjort i R (R Core Team, 2021).

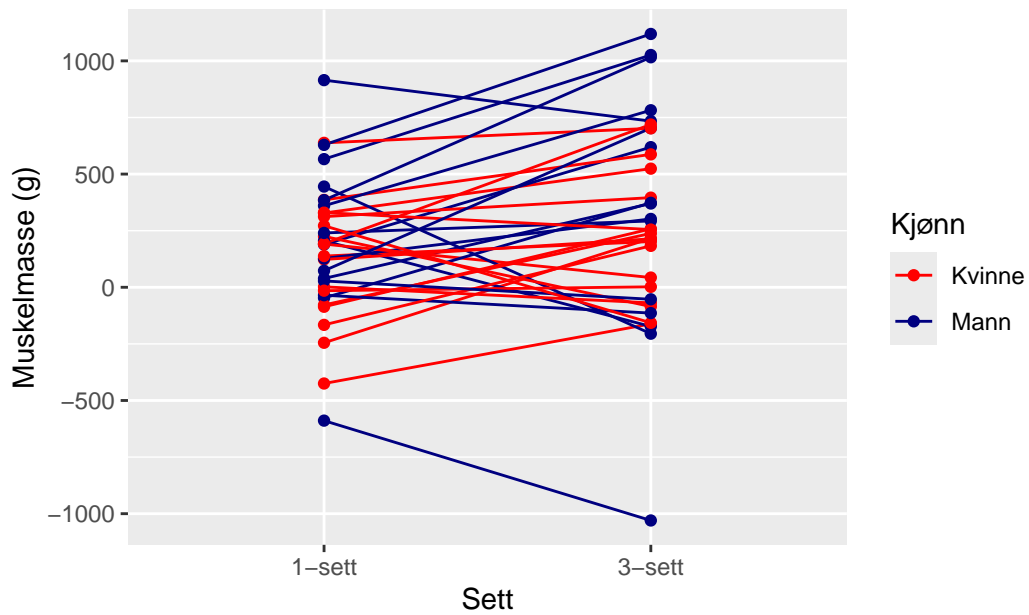
## 5.3 Resultater

### 5.3.1 Muskelmasse

Etter 12 ukers styrketrening viste analysene at det var en gjennomsnittlig forskjell i muskelmasse i gram mellom protokollene 1-sett versus 3-sett på 122.8 (95 CI: [8.6,237.0], *P*-verdi = 0.036,  $t_{33} = 2.19$ ) pre til post. For å se forskjellen mellom protokollene for hver deltaker, se Figure 5.1.

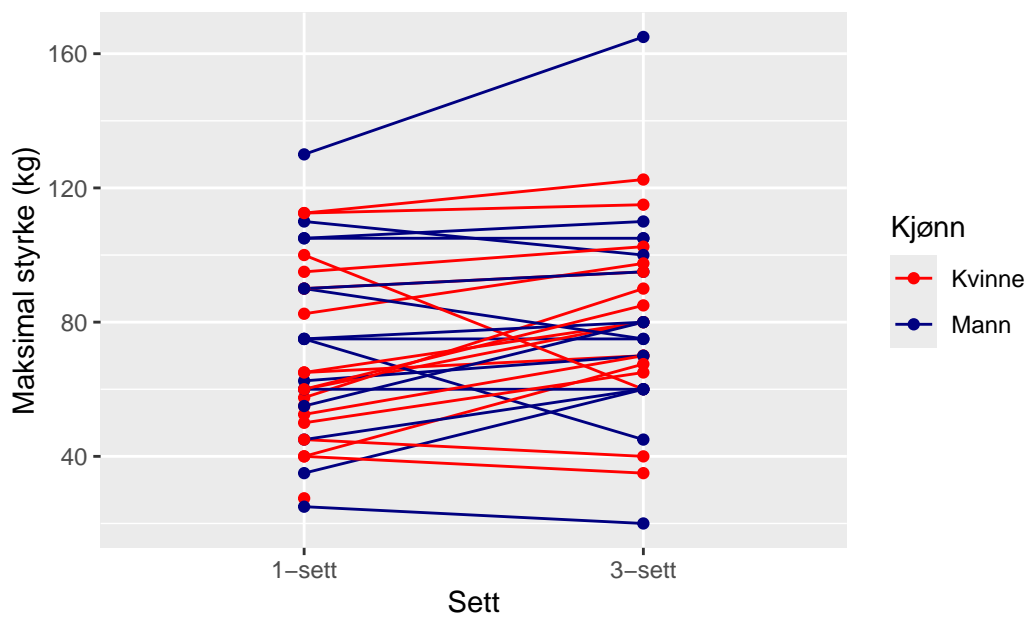
### 5.3.2 Maksimal styrke

Resultatene viste også en gjennomsnittlig forskjell i maksstyrke i kilogram mellom protokollene 1-sett versus 3-sett på 6.8 (95 CI: [0.7,12.9], *P*-verdi = 0.031,  $t_{30} = 2.26$ ) post til pre. For å se forskjellen mellom protokollene for hver deltaker, se Figure 5.2.



Figuren viser forskjell i muskelmasse mellom protokollene fra pre til post.

Figur 5.1



n viser forskjell i maksimal styrke i beinpress mellom protokollene fra pre til post.

Figur 5.2

## 5.4 Diskusjon

Denne studien viser at 3-sett med styrketrening på bein gir større økning i maksimal styrke og muskelmasse sammenlignet med 1-sett. Dette støtter min hypotese om at moderat volum gir større økning i maksimal styrke og muskelmasse enn lavt volum.

Tidligere forskning har som nevnt vist seg å være sprikende i sine resultater. Det finnes flere studier (Cannon & Marino, 2010), (Mitchell *et al.*, 2012) som ikke fant noen forskjell mellom ett sett og tre sett når det gjaldt økning i styrke eller muskelmasse. Men i meta-analyser gjort av (Krieger, 2009) og (Schoenfeld *et al.*, 2017) viser det seg at flere sett er bedre enn ett. Denne studien støtter meta-analysene og tilsier at et moderat treningsvolum kan være mer effektivt enn lavt volum for styrke og muskelvekst. Dette kan være til nytte for nybegynnere og folk som trener sporadisk styrketrening.

Designet i denne studien skiller seg fra tidligere studier og styrker funnene som taler for et mer moderat volum. Ved å tildele begge styrketreningsprotokollene til samme individ reduseres individforskjellene (Senn, 2002b). Dette betyr at man i større grad kan peke på effekten av treningsvolumet som begrunnelse for økning i styrke og muskelvekst og ikke personlige variabler som; personlig genetikk, kosthold eller treningsstatus. I tillegg ble 91% av øktene utført med oppfølging noe som reduserer risikoen for feil utførelse, og som gir lik kvalitet på treningen.

Samlet sett støtter funnene i denne studien hypotesen om at et moderat treningsvolum gir større økning i maksimal styrke og muskelmasse sammenlignet med et lavt volum. Dette kan ha praktisk betydning for å optimalisere treningsprogrammer, særlig for individer med lite erfaring i styrketrening. Likevel peker variasjonen i forskningsresultater på behovet for ytterligere forskning for å få klarhet i hvordan treningsvolum kan optimaliseres for ulike målgrupper og treningsmål.



## 6 Filosofi

### 6.0.1 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.

David Hume mente at for å trekke slutninger basert på induktiv logikk må man ha uniformitetsprinsippet, at naturen vil fortsette å reagere i fremtiden slik den har gjort i fortiden, som premiss. Hume hevdet at for å rasjonelt kunne bevise naturens uniformitet må induksjon begrunnes med induksjon. Dette ville da bli et sirkulært argument.

Ifølge Hume vil en sirkulær begrunnelse ikke være rasjonell da premisset for slutningen selv trenger begrunnelse. Derfor mener Hume at induksjon ikke kan begrunnes rasjonelt og at denne type konklusjoner ikke gir noe godt grunnlag for opphav til kunnskap. Han prøvde også å finne grunnlag for generalisering av kunnskap basert på tidligere erfaringer og observasjoner. Det klassiske eksempelet som tar for seg hvite svaner og om man kan si at alle svaner er hvite basert på tidligere observasjoner illustrer induksjonsproblemet. Her argumenterer man med at siden alle tidligere observerte svaner er hvite, er alle svaner hvite. I dette tilfellet generaliseres antagelser basert på tidligere erfaringer, men det trengs bare å observere en svane med annen farge for å avkrefte generaliseringen. Dette viser at induksjon som grunnlag for sannhet ikke holder mål og alltid vil være knyttet til usikkerhet (Vassend, 2024a).

Kunnskap er basert på erfaringer og observasjoner med uniformitetsprinsippet til grunn. At Hume ikke anerkjenner logiske slutninger som rasjonelle med bakgrunn i naturens uniformitet gir ikke grunnlag nok til å forkaste eller legge mindre vekt på fornuftige antakelser basert på erfaring. Humes argument var ment filosofisk og teoretisk. Han mente selv at en i hverdagslivet skulle følge sine logiske slutninger basert på erfaring. Det vil for eksempel, basert på erfaring om at gjenstander tar fyr, ikke være fornuftig å kaste glør fra peisen i restavfall. Uniformitetsprinsippet ligger til grunn, men ikke med en sannhet om at det tar fyr, men en sannsynlighet. Restavfall inneholder gjenstander som har vist seg å brenne før, derfor antar man at det kan begynne å brenne i fremtiden. Man unngår altså å kaste glør fra peisen i restavfallet nettopp fordi man antar at det kan ende katastrofalt ved at huset brenner ned. Man kan ikke si det for sikkert, men man tar ikke sjansen basert på tidligere erfaringer. Det ville blitt kaos om verden ikke skulle legge uniformitetsprinsippet til grunn for å veilede oss i antagelser om fremtiden. Men det kan ikke rasjonelt begrunnes som en sannhet, slik Hume har foreslått.

Å forsvare uniformsprinsippet med sannsynlighet og vitenskapelig suksess vil i følge Hume være å flytte problemet om induksjon, ikke løse det. Han mener at man ikke kan forsvare induksjon ved bruk av sannsynlighet. Dette fordi man for å sannsynliggjøre mønstre om fremtiden må anta at uniformitetsprinsippet ligger til grunn. Vitenskapelig suksess som har vist seg å forutsi og forklare fenomener rettferdiggjør heller ikke induksjon. Hume mener dette kan forklares ved å se på menneskets psykologi (Vassend, 2024a).

## **6.0.2 Gi en kort beskrivelse av falsifikasjonisme og si litt om hvorfor Karl Popper var motivert til å utvikle denne teorien. Presenter så ett problem med teorien og vurder hvorvidt problemet kan løses.**

### **6.0.2.1 Karl Popper - falsifikasjonismen**

Popper var av den oppfatning at Humes induksjonsproblemet ikke var til å løses (Vassend, 2024b). Popper mente induksjon var en fiasko og at vitenskapen ikke kunne verifisere noe som helst, bare falsifisere [Døderlein2007]. Han mente faktisk at bekreftelsen av hypoteser var et kjennetegn på pseudovitenskap!

Det kan tenkes at Karl Popper med sin uenighet i hva som ble sett på som vitenskap og sine tanker om induksjon, og i sin iver etter å skille vitenskapelige fra ikke-vitenskapelige utsagn, ble motivert til å utvikle sin teori om falsifikasjonisme. Popper, med utgangspunkt i sin teori om falsifiserbarhet, mente at ingen hypotese kan vise seg å være sann. En hypotese kan unnlates å bli forkastet, men ikke fordi den er sann eller usann, men fordi den ikke er blitt falsifisert. Den er altså ikke blitt bevist usann. En hypotese som er falsifisert kan derimot ikke beholdes da denne er bevist usann. Falsifiserbarheten av teorier og utsagn kan også graderes avhengig av hvor presise de er, og hvor mange prediksjoner de inneholder. Teori 1: ” vann koker alltid ved 100 grader celsius ved havnivå” er mer falsifiserbar enn teori 2: ” vann koker ved 100 grader”. Teori 1 har flere detaljer enn teori 2 som kan testes, noe som øker falsifiserbarheten i teori 1 versus teori 2 (Vassend, 2024b).

Pierre Duhem sitt logiske poeng fremhever en utfordring med falsifikasjon. Poenget til Duhem er at vitenskaplige hypoteser sjelden blir testet isolert, eller uten støttehypoteser. Han mener at når et resultat ikke stemmer overens med prediksjonene, er det ikke nødvendigvis hypotesen som er problemet, men muligens andre variabler, eller observasjonen. Dette kan være variabler som underliggende antagelser om hypotesen, feil på måleinstrument, eller til og med oppdagede variabler som påvirker resultatene. Dette betyr at en falsifisering ofte ikke entydig peker på hvilken del av eksperimentet som må forkastes (Vassend, 2024b). Duhem poengterer dermed at vitenskaplige testing er mer komplisert og en del av et større nettverk av teorier, noe som utfordrer Poppers idè om at en enkelt hypotese kan falsifiseres isolert. For å forkaste en hypotese kan det derfor kreve grundige revurderinger av tilhørende teorier og hjelpemodeller, og ikke en enkel slutning om at hypotesen er usann.

Poppers falsifikasjonsteori sier at hvis en hypotese er falsifisert så er den usann. Den kan derfor ikke være sann. En utfordring med denne teorien er at man for å hevde at noe er usant også må si at noe er sant. Det krever altså bekreftelse for å falsifisere. Dette blir vel en form for induktiv logikk med premiss om uniformitet. Man regner med at denne usanne hypotesen vil fortsette å være usann i fremtiden. Så da er kanskje ikke premisset om uniformitet så dumt likevel. Naturens lover viser seg å være konsekvente nok. Resultater som kommer fra teknologisk fremgang de siste x-antall årene vitner om nettopp gode nok metoder.

# 7 Labrapport

## 7.1 Introduksjon

Analyse av genuttrykk gjennom fluorescensbasert sanntids kvantitativ polymerasekjedereaksjon (qPCR) er en velkjent metode som brukes i en rekke treningsstudier (Kuang & Bishop, 2018). Denne metoden kvantifiserer uttrykket av målgener i biologiske prøver, som for eksempel fra blod eller muskelvev. qPCR-analyse er bredt anvendt, og det finnes en rekke protokoller og metoder for å gjennomføre analysen. I vår studie brukte vi SYBR Green-metoden, der et fluorescerende fargestoff binder seg til DNA under amplifikasjon. Denne fluorescensen gjør det mulig å følge reaksjonen i sanntid, ettersom signalintensiteten korrelerer med mengden DNA som er til stede (Kuang & Bishop, 2018). Fluorescensen når en forhåndsbestemt terskel, kjent som syklusterskelen (CT), som reflekterer nivået av genuttrykk. En lavere CT-verdi indikerer høyere genuttrykk, siden færre amplifikasjonssykluser er nødvendig for å nå terskelen (Livak & Schmittgen, 2001).

Innen treningsfysiologi brukes qPCR ofte for å undersøke endringer i genuttrykk som oppstår ved trening. I denne studien har vi som mål å teste primer-effektivitet og utføre målrettet amplifikasjon av cDNA ved bruk av spesifikke primere. I tillegg ønsker vi å undersøke hvordan styrketrening påvirker genuttrykket i ulike muskelfibertyper hos deltakere som gjennomførte et to ukers styrketreningsprogram. De ulike muskelfibertypene vi undersøkte var type I, type IIa og type IIx. Type I-fibre er kjent som langsomme muskelfibre og er svært oksidative og kan opprettholde kontraksjoner over lange perioder uten å bli slitne, mens type IIa- og type IIx-fibre har dårligere utholdende egenskaper, men kan produsere mer kraft over en kortere tidsperiode (Pette & Staron, 2000).

## 7.2 Metode

### 7.2.1 Materialer

- En sanntids PCR-maskin (QuantStudio 5)
- En qPCR-reaksjonsplate
- Nukleasefritt vann og pipettespisser
- SYBR-green Master Mix

Før eksperimentet forberedte laboratorielederen cDNA, som ble ekstrahert fra prøver samlet inn i en studie der deltakerne gjennomgikk et to ukers styrketreningsprogram.

Først opprettet vi en fortynningsserie for å teste primerne (Table 7.1). Vi flyttet 2 µl fra rør 1 til rør 2a, og 10 µl fra rør 1 til 2b, deretter vortexet vi rør 2a og 2b for å blande prøven og vannet (H<sub>2</sub>O). Neste steg var å flytte 2 µl fra 2a til 3a og 2 µl fra 2b til 3b, og deretter vortexe rør 3a og 3b. Til slutt flyttet vi 2 µl fra 3a til 4a og 2 µl fra 3b til 4b og vortexet rør 4a og 4b.

Tabell 7.1: Fortynningsserie

	1	2a	3a	4a	2b	3b	4b
Fortynning	1	1/10	1/100	1/1000	1/2	1/20	1/200
prøve	30µl	2µl	2µl	2µl	10µl	2µl	2µl
H <sub>2</sub> O	0µl	18µl	18µl	18µl	10µl	18µl	18µl

Vi laget deretter en master mix som besto av 250 µl SYBR-green, 50 µl primermikstur (MHC1, MHC2a, MHC2x, eller MHCb2m), og 100 µl H<sub>2</sub>O. Deretter fylte vi platen med primerspesifikk master mix i henhold til pipetteringsskjemaet vårt (Table 7.2). 8 µl av master mix ble tilsatt i brønnene sammen med 2 µl cDNA-prøve.

Tabell 7.2: Pipetteringsskjema

row	13	14	15	16	17	18	19	20	21	22	23
	Fp1	Fp2									
A	myhc 1	myhc 1			cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b
B	myhc 1	myhc 1			cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b
C	myhc 1	myhc 1			cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b
D	myhc 2a	myhc 2a									
E	myhc 2a	myhc 2a									
F	myhc 2a	myhc 2a									
G	myhc 2x	myhc 2x									
H	myhc 2x	myhc 2x									

I	myhc 2x	myhc 2x								
J	myhc	myhc	cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b	
K	myhc	myhc	cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b	
L	myhc	myhc	cmyc 1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b	

Platen ble deretter dekket med plast og sentrifugert ved 1200 o/min i 1 minutt. PCR-prøvene ble analysert ved hjelp av real-time PCR (Applied Biosystems 7500 Fast Real-Time PCR System, Life Technologies AS) og QuantStudio-programvare (Applied Biosystems, Waltham, MA, USA). PCR-prosessen besto av tre stadier: en “Hold-stadie,” en “PCR-stadie” og en “Melt curve-stadie.”

I det første trinnet økte temperaturen med 1,99 °C/s opp til 50 °C, hvor den ble holdt konstant i 2 minutter. Deretter økte temperaturen med 1,99 °C/s opp til 95 °C, hvor den forble konstant i 2 minutter. Deretter startet selve PCR-prosessen, som besto av 40 sykluser. Én syklus innebar 1 sekund ved 95 °C, etterfulgt av at temperaturen sank med en hastighet på 1,77 °C/s til 60 °C, hvor den ble holdt konstant i 30 sekunder. Etter hver syklus ble et bilde av fluorescensen i brønnene tatt.

Til slutt økte temperaturen med 1,99 °C/s opp til 95 °C og ble holdt konstant i 15 sekunder. Deretter sank temperaturen med en hastighet på 1,77 °C/s til 60 °C, hvor den ble holdt konstant i 1 minutt. Temperaturen økte så med en hastighet på 0,15 °C/s opp til 95 °C, hvor den ble holdt konstant i 15 sekunder. Etter at PCR-prosessen var fullført, hentet vi ut resultatene i form av CT-verdier.

## 7.2.2 Statistisk analyse

Analysen av dataen ble gjort i Excel (Microsoft Corporation, 2018) . Her beregnet vi gjennomsnittlig CT-verdi for triplikatene, samt base-10-logaritmen av prøvemengden. Deretter ble stigningen beregnet ved å plote log(prøvemengde) mot gjennomsnittlig CT-verdi. En lineær regresjon ble utført for å finne den best tilpassede linjen. Primer-effektiviteten ble deretter beregnet ut fra stigningen. Til slutt beregnet vi standardavviket (SD) og variasjonskoeffisienten (CV) mellom replikatene.

## 7.3 Resultater

### 7.3.1 Primereffektivitet

Tabell 7.3: Primereffektivitet

Dilution	Ct1	Ct2	Ct3	Avg.Ct	Sample.quan	Log...sample.quan.	Slope	Primer.Efficiency....
1	28.678	28.708	29.155	28.847	1.000	0.000	-2.6104	141.5901
1/2	29.414	29.62	29.264	29.433	0.500	-0.301		

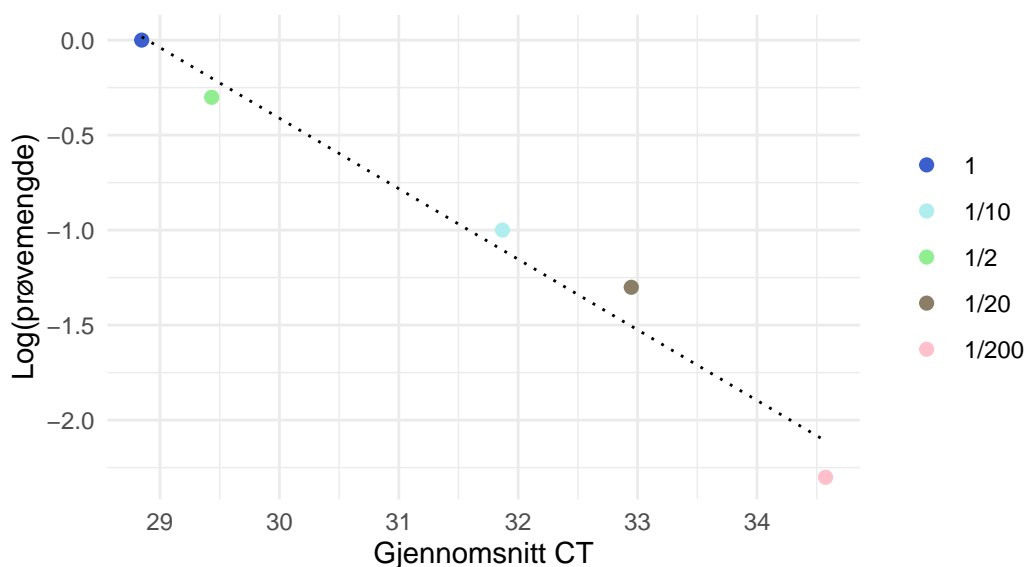
1/10	31.776	31.416	32.413	31.868	0.100	-1.000
1/20	33.241	32.653	Undetermined	32.947	0.050	-1.301
1/200	Undetermined	Undetermined	34.574	34.574	0.005	-2.301

En stigning på -2,6 indikerer at primereffektiviteten ikke er optimal, beregnet til å være 142 %. Observasjonen er ikke lineær (Table 7.3).

Tabell 7.4: Gjennomsnitt og variasjon

Samlet.prøve	Avg	SD	CV
1.0	28.847	0.2670490	0.93 %
0.5	29.433	0.1788764	0.61 %
0.1	31.868	0.5050010	1.58 %

De lave verdiene for standardavvik (SD) indikerer at målingene for de ulike fortynningene i den “samlede prøven” er konsistente og presise. Variasjonskoeffisienten (CV) viser at den 1/2-fortynnede prøven har den laveste variasjonen (0,61 %), mens den 1/10-fortynnede prøven har den høyeste variasjonen (1,58 %). Dette tyder på at målingene blir mindre presise ved høyere fortynning, se Table 7.4.



Figuren viser et lineært forholde mellom fortynning av prøver og syklusterskel. Fargede punkter er prøvenes fortynningsgrad. Svart stipled linje er en lineær regresjon tilpasset dataen.

## 7.4 Myosin Heavy Chain

I Table 7.5 ser vi at antall sykluser for å nå syklusterskelen (CT) har endret seg fra uke 0 til uke 2. For myhc1 gikk gjennomsnittet fra 19,7 til 19,3 sykluser. For myhc2a gikk det fra 20,9 til 19,4, og for myhc2x gikk det fra 26,6 til 24,4 sykluser. Dette indikerer at genuttrykket økte fra uke 0 til uke 2, da færre sykluser indikerer høyere genuttrykk (Livak & Schmittgen, 2001).

Tabell 7.5: CT-verdier

ID	Target_name	ct1	ct2	ct3	Snitt	Ref.gen	Ref.ct1	Ref.ct2	Ref.ct3	Ref.snitt	Delta.ct	Two.delta.delta.ct
FP6 R w0	myhc1	19.798	19.901	19.621	19.77333	b2m	24.670	24.513	24.691	24.625	-4.851	28.867
FP6 R w2pre	myhc1	18.944	19.240	19.861	19.34833	b2m	22.913	23.950	23.819	23.561	-4.212	18.536
FP6 R w0	mhc2a	21.029	21.247	20.627	20.96800	b2m	24.670	24.513	24.691	24.625	-3.657	12.613
FP6 R w2pre	mhc2a	19.549	19.304	19.580	19.47800	b2m	22.913	23.950	23.819	23.561	-4.083	16.944
FP6 R w0	myhc2x	27.019	26.898	25.907	26.60795	b2m	24.670	24.513	24.691	24.625	1.983	0.253
FP6 R w2pre	myhc2x	24.871	24.105	24.256	24.41062	b2m	22.913	23.950	23.819	23.561	0.850	0.555

I Table 7.6 ser vi at mengden genuttrykk for ulike muskelfibertyper har endret seg fra uke 0 til uke 2. Uttrykket av Muskelfibertype 1 (myhc1) sank fra 69 % til 51 %, mens Muskelfibertype 2a (myhc2a) økte fra 30 % til 47 %. Muskelfibertype 2x (myhc2x) økte fra 0,6 % til 1,5 %.

Tabell 7.6: Prosentvis fordeling av genuttrykk etter uke 0 og uke 2

100%	myhc1	myhc2a	myhc2x
41.733	69.17 %	30.22 %	0.61 %
36.034	51.44 %	47.02 %	1.54 %

## 7.5 Diskusjon

I en studie (Wilborn & Willoughby, 2004) fant man at etter 8 uker med tung styrketrening var genuttrykket for myhc 1 og 2a økt, men noe mindre for myhc 2x. Det er ikke studier på to uker med styrketrening som forklarer endringen i myhc genuttrykk, men vi spekulerer i at det vil være samme endringer som for (Wilborn & Willoughby, 2004), dog i mindre grad. Dette stemmer med våre funn da vi ser en økning i genuttrykk for både myhc 1 og myhc 2a, men vi har også motstridene resultat. Vi ser en økning i myhc 2x.

Økningen i myhc 2x, som er forbundet med de raskeste og mest eksplosive muskelfibrene, er uventet. Denne fibertypen reduseres vanligvis ved langvarig trening til fordel for mer utholdenhets- eller styrkeorienterte fibre (type I og IIa) (Pette & Staron, 2000). Det kan være at denne økningen er et resultat av tekniske feil, som for eksempel høy primer-effektivitet eller biologisk variasjon.



Primer effektiviteten er ikke optimal da den er kalkulert til 142%. Referanseverdiene som er på 90-110% indikerer at hver PCR syklus doubler mengden mål-DNA. Vår høye verdi tyder på ulike feilkilder underveis i forsøket som feil pipettering eller kontaminasjon, primere kan binde seg til flere plasser på DNA og primere kan binde seg til hverandre noe som kan gi falske resultat.

CV målingene forteller at det er høyere variabilitet i målingene når prøven blir mer fortynnet. Gjennomsnittet er lavest på den ufortynnede prøven noe som kan påvirke at CV er høyere enn den 1/2 fortynnede prøven da CV regnes ut fra standardavviket delt på gjennomsnittet som gjør at et lavere gjennomsnitt vil gi en høyere CV. Vi mistenker at dårlig pipettering er årsaken til en høyere CV for ufortynnet prøve enn 1/2 fortynnet prøve.

# Kilder

American College of Sports Medicine (2009). American college of sports medicine position stand. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* **41**, 687–708.

Anon. Login - eLabFTW — elab.inn.no.

Atkinson G & Nevill AM (2001). Selected issues in the design and analysis of sport performance research. *J Sports Sci* **19**, 811–827.

Bacon AP, Carter RE, Ogle EA & Joyner MJ (2013). VO2max trainability and high intensity interval training in humans: A meta-analysis. *PLoS One* **8**, e73182.

Bouchard C, Blair SN & Haskell WL (2012). Physical activity and health ed. Bouchard C, Blair SN & Haskell WL.

Breil FA, Weber SN, Koller S, Hoppeler H & Vogt M (2010). Block training periodization in alpine skiing: Effects of 11-day HIT on VO2max and performance. *Eur J Appl Physiol* **109**, 1077–1086.

Brosowsky N, Parshina O, Locicero A & Crump MJC (2020). Teaching undergraduate students to read empirical articles: An evaluation and revision of the QALMRI method. *PsyArXiv*.

Cannon J & Marino FE (2010). Early-phase neuromuscular adaptations to high- and low-volume resistance training in untrained young and older women. *J Sports Sci* **28**, 1505–1514.

Cohen J (1988). *Statistical power analysis for the behavioral sciences*, 2nd edn. Routledge Member of the Taylor; Francis Group, New York, NY.

Creswell JW (2008). Research design.

Currier BS, Mcleod JC, Banfield L, Beyene J, Welton NJ, D’Souza AC, Keogh JAJ, Lin L, Coletta G, Yang A, Colenso-Semple L, Lau KJ, Verboom A & Phillips SM (2023). Resistance training prescription for muscle strength and hypertrophy in healthy adults: A systematic review and bayesian network meta-analysis. *Br J Sports Med* **57**, 1211–1220.

- García-Pallarés J, García-Fernández M, Sánchez-Medina L & Izquierdo M (2010). Performance changes in world-class kayakers following two different training periodization models. *Eur J Appl Physiol* **110**, 99–107.
- Halperin I, Pyne DB & Martin DT (2015). Threats to internal validity in exercise science: A review of overlooked confounding variables. *Int J Sports Physiol Perform* **10**, 823–829.
- Hariton E & Locascio JJ (2018). Randomised controlled trials – the gold standard for effectiveness research. *BJOG* **125**, 1716–1716.
- Haun CT, Vann CG, Mobley CB, Osburn SC, Mumford PW, Roberson PA, Romero MA, Fox CD, Parry HA, Kavazis AN, Moon JR, Young KC & Roberts MD (2019). Pre-training skeletal muscle fiber size and predominant fiber type best predict hypertrophic responses to 6 weeks of resistance training in previously trained young men. *Front Physiol* **10**, 297.
- Haun CT, Vann CG, Mobley CB, Roberson PA, Osburn SC, Holmes HM, Mumford PM, Romero MA, Young KC, Moon JR, Gladden LB, Arnold RD, Israetel MA, Kirby AN & Roberts MD (2018). Effects of graded whey supplementation during extreme-volume resistance training. *Front Nutr* **5**, 84.
- Hopkins WG (2000a). Measures of reliability in sports medicine and science. *Sports Med* **30**, 1–15.
- Hopkins WG (2000b). Measures of reliability in sports medicine and science. *Sports Med* **30**, 1–15.
- Hulley SB, Cummings SR, Browner WS, Grady DG & Newman TB (2006). *Designing clinical research*, 3rd edn. Lippincott Williams; Wilkins, Philadelphia, PA.
- Knight KL (2010). Study/experimental/research design: Much more than statistics. *J Athl Train* **45**, 98–100.
- Krieger JW (2009). Single versus multiple sets of resistance exercise: A meta-regression. *J Strength Cond Res* **23**, 1890–1901.
- Kuang XY Jujiao & Bishop DJ (2018). Evaluation of iron status in non-dialysis chronic kidney disease patients. *Journal of Diabetic Association Medical College, Faridpur* **7**, 22–26.
- Livak KJ & Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* **25**, 402–408.
- Microsoft Corporation (2018). Microsoft excel. Available at: <https://office.microsoft.com/excel>.

- Mitchell CJ, Churchward-Venne TA, West DWD, Burd NA, Breen L, Baker SK & Phillips SM (2012). Resistance exercise load does not determine training-mediated hypertrophic gains in young men. *J Appl Physiol* **113**, 71–77.
- Newell J, Higgins D, Madden N, Cruickshank J, Einbeck J, McMillan K & McDonald R (2007). Software for calculating blood lactate endurance markers. *J Sports Sci* **25**, 1403–1409.
- Pette D & Staron RS (2000). Myosin isoforms, muscle fiber types, and transitions. *Microsc Res Tech* **50**, 500–509.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>.
- Rønnestad BR, Ellefsen S, Nygaard H, Zacharoff EE, Vikmoen O, Hansen J & Hallén J (2014a). Effects of 12 weeks of block periodization on performance and performance indices in well-trained cyclists. *Scand J Med Sci Sports* **24**, 327–335.
- Rønnestad BR, Hansen J & Ellefsen S (2014b). Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists. *Scand J Med Sci Sports* **24**, 34–42.
- Rønnestad BR, Hansen J, Thyli V, Bakken TA & Sandbakk Ø (2016). 5-week block periodization increases aerobic power in elite cross-country skiers. *Scand J Med Sci Sports* **26**, 140–146.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P & Cardona A (2012). Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682.
- Schoenfeld BJ, Ogborn D & Krieger JW (2017). Dose-response relationship between weekly resistance training volume and increases in muscle mass: A systematic review and meta-analysis. *J Sports Sci* **35**, 1073–1082.
- Senn SS (2002a). Cross-over trials in clinical research.
- Senn SS (2002b). *Cross-over trials in clinical research*, 2nd edn. John Wiley & Sons, Chichester, England.
- Spiegelhalter D (2019). The art of statistics.
- Sylta Ø, Tønnessen E, Hammarström D, Danielsen J, Skovereng K, Ravn T, Rønnestad BR, Sandbakk Ø & Seiler S (2016). The effect of different high-intensity periodization models

on endurance adaptations. *Med Sci Sports Exerc* **48**, 2165–2174.

Vassend O (2024a). KvantMet:vitenskapsfilosofi dag 1. Available at: [https://inn.instructure.com/courses/18083/files/2144313?module\\_item\\_id=629779](https://inn.instructure.com/courses/18083/files/2144313?module_item_id=629779).

Vassend O (2024b). KvantMet:vitenskapsfilosofi dag 2. Available at: [https://inn.instructure.com/courses/18083/files/2145719?module\\_item\\_id=630133](https://inn.instructure.com/courses/18083/files/2145719?module_item_id=630133).

Wilborn CD & Willoughby DS (2004). The role of dietary protein intake and resistance training on myosin heavy chain expression. *J Int Soc Sports Nutr* **1**, 27–34.